8-2021

Efficient estimation of integrated volatility functionals under general volatility dynamics

Jia LI
Singapore Management University, jiali@smu.edu.sg

Yunxiao Liu
University of North Carolina

Published in Econometric Theory, Volume 37, Issue 4, August 2021, pp. 664-707
DOI: 10.1017/S0266466620000274

EFFICIENT ESTIMATION OF INTEGRATED VOLATILITY FUNCTIONALS UNDER GENERAL VOLATILITY DYNAMICS*

JIA LI
Duke University
YUNXIAO LIU
University of North Carolina at Chapel Hill

We provide an asymptotic theory for the estimation of a general class of smooth nonlinear integrated volatility functionals. Such functionals are broadly useful for measuring financial risk and estimating economic models using high-frequency transaction data. The theory is valid under general volatility dynamics, which accommodates both Itô semimartingales (e.g., jump-diffusions) and long-memory processes (e.g., fractional Brownian motions). We establish the semiparametric efficiency bound under a nonstandard nonergodic setting with infill asymptotics, and show that the proposed estimator attains this efficiency bound. These results on efficient estimation are further extended to a setting with irregularly sampled data.

1. INTRODUCTION

A central problem in financial econometrics is the estimation of asset price volatility (Engle, 2004), which is greatly facilitated by using high-frequency data (Merton, 1980). A large literature has been devoted to estimating integrated volatility functionals using such data. Earlier work focused on the integrated covariance matrix (see Andersen and Bollerslev, 1998; Andersen et al., 2001; Andersen et al., 2003; Barndorff-Nielsen and Shephard, 2001, 2002a, 2002b, 2004a; Mancini, 2001, among others). The recent literature has shown great interest in the estimation of general nonlinear volatility functionals (see, e.g., Kristensen, 2010; Jacod and Rosenbaum, 2013; Renault, Sarisoy, and Werker, 2017) for two broad reasons: these functionals not only provide a battery of ex post risk measures (Mykland and Zhang, 2006, 2009; Todorov and Tauchen, 2012; Aït-Sahalia and Xiu, 2019; Kalnina and Xiu, 2017), but also play the role of moment conditions that are analogous to the classical Generalized Method of Moments

*We are grateful for comments from a co-editor and two anonymous referees, which have greatly improved the paper. We also thank Tim Bollerslev, Peter Hansen, Andrew Patton, Vladas Pipiras, George Tauchen, and Jean Jacod for helpful comments. J.L.'s research was partially supported by NSF Grant SES-1326819. Address correspondence to Jia Li, Department of Economics, Duke University, Durham, NC 27708, USA; e-mail: jl410@duke.edu.

(GMM) for estimating economic models (Li, Todorov, and Tauchen, 2016; Li and Xiu, 2016).

Regarding the basic problem with integrated covariance, the inference theory has been developed for essentially unrestricted volatility dynamics. [1] However, the analysis is notably more complicated, and less well understood, for general nonlinear volatility functionals. The early contributions of Barndorff-Nielsen and Shephard (2003) and Jacod (2008) proposed estimators that are formed as sums of nonlinear transforms of normalized returns. This approach implicitly imposes tight restrictions on the type of nonlinear functionals to be estimated, and the resulting estimator is not efficient. A more general approach is to first nonparametrically recover the spot covariance process (Foster and Nelson, 1996; Comte and Renault, 1998), and then construct a plug-in estimator for the integrated volatility functional. Along this line, Kristensen (2010) first derived the asymptotic distribution of a plug-in estimator under the assumption that the volatility process has differentiable paths; this early contribution thus, unfortunately, ruled out typical stochastic volatility models.[2] Jacod and Rosenbaum (2013) considered a more general case in which the spot volatility process is an Itô semimartingale.[3] They showed that the "raw" plug-in estimator carries a non-negligible high-order bias and proposed a feasible bias correction. The bias-corrected estimator is asymptotically mixed normal and attains the semiparametric efficiency bound derived in some specific settings considered by Clément, Delattre, and Gloter (2013) and Renault et al. (2017).

The aforementioned prior work, however, routinely imposes a strong restriction on the volatility dynamics, namely, the volatility process is assumed to be an Itô semimartingale; see, in particular, Jacod (2008) and Jacod and Rosenbaum (2013). Unlike the price process, which is necessarily a semimartingale under no-arbitrage (Delbaen and Schachermayer, 1994; Harrison and Kreps, 1979), economic theory is silent about whether the volatility should be a semimartingale. As a matter of fact, the semimartingale assumption on volatility rules out long-memory type dynamics. This restriction is thus undesirable from an econometric point of view, given the long tradition in time-series econometrics for studying long-memory dynamics (Granger, 1980; Geweke and Porter-Hudak, 1983). Furthermore, long-memory features in asset price volatility have been documented and analyzed by Ding, Granger, and Engle (1993), Baillie, Bollerslev, and Mikkelsen (1996), Andersen and Bollerslev (1997), Comte and Renault (1998), and Andersen et al. (2001) in the early literature; also see the recent analysis of McCloskey and Perron (2013) based on a robust method. As demonstrated theoretically by Granger (1980) and Comte and Renault (1996), long-memory properties arise naturally

[1] See, for example, Thm. 5.4.2 in Jacod and Protter (2012).

[2] As is standard in classical nonparametric analysis, the differentiability assumption allows one to use high-order kernels to reduce the asymptotic magnitude of bias due to the temporal (and stochastic) variation of volatility.

[3] Li, Todorov, and Tauchen (2017) extended the theory of Jacod and Rosenbaum (2013) to a more general class of functionals.

from the aggregation of heterogeneous micro series. In the same vein, Andersen and Bollerslev (1997) proposed a model with heterogeneous information arrivals, which was used to explain long-memory in asset price volatility.

Set against this background, we study the estimation of nonlinear volatility functionals in a general setting where the volatility contains (possibly) two additive components: one component is a general Itô semimartingale and the other exhibits long-memory. We refer to this model as the *long-memory Itô semimartingale (LMIS)* volatility model. Like the standard Itô semimartingale model, the LMIS model includes both a continuous local martingale component driven by a Brownian motion and a discontinuous component with arbitrary jump activity. With the long-memory component, the LMIS model also accommodates fractional Brownian motions (fBM) and, more generally, stochastic (Wiener and/or Skorokhod) integrals driven by fBMs and other multifractal processes used in financial risk management.

The main contribution of this paper is to develop an estimation theory for nonlinear integrated volatility functionals under the LMIS model, and to study the associated semiparametric efficiency. We first derive the asymptotic property of the estimator of Jacod and Rosenbaum (2013) in the general LMIS setting. Compared to prior work, the notable challenge here is that the volatility increment is no longer (only) dominated by the martingale-difference component, which thus requires a quite different (and more complicated) proof. Second, we derive the efficiency bound for the LMIS volatility model and show that the Jacod–Rosenbaum estimator attains this bound, and hence, is semiparametrically efficient for a large class of volatility models that is markedly more general than previously considered. Since our setting does not satisfy the local asymptotic normality (LAN) property, we cannot directly follow classical work on efficient estimation (see, e.g., Bickel et al., 1998). Instead, we implement Stein's insight (Stein, 1956) by constructing a class of parametric submodels. We show that the submodels satisfy the local asymptotic mixed normality (LAMN) property and then use the conditional convolution theorem (Jeganathan, 1982, 1983) to characterize the worst-case efficiency bound. Finally, the general version of our theoretical results are actually derived in the case when the high-frequency data are irregularly sampled. As a matter of fact, we propose a modification of the Jacod–Rosenbaum estimator so as to accommodate irregular sampling, which sets the current paper further apart from existing econometric work on general integrated volatility functionals.

Our result on the efficiency bound is novel and generalizes the recent work of Clément et al. (2013) and Renault et al. (2017) in important dimensions. Clément et al. (2013) considered a conditional Markov setting in which the volatility process is a continuous Itô process, and used a Malliavin calculus approach to derive the LAMN property. Their setting is restrictive since it excludes both long-memory dynamics and volatility jumps. Renault et al. (2017) used an alternative approach and studied a conditional LAN setting that allows for finite-variational volatility jumps and long-memory dynamics. Our approach is more similar to Renault et al.

(2017) in spirit, but we establish semiparametric efficiency bound that is explicitly applicable (rather than under high-level conditions) for the general LMIS volatility models under possibly irregular sampling.[4]

To the best of our knowledge, the current paper is the first one that provides an asymptotic theory for estimating nonlinear volatility functionals under the general LMIS volatility dynamics; moreover, we also address the issue of semiparametric efficiency in this general setting. Since most (if not all) volatility models in economics and finance are special cases of LMIS, the current paper provides a more complete understanding for the efficient estimation and inference of nonlinear volatility functionals.

This paper is organized as follows. We present the setting in Section 2. The baseline theory under regular sampling is presented in Section 3, and the extension to the case with irregular sampling is in Section 4. Section 5 presents simulation results. Section 6 concludes. All proofs are in the Appendix.

2. THE SETTING

2.1. Integrated Volatility Functionals and the LMIS Volatility Model

We start with describing the setting. Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ be a complete filtered probability space. All processes are assumed to be càdlàg and adapted. The d-dimensional (logarithmic) price process X is a semimartingale with the form

$$X_t = \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \tag{2.1}$$

where b is the drift process taking values in $\mathbb{R}^d$, σ is the stochastic co-volatility matrix process taking values in $\mathbb{R}^{d\times d}$, W is a d-dimensional Brownian motion, and J is a jump process. Let $c_t \equiv \sigma_t \sigma_t^\top$ denote the *spot covariance matrix* process, which takes values in $\mathcal{M}_d$, the space of $d \times d$ positive semidefinite matrices.[5]

In our baseline setting studied in Section 3, the data consist of high-frequency observations of X that are sampled on a regular time grid $\{i\Delta_n : 0 \leq i \leq [T/\Delta_n]\}$ for a fixed time span $[0,T]$, where the sampling interval $\Delta_n \to 0$ asymptotically as $n \to \infty$. Below, we denote the ith return of X by

$$\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}.$$

We use this baseline setting to first describe results under the LMIS volatility dynamics. An extension with irregular sampling is further developed in Section 4. We need some standard regularity conditions on X.

[4] We also note that Renault et al. (2017) only propose "nearly" efficient estimators. The notion of near efficiency is formalized via a sequential asymptotic embedding, in which the sample size and the bandwidth do not change simultaneously, but only sequentially. In contrast, our asymptotic theory is derived in a joint asymptotic setting in which the bandwidth changes together with sample size.

[5] The process of interest is c_t instead of σ_t. We do not assume the latter to be positive semidefinite. More generally, we may allow W to be a d'-dimensional Brownian motion, with σ_t taking values in the space of $d \times d'$ matrices for some $d' > d$.

Assumption H. Let $r \in [0,1)$ be a constant. The process X is an Itô semimartingale given by (2.1) with $J_t(\omega) = \int_0^t \int_{\mathbb{R}} \delta(\omega, s, z)\, \mu(ds, dz)$, where μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with compensator $\nu(dt, dz) = dt \otimes \lambda(dz)$ for some σ-finite measure $\lambda(\cdot)$ on $\mathbb{R}$, and $\delta : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}^d$ is a predictable function. There are a sequence of non-negative bounded λ-integrable functions $\mathcal{J}_m(\cdot)$ on $\mathbb{R}$ and a sequence of stopping times $(T_m)_{m\geq 1}$ increasing to ∞ , such that $\|\delta(\omega, t, z)\|^r \wedge 1 \leq \mathcal{J}_m(z)$ for all (ω, t, z) with $t \leq T_m(\omega)$. □

We are interested in estimating integrated volatility functionals of the form

$$S(g) \equiv \int_0^T g(c_s)\, ds, \tag{2.2}$$

for some three-time continuously differentiable transform $g : \mathcal{M}_d \mapsto \mathbb{R}$.[6] Many ex post risk measures of volatility can be written either as $S(g)$ or as a further nonlinear transform of such quantities, such as the integrated covariance matrix (Barndorff-Nielsen and Shephard, 2004a), diffusive return beta (Mykland and Zhang, 2009), correlation/leverage effect (Kalnina and Xiu, 2017), idiosyncratic variance (Mykland and Zhang, 2006), volatility Laplace transform (Todorov and Tauchen, 2012), variance betas (Li et al., 2016), eigenvalues (Aït-Sahalia and Xiu, 2019), among others. Moreover, general forms of $S(g)$ also serve as integrated moment conditions in specification tests and estimation problems in economic models (Li and Xiu, 2016).

Jacod and Rosenbaum (2013) propose an efficient estimator for integrated volatility functionals of the form (2.2). However, their asymptotic theory requires that the function $g(\cdot)$ and its derivatives have polynomial growth (see their Thm. 3.1), which can be restrictive for many applications. Li et al. (2017) and Li and Xiu (2016) provided a spatial localization argument to relax the polynomial growth condition, with the help of a mild sample path regularity on the spot covariance process stated in the following assumption. Below, for a compact set $\mathcal{K} \subset \mathcal{M}_d$ and a constant $\eta > 0$, we denote the "η-enlargement" of $\mathcal{K}$ by

$$\mathcal{K}^\eta \equiv \{M \in \mathcal{M}_d : \inf_{A \in \mathcal{K}} ||M - A|| < \eta\}.$$

Assumption C. There exist a localizing sequence of stopping times $(T_m)_{m\geq 1}$ and a sequence of convex compact subsets $\mathcal{K}_m \subseteq \mathcal{M}_d$ such that $c_t \in \mathcal{K}_m$ for $t \leq T_m$ and $g \in \mathcal{C}^3(\mathcal{K}_m^\eta)$, with the latter denoting the space of three-time continuously differentiable functions on $\mathcal{K}_m^\eta$ for some $\eta > 0$. [7] □

---

[6] We assume that $g(\cdot)$ is scalar-valued only for notational simplicity, because all results can be extended trivially for vector-valued $g(\cdot)$ by using the Cramér–Wold device.

[7] To illustrate how Assumption C allows us to relax the polynomial growth condition, consider an example with $g(x) = \log(x)$ and $d = 1$. In this case, $g(\cdot)$ does not have polynomial growth in that the log function is explosive near 0. On the other hand, if c_t and $1/c_t$ are locally bounded, then there exists a sequence of stopping times $(T_m)_{m\geq 1}$ increasing to infinity, such that $c_t \in [1/m, m]$ for $t \leq T_m$. Assumption C is satisfied for $\mathcal{K}_m = [1/m, m]$, because the log function is $\mathcal{C}^3$ on $(1/m - \eta, m + \eta)$ for η sufficiently small. See Sect. 4.4.1 of Jacod and Protter (2012) for additional details on localization.

As discussed in the Introduction, we aim to study the efficient estimation of $S(g)$ in a setting that accommodates both Itô semimartingale and long-memory type dynamics in the volatility. We consider the LMIS model given as follows

$$\sigma_t \equiv \sigma_{1,t} + \sigma_{2,t}, \tag{2.3}$$

where $\sigma_{1,t}$ and $\sigma_{2,t}$ are, respectively, the Itô semimartingale and the long-memory components that satisfy the following assumptions.

Assumption IS. The process $\left(\sigma_{1,t}\right)_{t\geq 0}$ is an Itô semimartingale of the form

$$\sigma_{1,t} = \sigma_{1,0} + \int_0^t b_s^{(\sigma_1)} ds + \int_0^t \sigma_s^{(\sigma_1)} dW_s + M_t^{(\sigma_1)} + \int_0^t \int_{\mathbb{R}} \delta^{(\sigma_1)}(s,z) 1_{\{||\delta^{(\sigma_1)}(s,z)||>1\}} \mu(ds,dz),$$

where $b^{(\sigma_1)}$ is $(d\times d)$-dimensional; $\sigma^{(\sigma_1)}$ is $(d\times d\times d)$-dimensional; $M^{(\sigma_1)}$ is a local martingale that is orthogonal to W with bounded jumps and its predictable quadratic variation process has the form $\langle M^{(\sigma_1)}, M^{(\sigma_1)}\rangle_t = \int_0^t q_s ds$ for some locally bounded process q; μ is Poisson random measure with compensator $\nu(ds,dz) = ds \otimes \lambda(dz)$ for some σ-finite measure $\lambda(\cdot)$ on $\mathbb{R}$; $\delta^{(\sigma_1)}$ is a $\mathbb{R}^{d\times d}$-valued predictable function such that for a sequence $(\mathcal{J}_m)_{m\geq 1}$ of nonnegative λ-integrable functions $\mathcal{J}_m$ on $\mathbb{R}$, $||\delta^{(\sigma_1)}(\omega,t,z)||^2 \wedge 1 \leq \mathcal{J}_m(z)$ for all $t \leq T_m$ and $z\in\mathbb{R}$, where $(T_m)_{m\geq 1}$ is a sequence of stopping times that increases to infinity.[8] □

Assumption LM. For some constant $\epsilon \in (0,1/2]$, a sequence $(T_m)_{m\geq 1}$ of stopping times that increases to infinity and a sequence $(K_m)_{m\geq 1}$ of positive constants, the following holds: for all $t,s \in [0,T]$, $\mathbb{E}[||\sigma_{2,t\wedge T_m} - \sigma_{2,s\wedge T_m}||^2] \leq K_m |t-s|^{1+2\epsilon}$. □

Assumption IS is the "standard" assumption on volatility in the study of nonlinear volatility functionals; see, for example, Jacod (2008) and Jacod and Rosenbaum (2013). This assumption accommodates typical jump-diffusion volatility models.[9] Importantly, it allows for volatility jumps with arbitrary activity. The jumps conveniently capture the "abrupt moves" of volatility in the short-run, which is well known to be empirically important.[10]

Assumption LM imposes the regularity that the long-memory component σ_2 needs to satisfy, namely, σ_2 is locally $(1/2+\epsilon)$-Hölder continuous under the L_2-norm. By Kolmogorov's continuity theorem, σ_2 necessarily has continuous

[8] The $M_t^{(\sigma_1)}$ process may contain a continuous local martingale driven by Brownian motions that are independent of W. It may also contain compensated "small" volatility jumps in the form of a purely discontinuous local martingale. We remind the reader that two local martingales are called orthogonal if their product is a local martingale (or equivalently, their predictable covariation process is identically zero). A local martingale is called purely discontinuous if it is orthogonal to all continuous local martingales. See Def. I.4.11 and Prop. I.4.15 in Jacod and Shiryaev (2003) for additional details.

[9] See, Singleton (2006) and Shepard (2005) and the many references therein.

[10] See, for example, Duffie, Pan, and Singleton (2000), Eraker, Johannes, and Polson (2003), Eraker (2004), Broadie, Chernov, and Johannes (2007), Todorov, Tauchen, and Grynkiv (2011), among others.

paths. This component is "smoother" than the Brownian component in the Itô semimartingale σ_1, which is locally (1/2)-Hölder continuous under the L_2 -norm; but the latter also enjoys the martingale property, which can compensate its lack of smoothness in our asymptotic analysis. We note that the index ϵ is related to the memory parameter in long-memory models. For example, if σ_2 is an fBM with Hurst index $H > 1/2$ (which corresponds to the case with long memory), then Assumption LM is verified for $\epsilon = H - 1/2$. In view of the connection between the fBM and traditional long-memory time series (see, e.g., Geweke and Porter-Hudak, 1983, Comte and Renault, 1996), ϵ can be further related to the memory parameter for fractionally integrated processes in discrete time models like those considered by Granger (1980) and Baillie et al. (1996). The scope of Assumption LM is actually far beyond the fBM; the somewhat technical discussion on general classes of examples is given in Section 2.2.

Finally, we note that Assumptions IS and LM both allow for unrestricted nonstationarity and dependence between the volatility and the price processes. In particular, we allow for leverage effect and price-volatility co-jumps. Overall, the LMIS model is quite general and is satisfied by almost all models in economics and finance.

2.2. Additional Examples of Long-Memory Type Volatility Dynamics

In this subsection, we discuss various examples that satisfy Assumption LM, with the leading one being the fractional stochastic volatility model introduced by Comte and Renault (1996, 1998). More generally, we show that Assumption LM is verified for stochastic (Wiener and/or Skorohod) integrals with respect to the fBM or other multifractal processes and it is also preserved under smooth transformations. Since Assumption LM can be verified component-by-component, here, we suppose that the process σ_2 is scalar-valued without loss of generality. The examples below can also be extended to even larger classes through a standard localization technique, because Assumption LM only requires some regularities up to a localizing sequence of stopping times. Below, W^H denotes an fBM with Hurst index $H > 1/2$ and $\|\cdot\|_2$ denotes the L_2 norm.

Example 1 (Wiener Integrals).
Analogous to volatility modeling based on Itô calculus, a natural way of building general stochastic volatility models from the fBM is via stochastic integration; see Comte and Renault (1996, 1998) for early contributions in economics and finance. Since the fBM is not a semimartingale when $H \neq 1/2$, the classical integration theory for semimartingales cannot be directly applied to the fBM. That being said, when the integrand f is deterministic and satisfies certain conditions, we can define Wiener integrals denoted by $\sigma_{2,t} = \int_0^t f(s)\, dW_s^H$. This stochastic integral is constructed as an isometry from $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ to the L_2 space, where

$$\|f\|_{\mathcal{H}} \equiv \left(\int_0^T \int_0^T f(u) f(v) \, |u - v|^{2H-2} \, du dv \right)^{1/2}$$

and $\mathcal{H} = \{f : \|f\|_{\mathcal{H}} < \infty\}$. If the function f is bounded on bounded sets (e.g., when $t \mapsto f(t)$ is càdlàg), we can use the isometry to verify $\|\sigma_{2,t} - \sigma_{2,s}\|_2 \le K\|W_t^H - W_s^H\|_2$. That is, $\sigma_{2,t}$ inherits the same Hölder continuity from the driving fBM, as required by Assumption LM.

Example 2 (Hermite Processes).
The fBM is a special case of Hermite processes of order one. A higher-order Hermite process is not Gaussian, but shares the same covariance function as the fBM. As a result, a Hermite process with Hurst index H also verifies Assumption LM for $\epsilon = H - 1/2$. In addition, Wiener integrals with respect to Hermite processes can be defined using the same isometry as described in Example 1 and, hence, inherits the same Hölder continuity from the driving Hermite process.

Example 3 (Skorokhod Integrals).
An undesirable limitation of Wiener integrals considered in Example 1 is that the integrand f needs to be deterministic or, more generally, independent of the driving fBM. In the more general case with dependent integrand, one can define $\sigma_{2,t} = \int_0^t f(s)\,dW_s^H$ as a Skorokhod integral using Malliavin calculus, which has become the cornerstone of the stochastic integration theory for fBM.[11] Under the mild conditions (which can be further weakened by localization) that $\sup_{t\in[0,T]} \mathbb{E}\,|f(t)| < \infty$ and $\sup_{t\in[0,T]} \mathbb{E}\left[\int_0^T |D_s f(t)|^2\,ds\right] < \infty$, where D denotes the (Malliavin) derivative operator associated with W^H, it can be shown[12]

$$\left\|\sigma_{2,t} - \sigma_{2,s}\right\|_2 \le K_H\,|t-s|^H,$$

for some constant K_H that only depends on the Hurst index H. Assumption LM is verified for $\epsilon = H - 1/2$.

Finally, we note that Assumption LM is preserved under smooth transformations. That is, if $\tilde{\sigma}_{2,t}$ satisfies Assumption LM for some stopping times $(\tilde{T}_m)_{m\ge1}$ and constants $(\tilde{K}_m)_{m\ge1}$, then $\sigma_{2,t} = h(\tilde{\sigma}_{2,t})$ also satisfies this assumption for any continuously differentiable function $h(\cdot)$. To see this, we note that we can choose a localizing sequence of stopping times (T'_m) such that $\tilde{\sigma}_{2,t}$ is bounded for $t \le T'_m$ (because $\tilde{\sigma}_{2,t}$ is continuous and, hence, locally bounded under Assumption LM). By the mean-value theorem, $\|\sigma_{2,t} - \sigma_{2,s}\| \le K'_m\|\tilde{\sigma}_{2,t} - \tilde{\sigma}_{2,s}\|$ for $t, s \le T'_m$, where $K'_m \equiv \sup_{t\le T\wedge T'_m} \left\|\partial h\left(\tilde{\sigma}_{2,t}\right)/\partial\tilde{\sigma}_{2,t}\right\|$ is finite. From here, it is easy to see that Assumption LM holds for $\sigma_{2,t}$ with respect to stopping times $T_m = \tilde{T}_m \wedge T'_m$ and constants $K_m = \tilde{K}_m K'_m$. It is interesting to note that this permanence property holds because we only require the Hölder continuity to hold up to a localizing sequence

[11] The Skorokhod integral is defined by considering the fBM as an isonormal Gaussian process and evaluating the associated divergence operator on the integrand. See Nualart (2005) for a comprehensive review about Malliavin calculus.

[12] This estimate follows from the calculations in the proof of Thm. 5 in Alòs and Nualart (2003) and our uniform bound on the moments of f and Df.

of stopping times; this highlights the usefulness of using this weaker requirement in our setting.

The permanence property discussed above can be used to extend the above examples to even larger classes. A basic, but important, example concerns log volatility models as follows. If $\sigma_{2,t} = \exp\left(\tilde{\sigma}_{2,t}\right)$ and the log-volatility process $\tilde{\sigma}_{2,t}$ verifies Assumption LM (e.g., $\tilde{\sigma}_{2,t}$ is generated from the examples above), then $\sigma_{2,t}$ also satisfies Assumption LM. In the same vein, we provide a more involved example concerning the solution of fBM-driven stochastic differential equations.

Example 4 (Ornstein–Uhlenbeck (OU) Process).
Comte and Renault (1996, 1998) modeled the log-volatility process $\tilde{\sigma}_{2,t} = \log\left(\sigma_{2,t}\right)$ as an OU process driven by the fBM. That is, $\tilde{\sigma}_{2,t} = e^{-\kappa t}(\xi + \gamma \int_0^t e^{\kappa s} dW_s^H)$, where ξ is the initial value and the parameters (κ, γ) control the mean reversion and "volatility-of-volatility," respectively. Note that the process $\sigma_{2,t}$ is a smooth transformation of the processes $e^{-\kappa t}$ and $\int_0^t e^{\kappa s} dW_s^H$. Since these processes verify Assumption LM (recall Example 1), $\sigma_{2,t}$ also satisfies Assumption LM in view of the aforementioned permanence property.

3. MAIN RESULTS UNDER REGULAR SAMPLING

3.1. The Estimation of Integrated Volatility Functionals

In this section, we consider the baseline setting with regular sampling. We now describe the Jacod–Rosenbaum estimator for the integrated volatility functional (2.2) and derive its asymptotic properties under the LMIS volatility model. The estimator is constructed in two steps. In the first step, we nonparametrically recover the spot covariance process. To this end, we choose an integer sequence k_n of bandwidths and a sequence u_n of truncation thresholds such that

$$k_n \asymp \Delta_n^{-\gamma} \text{ and } u_n \asymp \Delta_n^{\varpi} \tag{3.1}$$

for some constants $\gamma \in (0,1)$ and $\varpi \in (0,1/2)$. These tuning parameters will be further restricted in our asymptotic result below. The spot covariance matrix at time $i\Delta_n$ is then given by

$$\hat{c}_{i\Delta_n} \equiv \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \left(\Delta_{i+j}^n X\right) \left(\Delta_{i+j}^n X\right)^\top 1_{\{||\Delta_{i+j}^n X|| \leq u_n\}}. \tag{3.2}$$

We note that u_n determines the truncation threshold for eliminating jumps in X (Mancini, 2001). If X is continuous, then there is no need to truncate when forming $\hat{c}_{i\Delta_n}$ (by taking $u_n = \infty$).

In the second step, we construct the estimator for $S(g)$ as

$$S_n(g) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \left(g(\hat{c}_{i\Delta_n}) - \frac{1}{k_n} \mathbb{B}g\left(\hat{c}_{i\Delta_n}\right) \right), \tag{3.3}$$

where the correction function $\mathbb{B}g$ associated with g is given by

$$\mathbb{B}g(c) \equiv \frac{1}{2} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(c)(c^{jl}c^{km} + c^{jm}c^{kl}),$$

with c^{jk} denoting the (j,k) element of c and $\partial^2_{jk,lm} g(c)$ denoting the second-order partial derivative of g with respect to c^{jk} and c^{lm}. We note that the estimator $S_n(g)$ is formed as a bias-corrected version of the (raw) "plug-in" estimator $\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} g(\hat{c}_{i\Delta_n})$. The correction term is introduced to eliminate a high-order nonlinearity bias term that arises from the statistical error in the spot volatility estimation.

The estimator $S_n(g)$ was first proposed by Jacod and Rosenbaum (2013) for $\mathcal{C}^3$ (i.e., three-time continuously differentiable) $g(\cdot)$ functions.[13] These authors assume that the volatility process σ is an Itô semimartingale and derive a central limit theorem for it. The theory of Jacod and Rosenbaum (2013) requires some growth condition on the transformation $g(\cdot)$ that is restrictive in applications; for example, it excludes basic quantities like beta, correlation, or idiosyncratic variance. Such restriction is relaxed in Li et al. (2017) and Li and Xiu (2016) by using a uniform approximation result of spot volatility and a spatial localization argument; the resulting theory is applicable to essentially all $\mathcal{C}^3$ functions. We further extend these latter papers by allowing for the general LMIS volatility dynamics.

Theorem 1, presents the asymptotic distribution of $S_n(g)$. We denote by $\partial g(c)$ the $d \times d$ matrix whose (j,k) element is $\partial_{jk} g(c)$ and denote the matrix trace function by $\mathrm{Tr}[\,\cdot\,]$.

THEOREM 1. *Suppose Assumptions H, C, IS and LM, and (3.1) holds with*

$$\frac{r}{2} \vee \frac{1}{3} < \gamma < \frac{1}{2} \wedge \frac{2\epsilon}{1+2\epsilon}, \quad \frac{1-\gamma}{2-r} \le \varpi < \frac{1}{2}. \tag{3.4}$$

Then, the sequence of variables $\Delta_n^{-1/2}(S_n(g) - S(g))$ converges stably in law to a mixed centered Gaussian distribution with $\mathcal{F}$-conditional variance $V(g)$ given by

$$V(g) \equiv 2\int_0^T \mathrm{Tr}[c_s \partial g(c_s) c_s \partial g(c_s)]\, ds. \tag{3.5}$$

Comments. (i) Theorem 1 extends the results in Jacod and Rosenbaum (2013), Li et al. (2017), and Li and Xiu (2016) by establishing the asymptotic distribution of $S_n(g)$ under the LMIS volatility dynamics. In particular, in the absence of the long-memory component σ_2 (hence,

[13] Constructing the estimator clearly requires $g(\cdot)$ to be at least twice continuously differentiable. The $\mathcal{C}^3$ condition is used in the proof to control the approximation error of a second-order Taylor expansion. Technically speaking, one might relax this condition slightly by assuming that $g(\cdot)$ belongs to the Hölder class with index less than 3.

the lack of Assumption LM), Theorem 1 coincides with those in prior work.[14]

(ii) The "cost" of including the long-memory component is that we need an additional upper bound for the divergence rate of the local window size k_n, that is, $\gamma < 2\epsilon/(1+2\epsilon)$. This restriction is weaker when ϵ is larger, which corresponds to the case with "longer" memory. In the extreme case with $\epsilon = 1/2$ (i.e., σ_2 has locally Lipschitz path under the L_2-norm), this restriction is absent, because σ_2 then behaves essentially like a drift term. In this case, $\gamma < 1/2$ amounts to a standard "undersmoothing" condition, in view of the well-known fact that the optimal $\Delta_n^{-1/4}$-rate of convergence for the spot covariance estimation is attained with $k_n \asymp \Delta_n^{-1/2}$.

(iii) The condition (3.4) implicitly imposes a restriction on ϵ, that is, $\epsilon > 1/4$. In other words, σ_2 is Hölder-continuous under the L_2-norm with an index at least 3/4, which shows an apparent discrepancy relative to the (1/2)-Hölder continuity of the Itô semimartingale component. This "gap" arises as a compensation for the lack of martingale property in the long-memory component, whereas the martingale property is heavily exploited in previous work based on the Itô semimartingale volatility dynamics.[15] Although our proofs rely on the $\epsilon > 1/4$ condition, it is unclear whether it is necessary for the central limit theorem to hold. Further relaxation of this condition may be interesting for future research.

3.2. The Semiparametric Efficiency Bound

In this subsection, we establish the semiparametric efficiency bound for estimating $S(g)$ and show that the Jacod–Rosenbaum estimator attains this bound.[16] The result here is derived under regular sampling to highlight the key ideas, and will be extended in Section 4 to further accommodate a type of irregular sampling.

In contrast to the classical literature on efficient estimation (see, e.g., Bickel et al., 1998), our setting is more complicated in that the likelihood ratio generally

[14] It is instructive to recall the intuition concerning the restrictions on the local window k_n. Jacod and Rosenbaum (2013) require $\gamma > 1/3$ to eliminate a third-order nonlinearity bias, and impose the $\gamma < 1/2$ undersmoothing condition to eliminate biases from boundary effect, Brownian movements in volatility, and volatility jumps. Li et al. (2017) and Li and Xiu (2016) relax the polynomial growth restriction by using a uniform approximation result for the spot estimator, which requires the additional condition $\gamma > r/2$. Note that the latter condition is trivially satisfied when X is continuous or its jumps are finitely active (corresponding to $r = 0$).

[15] Specifically, a key source of bias in the nonparametric volatility estimation is the temporal variation of stochastic volatility. Although increments of the (Brownian) martingale volatility component are individually "large," they are serially uncorrelated, and hence, become small after time-aggregation.

[16] We consider the efficiency bound to be "semiparametric" in that the integrated volatility functional is finite-dimensional. The efficiency problem under consideration is the same as Renault et al. (2017), although these authors refer to their bound as "nonparametric."

does not satisfy the LAN property. Indeed, the asymptotic variances of estimators and information matrices depend on the realization of the volatility path, and hence, are random. For this reason, we cannot directly apply classical results in the literature on efficient estimation. Following Stein's insight (Stein, 1956), we start with the first principle by constructing a class of parametric submodels and compute the lower efficiency bound among them.

A similar approach has been adopted in the recent work of Clément et al. (2013) and Renault et al. (2017). Clément et al. (2013) considered a conditional Markov model for the volatility dynamics, but they rule out both long-memory behavior and volatility jumps, which are of great economic and econometric interest. Renault, Sarisoy, and Werker (2017) studied a conditional (pathwise) LAN setting that allows for long-memory volatility dynamics and finite-variational jumps. Their theory requires that volatility paths have locally bounded variance (see their Def. 1).

However, none of the aforementioned efficiency results can be directly invoked to cover our LMIS volatility model. For this reason, we now establish the efficiency bound in this general setting. Our approach is similar to that of Renault et al. (2017) but with a notable difference, theoretically speaking. That is, instead of targeting a conditional LAN property, we only establish a LAMN property for the likelihood ratios. The LAMN property is weaker than the conditional LAN property, but it is sufficient for establishing the efficiency bound.

Turning to the details, we start by constructing a class of parametric submodels with a one-dimensional parameter $a \in (-1,1)$ whose true value is 0. Let $\mathbb{DS}_d$ denote the space of càdlàg functions on $[0,T]$ that take values in the space of $d \times d$ symmetric real-valued matrices; let I_d denote the d-dimensional identity matrix. For each $h \in \mathbb{DS}_d$, we consider a parametric submodel for the observations $\left(\Delta_i^n X\right)_{1\leq i \leq [T/\Delta_n]}$ in which the processes $(b_t)_{t\geq 0}$ and $(J_t)_{t\geq 0}$ are observed and the spot covariance matrix is parametrized as

$$c(a,h)_t = \left(I_d + \frac{1}{2} a c_t h_t\right) c_t \left(I_d + \frac{1}{2} a c_t h_t\right)^\top, \tag{3.6}$$

where we remind the reader that a is the only unknown parameter in the submodel.

Two remarks on these submodels are in order. First, the "sandwich-form" parametrization (3.6) is designed to automatically ensure that the spot covariance matrix $c(a,h)_t$ is indeed positive semidefinite, without additional restrictions on the parameter space. Second, we only need to perturb the volatility path in certain "directions" of the form $c_t h_t$ for h_t symmetrically valued. This apparent restriction is actually without loss of generality, because it turns out that this restricted class already contains the least favorable model. We hence impose this restriction *a priori* so as to eliminate unnecessary analytical complications.

Unfortunately, we need to make some nontrivial restrictions in order to derive the LAMN property, which are similar to Renault et al. (2017).

Assumption LAMN. (i) The Brownian motion $(W_t)_{0\leq t\leq T}$ is independent of $\mathcal{G} = \sigma((b_t, c_t, J_t)_{0\leq t\leq T})$; (ii) $(\lambda_{\min}(c_t)^{-1})_{t\geq 0}$ is locally bounded, where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue. □

Assumption LAMN is restrictive because its condition (i) rules out the so-called leverage effect.[17] That said, we stress that we only need this restriction for computing the efficiency bound, while the limit theorems regarding our estimators are proved without this restriction. Under this condition, we can write the log likelihood ratio for each submodel indexed by $h \in \mathbb{DS}_d$ as

$$L_n(a\Delta_n^{1/2}, h) \equiv \log \frac{dP_n(a\Delta_n^{1/2}, h)}{dP_n^0},$$

where $P_n(a,h)$ is the $\mathcal{G}$-conditional distribution of $\left(\Delta_i^n X\right)_{1\leq i\leq [T/\Delta_n]}$ under the model (3.6), and P_n^0 is the true conditional distribution.[18]

We shall show that, for each $h \in \mathbb{DS}_d$, the corresponding submodel satisfies the LAMN property at $a = 0$. That is, for a sequence ξ_n of random variables and $\mathcal{G}$-measurable (random) information $\Gamma(h)$, we have

$$\begin{cases} L_n(a\Delta_n^{1/2}, h) = a\xi_n(h) - \dfrac{a^2}{2}\Gamma(h) + o_p(1), \\ (\xi_n(h), \Gamma(h)) \xrightarrow{\mathcal{L}} (\xi(h), \Gamma(h)), \end{cases} \tag{3.7}$$

where, conditionally on $\mathcal{G}$, the limiting variable $\xi(h)$ is centered Gaussian with variance $\Gamma(h)$ and $\xrightarrow{\mathcal{L}}$ denotes the convergence in law. Given the LAMN property, we can use the conditional convolution theorem (Jeganathan, 1982, 1983) to show that the lower efficiency bound for estimating the unknown parameter a in the submodel for a given h is $\Gamma(h)^{-1}$, and hence, the corresponding bound for estimating $S(g) = \int_0^T g(c(a,h)_s)\,ds$ is

$$\begin{aligned} \Sigma(h) &\equiv \left(\frac{\partial}{\partial a} \left[\int_0^T g(c(a,h)_s)\,ds \right]\Big|_{a=0} \right)^2 \Big/ \Gamma(h) \\ &= \left(\int_0^T \mathrm{Tr}[c_s \partial g(c_s) c_s h_s]\,ds \right)^2 \Big/ \Gamma(h). \end{aligned} \tag{3.8}$$

From here, we further compute the lower efficiency bound by searching for the least favorable submodel. Theorem 2, states our main efficiency result in the case with regular sampling.

[17] We note that a certain type of leverage effect can be accommodated by Clément et al. (2013), at the cost of ruling out jumps and long-memory in volatility. Extending Clément et al. (2013) to the more general setting with LMIS volatility would be interesting, but is also mathematically very challenging, because the approach of Clément et al. (2013) relies heavily on technical tools from Malliavin calculus. We leave this question to future research.

[18] Note that the likelihood ratio only depends on the $\mathcal{G}$-conditional distributions because the marginal distribution of (b, c, J) is canceled in the likelihood ratio.

THEOREM 2. *Under Assumptions H, IS, LM, and LAMN, the following statements hold:*

(a) *for each* $h \in \mathbb{DS}_d$*, the LAMN property (3.7) is satisfied at* $a = 0$ *with the information given by*

$$\Gamma(h) = \frac{1}{2}\int_0^T \mathrm{Tr}\left[c_s h_s c_s h_s\right] ds; \tag{3.9}$$

(b) *the lower efficiency bound for estimating* $S(g)$ *among all submodels indexed by* $h \in \mathbb{DS}_d$*, that is,* $\sup_{h\in\mathbb{DS}_d} \Sigma(h)$*, coincides with* $V(g)$ *given by (3.5);*

(c) *the supremum in* $\sup_{h\in\mathbb{DS}_d} \Sigma(h)$ *is attained at* $h^* = \partial g(c)$ *for every sample path.*

Comments. (i) Part (a) shows that each submodel (indexed by h) satisfies the LAMN property with random information given by (3.9). By the conditional convolution theorem, this result justifies $\Sigma(h)$ (recall (3.8)) as the efficiency bound for estimating $S(g)$ in the submodel.[19]

(ii) Part (b) computes the worst-case efficiency bound among all $h \in \mathbb{DS}_d$. Since the bound $\sup_{h\in\mathbb{DS}_d} \Sigma(h)$ is based on a specific class of submodels, it is possible that it is smaller than the true efficiency bound. We rule out this possibility by verifying that it is sharp; indeed, this bound coincides with the $\mathcal{F}$-conditional asymptotic variance of $S_n(g)$ given in Theorem 1. Since $S_n(g)$ does not depend on the information in the parametric submodel (including the information contained in $\mathcal{G}$), we conclude, *a fortiori* , that $\sup_{h\in\mathbb{DS}_d} \Sigma(h) = V(g)$ is indeed the efficiency bound and $S_n(g)$ is semiparametrically efficient.

(iii) Part (c) sheds light on the least favorable model. From the definitions of $\Sigma(h)$ and $V(g)$, it is easy to see that $\Sigma(h^*) = V(g)$. Hence, the least favorable model is chosen "in a random fashion," in that it depends on the realization of the volatility path. This phenomenon has also been recognized in Renault et al. (2017).

(iv) Finally, we remark on the relationship between Theorem 2 and its counterpart in Renault et al. (2017) (see Thm. 3.2 there) from a theoretical viewpoint. The key difference between these results is that we derive a convergence of the $\mathcal{G}$-conditional law of $\xi_n(h)$ in a weak sense, while Renault et al. (2017) derive a convergence in a strong sense. More precisely, in the proof of the LAMN property, we show that

$$\rho\left(\mathcal{L}_{\mathcal{G}}\left[\xi_n(h)\right], \mathcal{L}_{\mathcal{G}}\left[\xi(h)\right]\right) \xrightarrow{\mathbb{P}} 0, \tag{3.10}$$

[19]More precisely, the conditional convolution theorem shows that $\Sigma(h)$ is the efficiency bound in the submodel for any regular estimator. The regularity of the $S_n(g)$ estimator follows from Lem. 1 of Li et al. (2017).

where $\rho(\cdot,\cdot)$ denotes Prohorov's metric (for the weak convergence of measures) and $\mathcal{L}_{\mathcal{G}}[\cdot]$ denotes the $\mathcal{G}$-conditional law. It is known that (3.10) implies the $\mathcal{G}$-stable convergence in law of $\xi_n(h)$ towards $\xi(h)$, which in turn implies the convergence in (3.7).[20] In contrast, the pathwise argument underlying the conditional LAN setting of Renault et al. (2017) leads to

$$\rho\left(\mathcal{L}_{\mathcal{G}}[\xi_n(h)], \mathcal{L}_{\mathcal{G}}[\xi(h)]\right) \xrightarrow{a.s.} 0. \tag{3.11}$$

Given that (3.10) is much weaker than (3.11), it is not surprising that we can allow for a more general class of volatility models than Renault et al. (2017) without resorting to high-level assumptions that might be difficult to verify.

4. EXTENSION: THE CASE WITH IRREGULAR SAMPLING

In this section, we show that the results above (i.e., Theorems 1 and 2) are robust with respect to a type of irregular sampling. To proceed, we first need to adapt the Jacod–Rosenbaum estimator to the case with irregular sampling, and then establish its asymptotic distribution and semiparametric efficiency under the general LMIS volatility dynamics. This section presents our theoretical results in the most general form, which further sets our study apart from prior work (cf. Jacod and Rosenbaum (2013)).

We start with describing the irregular sampling scheme. We suppose that X is sampled at random times $0 = t(n,0) < t(n,1) < \cdots$ within the fixed time span $[0,T]$.[21] The return and the time span for the ith sampling interval are now given by, respectively,

$$\Delta_i^n X \equiv X_{t(n,i)} - X_{t(n,i-1)}, \quad \Delta_{n,i} \equiv t(n,i) - t(n,i-1).$$

The number of returns observed before time t is denoted by $N_{n,t} \equiv \sum_{i\geq 1} 1_{\{t(n,i)\leq t\}}$. Below, we denote by $\mathcal{H}_t$ the smallest filtration that contains $\mathcal{F}_t$ and with respect to which $\{t(n,i) : i \geq 1\}$ are stopping times. We also denote by $\mathcal{T}$ the σ-field generated by these sampling times. We impose the following regularity condition on the sampling scheme, which is inspired by the recent work of Jacod et al. (2017, 2019). Under irregular sampling, Δ_n no longer denotes "the" sampling duration, but only serves as an asymptotic device, which can be understood as an average measure of sampling durations.

[20] See, for example, Prop. 5 of Barndorff-Nielsen et al. (2008).

[21] We implicitly assume that individual assets are sampled at the same times. Hence, with nonsynchronous data, we need to sample the data sparsely so as to "align" the sampling times and mitigate the Epps effect in the estimation of the spot covariance matrix. Note that sparse sampling is also needed here to mitigate the effect of microstructure noise. The issues with noise and asynchronicity are more relevant for studying ultra high frequency data, which is beyond the scope of the current paper. See Zhang, Mykland, and Aït-Sahalia (2005) and Zhang (2011) for additional discussions on these issues.

Assumption IR. For a strictly positive continuous adapted process Λ and a sequence $(T_m)_{m\geq 1}$ of stopping times increasing to infinity, the following conditions hold: (i) for each $i \geq 1$, $\Delta_{n,i}$ is, conditionally on $\mathcal{H}_{t(n,i-1)}$, independent of $\mathcal{F}_T$; (ii) in restriction to the set $\{t(n,i-1) < T_m\}$, $\left|\mathbb{E}\left[\Delta_{n,i}\Lambda_{t(n,i-1)}|\mathcal{H}_{t(n,i-1)}\right] - \Delta_n\right| \leq K_{m,1}\Delta_n\kappa_n$ and $\mathbb{E}[\Delta_{n,i}^q|\mathcal{H}_{t(n,i-1)}] \leq K_{m,q}\Delta_n^q$ for all $q > 0$, where $(K_{m,q})_{m\geq 1, q>0}$ are constants and the sequence $\kappa_n = o(1)$; (iii) $N_{n,T} = O_p\left(\Delta_n^{-1}\right)$. □

This assumption accommodates time-changed regular sampling and Poisson-type sampling; see the examples in Jacod, Li, and Zheng (2017) for additional details. The process Λ in Assumption IR can be interpreted as the sampling intensity. Indeed, it can be shown that $\Delta_n N_t^n \xrightarrow{\mathbb{P}} \int_0^t \Lambda_s ds$ locally uniformly in t. Note that we can scale both Δ_n and Λ by the same multiplicative factor, which does not affect the applicability of the asymptotic result. Assumption IR allows the sampling intensity Λ to be dependent on the underlying processes. That said, condition (i) imposes a type of exogeneity on the sampling time; namely, given the current information, the next sampling time is independent of the future behavior of the underlying processes. Condition (ii) amounts to saying that $\Delta_{n,i}$ is expected to be approximately inversely proportional to the sampling intensity $\Lambda_{t(n,i-1)}$.

The estimator $S_n(g)$ in (3.3) can be naturally adapted with respect to the irregular sampling scheme as follows:

$$S_n'(g) \equiv \sum_{i=0}^{N_{n,T}-k_n} \left(g(\hat{c}_{t(n,i)}) - \frac{1}{k_n}\mathbb{B}g\left(\hat{c}_{t(n,i)}\right)\right)\Delta_{n,i+1}, \tag{4.1}$$

where the spot covariance matrix at time $t(n,i)$ is now estimated by

$$\hat{c}_{t(n,i)} \equiv \frac{1}{k_n}\sum_{j=1}^{k_n} \frac{\left(\Delta_{i+j}^n X\right)\left(\Delta_{i+j}^n X\right)^\top}{\Delta_{i+j}^n} 1_{\{\|\Delta_{i+j}^n X\| \leq u_{n,i+j}\}}, \tag{4.2}$$

with the tuning parameters satisfying

$$k_n \asymp \Delta_n^{-\gamma}, \quad u_{n,i} \asymp \Delta_{n,i}^{\varpi}. \tag{4.3}$$

Theorems 3 and 4, extend results in the previous section to the setting with irregular sampling. For these results to hold, we need to modify Assumptions LM and LAMN so as to accommodate random sampling times:

Assumption LM'. For some constant $\epsilon \in (0, 1/2]$, a sequence $(T_m)_{m\geq 1}$ of stopping times that increases to infinity and a sequence $(K_m)_{m\geq 1}$ of positive constants, the following holds: for $t, s \in [0, T]$, $\mathbb{E}[\left\|\sigma_{2,t\wedge T_m} - \sigma_{2,s\wedge T_m}\right\|^2 |\mathcal{T}] \leq K_m |t-s|^{1+2\epsilon}$. □

Assumption LAMN'. Assumption LAMN holds with $\mathcal{G}$ given by the σ-field generated by (b, c, J) and $\mathcal{T}$. □

THEOREM 3. *Suppose Assumptions H, C, IS, IR, LM′ and the tuning parameters in (4.3) satisfy (3.4). Then, the sequence of variables* $\Delta_n^{-1/2}(S'_n(g) - S(g))$ *converges stably in law to a mixed centered Gaussian distribution with* $\mathcal{F}$*-conditional variance* $V'(g)$ *given by*

$$V'(g) \equiv 2\int_0^T (\mathrm{Tr}[c_s \partial g(c_s) c_s \partial g(c_s)]/\Lambda_s)\, ds. \tag{4.4}$$

THEOREM 4. *Under Assumptions H, IS, IR, LM′, and LAMN′, the following statements hold:*

(a) *for each* $h \in \mathbb{DS}_d$*, the LAMN property (3.7) is satisfied at* $a = 0$ *with the information given by*

$$\Gamma(h) = \frac{1}{2}\int_0^T \mathrm{Tr}[c_s h_s c_s h_s]\,\Lambda_s ds; \tag{4.5}$$

(b) *the lower efficiency bound for estimating* $S(g)$ *among all submodels indexed by* $h \in \mathbb{DS}_d$*, that is,* $\sup_{h\in\mathbb{DS}_d} \Sigma(h)$*, coincides with* $V'(g)$ *given by (4.4);*

(c) *the supremum in* $\sup_{h\in\mathbb{DS}_d} \Sigma(h)$ *is attained at* $h^* = \partial g(c)/\Lambda$ *for every sample path.*

Comments. (i) Theorem 3 establishes the asymptotic distribution of the estimator given by (4.1). Under irregular sampling, the asymptotic variance of this estimator depends inversely on the sampling intensity Λ.

(ii) Theorem 4 shows that the parametric submodels satisfy the LAMN property, and the lower efficiency bound among them coincides with the asymptotic variance of our estimator. Hence, the estimator given by (4.1) is semiparametrically efficient.

(iii) Renault et al. (2017) first established the efficiency bound in the scalar case (see (3.18) there), which is the same as ours; we have extended the validity of this bound for the general LMIS volatility dynamics using a different theoretical approach. In addition, we show that this efficiency bound is sharp, as it is attained by the estimator $S'_n(g)$. This result verifies the conjecture of Renault et al. (2017), which is based on a "near efficiency" statement in a sequential asymptotic embedding. Our asymptotic result, in contrast, is derived in the usual asymptotic setting where tuning parameters change simultaneously with the sample size. Importantly, we note that our estimator is different from that in Renault et al. (2017), which is based on the kernel estimator of Kristensen (2010). The latter estimator does not involve bias-correction and requires the volatility process to have differentiable paths. Our bias-corrected estimator is valid for the general LMIS stochastic volatility dynamics without the (restrictive) differentiability assumption.

5. MONTE CARLO EXPERIMENTS

In this section, we examine the finite-sample performance of the asymptotic theory in Monte Carlo experiments. We simulate the log price process X and its volatility process σ according to the following:

$$dX_t = \sigma_t dW_t + dJ_t^X, \quad \sigma_t = \sigma_{1,t} + \sigma_{2,t},$$

where the jump component J^X is simulated as a compensated compound Poisson process with $\mathcal{N}(-0.01, 0.02^2)$ distributed jump size and intensity $\lambda = 36$ per year. The Itô semimartingale volatility component $\sigma_{1,t}$ is generated by

$$\sigma_{1,t} = \exp(-1.6 + F_t), \quad dF_t = -5F_t dt + 2d\widetilde{W}_t + dJ_t^F.$$

Here, $\widetilde{W}$ is another standard Brownian motion satisfying $\mathbb{E}[dW_t d\widetilde{W}_t] = -0.75dt$, which captures the so-called "leverage effect." The volatility jump process J^F is simulated as a compound Poisson process with the same jump arrivals as J^X, but with jump sizes exponentially distributed with mean 0.3, and hence, features price-volatility co-jumps. The long-memory volatility component σ_2 is simulated from the fractional stochastic volatility model studied by Comte and Renault (1998) as follows (see eqn. (2.1) of that paper):

$$\sigma_{2,t} = \exp(-2.5 + \kappa_t), \quad d\kappa_t = -\kappa_t dt + 0.5dW_t^{\epsilon}, \tag{5.1}$$

where W^{ϵ} is a fractional Brownian motion with Hurst index $\epsilon + 1/2$. To examine the effect of the memory parameter ϵ on the estimator's performance, we consider $\epsilon \in \{0.1, 0.25, 0.3, 0.4\}$. Note that our theory relies on the (sufficient) condition $\epsilon > 1/4$. But it is nevertheless interesting to examine the finite-sample behavior of the estimator when this condition does not hold (i.e., when $\epsilon = 0.1$ or 0.25).

We consider two sampling frequencies: $\Delta_n = 1$ or 5 min. To examine the robustness of the estimator with respect to the choice of local window size k_n, we consider $k_n \in \{40, 80, 120\}$ in the 1-min case, and $k_n \in \{15, 30, 45\}$ in the 5-min case. Note that in both cases, the largest value of k_n is three times of the smallest value, which provides a fairly large perturbation for the purpose of checking robustness. Following Bollerslev and Todorov (2011), we set the truncation threshold u_n on each day t adaptively as

$$u_n = 3.5 \times \sqrt{BV_t} \times \Delta_n^{0.49},$$

where BV_t is the day-t bipower variation (Barndorff-Nielsen and Shephard, 2004b). The sample span is fixed at $T = 21$ days and the total number of Monte Carlo trials is 1,000.

We consider the estimation of integrated quarticity $\int_0^t \sigma_t^4 ds$, which corresponds to $g(c) = c^2$. In this case, the bias-corrected estimator $S_n(g)$ has the form

$$S_n(g) = \left(1 - \frac{2}{k_n}\right) \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{c}_{i\Delta_n}^2.$$

Table 1. Monte Carlo results.

	Bias	RMSE	90%	95%	Bias	RMSE	90%	95%
				Panel A: Δ_n= *1 min*				
	$\epsilon = 0.1$				$\epsilon = 0.25$			
$k_n = 40$	−0.89	3.50	93.1	96.0	−0.79	3.16	92.9	97.3
$k_n = 80$	−0.81	3.50	90.8	95.4	−0.68	3.16	91.0	96.0
$k_n = 120$	−0.89	3.51	89.9	94.5	−0.73	3.17	89.1	95.3
	$\epsilon = 0.3$				$\epsilon = 0.4$			
$k_n = 40$	−0.90	3.17	92.5	95.9	−0.45	3.29	93.1	96.8
$k_n = 80$	−0.80	3.13	90.7	94.2	−0.38	3.29	91.0	95.8
$k_n = 120$	−0.84	3.14	89.6	93.8	−0.45	3.29	89.7	95.2
				Panel B: Δ_n= *5 min*				
	$\epsilon = 0.1$				$\epsilon = 0.25$			
$k_n = 15$	−2.36	8.20	94.2	96.9	−1.53	7.39	96.6	98.3
$k_n = 30$	−1.14	7.96	91.7	95.1	−0.45	7.11	93.2	96.9
$k_n = 45$	−1.07	7.89	89.9	93.7	−0.41	7.01	91.3	95.9
	$\epsilon = 0.3$				$\epsilon = 0.4$			
$k_n = 15$	−2.83	7.39	94.5	97.0	−1.95	7.37	95.3	97.6
$k_n = 30$	−1.70	7.20	91.7	94.9	−0.84	7.29	91.8	95.9
$k_n = 45$	−1.69	7.19	89.5	94.0	−0.80	7.18	90.7	94.8

Note: This table summarizes the performance of the S_n^* estimator for integrated quarticity in the Monte Carlo simulation. Panel A (resp. B) reports results for sampling frequency $\Delta_n = 1$ min (resp. 5 min). We consider different values of the memory parameter ϵ and local window size k_n. Bias and root mean-square-error (RSME) are in percentages, normalized by the true value of the estimand in each Monte Carlo trial. The columns labeled 90% and 95% report the coverage rates of the corresponding confidence intervals in percentages.

Following the suggestion from Jacod and Rosenbaum (2013), we further implement a finite-sample adjustment for the boundary effect by using

$$S_n^* = S_n(g) + \frac{k_n \Delta_n}{2}\left(\hat{c}_0^2 + \hat{c}_{[T/\Delta_n]-k_n}^2\right). \tag{5.2}$$

Note that the boundary-effect adjustment is of order $O_p(k_n\Delta_n) = o_p(\Delta_n^{1/2})$, and hence, S_n^* has the same asymptotic behavior as that of $S_n(g)$. We focus on S_n^* in the subsequent numerical work. The asymptotic variance of this estimator is $8\int_0^T \sigma_s^8 ds$, which can be consistently estimated by $8\Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \hat{c}_{i\Delta_n}^4$.

Table 1 presents the simulation results. Panel A of the table reports the results for 1-min sampling, including the bias, root mean-square-error (RMSE), and coverage rates of 90% and 95% level confidence intervals. In all cases, the biases (in relative percentage term with respect to the true value) are close to zero, with their magnitudes being less than 1%. From the RMSE column, we also see that the estimator is reasonably accurate. In addition, the coverage rates of the confidence intervals are close to their nominal levels. We are particularly interested in whether the estimator's performance is robust to the choice of k_n, and the findings seem satisfactory. It is also interesting to note that, although our theory requires "sufficiently long memory (i.e., $\epsilon > 1/4$)," the estimator and the associated confidence interval perform well when $\epsilon = 0.1$. That said, we do observe notably larger RMSEs in this case than the other cases with smoother long-memory components. The results in the case with 5-min sampling are qualitatively similar, but with larger bias and RMSE, which are expected for the smaller sample size.

6. CONCLUDING REMARKS

This paper establishes the stable convergence in law for the Jacod–Rosenbaum estimator of integrated volatility functionals under the general LMIS setting in which the stochastic volatility process contains both Itô semimartingale and long-memory components, whereas existing results only accommodate the former. Under additional sufficient conditions for the LAMN property, we show that the estimator attains the semiparametric efficiency bound under the LMIS volatility model. More general results under random sampling are also developed. For future work, our techniques may be used to extend related theories on integrated volatility functionals (e.g., alternative bias-correction methods, bootstrap inference, and empirical-process theory for a continuum of integrated functionals) to the setting with general LMIS volatility dynamics. Establishing the semiparametric efficiency bound under less restrictive conditions is also an important open question for the study of efficiency in the infill high-frequency setting.

REFERENCES

Aït-Sahalia, Y. & D. Xiu (2019) Principal component analysis of high-frequency data. *Journal of the American Statistical Association* 114(525), 287–303.

Alòs, E. & D. Nualart (2003) Stochastic integration with respect to the fractional Brownian motion. *Stochastics and Stochastic Reports* 75(3), 129–152.

Andersen, T. & T. Bollerslev (1997) Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *Journal of Finance* 52(3), 975–1005.

Andersen, T. & T. Bollerslev (1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39(4), 885–905.

Andersen, T.G., T. Bollerslev, F.X. Diebold, & P. Labys (2001) The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96(453), 42–55.

Andersen, T.G., T. Bollerslev, F.X. Diebold, & P. Labys (2003) Modeling and forecasting realized volatility. *Econometrica* 71(2), 579–625.

Baillie, R.T., T. Bollerslev, & H.O. Mikkelsen (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74(1), 3–30.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, & N. Shephard (2008) Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76(6), 1481–1536.

Barndorff-Nielsen, O.E. & N. Shephard (2001) Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics *Journal of the Royal Statistical Society, Series B* 63(2), 167–241.

Barndorff-Nielsen, O.E. & N. Shephard (2002a) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64(2), 253–280.

Barndorff-Nielsen, O.E. & N. Shephard (2002b) Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* 17(5), 457–477.

Barndorff-Nielsen, O.E. & N. Shephard (2003) Realized power variation and stochastic volatility models. *Bernoulli* 9(2), 243–265.

Barndorff-Nielsen, O.E. & N. Shephard (2004a) Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* 72(3), 885–925.

Barndorff-Nielsen, O.E. & N. Shephard (2004b) Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2(1), 1–37.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, & J.A. Wellner (1998) *Efficient and Adapative Estimation for Semiparametric Models*. Springer-Verlag.

Bollerslev, T. & V. Todorov (2011) Estimation of jump tails. *Econometrica* 79(6), 1727–1783.

Broadie, M., M. Chernov, & M. Johannes (2007) Model specification and risk premia: Evidence from futures options. *The Journal of Finance* 62(3), 1453–1490.

Clément, E., S. Delattre, & A. Gloter (2013) An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochastic Processes and their Applications* 123(7), 2500–2521.

Comte, F. & E. Renault (1996) Long memory continuous time models. *Journal of Econometrics* 73(1), 101–149.

Comte, F. & E. Renault (1998) Long memory in continuous-time stochastic volatility models. *Mathematical Finance* 8(4), 291–323.

Delbaen, F. & W. Schachermayer (1994) A general version of the fundamental theorem of asset pricing. *Mathematische Annalen* 300, 463–520.

Ding, Z., C.W. Granger, & R.F. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1(1), 83–106.

Duffie, D., J. Pan, & K.J. Singleton (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68(6), 1343–1376.

Engle, R.F. (2004) Risk and volatility: Econometric models and financial practice. *American Economic Review* 94(3), 405–420.

Eraker, B. (2004) Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance* 59(3), 1367–1403.

Eraker, B., M. Johannes, & N. Polson (2003) The impact of jumps in volatility and returns. *The Journal of Finance* 58(3), 1269–1300.

Foster, D.P. & D.B. Nelson (1996) Continuous record asymptotics for rolling sample variance estimators. *Econometrica* 64(1), 139–174.

Geweke, J. & S. Porter-Hudak (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4(4), 221–238.

Granger, C. (1980) Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14(2), 227–238.

Harrison, M. & D. Kreps (1979) Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20(3), 381–408.

Jacod, J. (2008) Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and their Applications* 118(4), 517–559.

Jacod, J., Y. Li, & X. Zheng (2017) Statistical properties of microstructure noise. *Econometrica* 85(4), 1133–1174.

Jacod, J., Y. Li, & X. Zheng (2019) Estimating the integrated volatility with tick observations. *Journal of Econometrics* 208(1), 80–100.

Jacod, J. & P. Protter (2012) *Discretization of Processes*. Springer.

Jacod, J. & M. Rosenbaum (2013) Quarticity and other functionals of volatility: Efficient estimation. *Annals of Statistics* 41(3), 1462–1484.

Jacod, J. & M. Rosenbaum (2015) Estimation of volatility functionals: The case of a a square root *n* window. In P.K. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, & J. Teichmann (eds.), *Large Deviations and Asymptotic Methods in Finance*, pp. 559–590. Springer International Publishing.

Jacod, J. & A.N. Shiryaev (2003) *Limit Theorems for Stochastic Processes, 2nd edition*. Springer-Verlag.

Jeganathan, P. (1982) On the asymptotic theory of estimation when the limit of the log-likelihood is mixed normal. *Sankhya: The Indian Journal of Statistics, Series A (1961–2002)* 44(2), 173–212.

Jeganathan, P. (1983) Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. *Sankhya: The Indian Journal of Statistics, Series A (1961–2002)* 45(1), 66–87.

Kalnina, I. & D. Xiu (2017) Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *Journal of the American Statistical Association* 112(517), 384–396.

Kristensen, D. (2010) Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* 26(1), 60–93.

Li, J., V. Todorov, & G. Tauchen (2016) Inference theory on volatility functional dependencies. *Journal of Econometrics* 193(1), 17–34.

Li, J., V. Todorov, & G. Tauchen (2017) Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics* 200(1), 36–47.

Li, J. & D. Xiu (2016) Generalized method of integrated moments for high-frequency data. *Econometrica* 84(4), 1613–1633.

Mancini, C. (2001) Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari* LXIV, 19–47.

McCloskey, A. & P. Perron (2013) Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory* 29(6), 1196–1237.

Merton, R.C. (1980) On estimating the expected return on the market. *Journal of Financial Economics* 8(4), 323–361.

Mykland, P. & L. Zhang (2006) ANOVA for diffusions and Ito processes. *Annals of Statistics* 34(4), 1931–1963.

Mykland, P. & L. Zhang (2009) Inference for continuous semimartingales observed at high frequency. *Econometrica* 77(5), 1403–1445.

Nualart, D. (2005) *The Malliavin Calculus and Related Topics, 2nd edition*. Springer-Verlag.

Renault, E., C. Sarisoy, & B.J. Werker (2017) Efficient estimation of integrated volatility and related processes. *Econometric Theory* 33(2), 439–478.

Shepard, N. (2005) *Stochastic Volatility: Selected Readings (Advanced Texts in Econometrics)*. Oxford University Press.

Singleton, K.J. (2006) *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*. Princeton University Press.

Stein, C. (1956) Efficient nonparametric testing and estimation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 187–195. University of California Press.

Todorov, V. & G. Tauchen (2012) The realized Laplace transform of volatility. *Econometrica* 80(3), 1105–1127.

Todorov, V., G. Tauchen, & I. Grynkiv (2011) Realized Laplace transforms for estimation of jump diffusive volatility models. *Journal of Econometrics* 164(2), 367–381.

van der Vaart, A. & J. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer-Verlag.

Zhang, L. (2011) Estimating covariation: Epps effect and microstructure noise. *Journal of Econometrics* 160(1), 33–47.

Zhang, L., P.A. Mykland, & Y. Aït-Sahalia (2005) A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100(472), 1394–1411.

APPENDIX

In this appendix, we prove the results in the main text. Since the results with regular sampling (Theorems 1 and 2) are special cases of those under irregular sampling, with $\Delta_{n,i} = \Delta_n$ and $\Lambda_t = 1$ identically, we only need to prove Theorems 3 and 4.

A.1 Notation and Preliminary Estimates

We start by introducing and recalling some notations. For a generic matrix A, we denote by A^{ij} its (i,j) element. Recall that $\mathcal{H}_t$ is the smallest filtration containing $\mathcal{F}_t$ and with respect to which each $t(n,i)$ is a stopping time, and that $\mathcal{T}$ is the σ-field generated by $\{t(n,i) : i \geq 0\}$. As shown by Lem. A.1 of Jacod et al. (2017), the semimartingale structure of X and σ_1 are preserved with respect to the extension from $(\mathcal{F}_t)$ to $(\mathcal{H}_t)$. We also consider a filtration $\mathcal{H}'_{t(n,i)}$ that is generated by $\mathcal{H}_{t(n,i)}$ and the variable $\Delta_{n,i+1}$. For notational simplicity, we write $\mathbb{E}_i^n[\cdot]$, $\mathbb{E}_i'^n[\cdot]$, $\mathbb{E}_{\mathcal{T}}[\cdot]$ and $\sum_i$ in place of $\mathbb{E}[\cdot|\mathcal{H}_{t(n,i)}]$, $\mathbb{E}[\cdot|\mathcal{H}'_{t(n,i)}]$, $\mathbb{E}[\cdot|\mathcal{T}]$ and $\sum_{i=0}^{N_{n,T}-k_n}$, respectively. Below, we use K to denote a generic constant that may change from line to line; we sometimes emphasize the dependence of this constant on some parameter q by writing K_q.

By a standard localization procedure (see Sect. 4.4.1 in Jacod and Protter, 2012), we can strengthen Assumptions H, C, IS, IR, LM$'$, and LAMN$'$ without loss of generality by further assuming the following for all $m \geq 1$: (i) $T_m = \infty$; (ii) $\mathcal{J}_m(\cdot) = \mathcal{J}_1(\cdot)$; (iii) $K_m = K_1$ and $K_{m,q} = K_{1,q}$ for all $q > 0$; (iv) $\mathcal{K}_m = \mathcal{K}_1$; (v) the processes $b, \sigma_1, \sigma_2, b^{(\sigma_1)}, \sigma^{(\sigma_1)}, q, \Lambda$ and Λ^{-1} are bounded; (vi) $\lambda_{\min}(c_t)$ is bounded away from 0 in Assumption LAMN$'$. In the proofs, below, it is understood that the strong versions of these assumptions are invoked.

We further introduce some quantities that are useful in the proofs. We denote the continuous part of X by

$$X'_t \equiv \int_0^t b_s ds + \int_0^t \sigma_s dW_s,$$

and complement the definition in (4.2) with

$$\hat{c}'_{t(n,i)} \equiv \frac{1}{k_n} \sum_{j=1}^{k_n} \frac{\Delta_{i+j}^n X' \Delta_{i+j}^n X'^\top}{\Delta_{n,i+j}}.$$

We then set

$$\begin{cases} \alpha_{n,i} \equiv \dfrac{\Delta_i^n X' \Delta_i^n X'^\top}{\Delta_{n,i}} - c_{t(n,i-1)}, \\ \tilde{c}_{n,i} \equiv \hat{c}'_{t(n,i)} - c_{t(n,i)} = \dfrac{1}{k_n} \displaystyle\sum_{j=1}^{k_n} (\alpha_{n,i+j} + (c_{t(n,i+j-1)} - c_{t(n,i)})). \end{cases} \quad \textbf{(A.1)}$$

For any process Z, we define $\eta_{t,s}(Z) \equiv \sup_{v \in (0,s]} ||Z_{t+v} - Z_t||^2$ and

$$\eta_{i,j}^n(Z) \equiv \sqrt{\mathbb{E}_{i-1}^n \left[\eta_{t(n,i-1),t(n,i+j-1)-t(n,i-1)}(Z)\right]}, \qquad \eta_i^n(Z) \equiv \eta_{i,k_n}^n(Z),$$

$$\eta_{i,j}'^n(Z) \equiv \sqrt{\mathbb{E}_{i-1}'^n \left[\eta_{t(n,i-1),t(n,i+j-1)-t(n,i-1)}(Z)\right]}.$$

Define the spot covariation of the continuous martingale parts of X and σ_1 and that for X and c_1, respectively, as

$$\left(c_t^{(X,\sigma_1)}\right)^{ijk} = \sum_{l=1}^{d} \sigma_t^{il}\left(\sigma_t^{(\sigma_1)}\right)^{jkl}, \quad \left(c_t^{(X,c_1)}\right)^{ijk} = \sum_{l=1}^{d} \sigma_t^{il}\left(\sigma_t^{(c_1)}\right)^{jkl}.$$

Note that both $c^{(X,\sigma_1)}$ and $c^{(X,c_1)}$ are càdlàg adapted. Below, we often use the following elementary identity for the increment of the process c: for $s > t$ (the order of the two times is crucial),

$$\begin{aligned} c_s - c_t = c_{1,s} - c_{1,t} + c_{2,s} - c_{2,t} \\ + \left(\sigma_{2,s} - \sigma_{2,t}\right)\sigma_{1,s}^{\top} + \sigma_{1,s}\left(\sigma_{2,s} - \sigma_{2,t}\right)^{\top} \\ + \left(\sigma_{1,s} - \sigma_{1,t}\right)\sigma_{2,t}^{\top} + \sigma_{2,t}\left(\sigma_{1,s} - \sigma_{1,t}\right)^{\top}, \end{aligned} \tag{A.2}$$

where $c_{j,t} = \sigma_{j,t}\sigma_{j,t}^{\top}$ for $j = 1,2$. We now collect some preliminary estimates that are used in the sequel.

LEMMA 1. *Under (the strengthened versions of) Assumptions H, C, IS, IR, LM′, the following inequalities hold for $p \geq 0$:*

(a) *for each $i \geq 0$, $u \geq 1$, and $q \geq 0$,*

$$\begin{aligned} &\mathbb{E}_{i-1}^{n}\left[\left\|\sigma_{1,t(n,i+u-1)} - \sigma_{1,t(n,i-1)}\right\|^q \Delta_{n,i}^{p}\right] \leq K_{p,q}(u\Delta_n)^{(q\wedge 2)/2}\Delta_n^p, \\ &\mathbb{E}_{i-1}^{n}\left[\left\|\alpha_{n,i}\right\|^q\right] \leq K_q, \\ &\mathbb{E}\left[\left\|\sigma_{2,t(n,i+u-1)} - \sigma_{2,t(n,i-1)}\right\|^q \Delta_{n,i}^{p}\right] \leq K_{p,q}\left(u\Delta_n\right)^{(q\wedge 2)(1/2+\epsilon)}\Delta_n^p, \\ &\mathbb{E}\left[\left\|c_{t(n,i+u-1)} - c_{t(n,i-1)}\right\|^q \Delta_{n,i}^{p}\right] \leq K_{p,q}(u\Delta_n)^{(q\wedge 2)/2}\Delta_n^p; \end{aligned}$$

(b) *we have*

$$\left\|\mathbb{E}_{i-1}^{\prime n}\left[\alpha_{n,i}\right]\right\| \leq K\Delta_{n,i} + \Delta_{n,i}^{1/2}\eta_{i,1}^{\prime n}(b) + K\mathbb{E}_{i-1}^{\prime n}\left[\Delta_{n,i}^{-1}\int_{t(n,i-1)}^{t(n,i)} ||\sigma_{2,s} - \sigma_{2,t(n,i-1)}||ds\right];$$

(c) *for any $q \geq 1$, $1 \leq u \leq k_n$,*

$$\begin{aligned} \mathbb{E}\left[\left\|\mathbb{E}_{i+u-1}^{\prime n}\left[\alpha_{n,i+u}\right]\right\|^q \Delta_{n,i+1}^{p}\right] \leq K_{p,q}\Delta_n^{(q/2)+p} + K_{p,q}\Delta_n^{(q\wedge 2)(1/2+\epsilon)+p}, \\ \mathbb{E}\left|\mathbb{E}_{i+u-1}^{\prime n}[\alpha_{n,i+u}^{lm}]\Delta_{n,i+1}\right| \leq K\Delta_n^{3/2}\left(\Delta_n^{1/2} + \mathbb{E}\left[\eta_{i+1}^{n}(b)\right] + \Delta_n^{\epsilon}\right); \end{aligned}$$

(d) *for any $q \geq 1$,*

$$\mathbb{E}\left[\left\|\tilde{c}_{n,i}\right\|^q \Delta_{n,i+1}^{p}\right] \leq K_{p,q}\Delta_n^p(k_n^{-q/2} + (k_n\Delta_n)^{(q\wedge 2)/2});$$

(e) *for any $u \geq 1$,*

$$\begin{aligned} &\mathbb{E}\left|\mathbb{E}_{i+u-1}^{\prime n}\left[\left(\alpha_{n,i+u}^{jk}\alpha_{n,i+u}^{lm} - (c_{t(n,i+u-1)}^{jl}c_{t(n,i+u-1)}^{km} + c_{t(n,i+u-1)}^{jm}c_{t(n,i+u-1)}^{kl})\right)\Delta_{n,i+1}^{p}\right]\right| \\ &\leq K_p\Delta_n^{(1/2)+p}. \end{aligned}$$

Proof. (a) When $u = 1$ and $q = 2$, the first assertion holds because

$$\begin{aligned} &\mathbb{E}_{i-1}^{n}\left[\left\|\sigma_{1,t(n,i)}-\sigma_{1,t(n,i-1)}\right\|^{2}\Delta_{n,i}^{p}\right] \\ &\quad = K\mathbb{E}_{i-1}^{n}\left[\mathbb{E}_{i-1}^{\prime n}\left[\left\|\sigma_{1,t(n,i)}-\sigma_{1,t(n,i-1)}\right\|^{2}\right]\Delta_{n,i}^{p}\right] \\ &\quad \leq K\mathbb{E}_{i-1}^{n}\left[\Delta_{n,i}^{1+p}\right] \leq K_{p}\Delta_{n}^{1+p}, \end{aligned} \tag{A.3}$$

where the first equality is by repeated conditioning; the first inequality is obtained by using a standard estimate for Itô semimartingales, recognizing that $\Delta_{n,i}$ is independent of $\mathcal{F}_T$ conditional on $\mathcal{H}_{t(n,i-1)}$; the second inequality is by Assumption IR. When $u > 1$, we observe

$$\begin{aligned} &\mathbb{E}_{i-1}^{n}\left[\left\|\sigma_{1,t(n,i+u-1)}-\sigma_{1,t(n,i-1)}\right\|^{2}\Delta_{n,i}^{p}\right] \\ &\quad \leq K\mathbb{E}_{i-1}^{n}\left[\Delta_{n,i}^{p}\mathbb{E}_{i}^{n}\left[\left\|\sigma_{1,t(n,i+u-1)}-\sigma_{1,t(n,i)}\right\|^{2}\right]\right] \\ &\qquad +K\mathbb{E}_{i-1}^{n}\left[\left\|\sigma_{1,t(n,i)}-\sigma_{1,t(n,i-1)}\right\|^{2}\Delta_{n,i}^{p}\right] \\ &\quad \leq Ku\Delta_{n}^{1+p}, \end{aligned} \tag{A.4}$$

where the first inequality is by the C_r inequality and repeated conditioning; the second inequality is derived by using the Itô semimartingale property of σ_1, Assumption IR and (A.3). Hence, the first assertion holds for $q = 2$. When $q \neq 2$, note that since σ_1 is bounded,

$$\begin{aligned} &\mathbb{E}_{i-1}^{n}\left[\left\|\sigma_{1,t(n,i+u-1)}-\sigma_{1,t(n,i-1)}\right\|^{q}\Delta_{n,i}^{p}\right] \\ &\leq K_{q}\mathbb{E}_{i-1}^{n}\left[\left\|\sigma_{1,t(n,i+u-1)}-\sigma_{1,t(n,i-1)}\right\|^{q\wedge 2}\Delta_{n,i}^{p}\right]. \end{aligned}$$

The first assertion for a general choice of q can be obtained by first applying (A.4) with p replaced with $2p/(q \wedge 2)$ and then using Jensen's inequality. The second assertion of part (a) can be proved in a very similar (actually simpler) way.

Turning to the third assertion, we note that,

$$\begin{aligned} \mathbb{E}_{\mathcal{T}}\left[\left\|\sigma_{2,t(n,i+u-1)}-\sigma_{2,t(n,i-1)}\right\|^{2}\right] &\leq K\left(t\left(n,i+u-1\right)-t\left(n,i-1\right)\right)^{1+2\epsilon} \\ &\leq K\left(\sum_{j=1}^{u-1}\Delta_{n,i+j}\right)^{1+2\epsilon}+K\Delta_{n,i}^{1+2\epsilon} \\ &\leq K\left(u-1\right)^{2\epsilon}\sum_{j=1}^{u-1}\Delta_{n,i+j}^{1+2\epsilon}+K\Delta_{n,i}^{1+2\epsilon}, \end{aligned} \tag{A.5}$$

where the first inequality is by Assumption LM′, the second inequality is by the C_r inequality and the third inequality is by Hölder's inequality. Therefore,

$$\begin{aligned} &\mathbb{E}\left[\left\|\sigma_{2,t(n,i+u-1)}-\sigma_{2,t(n,i-1)}\right\|^{2}\Delta_{n,i}^{p}\right] \\ &\quad \leq K\left(u-1\right)^{2\epsilon}\mathbb{E}\left[\sum_{j=1}^{u-1}\mathbb{E}_{i+j-1}^{n}\left[\Delta_{n,i+j}^{1+2\epsilon}\right]\Delta_{n,i}^{p}\right]+K\mathbb{E}\left[\Delta_{n,i}^{1+2\epsilon+p}\right] \\ &\quad \leq K_{p}\left(\left(u-1\right)^{1+2\epsilon}+1\right)\Delta_{n}^{1+2\epsilon+p} \\ &\quad \leq K_{p}\left(u\Delta_{n}\right)^{1+2\epsilon}\Delta_{n}^{p}, \end{aligned}$$

where the first inequality follows from (A.5) and repeated conditioning; the second inequality is by Assumption IR; the third inequality is obvious. This finishes the proof of the third assertion of part (a) when $q = 2$. By the boundedness of σ_2 and Jensen's inequality, we can further prove the case with a general q. The fourth assertion follows readily from the first and the third assertions.

(b) By Itô's formula, $\alpha_{n,i}^{jm} \equiv (\Delta_i^n X'^j \Delta_i^n X'^m)/\Delta_{n,i} - c_{t(n,i-1)}^{jm} = A_{n,i}^{jm} + B_{n,i}^{jm} + R_{n,i}$, where

$$
\begin{aligned}
A_{n,i}^{jm} \equiv\ & b_{t(n,i-1)}^{j} \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (X_s'^m - X_{t(n,i-1)}'^m) ds \\
& + b_{t(n,i-1)}^{m} \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (X_s'^j - X_{t(n,i-1)}'^j) ds \\
& + \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (X_s'^m - X_{t(n,i-1)}'^m)(b_s^j - b_{t(n,i-1)}^j) ds \\
& + \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (X_s'^j - X_{t(n,i-1)}'^j)(b_s^m - b_{t(n,i-1)}^m) ds,
\end{aligned}
\tag{A.6}
$$

$B_{n,i} \equiv \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (c_s - c_{t(n,i-1)}) ds$ and the remaining term $R_{n,i}$ is a martingale component that satisfies $\mathbb{E}_{i-1}'^n[R_{n,i}] = 0$ (the latter can be derived under Assumption IR).

Since $\Delta_{n,i}$ is independent of $\mathcal{F}_T$ conditional on $\mathcal{H}_{t(n,i-1)}$, we deduce that

$$
\left| \mathbb{E}_{i-1}'^n \left[\Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (X_s'^m - X_{t(n,i-1)}'^m) ds \right] \right| \leq K \Delta_{n,i}
\tag{A.7}
$$

by using the fact that X' is an Itô semimartingale. In addition, we observe that

$$
\begin{aligned}
& \left| \mathbb{E}_{i-1}'^n \left[\Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (X_s'^m - X_{t(n,i-1)}'^m)(b_s^j - b_{t(n,i-1)}^j) ds \right] \right| \\
& \leq \left(\mathbb{E}_{i-1}'^n \left[\Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (X_s'^m - X_{t(n,i-1)}'^m)^2 ds \right] \right)^{1/2} \\
& \qquad \times \left(\mathbb{E}_{i-1}'^n \left[\Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (b_s^j - b_{t(n,i-1)}^j)^2 ds \right] \right)^{1/2} \\
& \leq K \Delta_{n,i}^{1/2} \eta_{i,1}'^n(b),
\end{aligned}
\tag{A.8}
$$

where the first inequality is by the Cauchy–Schwarz inequality; the second inequality holds because of Assumption IR, the standard estimate $\mathbb{E}_{i-1}'^n[(X_s'^m - X_{t(n,i-1)}'^m)^2] \leq K(s - t(n,i-1))$ and the definition of $\eta_{i,1}'^n(b)$. From (A.6), (A.7), and (A.8), we deduce that

$\left\| \mathbb{E}_{i-1}'^n \left[A_{n,i} \right] \right\| \leq K \Delta_{n,i}^{1/2} \eta_{i,1}'^n(b) + K \Delta_{n,i}$.

To show the assertion of part (b), it remains to show that

$$
\left\| \mathbb{E}_{i-1}'^n [B_{n,i}] \right\| \leq K \Delta_{n,i} + K \mathbb{E}_{i-1}'^n \left[\Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} ||\sigma_{2,s} - \sigma_{2,t(n,i-1)}|| ds \right].
\tag{A.9}
$$

In view of (A.2), we can decompose $B_{n,i} = \sum_{k=1}^{6} B_{k,n,i}$ where

$$\begin{cases} B_{1,n,i} \equiv \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (c_{1,s} - c_{1,t(n,i-1)}) ds, \; B_{2,n,i} \equiv \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (c_{2,s} - c_{2,t(n,i-1)}) ds, \\ B_{3,n,i} \equiv \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (\sigma_{2,s} - \sigma_{2,t(n,i-1)}) \sigma_{1,s}^{\top} ds, \; B_{4,n,i} \equiv B_{3,n,i}^{\top}, \\ B_{5,n,i} \equiv \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} (\sigma_{1,s} - \sigma_{1,t(n,i-1)}) \sigma_{2,t(n,i-1)}^{\top} ds, \; B_{6,n,i} \equiv B_{5,n,i}^{\top}. \end{cases}$$

Since the process c_1 is an Itô semimartingale, $\|\mathbb{E}_{i-1}^{\prime n}[B_{1,n,i}]\| \leq K\Delta_{n,i}$ by an argument that is similar to (A.7). Similarly, noting that $\sigma_{2,t(n,i-1)}$ is bounded and $\mathcal{H}_{t(n,i-1)}$ -measurable, we deduce $\|\mathbb{E}_{i-1}^{\prime n}[(B_{5,n,i} + B_{6,n,i})]\| \leq K\Delta_{n,i}$. In addition, we note that $\|c_{2,s} - c_{2,t(n,i-1)}\| \leq K\|\sigma_{2,s} - \sigma_{2,t(n,i-1)}\|$, which further implies

$$\left\|\mathbb{E}_{i-1}^{\prime n}[B_{2,n,i}]\right\| \leq K\mathbb{E}_{i-1}^{\prime n}\left[\Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} ||\sigma_{2,s} - \sigma_{2,t(n,i-1)}|| ds\right].$$

Similarly, we can show that $\|\mathbb{E}_{i-1}^{\prime n}[B_{k,n,i}]\|$ is bounded by the majorant side of the above display for $k = 3$ and 4. Combining the above estimates, we obtain (A.9) as wanted.

(c) Note that $\eta_{i,1}^{\prime n}(b)$ is uniformly bounded. By part (b), the C_r inequality and Jensen's inequality,

$$\left\|\mathbb{E}_{i+u-1}^{\prime n}\left[\alpha_{n,i+u}\right]\right\|^q \leq K_q \Delta_{n,i+u}^{q/2}$$
$$+ K_q \mathbb{E}_{i+u-1}^{\prime n}\left[\Delta_{n,i+u}^{-1} \int_{t(n,i+u-1)}^{t(n,i+u)} ||\sigma_{2,s} - \sigma_{2,t(n,i+u-1)}||^q ds\right].$$

Hence,

$$\begin{aligned} &\mathbb{E}\left[\left\|\mathbb{E}_{i+u-1}^{\prime n}[\alpha_{n,i+u}]\right\|^q \Delta_{n,i+1}^{p}\right] \\ &\quad \leq K_{p,q}\Delta_n^{(q/2)+p} + K_q \mathbb{E}\left[\Delta_{n,i+1}^{p} \Delta_{n,i+u}^{-1} \int_{t(n,i+u-1)}^{t(n,i+u)} ||\sigma_{2,s} - \sigma_{2,t(n,i+u-1)}||^q ds\right] \\ &\quad \leq K_{p,q}\Delta_n^{(q/2)+p} + K_{p,q}\Delta_n^{(q\wedge 2)(1/2+\epsilon)+p}, \end{aligned}$$

where the second inequality is derived using arguments similar to the proof of the third inequality of part (a).

The second claim can be proved as follows:

$$\begin{aligned} &\mathbb{E}\left|\mathbb{E}_{i+u-1}^{\prime n}[\alpha_{n,i+u}^{lm}]\Delta_{n,i+1}\right| \\ &\quad \leq K\Delta_n^2 + K\mathbb{E}\left[\Delta_{n,i+u}^{1/2}\Delta_{n,i+1}\eta_{i+u,1}^{\prime n}(b)\right] \\ &\qquad + K\mathbb{E}\left[\Delta_{n,i+1}\Delta_{n,i+u}^{-1}\int_{t(n,i+u-1)}^{t(n,i+u)} ||\sigma_{2,s} - \sigma_{2,t(n,i+u-1)}|| ds\right] \\ &\quad \leq K\Delta_n^2 + K\Delta_n^{3/2}\mathbb{E}\left[\sqrt{\mathbb{E}_i^n\left[\left(\eta_{i+u,1}^{\prime n}(b)\right)^2\right]}\right] + K\Delta_n^{3/2+\epsilon} \\ &\quad \leq K\Delta_n^2 + K\Delta_n^{3/2}\mathbb{E}\left[\eta_{i+1}^n(b)\right] + K\Delta_n^{3/2+\epsilon}, \end{aligned}$$

where the first inequality is by Lemma 1(b) and Assumption IR; the second inequality is by the Cauchy–Schwarz inequality and Assumption LM′; the last inequality follows from the definition of $\eta_{i+1}^n(b)$.

(d) From the definition of $\tilde{c}_{n,i}$ in (A.1), we deduce

$$\mathbb{E}\left[\left\|\tilde{c}_{n,i}\right\|^q \Delta_{n,i+1}^p\right] \leq K_q \mathbb{E}\left[\left\|\frac{1}{k_n}\sum_{j=1}^{k_n}(c_{t(n,i+j-1)} - c_{t(n,i)})\right\|^q \Delta_{n,i+1}^p\right]$$
$$+ K_q \mathbb{E}\left[\left\|\frac{1}{k_n}\sum_{j=1}^{k_n}\alpha_{n,i+j}\right\|^q \Delta_{n,i+1}^p\right]. \tag{A.10}$$

The first term on the right-hand side of (A.10) can be bounded by

$$K_q \mathbb{E}\left[\frac{1}{k_n}\sum_{j=1}^{k_n}\left\|(c_{t(n,i+j-1)} - c_{t(n,i)}\right\|^q \Delta_{n,i+1}^p\right] \tag{A.11}$$

by Jensen's inequality. By Lemma 1(a) and the triangle inequality, this term can be further bounded by $K_{p,q}\,(k_n\Delta_n)^{(q\wedge 2)/2}\,\Delta_n^p$.

Turning to the second term on the right-hand side of (A.10), we first note that

$$\mathbb{E}\left[\left\|\frac{1}{k_n}\sum_{j=1}^{k_n}\alpha_{n,i+j}\right\|^q \Delta_{n,i+1}^p\right] \leq K_q \mathbb{E}\left[\left\|A_{1,n,i}\right\|^q \Delta_{n,i+1}^p\right] + K_q \mathbb{E}\left[\left\|A_{2,n,i}\right\|^q \Delta_{n,i+1}^p\right], \tag{A.12}$$

where

$$A_{1,n,i} \equiv \frac{1}{k_n}\sum_{j=1}^{k_n}\left(\alpha_{n,i+j} - \mathbb{E}_{i+j-1}^{\prime n}\left[\alpha_{n,i+j}\right]\right), \quad A_{2,n,i} \equiv \frac{1}{k_n}\sum_{j=1}^{k_n}\mathbb{E}_{i+j-1}^{\prime n}\left[\alpha_{n,i+j}\right].$$

Since $A_{1,n,i}$ is an average of martingale differences, by the Burkholder–Davis–Gundy inequality, Jensen's inequality, and the estimate $\mathbb{E}[\|\alpha_{n,i+j}\|^{2q}] \leq K_q$ (see Lemma 1(a)), we deduce $\mathbb{E}[\|A_{1,n,i}\|^{2q}] \leq K_q k_n^{-q}$. Then, by the Cauchy–Schwarz inequality and Assumption IR,

$$\mathbb{E}\left[\left\|A_{1,n,i}\right\|^q \Delta_{n,i+1}^p\right] \leq K_{p,q} k_n^{-q/2} \Delta_n^p.$$

Furthermore, by Jensen's inequality and part (c) (with the latter applied with $p = 0$),

$$\mathbb{E}\left[\left\|A_{2,n,i}\right\|^q \Delta_{n,i+1}^p\right] \leq K_q k_n^{-1} \sum_{j=1}^{k_n} \mathbb{E}\left[\left\|\mathbb{E}_{i+j-1}^{\prime n}\left[\alpha_{n,i+j}\right]\right\|^q \Delta_{n,i+1}^p\right]$$
$$\leq K_{p,q}\Delta_n^p\left(\Delta_n^{q/2} + \Delta_n^{(1/2+\epsilon)(q\wedge 2)}\right) \leq K_{p,q}\Delta_n^p\,(k_n\Delta_n)^{(q\wedge 2)/2}.$$

The assertion of part (d) then readily follows from these estimates.

(e) Recall that $\alpha_{n,i} = \left(\Delta_i^n X' \Delta_i^n X'^\top\right)/\Delta_{n,i} - c_{t(n,i-1)}$. Some straightforward algebra yields

$$
\begin{aligned}
&\alpha_{n,i}^{jk}\alpha_{n,i}^{lm} - \left(c_{t(n,i-1)}^{jl}c_{t(n,i-1)}^{km} + c_{t(n,i-1)}^{jm}c_{t(n,i-1)}^{kl}\right) \\
&\quad = \Delta_i^n X'^j \Delta_i^n X'^k \Delta_i^n X'^l \Delta_i^n X'^m / \Delta_{n,i}^2 \\
&\qquad - \left(c_{t(n,i-1)}^{jk}c_{t(n,i-1)}^{lm} + c_{t(n,i-1)}^{jl}c_{t(n,i-1)}^{km} + c_{t(n,i-1)}^{jm}c_{t(n,i-1)}^{kl}\right) \\
&\qquad - \alpha_{n,i}^{jk}c_{t(n,i-1)}^{lm} - \alpha_{n,i}^{lm}c_{t(n,i-1)}^{jk}.
\end{aligned}
$$

Following steps similar to Lem. 4.1 of Jacod and Rosenbaum (2013) and using Lemma 1(a), we deduce that

$$
\begin{aligned}
&\mathbb{E}\Big|\mathbb{E}_{i-1}^{\prime n}[(\Delta_i^n X'^j \Delta_i^n X'^k \Delta_i^n X'^l \Delta_i^n X'^m / \Delta_{n,i}^2 - (c_{t(n,i-1)}^{jk}c_{t(n,i-1)}^{lm} \\
&\quad + c_{t(n,i-1)}^{jl}c_{t(n,i-1)}^{km} + c_{t(n,i-1)}^{jm}c_{t(n,i-1)}^{kl}))\Delta_{n,i-u+1}^p]\Big| \le K\Delta_n^{(1/2)+p}.
\end{aligned}
$$

By part (c) with $q = 1$,

$$
\mathbb{E}\left|\mathbb{E}_{i-1}^{\prime n}\left[\left(\alpha_{n,i}^{jk}c_{t(n,i-1)}^{lm} + \alpha_{n,i}^{lm}c_{t(n,i-1)}^{jk}\right)\Delta_{n,i-u+1}^p\right]\right| \le K_p\Delta_n^{(1/2)+p}.
$$

The assertion of part (e) then readily follows these estimates.

□

A.2 Proofs of Theorem 1 and Theorem 3

We now prove Theorems 1 and 3. Since the former is a special case of the latter (with $\Delta_{n,i} = \Delta_n$ and $\Lambda_t = 1$ identically), we only need to prove the latter. The proof relies on a technical calculation, which we single out as Lemma 2. This lemma generalizes Lem. 4.3 of Jacod and Rosenbaum, 2013 in the current setting with general LMIS volatility dynamics and irregular sampling.

LEMMA 2. *Under the conditions of Theorem 3(a),*

$$
\begin{aligned}
&\mathbb{E}\left|\mathbb{E}_i^n\left[\left(\tilde{c}_{n,i}^{jk}\tilde{c}_{n,i}^{lm} - \frac{1}{k_n}\left(c_{t(n,i)}^{jl}c_{t(n,i)}^{km} + c_{t(n,i)}^{jm}c_{t(n,i)}^{kl}\right)\right)\Delta_{n,i+1}\right]\right| \\
&\quad \le K\Delta_n^{3/2}\left(k_n^{-1/2} + k_n\Delta_n^{1/2} + k_n^{1/2+\epsilon}\Delta_n^{\epsilon} + \mathbb{E}\left[\eta_{i+1}^n\right]\right),
\end{aligned}
$$

where $\eta_i^n \equiv \eta_i^n(b) + \eta_i^n(c^{(X,c_1)}) + \eta_i^n(c^{(X,\sigma_1)})$.

Proof. *Step 1*. We outline the proof in this step, while leaving some technical estimates in step 2 and step 3 below. For notational simplicity, we denote $\xi_i^n \equiv c_{t(n,i-1)}^{jl}c_{t(n,i-1)}^{km} + c_{t(n,i-1)}^{jm}c_{t(n,i-1)}^{kl}$ and $\zeta_{i,u}^n \equiv \alpha_{n,i+u} + (c_{t(n,i+u-1)} - c_{t(n,i)})$, so we can rewrite

$$
\tilde{c}_{n,i} = \frac{1}{k_n}\sum_{u=1}^{k_n}\zeta_{i,u}^n.
$$

We consider the decomposition $(\tilde{c}^{jk}_{n,i}\tilde{c}^{lm}_{n,i} - k_n^{-1}(c^{jl}_{t(n,i)}c^{km}_{t(n,i)} + c^{jm}_{t(n,i)}c^{kl}_{t(n,i)}))\Delta_{n,i+1} = A_{n,i} + B_{n,i} + B'_{n,i}$, where

$$\begin{cases} A_{n,i} \equiv \left(\frac{1}{k_n^2}\sum_{u=1}^{k_n}\zeta^{n,jk}_{i,u}\zeta^{n,lm}_{i,u} - \frac{1}{k_n}\xi^n_{i+1}\right)\Delta_{n,i+1}, \\ B_{n,i} \equiv \frac{\Delta_{n,i+1}}{k_n^2}\sum_{u=1}^{k_n-1}\sum_{v=u+1}^{k_n}\zeta^{n,lm}_{i,u}\zeta^{n,jk}_{i,v}, \quad B'_{n,i} \equiv \frac{\Delta_{n,i+1}}{k_n^2}\sum_{u=1}^{k_n-1}\sum_{v=u+1}^{k_n}\zeta^{n,jk}_{i,u}\zeta^{n,lm}_{i,v}. \end{cases} \tag{A.13}$$

We first consider $A_{n,i}$. Note that

$$\left|\zeta^{n,jk}_{i,u}\zeta^{n,lm}_{i,u} - \alpha^{jk}_{n,i+u}\alpha^{lm}_{n,i+u}\right| \leq K\left\|\alpha_{n,i+u}\right\|\left\|c_{t(n,i+u-1)} - c_{t(n,i)}\right\| + K\left\|c_{t(n,i+u-1)} - c_{t(n,i)}\right\|^2.$$

Hence, by the triangle inequality and the estimates in Lemma 1(a),

$$\begin{aligned} &\mathbb{E}\left|\mathbb{E}^n_i\left[A_{n,i} - \left(\frac{1}{k_n^2}\sum_{u=1}^{k_n}\alpha^{jk}_{n,i+u}\alpha^{lm}_{n,i+u} - \frac{1}{k_n}\xi^n_{i+1}\right)\Delta_{n,i+1}\right]\right| \\ &\leq \frac{K}{k_n^2}\sum_{u=1}^{k_n}\mathbb{E}\left[\left\|\alpha_{n,i+u}\right\|\left\|c_{t(n,i+u-1)} - c_{t(n,i)}\right\|\Delta_{n,i+1} + \left\|c_{t(n,i+u-1)} - c_{t(n,i)}\right\|^2\Delta_{n,i+1}\right] \\ &\leq K\Delta_n^{3/2}/k_n^{1/2}, \end{aligned} \tag{A.14}$$

and

$$\begin{aligned} &\mathbb{E}\left|\mathbb{E}^n_i\left[\left(\frac{1}{k_n^2}\sum_{u=1}^{k_n}\xi^n_{i+u} - \frac{1}{k_n}\xi^n_{i+1}\right)\Delta_{n,i+1}\right]\right| \\ &\leq \frac{1}{k_n^2}\sum_{u=1}^{k_n}\mathbb{E}\left[\left|\xi^n_{i+u} - \xi^n_{i+1}\right|\Delta_{n,i+1}\right] \leq K\Delta_n^{3/2}/k_n^{1/2}. \end{aligned} \tag{A.15}$$

By the triangle inequality, Lemma 1(e) and the fact that the conditional expectation operator is a contraction, we have

$$\mathbb{E}\left|\mathbb{E}^n_i\left[\frac{1}{k_n^2}\sum_{u=1}^{k_n}\left(\alpha^{jk}_{n,i+u}\alpha^{lm}_{n,i+u} - \xi^n_{i+u}\right)\Delta_{n,i+1}\right]\right| \leq K\Delta_n^{3/2}/k_n. \tag{A.16}$$

From (A.14), (A.15), and (A.16), we deduce

$$\mathbb{E}\left|\mathbb{E}^n_i\left[A_{n,i}\right]\right| \leq K\Delta_n^{3/2}/k_n^{1/2}.$$

In view of the symmetry between $B_{n,i}$ and $B'_{n,i}$ defined in (A.13), it remains to show that

$$\mathbb{E}\left|\mathbb{E}^n_i\left[B_{n,i}\right]\right| \leq K\Delta_n^{3/2}\left(k_n\Delta_n^{1/2} + k_n^{1/2+\epsilon}\Delta_n^{\epsilon} + \mathbb{E}\left[\eta^n_{i+1}\right]\right). \tag{A.17}$$

To do so, we first decompose, for $1 \leq u < v \leq k_n$,

$$\mathbb{E}^n_i\left[\zeta^{n,lm}_{i,u}\zeta^{n,jk}_{i,v}\Delta_{n,i+1}\right] = Z_{1,n,i} + Z_{2,n,i} + Z_{3,n,i} + Z_{4,n,i}, \tag{A.18}$$

where

$$\left\{\begin{array}{l} Z_{1,n,i} \equiv \mathbb{E}_i^n\left[\zeta_{i,u}^{n,lm}\Delta_{n,i+1}\left(\mathbb{E}_{i+u}^n\left(\zeta_{i,v}^{n,jk}\right)-\left(c_{t(n,i+u)}^{jk}-c_{t(n,i)}^{jk}\right)\right)\right], \\ Z_{2,n,i} \equiv \mathbb{E}_i^n\left[\alpha_{n,i+u}^{lm}\left(c_{t(n,i+u)}^{jk}-c_{t(n,i+u-1)}^{jk}\right)\Delta_{n,i+1}\right], \\ Z_{3,n,i} \equiv \mathbb{E}_i^n\left[\alpha_{n,i+u}^{lm}\left(c_{t(n,i+u-1)}^{jk}-c_{t(n,i)}^{jk}\right)\Delta_{n,i+1}\right], \\ Z_{4,n,i} \equiv \mathbb{E}_i^n\left[\left(c_{t(n,i+u-1)}^{lm}-c_{t(n,i)}^{lm}\right)\left(c_{t(n,i+u)}^{jk}-c_{t(n,i)}^{jk}\right)\Delta_{n,i+1}\right]. \end{array}\right. \quad \textbf{(A.19)}$$

We now provide estimates for the variables in (A.19). In steps 2 and 3, we shall show, respectively, that

$$\mathbb{E}\left|Z_{1,n,i}\right| \leq K\Delta_n^{3/2}\left(\mathbb{E}[\eta_{i+1}^n(b)]+k_n\Delta_n^{1/2}+k_n^{1/2+\epsilon}\Delta_n^{\epsilon}\right), \quad \textbf{(A.20)}$$

$$\mathbb{E}\left|Z_{2,n,i}\right| \leq K\Delta_n^{3/2}\left(\mathbb{E}[\eta_{i+1}^n(c^{(X,c_1)})]+\mathbb{E}[\eta_{i+1}^n(c^{(X,\sigma_1)})]+\Delta_n^{\epsilon}+\Delta_n^{1/2}\right). \quad \textbf{(A.21)}$$

Next, we observe

$$\begin{aligned} \mathbb{E}\left|Z_{3,n,i}\right| &= \mathbb{E}\left|\mathbb{E}_i^n\left[(c_{t(n,i+u-1)}^{jk}-c_{t(n,i)}^{jk})\mathbb{E}_{i+u-1}^{\prime n}[\alpha_{n,i+u}^{lm}]\Delta_{n,i+1}\right]\right| \\ &\leq K\mathbb{E}\left|\mathbb{E}_{i+u-1}^{\prime n}[\alpha_{n,i+u}^{lm}]\Delta_{n,i+1}\right| \\ &\leq K\Delta_n^{3/2}\left(\mathbb{E}[\eta_{i+1}^n(b)]+\Delta_n^{1/2}+\Delta_n^{\epsilon}\right), \end{aligned}$$

where the equality is by repeated conditioning; the first inequality holds because c is bounded; the second inequality follows directly from the second part of Lemma 1(c). Finally, by the Cauchy–Schwarz inequality and Lemma 1(a), we have $\mathbb{E}\left|Z_{4,n,i}\right| \leq Kk_n\Delta_n^2$. In view of (A.18), by the triangle inequality, these estimates further imply

$$\begin{aligned} &\mathbb{E}\left|\mathbb{E}_i^n\left[\zeta_{i,u}^{n,lm}\zeta_{i,v}^{n,jk}\Delta_{n,i+1}\right]\right| \\ &\quad \leq K\Delta_n^{3/2}\left(k_n\Delta_n^{1/2}+k_n^{1/2+\epsilon}\Delta_n^{\epsilon}+\mathbb{E}[\eta_{i+1}^n(b)]+\mathbb{E}[\eta_{i+1}^n(c^{(X,c_1)})]+\mathbb{E}[\eta_{i+1}^n(c^{(X,\sigma_1)})]\right), \end{aligned}$$

that is, (A.17) holds. The proof of Lemma 2 will be finished by showing (A.20) and (A.21), below.

Step 2. We show (A.20) in this step. We start by noting that $\zeta_{i,v}^{n,jk}-(c_{t(n,i+u)}^{jk}-c_{t(n,i)}^{jk}) = \alpha_{n,i+v}^{jk}+(c_{t(n,i+v-1)}^{jk}-c_{t(n,i+u)}^{jk})$. Hence, by the triangle inequality

$$\begin{aligned} &\left|\mathbb{E}_{i+u}^n\left[\zeta_{i,v}^{n,jk}-(c_{t(n,i+u)}^{jk}-c_{t(n,i)}^{jk})\right]\right| \\ &\quad \leq \left|\mathbb{E}_{i+u}^n\left[\alpha_{n,i+v}^{jk}\right]\right|+\left|\mathbb{E}_{i+u}^n\left[c_{t(n,i+v-1)}^{jk}-c_{t(n,i+u)}^{jk}\right]\right|. \end{aligned} \quad \textbf{(A.22)}$$

In addition, by repeated conditioning and Lemma 1(b),

$$\begin{aligned} \left|\mathbb{E}_{i+u}^n\left[\alpha_{n,i+v}^{jk}\right]\right| &\leq \mathbb{E}_{i+u}^n\left[\left|\mathbb{E}_{i+v-1}^{\prime n}\left[\alpha_{n,i+v}^{jk}\right]\right|\right] \\ &\leq K\mathbb{E}_{i+u}^n\left[\Delta_{n,i+v}^{1/2}\eta_{i+v,1}^{\prime n}(b)\right]+K\Delta_n \\ &\qquad +K\mathbb{E}_{i+u}^n\left[\Delta_{n,i+v}^{-1}\int_{t(n,i+v-1)}^{t(n,i+v)}||\sigma_{2,s}-\sigma_{2,t(n,i+v-1)}||ds\right]. \end{aligned} \quad \textbf{(A.23)}$$

Using an argument that is similar to the proof of (A.9), we further deduce that

$$\left|\mathbb{E}_{i+u}^{n}\left[c_{t(n,i+v-1)}^{jk}-c_{t(n,i+u)}^{jk}\right]\right| \\ \leq K\left(k_n\Delta_n+\mathbb{E}_{i+u}^{n}\left[\|\sigma_{2,t(n,i+v-1)}-\sigma_{2,t(n,i+u)}\|\right]\right). \tag{A.24}$$

In view of (A.22), (A.23), and (A.24), by repeated conditioning (note that $\zeta_{i,u}^{n,lm}$ is $\mathcal{H}_{t(n,i+u)}$-measurable) and the triangle inequality, we have

$$\left|Z_{1,n,i}\right| \leq K\left(W_{1,n,i}+W_{2,n,i}+W_{3,n,i}+W_{4,n,i}\right), \tag{A.25}$$

where

$$\begin{cases} W_{1,n,i}\equiv\mathbb{E}_{i}^{n}\left[\left|\zeta_{i,u}^{n,lm}\right|\Delta_{n,i+1}\mathbb{E}_{i+u}^{n}\left[\Delta_{n,i+v}^{1/2}\eta_{i+v,1}^{\prime n}(b)\right]\right], \\ W_{2,n,i}\equiv k_n\Delta_n\mathbb{E}_{i}^{n}\left[\left|\zeta_{i,u}^{n,lm}\right|\Delta_{n,i+1}\right], \\ W_{3,n,i}\equiv\mathbb{E}_{i}^{n}\left[\left|\zeta_{i,u}^{n,lm}\right|\Delta_{n,i+1}\Delta_{n,i+v}^{-1}\int_{t(n,i+v-1)}^{t(n,i+v)}\|\sigma_{2,s}-\sigma_{2,t(n,i+v-1)}\|ds\right], \\ W_{4,n,i}\equiv\mathbb{E}_{i}^{n}\left[\left|\zeta_{i,u}^{n,lm}\right|\Delta_{n,i+1}\|\sigma_{2,t(n,i+v-1)}-\sigma_{2,t(n,i+u)}\|\right]. \end{cases} \tag{A.26}$$

We now proceed to derive bounds for the terms in (A.26), starting with $W_{1,n,i}$ and $W_{2,n,i}$. By Lemma 1(a), the Cauchy–Schwarz inequality and Assumption IR,

$$\mathbb{E}_{i}^{n}\left[\left\|\zeta_{i,u}^{n}\right\|^{2}\Delta_{n,i+1}^{2}\right] \leq K\Delta_n^2. \tag{A.27}$$

In addition,

$$\sqrt{\mathbb{E}_{i}^{n}\left[\left(\mathbb{E}_{i+u}^{n}\left[\Delta_{n,i+v}^{1/2}\eta_{i+v,1}^{\prime n}(b)\right]\right)^{2}\right]} \\ \leq \Delta_n^{1/2}\sqrt{\mathbb{E}_{i}^{n}\left[\mathbb{E}_{i+u}^{n}\left[\left(\eta_{i+v,1}^{\prime n}(b)\right)^{2}\right]\right]} \leq \Delta_n^{1/2}\eta_{i+1}^{n}(b), \tag{A.28}$$

where the first inequality is by the Cauchy–Schwarz inequality and Assumption IR; the second inequality is by the tower property and the definition of $\eta_i^n(b)$. Then, by (A.27), (A.28), and the Cauchy–Schwarz inequality,

$$\mathbb{E}\left|W_{1,n,i}\right| \leq K\Delta_n^{3/2}\mathbb{E}\left[\eta_{i+1}^{n}(b)\right], \quad \mathbb{E}\left|W_{2,n,i}\right| \leq Kk_n\Delta_n^2. \tag{A.29}$$

Moreover,

$$\begin{aligned} &\mathbb{E}\left|W_{3,n,i}\right| \\ &\leq \left(\mathbb{E}\left[\left\|\zeta_{i,u}^{n}\right\|^{2}\Delta_{n,i+1}^{2}\right]\right)^{1/2}\left(\mathbb{E}\left[\left(\Delta_{n,i+v}^{-1}\int_{t(n,i+v-1)}^{t(n,i+v)}\|\sigma_{2,s}-\sigma_{2,t(n,i+v-1)}\|ds\right)^{2}\right]\right)^{1/2} \\ &\leq K\Delta_n\left(\mathbb{E}\left[\Delta_{n,i+v}^{-1}\int_{t(n,i+v-1)}^{t(n,i+v)}\mathbb{E}_{\mathcal{T}}\left[\|\sigma_{2,s}-\sigma_{2,t(n,i+v-1)}\|^{2}\right]ds\right]\right)^{1/2} \\ &\leq K\Delta_n\left(\mathbb{E}\left[\Delta_{n,i+v}^{1+2\epsilon}\right]\right)^{1/2} \leq K\Delta_n^{3/2+\epsilon}, \end{aligned} \tag{A.30}$$

where the first inequality is due to the Cauchy–Schwarz inequality; the second inequality is by (A.27) and repeated conditioning; the third inequality is obtained by Assumption LM′; the last inequality is due to Assumption IR. Similarly,

$$\mathbb{E}\left|W_{4,n,i}\right| \leq K\Delta_n (k_n\Delta_n)^{1/2+\epsilon}. \tag{A.31}$$

The claim (A.20) then readily follows from (A.25), (A.29), (A.30), and (A.31).

Step 3. In this step, we show (A.21). By repeated conditioning, we deduce

$$\mathbb{E}\left|Z_{2,n,i}\right| \leq \mathbb{E}\left|\mathbb{E}^n_{i+u-1}\left[\alpha^{lm}_{n,i+u}\left(c^{jk}_{t(n,i+u)} - c^{jk}_{t(n,i+u-1)}\right)\Delta_{n,i+1}\right]\right|. \tag{A.32}$$

In view of (A.2), we can decompose

$$\mathbb{E}^n_{i+u-1}\left[\alpha^{lm}_{n,i+u}\left(c_{t(n,i+u)} - c_{t(n,i+u-1)}\right)\Delta_{n,i+1}\right] = \sum_{q=1}^{6} D_{q,n,i}, \tag{A.33}$$

where

$$\begin{cases} D_{1,n,i} \equiv \mathbb{E}^n_{i+u-1}\left[\alpha^{lm}_{n,i+u}\left(c_{1,t(n,i+u)} - c_{1,t(n,i+u-1)}\right)\Delta_{n,i+1}\right], \\ D_{2,n,i} \equiv \mathbb{E}^n_{i+u-1}\left[\alpha^{lm}_{n,i+u}\left(c_{2,t(n,i+u)} - c_{2,t(n,i+u-1)}\right)\Delta_{n,i+1}\right], \\ D_{3,n,i} \equiv \mathbb{E}^n_{i+u-1}\left[\alpha^{lm}_{n,i+u}\left(\sigma_{2,t(n,i+u)} - \sigma_{2,t(n,i+u-1)}\right)\sigma^\top_{1,t(n,i+u)}\Delta_{n,i+1}\right], D_{4,n,i} \equiv D^\top_{3,n,i}, \\ D_{5,n,i} \equiv \mathbb{E}^n_{i+u-1}\left[\alpha^{lm}_{n,i+u}\left(\sigma_{1,t(n,i+u)} - \sigma_{1,t(n,i+u-1)}\right)\Delta_{n,i+1}\right]\sigma^\top_{2,t(n,i+u-1)}, D_{6,n,i} \equiv D^\top_{5,n,i}. \end{cases}$$

By the Cauchy–Schwarz inequality and Lemma 1(a), we have

$$\mathbb{E}\left[\left\|D_{2,n,i}\right\| + \left\|D_{3,n,i}\right\| + \left\|D_{4,n,i}\right\|\right] \leq K\Delta_n^{3/2+\epsilon}.$$

As for $D_{1,n,i}$, we have

$$\mathbb{E}\left[\left\|D_{1,n,i}\right\|\right] \leq K\Delta_n^{3/2}\left(\Delta_n^{1/2} + \mathbb{E}\left[\eta^n_{i+1}(c^{(X,c_1)})\right]\right),$$

which can be shown using the argument of Lem. 3.2(c) in Jacod and Rosenbaum (2015), for which we only need (i) the process $\sigma = \sigma_1 + \sigma_2$ to be (1/2)-Hölder continuous under the L_2 norm (see Lemma 1(a)), and (ii) the process c_1 is an Itô semimartingale. By using the same argument but with c_1 replaced by σ_1, we deduce

$$\mathbb{E}\left\|D_{5,n,i}\right\| + \mathbb{E}\left\|D_{6,n,i}\right\| \leq K\Delta_n^{3/2}\left(\Delta_n^{1/2} + \mathbb{E}\left[\eta^n_{i+1}(c^{(X,\sigma_1)})\right]\right).$$

Combining these estimates with (A.33), we see that

$$\begin{aligned} &\mathbb{E}\left|\mathbb{E}^n_{i+u-1}\left[\alpha^{lm}_{n,i+u}\left(c^{jk}_{t(n,i+u)} - c^{jk}_{t(n,i+u-1)}\right)\Delta_{n,i+1}\right]\right| \\ &\quad \leq K\Delta_n^{3/2}\left(\mathbb{E}[\eta^n_{i+1}(c^{(X,c_1)})] + \mathbb{E}[\eta^n_{i+1}(c^{(X,\sigma_1)})] + \Delta_n^{\epsilon} + \Delta_n^{1/2}\right). \end{aligned} \tag{A.34}$$

The claim (A.21) then follows from (A.32) and (A.34). □

Proof of Theorem 3. *Step 1*. In this step, we show that

$$\sup_{0\leq i\leq N_{n,T}} \left\|\hat{c}_{t(n,i)} - \bar{c}_{t(n,i)}\right\| = o_p(1), \tag{A.35}$$

where

$$\bar{c}_{t(n,i)} \equiv \frac{1}{k_n}\sum_{j=1}^{k_n}\frac{1}{\Delta_{n,i+j}}\int_{t(n,i+j-1)}^{t(n,i+j)} c_s ds$$

forms a type of average of the process c over $[t(n,i), t(n,i+k_n)]$. Since $N_{n,T} = O_p(\Delta_n^{-1})$, it suffices to show $\sup_{0\le i\le M/\Delta_n}\left\|\hat{c}_{t(n,i)} - \bar{c}_{t(n,i)}\right\| = o_p(1)$ for any fixed positive constant M. In addition, by using a polarization argument, we can and shall assume that X is one-dimensional without loss of generality for proving (A.35).

By Itô's formula,

$$\hat{c}'_{t(n,i)} - \bar{c}_{t(n,i)} = \frac{1}{k_n}\sum_{j=1}^{k_n}\frac{2}{\Delta_{n,i+j}}\int_{t(n,i+j-1)}^{t(n,i+j)}\left(X'_s - X'_{t(n,i+j-1)}\right)(b_s ds + \sigma_s dW_s). \tag{A.36}$$

Note that for $p \geq 1$,

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{\Delta_{n,i+j}}\int_{t(n,i+j-1)}^{t(n,i+j)}\left(X'_s - X'_{t(n,i+j-1)}\right)b_s ds\right|^p\right] \\
&\quad\le \mathbb{E}\left[\frac{1}{\Delta_{n,i+j}}\int_{t(n,i+j-1)}^{t(n,i+j)}\left|X'_s - X'_{t(n,i+j-1)}\right|^p |b_s|^p\, ds\right] \\
&\quad\le K_p\mathbb{E}\left[\Delta_{n,i+j}^{p/2}\right] \le K_p\Delta_n^{p/2},
\end{aligned} \tag{A.37}$$

where the first inequality follows from Jensen's inequality; the second inequality is obtained by using an argument similar to (A.8); the last inequality is by Assumption IR. Similarly,

$$\mathbb{E}_{i+j-1}^n\left[\left|\frac{1}{\Delta_{n,i+j}}\int_{t(n,i+j-1)}^{t(n,i+j)}\left(X'_s - X'_{t(n,i+j-1)}\right)\sigma_s dW_s\right|^p\right] \le K_p. \tag{A.38}$$

We further note that

$$\begin{aligned}
&\mathbb{E}_{i+j-1}^n\left[\frac{1}{\Delta_{n,i+j}}\int_{t(n,i+j-1)}^{t(n,i+j)}\left(X'_s - X'_{t(n,i+j-1)}\right)\sigma_s dW_s\right] \\
&\quad= \mathbb{E}_{i+j-1}^n\left[\mathbb{E}_{i+j-1}^{\prime n}\left[\frac{1}{\Delta_{n,i+j}}\int_{t(n,i+j-1)}^{t(n,i+j)}\left(X'_s - X'_{t(n,i+j-1)}\right)\sigma_s dW_s\right]\right] \\
&\quad= 0,
\end{aligned}$$

where the first equality is by repeated conditioning; the second equality holds because of the $\mathcal{H}_{t(n,i-1+j)}$-conditional independence between $\Delta_{n,i+j}$ and $\mathcal{F}_T$. This shows that $\Delta_{n,i+j}^{-1}\int_{t(n,i+j-1)}^{t(n,i+j)}(X'_s - X'_{t(n,i+j-1)})\sigma_s dW_s$ forms a martingale difference sequence with respect to $\mathcal{H}_{t(n,i+j)}$. Then, by (A.38) and the Burkholder–Davis–Gundy inequality, we deduce that

$$\mathbb{E}\left[\left|\frac{1}{k_n}\sum_{j=1}^{k_n}\frac{1}{\Delta_{n,i+j}}\int_{t(n,i+j-1)}^{t(n,i+j)}\left(X'_s - X'_{t(n,i+j-1)}\right)\sigma_s dW_s\right|^p\right] \le K_p k_n^{-p/2}. \tag{A.39}$$

Combining (A.37) and (A.39), we have, for all $p \geq 1$, $\|\hat{c}'_{t(n,i)} - \bar{c}_{t(n,i)}\|_p \leq K_p k_n^{-1/2}$, where $\|\cdot\|_p$ denotes the L_p norm. Then, by using a maximal inequality (see, e.g., Lem. 2.2.2 in van der Vaart and Wellner, 1996), we deduce

$$\left\| \max_{0 \leq i \leq M/\Delta_n} \left| \hat{c}'_{t(n,i)} - \bar{c}_{t(n,i)} \right| \right\|_p \leq K_p \Delta_n^{-1/p} k_n^{-1/2}.$$

By taking p sufficiently large, we deduce

$$\sup_{0 \leq i \leq M/\Delta_n} \left| \hat{c}'_{t(n,i)} - \bar{c}_{t(n,i)} \right| = o_p(1). \tag{A.40}$$

Next, we note that, for some deterministic sequence $\phi_n \to 0$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0 \leq i \leq M/\Delta_n} \left| \hat{c}_{t(n,i)} - \hat{c}'_{t(n,i)} \right| \right] \\
&\quad \leq \frac{1}{k_n} \sum_{i=0}^{\lceil M/\Delta_n \rceil} \mathbb{E}\left[\frac{1}{\Delta_{n,i}} \left| \left(\Delta_i^n X\right)^2 1_{\{|\Delta_i^n X| \leq u_{n,i}\}} - \left(\Delta_i^n X'\right)^2 \right| \right] \\
&\quad \leq \frac{1}{k_n} \sum_{i=0}^{\lceil M/\Delta_n \rceil} \mathbb{E}\left[\frac{1}{\Delta_{n,i}} \mathbb{E}_{i-1}'^n \left[\left| \left(\Delta_i^n X\right)^2 1_{\{|\Delta_i^n X| \leq u_{n,i}\}} - \left(\Delta_i^n X'\right)^2 \right| \right] \right] \\
&\quad \leq \frac{1}{k_n} \sum_{i=0}^{\lceil M/\Delta_n \rceil} \mathbb{E}\left[\Delta_{n,i}^{(2-r)\varpi} \phi_n \right] \\
&\quad = o\left(\Delta_n^{(2-r)\varpi - (1-\gamma)} \right) \to 0,
\end{aligned} \tag{A.41}$$

where the first inequality follows from the definitions of $\hat{c}_{t(n,i)}$ and $\hat{c}'_{t(n,i)}$; the second inequality is by repeated conditioning; the third inequality is obtained by first recognizing that $\Delta_{n,i}$ is independent of $\mathcal{F}_T$ conditional on $\mathcal{H}_{t(n,i-1)}$ and then using Lem. 13.2.6 of Jacod and Protter, 2012, where ϕ_n is implicitly defined; the last line follows from Assumption IR and the maintained condition on the tuning parameters. Combining (A.40) and (A.41), we deduce (A.35) as claimed.

Step 2. Given (A.35) and Assumption C, we can appeal to the spatial localization argument of Li et al., 2017 (see their Thm. 2), so as to assume that $g(\cdot)$ is compactly supported without loss of generality.

We consider the decomposition $\Delta_n^{-1/2}(S'_n(g) - S(g)) = \sum_{j=1}^{5} V_{j,n}$, where

$$V_{1,n} \equiv \Delta_n^{-1/2} \sum_i \left(g(\hat{c}_{t(n,i)}) - g(\hat{c}'_{t(n,i)}) - \frac{1}{k_n} \left(\mathbb{B}g(\hat{c}_{t(n,i)}) - \mathbb{B}g(\hat{c}'_{t(n,i)}) \right) \right) \Delta_{n,i+1},$$

$$V_{2,n} \equiv \Delta_n^{-1/2} \sum_i \int_{t(n,i)}^{t(n,i+1)} \left(g(c_{t(n,i)}) - g(c_s) \right) ds - \Delta_n^{-1/2} \int_{t(n,N_{n,T}-k_n+1)}^{T} g(c_s) ds,$$

$$V_{3,n} \equiv \Delta_n^{-1/2} \sum_i \sum_{l,m=1}^{d} \partial_{lm} g(c_{t(n,i)}) \frac{1}{k_n} \sum_{u=1}^{k_n} \left(c_{t(n,i+u-1)}^{lm} - c_{t(n,i)}^{lm} \right) \Delta_{n,i+1},$$

$$V_{4,n} \equiv \Delta_n^{-1/2} \sum_i \left(g(\hat{c}'_{t(n,i)}) - g(c_{t(n,i)}) - \sum_{l,m=1}^{d} \partial_{lm} g(c_{t(n,i)}) \tilde{c}_{n,i}^{lm} - \frac{1}{k_n} \mathbb{B}g(\hat{c}'_{t(n,i)}) \right) \Delta_{n,i+1},$$

$$V_{5,n} \equiv \Delta_n^{-1/2} k_n^{-1} \sum_i \left(\sum_{l,m=1}^{d} \partial_{lm} g(c_{t(n,i)}) \sum_{u=1}^{k_n} \alpha_{n,i+u}^{lm} \right) \Delta_{n,i+1}.$$

We prove the assertion of Theorem 3(a) by showing the following:

$$V_{j,n} = o_p(1) \text{ for } j = 1,2,3,4, \tag{A.42}$$

$$V_{5,n} \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V'(g)). \tag{A.43}$$

In this step, we show (A.42), leaving (A.43) to step 3, below. We note that $V_{1,n} = o_p(1)$ can be established by a straightforward adaption of "case $j = 1$" in (A.17) of Li and Xiu, 2016, upon using a polarization argument and the argument underlying (A.41); we omit the details for brevity.

We now turn to the term $V_{2,n}$. Since $k_n \Delta_n^{1/2} \to 0$, it is easy to see that

$$\Delta_n^{-1/2} \int_{t(n,N_{n,T}-k_n+1)}^{T} g(c_s) ds = o_p(1).$$

By a second-order Taylor expansion and (A.2), we decompose

$$\Delta_n^{-1/2} \sum_i \int_{t(n,i)}^{t(n,i+1)} \big(g(c_s) - g(c_{t(n,i)})\big) ds = \sum_{k=1}^{5} V_{2,n,k},$$

where, recalling that $\partial g(c)$ is the $d \times d$ matrix with its (l,m) element given by $\partial_{lm} g(c)$, we set

$$\begin{cases} V_{2,n,1} \equiv \Delta_n^{-1/2} \sum_i \int_{t(n,i)}^{t(n,i+1)} \mathrm{Tr}\big[\partial g(c_{t(n,i)})(c_{1,s} - c_{1,t(n,i)})\big] ds, \\ V_{2,n,2} \equiv \Delta_n^{-1/2} \sum_i \int_{t(n,i)}^{t(n,i+1)} \mathrm{Tr}\big[\partial g(c_{t(n,i)})(c_{2,s} - c_{2,t(n,i)})\big] ds, \\ V_{2,n,3} \equiv 2\Delta_n^{-1/2} \sum_i \int_{t(n,i)}^{t(n,i+1)} \mathrm{Tr}\Big[\partial g(c_{t(n,i)})(\sigma_{2,s} - \sigma_{2,t(n,i)})\sigma_{1,s}^{\top}\Big] ds, \\ V_{2,n,4} \equiv 2\Delta_n^{-1/2} \sum_i \int_{t(n,i)}^{t(n,i+1)} \mathrm{Tr}\Big[\partial g(c_{t(n,i)}) \big(\sigma_{1,s} - \sigma_{1,t(n,i)}\big) \sigma_{2,t(n,i)}^{\top}\Big] ds, \\ V_{2,n,5} \equiv \Delta_n^{-1/2} \frac{1}{2} \sum_i \int_{t(n,i)}^{t(n,i+1)} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 g(\xi_{n,i}(s))(c_s^{jk} - c_{t(n,i)}^{jk})(c_s^{lm} - c_{t(n,i)}^{lm}) ds, \end{cases}$$

for some mean values $\xi_{n,i}(s)$ between c_s and $c_{t(n,i)}$. It remains to show that $V_{2,n,k} = o_p(1)$ for $1 \leq k \leq 5$. By a standard Riemann approximation result for Itô semimartingales (see, e.g., Jacod and Protter, 2012, pp.153–154), here applied to the processes σ_1 and c_1, we can show that $V_{2,n,k} = o_p(1)$ for $k = 1$ and 4, for which we note that the $\mathcal{H}_{t(n,i)}$-measurability of $\partial g(c_{t(n,i)})$ and $\sigma_{2,t(n,i)}^{\top}$ is implicitly used. Furthermore, because of the boundedness of $\partial g(c_{t(n,i)})$ and Assumption LM$'$, we have $V_{2,n,2} = O_p(\Delta_n^{\epsilon}) = o_p(1)$; similarly, we deduce that $V_{2,n,3}$ is also $o_p(1)$. Finally, using similar techniques as above, we deduce that $V_{2,n,5} = O_p(\Delta_n^{1/2}) = o_p(1)$. These estimates readily imply (A.42) for the case $j = 2$.

Next, we show $V_{3,n} = o_p(1)$. In view of (A.2), we can decompose $V_{3,n} = \sum_{k=1}^4 V_{3,n,k}$, where

$$
\begin{cases}
V_{3,n,1} \equiv \Delta_n^{-1/2} \sum_i k_n^{-1} \mathrm{Tr} \left[\partial g(c_{t(n,i)}) \sum_{u=1}^{k_n} \left(c_{1,t(n,i+u-1)} - c_{1,t(n,i)} \right) \right] \Delta_{n,i+1}, \\
V_{3,n,2} \equiv \Delta_n^{-1/2} \sum_i k_n^{-1} \mathrm{Tr} \left[\partial g(c_{t(n,i)}) \sum_{u=1}^{k_n} \left(c_{2,t(n,i+u-1)} - c_{2,t(n,i)} \right) \right] \Delta_{n,i+1}, \\
V_{3,n,3} \equiv 2\Delta_n^{-1/2} \sum_i k_n^{-1} \mathrm{Tr} \left[\partial g(c_{t(n,i)}) \sum_{u=1}^{k_n} \left(\sigma_{2,t(n,i+u-1)} - \sigma_{2,t(n,i)} \right) \sigma_{1,t(n,i+u-1)}^\top \right] \Delta_{n,i+1}, \\
V_{3,n,4} \equiv 2\Delta_n^{-1/2} \sum_i k_n^{-1} \mathrm{Tr} \left[\partial g(c_{t(n,i)}) \sum_{u=1}^{k_n} \left(\sigma_{1,t(n,i+u-1)} - \sigma_{1,t(n,i)} \right) \sigma_{2,t(n,i)}^\top \right] \Delta_{n,i+1}.
\end{cases}
$$

We first analyze the term $V_{3,n,1}$. To simplify notations, we denote the ith summand of $V_{3,n,1}$ by $\chi_{n,i}$, so that we can write $V_{3,n,1} = \Delta_n^{-1/2} \sum_i \chi_{n,i}$. Since c_1 is an Itô semimartingale, we can follow similar steps underlying (A.7) to deduce that $|\mathbb{E}_i^n[\chi_{n,i}]| \le K k_n \Delta_n^2$ and use Lemma 1(a) to deduce $\mathbb{E}[\chi_{n,i}^2] \le K k_n \Delta_n^3$. Further note that $\chi_{n,i}$ is $\mathcal{H}_{t(n,i+k_n-1)}$-measurable. From these facts, we deduce

$$
\Delta_n^{-1/2} \sum_i \mathbb{E}_i^n \left[\chi_{n,i} \right] = O_p(k_n \Delta_n^{1/2}), \quad \Delta_n^{-1/2} \sum_i \left(\chi_{n,i} - \mathbb{E}_i^n \left[\chi_{n,i} \right] \right) = O_p(k_n \Delta_n^{1/2}).
$$

Hence, $V_{3,n,1} = o_p(1)$. Similarly, we can show that $V_{3,n,4} = o_p(1)$. By Lemma 1(a), $\mathbb{E}\left|V_{3,n,2}\right| \le K k_n^{1/2+\epsilon} \Delta_n^{\epsilon} \to 0$. Hence, $V_{3,n,2} = o_p(1)$. Similarly, we can show that $V_{3,n,3} = o_p(1)$. From these estimates, $V_{3,n} = o_p(1)$ readily follows.

We now show $V_{4,n} = o_p(1)$. To simplify notations, we denote by $\zeta_{n,i}$ the ith summand in $V_{4,n}$, so that we can rewrite $V_{4,n} = \Delta_n^{-1/2} \sum_i \zeta_{n,i}$. We further decompose $\zeta_{n,i} = \zeta'_{n,i} + \zeta''_{n,i}$, where

$$
\zeta'_{n,i} \equiv \frac{1}{2} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g\left(c_{t(n,i)}\right) \left(\tilde{c}_{n,i}^{jk} \tilde{c}_{n,i}^{lm} - \frac{1}{k_n} \left(c_{t(n,i)}^{jl} c_{t(n,i)}^{km} + c_{t(n,i)}^{jm} c_{t(n,i)}^{kl} \right) \right) \Delta_{n,i+1},
$$

and $\zeta''_{n,i}$ is defined implicitly through this decomposition. Since g is compactly supported (by spatial localization mentioned above), by using a mean-value expansion, it is easy to see that $|\zeta''_{n,i}| \le K(\left\|\tilde{c}_{n,i}\right\|^3 + k_n^{-1} \|\tilde{c}_{n,i}\|)\Delta_{n,i+1}$. By Lemma 1(d), we further have $\mathbb{E}|\zeta''_{n,i}| \le K(k_n^{-3/2} + k_n \Delta_n)\Delta_n$ and, hence,

$$
\Delta_n^{-1/2} \sum_i \zeta''_{n,i} = o_p(1). \tag{A.44}
$$

In addition, by Lemma 2, for any fixed constant $M > 0$,

$$
\mathbb{E}\left[\Delta_n^{-1/2} \sum_{i=0}^{\lceil M/\Delta_n \rceil} \left| \mathbb{E}_i^n \left[\zeta'_{n,i} \right] \right| \right] \le K \left(k_n^{-1/2} + k_n \Delta_n^{1/2} + k_n^{1/2+\epsilon} \Delta_n^{\epsilon} \right) + \Delta_n \sum_{i=0}^{\lceil M/\Delta_n \rceil} \mathbb{E}\left[\eta_{i+1}^n \right].
$$

Since the processes b, $c^{(X,\sigma_1)}$ and $c^{(X,c_1)}$ are càdlàg, by Lem. 4.2 of Jacod and Rosenbaum, 2013,

$$\Delta_n \sum_{i=0}^{\lceil M/\Delta_n \rceil} \mathbb{E}\left[\eta_{i+1}^n\right] \to 0. \tag{A.45}$$

From here, it follows that (recalling $N_{n,T} = O_p\left(\Delta_n^{-1}\right)$)

$$\Delta_n^{-1/2} \sum_i \left|\mathbb{E}_i^n\left[\zeta'_{n,i}\right]\right| = o_p(1). \tag{A.46}$$

Finally, we note that the sequence $\zeta'_{n,i} - \mathbb{E}_i^n[\zeta'_{n,i}]$ can only have nonzero autocovariance up to lag k_n. Hence, by the Cauchy–Schwarz inequality and Lemma 1 (d), for any fixed constant $M > 0$,

$$\begin{aligned}\mathbb{E}&\left[\left(\Delta_n^{-1/2} \sum_{i=0}^{\lceil M/\Delta_n \rceil} \left(\zeta'_{n,i} - \mathbb{E}_i^n\left[\zeta'_{n,i}\right]\right)\right)^2\right] \\ &\leq K k_n \Delta_n^{-1} \sum_{i=0}^{\lceil M/\Delta_n \rceil} \mathbb{E}\left[\left(\zeta'_{n,i}\right)^2\right] \leq K\left(k_n^{-1} + k_n^2 \Delta_n\right) \to 0,\end{aligned} \tag{A.47}$$

which, together with (A.46), implies $\Delta_n^{-1/2}\sum_i \zeta'_{n,i} = o_p(1)$. In view of (A.44), we further deduce $V_{4,n} = o_p(1)$ as claimed.

Step 3. Finally, we show that $V_{5,n} \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V'(g))$. We can rewrite

$$V_{5,n} = \Delta_n^{1/2} \sum_i \sum_{l,m=1}^d w_{n,i}^{lm} \alpha_{n,i}^{lm},$$

where (by convention, we set the summand to zero if $i - j \notin \{0, \ldots, N_{n,T} - k_n\}$)

$$w_{n,i}^{lm} \equiv \frac{1}{k_n} \sum_{j=1}^{k_n} \partial_{lm} g(c_{t(n,i-j)}) \frac{\Delta_{n,i+1-j}}{\Delta_n}.$$

Note that $w_{n,i}^{lm}\alpha_{n,i}^{lm}$ is $\mathcal{H}'_{t(n,i)}$-measurable and $w_{n,i}^{lm}$ is $\mathcal{H}'_{t(n,i-1)}$ measurable.

We also set

$$\tilde{w}_{n,i}^{lm} \equiv \partial_{lm} g(c_{t(n,i-1)}) \frac{1}{k_n} \sum_{j=1}^{k_n} \frac{\Delta_{n,i+1-j}}{\Delta_n}, \quad \bar{w}_{n,i}^{lm} \equiv \partial_{lm} g(c_{t(n,i-1)}) \frac{1}{k_n} \sum_{j=1}^{k_n} \frac{1}{\Lambda_{t(n,i-j)}}.$$

From Lemma 1(a) and the Cauchy–Schwarz inequality, it is easy to see

$$\begin{aligned}\left\|w_{n,i}^{lm} - \tilde{w}_{n,i}^{lm}\right\|_2 &\leq k_n^{-1} \sum_{j=1}^{k_n} \left\|\left(\partial_{lm} g(c_{t(n,i-j)}) - \partial_{lm} g(c_{t(n,i-1)})\right) \frac{\Delta_{n,i+1-j}}{\Delta_n}\right\|_2 \\ &\leq K\left(k_n \Delta_n\right)^{1/2}.\end{aligned} \tag{A.48}$$

Moreover,

$$\mathbb{E}\left[\left|\tilde{w}_{n,i}^{lm}-\bar{w}_{n,i}^{lm}\right|^2\right]=\mathbb{E}\left[\left|\partial_{lm}g(c_{t(n,i-1)})\right|^2\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\left(\frac{\Delta_{n,i+1-j}}{\Delta_n}-\frac{1}{\Lambda_{t(n,i-j)}}\right)\right)^2\right]$$
$$\leq K\mathbb{E}\left[\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\left(\frac{\Delta_{n,i+1-j}}{\Delta_n}-\frac{1}{\Lambda_{t(n,i-j)}}\right)\right)^2\right]$$
$$\leq K\mathbb{E}\left[\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\left(\mathbb{E}_{i-j}^n\left[\frac{\Delta_{n,i+1-j}}{\Delta_n}\right]-\frac{1}{\Lambda_{t(n,i-j)}}\right)\right)^2\right]$$
$$+K\mathbb{E}\left[\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\left(\frac{\Delta_{n,i+1-j}}{\Delta_n}-\mathbb{E}_{i-j}^n\left[\frac{\Delta_{n,i+1-j}}{\Delta_n}\right]\right)\right)^2\right]$$
$$\leq K\left(\kappa_n^2+k_n^{-1}\right), \tag{A.49}$$

where the first inequality is by the boundedness of $\partial_{lm}g(c_{t(n,i-1)})$; the second inequality is by the C_r-inequality; the last inequality follows from Assumption IR and the fact that $\Delta_{n,i+1-j}/\Delta_n-\mathbb{E}_{i-j}^n\left[\Delta_{n,i+1-j}/\Delta_n\right]$ forms a martingale difference sequence.

From (A.48) and (A.49), we deduce

$$\mathbb{E}\left[\left|w_{n,i}^{lm}-\bar{w}_{n,i}^{lm}\right|^2\right]\leq K\left(k_n\Delta_n+k_n^{-1}+\kappa_n^2\right). \tag{A.50}$$

We now proceed to prove the convergence (A.43). We first observe that by the $\mathcal{H}'_{t(n,i-1)}$-measurability of $w_{n,i}^{lm}$ and Lemma 1(b),

$$\begin{aligned}\mathbb{E}\left|\mathbb{E}_{i-1}'^n\left[w_{n,i}^{lm}\alpha_{n,i}^{lm}\right]\right|&=\mathbb{E}\left|w_{n,i}^{lm}\mathbb{E}_{i-1}'^n\left[\alpha_{n,i}^{lm}\right]\right|\\&\leq\mathbb{E}\left|w_{n,i}^{lm}\left(\Delta_{n,i}+\mathbb{E}_{i-1}'^n\left[\Delta_{n,i}^{-1}\int_{t(n,i-1)}^{t(n,i)}||\sigma_{2,s}-\sigma_{2,t(n,i-1)}||ds\right]\right)\right|\\&\quad+\mathbb{E}\left[\mathbb{E}_{i-k_n}^n\left|w_{n,i}^{lm}\Delta_{n,i}^{1/2}\eta_{i,1}'^n(b)\right|\right].\end{aligned} \tag{A.51}$$

In addition, we note that by Assumptions IR and LM′,

$$\begin{aligned}&\mathbb{E}\left[\left(w_{n,i}^{lm}\right)^2\right]\leq K,\quad\mathbb{E}_{i-k_n}^n\left[\left(w_{n,i}^{lm}\right)^2\Delta_{n,i}\right]\leq K\Delta_n,\\&\mathbb{E}\left[\left(\Delta_{n,i}+\mathbb{E}_{i-1}'^n\left[\Delta_{n,i}^{-1}\int_{t(n,i-1)}^{t(n,i)}||\sigma_{2,s}-\sigma_{2,t(n,i-1)}||ds\right]\right)^2\right]\leq K\Delta_n^{1+2\epsilon},\end{aligned} \tag{A.52}$$

and by the definition of $\eta_i^n(b)$,

$$\sqrt{\mathbb{E}_{i-k_n}^n\left[\left|\eta_{i,1}'^n(b)\right|^2\right]}\leq\eta_{i-k_n+1}^n(b). \tag{A.53}$$

By the Cauchy–Schwarz inequality, (A.52) and (A.53), we deduce from (A.51) that $\mathbb{E}\left|\mathbb{E}'^{n}_{i-1}\left[w^{lm}_{n,i}\alpha^{lm}_{n,i}\right]\right| \leq K\Delta_n^{1/2+\epsilon} + K\Delta_n^{1/2}\mathbb{E}\left[\eta^{n}_{i-k_n+1}(b)\right]$.

Hence,

$$\Delta_n^{1/2}\sum_i \sum_{l,m=1}^{d} \mathbb{E}'^{n}_{i-1}\left[w^{lm}_{n,i}\alpha^{lm}_{n,i}\right] = o_p(1).$$

In the sense of the above estimate, $w^{lm}_{n,i}\alpha^{lm}_{n,i}$ forms an array of approximate martingale differences (w.r.t. the filtration array $(\mathcal{H}'_{t(n,i)})$). We can then follow similar steps as in Lem. 4.5 of Jacod and Rosenbaum (2013) (by verifying the conditions of Thm. IX.7.28 in Jacod and Shiryaev, 2003) to prove the asserted convergence (A.43). For brevity, we only emphasize the key difference which concerns the derivation of the asymptotic variance. Note that by Lemma 1(e),

$$\begin{aligned}
&\Delta_n \sum_i \sum_{j,k,l,m=1}^{d} \mathbb{E}'^{n}_{i-1}\left[w^{jk}_{n,i}w^{lm}_{n,i}\alpha^{jk}_{n,i}\alpha^{lm}_{n,i}\right] \\
&\quad= \Delta_n \sum_i \sum_{j,k,l,m=1}^{d} w^{jk}_{n,i}w^{lm}_{n,i}\mathbb{E}'^{n}_{i-1}\left[\alpha^{jk}_{n,i}\alpha^{lm}_{n,i}\right] \\
&\quad= \Delta_n \sum_i \sum_{j,k,l,m=1}^{d} w^{jk}_{n,i}w^{lm}_{n,i}\left(c^{jl}_{t(n,i-1)}c^{km}_{t(n,i-1)} + c^{jm}_{t(n,i-1)}c^{kl}_{t(n,i-1)}\right) + o_p(1).
\end{aligned} \tag{A.54}$$

In addition, by the boundedness of c and $\bar{w}^{jk}_{n,i}$, as well as (A.50),

$$\begin{aligned}
&\Delta_n \sum_i \sum_{j,k,l,m=1}^{d} \left|\left(w^{jk}_{n,i}w^{lm}_{n,i} - \bar{w}^{jk}_{n,i}\bar{w}^{lm}_{n,i}\right)\left(c^{jl}_{t(n,i-1)}c^{km}_{t(n,i-1)} + c^{jm}_{t(n,i-1)}c^{kl}_{t(n,i-1)}\right)\right| \\
&\quad\leq K\Delta_n \sum_i \sum_{j,k,l,m=1}^{d} \left|\left(w^{jk}_{n,i}w^{lm}_{n,i} - \bar{w}^{jk}_{n,i}\bar{w}^{lm}_{n,i}\right)\right| \\
&\quad= O_p\left((k_n\Delta_n)^{1/2} + k_n^{-1/2} + \kappa_n\right) = o_p(1).
\end{aligned}$$

Hence, replacing $w^{jk}_{n,i}w^{lm}_{n,i}$ with $\bar{w}^{jk}_{n,i}\bar{w}^{lm}_{n,i}$ in (A.54) only leads to an $o_p(1)$ difference. Since Λ is continuous, we further have

$$\begin{aligned}
&\Delta_n \sum_i \sum_{j,k,l,m=1}^{d} \mathbb{E}'^{n}_{i-1}\left[w^{jk}_{n,i}w^{lm}_{n,i}\alpha^{jk}_{n,i}\alpha^{lm}_{n,i}\right] \\
&= \Delta_n \sum_i \sum_{j,k,l,m=1}^{d} \frac{\partial_{jk}g(c_{t(n,i-1)})\partial_{lm}g(c_{t(n,i-1)})\left(c^{jl}_{t(n,i-1)}c^{km}_{t(n,i-1)} + c^{jm}_{t(n,i-1)}c^{kl}_{t(n,i-1)}\right)}{\Lambda^2_{t(n,i-1)}} \\
&+o_p(1) \\
&\xrightarrow{\mathbb{P}} \int_0^T \sum_{j,k,l,m=1}^{d} \partial_{jk}g(c_s)\partial_{lm}g(c_s)\left(c^{jl}_s c^{km}_s + c^{jm}_s c^{kl}_s\right)\Lambda_s^{-1}\,ds,
\end{aligned}$$

where the convergence is by Lem. A.2 of Jacod et al., 2017. By some straightforward algebra, we see that the asymptotic variance identified in the above display coincides with $V'(g)$ defined in Theorem 3. This finishes the proof. □

A.3 Proofs of Theorem 2 and Theorem 4

Since Theorem 2 is a special case of Theorem 4, we only need to prove the latter. We shall use the following lemma, where part (a) is used to derive the asymptotic variance of the score $\xi_n(h)$ and part (b) is used to solve the functional maximization problem in the calculation of lower efficiency bound.

LEMMA 3. (a) *Let A be a $d \times d$ symmetric matrix and r be a d-dimensional Gaussian random variable with zero mean and covariance matrix Σ. Then the variance of $r^\top Ar$ is $2Tr[\Sigma A \Sigma A]$.*

(b) $\langle h, \tilde{h} \rangle_{\mathbb{DS}} \equiv \int_0^T Tr\left[c_s h_s c_s \tilde{h}_s\right] \Lambda_s ds$ *defines a semi-inner product for $h, \tilde{h} \in \mathbb{DS}_d$.*

Proof. The assertion in part (a) can be proved by direct calculation. Turning to part (b), it is easy to see that $\langle \cdot, \cdot \rangle_{\mathbb{DS}}$ is bilinear. By part (a), we see that Tr $[c_s h_s c_s h_s]$ is half of the variance of $r^\top h_s r$ for $r \sim \mathcal{N}(0, c_s)$ and, hence, is non-negative. Therefore, $\langle h, h \rangle_{\mathbb{DS}} \geq 0$ for every $h \in \mathbb{DS}_d$. This verifies that $\langle \cdot, \cdot \rangle_{\mathbb{DS}}$ is a semi-inner product on the vector space $\mathbb{DS}_d$. □

Proof of Theorem 4. (a) Fix some $h \in \mathbb{DS}_d$. We start by introducing some notation. From (3.6), we can decompose

$$c(\Delta_n^{1/2} a, h)_t - c_t = \Delta_n^{1/2} a \phi(h)_t + \Delta_n a^2 \psi(h)_t, \tag{A.55}$$

where $\phi(h)_t \equiv c_t h_t c_t$ and $\psi(h)_t \equiv c_t h_t c_t h_t c_t / 4$. We denote

$$\bar{b}_{n,i} \equiv \Delta_{n,i}^{-1} \int_{t(n,i-1)}^{t(n,i)} b_s ds,$$

and define $\bar{c}_{n,i}$ and $\overline{c(\Delta_n^{1/2} a, h)}_{n,i}$ in the similar way by replacing b with c and $c(\Delta_n^{1/2} a, h)$, respectively. We also denote

$$r_{n,i} \equiv \Delta_{n,i}^{-1/2} \int_{t(n,i-1)}^{t(n,i)} \sigma_s dW_s.$$

Below, we write $\sum_i$ in place of $\sum_{i=1}^{N_{n,T}}$ for notational simplicity.

With these notations, we can write $\Delta_i^n X = \Delta_{n,i} \bar{b}_{n,i} + \Delta_i^n J + \Delta_{n,i}^{1/2} r_{n,i}$. Recall that $\mathcal{G}$ is the σ-field generated by (b, J, σ) and the sampling times. Under Assumption LAMN$'$, we see that $\left(r_{n,i}\right)_{i \geq 1}$ are $\mathcal{G}$-conditionally independent centered Gaussian with covariance matrix $\overline{c(\Delta_n^{1/2} a, h)}_{n,i}$ under $P_n(\Delta_n^{1/2} a, h)$. We then see that the log conditional likelihood ratio has the following form:

$$L_n(a\Delta_n^{1/2},h) = \frac{1}{2}\sum_i \left(\log\det\left[\overline{c\left(\Delta_n^{1/2}a,h\right)_{n,i}}^{-1}\right] - \log\det\left[\bar{c}_{n,i}^{-1}\right]\right)$$
$$- \frac{1}{2}\sum_i r_{n,i}^{\top}\left(\overline{c\left(\Delta_n^{1/2}a,h\right)_{n,i}}^{-1} - \bar{c}_{n,i}^{-1}\right) r_{n,i}.$$

We further decompose $L_n(a\Delta_n^{1/2},h) = a\xi_n(h) + A_n(a,h) + B_n(a,h)$ where

$$\begin{cases} \xi_n(h) \equiv \dfrac{\Delta_n^{1/2}}{2}\sum_i \mathrm{Tr}\left[\bar{c}_{n,i}^{-1}\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}\left(r_{n,i}r_{n,i}^{\top} - \bar{c}_{n,i}\right)\right], \\ A_n(a,h) \equiv \sum_i \zeta(a,h)_{n,i}, \quad B_n(a,h) \equiv \sum_i \chi(a,h)_{n,i}, \\ \zeta(a,h)_{n,i} \equiv \dfrac{1}{2}\left(\log\det\left[\overline{c\left(\Delta_n^{1/2}a,h\right)_{n,i}}^{-1}\right] - \log\det\left[\bar{c}_{n,i}^{-1}\right]\right. \\ \qquad \left. -\mathrm{Tr}\left[\left(\overline{c\left(\Delta_n^{1/2}a,h\right)_{n,i}}^{-1} - \bar{c}_{n,i}^{-1}\right)\bar{c}_{n,i}\right]\right), \\ \chi(a,h)_{n,i} \equiv -\dfrac{1}{2}\mathrm{Tr}\left[\left(\overline{c\left(\Delta_n^{1/2}a,h\right)_{n,i}}^{-1} - \bar{c}_{n,i}^{-1} + \Delta_n^{1/2}a\bar{c}_{n,i}^{-1}\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}\right)\left(r_{n,i}r_{n,i}^{\top} - \bar{c}_{n,i}\right)\right]. \end{cases}$$

Clearly, the assertion of part (a) can be proved by showing

$$\xi_n(h) \xrightarrow{\mathcal{L}|\mathcal{G}} \mathcal{MN}(0,\Gamma(h)), \tag{A.56}$$

$$A_n(a,h) \xrightarrow{\mathbb{P}} -\frac{1}{2}a^2\Gamma(h), \tag{A.57}$$

$$B_n(a,h) \xrightarrow{\mathbb{P}} 0, \tag{A.58}$$

where $\xrightarrow{\mathcal{L}|\mathcal{G}}$ stands for the convergence in probability of $\mathcal{G}$-conditional distributions under the uniform metric.

We first show (A.56). Note that, conditional on $\mathcal{G}$, $\xi_n(h)$ is a sum of independent triangular array with mean zero and bounded fourth moments. Moreover, we observe

$$\begin{aligned} &\mathbb{E}\left[\xi_n(h)^2\middle|\mathcal{G}\right] \\ &= \frac{\Delta_n}{4}\sum_i Var\left[r_{n,i}^{\top}\bar{c}_{n,i}^{-1}\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}r_{n,i}\middle|\mathcal{G}\right] \\ &= \frac{\Delta_n}{2}\sum_i \mathrm{Tr}\left[\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}\right] \\ &= \frac{\Delta_n}{2}\sum_i \mathrm{Tr}\left[\phi(h)_{t(n,i-1)}c_{t(n,i-1)}^{-1}\phi(h)_{t(n,i-1)}c_{t(n,i-1)}^{-1}\right] + o_p(1) \\ &= \frac{1}{2}\int_0^T \mathrm{Tr}\left[\phi(h)_s c_s^{-1}\phi(h)_s c_s^{-1}\right]\Lambda_s ds + o_p(1), \end{aligned} \tag{A.59}$$

where the first equality follows from the $\mathcal{G}$-conditional independence of the summand; the second equality follows from Lemma 3; the third equality holds because the processes $\phi(h)$ and c are càdlàg; the last line is by Lem. A.2(a) of Jacod et al., 2017. Since $\phi(h)_t = c_t h_t c_t$ by definition, we further deduce $\mathbb{E}[\xi_n(h)^2 | \mathcal{G}] = \Gamma(h) + o_p(1)$. Hence, for any subsequence, we can find a further subsequence, along which $\mathbb{E}[\xi_n(h)^2 | \mathcal{G}] \to \Gamma(h)$ for almost every path. For each such path, we can use the Lindeberg–Lévy central limit theorem to deduce that $\xi_n(h)$ converges in distribution to $\mathcal{MN}(0, \Gamma(h))$ under the $\mathcal{G}$-conditional probability. By a reverse use of the subsequence argument, we get (A.56).

Next, we show (A.57). We note that the first and the second differentials of $\log\det(x)$ are $\mathrm{Tr}\left[x^{-1}dx\right]$ and $(-1/2)\mathrm{Tr}\left[x^{-1}dx x^{-1}dx\right]$, respectively. By a second-order Taylor expansion, we have

$$
\begin{aligned}
\log\det\left[\overline{c\left(\Delta_n^{1/2}a,h\right)}_{n,i}^{-1}\right]
&= \log\det\left[\bar{c}_{n,i}^{-1}\right] + \mathrm{Tr}\left[\bar{c}_{n,i}\left(\overline{c\left(\Delta_n^{1/2}a,h\right)}_{n,i}^{-1} - \bar{c}_{n,i}^{-1}\right)\right] \\
&\quad -\frac{1}{2}\mathrm{Tr}\left[\bar{c}_{n,i}\left(\overline{c\left(\Delta_n^{1/2}a,h\right)}_{n,i}^{-1} - \bar{c}_{n,i}^{-1}\right)\bar{c}_{n,i}\left(\overline{c\left(\Delta_n^{1/2}a,h\right)}_{n,i}^{-1} - \bar{c}_{n,i}^{-1}\right)\right] + \eta_{n,i},
\end{aligned}
$$

where the residual term $\eta_{n,i}$ satisfies $\left|\eta_{n,i}\right| \le K\Delta_n^{3/2}$. Hence,

$$
\zeta(a,h)_{n,i} = -\frac{1}{4}\mathrm{Tr}\left[\bar{c}_{n,i}\left(\overline{c\left(\Delta_n^{1/2}a,h\right)}_{n,i}^{-1} - \bar{c}_{n,i}^{-1}\right)\bar{c}_{n,i}\left(\overline{c\left(\Delta_n^{1/2}a,h\right)}_{n,i}^{-1} - \bar{c}_{n,i}^{-1}\right)\right] + \frac{1}{2}\eta_{n,i}.
$$

Further note that $d\left(x^{-1}\right) = -x^{-1}dx x^{-1}$. Then, by a Taylor expansion for the matrix inverse function, we have the following approximation

$$
\left|\zeta(a,h)_{n,i} + \frac{a^2\Delta_n}{4}\mathrm{Tr}\left[\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}\right]\right| \le K\Delta_n^{3/2}.
$$

Therefore,

$$
A_n(a,h) = -\frac{a^2\Delta_n}{4}\sum_i \mathrm{Tr}\left[\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}\right] + O_p(\Delta_n^{1/2}).
$$

By similar arguments underlying (A.59), we deduce

$$
A_n(a,h) \xrightarrow{\mathbb{P}} -\frac{a^2}{4}\int_0^T \mathrm{Tr}\left[\phi(h)_s c_s^{-1}\phi(h)_s c_s^{-1}\right]\Lambda_s ds = -\frac{1}{2}a^2\Gamma(h).
$$

Finally, we show (A.58). Note that, conditional on $\mathcal{G}$, $\chi(a,h)_{n,i}$ forms an independent array with zero mean. Moreover, by a Taylor expansion,

$$
\left\|\overline{c\left(\Delta_n^{1/2}a,h\right)}_{n,i}^{-1} - \bar{c}_{n,i}^{-1} + a\Delta_n^{1/2}\bar{c}_{n,i}^{-1}\overline{\phi(h)}_{n,i}\bar{c}_{n,i}^{-1}\right\| \le K\Delta_n,
$$

and, hence, $\mathbb{E}[\chi(a,h)_{n,i}^2 | \mathcal{G}] \le K\Delta_n^2$. Therefore,

$$
\mathbb{E}\left[B_n(a,h)^2\middle|\mathcal{G}\right] = \sum_i \mathbb{E}\left[\chi(a,h)_{n,i}^2\middle|\mathcal{G}\right] = O_p(\Delta_n) = o_p(1).
$$

From here, (A.58) readily follows. This finishes the proof of part (a).

(b,c) Recall the definition of the semi-inner product $\langle \cdot, \cdot \rangle_{\mathbb{DS}}$ from Lemma 3. We can rewrite $\Sigma(h)$ as

$$\Sigma(h) = \frac{2 \langle \partial g(c) / \Lambda, h \rangle_{\mathbb{DS}}^2}{\langle h, h \rangle_{\mathbb{DS}}}.$$

By the Cauchy–Schwarz inequality, we deduce that

$$\Sigma(h) \leq 2 \langle \partial g(c) / \Lambda, \partial g(c) / \Lambda \rangle_{\mathbb{DS}} = 2 \int_0^T \mathrm{Tr}\left[c_s \partial g(c_s) c_s \partial g(c_s)\right] \Lambda_s^{-1} ds,$$

where the inequality becomes an equality if $h = \partial g(c) / \Lambda$. □