When Moving-Average Models Meet High-Frequency Data: Uniform Inference on Volatility*

Rui Da† — University of Chicago
Dacheng Xiu‡ — University of Chicago

This version: May 8, 2021

Abstract

We conduct inference on volatility with noisy high-frequency data. We assume the observed transaction price follows a continuous-time Itô-semimartingale, contaminated by a discrete-time moving-average noise process associated with the arrival of trades. We estimate volatility, defined as the quadratic variation of the semimartingale, by maximizing the likelihood of a misspecified moving-average model, with its order selected based on an information criterion. Our inference is uniformly valid over a large class of noise processes whose magnitude and dependence structure vary with sample size. We show that the convergence rate of our estimator dominates $n^{1/4}$ as noise vanishes, and is determined by the selected order of noise dependence when noise is sufficiently small. Our implementation guarantees positive estimates in finite samples.

Keywords: QMLE, Serially Correlated Noise, Small Noise, Model Selection, Uniformity
JEL Codes: C13, C14, C55, C58.

*We are grateful to the editor, Ulrich Müller, and three anonymous referees for valuable comments. We also benefited tremendously from discussions with Yacine Aït-Sahalia, Torben Andersen, Roland Bénabou, Tim Bollerslev, Bruce Hansen, Christian Hansen, Peter Hansen, Jean Jacod, Per Mykland, Andrew Patton, Zhongjun Qu, Jeffrey Russell, Neil Shephard, Ruey Tsay, and Ken West, as well as seminar and conference participants at the University of Wisconsin, Imperial College London, UCL, Indiana University, Aarhus University, Tilburg University, the 9th SoFiE annual conference, the CEME Young Econometricians Workshop at Duke University, the conference on High Frequency Financial Econometrics at Rutgers, and the Market Microstructure and High Frequency Trading Conference at the University of Chicago. Xiu gratefully acknowledges computing support from the Research Computing Center at the University of Chicago and financial support from the Fama-Miller Center and the IBM Faculty Scholar Fund at the University of Chicago Booth School of Business, as well as an Azure Research Grant from Microsoft.

†Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: rui.da@chicagobooth.edu.
‡Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: dacheng.xiu@chicagobooth.edu.

1 Introduction

In this paper, we develop a simple estimator of volatility using high-frequency data in the presence of serially correlated, heteroscedastic, and endogenous microstructure noise. More importantly, we propose uniformly valid inference over a large class of noise processes that allows for simultaneously infinite-order autocorrelation and arbitrarily shrinking magnitude.

Hansen and Lunde (2006) provide empirical evidence that microstructure noise is quite small in Dow Jones Industrial Average stocks. To improve efficiency, one can consider a test of whether noise is present (or rely on an informal volatility signature plot), then decide whether to use a noise-robust estimator or the more efficient realized volatility estimator (Andersen, Bollerslev, Diebold, and Labys (2003)), which assumes noise is absent. Standard (pointwise) inference for this pre-testing approach, however, yields a misleading picture of the actual finite-sample behavior. Moreover, assuming the noise exists, a follow-up issue is to determine its dependence structure. An estimator robust to noise with long-range temporal dependence could be inefficient if the actual noise is simply i.i.d.. To strike a more desirable trade-off between efficiency and robustness, one can consider modeling the noise as a moving-average process and adopting information criteria to determine its order of dependence. Nevertheless, model-selection mistakes are inevitable in finite samples, so that pointwise inference is again unreliable. The lack of uniformity for pre-testing or post-selection estimators has been widely noted in the classic time-series setting by Shibata (1986); Pötscher (1991); Kabaila (1995); and Leeb and Pötscher (2005).

To remedy this issue, we develop uniformly valid inference in the spirit of Mikusheva (2012); Andrews and Cheng (2012); Andrews, Cheng, and Guggenberger (2020); and Belloni, Chernozhukov, and Hansen (2014) in different contexts, on volatility over a large class of MA(∞) models that allow for an asymptotically vanishing noise with a flexible dependence structure. Our inference is thereby more reliable than that of the realized volatility, which simply ignores the impact of small noise when it is difficult to detect. Our inference also allows for model-selection mistakes, which surely occur in the case of an MA(∞) data-generating process, and is therefore robust to the dependence structure of noise.

The crux of our uniformity results is that the convergence rate of our estimator depends on various sequences of noise DGPs. Similar to our estimator but in the case of white noise, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) show that as the variance of the noise vanishes, the convergence rate of the realized kernel estimator improves from $n^{1/4}$ to $n^{1/2}$ with the optimal choice of bandwidth. Their inference, however, is not uniformly valid because a gap remains between the small-noise regime they consider and the no-noise regime. More specifically, they require the noise variance to be greater than n^{-1}. This seemingly small gap is not innocuous, because in a finite sample, the noise magnitude may fall into this gap, resulting in a distortion in the prescribed asymptotic distribution. In contrast, Jacod and Protter (2011) (Theorem 16.5.7) study the pre-averaging estimator as the noise variance vanishes at the rate of $n^{-\eta}$ for $\eta \geq 0$. They do not, however, provide a uniformly valid inference procedure; Also, their inference requires knowledge of η, and the convergence rate of their estimator cannot exceed $n^{1/3}$.[1]

In the case of serially correlated noise, the unknown dependence structure may further plague the efficiency. The pre-averaging estimator by Jacod, Li, and Zheng (2019) and the flat-top realized kernel estimator by Varneskov (2016) converge at the rate of $n^{1/4}$. They do not consider alternative sequences of noise DGPs (particularly those in which noise magnitude and dependence structure interact) that may influence the convergence rate and validity of their inference. We investigate various noise DGPs with simultaneous infinite-order autocorrelation and arbitrarily shrinking magnitude. We show that the convergence rate of our estimator dominates $n^{1/4}$ as noise vanishes, and is determined by the order of noise dependence when noise is sufficiently small. While it is appealing to consider data-driven order selection, such as information criteria, for efficiency gains, in light of the critique by Leeb and Pötscher (2005), we adopt a slightly more conservative order selection procedure than the Akaike information criterion (AIC). As such, our inference remains uniformly valid with a slight efficiency loss only in the small-noise regime.

The literature on the estimation of quadratic variation using noisy high-frequency data is enormous. Earlier works mainly tackle a white microstructure noise.[2] In this paper, we target serially correlated noise using a likelihood-based approach. Hansen, Large, and Lunde (2008) first shed light on the asymptotic equivalence between the maximum likelihood estimator and MA filters. They implement the MA(q) estimator and demonstrate its desirable performance in extensive simulations with various noise models. Related work that discusses serially dependent noise also include Kalnina and Linton (2008); Bandi and Russell (2008); Aït-Sahalia, Mykland, and Zhang (2011); Hautsch and Podolskij (2013); Bibinger, Hautsch, Malec, and Reiß (2019); and Li, Laeven, and Vellekoop (2020). Their assumptions on noise, however, are more restrictive than in our setting.

Our paper is organized as follows. Section 2 sets up the model, Section 3 presents the main results, Section 4 provides simulation evidence, and Section 5 concludes. The appendix contains the proof of the main theorem, and the online supplemental appendix provides the proofs of the corollary, proposition, and technical lemmas.

2 Model Setup

We start with notation. For any matrix A, $A^\intercal$ denotes its transpose. We denote by $\delta_{i,j}$ the Kronecker delta. The imaginary unit and the indicator function are written as $\mathbb{i}$ and $\mathbb{1}_{\{\cdot\}}$, respectively. All

[1] Jacod and Protter (2011) impose an "essentially white" noise assumption, which requires that conditionally on the latent process, the noise is centered and independent, although not necessarily identically distributed. For the sake of presentation, we do not distinguish this type of noise from white noise, since it is also serially uncorrelated, unlike the "colored" noise setting we consider.

[2] Prominent estimators include, but are not limited to, two-scale or multi-scale estimators by Zhang, Mykland, and Aït-Sahalia (2005) and Zhang (2006); the realized kernel estimator and its extensions by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011); the pre-averaging estimator by Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Jacod, Podolskij, and Vetter (2010); the local method of moments estimator by Reiß (2011); and likelihood-based estimators by Aït-Sahalia, Mykland, and Zhang (2005) and Xiu (2010).

vectors are column vectors. We write (a, b, c) in place of $(a^\intercal, b^\intercal, c^\intercal)^\intercal$ for simplicity. d-dimensional vectors of 0s and 1s are written as 0_d and 1_d. We use $\|\cdot\|$ to denote the $\mathbb{L}^2$ norm. We use B to denote the backward (lag) operator associated with discrete-time time series. We use K as a generic positive constant that may vary from line to line but not depend on n. All limits are taken as $n \to \infty$. We use $\xrightarrow{\mathcal{L}}$ to denote convergence in law. We write $a_n \lesssim b_n$ if $a_n \leq K b_n$ for all n. We write $a_n \sim b_n$ if $a_n \lesssim b_n \lesssim a_n$. We use $a \vee b$ and $a \wedge b$ to denote $\max\{a, b\}$ and $\min\{a, b\}$, respectively. We use a superscript (n) to facilitate discussion of uniformity over different sequences of data-generating processes (DGPs) indexed by n.

At each stage $n \geq 1$, transaction prices $\widetilde{X}$ are observed at time points $0 = t_0 < t_1 < \ldots < t_{n_T} \leq T$, where T is fixed. Throughout, we assume n_T, the number of observations within $[0, T]$, is an observed random variable, whereas n is a non-observable mathematical abstraction. We let $\Delta_n = T/n_T$. We assume $\widetilde{X}_{t_i}$ comprises two components:

$$\widetilde{X}_{t_i} = X_{t_i} + U_i, \quad 0 \leq i \leq n_T,$$

where X_{t_i} is (the logarithm of) the efficient equilibrium price and U_i is the microstructure noise associated with the ith observation.

Specifically, with respect to the efficient price, we assume the following:

Assumption 1. *The logarithm of the efficient price process X_t is an Itô-semimartingale defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and satisfies*

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + (\delta \mathbb{1}_{\{|\delta| \leq 1\}}) \star (\underline{\mu} - \underline{\nu})_t + (\delta \mathbb{1}_{\{|\delta| > 1\}}) \star \underline{\mu}_t, \tag{2.1}$$

where μ_t and σ_t are adapted and locally bounded, W is a standard Brownian motion, and $\underline{\mu}$ is a Poisson random measure on $\mathbb{R}_+ \times E$, where E is a Polish space. The compensator $\underline{\nu}$ satisfies $\underline{\nu}(dt, du) = dt \otimes \lambda(du)$ for some σ-finite measure λ on E. Moreover, $|\delta(\omega, t, u)| \wedge 1 \leq \Gamma_m(u)$ for all (ω, t, u) with $t \leq \tau_m(\omega)$, where $\{\tau_m\}$ is a localizing sequence of stopping times and $\{\Gamma_m\}$ a sequence of deterministic functions satisfying $\int \Gamma_m^r(u) \lambda(du) < \infty$ for some $r \in [0, 1)$.

In addition, the process $Z_t = (\mu_t, \sigma_t^2)$ is also an Itô-semimartingale on the space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ with the form

$$Z_t = Z_0 + \int_0^t \widetilde{\mu}_s ds + \int_0^t \widetilde{\sigma}_s d\widetilde{W}_s + (\widetilde{\delta} \mathbb{1}_{\{|\widetilde{\delta}| \leq 1\}}) \star (\underline{\mu} - \underline{\nu})_t + (\widetilde{\delta} \mathbb{1}_{\{|\widetilde{\delta}| > 1\}}) \star \underline{\mu}_t, \tag{2.2}$$

where $\widetilde{\mu}_t$ and $\widetilde{\sigma}_t$ are locally bounded adapted processes, $\widetilde{W}$ is a multivariate Brownian motion, potentially correlated with W, and $\widetilde{\delta}$ is a predictable function such that for some deterministic function $\widetilde{\Gamma}_m(u)$, $\|\widetilde{\delta}(\omega, t, u)\| \wedge 1 \leq \widetilde{\Gamma}_m(u)$ for all $\omega \in \Omega$, $t \leq \tau_m(\omega)$, and $\int \widetilde{\Gamma}_m^2(u) \lambda(du) < \infty$.

Assumption 1 allows for the leverage effect and jumps in both the efficient price and its volatility. It accommodates most models in asset pricing and is commonly used to derive in-fill asymptotic results for high-frequency data—for example, Jacod and Protter (2011) and Aït-Sahalia and Jacod (2014), with notable exclusions of long-memory volatility models driven by fractional Brownian motions (Comte and Renault (1996, 1998)).

The parameter of interest is the quadratic variation of X (scaled by T^{-1}), which comprises both continuous and discontinuous components:

$$C_T = \frac{1}{T}\int_0^T \sigma_t^2 dt + \frac{1}{T}\sum_{0\le t\le T}(\Delta X_t)^2,$$

where $\Delta X_t = X_t - X_{t-}$. Although estimating the integrated volatility or the jump component of the quadratic variation is of tremendous interest, we do not pursue this agenda in this paper, in which we aim for a practical volatility estimate that depends on as few tuning parameters as possible.[3]

Next, we make an assumption on the arrival of trades:

Assumption 2. *For each $n \ge 1$, the sequence of observation times $\{t_i : i \ge 0\}$ satisfies $t_0 = 0$ and $t_i = t_{i-1} + \frac{T}{n}\xi_{t_{i-1}}\chi_i$, where the sequence $\{\chi_i : i \ge 1\}$ is i.i.d., $(0,\infty)$-valued, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, and independent of the σ-field $\mathcal{F}_\infty = \bigvee_{t>0}\mathcal{F}_t$, with $m_j = \mathbb{E}((\chi_i)^j) < \infty$ and $m_1 = 1$, for all $j > 0$. In addition, the process $\xi = (\xi_t)_{t\ge 0}$ is a nonnegative Itô-semimartingale defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ in the form of (2.2), such that neither ξ_t nor ξ_{t-} vanishes.*

Assumption 2 allows the arrival rate of transactions to depend on their prices through ξ_t. It also accommodates regular sampling, time-changed regular sampling, Poisson sampling, modulated Poisson sampling, and predictably modulated random-walk sampling schemes, as discussed in detail by Jacod, Li, and Zheng (2017).[4] We introduce here and below several stochastic processes, e.g., ξ_t and η_t, for which their driving Brownian motions (implicitly defined) are different from W in Assumption 1, but possibly correlated. Note that finding a single Poisson measure that drives the jumps of all processes involved is always possible.

Finally, we assume the noise is endogenous, heteroscedastic, and serially correlated.

Assumption 3. *For each $n \ge 1$, the noise sequence $\{U_i : i \ge 0\}$ consists of random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which have the following MA(∞) representation:*

$$U_i = \eta_{t_i}\iota^{(n)}\theta^{(n)}(\mathrm{B})\varepsilon_i, \quad \textit{with} \quad \theta^{(n)}(x) = 1 + \sum_{j=1}^{\infty}\theta_j^{(n)}x^j, \tag{2.3}$$

[3] To achieve robustness to serially correlated noise, prominent nonparametric estimators require three tuning parameters (see two such estimators in our simulation study). Exploring the finer structure of the quadratic variations would require at least one more—rendering these measures impractical to estimate, in particular for illiquid stocks.

[4] Our sampling scheme imposes that conditional on $\mathcal{F}_{t_{i-1}}$, t_i is independent of X_t for $t \ge t_{i-1}$, which we need to derive a desirable central limit theory. This assumption conforms with Assumption O(ii) of Jacod, Li, and Zheng (2017), but is more restrictive than those adopted by Li, Mykland, Renault, Zhang, and Zheng (2014) and Fukasawa (2010), both of which find an asymptotic bias in the CLT of the realized volatility estimator associated with their general sampling scheme (in the absence of microstructure noise). On a related note, Renault and Werker (2011) find that the instantaneous causality relations between price volatility and durations of trades could lead to severely biased volatility estimates.

where $\varepsilon_i \overset{i.i.d.}{\sim} (0,1)$*, defined on* $(\Omega, \mathcal{F}, \mathbb{P})$*, is independent of* $\mathcal{F}_\infty$ *and* $\{\chi_i : i \geq 1\}$*, and has finite moments of all orders;* $(\eta_t)_{t\geq 0}$ *is an* $(\mathcal{F}_t)$*-adapted nonnegative Itô-semimartingale that satisfies the same form of (2.2); and* $\iota^{(n)}$ *is a deterministic nonnegative number that characterizes the noise magnitude and satisfies* $\iota^{(n)} \leq K$*.*[5]

Assumption 3 accommodates several empirical features of the microstructure noise. The noise process depends on price X through η, since η_t is $\mathcal{F}_t$-adapted, which may be driven by a Brownian motion and a Poisson random measure that are correlated with X. Such dependence is potentially driven by comovement between the price and bid-ask spread or the discreteness of the observed price.[6] That said, this assumption implies zero correlation between any function of the path of X and U_i for each i—the key identifying assumption that separates efficient returns from noise. A fully specified structural microstructure model would be necessary, along with additional observables (e.g., bid-ask prices), if some non-vanishing correlation between X and U were allowed for. In this paper, we avoid imposing additional structural assumptions, and instead focus on the reduced-form model of $\widetilde{X}$, while being agnostic about the economic implications of reduced-form parameters; e.g., θ and ι^2. Many structural models yield specific reduced-form ARMA models of returns; for example, Hasbrouck (2007), with differences only in how the reduced-form parameters relate to structural parameters. Estimating and interpreting structural parameters in a microstructure model is interesting, but we leave this for future work.

The noise process features flexible serial correlations through its $\theta^{(n)}(\mathrm{B})\varepsilon$ component, as specified by an MA(∞) model. The next assumption spells out restrictions on its spectral density function, $g(\lambda; \theta^{(n)}) = |\theta^{(n)}(e^{\mathrm{i}\lambda})|^2$, such that the sequence of MA processes is uniformly invertible and their long-range serial dependence cannot be arbitrarily strong.

Assumption 4. *For each* $n \geq 1$*, the spectral density function of* $\theta^{(n)}(\mathrm{B})\varepsilon$ *satisfies, for some fixed* $\alpha > 3$*,*

$$\inf_\lambda g(\lambda; \theta^{(n)}) \geq \frac{1}{K} \quad \textit{and} \quad \left| \int_{-\pi}^{\pi} g(\lambda; \theta^{(n)}) e^{\mathrm{i}\lambda j} d\lambda \right| \leq K j^{-\alpha}, \quad \forall j \geq 0.$$

We next introduce our likelihood-based estimator.

3 Main Results

3.1 Likelihood-based Estimation

In contrast to existing nonparametric estimators, we construct a quasi-maximum likelihood estimator (QMLE) in the spirit of White (1982) by imposing a misspecified parametric model, for which the

[5] The probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ can be constructed more explicitly. Specifically, we define X, Z, ξ, and η (which satisfy the relevant assumptions) on a space $(\Omega_{(0)}, \mathcal{F}_\infty, (\mathcal{F}_t), \mathbb{P}_{(0)})$, and define $\{\chi_i\}$ and $\{\varepsilon_i\}$ on a different space $(\Omega_{(1)}, \mathcal{F}_{(1)}, \mathbb{P}_{(1)})$. We then set $\Omega = \Omega_{(0)} \times \Omega_{(1)}$, $\mathcal{F} = \mathcal{F}_\infty \otimes \mathcal{F}_{(1)}$, and $\mathbb{P}(d\omega_{(0)}, d\omega_{(1)}) = \mathbb{P}_{(0)}(d\omega_{(0)})\mathbb{P}_{(1)}(d\omega_{(1)})$.

[6] No one has yet built a perfect model that is meant to be a full describer of reality. Like Jacod, Li, and Zheng (2017) and Jacod, Li, and Zheng (2019), the type of rounding errors our setting allows for is still somewhat restrictive. In the simulation analysis, we investigate a particular form of rounding that appears more realistic but violates Assumption 3, and we show that the rounding effect is negligible.

likelihood function is available:

$$dX_t = \sigma dW_t; \quad U_i = \iota\theta(\mathrm{B})\varepsilon_i, \quad \text{with} \quad \theta(x) = 1 + \sum_{j=1}^{q} \theta_j x^j, \quad \text{and} \quad \varepsilon_i \sim \mathcal{N}(0,1).$$

In other words, we pretend the efficient price (in logarithm) is a Brownian motion with constant volatility but no drift, and that the noise follows a Gaussian MA(q) model with the order q to be determined. Under this model, the observed log-return vector $Y_n = (Y_{n,1}, Y_{n,2}, \dots, Y_{n,n_T})^\intercal$, which is defined as

$$Y_{n,i} = X_{t_i} - X_{t_{i-1}} + U_i - U_{i-1}, \quad 1 \le i \le n_T, \tag{3.4}$$

follows a reduced-form Gaussian MA($q+1$) model. Its $n_T \times n_T$ covariance matrix Σ_n is given by

$$\Sigma_n(\sigma^2, \iota^2, \theta) = \sigma^2 \Delta_n \mathbb{I}_n + \sum_{h=0}^{n_T-1} (2\gamma_h - \gamma_{h+1} - \gamma_{h-1}) \mathbb{G}_n^h, \tag{3.5}$$

where $(\mathbb{I}_n)_{ij} = \delta_{i,j}$, $(\mathbb{G}_n^h)_{ij} = \delta_{h,|i-j|}$, and γ_h is the h-th order autocovariance of U:

$$\gamma_h = \frac{\iota^2}{2\pi} \int_{-\pi}^{\pi} g(\lambda;\theta) e^{\mathbb{i}\lambda h} d\lambda, \quad \text{where} \quad g(\lambda;\theta) = \left|\theta(e^{\mathbb{i}\lambda})\right|^2. \tag{3.6}$$

Because θ is a nuisance parameter for volatility estimation and is unidentified if $\iota = 0$, we reparameterize the likelihood function in terms of strongly identified parameters (σ^2, γ):

$$L_n(\sigma^2, \gamma) = -\frac{1}{2} \log\det(\Sigma_n(\sigma^2, \gamma)) - \frac{1}{2} \mathrm{tr}(\Sigma_n(\sigma^2, \gamma)^{-1} Y_n Y_n^\intercal), \tag{3.7}$$

where $\Sigma_n(\sigma^2, \gamma) := \Sigma_n(\sigma^2, \iota^2, \theta)$ and γ is the $(q+1)$-dimensional vector of noise autocovariances.[7] We define $(\widehat{\sigma}_n^2(q), \widehat{\gamma}_n(q))$ as the maximizer of $L_n(\sigma^2, \gamma)$:

$$(\widehat{\sigma}_n^2(q), \widehat{\gamma}_n(q)) = \arg \max_{(\sigma^2,\gamma)\in\Pi_n(q)} L_n(\sigma^2, \gamma). \tag{3.8}$$

The parameter space of (σ^2, γ), denoted by $\Pi_n(q)$, can be derived from the usual condition that $(\sigma^2, \iota^2, \theta)$ satisfy, i.e., $\inf_\lambda f(\lambda;\gamma) \ge 0$, where $f(\lambda;\gamma) = \iota^2 g(\lambda;\theta)$. However, it is not ideal for reasons we now explain.

Aït-Sahalia and Xiu (2019) show that in the white-noise case, if the noise magnitude is small, the noise variance estimator $\widehat{\iota}_n^2$ hits the boundary zero, so that the asymptotic distribution of the volatility estimator, $\widehat{\sigma}_n^2$, becomes nonstandard. A similar yet more severe issue occurs here: The estimate $\widehat{\gamma}$ hits the boundary $\inf_\lambda f(\lambda;\widehat{\gamma}) = 0$ with nontrivial probability.

An easy solution in the white-noise case is to enlarge the parameter space of the nuisance parame-

[7] Note that $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_q)$, which is different from how vectors are typically indexed. For convenience, we often treat γ as an infinite dimensional vector, with 0s filled beyond the $(q+1)$th entry of γ when no ambiguity exists.

ter, allowing for negative values of $\widehat{\iota}_n^2$, so that the asymptotic distribution of $\widehat{\sigma}_n^2$ is not affected by confinement of the parameter space for ι^2. We adopt a similar strategy to enlarge the parameter space of (σ^2, γ) to $\{(\sigma^2, \gamma) \in \mathbb{R}^{q+2} : \inf_\lambda f(\lambda; \sigma^2, \gamma, \Delta_n) \geq 0\}$, where $f(\lambda; \sigma^2, \gamma, \Delta_n) = \sigma^2\Delta_n + |1 - e^{\mathrm{i}\lambda}|^2 f(\lambda; \gamma)$ is the spectral density of Y_n under the quasi model. In other words, the parameter space is enlarged such that only a reduced-form MA$(q+1)$ model of observed returns is required to be well defined (see Section 3.5), and that a well-defined decomposition of observed returns, as in (3.4), may not exist. On the other hand, the parameter space must be sufficiently "local" to the true value to avoid spurious estimates due to potential use of an overly flexible quasi-model (e.g., q is too large). For this purpose, we choose the following set that imposes constraints on the lower bound of the spectral density function and the decay of autocovariances:[8]

$$\Pi_n(q) = \Big\{(\sigma^2, \gamma) \in \mathbb{R}^{q+2} : \inf_\lambda f(\lambda; \sigma^2, \gamma, \Delta_n) \geq \frac{\Delta_n}{K}, \quad \sigma^2 + |\gamma_0| + \frac{\sum_{j=1}^\infty j^2|\gamma_j|}{\inf_\lambda |\sigma^2\Delta_n + f(\lambda; \gamma)|} \leq K\Big\}. \quad (3.9)$$

This parameter space depends on the order of the MA model, q, which we next discuss how to select.

3.2 Model Selection

To determine an appropriate order q, we use AIC, which in our setting can be written as

$$\mathrm{AIC}_n(q) = 2q - 2 \max_{(\sigma^2,\gamma)\in\Pi_n(q)} L_n(\sigma^2, \gamma).$$

Our choice of order q will be based on (but not necessarily identical to)

$$\widehat{q}_{n,\mathrm{AIC}} = \arg\min_{q \leq n_T^{1/3}} \mathrm{AIC}_n(q). \quad (3.10)$$

More generally, in Theorem 1 below, we spell out the conditions a desirable order $\widehat{q}_n$ must satisfy in order to accommodate uniformly valid inference on volatility for a large class of DGPs.

Similar to the case of AR(∞) in Shibata (1980), the upper bound on q precludes MA models with too many parameters from estimation. Asymptotically, this upper bound is not binding, because for all sequences of noise DGPs we consider, $\widehat{q}_{n,\mathrm{AIC}} = o_{\mathrm{P}}(n^{1/6})$—a claim we prove in Lemma B4 in the online supplemental appendix.

In a companion paper, Da and Xiu (2021) prove model selection consistency (based on BIC) and provide pointwise asymptotic inference on noise autocovariance parameters when the noise follows an MA$(q^\star)$ model with a finite $q^\star$. The pointwise asymptotic theory relies on this fixed DGP, as well as this unrealistic result of perfect model selection; hence, it provides a misleading picture of the actual finite-sample behavior of the inference. As shown in the classic time-series setting of Leeb and Pötscher (2005), conducting uniformly valid post-selection inference on parameters over a nontrivially large class of DGPs is generally impossible. For volatility estimation in our setting,

[8]The constraint (3.9) is essential for proofs. We do not find it critical to impose in our implementation.

however, uniformly valid inference is possible for a wide class of DGPs, which we turn to next.

3.3 Uniform Inference on Volatility

Obviously, the class of DGPs cannot be arbitrarily large, so we need restrictions on how the magnitude of the noise and its autocorrelation structure vary with sample size. We denote by $\kappa^{(n)}$ the ∞-dimensional vector of autocovariances of $\theta^{(n)}(\mathrm{B})\varepsilon$, whose components are given by

$$\kappa_j^{(n)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda;\theta^{(n)})e^{\mathrm{i}\lambda j}d\lambda, \quad j \geq 0. \tag{3.11}$$

The class of noise models we consider satisfies:

Assumption 5. *For any $0 < k < K$ and any sequence $\alpha_n \to \infty$, we have*

$$q_n^\star(k) = o(n^{1/3}(\iota^{(n)} \vee n^{-1/2})^{4/9}), \quad \psi_n^2 \sum_{j=q_n^\star(k)}^{\infty} |\kappa_j^{(n)}| = o\Big(\frac{q_n^\star(k)^{1/2}+\alpha_n}{n^{1/2}} + \frac{\sqrt{\iota^{(n)}}}{n^{1/4}}\Big),$$

where

$$q_n^\star(k) := \min q, \quad \textit{subject to} \quad n\psi_n^4 \sum_{j=q+1-\alpha_n}^{2q} |\widetilde{\kappa}_j^{(n)}|^2 \leq kq,$$

$\psi_n := (1+n^{-1/2}/\iota^{(n)})^{-1}$, *and* $\widetilde{\kappa}_j^{(n)} := \sum_{i=0}^{\infty}(i+1)\psi_n^i(2\kappa_{j+1+i}^{(n)} - \kappa_{j+i+2}^{(n)} - \kappa_{j+i}^{(n)})$.

Intuitively, $q_n^\star(k)$ mimics the "oracle" order that AIC selects. Assumption 5 effectively requires that this order cannot be too large (the first equation above) and that the approximation error induced by selection (the left-hand side of the second equation above) is asymptotically dominated by the estimation error (the right-hand side). Next, we provide two examples to demonstrate that the conditions in Assumption 5 are not restrictive from a practical point of view.

Example 1: Suppose $n^{1/2}\iota^{(n)} \to \infty$ and $\theta^{(n)}(\mathrm{B})\varepsilon$ follows an MA(∞) model with $|\kappa_j^{(n)}| \sim j^{-\alpha}$ for some $\alpha > 3 \vee \frac{2}{1+2\log \iota^{(n)}/\log n}$. It is easy to show that Assumption 5 holds, because

$$q_n^\star(k) \sim n^{1/(2\alpha)} \quad \text{and} \quad \psi_n^2 \sum_{j=q_n^\star(k)}^{\infty} |\kappa_j^{(n)}| \sim n^{-1/2+1/(2\alpha)} = o\big(n^{-1/4}(\iota^{(n)})^{1/2}\big).$$

Jacod, Li, and Zheng (2017) assume $|\kappa_j^{(n)}| \sim j^{-\alpha}$ with $\alpha > 3$ and a fixed $\iota^{(n)}$. Our condition further sheds light on a trade-off between $\iota^{(n)}$ and α: As $\iota^{(n)}$ shrinks, $|\kappa_j^{(n)}|$ must decay faster.

Example 2: Suppose $\theta^{(n)}(\mathrm{B})\varepsilon$ follows an arbitrary ARMA(p,q) process with finite p and q. Assumption 5 holds because in this case, as long as $\iota^{(n)} \lesssim 1$ (it can shrink arbitrarily fast),

$$q_n^\star(k) \lesssim \log n \quad \text{and} \quad \psi_n^2 \sum_{j=q_n^\star(k)}^{\infty} |\kappa_j^{(n)}| \lesssim n^{-1/2}\big(1+o(q_n^\star(k)^{1/2})\big).$$

Now we are ready to present the main theoretical result, based on which we build uniformly valid inference on volatility:

Theorem 1. *Suppose we select an order* $\widehat{q}_n = \widehat{q}_{n,\mathrm{AIC}} \vee \alpha_n$ *with* $\alpha_n = O(\log n)$. *Then, for all sequences of DGPs satisfying Assumptions 1 - 5, and as* $\widehat{q}_n \vee (n^{1/2}\iota^{(n)}) \to \infty$, *we have*[9]

$$\frac{\widehat{\sigma}^2(\widehat{q}_n) - C_T}{\sqrt{\mathrm{AVAR}(\widehat{q}_n, n)_T}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

where $\mathrm{AVAR}(q,n)_T$ *is given by*

$$\mathrm{AVAR}(q,n)_T = \frac{1}{n}\left[(4q+6)\,E(4,\xi)_T + \Delta_n^{-1/2}\zeta^{(n)}\left(5E(4,\xi)_T C_T^{-1/2} + C_T^{3/2}B(\xi)_T\right)\right], \tag{3.12}$$

$(\zeta^{(n)})^2$ *is the "long-run variance" of the general noise process, given by*

$$(\zeta^{(n)})^2 = (\iota^{(n)})^2 g(0,\theta^{(n)})\frac{\int_0^T \eta_s^2 \xi_s^{-1} ds}{\int_0^T \xi_s^{-1} ds},$$

$E(4,\xi)_T$ *is a general "quarticity" in the presence of random sampling and jumps, given by*

$$E(4,\xi)_T = \frac{1}{T}\int_0^T \xi_s \sigma_s^4 ds + \frac{1}{T}\sum_{s\le T}(\Delta X_s)^2(\xi_s\sigma_s^2 + \xi_{s-}\sigma_{s-}^2), \tag{3.13}$$

and

$$B(\xi)_T = \frac{2\int_0^T \eta_s^2\sigma_s^2 ds + \sum_{s\le T}(\Delta X_s)^2(\eta_s^2 + \eta_{s-}^2)}{C_T \times \int_0^T \eta_s^2 \xi_s^{-1} ds} + \frac{T\int_0^T \eta_s^4 \xi_s^{-1} ds}{\left(\int_0^T \eta_s^2 \xi_s^{-1} ds\right)^2}. \tag{3.14}$$

Combining with asymptotic variance estimators in Section 3.4, we immediately obtain:

Corollary 1. *Suppose the same assumptions as those in Theorem 1 hold. Let* $c_{1-\alpha} = F^{-1}(1-\alpha/2)$, *where* $F(\cdot)$ *is the standard Gaussian cumulative distribution function. We have*

$$\lim_{n\to\infty} \mathbb{P}\left(C_T \in CI_n(\alpha)\right) = 1 - \alpha,$$

where, using $\widehat{\zeta}_n^2 = \sum_{j=-\widehat{q}_n}^{\widehat{q}_n} \widehat{\gamma}_n(\widehat{q}_n)_{|j|}$, *the uniformly valid confidence interval is constructed as*

$$CI_n(\alpha) = \left[\widehat{\sigma}_n^2(\widehat{q}_n) \pm c_{1-\alpha} n_T^{-1/2}\sqrt{(4\widehat{q}_n + 6)\,\widehat{E}_n(4)_T + \frac{\widehat{\zeta}_n^2}{\Delta_n^{1/2}}\left(5\widehat{E}_n(4)_T \widehat{\sigma}_n^2(\widehat{q}_n)^{-1/2} + \widehat{\sigma}_n^2(\widehat{q}_n)^{3/2}\widehat{B}_n(\widehat{q}_n)_T\right)}\,\right].$$

To shed light on the asymptotic behavior of our estimator, we examine two special DGP sequences:

[9] To ensure $\widehat{q}_n \vee (n^{1/2}\iota^{(n)}) \to \infty$ holds without worrying about $n^{1/2}\iota^{(n)}$ and $\widehat{q}_{n,\mathrm{AIC}}$, we can select α_n such that $\alpha_n \to \infty$, say, $\alpha_n \sim \log n$ as in our implementation, although the statement of Theorem 1 does not require this.

i. Under $n^{1/2}\iota^{(n)}/(4\widehat{q}_n+6)\to\infty$,

$$\mathrm{AVAR}(\widehat{q}_n,n)_T = n^{-1/2}T^{-1/2}\zeta^{(n)}\left(5E(4,\xi)_T C_T^{-1/2} + C_T^{3/2}B(\xi)_T\right) + o_{\mathrm{P}}(n^{-1/2}\iota^{(n)}). \tag{3.15}$$

ii. Under $n^{1/2}\iota^{(n)}/(4\widehat{q}_n+6)\to 0$,[10]

$$\mathrm{AVAR}(\widehat{q}_n,n)_T = \frac{1}{n}(4\widehat{q}_n+6)E(4,\xi)_T + o_{\mathrm{P}}\Big(\frac{\widehat{q}_n+1}{n}\Big). \tag{3.16}$$

Case i describes the behavior of our estimator in the presence of "large" noise. The convergence rate is $(\iota^{(n)})^{-1/2}n^{1/4}$, which varies within $[n^{1/4}, n^{1/2}\widehat{q}_n^{-1/2}]$. This result echoes and extends that of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) for the realized kernel estimator whose rate varies within $[n^{1/4}, n^{1/2}]$ in the case of i.i.d. noise. As to the colored noise, our rate dominates $n^{1/4}$—the convergence rate of flat-top realized kernel and pre-averaging estimators by Varneskov (2016) and Jacod, Li, and Zheng (2019).

In the case of small noise (Case ii), the convergence rate is prescribed by $n^{1/2}\widehat{q}_n^{-1/2}$. When noise is absent, Case ii also shows that the efficiency loss compared with the realized volatility estimator is given by a factor $2\widehat{q}_n+3$, because realized volatility has knowledge of the absence of noise. Moreover, the bias of the realized volatility estimator is of order $(\iota^{(n)})^2 n$, which may not vanish in Case ii, because noise is not entirely negligible in this regime.

We now explain our choice of $\widehat{q}_n$. Recall that the noise-dependence structure follows MA(∞). Intuitively, a smaller choice of $\widehat{q}_n$ leads to a more efficient estimator at the risk of a larger bias due to model misspecification ($\widehat{q}_n<\infty$). In contrast to the somewhat ad hoc tuning parameters other approaches rely on, our estimate $\widehat{q}_{n,\mathrm{AIC}}$ is informative about the minimal order using which the model misspecification bias is negligible. The importance of this guidance on q is manifested in Case ii, in which the convergence rate clearly improves as $\widehat{q}_n$ decreases.

Nonetheless, instead of fully relying on $\widehat{q}_{n,\mathrm{AIC}}$, Theorem 1 requires the use of a certain $\widehat{q}_n = \widehat{q}_{n,\mathrm{AIC}}\vee\alpha_n$ that also approaches ∞ slowly if $n^{1/2}\iota^{(n)}$ is bounded, even when the true model may be of a finite order (and hence $\widehat{q}_{n,\mathrm{AIC}}$ is small). Indeed, if the true model is a finite-order MA(q), we can show that QMLE based on $\widehat{q}_{n,\mathrm{AIC}}$ can achieve a convergence rate as fast as $n^{1/2}$ in Case ii. However, the asymptotic distribution is highly nonstandard, because the model-selection bias is of an order comparable to the estimation error. For this reason, we intentionally inflate the order of the employed model, requiring $\widehat{q}_n\to\infty$, so that a standard asymptotic normal distribution is available in Case ii. α_n is the single tuning parameter required by our procedure. One possible choice of α_n is $\log n_T$, which (potentially) inflates $\widehat{q}_{n,\mathrm{AIC}}$ by $\log n_T$. The choice of $\widehat{q}_n$ does not affect the asymptotic variance in Case i, since (3.15) in fact does not rely on $\widehat{q}_n$, but it may hurt the efficiency of our estimator in Case ii. As a result, our rate in Case ii is strictly smaller than $n^{1/2}$ under the conditions

[10] In this case, Theorem 1 requires that $\widehat{q}_n$ approaches ∞, so that $4\widehat{q}_n+6$ and $\widehat{q}_n$ are in fact of the same order. That said, we prefer this small-sample adjustment that can be established in the case of a finite $\widehat{q}_n$.

in Theorem 1. This efficiency cost is in fact unavoidable for the sake of uniformity, because of the following "impossibility" result in the spirit of Leeb and Pötscher (2008).

To demonstrate this result, we consider a simple setting in which the noise process has no autocorrelation beyond the first lag (so that we use AIC to select q from $\{0, 1\}$), and the noise magnitude $(\iota^{(n)})^2$ is dominated by n^{-1} (so that the optimal rate of the volatility estimator is $n^{1/2}$). The next proposition shows that even with constant volatility, no uniformly consistent estimator exists for the cumulative distribution function $G_n(x)$, where

$$G_n(x) = \mathbb{P}\left(n^{1/2}(\widehat{\sigma}_n^2(\widehat{q}_{n,\mathrm{AIC}} \wedge 1) - C_T) \leq x\right).$$

Proposition 1. *For each $x \in \mathbb{R}$, there exists a DGP sequence satisfying Assumptions 1 - 5 with $\sigma_t^2 = C_T$ for some C_T fixed and all $t \in [0, T]$, $n(\iota^{(n)})^2 \leq K$, and a single parameter $\theta^{(n)}$, such that*

$$\liminf_{n\to\infty} \inf_{\widehat{G}_n(x)} \mathbb{P}\left(|\widehat{G}_n(x) - G_n(x)| > \frac{1}{K}\right) > 0,$$

where the infimum extends over all estimators $\widehat{G}_n(x)$ of $G_n(x)$.

On a different note, Theorem 1 establishes that our asymptotic distribution is conditionally Gaussian, which is typically the case for estimating quadratic variation in the presence of noise. Nonetheless, in the absence of noise, the limiting distribution for the realized volatility estimator is a mixture of Gaussian random variables with square root of uniform random variables around the jump times instead of Gaussian, unless the volatility and price processes do not jump together; see e.g., Theorem 5.4.2 of Jacod and Protter (2011). Because of this, their inference procedure is sufficiently complicated that simulations are entailed in order to achieve a sharp confidence interval; see page 349 of Aït-Sahalia and Jacod (2014). When $n^{1/2}\iota^{(n)} \leq K$, our asymptotic distribution would run into the same issue if $\widehat{q}_n$ is finite. Interestingly, while the requirement on $\widehat{q}_n \to \infty$ is motivated from uniformity considerations, we show that this condition also leads to a conditional Gaussian limiting distribution, which facilitates our inference procedure.

Another related point is that the asymptotic distribution of our volatility estimator does not depend on the true distribution of ε. This is not surprising in the case of large noise; see, e.g., Xiu (2010). However, when noise is small, it is possible that certain moments of ε might affect the asymptotic variance of volatility as the convergence rate improves. In this regime, the reason our volatility estimator has the same asymptotic variance regardless of the true distribution of ε is again due to $\widehat{q}_n \to \infty$.

3.4 Asymptotic Variance Estimators

In this section, we develop pre-averaging-based estimators of asymptotic variances. We need two sequences of integers k_n and k'_n, satisfying $k_n \sim n^{2/3}, k'_n \sim n^{7/8}$, and a nonzero real-valued function $g : \mathbb{R} \to \mathbb{R}$, supported on $[0, 1]$, which is continuous and piecewise C^1 with a piecewise Lipschitz derivative g' and $g(0) = g(1) = 0$. We also adopt a truncation strategy (Mancini (2001)) to handle jump-related quantities, for which we define:

$$v_n = \alpha(k_n\Delta_n)^{\varpi}, \quad \text{for some } \alpha > 0, \varpi \in (0, 1/2)\,.$$

We construct the estimator of $E(4,\xi)_T$ in (3.13) as $\widehat{E}_n(4)_T = \widehat{C}_n(4)_T + \widehat{D}_n(4)_T$ using the pre-averaging approach:

$$\widehat{C}_n(4)_T = \frac{1}{Tk_n^2\Delta_n\left(\int_0^1 g(s)^2 ds\right)^2} \sum_{m=1}^{n_T-2k_n} (\bar{Y}(g)_m^n)^2(\bar{Y}(g)_{m+k_n}^n)^2 \mathbb{1}_{\{|\bar{Y}(g)_m^n|\le v_n, |\bar{Y}(g)_{m+k_n}^n|\le v_n\}}, \tag{3.17}$$

$$\widehat{D}_n(4)_T = \frac{1}{Tk_n\int_0^1 g(s)^2 ds} \sum_{m=k_n'+1}^{n_T-k_n-k_n'} (\bar{Y}(g)_m^n)^2 \mathbb{1}_{\{|\bar{Y}(g)_m^n|>v_n\}} \left(\widetilde{c}(g)_m^n + \widetilde{c}(g)_{m-k_n'}^n\right), \tag{3.18}$$

where pre-averaged returns and spot volatilities are given by, respectively,

$$\bar{Y}(g)_i^n = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) Y_{n,i+j}, \quad \widetilde{c}(g)_i^n = \frac{1}{k_n' k_n\Delta_n\int_0^1 g(s)^2 ds} \sum_{m=1}^{k_n'} (\bar{Y}(g)_{i+m}^n)^2 \mathbb{1}_{\{|\bar{Y}(g)_{i+m}^n|\le v_n\}}. \tag{3.19}$$

These estimators are the same as those constructed by Aït-Sahalia and Xiu (2016) for i.i.d. noise. Despite their low convergence rate, these estimators are also consistent in this more general setting, because of the choice of a large local window size k_n which averages out the impact of the dependent noise. Because of the jump truncation, Assumption 1 imposes that $r < 1$, which is necessary for consistency.

Finally, we provide the estimator of $B(\xi)_T$ in (3.14) using

$$\widehat{B}_n(\widehat{q}_n)_T = \left|\frac{1}{\widehat{\sigma}_n^2(\widehat{q}_n)}\left(\frac{1}{\widehat{\gamma}_n(\widehat{q}_n)_0 - \widehat{\gamma}_n(\widehat{q}_n)_1}\right)(\widehat{B}_n'(1) + \widehat{B}_n'(2)) + \left(\frac{1}{\widehat{\gamma}_n(\widehat{q}_n)_0 - \widehat{\gamma}_n(\widehat{q}_n)_1}\right)^2 \widehat{B}_n'(3)\right| \wedge \log n_T, \tag{3.20}$$

where, with $\bar{Y}(g)_m^n$ and $\widetilde{c}(g)_m^n$ defined in (3.19),

$$\begin{aligned}
\widehat{B}_n'(1) &= \frac{1}{n_T}\sum_{m=1}^{n_T-k_n-k_n'} (Y_{n,m})^2 \widetilde{c}(g)_m^n, \\
\widehat{B}_n'(2) &= \frac{1}{2Tk_n\int_0^1 g(s)^2 ds}\sum_{m=1}^{n_T-k_n'} ((Y_{n,m})^2 + (Y_{n,m+k_n'})^2)(\bar{Y}(g)_m^n)^2 \mathbb{1}_{\{|\bar{Y}(g)_m^n|>v_n\}}, \\
\widehat{B}_n'(3) &= \frac{1}{4n_T}\sum_{m=1}^{n_T-k_n'} (Y_{n,m})^2 (Y_{n,m+k_n'})^2.
\end{aligned}$$

3.5 Implementation

We discuss the implementation of QMLE in this section. Apparently, directly calculating the inverse of $\Sigma_n(\sigma^2,\gamma)$ would be computationally expensive when evaluating the likelihood function at each stage of an optimization routine. To avoid this problem, the classic time-series literature adopts an approximation approach of Whittle (1951). Unfortunately, we can show the Whittle estimator is inconsistent in our in-fill asymptotic setting, even if the noise is i.i.d. Gaussian and the efficient price is a Brownian motion with constant volatility (hence, our QMLE is in fact the MLE).

We instead implement exact likelihood through the state-space representation of an MA model. To avoid the issue of weakly identified parameters, our implementation leverages an auxiliary reduced-form MA$(q+1)$ model of the observed noisy returns:

$$Y_{n,i} = \phi(B)\epsilon_i, \quad \text{with} \quad \phi(x) = 1 + \sum_{j=1}^{q+1} \phi_j x^j, \quad 1 \le i \le n, \quad \epsilon \sim \mathcal{N}(0,\chi^2). \tag{3.21}$$

Algorithm 1. *Our algorithm starts as follows:*

1. *Select the optimal order, $\widehat{q}_{n,\mathrm{AIC}}$, of the MA process (3.21) for Y_n using AIC, defined by (3.10) but rewritten in terms of χ^2 and ϕ.*

2. *Obtain exact quasi-likelihood estimates of $\widehat{\chi}^2$ and $\widehat{\phi}_j$ for $1 \le j \le \widehat{q}_n + 1$, using the state-space representation of (3.21) and Kalman filtering (see, e.g., Gardner, Harvey, and Phillips (1980)), where $\widehat{q}_n = \widehat{q}_{n,\mathrm{AIC}} \vee \log n_T$.*[11]

3. *Construct volatility and noise autocovariance estimators using the above estimates:*

$$\begin{aligned}
\widehat{\sigma}_n^2(\widehat{q}_n) =& \Delta_n^{-1}\widehat{\chi}^2 \left(1 + \sum_{j=1}^{\widehat{q}_n+1} \widehat{\phi}_j\right)^2, \\
\widehat{\gamma}_n(\widehat{q}_n)_j =& \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\widehat{\chi}^2 e^{\mathrm{i}j\lambda}}{|1-e^{\mathrm{i}\lambda}|^2} \left(\left|1 + \sum_{l=1}^{\widehat{q}_n+1} \widehat{\phi}_l e^{\mathrm{i}l\lambda}\right|^2 - \left(1 + \sum_{l=1}^{\widehat{q}_n+1} \widehat{\phi}_l\right)^2 \right) d\lambda, \quad 0 \le j \le \widehat{q}_n,
\end{aligned}$$

which are obtained by comparing different parameterizations of the return autocovariances.

Algorithm 1 is sufficient for estimating volatility and noise autocovariance. If we are further interested in (ι^2,θ), a unique solution $(\widehat{\iota}_n^2(q), \widehat{\theta}_n(q))$ exists with probability approaching one when noise is sufficiently large relative to sample size.[12] When noise is small, however, these parameters are weakly identified, and there may not be any solution such that $\widehat{\iota}_n^2(q)$ is positive and $\widehat{\theta}_n(q)$ is real.

[11] Packages of standard programming software (e.g., R and Matlab) are available that implement a likelihood estimator for MA models, despite the fact that some packages rely on Whittle approximations. Our codes are also available upon request.

[12] Da and Xiu (2021) suggest a Newton-Raphson algorithm based on Wilson (1969) to solve for $\widehat{q}_n+1$ model parameters (ι^2,θ) of the MA$(\widehat{q}_n)$ noise process from up-to-$\widehat{q}_n$th-order autocovariances $\widehat{\gamma}_n(\widehat{q}_n)_j$, $0 \le j \le \widehat{q}_n$.

4 Monte Carlo Simulations

In this section, we examine the finite-sample performance of our volatility estimator and compare it with alternative nonparametric estimators in the literature. Throughout, we fix $T = 1$ day and the average sampling frequency every 5 seconds. We conduct $1,000$ Monte Carlo trials in total.

We simulate X_t and σ_t^2 according to the same log-volatility model as in Li and Xiu (2016):

$$\begin{cases} dX_t &= (0.05 + 0.5\sigma_t^2)dt + \sigma_t dW_t + J^X dN_t, \\ \sigma_t^2 &= D_t \exp\left(-2.8 + 6F_t\right), \quad dF_t = -4F_t dt + 0.8d\widetilde{W}_t + J^F dN_t - 0.02\lambda_N dt, \end{cases} \tag{4.22}$$

where $\mathbb{E}[dW_t d\widetilde{W}_t] = -0.8dt$, $J^X \sim \mathcal{N}(0, 0.02^2)$, $J^F \sim \mathcal{N}(0.02, 0.02^2)$, N_t is a Poisson process with intensity $\lambda_N = 25$, and D_t captures the diurnal effect:

$$D_t = 0.75\exp(-10t/T) + 0.25\exp(-10(1 - t/T)) + 0.8.$$

The arrival of trades follows an inhomogeneous Poisson process with rate $nT^{-1}\xi_t^{-1} = nT^{-1}(1 + \cos(2\pi t/T)/2)$, so that fewer trades arrive in the middle of the day.

With respect to noise, we simulate an MA(∞) model with heteroscedastic variance:

$$U_i = \iota\eta_{t_i}(1 - 0.4B)^{-1}(1 + 0.2B)\varepsilon_i.$$

We vary the magnitude of the noise, ι, which takes values from: 10^{-4} (small noise), 5×10^{-4} (median noise), and 2.5×10^{-3} (large noise). These noise levels are common choices in the literature and also relevant for empirical data. η_t captures the heteroscedasticity of the noise, which follows

$$d\eta_t = 10 \times \left(\left(1 + 10^{-1}\cos(2\pi t/T)\right) - \eta_t\right)dt + 0.1dW_t, \tag{4.23}$$

where W_t is the same Brownian motion that drives X. We round the observed prices to the nearest cent: $\tilde{X}_{t_i} = \log\left([100 \times \exp(X_{t_i} + U_i)]\right) - \log 100$, where $[\cdot]$ means rounding to the nearest integer.[13]

We first compare the performance of the central limit theory using (i) (3.15), (ii) (3.16), and (iii) (3.13), given by Theorem 1. Recall that (i) works when noise is large and (ii) works when noise is small, whereas only (iii) works uniformly. Figure 1 compares the histograms of the standardized estimates of AIC*-QMLE that employ $\widehat{q}_{n,\mathrm{AIC}^\star} := \widehat{q}_{n,\mathrm{AIC}} \vee \log n_T$, using the corresponding asymptotic variances of these different scenarios. The histograms that correspond to (i) (resp. (ii)) on the top (resp. middle) panels do not match the standard normal density when noise is small (resp. large). By contrast, histograms on the bottom panel match the normal density uniformly well.

We then compare a variety of volatility estimators, including the usual realized volatility estimators using all returns and 5-minute returns; the MA(1)-based QMLE of Xiu (2010); the recent

[13] Although our theory does not allow for this type of rounding errors, we simulate this model to demonstrate that the effect of rounding appears negligible in a finite sample.

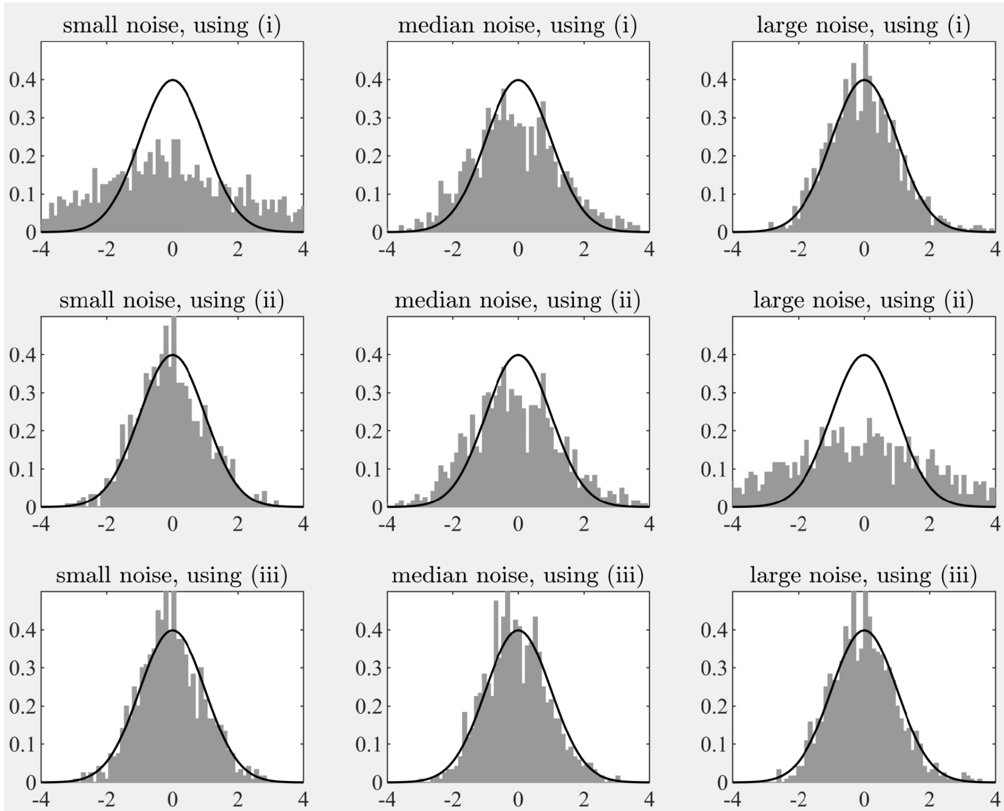

Figure 1: Histograms of Standardized Volatility Estimates

Note: This figure plots the histograms of standardized estimates of AIC*-QMLE using the central limit results given by (i) (3.15, top), (ii) (3.16, middle), and (iii) (3.13, bottom) of Theorem 1. Solid lines plot the density of the standard normal distribution. The noise magnitude parameter ι takes three values: 10^{-4} (small), 5×10^{-4} (median), and 2.5×10^{-3} (large).

pre-averaging estimator (PVG) proposed by Jacod, Li, and Zheng (2019); and the flat-top realized kernel estimator (FRK) proposed by Varneskov (2016). The last two estimators are consistent, even in the presence of colored noise. Other estimators that are robust to i.i.d. noise, such as two-scales realized variance and realized kernels, have very similar performance compared with the MA(1)-based QMLE, and hence are omitted.

To construct the PVG, Jacod, Li, and Zheng (2019) propose the following:

$$\widehat{\sigma}_n^{2,\mathrm{PVG}} = \left(T\sum_{i=0}^{h_n}(g_i^n)^2\right)^{-1}\left(\sum_{i=0}^{n_T-h_n}(\widetilde{Y}_i^n)^2 - n_T\sum_{j=-h_n''}^{h_n''}\widehat{\gamma}_{|j|}^{\mathrm{JLZ}}\sum_{i=0}^{h_n}\bar{g}_i^n\bar{g}_{i-j}^n\right),$$

where g is defined in Section 3.4, $g_i^n = g\big(\frac{i}{h_n}\big), \widetilde{Y}_i^n = \sum_{j=1}^{h_n-1} g_j^n Y_{n,i+j}, \bar{g}_i^n = g_{i+1}^n - g_i^n$, and $\widehat{\gamma}_j^{\mathrm{JLZ}}$ is the

Table 1: Simulation Results for Volatility Estimation

	Small Noise		Median Noise		Large Noise	
$\times 10^{-2}$	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
AIC*-QMLE	-0.02	1.09	-0.03	1.28	0.03	2.35
AIC-QMLE	0.06	0.75	0.11	1.40	0.07	2.34
PVG	-0.70	2.37	-0.24	2.34	2.08	3.50
FRK	-0.01	1.02	0.45	1.18	0.09	2.30
RV (5 min)	-0.02	1.68	1.62	2.58	42.24	43.44
MA(1) - QMLE	1.13	1.21	22.51	22.70	477.31	480.96
RV	3.11	3.12	47.99	48.02	1168.96	1169.42

Note: This table compares various volatility estimators in different simulation scenarios. "AIC*-QMLE" and "AIC-QMLE" are MA(q)-likelihood estimators using $\widehat{q}_{n,\mathrm{AIC}^\star} = \widehat{q}_{n,\mathrm{AIC}} \vee \log n_T$ and $\widehat{q}_{n,\mathrm{AIC}}$ as the selected orders, respectively. "PVG" refers to the pre-averaging estimator of Jacod, Li, and Zheng (2019). We report only PVG estimates with the smallest RMSE of all 27 versions, in which case $h_n = [n^{1/2}]$, $h'_n = 2$, and $h''_n = 3$ when noise is small; $h_n = [n^{1/2}]$, $h'_n = 2$, and $h''_n = 9$ in the case of a median noise level; and $h_n = [n^{1/2}]$, $h'_n = 2$, and $h''_n = 3$ when noise is large. "FRK" refers to the flat-top realized kernel estimator of Varneskov (2016). We only report FRK estimates with the smallest RMSE of all 27 versions, in which case $H_n = 25$, $\bar{h}_n = [0.5 \times H_n^{-0.6}]$, and $\bar{h}'_n = 0$ when noise is small; $H_n = 25$, $\bar{h}_n = [0.5 \times H_n^{-0.6}]$, and $\bar{h}'_n = 10$ in the case of a median noise level; and $H_n = 75$, $\bar{h}_n = [H_n^{-0.6}]$, and $\bar{h}'_n = 10$ when noise is large. "5-min RV" is the popular realized volatility estimator based on the 5-minute subsample. "MA(1)-QMLE" uses the MA(1) likelihood. "RV" is the realized volatility based on the full sample.

estimator of the jth autocovariance proposed by Jacod, Li, and Zheng (2017):

$$\widehat{\gamma}_j^{\mathrm{JLZ}} = \frac{1}{n_T} \sum_{i=0}^{n_T+1-j-4h'_n} \left(\widetilde{X}_{t_i} - \frac{1}{h'_n} \sum_{l=0}^{h'_n-1} \widetilde{X}_{t_{i+j+l+h'_n}} \right) \left(\widetilde{X}_{t_{i+j}} - \frac{1}{h'_n} \sum_{l=0}^{h'_n-1} \widetilde{X}_{t_{i+j+l+3h'_n}} \right).$$

The PVG estimator depends on three tuning parameters:

$$h'_n \sim \frac{1}{\Delta_n^{\eta}} \quad \text{with} \quad \frac{1}{2v+1} < \eta < \frac{1}{2}, \quad h_n \sim \frac{1}{\Delta_n^{1/2}}, \quad h''_n \sim \frac{1}{\Delta_n^{1/8}},$$

where v is the ρ-mixing exponent of ε. h'_n determines the local window size used to estimate realization of the noise, h_n is the usual local window size for averaging returns, and h''_n determines the maximum number of lags of nonzero noise autocovariances. Jacod, Li, and Zheng (2017) suggest $h'_n = 6$ in simulations with 1-second data. According to their criterion, when data are sampled at 5-second frequency, h'_n must be an even smaller integer in a finite sample, so we choose h'_n from $\{2, 4, 6\}$. Jacod, Li, and Zheng (2019) suggest $h''_n = 3$ and $h_n = [0.8 \times n^{1/2}]$ based on their simulation setting, so we choose h''_n from $\{3, 6, 9\}$ and h_n from $\{[0.6 \times n^{1/2}], [0.8 \times n^{1/2}], [1.0 \times n^{1/2}]\}$ for robustness. In total, we consider $3 \times 3 \times 3 = 27$ combinations of tuning parameters. To save space, we report only the best pre-averaging volatility estimate in terms of root-mean-square error (RMSE), despite the fact that this is not feasible beyond simulations and that the choice of tuning parameters matters quite a bit.

Regarding the FRK, Varneskov (2016) proposes the following:

$$\widehat{\sigma}_n^{2,\mathrm{FRK}} = \frac{1}{T}\gamma_0(Y_n^*) + \frac{1}{T}\sum_{h=1}^{\bar{h}_n}(\gamma_h(Y_n^*) + \gamma_{-h}(Y_n^*)) + \frac{1}{T}\sum_{h=\bar{h}_n+1}^{n_T-2\bar{h}_n'} k\left(\frac{h-\bar{h}_n}{H_n}\right)(\gamma_h(Y_n^*) + \gamma_{-h}(Y_n^*)),$$

where $k(\cdot)$ is the Parzen kernel, $\gamma_h(\cdot)$ is the hth-lag sample autocovariance function,

$$\gamma_h(Y_n^*) = \sum_{j=\bar{h}_n'+(0\vee h)}^{n_T-\bar{h}_n'+(0\wedge h)} Y_{n,j}^* Y_{n,j-h}^*,$$

and Y_n^* is a vector of returns after jittering,

$$Y_{n,j}^* = \begin{cases} \widetilde{X}_{t_{\bar{h}_n'}} - \dfrac{1}{\bar{h}_n'}\displaystyle\sum_{k=0}^{\bar{h}_n'-1}\widetilde{X}_{t_k}, & j = \bar{h}_n'; \\ -\widetilde{X}_{t_{n_T-\bar{h}_n'-1}} + \dfrac{1}{\bar{h}_n'}\displaystyle\sum_{k=n_T-\bar{h}_n'}^{n_T}\widetilde{X}_{t_k}, & j = n-\bar{h}_n'; \\ Y_{n,j}, & \text{Otherwise.} \end{cases}$$

There are also three tuning parameters:

$$H_n \sim \frac{1}{\Delta_n^{1/2}}, \quad \bar{h}_n \sim \Delta_n^{(\eta-1)/2} \text{ with } 0 < \eta < \frac{1+2\bar{r}}{2+2\bar{r}}, \quad \bar{h}_n' = \Delta_n^{-\varpi} \text{ with } \frac{1}{4} < \varpi < \frac{1}{2},$$

where $\bar{r}$ is the α-mixing exponent of ε. H_n is the usual bandwidth for kernel estimators, $\bar{h}_n$ controls the flatness of the kernel, and $\bar{h}_n'$ controls jittering. We choose H_n from $\{25, 50, 75\}$, $\bar{h}_n$ from $\{[0.5 \times H_n^{-0.6}], [1 \times H_n^{-0.6}], [1.5 \times H_n^{-0.6}]\}$, and $\bar{h}_n'$ from $\{0, 5, 10\}$. We report only the best estimate (in terms of the RMSE) of all 27 combinations.

Table 1 presents the comparison results.[14] When noise is small, all estimators work well, except that RV is a bit worse. This finding is not surprising, because all other estimators are somewhat robust to noise, and noise is present despite its small magnitude. When noise becomes larger and its dependence is thus more evident, the MA(1)-QMLE deteriorates, because it is only robust, essentially, to the white noise. The 5-min RV is better, but still has a substantial bias and a large RMSE, to the extent that using it is not reasonable in practice. AIC*- and AIC-QMLEs, FRK, and PVG are all well behaved, of which the PVG is the worst. That said, of all the estimators we compare, only AIC*-QMLE is uniformly valid. Our experiment also shows that FRK yields negative volatility estimates less often than the PVG estimator, whereas all other estimators guarantee positivity.

[14] We thank Yingying Li, Xinghua Zheng, and Rasmus Varneskov for sharing their codes with us.

5 Conclusion

We propose a simple volatility estimator based on the likelihood of an MA model, whose order is selected based on AIC. We establish uniformly valid inference on volatility over a large and flexible class of noise DGPs, featuring autocorrelations of an infinite order and an arbitrarily vanishing noise magnitude. The convergence rate of our estimator is greater than or equal to $n^{1/4}$, which depends on the noise magnitude and its dependence structure. Our estimator requires a single tuning parameter in order selection, and it always guarantees the positivity of volatility estimates. For these reasons, it delivers more desirable finite-sample performance than alternative nonparametric estimators.

References

Aït-Sahalia, Y., and J. Jacod (2014): *High Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351–416.

——— (2011): "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise," *Journal of Econometrics*, 160, 160–175.

Aït-Sahalia, Y., and D. Xiu (2016): "Increased Correlation Among Asset Classes: Are Volatility or Jumps to Blame?," *Journal of Econometrics*, 194, 205–219.

——— (2019): "A Hausman Test for the Presence of Market Microstructure Noise in High Frequency Data," *Journal of Econometrics*, 211, 176–205.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andrews, D. W. K., and X. Cheng (2012): "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80(5), 2153–2211.

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2020): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," *Journal of Econometrics*, 218(2), 496–531.

Bandi, F. M., and J. R. Russell (2008): "Microstructure Noise, Realized Volatility and Optimal Sampling," *Review of Economic Studies*, 75, 339–369.

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

——— (2011): "Subsampling Realised Kernels," *Journal of Econometrics*, 160(1), 204–219.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on Treatment Effects after Selection among High-Dimensional Controls," *Review of Economic Studies*, 81, 608–650.

Bibinger, M., N. Hautsch, P. Malec, and M. Reiss (2019): "Estimating the Spot Covariation of Asset Prices – Statistical Theory and Empirical Evidence," *Journal of Business & Economic Statistics*, 37(3), 419–435.

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

——— (1998): "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, 8, 291–323.

Da, R., and D. Xiu (2021): "Disentangling Autocorrelated Intraday Returns," Discussion paper, University of Chicago.

Fukasawa, M. (2010): "Realized volatility with stochastic sampling," *Stochastic Processes and Applications*, 120, 829–852.

Gardner, G., A. C. Harvey, and G. D. A. Phillips (1980): "Algorithm AS 154: An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(3), 311–322.

Hansen, P. R., J. Large, and A. Lunde (2008): "Moving average-based estimators of integrated variance," *Econometric Reviews*, 27, 79–111.

Hansen, P. R., and A. Lunde (2006): "Realized variance and market microstructure noise," *Journal of Business & Economic Statistics*, 24(2), 127–161.

Hasbrouck, J. (2007): *Empirical Market Microstructure*. Oxford University Press, New York, NY.

Hautsch, N., and M. Podolskij (2013): "Pre-Averaging Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence," *Journal of Business & Economic Statistics*, 31(2), 165–183.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J., Y. Li, and X. Zheng (2017): "Statistical Properties of Microstructure Noise," *Econometrica*, 85(4), 1133–1174.

——— (2019): "Estimating the Integrated Volatility with Tick Observations," *Journal of Econometrics*, 208, 80–100.

Jacod, J., M. Podolskij, and M. Vetter (2010): "Limit Theorems for Moving Averages of Discretized Processes Plus Noise," *Annals of Statistics*, 38, 1478–1545.

Jacod, J., and P. Protter (2011): *Discretization of Processes*. Springer-Verlag.

Kabaila, P. (1995): "The Effect of Model Selection on Confidence Regions and Prediction Regions," *Econometic Theory*, 11(3), 537–549.

Kalnina, I., and O. Linton (2008): "Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error," *Journal of Econometrics*, 147, 47–59.

Leeb, H., and B. M. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometic Theory*, 21(1), 21–59.

Leeb, H., and B. M. Pötscher (2008): "Can one estimate the unconditional distribution of post-model-selection estimators?," *Econometric Theory*, 24(2), 338–376.

Li, J., and D. Xiu (2016): "Generalized Method of Integrated Moments for High-Frequency Data," *Econometrica*, 84, 1613–1633.

Li, Y., P. A. Mykland, E. Renault, L. Zhang, and X. Zheng (2014): "Realized Volatility When Sampling Times are Possibly Endogenous," *Econometric Theory*, 30, 580–605.

Li, Z. M., R. J. Laeven, and M. H. Vellekoop (2020): "Dependent Microstructure Noise and Integrated Volatility Estimation from High-Frequency Data," *Journal of Econometrics*, 215(2), 536–558.

Mancini, C. (2001): "Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion," *Giornale dell'Istituto Italiano degli Attuari*, 64, 19–47.

Mikusheva, A. (2012): "One Dimensional Inference in Autoregressive Models with the Potential Presence of a Unit Root," *Econometrica*, 80(1), 173–212.

Pötscher, B. M. (1991): "Effects of Model Selection on Inference," *Econometic Theory*, 7(2), 163–185.

Reiss, M. (2011): "Asymptotic equivalence for inference on the volatility from noisy observations," *Annals of Statistics*, 39, 772–802.

Renault, E., and B. J. Werker (2011): "Causality Effects in Return Volatility Measures with Random Times," *Journal of Econometrics*, 160, 272–279.

Shibata, R. (1980): "Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process," *The Annals of Statistics*, 8(1), 147–164.

——— (1986): "Consistency of Model Selection and Parameter Estimation," *Journal of Applied Probability*, 23, 127–141.

Varneskov, R. T. (2016): "Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels," *Econometric Theory*, pp. 1–45.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

Whittle, P. (1951): *Hypothesis Testing in Time Series Analysis*. New York: Hafner Publishing Company.

Wilson, G. (1969): "Factorization of the Covariance Generating Function of a Pure Moving Average Process," *SIAM Journal on Numerica Analysis*, 6(1), 1–7.

Xiu, D. (2010): "Quasi-Maximum Likelihood Estimation of Volatility with High Frequency Data," *Journal of Econometrics*, 159, 235–250.

ZHANG, L. (2006): "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043.

ZHANG, L., P. A. MYKLAND, AND 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, 1394–1411.

ZYGMUND, A. (2002): *Trigonometric series*, vol. 1. Cambridge university press.

Appendix A Mathematical Proofs

As is typical in the literature, upon using a classical localization procedure (Section 4.4.1 of Jacod and Protter (2011)) we can strengthen the conditions introduced by Assumptions 1, 2, and 3 as follows:

Assumption A1. *There exist a constant $K > 0$ and non-negative functions Γ and $\widetilde{\Gamma}$, such that the processes X, μ, σ, ξ, ξ^{-1}, η, $\widetilde{\mu}$, $\widetilde{\sigma}$ are bounded by K, and that the functions δ and $\widetilde{\delta}$ satisfy $|\delta(u)| \leq \Gamma(u) \leq K$ and $\|\widetilde{\delta}(u)\| \leq \widetilde{\Gamma}(u) \leq K$. The ingredients of ξ and η (not written explicitly) also satisfy the same conditions as above.*

A.1 Notation

In this section we present the notation to be used throughout the proofs. Below we will introduce additional notation that applies only to the corresponding proofs unless otherwise indicated.

Part 1. We let X, Z, ξ, and η be defined on a sub-probability space $(\Omega_{(0)}, \mathcal{F}_\infty, (\mathcal{F}_t), \mathbb{P}_{(0)})$ of $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ as in footnote 5. We use $\xrightarrow{\mathcal{L}_s - \mathcal{F}_\infty}$ to denote stable convergence in law with respect to $\mathcal{F}_\infty$. For any $x = (x_1, x_2, \ldots, x_q) \in \mathbb{R}^q$, we set $x_j = 0$ for any $j \geq q+1$ and denote $\|x\|^2_{(q')} = \sum_{j=q'+1}^\infty x_j^2$, $\|x\|_{1,(q')} = \sum_{j=q'+1}^\infty |x_j|$, $\|x\|_1 = \|x\|_{1,(0)}$, and $1^\intercal \cdot x = \sum_{j=1}^\infty x_j$. For an integer i and a random variable x, we denote the i-th cumulant of x by $\mathrm{Cum}_i(x)$. Let $\mathcal{M}_d$ denote the set of all $d \times d$ matrices. For any m and h, let $O_m, \mathbb{D}^h_m, \mathbb{F}^h_m, \mathbb{I}_m \in \mathcal{M}_m$ be defined by $(\mathbb{I}_m)_{i,j} = \delta_{i,j}$, $(O_m)_{ij} = \sqrt{\frac{2}{m+1}} \sin \frac{ij\pi}{m+1}$, $(\mathbb{D}^h_m)_{ij} = \delta_{i,j}(2 - \delta_{h,0}) \cos \frac{hj\pi}{m+1}$, and $(\mathbb{F}^h_m)_{ij} = \mathbb{1}_{\{h=|i-j|\}} - \mathbb{1}_{\{h=i+j\}} - \mathbb{1}_{\{h=2m+2-(i+j)\}}$. We also introduce $\mathbb{I}_n, O_n, \mathbb{D}^h_n, \mathbb{F}^h_n \in \mathcal{M}_{n_T}$ (instead of $\mathcal{M}_n$) with similar entries. We let $n_d = \lfloor n^{7/8} \rfloor$, $J_d = \lfloor n_T / n_d \rfloor - 1$ and $n'_d = n_T - n_d J_d$. For any m, we define

$$D_m = \sum_{h=0}^\infty \gamma_h \mathbb{D}^h_m, \quad V_m = \sigma^2 \Delta_n \mathbb{I}_m + (2\mathbb{I}_m - \mathbb{D}^1_m) D_m, \quad \Omega_m = O_m V_m O_m, \quad \Omega_{D,n} = (\mathbb{I}_{J_d} \otimes \Omega_{n_d}) \oplus \Omega_{n'_d}.$$

Here the dependence of $(D_m, V_m, \Omega_m, \Omega_{D,n})$ on $(\sigma^2, \gamma, \Delta_n)$ is omitted.

Part 2. We use $\Delta^n_i A$ to denote $A_{t_i} - A_{t_{i-1}}$ when A is a continuous-time stochastic process and to denote $A_i - A_{i-1}$ when A is a discrete-time stochastic process. Further, for $j \geq 1$, we introduce $t(j)_i = t_{(j-1)n_d+i}$. When A is a continuous-time process, we let $A_C(j) = A_{t_{(j-1)n_d}}$, $A_{C,t} \coloneqq \sum_{j=1}^\infty A_C(j) \mathbb{1}_{\{t_{(j-1)n_d} \leq t < t_{jn_d}\}}$, and $A(j)_i = A_{t_{(j-1)n_d+i}}$. When A is a discrete-time process, we let $A(j)_i = A_{(j-1)n_d+i}$. In both cases, $A(j)$ can be regarded as a discrete-time process. We further let

$\varepsilon_C(j)_i := \varepsilon_{(j-1)n_d+i}$ for $i \geq 1$ and $\varepsilon_C(j)_i := \tilde{\varepsilon}(j)_i$ for $i < 1$, where $\{\tilde{\varepsilon}(j)_i : i \leq 0, j \geq 1\}$ is a set of standard normal random variables that are independent across (i,j) and with everything else. We define for all $i \geq 1$,

$$Y^C_{n,i} = \sigma_{C,t_{i-1}} \xi^{1/2}_{C,t_{i-1}} \xi^{-1/2}_{t_{i-1}} \Delta^n_i W + \eta_{C,t_{i-1}} \iota^{(n)} (1-\mathrm{B}) \theta^{(n)}(\mathrm{B}) \varepsilon_C(\lfloor i/n_d \rfloor + 1)_{i - \lfloor i/n_d \rfloor n_d}.$$

We define $n'_T = (J_d+2)n_d$ and define $Y^C_n \in \mathbb{R}^{n'_T}$ as $Y^C_n := (Y^C_{n,1}, \ldots, Y^C_{n,n'_T})^\intercal$. We also write

$$U^C(j)_i = \eta_C(j) \iota^{(n)} \theta^{(n)}(\mathrm{B}) \varepsilon_C(j)_i, \quad \Delta^n_i X^C(j) = \sigma_C(j) \xi^{1/2}_C(j) \xi^{-1/2}(j)_{i-1} \Delta^n_i W(j).$$

Next, for $r \geq 1$ we consider the successive jump times $\{T(r,m) : m \geq 1\}$ of the Poisson process $\underline{\mu}((0,T] \times \{z : 1/r < \Gamma(z) \leq 1/(r-1)\})$. Let $\{T_m\}_{m \geq 1}$ denote any reordering of $\{T(r,m) : r, m \geq 1, T(r,m) < T\}$ and $T_0 = 0$. We additionally set $P_r = \{r : \exists r' \geq 1, r' \leq r, T_m = T(r', m')\}$ and define:

$$\begin{cases} X^{J,r} = (\delta \mathbb{1}_{\{z : \Gamma(z) > 1/r\}}) \star \underline{\mu} \\ X^{J,r-} = (\delta \mathbb{1}_{\{z : \Gamma(z) \leq 1/r\}}) \star (\underline{\mu} - \underline{\nu}) \end{cases}, \quad \begin{cases} \mu^r_t = \mu_t - \int_{\{z : \Gamma(z) > 1/r, |\delta(t,z)| \leq 1\}} \delta(t,z) \lambda(dz) \\ X^{B,r}_t = X_0 + \int_0^t \mu^r_s ds + \int_0^t \sigma_s dW_s \end{cases},$$

$$\begin{cases} Y^{r-}_{n,j} = Y_{n,j} - \Delta^n_j X^{J,r} \\ Y^{r-}_n = (Y^{r-}_{n,1}, Y^{r-}_{n,2}, \ldots, Y^{r-}_{n,n})^\intercal \end{cases}, \quad \begin{cases} Y^{B,r}_{n,j} = \Delta^n_j X^{B,r} + \Delta^n_j U \\ Y^{B,r}_n = (Y^{B,r}_{n,1}, Y^{B,r}_{n,2}, \ldots, Y^{B,r}_{n,n})^\intercal \end{cases}.$$

Part 3. For $(m = n_d, 1 \leq j \leq J_d)$ or $(m = n'_d, j = J_d + 1)$, we define $\Omega^U_m(j) \in \mathcal{M}_m$ by

$$\begin{aligned} \Omega^U_m(j)_{ik} \quad = \quad & (\iota^{(n)})^2 (\eta(j)_i \eta(j)_k \kappa^{(n)}_{|i-k|} + \eta(j)_{i-1} \eta(j)_{k-1} \kappa^{(n)}_{|i-k|} \\ & - \eta(j)_i \eta(j)_{k-1} \kappa^{(n)}_{|i-k+1|} - \eta(j)_{i-1} \eta(j)_k \kappa^{(n)}_{|i-k-1|}). \end{aligned}$$

Using $\Omega^U_m(j)$, we define $\Omega^U_n = \left(\bigoplus_{j=1}^{J_d} \Omega^U_{n_d}(j) \right) \oplus \Omega^U_{n'_d}(J_d+1)$. For any n, let $\Omega^B_n, \Omega^J_n, \Omega^Y_n, \Omega^{Y,B}_n \in \mathcal{M}_{n_T}$ be defined by

$$(\Omega^B_n)_{ij} = \delta_{i,j} \int_{t_{i-1}}^{t_i} \sigma^2_s ds, \quad (\Omega^J_n)_{ij} = \delta_{i,j} \sum_{t_{i-1} < s \leq t_i} (\Delta X_s)^2, \quad \Omega^{Y,B}_n = \Omega^U_n + \Omega^B_n, \quad \Omega^Y_n = \Omega^{Y,B}_n + \Omega^J_n.$$

Then, for $j \geq 1$ we introduce an ∞-dimensional vector $\gamma_C(j) := (\gamma_C(j)_k)_{k \geq 0}$ with $\gamma_C(j)_k = (\iota^{(n)})^2 \eta^2_C(j) \kappa^{(n)}_k$, and a scalar $\zeta_C(j) := \sum_{k=-\infty}^{\infty} \gamma_C(j)_{|k|}$. Finally, we introduce

$$\Omega^C_n(j) = \sigma^2_C(j) \xi_C(j) \frac{T}{n} \mathbb{I}_{n_d}, \quad \Omega^{U,C}_n(j) = O_{n_d} (2\mathbb{I}_{n_d} - \mathbb{D}^1_{n_d}) D_{n_d}(\gamma_C(j)) O_{n_d},$$

based on which we define $\Omega^{Y,C}_n(j) = \Omega^C_n(j) + \Omega^{U,C}_n(j)$ and $\Omega^{Y,C}_n = \bigoplus_{j=1}^{J_d+2} \Omega^{Y,C}_n(j)$. For any n, let

$\Omega_n^{J,r}, \Omega_n^{J,r-}, \Omega_n^{Y,r-} \in \mathcal{M}_{n_T}$ be defined by

$$(\Omega_n^{J,r})_{ij} = \delta_{i,j} \sum_{m \in P_r : t_{i-1} \le T_m < t_i} (\Delta X_{T_m})^2, \quad \Omega_n^{J,r-} = \Omega_n^J - \Omega_n^{J,r}, \quad \Omega_n^{Y,r-} = \Omega_n^{Y,B} + \Omega_n^{J,r-}.$$

Part 4. We introduce shorthand notation $\mathcal{L}(A) = -\frac{1}{2}\log\det\Omega_{D,n} - \frac{1}{2}\mathrm{tr}(\Omega_{D,n}^{-1}A)$ and let

$$L_{A,n} = -\frac{1}{2}\log\det\Omega_n - \frac{1}{2}\mathrm{tr}(\Omega_n^{-1}Y_nY_n^\intercal), \quad L_{D,n} = \mathcal{L}(Y_nY_n^\intercal), \quad \bar{L}_n = \mathcal{L}(\Omega_n^Y),$$

$$L_n^{r-} = \mathcal{L}(Y_n^{r-}(Y_n^{r-})^\intercal), \quad \bar{L}_n^{r-} = \mathcal{L}(\Omega_n^{Y,r-}), \quad L_n^{B,r} = \mathcal{L}(Y_n^{B,r}(Y_n^{B,r})^\intercal_n), \quad \bar{L}_n^B = \mathcal{L}(\Omega_{D,n}, \Omega_n^{Y,B}),$$

where we omit the argument (σ^2, γ) of $(\Omega_n, \Omega_{D,n})$ and $(L_{A,n}, L_{D,n}, \bar{L}_n, L_n^{r-}, \bar{L}_n^{r-}, L_n^{B,r}, \bar{L}_n^B)$. Finally, we define

$$\bar{L}_n^\star(\sigma^2,\gamma) = -\frac{n_T}{4\pi}\int_{-\pi}^{\pi}\Big(\log f(\lambda;\sigma^2,\gamma,\Delta_n) + \frac{f(\lambda; C_T, \gamma^{(n)}, \Delta_n)}{f(\lambda;\sigma^2,\gamma,\Delta_n)}\Big)d\lambda,$$

$$\chi^2(\sigma^2,\gamma,\Delta_n) = \exp\Big(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda;\sigma^2,\gamma,\Delta_n)d\lambda\Big), \quad (\sigma^{(n)}(q)^2, \gamma^{(n)}(q)) = \arg\min_{(\sigma^2,\gamma)\in\Pi_n(q)} \bar{L}_n^\star(\sigma^2,\gamma).$$

Part 5. With any given n, (σ,γ), q, and $i \in \{b, s\}$, we define

$$\begin{aligned}
R_n(\sigma^2,\gamma,b) &= |\sigma^2 - C_T| + (\iota^{(n)})^{-2}\sup_\lambda |f(\lambda;\gamma) - f(\lambda;\gamma^{(n)})|,\\
R_n(\sigma^2,\gamma,s) &= \Delta_n^{-1}\sup_\lambda |f(\lambda;\sigma^2,\gamma,\Delta_n) - f(\lambda;C_T,\gamma^{(n)},\Delta_n)|,\\
\widehat{\mathcal{R}}_n(q,i) &= R_n(\widehat{\sigma}_n^2(q), \widehat{\gamma}_n(q), i), \quad \mathcal{R}^{(n)}(q,i) = R_n(\sigma^{(n)}(q)^2, \gamma^{(n)}(q), i).
\end{aligned}$$

Part 6. We introduce a framework to conduct reparameterization. To avoid ambiguity, throughout the proof we use $\Pi_n^{(\sigma^2,\gamma)}(q)$ to refer to the parameter space $\Pi_n(q)$ defined in (3.9). We start by introducing a bijection from $\Pi_n^{(\sigma^2,\gamma)}(q)$ to $\mathbb{R}^{q+2}$ denoted by $\beta_n(\sigma^2,\gamma)$. The inverse functions are denoted by $\sigma_n^2(\beta)$ and $\gamma_n(\beta)$. Choices of the functional form of β_n will only be specified when necessary and will typically vary across different scenarios. We set $\partial\sigma_n^2 := \partial\sigma_n^2(\beta)/\partial\beta$. Let $\widehat{\beta}_n(q), \beta^{(n)}(q) \in \mathbb{R}^{q+2}$ be defined as

$$\widehat{\beta}_n(q) = \beta_n(\widehat{\sigma}_n^2(q), \widehat{\gamma}_n(q)), \quad \beta^{(n)}(q) = \beta_n(\sigma^{(n)}(q)^2, \gamma^{(n)}(q)), \qquad \bar{\beta}^{(n)} = \beta_n(C_T, \gamma^{(n)}).$$

Let $\Pi_n^\beta(q) = \{\beta = (\beta_0, \beta_1, \ldots, \beta_{q+1})^\intercal \in \mathbb{R}^{q+2} : \beta = \beta_n(\sigma^2,\gamma) \text{ with } (\sigma^2,\gamma) \in \Pi_n^{(\sigma^2,\gamma)}(q)\}$. For any $\beta \in \Pi_n^\beta(q)$, and any $S_n \in \{f(\lambda;\ \cdot\ ,\ \cdot\ ,\Delta_n), L_n, L_{A,n}, L_{D,n}, \bar{L}_n, L_n^{r-}, \bar{L}_n^{r-}, L_n^{B,r}, \bar{L}_n^B, \bar{L}_n^\star, \Sigma_n, \Omega_n, \Omega_{D,n}, V_n\}$, we let $S_n(\beta) = S_n(\sigma^2,\gamma)$, with (σ^2,γ) satisfying $\beta = \beta_n(\sigma^2,\gamma)$. Furthermore, for $\beta \in \Pi_n^\beta(q)$ and $s \in \{A, D\}$, we define $\Xi_n(\beta), \Xi_{s,n}(\beta) \in \mathbb{R}^{q+2}$ and $\partial\Xi_{A,n}(\beta)_{ij}, \partial\bar{\Xi}_n^\star(\beta) \in \mathcal{M}_{q+2}$ such that

$$\big(\Xi_n(\beta)_j, \Xi_{s,n}(\beta)_j\big) = -\frac{1}{n}\frac{\partial}{\partial\beta_j}\big(L_n(\beta), L_{s,n}(\beta)\big), \ \big(\partial\Xi_{A,n}(\beta)_{ij}, \partial\bar{\Xi}_n^\star(\beta)_{ij}\big) = -\frac{1}{n}\frac{\partial^2}{\partial\beta_i\partial\beta_j}\big(\bar{L}_n^\star(\beta), L_{A,n}^\star(\beta)\big). \tag{A.1}$$

We let $\bar\eta := -\partial\bar\Xi_n^\star(\bar\beta^{(n)})^{-1}\partial\sigma_n^2$ and define $\bar\Xi_n(\beta), \Xi_n^{r-}(\beta), \bar\Xi_n^{r-}(\beta), \Xi_n^{B,r}(\beta), \bar\Xi_n^B(\beta) \in \mathbb{R}^{q+2}$ as

$$\big(\bar\Xi_n(\beta)_j, \Xi_n^{r-}(\beta)_j, \bar\Xi_n^{r-}(\beta)_j, \Xi_n^{B,r}(\beta)_j, \bar\Xi_n^B(\beta)_j\big) = -\frac{1}{n}\frac{\partial}{\beta_j}\big(\bar L_n(\beta), L_n^{r,-}(\beta), \bar L_n^{r,-}(\beta), L_n^{B,r}(\beta), \bar L_n^B(\beta)\big).$$

A.2 Proof of Lemma A1

Lemma A1. *Suppose Assumptions 1 - 4 hold and q_n is deterministic. We let $C(4,\xi)_T = \frac{1}{T}\int_0^T \xi_s\sigma_s^4 ds$ and $B(\xi,1)_T = 2\big(C_T\int_0^T \eta_s^2\xi_s^{-1}ds\big)^{-1}\int_0^T \eta_s^2\sigma_s^2 ds + \big(\int_0^T \eta_s^2\xi_s^{-1}ds\big)^{-2}T\int_0^T \eta_s^4\xi_s^{-1}ds$. We set $\beta_n(\sigma^2,\gamma) = (\sigma^2,\gamma)$. Let $\mathcal{U}^B$ be a random variable defined on an extension $(\overline\Omega, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{P}})$ of $(\Omega_{(0)}, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_{(0)})$, which, conditionally on $\mathcal{F}_\infty$, is standard normal. The central limit theorem*

$$\frac{n^{1/2}\bar\eta^\intercal\Xi_n^{B,r}(\beta^{(n)}(q_n)) - n^{1/2}\bar\eta^\intercal\bar\Xi_n^B(\beta^{(n)}(q_n))}{\sqrt{4q_nC(4,\xi)_T + \Delta_n^{-1/2}\zeta^{(n)}(5C(4,\xi)_TC_T^{-1/2} + C_T^{3/2}B(\xi,1)_T)}} \xrightarrow{\mathcal{L}_s-\mathcal{F}_\infty} \mathcal{U}^B \tag{A.2}$$

holds if either of the two following conditions is true:

(i) We have $n^{1/2}\iota^{(n)} \to \infty$, $\mathcal{R}^{(n)}(q_n,b) = o_{\mathrm{P}}(1)$, and $q_n \leq Kn^{1/3}$.

(ii) We have $n^{1/2}\iota^{(n)} \leq K$, $\mathcal{R}^{(n)}(q_n,s) = o_{\mathrm{P}}(1)$, $q_n \leq Kn^{1/3}$, and $q_n \to \infty$.

Proof. Step 1. In this step we make some preparations. We write $\Xi_n^{B,r}(\beta^{(n)}(q_n)) - \bar\Xi_n^B(\beta^{(n)}(q_n))$ explicitly as a partial sum of a triangular array. We drop the argument q_n of $\beta^{(n)}(q_n)$, $\sigma^{(n)}(q_n)^2$, and $\gamma^{(n)}(q_n)$ whenever there is no ambiguity. By definition we are able to write

$$\Xi_n^{B,r}(\beta^{(n)})^\intercal - \bar\Xi_n^B(\beta^{(n)})^\intercal = \frac{\partial}{\partial\beta}\mathrm{tr}\Big(-\frac{1}{2n}(Y_n^{B,r}(Y_n^{B,r})^\intercal - \Omega_n^{Y,B})\Omega_{D,n}(\beta^{(n)})^{-1}\Big). \tag{A.3}$$

In view of (A.3), we introduce for all $1 \leq j \leq J_d+1$,

$$\begin{cases} \mathcal{U}_n(j) = \sum_{i=1}^{q_n+2}\bar\eta_i\frac{\partial}{\partial\beta_i}\mathrm{tr}(\Omega_{n(j)}(\beta^{(n)})^{-1}Y_n^{B,r}(j)Y_n^{B,r}(j)^\intercal) \\ \overline{\mathcal{U}}_n(j) = \sum_{i=1}^{q_n+2}\bar\eta_i\frac{\partial}{\partial\beta_i}\mathrm{tr}(\Omega_{n(j)}(\beta^{(n)})^{-1}\Omega_n^{Y,B}(j)) \end{cases}, \quad n(j) = n_d + \delta_{j,J_d+1}(n_d' - n_d), \tag{A.4}$$

where $Y_n^{B,r}(j) = (Y_n^{B,r}(j)_1, \dots, Y_n^{B,r}(j)_{n(j)})^\intercal$ and $\Omega_n^{Y,B}(j) = \big((\Omega_n^{Y,B})_{(j-1)n_d,+k,(j-1)n_d+l}\big)_{1\leq k,l\leq n(j)}$. We see that (A.4) allows us to rewrite (A.3) as

$$\bar\eta\Xi_n^{B,r}(\beta^{(n)}) - \bar\eta\bar\Xi_n^B(\beta^{(n)}) = -\frac{1}{2n}\sum_{j=1}^{J_d+1}(\mathcal{U}_n(j) - \overline{\mathcal{U}}_n(j)). \tag{A.5}$$

Defining $Y_n^C(j) = (Y_n^C(j)_1, \dots, Y_n^C(j)_{n_d})^\intercal$, we further introduce for all $1 \leq j \leq J_d+2$,

$$\begin{cases} \mathcal{V}_n(j) = \sum_{i=1}^{q_n+2}\bar\eta_i\frac{\partial}{\partial\beta_i}\mathrm{tr}(\Omega_{n_d}(\beta^{(n)})^{-1}Y_n^C(j)Y_n^C(j)^\intercal) \\ \overline{\mathcal{V}}_n(j) = \sum_{i=1}^{q_n+2}\bar\eta_i\frac{\partial}{\partial\beta_i}\mathrm{tr}(\Omega_{n_d}(\beta^{(n)})^{-1}\Omega_n^{Y,C}(j)) \end{cases}. \tag{A.6}$$

Finally, we let $\bar{J}_d = n^{1/2}\iota^{(n)}+q_n$ and denote $\widetilde{n}_d(m) := (2^m-1)\lceil n_d \bar{J}_d^{-1}\rceil$ and $\bar{n}_d(m) = \widetilde{n}_d(m+1)-\widetilde{n}_d(m)$ for $m \geq 0$. To simplify exposition, we assume that there exists such an integer $\widetilde{J}_d$ that $\widetilde{n}_d(\widetilde{J}_d) = n_d$ (this is without loss of generality, as we have $\bar{J}_d \to \infty$ under either condition (i) or condition (ii) in the statement of the lemma). And we define, for two integers m and p,

$$\widetilde{O}(m,p)_{k,l} = \sum_{i=1}^{\bar{n}_d(m)} (O_{n_d})_{\widetilde{n}_d(m)+i,k}(O_{n_d})_{\widetilde{n}_d(m)+i,l}\Big(e^{-\mathrm{i}\pi\frac{(i-1/2)p}{\bar{n}_d(m)}} + e^{\mathrm{i}\pi\frac{(i-1/2)p}{\bar{n}_d(m)}}\Big), \tag{A.7}$$

$$\mathcal{R}_{f1}(j)_{k,l} = Y_n^C(j)_k Y_n^C(j)_l - \Omega_n^{Y,C}(j)_{k,l}, \quad \overline{\mathcal{R}}_f(j,m,p) = \sum_{k=1}^{n_d}\sum_{l=1}^{n_d} \widetilde{O}(m,p)_{k,l}\mathcal{R}_{f1}(j)_{k,l}. \tag{A.8}$$

Step 2. This step is devoted to proving asymptotic properties of various moments related to the quantity $\overline{\mathcal{R}}_f(j)$. Here $\overline{\mathcal{R}}_f(j)$ is a d-dimensional vector indexed by (m,p) (ordering does not matter), whose components are defined in (A.8). We let $\mathcal{F}^\varepsilon(j) = \sigma(\varepsilon_C(k)_i : i \leq n_d, k \leq j-1)$ be the σ-field generated by the sequence of all $\varepsilon_C(k)$ with $k \leq j-1$, and $\mathcal{F}^\chi(j) = \sigma(\chi_i : i \leq (j-1)n_d)$ be the σ-field generated by the sequence of all χ_i with $i \leq (j-1)n_d$, and denote $\mathcal{F}(j) = \mathcal{F}_{\tau(n,j-1)} \otimes \mathcal{F}^\chi(j) \otimes \mathcal{F}^\varepsilon(j)$. According to the definitions of Y_n^C and $\Omega_n^{Y,C}$, we have that for all $1 \leq j \leq J_d$, $\mathbb{E}(\mathcal{R}_{f1}(j)_{k,l}|\mathcal{F}(j)) = (\iota^{(n)})^2\eta_C^2(j)\sum_{h=0}^\infty(\kappa_h^{(n)} - \kappa_{h+1}^{(n)})(\mathbb{K}_{n_d}^h + \widetilde{\mathbb{K}}_{n_d}^h)_{k,l}$, where $\mathbb{K}_{n_d}^h$ and $\widetilde{\mathbb{K}}_{n_d}^h$ are $n_d \times n_d$ square matrices given by $(\mathbb{K}_{n_d}^h)_{k,l} = \mathbb{1}_{\{h=k+l\}} - \mathbb{1}_{\{h+1=k+l\}}$ and $(\widetilde{\mathbb{K}}_{n_d}^h)_{k,l} = (\mathbb{K}_{n_d}^h)_{n_d+1-k,n_d+1-l}$. According to Assumption 4, we have $\sum_{h=0}^\infty h^2|\kappa_h^{(n)}| \leq K$. Furthermore, the definition of O matrix indicates that for all $0 \leq m \leq J_d$, all $\widetilde{n}_d(m) \leq i \leq \widetilde{n}_d(m+1)$, and all $1 \leq k \leq n_d$,

$$|(O_{n_d})_{\widetilde{n}_d+i,k}| \lesssim (n_d^{-3/2}\bar{n}_d k) \wedge n_d^{-1/2}. \tag{A.9}$$

Here and in the rest of the proof we omit mentioning the argument m of $\widetilde{n}_d$ and $\bar{n}_d$ unless necessary. We also notice $|(O_{n_d})_{\widetilde{n}_d+i,n_d+1-k}| = |(O_{n_d})_{\widetilde{n}_d+i,k}|$, and therefore use Assumption A1 to conclude that for all $0 \leq m, m' \leq \widetilde{J}_d - 1$,

$$\sup_j \Big|\sum_{k=1}^{n_d}\sum_{l=1}^{n_d}(O_{n_d})_{\widetilde{n}_d(m)+i,k}(O_{n_d})_{\widetilde{n}_d(m')+i',l}\mathbb{E}(\mathcal{R}_{f1}(j)_{k,l}|\mathcal{F}(j))\Big| \leq K(\iota^{(n)})^2 n_d^{-3}\bar{n}_d(m)\bar{n}_d(m'). \tag{A.10}$$

We immediately obtain, using the definition of $\widetilde{O}(m,p)$ given in Step 1, that for all $0 \leq m \leq \widetilde{J}_d - 1$,

$$\sup_{j,p}|\mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)|\mathcal{F}(j))| \leq K(\iota^{(n)})^2 n_d^{-3}\bar{n}_d^3. \tag{A.11}$$

Now we study $\mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)\overline{\mathcal{R}}_f(j,m',p')|\mathcal{F}(j))$. We first introduce some shorthand notation:

$$\begin{aligned}
\mathcal{R}_{f2}(j)_{k,l} &= \mathbb{E}(\mathcal{R}_{f1}(j)_{k,l}|\mathcal{F}(j)), \\
\mathcal{R}_{f3}(j)_{k,l',l,k'} &= \mathcal{R}_{f2}(j)_{k,l'}\mathcal{R}_{f2}(j)_{l,k'} + \mathcal{R}_{f2}(j)_{k,l'}\Omega_n^{Y,C}(j)_{l,k'} + \Omega_n^{Y,C}(j)_{k,l'}\mathcal{R}_{f2}(j)_{l,k'},
\end{aligned}$$

$$\begin{aligned}
\mathcal{R}_{f4}(j)_{k,l,k',l'} &= (\iota^{(n)}\eta_C(j))^4 \mathrm{Cum}_4(\varepsilon) \sum_{p=-\infty}^{k\wedge l\wedge k'\wedge l'} \theta^{(n)}_{k-p}\theta^{(n)}_{l-p}\theta^{(n)}_{k'-p}\theta^{(n)}_{l'-p}, \\
\mathcal{R}_{f5}(j)_{k,l,k',l'} &= \sum_{s=k-1}^{k}\sum_{r=l-1}^{l}\sum_{s'=k'-1}^{k'}\sum_{r=l'-1}^{l'} (-1)^{k+k'+l+l'-(s+s'+r+r')}\mathcal{R}_{f4}(j)_{s,r,s',r'}, \\
\mathcal{R}_{f6}(j)_{k,l,k',l'} &= \mathcal{R}_{f2}(j)_{k,l}\mathcal{R}_{f2}(j)_{k',l'} + \mathcal{R}_{f3}(j)_{k,l',l,k'} + \mathcal{R}_{f3}(j)_{k,k',l,l'} + \mathcal{R}_{f5}(j)_{k,l,k',l'}.
\end{aligned}$$

It is easy to verify, again using the definitions of Y_n^C and $\Omega_n^{Y,C}$, that

$$\mathbb{E}(\mathcal{R}_{f1}(j)_{k,l}\mathcal{R}_{f1}(j)_{k',l'}|\mathcal{F}(j)) = \Omega_n^{Y,C}(j)_{k,l'}\Omega_n^{Y,C}(j)_{l,k'} + \Omega_n^{Y,C}(j)_{k,k'}\Omega_n^{Y,C}(j)_{l,l'} + \mathcal{R}_{f6}(j)_{k,l,k',l'}. \quad \text{(A.12)}$$

We observe from the definition of $\Omega_n^{Y,C}$ that

$$\sum_{k=1}^{n_d}\sum_{l=1}^{n_d}(O_{n_d})_{\widetilde{n}_d(m)+i,k}(O_{n_d})_{\widetilde{n}_d(m')+i',l}\Omega_n^{Y,C}(j)_{k,l} = \delta_{m,m'}\delta_{i,i'}f\Big(\cos\frac{(\widetilde{n}_d(m)+i)\pi}{n_d+1};j\Big), \quad \text{(A.13)}$$

where we use the shorthand notation $f(\lambda;j) \coloneqq f\left(\lambda;\sigma_C^2(j),\gamma_C(j),\xi_C(j)\frac{T}{n}\right)$ and Lemma A1 of Da and Xiu (2021). The function $f(\lambda;\sigma^2,\gamma,\Delta_n)$ is introduced in Section 3.1. We can further obtain, using Riemann approximation, that for some $\alpha_n \to 0$ and all $0 \le m \le \widetilde{J}_d - 1$,

$$\begin{aligned}
\sup_{j,p,p'}\Bigg| &\textstyle\sum_{i=1}^{\bar{n}_d}\left(e^{-\mathrm{i}\pi\frac{ip}{\bar{n}_d}} + e^{\mathrm{i}\pi\frac{(i-1)p}{\bar{n}_d}}\right)\left(e^{-\mathrm{i}\pi\frac{ip'}{\bar{n}_d}} + e^{\mathrm{i}\pi\frac{(i-1)p'}{\bar{n}_d}}\right)f\Big(\cos\frac{(\widetilde{n}_d+i)\pi}{n_d+1};j\Big)^2 \\
&\textstyle -\frac{4\bar{n}_d}{\pi}\int_0^\pi \cos(p\lambda)\cos(p'\lambda)\Big(\sigma_C^2(j)\xi_C(j)\frac{T}{n} + \frac{\zeta_C^2(j)}{\bar{J}_d^2}\left((2^m-1)\pi+2^m\lambda\right)^2\Big)^2 d\lambda\Bigg| \\
&\le \alpha_n\left(\bar{n}_d n^{-2} + \bar{n}_d 2^{4m}\bar{J}_d^{-4}(\iota^{(n)})^4\right) + K\bar{n}_d 2^{6m}\bar{J}_d^{-6}(\iota^{(n)})^4. \qquad \text{(A.14)}
\end{aligned}$$

Here we use Assumptions A1 and 4 and the definition of $\widetilde{n}_d$. On the other hand, (B.46) yields

$$\sup_j|\mathcal{R}_{f4}(j)_{k,l,k',l'}| \le \frac{K(\iota^{(n)})^4}{(|k-k'|^2+1)(|l-l'|^2+1)} \wedge \frac{K(\iota^{(n)})^4}{(|k-l|^2+1)(|k'-l'|^2+1)}. \quad \text{(A.15)}$$

In addition, we notice $|(O_{n_d})_{\widetilde{n}_d+i,k+1} - (O_{n_d})_{\widetilde{n}_d+i,k}| \le Kn_d^{-3/2}\bar{n}_d$ for all $0 \le m \le \widetilde{J}_d - 1$, again directly from the definition of O matrix. Combining this inequality with (A.9) and (A.15), we obtain, using the definition of $\widetilde{O}(m,p)$, that for all $0 \le m,m' \le \widetilde{J}_d - 1$,

$$\sup_{j,p,p'}\left|\sum_{k=1}^{n_d}\sum_{l=1}^{n_d}\sum_{k'=1}^{n_d}\sum_{l'=1}^{n_d}\widetilde{O}(m,p)_{k,l}\widetilde{O}(m',p')_{k',l'}\mathcal{R}_{f5}(j)_{k,l,k',l'}\right| \le K(\iota^{(n)})^4\bar{n}_d^3(m)\bar{n}_d^3(m')n_d^{-5}\log n. \quad \text{(A.16)}$$

Substituting (A.10), (A.13), (A.14), and (A.16) back into (A.12) and the definition of $\overline{\mathcal{R}}_f(j,m,p)$ in

(A.8), we obtain that for some $\alpha_n \to 0$ and all $0 \le m, m' \le \widetilde{J}_d - 1$,

$$\begin{aligned}
\sup_{j,p,p'} \Bigg| & \mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)\overline{\mathcal{R}}_f(j,m',p')|\mathcal{F}(j)) \\
& -\delta_{m,m'}\frac{8\bar{n}_d(m)}{\pi}\int_0^{\pi}\cos(p\lambda)\cos(p'\lambda)\Big(\sigma_C^2(j)\xi_C(j)\frac{T}{n}+\frac{\zeta_C^2(j)}{\bar{J}_d^2}\big((2^m-1)\pi+2^m\lambda\big)^2\Big)^2 d\lambda\Bigg| \\
& \le \alpha_n\Big(\delta_{m,m'}\bar{n}_d(m)\big(n^{-2}+2^{4m}\bar{J}_d^{-4}(\iota^{(n)})^4\big)\Big) \\
& \quad + K\Big((\iota^{(n)})^4\bar{n}_d^3(m)\bar{n}_d^3(m')n_d^{-5}\log n+\delta_{m,m'}\bar{n}_d(m)2^{6m}\bar{J}_d^{-6}(\iota^{(n)})^4\Big). \qquad (A.17)
\end{aligned}$$

Next, we study $\mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)^4|\mathcal{F}(j))$. We introduce some shorthand notation:

$$\begin{aligned}
\mathcal{R}_{f7}(j)_{k,k',l,l',r,r'} &= \operatorname{cum}_6(\varepsilon)(\iota^{(n)}\eta_C(j))^6 \sum_{p=-\infty}^{k\wedge k'\wedge l\wedge l'\wedge r\wedge r'} \theta_{k-p}^{(n)}\theta_{k'-p}^{(n)}\theta_{l-p}^{(n)}\theta_{l'-p}^{(n)}\theta_{r-p}^{(n)}\theta_{r'-p}^{(n)}, \\
\mathcal{R}_{f8}(j)_{k,k',l,l',r,r',s,s'} &= \operatorname{cum}_8(\varepsilon)(\iota^{(n)}\eta_C(j))^8 \sum_{p=-\infty}^{k\wedge l\wedge r\wedge s\wedge k'\wedge l'\wedge r'\wedge s'} \theta_{k-p}^{(n)}\theta_{l-p}^{(n)}\theta_{r-p}^{(n)}\theta_{s-p}^{(n)}\theta_{k'-p}^{(n)}\theta_{l'-p}^{(n)}\theta_{r'-p}^{(n)}\theta_{s'-p}^{(n)}.
\end{aligned}$$

As in (A.12), in the special case in which the third and fifth moments of ε vanish, and for all (j,m,p),

$$\mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)^4|\mathcal{F}(j)) = \sum_{s=2}^{4}\overline{\mathcal{R}}_{fs}(j,m,p). \qquad (A.18)$$

Here explicit expressions of $\overline{\mathcal{R}}_{fs}(j,m,p)$ are omitted for the sake of space. The key is that we split the left-hand side such that $\overline{\mathcal{R}}_{fs}(j,m,p)$ only depends on $\big(\widetilde{O}(m,p),\ \Sigma,\ \mathcal{R}_{f2}(j)\big)$ for $s=2$, on $\big(\widetilde{O}(m,p),\ \Sigma,\ \mathcal{R}_{f2}(j),\mathcal{R}_{f5}(j)\big)$ for $s=3$, and on $\big(\widetilde{O}(m,p),\ \Sigma,\ \mathcal{R}_{f2}(j),\mathcal{R}_{f7}(j),\mathcal{R}_{f8}(j)\big)$ for $s=4$. Here, and only here, we use Σ to denote a $n_d\times n_d$ matrix whose entry $\Sigma_{i,k}=\Omega_n^{Y,C}(j)_{i,k}+\mathcal{R}_{f2}(j)_{i,k}$. Nonzero third and fifth moments require only heavier notation. Importantly, here we only consider $\mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)^4|\mathcal{F}(j))$ under fixed m and p, as this is all we need. For $s=2$, applying (A.10), (A.13), and (A.14) immediately provides a bound on $\overline{\mathcal{R}}_{fs}(j,m,p)$. For $s=3$, $\mathcal{R}_{f5}(j)$ also appears. The same as proving (A.16), we can bound the product of $\mathcal{R}_{f5}(j)$ and O_{n_d} matrices using (A.9), (A.15), and $|(O_{n_d})_{\widetilde{n}_d+i,k+1}-(O_{n_d})_{\widetilde{n}_d+i,k}|\le Kn_d^{-3/2}\bar{n}_d$. For $s=4$, we need to analyze $\mathcal{R}_{f7}(j)$ and $\mathcal{R}_{f8}(j)$. Again, all we need is a bound similar to (A.15), which comes from trivially generalizing the proof of (B.46). Let us emphasize that results such as (A.9), (A.10), and (A.14) become simplified under our special case of fixed-m. We then have that for all $2\le s\le 4$ and fixed m and p, $\sup_j|\overline{\mathcal{R}}_{fs}(j,m,p)|\le K\bar{n}_d^2n^{-4}$, which, given (A.18), immediately leads to that for fixed m and p,

$$\sup_j \mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)^4|\mathcal{F}(j)) \le K\bar{n}_d^2 n^{-4}. \qquad (A.19)$$

Now we prove that for all (j, m, p),

$$\mathbb{E}\left(\overline{\mathcal{R}}_f(j,m,p)\big(M_{t_{jn_d}} - M_{t_{(j-1)n_d}}\big)|\mathcal{F}(j)\right) = 0, \tag{A.20}$$

for all M that is either W in Assumption 1 or a bounded $\mathcal{F}_t$-martingale orthogonal to W. We define

$$\begin{aligned}
\overline{\mathcal{R}}'_{f1}(j,m,p) &= \sum_{k=1}^{n_d}\sum_{l=1}^{n_d} \widetilde{O}(m,p)_{k,l}\Delta_k^n X^C(j)\Delta_l^n X^C(j), \\
\overline{\mathcal{R}}'_{f2}(j,m,p) &= \sum_{k=1}^{n_d}\sum_{l=1}^{n_d} \widetilde{O}(m,p)_{k,l}\Delta_k^n U^C(j)\Delta_l^n U^C(j), \\
\overline{\mathcal{R}}'_{f3}(j,m,p) &= \sum_{k=1}^{n_d}\sum_{l=1}^{n_d} \widetilde{O}(m,p)_{k,l}\Delta_k^n U^C(j)\Delta_l^n X^C(j).
\end{aligned}$$

Because $\Omega_n^{Y,C}(j)_{k,l}$ is $\mathcal{F}(j)$-measurable and M is a $\mathcal{F}_t$-martingale, the definition of $\overline{\mathcal{R}}_f(j,m,p)$ provided by (A.8) indicates that (A.20) comes directly from that for $s \in \{1,2,3\}$ and for all (j,m,p),

$$\mathbb{E}\left(\overline{\mathcal{R}}'_{fs}(j,m,p)\big(M_{t_{jn_d}} - M_{t_{(j-1)n_d}}\big)|\mathcal{F}(j)\right) = 0. \tag{A.21}$$

We let $\widetilde{\mathcal{F}}(j) = \mathcal{F}(j) \otimes \sigma(\varepsilon_C(j)_i : i \le n_d)$. From the definition of $\varepsilon_C(j)$, we clearly have that $\Delta_k^n U^C(j)$ is $\widetilde{\mathcal{F}}(j)$-measurable for $1 \le k \le n_d$. Therefore, given that $\varepsilon_C(j)$ is independent of $\mathcal{F}_\infty$ and $\{\chi_i : i \ge 1\}$, as required by Assumption 3, we conclude

$$\mathbb{E}\left(\Delta_k^n U^C(j)\Delta_l^n U^C(j)\big(M_{t_{jn_d}} - M_{t_{(j-1)n_d}}\big)|\widetilde{\mathcal{F}}(j)\right) = \Delta_k^n U^C(j)\Delta_l^n U^C(j)\mathbb{E}\left(M_{t_{jn_d}} - M_{t_{(j-1)n_d}}|\mathcal{F}(j)\right).$$

This immediately proves (A.21) for $s = 2$, as M is a $(\mathcal{F}_t)$-martingale. Since by definition $\varepsilon_C(j)$ is also independent of $\mathcal{F}^\varepsilon(j)$, we have $\mathbb{E}(\Delta_k^n U^C(j)|\mathcal{F}_\infty \otimes \mathcal{F}^\chi(j) \otimes \mathcal{F}^\varepsilon(j)) = 0$. Because $\Delta_l^n X^C(j)\big(M_{t_{jn_d}} - M_{t_{(j-1)n_d}}\big)$ is $\mathcal{F}_\infty \otimes \mathcal{F}^\chi(j)$-measurable and we have $\mathcal{F}(j) \subset \mathcal{F}_\infty \otimes \mathcal{F}^\chi(j) \otimes \mathcal{F}^\varepsilon(j)$, we readily obtain (A.21) for $s = 3$. Next, we observe that (A.21) obviously holds for $s = 1$ and $M = W$ as

$$\mathbb{E}\Big(\Delta_k^n X^C(j)\Delta_l^n X^C(j)\big(M_{t_{jn_d}} - M_{t_{(j-1)n_d}}\big)|\mathcal{F}(j)\Big) = \sigma_C(j)\mathbb{E}\Big(\Delta_k^n W(j)\Delta_l^n W(j)\big(W_{t_{jn_d}} - W_{t_{(j-1)n_d}}\big)|\mathcal{F}(j)\Big).$$

Each of $\Delta_k^n W(j)$, $\Delta_l^n W(j)$, and $\big(W_{t_{jn_d}} - W_{t_{(j-1)n_d}}\big)$ is an increment of the process of W after time $t_{(j-1)n_d}$; thus the $\mathcal{F}(j)$-conditional expectation of their product vanishes. Finally, we prove (A.21) for $s = 1$ when M is a bounded $(\mathcal{F}_t)$-martingale orthogonal to W. We define

$$\overline{\mathcal{R}}'_{f4}(s;j,m,p) = \int_{t_{(j-1)n_d}}^{s} \overline{\mathcal{R}}'_{f5}(u;j,m,p)dW_u, \quad \overline{\mathcal{R}}'_{f5}(u;j,m,p) = \int_{t_{(j-1)n_d}}^{u} \widetilde{O}(m,p)_{k(s),l(r)}dW_r,$$

where $k(s) = \min\{k : t(j)_k > s\}$ and $l(r) = \min\{l : t(j)_l > r\}$. Because we have $\overline{\mathcal{R}}'_{f1}(j,m,p) = 2\int_{t_{(j-1)n_d}}^{t_{jn_d}} \overline{\mathcal{R}}'_{f4}(s;j,m,p)dW_s$ and $\overline{\mathcal{R}}'_{f4}(s;j,m,p)$ is $\mathcal{F}_s$-measurable, the fact that M is orthogonal to

W directly indicates $\mathbb{E}\big(\overline{\mathcal{R}}'_{f1}(j,m,p)\big(M_{t_{jn_d}} - M_{t_{(j-1)n_d}}\big)|\mathcal{F}(j)\big) = 0$ and we have proved (A.20).

Step 3. In this step we prove a central limit theorem of the quantity $\sum_{j=1}^{J_d+2}(\mathcal{V}_n(j) - \overline{\mathcal{V}}_n(j))$. We first define $n_t = \max\{i : t_i \le t\}$ and $J(t) = \lfloor n_t/n_d \rfloor + 1$. In view of (A.11) and (A.19) proved in Step 2, we have that for all fixed (m,p) and all $t > 0$,

$$
\begin{aligned}
\sup_{s\le t}\left|(n\bar{J}_d)^{1/2}\sum_{j=1}^{J(s)}\mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)|\mathcal{F}(j))\right| &\le KJ(t)n^{-1/2} = o_{\mathrm{P}}(1), &\text{(A.22)}\\
(n\bar{J}_d)^2\sum_{j=1}^{J(t)}\mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)^4|\mathcal{F}(j)) &\le KJ(t)n^{-1/4} = o_{\mathrm{P}}(1). &\text{(A.23)}
\end{aligned}
$$

The last steps in both lines come from (B.2). In view of the subsequence argument (see, e.g., Andrews and Cheng (2012)), we suppose without loss of generality that both $n^{1/2}\iota^{(n)}\bar{J}_d^{-1}$ and $\sum_{j=-\infty}^{\infty}\kappa_{|j|}^{(n)}$ have limits as $n \to \infty$. Next, we introduce shorthand notation $a = \lim_{n\to\infty} n^{1/2}\iota^{(n)}\bar{J}_d^{-1}$, $\zeta_\kappa^2 = \lim_{n\to\infty}\sum_{j=-\infty}^{\infty}\kappa_{|j|}^{(n)}$, and $A(\lambda)_t = \sigma_t^2\xi_t T + a^2\eta_t^2\zeta_\kappa^2\lambda^2$. We observe that for fixed (m,p) and (m',p'), (A.17) indicates that for some $\alpha_n \to 0$

$$
\begin{aligned}
\sup_j \Big|&\mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)\overline{\mathcal{R}}_f(j,m',p')|\mathcal{F}(j))\\
&-4\delta_{m,m'}\frac{\bar{n}_d}{n^2}\int_0^\pi \cos(p\lambda)\cos(p'\lambda)A\big((2^m-1)\pi + 2^m\lambda\big)^2_{\tau(n,j-1)}d\lambda\Big| \le \alpha_n\bar{n}_d n^{-2}. \qquad \text{(A.24)}
\end{aligned}
$$

On the other hand, we notice that for fixed m and $m\pi \le \lambda \le (m+1)\pi$,

$$
\begin{aligned}
&\mathbb{E}\Big|n_d A(\lambda)^2_{\tau(n,j-1)} - \frac{n}{T}\int_{\tau(n,j-1)}^{\tau(n,j)} A(\lambda)_t^2\xi_t^{-1}dt\Big|\\
\le\ & K\mathbb{E}\Big|\sum_{i=1}^{n_d}(\chi(j)_i - 1) + \sum_{i=1}^{n_d}\big(\xi_{t_{(j-1)n_d+(i-1)}} - \xi_C(j)\big)\chi(j)_i\Big| + o(n_d) = o(n_d). \qquad \text{(A.25)}
\end{aligned}
$$

The first inequality comes from that $A(\lambda)_t$ is locally bounded by Assumption A1, and the observation scheme described by Assumption 2. We also use that $\mathbb{E}(\sup_{\tau(n,j-1)\le t\le\tau(n,j)}|A(\lambda)_t - A(\lambda)_{\tau(n,j-1)}|^2) = o(1)$, indicated by (B.33). To obtain the last equality we also use that $\chi(j)_i$ is i.i.d. and has zero mean and bounded second moment, from Assumption 2. Combining (A.24) and (A.25), and noticing that the boundary summand is negligible, we obtain for all fixed (m,p) and (m',p') and all fixed $t > 0$,

$$
\begin{aligned}
&\frac{n\bar{J}_d}{4}\sum_{j=1}^{J(t)}\Big(\mathbb{E}(\overline{\mathcal{R}}_f(j,m,p)\overline{\mathcal{R}}_f(j,m',p')|\mathcal{F}(j)) - \mathbb{E}\left(\overline{\mathcal{R}}_f(j,m,p)|\mathcal{F}(j)\right)\mathbb{E}(\overline{\mathcal{R}}_f(j,m',p')|\mathcal{F}(j))\Big)\\
=\ &\frac{2^{m+1}\delta_{m,m'}}{\pi T}\int_0^\pi \cos(p\lambda)\cos(p'\lambda)\int_0^t A\big((2^m-1)\pi + 2^m\lambda\big)^2_s\xi_s^{-1}ds\,d\lambda + o_{\mathrm{P}}(1). \qquad \text{(A.26)}
\end{aligned}
$$

Here we use (A.11) to bound the second term in the summand. Further, we notice that $\overline{\mathcal{R}}_f(j,m,p)$ is $\mathcal{F}(j)$-measurable for each (m,p) and n and that $J(t)$ is a $(\mathcal{F}(j))$-stopping time for each n. Finally, we observe that $\int_0^t A(m,\lambda)_s^2 ds$ is $\mathcal{F}_\infty$-measurable and continuous in t. Therefore, in view of (A.22), (A.26), (A.23), and (A.21), Theorem 2.2.15 of Jacod and Protter (2011) indicates that for all $t > 0$,

$$\frac{(n\bar{J}_d)^{1/2}}{2}\sum_{j=1}^{J(t)}\overline{\mathcal{R}}_f(j) \overset{\mathcal{L}_s-\mathcal{F}_\infty}{\longrightarrow} \widetilde{\mathcal{R}}_f(t), \tag{A.27}$$

where $\widetilde{\mathcal{R}}_f(t)$ is a continuous process defined on the extended space $(\overline{\Omega},\overline{\mathcal{F}},(\overline{\mathcal{F}}_t)_{t\geq 0},\overline{\mathbb{P}})$ and which, conditionally on $\mathcal{F}_\infty$, is a centered Gaussian process with independent increments whose component $\widetilde{\mathcal{R}}_f(t,m,p)$ satisfies

$$\begin{aligned}&\overline{\mathbb{E}}\big(\widetilde{\mathcal{R}}_f(t,m,p)\widetilde{\mathcal{R}}_f(t,m',p')\big|\mathcal{F}_\infty\big)\\ &\qquad= \frac{2^{m+1}\delta_{m,m'}}{\pi T}\int_0^\pi \cos(p\lambda)\cos(p'\lambda)\int_0^t A\big((2^m-1)\pi+2^m\lambda\big)_s^2\xi_s^{-1}dsd\lambda.\end{aligned} \tag{A.28}$$

Now we relate $\sum_{j=1}^{J(t)}\overline{\mathcal{R}}_f(j)$ to $\sum_{j=1}^{J_d+2}\big(\mathcal{V}_n(j)-\overline{\mathcal{V}}_n(j)\big)$. Define Θ, $\widetilde{\Theta}\in\mathcal{M}_{n_d}$ and $\widetilde{\Theta}(m,p)\in\mathbb{R}$ as

$$\Theta = \sum_{i=1}^{q_n+2}\bar{\eta}_i\frac{\partial\Omega_{n_d}(\beta^{(n)})^{-1}}{\partial\beta_i}, \qquad \widetilde{\Theta} = O_{n_d}\Theta O_{n_d}, \tag{A.29}$$

$$\widetilde{\Theta}(m,p) = \frac{1}{4\bar{n}_d(m)}\sum_{i=1}^{\bar{n}_d(m)}\widetilde{\Theta}_{\widetilde{n}_d(m)+i,\widetilde{n}_d(m)+i}\Big(e^{\mathbb{i}\pi\frac{(i-1/2)p}{\bar{n}_d(m)}}+e^{-\mathbb{i}\pi\frac{(i-1/2)p}{\bar{n}_d(m)}}\Big). \tag{A.30}$$

Following the same reasoning as in the derivation of (B.25), we obtain that for fixed (m,p) and under $\lim_{n\to\infty} n^{1/2}\iota^{(n)}\bar{J}_d^{-1}\to a$,

$$(n\bar{J}_d)^{-1}\widetilde{\Theta}(m,p) = \widetilde{\Theta}^*(m,p) + o_{\mathrm{P}}(1). \tag{A.31}$$

Here the probability limit $\widetilde{\Theta}^*(m,p)$ is given by $\widetilde{\Theta}^*(m,p) = \frac{1}{\pi}\int_0^\pi \widetilde{\Theta}^*\big((2^m-1)\pi+2^m\lambda\big)\cos(p\lambda)d\lambda$, where $\widetilde{\Theta}^*(\lambda) = \frac{2}{T}(1+\alpha^2a^2\lambda^2)^{-2}\big(2\alpha a\cos\big((1-a)\lambda\big) - \frac{\alpha^2a^2\lambda^2-1}{\lambda}\sin\big((1-a)\lambda\big)\big)$, with the short-hand notation $\alpha = C_T^{-1/2}T^{-1}\zeta_\kappa\big(\int_0^T\eta_s^2\xi_s^{-1}ds\big)^{1/2}$. We also use that either condition (i) or condition (ii) holds. Next, we define

$$\mathcal{R}_g\big(M,\{P_m\}\big) = \sum_{m=0}^{M}\sum_{p=1-P_m}^{P_m}\widetilde{\Theta}(m,p)\sum_{j=1}^{J(T)}\overline{\mathcal{R}}_f(j,m,p). \tag{A.32}$$

Because of (A.31) and the fact that $\widetilde{\Theta}^*(m,p)$ is $\mathcal{F}_\infty$-measurable, the $\mathcal{F}_\infty$-stable convergence in law of $\sum_{j=1}^{J(T)}\overline{\mathcal{R}}_f(j)$ stated in (A.27) immediately indicates that for all fixed $(M,\{P_m\})$,

$$\frac{1}{2}(n\bar{J}_d)^{-1/2}\mathcal{R}_g\big(M,\{P_m\}\big) \overset{\mathcal{L}_s-\mathcal{F}_\infty}{\longrightarrow} \widetilde{\mathcal{R}}_g\big(M,\{P_m\}\big). \tag{A.33}$$

Here $\widetilde{\mathcal{R}}_g\big(M,\{P_m\}\big)$, conditionally on $\mathcal{F}_\infty$, is a centered Gaussian $\mathbb{R}$-valued process with independent increments satisfying

$$\begin{aligned}&\overline{\mathbb{E}}\big(\widetilde{\mathcal{R}}_g(M,\{P_m\})^2\big|\mathcal{F}_\infty\big)\\&\qquad=\frac{2}{\pi T}\sum_{m=0}^{M}2^m\int_0^\pi\Big(\sum_{p=1-P_m}^{P_m}\widetilde{\Theta}^*(m,p)\cos(p\lambda)\Big)^2\int_0^T A\big((2^m-1)\pi+2^m\lambda\big)_s^2\xi_s^{-1}dsd\lambda.\end{aligned}$$

According to Theorem II.8.1 of Zygmund (2002), we have $\lim_{P\to\infty}\sum_{p=1-P}^{P}\widetilde{\Theta}^*(m,p)\cos(p\lambda)=\widetilde{\Theta}^*\big((2^m-1)\pi+2^m\lambda\big)$. This indicates

$$\begin{aligned}\lim_{M\to\infty}\lim_{\underline{P}\to\infty}\overline{\mathbb{E}}(\widetilde{\mathcal{R}}_g(M,\{P_m\})^2|\mathcal{F}_\infty) &= \frac{2}{\pi T}\int_0^T\int_0^\infty\widetilde{\Theta}^*(\lambda)^2A(\lambda)_s^2\xi_s^{-1}dsd\lambda.\\ &= \big(5\alpha a+4(1-a)\big)C(4,\xi)_T+\alpha a C_T^2B(\xi,1)_T. \end{aligned}\tag{A.34}$$

Here $\underline{P}$ stands for $\min_{0\le m\le M}P_m$, and for the second equality we use

$$\frac{8}{\pi}\int_0^\infty\lambda^{2k}(1+\lambda^2)^{-4}\Big(2\cos(b\lambda)-\frac{\lambda^2-1}{\lambda}\sin(b\lambda)\Big)^2d\lambda=\begin{cases}5+4b, & k=0\\ 1, & k\in\{1,2\}\end{cases}.\tag{A.35}$$

From the construction of $\widetilde{\Theta}(m,p)$, $\overline{\mathcal{R}}_f(j,m,p)$, $\mathcal{R}_g(M,P)$, $J(t)$, $\mathcal{V}_n(j)$, and $\overline{\mathcal{V}}_n(j)$, we directly have

$$\sum_{j=1}^{J_d+2}\big(\mathcal{V}_n(j)-\overline{\mathcal{V}}_n(j)\big)=\mathcal{R}_g\big(\widetilde{J}_d-1,\{\bar{n}_d(m)\}\big).\tag{A.36}$$

Using the bounds on $\widetilde{\Theta}(m,p)$ provided by (B.24) and (B.25) and the conditional moments of $\overline{\mathcal{R}}_f(j,m,p)$ provided by (A.11) and (A.17), plus Cauchy-Schwarz inequality, we are able to show that for all $\epsilon>0$, there exist fixed M^* and P^* such that for n large enough and all fixed $M\ge M^*$, $M'\le M^*$, and $\min_{0\le m\le M'}P'_m\ge P^*$,

$$(n\bar{J}_d)^{-1/2}\Big(\mathbb{E}|\mathcal{R}_g(\widetilde{J}_d-1,\{\bar{n}_d\})-\mathcal{R}_g(M,\{\bar{n}_d\})|+\mathbb{E}|\mathcal{R}_g(M',\{\bar{n}_d\})-\mathcal{R}_g(M',\{P'_m\})|\Big)<\varepsilon.\tag{A.37}$$

Here we also use that $\bar{J}_d\to\infty$ under either condition (i) or (ii). Again we omit writing the m-dependence of $\bar{n}_d$. Combining (A.33), (A.36), and (A.37) immediately leads to $\mathcal{F}_\infty$-stable convergence in law of $\frac{1}{2}(n\bar{J}_d)^{-1/2}\sum_{j=1}^{J_d+2}\big(\mathcal{V}_n(j)-\overline{\mathcal{V}}_n(j)\big)$ with asymptotic variance provided by (A.34).

Step 4. This step concludes. According to Lemma B2, we have

$$\sum_{j=1}^{J_d}(\mathcal{U}_n(j)-\overline{\mathcal{U}}_n(j)-\mathcal{V}_n(j)+\overline{\mathcal{V}}_n(j))=o_{\mathrm{P}}(n^{1/2}(q_n+1)^{1/2}+n^{3/4}(\iota^{(n)})^{1/2}).$$

Showing that $\mathcal{U}_n(J_d+1)-\overline{\mathcal{U}}_n(J_d+1)$ and $\mathcal{V}_n(j)-\overline{\mathcal{V}}_n(j)$ with $j\in\{J_d+1,J_d+2\}$ also satisfy the same bound is only a matter of notation. At this stage, given the $\mathcal{F}_\infty$-stable convergence in law of $\frac{1}{2}(n\bar{J}_d)^{-1/2}\sum_{j=1}^{J_d+2}(\mathcal{V}_n(j)-\overline{\mathcal{V}}_n(j))$ proved in the last step, plus using the relation (A.5) and the estimate (B.2), we have proved the current lemma. ∎

A.3 Proof of Lemma A2

Lemma A2. *Suppose Assumptions 1 - 4 hold and q_n is deterministic. We set $\beta_n(\sigma^2,\gamma)=(\sigma^2,\gamma)$. Let $\mathcal{U}^D$ be a random variable defined on the extension $(\overline{\Omega},\overline{\mathcal{F}},(\overline{\mathcal{F}}_t)_{t\geq 0},\overline{\mathbb{P}})$ of $(\Omega_{(0)},\mathcal{F}_\infty,(\mathcal{F}_t)_{t\geq 0},\mathbb{P}_{(0)})$ introduced in Lemma A1, which, conditionally on $\mathcal{F}_\infty$, is standard normal. The central limit theorem*

$$\frac{n^{1/2}\bar{\eta}^\intercal\Xi_{D,n}(\beta^{(n)}(q_n))}{\sqrt{4q_nE(4,\xi)_T+\Delta_n^{-1/2}\zeta^{(n)}(5E(4,\xi)_TC_T^{-1/2}+C_T^{3/2}B(\xi)_T)}}\stackrel{\mathcal{L}_s-\mathcal{F}_\infty}{\longrightarrow}\mathcal{U}^D \tag{A.38}$$

holds if either of the two following conditions is true:

(i) We have $n^{1/2}\iota^{(n)}\to\infty$, $\mathcal{R}^{(n)}(q_n,b)=o_{\mathrm{P}}(1)$, and $q_n\leq Kn^{1/3}$.

(ii) We have $n^{1/2}\iota^{(n)}\leq K$, $\mathcal{R}^{(n)}(q_n,s)=o_{\mathrm{P}}(1)$, $q_n\leq Kn^{1/3}$, and $q_n\to\infty$.

Proof. Step 1. We start with some notation. Let $(j(m),k(m))$ be integers such that $t(j(m))_{k(m)-1}<T_m<t(j(m))_{k(m)}$ and $k(m)\leq n_d$. Then, making use of the notation $\bar{J}_d$ defined before (A.30), we introduce for each $m\geq 1$ three sequences of random vectors

$$\begin{aligned}
\mathcal{R}_{a1}(m,\alpha;v) &= (\alpha\bar{J}_d/n)^{-1/2}\int_{T_m+\alpha v\bar{J}_d/n}^{T_m+\alpha(v+1)\bar{J}_d/n}dW_s,\\
\mathcal{R}_{a2}(m;u,p) &= (n\bar{J}_d)^{1/2}\sum_{l=k(m)+1}^{n_d-1}\Delta\widetilde{O}(u,p)_{k(m),l}\sum_{i=-\infty}^{l}\theta_{l-i}^{(n)}\varepsilon(j(m))_l,\\
\mathcal{R}_{a3}(m;u,p) &= (n\bar{J}_d)^{1/2}\sum_{l=1}^{k(m)}\Delta\widetilde{O}(u,p)_{k(m),l}\sum_{i=-\infty}^{l}\theta_{l-i}^{(n)}\varepsilon(j(m))_l.
\end{aligned}$$

Here the vector $\mathcal{R}_{a1}(m)$ is indexed by a integer v and depends on a positive number α, while the vectors $\mathcal{R}_{a2}(m)$ and $\mathcal{R}_{a3}(m)$ are indexed by integers (u,p) (the ordering does not matter). The matrix $\widetilde{O}(u,p)$ is defined in (A.7) and $\Delta\widetilde{O}(u,p)_{i,l}:=\widetilde{O}(u,p)_{i,l+1}-\widetilde{O}(u,p)_{i,l}$. This step is devoted to proving a central limit theorem related to $\mathcal{R}_{as}(m)$ for $s\in\{1,2,3\}$. For this purpose, we set three independent sequences of i.i.d. random vectors $(\mathcal{U}_{m1},\mathcal{U}_{m2},\mathcal{U}_{m3})_{m\geq 1}$, all defined on the extended probability space $(\overline{\Omega},\overline{\mathcal{F}},(\overline{\mathcal{F}}_t)_{t\geq 0},\overline{\mathbb{P}})$, independent of $\mathcal{F}_\infty$, $\{\varepsilon_i\}$, and $\widetilde{\mathcal{R}}_g(M,P)$ introduced in the proof of Lemma A1. Moreover, for each $m\geq 1$, $\mathcal{U}_{m1}$, $\mathcal{U}_{m2}$, and $\mathcal{U}_{m3}$ are all centered Gaussian whose components satisfy

$$\overline{\mathbb{E}}(\mathcal{U}_{m1}(v)\mathcal{U}_{m1}(v'))=\delta_{v,v'},$$

$$\overline{\mathbb{E}}(\mathcal{U}_{ms}(u,p)\mathcal{U}_{ms}(u',p'))=\frac{2^{u+1}\zeta_\kappa^2\delta_{u,u'}}{\pi}\int_0^\pi\cos(p\lambda)\cos(p'\lambda)\big((2^m-1)\pi+2^m\lambda\big)^2d\lambda,\quad s\in\{2,3\}.$$

We aim to show that for all positive fixed α and as $n \to \infty$,

$$\big(\widetilde{\mathcal{R}}_f^n, (\mathcal{R}_{a1}(m,\alpha), \mathcal{R}_{a2}(m), \mathcal{R}_{a3}(m))_{m\geq 1}\big) \overset{\mathcal{L}_s - \mathcal{F}_\infty}{\longrightarrow} \big(\widetilde{\mathcal{R}}_f, (\mathcal{U}_{m1}, \mathcal{U}_{m2}, \mathcal{U}_{m3})_{m\geq 1}\big), \tag{A.39}$$

where we denote by $\widetilde{\mathcal{R}}_f^n$ the process on the left-hand side of (A.27), i.e., $\widetilde{\mathcal{R}}_f^n(t) = \frac{(n\bar{J}_d)^{1/2}}{2} \sum_{j=1}^{J(t)} \overline{\mathcal{R}}_f(j)$, and $\widetilde{\mathcal{R}}_f$ is defined after (A.27). We emphasize that $\mathcal{R}_{as}(m)$, for all $s \in \{1,2,3\}$, are also n-dependent. We denote by $\bar{T}$ the l-dimensional vector $\bar{T} = (T_1, \ldots, T_M)$. For any $l \geq 1$, we introduce the random set $B_l = \cup_{m=1}^M [(T_m - 1/l)_+, T_m + 1/l]$ and the random family $\mathcal{J}(n,l)$ for all integers $j \geq 1$ such that $(\tau(n,j-1), \tau(n,j-1) + n^{-1/16}) \cap B_l \neq \emptyset$. We also denote by $(\mathcal{G}_t)$ the smallest discrete-time filtration containing $(\mathcal{F}_t)$ and such that $\bar{T}$ is $\mathcal{G}_0$-measurable. Because the process W introduced in Assumption 1 is a $(\mathcal{G}_t)$-Brownian motion, and $\{s \in B_l\} \in \mathcal{G}_0$ for all s, we can define the processes $W(l)_t$ and $\widetilde{\mathcal{R}}_f^{n,(l)}(t)$ as

$$W(l)_t = \int_0^t \mathbb{1}_{B(l)}(s) dW_s \quad \text{and} \quad \widetilde{\mathcal{R}}_f^{n,(l)}(t) = \frac{(n\bar{J}_d)^{1/2}}{2} \sum_{j=1}^{J(t)} \overline{\mathcal{R}}_f(j) \mathbb{1}_{\{j \notin \mathcal{J}(n,l)\}}.$$

Furthermore, we introduce $\widetilde{\mathcal{R}}_f^{(l)}(t)$ as a continuous process defined on the extended probability space $(\bar{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{P}})$, which, conditionally on $\mathcal{F}_\infty$, is a centered Gaussian $\mathbb{R}^d$-valued process with independent increments whose components $\widetilde{\mathcal{R}}_f^{(l)}(t,m,p)$ satisfy

$$\overline{\mathbb{E}}(\widetilde{\mathcal{R}}_f^{(l)}(t,m,p)\widetilde{\mathcal{R}}_f^{(l)}(t,m',p')|\mathcal{F}_\infty) = \frac{2\delta_{m,m'}}{\pi T} \int_0^\pi \cos(p\lambda)\cos(p'\lambda) \int_0^t A(m\pi + \lambda)_s^2 \mathbb{1}_{B(l)}(s) ds d\lambda.$$

Since B_l decreases to the finite set $\{S_1, \ldots, S_l\}$, we clearly have

$$\mathbb{E}\Big(\sup_{s\leq t} \|\widetilde{\mathcal{R}}_f^{(l)}(s) - \widetilde{\mathcal{R}}_f(s)\|\Big) \to 0, \quad \text{as } l \to \infty. \tag{A.40}$$

According to (A.24), the definition of $\mathcal{J}(n,l)$, and Assumption 4, we have for all fixed t and all Lipschitz and bounded function $F(x)$ that depends on the restriction of process x to $[0,t]$,

$$\lim_{j\to\infty} \limsup_{n\to\infty} \mathbb{E}\big|F(\widetilde{\mathcal{R}}_f^{n,(l)}) - F(\widetilde{\mathcal{R}}_f^n)\big| = 0. \tag{A.41}$$

By virtue of the properties of the Skorokhod topology and the definition of stable convergence in law, (A.40) and (A.41) indicate that (A.39) directly comes from the fact that for each l large enough,

$$\big(\widetilde{\mathcal{R}}_f^{n,(l)}, (\mathcal{R}_{a1}(m,\alpha), \mathcal{R}_{a2}(m), \mathcal{R}_{a3}(m))_{m\geq 1}\big) \overset{\mathcal{L}_s - \mathcal{F}_\infty}{\longrightarrow} \big(\widetilde{\mathcal{R}}_f^{(l)}, (\mathcal{U}_{m1}, \mathcal{U}_{m2}, \mathcal{U}_{m3})_{m\geq 1}\big). \tag{A.42}$$

We denote by $\Omega(n,r)$ the set of all ω such that for any $m, m' \in P_r$, it holds that $\tau(n,1) \leq T_m(\omega) \leq \tau(n, J_d - 1)$, and $|T_m(\omega) - T_{m'}(\omega)| \geq n^{-1/16}$, and also $|T_m(\omega) - \tau(n,j)| \geq n^{-1/4}$ for all $1 \leq j \leq J_d - 1$. For positive integer l, we define $j'(m,l) = \max_j\{\tau(n,j-1) + n^{-1/16} < (T_m - 1/l)_+\}$ and denote by $\Omega(n,r,l)$ the set of all $\omega \in \Omega(n,r)$ such that $\sum_{i=1}^{\lfloor n^{7/8} \rfloor} \chi_{i+\tau(n,j'(m,l))} \leq 2n^{7/8}$ for all $m \geq 1$. Since the set $\{T_m : m \in P_r\}$ is finite, in view of Assumption 2, we have $\lim_{n\to\infty} \mathbb{P}(\Omega(n,r)) = 1$ and $\lim_{n\to\infty} \mathbb{P}(\Omega(n,r,l)) = 1$. In view of the definitions of $\varepsilon_C(j)_l$ and of Poisson random measure and in restriction to $\Omega(n,r)$, we obtain that $\mathcal{R}_{as}(m)$ are independent across $s \in \{1,2,3\}$ and that $\mathcal{R}_{a2}(m)$ and $\mathcal{R}_{a3}(m)$ are independent of $\widetilde{\mathcal{R}}_f^{n,(l)}$. On the other hand, $((\mathcal{R}_{a2}(m), \mathcal{R}_{a3}(m))_{m\geq 1}) \overset{\mathcal{L}_s - \mathcal{F}_\infty}{\longrightarrow} ((\mathcal{U}_{m2}, \mathcal{U}_{m3})_{m\geq 1})$ comes from the definition of matrix $\widetilde{O}(u,p)$, Assumption 4, the Lindeberg–Lévy central limit theorem, and the fact that $\bar{J}_d \to \infty$ under either condition (i) or condition (ii) (in the statement of the current lemma). We therefore only need to show for each l large enough,

$$\big(\widetilde{\mathcal{R}}_f^{n,(l)}, (\mathcal{R}_{a1}(m,\alpha))_{m\geq 1}\big) \overset{\mathcal{L}_s - \mathcal{F}_\infty}{\longrightarrow} \big(\widetilde{\mathcal{R}}_f^{(l)}, (\mathcal{U}_{m1})_{m\geq 1}\big).$$

From the definition of stable convergence in law, we only need to show that for all bounded $\mathcal{F}_\infty$-measurable variable Z, all fixed t, and all Lipschitz and bounded function $F(x)$ that depends on the restriction of process x to $[0,t]$,

$$\mathbb{E}\Big(ZF(\widetilde{\mathcal{R}}_f^{n,(l)}) \prod_{m=1}^{M} F_m(\mathcal{R}_{a1}(m,\alpha))\Big) \to \overline{\mathbb{E}}\Big(ZF(\widetilde{\mathcal{R}}_f^{(l)}) \prod_{m=1}^{M} F_m(\mathcal{U}_{m1})\Big).$$

For this purpose, we introduce the σ-fields $\mathcal{H}^{W(l)}$ generated by the variable $W(l)_s$ for $s \geq 0$, and the filtration $(\mathcal{G}(l)_t)$ that is the smallest filtration containing $(\mathcal{G}_t)$ and such that $\mathcal{H}^{W(l)} \in \mathcal{G}(l)_0$. We denote by $\mathbb{Q}_\omega(\cdot)$ the $\mathcal{G}(l)_0$-conditional version of probability $\mathbb{P}$ on $(\Omega, \mathcal{F})$ and $\overline{\mathbb{Q}}_\omega(\cdot)$ the $\mathcal{G}(l)_0$-conditional probability of the extended space $(\bar{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$. We observe that, in restriction to $\Omega(n,r,l)$ and for n large enough, the $\mathcal{G}(l)_{\tau(n,j-1)} \otimes \mathcal{F}^\varepsilon(j) \otimes \mathcal{F}^\chi(j)$-conditional distribution of $\overline{\mathcal{R}}_f(j)$ is the same as its $\mathcal{F}(j)$-conditional distribution for all $j \notin \mathcal{J}(n,l)$, where $\mathcal{F}^\varepsilon(j)$ and $\mathcal{F}^\chi(j)$ are defined after (A.8). We can therefore repeat the proof of Lemma A1 to obtain that $\widetilde{\mathcal{R}}_f^{n,(l)}$ converges stably in law under the measure $\mathbb{Q}_\omega$ toward $\widetilde{\mathcal{R}}_f^{(l)}$. On the other hand, apparently for n large enough, $\mathcal{R}_{a1}(m,\alpha)$ are $\mathcal{G}(l)_0$-measurable. Therefore, it is sufficient to prove that for all bounded $\mathcal{G}(l)_0$-measurable random variables Z',

$$\mathbb{E}\Big(Z' \prod_{m=1}^{M} F_m(\mathcal{R}_{a1}(m,\alpha))\Big) \to \overline{\mathbb{E}}\Big(Z' \prod_{m=1}^{M} F_m(\mathcal{U}_{m1})\Big). \tag{A.43}$$

We observe that the argument of step 6 of the proof of Theorem 4.3.1 of Jacod and Protter (2011) can directly apply here and, as a result, we only need to prove (A.43) for Z' that is $\mathcal{G}_0$-measurable. For each m and n, $\mathcal{R}_{a1}(m,\alpha)$ is a centered Gaussian random vector independent of $\mathcal{G}_0$ with identity covariance matrix by virtue of the definition of Poissson random measure. Moreover, $\mathcal{R}_{a1}(m,\alpha)$ is independent across m in restriction to $\Omega(n,r)$. Therefore (A.43) indeed holds for Z' that is $\mathcal{G}_0$-measurable and we arrive at (A.39).

Step 2. In the sequel we omit writing the argument q_n of $\beta^{(n)}$. We start by defining

$$\begin{aligned}
\mathcal{R}_{b1}(r) &= \bar{\eta}^\intercal \Xi_n^{r,-}(\beta^{(n)}) - \bar{\eta}^\intercal \bar{\Xi}_n^{r,-}(\beta^{(n)}) - \bar{\eta}^\intercal \Xi_n^{B,r}(\beta^{(n)}) + \bar{\eta}^\intercal \bar{\Xi}_n^{B}(\beta^{(n)}), && \text{(A.44)}\\
\mathcal{R}_{b2}(r) &= -\frac{1}{n}\sum_{m\in P_r} \Delta X_{T_m} \sum_{l=1}^{n_d} \Theta_{k(m),l} \Delta_l^n X^{J,r-}(j(m)), && \text{(A.45)}\\
\mathcal{R}_{b3}(r) &= -\frac{1}{n}\sum_{m\in P_r} \Delta X_{T_m} \sum_{l=1}^{n_d} \Theta_{k(m),l} Y_n^{B,r}(j(m))_l. && \text{(A.46)}
\end{aligned}$$

Here Θ is introduced in (A.29). It is straightforward to verify

$$\bar{\eta}^\intercal \Xi_{D,n}(\beta^{(n)}) = \sum_{s=1}^{3} \mathcal{R}_{bs}(r) + (\bar{\eta}^\intercal \Xi_n^{B,r}(\beta^{(n)}) - \bar{\eta}^\intercal \bar{\Xi}_n^{B}(\beta^{(n)})) + \bar{\eta}^\intercal \bar{\Xi}_n(\beta^{(n)}). \tag{A.47}$$

Moreover, we note that the structure of $\mathcal{R}_{b1}(r)$ is the same as that of $\sum_j(\mathcal{U}_n(j) - \overline{\mathcal{U}}_n(j) - \mathcal{V}_n(j) + \overline{\mathcal{V}}_n(j))$ introduced in step 1 of the proof of Lemma A1. And the structure of $\sum_{l=1}^{n_d} \Theta_{k(m),l}\Delta_l^n X^{J,r-}(j(m))$ is very similar to that of $\mathcal{R}_{c3}(j)$ introduced above (B.16). Therefore, by repeating the proof of Lemma B2, plus using the fact that the set $\{T_m : m \in P_r\}$ is finite and that

$$\lim_{r\to\infty} \limsup_{n\to\infty} \mathbb{E}(n|\Delta_i^n X^{J,r-}|^2) \le K \lim_{r\to\infty} \int_{\{z:\Gamma(z)\le 1/r\}} \Gamma(z)^2 \lambda(dz) = 0,$$

from Assumption A1, we obtain that for $s \in \{1, 2\}$,

$$\lim_{r\to\infty} \limsup_{n\to\infty} \mathbb{E}|\mathbb{1}_{\Omega_n'} \mathcal{R}_{bs}(r)| = o(n^{-1/2}\bar{J}_d^{1/2}), \tag{A.48}$$

where recall that Ω_n' is introduced before (B.2). On the other hand, by the definitions of $\bar{\Xi}_n(\beta)$ and $\bar{\Xi}_n^\star$, we conclude

$$\bar{\eta}^\intercal \bar{\Xi}_n(\beta^{(n)}) = \bar{\eta}^\intercal \bar{\Xi}_n^\star(\beta^{(n)}) + o_{\mathrm{P}}(n^{-1/2}\bar{J}_d^{1/2}) = o_{\mathrm{P}}(n^{-1/2}\bar{J}_d^{1/2}). \tag{A.49}$$

The first equality comes from calculations using Lemma A2 of Da and Xiu (2021). The second equality comes from $\bar{\Xi}_n^\star(\beta^{(n)}) = 0$, because of the definitions of $\bar{\Xi}_n^\star$ and $\beta^{(n)}$, Assumption 4, and (B.3). Now we decompose $\mathcal{R}_{b3}(r)$. We need some notation:

$$\begin{aligned}
\mathcal{R}_{a4}(m;u,p) &= (n\bar{J}_d)^{1/2} \sum_{l=k(m)+1}^{n_d} \widetilde{O}(u,p)_{k(m),l} \Delta_l^n W(j(m)),\\
\mathcal{R}_{a5}(m;u,p) &= (n\bar{J}_d)^{1/2} \sum_{l=1}^{k(m)-1} \widetilde{O}(u,p)_{k(m),l} \Delta_l^n W(j(m)),\\
\mathcal{R}_{a6}(m;u,p) &= \iota^{(n)}\eta_{T_m}\mathcal{R}_{a2}(m;u,p) + \iota^{(n)}\eta_{T_m-}\mathcal{R}_{a3}(m;u,p) + \sigma_{T_m}\mathcal{R}_{a4}(m;u,p) + \sigma_{T_m-}\mathcal{R}_{a5}(m;u,p),
\end{aligned}$$

$$
\begin{aligned}
\mathcal{R}_{b4}(m;u,p) &= \sum_{l=1}^{n_d} \widetilde{O}(u,p)_{k(m),l} Y_n^{B,r}(j(m))_l, \\
\mathcal{R}_{b5}(m;u,p) &= \widetilde{O}(u,p)_{k(m),k(m)} \int_{t(j(m))_{k(m)-1}}^{t(j(m))_{k(m)}} \sigma_s dW_s, \\
\mathcal{R}_{b6}(m;u,p) &= \sum_{l=1}^{n_d} \widetilde{O}(u,p)_{k(m),l} \int_{t(j(m))_{l-1}}^{t(j(m))_l} \mu_s^r ds, \\
\mathcal{R}_{b7}(m;u,p) &= \sum_{l=k(m)+1}^{n_d} \widetilde{O}(u,p)_{k(m),l} \int_{t(j(m))_{l-1}}^{t(j(m))_l} (\sigma_s - \sigma_{T_m}) dW_s, \\
\mathcal{R}_{b8}(m;u,p) &= \sum_{l=1}^{k(m)-1} \widetilde{O}(u,p)_{k(m),l} \int_{t(j(m))_{l-1}}^{t(j(m))_l} (\sigma_s - \sigma_{T_m-}) dW_s, \\
\mathcal{R}_{b9}(m,u,p) &= \widetilde{O}(u,p)_{k(m),n_d} U(j(m))_{n_d} - \widetilde{O}(u,p)_{k(m),1} U(j(m))_0, \\
\mathcal{R}_{b10}(m;u,p) &= -\sum_{l=k(m)}^{n_d-1} \Delta\widetilde{O}(u,p)_{k(m),l} \iota^{(n)} (\eta(j(m))_l - \eta_{T_m}) \sum_{i=-\infty}^{l} \theta_{l-i}^{(n)} \varepsilon(j(m))_l, \\
\mathcal{R}_{b11}(m;u,p) &= -\sum_{l=1}^{k(m)-1} \Delta\widetilde{O}(u,p)_{k(m),l} \iota^{(n)} (\eta(j(m))_l - \eta_{T_m-}) \sum_{i=-\infty}^{l} \theta_{l-i}^{(n)} \varepsilon(j(m))_l.
\end{aligned} \tag{A.50}
$$

These definitions lead to $\mathcal{R}_{b4}(m;u,p) = \sum_{s=5}^{10} \mathcal{R}_{bs}(m;u,p) + (n\bar{J}_d)^{-1/2} \mathcal{R}_{a6}(m;u,p)$. On the other hand, using the definition of $\widetilde{\Theta}(u,p)$ from (A.30), we have

$$
\mathcal{R}_{b3}(r) = -\frac{1}{n} \sum_{m \in P_r} \Delta X_{T_m} \sum_{u=0}^{\widetilde{J}_d - 1} \sum_{p=1-\bar{n}_d}^{\bar{n}_d} \widetilde{\Theta}(u,p) \mathcal{R}_{b4}(m;u,p). \tag{A.51}
$$

Here $\widetilde{J}_d$ and $\bar{n}_d$ are defined in the paragraph before (A.30). We now prove that $\mathcal{R}_{bs}(m;u,p)$ for all $s \in \{5,6,7,8,9,10,11\}$ are asymptotically negligible terms. Using Assumptions A1 and 4, Burkholder-Davis-Gundy, that $\widetilde{O}(u,p)$ is deterministic, and (B.18), we have for all $0 \le u \le \widetilde{J}_d - 1$ and $m \in P_r$,

$$
\begin{aligned}
&\sup_p \mathbb{E}|\mathbb{1}_{\Omega(n,r)} \mathcal{R}_{b5}(m;u,p)| \le K n^{-1/2} n_d^{-1} \bar{n}_d, && \text{(A.52)} \\
&\sup_p \mathbb{E}|\mathbb{1}_{\Omega(n,r)} \mathcal{R}_{b6}(m;u,p)| \le K n^{-1} \log n, \\
&\sup_p \mathbb{E}|\mathbb{1}_{\Omega(n,r)} \mathcal{R}_{b7}(m;u,p)^2| \le K n^{-1} n_d^{-1} \bar{n}_d \mathbb{E}\Big(\sup_{T_m \le s \le t(j(m)+1)} |\sigma_s - \sigma_{T_m}|^4 \Big)^{1/2}, && \text{(A.53)} \\
&\sup_p \mathbb{E}|\mathbb{1}_{\Omega(n,r)} \mathcal{R}_{b8}(m;u,p)^2| \le K n^{-1} n_d^{-1} \bar{n}_d \mathbb{E}\Big(\sup_{t(j(m)) \le s < T_m -} |\sigma_s - \sigma_{T_m-}|^4 \Big)^{1/2}, && \text{(A.54)} \\
&\sup_p \mathbb{E}|\mathbb{1}_{\Omega(n,r)} \mathcal{R}_{b9}(m,u,p)| \le K n_d^{-2} \bar{n}_d^2 \iota^{(n)}, \\
&\sup_p \mathbb{E}|\mathbb{1}_{\Omega(n,r)} \mathcal{R}_{b10}(m;u,p)^2| \le K n_d^{-3} \bar{n}_d^3 (\iota^{(n)})^2 \mathbb{E}\Big(\sup_{T_m \le s \le t(j(m)+1)} |\xi_s - \xi_{T_m}|^4 \Big)^{1/2}, && \text{(A.55)}
\end{aligned}
$$

$$\sup_p \mathbb{E}|\mathbb{1}_{\Omega(n,r)}\mathcal{R}_{b11}(m;u,p)^2| \leq Kn_d^{-3}\bar{n}_d^3(\iota^{(n)})^2\mathbb{E}\Big(\sup_{t(j(m))\leq s<T_m-}|\xi_s-\xi_{T_m-}|^4\Big)^{1/2}. \quad \text{(A.56)}$$

For the last three inequalities, we additionally use $|(O_{n_d})_{\widetilde{n}_d+i,k+1}-(O_{n_d})_{\widetilde{n}_d+i,k}| \leq Kn_d^{-3/2}\bar{n}_d$. Since all of the terms on the right-hand side of (A.53) – (A.56) go to zero from the dominated convergence theorem and the fact that both σ and ξ are càdlàg, we obtain that for all $0 \leq u \leq \widetilde{J}_d - 1$ and some $\alpha_n \to 0$,

$$\sup_{m,p}\mathbb{E}\big|\mathbb{1}_{\Omega(n,r)}\big(\mathcal{R}_{b4}(m;u,p)-(n\bar{J}_d)^{-1/2}\mathcal{R}_{a6}(m;u,p)\big)\big| \leq Kn^{-1/2}n_d^{-1}\bar{n}_d + \alpha_n n^{-1/2}n_d^{-1/2}\bar{n}_d^{1/2}. \quad \text{(A.57)}$$

In view of (B.24), the relation (A.51) and the bound (A.57) lead to

$$\mathcal{R}_{b3}(r) = -\frac{1}{n}(n\bar{J}_d)^{-1/2}\sum_{m\in P_r}\Delta X_{T_m}\sum_{u=0}^{\widetilde{J}_d-1}\sum_{p=1-\bar{n}_d}^{\bar{n}_d}\widetilde{\Theta}(u,p)\mathcal{R}_{a6}(m;u,p) + o_{\mathrm{P}}(n^{-1/2}\bar{J}_d^{1/2}). \quad \text{(A.58)}$$

That $q_n \to \infty$ when $n^{1/2}\iota^{(n)} \leq K$, required by condition (ii), is essential for deriving (A.58). Indeed, without $\bar{J}_d := n^{1/2}\iota^n + q_n \to \infty$, the first term on the right side of (A.57), which comes from $\mathcal{R}_{b5}(m;u,p)$ (bounded by (A.52) and defined by (A.50)), would not be $o_{\mathrm{P}}(n^{-1/2}\bar{J}_d^{1/2})$ as in (A.58), and would generate a non-Gaussian asymptotic distribution of $\mathcal{R}_{b3}(r)$.

Step 3. Given (A.58), we now study $\mathcal{R}_{a6}(m;u,p)$, which is a linear combination of $\mathcal{R}_{as}(m;u,p)$ with $s \in \{2,3,4,5\}$. In view of the definition of $\widetilde{O}(u,p)$, and using Burkholder-Davis-Gundy and (B.33), we obtain with direct calculation that for all fixed (u,p),

$$\lim_{\alpha\to 0^+,\alpha\bar{v}\to\infty}\limsup_{n\to\infty}\mathbb{E}\Big|\mathbb{1}_{\Omega(n,r)}\Big(\mathcal{R}_{a4}(m;u,p)-\sum_{v=0}^{\bar{v}}\mathcal{R}_{a1}(m,\alpha;v)\mathcal{R}_{a7+}(m,\alpha;p,v,u)\Big)\Big| = 0, \quad \text{(A.59)}$$

$$\lim_{\alpha\to 0^+,\alpha\bar{v}\to\infty}\limsup_{n\to\infty}\mathbb{E}\Big|\mathbb{1}_{\Omega(n,r)}\Big(\mathcal{R}_{a5}(m;u,p)-\sum_{v=-\bar{v}}^{-1}\mathcal{R}_{a1}(m,\alpha;v)\mathcal{R}_{a7-}(m,\alpha;p,v,u)\Big)\Big| = 0, \quad \text{(A.60)}$$

where we use the shorthand notation

$$\mathcal{R}_{a7\pm}(m,\alpha;p,v,u) = 2^{u+1}\alpha^{1/2}\pi^{-1}\int_0^\pi \cos(p\lambda)\cos\big(\alpha v((2^u-1)\pi+2^u\lambda)(\xi_{T_m\pm}T)^{-1}\big)d\lambda.$$

(Note $\xi_{T_m+} = \xi_{T_m}$ as ξ is càdlàg.) Combined with (A.39), we obtain

$$\big(\widetilde{\mathcal{R}}_f^n,(\mathcal{R}_{a2}(m),\mathcal{R}_{a3}(m),\mathcal{R}_{a4}(m),\mathcal{R}_{a5}(m))_{m\geq 1}\big) \overset{\mathcal{L}_s-\mathcal{F}_\infty}{\longrightarrow} \big(\widetilde{\mathcal{R}}_f,(\mathcal{U}_{m2},\mathcal{U}_{m3},\mathcal{U}_{m4},\mathcal{U}_{m5})_{m\geq 1}\big). \quad \text{(A.61)}$$

Here $(\mathcal{U}_{m4},\mathcal{U}_{m5})_{m\geq 1}$ are two sequences of random vectors defined on $\big(\overline{\Omega},\overline{\mathcal{F}},(\overline{\mathcal{F}}_t)_{t\geq 0},\overline{\mathbb{P}}\big)$. Conditionally on $\mathcal{F}_\infty$, $(\mathcal{U}_{m4})_{m\geq 1}$ and $(\mathcal{U}_{m5})_{m\geq 1}$ are mutually independent, independent of $(\widetilde{\mathcal{R}}_f,(\mathcal{U}_{m2},\mathcal{U}_{m3})_{m\geq 1})$, and independent across m. For each $m \geq 1$, $\mathcal{U}_{m4}$ and $\mathcal{U}_{m5}$, conditionally on $\mathcal{F}_\infty$, are both centered

Gaussian whose components satisfy, for $s \in \{4, 5\}$,

$$\overline{\mathbb{E}}(\mathcal{U}_{ms}(u,p)\mathcal{U}_{ms}(u',p')) = (1+\delta_{p,0})2^u \delta_{u,u'}\delta_{p,p'}T(\xi_{T_m}\mathbb{1}_{\{s=4\}} + \xi_{T_m-}\mathbb{1}_{\{s=5\}}).$$

On the other hand, following the same reasoning in deriving for (A.57), we obtain for all $0 \le u \le \widetilde{J}_d - 1$ that $\sup_{m,p}\sum_{s=2}^{5}\mathbb{E}|\mathbb{1}_{\Omega(n,r)}\mathcal{R}_{as}(m;u,p)| \le K\bar{J}_d n_d^{-1/2}\bar{n}_d^{1/2}$. Therefore, in view of the central limit theorem (A.61), the bounds on $\widetilde{\Theta}(u,p)$ provided by (B.24) and (B.25), the $\mathcal{F}_\infty$-measurable asymptotic limit of $\widetilde{\Theta}(u,p)$ for fixed (u,p) provided by (A.31) under either condition (i) or condition (ii), the integral (A.35), and the definition of $\mathcal{R}_{a6}(m;u,p)$ as a weighted sum of $\mathcal{R}_{as}(m;u,p)$ with $s \in \{2,3,4,5\}$ and $\mathcal{F}_\infty$-measurable weighting coefficients, we obtain that under $n^{1/2}\iota^{(n)}\bar{J}_d^{-1} \to a$,

$$\big(\widetilde{\mathcal{R}}_f^n, (\mathcal{R}_{a8}(m))_{m\ge1}\big) \overset{\mathcal{L}_s-\mathcal{F}_\infty}{\longrightarrow} \big(\widetilde{\mathcal{R}}_f, (\mathcal{U}_{m8})_{m\ge1}\big), \tag{A.62}$$

where the shorthand notation $\mathcal{R}_{a8}(m)$ stands for

$$\mathcal{R}_{a8}(m) = (n\bar{J}_d)^{-1}\sum_{u=0}^{\bar{J}_d-1}\sum_{p=1-\bar{n}_d(u)}^{\bar{n}_d(u)}\widetilde{\Theta}(u,p)\mathcal{R}_{a6}(m;u,p).$$

Here $(\mathcal{U}_{m8})_{m\ge1}$ is a sequence of random variables defined on $\big(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\ge0}, \overline{\mathbb{P}}\big)$, which, conditionally on $\mathcal{F}_\infty$, is independent of $\widetilde{\mathcal{R}}_f$ and across m, is centered Gaussian for each $m \ge 1$, and satisfies

$$\overline{\mathbb{E}}(\mathcal{U}_{m8}^2|\mathcal{F}_\infty) = \frac{1}{T}\big(5\alpha a + 4(1-a)\big)\big(\xi_{T_m}\sigma_{T_m}^2 + \xi_{T_m-}\sigma_{T_m-}^2\big) + \frac{1}{T^2}\alpha^{-1}a\zeta_\kappa^2\big(\eta_{T_m}^2 + \eta_{T_m-}^2\big). \tag{A.63}$$

The shorthand notation α is introduced after (A.28). The analysis from (A.32) to the end of the proof of Lemma A1 indicates that we can replace $\widetilde{\mathcal{R}}_f^n$ and $\widetilde{\mathcal{R}}_f$ in (A.62) with, respectively, $(n/\bar{J}_d)^{1/2}\big(\bar{\eta}^\intercal\Xi_n^{B,r}(\beta^{(n)}) - \bar{\eta}^\intercal\bar{\Xi}_n^B(\beta^{(n)})\big)$ and $\mathcal{U}_7$, where $\mathcal{U}_7$ is a random variable on $\big(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\ge0}, \overline{\mathbb{P}}\big)$, which, conditionally on $\mathcal{F}_\infty$, is independent of $(\mathcal{U}_{m8})_{m\ge1}$ and is centered Gaussian with $\overline{\mathbb{E}}(\mathcal{U}_7^2|\mathcal{F}_\infty)$ given by the right-hand side of (A.34). Therefore, in view of (A.47), (A.48), (A.49), (A.58), and (A.63), plus using Assumption 2 and the definitions of $E(4,\xi)_T$ and $B(\xi)_T$, we have already obtained the desired central limit theorem of $\bar{\eta}^\intercal\Xi_{D,n}(\beta^{(n)}(q_n))$. ∎

A.4 Proof of Theorem 1

Proof. We let $F(\cdot)$ be the standard Gaussian CDF. Furthermore, to simplify notation, we set

$$\mathbb{P}_n(\alpha;\iota^{(n)},\theta^{(n)}) = \mathbb{P}\left(\frac{n^{1/2}(\widehat{\sigma}_n^2(\widehat{q}_n) - C_T)}{\sqrt{4\widehat{q}_n E(4,\xi)_T + \Delta_n^{-1/2}\zeta^{(n)}(5E(4,\xi)_T C_T^{-1/2} + 3C_T^{3/2}B(\xi)_T)}} \in (-\infty, F^{-1}(\alpha))\right).$$

We denote by $\mathcal{P}$ the collection of all DGP sequences $\{\mathbb{P}, \iota^{(n)}, \theta^{(n)}\}_{n\ge1}$ that satisfy Assumptions 1 - 5. For any set Γ consisting of parameter sequences $\{\iota^{(n)}\}_{n\ge1}$, let $\mathcal{P}(\Gamma)$ be the collection of all DGP sequences $\{\mathbb{P}, \iota^{(n)}, \theta^{(n)}\}_{n\geq 1}$ within $\mathcal{P}$ under which $\{\iota^{(n)}\} \in \Gamma$. Note that Theorem 1 is equivalent to

$$\inf_{\{\mathbb{P},\iota^{(n)},\theta^{(n)}\}\in\mathcal{P}} \liminf_{n\to\infty} \mathbb{P}_n(\alpha;\iota^{(n)},\theta^{(n)}) = \alpha = \sup_{\{\mathbb{P},\iota^{(n)},\theta^{(n)}\}\in\mathcal{P}} \limsup_{n\to\infty} \mathbb{P}_n(\alpha;\iota^{(n)},\theta^{(n)}). \tag{A.64}$$

To proceed, we introduce two sets of parameter sequences: $\Gamma_1(a^2) = \{\{\iota^{(n)} : n \geq 1\} : n^{1/2}\iota^{(n)} \leq a^2\}$ and $\Gamma_2 = \{\{\iota^{(n)} : n \geq 1\} : n^{1/2}\iota^{(n)} \to \infty\}$. In view of Lemma A2 and Lemmas B1, B3, and B4, we have for any $0 \leq a^2 < \infty$,

$$\inf_{\{\mathbb{P},\iota^{(n)},\theta^{(n)}\}\in\mathcal{P}(\Gamma_1(a^2)\cup\Gamma_2)} \liminf_{n\to\infty} \mathbb{P}_n(\alpha;\iota^{(n)},\theta^{(n)}) = \alpha. \tag{A.65}$$

We show the first equality in (A.64); showing the second is a simple repetition. Let $\{\check{\mathbb{P}}, \check{\iota}^{(n)}, \check{\theta}^{(n)}\} \in \mathcal{P}$ be a DGP sequence such that $\liminf_{n\to\infty} \check{\mathbb{P}}_n(\alpha;\check{\iota}^{(n)},\check{\theta}^{(n)}) = \inf_{\mathcal{P}} \liminf_{n\to\infty} \mathbb{P}_n(\alpha;\iota^{(n)},\theta^{(n)})$. Such a sequence always exists. Let $\{u_n\}$ be a subsequence of $\{n\}$ such that $\lim_{n\to\infty} \check{\mathbb{P}}_{u_n}(\alpha;\check{\iota}^{(u_n)},\check{\theta}^{(u_n)}) = \inf_{\mathcal{P}} \liminf_{n\to\infty} \mathbb{P}_n(\alpha;\iota^{(n)},\theta^{(n)})$. Such a sequence also always exists. The first equality in (A.64) follows once we show that there exists a subsequence $\{v_n\}$ of $\{u_n\}$ such that

$$\lim_{n\to\infty} \check{\mathbb{P}}_{v_n}(\alpha;\check{\iota}^{(v_n)},\check{\theta}^{(v_n)}) = \alpha. \tag{A.66}$$

To this end, we explicitly construct $\{v_n\}$. Let $\Gamma_3 = \{\{\iota^{(n)} : n \geq 1\} : \limsup_{n\to\infty} n^{1/2}\iota^{(n)} < \infty\}$ and $\Gamma_4 = \{\{\iota^{(n)} : n \geq 1\} : \limsup_{n\to\infty} n^{1/2}\iota^{(n)} = \infty\}$. Observe that by construction we have $\{\check{\iota}^{(u_n)}\} \in \Gamma_3 \cup \Gamma_4$. If $\{\check{\iota}^{(u_n)}\} \in \Gamma_3$, we can always find some subsequence $\{v_n\}$ of $\{u_n\}$ such that $\{\check{\iota}^{(v_n)}\} \in \Gamma_1(a^2)$ with $a^2 < \infty$. If $\{\check{\iota}^{(u_n)}\} \in \Gamma_4$, then we can always find some subsequence $\{v_n\}$ of $\{u_n\}$ such that $\{\check{\iota}^{(v_n)}\} \in \Gamma_2$. Thus, to prove the first equality in (A.64), it suffices to demonstrate that (A.66) holds for any $\{\check{\mathbb{P}}, \check{\iota}^{(v_n)}, \check{\theta}^{(v_n)}\} \in \mathcal{P}(\Gamma_1(a^2) \cup \Gamma_2)$ with any $\{v_n\}$ and $a^2 \in [0,\infty)$. Given (A.65), this can be proved by showing that for any such $\{\check{\iota}^{(v_n)}\}$, there always exists a sequence $\{\iota_*^{(n)}\}$ such that

$$\{\iota_*^{(n)}\} \in \Gamma_1(a^2) \cup \Gamma_2 \text{ and } \iota_*^{(v_n)} = \check{\iota}^{(v_n)}, \forall n \geq 1. \tag{A.67}$$

To do so, we construct another sequence. If $v_n^{1/2}\check{\iota}^{(v_n)} \leq a^2$, we construct $\{\iota_*^{(k)} : k \geq 1\}$ as follows: (i) $\forall k = v_n$, define $\iota_*^{(k)} = \check{\iota}^{(k)}$, and (ii) $\forall k \in (v_n, v_{n+1})$, define $\iota_*^{(k)} = \check{\iota}^{(k)}{v_n}^{1/2}k^{-1/2}$. We observe that $\{\widetilde{\iota}^{(k)}\} \in \Gamma_1(a^2)$ with $a^2 \in (0,\infty)$. In the same way, we can construct a $\{\iota_*^{(n)}\}$ that satisfies (A.67) under $v_n^{1/2}\check{\iota}^{(v_n)} \to \infty$. So (A.67) is satisfied and we obtain the first equality in (A.64). The proof ends. ∎