Disentangling Autocorrelated Intraday Returns

Rui Da*
Indiana University

Dacheng Xiu†
University of Chicago and NBER

This version: June 20, 2025

Abstract

We propose a semiparametric approach to disentangling the autocovariance of equity returns at high frequency. We assume the observed price consists of an efficient component that follows a nonparametric continuous-time Itô-semimartingale, along with a market microstructure component that follows a discrete-time moving-average model. Our quasi-likelihood procedure relies on a misspecified moving-average model selected by information criteria. We establish the model-selection consistency, provide a central limit theory on autocovariance parameters, and show their consistency uniformly over a large class of models that allow for an arbitrary noise magnitude and a flexible dependence structure. We also provide a quadratic representation of the likelihood estimator, which sheds light on its connection with nonparametric kernel estimators. Our simulation evidence suggests that our estimator outperforms the nonparametric alternatives particularly when noise magnitude is small. We apply this estimator to S&P 1500 index constituents, and find that in recent years the microstructure friction has become smaller but existed in 5-minute returns, particularly in small caps, and that the average duration of autocorrelations for large caps has shrunk considerably to merely 10 seconds.

Keywords: Moving-Average Models, Noise Autocorrelations, QMLE, Small Noise

*Address: 1275 E 10th St, Bloomington, IN 47405. E-mail: ruida@iu.edu.
†Address: 5807 S Woodlawn Ave, Chicago, IL 60637, USA. E-mail: dacheng.xiu@chicagobooth.edu.

1 Introduction

Autocorrelations in stock returns are ubiquitous. The earlier literature regards such autocorrelations as evidence against market efficiency. Nonetheless, as market efficiency has improved over past decades, autocorrelations have remained a salient feature of intraday stock returns sampled at sufficiently high frequencies. The modern view of such autocorrelations is that they arise from market microstructure frictions, such as bid-ask bounces, nonsynchronous trading, price discreteness, etc, which coalesce into efficient equilibrium prices and lead to the convoluted dynamics of returns.

To disentangle the observed autocorrelations in intraday returns, we model the transaction price as a discretized continuous-time semimartingale process plus a discrete-time moving-average process. The former represents the efficient price process that features return heteroscedasticity in the form of stochastic volatility and jumps, but does not contribute to any autocovariance; the latter serves as a reduced-form description of the microstructure friction that is the main driver behind the observed autocovariances.

To conduct inference on various model components and parameters, we construct a tractable quasi-maximum likelihood estimator (QMLE), pretending that the transaction price arrives regularly and comprises a Brownian motion with constant volatility and an MA(q) noise. We select q based on the Akaike/Bayesian information criteria (AIC/BIC). While our estimator shares the same likelihood with that from an MA($q+1$) model, our asymptotic design is in-fill, i.e., the number of observations increases within a fixed window–say, a trading day–which renders our analysis rather different from the usual long-span asymptotics in the classic time-series setting.

In a related paper, Da and Xiu (2021) show how to conduct uniformly valid inference on volatility over a large class of MA(∞) models that allow for an asymptotically vanishing noise with a flexible dependence structure. In this paper, our main objective is to develop asymptotic properties of the estimator for noise parameters. When the noise data-generating process (DGP) follows a finite-order moving-average model, we show that our quasi-likelihood estimator, combined with BIC, recovers the true model asymptotically, is consistent with respect to the noise parameters, and achieves a pointwise central limit theory at the usual rate of $n^{1/2}$. Moreover, we develop uniform consistency results when noise follows an MA(∞) process. As alternatives to our semiparametric approach, Jacod et al. (2017) and Li and Linton (2021) provide nonparametric estimators of the serial correlations of the microstructure noise based on local averaging and differencing strategies, respectively. They focus on the case in which noise is large, whereas we also allow for vanishing noise. More importantly, our likelihood-based approach provides a benchmark on the efficiency of noise parameters.

We apply our estimator to analyze all intraday returns of S&P 1,500 index constituents from 1996 to 2016. Several interesting findings emerge. The microstructure noise is present in 5-minute returns, at least for small and mid caps, though it is an order of magnitude smaller in recent years than at the beginning of the sample, thanks to the improvement in market efficiency. For a sizable portion of stock-day pairs, it appears that the noise is either absent or approximately follows an i.i.d. assumption. For the remaining stocks with autocorrelated noise, the duration of autocorrelations has been on the decline, from several minutes in 1996 to merely 10 seconds on average for large caps and 100 seconds for small caps in 2016.

Empirical evidence of autocorrelations in the returns of transaction prices goes back to as early as Niederhoffer and Osborne (1966), Simmons (1971), and Garbade and Lieber (1977). Among others, Hasbrouck and Ho (1987) document positive autocorrelations in intraday stock returns, in returns of quote midpoints, and in the arrival of buy and sell orders. They thus propose a model of the return-generating process, which is observationally equivalent to an ARMA$(2, 2)$ model. While classical time-series models such as ARMA are convenient for dependent data, they are not appropriate for intraday returns because of the heteroscedasticity in returns.

Why do higher-order autocorrelations of returns exist? There are many economic hypotheses, such as strategic order splitting (Garbade and Lieber (1977)); optimal control of execution cost (Bertsimas and Lo (1998)); price impact and inventory control (Kyle (1985), Amihud and Mendelson (1980)); the crowd effect or herding (Tóth et al. (2015)); and high-frequency trading Brogaard et al. (2014). Our objective here is modest. We aim to estimate parameters in a general class of reduced-form models, since many structural economic models yield specific reduced-form models–see, for example, Hasbrouck (2007)–with differences only in how the reduced-form parameters relate to structural parameters.

There is an enormous literature on the estimation of quadratic variation or its components using noisy high-frequency data; e.g., the two-scale or multi-scale estimators by Zhang et al. (2005) and Zhang (2006); the realized kernel estimator and its extensions by Barndorff-Nielsen et al. (2008) and Barndorff-Nielsen et al. (2011); the pre-averaging estimator by Jacod et al. (2009) and Jacod et al. (2010); the quasi-maximum likelihood estimator (QMLE) by Xiu (2010); and the local method of moments estimator by Reiß (2011). An "essentially" white noise assumption is most common in this strand of the literature, with the exception of Jacod et al. (2019), Varneskov (2016), and Da and Xiu (2021), who tackle general colored-noise processes for the purpose of volatility estimation. Related work also include Aït-Sahalia et al. (2005), Aït-Sahalia et al. (2011), Kalnina and Linton (2008), and Bibinger et al. (2019). Unlike the above papers, which treat noise as nuisance parameters in the estimation of quadratic variation, our target here is mainly the temporal dependence of intraday returns beyond the first-order autocorrelations. Chang et al. (2018) also focus on analyzing the statistical properties of the noise process and propose an estimator of noise density and noise moments in an i.i.d. noise setting.

The paper is organized as follows. Section 2 sets up the model. Section 3 introduces the estimator and provides the main asymptotic results. Section 4 reports Monte Carlo simulations. We analyze volatilities and noise for S&P Composite 1,500 index constituents in Section 5, and Section 6 concludes. The online supplemental appendix contains all mathematical proofs.

2 Model Assumptions

We assume that transaction prices $\widetilde{X}$ are observed at t_i, for $i = 1, 2, \ldots, n_T$, within a fixed window $[0, T]$. They comprise two components: $\widetilde{X}_{t_i} = X_{t_i} + U_i$, where X_{t_i} is (the logarithm of) the efficient equilibrium price and U_i is the microstructure noise associated with the ith observation. Furthermore, the efficient price satisfies:

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*

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

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, $|\delta(\omega, t, u)| \wedge 1 \le \Gamma_m(u)$ for all (ω, t, u) with $t \le \tau_m(\omega)$, where $\{\tau_m\}$ is a localizing sequence of stopping times and $\{\Gamma_m\}$ a sequence of deterministic functions satisfying $\int \Gamma_m^2(u)\lambda(du) < \infty$.

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

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

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 $\widetilde{\Gamma}_m(u)$, $\|\widetilde{\delta}(\omega, t, u)\| \wedge 1 \le \widetilde{\Gamma}_m(u)$ for all $\omega \in \Omega$, $t \le \tau_m(\omega)$, and $\int \widetilde{\Gamma}_m^2(u)\lambda(du) < \infty$.

While the efficient prices are defined in continuous time, we only observe their noisy versions at discrete time points. We now describe the assumption of the arrival times of transactions:

Assumption 2. *The sequence of observation times* $\{t_i : i \geq 0\}$ *satisfies* $t_0 = 0$ *and* $t_i = t_{i-1} + \frac{T}{n}\xi_{t_{i-1}}\chi_i$, *where the sequence* $\{\chi_i : i \geq 1\}$ *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* $\xi = (\xi_t)_{t\geq 0}$ *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.*

The intervals between adjacent transactions are determined by a continuous-time process, ξ_t, and a discrete-time process, χ_i, jointly. This assumption allows for dependence between trading times and the underlying driving forces of efficient prices, and thereby accommodates a large class of sampling schemes; see Jacod et al. (2017) for detailed discussions.

Next, we impose a discrete-time moving-average process for the microstructure noise to capture the potential temporal dependence in the transaction prices:[1]

Assumption 3. *The noise sequence* $\{U_i : i \geq 0\}$ *consists of random variables defined on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *such that* $\{U_i : i \geq 0\}$ *has an MA*(∞) *representation with mean* 0*:*

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

where B *is the lag operator;* $\varepsilon_i \overset{i.i.d.}{\sim} (0, 1)$, *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* $\iota^{(n)} \leq K$.

The noise again depends on a continuous-time process η_t and a discrete-time moving-average process U. The former introduces dependence between noise and the underlying efficient price, whereas the latter dictates the temporal dependence of the noise. Combining the two allows for heteroscedastic, temporally dependent, and endogenous noise.

The parameters of interest in our study are autocovariances $\{\gamma_j^{(n)} : j \geq 0\}$ and autocorrelations $\{\rho_j^{(n)} : j \geq 1\}$ of the noise process, defined as

$$\gamma_j^{(n)} = (\iota^{(n)})^2 \frac{\int_0^T \eta_s^2 \xi_s^{-1} ds}{\int_0^T \xi_s^{-1} ds} \times \kappa_j^{(n)}, \quad j \geq 0, \quad \text{and} \quad \rho_j^{(n)} = \kappa_j^{(n)}/\kappa_0^{(n)}, \quad j \geq 1, \tag{2.4}$$

[1] We use a superscript (n) on noise parameters to facilitate discussion of uniformity over different sequences of data-generating processes (DGPs) of noise indexed by n. n is a nonobservable mathematical abstraction. All limits are taken as $n \to \infty$. K is a generic n-independent positive constant that may vary from line to line.

where $\kappa_j^{(n)}$ is given by

$$\kappa_j^{(n)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda;\theta^{(n)})e^{\mathbb{i}\lambda j}d\lambda, \quad j \geq 0, \tag{2.5}$$

and $g(\lambda;\theta^{(n)}) = |\theta^{(n)}(e^{\mathbb{i}\lambda})|^2$ is the spectral density of $\theta^{(n)}(\mathrm{B})\varepsilon$. While the autocovariances depend on the underlying processes η_t and ξ_t that drive the sampling times and noise magnitudes, respectively, the autocorrelations are entirely determined by the set of parameters $\{\theta_j^{(n)} : j = 1, 2, \ldots, \infty\}$ in the MA process.

Finally, we need some regularity assumption on the behavior of the spectral density of the noise process so that it is uniformly invertible and its long range dependence cannot be overly strong.

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

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

3 Main Results

In what follows, we will discuss the constructed estimators and their asymptotic properties.

3.1 Quasi-likelihood Estimation

To estimate volatility, Da and Xiu (2021) propose a quasi-likelihood approach based on a misspecified model for observed returns. We adopt the same estimator here, but focus on the noise parameters. Specifically, we pretend that the efficient price X (in logarithm) is a Brownian motion with constant volatility but no drift, and that the noise U follows a Gaussian MA(q) model with the order q to be determined:

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

Under this model, the observed log-return vector $Y_n = (Y_{n,1}, Y_{n,2}, \ldots, Y_{n,n_T})^\intercal$,

$$Y_{n,i} = X_{t_i} - X_{t_{i-1}} + U_i - U_{i-1}, \quad 1 \leq i \leq n_T. \tag{3.6}$$

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

$$\Sigma_n(\sigma^2, \iota^2, \theta) = \sigma^2 \Delta_n \mathbb{I}_n + \sum_{h=0}^{n_T-1} (2\gamma_h - \gamma_{h+1} - \gamma_{h-1}) \mathbb{G}_n^h, \tag{3.7}$$

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:

$$\gamma_h = \frac{\iota^2}{2\pi} \int_{-\pi}^{\pi} g(\lambda; \theta) e^{\mathrm{i}\lambda h} d\lambda, \quad \text{where} \quad g(\lambda; \theta) = \left|\theta(e^{\mathrm{i}\lambda})\right|^2. \tag{3.8}$$

Since we are interested in the noise autocovariances, we reparameterize the likelihood function in terms of (σ^2, γ):

$$L_n(\sigma^2, \gamma) = -\frac{1}{2} \log \det(\Sigma_n(\sigma^2, \gamma)) - \frac{1}{2} \mathrm{tr}(\Sigma_n(\sigma^2, \gamma)^{-1} Y_n Y_n^{\intercal}), \tag{3.9}$$

where $\Sigma_n(\sigma^2, \gamma) := \Sigma_n(\sigma^2, \iota^2, \theta)$ and γ is the $(q+1)$-dimensional vector of the noise autocovariances.

We define $(\widehat{\sigma}_n^2(q), \widehat{\gamma}_n(q))$ as the maximizer of $L_n(\sigma^2, \gamma)$:

$$(\widehat{\sigma}_n^2(q), \widehat{\gamma}_n(q)) = \arg \max_{(\sigma^2, \gamma) \in \Pi_n(q)} L_n(\sigma^2, \gamma), \tag{3.10}$$

where, following Da and Xiu (2021), the parameter space $\Pi_n(q)$ is defined as

$$\Pi_n(q) = \Big\{(\sigma^2, \gamma) \in \mathbb{R}^{q+2} : \inf_{\lambda} f(\lambda; \sigma^2, \gamma, \Delta_n) \geq \frac{\Delta_n}{K}, \quad \sigma^2 + |\gamma_0| + \frac{\sum_{j=1}^{\infty} j^2 |\gamma_j|}{\inf_{\lambda} |\sigma^2 \Delta_n + f(\lambda; \gamma)|} \leq K \Big\}. \tag{3.11}$$

Here $f(\lambda; \sigma^2, \gamma, \Delta_n)$ stands for the spectral density of Y_n under the quasi-model: $f(\lambda; \sigma^2, \gamma, \Delta_n) = \sigma^2 \Delta_n + (2 - 2\cos\lambda) f(\lambda; \gamma)$, with $f(\lambda; \gamma) = \sum_{j=-\infty}^{\infty} \gamma_{|j|} e^{\mathrm{i}j\lambda}$.

To determine an appropriate order q, we use information criteria, such as BIC, which in our setting can be written as

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

Our choice of order q will be based on

$$\widehat{q}_n = \arg \min_{q \leq n_T^{1/3}} \mathrm{BIC}_n(q). \tag{3.12}$$

We can define a similar criterion based on AIC, by replacing $q \log n_T$ above by $2q$. Hannan (1980) shows that using BIC results in consistent order selection for ARMA models. We demonstrate that a similar result with BIC also holds in our setting. We will therefore focus on BIC in the following discussion.

3.2 Implementation

We implement the exact likelihood via an auxiliary reduced-form MA($q+1$) model of the observed noisy returns:

$$Y_{n,i} = \phi(B)\epsilon_i, \quad \text{with} \quad \phi(x) = 1 + \sum_{j=1}^{q+1} \phi_j x^j, \quad 1 \le i \le n, \quad \epsilon \sim \mathcal{N}(0, \chi^2). \tag{3.13}$$

Algorithm 1. *Our algorithm starts as follows:*

1. *Select the optimal order, $\widehat{q}_n$, of the MA process (3.13) for Y_n using BIC, defined by (3.12) but rewritten equivalently 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 (3.13) and Kalman filtering,*

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

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

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

4. *Solve $\widehat{q}_n + 1$ nonlinear equations for $\widehat{q}_n + 1$ model parameters $(\widehat{\iota}_n^2(\widehat{q}_n), \widehat{\theta}_n(\widehat{q}_n))$ from $\widehat{\gamma}_n(\widehat{q}_n)$ obtained in Step 3:*

$$\widehat{\gamma}_n(\widehat{q}_n)_j = \widehat{\iota}_n^2(\widehat{q}_n) \sum_{l=0}^{\widehat{q}_n - j} \widehat{\theta}_n(\widehat{q}_n)_l \widehat{\theta}_n(\widehat{q}_n)_{l+j}, \quad 0 \le j \le \widehat{q}_n. \tag{3.14}$$

   A Newton-Raphson algorithm that converges quadratically is available from Wilson (1969).

Effectively, Step 4 is to find $\widehat{q}_n + 1$ model parameters of the MA($\widehat{q}_n$) noise process from up-to-$\widehat{q}_n$th-order autocovariances $\widehat{\gamma}_n(\widehat{q}_n)_j$, $0 \le j \le \widehat{q}_n$. This practice is common in the classic time-series analysis. For instance, Box et al. (2007) recommend using this algorithm to find initial values based on autocovariances for the maximum likelihood estimation of an MA model.

Step 3 is sufficient for volatility and noise autocovariance estimation, and it is rather simple to implement. If one is further interested in (ι^2, θ), a unique solution $(\widehat{\iota}_n^2(q), \widehat{\theta}_n(q))$ exists from Step 4, with probability approaching 1 when noise is sufficiently large relative to the sample size. When noise is small, however, these parameters are weakly identified, and (3.14) may have no solution such that $\widehat{\iota}_n^2(q)$ is positive and $\widehat{\theta}_n(q)$ is real.

3.3 Model Selection Consistency

We now discuss the asymptotic properties of the proposed estimators. The asymptotic analysis here is more involved than the classic time-series analysis, because the DGP of observed returns is misspecified. Moreover, the asymptotic design is in-fill, so that not only the dimensions, but also the entries of the covariance matrix Σ_n in the quasi-likelihood, depend on the sample size n_T; see (3.7). Consequently, prior results from classic time-series studies are not applicable. Even worse, the quasi-likelihood estimator does not have an explicit form.

We start with a model selection consistency result based on BIC, which allows us to conduct pointwise inference on autocovariance parameters. We thereby impose a finite-order moving-average model for the DGP of noise. In an in-fill asymptotic experiment, imposing a finite-order MA model for noise independent of the sampling frequency might appear ambiguous, in that observations are filled in between adjacent ones and the dependence structure changes as the sampling frequency approaches 0. However, as Jacod et al. (2017) argue, the frequency of observations in practice is fixed by the available data and does not really go to 0. Therefore, the interpretation of the asymptotic design is that the frequency of our observations is "high enough" to consider that we are "almost" in the asymptotic regime.

Theorem 1. *Suppose Assumptions 1 - 4 hold. We further assume a non-vanishing noise process with an exact MA($q^\star$) structure, i.e., $\iota^{(n)} \geq K^{-1}$ and $\theta^{(n)} \in \mathbb{R}^{q^\star}$ for all $n \geq 1$ and $\sqrt{n}(\log n)^{-1}|\theta_{q^\star}^{(n)}| \to \infty$, for some fixed $q^\star \geq 0$. Then it holds that*

$$\lim_{n\to\infty} \mathbb{P}(\widehat{q}_n = q^\star) = 1.$$

As the sample size increases, the likelihood is asymptotically dominated by that of the noise component. Therefore, the same intuition from the classic time-series result applies here. The likelihood estimator effectively minimizes the Kullback-Leibler divergence, but only when the selected order is no smaller than the truth. Moreover, the BIC imposes a penalty just large enough to rule out orders that are greater than the truth asymptotically. The combination of these two results leads to the desired consistency in model selection.

3.4 Inference on Noise Autocovariances and Autocorrelations

Recall that in (3.10) and Step 3 of Algorithm 1, we defined and implemented estimators of noise autocovariances. We now propose estimators of autocorrelations, denoted by $\widehat{\rho}_n(\widehat{q}_n)$, which are defined as follows.

If (3.14) has a solution such that $\widehat{\iota}_n^2(\widehat{q}_n)$ is positive and $\widehat{\theta}_n(\widehat{q}_n)$ is real, we set[2]

$$\widehat{\rho}_n(\widehat{q}_n)_j = \frac{\widehat{\gamma}_n(\widehat{q}_n)_j}{\widehat{\gamma}_n(\widehat{q}_n)_0}, \quad j \geq 1.$$

Otherwise, we set

$$\widehat{\rho}_n(\widehat{q}_n) = 0.$$

In light of their definitions, we can regard these estimators as "hard-thresholding" estimators, in that higher-order autocovariance and autocorrelation estimates are truncated to zero beyond the selected order $\widehat{q}_n$.

Next, we prove the pointwise central limit theorem for estimators of noise autocovariances in the finite-order moving average model. The corresponding result for autocorrelations follows straightforwardly.

Theorem 2. *Suppose Assumptions 1 - 4 hold. We further assume $\iota^{(n)} \geq K^{-1}$ and $\theta^{(n)} \in \mathbb{R}^{q^\star}$ for all $n \geq 1$ and some fixed $q^\star \geq 0$. Let $\gamma^{(n)}$ be the $(q^\star+1)$-dimensional vector of up-to-$q^\star$th-order autocovariances of U, whose components are defined in equation (2.4).[3] Assume there exists a $(q^\star+1)$-dimensional vector $\gamma^\star$ such that $\gamma^{(n)} - \gamma^\star = o_{\mathrm{P}}(1)$. Then it holds that[4]*

$$n^{1/2}\left(\widehat{\gamma}_n(q^\star) - \gamma^{(n)}\right) \xrightarrow{\mathcal{L}_s - \mathcal{F}_\infty} \mathcal{MN}\left(0_{q^\star+1}, \mathrm{AVAR}_1\right),$$

where

$$\mathrm{AVAR}_1 = \left(2W(\gamma^\star)^{-1} + \gamma^\star\gamma^{\star\intercal} cum_4(\varepsilon)\right)\frac{T\int_0^T \eta_s^4\xi_s^{-1}ds}{\left(\int_0^T \eta_s^2\xi_s^{-1}ds\right)^2},$$

$cum_4(\varepsilon)$ denotes the fourth cumulant of ε,

$$W(\gamma) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial \log f(\lambda;\gamma)}{\partial\gamma}\right)^\intercal \frac{\partial \log f(\lambda;\gamma)}{\partial\gamma}d\lambda.$$

[2]Estimates of autocovariances and autocorrelations are, of course, zero beyond the $\widehat{q}_n$-th lag.

[3]Recall that the vectors $\gamma^{(n)}$ and $\gamma^\star$ are indexed from 0. We refer to $\gamma^{(n)}$ here as a $(q^\star+1)$-dimensional vector simply because $\gamma_j^{(n)} = 0$ for all $j > q^\star$, since $\theta^{(n)} \in \mathbb{R}^{q^\star}$. For this reason, in most of our discussions, we do not distinguish it from an ∞-dimensional vector. The same applies to other ∞-dimensional vectors.

[4]Here and throughout the appendix, $\xrightarrow{\mathcal{L}_s - \mathcal{F}_\infty}$ stands for stable convergence in law with respect to $\mathcal{F}_\infty$.

This result shows that our estimator achieves the best convergence rate possible—$n^{1/2}$. In addition, the nonparametric estimation of volatility, which serves as a nuisance parameter here, does not influence the asymptotic variance of noise parameters. In fact, the asymptotic variance has the same form as in the classic time-series analysis–e.g., Brockwell and Davis (1991)–barring η and ξ terms, which are irrelevant in discrete time settings, as if the observed prices were purely made of noise. This further suggests that when ε indeed follows a Gaussian distribution, our estimator achieves the optimal efficiency.

The next corollary presents the central limit result for autocorrelations:

Corollary 1. *Suppose the same assumptions as those in Theorem 2 hold. Let $\rho^{(n)}$ be the $q^\star$ vector of up-to-$q^\star$th-order autocorrelations of U whose components are defined in equation (2.4). Then it holds that*

$$n^{1/2}\left(\widehat{\rho}_n(q^\star) - \rho^{(n)}\right) \xrightarrow{\mathcal{L}_s-\mathcal{F}_\infty} \mathcal{MN}\left(0_{q^\star}, \mathrm{AVAR}_2\right),$$

where the ijth entry of the $q^\star \times q^\star$ matrix $AVAR_2$ is given by

$$(\mathrm{AVAR}_2)_{ij} = \frac{\gamma_i^\star\gamma_j^\star}{\gamma_0^{4\star}}(\mathrm{AVAR}_1)_{11} + \frac{1}{\gamma_0^{2\star}}(\mathrm{AVAR}_1)_{i+1,j+1} - \frac{\gamma_i^\star}{\gamma_0^{3\star}}(\mathrm{AVAR}_1)_{1,j+1} - \frac{\gamma_j^\star}{\gamma_0^{3\star}}(\mathrm{AVAR}_1)_{1,i+1}.$$

Next, we construct an estimator of the asymptotic variance, AVAR_1, in Theorem 2, which naturally leads to an estimator for AVAR_2 in Corollary 1.

Proposition 1. *Suppose the same assumptions as those in Theorem 2 hold. Define*

$$\widehat{\mathrm{AVAR}}_1 = \left(2W(\widehat{\gamma}_n(\widehat{q}_n))^{-1} + \widehat{\gamma}_n(\widehat{q}_n)\widehat{\gamma}_n(\widehat{q}_n)^\intercal \widehat{cum}_4(\varepsilon)\right)\left(\widehat{\gamma}_n(\widehat{q}_n)_0 - \widehat{\gamma}_n(\widehat{q}_n)_1\right)^{-2}\widehat{B}_n,$$

where, with $k_n \sim \log n$,

$$\widehat{cum}_4(\varepsilon) = k_n\widehat{B}_n^{-1}\widehat{B}_n' - 2k_n - \left(\widehat{\gamma}_n(\widehat{q}_n)_0 - \widehat{\gamma}_n(\widehat{q}_n)_1\right)^{-2}\frac{1}{\pi}\int_{-\pi}^{\pi} f(\lambda;\widehat{\gamma}_n(\widehat{q}_n))^2(1-\cos\lambda)^2 d\lambda,$$

$$\widehat{B}_n = \frac{1}{4n_T k_n}\sum_{i=1}^{n_T-2k_n} Y_{n,i}^2 \sum_{j=k_n+1}^{2k_n} Y_{n,i+j}^2, \quad \text{and} \quad \widehat{B}_n' = \frac{1}{4n_T k_n}\sum_{i=1+k_n}^{n_T-k_n} Y_{n,i}^2 \sum_{j=-k_n}^{k_n} Y_{n,i+j}^2.$$

Then, we have

$$\left\|\widehat{\mathrm{AVAR}}_1 - n_T n^{-1}\mathrm{AVAR}_1\right\| = o_{\mathrm{P}}(1).$$

With this proposition in place, we can build confidence intervals for noise autocovariances and autocorrelations using $n_T^{-1}\widehat{\mathrm{AVAR}}_1$, which does not involve the unobservable scalar n in the CLT.

3.5 Uniform Consistency of Noise Autocovariances and Autocorrelations

The asymptotic inference established here is pointwise, in the sense that it does not allow for model-selection mistakes. As pointed out by Leeb and Pötscher (2005), model selection errors matter in finite samples, to the extent that the prescribed asymptotic distribution could be seriously distorted. Moreover, uniformly valid inference is generally not available.

That said, we establish a uniform consistency result for $\widehat{\gamma}_n(\widehat{q}_n)$ and $\widehat{\rho}_n(\widehat{q}_n)$ with respect to $\gamma^{(n)}$ and $\rho^{(n)}$ under $\mathbb{L}^2$-norm, where all vectors are regarded as ∞-dimensional. This result sheds light on the asymptotic behavior of these estimators when noise DGPs are allowed to vary within a larger class beyond MA(q), allowing for a vanishing magnitude and a more flexible dependence structure. We characterize the class of noise DGPs we consider in the next assumption.

Assumption 5. *Define* $q_n^\star(k) := \min q$*, subject to* $n\psi_n^4 \sum_{j=q}^{2q} |\widetilde{\kappa}_j^{(n)}|^2 \leq kq \log n$*, where* $\psi_n := (1 + n^{-1/2}/\iota^{(n)})^{-1}$ *and* $\widetilde{\kappa}_j^{(n)} := \sum_{i=0}^{\infty}(i+1)\psi_n^i(2\kappa_{j+1+i}^{(n)} - \kappa_{j+i+2}^{(n)} - \kappa_{j+i}^{(n)})$*. We assume for any* $0 < k < K$*,*

$$q_n^\star(k) = o(n^{1/3}(\iota^{(n)} \vee n^{-1/2})^{4/9}), \quad \textit{and} \quad n\psi_n^4 \sum_{j=q_n^\star(k)}^{\infty} |\kappa_j^{(n)}|^2 = O\big(q_n^\star(k) \log n\big).$$

Intuitively, $q_n^\star(k)$ mimics the "oracle" order that BIC selects. Effectively, Assumption 5 requires that this order cannot be too large and imposes an upper bound on the approximation error induced by a selected MA model. Nevertheless, these conditions in Assumption 5 are not restrictive. They accommodate common processes such as MA(∞), with $|\kappa_j^{(n)}| \sim j^{-\alpha}$ for some $\alpha > 3 \vee \frac{3}{2+4\log \iota^{(n)}/\log n}$, as well as any finite order ARMA(p,q) with an arbitrarily shrinking noise magnitude $\iota^{(n)} \lesssim 1$.

We are now ready to present the uniform consistency result for autocovariances and autocorrelations:

Theorem 3. *For any sequence of DGPs that satisfies Assumptions 1 - 5, we have*

$$\big\|\widehat{\gamma}^{(n)}(\widehat{q}_n) - \gamma^{(n)}\big\|^2 = O_{\mathrm{P}}\Big(n^{-1}(\iota^{(n)})^4(\widehat{q}_n+1)^2\log n + n^{-3}(n^{1/2}\iota^{(n)}+1)(\widehat{q}_n+1)^4\log n\Big).$$

If, in addition, we assume $\iota^{(n)} \geq Kn^{-2/3}(\log n)^{1/4}$*, it holds that*

$$\big\|\widehat{\rho}_n(\widehat{q}_n) - \rho^{(n)}\big\|^2 = O_{\mathrm{P}}\Big((\iota^{(n)})^{-4}\big\|\widehat{\gamma}^{(n)}(\widehat{q}_n) - \gamma^{(n)}\big\|^2\Big).$$

In general, the autocorrelation $\rho^{(n)}$ is weakly identified in the presence of small noise. The last part of Theorem 3 rules out this scenario, restricting the class of DGPs such that the noise variance cannot be too small.

Whereas consistent estimation of autocorrelations requires a more restrictive class of DGPs, Theorem 3 allows for arbitrarily small and vanishing noise for autocovariances. The case of small noise is highly relevant in practice, as shown from our empirical study below. Our result is complementary to the asymptotic theory developed by Jacod et al. (2017) and Li and Linton (2021), who focus on the case of non-vanishing noise.

3.6 Quadratic Representation

The QMLE estimator appears to have a rather different structure compared with alternative nonparametric estimators in the literature, e.g., realized kernels, which can be regarded as quadratic estimators. In this section, we propose an alternative but equivalent quadratic form of the QMLE, which sheds light on its connection with and distinction from these quadratic estimators. We do so for both volatility and noise autocovariance estimators.

Theorem 4. *Suppose the same assumptions as those in Theorem 2 hold and that $\gamma^{(n)} = \gamma^\star$. The QMLE $(\widehat{\sigma}_n^2(q^\star), \widehat{\gamma}_n(q^\star))$ satisfies that for $0 \le j \le q^\star$,*

$$\widehat{\sigma}_n^2(q^\star) = Y_n^\intercal \mathcal{W}_n(\widehat{\sigma}_n^2(q^\star), \widehat{\gamma}_n(q^\star); 1) Y_n, \quad \widehat{\gamma}_n(q^\star)_j = Y_n^\intercal \mathcal{W}_n(\widehat{\sigma}_n^2(q^\star), \widehat{\gamma}_n(q^\star); j+2) Y_n, \tag{3.15}$$

where the set of $n_T \times n_T$ weighting matrices $\mathcal{W}_n(\sigma^2, \gamma; l)$, $l = 1, 2, \ldots, q^\star + 2$, is defined by[5]

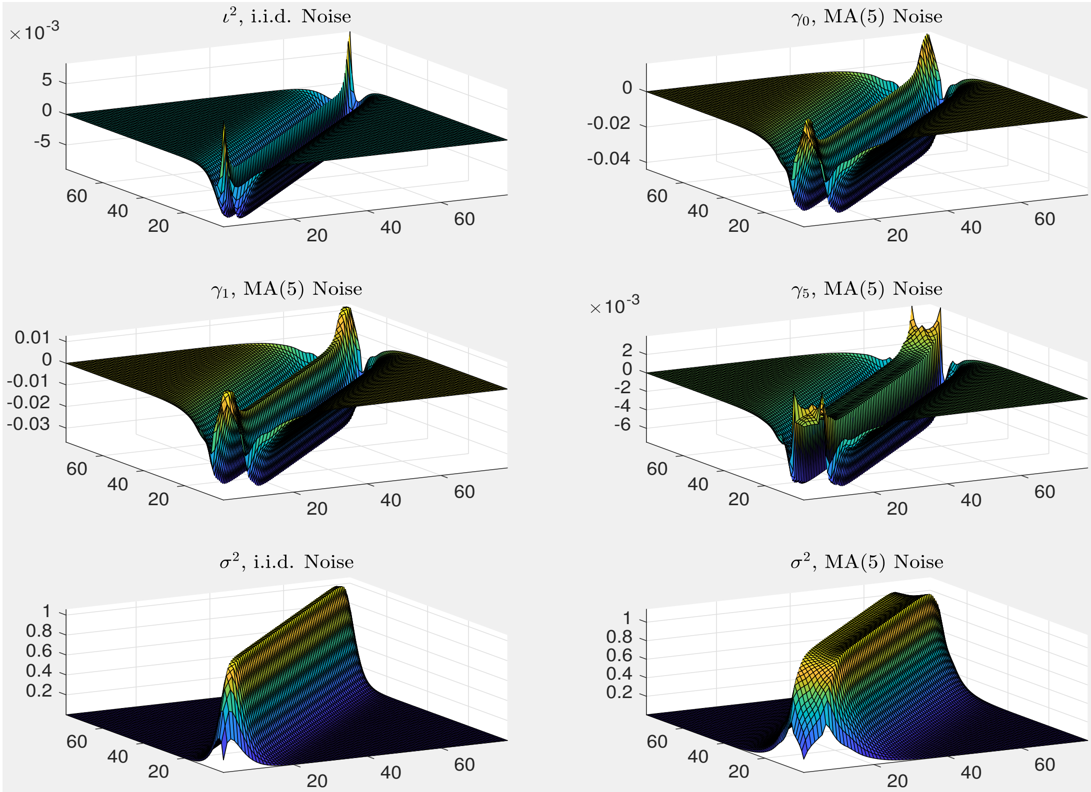
$$vec(\mathcal{W}_n(\sigma^2, \gamma; l)) = \Sigma_n^{-1}(\sigma^2, \gamma) \frac{\partial \Sigma_n(\sigma^2, \gamma)}{\partial(\sigma^2, \gamma)} \Sigma_n^{-1}(\sigma^2, \gamma) \widetilde{W}_n^{-1}(\sigma^2, \gamma)(0_{l-1}, 1, 0_{q^\star+2-l}),$$

with $\Sigma_n(\sigma^2, \gamma)$ given by (3.7), and the $(q^\star + 2) \times (q^\star + 2)$ matrix $\widetilde{W}_n(\sigma^2, \gamma)$ given by

$$\widetilde{W}_n(\sigma^2, \gamma)_{i,j} = tr\left(\Sigma_n^{-1}(\sigma^2, \gamma) \frac{\partial \Sigma_n(\sigma^2, \gamma)}{\partial(\sigma^2, \gamma)_i} \Sigma_n^{-1}(\sigma^2, \gamma) \frac{\partial \Sigma_n(\sigma^2, \gamma)}{\partial(\sigma^2, \gamma)_j} \right).$$

Theorem 4 shows that the QMLE can be written as an iterative quadratic estimator. It also suggests an alternative algorithm for estimation. With some initial values given, we can iteratively update parameters via equations given by (3.15) until convergence. Figure 1 plots these weighting matrices for both volatility and noise parameters, and compares them in the case of i.i.d. and MA(5) noises. The noise weighting matrices feature a "W" shape along the diagonal, and the magnitude of weighting matrices for autocovariance decays as their order increases. With respect to the volatility estimator, the bottom panel shows notable "flatness"

[5] 0_d is the d-dimensional vector of 0s. All vectors are column vectors. We write (a, b, c) in place of $(a^\intercal, b^\intercal, c^\intercal)$ for simplicity.



Note: This figure compares weighting matrices $\mathcal{W}$s in the quadratic representations of the QMLE for σ^2 and ι^2 in the case of i.i.d. noise, as well as those matrices for σ^2, γ_0, γ_1, and γ_5 in the case of MA(5) noise. We scale the volatility weighting matrices by T. In both cases, we fix $\sigma^\star = 0.3$, $\iota^\star = 0.005$, $\Delta = 5$ minutes, $T = 1$ day. The moving-average parameters of the MA(5) process are given by $\theta^\star = (0.25, 0.2, 0.15, 0.1, 0.05)$.

Figure 1: Quadratic Representations of the Estimators

at the top of the volatility weighting matrix for the MA(5) model, which helps cancel out the impact of dependent noise. This patten motivates us to investigate the connection between the QMLE and the flat-top realized kernel introduced by Varneskov (2016) to the high-frequency environment in the context of volatility estimation. We also provide an equivalent kernel for autocovariances.

Theorem 5. *Suppose the same assumptions as those in Theorem 2 hold. In addition, suppose* $q \geq 0$ *is fixed and* $(\sigma^2, \gamma) \in \Pi_n(q)$ *such that* $K^{-1} \leq \inf_\lambda f(\lambda;\gamma) \leq \sup_\lambda f(\lambda;\gamma) \leq K$. *Then for all* $n^{1/2+\alpha} \leq i, j \leq n - n^{1/2+\alpha}$ *with* $0 < a < \frac{1}{2}$, *the weighting matrix* $\mathcal{W}_n(\sigma^2, \gamma; l)$ *satisfies for* $l \geq 1$,

$$(i) \quad \mathcal{W}_n(\sigma^2,\gamma;1)_{i,j} = T^{-1}k\big(H_n^{-1}|i-j|\big)(1+o(1)), \quad \mathcal{W}_n(\sigma^2,\gamma;l)_{i,j} = \lambda_l \widetilde{k}\big(H_n^{-1}|i-j|\big) + O(1);$$

$$(ii) \quad \sup_{|i-j|\leq q+1} \left| \mathcal{W}_n(\sigma^2,\gamma;1)_{i,j} - \mathcal{W}_n(\sigma^2,\gamma;1)_{i,i} \right| = O(\Delta_n^{3/2});$$

$$(iii) \quad \sup_{|i-j|\leq q+1} \left| \mathcal{W}_n(\sigma^2,\gamma;l)_{i,j} + \mathbb{1}_{\{l\leq|i-j|+1\}} \frac{|i-j|+2-l}{2n_T} \right| = O(\Delta_n^{3/2}),$$

where the implied equivalent kernels are $k(x) = (1+x)e^{-x}$ *and* $\widetilde{k}(x) = xe^{-x}$*, the implied bandwidth is* $H_n = \zeta\sigma^{-1}\Delta_n^{-1/2} + O(1)$ *with* $\zeta^2 = \sum_{|j|\leq q}\gamma_{|j|}$*, and* $\lambda_l = (2\sigma\zeta^3\Delta_n^{1/2}n_T)^{-1}\sum_{r=1}^{q+1}(2-\delta_{r,1})W(\gamma)^{-1}_{r,l-1}$*, with* $W(\gamma)$ *defined in Theorem 2.*

Theorem 5 suggests that the bulk of the QMLE weighting matrices can be approximately written as that of a nonparametric kernel estimator with an implicit bandwidth. Despite this equivalence, it is more convenient to implement the QMLE using Algorithm 1 in Section 3.2, which does not require tuning parameters barring order selection, or any special adjustment to the border effect. Also note that this equivalence result is only established under the assumption that the spectral density of the noise (and hence its magnitude) is bounded from below, which rules out the case of small noise.

4 Monte Carlo Simulations

We examine the finite-sample performance of the estimators in a variety of simulation settings. Throughout we fix $T = 1$ day and the average sampling frequency every 5 seconds. We have $1,000$ Monte Carlo trials in total.

4.1 Verification of the Asymptotic Results

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

$$\begin{cases} dX_t &= (0.05+0.5\sigma_t^2)dt + \sigma_t dW_t + J^X dN_t, \\ \sigma_t^2 &= D_t \exp\left(-2.8+6F_t\right), \quad dF_t = -4F_t dt + 0.8d\widetilde{W}_t + J^F dN_t - 0.02\lambda_N dt, \end{cases} \tag{4.16}$$

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 the noise, we start with an MA(5) model of U with $\theta^\star = (0.25, 0.2, 0.15, 0.1, 0.05)$, innovation ε_i being Student's t-distribution with 7 degrees of freedom, $\iota = 2.5 \times 10^{-3}$, and η_t following

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

where W_t is the same Brownian motion that drives X. We also round the observed prices to the nearest cent: $\tilde{X}_t = \log\left([100 \times \exp(X_t)]\right) - \log 100$, where $[\cdot]$ means rounding to the nearest integer.[6]

We first assume that the correct order, namely 5, is known, so that we can verify the CLTs for noise autocovariances given in Section 3.4 without worrying about model selection mistakes. Figure 2 provides the histograms of the standardized estimates for $\widehat{\gamma}_k(q)$, $k = 0, 2, \ldots, 5$, using estimated asymptotic variances. All histograms match the standard normal density.

4.2 Comparison with Alternative Estimators

We then compare our estimators of noise autocorrelations against alternative nonparametric estimators by Jacod et al. (2017) (JLZ) and Li and Linton (2021) (ReMeDI) in a more challenging MA(∞) setting in which $\theta(B) = (1 - 0.4B)^{-1}(1 + 0.2B)$. To demonstrate the effect of small noise, we consider three different scenarios for the magnitude of the noise, ι, which takes values from 10^{-4} (small noise) to 5×10^{-4} (median noise) and 2.5×10^{-3} (large noise). Our estimator uses either AIC or BIC for model selection, whereas nonparametric estimators involve a tuning parameter.

Jacod et al. (2017) propose to estimate autocovariances, γ, by approximating efficient prices using their local averages:

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

Here h_n is a sequence of integers satisfying $h_n \sim n^{-\eta}$ with $\frac{1}{2v+1} < \eta < \frac{1}{2}$, where v is the ρ-mixing exponent of ε. It determines the local window size used to estimate realization of the noise. Their paper selects $h_n = 6$ 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 report the autocorrelation estimates for $h_n = 2, 4$, and 6.

[6]Our theory does not allow for this type of rounding errors. We simulate this model to demonstrate that the rounding effect appears negligible.

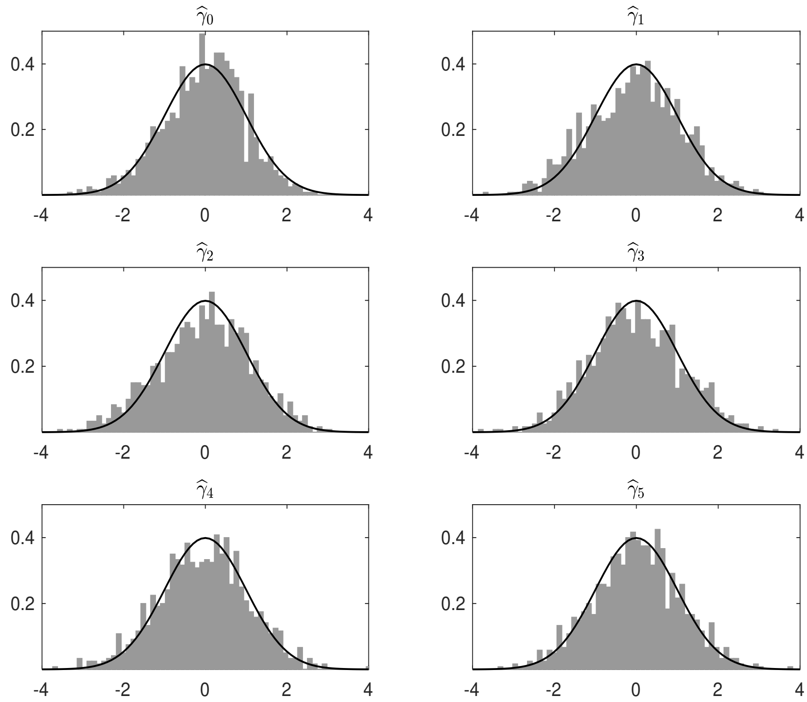


Figure 2: Histograms of the Standardized Parameter Estimates

Note: This figure plots the histograms of the standardized estimates for $\widehat{\gamma}_k(q)$, $k = 0, 1, \ldots, 5$, along with the density of the standard normal distribution. The noise is simulated from an MA(5) model with $\theta^\star = (0.25, 0.2, 0.15, 0.1, 0.05)$ and $\iota^\star = 2.5 \times 10^{-3}$. The order of the MA model is known prior to estimation.

Li and Linton (2021) suggest an alternative construction that takes the differences of log prices over longer horizons to dampen the impact of efficient prices:

$$\widehat{\gamma}_j^{\text{ReMeDI}} = -\frac{1}{n_T} \sum_{i=1}^{n_T - 2k_n - j} (\widetilde{X}_{i+k_n} - \widetilde{X}_i)(\widetilde{X}_{i+j+2k_n} - \widetilde{X}_{i+j+k_n}),$$

where k_n is a tuning parameter that satisfies: $k_n \to \infty$, $k_n n^{-\eta} \to 0$, for $\frac{1}{2v} < \eta < \frac{1}{3}$. We select $k_n = k_n' \log n$, where $k_n' = 0.5, 1$, and 2 in simulations.

With autocovariances given, the autocorrelations can thereby be estimated accordingly: $\widehat{\rho}_j^{\text{JLZ}} = \widehat{\gamma}_j^{\text{JLZ}} / \widehat{\gamma}_0^{\text{JLZ}}$ and $\widehat{\rho}_j^{\text{ReMeDI}} = \widehat{\gamma}_j^{\text{ReMeDI}} / \widehat{\gamma}_0^{\text{ReMeDI}}$. We prefer autocorrelations (to autocovariances) because their scale is interpretable. However, we find it necessary to winsorize the estimated autocorrelations for AIC-based QMLE and both nonparametric estimators, when the noise magnitude is small, to ensure that their estimates are within the natural bound $[-1, 1]$.[7]

Table 1: Simulation Results for Noise Autocorrelation Estimation

		QMLE	QMLE	JLZ	JLZ	JLZ	ReMeDI	ReMeDI	ReMeDI
		BIC	AIC	$h_n = 2$	$h_n = 4$	$h_n = 6$	$k'_n = 0.5$	$k'_n = 1$	$k'_n = 2$
		Panel A: Small Noise							
ρ_1	BIAS	-0.309	-0.195	0.586	0.639	0.657	-0.082	0.224	0.387
	RMSE	0.310	0.432	0.587	0.639	0.657	0.577	0.603	0.617
ρ_3	BIAS	-0.163	-0.101	0.716	0.775	0.795	0.008	0.315	0.477
	RMSE	0.163	0.327	0.718	0.775	0.795	0.600	0.691	0.723
ρ_5	BIAS	-0.044	-0.027	0.821	0.885	0.906	0.048	0.369	0.536
	RMSE	0.044	0.242	0.824	0.886	0.906	0.633	0.757	0.801
		Panel B: Median Noise							
ρ_1	BIAS	-0.093	-0.017	0.153	0.300	0.379	-0.006	-0.016	-0.035
	RMSE	0.150	0.075	0.185	0.312	0.386	0.098	0.150	0.251
ρ_3	BIAS	-0.063	-0.010	0.188	0.364	0.459	0.000	-0.012	-0.034
	RMSE	0.094	0.054	0.227	0.379	0.468	0.111	0.170	0.280
ρ_5	BIAS	-0.029	-0.004	0.217	0.416	0.524	0.003	-0.002	-0.033
	RMSE	0.039	0.036	0.262	0.434	0.535	0.115	0.186	0.300
		Panel C: Large Noise							
ρ_1	BIAS	-0.009	-0.001	-0.055	0.010	0.045	0.000	-0.003	-0.001
	RMSE	0.040	0.020	0.073	0.063	0.083	0.036	0.039	0.043
ρ_3	BIAS	-0.008	0.000	-0.066	0.012	0.055	0.001	-0.002	-0.001
	RMSE	0.029	0.019	0.088	0.075	0.100	0.042	0.043	0.048
ρ_5	BIAS	-0.011	0.000	-0.075	0.014	0.062	0.001	0.002	0.000
	RMSE	0.028	0.018	0.101	0.086	0.114	0.044	0.046	0.051

Note: This table compares estimators of 1st-, 3rd-, and 5th-order autocorrelations (ρ_1, ρ_3, ρ_5) in three scenarios of noise magnitude. "QMLE" is an MA($\widehat{q}_n$)-likelihood estimators using either BIC or AIC for order selection. "JLZ" refers to the nonparametric estimator of Jacod et al. (2017). "ReMeDI" refers to the nonparametric estimator of Li and Linton (2021). We report three choices of h_n and k'_n for comparison. The AIC-QMLE, JLZ, and ReMeDI estimates of autocorrelations are winsorized so that their magnitude stays within $[-1, 1]$. The true 1st, 3rd-, and 5th-order autocorrelations are 0.308, 0.163, and 0.04, respectively.

Table 1 provides comparison results for autocorrelations among QMLE, JLZ, and ReMeDI estimators across various noise magnitudes. Several points are worth making. For large noise, all estimators work reasonably well, but QMLEs generally outperform nonparametric estimators in terms of RMSE because they are more efficient. AIC slightly outperforms

[7] If a correlation estimate exceeds 1 (resp. -1), we reset it to be 1 (resp. -1).

BIC, and ReMeDI appears to outperform JLZ. The latter suffers from a large finite sample bias. In the small noise regime, nonetheless, the biases and RMSEs for both nonparametric estimators deteriorate substantially. For estimation of noise autocovariances, "signal' is the microstructure friction, whereas "noise" is the efficient price. When the signal-to-noise ratio is too low, the error due to estimation is too large to justify doing so. In contrast, the QMLEs either conclude that noise is absent (i.e., θ and ι^2 are not available), in which case all autocorrelations are zeros, or select an MA model with a certain $\widehat{q}_n$, so that any autocorrelation beyond the $\widehat{q}_n$-th order is zero. Because of the rapid decay in autocorrelations and small noise magnitude, 0 is often a better estimate in terms of RMSE than nonparametric estimates, and in particular for larger lags. Comparing AIC with BIC, the latter is more conservative, as it essentially yields 0 autocorrelation estimates for almost all Monte Carlo replications, whereas the former produces many nontrivial estimates. However, doing so seems to increase AIC's RMSE, and AIC does require winsorization for about 5.3% of sample paths, compared with 20.9% for ReMeDi and 4.0% for JLZ. BIC needs no adjustment.

5 Empirical Analysis of U.S. Equity

To demonstrate the empirical relevance of the proposed approach, we conduct a large-scale study of noise autocovariances for S&P 1500 index constituents from January 1, 1996, to December 31, 2016. There are approximately 1,500 tickers every day, and about 3,500 tickers in total due to changes in index constituents. To illustrate, we summarize cross-sectional findings here though all estimates are available upon request. We use BIC-QMLE for noise-related parameters because of the model selection consistency result discussed earlier. We also report volatility estimation results, but with AIC*-QMLE, as suggested by Da and Xiu (2021).

We download the trades and quotes of all equities at their highest frequency available (up to a millisecond after January 1, 2007, and a microsecond from July 27, 2015) from the TAQ database.[8] Next, we remove trades and quotes with special condition codes or suffix codes, as well as those that occur outside regular trading hours.[9] We then construct national best bid and offer (NBBO) data using quotes from all exchanges at a 1-second frequency.[10]

[8]Because companies change their tickers from time to time for mergers, acquisitions, or other reasons, the same ticker in the TAQ database may correspond to different stocks. We therefore keep track of these changes and use CRSP PERMINOs to index all stocks that do not change over time.

[9]We remove trades and quotes with condition codes Z, B, U, T, L, G, W, K, J and corresponding odd-lot trades, which have an additional letter I, as well as those with non-empty suffix codes (preferred shares). We identify opening trades as those with condition codes O, Q, OI, or QI; closing trades with 6, M, 6I, or MI; and remove all trades beyond the window of opening and closing time points. We only keep trades with correction indicator 00 or 01.

[10]We construct NBBOs from the millisecond dataset by adapting the SAS codes from `https://wrds-web.`

We then match trades with NBBOs by their recorded time points and remove those trades that are outside the range of the corresponding NBBOs.[11] Our approach is less aggressive than that of Barndorff-Nielsen et al. (2009), in that we maintain trades and quotes from all exchanges, whereas they retain only entries originating from a single exchange. Next, we remove redundant trades, retaining only nonzero returns.[12] This step helps alleviate model misspecification due, for example, to the effect of rounding, latency or delay across exchanges, and so on. Finally, we remove any stock days that have fewer than 12 observations after cleaning.

We start by examining the time-series behavior of volatility and microstructure noise. The upper panel of Figure 3 presents the time series of volatility estimates for constituents of each of the three indices, respectively. The lower panel provides the time series of noise-variance estimates among those constituents whose estimates are available. We use lines to represent the median and shaded areas to represent the lower and upper quartiles in the cross-section. We also smooth these time series using equal weights over a monthly moving window. Although considerable cross-sectional variation is present, the median volatility estimates among constituents of all three indices share a pattern similar to what we usually find from the volatility of the S&P 500 index. That said, the small caps are on average more volatile than the large caps, with the mid caps in between. As to the noise, there is a clear declining pattern in its order of magnitude over time across the entire universe, which is likely because of the improvement in market efficiency. Not surprisingly, the small caps have the largest noise, followed by the mid cap and then the large cap.

Next, we focus on the dependence structure of the noise. As the left panels of Figure 4 show, around 30%-60% of stocks have noise that is too small to be estimated. This percentage is higher for large caps than for small caps. For a large percentage of stock-day pairs, the selected orders based on the BIC are 0, so that i.i.d. noise assumption is reasonable for them. That said, about 10%-30% of stock-day pairs remain for which BIC prefers a few more lags. For BIC to select more than 6 lags is rare. We also find more stock-days in 2016 with selected orders greater than or equal to 1, compared with earlier years, particularly for large caps. This finding is due to the availability of data sampled at a frequency even higher than every second, for which we expect to see more autocorrelated lags.

To shed further light on this point, we provide in the right panels of Figure 4 histograms of the durations of autocorrelations for those tickers with selected lags greater than or equal to

---

`wharton.upenn.edu/wrds/research/applications/microstructure/NBBO%20derivation/`. Although this database has more precise timestamps, we do not construct NBBOs at any frequency higher than every second.

[11]For trades that are observed at millisecond or microsecond intervals, we match them with the NBBOs of the previous second. Our SAS codes for cleaning the data are available upon request.

[12]This step is called "tick-time sampling" by Griffin and Oomen (2008).

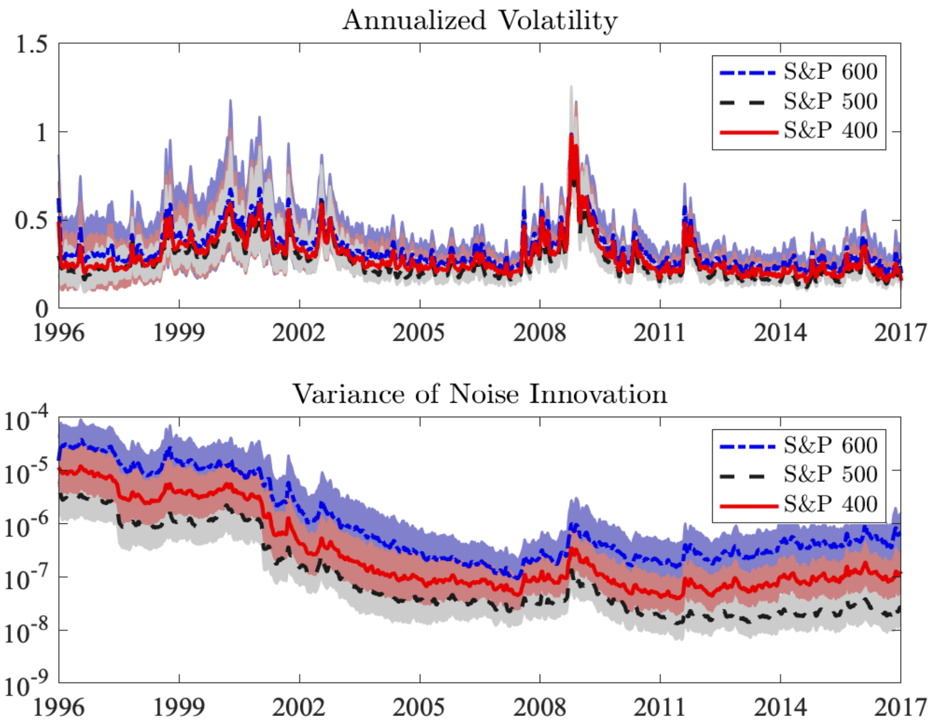


Figure 3: Time Series of Volatility and Noise-innovation Variance

Note: The upper panel compares the cross-sectional median (lines), lower, and upper quartiles (shaded areas) of the annualized volatility estimates for S&P Composite 1500 Index constituents (using Algorithm 1.3), and the lower panel presents the variance estimates of noise innovation (using Algorithm 1.4) for those constituents that have large-enough noise. The time series are smoothed with equal weights over a moving window of 21 days. The y-axis of the lower panel is transformed to the logarithm scale for the sake of presentation.

1. Duration is defined in terms of seconds as the product of the selected order and the average trading frequency for each stock-day pair. We find that estimated durations are much shorter for large-cap stocks than for smaller caps. Moreover, the average duration of autocorrelations has been decreasing in the past two decades. For instance, the average duration of large caps has decreased from $10^2 \sim 10^3$ to merely 10 seconds.

Finally, we discuss the importance of modeling the microstructure noise through the lens of volatility inference. While there exist informal volatility signature plot or more formal tests of microstructure noise (Aït-Sahalia and Xiu (2019)), such pre-testing-based approaches do not deliver correct volatility inference due to uniformity concerns when noise exists but is too small to be detected. We compare the biases and RMSEs of the popular realized volatility estimator

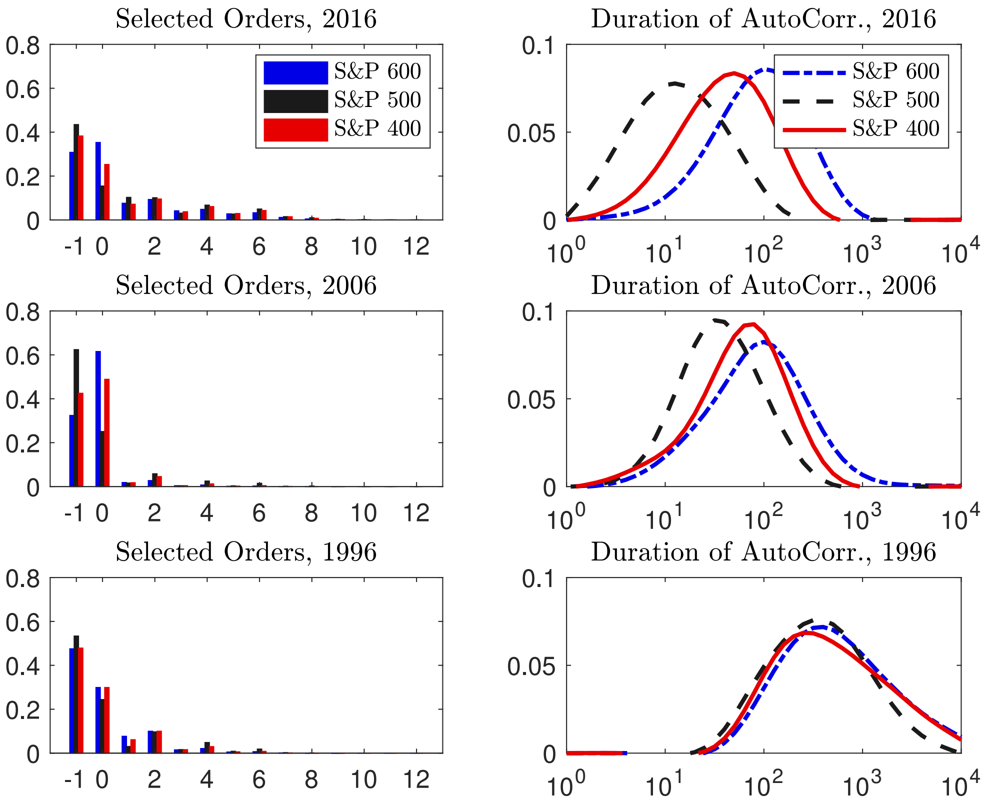


Figure 4: Selected Orders and Durations of Autocorrelations

Note: Left panels provide the frequencies of selected orders using BIC for each stock-day pair in 1996, 2006, and 2016, respectively. "-1" represents the case of small noise, i.e., the stock-day pair for which no reliable estimate of noise variance exists. "0" represents the case of i.i.d. noise, whereas other values are the selected orders of MA processes. Panels on the right provide the corresponding (fitted) histograms of the durations of autocorrelations in the case of dependent noise. Duration in terms of seconds is defined as the product of the selected order and the average trading frequency for each stock-day pair. The x-axis is transformed to a logarithmic scale for the sake of presentation.

and the QMLE, to indirectly shed light on the influence of noise. The former estimator, based on data sampled at a prespecified frequency—say, every 5 or 15 minutes—is most commonly adopted in practice.

The left panels of Figure 5 compare the cross-sectional medians of realized volatility estimates based on 5-minute and 15-minute subsamples, respectively, with the corresponding medians of the QMLEs. Remarkably, on average, a large upward bias associated with the former estimates is present, potentially due to the presence of noise at the 5-minute frequency. The biases are substantial–over 160% for small caps–compared with noise-robust QMLEs in earlier years. The biases have been decreasing over the past two decades, with a slight increase post-2008. Biases of the small caps are more evident than those of the large caps. On average, the large caps are traded more frequently than every 5 minutes, so their biases in the cross-sectional medians are almost indistinguishable from zero post-2002. This finding does not imply that every 5 minutes is a safe frequency for each individual constituent of the S&P 500 index. At a 15-minute frequency, the biases are clearly smaller–though they have not completely vanished, even in 2016–for these median estimates. The right panels of Figure 5 compare the ratios of standard errors between the 5-minute (resp. 15-minute) realized volatility estimator and the QMLE using the entire sample. The larger the ratio, the greater the efficiency loss for the realized volatility. We only report results for 2016, because the quality of the realized volatility estimator is best. We find that when the sampling frequency reaches every 15 minutes, most of the ratios are greater than 1, with some being as large as 10–in particular, for S&P 500 constituents–which suggests substantial efficiency losses.

To sum up, without accounting for noise, the realized volatility estimator faces a bias and variance dilemma. Estimates using 5-minute data are subject to severe biases, whereas 15-minute estimates suffer from considerable efficiency losses. Additionally, the standard errors could still be understated because the noise might not be sufficiently small to the extent that it can be safely ignored.

6 Conclusion

We propose a semiparametric approach to disentangling autocovariances and autocorrelations due to the microstructure frictions associated with observed prices. Our approach resembles a threshold estimator, which gives zero autocovariance estimates beyond the lag selected by the information criteria. This feature delivers superior performance in the finite sample, particularly when noise is relatively small, compared with alternative nonparametric estimators. Our empirical study of S&P 1500 stocks finds that the microstructure noise has shrunk by several orders of magnitude and that its autocovariances have faded more rapidly in recent years than earlier. These findings indicate that market efficiency has improved substantially, potentially due to the popularity of electronic and algorithmic trading. In a cross-sectional comparison, the autocovariances of small-cap stocks tend to persist for a longer period than the large caps, perhaps due to limits to arbitrage or for liquidity reasons.

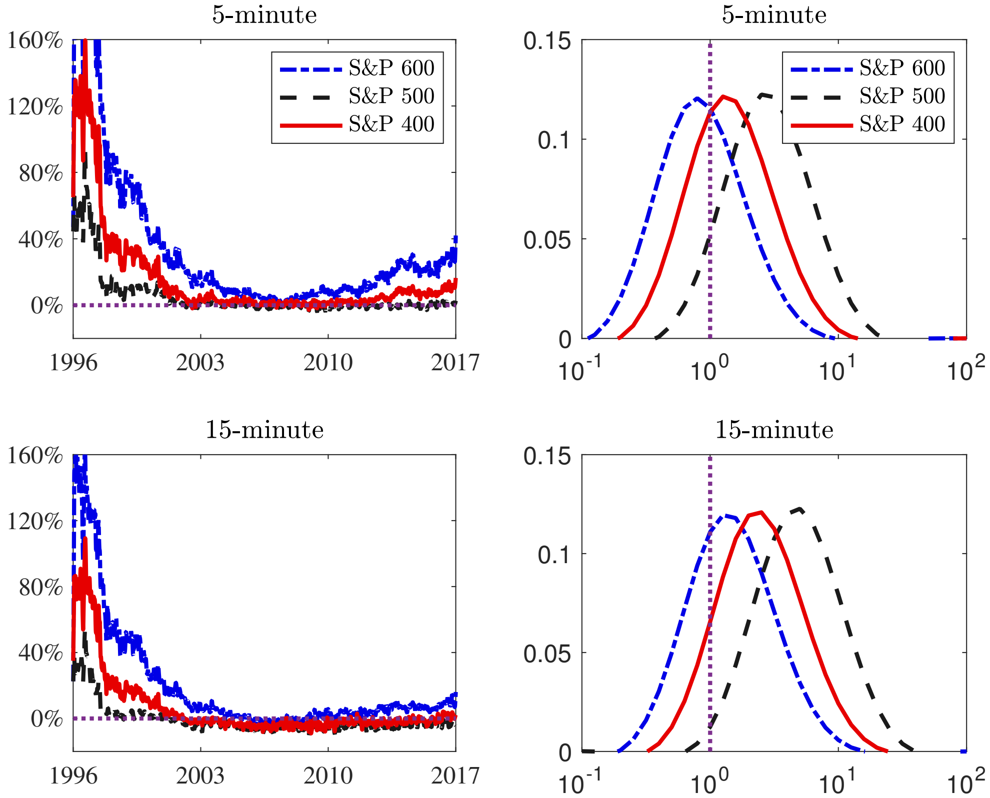


Figure 5: Relative Biases and Standard Errors of the Realized Volatility against QMLE

Note: The right panels plot percentage biases in the cross-sectional medians of 5-minute and 15-minute realized volatility estimates, respectively, relative to their corresponding QMLEs using the entire sample. Time series are smoothed with equal weights over a moving window of 21 days. The right panels provide the histograms of the ratios of standard errors between the 5-minute (resp. 15-minute) realized volatility estimator and the QMLE, for each stock-day pair in 2016. The x-axes on the right panels are transformed to the a logarithmic scale for the sake of presentation.

References

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.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics 160*, 160–175.

Aït-Sahalia, Y. and D. Xiu (2019). A Hausman test for the presence of market microstructure noise in high frequency data. *Journal of Econometrics 211*, 176–205.

Amihud, Y. and H. Mendelson (1980). Dealership market: Market-making and inventory. *Journal of Financial Economics 8*, 31–53.

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.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal 12*, 1–32.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2011). Subsampling realised kernels. *Journal of Econometrics 160*(1), 204–219.

Bertsimas, D. and A. W. Lo (1998). Optimal control of execution costs. *Journal of Financial Markets 1*, 1–50.

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.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (2007). *Time Series Analysis: Forecasting and Control* (4th ed.). Wiley.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (Second ed.). New York: Springer-Verlag.

Brogaard, J. A., T. Hendershott, and R. Riordan (2014). High frequency trading and price discovery. *Review of Financial Studies 27*, 2267–2306.

Chang, J., A. Delaigle, P. Hall, and C. Tang (2018). A frequency domain analysis of the error distribution from noisy high-frequency data. *Biometrika 105*(2), 353–369.

Da, R. and D. Xiu (2021). When moving-average models meet high-frequency data: Uniform inference on volatility. *Econometrica, forthcoming*.

Garbade, K. and Z. Lieber (1977). On the independence of transactions on the new york stock exchange. *Journal of Banking and Finance 1*(151-172).

Griffin, J. E. and R. C. Oomen (2008). Sampling returns for realized variance calculations: Tick time or transaction time? *Econometric Reviews 27*, 230–253.

Hannan, E. J. (1980). The estimation of the order of an arma process. *The Annals of Statistics*, 1071–1081.

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

Hasbrouck, J. and T. S. Y. Ho (1987). Order arrival, quote behavior, and the return-generating process. *Journal of Finance 42*(4), 1035–1048.

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.

Jacod, J., Y. Li, and X. Zheng (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.

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.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*, 1315–1336.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometic Theory 21*(1), 21–59.

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

Li, M. and O. Linton (2021). A remedi for microstructure noise. Technical report, Cambridge University.

Niederhoffer, V. and M. Osborne (1966). Market making and reversal on the stock exchange. *Journal of American Statistical Association 61*(316), 897–916.

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

Simmons, D. M. (1971). Common-stock transaction sequences and the random-walk model. *Operations Research 19*(4), 845–861.

Tóth, B., I. Palit, F. Lillo, and J. D. Farmer (2015). Why is equity order flow so persistent? *Journal of Economic Dynamics & Control 51*, 218–239.

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

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.

Web Supplement to

Disentangling Autocorrelated Intraday Returns

Rui Da[*] Dacheng Xiu[†]

Indiana University University of Chicago and NBER

This version: June 20, 2025

Abstract

This supplement contains mathematical proofs.

Appendix A Proofs of Technical Lemmas

A.1 Notation

In this section, we prepare 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. 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 $\mathbb{P}(d\omega_{(0)}, d\omega_{(1)}) = \mathbb{P}_{(0)}(d\omega_{(0)})\mathbb{P}_{(1)}(d\omega_{(1)})$.

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

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

[*]Address: 1275 E 10th St, Bloomington, IN 47405. E-mail: ruida@iu.edu.

[†]Address: 5807 S Woodlawn Ave, Chicago, IL 60637, USA. E-mail: dacheng.xiu@chicagobooth.edu.

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 $\Delta_i^n A$ 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 $\varepsilon_C(j)_i \coloneqq \varepsilon_{(j-1)n_d+i}$ for $i \geq 1$ and $\varepsilon_C(j)_i \coloneqq \tilde{\varepsilon}(j)_i$ 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 $i \geq 1$,

$$U^C(j)_i = \eta_C(j)\iota^{(n)}\theta^{(n)}(\mathrm{B})\varepsilon_C(j)_i,$$

and write $U^C(j) \coloneqq (U^C(j)_1, \ldots, U^C(j)_{n_d})^\intercal$.

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

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

Using $\Omega_m^U(j)$, we define $\Omega_n^U = \left(\bigoplus_{j=1}^{J_d} \Omega_{n_d}^U(j)\right) \oplus \Omega_{n_d'}^U(J_d+1)$. For any n, let $\Omega_n^B, \Omega_n^J, \Omega_n^Y, \Omega_n^{Y,B} \in \mathcal{M}_{n_T}$ be defined by

$$(\Omega_n^B)_{ij} = \delta_{i,j}\int_{t_{i-1}}^{t_i} \sigma_s^2 ds, \quad (\Omega_n^J)_{ij} = \delta_{i,j}\sum_{t_{i-1}<s\leq t_i} (\Delta X_s)^2, \quad \Omega_n^{Y,B} = \Omega_n^U + \Omega_n^B, \quad \Omega_n^Y = \Omega_n^{Y,B} + \Omega_n^J.$$

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

$$\Omega_n^{U,C}(j) = O_{n_d}(2\mathbb{I}_{n_d} - \mathbb{D}_{n_d}^1)D_{n_d}(\gamma_C(j))O_{n_d},$$

where D_{n_d} is defined in (A.1), whose dependence on $\gamma_C(j)$ is made explicit.

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

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

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

$$\begin{aligned}&\bar{L}_n^\star(\sigma^2,\gamma) = -\frac{n_T}{4\pi}\int_{-\pi}^{\pi}\left(\log f(\lambda;\sigma^2,\gamma,\Delta_n) + \frac{f(\lambda; C_T, \gamma^{(n)}, \Delta_n)}{f(\lambda;\sigma^2,\gamma,\Delta_n)}\right)d\lambda, \\ &\chi^2(\sigma^2,\gamma,\Delta_n) = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda;\sigma^2,\gamma,\Delta_n)d\lambda\right).\end{aligned}$$

Part 5. With any given n, (σ, γ), and q, we define

$$R_n(\sigma^2, \gamma) = |\sigma^2 - C_T| + \sup_\lambda |f(\lambda; \gamma) - f(\lambda; \gamma^{(n)})|,$$
$$\widehat{\mathcal{R}}_n(q) = R_n(\widehat{\sigma}_n^2(q), \widehat{\gamma}_n(q)), \quad \mathcal{R}^{(n)}(q) = R_n(\sigma^{(n)}(q)^2, \gamma^{(n)}(q)).$$

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 (3.11). We let

$$(\sigma^{(n)}(q)^2, \gamma^{(n)}(q)) = \arg \min_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q)} \bar{L}_n^\star(\sigma^2, \gamma).$$

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 $\partial\sigma_n^2 \coloneqq \partial\sigma_n^2(\beta)/\partial\beta$. 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 $S_n \in \{f(\lambda;\ \cdot\ ,\ \cdot\ , \Delta_n), L_n, L_{A,n}, L_{D,n}, \bar{L}_n, \bar{L}_n^\star, \Sigma_n, \Omega_n, \Omega_{D,n}, V_n\}$, 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 $\Xi_n(\beta), \bar{\Xi}_n(\beta), \Xi_{s,n}(\beta) \in \mathbb{R}^{q+2}$ and $\partial\bar{\Xi}_n^\star(\beta) \in \mathcal{M}_{q+2}$ such that

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

and we write $\bar{\eta} \coloneqq -\partial\bar{\Xi}_n^\star(\bar{\beta}^{(n)})^{-1}\partial\sigma_n^2$.

A.2 Proofs of Lemmas

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

Assumption A1. *There exist a constant $K > 0$ and nonnegative 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 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.*

Lemma A1. *For all integers m and h satisfying $0 \leq h \leq m$, it holds that $\mathbb{D}_m^h = O_m\mathbb{F}_m^h O_m$, where $\mathbb{F}_m^h \in \mathcal{M}_m$ given by*

$$(\mathbb{F}_m^h)_{ij} = \mathbb{1}_{\{h=|i-j|\}} - \mathbb{1}_{\{h=i+j\}} - \mathbb{1}_{\{h=2m+2-(i+j)\}}. \tag{A.3}$$

Proof. The lemma can be verified with straightforward algebra. ∎

Lemma A2. *Suppose $m\Delta_n^{1/2+\alpha} \to \infty$ for some fixed $\alpha > 0$. Define $\mathbb{F}_m^h$ by (A.3). It holds that for $v \in \{0, 1\}$,*

$$V_m^{-1}(\sigma^2, \gamma, \Delta_n) D_m^{-v}(\gamma) = \sum_{h=0}^{m+1} \rho_h(\sigma^2, \gamma, \Delta_n, v) \mathbb{D}_m^h$$

$$\text{and} \qquad \Omega_m^{-1}(\sigma^2, \gamma, \Delta_n) O_m D_m^{-v}(\gamma) O_m = \sum_{h=0}^{m+1} \rho_h(\sigma^2, \gamma, \Delta_n, v) \mathbb{F}_m^h.$$

Here $\rho_h(\sigma^2, \gamma, \Delta_n)$ satisfies that, for all sequences of parameters $\{(\sigma_n^2, \gamma_n) \in \Pi_n^{(\sigma^2,\gamma)}(q_n) : n \geq 1\}$ and all $\{q_n\}$, (i) under $\Delta_n^{-1}\chi^2(\sigma_n^2, \gamma_n, \Delta_n) \to \infty$ and for $v \in \{0, 1\}$,

$$\zeta_n^{2v} \rho_h(\sigma_n^2, \gamma_n, \Delta_n, v) = \frac{1 - z_n^*}{\sigma^2 \Delta_n (1 + z_n^*)} (z_n^*)^{|h|} + O\Big((z_n^*)^{|h|} \chi_n^{-2} \log(\Delta_n^{-1/2} \chi_n) + \frac{1}{\chi_n^2} \wedge \frac{1}{h^2 \Delta_n}\Big), \quad \text{(A.4)}$$

$$\begin{aligned} \zeta_n^{2v} \Big(\rho_h(\sigma_n^2, \gamma_n, \Delta_n, v) - \rho_{h+1}(\sigma_n^2, \gamma_n, \Delta_n, v)\Big) &= \frac{(1 - z_n^*)^2}{\sigma^2 \Delta_n (1 + z_n^*)} (z_n^*)^h \\ &+ O\Big(\Delta_n^{1/2} \chi_n^{-1} \Big(\frac{1}{\chi_n^2} \wedge \frac{1}{h^2 \Delta_n}\Big) + ((z_n^*)^h \chi_n^{-2} \big(\Delta_n^{1/2} \chi_n^{-1} \log(\Delta_n^{-1/2} \chi_n) + h^{-1}\big)\Big), \quad \text{(A.5)} \end{aligned}$$

$$\begin{aligned} &\zeta_n^{2v} \Big(2\rho_{h+1}(\sigma_n^2, \gamma_n, \Delta_n, v) - \rho_{h+2}(\sigma_n^2, \gamma_n, \Delta_n, v) - \rho_h(\sigma_n^2, \gamma_n, \Delta_n, v)\Big) \\ &\quad = -\frac{(1 - z_n^*)^3}{\sigma^2 \Delta_n (1 + z_n^*)} (z_n^*)^h + O\Big(\frac{1}{h^2 \chi_n^2} + (z_n^*)^h \Delta_n \chi_n^{-4} \log(\Delta_n^{-1/2} \chi_n)\Big), \end{aligned} \quad \text{(A.6)}$$

where $(z_n^*, \zeta_n^2, \chi_n^2)$ does not depend on h and is given by

$$z_n^* = 1 - \frac{\sigma_n \Delta_n^{1/2}}{\zeta_n} + o(\Delta_n^{1/2} \chi_n^{-1}), \quad \zeta_n^2 = \sum_{j=-\infty}^{\infty} \gamma_{n,|j|}, \quad \chi_n^2 = \chi^2(\sigma_n^2, \gamma_n, \Delta_n);$$

and (ii) under $\Delta_n^{-1}\chi^2(\sigma_n^2, \gamma_n, \Delta_n) \leq K$,

$$\rho_h(\sigma_n^2, \gamma_n, \Delta_n, 0) = O\big(h^{-2} \Delta_n^{-1}\big). \quad \text{(A.7)}$$

Proof. Step 1. (Main proof) Given the expression of $V_m^{-1} D_m^{-v}$, that of $\Omega_m^{-1} O_m D_m^{-v} O_m$ directly follows by applying Lemma A1. Hence it suffices to analyze $V_m^{-1} D_m^{-v}$. First, for all $z \in \mathbb{C}$ with $z \neq 0$, we define

$$\mathcal{V}(z; \sigma^2, \gamma, \Delta_n) := \sigma^2 \Delta_n + (2 - z - z^{-1}) f(z; \gamma), \quad \text{with} \quad f(z; \gamma) = \sum_{j=-\infty}^{\infty} \gamma_{|j|} z^j. \quad \text{(A.8)}$$

We also define

$$\check{\rho}_h(\sigma^2, \gamma, \Delta_n, v) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{\mathbb{i}h\lambda} d\lambda}{\mathcal{V}(e^{\mathbb{i}\lambda}; \sigma^2, \gamma, \Delta_n) f^v(z; \gamma)}. \tag{A.9}$$

In the remaining steps, we prove a key property whereby $\check{\rho}_h(\sigma^2, \gamma, \Delta_n)$ satisfies (A.4), (A.5), (A.6), and (A.7) (of course, we will replace ρ_h with $\check{\rho}_h$ in those two equations) for all $\{(\sigma_n^2, \gamma_n) \in \Pi_n^{(\sigma^2,\gamma)}(q_n) : n \geq 1\}$. Now we demonstrate that this property directly leads to what this lemma claims. In view of (A.8) and the definitions of V_m and $\mathbb{D}_m^h$, we have

$$V_m(\sigma^2, \gamma, \Delta_n)_{j,j} D_m^v(\gamma)_{j,j} = \mathcal{V}\left(e^{\mathbb{i}\frac{j\pi}{m+1}}; \sigma^2, \gamma, \Delta_n\right) f^v\left(e^{\mathbb{i}\frac{j\pi}{m+1}}; \gamma\right), \quad \forall 1 \leq j \leq m.$$

Because $V_m(\sigma^2, \gamma, \Delta_n)$ and $D_m(\gamma)$ are diagonal by construction, we have

$$(V_m^{-1}(\sigma^2, \gamma, \Delta_n) D_m^{-v}(\gamma))_{i,j} = \frac{\delta_{i,j}}{\mathcal{V}\left(e^{\mathbb{i}\frac{j\pi}{m+1}}; \sigma^2, \gamma, \Delta_n\right) f^v\left(e^{\mathbb{i}\frac{j\pi}{m+1}}; \gamma\right)}.$$

Moreover, since we have $(\mathbb{D}_m^h)_{i,j} = \delta_{i,j}(2 - \delta_{h,0}) \cos \frac{hj\pi}{m+1}$, we only need to show that there exists some $\rho_h(\sigma^2, \gamma, \Delta_n)$ that satisfies

$$\text{that (A.4), (A.5), (A.6), and (A.7) for all } \{(\sigma_n^2, \gamma_n) \in \Pi_n^{(\sigma^2,\gamma)}(q_n) : n \geq 1\} \text{ is true,} \tag{A.10}$$

and that

$$\frac{1}{\mathcal{V}\left(e^{\mathbb{i}\frac{j\pi}{m+1}}; \sigma^2, \gamma, \Delta_n\right) f^v\left(e^{\mathbb{i}\frac{j\pi}{m+1}}; \gamma\right)} = \sum_{h=0}^{m+1} \rho_h(\sigma^2, \gamma, \Delta_n, v) \cos \frac{hj\pi}{m+1}. \tag{A.11}$$

We now prove this is true. Using the definition of $\check{\rho}_h(\sigma^2, \gamma, \Delta_n, v)$ from (A.9), we can write

$$\begin{aligned}
\frac{1}{\mathcal{V}\left(e^{\mathbb{i}\frac{j\pi}{m+1}}; \sigma^2, \gamma, \Delta_n\right) f^v\left(e^{\mathbb{i}\frac{j\pi}{m+1}}; \gamma\right)} &= \sum_{h=0}^{\infty} \check{\rho}_h(\sigma^2, \gamma, \Delta_n, v)(2 - \delta_{h,0}) \cos \frac{hj\pi}{m+1} \\
&= -\check{\rho}_0(\sigma^2, \gamma, \Delta_n, v) \\
&\quad + 2\sum_{h=0}^{m} \sum_{k=0}^{\infty} \check{\rho}_{h+2k(m+1)}(\sigma^2, \gamma, \Delta_n, v) \cos \frac{hj\pi}{m+1} \\
&\quad + 2\sum_{h=1}^{m+1} \sum_{k=0}^{\infty} \check{\rho}_{m+1-h+(2k+1)(m+1)}(\sigma^2, \gamma, \Delta_n, v) \cos \frac{hj\pi}{m+1}.
\end{aligned}$$

Here the last equality comes from basic properties of sine and cosine functions. This indicates that (A.11) indeed holds with $\{\rho_h\}_{h=0}^{m+1}$ given by

$$\rho_0(\sigma^2, \gamma, \Delta_n, v) = \check{\rho}_0 + 2\sum_{k=1}^{\infty} \check{\rho}_{2k(m+1)}, \qquad \rho_{m+1}(\sigma^2, \gamma, \Delta_n, v) = \sum_{k=0}^{\infty} \check{\rho}_{(2k+1)(m+1)}, \tag{A.12}$$

$$\rho_h(\sigma^2,\gamma,\Delta_n,v) \;=\; \sum_{k=0}^{\infty}\left(\check{\rho}_{h+2k(m+1)}+\check{\rho}_{m+1-h+(2k+1)(m+1)}\right),\quad \forall 1\le h\le m. \tag{A.13}$$

Here we omit the argument $(\sigma^2,\gamma,\Delta_n,v)$ of $\check{\rho}_h$. Suppose $\check{\rho}_h(\sigma^2,\gamma,\Delta_n,v)$ satisfies (A.10). Then, given that $m\Delta_n^{1/2+\alpha}\to\infty$ for a fixed $\alpha>0$, we have that $\{\rho_h\}_{h=0}^{m+1}$ defined by (A.12) and (A.13) also satisfies (A.10), which proves the current lemma. Now we move forward to show that $\check{\rho}_h(\sigma^2,\gamma,\Delta_n,v)$ indeed satisfies (A.10).

Step 2. (Characterization of $\check{\rho}$) In this step, we connect the behavior of $\check{\rho}$ with properties of $\mathcal{V}$ using the definition (A.9). We start with a decomposition. We write that for all $p\ge 1$,

$$\mathcal{V}(z;\sigma^2,\gamma,\Delta_n)=\mathcal{V}(z;\sigma^2,\widetilde{\gamma}(p,\gamma),\Delta_n)+\mathcal{V}(z;0,\widetilde{\gamma}(-p,\gamma),\Delta_n),$$

where $\widetilde{\gamma}(p,\gamma)$ and $\widetilde{\gamma}(-p,\gamma)$ are shorthand notation defined by $\widetilde{\gamma}(p,\gamma)=(\gamma_0,\gamma_1,\ldots,\gamma_p,0,\ldots,0)^\intercal$ and $\widetilde{\gamma}(-p,\gamma)=\gamma-\widetilde{\gamma}(p,\gamma)$. In other words, $\widetilde{\gamma}(p,\gamma)$ represents the first $p+1$ components of γ, while $\widetilde{\gamma}(-p,\gamma)$ captures the remaining ones. The decomposition directly comes from the fact that $\mathcal{V}$ is linear in γ. In the rest of the proof, for notational simplicity, we write

$$\mathcal{V}(z;\Delta_n)=\mathcal{V}(z;\sigma^2,\gamma,\Delta_n),\quad \mathcal{V}(z;\Delta_n,p)=\mathcal{V}(z;\sigma^2,\widetilde{\gamma}(p,\gamma),\Delta_n),$$

$$\mathcal{V}(z;\Delta_n,-p)=\mathcal{V}(z;\sigma^2,\widetilde{\gamma}(-p,\gamma),\Delta_n),\quad \text{and}\quad f(z;p)=f(z;\widetilde{\gamma}(p,\gamma)).$$

We can now write that for all $|z|=1$,

$$\mathcal{V}(z;\Delta_n)^{-1}=\mathcal{V}(z;\Delta_n,p_n)^{-1}\left[1+\sum_{j=1}^{\infty}\left(-\frac{\mathcal{V}(z;\Delta_n,-p_n)}{\mathcal{V}(z;\Delta_n,p_n)}\right)^j\right]. \tag{A.14}$$

For a positive sequence p_n, let $\{\check{\rho}_h(\Delta_n,p_n,v)\}_{h=-\infty}^{\infty}$ and $\{\check{\rho}_h(\Delta_n,-p_n)\}_{h=-\infty}^{\infty}$ be the Fourier coefficients of, respectively, $\frac{1}{\mathcal{V}(e^{\mathrm{i}\lambda};\Delta_n,p_n)f^v(e^{\mathrm{i}\lambda};\gamma)}$ and $\sum_{j=1}^{\infty}\left(-\frac{\mathcal{V}(z;\Delta_n,-p_n)}{\mathcal{V}(z;\Delta_n,p_n)}\right)^j$:

$$\check{\rho}_h(\Delta_n,p_n,v)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-\mathrm{i}h\lambda}}{\mathcal{V}(e^{\mathrm{i}\lambda};\Delta_n,p_n)f^v(e^{\mathrm{i}\lambda};\gamma)}d\lambda, \tag{A.15}$$

$$\check{\rho}_h(\Delta_n,-p_n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-\mathrm{i}h\lambda}\sum_{j=1}^{\infty}\left(-\frac{\mathcal{V}(e^{\mathrm{i}\lambda};\Delta_n,-p_n)}{\mathcal{V}(e^{\mathrm{i}\lambda};\Delta_n,p_n)}\right)^j d\lambda. \tag{A.16}$$

In view of (A.9), (A.14), (A.15), and (A.16), we have

$$\check{\rho}_h(\sigma^2,\gamma,\Delta_n,v)=\check{\rho}_h(\Delta_n,p_n,v)+\sum_{j=-\infty}^{\infty}\check{\rho}_j(\Delta_n,p_n,v)\check{\rho}_{h-j}(\Delta_n,-p_n). \tag{A.17}$$

Step 3. (Implication of $(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}$) The definition of $\Pi_n^{(\sigma^2,\gamma)}$ indicates

$$\frac{1}{K}\leq\sigma^2\leq K, \tag{A.18}$$

where the first inequality is because $\sigma^2 = f(0;\sigma^2,\gamma,\Delta_n)$ and the second is obvious. Given the positivity of σ^2, the second inequality in (3.11) requires $\frac{\sum_{j=1}^{\infty} j^2|\gamma_j|}{\inf_\lambda |\sigma^2\Delta_n+f(\lambda;\gamma)|}\leq K$. This further indicates

$$\sup_\lambda\left|\frac{d}{d\lambda}\log|\sigma^2\Delta_n+f(\lambda;\gamma)|\right|\leq K \quad\text{and}\quad \sup_\lambda\left|\frac{1}{\sigma^2\Delta_n+f(\lambda;\gamma)}\frac{d^2}{d\lambda^2}f(\lambda;\gamma)\right|\leq K. \tag{A.19}$$

Because of the periodicity of $\log|\sigma^2\Delta_n+f(\lambda;\gamma)|$, the first inequality of (A.19) indicates that

$$\sup_\lambda\log|\sigma^2\Delta_n+f(\lambda;\gamma)|-\inf_\lambda\log|\sigma^2\Delta_n+f(\lambda;\gamma)|\leq K. \tag{A.20}$$

Using $\sigma^2>0$ and $\sigma^2\Delta_n+4f(-\pi;\gamma)>0$, both of which come from the first inequality in (3.11), we conclude $\sigma^2\Delta_n+f(-\pi;\gamma)>0$. This indicates, in view of the fact that $\log x$ diverges as $x\to 0$ and that $\log|\sigma^2\Delta_n+f(\lambda;\gamma)|$ has a bounded derivative,

$$\inf_\lambda(\sigma^2\Delta_n+f(\lambda;\gamma))>0 \quad\text{and}\quad \inf_\lambda(\sigma^2\Delta_n+f(\lambda;\gamma))\geq\frac{1}{K}\sup_\lambda(\sigma^2\Delta_n+f(\lambda;\gamma)). \tag{A.21}$$

The two inequalities in (A.21), plus the positivity of σ^2, indicates that

$$\sup_\lambda f(\lambda;\gamma)\leq K\inf_\lambda(\sigma^2\Delta_n+f(\lambda;\gamma))\leq K\int_{-\pi}^{\pi}\sigma^2\Delta_n+f(\lambda;\gamma)d\lambda\leq\sigma^2\Delta_n+\gamma_0\leq K,$$

where the last inequality comes from the second inequality in (3.11). Furthermore, using (A.21), straightforward algebra shows that uniformly over $-\pi\leq\lambda\leq\pi$,

$$K^{-1}\chi^2\leq\sigma^2\Delta_n+f(\lambda;\gamma)\leq K\chi^2 \quad\text{and}\quad \sum_{j=1}^{\infty}j^2|\gamma_j|\leq K\chi^2, \tag{A.22}$$

where $\chi^2=\chi^2(\sigma^2,\gamma,\Delta_n)$. We emphasize that all the Ks involved in the current step are constants that do not depend on n, and therefore all the bounds here hold uniformly over all $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q_n):n\geq 1\}$.

Step 4. (The case $\Delta_n^{-1}\chi_n^2\to\infty$: Properties of $\mathcal{V}(z;\Delta_n,p)$, part 1) Throughout the rest of the proof, we suppress the subscript $_n$ of (σ_n^2,γ_n) whenever possible. In the case of $\Delta_n^{-1}\chi_n^2\to\infty$, we first prove that for each n sufficiently large and p_n satisfying $p_n\Delta_n^{1/2}\chi_n^{-1}\leq K$, there exists a unique complex number z_n^* such that

$$1-K^{-1}p_n^{-1}\leq|z_n^*|\leq 1 \quad\text{and}\quad \mathcal{V}(z_n^*;\Delta_n,p_n)=0. \tag{A.23}$$

In other words, asymptotically, z_n^* is the solution of $\mathcal{V}(z; \Delta_n, p_n) = 0$, which is closest to the unit circle in the complex plane. We first show there exists a unique real solution within $[1 - K^{-1}p_n^{-1}, 1]$. We can calculate

$$\frac{1}{1-z}\frac{d}{dz}\mathcal{V}(z; \Delta_n, p_n) = \frac{1+z}{z^2}f(z; p_n) - \frac{1-z}{z}\frac{d}{dz}f(z; p_n). \tag{A.24}$$

Moreover, we have that for n sufficiently large and uniformly over $z \in (1 - K^{-1}p_n^{-1}, 1)$,

$$f(z; p_n) \geq f(1; p_n) - \sum_{j=1}^{p_n} |\gamma_j| \times |z^j + z^{-j} - 2| \geq K^{-1}\chi_n^2 - Kp_n^{-2}\sum_{j=1}^{p_n} |\gamma_j| j^2 \geq K^{-1}\chi_n^2. \tag{A.25}$$

Here the first inequality comes from the triangular inequality, the second comes from the fact that the highest power terms in $f(z; p_n)$ are z^{p_n} and z^{-p_n}, and the last from the second part of (A.22). Furthermore, for n sufficiently large and uniformly over $z \in (1 - K^{-1}p_n^{-1}, 1)$,

$$\left|\frac{d}{dz}f(z; p_n)\right| \leq \sum_{j=1}^{p_n} j|\gamma_j||z^{j-1} - z^{-j-1}| \leq Kp_n^{-1}\sum_{j=1}^{p_n} j^2|\gamma_j| \leq Kp_n^{-1}, \tag{A.26}$$

where once again the first inequality comes from the triangular inequality, the second from the fact that the highest power terms in $f(z; p_n)$ are z^{p_n} and z^{-p_n}, and the last from the second part of (A.22). Plugging (A.25) and (A.26) back into (A.24), we obtain that

$$\inf_{z \in (1-K^{-1}p_n^{-1}, 1)} \frac{1}{1-z}\frac{d}{dz}\mathcal{V}(z; \Delta_n, p_n) \geq K^{-1}\chi_n^2. \tag{A.27}$$

In view of $\mathcal{V}(1; \Delta_n, p_n) = \sigma^2\Delta_n$ and $\sigma^2 \in (K^{-1}, K)$, as shown in (A.18), plus applying mean value theorem, (A.27) readily leads to the existence and uniqueness of the real solution within $[1 - K^{-1}p_n^{-1}, 1]$. Now we show that any z_n^* that satisfies (A.23) must be real for large n. Suppose we write $z_n^* = |z_n^*|e^{\mathbb{i}\varphi_n^*}$ with $\varphi \in [0, 2\pi)$. We prove $\varphi_n^* = 0$ by contradiction. In view of the fact that for all $|z| = 1$, $\mathcal{V}(z; \Delta_n, p_n) \geq K^{-1}\Delta_n > 0$ by construction, it suffices to show that $\varphi_n^* \neq 0$ indicates that $|z_n^*| = 1$. The imaginary part of $\mathcal{V}(z_n^*; \Delta_n, p_n)$ is

$$\mathrm{Im}(\mathcal{V}(z_n^*; \Delta_n, p_n)) = \mathcal{R}_a(z_n^*) + \mathcal{R}_b(z_n^*),$$

where we use the shorthand notation

$$\mathcal{R}_a(z) = -\sin\varphi\left(|z| - \frac{1}{|z|}\right)\mathrm{Re}(f(z, p_n)) \quad \text{and} \quad \mathcal{R}_b(z) = \cos\varphi\left(2 - |z| - \frac{1}{|z|}\right)\mathrm{Im}(f(z, p_n)),$$

with $\varphi \in [0, 2\pi)$ and $e^{\mathbb{i}\varphi} = z/|z|$. We notice that uniformly over $z \in \{z : |z| \in (1 - K^{-1}p_n^{-1}, 1)\}$,

$$\mathrm{Re}(f(z, p_n)) \geq K^{-1} \quad \text{and} \quad |\mathrm{Im}(f(z, p_n))| \leq (1 - |z|) \times |\sin\varphi|, \quad \text{with} \quad e^{\mathbb{i}\varphi} = z/|z|,$$

which can be shown by the same argument that justifies (A.25) and (A.26). This result, plus the proximity of $|z_n^*|$ to one by construction, immediately indicates that $\mathcal{R}_a$ dominates $\mathcal{R}_b$ asymptotically:

$$\sup_{z:|z|\in(1-K^{-1}p_n^{-1},1)} \left|\frac{\mathcal{R}_b(z)}{\mathcal{R}_a(z)}\right| \to 0.$$

On the other hand, we obviously have $\mathrm{Im}(\mathcal{V}(z_n^*;\Delta_n,p_n)) = 0$ because $\mathcal{V}(z_n^*;\Delta_n,p_n) = 0$, which further requires $\mathcal{R}_a(z_n^*) = 0$. Therefore $\varphi_n^* \neq 0$ necessarily indicates $|z_n^*| = 1$. Contradiction is established and the fact that z_n^* is real is proved. Finally, we derive the expression of z_n^*. We write $z_n^* = 1 + a\Delta_n^{1/2}\chi_n^{-1} + b\Delta_n\chi_n^{-2} + \dots$ and match the coefficients to let $\mathcal{V}(z_n^*;\Delta_n,p_n) = 0$, which gives the explict expression of z_n^*, up to $o(\Delta_n^{1/2}\chi_n^{-1})$:

$$z_n^* = 1 - \frac{\sigma_n\Delta_n^{1/2}}{\zeta_n} + o(\Delta_n^{1/2}\chi_n^{-1}). \tag{A.28}$$

Step 5. (The case $\Delta_n^{-1}\chi_n^2 \to \infty$: Properties of $\mathcal{V}(z;\Delta_n,p)$, part 2) Now we study the properties of $\mathcal{V}(z;\Delta_n,p)$ beyond its closest-to-unit-circle root for the case of $\Delta_n^{-1}\chi_n^2 \to \infty$. Given (A.28) and noticing that $\mathcal{V}(z;\Delta_n,p_n) = 0$ indicates $\mathcal{V}(z^{-1};\Delta_n,p_n) = 0$, we can introduce $\widetilde{\mathcal{V}}(z;\Delta_n,p_n)$ defined by

$$(z - z_n^*)\left(\frac{1}{z} - z_n^*\right)\widetilde{\mathcal{V}}(z;\Delta_n,p_n) = \mathcal{V}(z;\Delta_n,p_n), \tag{A.29}$$

where z can be any nonzero complex number. In other words, $\widetilde{\mathcal{V}}(z;\Delta_n,p_n)$ can be understood as capturing the roots of $\mathcal{V}(z;\Delta_n,p_n)$ other than z_n^* and $1/z_n^*$. We now analyze its properties; again, $p_n\Delta_n^{1/2}\chi_n^{-1} \le K$. We first claim that uniformly over $-\pi \le \lambda \le \pi$,

$$K^{-1} \le \frac{(e^{\mathbb{i}\lambda} - z_n^*)(e^{-\mathbb{i}\lambda} - z_n^*)}{\Delta_n\chi_n^{-2} + 1 - \cos\lambda} \le K \quad \text{and} \quad K^{-1} \le \frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p_n)}{\Delta_n + \chi_n^2(1-\cos\lambda)} \le K, \tag{A.30}$$

where the first result is obvious for the expression of z_n^* from (A.28), while the second result can be easily verified using the first part of (A.22). Combined with the construction of $\widetilde{\mathcal{V}}(z;\Delta_n,p_n)$ from (A.29), we obtain that uniformly over $-\pi \le \lambda \le \pi$,

$$K^{-1}\chi_n^2 \le \widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda};\Delta_n,p_n) \le K\chi_n^2. \tag{A.31}$$

In other words, $\widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda})$ is uniformly of the same order of χ_n^2. Now we bound the derivatives of $\widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda})$. We can write

$$\widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda};\Delta_n,p_n) = h(\lambda;\sigma^2,\gamma,\Delta_n) + (z_n^*)^{-1}f(e^{\mathbb{i}\lambda};p_n),$$

where we introduce the shorthand notation $h(\lambda;\sigma^2,\gamma,\Delta_n) := \frac{\sigma^2\Delta_n - (1-z^*)^2(z^*)^{-1}f(e^{\mathbb{i}\lambda};p_n)}{(1-z^*)^2+(2-2\cos\lambda)z^*}$. Because the second part of (A.22), whereby the first- and second-order derivatives of $f(\lambda;\gamma)$ are bounded by $K\chi_n^2$, plus the fact that $|(z_n^*)^{-1}|$ is bounded because of (A.28), it obviously holds that uniformly over

$-\pi \leq \lambda \leq \pi$,

$$\left|\frac{d}{d\lambda}\widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda})\right| \leq K\chi_n^2 \quad \text{and} \quad \left|\frac{d^2}{d\lambda^2}\widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda})\right| \leq K\chi_n^2, \tag{A.32}$$

as long as we show that uniformly over $-\pi \leq \lambda \leq \pi$,

$$\left|\frac{d}{d\lambda}h(\lambda;\sigma^2,\gamma,\Delta_n)\right| \leq K\chi_n^2 \quad \text{and} \quad \left|\frac{d^2}{d\lambda^2}h(\lambda;\sigma^2,\gamma,\Delta_n)\right| \leq K\chi_n^2. \tag{A.33}$$

We first explicitly calculate these two derivatives of h. Some algebra can show that

$$\frac{d}{d\lambda}h(\lambda;\sigma^2,\gamma,\Delta_n) \tag{A.34}$$

$$= \frac{-(1-z_n^*)^2(z_n^*)^{-1}}{(1-z_n^*)^2+(2-2\cos\lambda)z_n^*}\frac{d}{d\lambda}f(e^{\mathbb{i}\lambda};p_n)$$
$$+(\sigma^2\Delta_n-(1-z_n^*)^2(z_n^*)^{-1}f(e^{\mathbb{i}\lambda};p_n))\frac{d}{d\lambda}\frac{1}{(1-z_n^*)^2+(2-2\cos\lambda)z_n^*} \tag{A.35}$$

and

$$\frac{d^2}{d\lambda^2}h(\lambda;\sigma^2,\gamma,\Delta_n) \tag{A.36}$$

$$= \frac{-(1-z_n^*)^2(z_n^*)^{-1}}{(1-z_n^*)^2+(2-2\cos\lambda)z_n^*}\frac{d^2}{d\lambda^2}f(e^{\mathbb{i}\lambda};p_n)$$
$$-2(1-z_n^*)^2(z_n^*)^{-1}\left(\frac{d}{d\lambda}\frac{1}{(1-z_n^*)^2+(2-2\cos\lambda)z_n^*}\right)\frac{d}{d\lambda}f(e^{\mathbb{i}\lambda};p_n)$$
$$+(\sigma^2\Delta_n-(1-z_n^*)^2(z_n^*)^{-1}f(e^{\mathbb{i}\lambda};p_n))\frac{d^2}{d\lambda^2}\frac{1}{(1-z_n^*)^2+(2-2\cos\lambda)z_n^*}. \tag{A.37}$$

Using the expression of z^* from (A.28), we have that uniformly over $-\pi \leq \lambda \leq \pi$,

$$\left|\frac{(1-z_n^*)^2(z_n^*)^{-1}}{(1-z_n^*)^2+(2-2\cos\lambda)z_n^*}\right| \leq K, \quad \left|\frac{d}{d\lambda}\frac{1}{(1-z_n^*)^2+(2-2\cos\lambda)z_n^*}\right| \leq K\frac{|\lambda|}{(\Delta_n\chi_n^{-2}+2-2\cos\lambda)^2},$$

and

$$\left|\frac{d^2}{d\lambda^2}\frac{1}{(1-z_n^*)^2+(2-2\cos\lambda)z_n^*}\right| \leq K\frac{1}{(\Delta_n\chi_n^{-2}+2-2\cos\lambda)^2}+K\frac{\lambda^2}{(\Delta_n\chi_n^{-2}+2-2\cos\lambda)^3}.$$

Plugging these bounds back into the expressions of the derivatives of $h(\lambda;\sigma^2,\gamma,\Delta_n)$ provided by (A.35) and (A.37), plus the fact that the first- and second-order derivatives of $f(e^{\mathbb{i}\lambda};p_n)$ are bounded by $K\chi_n^2$, as indicated by the second part of (A.22), plus the magnitude of $(1-z_n^*)$ and $(z_n^*)^{-1}$ indicated by the expression of z_n^* from (A.28), we obtain

$$\left|\frac{d}{d\lambda}h(\lambda;\sigma^2,\gamma,\Delta_n)\right| \leq K\chi_n^2 + K\left|\lambda\frac{\sigma^2\Delta_n-(1-z_n^*)^2(z_n^*)^{-1}f(e^{\mathbb{i}\lambda};p_n)}{(\Delta_n\chi_n^{-2}+2-2\cos\lambda)^2}\right|$$

and

$$\left|\frac{d^2}{d\lambda^2}h(\lambda;\sigma^2,\gamma,\Delta_n)\right| \le K\chi_n^2 + K\Delta_n\chi_n^{-2}\frac{|\lambda|}{(\Delta_n\chi_n^{-2}+2-2\cos\lambda)^2}\left|\frac{d}{d\lambda}f(e^{\mathbb{i}\lambda};p_n)\right|$$
$$+K\left|\frac{\sigma^2\Delta_n-(1-z_n^*)^2(z_n^*)^{-1}f(e^{\mathbb{i}\lambda};p_n)}{(\Delta_n\chi_n^{-2}+2-2\cos\lambda)^2}\right|$$
$$+K\left|\lambda^2\frac{\sigma^2\Delta_n-(1-z_n^*)^2(z_n^*)^{-1}f(e^{\mathbb{i}\lambda};p_n)}{(\Delta_n\chi_n^{-2}+2-2\cos\lambda)^3}\right|.$$

These two results indicate that to prove (A.33), it is sufficient to show that uniformly over $-\pi \le \lambda \le \pi$,

$$|\sigma^2\Delta_n-(1-z_n^*)^2(z_n^*)^{-1}f(e^{\mathbb{i}\lambda};p_n)| \le K\Delta_n(\lambda^2+\Delta_n\chi_n^{-2}) \quad \text{and} \quad \left|\frac{d}{d\lambda}f(e^{\mathbb{i}\lambda};p_n)\right| \le \chi_n^2|\lambda|. \tag{A.38}$$

The second part of (A.38) comes from the fact that $f(e^{\mathbb{i}\lambda};p_n)$ is a differentiable even function, so $\frac{d}{d\lambda}f(1;p_n)=0$ and the fact that $\left|\frac{d^2}{d\lambda^2}f(e^{\mathbb{i}\lambda};p_n)\right| \le K\chi_n^2$ uniformly over $-\pi\le\lambda\le\pi$ is indicated by the second part of (A.22). Now we show the first part of (A.38). We recall that the definition of z_n^* requires that

$$\sigma^2\Delta_n-(1-z_n^*)^2(z_n^*)^{-1}f(z_n^*;p_n)=0.$$

This indicates that

$$\sigma^2\Delta_n-(1-z_n^*)^2(z_n^*)^{-1}f(e^{\mathbb{i}\lambda};p_n) = -(1-z_n^*)^2(z_n^*)^{-1}(f(e^{\mathbb{i}\lambda};p_n)-f(z_n^*;p_n)). \tag{A.39}$$

In view of $(1-z_n^*)^2 \le K\Delta_n\chi_n^{-2}$ from the expression of z_n^* given by (A.28), plus the traingular inequality, the first part of (A.38) will then come from (A.39) because uniformly over $-\pi\le\lambda\le\pi$,

$$|f(e^{\mathbb{i}\lambda};p_n)-f(1;p_n)| \le K\lambda^2\sum_{j=1}^{p_n}j^2|\gamma_j| \le K\lambda^2\chi_n^2, \tag{A.40}$$

$$|f(z_n^*;p_n)-f(1;p_n)| \le K\sum_{j=1}^{p_n}|\gamma_j||2-(z_n^*)^j-(z_n^*)^{-j}| \le K\Delta_n\chi_n^{-2}\sum_{j=1}^{\infty}j^2|\gamma_j| \le K\Delta_n\chi_n^{-2}. \tag{A.41}$$

The first inequality in (A.40) is obvious from the definition of $f(z;p_n)$ given in Step 2, while the second arises from the second part of (A.22). On the other hand, the first inequality in (A.41) is also obvious from the definition of $f(z;p_n)$, the second is from the definition of z_n^*, and the last is once again from the second part of (A.22).

Step 6. (The case of $\Delta_n^{-1}\chi_n^2\to\infty$: Properties of $\mathcal{V}(z;\Delta_n,-p)$) Now we study the properties of

$\mathcal{V}(z;\Delta_n,-p)$ for the case of $\Delta_n^{-1}\chi_n^2\to\infty$. We notice that

$$\frac{\mathcal{V}(z;\Delta_n,-p)}{\mathcal{V}(z;\Delta_n,p)}=\frac{(2-z-z^{-1})f(z;-p)}{\sigma^2\Delta_n+(2-z-z^{-1})f(z;p)}=\frac{f(z;-p)}{f(z;p)}-\frac{\sigma^2\Delta_n\frac{f(z;-p)}{f(z;p)}}{\sigma^2\Delta_n+(2-z-z^{-1})f(z;p)}.$$

Therefore, we have

$$\sup_\lambda\left|\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}\right|\le K\chi_n^{-2}\sup_\lambda f(e^{\mathbb{i}\lambda};-p)\le Kp^{-2},\tag{A.42}$$

where the last inequality comes from the second part of (A.22). Further, using the second part of (A.22) and with direct calculations, we obtain that uniformly over $-\pi\le\lambda\le\pi$,

$$\left|\frac{d\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}{d\lambda}\right|\le K\chi_n^2|\lambda|,\quad\left|\frac{d\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{d\lambda}\right|\le K\chi_n^2(\lambda^2p^{-1}+|\lambda|p^{-2}),\tag{A.43}$$

$$\left|\frac{d^2\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}{d\lambda^2}\right|\le K\chi_n^2\quad\text{and}\quad\left|\frac{d^2\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{d\lambda^2}\right|\le K\chi_n^2(\lambda^2+|\lambda|p^{-1}+p^{-2}).\tag{A.44}$$

On the other hand, we can caluculate

$$\frac{d}{d\lambda}\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}=-\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)^2}\frac{d}{d\lambda}\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)+\frac{1}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}\frac{d}{d\lambda}\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)\tag{A.45}$$

and

$$\begin{aligned}\frac{d^2}{d\lambda^2}\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}&=-\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)^2}\frac{d^2\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}{d\lambda^2}\\&\quad-\frac{2}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)^2}\frac{d\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}{d\lambda}\frac{d\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{d\lambda}\\&\quad+\frac{1}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}\frac{d^2\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{d\lambda^2}.\end{aligned}\tag{A.46}$$

Plugging (A.30) (A.42), (A.43), and (A.44) back into (A.45) and (A.46), we obtain

$$\left|\frac{d}{d\lambda}\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}\right|\le Kp^{-1}\frac{\lambda^2+\lambda p^{-1}}{\Delta_n\chi_n^{-2}+\lambda^2}\tag{A.47}$$

and

$$\left|\frac{d^2}{d\lambda^2}\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}\right|\le K\frac{p^{-2}+\lambda^2}{\Delta_n\chi_n^{-2}+\lambda^2}+K\frac{|\lambda|^3p^{-1}+\lambda^2p^{-2}}{(\Delta_n\chi_n^{-2}+\lambda^2)^2}.\tag{A.48}$$

Step 7. (The case of $\Delta_n^{-1}\chi_n^2\le K$: Properties of $\mathcal{V}(z;\Delta_n)$) Now we study the properties of $\mathcal{V}(z;\Delta_n)$ for the case of $\Delta_n^{-1}\chi_n^2\le K$. We observe that uniformly over $-\pi\le\lambda\le\pi$,

$$K^{-1}\Delta_n\le\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n)\le K\Delta_n.$$

Indeed, the first inequality comes from the definition of $\Pi_n^{(\sigma^2,\gamma)}$ specified by (3.11). The second inequality comes from

$$\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n) \le \sigma^2\Delta_n + K|f(\lambda;\gamma)| \le K\Delta_n, \tag{A.49}$$

in which the first inequality is obvious, given the definition of $\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n)$, and the second comes from (A.18), the first part of (A.22), and $\Delta_n^{-1}\chi_n^2 \le K$. Using the second part of (A.22), we obtain, under the case of $\Delta_n^{-1}\chi_n^2 \le K$ and uniformly over $-\pi \le \lambda \le \pi$,

$$|f(\lambda;\gamma)| \le K\Delta_n, \quad \left|\frac{df(\lambda;\gamma)}{d\lambda}\right| \le K\Delta_n, \quad \left|\frac{d^2 f(\lambda;\gamma)}{d^2\lambda}\right| \le K\Delta_n.$$

These bounds immediately give that uniformly over $-\pi \le \lambda \le \pi$,

$$\left|\frac{d\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n)}{d\lambda}\right| \le K|f(\lambda;\gamma)| + K\left|\frac{df(\lambda;\gamma)}{d\lambda}\right| \le K\Delta_n \tag{A.50}$$

and

$$\left|\frac{d^2\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n)}{d\lambda^2}\right| \le K|f(\lambda;\gamma)| + K\left|\frac{df(\lambda;\gamma)}{d\lambda}\right| + K\left|\frac{d^2 f(\lambda;\gamma)}{d^2\lambda}\right| \le K\Delta_n. \tag{A.51}$$

Step 8. (Properties of $\check{\rho}$) In this step we show that $\check{\rho}$ satisfies (A.10). We start with the case of $\Delta_n^{-1}\chi_n^2 \to \infty$. We first understand that $\check{\rho}_h(\Delta_n, p_n)$ and $\check{\rho}_h(\Delta_n, -p_n)$ are defined by (A.15) and (A.16). We start by introducing $\{\widetilde{\rho}_h(\Delta_n, p_n, v)\}_{h=-\infty}^{\infty}$ as the Fourier coefficients of $\frac{1}{\widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda};\Delta_n,p_n)f^v(e^{\mathbb{i}\lambda};\gamma)}$:

$$\widetilde{\rho}_h(\Delta_n, p_n, v) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-\mathbb{i}h\lambda}}{\widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda};\Delta_n,p_n)f^v(e^{\mathbb{i}\lambda};\gamma)}d\lambda. \tag{A.52}$$

Using the properties of $\widetilde{\mathcal{V}}(z;\Delta_n,p_n)$ specified by (A.31) and (A.32), plus the properties of $f(e^{\mathbb{i}\lambda};\gamma)$ provided by (A.22) and (A.18) (notice that $f(e^{\mathbb{i}\lambda};\gamma)$ here is the same as $f(\lambda;\gamma)$ in (A.22) and (A.18), as we slightly abuse the notation), we obtain for $p_n\Delta_n^{1/2}\chi_n^{-1} \le K$,

$$\sup_{-\pi\le\lambda\le\pi}\left|\frac{d^2}{d\lambda^2}\frac{1}{\widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda};\Delta_n,p_n)f^v(e^{\mathbb{i}\lambda};\gamma)}\right| \le K\chi_n^{-2-2v}.$$

Let us emphasize that in view of the previous steps, obviously this result holds uniformly over $\{(\sigma_n^2,\gamma_n) \in \Pi_n^{(\sigma^2,\gamma)}(q_n) : n \ge 1\}$; so do all of the relevant results throughout the proof. We omit mentioning this afterward. According to the well-known results on how the smoothness of a function affects the order of magnitude of its Fourier coefficients (see, e.g., the proof of Theorem II.4.7 on Page 46 of **?**), we have, for $p_n\Delta_n^{1/2}\chi_n^{-1} \le K$,

$$|\widetilde{\rho}_h(\Delta_n, p_n, v)| \le K\chi_n^{-2-2v}h^{-2}. \tag{A.53}$$

Further, we notice that from (A.29) and (A.52) we have

$$\sum_{j=-\infty}^{\infty}\widetilde{\rho}_j(\Delta_n,p_n,v)=\sum_{j=-\infty}^{\infty}\widetilde{\rho}_j(\Delta_n,p_n,v)e^{\mathrm{i}j0}=\frac{1}{\widetilde{\mathcal{V}}(1;\Delta_n,p_n)f^v(1,\gamma)}=\frac{(1-z_n^*)^2}{\sigma^2\Delta_n\zeta_n^{2v}}. \tag{A.54}$$

On the other hand, in view of the decomposition of $\mathcal{V}(e^{\mathrm{i}\lambda};\Delta_n,p_n)$ using $\widetilde{\mathcal{V}}(z;\Delta_n,p_n)$ and the definition of $\check{\rho}_h(\Delta_n,p_n,v)$ given in (A.15) as the coefficients of the Laurent expansion of $\frac{1}{\mathcal{V}(e^{\mathrm{i}\lambda};\Delta_n,p_n)f^v(e^{\mathrm{i}\lambda};\gamma)}$, we can write

$$\check{\rho}_h(\Delta_n,p_n,v)=\frac{1}{1-(z_n^*)^2}\sum_{j=-\infty}^{\infty}\widetilde{\rho}_j(\Delta_n,p_n,v)(z_n^*)^{|j-h|}. \tag{A.55}$$

Therefore, using (A.53) and (A.54), plus the expression z_n^* from (A.28) and the bound on σ^2 from (A.18), we can write

$$\begin{aligned}\left|\check{\rho}_h(\Delta_n,p_n,v)-\frac{(1-z_n^*)^2}{\sigma^2\Delta_n\zeta_n^2}\frac{(z_n^*)^{|h|}}{1-(z_n^*)^2}\right| &= \frac{1}{1-(z_n^*)^2}\left|\sum_{j=-\infty}^{\infty}\widetilde{\rho}_j(\Delta_n,p_n,v)((z_n^*)^{|h-j|}-(z_n^*)^{|h|})\right|\\ &\le K(z_n^*)^{|h|}\chi_n^{-2-2v}\log(\Delta_n^{-1/2}\chi_n).\end{aligned} \tag{A.56}$$

Now we move to $\check{\rho}_h(\Delta_n,-p_n)$. The properties of $\mathcal{V}(z;\Delta_n,-p)$ provided by Step 6, including (A.42), (A.47), and (A.48), indicate that under $p_n\Delta_n^{-1/2}\chi_n\to 1$,

$$\int_{-\pi}^{\pi}\left|\frac{d^2}{d\lambda^2}\sum_{j=1}^{\infty}\left(-\frac{\mathcal{V}(e^{\mathrm{i}\lambda};\Delta_n,-p_n)}{\mathcal{V}(e^{\mathrm{i}\lambda};\Delta_n,p_n)}\right)^j\right|d\lambda\le K.$$

Following the proof of Theorem II.4.7 of **?**, plus the definition of $\check{\rho}_h(\Delta_n,-p_n)$ given by (A.16), plus (A.42), this immediately indicates that under $p_n\Delta_n^{-1/2}\chi_n\to 1$,

$$|\check{\rho}_h(\Delta_n,-p_n)|\le K((\Delta_n\chi_n^{-2})\wedge h^{-2}). \tag{A.57}$$

Now we combine (A.56), (A.57), and (A.17) to obtain

$$\check{\rho}_h(\sigma^2,\gamma,\Delta_n,v)=\check{\rho}_h(\Delta_n,p_n,v)+O\left(\frac{1}{\chi_n^{2+2v}}\wedge\frac{1}{h^2\Delta_n\chi_n^{2v}}\right),$$

which, combined with (A.56) again, is exactly (A.4). Now we prove (A.5). In view of (A.54), we can write

$$\begin{aligned}\check{\rho}_h(\Delta_n,p_n,v)-\check{\rho}_{h+1}(\Delta_n,p_n,v) &= \frac{1}{1-(z_n^*)^2}\sum_{j=h+1}^{\infty}\widetilde{\rho}_j(\Delta_n,p_n,v)(z_n^*)^{j-h-1}(z_n^*-1)\\ &\quad+\frac{1}{1-(z_n^*)^2}\sum_{j=-\infty}^{h}\widetilde{\rho}_j(\Delta_n,p_n,v)(z_n^*)^{h-j}(1-z_n^*).\end{aligned}$$

This indicates, using (A.53), the expression of z_n^* from (A.28), and the bound on σ^2 from (A.18), that

$$\begin{aligned}\zeta_n^{2v}\Big(\check\rho_h(\Delta_n,p_n,v)-\check\rho_{h+1}(\Delta_n,p_n,v)\Big) &= \frac{(1-z_n^*)^2}{\sigma^2\Delta_n(1+z_n^*)}(z_n^*)^h+O\Big(\Delta_n^{1/2}\chi_n^{-1}\Big(\frac{1}{\chi_n^2}\wedge\frac{1}{h^2\Delta_n}\Big)\Big)\\ &\quad +O\Big((z_n^*)^h\chi_n^{-2}\big(\Delta_n^{1/2}\chi_n^{-1}\log(\Delta_n^{-1/2}\chi_n)+h^{-1}\big)\Big).\end{aligned}$$

Provided this result, we can immediately verify (A.5) by observing that

$$\begin{aligned}\check\rho_h(\sigma^2,\gamma,\Delta_n,v)-\check\rho_{h+1}(\sigma^2,\gamma,\Delta_n,v) &= \check\rho_h(\Delta_n,p_n,v)-\check\rho_{h+1}(\Delta_n,p_n,v)\\ &\quad+\sum_{j=-\infty}^{\infty}\Big(\check\rho_j(\Delta_n,p_n,v)-\check\rho_{j+1}(\Delta_n,p_n,v)\Big)\check\rho_{h-j}(\Delta_n,-p_n)\end{aligned}$$

and using the bound on $\check\rho_h(\Delta_n,-p_n)$ given by (A.57). We can prove (A.6) in the same way. Indeed, (A.54) and (A.55) indicate that

$$\begin{aligned}&2\check\rho_{h+1}(\Delta_n,p_n,v)-\check\rho_{h+2}(\Delta_n,p_n,v)-\check\rho_h(\Delta_n,p_n,v)\\ &\quad=-\frac{(1-z_n^*)^3}{\sigma^2\Delta_n(1+z_n^*)}(z_n^*)^h-\frac{1-z_n^*}{1+z_n^*}(z_n^*)^h\sum_{j=h+2}^{\infty}\widetilde\rho_j(\Delta_n,p_n,v)\Big((z_n^*)^{j-2h-1}-1\Big)\\ &\qquad-\frac{1-z_n^*}{1+z_n^*}(z_n^*)^h\widetilde\rho_{h+1}(\Delta_n,p_n,v)\Big((z_n^*)^{1-h}\frac{2}{z_n^*-1}-1\Big)\\ &\qquad-\frac{1-z_n^*}{1+z_n^*}(z_n^*)^h\sum_{j=-\infty}^{h}\widetilde\rho_j(\Delta_n,p_n,v)\Big((z_n^*)^{1-j}-1\Big).\end{aligned}$$

Using (A.53), the expression of z_n^* from (A.28), and the bound on σ^2 from (A.18), we obtain

$$\begin{aligned}&\zeta_n^{2v}\Big(2\check\rho_{h+1}(\Delta_n,p_n,v)-\check\rho_{h+2}(\Delta_n,p_n,v)-\check\rho_h(\Delta_n,p_n,v)\Big)\\ &\quad=-\frac{(1-z_n^*)^3}{\sigma^2\Delta_n(1+z_n^*)}(z_n^*)^h+O\Big((z_n^*)^h\Delta_n\chi_n^{-4}\log(\Delta_n^{-1/2}\chi_n)+\chi_n^{-2}h^{-2}\Big).\end{aligned}$$

Combining this result with the equality

$$\begin{aligned}&2\check\rho_{h+1}(\sigma^2,\gamma,\Delta_n,v)-\check\rho_h(\sigma^2,\gamma,\Delta_n,v)-\check\rho_{h+1}(\sigma^2,\gamma,\Delta_n,v)\\ &\quad=2\check\rho_{h+1}(\Delta_n,p_n,v)-\check\rho_{h+2}(\Delta_n,p_n,v)-\check\rho_h(\Delta_n,p_n,v)\\ &\qquad+\sum_{j=-\infty}^{\infty}\Big(2\check\rho_{h+1}(\Delta_n,p_n,v)-\check\rho_{h+2}(\Delta_n,p_n,v)-\check\rho_h(\Delta_n,p_n,v)\Big)\check\rho_{h-j}(\Delta_n,-p_n),\end{aligned}$$

which comes immediately from (A.17), and the bound on $\check\rho_h(\Delta_n,-p_n)$ given by (A.57), we readily obtain (A.6). We move forward to the case of $\Delta_n^{-1}\chi_n^2\le K$ and prove (A.7). This is relatively straightforward. Given (A.49), (A.50), and (A.51), it follows that

$$\sup_{\lambda}\left|\frac{d^2}{d\lambda^2}\frac{1}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n)}\right| \leq K\Delta_n^{-1}.$$

Following the proof of Theorem II.4.7 of **?**, we immediately obtain

$$\check{\rho}_h(\sigma^2,\gamma,\Delta_n,0) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-\mathbb{i}h\lambda}}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n)}d\lambda = O\left(\frac{1}{h^2\Delta_n}\right). \tag{A.58}$$

∎

Lemma A3. *Suppose $m\Delta_n^{1/2+\alpha}\to\infty$ for some fixed $\alpha>0$. Also suppose $(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q)$ with q fixed and $\frac{1}{K}\leq\inf_\lambda f(\lambda;\gamma)\leq\sup_\lambda f(\lambda;\gamma)\leq K$. Let $\zeta^2=f(0,\gamma)$. As $n\to\infty$, it holds that*

$$V_n(\sigma^2,\gamma,\Delta_n) = \frac{1}{\widetilde{\zeta}^2}V_n(\widetilde{\sigma}^2,\widetilde{\zeta}^2,\Delta_n)O_nD_n(\widetilde{\gamma})O_n,\ \textit{with}\ |\sigma^2-\widetilde{\sigma}^2|+\|\widetilde{\gamma}-\gamma\|+|\widetilde{\zeta}^2-\zeta^2|\lesssim n^{-1/2},$$

and

$$\Omega_n^{-1}(\sigma^2,\zeta^2,0)_{ij} = b_n(z_n^{|i-j|}-z_n^{i+j}-z_n^{2n+2-i-j})+O(n^{-\infty}),$$

where b_n and z_n do not depend on i or j and satisfy

$$b_n=\frac{1}{2\sigma\zeta\Delta_n^{1/2}}+O(1),\quad z_n=1-\frac{\sqrt{\sigma^2\Delta_n}}{\zeta}+O(n^{-1}).$$

Proof. The desired results follow from a simplified version of the proof of Lemma A2. Concretely, the fact that q is finite allows us to solve explicitly the $q+1$ zeros of $\mathcal{V}(z;\sigma^2,\gamma,\Delta_n)$, up to $O(n^{-1})$. Hence, we directly obtain the first result using steps 1, 3, and 4 of the proof of A2. The second result follows by also applying Lemma A1. ∎

Lemma A4. *Suppose $m\Delta_n^{1/2+\alpha}\to\infty$ for some fixed $\alpha>0$. Let $\zeta^2=f(0;\gamma)$. Omitting the argument $(\sigma^2,\gamma,\Delta_n)$ of V_m and D_m, it holds that, for all $(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n)$ satisfying $\sup_\lambda f(\lambda;\gamma)\sim\inf_\lambda f(\lambda;\gamma)$ and $\Delta_n^{-1}\gamma_0\to\infty$,*

(i) For $j\in\{1,2,3,4\}$,

$$tr(V_m^{-j})=\frac{\lambda_j m}{\zeta(\sigma^2\Delta_n)^{j-1/2}}+o(m\Delta_n^{-j+1/2}),\quad \textit{with}\quad \lambda_1=\frac{1}{2},\quad\lambda_2=\frac{1}{4},\quad\lambda_3=\frac{3}{16},\quad\lambda_4=\frac{5}{32}.$$

(ii) For $0\leq i,j\leq q_n$, and $v\in\{0,1\}$,

$$\frac{1}{m}tr\bigg(V_m^{-1}\frac{\partial V_m}{\partial\gamma_i}\Big(V_m^{-1}\frac{\partial V_m}{\partial\gamma_j}\Big)^v\bigg)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial\log f(\lambda;\gamma)}{\partial\gamma_i}\Big(\frac{\partial\log f(\lambda;\gamma)}{\partial\gamma_j}\Big)^v d\lambda+o(1).$$

Proof. The current lemma follows from straightforward algebra using Lemma A2. ∎

Lemma A5. *Suppose Assumptions 1 - 4 hold. For all sequences* $\{q_n\}$ *and under* $\iota^{(n)}\Delta_n^{-1/2} \to \infty$*, it holds that for all* $0 \le j \le q_n$*,*

$$\left|\sigma^{(n)}(q_n)^2 - C_T\right| \lesssim K\|\gamma^{(n)}\|_{1,(q_n)}/(\iota^{(n)})^2, \quad \left|\gamma^{(n)}(q_n)_j - \gamma_j^{(n)}\right| \lesssim \left(\frac{\Delta_n^{3/4} q_n}{(\iota^{(n)})^{3/2}} + 1\right)\|\gamma^{(n)}\|_{(q_n)}.$$

Proof. Step 1. (Characterization of $\sigma^{(n)}(q_n)^2 - C_T$) Throughout the proof, we omit writing the subscript $_n$ of q_n. The inequality obviously holds, by the definition of $\Pi_n^{(\sigma^2,\gamma)}(q)$, if $\|\gamma^{(n)}\|_{(q_n)}/(\iota^{(n)})^2 \ge \frac{1}{K}$. The subsequence argument indicates that we only need to consider the case in which $\|\gamma^{(n)}\|_{1,(q_n)}/(\iota^{(n)})^2 = o(1)$. The definitions of $\Pi_n^{(\sigma^2,\gamma)}(q)$ and of $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$ and Assumption 4 indicate that $\frac{\partial \bar{L}_n^\star}{\partial \gamma_j} = 0$, $\forall\ 0 \le j \le q$ and $\frac{\partial \bar{L}_n^\star}{\partial \sigma^2} = 0$. Now we solve these first-order conditions explicitly. From the definition of $\bar{L}_n^\star(\sigma^2, \gamma)$, we have that for $0 \le j \le q$,

$$-\frac{2}{n_T}\frac{\partial \bar{L}_n^\star(\sigma^2,\gamma)}{\partial \sigma^2} = \frac{\Delta_n}{2\pi}\int_{-\pi}^{\pi} \frac{f(\lambda; C_T, \gamma^{(n)}, \Delta_n) - f(\lambda; \sigma^2, \gamma, \Delta_n)}{f^2(\lambda; \sigma^2, \gamma, \Delta_n)} d\lambda, \tag{A.59}$$

$$-\frac{2}{n_T}\frac{1}{2-\delta_{0,j}}\frac{\partial \bar{L}_n^\star(\sigma^2,\gamma)}{\partial \gamma_j} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f(\lambda; C_T, \gamma^{(n)}, \Delta_n) - f(\lambda; \sigma^2, \gamma, \Delta_n)}{f^2(\lambda; \sigma^2, \gamma, \Delta_n)}(2 - 2\cos\lambda)\cos j\lambda d\lambda \tag{A.60}$$

Substituting the definition of f back into (A.59) and (A.60), we can write that for $0 \le j \le q$,

$$-\frac{2}{n_T}\frac{\partial \bar{L}_n^\star(\sigma^2,\gamma)}{\partial \sigma^2} = c(\sigma^2,\gamma)_{1,1}(C_T - \sigma^{(n)}(q)^2) + \sum_{k=0}^{\infty} c(\sigma^2,\gamma)_{1,2+k}(\gamma_k^{(n)} - \gamma_k), \tag{A.61}$$

$$-\frac{2}{n_T}\frac{\partial \bar{L}_n^\star(\sigma^2,\gamma)}{\partial \gamma_j} = c(\sigma^2,\gamma)_{j+2,1}(C_T - \sigma^{(n)}(q)^2) + \sum_{j=0}^{\infty} c(\sigma^2,\gamma)_{2+j,2+k}(\gamma_k^{(n)} - \gamma_k). \tag{A.62}$$

Here we use the shorthand notation

$$\begin{aligned}
c(\sigma^2,\gamma)_{1,1} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Delta_n^2}{f^2(\lambda; \sigma^2, \gamma, \Delta_n)} d\lambda, \\
c(\sigma^2,\gamma)_{j+2,1} = c(\sigma^2,\gamma)_{1,2+j} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Delta_n(2 - 2\cos\lambda)(2 - \delta_{j,0})\cos j\lambda}{f^2(\lambda; \sigma^2, \gamma, \Delta_n)} d\lambda, \\
c(\sigma^2,\gamma)_{j+2,k+2} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{(2 - 2\cos\lambda)^2(2 - \delta_{j,0})\cos j\lambda(2 - \delta_{k,0})\cos k\lambda}{f^2(\lambda; \sigma^2, \gamma, \Delta_n)} d\lambda.
\end{aligned}$$

Recall that equations (A.61) and (A.62), evaluated at $(\sigma^2, \gamma) = (\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$ are the first-order conditions we are solving. In view of the form of (A.61) and (A.62), plus the fact that $\gamma^{(n)}(q)_j = 0$ for $j \ge q + 1$, we introduce a $(q + 2)$-dimensional vector A defined as the first $q + 2$ components of $(C_T - \sigma^{(n)}(q)^2, \gamma^{(n)} - \gamma^{(n)}(q))$. We can write the first-order conditions in a compact form:

$$\sum_{k=1}^{q+2} c_{j,k}A_k + \sum_{k=q+3}^{\infty} c_{j,k}\gamma_{k-2}^{(n)} = 0$$

for all $1 \le j \le q+2$. Here we omit the argument $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$ of c for cleaner exposition. This system of $q+2$ equations can be regarded as a matrix equation satisfied by vector A. Indeed, we can write $CA + B = 0$, where C is a $(q+2) \times (q+2)$ matrix and B is a $(q+2)$-dimensional vector and their entries are given by $C_{j,k} = c(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))_{j,k}$, and $B_j = \sum_{k=q+3}^{\infty} c(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))_{j,k} \gamma_{k-2}^{(n)}$. We can invert the equation $CA+B=0$ and obtain an explicit characterization of $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$:

$$\sigma^{(n)}(q)^2 - C_T = \sum_{k=1}^{q+2} (C^{-1})_{1,k} B_k, \quad \gamma^{(n)}(q)_j - \gamma_j^{(n)} = \sum_{k=1}^{q+2} (C^{-1})_{j+2,k} B_k. \tag{A.63}$$

We note that here we have only solved $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$ partially, because vector B and matrix C depend on $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$. But the goal here is to show that $\sigma^{(n)}(q)^2 - C_T$ is small, and such partial solution turns out to be sufficient.

Step 2. (Behavior of χ^2) It is clear from (A.63) that we understand the properies of c and C^{-1} to provide a bound on $\sigma^{(n)}(q_n)^2 - C_T$. Before that, we first understand the behavior of

$$\chi^{(n)}(q)^2 := \chi^2(\sigma^{(n)}(q)^2, \gamma^{(n)}(q), \Delta_n),$$

which is technically necessary. We introduce shorthand $\gamma_{(q)}^{(n)}$ to be the $(q+1)$-dimensional vector consisting of the first $q+1$ components of $\gamma^{(n)}$. Because of the properties of $\theta^{(n)}$ specified by Assumption 4 and the definitions of $\gamma^{(n)}$ and $\Pi_n^{(\sigma^2,\gamma)}(q)$, we have $(C_T, \gamma_{(q)}^{(n)}) \in \Pi_n^{(\sigma^2,\gamma)}(q)$. Therefore the first part of (A.22) indicates that uniformly over $-\pi \le \lambda \le \pi$,

$$C_T \Delta_n + f(\lambda; \gamma_{(q)}^{(n)}) \sim \chi^2(C_T, \gamma_{(q)}^{(n)}, \Delta_n). \tag{A.64}$$

In view of the definition of $\chi^2(\sigma^2, \gamma, \Delta_n)$, we conclude that under $\Delta_n^{-1/2} \iota^{(n)} \to \infty$,

$$\chi^2(C_T, \gamma_{(q)}^{(n)}, \Delta_n) \sim (\iota^{(n)})^2. \tag{A.65}$$

Because we are considering $\|\gamma^{(n)}\|_{1,(q)} (\iota^{(n)})^{-2} = o(1)$, as we discussed in Step 1, and the fact that (A.64) and (A.65) jointly indicate $f(\lambda; \gamma^{(n)}) \ge K^{-1} (\iota^{(n)})^2$ under $\Delta_n^{-1/2} \iota^{(n)} \to \infty$ and uniformly over $-\pi \le \lambda \le \pi$, $f(\lambda; \gamma_{(q)}^{(n)}) = f(\lambda; \gamma^{(n)}) - \sum_{j=q+1}^{\infty} 2\gamma_j^{(n)} \cos\lambda \ge K^{-1} (\iota^{(n)})^2$, which further gives

$$\frac{f(\lambda; C_T, \gamma^{(n)}, \Delta_n)}{f(\lambda; C_T, \gamma_{(q)}^{(n)}, \Delta_n)} = 1 + \frac{(2 - 2\cos\lambda) \sum_{j=q+1}^{\infty} 2\gamma_j^{(n)} \cos\lambda}{\sigma^2 \Delta_n + (2 - 2\cos\lambda) f(\lambda; \gamma_{(q)}^{(n)})} = 1 + o(1).$$

Given this result, and in view of the definition of $\bar{L}_n^\star(\sigma^2, \gamma)$, we can write $2n_T^{-1} \bar{L}_n^\star(C_T, \gamma_{(q)}^{(n)}) = -\log \chi^2(C_T, \gamma_{(q)}^{(n)}, \Delta_n) - 1 + o(1)$. Since $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$ is constructed as the maximizer of $\bar{L}_n^\star$ over

$\Pi_n^{(\sigma^2,\gamma)}(q)$, we have

$$2n_T^{-1}\bar{L}_n^{\star}(\sigma^{(n)}(q)^2,\gamma^{(n)}(q)) \geq -\log\chi^2(C_T,\gamma_{(q)}^{(n)},\Delta_n)-1+o(1). \tag{A.66}$$

On the other hand, starting from the definition of $\bar{L}_n^{\star}$, we can write

$$2n_T^{-1}\bar{L}_n^{\star}(\sigma^{(n)}(q)^2,\gamma^{(n)}(q)) \leq -\log\chi^2(\sigma^{(n)}(q)^2,\gamma^{(n)}(q),\Delta_n)-K^{-1}\frac{\chi^2(C_T,\gamma_{(q)}^{(n)},\Delta_n)}{\chi^2(\sigma^{(n)}(q)^2,\gamma^{(n)}(q),\Delta_n)}, \tag{A.67}$$

where the last step comes from the first inequality in (A.64) and the fact that it holds uniformly over $-\pi\leq\lambda\leq\pi$ that $0<f(\lambda;\sigma^{(n)}(q)^2,\gamma^{(n)}(q),\Delta_n)\leq K\chi^2(\sigma^{(n)}(q)^2,\gamma^{(n)}(q),\Delta_n)$, which is indicated by the first part of (A.22), since we obviously have $(\sigma^{(n)}(q)^2,\gamma^{(n)}(q))\in\Pi_n^{(\sigma^2,\gamma)}(q)$. Combining (A.66) and (A.67) indicates that $\log\frac{\chi^2(\sigma^{(n)}(q)^2,\gamma^{(n)}(q),\Delta_n)}{\chi^2(C_T,\gamma_{(q)}^{(n)},\Delta_n)}+K^{-1}\frac{\chi^2(C_T,\gamma_{(q)}^{(n)},\Delta_n)}{\chi^2(\sigma^{(n)}(q)^2,\gamma^{(n)}(q),\Delta_n)}\leq 1+o(1)$. Combined with (A.66) and (A.67), this indicates that we must have

$$\chi^2(\sigma^{(n)}(q)^2,\gamma^{(n)}(q),\Delta_n)\sim\chi^2(C_T,\gamma_{(q)}^{(n)},\Delta_n). \tag{A.68}$$

Step 3. (Properties of c) In this step we analyze the properies of c. We first express c in terms of Fourier coefficients of various functions. Fourier analysis states that $\frac{1}{f(\lambda;\sigma^2,\gamma,\Delta_n)}=\sum_{j=-\infty}^{\infty}\rho_j e^{\mathbb{i}j\lambda}$, with $\rho_j:=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-\mathbb{i}j\lambda}d\lambda}{f(\lambda;\sigma^2,\gamma,\Delta_n)}$. This allows us to write $c(\sigma^2,\gamma)_{1,1}=\Delta_n^2\sum_{j=-\infty}^{\infty}\rho_j^2$. Now we move to $c(\sigma^2,\gamma)_{1,2+j}$ and $c(\sigma^2,\gamma)_{j+2,k+2}$. We first show some auxiliary results. We introduce for $p\in\{1,2\}$, $\rho_j^{(\gamma,p)}:=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-\mathbb{i}j\lambda}d\lambda}{f^p(\lambda;\gamma)}$. We can calculate with some algebra that for $p\in\{1,2\}$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{\mathbb{i}i\lambda}}{f^p(\lambda;\gamma)}d\lambda = \rho_i^{(\gamma,p)}, \tag{A.69}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f(\lambda;\sigma^2,\gamma,\Delta_n)}\frac{e^{\mathbb{i}i\lambda}}{f^p(\lambda;\gamma)}d\lambda = \sum_{j=-\infty}^{\infty}\rho_j^{(\gamma,p)}\rho_{i+j}, \tag{A.70}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f^2(\lambda;\sigma^2,\gamma,\Delta_n)}\frac{e^{\mathbb{i}i\lambda}}{f^p(\lambda;\gamma)}d\lambda = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\rho_j^{(\gamma,p)}\rho_k\rho_{i+j+k}. \tag{A.71}$$

Next, from the definition of $f(\lambda;\sigma^2,\gamma,\Delta_n)$ we notice that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{(2-2\cos\lambda)\cos i\lambda}{f^2(\lambda;\sigma^2,\gamma,\Delta_n)}d\lambda=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f(\lambda;\sigma^2,\gamma,\Delta_n)-\sigma^2\Delta_n}{f^2(\lambda;\sigma^2,\gamma,\Delta_n)}\frac{e^{\mathbb{i}i\lambda}}{f(\lambda;\gamma)}d\lambda,$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{(2-2\cos\lambda)^2\cos j\lambda\cos k\lambda}{f^2(\lambda;\sigma^2,\gamma,\Delta_n)}d\lambda=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{(f(\lambda;\sigma^2,\gamma,\Delta_n)-\sigma^2\Delta_n)^2}{f^2(\lambda;\sigma^2,\gamma,\Delta_n)}\frac{e^{\mathbb{i}(i+j)\lambda}+e^{\mathbb{i}(i-j)\lambda}}{2f^2(\lambda;\gamma)}d\lambda.$$

With these two equalities, plus (A.69) - (A.71), we are able to write

$$
\begin{aligned}
c(\sigma^2,\gamma)_{1,2+j} &= \Delta_n(2-\delta_{j,0})\Big(\sum_{k=-\infty}^{\infty}\rho_j^{(\gamma,1)}\rho_{j+k}-\sigma^2\Delta_n\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\rho_k^{(\gamma,1)}\rho_l\rho_{j+k+l}\Big),\\
c(\sigma^2,\gamma)_{i+2,j+2} &= \frac{1}{2}(2-\delta_{i,0})(2-\delta_{j,0})\Big((\rho_{i+j}^{(\gamma,2)}+\rho_{i-j}^{(\gamma,2)})-2\sigma^2\Delta_n\sum_{l=-\infty}^{\infty}\rho_l^{(\gamma,2)}(\rho_{i+j+l}+\rho_{i-j+l})\\
&\quad+(\sigma^2\Delta_n)^2\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\rho_k^{(\gamma,2)}\rho_l(\rho_{i+j+k+l}+\rho_{i-j+k+l})\Big).
\end{aligned}
$$

At this stage, it is quite clear that in order to bound c, we only need to control the behavior of ρ and $\rho^{(\gamma,p)}$. To emphasize the dependence, we write $\rho_h(\sigma^2,\gamma,\Delta_n)$ and $\rho_h^{(\gamma,p)}(\gamma)$. Scrutiny of the proof of Lemma A2 reveals that $\rho_h(\sigma^2,\gamma,\Delta_n)$ is exactly $\breve{\rho}_h(\sigma^2,\gamma,\Delta_n)$ defined in (A.9). Therefore, we have under $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q):n\geq 1\}$ and $\Delta_n^{-1}\chi^2(\sigma_n^2,\gamma_n,\Delta_n)\to\infty$, $\rho_h(\sigma_n^2,\gamma_n,\Delta_n)$ satisfies (A.4). It is worth pointing out that ρ_h here and ρ_h in Lemma A2 are not exactly the same, but are very close and both satisfy (A.4). See the ending part of step 1 of the proof of Lemma A2 for details. On the other hand, from the construction of $\Pi_n^{(\sigma^2,\gamma)}(q)$, we know under $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q):n\geq 1\}$ and $\Delta_n^{-1}\chi^2(\sigma_n^2,\gamma_n,\Delta_n)\to\infty$ that uniformly over $-\pi\leq\lambda\leq\pi$,

$$
K^{-1}\chi_n^2\leq f(\lambda;\gamma_n)\leq K\chi_n^2\quad\text{and}\quad\sum_{h=0}^{\infty}h^2|(\gamma_n)_h|\leq K\chi_n^2,\quad\text{with}\quad\chi_n^2:=\chi^2(\sigma_n^2,\gamma_n,\Delta_n),
$$

where both results come from (A.18) and (A.22). An immediate result is that under the same condition and for $p\in\{1,2\}$, we can control the order of magnitude of the Fourier coefficients $f^{-p}(\lambda;\gamma_n)$ (see, e.g., the proof of Theorem II.4.7 of ?), i.e., $\rho_h^{(\gamma,p)}(\gamma_n)$ as

$$
\sum_{h=0}^{\infty}(h+1)^2\chi_n^{2p}|\rho_h^{(\gamma,p)}(\gamma_n)|\leq K. \tag{A.72}
$$

Given the bounds on $\rho_h(\sigma_n^2,\gamma_n,\Delta_n)$ and $\rho_h^{(\gamma,p)}(\gamma_n)$, plus $\frac{1}{K}\leq\sigma_n^2\leq K$ as indicated by (A.18), some algebra leads to the following technical results, again under the condition $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q):n\geq 1\}$ and $\Delta_n^{-1}\chi^2(\sigma_n^2,\gamma_n,\Delta_n)\to\infty$ and for $p\in\{1,2\}$,

$$
\begin{aligned}
\sum_{l=-\infty}^{\infty}|\rho_l\rho_{i+l}| &\lesssim \frac{e^{-|i|\Delta_n^{1/2}\chi_n^{-1}}}{\Delta_n^{3/2}\chi_n}+\frac{1}{\Delta_n}\Big(\frac{1}{\chi_n^2}\wedge\frac{1}{i^2\Delta_n}\Big),\\
\chi_n^{2p}\sum_{j=-\infty}^{\infty}|\rho_j^{(\gamma,p)}\rho_{i+j}| &\lesssim \frac{1}{\Delta_n^{1/2}\chi_n}\Big(e^{-|i|\Delta_n^{1/2}\chi_n^{-1}}+\frac{\Delta_n^{-1/2}\chi_n}{i^2}\wedge 1\Big),\\
\chi_n^{2p}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|\rho_k^{(\gamma,p)}\rho_l\rho_{i+k+l}| &\lesssim \frac{1}{\Delta_n^{3/2}\chi_n}\Big(e^{-|i|\Delta_n^{1/2}\chi_n^{-1}}+\frac{\Delta_n^{-1/2}\chi_n}{i^2}\wedge 1\Big),
\end{aligned}
$$

where $\chi_n^2 \coloneqq \chi^2(\sigma_n^2, \gamma_n, \Delta_n)$. Therefore, using the previous expressions of c in terms of ρ and $\rho^{(\gamma,p)}$, plus that $\Delta_n^{-1}\chi^2(\sigma^{(n)}(q)^2, \gamma^{(n)}(q), \Delta_n) \to \infty$ as required by (A.68) and $\Delta_n^{-1/2}\iota^{(n)} \to \infty$, we are able to write

$$
\begin{aligned}
|c(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))_{1,1}| &\lesssim \Delta_n^{1/2}(\iota^{(n)})^{-1}, \\
(\iota^{(n)})^2|c(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))_{1,2+j}| &\lesssim \frac{\Delta_n^{1/2}}{\iota^{(n)}} e^{-|j|\Delta_n^{1/2}(\iota^{(n)})^{-1}} + \frac{1}{j^2} \wedge \frac{\Delta_n^{1/2}}{\iota^{(n)}}, \qquad (A.73) \\
(\iota^{(n)})^4|c(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))_{i+2,j+2}| &\lesssim |\rho_{i+j}^{(\gamma,2)} + \rho_{i-j}^{(\gamma,2)}| \\
&\quad + \frac{\Delta_n^{1/2}}{\iota^{(n)}}\Big(e^{-|i+j|\Delta_n^{1/2}(\iota^{(n)})^{-1}} + e^{-|i-j|\Delta_n^{1/2}(\iota^{(n)})^{-1}}\Big). \qquad (A.74)
\end{aligned}
$$

Step 4. (Properties of C^{-1}: Special case) Now we invert matrix C. We define a $(q+2) \times (q+2)$ matrix $C(\sigma^2, \gamma)$ whose entries are

$$
C(\sigma^2, \gamma)_{i,j} = c(\sigma^2, \gamma)_{i,j}, \quad 1 \le i, j \le q+2. \tag{A.75}
$$

Obviously, the matrix C appearing in (A.63) is just $C(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$. We note that by definition $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q)) \in \Pi_n^{(\sigma^2,\gamma)}(q)$. Also, we have $\Delta_n^{-1/2}\chi(\sigma^{(n)}(q)^2, \gamma^{(n)}(q), \Delta_n) \to \infty$, as indicated by (A.68) and $\Delta_n^{-1/2}\iota^{(n)} \to \infty$. It is hence more than enough to calculate $C^{-1}(\sigma_n^2, \gamma_n)$ for all $\{(\sigma_n^2, \gamma_n) \in \Pi_n^{(\sigma^2,\gamma)}(q) : n \ge 1\}$ satisfying $\Delta_n^{-1/2}\chi_n \to \infty$, where we recall that $\chi_n^2 \coloneqq \chi^2(\sigma_n^2, \gamma_n, \Delta_n)$. The current step considers this problem under the following restriction:

$$
(\gamma_n)_j = 0 \text{ for all } n \ge 1 \text{ and } j \ge \lceil \Delta_n^{-1/2}\chi_n \rceil + 1. \tag{A.76}
$$

The next step will show that restriction (A.76) is innocuous. To be able to invert C, an ∞-dimensional matrix, we introduce a reparameterization scheme. Concretely, with a scalar z and a $(q+1)$-dimensional vector ϕ, we define $f(\lambda; \phi) = \sum_{j=0}^q (2 - \delta_{j,0})\phi_j \cos j\lambda$ and $f(\lambda; z, \phi) = |1 - ze^{\mathrm{i}\lambda}|^2 f(\lambda; \phi)$. We connect the parameterization (z, ϕ) to (σ_n^2, γ_n), which belongs to $\Pi_n^{(\sigma^2,\gamma)}(q)$, by requiring

$$
z = z(\sigma_n^2, \gamma_n) \coloneqq z_n^* \quad \text{and} \quad \phi_j = \phi_j(\sigma_n^2, \gamma_n) \coloneqq \widetilde{\rho}_h(\Delta_n, \lceil \Delta_n^{-1/2}\chi_n \rceil), \tag{A.77}
$$

where z_n^* is defined in (A.23), with the argument p_n in $\mathcal{V}(z_n^*; \Delta_n, p_n)$ set as $\lceil \Delta_n^{-1/2}\chi_n \rceil$, and $\widetilde{\rho}_h$ is defined by (A.52), both appearing in the proof of Lemma A2. Because p_n satisfies $p_n \le K\Delta_n^{-1/2}\chi_n$, plus $(\sigma_n^2, \gamma_n) \in \Pi_n^{(\sigma^2,\gamma)}(q)$ and $\Delta_n^{-1/2}\chi_n \to \infty$, the results obtained in step 4 and step 5 of the proof of Lemma A2 hold. Of these, we use the expression z_n^* provided by (A.28) and the properties of $\widetilde{\mathcal{V}}$ given by (A.31) and (A.32). Because $(\gamma_n)_j = 0$ for all $j \ge p_n$ as required by (A.76), we have, as made obvious by step 2 of the proof of Lemma A2 and the definition of (z, ϕ) given in (A.77), that

$$
f(\lambda; z, \phi) = f(\lambda; \sigma_n^2, \gamma_n, \Delta_n). \tag{A.78}
$$

We define $C(z, \phi)$ as a $(q+2) \times (q+2)$ matrix given by

$$C(z,\phi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\partial \log f(\lambda; z, \phi)}{\partial (z, \phi)} \right)^{\intercal} \frac{\partial \log f(\lambda; z, \phi)}{\partial (z, \phi)} d\lambda. \tag{A.79}$$

We further partition matrix $C(z, \phi)$ into four submatrices:

$$C(z,\phi) = \begin{pmatrix} C_{zz} & C_{z\phi} \\ C_{\phi z} & C_{\phi\phi} \end{pmatrix}. \tag{A.80}$$

Here we require C_{zz} to be a scalar. Such partition leads to uniquely defined submatrices. We can write, for all $0 \le i, j \le q$,

$$\begin{aligned}
c(z,\phi)_{1,1} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\partial \log |1 - z e^{\mathbb{i}\lambda}|^2}{\partial z} \right)^2 d\lambda = \frac{2}{1 - z^2}, \\
c(z,\phi)_{1,j+2} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \log |1 - z e^{\mathbb{i}\lambda}|^2}{\partial z} \frac{\partial \log f(\lambda; \phi)}{\partial \phi_j} d\lambda, \\
c(z,\phi)_{i+2,j+2} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \log f(\lambda;\phi)}{\partial \phi_i} \frac{\partial \log f(\lambda;\phi)}{\partial \phi_j} d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{(2 - \delta_{i,0}) \cos i\lambda}{f(\lambda;\phi)} \frac{(2 - \delta_{j,0}) \cos j\lambda}{f(\lambda;\phi)} d\lambda.
\end{aligned}$$

Because of $f(\lambda; \phi) = \widetilde{\mathcal{V}}(e^{\mathbb{i}\lambda}; \Delta_n, q)$ as indicated by (A.77) and the properties of $\widetilde{\mathcal{V}}$ given by (A.31) and(A.32), we have

$$K^{-1} \chi_n^2 \le f(\lambda; \phi) \le K \chi_n^2 \quad \text{and} \quad \left| \frac{d^2 f(\lambda;\phi)}{d\lambda^2} \right| \le K \chi_n^2. \tag{A.81}$$

Therefore, in view of the expression of z_n^* provided by (A.28), all the entries of C can be rewritten, for all $0 \le i, j \le q$, as

$$C_{zz} = \frac{2}{1 - z^2}, \quad (C_{z\phi})_{j+1} = -\frac{2 z^j}{f(0;\phi)} + o(\chi_n^{-2}), \tag{A.82}$$

$$(C_{\phi\phi})_{i+1,j+1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{(2 - \delta_{i,0}) \cos i\lambda}{f(\lambda;\phi)} \frac{(2 - \delta_{j,0}) \cos j\lambda}{f(\lambda;\phi)} d\lambda. \tag{A.83}$$

We introduce an auxiliary $(q+1) \times (q+1)$ matrix $\widetilde{C}_{\phi\phi}$:

$$(\widetilde{C}_{\phi\phi})_{i+1,j+1} := \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(\lambda;\phi) \cos i\lambda \cos j\lambda d\lambda. \tag{A.84}$$

In view of the expressions of $C_{\phi\phi}$ provided by (A.83), we have

$$(C_{\phi\phi} \widetilde{C}_{\phi\phi})_{i+1,j+1} = \frac{(2 - \delta_{i,0})}{(2\pi)^2} \sum_{k=-q}^{q} \int_{-\pi}^{\pi} f^2(\lambda;\phi) \cos i\lambda e^{\mathbb{i}k\lambda} d\lambda \int_{-\pi}^{\pi} \frac{\cos j\lambda'}{f^2(\lambda';\phi)} e^{-\mathbb{i}k\lambda'} d\lambda'.$$

It is straighforward to calculate, using the orthogonality of complex exponentials, that

$$\frac{(2-\delta_{i,0})}{(2\pi)^2}\sum_{k=-\infty}^{\infty}\int_{-\pi}^{\pi} f^2(\lambda;\phi)\cos i\lambda e^{\mathrm{i}k\lambda}d\lambda\int_{-\pi}^{\pi}\frac{\cos j\lambda'}{f^2(\lambda';\phi)}e^{-\mathrm{i}k\lambda'}d\lambda' = \delta_{i,j}.$$

On the other hand, according to (A.81) and the proof of Theorem II.4.7 of **?**, we have for (z,ϕ) satisfying (A.77), under $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q) : n\geq 1\}$ and $\Delta_n^{-1/2}\chi_n\to\infty$, and for $k\geq q+1$ and $0\leq i\leq q$,

$$\left|\chi_n^{-4}\int_{-\pi}^{\pi} f^2(\lambda;\phi)\cos i\lambda e^{\mathrm{i}k\lambda}d\lambda\right| + \left|\chi_n^{4}\int_{-\pi}^{\pi} f^{-2}(\lambda;\phi)\cos i\lambda e^{\mathrm{i}k\lambda}d\lambda\right| \leq \frac{K}{(k-i)^2}.$$

We hence are able to write $C_{\phi\phi}\widetilde{C}_{\phi\phi} = \mathbb{I}_{q+1}+A$, with $|A_{i,j}|\leq K\frac{q+1-i\vee j}{(q+1-i)^2(q+1-j)^2}$, which immediately gives

$$C_{\phi\phi}^{-1} = \widetilde{C}_{\phi\phi} - \widetilde{C}_{\phi\phi}(\mathbb{I}_{q+1}+A)^{-1}A. \tag{A.85}$$

Because of (A.81) and Proposition 4.5.3 in Brockwell and Davis (1991), we have $\chi_n^4 C_{\phi\phi}\sim\mathbb{I}_{q+1}$ and $\chi_n^{-4}\widetilde{C}_{\phi\phi}\sim\mathbb{I}_{q+1}$, where $\sim$ is defined based on Loewner partial order. We hence have $(\mathbb{I}_{q+1}+A)(\mathbb{I}_{q+1}+A)^\intercal\sim\mathbb{I}_{q+1}$ and therefore $((\mathbb{I}_{q+1}+A)(\mathbb{I}_{q+1}+A)^\intercal)^{-1}\sim\mathbb{I}_{q+1}$. This allows us to conclude that

$$\|\chi_n^{-4}\widetilde{C}_{\phi\phi}(\mathbb{I}_{q+1}+A)^{-1}\|\leq K,$$

where $\|\cdot\|$ stands for the matrix operator norm. We hence have, by Cauchy-Schwarz inequality,

$$\sum_{i=0}^{q}|\chi_n^{-4}(\widetilde{C}_{\phi\phi}(\mathbb{I}_{q+1}+A)^{-1}A)_{i+1,j+1}|\leq K\Big(q\sum_{i=0}^{q}A_{i+1,j+1}^2\Big)^{1/2}\leq\frac{K\sqrt{q}}{(q+1-j)^2}. \tag{A.86}$$

Because of (A.80), block matrix inversion states that for $0\leq i,j\leq q$,

$$\begin{aligned}
C^{-1}(z,\phi)_{1,1} &= (C_{zz}-C_{z\phi}C_{\phi\phi}^{-1}C_{\phi z})^{-1},\\
C^{-1}(z,\phi)_{1,j+2} &= -(C_{zz}-C_{z\phi}C_{\phi\phi}^{-1}C_{\phi z})^{-1}(C_{z\phi}C_{\phi\phi}^{-1})_{j+1},\\
C^{-1}(z,\phi)_{i+2,j+2} &= (C_{\phi\phi}^{-1})_{i+1,j+1}+(C_{z\phi}C_{\phi\phi}^{-1})_{i+1}(C_{zz}-C_{z\phi}C_{\phi\phi}^{-1}C_{\phi z})^{-1}(C_{z\phi}C_{\phi\phi}^{-1})_{j+1}.
\end{aligned}$$

We therefore have obtained all the elements needed for calculation of $C^{-1}(z,\phi)$. Indeed, we have C_{zz} and $C_{z\phi}$, provided by (A.82), and $C_{\phi\phi}^{-1}$, provided by (A.85) and (A.86). It is straightforward to calculate that for (z,ϕ) satisfying (A.77) and under $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q) : n\geq 1\}$ and $\Delta_n^{-1/2}\chi_n\to\infty$,

$$\begin{aligned}
C^{-1}(z,\phi)_{1,1} &= (C_{zz}-C_{z\phi}C_{\phi\phi}^{-1}C_{\phi z})^{-1} = \frac{1-z^2}{2z^{2q}}+o(\Delta_n^{1/2}\chi_n^{-1}),\\
C^{-1}(z,\phi)_{1,j+2} &= \frac{1-z^2}{2z^{2q}}f(0;\phi)z^j+O\Big(\frac{\Delta_n^{1/2}\chi_n^{-1}\sqrt{q}}{(q+1-j)^2}\Big),
\end{aligned}$$

$$C^{-1}(z,\phi)_{i+2,j+2} \;=\; \widetilde{C}_{\phi\phi} - \widetilde{C}_{\phi\phi}(\mathbb{I}_{q+1}+A)^{-1}A + \frac{1-z^2}{2z^{2q}}f^2(0;\phi)z^{i+j} + O\Big(\frac{\Delta_n^{1/2}\chi_n^{-1}\sqrt{q}}{(q+1-i\vee j)^2}\Big).$$

Now we move on to calculate $C^{-1}(\sigma^2,\gamma)$. In view of the definition of c specified in step 1, we realize that

$$C(\sigma^2,\gamma) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big(\frac{\partial \log f(\lambda;\sigma^2,\gamma,\Delta_n)}{\partial(\sigma^2,\gamma)}\Big)^{\intercal}\frac{\partial \log f(\lambda;\sigma^2,\gamma,\Delta_n)}{\partial(\sigma^2,\gamma)}d\lambda. \tag{A.87}$$

According to the inverse function theorem, plus the equality $f(\lambda;z,\phi) = f(\lambda;\sigma^2,\gamma,\Delta_n)$ from (A.78) and the definition of $C(z,\phi)$ specified in (A.79), we have that under the restriction (A.76),

$$C^{-1}(\sigma_n^2,\gamma_n) = \frac{\partial(\sigma_n^2,\gamma_n)}{\partial(z,\phi)}C^{-1}(z,\phi)\Big(\frac{\partial(\sigma_n^2,\gamma_n)}{\partial(z,\phi)}\Big)^{\intercal}.$$

The mapping between (σ_n^2,γ_n) and (z,ϕ) specified in (A.77) indicates that

$$\frac{\partial\sigma_n^2}{\partial z} = -2(1-z)f(0;\phi), \quad \frac{\partial\sigma_n^2}{\partial\phi_k} = (2-\delta_{k,0})(1-z)^2,$$

$$\frac{\partial(\gamma_n)_j}{\partial\phi_k} = z\delta_{k,j} - (1-z)^2(k-j)_+, \quad \frac{\partial(\gamma_n)_j}{\partial z} = \phi_j + (1-z)\sum_{k=0}^{q}\phi_k(k-j)_+.$$

Here and below, we use $x_+ = \max\{x,0\}$. Some algebra yields that under $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q): n\geq 1\}$, $\Delta_n^{-1/2}\chi_n\to\infty$, and the restriction (A.76), and for all $0\leq j\leq q$,

$$C^{-1}(\sigma_n^2,\gamma_n)_{1,1} \;=\; 4\Delta_n^{-2}(1-z)^3f^2(0;\phi) + 2\Delta_n^{-2}(1-z)^4qf^2(0;\phi) + o\Big(\Delta_n^{-1/2}\chi_n + q\Big), \tag{A.88}$$

$$\begin{aligned} C^{-1}(\sigma_n^2,\gamma_n)_{1,j+2} \;=\;& -\Delta_n^{-1}(1-z)^2f^2(0;\phi)\Big(1+2(1-z)(q-j)+\frac{1}{2}(1-z)^2(q-j)^2\Big)\\ &+O\Big(\chi_n^2(j+1)^{-2}\Big) + o\Big(\chi_n^2+\Delta_n(q-j)^2\Big), \end{aligned} \tag{A.89}$$

$$\begin{aligned} C^{-1}(\sigma_n^2,\gamma_n)_{i+2,j+2} \;=\;& \big(1+o(1)\big)\Big[\frac{1}{2\pi}\int_{-\pi}^{\pi}f^2(\lambda;\phi)\cos i\lambda\cos j\lambda d\lambda\\ &+f^2(0;\phi)\Big(\frac{1}{2}(1-z)^4\sum_{k=0}^{q}(k-i)_+(k-j)_+ - \frac{1}{2}z(1-z)^2|i-j|\\ &+(1-z^2)\big(z+(q-j+1)(1-z)\big)\big(z+(q-i+1)(1-z)\big)\Big)\Big]. \end{aligned} \tag{A.90}$$

Step 5. (Properties of C^{-1}: General case) This step shows that the restriction (A.76) is not needed to obtain (A.88), (A.89), and (A.90), given $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q): n\geq 1\}$ and $\Delta_n^{-1/2}\chi_n\to\infty$. Clearly, we only need to consider the case in which $q > \lceil\Delta_n^{-1/2}\chi_n\rceil$, otherwise the restiction (A.76) is automatically satisfied from the definition of $\Pi_n^{(\sigma^2,\gamma)}(q)$. Using the fact that $\frac{\partial f(\lambda;\sigma^2,\gamma,\Delta_n)}{\partial(\sigma^2,\gamma)}$ does not depend on (σ^2,γ), we can write

$$C(\sigma^2,\gamma) \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big(\frac{\partial\log f(\lambda;\sigma^2,\widetilde{\gamma}(p,\gamma),\Delta_n)}{\partial(\sigma^2,\gamma)}\Big)^{\intercal}\frac{\partial\log f(\lambda;\sigma^2,\widetilde{\gamma}(p,\gamma),\Delta_n)}{\partial(\sigma^2,\gamma)}$$

$$\times\left[1+\sum_{j=1}^{\infty}\Big(-\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}\Big)^j\right]^2 d\lambda. \quad \text{(A.91)}$$

Here we recall that $\widetilde{\gamma}(p,\gamma)=(\gamma_0,\gamma_1,\dots,\gamma_p,0,\dots,0)^\intercal$, $\mathcal{V}(z;\Delta_n,p)$, and $\mathcal{V}(z;\Delta_n,-p)$ are all introduced in step 2 of the proof of Lemma A2 and we use (A.14). We observe the apparent fact that without the multiplicative term $\Big[1+\sum_{j=1}^{\infty}\big(-\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}\big)^j\Big]^2$, the right-hand side of (A.91) would just be $C(\sigma^2,\widetilde{\gamma}(p,\gamma))$. We can then conclude, following the proof of Proposition 4.5.3 in Brockwell and Davis (1991) and using the bound on $\Big|\frac{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,-p)}{\mathcal{V}(e^{\mathbb{i}\lambda};\Delta_n,p)}\Big|$ provided by (A.42), that

$$(1-Kp^{-2})C(\sigma^2,\widetilde{\gamma}(p,\gamma))\le C(\sigma^2,\gamma)\le(1+Kp^{-2})C(\sigma^2,\widetilde{\gamma}(p,\gamma)), \quad \text{(A.92)}$$

where $\le$ is the Loewner partial order. We let $p_n=\lceil\Delta_n^{-1/2}\chi_n\rceil$. Then $\widetilde{\gamma}(p_n,\gamma_n)$ satisfies (A.76). Since $\Delta_n^{-1/2}\chi_n\to\infty$, we have $p_n\to\infty$ and hence $\{(\sigma_n^2,\widetilde{\gamma}(p_n,\gamma_n))\in\Pi_n^{(\sigma^2,\gamma)}(q):n\ge1\}$ and $\Delta_n^{-1}\chi^2(\sigma_n^2,\widetilde{\gamma}(p_n,\gamma_n),\Delta_n)\to\infty$. Therefore, the expressions (A.88), (A.89), and (A.90) in step 4 apply to $C^{-1}(\sigma_n^2,\widetilde{\gamma}(p_n,\gamma_n))$. If we further apply (A.92), plus that $f(\lambda;\widetilde{\gamma}(p_n,\gamma_n))/f(\lambda;\gamma_n)\to1$ uniformly over λ, we have $C(\sigma^2,\gamma)^{-1}=C(\sigma^2,\widetilde{\gamma}(p,\gamma))^{-1}\sum_{j=0}^{\infty}\big(\big[C(\sigma^2,\gamma)-C(\sigma^2,\widetilde{\gamma}(p,\gamma))\big]C(\sigma^2,\gamma)^{-1}\big)^j$ and direct calculation leads to the fact that (A.88), (A.89), and (A.90) indeed hold, for all parameter sequences $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q):n\ge1\}$ satisfying $\Delta_n^{-1/2}\chi_n\to\infty$.

Step 6. (Bound on $\sigma^{(n)}(q_n)^2-C_T$) Given the relation (A.68) and $\Delta_n^{-1/2}\iota^{(n)}\to\infty$, we immediately obtain $\Delta_n^{-1}\chi^2(\sigma^{(n)}(q)^2,\gamma^{(n)}(q),\Delta_n)\to\infty$. Also, we have $(\sigma^{(n)}(q)^2,\gamma^{(n)}(q))\in\Pi_n^{(\sigma^2,\gamma)}(q)$ by definition. Thus, in view of step 5, the relations (A.88), (A.89), and (A.90) apply to $C^{-1}(\sigma^{(n)}(q)^2,\gamma^{(n)}(q))$ (of course, now with $z=z(\sigma^{(n)}(q)^2,\gamma^{(n)}(q))$ and $\phi_j=\phi_j(\sigma^{(n)}(q)^2,\gamma^{(n)}(q))$). Using the expression of z_n^* given by (A.23) (see explanation after (A.77)), the bound on $f(\lambda;\phi)$ provided by (A.81), and the relation (A.68), we can write for $0\le j\le q$,

$$\begin{aligned}
C^{-1}(\sigma^{(n)}(q)^2,\gamma^{(n)}(q))_{1,1} &= 4\Delta_n^{-1/2}C_T^{3/2}f^{1/2}(0;\phi)+2C_T^2q+o\Big(\Delta_n^{-1/2}\iota^{(n)}+q\Big), && \text{(A.93)}\\
C^{-1}(\sigma^{(n)}(q)^2,\gamma^{(n)}(q))_{1,j+2} &= -C_Tf(0;\phi)\Big(1+2(1-z)(q-j)+\frac{1}{2}(1-z)^2(q-j)^2\Big)\\
&\quad +O\Big((\iota^{(n)})^2(j+1)^{-2}\Big)+o\Big((\iota^{(n)})^2+\Delta_n(q-j)^2\Big), && \text{(A.94)}\\
C^{-1}(\sigma^{(n)}(q)^2,\gamma^{(n)}(q))_{i+2,j+2} &= (1+o(1))\Big[\frac{1}{2\pi}\int_{-\pi}^{\pi}f^2(\lambda;\phi)\cos i\lambda\cos j\lambda d\lambda\\
&\quad +f^2(0;\phi)\Big(\frac{1}{2}(1-z)^4\sum_{k=0}^{q}(k-i)_+(k-j)_+-\frac{1}{2}z(1-z)^2|i-j|\\
&\quad +(1-z^2)\big(z+(q-j+1)(1-z)\big)\big(z+(q-i+1)(1-z)\big)\Big)\Big]. && \text{(A.95)}
\end{aligned}$$

In view of the bounds on c provided by (A.73), (A.74), and (A.72) and the relation (A.68), we have for all $2 \le i \le q+2$ and $k \ge q+3$,

$$|C^{-1}_{1,1}c_{1,k}|+\sum_{j=2}^{q+2}|C^{-1}_{1,j}c_{j,k}| \lesssim (\iota^{(n)})^{-2}, \quad |C^{-1}_{i,1}|\Big(\sum_{l=q+3}^{\infty}|c_{1,l}|^2\Big)^{1/2}+\sum_{j=2}^{q+2}|C^{-1}_{i,j}|\Big(\sum_{l=q+3}^{\infty}|c_{j,l}|^2\Big)^{1/2} \lesssim \frac{\Delta_n^{3/4}q}{(\iota^{(n)})^{3/2}}+1,$$

where C^{-1} and c are evaluated at $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$. Subsituting this result back into (A.63) immediately proves the lemma. ∎

Lemma A6. *Suppose Assumptions 1 - 4 hold. For all sequences $\{q_n\}$ and under $\Delta_n^{-1/2}\iota^{(n)} \le K$, it holds that for all $0 \le j \le q_n$,*

$$\big|\sigma^{(n)}(q_n)^2 - C_T\big| \le K\Delta_n^{-1}\big\|\gamma^{(n)}\big\|_{1,(q_n)}, \quad \big|\gamma^{(n)}(q_n)_j - \gamma_j^{(n)}\big| \lesssim (q_n+1)\big\|\gamma^{(n)}\big\|_{(q_n)}.$$

Proof. Step 1. (Characterization of $\sigma^{(n)}(q_n)^2 - C_T$) Throughout the proof, we omit writing the subscript $_n$ of q_n. We set the bijection β_n as

$$\beta_n(\sigma^2,\gamma)_j = \frac{\Delta_n^{-1}}{2\pi}\int_{-\pi}^{\pi} f(\lambda;\sigma^2,\gamma,\Delta_n)e^{\mathrm{i}j\lambda}d\lambda, \quad 0 \le j \le q+1. \tag{A.96}$$

In view of (A.96), we have $\sigma^{(n)}(q)^2 = \sum_{j=-q-1}^{q+1}\beta^{(n)}(q)_{|j|}$ and $\gamma^{(n)}(q)_j = -\Delta_n\sum_{i=j+1}^{q+1}(i-j)\beta^{(n)}(q)_i$. The current lemma, therefore, would follow from

$$\|\beta^{(n)}(q) - \bar{\beta}^{(n)}\|_1 \le K\Delta_n^{-1}\|\gamma^{(n)}\|_{(q)}. \tag{A.97}$$

Trivially, (A.97) holds under $\Delta_n^{-1}\|\gamma^{(n)}\|_{1,(q_n)} \ge \frac{1}{K}$, given $\Delta_n^{-1/2}\iota^{(n)} \le K$ and Assumption 4. The subsequence argument indicates that we only need to consider the case in which $\|\gamma^{(n)}\|_{1,(q)} = o(\Delta_n)$. In view of the definitions of $\bar{L}_n^{\star}(\beta)$, $\beta^{(n)}(q)$, and $\Pi_n^{\beta}(q)$, and Assumption 4, we have for all $0 \le j \le q+1$,

$$\frac{\partial}{\partial\beta_j}\bar{L}_n^{\star}(\beta^{(n)}(q)) = 0. \tag{A.98}$$

On the other hand, using (A.96), we obtain that for all $0 \le j \le q+1$,

$$-\frac{2}{n_T}\frac{1}{2-\delta_{0,j}}\frac{\partial\bar{L}_n^{\star}(\beta)}{\partial\beta_j} = A(\beta)_j + B(\beta)_j, \tag{A.99}$$

where $A(\beta)_j = \sum_{k=0}^{q+1}(2-\delta_{k,0})c(\beta)_{j,k}(\bar{\beta}_k^{(n)} - \beta_k)$ and $B(\beta)_j = \sum_{k=q+2}^{\infty}2c(\beta)_{j,k}\bar{\beta}_k^{(n)}$. Here we use the shorthand notation $c(\beta)_{j,k} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\cos j\lambda\cos k\lambda}{f^2(\lambda;\beta)}d\lambda$, where $f(\lambda;\beta) = \sum_{j=-\infty}^{\infty}\beta_{|j|}e^{\mathrm{i}j\lambda}$. In view of (A.99), if we let $C(\beta)$ be the $(q+2)\times(q+2)$ matrix whose entries are $C(\beta)_{i,j} = c(\beta)_{i,j}$ for all

$0 \le i, j \le q+1$, the first-order condition (A.98) can be rewritten as

$$\bar{\beta}_j^{(n)} - \beta^{(n)}(q)_j = \frac{2}{2-\delta_{j,0}} \sum_{i=0}^{q+1} C^{-1}(\beta^{(n)}(q))_{i,j} \sum_{k=q+2}^{\infty} c(\beta^{(n)}(q))_{i,k} \beta_k^{(n)}. \tag{A.100}$$

Step 2. (Properties of c and C^{-1}) In this step we provide bounds on c and C^{-1}, from which we would immediately prove the current lemma. We first note the connection between $\beta^{(n)}(q)$ and $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))$ and that $(\sigma^{(n)}(q)^2, \gamma^{(n)}(q)) \in \Pi_n^{(\sigma^2,\gamma)}(q)$. Then according to (A.18) and (A.22), Assumption 4 and the construction of $\Pi_n^{(\sigma^2,\gamma)}(q)$, which we recall comes from (3.11), indicate that under $\Delta_n^{-1/2}\iota^{(n)} \le K$ and $\|\gamma^{(n)}\|_{1,(q)} = o(\Delta_n)$, for n large enough and uniformly over λ,

$$f(\lambda; \beta^{(n)}(q)) \sim 1 \quad \text{and} \quad \frac{d^2 f(\lambda; \beta^{(n)}(q))}{d\lambda^2} \le K. \tag{A.101}$$

An immediate result of (A.101) is that, following the proof of Theorem II.4.7 of **?** , the definition of $c(\beta)_{j,k}$ as Fourier coefficients of $f^{-2}(\lambda;\beta)$ indicates that

$$|c(\beta^{(n)}(q))_{j,k}| \le K(|j-k|+1)^{-2}. \tag{A.102}$$

Now we analyze C^{-1}. We introduce an auxiliary $(q+2)\times(q+2)$ matrix $\widetilde{C}(\beta)$:

$$\widetilde{C}(\beta)_{i,j} := \frac{1}{2\pi}\int_{-\pi}^{\pi} f^2(\lambda;\beta)(2-\delta_{i,0})(2-\delta_{j,0})\cos i\lambda \cos j\lambda d\lambda, \quad \text{with} \quad 0 \le i, j \le q+1.$$

If we compare (A.101) with (A.81), we can repeat the derivation of (A.85) and (A.86) and conclude that $\sum_{j=0}^{q+1} |\widetilde{C}(\beta^{(n)}(q))| \le K$ for all $0 \le i \le q+1$ and that

$$C^{-1}(\beta^{(n)}(q)) = \widetilde{C}(\beta^{(n)}(q)) - \widetilde{C}(\beta^{(n)}(q))(\mathbb{I}_{q+2} + A)^{-1}A, \tag{A.103}$$

where the $(q+2)\times(q+2)$ matrix A satisfies $\sum_{i=0}^{q+1} |(\widetilde{C}(\beta^{(n)}(q))(\mathbb{I}_{q+1}+A)^{-1}A)_{i,j}| \le \frac{K\sqrt{q}}{(q+2-j)^2}$. We then immediately have for all $0 \le i \le q+1$,

$$\sum_{j=0}^{q+1} |C^{-1}(\beta^{(n)}(q))_{i,j}| \le K. \tag{A.104}$$

Applying Hölder's inequality to the expression of $\beta_j^{(n)} - \beta^{(n)}(q)_j$ provided in (A.100) leads to

$$\sum_{j=0}^{q+1} |\bar{\beta}_j^{(n)} - \beta^{(n)}(q)_j| \le 2\|\bar{\beta}^{(n)}\|_{(q+1)} \sum_{j=0}^{q+1}\sum_{i=0}^{q+1} |C^{-1}(\beta^{(n)}(q))_{i,j}| \Big(\sum_{k=q+2}^{\infty} |c(\beta^{(n)}(q))_{i,k}|^2 \Big)^{1/2}.$$

Therefore, in view of (A.102) and (A.104), we have already proved (A.97), and thereby the current lemma, by observing that $\|\bar{\beta}^{(n)}\|_{1,(q+1)} = O(\Delta_n^{-1}\|\gamma^{(n)}\|_{1,(q)})$. ∎

Lemma A7. *Suppose Assumptions 1 - 4 hold and $q_n \le K n^{1/3}$. It holds that under either $n^{1/2}\iota^{(n)} \to \infty$ or $n^{1/2}\iota^{(n)} \le K$,*

$$\bar{\eta}^\intercal \Xi_n(\beta^{(n)}(q_n)) - \bar{\eta}^\intercal \Xi_{A,n}(\beta^{(n)}(q_n)) = o_{\mathrm{P}}\left(n^{-1/2}\sqrt{q_n+1} + n^{-1/4}\sqrt{\iota^{(n)}}\right). \tag{A.105}$$

Proof. Step 1. (Main proof) We only consider the case in which $\iota^{(n)} \ge K^{-1}$, as the problem gets harder as noise becomes larger. Intuitively, when noise becomes small enough ($\Delta_n^{-1}(\iota^{(n)})^2 \le K$), the data-generating process is the same as that of classic time-series models. In this case, (A.105) becomes

$$\bar{\eta}^\intercal \Xi_n(\beta^{(n)}(q_n)) - \bar{\eta}^\intercal \Xi_{A,n}(\beta^{(n)}(q_n)) = o_{\mathrm{P}}(n^{-1/4}). \tag{A.106}$$

First, we define Ω_n' as the set of all ω such that $K^{-1} \le n\Delta_n \le K$ (it should not be confused with the matrix Ω_n) and observe that

$$n^{-1}n_t = \frac{1}{T}\int_0^t \xi_s^{-1} ds + o_{\mathrm{P}}(1) \quad \text{and} \quad \lim_{n\to\infty} \mathbb{P}(\Omega_n') = 1, \tag{A.107}$$

which are direct results of Lemma 14.1.5 of ? and Assumption 2. Then we let the bijection β_n be the identity function. For this choice of β_n, we have $\partial\sigma_n^2 = (1, 0_{q+1})$. Moreover, we observe that $\partial\bar{\Xi}_n^\star(\beta^{(n)}(q_n)) = 2n_T^{-1} n C(\sigma^{(n)}(q_n), \gamma^{(n)}(q_n))$, where the matrix C is introduced in (A.75) and satisfies (A.87). In particular, (A.93) and (A.94) indicate that in restriction to Ω_n' we have $|\partial\bar{\Xi}_n^\star(\beta^{(n)}(q_n))_{1,1}^{-1}| \le K n^{1/2}$ and $|\partial\bar{\Xi}_n^\star(\beta^{(n)}(q_n))_{1,j}^{-1}| \le K$ for $2 \le j \le q_n+2$. Therefore, according to (A.107), for showing (A.106) it is sufficient to prove that

$$\mathbb{E}\left|\mathbb{1}_{\Omega_n'}\left(\Xi_n(\beta^{(n)}(q_n))_1 - \Xi_{A,n}(\beta^{(n)}(q_n))_1\right)\right| = o(n^{-3/4}), \tag{A.108}$$

and for each $2 \le j \le q_n + 2$,

$$\mathbb{E}\left|\mathbb{1}_{\Omega_n'}\left(\Xi_n(\beta^{(n)}(q_n))_j - \Xi_{A,n}(\beta^{(n)}(q_n))_j\right)\right| = o(n^{-3/4}). \tag{A.109}$$

Now we show (A.108). The same reasoning proves (A.109). Below, we suppress the dependence of $\beta^{(n)}$, $\sigma^{(n)}$, and $\gamma^{(n)}$ on q_n. We can rewrite (A.108) as

$$\mathbb{E}\Big|\underbrace{\mathbb{1}_{\Omega_n'}\frac{\partial}{\partial\beta_1}\log\det(\Sigma_n(\beta^{(n)})\Omega_n(\beta^{(n)})^{-1})}_{=:\mathcal{R}_a} + n^{-1/4}\underbrace{\mathbb{1}_{\Omega_n'}Y_n^\intercal\frac{\partial}{\partial\beta_1}(\Sigma_n(\beta^{(n)})^{-1} - \Omega_n(\beta^{(n)})^{-1})Y_n}_{=:\mathcal{R}_b}\Big| = o(n^{1/4}). \tag{A.110}$$

We hence only need prove $\mathbb{E}|\mathcal{R}_a| = o(n^{1/4})$ and $\mathbb{E}|\mathcal{R}_b| = o(n^{1/4})$. Observe that $\frac{\partial}{\partial\beta_1}\Sigma_n(\beta) = \frac{\partial}{\partial\beta_1}\Omega_n(\beta) = \Delta_n$. Then we can write

$$\mathcal{R}_a = \mathbb{1}_{\Omega_n'}\Delta_n \mathrm{tr}(\Sigma_n(\beta^{(n)})^{-1} - \Omega_n(\beta^{(n)})^{-1}), \quad \mathcal{R}_b = -\mathbb{1}_{\Omega_n'}\Delta_n Y_n^\intercal(\Sigma_n(\beta^{(n)})^{-2} - \Omega_n(\beta^{(n)})^{-2})Y_n. \tag{A.111}$$

Here we use $\log\det A = \operatorname{tr}\log A$. The challenge we face is that we do not have an analytical expression for Σ_n^{-1}. However, we observe that

$$\Sigma_n^{-1} = \Omega_n^{-1} - \Omega_n^{-1}R_n\Omega_n^{-1} + \Omega_n^{-1}R_n\Sigma_n^{-1}R_n\Omega_n^{-1}, \quad \text{with} \quad R_n(\beta) := \Sigma_n(\beta) - \Omega_n(\beta). \tag{A.112}$$

Although in the last term on the right-hand side, Σ_n^{-1} still appears; later we show that we can replace it with Ω_n^{-1} for the purpose of bounding $\mathcal{R}_a$ and $\mathcal{R}_b$. Now we apply (A.112) to (A.111). Introduce simplifying notation

$$\mathcal{R}_{a1}(\beta) := \operatorname{tr}(\Omega_n^{-1}R_n\Omega_n^{-1}), \quad \mathcal{R}_{a2}(\beta) := \operatorname{tr}(\Omega_n^{-1}R_n\Sigma_n^{-1}R_n\Omega_n^{-1}), \quad \mathcal{R}_{b1}(\beta) := Y_n^\intercal\Omega_n^{-1}R_n\Omega_n^{-2}Y_n,$$

$$\mathcal{R}_{b2}(\beta) := Y_n^\intercal\Omega_n^{-1}R_n\Sigma_n^{-1}R_n\Omega_n^{-2}Y_n, \quad \text{and} \quad \mathcal{R}_{b3}(\beta) := Y_n^\intercal\Omega_n^{-1}R_n\Sigma_n^{-2}R_n\Omega_n^{-1}Y_n. \tag{A.113}$$

Here we drop the argument β of Ω_n and Σ_n. Then we can rewrite (A.111) as

$$\Delta_n^{-1}\mathcal{R}_a = \mathbb{1}_{\Omega_n'}\big(-\mathcal{R}_{a1}(\beta^{(n)}) + \mathcal{R}_{a2}(\beta^{(n)})\big), \quad \Delta_n^{-1}\mathcal{R}_b = \mathbb{1}_{\Omega_n'}\big(2\mathcal{R}_{b1}(\beta^{(n)}) - 2\mathcal{R}_{b2}(\beta^{(n)}) - \mathcal{R}_{b3}(\beta^{(n)})\big).$$

In view of the triangle inequality, the desired result (A.110) follows from the fact that for all $A \in \{\mathcal{R}_{a1}, \mathcal{R}_{a2}, \mathcal{R}_{b1}, \mathcal{R}_{b2}, \mathcal{R}_{b3}\}$, $\mathbb{E}|A(\beta^{(n)})| = o(n^{5/4})$.

Step 2. (Bounds of Σ) In this step we prove

$$\Sigma_n((\sigma^{(n)})^2, \gamma_0^{(n)}) \leq K\Sigma_n((\sigma^{(n)})^2, \gamma^{(n)}). \tag{A.114}$$

Namely, we bound $\Sigma_n((\sigma^{(n)})^2, \gamma^{(n)})$ from below by $\Sigma_n((\sigma^{(n)})^2, \gamma_0^{(n)})$. For all $x = (x_1, x_2, \ldots, x_{n_T})^\intercal \in \mathbb{R}^{n_T}$, define $\widetilde{x} = (\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_{n_T+1})^\intercal \in \mathbb{R}^{n_T+1}$ by $\widetilde{x}_j = x_{j-1} - x_j$ with $x_0 = x_{n_T+1} = 0$. We deduce (A.114) from

$$x^\intercal\Sigma_n((\sigma^{(n)})^2, \gamma^{(n)})x = (\sigma^{(n)})^2\Delta_n\|x\|^2 + \widetilde{x}^\intercal\Gamma(\gamma^{(n)})_{n+1}\widetilde{x} \quad \text{and} \quad \Gamma(\gamma^{(n)})_{n+1} \sim \gamma_0^{(n)}\mathbb{I}_{n+1}. \tag{A.115}$$

Here we define $\Gamma(\gamma^{(n)})_{n+1} \in \mathcal{M}_{n_T+1}$ by $(\Gamma(\gamma^{(n)})_{n+1})_{ij} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda; \gamma^{(n)})e^{\mathrm{i}(i-j)\lambda}d\lambda$. $\mathbb{I}_{n+1}$ is the $(n_T+1)\times(n_T+1)$ identity matrix. The first claim in (A.115) holds by definition of Σ_n. The second claim in (A.115) comes from Proposition 4.5.3 in Brockwell and Davis (1991) by applying (A.18) and (A.22).

Step 3. (Useful estimates) We provide some estimates in this step. We start by introducing some notation:

$$A(1)_j := (\Omega_n^{-1}Y_n)_j, \quad A(2)_{ij} := (\Omega_n^{-1})_{i,j} - (\Omega_n^{-1})_{i,j+1}, \quad \text{and} \quad A(3)_{ij} := A(2)_{ij} - A(2)_{i+1,j}.$$

We have the following four estimates:

$$\begin{cases} \mathbb{E}(\mathbb{1}_{\Omega'_n} A(1)^2_{h_n}) \le K h_n, & \mathbb{E}(\mathbb{1}_{\Omega'_n} (A(1)_{h_n} - A(1)_{h_n+1})^2) \le K, \\ |\mathbb{1}_{\Omega'_n} (\Omega_n^{-1})_{1h_n}| \le K, & |\mathbb{1}_{\Omega'_n} A(3)_{h_n,l_n}| \le K n^{-1/2} + K \mathbb{1}_{\{|h_n - l_n| \le 2\}}. \end{cases}$$

All of the estimates are direct results of Lemma A2. Note $A(3)_{i,j}$ is a linear combination of four entries of Ω_n^{-1}. Due to such a combination, for $|h_n - l_n| \ge 3$, the magnitude of $A(3)_{h_n,l_n}$ is reduced by a factor of n compared with $(\Omega_n^{-1})_{i,j} \sim \sqrt{n}$.

Step 4. (Bound on $\mathcal{R}_{b2}(\beta^{(n)})$) In this step we show that $\mathbb{E}|\mathcal{R}_{b2}(\beta^{(n)})| = o(n^{5/4})$. We obtain results in other cases following the same reasoning. Note that Σ_n is positive definite, as it is a covariance matrix. Then the definition of $\mathcal{R}_{b2}(\beta)$ given by (A.113) and the Cauchy-Schwarz inequality indicate that $\mathbb{E}|\mathcal{R}_{b2}(\beta^{(n)})| = o(n^{5/4})$ follows from

$$\mathbb{E}\big|\mathbb{1}_{\Omega'_n} Y_n^\intercal \Omega_n^{-1} R_n \Sigma_n^{-1} R_n \Omega_n^{-1} Y_n\big| = o(n^{1/4}) \quad \text{and} \quad \mathbb{E}\big|\mathbb{1}_{\Omega'_n} Y_n^\intercal \Omega_n^{-2} R_n \Sigma_n^{-1} R_n \Omega_n^{-2} Y_n\big| = o(n^{9/4}). \tag{A.116}$$

Here Ω_n, R_n, and Σ_n are evaluated at $\beta^{(n)}$. Now we show the first claim in (A.116). The second comes from the same reasoning. In Step 1 we state that we can replace Σ_n^{-1} with Ω_n^{-1}. Obviously, we can do so if $\Sigma_n^{-1} \le K\Omega_n^{-1}$. Here $\le$ stands for Loewner partial order. Indeed, Corollary 7.7.4 in **?** states that for any two Hermitian matrices A and B, if $A < B$, then $A^{-1} > B^{-1}$. Hence, we only need $\Omega_n \le K\Sigma_n$. On the one hand, we have (A.114). On the other hand, in view of the definitions of Ω_n and V_n, we conclude that

$$\Omega_n((\sigma^{(n)})^2, \gamma^{(n)}) \lesssim O_n V_n((\sigma^{(n)})^2, \gamma_0^{(n)}) O_n = \Omega_n((\sigma^{(n)})^2, \gamma_0^{(n)}) = \Sigma_n((\sigma^{(n)})^2, \gamma_0^{(n)}). \tag{A.117}$$

Here the first equality can be verified using Lemma A1. Given (A.114) and (A.117), we have $\Omega_n((\sigma^{(n)})^2, \gamma^{(n)}) \le K\Sigma_n((\sigma^{(n)})^2, \gamma^{(n)})$. Therefore, the first claim in (A.116) follows from

$$\mathbb{E}\big|\mathbb{1}_{\Omega'_n} Y_n^\intercal \Omega_n^{-1} R_n \Omega_n^{-1} R_n \Omega_n^{-1} Y_n\big| = o(n^{1/4}), \tag{A.118}$$

with Ω_n, R_n and Σ_n evaluated at $\beta^{(n)}$.

Step 5. (Proof of (A.118)) First, we derive the expression of R_n. Using Lemma A1 and the definition of $\{\mathbb{F}_n^h : 0 \le h \le n\}$ given by (A.3) therein, we write

$$\Omega_n(\sigma^2, \gamma) = \sigma^2 \Delta_n \mathbb{I}_n + \sum_{h=0}^{q_n} \gamma_h (2\mathbb{F}_n^h - \mathbb{F}_n^{h+1} - \mathbb{F}_n^{h-1}). \tag{A.119}$$

Here $\mathbb{F}_n^h = 0_{n\times n}$ for $h = -1$ by convention. On the other hand, we rewrite Σ_n defined by (3.7) as

$$\Sigma_n(\sigma^2, \gamma) = \sigma^2 \Delta_n \mathbb{I}_n + \sum_{h=0}^{q_n} \gamma_h (2\mathbb{G}_n^h - \mathbb{G}_n^{h+1} - \mathbb{G}_n^{h-1}). \tag{A.120}$$

To write R_n in a more compact form, define $\mathbb{K}_n^h, \widetilde{\mathbb{K}}_n^h \in \mathcal{M}_{n_T}$ by $(\mathbb{K}_n^h)_{ij} = \mathbb{1}_{\{h=i+j\}} - \mathbb{1}_{\{h+1=i+j\}}$ and $(\widetilde{\mathbb{K}}_n^h)_{ij} = (\mathbb{K}_n^h)_{n+1-i,n+1-j}$. Obviously, $\mathbb{K}_n^h + \widetilde{\mathbb{K}}_n^h = \mathbb{G}_n^h - \mathbb{G}_n^{h+1} - \mathbb{F}_n^h + \mathbb{F}_n^{h+1}$; hence (A.119) and (A.120) lead to

$$R_n(\beta^{(n)}) = \Sigma_n(\beta^{(n)}) - \Omega_n(\beta^{(n)}) = \sum_{h=0}^{q_n-1} (\gamma_h^{(n)} - \gamma_{h+1}^{(n)})(\mathbb{K}_n^h + \widetilde{\mathbb{K}}_n^h). \tag{A.121}$$

Here $\gamma_{q_n+1}^{(n)} = 0$ by convention. Next, we apply (A.121) to (A.118). In view of the symmetry between $\mathbb{K}_n^h$ and $\widetilde{\mathbb{K}}_n^h$, we can replace $R_n(\beta^{(n)})$ in (A.118) with $\widetilde{R}_n(\beta^{(n)}) \coloneqq \sum_{h=0}^{q_n-1}(\gamma_h^{(n)} - \gamma_{h+1}^{(n)})\mathbb{K}_n^h$. Then (A.118) becomes

$$\mathbb{E}\Big|\mathbb{1}_{\Omega_n'} \sum_{h=0}^{q_n-1}\sum_{l=0}^{q_n-1} (\gamma_h^{(n)} - \gamma_{h+1}^{(n)})(\gamma_l^{(n)} - \gamma_{l+1}^{(n)}) Y_n^\intercal \Omega_n^{-1}\mathbb{K}_n^h\Omega_n^{-1}\mathbb{K}_n^l\Omega_n^{-1}Y_n\Big| = o(n^{1/4}). \tag{A.122}$$

From the definition of $\gamma^{(n)}$ and applying Hölder's inequality, we only need prove for all $n^{1/3}h_n \le K$ and $n^{1/3}l_n \le K$ that

$$\mathbb{E}\big|\mathbb{1}_{\Omega_n'} Y_n^\intercal \Omega_n^{-1}\mathbb{K}_n^{h_n}\Omega_n^{-1}\mathbb{K}_n^{l_n}\Omega_n^{-1}Y_n\big| \le K n^{1/4} h_n l_n. \tag{A.123}$$

In view of the notation introduced in Step 3, plus the definition of $\mathbb{K}_n^h$, we have that

$$\begin{aligned} Y_n^\intercal \Omega_n^{-1}\mathbb{K}_n^h\Omega_n^{-1}\mathbb{K}_n^l\Omega_n^{-1}Y_n \quad &= \quad \sum_{i=1}^{h-1}\sum_{j=1}^{l-1} A(1)_i A(1)_j A(3)_{h-i,h-j} - A(1)_h \sum_{j=1}^{l-1} A(1)_j A(2)_{1,l-j} \\ &\quad - A(1)_l \sum_{i=1}^{h-1} A(1)_i A(2)_{1,h-j} + A(1)_h A(1)_l (\Omega_n^{-1})_{11}. \end{aligned} \tag{A.124}$$

We hence deduce (A.123) by applying the Cauchy-Schwarz inequality and the four estimates provided in Step 3 to (A.124). ∎

Lemma A8. *Suppose Assumptions 1 - 4 hold and $q_n \le K n^{1/3}$. It holds that under either $n^{1/2}\iota^{(n)} \to \infty$ or $n^{1/2}\iota^{(n)} \le K$,*

$$\bar{\eta}^\intercal \Xi_{A,n}(\beta^{(n)}(q_n)) - \bar{\eta}^\intercal \Xi_{D,n}(\beta^{(n)}(q_n)) = o_{\mathrm{P}}\left(n^{-1/2}\sqrt{q_n+1} + n^{-1/4}\sqrt{\iota^{(n)}}\right). \tag{A.125}$$

Proof. All of the contents of the proof of Lemma A7 until (A.111) remain valid if we replace Σ_n with Ω_n, Ω_n with $\Omega_{D,n}$, and (A.105) with (A.125). Unlike the situation there, we do know the analytical expressions of both Ω_n^{-1} and $\Omega_{D,n}^{-1}$, as given by Lemma A2. Note that $\Omega_{D,n}$ is a block-diagonal matrix and we apply Lemma A2 to each block. Instead of (A.112), we use

$$\Omega_n^{-1} = \Omega_{D,n}^{-1} - \Omega_n^{-1} R_n \Omega_{D,n}^{-1} \quad \text{and} \quad R_n \coloneqq \Omega_n - \Omega_{D,n}.$$

The justification for doing so is the same as the one stated at the end of the proof of Lemma A7. Indeed, by definition, R_n here has only nonzero entries near the top-left or right-bottom corners of the blocks $\Omega_{D,n}$ consists of. According to Lemma A2, $(\Omega_{D,n}^{-1})_{i,j}$ shrinks when either i or j approaches the borders of those blocks. Moreover, locally R_n also maintains a structure similar to (A.121); See the comment at the end of the proof of Lemma A7. Hence, we obtain (A.125) following the same reasoning as in the proof of Lemma A7. Note that we will skip Steps 2 and 4 there, since we know both $\Omega_{D,n}^{-1}$ and Ω_n^{-1}. ∎

Lemma A9. *Suppose Assumptions 1 - 4 hold. It holds that for all sequences $q_n \le K n^{1/3}$ and under either $n^{1/2}\iota^{(n)} \to \infty$ or $n^{1/2}\iota^{(n)} \le K$,*

$$\sup_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n)} \left| \frac{L_n(\sigma^2,\gamma) - \bar{L}_n^\star(\sigma^2,\gamma)}{\bar{L}_n^\star(C_T,\gamma^{(n)}) - \bar{L}_n^\star(\sigma^2,\gamma) + n} \right| = o_{\mathrm{P}}(1). \tag{A.126}$$

Proof. Step 1. (Technical results) We start by defining a family of subsets of $\Pi_n^{(\sigma^2,\gamma)}(q)$ indexed by α_1 and α_2:

$$\Pi_n^{(\sigma^2,\gamma)}(q,\alpha_1,\alpha_2) = \{(\sigma^2,\gamma) \in \Pi_n^{(\sigma^2,\gamma)}(q) : \alpha_1 \le \chi^2(\sigma^2,\gamma,\Delta_n) \le \alpha_2\}.$$

Obviously, $\Pi_n^{(\sigma^2,\gamma)}(q) = \Pi_n^{(\sigma^2,\gamma)}(q,0,\alpha) \cup \Pi_n^{(\sigma^2,\gamma)}(q,\alpha,\infty)$ for all α. In this step we aim to prove that for all $\alpha_n \to \infty$ and all K fixed,

$$\sup_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,\alpha_n\Delta_n,\infty)} \left| \frac{L_{A,n}(\sigma^2,\gamma) - \bar{L}_n^\star(\sigma^2,\gamma)}{\bar{L}_n^\star(C_T,\gamma^{(n)}) - \bar{L}_n^\star(\sigma^2,\gamma) + n} \right| = o_{\mathrm{P}}(1), \tag{A.127}$$

$$\text{and} \qquad \sup_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,0,K\Delta_n)} \left| \frac{L_{A,n}(\sigma^2,\gamma) - \bar{L}_n^\star(\sigma^2,\gamma)}{\bar{L}_n^\star(C_T,\gamma^{(n)}) - \bar{L}_n^\star(\sigma^2,\gamma) + n} \right| = o_{\mathrm{P}}(1). \tag{A.128}$$

We consider the case in which $n^{1/2}\iota^{(n)} \le K$. The case in which $n^{1/2}\iota^{(n)} \to \infty$ follows from the same reasoning. Straightforwardly, it holds that in restriction to Ω_n',

$$\sup_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,\alpha_n\Delta_n,\infty)} \left| \bar{L}_n^\star(C_T,\gamma^{(n)}) - \bar{L}_n^\star(\sigma^2,\gamma) - \frac{n_T}{2}\log\frac{\chi^2(\sigma^2,\gamma,\Delta_n)}{\chi^2(C_T,\gamma^{(n)},\Delta_n)} \right| \le Kn.$$

Because $\chi^2(\sigma^2,\gamma,\Delta_n) \ge \alpha_n\Delta_n$ and $\chi^2(C_T,\gamma^{(n)},\Delta_n) \in (K^{-1}\Delta_n, K\Delta_n)$, it holds that in restriction to Ω_n',

$$\inf_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,\alpha_n\Delta_n,\infty)} \frac{1}{n_T}|\bar{L}_n^\star(C_T,\gamma^{(n)}) - \bar{L}_n^\star(\sigma^2,\gamma)| \to \infty.$$

Hence, in view of the triangle inequality and the definitions of $L_{A,n}$ and $\bar{L}_n^\star$, plus using (A.107), to obtain (A.127) it is enough to show that

$$\sup_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,\alpha_n\Delta_n,\infty)} Y_n^\intercal \Omega_n(\sigma^2,\gamma)^{-1} Y_n = O_{\mathrm{P}}(n) \tag{A.129}$$

$$\text{and} \qquad \sup_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,\alpha_n\Delta_n,\infty)} \frac{\mathbb{1}_{\Omega_n'}}{2\pi}\int_{-\pi}^{\pi}\frac{f(\lambda;C_T,\gamma^{(n)},\Delta_n)}{f(\lambda;\sigma^2,\gamma,\Delta_n)}d\lambda \le K.$$

The second bound is obviously true and we now show the first bound (A.129). Lemma A2 states that we can write

$$Y_n^\intercal \Omega_n(\sigma^2,\gamma)^{-1}Y_n = Y_n^\intercal \sum_{h=0}^{\infty}\rho_h(\sigma^2,\gamma,\Delta_n)\mathbb{F}_n^h Y_n, \tag{A.130}$$

where $\rho_h(\sigma^2,\gamma,\Delta_n)$ satisfies (A.4) uniformly over $(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,\alpha_n\Delta_n,\infty)$. On the other hand, with $n^{1/2}\iota^{(n)}\le K$ and Assumption 4, we conclude that

$$\mathbb{E}|Y_n^\intercal \mathbb{F}_n^h Y_n| \le Kh^{-2}+Kn^{-1/2}. \tag{A.131}$$

The combination of (A.130), (A.4), (A.131), and (A.107), plus Hölder's inequality, readily yields (A.129) and hence (A.127) is proved. We now prove (A.128). Since $\bar{L}_n^\star(C_T,\gamma^{(n)})-\bar{L}_n^\star(\sigma^2,\gamma)$ is always nonnegative, in view of the definitions of $L_{A,n}$ and $\bar{L}_n^\star$, (A.128) directly comes from

$$\sup_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,0,K\Delta_n)}\Big|Y_n^\intercal\Omega_n(\sigma^2,\gamma)^{-1}Y_n-\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f(\lambda;C_T,\gamma^{(n)},\Delta_n)}{f(\lambda;\sigma^2,\gamma,\Delta_n)}d\lambda\Big| = o_{\mathrm{P}}(n). \tag{A.132}$$

The uniform convergence (A.132) comes from establishing pointwise convergence and stochastic equicontinuity, following the same reasoning as for Theorem 2.1 and Corollary 2.2 in ?. Applying steps 1 and 7 of the proof of Lemma A2, we have, for all deterministic $\{(\sigma_n^2,\gamma_n)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,0,K\Delta_n): n\ge 1\}$ and under $n^{1/2}\iota^{(n)}\le K$,

$$Y_n^\intercal\Omega_n(\sigma_n^2,\gamma_n)^{-1}Y_n-\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f(\lambda;C_T,\gamma^{(n)},\Delta_n)}{f(\lambda;\sigma_n^2,\gamma_n,\Delta_n)}d\lambda = o_{\mathrm{P}}(n).$$

On the other hand, using (A.131) and (A.7), plus Assumption 2, we can repeat the reasoning for deriving (A.129) to conclude that for all $0\le j\le q_n$ and with Ω_n' introduced above (A.107),

$$\mathbb{E}\Big(\mathbb{1}_{\Omega_n'}\sup_{(\sigma^2,\gamma)}\Big|\frac{\partial}{\partial\sigma^2}Y_n^\intercal\Omega_n(\sigma^2,\gamma)^{-1}Y_n\Big|\Big)\le Kn, \quad \mathbb{E}\Big(\mathbb{1}_{\Omega_n'}\sup_{(\sigma^2,\gamma)}\Big|\Delta_n\frac{\partial}{\partial\gamma_j}Y_n^\intercal\Omega_n(\sigma^2,\gamma)^{-1}Y_n\Big|\Big)\le Kn,$$

$$\sup_{(\sigma^2,\gamma)}\Big|\mathbb{1}_{\Omega_n'}\frac{\partial}{\partial\sigma^2}\int_{-\pi}^{\pi}\frac{f(\lambda;C_T,\gamma^{(n)},\Delta_n)}{f(\lambda;\sigma_n^2,\gamma_n,\Delta_n)}d\lambda\Big|\le K, \quad \sup_{(\sigma^2,\gamma)}\Big|\mathbb{1}_{\Omega_n'}\Delta_n\frac{\partial}{\partial\gamma_j}\int_{-\pi}^{\pi}\frac{f(\lambda;C_T,\gamma^{(n)},\Delta_n)}{f(\lambda;\sigma_n^2,\gamma_n,\Delta_n)}d\lambda\Big|\le K,$$

from which the stochastic equicontinuity follows. Here the range over which the supremums are taken is $(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n,0,K\Delta_n)$ and the additional factor Δ_n compared with Assumption 3A in ? arises because of Assumption 4 and the definition of $\Pi_n^{(\sigma^2,\gamma)}(q_n,0,K\Delta_n)$.

Step 2. (Conclusion) In view of (A.127) and (A.128), plus using the subsequence argument, we obtain

$$\sup_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n)}\left|\frac{L_{A,n}(\sigma^2,\gamma)-\bar{L}_n^{\star}(\sigma^2,\gamma)}{\bar{L}_n^{\star}(C_T,\gamma^{(n)})-\bar{L}_n^{\star}(\sigma^2,\gamma)+n}\right| = o_{\mathrm{P}}(n). \tag{A.133}$$

Further, we have, following the reasoning in the proof of Lemma A7 and using Lemma A2, that

$$\sup_{(\sigma^2,\gamma)\in\Pi_n^{(\sigma^2,\gamma)}(q_n)}\left|\frac{L_n(\sigma^2,\gamma)-L_{A,n}(\sigma^2,\gamma)}{\bar{L}_n^{\star}(C_T,\gamma^{(n)})-\bar{L}_n^{\star}(\sigma^2,\gamma)+n}\right| = o_{\mathrm{P}}(1). \tag{A.134}$$

Note that the bound we require here is less sharp and that only Σ_n^{-1} and Ω_n^{-1} themselves are involved, as we do not take derivatives here. The lemma is a direct result of (A.133) and (A.134). ∎

Lemma A10. *Suppose Assumptions 1 - 4 hold. For all sequences $\{q_n\}$ and $\{q_n'\}$, it holds that with probability approaching one,*

$$\frac{1}{n_T}\bar{L}_n^{\star}(\sigma^{(n)}(q_n)^2,\gamma^{(n)}(q_n)) - \frac{1}{n_T}\bar{L}_n^{\star}(\sigma^{(n)}(q_n')^2,\gamma^{(n)}(q_n')) \sim \psi_n^4(\|\widetilde{\kappa}^{(n)}\|^2_{(q_n')} - \|\widetilde{\kappa}^{(n)}\|^2_{(q_n)}).$$

Proof. We define $C_{i,j} = \pi^{-1}\int_{-\pi}^{\pi}|1-\psi_n e^{\mathrm{i}\lambda}|^{-4}\cos i\lambda\cos j\lambda d\lambda$. It holds that

$$(1-\psi_n^2)^2 C_{i,j} = \psi_n^{|i-j|}|i-j| + \psi_n^{|i+j|}|i+j| + \frac{1+\psi_n^2}{1-\psi_n^2}(\psi_n^{|i-j|}+\psi_n^{|i+j|}).$$

We introduce m-dimensional matrices C_m, $\mathbb{J}_m^h$, and $\mathbb{K}_m^h$ with entries given by $C_{i,j}$, $(\mathbb{J}_m^h)_{i,j} = \mathbb{1}_{\{|i-j|=h\}}$, and $(\mathbb{K}_m^h)_{i,j} = \mathbb{1}_{\{i+j=h\}}+\mathbb{1}_{\{2m+2-i-j=h\}}$. We further define $\bar{C}_m = (1-\psi_n^2)^2 C_m$, $\widetilde{C}_m = \sum_{h=0}^{\infty}\mathbb{J}_m^h\psi_n^h(h+(1+\psi_n^2)(1-\psi_n^2)^{-1})$, and $\check{C}_m = \sum_{h=0}^{\infty}\mathbb{K}_m^h\psi_n^h(h+(1+\psi_n^2)(1-\psi_n^2)^{-1})$. From Lemma A1, it follows that

$$(O_m(\widetilde{C}_m-\check{C}_m)O)_{i,j} = \delta_{i,j}\sum_{h=-\infty}^{\infty}\psi_n^{|h|}(|h|+(1+\psi_n^2)(1-\psi_n^2)^{-1})\cos\frac{hj\pi}{m+1}.$$

By direct calculations, we obtain $\widetilde{C}_m \geq \check{C}_m$ for m sufficiently large. On the other hand, it holds that for all m-dimensional vector β,

$$\beta^{\intercal}\widetilde{C}_m\beta = (1-\psi_n^2)^2\sum_{j=-\infty}^{\infty}\left|\sum_{k=0}^{\infty}(k+1)\psi_n^k\beta_{j+k}\right|^2, \tag{A.135}$$

$$\beta^{\intercal}\bar{C}_m\beta = (1-\psi_n^2)^2\sum_{j=-\infty}^{\infty}\left|\sum_{k=0}^{\infty}(k+1)\psi_n^k\beta_{|j+k|}(1+\delta_{j+k,0})\right|^2. \tag{A.136}$$

Here we take $\beta_j = 0$ for all $j<0$. Because (A.135) and (A.136) hold for arbitrary m, and comparing $\widetilde{C}_m-\check{C}_m$ with $\bar{C}_m$, we prove the lemma by using the fact that $f(\lambda;\sigma^{(n)}(q_n)^2,\gamma^{(n)}(q_n),\Delta_n) \sim ((\iota^{(n)})^2+n^{-1})|1-\psi_n e^{\mathrm{i}\lambda}|^2$ with probability approaching one. ∎

Appendix B Proof of Main Results

B.1 Proof of Theorem 1

Proof. The assumptions of Theorem 1 lead to the fact that $\bar{L}_n^\star(C_T, \gamma^{(n)}) - \bar{L}_n^\star(\sigma^{(n)}(q)^2, \gamma^{(n)}(q)) = 0$ for all $q \geq q^\star$. Then, in view of the proof of Lemma B4 in the online appendix of Da and Xiu (2021), we have

$$\bar{L}_n^\star(C_T, \gamma^{(n)}) - \bar{L}_n^\star(\sigma^{(n)}(\widehat{q}_n)^2, \gamma^{(n)}(\widehat{q}_n)) \leq \frac{\log n}{2}(q^\star - \widehat{q}_n) + O(\widehat{q}_n + \log n). \tag{B.1}$$

As an immediate result, we obtain that for n sufficiently large, $\widehat{q}_n \leq q^\star$. On the other hand, we have $(\log n)^{-1}\big(\bar{L}_n^\star(C_T, \gamma^{(n)}) - \bar{L}_n^\star(\sigma^{(n)}(q)^2, \gamma^{(n)}(q))\big) \to \infty$ for all $q \leq q^\star - 1$, according to the assumption that $\sqrt{n}(\log n)^{-1}|\theta_{q^\star}^{(n)}| \to \infty$. Then (B.1) indicates that for n sufficiently large, $\widehat{q}_n \geq q^\star$ and we conclude the proof. ∎

B.2 Proofs of Theorem 2, Corollary 1, and Proposition 1

Lemma B1. *Suppose the same assumptions as those in Theorem 2 hold. Then it holds that*

$$\widehat{\mathcal{R}}_n(q^\star) = o_{\mathrm{P}}(1) \quad \textit{and} \quad \mathcal{R}^{(n)}(q^\star) = o_{\mathrm{P}}(1).$$

Proof. We start by proving the convergence of $\widehat{\mathcal{R}}_n(q_n^\star)$. From Lemma A9 it directly follows that

$$\bar{L}_n^\star(C_T, \gamma^{(n)}) - \bar{L}_n^\star(\widehat{\sigma}_n^2(\widehat{q}_n), \widehat{\gamma}_n(\widehat{q}_n)) = o_{\mathrm{P}}(n). \tag{B.2}$$

Since both (σ_n^2, γ_n) and $(C_T, \gamma^{(n)})$ belong to $\Pi_n^{(\sigma^2,\gamma)}$, according to Theorem 4.1.1, Proposition 4.5.3, Proposition 3.2.1, and Theorem 3.1.2 in Brockwell and Davis (1991), there exist unique (χ_n^2, ϕ_n) and $((\chi^{(n)})^2, \phi^{(n)})$ such that for all $-\pi \leq \lambda \leq \pi$,

$$f(\lambda; \widehat{\sigma}_n^2(\widehat{q}_n), \widehat{\gamma}_n(\widehat{q}_n), \Delta_n) = \chi_n^2 g(\lambda; \phi_n) \quad \text{and} \quad f(\lambda; C_T, \gamma^{(n)}, \Delta_n) = (\chi^{(n)})^2 g(\lambda; \phi^{(n)}), \tag{B.3}$$

$$1 + \inf_{z\in\mathbb{C}, |z|\leq 1} \sum_{j=1}^{\infty} \phi_{n,j} z^j > 0 \quad \text{and} \quad 1 + \inf_{z\in\mathbb{C}, |z|\leq 1} \sum_{j=1}^{\infty} \phi_j^{(n)} z^j > 0. \tag{B.4}$$

In view of (B.3) and the definition of $\bar{L}_n^\star$, the bound (B.2) can be rewritten in terms of (χ_n^2, ϕ_n) and $((\chi^{(n)})^2, \phi^{(n)})$, which leads to

$$\log \frac{\chi_n^2}{(\chi^{(n)})^2} = o_{\mathrm{P}}(1) \quad \text{and} \quad \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f(\lambda; C_T, \gamma^{(n)}, \Delta_n)}{f(\lambda; \sigma_n^2, \gamma_n, \Delta_n)} d\lambda - 1 = o_{\mathrm{P}}(1). \tag{B.5}$$

Here we use (A.107) and the fact that $(2\pi)^{-1}\int_{-\pi}^{\pi} g(\lambda; \phi^{(n)})/g(\lambda; \phi_n) d\lambda \geq 1$, indicated by (B.4). With $\chi^{(n)}$ calculated using Assumption 4, the first part of (B.5) indicates that $\log \chi_n^2 = \log(\iota^{(n)})^2 + o_{\mathrm{P}}(1)$. Substituting the estimate of χ_n^2 back into (A.22), plus using the second part of (B.5), plus (A.107), immediately allows us to prove the convergence of $\widehat{\mathcal{R}}_n(q_n^\star)$. On the other hand, the convergence of

$\mathcal{R}^{(n)}(q^\star) = o_{\mathrm{P}}(1)$ follows directly from Lemma A5. We conclude the proof. ∎

Proofs of Theorem 2 and Corollary 1. Step 1. (Technical preparation) In this proof the dependence of $\beta^{(n)}$ on q is suppressed and here we take $q = q^\star$. We set $\beta_n(\sigma^2, \gamma) = (\sigma^2, \gamma)$. We start by introducing $(q^\star + 2) \times (q^\star + 2)$ matrices $\partial \Xi_n(\beta_n, \beta_n', k)$ with $k \in \{1, 2\}$ and $\beta_n, \beta_n' \in \Pi_n^\beta(q^\star)$, defined, for $0 \le i, j \le q^\star + 1$, by

$$\partial \Xi_n(\beta_n, \beta_n'; 1)_{i,j} = \frac{1}{2n} \mathrm{tr}\left(\frac{\partial \log \Omega_n(\beta_n)}{\partial \beta_i} \frac{\partial \log \Omega_n(\beta_n')}{\partial \beta_j}\right), \tag{B.6}$$

$$\partial \Xi_n(\beta_n, \beta_n'; 2)_{i,j} = \frac{1}{4n} \mathrm{tr}\left(\frac{\partial \log \Omega_n(\beta_n)}{\partial \beta_i} \frac{\partial \log \Omega_n(\beta_n')}{\partial \beta_j} (\Omega_n(\beta_n)^{-1} + \Omega_n(\beta_n')^{-1}) Y_n Y_n^\intercal\right). \tag{B.7}$$

We further denote $\partial \Xi_n(\beta_n; j) := \partial \Xi_n(\beta_n, \beta_n; j)$. In addition, we use $\partial \Xi_n(\bar{\beta}^{(n)}, q^\star; j)$ and $\partial \bar{\Xi}_n^\star(\bar{\beta}^{(n)}, q^\star)$, respectively, to denote the $(q^\star + 2) \times (q^\star + 2)$ matrices with entries defined by (B.6) and (B.7) and with entries defined by (A.2). On the other hand, we let $\{\check{\beta}_n \in \Pi_n^\beta(q^\star) : n \ge 1\}$ be a sequence of $(q^\star + 2)$-dimensional random vectors that satisfies the equation $\Xi_n(\check{\beta}_n) = 0_{q^\star+2}$, and the condition whereby $\sup_\lambda |f(\lambda; \check{\beta}_n, \Delta_n) f(\lambda; \bar{\beta}^{(n)}, \Delta_n)^{-1} - 1| = o_{\mathrm{P}}(1)$ holds. In view of the definition of $\partial \Xi_n(\beta_n, \beta_n'; j)$ introduced in (B.6) and (B.7), plus applying the rules of matrix differentiation, in particular that $\Omega_n(\beta)$ and $\Omega_n(\beta')$ commute for all (β, β'), we observe that

$$\check{\beta}_n - \beta^{(n)} = (2\partial \Xi_n(\check{\beta}_n, \beta^{(n)}; 2) - \partial \Xi_n(\check{\beta}_n, \beta^{(n)}; 1))^{-1} (\Xi_{A,n}(\check{\beta}_n) - \Xi_{A,n}(\beta^{(n)})). \tag{B.8}$$

On the other hand, using $\mathbb{D}_m^j = O_m \mathbb{F}_m^j O_m$ and the connection between matrix V_m and spectral density $f(\lambda; \beta, \Delta_n)$ and the positivity of both following the reasoning of step 1 of the proof of Lemma A2, plus the fact that $\partial f(\lambda; \beta, \Delta_n)/\partial \beta$ does not depend on β, we have, for all $\alpha_n \to 0$ and $j \in \{1, 2\}$, and under that $\sup_\lambda |f(\lambda; b_n, \Delta_n) f(\lambda; \bar{\beta}^{(n)}, \Delta_n)^{-1} - 1| \to 0$ for $b_n \in \{\check{\beta}_n, \beta^{(n)}\}$,

$$\begin{cases} (1 - \alpha_n) \partial \Xi_n(\bar{\beta}^{(n)}, q^\star; j) \le \partial \Xi_n(\check{\beta}_n, \beta^{(n)}; j) \le (1 + \alpha_n) \partial \Xi_n(\bar{\beta}^{(n)}, q^\star; j) \\ (1 - \alpha_n) \partial \bar{\Xi}_n^\star(\bar{\beta}^{(n)}, q^\star) \le \partial \Xi_n(\bar{\beta}^{(n)}, q^\star; 1) \le (1 + \alpha_n) \partial \bar{\Xi}_n^\star(\bar{\beta}^{(n)}, q^\star) \end{cases}. \tag{B.9}$$

Furthermore, using Lemma A2, we can derive $\mathbb{E}\big|\mathbb{1}_{\Omega_n'}\big(\mathrm{tr}(\Omega_n(\bar{\beta}^{(n)})^{-1} Y_n Y_n^\intercal - \mathbb{I}_n)\big)^2\big| \le Kn$, which, combined with (A.22) and (A.107), leads to the fact that for some $\alpha_n \to 0$,

$$\lim_{n\to\infty} \mathbb{P}((1 - \alpha_n) \partial \Xi_n(\bar{\beta}^{(n)}, q^\star; 1) \le \partial \Xi_n(\bar{\beta}^{(n)}, q^\star; 2) \le (1 + \alpha_n) \partial \Xi_n(\bar{\beta}^{(n)}, q^\star; 1)) = 1. \tag{B.10}$$

Step 2. (Main proof) We define $Y_n^C(j) = (Y_n^C(j)_1, \dots, Y_n^C(j)_{n_d})^\intercal$ and introduce, for all $1 \le i \le q^\star + 2$ and $j \ge 1$,

$$\mathcal{V}_n(j)_i = -\frac{1}{2n} \frac{\partial}{\partial \beta_i} \mathrm{tr}(\Omega_{n_d}(\beta^{(n)})^{-1} U^C(j) U^C(j)^\intercal), \quad \bar{\mathcal{V}}_n(j)_i = -\frac{1}{2n} \frac{\partial}{\partial \beta_i} \mathrm{tr}(\Omega_{n_d}(\beta^{(n)})^{-1} \Omega_n^{U,C}(j)). \tag{B.11}$$

Using Lemmas A3 and A4, we obtain that $\Xi_{D,n}(\beta^{(n)})_1 = o_{\mathrm{P}}(n^{-1/2})$, and that for all $2 \le i \le q^\star + 2$,

$$\Xi_{D,n}(\beta^{(n)})_i - \bar{\Xi}_n(\beta^{(n)})_i - \sum_{j=1}^{J_d} (\mathcal{V}_n(j)_i - \bar{\mathcal{V}}_n(j)_i) = o_{\mathrm{P}}(n^{-1/2}) \quad \text{and} \quad \bar{\Xi}_n(\beta^{(n)})_i = o_{\mathrm{P}}(n^{-1/2}).$$

Here we use the well-known result (see Section 2.1.5 of ?) that under Assumption A1 and for two finite stopping times $S \le S'$ and some $p \ge 0$, and for a process A that is one of μ, σ, ξ, ξ^{-1}, and η,

$$\mathbb{E}\big(\sup_{S \le s \le S'} (\|A_s - A_S\|^p) | \mathcal{F}_S\big) \le \mathbb{E}((S' - S)^{1 \wedge (p/2)} | \mathcal{F}_S). \tag{B.12}$$

We let $\mathcal{F}^\varepsilon(j) = \sigma(\varepsilon_C(k)_i : i \le n_d, k \le j-1)$ be the σ-field generated by the sequence of all $\varepsilon_C(k)$ with $k \le j-1$, and $\mathcal{F}^\chi(j) = \sigma(\chi_i : i \le (j-1)n_d)$ be the σ-field generated by the sequence of all χ_i with $i \le (j-1)n_d$, and denote $\mathcal{F}(j) = \mathcal{F}_\infty \otimes \mathcal{F}^\varepsilon(j) \otimes \bigvee_{k \ge 0} \mathcal{F}^\chi(k)$. From direct calculations we have that for all $2 \le i \le q^\star + 2$,

$$n^{1/2} \sum_{j=1}^{J_d} \mathbb{E}(\mathcal{V}_n(j)_i - \bar{\mathcal{V}}_n(j)_i | \mathcal{F}(j)) = o_{\mathrm{P}}(1), \quad n^2 \sum_{j=1}^{J_d} \mathbb{E}((\mathcal{V}_n(j)_i - \bar{\mathcal{V}}_n(j)_i)^4 | \mathcal{F}(j)) = o_{\mathrm{P}}(1). \tag{B.13}$$

And for all $2 \le i, i' \le q^\star + 2$, it holds that

$$\begin{aligned} &n^{3/2} n_T^{-1} \sum_{j=1}^{J_d} \mathbb{E}((\mathcal{V}_n(j)_i - \bar{\mathcal{V}}_n(j)_i)(\mathcal{V}_n(j)_{i'} - \bar{\mathcal{V}}_n(j)_{i'}) | \mathcal{F}(j)) \\ &\qquad = \Big(\frac{1}{2} W(\gamma^\star) + \frac{\mathrm{cum}_4(\varepsilon)}{4} W(\gamma^\star) \gamma^\star \gamma^{\star\intercal} W(\gamma^\star)\Big) \frac{T \int_0^T \eta_s^4 \xi_s^{-1} ds}{\big(\int_0^T \eta_s^2 \xi_s^{-1} ds\big)^2} + o_{\mathrm{P}}(1). \end{aligned} \tag{B.14}$$

Here we use Lemmas A3 and A4, Assumption A1, and the fact that $\varepsilon_C(j)$ is independent of $\mathcal{F}(j')$ for all $j' \le j$. Because of the definition of stable convergence and the fact that $\mathcal{F}_\infty \subset \mathcal{F}(j)$, in view of (B.13) and (B.14), we readily obtain

$$n^{3/2} n_T^{-1} (0_{q^\star+1} : \mathbb{I}_{q^\star+1}) \Xi_{D,n}(\beta^{(n)}) \xrightarrow{\mathcal{L}-\mathcal{F}_\infty} \mathcal{U}, \tag{B.15}$$

where the stable convergence in law is with respect to $\mathcal{F}_\infty$. Here $\mathcal{U}$ is a $(q^\star+1)$-dimensional random vector defined on an extension $(\bar{\Omega}, \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t \ge 0}, \bar{\mathbb{P}})$ of $(\Omega_{(0)}, \mathcal{F}_\infty, (\mathcal{F}_t)_{t \ge 0}, \mathbb{P}_{(0)})$ and which, conditionally on $\mathcal{F}_\infty$, is centered Gaussian and satisfies

$$\bar{\mathbb{E}}(\mathcal{U}\mathcal{U}^\intercal | \mathcal{F}_\infty) = \Big(\frac{1}{2} W(\gamma^\star) + \frac{\mathrm{cum}_4(\varepsilon)}{4} W(\gamma^\star) \gamma^\star \gamma^{\star\intercal} W(\gamma^\star)\Big) \frac{T \int_0^T \eta_s^4 \xi_s^{-1} ds}{\big(\int_0^T \eta_s^2 \xi_s^{-1} ds\big)^2}.$$

On the other hand, straightforward algebra leads to

$$\begin{aligned}
n^{-1}n_T\partial\Xi_n(\bar\beta^{(n)},q^\star)^{-1}_{1,1} &= 2\sigma^3\zeta\Delta_n^{-1/2}+O(1),\\
n^{-1}n_T\partial\Xi_n(\bar\beta^{(n)},q^\star)^{-1}_{1,k+1} &= -\frac{2\sigma^2}{\zeta^2}\sum_{r=1}^{q+1}(2-\delta_{r,1})W(\gamma)^{-1}_{r,k}+O(\Delta_n^{1/2}),\\
n^{-1}n_T\partial\Xi_n(\bar\beta^{(n)},q^\star)^{-1}_{j+1,k+1} &= 2W(\gamma)^{-1}_{j,k}+O(\Delta_n^{1/2}).
\end{aligned}$$

Hence, (B.9) and (B.10) jointly indicate that for all $2\le j\le q^\star+2$,

$$\check\beta_{n,j}-\beta_j^{(n)}=-(0_{j-1},1,0_{q^\star+2-j})^\intercal\partial\Xi_n(\bar\beta^{(n)},q^\star)^{-1}\Xi_{D,n}(\beta^{(n)})+o_{\mathrm{P}}(n^{-1/2}). \tag{B.16}$$

Here we also use Lemmas A7 and A8, (A.109), and the relation (B.8). At this stage, in view of the fact that by definition $(\widehat\sigma_n^2(q_n),\widehat\gamma_n(q_n))$ maximizes $L_n(\sigma^2,\gamma)$ over $\Pi_n^{(\sigma,\gamma^2)}(q_n)$ and the definition of $\check\beta_n$, plus Lemma B1, the combination of (B.15) and (B.16) proves the theorem. Applying continuous mapping theorem, we obtain the corollary. ∎

Proof of Proposition 1. In view of Assumption A1 and (B.12), a Riemann sum argument leads to

$$\begin{aligned}
\frac{1}{4n_T}\sum_{i=1+k_n}^{n_T-k_n}(\Delta_i^nU)^2\sum_{j=-k_n}^{k_n}(\Delta_{i+j}^nU)^2 &= \frac{\big(\int_0^T\eta_s^4\xi_s^{-1}ds\big)\big(\int_0^T\xi_s^{-1}ds\big)}{\big(\int_0^T\eta_s^2\xi_s^{-1}ds\big)^2}\Bigg(\big(2k_n+\mathrm{cum}_4(\varepsilon)\big)(\gamma_0^\star-\gamma_1^\star)^2\\
&\quad+\frac{1}{\pi}\int_{-\pi}^{\pi}f(\lambda;\gamma^\star)^2(1-\cos\lambda)^2d\lambda\Bigg)+o_{\mathrm{P}}(1),\\
\frac{1}{4n_T}\sum_{i=1}^{n_T-2k_n}(\Delta_i^nU)^2\sum_{j=k_n+1}^{2k_n}(\Delta_{i+j}^nU)^2 &= \frac{\big(\int_0^T\eta_s^4\xi_s^{-1}ds\big)\big(\int_0^T\xi_s^{-1}ds\big)}{\big(\int_0^T\eta_s^2\xi_s^{-1}ds\big)^2}k_n(\gamma_0^\star-\gamma_1^\star)^2+o_{\mathrm{P}}(1).
\end{aligned}$$

On the other hand, from Theorem 3.3.1 of **?** it follows that $\sum_{i=1}^{n_T}(\Delta_i^nX)^4=O_{\mathrm{P}}(1)$. Given the consistency of $\widehat\gamma_n$ and Assumption 4, applying Cauchy-Schwarz and Jensen's inequalities proves the proposition. ∎

B.3 Proof of Theorem 3

Proof. In view of the proof of Lemma B4 in the online supplemental appendix of Da and Xiu (2021), we have that under either $n^{1/2}\iota^{(n)}\to\infty$ or $n^{1/2}\iota^{(n)}\le K$, and for all $a_n\to\infty$ and all fixed $0<k<K$,

$$\begin{aligned}
&\bar L_n^\star\big(\sigma^{(n)}(q_n^\star(k))^2,\gamma^{(n)}(q_n^\star(k))\big)-\bar L_n^\star\big(\sigma^{(n)}(\widehat q_n)^2,\gamma^{(n)}(\widehat q_n)^2\big)\\
&\qquad=\frac{\log n}{2}\big(q_n^\star(k)-\widehat q_n\big)+o_{\mathrm{P}}\big(q_n^\star(k)+a_n\big)+O_{\mathrm{P}}\big(|q_n^\star(k)-\widehat q_n|\big).
\end{aligned}\tag{B.17}$$

The definition of $q_n^\star(k)$, combined with Lemma A10 and (B.17), indicates that there exists a fixed k such that $q_n^\star(k)-\widehat q_n\le 1$ with probability approaching one. Further, in view of Lemma A10 and using the bound on $n\psi_n^4 \sum_{j=q_n^\star(k)}^\infty |\kappa_j^n|^2$, we obtain that $n\psi_n^4 \sum_{j=\widehat{q}_n+1}^\infty |\kappa_j^n|^2 = O_{\mathrm{P}}\big((\widehat{q}_n+1)\log n\big)$ and that $\widehat{q}_n \le O_{\mathrm{P}}(q_n^\star(k)+1)$. Hence, it follows from Lemmas A5 and A6 that for all $0 \le j \le \widehat{q}_n$,

$$\big|\widehat{\gamma}^{(n)}(\widehat{q}_n)_j - \gamma_j^{(n)}\big|^2 = O_{\mathrm{P}}\Big(n^{-1}(\iota^{(n)})^4(\widehat{q}_n+1)\log n + n^{-3}(n^{1/2}\iota^{(n)}+1)(\widehat{q}_n+1)^3\log n\Big).$$

Here we also use the proof of Lemma B3 of Da and Xiu (2021). The bound on $\|\widehat{\gamma}^{(n)}(\widehat{q}_n) - \gamma^{(n)}\|$ directly follows. Continuous mapping theorem leads to the bound on $\|\widehat{\rho}^{(n)}(\widehat{q}_n) - \rho^{(n)}\|$. ∎

B.4 Proof of Theorem 4

Proof. We observe that for all $(\sigma^2,\gamma) \in \Pi_n^{(\sigma^2,\gamma)}(q)$,

$$\mathbb{I}_n = \frac{\partial \log \Sigma_n(\sigma^2,\gamma)}{\partial \sigma^2}\sigma^2 + \sum_{j=0}^q \frac{\partial \log \Sigma_n(\sigma^2,\gamma)}{\partial \gamma_j}\gamma_j.$$

Thus, the theorem directly follows from the fact that for all finte q,

$$\mathrm{tr}\left(\frac{\partial \log \Sigma_n(\widehat{\sigma}^2(q),\widehat{\gamma}(q))}{\partial(\sigma^2,\gamma)_j}\right) = -\mathrm{tr}\left(\frac{\partial \Sigma_n^{-1}(\widehat{\sigma}^2(q),\widehat{\gamma}(q))}{\partial(\sigma^2,\gamma)_j} Y_n Y_n^\intercal\right), \quad 1 \le j \le q+2,$$

which are the first-order conditions. ∎

B.5 Proof of Theorem 5

Proof. Step 1. (Technical preparation) We define $(q+2)\times(q+2)$ matrix $\bar{W}_n(\sigma^2,\gamma)$ and $n_T \times n_T$ matrices $\mathcal{R}(k)$, with $1 \le k \le q+2$, as

$$\begin{aligned}
\bar{W}_n(\sigma^2,\gamma) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial \log f(\lambda;\sigma^2,\gamma,\Delta_n)}{\partial(\sigma^2,\gamma)}\right)^\intercal \frac{\partial \log f(\lambda;\sigma^2,\gamma,\Delta_n)}{\partial(\sigma^2,\gamma)} d\lambda,\\
\mathcal{R}(k) &= \delta_{k,1}\frac{\partial \Sigma_n(\sigma^2,\gamma)^{-1}}{\partial \sigma^2} + \mathbb{1}_{\{k\ge 2\}}\frac{\partial \Sigma_n(\sigma^2,\gamma)^{-1}}{\partial \gamma_{k-2}}.
\end{aligned}$$

Throughout the proof we omit the argument of $\bar{W}_n(\sigma^2,\gamma)$ and $\widetilde{W}_n(\sigma^2,\gamma)$. We can calculate, using Lemmas A1, A3, and A4,

$$(\bar{W}_n^{-1})_{1,1} = 4\sigma^3\zeta\Delta_n^{-1/2} + O(1), \quad (\bar{W}_n^{-1})_{1,k+1} = -\frac{\sigma^2}{\zeta^2}\sum_{r=1}^{q+1}(2-\delta_{r,1})W(\gamma)_{r,k}^{-1} + O(\Delta_n^{1/2}), \tag{B.18}$$

$$(\bar{W}_n^{-1})_{j+1,k+1} = W(\gamma)_{j,k}^{-1} + O(\Delta_n^{1/2}), \quad (n_T\widetilde{W}_n^{-1})_{j,k} - (\bar{W}_n^{-1})_{j,k} = O(\Delta_n + \delta_{j,1}\delta_{k,1}). \tag{B.19}$$

On the other hand, also using $n^{1/2+\alpha} \le i \le n - n^{1/2+\alpha}$, we have, for $r \ge 0$,

$$\mathcal{R}(1)_{i,i} = -(4\zeta\sigma^3\Delta_n^{1/2})^{-1} + O(1), \quad \mathcal{R}(k)_{i,i} = -(2-\delta_{k,2})(4\sigma\zeta^3\Delta_n^{1/2})^{-1} + O(1), \tag{B.20}$$

$$\mathcal{R}(k)_{i,i+r} - \mathcal{R}(k)_{i,i} = \delta_{k,1}\frac{1}{4\zeta^2\sigma^2}\left(\frac{1-z_n^r}{1-z_n} - rz_n^r\right) + \mathbb{1}_{\{k\geq 2\}}\frac{2-\delta_{k,2}}{4\zeta^4}\left(\frac{1-z_n^r}{1-z_n} + rz_n^r\right) + O(1), \quad \text{(B.21)}$$

where z_n is defined in Lemma A3. If we further restrict $0 \leq r \leq K$, then it holds that

$$\mathcal{R}(1)_{i,i+r} - \mathcal{R}(1)_{i,i} = \frac{\Delta_n^{1/2} r^2}{8\sigma\zeta^3} + O(\Delta_n), \quad \text{(B.22)}$$

$$\mathcal{R}(k)_{i,i+r} - \mathcal{R}(k)_{i,i} = -\frac{1}{4\pi}\sum_{s=0}^{r-1}(2-\delta_{s,0})(r-s)\int_{-\pi}^{\pi}\frac{\partial f^{-1}(\lambda;\gamma)}{\partial\gamma_{k-1}}e^{\mathrm{i}\lambda s}d\lambda + O(\Delta_n^{1/2}). \quad \text{(B.23)}$$

Step 2 (Main proof) Using $(2\pi)^{-1}\int_{-\pi}^{\pi}(\partial f^{-1}(\lambda;\gamma)/\partial\gamma_{k-1})e^{\mathrm{i}\lambda s}d\lambda = (1-\delta_{s,0}/2)W(\gamma)_{k,s+1}$, we derive from (B.23) and the second part of (B.18) that for $0 \leq r \leq K$,

$$\sum_{k=2}^{q+2}(\bar{W}_n^{-1})_{1,k}(\mathcal{R}(k)_{i,i+r} - \mathcal{R}(k)_{i,i}) = -\frac{\sigma^2 r^2}{2\zeta^2} + O(\Delta_n^{1/2}).$$

Combined with (B.22) and the first part of (B.18), we prove claim (i). Using (B.20) and the second part of (B.19), we have, for $2 \leq l \leq q+2$ and $0 \leq r \leq K$,

$$\sum_{k=2}^{q+2}(\bar{W}_n^{-1} - n_T\widetilde{W}_n^{-1})_{l,k}\mathcal{R}(k)_{i,i+r} = O(\Delta_n^{1/2}). \quad \text{(B.24)}$$

In view of Lemmas A1 and A3, it holds by definition that

$$\mathcal{R}(k)_{i,i+r} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\delta_{k,1}\frac{\partial}{\partial\sigma^2} + \mathbb{1}_{\{k\geq 2\}}\frac{\partial}{\partial\gamma_{k-2}}\right)f^{-1}(\lambda;\sigma^2,\gamma,\Delta_n)\cos r\lambda d\lambda + O(\Delta_n^{1/2}).$$

Hence, by observing that $\cos r\lambda = \Delta_n^{-1}\partial f(\lambda;\sigma^2,\gamma,\Delta_n)/\partial\sigma^2 - \frac{1}{2}\sum_{k=0}^{r}(r-k)\partial f(\lambda;\sigma^2,\gamma,\Delta_n)/\partial\gamma_k$ for $0 \leq r \leq q+1$, we have, for $2 \leq l \leq q+2$ and $0 \leq r \leq K$,

$$\sum_{k=2}^{q+2}(\bar{W}_n^{-1})_{l,k}\mathcal{R}(k)_{i,i+r} = \mathbb{1}_{\{l\leq r+1\}}\frac{r+2-l}{2} + O(\Delta_n^{1/2}). \quad \text{(B.25)}$$

Combining (B.24) and (B.25) proves claim (ii). Claim (iii) comes directly from the expressions of $(\bar{W}_n^{-1})_{j,k}$ provided by (B.18) and (B.19) and the expression of $\mathcal{R}(k)_{i,i+r} - \mathcal{R}(k)_{i,i}$ provided by (B.22). For the first part of claim (iii) we additionally use (B.20) to obtain $\mathcal{W}(\sigma^2,\gamma;1)_{i,i}$, whereas for its second part we use claim (ii). ∎