TERG

Discussion Paper No. 239

Broadband semiparametric estimation of the long-memory parameter by the likelihood-based FEXP approach

Masaki Narukawa
and
Yasumasa Matsuda

November 2008

TOHOKU ECONOMICS RESEARCH GROUP

GRADUATE SCHOOL OF ECONOMICS AND
MANAGEMENT TOHOKU UNIVERSITY
KAWAUCHI, AOBA-KU, SENDAI,
980-8576 JAPAN

Broadband semiparametric estimation of the long-memory parameter by the likelihood-based FEXP approach

Masaki Narukawa
and
Yasumasa Matsuda

Tohoku University

Abstract

This paper proposes a semiparametric estimator of the long-memory parameter to fit a fractional exponential (FEXP) model by a likelihood-based approach. We establish that our proposed estimator is more efficient than the FEXP estimator proposed independently by Moulines and Soulier (1999) and Hurvuch and Brodsky (2001), and has the same asymptotic variance as the fractionally differenced autoregressive (FAR) estimator proposed by Bhansali et al. (2006) without pooling the periodogram. The Monte Carlo studies suggest that our estimator outperforms the FEXP estimator or is not inferior to the Gaussian semiparametric estimator (GSE) and will be also empirically effective in non-Gaussian processes.

1 Introduction

In this paper, we consider a covariance stationary process $\{X_t\}_{t\in\mathbb{Z}}$ with spectral density of the following form:

$$f(\lambda) = |1 - e^{i\lambda}|^{-2d} f^*(\lambda), \qquad \lambda \in [-\pi, \pi]. \tag{1}$$

where $-1/2 < d < 1/2$ and $f^*(\lambda)$ is an even, non-negative, continuous and bounded function such that $f^*(0) \neq 0$. The memory parameter d governs the behavior of the spectral density in a neighborhood of the zero frequency, so that in the case $0 < d < 1/2$, the process $\{X_t\}_{t\in\mathbb{Z}}$ is said to be long-range dependent, whereas the case $-1/2 < d < 0$ corresponds to the antipersistence where the spectral density at zero frequency is zero but the process is invertible. $f^*(\lambda)$ controls the short-memory behavior, and so the case $d = 0$ corresponds to the short-range or weak dependence as usual. Such long-memory models have recently been applied to many fields and the importance of that have rapidly increased (see, e.g. Robinson (1994), Beran (1994) or Doukhan et al. (2003)).

If properly finite dimensional parameterization of $f^*(\lambda)$ is assumed, the parameters of $f(\lambda)$ will be estimated using the parametric model aprroach, such as Fox and Taqqu (1986), Dalhaus (1989), Giraitis and Surgailis (1990) and Hosoya (1997) among others, which can be consisitent and asymptotically efficient estimator. However, the estimator may be inconsistent if the parameterization is misspecified, because the above argument is only provided that the parameterization is correctly specified. To avoid this drawback and because the most interest is usually in the estimation of the memory parameter d, we require semiparametric estimation of d. Among such estimations, to only take into account the behavior of $f^*(\lambda)$ at a neighborhood of the zero frequency is called *local methods* or *narrowband* because the frequencies used in estimation are restricted or trimmed to some extent around the zero frequency. Narrowband semiparametric estimators include, for example, the *GPH* estimator introduced by Geweke and Poter-Hudak (1983) and later exhaustively investigated by Robinson (1995a) and Hurvich et al.

(1998), or the *Gaussian semiparametric estimator* (GSE) proposed by Künsch (1987) and theoretically established by Robinson (1995b). The optimal choice of the trimming number or bandwidth for local methods is a rather complicated problem, so that it seems that the critical solutions to this problem are not established, though some approaches, such as *plug-in* methods (see, e.g. Hurvich and Deo (1999), Henry (2001)) or *adaptive estimation* (see, e.g. Giraitis et al. (2000)), have been proposed. Another semiparametric estimation is *global methods*, using the whole frequency range by assuming a regularity condition on $f^*(\lambda)$, and the term of *broadband* comes from such fact.

Broadband semiparametric estimators include two approaches, one of which is a *fractioanl exponential* (FEXP) approach and the other is a *fractionally differenced autoregressive* (FAR) approach. The FEXP approach is to fit a FEXP model to the log-periodogram regression at all Fourier frequencies by a least-squares procedure. The FEXP model is based on a Fourier series expansion of the logarithm of the short-memory component as follows: under appropriate regularity condition, $l^*(\lambda) = \log f^*(\lambda)$, may be expanded on the cosine basis,

$$l^*(\lambda) = \sum_{j=0}^{\infty} \theta_j h_j(\lambda), \qquad h_j(\lambda) = \cos(j\lambda). \tag{2}$$

Then, $\log f(\lambda)$ is given by $\log f(\lambda) = dg(\lambda) + l^*(\lambda)$, where $g(\lambda) = -2\log|1 - e^{i\lambda}|$. The class of which the expansion of $l^*(\lambda)$ is a finite number of cosin bases called FEXP models by Beran (1993), generalizing exponential models proposed by Bloomfield (1973). When a FEXP model is not regarded as a finite expansion a priori but a truncated expansion of the infinite expansion of $\log f(\lambda)$ and order of the truncation tends to infinity as $n \to \infty$, the FEXP approach is semiparametric. Such approach with the least-squares fitting is discussed by Robinson (1994) and theoretically investigated independently by Moulines and Soulier (1999) and Hurvich and Brodsky (2001). On the other hand, the FAR approach is to assume that the true spectral density obeys FAR(p, d) where the AR order p tends to infinity as $n \to \infty$, and so the estimator of d is obtained from fitting a FAR(p, d) model by a Whittle likelihood procedure. Such approach has been proposed by Bhansali et al. (2006) and they established the asymptotic properties of the FAR estimator of d. They also showed that the asymptotic variance of the estimator of d is 1, which implies that the FAR approach is more efficient than the FEXP approach, though the asymptotic variance of the FEXP estimator shown by Moulines and Soulier (1999) is dependent on pooling number J and theoretically tends to 1 only as $J \to \infty$.

The purpose of this paper is to propose a more efficient broadband semiparametric estimator for the FEXP model by a likelihood-based approach, called the *likelihood-based* FEXP approach. Since the distribution of the error terms in the log-periodogram regression obviously deviates from the normal distribution, the least-squares fitting of a FEXP model is not necessarily efficient as the above fact. In the short-range dependence context, the efficiency of the log-periodogram regression can be improved by the *maximum likelihood estimation* as in Fan and Kreutzberger (1998). This motivates us to apply the *maximum likelihood estimation* to fitting a FEXP model in the long-range dependence context. We establish that our estimator is consistent and asymptotically normal, and achieves the same asymptotic variance as Bhansali et al. (2006) without pooling, which suggests that the likelihood-based FEXP approach improves the efficiency of estimators and will give the asymptotically efficient estimator in global methods.

The paper is organized as follows. Section 2 describes the estimation procedure. In Section3, we state the assumptions and the asymptotic properties of our estimator. In Section 4, we provide a small Monte Carlo simulation to support the finite sample performance of our estimator by comparing with the other estimators (Robinson (1995b), Moulines and Soulier (1999) or Hurvich and Brodsky (2001)) and show an application for financial time series. Section 5 describes the concluding remarks. In Section 6, we provide the proofs of the main results given in Section 3.

2 Semiparametric estimation of the long-memory parameter

Suppose that $\{X_t\}_{t\in\mathbb{Z}}$ is a stationary Gaussian process and its spectral density $f(\lambda)$ is given by (1). Denoting $K_n = [n/2]$, the periodogram of $\{X_t\}$ is given by

$$I_n(\omega_k) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t e^{it\omega_k}\right|^2, \qquad k = 1, \cdots, 2K_n,$$

where $\omega_k = 2\pi k/n$ is the kth Fourier frequency.

The log-periodogram regression is based on the following identity:

$$\begin{aligned} Y_k &= \log f(\omega_k) + \log\left(\frac{I_n(\omega_k)}{f(\omega_k)}\right) \\ &= dg(\omega_k) + l^*(\omega_k) + \varepsilon_k, \end{aligned} \tag{3}$$

where $Y_k = \log I_n(\omega_k)$ and $\varepsilon_k = \log(I_n(\omega_k)/f(\omega_k))$. In the short-range dependence context, it is well known that the error terms ε_k are asymptotically independent and distributed as $\log(\frac{1}{2}\chi_2^2)$, where χ_2^2 is distributed as a central *chi-square* with 2 degrees of freedom (see, e.g. Brockwell and Davis, 1991, Theorem 10.3.2). We could, therefore, approximately regard $\{\varepsilon_k\}_{1\le k\le K_n}$ as K_n independently and identically distributed $\log(\frac{1}{2}\chi_2^2)$. As for the long-range dependence, it has been first shown by Künsch (1986) and later Hurvich and Beltrao (1993) and Robinson (1995a) that the normalized periodograms $I_n(\omega_k)/f(\omega_k)$ are asymptotically neither independent nor identically distributed, so that ε_k no longer have the above properties. However, the following decomposition of ε_k, which is derived from Theorem2 in Moulines and Soulier (1999), shows that the log-periodogram regression at all Fourier frequencies is relevant in the long-range dependence context. Under a Gaussianity of $\{X_t\}_{t\in\mathbb{Z}}$ and a global smoothness condition on f^* (corresponding to Assumption 2 in Section 3 below), there exists a constant $C < \infty$, such that for all $1 \le k \le K_n$,

$$\begin{aligned} \varepsilon_k &= \eta_k + r_k, \\ |r_k| &\le C\log(1+k)/k, \ w.p.1, \qquad (4)\\ |\mathrm{cov}(\eta_k, \eta_l)| &\le C\log^2(l)\, k^{-2|d|}\, l^{2|d|-2}, \qquad (5) \end{aligned}$$

with η_k is distributed as $\log(\frac{1}{2}\chi_2^2)$, and $E(\eta_k) = \psi(1)$ and $\mathrm{Var}(\eta_k) = \psi'(1)$, where $\psi(z)$ and $\psi'(z)$ denotes the *digamma function* and *trigamma function*, respectively.

Denote $g_k = g(\omega_k)$, $h_{j,k} = h_j(\omega_k)$ and $l^*_{p,k} = \sum_{j=p}^{\infty}\theta_j h_{j,k}$. By the Fourier expansion (2), the regression equation (3) is

$$Y_k = dg_k + \sum_{j=0}^{p-1}\theta_j h_{j,k} + l^*_{p,k} + \varepsilon_k = dg_k + \boldsymbol{\theta}'\mathbf{h}_{p,k} + l^*_{p,k} + \varepsilon_k,$$

where $\boldsymbol{\theta} = (\theta_0, \cdots, \theta_{p-1})'$, $\mathbf{h}_{p,k} = (h_{0,k}, \cdots, h_{p-1,k})'$ and $\boldsymbol{\theta}'\mathbf{h}_{p,k}$ is the truncated Fourier expansion or the FEXP of order p. Since our estimation is semiparametric, the truncated order p depends on n and $\lim_{n\to\infty} p_n = \infty$, which implies $l^*_{p,k}$ is an asymptically neglligible term and we can write the rgression equation as

$$Y_k \approx dg_k + \boldsymbol{\theta}'\mathbf{h}_{p,k} + \varepsilon_k. \tag{6}$$

This shows that the log-periodogram is approximated by the FEXP at all Forier frequancies, and Moulines and Soulier (1999) and Hurvich and Brodsky (2001) have independently proposed the least-squares estimation of the parameter $(d, \boldsymbol{\theta}')'$. The least-squares fitting is, however, regarded as assuming that $\{\varepsilon_k\}_{1\le k\le K_n}$ are indepedently and normally distributed with mean 0 and variance $\psi'(1)$ in view of the *maximum likelihood estimation*, which does not take into account the asymmetry of $\log(\frac{1}{2}\chi_2^2)$. Since this fact may imply the efficeiency loss of the estimators in the FEXP approach and η_k is distributed as $\log(\frac{1}{2}\chi_2^2)$, we will construct the likelihood based estimation of these parameter by pretending that $\{\varepsilon_k\}_{1\le k\le K_n}$ are indepedently and *log-gamma* distributed, which has the probability density function (see, e.g. Kotz and Nadarajah, 2000, p.48)

$$f_\varepsilon(x) = \exp\{-\exp(x) + x\}.$$

Such approach has been introduced by Fan and Kreutzberger (1998) for spectral density estimation in the short-range dependence context. For the spectral density $f(\lambda)$ in (1), the log-likelihood function associated with the regression equation and the *log-gamma* pdf is given by

$$\mathcal{L}_n(d,\boldsymbol{\theta}) = \frac{1}{K_n}\sum_{k=1}^{K_n}\left\{Y_k - dg_k - \boldsymbol{\theta}'\mathbf{h}_{p,k} - \exp\Big(Y_k - dg_k - \boldsymbol{\theta}'\mathbf{h}_{p,k}\Big)\right\}. \tag{7}$$

The maximum likelihood estimator $(\hat{d}_n, \hat{\boldsymbol{\theta}}_n')'$ is obtained by maximizing the function $\mathcal{L}_n(d,\boldsymbol{\theta})$. Since the form of $\mathcal{L}_n(d,\boldsymbol{\theta})$ is a non-linear but strictly concave function, the maximization of $\mathcal{L}_n(d,\boldsymbol{\theta})$ is easily implemented. The zero vector or the least-squares estimators work well as the initial estimators for the non-linear optimization.

3 Consistency and asymptotic distribution of the estimator

We now precisely state our assumptions below, which are required to derive the asymptotic properties of the estimator in the previous section.

Assumption 1. The process $\{X_t\}_{t\in\mathbb{Z}}$ is Gaussian.

Assumption 2. The spectral density f of the process $\{X_t\}_{t\in\mathbb{Z}}$ satisfies

$$f(\lambda) = \left|1 - e^{i\lambda}\right|^{-2d} f^*(\lambda), \qquad \lambda \in [-\pi,\pi],$$

where $0 \le d < 1/2$, and f^* is positive and differentiable on $[-\pi,\pi]\setminus\{0\}$ with

$$\forall\,\lambda \in [-\pi,\ \pi]\setminus\{0\}, \qquad \left|\frac{df^*(\lambda)}{d\lambda}\right| \le \frac{C}{|\lambda|},$$

for some finite constant $C < \infty$.

Assumption 3. $l^*(\lambda)$ has a convergent Fourier expansion,

$$l^*(\lambda) = \sum_{j=0}^{\infty}\theta_j h_j(\lambda), \qquad \lambda \in [-\pi,\pi],$$

with $\sum_{j=0}^{\infty} j^\alpha|\theta_j| < \infty$ for some real $\alpha > 1$.

Assumption 4a. Suppose that $\{p_n\}$ is an increasing sequence of integers such that $\lim_{n\to\infty} p_n = \infty$ and

$$\lim_{n\to\infty}\frac{p_n}{n} = 0.$$

Assumption 4b. Suppose that $\{p_n\}$ is an increasing sequence of integers such that $\lim_{n\to\infty} p_n = \infty$ and

$$\lim_{n\to\infty}\frac{p_n^2\log^2(n)}{n} = 0, \qquad \lim_{n\to\infty}\frac{n}{p_n^{2\alpha}} = 0.$$

Assumption 5. Let the parameter space Θ be the set of points $(d, \boldsymbol{\theta}') = (d, \theta_0, \theta_1, \cdots)$, satisfying $d \in [0, \frac{1}{2}]$ and $D = \left\{\boldsymbol{\theta} \in \mathbb{R}^\infty \mid \forall\, j,\ |\theta_j| \le K\, j^{-\delta}\right\}$ for some real $\delta > 1$ and finite constant $K < \infty$. Suppose that the true parameter $(d_0, \boldsymbol{\theta}_0') = (d^0, \theta_0^0, \theta_1^0, \cdots)$ is in the interior of the set Θ.

Let us give some comments on Assumptions. Gaussianity of the time series, as in Robinson (1995a) and Moulines and Soulier (1999), removes the complexity to evaluate the high order moment of non-linear transformation of the periodogram. Assumptions 1 and 2 are also necessary to derive (4) and (5). Assumption 5 is required to ensure compactness of the infinitely dimensional parameter space Θ.

The consitency of our likelihood-based FEXP estimator of the long-memory parameter d is shown by the following theorem and the proof is given in Section 6.

Theorem 1. Under Assumptions 1-3, 4a, 5, for the estimator $\hat{d}_n$ that maximizes the log-likelihood function (7),

$$\hat{d}_n \xrightarrow{p} d_0 \quad \text{as } n \to \infty.$$

Next, we state the asymptotic normality of our likelihood-based FEXP estimator of the long-memory parameter d.

Theorem 2. Under Assumptions 1-3, 4b, 5,

$$\sqrt{\frac{n}{p_n}}\left(\hat{d}_n - d_0\right) \xrightarrow{d} N(0,\ 1) \quad \text{as } n \to \infty.$$

The proof is given in Section 6. It should be noted that our estimator is asymptotically more efficient than that of Moulines and Soulier (1999), though the asymptotic variance of these authors is $J\psi'(J)$, where J is a pooling number, and $J\psi'(J) \to 1$ as $J \to \infty$. Since our likelihood-based FEXP estimator attains the asymptotic variance 1 without pooling, it is not necessary for our approach. Also, our estimatior achieves the same asymptotic variance as Bhansali et al. (2006), though they did not assume that the process is Gaussian. Our likelihood-based FEXP approach may possibly be extended to non-Gaussian case, by applying the same argument in Hurvich et al. (2002) which has established the asymptotic theory of the FEXP approach for a linear process by using the method in Fay and Soulier (2001). However, in this paper, we do not pursue the theoretical extension but the empirical demonstrations for non-Gaussian cases in the next section.

By the virtue of Theorem 2, we can construct LM statistic for the testing problem that

$$H_0 : d = 0 \quad \textit{against} \quad H_1 : d > 0. \tag{8}$$

Let $\beta = (d, \boldsymbol{\theta}')'$ and define

$$S_n(\tilde{\boldsymbol{\beta}}) = \left.\frac{\partial \mathcal{L}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{H_0}, \qquad \Omega_n(\tilde{\boldsymbol{\beta}}) = \left.\frac{\partial^2 \mathcal{L}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right|_{H_0}.$$

Then the standardized LM statistic for the one-sided testing problem is given by

$$t_{LM} = \sqrt{K_n}\frac{S_{n,1}(\tilde{\boldsymbol{\beta}})}{\sqrt{[\Omega_n(\tilde{\boldsymbol{\beta}})]_{11}}},$$

where $S_{n,1}(\tilde{\boldsymbol{\beta}})$ is the first element of $S_n(\tilde{\boldsymbol{\beta}})$, and $[\Omega_n(\tilde{\boldsymbol{\beta}})]_{11}$ is the $(1, 1)$th element of $\Omega_n(\tilde{\boldsymbol{\beta}})$.

Corollary. Under the null hypothesis H_0,

$$t_{LM} \xrightarrow{d} N(0,\ 1) \quad \text{as } n \to \infty.$$

The proof is given in Section 6. Corollary enables us to test the statistical hypothesis for the long-range dependence, which will be extended to the testing of *fractional unit roots* in a frequency domain, such as Lobato and Robinson (1998).

Remark. Theorem 1, 2 and Corollary are only established in the cases of the long-range dependence, $0 < d < 1/2$ and the short-range dependence, $d = 0$, while Moulines and Soulier (1999) have established the asymptotic theory of the FEXP estimator for $-1/2 < d < 1/2$. However, we can prove these theorems for the antipersistence, $-1/2 < d < 0$, in the same way, by considering the parameter space $d \in [-\frac{1}{2}, 0]$ instead of $d \in [0, \frac{1}{2}]$.

4 Empirical studies

This section investigates the finite sample behavior of our estimator and compares that with the existing semiparametric estimators of Robinson (1995b) and Moulines and Soulier (1999) in a Monte-Carlo simulation, and moreover we show a simple application for real data analysis. Note that the proposed estimator in this paper is denoted as $\hat{d}_{MLE}$.

4.1 Simulation

We now introduce two other semiparametric estimators for comparison purposes. One is the FEXP estimator based on the least-squares fitting of a FEXP model, as referred in Section 2, and the estimator of d is given by

$$\hat{d}_{LSE} = \arg \min_{d,\theta_0,\cdots,\theta_{p-1}} \sum_{k=1}^{K_n} \left\{ Y_k - dg_k - \boldsymbol{\theta}' \mathbf{h}_{p,k} \right\}^2 .$$

The other is a Gaussian semiparametric estimator (GSE) and the estimator of d is defined by Robinson (1995b) as

$$\hat{d}_{GSE} = \arg \min_{d \in (-1/2, 1/2)} \left(\log \left\{ \frac{1}{m} \sum_{k=1}^{m} \omega_k^{2d} I_n(\omega_k) \right\} - \frac{2d}{m} \sum_{k=1}^{m} \log \omega_k \right),$$

where m is a bandwidth or trimming parameter for local methods. Robinson (1995b) suggests that $\hat{d}_{GSE}$ can be identified with the most efficient semiparametric estimator of d in local methods, and so $\hat{d}_{GSE}$ is the benchmark in local methods for comparison purposes. It is noted that since pooling is also irrelevant for $\hat{d}_{GSE}$ that is based on the Whittle likelihood, pooling is not considered in the simulation below.

The design of data generating process (DGP) for our Monte-Carlo simulation is based on, such as Diebold and Rudebusch (1991), the Choleski decomposition of the autocovariance function of *fractionally integrated noise*, defined as

$$(1 - B)^d X_t = \varepsilon_t,$$

where B is the backward shift operator. By using the above, we generate two cases of a ARFIMA$(1, 0.3, 1)$ process $\{X_t\}_{t \in \mathbb{Z}}$ as follows:

$$(1 - \phi B)(1 - B)^{0.3} X_t = (1 + \theta B)\varepsilon_t,$$

where $\{\varepsilon_t\}$ are designed to be independent (i) *standard normal* and (ii) *Uniform* $[-\sqrt{3}, \sqrt{3}]$. The uniform distribution is selected as a typical example of non-Gaussian distribution to examine whether $\hat{d}_{MLE}$ is empirically robust in the comparison with $\hat{d}_{LSE}$, as in Velasco (2000). The values for (ϕ, θ) are $(0.4, 0)$ and $(0.2, 0.6)$. The length of the series is $T = 501$ and 1000 independent replications of each time series are generated. We consider the following cases for the truncated order p and the bandwidth parameter m as $p = 1, \cdots, 6$ and $m = [T^\delta]$ with $\delta = 0.3, \cdots, 0.8$, respectively.

Tables 1 and 2 report the results of our Monte Carlo simulation in Gaussian case, where *Bias*, *Std.dev* and *MSE* denote the bias, the standard deviation and the mean squared error of the estimators, calculated across replications, respectively. Also, *Size* and *Power* are each calculated as the empirical size when $d = 0$ and the power when $d = 0.3$ at the nominal 5% significant level for one-sided *LM* statistic and t-statistic corresponding to (8). Our estimator $\hat{d}_{MLE}$ is the least bias in comparison with the two other estimators in all the cases when both p and δ are chosen at the minimum of bias. Similarly, in the view of *MSE*, $\hat{d}_{MLE}$ is less than $\hat{d}_{LSE}$ in all the cases and less than $\hat{d}_{GSE}$ for $(\phi, \theta) = (0.4, 0.0)$, whereas $\hat{d}_{GSE}$ has less *MSE* for $(\phi, \theta) = (0.2, 0.6)$, when the optimal p and δ are chosen, respectively. As for *Size* and *Power*, the same tendency as the above is indicated and it is obvious that our one-sided *LM* statistic with $\hat{d}_{MLE}$ is more powerful than that of t-statistic with $\hat{d}_{LSE}$.

Thus, our estimator $\hat{d}_{MLE}$ is clearly more desirable than $\hat{d}_{LSE}$, as established by the asymptotic properties in Section 3, and moreover the performance of $\hat{d}_{MLE}$ is not inferior to $\hat{d}_{GSE}$ at the optimal p and δ. However, it should be noted that the optimal p or δ based on the minimum of MSE is not always appropriate with respect to *Bias* or *Size*, which implies that the optimal choice of p or δ are critical for these semiparametric estimators but a considerably complicated problem.

Tables 3 and 4 indicate that replacing Gaussian innovations with non-Gaussian ones that are uniformly distributed does not essentially affect the performance of the estimators. Although there is no theoretical results of the likelihood-based FEXP estimation with non-Gaussian innovations, these results

Table 1. $(\phi, \theta) = (0.4, 0.0)$ with *Gaussian* innovations.

$\hat{d}_{MLE}$

p	Bias	Std.dev	MSE	Size	Power
1	0.32379	0.04143	0.10655	1.000	1.000
2	0.08839	0.06450	0.01197	0.380	1.000
3	0.02384	0.08397	0.00762	0.078	0.983
4	-0.00032	0.10202	0.01041	0.033	0.899
5	-0.00864	0.11936	0.01432	0.035	0.800
6	-0.01748	0.13642	0.01892	0.033	0.699

$\hat{d}_{LSE}$

p	Bias	Std.dev	MSE	Size	Power
1	0.29438	0.04522	0.08871	1.000	1.000
2	0.09439	0.07823	0.01503	0.339	0.997
3	0.03534	0.10169	0.01159	0.090	0.948
4	0.01723	0.12463	0.01583	0.055	0.806
5	0.01112	0.14610	0.02147	0.049	0.685
6	0.00468	0.16746	0.02807	0.046	0.587

$\hat{d}_{GSE}$

δ	Bias	Std.dev	MSE	Size	Power
0.3	-0.05391	0.39440	0.15846	0.005	0.043
0.4	-0.02549	0.21025	0.04486	0.009	0.272
0.5	0.00258	0.13550	0.01837	0.027	0.659
0.6	0.03250	0.08997	0.00915	0.073	0.956
0.7	0.09594	0.06556	0.01350	0.394	1.000
0.8	0.19827	0.04726	0.04154	0.992	1.000

suggest that our approach will be also effective in non-Gaussian cases and give a robust estimator.

The question about the optimal choice of the truncated order p has been solved, in view of *model selection*, for the FEXP approach by Moulines and Soulier (2000) based on Mallow's C_p criterion and Hurvich (2001) based on Mallow's C_L criterion. These criterion, however, were constructed for *linear-regression models* and cannot apply for our likelihood-based FEXP approach. Instead, we empirically propose a data-driven selection of p based on *Akaike Information Criterion* (AIC, Akaike (1973)), which is the model selection criterion constructed from the *log-likelihood*. Our proposed method is to choose so as to minimize

$$AIC(p) = -2\Big\{K_n \mathcal{L}_n(\hat{d}, \hat{\boldsymbol{\theta}}) - (p+1)\Big\},$$

where $\mathcal{L}_n(\hat{d}, \hat{\boldsymbol{\theta}})$ is the estimated log-likelihood function (7), over the set of all examined values for p. We do not pursue the theoretical validity of this criterion in our context but show the numerical results by the simulation below.

Table 5 reports the rate of p in the same DGP, length of series and replications as the above with Gaussian innovations and the examined values of p are $p = 1, \cdots, 10$. In the case $(\phi, \theta) = (0.4, 0.0)$, although the optimal p is $p = 3$ or $p = 4$ from Table 1, Table 5 indicates that about half of the replications select $p = 2$, which suggests that the selected p based on $AIC(p)$ will be somewhat biased downward in this case. As for the case $(\phi, \theta) = (0.2, 0.6)$, the selected p is a little dispersive but gather around $p = 3$

Table 2. $(\phi, \theta) = (0.2, 0.6)$ with *Gaussian* innovations.

$\hat{d}_{MLE}$

p	Bias	Std.dev	MSE	Size	Power
1	0.52428	0.05024	0.27740	1.000	1.000
2	-0.10474	0.06445	0.01513	0.000	0.933
3	0.05477	0.08429	0.01010	0.145	0.993
4	-0.03638	0.10212	0.01175	0.013	0.835
5	0.00134	0.11943	0.01427	0.042	0.823
6	-0.02552	0.13650	0.01928	0.026	0.679

$\hat{d}_{LSE}$

p	Bias	Std.dev	MSE	Size	Power
1	0.48299	0.04553	0.23535	1.000	1.000
2	-0.10060	0.07877	0.01633	0.000	0.811
3	0.06504	0.10177	0.01459	0.152	0.967
4	-0.01796	0.12474	0.01588	0.018	0.731
5	0.02061	0.14625	0.02181	0.056	0.713
6	-0.00316	0.16772	0.02814	0.042	0.577

$\hat{d}_{GSE}$

δ	Bias	Std.dev	MSE	Size	Power
0.3	-0.05477	0.39472	0.15880	0.006	0.042
0.4	-0.02787	0.21027	0.04499	0.009	0.271
0.5	-0.00392	0.13547	0.01837	0.023	0.644
0.6	0.01456	0.08988	0.00829	0.049	0.947
0.7	0.05950	0.06522	0.00779	0.217	0.999
0.8	0.18166	0.04746	0.03525	0.971	1.000

or $p = 4$ and, in fact, around 80 percent of the replications is within 3, 4 and 5, where *Bias* and *MSE* of $\hat{d}_{MLE}$ are comparatively less than that of the other p. From these results, the choice of p based on *AIC(p)* will be empirically considered as useful for the optimal selection, though such p can be slightly smaller than the optimal p in some cases.

4.2 Application

This section shows the simple application of semiparametric estimation of the long-range dependence to financial time series, which is the long-range dependence in volatility. We now introduce the return series $\{r_t\}$ from a financial asset at time t as follows:

$$r_t = \log(p_t) - \log(p_{t-1}),$$

where p_t is the price for a financial asset at time t. To investigate the long-range dependence in volatility, the most simple approach is to semiparametrically estimate the long-memory parameter d by using squared return series $\{r_t^2\}$. We study such point with our likelihood-based FEXP and the FEXP approach for the returns from stock and foreign exchange rates, respectively.

The data studied is daily series of Dow Jones Industrial average (DJI) obtained from *yahoo finance* for stock market and of Euro/Dollar rate for exchange rates obtained from *Bank of Japan*, respectively.

Table 3. $(\phi, \theta) = (0.4, 0.0)$ with *Uniform* innovations.

$\hat{d}_{MLE}$

p	Bias	Std.dev	MSE	Size	Power
1	0.32434	0.04248	0.10700	1.000	1.000
2	0.08929	0.06629	0.01237	0.391	1.000
3	0.02346	0.08636	0.00801	0.071	0.981
4	-0.00124	0.10303	0.01062	0.044	0.895
5	-0.00973	0.11746	0.01389	0.033	0.800
6	-0.01407	0.13556	0.01858	0.029	0.703

$\hat{d}_{LSE}$

p	Bias	Std.dev	MSE	Size	Power
1	0.29452	0.04772	0.08902	1.000	1.000
2	0.09083	0.08046	0.01473	0.306	0.998
3	0.03309	0.10836	0.01284	0.092	0.936
4	0.01177	0.13005	0.01705	0.060	0.796
5	0.00509	0.15038	0.02264	0.053	0.684
6	0.00392	0.17237	0.02973	0.044	0.568

$\hat{d}_{GSE}$

δ	Bias	Std.dev	MSE	Size	Power
0.3	-0.05317	0.40499	0.16684	0.002	0.047
0.4	-0.02573	0.22048	0.04928	0.015	0.312
0.5	0.00228	0.14167	0.02008	0.027	0.652
0.6	0.03140	0.09237	0.00952	0.068	0.952
0.7	0.09539	0.06751	0.01366	0.384	1.000
0.8	0.19977	0.04874	0.04228	0.989	1.000

Table 6 reports the semiparametric estimators of d and the values in the parenthesis below are one-sided LM-statistic and t-statistic for testing the hypothesis of (8), respectively. We also present the each estimators of the return series r_t for comparison purposes. $\hat{d}_{MLE}$ and LM-statistic tend to take the larger values than $\hat{d}_{LSE}$ and t-statistic, which implies that our approach is more powerful. The long-range

Table 4. $(\phi, \theta) = (0.2, 0.6)$ with *Uniform* innovations.

$\hat{d}_{MLE}$

p	Bias	Std.dev	MSE	Size	Power
1	0.52539	0.05130	0.27867	1.000	1.000
2	-0.10350	0.06543	0.01500	0.046	0.998
3	0.05407	0.08697	0.01049	0.043	0.965
4	-0.03694	0.10282	0.01194	0.033	0.883
5	0.00003	0.11757	0.01382	0.030	0.797
6	-0.02221	0.13548	0.01885	0.027	0.700

$\hat{d}_{LSE}$

p	Bias	Std.dev	MSE	Size	Power
1	0.48265	0.04802	0.23526	1.000	1.000
2	-0.10352	0.08041	0.01718	0.000	0.796
3	0.06203	0.10852	0.01562	0.142	0.957
4	-0.02304	0.13012	0.01746	0.030	0.711
5	0.01443	0.15060	0.02289	0.058	0.701
6	-0.00375	0.17183	0.02954	0.037	0.559

$\hat{d}_{GSE}$

δ	Bias	Std.dev	MSE	Size	Power
0.3	-0.05399	0.40493	0.16689	0.002	0.047
0.4	-0.02809	0.22045	0.04939	0.015	0.309
0.5	-0.00423	0.14160	0.02007	0.027	0.636
0.6	0.01340	0.09224	0.00869	0.052	0.939
0.7	0.05901	0.06698	0.00797	0.212	1.000
0.8	0.18306	0.04884	0.03590	0.973	1.000

Table 5. Rates of selected p with *Gaussian* innovations in 1000 replications.

p	$(\phi, \theta) = (0.4, 0.0)$	$(\phi, \theta) = (0.2, 0.6)$
1	2	0
2	507	21
3	269	342
4	94	344
5	41	121
6	26	63
7	15	29
8	18	37
9	16	23
10	12	20

Table 6. Estimators of the long-memory parameter in return and squared return series.

DJI average (T=900)	$\hat{d}_{MLE}$	$\hat{d}_{LSE}$
r_t	-0.0697	-0.11077
	(-2.7338)	(-3.0330)
r_t^2	0.48005	0.423883
	(7.0329)	(4.6328)
Euro/Dollar rate (T=817)	$\hat{d}_{MLE}$	$\hat{d}_{LSE}$
r_t	0.01349	0.05671
	(0.47397)	(1.4705)
r_t^2	0.23781	0.154632
	(5.1394)	(2.6555)

dependence is detected only in the squared series, which suggests that the volatility of both series has the long-range dependence. Note that it is necessary for estimation of the long-memory parameter in volatility to consider a more elaborate model such as a *long-memory stochastic volatility* (LMSV) or a *fractionally integrated exponential generalized autoregressive conditional heteroskedasticity* (FIE-GARCH) model, and more specifically see, e.g. Deo and Hurvich (2003), Hurvich et al. (2005) and the references therein.

5 Concluding remarks

In this paper, we have proposed the broadband semiparametric estimator of the long-memory parameter of a long-range dependent time series using the likelihood-based FEXP approach and established the asymptotic properties of the estimator. We have also shown that the likelihood-based FEXP approach gives the more efficient semiparametric estimator than that of the FEXP approach, which achieves the same asymptotic variance as Bhansali et al. (2006) without pooling the preriodogram.

The simulation studies have supported the validity of the theoretical results of our estimator and shown that *AIC(p)* is effective for the data-driven selection of the truncated order p. Moreover the likelihood-based FEXP approach is not only useful and preferable in Gaussian case but also empirically robust in non-Gaussian case, though there is no theoretical results of our estimator in non-Gaussian processes.

The asymptotic theory of our estimator in Theorem 1 and 2 depends on the assumption that the processes are Gaussian. We have not pursued the non-Gaussian extension in this paper, so that an important task remained open is to prove without the Gaussian assumption. The theoretical validity of the model selection based on *AIC(p)* for the likelihood-based FEXP approach have not been established, which is also left to the future study.

6 Appendix

This section proves Theorem 1, 2 and Corollary, where Lemmas employed in the proofs are given in Section 6.2. Throughout this section, $C, C^*, C_1, C_2, \cdots$, are a positive generic constant term, but not always the same one in each context.

6.1 Proofs of Theorem 1, 2 and Corollary

Proof of Theorem 1. To prove Theorem 1, we shall follow Lemma 2 in Walker (1964) adapted to the infinitely dimensional compact parameter space.

Throughout the proofs, let the parameter space Θ be a metric space with the metric defined as

$$\|(d,\boldsymbol{\theta}')\| = |d| + \sum_{j=0}^{\infty} |\theta_j|,$$

which is induced by l_1-norm. It follows from Assumption 5 that the metric space Θ equipped with l_1-norm is *totally bounded* and *complete*. Therefore, it is compact (see, e.g. Kolmogorov and Fomin, 1970, p.100).

Let $(d_0,\boldsymbol{\theta}_0') = (d^0,\theta_0^0,\theta_1^0,\cdots,\theta_{p-1}^0,\cdots)$ be the true parameter values, and $(d,\boldsymbol{\theta}') = (d,\theta_0,\theta_1,\cdots,\theta_{p-1},0,0,\cdots)$ be any admissible values of the truncated order p. By equation (6), the log-likelihood function (7) rewrite,

$$\begin{aligned}\mathcal{L}_n(d,\boldsymbol{\theta}) = \frac{1}{K_n}\sum_{k=1}^{K_n}\Big\{&(d_0-d)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_p)'\mathbf{h}_{p,k} + l^*_{(0),p,k} + \eta_k + r_k\\ &- \exp\Big((d_0-d)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_p)'\mathbf{h}_{p,k} + l^*_{(0),p,k} + \eta_k + r_k\Big)\Big\}. \qquad (9)\end{aligned}$$

where $\boldsymbol{\theta}_{0,p} = (\theta_0^0,\theta_1^0,\cdots,\theta_{p-1}^0)'$, $\boldsymbol{\theta}_p = (\theta_0,\theta_1,\cdots,\theta_{p-1})'$ and $l^*_{(0),p,k} = \sum_{j=p}^{\infty}\theta_j^0 h_{j,k}$.

We must first show that these terms satisfy the following properties: as $n\to\infty$,

$$\mathcal{L}_n(d,\boldsymbol{\theta}) \xrightarrow{p} \mathcal{L}_\infty(d,\boldsymbol{\theta}),$$

where

$$\mathcal{L}_\infty(d,\boldsymbol{\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big\{(d_0-d)g(\lambda) + l^*_{(0)}(\lambda) - l^*(\lambda) - \exp\Big((d_0-d)g(\lambda) + l^*_{(0)}(\lambda) - l^*(\lambda)\Big) + \psi(1)\Big\}d\lambda,$$

with $l^*_{(0)}(\lambda) = \sum_{j=0}^{\infty}\theta_j^0 h_j(\lambda)$. By Lemma 3 and the square integrability of $g(\lambda)$, the expectation of $\mathcal{L}_n(d,\boldsymbol{\theta})$ is evaluated as

$$\begin{aligned}E\Big(\mathcal{L}_n(d,\boldsymbol{\theta})\Big) &= \frac{1}{K_n}\sum_{k=1}^{K_n}\Big\{(d_0-d)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_p)'\mathbf{h}_{p,k} + l^*_{(0),p,k} + \psi(1)\\ &\quad - \exp\Big((d_0-d)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_p)'\mathbf{h}_{p,k} + l^*_{(0),p,k}\Big)\Big\}\\ &\quad + \frac{1}{K_n}\sum_{k=1}^{K_n} O\Big(\frac{\log(1+k)}{k}\Big) - \frac{1}{K_n}\sum_{k=1}^{K_n} O\left\{\Big(2\sin\Big(\frac{\omega_k}{2}\Big)\Big)^{-2\mu}\right\} O\Big(\frac{\log(1+k)}{k}\Big)\\ &= \frac{1}{K_n}\sum_{k=1}^{K_n}\Big\{(d_0-d)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_p)'\mathbf{h}_{p,k} + O(p^{-\alpha}) + \psi(1)\\ &\quad + \exp\Big((d_0-d)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_p)'\mathbf{h}_{p,k}+\Big)\Big\} + O\Big(\frac{\log^2(n)}{n}\Big) - O\Big(n^{2\mu}\Big)\cdot O\Big(\frac{\log^2(n)}{n}\Big)\\ &= \frac{1}{K_n}\sum_{k=1}^{K_n}\Big\{(d_0-d)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_p)'\mathbf{h}_{p,k}\\ &\quad - \exp\Big((d_0-d)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_p)'\mathbf{h}_{p,k}\Big) + \psi(1)\Big\} + O(p^{-\alpha}) + O\Big(\frac{\log^2(n)}{n^{1-2\mu}}\Big), \qquad (10)\end{aligned}$$

where $|d_0 - d| < \mu < 1/2$, because $|d| < 1/2$ and d has the same sign as d_0, so that $0 < 1 - 2\mu < 1$. The second equality of the expectation follows from $\lim_{x\to 0}\sin(x)/x = 1$, Assumption 3 that yields $l^*_{(0),p,k} = O(p^{-\alpha})$, and Taylor's expansion of $\exp(O(p^{-\alpha}))$. In addition to the previous Lemma, by Lemma 2, 4, 5 and Cauchy-Schwarz inequality, the variance of $\mathcal{L}_n(d,\boldsymbol{\theta})$ is evaluated as

$$
\begin{aligned}
\mathrm{Var}\Big(\mathcal{L}_n(d,\boldsymbol{\theta})\Big) &= E\left\{\frac{1}{K_n}\sum_{k=1}^{K_n}(\eta_k - E(\eta_k))\right\}^2 + E\left\{\frac{1}{K_n}\sum_{k=1}^{K_n}(r_k - E(r_k))\right\}^2 \\
&\quad + E\left\{\frac{1}{K_n}\sum_{k=1}^{K_n}\exp\Big((d_0-d)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_p)'\mathbf{h}_{p,k} + l^*_{(0),p,k}\Big)\right. \\
&\qquad \left.\times\Big(\exp(\eta_k) - E(\exp(\eta_k))\Big)\left(1 + O\left(\frac{\log(1+k)}{k}\right)\right)\right\}^2 + \mathit{Cross\ terms} \\
&\le \frac{1}{K_n^2}\cdot O(n) + \frac{1}{K_n^2}\cdot O\Big(\log^4(n)\Big) \\
&\quad + \frac{1}{K_n}\cdot O\Big(n^{2\mu}\Big)\cdot\frac{1}{K_n}\sum_{k=1}^{K_n} O\left\{\left(2\sin\left(\frac{\omega_k}{2}\right)\right)^{-2\mu}\right\} + \frac{1}{K_n^2}\cdot O\left\{\left(n^{2\mu}\right)^2\cdot\log^3(n)\right\} \\
&= O\left(\frac{1}{n}\right) + O\left(\frac{\log^4(n)}{n^2}\right) + O\left(\frac{1}{n^{1-2\mu}}\right) + O\left(\frac{\log^3(n)}{n^{2-4\mu}}\right) = O\left(\frac{\log^3(n)}{n^{2-4\mu}}\right), \qquad (11)
\end{aligned}
$$

The second equality follows from the integrability of $\sin^{\nu}(x)$ for $\nu > -1$ on $(0,\pi/2)$ (see, e.g. Gradshteyn and Ryzhik, 1965, p.369). By the condition of μ and Assumption 4a, (10) and (11) yield $\mathcal{L}_n(d,\boldsymbol{\theta}) \xrightarrow{p} \mathcal{L}_\infty(d,\boldsymbol{\theta})$, as $n\to\infty$.

Next, let $(d_1,\boldsymbol{\theta}_1') \neq (d_0,\boldsymbol{\theta}_0')$. Substituting these parameter for (9), respectively, as $n\to\infty$, $\mathcal{L}_n(d_0,\boldsymbol{\theta_0}) \xrightarrow{p} \mathcal{L}_\infty(d_0,\boldsymbol{\theta}_0)$, where $\mathcal{L}_\infty(d_0,\boldsymbol{\theta}_0) = \psi(1) - 1$, and $\mathcal{L}_n(d_1,\boldsymbol{\theta}_1) \xrightarrow{p} \mathcal{L}_\infty(d_1,\boldsymbol{\theta}_1)$, where

$$
\mathcal{L}_\infty(d_1,\boldsymbol{\theta}_1) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{(d_0-d_1)g(\lambda) + l^*_{(0)}(\lambda) - l^*_{(1)}(\lambda) - \exp\Big((d_0-d_1)g(\lambda) + l^*_{(0)}(\lambda) - l^*_{(1)}(\lambda)\Big) + \psi(1)\right\}d\lambda.
$$

Since $\{h_0(\lambda), h_1(\lambda), \cdots\}$ is an orthogonal system in $\mathbb{L}^2$, the identifiability condition of the parameter $(d,\boldsymbol{\theta}')$ is satisfied, which yields

$$
\begin{aligned}
\mathcal{L}_\infty(d_0,\boldsymbol{\theta}_0) - \mathcal{L}_\infty(d_1,\boldsymbol{\theta}_1) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\exp\Big((d_0-d_1)g(\lambda) + l^*_{(0)}(\lambda) - l^*_{(1)}(\lambda)\Big)\right. \\
&\qquad \left. - \Big((d_0-d_1)g(\lambda) + l^*_{(0)}(\lambda) - l^*_{(1)}(\lambda)\Big) - 1\right\}d\lambda \\
&> 0.
\end{aligned}
$$

It follows that, for any positive constant $G[(d_0,\boldsymbol{\theta}_0),(d_1,\boldsymbol{\theta}_1)]$ less than $\mathcal{L}_\infty(d_0,\boldsymbol{\theta}_0) - \mathcal{L}_\infty(d_1,\boldsymbol{\theta}_1)$,

$$
\lim_{n\to\infty} P\Big\{\mathcal{L}_n(d_1,\boldsymbol{\theta}_1) - \mathcal{L}_n(d_0,\boldsymbol{\theta}_0) < -G[(d_0,\boldsymbol{\theta}_0),(d_1,\boldsymbol{\theta}_1)]\Big\} = 1.
$$

For any $\delta > 0$, there exists $H_{n,\delta}$ such that, for any $(d_1,\boldsymbol{\theta}_1')$ and $(d_2,\boldsymbol{\theta}_2')$ that satisfies $\| (d_1,\boldsymbol{\theta}_1') - (d_2,\boldsymbol{\theta}_2') \| < \delta$,

$$
\Big|\mathcal{L}_n(d_1,\boldsymbol{\theta}_1) - \mathcal{L}_n(d_2,\boldsymbol{\theta}_2)\Big| < H_{n,\delta},
$$

where $H_{n,\delta}$ is defined as

$$
H_{n,\delta} := \frac{1}{K_n}\sum_{k=1}^{K_n}\Big\{\Big|(d_2-d_1)g_k\Big| + \Big|(\boldsymbol{\theta}_{2,p}-\boldsymbol{\theta}_{1,p})'\mathbf{h}_{p,k}\Big|
$$

$$+ I_n(\omega_k)\Big|\exp\Big(-d_2 g_k - \boldsymbol{\theta}'_{2,p}\mathbf{h}_{p,k}\Big) - \exp\Big(-d_1 g_k - \boldsymbol{\theta}'_{1,p}\mathbf{h}_{p,k}\Big)\Big|\Bigg\},$$

and has the following two properties:

$$\lim_{\delta\to 0} E(H_{n,\delta}) = 0, \tag{12}$$

because, by Lemma 3 and Taylor's expansion,

$$\begin{aligned}
E(H_{n,\delta}) = {} & \frac{1}{K_n}\sum_{k=1}^{K_n}\Big\{\Big|(d_2-d_1)g_k\Big| + \Big|(\boldsymbol{\theta}_{2,p}-\boldsymbol{\theta}_{1,p})'\mathbf{h}_{p,k}\Big|\Big\} \\
& + \frac{1}{K_n}\sum_{k=1}^{K_n}\Big(1 + O\Big(\frac{\log(1+k)}{k}\Big)\Big)\Big(1 + O\Big(p^{-\alpha}\Big)\Big) \\
& \times \exp\Big(d_0 g_k + \boldsymbol{\theta}'_{0,p}\mathbf{h}_{p,k}\Big)\Big|\exp\Big(-d_2 g_k - \boldsymbol{\theta}'_{2,p}\mathbf{h}_{p,k}\Big) - \exp\Big(-d_1 g_k - \boldsymbol{\theta}'_{1,p}\mathbf{h}_{p,k}\Big)\Big| \\
\leq {} & \frac{1}{K_n}\sum_{k=1}^{K_n}\Big|d_2-d_1\Big|\Big|g_k\Big| + \frac{1}{K_n}\sum_{k=1}^{K_n}\Big|\boldsymbol{\theta}_{2,p}-\boldsymbol{\theta}_{1,p}\Big|'\Big|\mathbf{h}_{p,k}\Big| \\
& + \frac{C_1}{K_n}\sum_{k=1}^{K_n}\Big|(d_1-d_2)g_k\Big|\exp\Big((d_0-d^*)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_{2,p})'\mathbf{h}_{p,k}\Big) \\
& + \frac{C_2}{K_n}\sum_{k=1}^{K_n}\Big|(\boldsymbol{\theta}_{2,p}-\boldsymbol{\theta}_{1,p})'\mathbf{h}_{p,k}\Big|\exp\Big((d_0-d_1)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}^*_p)'\mathbf{h}_{p,k}\Big) \\
\leq {} & \Big|d_1-d_2\Big|\cdot C_1 + \Big|d_1-d_2\Big|\cdot\frac{C_2}{K_n}\sum_{k=1}^{K_n}\exp\Big((d_0-d^*)g_k\Big)\Big|g_k\Big| + \Big\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\Big\|\cdot\frac{C_3}{K_n}\sum_{k=1}^{K_n}\exp\Big((d_0-d^*)g_k\Big) \\
\leq {} & \Big|d_1-d_2\Big|\cdot C_1 + \Big\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\Big\|\cdot C_2,
\end{aligned}$$

where d^* or $\boldsymbol{\theta}^*_p$ lie between d_1 and d_2 or $\boldsymbol{\theta}_{1,p}$ and $\boldsymbol{\theta}_{2,p}$, respectively. The second inequality follows from the square integrability of $g(\lambda)$ and the last one follows from the integrability of $\sin^{\nu}(x)$ and $\sin^{\nu}(x)\log(\sin(x))$ for $\nu > -1$ on $(0, \pi/2)$, respectively (see, e.g. Gradshteyn and Ryzhik, 1965, p.369 and p.588). And the other property is given by

$$\lim_{n\to\infty} \mathrm{Var}(H_{n,\delta}) = 0, \tag{13}$$

because, by the same argument of (11),

$$\begin{aligned}
\mathrm{Var}(H_{n,\delta}) \leq {} & \frac{C_1}{K_n^2}E\left\{\sum_{k=1}^{K_n}\exp\Big((d_0-d_1)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_{1,p})'\mathbf{h}_{p,k}\Big)\Big(\exp(\eta_k) - E(\exp(\eta_k))\Big)\right\}^2 \\
& + \frac{C_2}{K_n^2}E\left\{\sum_{k=1}^{K_n}\exp\Big((d_0-d_2)g_k + (\boldsymbol{\theta}_{0,p}-\boldsymbol{\theta}_{2,p})'\mathbf{h}_{p,k}\Big)\Big(\exp(\eta_k) - E(\exp(\eta_k))\Big)\right\}^2 + \mathit{Cross\ term} \\
= {} & O\Big(\frac{\log^3(n)}{n^{2-4\mu}}\Big).
\end{aligned}$$

Now, (12) and (13) correspond to (16) and (17) in Walker (1964), respectively, which implies that there exists $\delta > 0$ such that

$$\lim_{n\to\infty} P\Big\{H_{n,\delta} < G[(d_0, \boldsymbol{\theta}_0), (d_1, \boldsymbol{\theta}_1)]\Big\} = 1.$$

Then, since the parameter space Θ is compact in $\mathbb{R}^\infty$ by Assumption 5, Theorem 1 follows from Lemma 2 in Walker (1964). □

Proof of Theorem 2. Denoting $\boldsymbol{\beta} = (d, \theta_0, \theta_1, \cdots, \theta_{p-1})$, by the mean value theorem,

$$0 = \frac{\partial \mathcal{L}_n(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} + \frac{\partial^2 \mathcal{L}_n(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0),$$

where $\boldsymbol{\beta}^*$ lies between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}_n$ with $\boldsymbol{\beta}_0 = (d^0, \theta_0^0, \theta_1^0, \cdots, \theta_{p-1}^0)$ and $\hat{\boldsymbol{\beta}}_n = (\hat{d}_n, \hat{\theta}_0, \hat{\theta}_1, \cdots, \hat{\theta}_{p-1})$. We obtain

$$\sqrt{\frac{K_n}{p}}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \Omega_n^{-1}(\boldsymbol{\beta}^*)\Lambda_n(\boldsymbol{\beta}_0),$$

with

$$\Lambda_n(\boldsymbol{\beta}_0) = \sqrt{\frac{K_n}{p}} \frac{\partial \mathcal{L}_n(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}, \qquad \Omega_n(\boldsymbol{\beta}^*) = -\frac{\partial^2 \mathcal{L}_n(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}.$$

Now, by Theorem 1 and Lemma 6,

$$\Omega_n^{-1}(\boldsymbol{\beta}^*)\Lambda_n(\boldsymbol{\beta}_0) = \Gamma_n^{-1}(p)\Lambda_n(\boldsymbol{\beta}_0) + O_p\left(p^{-\alpha} + \frac{\log(n)}{n^{1/2}}\right), \tag{14}$$

where

$$\Gamma_n(p) = \frac{1}{K_n} \sum_{k=1}^{K_n} \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}',$$

with $\mathbf{h}_p(\lambda) = (h_0(\lambda), \cdots, h_{p-1}(\lambda))'$. To prove Theorem 2, we only need the first element of $\Omega_n^{-1}(\boldsymbol{\beta}^*)\Lambda_n(\boldsymbol{\beta}_0)$ because that corresponds to the long-memory parameter d. Let $\boldsymbol{\gamma}_n(p)$ denotes the first row of $\Gamma_n^{-1}(p)$, then by the equation (14)

$$\sqrt{\frac{K_n}{p}}(\hat{d}_n - d_0) = \boldsymbol{\gamma}_n'(p)\Lambda_n(\boldsymbol{\beta}_0) + O_p\left(p^{1-\alpha} + \frac{p\log(n)}{n^{1/2}}\right). \tag{15}$$

It should be noted that Assumption 4b and $\alpha > 1$ ensures that $O_p(\cdot)$ in equation (14) and (15) are asymptotically negligible, respectively.

Define Z_k as a zero-mean process that $Z_k = \exp(\eta_k) - 1$. It follows from $\mathcal{L}_n(d, \boldsymbol{\theta})$ and Lemma 3 that

$$\begin{aligned}
\boldsymbol{\gamma}_n'(p)\Lambda_n(\boldsymbol{\beta}_0) &= \frac{1}{\sqrt{pK_n}} \sum_{k=1}^{K_n} \left\{ \exp\left(l_{p,k}^* + \eta_k + r_k\right) - 1 \right\} \boldsymbol{\gamma}_n'(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \\
&= \frac{1}{\sqrt{pK_n}} \sum_{k=1}^{K_n} \exp(l_{p,k}^*) Z_k \boldsymbol{\gamma}_n'(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} + \frac{1}{\sqrt{pK_n}} \sum_{k=1}^{K_n} \left(\exp(l_{p,k}^*) - 1\right) \boldsymbol{\gamma}_n'(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \\
&\quad + \frac{1}{\sqrt{pK_n}} \sum_{k=1}^{K_n} \exp(l_{p,k}^*) \exp(\eta_k) O\left(\frac{\log(1+k)}{k}\right) \boldsymbol{\gamma}_n'(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \\
&= (1 + O(p^{-\alpha})) \frac{1}{\sqrt{pK_n}} \sum_{k=1}^{K_n} Z_k \boldsymbol{\gamma}_n'(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} + O(p^{-\alpha}) \frac{1}{\sqrt{pK_n}} \sum_{k=1}^{K_n} \boldsymbol{\gamma}_n'(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \\
&\quad + (1 + O(p^{-\alpha})) \frac{1}{\sqrt{pK_n}} \sum_{k=1}^{K_n} \exp(\eta_k) O\left(\frac{\log(1+k)}{k}\right) \boldsymbol{\gamma}_n'(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}
\end{aligned}$$

$$=: (1+O(p^{-\alpha}))\Upsilon_1 + \Upsilon_2 + (1+O(p^{-\alpha}))\Upsilon_3,$$

say. The second equality follows from Taylor's expansion of $\exp(l^*_{p,k})$. Now we derive the asymptotic property of these three terms. Using the fact that the vector $(g_k, \mathbf{h}'_{p,k})'$ is deterministic sequence, Lemma 7(i) and Cauchy-Schwarz inequality,

$$\begin{aligned}
\Upsilon_2 &= \sqrt{\frac{K_n}{p}} \cdot O(p^{-\alpha}) \cdot \frac{1}{K_n} \sum_{k=1}^{K_n} \boldsymbol{\gamma}'_n(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \\
&\le O\Big(\frac{\sqrt{n}}{p^{1/2+\alpha}}\Big) \cdot \sqrt{\frac{p}{2} + O\Big(\frac{1}{p^2}\Big)} = O\Big(\frac{\sqrt{n}}{p^{\alpha}}\Big),
\end{aligned}$$

so that Assumption 4b ensures that $\Upsilon_2 = o(1)$ when p tends to infinity. By Lemma 1(ii), 7(ii) and $|h_{j,k}| \le 1$, the expectation of Υ_3 is evaluated as

$$\begin{aligned}
E(\Upsilon_3) &= \frac{1}{\sqrt{pK_n}} \sum_{k=1}^{K_n} O\Big(\frac{\log(1+k)}{k}\Big) \boldsymbol{\gamma}'_n(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \\
&= \frac{1}{\sqrt{pn}} \cdot O\big(p\log(n)\big) \cdot \sum_{k=1}^{K_n} O\Big(\frac{\log(1+k)}{k}\Big) \\
&= O\Big(\frac{p^{1/2}\log^3(n)}{n^{1/2}}\Big),
\end{aligned}$$

and moreover, by Lemma 5(iv),

$$\begin{aligned}
\mathrm{Var}(\Upsilon_3) &= \frac{1}{pK_n} E\left\{ \sum_{k=1}^{K_n} \Big(\exp(\eta_k) - E(\exp(\eta_k))\Big) O\Big(\frac{\log(1+k)}{k}\Big) \boldsymbol{\gamma}'_n(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \right\}^2 \\
&= \frac{1}{pn} \Big\{ O\big(p\log(n)\big) \Big\}^2 E\left\{ \sum_{k=1}^{K_n} \Big(\exp(\eta_k) - E(\exp(\eta_k))\Big) O\Big(\frac{\log(1+k)}{k}\Big) \right\}^2 \\
&= O\Big(\frac{p\log^2(n)}{n}\Big).
\end{aligned}$$

Thus, under Assumption 4b, we have $\Upsilon_3 = o_p(1)$.

Next, to derive the asymptotic property of Υ_1, letting $\xi_{p,k}$ be a normalized process that

$$\xi_{p,k} = \frac{1}{\sqrt{K_n}} \Big(\boldsymbol{\gamma}'_n(p)\Gamma_n(p)\boldsymbol{\gamma}_n(p)\Big)^{-1/2} \left\{ \boldsymbol{\gamma}'_n(p) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \right\},$$

Υ_1 rewrites

$$\Upsilon_1 = \frac{1}{\sqrt{p}} \Big(\boldsymbol{\gamma}'_n(p)\Gamma_n(p)\boldsymbol{\gamma}_n(p)\Big)^{1/2} \sum_{k=1}^{K_n} \xi_{p,k} Z_k.$$

The definition of $\xi_{p,k}$ shows $\sum_{k=1}^{K_n} \xi^2_{p,k} = 1$ and, by Lemma 1(ii) and 7(ii),

$$\max_{1\le k\le K_n} |\xi_{p,k}| = \frac{1}{\sqrt{K_n}} \cdot \frac{1}{\sqrt{p}} \cdot O\big(p\log(n)\big) = O\Big(\frac{p^{1/2}\log(n)}{n^{1/2}}\Big),$$

so that, under Assumption 4b,

$$\lim_{n\to\infty} \max_{1\le k\le K_n} |\xi_{p,k}| = 0.$$

Suppose that there exists a non-decreasing sequence of integers $v_n < K_n$ such that $\lim_{n\to\infty} v_n = \infty$ and

$$\lim_{n\to\infty} \frac{v_n}{n} = 0, \qquad \lim_{n\to\infty} \frac{\sqrt{n}\log(n)}{v_n} = 0.$$

Then, by Lemma 1(ii) and 7(ii),

$$\sum_{k=1}^{v_n} \xi_{p,k}^2 = \frac{1}{K_n} \sum_{k=1}^{v_n} \left\{ O\left(p^2 \log^2(n)\right) \right\}^{-1} \left\{ O\left(p^2 \log^2(n)\right) \right\} = O\left(\frac{v_n}{n}\right),$$

and, by Cauchy-Schwarz inequality,

$$\sum_{k=v_n+1}^{K_n} \frac{|\xi_{p,k}| \log(v_n)}{v_n} \le \frac{\sqrt{n}\log(v_n)}{v_n} \cdot \left(\boldsymbol{\gamma}_n'(p)\Gamma_n(p)\boldsymbol{\gamma}_n(p)\right)^{-1/2} \left(\boldsymbol{\gamma}_n'(p)\Gamma_n(p)\boldsymbol{\gamma}_n(p)\right)^{1/2} = \frac{\sqrt{n}\log(v_n)}{v_n}.$$

It follows from the condition of v_n that

$$\lim_{n\to\infty} \sum_{k=1}^{v_n} \xi_{p,k}^2 = 0, \tag{16}$$

$$\lim_{n\to\infty} \sum_{k=v_n+1}^{K_n} \frac{|\xi_{p,k}| \log(v_n)}{v_n} = 0. \tag{17}$$

(16) and (17) correspond to (4.3) and (4.4) in Soulier (2001), respectively, and Z_k is a function with $\mathrm{Var}(Z_k) = 1$ and *Hermite rank* at least 1 because $E(Z_k) = 0$, which yields

$$\sum_{k=1}^{K_n} \xi_{p,k} Z_k \xrightarrow{d} N(0,\ 1) \quad \text{as } n \to \infty,$$

by applying Theorem 4.1 in Soulier (2001). It is easily seen from Lemma 7(i) that

$$\frac{1}{\sqrt{p}} \left(\boldsymbol{\gamma}_n'(p)\Gamma_n(p)\boldsymbol{\gamma}_n(p)\right)^{1/2} \to \frac{1}{\sqrt{2}} \quad \text{as } n \to \infty.$$

Therefore, we have

$$\Upsilon_1 \xrightarrow{d} N\left(0,\ \tfrac{1}{2}\right) \quad \text{as } n \to \infty. \qquad \square$$

Proof of Corollary. Under the null hypothesis, by Theorem 1, Lemma 6 and Assumption 4b, t_{LM} rewrites

$$\begin{aligned} t_{LM} &= \sqrt{K_n}\left[\Omega_n^{1/2}(\tilde{\boldsymbol{\beta}})\Omega_n^{-1}(\tilde{\boldsymbol{\beta}})S_n(\tilde{\boldsymbol{\beta}})\right]_{11} = \sqrt{K_n}\left[\tilde{\Gamma}_n^{1/2}(p)\tilde{\Gamma}_n^{-1}(p)S_n(\tilde{\boldsymbol{\beta}})\right]_{11} + o_p(1) \\ &= \left[\tilde{\Gamma}_n^{1/2}(p)\left\{\tilde{\Gamma}_n^{-1}(p)\tilde{\Gamma}_n(p)\tilde{\Gamma}_n^{-1}(p)\right\}^{1/2}\right]_{11} \left(\tilde{\boldsymbol{\gamma}}_n{}'(p)\tilde{\Gamma}_n^{-1}(p)\tilde{\boldsymbol{\gamma}}_n(p)\right)^{-1/2} \sqrt{K_n}S_n(\tilde{\boldsymbol{\beta}}) + o_p(1), \end{aligned}$$

where $\tilde{\Gamma}(p)$ and $\tilde{\boldsymbol{\gamma}}(p)$ correspond to $\Gamma(p)$ and $\boldsymbol{\gamma}(p)$, respectively. It is easily seen that

$$\left[\tilde{\Gamma}_n^{1/2}(p)\left\{\tilde{\Gamma}_n^{-1}(p)\tilde{\Gamma}_n(p)\tilde{\Gamma}_n^{-1}(p)\right\}^{1/2}\right]_{11} = 1,$$

and $\sqrt{K_n}S_n(\tilde{\boldsymbol{\beta}})$ corresponds to $\sum_{k=1}^{K_n} \xi_{p,k}Z_k$, so that the proof of Corollary immediately follows from that of Theorem 2. $\square$

6.2 Lemmas

***Lemma* 1.**

(i) Let $g^*_{p,k} = \sum_{j=p+1}^{\infty} 2h_{j,k}/j$, for all $1 \leq k \leq K_n$. Under Assumption 2,

$$\left|g^*_{p,k}\right| \leq \frac{4K_n}{p(2k-1)}.$$

(ii) Under Assumption 2,

$$\max_{1 \leq k \leq K_n} |g_k| = O\left(\log(n)\right).$$

Proof of Lemma 1(i). It is well known that $g(\lambda) = \sum_{j=1}^{\infty} 2h_j(\lambda)/j$ and the proof immediately follows from that of equation (17) in Moulines and Soulier (2000).

Proof of Lemma 1(ii). Setting $p = K_n$ in *Lemma* 1(*i*), for all $1 \leq k \leq K_n$,

$$\left|g^*_{K_n,k}\right| \leq \frac{4}{2k-1}.$$

Then,

$$|g_k| = \left|\sum_{j=1}^{K_n} \frac{2h_{j,k}}{j} + \sum_{j=K_n+1}^{\infty} \frac{2h_{j,k}}{j}\right| \leq \sum_{j=1}^{K_n} \frac{2}{j} + \frac{4}{2k-1},$$

and we have

$$\max_{1 \leq k \leq K_n} |g_k| = \sum_{j=1}^{K_n} \frac{2}{j} + 4 = O\left(\sum_{j=1}^{K_n} \frac{1}{j}\right),$$

which leads to *Lemma* 1(*ii*) by approximation of sums by integrals. □

***Lemma* 2.** Under Assumption 1 and 2, for any sequences of reals φ_k,

$$\sum_{k=1}^{K_n} \sum_{l=k+1}^{K_n} |\varphi_k||\varphi_l||\mathrm{cov}(\eta_k, \eta_l)| = O\Big(\varphi_*^2 \log^3(n)\Big),$$

where $\varphi_* = \max_{1 \leq k \leq K_n} |\varphi_k|$. Note that this lemma is the same as (3.6) in Moulines and Soulier (1999).

Proof of Lemma 2. (5) and approximation of sums by integrals yield

$$\begin{aligned}
\sum_{k=1}^{K_n} \sum_{l=k+1}^{K_n} |\varphi_k||\varphi_l||\mathrm{cov}(\eta_k, \eta_l)| &\leq C \sum_{k=1}^{K_n} \sum_{l=k+1}^{K_n} \varphi_*^2 \log^2(l)\, k^{-2|d|}\, l^{2|d|-2} \\
&\leq \sum_{k=1}^{K_n} \varphi_*^2\, k^{-2|d|}\, O\left(\frac{\log^2(k)}{k^{1-2|d|}}\right) \\
&= O\Big(\varphi_*^2 \log^3(n)\Big),
\end{aligned}$$

because integration by parts yields

$$\begin{aligned}
\sum_{l=k+1}^{K_n} \log^2(l)\, l^{2|d|-2} &\leq \int_{k+1}^{n} \log^2(x)\, x^{2|d|-2} dx \\
&= \left[\frac{1}{2|d|-1} x^{2|d|-1} \log^2(x)\right]_{k+1}^{n} - o\left(\int_{k+1}^{n} \log^2(x)\, x^{2|d|-2} dx\right)
\end{aligned}$$

$$= O\left(\frac{\log^2(k)}{k^{1-2|d|}}\right),$$

with $2|d| - 1 < 0$. □

***Lemma* 3.** Under Assumption 1 and 2, for all $1 \le k \le K_n$,

$$\exp(r_k) = 1 + O\left(\frac{\log(1+k)}{k}\right), \quad w.p.1\,.$$

Proof of Lemma 3. By Lemma 1-3 in Moulines and Soulier (1999), r_k is given by

$$r_k = \log(1+\zeta_1) + \log(1+\zeta_2) + \log(1+\zeta_3),$$

where

$$|\zeta_1| \le \frac{C_1 \log(1+k)}{k}, \qquad |\zeta_2| \le \frac{C_2 \log(1+k)}{k}, \qquad |\zeta_3| \le \frac{C_3 \log(1+k)}{k},$$

$w.p.1$, for all $1 \le k \le K_n$. Taking exponential on both sides, which yields

$$\begin{aligned}\exp(r_k) &= (1+\zeta_1)(1+\zeta_2)(1+\zeta_3)\\ &\le \left(1 + \frac{C_1 \log(1+k)}{k}\right)\left(1 + \frac{C_2 \log(1+k)}{k}\right)\left(1 + \frac{C_3 \log(1+k)}{k}\right)\\ &\le \left(1 + \frac{C^* \log(1+k)}{k}\right)^3 = 1 + O\left(\frac{\log(1+k)}{k}\right) + o\left(\frac{\log(1+k)}{k}\right),\end{aligned}$$

because $\log(1+k)/k < 1$ for $k \in \mathbb{N}$. □

***Lemma* 4.** Under Assumption 1 and 2, there exists a constant $C^* < \infty$, such that for all $1 \le k < l \le K_n$,

$$\left|\text{cov}(\exp(\eta_k), \exp(\eta_l))\right| \le C^* \log^2(l)\, k^{-2|d|}\, l^{2|d|-2}.$$

Proof of Lemma 4. Using the same argument as the proof of Theorem 2 in Moulines and Soulier (1999), we can define η_k as $\eta_k = \log(W_k' W_k) - \log(2)$, where W_k is a standard 2-dimensional Gaussian vector. Then,

$$\exp(\eta_k) - E(\exp(\eta_k)) = \frac{1}{2} W_k' W_k - 1.$$

Define a function $\phi : \mathbb{R}^2 \to \mathbb{R}$ as

$$\phi(X) = \frac{1}{2} X' X - 1, \quad X \in \mathbb{R}^2,$$

and which implies that for all $1 \le k \le K_n$,

$$E(\phi(W_k)) = E(\exp(\eta_k) - E(\exp(\eta_k))) = 0,$$
$$E(\phi^2(W_k)) = \text{Var}(\exp(\eta_k)) = 1.$$

Since $\phi(\cdot)$ is an even function and $E(\phi(W_k)) = 0$, the *Hermite rank* of $\phi(\cdot)$ is 2.

Next, by applying Corollary 2.1 in Soulier (2001) or Theorem 4 in Moulines and Soulier (1999), there exists a constant C^* such that

$$\left|E\left(\prod_{i=1}^{2} \phi(X_i)\right)\right| \le C^* \prod_{i=1}^{2} \left(E(\phi^2(X_i))\right)^{1/2} \rho^{\tau},$$

where $X_i = (X_{i,1}, X_{i,2})'$ is a standard 2-dimensional Gaussian vector and τ is *Hermite rank* of a function $\phi(\cdot)$, and

$$\rho = \max_{1 \le i \ne i' \le 2,\, 1 \le j, j' \le 2} \left|E(X_{i,j} X_{i',j'})\right|.$$

It follows from $E(\phi^2(W_k)) = 1$ and Lemma 4 in Moulines and Soulier (1999) that

$$C^* \prod_{i=1}^{2} \Big(E(\phi^2(X_i))\Big)^{1/2} \rho^2 \leq C^* \Big(\log(l) k^{-|d|} l^{|d|-1}\Big)^2,$$

for all $1 \leq k < l \leq K_n$. □

***Lemma* 5.** Under Assumption 1 and 2, for any non-decreasing sequence of integers $w_n \leq K_n$,

$$\textit{(i)} \quad E\left(\sum_{k=1}^{w_n} (\eta_k - E(\eta_k))\right)^2 = O(w_n).$$

$$\textit{(ii)} \quad E\left(\sum_{k=1}^{w_n} (r_k - E(r_k))\right)^2 = O\left(\log^4(w_n)\right).$$

$$\textit{(iii)} \quad E\left(\sum_{k=1}^{w_n} \Big(\exp(\eta_k) - E(\exp(\eta_k))\Big)\right)^2 = O(w_n).$$

$$\textit{(iv)} \quad E\left(\sum_{k=1}^{w_n} \Big(\exp(\eta_k) - E(\exp(\eta_k))\Big) O\left(\frac{\log(1+k)}{k}\right)\right)^2 = O(1).$$

Proof of Lemma 5(i). By Lemma 2,

$$\begin{aligned} E\left(\sum_{k=1}^{w_n} (\eta_k - E(\eta_k))\right)^2 &= \sum_{k=1}^{w_n} E(\eta_k - E(\eta_k))^2 + 2 \sum_{k=1}^{w_n} \sum_{l=k+1}^{w_n} E(\eta_k - E(\eta_k))(\eta_l - E(\eta_l)) \\ &\leq \sum_{k=1}^{w_n} \mathrm{Var}(\eta_k) + 2 \sum_{k=1}^{w_n} \sum_{l=k+1}^{w_n} \left|\mathrm{cov}(\eta_k, \eta_l)\right| \\ &= O(w_n) + O(\log^3(w_n)). \quad \square \end{aligned}$$

Proof of Lemma 5(ii). It follows from (4),

$$\begin{aligned} E\left(\sum_{k=1}^{w_n} (r_k - E(r_k))\right)^2 &= \sum_{k=1}^{w_n} E(r_k - E(r_k))^2 + 2 \sum_{k=1}^{w_n} \sum_{l=k+1}^{w_n} E(r_k - E(r_k))(r_l - E(r_l)) \\ &\leq C_1 \sum_{k=1}^{w_n} \frac{\log^2(1+k)}{k^2} + C_2 \sum_{k=1}^{w_n} \sum_{l=1}^{w_n} \left(\frac{\log(1+k)}{k}\right)\left(\frac{\log(1+l)}{l}\right) \\ &= O(1) + O\left(\log^2(w_n)\right) \cdot O\left(\log^2(w_n)\right) = O\left(\log^4(w_n)\right), \end{aligned}$$

because $\sum \log^2(1+k)/k^2$ is convergent series. □

Proof of Lemma 5(iii). It follows from Lemma 2 and 4 that

$$\begin{aligned} E\left(\sum_{k=1}^{w_n} \Big(\exp(\eta_k) - E(\exp(\eta_k))\Big)\right)^2 &= \sum_{k=1}^{w_n} \mathrm{Var}(\exp(\eta_k)) + 2 \sum_{k=1}^{w_n} \sum_{l=k+1}^{w_n} \mathrm{cov}(\exp(\eta_k), \exp(\eta_l)) \\ &\leq w_n + C^* \sum_{k=1}^{w_n} \sum_{l=k+1}^{w_n} \log^2(l)\, k^{-2|d|}\, l^{2|d|-2} \\ &= O(w_n) + O(\log^3(w_n)). \quad \square \end{aligned}$$

Proof of Lemma 5(iv). By Lemma 4,

$$
\begin{aligned}
E&\left(\sum_{k=1}^{w_n}\Big(\exp(\eta_k)-E(\exp(\eta_k))\Big)O\Big(\frac{\log(1+k)}{k}\Big)\right)^2\\
&=\sum_{k=1}^{w_n}\mathrm{Var}(\exp(\eta_k))\Big\{O\Big(\frac{\log(1+k)}{k}\Big)\Big\}^2\\
&\quad+2\sum_{k=1}^{w_n}\sum_{l=k+1}^{w_n}\mathrm{cov}\Big(\exp(\eta_k),\exp(\eta_l)\Big)O\Big(\frac{\log(1+k)}{k}\Big)O\Big(\frac{\log(1+l)}{l}\Big)\\
&\le C^*\sum_{k=1}^{w_n}\frac{\log^2(1+k)}{k^2},
\end{aligned}
$$

because, by Cauchy-Schwarz inequality twice,

$$
\begin{aligned}
&\sum_{k=1}^{w_n}\sum_{l=k+1}^{w_n}C^*\log^2(l)\,k^{-2|d|}\,l^{2|d|-2}O\Big(\frac{\log(1+k)}{k}\Big)O\Big(\frac{\log(1+l)}{l}\Big)\\
&\le\sum_{k=1}^{w_n}k^{-2|d|}O\Big(\frac{\log(1+k)}{k}\Big)\left(\sum_{l=1}^{w_n}O\Big(\frac{\log^2(1+l)}{l^2}\Big)\right)^{1/2}\left(\sum_{l=k+1}^{w_n}\log^4(l)l^{4|d|-4}\right)^{1/2}\\
&\le\left(\sum_{k=1}^{w_n}O\Big(\frac{\log^2(1+k)}{k^2}\Big)\right)^{1/2}\left(\sum_{l=1}^{w_n}O\Big(\frac{\log^2(1+l)}{l^2}\Big)\right)^{1/2}\left\{\sum_{k=1}^{w_n}\left(\sum_{l=k+1}^{w_n}\log^4(l)l^{4|d|-4}\right)\right\}^{1/2}\\
&\le\left(\sum_{k=1}^{w_n}O\Big(\frac{\log^2(1+k)}{k^2}\Big)\right)\left(\sum_{l=1}^{w_n}\log^4(l)l^{4|d|-3}\right)^{1/2},
\end{aligned}
$$

where $\sum\log^4(l)\,l^{4|d|-3}$ with $|d|<1/2$ is convergent series. □

***Lemma* 6.** Under Assumption 1-3,

$$
\Omega_n(\boldsymbol{\beta}_0)=\Gamma_n(p)+O_p\left(p^{-\alpha}+\frac{\log(n)}{n^{1/2}}\right),
$$

where

$$
\Omega_n(\boldsymbol{\beta}_0)=-\frac{\partial^2\mathcal{L}_n(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'},\qquad \Gamma_n(p)=\frac{1}{K_n}\sum_{k=1}^{K_n}\begin{pmatrix}g_k\\ \mathbf{h}_{p,k}\end{pmatrix}\begin{pmatrix}g_k\\ \mathbf{h}_{p,k}\end{pmatrix}',
$$

with $\mathbf{h}_p(\lambda)=(h_0(\lambda),\cdots,h_{p-1}(\lambda))'$.

Proof of Lemma 6. By Lemma 3 and Taylor's expansion of $\exp(l^*_{p,k})$,

$$
\begin{aligned}
\Omega_n(\boldsymbol{\beta}_0)&=\frac{1}{K_n}\sum_{k=1}^{K_n}\Big\{\exp(l^*_{p,k}+\eta_k+r_k)\Big\}\begin{pmatrix}g_k\\ \mathbf{h}_{p,k}\end{pmatrix}\begin{pmatrix}g_k\\ \mathbf{h}_{p,k}\end{pmatrix}'\\
&=(1+O(p^{-\alpha}))\frac{1}{K_n}\sum_{k=1}^{K_n}\exp(\eta_k)\begin{pmatrix}g_k\\ \mathbf{h}_{p,k}\end{pmatrix}\begin{pmatrix}g_k\\ \mathbf{h}_{p,k}\end{pmatrix}'\\
&\quad+(1+O(p^{-\alpha}))\frac{1}{K_n}\sum_{k=1}^{K_n}\exp(\eta_k)O\Big(\frac{\log(1+k)}{k}\Big)\begin{pmatrix}g_k\\ \mathbf{h}_{p,k}\end{pmatrix}\begin{pmatrix}g_k\\ \mathbf{h}_{p,k}\end{pmatrix}'.
\end{aligned}
$$

Now using the fact that by Lemma 1(ii),

$$\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}' = O \begin{pmatrix} \log^2(n) & \log(n) & \dots & \log(n) \\ \log(n) & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ \log(n) & 1 & \dots & 1 \end{pmatrix}, \tag{18}$$

the expectation of the second term is given by

$$\frac{1}{K_n} \sum_{k=1}^{K_n} O\left(\frac{\log(1+k)}{k}\right) \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}' = O \begin{pmatrix} \frac{\log^4(n)}{n} & \frac{\log^3(n)}{n} & \dots & \frac{\log^3(n)}{n} \\ \frac{\log^3(n)}{n} & \frac{\log^2(n)}{n} & \dots & \frac{\log^2(n)}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\log^3(n)}{n} & \frac{\log^2(n)}{n} & \dots & \frac{\log^2(n)}{n} \end{pmatrix}.$$

We therefore have

$$E\Big(\Omega_n(\boldsymbol{\beta}_0)\Big) = (1 + O(p^{-\alpha}))\Gamma_n(p) + O\left(\frac{\log^4(n)}{n}\right)\mathbf{1}_{p+1}\mathbf{1}'_{p+1},$$

where $\mathbf{1}_{p+1}$ is a column vector of $(p+1)$ ones.

Next, the variance of $\Omega_n(\boldsymbol{\beta}_0)$ is decomposed into the two terms as follows:

$$\begin{aligned}
\mathrm{Var}\Big(\Omega_n(\boldsymbol{\beta}_0)\Big) &= \Big(1 + O(p^{-\alpha})\Big)^2 \frac{1}{K_n^2} E\left\{\sum_{k=1}^{K_n}\Big(\exp(\eta_k)\exp(r_k) - E(\exp(\eta_k)\exp(r_k))\Big)\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}'\right\}^2 \\
&\le \frac{1}{K_n^2} E\left\{\sum_{k=1}^{K_n} 2\Big(\exp(\eta_k) - E(\exp(\eta_k))\Big)\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}'\right\}^2 \\
&= \frac{2}{K_n^2}\sum_{k=1}^{K_n} \mathrm{Var}\Big(\exp(\eta_k)\Big)\left\{\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}'\right\}^2 \\
&\quad + \frac{4}{K_n^2}\sum_{k=1}^{K_n}\sum_{l=k+1}^{K_n} \mathrm{cov}\Big(\exp(\eta_k), \exp(\eta_l)\Big)\left\{\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}'\right\}\left\{\begin{pmatrix} g_l \\ \mathbf{h}_{p,l} \end{pmatrix}\begin{pmatrix} g_l \\ \mathbf{h}_{p,l} \end{pmatrix}'\right\} \\
&=: 2\Phi_1 + 4\Phi_2,
\end{aligned}$$

say. The inequality follows from Lemma 3 and $\log(1+k)/k < 1$ for $k \in \mathbb{N}$. By (18) and the same argument over the proof of Lemma 7(i) below,

$$\begin{aligned}
\Phi_1 &= \frac{1}{n} \cdot O\begin{pmatrix} \log^2(n) & \log(n) & \dots & \log(n) \\ \log(n) & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ \log(n) & 1 & \dots & 1 \end{pmatrix} \times \frac{1}{K_n}\sum_{k=1}^{K_n}\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}\begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}' \\
&= O\begin{pmatrix} \frac{\log^2(n)}{n} & \frac{\log(n)}{n} & \dots & \frac{\log(n)}{n} \\ \frac{\log(n)}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\log(n)}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}
\end{aligned}$$

$$
\times \begin{pmatrix} O(1) + O\left(\frac{\log^2(n)}{n}\right) & O\left(\frac{\log(n)}{n}\right) & 1 + O\left(\frac{\log(n)}{n}\right) & \dots & \frac{1}{p-1} + O\left(\frac{\log(n)}{n}\right) \\ O\left(\frac{\log(n)}{n}\right) & 1 & 0 & \dots & 0 \\ 1 + O\left(\frac{\log(n)}{n}\right) & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p-1} + O\left(\frac{\log(n)}{n}\right) & 0 & 0 & \dots & 1 \end{pmatrix}
$$

$$
= O\begin{pmatrix} \frac{\log^2(n)}{n} + \frac{\log(n)\log(p)}{n} & \frac{\log^3(n)}{n^2} + \frac{\log(n)}{n} & \frac{\log^2(n)}{n} + \frac{\log(n)}{n} & \dots & \frac{\log^2(n)}{(p-1)n} + \frac{\log(n)}{n} \\ \frac{\log(n)}{n} + \frac{\log(p)}{n} & \frac{\log^2(n)}{n} + \frac{1}{n} & \frac{\log(n)}{n} + \frac{1}{n} & \dots & \frac{\log(n)}{(p-1)n} + \frac{1}{n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\log(n)}{n} + \frac{\log(p)}{n} & \frac{\log^2(n)}{n} + \frac{1}{n} & \frac{\log(n)}{n} + \frac{1}{n} & \dots & \frac{\log(n)}{(p-1)n} + \frac{1}{n} \end{pmatrix}.
$$

By the same way of (18),

$$
\left\{ \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix} \begin{pmatrix} g_k \\ \mathbf{h}_{p,k} \end{pmatrix}' \right\}^2 = O\begin{pmatrix} \log^4(n) + p\log^2(n) & \log^3(n) + p\log(n) & \dots & \log^3(n) + p\log(n) \\ \log^3(n) + p\log(n) & \log^2(n) + p & \dots & \log^2(n) + p \\ \vdots & \vdots & \ddots & \vdots \\ \log^3(n) + p\log(n) & \log^2(n) + p & \dots & \log^2(n) + p \end{pmatrix},
$$

so that, by Lemma 2 and 4,

$$
\Phi_2 = O\left(\frac{\log^3(n)}{n^2}\right) \cdot O\begin{pmatrix} \log^4(n) + p\log^2(n) & \log^3(n) + p\log(n) & \dots & \log^3(n) + p\log(n) \\ \log^3(n) + p\log(n) & \log^2(n) + p & \dots & \log^2(n) + p \\ \vdots & \vdots & \ddots & \vdots \\ \log^3(n) + p\log(n) & \log^2(n) + p & \dots & \log^2(n) + p \end{pmatrix}
$$

$$
= O\begin{pmatrix} \frac{\log^7(n)}{n^2} + \frac{p\log^5(n)}{n^2} & \frac{\log^6(n)}{n^2} + \frac{p\log^4(n)}{n^2} & \dots & \frac{\log^6(n)}{n^2} + \frac{p\log^4(n)}{n^2} \\ \frac{\log^6(n)}{n^2} + \frac{p\log^4(n)}{n^2} & \frac{\log^5(n)}{n^2} + \frac{p\log^3}{n^2} & \dots & \frac{\log^5(n)}{n^2} + \frac{p\log^3}{n^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\log^6(n)}{n^2} + \frac{p\log^4(n)}{n^2} & \frac{\log^5(n)}{n^2} + \frac{p\log^3}{n^2} & \dots & \frac{\log^5(n)}{n^2} + \frac{p\log^3}{n^2} \end{pmatrix}.
$$

Therefore, the variance of $\Omega_n(\boldsymbol{\beta}_0)$ is evaluated as

$$
\mathrm{Var}\left(\Omega_n(\boldsymbol{\beta}_0)\right) = O\left(\frac{\log^2(n)}{n}\right)\mathbf{1}_{p+1}\mathbf{1}'_{p+1}. \quad \square
$$

***Lemma* 7.**

(i) Let $[\Gamma_n^{-1}(p)]_{11}$ denote the $(1,1)$th element in $\Gamma_n^{-1}(p)$. Under Assumption 3, $[\Gamma_n^{-1}(p)]_{11}$ is evaluated as

$$
\left[\Gamma_n^{-1}(p)\right]_{11} = \frac{p}{2} + O\left(\frac{1}{p^2}\right) + O\left(\frac{p\log^2(n)}{n^2}\right) + O\left(\frac{\log^2(n)}{n}\right).
$$

(ii) Let $\boldsymbol{\gamma}_n(p)$ denote the first row of $\Gamma_n^{-1}(p)$. Under Assumption 3, $\boldsymbol{\gamma}_n(p)$ is evaluated as

$$
\boldsymbol{\gamma}_n(p) = [\Gamma_n^{-1}(p)]_{11} \cdot \left(1 \quad O\left(\frac{\log(n)}{n}\right) \quad -1 + O\left(\frac{\log(n)}{n}\right) \quad -\frac{1}{2} + O\left(\frac{\log(n)}{n}\right) \quad \cdots \quad -\frac{1}{p-1} + O\left(\frac{\log(n)}{n}\right)\right)'.
$$

Proof of Lemma 7(i). Since $\{h_0(\lambda), h_1(\lambda), \cdots\}$ is an orthogonal system in $\mathbb{L}^2$ and, by construction, $g_k = g_{2K_n-k+1}$ and $h_{j,k} = h_{j,2K_n-k+1}$, $\Gamma_n(p)$ is given by

$$\Gamma_n(p) = \frac{1}{2}\begin{pmatrix} \frac{1}{K_n}\sum_{k=1}^{2K_n} g_k^2 & \frac{1}{K_n}\sum_{k=1}^{2K_n} g_k\, h_{0,k} & \dots & \frac{1}{K_n}\sum_{k=1}^{2K_n} g_k\, h_{p-1,k} \\ \frac{1}{K_n}\sum_{k=1}^{2K_n} h_{0,k}\, g_k & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{K_n}\sum_{k=1}^{2K_n} h_{p-1,k}\, g_k & 0 & \dots & 1 \end{pmatrix}$$

$$=: \frac{1}{2}\begin{pmatrix} \gamma_{11} & \boldsymbol{\gamma}'_{p1} \\ \boldsymbol{\gamma}_{p1} & I_{p,p} \end{pmatrix},$$

say, where $I_{p,p}$ is the $(p \times p)$ identity matrix. Then inverse of a partitioned matrix yields

$$\left[\Gamma_n^{-1}(p)\right]_{11} = 2\left(\gamma_{11} - \boldsymbol{\gamma}'_{p1}\boldsymbol{\gamma}_{p1}\right)^{-1}.$$

Now by Lemma 1(i) and approximation of sums by integrals, γ_{11} is evaluated as

$$\begin{aligned}
\gamma_{11} &\le \frac{1}{K_n}\sum_{k=1}^{2K_n}\left\{\sum_{i=1}^{2K_n}\frac{2\cos(i\omega_k)}{i} + \frac{4}{2k-1}\right\}^2 \\
&= \frac{1}{K_n}\sum_{k=1}^{2K_n}\left(\sum_{i=1}^{2K_n}\frac{2\cos(i\omega_k)}{i}\right)\left(\sum_{l=1}^{2K_n}\frac{2\cos(i\omega_k)}{l}\right) \\
&\quad + \frac{1}{K_n}\sum_{k=1}^{2K_n}\frac{8}{2k-1}\sum_{i=1}^{2K_n}\frac{2\cos(i\omega_k)}{i} + \frac{1}{K_n}\sum_{k=1}^{2K_n}\left(\frac{4}{2k-1}\right)^2 \\
&\le \frac{1}{K_n}\sum_{i=1}^{2K_n}\frac{4}{i^2}\sum_{k=1}^{2K_n}\cos^2(i\omega_k) + \frac{2}{K_n}\sum_{i=1}^{2K_n}\sum_{l=1}^{2K_n}\frac{4}{il}\sum_{k=1}^{2K_n}\cos(i\omega_k)\cos(l\omega_k) \\
&\quad + \frac{1}{K_n}\left(\sum_{k=1}^{2K_n}\frac{8}{2k-1}\right)\left(\sum_{i=1}^{2K_n}\frac{1}{i}\right) + \frac{C}{K_n}\sum_{k=1}^{2K_n}\frac{1}{k^2} \\
&= 4\sum_{i=1}^{2K_n}\frac{1}{i^2} + O\left(\frac{\log^2(n)}{n}\right) + O\left(\frac{1}{n}\right) = \frac{2}{3}\pi^2 + O\left(\frac{\log^2(n)}{n}\right).
\end{aligned}$$

By the same argument,

$$\begin{aligned}
\boldsymbol{\gamma}'_{p1}\boldsymbol{\gamma}_{p1} &= \sum_{j=0}^{p-1}\left\{\frac{1}{K_n}\sum_{k=1}^{2K_n} g_k h_{j,k}\right\}^2 \\
&\le \left\{\frac{1}{K_n}\sum_{k=1}^{2K_n}\frac{4}{2k-1}\right\}^2 + \sum_{j=1}^{p-1}\left\{\frac{1}{K_n}\sum_{i=1}^{2K_n}\sum_{k=1}^{2K_n}\frac{2\cos(i\omega_k)\cos(j\omega_k)}{i} + \frac{1}{K_n}\sum_{k=1}^{2K_n}\frac{4\cos(j\omega_k)}{2k-1}\right\}^2 \\
&= \left\{O\left(\frac{\log(n)}{n}\right)\right\}^2 + \sum_{j=1}^{p-1}\left\{\frac{2}{j} + O\left(\frac{\log(n)}{n}\right)\right\}^2 = 4\sum_{j=1}^{p-1}\frac{1}{j^2} + O\left(\frac{p\log^2(n)}{n^2}\right) \\
&= \frac{2}{3}\pi^2 - \frac{4}{p} + O\left(\frac{1}{p^2}\right) + O\left(\frac{p\log^2(n)}{n^2}\right).
\end{aligned}$$

Consequently we have

$$\left[\Gamma_n^{-1}(p)\right]_{11} = \frac{p}{2} + O\left(\frac{1}{p^2}\right) + O\left(\frac{p\log^2(n)}{n^2}\right) + O\left(\frac{\log^2(n)}{n}\right). \qquad \square$$

Proof of Lemma 7(ii). By inverse of a partitioned matrix,

$$\boldsymbol{\gamma}_n(p) = \left([\Gamma_n^{-1}(p)]_{11} \quad -[\Gamma_n^{-1}(p)]_{11}\boldsymbol{\gamma}'_{p1}I_{p,p}^{-1}\right)',$$

where $\boldsymbol{\gamma}_{p1}$ and $I_{p,p}$ are the same notation as the proof of Lemma 7(i), respectively. It is easily seen from the proof of Lemma 7(i) that

$$\boldsymbol{\gamma}'_{p1}I_{p,p}^{-1} = \left(O\left(\frac{\log(n)}{n}\right) \quad 1 + O\left(\frac{\log(n)}{n}\right) \quad \frac{1}{2} + O\left(\frac{\log(n)}{n}\right) \quad \cdots \quad \frac{1}{p-1} + O\left(\frac{\log(n)}{n}\right)\right)'. \quad \square$$

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd Inter. Symp. on Information Theory* (B.N. Petrov and F. Csaki, eds.), 267-281. Akademiai Kiado, Budapest.

[2] Beran, J. (1993). Fitting long-memory models by generalized linear regression. *Biometrika* **80**, 817-822.

[3] Beran, J. (1994). *Statistics for Long-Momory Processes*. Chapman and Hall, New York.

[4] Bhansali, R. J., Giraitis, L. and Kokoszka, P. S. (2006). Estimation of the memory parameter by fitting fractionally differenced autoregressive models. *J. Multivariate Anal.* **94**, 2101-2130.

[5] Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60**, 217-226.

[6] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods* (2nd eds.). Springer-Verlag, New York.

[7] Dahlhaus, R. (1989). Efficient paramter estimation for self-similar processes. *Ann. Statist* . **17**, 1749-1766.

[8] Deo, R. S. and Hurvich, C. M. (2003). Estimation of long memory in volatility. In *Theory and Application of Long-Range Dependence* (P. Doukhan, G. Oppenheim and M. S. Taqqu, eds.), 313-324. Birkhäuser, Boston.

[9] Diebold, F. X. and Rudebusch, G.D. (1991). On the power of Dickey-Fuller tests against fractional alternatives. *Economics Letters* **35**, 155-160.

[10] Doukhan, P., Oppenheim, G. and Taqqu, M. S., eds. (2003). *Theory and Application of Long-Range Dependence*. Birkhäuser, Boston.

[11] Fan, J. and Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation. *Scand. J. Statist.* **25**, 359-369.

[12] Fay, G. and Soulier, P. (2001). The periodogram of an i.i.d.sequence. *Stochastic Process. Appl.* **92**, 315-343.

[13] Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14**, 517-532.

[14] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.* **4**, 221-238.

[15] Giraitis, L. and Surgailis, D. (1990). A central limit theorems for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittles's estimate. *Probab. Theory and Related Fields* **86**, 87-104.

[16] Giraitis, L., Robinson, P. M. and Samarov, A. (2000). Adaptive semiparametric estimation of the memory parameter. *J. Multivariate Anal.* **72**, 183-207.

[17] Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Table of Integrals, Series, and Products* (4th eds.). Academic Press, New York.

[18] Henry, M. (2001). Robust automatic bandwidth for long memory. *J. Time Ser. Anal.* **22**, 293-316.

[19] Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models. *Ann. Statist.* **25**, 105-137.

[20] Hurvich, C. M. (2001). Model selection for broadband semiparametric estimation of long memory in time series. *J. Time Ser. Anal.* **22**, 679-709.

[21] Hurvich, C. M. and Beltrao, K. I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. *J. Time Ser. Anal.* **14**, 455-472.

[22] Hurvich, C. M. and Brodsky, J. (2001). Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. *J. Time Ser. Anal.* **22**, 221-249.

[23] Hurvich, C. M. and Deo, R. (1999). Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series. *J. Time Ser. Anal.* **20**, 331-341.

[24] Hurvich, C. M., Deo, R. and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *J. Time Ser. Anal.* **19**, 19-46.

[25] Hurvich, C. M., Moulines, E. and Soulier, P. (2002). The FEXP estimator for potentially non-stationary linear time series. *Stochastic Process. Appl.* **97**, 307-340.

[26] Hurvich, C. M., Moulines, E. and Soulier, P. (2005). Estimating long memory in volatility. *Econometrica* **73**, 1283-1328.

[27] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions*. Imperial College Press, London.

[28] Kolmogorov, A. N. and Fomin, S. V. (1970). *Introductory Real Analysis*. Prentice-Hall, Englewood Cliffs, N.J.

[29] Künsch, H. R. (1986). Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.* **23**, 1025-1030.

[30] Künsch, H. R. (1987). Statistical aspects of self similar processes. In *Proceedings of the First World Congress of the Bernoulli Society* (Yu. Prohorov and V. V. Sazanov, eds.) **1**, 67-74. VNU Science Press, Utrecht.

[31] Lobato, I. N. and Robinson, P. M. (1998). A nonparametric test for I(0). *Rev. Econ. Studies.* **65**, 475-495.

[32] Moulines, E. and Soulier, P. (1999). Broadband log-periodogram regression of time series with long-range dependence. *Ann. Statist.* **27**, 1415-1439.

[33] Moulines, E. and Soulier, P. (2000). Data driven order selection for projection estimation of the spectral density of time series with long range dependence. *J. Time Ser. Anal.* **21**, 193-218.

[34] Robinson, P. M. (1994). Time series with strong dependence. In *Advances in Econometrics: Sixth World Congress* (C. A. Sims, eds.) **1**, 47-95. Cambridge University Press, New York.

[35] Robinson, P. M. (1995a). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23**, 1048-1072.

[36] Robinson, P. M. (1995b). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23**, 1630-1661.

[37] Soulier, P. (2001). Moment bounds and central limit theorem for functions of Gaussian vectors. *Statistics and Probability Letters* **54**, 193-203.

[38] Velasco, C. (2000). Non-Gaussian log-periodogram regression. *Econometric Theory* **16**, 44-79.

[39] Walker, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Aust. Math. Soc.* **4**, 363-384.