ISSN 1066-5307, Mathematical Methods of Statistics, 2015, Vol. 24, No. 4, pp. 309–319. © Allerton Press, Inc., 2015.

Variance Inequalities for Quadratic Forms with Applications

P. Yaskov[1*]

[1] *Steklov Math. Inst. RAS, Moscow, Russia;*
Nat. Univ. of Sci. and Technol. MISIS, Moscow, Russia

Received August 20, 2015

Abstract—We obtain variance inequalities for quadratic forms of weakly dependent random variables with bounded fourth moments. We also discuss two applications. Namely, we use these inequalities for deriving the limiting spectral distribution of a random matrix and estimating the long-run variance of a stationary time series.

Keywords: quadratic forms, moment inequalities, random matrices.

2000 Mathematics Subject Classification: 60F26, 60G10, 60B20.

DOI: 10.3103/S1066530715040055

1. INTRODUCTION

Moment inequalities for quadratic forms constitute a powerful tool in time series analysis and the random matrix theory. In particular, they are used in the study of consistency and optimality properties of spectral density estimates (see, e.g., Section V.4 in Hannan [9]) as well as provide low-level conditions under which the limiting spectral distribution of a random matrix can be derived (see, e.g., Chapter 19 in Pastur and Shcherbina [19]).

When the random variables $\{X_i\}_{i=1}^n$ are independent, the moment inequalities for quadratic forms $\sum_{i,j=1}^n a_{ij}X_iX_j$ are well studied (see, e.g., Lemma B.26 in Bai and Silverstein [2] and Chen [5]). In the time series context, similar inequalities were obtained by many authors in connection with spectral density estimation and long-run variance estimation (see, e.g., Chapter VI in White [21], Sections 6 and 7 in Wu and Xiao [22] and the references therein). In particular, high-order moment inequalities for causal time series were obtained by Wu and Xiao [22].

In this paper we study variance inequalities for quadratic forms $\sum_{i,j=1}^n a_{ij}X_iX_j$ of weakly dependent random variables $\{X_i\}_{i=1}^n$ with bounded fourth moments. Our assumptions deal with covariances of X_i's products up to the fourth order only and are closely related to the classical fourth-order cumulant condition for a stationary time series (see Theorem V.4 in Hannan [9] and Assumption A in Andrews [1]). These assumptions can be easily verified under standard weak dependence conditions (e.g., strong mixing). We also demonstrate how our results can be applied in the random matrix theory and time series analysis.

The paper is structured as follows. The main results are given in Section 2. Section 3 is devoted to applications. Section 4 contains the proofs.

2. MAIN RESULTS

Let $\{X_k\}_{k=1}^\infty$ be a sequence of centered random variables and let $\{\varphi_k\}_{k=1}^\infty$ be a nonincreasing sequence of nonnegative numbers such that, for all $i<j<k<l$,

$$|\operatorname{Cov}(X_i, X_jX_kX_l)| \le \varphi_{j-i}, \qquad |\operatorname{Cov}(X_iX_jX_k, X_l)| \le \varphi_{l-k}, \tag{1}$$

$$|\operatorname{Cov}(X_iX_j, X_kX_l)| \le \varphi_{k-j}, \qquad \text{and} \qquad |\operatorname{Cov}(X_i, X_j)| \le \varphi_{j-i}. \tag{2}$$

*E-mail: yaskov@mi.ras.ru

Assumptions of this kind go back to Révész [20] and were studied by Komlós and Révész [13], Gaposhkin [7], and Longecker and Serfling [14] (see also Móricz [16], Sections 4.3 and 4.4 in Doukhan et al. (2007)).

For simplicity, we will further assume that $\mathbb{E}X_k^2 \le 1$, $k \ge 1$. Define $\mathbf{x}_p = (X_1, \dots, X_p)$ for $p \ge 1$,

$$\Phi_0 = \sup\{\mathbb{E}X_1^4, \mathbb{E}X_2^4, \dots\}, \quad \text{and} \quad \Phi_1 = \sum_{k=1}^{\infty} k\varphi_k.$$

Theorem 2.1. *There is a universal constant $C > 0$ such that, for any $a \in \mathbb{R}^p$ and all $p \times p$ matrices A with zero diagonal,*

$$\mathbb{E}(\mathbf{x}_p^\top a)^4 \le C(\Phi_0 + \Phi_1)(a^\top a)^2 \qquad \textit{and} \qquad \mathrm{Var}(\mathbf{x}_p^\top A\mathbf{x}_p) \le C(\Phi_0 + \Phi_1)\,\mathrm{tr}(AA^\top).$$

A version of the first inequality in Theorem 2.1 is proved by Komlós and Révész [13], Gaposhkin [7], and Longecker and Serfling [14]. The second inequality is new.

Let now ϕ_k, $k \ge 1$, satisfy

$$\mathrm{Cov}(X_i^2, X_j^2) \le \phi_{j-i} \quad \text{for all } i < j. \tag{3}$$

Define

$$\Phi_2 = \sum_{k=1}^{\infty} \phi_k.$$

Theorem 2.2. *There is a universal constant $C > 0$ such that, for all $p \times p$ matrices A,*

$$\mathrm{Var}(\mathbf{x}_p^\top A\mathbf{x}_p) \le C(\Phi_0 + \Phi_1 + \Phi_2)\,\mathrm{tr}(AA^\top).$$

Let us give two examples of $\{X_k\}_{k=1}^{\infty}$ that satisfy (1), (2), and (3).

Example 1. Let $\{X_k\}_{k=1}^{\infty}$ be a martingale difference sequence with bounded 4th moments. Then (1) and (2) hold for $\varphi_k = 0$, $k \ge 1$, and $\Phi_1 = 0$. However, in general, there are no such ϕ_k, $k \ge 1$, that (3) holds and $\Phi_2 < \infty$. This explains why we introduce two sets of coefficients $\{\varphi_k\}_{k=1}^{\infty}$ and $\{\phi_k\}_{k=1}^{\infty}$. If, in addition, $\{X_k^2 - \mathbb{E}X_k^2\}_{k=1}^{\infty}$ is a martingale difference sequence, then, of course, $\phi_k = 0$, $k \ge 1$, and $\Phi_2 = 0$.

Example 2. Let $\{X_k\}_{k=1}^{\infty}$ be strongly mixing random variables with mixing coefficients $(\alpha_k)_{k=1}^{\infty}$, zero mean, and bounded moments of order 4δ for some $\delta > 1$. Then (1)–(3) hold for $\varphi_k = \phi_k = C\alpha_k^{(\delta-1)/\delta}$ and large enough $C > 0$ (see, e.g., Corollary A.2 in Hall and Heyde [8]). One can give similar bounds for other weak dependence conditions.

Remark 2.3. We believe that higher-order moment inequalities for quadratic forms $\mathbf{x}_p^\top A\mathbf{x}_p$ can be derived under similar conditions on $\mathrm{Cov}(X_{i_1} \dots X_{i_k}, X_{i_{k+1}} \dots X_{i_p})$ for $i_1 < \dots < i_p$, $k = 1, \dots, p-1$, and $p > 4$. However the proofs are quite technical even in the case of the second-order inequalities and we leave this question for future research.

Consider the special case when X_k, $k \ge 1$, are centered orthonormal random variables. In this case, (1) and (2) reduce to

$$|\mathbb{E}X_iX_jX_kX_l| \le \min\{\varphi_{j-i}, \varphi_{k-j}, \varphi_{l-k}\}, \quad i < j < k < l.$$

Let $\mathbf{y}_p = (Y_1, \dots, Y_p)$, where each Y_j can be written as $\sum_{k=1}^{\infty} a_kX_k$ in L_2 for some $a_k \in \mathbb{R}$, $k \ge 1$, with $\sum_{k=1}^{\infty} a_k^2 < \infty$.

Corollary 2.4. *Let $\Sigma_p = \mathbb{E}\mathbf{y}_p\mathbf{y}_p^\top$. Then there is $C > 0$ such that, for any $p \times p$ matrix A,*

$$\mathrm{Var}(\mathbf{y}_p^\top A\mathbf{y}_p) \le C(\Phi_0 + \Phi_1 + \Phi_2)\,\mathrm{tr}(\Sigma_pA\Sigma_pA^\top).$$

If $\{X_k\}_{k=1}^{\infty}$ are independent standard normal variables and A is a $p \times p$ symmetric matrix, then $\mathrm{Var}(\mathbf{y}_p^\top A\mathbf{y}_p) = 2\,\mathrm{tr}((\Sigma_pA)^2)$ (see, e.g., Lemma 2.3 in Magnus [15]). Thus, Corollary 2.4 delivers an optimal estimate of the variance up to the factor $C(\Phi_0 + \Phi_1 + \Phi_2)$.

3. APPLICATIONS

In this section we discuss two applications of the obtained inequalities.

Our first application will be in the random matrix theory. For each $p, n \geq 1$, let $\mathbf{Y}_{pn}$ be a $p \times n$ random matrix whose columns are independent copies of $\mathbf{y}_p$, where $\mathbf{y}_p$ is given either in Corollary 2.4, or $\mathbf{y}_p = \mathbf{x}_p$ for $\mathbf{x}_p$ in Theorem 2.2.

Theorem 3.1. *Let $\Phi_0, \Phi_1, \Phi_2 < \infty$. If the following conditions hold*

(1) *$p = p(n)$ is such that $p/n \to c$ for some $c > 0$,*

(2) *the spectral norm of $\Sigma_p = \mathbb{E}\mathbf{y}_p\mathbf{y}_p^\top$ is bounded over p,*

(3) *the empirical spectral distribution of Σ_p's eigenvalues has a weak limit $P(d\lambda)$,*

then, with probability one, the empirical spectral distribution of $n^{-1}\mathbf{Y}_{pn}\mathbf{Y}_{pn}^\top$'s eigenvalues has a weak limit whose Stieltjes transform $m = m(z)$ satisfies

$$m(z) = \int_0^\infty \frac{P(d\lambda)}{\lambda(1 - c - czm(z)) - z}, \qquad z \in \mathbb{C}, \quad \Im(z) > 0.$$

The next application concerns the long-run variance estimation. First, let us recall the generic form of the central limit theorem for a weakly stationary time series $(X_t)_{t=-\infty}^\infty$:

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad n \to \infty,$$

where $\overline{X}_n = n^{-1}\sum_{t=1}^n X_t$, $\mathbb{E}X_t = \mu$, and σ^2 is the long-run variance of $(X_t)_{t=-\infty}^\infty$, i.e.,

$$\sigma^2 = \sum_{j=-\infty}^{\infty} \mathrm{Cov}(X_t, X_{t+j}).$$

This theorem can be proved under various weak dependence assumptions (see, e.g., the books by Doukhan et al. [6] and Bulinski and Shashkin [4]). In statistical applications, this theorem takes the form

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\widehat{\sigma}_n} \xrightarrow{d} \mathcal{N}(0, 1), \quad n \to \infty,$$

where $\widehat{\sigma}_n^2$ is a consistent estimator of σ^2. Recall also that σ^2 can be written as $\sigma^2 = 2\pi f(0)$, where $f = f(x)$, $x \in [-\pi, \pi)$, is the spectral density of $(X_t)_{t=-\infty}^\infty$. Therefore long-run variance estimation is closely related to spectral density estimation.

There are a number of papers that study consistency and optimality properties of long-run variance estimators (see, e.g., Andrews [1], Hansen [10], de Jong and Davidson [12], and Jansson [11] among others). When $\mu = 0$, a typical estimator has the form

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{s,t=1}^n K\left(\frac{|s-t|}{m}\right) X_s X_t, \tag{4}$$

where $K = K(x)$, $x \geq 0$, is a kernel function. Standard assumptions on $K = K(x)$ include

(a) $K(0) = 1$, K is continuous at $x = 0$, and $\sup_{x\geq 0}|K(x)| < \infty$,

(b) $\int_0^\infty \bar{K}^2(x)\,dx < \infty$ for $\bar{K}(x) = \sup_{y \geq x}|K(y)|$,

(c) $k_q = \lim_{x\to 0+} x^{-q}(K(x) - 1)$ exists for some $q > 0$.

Assumptions (a)–(b) are inspired by Assumption 3 of Jansson [11]. However, (b) is weaker than Assumption 3(ii) in [11], where the integrability of $\bar{K}$ is assumed. To our knowledge, the weakest alternative to (b) considered in the literature is the integrability of K^2. However, as discussed in

Jansson [11], many previous results (Andrews [1], Hansen [10], among others) are incorrect as they stated and need stronger alternatives to the integrability of K^2. Assumption (c) is classical and goes back to Parzen [18] (see also Andrews [1]).

Let further $(X_t)_{t=-\infty}^{\infty}$ be a centered weakly stationary time series that satisfies conditions from Section 2 (in particular, $\mathbb{E}X_t^2 \le 1$). Our first result is the consistency of $\widehat{\sigma}_n^2$.

Theorem 3.2. *Let $K = K(x)$ satisfy (a)–(b). If Φ_0, Φ_1, Φ_2 are finite, then $\widehat{\sigma}_n^2 \to \sigma^2$ in the mean square as $m, n \to \infty$ and $m = o(n)$.*

The dependence $m = o(n)$ is optimal. This can be seen by taking a Gaussian white noise $(X_t)_{t=-\infty}^{\infty}$ and showing that $\mathrm{Var}(\widehat{\sigma}^2) \nrightarrow 0$ when $m/n \nrightarrow 0$ due to the variance formula for Gaussian quadratic forms given at the end of Section 2. Andrews [1] following Hannan [9] proved consistency of $\widehat{\sigma}_n^2$ under the cumulant condition

$$\sum_{j,k,l=1}^{\infty} \sup_{t\ge 1} |\kappa(X_t, X_{t+j}, X_{t+k}, X_{t+l})| < \infty. \tag{5}$$

Here $\kappa(X_i, X_j, X_k, X_l)$ is the fourth-order cumulant that is equal to

$$\mathbb{E}X_iX_jX_kX_l - \mathbb{E}X_iX_j\mathbb{E}X_kX_l - \mathbb{E}X_iX_k\mathbb{E}X_jX_l - \mathbb{E}X_iX_l\mathbb{E}X_k\mathbb{E}X_j$$

when each X_t has zero mean. By Lemma 1 of Andrews [1], (5) holds when $(X_t)_{t=-\infty}^{\infty}$ is a strongly mixing sequence with mixing coefficients $(\alpha_k)_{k=1}^{\infty}$ satisfying

$$\sum_{k=1}^{\infty} k^2 \alpha_k^{(\delta-1)/\delta} < \infty$$

and bounded moments of order 4δ for some $\delta > 1$. By Example 2, our Theorem 3.2 is applicable whenever $\sum_{k=1}^{\infty} k\alpha_k^{(\delta-1)/\delta} < \infty$.

The cumulant condition allows us to calculate the limit of the mean squared error (MSE) of $\widehat{\sigma}_n^2$ explicitly. We cannot do it under our assumptions. However, we can give an upper bound for MSE, which is very similar to the exact limit (see Proposition 1 in Andrews [1]).

Theorem 3.3. *Under the conditions of Theorem 3.2, let (c) hold for some $q > 0$ and*

$$\textit{the series } \Gamma_q = \sum_{j=1}^{\infty} j^q \,\mathrm{Cov}(X_t, X_{t+j}) \textit{ converges absolutely.}$$

Then there is an absolute constant $C > 0$ such that, as $m, n \to \infty$,

$$\mathbb{E}|\widehat{\sigma}_n^2 - \sigma^2|^2 \le C(\Phi_0 + \Phi_1 + \Phi_2)\frac{m}{n}\int_0^{\infty} \bar{K}^2(x)\,dx + \frac{4(k_q\Gamma_q)^2}{m^{2q}} + o(m^{-2q}) + O(n^{-1}).$$

4. PROOFS

Below we assume that Φ_0, Φ_1, Φ_2 are finite otherwise all bounds become trivial.

Proof of Theorem 2.1. To prove the first inequality, we reproduce the proof given in Gaposhkin [7] with the only difference that we derive explicit constants in his inequality. Note first that, as $\mathbb{E}X_i = 0$ for all $i \ge 1$, it follows from (1) that

$$|\mathbb{E}X_iX_jX_kX_l| \le \min\{\varphi_{j-i}, \varphi_{l-k}\}, \quad i < j < k < l. \tag{6}$$

Write $a = (a_1, \ldots, a_p)$. Using Lemma 1 in Moricz [16] with $p = 2$ and $r = 4$, we get

$$|(\mathbf{x}_p^\top a)^4 - 24T| \le C_0(S^4 + |\mathbf{x}_p^\top a|^3 S),$$

where $C_0 > 0$ is a universal constant,

$$T = \sum_{i<j<k<l} a_i a_j a_k a_l X_i X_j X_k X_l, \quad S = \Big(\sum_{i=1}^{p} a_i^2 X_i^2\Big)^{1/2},$$

here and in what follows, i, j, k, l are any numbers in $\{1, \dots, p\}$. By Hölder's inequality,

$$\mathbb{E}(\mathbf{x}_p^\top a)^4 \le 24\mathbb{E}T + C_0\big(\mathbb{E}S^4 + (\mathbb{E}S^4)^{1/4}(\mathbb{E}|\mathbf{x}_p^\top a|^4)^{3/4}\big).$$

By (6),

$$|\mathbb{E}T| \le \sum_{i<j<k<l} |a_i a_j a_k a_l| \min\{\varphi_{j-i}, \varphi_{l-k}\} \le \frac{1}{4} \sum_{i<j<k<l} (a_i^2 + a_j^2)(a_k^2 + a_l^2) \min\{\varphi_{j-i}, \varphi_{l-k}\}.$$

We estimate only the term

$$J = \sum_{i<j<k<l} a_i^2 a_k^2 \min\{\varphi_{j-i}, \varphi_{l-k}\}.$$

Other terms with $a_j^2 a_k^2$, $a_i^2 a_l^2$, and $a_j^2 a_l^2$ instead of $a_i^2 a_k^2$ can be estimated similarly. We have

$$J \le \sum_{i<k} a_i^2 a_k^2 \sum_{q,r=1}^{\infty} \min\{\varphi_q, \varphi_r\}$$

and

$$\sum_{q,r=1}^{\infty} \min\{\varphi_q, \varphi_r\} \le \sum_{q=1}^{\infty} \Big(q\varphi_q + \sum_{r=q+1}^{\infty} \varphi_r\Big) = \Phi_1 + \sum_{r=2}^{\infty}\sum_{q=1}^{r-1} \varphi_r \le 2\Phi_1. \tag{7}$$

As a result,

$$J \le 2\Phi_1 \sum_{i<k} a_i^2 a_k^2 \le \Phi_1 \Big(\sum_{i=1}^{p} a_i^2\Big)^2 \qquad \text{and} \qquad |\mathbb{E}T| \le \Phi_1 \|a\|^4.$$

Let us also note that

$$\mathbb{E}S^4 = \sum_{i,j=1}^{p} a_i^2 a_j^2 \mathbb{E}X_i^2 X_j^2 \le \Phi_0 \|a\|^4.$$

Combining the above inequalities, we infer that

$$\mathbb{E}(\mathbf{x}_p^\top a)^4 \le (24 + C_0)(\Phi_0 + \Phi_1)\|a\|^4 + C_0[(\Phi_0 + \Phi_1)\|a\|^4]^{1/4}[\mathbb{E}(\mathbf{x}_p^\top a)^4]^{3/4}.$$

Put $R = [\mathbb{E}(\mathbf{x}_p^\top a)^4/(\Phi_0 + \Phi_1)]^{1/4}/\|a\|$. Then $R^4 \le 24 + C_0 + C_0 R^3$. Therefore, $R \le R_0$, where $R_0 > 0$ is the largest positive root of the equation $x^4 = 24 + C_0 + C_0 x^3$. Finally, we obtain

$$\mathbb{E}|\mathbf{x}_p^\top a|^4 \le R_0^4(\Phi_0 + \Phi_1)\|a\|^4.$$

We now verify the second inequality. First, note that, for $i < j$, $|\operatorname{Cov}(X_i, X_j)| = |\mathbb{E}X_i X_j| \le \sqrt{\mathbb{E}X_i^2 \mathbb{E}X_j^2} \le 1$. In addition, for $i < j < k < l$,

$$|\operatorname{Cov}(X_i X_j, X_k X_l)| \le 2\min\{\varphi_{j-i}, \varphi_{k-j}, \varphi_{l-k}\} \quad \text{and} \quad I \le 2\min\{\varphi_{j-i}, \varphi_{l-k}\}, \tag{8}$$

where I is equal to $|\operatorname{Cov}(X_i X_k, X_j X_l)|$ or $|\operatorname{Cov}(X_i X_l, X_j X_k)|$. Indeed, by (2) and (6),

$$\begin{aligned}|\operatorname{Cov}(X_i X_j, X_k X_l)| &\le \min\big\{\varphi_{k-j}, |\mathbb{E}X_i X_j X_k X_l| + |\mathbb{E}X_i X_j \mathbb{E}X_k X_l|\big\} \\ &\le \min\big\{\varphi_{k-j}, 2\min\{\varphi_{j-i}, \varphi_{l-k}\}\big\} \\ &\le 2\min\{\varphi_{j-i}, \varphi_{k-j}, \varphi_{l-k}\},\end{aligned}$$

and, by the monotonicity of φ_k,

$$\begin{aligned}|\operatorname{Cov}(X_iX_k, X_jX_l)| &\le |\mathbb{E}X_iX_k\mathbb{E}X_jX_l| + |\mathbb{E}X_iX_jX_kX_l|\\ &\le \min\{\varphi_{k-i}, \varphi_{l-j}\} + \min\{\varphi_{j-i}, \varphi_{l-k}\}\\ &\le 2\min\{\varphi_{j-i}, \varphi_{l-k}\}.\end{aligned}$$

A similar bound holds for $\operatorname{Cov}(X_iX_l, X_jX_k)$.

Let $A = (a_{ij})_{i,j=1}^p$ and $a_{ii} = 0$, $1 \le i \le p$. Set $B = (A^\top + A)/2$. Then $\mathbf{x}_p^\top A\mathbf{x}_p = \mathbf{x}_p^\top B\mathbf{x}_p$ and

$$\operatorname{tr}(BB^\top) = \sum_{i,j=1}^p \Big(\frac{a_{ij}+a_{ji}}{2}\Big)^2 \le \sum_{i,j=1}^p \frac{a_{ij}^2 + a_{ji}^2}{2} = \sum_{i,j=1}^p a_{ij}^2 = \operatorname{tr}(AA^\top). \tag{9}$$

Hence we may assume w.l.o.g. that $A = A^\top$. Then

$$\begin{aligned}\operatorname{Var}(\mathbf{x}_p^\top A\mathbf{x}_p) &= 4\operatorname{Var}\Big(\sum_{i=1}^{p-1} X_i \sum_{k=i+1}^p a_{ik}X_k\Big)\\ &= 4\sum_{i=1}^{p-1}\operatorname{Var}\Big(X_i\sum_{k=i+1}^p a_{ik}X_k\Big) + 8\sum_{i<j}\operatorname{Cov}\Big(X_i\sum_{k=i+1}^p a_{ik}X_k, X_j\sum_{k=j+1}^p a_{jk}X_k\Big)\\ &= 4I_1 + 8I_2 + 8I_3 + 8I_4,\end{aligned}$$

where

$$\begin{aligned}I_1 &= \sum_{i=1}^{p-1}\operatorname{Var}\Big(X_i\sum_{k=i+1}^p a_{ik}X_k\Big),\\ I_2 &= \sum_{i<j}\operatorname{Cov}\Big(X_i\sum_{k=i+1}^{j-1} a_{ik}X_k, X_j\sum_{k=j+1}^p a_{jk}X_k\Big),\\ I_3 &= \sum_{i<j} a_{ij}\operatorname{Cov}\Big(X_iX_j, X_j\sum_{k=j+1}^p a_{jk}X_k\Big),\\ I_4 &= \sum_{i<j}\operatorname{Cov}\Big(X_i\sum_{k=j+1}^p a_{ik}X_k, X_j\sum_{k=j+1}^p a_{jk}X_k\Big),\end{aligned}$$

and the sums over the empty set are zeros.

Control of I_1. By the Cauchy–Schwarz inequality and the first inequality in Theorem 2.1,

$$\begin{aligned}I_1 &\le \sum_{i=1}^{p-1}\sqrt{\mathbb{E}X_i^4}\Big(\mathbb{E}\Big|\sum_{k=i+1}^p a_{ik}X_k\Big|^4\Big)^{1/2}\\ &\le C(\Phi_0+\Phi_1)\sum_{i=1}^{p-1}\sum_{k=i+1}^p a_{ik}^2 = C(\Phi_0+\Phi_1)\frac{\operatorname{tr}(A^2)}{2}.\end{aligned}$$

Control of I_2. By the Cauchy–Schwarz inequality and (8),

$$\begin{aligned}I_2 &\le \sum_{i<k<j<l} |a_{ik}a_{jl}|\,|\operatorname{Cov}(X_iX_k, X_jX_l)|\\ &\le 2\sum_{i<k<j<l} |a_{ik}a_{jl}|\min\{\varphi_{k-i}, \varphi_{j-k}, \varphi_{l-j}\} \le I_5 + I_6,\end{aligned}$$

where

$$I_5 = \sum_{i<k<j<l} a_{ik}^2 \min\{\varphi_{j-k}, \varphi_{l-j}\}, \qquad I_6 = \sum_{i<k<j<l} a_{jl}^2 \min\{\varphi_{k-i}, \varphi_{j-k}\}.$$

Additionally, by (7),

$$I_5 \le \sum_{i<k} a_{ik}^2 \sum_{q,r=1}^{\infty} \min\{\varphi_q, \varphi_r\} \le \frac{\operatorname{tr}(A^2)}{2}\,(2\Phi_1) = \operatorname{tr}(A^2)\Phi_1.$$

We similarly derive that $I_6 \le \operatorname{tr}(A^2)\Phi_1$. Hence $I_2 \le 2\operatorname{tr}(A^2)\Phi_1$.

Control of I_3. By the Cauchy–Schwarz inequality and the first inequality in Theorem 2.1,

$$\begin{aligned} I_3 &= \sum_{j=2}^{p-1} \operatorname{Cov}\Big(X_j \sum_{i=1}^{j-1} a_{ij}X_i, X_j \sum_{k=j+1}^{p} a_{jk}X_k\Big) \\ &\le \sum_{j=2}^{p-1} \Big(\mathbb{E}X_j^2\Big|\sum_{i=1}^{j-1} a_{ij}X_i\Big|^2\Big)^{1/2}\Big(\mathbb{E}X_j^2\Big|\sum_{k=j+1}^{p} a_{jk}X_k\Big|^2\Big)^{1/2} \\ &\le \sum_{j=2}^{p-1} \sqrt{\mathbb{E}X_j^4}\Big[\mathbb{E}\Big(\sum_{i=1}^{j-1} a_{ij}X_i\Big)^4 \mathbb{E}\Big(\sum_{k=j+1}^{p} a_{jk}X_k\Big)^4\Big]^{1/4} \\ &\le \sqrt{C(\Phi_0+\Phi_1)}(I_7+I_8)/2, \end{aligned}$$

where

$$I_7 = \sum_{j=2}^{p-1}\Big[\mathbb{E}\Big(\sum_{i=1}^{j-1} a_{ij}X_i\Big)^4\Big]^{1/2}, \qquad I_8 = \sum_{j=2}^{p-1}\Big[\mathbb{E}\Big(\sum_{k=j+1}^{p} a_{jk}X_k\Big)^4\Big]^{1/2}.$$

By the first inequality in Theorem 2.1,

$$I_7 \le K\sum_{j=2}^{p-1}\sum_{i=1}^{j-1} a_{ij}^2 \le \frac{K\operatorname{tr}(A^2)}{2}, \qquad I_8 \le K\sum_{j=2}^{p-1}\sum_{k=j+1}^{p} a_{jk}^2 \le \frac{K\operatorname{tr}(A^2)}{2},$$

where $K = \sqrt{C(\Phi_0+\Phi_1)}$. As a result, $I_3 \le C(\Phi_0+\Phi_1)\operatorname{tr}(A^2)/2$.

Control of I_4. We have $I_4 = I_9 + I_{10} + I_{11}$, where

$$I_9 = \sum_{i<j<k} \operatorname{Cov}(a_{ik}X_iX_k, a_{jk}X_jX_k), \qquad I_{10} = \sum_{i<j<k<l} a_{ik}a_{jl}\operatorname{Cov}(X_iX_k, X_jX_l),$$

$$I_{11} = \sum_{i<j<k<l} a_{il}a_{jk}\operatorname{Cov}(X_iX_l, X_jX_k).$$

By the first inequality in Theorem 2.1,

$$\begin{aligned} I_9 &= \frac{1}{2}\sum_{k=3}^{p}\operatorname{Var}\Big(X_k\sum_{i=1}^{k-1} a_{ik}X_i\Big) - \frac{1}{2}\sum_{k=3}^{p}\sum_{i=1}^{k-1}\operatorname{Var}(a_{ik}X_iX_k) \\ &\le \frac{1}{2}\sum_{k=3}^{p}\Big[\mathbb{E}X_k^4\mathbb{E}\Big(\sum_{i=1}^{k-1} a_{ik}X_i\Big)^4\Big]^{1/2} \le C(\Phi_0+\Phi_1)\sum_{k=3}^{p}\sum_{i=1}^{k-1}\frac{a_{ik}^2}{2} \\ &\le C(\Phi_0+\Phi_1)\,\frac{\operatorname{tr}(A^2)}{4}. \end{aligned}$$

Let us now estimate I_{10} and I_{11}. By (8),

$$I_{10} \le 2 \sum_{i<j<k<l} |a_{ik}a_{jl}| \min\{\varphi_{j-i}, \varphi_{l-k}\} \qquad \text{and} \qquad I_{11} \le 2 \sum_{i<j<k<l} |a_{il}a_{jk}| \min\{\varphi_{j-i}, \varphi_{l-k}\}.$$

By the Cauchy–Schwarz inequality, $I_{10} \le I_{12} + I_{13}$ and $I_{11} \le I_{14} + I_{15}$ with

$$I_{12} = \sum_{i<j<k<l} a_{ik}^2 \min\{\varphi_{j-i}, \varphi_{l-k}\}, \qquad I_{13} = \sum_{i<j<k<l} a_{jl}^2 \min\{\varphi_{j-i}, \varphi_{l-k}\},$$

$$I_{14} = \sum_{i<j<k<l} a_{il}^2 \min\{\varphi_{j-i}, \varphi_{l-k}\}, \qquad I_{15} = \sum_{i<j<k<l} a_{jk}^2 \min\{\varphi_{j-i}, \varphi_{l-k}\}.$$

As before, we have

$$I_{12} \le \sum_{i<k} a_{ik}^2 \sum_{q,r=1}^{\infty} \min\{\varphi_q, \varphi_r\} \le \operatorname{tr}(A^2)\Phi_1.$$

By the same arguments, I_{13}, I_{14}, and I_{15} can be bounded from above by $\operatorname{tr}(A^2)\Phi_1$. Thus, we conclude that $I_{10} + I_{11} \le 4\operatorname{tr}(A^2)\Phi_1$.

Combining the above inequalities, we get $\operatorname{Var}(\mathbf{x}_p^\top A\mathbf{x}_p) \le C(\Phi_0 + \Phi_1)\operatorname{tr}(A^2)$ for a universal constant $C > 0$. □

Proof of Theorem 2.2. Let $A = (a_{ij})_{i,j=1}^p$ and D be the $p \times p$ diagonal matrix with diagonal entries $a_{11}, \dots, a_{pp}$. By Theorem 2.1,

$$\operatorname{Var}(\mathbf{x}_p^\top (A - D)\mathbf{x}_p) \le C(\Phi_0 + \Phi_1)\operatorname{tr}((A - D)(A - D)^\top).$$

In addition,

$$\operatorname{Var}(\mathbf{x}_p^\top A\mathbf{x}_p) \le 2\operatorname{Var}(\mathbf{x}_p^\top D\mathbf{x}_p) + 2\operatorname{Var}(\mathbf{x}_p^\top (A - D)\mathbf{x}_p).$$

Noting that

$$\operatorname{tr}(AA^\top) = \operatorname{tr}\left((A - D)(A - D)^\top\right) + \operatorname{tr}(D^2),$$

we only need to bound $\operatorname{Var}(\mathbf{x}_p^\top D\mathbf{x}_p)$ from above by $\operatorname{tr}(D^2)$ up to a constant factor. Write $D = D_1 - D_2$, where D_i are diagonal matrices with nonnegative diagonal entries and $\operatorname{tr}(D^2) = \operatorname{tr}(D_1^2) + \operatorname{tr}(D_2^2)$. By the Cauchy–Schwarz inequality,

$$\operatorname{Var}(\mathbf{x}_p^\top D\mathbf{x}_p) \le 2\sum_{i=1}^{2} \operatorname{Var}(\mathbf{x}_p^\top D_i\mathbf{x}_p).$$

Hence we may assume w.l.o.g. that diagonal elements of D are nonnegative.

We see that

$$\begin{aligned}
\operatorname{Var}(\mathbf{x}_p^\top D\mathbf{x}_p) &= \operatorname{Var}\Big(\sum_{i=1}^{p} a_{ii}X_i^2\Big) \\
&= \sum_{i=1}^{p} a_{ii}^2 \operatorname{Var}(X_i^2) + \sum_{i \ne j} a_{ii}a_{jj} \operatorname{Cov}(X_i^2, X_j^2) \\
&\le \Phi_0 \sum_{i=1}^{n} a_{ii}^2 + \sum_{i \ne j} a_{ii}a_{jj}\phi_{|i-j|}
\end{aligned}$$

and, as a result,

$$\operatorname{Var}(\mathbf{x}_p^\top D\mathbf{x}_p) \le \Phi_0 \operatorname{tr}(D^2) + \sum_{i \ne j} \frac{a_{ii}^2 + a_{jj}^2}{2}\phi_{|i-j|}$$

$$\begin{aligned}&\le \Phi_0 \operatorname{tr}(D^2) + \sum_{i=1}^{p} a_{ii}^2 \sum_{j:j\neq i} \phi_{|i-j|}\\ &\le 2\operatorname{tr}(D^2)\Big(\Phi_0 + \sum_{k=1}^{\infty}\phi_k\Big)\\ &= 2(\Phi_0+\Phi_2)\operatorname{tr}(D^2).\end{aligned}$$

Combining the above bounds, we get the desired inequality. □

Proof of Corollary 2.4. By the definition of $\mathbf{y}_p$, $\Gamma_n\mathbf{x}_n \to \mathbf{y}_p$ in probability and in the mean square as $n\to\infty$ for some $p\times n$ matrices Γ_n and $\mathbf{x}_n = (X_1,\dots,X_n)$. Since $X_k, k\ge 1$, are orthonormal, we have

$$\Gamma_n\Gamma_n^\top = \mathbb{E}(\Gamma_n\mathbf{x}_n)(\Gamma_n\mathbf{x}_n)^\top \to \mathbb{E}\mathbf{y}_p\mathbf{y}_p^\top = \Sigma_p,$$

$$\mathbf{x}_n^\top(\Gamma_n^\top A\Gamma_n)\mathbf{x}_n = (\Gamma_n\mathbf{x}_n)^\top A(\Gamma_n\mathbf{x}_n) \to \mathbf{y}_p^\top A\mathbf{y}_p \quad \text{in probability,}$$

and

$$\begin{aligned}\mathbb{E}\mathbf{x}_n^\top(\Gamma_n^\top A\Gamma_n)\mathbf{x}_n &= \operatorname{tr}(\Gamma_n^\top A\Gamma_n)\\ &= \operatorname{tr}(\Gamma_n\Gamma_n^\top A) \to \operatorname{tr}(\Sigma_p A) = \mathbb{E}\mathbf{y}_p^\top A\mathbf{y}_p\end{aligned}$$

as $n\to\infty$. We need the following version of Fatou's lemma:

$$\text{if } \xi_n \to \xi \text{ in probability, then } \mathbb{E}|\xi| \le \liminf_{n\to\infty}\mathbb{E}|\xi_n|.$$

By this lemma and Theorem 2.2,

$$\begin{aligned}\mathbb{E}|\mathbf{y}_p^\top A\mathbf{y}_p - \operatorname{tr}(\Sigma_p A)|^2 &\le \liminf_{n\to\infty}\mathbb{E}|\mathbf{x}_n^\top(\Gamma_n^\top A\Gamma_n)\mathbf{x}_n - \operatorname{tr}(\Gamma_n^\top A\Gamma_n)|^2\\ &\le \liminf_{n\to\infty} C(\Phi_0+\Phi_1+\Phi_2)\operatorname{tr}(\Gamma_n^\top A\Gamma_n\Gamma_n^\top A^\top\Gamma_n).\end{aligned}$$

Note that

$$\operatorname{tr}(\Gamma_n^\top A\Gamma_n\Gamma_n^\top A^\top\Gamma_n) = \operatorname{tr}(\Gamma_n\Gamma_n^\top A\Gamma_n\Gamma_n^\top A^\top) \to \operatorname{tr}(\Sigma_p A\Sigma_p A^\top).$$

□

Proof of Theorem 3.1. Denote the spectral norm of a matrix A by $\|A\|$. Recall that $\|A\| = \sqrt{\|AA^\top\|} = \sqrt{\|A^\top A\|}$. In addition, let $A^{1/2}$ be the principal square root of a square positive semi-definite matrix A. By Theorem 1.1 in Bai and Zhou [3], we will prove the theorem by checking that $\operatorname{Var}(\mathbf{y}_p^\top A_p\mathbf{y}_p) = o(p^2)$ as $p\to\infty$ for any sequence $(A_p)_{p=1}^\infty$ with $\|A_p\| = O(1)$, where A_p is a $p\times p$ matrix.

First, let $\mathbf{y}_p$ be as in Corollary 2.4. Then

$$\operatorname{tr}(\Sigma_p A_p\Sigma_p A_p^\top) = \operatorname{tr}(\Sigma_p^{1/2}A_p\Sigma_p A_p^\top\Sigma_p^{1/2}) = \operatorname{tr}(Q\Sigma_p Q^\top)$$

with $Q = \Sigma_p^{1/2}A_p$. If I_p is the $p\times p$ identity matrix, then $\|\Sigma_p\|I_p - \Sigma_p$ is positive semi-definite and, as a result, $Q(\|\Sigma_p\|I_p - \Sigma_p)Q^\top$ is positive semi-definite for any Q. Hence

$$\begin{aligned}\operatorname{tr}(Q^\top\Sigma_p Q) &\le \|\Sigma_p\|\operatorname{tr}(QQ^\top) = \|\Sigma_p\|\operatorname{tr}(Q^\top Q)\\ &= \|\Sigma_p\|\operatorname{tr}(A_p^\top\Sigma_p A_p) \le \|\Sigma_p\|^2\operatorname{tr}(A_p^\top A_p)\end{aligned}$$

and $\operatorname{tr}(A_p^\top A_p) \le \|A_p\|^2 p$. Therefore, by Corollary 2.4,

$$\operatorname{Var}(\mathbf{y}_p^\top A_p\mathbf{y}_p) \le C(\Phi_0+\Phi_1+\Phi_2)\|A_p\|^2\|\Sigma_p\|^2 p = o(p^2)$$

whenever $\|A_p\| = O(1)$. The case $\mathbf{y}_p = \mathbf{x}_p$ with $\mathbf{x}_p$ as in Theorem 2.2 can be considered along the same lines due to the inequality $\operatorname{tr}(A_pA_p^\top) \le \|A_p\|^2 p$. □

Proof of Theorem 3.2. Since $\Phi_1 < \infty$, we have

$$\sum_{j=1}^{\infty} j|C(j)| < \infty \tag{10}$$

and σ^2 is well defined, where $C(j) = \operatorname{Cov}(X_t, X_{t+j})$, $j \in \mathbb{Z}$. We also have the bias-variance decomposition

$$\mathbb{E}(\widehat{\sigma}^2 - \sigma^2)^2 = \operatorname{Var}(\widehat{\sigma}^2) + (\mathbb{E}\widehat{\sigma}^2 - \sigma^2)^2.$$

First, let us estimate the bias term $\mathbb{E}\widehat{\sigma}^2 - \sigma^2$. Using $K(0) = 1$, and $C(j) = C(-j)$, $j \in \mathbb{Z}$, we get

$$\begin{aligned}\mathbb{E}\widehat{\sigma}^2 - \sigma^2 &= \sum_{j=-n}^{n} \Big(1 - \frac{|j|}{n}\Big) K\Big(\frac{|j|}{m}\Big) C(j) - \sum_{j=-\infty}^{\infty} C(j) \\ &= 2\sum_{j=1}^{n} \Big(K\Big(\frac{j}{m}\Big) - 1\Big) C(j) - \frac{2}{n}\sum_{j=1}^{n} K\Big(\frac{j}{m}\Big) jC(j) - 2\sum_{j>n} C(j).\end{aligned}$$

Now, setting $M = \sup_{x\geq 0} |K(x)|$,

$$\Big|\sum_{j=1}^{n} K\Big(\frac{j}{m}\Big) jC(j)\Big| \leq M \sum_{j=1}^{\infty} j|C(j)| = O(1).$$

Additionally,

$$\Big|\sum_{j>n} C(j)\Big| \leq \frac{1}{n}\sum_{j>n} j|C(j)| = o(1/n).$$

Combining these relations yields

$$\mathbb{E}\widehat{\sigma}^2 - \sigma^2 = 2\sum_{j=1}^{n} \Big(K\Big(\frac{j}{m}\Big) - 1\Big) C(j) + O(1/n) = o(1), \tag{11}$$

where the last equality follows from (a) and (10).

Now, consider the variance term $\operatorname{Var}(\widehat{\sigma}^2)$. By Theorem 2.2,

$$\operatorname{Var}(\widehat{\sigma}^2) \leq \frac{C_0}{n^2} \sum_{s,t=1}^{n} K^2\Big(\frac{|s-t|}{m}\Big),$$

where $C_0 = C(\Phi_0 + \Phi_1 + \Phi_2)$ with C given in Theorem 2.2. Using that $\bar{K}(x) = \sup_{y\geq x} |K(y)|$ is a nondecreasing function in $L_2(\mathbb{R})$, we derive

$$\begin{aligned}\frac{1}{mn}\sum_{s,t=1}^{n} K^2\Big(\frac{|s-t|}{m}\Big) &\leq \frac{1}{m} + \frac{2}{m}\sum_{j=1}^{n} \bar{K}^2\Big(\frac{j}{m}\Big) \\ &\leq \frac{1}{m} + 2\int_0^{\infty} \bar{K}^2(x)\,dx.\end{aligned}$$

As a result,

$$\operatorname{Var}(\widehat{\sigma}^2) \leq \frac{C_0}{n} + \frac{2C_0 m}{n}\int_0^{\infty} \bar{K}^2(x)\,dx \tag{12}$$

and $\operatorname{Var}(\widehat{\sigma}^2) = o(1)$ whenever $m, n \to \infty$ and $m/n \to 0$.

Combining the above bounds for the bias and variance, we finish the proof. ☐

Proof of Theorem 3.3. The proof follows the same lines as the proof of Theorem 3.2. We only need to note the following. If (c) holds for some $q > 0$, then, by (c) and the boundedness of K, $x^{-q}(K(x)-1)$ is bounded on $\mathbb{R}_+$. Therefore, by (a), (c), and the absolute convergence of $\sum_{j\geq 1} j^q C(j)$,

$$m^q \sum_{j=1}^{n} \Big(K\Big(\frac{j}{m}\Big) - 1\Big) C(j) = \sum_{j=1}^{n} \frac{K(j/m)-1}{(j/m)^q}\, j^q C(j)$$
$$= k_q \sum_{j=1}^{\infty} j^q C(j) + o(1).$$

By (11) and (12), this relation yields the desired bound. □

ACKNOWLEDGMENTS

The work is supported by the Russian Science Foundation under grant 14-21-00162.

REFERENCES

1. D. Andrews, Heteroskedastic and autocorrelation consistent covariance matrix estimation. Econometrica **59**, (1991), 817–858.
2. Z. Bai and J. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. (Springer, New York, 2010).
3. Z. Bai and W. Zhou, "Large Sample Covariance Matrices without Independence Structures in Columns", Statist. Sinica **18**, (2008), 425–442.
4. A. Bulinski and A. Shashkin, *Limit Theorems for Associated Random Fields and Related Systems* (World Scientific, 2007).
5. X. Chen, "A Note on Moment Inequality for Quadratic Forms", Statist. and Probab. Lett. **92**, 83–88 (2014).
6. J. Dedecker, P. Doukhan, G. Lang, R. José Rafael León, S. Louhichi, and C. Prieur, *Weak Dependence: With Examples and Applications* (Springer, New York, 2007).
7. V. F. Gaposhkin, "On the Convergence of Series of Weakly Multiplicative Systems of Functions", Math. USSR-Sb. **18**, 361–371 (1972).
8. P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application* (Academic Press, Boston, 1980).
9. E. J. Hannan, *Multiple Time Series* (Wiley, New York, 1970).
10. B. E. Hansen, "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes", Econometrica **60**, 967–972 (1992).
11. M. Jansson, "Consistent Covariance Matrix Estimation for Linear Processes", Econometric Theory **18**, 1449–1459 (2002).
12. R. M. Jong and J. Davidson, "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices", Econometrica **68**, (2000), 407–423.
13. J. Komlós and P. Révész, "Remark to a Paper by Gaposhkin", Acta Sci. Math. **33**, 237–241 (1972).
14. M. Longecker and R. J. Serfling, "Moment Inequalities for S_n under General Dependence Restrictions, with Applications", Z. Wahrsch. Verw. Gebiete **43**, 1–21 (1978).
15. J. R. Magnus, "The Moments of Products of Quadratic Forms in Normal Variables", Statistica Neerlandica **32**, 201–210 (1978).
16. F. Moricz, "On the Convergence Properties of Weakly Multiplicative Systems", Acta Sci. Math. **38**, 127–144 (1976).
17. A. Pajor and L. Pastur, "On the Limiting Empirical Measure of Eigenvalues of the Sum of Rank One Matrices with Log-Concave Distribution", Studia Math. **195**, 11–29 (2009).
18. E. Parzen, "On Consistent Estimates of the Spectrum of a Stationary Time Series", Ann. Math. Statist. **28**, 921–932 (1957).
19. L. Pastur and M. Shcherbina, *Eigenvalue distribution of large random matrices*, in *Mathematical Surveys and Monographs* (American Mathematical Society, Providence, RI, 2011), Vol. 171.
20. P. Révész, "M-Mixing Systems", Acta Sci. Math. **20**, 431–442 (1969).
21. H. White, *Asymptotic Theory for Econometricians* (Academic Press, Orlando, Fla, 2000).
22. W. B. Wu and H. Xiao, "Covariance Matrix Estimation for Stationary Time Series", Ann. Statist. **40**, 466–493 (2012).