mathematics

MDPI

Article

Asymptotic Normality in Linear Regression with Approximately Sparse Structure

Saulius Jokubaitis † and **Remigijus Leipus** *,†

Faculty of Mathematics and Informatics, Institute of Applied Mathematics, Vilnius University, Naugarduko 24, LT-03225 Vilnius, Lithuania; saulius.jokubaitis@mif.vu.lt
* Correspondence: remigijus.leipus@mif.vu.lt
† These authors contributed equally to this work.

Abstract: In this paper, we study the asymptotic normality in high-dimensional linear regression. We focus on the case where the covariance matrix of the regression variables has a KMS structure, in asymptotic settings where the number of predictors, p, is proportional to the number of observations, n. The main result of the paper is the derivation of the exact asymptotic distribution for the suitably centered and normalized squared norm of the product between predictor matrix, $\mathbb{X}$, and outcome variable, Y, i.e., the statistic $\|\mathbb{X}'Y\|_2^2$, under rather unrestrictive assumptions for the model parameters β_j. We employ variance-gamma distribution in order to derive the results, which, along with the asymptotic results, allows us to easily define the exact distribution of the statistic. Additionally, we consider a specific case of approximate sparsity of the model parameter vector β and perform a Monte Carlo simulation study. The simulation results suggest that the statistic approaches the limiting distribution fairly quickly even under high variable multi-correlation and relatively small number of observations, suggesting possible applications to the construction of statistical testing procedures for the real-world data and related problems.

Keywords: linear regression; sparsity; asymptotic normality; variance-gamma distribution

MSC: 60F05; 62E20; 62J99

check for updates

Citation: Jokubaitis, S.; Leipus, R. Asymptotic Normality in Linear Regression with Approximately Sparse Structure. *Mathematics* **2022**, *10*, 1657. https://doi.org/10.3390/math10101657

Academic Editors: Andrzej Sokołowski and Seifedine Kadry

Received: 1 March 2022
Accepted: 7 May 2022
Published: 12 May 2022

1. Introduction

Consider a linear regression model

$$Y = \mathbb{X}\beta + \varepsilon, \tag{1}$$

where $Y := (y_1, \ldots, y_n)' \in \mathbb{R}^{n\times 1}$ are n observations of outcome and $\mathbb{X} = (X_1, \ldots, X_n)' \in \mathbb{R}^{n\times p}$ are p-dimensional predictors with $X_1, \ldots, X_n$ being i.i.d. $p \times 1$ random vectors $X_i = (X_{1,i}, \ldots, X_{p,i})'$, which are normally distributed with zero mean and the covariance matrix Σ, denoted $X_i \overset{d}{=} \mathcal{N}_p(0, \Sigma)$. We assume that the covariance matrix Σ has a form

$$\Sigma = (\varrho^{|i-j|})_{i,j=1}^{p} = \begin{bmatrix} 1 & \varrho & \cdots & \varrho^{p-1} \\ \varrho & 1 & \cdots & \varrho^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho^{p-1} & \varrho^{p-2} & \cdots & 1 \end{bmatrix}, \tag{2}$$

if $0 < |\varrho| < 1$ and $\Sigma = I_p$ if $\varrho = 0$ (here and below I_p denotes the $p \times p$ identity matrix). This matrix is often called the Kac–Murdock–Szegő (KMS) matrix, originally introduced in [1]. As the autocorrelation matrix of corresponding causal AR(1) processes, the KMS matrix is positive definite and is considered due to the wide array of applications in the literature and its well known spectral properties (see, e.g., [2] for a thorough literature review). When carefully chosen, such a structure could well-approximate a wide array of possible covariance structures (see, e.g., [3] for a more general approach with various Toeplitz covariance structures). Furthermore, $\varepsilon := (\varepsilon_1, \ldots, \varepsilon_n)' \in \mathbb{R}^{n \times 1} \stackrel{d}{=} \mathcal{N}_n(0, \sigma_\varepsilon^2 I_n)$ are unobserved i.i.d. errors with $\mathbb{E}\varepsilon_i = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma_\varepsilon^2 > 0$, and $\beta := (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^{p \times 1}$ is an unknown p-dimensional parameter. In practice, the assumption is that $\mathbb{E}X_i = 0$ can be untenable, and it may be appropriate to add an intercept to the linear model (1); however, for simplicity, throughout this paper we will assume that the intercept is known and the variables are centered. Similar settings are considered when dealing with certain geospatial data, longitudinal studies, microarray data, and research on approximate message passing algorithms (see, e.g., [4–9]).

This paper is concerned with the derivation of the exact asymptotic distribution for the suitably centered and normalized squared norm $\|\mathbb{X}'Y\|_2^2$ under the assumption of the KMS type covariance structure in (2), where p and n are assumed to be large. Throughout the paper, we assume that $p, n \to \infty$ and $p/n \to c \in (0, \infty)$. We are particularly interested in cases where $p > n$. Statistics of such form arise in various applications in the context of high-dimensional linear regression, and under normality assumptions general results can be derived using random matrix theory through Wishart distributions (see, e.g., [7,10–12]). Dealing with such statistics typically require strong restrictions on the model parameters β; however, in this paper, we only require that $\|\beta\|_2^2 < \infty$ is satisfied. Moreover, our results could be extended by using β-generating functions (e.g., parameters of FARIMA models). In comparison to the related papers, ref. [12] assumes exact sparsity, while [7,10] require approximate sparsity.

We approach the problem following an observation by [13] that the distribution of product of Gaussian random variables admits a variance-gamma distribution, which results in a set of attractive properties. We contribute to the literature on variance-gamma distribution by extending the results by [14–16]. We demonstrate that, along with the derivation of the asymptotic distribution of $\|\mathbb{X}'Y\|_2^2$, this approach allows us to define the exact distribution of the statistic given any fixed values p, n, which can be expressed through a combination of gamma and normal random variables. In the related literature we were not able to find results for the exact distribution and asymptotic analysis of the statistic $\|\mathbb{X}'Y\|_2^2$ based on the variance-gamma distribution. Furthermore, we deem that such a result is much easier to work with than when considering the characteristic or density functions of $\|\mathbb{X}'Y\|_2^2$ straightforwardly. Therefore, in addition to the ℓ_2-norm statistic, we argue that the obtained results can be easily extended towards alternative forms of the statistic, e.g., by using a different norm, which would reduce the problem to manipulating variance-gamma distribution, thus suggesting possible further research cases and useful extensions.

Additionally, we examine a specific case of parameter β by considering $\beta_j = j^{-1}$, $j \geq 1$. Similar structures of the vector β are often found in the literature when approximate sparsity of the coefficients in the linear regression model (1) is assumed. See, e.g., [17,18] for a broader view towards sparsity requirements and its implications to specific high-dimensional algorithms; refs. [19,20] for model selection problems in autoregressive time series models; refs. [21–28] for applications on inference of high-dimensional models and high-dimensional instrumental variable (IV) regression models; or [29–33] for recent applications of high-dimensional and sparse methods with financial and economic data. Performing Monte Carlo simulations, we find that the empirical distributions of the corresponding statistic approach the limiting distribution reasonably quickly even for large values of ϱ and c. These results suggest that the assumption of sparse structure can be included in the applications and statistical tests, thus, could be further extended following the literature on testing for sparsity or construction of signal-to-noise ratio estimators (see, e.g., [7,10–12]).

In this paper, $\stackrel{d}{=}$, $\stackrel{d}{\to}$ and $\stackrel{\mathbb{P}}{\to}$ denote the equality of distributions, convergence of distributions and convergence in probability, respectively. The notation of C represents a generic positive constant which may assume different values at various locations, and $\mathbf{1}_A$ denotes the indicator function of a set A.

The structure of the paper is as follows. In Section 2, we present the main results of the paper. In Section 3, we present useful properties of variance-gamma distribution, which are used in Section 4 in order to prove some auxiliary results. In Section 5, we present the proof of the main result. Finally, in Section 6, we provide an example of the main result under imposed approximate sparsity assumption for the parameter β of the model (1). Technical results are presented in Appendix A, while, for brevity, some straightforward yet tedious proofs are presented in the Supplementary material.

2. Main Results

In this section we formulate the main results on the normality of statistic $\|\mathbb{X}'Y\|_2^2$. Introduce the notations:

$$\kappa_{1,p} := \sum_{k=1}^{p}\sum_{l=1}^{p}\beta_k\beta_l\varrho^{|k-l|}, \tag{3}$$

$$\kappa_{2,p} := \sum_{k=1}^{p}\Big(\sum_{l=1}^{p}\beta_l\varrho^{|k-l|}\Big)^2, \tag{4}$$

$$\kappa_{3,p} := \sum_{k,l,j,j'=1}^{p}\beta_j\beta_{j'}\varrho^{|k-j|}\varrho^{|l-j'|}\varrho^{|k-l|}. \tag{5}$$

It can be observed that under $\sum_{j=1}^{\infty}\beta_j^2 < \infty$, there exist limits

$$\kappa_i = \lim_{p\to\infty}\kappa_{i,p},\ i = 1,2,3.$$

Additionally, $\kappa_{2,p} \geq 0$. Since $(\varrho^{|i-j|})_{i,j=1}^{p}$ is positive semi-definite, $\kappa_{i,p} \geq 0$, $i = 1,3$. Indeed, $\sum_{k,l=1}^{p}\varrho^{|k-l|}a_k a_l \geq 0$, thus it suffices to take $a_k = \beta_k$ for $i = 1$ and $a_k = \sum_{j=1}^{p}\beta_j\varrho^{|k-j|}$ for $i = 3$.

Our first main result is the following theorem.

Theorem 1. *Assume the model in (1) with covariance structure in (2). Let $n \to \infty$ and let $p = p_n$ satisfy*

$$p \to \infty, \qquad \frac{p}{n} \to c \in (0,\infty). \tag{6}$$

Let also the β_j satisfy

$$\sum_{j=1}^{\infty}\beta_j^2 < \infty. \tag{7}$$

Then

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2\kappa_{2,p} - pn(\kappa_{1,p}+\sigma_\varepsilon^2)}{n^{3/2}} \xrightarrow{d} \mathcal{N}(0,s^2), \tag{8}$$

where variance s^2 has the structure

$$s^2 = 4\kappa_2^2 + 4(\kappa_1+\sigma_\varepsilon^2)(2\kappa_2 c+\kappa_3) + 2c(\kappa_1+\sigma_\varepsilon^2)^2\Big(c+\frac{1+\varrho^2}{1-\varrho^2}\Big). \tag{9}$$

Our second main result deals with the case where the centering sequence in (8) is modified to include the limiting values of $\kappa_{i,p}$, $i = 1,2$.

Theorem 2. *Let the assumptions of Theorem 1 hold. In addition, assume that* $\sum_{j=p+1}^{\infty}\beta_j^2 = o(p^{-1/2})$ *and* $\sup_{j\ge 1}|\beta_j|j^{\alpha} < \infty$ *with* $\alpha > 1/2$. *Then,*

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2\big(\kappa_2 + c(\kappa_1 + \sigma_\varepsilon^2)\big)}{n^{3/2}} \quad \xrightarrow{d} \quad \mathcal{N}(0, s^2). \tag{10}$$

The proofs of these theorems are given in Section 5.

Remark 1. *For alternative expressions of* κ_1, κ_2 *and* κ_3, *see Lemma 5 below.*

Define

$$\beta(x) \quad := \quad \sum_{j=1}^{\infty}\beta_j^2 x^j, \ \ |x| \le 1.$$

The following corollary deals with the case when $\varrho = 0$, i.e., $\Sigma = I_p$. The result follows from Theorem 2, noting that in this case $\kappa_i = \beta(1)$, $i = 1, 2, 3$.

Corollary 1. *Assume a model (1) with covariance structure* $\Sigma = I_p$. *Let assumptions (6) and (7) be satisfied. In addition, assume that* $\sum_{j=p+1}^{\infty}\beta_j^2 = o(p^{-1/2})$ *and* $\sup_{j\ge 1}|\beta_j|j^{\alpha} < \infty$ *with* $\alpha > 1/2$. *Then,*

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2(\beta(1)(1+c) + c\sigma_\varepsilon^2)}{n^{3/2}} \quad \xrightarrow{d} \quad \mathcal{N}(0, s^2), \tag{11}$$

where

$$s^2 \quad = \quad 2\beta(1)^2(4 + 5c + c^2) + 4\beta(1)\sigma_\varepsilon^2(1 + 3c + c^2) + 2\sigma_\varepsilon^4(c + c^2). \tag{12}$$

3. Properties of the Variance-Gamma Distribution

In this section, we provide some properties of the variance-gamma distribution, which will be used in the following proofs.

Recall that the variance-gamma distribution with parameters $r > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$ and $\mu \in \mathbb{R}$ has density

$$f^{\mathrm{VG}}(x) = \frac{1}{\sigma\sqrt{\pi}\Gamma(r/2)}\,\mathrm{e}^{\theta(x-\mu)/\sigma^2}\left(\frac{|x-\mu|}{2\sqrt{\theta^2+\sigma^2}}\right)^{(r-1)/2} K_{(r-1)/2}\left(\frac{\sqrt{\theta^2+\sigma^2}}{\sigma^2}\,|x-\mu|\right), \tag{13}$$

where $x \in \mathbb{R}$, $K_\nu(x)$ is the modified Bessel function of the second kind. For a random variable Q with density (13), we write $Q \overset{d}{=} \mathrm{VG}(r, \theta, \sigma, \mu)$. Let $\Gamma(a, b)$, $a > 0$, $b > 0$, denote the gamma distribution with density

$$f^{\mathrm{G}}(x) \quad = \quad \frac{b^a}{\Gamma(a)}\,x^{a-1}\mathrm{e}^{-bx}, \ \ x > 0.$$

It holds that

$$Q \quad \overset{d}{=} \quad \mu + \theta W_r + \sigma\sqrt{W_r}U, \tag{14}$$

where $W_r \overset{d}{=} \Gamma(r/2, 1/2)$, $U \overset{d}{=} \mathcal{N}(0, 1)$, W_r and U are independent. The characteristic function of $Q \overset{d}{=} \mathrm{VG}(r, \theta, \sigma, \mu)$ has a form (see, e.g., [34,35])

$$\varphi_Q(t) \quad = \quad \frac{\mathrm{e}^{\mathrm{i}\mu t}}{(1 + \sigma^2 t - 2\mathrm{i}\theta t)^{r/2}}, \ \ t \in \mathbb{R}. \tag{15}$$

We note the following properties of the variance-gamma distribution.

(i) If $Q_1 \stackrel{d}{=} \mathrm{VG}(r_1,\theta,\sigma,\mu_1)$ and $Q_2 \stackrel{d}{=} \mathrm{VG}(r_2,\theta,\sigma,\mu_2)$ are independent random variables then

$$Q_1+Q_2 \stackrel{d}{=} \mathrm{VG}(r_1+r_2,\theta,\sigma,\mu_1+\mu_2).$$

(ii) If $Q \stackrel{d}{=} \mathrm{VG}(r,\theta,\sigma,\mu)$, then for any $a>0$

$$aQ \stackrel{d}{=} \mathrm{VG}(r,a\theta,a\sigma,a\mu).$$

The following proposition is crucial for our purposes.

Proposition 1. (i) *If* $(\xi_1,\xi_2)' \stackrel{d}{=} \mathcal{N}_2(0,\Sigma)$, *where* $\Sigma = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, *then*

$$\xi_1\xi_2 \stackrel{d}{=} \mathrm{VG}(1,\varrho\sigma_1\sigma_2,\sqrt{1-\varrho^2}\sigma_1\sigma_2,0).$$

(ii) *If* $(\xi_{1j},\xi_{2j})'$, $j=1,\dots,n$, *are i.i.d. random vectors with common distribution* $\mathcal{N}_2(0,\Sigma)$, *then*

$$\sum_{j=1}^{n}\xi_{1j}\xi_{2j} \stackrel{d}{=} \mathrm{VG}(n,\varrho\sigma_1\sigma_2,\sqrt{1-\varrho^2}\sigma_1\sigma_2,0)$$

and

$$\sum_{j=1}^{n}\xi_{1j}\xi_{2j} \stackrel{d}{=} \sigma_1\sigma_2(\varrho W_n+\sqrt{1-\varrho^2}\sqrt{W_n}U),$$

where $W_n \stackrel{d}{=} \Gamma(n/2,1/2)$ *and* $U \stackrel{d}{=} \mathcal{N}(0,1)$ *are independent random variables.*
(iii) *Assume that* $(\xi_{1j}^{(1)},\dots,\xi_{1j}^{(p)},\xi_{2j})'$, $j=1,\dots,n$, *are i.i.d. copies of* $(\xi_1^{(1)},\dots,\xi_1^{(p)},\xi_2)' \stackrel{d}{=} \mathcal{N}_{p+1}(0,\Sigma^{(p)})$ *and let* $\varrho^{(kl)} := \mathrm{Corr}(\xi_1^{(k)},\xi_1^{(l)})$, $\varrho^{(k)} := \mathrm{Corr}(\xi_1^{(k)},\xi_2)$, $(\sigma_1^{(k)})^2 := \mathrm{Var}(\xi_1^{(k)})$, $\sigma_2^2 := \mathrm{Var}(\xi_2)$, $k,l=1,\dots,p$. *Then*

$$\begin{pmatrix} \sum_{j=1}^{n}\xi_{1j}^{(1)}\xi_{2j} \\ \vdots \\ \sum_{j=1}^{n}\xi_{1j}^{(p)}\xi_{2j} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \sigma_1^{(1)}\sigma_2(\varrho^{(1)}W_n+\sqrt{1-(\varrho^{(1)})^2}\sqrt{W_n}U_1) \\ \vdots \\ \sigma_1^{(p)}\sigma_2(\varrho^{(p)}W_n+\sqrt{1-(\varrho^{(p)})^2}\sqrt{W_n}U_p) \end{pmatrix},$$

where $(U_1,\dots,U_p)' \stackrel{d}{=} \mathcal{N}_p(0,\Sigma_U)$, $\Sigma_U=(\sigma_U^{(kl)})$ *with*

$$\sigma_U^{(k,l)} = \mathbb{E}U_kU_l = \frac{\varrho^{(kl)}-\varrho^{(k)}\varrho^{(l)}}{\sqrt{1-(\varrho^{(k)})^2}\sqrt{1-(\varrho^{(l)})^2}},\quad k,l=1,\dots,p. \tag{16}$$

Proof. The statements in (i), (ii) are well known, see e.g., [16]. The proof of part (iii) follows from Lemma 1. □

Lemma 1. *Assume that* $(\xi_1^{(1)}, \dots, \xi_1^{(p)}, \xi_2)'$ *has distribution* $\mathcal{N}_{p+1}(0, \Sigma^{(p)})$ *and let* $\varrho^{(kl)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_1^{(l)})$, $\varrho^{(k)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_2)$, $(\sigma_1^{(k)})^2 := \mathrm{Var}(\xi_1^{(k)})$, $\sigma_2^2 := \mathrm{Var}(\xi_2)$, $k, l = 1, \dots, p$. *Then*

$$\begin{pmatrix} \xi_1^{(1)}\xi_2 \\ \vdots \\ \xi_1^{(p)}\xi_2 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \sigma_1^{(1)}\sigma_2\big(\varrho^{(1)}W_1 + \sqrt{1-(\varrho^{(1)})^2}\sqrt{W_1}U_1\big) \\ \vdots \\ \sigma_1^{(p)}\sigma_2\big(\varrho^{(p)}W_1 + \sqrt{1-(\varrho^{(p)})^2}\sqrt{W_1}U_p\big) \end{pmatrix},$$

where $W_1 \stackrel{d}{=} \Gamma(1/2, 1/2)$, $(U_1, \dots, U_p)'$ *is, independent of* W_1*, zero mean normal vector with covariances in* (16)*.*

Proof. It suffices to prove that for any $(t_1, \dots, t_p) \in \mathbb{R}^p$ it holds

$$\Big(\sum_{k=1}^p t_k \xi_1^{(k)}\Big)\xi_2 \stackrel{d}{=} \sigma_2 \sum_{k=1}^p t_k \sigma_1^{(k)}\big(\varrho^{(k)}W_1 + \sqrt{1-(\varrho^{(k)})^2}\sqrt{W_1}U_k\big). \tag{17}$$

Since

$$\sum_{k=1}^p t_k \xi_1^{(k)} \stackrel{d}{=} \mathcal{N}\Big(0, \sum_{k,l=1}^p t_k t_l \varrho^{(kl)}\sigma_1^{(k)}\sigma_1^{(l)}\Big), \ \xi_2 \stackrel{d}{=} \mathcal{N}(0, \sigma_2^2),$$

by Proposition 1(i) we obtain that

$$\Big(\sum_{k=1}^p t_k \xi_1^{(k)}\Big)\xi_2 \stackrel{d}{=} \mathrm{VG}\Big(1, \sigma_2 \sum_{k=1}^p t_k \varrho^{(k)}\sigma_1^{(k)}, \sigma_2\sqrt{\sum_{k,l=1}^p t_k t_l \sigma_1^{(k)}\sigma_1^{(l)}(\varrho^{(kl)} - \varrho^{(k)}\varrho^{(l)})}, 0\Big). \tag{18}$$

For the right-hand side of (17) write

$$\begin{aligned} &\sigma_2 \sum_{k=1}^p t_k \sigma_1^{(k)}\big(\varrho^{(k)}W_1 + \sqrt{1-(\varrho^{(k)})^2}\sqrt{W_1}U_k\big) \\ &= \Big(\sigma_2 \sum_{k=1}^p t_k \sigma_1^{(k)}\varrho^{(k)}\Big)W_1 + \Big(\sigma_2 \sum_{k=1}^p t_k \sigma_1^{(k)}\sqrt{1-(\varrho^{(k)})^2}U_k\Big)\sqrt{W_1}. \end{aligned}$$

Here, by (16),

$$\begin{aligned} \sigma_2 \sum_{k=1}^p t_k \sigma_1^{(k)}\sqrt{1-(\varrho^{(k)})^2}U_k &\stackrel{d}{=} \sigma_2\Big(\sum_{k,l=1}^p t_k t_l \sigma_1^{(k)}\sigma_1^{(l)}\sqrt{1-(\varrho^{(k)})^2}\sqrt{1-(\varrho^{(l)})^2}\mathbb{E}(U_kU_l)\Big)^{1/2}U_1 \\ &= \sigma_2\Big(\sum_{k,l=1}^p t_k t_l \sigma_1^{(k)}\sigma_1^{(l)}(\varrho^{(kl)} - \varrho^{(k)}\varrho^{(l)})\Big)^{1/2}U_1. \end{aligned}$$

Note that $U_1 \stackrel{d}{=} \mathcal{N}(0,1)$. So that,

$$\begin{aligned} &\sigma_2 \sum_{k=1}^p t_k \sigma_1^{(k)}\big(\varrho^{(k)}W_1 + \sqrt{1-(\varrho^{(k)})^2}\sqrt{W_1}U_k\big) \\ &\stackrel{d}{=} \Big(\sigma_2 \sum_{k=1}^p t_k \sigma_1^{(k)}\varrho^{(k)}\Big)W_1 + \sigma_2\Big(\sum_{k,l=1}^p t_k t_l \sigma_1^{(k)}\sigma_1^{(l)}(\varrho^{(kl)} - \varrho^{(k)}\varrho^{(l)})\Big)^{1/2}\sqrt{W_1}U_1, \end{aligned}$$

which, by representation (14), has the same VG distribution as that in (18). This proves (17). □

4. Some Auxiliary Lemmas

In this section we establish some auxiliary results that will be used in the proofs of Theorems 1 and 2. Here and throughout the paper we remove the upper indices when working with triangular schemes of random variables, e.g., $(V_1, \dots, V_p) \equiv (V_1^{(p)}, \dots, V_p^{(p)})$, whenever it is clear from the context.

Lemma 2. *Let* $V = (V_1, \dots, V_p)' \overset{d}{=} \mathcal{N}_p(0, \Sigma_V^{(p)})$, *where* $\Sigma_V^{(p)}$ *is positive definite covariance matrix and* $\operatorname{tr}((\Sigma_V^{(p)})^2) = o(p^2)$, $p \to \infty$. *Then*

$$\frac{1}{p}\sum_{k=1}^{p}\left(V_k^2 - \mathbb{E}V_k^2\right) \overset{\mathbb{P}}{\to} 0 \ \ as \ \ p \to \infty. \tag{19}$$

If, in addition, $p^{-1}\operatorname{tr}(\Sigma_V^{(p)}) \to 1$, *then*

$$\frac{1}{p}\sum_{k=1}^{p} V_k^2 \overset{\mathbb{P}}{\to} 1 \ \ as \ \ p \to \infty. \tag{20}$$

Proof. Due to the Spectral Theorem, we have

$$V'V = \sum_{k=1}^{p} V_k^2 \overset{d}{=} \sum_{j=1}^{p} \lambda_j^{(p)} \tilde{Z}_j^2, \tag{21}$$

where $\tilde{Z}_j$ are i.i.d. standard normal variables and $\lambda_1^{(p)}, \dots, \lambda_p^{(p)}$ are the eigenvalues of $\Sigma_V^{(p)}$. Observe from (21) that

$$\mathbb{E}V'V = \sum_{j=1}^{p} \lambda_j^{(p)} = \operatorname{tr}(\Sigma_V^{(p)}), \tag{22}$$

$$\operatorname{Var}(V'V) = \operatorname{Var}\Big(\sum_{j=1}^{p} \lambda_j^{(p)} \tilde{Z}_j^2\Big) = 2\sum_{j=1}^{p} (\lambda_j^{(p)})^2 = 2\operatorname{tr}((\Sigma_V^{(p)})^2). \tag{23}$$

Thus, by (22) and (23), for any $\epsilon > 0$

$$\mathbb{P}\Big(\Big|\frac{1}{p}(V'V - \mathbb{E}V'V)\Big| > \epsilon\Big) \le \frac{\operatorname{Var}(V'V)}{p^2\epsilon^2} \to 0, \ \ p \to \infty,$$

and the relation in (19) follows due to assumption $\operatorname{tr}((\Sigma_V^{(p)})^2) = o(p^2)$. Finally, if $p^{-1}\operatorname{tr}(\Sigma_V^{(p)}) \to 1$, by (22), the result (19) leads to (20). □

Remark 2. *The assumption on matrix* $\Sigma_V = \Sigma_V^{(p)}$ *in Lemma 2, requiring that* $\operatorname{tr}(\Sigma_V^2) = o(p^2)$, *is not overly restrictive: assume, for example, that* $\Sigma_V = (\sigma^{(i,j)})$ *is any KMS type covariance matrix, as in (2). Then, it can be seen that*

$$\begin{aligned}\operatorname{tr}(\Sigma_V^2) &= \sum_{i,j=1}^{p} (\sigma^{(i,j)})^2 = \sum_{i,j=1}^{p} \varrho^{2|i-j|} \\ &= \sum_{|m|<p} (p - |m|)\varrho^{2|m|} \le p \sum_{|m|<p} |m|\varrho^{2|m|} = \mathcal{O}(p).\end{aligned}$$

Lemma 3. *Assume that* $\tilde{Z}_1, \tilde{Z}_2, \dots$ *are i.i.d.* $\mathcal{N}(0,1)$ *random variables. For any* $p \in \mathbb{N}$ *define*

$$\zeta_j^{(p)} := \nu_j^{(p)}(\tilde{Z}_j^2 - 1) + \gamma_j^{(p)}\sqrt{p}\tilde{Z}_j, \ \ j = 1, \dots, p, \tag{24}$$

where $\nu_j^{(p)}$, $j = 1, \ldots, p$, *are positive scalars, and* $\gamma_j^{(p)}$, $j = 1, \ldots, p$, *are real scalars, such that*

$$\sum_{j=1}^{p} (\nu_j^{(p)})^3 = o\left(\left(\sum_{j=1}^{p} \mathrm{Var}(\zeta_j^{(p)}) \right)^{3/2} \right), \tag{25}$$

$$p \sum_{j=1}^{p} (\gamma_j^{(p)})^2 \nu_j^{(p)} = o\left(\left(\sum_{j=1}^{p} \mathrm{Var}(\zeta_j^{(p)}) \right)^{3/2} \right) \tag{26}$$

with $\mathrm{Var}(\zeta_j^{(p)}) = 2(\nu_j^{(p)})^2 + p(\gamma_j^{(p)})^2$. *Then, as* $p \to \infty$,

$$\frac{\sum_{j=1}^{p} \zeta_j^{(p)}}{\sqrt{\sum_{j=1}^{p} \mathrm{Var}(\zeta_j^{(p)})}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{27}$$

Proof. The proof uses the method of cumulants and is structured as follows:

(i) We establish the moment-generating function of $\zeta_j^{(p)}$, $M_{\zeta_j^{(p)}}(t) := \mathbb{E}e^{t\zeta_j^{(p)}}$, and $\log\left(M_{\zeta_j^{(p)}}(t)\right)$;

(ii) We find $G(t;p)$ which corresponds to the cumulant generating function of the sum $\sum_{j=1}^{p} \zeta_j^{(p)}$;

(iii) We find $K(t;p) := G\left(\frac{t}{\sqrt{\sum_{j=1}^{p}(2(\nu_j^{(p)})^2 + p(\gamma_j^{(p)})^2)}}; p \right)$, which corresponds to the cumulant generating function of the left hand side of (27);

(iv) Finally, in order to prove (27), we show that the cumulants $\varkappa_j^{(p)}$, generated by $K(t;p)$, satisfy $\varkappa_1^{(p)} = 0$, $\varkappa_2^{(p)} = 1$ and $\varkappa_d^{(p)} \to 0$, $d = 3, 4, \ldots$, as $p \to \infty$.

Step 1. First, rewrite

$$\zeta_j^{(p)} = \nu_j^{(p)} \left(\tilde{Z}_j + \frac{\gamma_j^{(p)} \sqrt{p}}{2\nu_j^{(p)}} \right)^2 - \nu_j^{(p)} - \frac{(\gamma_j^{(p)})^2 p}{4\nu_j^{(p)}}. \tag{28}$$

Here, $\psi_j^{(p)} := \left(\tilde{Z}_j + \frac{\gamma_j^{(p)} \sqrt{p}}{2\nu_j^{(p)}} \right)^2$ has a noncentral chi-squared distribution with the following moment-generating function:

$$M_{\psi_j^{(p)}}(t) := \mathbb{E}e^{t\psi_j^{(p)}} = (1-2t)^{-1/2} \exp\left\{ \left(\frac{\gamma_j^{(p)}}{2\nu_j^{(p)}} \right)^2 tp(1-2t)^{-1} \right\}, \quad |t| < \frac{1}{2}. \tag{29}$$

Therefore, by (28) and (29),

$$\begin{aligned} M_{\zeta_j^{(p)}}(t) &= M_{\psi_j^{(p)}}(\nu_j^{(p)} t) \exp\left\{ -t\nu_j^{(p)} - tp\left(\frac{\gamma_j^{(p)}}{2\nu_j^{(p)}} \right)^2 \right\} \\ &= (1 - 2\nu_j^{(p)} t)^{-\frac{1}{2}} \exp\left\{ \frac{(\gamma_j^{(p)})^2}{4\nu_j^{(p)}} tp(1 - 2\nu_j^{(p)} t)^{-1} - t\left(\nu_j^{(p)} + \frac{(\gamma_j^{(p)})^2 p}{4\nu_j^{(p)}} \right) \right\}, \end{aligned}$$

for $|t| < (2\nu_j^{(p)})^{-1}$, and

$$\log\left(M_{\zeta_j^{(p)}}(t)\right) = \left(\frac{\gamma_j^{(p)}}{2\nu_j^{(p)}}\right)^2 pt\nu_j^{(p)}\left(1-2\nu_j^{(p)}t\right)^{-1} - \frac{1}{2}\log\left(1-2\nu_j^{(p)}t\right) - t\left(\nu_j^{(p)} + \frac{(\gamma_j^{(p)})^2 p}{4\nu_j^{(p)}}\right)$$
$$= \frac{1}{2}\left((\gamma_j^{(p)})^2 p + 2(\nu_j^{(p)})^2\right)t^2 + \frac{(\gamma_j^{(p)})^2 p}{2}\sum_{k=3}^{\infty} t^k 2^{k-2}(\nu_j^{(p)})^{k-2} + \frac{1}{2}\sum_{k=3}^{\infty}\frac{2^k(\nu_j^{(p)})^k t^k}{k}.$$

Step 2. Since $\zeta_1^{(p)},\ldots,\zeta_j^{(p)}$ are independent, we have that

$$G(t;p) = \sum_{j=1}^{p}\log M_{\zeta_j^{(p)}}(t) = \frac{t^2}{2}\sum_{j=1}^{p}\left((\gamma_j^{(p)})^2 p + 2(\nu_j^{(p)})^2\right)$$
$$+ \frac{p}{2}\sum_{k=3}^{\infty}2^{k-2}t^k\sum_{j=1}^{p}(\gamma_j^{(p)})^2(\nu_j^{(p)})^{k-2} + \frac{1}{2}\sum_{k=3}^{\infty}\frac{2^k}{k}t^k\sum_{j=1}^{p}(\nu_j^{(p)})^k.$$

Step 3. It can be observed that

$$K(t;p) = G\left(\frac{t}{\sqrt{\sum_{j=1}^{p}\left(2(\nu_j^{(p)})^2 + p(\gamma_j^{(p)})^2\right)}};p\right)$$
$$= \frac{t^2}{2} + \frac{1}{2}\sum_{k=3}^{\infty}2^{k-2}t^k\frac{p\sum_{j=1}^{p}(\gamma_j^{(p)})^2(\nu_j^{(p)})^{k-2}}{\left(\sum_{j=1}^{p}(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p)\right)^{k/2}}$$
$$+ \frac{1}{2}\sum_{k=3}^{\infty}\frac{2^k}{k}t^k\frac{\sum_{j=1}^{p}(\nu_j^{(p)})^k}{\left(\sum_{j=1}^{p}(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p)\right)^{k/2}} = \sum_{k=1}^{\infty}\varkappa_k^{(p)}\frac{t^k}{k!},$$

where $\varkappa_1^{(p)} = 0$, $\varkappa_2^{(p)} = 1$, and for $k \geq 3$,

$$\varkappa_k^{(p)} = \frac{k!2^{k-3}p\sum_{j=1}^{p}(\gamma_j^{(p)})^2(\nu_j^{(p)})^{k-2} + (k-1)!2^{k-1}\sum_{j=1}^{p}(\nu_j^{(p)})^k}{\left(\sum_{j=1}^{p}(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p)\right)^{k/2}}. \tag{30}$$

Step 4. In order to prove that (27) holds, it remains to show that, as $p \to \infty$, $\varkappa_d^{(p)} \to 0$ for all $d \geq 3$. By (30), it is equivalent to showing that for any fixed $k \geq 3$, as $p \to \infty$,

$$\frac{\sum_{j=1}^{p}(\nu_j^{(p)})^k}{\left(\sum_{j=1}^{p}\left(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p\right)\right)^{k/2}} \to 0, \tag{31}$$
$$\frac{p\sum_{j=1}^{p}(\gamma_j^{(p)})^2(\nu_j^{(p)})^{k-2}}{\left(\sum_{j=1}^{p}\left(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p\right)\right)^{k/2}} \to 0. \tag{32}$$

In order to prove (31) we use induction. The case for $k = 3$ holds by assumption. Assuming that (31) holds for fixed $k \geq 3$, we have

$$\begin{aligned}\frac{\sum_{j=1}^{p}(\nu_j^{(p)})^{k+1}}{\left(\sum_{j=1}^{p}\left(2(\nu_j^{(p)})^2+(\gamma_j^{(p)})^2 p\right)\right)^{(k+1)/2}} &\leq \frac{\left(\sum_{j'=1}^{p}(\nu_{j'}^{(p)})^2\right)^{1/2}\sum_{j=1}^{p}(\nu_j^{(p)})^k}{\left(\sum_{j=1}^{p}\left(2(\nu_j^{(p)})^2+(\gamma_j^{(p)})^2 p\right)\right)^{(k+1)/2}} \\ &\leq \frac{\left(\sum_{j'=1}^{p}\left(2(\nu_{j'}^{(p)})^2+(\gamma_{j'}^{(p)})^2 p\right)\right)^{1/2}\sum_{j=1}^{p}(\nu_j^{(p)})^k}{\left(\sum_{j=1}^{p}\left(2(\nu_j^{(p)})^2+(\gamma_j^{(p)})^2 p\right)\right)^{(k+1)/2}} \\ &= \frac{\sum_{j=1}^{p}(\nu_j^{(p)})^k}{\left(\sum_{j=1}^{p}\left(2(\nu_j^{(p)})^2+(\gamma_j^{(p)})^2 p\right)\right)^{k/2}} \quad \to \quad 0,\end{aligned}$$

concluding that (31) holds for all $k \geq 3$. The proof for (32) is analogous: the case for $k = 3$ holds by assumption; thus, we repeat the same arguments as with (31) and conclude that (32) holds for all $k \geq 3$. This concludes the proof of the lemma. □

5. Proof of the Main Results

In this section we give the proofs of Theorems 1 and 2. Throughout the proofs, we express corresponding constants in terms of $\kappa_{i,p}$ and κ_i, $i = 1,2,3$, introduced in (3)–(5). Recall that $\kappa_{i,p} \geq 0$, and, by Remark 3, $\kappa_i < \infty$, for $i = 1,2,3$.

Proof of Theorem 1. Write

$$\|\mathbb{X}'Y\|_2^2 \quad = \quad H_1^2 + \cdots + H_p^2 \; =: \; H,$$

where

$$H_k \quad := \quad \sum_{j=1}^{n} X_{k,j}\left(\sum_{l=1}^{p}\beta_l X_{l,j} + \varepsilon_j\right), \quad k = 1,\ldots,p.$$

Denote $Z_j := \sum_{l=1}^{p}\beta_l X_{l,j} + \varepsilon_j, j = 1,\ldots,n$. By covariance structure (2) and $X_{k,j} \stackrel{d}{=} \mathcal{N}(0,1)$, $\varepsilon_j \stackrel{d}{=} \mathcal{N}(0,\sigma_\varepsilon^2)$, we have $Z_j \stackrel{d}{=} \mathcal{N}(0,\sigma_Z^2)$, where $\sigma_Z^2 = \sum_{l,l'=1}^{p}\beta_l\beta_{l'}\varrho^{|l-l'|} + \sigma_\varepsilon^2$ and $\mathrm{Cov}(X_{k,j},Z_j) = \sum_{l=1}^{p}\beta_l\varrho^{|k-l|}$.

Applying Proposition 1(iii) with $\xi_{1j}^{(k)} = X_{k,j}$, $\xi_{2j} = Z_j$, and $\sigma_1^{(k)} = 1$, $\sigma_{2,p} = \sigma_Z$, $\theta_k^{(p)} := \varrho^{(k)} = \sigma_Z^{-1}\sum_{l=1}^{p}\beta_l\varrho^{|k-l|}$, where $\varrho^{(kl)} = \varrho^{|k-l|}$, we obtain that

$$\|\mathbb{X}'Y\|_2^2 \quad \stackrel{d}{=} \quad \sigma_{2,p}^2\sum_{k=1}^{p}\left(\theta_k^{(p)}W_n + \sqrt{1-(\theta_k^{(p)})^2}\sqrt{W_n}U_k\right)^2,$$

where $W_n \stackrel{d}{=} \Gamma(n/2,1/2)$ and $(U_1,\ldots,U_p)' \stackrel{d}{=} \mathcal{N}_p(0,\Sigma_U^{(p)})$ with $\Sigma_U^{(p)} = (\sigma_U^{(k,l)})$ defined as (see (16)):

$$\sigma_U^{(k,l)} \quad = \quad \frac{\varrho^{|k-l|} - \theta_k^{(p)}\theta_l^{(p)}}{\sqrt{1-(\theta_k^{(p)})^2}\sqrt{1-(\theta_l^{(p)})^2}}, \quad k,l = 1,\ldots,p. \tag{33}$$

By expanding the square we can write

$$\begin{aligned}\|\mathbb{X}'Y\|_2^2 \quad \stackrel{d}{=} \quad & \sigma_{2,p}^2\Big((W_n - \mathbb{E}W_n + \mathbb{E}W_n)^2\sum_{k=1}^{p}(\theta_k^{(p)})^2 + 2W_n^{3/2}\sum_{k=1}^{p}\theta_k^{(p)}\sqrt{1-(\theta_k^{(p)})^2}U_k \\ & + (W_n - \mathbb{E}W_n)\sum_{k=1}^{p}\left(1-(\theta_k^{(p)})^2\right)U_k^2 + \mathbb{E}W_n\sum_{k=1}^{p}\left(1-(\theta_k^{(p)})^2\right)U_k^2\Big).\end{aligned}$$

By further rearranging the right-hand side, we have

$$\frac{\|\mathbb{X}'Y\|_2^2}{n^{3/2}} \overset{d}{=} I_1 + I_2 + I_3 + I_4, \tag{34}$$

where

$$I_1 := \frac{\sigma_{2,p}^2}{n^{3/2}}(W_n - \mathbb{E}W_n)^2 \sum_{k=1}^{p} (\theta_k^{(p)})^2, \tag{35}$$

$$I_2 := \frac{\sigma_{2,p}^2}{n^{3/2}}(W_n - \mathbb{E}W_n)\Big(2\mathbb{E}W_n \sum_{k=1}^{p} (\theta_k^{(p)})^2 + \sum_{k=1}^{p}(1-(\theta_k^{(p)})^2)U_k^2\Big), \tag{36}$$

$$I_3 := \frac{\sigma_{2,p}^2}{n^{3/2}} 2W_n^{3/2} \sum_{k=1}^{p} \theta_k^{(p)} \sqrt{1-(\theta_k^{(p)})^2}U_k + \frac{\sigma_{2,p}^2}{n^{3/2}} \mathbb{E}W_n \sum_{k=1}^{p} ((1-(\theta_k^{(p)})^2)U_k^2 - 1), \tag{37}$$

$$I_4 := \frac{\sigma_{2,p}^2}{n^{3/2}} \Big(p\mathbb{E}W_n + (\mathbb{E}W_n)^2 \sum_{k=1}^{p} (\theta_k^{(p)})^2\Big). \tag{38}$$

We will show that, as $p, n \to \infty$, $p/n \to c \in (0, \infty)$, the term $I_1 = o_P(1)$, while the terms I_2 and I_3 are asymptotically normal. More precisely, we will show that $I_2 \xrightarrow{d} \mathcal{N}(0, s_1^2)$ and $I_3 \xrightarrow{d} \mathcal{N}(0, s_2^2)$, where s_1^2 and s_2^2 are given by (44) and (62) below. Here, since W_n and $(U_1, \ldots, U_p)'$ are mutually independent for each n, it follows that $I_2 + I_3 \xrightarrow{d} \mathcal{N}(0, s_1^2 + s_2^2)$. Finally, the term I_4 defines the mean of the statistic, i.e.,

$$\frac{\|\mathbb{X}'Y\|_2^2}{n^{3/2}} - I_4 \xrightarrow{d} \mathcal{N}(0, s_1^2 + s_2^2). \tag{39}$$

Thus, we will conclude by establishing that $I_4 = \sqrt{n}(\kappa_{2,p} + pn^{-1}(\kappa_{1,p} + \sigma_\varepsilon^2))$, while $s_1^2 + s_2^2 = s^2$, as in the statement of the theorem.

First, consider I_1 defined in (35). We will show that $I_1 = o_P(1)$. Denote

$$c_2 := \lim_{p\to\infty} \sum_{k=1}^{p} (\theta_k^{(p)})^2 = (\kappa_1 + \sigma_\varepsilon^2)^{-1}\kappa_2, \quad \sigma_2^2 := \lim_{p\to\infty} \sigma_{2,p}^2 = \kappa_1 + \sigma_\varepsilon^2. \tag{40}$$

It is clear that $c_2 < \infty$ and $\sigma_2^2 < \infty$. Recall that, by CLT,

$$\frac{W_n - \mathbb{E}W_n}{n^{1/2}} \xrightarrow{d} \mathcal{N}(0, 2). \tag{41}$$

Therefore,

$$I_1 = \mathcal{O}(1)n^{-1/2}\Big(\frac{W_n - \mathbb{E}W_n}{n^{1/2}}\Big)^2 = o(1)\mathcal{O}_P(1) = o_P(1). \tag{42}$$

Second, consider I_2, defined in (36). We will show that

$$I_2 \xrightarrow{d} \mathcal{N}(0, s_1^2) \tag{43}$$

with s_1^2 given by

$$s_1^2 = 2\sigma_2^4(2c_2 + c)^2 = 8\kappa_2^2 + 8c(\kappa_1 + \sigma_\varepsilon^2)\kappa_2 + 2c^2(\kappa_1 + \sigma_\varepsilon^2)^2. \tag{44}$$

Rewrite

$$I_2 = \sigma_{2,p}^2 \frac{W_n - \mathbb{E}W_n}{n^{1/2}} \Big(\frac{2\mathbb{E}W_n}{n} \sum_{k=1}^{p} (\theta_k^{(p)})^2 + \frac{1}{n} \sum_{k=1}^{p} (1-(\theta_k^{(p)})^2)U_k^2\Big). \tag{45}$$

Applying (40) and (41) for the outer term of (45), we obtain

$$\sigma_{2,p}^2 \frac{W_n - \mathbb{E}W_n}{n^{1/2}} \quad \xrightarrow{d} \quad \mathcal{N}(0, 2\sigma_2^4).$$

We will show that the inner term of (45) approaches $2c_2 + c$. Since $\mathbb{E}W_n = n$, by (40) and assumption $p/n \to c$ it suffices to prove the convergence

$$\frac{1}{p}\sum_{k=1}^{p}(1-(\theta_k^{(p)})^2)U_k^2 \quad \xrightarrow{\mathbb{P}} \quad 1. \tag{46}$$

Denote matrix

$$A \quad := \quad \operatorname{diag}\left(1-(\theta_1^{(p)})^2, \, \ldots \, , 1-(\theta_p^{(p)})^2\right). \tag{47}$$

To prove (46), we apply Lemma 2 with $V_j \;=\; \sqrt{1-(\theta_j^{(p)})^2}U_j$, $j \;=\; 1,\ldots,p$, and $\Sigma_V^{(p)} = A^{1/2}\Sigma_U A^{1/2}$. The conditions of Lemma 2 will hold if $\operatorname{tr}((A^{1/2}\Sigma_U A^{1/2})^2) = \mathcal{O}(p)$ and $p^{-1}\operatorname{tr}(A^{1/2}\Sigma_U A^{1/2}) \to 1$, as $p \to \infty$. Observe, that

$$\begin{aligned}
\operatorname{tr}((A^{1/2}\Sigma_U A^{1/2})^2) &= \operatorname{tr}((A\Sigma_U)^2) \\
&= \sum_{k=1}^{p}\sum_{k'=1}^{p}(1-(\theta_k^{(p)})^2)(1-(\theta_{k'}^{(p)})^2)(\sigma_U^{(k,k')})^2 \\
&= \sum_{k=1}^{p}\sum_{k'=1}^{p}\left(\varrho^{2|k-k'|} - 2\varrho^{|k-k'|}\theta_k^{(p)}\theta_{k'}^{(p)} + (\theta_k^{(p)})^2(\theta_{k'}^{(p)})^2\right) \\
&= \sum_{k=1}^{p}\sum_{k'=1}^{p}\varrho^{2|k-k'|} - 2\left(\kappa_{1,p}+\sigma_\varepsilon^2\right)^{-1}\kappa_{3,p} + \left(\kappa_{1,p}+\sigma_\varepsilon^2\right)^{-2}\kappa_{2,p}^2 \\
&= \sum_{k=1}^{p}\sum_{k'=1}^{p}\varrho^{2|k-k'|} + o(p) \quad \sim \quad p\,\frac{1+\varrho^2}{1-\varrho^2},
\end{aligned} \tag{48}$$

since $\kappa_i < \infty$, $i = 1,2,3$ and $\kappa_{1,p} \geq 0$. Here we used (40) and the observation that

$$\sum_{k=1}^{p}\sum_{k'=1}^{p}\varrho^{|k-k'|}\theta_k^{(p)}\theta_{k'}^{(p)} \quad = \quad \frac{\kappa_{3,p}}{\kappa_{1,p}+\sigma_\varepsilon^2} \quad \to \quad \frac{\kappa_3}{\kappa_1+\sigma_\varepsilon^2}, \quad \text{as } p \to \infty. \tag{49}$$

Similarly, we have

$$\frac{1}{p}\operatorname{tr}(A^{1/2}\Sigma_U A^{1/2}) \quad = \quad \frac{1}{p}\sum_{k=1}^{p}(1-(\theta_k^{(p)})^2) \quad = \quad 1 \; - \; \frac{\kappa_{2,p}}{p(\kappa_{1,p}+\sigma_\varepsilon^2)} \quad \to \quad 1,$$

since, by Lemma A4, $\kappa_{2,p} = o(p)$, while $\kappa_{1,p} \geq 0$, $\kappa_1 < \infty$. This concludes the proof of (46).

Next, consider I_3, defined by (37). We will show that

$$I_3 \quad \xrightarrow{d} \quad \mathcal{N}(0, s_2^2), \tag{50}$$

with s_2^2 defined in (62). Write

$$I_3 \quad = \quad \sigma_{2,p}^2\left(2\frac{W_n^{3/2}}{n^{3/2}}\mathbf{b}'U + n^{-1/2}(U'AU - p)\right),$$

where $U \;=\; (U_1,\ldots,U_p)'$, A is defined by (47), and

$$\mathbf{b} \quad = \quad \left(\theta_1^{(p)}\sqrt{1-(\theta_1^{(p)})^2}, \ldots, \theta_p^{(p)}\sqrt{1-(\theta_p^{(p)})^2}\right)'.$$

Observe that $n^{-3/2}W_n^{3/2} \xrightarrow{P} 1$ due to the Law of Large Numbers. Thus, since W_n and U are independent for any n and $p/n \to c$, it follows that

$$I_3 \quad = \quad \sigma_{2,p}^2\Big(2\mathbf{b}'U + \sqrt{\frac{c}{p}}(U'AU - p)\Big) + o_P(1). \tag{51}$$

First, we consider the inner term of (51) and show that as $p \to \infty$,

$$2\mathbf{b}'U \;+\; \sqrt{\frac{c}{p}}(UAU' - p) \;\xrightarrow{d}\; V_2, \tag{52}$$

where $V_2 \stackrel{d}{=} \mathcal{N}(0, \sigma_2^{-4}s_2^2)$. Then, (50) readily follows from (51).

Recall, that $U \stackrel{d}{=} \mathcal{N}_p(0, \Sigma_U)$, $\Sigma_U > 0$. Further, let $\tilde{Z} \stackrel{d}{=} \mathcal{N}_p(0, I_p)$. Clearly, one has that $U \stackrel{d}{=} \Sigma_U^{1/2}\tilde{Z}$, where $\Sigma_U^{1/2}$ denotes the symmetric square root of Σ_U. By the Spectral Theorem, we construct $V := P'\tilde{Z}$, where $V \stackrel{d}{=} \mathcal{N}_p(0, I_p)$ and P is an orthogonal matrix that diagonalizes $\Sigma_U^{1/2}A\Sigma_U^{1/2}$, such that $P'\Sigma_U^{1/2}A\Sigma_U^{1/2}P = \Lambda$, with $\Lambda = \mathrm{diag}(\lambda_1^{(p)}, \ldots, \lambda_p^{(p)})$ comprised of the eigenvalues of $\Sigma_U^{1/2}A\Sigma_U^{1/2}$. Then,

$$\begin{aligned}
\frac{\sqrt{c}}{\sqrt{p}}(U'AU - p) + 2\mathbf{b}'U \quad &\stackrel{d}{=} \quad \frac{\sqrt{c}}{\sqrt{p}}(V'\Lambda V - p) + 2\mathbf{b}'\Sigma_U^{1/2}PV \\
&= \quad \frac{\sqrt{c}}{\sqrt{p}}\Bigg(\sum_{j=1}^{p}\Big(\lambda_j^{(p)}(V_j^2 - 1) + g_j^{(p)}\sqrt{p}V_j\Big)\Bigg) \\
&=: \quad \frac{\sqrt{c}}{\sqrt{p}}\sum_{j=1}^{p}\widetilde{V}_j^{(p)},
\end{aligned} \tag{53}$$

where $(g_1^{(p)}, \ldots, g_p^{(p)}) = 2c^{-1/2}\mathbf{b}'\Sigma_U^{1/2}P$, and

$$\widetilde{V}_j^{(p)} \quad := \quad \lambda_j^{(p)}(V_j^2 - 1) \;+\; g_j^{(p)}\sqrt{p}V_j, \quad j = 1, \ldots, p. \tag{54}$$

Clearly, $\mathbb{E}\widetilde{V}_j^{(p)} = 0$ and $\mathbb{E}(\widetilde{V}_j^{(p)})^2 = 2(\lambda_j^{(p)})^2 + (g_j^{(p)})^2 p$. Therefore, proving the result (52) is equivalent to showing:

$$\frac{\sqrt{c}}{\sqrt{p}}\sum_{j=1}^{p}\widetilde{V}_j^{(p)} \quad \xrightarrow{d} \quad \mathcal{N}(0, \sigma_2^{-4}s_2^2), \tag{55}$$

where

$$\sigma_2^{-4}s_2^2 \;=\; c\lim_{p\to\infty} p^{-1}\sum_{j=1}^{p}\mathbb{E}(\widetilde{V}_j^{(p)})^2 \;=\; 2c\lim_{p\to\infty} p^{-1}\sum_{j=1}^{p}(\lambda_j^{(p)})^2 + c\lim_{p\to\infty}\sum_{j=1}^{p}(g_j^{(p)})^2. \tag{56}$$

We prove (55) by applying Lemma 3 with $\nu_j^{(p)} = \lambda_j^{(p)}$ as the eigenvalues of $\Sigma_U^{1/2}A\Sigma_U^{1/2}$ and $\gamma_j^{(p)} = g_j^{(p)}$. By the conditions of Lemma 3, we need to show that the following holds

$$\sum_{j=1}^{p}(\lambda_j^{(p)})^3 + p\sum_{j=1}^{p}(g_j^{(p)})^2\lambda_j^{(p)} \quad = \quad o\Bigg(\Bigg(\sum_{j=1}^{p}\big(2(\lambda_j^{(p)})^2 + (g_j^{(p)})^2 p\big)\Bigg)^{3/2}\Bigg). \tag{57}$$

First, observe that $p^{-1}\sum_{j=1}^{p}(2(\lambda_j^{(p)})^2+(g_j^{(p)})^2p) \to C \in (0,\infty)$. Indeed, we have that $\sum_{j=1}^{p}(g_j^{(p)})^2 \to C_g \in (0,\infty)$, since

$$\begin{aligned}
\sum_{j=1}^{p}(g_j^{(p)})^2 &= 4c^{-1}(\mathbf{b}'\Sigma_U^{1/2}P)(\mathbf{b}'\Sigma_U^{1/2}P)' = 4c^{-1}\mathbf{b}'\Sigma_U\mathbf{b} \\
&= 4c^{-1}\sum_{j=1}^{p}\sum_{j'=1}^{p}\theta_j^{(p)}\theta_{j'}^{(p)}\sqrt{1-(\theta_j^{(p)})^2}\sqrt{1-(\theta_{j'}^{(p)})^2}\sigma_U^{(j,j')} \\
&= 4c^{-1}\sum_{j=1}^{p}\sum_{j'=1}^{p}\theta_j^{(p)}\theta_{j'}^{(p)}\left(\varrho^{|j-j'|}-\theta_j^{(p)}\theta_{j'}^{(p)}\right) \\
&\to 4c^{-1}\left(\kappa_1+\sigma_\varepsilon^2\right)^{-1}\kappa_3-4c^{-1}\left(\kappa_1+\sigma_\varepsilon^2\right)^{-2}\kappa_2^2 = C_g
\end{aligned} \tag{58}$$

by (40) and (49).

Next, by (48), we find that $p^{-1}\sum_{j=1}^{p}(\lambda_j^{(p)})^2 \to C_\lambda \in (0,\infty)$. Indeed, by (48), we have

$$\begin{aligned}
\sum_{j=1}^{p}(\lambda_j^{(p)})^2 &= \operatorname{tr}((\Sigma_U^{1/2}A\Sigma_U^{1/2})^2) = \operatorname{tr}((\Sigma_U A)^2) \\
&= \sum_{j=1}^{p}\sum_{j'=1}^{p}\varrho^{2|j-j'|}+o(p) \sim p\,\frac{1+\varrho^2}{1-\varrho^2}.
\end{aligned} \tag{59}$$

Thus, by (58) and (59), it follows that $p^{-1}\sum_{j=1}^{p}(2c(\lambda_j^{(p)})^2+(g_j^{(p)})^2p) \to C \in (0,\infty)$ and condition (57) reduces to:

$$\sum_{j=1}^{p}(\lambda_j^{(p)})^3+p\sum_{j=1}^{p}(g_j^{(p)})^2\lambda_j^{(p)} = o(p^{3/2}). \tag{60}$$

We show that (60) holds. For the first term of (60), we have

$$\begin{aligned}
\sum_{j=1}^{p}(\lambda_j^{(p)})^3 &= \operatorname{tr}((\Sigma_U^{1/2}A\Sigma_U^{1/2})^3) = \operatorname{tr}((\Sigma_U A)^3) \\
&= \sum_{i,j,k=1}^{p}\left(1-(\theta_i^{(p)})^2\right)\left(1-(\theta_k^{(p)})^2\right)\left(1-(\theta_j^{(p)})^2\right)\sigma_U^{(i,j)}\sigma_U^{(i,k)}\sigma_U^{(k,j)} \\
&= \sum_{i,j,k=1}^{p}\left(\varrho^{|i-j|}+\theta_i^{(p)}\theta_j^{(p)}\right)\left(\varrho^{|i-k|}+\theta_i^{(p)}\theta_k^{(p)}\right)\left(\varrho^{|k-j|}+\theta_k^{(p)}\theta_j^{(p)}\right) \\
&= o(p^{3/2}),
\end{aligned} \tag{61}$$

where the last equality follows from Lemma A5. For the second term of (60), observe that by Hölder's inequality and (61),

$$\begin{aligned}
p\sum_{j=1}^{p}(g_j^{(p)})^2\lambda_j^{(p)} &\le p\left(\sum_{j=1}^{p}|g_j^{(p)}|^3\right)^{2/3}\left(\sum_{j=1}^{p}(\lambda_j^{(p)})^3\right)^{1/3} \\
&= p^{3/2}\mathcal{O}(1)\left(\frac{\sum_{j=1}^{p}(\lambda_j^{(p)})^3}{p^{3/2}}\right)^{1/3} = o(p^{3/2}).
\end{aligned}$$

This concludes with (60), ensuring that the conditions of Lemma 3 hold.

Now we can establish the expression for s_2^2. By (40), (56), (58) and (59),

$$
\begin{aligned}
s_2^2 &= \sigma_2^4 \lim_{p\to\infty} \sum_{j=1}^{p} \left(2p^{-1}c(\lambda_j^{(p)})^2 + c(g_j^{(p)})^2\right) \\
&= \sigma_2^4 \lim_{p\to\infty} \frac{2c}{p}\left(\sum_{k=1}^{p}\sum_{k'=1}^{p} \varrho^{2|k-k'|} + o(p)\right) + 4\sigma_2^4\left(\kappa_1+\sigma_\varepsilon^2\right)^{-1}\kappa_3 - 4\sigma_2^4\left(\kappa_1+\sigma_\varepsilon^2\right)^{-2}\kappa_2^2 \\
&= 2c\,\frac{1+\varrho^2}{1-\varrho^2}\,(\kappa_1+\sigma_\varepsilon^2)^2 + 4(\kappa_1+\sigma_\varepsilon^2)\kappa_3 - 4\kappa_2^2. \qquad (62)
\end{aligned}
$$

By (44) and (62), recalling that $s^2 = s_1^2 + s_2^2$, we have that

$$
s^2 = 4\kappa_2^2 + 4(\kappa_1+\sigma_\varepsilon^2)(2\kappa_2 c + \kappa_3) + 2c(\kappa_1+\sigma_\varepsilon^2)^2\left(c + \frac{1+\varrho^2}{1-\varrho^2}\right). \qquad (63)
$$

Finally, consider I_4, defined by (38). Since $\mathbb{E}W_n = n$, we have that

$$
I_4 = \frac{\kappa_{1,p}+\sigma_\varepsilon^2}{n^{3/2}}\left(n^2\,\frac{\kappa_{2,p}}{\kappa_{1,p}+\sigma_\varepsilon^2} + pn\right) = \sqrt{n}\left(\kappa_{2,p} + \frac{p}{n}(\kappa_{1,p}+\sigma_\varepsilon^2)\right). \qquad (64)
$$

By (34), having established four parts by (35)–(38), we proved that (39) holds due to (42), (43), (50), (62), with terms (63) and (64), as in the statement of the theorem, thus concluding the proof. □

Before proceeding with the proof of Theorem 2, we establish the following lemma that ensures $\mathcal{O}(p^{-1/2})$ convergence rate for $\kappa_{1,p}$ and $\kappa_{2,p}$, appearing in Theorem 1, under additional restrictions for the parameters β_j.

Lemma 4. *Assume that* $\sum_{j=p+1}^{\infty}\beta_j^2 = o(p^{-1/2})$ *and* $\sup_{j\ge1}|\beta_j|j^{\alpha} < \infty$, $\alpha > 1/2$, *and* $|\varrho| < 1$. *Then,*

(i) $\kappa_1 = \kappa_{1,p} + o(p^{-1/2})$,
(ii) $\kappa_2 = \kappa_{2,p} + o(p^{-1/2})$.

Proof. For the proof see Supplementary Materials, Section S1. □

Proof of Theorem 2. Rewrite the left-hand side of (10) as follows:

$$
\begin{aligned}
\frac{\|\mathbb{X}'Y\|_2^2 - n^2(\kappa_2 + c(\kappa_1+\sigma_\varepsilon^2))}{n^{3/2}} &= \frac{\|\mathbb{X}'Y\|_2^2 - n^2\kappa_{2,p} - pn(\kappa_{1,p}+\sigma_\varepsilon^2)}{n^{3/2}} \\
&\quad + \sqrt{n}(\kappa_{2,p}-\kappa_2) + \sqrt{n}c(\kappa_{1,p}-\kappa_1) + o(1).
\end{aligned}
$$

It remains to apply Lemma 4 and Theorem 1 in order to conclude the proof of the theorem. □

We end this section by deriving two supporting results that allows us to derive convenient alternative expressions for the terms κ_1, κ_2 and κ_3. For this, we introduce functions $\beta(\cdot)$ and $b(\cdot)$ by Definition 3 below, which, under the assumptions of Theorem 1 and a given structure of β_j's, requires only to evaluate the terms $\beta(1), \beta(\varrho), \beta(\varrho^2)$ and $b_1(\varrho), b_2(\varrho)$. Then, due to Lemma 5 below, the expressions for κ_1, κ_2 and κ_3 easily follow.

Definition 3. *Assume that* $\sum_{j=1}^{\infty} \beta_j^2 < \infty$ *and* $|\varrho| \le 1$. *Define,*

$$\beta(\varrho) := \sum_{j=1}^{\infty} \beta_j^2 \varrho^j, \tag{65}$$

$$b_1(\varrho) := \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'-j}, \tag{66}$$

$$b_2(\varrho) := \sum_{j=2}^{\infty} \sum_{j'=1}^{j-1} \beta_j \beta_{j'} \varrho^{j+j'}, \tag{67}$$

and define the following quantities which involve derivatives of (65)–(67)*:*

$$\beta^{(1)}(\varrho) := \varrho \frac{\mathrm{d}\beta(\varrho)}{\mathrm{d}\varrho} = \sum_{j=1}^{\infty} j \beta_j^2 \varrho^j, \tag{68}$$

$$b_1^{(1)}(\varrho) := \varrho \frac{\mathrm{d}b_1(\varrho)}{\mathrm{d}\varrho} = \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'-j}(j'-j), \tag{69}$$

$$b_2^{(1)}(\varrho) := \varrho \frac{\mathrm{d}b_2(\varrho)}{\mathrm{d}\varrho} = \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'+j}(j'+j), \tag{70}$$

$$b^{(2)}(\varrho) := \varrho^2 \frac{\mathrm{d}^2 b_1(\varrho)}{\mathrm{d}\varrho^2} + b_1^{(1)}(\varrho) = \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'-j}(j'-j)^2. \tag{71}$$

Note that, by the rules of differentiation of power series, the functions (68)–(71) are well defined.

Lemma 5. *Let the assumptions of Theorem 1 hold. Let* κ_1, κ_2 *and* κ_3 *be given by* (3)–(5)*, respectively. Then, under notation in Definition 3, the following identities hold:*

(i) $\kappa_1 = \beta(1) + 2b_1(\varrho)$,

(ii) $\kappa_2 = \beta(1)\frac{1+\varrho^2}{1-\varrho^2} - \beta(\varrho^2)\frac{1}{1-\varrho^2} + 2\Big(b_1^{(1)}(\varrho) + b_1(\varrho)\frac{1+\varrho^2}{1-\varrho^2} - b_2(\varrho)\frac{1}{1-\varrho^2}\Big)$,

(iii) $\kappa_3 = \frac{1}{(1-\varrho^2)^2}\big((1+4\varrho^2+\varrho^4)(\beta(1)+2b_1(\varrho)) - (1+3\varrho^2)(\beta(\varrho^2)+2b_2(\varrho))\big) + \frac{1}{1-\varrho^2}\big(3b_1^{(1)}(\varrho)(1+\varrho^2) - 2\big(b_2^{(1)}(\varrho)+\beta^{(1)}(\varrho^2)\big)\big) + b^{(2)}(\varrho).$

Proof. See the proof in Appendix A.2. □

Remark 3. *From the assumptions of Definition 3 it follows that* $\beta(1), |\beta(\varrho)|, |b_1(\varrho)|, |b_2(\varrho)| < \infty$ *for* $|\varrho| < 1$. *Thus, it follows from Lemma 5 that* $\kappa_i < \infty$, $i = 1, 2, 3$.

Proof of Remark 3. Cases for $\beta(1)$ and $\beta(\varrho)$ follow straightforwardly from the assumptions. Consider $b_1(\varrho)$. Note that for any p,

$$\begin{aligned} |b_1(\varrho)| &\le \sum_{l_1,l_2=1}^{\infty} |\beta_{l_1}||\beta_{l_2}||\varrho|^{|l_1-l_2|} = \sum_{l_1,l_2=1}^{\infty} \big(|\beta_{l_1}||\varrho|^{|l_1-l_2|/2}\big)\big(|\beta_{l_2}||\varrho|^{|l_1-l_2|/2}\big) \\ &\le (1/2) \sum_{l_1,l_2=1}^{\infty} \big(\beta_{l_1}^2|\varrho|^{|l_1-l_2|} + \beta_{l_2}^2|\varrho|^{|l_1-l_2|}\big) \\ &= \sum_{l_1=1}^{\infty} \beta_{l_1}^2 \sum_{l_2=1}^{\infty} |\varrho|^{|l_1-l_2|} \le \beta(1)\frac{1+|\varrho|}{1-|\varrho|} < \infty \end{aligned}$$

by (S9). In a similar manner, it can be seen that $|b_2(\varrho)| \leq \beta(1)\frac{|\varrho|}{1-|\varrho|}$. $\square$

6. Approximate Sparsity: An Example

In this section, we study the case when coefficients β_j decay hyperbolically, i.e., $\beta_j = j^{-1}, j \geq 1$. This assumption is analogous to the assumption of approximate sparsity, as defined by [21]. The authors of the aforementioned paper note that for approximately sparse models, the regression function can be well approximated by a linear combination of relatively few important regressors, which is one of the reasons of popularity of variable selection approaches such as LASSO ([36]) and its modifications (see, e.g., [37–39]). At the same time, approximate sparsity allows all coefficients β_j to be nonzero, which is a more plausible assumption in many real world settings.

In order to derive the quantities in Theorem 2, we apply the results of Lemma 5. For this, we establish the expressions for the quantities in Definition 3.

Define the real dilogarithm function (see, e.g., [40]):

$$\mathrm{Li}_2(x) = -\int_0^x \frac{\log(1-u)}{u}\,\mathrm{d}u, \quad x \leq 1, \quad x \in \mathbb{R}. \tag{72}$$

(Here and below, $\int_0^x = -\int_x^0$ if $x \leq 0$.) For $|x| \leq 1$ the real dilogarithm has a series representation,

$$\mathrm{Li}_2(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^2}. \tag{73}$$

Then,

$$\beta(1) = \sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}, \qquad \beta(\varrho) = \sum_{j=1}^{\infty} \frac{\varrho^j}{j^2} = \mathrm{Li}_2(\varrho).$$

Additionally, we have

$$\frac{\mathrm{d}}{\mathrm{d}\varrho}\mathrm{Li}_2(\varrho) = -\frac{\log(1-\varrho)}{\varrho}. \tag{74}$$

Thus, by (68) and (74), we establish

$$\beta^{(1)}(\varrho) = \varrho\frac{\mathrm{d}}{\mathrm{d}\varrho}\beta(\varrho) = \varrho\frac{\mathrm{d}}{\mathrm{d}\varrho}\mathrm{Li}_2(\varrho) = -\log(1-\varrho).$$

Next, note that

$$\begin{aligned} b_1(\varrho) &= \sum_{i=2}^{\infty}\sum_{j=1}^{i-1} \frac{\varrho^{i-j}}{ij} = \sum_{i=2}^{\infty}\sum_{k=1}^{i-1} \frac{\varrho^k}{i(i-k)} \\ &= \sum_{k=1}^{\infty} \varrho^k \sum_{i=k+1}^{\infty} \frac{1}{i(i-k)} = \sum_{k=1}^{\infty} \frac{\varrho^k}{k} \sum_{l=1}^{k} \frac{1}{l} \\ &= \sum_{l=1}^{\infty} \frac{1}{l} \sum_{k=l}^{\infty} \frac{\varrho^k}{k} = \sum_{l=1}^{\infty} \frac{1}{l} \int_0^{\varrho} \frac{x^{l-1}}{1-x}\,\mathrm{d}x \\ &= -\int_0^{\varrho} \frac{\log(1-x)}{x(1-x)}\,\mathrm{d}x = \frac{\log^2(1-\varrho)}{2} + \mathrm{Li}_2(\varrho), \end{aligned} \tag{75}$$

where we have used identities

$$\sum_{i=k+1}^{\infty} \frac{1}{i(i-k)} = \frac{1}{k}\sum_{l=1}^{k} \frac{1}{l},\ k \geq 1, \qquad \sum_{k=l}^{\infty} \frac{\varrho^k}{k} = \int_0^{\varrho} \frac{x^{l-1}}{1-x}\,\mathrm{d}x$$

and (72). Then, by (69), (74) and (75),

$$b_1^{(1)}(\varrho) \;=\; \varrho\frac{\mathrm{d}}{\mathrm{d}\varrho}b_1(\varrho) \;=\; -\frac{\log(1-\varrho)}{1-\varrho},$$

whereas by (71),

$$b^{(2)}(\varrho) \;=\; \varrho^2\frac{\mathrm{d}^2 b_1(\varrho)}{\mathrm{d}\varrho^2} + b_1^{(1)}(\varrho) \;=\; \frac{\varrho-\varrho\log(1-\varrho)}{(1-\varrho)^2}.$$

Furthermore, note that

$$\begin{aligned}
b_2(\varrho) &= \sum_{i=2}^{\infty}\sum_{j=1}^{i-1}\frac{\varrho^{i+j}}{ij} = \sum_{i=2}^{\infty}\frac{\varrho^i}{i}\sum_{j=1}^{i-1}\frac{\varrho^j}{j} = \sum_{i=2}^{\infty}\frac{\varrho^i}{i}\int_0^{\varrho}\sum_{j=1}^{i-1}x^{j-1}\,\mathrm{d}x \\
&= \sum_{i=1}^{\infty}\frac{\varrho^{i+1}}{i+1}\int_0^{\varrho}\frac{1-x^i}{1-x}\,\mathrm{d}x \\
&= -\log(1-\varrho)\Big(\sum_{i=1}^{\infty}\frac{\varrho^i}{i}-\varrho\Big) - \int_0^{\varrho}\Big(\sum_{i=1}^{\infty}\frac{\varrho^i}{i}\frac{x^{i-1}}{1-x} - \varrho\frac{1}{1-x}\Big)\,\mathrm{d}x \\
&= -\log(1-\varrho)\sum_{i=1}^{\infty}\frac{\varrho^i}{i} - \int_0^{\varrho}\sum_{i=1}^{\infty}\frac{(\varrho x)^i}{i}\frac{1}{x(1-x)}\,\mathrm{d}x \\
&= \log^2(1-\varrho) + \int_0^{\varrho}\frac{\log(1-\varrho x)}{x(1-x)}\,\mathrm{d}x \\
&= \frac{1}{2}\big(\log^2(1-\varrho) - \mathrm{Li}_2(\varrho^2)\big),
\end{aligned} \tag{76}$$

where the last equality follows from Lemma A1. Next, by (69), (74) and (76) we have

$$b_2^{(1)}(\varrho) \;=\; \log\Big(1-\varrho^2\Big) - \frac{\varrho\log(1-\varrho)}{1-\varrho}.$$

Thus, we can apply Lemma 5(i) and arrive at the following expression for κ_1:

$$\kappa_1 \;=\; \frac{\pi^2}{6} \;+\; \log^2(1-\varrho) \;+\; 2\mathrm{Li}_2(\varrho). \tag{77}$$

Similarly, for κ_2, by collecting and simplifying the terms, by Lemmas 5(ii) and A1, we have

$$\begin{aligned}
\kappa_2 &= \frac{1+\varrho^2}{1-\varrho^2}\Big(\frac{\pi^2}{6}+2\mathrm{Li}_2(\varrho)\Big) \;-\; \frac{2\log(1-\varrho)}{1-\varrho} \;+\; \log^2(1-\varrho)\frac{\varrho^2}{1-\varrho^2} \\
&= \frac{1}{1-\varrho^2}\big((1+\varrho^2)\kappa_1 - \log^2(1-\varrho) - 2(1+\varrho)\log(1-\varrho)\big).
\end{aligned} \tag{78}$$

Lastly, for κ_3, by Lemma 5(iii), through simplification of terms, we get

$$\begin{aligned}
\kappa_3 &= \frac{1}{(1-\varrho^2)^2}\Big((1+4\varrho^2+\varrho^4)\Big(\frac{\pi^2}{6}+2\mathrm{Li}_2(\varrho)\Big) + \log^2(1-\varrho)\varrho^2(1+\varrho^2) \\
&\quad - (3-\varrho+4\varrho^2)(1+\varrho)\log(1-\varrho) + \varrho(1+\varrho)^2\Big) \\
&= \kappa_2\frac{1+3\varrho^2}{1-\varrho^2} + \frac{1}{(1-\varrho^2)^2}\Big((-1+\varrho+2\varrho^2)(1+\varrho)\log(1-\varrho) + \varrho(1+\varrho)^2 - 2\varrho^4\kappa_1\Big).
\end{aligned} \tag{79}$$

This allows us to apply Theorem 2 under the considered specification of the parameter β and conclude with the following corollary.

Corollary 2. *Assume a model* (1) *with* (2) *covariance structure and consider* $\beta_j := j^{-1}, j = 1, \ldots, p$. *Let* $p = p_n$ *satisfies*

$$p \to \infty, \qquad \frac{p}{n} \to c \in (0, \infty).$$

Then

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2(\kappa_2 + c(\kappa_1 + \sigma_\varepsilon^2))}{n^{3/2}} \quad \xrightarrow{d} \quad \mathcal{N}(0, s^2), \tag{80}$$

where

$$s^2 \quad = \quad 4\kappa_2^2 + 4(\kappa_1 + \sigma_\varepsilon^2)(2\kappa_2 c + \kappa_3) + 2c(\kappa_1 + \sigma_\varepsilon^2)^2\Big(c + \frac{1+\varrho^2}{1-\varrho^2}\Big), \tag{81}$$

and κ_1, κ_2 *and* κ_3 *are defined by* (77)–(79), *respectively.*

In order to illustrate the results of Corollary 2, we end this section with a Monte Carlo simulation study where we generate 1000 independent replications of the statistic $\|\mathbb{X}'Y\|_2^2$. The data is generated following the assumptions of Corollary 2. We consider the following parameter values: $p = 100, 500, 1000, 1500, 2000, 3000$, $c = 1, 2, 5, 10$, $\sigma_\varepsilon^2 = 1, 2, 4, 10$. Due to the large number of resulting figures, we present only selected cases in Figures 1–9, which demonstrate certain disparities in greater detail. Figures show the empirical cumulative distribution function (CDF) and the empirical probability density function (PDF), together with the limiting CDF and PDF of $\mathcal{N}(0, s^2)$ by (80) for different parameter combinations. In addition, we present the corresponding Q-Q plots in order to inspect the tails of the resulting distributions in greater detail.

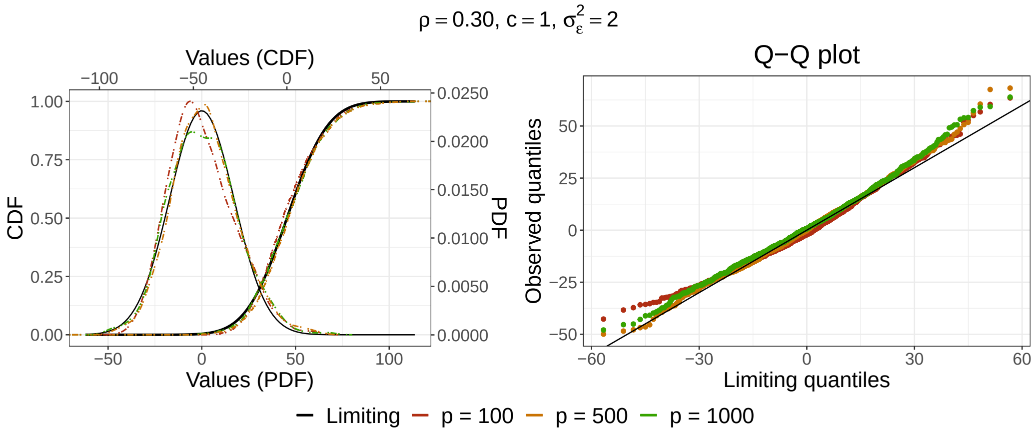

Figure 1. Comparison of the PDF and CDF (**left**) and the corresponding Q-Q plots (**right**) after 1000 replications from the Monte Carlo simulation of the statistic (80) with the limiting distribution $\mathcal{N}(0, s^2)$ by the Corollary 2 (in black) for $\varrho = 0.3$, $c = 1$, $\sigma_\varepsilon^2 = 2$ and $p = 100, 500, 1000$.

We find that for relatively small values of ϱ, the observed distribution of the statistic is fairly close to the limiting distribution even for small values of p, n and larger σ_ε^2, c (see, e.g., Figures 1–4). However, slower convergence is more evident with increasing values of ϱ. Furthermore, for moderate values of $\varrho, c, \sigma_\varepsilon^2$, only with larger values of p we observe adequate convergence towards the limiting distribution (see Figures 5 and 6). Similar behaviour is observed when the relation between the parameters $\varrho, c, \sigma_\varepsilon^2$ is appropriately controlled: e.g., in Figure 7, we see comparable results to those presented by Figure 6, where the effect of the increase in parameter value ϱ is countered by a smaller value of σ_ε^2. Alternatively, analogous effects can be achieved when reducing the values of c, instead.

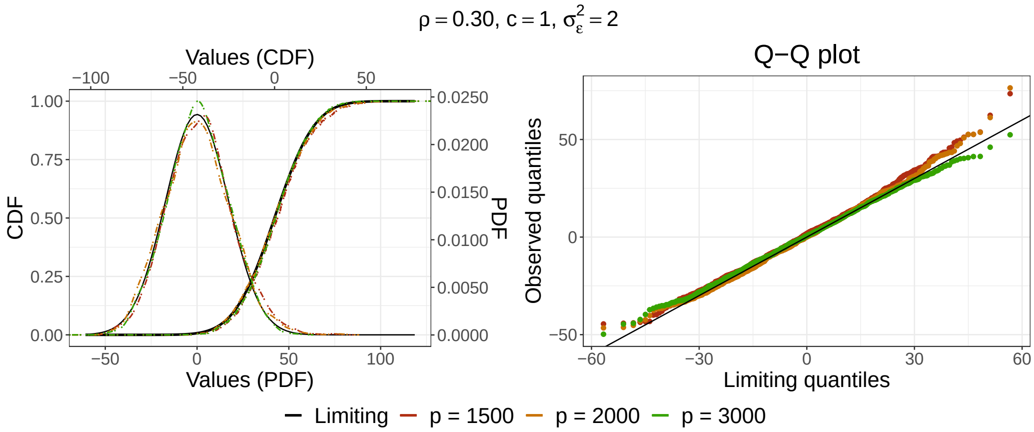

Figure 2. Comparison of the PDF and CDF (**left**) and the corresponding Q-Q plots (**right**) after 1000 replications from the Monte Carlo simulation of the statistic (80) with the limiting distribution $\mathcal{N}(0, s^2)$ by the Corollary 2 (in black) for $\varrho = 0.3$, $c = 1$, $\sigma_\varepsilon^2 = 2$ and $p = 1500, 2000, 3000$.

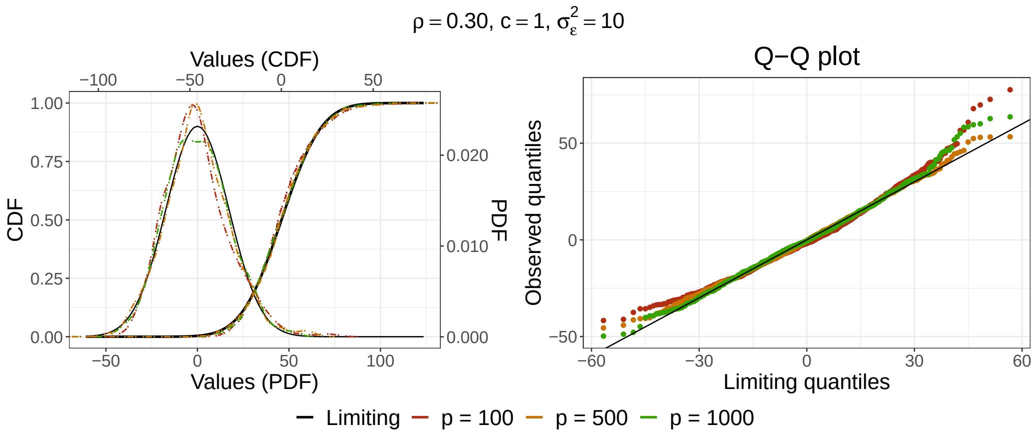

Figure 3. Comparison of the PDF and CDF (**left**) and the corresponding Q-Q plots (**right**) after 1000 replications from the Monte Carlo simulation of the statistic (80) with the limiting distribution $\mathcal{N}(0, s^2)$ by the Corollary 2 (in black) for $\varrho = 0.3$, $c = 1$, $\sigma_\varepsilon^2 = 10$ and $p = 100, 500, 1000$.

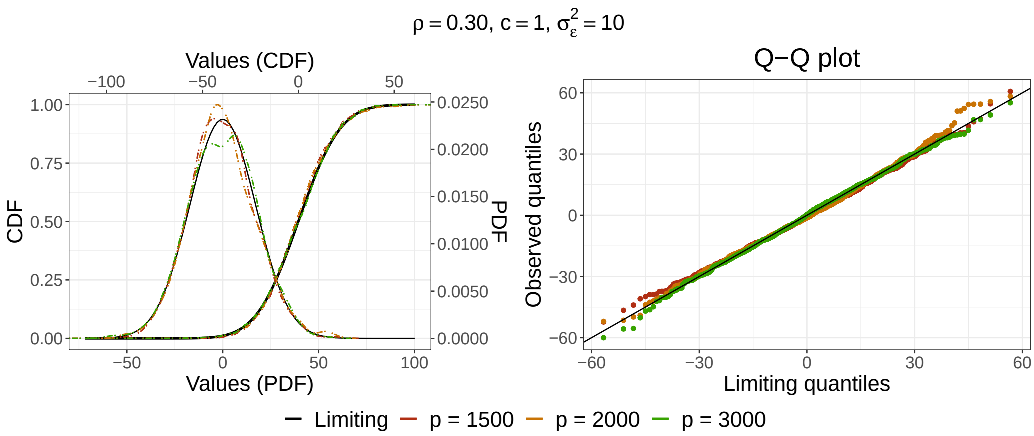

Figure 4. Comparison of the PDF and CDF (**left**) and the corresponding Q-Q plots (**right**) after 1000 replications from the Monte Carlo simulation of the statistic (80) with the limiting distribution $\mathcal{N}(0, s^2)$ by the Corollary 2 (in black) for $\varrho = 0.3$, $c = 1$, $\sigma_\varepsilon^2 = 10$ and $p = 1500, 2000, 3000$.

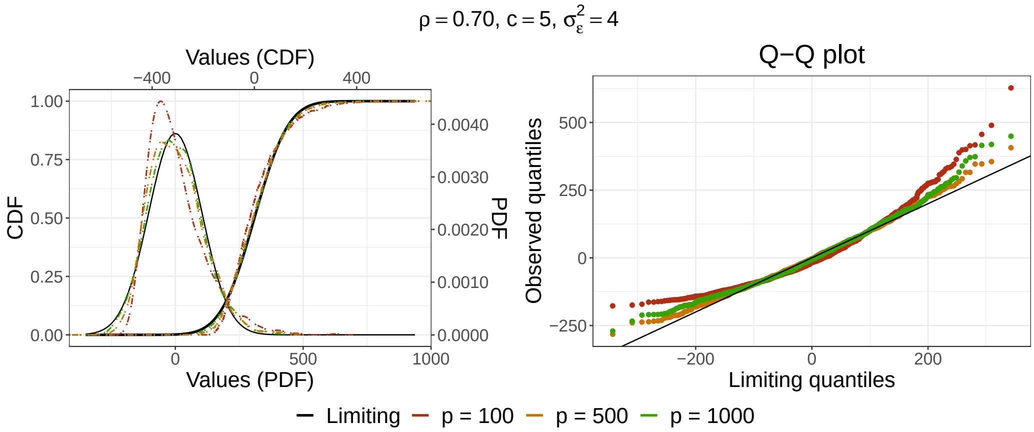

Figure 5. Comparison of the PDF and CDF (**left**) and the corresponding Q-Q plots (**right**) after 1000 replications from the Monte Carlo simulation of the statistic (80) with the limiting distribution $\mathcal{N}(0, s^2)$ by the Corollary 2 (in black) for $\varrho = 0.7$, $c = 5$, $\sigma_\varepsilon^2 = 4$ and $p = 100, 500, 1000$.

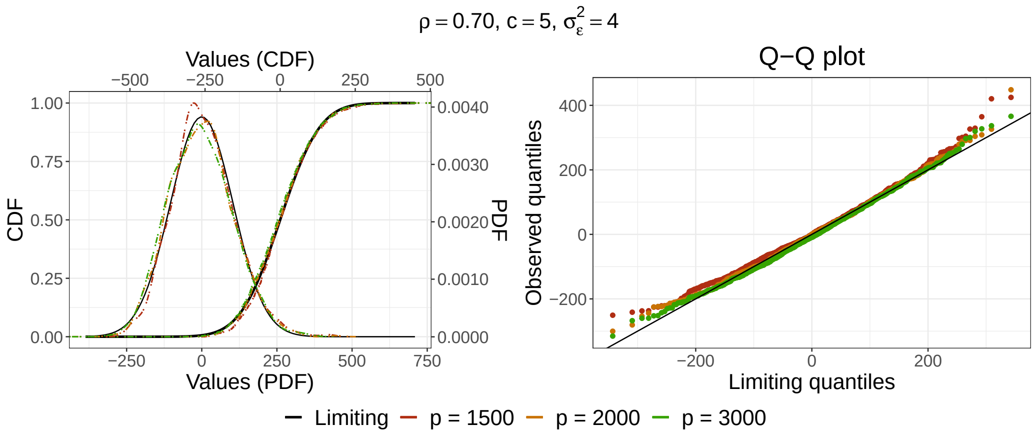

Figure 6. Comparison of the PDF and CDF (**left**) and the corresponding Q-Q plots (**right**) after 1000 replications from the Monte Carlo simulation of the statistic (80) with the limiting distribution $\mathcal{N}(0, s^2)$ by the Corollary 2 (in black) for $\varrho = 0.7$, $c = 5$, $\sigma_\varepsilon^2 = 4$ and $p = 1500, 2000, 3000$.

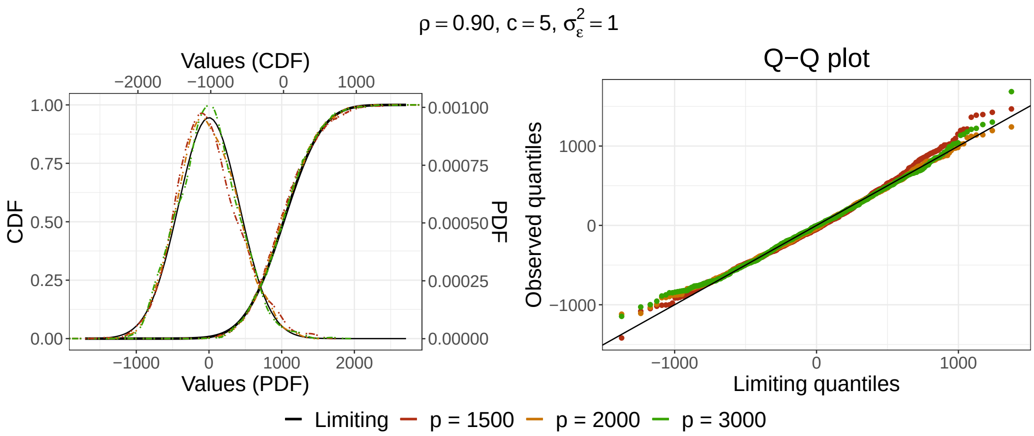

Figure 7. Comparison of the PDF and CDF (**left**) and the corresponding Q-Q plots (**right**) after 1000 replications from the Monte Carlo simulation of the statistic (80) with the limiting distribution $\mathcal{N}(0, s^2)$ by the Corollary 2 (in black) for $\varrho = 0.9$, $c = 5$, $\sigma_\varepsilon^2 = 1$ and $p = 1500, 2000, 3000$.

Finally, slow convergence is observed for large values of $\varrho, c, \sigma_\varepsilon^2$, as expected (see Figures 8 and 9). In such cases, the simulation results suggest that even larger values of p, n would be needed for more accurate results.

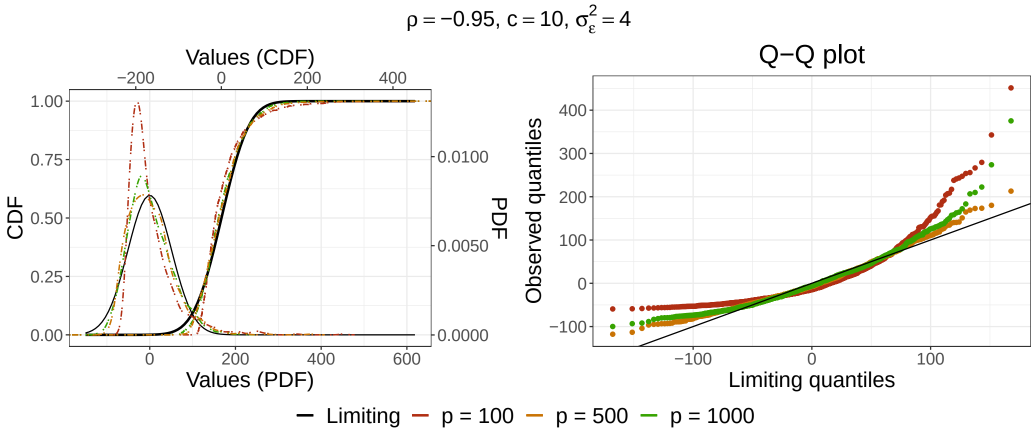

Figure 8. Comparison of the PDF and CDF (**left**) and the corresponding Q-Q plots (**right**) after 1000 replications from the Monte Carlo simulation of the statistic (80) with the limiting distribution $\mathcal{N}(0, s^2)$ by the Corollary 2 (in black) for $\varrho = -0.95$, $c = 10$, $\sigma_\varepsilon^2 = 4$ and $p = 100, 500, 1000$.

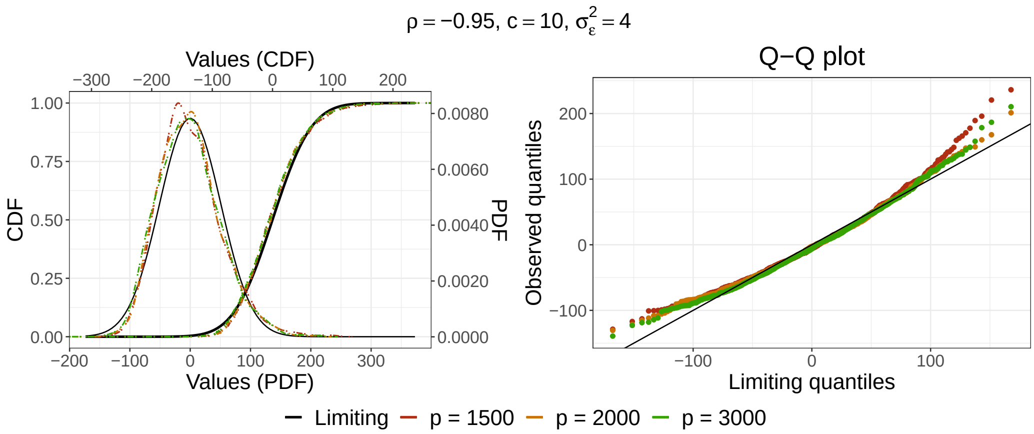

Figure 9. Comparison of the PDF and CDF (**left**) and the corresponding Q-Q plots (**right**) after 1000 replications from the Monte Carlo simulation of the statistic (80) with the limiting distribution $\mathcal{N}(0, s^2)$ by the Corollary 2 (in black) for $\varrho = -0.95$, $c = 10$, $\sigma_\varepsilon^2 = 4$ and $p = 1500, 2000, 3000$.

7. Discussion

In this paper, we consider a specific KMS covariance structure due to its attractive properties and wide application possibilities for working with real world datasets. Moreover, our results could be extended further by considering a wider family of Toeplitz covariance structures. For instance, under specific constraints, one could employ the approaches proposed in [3] in order to extend the application of our results towards more complex covariance structures of the data.

Furthermore, for future work, it would be interesting to expand and examine the results by removing the assumption of independence between the observations X_i, $i = 1, \ldots, n$.

Finally, in this paper we have established both the exact and the asymptotic distributions of the statistic $\|\mathbb{X}'Y\|_2^2$ (see (34) and (8), (10)). Both distributions could be used for estimating β, σ_ε^2 or related measures (e.g., by applying the method of moments or maximum likelihood estimation) in future research. Such research direction could open up interesting avenues when compared with popular LASSO type methods in high-dimensional linear regression. Similar approach is taken by [10], who construct maximum likelihood estimators for the signal strength $\|\beta\|_2^2$ in a high-dimensional regression context. Note that the results by [10] are achieved under certain strong restrictions, which are consistent with the related literature (see, e.g., [7,41,42]). In our case, we impose weaker assumptions; therefore, both our asymptotic or exact results could be used in order to extend the approaches in the aforementioned literature.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math10101657/s1, Proof of Lemmas 4, A2, A5, Proof of result (A9) of Lemma 5(ii), Proof of results (A13)–(A14) of Lemma 5(iii).

Author Contributions: Conceptualization, S.J., R.L.; methodology, S.J., R.L.; investigation, S.J., R.L.; writing—original draft preparation, S.J., R.L.; writing—review and editing, S.J., R.L.; visualization, S.J., R.L. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by grant No. S-MIP-20-16 from the Research Council of Lithuania.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the anonymous Referees for their very constructive and detailed comments and suggestions on the first version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Throughout the proofs we use the notation C to mark generic constants, the specific values of which can change from line to line.

Appendix A.1. Technical Lemmas

Lemma A1. *Assume that* $|\varrho| < 1$. *Then,*

$$\int_0^{\varrho} \frac{\log(1-\varrho x)}{x(1-x)}\,\mathrm{d}x \;=\; -\frac{1}{2}\big(\mathrm{Li}_2(\varrho^2) + \log^2(1-\varrho)\big),$$

where Li_2 *denotes the real dilogarithm function. (Recall, that for* $\varrho < 0$*, by* $\int_0^{\varrho}$ *we denote* $-\int_{\varrho}^{0}$*.)*

Proof. Write,

$$\int_0^{\varrho} \frac{\log(1-\varrho x)}{x(1-x)}\,\mathrm{d}x \;=\; \int_0^{\varrho} \frac{\log(1-\varrho x)}{x}\,\mathrm{d}x + \int_0^{\varrho} \frac{\log(1-\varrho x)}{1-x}\,\mathrm{d}x.$$

By (72), we have

$$\int_0^{\varrho} \frac{\log(1-\varrho x)}{x}\,\mathrm{d}x \;=\; -\mathrm{Li}_2(\varrho^2). \tag{A1}$$

It remains to show that

$$\int_0^{\varrho} \frac{\log(1-\varrho x)}{1-x}\,\mathrm{d}x \;=\; \frac{1}{2}\big(\mathrm{Li}_2(\varrho^2) - \log^2(1-\varrho)\big). \tag{A2}$$

Indeed, by substitution $v = \varrho - \varrho x$, we have

$$\begin{aligned}\int_0^{\varrho} \frac{\log(1-\varrho x)}{1-x}\,\mathrm{d}x &= \int_{\varrho-\varrho^2}^{\varrho} \frac{\log(1-\varrho+v)}{v}\,\mathrm{d}v \\ &= \int_{\varrho-\varrho^2}^{\varrho} \frac{\log(1+\frac{v}{1-\varrho})}{v}\,\mathrm{d}v \;-\; \log^2(1-\varrho).\end{aligned} \tag{A3}$$

Further, by substitution $w = -\frac{v}{1-\varrho}$, we have

$$\begin{aligned}\int_{\varrho-\varrho^2}^{\varrho} \frac{\log(1+\frac{v}{1-\varrho})}{v}\,dv &= -\int_{-\frac{\varrho}{1-\varrho}}^{-\varrho} \frac{\log(1-w)}{w}\,dw \\ &= \mathrm{Li}_2(-\varrho) - \mathrm{Li}_2\left(-\frac{\varrho}{1-\varrho}\right) \\ &= \mathrm{Li}_2(-\varrho) + \mathrm{Li}_2(\varrho) + \frac{1}{2}\log^2(1-\varrho) \qquad \text{(A4)} \\ &= \frac{1}{2}\left(\mathrm{Li}_2(\varrho^2) + \log^2(1-\varrho)\right), \qquad \text{(A5)}\end{aligned}$$

where for (A4) and (A5) we apply the easily verifyable identities (see, e.g., [43]):

$$\begin{aligned}\mathrm{Li}_2\left(\frac{x}{x-1}\right) &= -\mathrm{Li}_2(x) - \frac{1}{2}\log^2(1-x), \quad x<1, \\ \mathrm{Li}_2(x) + \mathrm{Li}_2(-x) &= \frac{1}{2}\mathrm{Li}_2(x^2), \quad |x|<1.\end{aligned}$$

Thus, (A3) and (A5) imply (A2), which concludes the proof. □

Lemma A2. *Assume that $\sum_{j=1}^{\infty}\beta_j^2 < \infty$ and $|\varrho| < 1$. Then, the following inequalities hold:*

(i) $$\left|\sum_{l=p+1}^{\infty}\sum_{l'=l+1}^{\infty}\beta_l\beta_{l'}\varrho^{l'-l}\right| \le C\sum_{l=p+1}^{\infty}\beta_l^2.$$

(ii) $$\left|\sum_{l=p+1}^{\infty}\sum_{l'=l+1}^{\infty}\beta_l\beta_{l'}\varrho^{l'-l}(l'-l)\right| \le C\sum_{l=p+1}^{\infty}\beta_l^2.$$

(iii) $$\left|\sum_{l=1}^{p}\sum_{l'=p+1}^{\infty}\beta_l\beta_{l'}\varrho^{l'-l}\right| \le C\sum_{l=p+1}^{\infty}\beta_l^2.$$

(iv) $$\left|\sum_{l=1}^{p}\sum_{l'=p+1}^{\infty}\beta_l\beta_{l'}\varrho^{l'+l}\right| \le C\sum_{l=p+1}^{\infty}\beta_l^2.$$

Proof. See the proof in Supplementary Materials, Section S2. □

Lemma A3. *Assume that $\sup_{j\ge1}|\beta_j|j^{\alpha} < \infty$, $\alpha > 1/2$ and that $|\varrho| < 1$. Then,*

$$\left|\sum_{j=1}^{p}\beta_j\varrho^{p-j}\right| = o(p^{-1/4}).$$

Proof. We have

$$\begin{aligned}\left|\sum_{j=1}^{p}\beta_j\varrho^{p-j}\right| &\le \sum_{j=1}^{\lfloor\sqrt{p}\rfloor}|\beta_j||\varrho|^{p-j} + \sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\beta_j||\varrho|^{p-j} \\ &\le \sup_{j\ge1}|\beta_j|\sum_{j=1}^{\lfloor\sqrt{p}\rfloor}|\varrho|^{p-j} + p^{-\alpha/2}\sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\beta_j|p^{\alpha/2}|\varrho|^{p-j} \\ &\le \sup_{j\ge1}|\beta_j|\sum_{j=1}^{\lfloor\sqrt{p}\rfloor}|\varrho|^{p-j} + p^{-\alpha/2}\sup_{j\ge1}|\beta_j|j^{\alpha}\sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\varrho|^{p-j} \\ &\le C\left(\sum_{j=1}^{\lfloor\sqrt{p}\rfloor}|\varrho|^{p-j} + p^{-\alpha/2}\sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\varrho|^{p-j}\right) \\ &\le C\left(|\varrho|^{p-\lfloor\sqrt{p}\rfloor} + p^{-\alpha/2}\right).\end{aligned}$$

Here we used the fact that $\sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\varrho|^{p-j} \to (1-|\varrho|)^{-1} < \infty$. Thus,

$$p^{1/4}\left|\sum_{j=1}^{p}\beta_j\varrho^{p-j}\right| \le C\left(p^{1/4}|\varrho|^{p-\lfloor\sqrt{p}\rfloor} + p^{\frac{1}{4}-\frac{\alpha}{2}}\right) \to 0. \tag{A6}$$

□

Remark A1. *The assumption* $\sup_{j\ge1}|\beta_j|j^{\alpha} < \infty$*, for* $\alpha > 1/2$*, implies that* $\sum_{j=1}^{\infty}\beta_j^2 < \infty$*:*

$$\sum_{j=1}^{\infty}\beta_j^2 = \sum_{j=1}^{\infty}\beta_j^2 j^{2\alpha}j^{-2\alpha} \le \sup_{j\ge1}\beta_j^2 j^{2\alpha}\sum_{k=1}^{\infty}k^{-2\alpha} < \infty.$$

Lemma A4. *Assume that the assumptions of Theorem 1 hold. Then,*

$$\kappa_{2,p} = o(p).$$

Proof. Observe, that

$$\begin{aligned}\kappa_{2,p} = \sum_{k=1}^{p}\left(\sum_{l=1}^{p}\beta_l\varrho^{|k-l|}\right)^2 &= \sum_{k=1}^{p}\sum_{l_1,l_2=1}^{p}\beta_{l_1}\beta_{l_2}\varrho^{|k-l_1|+|k-l_2|} \\ &\le \sum_{l_1,l_2=1}^{p}|\beta_{l_1}||\beta_{l_2}|\sum_{k=1}^{p}|\varrho|^{|k-l_1|+|k-l_2|} \\ &\le C\left(\sum_{l=1}^{p}|\beta_{l_1}|\right)^2 \qquad (A7) \\ &= o(p)\end{aligned}$$

where (A7) follows from (S9). Meanwhile, $\sum_{l=1}^{p}|\beta_{l_1}| = o(p^{1/2})$, since

$$\begin{aligned}\sum_{l=1}^{p}|\beta_l| &= \sum_{l=1}^{\lfloor p^{1/2}\rfloor}|\beta_l| + \sum_{l=\lfloor p^{1/2}\rfloor+1}^{p}|\beta_l| \\ &\le p^{1/4}\left(\sum_{l=1}^{\infty}\beta_l^2\right)^{1/2} + p^{1/2}\left(\sum_{l=\lfloor p^{1/2}\rfloor+1}^{\infty}\beta_l^2\right)^{1/2} = o(p^{1/2}).\end{aligned}$$

□

Lemma A5. *Assume that* $\sum_{j=1}^{\infty}\beta_j^2 < \infty$ *and* $|\varrho| < 1$*. Define* $\theta_k^{(p)} = \sum_{j=1}^{p}\beta_j\varrho^{|k-j|}$*. Then,*

$$\left|\sum_{i,j,k=1}^{p}\left(\varrho^{|i-j|}+\theta_i^{(p)}\theta_j^{(p)}\right)\left(\varrho^{|i-k|}+\theta_i^{(p)}\theta_k^{(p)}\right)\left(\varrho^{|k-j|}+\theta_k^{(p)}\theta_j^{(p)}\right)\right| = o(p^{3/2}). \tag{A8}$$

Proof. See the proof in Supplementary Material, Section S3. □

Appendix A.2. Proof of Lemma 5

Here and throughout the proof we employ the notation as in Definition 3.
(i) Note that, by (65) and (67), we have

$$\kappa_{1,p} = \sum_{k=1}^{p}\beta_k^2 + 2\sum_{k=2}^{p}\sum_{l=1}^{k-1}\beta_k\beta_l\varrho^{k-l} \to \beta(1) + 2b_1(\varrho) \quad \text{as } p\to\infty.$$

(ii) Write

$$\kappa_{2,p} = \sum_{l=1}^{p}\sum_{k=1}^{p}\beta_l^2\varrho^{2|k-l|} + 2\sum_{l'>l}\sum_{k=1}^{p}\beta_l\beta_{l'}\varrho^{|k-l|}\varrho^{|k-l'|}.$$

From here, it can be seen that

$$\kappa_{2,p} \to \beta(1)\frac{1+\varrho^2}{1-\varrho^2} - \beta(\varrho^2)\frac{1}{1-\varrho^2} + 2\Big(b_1^{(1)}(\varrho) + b_1(\varrho)\frac{1+\varrho^2}{1-\varrho^2} - b_2(\varrho)\frac{1}{1-\varrho^2}\Big). \quad \text{(A9)}$$

Technical details of the proof of (A9) are presented in Supplementary Materials, Section S4.

(iii) Consider

$$\kappa_{3,p} = \sum_{l=1}^{p}\beta_l^2 J_1(l) + 2\sum_{l<l'}\beta_l\beta_{l'}J_2(l,l'), \quad \text{(A10)}$$

where

$$J_1(l) := \sum_{k,k'=1}^{p}\varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l|}\mathbf{1}_{\{l=l'\}}, \quad \text{(A11)}$$

$$J_2(l,l') := \sum_{k,k'=1}^{p}\varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbf{1}_{\{l<l'\}}. \quad \text{(A12)}$$

Then, as $p \to \infty$, using the notation in Definition 3, we have that

$$\sum_{l=1}^{p}\beta_l^2 J_1(l) \to \beta(1)\frac{1+4\varrho^2+\varrho^4}{(1-\varrho^2)^2} - \beta(\varrho^2)\frac{1+3\varrho^2}{(1-\varrho^2)^2} - \frac{2}{1-\varrho^2}\beta^{(1)}(\varrho^2), \quad \text{(A13)}$$

and

$$\sum_{l'>l}\beta_l^2 J_2(l,l') \to \frac{1}{2(1-\varrho^2)^2}\big(b^{(2)}(\varrho)(1-\varrho^2)^2 + 3b_1^{(1)}(\varrho)(1-\varrho^4) + 2b_1(\varrho)(1+4\varrho^2+\varrho^4) - 2b_2^{(1)}(\varrho)(1-\varrho^2) - 2b_2(\varrho)(1+3\varrho^2)\big). \quad \text{(A14)}$$

Technical details of the proof of (A13)–(A14) are omitted here and presented in the Supplementary Materials, Section S5. This concludes the proof.

References

1. Kac, M.; Murdock, W.; Szegő, G. On the eigen-values of certain Hermitian forms. *J. Linear Ration. Mech. Anal.* **1953**, *2*, 767–800.
2. Fikioris, G. Spectral properties of Kac–Murdock–Szegö matrices with a complex parameter. *Linear Algebra Appl.* **2018**, *553*, 182–210.
3. Yang, Y.; Zhou, J.; Pan, J. Estimation and optimal structure selection of high-dimensional Toeplitz covariance matrix. *J. Multivar. Anal.* **2021**, *184*, 104739, https://doi.org/10.1016/j.jmva.2021.104739.
4. Liang, K.Y.; Zeger, S.L. Longitudinal data analysis using generalized linear models. *Biometrika* **1986**, *73*, 13–22.
5. Rangan, S. Generalized approximate message passing for estimation with random linear mixing. In Proceedings of the 2011 IEEE International Symposium on Information Theory Proceedings, St. Petersburg, Russia, 31 July–5 August 2011; pp. 2168–2172.
6. Vila, J.P.; Schniter, P. Expectation-maximization Gaussian-mixture approximate message passing. *IEEE Trans. Signal Process.* **2013**, *61*, 4658–4672.
7. Dicker, L.H. Variance estimation in high-dimensional linear models. *Biometrika* **2014**, *101*, 269–284, https://doi.org/10.1093/biomet/ast065.
8. Diggle, P.J.; Giorgi, E. *Model-Based Geostatistics for Global Public Health: Methods and Applications*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2019.
9. Patil, A.R.; Kim, S. Combination of ensembles of regularized regression models with resampling-based lasso feature selection in high dimensional data. *Mathematics* **2020**, *8*, 110.
10. Dicker, L.H.; Erdogdu, M.A. Maximum likelihood for variance estimation in high-dimensional linear models. In Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, Cadiz, Spain, 9–11 May 2016.
11. Carpentier, A.; Verzelen, N. Adaptive estimation of the sparsity in the Gaussian vector model. *Ann. Stat.* **2019**, *47*, 93–126.

12. Carpentier, A.; Verzelen, N. Optimal sparsity testing in linear regression model. *Bernoulli* **2021**, *27*, 727–750.
13. Gaunt, R.E. Rates of Convergence of Variance-Gamma Approximations via Stein's Method. Ph.D. Thesis, The Queen's College, University of Oxford, Oxford, UK, 2013.
14. Gaunt, R.E. Variance-Gamma approximation via Stein's method. *Electron. J. Probab.* **2014**, *19*, 1–33.
15. Gaunt, R.E. Products of normal, beta and gamma random variables: Stein operators and distributional theory. *Braz. J. Probab. Stat.* **2018**, *32*, 437–466.
16. Gaunt, R.E. A note on the distribution of the product of zero-mean correlated normal random variables. *Stat. Neerl.* **2019**, *73*, 176–179, https://doi.org/10.1111/stan.12152.
17. Ing, C.K. Model selection for high-dimensional linear regression with dependent observations. *Ann. Stat.* **2020**, *48*, 1959–1980, https://doi.org/10.1214/19-AOS1872.
18. Cha, J.; Chiang, H.D.; Sasaki, Y. Inference in high-dimensional regression models without the exact or L^p sparsity. *arXiv* **2021**, arXiv:2108.09520.
19. Shibata, R. Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process. *Ann. Stat.* **1980**, *8*, 147–164.
20. Ing, C.K. Accumulated Prediction Errors, Information Criteria and Optimal Forecasting for Autoregressive Time Series. *Ann. Stat.* **2007**, *35*, 1238–1277.
21. Belloni, A.; Chen, D.; Chernozhukov, V.; Hansen, C. Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica* **2012**, *80*, 2369–2429, https://doi.org/10.3982/ECTA9626.
22. Javanmard, A.; Montanari, A. Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **2014**, *15*, 2869–2909.
23. Zhang, C.H.; Zhang, S.S. Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2014**, *76*, 217–242.
24. Caner, M.; Kock, A.B. Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative Lasso. *J. Econom.* **2018**, *203*, 143–168.
25. Belloni, A.; Chernozhukov, V.; Chetverikov, D.; Hansen, C.; Kato, K. High-dimensional econometrics and regularized GMM. *arXiv* **2018**, arXiv:1806.01888.
26. Gold, D.; Lederer, J.; Tao, J. Inference for high-dimensional instrumental variables regression. *J. Econom.* **2020**, *217*, 79–111.
27. Ning, Y.; Peng, S.; Tao, J. Doubly Robust Semiparametric Difference-in-Differences Estimators with High-Dimensional Data. *arXiv* **2020**, arXiv:2009.03151.
28. Guo, Z.; Ćevid, D.; Bühlmann, P. Doubly Debiased Lasso: High-Dimensional Inference under Hidden Confounding. *arXiv* **2021**, arXiv:2004.03758.
29. Dai, Z.; Li, T.; Yang, M. Forecasting stock return volatility: The role of shrinkage approaches in a data-rich environment. *J. Forecast.* **2021**, 1–17, https://doi.org/10.1002/for.2841.
30. Dai, Z.; Zhu, H.; Zhang, X. Dynamic spillover effects and portfolio strategies between crude oil, gold and Chinese stock markets related to new energy vehicle. *Energy Econom.* **2022**, *109*, 105959.
31. Dai, Z.; Zhu, H. Time-varying spillover effects and investment strategies between WTI crude oil, natural gas and Chinese stock markets related to belt and road initiative. *Energy Econ.* **2022**, *108*, 105883.
32. Sánchez García, J.; Cruz Rambaud, S. Machine Learning Regularization Methods in High-Dimensional Monetary and Financial VARs. *Mathematics* **2022**, *10*, 877.
33. Yi, J.; Tang, N. Variational Bayesian inference in high-dimensional linear mixed models. *Mathematics* **2022**, *10*, 463.
34. Madan, D.B.; Carr, P.P.; Chang, E.C. The Variance Gamma Process and Option Pricing. *Rev. Financ.* **1998**, *2*, 79–105.
35. Kotz, S.; Kozubowski, T.; Podgórski, K. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*; Birkhäuser: Boston, MA, USA, 2001.
36. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288.
37. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429.
38. Meinshausen, N. Relaxed Lasso. *Comput. Stat. Data Anal.* **2007**, *52*, 374–393, https://doi.org/10.1016/j.csda.2006.12.019.
39. Belloni, A.; Chernozhukov, V.; Wang, L. Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **2011**, *98*, 791–806.
40. Morris, R. The Dilogarithm Function of a Real Argument. *Math. Comput.* **1979**, *33*, 778–787.
41. Bayati, M.; Erdogdu, M.A.; Montanari, A. Estimating lasso risk and noise level. In Proceedings of the Advances in Neural Information Processing Systems: 27th Annual Conference on Neural Information, Processing Systems 2013, Lake Tahoe, NV, USA, 5–10 December 2013; Volume 26.
42. Janson, L.; Barber, R.F.; Candes, E. EigenPrism: Inference for high dimensional signal-to-noise ratios. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2017**, *79*, 1037–1065.
43. Maximon, L.C. The dilogarithm function for complex argument. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **2003**, *459*, 2807–2819.