Journal of Multivariate Analysis 164 (2018) 22–39

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Asymptotic normality of quadratic forms with random vectors of increasing dimension

Hanxiang Peng [a], Anton Schick [b,*]

[a] Indiana University Purdue University at Indianapolis, Department of Mathematical Sciences Indianapolis, IN 46202-3267, USA
[b] Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902-6000, USA

ARTICLE INFO

Article history:
Received 28 February 2017
Available online 13 November 2017

AMS subject classifications:
G2E20
G2H15

Keywords:
Chi-square test with increasing number of cells
Empirical likelihood
Equal marginals
Independence of components of high-dimensional normal random vectors
Lindeberg condition
Martingale central limit theorem

ABSTRACT

This paper provides sufficient conditions for the asymptotic normality of quadratic forms of averages of random vectors of increasing dimension and improves on conditions found in the literature. Such results are needed in applications of Owen's empirical likelihood when the number of constraints is allowed to grow with the sample size. Indeed, the results of this paper are already used in Peng and Schick (2013) for this purpose. We also demonstrate how our results can be used to obtain the asymptotic distribution of the empirical likelihood with an increasing number of constraints under contiguous alternatives. In addition, we discuss potential applications of our result. The first example focuses on a chi-square test with an increasing number of cells. The second example treats testing for the equality of the marginal distributions of a bivariate random vector. The third example generalizes a result of Schott (2005) by showing that a standardized version of his test for diagonality of the dispersion matrix of a normal random vector is asymptotically standard normal even if the dimension increases faster than the sample size. Schott's result requires the dimension and the sample size to be of the same order.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Let r_n be positive integers that tend to infinity with n. Let $\xi_{n,1}, \ldots, \xi_{n,n}$ be independent and identically distributed r_n-dimensional random vectors with mean $\mathrm{E}(\xi_{n,1}) = 0$ and dispersion matrix $V_n = \mathrm{E}(\xi_{n,1}\xi_{n,1}^\top)$. Let $|x|$ denote the Euclidean norm of a vector x. We are interested in the asymptotic behavior of $|\tilde{\xi}_n + \mu_n|^2$ with μ_n an r_n-dimensional vector and $\tilde{\xi}_n$ the r_n-dimensional random vector defined by

$$\tilde{\xi}_n = n^{-1/2} \sum_{j=1}^{n} \xi_{n,j}.$$

More precisely, we are looking for conditions that imply the asymptotic normality

$$\frac{|\tilde{\xi}_n + \mu_n|^2 - |\mu_n|^2 - \mathrm{tr}(V_n)}{\sqrt{2\,\mathrm{tr}(V_n^2)}} \rightsquigarrow \mathcal{N}(0, 1). \tag{1}$$

* Corresponding author.
E-mail addresses: hanxpeng@iupui.edu (H. Peng), anton@math.binghamton.edu (A. Schick).

https://doi.org/10.1016/j.jmva.2017.11.002
0047-259X/© 2017 Elsevier Inc. All rights reserved.

Of special interest is the case when μ_n is the zero vector and V_n is idempotent with rank q_n tending to infinity. Then (1) simplifies to

$$\frac{|\tilde{\xi}_n|^2 - q_n}{\sqrt{2q_n}} \rightsquigarrow \mathcal{N}(0, 1). \tag{2}$$

In particular, if μ_n is the zero vector and V_n equals I_{r_n}, the $r_n \times r_n$ identity matrix, then (2) becomes

$$\frac{|\tilde{\xi}_n|^2 - r_n}{\sqrt{2r_n}} \rightsquigarrow \mathcal{N}(0, 1). \tag{3}$$

Such results are needed to obtain the asymptotic behavior of the likelihood ratio statistic in exponential families of increasing dimensions and to study the behavior of Owen's empirical likelihood when the data dimension is allowed to increase with the sample size. The former was done by Portnoy [15], who proved (3) under the assumption that the sixth moments of the coordinates of $\xi_{n,1}$ are uniformly bounded. The latter was studied in [9] by Hjort, McKeague and Van Keilegom, who relied on Portnoy's result and in [3] by Chen, Peng and Qin, who relied on results and structural assumptions of [1]. We are interested in verifying (1) under weaker moment assumptions than used by these authors.

The results of this paper play key roles in developing the theory in [14]. Theorem 2 is used there to improve on results in [9], to extend these further to allow for infinitely many constraints that depend on nuisance parameters, and to obtain results similar to those in [3] without their structural assumptions. The results in Section 4 are also used in [14] to address the asymptotic distribution of the empirical likelihood with an increasing number of constraints under contiguous alternatives; see their Remark 2.2 and Remark 7.3. While these applications were the main motivation for the present paper, our results go well beyond this. Several applications to testing problems will be discussed here. The first one focuses on a chi-square test with an increasing number of cells. The second one treats testing for the equality of the marginal distributions of a bivariate random vector. The third one shows that Schott's [19] test for diagonality of the dispersion matrix of a normal random vector remains valid if the dimension of the random vector increases at a faster rate than the sample size, see Theorem 4 in Section 8.

The left-hand side of (1) can be written as the sum $T_{n,1} + T_{n,2} + T_{n,3}$ of the terms

$$T_{n,1} = \frac{1}{n}\sum_{j=1}^{n} \frac{|\xi_{n,j}|^2 - \mathrm{tr}(V_n)}{\sqrt{2\mathrm{tr}(V_n^2)}}, \quad T_{n,2} = \frac{2\tilde{\xi}_n^\top \mu_n}{\sqrt{2\mathrm{tr}(V_n^2)}} \quad \text{and} \quad T_{n,3} = \frac{2}{n}\sum_{1\le i<j\le n} \frac{\xi_{n,i}^\top \xi_{n,j}}{\sqrt{2\mathrm{tr}(V_n^2)}}.$$

The leading term in this expansion is $T_{n,3}$. Indeed, we will give conditions that let us show that $T_{n,1}$ and $T_{n,2}$ converge to zero in probability and let us use a martingale central limit theorem to establish the asymptotic normality of $T_{n,3}$. In contrast to Portnoy [15] who used a martingale central limit theorem to deal with the sum $T_{n,1}+T_{n,3}$, we do so only for the term $T_{n,3}$ and use a different argument to control $T_{n,1}$. Our approach is in the spirit of Guttorp and Lockhart [6], who treated quadratic forms with a random vector of independent components and not necessarily an average. They separated on-diagonal elements and off-diagonal elements into statistics similar to our $T_{n,1}$ and $T_{n,3}$, and treated these two statistics separately relying on work of Rotar [17,16,18]. The use of martingale arguments to deal with quadratic forms with weight matrices with vanishing diagonal is common; see, e.g., the recent work [5].

We achieve our goal by proving two central limit theorems. Both of our theorems require the conditions

$$\mu_n^\top V_n \mu_n = o\{\mathrm{tr}(V_n^2)\} \tag{C1}$$

and

$$\mathrm{tr}(V_n^4) = o\{\mathrm{tr}^2(V_n^2)\}. \tag{C2}$$

Condition (C1) implies that $T_{n,2}$ converges to zero in probability in view of the identity $\mathrm{E}(T_{n,2}^2) = 2\mu_n^\top V_n \mu_n / \mathrm{tr}(V_n^2)$. Condition (C2) is used in the application of the martingale central limit theorem to $T_{n,3}$. This condition already appeared in [4], where one can also find examples of dispersion matrices satisfying (C2). These authors combine (C2) with $\mathrm{tr}(V_n^2) \to \infty$ and the structural assumptions on their random vectors from [1] to derive asymptotic normality of some statistics arising in their tests for diagonality of high-dimensional dispersion matrices.

Of course, (C1) is always met if μ_n is the zero vector. Condition (C2) is met if V_n is idempotent with rank q_n tending to infinity. In particular, it is met if $V_n = I_{r_n}$. In this case, (C1) is implied by $|\mu_n|^2 = o(r_n)$. Condition (C2) is also met if V_n satisfies

$$0 < \lambda = \inf_n \inf_{|u|=1} u^\top V_n u \le \sup_n \sup_{|u|=1} u^\top V_n u = \Lambda < \infty. \tag{4}$$

Indeed, in this case we have $\mathrm{tr}(V_n^4) \le \Lambda^4 r_n$ and $\mathrm{tr}(V_n^2) \ge \lambda^2 r_n$. Under (4), (C1) is implied by $|\mu_n|^2 = o(r_n)$. The sufficiency of (4) for (C2) was already observed in [4]. Finally, in view of the inequality $\mathrm{tr}(V_n^4) \le \rho_n^2 \mathrm{tr}(V_n^2)$ with ρ_n the largest eigenvalue of V_n, a sufficient condition for (C2) is $\rho_n^2 = o\{\mathrm{tr}(V_n^2)\}$. In particular, (C2) holds if ρ_n is bounded and $\mathrm{tr}(V_n^2)$ tends to infinity. These sufficient conditions were used in an earlier version of the present paper, referred to as Peng and Schick (2012) in [14].

The first theorem uses the following growth conditions.

$$\mathrm{var}(|\xi_{n,1}|^2) = o\{n\,\mathrm{tr}(V_n^2)\}, \tag{5}$$

$$\mathrm{var}(|V_n^{1/2}\xi_{n,1}|^2) = o\{n\,\mathrm{tr}^2(V_n^2)\}, \tag{6}$$

$$\mathrm{E}(|\xi_{n,1}^\top \xi_{n,2}|^4) = o\{n^2\,\mathrm{tr}^2(V_n^2)\}. \tag{7}$$

Theorem 1. *Suppose (C1) and (C2) hold. Then (5)–(7) imply (1).*

The proof of the theorem shows that (C1) is only used to show $T_{n,2} = o_p(1)$ and (5) is only used to show $T_{n,1} = o_p(1)$. Thus we have the following corollary.

Corollary 1. *Suppose (C2) holds. Then (6) and (7) imply $T_{n,3} \rightsquigarrow \mathcal{N}(0,1)$.*

The growth conditions (5)–(7) are implied by

$$\mathrm{E}(|\xi_{n,1}|^4) = o\{n\,\mathrm{tr}(V_n^2)\}. \tag{8}$$

Thus we have the following corollary.

Corollary 2. *Suppose (C1) and (C2) hold. Then (8) implies (1).*

Suppose the components of $\xi_{n,1}$ are independent and identically distributed with mean 0, standard deviation σ, and finite positive fourth moment τ. Then we have $V_n = \sigma^2 I_{r_n}$ and calculate $\mathrm{tr}(V_n^2) = \sigma^4 r_n$, $\mathrm{tr}(V_n^4) = \sigma^8 r_n$ and

$$\mathrm{E}(|\xi_{n,1}|^4) = r_n\tau + r_n(r_n-1)\sigma^4.$$

Thus (C2) holds and (8) requires $r_n = o(n)$. On the other hand we have

$$\mathrm{var}(|\xi_{n,1}|^2) = r_n(\tau-\sigma^4), \quad \mathrm{var}(|V_n^{1/2}\xi_{n,1}|^2) = \sigma^4 r_n(\tau-\sigma^4) \quad \text{and} \quad \mathrm{E}(|\xi_{n,1}^\top\xi_{n,2}|^4) = r_n\tau^2 + 3r_n(r_n-1)\sigma^8.$$

Thus (5)–(7) impose no restrictions on the growth of r_n.

For an r-dimensional random vector X, the Cauchy–Schwarz inequality yields the bound

$$\mathrm{E}(|X|^4) = \mathrm{E}\left\{\left(\sum_{i=1}^r X_i^2\right)^2\right\} \le \mathrm{E}\left(r\sum_{j=1}^r X_i^4\right) = r\sum_{i=1}^r \mathrm{E}(X_i^4).$$

If V_n equals the identity matrix I_{r_n}, then $\mathrm{tr}(V_n^2) = \mathrm{tr}(V_n) = r_n$. Thus we obtain the following important special case.

Corollary 3. *Suppose $V_n = I_{r_n}$ and $\mu_n = 0$ hold. Then (3) is implied by $\sum_{i=1}^{r_n}\mathrm{E}(\xi_{n,1,i}^4) = o(n)$. In particular, if the condition*

$$\max_{1\le i\le r_n} \mathrm{E}(\xi_{n,1,i}^4) = O(1) \tag{9}$$

holds, then $r_n = o(n)$ implies (3).

The next theorem uses

$$\frac{1+\mathrm{tr}(V_n)}{\mathrm{tr}(V_n^2)} = O(1) \tag{C3}$$

and the Lindeberg condition,

$$\forall_{\epsilon>0} \quad L_n(\epsilon) = \mathrm{E}(|\xi_{n,1}|^2 \mathbf{1}\{|\xi_{n,1}| > \epsilon\sqrt{n}\}) \to 0 \tag{L}$$

to obtain the desired result.

Theorem 2. *Suppose (C1)–(C3) hold. Then (L) implies (1).*

Note that (C3) holds if V_n is idempotent or if V_n satisfies (4). Sufficient conditions for the Lindeberg condition can be derived from moment conditions on $|\xi_{n,1}|$. This is demonstrated in Section 3. There we also discuss the results in more detail and compare our results with those in the literature. A first example with simulations is given in Section 2 where a chi-square goodness-of-fit test is discussed and shown to be valid even if the number of cells increases almost as fast as the sample size. The results of a simulation study that addresses the quality of the normal approximation are presented in Section 4. There we looked at normal random vectors with moving average or autoregressive dispersion structure. Section 5 illustrates how our results can be used to give the asymptotic behavior under contiguous alternatives. In Section 6 we discuss potential applications of our results in generality. We illustrate concrete applications in the following two sections. A test for the equality of the marginal distributions of a bivariate random vector is treated in Section 7. In Section 8 a standardized version of Schott's [19] test for diagonality of a dispersion matrix of a normal random vector is shown to be asymptotically standard normal even if the dimension of the random vector grows at a faster rate than the sample size. Section 9 gives technical details needed in the proofs. The proofs of the theorems are given in Section 10.

Table 1
Simulated sizes of the test $\delta_{n,n/5}$ for selected values of n and α.

α	n									
	40	60	80	100	120	140	160	180	200	400
.10	.0910	.0941	.0985	.0926	.0936	.1040	.0998	.0949	.0956	.1028
.05	.0449	.0464	.0460	.0476	.0503	.0498	.0501	.0495	.0466	.0525
.01	.0090	.0109	.0098	.0102	.0096	.0100	.0108	.0096	.0098	.0114

2. A chi-square test with an increasing number of cells

Let $X_1, \ldots, X_n$ be independent random variables with common distribution function F. To test the null hypothesis that F equals a specified *continuous* distribution function F_0, we can use the test statistic

$$T_{n,r} = \sum_{i=1}^{r} \frac{(N_i - n/r)^2}{n/r},$$

where, for each $i \in \{1, \ldots, r\}$,

$$N_i = \sum_{j=1}^{n} \mathbf{1}\left\{\frac{i-1}{r} < F_0(X_j) \le \frac{i}{r}\right\},$$

and reject the null hypothesis if $T_{n,r}$ exceeds $\chi_{1-\alpha}(r-1)$, the $(1-\alpha)$-quantile of the chi-square distribution with $r-1$ degrees of freedom. This test has asymptotic size α. This follows from the fact that under the null hypothesis the test statistic is asymptotically chi-square with r degrees of freedom.

Can we let r grow with n and still maintain the asymptotic size of this test? The answer is yes. More precisely, we have the following result. The test

$$\delta_{n,r_n} = \mathbf{1}\{T_{n,r_n} > \chi_{1-\alpha}(r_n - 1)\}$$

has asymptotic size α as long as r_n tends to infinity at a rate slower than n, i.e., $r_n = o(n)$. The proof of this claim is based on the observation that a chi-square random variable with m degrees of freedom is approximately normal with mean m and variance $2m$ for large m. This result is a consequence of the central limit theorem and the fact that a chi-square random variable with m degrees of freedom has the same distribution as a sum of m independent chi-square random variables with one degree of freedom. Thus our claim can be verified by showing the asymptotic normality result

$$\frac{T_{n,r_n} - (r_n - 1)}{\sqrt{2(r_n - 1)}} \rightsquigarrow \mathcal{N}(0, 1). \tag{10}$$

We note that T_{n,r_n} equals $n|\tilde{\xi}_n|^2$ if we take ξ_{nj} to be the r_n-dimensional random vector whose ith coordinate is $\sqrt{r_n}\mathbf{1}\{i-1 < r_n F_0(X_j) \le i\} - 1/\sqrt{r_n}$. These random vectors are independent and identically distributed with mean vector 0 and dispersion matrix $V_n = I_{r_n} - 1/r_n J_{r_n}$, with J_r the $r \times r$ matrix with all its entries equal to 1, and satisfy $|\xi_{nj}|^2 = r_n - 1$ almost surely. The matrix V_n is idempotent with trace $r_n - 1$ and $\mathrm{E}(|\xi_{n1}|^4)$ equals $(r_n - 1)^2$. Consequently, the assumptions of Corollary 2 are met with $\mu_n = 0$ if $r_n = o(n)$ holds, and this corollary gives the desired (10) for such r_n.

We have run some simulations to assess this result. In the simulations F_0 was taken to be $\mathcal{U}(0, 1)$, i.e., the distribution function of a uniform random variable on $(0, 1)$, and r_n was chosen to be $n/5$. We generated 25,000 independent copies of $T_{n,n/5}$ for several choices of n. Fig. 1 gives Q–Q plots. These show that the chi-square approximation is quite good. Table 1 reports the simulated size of the test for three choices of α.

3. Discussion of the results

We begin by addressing sufficient conditions for the Lindeberg condition.

Remark 1. Let ψ be an increasing function on $[0, \infty)$ with $\psi(0) = 0$ and $\psi(x) \to \infty$ as $x \to \infty$. Then $L_n(\epsilon)$ is bounded by $\mathrm{E}\{|\xi_{n,1}|^2 \psi(|\xi_{n,1}|)\}/\psi(\epsilon\sqrt{n})$. The choice $\psi(x) = \ln(1+x)$ yields that (L) is implied by $\mathrm{E}\{|\xi_{n,1}|^2 \ln(1 + |\xi_{n,1}|)\} = o(\ln n)$. The choice $\psi(x) = x^\delta$, with $\delta > 0$, yields that the Lindeberg condition (L) holds whenever $\mathrm{E}(|\xi_{n,1}|^{2\alpha}) = o(n^{\alpha-1})$, for some $\alpha > 1$. In particular, if $\mathrm{E}(|\xi_{n,1}|^{2\alpha}) = O(r_n^\alpha)$ holds for some $\alpha > 1$, then (L) is implied by $r_n = o(n^{1-1/\alpha})$.

Let us now specialize our results to the case when μ_n is the zero vector and V_n is an idempotent matrix with rank q_n tending to infinity. In this case (C1)–(C3) hold and (1) simplifies to (2).

Corollary 4. *Suppose V_n is idempotent with rank q_n tending to infinity. Then the following are true.*

(a) *The growth conditions* $\mathrm{var}(|\xi_{n,1}|^2) = o(nq_n)$, $\mathrm{var}(|V_n^{1/2}\xi_{n,1}|^2) = o(nq_n^2)$ *and* $\mathrm{E}(|\xi_{n,1}^\top \xi_{n,2}|^4) = o(n^2 q_n^2)$ *imply* (2).
(b) *The moment condition* $\mathrm{E}(|\xi_{n,1}|^4) = o(nq_n)$ *implies* (2).

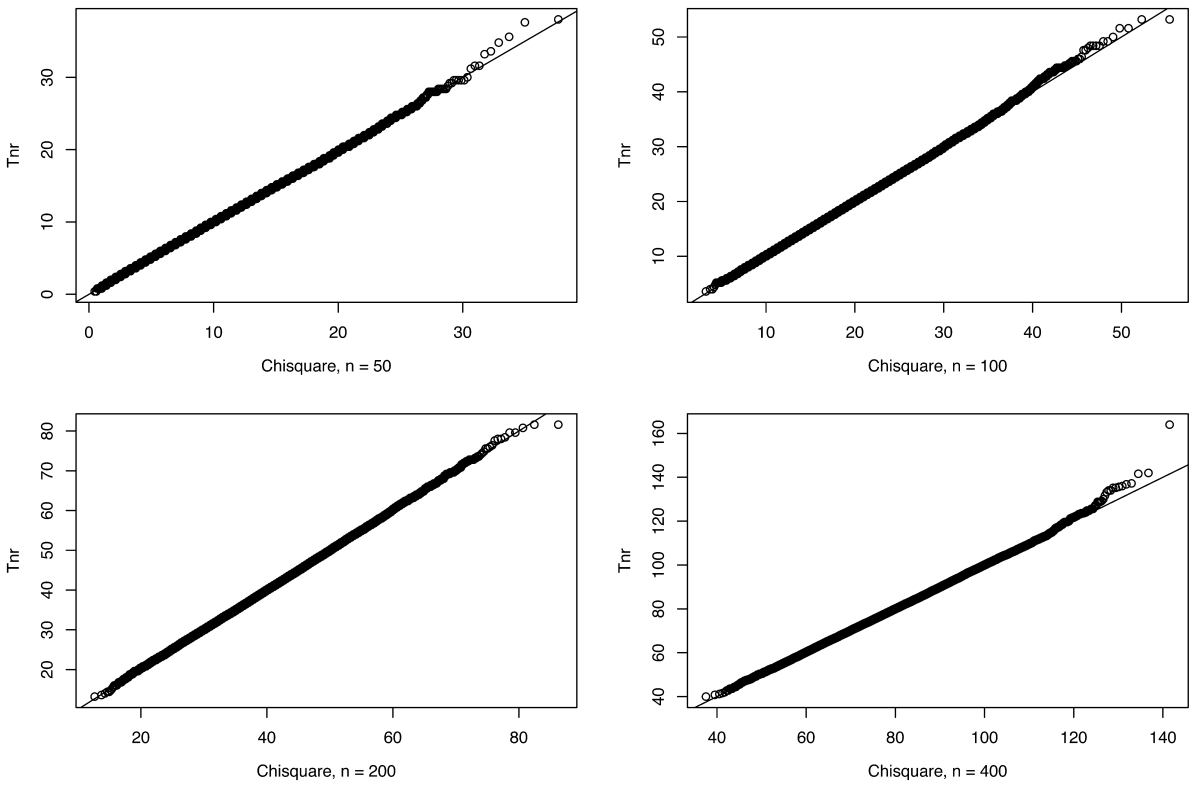


Fig. 1. Q–Q plots of the simulated values of $T_{n,n/5}$ against the chi-square distribution with $n/5-1$ degrees of freedom for $n=50, 100, 200, 400$.

(c) *The Lindeberg condition* (L) *implies* (2).
(d) *If* $\mathrm{E}(|\xi_{n,1}|^{2\alpha})=O(r_n^\alpha)$ *and* $r_n=o(n^{1-1/\alpha})$ *hold for some* $\alpha>1$, *then* (2) *holds.*

For its importance we formulate the special case $V_n=I_{r_n}$.

Corollary 5. *Suppose* V_n *equals* I_{r_n}. *Then the following are true.*

(a) $\mathrm{var}(|\xi_{n,1}|^2)=o(nr_n)$ *and* $\mathrm{E}(|\xi_{n,1}^\top\xi_{n,2}|^4)=o(n^2r_n^2)$ *imply* (3).
(b) *The moment condition* $\mathrm{E}(|\xi_{n,1}|^4)=o(nr_n)$ *implies* (3). *In particular,* $\mathrm{E}(|\xi_{n,1}|^4)=O(r_n^2)$ *and* $r_n=o(n)$ *imply* (3).
(c) *The Lindeberg condition* (L) *implies* (3).
(d) *If* $\mathrm{E}(|\xi_{n,1}|^{2\alpha})=O(r_n^\alpha)$ *and* $r_n=o(n^{1-1/\alpha})$ *hold for some* $\alpha>1$, *then* (3) *holds.*

When Peng and Schick [14] refer to part (c) of Corollary 3 of the preprint Peng and Schick (2012), which is an earlier version of this paper, they refer to part (c) of the above corollary.

Note that in the case $\mathrm{E}(|\xi_{n,1}|^4)=O(r_n^2)$ part (b) allows for larger r_n than part (d). More precisely, part (b) requires $r_n=o(n)$, while part (d) requires $r_n=o(n^{1/2})$.

Remark 2. In his Theorem 4.1, Portnoy [15] obtains the conclusion (3) in the case $V_n=I_{r_n}$ under the growth condition $r_n/n\to 0$ and the assumption that the coordinates $\xi_{n,1,i}$ of $\xi_{n,1}$ have a uniformly bounded sixth moment, viz.

$$\max_{1\le i\le r_n}\mathrm{E}(\xi_{n,1,i}^6)=O(1).$$

The last condition implies (9). Thus his result is a special case of Corollary 3.

Remark 3. Assume that $\xi_{n,1}=V_nZ_n$ for some symmetric idempotent matrix V_n with rank q_n tending to infinity and some random vector Z_n satisfying $\mathrm{E}(Z_n)=0$, $\mathrm{E}(Z_nZ_n^\top)=I_{r_n}$,

$$\zeta_n=\max_{1\le i\le r_n}\mathrm{E}(Z_{n,i}^4)=o(n), \tag{11}$$

and

$$\mathrm{E}(Z_{n,i}^{\alpha_1}Z_{n,j}^{\alpha_2}Z_{n,k}^{\alpha_3}Z_{n,\ell}^{\alpha_4}) = \mathrm{E}(Z_{n,i}^{\alpha_1})\mathrm{E}(Z_{n,j}^{\alpha_2})\mathrm{E}(Z_{n,k}^{\alpha_3})\mathrm{E}(Z_{n,\ell}^{\alpha_4}) \tag{12}$$

for distinct indices i, j, k, ℓ and non-negative integers $\alpha_1, \ldots, \alpha_4$ that sum to 4. The above conditions generalize those in [3] with our V_n equal to their $\Gamma_n^\top(\Gamma_n\Gamma_n^\top)^{-1}\Gamma_n$. These authors require instead of (11) the stronger $\mathrm{E}(Z_{n,1}^4) = \cdots = \mathrm{E}(Z_{n,r_n}^4) = \beta$ for some β. Relying on results of Bai and Saranadasa [1], they obtain (2) under the condition that $q_n = O(n)$. We shall show

$$\mathrm{var}(|V_n^{1/2}\xi_{n,1}|^2) = \mathrm{var}(|\xi_{n,1}|^2) \le (2+\zeta_n)q_n = o(nq_n) \tag{13}$$

and

$$\mathrm{E}(|\xi_{n,1}^\top\xi_{n,2}|^4) \le 3(q_n^2 + 2q_n + \zeta_n q_n) + \zeta_n(3+\zeta_n)q_n = o(n^2q_n^2). \tag{14}$$

Thus we obtain (2) from part (a) of Corollary 4 without their restrictions.

Note that the right-hand side in (12) equals zero if at least one of $\alpha_1, \ldots, \alpha_4$ equals one and that (12) yields $\mathrm{E}(Z_{n,i}^2Z_{n,j}^2) = \mathrm{E}(Z_{n,i}^2)\mathrm{E}(Z_{n,j}^2) = 1$ for $i \ne j$. Thus we calculate

$$\begin{aligned}
\mathrm{E}(|\xi_{n,1}|^4) &= \mathrm{E}\{(Z_n^\top V_nZ_n)^2\} = \sum_{i=1}^{r_n}\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\sum_{\ell=1}^{r_n}\mathrm{E}(Z_{n,i}V_{n,i,j}Z_{n,j}Z_{n,k}V_{n,k,\ell}Z_{n,\ell}) \\
&= \sum_{i\ne k}V_{n,i,i}V_{n,k,k} + \sum_{i\ne j}2V_{n,i,j}V_{n,i,j} + \sum_i \mathrm{E}(Z_{n,i}^4)V_{n,i,i}^2 \\
&= \{\mathrm{tr}(V_n)\}^2 + 2\,\mathrm{tr}(V_n) + \sum_{i=1}^{r_n}\{\mathrm{E}(Z_{n,i}^4) - 3\}V_{n,i,i}^2, \\
\mathrm{E}\{(Z_n^\top\xi_{n,2})^4\} &= \sum_{i=1}^{r_n}\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\sum_{\ell=1}^{r_n}\mathrm{E}(Z_{n,i}\xi_{n,2,i}Z_{n,j}\xi_{n,2,j}Z_{n,k}\xi_{n,2,k}Z_{n,\ell}\xi_{n,2,\ell}) \le 3\mathrm{E}(|\xi_{n,2}|^4) + \zeta_n\sum_{i=1}^{r_n}\mathrm{E}(\xi_{n,2,i}^4), \\
\sum_{\nu=1}^{r_n}\mathrm{E}(\xi_{n,1,\nu}^4) &= \sum_{\nu=1}^{r_n}\sum_{i=1}^{r_n}\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\sum_{\ell=1}^{r_n}\mathrm{E}(V_{n,\nu,i}Z_{n,i}V_{n,\nu,j}Z_{n,j}V_{n,\nu,k}Z_{n,k}V_{n,\nu,\ell}Z_{n,\ell}) \\
&\le \sum_{\nu=1}^{r_n}\left\{3\left(\sum_{i=1}^{r_n}V_{n,\nu,i}^2\right)^2 + \zeta_n\sum_{i=1}^{r_n}V_{n,\nu,i}^4\right\} \le 3\,\mathrm{tr}(V_n) + \zeta_n\,\mathrm{tr}(V_n).
\end{aligned}$$

Here we used the identity

$$\sum_{j=1}^{r_n}V_{n,i,j}^2 = \sum_{j=1}^{r_n}V_{n,i,j}V_{n,j,i} = V_{n,i,i}$$

and the following inequalities, valid for all $i, j \in \{1, \ldots, r_n\}$,

$$0 \le V_{n,i,i} \le 1 \quad \text{and} \quad V_{n,i,j}^2 \le 1.$$

Using the identities $|\xi_{n,1}|^2 = |V_n^{1/2}\xi_{n,1}|^2$ and $\xi_{n,1}^\top\xi_{n,2} = Z_n^\top\xi_{n,2}$ we obtain (13) and (14). Note also that $\mathrm{E}(|\xi_{n,1}|^2)$ equals $\mathrm{tr}(V_n)$.

Remark 4. Our results are motivated by recent results on extending Owen's [10–12] empirical likelihood approach to allow for an increasing number of constraints; see [9,3]. The empirical likelihood for this case is given by

$$\mathscr{R}_n = \sup\left\{\prod_{j=1}^n n\pi_j : 0 \le \pi_j, \sum_{j=1}^n\pi_j = 1, \sum_{j=1}^n\pi_jX_{n,j} = 0\right\}$$

where $X_{n,1}, \ldots, X_{n,n}$ are independent and identically distributed r_n-dimensional random variables with mean $\mathrm{E}(X_{n,1}) = 0$ and invertible dispersion matrix W_n. It is equivalent to

$$\mathscr{R}_n = \sup\left\{\prod_{j=1}^n n\pi_j : 0 \le \pi_j, \sum_{j=1}^n\pi_j = 1, \sum_{j=1}^n\pi_j\xi_{n,j} = 0\right\}$$

with $\xi_{n,j} = W_n^{-1/2}X_{n,j}$. The goal is to show that $-2\ln\mathscr{R}_n$ is approximately a chi-square random variable with r_n degrees of freedom. This is done by showing the asymptotic normality result

$$\frac{-2\ln\mathscr{R}_n - r_n}{\sqrt{2r_n}} \rightsquigarrow \mathcal{N}(0,1). \tag{15}$$

Table 2
Empirical sizes based on 1000 repetitions of the Kolmogorov–Smirnov test for normality at the .05 level using 500 standardized chi-square random variables with r degrees of freedom.

Degrees of freedom	50	100	150	200	250	300	350	400	450	500
Rejection frequency	0.193	0.103	0.098	0.081	0.082	0.073	0.069	0.072	0.060	0.058

This result is typically achieved in two steps. The first step establishes the approximation

$$-2\ln\mathscr{R}_n - |\tilde{\xi}_n|^2 = o_p(r_n^{1/2}),$$

and the second step obtains the asymptotic normality result (3).

Theorem 4.1 in [9] claims (15) under the assumptions that the qth moments of the coordinates of $X_{n,1}$ are uniformly bounded for some $q > 2$, that the eigenvalues of W_n are bounded and bounded away from zero, and that the dimension r_n satisfies

$$r_n^{3+6/(q-2)} = r_n^{3q/(q-2)} = o(n). \tag{16}$$

Its proof, however, is valid for the case $q \geq 6$ only, as the authors rely on Portnoy's [15] asymptotic normality result mentioned in Remark 2. With C a bound on the largest eigenvalue of $W_n^{-1/2}$ and B a bound on the qth moments of the coordinates of $X_{n,1}$, their assumptions imply

$$\mathrm{E}(|\xi_{n,1}|^q) \leq C^q \mathrm{E}(|X_{n,1}|^q) = C^q r_n^{q/2} \mathrm{E}\left\{\left(\frac{1}{r_n}\sum_{j=1}^{n} X_{n,1,j}^2\right)^{q/2}\right\} \leq C^q r_n^{q/2-1} \sum_{j=1}^{r_n} \mathrm{E}(|X_{n,1,j}|^q) = C^q B r_n^{q/2}.$$

Thus the required asymptotic normality follows from part (d) of Corollary 5 with $\alpha = q/2$. Note that their requirement (16) on r_n implies

$$r_n = o(n^{(q-2)/(3q)}) = o(n^{2(\alpha-1)/6\alpha}) = o(n^{1/3-1/(3\alpha)})$$

as needed. This closes the gap in Theorem 4.1 of [9]. This was already mentioned in [14] with a reference to an earlier version of the present paper. We should point out that Theorem 3 in [2] closes this gap as well. These authors treat generalized empirical likelihood ratios with dependent data. Theorem 2 is used in [14] to allow for larger r_n than mentioned above, see their Theorems 7.2–7.3.

Remark 5. Suppose $\mathrm{E}(|\xi_{n,1}|^{2\alpha}) = O(r_n^\alpha)$ holds for some $\alpha \geq 2$. Then the moment inequality yields $\mathrm{E}(|\xi_{n,1}|^4) \leq \mathrm{E}(|\xi_{n,1}|^{2\alpha})^{2/\alpha} = O(r_n^2)$. In this case, part (b) of Corollary 5 allows for larger r_n than part (d).

4. Simulations

We performed a simulation study to assess the asymptotic normality of the statistic in (1) with $\mu_n = 0$. We looked at the case when $\xi_{n,1}$ was a centered normal random vector and considered two types of parametric dispersion matrices. The first type uses a dispersion matrix with moving average structure with entries

$$\phi_\theta(i,j) = (1+\theta^2)\mathbf{1}\{i=j\} + \theta\mathbf{1}\{|i-j|=1\}, \quad i,j \in \{1,\ldots,r_n\},$$

and the second type an autoregressive dispersion matrix with entries

$$\rho_\theta(i,j) = \theta^{|i-j|}/(1-\theta^2), \quad i,j \in \{1,\ldots,r_n\}.$$

We chose the parameter θ to take the values 1/4, 1/2 and 3/4.

For each type of dispersion matrix and parameter value θ, we generated 500 copies of the statistic and performed a Kolmogorov–Smirnov test for a combination of sample size n and dimension r_n. We repeated this 1000 times and recorded the percentage of rejections at the .05 level. We selected the sample sizes $n = 300, 600, 900, 1200, 1500$ and chose the dimensions $r_n = n(i/6)$ for all $i \in \{1,\ldots,6\}$.

The models coincide when the parameter θ equals zero as then both covariance matrices reduce to identity matrices. In this case the statistic $|\tilde{\xi}_n|^2$ has a chi-square distribution with r_n degrees of freedom. Table 2 reports the rejection frequency based on 1000 repetitions of the Kolmogorov–Smirnov test for normality at the .05 level based on 500 copies of a standardized chi-square random variable with r degrees of freedom.

Table 3 reports the percentage of rejections for these combinations of n and r_n and the two dispersion structures with the above three choices of parameters. The table suggests the following. The normal approximation requires rather large n and r_n. The normal approximation is worse for larger values of the parameter and is worse under the autoregressive structure than under the moving average structure.

Table 3
Empirical sizes based on 1000 repetitions of the Kolmogorov–Smirnov test for normality at the level .05 using 500 copies of the statistic in Eq. (3) for selected sample sizes n, dimensions r_n and two types of parametric dispersion structures each with three choices of parameters.

θ	n	$r_n = n/6$	$r_n = n/3$	$r_n = n/2$	$r_n = 2n/3$	$r_n = 5n/6$	$r_n = n$
Moving average structure							
	300	0.219	0.132	0.114	0.107	0.088	0.078
	600	0.147	0.092	0.072	0.065	0.070	0.066
.25	900	0.122	0.075	0.067	0.066	0.069	0.065
	1200	0.087	0.065	0.065	0.069	0.062	0.061
	1500	0.085	0.065	0.056	0.060	0.054	0.055
	300	0.305	0.171	0.129	0.113	0.109	0.090
	600	0.164	0.099	0.084	0.070	0.063	0.069
.50	900	0.130	0.091	0.072	0.065	0.066	0.061
	1200	0.111	0.090	0.063	0.062	0.061	0.060
	1500	0.089	0.068	0.068	0.056	0.046	0.048
	300	0.345	0.179	0.146	0.120	0.097	0.095
	600	0.190	0.121	0.093	0.075	0.056	0.065
.75	900	0.143	0.108	0.078	0.071	0.081	0.065
	1200	0.121	0.087	0.071	0.075	0.068	0.057
	1500	0.101	0.077	0.073	0.082	0.072	0.070
Autoregressive structure							
	300	0.283	0.155	0.116	0.102	0.086	0.083
	600	0.148	0.106	0.090	0.084	0.068	0.059
.25	900	0.124	0.075	0.065	0.058	0.047	0.054
	1200	0.107	0.066	0.056	0.055	0.048	0.065
	1500	0.090	0.063	0.054	0.058	0.055	0.057
	300	0.637	0.272	0.187	0.179	0.140	0.127
	600	0.444	0.153	0.122	0.104	0.092	0.079
.50	900	0.405	0.125	0.098	0.095	0.083	0.084
	1200	0.342	0.080	0.076	0.073	0.065	0.070
	1500	0.323	0.094	0.065	0.068	0.073	0.066
	300	0.996	0.713	0.468	0.376	0.297	0.240
	600	0.961	0.559	0.313	0.188	0.154	0.141
.75	900	0.935	0.444	0.228	0.158	0.145	0.132
	1200	0.937	0.399	0.179	0.143	0.118	0.104
	1500	0.943	0.382	0.188	0.122	0.096	0.087

5. Asymptotic behavior under local alternatives

Let $(\mathscr{X}, \mathscr{S}, Q)$ be a probability space and w_n be a function from $\mathscr{X}$ into $\mathbb{R}^{r_n}$ satisfying $\int w_n\,dQ = 0$, $\int |w_n|^2\,dQ < \infty$ and

$$\forall_{\epsilon>0} \quad \Lambda_n(\epsilon) = \int |w_n|^2 \mathbf{1}\{|w_n| > \epsilon\sqrt{n}\}\,dQ \to 0. \tag{17}$$

Assume also that the matrix $W_n = \int w_n w_n^\top\,dQ$ satisfies

$$\lambda_n = \sup_{|u|=1} u^\top W_n u = \sup_{|u|=1} \int (u^\top w_n)^2\,dQ = O(1), \quad \operatorname{tr}(W_n^2) \to \infty \quad \text{and} \quad \operatorname{tr}(W_n) = O\{\operatorname{tr}(W_n^2)\}.$$

It then follows from Theorem 2 that

$$\frac{|n^{-1/2}\sum_{j=1}^n w_n(X_j)|^2 - \operatorname{tr}(W_n)}{\sqrt{2\operatorname{tr}(W_n^2)}} \rightsquigarrow \mathcal{N}(0,1)$$

if $X_1, \ldots, X_n$ are independent $\mathscr{X}$-valued random variables with distribution Q.

The next theorem answers the question of what happens if we slightly perturb the distribution Q. Let h denote a measurable function satisfying $\int h\,dQ = 0$ and $\int h^2\,dQ < \infty$ and set

$$h_n = h\mathbf{1}\{|h| < c_n n^{1/2}/s_n\} - \int h\mathbf{1}\{|h| < c_n n^{1/2}/s_n\}\,dQ$$

with $0 < c_n < 1/2$, $1 \le s_n$, $c_n \to 0$, $s_n^2 = o\{\operatorname{tr}(W_n^2)\}$ and $c_n n^{1/2}/s_n \to \infty$. Let $Q_{n,h}$ denote the probability measure with density $1 + n^{-1/2}s_n h_n$ with respect to Q. By construction, we have

$$\int |n^{1/2}/s_n(\sqrt{1 + n^{-1/2}s_n h_n} - 1) - h/2|^2\,dQ \to 0. \tag{18}$$

If $s_n = 1$, this implies that the product measures $Q_{n,h}^n$ and Q^n are mutually contiguous. Set

$$\mu_n(h) = \int w_n h \, dQ \quad \text{and} \quad \Delta_n = n^{-1/2} s_n \int |w_n|^2 h_n \, dQ.$$

Theorem 3. *Let $X_{n,1}, \ldots, X_{n,n}$ be independent $\mathscr{X}$-valued random variables with distribution $Q_{n,h}$. Then we have the asymptotic normality result*

$$\frac{|n^{-1/2} \sum_{j=1}^n w_n(X_{n,j})|^2 - s_n^2 |\mu_n(h)|^2 - \operatorname{tr}(W_n) - \Delta_n}{\sqrt{2 \operatorname{tr}(W_n^2)}} \rightsquigarrow \mathcal{N}(0, 1). \tag{19}$$

In the case $s_n = 1$, this simplifies to

$$\frac{|n^{-1/2} \sum_{j=1}^n w_n(X_{n,j})|^2 - |\mu_n(h)|^2 - \operatorname{tr}(W_n)}{\sqrt{2 \operatorname{tr}(W_n^2)}} \rightsquigarrow \mathcal{N}(0, 1). \tag{20}$$

Proof.

Taking $\nu_n = \mu_n(h_n)$ and $\xi_{n,j} = w_n(X_{n,j}) - n^{-1/2} s_n \nu_n$, we can write

$$\Big| n^{-1/2} \sum_{j=1}^n w_n(X_{n,j}) \Big|^2 = |\tilde{\xi}_n + s_n \nu_n|^2.$$

The dispersion matrix of $\xi_{n,1}$ is given by $V_n = \tilde{W}_n - n^{-1} s_n^2 \nu_n \nu_n^\top$, where

$$\tilde{W}_n = \int w_n w_n^\top \, dQ_{n,h} = W_n + n^{-1/2} s_n \int w_n w_n^\top h_n \, dQ.$$

By construction, $|n^{-1/2} s_n h_n|$ is bounded by $2c_n$. Thus, for $k \in \{1, 2\}$, we have the inequality

$$(1 - 2c_n)^k \operatorname{tr}(W_n^k) \le \operatorname{tr}(\tilde{W}_n^k) \le (1 + 2c_n)^k \operatorname{tr}(W_n^k)$$

and obtain

$$\frac{\operatorname{tr}(\tilde{W}_n)}{\operatorname{tr}(W_n)} \to 1 \quad \text{and} \quad \frac{\operatorname{tr}(\tilde{W}_n^2)}{\operatorname{tr}(W_n^2)} \to 1.$$

Since $\operatorname{tr}(W_n^2) \le \lambda_n \operatorname{tr}(W_n)$, we also have $\operatorname{tr}(W_n) \to \infty$.

The requirements on the sequences c_n and s_n imply $n^{-1/2} s_n = o(c_n) = o(1)$. Using this and the above, we find

$$\sup_{|u|=1} u^\top V_n u \le \sup_{|u|=1} u^\top \tilde{W}_n u \le (1 + 2c_n) \sup_{|u|=1} u^\top W_n u = O(\lambda_n) = O(1) \tag{21}$$

$$|\nu_n|^2 = \sup_{|u|=1} (u^\top \nu_n)^2 \le \int h_n^2 \, dQ \sup_{|u|=1} \int (u^\top w_n)^2 \, dQ \le \lambda_n \int h^2 \, dQ = O(1), \tag{22}$$

$$|\nu_n - \mu_n(h)|^2 \le \lambda_n \Big| \int (h_n - h) W_n^{-1/2} w_n \, dQ \Big|^2 \le \lambda_n \int (h_n - h)^2 \, dQ \to 0, \tag{23}$$

$$\operatorname{tr}(V_n) = \operatorname{tr}(\tilde{W}_n) - n^{-1} s_n^2 |\nu_n|^2 = \operatorname{tr}(W_n) + o\{\operatorname{tr}(W_n)\}. \tag{24}$$

$$\operatorname{tr}(V_n^2) = \operatorname{tr}(\tilde{W}_n^2) - 2n^{-1} s_n^2 \nu_n^\top \tilde{W}_n \nu_n + n^{-2} s_n^4 |\nu_n|^4 = \operatorname{tr}(W_n^2) + o\{\operatorname{tr}(W_n^2)\}. \tag{25}$$

In the last step of (23) we used Bessel's inequality. It is now easy to see that conditions (C1)–(C3) hold with $\mu_n = s_n \nu_n$. Indeed, (C1) follows from (21) and (22), (C2) follows from (21), (25) and $\operatorname{tr}(W_n^2) \to \infty$, and (C3) follows from (24), (25) and $\operatorname{tr}(W_n) = O\{\operatorname{tr}(W_n^2)\}$. Finally, using (17), $|n^{-1/2} s_n \nu_n| = o(1)$ and the bound $n^{-1/2} s_n |h_n| \le 1$, we derive the Lindeberg condition (L). Thus Theorem 2 yields

$$\frac{|n^{-1/2} \sum_{j=1}^n w_n(X_{n,j})|^2 - s_n^2 |\nu_n|^2 - \operatorname{tr}(V_n)}{\sqrt{2 \operatorname{tr}(V_n^2)}} \rightsquigarrow \mathcal{N}(0, 1).$$

The desired result (19) follows from this, (23), (25) and the fact that $\operatorname{tr}(V_n) = \operatorname{tr}(W_n) + \Delta_n + o(1)$.

In the case $s_n = 1$, we have the bound

$$\begin{aligned} \int |w_n|^2 |n^{-1/2} h_n| \, dQ &\le 2c_n \Lambda_n(\epsilon) + \epsilon \int |w_n| |h_n| \mathbf{1}\{|w_n| \le \epsilon \sqrt{n}\} \, dQ \\ &\le 2c_n \Lambda_n(\epsilon) + \epsilon \Big(\int h_n^2 \, dQ \int |w_n|^2 \, dQ \Big)^{1/2}, \quad \epsilon > 0. \end{aligned}$$

This bound and $\operatorname{tr}(W_n) = O\{\operatorname{tr}(W_n^2)\}$ yield $\Delta_n = o[\{\operatorname{tr}(W_n^2)\}^{1/2}]$ and hence (20). □

Remark 6. Let $X_{n,1}, \ldots, X_{n,n}$ be independent $\mathscr{X}$-valued random variables with distribution $Q_{n,h}$ for $s_n = 1$. Consider the empirical likelihood

$$\mathscr{R}_n = \sup\left\{\prod_{j=1}^n n\pi_j : 0 \le \pi_j, \sum_{j=1}^n \pi_j = 1, \sum_{j=1}^n \pi_j v_n(X_{n,j}, X_{n,1}, \ldots, X_{n,n}) = 0\right\}$$

with v_n a measurable function from $\mathscr{X}^{n+1}$ into $\mathbb{R}^{r_n}$. Suppose that

$$-2\ln\mathscr{R}_n - \Big|n^{-1/2}\sum_{j=1}^n w_n(X_{n,j})\Big|^2 = o_p[\{\mathrm{tr}(W_n^2)\}^{1/2}]$$

when $h = 0$. By contiguity, this then also holds if $h \neq 0$ and we obtain

$$\frac{-2\ln\mathscr{R}_n - |\mu_n(h)|^2 - \mathrm{tr}(W_n)}{\sqrt{2\,\mathrm{tr}(W_n^2)}} \rightsquigarrow \mathcal{N}(0,1).$$

If W_n is idempotent with rank q_n tending to infinity, this simplifies to

$$\frac{-2\ln\mathscr{R}_n - |\mu_n(h)|^2 - q_n}{\sqrt{2q_n}} \rightsquigarrow \mathcal{N}(0,1)$$

and may be interpreted as $-2\ln\mathscr{R}_n$ being approximately a non-central chi-square random variable with q_n degrees of freedom and non-centrality parameter $|\mu_n(h)|$.

Remark 7. In the previous remark $Q_{n,h}$ was chosen to have density $1 + n^{-1/2}h_n$. By (18) this implies that

$$\int |n^{1/2}(\sqrt{dQ_{n,h}} - \sqrt{dQ}) - h/2\sqrt{dQ}|^2 \to 0. \tag{26}$$

The results of the previous remark remain true under the more general condition (26).

6. Applications

In applications, the quadratic form $|\tilde{\xi}_n|^2$ will often serve as an approximation to a more complicated statistic S_n. More precisely, suppose that we have the expansion

$$S_n = |\tilde{\xi}_n|^2 + o_p(r_n^{1/2}), \tag{27}$$

then the asymptotic normality result (3) implies the same asymptotic normality result for S_n,

$$\frac{S_n - r_n}{\sqrt{2r_n}} \rightsquigarrow \mathcal{N}(0,1). \tag{28}$$

We have already encountered this concept in Remark 4.

Of special interest is the case $\xi_{nj} = W_n^{-1/2}w_n(Z_j)$, where $Z_1, \ldots, Z_n$ are k-dimensional random vectors with common distribution Q and w_n is a measurable function from $\mathbb{R}^k$ into $\mathbb{R}^{r_n}$ such that $w_n(Z_1)$ has mean $\int w_n\,dQ = 0$ and dispersion matrix $W_n = \int w_n w_n^\top\,dQ$ which satisfies

$$0 < \inf_n \inf_{|u|=1} u^\top W_n u \le \sup_n \sup_{|u|=1} u^\top W_n u < \infty. \tag{29}$$

Suppose also that $\mathrm{E}(|w_n(Z_1)|^4) = O(r_n^2)$ and $r_n = o(n)$. It then follows from part (b) of Corollary 3 that (3) holds.

Now let $\hat{w}_n$ denote an estimator of w_n and set

$$\hat{T}_n = n^{-1/2}\sum_{j=1}^n \hat{w}_n(Z_j), \quad \hat{W}_n = \frac{1}{n}\sum_{j=1}^n \hat{w}_n(Z_j)\hat{w}_n^\top(Z_j) \quad \text{and} \quad \bar{W}_n = \frac{1}{n}\sum_{j=1}^n w_n(Z_j)w_n^\top(Z_j).$$

Now consider the statistic $S_n = \hat{T}_n^\top \hat{W}_n^{-1}\hat{T}_n$. In this setting, (27) follows from the statements

$$\Big|n^{-1/2}\sum_{j=1}^n\{\hat{w}_n(Z_j) - w_n(Z_j)\}\Big|^2 = o_p(1) \tag{30}$$

and

$$|\hat{W}_n - W_n|_o = \sup_{|u|=1}|u^\top(\hat{W}_n - W_n)u| = o_p(r_n^{-1/2}). \tag{31}$$

These statements typically require additional restrictions on the rate of growth of r_n. We have $|\bar{W}_n - W_n|_o = O_p(r_n/n^{1/2})$ as $\mathrm{E}(|\bar{W}_n - W_n|_0^2) \le \mathrm{E}(|\bar{W}_n - W_n|^2) \le \mathrm{E}(|w_n(Z_1)|^4)/n = O(r_n^2/n)$ and $|\hat{W}_n - \bar{W}_n|_o \le D_n + 2|\bar{W}_n|_o^{1/2} D_n^{1/2}$ with

$$D_n = \frac{1}{n}\sum_{j=1}^n |\hat{w}_n(Z_j) - w_n(Z_j)|^2.$$

Hence (31) is implied by $r_n^3 = o(n)$ and $D_n = o_p(r_n^{-1})$. Let us summarize our findings.

Proposition 1. *Suppose w_n is given as above and $r_n^3 = o(n)$, $D_n = o_p(r_n^{-1})$ and (30) hold. Then we have the asymptotic normality result (28) with $S_n = \hat{T}_n \hat{W}_n^{-1} \hat{T}_n$.*

7. Testing for equal marginals

Let us illustrate the result of the previous section by means of an example, namely testing for the equality of the marginal distributions of a bivariate random vector. Let the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent copies of a bivariate random vector (X, Y). We want to test whether the marginal distributions are the same. This is of importance when X denotes pre-treatment and Y post-treatment measurement. Equality of the marginal distributions indicates that there is no treatment effect. Assume that the marginal distribution functions F (of X) and G (of Y) are continuous.

Let us set $H = (F + G)/2$. We can estimate H by the pooled empirical distribution function, defined for all $x \in \mathbb{R}$, by

$$\mathbb{H}(x) = \frac{1}{n}\sum_{j=1}^n (\mathbf{1}\{X_j \le x\} + \mathbf{1}\{Y_j \le x\})/2.$$

Assume from now on that F equals G so that the null hypothesis holds. Then we have $H = F = G$ and $\mathrm{E}\{a(X) - a(Y)\} = 0$ for every $a \in L_{2,0}(H)$ where

$$L_{2,0}(H) = \left\{ a \in \ell_2(H) : \int a\, dH = 0 \right\}.$$

We also impose the condition

$$\inf_{a \in A} \mathrm{E}\{|a(X) - a(Y)|^2\} > 0 \tag{32}$$

with $A = \{a \in L_{2,0}(H) : \int a^2\, dH = 1\}$ the unit sphere in $L_{2,0}(H)$.

Let $\psi_1, \psi_2, \ldots$ denote an orthonormal basis of $L_{2,0}(U)$, where U is the uniform distribution on $[0, 1]$. Since H is continuous, the functions $\psi_1 \circ H, \psi_2 \circ H, \ldots$ form an orthonormal basis of $L_{2,0}(H)$. We shall work with the trigonometric basis defined, for all $x \in [0, 1]$ and $k \in \{1, 2, \ldots\}$, by $\psi_k(x) = \sqrt{2}\cos(\pi k x)$, because these functions are bounded and have bounded derivatives. Let $v_n = (\psi_1, \ldots, \psi_{r_n})^\top$ and set, for all $x, y \in \mathbb{R}$,

$$w_n(x, y) = v_n \circ H(x) - v_n \circ H(y) \quad \text{and} \quad \hat{w}_n(x, y) = v_n \circ \mathbb{H}(x) - v_n \circ \mathbb{H}(y).$$

It follows from (32) that the dispersion matrix $W_n = \mathrm{E}(w_n(X_1, Y_1) w_n^\top(X_1, Y_1))$ of $w_n(X_1, Y_1)$ satisfies

$$0 < \inf_{a \in A} \mathrm{E}\{|a(X) - a(Y)|^2\} \le u^\top W_n u \le 4, \quad |u| = 1.$$

To see this use the fact that $u^\top v_n$ belongs to A for each unit vector u. Thus (29) holds. Since $|w_n| \le 2|v_n| \le 2\sqrt{2r_n}$, we obtain $\mathrm{E}\{|w_n(X, Y)|^4\} = O(r_n^2)$.

If H and W_n were known, we could work with the test statistic $|W_n^{-1/2} T_n|^2 = T_n^\top W_n^{-1} T_n$, where T_n is the statistic $n^{-1/2}\sum_{j=1}^n w_n(X_j, Y_j)$. Since H and W_n are unknown, we work instead with $\hat{T}_n^\top \hat{W}_n^{-1} \hat{T}_n$ where

$$\hat{T}_n = n^{-1/2}\sum_{j=1}^n \hat{w}_n(X_j, Y_j) \quad \text{and} \quad \hat{W}_n = \frac{1}{n}\sum_{j=1}^n \hat{w}_n(X_j, Y_j)\hat{w}_n^\top(X_j, Y_j).$$

Using $|v_n'|^2 \le 2\pi^2 r_n^3$, we have

$$D_n = \frac{1}{n}\sum_{j=1}^n |\hat{w}_n(X_j, Y_j) - w_n(X_j, Y_j)|^2 \le 8\pi^2 r_n^3 \sup_{t \in \mathbb{R}} |\mathbb{H}(t) - H(t)|^2 = O_p(r_n^3/n).$$

This implies $D_n = o_p(r_n^{-1})$ if $r_n^4 = o(n)$. Finally (30) holds as shown in [13], pp. 403–404, if $r_n^3 = o(n)$. Indeed it follows from there that

$$\sum_{k=1}^{r_n} \mathrm{E}\left\{\left(\frac{1}{n}\sum_{j=1}^{n}\hat{w}_n(X_j,Y_j)-w_n(X_j,Y_j)\right)^2\right\} \le \frac{48\pi^2 r_n^3}{(n-1)^2}+3\left\{\frac{8\pi^2 r_n^3}{n(n-1)}+\frac{32\pi^2 r_n^3(n-1)}{n(n-1)^2}\right\}$$

Thus we have proved the following result.

Corollary 6. *Suppose F equals G and (32) holds. Then we have the asymptotic normality result*

$$\frac{\hat{T}_n^\top \hat{W}_n^{-1}\hat{T}_n - r_n}{\sqrt{2r_n}} \rightsquigarrow \mathcal{N}(0,1)$$

provided r_n tends to infinity and r_n^4/n tends to zero.

This result shows that the test which rejects the null hypothesis if $\hat{T}_n^\top \hat{W}_n^{-1}\hat{T}_n$ exceeds the $(1-\alpha)$-quantile of the chi-square distribution with r_n degrees of freedom has asymptotic size α.

We conducted a small simulation study to investigate the power of this test. We first looked at data from a bivariate normal distribution with parameters $(0,\theta,1,\sigma^2,\rho)$ where the first two coordinates refer to the means, the third and fourth to the variances, and the fifth to the correlation coefficient. By (a) (p. 394) of [13] the bivariate normal model satisfies the condition (32). We simulated the power for some choices of θ, σ^2 and ρ, namely $\theta \in \{.0, .2, .4\}$, $\sigma^2 = .7, 1, 1.3$ and $\rho \in \{.5, .8\}$, for the sample sizes $n \in \{50, 100, 150\}$ and for the values $r_n \in \{1, 2, 3, 4\}$. In each case the power was estimated based on 10,000 repetitions using a significance level of $\alpha = .05$. The results are reported in Table 4 for the above mentioned values of θ, σ^2, ρ, n and r_n. The rows corresponding to the value $(\theta,\sigma^2) = (0,1)$ refer to the null hypothesis and are set in boldface. We see from the table that the power is larger for the larger value of ρ.

We also generated data from the Farlie–Gumbel–Morgenstern copula model with marginals F and G possessing densities f and g, respectively. The density for this model is given, for all $x, y \in \mathbb{R}$, by

$$p_{\gamma,F,G}(x,y) = [1+\gamma\{1-2F(x)\}\{1-2G(y)\}]f(x)g(y),$$

where γ is a number in the interval $(-1,1)$. As shown in [13], the density $p_{\gamma,F,G}$ satisfies condition (32). For our simulation we took $\gamma \in \{.5, .8\}$, F to be the logistic distribution function, $F(x) = 1/(1+e^{-x})$, and G of the form $G(x) = F\{(x-\theta)/\sigma\}$ for some selected values of θ and σ, namely $\theta \in \{0, .2, .4\}$ and $\sigma \in \{.8, 1, 1.2\}$. We again estimated the powers using 10,000 repetitions. Table 5 reports the simulated powers of the test for the above combinations of values of θ, σ and $\gamma \in \{.5, .8\}$, for sample sizes $n \in \{50, 100, 150\}$ and nominal level of significance .05. The rows with $(\theta,\sigma) = (0,1)$ correspond to the null hypothesis. From the table it appears that the power is slightly larger for the larger value of γ.

8. Testing diagonality of a covariance matrix of a normal random vector with increasing dimension

Let $Z_1, \ldots, Z_{q+1}$ be independent m-dimensional $\mathcal{N}(\mu, \Sigma)$ random vectors. To test whether Σ is diagonal, we introduce the test statistic

$$S(m,q) = \frac{2}{m}\sqrt{\frac{q^2(q+2)}{4(q-1)}}\left\{\sum_{1\le i<j\le m} r_{ij}^2 - \frac{m(m-1)}{2q}\right\},$$

where r_{ij} denotes the sample correlation coefficient of the ith and jth row of the $m\times(q+1)$ random matrix $[Z_1, \ldots, Z_{q+1}]$. This is a standardized version of the test statistic studied by Schott [19]. His n corresponds to our q. We changed the asymptotic variance by replacing his $m(m-1)$ by m^2. This was done for notational convenience. He showed that under the condition $m/q \to \gamma \in (0,\infty)$ his (non-standardized) test statistic is asymptotically normal with mean 0 and variance γ^2 if the null hypothesis of diagonality holds. Here we shall show that the standardized version given above is asymptotically standard normal even if $m/q \to \infty$. More precisely, we shall prove the following result.

Theorem 4. *Suppose Σ is diagonal and m and q tend to infinity. Then $S(m,q)$ is asymptotically standard normal.*

Proof. Without loss of generality we may and do assume in the following that μ is the zero vector and Σ is the $m\times m$ identity matrix. Then the rows $\tilde{Z}_1, \ldots, \tilde{Z}_m$ of the $m\times(q+1)$ matrix $[Z_1, \ldots, Z_n]$ are independent $(q+1)$-dimensional standard normal random vectors. Let J_k denote the $k\times k$ matrix with all entries equal to $1/k$. Then, with $A_{q+1} = I_{q+1} - J_{q+1}$, we can write, for all $1 \le i < j \le m$,

$$r_{ij} = \frac{\tilde{Z}_i^\top A_{q+1}\tilde{Z}_j}{|A_{q+1}\tilde{Z}_i|\,|A_{q+1}\tilde{Z}_j|}.$$

The matrix A_{q+1} is idempotent with rank q and thus can be written as $Q^\top \Delta Q$ with Q orthogonal and Δ the diagonal matrix with the first q diagonal entries equals to 1 and the last entry equal to 0. This shows that

$$r_{ij} = \frac{\bar{Z}_i^\top \bar{Z}_j}{|\bar{Z}_i|\,|\bar{Z}_j|} = S_i^\top S_j,$$

Table 4
Simulated power for bivariate normal data with $\alpha = .05$ for selected values of θ, σ^2, ρ, n and $r_n \in \{1, \ldots, 4\}$. The simulations are based on 10,000 repetitions. Numbers in boldface refer to the null hypothesis.

n	θ	σ^2	$\rho = .5$				$\rho = .8$			
			1	2	3	4	1	2	3	4
	0.0	0.7	0.051	0.127	0.105	0.103	0.060	0.186	0.145	0.161
	0.0	**1.0**	**0.049**	**0.044**	**0.047**	**0.042**	**0.050**	**0.045**	**0.042**	**0.037**
	0.0	1.4	0.051	0.117	0.097	0.095	0.052	0.164	0.130	0.144
	0.2	0.7	0.276	0.290	0.240	0.224	0.531	0.539	0.484	0.453
50	0.2	1.0	0.244	0.180	0.161	0.134	0.468	0.341	0.319	0.261
	0.2	1.4	0.210	0.225	0.187	0.171	0.409	0.433	0.380	0.352
	0.4	0.7	0.778	0.724	0.684	0.629	0.982	0.971	0.972	0.956
	0.4	1.0	0.714	0.590	0.554	0.484	0.969	0.925	0.928	0.888
	0.4	1.4	0.613	0.562	0.517	0.471	0.927	0.895	0.886	0.859
	0.0	0.7	0.052	0.227	0.188	0.206	0.052	0.353	0.291	0.359
	0.0	**1.0**	**0.048**	**0.049**	**0.047**	**0.044**	**0.053**	**0.049**	**0.050**	**0.048**
	0.0	1.4	0.054	0.199	0.160	0.180	0.057	0.313	0.259	0.313
	0.2	0.7	0.505	0.564	0.514	0.499	0.836	0.876	0.861	0.862
100	0.2	1.0	0.440	0.341	0.314	0.274	0.778	0.676	0.671	0.610
	0.2	1.4	0.370	0.450	0.403	0.400	0.691	0.756	0.727	0.730
	0.4	0.7	0.972	0.967	0.965	0.954	1.000	1.000	1.000	1.000
	0.4	1.0	0.950	0.899	0.891	0.856	1.000	1.000	1.000	0.999
	0.4	1.4	0.908	0.897	0.879	0.859	0.998	0.998	0.998	0.997
	0.0	0.7	0.050	0.330	0.280	0.328	0.055	0.506	0.431	0.542
	0.0	**1.0**	**0.048**	**0.052**	**0.049**	**0.047**	**0.052**	**0.052**	**0.051**	**0.051**
	0.0	1.4	0.052	0.296	0.249	0.283	0.052	0.456	0.386	0.490
	0.2	0.7	0.677	0.758	0.728	0.727	0.945	0.973	0.974	0.974
150	0.2	1.0	0.598	0.486	0.465	0.418	0.918	0.859	0.863	0.823
	0.2	1.4	0.516	0.629	0.578	0.589	0.856	0.919	0.908	0.915
	0.4	0.7	0.997	0.997	0.997	0.996	1.000	1.000	1.000	1.000
	0.4	1.0	0.993	0.983	0.982	0.972	1.000	1.000	1.000	1.000
	0.4	1.4	0.982	0.983	0.981	0.974	1.000	1.000	1.000	1.000

Table 5
Simulated power in the FGM copula with logistic marginals with $\alpha = .05$ for selected values of θ, σ, γ, n, and $r_n \in \{1, \ldots, 4\}$. The simulations are based on 10,000 repetitions. Numbers in boldface refer to the null hypothesis.

n	θ	σ	$\gamma = .5$				$\gamma = .8$			
			1	2	3	4	1	2	3	4
	0.0	0.8	0.048	0.138	0.112	0.109	0.052	0.142	0.116	0.112
	0.0	**1.0**	**0.050**	**0.050**	**0.049**	**0.047**	**0.052**	**0.049**	**0.046**	**0.042**
	0.0	1.2	0.056	0.110	0.092	0.084	0.056	0.111	0.093	0.086
	0.2	0.8	0.110	0.179	0.143	0.134	0.116	0.189	0.155	0.140
50	0.2	1.0	0.094	0.078	0.070	0.062	0.101	0.075	0.070	0.063
	0.2	1.2	0.080	0.127	0.106	0.100	0.086	0.134	0.112	0.102
	0.4	0.8	0.274	0.309	0.253	0.225	0.305	0.330	0.268	0.244
	0.4	1.0	0.231	0.169	0.137	0.118	0.257	0.185	0.153	0.131
	0.4	1.2	0.206	0.219	0.177	0.161	0.215	0.227	0.182	0.164
	0.0	0.8	0.051	0.252	0.212	0.218	0.055	0.257	0.214	0.217
	0.0	**1.0**	**0.050**	**0.051**	**0.048**	**0.046**	**0.050**	**0.049**	**0.047**	**0.046**
	0.0	1.2	0.050	0.178	0.149	0.154	0.053	0.183	0.153	0.151
	0.2	0.8	0.166	0.350	0.296	0.290	0.186	0.361	0.307	0.298
100	0.2	1.0	0.145	0.110	0.093	0.083	0.148	0.115	0.097	0.087
	0.2	1.2	0.121	0.239	0.202	0.194	0.136	0.249	0.204	0.199
	0.4	0.8	0.493	0.586	0.517	0.491	0.543	0.612	0.548	0.523
	0.4	1.0	0.417	0.321	0.270	0.237	0.468	0.354	0.309	0.267
	0.4	1.2	0.356	0.416	0.353	0.326	0.388	0.450	0.388	0.364
	0.0	0.8	0.055	0.377	0.318	0.333	0.054	0.366	0.308	0.322
	0.0	**1.0**	**0.050**	**0.049**	**0.052**	**0.050**	**0.051**	**0.054**	**0.049**	**0.048**
	0.0	1.2	0.052	0.267	0.221	0.228	0.049	0.255	0.222	0.227
	0.2	0.8	0.226	0.498	0.433	0.434	0.255	0.511	0.446	0.452
150	0.2	1.0	0.187	0.151	0.121	0.112	0.211	0.157	0.139	0.123
	0.2	1.2	0.163	0.351	0.294	0.289	0.181	0.364	0.310	0.310
	0.4	0.8	0.661	0.778	0.718	0.702	0.722	0.814	0.758	0.740
	0.4	1.0	0.577	0.463	0.407	0.354	0.635	0.524	0.460	0.412
	0.4	1.2	0.497	0.598	0.527	0.508	0.545	0.631	0.569	0.543

where $\bar{Z}_k$ denotes the q-dimensional random vector formed by the first q components of $Q\tilde{Z}_k$ and $S_k = \bar{Z}_k/|\bar{Z}_k|$ for all $k \in \{1,\dots,m\}$. Note that $S_1,\dots,S_m$ are independent q-dimensional random vectors each uniformly distributed on the unit sphere $\{x \in \mathbb{R}^q : |x| = 1\}$ in $\mathbb{R}^q$. This is as in [19]. We can write

$$r_{ij}^2 - \frac{1}{q} = \sum_{k=1}^{q}\sum_{\ell=1}^{q} S_{ik}S_{jk}S_{i\ell}S_{j\ell} - \frac{1}{q} = \sum_{k=1}^{q}\left(S_{ik}^2 - \frac{1}{q}\right)\left(S_{jk}^2 - \frac{1}{q}\right) + \sum_{1\le k<\ell\le q} 2S_{ik}S_{i\ell}S_{jk}S_{j\ell} = X_i^\top X_j + Y_i^\top Y_j,$$

where, for each $i \in \{1,\dots,m\}$, $X_i = (S_{i1}^2 - 1/q,\dots,S_{iq}^2 - 1/q)^\top$ and Y_i is the $q(q-1)/2$-dimensional random vector with entries $\sqrt{2}S_{ik}S_{i\ell}$ with $1 \le k < \ell \le q$. Next, we introduce the matrix

$$V = \frac{2}{q(q+2)}\begin{bmatrix} A_q & 0 \\ 0 & I_{q(q-1)/2}\end{bmatrix}.$$

Note that $(1/2)q(q+2)V$ is idempotent with rank $q-1+q(q-1)/2 = (q-1)(q+2)/2$. Thus we calculate

$$\operatorname{tr}(V) = \frac{q-1}{q}, \quad \operatorname{tr}(V^2) = \frac{2(q-1)}{q^2(q+2)} \quad \text{and} \quad \operatorname{tr}(V^4) = \frac{8(q-1)}{q^4(q+2)^3}.$$

In view of this we can express the test statistics as

$$S(m,q) = \frac{2}{m}\sum_{1\le i<j\le m}\frac{T_i^\top T_j}{\sqrt{2\operatorname{tr}(V^2)}}$$

with $T_i = (X_i^\top, Y_i^\top)^\top$. We shall show that T_1 is centered and has dispersion matrix V, and that the following statements hold:

$$\operatorname{var}(|V^{1/2}T_1|^2) = \operatorname{var}(T_1^\top V T_1) = 0 \tag{33}$$

and

$$\mathrm{E}\{(T_1^\top T_2)^4\} = O(1/q^4). \tag{34}$$

Then Corollary 1 applied with $n = m$, $r_n = q(q+1)/2$ and $\xi_{n,i} = T_i$ yields the desired result.

We write $X_1 = (U_1^2 - 1/q,\dots,U_q^2 - 1/q)$ and denote the entries of Y_1 by U_kU_ℓ for $1 \le k < \ell \le q$, where $U = (U_1,\dots,U_q)^\top$ is uniformly distributed on the unit sphere in $\mathbb{R}^q$. Then the random variables $\operatorname{sign}(U_1),\dots,\operatorname{sign}(U_r)$ are independent Rademacher random variables and $(\operatorname{sign}(U_1),\dots,\operatorname{sign}(U_q))^\top$ is independent of $(U_1^2,\dots,U_q^2)^\top$ which has a Dirichlet distribution with parameter $(1/2,\dots,1/2)$. From this we derive that $\mathrm{E}(U_{i_1}^{\alpha_{i_1}}\cdots U_{i_r}^{\alpha_{i_r}}) = 0$ holds for all distinct indices $i_1,\dots,i_r$ and integers $\alpha_1,\dots,\alpha_r$ as long as one among the αs is odd. In particular, we have

$$\mathrm{E}(U_1U_2) = 0, \quad \mathrm{E}(U_1U_2^2U_3) = 0, \quad \mathrm{E}(U_1U_2U_3U_4) = 0, \quad \mathrm{E}(U_1U_2^3) = 0,$$

$$\mathrm{E}(U_1^2) = \frac{1}{q}, \quad \mathrm{E}(U_1^4) = \frac{3}{q(q+2)} \quad \text{and} \quad \mathrm{E}(U_1^2U_2^2) = \frac{1}{q(q+2)}.$$

This shows that $\mathrm{E}(T_1) = 0$ and has dispersion matrix V. Since $|U| = 1$, we find $J_qX_1 = 0$,

$$X_1^\top A_q X_1 = X_1^\top X_1 = |X_1|^2 = \sum_{k=1}^{q} U_k^4 - \frac{1}{q} \tag{35}$$

and

$$Y_1^\top Y_1 = \sum_{k\ne\ell} U_k^2U_\ell^2 = 1 - \sum_{k=1}^{n} U_k^4$$

and thus obtain the identity

$$T_1^\top V T_1 = \frac{2}{q(q+2)}(X_1^\top A_q X_1 + Y_1^\top Y_1) = \frac{2(q-1)}{q^2(q+2)}.$$

This yields (33). Finally, using the inequality $(a+b)^4 \le 8(a^4+b^4)$, the Cauchy–Schwarz inequality, the identity (35) and the independence of T_1 and T_2, we derive

$$\mathrm{E}\{(T_1^\top T_2)^4\} \le 8\mathrm{E}\{(X_1^\top X_2)^4\} + 8\mathrm{E}\{(Y_1^\top Y_2)^4\},$$

$$\mathrm{E}\{(X_1^\top X_2)^4\} \le \mathrm{E}(|X_1|^4|X_2|^4) = \{\mathrm{E}(|X_1|^4)\}^2,$$

$$\mathrm{E}(|X_1|^4) \le \mathrm{E}\left\{\left(\sum_{k=1}^{q} U_k^4\right)^2\right\} \le q^2\mathrm{E}(U_1^8) = \frac{105q^2}{q(q+2)(q+4)(q+6)} \le \frac{105}{q^2},$$

Table 6
Simulated significance levels and powers for $S(m, q)$ when $\alpha = 0.05$.

	$\theta = 0$			$\theta = \pi/21$		
	$m = 1000$	$m = 5000$	$m = 10{,}000$	$m = 1000$	$m = 5000$	$m = 10{,}000$
$q + 1 = 100$	.0530	.0490	.0510	.5750	.5740	.5700
$q + 1 = 150$	.0520	.0595	.0530	.8810	.8955	.8875
$q + 1 = 200$	.0500	.0495	.0475	.9865	.9885	.9865

and

$$\mathrm{E}\{(Y_1^\top Y_2)^4\} \le \sum_{k_1\ne\ell_1}\sum_{k_2\ne\ell_2}\sum_{k_3\ne\ell_3}\sum_{k_4\ne\ell_4}\{\mathrm{E}(U_{k_1}U_{\ell_1}U_{k_2}U_{\ell_2}U_{k_3}U_{\ell_3}U_{k_4}U_{\ell_4})\}^2. \tag{36}$$

The only non-zero expectations of the right-hand side of (36) equal $\mathrm{E}(U_1^4U_2^4)$, $\mathrm{E}(U_1^2U_2^2U_3^4)$ or $\mathrm{E}(U_1^2U_2^2U_3^2U_4^2)$ all of which are bounded by $\mathrm{E}(U_1^8)$, and the number of non-zero expectations is of order q^4. This shows that $\mathrm{E}\{(Y_1^\top Y_2)^4\} = O(1/q^4)$. The above show that (34) holds. □

We performed a small simulation study to evaluate this test. We generated m-dimensional normal random vectors with mean vector zero and dispersion matrix $\Sigma(\theta)$ with $\Sigma_{i,i}(\theta) = 1$ and $\Sigma_{i,j}(\theta) = \sin(\theta)\cos(\theta)$ for $|i-j| = 1$ and zero for $|i-j| > 1$. Table 6 gives simulated significance levels ($\theta = 0$) and powers ($\theta = \pi/21$) for selected values of $q + 1$ (namely 100, 150 and 200) and of dimensions m (namely 1000, 5000, 10,000). The results are based on 2000 repetitions.

9. An auxiliary lemma

Our proofs of the theorems will rely on the following simple lemma.

Lemma 1. *Let $X_1, \dots, X_m$ be independent and identically distributed random vectors with zero mean and dispersion matrix V and, for each $k \in \{1, \dots, m\}$, set $S_k = X_1 + \cdots + X_k$. Then one has, for all $k \in \{1, \dots, m\}$, $\mathrm{E}(|S_k|^2) = k\mathrm{E}(|X_1|^2) = k\,\mathrm{tr}(V)$ and, for all $\epsilon > 0$, $\epsilon^2 \Pr(\max_{1\le k\le m}|S_k| > \epsilon) \le \mathrm{E}(|S_m|^2) = m\,\mathrm{tr}(V)$. If also $\mathrm{E}(|X_1|^4)$ is finite, then one has*

$$\forall_{k\in\{1,\dots,m\}} \quad \mathrm{var}(|S_k|^2) = 2k(k-1)\,\mathrm{tr}(V^2) + k\mathrm{var}(|X_1|^2)$$

and

$$\mathrm{var}\left(\sum_{k=1}^m |S_k|^2\right) \le 2m^4\mathrm{tr}(V^2) + 2m^3\mathrm{var}(|X_1|^2).$$

Proof. The first inequality is the Kolmogorov inequality for random vectors. Let $X = X_1$ and $Y = X_2$. Then

$$\mathrm{E}(|X|^2) = \mathrm{E}\{\mathrm{tr}(XX^\top)\} = \mathrm{tr}\{\mathrm{E}(XX^\top)\} = \mathrm{tr}(V)$$

and

$$\mathrm{E}\{(X^\top Y)^2\} = \mathrm{E}\{\mathrm{tr}(X^\top YY^\top X)\} = \mathrm{E}\{\mathrm{tr}(YY^\top XX^\top)\} = \mathrm{tr}\{\mathrm{E}(YY^\top)\mathrm{E}(XX^\top)\} = \mathrm{tr}(V^2).$$

Using independence we calculate

$$\begin{aligned}
\mathrm{E}(|S_k|^2) &= \sum_{i=1}^k\sum_{j=1}^k \mathrm{E}(X_i^\top X_j) = \sum_{i=1}^k \mathrm{E}(|X_i|^2) = k\mathrm{E}(|X|^2) = k\,\mathrm{tr}(V),\\
\mathrm{E}(|S_k|^4) &= \sum_{i=1}^k\sum_{j=1}^k\sum_{\ell=1}^k\sum_{p=1}^k \mathrm{E}(X_i^\top X_jX_l^\top X_p) = 4\sum_{1\le i<j\le k}\mathrm{E}\{(X_i^\top X_j)^2\} + \sum_{i=1}^k \mathrm{E}(|X_i|^4) + 2\sum_{1\le i<p\le k}\mathrm{E}(|X_i|^2)\mathrm{E}(|X_p|^2)\\
&= 2k(k-1)\mathrm{tr}(V^2) + k\mathrm{E}(|X|^4) + k(k-1)\{\mathrm{tr}(V)\}^2,
\end{aligned}$$

and hence we obtain the desired form of $\mathrm{var}(|S_k|^2)$. It is easy to see that the covariance of $|S_i|^2$ and $|S_j|^2$ equals the variance of $|S_{\min(i,j)}|^2$. Thus we obtain

$$\mathrm{var}\left(\sum_{k=1}^m |S_k|^2\right) = \sum_{k=1}^m \mathrm{var}(|S_k|^2)\{1 + 2(m-k)\} = \sum_{k=1}^m \{1 + 2(m-k)\}k\{2(k-1)\mathrm{tr}(V^2) + \mathrm{var}(|X|^2)\}$$

and hence the desired bound on the variance of $|S_1|^2 + \cdots + |S_m|^2$. □

10. Proof of Theorems 1 and 2

To simplify notation we abbreviate $\xi_{n,j}$ by ξ_j and $\{\mathrm{tr}(V_n^2)\}^{1/2}$ by σ_n and introduce r_n-dimensional random vectors $D_0 = 0$ and, for all $j \in \{1, \ldots, n\}$,

$$D_j = \frac{\sqrt{2}}{n\sigma_n} \sum_{i=1}^{j} \xi_i.$$

Then we can write the terms $T_{n,1}$ and $T_{n,3}$ from the Introduction as

$$T_{n,1} = \frac{1}{n\sigma_n} \sum_{j=1}^{n} \{|\xi_j|^2 - \mathrm{E}(|\xi_j|^2)\} \quad \text{and} \quad T_{n,3} = \sum_{j=1}^{n} D_{j-1}^\top \xi_j.$$

By what we have shown there, the desired result follows if we verify that $T_{n,1}$ converges to zero in probability and that $T_{n,3}$ is asymptotically standard normal. The latter follows from the martingale central limit theorem (see, e.g., part (a) of Theorem 2.5 of [8], or Corollary 3.1 in [7] and the ensuing remarks) if we verify that

$$\forall_{j\in\{1,\ldots,n\}} \quad \mathrm{E}_{j-1}(D_{j-1}^\top \xi_j) = 0, \tag{37}$$

$$\sum_{j=1}^{n} \mathrm{E}_{j-1}\{(D_{j-1}^\top \xi_j)^2\} = 1 + o_p(1) \tag{38}$$

and, for $\epsilon > 0$,

$$\sum_{j=1}^{n} \mathrm{E}_{j-1}(|D_{j-1}^\top \xi_j|^2 \mathbf{1}\{|D_{j-1}^\top \xi_j| > \epsilon\}) = o_p(1), \tag{39}$$

with E_{j-1} the conditional expectation given $\xi_1, \ldots, \xi_{j-1}$. Of course, (37) is a simple consequence of the independence of the random vectors $\xi_1, \ldots, \xi_n$.

Proof of Theorem 1. We have $T_{n,1} = o_p(1)$ in view of (5) and the identity

$$\mathrm{E}(T_{n,1}^2) = \frac{\mathrm{var}(|\xi_1|^2)}{n\sigma_n^2}.$$

The left-hand side of (38) equals

$$S_n = \sum_{j=1}^{n} D_{j-1}^\top V_n D_{j-1} = \frac{2}{n^2\sigma_n^2} \sum_{j=1}^{n-1} \Big| \sum_{i=1}^{j} V_n^{1/2} \xi_i \Big|^2.$$

Note that the random vector $V_n^{1/2}\xi_1$ is centered and has dispersion matrix V_n^2. Thus, with the aid of Lemma 1, (C2) and (6), we find

$$\mathrm{E}(S_n) = \sum_{j=1}^{n-1} \frac{2j}{n^2} = \frac{n-1}{n} \to 1$$

and

$$\mathrm{var}(S_n) \le \frac{8\mathrm{tr}(V_n^4)}{\sigma_n^4} + \frac{8\mathrm{var}(|V^{1/2}\xi_1|^2)}{n\sigma_n^4} \to 0.$$

This shows that $S_n = 1 + o_p(1)$. Finally, the expected value of the left-hand side of (39) is bounded by U_n/ϵ^2 with

$$U_n = \sum_{j=1}^{n} \mathrm{E}(|D_{j-1}^\top \xi_j|^4) = \frac{4}{n^4\sigma_n^4} \sum_{j=2}^{n} \mathrm{E}\left\{ \left(\sum_{i=1}^{j-1} \xi_i^\top \xi_j \right)^4 \right\}.$$

Conditioning in the expectation with index j on ξ_j, we obtain with the aid of Lemma 1,

$$\begin{aligned} U_n &\le \frac{4}{n^4\sigma_n^4} \sum_{j=2}^{n} \Big[3(j-1)(j-2)\mathrm{E}\{(\xi_1^\top V_n \xi_1)^2\} + (j-1)\mathrm{E}\{(\xi_1^\top \xi_2)^4\} \Big] \\ &\le \frac{4}{n^4\sigma_n^4} \Big[3n^3 \Big\{ \mathrm{var}(|V_n^{1/2}\xi_1|^2) + \mathrm{tr}^2(V_n^2) \Big\} + n^2 \mathrm{E}(|\xi_1^\top \xi_2|^4) \Big]. \end{aligned}$$

It follows from (6) and (7) that U_n converges to zero. This proves (39) and completes the proof of Theorem 1. □

Proof of Theorem 2. By the Lindeberg condition (L), there are positive numbers ϵ_n such that $\epsilon_n \to 0$ and $L_n(\epsilon_n) \to 0$. We can write $T_{n,1} = R_{n,1} + R_{n,2}$, where

$$R_{n,1} = \frac{1}{n\sigma_n} \sum_{j=1}^{n} \Big[|\xi_j|^2 \mathbf{1}\{|\xi_j| \le \epsilon_n \sqrt{n}\} - \mathrm{E}(|\xi_j|^2 \mathbf{1}\{|\xi_{n,j}| \le \epsilon_n \sqrt{n}\}) \Big],$$

$$R_{n,2} = \frac{1}{n\sigma_n} \sum_{j=1}^{n} \Big[|\xi_j|^2 \mathbf{1}\{|\xi_j| > \epsilon_n \sqrt{n}\} - \mathrm{E}(|\xi_j|^2 \mathbf{1}\{|\xi_j| > \epsilon_n \sqrt{n}\}) \Big],$$

and calculate $\mathrm{E}(|R_{n,2}|) \le 2L_n(\epsilon_n)/\sigma_n$ and

$$\mathrm{E}(R_{n,1}^2) \le \frac{\mathrm{E}(|\xi_1|^4 \mathbf{1}\{|\xi_1| \le \epsilon_n \sqrt{n}\})}{n\sigma_n^2} \le \frac{\epsilon_n^2 \mathrm{E}(|\xi_1|^2)}{\sigma_n^2} = \epsilon_n^2 \frac{\mathrm{tr}(V_n)}{\mathrm{tr}(V_n^2)}.$$

This shows that $T_{n,1} = o_p(1)$.

Next, we show

$$D_n^* = \max_{1 \le j \le n} \sqrt{n}\, |D_{j-1}| = O_p(1) \quad \text{and} \quad \sum_{j=1}^{n} |D_{j-1}|^2 = O_p(1). \tag{40}$$

Indeed, with the help of Lemma 1 we obtain, for all $K > 0$,

$$\Pr(D_n^* > \sqrt{2}K) = \Pr\left(\max_{1 \le j \le n} \Big| \sum_{i=1}^{j-1} \xi_i \Big| > K\sigma_n \sqrt{n} \right) \le \frac{1}{K^2} \frac{\mathrm{tr}(V_n)}{\mathrm{tr}(V_n^2)},$$

and

$$\sum_{j=1}^{n} \mathrm{E}(|D_{j-1}|^2) = \frac{2}{n^2\sigma_n^2} \sum_{j=1}^{n} (j-1)\mathrm{tr}(V_n) \le \frac{\mathrm{tr}(V_n)}{\mathrm{tr}(V_n^2)}.$$

The statements (40) imply (39), since the left-hand side of (39) is bounded by

$$\sum_{j=1}^{n} |D_{j-1}|^2 \int y^2 \mathbf{1}\{|D_{j-1}| y > \epsilon\}\, dF_n(y) \le \int y^2 \mathbf{1}\{D_n^* y > \epsilon \sqrt{n}\}\, dF_n(y) \sum_{j=1}^{n} |D_{j-1}|^2,$$

where F_n is the distribution of $|\xi_1|$.

Finally, we obtain (38) by verifying

$$S_n = \frac{2}{n^2\sigma_n^2} \sum_{j=2}^{n} \Big| \sum_{i=1}^{j-1} V_n^{1/2} \xi_i \Big|^2 = 1 + o_p(1). \tag{41}$$

For this we write $\xi_j = X_j + Y_j$ with

$$X_j = \xi_j \mathbf{1}\{|\xi_j| \le \epsilon_n \sqrt{n}\} - \mathrm{E}(\xi_j \mathbf{1}\{|\xi_j| \le \epsilon_n \sqrt{n}\}),$$

$$Y_j = \xi_j \mathbf{1}\{|\xi_j| > \epsilon_n \sqrt{n}\} - \mathrm{E}(\xi_j \mathbf{1}\{|\xi_j| > \epsilon_n \sqrt{n}\}).$$

In view of the Cauchy–Schwarz inequality, the desired (41) follows from the statements

$$S_{n,1} = \frac{2}{n^2\sigma_n^2} \sum_{i=1}^{n} \Big| \sum_{i=1}^{j-1} V_n^{1/2} X_j \Big|^2 = 1 + o_p(1)$$

and

$$S_{n,2} = \frac{2}{n^2\sigma_n^2} \sum_{j=1}^{n} \Big| \sum_{i=1}^{j-1} V_n^{1/2} Y_j \Big|^2 = o_p(1).$$

The latter follows from the bound

$$\mathrm{E}(S_{n,2}) = \frac{2}{n^2\sigma_n^2} \sum_{j=1}^{n} (j-1)\mathrm{E}(|V_n^{1/2} Y_1|^2) \le \frac{\mathrm{tr}(V_n)}{\mathrm{tr}(V_n^2)} \mathrm{E}(|Y_1|^2) \le \frac{\mathrm{tr}(V_n)}{\mathrm{tr}(V_n^2)} L_n(\epsilon_n).$$

The former follows if we show $\mathrm{E}(S_{n,1}) \to 1$ and $\mathrm{var}(S_{n,1}) \to 0$. We calculate

$$\mathrm{E}(S_{n,1}) = \frac{(n-1)}{n\sigma_n^2} \mathrm{E}(|V_n^{1/2} X_1|^2) = \frac{(n-1)}{n} \frac{\mathrm{tr}(W_n)}{\sigma_n^2}$$

with $W_n = V_n^{1/2}\mathrm{E}(X_1X_1^\top)V_n^{1/2}$ the dispersion matrix of $V_n^{1/2}X_1$. We have the identity

$$W_n = V_n^{1/2}\{V_n - \mathrm{E}(Z_nZ_n^\top) - \mathrm{E}(Z_n)\mathrm{E}(Z_n^\top)\}V_n^{1/2}$$

with $Z_n = \xi_1\{\mathbf{1}|\xi_1| > \sqrt{n}\}$ and obtain the inequality

$$\mathrm{tr}(V_n^2) - 2L_n(\epsilon_n)\mathrm{tr}(V_n) \le \mathrm{tr}(W_n) \le \mathrm{tr}(V_n^2).$$

This lets us conclude $\mathrm{E}(S_{n,1}) \to 1$. Lemma 1 and the inequalities $\mathrm{tr}(W_n^2) \le \mathrm{tr}(V_n^4)$, $\mathrm{E}(|X_1|^2) \le \mathrm{tr}(V_n)$ and

$$|V_n^{1/2}X_1|^4 = (X_1^\top V_nX_1)^2 \le |X_1|^2|V_nX_1|^2 \le \sigma_n^2|X_1|^4 \le \sigma_n^2 4\epsilon_n^2 n|X_1|^2$$

yield

$$\mathrm{var}(S_{n,1}) \le \frac{8}{n^4\sigma_n^4}(n^4\mathrm{tr}(W_n^2) + n^3\mathrm{E}(|V_n^{1/2}X_1|^4]) \le \frac{8\mathrm{tr}(V_n^4)}{\mathrm{tr}^2(V_n^2)} + \frac{32\epsilon_n^2\mathrm{tr}(V_n)}{\mathrm{tr}(V_n^2)}.$$

This completes the proof of Theorem 2. □

Acknowledgments

The authors would like to thank the Editor-in-Chief, the Associate Editor and the two reviewers for their valuable comments and suggestions. The research of Hanxiang Peng was partially supported by National Science Foundation Grant DMS 0940365. The research of Anton Schick was partially supported by National Science Foundation Grant DMS 0906551.

References

[1] Z. Bai, H. Saranadasa, Effect of high dimension: By an example of a two sample problem, Statist. Sinica 6 (1996) 311–329.
[2] J. Chang, S.X. Chen, X. Chen, High dimensional generalized empirical likelihood for moment restrictions for dependent data, J. Econometrics 185 (2015) 283–304.
[3] S.X. Chen, L. Peng, Y.L. Qin, Effects of data dimension on empirical likelihood, Biometrika 96 (2009) 711–722.
[4] S.X. Chen, L.X. Zhang, P.S. Zhong, Tests for high-dimensional covariance matrices, J. Amer. Statist. Assoc. 105 (2010) 810–819.
[5] L. Giraitis, M. Taniguchi, M.S. Taqqu, Asymptotic normality of quadratic forms with martingale differences, Stat. Inference Stoch. Process. 20 (2017) 315–327.
[6] P. Guttorp, R.A. Lockhart, On the asymptotic distribution of quadratic forms and uniform order statistics, Ann. Statist. 16 (1988) 433–449.
[7] P. Hall, C.C. Heyde, Martingale Limit Theory and its Application, Academic Press, 1980.
[8] I. Helland, Central limit theorems for martingales with discrete or continuous time, Scand. J. Stat. 9 (1982) 79–94.
[9] N.L. Hjort, I.W. McKeague, I. Van Keilegom, Extending the scope of empirical likelihood, Ann. Statist. 37 (2009) 1079–1111.
[10] A. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (1988) 237–249.
[11] A. Owen, Empirical likelihood ratio confidence regions, Ann. Statist. 18 (1990) 90–120.
[12] A. Owen, Empirical Likelihood, Chapman & Hall/CRC, London, 2001.
[13] H. Peng, A. Schick, Efficient estimation of linear functionals of a bivariate distribution with equal, but unknown, marginals: The least squares approach, J. Multivariate Anal. 95 (2005) 385–409.
[14] H. Peng, A. Schick, An empirical likelihood approach to goodness of fit testing, Bernoulli 19 (2013) 2250–2276.
[15] S. Portnoy, Behavior of likelihood methods for exponential families when the number of parameters tends to infinity, Ann. Statist. 16 (1988) 356–366.
[16] V.I. Rotar, On the distribution of a quadratic form in many random variables, Theory Probab. Appl. 20 (1973) 880–882.
[17] V.I. Rotar, Some limit theorems for polynomials of second degree, Theory Probab. Appl. 18 (1973) 499–507.
[18] V.I. Rotar, Limit theorems for polylinear forms, J. Multivariate Anal. 9 (1979) 511–530.
[19] J.R. Schott, Testing for complete independence in high dimensions, Biometrika 92 (2005) 951–956.