Exponential bounds for regularized Hotelling's T_n^2 statistic in high dimension

El Mehdi Issouani[a,*], Patrice Bertail[a], Emmanuelle Gautherat[b]

[a]*MODAL'X UMR 9023, UPL University Paris Nanterre.*
[b]*Université de Reims Champagne Ardenne, REGARDS EA 6292, 51097 Reims, France*

Abstract

We obtain exponential inequalities for regularized Hotelling's T_n^2 statistics, that take into account the potential high dimensional aspects of the problem. We explore the finite sample properties of the tail of these statistics by deriving exponential bounds for symmetric distributions and also for general distributions under weak moment assumptions (we never assume exponential moments). For this, we use a penalized estimator of the covariance matrix and propose an optimal choice for the penalty coefficient.

Keywords: Eigenvalues, exponential bounds, high dimension, Hotelling statistics, penalized covariance, regularized covariance matrix, self-normalization

1. Introduction

In many applications (for instance in genomics or natural language processing), the dimension of the parameter of interest q is large in comparison to the sample size n and sometimes is increasing with n. Consider for instance the problem of estimating or testing a mean of variables in $\mathbb{R}^q$, with $q > n$; in that case, the empirical covariance matrix is not full rank and does not even converge to the true one when n tends to infinity and is ill-conditioned (see Johnstone (2001) [8]). As a consequence, the usual Hotelling's T_n^2 tests in a large dimension framework are no longer valid. It is thus important to construct estimators and testing procedures that take into account the high dimensional aspects of the problem (as done for instance in Ledoit and Wolf (2000, 2022) [11, 12], see also the references therein). One relevant proposition which has been developed in the statistical literature is to use a penalized estimator of the covariance matrix which is non-singular and to use this matrix in tests. In that spirit, Chen et al. (2011) [5] have obtained asymptotically valid regularized Hotelling's T_n^2 tests for the mean in the Gaussian case in a high dimensional framework, when n and $q \equiv q(n)$ tend to infinity at some specific rate. Li et al. (2020) [14] have extended these results to some specific sub-gaussian distribution. The purpose of this paper is to further explore the finite sample properties of such tests by deriving exponential bounds of some correctly regularized Hotelling's T_n^2 under general distributions, including ones with very few moments.

Such bounds allow to build conservative confidence regions for the parameter of interest. They are also of interest in statistical learning to control risk even with unbounded loss functions. For this, we derive exponential bounds for some regularized Hotelling's T_n^2 statistics in the spirit of Bertail et al (2008) [2], who obtained bounds for self-normalized quadratic forms or the Hotelling's T_n^2 statistic when $q < n$. We show that for symmetric distributions, only moments of order 2 are needed and we only assume the existence of moments of order 8 for general distributions.

Let $Z, Z_1, \ldots, Z_n$ be i.i.d. centered random vectors with probability distribution P, defined on a probability space $(\Omega, A, \mathbb{P})$ with values in $\left(\mathbb{R}^{q(n)}, \mathcal{B}, P\right)$ endowed with the L_2 norm $\|.\|_2$. We denote $\mathbb{E}$ the expectation under P. Put $Z^{(n)} = (Z_i)_{1\le i\le n}$. As n and $q(n)$ go to infinity, notice that actually $(Z^{(n)})_n$ defines a triangular array of random variables with varying dimensions. However, since we are interested in finite sample properties, we will drop the dependence in n. In particular, we use q instead of $q(n)$. But keep in mind that q is a function of n in an asymptotic framework.

*Corresponding author. Email address: elmehdi.issouani@gmail.com

The covariance matrix of the observation is given by $S^2 = \mathbb{E}(ZZ')$, where we denote by Z' the transpose of Z and S the square root of S^2. Denote by $\bar{Z}_n = n^{-1}\sum_{i=1}^n Z_i$ the sample mean. The sample covariance matrix is defined here by

$$S_n^2\left(Z^{(n)}\right) = \frac{1}{n}\sum_{i=1}^n Z_i Z_i'.$$

To simplify notations, we denote the sample covariance matrix of Z_i's by S_n^2 when there is no confusion. Notice that we do not center by the empirical mean.

We recall that Hotelling's T_n^2, which can be seen as a quadratic form of self-normalized sums, is given by

$$T_n^2 = n\bar{Z}_n' S_n^{-2}\bar{Z}_n,$$

when $q < n$ and $S_n^{-2} = \left(S_n^2\right)^{-1}$. For some nonnegative real numbers, ρ_1 and ρ_2, define Σ_n^2 the linear combination of the identity matrix with the sample covariance matrix

$$\Sigma_n^2 \equiv \Sigma_n^2(\rho_1, \rho_2) = \rho_1 I_q + \rho_2 S_n^2,$$

with I_q the identity matrix of size q. For $\rho_1 = 0$ and $\rho_2 = 1$, $\Sigma_n^2(0, 1) = S_n^2$ is the empirical covariance matrix, which is singular for $q > n$. When $\rho_2 = 1$ and $\rho_1 > 0$ (and small), Σ_n^2 corresponds to a Tikhonov regularization of the sample covariance matrix: see Tikhonov (1963) [20]. It is precisely this estimator which is used in the tests proposed by Chen et al (2011) [5]. However, it is shown in Ledoit and Wolf (2000) [11] that if one chooses adequately ρ_1 and ρ_2 then one can obtain a well-conditioned estimator of the covariance matrix which is invertible and more accurate than the sample covariance for some L_2-distance.
We denote by Σ^2 the expectation of Σ_n^2, which is given by

$$\Sigma^2 \equiv \Sigma^2(\rho_1, \rho_2) = \rho_1 I_q + \rho_2 S^2.$$

Actually, such modification ensures that we can control the distance between Σ_n^2 and S^2: this will be fundamental to obtain exponential bounds.

In the following, we are interested in the Hotelling's T_n^2 statistic with a linear combination of the sample covariance and the identity, that we now call the regularized Hotelling's T_n^2 statistic defined by

$$T_n^2(\rho_1, \rho_2) = n\bar{Z}_n'\Sigma_n^{-2}(\rho_1, \rho_2)\bar{Z}_n$$

generalizing the proposal of Chen et al (2011)[5].

In the framework of high dimension, such quantities also appear naturally when studying empirical likelihood under a lot of constraints, penalized in its dual form by an L2-norm: see for instance Newey and Smith (2004) [16], Lahiri and Mukhopadhyay (2012), [9], Carrasco and Kontchoni (2017) [3] among others.

When $q < n$, exponential bounds for $T_n^2(0, 1)$ (that is, with the empirical covariance matrix instead of a regularized one) have been obtained by Bertail et al (2008) [2]. Their exponential bound is controlled by two terms: (1) an exponential term corresponding to a "Hoeffding" or Pinelis (1994) [18] type of inequality applied to a symmetrized version of the observations and (2) an exponential bound which essentially controls the minimum eigenvalue of the sample covariance matrix and the proximity of S_n^2 to S^2. However, for $q \geq n$ such inequality can not hold since in that case the minimum eigenvalues of S_n^2 is always 0. Moreover, it can easily be seen from the results of [2] that the bound becomes very bad when $q > n$ or/and when q and n are of the same order. We obtain in this paper general results with an adequate choice of ρ_1 and ρ_2 when q is bigger than n and when q and n are such that $\frac{q}{n} \to l \in]0, \infty[$.

The paper is divided into four parts including this introduction. In the second part, we recall some known exponential inequalities for $q < n$ under weak moments assumptions. Then we obtain an oracle exponential inequality for the regularized Hotteling's T_n^2, assuming that the values ρ_1 and ρ_2 are fixed and known. Some interesting sharp bounds which may be useful in statistical learning assuming symmetry are obtained for any n and q large. We then establish a general inequality for $q = O(n)$ for non-symmetric distributions under a few moments' assumptions. In the third part, we estimate the optimal values ρ_1^* and ρ_2^* and show that the inequality remains valid up to some additional small terms controlling the concentration of these estimators around their true value. We illustrate our results with some simulations in the supplementary material.

2. Oracle exponential bounds for regularized Hotelling's T_n^2

In the following, we define the penalized Hotelling's T_n^2 as the particular regularized Hotelling's statistic $T_n^2(\rho, 1)$ with $\rho \geq 0$. The aim of this section is to establish an oracle exponential inequality of the distribution of the penalized Hotelling's T_n^2 in the case $q \geq n$ and when the distribution of the data is symmetric (Theorem 1 and Theorem 2) as well as in the general case, that is when the distribution is not necessarily symmetric (Theorem 3).

2.1. Known bounds for Hotelling's T_n^2

Some bounds for T_n^2 or self-normalized sums may be quite easily obtained in the symmetric case (that is for random variables having a symmetric distribution see Pinelis (1994) [18]) and are well-known in the unidimensional case $q = 1$. In non-symmetric and/or multidimensional cases with $q < n$, these bounds are new and not trivial to prove. One of the main tools for obtaining exponential inequalities in various settings is the famous Hoeffding inequality (see Hoeffding (1994) [7]). For centered independent real random variables $Y_1, \ldots, Y_n$, that are bounded, say $|Y_i| < 1$, for all $i \in \{1, \ldots, n\}$, we have, for $a_i \in [-1, 1]$ such that $\sum a_i^2 = 1$,

$$\forall t > 0, \quad \mathbb{P}\left(\left(\sum_{i=1}^{n} a_i Y_i\right)^2 \geq t\right) \leq 2\exp\left(-\frac{t}{2}\right).$$

A direct application of this inequality to self-normalized sums (via a randomization step introducing independent Rademacher r.v.'s ε_i) yields that, for independent real ($q = 1$) symmetric random variables Z_i, $i \in \{1, \ldots, n\}$ and not necessarily bounded (nor identically distributed). Indeed, we have by putting $Y_i = \varepsilon_i$ and $a_i = \frac{Z_i}{(\sum Z_i^2)^{1/2}}$; for $t > 0$,

$$\mathbb{P}\left(T_n^2 \geq t\right) = \mathbb{P}\left(\frac{\left(\sum_{i=1}^n Z_i\right)^2}{\sum_{i=1}^n Z_i^2} \geq t\right) = \mathbb{P}\left(\frac{\left(\sum_{i=1}^n Z_i \varepsilon_i\right)^2}{\sum_{i=1}^n Z_i^2} \geq t\right) = \mathbb{E}\left[\mathbb{P}\left(\frac{\left(\sum_{i=1}^n Z_i \varepsilon_i\right)^2}{\sum_{i=1}^n Z_i^2} \geq t \,\middle|\, (Z_i)_{i\in\{1,\ldots,n\}}\right)\right] \leq 2\exp\left(-\frac{t}{2}\right).$$

Pinelis (1994) [18] has obtained with a different technique, a sharp χ^2 type of bounds which generalizes this kind of results for multivariate data when $q < n$. He proved that, if Z has a symmetric distribution, without any moment assumption on the variables Z_i, then one has

$$\forall t > 0, \quad \mathbb{P}\left(T_n^2 \geq t\right) \leq \frac{2e^3}{9}\overline{F}_q(t), \tag{1}$$

where $F_q(t)$ is the cumulative distribution function (cdf) of a $\chi^2(q)$ distribution. The density is denoted by f_q. A crude approximation yields that for t large enough,

$$\mathbb{P}\left(T_n^2 \geq t\right) \leq \frac{e^3}{9}\frac{2^{2-\frac{q}{2}}}{\Gamma(\frac{q}{2})} t^{\frac{q}{2}-1}\exp(-t/2),$$

where $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt$ is the gamma function.

Notice that, for $q = 1$ this bound (only valid for large t) is better than the crude Hoeffding bound since we recover the missing factor $\frac{1}{\sqrt{t}}$ in front of the exponential (see Talagrand (1995) [19]). When $q < n$, using a multidimensional version of Panchenko's symmetrization lemma (Panchenko (2003) [17]) Bertail et al (2008) [2] have obtained an exponential inequality for the general distribution of Z with finite kurtosis $\gamma_4 = \mathbb{E}\left(\left\|S^{-1}Z\right\|_2^4\right)$. More precisely, they establish that under $0 < \gamma_4 < \infty$,

(i) for $t > nq$, $\mathbb{P}\left(T_n^2 \geq t\right) = 0$.

(ii) for any $a > 1$, and any nonnegative t such that $2q(1 + a) \leq t \leq nq$, the following bound holds for some is an explicit constant $C(q)$,

$$\mathbb{P}(T_n^2 \geq t) \leq \frac{2e^3}{9\Gamma(\frac{q}{2}+1)}\left(\frac{t - q(1+a)}{2(1+a)}\right)^{\frac{q}{2}} \exp\left(-\frac{t - q(1+a)}{2(1+a)}\right) + C(q)n^{3-\frac{6}{q+1}}\exp\left(-\frac{n\left(1-\frac{1}{a}\right)^2}{\gamma_4(q+1)}\right).$$

The first term is essentially equivalent to the tail of a $\chi^2(q)$ distribution (up to an explicit constant), while the second term controls the speed of convergence of S_n^2 to S^2, when $\gamma_4 < \infty$. The constant a controls the balance between these two terms on the right-hand side of the inequality and may be optimized. Notice that this second exponential term is small when $q << n$ but explodes in n^3 if $q/n \to l > 0$ for large n, making this bound totally useless in that case.

In the general multidimensional framework considered in Bertail et al (2008) [2] and in this paper, the main difficulty is to keep the self-normalized structure when symmetrizing the initial sum. In the next sections, the results of Bertail et al (2008) [2] obtained for $q < n$ are extended to the case $q \geq n$ by using a regularized version of S_n^2. This inequality is based on an appropriate diagonalization of the regularized sample covariance matrix which allows applying Pinelis (1994)'s inequality [18] (see section 2.2). This crude inequality is refined in section 2.3. When dealing with the general case (see section 2.4), we establish first a multivariate symmetrization lemma 3 in the spirit of Panchenko (2003) [17]. This symmetrization partially destroys the self-normalized structure (the normalization is then $\Sigma_n^2+\Sigma^2 = 2\Sigma_n^2+(\Sigma^2-\Sigma_n^2)$ instead of the expected normalization Σ_n^2), but the right standardization can be recovered (up to the factor 2) by obtaining a lower tail control of the distance between Σ_n^2 and Σ^2. To control this distance and make it as small as possible we will use the results of Ledoit and Wolf (2000) [11].

2.2. Bounds for regularized Hotelling's T_n^2 in a symmetric framework

We now obtain a simple inequality for the regularized Hotelling's T_n^2 in the symmetric case, based on previous results by Pinelis (1994) [18]. It essentially shows that the tail of the regularized Hotelling's T_n^2 is controlled by the tail of a $\chi^2(n)$ distribution.

Theorem 1. *Assume that Z has a symmetric distribution with finite covariance matrix then, without any additional moment assumption, we have, for any $n > 1$, for $t > n$, for any $\rho_1, \rho_2 > 0$,*

$$\mathbb{P}\left(T_n^2\left(\frac{\rho_1}{\rho_2}, 1\right) \geq t\right) = \mathbb{P}\left(n\bar{Z}_n'\Sigma_n^{-2}(\rho_1,\rho_2)\bar{Z}_n \geq \frac{t}{\rho_2}\right) \quad \leq \quad \frac{2e^3}{9}\bar{F}_n(t) \leq \frac{2e^3}{9}\exp\left(-\frac{(t-n)^2}{4t}\right), \tag{2}$$

where F_n is the cdf of a $\chi^2(n)$ distribution.
Moreover, for any $\rho > 0$, we have

$$\mathbb{P}\left(\frac{T_n^2(\rho,1)-n}{\sqrt{2n}} \geq t\right) = \mathbb{P}\left(\frac{n\bar{Z}_n'\Sigma_n^{-2}(\rho,1)\bar{Z}_n - n}{\sqrt{2n}} \geq t\right) \quad \leq \quad \frac{2e^3}{9}\exp\left(\frac{-t^2}{2\left(1+\sqrt{2}\frac{t}{\sqrt{n}}\right)}\right).$$

The inequality (2) yields a control of $T_n^2(\rho_1,\rho_2) = n\bar{Z}_n'\Sigma_n^{-2}(\rho_1,\rho_2)\bar{Z}_n$, when using a linear shrinkage estimator of the variance. This in turn can be simplified in (3), to a truly penalized Hotelling's T_n^2. Note that for any $\rho_1,\rho_2 > 0$,

$$\frac{\Sigma_n^2(\rho_1,\rho_2)}{\rho_2} = \frac{\rho_2 S_n^2 + \rho_1 I_q}{\rho_2} = S_n^2 + \frac{\rho_1}{\rho_2} I_q$$

and for any $\rho > 0$,

$$\Sigma_n^2(\rho,1) = S_n^2 + \rho I_q$$

is a penalized estimator of the covariance matrix. Inequality (3) can be interpreted as a Bernstein-type inequality.

Remark: These inequalities hold for any choice of ρ_1 and ρ_2. However for the inequalities to be sharp, ρ_1 and ρ_2 should be chosen adequately. First from the proof of Theorem 1, we see that the inequality is sharp only when ρ_1 is close to 0, which is in accordance with what we know about Tikhonov regularisation (1963) [20]. Actually when ρ_1 tends to $0, \Sigma_n^{-2}(\rho_1,\rho_2)$ is going to be identical to $\frac{1}{\rho_2}(S_n^2)^-$ where $(A)^-$ is the Moore-Penrose or generalized inverse of A (which is unique for symmetric matrices). Notice that the proof of the theorem and the inequality remain valid if we use $n\bar{Z}_n'\left(S_n^2\right)^-\bar{Z}_n$ rather than $n\bar{Z}_n'\Sigma_n^{-2}(\rho_1,\rho_2)\bar{Z}_n$. In the procedure of Chen et al. (2011) [5] this means that asymptotically there is no difference between standardizing by the regularized variance or by the generalized inverse of the covariance matrix. The regularization just serves as a trick to approximate the generalized inverse. However, the finite sample properties of the regularized Hotelling's T_n^2 will strongly depend on the choice of ρ_1 and ρ_2.

2.3. An improved bound for penalized Hotelling's T_n^2 in the symmetric case

It can be seen from the proof of Theorem 1 that the penalized Hotelling's T_n^2 statistic essentially behaves like a weighted sum of asymptotically χ^2 random variables. This also explains the results of Chen et al. (2015) [5]. Actually, we can obtain a bound for this quantity relying on the results of Pinelis (1994) [18] and Laurent and Massart (2000) [10] (see p.24 of their paper) who control the tail of the weighted sum of independent $\chi^2(1)$ random variables.
Let $\lambda = (\lambda_j)_{j=1,\ldots,q} \in \mathbb{R}_+^q$ be the eigenvalues of S_n^2 (ordered in a increasing order). We define for any $\rho_1, \rho_2 > 0$, the following effective dimensions (see [5] for other expressions of these quantities):

$$\Theta_1(\lambda,\rho_1,\rho_2) = \sum_{j=1}^{\inf(n,q)} \frac{\lambda_j}{\rho_1+\rho_2\lambda_j} \ ; \quad \Theta_2(\lambda,\rho_1,\rho_2) = \sqrt{\sum_{j=1}^{\inf(n,q)} \frac{\lambda_j^2}{(\rho_1+\rho_2\lambda_j)^2}} \ ; \quad \Theta_\infty(\lambda,\rho_1,\rho_2) = \sup_{1\le j\le \inf(n,q)} \left(\frac{\lambda_j}{\rho_1+\rho_2\lambda_j}\right).$$

In the next result, we obtain a sharp bound for regularized and penalized Hotelling's T_n^2. Notice that, in that case, the recentering factor depends on $\Theta_1(\lambda,\rho_1,\rho_2)$ and is random. In the proof of Theorem 2.1, this value is essentially bounded by n/ρ_2, which is a very bad approximation when ρ_2 is small. Theorem 2 tells that, for $q \ge n$, the tail of the regularized Hotelling's T_n^2 statistic behaves as the weighted sum of n independent $\chi^2(1)$ r.v.'s where the weights are given by the random factors $\frac{\lambda_j}{\rho_1+\rho_2\lambda_j}$. We get some Bernstein bounds for this weighted sum by first randomizing by some independent Gaussian r.v.'s, then conditioning on the data and applying Laurent and Massart (2000)'s Bernstein inequality [10]. This inequality in turn can be transformed into some exact bounds for the statistics of interest.

Theorem 2. *Assume that Z has a symmetric distribution then, without any moment assumption, we have, for any $n > 1$ and $q > 0$, for any $t > 0$ and for any $\rho_1, \rho_2 > 0$,*

$$\mathbb{P}\left(\frac{T_n^2(\rho_1,\rho_2) - \Theta_1(\lambda,\rho_1,\rho_2)}{\sqrt{2\Theta_2(\lambda,\rho_1,\rho_2)^2}} \ge \sqrt{2}\left(\sqrt{t} + \frac{\Theta_\infty(\lambda,\rho_1,\rho_2)}{\Theta_2(\lambda,\rho_1,\rho_2)} t\right)\right) \le C\exp(-t), \quad \textit{with } C = 3824.$$

Or equivalently, we have for the penalized Hotelling's statistic, for $n > 1$ and $q > 0$, for any $t > 0$ and, for any $\rho > 0$,

$$\mathbb{P}\left(\frac{T_n^2(\rho,1) - \Theta_1(\lambda,\rho,1)}{\Theta_2(\lambda,\rho,1)} \ge \sqrt{2t} + \frac{\Theta_\infty(\lambda,\rho,1)}{\Theta_2(\lambda,\rho,1)} t\right) \le C\exp\left(-\frac{t}{2}\right).$$

In the symmetric case, this theorem enables us to easily obtain confidence regions of level $1-\delta$, for $\delta \in [0,1]$ for the regularized Hotelling's statistic, as stated in the following corollary.

Corollary 1. *Put $c(\delta) = \log \frac{C}{\delta}$ with $C = 3824$. Then, for any $n > 1$ and $q \ge 1$, for any $t > 0$ and for any $\rho_1, \rho_2 > 0$, with probability $1-\delta$, we have*

$$T_n^2(\rho_1,\rho_2) \le \Theta_1(\lambda,\rho_1,\rho_2) + 2\Theta_2(\lambda,\rho_1,\rho_2)\left(\sqrt{c(\delta)} + \frac{\Theta_\infty(\lambda,\rho_1,\rho_2)}{\Theta_2(\lambda,\rho_1,\rho_2)} c(\delta)\right),$$

The proof of this corollary is left to the reader. This result holds for any n and q. When $q \le n$ is large, we can actually put $\rho_1 = 0$ and get some Pinelis' type bounds (when the χ^2 distribution tail is itself approximated by a Gaussian tail).

The constant C comes from a result of Chasapis and al (2022) [4] who extended a result of Pinelis [18] (1994). Indeed they state that, when symmetrizing, for smooth functions of quadratic forms, Rademacher variables may be replaced by standard normal variables. However, their constant is clearly not optimal and we expect the optimal C to be $2e^3/9$ as in Pinelis [18] (1994).

The bounds in Theorem 2 and Corollary 1 can be used in practice for testing purposes in particular in anomaly detection in statistical learning. See for instance the literature on intrusion detection systems using multivariate control charts based on Hotelling T_n^2 (for instance Tracy et al. (1992) [21] and further works by these authors).

2.4. Bounds for regularized Hotelling's T_n^2 for non symmetric distribution

We now consider Z with a general (not necessarily symmetric) distribution. We will later prove a symmetrization lemma that generalizes the one obtained in Bertail et al. (2008) [2]. In the following, we also use the results of Ledoit and Wolf (2000) [11] to optimally control the distance between $\Sigma_n^2(\rho_1, \rho_2)$ and S^2. For this, consider the modified Frobenius scalar product between matrices and the corresponding norm given by

$$\langle A, B\rangle = \frac{Tr(AB')}{q} \qquad \text{and} \quad \|A\|^2 = \langle A, A\rangle = \frac{Tr(AA')}{q}.$$

Note that dividing the standard Frobenius scalar product by q enables the norm of the identity I_q to be equal to 1, which is more convenient. In the following, we extend this modified Frobenius norm to vectors by considering, for any vector $Z \in \mathbb{R}^q$,

$$\|Z\|^2 = Tr(ZZ')/q.$$

2.4.1. Additional notations and hypotheses

Put $S^2 = \left(\sigma_{kj}\right)_{1\le k,j\le q}$ and consider Λ the diagonal matrix of the eigenvalues of S^2 and O the matrix of associated eigenvectors. The eigenvalues are denoted $\mu_1, \dots, \mu_q$ with $\mu_1 \le \mu_2 \le \cdots \le \mu_q$. We have $S^2 = O'\Lambda^2 O$. Now, for $i \in \{1, \dots, n\}$, we define $Y_i = OZ_i$ with $Y_i = \left(Y_{i,1}, \dots, Y_{i,q}\right)'$.

In order to provide a well-conditioned estimator for large dimensional covariance matrices, Ledoit and Wolf (2000) [11] have studied the minimum of $\mathbb{E}\left(\left\|\Sigma_n^2(\rho_1, \rho_2) - S^2\right\|^2\right)$. This minimization can be seen as a projection problem in the Hilbert space of random matrices, equipped with the inner product $\langle A, B\rangle_{\mathcal{H}} = \mathbb{E}\left[\langle A, B\rangle\right]$ with associated norm $\|.\|_{\mathcal{H}}^2 = \mathbb{E}\|.\|^2$.

We assume the four following assumptions:

(A_1) $\exists K_0, K_1 > 0$ such that, for any n and any $q \ge n$, $K_0 \le \frac{q}{n} \le K_1$.

(A_2) $\exists K_2 > 0$ such that, for any n and any $q \ge n$, $\frac{1}{q}\sum_{j=1}^q \mathbb{E}\left[Y_{1,j}^8\right] \le K_2$.

(A_3) $\exists K_3 > 0$ such that for any n and any $q \ge n$, $\frac{1}{K_3} < \mu_1 \le \mu_q < K_3$.

(A_4) $\exists K_4 > 0$ such that for any n and any $q \ge n$,

$$\nu = \frac{q^2}{n^2} \times \frac{\sum_{(i,j,k,l)\in \mathbf{Q}} \left(\mathrm{Cov}\left(Y_{1,i}Y_{1,j}, Y_{1,k}Y_{1,l}\right)\right)^2}{Card(\mathbf{Q})} \le \frac{K_4}{n},$$

where $\mathbf{Q}$ denotes the set of all the quadruples that are made of four distinct integers between 1 and q.

Remarks: (A_2) and (A_4) are already assumed in Ledoit and Wolf (2000) [11]. First assumption (A_1) essentially means that $q = q(n)$ is of the same order as n. (A_2) states that the moment of order 8 is bounded in average: this condition holds if the following moment of order 8, $\frac{1}{q}\sum_{j=1}^q \mathbb{E}\left[Z_{1,j}^8\right]$ is finite (by sub-multiplicative inequality and the fact that $\|O\| = 1$). This is a weak condition: we do not require exponential moments and allow for fat tail behavior of the sample. (A_3) ensures that the largest and the smallest eigenvalue of the true covariance matrix are bounded. This rules out the case when the components of the vector Z are too correlated: consider for instance the degenerate case where S^2 is a matrix full of 1, then in that case the smallest eigenvalue is 0 and the largest is q. The case of a vector with long memory components is studied in Merlevède et al. (2019) [15] : they show that the largest eigenvalue is unbounded. Thus this case does not enter our framework. Assumption (A_4) is immediate in the Gaussian case, since $\nu = 0$ because of the rotation which makes the $Y_{1,j}$'s $j \in \{1, \dots, q\}$ independent. Obviously, for $(Z_{1,j})_j$ independent, $\nu = 0$ as well. More generally if the components of the vector satisfy some adequate α-mixing conditions, then the sum in the hypothesis (A_4) can be seen as a sum of cumulants and may also be controlled using the arguments of Doukhan and León (1989) [6].

2.4.2. Inequalities for random variables with a general distribution

The next Theorem 3 extends Theorem 1 to general distributions which are not necessarily symmetric. From now on, following Ledoit and Wolf (2000) [11], we denote ρ_1^* and ρ_2^* the optimal values defined as the minimum arguments of $\mathbb{E}\left\|\Sigma_n^2(\rho_1, \rho_2) - S^2\right\|^2$. Ledoit and Wolf (2000) [11] have obtained

$$\rho_1^* = \frac{\beta^2}{\delta^2}\sigma^2 \quad \text{and} \quad \rho_2^* = \frac{\alpha^2}{\delta^2}, \text{ with}$$

$$\sigma^2 = \left\langle S^2, I_q \right\rangle ; \quad \alpha^2 = \left\|S^2 - \sigma^2 I_q\right\|^2 ; \quad \beta^2 = \mathbb{E}\left\|S_n^2 - S^2\right\|^2 \quad \text{and} \quad \delta^2 = \alpha^2 + \beta^2 = \mathbb{E}\left\|S_n^2 - \sigma^2 I_q\right\|^2.$$

Now, we define, for $\alpha^2 \neq 0$, $\rho^* = \frac{\rho_1^*}{\rho_2^*} = \frac{\beta^2}{\alpha^2}\sigma^2$, which yields the optimal penalized estimator of S_n^2:

$$\Sigma_n^{*2} = \frac{\Sigma_n^2\left(\rho_1^*, \rho_2^*\right)}{\rho_2^*} = S_n^2 + \rho^* I_q.$$

If $\alpha^2 = 0$, take $\Sigma_n^{*2} = \sigma^2 I_q$ (in that case we will just need to estimate σ^2).

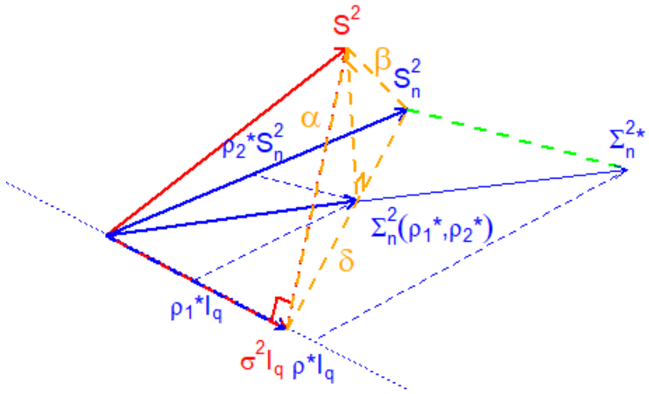

Fig. 1: True covariance S^2, sample covariance S_n^2, and $\Sigma_n^2(\rho_1^*, \rho_2^*)$, Σ_n^{2*} respectively regularized and penalized sample covariance

.

In Figure 1, the scalar product is $\langle , \rangle_{\mathcal{H}}$ with its associated norm. We represent $\Sigma_n^2(\rho_1^*, \rho_2^*)$, the optimal combination of S_n^2 and I_q defined by orthogonal projection of the true covariance matrix S^2 on the random vector-space generated by S_n^2 and I_q. Thus $\Sigma_n^{2*} = \Sigma_n^2(\rho^*, 1)$ is the penalization of S_n^2 by I_q with $\rho^* = \frac{\rho_1^*}{\rho_2^*}$. The green dashed line represents the set of penalized estimators $\Sigma_n^2(\rho, 1)$ for which we obtain universal bounds in Theorem 3.

Theorem 3. *Assume that Z has a general distribution with finite variance S^2. Assume in addition that assumptions (A_1) to (A_3) hold. Put $a^* = 1 + \frac{K_3}{\rho^*}$. Then we have, for any $n > 1$, for any $q \geq n$, and for $t > 2n$,*

$$\mathbb{P}\left(T_n^2(\rho^*, 1) \geq (1 + a^*)\, t\right) = \mathbb{P}\left[n\bar{Z}_n' \Sigma_n^{*-2} \bar{Z}_n \geq (1 + a^*)\, t\right] \quad \leq \quad \frac{2e^3}{9}\left(\frac{t-n}{2}\right)^{\frac{n}{2}} \frac{\exp\left(-\frac{t-n}{2}\right)}{\Gamma\left(\frac{n}{2}+1\right)},$$

Remark: Here the bounding function for large t behaves like a centered $\chi^2(n)$ distribution, up to the factor $\frac{2e^3}{9}$. The term $(1 + a^*)$ ensures that the smallest eigenvalue of Σ_n^{*2} does not contribute to the inequality.
Notice that the inequality is still valid when using Σ_n^2, the regularized version of S_n^2 instead of the penalized version Σ_n^{*2}, up to a small modification of the bound $(1 + a^*)t$ by the factor $1/\rho_2^*$: for $n > 1$, $q \geq n$, for any $t > 2n$

$$\mathbb{P}\left(T_n^2(\rho_1^*, \rho_2^*) \geq \frac{1}{\rho_2^\star}(1 + a^*)\, t\right) \leq \frac{2e^3}{9}\left(\frac{t-n}{2}\right)^{\frac{n}{2}} \frac{\exp\left(-\frac{t-n}{2}\right)}{\Gamma\left(\frac{n}{2}+1\right)}.$$

3. Inequality with estimated parameters

We have proved an exponential inequality for the penalized Hotelling's T_n^2 with theoretical values a^* and ρ^*. In practice these values are unknown. In this section, we estimate these quantities and obtain an inequality for the penalized Hotelling's T_n^2 with estimated parameters.

We first recall several results of Ledoit and Wolf (2000) [11] on the asymptotic behavior of regularized empirical covariance estimator Σ_n^2. Lemma 1 and proposition 1 below summarize these results with our notations and are proved by Ledoit and Wolf (2000) [11] in different lemmas and a theorem of their paper.

We use the same assumptions as in Ledoit and Wolf (2000) [11]: $\xrightarrow{L_4}$ denotes the fourth-moment convergence as n goes to infinity, i.e.

$$U_n \xrightarrow{L_4} U \Longleftrightarrow \mathbb{E}\left[(U_n - U)^4\right] \underset{n\to\infty}{\longrightarrow} 0.$$

Ledoit and Wolf (2000) [11] essentially have shown that L_4-consistent estimators for σ^2, α^2, β^2 and δ^2 are simply given by their empirical counterparts that is

$$\hat{\sigma}_n^2 = \left\langle S_n^2, I_q \right\rangle \;\; ; \quad \hat{\delta}_n^2 = \left\| S_n^2 - \hat{\sigma}_n^2 I_q \right\|^2 \;\; ; \quad \hat{\alpha}_n^2 = \hat{\delta}_n^2 - \hat{\beta}_n^2 \quad \text{with } \bar{\beta}_n^2 = \frac{1}{n^2}\sum_{i=1}^{n} \left\| Z_i (Z_i)' - S_n^2 \right\|^2 \text{ and } \hat{\beta}_n^2 = \min\left(\bar{\beta}_n^2, \hat{\delta}_n^2\right)$$

Lemma 1. *[Ledoit and Wolf (2000) [11] lemma 3.2, lemma 3.3, lemma 3.4, lemma 3.5] Under assumptions (A_1) to (A_4), we have*

1. *σ^2, α^2 and β^2 remain bounded (as n and q tend to ∞).*
2. *For all n, $\mathbb{E}\left[\hat{\sigma}_n^2\right] = \sigma^2$, and $\hat{\sigma}_n^2 - \sigma^2 \xrightarrow{L_4} 0$ and $\hat{\sigma}_n^4 - \sigma^4 \xrightarrow{L_4} 0$.*
3. $\hat{\delta}_n^2 - \delta^2 \xrightarrow{L_4} 0$. 4. $\bar{\beta}_n^2 - \beta^2 \xrightarrow{L_4} 0$ *and* $\hat{\beta}_n^2 - \beta^2 \xrightarrow{L_4} 0$. 5. $\hat{\alpha}_n^2 - \alpha^2 \xrightarrow{L_4} 0$.

After replacing the unobservable scalars σ^2, α^2, β^2 and δ^2 by their sample counterparts in the formula of Σ_n^2, Ledoit and Wolf obtained an estimation of the regularized empirical covariance matrix say

$$\hat{\Sigma}_n^2 = \frac{\hat{\beta}_n^2}{\hat{\delta}_n^2}\hat{\sigma}_n^2 I_q + \frac{\hat{\alpha}_n^2}{\hat{\delta}_n^2} S_n^2.$$

Ledoit and Wolf (2000) [11] have shown that $\hat{\Sigma}_n^2$ and Σ_n^2 are asymptotically equivalent in the modified Frobenius norm.

Proposition 1. *[Ledoit and Wolf (2000) [11], Theorem 3.2] Under the assumptions* (A_1)-(A_4)*, we have*

1. $\underset{n\to\infty}{lim} \mathbb{E}\left\|\hat{\Sigma}_n^2 - \Sigma_n^2\right\|^2 = 0.$
2. *Moreover, $\hat{\Sigma}_n^2$ has the same asymptotic expected loss (or risk) as Σ_n^2 i.e.*

$$\underset{n\to\infty}{lim} \mathbb{E}\left\|\hat{\Sigma}_n^2 - \Sigma^2\right\|^2 - \mathbb{E}\left\|\Sigma_n^2 - \Sigma^2\right\|^2 = 0.$$

In the same way as Ledoit and Wolf (2000) [11] we define the optimal coefficients ρ_1^* and ρ_2^*. They are estimated respectively by $\hat{\rho}_1^*$ and $\hat{\rho}_2^*$, where $\hat{\rho}_1^* = \frac{\hat{\beta}_n^2}{\hat{\delta}_n^2}\hat{\sigma}_n^2$ and $\hat{\rho}_2^* = \frac{\hat{\alpha}_n^2}{\hat{\delta}_n^2}$. Now, if $\hat{\alpha}_n^2 \neq 0$, we introduce $\hat{\Sigma}_n^{2*}$ the "estimated optimal" penalized version of S_n^2 given by

$$\hat{\Sigma}_n^{*2} = \Sigma_n^2\left(\frac{\hat{\rho}_1^*}{\hat{\rho}_2^*}, 1\right) = S_n^2 + \hat{\rho}_n^* I_q, \quad \text{where } \hat{\rho}_n^* = \frac{\hat{\beta}_n^2 \hat{\sigma}_n^2}{\hat{\alpha}_n^2}.$$

Similarly the unobservable threshold constant a^* introduced in theorem 3 is estimated by $\hat{a}_n^* = 1 + \frac{K_3}{\hat{\rho}_n^*}$. The principle in Figure 2 is similar to the one in Figure 1 except that $\hat{\Sigma}_n^2$ is determined first so that the regularized estimator belongs to the yellow line and the optimal estimator $\Sigma_n^2 = \Sigma_n^2\left(\hat{\rho}_1^*, \hat{\rho}_2^*\right)$ is the closest value to S^2 on this line. This difference induces an additional error term in our inequalities.

Theorem 4 establishes an exponential bound for the penalized self-normalized sums, when Σ_n^{*2} is replaced by the estimator $\hat{\Sigma}_n^{2*}$ and a^* by $\hat{a}_n^*$, up to a small error term that we control explicitly.

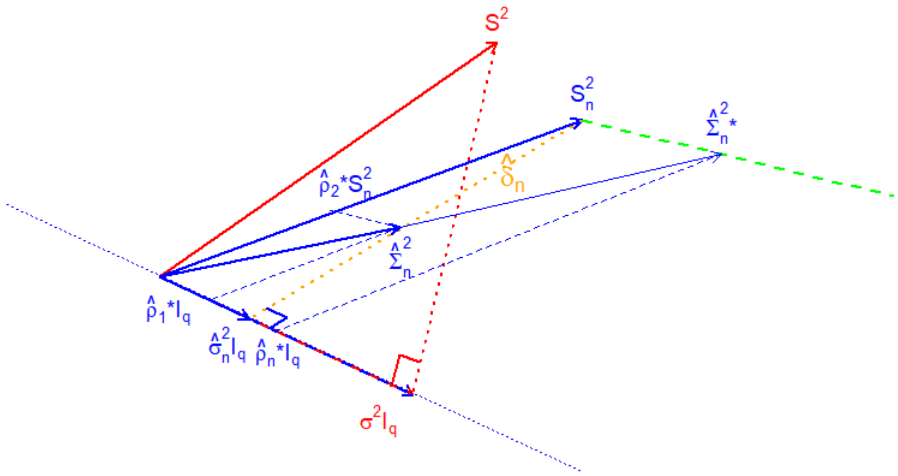

Fig. 2: True covariance S^2, sample covariance S_n^2, regularized and penalized estimators of S_n^2, respectively $\hat{\Sigma}_n^2$ and $\hat{\Sigma}_n^{*2}$.

Theorem 4. *Under the assumptions* (A_1) *to* (A_4)*, we have, for any* $n > 1$*, for any* $q > n$*, for any* $t > 2n$ *and for any small value of* $\epsilon > 0$*,*

$$\mathbb{P}\left(T_n^2(\widehat{\rho}_n^*, 1) \geq t\left(1 + \hat{a}_n^* + 2\epsilon\right)\right) = \mathbb{P}\left(n\bar{Z}_n'\hat{\Sigma}_n^{*-2}\bar{Z}_n \geq t\left(1 + \hat{a}_n^* + 2\epsilon\right)\right) \leq \frac{2e^3}{9}\left(\frac{t-n}{2}\right)^{\frac{n}{2}} \frac{e^{-\frac{t-n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)} + \frac{C(\epsilon)}{n\epsilon}, \quad (3)$$

where $\hat{a}_n^* = 1 + \frac{K_3}{\hat{\rho}_n^*}$*, and* $C(.)$ *is a real nonnegative function, independent of n, defined by*

$$C(\epsilon) = 4K_1\sqrt{K_2}\left(2 + \frac{1}{q} + K_1\right) + 2K_1 G\left(\sqrt{\frac{\epsilon}{2K_1}}\right) + \frac{4K_1^2\sigma^4}{\epsilon} G\left(\frac{\epsilon}{2\sigma^2 K_1}\right) + \frac{K_3^2}{\epsilon} G\left(\frac{\epsilon}{K_3}\right).$$

The function G is defined explicitly in lemma 7. Notice that $C(\epsilon)/\epsilon$ *explodes when* ϵ *goes to 0.*

These results essentially show that we have a $\chi^2(n)$ control in the tail of the distribution, for a threshold larger than $2n(1 + \hat{a}_n^*)$ (recall that $2n$ is the variance of a $\chi^2(n)$ distribution). The loss $(1 + \hat{a}_n^*)$ is essentially due to the correlation between the components of Z and the deviation from homoscedasticity. The value of ϵ can not be too small but can be optimized by balancing the two terms in the inequality. For a given ϵ and a given level δ it is possible to solve numerically the second term of the inequality (3) equal to delta to get a valid bound for the Hotelling's T^2 for any n and q.

Acknowledgment

This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023- 01).

Credit authorship contribution statement

The three authors contributed equally. This paper is a part of the PhD thesis of El Mehdi ISSOUANI.

4. Appendix

In the first part of this section we provide all the proofs of Theorem 1, 2, 3 and 4 given in the sections 2, 3 and 4. In the second part of the appendix, we detail all the calculations to obtain an explicit constant $C(\epsilon)$ appearing in Theorem 4 when replacing the true quantities by their empirical estimators.

Proofs of the theorems

We set some notations that we will consider in the following proofs. S_n^2 is a symmetric and diagonalizable matrix. Let's denote by O_n an orthogonal matrix in $\mathcal{M}_q(\mathbb{R})$ such that $S_n^2 = O_n'\Lambda_n^2 O_n$ where Λ_n^2 is a diagonal matrix and

$$\Lambda_n^2 = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_n & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \quad \text{for any } q > n.$$

Put $\hat{Y}_i = O_n Z_i$ with $\hat{Y}_i = \left(\hat{Y}_{i,1}\ldots,\hat{Y}_{i,q}\right)'$. Let $\lambda_1 \le \cdots \le \lambda_q$ denote eigenvalues of S_n^2.

4.1. Proof of theorem 1 and 2

We first establish a simple inequality for the penalized Hotelling's T_n^2 in the symmetric case, based on previous results by Pinelis [18]. The idea of the theorem is to use a rotation trick of the Z_i that allows us to return to the "small" dimension case given by Pinelis. This yields a bound given by the survival function of a χ^2 with n degrees of freedom.

Proof of theorem 1. Note that Vectors $\hat{Y}_i$ remain symmetric in distribution and uncorrelated. It is easy to see that, by construction, the empirical covariance matrix of the $\hat{Y}_1,\ldots,\hat{Y}_n$ is

$$\frac{1}{n}\sum_{i=1}^n \hat{Y}_i\hat{Y}_i' = \frac{1}{n}\sum_{i=1}^n O_n Z_i Z_i' O_n' = O_n S_n^2 O_n' = \Lambda_n^2.$$

This implies that, for any vector $\hat{Y}_i$, their coordinates for $j \ge n+1$ are zero. Indeed, for $j \ge n+1$, $n^{-1}\sum_{i=1}^n \hat{Y}_{i,j}^2 = 0$, implies in turn that each $\hat{Y}_{i,j} = 0$, for $j \in \{n+1,\ldots,q\}$ and $i \in \{1,\ldots,n\}$. Define $\tilde{Y}_i$ the n-dimensional vector version of $\hat{Y}_i$ with these non-zero components, that is to say $\forall j \le n,\ \tilde{Y}_{i,j} = \hat{Y}_{i,j}$ and their corresponding empirical mean $\bar{\tilde{Y}}_n$ on the collection $\tilde{Y}^{(n)} = \left(\tilde{Y}_i\right)_{1\le i\le n}$. Thus, for all $\rho_2 > 0$, we have

$$\begin{aligned} n\bar{Z}_n'\Sigma_n^{-2}(\rho_1,\rho_2)\bar{Z}_n &= n\left(\frac{1}{n}\sum_{i=1}^n \hat{Y}_i'\right)\left(\rho_1 I_q + \rho_2\Lambda_n^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \hat{Y}_i\right) \\ &= n\sum_{j=1}^n \frac{\left(n^{-1}\sum_{i=1}^n \hat{Y}_{i,j}\right)^2}{\rho_1+\rho_2\lambda_j} \le n\sum_{j=1}^n \frac{\left(n^{-1}\sum_{i=1}^n \hat{Y}_{i,j}\right)^2}{\rho_2\lambda_j} \le \frac{1}{\rho_2}\sum_{j=1}^n \frac{\left(n^{-1/2}\sum_{i=1}^n \hat{Y}_{i,j}\right)^2}{\lambda_j}. \end{aligned}$$

As $\lambda_j = n^{-1}\sum_{i=1}^n \hat{Y}_{i,j}^2$, we have reduced the problem to the sum of n self normalized sums, which can be seen as Hotelling's T_n^2 of symmetric random variables in $\mathbb{R}^n$. In other words, $n\bar{Z}_n'\Sigma_n^{-2}(\rho_1,\rho_2)\bar{Z}_n \le \frac{1}{\rho_2} n\bar{\tilde{Y}}_n' S_n^{-2}\left(\tilde{Y}^{(n)}\right)\bar{\tilde{Y}}_n$. Thus, by applying Pinelis' equation (1) [18], we have

$$\forall t > 0,\quad \mathbb{P}\left(n\bar{Z}_n'\Sigma_n^{-2}\bar{Z}_n \ge \frac{t}{\rho_2}\right) \le \frac{2e^3}{9}\bar{F}_n(t)\ .$$

Recall that, if $N_1,\ldots,N_n$ are independent $N(0,1)$ random variables, then by Lemma 1 of Laurent and Massart (2000) [10], one has, for $u > 0$,

$$\mathbb{P}\left(\frac{\sum_{i=1}^n N_i^2 - n}{\sqrt{2n}} \ge \sqrt{2}(\sqrt{u} + \frac{u}{\sqrt{n}})\right) \le e^{-u}\ .$$

By inverting the polynomial in $\sqrt{u}$, this is a Bernstein type inequality for i.i.d random variables

$$\mathbb{P}\left(\frac{\sum_{i=1}^n N_i^2 - n}{\sqrt{2n}} \ge v\right) \le \exp\left(-\frac{2v^2}{\left(1+\sqrt{1+2\sqrt{2}\frac{v}{\sqrt{n}}}\right)^2}\right) \le \exp\left(-\frac{v^2}{2(1+\sqrt{2}\frac{v}{\sqrt{n}})}\right).$$

It follows that, for $t > n$,

$$\bar{F}_n(t) = \mathbb{P}\left(\frac{\sum_{i=1}^n N_i^2 - n}{\sqrt{2n}} \geq \frac{t-n}{\sqrt{2n}}\right) \quad \leq \quad \exp\left(-\frac{(t-n)^2}{4t}\right).$$

□

Proof of theorem 2. Recall that : $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$ with $Z_i \in \mathbb{R}^q$. Introduce independent Rademacher r.v.'s ε_i taking the values ± 1 with probability $1/2$. Define $\bar{Z}_n^\epsilon = \frac{1}{n}\sum_{i=1}^n \epsilon_i Z_i$. Then, in the symmetric case considered here, $\bar{Z}_n$ and $\bar{Z}_n^\epsilon$ have the same distribution. Now write

$$n\bar{Z}_n^{\epsilon\prime}\Sigma_n^{-2}(\rho_1,\rho_2)\bar{Z}_n^\epsilon = n\left(\frac{1}{n}\sum_{i=1}^n \epsilon_i \hat{Y}_i'\right)\left(\rho_1 I_q + \rho_2\Lambda_n^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \epsilon_i \hat{Y}_i\right) = \epsilon' V V' \epsilon \quad (4)$$

where $\hat{Y} = (\hat{Y}_1,\ldots,\hat{Y}_n)'$, $\epsilon = (\epsilon_1,\ldots,\epsilon_n)'$ and $V = \frac{1}{\sqrt{n}}\hat{Y}\left(\rho_1 I_q + \rho_2\Lambda_n^2\right)^{-1/2}$.

Chasapis and al (2022) [4] obtain an extension of Pinelis' result [18] stating that for smooth functions of quadratic forms, Rademacher variables may be replaced by standard normal variables. More precisely, define the Euclidian norm $\|x\|_2 = \sqrt{\langle x, x\rangle}$ and consider $\xi_1,\ldots,\xi_n$ independent standard Gaussian random variables. Then, for any $t \geq 0$, for any vectors $v_1,\ldots,v_n$ in $\mathbb{R}^q$, we have

$$\mathbb{P}\left[\|\epsilon_1 v_1 + \cdots + \epsilon_n v_n\|_2 \geq t\right] \leq C\mathbb{P}\left[\|\xi_1 v_1 + \cdots + \xi_n v_n\|_2 \geq t\right] \quad \text{with } C = 3824.$$

Since we have

$$\epsilon_1 v_1 + \cdots + \epsilon_n v_n = \epsilon' V$$

where V is the matrix of vectors $v_i = \left(v_{i1},\ldots,v_{iq}\right)$ corresponding to the rows, we can rewrite

$$\|\epsilon_1 v_1 + \cdots + \epsilon_n v_n\|_2^2 = \left\|\epsilon' V\right\|_2^2 = \epsilon' V V' \epsilon.$$

It follows that, for any $u > 0$,

$$\mathbb{P}\left[\epsilon' V V' \epsilon \geq u\right] \leq C\mathbb{P}\left[\xi' V V' \xi \geq u\right]$$

By conditioning according to $\hat{Y}_i$'s and using equation (4), we have, for any $u > 0$ and, for any $\rho_1, \rho_2 > 0$,

$$\begin{aligned}\mathbb{P}\left[n\bar{Z}_n^{\epsilon\prime}\Sigma_n^{-2}(\rho_1,\rho_2)\bar{Z}_n^\epsilon \geq u\right] &= \mathbb{E}\left[\mathbb{P}\left(\epsilon' V V' \epsilon \geq u \;\middle|\; \hat{Y}_1,\ldots,\hat{Y}_n\right)\right] \\ &\leq C\mathbb{E}\left[\mathbb{P}\left(\xi' V V' \xi \geq u \;\middle|\; \hat{Y}_1,\ldots,\hat{Y}_n\right)\right].\end{aligned}$$

Moreover recall from the preceding proof that we have

$$\begin{aligned}n\bar{Z}_n^{\epsilon\prime}\Sigma_n^{-2}(\rho_1,\rho_2)\bar{Z}_n^\epsilon &= n\left(\frac{1}{n}\sum_{i=1}^n \epsilon_i \hat{Y}_i'\right)\left(\rho_1 I_q + \rho_2\Lambda_n^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \epsilon_i \hat{Y}_i\right) = n\sum_{j=1}^{\inf(q,n)} \frac{\left(n^{-1}\sum_{i=1}^n \epsilon_i \hat{Y}_{i,j}\right)^2}{\rho_1 + \rho_2\lambda_j} \\ &= n\sum_{j=1}^{\inf(q,n)} \frac{\left(n^{-1}\sum_{i=1}^n \epsilon_i \hat{Y}_{i,j}\right)^2}{\lambda_j}\frac{\lambda_j}{\rho_1 + \rho_2\lambda_j}\end{aligned}$$

We obtain

$$\mathbb{P}\left[n\bar{Z}_n^{\epsilon\prime}\left(\Sigma_n^2(\rho_1,\rho_2)\right)^{-1}\bar{Z}_n^\epsilon > u\right] \leq C\mathbb{E}\left[\mathbb{P}\left(n\sum_{j=1}^{\inf(q,n)} \frac{\left(n^{-1}\sum_{i=1}^n \xi_i \hat{Y}_{i,j}\right)^2}{\lambda_j}\frac{\lambda_j}{\rho_1 + \rho_2\lambda_j} > u \;\middle|\; \hat{Y}_1,\ldots,\hat{Y}_n\right)\right]. \quad (5)$$

Let us work now conditionally to $\hat{Y}_1, \ldots, \hat{Y}_n$. Put $K_j = \sqrt{n}\left(n^{-1}\sum_{i=1}^n \xi_i \hat{Y}_{i,j}\right)/\sqrt{\lambda_j}$ for $j \in \{1, \ldots, \inf(q,n)\}$. Thus

$$\text{for any } j \neq k \quad \mathrm{Cov}\left(K_j, K_k \,\middle|\, \hat{Y}_1, \ldots, \hat{Y}_n\right) = \mathrm{Cov}\left(\sqrt{n}\frac{n^{-1}\sum_{i=1}^n \xi_i \hat{Y}_{i,j}}{\sqrt{\lambda_j}}, \sqrt{n}\frac{n^{-1}\sum_{i=1}^n \xi_i \hat{Y}_{i,k}}{\sqrt{\lambda_k}} \,\middle|\, \hat{Y}_1, \ldots, \hat{Y}_n\right)$$
$$= \frac{1}{n}\sum_{i=1}^n \frac{\hat{Y}_{i,j}\hat{Y}_{i,k}}{\sqrt{\lambda_j \lambda_k}} = 0.$$

Since $K = \left(K_1, \ldots, K_{\inf(q,n)}\right)$ is a Gaussian vector (as a linear combination of independent variables) it follows that $K_1^2, \ldots, K_{\inf(q,n)}^2$ are iid $\chi^2(1)$.

Now, consider the vector $b = \left(b_1, \ldots, b_q\right)$ with nonnegative components (conditionally to $\hat{Y}_{i,j}$'s) defined by

$$b_j = \frac{\lambda_j}{\rho_1 + \rho_2 \lambda_j}.$$

A direct application of Laurent and Massart, lemma [10] to $\sum_{j=1}^{\inf(q,n)} b_j \left(K_j^2 - 1\right)$ gives for any $u > 0$

$$\mathbb{P}\left(\sum_{j=1}^{\inf(q,n)} b_j\left(K_j^2 - 1\right) > 2\|b\|_2\sqrt{u} + 2\|b\|_\infty u\right) \leq \exp(-u).$$

In other words, for any $u > 0$, we have

$$\mathbb{P}\left(\frac{\sum_{j=1}^{\inf(q,n)} b_j K_j^2 - \|b\|_1}{\sqrt{2\|b\|_2^2}} > \sqrt{2}\sqrt{u} + \sqrt{2}\frac{\|b\|_\infty}{\|b\|_2}u\right) \leq \exp(-u) \tag{6}$$

Now by combining (5) and (6) we obtain the following result for the recentered version of our quantity of interest,

$$\mathbb{P}\left(\frac{n\bar{Z}_n^{\epsilon\prime}\Sigma_n^{-2}(\rho_1,\rho_2)\bar{Z}_n^{\epsilon} - \|b\|_1}{\sqrt{2\|b\|_2^2}} > \sqrt{2}\sqrt{u} + \sqrt{2}\frac{\|b\|_\infty}{\|b\|_2}u\right)$$
$$\leq C\mathbb{E}\left[\mathbb{P}\left(\frac{\sum_{j=1}^{\inf(q,n)} b_j K_j^2 - \|b\|_1}{\sqrt{2\|b\|_2^2}} > \sqrt{2}\sqrt{u} + \sqrt{2}\frac{\|b\|_\infty}{\|b\|_2}u \,\middle|\, \left(\hat{Y}_{1,j}, \ldots, \hat{Y}_{n,j}\right)_{j\in\{1,\ldots,\inf(q,n)\}}\right)\right]$$
$$\leq C\exp(-u).$$

The result of the theorem follows by noticing that $\|b\|_k = \Theta_k(\lambda, \rho_1, \rho_2), \quad k \in \{1, 2, \infty\}$ □

4.2. Proof of theorem 3

4.2.1. A symmetrization lemma adapted to χ^2 distribution

The following lemma ensures that, if we have a $\chi^2(k)$ type of control for the tail of a random variable ν, which stochastically dominates some random variable ξ, then we are also able to control the tail of ξ. For large values, this tail is essentially the same as the one of a $\chi^2(k)$ distribution. We use exactly the same ideas as in Panchenko's lemma 1 and corollary 1 (which assumes an exponential control of the tail of the distribution of the variable ν).

Lemma 2. *Let ν and ξ be two real r.v.'s. For $a \in \mathbb{R}$, put $\Phi_a(x) = \max(x - a; 0)$. Assume that:*

(i) for any $a \in \mathbb{R}$,

$$\mathbb{E}\Phi_a(\xi) \leq \mathbb{E}\Phi_a(\nu)$$

(ii) there exists k and constants $C_1 > 0, c_1 > 0$, such that for any $t > 0$

$$\mathbb{P}(\nu \geq t) \leq C_1 \bar{F}_k(c_1 t)$$

then, for $t > 2k/c_1$, *we have*

$$\mathbb{P}(\xi \geq t) \leq \mathsf{C}_1 \left(\frac{c_1 t - k}{2}\right)^{\frac{k}{2}} \frac{e^{-\frac{c_1 t - k}{2}}}{\Gamma\left(\frac{k}{2}+1\right)}$$

and, for $t > k/c_1$, *we also get*

$$\mathbb{P}(\xi \geq t) \leq \mathsf{C}_1 \bar{F}_{k+2}(c_1 t - k).$$

Proof of lemma 2 . We follow the lines of the proof of Panchenko's lemma, with a function Φ_a with $a = t - \frac{k}{c_1}$ given by $\Phi(x) = \max(x - t + k/c_1; 0)$, for $t > k/c_1$. Remark that $\Phi(x)$ is convex, nondecreasing and that $\Phi(0) = 0$ and $\Phi(t) = k/c_1$. We thus have by Markov's inequality

$$\begin{aligned}\mathbb{P}(\xi \geq t) &\leq \frac{E\Phi(\xi)}{\Phi(t)} \leq \frac{E\Phi(\nu)}{\Phi(t)} \leq \frac{1}{\Phi(t)}\left(\Phi(0) + \int_{t-k/c_1}^{+\infty} \Phi'(x)\,\mathbb{P}(\nu \geq x)\,dx\right) \\ &\leq \mathsf{C}_1 \frac{c_1}{k} \int_{t-k/c_1}^{+\infty} \overline{F}_k(c_1 x)\,dx.\end{aligned}$$

By integration by parts, we get

$$\int_{t-k/c_1}^{+\infty} \bar{F}_k(c_1 x)\,dx = \int_{t-k/c_1}^{+\infty} c_1 x f_k(c_1 x)\,dx - (t - k/c_1)\int_{t-k/c_1}^{+\infty} c_1 f_k(c_1 x)\,dx.$$

Recall that

$$f_k(u) = \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} u^{\frac{k}{2}-1} \exp(-\frac{u}{2}),$$

we thus have

$$\begin{aligned}\frac{c_1}{k}\int_{t-k/c_1}^{+\infty} c_1 x f_k(c_1 x)\,dx &= \frac{c_1}{2^{k/2+1}\frac{k}{2}\Gamma(\frac{k}{2})}\int_{t-k/c_1}^{+\infty} (c_1 x)^{\frac{k+2}{2}-1} \exp(-\frac{c_1 x}{2})dx \\ &= \overline{F}_{k+2}(c_1 t - k).\end{aligned}$$

It follows by straightforward calculations that, for $t > k/c_1$,

$$\mathbb{P}(\xi \geq t) \leq \mathsf{C}_1\left(\overline{F}_{k+2}(c_1 t - k) - \frac{c_1 t - k}{k}\overline{F}_k(c_1 t - k)\right).$$

Using the recurrence relation 26.4.8 of Abramovitch and Stegun ([1], page 941), for $u \geq 2k$,

$$\begin{aligned}\mathsf{C}_1\left(\overline{F}_{k+2}(u-k) - \tfrac{u-k}{k}\overline{F}_k(u-k)\right) &\leq \mathsf{C}_1\left(\overline{F}_{k+2}(u-k) - \overline{F}_k(u-k)\right) \\ &\leq \left(\tfrac{(u-k)}{2}\right)^{k/2} \tfrac{\mathsf{C}_1 e^{-\frac{(u-k)}{2}}}{\Gamma(\frac{k}{2}+1)}.\end{aligned}$$

We get with $u = c_1 t$, for $t \geq 2k/c_1$,

$$\mathbb{P}(\xi \geq t) \leq \left(\frac{(c_1 t - k)}{2}\right)^{k/2} \frac{\mathsf{C}_1 e^{-\frac{(c_1 t - k)}{2}}}{\Gamma\left(\frac{k}{2}+1\right)}.$$

Moreover, for $t > k/c_1$ we have $\mathbb{P}(\xi \geq t) \leq \mathsf{C}_1\left(\overline{F}_{k+2}(c_1 t - k)\right)$.

Notice that we only lose 2 degrees of freedom in this case. It will not be important if k is large, typically of the order of n in our case. □

4.2.2. Extension of Panchenko symmetrization lemma (see [17] Corollary 1, p. 2069)

Let $\mathcal{J}_q = \{u \in \mathbb{R}^q, \|u\|_2 = 1\}$ be the unit circle of $\mathbb{R}^q$. Let $X^{(n)} = (X_i)_{1\le i\le n}$ be an independent copy of $Z^{(n)} = (Z_i)_{1\le i\le n}$. Since $q > n$, the matrix $S_n^2\left(Z^{(n)} - X^{(n)}\right) = \frac{1}{n}\sum_{i=1}^n (Z_i - X_i)(Z_i - X_i)'$ is not invertible. We derive from $S_n^2\left(Z^{(n)} - X^{(n)}\right)$ the corresponding penalized empirical covariance matrix

$$\tilde{\Sigma}_n^2 = 2\rho_1 I_q + \rho_2 S_n^2\left(Z^{(n)} - X^{(n)}\right)$$

It is easy to see that

$$\mathbb{E}\left(S_n^2\left(Z^{(n)} - X^{(n)}\right)\right) = 2S^2 \text{ and } \mathbb{E}\left(S_n^2\left(Z^{(n)} - X^{(n)}\right) \mid Z^{(n)}\right) = S_n^2 + S^2.$$

Since $\tilde{\Sigma}_n^2 = \tilde{\rho}_1 I_q + \tilde{\rho}_2 S_n^2\left(Z^{(n)} - X^{(n)}\right)$, we get that

$$\mathbb{E}\left(\tilde{\Sigma}_n^2 \mid Z^{(n)}\right) = \tilde{\rho}_1 I_q + \tilde{\rho}_2\left(S_n^2 + S^2\right) = 2\rho_1 I_q + \rho_2\left(S_n^2 + S^2\right).$$

As a consequence, define

$$\tilde{\beta}^2 = \mathbb{E}\left(\left\|S_n^2\left(Z^{(n)} - X^{(n)}\right) - 2S^2\right\|^2\right) = \mathbb{E}\left(\left\|S_n^2\left(Z^{(n)}\right) - S^2\right\|^2\right) + \mathbb{E}\left(\left\|S_n^2\left(X^{(n)}\right) - S^2\right\|^2\right) = 2\beta^2.$$

Similarly, put

$$\tilde{\alpha}^2 = 2\alpha^2;\ \tilde{\delta} = 2\delta^2 \text{ and } \tilde{\sigma}^2 = \left\langle 2S^2, I_n\right\rangle = 2\sigma^2$$

then we have

$$\tilde{\rho}_1 = \frac{\tilde{\alpha}^2}{\tilde{\delta}^2}\tilde{\sigma}^2 = 2\frac{\alpha^2}{\delta^2}\sigma^2 = 2\rho_1 \quad \text{and } \tilde{\rho}_2 = \frac{\tilde{\beta}^2}{\tilde{\delta}^2} = \frac{\beta^2}{\delta^2} = \rho_2.$$

It thus follows with this natural choice of $\widetilde{\rho}_1$ and $\widetilde{\rho}_2$ that we have

$$\mathbb{E}\left(\tilde{\Sigma}_n^2 \mid Z^{(n)}\right) = \Sigma_n^2 + \Sigma^2 \quad \text{and } \mathbb{E}\left(\tilde{\Sigma}_n^2\right) = 2(\rho_1 I_q + \rho_2 S^2) = 2\Sigma^2$$

The following lemma and its proof is an extension of corollary 1 of Panchenko (2003) (see [17]) with some adaptations to the multidimensional χ^2 case. See also Bertail et al. (2008) [2] for the non penalized version of this result for $q < n$.

Lemma 3. *If there exists $k \in \mathbb{N}^*$, $C_2 > 0$ and $c_2 > 0$ such that, for all $t \ge 0$,*

$$\mathbb{P}\left(\sup_{u\in\mathcal{J}_q}\left(\frac{\sqrt{n}u'\left(\bar{Z}_n - \bar{X}_n\right)}{\sqrt{u'\tilde{\Sigma}_n^2 u}}\right) \ge \sqrt{t}\right) \le C_2\bar{F}_k(c_2 t),$$

then, for all $t \ge 2k/c_2$,

$$\mathbb{P}\left(\sup_{u\in\mathcal{J}_q}\left(\frac{\sqrt{n}u'\bar{Z}_n}{\sqrt{u'\left(\Sigma_n^2 + \Sigma^2\right)u}}\right) \ge \sqrt{t}\right) \le C_2\left(\frac{(c_2 t - k)}{2}\right)^{k/2}\frac{e^{-\frac{(c_2 t - k)}{2}}}{\Gamma\left(\frac{k}{2} + 1\right)}$$

and, for all $t \ge k/c_2$,

$$\mathbb{P}\left(\sup_{u\in\mathcal{J}_q}\left(\frac{\sqrt{n}u'\bar{Z}_n}{\sqrt{u'\left(\Sigma_n^2 + \Sigma^2\right)u}}\right) \ge \sqrt{t}\right) \le C_2\overline{F}_{k+2}\left(c_2 t - k\right)$$

Proof of Lemma 3 . Denote

$$A_n\left(Z^{(n)}\right) = n\sup_{u\in\mathcal{J}_q}\sup_{b>0}\left\{\mathbb{E}\left[4b\left(u'\left(\bar{Z}_n - \bar{X}_n\right) - bu'\tilde{\Sigma}_n^2 u\right) \mid Z^{(n)}\right]\right\}$$

and

$$C_n\left(Z^{(n)}, X^{(n)}\right) = n\sup_{u\in\mathcal{J}_q}\sup_{b>0}\left\{4b\left(u'\left(\bar{Z}_n - \bar{X}_n\right) - bu'\tilde{\Sigma}_n^2 u\right)\right\}$$

We have by Jensen's inequality, that for any convex function ϕ

$$\phi\left(A_n\left(Z^{(n)}\right)\right) \leq \mathbb{E}\left[\phi\left(C_n\left(Z^{(n)}, X^{(n)}\right)\right) \mid Z^{(n)}\right] \tag{7}$$

Finally, we can rewrite $A_n\left(Z^{(n)}\right)$ and $C_n\left(Z^{(n)}, X^{(n)}\right)$ in an explicit form of self-normalized sums by maximizing according to b, the two expressions above, which leads to

$$A_n\left(Z^{(n)}\right) = \sup_{u\in\mathcal{J}_q}\left\{\left(\frac{\sqrt{n}u'\bar{Z}_n}{\sqrt{\tilde{\rho}_1 + \tilde{\rho}_2 u'\left(S_n^2 + S^2\right)u}}\right)^2\right\} = \sup_{u\in\mathcal{J}_q}\left\{\left(\frac{\sqrt{n}u'\bar{Z}_n}{\sqrt{u'\Sigma_n^2 u + u'\Sigma^2 u}}\right)^2\right\}$$

Similarly, we have

$$C_n\left(Z^{(n)}, X^{(n)}\right) = \sup_{u\in\mathcal{J}_q}\left\{\left(\frac{\sqrt{n}u'\left(\bar{Z}_n - \bar{X}_n\right)}{\sqrt{u'\tilde{\Sigma}_n^2 u}}\right)^2\right\}$$

Now we conclude by applying lemma 1 to the inequality (7) with these expressions of $A_n\left(Z^{(n)}\right)$ and $C_n\left(Z^{(n)}, X^{(n)}\right)$ with $\mathsf{C}_2 = \mathsf{C}_1$ and $c_2 = c_1$. □

Proof of theorem 3 . We now control the Hotelling's T_n^2 in the general case, by cutting its distribution tail into two parts. The first part allows us to get back to the expression above $\sup_{u\in\mathcal{J}_q}\left\{\left(\frac{\sqrt{n}u'\bar{Z}_n}{\sqrt{u'\Sigma_n^2 u+u'\Sigma^2 u}}\right)^2\right\}$ controlled by Lemma 2. The second term is controlled by the largest eigenvalue of S^2.

Let

$$B_n = \sup_{u\in\mathcal{J}_q}\left\{\frac{u'\bar{Z}_n}{\sqrt{u'\Sigma_n^2 u}}\right\}.$$

Notice that by construction we have, for any $t > 0$, (and particularly for any $t > 2n$)

$$\left\{n\bar{Z}_n'\Sigma_n^{-2}\bar{Z}_n \geq t\right\} = \left\{n^{1/2}B_n \geq \sqrt{t}\right\}.$$

To transform the penalized self-normalised sum from the expression $n\bar{Z}_n'\left(\Sigma_n^2\right)^{-1}\bar{Z}_n$ to its "pseudo" version with the wrong normalization, $\sup_{u\in\mathcal{J}_q}\left\{\left(\frac{\sqrt{n}u'\bar{Z}_n}{\sqrt{u'\Sigma_n^2 u+u'\Sigma^2 u}}\right)^2\right\}$, let us introduce D_n defined by

$$D_n = \sup_{u\in\mathcal{J}_q}\left\{\sqrt{1 + \frac{u'\Sigma^2 u}{u'\Sigma_n^2 u}}\right\} = \sup_{u\in\mathcal{J}_q}\left\{\sqrt{1 + \frac{u'\left(\rho_1 I_q + \rho_2 S^2\right)u}{u'\left(\rho_1 I_q + \rho_2 S_n^2\right)u}}\right\}.$$

First, notice that we have

$$\begin{aligned}
\sqrt{n}\frac{B_n}{D_n} = \sup_{u\in\mathcal{J}_q}\left\{\frac{u'\bar{Z}_n}{\sqrt{u'\Sigma_n^2 u}}\right\}\inf_{u\in\mathcal{J}_q}\left\{\left(\sqrt{1 + \frac{u'\Sigma^2 u}{u'\Sigma_n^2 u}}\right)^{-1}\right\} &\leq \sup_{u\in\mathcal{J}_q}\left(\frac{u'\bar{Z}_n}{\sqrt{u'\Sigma_n^2 u}}\left(\sqrt{1 + \frac{u'\Sigma^2 u}{u'\Sigma_n^2 u}}\right)^{-1}\right) \\
&\leq \sqrt{\sup_{u\in\mathcal{J}_q}\left\{\left(\frac{\sqrt{n}u'\bar{Z}_n}{\sqrt{u'\Sigma_n^2 u+u'\Sigma^2 u}}\right)^2\right\}},
\end{aligned} \tag{8}$$

for which we have an exponential bound by Lemma 3 and theorem 1.

Thus by splitting the probability according to the event $\{D_n^2 \geq 1 + a\}$, for $a > 1$ and, for any $t > 2n$, we have

$$\begin{aligned}
\mathbb{P}\left(n\bar{Z}_n'\Sigma_n^{-2}\bar{Z}_n \geq t\right) &\leq \mathbb{P}\left(B_n \geq \sqrt{\frac{t}{n}},\ D_n \leq \sqrt{1+a}\right) + \mathbb{P}\left(D_n \geq \sqrt{1+a}\right) \\
&\leq \mathbb{P}\left(\frac{B_n}{D_n} \geq \sqrt{\frac{t}{n(1+a)}}\right) + \mathbb{P}\left(D_n \geq \sqrt{1+a}\right).
\end{aligned} \tag{9}$$

So now, it remains to treat the second term in the right-hand side of inequality (9). Notice that we have, for $a > 1$,

$$\left\{D_n \geq \sqrt{1+a}\right\} = \left\{\sup_{u\in\mathcal{J}_q}\left(\frac{u'\Sigma^2 u}{u'\Sigma_n^2 u}\right) \geq a\right\} = \left\{\inf_{u\in\mathcal{J}_q}\left(\frac{u'\Sigma_n^2 u}{u'\Sigma^2 u}\right) \leq \frac{1}{a}\right\}.$$

First, if $S^2 = \sigma^2 I_q$ is diagonal, then we have

$$u'\Sigma^2 u = u'(\rho_1 I_q + \rho_2\sigma^2 I_q)u = \rho_1 + \rho_2\sigma^2.$$

Since

$$\inf_{u\in\mathcal{J}_q}\left(u'\Sigma_n^2 u\right) = \inf_{u\in\mathcal{J}_q}\left(u'(\rho_1 I_q + \rho_2 S_n^2\right)u) = \rho_1,$$

if we choose a such that $a > (\rho_1 + \rho_2\sigma^2)/\rho_1$, then we have

$$\mathbb{P}\left[D_n \geq \sqrt{1+a}\right] \leq \mathbb{P}\left(\frac{\inf_{u\in\mathcal{J}_q}\left(u'\Sigma_n^2 u\right)}{\rho_1 + \rho_2\sigma^2} \leq \frac{1}{a}\right) = 0.$$

Remark that, in this case, we have $\rho_1^* = \sigma^2$ and $\rho_2^* = 0$ and it follows that the inequality is true for any $a > 1$. Notice that the proximity between S^2 and $\sigma^2 I_q$ is precisely controlled by the term $\alpha^2 = \left\|S^2 - \sigma^2 I_q\right\|$.

Now consider the general case. First, notice that

$$\begin{aligned}
\inf_{u\in\mathcal{J}_q}\left(\frac{u'\Sigma_n^2 u}{u'\Sigma^2 u}\right) &= \inf_{u\in\mathcal{J}_q}\left(u'\Sigma^{-1}\Sigma_n^2\Sigma^{-1}u\right) = \inf_{u\in\mathcal{J}_q}\left(\frac{u'\Sigma^{-1}}{\left\|\Sigma^{-1}u\right\|_2}\Sigma_n^2\frac{\Sigma^{-1}u}{\left\|\Sigma^{-1}u\right\|_2}\left\|\Sigma^{-1}u\right\|_2^2\right) \\
&\geq \inf_{v\in\mathcal{J}_q}\left(v'\Sigma_n^2 v\right) \times \inf_{u\in\mathcal{J}_q}\left(u'\Sigma^{-2}u\right), \quad \text{with } v = \frac{\Sigma^{-1}u}{\left\|\Sigma^{-1}u\right\|_2} \\
&\geq \rho_1\mu_1(\Sigma^{-2}) = \frac{\rho_1}{\mu_q(\Sigma^2)}.
\end{aligned}$$

Now, using the optimal values ρ_1^* and ρ_2^*, we have the decomposition

$$\Sigma^2\left(\rho_1^*, \rho_2^*\right) = \rho_1^* I_q + \rho_2^* S^2,$$

it follows that we get

$$\mu_q(\Sigma^2\left(\rho_1^*, \rho_2^*\right)) = \rho_1^* + \rho_2^*\mu_q(S^2)$$

and

$$\inf_{u\in\mathcal{J}_q}\left(\frac{u'\Sigma_n^2\left(\rho_1^*, \rho_2^*\right)u}{u'\Sigma^2\left(\rho_1^*, \rho_2^*\right)u}\right) \geq \frac{\rho_1^*}{\rho_1^* + \rho_2^*\mu_q(S^2)}.$$

It follows that if we choose a such that

$$\frac{1}{a} < \frac{1}{1 + \frac{\mu_q(S^2)}{\rho^*}}$$

and, since $a^* = 1 + \frac{K_3}{\rho^*} > 1 + \frac{\mu_q(S^2)}{\rho^*}$ by the assumption (A_3), then, if $a \geq a^*$, we get

$$\mathbb{P}\left(D_n \geq \sqrt{1+a}\right) = 0. \tag{10}$$

As a consequence, we obtain an exponential inequality for any value $a \geq a^*$. Combining (9) and (10), we get, for any $a \geq a^*$,

$$\forall t > 2n, \quad \mathbb{P}\left(n\bar{Z}_n'\Sigma_n^{-2}\bar{Z}_n \geq t\,(1+a)\right) \leq \mathbb{P}\left(\sqrt{n}\frac{B_n}{D_n} \geq \sqrt{t}\right). \tag{11}$$

Let $X^{(n)} = (X_i)_{1\le i\le n}$ be an independent copy of $Z^{(n)} = (Z_i)_{1\le i\le n}$. Applying theorem 1 to $(Z_i - X_i)_{1\le i\le n}$ which is symmetric, we obtain

$$\mathbb{P}\left(\sup_{u\in\mathcal{J}_q}\left(\frac{\sqrt{n}u'\left(\bar{Z}_n - \bar{X}_n\right)}{\sqrt{u'\tilde{\Sigma}_n^2 u}}\right) \ge \sqrt{t}\right) \le \frac{2e^3}{9}\bar{F}_n(t),$$

Thus, applying the lemma 3 to the inequality above implies that, for all $t \ge 2n$,

$$\mathbb{P}\left(\sup_{u\in\mathcal{J}_q}\left(\frac{\sqrt{n}u'\bar{Z}_n}{\sqrt{u'\left(\Sigma_n^2 + \Sigma^2\right)u}}\right) \ge \sqrt{t}\right) \le \frac{2e^3}{9}\left(\frac{(t-n)}{2}\right)^{n/2}\frac{e^{-\frac{(t-n)}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}. \tag{12}$$

Finally by combining expressions (8), (11) and (12), the result holds. □

Proof of Theorem 4

The following lemmas will allow us to control explicitly the deviation $\mathbb{P}\left[\left|\frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*}\right| > \epsilon\right]$ for small positive values of ϵ.

Lemma 4. (Inversion) *Let $w > 0$, and consider $(W_n)_{n\ge1}$ a sequence of positive random variables. Assume that there exists a nonnegative constant C_3, such that $\forall\epsilon > 0, \exists N > 0, \forall n > N$,*

$$\mathbb{P}\left(|W_n - w| > \epsilon\right) \le \frac{C_3}{n}\frac{1}{\epsilon^2}.$$

Then there exists a function $C_{3;1/w}$ nonnegative, such that $\forall\epsilon > 0, \forall n > N$

$$\mathbb{P}\left(\left|\frac{1}{W_n} - \frac{1}{w}\right| > \epsilon\right) \quad\le\quad \frac{C_{3;1/w}(\epsilon)}{n\epsilon^2},$$

where $C_{3;1/w}(\epsilon) = \frac{C_3}{w^4}\left(1 + (w\epsilon)^{2/5}\right)^5$.

Proof of Lemma 4 . Since $w > 0$, we have

$$\mathbb{P}\left(\left|\frac{1}{W_n} - \frac{1}{w}\right| > \frac{\epsilon}{w}\right) = \mathbb{P}\left(\left|\frac{w}{W_n} - 1\right| > \epsilon\right)$$

Now, $\forall\eta \in]0, w[$ we get

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{W_n} - \frac{1}{w}\right| > \frac{\epsilon}{w}\right) &\le \mathbb{P}\left(\left|\frac{w}{W_n} - 1\right| > \epsilon,\ |W_n - w| \le \eta\right) + \mathbb{P}\left(|W_n - w| > \eta\right)\\ &\le \text{(I)} + \text{(II)}.\end{aligned}$$

On the interval $[w-\eta; w+\eta]$, $f : x \mapsto \frac{w}{x}$ is Lipschitz with $|f'(x)| \le \frac{w}{(w-\eta)^2}$, thus we obtain

$$\forall W_n \in [w-\eta; w+\eta],\ \left|\frac{w}{W_n} - 1\right| \le \frac{w}{(w-\eta)^2}|W_n - w|.$$

$$\forall\eta \in]0; w[,\ (I) \le \mathbb{P}\left(\frac{w}{(w-\eta)^2}|W_n - w| > \epsilon\right) \le \frac{C_3}{n} \times \frac{w^2}{\epsilon^2(w-\eta)^4}$$

and since

$$\forall\eta \in]0; w[,\ (II) \le \frac{C_3}{n} \times \frac{1}{\eta^2},$$

it follows that

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{W_n}-\frac{1}{w}\right|>\frac{\epsilon}{w}\right) &\leq \frac{\mathsf{C}_3}{n}\times\frac{w^2}{\epsilon^2(w-\eta)^4}+\frac{\mathsf{C}_3}{n}\times\frac{1}{\eta^2}\leq\frac{\mathsf{C}_3}{n}\min_{\eta\in]0;w[}\left\{\frac{w^2}{\epsilon^2\left(1-\frac{\eta}{w}\right)^4w^4}+\frac{1}{w^2\left(\frac{\eta}{w}\right)^2}\right\}\\ &\overset{\alpha=\frac{\eta}{w}}{\leq} \frac{\mathsf{C}_3}{nw^2}\min_{\alpha\in]0;1[}\left\{\frac{1}{\epsilon^2(1-\alpha)^4}+\frac{1}{\alpha^2}\right\}\leq\frac{\mathsf{C}_3}{nw^2}\min_{\alpha\in]0;1[}\left\{\frac{1}{\epsilon^2(1-\alpha)^4}+\frac{1}{\alpha^4}\right\}\\ &\leq \frac{\mathsf{C}_3}{nw^2}\left(1+\epsilon^{-2/5}\right)^5.\end{aligned}$$

Setting $\epsilon'=\frac{\epsilon}{w}$ and $\mathsf{C}_{3;1/w}(\epsilon')=\epsilon'^2\times\frac{\mathsf{C}_3}{w^2}\left(1+(w\epsilon')^{-2/5}\right)^5=\frac{\mathsf{C}_3}{w^4}\left(1+(w\epsilon')^{2/5}\right)^5$, the result holds. □

Lemma 5. (Product) *Consider u, v two positive scalars, and (U_n), (V_n) some random sequences. Assume that there exists nonnegative constants $\tilde{\mathsf{C}}_4$ and $\check{\mathsf{C}}_4$ such that $\forall\epsilon>0,\forall n\geq 1$:*

$$\mathbb{P}(|U_n-u|>\epsilon)\leq\frac{\tilde{\mathsf{C}}_4}{n}\frac{1}{\epsilon^2}\quad\textit{and}\quad\mathbb{P}(|V_n-v|>\epsilon)\leq\frac{\check{\mathsf{C}}_4}{n}\frac{1}{\epsilon^2}.$$

Then there exists a function $\mathsf{C}_{4;uv}$ such that $\forall\epsilon>0$,

$$\mathbb{P}(|U_nV_n-uv|>\epsilon)\leq\frac{\mathsf{C}_{4;uv}(\epsilon)}{n}\frac{1}{\epsilon^2},$$

where $\mathsf{C}_{4;uv}(\epsilon)=\tilde{\mathsf{C}}_4\left(\frac{2uv+\epsilon}{u}\right)^2+\check{\mathsf{C}}_4(2u)^2$ is a positive function of ϵ depending on u, v, $\tilde{\mathsf{C}}_4$ and $\check{\mathsf{C}}_4$.

Proof of Lemma 5. By straightforward inequalities, we get

$$\begin{aligned}\mathbb{P}(|U_nV_n-uv|>\epsilon)=\mathbb{P}(|U_nV_n-uV_n+uV_n-uv|>\epsilon) &\leq \mathbb{P}\left(V_n|U_n-u|>\frac{\epsilon}{2},\,u|V_n-v|\leq\frac{\epsilon}{2}\right)+\mathbb{P}\left(u|V_n-v|>\frac{\epsilon}{2}\right)\\ &\leq \mathbb{P}\left(\left(v+\frac{\epsilon}{2u}\right)|U_n-u|>\frac{\epsilon}{2}\right)+\mathbb{P}\left(|V_n-v|>\frac{\epsilon}{2u}\right)\\ &\leq \mathbb{P}\left(|U_n-u|>\frac{\epsilon u}{2uv+\epsilon}\right)+\mathbb{P}\left(|V_n-v|>\frac{\epsilon}{2u}\right)\\ &\leq \frac{\tilde{\mathsf{C}}_4}{n}\left(\frac{2uv+\epsilon}{\epsilon u}\right)^2+\frac{\check{\mathsf{C}}_4}{n}\left(\frac{2u}{\epsilon}\right)^2\leq\frac{\mathsf{C}_{4;uv}(\epsilon)}{n}\frac{1}{\epsilon^2}.\end{aligned}$$

□

Lemma 6. Proximity between $\sigma^2,\alpha^2,\beta^2,\delta^2$ and their estimators *Let $u^2\in\left\{\sigma^2,\alpha^2,\beta^2,\delta^2\right\}$ be one of these quantities of interest and $\hat{u}_n^2$ its corresponding estimator. Then $\forall n\geq 1$ and $\forall\epsilon>0$, we have :*

$$\mathbb{P}\left(\left|\hat{u}_n^2-u^2\right|>\epsilon\right)\leq\frac{C_{u^2}(\epsilon)}{n\epsilon^2},$$

with

- $C_{\sigma^2}=\sqrt{K_2}$ *for the case where* $u^2=\sigma^2$ *and* $\hat{u}_n^2=\sigma_n^2$,
- $C_{\delta^2}=2K_4+(100+K_1^2)K_2+2^4\sqrt{6}K_2^{5/4}+4K_2^{3/2}+2^23K_2^2+4K_2^{1/2}\left(K_2^{1/4}+2\sqrt{6}\right)\sqrt{K_1^2K_2+4K_2(1+3K_2)+2K_4}$ *for the case where* $u^2=\delta^2$ *and* $\hat{u}_n^2=\delta_n^2$,
- $C_{\beta^2}(\epsilon)=4K_1^2\sqrt{K_2}+C_{\delta^2}+2K_1\sqrt{K_2}\,\epsilon$ *for the case where* $u^2=\beta^2$ *and* $\hat{u}_n^2=\beta_n^2$,
- $C_{\alpha^2}(\epsilon)=2^3C_{\delta^2}+2^4K_1^2\sqrt{K_2}+2^2K_1\sqrt{K_2}\,\epsilon$. *for the case where* $u^2=\alpha^2$ *and* $\hat{u}_n^2=\alpha_n^2$.

Proof of Lemma 6.

Consider $\hat{\sigma}_n^2$ and σ^2.

Recall that $\hat{\sigma}_n^2 = \frac{1}{q}\sum_{j=1}^q \left(\frac{1}{n}\sum_{i=1}^n y_{ij}^2\right)$ and $\sigma^2 = \frac{1}{q}\sum_{j=1}^q \mathbb{E}\left[y_{1j}^2\right] = \frac{1}{q}\sum_{j=1}^q \mu_j$.

Following the ideas of Ledoit and Wolf [11] who obtain the convergence of the fourth order moment, we rather control the second order moment as follows :

$$\begin{aligned}\mathbb{E}\left[\left(\hat{\sigma}_n^2-\sigma^2\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^q\frac{1}{n}\sum_{i=1}^n\left(y_{ij}^2-\mu_j\right)\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n\frac{1}{q}\sum_{j=1}^q\left(y_{ij}^2-\mu_j\right)\right)^2\right]\\ &= \frac{1}{n^2}\sum_{i_1=1}^n\sum_{i_2=1}^n\mathbb{E}\left[\frac{1}{q}\sum_{j=1}^q\left(y_{i_1 j}^2-\mu_j\right)\times\frac{1}{q}\sum_{j=1}^q\left(y_{i_2 j}^2-\mu_j\right)\right].\end{aligned}$$

This last expression is equal to zero for any $i_1 \neq i_2$ because of the independence between observations. Thus we get

$$\begin{aligned}\mathbb{E}\left[\left(\hat{\sigma}_n^2-\sigma^2\right)^2\right] &= \frac{1}{n^2}\sum_{i=1}^n\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^q\left(y_{1j}^2-\mu_j\right)\right)^2\right]\\ &= \frac{1}{n}\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^q\left(y_{1j}^2-\mu_j\right)\right)^2\right] = \frac{1}{n}\left(\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^q y_{1j}^2\right)^2\right]-\left(\mathbb{E}\left[\frac{1}{q}\sum_{j=1}^q y_{1j}^2\right]\right)^2\right)\\ &\leq \frac{1}{n}\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^q y_{1j}^2\right)^2\right] \leq \frac{1}{n}\left(\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^q y_{1j}^2\right)^4\right]\right)^{1/2} \leq \frac{1}{n}\left(\frac{1}{q}\sum_{j=1}^q\mathbb{E}\left[y_{1j}^8\right]\right)^{1/2}.\end{aligned}$$

Therefore, using the second assumption (A_2), one gets

$$\mathbb{E}\left[\left(\hat{\sigma}_n^2-\sigma^2\right)^2\right] \leq \frac{\sqrt{K_2}}{n}. \tag{13}$$

Finally, we have by Markov inequality the bound

$$\mathbb{P}\left[\left|\hat{\sigma}_n^2-\sigma^2\right| > \epsilon\right] \leq \frac{\mathbb{E}\left[\left(\hat{\sigma}_n^2-\sigma^2\right)^2\right]}{\epsilon^2} \leq \frac{\sqrt{K_2}}{n\epsilon^2}.$$

Consider $\hat{\delta}_n^2$ and δ^2.

Combining the expressions (A.2) and (A.3) on page 394 in Ledoit and Wolf ([11]) we get

$$\hat{\delta}_n^2-\delta^2 = \left(\hat{\sigma}_n^2-\sigma^2\right)^2 - 2\sigma^2\left(\hat{\sigma}_n^2-\sigma^2\right) + \left\|S_n^2\right\|^2 - \mathbb{E}\left(\left\|S_n^2\right\|^2\right).$$

Similarly using their expressions, from page 394 (A.4) to page 399, and page 390 (A.1), we have respectively the inequalities

$$\mathrm{Var}\left(\left\|S_n^2\right\|^2\right) \leq \frac{1}{n}\left(K_1^2K_2 + 4K_2\left(1+3K_2\right)+2K_4\right) \quad \text{and} \quad \sigma^2 \leq \sqrt{K_2}.$$

Combining these expressions with Bienaymé-Tchebychev, Markov and Cauchy-Schwartz inequalities, we obtain a control of $\mathbb{P}\left(\left|\hat{\delta}_n^2-\delta^2\right| > \epsilon\right)$ by a function of n, ϵ, A_2, A_4 and $\mathrm{Var}(\|S_n^2\|^2)$ where $A_k = \mathbb{E}\left(\left|\hat{\sigma}_n^2-\sigma^2\right|^k\right)$. Indeed we have, by Markov inequality, for all $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\left|\hat{\delta}_n^2-\delta^2\right| > \epsilon\right) &\leq \frac{1}{\epsilon^2}\left\{\mathbb{E}\left[\left(\hat{\sigma}_n^2-\sigma^2\right)^4\right] + 4\sigma^4\mathbb{E}\left[\left(\hat{\sigma}_n^2-\sigma^2\right)^2\right] + \mathbb{E}\left[\left(\left\|S_n^2\right\|^2-\mathbb{E}\left\|S_n^2\right\|^2\right)^2\right] + 4\sigma^2\mathbb{E}\left[\left|\hat{\sigma}_n^2-\sigma^2\right|^3\right]\right.\\ &\quad \left. +4\sigma^2\mathbb{E}\left[\left|\hat{\sigma}_n^2-\sigma^2\right|\left(\left\|S_n^2\right\|^2-\mathbb{E}\left(\left\|S_n^2\right\|^2\right)\right)\right] + 2\mathbb{E}\left[\left(\hat{\sigma}_n^2-\sigma^2\right)^2\left|\left\|S_n^2\right\|^2-\mathbb{E}\left(\left\|S_n^2\right\|^2\right)\right|\right]\right\}\\ &\leq \frac{1}{\epsilon^2}\left\{A_4 + 4\sigma^4A_2 + \mathrm{Var}\left(\left\|S_n^2\right\|^2\right) + 4\sigma^2\sqrt{A_2A_4} + 4\sigma^2\sqrt{A_2\mathrm{Var}\left(\left\|S_n^2\right\|^2\right)} + 2\sqrt{A_4\mathrm{Var}\left(\left\|S_n^2\right\|^2\right)}\right\}.\end{aligned}$$

Now by some previous controls established by Ledoit and Wolf ([11], page 394) we have

$$A_4 \leq \frac{96K_2}{n}\ ; \quad \mathrm{Var}\left(\left\|S_n^2\right\|^2\right) \leq \frac{1}{n}\left(K_1^2K_2 + 4K_2\,(1+3K_2) + 2K_4\right) = \frac{1}{n}K \quad \text{and}\ \ \sigma^2 \leq \sqrt{K_2}.$$

Using the control stated in (13), $A_2 \leq \sqrt{K_2}/n$, we can easily get the explicit constant C_{δ^2} as a function of K_1, K_2, and K_4. For all $\epsilon > 0$, for all $n \in \mathbb{N}^*$, we have

$$\begin{aligned}
\mathbb{P}\left(\left|\hat{\delta}_n^2 - \delta^2\right| > \epsilon\right) &\leq \frac{1}{n\epsilon^2}\left[96K_2 + 4K_2K_2^{1/2} + K + 4K_2^{1/2}\sqrt{96K_2^{1/2}K_2} + 4K_2^{1/2}\sqrt{K_2^{1/2}K} + 2\sqrt{96K_2K}\right] \\
&\leq \frac{1}{n\epsilon^2}\left\{2K_4 + (100+K_1^2)K_2 + 2^4\sqrt{6}K_2^{5/4} + 4K_2^{3/2} + 2^2 3K_2^2\right. \\
&\quad \left. + 4K_2^{1/2}\left(K_2^{1/4} + 2\sqrt{6}\right)\sqrt{K_1^2K_2 + 4K_2\,(1+3K_2) + 2K_4}\right\} \leq \frac{\mathsf{C}_{\delta^2}}{n\epsilon^2}.
\end{aligned}$$

Consider $\hat{\beta}_n^2$ and β^2.

Since $\delta^2 = \alpha^2 + \beta^2$ yielding $\delta^2 \geq \beta^2$, Ledoit and Wolf showed ([11], proof of Lemma 3.4 page 401, lines from -12 to -6) that

$$-\max\left(|\bar{\beta}_n^2 - \beta^2|, |\hat{\delta}_n^2 - \delta^2|\right) \leq \hat{\beta}_n^2 - \beta^2 \leq |\bar{\beta}_n^2 - \beta^2|.$$

From this we deduce

$$|\hat{\beta}_n^2 - \beta^2| \quad \leq \quad \max\left\{\max\left(|\bar{\beta}_n^2 - \beta^2|, |\hat{\delta}_n^2 - \delta^2|\right), |\bar{\beta}_n^2 - \beta^2|\right\} \leq \max\left(|\bar{\beta}_n^2 - \beta^2|, |\hat{\delta}_n^2 - \delta^2|\right).$$

Controlling $|\bar{\beta}_n^2 - \beta^2|$ leads to a control for $|\hat{\delta}_n^2 - \delta^2|$ and $|\hat{\beta}_n^2 - \beta^2|$. By the same arguments as in [11] (proof of Lemma 3.4, page 399, equation (A.7)), we have the following expression

$$\bar{\beta}_n^2 - \beta^2 = \frac{1}{n}\|S_n^2 - S^2\|^2 + \left(\frac{1}{n^2}\sum_{i=1}^n \|Z_iZ_i^{'} - S^2\|^2 - \mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^n \|Z_iZ_i^{'} - S^2\|^2\right]\right).$$

Now, splitting the probability into two terms, on the one hand, using Markov inequality on the first term and applying Bienaymé-Tchebychev inequality to the second term, we get

$$\mathbb{P}\left(|\bar{\beta}_n^2 - \beta^2| > \epsilon\right) \leq \frac{2}{\epsilon}\mathbb{E}\left(\frac{1}{n}\|S_n^2 - S^2\|^2\right) + \frac{4}{\epsilon^2}\mathrm{Var}\left(\frac{1}{n^2}\sum_{i=1}^n \|Z_iZ_i^{'} - S^2\|^2\right).$$

Following Ledoit and Wolf ([11], proof of Lemma 3.1 page 391 line +5), we have

$$\mathbb{E}\left(\left\|S_n^2 - S^2\right\|^2\right) \leq K_1\sqrt{K_2}.$$

Moreover, we have (in the proof of Lemma 3.4, page 401 line +3)

$$\mathrm{Var}\left(\frac{1}{n^2}\sum_{i=1}^n \left\|Z_iZ_i' - S^2\right\|^2\right) \leq K_1^2\sqrt{K_2}/n.$$

We obtain

$$\mathbb{P}\left(\left|\bar{\beta}_n^2 - \beta^2\right| > \epsilon\right) \leq \frac{2}{\epsilon}\frac{K_1\sqrt{K_2}}{n} + \frac{4}{\epsilon^2}\frac{K_1^2\sqrt{K_2}}{n}.$$

Finally, with $\mathbb{P}\left(\left|\hat{\delta}_n^2 - \delta^2\right| > \epsilon\right) \leq \frac{C_{\delta^2}}{n\epsilon^2}$ and

$$\mathbb{P}\left(\left|\hat{\beta}_n^2 - \beta^2\right| > \epsilon\right) \leq \mathbb{P}\left(\left|\bar{\beta}_n^2 - \beta^2\right| > \epsilon\right) + \mathbb{P}\left(\left|\hat{\delta}_n^2 - \delta^2\right| > \epsilon\right),$$

we obtain

$$\mathbb{P}\left(\left|\hat{\beta}_n^2 - \beta^2\right| > \epsilon\right) \leq \frac{1}{n\epsilon^2}\left(4K_1^2\sqrt{K_2} + \mathsf{C}_{\delta^2} + 2K_1\sqrt{K_2}\,\epsilon\right) \leq \frac{C_{\beta^2}(\epsilon)}{n\epsilon^2}.$$

Remark that $C_{\beta^2}(\epsilon)$ tends to $4K_1^2\sqrt{K_2} + \mathsf{C}_{\delta^2}$ when ϵ tends to 0.

Consider $\hat{\alpha}_n^2$ and α^2.
Since we have $\hat{\alpha}_n^2 = \hat{\delta}_n^2 - \hat{\beta}_n^2$ and $\alpha^2 + \beta^2 = \delta^2$, one can easily see that $\hat{\alpha}_n^2 - \alpha^2 = \hat{\delta}_n^2 - \hat{\beta}_n^2 - \delta^2 + \beta^2$. For all $\epsilon > 0$, we get

$$\begin{aligned}\mathbb{P}\left(\left|\hat{\alpha}_n^2 - \alpha^2\right| > \epsilon\right) &\leq \mathbb{P}\left(\left|\hat{\delta}_n^2 - \delta^2\right| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\left|\hat{\beta}_n^2 - \beta^2\right| > \frac{\epsilon}{2}\right) \\ &\leq \frac{2^2\mathsf{C}_{\delta^2}}{n\epsilon^2} + \frac{2^2\mathsf{C}_{\beta^2}(\epsilon/2)}{n\epsilon^2} \leq \frac{1}{n\epsilon^2}\left(2^3\mathsf{C}_{\delta^2} + 2^4K_1^2\sqrt{K_2} + 2^2K_1\sqrt{K_2}\,\epsilon\right) \leq \frac{C_{\alpha^2}(\epsilon)}{n\epsilon^2}.\end{aligned}$$

Remark that $C_{\alpha^2}(\epsilon)$ tends to $2^3\mathsf{C}_{\delta^2} + 2^4K_1^2\sqrt{K_2}$ when ϵ tends to 0. □

In the next lemma 7, we control the proximity between $1/\hat{\rho}_n^*$ and $1/\rho^*$, that we denote $g_n(\epsilon)$ and show that it is of order $O(1/n)$. For this, we first apply product lemma 5 to $\hat{\beta}_n^2$ and $\hat{\sigma}_n^2$. Then, we apply the inverse lemma 4 to $\hat{\beta}_n^2\hat{\sigma}_n^2$. Finally, we use another time product lemma 5 applied to $\hat{\alpha}_n^2$ and $1/\hat{\beta}_n^2\hat{\sigma}_n^2$.

Lemma 7. Proximity between $1/\rho^*$ and $1/\hat{\rho}_n^*$
For any $\epsilon > 0$, we have

$$g_n(\epsilon) = \mathbb{P}\left(\left|\frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*}\right| > \epsilon\right) \leq \frac{G(\epsilon)}{n\epsilon^2}$$

with C_{β^2} and C_{α^2} defined in lemma 6 and

$$G(\epsilon) = \mathsf{C}_{3;1/\beta^2\sigma^2}(\epsilon)\left(2\alpha^2 + \epsilon\beta^2\sigma^2\right)^2 + \frac{2^2C_{\alpha^2}(\epsilon)}{\beta^4\sigma^4}$$

$$\mathsf{C}_{3;1/\beta^2\sigma^2}(\epsilon) = \left[K_2^{1/2}\frac{\left(2\sigma^2\beta^2 + \epsilon\right)^2}{\beta^8\sigma^{12}} + \frac{2^2C_{\beta^2}(\epsilon)}{\beta^8\sigma^4}\right]\left(1 + \left(\beta^2\sigma^2\epsilon\right)^{2/5}\right)^5.$$

Remark : the function $\mathsf{C}_{3;1/\beta^2\sigma^2}(\epsilon)$ may be clearly bounded by a polynomial of degree 4 in ϵ. As a consequence, the function $G(\epsilon)$ may be bounded by a polynomial of degree 6.

Proof of Lemma 7. We apply the product lemma 5 to obtain a control for $\hat{\beta}_n^2\hat{\sigma}_n^2$ thanks to lemma 6 which gives us some control of $\hat{\sigma}_n^2$ and $\hat{\beta}_n^2$. For all $\epsilon > 0$, one gets

$$\mathbb{P}\left(\left|\hat{\beta}_n^2\hat{\sigma}_n^2 - \beta^2\sigma^2\right| > \epsilon\right) \leq \frac{C_{4;\beta^2\sigma^2}(\epsilon)}{n\epsilon^2}, \tag{14}$$

with

$$C_{4;\sigma^2\beta^2}(\epsilon) = K_2^{1/2}\left(\frac{2\sigma^2\beta^2 + \epsilon}{\sigma^2}\right)^2 + C_{\beta^2}(\epsilon)\left(2\sigma^2\right)^2.$$

We now apply the inverse lemma 4 with inequality (14) and obtain a control of $1/\hat{\beta}_n^2\hat{\sigma}_n^2$. That is, for all $\epsilon > 0$, we have

$$\mathbb{P}\left(\left|\frac{1}{\hat{\beta}_n^2\hat{\sigma}_n^2} - \frac{1}{\beta^2\sigma^2}\right| > \epsilon\right) \leq \frac{C_{3;1/\beta^2\sigma^2}(\epsilon)}{n\epsilon^2},$$

with $\mathsf{C}_{3;1/\beta^2\sigma^2}$ defined by

$$\mathsf{C}_{3;1/\beta^2\sigma^2}(\epsilon) \quad = \quad \frac{\mathsf{C}_{4;\beta^2\sigma^2}(\epsilon)}{\beta^8\sigma^8}\left(1 + \left(\beta^2\sigma^2\epsilon\right)^{2/5}\right)^5.$$

Applying the product lemma 5 with $u = 1/(\beta^2\sigma^2)$ and $v = \alpha^2$, we obtain

$$\mathsf{C}_{4;1/\rho^*}(\epsilon) = \mathsf{C}_{3;1/\beta^2\sigma^2}(\epsilon)\left(2\alpha^2 + \epsilon\beta^2\sigma^2\right)^2 + \mathsf{C}_{\alpha^2}(\epsilon)\frac{2^2}{\beta^4\sigma^4}.$$

Remark that when ϵ tends to 0, $\mathsf{C}_{4;1/\rho^*}$ tends to

$$\frac{2^4\alpha^4 K_2^{1/2}}{\beta^4\sigma^8} + \frac{2^6\alpha^4\left(2^2K_1^2\sqrt{K_2}+\mathsf{C}_{\delta^2}\right)}{\beta^8\sigma^4} + \frac{2^5\left(2K_1^2\sqrt{K_2}+\mathsf{C}_{\delta^2}\right)}{\beta^4\sigma^4}$$

$\square$

Proof of theorem 4. Recall that $\hat{a}_n^* = 1 + \frac{K_3}{\hat{\rho}_n^*}$ and $a^* = 1 + \frac{K_3}{\rho^*}$. For any $u > 2n$, we have for $\epsilon > 0$

$$\mathbb{P}\left(n\bar{Z}_n'\hat{\Sigma}_n^{*-2}\bar{Z}_n \geq u\left(1+\hat{a}_n^*+2\epsilon\right)\right) \leq \mathbb{P}\left(n\bar{Z}_n'\hat{\Sigma}_n^{*-2}\bar{Z}_n \geq u\left(1+a^*+\epsilon\right)\right) + \quad \mathbb{P}\left(|\hat{a}_n - a^*| \geq \epsilon\right) \quad \leq \quad \text{(I)} + \text{ (II)}.$$

We start by establishing a control for (I). Define $\Delta_n = n\bar{Z}_n'\left(\hat{\Sigma}_n^{*-2} - \Sigma_n^{*-2}\right)\bar{Z}_n$, then we have

$$\text{(I)} \quad = \quad \mathbb{P}\left(n\bar{Z}_n'\Sigma_n^{*-2}\bar{Z}_n + \Delta_n \geq u\left(1+a^*+\epsilon\right)\right).$$

Since $u > 2n > n$, we have

$$\begin{aligned} \text{(I)} \quad &\leq \quad \mathbb{P}\left(n\bar{Z}_n'\Sigma_n^{*-2}\bar{Z}_n + \Delta_n \geq u\left(1+a^*+\epsilon\right), |\Delta_n| \leq \epsilon n\right) + \mathbb{P}\left(|\Delta_n| > \epsilon n\right) \\ &\leq \quad \mathbb{P}\left(n\bar{Z}_n'\Sigma_n^{*-2}\bar{Z}_n \geq u\left(1+a^*+\epsilon\right) - \epsilon n\right) + \mathbb{P}\left(|\Delta_n| > \epsilon n\right) \\ &\leq \quad \mathbb{P}\left(n\bar{Z}_n'\Sigma_n^{*-2}\bar{Z}_n \geq u\left(1+a^*\right)\right) + \mathbb{P}\left(|\Delta_n| > \epsilon n\right). \end{aligned} \tag{15}$$

Theorem 3 gives us an exponential bound controlling the first term of the right hand of the inequality when $a = a^*$ and $u > 2n$.
Now use the following matrix factorisation $A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}$ to control the second term in the right hand with $A = \hat{\Sigma}_n^{*2}$ and $B = \Sigma_n^{*2}$. It is easy to see that $B - A = (\rho^* - \hat{\rho}_n^*)I_q$, then we obtain

$$\begin{aligned} \Delta_n = Tr(\Delta_n) \quad &= \quad Tr\left(n\bar{Z}_n'\left(\hat{\Sigma}_n^{*-2} - \Sigma_n^{*-2}\right)\bar{Z}_n\right) \\ &= \quad n\left(\rho^* - \hat{\rho}_n^*\right)Tr\left(\bar{Z}_n'\hat{\Sigma}_n^{*-2}\Sigma_n^{*-2}\bar{Z}_n\right). \end{aligned}$$

Recall that

$$\Sigma_n^{*2} = S_n^2 + \rho^* I_q = O_n'\Lambda_n^2 O_n + \rho^* I_q = O_n'\left(\Lambda_n^2 + \rho^* I_q\right)O_n.$$

then using the same rotation matrix O_n, we obtain $\hat{\Sigma}_n^{*-2}\Sigma_n^{*-2} = O_n'DO_n$, with

$$D = \begin{pmatrix} \frac{1}{(\lambda_1+\rho^*)(\lambda_1+\hat{\rho}_n^*)} & & & & & \\ & \ddots & & & 0 & \\ & & \frac{1}{(\lambda_n+\rho^*)(\lambda_n+\hat{\rho}_n^*)} & & & \\ & & & \frac{1}{\rho^*\hat{\rho}_n^*} & & \\ & 0 & & & \ddots & \\ & & & & & \frac{1}{\rho^*\hat{\rho}_n^*} \end{pmatrix}.$$

It follows that

$$\Delta_n \quad = \quad n\left(\rho^* - \hat{\rho}_n^*\right)Tr\left(\bar{Z}_n'O_n'D^{\frac{1}{2}}D^{\frac{1}{2}}O_n\bar{Z}_n\right) = \left(\rho^* - \hat{\rho}_n^*\right)Tr\left(\left(D^{\frac{1}{2}}n^{\frac{1}{2}}O_n\bar{Z}_n\right)'\left(D^{\frac{1}{2}}n^{\frac{1}{2}}O_n\bar{Z}_n\right)\right) = \left(\rho^* - \hat{\rho}_n^*\right)\left\|D^{\frac{1}{2}}n^{\frac{1}{2}}\bar{Y}_n\right\|_2^2.$$

Since, for any x in $\mathbb{R}^q$, $\|D^{\frac{1}{2}}x\|_2^2 \leq \frac{1}{\rho^*\hat{\rho}_n^*}\|x\|_2^2$, and because we have $\|x\|_2^2 = q\|x\|^2$, we get

$$|\Delta_n| \quad \leq \quad \frac{\left|\rho^* - \hat{\rho}_n^*\right|}{\rho^*\hat{\rho}_n^*}\left\|n^{\frac{1}{2}}\bar{Y}_n\right\|_2^2 \leq \left|\frac{1}{\rho^*} - \frac{1}{\hat{\rho}_n^*}\right| q \left\|n^{\frac{1}{2}}\bar{Y}_n\right\|^2.$$

Lemma 7 gives a control of the first term on the right-hand side of this inequality so that it is sufficient to control the second term. Write

$$\left\|n^{\frac{1}{2}}\bar{Y}_n\right\|^2 = \frac{1}{qn}\sum_{j=1}^{q}\left(\sum_{i=1}^{n} Y_{i,j}\right)^2 \quad = \quad \frac{1}{qn}\sum_{j=1}^{q}\sum_{i=1}^{n} Y_{i,j}^2 + \frac{1}{qn}\sum_{j=1}^{q}\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i'\neq i}}^{n} Y_{i,j}Y_{i',j} = \mathrm{I}_1 + \mathrm{I}_2$$

Since $\mathbb{E}\left(\mathrm{I}_1\right) = \mathbb{E}\left(\frac{1}{qn}\sum_{j=1}^{q}\sum_{i=1}^{n} Y_{i,j}^2\right) = \sigma^2$, use Bienaymé-Tchebychev inequality and the independence of the Y_i's to get

$$\begin{aligned}\mathbb{P}\left(\frac{1}{qn}\sum_{j=1}^{q}\sum_{i=1}^{n} Y_{i,j}^2 - \sigma^2 > \frac{\epsilon}{2}\right) \quad &\leq \quad \mathbb{P}\left(\left|\frac{1}{qn}\sum_{j=1}^{q}\sum_{i=1}^{n} Y_{i,j}^2 - \sigma^2\right| > \frac{\epsilon}{2}\right) \leq \frac{4}{\epsilon^2}\mathrm{Var}\left(\frac{1}{qn}\sum_{j=1}^{q}\sum_{i=1}^{n} Y_{i,j}^2\right) \\ &\leq \quad \frac{4}{\epsilon^2}\frac{1}{nq^2}\mathbb{E}\left(\left(\sum_{j=1}^{q} Y_{1,j}^2\right)^2\right). \end{aligned} \tag{16}$$

Then, by hypothesis (A_2), we have $\mathbb{E}\left(\frac{1}{q}\sum_{j=1}^{q} Y_{1,j}^4\right) \leq \sqrt{K_2}$. Then, by Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}\frac{1}{nq^2}\mathbb{E}\left(\left(\sum_{j=1}^{q} Y_{1,j}^2\right)^2\right) \quad &\leq \quad \frac{1}{nq}\mathbb{E}\left(\frac{1}{q}\sum_{j=1}^{q} Y_{1,j}^4\right) + \frac{1}{nq^2}\sum_{j=1}^{q}\sum_{\substack{k=1\\ k\neq j}}^{q}\mathbb{E}\left(Y_{1,j}^2 Y_{1,k}^2\right) \\ &\leq \quad \frac{1}{nq}\sqrt{K_2} + \frac{1}{nq^2}\sum_{j=1}^{q}\sum_{\substack{k=1\\ k\neq j}}^{q}\sqrt{\mathbb{E}\left(Y_{1,j}^4\right)}\sqrt{\mathbb{E}\left(Y_{1,k}^4\right)} \\ &\leq \quad \frac{1}{nq}\sqrt{K_2} + \frac{1}{n}\left(\frac{1}{q}\sum_{j=1}^{q}\sqrt{\mathbb{E}\left(Y_{1,j}^4\right)}\right)^2 \leq \frac{1}{nq}\sqrt{K_2} + \frac{1}{n}\mathbb{E}\left(\frac{1}{q}\sum_{j=1}^{q} Y_{1,j}^4\right) \\ &\leq \quad \frac{1}{n}\sqrt{K_2}\left(\frac{1}{q}+1\right). \end{aligned} \tag{17}$$

Finally, combining inequalities (16, 17), we get the following control for I_1, for $\eta > 0$

$$\mathbb{P}\left(\mathrm{I}_1 - \mathbb{E}(\mathrm{I}_1) > \frac{\eta}{2}\right) \quad \leq \quad \frac{4}{\eta^2}\frac{1}{n}\sqrt{K_2}\left(\frac{1}{q}+1\right). \tag{18}$$

Now, we focus on I_2. Using the independence between the observations Y_i's, we have

$$\mathbb{E}\left(\mathrm{I}_2\right) = \mathbb{E}\left(\frac{1}{qn}\sum_{j=1}^{q}\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i'\neq i}}^{n} Y_{i,j}Y_{i',j}\right) = 0.$$

By Bienaymé-Tchebychev inequality, we have, for $\eta > 0$

$$\mathbb{P}\left(\mathrm{I}_2 > \frac{\eta}{2}\right) \quad \leq \quad \frac{4}{\eta^2}\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^{q}\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i'\neq i}}^{n} Y_{i,j}Y_{i',j}\right)^2\right]. \tag{19}$$

Furthermore, since $\frac{1}{n} = \frac{(n-1)^2}{4}\left(\frac{2}{n(n-1)}\right)^2$, we can express the expectation above as the expectation of a U-statistic

$$\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^{q}\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i'\neq i}}^{n} Y_{i,j}Y_{i',j}\right)^2\right] = \frac{(n-1)^2}{4}\mathbb{E}\left[\left(\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i'\neq i}}^{n}\frac{1}{q}\sum_{j=1}^{q} Y_{i,j}Y_{i',j}\right)^2\right].$$

More precisely, this is a U-statistic of degree 2 with kernel $w(Y_i, Y_{i'}) = \frac{1}{q}\sum_{j=1}^{q} Y_{i,j}Y_{i',j}$, with $\mathbb{E}[w(Y_i, Y_{i'})] = 0$ and degenerated gradients

$$\mathbb{E}[\,w(Y_i, Y_{i'})\,|\,Y_i] = 0 \quad \text{and} \quad \mathbb{E}[\,w(Y_i, Y_{i'})\,|\,Y_{i'}] = 0,$$

Using the expression of the variance of this U-statistic as given in Lee (2019) [13], it follows that

$$\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^{q}\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i'\neq i}}^{n} Y_{i,j}Y_{i',j}\right)^2\right] = \frac{(n-1)^2}{4}\frac{1}{\frac{n(n-1)}{2}}\binom{n-2}{0}\mathrm{Var}\left(w(Y_i, Y_{i'})\right) \quad = \quad \frac{n-1}{2n}\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^{q} Y_{1,j}Y_{2,j}\right)^2\right]. \tag{20}$$

Now, we have by independence

$$\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^{q} Y_{1,j}Y_{2,j}\right)^2\right] \quad = \quad \mathbb{E}\left[\frac{1}{q^2}\sum_{j=1}^{q}\sum_{k=1}^{q} Y_{1,j}Y_{2,j}Y_{1,k}Y_{2,k}\right] = \frac{1}{q^2}\sum_{j=1}^{q}\sum_{k=1}^{q}\left[\mathbb{E}\left(Y_{1,j}Y_{1,k}\right)\right]^2.$$

Recall that $\mathbb{E}\left[Y_{1,j}Y_{1,k}\right] = 0$ if $j \neq k$. By using Hölder inequalities repetitively and by hypothesis (A_2), we have $\frac{1}{q}\sum_{j=1}^{q}\left[\mathbb{E}\left(Y_{1,j}^2\right)\right]^2 \leq \left(\frac{1}{q}\sum_{j=1}^{q}\mathbb{E}\left(Y_{1,j}^8\right)\right)^{\frac{1}{2}} \leq K_2^{\frac{1}{2}}$, yielding

$$\mathbb{E}\left[\left(\frac{1}{q}\sum_{j=1}^{q} Y_{1,j}Y_{2,j}\right)^2\right] \leq \frac{1}{q}\sqrt{K_2}. \tag{21}$$

Finally, combining equations (19,20) and (21), we obtain a control for I_2 as follows

$$\mathbb{P}\left(\mathrm{I}_2 > \frac{\eta}{2}\right) \quad = \quad \mathbb{P}\left(\frac{1}{qn}\sum_{j=1}^{q}\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i'\neq i}}^{n} Y_{i,j}Y_{i',j} > \frac{\eta}{2}\right) \leq \frac{1}{\eta^2}\frac{2(n-1)}{qn}\sqrt{K_2}.$$

Finally, assumption (A_1) implies

$$\begin{aligned}\mathbb{P}\left(|\Delta_n| > \epsilon n\right) \quad &= \quad \mathbb{P}\left(q\left|\frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*}\right|\left\|n^{1/2}\bar{Y}_n\right\|^2 > \epsilon n\right) \leq \mathbb{P}\left(\left\|n^{\frac{1}{2}}\bar{Y}_n\right\|^2\left|\frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*}\right| > \frac{\epsilon}{K_1}\right)\\ &\leq \quad \mathbb{P}\left(\left(\left\|n^{\frac{1}{2}}\bar{Y}_n\right\|^2 - \sigma^2\right)\left|\frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*}\right| > \frac{\epsilon}{2K_1}\right) + \mathbb{P}\left(\left|\frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*}\right| > \frac{\epsilon}{2\sigma^2 K_1}\right).\end{aligned}$$

Using the fact that $\mathbb{P}(AB > \epsilon) \leq \mathbb{P}(A > \sqrt{\epsilon}) + \mathbb{P}(B > \sqrt{\epsilon})$, and the definition of the function g_n in lemma 7, we have

$$\begin{aligned}\mathbb{P}\left(|\Delta_n| > \epsilon n\right) \quad &\leq \quad \mathbb{P}\left(\left\|n^{\frac{1}{2}}\bar{Y}_n\right\|^2 - \sigma^2 > \sqrt{\frac{\epsilon}{2K_1}}\right) + g_n\left(\sqrt{\frac{\epsilon}{2K_1}}\right) + g_n\left(\frac{\epsilon}{2\sigma^2 K_1}\right)\\ &\leq \quad \mathbb{P}\left(\mathrm{I}_1 - \sigma^2 > \frac{1}{2}\sqrt{\frac{\epsilon}{2K_1}}\right)) + \mathbb{P}\left(\mathrm{I}_2 > \frac{1}{2}\sqrt{\frac{\epsilon}{2K_1}}\right) + g_n\left(\sqrt{\frac{\epsilon}{2K_1}}\right) + g_n\left(\frac{\epsilon}{2\sigma^2 K_1}\right).\end{aligned}$$

Therefore, by inequalities (18) and (22), considering $\eta = \sqrt{\frac{\epsilon}{2K_1}}$, we get

$$\begin{aligned}\mathbb{P}\left(|\Delta_n| > \epsilon n\right) \quad &\leq \quad \frac{4}{\left(\sqrt{\frac{\epsilon}{2K_1}}\right)^2} \times \left[\frac{\sqrt{K_2}}{n}\left(\frac{1}{q} + 1\right) + \frac{1}{2}\frac{n-1}{n}\frac{\sqrt{K_2}}{q}\right] + g_n\left(\sqrt{\frac{\epsilon}{2K_1}}\right) + g_n\left(\frac{\epsilon}{2\sigma^2 K_1}\right)\\ &\leq \quad \frac{4K_1\sqrt{K_2}}{\epsilon n}\left(2 + \frac{1}{q} + K_1\right) + g_n\left(\sqrt{\frac{\epsilon}{2K_1}}\right) + g_n\left(\frac{\epsilon}{2\sigma^2 K_1}\right).\end{aligned}$$

We now complete the proof of the theorem by handling the term (II). By lemma 7, we get

$$\mathbb{P}\left(|\hat{a}_n - a^*| > \epsilon\right) = \mathbb{P}\left(\left|\frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*}\right| > \frac{\epsilon}{K_3}\right) = g_n\left(\frac{\epsilon}{K_3}\right). \tag{22}$$

With inequalities (15), (15), (22), and (22), and using the expression of G to bound g_n given in lemma 7, we obtain

$$\begin{aligned}\mathbb{P}\left(n\bar{Z}_n'\hat{\Sigma}_n^{*-2}\bar{Z}_n \geq u\left(1+\hat{a}_n^*+2\epsilon\right)\right) &\leq \mathbb{P}\left(n\bar{Z}_n'\Sigma_n^{*-2}\bar{Z}_n \geq u\left(1+a^*\right)\right) + \frac{4K_1\sqrt{K_2}}{\epsilon n}\left(2+\frac{1}{q}+K_1\right) \\ &+ g_n\left(\sqrt{\frac{\epsilon}{2K_1}}\right) + g_n\left(\frac{\epsilon}{2\sigma^2 K_1}\right) + g_n\left(\frac{\epsilon}{K_3}\right) \\ &\leq \frac{2e^3}{9}\left(\frac{u-n}{2}\right)^{\frac{n}{2}}\frac{e^{-\frac{u-n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)} + \frac{1}{n}\frac{C\left(\epsilon\right)}{\epsilon},\end{aligned}$$

where $C(\epsilon)$ is independent of n such that

$$C\left(\epsilon\right) = 4K_1\sqrt{K_2}\left(2+\frac{1}{q}+K_1\right) + 2K_1 G\left(\sqrt{\frac{\epsilon}{2K_1}}\right) + \frac{4K_1^2\sigma^4}{\epsilon}G\left(\frac{\epsilon}{2\sigma^2K_1}\right) + \frac{K_3^2}{\epsilon}G\left(\frac{\epsilon}{K_3}\right).$$

□

References

[1] M. Abramovitch, I. Stegun, Handbook of mathematical tables, National Bureau of Standards, Washington, DC (1970).
[2] P. Bertail, E. Gautherat, H. Harari-Kermadec, Exponential bounds for multivariate self-normalized sums, Electronic Communications in Probability 13 (2008) 628–640.
[3] M. Carrasco, R. Kotchoni, Regularized generalized empirical likelihood estimators, Preprint (2017).
[4] G. Chasapis, R. Liu, T. Tkocz, Rademacher–gaussian tail comparison for complex coefficients and related problems, Proceedings of the American Mathematical Society 150 (2022) 1339–1349.
[5] L. Chen, D. Paul, R. L. Prentice, P. Wang, A regularized hotelling's t 2 test for pathway analysis in proteomic studies, Journal of the American Statistical Association 106 (2011) 1345–1360.
[6] P. Doukhan, J. R. León, Cumulants for stationary mixing random sequences and applications to empirical spectral density, Probab. Math. Stat 10 (1989) 11–26.
[7] W. Hoeffding, Probability inequalities for sums of bounded random variables, in: The collected works of Wassily Hoeffding, Springer, 1994, pp. 409–426.
[8] I. M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, The Annals of statistics 29 (2001) 295–327.
[9] S. Lahiri, S. Mukhopadhyay, A penalized empirical likelihood method in high dimensions, The Annals of Statistics 40 (2012) 2511 – 2540.
[10] B. Laurent, P. Massart, Adaptive estimation of a quadratic functional by model selection, Annals of Statistics 28 (2000) 1302–1338.
[11] O. Ledoit, M. Wolf, A well conditioned estimator for large dimensional covariance matrices, Journal of Multivariate Analysis 88 (2000) 365–411.
[12] O. Ledoit, M. Wolf, Quadratic shrinkage for large covariance matrices, Bernoulli 28(3) (2022) 1519–1547.
[13] A. Lee, U-statistics: Theory and Practice, Routledge, 2019.
[14] H. Li, A. Aue, D. Paul, J. Peng, P. Wang, An adaptable generalization of hotelling's t^2 test in high dimension, The Annals of Statistics 48 (2020) 1815–1847.
[15] F. Merlevède, J. Najim, P. Tian, Unbounded largest eigenvalue of large sample covariance matrices: Asymptotics, fluctuations and applications, Linear Algebra and its Applications 577 (2019) 317–359.
[16] W. K. Newey, R. J. Smith, Higher order properties of gmm and generalized empirical likelihood estimators, Econometrica 72 (2017) 219–255.
[17] D. Panchenko, Symmetrization approach to concentration inequalities for empirical processes, Annals of Probability 31 (2003) 2068–2081.
[18] I. Pinelis, Extremal probabilistic problems and hotelling's t2 test under a symmetry condition, The Annals of Statistics 22 (1994) 357–368.
[19] M. Talagrand, The missing factor in hoeffding's inequalities, Annales de l'IHP Probabilités et statistiques 31 (1995) 689–702.
[20] A. N. Tikhonov, On the regularization of ill-posed problems, in: Doklady Akademii Nauk, volume 153, Russian Academy of Sciences, pp. 49–52.
[21] N. D. Tracy, J. C. Young, R. L. Mason, Multivariate control charts for individual observations, Journal of Quality Technology 24 (1992) 88–95.

5. Supplementary material - Simulations

In this supplementary material, we explore graphically for different distributions, how the dependence structure of the observations and the distance to homoscedasticity impact the penalization constants and the tail of the T^2 Hotelling distribution.
We generate Gaussian random variables with a given covariance structure corresponding respectively to the following scenarios:

- **scenario a)** the components $Z_{i,j}$, $j \in \{1, \ldots, q\}$ are independent with variance $\sigma_{j,j}$, that is Z_i are i.i.d $N(0, S^2)$ with $S^2 = diag(\sigma_{j,j})_{1 \le j \le q}$ for $i \in \{1, \ldots, n\}$. The $\sigma_{j,j}$ are themselves generated randomly in a $LN(0, \eta^2)$. We actually expect the variance of the eigenvalues to have a strong influence on the penalized term. The variance η^2 is calibrated for comparison with the dependent case and chosen equal to $log(1 + \sqrt{1 + 4\alpha^2}) - log(2)$ to ensure that the distance between S^2 and $\sigma^2 I_q$ is indeed equal to α^2 (which is chosen the same in the dependent case).
- **scenario b)** the r.v. Z_i's are i.i.d $N(0, S^2)$ with S^2 given by a Toeplitz matrix of the form

$$S^2 = \begin{pmatrix} 1 & s & s^2 & \cdots & s^{q-2} & s^{q-1} \\ s & 1 & s & \ddots & \ddots & s^{q-2} \\ s^2 & s & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & \ddots & s^2 \\ s^{q-2} & \ddots & \ddots & \ddots & 1 & s \\ s^{q-1} & s^{q-2} & \cdots & s^2 & s & 1 \end{pmatrix}.$$

  Up to a constant, this is the covariance matrix of a stationary $AR(1)$ process with auto-regressive parameter s. This parameter s is thus a dependence parameter in $]-1, 1[$ allowing the components of the observations to exhibit more or less dependence.

Notice that in our framework the quantity α^2 is a measure of the complexity of the problem. Actually, if $\alpha^2 = 0$, we can directly use the identity matrix instead of the empirical variance and there is no need for penalizing. For this reason, we are going to compare our simulation results for some given fixed values of α^2 respectively in the dependent and independent cases. For that, we now consider four simulation cases:

(i) scenario a) with α^2 close to 1.10 (note that actually the value of α^2 depends on q but is close to this value in all simulations) corresponding to a standard deviation $\eta = 0.71$;

(ii) scenario b) with the same values of α^2 as in (i) corresponding to a dependence parameter $s = 0.6$;

(iii) scenario a) with α^2 equal respectively to 35.74, 55.63, 67.12, 74.19, 78.83, 82.04, 84.37, 86.13 corresponding to η between 1.89 and 2.11 respectively for the value of $q \in \{50, 100, 150, 200, \ldots, 400\}$;

(iv) scenario b) with the same values of α^2 as in (iii) corresponding to a dependence parameter $s = 0.99$.

For each set of parameters, (i) to (iv), for $n \in \{50, 75, 100, \ldots, 200\}$, we generate n r.v.'s of size $q \in \{50, 100, 150, \ldots, 400\}$ with $q \ge n$. The procedure is repeated $K = 999$ times independently to obtain Monte-Carlo approximations respectively of the distributions of the penalized T_n^2-Hotelling's statistic (with estimated parameters) and the distribution of the penalizing parameter $\hat{\rho}_n^*$.

The graphics in Figure 3 compare the distribution of $\hat{\rho}^*$ for case (i) (independent case, first column) and case (ii) (dependent case, second column of the panel) respectively.
- on the first row: for fixed sample size $n = 50$ and varying $q's$ equal 50, 200 and 400,
- on the second row : for $q = n$ equal to $50, 100, 200$,
- on the last row shows this distribution when $q = 2n$ and n is equal respectively to $50, 100, 200$.

The figures in panel 3 show that the dependence structure tends to lead to smaller penalization constants. By comparing the rows, it seems that there is a proportionality between the penalization parameter $\hat{\rho}^*$ and q/n.

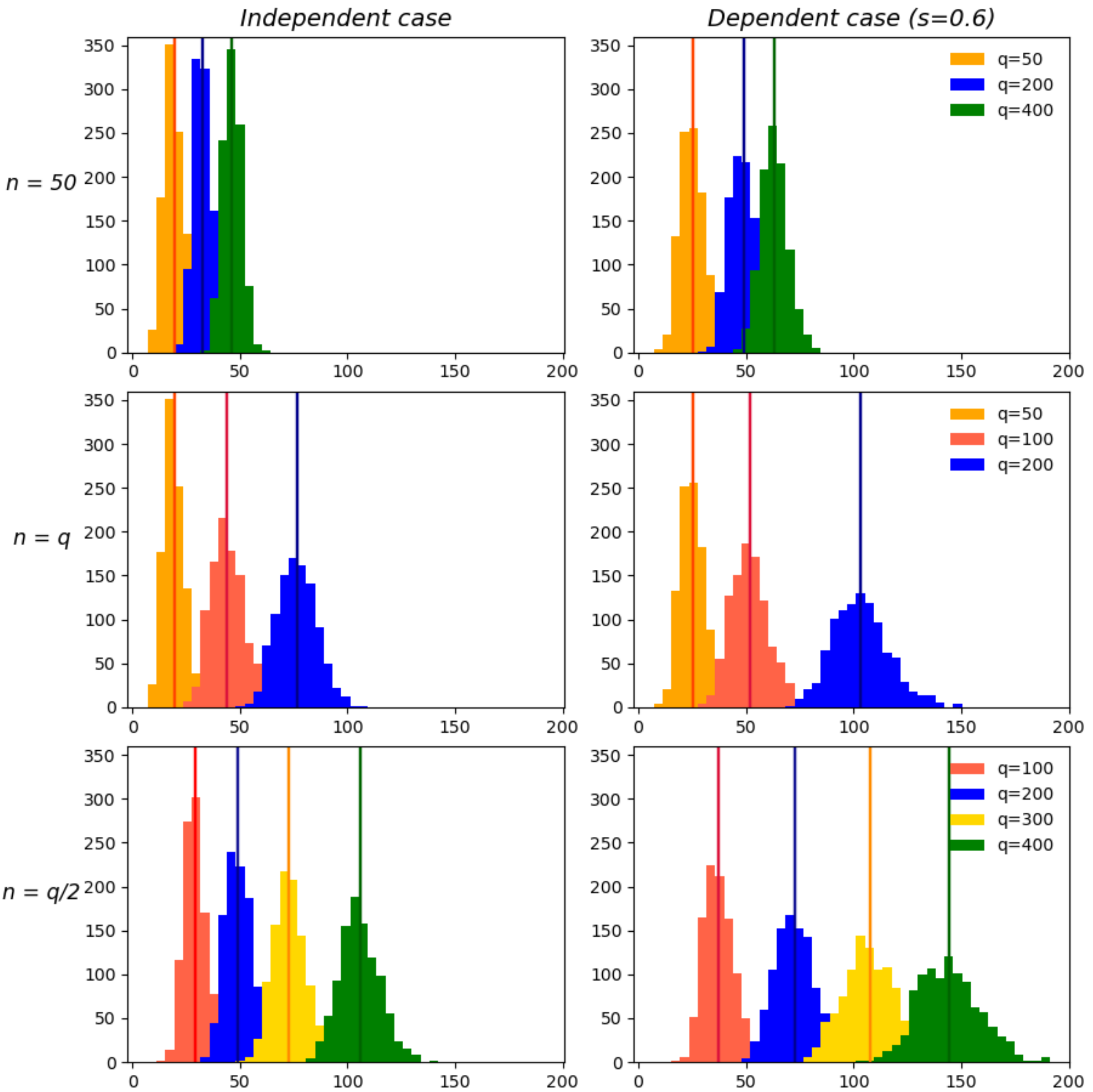

Fig. 3: Distributions of $\hat{\rho}^*$, independent (first column, case (i)), dependent with s=0.6 (second column, case (ii)). Vertical lines represent the empirical mean of the corresponding distribution.

In the independent case, it seems to be of the order $2q/n$ up to some factor probably depending on the variance of the eigenvalues of the matrix. Notice that when $q = n$ the center of the distribution is rather stable but with a smaller variance as n grows. In the dependent case, the "optimal" penalization can decrease drastically even if the value of α is fixed but is even more stable (in mean). This can be explained by the fact that we have

$$\alpha^2 = \|S^2 - \sigma^2 I_q\|^2 = \frac{1}{q}\sum_{k=1}^{q}\left(\sigma_k^2 - \sigma^2\right)^2 + \frac{2}{q}\sum_{k,j<k}^{q}\mathrm{Cov}^2(Z_{1,k}, Z_{1,j}).$$

In the independent case, α^2 is essentially the empirical variance of the eigenvalues. But in the dependent case, the

covariance terms clearly increase which induces a reduction of the penalizing term since $\rho^* = \frac{\beta^2}{\alpha^2}\sigma^2$.
Now, we focus on the distribution of the optimal penalization when there is a strong dependent component. The graphics panel in Figure 4 compares the distribution of $\hat{\rho}^*$ for case (iii) (independent case, first column) and case (iv) (dependent case, second column) respectively on the first row for fixed sample size $n = 50$ and varying $q's$, on the second row for $q = n$ varying in $50, 100, 200$. Finally, the last row shows this distribution when $q = 2n$ and n varies in $50, 100, 200$. Figure 4 compares the distribution of the "optimal" estimated penalty for identical values

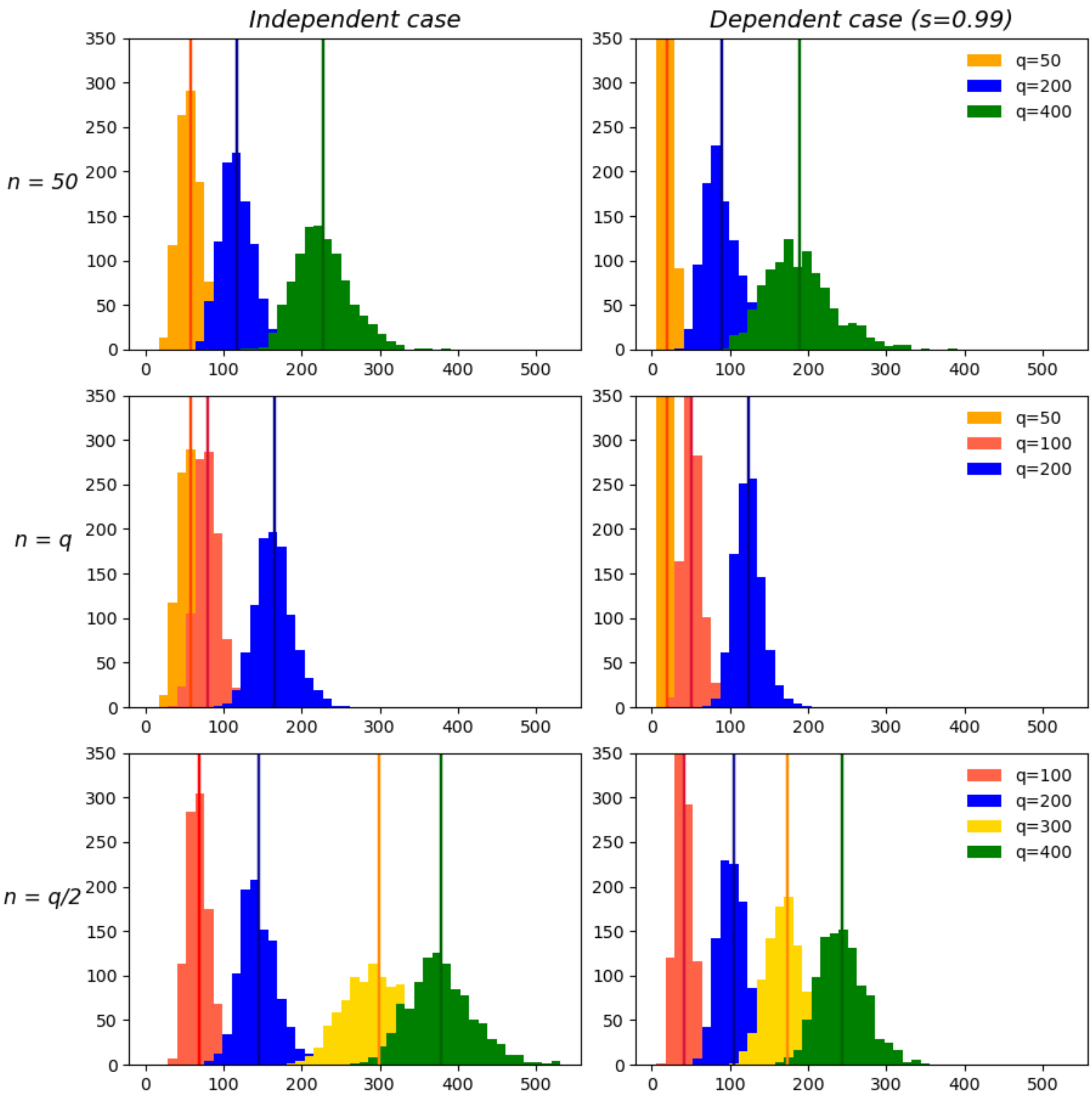

Fig. 4: Distributions of $\hat{\rho}^*$, independent (first column, case (iii)), dependent with s=0.99 (second column, case (iv)) with the same α^2. Vertical lines represent the empirical mean of the corresponding distribution.

of α^2 (depending on q) for the two scenarios, that is, the left (i.i.d.) and the right column (dependent case) and for different values of q. α^2 is equal respectively to $35.74, 55.63, 67.12, 74.19, 78.83, 82.04, 84.37, 86.13$ for the values of

q equal to $50, 100, 150, 200, \ldots, 400$. We see that, even for an identical value of α^2, i.e. for a given distance between the true covariance matrix and the diagonal matrix, the distribution of optimal penalty term is systematically more concentrated around smaller values in the dependent case (second column). This conclusion is true for all values of q and n. In other words, the stronger the dependence, the smaller the optimal penalty term.
Recall that, in Figure 3, we consider a fixed value $\alpha^2 = 1, 10$ for all values of q. The comparison of Figures 3 and 4 shows that when the α^2 term is big, this leads to a smaller penalization term. Furthermore, this penalization becomes smaller when q grows with n. This is quite in contradiction with the practice which suggests using a penalization of the order $2q/n$ as noticed in Figure 3. The distance to the homoscedastic framework has thus a very strong impact on the penalty.
The following Figure 5 and Figure 6 give the histogram of the penalized Hotelling's statistic obtained by $K = 999$ Monte-Carlo simulations, respectively for the independent and dependent case but with the same α^2. We present first the case for $s = 0.6$ (Figure 5) and then the case $s = 0.99$ (Figure 6).

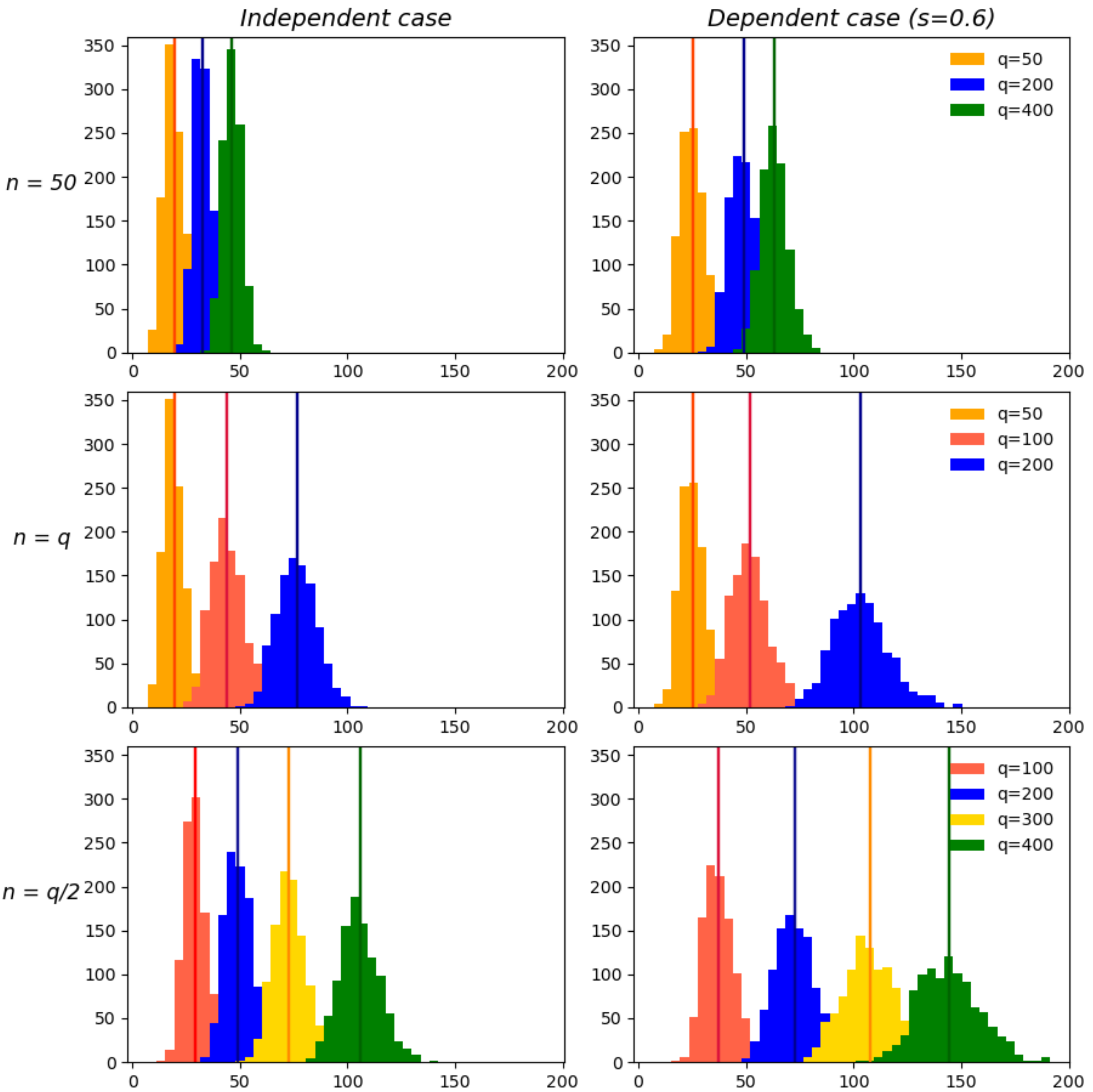

Fig. 5: Distributions of $T_n^2(\widehat{\rho}_n^\star, 1)$, the penalized Hotelling's statistic, in independent (first column, case (i)) and dependent with s=0.6 frameworks (second column, case (ii)).

Compare figures 5 and 6, focusing first on the first column corresponding to the independent case. We see the importance of the value α (the distance to homoscedasticity) in the distribution. Increasing α^2 tends to lead to a smaller penalization and to a less precise approximation of the covariance matrix yielding a shift of the distribution of the Hotelling's statistic on the right. Comparing the two columns (independent and dependent case), we see that the distributions are centered around quite similar values but tend to be more concentrated in the independent case. Increasing the value of α^2 in figure 6 tends to reverse this phenomenon. These figures also emphasize the role of the ratio q/n.

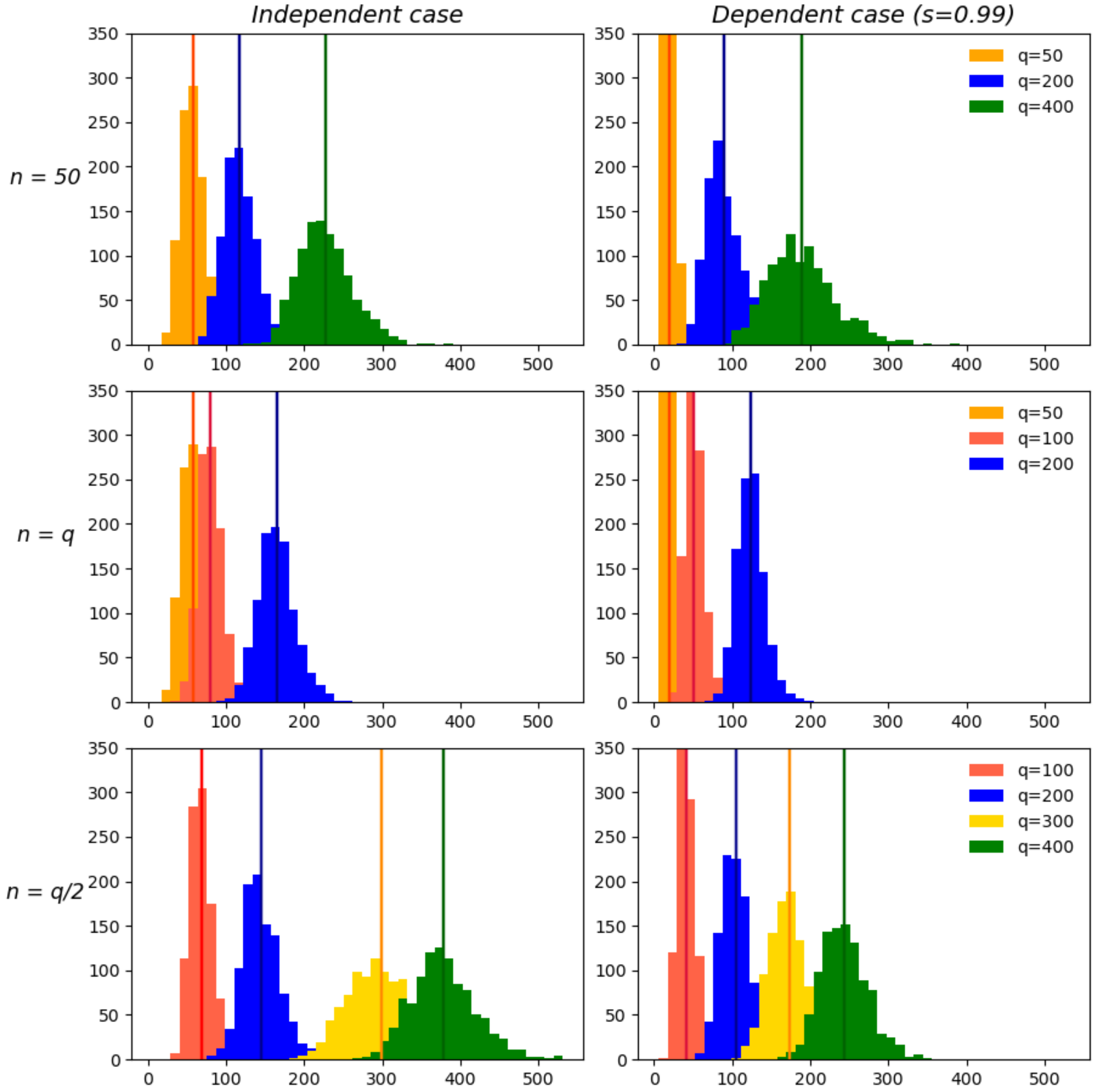

Fig. 6: Distributions of $T_n^2(\widehat{\rho}_n^\star, 1)$, the penalized Hotelling's statistic, in independent (first column, case (iii)) and dependent with s=0.99 frameworks (second column, case (iv))

Figures 7 and 8 show the comparison between the survival function of $T_n^2(\widehat{\rho}_n^*, 1)/(1 + \hat{a}_n^*)$, the penalized Hotelling's statistic reduced by $1 + \hat{a}_n^\star$ compared to the bound obtained in the Theorem 4. These figures show clearly that the bounds we obtained are too conservative. Curiously the bounds seem to be better when the dependence is very strong.

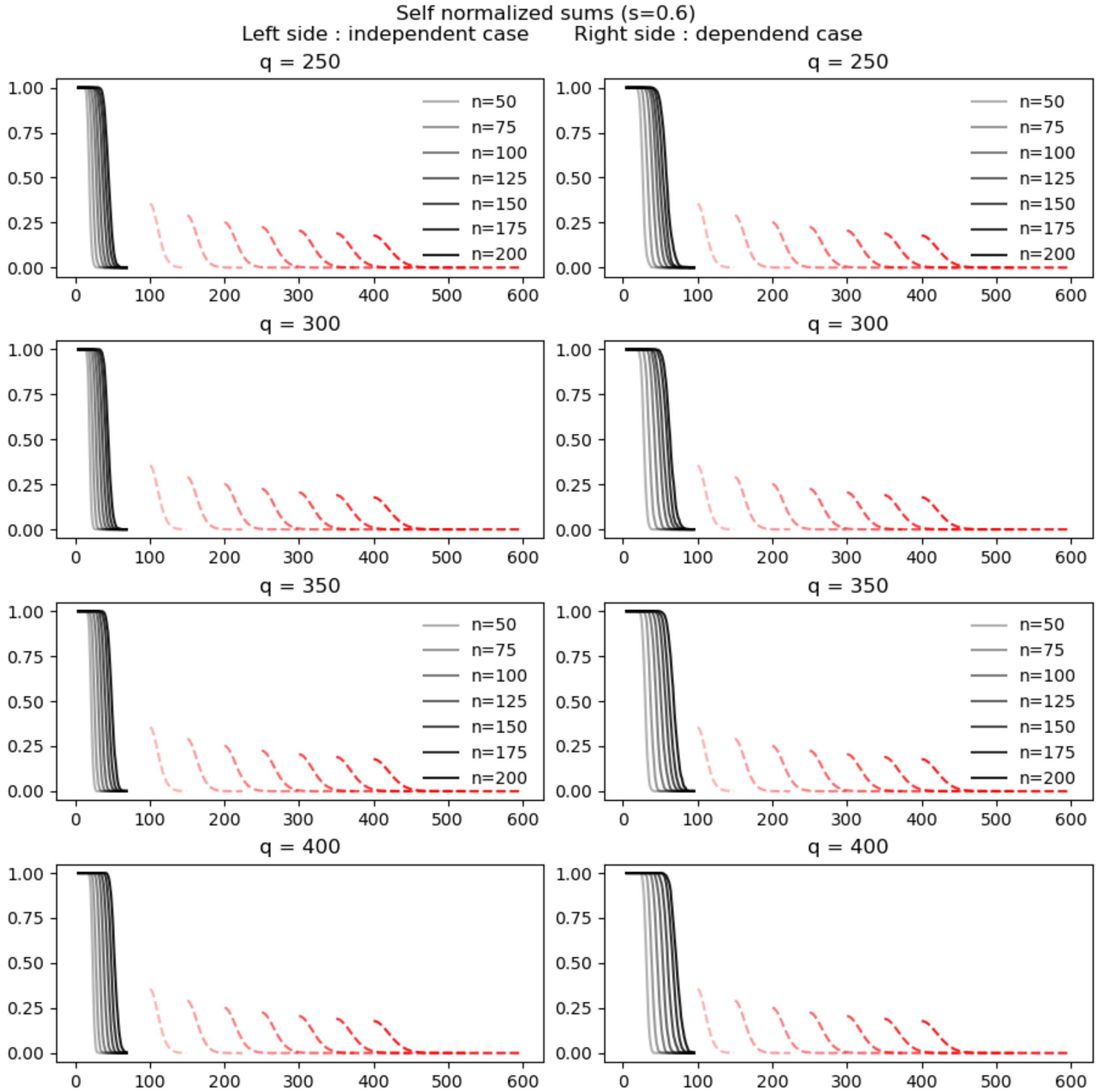

Fig. 7: Comparison of the true tail of the penalized Hotelling's statistic and the tail given by the bound for different values of n, q. $s = 0.6$ in the right column. The red dotted lines refer to the bounds for the ordered corresponding n.

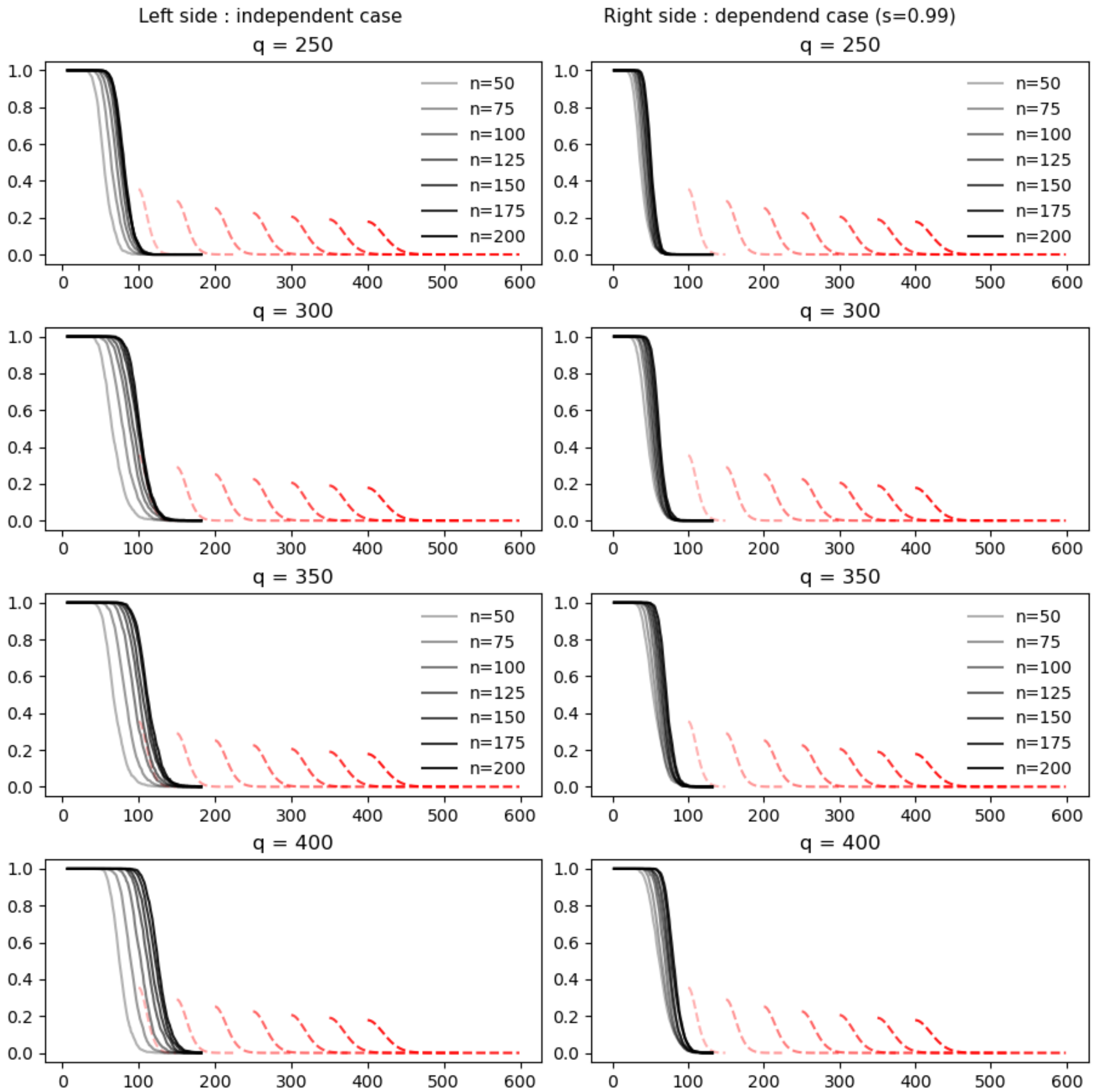

Fig. 8: Comparison of the true tail of the penalized Hotelling's statistic and the tail given by the bound for different values of n, q. $s = 0.99$ in the right column. The red dotted lines refer to the bounds for the ordered corresponding n.

From this simulation study, we conclude that our bounds give some interesting information both on the optimal penalty that one may choose and on the order of the bounds. However, there is still room for improvement.