INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

Confidence Sets for the Sample Average Approximation of Stochastic Discrete Optimization Problems

Mario Martinoli [a]
Raffaello Seri [b]
Samuele Tonati [b]

[a] Institute of Economics and L'EmbeDS, Scuola Superiore Sant'Anna, Pisa, Italy
[b] InsIDE and DiECO, University of Insubria, Italy

2026/09 **March 2026**

ISSN(ONLINE): 2284-0400
DOI: 10.57838/sssa/7vnw-1715

Confidence Sets for the Sample Average Approximation of Stochastic Discrete Optimization Problems

Mario Martinoli[1], Raffaello Seri[2], and Samuele Tonati[3]

[1]Institute of Economics and L'EMbeDS, Sant'Anna School of Advanced Studies, Italy
[2]InsIDE Lab, DiECO, University of Insubria, Italy, raffaello.seri@uninsubria.it (corr. author)
[3]InsIDE Lab, DiECO, University of Insubria, Italy

Abstract

Purpose - We propose a method to build confidence sets for the solutions of stochastic discrete optimization problems solved through the sample average approximation method.

Design/methodology/approach - By combining the concept of Model Confidence Set (MCS) with shrinkage estimation of large covariance matrices, we accommodate sampling mechanisms that allow for arbitrary dependence across alternatives, even when the number of alternatives is larger than the sample size, and deliver confidence sets asymptotically containing the solution set with probability at least $1-\alpha$ for predetermined α.

Findings - We derive bounds for the error induced by replacing the true covariance matrix with an estimator and characterize the impact of this error on the asymptotic distribution of the MCS test statistics. We test the theoretical properties of our set estimator in finite samples through an extensive Monte Carlo experiment involving the computation of the covariance matrix using different shrinkage estimators.

Research limitations/implications - This research is the first to provide generally applicable measures of uncertainty in discrete optimization.

Practical implications - Whenever a stochastic discrete optimization problem is solved using the sample average approximation method, the confidence set should be reported alongside the solution in order to provide a measure of uncertainty.

Originality/Value - The main contribution of the paper is to offer, for the first time, a method for computing confidence sets for the solutions of stochastic discrete optimization problems. We also derive a bound on the accuracy of the asymptotic distribution for a class of test statistics involving covariance matrices estimated with non-standard estimators.

Keywords: Stochastic discrete optimization; Sample average approximation; Confidence set; Shrinkage estimation; Covariance structure

1 Introduction

Many situations arising in decision theory manifest as *discrete optimization* problems, which consist in obtaining the optimal decision(s) among a finite set of choices. A particularly interesting case is the one of *stochastic discrete optimization*, appearing when the objective function can be expressed as the average value of a random variable. Given a discrete set $\mathcal{M}^0$ of possible decisions, we aim at identifying the set of alternatives, called $\mathcal{M}^\star$, minimizing

$$\mu_i = \mathbb{E}X\left(i,\omega\right),\ i \in \mathcal{M}^0,$$

where $X\left(i,\omega\right)$ is a random variable—or a transformation of it. Whenever the computation of the integral appearing in the objective functions is daunting, a way to approximate the optimal alternative(s) is the *sample average approximation* (SAA) method (Kleywegt *et al.*, 2002), in which, for any i, the expected value is replaced by the empirical mean $\widehat{\mu}_i := \frac{1}{n}\sum_{j=1}^{n} X_j\left(i,\omega\right)$ based on a sample $\left\{X_1\left(i,\omega\right), X_2\left(i,\omega\right), \dots\right\}$ of independent realizations of $X\left(i,\omega\right)$. The random variables for different values of i are generally correlated with a covariance matrix $\mathbf{\Sigma}$. This method of approximation fits into the framework of statistical estimation for discrete parameter models covered in Choirat and Seri (2012), but this does not help in providing a method for the quantification of the uncertainty in the identification of the optimal solutions. The objective of this paper is to provide a method to compute a confidence set around the set of optima $\mathcal{M}^\star$ of the SAA of a stochastic discrete optimization.

The first attempts to provide confidence sets for discrete optimization problems can be attributed to Futschik and Pflug (1995) and Futschik and Pflug (1997) and are based on the subset selection procedure of Gupta (1965). The authors propose a two-step procedure that exploits independent portions of the data to sequentially exclude suboptimal alternatives. Specifically, in the first step they define an inclusion criterion to build a level-α_1 confidence set $\mathcal{C}_1$. In the second stage, they build a level-α_2 confidence set $\mathcal{C}_2$ using a different inclusion criterion. If $(1-\alpha_1)(1-\alpha_2) = 1-\alpha$, the final level-$\alpha$ confidence region is obtained as $\mathcal{C}_1 \cap \mathcal{C}_2$. Although the method of Futschik and Pflug (1995) is appealing, it has two drawbacks. First, it assumes that the sample means are computed using independent samples for each alternative, a condition that is clearly violated in the situation usually corresponding to the SAA, that we deem common sampling (see Section 4.5). Second, it provides a *level-α confidence set*, but not a strong one, which does not ensure that all optimal alternatives are covered with a predefined probability. Rather, it only guarantees that each member of the solution set $\mathcal{M}^\star$ belongs to the confidence region $\mathcal{C}$ with probability $1-\alpha$ (see Section 3.1 for a formal definition).

In this paper, we offer solutions to both these issues. In particular, we allow for potentially full covariance structures among the estimators $\widehat{\mu}_i$ for $i \in \mathcal{M}^0$, which show up when the random variables are the same (or are correlated) across the different alternatives as in the SAA approach (see, e.g., Kleywegt *et al.*, 2002), and we deliver *strong level-α confidence sets*. These two issues, namely the construction of strong confidence sets and the extension to more general dependence structures among the simulated alternative values, are addressed using two recent contributions from the econometric and statistical literature.

First, the confidence regions for the best alternative(s) are computed exploiting the so-called Model Confidence Set (MCS, Hansen *et al.*, 2011; Seri *et al.*, 2021). The MCS allows us to define confidence sets for the whole solution set $\mathcal{M}^\star \subseteq \mathcal{M}^0$. In this way, we can identify all the optimal alternatives that are statistically indistinguishable. The case of independent sampling, in which the sample means $\widehat{\mu}_i$ are independent across i, has been treated in Seri *et al.*, (2021) and Martinoli *et al.*, (2024), who propose to use the MCS to calibrate and validate the parameters of agent-based models (ABM). That situation can be seen as a special case of the present work where the covariance matrix $\mathbf{\Sigma}$ is diagonal. The MCS can also be constructed from t-statistics as in the case of Hansen *et al.*, (2011, Sec. 3.1.2) or Barde (2025), so as to bypass the estimation of $\mathbf{\Sigma}$. We leave this extension to future research.

Second, to build the confidence sets, one must know the structure of the covariance matrix $\mathbf{\Sigma}$ of the random variables sampled across alternatives. This covariance matrix is usually computed using the classical estimator $\widehat{\mathbf{\Sigma}}_0$ of the sample covariance. However, when the number of alternative m_0 is at least as large as the number of simulations n, $\widehat{\mathbf{\Sigma}}_0$ turns out to be singular, and even when n and m_0 are comparable $\widehat{\mathbf{\Sigma}}_0$ is ill conditioned. In that case, we suggest to compute $\widehat{\mathbf{\Sigma}}_0$ using a shrinkage estimator (Ledoit and Wolf, 2003; Ledoit and Wolf, 2004a; Ledoit and Wolf, 2004b; Ledoit and Wolf, 2020).

From a theoretical point of view, we provide two main innovations that may be of independent interest. First, we characterize the error of replacing the true matrix $\mathbf{\Sigma}$ with an estimator $\widehat{\mathbf{\Sigma}}$. Second, we investigate the impact of this error on the asymptotic distribution of the MCS test statistics, for both the independent- and common-sampling case.

We test the small-sample properties of our approach through an extensive simulation experiment. The Monte Carlo experiment is performed under different parametrizations and the regularized covariance matrix is computed using different estimators (Ledoit and Wolf, 2004a; Ledoit and Wolf, 2003; Ledoit and Wolf, 2004b; Ledoit and Wolf, 2020).

The rest of the paper is organized as follows. In Section 2, we introduce some notation. In Section 3, we describe our statistical framework. The main results of the paper are gathered in Section 4. In Section 5, we provide a Monte Carlo experiment to test the finite-sample properties of our estimator. Section 6 concludes. The proofs of the theoretical results are deferred to Section 7.

2 Notation

Capital bold letters, such as $\mathbf{A}$, denote matrices while lowercase bold letters, such as $\mathbf{a}$, usually denote vectors. The i-th element of vector $\mathbf{a}$ is generally denoted a_i. $\mathbf{u}_n$ is a n-vector composed of ones. $\mathbf{I}_n$ is the $(n \times n)$-identity matrix. $\mathbf{U}_n$ is a $(n \times n)$-matrix composed of ones. $\mathbf{e}_{i,n}$ is a n-vector of zeros with a one in the i-th position; when the length is clear from the context we simply use $\mathbf{e}_i$. $\mathbf{0}_{m\times n}$ is a $(m \times n)$-matrix composed of zeros. We do not indicate the dimensions when they are clear from the context. $\operatorname{diag}(\mathbf{a})$ is a diagonal matrix with $\mathbf{a}$ on its diagonal. $\mathbf{A}'$ and $\mathbf{A}^{-1}$ are respectively the transpose and the inverse of the matrix $\mathbf{A}$, provided they exist. $\operatorname{rank}(\mathbf{A})$ is the rank of the matrix $\mathbf{A}$, $\operatorname{tr}(\mathbf{A})$ its trace, $\lambda_i(\mathbf{A})$ its i-th eigenvalue, $\lambda_{\max}(\mathbf{A})$ its largest eigenvalue, $\lambda_{\min}(\mathbf{A})$ its smallest eigenvalue, $\sigma_i(\mathbf{A})$ its i-th singular value. For a positive semidefinite matrix $\mathbf{A}$, $\mathbf{A}^{\frac{1}{2}}$ is its square root, i.e. the positive semidefinite matrix such that $\mathbf{A}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}} = \left(\mathbf{A}^{\frac{1}{2}}\right)'\mathbf{A}^{\frac{1}{2}} = \mathbf{A}$, and $\mathbf{A}^{-\frac{1}{2}}$ is its inverse. The element of $\mathbf{A}$ in position (i,j) is denoted as $\mathbf{A}_{ij}$ or $[\mathbf{A}]_{ij}$; the matrix with generic element a_{ij} is denoted $[a_{ij}]$. For a $(m \times n)$-matrix $\mathbf{A}$, $\|\mathbf{A}\| := \sqrt{\lambda_{\max}(\mathbf{A}'\mathbf{A})}$ denotes the spectral norm, $\|\mathbf{A}\|_F := \sqrt{\operatorname{tr}(\mathbf{A}'\mathbf{A})}$ the Schatten 2-norm or Frobenius norm, $\|\mathbf{A}\|_* := \sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{A})$ the Schatten 1-norm or nuclear norm. A Gaussian random variable with mean μ and variance σ^2 is denoted as $\mathcal{N}(\mu,\sigma^2)$, while $\mathcal{E}(\lambda)$ indicates an exponential random variable with parameter λ.

3 The Context

3.1 *Definitions*

We consider the problem

$$\begin{cases} \min_{i\in\mathcal{M}^0} \mu_i = \mathbb{E}X(i,\omega), \\ \sharp(\mathcal{M}^0) = m_0 < \infty. \end{cases}$$

In the following, we define the *solution set* $\mathcal{M}^\star := \left\{j \in \mathcal{M}^0 : j = \arg\min_{i\in\mathcal{M}^0} \mu_i\right\}$.

We define three types of confidence sets $\mathcal{C}$. A *weak level-α confidence set* is such that

$$\mathbb{P}\{\mathcal{C} \cap \mathcal{M}^\star \neq \emptyset\} \geq 1-\alpha;$$

a *level-α confidence set* is such that

$$\mathbb{P}\{j \in \mathcal{C}\} \geq 1-\alpha, \text{ for all } j \in \mathcal{M}^\star;$$

a *strong level-α confidence set* is such that

$$\mathbb{P}\{\mathcal{M}^\star \subseteq \mathcal{C}\} \geq 1-\alpha.$$

The strong definition is the one used in a different set-valued setting in Seri and Choirat (2004), Jankowski and Stanberry (2012), and Choirat and Seri (2014, Definition 4).

3.2 *Estimation*

We write $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_{m_0})'$ and we consider the estimators $\widehat{\mu}_i := \frac{1}{n}\sum_{j=1}^n X_j(i,\omega)$ and $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1,\ldots,\widehat{\mu}_m)'$. Then, $\boldsymbol{\Sigma} = \mathbb{V}(\widehat{\boldsymbol{\mu}})$ and let $\widehat{\boldsymbol{\Sigma}}$ be an estimator of the covariance matrix $\boldsymbol{\Sigma}$ (see below for some alternatives). Two extreme situations are of interest. In between these two extreme cases, several other possibilities arise that we will not consider here.

First, the random variables are sampled separately and independently across the values $i \in \mathcal{M}^0$:

$$X_j(i,\omega) = \xi_{ij}(\omega).$$

This situation is the one observed, e.g., by Seri *et al.*, (2021) in agent-based models, in which resampling the same variables $\xi_j(\omega)$ for different values of i is impossible. In this case, the matrix $\boldsymbol{\Sigma}$ is diagonal and the most natural estimator $\widehat{\boldsymbol{\Sigma}}$, obtained populating the diagonal with $\widehat{\sigma}_i^2 := \frac{1}{n}\sum_{j=1}^n X_j^2(i,\omega) - \widehat{\mu}_i^2$ for $i \in \mathcal{M}^0$ and setting the extra-diagonal elements to 0, is perfectly feasible.

Second, the random variables are the same for all values $i \in \mathcal{M}^0$, but are modified by functions F_i:

$$X_j(i,\omega) = F_i(\xi_j(\omega)),$$

where $\xi_j(\omega)$ is allowed to be a vector. In this case, the matrix $\boldsymbol{\Sigma}$ is full, and any estimator $\widehat{\boldsymbol{\Sigma}}$ should contain also the covariances between the values of $\widehat{\mu}_i$. This situation arises, e.g., when the random variables come from a real-world problem, as in the case in M-estimation in discrete parameter problems (Choirat and Seri, 2012), but it is also the basis of the SAA method as covered in Kleywegt *et al.*, (2002) (see (Hess and Seri, 2019), for an approach covering both statistical estimation and SAA in the case of continuous optimization). The most natural approach is to estimate $\boldsymbol{\Sigma}$, if necessary for the method under scrutiny, using the classical estimator of the sample covariance $\widehat{\boldsymbol{\Sigma}}_0 = [\widehat{\sigma}_{ik}]$, where $\widehat{\sigma}_{ik} := \frac{1}{n}\sum_{j=1}^{n} X_j(i,\omega) X_j(k,\omega) - \widehat{\mu}_i\widehat{\mu}_k$ (note that $\widehat{\sigma}_i^2 = \widehat{\sigma}_{ii}$). The problem of this approach is that the matrix $\widehat{\boldsymbol{\Sigma}}_0$ is singular whenever $n \le m_0$, a situation that may arise when m_0 is very large. Two approaches can be used to solve this problem:

- we could estimate $\boldsymbol{\Sigma}$ using some estimators $\widehat{\boldsymbol{\Sigma}}$ for large covariance matrices, such as the shrinkage ones proposed in Ledoit and Wolf (2003), Ledoit and Wolf (2004a), Ledoit and Wolf (2004b), and Ledoit and Wolf (2020);
- we could resort to ad hoc rules that do not use the estimator $\widehat{\boldsymbol{\Sigma}}_0$ (just for future reference, we could use the elimination rules based on t-tests in (Hansen *et al.*, 2011; Barde, 2025)).

4 The Construction of Confidence Sets

4.1 The Theory

In the following, $\mathcal{M}$ denotes a generic set of discrete parameters, that may coincide or not with $\mathcal{M}^0$. The procedure outlined in Hansen *et al.*, (2011) is based on an equivalence test $\delta_{\mathcal{M}}$ and a selection rule $e_{\mathcal{M}}$, that are associated to the set $\mathcal{M}$.

The test $\delta_{\mathcal{M}}$ is used to test the null hypothesis

$$\mathsf{H}_{0,\mathcal{M}} : \mu_i = \mu_j, \forall i, j \in \mathcal{M}.$$

Note that $\mathsf{H}_{0,\mathcal{M}^\star}$ is true while $\mathsf{H}_{0,\mathcal{M}}$ is false whenever $\mathcal{M} \neq \mathcal{M}^\star$. The alternative hypothesis is

$$\mathsf{H}_{A,\mathcal{M}} : \exists i, j \in \mathcal{M} \text{ such that } \mu_i \neq \mu_j.$$

When the test rejects the null hypothesis, we set $\delta_{\mathcal{M}} = 1$, while, when it does not reject it, we set $\delta_{\mathcal{M}} = 0$.

The elimination rule $e_{\mathcal{M}}$ is used to eliminate an element from $\mathcal{M}$ when $\delta_{\mathcal{M}} = 1$. We suppose that $e_{\mathcal{M}}$ takes its values in $\mathcal{M}$, so that the elimination rule forces us to pass from $\mathcal{M}$ to $\mathcal{M} \backslash e_{\mathcal{M}}$.

We start from the set $\mathcal{M} = \mathcal{M}^0 := \{1, \ldots, m_0\}$ and perform the test $\delta_{\mathcal{M}} = \delta_{\mathcal{M}^0}$. If the test leads to a rejection, we perform an elimination step $e_{\mathcal{M}} = e_{\mathcal{M}^0}$ to get a new set $\mathcal{M}_1$. We repeat the process until the test $\delta_{\mathcal{M}}$ does not reject the null hypothesis. We denote with $\widehat{\mathcal{M}}^\star$ the final set of models. When all the tests are performed at the same significance level α, we can explicitly write $\widehat{\mathcal{M}}^\star = \widehat{\mathcal{M}}^\star_{1-\alpha}$.

For any $\mathcal{M} \subset \mathcal{M}^0$, we need the following assumptions:

B1 $\limsup_{n\to\infty} \mathbb{P}\{\delta_{\mathcal{M}} = 1 \,|\mathsf{H}_{0,\mathcal{M}}\} \le \alpha$.

B2 $\lim_{n\to\infty} \mathbb{P}\{\delta_{\mathcal{M}} = 1 \,|\mathsf{H}_{A,\mathcal{M}}\} = 1$.

B3 $\lim_{n\to\infty} \mathbb{P}\{e_{\mathcal{M}} \in \mathcal{M}^\star \,|\mathsf{H}_{A,\mathcal{M}}\} = 0$.

Under these conditions, the following results hold (see Theorem 1 in (Hansen *et al.*, 2011, p. 459) and Theorem 1 below):

- $\lim_{n\to\infty} \mathbb{P}\left\{\mathcal{M}^\star \subseteq \widehat{\mathcal{M}}^\star_{1-\alpha}\right\} \ge 1 - \alpha$,
- $\lim_{n\to\infty} \mathbb{P}\left\{i \in \widehat{\mathcal{M}}^\star_{1-\alpha}\right\} = 0, \quad i \notin \mathcal{M}^\star$.

We can reasonably suppose that the confidence interval obtained in this way is conservative. Indeed, Corollary 1 in Hansen *et al.*, (2011, p. 460) shows that, when $\mathcal{M}^\star$ is a singleton, $\lim_{n\to\infty} \mathbb{P}\left\{\mathcal{M}^\star = \widehat{\mathcal{M}}^\star_{1-\alpha}\right\} = 1$.

4.2 The Implementation

In order to build a MCS, we have to choose a test procedure $\delta_{\mathcal{M}}$ and an elimination procedure $e_{\mathcal{M}}$.

As a test procedure $\delta_{\mathcal{M}}$, we suppose to estimate the mean μ_i through $\widehat{\mu}_i$ and the variance $\sigma_i^2 := \mathbb{V}(\widehat{\mu}_i)$. We will need the following assumption.

A1 For $j = 1, \ldots, n$ the vectors $\mathbf{x}_j := [X_j(i,\omega)]_i$ are independent and identically distributed. The mean $\mathbb{E}X_j(i,\omega)$ exists and is finite for each $j \in \mathcal{M}^0$.

A2 The variances $\sigma_i^2 = \mathbb{V}\left[X_j\left(i,\omega\right)\right]$ are finite for any $i \in \mathcal{M}^0$.

Consider the matrix $\mathbf{A}$ defined by

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & -1 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_{m-1} & -\mathbf{I}_{m-1} \end{bmatrix}.$$

The null hypothesis $\mathsf{H}_{0,\mathcal{M}}$ is that $\mathbf{A}\boldsymbol{\mu} = \mathbf{0}_{1,m-1}$. The test statistic is

$$W_{\mathcal{M}} = n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right).$$

Remark that the test is invariant with respect to the order of the models and, in particular, with respect to which one is chosen as the first one. Moreover, one could choose $\mathbf{A}$ as a first-difference $((m-1)\times m)$-matrix.

As an elimination procedure $e_{\mathcal{M}}$, we choose the index $j \in \mathcal{M}$ with the largest value $\widehat{\mu}_j$, i.e. we identify $e_{\mathcal{M}} := \arg\max_{j\in\mathcal{M}} \widehat{\mu}_j$.

The following result guarantees that the procedure works as desired.

Theorem 1. *(Theorem 1 in (Hansen et al., 2011, p. 459), Theorem 2 in (Seri et al., 2021, p. 71)) Let the test procedure $\delta_{\mathcal{M}}$ be based on the test statistic $W_{\mathcal{M}}$, with asymptotic distribution $W_{\mathcal{M}} \xrightarrow{\mathcal{D}} \chi^2_{m-1}$, and let the elimination procedure $e_{\mathcal{M}}$ be based on the elimination from $\mathcal{M}$ of the index $j \in \mathcal{M}$ with the largest value $\widehat{\mu}_j$. Then, under A1 and A2, we have*

- $\lim_{n\to\infty} \mathbb{P}\left\{\mathcal{M}^\star \subseteq \widehat{\mathcal{M}}^\star_{1-\alpha}\right\} \geq 1-\alpha$,
- $\lim_{n\to\infty} \mathbb{P}\left\{i \in \widehat{\mathcal{M}}^\star_{1-\alpha}\right\} = 0 \quad i \notin \mathcal{M}^\star$.

An alternative way to see the procedure is as follows. We set $\mathcal{M}_1 := \mathcal{M}^0$, and we define a sequence of subsets of $\mathcal{M}^0$ through the elimination rule as

$$\mathcal{M}_{i+1} = \mathcal{M}_i \backslash e_{\mathcal{M}_i} \qquad i = 1,\ldots,m_0-1,$$

or

$$\mathcal{M}_i = \left\{e_{\mathcal{M}_i}, e_{\mathcal{M}_{i+1}}, \ldots, e_{\mathcal{M}_{m_0}}\right\}.$$

In our case, this amounts to ordering the elements $\mathcal{M}^0$ according to the value of $\widehat{\mu}_j$, from the largest to the smallest.

To each element $e_{\mathcal{M}_i}$, we can associate the p-value of the test procedure $\delta_{\mathcal{M}_i}$ to test the null hypothesis $\mathsf{H}_{0,\mathcal{M}_i}$. We call this p-value $p_{\mathsf{H}_{0,\mathcal{M}_i}}$, with the convention that $p_{\mathsf{H}_{0,\mathcal{M}_{m_0}}} \equiv 1$. These p-values are not necessarily decreasing in i. However, it is possible to define an MCS p-value as:

$$\widehat{p}_{e_{\mathcal{M}_j}} := \max_{i\leq j} p_{\mathsf{H}_{0,\mathcal{M}_i}}.$$

The interest of the MCS p-values $\widehat{p}_{e_{\mathcal{M}_j}}$, for $j = 1,\ldots,m_0$, is that $i \in \widehat{\mathcal{M}}^\star_{1-\alpha}$ if and only if $\widehat{p}_i \geq \alpha$. This allows us to compute the MCS over a range of values α. In this case, the MCS can be used to assess the stability of the optimal solution.

4.3 Error in Covariance Matrix Computation

The previous procedure requires the replacement of the true covariance matrix with an estimator. Now, while the classical estimator $\widehat{\boldsymbol{\Sigma}}_0$ is consistent when $n\to\infty$ with fixed m, we will deal with some other estimators $\widehat{\boldsymbol{\Sigma}}$ whose properties are not as well known as those of $\widehat{\boldsymbol{\Sigma}}_0$. The effect of this replacement has to be investigated for a number of reasons. First, our tests use non-standard estimators of the covariance matrix, and the impact this may have is not obvious. Second, the initial tests are carried out on very large choice sets, in which m has a dimension often comparable to n. Although we do not explicitly treat m as an asymptotic parameter in the procedure for constructing the confidence set, assessing the impact of a large m can be relevant.

The next result provides an upper bound on the error of replacing the true matrix $\boldsymbol{\Sigma}$ with an estimator $\widehat{\boldsymbol{\Sigma}}$. The bound is probably not tight but suggestive of the true dependence.

Theorem 2. *The following results hold.*
(i) We have

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(\chi^2_{m-1}\leq t\right)\right|$$
$$\leq R_{m,n}:=\begin{cases}O\left(n^{-\frac{1}{2}}\right), & \text{for } m\geq 2 \text{ if } \mathbb{E}\left(\left(\mathbf{A}\mathbf{X}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\mathbf{X}\right)\right)^{\frac{3}{2}}<\infty,\\ O\left(n^{-1}\right), & \text{for } m\geq 10 \text{ if } \mathbb{E}\left(\left(\mathbf{A}\mathbf{X}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\mathbf{X}\right)\right)^{2}<\infty.\end{cases}$$

(ii) Let λ be a deterministic quantity such that $\left\|\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\|\leq\lambda$ and suppose that $m>2$, $\lambda\leq\frac{1}{2}$. Then,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|$$
$$\leq 3R_{m,n}+\frac{(m-2)^{\frac{m-1}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-5}{2}}\Gamma\left(\frac{m-1}{2}\right)}\lambda\sqrt{-W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right)}$$
$$+\frac{\Gamma\left(\frac{m-1}{2},-\frac{m-2}{2}W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right)\right)}{\Gamma\left(\frac{m-1}{2}\right)}$$
$$\simeq 3R_{m,n}+\frac{(m-2)^{\frac{m-2}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-6}{2}}\Gamma\left(\frac{m-1}{2}\right)}\lambda\left|\ln\lambda\right|^{\frac{1}{2}}.$$

where $\Gamma(a,x)$ is the upper incomplete gamma function and $W_{-1}(x)$ is the lower branch of the Lambert W function.
(iii) For $m>2$, $\mu\leq\frac{1}{2}$ and any $q>0$,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|$$
$$\leq 3R_{m,n}+\frac{(m-2)^{\frac{m-1}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-5}{2}}\Gamma\left(\frac{m-1}{2}\right)}\mu\sqrt{-W_{-1}\left(-\frac{(2\mu)^{\frac{2}{m-2}}}{e}\right)}$$
$$+\frac{\Gamma\left(\frac{m-1}{2},-\frac{m-2}{2}W_{-1}\left(-\frac{(2\mu)^{\frac{2}{m-2}}}{e}\right)\right)}{\Gamma\left(\frac{m-1}{2}\right)}+\frac{\mathbb{E}\left\|\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\|^q}{\mu^q}$$
$$\simeq 3R_{m,n}+\frac{(m-2)^{\frac{m-2}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-6}{2}}\Gamma\left(\frac{m-1}{2}\right)}\mu\left|\ln\mu\right|^{\frac{1}{2}}+\frac{\mathbb{E}\left\|\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\|^q}{\mu^q}.$$

(iv) Provided m is fixed, the previous bounds also apply to

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(\chi^2_{m-1}\leq t\right)\right|$$

with $3R_{m,n}$ replaced by $4R_{m,n}$.

Remark 3. *(i) The discrepancy in the convergence rate $R_{m,n}$ for small and large m is a well-known phenomenon, already present in Esséen (1945) and Jensen (1977). The convergence rates in these papers do not correspond exactly with those in the statement of the theorem. See Götze and Ulyanov (2003) for more details in the case of quadratic forms in averages of lattice-valued random vectors.*

Corollary 4. *In Theorem 2, one can use the inequality*

$$\left\|\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\|\leq 2^{\frac{1}{2}}(m-1)^{\frac{3}{4}}\left(\frac{\sum_{i=1}^m\left|\lambda_i\left(\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\right)\right|}{\lambda_{\min}\left(\widehat{\boldsymbol{\Sigma}}\right)}\right)^{\frac{1}{2}}.$$

4.4 The Case of Independent Sampling

In the case of independent sampling for each i, we have

$$\mathbf{\Sigma} := \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_m^2 \end{bmatrix}.$$

Let $\mathbf{\Sigma}_{-1}$ be the $((m-1)\times(m-1))$-matrix obtained from $\mathbf{\Sigma}$ removing the first line and column. Let $\widehat{\mathbf{\Sigma}}$ be the estimator with zeros off the diagonal and σ_i^2 replaced by $\widehat{\sigma}_i^2$. Then, we have

$$W_{\mathcal{M}} = n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left[\widehat{\sigma}_1^2\mathbf{U}_{m-1} + \widehat{\mathbf{\Sigma}}_{-1}\right]^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right).$$

The following result provides a convergence rate that supplements the analysis in Seri *et al.*, (2021).

Theorem 5. *If $m \geq 10$, under the conditions of Theorem 2,*

$$\begin{aligned}&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\mathbf{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|\\ &\lesssim \frac{4\sqrt{2}\,(m-1)\sum_{i=1}^{m}\sqrt{\mu_{4i}-\sigma_i^4}}{m\sqrt{\pi n}\,\min_{1\leq i\leq m}\sigma_i^2},\end{aligned}$$

where μ_{4i} is the fourth central moment of the distribution of $X_j(i,\omega)$, provided $\sup_n \mathbb{E}\frac{\left|\sigma_j^2-\widehat{\sigma}_j^2\right|^{1+\varepsilon}}{\min_{1\leq i\leq m}\widehat{\sigma}_i^{2+2\varepsilon}} < \infty$ for any j. *The same bound holds for*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(\chi_{m-1}^2\leq t\right)\right|.$$

Remark 6. *The bound for $m < 10$ can be obtained inspecting the proof of the theorem.*

4.5 The Case of Common Sampling

In the case where the random variables are the same across the different alternatives, one must estimate all the elements of the sample covariance matrix and the rank of the resulting sample covariance matrix is bounded by the minimum of the number of simulations n minus 1. If $m > n-1$, the matrix $\widehat{\mathbf{\Sigma}}$ will therefore be singular. Moreover, when n is larger than m but the orders of magnitude are similar, the sample covariance matrix may be ill-conditioned. As these target matrices are not, in general, sparse, the estimators based on the approaches reviewed in Fan *et al.*, (2016), the rank-based method and the factor-model-based method, are not available. A possible solution is the use of a shrinkage method. In what follows, we consider three different estimators.

First, one could use a linear shrinkage estimator in which the eigenvalues of the sample covariance matrix are transformed by a linear function. The estimator in Ledoit and Wolf (2004a) is obtained as a linear combination of the classical covariance matrix estimator and an identity matrix. The result is clearly a classical regularized matrix with a modified diagonal, but the choice of the regularization parameter is data-driven and based on the minimization of a loss function. The estimator in Ledoit and Wolf (2003) and Ledoit and Wolf (2004b) is given by a linear combination of the classical covariance matrix estimator and the constant correlation matrix.

Second, one could use a nonlinear shrinkage estimator, where the eigenvalues of the sample covariance matrix are transformed elementwise by a nonlinear function. As the whole procedure is likely to be computationally intensive, and other approaches like QuEST and NERCOME are probably too demanding, we propose to use the analytical estimator of Ledoit and Wolf (2020).

In what follows, we use an index 0 to denote the classical sample covariance matrix $\widehat{\mathbf{\Sigma}}_0$ and the derived quantities. We start by identifying the eigendecomposition $\widehat{\mathbf{\Sigma}}_0 = \mathbf{U}\,\mathrm{diag}\left(\boldsymbol{\lambda}_0\right)\,\mathbf{U}'$, where $\mathbf{U}$ is a matrix containing the eigenvectors and $\boldsymbol{\lambda}_0$ a vector containing the eigenvalues. Both can depend on n and m. The eigenvalues $\boldsymbol{\lambda}_0 = (\lambda_{01},\ldots,\lambda_{0m})$ are listed in non-increasing order, with corresponding eigenvectors $\mathbf{U} = [\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_m]$. The regularized estimator is denoted $\widehat{\mathbf{\Sigma}}$.

4.5.1 Ledoit and Wolf (2004a)

We recall that the m-vector $\mathbf{x}_j$ is such that its i-th element is $X_j(i,\omega) = F_i(\xi_j(\omega))$, and we define $\widetilde{\mathbf{x}}_j := \mathbf{x}_j - \widehat{\mu}_i \mathbf{u}$. The estimator of the covariance matrix is

$$\widehat{\mathbf{\Sigma}} = \frac{\widehat{\alpha}^2}{\widehat{\delta}^2}\widehat{\mathbf{\Sigma}}_0 + \frac{\widehat{\beta}^2}{\widehat{\delta}^2}\widehat{\overline{\sigma}}^2\mathbf{I}$$

where

$$\begin{aligned}
\widehat{\overline{\sigma}}^2 &= \frac{1}{m}\operatorname{tr}\left(\widehat{\mathbf{\Sigma}}_0\right),\\
\widehat{\delta}^2 &= \frac{1}{m}\left\|\widehat{\mathbf{\Sigma}}_0 - \widehat{\mu}\mathbf{I}\right\|_F^2,\\
\widehat{\beta}^2 &= \min\left\{\widehat{\delta}^2, \frac{1}{mn^2}\sum_{k=1}^{n}\left\|\widetilde{\mathbf{x}}_k\widetilde{\mathbf{x}}_k' - \widehat{\mathbf{\Sigma}}_0\right\|_F^2\right\},\\
\widehat{\alpha}^2 &= \widehat{\delta}^2 - \widehat{\beta}^2.
\end{aligned}$$

Using the orthogonality of the eigenvectors, we can write $\widehat{\mathbf{\Sigma}} = \mathbf{U}\operatorname{diag}(\boldsymbol{\lambda})\mathbf{U}'$, where

$$\operatorname{diag}(\boldsymbol{\lambda}) = \frac{\widehat{\alpha}^2}{\widehat{\delta}^2}\operatorname{diag}(\boldsymbol{\lambda}_0) + \frac{\widehat{\beta}^2}{\widehat{\delta}^2}\widehat{\overline{\sigma}}^2\mathbf{I}$$

or

$$\lambda_j = \frac{\widehat{\alpha}^2}{\widehat{\delta}^2}\lambda_{0j} + \frac{\widehat{\beta}^2}{\widehat{\delta}^2}\widehat{\overline{\sigma}}^2.$$

The regularized eigenvalues are an affine function of the non-regularized ones, and a linear convex combination of the original eigenvalues and their average value $\widehat{\overline{\sigma}}^2$, that is also the average value of the variances.

The implementation that we use is the one in ScikitLearn (see (Oriol and Miot, 2025), for more details on the differences).

4.5.2 Ledoit and Wolf (2003) and Ledoit and Wolf (2004b)

For each covariance $\widehat{\sigma}_{ij}$, we define the corresponding correlation $\widehat{\rho}_{ij} = \frac{\widehat{\sigma}_{ij}}{\sqrt{\widehat{\sigma}_i^2\widehat{\sigma}_j^2}}$. We introduce the average correlation

$$\widehat{\overline{\rho}} = \frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\widehat{\rho}_{ij},$$

and an equicorrelated matrix $\widehat{\mathbf{\Phi}}$ such that

$$\begin{aligned}
\left[\widehat{\mathbf{\Phi}}\right]_{(i,i)} &:= \widehat{\sigma}_i^2,\\
\left[\widehat{\mathbf{\Phi}}\right]_{(i,j)} &:= \widehat{\overline{\rho}}\sqrt{\widehat{\sigma}_i^2\widehat{\sigma}_j^2}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\widehat{\pi}_{ij} &= \frac{1}{n}\sum_{k=1}^{n}\left\{\left[X_k\left(i,\omega\right)-\widehat{\mu}_i\right]\left[X_k\left(j,\omega\right)-\widehat{\mu}_j\right]-\widehat{\sigma}_{ij}\right\}^2, \\
\widehat{\pi} &= \sum_{i=1}^{m}\sum_{j=1}^{m}\widehat{\pi}_{ij} = \frac{1}{n}\sum_{k=1}^{n}\left\|\widetilde{\mathbf{x}}_k\widetilde{\mathbf{x}}_k' - \widehat{\mathbf{\Sigma}}_0\right\|_F^2, \\
\widehat{\vartheta}_{ii,ij} &= \frac{1}{n}\sum_{k=1}^{n}\left\{\left[X_k\left(i,\omega\right)-\widehat{\mu}_i\right]^2-\widehat{\sigma}_i^2\right\} \\
&\quad \cdot\left\{\left[X_k\left(i,\omega\right)-\widehat{\mu}_i\right]\left[X_k\left(j,\omega\right)-\widehat{\mu}_j\right]-\widehat{\sigma}_{ij}\right\}, \\
\widehat{\vartheta}_{jj,ij} &= \frac{1}{n}\sum_{k=1}^{n}\left\{\left[X_k\left(j,\omega\right)-\widehat{\mu}_j\right]^2-\widehat{\sigma}_j^2\right\} \\
&\quad \cdot\left\{\left[X_k\left(i,\omega\right)-\widehat{\mu}_i\right]\left[X_k\left(j,\omega\right)-\widehat{\mu}_j\right]-\widehat{\sigma}_{ij}\right\}, \\
\widehat{\rho} &= \sum_{i=1}^{m}\widehat{\pi}_{ii} + \sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m}\frac{\widehat{\bar{\rho}}}{2}\left(\sqrt{\frac{\widehat{\sigma}_j^2}{\widehat{\sigma}_i^2}}\widehat{\vartheta}_{ii,ij} + \sqrt{\frac{\widehat{\sigma}_i^2}{\widehat{\sigma}_j^2}}\widehat{\vartheta}_{jj,ij}\right), \\
\widehat{\gamma} &= \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\widehat{\bar{\rho}}\sqrt{\widehat{\sigma}_i^2\widehat{\sigma}_j^2}-\widehat{\sigma}_{ij}\right)^2, \\
\widehat{\kappa} &= \frac{\widehat{\pi}-\widehat{\rho}}{\widehat{\gamma}}, \\
\widehat{\delta} &= \max\left\{0, \min\left\{\frac{\widehat{\kappa}}{n}, 1\right\}\right\}.
\end{aligned}$$

At last, the estimator is

$$\widehat{\mathbf{\Sigma}} = \left(1-\widehat{\delta}\right)\widehat{\mathbf{\Sigma}}_0 + \widehat{\delta}\widehat{\mathbf{\Phi}}.$$

4.5.3 Ledoit and Wolf (2020)

This estimator is similar to the one of Ledoit and Wolf (2004a), but it replaces the affine transformation $\lambda_{0j} \mapsto \lambda_j$ of that method with a nonlinear transformation. The final estimator is $\widehat{\mathbf{\Sigma}} = \mathbf{U}\mathrm{diag}\left(\boldsymbol{\lambda}\right)\mathbf{U}'$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$, and the function $\lambda_{0j} \mapsto \lambda_j$ is explained below.

We consider the high-dimensional limit in which $n, m \to \infty$ and $m/n \to \gamma \in (0,1) \cup (1,\infty)$, with $m \neq n$. With the notation $h_j := n^{-1/3}\lambda_{0j}$ and $[y]^+ := \max\{y, 0\}$, we define two functions

$$\begin{aligned}
a\left(\lambda,\boldsymbol{\lambda}\right) &:= \frac{1}{m}\sum_{j=[m-n]^+ +1}^{m}\left\{-\frac{3\left(\lambda-\lambda_j\right)}{10\pi h_j^2}+\frac{3}{4\sqrt{5}\pi h_j}\right. \\
&\quad \left.\cdot\left[1-\frac{1}{5}\left(\frac{\lambda-\lambda_j}{h_j}\right)^2\right]\log\left|\frac{\sqrt{5}h_j-\lambda+\lambda_j}{\sqrt{5}h_j+\lambda-\lambda_j}\right|\right\}, \\
b\left(\lambda,\boldsymbol{\lambda}\right) &:= \frac{1}{m}\sum_{j=[m-n]^+ +1}^{m}\frac{3}{4\sqrt{5}h_j}\left[1-\frac{1}{5}\left(\frac{\lambda-\lambda_j}{h_j}\right)^2\right]^+.
\end{aligned}$$

The shrunken eigenvalues λ_i are then defined as

$$\lambda_i := \begin{cases} \dfrac{\lambda_{0i}}{\left[\pi\frac{m}{n}\lambda_{0i}b(\lambda_{0i},\boldsymbol{\lambda}_0)\right]^2+\left[1-\frac{m}{n}-\pi\frac{m}{n}\lambda_{0i}a(\lambda_{0i},\boldsymbol{\lambda}_0)\right]^2}, & \lambda_{0i} > 0, \\ \dfrac{1}{\pi\left(\frac{m}{n}-1\right)\frac{m}{n}a(0,\boldsymbol{\lambda}_0)}, & \lambda_{0i} = 0. \end{cases}$$

4.5.4 Implementation of the Procedure

We define

$$W_{\mathcal{M}} = n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right),$$

where $\widehat{\mathbf{\Sigma}}$ is one of the estimators above.

The application of Theorem 2 clashes with the fact that no complete convergence rate theory seems to exist for all the covariance estimators that we use. However, we obtain a generic result on the rate of convergence under the hypothesis that $\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} = O_{\mathbb{P}}\left(n^{-\alpha}m^{\beta}\right)$. Despite m is not an asymptotic parameter, the dependence should be indicative of the degradation of the convergence rate for large m.

Theorem 7. *If* $\mathbb{E}\frac{\left|\lambda_i\left(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right)\right|}{\lambda_{\min}\left(\widehat{\mathbf{\Sigma}}\right)} \leq Cn^{-\alpha}m^{\beta}$,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\mathbf{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|$$
$$\lesssim 3R_{m,n} + 8C\frac{m^{\frac{3}{2}+\beta}}{n^{\alpha}}.,$$
$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(\chi^2_{m-1}\leq t\right)\right|$$
$$\lesssim 4R_{m,n} + 8C\frac{m^{\frac{3}{2}+\beta}}{n^{\alpha}}.$$

Remark 8. *(i) The impact on m on the final result is probably overestimated, as it depends on the use of Corollary 4.*
(ii) The properties of the matrix $\widehat{\mathbf{\Sigma}}$ *as an approximation to* $\mathbf{\Sigma}$ *are better when m is small and, as a result, the use of the asymptotic distribution* χ^2_{m-1} *for* $W_{\mathcal{M}}$ *is more precise. Now, the first steps of the procedure of construction of the model confidence set are performed for large sets* $\mathcal{M}$*, while the last ones are operating on smaller sets* $\mathcal{M}$*. For this reason, we expect the effect of the use of* $\widehat{\mathbf{\Sigma}}$ *to be less important than it could be, because in the first steps the elimination of the least performing alternatives will be almost automatic.*
(iii) The matrix $\widehat{\mathbf{\Sigma}}$ *can be built once and for all at the beginning of the procedure, and submatrices can be selected at each step, or can be computed from scratch at each step. Despite the second method is more time-consuming, we use it because it is sensible to suppose that it provides better performances.*
(iv) For Ledoit and Wolf (2004a), we are able to obtain a convergence rate. If $\frac{m^5}{n} = o(1)$,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\mathbf{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|$$
$$= O\left(R_{m,n} + \frac{m^{\frac{5}{2}}}{n^{\frac{1}{2}}}\right).$$

5 Monte Carlo Experiment

The previous analysis shows what is the effect of using a non-standard estimator for the covariance matrix on the single test, but studying theoretically the impact of this choice on the characteristics of the confidence set is beyond the scope of this paper. In order to do so, in this section, we provide a Monte Carlo experiment to test the performance of our methodology in different settings.

Specifically, we consider the case in which the random variables are the same across the different alternatives, and we estimate the regularized covariance matrix $\widehat{\mathbf{\Sigma}}$ using three different techniques: (i) the linear shrinkage estimator proposed by Ledoit and Wolf (2004a), as implemented in Scikit-Learn (see (Oriol and Miot, 2025)), (ii) the linear shrinkage estimator of Ledoit and Wolf (2003) and Ledoit and Wolf (2004b), and (iii) the analytical nonlinear shrinkage estimator of Ledoit and Wolf (2020).

The simulation experiment follows the steps below:

1. we define a set of m points, indexed by $i = 1, \ldots, m$, corresponding to the possible alternatives;

2. we select as $\mathcal{M}^{\star}$ a subset of cardinality $m^{\star} = \lceil p \cdot m \rceil$ of the previous set corresponding to the true optima, e.g., when $m = 100$ and $p = 5\%$, the number of true minima is $m^{\star} = 5$;

3. we define the true means as

$$\mu_i = \begin{cases} 0, & \text{if } i = 1, \ldots, m^{\star}, \\ 0.01\left(i - m^{\star}\right), & \text{if } i = m^{\star}+1, \ldots, m, \end{cases}$$

so that the $m^\star$ alternatives are tied at the minimum and the remaining ones follow in strictly increasing order;

4. we sample independently a random variable ξ_j for any $j = 1, \ldots, n$, and a random variable ξ_{ij} for any $j = 1, \ldots, n$ and $i = 1, \ldots, m$;
5. for any $j = 1, \ldots, n$ and $i = 1, \ldots, m$, we compute

$$X_j(i) = a_{ij}\xi_j + b_{ij}\xi_{ij} + \mu_i,$$

   where ξ_j, ξ_{ij}, a_{ij} and b_{ij} take these values:

   (a) $\xi_j \sim \mathcal{N}(0,1)$, $\xi_{ij} \sim \mathcal{N}(0,1)$, $a_{ij} = 0.2$ and $b_{ij} = 0.5$;
   (b) $\xi_j \sim \mathcal{N}(0,1)$, $\xi_{ij} \sim \mathcal{N}(0,1)$, $a_{ij} = 0.5$ and $b_{ij} = 0.2$;
   (c) $\xi_j \sim \mathcal{N}(0,1)$, $\xi_{ij} \sim \mathcal{N}(0,1)$, a_{ij}^2 increases linearly from 0.2^2 to 0.5^2 and b_{ij}^2 decreases linearly from 0.5^2 to 0.2^2;
   (d) $\xi_j \sim \mathcal{N}(0,1)$, $\xi_{ij} \sim \mathcal{N}(0,1)$, a_{ij}^2 decreases linearly from 0.5^2 to 0.2^2 and b_{ij}^2 increases linearly from 0.2^2 to 0.5^2;
   (e) $-\xi_j - 1 \sim \mathcal{E}(1)$, $-\xi_{ij} - 1 \sim \mathcal{E}(1)$, $a_{ij} = 0.2$ and $b_{ij} = 0.5$;
   (f) $\xi_j + 1 \sim \mathcal{E}(1)$, $\xi_{ij} + 1 \sim \mathcal{E}(1)$, $a_{ij} = 0.2$ and $b_{ij} = 0.5$;
6. for each alternative i, we compute the sample average

$$\widehat{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} X_j(i),$$

   the centered observations $\widetilde{X}_j(i) = X_j(i) - \widehat{\mu}_i$, the m-vector $\widetilde{\mathbf{x}}_j$ whose i-th element is $\widetilde{X}_j(i)$, and the classical sample covariance matrix

$$\widehat{\boldsymbol{\Sigma}}_0 = \frac{1}{n}\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j',$$

   which is singular for $m > n$;
7. for each case, we apply the MCS procedure:
   (a) starting from the set $\mathcal{M}^0 = \{1, \ldots, m_0\}$, we order the alternatives by $\widehat{\mu}_i$, so that $\widehat{\mu}_1 \leq \widehat{\mu}_2 \leq \cdots \leq \widehat{\mu}_{m_0}$;
   (b) we compute the covariance matrix using one of four different estimators:
      i. the classical sample covariance matrix, $\widehat{\boldsymbol{\Sigma}}_0$;
      ii. the linear shrinkage estimator in Ledoit and Wolf (2004a) (see Section 4.5.1);
      iii. the linear shrinkage estimator in Ledoit and Wolf (2003) and Ledoit and Wolf (2004b) (see Section 4.5.2);
      iv. the nonlinear shrinkage estimator in Ledoit and Wolf (2020) (see Section 4.5.3);
   (c) we define the elimination rule $e_{\mathcal{M}}$, i.e. the index $j \in \mathcal{M}$ with the largest value $\widehat{\mu}_j$, and the test $W_{\mathcal{M}}$;
   (d) we test, with level $1-\alpha$, that $\widehat{\mu}_i = \widehat{\mu}_j\ \forall i, j \in \mathcal{M}$. If $\widehat{p}_{e_{\mathcal{M}_1}} \geq \alpha$, the procedure is over. Otherwise, we use $e_{\mathcal{M}}$ to remove an alternative from $\mathcal{M}^0$ and we get $\mathcal{M}^1 = \{1, \ldots, m_0 - 1\}$;
   (e) we apply the steps from (b) to (d) on the set $\mathcal{M}^1$, and if $\widehat{p}_{e_{\mathcal{M}_2}} < \alpha$ we use again $e_{\mathcal{M}}$ to obtain $\mathcal{M}^2 = \{1, \ldots, m_0 - 2\}$. We continue the procedure until $\mathsf{H}_{0,\mathcal{M}}$ is not rejected.

The algorithm is replicated $R = 100{,}000$ times using 27 different configurations and $\alpha = 0.05$. The results of the simulation experiment are exposed in Tables I-VI.

Now we discuss the rationale behind the choices of (a)-(f) in step 5. For the six cases, we keep the variances σ_i^2 constant across alternatives, so that differences in the results cannot be imputed to heteroskedasticity. Both (a) and (b) feature exchangeable or equicorrelated Gaussian random variables, but the common factor takes $4/29$ of the total variance in (a) and $25/29$ in (b). Therefore, (a) is nearer to the case of independent sampling, while (b) is nearer to the case in which errors are the same across the alternatives. The two cases (c) and (d) keep the total variance constant for each alternative, but provide two distinct situations in which the minima are respectively less or more strongly correlated than the rest. At last, (e) and (f) provide a version of (a) in which $\widehat{\mu}_i$ is skewed respectively towards negative or positive values.

Table I: Monte Carlo results with common variance 0.2^2 and idiosyncratic variance 0.5^2.

m	n	p	$\widehat{\boldsymbol{\Sigma}}_0$			LW (2004a)			LW (2004b)			LW (2020)		
			Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR
10	20	0.05	0.9534	9.0250	0.8968	0.9958	9.8818	0.9873	0.9970	9.9087	0.9902	0.9990	9.9731	0.9971
10	20	0.1	0.9530	9.0254	0.8969	0.9957	9.8831	0.9875	0.9970	9.9109	0.9904	0.9991	9.9732	0.9971
10	20	0.2	0.8987	9.0347	0.8930	0.9910	9.8885	0.9873	0.9930	9.9121	0.9899	0.9981	9.9743	0.9970
10	200	0.05	0.9967	9.1952	0.9109	0.9972	9.3271	0.9255	0.9978	9.3374	0.9266	0.9976	9.3698	0.9302
10	200	0.1	0.9964	9.1973	0.9112	0.9976	9.3297	0.9258	0.9977	9.3368	0.9266	0.9976	9.3657	0.9298
10	200	0.2	0.9888	9.3055	0.9147	0.9918	9.4249	0.9292	0.9920	9.4344	0.9304	0.9923	9.4608	0.9336
10	2000	0.05	0.9976	4.2841	0.3652	0.9978	4.3135	0.3684	0.9973	4.3073	0.3678	0.9976	4.3043	0.3674
10	2000	0.1	0.9975	4.2960	0.3665	0.9975	4.3077	0.3678	0.9976	4.3087	0.3679	0.9974	4.3030	0.3673
10	2000	0.2	0.9777	4.9474	0.3712	0.9788	4.9662	0.3735	0.9782	4.9612	0.3729	0.9776	4.9575	0.3725
50	20	0.05	1.0000	48.9074	0.9768	0.9890	34.8579	0.6781	0.9925	35.9011	0.7002	0.9989	41.6028	0.8214
50	20	0.1	0.9999	48.9234	0.9761	0.9764	35.5073	0.6785	0.9825	36.5117	0.7007	0.9975	42.0627	0.8237
50	20	0.2	0.9996	48.9145	0.9729	0.9339	38.0087	0.7021	0.9479	38.9975	0.7264	0.9899	43.7937	0.8451
50	200	0.05	0.9817	12.9743	0.2127	0.9861	13.4498	0.2227	0.9866	13.4888	0.2235	0.9881	13.7739	0.2295
50	200	0.1	0.9619	14.3427	0.2086	0.9720	14.8405	0.2194	0.9718	14.8676	0.2200	0.9764	15.2311	0.2278
50	200	0.2	0.9038	18.5823	0.2176	0.9330	19.1676	0.2312	0.9320	19.1933	0.2319	0.9528	19.7774	0.2459
50	2000	0.05	0.9700	5.8077	0.0605	0.9702	5.8168	0.0607	0.9699	5.8173	0.0607	0.9700	5.8118	0.0606
50	2000	0.1	0.9561	7.7057	0.0614	0.9575	7.7169	0.0616	0.9567	7.7152	0.0616	0.9578	7.7216	0.0617
50	2000	0.2	0.9347	12.7078	0.0699	0.9356	12.7298	0.0703	0.9366	12.7268	0.0702	0.9382	12.7391	0.0705
250	20	0.05	1.0000	248.9788	0.9957	0.8958	40.9258	0.1183	0.9156	41.9770	0.1227	0.9774	47.4295	0.1454
250	20	0.1	1.0000	248.9868	0.9955	0.7536	50.8035	0.1162	0.7953	51.8924	0.1207	0.9196	56.4091	0.1399
250	20	0.2	1.0000	248.9949	0.9949	0.5101	73.6969	0.1229	0.5809	75.0306	0.1287	0.7645	78.7618	0.1454
250	200	0.05	1.0000	248.6845	0.9944	0.9085	21.9665	0.0383	0.9107	21.9968	0.0384	0.9456	25.4400	0.0528
250	200	0.1	1.0000	248.6888	0.9942	0.8223	33.5526	0.0391	0.8274	33.6026	0.0393	0.9314	37.6776	0.0567
250	200	0.2	1.0000	248.6931	0.9935	0.6281	57.6224	0.0414	0.6367	57.6938	0.0417	0.9663	64.0587	0.0705
250	2000	0.05	0.9276	15.7474	0.0120	0.9284	15.7671	0.0121	0.9291	15.7646	0.0121	0.9297	15.7784	0.0121
250	2000	0.1	0.9070	27.8927	0.0134	0.9093	27.9048	0.0135	0.9099	27.9115	0.0135	0.9196	27.9875	0.0138
250	2000	0.2	0.8807	53.0600	0.0162	0.8848	53.0878	0.0163	0.8866	53.0923	0.0163	0.9168	53.3189	0.0172

Notes: Cov. stands for coverage, FPR stands for false positive rate and MCS size is the average cardinality of the confidence set.

Table II: Monte Carlo results with common variance 0.5^2 and idiosyncratic variance 0.2^2.

m	n	p	$\widehat{\boldsymbol{\Sigma}}_0$			LW (2004a)			LW (2004b)			LW (2020)		
			Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR
10	20	0.05	0.9778	8.4079	0.8256	0.9992	9.8493	0.9833	0.9993	9.8717	0.9858	0.9993	9.8914	0.9880
10	20	0.1	0.9779	8.3945	0.8241	0.9993	9.8539	0.9838	0.9993	9.8699	0.9856	0.9993	9.8897	0.9878
10	20	0.2	0.9391	8.4897	0.8192	0.9974	9.8675	0.9838	0.9981	9.8853	0.9859	0.9980	9.9057	0.9885
10	200	0.05	0.9971	5.2992	0.4780	0.9978	5.5621	0.5071	0.9980	5.5734	0.5084	0.9971	5.4107	0.4904
10	200	0.1	0.9969	5.3096	0.4792	0.9981	5.5653	0.5075	0.9981	5.5821	0.5093	0.9971	5.4287	0.4924
10	200	0.2	0.9803	5.8692	0.4862	0.9854	6.1194	0.5168	0.9874	6.1356	0.5186	0.9824	5.9909	0.5011
10	2000	0.05	0.9998	1.8539	0.0949	0.9998	1.8658	0.0962	0.9998	1.8639	0.0960	0.9998	1.8533	0.0948
10	2000	0.1	0.9999	1.8549	0.0949	0.9999	1.8630	0.0959	0.9999	1.8681	0.0965	0.9998	1.8545	0.0950
10	2000	0.2	0.9591	2.7301	0.0964	0.9602	2.7404	0.0975	0.9585	2.7393	0.0976	0.9594	2.7330	0.0967
50	20	0.05	1.0000	48.5984	0.9702	0.9937	18.5155	0.3302	0.9951	18.8443	0.3372	0.9999	23.6327	0.4389
50	20	0.1	1.0000	48.6119	0.9692	0.9857	19.6020	0.3248	0.9875	19.8651	0.3306	0.9996	24.3565	0.4302
50	20	0.2	0.9999	48.6046	0.9651	0.9469	23.2687	0.3333	0.9530	23.5362	0.3398	0.9970	27.2016	0.4301
50	200	0.05	0.9708	6.6637	0.0787	0.9790	6.9027	0.0836	0.9801	6.9234	0.0840	0.9751	6.8058	0.0816
50	200	0.1	0.9519	8.4636	0.0783	0.9659	8.7215	0.0836	0.9669	8.7387	0.0840	0.9625	8.6624	0.0825
50	200	0.2	0.9099	13.2985	0.0855	0.9398	13.6099	0.0921	0.9403	13.6204	0.0924	0.9471	13.6951	0.0941
50	2000	0.05	0.9544	3.7124	0.0164	0.9556	3.7220	0.0165	0.9553	3.7202	0.0165	0.9542	3.7125	0.0164
50	2000	0.1	0.9487	5.7362	0.0179	0.9503	5.7463	0.0181	0.9500	5.7471	0.0181	0.9495	5.7392	0.0179
50	2000	0.2	0.9383	10.8036	0.0222	0.9421	10.8243	0.0226	0.9414	10.8215	0.0225	0.9410	10.8163	0.0224
250	20	0.05	1.0000	248.9394	0.9955	0.9166	25.7462	0.0542	0.9239	26.0345	0.0554	0.9933	29.3779	0.0691
250	20	0.1	1.0000	248.9397	0.9953	0.7716	36.2764	0.0519	0.7895	36.5974	0.0532	0.9685	39.5694	0.0649
250	20	0.2	1.0000	248.9434	0.9947	0.5298	58.7570	0.0524	0.5539	59.1490	0.0538	0.8947	63.2986	0.0672
250	200	0.05	1.0000	248.2357	0.9926	0.9249	16.5781	0.0155	0.9268	16.5975	0.0156	0.9433	19.6328	0.0283
250	200	0.1	1.0000	248.2124	0.9921	0.8634	28.4474	0.0162	0.8674	28.4803	0.0163	0.9501	31.8606	0.0308
250	200	0.2	1.0000	248.2344	0.9912	0.6501	52.5192	0.0162	0.6589	52.5745	0.0164	0.9834	57.4029	0.0371
250	2000	0.05	0.9360	13.8488	0.0039	0.9391	13.8598	0.0040	0.9401	13.8659	0.0040	0.9394	13.8652	0.0040
250	2000	0.1	0.9272	25.9559	0.0047	0.9300	25.9764	0.0048	0.9294	25.9736	0.0048	0.9379	26.0057	0.0049
250	2000	0.2	0.9109	51.0655	0.0060	0.9181	51.1002	0.0061	0.9171	51.0983	0.0061	0.9406	51.2237	0.0066

Notes: Cov. stands for coverage, FPR stands for false positive rate and MCS size is the average cardinality of the confidence set.

Table III: Monte Carlo results with common variance increasing linearly from 0.2^2 to 0.5^2 and idiosyncratic variance decreasing linearly from 0.5^2 to 0.2^2.

m	n	p	$\widehat{\mathbf{\Sigma}}_0$			LW (2004a)			LW (2004b)			LW (2020)		
			Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR
10	20	0.05	0.9734	8.9464	0.8859	0.9987	9.8432	0.9827	0.9991	9.8702	0.9857	0.9994	9.9442	0.9939
10	20	0.1	0.9741	8.9578	0.8871	0.9986	9.8418	0.9826	0.9990	9.8717	0.9859	0.9994	9.9456	0.9940
10	20	0.2	0.9369	8.9876	0.8824	0.9969	9.8524	0.9820	0.9973	9.8778	0.9851	0.9986	9.9467	0.9935
10	200	0.05	0.9984	8.6364	0.8487	0.9990	8.8001	0.8668	0.9989	8.8069	0.8676	0.9988	8.8304	0.8702
10	200	0.1	0.9983	8.6458	0.8497	0.9988	8.7810	0.8647	0.9991	8.8086	0.8677	0.9987	8.8355	0.8708
10	200	0.2	0.9914	9.0046	0.8767	0.9943	9.1152	0.8902	0.9946	9.1274	0.8916	0.9939	9.1542	0.8951
10	2000	0.05	0.9997	2.3494	0.1499	0.9996	2.3532	0.1504	0.9997	2.3546	0.1505	0.9997	2.3499	0.1500
10	2000	0.1	0.9996	2.3428	0.1492	0.9997	2.3500	0.1500	0.9997	2.3565	0.1508	0.9996	2.3480	0.1498
10	2000	0.2	0.9660	3.5414	0.1969	0.9673	3.5496	0.1978	0.9685	3.5563	0.1985	0.9662	3.5456	0.1974
50	20	0.05	1.0000	48.8716	0.9760	0.9971	24.8274	0.4645	0.9979	25.4430	0.4776	1.0000	38.3325	0.7518
50	20	0.1	1.0000	48.8729	0.9750	0.9912	25.9676	0.4662	0.9933	26.5133	0.4782	1.0000	39.0957	0.7577
50	20	0.2	1.0000	48.8789	0.9720	0.9523	29.5898	0.4912	0.9597	30.2188	0.5067	0.9999	41.9025	0.7976
50	200	0.05	0.9762	7.8258	0.1033	0.9829	8.0689	0.1083	0.9846	8.0947	0.1088	0.9818	8.0228	0.1073
50	200	0.1	0.9635	9.8971	0.1099	0.9734	10.1366	0.1149	0.9733	10.1531	0.1153	0.9748	10.1127	0.1144
50	200	0.2	0.9354	15.3966	0.1371	0.9521	15.6638	0.1431	0.9540	15.6889	0.1437	0.9649	15.7844	0.1458
50	2000	0.05	0.9563	3.9119	0.0206	0.9583	3.9197	0.0207	0.9582	3.9242	0.0208	0.9574	3.9038	0.0204
50	2000	0.1	0.9528	6.0500	0.0248	0.9528	6.0598	0.0250	0.9533	6.0580	0.0249	0.9530	6.0435	0.0246
50	2000	0.2	0.9461	11.4034	0.0370	0.9463	11.4047	0.0370	0.9477	11.4120	0.0371	0.9474	11.3927	0.0367
250	20	0.05	1.0000	248.9809	0.9957	0.9075	26.5596	0.0577	0.9161	26.8744	0.0590	0.9962	32.0618	0.0804
250	20	0.1	1.0000	248.9898	0.9955	0.6933	36.7071	0.0546	0.7160	37.0968	0.0561	0.9808	42.9997	0.0801
250	20	0.2	1.0000	248.9888	0.9949	0.2895	56.8921	0.05153	0.3146	57.4715	0.0532	0.9259	68.1605	0.0912
250	200	0.05	1.0000	248.2661	0.9927	0.9284	17.0745	0.0176	0.9318	17.0998	0.0177	0.9510	20.4990	0.0319
250	200	0.1	1.0000	248.2677	0.9923	0.8673	29.2154	0.0196	0.8685	29.2369	0.0197	0.9542	32.9632	0.0356
250	200	0.2	1.0000	248.2639	0.9913	0.6211	53.4324	0.0214	0.6237	53.4732	0.0215	0.9796	59.1927	0.0461
250	2000	0.05	0.9390	14.0224	0.0047	0.9386	14.0259	0.0047	0.9417	14.0300	0.0047	0.9405	14.0071	0.0046
250	2000	0.1	0.9307	26.2874	0.0062	0.9340	26.3049	0.0062	0.9349	26.3046	0.0062	0.9394	26.2896	0.0061
250	2000	0.2	0.9169	51.6854	0.0091	0.9219	51.7154	0.0092	0.9209	51.7089	0.0092	0.9373	51.7412	0.0092

Notes: Cov. stands for coverage, FPR stands for false positive rate and MCS size is the average cardinality of the confidence set.

Table IV: Monte Carlo results with common variance decreasing linearly from 0.5^2 to 0.2^2 and idiosyncratic variance increasing linearly from 0.2^2 to 0.5^2.

m	n	p	$\widehat{\boldsymbol{\Sigma}}_0$			LW (2004a)			LW (2004b)			LW (2020)		
			Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR
10	20	0.05	0.9308	8.8629	0.8813	0.9885	9.7887	0.9778	0.9907	9.8235	0.9814	0.9960	9.9246	0.9921
10	20	0.1	0.9307	8.8678	0.8819	0.9886	9.7913	0.9781	0.9905	9.8268	0.9818	0.9962	9.9257	0.9922
10	20	0.2	0.8609	8.8751	0.8784	0.9762	9.8002	0.9782	0.9812	9.8336	0.9817	0.9924	9.9309	0.9924
10	200	0.05	0.9926	8.7075	0.8572	0.9935	8.8090	0.8684	0.9939	8.8316	0.8709	0.9937	8.8584	0.8739
10	200	0.1	0.9930	8.7129	0.8578	0.9941	8.8184	0.8694	0.9936	8.8328	0.8710	0.9938	8.8562	0.8736
10	200	0.2	0.9773	8.8246	0.8560	0.9804	8.9267	0.8684	0.9805	8.9471	0.8709	0.9812	8.9867	0.8758
10	2000	0.05	0.9976	3.9322	0.3261	0.9974	3.9331	0.3262	0.9975	3.9470	0.3277	0.9971	3.9294	0.3258
10	2000	0.1	0.9973	3.9225	0.3250	0.9975	3.9361	0.3265	0.9981	3.9358	0.3264	0.9973	3.9397	0.3269
10	2000	0.2	0.9746	4.4906	0.3145	0.9760	4.5052	0.3162	0.9746	4.5034	0.3161	0.9745	4.4974	0.3154
50	20	0.05	0.9999	48.8717	0.9759	0.9520	27.2111	0.5162	0.9580	27.9436	0.5317	0.9959	38.5295	0.7560
50	20	0.1	0.9997	48.8782	0.9751	0.9020	27.7467	0.5079	0.9136	28.4378	0.5229	0.9911	38.9301	0.7542
50	20	0.2	0.9978	48.8803	0.9721	0.7355	29.6873	0.5008	0.7609	30.3762	0.5170	0.9685	40.6761	0.7677
50	200	0.05	0.9782	12.4363	0.2013	0.9829	12.7082	0.2070	0.9820	12.7271	0.2074	0.9836	13.0745	0.2147
50	200	0.1	0.9550	13.7413	0.1954	0.9625	14.0217	0.2015	0.9627	14.0469	0.2020	0.9694	14.4980	0.2119
50	200	0.2	0.8873	17.7788	0.1981	0.9050	18.0719	0.2047	0.9075	18.1054	0.2055	0.9359	18.8937	0.2243
50	2000	0.05	0.9676	5.6826	0.0579	0.9673	5.6887	0.0581	0.9695	5.7000	0.0582	0.9683	5.7020	0.0583
50	2000	0.1	0.9548	7.5510	0.0580	0.9553	7.5557	0.0581	0.9549	7.5598	0.0582	0.9552	7.5764	0.0586
50	2000	0.2	0.9288	12.4467	0.0635	0.9331	12.4563	0.0636	0.9309	12.4567	0.0637	0.9320	12.5034	0.0649
250	20	0.05	1.0000	248.9727	0.9957	0.8450	38.2673	0.1074	0.8664	39.2234	0.1113	0.9735	46.2098	0.1402
250	20	0.1	1.0000	248.9736	0.9954	0.6226	47.1783	0.1012	0.6688	48.1403	0.1050	0.9069	55.0121	0.1339
250	20	0.2	1.0000	248.9870	0.9949	0.2454	66.8142	0.0953	0.2932	67.9625	0.0995	0.7324	76.7938	0.1358
250	200	0.05	1.0000	248.7165	0.9946	0.8987	21.6267	0.0369	0.8986	21.6543	0.0371	0.9399	25.5964	0.0535
250	200	0.1	1.0000	248.7201	0.9943	0.7896	32.9362	0.0366	0.7927	32.9831	0.0368	0.9281	37.5775	0.0563
250	200	0.2	1.0000	248.7209	0.9936	0.4978	56.0585	0.0356	0.5056	56.1389	0.0359	0.9630	63.6490	0.0685
250	2000	0.05	0.9247	15.6798	0.0117	0.9289	15.6919	0.0118	0.9275	15.7001	0.0118	0.9303	15.7321	0.0119
250	2000	0.1	0.9036	27.7584	0.0128	0.9060	27.7770	0.0129	0.9054	27.7818	0.0129	0.9158	27.8858	0.0133
250	2000	0.2	0.8730	52.8030	0.0149	0.8756	52.8245	0.0150	0.8767	52.8253	0.0150	0.9138	53.1411	0.0163

Notes: Cov. stands for coverage, FPR stands for false positive rate and MCS size is the average cardinality of the confidence set.

Table V: Monte Carlo results with common variance 0.2^2 and idiosyncratic variance 0.5^2. The additive error is given by the weighted sum of two centered exponential random variables with changed sign.

m	n	p	$\widehat{\boldsymbol{\Sigma}}_0$			LW (2004a)			LW (2004b)			LW (2020)		
			Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR
10	20	0.05	0.9589	9.0212	0.8958	0.9974	9.9071	0.9900	0.9983	9.9315	0.9926	0.9982	9.9462	0.9942
10	20	0.1	0.9580	9.0179	0.8955	0.9977	9.9068	0.9899	0.9980	9.9292	0.9924	0.9984	9.9455	0.9941
10	20	0.2	0.9050	9.0215	0.8902	0.9941	9.9111	0.9897	0.9957	9.9341	0.9923	0.9962	9.9474	0.9939
10	200	0.05	0.9976	9.2400	0.9158	0.9986	9.4198	0.9357	0.9985	9.4181	0.9355	0.9985	9.3893	0.9323
10	200	0.1	0.9979	9.2384	0.9156	0.9985	9.4121	0.9348	0.9987	9.4262	0.9364	0.9983	9.3902	0.9324
10	200	0.2	0.9933	9.3482	0.9194	0.9952	9.4989	0.9380	0.9954	9.5056	0.9388	0.9945	9.4832	0.9361
10	2000	0.05	0.9977	4.3240	0.3696	0.9978	4.3470	0.3721	0.9980	4.3472	0.3721	0.9979	4.3240	0.3696
10	2000	0.1	0.9977	4.3125	0.3683	0.9981	4.3519	0.3726	0.9981	4.3496	0.3724	0.9980	4.3293	0.3701
10	2000	0.2	0.9791	4.9832	0.3755	0.9805	5.0067	0.3783	0.9812	5.0069	0.3782	0.9786	4.9816	0.3754
50	20	0.05	1.0000	48.9221	0.9771	0.9990	37.2945	0.7297	0.9994	38.2284	0.7496	1.0000	41.5881	0.8210
50	20	0.1	1.0000	48.9260	0.9761	0.9974	37.8274	0.7296	0.9984	38.7355	0.7497	0.9998	42.0499	0.8233
50	20	0.2	1.0000	48.9196	0.9730	0.9908	39.9883	0.7500	0.9940	40.8769	0.7721	0.9985	43.6272	0.8407
50	200	0.05	0.9879	13.3736	0.2210	0.9927	13.9620	0.2334	0.9931	13.9877	0.2339	0.9921	14.0475	0.2352
50	200	0.1	0.9744	14.7491	0.2173	0.9843	15.3255	0.2299	0.9848	15.3821	0.2311	0.9832	15.4553	0.2328
50	200	0.2	0.9292	18.9728	0.2264	0.9598	19.6567	0.2426	0.9583	19.6917	0.2435	0.9644	19.9451	0.2497
50	2000	0.05	0.9716	5.8430	0.0612	0.9729	5.8569	0.0615	0.9735	5.8683	0.0617	0.9714	5.8439	0.0612
50	2000	0.1	0.9588	7.7496	0.0623	0.9615	7.7701	0.0627	0.9597	7.7713	0.0627	0.9600	7.7487	0.0622
50	2000	0.2	0.9385	12.7582	0.0709	0.9427	12.7793	0.0713	0.9419	12.7831	0.0714	0.9417	12.7687	0.0711
250	20	0.05	1.0000	248.9865	0.9957	0.9864	44.2637	0.1320	0.9910	45.3745	0.1366	0.9981	48.5900	0.1502
250	20	0.1	1.0000	248.9950	0.9955	0.9530	53.9919	0.1291	0.9678	55.1456	0.1342	0.9867	57.5093	0.1446
250	20	0.2	1.0000	248.9937	0.9950	0.8541	76.9160	0.1357	0.8996	78.2438	0.1419	0.9372	79.6777	0.1488
250	200	0.05	1.0000	248.6794	0.9944	0.9425	22.4670	0.0402	0.9441	22.5090	0.0404	0.9584	25.7361	0.0539
250	200	0.1	1.0000	248.6798	0.9941	0.8824	34.0862	0.0410	0.8847	34.1320	0.0412	0.9492	37.8406	0.0573
250	200	0.2	1.0000	248.6848	0.9934	0.7344	58.3387	0.0438	0.7440	58.4228	0.0441	0.9751	63.9495	0.0699
250	2000	0.05	0.9335	15.8026	0.0122	0.9365	15.8283	0.0123	0.9352	15.8226	0.0123	0.9350	15.8152	0.0122
250	2000	0.1	0.9119	27.9309	0.0135	0.9177	27.9664	0.0137	0.9173	27.9670	0.0137	0.9233	28.0006	0.0138
250	2000	0.2	0.8877	53.1024	0.0163	0.8934	53.1430	0.0165	0.8930	53.1469	0.0165	0.9207	53.3189	0.0172

Notes: Cov. stands for coverage, FPR stands for false positive rate and MCS size is the average cardinality of the confidence set.

Table VI: Monte Carlo results with common variance 0.2^2 and idiosyncratic variance 0.5^2. The additive error is given by the weighted sum of two centered exponential random variables.

m	n	p	$\widehat{\boldsymbol{\Sigma}}_0$			LW (2004a)			LW (2004b)			LW (2020)		
			Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR	Cov.	MCS Size	FPR
10	20	0.05	0.9220	8.5734	0.8502	0.9966	9.9063	0.9900	0.9978	9.9288	0.9923	0.9960	9.9114	0.9906
10	20	0.1	0.9231	8.5813	0.8509	0.9973	9.9099	0.9903	0.9975	9.9273	0.9922	0.9959	9.9113	0.9906
10	20	0.2	0.8434	8.5763	0.8442	0.9926	9.9112	0.9899	0.9946	9.9321	0.9922	0.9911	9.9121	0.9903
10	200	0.05	0.9940	9.0454	0.8946	0.9967	9.2976	0.9223	0.9963	9.3123	0.9240	0.9960	9.2687	0.9192
10	200	0.1	0.9935	9.0450	0.8946	0.9966	9.3007	0.9227	0.9968	9.2993	0.9225	0.9959	9.2680	0.9191
10	200	0.2	0.9807	9.1629	0.8980	0.9892	9.4032	0.9268	0.9896	9.4136	0.9281	0.9880	9.3836	0.9245
10	2000	0.05	0.9972	4.2496	0.3614	0.9974	4.2869	0.3655	0.9976	4.2882	0.3656	0.9976	4.2636	0.3629
10	2000	0.1	0.9974	4.2496	0.3614	0.9974	4.2851	0.3653	0.9974	4.2790	0.3646	0.9973	4.2629	0.3629
10	2000	0.2	0.9760	4.9094	0.3667	0.9768	4.9292	0.3691	0.9778	4.9354	0.3697	0.9769	4.9258	0.3686
50	20	0.05	0.9998	48.9416	0.9775	0.9664	32.5778	0.6301	0.9726	33.7290	0.6544	0.9891	38.2962	0.7512
50	20	0.1	0.9995	48.9361	0.9764	0.9354	33.3947	0.6325	0.9461	34.4863	0.6565	0.9757	38.8407	0.7526
50	20	0.2	0.9971	48.9378	0.9735	0.8470	36.3880	0.6643	0.8706	37.4071	0.6890	0.9344	41.1237	0.7799
50	200	0.05	0.9642	12.2692	0.1981	0.9797	13.0530	0.2144	0.9799	13.1089	0.2156	0.9779	13.2365	0.2183
50	200	0.1	0.9338	13.6565	0.1942	0.9591	14.4480	0.2110	0.9599	14.4936	0.2120	0.9615	14.7160	0.2170
50	200	0.2	0.8504	17.8405	0.2014	0.9059	18.7987	0.2230	0.9087	18.8268	0.2235	0.9269	19.3450	0.2360
50	2000	0.05	0.9673	5.7500	0.0594	0.9678	5.7740	0.0599	0.9683	5.7755	0.0599	0.9669	5.7706	0.0598
50	2000	0.1	0.9520	7.6498	0.0603	0.9557	7.6786	0.0608	0.9543	7.6752	0.0608	0.9534	7.6788	0.0609
50	2000	0.2	0.9272	12.6427	0.0686	0.9330	12.6838	0.0694	0.9316	12.6837	0.0694	0.9332	12.7058	0.0699
250	20	0.05	1.0000	248.9842	0.9957	0.7810	38.7166	0.1097	0.8097	39.8107	0.1141	0.8792	43.1292	0.1277
250	20	0.1	1.0000	248.9873	0.9955	0.5857	49.1195	0.1100	0.6331	50.3182	0.1149	0.7164	52.5491	0.1241
250	20	0.2	1.0000	248.9866	0.9949	0.3465	72.7915	0.1209	0.4041	74.1506	0.1266	0.4658	75.6421	0.1327
250	200	0.05	1.0000	248.6946	0.9945	0.8751	21.6025	0.0370	0.8813	21.6511	0.0372	0.9121	25.0498	0.0513
250	200	0.1	1.0000	248.6827	0.9941	0.7768	33.2239	0.0380	0.7817	33.2794	0.0382	0.8955	37.4205	0.0558
250	200	0.2	1.0000	248.6998	0.9935	0.5862	57.3958	0.0410	0.5917	57.4683	0.0412	0.9367	63.8888	0.0698
250	2000	0.05	0.9208	15.6852	0.0118	0.9222	15.7076	0.0119	0.9237	15.7152	0.0119	0.9273	15.7539	0.0120
250	2000	0.1	0.8972	27.8226	0.0132	0.9028	27.8572	0.0133	0.9018	27.8549	0.0133	0.9141	27.9579	0.0137
250	2000	0.2	0.8664	52.9704	0.0159	0.8752	53.0219	0.0161	0.8751	53.0289	0.0161	0.9111	53.2963	0.0171

Notes: Cov. stands for coverage, FPR stands for false positive rate and MCS size is the average cardinality of the confidence set.

We now discuss the results of the Monte Carlo experiments. The coverage (Cov.) represents the probability $\mathbb{P}\left\{\mathcal{M}^\star \subseteq \widehat{\mathcal{M}}^\star_{1-\alpha}\right\}$ that the solution set is contained in the confidence set. The false positive rate (FPR) is the proportion of non-optimal values in the confidence set. Finally, the MCS size (MCS Size) is the average cardinality of the confidence set across replications.

Let us start by comparing Tables I-VI. The redistribution of the variance between the common and idiosyncratic part directly affects the MCS size and FPR, while the effect on the coverage is more nuanced. When the common components is larger (Table II), the effect of the regularization is more evident, reflecting in lower MCS size and FPR. This can be attributed to the fact that the m alternatives are more correlated and $\boldsymbol{\Sigma}$ approaches a low-rank structure. We observe the same phenomena in the dynamic settings, where the common/idiosyncratic component of the variance changes linearly (Tables III and IV). It is worth noting that the shrinkage estimators are more effective when the common component of the variance increases linearly and the idiosyncratic components decreases linearly, providing lower values of the MCS size and FPR as the number of alternatives and sample size increase while maintaining a good coverage. When the additive error is given by the weighted sum of two centered exponential random variables (Tables V and VI), the results are close to the ones in Table I. Therefore, our approach is robust to some degree of asymmetry.

When m is small, namely $m = 10$, all the estimators provide similar outcomes. Notice that, when the ratio m/n is small, e.g., $m = 10$ and $n = 2000$, $\widehat{\boldsymbol{\Sigma}}_0$ is already a good estimator as it yields coverage close to the nominal level and small MCS size. This is in line with Theorem 2. On the other hand, when $m > n$, the use of the classical sample covariance matrix estimator leads us to include almost all models in the final set (i.e. MCS size $\simeq m$ and FPR $\simeq 1$), as $\widehat{\boldsymbol{\Sigma}}_0$ is singular or ill-conditioned and the asymptotic distribution $W_{\mathcal{M}} \xrightarrow{\mathcal{D}} \chi^2_{m-1}$ is not a good approximation of the finite-sample one. Accordingly, the adoption of the shrinkage estimators improves the MCS size and FPR. This is also coherent with Theorem 2, which gives an upper bound on the error of replacing the true matrix $\boldsymbol{\Sigma}$ with an estimator $\widehat{\boldsymbol{\Sigma}}$, and Corollary 4, which links this approximation error to the spectral deviation between $\boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Sigma}}$. In fact, since the regularization reduces the dispersion of the eigenvalues and improves conditioning, the deviation between $\boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Sigma}}$ and the distributional error of the test statistic decrease.

In high-dimensional settings, i.e. large m and small n, the three shrinkage estimators perform surprisingly well, probably because the large number of alternatives m replaces the information lacking due to the small number of realizations n, and provide similar coverage and FPR. However, whereas the linear shrinkage estimators of Ledoit and Wolf (2004a) and Ledoit and Wolf (2003) and Ledoit and Wolf (2004b) produce smaller MCS size and lower FPR than the nonlinear shrinkage estimator of Ledoit and Wolf (2020), the latter performs better in terms of coverage. This is consistent with the motivation of approximating the optimal nonlinear eigenvalue transformation to better control the eigenstructure entering the bounds in Theorem 2. In more balanced designs, i.e. $m < n$, differences between shrinkage schemes are less evident, as $\widehat{\boldsymbol{\Sigma}}_0$ is better conditioned.

Finally, we highlight a trade-off between coverage, FPR, and MCS size. For the classical sample covariance estimator $\widehat{\boldsymbol{\Sigma}}_0$, when $m/n > 1$ the coverage approaches 1 but the MCS size and FPR increase. Moreover, when m, n are kept fixed and the percentage p of true minima among the possible alternatives increases, the coverage decreases and the MCS size and FPR increase. For shrinkage estimators, we observe that: (i) for fixed $m > 10$ and growing n, the coverage increases and the MCS size and FPR drop; (ii) for fixed m, n and increasing p, all the indicators worsen, i.e. the coverage becomes smaller and the MCS size and FPR increase. These results match the theoretical expectation that strong sets satisfying $\mathbb{P}\left\{\mathcal{M}^\star \subseteq \widehat{\mathcal{M}}^\star_{1-\alpha}\right\} \geq 1-\alpha$ can be conservative in finite samples. As n increases, nonoptimal elements are discarded with probability approaching 1 and the MCS converges to $\mathcal{M}^\star$, which is the statement of Theorem 1.

The effects of shrinkage estimation can be also illustrated using some heatmaps. Figures 1-6 show the ratio between the MCS size and the number of true minima for different combinations of m, n and p. We interpret it as a measure of misclassification, since it tells us how big is the MCS with respect to the target value $m^\star$.

We study the efficiency index $\text{MCS size}/m^\star$ along different dimensions. First, we note that its behavior is very similar across specifications of error distribution and composition of common and idiosyncratic variance. Second, for fixed m, n and increasing p, the efficiency index improves, thereby providing values close to one for large values of m and n. This is in line with our expectations as the higher is the number of true minima, the lower is the probability to include wrong alternatives in the MCS. Third, the efficiency index improves for fixed m and increasing n, respecting the theoretical results. Finally, for small values of p, the shrinkage estimators are

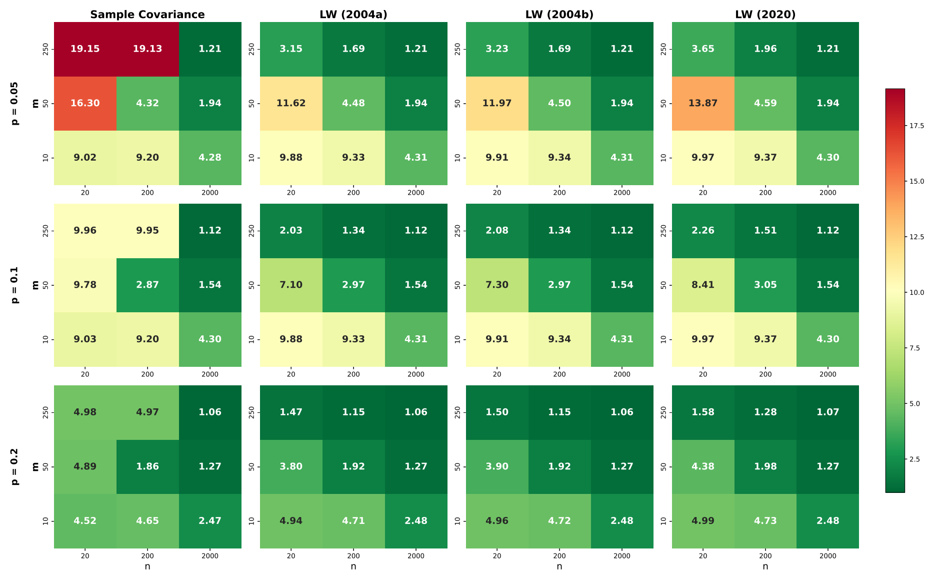

Figure 1: $^{\text{MCS size}}/_{m^\star}$ with common variance 0.2^2 and idiosyncratic variance 0.5^2.

Figure 2: $^{\text{MCS size}}/_{m^\star}$ with common variance 0.5^2 and idiosyncratic variance 0.2^2.

particularly effective in ameliorating the performance of the efficiency index.

In Figures 7-12 we depict the FPR for different combinations of m, n and p. As previously outlined, the effect of the shrinkage estimators is evident when $^m/_n > 1$.

6 Conclusions

In this paper, we suggest a method to compute a confidence set around the set $\mathcal{M}^\star$ of optimal solutions of a discrete optimization problem as tackled through the sample average approximation method, by combining model confidence sets (Hansen *et al.*, 2011; Seri *et al.*, 2021) and shrinkage estimation (Ledoit and Wolf, 2003; Ledoit and Wolf, 2004a; Ledoit and Wolf, 2004b; Ledoit and Wolf, 2020).

Specifically, we improve the method originally proposed by Futschik and Pflug (1995) and Futschik and Pflug (1997) along two directions: (i) we allow for potentially full covariance structures among the estimators $\widehat{\mu}_i$, a situation that arises in the SAA approach; (ii) we provide *strong level-α confidence sets*. The MCS allows us to identify all the optimal alternatives that are statistically indistinguishable, while the shrinkage estimators allow us to estimate the covariance matrix when the number of alternatives is at least large as the number of simulations.

Figure 3: MCS size$/m^\star$ with common variance increasing linearly from 0.2^2 to 0.5^2 and idiosyncratic variance decreasing linearly from 0.5^2 to 0.2^2.

We rigorously characterize the error of replacing the true matrix $\mathbf{\Sigma}$ with an estimator $\widehat{\mathbf{\Sigma}}$ and we provide its impact on the asymptotic distribution. We study the finite-sample properties of our method through an extensive Monte Carlo experiment, showing that the simulation outcomes align with the theoretical results.

The MCS can also be constructed from t-statistics, thereby avoiding the estimation of the covariance matrix $\mathbf{\Sigma}$. We leave the theoretical and practical investigation of this case for future research.

7 Proofs

Proof of Theorem 2. (i) We bound the distance in distribution between functions of $\widehat{\boldsymbol{\mu}}$ and Gaussian vectors. Let $\mathbf{Y} := \frac{1}{n^{\frac{1}{2}}} \sum_{j=1}^{n} (\mathbf{A}\mathbf{X}_j)' (\mathbf{A}\mathbf{\Sigma}\mathbf{A}')^{-\frac{1}{2}}$. We first note that, for $\mathbf{G} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, all these expressions are equivalent:

$$\begin{aligned}\Delta_{m,n} &:= \sup_{t\in\mathbb{R}} |\mathbb{P}(\|\mathbf{Y}\| \leq t) - \mathbb{P}(\|\mathbf{G}\| \leq t)| = \sup_{t\in\mathbb{R}} |\mathbb{P}(\mathbf{Y} \in \mathsf{B}_t) - \mathbb{P}(\mathbf{G} \in \mathsf{B}_t)| \\ &= \sup_{t\in\mathbb{R}} \left|\mathbb{P}\left(\|\mathbf{Y}\|^2 \leq t\right) - \mathbb{P}\left(\chi^2_{m-1} \leq t\right)\right| = \sup_{t\in\mathbb{R}} |\mathbb{P}(\|\mathbf{Y}\| \leq t) - \mathbb{P}(\chi_{m-1} \leq t)|\,.\end{aligned}$$

We can then apply Theorem 1.1 in Bentkus (2003) to the sum $\mathbf{Y}$:

$$\begin{aligned}&\sup_{t\in\mathbb{R}} |\mathbb{P}(\mathbf{Y} \in \mathsf{B}_t) - \mathbb{P}(\mathbf{G} \in \mathsf{B}_t)| \\ &\leq \frac{c\mathbb{E}\left\|(\mathbf{A}\mathbf{X})' (\mathbf{A}\mathbf{\Sigma}\mathbf{A}')^{-\frac{1}{2}}\right\|_2^3}{\sqrt{n}} \leq \frac{c\mathbb{E}\left((\mathbf{A}\mathbf{X})' (\mathbf{A}\mathbf{\Sigma}\mathbf{A}')^{-1} (\mathbf{A}\mathbf{X})\right)^{\frac{3}{2}}}{\sqrt{n}},\end{aligned}$$

where c is an absolute constant. If $m \geq 10$, however, we can apply Theorem 9.2 in Bentkus and Götze (1999). We take $d = m-1$, $\phi_1 \equiv 0$, $q_\ell = 1\{\ell \leq m-1\}, \forall \ell \in \mathbb{N}$, $a_j = 0, \forall j$, and $\beta_k = 0, \forall k > 0$:

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left(\frac{\|\mathbf{Y}\|^2 - \nu}{2} \leq t\right) - \mathbb{P}\left(\frac{\chi^2_{m-1} - (m-1)}{2} \leq t\right)\right| \leq \frac{\exp(c\sigma)\,\gamma_{2,2}}{n\sigma^4}.$$

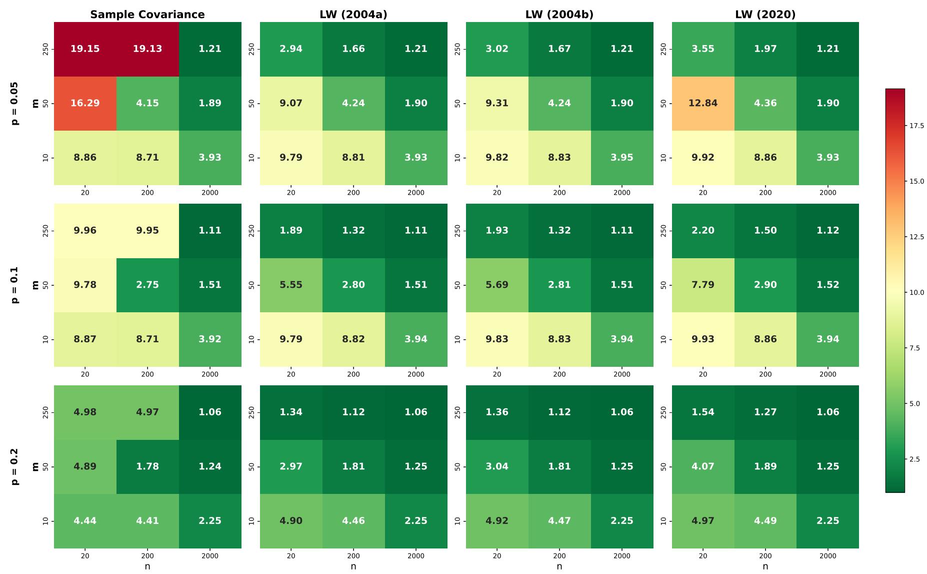

Figure 4: ${}^{\text{MCS size}}/_{m^\star}$ with common variance decreasing linearly from 0.5^2 to 0.2^2 and idiosyncratic variance increasing linearly from 0.2^2 to 0.5^2.

Here,

$$\begin{aligned}
\nu &= \mathbb{E}\left((\mathbf{AX}_1)'\left(\mathbf{A\Sigma A}'\right)^{-1}(\mathbf{AX}_1)\right) = m-1,\\
\gamma_{2,2} &= \mathbb{E}\left(\mathbb{E}\left\{\left[(\mathbf{AX}_1)'\left(\mathbf{A\Sigma A}'\right)^{-1}(\mathbf{AX}_2)\right]^2 |\mathbf{X}_1\right\}\right)^2\\
&= \mathbb{E}\left((\mathbf{AX}_1)'\left(\mathbf{A\Sigma A}'\right)^{-1}(\mathbf{AX}_1)\right)^2,\\
\sigma^2 &= \mathbb{E}\left((\mathbf{AX}_1)'\left(\mathbf{A\Sigma A}'\right)^{-1}(\mathbf{AX}_2)\right)^2\\
&= \mathbb{E}\left((\mathbf{AX}_1)'\left(\mathbf{A\Sigma A}'\right)^{-1}(\mathbf{AX}_1)\right) = m-1.
\end{aligned}$$

At last,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\|\mathbf{Y}\|^2 \le t\right) - \mathbb{P}\left(\chi^2_{m-1} \le t\right)\right| \le \frac{\exp\left(c\sqrt{m-1}\right)}{n\,(m-1)^2}\mathbb{E}\left((\mathbf{AX})'\left(\mathbf{A\Sigma A}'\right)^{-1}(\mathbf{AX})\right)^2.$$

We denote the upper bounds on $\Delta_{m,n}$ as $R_{m,n}$.

(ii) The proof is organized in several steps.

First, we start with a bound in the spirit of Lemma 2, p. 402, in Le Cam (1986), or Lemme 5.3, p. 213, in Foata and Fuchs (1996). Let $\mathsf{B}_t := \{x : \|x\| \le t\}$ be the ball of radius t. We have

$$\begin{aligned}
\{\mathbf{X}\in\mathsf{B}_r\} &= \{\mathbf{X}\in\mathsf{B}_r, \|\mathbf{X}-\mathbf{Y}\|\le\eta\}\cup\{\mathbf{X}\in\mathsf{B}_r, \|\mathbf{X}-\mathbf{Y}\|>\eta\}\\
&\subseteq \{\mathbf{Y}\in\mathsf{B}_{r+\eta}\}\cup\{\|\mathbf{X}-\mathbf{Y}\|>\eta\},\\
\mathbb{P}\{\mathbf{X}\in\mathsf{B}_r\} &\le \mathbb{P}\{\mathbf{Y}\in\mathsf{B}_{r+\eta}\} + \mathbb{P}\{\|\mathbf{X}-\mathbf{Y}\|>\eta\},
\end{aligned}$$

and

$$\mathbb{P}\{\mathbf{Y}\in\mathsf{B}_{r-\eta}\} \le \mathbb{P}\{\mathbf{X}\in\mathsf{B}_r\} + \mathbb{P}\{\|\mathbf{X}-\mathbf{Y}\|>\eta\}.$$

Then,

$$\mathbb{P}\{\mathbf{Y}\in\mathsf{B}_{r-\eta}\} - \mathbb{P}\{\|\mathbf{X}-\mathbf{Y}\|>\eta\} \le \mathbb{P}\{\mathbf{X}\in\mathsf{B}_r\} \le \mathbb{P}\{\mathbf{Y}\in\mathsf{B}_{r+\eta}\} + \mathbb{P}\{\|\mathbf{X}-\mathbf{Y}\|>\eta\}.$$

At the same time,

$$\mathbb{P}\{\mathbf{Y}\in\mathsf{B}_{r-\eta}\} \le \mathbb{P}\{\mathbf{Y}\in\mathsf{B}_r\} \le \mathbb{P}\{\mathbf{Y}\in\mathsf{B}_{r+\eta}\}.$$

Therefore,

$$\begin{aligned}
&-\mathbb{P}\{\mathbf{Y}\in\mathsf{B}_{r+\eta}\backslash\mathsf{B}_{r-\eta}\} - \mathbb{P}\{\|\mathbf{X}-\mathbf{Y}\|>\eta\}\\
&\le \mathbb{P}\{\mathbf{X}\in\mathsf{B}_r\} - \mathbb{P}\{\mathbf{Y}\in\mathsf{B}_r\}\\
&\le \mathbb{P}\{\mathbf{Y}\in\mathsf{B}_{r+\eta}\backslash\mathsf{B}_{r-\eta}\} + \mathbb{P}\{\|\mathbf{X}-\mathbf{Y}\|>\eta\},
\end{aligned}$$

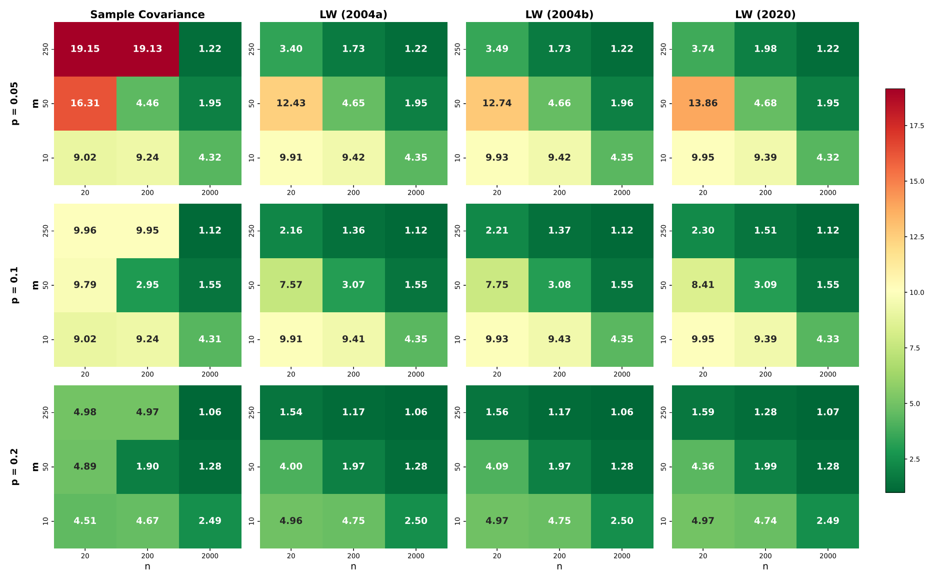

Figure 5: $^{\text{MCS size}}/_{m^\star}$ with common variance 0.2^2 and idiosyncratic variance 0.5^2. The additive error is given by the weighted sum of two centered exponential random variables with changed sign.

and

$$|\mathbb{P}\{\mathbf{X}\in\mathsf{B}_r\}-\mathbb{P}\{\mathbf{Y}\in\mathsf{B}_r\}|\leq\mathbb{P}\{\mathbf{Y}\in\mathsf{B}_{r+\eta}\backslash\mathsf{B}_{r-\eta}\}+\mathbb{P}\{\|\mathbf{X}-\mathbf{Y}\|>\eta\}.$$

Then, we have

$$|\mathbb{P}\{\|\mathbf{X}\|\leq r\}-\mathbb{P}\{\|\mathbf{Y}\|\leq r\}|\leq\mathbb{P}\{\|\mathbf{Y}\|\in(r-\eta,r+\eta]\}+\mathbb{P}\{\|\mathbf{X}-\mathbf{Y}\|>\eta\}. \tag{1}$$

Now, we identify

$$\|\mathbf{X}\|^2=n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)=\left\|\frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n}\left(\mathbf{A}\mathbf{X}_j\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}\right\|^2,$$

$$\|\mathbf{Y}\|^2=n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)=\left\|\frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n}\left(\mathbf{A}\mathbf{X}_j\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-\frac{1}{2}}\right\|^2.$$

Second, we start from the first term in (1):

$$\begin{aligned}\mathbb{P}\{\|\mathbf{Y}\|\in(r-\eta,r+\eta]\}&\leq|\mathbb{P}\{\|\mathbf{Y}\|\in(r-\eta,r+\eta]\}-\mathbb{P}\{\|\mathbf{G}\|\in(r-\eta,r+\eta]\}|\\&\quad+\mathbb{P}\{\|\mathbf{G}\|\in(r-\eta,r+\eta]\}\\&\leq 2R_{m,n}+2\eta\sup_{x\in\mathbb{R}}f_{\chi_{m-1}}(x)\\&\leq 2R_{m,n}+2\eta\frac{(m-2)^{\frac{m-2}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-3}{2}}\Gamma\left(\frac{m-1}{2}\right)}.\end{aligned}$$

Third, we consider the second term. Taking $\lambda:=\left\|\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\|$ and $\mathbf{G}\sim$

Figure 6: $^{\text{MCS size}}/_{m^\star}$ with common variance 0.2^2 and idiosyncratic variance 0.5^2. The additive error is given by the weighted sum of two centered exponential random variables.

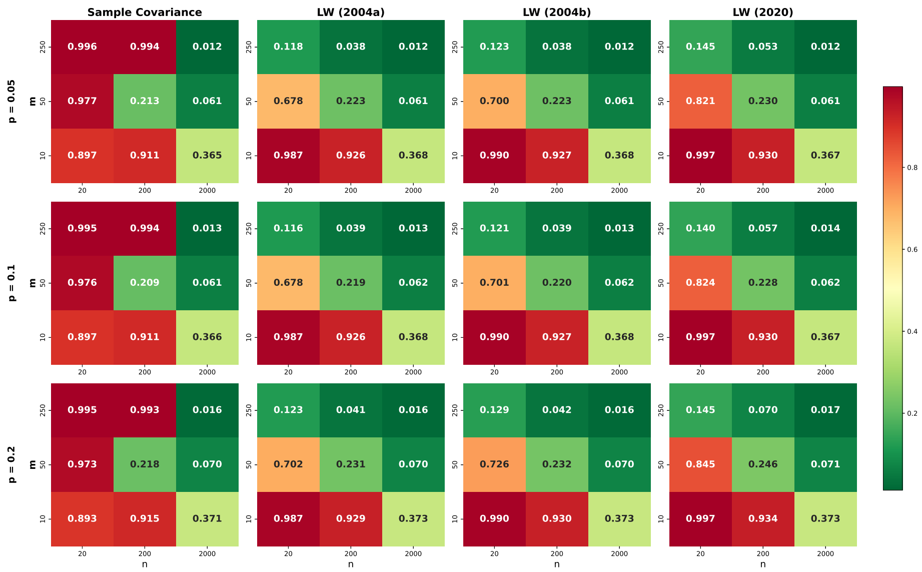

Figure 7: FPR with common variance 0.2^2 and idiosyncratic variance 0.5^2.

$\mathcal{N}(\mathbf{0}, \mathbf{I})$, we write

$$\begin{aligned}
\mathbb{P}\{\|\mathbf{X}-\mathbf{Y}\|>\eta\} &= \mathbb{P}\left\{\left\|\frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n}(\mathbf{A}\mathbf{X}_j)'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{-\frac{1}{2}}\left[(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right]\right\|>\eta\right\} \\
&\le \mathbb{P}\left\{\left\|\frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n}(\mathbf{A}\mathbf{X}_j)'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{-\frac{1}{2}}\right\|>\frac{\eta}{\lambda}\right\} \\
&\le \mathbb{P}\left\{\left\|\left(\mathbf{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{G}\right)'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{-\frac{1}{2}}\right\|>\frac{\eta}{\lambda}\right\}+R_{m,n} \\
&= \mathbb{P}\left\{\left(\mathbf{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{G}\right)'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{-1}\left(\mathbf{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{G}\right)>\frac{\eta^2}{\lambda^2}\right\}+R_{m,n} \\
&= \mathbb{P}\left\{\chi_{m-1}>\frac{\eta}{\lambda}\right\}+R_{m,n}. \qquad (2)
\end{aligned}$$

Figure 8: FPR with common variance 0.5^2 and idiosyncratic variance 0.2^2.

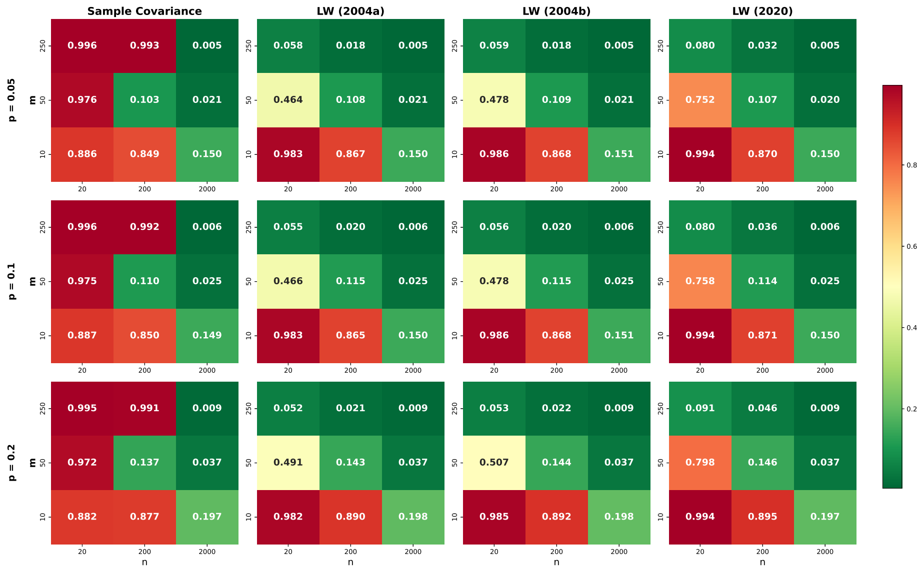

Figure 9: FPR with common variance increasing linearly from 0.2^2 to 0.5^2 and idiosyncratic variance decreasing linearly from 0.5^2 to 0.2^2.

At last, we are left with a bound of the form

$$\Delta_{m,n} \leq 3R_{m,n} + 2\eta \frac{(m-2)^{\frac{m-2}{2}} e^{-\frac{m-2}{2}}}{2^{\frac{m-3}{2}} \Gamma\left(\frac{m-1}{2}\right)} + 1 - F_{\chi_{m-1}}\left(\frac{\eta}{\lambda}\right).$$

We maximize this bound by equating to zero its derivative with respect to η:

$$2\frac{(m-2)^{\frac{m-2}{2}} e^{-\frac{m-2}{2}}}{2^{\frac{m-3}{2}} \Gamma\left(\frac{m-1}{2}\right)} - \frac{1}{\lambda} f_{\chi_{m-1}}\left(\frac{\eta}{\lambda}\right) = 0,$$

$$2\frac{(m-2)^{\frac{m-2}{2}} e^{-\frac{m-2}{2}}}{2^{\frac{m-3}{2}} \Gamma\left(\frac{m-1}{2}\right)} - \frac{1}{2^{\frac{m-3}{2}} \Gamma\left(\frac{m-1}{2}\right)} \frac{\eta^{m-2}}{\lambda^{m-1}} e^{-\frac{\eta^2}{2\lambda^2}} = 0,$$

$$-\frac{(2\lambda)^{\frac{2}{m-2}}}{e} = -\frac{\eta^2}{(m-2)\,\lambda^2} e^{-\frac{\eta^2}{(m-2)\lambda^2}},$$

$$\eta^2 = -(m-2)\,\lambda^2 W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right),$$

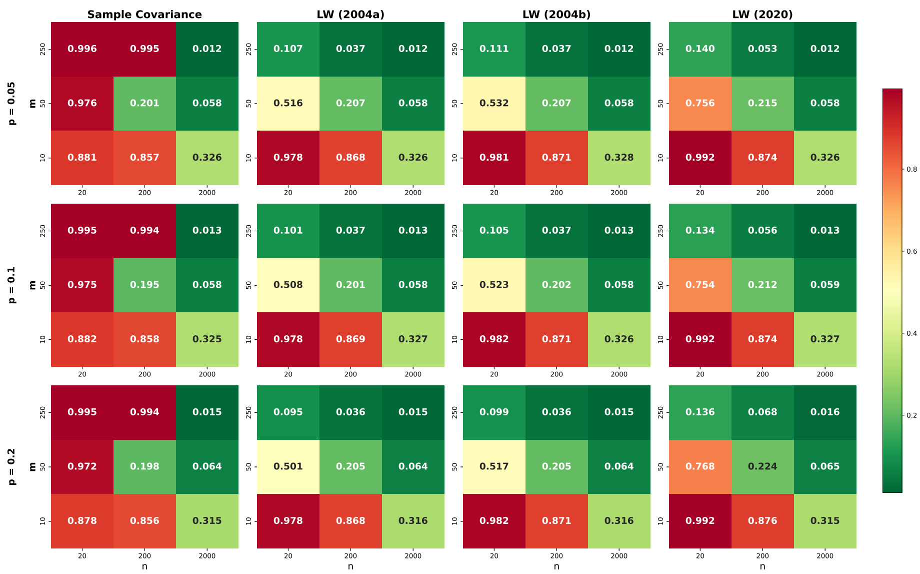

Figure 10: FPR with common variance decreasing linearly from 0.5^2 to 0.2^2 and idiosyncratic variance increasing linearly from 0.2^2 to 0.5^2.

where W_{-1} is the lower branch of the Lambert W function, valid only when its argument is larger than or equal to $-\frac{1}{e}$, or $\lambda \le \frac{1}{2}$. Note that, by using the properties of W_{-1},

$$\eta^2 \simeq -(m-2)\,\lambda^2 \ln\left(\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right) \simeq 2\lambda^2\,|\ln\lambda|\,.$$

The second derivative is

$$\begin{aligned}
&\frac{1}{2^{\frac{m-3}{2}}\Gamma\left(\frac{m-1}{2}\right)}\left[\frac{\eta^2}{\lambda^2}-(m-2)\right]\frac{\eta^{m-3}}{\lambda^{m-1}}e^{-\frac{\eta^2}{2\lambda^2}} \\
&=\frac{(m-2)^{\frac{m-2}{2}}\,e^{-\frac{m-2}{2}}}{2^{\frac{m-5}{2}}\Gamma\left(\frac{m-1}{2}\right)\eta}\left[\frac{\eta^2}{\lambda^2}-(m-2)\right] \\
&=\frac{(m-2)^{\frac{m-2}{2}+1-\frac{1}{2}}\,e^{-\frac{m-2}{2}}}{2^{\frac{m-5}{2}}\Gamma\left(\frac{m-1}{2}\right)\lambda\sqrt{-W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right)}}\left[-W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right)-1\right] \\
&\simeq\frac{(m-2)^{\frac{m-2}{2}}\,e^{-\frac{m-2}{2}}}{2^{\frac{m-6}{2}}\Gamma\left(\frac{m-1}{2}\right)\lambda}\sqrt{|\ln\lambda|}>0,
\end{aligned}$$

confirming that this is indeed a minimum. The final solution is

$$\begin{aligned}
\Delta_{m,n} \le 3R_{m,n} &+ 2\frac{(m-2)^{\frac{m-1}{2}}\,e^{-\frac{m-2}{2}}}{2^{\frac{m-3}{2}}\Gamma\left(\frac{m-1}{2}\right)}\lambda\sqrt{-W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right)} \\
&+\frac{\Gamma\left(\frac{m-1}{2},-\frac{m-2}{2}W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right)\right)}{\Gamma\left(\frac{m-1}{2}\right)}.
\end{aligned}$$

We can investigate the asymptotic behavior using the formulas

$$\begin{aligned}
W_{-1}(x) &\simeq \ln(-x)-\ln(-\ln(-x))\,, \\
\Gamma(a,x) &\simeq x^{a-1}e^{-x},
\end{aligned}$$

from which

$$W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right) \simeq \frac{2}{m-2}\ln(2\lambda)-1-\ln\left(\frac{2}{m-2}\right)-\ln(-\ln\lambda)\,,$$

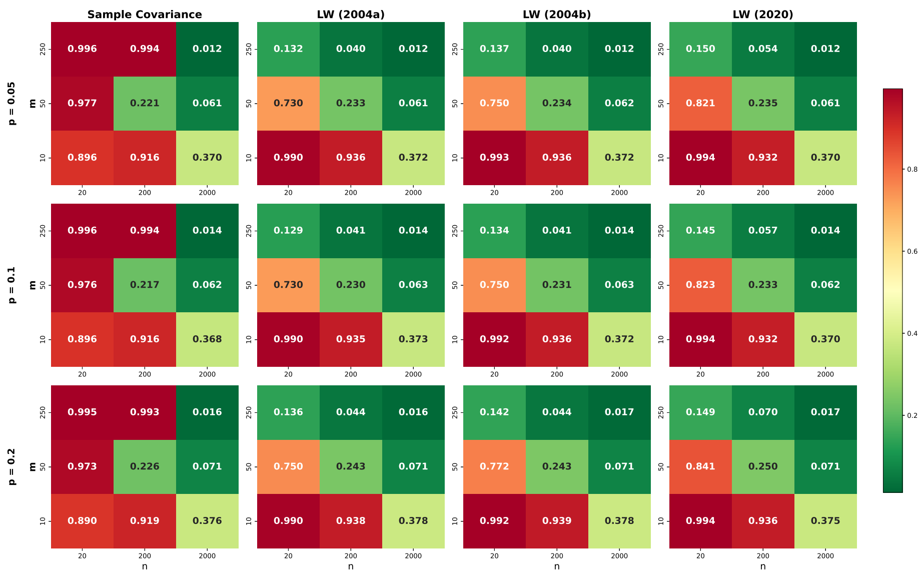

Figure 11: FPR with common variance 0.2^2 and idiosyncratic variance 0.5^2. The additive error is given by the weighted sum of two centered exponential random variables with changed sign.

and

$$\Delta_{m,n} \leq 3R_{m,n} + 2\frac{(m-2)^{\frac{m-1}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-3}{2}}\Gamma\left(\frac{m-1}{2}\right)}\lambda\sqrt{-W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right)}$$
$$+\frac{\Gamma\left(\frac{m-1}{2}, -\frac{m-2}{2}W_{-1}\left(-\frac{(2\lambda)^{\frac{2}{m-2}}}{e}\right)\right)}{\Gamma\left(\frac{m-1}{2}\right)}.$$
$$\simeq 3R_{m,n} + \frac{(m-2)^{\frac{m-2}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-6}{2}}\Gamma\left(\frac{m-1}{2}\right)}\lambda\left|\ln\lambda\right|^{\frac{1}{2}}.$$

(iii) We start from the third step of (i). We notice that

$$\begin{aligned}
&\mathbb{P}\left\{\|\mathbf{X}-\mathbf{Y}\| > \eta\right\} \\
&= \mathbb{P}\left\{\left\|\frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n}(\mathbf{A}\mathbf{X}_j)'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{-\frac{1}{2}}\left[(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right]\right\| > \eta\right\} \\
&\leq \mathbb{P}\left\{\left\|\frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n}(\mathbf{A}\mathbf{X}_j)'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{-\frac{1}{2}}\right\|\left\|(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\| > \frac{\eta}{\mu}\mu\right\} \\
&= \mathbb{P}\left\{\ln\left\|\frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n}(\mathbf{A}\mathbf{X}_j)'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{-\frac{1}{2}}\right\| + \ln\left\|(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\| > \ln\frac{\eta}{\mu}+\ln\mu\right\} \\
&\leq \mathbb{P}\left\{\left\|\frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n}(\mathbf{A}\mathbf{X}_j)'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{-\frac{1}{2}}\right\| > \frac{\eta}{\mu}\right\} + \mathbb{P}\left\{\left\|(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\| > \mu\right\} \\
&\leq \mathbb{P}\left\{\chi_{m-1} > \frac{\eta}{\mu}\right\} + R_{m,n} + \mathbb{P}\left\{\left\|(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\| > \mu\right\}
\end{aligned}$$

where we have followed (2). We majorize the last term through the moment bound (see Philips and Nelson (1995)):

$$\mathbb{P}\left\{\left\|(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\| > \mu\right\} \leq \frac{\mathbb{E}\left\|(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\|^q}{\mu^q}.$$

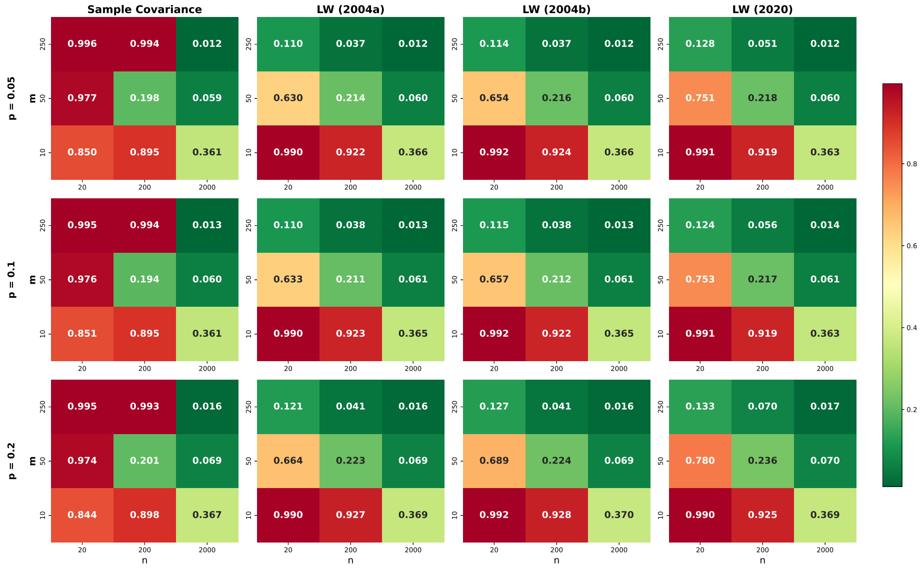

Figure 12: FPR with common variance 0.2^2 and idiosyncratic variance 0.5^2. The additive error is given by the weighted sum of two centered exponential random variables.

We are left with

$$\Delta_{m,n} \leq 3R_{m,n} + 2\eta \frac{(m-2)^{\frac{m-2}{2}} e^{-\frac{m-2}{2}}}{2^{\frac{m-3}{2}}\Gamma\left(\frac{m-1}{2}\right)} + 1 - F_{\chi_{m-1}}\left(\frac{\eta}{\mu}\right) + \frac{\mathbb{E}\left\|(\mathbf{A\Sigma A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}} - \mathbf{I}\right\|^q}{\mu^q}.$$

The first term is majorized as in (i).

(iv) For the fourth result, we use

$$\begin{aligned}
&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\|\mathbf{X}\|^2 \leq t\right) - \mathbb{P}\left(\chi^2_{m-1} \leq t\right)\right| \\
&\leq \sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\|\mathbf{X}\|^2 \leq t\right) - \mathbb{P}\left(\|\mathbf{Y}\|^2 \leq t\right)\right| + \sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\|\mathbf{Y}\|^2 \leq t\right) - \mathbb{P}\left(\chi^2_{m-1} \leq t\right)\right| \\
&\leq \sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\|\mathbf{X}\|^2 \leq t\right) - \mathbb{P}\left(\|\mathbf{Y}\|^2 \leq t\right)\right| + R_{m,n}.
\end{aligned}$$

QED

Proof of Corollary 4. We have

$$\begin{aligned}
\left\|(\mathbf{A\Sigma A'})^{\frac{1}{2}}\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{-\frac{1}{2}}-\mathbf{I}\right\| &\le \left\|(\mathbf{A\Sigma A'})^{\frac{1}{2}}-\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{\frac{1}{2}}\right\|\left\|\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{-\frac{1}{2}}\right\| \\
&\le \left\|(\mathbf{A\Sigma A'})^{\frac{1}{2}}-\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{\frac{1}{2}}\right\|_F\left\|\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{-\frac{1}{2}}\right\| \\
&\le \left\|\mathbf{A\Sigma A'}-\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right\|_*^{\frac{1}{2}}\left\|\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{-\frac{1}{2}}\right\| \\
&= (m-1)^{\frac{1}{4}}\left\|\mathbf{A}\left(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\right)\mathbf{A'}\right\|_F^{\frac{1}{2}}\left\|\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{-\frac{1}{2}}\right\| \\
&\le 2^{\frac{1}{2}}(m-1)^{\frac{3}{4}}\left\|\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\right\|_*^{\frac{1}{2}}\left\|\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{-\frac{1}{2}}\right\| \\
&\le \frac{2^{\frac{1}{2}}(m-1)^{\frac{3}{4}}\left\|\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\right\|_*^{\frac{1}{2}}}{\lambda_{\min}^{\frac{1}{2}}(\mathbf{AA'})\,\lambda_{\min}^{\frac{1}{2}}\left(\widehat{\mathbf{\Sigma}}\right)} \\
&= \frac{2^{\frac{1}{2}}(m-1)^{\frac{3}{4}}\left\|\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\right\|_*^{\frac{1}{2}}}{\lambda_{\min}^{\frac{1}{2}}(\mathbf{AA'})\,\lambda_{\min}^{\frac{1}{2}}\left(\widehat{\mathbf{\Sigma}}\right)} \\
&\le 2^{\frac{1}{2}}(m-1)^{\frac{3}{4}}\left(\frac{\sum_{i=1}^m\left|\lambda_i\left(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\right)\right|}{\lambda_{\min}\left(\widehat{\mathbf{\Sigma}}\right)}\right)^{\frac{1}{2}},
\end{aligned}$$

where the first and the fifth steps come from the sub-additivity of the norms, the second step from the inequality $\|\mathbf{M}\|\le\|\mathbf{M}\|_F$, the third step from the Powers–Størmer inequality, the fourth from the inequality $\|\mathbf{M}\|_*\le\sqrt{\operatorname{rank}(\mathbf{M})}\,\|\mathbf{M}\|_F$, the sixth from the equality $\|\mathbf{A}\|_F=\sqrt{2(m-1)}$ and the inequality $\|\mathbf{M}\|_F\le\|\mathbf{M}\|_*$, the seventh from the definition of spectral norm, the eighth from $\lambda_{\min}(\mathbf{AA'})=1$ and $\|\mathbf{M}\|_*=\sum_{i=1}^m|\lambda_i(\mathbf{M})|$. QED

Proof of Theorem 5. The error in the approximation can be obtained directly from Theorem 2. Taking $q=2$, we have

$$\begin{aligned}
&\left\|(\mathbf{A\Sigma A'})^{\frac{1}{2}}\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{-\frac{1}{2}}-\mathbf{I}\right\|^2 \\
&= \left\|\left(\sigma_1^2\mathbf{U}_{m-1}+\mathbf{\Sigma}_{-1}\right)^{\frac{1}{2}}\left(\widehat{\sigma}_1^2\mathbf{U}_{m-1}+\widehat{\mathbf{\Sigma}}_{-1}\right)^{-\frac{1}{2}}-\mathbf{I}\right\|^2 \\
&\le \left\|\left(\sigma_1^2\mathbf{U}_{m-1}+\mathbf{\Sigma}_{-1}\right)^{\frac{1}{2}}-\left(\widehat{\sigma}_1^2\mathbf{U}_{m-1}+\widehat{\mathbf{\Sigma}}_{-1}\right)^{\frac{1}{2}}\right\|^2\left\|\left(\widehat{\sigma}_1^2\mathbf{U}_{m-1}+\widehat{\mathbf{\Sigma}}_{-1}\right)^{-\frac{1}{2}}\right\|^2 \\
&\le \left\|\left(\sigma_1^2\mathbf{U}_{m-1}+\mathbf{\Sigma}_{-1}\right)^{\frac{1}{2}}-\left(\widehat{\sigma}_1^2\mathbf{U}_{m-1}+\widehat{\mathbf{\Sigma}}_{-1}\right)^{\frac{1}{2}}\right\|_F^2\left\|\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{-\frac{1}{2}}\right\|^2 \\
&\le \frac{\left\|\left(\sigma_1^2\mathbf{U}_{m-1}+\mathbf{\Sigma}_{-1}\right)-\left(\widehat{\sigma}_1^2\mathbf{U}_{m-1}+\widehat{\mathbf{\Sigma}}_{-1}\right)\right\|_*}{\lambda_{\min}\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)} \\
&\le \frac{\left\|\left(\sigma_1^2-\widehat{\sigma}_1^2\right)\mathbf{U}_{m-1}\right\|_*+\left\|\mathbf{\Sigma}_{-1}-\widehat{\mathbf{\Sigma}}_{-1}\right\|_*}{\lambda_{\min}(\mathbf{AA'})\,\lambda_{\min}\left(\widehat{\mathbf{\Sigma}}\right)} \\
&= \frac{\left|\sigma_1^2-\widehat{\sigma}_1^2\right|(m-1)+\sum_{i=2}^m\left|\sigma_i^2-\widehat{\sigma}_i^2\right|}{\lambda_{\min}\left(\widehat{\mathbf{\Sigma}}\right)} = \frac{\left|\sigma_1^2-\widehat{\sigma}_1^2\right|(m-2)+\sum_{i=1}^m\left|\sigma_i^2-\widehat{\sigma}_i^2\right|}{\lambda_{\min}\left(\widehat{\mathbf{\Sigma}}\right)}
\end{aligned}$$

where the first and third steps come from the equalities $\mathbf{A\Sigma A'}=\sigma_1^2\mathbf{U}_{m-1}+\mathbf{\Sigma}_{-1}$ and $\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}=\widehat{\sigma}_1^2\mathbf{U}_{m-1}+\widehat{\mathbf{\Sigma}}_{-1}$, the second step from the sub-additivity of the norm, the fourth step from the Powers–Størmer inequality, the fifth from the triangular inequality, the sixth from the definitions of the norms. As the choice of the first element is arbitrary, the bound should hold as

$$\left\|(\mathbf{A\Sigma A'})^{\frac{1}{2}}\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A'}\right)^{-\frac{1}{2}}-\mathbf{I}\right\|^2\le\frac{\left|\sigma_j^2-\widehat{\sigma}_j^2\right|(m-2)+\sum_{i=1}^m\left|\sigma_i^2-\widehat{\sigma}_i^2\right|}{\lambda_{\min}\left(\widehat{\mathbf{\Sigma}}\right)},$$

for a generic j and, therefore, also for the mean over j:

$$\begin{aligned}&\left\|\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{\frac{1}{2}}\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\|^2\\ &\leq \frac{1}{m}\sum_{j=1}^{m}\frac{\left|\sigma_j^2-\widehat{\sigma}_j^2\right|(m-2)+\sum_{i=1}^m\left|\sigma_i^2-\widehat{\sigma}_i^2\right|}{\lambda_{\min}\left(\widehat{\boldsymbol{\Sigma}}\right)}\\ &=\frac{2(m-1)}{m}\frac{\sum_{i=1}^m\left|\sigma_i^2-\widehat{\sigma}_i^2\right|}{\lambda_{\min}\left(\widehat{\boldsymbol{\Sigma}}\right)}.\end{aligned}$$

Then, we have

$$\begin{aligned}&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|\\ &\lesssim 3R_{m,n}+\frac{(m-2)^{\frac{m-2}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-6}{2}}\Gamma\left(\frac{m-1}{2}\right)}\mu\left|\ln\mu\right|^{\frac{1}{2}}+\frac{2(m-1)}{m\mu^2}\mathbb{E}\frac{\sum_{i=1}^m\left|\sigma_i^2-\widehat{\sigma}_i^2\right|}{\min_{1\leq i\leq m}\widehat{\sigma}_i^2}\end{aligned}$$

where $\sqrt{n}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\to_{\mathcal{D}}\mathcal{N}\left(0,\mu_{4i}-\sigma_i^4\right)$. Under $\sup_n\mathbb{E}\frac{\left|\sigma_i^2-\widehat{\sigma}_i^2\right|^{1+\varepsilon}}{\min_{1\leq i\leq m}\widehat{\sigma}_i^{2+2\varepsilon}}<\infty$,

$$\mathbb{E}\frac{\sqrt{n}\left|\sigma_i^2-\widehat{\sigma}_i^2\right|}{\min_{1\leq i\leq m}\widehat{\sigma}_i^2}\to\frac{\sqrt{\mu_{4i}-\sigma_i^4}\,\mathbb{E}\left|\mathcal{N}(0,1)\right|}{\min_{1\leq i\leq m}\sigma_i^2}=\frac{\sqrt{2\left(\mu_{4i}-\sigma_i^4\right)}}{\sqrt{\pi}\min_{1\leq i\leq m}\sigma_i^2}.$$

At last,

$$\begin{aligned}&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|\\ &\lesssim 3R_{m,n}+\frac{(m-2)^{\frac{m-2}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-6}{2}}\Gamma\left(\frac{m-1}{2}\right)}\mu\left|\ln\mu\right|^{\frac{1}{2}}+\frac{2(m-1)\sum_{i=1}^m\sqrt{2\left(\mu_{4i}-\sigma_i^4\right)}}{\mu^2m\sqrt{\pi n}\min_{1\leq i\leq m}\sigma_i^2}.\end{aligned}$$

Therefore, by equating the two terms depending on μ and solving for μ, we have

$$\begin{aligned}\mu&=\exp\left\{\frac{1}{6}W_0\left(\frac{(m-1)^2 2^{m-2}3\Gamma^2\left(\frac{m-1}{2}\right)\left[\sum_{i=1}^m\sqrt{\mu_{4i}-\sigma_i^4}\right]^2}{m^2(m-2)^{m-2}e^{-m+2}\pi n\min_{1\leq i\leq m}\sigma_i^4}\right)\right\}\\ &\simeq\exp\left\{\frac{(m-1)^2 2^{m-3}\Gamma^2\left(\frac{m-1}{2}\right)\left[\sum_{i=1}^m\sqrt{\mu_{4i}-\sigma_i^4}\right]^2}{m^2(m-2)^{m-2}e^{-m+2}\pi n\min_{1\leq i\leq m}\sigma_i^4}-O\left(n^{-2}\right)\right\}.\end{aligned}$$

Therefore,

$$\begin{aligned}&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|\\ &\lesssim 3R_{m,n}+\frac{4\sqrt{2}(m-1)\sum_{i=1}^m\sqrt{\mu_{4i}-\sigma_i^4}}{m\sqrt{\pi n}\min_{1\leq i\leq m}\sigma_i^2}.\end{aligned}$$

The error is, therefore, $O\left(n^{-\frac{1}{2}}\right)$. If $m\geq 10$,

$$\begin{aligned}&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|\\ &\lesssim\frac{4\sqrt{2}(m-1)\sum_{i=1}^m\sqrt{\mu_{4i}-\sigma_i^4}}{m\sqrt{\pi n}\min_{1\leq i\leq m}\sigma_i^2}.\end{aligned}$$

QED

Proof of Theorem 7. From Theorem 2 with $q=2$ and Corollary 4,

$$\begin{aligned}&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|\\ &\lesssim 3R_{m,n}+\frac{(m-2)^{\frac{m-2}{2}}e^{-\frac{m-2}{2}}}{2^{\frac{m-6}{2}}\Gamma\left(\frac{m-1}{2}\right)}\mu\left|\ln\mu\right|^{\frac{1}{2}}+\frac{2(m-1)^{\frac{3}{2}}}{\mu^2}\mathbb{E}\frac{\sum_{i=1}^m\left|\lambda_i\left(\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\right)\right|}{\lambda_{\min}\left(\widehat{\boldsymbol{\Sigma}}\right)}.\end{aligned}$$

If $\mathbb{E}\frac{\left|\lambda_i\left(\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\right)\right|}{\lambda_{\min}\left(\widehat{\boldsymbol{\Sigma}}\right)} \leq Cn^{-\alpha}m^{\beta}$,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)-\mathbb{P}\left(n\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)'\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\mu}}\right)\leq t\right)\right|$$
$$\lesssim 3R_{m,n}+C_m\mu\left|\ln\mu\right|^{\frac{1}{2}}+\frac{2Cn^{-\alpha}m^{\frac{3}{2}+\beta}}{\mu^2}.$$

We equate the last two terms in parenthesis. If $\mu>1$, we take $|\ln\mu|=\ln\mu$ and the solution is

$$\mu=\exp\left[\frac{1}{6}W_0\left(\frac{96C^2m^{3+2\beta}}{C_m^2n^{2\alpha}}\right)\right]\simeq\exp\left(\frac{16C^2m^{3+2\beta}}{C_m^2n^{2\alpha}}\right),$$

giving

$$3R_{m,n}+8C\frac{m^{\frac{3}{2}+\beta}}{n^{\alpha}}.$$

If $\mu<1$, we take $|\ln\mu|=-\ln\mu$ and we get two solutions. One is

$$\mu=\exp\left[\frac{1}{6}W_0\left(-\frac{96C^2m^{3+2\beta}}{C_m^2n^{2\alpha}}\right)\right]\simeq\exp\left(-\frac{16C^2m^{3+2\beta}}{C_m^2n^{2\alpha}}\right).$$

The bound using this choice of μ has the same convergence to 0 as the one above. The other is

$$\mu=\exp\left[\frac{1}{6}W_{-1}\left(-\frac{96C^2m^{3+2\beta}}{C_m^2n^{2\alpha}}\right)\right]\simeq\left(\frac{48C^2m^{3+2\beta}}{\alpha C_m^2n^{2\alpha}}\right)^{\frac{1}{6}}\left(\ln n\right)^{-\frac{1}{6}},$$

where we have used $W_{-1}(x)\simeq\ln(-x)-\ln(-\ln(-x))$. This leads to

$$3R_{m,n}+\frac{2^{\frac{2}{3}}\alpha^{\frac{1}{3}}C_m^{\frac{2}{3}}C^{\frac{1}{3}}m^{\frac{3+2\beta}{6}}}{3n^{\frac{\alpha}{3}}}\left(\ln n\right)^{\frac{1}{3}},$$

that is slower. QED

Proof of Remark 8. In the first part of this proof, we use the notation of Ledoit and Wolf (2004a). From Ledoit and Wolf (2004a, p. 367), $\|\mathbf{A}\|_n^2=\operatorname{tr}\left(\mathbf{A}\mathbf{A}'\right)/m$. From Theorem 3.2 in the same source,

$$\mathbb{E}\left\|S_n^*-\Sigma_n\right\|_n^2-\mathbb{E}\left\|\Sigma_n^*-\Sigma_n\right\|_n^2\to 0.$$

By Theorem 2.1 in Ledoit and Wolf (2004a), $\mathbb{E}\left\|\Sigma_n^*-\Sigma_n\right\|_n^2=\frac{\alpha_n^2\beta_n^2}{\delta_n^2}$. Using Section 4.1 in Oriol and Miot (2025) as a guide, we have

$$\begin{aligned}
\mu_n&=\frac{\operatorname{tr}\left(\Sigma_n\right)}{m_n}\asymp 1,\\
\alpha_n^2&=\left\|\Sigma_n-\mu_nI_n\right\|_n^2=\left\|\Sigma_n\right\|_n^2-\mu_n^2\asymp m,\\
\beta_n^2&=\mathbb{E}\left\|\Sigma_n-S_n\right\|_n^2=\frac{m+1}{n-1}\mu_n^2+\frac{1}{n-1}\alpha_n^2\asymp\frac{m}{n},\\
\delta_n^2&=\mathbb{E}\left\|S_n-\mu_nI_n\right\|_n^2=\alpha_n^2+\beta_n^2\asymp m+\frac{m}{n}.
\end{aligned}$$

As a result,

$$\mathbb{E}\left\|S_n^*-\Sigma_n\right\|_n^2\simeq\mathbb{E}\left\|\Sigma_n^*-\Sigma_n\right\|_n^2=\frac{\alpha_n^2\beta_n^2}{\delta_n^2}\asymp\frac{m}{n}.$$

Using our notation with the usual Frobenius norm,

$$\left\|\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\right\|_F=O_{\mathbb{P}}\left(\frac{m}{\sqrt{n}}\right).$$

On the other hand,

$$\begin{aligned}\left\|(\mathbf{A\Sigma A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\| &\leq \left\|\mathbf{A}\left(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\right)\mathbf{A}'\right\|_*^{\frac{1}{2}}\left\|\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}\right\| \\ &\leq (m-1)^{\frac{1}{4}}\left\|\mathbf{A}\left(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\right)\mathbf{A}'\right\|_F^{\frac{1}{2}}\left\|\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}\right\| \\ &\leq \frac{(m-1)^{\frac{1}{4}}\left\|\mathbf{A}\left(\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\right)\mathbf{A}'\right\|_F^{\frac{1}{2}}}{\lambda_{\min}^{\frac{1}{2}}\left(\widehat{\mathbf{\Sigma}}\right)} \\ &\leq \frac{\sqrt{2}\,(m-1)^{\frac{3}{4}}\left\|\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\right\|_F^{\frac{1}{2}}}{\lambda_{\min}^{\frac{1}{2}}\left(\widehat{\mathbf{\Sigma}}\right)},\end{aligned}$$

where the first step derives from the proof of Corollary 4, the second step uses the inequality $\|\mathbf{M}\|_* \leq \sqrt{\operatorname{rank}(\mathbf{M})}\,\|\mathbf{M}\|_F$, the third comes from standard properties of the spectral norm, the fourth from the subadditivity of the nuclear norm and from the fact that $\|\mathbf{A}\|_F = \sqrt{2\,(m-1)}$. As a result,

$$\left\|(\mathbf{A\Sigma A}')^{\frac{1}{2}}\left(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}'\right)^{-\frac{1}{2}}-\mathbf{I}\right\| = O\left(\frac{m^{\frac{5}{4}}}{n^{\frac{1}{4}}}\right).$$

The rest of the proof follows the one of Theorem 7. QED

Acknowledgments
Funding from the MUR (Italian Ministry of University and Research) Project 'Dipartimenti d'Eccellenza 2023-2027' is gratefully acknowledged.

Disclosure statement:
No potential conflict of interest is reported by the authors.

References

Barde, S. (2025). "Large-scale model comparison with fast model confidence sets", *Journal of Econometrics*, Vol. 253, p. 106123. DOI: 10.1016/j.jeconom.2025.106123.

Bentkus, V. (2003). "On the dependence of the Berry-Esseen bound on dimension", *Journal of Statistical Planning and Inference*, Vol. 113 No. 2, pp. 385–402. DOI: 10.1016/S0378-3758(02)00094-0.

Bentkus, V. and Götze, F. (1999). "Optimal Bounds in Non-Gaussian Limit Theorems for *U*-Statistics", *The Annals of Probability*, Vol. 27 No. 1, pp. 454–521. DOI: 10.1214/aop/1022677269.

Choirat, C. and Seri, R. (2012). "Estimation in Discrete Parameter Models", EN. *Statistical Science*, Vol. 27 No. 2, pp. 278–293. DOI: 10.1214/11-STS371.

— (2014). "Bootstrap Confidence Sets for the Aumann Mean of a Random Closed Set", *Computational Statistics and Data Analysis*, Vol. 71 No. SI, pp. 803–817. DOI: 10.1016/j.csda.2012.10.015.

Esséen, C.-G. (1945). "Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law", *Acta Mathematica*, Vol. 77, pp. 1–125. DOI: 10.1007/BF02392223.

Fan, J., Liao, Y., and Liu, H. (2016). "An overview of the estimation of large covariance and precision matrices", *The Econometrics Journal*, Vol. 19 No. 1, pp. C1–C32. DOI: 10.1111/ectj.12061.

Foata, D. and Fuchs, A. (1996). *Calcul des Probabilités*, Masson.

Futschik, A. and Pflug, G. (1995). "Confidence sets for discrete stochastic optimization", *Annals of Operations Research*, Vol. 56 No. 1, pp. 95–108. DOI: 10.1007/BF02031702.

— (1997). "Optimal allocation of simulation experiments in discrete stochastic optimization and approximative algorithms", *European Journal of Operational Research*, Vol. 101 No. 2, pp. 245–260. DOI: 10.1016/S0377-2217(96)00396-7.

Götze, F. and Ulyanov, V. (2003). *Asymptotic Distribution of χ^2-type Statistics*, tech. rep. No. 03-033. Research Group "Spectral Analysis, Asymptotic Distributions and Stochastic Dynamics". Bielefeld University.

Gupta, S. S. (1965). "On Some Multiple Decision (Selection and Ranking) Rules", *Technometrics*, Vol. 7 No. 2, pp. 225–245. DOI: 10.1080/00401706.1965.10490251.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). "The Model Confidence Set", *Econometrica*, Vol. 79 No. 2, pp. 453–497. DOI: 10.3982/ECTA5771.

Hess, C. and Seri, R. (2019). "Generic Consistency for Approximate Stochastic Programming and Statistical Problems", en. *SIAM Journal on Optimization*, Vol. 29 No. 1, pp. 290–317. DOI: 10.1137/17M1156769.

Jankowski, H. and Stanberry, L. (2012). "Confidence Regions for Means of Random Sets Using Oriented Distance Functions", *Scandinavian Journal of Statistics*, Vol. 39 No. 2, pp. 340–357. DOI: 10.1111/j.1467-9469.2011.00753.x.

Jensen, D. R. (1977). "On approximating the distributions of Friedman's χ_r^2 and related statistics", *Metrika*, Vol. 24 No. 1, pp. 75–85. DOI: 10.1007/BF01893394.

Kleywegt, A. J., Shapiro, A., and Homem-de-Mello, T. (2002). "The Sample Average Approximation Method for Stochastic Discrete Optimization", *SIAM Journal on Optimization*, Vol. 12 No. 2, pp. 479–502. DOI: 10.1137/S1052623499363220.

Le Cam, L. M. (1986). *Asymptotic methods in statistical decision theory*, Springer Series in Statistics. Springer-Verlag, pp. xxvi+742. DOI: 10.1007/978-1-4612-4946-7.

Ledoit, O. and Wolf, M. (2003). "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection", *Journal of Empirical Finance*, Vol. 10 No. 5, pp. 603–621. DOI: 10.1016/S0927-5398(03)00007-0.

— (2004a). "A well-conditioned estimator for large-dimensional covariance matrices", *Journal of Multivariate Analysis*, Vol. 88 No. 2, pp. 365–411. DOI: 10.1016/S0047-259X(03)00096-4.

— (2004b). "Honey, I Shrunk the Sample Covariance Matrix", *The Journal of Portfolio Management*, Vol. 30 No. 4, pp. 110–119. DOI: 10.3905/jpm.2004.110.

— (2020). "Analytical Nonlinear Shrinkage of Large-Dimensional Covariance Matrices", *The Annals of Statistics*, Vol. 48 No. 5, pp. 3043–3065. DOI: 10.1214/19-AOS1921.

Martinoli, M., Moneta, A., and Pallante, G. (2024). "Calibration and validation of macroeconomic simulation models by statistical causal search", *Journal of Economic Behavior & Organization*, Vol. 228, p. 106786. DOI: 10.1016/j.jebo.2024.106786.

Oriol, B. and Miot, A. (2025). "Ledoit-Wolf linear shrinkage with unknown mean", *Journal of Multivariate Analysis*, Vol. 208, p. 105429. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2025.105429.

Philips, T. K. and Nelson, R. (1995). "The Moment Bound is Tighter than Chernoff's Bound for Positive Tail Probabilities", *The American Statistician*, Vol. 49 No. 2, pp. 175–178. DOI: 10.1080/00031305.1995.10476137.

Seri, R. and Choirat, C. (2004). "Confidence Sets for the Aumann Mean of a Random Closed Set", *Computational Science and Its Applications – ICCSA 2004*. Ed. by Laganá, A. *et al.*, Springer Berlin Heidelberg: Berlin, Heidelberg, pp. 298–307. DOI: 10.1007/978-3-540-24767-8_31.

Seri, R. *et al.*, (2021). "Model calibration and validation via confidence sets", *Econometrics and Statistics*, Vol. 20, pp. 62–86. DOI: 10.1016/j.ecosta.2020.01.001.