A Robust Statistics Approach to Minimum Variance Portfolio Optimization

Liusha Yang, Romain Couillet, and Matthew R. McKay

Abstract—We study the design of portfolios under a minimum risk criterion. The performance of the optimized portfolio relies on the accuracy of the estimated covariance matrix of the portfolio asset returns. For large portfolios, the number of available market returns is often of similar order to the number of assets, so that the sample covariance matrix performs poorly as a covariance estimator. Additionally, financial market data often contain outliers which, if not correctly handled, may further corrupt the covariance estimation. We address these shortcomings by studying the performance of a hybrid covariance matrix estimator based on Tyler's robust M-estimator and on Ledoit-Wolf's shrinkage estimator while assuming samples with heavy-tailed distribution. Employing recent results from random matrix theory, we develop a consistent estimator of (a scaled version of) the realized portfolio risk, which is minimized by optimizing online the shrinkage intensity. Our portfolio optimization method is shown via simulations to outperform existing methods both for synthetic and real market data.

Index Terms—Minimum variance portfolio, Tyler's M-estimator, shrinkage estimator, random matrix theory.

I. INTRODUCTION

THE theory of portfolio optimization is generally associated with the classical mean-variance optimization framework of Markowitz [1]. The pitfalls of the mean-variance analysis are mainly related to its sensitivity to the estimation error of the means and covariance matrix of the asset returns. It is nonetheless argued that estimates of the covariance matrix are more accurate than those of the expected returns [2], [3]. Thus, many studies concentrate on improving the performance of the global minimum variance portfolio (GMVP), which provides the lowest possible portfolio risk and involves only the covariance matrix estimate.

The frequently used covariance estimator is the well-known sample covariance matrix (SCM). However, covariance esti-mates for portfolio optimization commonly involve few historical observations of sometimes up to a thousand assets. In such a case, the number of independent samples n may be small compared to the covariance matrix dimension N, which suggests a poor performance of the SCM. The impact of the estimation error on the out-of-sample performance of the GMVP based on the SCM has already been analyzed in [4]–[7].

In the finance literature, several approaches have been proposed to get around the problem of the scarcity of samples. One approach is to impose some factor structure on the estimator of the covariance matrix [8], [9], which reduces the number of parameters to be estimated. A second approach is to use as a covariance matrix estimator a weighted average of the sample covariance matrix and another estimator, such as the 1-factor covariance matrix or the identity matrix [10], [11]. A third approach is a nonlinear shrinkage estimation approach [12], which modifies each eigenvalue of the SCM under the framework of Markowitz's portfolio selection. A fourth approach comprises eigenvalue clipping methods [13]–[15] whose underlying idea is to 'clean' the SCM by filtering noisy eigenvalues claimed to convey little valuable information. This approach has also been employed recently in proposing novel vaccine design strategies for infectious diseases [16], [17], and its theoretical foundations have been examined in [18]. A fifth method employs a bootstrap-corrected estimator for the optimal return and its asset allocation, which reduces the error of over-prediction of the in-sample return by bootstrapping [6]. In contrast to all of these methods (which aim to improve the covariance matrix estimate), alternative methods have also been proposed which directly impose various constraints on the portfolio weights, such as a no-shortsale constraint [3], a L_1 norm constraint and a L_2 norm constraint [19], [20]. By bounding directly the portfolio-weight vector, it is demonstrated that the estimation error can be reduced, particularly when the portfolio size is large [19].

In addition to the problem of sample deficiency, it is often the case that the return observations exhibit impulsiveness and local loss of stationarity [21], which is not addressed by the methods mentioned above and leads to performance degradation. The field of robust estimation [22]–[25] intends to deal with this problem. However, classical robust covariance estimators generally require $n \gg N$ and do not perform well (or are not even defined) when $n \simeq N$, making them unsuitable for many modern applications. Recent works [26]–[32] based on random matrix theory have therefore considered robust estimation in the $n \simeq N$ regime. Two hybrid robust shrinkage covariance matrix estimates have been proposed in parallel in [29], [30] and in [31], respectively, both of which estimators are built upon Tyler's robust M-estimatior [23] and Ledoit-Wolf's

L. Yang and M. R. McKay are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: lyangag@connect.ust.hk; eemckay@ust.hk).

R. Couillet is with the Laboratoire de Signaux et Systmes (L2S, UMR8506), CNRS-CentraleSupélec-Université Paris-Sud, 91192 Gif-sur-Yvette, France (e-mail: romain.couillet@centralesupelec.fr).

shrinkage approach [11]. In [32], the authors show, by means of random matrix theory, that in the large n, N regime and under the assumption of elliptical vector observations, the estimators in [29], [30] and [31] perform essentially the same and can be analyzed thanks to their asymptotic closeness to well-known random matrix models. Therefore, in this paper, we concentrate on the estimator studied in [29], [30], which we denote by $\hat{\mathbf{C}}_{\text{ST}}$ (ST standing for shrinkage Tyler). Namely, for independent samples $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^N$ with zero mean, $\hat{\mathbf{C}}_{\text{ST}}$ is the unique solution to the fixed-point equation

$$\hat{\mathbf{C}}_{\text{ST}}(\rho) = (1 - \rho) \frac{1}{n} \sum_{t=1}^{n} \frac{\mathbf{x}_t \mathbf{x}_t^T}{\frac{1}{N} \mathbf{x}_t^T \hat{\mathbf{C}}_{\text{ST}}^{-1}(\rho) \mathbf{x}_t} + \rho \mathbf{I}_N$$

for any $\rho \in (\max\{1 - \frac{N}{n}, 0\}, 1]$. It should be noted that the shrinkage structure even allows $n < N$.

This paper designs a novel minimum variance portfolio optimization strategy based on $\hat{\mathbf{C}}_{\text{ST}}$ with a risk-minimizing (instead of Frobenius norm minimizing as in [32]) shrinkage parameter ρ. This provides an advance over recent techniques, such as the GMVP optimization strategy in [33] that is based on the non-robust shrinkage covariance estimator with a risk-minimizing shrinkage parameter ρ. We first characterize the out-of-sample risk of the minimum variance portfolio with plug-in ST for all ρ within a specified range. This is done by analyzing the uniform convergence of the achieved realized risk on ρ in the double limit regime, where $N, n \to \infty$, with $c_N = N/n \to c \in (0, \infty)$. We subsequently provide a consistent estimator of the realized portfolio risk (or, more precisely, a scaled version of it) that is defined only in terms of the observed returns. Based on this, we obtain a risk-optimized ST covariance estimator by optimizing online over ρ, and thus our optimized portfolio.

The proposed portfolio selection, which as a consequence is the first to integrate both robustness and risk minimal objective, is shown to achieve superior performance over the competing methods in [11], [31]–[33] in minimizing the realized portfolio risk under the GMVP framework for impulsive data. The outperformance of our portfolio optimization strategy compared to other methods is demonstrated through Monte Carlo simulations with elliptically distributed samples, as well as with real data of historical (daily) stock returns from Hong Kong's Hang Seng Index (HSI).

Notations

Boldface upper case letters denote matrices, boldface lower case letters denote column vectors, and standard lower case letters denote scalars. $(\cdot)^T$ denotes transpose. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix and $\mathbf{1}_N$ denotes an N-dimensional vector with all entries equal to one. $\text{tr}[\cdot]$ denotes the matrix trace operator. $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex fields of dimension specified by a superscript. $\|\cdot\|$ denotes the Euclidean norm for vectors and the spectral norm for matrices. The Dirac measure at point x is denoted by $\boldsymbol{\delta}_x$. The ordered eigenvalues of a symmetric matrix $\mathbf{X}$ of size $N \times N$ are denoted by $\lambda_1(\mathbf{X}) \leq \cdots \leq \lambda_N(\mathbf{X})$, and the cardinality of a set $\mathcal{C} \subset \mathbb{R}$ is denoted by $|\mathcal{C}|$. Letting $\mathbf{U}, \mathbf{V}$ be symmetric $N \times N$ matrices, we write $\mathbf{U} \succcurlyeq \mathbf{V}$ if $\mathbf{U} - \mathbf{V}$ is positive semidefinite.

II. DATA MODEL AND PROBLEM FORMULATION

We consider a time series comprising $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^N$ logarithmic returns of N financial assets. We assume the $\mathbf{x}_t$ to be independent and identically distributed (i.i.d.) with

$$\mathbf{x}_t = \boldsymbol{\mu} + \sqrt{\tau_t} \mathbf{C}_N^{1/2} \mathbf{y}_t, \quad t = 1, 2, \ldots, n, \quad (1)$$

where $\boldsymbol{\mu} \in \mathbb{R}^N$ is the mean vector of the asset returns, τ_t is a real, positive random variable, $\mathbf{C}_N \in \mathbb{R}^{N \times N}$ is positive definite and $\mathbf{y}_t \in \mathbb{R}^N$ is a zero mean unitarily invariant random vector with norm $\|\mathbf{y}_t\|^2 = N$, independent of the τ_i's. It is assumed that $\boldsymbol{\mu}$ and $\mathbf{C}_N$ are time-invariant over the observation period. Denote $\mathbf{z}_t = \mathbf{C}_N^{1/2} \mathbf{y}_t$. The model (1) for $\mathbf{x}_t$ embraces in particular the class of elliptical distributions, including the multivariate normal distribution, exponential distribution and the multivariate Student-T distribution as special cases. This model for $\mathbf{x}_t$ leads to tractable and adoptable design solutions and is a commonly used approximation of the impulsive nature of financial data [10].

Let $\mathbf{h} \in \mathbb{R}^N$ denote the portfolio selection, i.e., the vector of asset holdings in units of currency normalized by the total outstanding wealth, satisfying $\mathbf{h}^T \mathbf{1}_N = 1$. In this paper, short-selling is allowed, and thus the portfolio weights may be negative. Then the portfolio variance (or risk) over the investment period of interest is defined as $\sigma^2(\mathbf{h}) = E[|\mathbf{h}^T \mathbf{x}_t|^2] = \mathbf{h}^T \mathbf{C}_N \mathbf{h}$ [1]. Accordingly, the GMVP selection problem can be formulated as the following quadratic optimization problem with a linear constraint:

$$\min_{\mathbf{h}} \quad \sigma^2(\mathbf{h}) \quad \text{s.t.} \quad \mathbf{h}^T \mathbf{1}_N = 1.$$

This has the well-known solution

$$\mathbf{h}_{\text{GMVP}} = \frac{\mathbf{C}_N^{-1} \mathbf{1}_N}{\mathbf{1}_N^T \mathbf{C}_N^{-1} \mathbf{1}_N}$$

and the corresponding portfolio risk is

$$\sigma^2(\mathbf{h}_{\text{GMVP}}) = \frac{1}{\mathbf{1}_N^T \mathbf{C}_N^{-1} \mathbf{1}_N}. \quad (2)$$

Here, (2) represents the theoretical minimum portfolio risk bound, attained upon knowing the covariance matrix $\mathbf{C}_N$ exactly. In practice, $\mathbf{C}_N$ is unknown, and instead we form an estimate, denoted by $\hat{\mathbf{C}}_N$. Thus, the GMVP selection based on the plug-in estimator $\hat{\mathbf{C}}_N$ is given by

$$\hat{\mathbf{h}}_{\text{GMVP}} = \frac{\hat{\mathbf{C}}_N^{-1} \mathbf{1}_N}{\mathbf{1}_N^T \hat{\mathbf{C}}_N^{-1} \mathbf{1}_N}.$$

The quality of $\hat{\mathbf{h}}_{\text{GMVP}}$, implemented based on the in-sample covariance prediction $\hat{\mathbf{C}}_N$, can be measured by its achieved out-of-sample (or "realized") portfolio risk:

$$\sigma^2(\hat{\mathbf{h}}_{\text{GMVP}}) = \frac{\mathbf{1}_N^T \hat{\mathbf{C}}_N^{-1} \mathbf{C}_N \hat{\mathbf{C}}_N^{-1} \mathbf{1}_N}{(\mathbf{1}_N^T \hat{\mathbf{C}}_N^{-1} \mathbf{1}_N)^2}.$$

The goal is to construct a good estimator $\hat{\mathbf{C}}_N$, and consequently $\hat{\mathbf{h}}_{\text{GMVP}}$, which minimizes this quantity.

Note that, for the naive uniform diversification rule, $\mathbf{h} = \frac{1}{N}\mathbf{1}_N$. This is equivalent to setting $\hat{\mathbf{C}}_N = \mathbf{I}_N$, and yields the realized portfolio risk: $\frac{\mathbf{1}_N^T \mathbf{C}_N \mathbf{1}_N}{N^2}$. Interestingly, this extremely simple strategy has been shown in [34] to outperform numerous optimized models and will serve as a benchmark in our work.

III. Novel Covariance Estimator and Portfolio Design for Minimizing Risk

A. Tyler's Robust M-Estimator With Linear Shrinkage

Consider the ST covariance matrix estimate introduced in [29], [30], built upon both Tyler's M-estimate [23] and the Ledoit-Wolf shrinkage estimator [11]. This estimator accounts for the scarcity of samples, even allowing $N > n$, and exhibits robustness to outliers or impulsive samples, e.g., elliptically distributed data. It is defined as the unique solution to the following fixed-point equation for $\rho \in (\max\{0, 1 - n/N\}, 1]$:

$$\hat{\mathbf{C}}_{\mathrm{ST}}(\rho) = (1-\rho)\frac{1}{n}\sum_{t=1}^{n}\frac{\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^T}{\frac{1}{N}\tilde{\mathbf{x}}_t^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\tilde{\mathbf{x}}_t} + \rho\mathbf{I}_N \qquad (3)$$

where $\tilde{\mathbf{x}}_t = \mathbf{x}_t - \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$.

Since with probability one the $\mathbf{x}_t$ are linearly independent, $\hat{\mathbf{C}}_{\mathrm{ST}}(\rho)$ is almost surely defined for each N and n [29, Theorem III.1]. The corresponding GMVP selection is

$$\hat{\mathbf{h}}_{\mathrm{ST}}(\rho) = \frac{\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{1}_N}{\mathbf{1}_N^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{1}_N}$$

with realized portfolio risk

$$\sigma^2\left(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)\right) = \frac{\mathbf{1}_N^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{C}_N \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{1}_N}{(\mathbf{1}_N^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{1}_N)^2}. \qquad (4)$$

Our goal is to optimize ρ online such that (4) is minimum. However, since (4) involves $\mathbf{C}_N$ which is unobservable, this equation cannot be optimized directly. We tackle this problem by obtaining a consistent estimator for a scaled version of the realized risk (4) as n and N go to infinity at the same rate. Contrary to classical asymptotic theory for time series analysis and mathematical statistics, which typically deal with the case of N fixed and $n \to \infty$, a double-limiting condition is of more relevance for large portfolio problems, where n is comparable to N. To this end, following [33], we first derive a deterministic asymptotic equivalent of (4) and then provide a consistent estimator based on this.

B. Deterministic Equivalent of the Realized Portfolio Risk

For our asymptotic analysis, we assume the following:
Assumption 1:
a. As $N, n \to \infty$, $N/n = c_N \to c \in (0, \infty)$.
b. The τ_t, $t = 1, \ldots, n$ are i.i.d. $\tau_1, \ldots, \tau_n \geq \xi$ a.s. for some $\xi > 0$ and $E[\tau_1] < \infty$.[1]
c. Denoting $0 < \lambda_1 \leq \cdots \leq \lambda_N$ the ordered eigenvalues of $\mathbf{C}_N$, as $N, n \to \infty$, $\nu_N \triangleq \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\delta}_{\lambda_i}$ satisfies $\nu_N \to$

[1]For technical reasons, made explicit in the appendix, we require the quantities $\tilde{\mathbf{z}}_t = \mathbf{z}_t - \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\sqrt{\frac{\tau_i}{\tau_t}}$ to have controllable norms. This imposes the constraint $\tau_t \geq \xi > 0$ which might be possible to relax at the expense of increased mathematical complexity.

ν weakly with $\nu \neq \boldsymbol{\delta}_0$ almost everywhere. In addition, $\limsup_N \lambda_N < \infty$.

We also introduce some further definitions, which will arise in our asymptotic analysis. For $\rho \in (\max(0, 1 - c^{-1}), 1]$, define γ the unique positive solution to

$$1 = \int \frac{t}{\gamma\rho + (1-\rho)t}\nu(dt) \qquad (5)$$

and

$$\beta = \int \frac{c\gamma^2 t^2}{(\gamma\rho + (1-\rho)t)^2}\nu(dt).$$

The following theorem presents our first key result: a deterministic characterization of the asymptotic realized portfolio risk achieved with $\hat{\mathbf{C}}_{\mathrm{ST}}(\rho)$.

Theorem 1: Let Assumption 1 hold. For $\varepsilon \in (0, \min\{1, c^{-1}\})$, define $\mathcal{R}_\varepsilon = [\varepsilon + \max\{0, 1 - c^{-1}\}, 1]$. Then, as $N, n \to \infty$,

$$\sup_{\rho \in \mathcal{R}_\varepsilon}\left|N\sigma^2\left(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)\right) - N\bar{\sigma}^2(\rho)\right| \xrightarrow{\text{a.s.}} 0 \qquad (6)$$

where

$$\bar{\sigma}^2(\rho) = \frac{\gamma^2}{\gamma^2 - \beta(1-\rho)^2} \times$$
$$\frac{\mathbf{1}_N^T\left(\frac{1-\rho}{\gamma}\mathbf{C}_N + \rho\mathbf{I}_N\right)^{-1}\mathbf{C}_N\left(\frac{1-\rho}{\gamma}\mathbf{C}_N + \rho\mathbf{I}_N\right)^{-1}\mathbf{1}_N}{\left(\mathbf{1}_N^T\left(\frac{1-\rho}{\gamma}\mathbf{C}_N + \rho\mathbf{I}_N\right)^{-1}\mathbf{1}_N\right)^2}.$$

Proof: The proof relies on a conjunction of methods [27] and [33] with an additional technical difficulty introduced by re-centering the data via subtracting their sample mean. Details are provided in Appendix B.

Remark 1: In Theorem 1, the set $\mathcal{R}_\varepsilon$ excludes the region $[0, \varepsilon + \max\{0, 1 - c^{-1}\})$. As we handle the uniformity of the convergence (6), the proof of Theorem 1 requires us to work on sequences $\{\rho_n\}_{n=1}^{\infty}$ of ρ. It is however difficult to handle the limit $\left|N\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho_n)) - N\bar{\sigma}^2(\rho_n)\right|$ for a sequence $\{\rho_n\}_{n=1}^{\infty}$ with $\rho_n \to 0$. This follows from the same reasoning as that in [32] (see (5) and (6) in Section 5.1 of [32] as well as (12) in Appendix A where $\rho_n \to \rho_0 > 0$ is necessary to ensure $e^+ < 1$). In the subsequent results, $\rho \in \mathcal{R}_\varepsilon$ is also required for the same reason.

Theorem 1 enables us to analyze the convergence of the realized portfolio risk in the regime of Assumption 1–a for $\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)$. In order to calibrate the shrinkage parameter ρ for optimum GMVP performance, only the available sample data and certainly not the unknown $\mathbf{C}_N$ can be used. This is the objective of the subsequent section.

C. Consistent Estimation of Scaled Realized Portfolio Risk

Based on the observable data only, we can obtain an estimator of a scaled version of the realized portfolio risk, $\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho))/\kappa$, where we define $\kappa \triangleq \int t\nu(dt)$. We begin with the following lemma that provides a consistent estimator of γ, scaled by $1/\kappa$, which is denoted as $\hat{\gamma}_{\mathrm{sc}}$ ("sc" standing for "scaled").

Algorithm 1: Proposed algorithm for GMVP optimization

1) Compute the optimized shrinkage parameter via a numerical search

$$\rho^o = \underset{\rho \in [\varepsilon + \max\{0, 1 - c_N^{-1}\}, 1]}{\arg\min} \left\{ \hat{\sigma}_{\text{sc}}^2(\rho) \right\}.$$

2) Form the risk-minimizing ST estimator $\hat{\mathbf{C}}_{\text{ST}}^o$, the unique solution to

$$\hat{\mathbf{C}}_{\text{ST}}^o = (1 - \rho^o) \frac{1}{n} \sum_{t=1}^{n} \frac{\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^T}{\frac{1}{N} \tilde{\mathbf{x}}_t^T \hat{\mathbf{C}}_{\text{ST}}^{o-1} \tilde{\mathbf{x}}_t} + \rho^o \mathbf{I}_N.$$

3) Construct the optimized portfolio

$$\hat{\mathbf{h}}_{\text{ST}}^o = \frac{\hat{\mathbf{C}}_{\text{ST}}^{o-1} \mathbf{1}_N}{\mathbf{1}_N^T \hat{\mathbf{C}}_{\text{ST}}^{o-1} \mathbf{1}_N}.$$

Lemma 1: Under the settings of Theorem 1, as $N, n \to \infty$,

$$\sup_{\rho \in \mathcal{R}_\varepsilon} |\hat{\gamma}_{\text{sc}} - \gamma/\kappa| \xrightarrow{\text{a.s.}} 0 \qquad (7)$$

where

$$\hat{\gamma}_{\text{sc}} = \frac{1}{1 - (1 - \rho)c_N} \frac{1}{n} \sum_{t=1}^{n} \frac{\tilde{\mathbf{x}}_t^T \hat{\mathbf{C}}_{\text{ST}}^{-1}(\rho) \tilde{\mathbf{x}}_t}{\|\tilde{\mathbf{x}}_t\|^2}.$$

Proof: See Appendix C.

The following theorem provides a consistent estimator of $\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho))$, scaled by $1/\kappa$, which is denoted as $\hat{\sigma}_{\text{sc}}^2(\rho)$. This is our second main result.

Theorem 2: Under the settings of Theorem 1, as $N, n \to \infty$,

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \left| N \hat{\sigma}_{\text{sc}}^2(\rho) - \frac{1}{\kappa} N \sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho)) \right| \xrightarrow{\text{a.s.}} 0$$

where $\hat{\sigma}_{\text{sc}}^2(\rho)$ is defined in (8) at the bottom of the page.

Proof: See Appendix D.

Note that, since κ is independent of ρ, the same ρ minimizes both $\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho))$ and $\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho))/\kappa$.

The following corollary of Theorem 2 is of fundamental importance, which demonstrates that choosing ρ to minimize $\hat{\sigma}_{\text{sc}}^2(\rho)$ is asymptotically equivalent to minimizing the unobservable $\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho))$.

Corollary 1: Denote ρ^o and ρ^* the minimizers of $\hat{\sigma}_{\text{sc}}^2(\rho)$ and $\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho))$ over $\mathcal{R}_\varepsilon$, respectively. Then, under the settings of Theorem 1 and Theorem 2, as $N, n \to \infty$,

$$\left| N \sigma^2 \left(\hat{\mathbf{h}}_{\text{ST}}(\rho^o) \right) - N \sigma^2 \left(\hat{\mathbf{h}}_{\text{ST}}(\rho^*) \right) \right| \xrightarrow{\text{a.s.}} 0.$$

Proof: See Appendix E.

With this result, the GMVP optimization problem is now reduced to the minimization of $\hat{\sigma}_{\text{sc}}^2(\rho)$, which can be done with a simple numerical search.

To summarize, given n past return observations of N assets, our proposed algorithm to construct a portfolio with minimal risk can be described as in Algorithm 1.

IV. SIMULATION RESULTS

We use both synthetic data and real market data to show the performance of $\hat{\mathbf{C}}_{\text{ST}}^o$ compared to the following competing methods.

1) $\hat{\mathbf{C}}_{\text{P}}$, referred to as the Abramovich-Pascal estimate from [32];
2) $\hat{\mathbf{C}}_{\text{C}}$, referred to as the Chen estimate from [32];
3) $\hat{\mathbf{C}}_{\text{C2}}$, the oracle estimator in [31], which has the same structure as $\hat{\mathbf{C}}_{\text{C}}$, but resorts to solving an approximate problem of minimizing the Frobenius distance to find the optimal shrinkage;
4) $\hat{\mathbf{C}}_{\text{LW}}$, the Ledoit-Wolf shrinkage estimator in [11];
5) $\hat{\mathbf{C}}_{\text{R}}$, the Rubio estimator proposed in [33], which has the same structure as $\hat{\mathbf{C}}_{\text{LW}}$, but with ρ calibrated based on the GMVP framework, as in the present article.

A. Synthetic Data Simulations

The synthetic data are generated i.i.d. from a multivariate Student-T distribution, where $\sqrt{\tau_t} = \sqrt{d/\chi_d^2}$, $d = 3$ and χ_d^2 is a Chi-square random variable with d degree of freedom. We set $N = 200$. The mean vector $\boldsymbol{\mu}$ can be set arbitrarily since it is discarded by the empirical mean, having no impact on the covariance estimates. We assume the population covariance matrix $\mathbf{C}_N$ is based on a one-factor return structure [35]: $\mathbf{C}_N = \mathbf{b}\mathbf{b}^T \sigma^2 + \boldsymbol{\Sigma}$, where $\sigma = 0.16$. The factor loadings $\mathbf{b} \in \mathbb{R}^N$ are evenly spread between 0.5 and 1.5. The residual variance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ is set to be diagonal and proportional to the identity matrix: $\boldsymbol{\Sigma} = \sigma_r^2 \mathbf{I}$, where $\sigma_r = 0.2$.

Fig. 1 illustrates the performance of different estimation approaches in terms of the realized risk, averaged over 200 Monte Carlo simulations. The risk bound is computed by (2), the theoretical minimum portfolio risk. Compared to other methods, our proposed estimator $\hat{\mathbf{C}}_{\text{ST}}^o$ achieves the smallest realized risk for both $n \leq N$ and $n > N$. For $n = 300$, there is about a 11.7% reduction in risk compared with $\hat{\mathbf{C}}_{\text{R}}$ (i.e., 3.0×10^{-3} compared to 3.4×10^{-3}) and about a 38.7% reduction in risk over $\hat{\mathbf{C}}_{\text{C}}$ (i.e., 3.0×10^{-3} compared to 4.9×10^{-3}). We omit the realized risks achieved by $\hat{\mathbf{C}}_N = \mathbf{I}_N$ as they are uniformly more than five times as large as those achieved by the other methods.

It is interesting to compare the optimized ρ of $\hat{\mathbf{C}}_{\text{ST}}^o$ and $\hat{\mathbf{C}}_{\text{P}}$. They are both solutions of (3), but with ρ optimized under different metrics: minimizing the risk and minimizing

---

$$\hat{\sigma}_{\text{sc}}^2(\rho) = \frac{\hat{\gamma}_{\text{sc}}}{(1 - \rho) - (1 - \rho)^2 c_N} \frac{\mathbf{1}_N^T \hat{\mathbf{C}}_{\text{ST}}^{-1}(\rho) \left(\hat{\mathbf{C}}_{\text{ST}}(\rho) - \rho \mathbf{I}_N \right) \hat{\mathbf{C}}_{\text{ST}}^{-1}(\rho) \mathbf{1}_N}{(\mathbf{1}_N^T \hat{\mathbf{C}}_{\text{ST}}^{-1}(\rho) \mathbf{1}_N)^2}. \qquad (8)$$
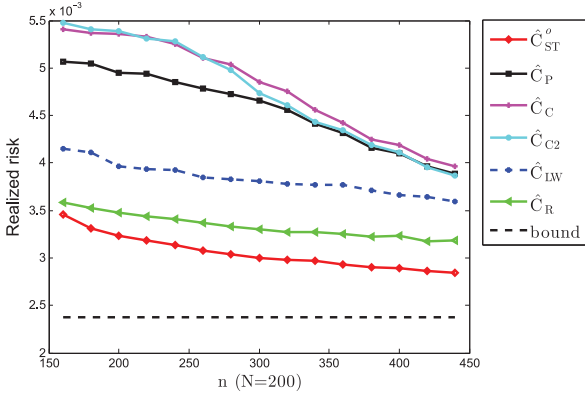
Fig. 1. The average realized portfolio risk of different covariance estimators in the GMVP framework using synthetic data.
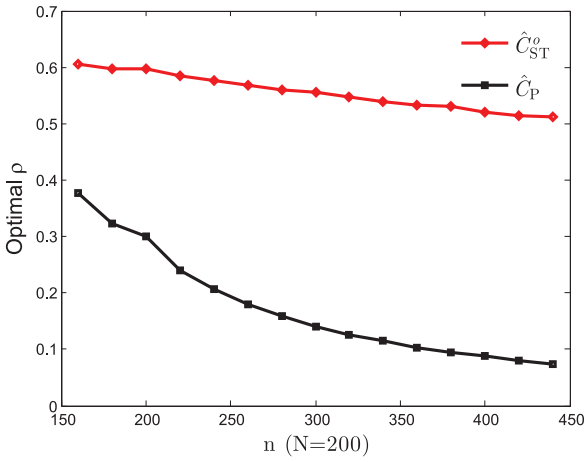

Fig. 2. The optimal shrinkage parameters of $\hat{\mathbf{C}}_{\mathrm{ST}}^{o}$ and $\hat{\mathbf{C}}_{\mathrm{P}}$ in the synthetic data simulation.
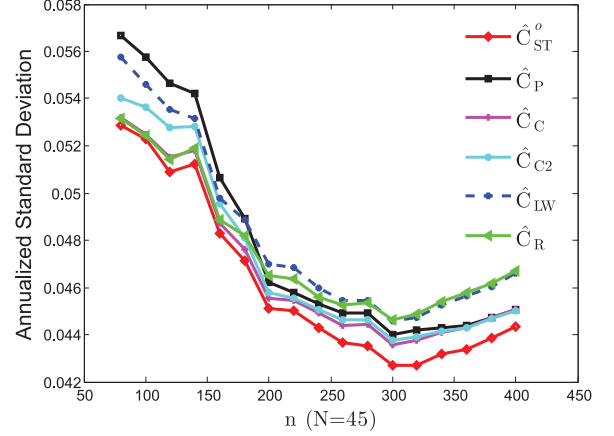

Fig. 3. Realized portfolio risks achieved out-of-sample over 736 days of HSI real market data (from Jan. 3, 2011 to Dec. 31, 2013) by a GMVP implemented using different covariance estimators.

the Frobenius distance, respectively. As shown in Fig. 2, the optimal shrinkage parameter varies under different metrics. Interestingly, optimizing ρ under the risk function as opposed to the Frobenius distance leads to more aggressive shrinkage (regularization) towards the identity matrix, thus producing a portfolio allocation which is closer to the uniform allocation policy.

B. Real Market Data Simulations

We now investigate the out-of-sample portfolio performance of the different estimators with the real market data. We consider the stocks comprising the HSI. In particular, we use the dividend-adjusted daily closing prices downloaded from the Yahoo Finance database to obtain the continuously compounded (logarithmic) returns for the 45 constituents of the HSI over $L = 736$ working days, from Jan. 3, 2011 to Dec. 31, 2013 (excluding weekends and public holidays).

As conventionally done in the financial literature, the out-of-sample evaluation is defined in terms of a rolling window method. At a particular day t, we use the previous n days (i.e., from $t - n$ to $t - 1$) as the training window for covariance estimation and construct the portfolio selection $\hat{\mathbf{h}}_{\mathrm{GMVP}}$. We then use $\hat{\mathbf{h}}_{\mathrm{GMVP}}$ to compute the portfolio returns in the following 10 days. Next the window is shifted 10 days forward and the

portfolio returns for another 10 days are computed. This procedure is repeated until the end of the data. The realized risk is computed conventionally as the annualized sample standard deviation of the corresponding GMVP returns. In our tests, different training window lengths are considered.

Fig. 3 shows that the proposed $\hat{\mathbf{C}}_{\mathrm{ST}}^{o}$ achieves the smallest realized risk. It outperforms the other methods over the entire span of considered estimation windows. The realized risk achieved by $\hat{\mathbf{C}}_{N} = \mathbf{I}_{N}$ is also omitted here because it is more than double as those achieved by the competing methods. When the estimation window is too long (e.g., greater than 320 days), we observe that the performance starts to systematically degrade. This is presumably due to a lack of stationarity in the data over such long durations. This highlights an interesting phenomenon worthy of further consideration, but a detailed study falls beyond the scope of the current contribution.

When the estimation window length is 300, the lowest risk is achieved by $\hat{\mathbf{C}}_{\mathrm{ST}}^{o}$. Table I presents the risks obtained by the different covariance estimators at the optimal estimation window length of 300. We also test whether the pairwise differences between the portfolio variance achieved by $\hat{\mathbf{C}}_{\mathrm{ST}}^{o}$ and each benchmark strategy are statistically different from zero. We follow the method described in [36] and [37] and employ a studentized version of the circular block bootstrap [38] to do the test. The p-values are computed under the null hypothesis that the portfolio variance achieved by a particular benchmark covariance matrix estimator is equal to that achieved by $\hat{\mathbf{C}}_{\mathrm{ST}}^{o}$. We use a block length $b = 5^{2}$ and base our reported p-values on 2000 bootstrap iterations. In the row reporting the risks, statistically significant outperformance of $\hat{\mathbf{C}}_{\mathrm{ST}}^{o}$ over other methods is denoted by asterisks: ** denotes significance at the 0.01 level ($p < 0.01$) and * denotes significance at the 0.05 level ($p < 0.05$). It can be seen from Table I that the outperformance of our proposed method is statistically significant, with $p < 0.05$ in all cases.

As a further comparison to investigate the performance with finer temporal resolution than that in Fig. 3, we carry out a

---

[2]We also compute the p-values when the block lengths are $b = 1$ and $b = 10$. The interpretation of the results does not change for $b = 1$, $b = 5$, or $b = 10$.

TABLE I
REALIZED PORTFOLIO RISKS (ANNUALIZED STANDARD DEVIATIONS) AND THE CORRESPONDING
P-VALUES UNDER DIFFERENT COVARIANCE MATRIX ESTIMATORS

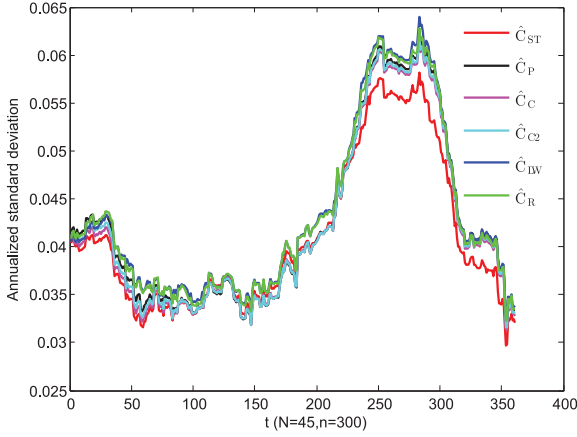| Dataset	Statistic	$\hat{\mathbf{C}}_{\mathrm{ST}}^o$	$\hat{\mathbf{C}}_{\mathrm{P}}$	$\hat{\mathbf{C}}_{\mathrm{C}}$	$\hat{\mathbf{C}}_{\mathrm{C2}}$	$\hat{\mathbf{C}}_{\mathrm{LW}}$	$\hat{\mathbf{C}}_{\mathrm{R}}$	$\mathbf{I}_N$
HSI	Risk (n=300)	0.0419	0.0433**	0.0428*	0.0430*	0.0438**	0.0439**	0.1112**
	p-value	1.000	0.009	0.028	0.041	0.001	0.001	0.000

Fig. 4. Annualized rolling-window standard deviations of the most recent 70 out-of-sample log returns for the GMVP based on different covariance matrix estimators.

rolling-window analysis on the realized risks. Under the optimal estimation window length of 300, we obtain 436 out-of-sample portfolio returns. From the start of the data, we use the most recent 70 out-of-sample portfolio returns to compute the (annualized) standard deviations of the GMVP. Shifting one day forward, we repeat this procedure until the end of the portfolio returns. For each covariance matrix estimator, this results in 367 risk measurements, which are then displayed in a time series plot, Fig. 4. We find that 69.2% of the time, $\hat{\mathbf{C}}_{\mathrm{ST}}^o$ achieves the lowest risk among all alternative methods. In addition, during the period of high volatility, that is, when $230 < t < 300$, $\hat{\mathbf{C}}_{\mathrm{ST}}^o$ exhibits the greatest outperformance. This justifies that our proposed GMVP optimization strategy is robust to market fluctuations and even possibly to outliers.

V. CONCLUSIONS

We have proposed a novel minimum-variance portfolio optimization strategy based on a robust shrinkage covariance estimator with a shrinkage parameter calibrated to minimize the realized portfolio risk. Our strategy has been shown to be robust to finite-sampling effects as well as to the impulsive characteristics of the data. It has been demonstrated that our approach outperforms more standard techniques in terms of the realized portfolio risk, both for synthetic data and for real historical stock returns from Hong Kong's HSI. Although we base our analysis on the assumption of the absence of the outliers, a recent study [39] has shown that the robust covariance estimator $\hat{\mathbf{C}}_{\mathrm{ST}}$ is resilient to arbitrary outliers by appropriately weighting good versus outlying data. This is somewhat confirmed by our real data tests and is worth investigating further.

Even though GMVP is not an optimal portfolio in terms of the Sharpe ratio or return maximization at a given level of risk, many empirical studies [40], [41] has shown that an investment

in the GMVP often yields better out-of-sample results than other mean-variance portfolios, because of the poor estimates of the means of the asset returns. Therefore, besides the robust estimation of the covariance matrix, it would be of interest to take into account the robust estimation of the means and further develop robust approaches to the various portfolio optimization strategies that involve both the estimates of the means and the covariance matrix of the asset returns, such as Sharpe ratio maximization or Markowitz's mean-variance portfolio optimization. These considerations are left to future work.

APPENDIX A
PRELIMINARY RESULTS

In this appendix we provide some preparatory lemmas that are essential for the proof of the main theorems. From now on, for readability, we discard all unnecessary indices ρ when no confusion is possible.

We start by rewriting $\hat{\mathbf{C}}_{\mathrm{ST}}$ in a more convenient form. Denoting

$$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \frac{1}{n}\mathbf{Z}_N\sqrt{\frac{\boldsymbol{\tau}}{\tau_t}}, \quad t = 1, 2, \ldots, n,$$

with $\sqrt{\boldsymbol{\tau}} = (\sqrt{\tau_1}, \ldots, \sqrt{\tau_n})^T$ and $\mathbf{Z}_N = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$, after some basic algebra, we obtain

$$\hat{\mathbf{C}}_{\mathrm{ST}} = (1-\rho)\frac{1}{n}\sum_{t=1}^n \frac{\tilde{\mathbf{z}}_t\tilde{\mathbf{z}}_t^T}{\frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\tilde{\mathbf{z}}_t} + \rho\mathbf{I}_N.$$

Denoting $\hat{\mathbf{C}}_{(t)} \triangleq \hat{\mathbf{C}}_{\mathrm{ST}} - (1-\rho)\frac{1}{n}\frac{\tilde{\mathbf{z}}_t\tilde{\mathbf{z}}_t^T}{\frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\tilde{\mathbf{z}}_t}$ and using $(\mathbf{A}+r\boldsymbol{v}\boldsymbol{v}^T)^{-1}\boldsymbol{v} = \mathbf{A}^{-1}\boldsymbol{v}/(1+r\boldsymbol{v}^T\mathbf{A}^{-1}\boldsymbol{v})$ for positive definite matrix $\mathbf{A}$, vector $\boldsymbol{v}$ and scalar $r > 0$, we have

$$\frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\tilde{\mathbf{z}}_t = \frac{\frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{(t)}^{-1}\tilde{\mathbf{z}}_t}{1 + (1-\rho)c_N\frac{\frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{(t)}^{-1}\tilde{\mathbf{z}}_t}{\frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\tilde{\mathbf{z}}_t}}$$

so that

$$\frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\tilde{\mathbf{z}}_t = (1 - (1-\rho)c_N)\frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{(t)}^{-1}\tilde{\mathbf{z}}_t \quad (9)$$

and we can rewrite $\hat{\mathbf{C}}_{\mathrm{ST}}$ as

$$\hat{\mathbf{C}}_{\mathrm{ST}} = \frac{1-\rho}{1-(1-\rho)c_N}\frac{1}{n}\sum_{t=1}^n \frac{\tilde{\mathbf{z}}_t\tilde{\mathbf{z}}_t^T}{\frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{(t)}^{-1}\tilde{\mathbf{z}}_t} + \rho\mathbf{I}_N.$$

For $t \in \{1, \ldots, n\}$, denote $\hat{d}_t(\rho) \triangleq \frac{1}{N}\tilde{\mathbf{z}}_t^T\hat{\mathbf{C}}_{(t)}^{-1}\tilde{\mathbf{z}}_t$. The following lemma gives a deterministic approximation of $\hat{d}_t(\rho)$, which later helps to show that, up to scaling, $\hat{\mathbf{C}}_{\mathrm{ST}}$ is somewhat similar to $\sum_{t=1}^n \mathbf{z}_t\mathbf{z}_t^T$, which is not observable.

Lemma 2: Under the settings of Theorem 1, as $N, n \to \infty$,

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \max_{1 \le t \le n} \left| \hat{d}_t(\rho) - \gamma(\rho) \right| \xrightarrow{\text{a.s.}} 0.$$

Proof: This is proved via a contradiction argument, which follows along lines similar to the proof in [32]. The main difference lies in that we re-center the sample data by subtracting the sample mean, while the samples are assumed to be zero mean in [32]. By subtracting the sample mean, the re-centered data are correlated and some $\sqrt{\tau_t}$ terms still remain in $\hat{\mathbf{C}}_{\mathrm{ST}}$, which introduces new technical difficulties.

Assuming (by relabelling) that $\hat{d}_1(\rho) \leq \cdots \leq \hat{d}_n(\rho)$, we first prove that for any fixed $\ell > 0$, $\hat{d}_n(\rho)$ is bounded above by $\gamma(\rho) + \ell$ for all large n, uniformly on $\rho \in \mathcal{R}_\varepsilon$. Since $\mathbf{U} \succcurlyeq \mathbf{V}$ $\Rightarrow \mathbf{V}^{-1} \succcurlyeq \mathbf{U}^{-1}$, for positive definite matrices $\mathbf{U}$ and $\mathbf{V}$, we obtain

$$
\hat{d}_n(\rho) = \frac{1}{N}\tilde{\mathbf{z}}_n^T \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n}\sum_{t=1}^{n-1} \frac{\tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T}{\hat{d}_t(\rho)} + \rho \mathbf{I}_N \right)^{-1} \tilde{\mathbf{z}}_n
$$

$$
\leq \frac{1}{N}\tilde{\mathbf{z}}_n^T \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n}\sum_{t=1}^{n-1} \frac{\tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T}{\hat{d}_n(\rho)} + \rho \mathbf{I}_N \right)^{-1} \tilde{\mathbf{z}}_n.
$$

Since $\tilde{\mathbf{z}}_n \neq 0$ with probability one, this implies

$$
1 \leq \frac{1}{N}\tilde{\mathbf{z}}_n^T \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n}\sum_{t=1}^{n-1} \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T + \hat{d}_n(\rho)\rho \mathbf{I}_N \right)^{-1} \tilde{\mathbf{z}}_n.
\tag{10}
$$

Assume that there exists a sequence $\{\rho_n\}_{n=1}^\infty$ over which $\hat{d}_n(\rho_n) > \gamma(\rho_n) + \ell$ infinitely often, for some fixed $\ell > 0$. Since $\{\rho_n\}_{n=1}^\infty$ is bounded, it has a limit point $\rho_0 \in \mathcal{R}_\varepsilon$. Let us restrict ourselves to such a subsequence on which $\rho_n \to \rho_0 > 0$ and $\hat{d}_n(\rho_n) > \gamma(\rho_n) + \ell$. On this subsequence, from (10), we have $\widetilde{m}_{N,n} \geq 1$, where $\widetilde{m}_{N,n} = \frac{1}{N}\tilde{\mathbf{z}}_n^T \widetilde{\mathbf{M}}_{N,n} \tilde{\mathbf{z}}_n$ and $\widetilde{\mathbf{M}}_{N,n} = \left(\frac{1-\rho_n}{1-(1-\rho_n)c_N} \frac{1}{n}\sum_{t=1}^n \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T + (\gamma(\rho_n) + \ell)\rho_n \mathbf{I}_N \right)^{-1}$.

The quadratic form $\widetilde{m}_{N,j}$ is amenable to large random matrix analysis. The first step is to remove the effect of the sample mean. Denote $m_{N,j} = \frac{1}{N}\mathbf{z}_j^T \mathbf{M}_{N,j} \mathbf{z}_j$ and $\mathbf{M}_{N,j} = \left(\frac{1-\rho_n}{1-(1-\rho_n)c_N} \frac{1}{n}\sum_{t \neq j} \mathbf{z}_t \mathbf{z}_t^T + (\gamma(\rho_n) + \ell)\rho_n \mathbf{I}_N \right)^{-1}$. We have in particular:

Proposition 1: As $N, n \to \infty$,

$$
\max_{1 \leq j \leq n} |\widetilde{m}_{N,j} - m_{N,j}| \xrightarrow{a.s.} 0.
\tag{11}
$$

Proof: See Appendix F.

Remark 2: In Proposition 1, Assumption 1–b is necessary; that is, i.i.d. $\tau_1, \ldots, \tau_n \geq \xi$ a.s. for some $\xi > 0$ and $E[\tau_1] < \infty$. It guarantees that for $t = 1, \ldots, n$, the norm of $\tilde{\mathbf{z}}_t$ does not go off to infinity, recalling that $\tilde{\mathbf{z}}_t = \mathbf{z}_t - \frac{1}{n}\mathbf{Z}_N \sqrt{\frac{\tau}{\tau_t}}$.

By Proposition 1, we have $|\widetilde{m}_{N,n} - m_{N,n}| \xrightarrow{a.s.} 0$. This allows us to follow the proof in [32], which deals with data with mean zero.

To proceed, assume first $\rho_0 \neq 1$. From the proof of Theorem 1 in [32],

$$
m_{N,n} \xrightarrow{a.s.}
$$

$$
\frac{1-(1-\rho_0)c}{1-\rho_0}\delta\left(-(\gamma(\rho_0)+\ell)\rho_0 \frac{1-(1-\rho_0)c}{1-\rho_0} \right) \triangleq m^+, \tag{12}
$$

where, for $x < 0$, $\delta(x)$ is the unique positive solution to

$$
\delta(x) = \int \frac{t}{-x + \frac{t}{1+c\delta(x)}}\nu(dt).
$$

Together with $|\widetilde{m}_{N,n} - m_{N,n}| \xrightarrow{a.s.} 0$, we have

$$
|\widetilde{m}_{N,n} - m^+| \xrightarrow{a.s.} 0.
\tag{13}
$$

It was demonstrated in [32] that $m^+ < 1$. But this is in contradiction with $\widetilde{m}_{N,n} \geq 1$.

Now assume $\rho_0 = 1$. According to [32], $m_{N,n} \xrightarrow{a.s.} \frac{1}{1+\ell} < 1$. Then

$$
\left| \widetilde{m}_{N,n} - \frac{1}{1+\ell} \right| \xrightarrow{a.s.} 0,
$$

but $\frac{1}{1+\ell} < 1$, again raising a contradiction with $\widetilde{m}_{N,n} \geq 1$.

Hence, for all large n, there is no converging subsequence of ρ_n (and thus no subsequence of ρ_n) for which $\hat{d}_n(\rho_n) > \gamma(\rho_n) + \ell$ infinitely often. Therefore $\hat{d}_n(\rho) \leq \gamma(\rho) + \ell$ for all large n a.s., uniformly on $\rho \in \mathcal{R}_\varepsilon$.

The same reasoning holds for $\hat{d}_1(\rho)$, which can be proved greater than $\gamma(\rho) - \ell$ for all large n uniformly on $\rho \in \mathcal{R}_\varepsilon$. Following the same arguments in [32], since $\ell > 0$ is arbitrary, from the ordering of the $\hat{d}_t(\rho)$, we have proved that $\sup_{\rho \in \mathcal{R}_\varepsilon} \max_{1 \leq t \leq n} |\hat{d}_t(\rho) - \gamma(\rho)| \xrightarrow{a.s.} 0$. ∎

The following three lemmas, Lemma 3, 4 and 5 show that functionals of Tyler's estimator asymptotically perform similar to functionals of $\frac{1}{n}\sum_{t=1}^n \mathbf{z}_t \mathbf{z}_t^T$ or $\frac{1}{n}\sum_{t=1}^n \mathbf{y}_t \mathbf{y}_t^T$. They are used as an intermediate step for the development of the asymptotic deterministic equivalent of the risk function. Using existing results in [33], quoted as Lemma 6 in this paper, we can then obtain our main theorems.

For notational convenience, we denote $k = k(\rho) \triangleq \frac{1-\rho}{1-(1-\rho)c}$. Also recall that γ is the unique positive solution to $1 = \int \frac{t}{\gamma\rho + (1-\rho)t}\nu(dt)$. Assuming $\mathbf{A}_N \in \mathbb{R}^{N \times N}$ is a deterministic symmetric nonnegative definite matrix, for some $\eta > 0$, define $\mathcal{D} = \begin{cases} [0, \infty) & \text{if } \liminf_N \lambda_1(\mathbf{A}_N) > 0 \\ [\eta, \infty) & \text{otherwise} \end{cases}$, and further define that, for $\rho \in \mathcal{R}_\varepsilon$ and $w \in \mathcal{D}$,

$$
\widetilde{\mathbf{R}}_N = \left(\mathbf{A}_N + (1-\rho)\frac{1}{n}\sum_{t=1}^n \frac{\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^T}{\frac{1}{N}\tilde{\mathbf{x}}_t^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1} \tilde{\mathbf{x}}_t} + w\mathbf{I}_N \right)^{-1}
$$

$$
\widetilde{\mathbf{S}}_N = \left(\mathbf{A}_N + \frac{k}{\gamma}\frac{1}{n}\sum_{t=1}^n \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T + w\mathbf{I}_N \right)^{-1}
$$

$$
\mathbf{S}_N = \left(\mathbf{A}_N + \frac{k}{\gamma}\frac{1}{n}\sum_{t=1}^n \mathbf{z}_t \mathbf{z}_t^T + w\mathbf{I}_N \right)^{-1}.
$$

Then we introduce the following lemma.

Lemma 3: Assume $\mathbf{a}_N \in \mathbb{R}^N$ is a deterministic vector with $\limsup_N \|\mathbf{a}_N\|^2 < \infty$. Under the settings of Theorem 1, as $N, n \to \infty$,

$$
\sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} \left| \mathbf{a}_N^T \widetilde{\mathbf{R}}_N \mathbf{a}_N - \mathbf{a}_N^T \mathbf{S}_N \mathbf{a}_N \right| \xrightarrow{a.s.} 0.
\tag{14}
$$

Proof: Define

$$\hat{\mathbf{B}}_N(\rho) = \frac{k}{\gamma(\rho)} \frac{1}{n} \sum_{t=1}^{n} \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T,$$

$$\hat{\mathbf{D}}_N(\rho) = (1-\rho) \frac{1}{n} \sum_{t=1}^{n} \frac{\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^T}{\frac{1}{N} \tilde{\mathbf{x}}_t^T \hat{\mathbf{C}}_{ST}^{-1} \tilde{\mathbf{x}}_t}$$

$$\overset{(a)}{=} \frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{t=1}^{n} \frac{\tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T}{\hat{d}_t(\rho)}$$

where (a) uses the identity (9). Denote

$$\Delta \triangleq \mathbf{a}_N^T \widetilde{\mathbf{R}}_N \mathbf{a}_N - \mathbf{a}_N^T \widetilde{\mathbf{S}}_N \mathbf{a}_N$$

$$\overset{(a)}{=} \mathbf{a}_N^T \widetilde{\mathbf{R}}_N \left(\hat{\mathbf{B}}_N(\rho) - \hat{\mathbf{D}}_N(\rho) \right) \widetilde{\mathbf{S}}_N \mathbf{a}_N$$

where (a) uses the identity that $\mathbf{U}^{-1} - \mathbf{V}^{-1} = \mathbf{U}^{-1}(\mathbf{V} - \mathbf{U})\mathbf{V}^{-1}$ for invertible $\mathbf{U}, \mathbf{V}$ matrices. We first prove that as $N, n \to \infty$, $\sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} |\Delta| \overset{\text{a.s.}}{\longrightarrow} 0$.

As $N, n \to \infty$, using the definition of k,

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \left\| \hat{\mathbf{D}}_N(\rho) - \hat{\mathbf{B}}_N(\rho) \right\|$$

$$\leq \left\| \frac{1}{n} \sum_{t=1}^{n} \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T \right\| \sup_{\rho \in \mathcal{R}_\varepsilon} \max_{1 \leq t \leq n} \frac{1-\rho}{1-(1-\rho)c} \left| \frac{\hat{d}_t(\rho) - \gamma(\rho)}{\gamma(\rho)\hat{d}_t(\rho)} \right|. \quad (15)$$

We will show that the RHS of (15) goes to 0 a.s. Recalling Lemma 2, this follows upon showing that $\limsup_n \left\| \frac{1}{n} \sum_{t=1}^{n} \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T \right\| < \infty$ a.s. To this end, recall that $\tilde{\mathbf{z}}_t = \mathbf{z}_t - \frac{1}{n} \mathbf{Z}_N \sqrt{\frac{\tau}{\tau_t}}$. Then

$$\frac{1}{n} \sum_{t=1}^{n} \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T$$

$$= \frac{1}{n} \sum_{t=1}^{n} \mathbf{z}_t \mathbf{z}_t^T - \frac{1}{n} \sum_{t=1}^{n} \mathbf{z}_t \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right)^T - \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right) \mathbf{z}_t^T$$

$$+ \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right) \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right)^T. \quad (16)$$

We will show that the spectral norm of each term on the RHS of (16) is bounded for all large n a.s.

First, from Assumption 1-c. and [42], we have $\limsup_n \left\| \frac{1}{n} \sum_{t=1}^{n} \mathbf{z}_t \mathbf{z}_t^T \right\| < \infty$ a.s. Next, for the second and the third terms on the RHS of (16),

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \mathbf{z}_t \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right)^T \right\| = \left\| \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right) \mathbf{z}_t^T \right\|$$

$$= \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathbf{y}_t}{\sqrt{\tau_t}} \right)^T \mathbf{C}_N \left(\frac{1}{n} \sum_{t=1}^{n} \mathbf{y}_t \sqrt{\tau_t} \right)$$

$$\leq, \|\mathbf{C}_N\| \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathbf{y}_t}{\sqrt{\tau_t}} \right)^T \left(\frac{1}{n} \sum_{t=1}^{n} \mathbf{y}_t \sqrt{\tau_t} \right).$$

By the law of large numbers, as $N, n \to \infty$,

$$\left| \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathbf{y}_t}{\sqrt{\tau_t}} \right)^T \left(\frac{1}{n} \sum_{t=1}^{n} \mathbf{y}_t \sqrt{\tau_t} \right) - c \right| \overset{\text{a.s.}}{\longrightarrow} 0.$$

According to Assumption 1-a and Assumption 1-c, we can see that $\limsup_n \left\| \frac{1}{n} \sum_{t=1}^{n} \mathbf{z}_t \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right)^T \right\| = \limsup_n \left\| \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right) \mathbf{z}_t^T \right\| \leq c \|\mathbf{C}_N\| < \infty$ a.s.

For the fourth term, with Assumption 1-b, we have

$$\limsup_n \left\| \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right) \left(\frac{1}{n} \mathbf{Z}_N \frac{\sqrt{\tau}}{\sqrt{\tau_t}} \right)^T \right\|$$

$$\leq \limsup_n \left\{ \left\| \frac{1}{n} \mathbf{Z}_N^T \mathbf{Z}_N \right\| \left(\frac{1}{n} \sum_{t=1}^{n} \tau_t \right) \left(\frac{1}{n} \sum_{t=1}^{n} \frac{1}{\tau_t} \right) \right\} < \infty \text{ a.s.}$$

Therefore, $\limsup_n \left\| \frac{1}{n} \sum_{t=1}^{n} \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T \right\| < \infty$ a.s. Together with Lemma 2, from (15), we have

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \left\| \hat{\mathbf{B}}_N(\rho) - \hat{\mathbf{D}}_N(\rho) \right\| \overset{\text{a.s.}}{\longrightarrow} 0. \quad (17)$$

Note that $w \in \mathcal{D}$ ensures $\limsup_N \sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} \left\| \widetilde{\mathbf{R}}_N \right\| < \infty$ and $\limsup_N \sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} \left\| \widetilde{\mathbf{S}}_N \right\| < \infty$.

Together with (17) and $\|\mathbf{a}_N\|^2 < \infty$, we have

$$\sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} |\Delta| \leq \|\mathbf{a}_N\|^2 \sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} \left\| \widetilde{\mathbf{R}}_N \right\| \sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} \left\| \widetilde{\mathbf{S}}_N \right\|$$

$$\times \sup_{\rho \in \mathcal{R}_\varepsilon} \left\| \frac{k}{\gamma} \frac{1}{n} \sum_{t=1}^{n} \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T - (1-\rho) \frac{1}{n} \sum_{t=1}^{n} \frac{\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^T}{\frac{1}{N} \tilde{\mathbf{x}}_t^T \hat{\mathbf{C}}_{ST}^{-1} \tilde{\mathbf{x}}_t} \right\| \overset{\text{a.s.}}{\longrightarrow} 0.$$

Following the same reasoning as that of Proposition 1, we have

$$\sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} \left| \mathbf{a}_N^T \widetilde{\mathbf{S}}_N \mathbf{a}_N - \mathbf{a}_N^T \mathbf{S}_N \mathbf{a}_N \right| \overset{\text{a.s.}}{\longrightarrow} 0.$$

Together with $\sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} |\Delta| \overset{\text{a.s.}}{\longrightarrow} 0$, we obtain (14). ∎

Define $\mathbf{W}_N = \left(\mathbf{A}_N + \frac{k}{\gamma} \frac{1}{n} \sum_{t=1}^{n} \mathbf{y}_t \mathbf{y}_t^T + w \mathbf{I}_N \right)^{-1}$ and

$$\widetilde{\mathbf{W}}_N = \left(\mathbf{A}_N + (1-\rho) \frac{1}{n} \sum_{t=1}^{n} \frac{\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t^T}{\frac{1}{N} \tilde{\mathbf{y}}_t^T \mathbf{C}_N^{1/2} \hat{\mathbf{C}}_{ST}^{-1} \mathbf{C}_N^{1/2} \tilde{\mathbf{y}}_t} + w \mathbf{I}_N \right)^{-1}$$

where $\tilde{\mathbf{y}}_t = \mathbf{y}_t - \frac{1}{n} \sum_{i=1}^{n} \mathbf{y}_i$. We introduce the following lemma.

Lemma 4: Under the settings of Lemma 3, as $N, n \to \infty$,

$$\sup_{\rho \in \mathcal{R}_\varepsilon, w \in \mathcal{D}} \left| \mathbf{a}_N^T \widetilde{\mathbf{W}}_N \mathbf{a}_N - \mathbf{a}_N^T \mathbf{W}_N \mathbf{a}_N \right| \overset{\text{a.s.}}{\longrightarrow} 0. \quad (18)$$

Proof: The derivation is similar to that of (14).

Lemma 5: Under the settings of Lemma 3 and assuming $\mathbf{A}_N = \mathbf{0}$, as $N, n \to \infty$,

$$\sup_{\rho \in \mathcal{R}_\varepsilon, w \in [\eta, \infty)} \left| \mathbf{a}_N^T (1-\rho) \frac{1}{n} \sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T}{\frac{1}{N} \tilde{\mathbf{x}}_i^T \hat{\mathbf{C}}_{ST}^{-1} \tilde{\mathbf{x}}_i} \widetilde{\mathbf{R}}_N^2 \mathbf{a}_N \right.$$

$$\left. - \mathbf{a}_N^T \frac{k}{\gamma} \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^T \mathbf{S}_N^2 \mathbf{a}_N \right| \overset{\text{a.s.}}{\longrightarrow} 0. \quad (19)$$

Proof: We first notice that

$$
\mathbf{a}_N^T (1-\rho) \frac{1}{n} \sum_{i=1}^n \frac{\tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T}{\frac{1}{N}\tilde{\mathbf{x}}_i^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1} \tilde{\mathbf{x}}_i} \widetilde{\mathbf{R}}_N^2 \mathbf{a}_N
$$

$$
= -\frac{\mathrm{d}}{\mathrm{d}w}\left[\mathbf{a}_N^T (1-\rho)\frac{1}{n}\sum_{i=1}^n \frac{\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T}{\frac{1}{N}\tilde{\mathbf{x}}_i^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\tilde{\mathbf{x}}_i}\widetilde{\mathbf{R}}_N \mathbf{a}_N\right]
$$

$$
= -\frac{\mathrm{d}}{\mathrm{d}w}\left(\mathbf{a}_N^T \mathbf{a}_N - w\mathbf{a}_N^T\widetilde{\mathbf{R}}_N\mathbf{a}_N\right)
$$

$$
= \mathbf{a}_N^T \widetilde{\mathbf{R}}_N \mathbf{a}_N + w\frac{\mathrm{d}}{\mathrm{d}w}\left(\mathbf{a}_N^T\widetilde{\mathbf{R}}_N\mathbf{a}_N\right).
$$

Following similar steps, we also have

$$
\mathbf{a}_N^T \frac{k}{\gamma}\frac{1}{n}\sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i^T \mathbf{S}_N^2 \mathbf{a}_N = \mathbf{a}_N^T \mathbf{S}_N \mathbf{a}_N + w\frac{\mathrm{d}}{\mathrm{d}w}\left(\mathbf{a}_N^T \mathbf{S}_N\mathbf{a}_N\right).
$$

The almost sure convergence (14) in Lemma 3 when extended to $w \in \mathbb{C}$ is uniform on any bounded region of $(\mathbb{C}-\mathbb{R}) \cup \mathcal{D}$, and the functionals of w in (14) are analytic. Thus, by the Weierstrass convergence theorem [43], the following holds:

$$
\sup_{\rho\in\mathcal{R}_\varepsilon, w\in\mathcal{D}}\left|\frac{\mathrm{d}}{\mathrm{d}w}\left(\mathbf{a}_N^T\widetilde{\mathbf{R}}_N\mathbf{a}_N\right) - \frac{\mathrm{d}}{\mathrm{d}w}\left(\mathbf{a}_N^T\mathbf{S}_N\mathbf{a}_N\right)\right| \xrightarrow{\text{a.s.}} 0.
$$

Together with Lemma 3, we obtain (19). ∎

Lemma 6 [33, Appendix I-B]: Under the settings of Lemma 3, as $N, n \to \infty$,

$$
\sup_{\rho\in\mathcal{R}_\varepsilon, w\in\mathcal{D}}\left|\mathbf{a}_N^T\mathbf{S}_N\mathbf{a}_N - \mathbf{a}_N^T\mathbf{T}_N\mathbf{a}_N\right| \xrightarrow{\text{a.s.}} 0 \qquad (20)
$$

where $\mathbf{T}_N = \left(\mathbf{A}_N + \frac{k}{(\gamma+e_N(w)k)}\mathbf{C}_N + w\mathbf{I}_N\right)^{-1}$ and, for each $w \in \mathcal{D}$, $e_N(w)$ is the unique positive solution to the following equation:

$$
e_N(w) = \frac{1}{n}\mathrm{tr}\left[\mathbf{C}_N\left(\mathbf{A}_N + \frac{k}{(\gamma+e_N(w)k)}\mathbf{C}_N + w\mathbf{I}_N\right)^{-1}\right].
$$

Moreover, when $\mathbf{A}_N = \mathbf{0}$, we have

$$
\sup_{\rho\in\mathcal{R}_\varepsilon, w\in[\eta,\infty)}\left|\mathbf{a}_N^T\frac{k}{\gamma}\frac{1}{n}\sum_{t=1}^n \mathbf{z}_t\mathbf{z}_t^T \mathbf{S}_N^2 \mathbf{a}_N\right.
$$

$$
\left. -\frac{k\gamma}{(\gamma+e_N(w)k)^2}\frac{\mathbf{a}_N^T\mathbf{C}_N\mathbf{T}_N^2\mathbf{a}_N}{1-\frac{k^2}{(\gamma+e_N(w)k)^2}\frac{1}{n}\mathrm{tr}[\mathbf{C}_N^2\mathbf{T}_N^2]}\right| \xrightarrow{\text{a.s.}} 0.
$$

∎

APPENDIX B
PROOF OF THEOREM 1

First consider the (re-scaled) realized portfolio risk:

$$
N\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}) = \frac{\frac{1}{N}\mathbf{1}_N^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{C}_N\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{1}_N}{(\frac{1}{N}\mathbf{1}_N^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{1}_N)^2}. \qquad (21)
$$

For the denominator, Lemma 3 and Lemma 6 imply

$$
\sup_{\rho\in\mathcal{R}_\varepsilon}\left|\frac{1}{N}\mathbf{1}_N^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{1}_N - \frac{1}{N}\mathbf{1}_N^T\left(\frac{1-\rho}{\gamma}\mathbf{C}_N + \rho\mathbf{I}_N\right)^{-1}\mathbf{1}_N\right| \xrightarrow{\text{a.s.}} 0.
$$

Note that in this case, $\mathbf{A}_N = \mathbf{0}$, $\mathbf{a}_N = \frac{1}{\sqrt{N}}\mathbf{1}_N$ and $w = \rho$, which leads to $|e_N(\rho) - c\gamma| \xrightarrow{\text{a.s.}} 0$ when $N, n \to \infty$. The

derivation is based on Assumption 1-c and the definition of γ in (5).

For the numerator, we rewrite it as

$$
\frac{1}{N}\mathbf{1}_N^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{C}_N\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{1}_N
$$

$$
= \frac{1}{N}\mathbf{1}_N^T\mathbf{C}_N^{-1/2}(\mathbf{C}_N^{-1/2}\hat{\mathbf{C}}_{\mathrm{ST}}\mathbf{C}_N^{-1/2})^{-2}\mathbf{C}_N^{-1/2}\mathbf{1}_N,
$$

which, upon substituting the RHS of (3) for $\hat{\mathbf{C}}_{\mathrm{ST}}$ and setting $\mathbf{A}_N = \rho\mathbf{C}_N^{-1}$ in $\widetilde{\mathbf{W}}_N$, yields

$$
\frac{1}{N}\mathbf{1}_N^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{C}_N\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{1}_N
$$

$$
= -\frac{\mathrm{d}}{\mathrm{d}w}\left[\frac{1}{N}\mathbf{1}_N^T\mathbf{C}_N^{-1/2}\widetilde{\mathbf{W}}_N\mathbf{C}_N^{-1/2}\mathbf{1}_N\right]\bigg|_{w=0}.
$$

Setting $\mathbf{A}_N = \rho\mathbf{C}_N^{-1}$ and $\mathbf{a}_N^T = \frac{1}{\sqrt{N}}\mathbf{C}_N^{-1/2}\mathbf{1}_N$ in (20), as well as $\mathbf{A}_N = \rho\mathbf{C}_N^{-1}$ in $\mathbf{W}_N$, yields

$$
\sup_{\rho\in\mathcal{R}_\varepsilon, w\in[0,\infty)}\left|\frac{1}{N}\mathbf{1}_N^T\mathbf{C}_N^{-1/2}\mathbf{W}_N\mathbf{C}_N^{-1/2}\mathbf{1}_N\right.
$$

$$
\left. -\frac{1}{N}\mathbf{1}_N^T\mathbf{C}_N^{-1/2}\mathbf{J}_N\mathbf{C}_N^{-1/2}\mathbf{1}_N\right| \xrightarrow{\text{a.s.}} 0 \quad (22)
$$

where $\mathbf{J}_N = \left(\rho\mathbf{C}_N^{-1} + \left(\frac{k}{(\gamma+\tilde{e}_N(w)k)} + w\right)\mathbf{I}_N\right)^{-1}$ and for each $w \in [0, \infty)$, $\tilde{e}_N(w)$ is the unique positive solution to the following equation:

$$
\tilde{e}_N(w) = \frac{1}{n}\mathrm{tr}\left[\left(\rho\mathbf{C}_N^{-1} + \left(\frac{k}{\gamma+\tilde{e}_N(w)k} + w\right)\mathbf{I}_N\right)^{-1}\right].
$$

Lemma 4 and the convergence (22) imply

$$
\sup_{\rho\in\mathcal{R}_\varepsilon, w\in[0,\infty)}\left|\frac{1}{N}\mathbf{1}_N^T\mathbf{C}_N^{-1/2}\widetilde{\mathbf{W}}_N\mathbf{C}_N^{-1/2}\mathbf{1}_N\right.
$$

$$
\left. -\frac{1}{N}\mathbf{1}_N^T\mathbf{C}_N^{-1/2}\mathbf{J}_N\mathbf{C}_N^{-1/2}\mathbf{1}_N\right| \xrightarrow{\text{a.s.}} 0.
$$

Following the same reasoning as for the proof of Lemma 5, the convergence of the derivatives holds such that at $w = 0$ by the Weierstrass convergence theorem,

$$
\sup_{\rho\in\mathcal{R}_\varepsilon}\left|\frac{\mathrm{d}}{\mathrm{d}w}\left(\frac{1}{N}\mathbf{1}_N^T\mathbf{C}_N^{-1/2}\widetilde{\mathbf{W}}_N\mathbf{C}_N^{-1/2}\mathbf{1}_N\right)\right|_{w=0}
$$

$$
\left. -\frac{\mathrm{d}}{\mathrm{d}w}\left(\frac{1}{N}\mathbf{1}_N^T\mathbf{C}_N^{-1/2}\mathbf{J}_N\mathbf{C}_N^{-1/2}\mathbf{1}_N\right)\right|_{w=0}\right| \xrightarrow{\text{a.s.}} 0.
$$

With (23), shown at the bottom of the next page, and $|\tilde{e}_N(0) - c\gamma| \xrightarrow{\text{a.s.}} 0$ when $N, n \to \infty$, we have

$$
\sup_{\rho\in\mathcal{R}_\varepsilon}\left|\frac{1}{N}\mathbf{1}_N^T\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{C}_N\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}\mathbf{1}_N - \frac{\gamma^2}{\gamma^2-\beta(1-\rho)^2}\frac{1}{N}\mathbf{1}_N^T\right.
$$

$$
\left. \times \left(\frac{1-\rho}{\gamma}\mathbf{C}_N + \rho\mathbf{I}_N\right)^{-1}\mathbf{C}_N\left(\frac{1-\rho}{\gamma}\mathbf{C}_N + \rho\mathbf{I}_N\right)^{-1}\mathbf{1}_N\right|
$$

$$
\xrightarrow{\text{a.s.}} 0. \qquad (24)
$$

Equipped with the asymptotic equivalences of the denominator and numerator of (21), we prove Theorem 1.

APPENDIX C
PROOF OF LEMMA 1

First notice that

$$\hat{\gamma}_{\text{sc}} = \frac{1}{1-(1-\rho)c_N} \frac{1}{N} \frac{1}{n} \sum_{t=1}^{n} \frac{\tilde{\mathbf{x}}_t^T \hat{\mathbf{C}}_{\text{ST}}^{-1}(\rho) \tilde{\mathbf{x}}_t}{\frac{1}{N}\|\tilde{\mathbf{x}}_t\|^2}$$

$$= \frac{1}{1-(1-\rho)c_N} \frac{1}{N} \frac{1}{n} \sum_{t=1}^{n} \frac{\tilde{\mathbf{z}}_t^T \hat{\mathbf{C}}_{\text{ST}}^{-1}(\rho) \tilde{\mathbf{z}}_t}{\frac{1}{N}\|\tilde{\mathbf{z}}_t\|^2}.$$

It has been shown in Lemma 2 that

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \max_{1 \le t \le n} \left| \hat{d}_t(\rho) - \gamma(\rho) \right| \xrightarrow{\text{a.s.}} 0$$

here $\hat{d}_t(\rho) = \frac{1}{1-(1-\rho)c_N} \frac{1}{N} \tilde{\mathbf{z}}_t^T \hat{\mathbf{C}}_{\text{ST}}^{-1}(\rho) \tilde{\mathbf{z}}_t$. Therefore, to prove the convergence (7), it is left to show that $\frac{1}{N}\|\tilde{\mathbf{z}}_t\|^2 \xrightarrow{\text{a.s.}} \kappa$.

We start by writing

$$\frac{1}{N}\|\tilde{\mathbf{z}}_t\|^2 = \frac{1}{N}\left(\mathbf{z}_t^T \mathbf{z}_t - \frac{1}{n}\mathbf{z}_t^T \mathbf{Z}_N \frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_t}} - \frac{1}{n}\frac{\sqrt{\boldsymbol{\tau}}^T}{\sqrt{\tau_t}} \mathbf{Z}_N^T \mathbf{z}_t \right.$$
$$\left. + \frac{1}{n^2}\frac{\sqrt{\boldsymbol{\tau}}^T}{\sqrt{\tau_t}} \mathbf{Z}_N^T \mathbf{Z}_N \frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_t}} \right). \quad (25)$$

Since the second and the third term on the RHS of (25) are the same, we analyze the second term only. It can be rewritten as

$$\frac{1}{N}\frac{1}{n}\mathbf{z}_t^T \mathbf{Z}_N \frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_t}} = \frac{1}{N}\frac{1}{n}\mathbf{z}_t^T \mathbf{z}_t + \frac{1}{N}\frac{1}{n}\mathbf{z}_t^T \mathbf{Z}_N^{(t)} \frac{\sqrt{\boldsymbol{\tau}^{(t)}}}{\sqrt{\tau_t}}$$

where $\mathbf{Z}_N^{(t)}$ is the matrix with the tth column removed from $\mathbf{Z}_N$ and $\sqrt{\boldsymbol{\tau}^{(t)}}$ is the vector with the tth entry removed from $\sqrt{\boldsymbol{\tau}}$. Since $\mathbf{z}_t$ is independent of $\frac{1}{n}\mathbf{Z}_N^{(t)}\frac{\sqrt{\boldsymbol{\tau}^{(t)}}}{\sqrt{\tau_t}}$, we have $\frac{1}{n}\mathbf{z}_t^T \mathbf{Z}_N^{(t)} \frac{\sqrt{\boldsymbol{\tau}^{(t)}}}{\sqrt{\tau_t}} \xrightarrow{\text{a.s.}} 0$. Together with $\frac{1}{n}\mathbf{z}_t^T \mathbf{z}_t = O(1)$ a.s., we have $\frac{1}{N}\frac{1}{n}\mathbf{z}_t^T \mathbf{Z}_N \frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_t}} \xrightarrow{\text{a.s.}} 0$.

For the last term in (25),

$$\limsup_n \left| \frac{1}{n^2}\frac{\sqrt{\boldsymbol{\tau}}^T}{\sqrt{\tau_t}} \mathbf{Z}_N^T \mathbf{Z}_N \frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_t}} \right|$$
$$\le \limsup_n \left\| \frac{1}{n}\mathbf{Z}_N^T \mathbf{Z}_N \right\| \frac{1}{\tau_t}\frac{1}{n}\sum_{i=1}^{n} \tau_i < \infty.$$

Thus, $\frac{1}{N}\frac{1}{n^2}\frac{\sqrt{\boldsymbol{\tau}}^T}{\sqrt{\tau_t}} \mathbf{Z}_N^T \mathbf{Z}_N \frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_t}} \xrightarrow{\text{a.s.}} 0$.

Since the last three terms on the RHS of (25) vanish with large n, we obtain $\frac{1}{N}\left| \|\tilde{\mathbf{z}}_t^2\| - \|\mathbf{z}_t^2\| \right| \xrightarrow{\text{a.s.}} 0$. Therefore, as $\frac{1}{N}\|\mathbf{z}_t\|^2 \xrightarrow{\text{a.s.}} \kappa$, we obtain $\frac{1}{N}\|\tilde{\mathbf{z}}_t\|^2 \xrightarrow{\text{a.s.}} \kappa$ and the convergence (7) unfolds.

APPENDIX D
PROOF OF THEOREM 2

According to Lemma 5 and Lemma 6, in which we set $\mathbf{A}_N = \mathbf{0}$, $w = \rho$ and $\mathbf{a}_N = \frac{1}{\sqrt{N}}\mathbf{1}_N$, the convergence (26) shown at the bottom of the page, holds.

As $|e_N(\rho) - c\gamma| \xrightarrow{\text{a.s.}} 0$ when $N, n \to \infty$, we substitute $c\gamma$ for $e_N(\rho)$ in (26), giving

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \left| \frac{1}{N}\mathbf{1}_N^T \hat{\mathbf{C}}_{\text{ST}}^{-1} \left(\hat{\mathbf{C}}_{\text{ST}} - \rho\mathbf{I}_N \right) \hat{\mathbf{C}}_{\text{ST}}^{-1}\mathbf{1}_N - \frac{(1-\rho) - (1-\rho)^2 c}{\gamma} \right.$$
$$\left. \times \frac{\gamma^2}{\gamma^2 - \beta(1-\rho)^2} \frac{1}{N}\mathbf{1}_N^T \mathbf{C}_N \left(\frac{1-\rho}{\gamma}\mathbf{C}_N + \rho\mathbf{I}_N \right)^{-2} \mathbf{1}_N \right| \xrightarrow{\text{a.s.}} 0.$$

With respect to the asymptotic equivalence in (24) and upon substituting $\hat{\gamma}_{\text{sc}}$ for γ/κ, we obtain

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \left| \frac{1}{\kappa N}\mathbf{1}_N^T \hat{\mathbf{C}}_{\text{ST}}^{-1} \mathbf{C}_N \hat{\mathbf{C}}_{\text{ST}}^{-1}\mathbf{1}_N - \frac{\hat{\gamma}_{\text{sc}}}{(1-\rho) - (1-\rho)^2 c_N} \right.$$
$$\left. \times \frac{1}{N}\mathbf{1}_N^T \hat{\mathbf{C}}_{\text{ST}}^{-1} \left(\hat{\mathbf{C}}_{\text{ST}} - \rho\mathbf{I}_N \right) \hat{\mathbf{C}}_{\text{ST}}^{-1}\mathbf{1}_N \right| \xrightarrow{\text{a.s.}} 0.$$

Thus we obtain the consistent estimator of $\frac{1}{\kappa}\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho))$ in Theorem 2.

APPENDIX E
PROOF OF COROLLARY 1

According to Theorem 2, we have

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \left| N\hat{\sigma}_{\text{sc}}^2(\rho) - \frac{1}{\kappa}N\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho)) \right| \xrightarrow{\text{a.s.}} 0.$$

Then, the following holds true

$$N\hat{\sigma}_{\text{sc}}^2(\rho^o) \le N\hat{\sigma}_{\text{sc}}^2(\hat{\mathbf{h}}_{\text{ST}}(\rho^*))$$

$$\frac{1}{\kappa}N\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho^*)) \le \frac{1}{\kappa}N\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho^o))$$

$$N\hat{\sigma}_{\text{sc}}^2(\rho^o) - \frac{1}{\kappa}N\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho^o)) \le \sup_{\rho \in \mathcal{R}_\varepsilon} \left| N\hat{\sigma}_{\text{sc}}^2(\rho) - \frac{1}{\kappa}N\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho)) \right|$$
$$\xrightarrow{\text{a.s.}} 0$$

$$N\hat{\sigma}_{\text{sc}}^2(\rho^*) - \frac{1}{\kappa}N\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho^*)) \le \sup_{\rho \in \mathcal{R}_\varepsilon} \left| N\hat{\sigma}_{\text{sc}}^2(\rho) - \frac{1}{\kappa}N\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho)) \right|$$
$$\xrightarrow{\text{a.s.}} 0.$$

$$\frac{\text{d}}{\text{d}w}\left[\frac{1}{N}\mathbf{1}_N^T \mathbf{C}_N^{-1/2} \mathbf{J}_N \mathbf{C}_N^{-1/2}\mathbf{1}_N \right]\bigg|_{w=0} = \frac{-\frac{1}{N}\mathbf{1}_N^T \mathbf{C}_N^{1/2}\left(\frac{k}{\gamma+\tilde{e}_N(0)k}\mathbf{C}_N + \rho\mathbf{I}_N \right)^{-2} \mathbf{C}_N^{1/2}\mathbf{1}_N}{1 - \frac{k^2}{(\gamma+\tilde{e}_N(0)k)^2}\frac{1}{n}\text{tr}\left[\mathbf{C}_N^2\left(\left(\frac{k}{\gamma+\tilde{e}_N(0)k} \right)\mathbf{C}_N + \rho\mathbf{I}_N \right)^{-2} \right]} \quad (23)$$

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \left| \frac{1}{N}\mathbf{1}_N^T \hat{\mathbf{C}}_{\text{ST}}^{-1} \left(\hat{\mathbf{C}}_{\text{ST}} - \rho\mathbf{I}_N \right) \hat{\mathbf{C}}_{\text{ST}}^{-1}\mathbf{1}_N - \frac{k\gamma}{(\gamma+e_N(\rho)k)^2} \cdot \frac{\frac{1}{N}\mathbf{1}_N^T \mathbf{C}_N\left(\frac{k}{(\gamma+e_N(\rho)k)}\mathbf{C}_N + \rho\mathbf{I}_N \right)^{-2}\mathbf{1}_N}{1 - \frac{k^2}{(\gamma+e_N(\rho)k)^2}\frac{1}{n}\text{tr}\left[\mathbf{C}_N^2\left(\frac{k}{(\gamma+e_N(\rho)k)}\mathbf{C}_N + \rho\mathbf{I}_N \right)^{-2} \right]} \right| \xrightarrow{\text{a.s.}} 0$$

$$(26)$$

These four relations together ensure that

$$|N\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho^o)) - N\sigma^2(\hat{\mathbf{h}}_{\text{ST}}(\rho^*))| \xrightarrow{\text{a.s.}} 0.$$

APPENDIX F
PROOF OF PROPOSITION 1

Denote

$$|\widetilde{m}_{N,j} - m_{N,j}| = |-A - B + C - D|,$$

where $1 \le j \le n$ and

$$A \triangleq \frac{1}{N}\frac{1}{n}\frac{\sqrt{\boldsymbol{\tau}}^T}{\sqrt{\tau_j}}\mathbf{Z}_N^T\mathbf{M}_{N,j}\mathbf{z}_j \quad B \triangleq \frac{1}{N}\mathbf{z}_j^T\mathbf{M}_{N,j}\frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_j}}$$

$$C \triangleq \frac{1}{N}\frac{1}{n}\frac{\sqrt{\boldsymbol{\tau}}^T}{\sqrt{\tau_j}}\mathbf{Z}_N^T\mathbf{M}_{N,j}\frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_j}}$$

$$D \triangleq \frac{1}{N}\tilde{\mathbf{z}}_j^T\widetilde{\mathbf{M}}_{N,j}\frac{1-\rho_n}{1-(1-\rho_n)c_N}\frac{1}{n}\left(\frac{1}{n^2}\sum_{t\ne j}\mathbf{Z}_N\frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_t}}\frac{\sqrt{\boldsymbol{\tau}}^T}{\sqrt{\tau_t}}\mathbf{Z}_N^T\right.$$

$$\left.- \frac{1}{n}\sum_{t\ne j}\mathbf{z}_t\frac{\sqrt{\boldsymbol{\tau}}^T}{\sqrt{\tau_t}}\mathbf{Z}_N^T - \frac{1}{n}\sum_{t\ne j}\mathbf{Z}_N\frac{\sqrt{\boldsymbol{\tau}}}{\sqrt{\tau_t}}\mathbf{z}_t^T\right)\mathbf{M}_{N,j}\tilde{\mathbf{z}}_j.$$

We wish to prove that

$$E[|\widetilde{m}_{N,j} - m_{N,j}|^p] \le \frac{K_p}{N^p} \tag{27}$$

for some constants $p \ge 1$, where K_p depends on p but not on N. Then, taking $p \ge 2$, along with the union bound, the Markov inequality, and the Borel-Cantelli lemma, completes the proof of Proposition 1.

Using the Minkowski inequality, we have

$$E[|\widetilde{m}_{N,j} - m_{N,j}|^p] \le$$
$$(E^{1/p}[|A|^p] + E^{1/p}[|B|^p] + E^{1/p}[|C|^p] + E^{1/p}[|D|^p])^p.$$

Thus, to prove (27), it is enough to show that $E[|A|^p] \le \frac{K_{pA}}{N^p}$, $E[|B|^p] \le \frac{K_{pB}}{N^p}$, $E[|C|^p] \le \frac{K_{pC}}{N^p}$ and $E[|D|^p] \le \frac{K_{pD}}{N^p}$.

Moments of $|A|$, $|B|$ and $|C|$:

Start by noting that

$$E[|A|^p] = \frac{1}{N^p}\frac{1}{n^p}E\left[\left|\mathbf{z}_j^T\mathbf{M}_{N,j}\left(\mathbf{z}_j\mathbf{z}_j^T + \mathbf{z}_j\frac{\sqrt{\boldsymbol{\tau}^{(j)}}^T}{\sqrt{\tau_j}}\mathbf{Z}_N^{(j)T}\right.\right.\right.$$

$$\left.\left.\left.+ \mathbf{Z}_N^{(j)}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}}{\sqrt{\tau_j}}\mathbf{z}_j^T + \mathbf{Z}_N^{(j)}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}}{\sqrt{\tau_j}}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}^T}{\sqrt{\tau_j}}\mathbf{Z}_N^{(j)T}\right)\mathbf{M}_{N,j}\mathbf{z}_j\right|^{p/2}\right]$$

$$\overset{(a)}{\le} A_1 + A_2 + A_3 + A_4,$$

where (a) follows from Jensen's inequality and

$$A_1 = \frac{4^{p/2-1}}{N^p n^p}E\left[\left|\mathbf{z}_j^T\mathbf{M}_{N,j}\mathbf{z}_j\mathbf{z}_j^T\mathbf{M}_{N,j}\mathbf{z}_j\right|^{p/2}\right]$$

$$A_2 = \frac{4^{p/2-1}}{N^p n^p}E\left[\left|\mathbf{z}_j^T\mathbf{M}_{N,j}\mathbf{Z}_N^{(j)}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}}{\sqrt{\tau_j}}\right.\right.$$

$$\left.\left.\times \frac{\sqrt{\boldsymbol{\tau}^{(j)}}^T}{\sqrt{\tau_j}}\mathbf{Z}_N^{(j)T}\mathbf{M}_{N,j}\mathbf{z}_j\right|^{p/2}\right]$$

$$A_3 = \frac{4^{p/2-1}}{N^p n^p}E\left[\left|\mathbf{z}_j^T\mathbf{M}_{N,j}\mathbf{Z}_N^{(j)}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}}{\sqrt{\tau_j}}\mathbf{z}_j^T\mathbf{M}_{N,j}\mathbf{z}_j\right|^{p/2}\right]$$

$$A_4 = \frac{4^{p/2-1}}{N^p n^p}E\left[\left|\mathbf{z}_j^T\mathbf{M}_{N,j}\mathbf{z}_j\frac{\sqrt{\boldsymbol{\tau}^{(j)}}^T}{\sqrt{\tau_j}}\mathbf{Z}_N^{(j)T}\mathbf{M}_{N,j}\mathbf{z}_j\right|^{p/2}\right].$$

For term A_1,

$$A_1 = \frac{1}{N^p}\frac{1}{n^p}\cdot 4^{p/2-1}E\left[\left|\mathbf{z}_j^T\mathbf{M}_{N,j}\mathbf{z}_j\right|^p\right]$$

$$\le \frac{1}{N^p}\frac{1}{n^p}\cdot 4^{p/2-1}E\left[\|\mathbf{y}_j\|^{2p}\|\mathbf{M}_{N,j}\|^p\|\mathbf{C}_N\|^p\right].$$

Note also that

$$\|\mathbf{M}_{N,j}\|^p \le \frac{1}{(\gamma(\rho_n)+\ell)^p\rho_n^p},$$

and by Minkowsky's inequality,

$$E[\|\mathbf{y}_j\|^{2p}] = E(\sum_{i=1}^N y_{i,j}^2)^p \le N^p E|y_{1,j}|^{2p} \le K_p N^p.$$

Thus

$$A_1 \le \frac{1}{N^p}\frac{K_p\|\mathbf{C}_N\|^p 4^{p/2-1}c_N^p}{(\gamma(\rho_n)+\ell)^p\rho_n^p} \le \frac{K_{pA_1}}{N^p}.$$

Now consider A_2:

$$A_2 \overset{(a)}{\le} \frac{2^{3p/2-3}}{N^p n^p}\left(E\left[\left|\mathbf{z}_j^T\mathbf{Q}_N\mathbf{z}_j - \text{tr}(\mathbf{Q}_N)\right|^{p/2}\right] + E\left[\left|\text{tr}(\mathbf{Q}_N)\right|^{p/2}\right]\right)$$

$$\overset{(b)}{\le} \frac{K_p}{N^p n^p}E\left[\left(E^{p/4}|z_{1,j}|^4\text{tr}[\mathbf{Q}_N\mathbf{Q}_N^T]\right)^{p/4}\right.$$

$$\left.+ E|z_{1,j}|^p\text{tr}\left[(\mathbf{Q}_N\mathbf{Q}_N^T)^{p/4}\right] + E|\text{tr}(\mathbf{Q}_N)|^{p/2}\right]$$

$$= \frac{K_p}{N^p n^p}\left(E^{p/4}|z_{1,j}|^4 + E|z_{1,j}|^p + 1\right)E\|\mathbf{M}_{N,j}\mathbf{Z}_N^{(j)}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}}{\sqrt{\tau_j}}\|^p$$

$$\le \frac{K_p}{N^p n^p}\left(E^{p/4}|z_{1,j}|^4 + E|z_{1,j}|^p + 1\right)$$

$$\times E\left(\|\mathbf{M}_{N,j}\|^p\|\mathbf{C}_N\|^{p/2}\left\|\mathbf{Y}_N^{(j)}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}}{\sqrt{\tau_j}}\right\|^p\right)$$

$$\le \frac{1}{N^p}\frac{K_p K_{\mathbf{C}_N}^{p/2}\left(E^{p/4}|z_{1,j}|^4 + E|z_{1,j}|^p + 1\right)}{(\gamma(\rho_n)+\ell)^p\rho_n^p}E\left\|\frac{1}{n}\mathbf{Y}_N^{(j)}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}}{\sqrt{\tau_j}}\right\|^p$$

$$\le K_{pA_2}/N^p,$$

where $\mathbf{Q}_N = \mathbf{M}_{N,j}\mathbf{Z}_N^{(j)}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}}{\sqrt{\tau_j}}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}^T}{\sqrt{\tau_j}}\mathbf{Z}_N^{(j)T}\mathbf{M}_{N,j}$, (a) follows from Jensen's inequality and (b) follows from the trace lemma [44, Lemma B.26].

For A_3,

$$A_3 \le \frac{4^{p/2-1}}{N^p n^p}E^{1/2}\left[\left|\mathbf{z}_j^T\mathbf{M}_{N,j}\mathbf{Z}_N^{(j)}\frac{\sqrt{\boldsymbol{\tau}^{(j)}}}{\sqrt{\tau_j}}\right|^p\right]E^{1/2}\left[\left|\mathbf{z}_j^T\mathbf{M}_{N,j}\mathbf{z}_j\right|^p\right].$$

As we have $A_1 \le K_{pA_1}/N^p$ and $A_2 \le K_{pA_2}/N^p$, we obtain $A_3 \le K_{pA_3}/N^p$. Following the same reasoning as for A_3, we also get $A_4 \le K_{pA_4}/N^p$.

Therefore, we obtain

$$E[|A|^p] \le A_1 + A_2 + A_3 + A_4 \le K_{pA}/N^p.$$

The same reasoning holds for $E[|B|^p]$, giving $E[|B|^p] \leq K_{pB}/N^p$. As for the moments of $|C|$, it is similar to how we dealt with A_1:

$$
\begin{aligned}
E[|C|^p] &\leq \frac{1}{N^p} E\left[\left\|\frac{1}{n}\mathbf{Y}_N \frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right\|^{2p} \|\mathbf{C}_N\|^p \|\mathbf{M}_{N,j}^p\|\right] \\
&\leq \frac{\|\mathbf{C}_N\|^p}{N^p(\gamma(\rho_n)+\ell)^p \rho_n^p} E\left[\left\|\frac{1}{n}\mathbf{Y}_N \frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right\|^{2p}\right] \\
&\leq \frac{K_{pC}}{N^p}.
\end{aligned}
$$

A. Moments of $|D|$

We denote $\frac{1}{\sqrt{\tau}} = \left(\frac{1}{\sqrt{\tau_1}}, \ldots, \frac{1}{\sqrt{\tau_n}}\right)$ and rewrite D as

$$D = D_1 + D_2 + D_3$$

with D_1, D_2 and D_3 the bottom of the page. We aim to prove $E[|D|^p] \leq K_{pD}/N^p$, which is achieved by proving $E[|D_1|^p] \leq K_{pD_1}/N^p$, $E[|D_2|^p] \leq K_{pD_2}/N^p$ and $E[|D_3|^p] \leq K_{pD_3}/N^p$.

Let's first analyze D_1 with the analysis of D_2 and D_3 following similarly. We obtain

$$
\begin{aligned}
E[|D_1|^p] &\overset{(a)}{\leq} \frac{1}{N^p}\left(\frac{1-\rho_n}{1-(1-\rho_n)c_N}\right)^p E^{1/2}[|D_{1a}|^{2p}] E^{1/2}[|D_{1b}|^{2p}] \\
&\overset{(b)}{\leq} \frac{1}{N^p}\left(\frac{1-\rho_n}{1-(1-\rho_n)c_N}\right)^p \\
&\quad \times (E^{1/2p}[|D_{1c}|^{2p}] + E^{1/2p}[|D_{1d}|^{2p}])^p E^{1/2}[|D_{1b}|^{2p}] \quad (28)
\end{aligned}
$$

where (a) follows from the Cauchy-Schwarz inequality, (b) follows from Minkowsky's inequality, and

$$
\begin{aligned}
D_{1a} &= \left(\mathbf{z}_j - \frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right)^T \widetilde{\mathbf{M}}_{N,j}\frac{1}{n^2}\frac{1}{\sqrt{\tau}}^T\frac{1}{\sqrt{\tau}}\mathbf{Z}_N\sqrt{\tau} \\
D_{1b} &= \frac{1}{n}\sqrt{\tau}^T\mathbf{Z}_N^T\mathbf{M}_{N,j}\left(\mathbf{z}_j - \frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right)
\end{aligned}
$$

$$
\begin{aligned}
D_{1c} &= \mathbf{z}_j^T\widetilde{\mathbf{M}}_{N,j}\frac{1}{n^2}\frac{1}{\sqrt{\tau}}^T\frac{1}{\sqrt{\tau}}\mathbf{Z}_N\sqrt{\tau} \\
D_{1d} &= \left(\frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right)^T\widetilde{\mathbf{M}}_{N,j}\frac{1}{n^2}\frac{1}{\sqrt{\tau}}^T\frac{1}{\sqrt{\tau}}\mathbf{Z}_N\sqrt{\tau}.
\end{aligned}
$$

Our aim is to prove that $E[|D_{1b}|^{2p}] \leq K_{pb}$, $E[|D_{1c}|^{2p}] \leq K_{pc}$ and $E[|D_{1d}|^{2p}] \leq K_{pd}$. Following the analysis of $E[|A|^p]$ and $E[|C|^p]$, we obtain $E[|D_{1b}|^{2p}] \leq K_{pb}$. For $E[|D_{1d}|^{2p}]$, we have

$$
\begin{aligned}
&E[|D_{1d}|^{2p}] \\
&\leq \frac{1}{n^{2p}\tau_j^p}\left\|\frac{1}{\sqrt{\tau}}\right\|^{4p} E\left[\left\|\widetilde{\mathbf{M}}_{N,j}\right\|^{2p}\|\mathbf{C}_N\|^{2p}\left\|\frac{1}{n}\mathbf{Y}_N\sqrt{\tau}\right\|^{4p}\right] \\
&\leq \frac{\|\mathbf{C}_N\|^{2p}}{n^{2p}\tau_j^p(\gamma(\rho_n)+\ell)^{2p}\rho_n^{2p}}\left\|\frac{1}{\sqrt{\tau}}\right\|^{4p} E\left\|\frac{1}{n}\mathbf{Y}_N\sqrt{\tau}\right\|^{4p} \\
&\leq K_{pd}.
\end{aligned}
$$

Let us now establish the inequality for D_{1c}. We can see that $\mathbf{z}_j$ is not independent of $\widetilde{\mathbf{M}}_{N,j}$, thus we cannot follow the same procedure as for our analysis of A to determine the order of $E[|D_{1c}|^{2p}]$. Instead, we divide $\widetilde{\mathbf{M}}_{N,j}$ into two parts, one that is independent of $\mathbf{z}_j$ and the other the remainder.

We first write $\sum_{t\neq j}\tilde{\mathbf{z}}_t\tilde{\mathbf{z}}_t^T = \mathbf{E} + \mathbf{F}$, where $\mathbf{E}$ and $\mathbf{F}$ are defined as shown at the bottom of the page. Note that $\mathbf{E}$ is independent of $\mathbf{z}_j$ and $\mathbf{F}$ is not. Then D_{1c} can be rewritten as shown in (29) at the top of the next page.

Using Jensen's inequality,

$$E[|D_{1c}|^{2p}] \leq 2^{2p-1}(E[|G|^{2p}] + E[|H|^{2p}]),$$

where G and H are the two terms on the RHS of (29). Next we can use the same technique as used in Appendix F-A to prove that $E[|G|^{2p}] \leq K_{pG}$ and $E[|H|^{2p}] \leq K_{pH}$. Therefore, we obtain $E[|D_{1c}|^{2p}] \leq K_{pc}$.

Thus far, we have proven that $E[|D_{1b}|^{2p}] \leq K_{pb}$, $E[|D_{1c}|^{2p}] \leq K_{pc}$, and $E[|D_{1d}|^{2p}] \leq K_{pd}$. Coming back to (28), we obtain $E[|D_1|^p] \leq K_{pD_1}/N^p$.

$$
\begin{aligned}
D_1 &= \frac{1}{N}\left(\mathbf{z}_j - \frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right)^T\widetilde{\mathbf{M}}_{N,j}\frac{1-\rho_n}{1-(1-\rho_n)c_N}\frac{1}{n}\left(\frac{1}{n^2}\frac{1}{\sqrt{\tau}}^T\frac{1}{\sqrt{\tau}}\mathbf{Z}_N\sqrt{\tau}\sqrt{\tau}^T\mathbf{Z}_N^T\right)\mathbf{M}_{N,j}\left(\mathbf{z}_j - \frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right) \\
D_2 &= \frac{1}{N}\left(\mathbf{z}_j - \frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right)^T\widetilde{\mathbf{M}}_{N,j}\frac{1-\rho_n}{1-(1-\rho_n)c_N}\frac{1}{n}\left(-\frac{1}{n}\mathbf{Z}_N\frac{1}{\sqrt{\tau}}\sqrt{\tau}^T\mathbf{Z}_N^T\right)\mathbf{M}_{N,j}\left(\mathbf{z}_j - \frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right) \\
D_3 &= \frac{1}{N}\left(\mathbf{z}_j - \frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right)^T\widetilde{\mathbf{M}}_{N,j}\frac{1-\rho_n}{1-(1-\rho_n)c_N}\frac{1}{n}\left(-\frac{1}{n}\mathbf{Z}_N\sqrt{\tau}\frac{1}{\sqrt{\tau}}^T\mathbf{Z}_N^T\right)\mathbf{M}_{N,j}\left(\mathbf{z}_j - \frac{1}{n}\mathbf{Z}_N\frac{\sqrt{\tau}}{\sqrt{\tau_j}}\right).
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{E} &= \mathbf{Z}_N^{(j)}\mathbf{Z}_N^{(j)^T} - \frac{1}{n}\mathbf{Z}_N^{(j)}\frac{1}{\sqrt{\tau^{(j)}}}\left(\mathbf{Z}_N^{(j)}\sqrt{\tau^{(j)}}\right)^T - \frac{1}{n}\mathbf{Z}_N^{(j)}\sqrt{\tau^{(j)}}\left(\mathbf{Z}_N^{(j)}\frac{1}{\sqrt{\tau}}\right)^T + \frac{1}{n^2}\left(\frac{1}{\sqrt{\tau^{(j)}}}\right)^T\frac{1}{\sqrt{\tau^{(j)}}}\mathbf{Z}_N^{(j)}\sqrt{\tau^{(j)}}\left(\mathbf{Z}_N^{(j)}\sqrt{\tau^{(j)}}\right)^T \\
\mathbf{F} &= -\frac{1}{n}\mathbf{Z}_N^{(j)}\frac{1}{\sqrt{\tau^{(j)}}}\sqrt{\tau_j}\mathbf{z}_j^T - \frac{1}{n}\sqrt{\tau_j}\mathbf{z}_j\left(\mathbf{Z}_N^{(j)}\frac{1}{\sqrt{\tau^{(j)}}}\right)^T + \frac{\tau_j}{n^2}\left(\frac{1}{\sqrt{\tau^{(j)}}}\right)^T\frac{1}{\sqrt{\tau^{(j)}}}\mathbf{z}_j\mathbf{z}_j^T + \frac{1}{n^2}\left(\frac{1}{\sqrt{\tau^{(j)}}}\right)^T\frac{1}{\sqrt{\tau^{(j)}}}\mathbf{Z}_N^{(j)}\sqrt{\tau^{(j)}}\sqrt{\tau_j}\mathbf{z}_j^T \\
&\quad + \frac{1}{n^2}\left(\frac{1}{\sqrt{\tau^{(j)}}}\right)^T\frac{1}{\sqrt{\tau^{(j)}}}\sqrt{\tau_j}\mathbf{z}_j\left(\mathbf{Z}_N^{(j)}\sqrt{\tau^{(j)}}\right)^T.
\end{aligned}
$$

$$D_{1c} = \mathbf{z}_j \left(\frac{1 - \rho_n}{1 - (1 - \rho_n)c_N} \frac{1}{n} \mathbf{E} + (\gamma(\rho_n) + \ell)\rho_n \mathbf{I}_N \right)^{-1} \frac{1}{n^2} \frac{1}{\sqrt{\tau}}^T \frac{1}{\sqrt{\tau}} \mathbf{Z}_N \sqrt{\tau} + \mathbf{z}_j \widetilde{\mathbf{M}}_{N,j} \left(-\frac{1 - \rho_n}{1 - (1 - \rho_n)c_N} \frac{1}{n} \mathbf{F} \right)$$

$$\times \left(\frac{1 - \rho_n}{1 - (1 - \rho_n)c_N} \frac{1}{n} \mathbf{E} + (\gamma(\rho_n) + \ell)\rho_n \mathbf{I}_N \right)^{-1} \frac{1}{n^2} \frac{1}{\sqrt{\tau}}^T \frac{1}{\sqrt{\tau}} \mathbf{Z}_N \sqrt{\tau}. \tag{29}$$

Following similar arguments to our analysis of $E[|D_1|^p]$, we can also obtain $E[|D_2|^p] \leq K_{pD_2}/N^p$ and $E[|D_3|^p] \leq K_{pD_3}/N^p$. As $D = D_1 + D_2 + D_3$, by Minkowsky's inequality, we obtain $E[|D|^p] \leq K_{pD}/N^p$.

REFERENCES

[1] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, Mar. 1952.

[2] R. C. Merton, "On estimating the expected return on the market: An exploratory investigation," *J. Finan. Econ.*, vol. 8, no. 4, pp. 323–361, Dec. 1980.

[3] R. Jagannathan and T. Ma, "Risk reduction in large portfolios: Why imposing the wrong constraints helps," *J. Finance*, vol. 58, no. 4, pp. 1651–1684, Aug. 2003.

[4] N. E. Karoui *et al.*, "High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation," *Ann. Statist.*, vol. 38, no. 6, pp. 3487–3566, 2010.

[5] N. E. Karoui, "On the realized risk of high-dimensional Markowitz portfolios," *SIAM J. Finan. Math.*, vol. 4, no. 1, pp. 737–783, 2013.

[6] Z. Bai, H. Liu, and W.-K. Wong, "Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory," *Math. Finance*, vol. 19, no. 4, pp. 639–667, 2009.

[7] J.-P. Bouchaud and M. Potters, "Financial applications of random matrix theory: A short review," in *The Oxford Handbook of Random Matrix Theory*, G. Akemann, J. Baik, and P. D. Francesco, Eds. Oxford, U.K.: Oxford Univ. Press, 2011, ch. 40.

[8] L. K. Chan, J. Karceski, and J. Lakonishok, "On portfolio optimization: Forecasting covariances and choosing the risk model," *Rev. Finan. Stud.*, vol. 12, no. 5, pp. 937–974, 1999.

[9] J. Fan, Y. Fan, and J. Lv, "High dimensional covariance matrix estimation using a factor model," *J. Econometr.*, vol. 147, no. 1, pp. 186–197, Nov. 2008.

[10] O. Ledoit and M. Wolf, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *J. Empir. Finance*, vol. 10, no. 5, pp. 603–621, Dec. 2003.

[11] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivar. Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.

[12] O. Ledoit and M. Wolf, Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks 2014, SSRN 2383361.

[13] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, "Noise dressing of financial correlation matrices," *Phys. Rev. Lett.*, vol. 83, no. 7, p. 1467, Aug. 1999.

[14] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud, "Random matrix theory and financial correlations," *Int. J. Theoret. Appl. Finance*, vol. 3, no. 3, pp. 391–397, Jul. 2000.

[15] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, "Random matrix approach to cross correlations in financial data," *Phys. Rev.*, vol. 65, no. 6, p. 066126, Jun. 2002.

[16] V. Dahirel *et al.*, "Coordinate linkage of HIV evolution reveals regions of immunological vulnerability," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 28, pp. 11530–11535, Jul. 2011.

[17] A. Quadeer, R. Louie, K. Shekhar, A. Chakraborty, I. Hsing, and M. McKay, "Statistical linkage of mutations in the non-structural proteins of hepatitis C virus exposes targets for immunogen design," *J. Virol.*, vol. 88, no. 13, pp. 7628–7644, Apr. 2014.

[18] D. L. Donoho, M. Gavish, and I. M. Johnstone, "Optimal shrinkage of eigenvalues in the spiked covariance model," 2013 [Online]. Available: ArXiv:1311.0851

[19] J. Fan, J. Zhang, and K. Yu, "Vast portfolio selection with gross-exposure constraints," *J. Amer. Statist. Assoc.*, vol. 107, no. 498, pp. 592–606, 2012.

[20] V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal, "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," *Manage. Sci.*, vol. 55, no. 5, pp. 798–812, Mar. 2009.

[21] R. Cont, "Empirical properties of asset returns: Stylized facts and statistical issues," *Quant. Finance*, vol. 1, no. 2, pp. 223–236, 2001.

[22] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 1964.

[23] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 1987.

[24] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Ann. Statist.*, vol. 4, no. 1, pp. 51–67, 1976.

[25] J. T. Kent and D. E. Tyler, "Redescending M-estimates of multivariate location and scatter," *Ann. Statist.*, vol. 19, no. 4, pp. 2102–2119, 1991.

[26] R. Couillet, F. Pascal, and J. W. Silverstein, "Robust estimates of covariance matrices in the large dimensional regime," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7269–7278, Sep. 2014.

[27] R. Couillet, F. Pascal, and J. W. Silverstein, "The random matrix regime of Maronna's M-estimator with elliptically distributed samples," *J. Mult. Anal.*, vol. 139, pp. 56–78, 2015.

[28] T. Zhang, X. Cheng, and A. Singer, "Marchenko-Pastur law for Tyler's and Maronna's M-estimators," 2014, ArXiv:1401.3424 [Online]. Available: http://arxiv.org/abs/1401.3424

[29] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to STAP detection problem," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5640–5651, Nov. 2014.

[30] Y. Abramovich and N. K. Spencer, "Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, vol. 3, pp. 1105–1108.

[31] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4097–4107, Sep. 2011.

[32] R. Couillet and M. R. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *J. Mult. Anal.*, vol. 131, pp. 99–120, 2014.

[33] F. Rubio, X. Mestre, and D. P. Palomar, "Performance analysis and optimal selection of large minimum variance portfolios under estimation risk," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 337–350, Aug. 2012.

[34] V. DeMiguel, L. Garlappi, and R. Uppal, "Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?," *Rev. Finan. Stud.*, vol. 22, no. 5, pp. 1915–1953, 2009.

[35] A. C. MacKinlay and L. Pastor, "Asset pricing models: Implications for expected returns and portfolio selection," *Rev. Finan. Stud.*, vol. 13, no. 4, pp. 883–916, 2000.

[36] O. Ledoit and M. Wolf, "Robust performance hypothesis testing with the Sharpe ratio," *J, Empir. Finance*, vol. 15, no. 5, pp. 850–859, Dec. 2008.

[37] O. Ledoit and M. Wolf, "Robust performances hypothesis testing with the variance," *Wilmott*, vol. 2011, no. 55, pp. 86–89, 2011.

[38] D. N. Politis and J. P. Romano, "The stationary bootstrap," *J. Amer. Statist. Assoc.*, vol. 89, no. 428, pp. 1303–1313, 1994.

[39] D. Morales-Jimenez, R. Couillet, and M. R. McKay, "Large dimensional analysis of robust M-estimators of covariance with outliers," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5784–5797, Nov. 2015.

[40] P. Jorion, "Bayesian and CAPM estimators of the means: Implications for portfolio selection," *J. Bank. Finan.*, vol. 15, no. 3, pp. 717–727, Jun. 1991.

[41] V. K. Chopra and W. T. Ziemba, "The effect of errors in means, variances, and covariances on optimal portfolio choice," *J. Portfolio Manage.*, vol. 19, no. 2, pp. 6–11, Dec. 1993.

[42] Z. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 26, no. 1, pp. 316–345, 1998.

[43] L. Ahlfors, *Complex Analysis: An Introductin to the Theory of Analytic Functions of one Complex Variable*, 2nd ed. New York, NY, USA: McGraw-Hill, 1978.

[44] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, ser. Statistics, 2nd ed. New York, NY, USA: Springer-Verlag, 2009.

Romain Couillet received the M.Sc. degree in mobile communications from the Eurecom Institute and the M.Sc. degree in communication systems from Telecom ParisTech, France, in 2007. He received the Ph.D. degree from Supelec, France, November 2010.

From 2007 to 2010, he was with ST-Ericsson as an Algorithm Development Engineer on the Long Term Evolution Advanced project. He is currently an assistant professor with the Telecommunication Department of CentraleSupélec, France. His research topics are in random matrix theory applied to wireless communications, signal processing, and statistics.

Liusha Yang received the B.S. degree in communication engineering from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2012.

Currently, she is pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST). Her research interests are in random matrix theory, with applications in signal processing and financial engineering.

Matthew R. McKay received the Ph.D. degree from the University of Sydney, Australia.

He joined the Hong Kong University of Science and Technology (HKUST), where he is currently the Hari Harilela Associate Professor of Electronic and Computer Engineering. His research interests include communications, signal processing, random matrix theory, and associated applications. Most recently, he has developed a keen interest in the interdisciplinary areas of computational immunology and financial engineering.