Minimum Variance Portfolio Optimization with Robust Shrinkage Covariance Estimation

Liusha Yang*, Romain Couillet†, Matthew R. McKay*

*Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong
†Telecommunication Department, Supélec, France

Abstract—We study the design of portfolios under a minimum risk criterion. The performance of the optimized portfolio relies on the accuracy to the estimated covariance matrix of portfolio asset returns. For large portfolios, the sample size is often of similar order to the number of assets, and the traditional sample covariance matrix performs poorly. Additionally, financial market data often involve outliers and exhibit heavy-tails, which, if not correctly handled, may further corrupt the covariance estimation. We aim to address these problems by studying the performance of a hybrid covariance matrix estimator based on Tyler's robust M-estimator and on Ledoit-Wolf's shrinkage estimator. Employing recent results from random matrix theory, we develop a consistent estimator of a scaled version of the portfolio risk, based on which, the shrinkage intensity is directly optimized to minimize the risk. Our portfolio optimization method is shown via simulations to outperform existing methods both for synthetic data and for a real market data set from Hang Seng Index.

I. INTRODUCTION

The theory of portfolio optimization is generally associated with the classical mean-variance optimization framework of Markowitz [1]. The pitfalls of 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 argued [2] that estimates of the covariance matrix are more accurate than those of the expected returns. Thus, many studies concentrate on improving the performance of the global minimum variance portfolio (GMVP), which involves only the covariance matrix estimate.

The most often used covariance estimator is the well-known sample covariance matrix (SCM). However, it shows good performance only when the number of samples n is much larger than the dimension of the matrix N, or when the underlying multivariate distribution is known to be Gaussian. However, in financial data analysis, often neither of these assumptions holds true. More specifically, covariance estimates for portfolio optimization commonly involve few historical observations of sometimes up to a thousand assets. In addition, the return observations often exhibit heavy-tails and contain outliers [3], which strongly degrade the performance of SCM [4].

The field of robust covariance estimation intends to deal with samples that are non-Gaussian or contain outliers [5–8]. However, classical robust covariance estimators generally

require $n \gg N$ and do not perform well when $n \simeq N$. Recent works have therefore considered robust estimation in the $n \simeq N$ regime [4, 9, 10]. In these papers, a hybrid covariance estimator based on Tyler's robust M-estimation [6] and Ledoit-Wolf's shrinkage approach [11] was proposed. We shall refer to this "shrinkage-Tyler" estimator as ST for short in the remainder. It is capable to deal with the problems that are high dimensional, non-Gaussian and contain outliers. In [12], the authors show, by means of random matrix theory, that, in the large n, N regime, the estimators [4, 9] are essentially the same and can be analyzed thanks to their proximity to well known random matrix models. An application to Frobenius-norm distance minimization is provided to optimally select the shrinkage parameter in [9, 12]. However, this is suboptimal in terms of minimizing the portfolio risk under the GMVP framework.

This paper aims at designing a novel covariance estimation technique for optimizing large portfolios in the presence of heavy-tailed data or outliers, and under practical conditions in which the number of observed samples is of similar order to the number of assets in the portfolio. This covariance estimation approach is based on the ST estimator with risk-minimizing shrinkage parameter.

To facilitate our new estimator design, we first characterize the out-of-sample performance of minimum variance portfolios based on ST by analyzing the convergence of the achieved realized risk in the double limit regime, i.e., $N, n \to \infty$, 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 this) that is defined in terms of only the observed returns. Based on this, we obtain our risk-optimized ST covariance estimation, and thus our optimized portfolio design.

The proposed portfolio selection is shown to achieve superior performance over other competing methods in minimizing the realized portfolio risk under the GMVP framework. These results are established through Monte Carlo simulations with samples distributed within the elliptical family, as well as through an example real data test using historical (daily) stock returns from the Hang Seng Index (HSI).

II. DATA MODEL AND PROBLEM FORMULATION

We consider a time series comprising logarithmic returns of N financial assets with n independent and identically

distributed (i.i.d.) observations. It is generated by the following model:

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

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 $\|\mathbf{y}_t\|^2 = N$. It is assumed that $\boldsymbol{\mu}$ and $\mathbf{C}_N$ are time-invariant. Denote $\mathbf{z}_t = \mathbf{C}_N^{\frac{1}{2}}\mathbf{y}_t$. We assume that $\sqrt{\tau_t}$ and $\mathbf{z}_t$ are statistically independent. The model (1) for $\mathbf{x}_t$ is quite general, embracing in particular the class of elliptical distributions, including the multivariate normal distribution, exponential distribution as well as the multivariate Student-T distribution as special cases. The capability of modeling heavy-tails makes this appealing in simulating financial stock return data [13]. It is worth mentioning that in reality the observed return $\mathbf{x}_t$ may not be i.i.d.; however this assumption leads to tractable design solutions, and it is a commonly used approximation [14].

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, such that $\mathbf{h}^T \mathbf{1}_N = 1$. Then the portfolio variance (or risk) over the investment period of interest is given by $\sigma^2(\mathbf{h}) = \mathbf{h}^T \mathbf{C}_N \mathbf{h}$. Accordingly, the GMVP selection problem can be formulated as the following quadratic optimization problem with a linear constraint:

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

The solution to (2) is $\mathbf{h}_{\mathrm{GMVP}} = \frac{\mathbf{C}_N^{-1}\mathbf{1}_N}{\mathbf{1}_N^T \mathbf{C}_N^{-1}\mathbf{1}_N}$, and the corresponding portfolio risk is:

$$\sigma^2(\mathbf{h}_{\mathrm{GMVP}}) = \frac{1}{\mathbf{1}_N^T \mathbf{C}_N^{-1}\mathbf{1}_N}. \qquad (3)$$

Here, (3) represents the theoretical minimum portfolio risk bound, attained upon knowing the true 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 $\hat{\mathbf{C}}_N$ is given by

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

The quality of $\hat{\mathbf{h}}_{\mathrm{GMVP}}$, 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:

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

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

Note that, for the naïve uniform weighting portfolio 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 [15] 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 introduced in [4] and [10], which applied the Ledoit-Wolf shrinkage method to Tyler's robust M-estimator [6]. 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 $\rho \in (\max\{0, 1 - c_N^{-1}\}, 1]$:

$$\hat{\mathbf{C}}_{\mathrm{ST}}(\rho) = (1 - \rho)\frac{1}{n}\sum_{t=1}^{n} \frac{\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^T}{\frac{1}{N}\tilde{\mathbf{x}}_t^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\tilde{\mathbf{x}}_t} + \rho\mathbf{I}_N \qquad (4)$$

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 linear independent, $\hat{\mathbf{C}}_{\mathrm{ST}}(\rho)$ is unequivocally defined with probability one [4, 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:

$$\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)) = \frac{\mathbf{1}_N^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{C}_N \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{1}_N}{(\mathbf{1}_N^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{1}_N)^2}. \qquad (5)$$

Our goal is to optimize ρ such that (5) is minimum. However, this cannot be done directly since it involves $\mathbf{C}_N$, which is unobservable. Also note that the naïve approach of simply replacing $\mathbf{C}_N$ with $\hat{\mathbf{C}}_{\mathrm{ST}}(\rho)$ in (5) would yield the so-called "in-sample risk", which underestimates the realized portfolio risk, leading to overly-optimistic investment decisions [16].

We tackle this problem by obtaining a consistent estimator for a scaled version of the realized risk (5) as n and N go to infinity together. Classical asymptotic theory for time series analysis and mathematical statistics typically deals with the case of N fixed and $n \to \infty$. However, a double-limiting condition is of more relevance for large portfolio problems, where n is comparable to N. To this end, we first derive a deterministic asymptotic equivalent of (5) and then provide a consistent estimator based on this.

B. Asymptotic deterministic equivalence 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 positive random scalars τ_t, $t = 1, ..., n$ are i.i.d., and $E[\tau_1], E[\frac{1}{\tau_1}] < \infty$;

c. The spectral norm of $\mathbf{C}_N$ satisfies $\limsup_N \|\mathbf{C}_N\| < \infty$;

d. Denoting $\lambda_1 \leq ... \leq \lambda_N$ the ordered eigenvalues of $\mathbf{C}_N$, $\nu_N \triangleq \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\delta}_{\lambda_i}$ satisfies $\nu_N \to \nu$ weakly with $\nu \neq \boldsymbol{\delta}_0$ almost everywhere, where $\boldsymbol{\delta}_x$ is the Dirac measure at x.

$$\bar{\sigma}^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)) = \frac{1}{1 - \frac{\beta k(\rho)^2}{(\gamma + \alpha k(\rho))^2}} \frac{\mathbf{1}_N^T \left(\frac{k(\rho)}{(\gamma + \alpha k(\rho))} \mathbf{C}_N + \rho \mathbf{I}_N \right)^{-1} \mathbf{C}_N \left(\frac{k(\rho)}{(\gamma + \alpha k(\rho))} \mathbf{C}_N + \rho \mathbf{I}_N \right)^{-1} \mathbf{1}_N}{\left(\mathbf{1}_N^T \left(\frac{k(\rho)}{(\gamma + \alpha k(\rho))} \mathbf{C}_N + \rho \mathbf{I}_N \right)^{-1} \mathbf{1}_N \right)^2} \tag{6}$$

We also introduce some further definitions. First denote $k(\rho) = \frac{1-\rho}{1-(1-\rho)c}$. Define α as the unique positive solution to:

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

with γ the unique positive solution to the equation in γ

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

We also define

$$\beta = \frac{1}{n} \mathrm{tr} \left[\mathbf{C}_N^2 \left(\frac{k(\rho)}{(\gamma + \alpha k(\rho))} \mathbf{C}_N + \rho \mathbf{I}_N \right)^{-2} \right].$$

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 $\hat{\mathcal{R}}_\varepsilon = [\varepsilon + \max\{0, 1 - c^{-1}\}, 1]$. Then,*

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

where $\bar{\sigma}^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho))$ is defined in (6) at the top of this page.

The proof of Theorem 1 draws from the asymptotic properties of the ST estimator developed in [12], and follows a similar methodology to that in [16] to derive the asymptotic equivalent of the realized risk. The key challenges lie in first obtaining similar asymptotic properties in [12] with non-zero mean samples, and second, in contrast to [16], adopting the ST estimator instead of the SCM to the risk function, to study the convergence of the risk in the double limit. The details will be included in a full version of this paper.

Remark 1. *In Theorem 1 as well as the subsequent results, the set $\hat{\mathcal{R}}_\varepsilon$ excludes the region $[0, \varepsilon)$. This follows the same reasoning as that in [12]. The behavior of any solution $\hat{\mathbf{C}}_{\mathrm{ST}}(0)$ to (4) in the large N, n regime remains difficult to handle with our proof technique.*

C. Consistent estimation of the realized portfolio risk

Theorem 1 enables us to analyze the convergence of the realized portfolio risk in the double-limit regime based on the aforementioned ST. 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. Denote $\kappa = \int t\nu(dt)$. Based on observable data only, we can obtain an estimator of a scaled version of the realized portfolio

risk, $\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho))/\kappa$. We begin with the following lemmas that provide consistent estimators of γ/κ and α/κ.

Lemma 1. *Under Assumption 1,*

$$\sup_{\rho \in \hat{\mathcal{R}}_\varepsilon} |\hat{\gamma}_{\mathrm{sc}} - \gamma/\kappa| \xrightarrow{\mathrm{a.s.}} 0,$$

where

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

and $\mathcal{B} = \{t : \|\tilde{\mathbf{x}}_t\|^2 > \xi\}$ with $\xi > 0$ sufficiently small.

The proof of Lemma 1 is largely based on [12, Theorem 1]. The requirement of $t \in \mathcal{B}$ is essential. As we adapt the ST to the non-zero mean data by subtracting the empirical mean in (4), the asymptotic results in [12, Theorem 1] need to be treated carefully. They may not hold true for some $\tilde{x}_t$ that contain τ_t which go to zero in the limit.

Lemma 2. *Under Assumption 1,*

$$\sup_{\rho \in \hat{\mathcal{R}}_\varepsilon} |\hat{\alpha}_{\mathrm{sc}} - \alpha/\kappa| \xrightarrow{\mathrm{a.s.}} 0,$$

where

$$\hat{\alpha}_{\mathrm{sc}} = \frac{\hat{\gamma}_{\mathrm{sc}} \frac{1}{N} \mathrm{tr} \left[\mathbf{I}_N - \rho \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho) \right]}{\hat{k}(\rho) \left(\frac{n}{N} - \frac{1}{N} \mathrm{tr} \left[\mathbf{I}_N - \rho \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho) \right] \right)}$$

and $\hat{k}(\rho) = \frac{1-\rho}{1-(1-\rho)c_N}$.

Proof details are omitted here due to space limitation.

Lemma 1 and Lemma 2 provide consistent estimators of scaled γ and α, with a scale factor κ. They play an important role in developing a consistent estimator of $\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}})/\kappa$, which is provided in the following theorem.

Theorem 2. *Under Assumption 1,*

$$\sup_{\rho \in \hat{\mathcal{R}}_\varepsilon} \left| \hat{\sigma}_{\mathrm{sc}}^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)) - \frac{1}{\kappa} \sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)) \right| \xrightarrow{\mathrm{a.s.}} 0, \tag{7}$$

where

$$\hat{\sigma}_{\mathrm{sc}}^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)) = \frac{(\hat{\gamma}_{\mathrm{sc}} + \hat{\alpha}_{\mathrm{sc}} \hat{k}(\rho))^2}{\hat{k}(\rho) \hat{\gamma}_{\mathrm{sc}}} \frac{\mathbf{1}_N^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho) \left(\hat{\mathbf{C}}_{\mathrm{ST}}(\rho) - \rho \mathbf{I}_N \right) \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho) \mathbf{1}_N}{(\mathbf{1}_N^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho) \mathbf{1}_N)^2}.$$

The proof follows by virtue of the asymptotic equivalence (8) at the top of the next page (derived in the full paper), and by substituting $\hat{k}(\rho)$, $\hat{\gamma}_{\mathrm{sc}}$ and $\hat{\alpha}_{\mathrm{sc}}$ for $k(\rho)$, γ and α on the left hand side of (8).

From our derivation, we can only obtain a consistent estimator of a scaled version of the realized risk, i.e., $\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho))/\kappa$.

$$\sup_{\rho \in \hat{\mathcal{R}}_\varepsilon} \left| \frac{k(\rho)\gamma}{(\gamma + \alpha k(\rho))^2}\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)) - \frac{\mathbf{1}_N^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\left(\hat{\mathbf{C}}_{\mathrm{ST}}(\rho) - \rho\mathbf{I}_N\right)\hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{1}_N}{(\mathbf{1}_N^T \hat{\mathbf{C}}_{\mathrm{ST}}^{-1}(\rho)\mathbf{1}_N)^2} \right| \xrightarrow{\text{a.s.}} 0 \tag{8}$$

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

Remark 2. *As a corollary of Theorem 2, it can be shown that* $|\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho^o)) - \sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho^*))| \xrightarrow{\text{a.s.}} 0$ *where ρ^o is the minimizer of $\hat{\sigma}_{\mathrm{sc}}^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho))$ and ρ^* is the minimizer of $\sigma^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho))$.*

The problem of obtaining the best asset allocation, as measured by the minimum realized portfolio risk, is now reduced to minimizing $\hat{\sigma}_{\mathrm{sc}}^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho))$ with regard to ρ. This can be done with a simple numerical search.

D. Summary of the proposed portfolio optimization algorithm

Given n past return observations of N assets, our proposed algorithm to construct a portfolio with minimal risk involves three main steps:

1) Compute the optimized shrinkage parameter via a numerical search

$$\rho^o = \underset{\rho \in (\max\{0, 1-c_N^{-1}\}, 1]}{\arg\min} \hat{\sigma}_{\mathrm{sc}}^2(\hat{\mathbf{h}}_{\mathrm{ST}}(\rho)).$$

2) Form the risk-minimizing ST estimator $\hat{\mathbf{C}}_{\mathrm{ST}}^o$. It is the unique solution to the following fixed-point equation:

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

3) Construct the optimized portfolio:

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

IV. SIMULATION RESULTS

We use both synthetic data and real market data to show the performance of $\hat{\mathbf{C}}_{\mathrm{ST}}^o$ compared to other competing methods. We consider three ST estimators, each designed to calibrate ρ under a Frobenius norm minimization metric. These include $\hat{\mathbf{C}}_{\mathrm{P}}$, referred to as the Pascal estimate in [12]; $\hat{\mathbf{C}}_{\mathrm{C}}$, referred to as the the Chen estimate in [12]; and $\hat{\mathbf{C}}_{\mathrm{C2}}$, the oracle estimator in [9], which has the same structure as $\hat{\mathbf{C}}_{\mathrm{C}}$, but resorts to solving an approximate problem of minimizing the Frobenius distance to find the optimal shrinkage. Additionally, we consider the Ledoit-Wolf shrinkage estimator $\hat{\mathbf{C}}_{\mathrm{LW}}$ [11], and the Ledoit-Wolf shrinkage estimator $\hat{\mathbf{C}}_{\mathrm{F}}$ [16] designed to calibrate ρ based on the GMVP framework, as in the current paper.

A. Synthetic data simulations

The synthetic data are generated i.i.d. by a multivariate Student-T distribution, where $\sqrt{\tau_t}$ in (1) is a function of a Chi-square random variable $\sqrt{\tau_t} = \sqrt{d/\chi_d^2}$. The degree-of-freedom d is set to 3 and the dimensionality N is chosen to

be 200. The mean vector $\boldsymbol{\mu}$ can be set arbitrarily since it is subtracted with 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 [15]: $\mathbf{C}_N = \mathbf{b}\mathbf{b}^T\sigma^2 + \boldsymbol{\Sigma}$, where the standard deviation σ of the factor is 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 a diagonal matrix, whose elements are drawn from a uniform distribution with support $[0.1, 0.3]$ and are fixed throughout the simulations.

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 (3), the theoretical minimum portfolio risk. Comparing with other methods, our proposed estimator $\hat{\mathbf{C}}_{\mathrm{ST}}^o$ achieves the smallest realized risk in both cases of $n \leq N$ and $n > N$. We omit the realized risks achieved by $\mathbf{I}_N$ as they are uniformly more than 5 times larger than those achieved by other competing methods. Additionally, it is interesting to compare the optimized ρ of $\hat{\mathbf{C}}_{\mathrm{ST}}^o$ and $\hat{\mathbf{C}}_{\mathrm{P}}$. They are both solutions of (4) with ρ optimized under the metrics of minimizing the risk and minimizing the Frobenius distance respectively. Demonstrated in Fig. 2, the optimal shrinkage parameter varies under different metrics.
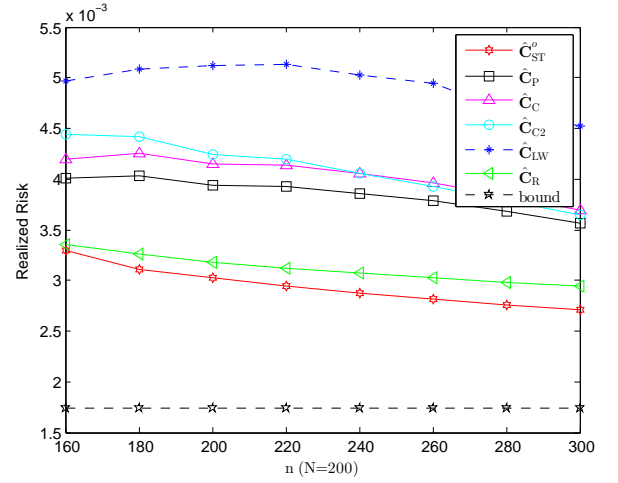

Fig. 1. The average realized portfolio risk of different covariance estimators in the GMVP framework using synthetic data.

B. Real market data simulations

Now, with real market data, we investigate the out-of-sample portfolio performance for different estimators. We consider the stocks conforming the HSI. In particular, we use the divident-adjusted daily closing prices downloaded from the Yahoo
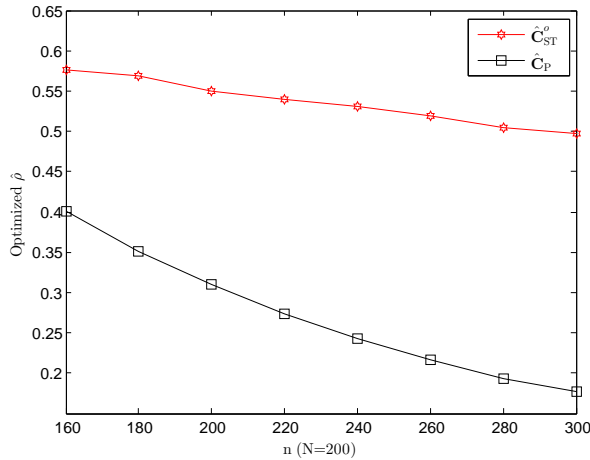
Fig. 2. The optimal shrinkage parameters of $\hat{\mathbf{C}}^o_{\mathrm{ST}}$ and $\hat{\mathbf{C}}_{\mathrm{P}}$ in the synthetic data simulation.
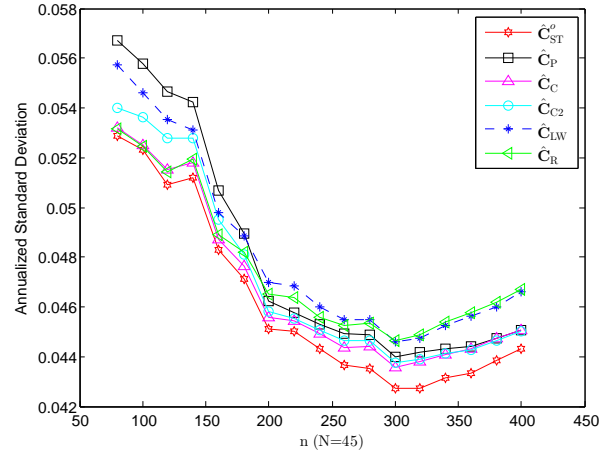

Fig. 3. Realized portfolio risks achieved out-of-sample over 736 days of HSI real market data (from Jan. 3, 2011 to Dec. 31, 2013) by a GMVP implemented using different covariance estimators.

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 the 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., $t-n$ 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 20 days. Next the window is shifted 20 days forward and the portfolio returns for another 20 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.

We can see from Fig. 3 that our proposed $\hat{\mathbf{C}}^o_{\mathrm{ST}}$ achieves the smallest realized risk. It outperforms the other competing methods over the entire span of estimation windows considered. The realized risk achieved by $\mathbf{I}_N$ is also omitted here because it is more than 2 times larger than those achieved by the competing methods. When the estimation window is too long (e.g., greater than 320 days), observe that the performance degrades. This is presumably due to a lack of stationarity in the data over such long durations. This is an interesting phenomena worthy of further investigation.

V. Conclusions

We proposed a novel minimum-variance portfolio optimization strategy, based on employing a robust shrinkage covariance estimator with the shrinkage parameter calibrated to minimize the realized portfolio risk. Our strategy was shown to be robust to the finite-sampling effects as well as to non-Gaussian samples. It was demonstrated that our approach outperformed other competing techniques in terms of the realized portfolio risk both for synthetic data and for real historical stock returns from the HSI.

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. Financ. Econ.*, vol. 8, no. 4, pp. 323–361, Dec. 1980.

[3] M. Loretan and P. C. Phillips, "Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets," *J. Empir. Finance*, vol. 1, no. 2, pp. 211–248, Jan. 1994.

[4] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its pplication to STAP detection problem," *Submitted for publication*, 2013. [Online]. Available: http://arxiv.org/abs/1311.6567

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

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

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

[8] 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.

[9] 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, Sept. 2011.

[10] 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)*, vol. 3, Honolulu, HI, Apr. 2007, pp. 1105–1108.

[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] R. Couillet and M. R. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *to appear in J. Mult. Anal.*, 2013. [Online]. Available: arXiv:1401.4083

[13] D. Ruppert, *Statistics and Data Analysis for Financial Engineering*. New York: Springer, 2010.

[14] 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.

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

[16] 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.