Sub-Gaussian Estimation of the Scatter Matrix in Ultra-High Dimensional Elliptical Factor Models with $2 + \varepsilon$th Moment

Yi Ding * Xinghua Zheng †

February 5, 2025

Abstract

We study the estimation of scatter matrices in elliptical factor models with $2 + \varepsilon$th moment. For such heavy-tailed data, robust estimators like the Huber-type estimator in Fan et al. (2018) cannot achieve a sub-Gaussian convergence rate. In this paper, we develop an idiosyncratic-projected self-normalization method to remove the effect of the heavy-tailed scalar component and propose a robust estimator of the scatter matrix that achieves the sub-Gaussian rate under an ultra-high dimensional setting. Such a high convergence rate leads to superior performance in estimating high-dimensional global minimum variance portfolios.

Keywords: High-dimension, elliptical model, factor model, scatter matrix, robust estimation

*Faculty of Business Administration, University of Macau, Macau. Research is supported in part by NSFC 72101226 of China, and SRG2023-00001-FBA, MYRG-GRG2023-00119-FBA of the University of Macau. Email: yiding@um.edu.mo

†Department of ISOM, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Research is supported in part by RGC grants GRF-16304521 and GRF 15302321 of the HKSAR. Email: xhzheng@ust.hk

1 Introduction

1.1 Motivation

Elliptical models are widely used to model financial returns (Zheng and Li (2011); Han and Liu (2014)), gene microarrays (Avella-Medina et al. (2018)) and brain images (Han and Liu (2018)) due to their flexibility in capturing heavy-tailedness and tail-dependences. An elliptical model is typically defined as follows:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sqrt{N}\xi_t \boldsymbol{\Sigma}_0^{1/2}\mathbf{u}_t, \quad t = 1, ..., T, \tag{1.1}$$

where $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)^\intercal$ is the expectation, $\xi_t \geq 0$ is a random scalar variable, $\boldsymbol{\Sigma}_0$ is a $N \times N$ positive-definite matrix, $\mathbf{u}_t$ is a length-N vector uniformly distributed on the unit sphere $\mathcal{S}^{N-1}$ and independent of ξ_t, and T stands for the sample size. The covariance matrix of $(\mathbf{y}_t)$ is $\boldsymbol{\Sigma} = E(\xi_t^2)\boldsymbol{\Sigma}_0$. Usually, $\boldsymbol{\Sigma}_0$ is referred to as the scatter matrix (Fan et al. (2018)). For identification purposes, $\boldsymbol{\Sigma}_0$ can be normalized to have a trace of N.

High-dimensional data such as financial returns usually exhibit a factor structure, which leads to spiked eigenvalues. Specifically, suppose that the number of factors is K, then the eigenvalues of $\boldsymbol{\Sigma}$, denoted by $\lambda_1 \geq ... \geq \lambda_N$, satisfy that

$$\lambda_i \asymp N \text{ for } 1 \leq i \leq K, \quad \text{ and } \lambda_i \asymp 1 \text{ for } K < i \leq N, \tag{1.2}$$

where for any two sequences (a_n) and (b_n), we write $a_n \asymp b_n$ if $b_n/c \leq a_n \leq cb_n$ for some constant $c \geq 1$.

It is worth mentioning that in the elliptical model (1.1), the scatter matrix $\boldsymbol{\Sigma}_0$ determines the cross-sectional dependence and plays a central role in various applications. For example, the eigenstructure of $\boldsymbol{\Sigma}_0$ governs the principal component analysis (PCA) of $(\mathbf{y}_t)$. In financial applications, the inverse of the scatter matrix, $\boldsymbol{\Sigma}_0^{-1}$, is used in constructing the global minimum variance portfolio (MVP). The scalar component, $E(\xi_t^2)$, on the other hand, is not relevant. Specifically, suppose that there are N assets, whose returns have a covariance matrix of $\boldsymbol{\Sigma}$. Then the MVP solves

$$\min \mathbf{w}^\intercal \boldsymbol{\Sigma} \mathbf{w}, \text{ subject to } \mathbf{w}^\intercal \mathbf{1} = 1, \tag{1.3}$$

where $\mathbf{w}$ is a length N vector of portfolio weights, and $\mathbf{1}$ is a length N vector of ones. The solution of the MVP is

$$\mathbf{w}^* = \frac{1}{\mathbf{1}^\intercal \boldsymbol{\Sigma}^{-1} \mathbf{1}} \boldsymbol{\Sigma}^{-1} \mathbf{1}.$$

It is straightforward to see that under elliptical models,

$$\mathbf{w}^* = \frac{1}{\mathbf{1}^\intercal \boldsymbol{\Sigma}_0^{-1} \mathbf{1}} \boldsymbol{\Sigma}_0^{-1} \mathbf{1}.$$

In other words, using $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}_0^{-1}$ are equivalent in constructing the MVP.

1.2 Existing Scatter Matrix Estimators in Elliptical Models When $N = O(T)$

Various methods have been proposed for estimating the scatter matrix in elliptical models. Popular estimators include the minimum covariance determinant (MCD) and minimum volume ellipsoid (MVE) estimators (Rousseeuw (1984, 1985)), Maronna's M-estimator (Maronna (1976)) and Tyler's M-estimator (TME, Tyler (1987)). However, these estimators cannot handle the case when $N > T$; see, for example, the discussions in Van Aelst and Rousseeuw (2009), Zhang (2016); Zhang et al. (2016)

and Hubert et al. (2018).

Sun et al. (2014) propose a regularized TME for the case when $N > T$. The regularized TME is obtained by first solving

$$\widehat{\mathbf{\Sigma}}(\alpha) = \frac{1}{1+\alpha}\frac{N}{T}\sum_{i=1}^{N}\frac{\mathbf{y}_t\mathbf{y}_t^\intercal}{\mathbf{y}_t^\intercal\widehat{\mathbf{\Sigma}}(\alpha)^{-1}\mathbf{y}_t} + \frac{\alpha}{1+\alpha}\mathbf{I}, \tag{1.4}$$

where α is a sufficiently large regularization parameter, and then normalizing $\widehat{\mathbf{\Sigma}}(\alpha)$ as follows:

$$\widehat{\mathbf{\Sigma}}(\alpha)_{Reg-TME} = \frac{N}{\mathrm{tr}\left(\widehat{\mathbf{\Sigma}}(\alpha) - \alpha/(1+\alpha)\mathbf{I}\right)}\left(\widehat{\mathbf{\Sigma}}(\alpha) - \frac{\alpha}{1+\alpha}\mathbf{I}\right). \tag{1.5}$$

When $\alpha = 0$ in (1.4), the regularized TME reduces to the original TME.

Goes et al. (2020) show that, when $\mathbf{\Sigma}_0$ has bounded eigenvalues, thresholding the (regularized) TME achieves the sub-Gaussian convergence rate for heavy-tailed elliptical distribution. However, their results are only for the case when $N = O(T)$. Furthermore, when the mean $\boldsymbol{\mu}$ in (1.1) is not zero, symmetrization needs to be applied, which reduces the sample size by half.

1.3 Existing Covariance Matrix Estimators under Heavy-tailedness When $N \gg T$

1.3.1 Estimators when There is No Factor Structure

Several robust estimators of the covariance matrix have been proposed in the literature. These include Huber-type estimators (Minsker (2018); Avella-Medina et al. (2018); Minsker and Wei (2020)), and truncation or shrinkage methods (Fan et al. (2021)). These estimators can handle the case when $N \gg T$ but typically need that data have finite 4th moment. Avella-Medina et al. (2018) study the high-dimensional covariance matrix estimation for cases where data have finite 4th or $2+\varepsilon$th moment with $\varepsilon < 2$. Under sparsity assumptions on the covariance matrix or its inverse matrix, their approach involves applying thresholding techniques (Cai and Liu (2011); Cai et al. (2016)) on a robust pilot matrix. The pilot matrix combines a Huber-type M-estimator (Huber (1964)) of variances and a rank-based robust correlation matrix estimator such as the marginal Kendall's tau (Liu et al. (2012); Xue and Zou (2012)). Under the high-dimensional setting where the dimension N and the sample size T both go to infinity, if the data have 4th moment, Avella-Medina et al. (2018) show that their estimator achieves the convergence rate of $\sqrt{(\log N)/T}$, which is the same as in the sub-Gaussian case. However, for the case when only $2+\varepsilon$th moment exists, the convergence rate reduces to $((\log N)/T)^{\varepsilon/(2+\varepsilon)}$; see Proposition 6 therein.

1.3.2 Estimators when There is Factor Structure

If we denote by $\phi_1, ..., \phi_N$ the eigenvectors corresponding to $\lambda_1 \geq ... \geq \lambda_N$, then we have

$$\mathbf{\Sigma} = \sum_{i=1}^{N} \lambda_i \phi_i \phi_i^\intercal = \sum_{i=1}^{K} \lambda_i \phi_i \phi_i^\intercal + \sum_{i=K+1}^{N} \lambda_i \phi_i \phi_i^\intercal = \mathbf{\Gamma}_K \mathbf{\Lambda}_K \mathbf{\Gamma}_K^\intercal + \mathbf{\Sigma}_u, \tag{1.6}$$

where $\mathbf{\Gamma}_K = (\phi_1, ..., \phi_K)$, $\mathbf{\Lambda}_K = \text{diag}(\lambda_1, ..., \lambda_K)$, and $\mathbf{\Sigma}_u = \sum_{i=K+1}^{N} \lambda_i \phi_i \phi_i^\intercal$. Under the factor model setting (1.2), further assuming that $\mathbf{\Sigma}_u$ is sparse, Fan et al. (2013) propose a Principal Orthogonal ComplEment Thresholding (POET) estimator that thresholds the sample analogue of $\mathbf{\Sigma}_u$. Specifically, suppose that $\mathbf{S}$ is the sample covariance matrix of $(\mathbf{y}_t)_{t=1,..,T}$. In analogous to (1.6), decompose $\mathbf{S}$ as $\mathbf{S} = \widehat{\mathbf{\Gamma}}_K \widehat{\mathbf{\Lambda}}_K \widehat{\mathbf{\Gamma}}_K^\intercal + \widehat{\mathbf{\Sigma}}_u$, where $\widehat{\mathbf{\Lambda}}_K$ and $\widehat{\mathbf{\Gamma}}_K$ consist of the leading K eigenvalues of the sample covariance matrix $\mathbf{S}$ and the corresponding eigenvectors, respectively. The

POET estimator is given by $\widehat{\mathbf{\Sigma}}^{\tau} = \widehat{\mathbf{\Gamma}}_K \widehat{\mathbf{\Lambda}}_K \widehat{\mathbf{\Gamma}}_K^{\intercal} + \widehat{\mathbf{\Sigma}}_u^{\tau}$, where $\widehat{\mathbf{\Sigma}}_u^{\tau}$ is obtained by applying the adaptive thresholding method (Cai and Liu (2011)) to $\widehat{\mathbf{\Sigma}}_u$. This study focuses on the case where data are sub-Gaussian or sub-Exponential.

Fan et al. (2018) extend the POET estimator to elliptical factor models with a 4th moment condition. They propose a generic POET procedure to estimate the high-dimensional covariance matrix. The approach relies on three components, $\widehat{\mathbf{\Sigma}}$, $\widehat{\mathbf{\Lambda}}_K$ and $\widehat{\mathbf{\Gamma}}_K$, which are robust pilot estimators of the covariance matrix $\mathbf{\Sigma}$, its leading eigenvalues $\mathbf{\Lambda}_K$ and eigenvectors $\mathbf{\Gamma}_K$, respectively. The estimators are constructed as follows. The pilot covariance matrix estimator, $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{D}}\widehat{\mathbf{R}}\widehat{\mathbf{D}}$, where $\widehat{\mathbf{D}}$ is a Huber estimator of the standard deviations, and $\widehat{\mathbf{R}}$ is a Marginal Kendall's estimator of the correlation matrix. The pilot estimator of the leading eigenvalues, $\widehat{\mathbf{\Lambda}}_K$, is the K largest eigenvalues of $\widehat{\mathbf{\Sigma}}$. The pilot estimator of the leading eigenvectors, $\widehat{\mathbf{\Gamma}}_K$, is taken to be the leading eigenvectors of the Spatial Kendall's tau estimator:

$$\widehat{\mathbf{\Sigma}}_2 = \frac{2}{T(T-1)} \sum_{t<t'} \frac{(\mathbf{y}_t - \mathbf{y}_{t'})(\mathbf{y}_t - \mathbf{y}_{t'})^{\intercal}}{\|\mathbf{y}_t - \mathbf{y}_{t'}\|_2^2}. \tag{1.7}$$

See Sections 4.2–4.4 in Fan et al. (2018) for details.

Based on the three pilot components, the generic POET estimator of the covariance matrix is given by

$$\widehat{\mathbf{\Sigma}}^{\tau} = \widehat{\mathbf{\Gamma}}_K \widehat{\mathbf{\Lambda}}_K \widehat{\mathbf{\Gamma}}_K^{\intercal} + \widehat{\mathbf{\Sigma}}_u^{\tau},$$

where $\widehat{\mathbf{\Sigma}}_u^{\tau}$ is obtained by applying the adaptive thresholding method to $\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Gamma}}_K \widehat{\mathbf{\Lambda}}_K \widehat{\mathbf{\Gamma}}_K^{\intercal}$.

The theoretical properties of the generic POET procedure crucially depend on these pilot components. Fan et al. (2018) show that the proposed pilot components have sub-Gaussian convergence rate under the 4th moment condition:

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} = O_p\left(\sqrt{\frac{\log N}{T}}\right), \tag{1.8}$$

$$\|\widehat{\mathbf{\Lambda}}_K \mathbf{\Lambda}_K^{-1} - \mathbf{I}\|_{\max} = O_p\left(\sqrt{\frac{\log N}{T}}\right), \quad \text{and} \tag{1.9}$$

$$\|\widehat{\mathbf{\Gamma}}_K - \mathbf{\Gamma}_K\|_{\max} = O_p\left(\sqrt{\frac{\log N}{TN}}\right), \tag{1.10}$$

where $\|A\|_{\max} = \max_{ij} |a_{ij}|$ for any matrix $A = (a_{ij})$. When only $2 + \varepsilon$th moment exists, the convergence rates in equations (1.8) and (1.9) reduce to $((\log N)/T)^{\varepsilon/(2+\varepsilon)}$.

1.4 Our Contributions

Infinite kurtosis and cross-sectional tail dependence are commonly observed in financial data. Stock returns contain large and dependent jumps (e.g., Li et al. (2017); Ding et al. (2023); Jacod et al. (2024)), which have slowly decaying tails that suggest infinite 4th moment (Bollerslev et al. (2013)). For data with only $2+\varepsilon$th moment, under the ultra-high dimensional setting when $N \gg T$, the aforementioned estimation methods cannot achieve the sub-Gaussian rate. An important question is:

Is there a scatter matrix estimator that can achieve the sub-Gaussian rate when only $2 + \varepsilon$th moment exists under the setting when $N \gg T$?

In this paper, we provide an affirmative answer to the question. Our approach critically relies on a consistent estimator of the random scalar process (ξ_t) in equation (1.1). The estimator allows us to remove the effect of (ξ_t) in $(\mathbf{y}_t)$ and achieve the desired sub-Gaussian rate. The intuition behind our proposed estimator $(\widehat{\xi}_t)$ is as follows.

Let us start with a motivating case when there is no factor and the mean is zero. Under this case, Zheng and Li (2011) develop a self-normalization approach to remove (ξ_t) in elliptical models. Specifically, they show that (ξ_t) can be consistently estimated by $(\|\mathbf{y}_t\|_2/\sqrt{N})$ as $N \to \infty$. It follows that $\mathbf{y}_t/(\|\mathbf{y}_t\|_2/\sqrt{N}) \approx \mathbf{\Sigma}_0^{1/2}\mathbf{u}_t$.

Under the elliptical factor model, the consistency of $(\|\mathbf{y}_t\|_2/\sqrt{N})$ in estimating (ξ_t) no longer holds. To overcome this challenge, we develop an idiosyncratic projected self-normalization (IPSN) approach. In short, we use the idiosyncratic component that are absent of the strong cross-sectional dependence to perform the self-normalization.

Finally, in the case when $\boldsymbol{\mu} = E(\mathbf{y}_t) \neq 0$, we need to demean the data. Due to the heavy-tailedness, the sample mean is not accurate enough. Instead, we use a Huber-type robust estimator, $\widehat{\boldsymbol{\mu}}$. The details of our estimators are given in Section 2.2.

We show that our IPSN approach is consistent in estimating (ξ_t), enabling us to remove the effect of (ξ_t) in $(\mathbf{y}_t)$ when estimating the scatter matrix. We apply POET to $((\mathbf{y}_t - \widehat{\boldsymbol{\mu}})/\widehat{\xi}_t)$ to obtain the scatter matrix estimator, denoted by POET-IPSN. We show that POET-IPSN achieves the sub-Gaussian convergence rate in estimating the scatter matrix under only $2+\varepsilon$th moment condition.

We also construct consistent estimator of $E(\xi_t^2)$ based on $(\|\mathbf{y}_t - \widehat{\boldsymbol{\mu}}\|_2^2)$, which we show achieves the same converge rate as the oracle estimator based on the unobserved (ξ_t) process. Combining our proposed POET-IPSN estimator of $\mathbf{\Sigma}_0$ with the estimator of $E(\xi_t^2)$ yields an estimator of $\mathbf{\Sigma}$, which we show achieves a faster convergence rate than the generic POET estimator in Fan et al. (2018).

The favorable properties of our proposed scatter matrix estimator are clearly demonstrated in the numerical studies. Simulation studies show that the POET-IPSN estimator performs robustly well under various heavy-tailedness settings and achieves a higher estimation accuracy than the robust generic POET estimator in Fan et al. (2018) and the POET estimator in Fan et al. (2013). Empirically, we apply the POET-

IPSN estimator to construct minimum variance portfolios using S&P 500 Index constituent stocks. We find that our portfolio has a significantly lower out-of-sample risk than various benchmark portfolios.

The rest of this paper is organized as follows. We present our theoretical results in Section 2. Simulations and empirical studies are presented in Sections 3 and 4, respectively. We conclude in Section 5. The proofs of the theorems and propositions are collected in Appendix A of the Supplementary Material.

We use the following notation throughout the paper. For any matrix $\mathbf{A} = (A_{ij})$, its spectral norm is defined as $\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2 \le 1} \sqrt{\mathbf{x}^\intercal \mathbf{A}^\intercal \mathbf{A}\mathbf{x}}$, where $\|\mathbf{x}\|_2 = \sqrt{\sum x_i^2}$ for any vector $\mathbf{x} = (x_i)$; the Frobenius norm is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$; and the relative Frobenius norm is defined as $\|\mathbf{A}\|_{\mathbf{\Sigma}} = \|\mathbf{\Sigma}^{-1/2}\mathbf{A}\mathbf{\Sigma}^{-1/2}\|_F/\sqrt{N}$, where $\mathbf{\Sigma}$ is a $N \times N$ positive-definite matrix. Finally, c, C, ..., etc., denote generic constants that may vary from place to place.

2 Main Results

2.1 Settings and Assumptions

Suppose that $(\mathbf{y}_t)$ follows the elliptical model (1.1). We make the following assumptions.

Assumption 1 *(ξ_t) are i.i.d. and independent of $(\mathbf{u}_t)$. Moreover, there exist constants $0 < c < C < \infty$ such that $E(\xi_t^{2+\varepsilon}) < C$ for some $0 < \varepsilon < 2$ and $P(|\xi_t| > c) = 1$.*

Assumption 2 *There exist constants $0 < c < C < \infty$ such that $c < \min_{1\le i\le N}(\mathbf{\Sigma}_0)_{ii} \le \max_{1\le i\le N}(\mathbf{\Sigma}_0)_{ii} < C$ for all N.*

Write the eigendecomposition of $\mathbf{\Sigma}_0$ as $\mathbf{\Sigma}_0 = \mathbf{\Gamma}\mathbf{\Lambda}_0\mathbf{\Gamma}^\intercal$, where $\mathbf{\Lambda}_0 = \mathrm{diag}(\lambda_{0;1}, \lambda_{0;2}, ..., \lambda_{0;N})$,

$\lambda_{0;1} \geq \lambda_{0;2} ... \geq \lambda_{0;N}$ are the eigenvalues of $\mathbf{\Sigma}_0$, and $\mathbf{\Gamma} = (\phi_1, ..., \phi_N)$ is the corresponding matrix of eigenvectors. We assume that factors exist so that the following eigen-gap condition holds.

Assumption 3 *There exist a $K \in \mathbb{N}$ and $0 < \delta < 1$ such that $\delta N \leq \lambda_{0;i} \leq N/\delta$ for $i = 1, ..., K$, and $0 < \delta \leq \lambda_{0;i} \leq 1/\delta$ for $i = K+1, ..., N$.*

Assumptions 1–3 describe the setting of elliptical factor models (Fan et al. (2018)). The latent component ξ_t drives the heteoskedasticity in both factor and idiosyncratic components. This setting is in line with the widely documented co-movement feature in stock volatilities (Susmel and Engle 1994) and idiosyncratic volatilities (Herskovic et al. 2016). In particular, Ding et al. (2024) show that a multiplicative volatility factor model, which includes a single multiplicative volatility factor, describes the co-movement among the stock volatilities well. Assumption 3 describes statistical factor models with spiked eigenvalues; see, e.g., Fan et al. (2013); Ding et al. (2021).

Widely used approaches for estimating the number of factors such as those developed in Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) require the existence of 4th moment. Under the elliptical factor model with $2+\varepsilon$th moment, the number of factors can be consistently estimated using the robust estimator in Yu et al. (2019). It follows that estimating the number of factors does not affect the convergence rates in estimating $\mathbf{\Sigma}_0$ or $\mathbf{\Sigma}$. For this reason, in the following analysis, we assume that the number of factors is known.

2.2 Idiosyncratic-Projected Self-Normalized Estimator of Scatter Matrix and Covariance Matrix

Observe that under model (1.1), we can write $\mathbf{y}_t$ as follows:

$$\mathbf{y}_t = \boldsymbol{\mu} + \frac{\sqrt{N}\xi_t}{\|\mathbf{z}_t\|_2}\boldsymbol{\Sigma}_0^{1/2}\mathbf{z}_t, \tag{2.1}$$

where $\mathbf{z}_t \sim N(0, \mathbf{I}_N)$, and ξ_t and $\mathbf{z}_t$ are independent. The heavy-tailedness of $(\mathbf{y}_t)$ is solely determined by (ξ_t). Motivated by this observation, we separately estimate $E(\xi_t^2)$ and $\boldsymbol{\Sigma}_0$. To do so, we develop an idiosyncratic-projected self-normalization (IPSN) approach. We explain the construction of our proposed estimator below.

Step I. Estimate $\boldsymbol{\mu}$

Given that $(\mathbf{y}_t)$ is heavy-tailed, we estimate $\boldsymbol{\mu}$ using a Huber estimator, a widely used estimator for heavy-tailed distributions (Huber (1964)). Specifically, the estimator $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_i)_{1\le i\le N}$ solves

$$\sum_{t=1}^{T} \Psi_H(y_{i,t} - \widehat{\mu}_i) = 0, \quad \text{for} \quad i = 1, ..., N, \tag{2.2}$$

where $\Psi_H(x) = \min\Big(H, \max(-H, x)\Big)$ is the Huber function. H is a tuning parameter that goes to infinity with T and satisfies $H = O(\sqrt{T/(\log N)})$. The tuning parameter H can be chosen adaptively; see Sun et al. (2020).

Step II. Construct idiosyncratic-projected self-normalized variables $(\widehat{\mathbf{X}}_t)$

In Step II, we normalize observations to remove the effect of (ξ_t). Under the elliptical factor model, robust consistent estimation of the leading eigenvectors, and hence its orthogonal space can be achieved using spatial Kendall's tau matrix (see, e.g., Fan et al. (2018)). Specifically, we first compute the spatial Kendall's tau matrix (1.7), get its leading eigenvectors, denoted by $\widehat{\boldsymbol{\Gamma}}_{KED}$, and use them as a robust estimator of

$\mathbf{\Gamma}_K = (\phi_1, ..., \phi_K)$. Then, we define a $(N-K) \times N$ projection matrix $\widehat{\mathbf{P}}_I$ as follows:

$$\widehat{\mathbf{P}}_I^\intercal = \text{Null}(\widehat{\mathbf{\Gamma}_{KED}}),$$

that is, $\widehat{\mathbf{P}}_I^\intercal$ is the $N \times (N-K)$ matrix that spans the null space of $\widehat{\mathbf{\Gamma}_{KED}}$ and satisfies $\widehat{\mathbf{P}}_I \widehat{\mathbf{\Gamma}_{KED}} = \mathbf{0}$ and $\widehat{\mathbf{P}}_I \widehat{\mathbf{P}}_I^\intercal = \mathbf{I}_{N-K}$. Next, we define the idiosyncratic projected self-normalized variables $(\widehat{\mathbf{X}}_t)$:

$$\widehat{\mathbf{X}}_t = \sqrt{N} \frac{\mathbf{y}_t - \widehat{\boldsymbol{\mu}}}{\|\widehat{\mathbf{P}}_I(\mathbf{y}_t - \widehat{\boldsymbol{\mu}})\|_2}, \qquad t = 1, ..., T. \tag{2.3}$$

The normalization terms $(\|\widehat{\mathbf{P}}_I(\mathbf{y}_t - \widehat{\boldsymbol{\mu}})\|_2)$ are approximately proportional to (ξ_t), hence $(\widehat{\mathbf{X}}_t)$ are approximately proportional to $(\mathbf{\Sigma}_0^{1/2}\mathbf{u}_t)$.

The goal of the idiosyncratic projection is to remove the strong cross-sectional dependence in the data, thereby enabling consistent estimation of the scalar component (ξ_t). This step is crucial. As can be seen in Section 3.3 below, when there is strong cross-sectional dependence, the self-normalization approach in Zheng and Li (2011) no longer works in estimating the time-varying scalar component (ξ_t).

Step III. Construct pilot estimators of $\mathbf{\Sigma}_0$

Define

$$\widehat{\mathbf{\Sigma}}_0 = \frac{\widehat{\eta}}{T} \sum_{t=1}^{T} \widehat{\mathbf{X}}_t \widehat{\mathbf{X}}_t^\intercal, \tag{2.4}$$

where

$$\widehat{\eta} = \frac{N}{\text{tr}(\sum_{t=1}^{T} \widehat{\mathbf{X}}_t \widehat{\mathbf{X}}_t^\intercal / T)}. \tag{2.5}$$

Essentially, we normalize $\mathbf{y}_t - \widehat{\boldsymbol{\mu}}$ with

$$\widehat{\xi}_t = \frac{\|\widehat{\mathbf{P}}_I(\mathbf{y}_t - \widehat{\boldsymbol{\mu}})\|_2}{\sqrt{\widehat{\eta}}}. \tag{2.6}$$

By the definition of $\widehat{\eta}$ in equation (2.5), we ensure that $\mathrm{tr}(\widehat{\mathbf{\Sigma}}_0) = N$.

Write the eigendecomposition of $\widehat{\mathbf{\Sigma}}_0$ as $\widehat{\mathbf{\Sigma}}_0 = \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Lambda}}_0\widehat{\mathbf{\Gamma}}^\intercal$, where $\widehat{\mathbf{\Lambda}}_0 = \mathrm{diag}(\widehat{\lambda}_{0;1}, ..., \widehat{\lambda}_{0;N})$ is the matrix of eigenvalues with $\widehat{\lambda}_{0;1} \geq \widehat{\lambda}_{0;2} \geq ... \geq \widehat{\lambda}_{0;N}$, and $\widehat{\mathbf{\Gamma}} = (\widehat{\phi}_1, ..., \widehat{\phi}_N)$ is the matrix of eigenvectors. We then use the leading eigenvectors and leading eigenvalues of $\widehat{\mathbf{\Sigma}}_0$ as the pilot estimators:

$$\widehat{\mathbf{\Gamma}}_K = (\widehat{\phi}_1, ..., \widehat{\phi}_K), \quad \text{and} \quad \widehat{\mathbf{\Lambda}}_{0K} = \mathrm{diag}(\widehat{\lambda}_{0;1}, ..., \widehat{\lambda}_{0;K}).$$

The estimators $\widehat{\mathbf{\Sigma}}_0$, $\widehat{\mathbf{\Gamma}}_K$ and $\widehat{\mathbf{\Lambda}}_{0K}$ will be used in the POET procedure for estimating $\mathbf{\Sigma}_0$.

In the generic POET procedure proposed in Fan et al. (2018), the leading eigenvectors are estimated using the spatial Kendall's tau, and pilot components of the covariance matrix and its leading eigenvalues are from a Huber estimator of the covariance matrix. Due to that these estimators are from different sources, the resulting pilot idiosyncratic covariance matrix estimator is not guaranteed to be positive semi-definite. Our approach, on the other hand, does not have such an issue because our pilot estimators of the leading eigenvalues and leading eigenvectors are from the eigendecomposition of a same covariance matrix estimator.

Step IV. POET estimators of $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}$

Finally, we estimate the scatter matrix $\mathbf{\Sigma}_0$ and the covariance matrix $\mathbf{\Sigma}$. To estimate $\mathbf{\Sigma}_0$, we apply the POET procedure on our proposed idiosyncratic-projected self-normalized pilot estimators, $\widehat{\mathbf{\Sigma}}_0$, $\widehat{\mathbf{\Lambda}}_{0K}$ and $\widehat{\mathbf{\Gamma}}_K$ from Step III. Specifically, the estimator of $\mathbf{\Sigma}_0$ is

$$\widehat{\mathbf{\Sigma}}_0^\tau = \widehat{\mathbf{\Gamma}}_K\widehat{\mathbf{\Lambda}}_{0K}\widehat{\mathbf{\Gamma}}_K^\intercal + \widehat{\mathbf{\Sigma}}_{0u}^\tau, \tag{2.7}$$

where $\widehat{\mathbf{\Sigma}}_{0u}^\tau$ is obtained by applying the adaptive thresholding method (Cai and Liu (2011)) to $\widehat{\mathbf{\Sigma}}_0 - \widehat{\mathbf{\Gamma}}_K\widehat{\mathbf{\Lambda}}_{0K}\widehat{\mathbf{\Gamma}}_K^\intercal$.

On the other hand, if one is also interested in estimating the covariance matrix, then one further estimates $E(\xi_t^2)$ by the following robust estimator $\widehat{E(\xi_t^2)}$ that solves

$$\sum_{t=1}^{T} \Psi_H \left(\frac{\|\mathbf{y}_t - \widehat{\boldsymbol{\mu}}\|_2^2}{N} - \widehat{E(\xi_t^2)} \right) = 0, \tag{2.8}$$

where $H \asymp T^{\min(1/(1+\varepsilon/2),1/2)}$. Again, the tuning parameter H can be chosen adaptively. The estimator of $\boldsymbol{\Sigma}$ is then simply

$$\widehat{\boldsymbol{\Sigma}}^{\tau} = \widehat{E(\xi_t^2)}\,\widehat{\boldsymbol{\Sigma}}_0^{\tau}. \tag{2.9}$$

The idiosyncratic covariance matrix estimator is $\widehat{\boldsymbol{\Sigma}}_u^{\tau} = \widehat{E(\xi_t^2)}\,\widehat{\boldsymbol{\Sigma}}_{0u}^{\tau}$. It can be seen that $\widehat{\boldsymbol{\Sigma}}^{\tau}$ is equivalent to the generic POET estimator using the following pilot estimators: $\widehat{\boldsymbol{\Sigma}} =: \widehat{E(\xi_t^2)}\,\widehat{\boldsymbol{\Sigma}}_0$, $\widehat{\boldsymbol{\Lambda}}_K =: \widehat{E(\xi_t^2)}\,\widehat{\boldsymbol{\Lambda}}_{0K}$, and $\widehat{\boldsymbol{\Gamma}}_K$.

We summarize our IPSN approach in the following algorithm.

Algorithm:	Idiosyncratic-Projected Self-Normalization (IPSN) for Scatter Matrix and Covariance Matrix Estimation
Input:	$(\mathbf{y}_t)_{t\le T}$, K.
Output:	$\widehat{\boldsymbol{\Sigma}}_0^{\tau}$, $\widehat{\boldsymbol{\Sigma}}^{\tau}$.
Step I.	Compute $\widehat{\boldsymbol{\mu}}$ via (2.2).
Step II.	Compute $(\widehat{\mathbf{X}}_t)_{t\le T}$ via (2.3).
Step III.	Compute $\widehat{\eta}$ and $(\widehat{\xi}_t)_{t\le T}$ via (2.5) and (2.6), respectively. Compute $\widehat{\boldsymbol{\Sigma}}_0$ via (2.4), its first K eigenvalues $\widehat{\boldsymbol{\Lambda}}_{0K}$, and the corresponding eigenvectors $\widehat{\boldsymbol{\Gamma}}_K$.
Step IV.	Compute $\widehat{E(\xi_t^2)}$ via (2.8).
Step V.	Compute $\widehat{\boldsymbol{\Sigma}}_0^{\tau}$ via (2.7), and $\widehat{\boldsymbol{\Sigma}}^{\tau}$ via (2.9).

2.3 Theoretical Properties

In this subsection, we present the asymptotic properties of our proposed estimators.

2.3.1 Convergence Rates of Idiosyncratic-Projected Self-Normalized Pilot Estimators

We start with the properties of the idiosyncratic-projected self-normalization. The key to the success of our approach lies in approximating (ξ_t) with $(\widehat{\xi}_t)$ defined in equation (2.6). Intuitively, if $(\widehat{\xi}_t)$ is close to (ξ_t), then $((\mathbf{y}_t - \widehat{\boldsymbol{\mu}})/\widehat{\xi}_t)$ is close to $\sqrt{N}\boldsymbol{\Sigma}_0^{1/2}\mathbf{u}_t$, which is absent of the heavy-tailed scalar process (ξ_t). The next proposition gives the properties of $(\widehat{\xi}_t)$.

Proposition 1 *Under Assumptions 1–3, if in addition,* $N, T \to \infty$ *and satisfy* $(\log(N))^{2+\gamma} = o(T)$ *for some* $\gamma > 0$ *and* $\log(T) = O(\log(N))$*, then there exists a constant* $c_1 > 0$ *such that*

$$P\left(\min_{1\le t\le T} \widehat{\xi}_t > c_1\right) \to 1, \tag{2.10}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{\xi_t^2}{\widehat{\xi}_t^2} - 1\Big)^2 = O_p\Big(\frac{\log N}{N} + \frac{\log N}{T}\Big). \tag{2.11}$$

Equation (2.10) guarantees that, with probability approaching one, $(\widehat{\xi}_t)$ is bounded away from zero, hence the normalization based on $(\widehat{\xi}_t)$ is well-behaved. Furthermore, equation (2.11) guarantees the consistency of $(\widehat{\xi}_t)$ in estimating (ξ_t). We need both $N, T \to \infty$ to estimate (ξ_t) consistently, and N can be much larger than T.

The next theorem gives the convergence rate of the IPSN pilot estimators, $\widehat{\boldsymbol{\Sigma}}_0$, $\widehat{\boldsymbol{\Lambda}}_{0K}$, and $\widehat{\boldsymbol{\Gamma}}_K$.

Theorem 1 *Under Assumptions 1–3, if in addition, $N, T \to \infty$ and satisfy $(\log(N))^{2+\gamma} = o(T)$ for some $\gamma > 0$ and $\log(T) = O(\log(N))$, then*

$$\begin{aligned}
&\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\|_{\max} = O_p\left(\sqrt{\frac{\log N}{N}} + \sqrt{\frac{\log N}{T}}\right), \\
&\|\widehat{\mathbf{\Lambda}}_{0K}\mathbf{\Lambda}_{0K}^{-1} - \mathbf{I}\|_{\max} = O_p\left(\sqrt{\frac{\log N}{N}} + \sqrt{\frac{\log N}{T}}\right), \quad \text{and} \\
&\|\widehat{\mathbf{\Gamma}}_K - \mathbf{\Gamma}_K\|_{\max} = O_p\left(\sqrt{\frac{\log N}{N^2}} + \sqrt{\frac{\log N}{TN}}\right).
\end{aligned} \tag{2.12}$$

Theorem 1 states that the pilot estimators of the scatter matrix achieve the sub-Gaussian convergence rates under $2+\varepsilon$th moment assumption. In contrast, when only $2+\varepsilon$th moment exists, the pilot estimator $\widehat{\mathbf{\Sigma}}$ in Fan et al. (2018) has a convergence rate of only $O_p(((\log N)/T)^{\varepsilon/(2+\varepsilon)})$.

2.3.2 Convergence Rates of POET Estimators based on IPSN

Under the factor model specified in Assumption 3, we can write $\mathbf{\Sigma}_0 = \mathbf{B}_0\mathbf{B}_0^\intercal + \mathbf{\Sigma}_{0u}$, where $\mathbf{B}_0$ is a $N \times K$ matrix, and $\mathbf{\Sigma}_{0u}$ is the idiosyncratic component of the scatter matrix $\mathbf{\Sigma}_0$.

We consider factor models with conditional sparsity. Specifically, we assume that $\mathbf{\Sigma}_{0u}$ belongs to the following class of sparse matrices: for some $q \in [0, 1]$, $c > 0$, and $s_0(N) < \infty$,

$$\begin{aligned}
&\mathcal{U}_q\big(s_0(N)\big) \\
=&\left\{\mathbf{\Sigma} : \mathbf{\Sigma} = (\sigma_{ij}) \text{ is positive semi-definite, } \max_i \sigma_{ii} \le c, \text{and } \max_i \sum_{j=1}^{N} \sigma_{ij}^q \le s_0(N)\right\}.
\end{aligned} \tag{2.13}$$

Assumption 4 $\mathbf{\Sigma}_{0u} \in \mathcal{U}_q\big(s_0(N)\big)$ *for some* $q \in [0, 1]$.

The sparsity condition in Assumption 4 is the same as in Fan et al. (2018). It is typically assumed for estimating a high-dimensional covariance matrix under factor models, with the intuition that after removing the common factor component, it is reasonable to assume that the remaining components would be only weakly dependent.

Denote $\omega_T = \sqrt{(\log N)/N} + \sqrt{(\log N)/T}$. The next theorem gives the convergence rate of our proposed POET estimator of the scatter matrix.

Theorem 2 *Under the assumptions of Theorem 1 and Assumption 4, if in addition,* $s_0(N)\omega_T^{1-q} = o(1)$, *then*

$$\begin{aligned} &\|(\widehat{\mathbf{\Sigma}}_0^{\tau})^{-1} - (\mathbf{\Sigma}_0)^{-1}\|_2 = O_p\Big(s_0(N)\omega_T^{1-q}\Big), \\ &\|\widehat{\mathbf{\Sigma}}_0^{\tau} - \mathbf{\Sigma}_0\|_{\mathbf{\Sigma}_0} = O_p\big(\sqrt{N}\omega_T^2 + s_0(N)\omega_T^{1-q}\big), \quad \textit{and} \\ &\|\widehat{\mathbf{\Sigma}}_{0u}^{\tau} - \mathbf{\Sigma}_{0u}\|_2 = O_p\Big(s_0(N)\omega_T^{1-q}\Big) = \|(\widehat{\mathbf{\Sigma}}_{0u}^{\tau})^{-1} - \mathbf{\Sigma}_{0u}^{-1}\|_2. \end{aligned} \tag{2.14}$$

Theorem 2 asserts that in estimating the scatter matrix, its idiosyncratic components and their inverse matrices, under only $2+\varepsilon$th moment condition, our POET-IPSN estimator achieves the same convergence rate as the sub-Gaussian case (Theorems 3.1 and 3.2 in Fan et al. (2013)).

Finally, about the estimation of the covariance matrix, we need the estimator $\widehat{E(\xi_t^2)}$ defined in equation (2.8). The next proposition gives the convergence rate of $\widehat{E(\xi_t^2)}$.

Proposition 2 *Under Assumptions 1–3, if in addition,* $T \to \infty$ *and* $\log(N) = O(T)$, *then*

$$|\widehat{E(\xi_t^2)} - E(\xi_t^2)| = O_p\left(\frac{1}{T^{\varepsilon/(2+\varepsilon)}} + \sqrt{\frac{\log N}{T}}\right).$$

The error term $1/T^{\varepsilon/(2+\varepsilon)}$ is from robust Huber estimation based on the infeasible series (ξ_t), and the error term $\sqrt{(\log N)/T}$ comes from the estimation error of $\widehat{\boldsymbol{\mu}}$. When $\log N = o(T^{(2-\varepsilon)/(2+\varepsilon)})$, the convergence rate of $\widehat{E(\xi_t^2)}$ is the same as the robust Huber estimator applied to the unobserved series (ξ_t).

The next proposition gives the convergence rates of $\widehat{\boldsymbol{\Sigma}}^{\tau}$ and $\widehat{\boldsymbol{\Sigma}}_u^{\tau}$ in estimating the covariance matrix and the idiosyncratic covariance matrix, respectively.

Proposition 3 *Under the assumptions of Theorem 1 and Assumption 4, if in addition,* $s_0(N)\omega_T^{1-q} = o(1)$, *then*

$$\begin{aligned}
\|(\widehat{\boldsymbol{\Sigma}}^{\tau})^{-1} - \boldsymbol{\Sigma}^{-1}\|_2 &= O_p\left(s_0(N)\omega_T^{1-q} + \frac{1}{T^{\varepsilon/(2+\varepsilon)}} \right), \\
\|\widehat{\boldsymbol{\Sigma}}^{\tau} - \boldsymbol{\Sigma}\|_{\boldsymbol{\Sigma}} &= O_p\left(\sqrt{N}\omega_T^2 + s_0(N)\omega_T^{1-q} + \frac{1}{T^{\varepsilon/(2+\varepsilon)}} \right), \quad \textit{and} \qquad (2.15) \\
\|\widehat{\boldsymbol{\Sigma}}_u^{\tau} - \boldsymbol{\Sigma}_u\|_2 &= O_p\left(s_0(N)\omega_T^{1-q} + \frac{1}{T^{\varepsilon/(2+\varepsilon)}} \right) = \|(\widehat{\boldsymbol{\Sigma}}_u^{\tau})^{-1} - \boldsymbol{\Sigma}_u^{-1}\|_2.
\end{aligned}$$

One can show that when only $2+\varepsilon$th moment exists, the generic POET estimator proposed in Fan et al. (2018) has a lower convergence rate. Specifically, it has a convergence rate with the term ω_T^2 replaced by $\widetilde{\omega}_T^2$ and the term ω_T^{1-q} replaced by $\widetilde{\omega}_T^{2\varepsilon(1-q)/(2+\varepsilon)}$, where $\widetilde{\omega}_T = \sqrt{\log N/T} + 1/\sqrt{N}$.

3 Simulation Studies

3.1 Simulation Setting

To specify the scatter matrix, $\boldsymbol{\Sigma}_0$, we simulate $\mathbf{B} = (b_{ik})_{1\le i\le N, 1\le k\le K}$, where $b_{ik} \underset{\text{i.i.d.}}{\sim} N(0, s_i)$, $K = 3$, $s_1 = 1$, $s_2 = 0.75^2$, and $s_3 = 0.5^2$. We then set $\boldsymbol{\Sigma}_0 = N(\mathbf{B}\mathbf{B}^{\intercal} + \mathbf{I}_N)/\operatorname{tr}((\mathbf{B}\mathbf{B}^{\intercal} + \mathbf{I}_N))$.

About the expectation, $\boldsymbol{\mu}$, we simulate $\boldsymbol{\mu} = (\mu_i) \underset{\text{i.i.d.}}{\sim} N(0,1)$. Note that the estimators of the scatter matrices and covariance matrices are location invariant, hence the results will not vary for any arbitrary $\boldsymbol{\mu}$ or by setting $\boldsymbol{\mu} = \mathbf{0}$.

For the scalar process (ξ_t), we generate it from a Pareto distribution with $P(\xi_t > x) = (x_m/x)^\alpha$ for $x > x_m$, where x_m is chosen such that $E(\xi_t^2) = 1$. We set $\alpha = 2 + \varepsilon$ with $\varepsilon = 2$ or 0.2. When $\varepsilon = 2$, (ξ_t) has finite moments with any order below 4, and when $\varepsilon = 0.2$, (ξ_t) does not have a finite third moment.

We then generate $(\mathbf{y}_t)$ from $\mathbf{y}_t = \xi_t \sqrt{N} \boldsymbol{\Sigma}_0^{1/2} \mathbf{z}_t / \|\mathbf{z}_t\|_2$, $t = 1, ..., T$, where $\mathbf{z}_t \underset{\text{i.i.d.}}{\sim} N(0, \mathbf{I})$. The dimension and the sample size N and T are set to be $N = 500$ and $T = 250$, respectively.

3.2 Scatter Matrix and Covariance Matrix Estimators and Evaluation Metrics

We evaluate the performance of the following scatter matrix and covariance matrix estimators: the POET estimators based on our idiosyncratic-projected self-normalization approach, denoted by POET-IPSN; the robust generic POET estimators in Fan et al. (2018), denoted by GPOET-FLW; and the original POET estimator based on the sample covariance matrix (Fan et al. (2013)), denoted by POET-S. In addition, we include the regularized TME estimator of the scatter matrix in Goes et al. (2020) as the pilot estimator and apply the POET procedure to the regularized TME. Denote this new estimator by POET-RegTME. When implementing the Regularized TME, we first conduct symmetrization to remove the mean, $\tilde{\mathbf{y}}_t = \mathbf{y}_{2t+1} - \mathbf{y}_{2t+2}$ for $t = 1, ..., [T/2]$. We then perform the regularized TME on the symmetrized data. Based on Theorems 1-2 of Goes et al. (2020), we set the regularization parameter α in (1.4) and (1.5) to be $\max(0.1, 1.1(\gamma - 1 + \hat{s}_{\max}(1 + \sqrt{\gamma})^2))$ with $\gamma = N/(2T)$, and

$\hat{s}_{\max}$ is the spectral norm of the sample covariance matrix of $\tilde{\mathbf{y}}_t$.

We use the same evaluation metrics as in Fan et al. (2018). Specifically, errors in estimating $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}$ are measured by the difference in the relative Frobenius norm:

$$\|\widehat{\boldsymbol{\Sigma}}_0^\tau - \boldsymbol{\Sigma}_0\|_{\boldsymbol{\Sigma}_0} = \frac{1}{\sqrt{N}}\|\boldsymbol{\Sigma}_0^{-1/2}(\widehat{\boldsymbol{\Sigma}}_0^\tau - \boldsymbol{\Sigma}_0)\boldsymbol{\Sigma}_0^{-1/2}\|_F,$$
$$\|\widehat{\boldsymbol{\Sigma}}^\tau - \boldsymbol{\Sigma}\|_{\boldsymbol{\Sigma}} = \frac{1}{\sqrt{N}}\|\boldsymbol{\Sigma}^{-1/2}(\widehat{\boldsymbol{\Sigma}}^\tau - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1/2}\|_F.$$

Errors in estimating the matrices $\boldsymbol{\Sigma}_{0u}$, $\boldsymbol{\Sigma}_u$, $\boldsymbol{\Sigma}_0^{-1}$, and $\boldsymbol{\Sigma}^{-1}$ are measured by the difference in spectral norm.

3.3 Simulation Results

First, to illustrate how well our approach works in estimating the scalar process (ξ_t), we compute the estimator $(\widehat{\xi_t})$ from our IPSN approach in equation (2.6) and obtain the ratio $(\widehat{\xi_t}/\xi_t)$ from one realization. We compare the results with the estimator in Zheng and Li (2011) (denoted by ZL11). The results are shown in Figure 1.

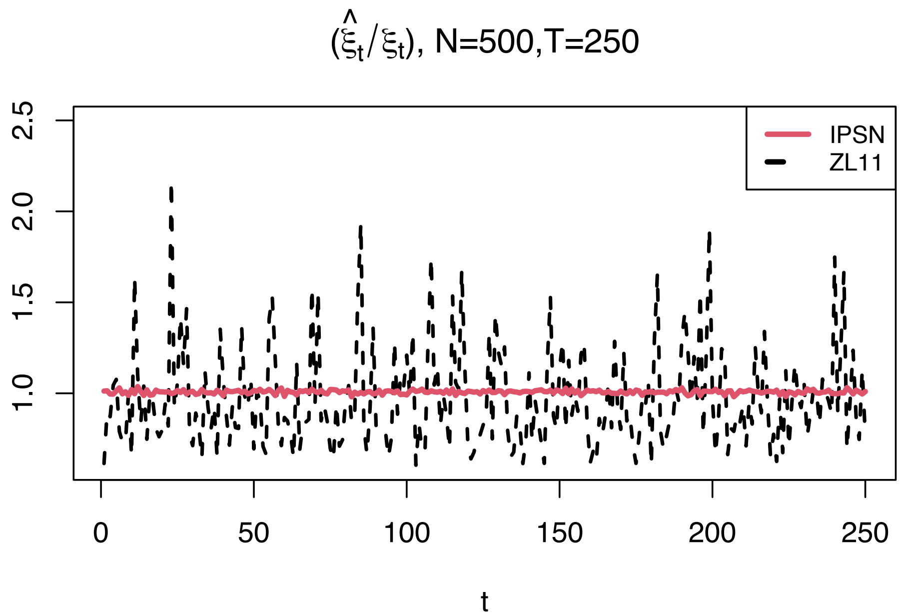

Figure 1: The ratios between $(\widehat{\xi}_t)$ and (ξ_t) from one simulation. We compare our proposed IPSN estimator with the estimator (ZL11) in Zheng and Li (2011). The process (ξ_t) is generated from Pareto distribution $P(\xi_t > x) = (x_m/x)^\alpha$ with the shape parameter $\alpha = 2.2$.

We see from Figure 1 that the ratios between our estimated and the true (ξ_t)'s are remarkably close to one for all t. In contrast, due to the factor structure in the data, the estimator in Zheng and Li (2011) is not consistent, reflected in that the ratios between the estimated (ξ_t) process and the true one can be far away from one.

Next, we summarize the performance of the pilot estimators. We include our IPSN pilot estimators and the following benchmark methods for comparison: the robust estimators of Fan et al. (2018) (FLW), the sample covariance matrix with its leading eigenvectors and eigenvalues (SAMPLE), as well as the regularized TME estimator (Reg-TME). For the benchmark methods, we get the pilot estimators for $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Lambda}_{0K}$ by normalizing $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Lambda}}_K$ with $N/\operatorname{tr}(\widehat{\boldsymbol{\Sigma}})$. For the pilot estimators of $\widehat{\boldsymbol{\Sigma}}_0$ and $\widehat{\boldsymbol{\Sigma}}$, we

Table 1: Performance of the pilot estimators. The heavy-tailedness index is set to be $\varepsilon = 2$ or 0.2. The dimension and the sample size are $N = 500$ and $T = 250$, respectively. Reported values are mean and standard deviations (in parentheses) from 100 replications for the error in maximum norm in estimating $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}$, and ($\sqrt{N}\times$) $\mathbf{\Gamma}_K$, and the maximum norm error in ratios in estimating $\mathbf{\Lambda}_{0K}$ and $\mathbf{\Lambda}_K$. We include our IPSN and the following benchmark methods: the robust method proposed in Fan et al. (2018) (FLW), the sample covariance matrix (SAMPLE), and the Regularized TME estimator (Reg-TME).

	$\mathbf{\Sigma}_0$	$\mathbf{\Sigma}$	$\mathbf{\Lambda}_{0K}$	$\mathbf{\Lambda}_K$	$\mathbf{\Gamma}_K$
$\varepsilon = 2$					
IPSN	0.482	0.676	0.100	0.165	0.592
	(0.108)	(0.273)	(0.048)	(0.099)	(0.177)
FLW	0.706	0.921	0.111	0.181	0.620
	(0.392)	(0.760)	(0.053)	(0.122)	(0.178)
SAMPLE	0.728	0.923	0.135	0.204	0.872
	(0.394)	(0.768)	(0.076)	(0.147)	(0.488)
Reg-TME	0.891	–	0.232	–	2.755
	(0.149)	–	(0.051)	–	(2.406)
$\varepsilon = 0.2$					
IPSN	0.485	2.105	0.100	0.623	0.598
	(0.109)	(2.706)	(0.048)	(0.886)	(0.182)
FLW	1.820	3.371	0.141	0.656	0.641
	(1.289)	(9.982)	(0.062)	(0.891)	(0.185)
SAMPLE	1.913	3.406	0.326	0.882	2.135
	(1.313)	(9.987)	(0.248)	(2.350)	(1.147)
Reg-TME	0.884	–	0.229	–	3.007
	(1.150)	–	(0.052)	–	(2.436)

compute the error in the maximum norm. For the estimators of $\mathbf{\Gamma}_K$, we evaluate the maximum error scaled by $\sqrt{N}$, $(\|\widehat{\mathbf{\Gamma}}_K - \mathbf{\Gamma}_K\|_{\max})\sqrt{N}$. For the estimators of $\mathbf{\Lambda}_{0,K}$ and $\mathbf{\Lambda}_K$, we report the errors in ratios: $\|\widehat{\mathbf{\Lambda}}_{0,K}\mathbf{\Lambda}_{0K}^{-1} - \mathbf{I}\|_{\max}$, and $\|\widehat{\mathbf{\Lambda}}_K\mathbf{\Lambda}_K^{-1} - \mathbf{I}\|_{\max}$. In Table 1, we report the performance of the pilot estimators from 100 replications. We see from Table 1 that our IPSN estimators outperform the other estimators for all five targets. The advantage of IPSN over FLW and SAMPLE is more salient in the estimation of the scatter matrices and with increasing heavy-tailedness in the data. In particular, for the pilot estimators of the scatter matrix and its leading eigenvectors and eigenvalues, when ε changes from 2 to 0.2, that is, when data get more heavy-tailed, the estimation errors for IPSN remain almost the same. In contrast, the errors in the FLW and SAMPLE estimators increase substantially. The Reg-TME estimator is also robust to heavy-tailedness, but it underperforms our IPSN estimator.

For pilot estimators of the covariance matrix, the error of our IPSN estimator grows with the heavy-tailedness of the model, but it is still significantly lower than the other methods.

Finally, we summarize the performance in estimating $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}$ and related components in Table 2. We see from Table 2 that our POET-IPSN estimators deliver the lowest estimation error in for all targets. Compared to GPOET-FLW or POET-S, the advantage of POET-IPSN is especially evident for estimating $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}_{0u}$, their inverse matrices, and for the more heavy-tailed setting when $\varepsilon = 0.2$. When estimating the scatter matrix, its idiosyncratic component and their inverses, the performance of POET-IPSN is robust for different heavy-tailedness conditions. When ε decreases from 2 to 0.2, the data become more heavy-tailed, but the estimation errors of POET-IPSN remain largely the same. In contrast, the errors of the GPOET-FLW and POET-S estimators are substantially larger in the more heavy-tailed setting. The POET-RegTME also performs robustly under different heavy-tailedness conditions, but it underperforms POET-IPSN in all measures. In the estimation of $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^{-1}$, POET-IPSN also has lower estimation errors than the benchmark methods.

In summary, the simulation results validate the theoretical properties of our approach and demonstrate its substantial advantage in estimating the scatter matrix and its inverse matrix. Such advantages are particularly valuable in applications where only the scatter matrix matters, as we will show in the empirical studies below.

Table 2: Performance in estimating $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}$ and related components. The heavy-tailedness index is set to be $\varepsilon = 2$ or 0.2. The dimension and the sample size are $N = 500$ and $T = 250$, respectively. Reported values are mean and standardized deviation (in parentheses) from 100 replications for the error in relative Frobenius norm in estimating $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}$, and the error in ℓ_2 norm in estimating $\mathbf{\Sigma}_{0u}$, $\mathbf{\Sigma}_{0u}$, $\mathbf{\Sigma}_0^{-1}$ and $\mathbf{\Sigma}^{-1}$. We include our POET-IPSN and the following benchmark methods: the robust method GPOET-FLW and the original POET based on sample covariance matrix (POET-S).

	$\mathbf{\Sigma}_0$	$\mathbf{\Sigma}$	$\mathbf{\Sigma}_{0u}$	$\mathbf{\Sigma}_u$	$\mathbf{\Sigma}_0^{-1}$	$\mathbf{\Sigma}^{-1}$	$\mathbf{\Sigma}_{0u}^{-1}$	$\mathbf{\Sigma}_u^{-1}$
$\varepsilon = 2$								
POET-IPSN	0.241	0.091	0.107	0.122	1.023	2.894	1.029	1.161
	(0.008)	(0.020)	(0.012)	(0.034)	(0.209)	(0.396)	(0.210)	(0.301)
GPOET-FLW	0.328	0.127	0.545	0.632	1.533	2.995	1.542	1.621
	(0.103)	(0.085)	(0.768)	(1.360)	(0.354)	(0.290)	(0.357)	(0.318)
POET-S	0.479	0.186	0.206	0.224	1.423	3.235	1.430	1.494
	(0.483)	(0.278)	(0.239)	(0.318)	(0.437)	(0.284)	(0.439)	(0.289)
POET-RegTME	0.523	–	0.302	–	1.299	–	1.300	–
	(0.038)	–	(0.035)	–	(0.111)	–	(0.112)	–
$\varepsilon = 0.2$								
POET-IPSN	0.242	0.219	0.107	0.266	1.023	7.155	1.028	5.448
	(0.009)	(0.289)	(0.013)	(0.371)	(0.210)	(2.567)	(0.211)	(2.417)
GPOET-FLW	0.707	0.375	2.245	3.098	4.822	11.321	6.839	12.611
	(0.394)	(1.218)	(2.428)	(15.229)	(10.292)	(10.969)	(20.067)	(27.690)
POET-S	3.331	1.381	0.549	0.412	4.925	11.616	4.972	9.888
	(4.122)	(5.096)	(0.567)	(0.434)	(11.121)	(2.812)	(11.203)	(2.996)
POET-RegTME	0.519	–	0.302	–	1.292	–	1.295	
	(0.037)	–	(0.037)	–	(0.100)	–	(0.101)	

4 Empirical Studies

4.1 Global Minimum Variance Portfolio Optimization

The estimation of the minimum variance portfolio (MVP) in the high-dimensional setting has drawn considerable attention in recent years. The MVP solves (1.3) and is widely used to measure the performance of covariance matrix estimators (Fan et al. (2012); Ledoit and Wolf (2017); Ding et al. (2021)). Specifically, given an estimator of the inverse of the scatter matrix or the covariance matrix, $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ or $\widehat{\boldsymbol{\Sigma}}^{-1}$, one can estimate the MVP by

$$\widehat{\mathbf{w}}^* = \frac{1}{\mathbf{1}^\intercal \widehat{\boldsymbol{\Sigma}}_0^{-1} \mathbf{1}} \widehat{\boldsymbol{\Sigma}}_0^{-1} \mathbf{1}, \quad \text{or} \quad \widehat{\mathbf{w}}^* = \frac{1}{\mathbf{1}^\intercal \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}} \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}. \tag{4.1}$$

We can then evaluate the scatter matrix or covariance matrix estimator based on the risk of such constructed portfolio. In this section, we conduct empirical analysis to evaluate the performance of our proposed estimator in the MVP optimization.

4.2 Data and Compared Methods

We use the daily returns of S&P 500 Index constituent stocks between January 1995 and December 2023 to construct MVPs. At the beginning of each month, we use the historical returns of the stocks that stayed in the S&P 500 Index for the past five years to estimate the portfolio weights. We evaluate the risk of the portfolios based on their out-of-sample daily returns from January 2000 to December 2023.

We estimate the MVP using equation (4.1) with the following estimators: our POET estimator of the scatter matrix based on idiosyncratic-projected self-normalization, POET-IPSN; the GPOET-FLW estimator; the original POET estimator based on the sample covariance matrix, POET-S; and the POET estimator based on the regularized TME, POET-RegTME. When performing the generic POET procedure, the tuning parameter is chosen by cross-validation with the criterion of minimizing the out-of-sample risk. The number of factors is estimated by the method in Yu et al. (2019). We also include the equal-weight portfolio $(1/N, ..., 1/N)^\intercal$ as a benchmark portfolio, denoted by EW.

4.3 Out-of-Sample Performance

We use the standard deviation of out-of-sample returns to evaluate the risk of portfolios. In addition, we perform the following test to evaluate the statistical significance of the differences in risks between our POET-IPSN method and benchmark portfolios:

$$H_0 : \sigma \geq \sigma_0 \quad \text{vs.} \quad H_1 : \sigma < \sigma_0, \tag{4.2}$$

where σ denotes the standard deviation of the POET-IPSN portfolio and σ_0 denotes the standard deviation of a benchmark portfolio. To carry out the test, we adopt the method in Ledoit and Wolf (2011) and use the heteroskedasticity-autocorrelation-consistent (HAC) standard deviation estimator of the test statistic therein.

Table 3 reports the risks of various portfolios. We see from Table 3 that our POET-IPSN portfolio achieves a substantially lower risk than benchmark portfolios for the period between 2000 and 2023. The statistical test results suggest that the differences in the risk between our portfolio and benchmark portfolios are all statistically significantly negative. In addition, when checking the risk performance in different years, we see that the POET-IPSN portfolio performs robustly well. It outperforms the benchmark portfolios by achieving the lowest risk in the majority of the years (18 out of 24 years). The statistical test results further show that the difference between the risks of our POET-IPSN portfolio and other portfolios is statistically significantly negative for most of the years.

5 Conclusion

We propose an idiosyncratic-projected self-normalization (IPSN) approach to estimate the scatter matrix under an ultra-high dimensional elliptical factor model set-

Table 3: Out-of-sample risks of the portfolios. We report the annualized standard deviations. We construct the MVP portfolio using equation (4.1) with the covariance matrix estimated using POET based on our idiosyncratic-projected self-normalized method (POET-IPSN), the robust method GPOET-FLW, the sample covariance matrix (POET-S), and the regularized TME (POET-RegTME). We also include the equal weight (EW) portfolio as a benchmark. The out-of-sample period is from January 2000 to December 2023. The risk is evaluated using out-of-sample daily returns of the portfolios. For all methods, the portfolios weights are estimated based on prior five-year daily excess returns of S&P 500 Index constituent stocks, reestimated monthly. The symbols *, ** and *** indicate statistical significance for testing (4.2) at 5%, 1% and 0.1% levels, respectively.

	Out-of-sample portfolio risk				
Period	2000–2023	2000	2001	2002	2003
POET-IPSN	0.121	0.154	0.116	0.140	0.090
GPOET-FLW	0.150 ***	0.168 **	0.167 ***	0.173 ***	0.121 ***
POET-S	0.150 ***	0.152	0.121	0.140	0.095 **
POET-RegTME	0.156 ***	0.203 ***	0.168 ***	0.165 ***	0.133 ***
EW	0.217 ***	0.192 ***	0.208 ***	0.269 ***	0.179 ***
Period	2004	2005	2006	2007	2008
POET-IPSN	0.081	0.084	0.068	0.081	0.208
GPOET-FLW	0.095 ***	0.100 ***	0.089 ***	0.100 ***	0.224
POET-S	0.091 ***	0.091 ***	0.074**	0.087 ***	0.188
POET-RegTME	0.122 ***	0.105 ***	0.064	0.092 **	0.219
EW	0.128 ***	0.115 ***	0.113 ***	0.166 ***	0.460 ***
Period	2009	2010	2011	2012	2013
POET-IPSN	0.133	0.083	0.103	0.082	0.084
GPOET-FLW *	0.228 ***	0.094 **	0.108	0.073	0.081
POET-S	0.189 ***	0.089 *	0.114 **	0.084	0.089
POET-RegTME	0.163 ***	0.090 **	0.108	0.077	0.083
EW	0.357 ***	0.210 ***	0.271 ***	0.148***	0.123 ***
Period	2014	2015	2016	2017	2018
POET-IPSN	0.076	0.096	0.099	0.072	0.105
GPOET-FLW	0.081	0.114 ***	0.128 ***	0.076	0.116 *
POET-S	0.089	0.120 ***	0.116 **	0.073	0.119 ***
POET-RegTME	0.079	0.109 **	0.119 ***	0.085 **	0.129 ***
EW	0.119 ***	0.158 ***	0.157 ***	0.077	0.161 ***
Period	2019	2020	2021	2022	2023
POET-IPSN	0.097	0.269	0.106	0.141	0.125
GPOET-FLW	0.131 ***	0.317 **	0.184 ***	0.179 **	0.176 ***
POET-S	0.101	0.413 ***	0.194 ***	0.226 ***	0.168 ***
POET-RegTME	0.141 ***	0.361 **	0.171 ***	0.229 **	0.171 ***
EW	0.137 ***	0.404 ***	0.144 ***	0.239 ***	0.152 ***

ting. We show that our estimator achieves the sub-Gaussian convergence rate under only $2+\varepsilon$th moment condition and a setting where N can be much larger than T. Moreover, under a conditional sparsity assumption, we develop a POET estimator of the scatter matrix, and show that it achieves the same convergence rate as in the sub-Gaussian case. Numerical studies demonstrate the clear advantages of our proposed estimators over various benchmark methods.

Supplementary Materials

Supplement to "**Sub-Gaussian estimation of the scatter matrix in ultra-high dimensional elliptical factor models with $2+\varepsilon$th moment.**"

This supplement contains the proofs of Theorems 1–2 and Propositions 1–3.

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Avella-Medina, M., Battey, H. S., Fan, J., and Li, Q. (2018). Robust estimation of high-dimensional covariance and precision matrices. *Biometrika*, 105(2):271–284.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bollerslev, T., Todorov, V., and Li, S. Z. (2013). Jump tails, extreme dependencies, and the distribution of stock returns. *Journal of Econometrics*, 172(2):307–324.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684.

Cai, T. T., Liu, W., and Zhou, H. H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *The Annals of Statistics*, 44(2):455–488.

Chatterjee, S. (2013). Assumptionless consistency of the lasso. *arXiv preprint arXiv:1303.5817*.

Ding, Y., Engle, R., Li, Y., and Zheng, X. (2024). Factor modeling for volatility.

Ding, Y., Li, Y., Liu, G., and Zheng, X. (2023). Stock co-jump networks. *Journal of Econometrics*.

Ding, Y., Li, Y., and Zheng, X. (2021). High dimensional minimum variance portfolio estimation under statistical factor models. *Journal of Econometrics*, 222(1):502–515.

Fan, J., Li, Y., and Yu, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association*, 107(497):412–428.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.

Fan, J., Liu, H., and Wang, W. (2018). Large covariance estimation through elliptical factor models. *The Annals of Statistics*, 46(4):1383.

Fan, J., Wang, W., and Zhu, Z. (2021). A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery. *The Annals of Statistics*, 49(3):1239.

Goes, J., Lerman, G., and Nadler, B. (2020). Robust sparse covariance estimation by thresholding tylerâĂŹs m-estimator. *Annals of Statistics*, 48(1):86–110.

Han, F. and Liu, H. (2014). Scale-invariant sparse pca on high-dimensional meta-elliptical data. *Journal of the American Statistical Association*, 109(505):275–287.

Han, F. and Liu, H. (2018). Eca: High-dimensional elliptical component analysis in non-gaussian distributions. *Journal of the American Statistical Association*, 113(521):252–268.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, 119(2):249–283.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, pages 73–101.

Hubert, M., Debruyne, M., and Rousseeuw, P. J. (2018). Minimum covariance determinant and extensions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(3):e1421.

Jacod, J., Lin, H., and Todorov, V. (2024). Systematic jump risk. *The Annals of Applied Probability*, 34(5):4342–4386.

Ledoit, O. and Wolf, M. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, 2011(55):86–89.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*, 30(12):4349–4388.

Li, J., Todorov, V., and Tauchen, G. (2017). Jump regressions. *Econometrica*, 85(1):173–195.

Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. (2012). High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326.

Maronna, R. A. (1976). Robust m-estimators of multivariate location and scatter. *The Annals of Statistics*, pages 51–67.

Minsker, S. (2018). Sub-gaussian estimators of the mean of a random matrix with heavy-tailed entries. *The Annals of Statistics*, 46(6A):2871–2903.

Minsker, S. and Wei, X. (2020). Robust modifications of u-statistics and applications to covariance estimation problems. *Bernoulli*, 26(1):694–727.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388):871–880.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, 8(283-297):37.

Sun, Q., Zhou, W.-X., and Fan, J. (2020). Adaptive huber regression. *Journal of the American Statistical Association*, 115(529):254–265.

Sun, Y., Babu, P., and Palomar, D. P. (2014). Regularized tyler's scatter estimator: Existence, uniqueness, and algorithms. *IEEE Transactions on Signal Processing*, 62(19):5143–5156.

Susmel, R. and Engle, R. F. (1994). Hourly volatility spillovers between international equity markets. *Journal of international Money and Finance*, 13(1):3–25.

Tyler, D. E. (1987). A distribution-free m-estimator of multivariate scatter. *The Annals of Statistics*, pages 234–251.

Van Aelst, S. and Rousseeuw, P. (2009). Minimum volume ellipsoid. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1):71–82.

Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40(5):2541–2571.

Yu, L., He, Y., and Zhang, X. (2019). Robust factor number specification for large-dimensional elliptical factor model. *Journal of Multivariate Analysis*, 174:104543.

Zhang, T. (2016). Robust subspace recovery by tyler's m-estimator. *Information and Inference: A Journal of the IMA*, 5(1):1–21.

Zhang, T., Cheng, X., and Singer, A. (2016). Marčenko–pastur law for tyler's m-estimator. *Journal of Multivariate Analysis*, 149:114–123.

Zheng, X. and Li, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. *The Annals of Statistics*, 39(6):3121–3151.