Large-Dimensional Factor Analysis Without Moment Constraints

Yong He, Xinbing Kong, Long Yu, and Xinsheng Zhang

ABSTRACT

Large-dimensional factor model has drawn much attention in the big-data era, to reduce the dimensionality and extract underlying features using a few latent common factors. Conventional methods for estimating the factor model typically requires finite fourth moment of the data, which ignores the effect of heavy-tailedness and thus may result in unrobust or even inconsistent estimation of the factor space and common components. In this article, we propose to recover the factor space by performing principal component analysis to the spatial Kendall’s tau matrix instead of the sample covariance matrix. In a second step, we estimate the factor scores by the ordinary least square regression. Theoretically, we show that under the elliptical distribution framework the factor loadings and scores as well as the common components can be estimated consistently without any moment constraint. The convergence rates of the estimated factor loadings, scores, and common components are provided. The finite sample performance of the proposed procedure is assessed through thorough simulations. An analysis of a financial dataset of asset returns shows the superiority of the proposed method over the classical principle component analysis method.

1. Introduction

Factor model is a classical statistical model that serves as an important dimension reduction tool by characterizing the dependency structure of variables via a few latent factors. In the “big-data era” where more and more variables are recorded and stored, large-dimensional approximate factor model is drawing growing attention as it provides an effective way of summarizing information from large datasets. The large-dimensional approximate factor models are widely used in genomics, neuroscience, computer science, and financial economics. Theoretical analysis of large-dimensional approximate factor models has been studied by many researchers. Existing factor analysis procedures mainly fall into two categories: the principle component analysis (PCA) approach and the maximum likelihood estimation (MLE) method. The PCA-based method is easy to implement and provides consistent estimators for the factors and factor loadings when both the cross-section and time dimension are large. Representative works include, but not limited to, Bai and Ng (2002), Stock and Watson (2002a, 2002b), Bai (2003), Onatski (2009), Ahn and Horenstein (2013), Fan, Liao, and Mincheva (2013), and Trapani (2018). It turns out that the PCA approach is equivalent to the least square optimization. The MLE-based method is more efficient than the PCA-based approach but is also computationally more suffering. Representative works, to name a few, are Bai and Li (2012, 2014, 2016).

However, the aforementioned works all assume that the fourth moments (or even higher moments) of factors and idiosyncratic errors are bounded such that the least-squares regression, or MLE can be applied. This assumption is really an idealization of the complex random real world. Heavy-tailed data are often encountered in scientific fields such as financial engineering and biomedical imaging. In finance, Fama (1963) discussed the power law behavior of asset returns. Cont (2001) provided extensive empirical evidence of heavy-tailedness in financial returns. Jing, Kong, and Liu (2012) and Kong, Liu, and Jing (2015) even suggested to model the log price dynamics of an asset by pure jump processes without any moment conditions. Thus, it is imperative to develop estimation procedures that are robust to heavy-tailedness for large-dimensional factor models. After resubmitting the article, we are notified by one reviewer an interesting concurrent working article by Chen, Dolado, and Gonzalo (2019), which considered a quantile factor model without requiring any moment of the idiosyncratic errors. Our article differs from theirs in several aspects. First, their estimation is more computationally involved while ours is simple to use; second, their article imposed restrictions on the error density function while our article assumed elliptical condition on the error distribution; third, the factor series is restricted to be bounded in minimizing the loss function in that article while ours are random and could explode as n goes to infinity.

As an illustration, we check the sensitivity of the PCA (or least square optimization) method to the heavy-tailedness of the factor and idiosyncratic errors with a synthetic dataset. We generate the factors and idiosyncratic errors from joint normal, $t_3$, $t_2$, and $t_1$ distributions that will be described in detail in Section 4. Figure 1 depicts the boxplots of the factor loading and factor score estimation errors based on 1000 replications. We observe that the PCA results in bigger biases and higher

ARTICLE HISTORY

Received August 2019
Accepted August 2020

KEYWORDS

Elliptical factor model; Multivariate Kendall’s tau matrix; Ordinary least square regression
dispersions as the distribution tails become heavier. This is also consistent with the fourth moment condition imposed in existing articles.

In this article, we propose a robust two-step (RTS) procedure to estimate the factor loadings, scores, and common components without any moment constraint under the framework of elliptical distributions (FED). The FED assumes that the factors and the idiosyncratic errors jointly follow an elliptical distribution, which covers a large class of heavy-tailed distributions such as t-distribution. The FED is drawing growing attention as an important tool to simultaneously simplify the structure and capture the heavy-tailedness of the data. For example, Fan, Liu, and Wang (2018) considered large-scale covariance estimation and the indicator function. Let diag

\[ \mathbf{C} \]  

and the idiosyncratic errors jointly follow an elliptical distribution, the estimated factor scores are consistent

\[ \mathbf{C} \]  

and Chroux, Ollila, and Oja (2002). For high-dimensional settings, Han and Liu (2018) studied the eigen-analysis of spatial Kendall’s tau matrix for elliptical distributions. Similar rank-based methods in high dimensions are also discussed in Han and Liu (2012, 2014). The mentioned literatures provide a sound framework of constructing Bernstein’s type concentration inequalities for high-dimensional matrix-form U-statistics, but none focused on the factor models. As a result, the existing convergence rates for robust principal component analysis are not optimal if the observed vectors contain some low-rank factor structures. To the best of our knowledge, Fan, Liu, and Wang (2018) is the first to consider factor structures with spatial Kendall’s tau matrix in high dimensions, which proposed the elliptical factor model exactly the same as the model considered in the current article. However, the motivations of the two works are quite different. Fan, Liu, and Wang (2018) focused on robust covariance estimation while our work focused on robust estimation for the factor loadings and scores.

The contributions of the current article lie in the following aspects. First, it is the first to estimate the factor scores and loadings using spatial Kendall’s tau matrix in high-dimensional settings. The proposed method is computationally efficient and easy to implement. Second, our theoretical analysis shows that the proposed robust estimators are consistent without any moment constraints on the underlying distributions, which makes the method applicable to analyzing heavy-tailed datasets such as financial returns and macroeconomic indicators. Third, the convergence rates of the RTS estimators are the same as those in Bai (2003), which are faster than those obtained in Fan, Liu, and Wang (2018) and Chen, Dolado, and Gonzalo (2019). We assumed temporal independence of the data for technical simplicity. We leave the extension to the dependence case in our future work. We overcome two major challenges in the technical proofs: (1) the summing terms in the sample spatial Kendall’s tau matrix are dependent, which makes the typical Bernstein’s inequalities unapplicable; (2) the sample spatial Kendall’s tau matrix is a nonlinear function of the observed vectors, which in essence makes the theoretical analysis more challenging.

We introduce the notations adopted throughout the article. For any vector \( \mathbf{\mu} = (\mu_1, \ldots, \mu_p)^\top \in \mathbb{R}^p \), let \( \| \mathbf{\mu} \|_2 = (\sum_{i=1}^{p} \mu_i^2)^{1/2} \), \( \| \mathbf{\mu} \|_{\infty} = \max_i |\mu_i| \). For a real number \( a \), denote \( [a] \) as the largest integer smaller than or equal to \( a \). Let \( I(\cdot) \) be the indicator function. Let \( \text{diag}(a_1, \ldots, a_p) \) be a \( p \times p \) diagonal
matrix, whose diagonal entries are $a_1, \ldots, a_p$. For a matrix $A$, let $A_{ij}$ (or $A_{ji}$) be the $ij$ entry of $A$, $A^\top$ the transpose of $A$, $\text{Tr}(A)$ the trace of $A$, $\text{rank}(A)$ the rank of $A$ and $\text{diag}(A)$ a vector composed of the diagonal elements of $A$. Denote $\lambda_j(A)$ as the $j$th largest eigenvalue of a nonnegative definite matrix $A$, and let $\|A\|$ be the spectral norm of matrix $A$ and $\|A\|_F$ be the Frobenius norm of $A$. For two series of random variables, $X_n$ and $Y_n$, $X_n \overset{d}{=} Y_n$ means $X_n = \text{Op}(Y_n)$ and $Y_n = \text{Op}(X_n)$. For two random variables (vectors) $X$ and $Y$, $\|X - Y\|$ means the distributions of $X$ and $Y$ are the same. Let $I$ be a vector with all elements 1. The constants $c, C_1, C_2$ in different lines can be nonidentical.

The rest of the article proceeds as follows. In Section 2, we introduce the setup assumptions and multivariate Kendall's tau matrix. Estimators of the factor loadings, scores, and common components are also provided. In Section 3, we establish the consistency including the convergence rate for the estimated factor loadings, scores, and common components. Section 4 is devoted to a thorough numerical study. A real financial dataset of asset returns is analyzed in Section 5. We discuss the possible future research directions and conclude the article in Section 6. The proofs of the main theorems are collected in the Appendix and additional details are put in the supplementary materials.

2. Methodology

2.1. Elliptical Distribution and Spatial Kendall’s Tau Matrix

Consider the large-dimensional factor model for a large panel dataset $\{y_{it}\}_{t \leq n, i \leq p}$.

$$y_{it} = \mu_i^\top f_t + \epsilon_{it}, \quad i \leq p, \quad t \leq n,$$

or in vector form, $y_t = (y_{t1}, \ldots, y_{tp})^\top$, $f_t \in \mathbb{R}^m$ are the unobserved factors, $L = (l_{11}, \ldots, l_{pq})^\top$ is the factor loading matrix, and $\epsilon_t = (\epsilon_{t1}, \ldots, \epsilon_{tp})^\top$ represents the idiosyncratic errors. The term $\mu_i = \mu_i^\top f_i$ is referred to as the common component of $y_{it}$. For the large-dimensional approximate factor model introduced in Chamberlain and Rothschild (1983), $\epsilon_t$ is assumed to be cross-sectionally weakly dependent.

As mentioned in the introduction and precisely stated in Assumption A, we assume that $(f_{i1}^\top, \epsilon_{i1}^\top)^\top$ is a series of temporally independent and identically distributed random vectors generated from an elliptical distribution. For a random vector $Z = (Z_1, \ldots, Z_p)^\top$ following an elliptical distribution, denoted by $Z \sim \text{ED}(\mu, \Sigma, \xi)$, we mean that

$$Z \overset{d}{=} \mu + \xi A U,$$

where $\mu \in \mathbb{R}^p$, $U$ is a random vector uniformly distributed on the unit sphere $S^{q-1}$ in $\mathbb{R}^q$, $\xi \geq 0$ is a scalar random variable independent of $U$, $A \in \mathbb{R}^{p \times q}$ is a deterministic matrix satisfying $AA^\top = \Sigma$ with $\Sigma$ called scatter matrix whose rank is $q$. Let the scatter matrices of $f_t$ and $\epsilon_t$ be $\Sigma_f$ and $\Sigma_\epsilon$, respectively. If $\Sigma_\epsilon$ is sparse, the model (2.1) is indeed an approximate factor model including the strict factor model, in which $\Sigma_\epsilon$ is diagonal, as a special case. Another equivalent characterization of the elliptical distribution is by its characteristic function, which has the form $\exp(it^\top \mu \psi(t^\top \Sigma t))$, where $\psi(\cdot)$ is a properly defined function and $i = \sqrt{-1}$. Elliptical family includes many distributions as special examples, such as multivariate Gaussian distribution, t-distribution, logistic distribution, and symmetric Laplace distribution, but it rules out $\alpha$-stable distribution for $\alpha \neq 1$.

The factor loadings and scores, $L$ and $f_t$, are not separately identifiable as they are unobservable. For an arbitrary $m \times m$ invertible matrix $H$, one can always have $L^\top = LH$ and $f_t^\top = H^{-1} f_t$, such that $L^\top f_t^\top = L f_t$. For reason of identifiability, we impose the following constraints:

$$\Sigma_f = I_m \quad \text{and} \quad \|\text{diag}(\Sigma_f)\|_\infty = 1.$$

This identification condition is also used in Han and Liu (2014) and Yu, He, and Zhang (2019), and it is not unique and one may refer to Bai and Li (2012) for more detailed discussion on identification issues.

It is worthy of pointing out that elliptical distributions have some nice properties as Gaussian distributions, for example, the marginal distributions, conditional distributions, and distributions of linear combinations of elliptical vectors are also elliptical. Thus, for the factor model (2.1) under the FED condition, the scatter matrix of $y_t$, $\Sigma_f$, is composed of a low-rank part $LL^\top$ and a sparse part $\Sigma_\epsilon$, that is, $\Sigma_f = LL^\top + \Sigma_\epsilon$. For Gaussian distribution, $\Sigma_f$ is simply the population covariance matrix of $y_t$. For non-Gaussian distributions, especially distributions with infinite variances, the scatter matrix is still a measure of the dispersion of a random vector. So, naturally the eigenspace of the scatter matrix $\Sigma_f$ sheds light into the recovery of the factor space, but this cannot be reached by performing PCA to the sample covariance matrix because the covariance is meaningless for pairs of random variables with infinite variances. To tackle with this difficulty, we introduce the population spatial Kendall’s tau matrix. Let $X \sim \text{ED}(\mu, \Sigma, \xi)$ and $\tilde{X}$ be an independent copy of $X$. The population spatial Kendall’s tau matrix is defined as

$$K = E \left\{ \frac{(X - \tilde{X})(X - \tilde{X})^\top}{\|X - \tilde{X}\|_2^2} \right\}.$$

$K$ can be estimated by a second-order $U$-statistic. Specifically, assume $\{X_1, \ldots, X_n\}$ is a series of $n$ independent data points following the distribution $X \sim \text{ED}(\mu, \Sigma, \xi)$. The sample version spatial Kendall’s tau matrix is

$$\hat{K} = \frac{2}{n(n - 1)} \sum_{t < t'} \frac{(X_t - X_{t'})(X_t - X_{t'})^\top}{\|X_t - X_{t'}\|_2^2}.$$

The spatial Kendall’s tau matrix was first introduced in Choi and Marden (1998) and has been used for covariance matrix estimation in Visuri, Koivunen, and Oja (2000) and Fan, Liu, and Wang (2018) and principal component estimation in Han and Liu (1999) and Han and Liu (2018). A critical result is that the spatial Kendall’s tau matrix $K$ shares the same ordering of eigenvalues and the same eigenspace as those of the scatter matrix $\Sigma$. We cite this result directly without proof in Lemma 2.1.

**Lemma 2.1.** Let $X$ be a continuous elliptically distributed random vector, that is, $X \sim \text{ED}(\mu, \Sigma, \xi)$ with $P(\xi = 0) = 0$ and...
Let $K$ be the population multivariate Kendall's tau statistic. Further assume that $\text{rank}(\Sigma) = q$, we have

$$\lambda_i(K) = E\left( \frac{\lambda_i(\Sigma) g_i^2}{\lambda_1(\Sigma) g_1^2 + \cdots + \lambda_q(\Sigma) g_q^2} \right),$$

where $g = (g_1, \ldots, g_q)^T \sim N(0, I)$, and in addition $K$ and $\Sigma$ share the same eigenspace with the same descending order of the eigenvalues.

The proof of Lemma 2.1 can be found in Han and Liu (2018). By Lemma 2.1, estimating the eigenvectors of $\Sigma$ is equivalent to estimating those of $K$, and thus $\tilde{K}$ fits the goal of estimating the eigenvectors of $\Sigma$.

### 2.2. Robust Two-Step Estimation Procedure

In this section, we introduce an innovative two-step estimation procedure for large-dimensional elliptical factor model. In the first step, we propose to estimate $L$ by the eigenvectors of the spatial Kendall's tau matrix. First, we estimate the spatial Kendall's tau matrix of $y_t$ by

$$\tilde{K}_y = \frac{2}{n(n-1)} \sum_{i<j} (y_{it} - y_{jt})(y_{it} - y_{jt})^T, \quad (2.2)$$

As the eigenvectors of the spatial Kendall’s tau matrix $K_y$ is identical to the eigenvectors of the scatter matrix $\Sigma_y$, we thus estimate the factor loading matrix $L$ by $\sqrt{p}$ times the leading $m$ eigenvectors of $\tilde{K}_y$. In detail, let $(\tilde{f}_1, \ldots, \tilde{f}_m)$ be the leading $m$ eigenvectors of $\tilde{K}_y$ and let $\tilde{F} = (\tilde{f}_1, \ldots, \tilde{f}_m)$. We take $\tilde{L} = \sqrt{p}\tilde{F}$ as the estimator of the factor loading matrix $L$. The number of factors $m$ is relatively small compared with $n$ and $p$. We first assume that $m$ is known and fixed. If $m$ is unknown, we can estimate $m$ consistently as in Yu, He, and Zhang (2019).

In a second step, we estimate the factors $(f_{it}, t = 1, \ldots, n)$ by regressing $y_t$ on $\tilde{L} f_t$ is estimated by the following least square optimization,

$$\hat{f}_t = \arg \min_{\beta_t \in \mathbb{R}^m} \sum_{i=1}^p (y_{it} - \tilde{L}^T \beta_t)^2, \quad t = 1, \ldots, n, \quad (2.3)$$

where $\tilde{L}^T_i$ is the $i$th row of $\tilde{L}$, that is, $\tilde{L} = (\tilde{L}_1, \ldots, \tilde{L}_p)^T$. For conventional factor model, when both $n$ and $p$ are large, the factor loadings and the factors can be estimated by PCA, which is equivalent to solving a double least-square regression problem, see Bai and Ng (2002) or (2.4) in Fan, Liu, and Wang (2018). The two-step estimation procedure is motivated by the idea of the regression formulation.

### 3. Theoretical Results

In this section, we investigate the theoretical properties of the proposed estimators $\hat{L}$ and $\hat{F} = (\hat{f}_1, \ldots, \hat{f}_n)^T$. We need the following technical assumptions.

**Assumption A.** We assume that

$$\begin{pmatrix} f_{it} \\ \epsilon_{it} \end{pmatrix} = \zeta_i \begin{pmatrix} I_m \\ 0 \end{pmatrix} g_i / \|g_i\|,$$

where $\zeta_i$’s are independent samples of a scalar random variable $\zeta$, and $g_i$’s are independent Gaussian samples from $g \sim N(0, I)$, and in addition $K$ and $\Sigma$ share the same eigenspace with the same descending order of the eigenvalues.

**Theorem 3.1.** Under Assumptions A–C, there exist a series of matrices $\tilde{H}$ (dependent on $n, p$, and $\hat{L}$) so that $\hat{L}^T V \hat{H} \overset{D}{\rightarrow} I_m$ as $\min(n, p) \to \infty$ and

$$\frac{1}{p} \|\hat{L} - L\|_F^2 = O_p\left( \frac{1}{n} + \frac{1}{p} \right).$$

In Theorem 3.1, we obtain the same convergence rate of the estimated factor loadings as that in Bai (2003). However, we impose no moment constrains on the factors and idiosyncratic errors. In the following theorem, we establish the convergence rate of the estimated factor scores $f_t$.

**Theorem 3.2.** Assume that Assumptions A–C hold and $\min(n, p) \to \infty$, then for any $t \leq n$,

$$\|\hat{f}_t - f_t\| = O_p\left( \frac{1}{p} + \frac{1}{n^2} \right).$$
By the results in Theorems 3.1 and 3.2, we finally show that the estimated common components are consistent to the true ones.

**Theorem 3.3.** Assume that Assumptions A–C hold and \( \min(n,p) \to \infty \), then for any \( t \leq n \),

\[
\frac{1}{p} \left\| \hat{L}_t f_t - L f_t \right\|^2 = O_p \left( \frac{1}{n} + \frac{1}{p} \right).
\]

As far as we know, this is the first time that consistent estimators for the factor loadings, scores, and common components are proposed without any moment constraints. Under the elliptical assumption containing heavy-tailed cases, our RTS estimators converge at the same rates as those of the PCA estimators with finite fourth moment constraints on the factors and errors, see Bai (2003).

### 4. Simulation Study

In this section, we conduct thorough simulation studies to compare the RTS estimator with the conventional PCA method. We use similar data-generating models as in Ahn and Horenstein (2013), Xia, Liang, and Wu (2017), and Yu, He, and Zhang (2019). We generate the data from the following model,

\[
y_{it} = \sum_{j=1}^{m} L_{ij} f_{jt} + \sqrt{\theta} \epsilon_{it}, \quad u_{it} = \sqrt{1 - \rho^2} \tilde{e}_{it},
\]

\[
e_{it} = \rho e_{i,t-1} + (1 - \beta) v_{it} + \sum_{l=\max(i-t,1)}^{\min(i+t,p)} \beta v_{lt},
\]

\[t = 1, \ldots, n, \quad i = 1, \ldots, p,
\]

where \( f_t = (f_{t1}, \ldots, f_{tm})^\top \) and \( v_t = (v_{t1}, \ldots, v_{tp})^\top \) are jointly generated from elliptical distributions. We let \( L_i \) be independently drawn from the standard normal distribution. The parameter \( \theta \) controls the signal-to-noise ratio (SNR), \( \rho \) controls the serial correlations of idiosyncratic errors, and \( \beta \) and \( \beta \) control the cross-sectional correlations. We point out that although we assume \( y_i \)'s are temporally independent theoretically in Assumption A, we allow \( u_t \) to be serially correlated in the simulation studies.

Before we give the data generating scenarios, we first review the multivariate \( t \) distribution. The probability distribution function (PDF) of a \( d \)-dimensional multivariate \( t \) distribution \( t_n(\mu, \Sigma) \) is

\[
\frac{\Gamma((v + d)/2)}{\Gamma(v/2)\sqrt{2\pi^d|\Sigma|/|v|}} \left( 1 + \frac{1}{v} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right)^{-(v+1)/2},
\]

where \( \Gamma(\cdot) \) is the gamma function. In fact, multivariate \( t \) distribution with \( v = 1 \) is the multivariate Cauchy distribution that has no finite mean. We also consider the following data generating scenarios in the simulation studies.

**Scenario A.** Set \( m = 3, \theta = 1, \rho = \beta = J = 0 \), \( (\rho, p) = \{(150, 100), (250, 100), (250, 150), (250, 200)\} \), \( (f^\top_t, v^\top_t) \) are generated in the following ways: (i) \((f^\top_t, v^\top_t)\) are iid jointly elliptical random samples from multivariate Gaussian distributions \( N(0, I_{p+m}) \); (ii) \((f^\top_t, v^\top_t)\) are iid jointly elliptical random samples from multivariate centralized \( t \) distributions \( t_v(0, I_{p+m}) \) with \( v = 3, 2, 1 \); (iii) \((f^\top_t, v^\top_t)\) are iid samples from multivariate skewed-\( t \) distribution; (iv) \((f^\top_t, v^\top_t)\) are iid samples from multivariate Gaussian distributions \( N(0, I_{p+m}) \) while the elements of \( v_t \) are iid random samples from symmetric \( \alpha \)-stable distribution \( ST_\alpha(\beta, \gamma, \Delta) \) with skewness parameter \( \beta = 0 \), scale parameter \( \gamma = 1 \) and location parameter \( \delta = 0, \alpha = 1.8 \).

**Scenario B.** Set \( m = 3, \theta = 1, \rho = 0.5, \beta = 0.2, J = \max(10, p/20) \), \((f^\top_t, v^\top_t)\) are generated in the same ways as in Scenario A. \( (\rho, n) = \{(150, 100), (250, 100), (250, 150), (250, 200)\} \).

**Scenario C.** Set \( m = 3, \theta = 1, \rho = 0.5, \beta = 0.2, J = \max(10, p/20) \), \((\rho, n) = \{(250, 100), (250, 150), (f^\top_t, v^\top_t)\}) \) are iid jointly elliptical random vectors from multivariate Gaussian \( N(0, D) \) and multivariate centralized \( t \) distribution \( t_v(0, D) \) with \( v = 3 \), where \( D \) is a diagonal matrix of dimension \((p + m) \times (p + m)\) with \( D_{ii} = 1, i \neq 3 \) and \( D_{33} = \text{SNR} \) with SNR from \[0.7, 0.6, 0.5, 0.4]\).

In Scenario A, the setting perfectly fits to our assumption with no serial correlations of idiosyncratic errors and \((f^\top_t, v^\top_t)\) are from light-tailed Gaussian \( N(0, I_{p+m}) \) or heavy-tailed \( t_v(0, I_{p+m}) \) with \( v = 3, 2, 1 \). Note that when \( v = 1 \), it is indeed the Cauchy distribution which does not have finite moments of order greater than or equal to one. We also consider the skewed \( t_3 \) and \( \alpha \)-stable distributions to gauge how sensitive the method is to the elliptical distribution assumption. We generate the multivariate skewed \( t_3 \) random samples from \[ST_{N+\alpha}(k = 0, \Omega = I, \alpha = 20, \nu = 3)\] by function \texttt{msnvm} in R package \texttt{Multivar}. Scenario B is a simple case containing both serially and cross-sectionally correlated errors from Gaussian distribution, \( t \) distribution with degree 1,2,3, skewed \( t_3 \) distribution or \( \alpha \)-stable distribution. Scenario C corresponds to a case where both serially and cross-sectionally correlated errors, and strong and weak factors exist. To evaluate the empirical performance of different methods, we consider the following indices: the median of the normalized estimation errors for common components in terms of the matrix Frobenius norm, denoted as MEE-CC; the average estimation error for the factor loading matrices, denoted as AVE-FL; and the average estimation error for the factor score matrices, denoted as AVE-FS. Specifically,

\[
\text{MEE-CC} = \text{median}\left\{\|\hat{L}_t F_t - LF_t\|_F^2 / \|LF_t\|_F^2, r = 1, \ldots, R\right\},
\]

\[
\text{AVE-FL} = \frac{1}{R} \sum_{r=1}^{R} D(\hat{L}_r, L), \quad \text{AVE-FS} = \frac{1}{R} \sum_{r=1}^{R} D(\hat{F}_r, F),
\]

where \( R \) is the replication times, \( \hat{L}_r \) and \( \hat{F}_r \) are the estimators for the \( r \)th replication, and for two orthogonal matrices \( O_1 \) and \( O_2 \) of sizes \( p \times q_1 \) and \( p \times q_2 \),

\[
D(O_1, O_2) = \left( 1 - \frac{1}{\max(q_1, q_2)} \text{Tr}(O_1 O_1^\top O_2 O_2^\top) \right)^{1/2}.
\]

The Gram–Schmidt orthonormal transformation can be used when \( O_1 \) and \( O_2 \) are not column-orthogonal matrices. In fact, \( D(O_1, O_2) \) measures the distance between the column
spaces of $O_1$ and $O_2$, and it is a quantity between 0 and 1. It is equal to 0 if the column spaces of $O_1$ and $O_2$ are the same and 1 if they are orthogonal. As the factor loading matrix and factor score matrix are not separately identifiable, $D(\cdot,\cdot)$ particularly suits to quantify the accuracy of factor loading/score matrices estimation.

The simulation results for Scenario A, Scenario B, and Scenario C are reported in Tables 1–3, respectively. For Scenario A, from Table 1, we can see that in Gaussian setting, PCA performs slightly better than the RTS in terms of MEE_CC, AVE_FL, and AVE_FS while the performances of the two methods are still comparable. In the heavy-tailed and $\alpha$-stable RTS performs better. In the skewed- $\alpha$-stable RTS is a safe replacement of the conventional PCA method in practice.

The raw dataset is a “105 weeks” $\times$ “100 shares” panel without missing values. We first calculate the sample autocorrelation functions, which indicate that no significant serial correlations exist for most of the weekly return series. We also performed the augmented Dickey–Fuller tests and found that all the series are stationary.

We use the centralized log returns to do factor analysis. We leave the centralized log returns unscaled since volatilities of all assets are themselves very informative in portfolio allocation, risk management and derivatives pricing. As for the factor number, the “eigenvalue-ratio” method proposed by Ahn and Horenstein (2013) and its robust version proposed by Yu, He, and Zhang (2019) both lead to an estimate of just 1 common factor. Inspired by the Fama–French 3 factor model, we also consider $m = 3$ in this example for comparison.

Table 1. Simulation results for Scenario A, the values in the parentheses are the interquartile ranges for MEE-CC and standard deviations for AVE-FL and AVE-FS.

| Type           | Method | $(p, n) = (150, 100)$ | $(p, n) = (250, 100)$ |
|----------------|--------|----------------------|----------------------|
| $\mathcal{N}(0, D)$ | RTS    | 0.02(0.00)           | 0.01(0.00)           |
|                | PCA    | 0.01(0.00)           | 0.00(0.00)           |
| $t_3(0, I_{p+m})$ | RTS    | 0.02(0.00)           | 0.01(0.00)           |
|                | PCA    | 0.01(0.00)           | 0.00(0.00)           |
| $t_2(0, I_{p+m})$ | RTS    | 0.02(0.00)           | 0.01(0.00)           |
|                | PCA    | 0.01(0.00)           | 0.00(0.00)           |
| $t_1(0, I_{p+m})$ | RTS    | 0.02(0.00)           | 0.01(0.00)           |
|                | PCA    | 0.01(0.00)           | 0.00(0.00)           |
| Skewed $t_3$   | RTS    | 0.02(0.00)           | 0.01(0.00)           |
|                | PCA    | 0.01(0.00)           | 0.00(0.00)           |
| $\alpha$-stable| RTS    | 0.06(0.02)           | 0.05(0.02)           |
|                | PCA    | 0.14(0.70)           | 0.37(0.22)           |

5. Real Example: S&P 100 Weekly Returns Panel

In this section, we apply the proposed method to study the weekly share returns of Standard & Poor 100 component companies during the period from January 1, 2018 to December 31, 2019. Details of the dataset are available upon request, including the symbol list and names of the corresponding companies. The raw dataset is a “105 weeks” $\times$ “100 shares” panel without missing values. We first calculate the sample autocorrelation functions, which indicate that no significant serial correlations exist for most of the weekly return series. We also performed the augmented Dickey–Fuller tests and found that all the series are stationary.
Given by matrix \( \Sigma \) of the share returns, a risk-minimization portfolio is given by

\[
\mathbf{w}_{\text{opt}} = \arg \min_{\mathbf{w}} \mathbf{w}^\top \Sigma \mathbf{w} = \frac{\Sigma^{-1} \mathbf{1}^\top \mathbf{1}}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}},
\]

where \( \mathbf{w}_{\text{opt}} \) determines the optimal weights on the shares. Such a risk-minimization portfolio strategy is justified in both Chamberlain (1983) and Owen and Rabinovich (2012) under elliptical distributions, which show that various features of portfolio analysis including mutual fund separation theorems and the capital asset pricing model hold for all elliptical distributions. Since the scatter matrix \( \Sigma \) is unknown in practice, we use the factor model with PCA or RTS method to estimate it. In detail,

### Table 2. Simulation results for Scenario B, the values in the parentheses are the interquartile ranges for MEE-CC and standard deviations for AVE-FL and AVE-FS.

| Type          | Method | \((p, n) = (150, 100)\) | \((p, n) = (250, 100)\) |
|---------------|--------|-------------------------|-------------------------|
| \(N(0, I_p + \mathbf{m})\) | RTS    | 0.02(0.01)              | 0.11(0.02)              |
|               | PCA    | 0.01(0.00)              | 0.09(0.01)              |
| \(t_3(0, I_p + \mathbf{m})\) | RTS    | 0.02(0.01)              | 0.13(0.02)              |
|               | PCA    | 0.04(0.03)              | 0.12(0.07)              |
| \(t_2(0, I_p + \mathbf{m})\) | RTS    | 0.03(0.02)              | 0.15(0.03)              |
|               | PCA    | 0.09(0.11)              | 0.31(0.13)              |
| \(t_1(0, I_p + \mathbf{m})\) | RTS    | 0.08(0.16)              | 0.29(0.14)              |
|               | PCA    | 0.28(0.30)              | 0.55(0.13)              |
| Skewed \(t_3\) | RTS    | 0.02(0.01)              | 0.12(0.02)              |
|               | PCA    | 0.05(0.04)              | 0.21(0.07)              |
| \(\alpha\)-stable | RTS    | 0.14(0.14)              | 0.30(0.10)              |
|               | PCA    | 0.35(0.78)              | 0.46(0.18)              |

### Table 3. Simulation results for Scenario C, the values in the parentheses are the interquartile ranges for MEE-CC and standard deviations for AVE-FL and AVE-FS.

| SNR | Type          | Method | \((p, n) = (250, 150)\) | \((p, n) = (250, 200)\) |
|-----|---------------|--------|-------------------------|-------------------------|
| 0.4 | \(N(0, I_p + \mathbf{m})\) | RTS    | 0.02(0.01)              | 0.14(0.02)              |
|     |               | PCA    | 0.02(0.01)              | 0.13(0.02)              |
| 0.5 | \(N(0, I_p + \mathbf{m})\) | RTS    | 0.03(0.01)              | 0.16(0.03)              |
|     |               | PCA    | 0.05(0.06)              | 0.26(0.10)              |
| 0.6 | \(N(0, I_p + \mathbf{m})\) | RTS    | 0.03(0.01)              | 0.15(0.02)              |
|     |               | PCA    | 0.05(0.04)              | 0.23(0.09)              |
| 0.7 | \(N(0, I_p + \mathbf{m})\) | RTS    | 0.02(0.00)              | 0.12(0.01)              |
|     |               | PCA    | 0.02(0.00)              | 0.12(0.01)              |

First, we design a rolling scheme to evaluate the PCA and RTS methods based on their performances in the annual return by constructing risk-minimization portfolios. Given the scatter matrix \( \Sigma \) of the share returns, a risk-minimization portfolio is given by

\[
\mathbf{w}_{\text{opt}} = \arg \min_{\mathbf{w}} \mathbf{w}^\top \Sigma \mathbf{w} = \frac{\Sigma^{-1} \mathbf{1}^\top \mathbf{1}}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}},
\]
at the beginning of each week $t$ in the year 2019, we recursively use the returns during the past 52 weeks ($52 \times 100$ panel) to train factor models either by PCA or RTS method. The estimated common components and idiosyncratic errors are recorded as $\hat{\chi}_t$ and $\hat{E}_t$, both of dimension $52 \times 100$. Then, we empirically estimate the scatter matrix of the 100 variables at week $t$ by

$$
\hat{\Sigma}_t = \frac{1}{52} \hat{\chi}_t^\top \hat{\chi}_t + \text{HardThresh}\left(\frac{1}{52} \hat{E}_t^\top \hat{E}_t\right),
$$

where $\text{HardThresh}(\hat{E}_t^\top \hat{E}_t/52)$ denotes the hard-threshold estimator proposed by Bickel and Levina (2008). The portfolio weights $\hat{w}_t$ are calculated using $\hat{\Sigma}_t$ and the return of the portfolio at week $t$ is $\hat{w}_t^\top x_t$, where $x_t$ is the raw return vector at week $t$. Figure 2 shows the net value curves of this strategy during the year 2019 by ignoring transaction cost and liquidity risk. It can be seen that when the factor models are trained by RTS method, the annual return of this portfolio is higher than that by PCA, regardless of taking $m = 1$ or $m = 3$.

We then further compare the RTS and PCA methods by their sensitivity to outliers in this real example. The sensitivity is evaluated by the variation of estimated loading space $D(\hat{L}^\text{new}, \hat{L}^\text{old})$, if we randomly select a small proportion of the demeaned log returns in the $105 \times 100$ panel and double their values. We repeat the procedure for 100 times to reduce randomness, and report the mean variation in Figure 3 with various contamination levels. It is seen that the estimated loading space can vary a lot with just a small number of outliers. When the contamination level grows, the discrepancy between the loading spaces $\hat{L}^\text{new}, \hat{L}^\text{old}$ becomes larger, and the phenomenon is more obvious in the case $m = 3$ compared with $m = 1$. However, the RTS method is less sensitive to outliers than the PCA method in both cases, which is expected as the RTS is more robust.

6. Discussion

We proposed a robust two-step estimation procedure for large-dimensional elliptical factor model. In the first step, we estimate the factor loadings by the leading eigenvectors of the spatial Kendall’s tau matrix. In the second step, we resort to ordinary least squares regression to estimate the factor scores. We prove the consistency of the proposed estimators for factor loadings, scores, and common components. Numerical studies show that
the proposed procedure works comparably with the conventional PCA method when data are from Gaussian distribution while performs much better when data are heavy-tailed, which indicates that the proposed RTS procedure can be used as a safe replacement of the conventional PCA method. In the future, we aim to propose a robust procedure for more general heavy-tailed data without the constraint of elliptical distribution. In fact, the elliptical assumption exerts a shape constraint on the distribution of the factors and the idiosyncratic errors, which may also constrain the real application.

Appendix A: Proofs of Main Theorems

We first present three useful lemmas before we give the detailed proofs of main theorems. In the proofs, $c$ denotes some generic finite constant. We denote a random matrix of fixed dimensions as $o_p(1)$ or $O_p(1)$ when all of its entries are $o_p(1)$ or $O_p(1)$.

Lemma A.1. Assume that $\mathbf{g} = (g_1, \ldots, g_p)^T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, then for any $i, j$, we have

$$E \frac{\|g_i g_j\|}{\|\mathbf{g}\|^2} = 0.$$

Proof. Without loss of generality, we take $i = 1, j = 2$ for example. Define

$$\tilde{g}_1 = \frac{1}{\sqrt{2}}(g_1 + g_2), \quad \tilde{g}_2 = \frac{1}{\sqrt{2}}(g_1 - g_2),$$

then it is easy to verify that

$$(\tilde{g}_1, \tilde{g}_2, \ldots, \tilde{g}_p) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p),$$

while by symmetry property,

$$2E \frac{\tilde{g}_1 \tilde{g}_2}{\|\mathbf{g}\|^2} = E \frac{\tilde{g}_1^2 - \tilde{g}_2^2}{\tilde{g}_1^2 + \tilde{g}_2^2 + \sum_{i=3}^p \delta_i^2} = 0.$$

Lemma A.2. Assume that $\mathbf{g} = (g_1, \ldots, g_p)^T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, then for any $q \times p$ deterministic matrix $\mathbf{A}$,

$$E \frac{\|\mathbf{A} \mathbf{g}\|^2}{\|\mathbf{g}\|^2} = \frac{1}{p} \|\mathbf{A}\|_F^2.$$

Proof. It is sufficient to prove that the lemma holds with $q = 1$. Given a $p$-dimensional deterministic vector $\mathbf{a}$, we have

$$E \frac{(\mathbf{a}^T \mathbf{g})^2}{\|\mathbf{g}\|^2} = \sum_{i=1}^p E \frac{(a_i g_i)^2}{\|\mathbf{g}\|^2} + \sum_{i \neq j} E \frac{a_i g_i \times a_j g_j}{\|\mathbf{g}\|^2}.$$

By Lemma A.1, the second term is 0. For any $i = 1, \ldots, p$ we have

$$E \frac{(a_i g_i)^2}{\|\mathbf{g}\|^2} = a_i^2 E \frac{g_i^2}{\|\mathbf{g}\|^2} = \frac{a_i^2}{p},$$

which concludes the lemma.

Lemma A.3. Under Assumptions A–C, as $\min[n, p] \to \infty$ we have

$$\{\lambda_j(\hat{\mathbf{K}}_x) \asymp m^{-1}, \quad j \leq m, \}$$

$$\{\lambda_j(\hat{\mathbf{K}}_x) = o_p(1), \quad j > m. \}$$

Proof. It is adapted from Lemma 3.1 and Lemma A.1 in Yu, He, and Zhang (2019), so we omit the proof here.

Proof of Theorem 3.1. Define $\hat{\mathbf{A}}$ as the diagonal matrix composed of the leading $m$ eigenvalues of $\hat{\mathbf{K}}_x$. Lemma A.3 implies that $\hat{\mathbf{A}}$ is asymptotically invertible, $\|\hat{\mathbf{A}}\|_F = O_p(1)$ and $\|\hat{\mathbf{A}}^{-1}\|_F = O_p(1)$. Because $\hat{\mathbf{L}} = \sqrt{p} \hat{\mathbf{T}}$ and $\hat{\mathbf{T}}$ is composed of the leading eigenvectors of $\hat{\mathbf{K}}_x$, we have

$$\hat{\mathbf{K}}_x \hat{\mathbf{L}} = \hat{\mathbf{L}} \hat{\mathbf{A}}.$$

Expand $\hat{\mathbf{K}}_x$ by its definition, then

$$\hat{\mathbf{K}}_x \hat{\mathbf{L}} = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} (y_i - y_j)(y_i - y_j)^T \hat{\mathbf{L}}$$

$$= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left[ \hat{\mathbf{L}}(f_i - f_j) + (\epsilon_i - \epsilon_j) \right] (y_i - y_j)^2 \hat{\mathbf{L}}$$

$$= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left[ \hat{\mathbf{L}}(f_i - f_j)^T \hat{\mathbf{L}}^{-1} \hat{\mathbf{L}}(f_i - f_j) \right] (y_i - y_j)^2 \hat{\mathbf{L}}$$

$$+ \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left[ (\epsilon_i - \epsilon_j) \right] (y_i - y_j)^2 \hat{\mathbf{L}}$$

For the ease of notations, we denote

$$\mathbf{M}_1 = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left[ (f_i - f_j)^T \right] (y_i - y_j)^2$$

$$\mathbf{M}_2 = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left[ (\epsilon_i - \epsilon_j) \right] (y_i - y_j)^2$$

$$\mathbf{M}_3 = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left[ (f_i - f_j)^T \right] (y_i - y_j)^2$$

$$\mathbf{M}_4 = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left[ (\epsilon_i - \epsilon_j) \right] (y_i - y_j)^2$$

and let $\hat{\mathbf{H}} = \mathbf{M}_1 \hat{\mathbf{L}}^T \hat{\mathbf{A}}^{-1}$, then

$$\hat{\mathbf{L}} - \hat{\mathbf{L}} \mathbf{H} = (\mathbf{M}_2 \hat{\mathbf{L}}^T \hat{\mathbf{L}} + \mathbf{M}_3 \mathbf{L}^T \mathbf{L} + \mathbf{M}_4 \mathbf{L}^T \mathbf{L}^{-1}).$$

(A.1)

Lemmas S2–S4 in the supplementary materials show that

$$\|\mathbf{M}_2\|_F^2 = O_p \left( \frac{1}{np} + \frac{1}{p^3} \right), \quad \|\mathbf{M}_3\|_F^2 = O_p \left( \frac{1}{np} + \frac{1}{p^3} \right),$$

while

$$\frac{1}{p} \|\mathbf{M}_4\|_F^2 = O_p \left( \frac{1}{np} + \frac{1}{p^3} \right) + o_p(1) \times \frac{1}{p} \|\hat{\mathbf{L}} - \hat{\mathbf{L}} \mathbf{H}\|_F^2.$$

Therefore, by Cauchy–Schwartz inequality and triangular inequality, it is easy to prove that

$$\frac{1}{p} \|\hat{\mathbf{L}} - \hat{\mathbf{L}} \mathbf{H}\|_F^2 = O_p \left( \frac{1}{np} + \frac{1}{p^3} \right) + o_p(1) \times \frac{1}{p} \|\hat{\mathbf{L}} - \hat{\mathbf{L}} \mathbf{H}\|_F^2.$$
and the convergence rate for $\hat{L}$ follows directly. To complete the proof, it remains to show $\hat{H}^\top V\hat{H} \overset{p}{\to} I_m$. By Cauchy–Schwartz inequality,

\[
\begin{align*}
\left\| \frac{1}{p} (L - \hat{L}) \right\|_F &\leq \frac{\left\| (L - \hat{L})_1 \right\|_F}{\sqrt{p}} \left\| (L - \hat{L})_1 \right\|_F = o_p(1), \\
\left\| \frac{1}{p} (L - \hat{L}) \right\|_F &\leq o_p(1).
\end{align*}
\]

Note that $p^{-1} \hat{L}^\top \hat{L} = I_m$ and $p^{-1} L^\top L \to V$, thus it holds that

\[
\frac{1}{p} L^\top L = V\hat{H} + o_p(1), \quad I_m = \frac{1}{p} \hat{L}^\top \hat{L} + o_p(1),
\]

which further implies $\hat{H}^\top V\hat{H} \overset{p}{\to} I_m$, and concludes the theorem.

\[\square\]

**Proof of Theorem 3.2.** Because $\hat{H}^\top V\hat{H} \overset{p}{\to} I_m$, we have $\|\hat{H}\|_F = O_p(1)$ and $H$ is invertible with probability approaching to 1. By our robust estimation procedure,

\[
\begin{align*}
\hat{f}_t &= \frac{1}{p} \hat{L}^\top \left( L - \hat{L} \right) \hat{y}_t = \frac{1}{p} L^\top (L f_t + \epsilon_t) \\
&= \frac{1}{p} L^\top \left( \hat{L}^\top \hat{L}^{-1} - (L\hat{L}^{-1} - L) \right) f_t + \frac{1}{p} L^\top \epsilon_t.
\end{align*}
\]

Note that $p^{-1} L^\top L = I_m$, then

\[
\hat{f}_t - \hat{f}_t - f_t = \frac{1}{p} L^\top (L - \hat{L}) \hat{y}_t - f_t + \frac{1}{p} \hat{L} - L \hat{L}^{-1} \epsilon_t + \frac{1}{p} \hat{L}^\top L \epsilon_t.
\]

Lemmas 5 and 6 in our supplementary materials show that

\[
\begin{align*}
\left\| \frac{1}{p} L^\top (L - \hat{L}) \hat{y}_t \right\|_F^2 &= O_p \left( \frac{1}{n^2} + \frac{1}{p^2} \right), \\
\left\| \frac{1}{p} \hat{L} - L \hat{L}^{-1} \epsilon_t \right\|_F^2 &= O_p \left( \frac{n^2}{1 + p^2} \right).
\end{align*}
\]

Meanwhile, for any $t \leq n$, by Assumption A and Lemma A.2 we have

\[
\|f_t\|^2 = \left( \frac{\zeta_t}{p} \right)^2 \frac{1}{\sqrt{p}} \|P_m \|g_t\|_F^2 = O_p(1).
\]

Similarly, it is not hard to prove that for any $t \leq n$,

\[
\left\| \frac{1}{p} L^\top \epsilon_t \right\|_F^2 = \frac{1}{p^2} \left( \frac{\zeta_t}{\sqrt{p}} \right)^2 \|P^\top A_t g_t\|_F^2 = O_p \left( \frac{1}{p^2} \right).
\]

Hence, by Cauchy–Schwartz inequality and triangular inequality, we have for any $t \leq n$

\[
\|\hat{H} f_t - f_t\|^2 = O_p(n^{-2} + p^{-1}),
\]

which concludes the theorem.

\[\square\]

**Proof of Theorem 3.3.** By Theorems 3.1 and 3.2, we already have

\[
\frac{1}{p} \left\| L - \hat{L} \right\|_F^2 = O_p \left( \frac{1 + \frac{1}{n}}{p} \right), \quad \text{and} \quad \|\hat{H} f_t - f_t\|^2 = O_p \left( \frac{1}{p} + \frac{1}{n^2} \right)
\]

for any $t \leq n$.

Hence, by triangular inequality and Cauchy–Schwartz inequality, we have for any $t \leq n$

\[
\frac{1}{p} \|\hat{L} f_t - f_t\|^2 = \frac{1}{p} \|\hat{L}^\top \hat{L} f_t - \hat{L}^\top f_t + \hat{L}^\top f_t - \hat{L} f_t\|^2 \\
\leq \left( \frac{2}{p} \|\hat{L}^\top \hat{L}^{-1} \right\| \left\| \hat{H} f_t - f_t \right\|^2 \\
+ \left( \frac{2}{p} \|\hat{L}^\top \hat{L}^{-1} - L \|_F^2 \right) \|f_t\|^2
\]

\[
= O_p \left( \frac{n^2}{p} + \frac{1}{p^2} \right),
\]

which concludes the theorem.

\[\square\]

**Supplementary Materials**

The supplementary material contains some useful lemmas which are of their own interest and the detailed proof of the main theorems.

**Acknowledgments**

We thank the AE and two anonymous reviewers for their helpful comments which improve the article a lot.

**Funding**

The authors gratefully acknowledge National Science Foundation of China (grant nos. 11801316, 71971118, 11971116), Natural Science Foundation of Shandong Province (grant no. ZR2019QA002), National Statistical Scientific Research Project (grant no. 2018LY63), Qinglan Project and NSF (18KJB520025) of Jiangsu High Institution.

**References**

Ahn, S. C., and Hereonstein, A. R. (2013), "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227. [302,305,306,307]

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171. [302,303,304,305]

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [302,305]

Bickel, P. J., and Levina, E. (2008), "Covariance Regularization by Thresholding," *The Annals of Statistics*, 36, 2577–2604. [309]

Chamberlain, G. (1983), "A Characterization of the Distributions That Imply Mean–Variance Utility Functions," *Journal of Economic Theory*, 29, 975–988. [308]

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure, and Mean–Variance Analysis of Large Asset Markets," *Econometrica*, 51, 1281–1304. [304]

Chen, L., Dolado, J. I., and Gonzalo, J. (2019), "Quantile Factor Models," arXiv no. 1911.02173. [302,303]

Choi, K., and Marden, J. (1998), "A Multivariate Version of Kendall’s τ," *Journal of Nonparametric Statistics*, 9, 261–293. [303,304]

Cont, R. (2001), "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues," *Quantitative Finance*, 1, 223–236. [302]

Croux, C., Ollila, E., and Oja, H. (2002), "Sign and Rank Covariance Matrices: Statistical Properties and Application to Principal Components Analysis," in *Statistical Data Analysis Based on the L1-Norm and Related Methods*, ed. Y. Dodge, Basel: Birkhäuser, pp. 257–269. [303]

Fama, E. F. (1963), "Mandelbrot and the Stable Paretian Hypothesis," *Journal of Business*, 36, 420–429. [302]

Fan, J., Liao, Y., and Mincheva, M. (2013), "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society, Series B*, 75, 603–680. [302,305]

Fan, J., Liu, H., and Wang, W. (2018), "Large Covariance Estimation Through Elliptical Factor Models," *The Annals of Statistics*, 46, 1383–1414. [303,304,305]

Fan, H., and Liu, H. (2012), "Semiparametric Principal Component Analysis," in *Advances in Neural Information Processing Systems*, pp. 171–179. [303]

(2014), "Scale-Invariant Sparse PCA on High-Dimensional Meta-Elliptical Data," *Journal of the American Statistical Association*, 109, 275–287. [303,304]
Y. HE ET AL. (2018), “ECA: High-Dimensional Elliptical Component Analysis in Non-Gaussian Distributions,” *Journal of the American Statistical Association*, 113, 252–268. [303,304,305]

Jing, B.-Y., Kong, X.-B., and Liu, Z. (2012), “Modeling High-Frequency Financial Data by Pure Jump Processes,” *The Annals of Statistics*, 40, 759–784. [302]

Kong, X.-B., Liu, Z., and Jing, B.-Y. (2015), “Testing for Pure-Jump Processes for High-Frequency Data,” *The Annals of Statistics*, 43, 847–877. [302]

Marden, J. I. (1999), “Some Robust Estimates of Principal Components,” *Statistics & Probability Letters*, 43, 349–359. [303,304]

Onatski, A. (2009), “Testing Hypotheses About the Number of Factors in Large Factor Models,” *Econometrica*, 77, 1447–1479. [302]

Owen, J., and Rabinovitch, R. (2012), “On the Class of Elliptical Distributions and Their Applications to the Theory of Portfolio Choice,” *Journal of Finance*, 38, 745–752. [308]

Stock, J. H., and Watson, M. W. (2002a), “Forecasting Using Principal Components From a Large Number of Predictors,” *Journal of the American Statistical Association*, 97, 1167–1179. [302]

——— (2002b), “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business & Economic Statistics*, 20, 147–162. [302]

Trapani, L. (2018), “A Randomised Sequential Procedure to Determine the Number of Factors,” *Journal of the American Statistical Association*, 113, 1341–1349. [302,305]

Visuri, S., Koivunen, V., and Oja, H. (2000), “Sign and Rank Covariance Matrices,” *Journal of Statistical Planning & Inference*, 91, 557–575. [303,304]

Xia, Q., Liang, R., and Wu, J. (2017), “Transformed Contribution Ratio Test for the Number of Factors in Static Approximate Factor Models,” *Computational Statistics & Data Analysis*, 112, 235–241. [306]

Yu, L., He, Y., and Zhang, X. (2019), “Robust Factor Number Specification for Large-Dimensional Elliptical Factor Model,” *Journal of Multivariate Analysis*, 174, 104543. [303,304,305,306,307,310]