


Article

Simultaneous Inference for High-Dimensional Approximate Factor Model

Yong Wang ¹ and Xiao Guo ^{2,*} 

¹ Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China; wy2608@mail.ustc.edu.cn

² International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230026, China

* Correspondence: xiaoguo@ustc.edu.cn

Received: 30 September 2020; Accepted: 2 November 2020; Published: 5 November 2020



Abstract: This paper studies simultaneous inference for factor loadings in the approximate factor model. We propose a test statistic based on the maximum discrepancy measure. Taking advantage of the fact that the test statistic can be approximated by the sum of the independent random variables, we develop a multiplier bootstrap procedure to calculate the critical value, and demonstrate the asymptotic size and power of the test. Finally, we apply our result to multiple testing problems by controlling the family-wise error rate (FWER). The conclusions are confirmed by simulations and real data analysis.

Keywords: high-dimensional factor model; multiple testing; multiplier bootstrap; simultaneous inference

1. Introduction

The high-dimensional factor model is becoming more and more important in different scientific areas including finance and macroeconomics. For example, the data in the World Bank contain two-hundred countries over forty years and the number of stocks can be in the thousands which is larger than or of the same order of the sample size for portfolio allocation. Due to its broad applications, much efforts have been devoted to analyzing factor model in different aspects. Examples include estimation of factors and loadings for latent factor model [1,2], covariance matrix estimation [3–6], and simultaneous inference for factor loadings of dynamic factor model [7,8], among others.

This work focuses on the simultaneous inference for the loading matrix with observed factors, which is an important issue in the analysis of approximate factor models. For example, in the study of gene expression genomics, it is commonly assumed that each gene is associated with only a few factors. For example, the authors of [9] showed that several oncogenes are related to Rb/E2F pathway rather than any other pathways. The authors of [10] also considered sparse loading matrix for gene expression data. Therefore, it is necessary to test the sparsity of the factor loadings. In the literature, some inference procedures have been proposed for latent factor models. For example, in low-dimensional setting, the authors of [11] considered testing the homogeneity assumption, i.e., the loadings associated to a factor are identical for all variables. The same testing problem has been considered by the authors of [12] in high-dimensional situation. As for observed factors, to the best of our knowledge, very limited work has been conducted.

Inference for the factor loadings with observed factors is not trivial. The approaches for latent factors cannot be directly applied to observed factors. The major difficulty is due to the high dimensionality,

which poses significant challenges in deriving the asymptotic null limiting distribution of the test statistic. We propose a test statistic based on the maximum discrepancy measure. The distribution of this statistic is attractive in high-dimensional statistical inference such as model selection, simultaneous inference, and multiple testing. Examples include the works in [13–17], among others.

We use the multiplier bootstrap procedure to obtain the critical value of our test statistic. Based on the fact that the test statistic can be approximated by the sum of the independent random variables, we show that the proposed multiplier bootstrap method consistently approximates the null limiting distribution of the test statistic, and thus the testing procedure achieves the prespecified significance level asymptotically. There are some related works applying multiplier bootstrap method to high-dimensional inference; see in [16,18,19], among others. However, their procedures require sparsity assumption on the parameters and cannot be directly applied to factor model. Compared with the works with latent factors, we do not require homogeneity constraints or sparsity on the model and our procedure is adaptive to high-dimensional regime.

Another application of our procedure is the multiple testing problem. Combining the multiplier bootstrap method with step-down procedure proposed by [17], we show that our procedure has a strong control of the family-wise error rate (FWER). Our method is asymptotically non-conservative as compared to the Bonferroni–Holm procedure since the correlation among the test statistics has been taken into account. We also want to point out that any procedure controlling the FWER will also control the false discovery rate [20] when there exist some true discoveries.

The rest of the paper is organized as follows. In Section 2.1, we develop the multiplier bootstrap procedure for simultaneous test of parameters for a single factor and demonstrate its asymptotic level and power. In Section 2.2, we give the result of simultaneous test of parameters for multiple factors. Section 3 discusses the multiple testing problem by combining the multiplier bootstrap procedure with the step-down method proposed by [17]. Section 4 investigates the numerical performance of the proposed test by simulations. We also conduct real data analysis on portfolio risk of S&P stocks via Fama–French model in Section 5. The proofs of the main results are given in Appendix A.

Finally, we introduce some notation. For set S , let $|S|$ denote the cardinality of S . Let $\mathbf{0}_p \in \mathbb{R}^p$ be the vector of zeros. For $p \times p$ matrix $\mathbf{A} = (a_{ij})_{i,j=1}^p$, denote by $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ the minimum and maximum eigenvalues of \mathbf{A} , respectively. The matrix element-wise maximum norm and L_2 norm are defined as $\|\mathbf{A}\|_{\infty} = \max_{1 \leq i,j \leq p} |a_{ij}|$ and $\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, respectively. For $a = (a_1, \dots, a_p)' \in \mathbb{R}^p$ and $q > 0$, denote by $\|a\|_q = (\sum_{i=1}^p |a_i|^q)^{1/q}$ and $\|a\|_{\infty} = \max_{1 \leq j \leq p} |a_j|$. Let $\mathbf{v}_i \in \mathbb{R}^K$ be the i th column of the $K \times K$ identity matrix. We write $a_t \lesssim b_t$ if a_t is smaller than or equal to b_t up to a universal positive constant. For $a, b \in \mathbb{R}$, we write $a \vee b = \max\{a, b\}$. For two sets, A and B , $A \ominus B$ denotes their symmetric difference, that is, $A \ominus B = (A \setminus B) \cup (B \setminus A)$.

2. Methodology

2.1. Simultaneous Test for a Single Factor

We consider the factor model defined as follows,

$$y_{it} = \mathbf{b}_i' \mathbf{f}_t + u_{it}, \quad i = 1, \dots, p \text{ and } t = 1, \dots, T, \quad (1)$$

where y_{it} is the observed response for the i th variable at time t , $\mathbf{b}_i \in \mathbb{R}^K$ is the unknown vector of factor loadings, $\mathbf{f}_t \in \mathbb{R}^K$ is the observed vector of common factors, and u_{it} is the latent error. Here, K is a fixed integer denoting the number of factors, p is the number of variables, and T denotes the sample size. Model (1) is commonly used in finance and macroeconomics, see, e.g., in [3,4,21], among others.

Denote by $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_p)'$, $\mathbf{y}_t = (y_{1t}, \dots, y_{pt})'$ and $\mathbf{u}_t = (u_{1t}, \dots, u_{pt})'$, then model (1) can be re-expressed as

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t. \tag{2}$$

We first focus on testing the coefficients $b_{ik} = \mathbf{b}'_i \mathbf{v}_k$ corresponding to a single factor, i.e., the k th factor. Specifically, we consider the following simultaneous testing problem that given $k = 1, \dots, K$,

$$H_{0,G} : b_{ik} = b_{ik}^{\text{null}} \quad \text{for all } i \in G \quad \text{versus} \quad H_{1,G} : b_{ik} \neq b_{ik}^{\text{null}} \quad \text{for some } i \in G, \tag{3}$$

where G is a subset of $\{1, \dots, p\}$ and b_{ik}^{null} are prespecified values. For example, if b_{ik}^{null} are 0, then the hypotheses are able to test whether the variables with indices in G are significantly associated with the k th factor. Throughout the paper, $|G|$ is allowed to grow as fast as p , which may have exponential growth with T as in Assumption 3.

The ordinary least squares (OLS) estimator $\hat{\mathbf{B}} = \mathbf{Y}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}$ is applied to estimate \mathbf{B} , where $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$ and $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)'$. Therefore,

$$\hat{\mathbf{B}} - \mathbf{B} = \left(\sum_{t=1}^T \mathbf{u}_t \mathbf{f}'_t \right) \left(\sum_{t=1}^T \mathbf{f}_t \mathbf{f}'_t \right)^{-1}. \tag{4}$$

We propose the following test statistic for $H_{0,G}$,

$$M_{T,k} = \max_{i \in G} \sqrt{T} |\hat{b}_{ik} - b_{ik}^{\text{null}}|,$$

where $(\hat{b}_{ik})_{i \leq p, k \leq K} = \hat{\mathbf{B}}$. For each $i \in G$, the asymptotic normality of \hat{b}_{ik} is straightforward due to the central limit theorem. However, when $|G|$ diverges with p , it is very challenging to demonstrate the existence of the limiting distribution of $M_{T,k}$. In order to approximate the asymptotic distribution of $M_{T,k}$, we will use the multiplier bootstrap method. From (4), we know

$$\sqrt{T}(\hat{b}_{ik} - b_{ik}) = \frac{1}{\sqrt{T}} \sum_{t=1}^T u_{it} \mathbf{f}'_t \hat{\mathbf{\Omega}}_f \mathbf{v}_k = \frac{1}{\sqrt{T}} \sum_{t=1}^T \hat{\xi}_{it}, \tag{5}$$

where $\hat{\xi}_{it} = u_{it} \mathbf{f}'_t \hat{\mathbf{\Omega}}_f \mathbf{v}_k$ and $\hat{\mathbf{\Omega}}_f = (\sum_{t=1}^T \mathbf{f}_t \mathbf{f}'_t / T)^{-1}$.

In order to apply the multiplier bootstrap procedure, we need to approximate $\sum_{t=1}^T \hat{\xi}_{it} / \sqrt{T}$ by sum of independent random variables. As $\hat{\mathbf{\Omega}}_f$ is consistent for $\mathbf{\Omega}_f = \{E(\mathbf{f}_t \mathbf{f}'_t)\}^{-1}$, we can replace the former with the latter in $\hat{\xi}_{it}$, and define $\xi_{it} = u_{it} \mathbf{f}'_t \mathbf{\Omega}_f \mathbf{v}_k$. Then, for each $i \in G$, $\{\xi_{it}\}_{t \geq 1}$ are i.i.d. and $\sum_{t=1}^T \xi_{it} / \sqrt{T}$ well approximates $\sum_{t=1}^T \hat{\xi}_{it} / \sqrt{T}$.

We then apply the multiplier bootstrap procedure to approximate the distribution of $\max_{i \in G} |\sum_{t=1}^T \xi_{it} / \sqrt{T}|$. Denote by $\mathbf{\Sigma}_u = (\sigma_{ij})_{p \times p}$ the covariance matrix of \mathbf{u}_t , and hence $\text{cov}(\xi_{it}, \xi_{jt}) = \mathbf{\Omega}_f(k, k) \sigma_{ij}$, where $\mathbf{\Omega}_f(k, k) = \mathbf{v}'_k \mathbf{\Omega}_f \mathbf{v}_k$. We know that $\hat{\mathbf{\Omega}}_f(k, k) = \mathbf{v}'_k \hat{\mathbf{\Omega}}_f \mathbf{v}_k$ is \sqrt{T} -consistent for $\mathbf{\Omega}_f(k, k)$. To estimate σ_{ij} , we first calculate the residuals

$$\hat{u}_{it} = y_{it} - \hat{\mathbf{b}}'_i \mathbf{f}_t.$$

Denote by $\hat{\mathbf{u}}_t = (\hat{u}_{1t}, \dots, \hat{u}_{pt})'$, then the error covariance matrix is estimated by

$$\hat{\mathbf{\Sigma}}_u = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}'_t = (\hat{\sigma}_{ij})_{i \leq p, j \leq p}.$$

Let $\{e_t\}_{t=1}^T$, a sequence of i.i.d. $N(0, 1)$ independent of $\{\mathbf{y}_t, \mathbf{f}_t\}_{t=1}^T$, be the multiplier random variables. Then the multiplier bootstrap statistic is defined as

$$W_{T,k} = \max_{i \in G} T^{-1/2} \sqrt{\widehat{\Omega}_f(k, k)} \left| \sum_{t=1}^T \widehat{u}_{it} e_t \right|.$$

Conditioning on $\{\mathbf{y}_t, \mathbf{f}_t\}_{t=1}^T$, the covariance of $T^{-1/2} \sum_{t=1}^T \sqrt{\widehat{\Omega}_f(k, k)} \widehat{u}_{it} e_t$ and $T^{-1/2} \sum_{t=1}^T \sqrt{\widehat{\Omega}_f(k, k)} \widehat{u}_{jt} e_t$ is $\widehat{\Omega}_f(k, k) \widehat{\sigma}_{ij}$, which can sufficiently approximate the covariance between ζ_{it} and ζ_{jt} . Then, the bootstrap critical value can be obtained via

$$c_{W_{T,k}}(\alpha) = \inf\{t \in \mathbb{R} : P(W_{T,k} \leq t | (\mathbf{Y}, \mathbf{F})) \geq 1 - \alpha\}.$$

$c_{W_{T,k}}(\alpha)$ is calculated by generating $\{e_t\}_{t=1}^T$ repeatedly. In our simulations and real data, we conduct bootstrap 500 times. We now present some technical assumptions.

- Assumption 1.** (i) $\{\mathbf{f}_t, \mathbf{u}_t\}_{t \geq 1}$ are i.i.d. with $E(\mathbf{u}_t) = \mathbf{0}_p$ and $\Sigma_u = \text{cov}(\mathbf{u}_t)$.
 (ii) There exist constants c_1, c_2 such that $0 < c_1 < \lambda_{\min}(\Sigma_u) < \lambda_{\max}(\Sigma_u) < c_2 < \infty$.
 (iii) $\{\mathbf{u}_t\}_{t \geq 1}$ and $\{\mathbf{f}_t\}_{t \geq 1}$ are independent.

Assumption 2. There exist positive constants r_1, r_2, b_1, b_2 , such that for any $s > 0, t \leq T, i \leq p$ and $j \leq K$,

$$P(|u_{it}| > s) \leq \exp\{-(s/b_1)^{r_1}\}, \quad P(|f_{jt}| > s) \leq \exp\{-(s/b_2)^{r_2}\}.$$

The ‘‘i.i.d.’’ condition in Assumption 1 is commonly considered in the literature for high-dimensional inference, see, e.g., in [16]. Assumption 1 (ii) allows the bounded eigenvalue of the error covariance matrix. As noted in [22], such assumption is satisfied by two situations: (1) $\text{cov}(U_1, \dots, U_p)$, where $\{U_i, i \geq 1\}$ is a stationary ergodic process with spectral density $f, 0 < c_1 < f < c_2$ and (2) $\text{cov}(X_1, \dots, X_p)$ where $X_i = U_i + V_i, i = 1, \dots, p, \{U_i\}$ is a stationary process as above and $\{V_i\}$ is a noise process independent of $\{U_i\}$. In Example 1 in [22], they demonstrated that ARMA(r, q) process satisfies Assumption 1 (ii). Furthermore, this assumption is commonly considered in the literature, see, e.g., in [4,15].

Assumption 2 allows the application of the large deviation theory to $(1/T) \sum_{t=1}^T u_{it} u_{jt} - \sigma_{ij}$ and $(1/T) \sum_{t=1}^T u_{it} f_{jt}$. In this paper, we assume that \mathbf{f}_t and \mathbf{u}_t have exponential-type tails. Let $\gamma_1^{-1} = 3r_1^{-1}$ and $\gamma_2^{-1} = 1.5r_1^{-1} + 1.5r_2^{-1}$.

Assumption 3. Suppose $\gamma_1 < 1, \gamma_2 < 1$ and there exists a constant $c_1 > 0$, such that $(\log p)^\gamma = o(T)$, where $\gamma = \max\{2/\gamma_1 - 1, 2/\gamma_2 - 1, 7 + c_1\}$.

Assumption 4. There exists a constant $C > 0$ such that $\lambda_{\max}(\Omega_f) < C$.

Assumption 3 is needed in Bernstein-type inequality [23] and commonly assumed in the literature for Gaussian approximation theory. Assumption 4 is also reasonable by bounding the eigenvalues of Ω_f .

Theorem 1. Under Assumptions 1–4, we have

$$\sup_{\alpha \in (0,1)} \left| P\left(\max_{i \in G} \sqrt{T} |\widehat{b}_{ik} - b_{ik}| > c_{W_{T,k}}(\alpha) \right) - \alpha \right| = o(1).$$

Theorem 1 demonstrates that the multiplier bootstrap critical value $c_{W_{T,k}}(\alpha)$ well approximates the quantile of the test statistic. It is worth mentioning that our method does not require any sparsity assumption on either Σ_u or \mathbf{B} .

The proof of Theorem 1 depends on the two results: (1) $\max_{i \in G} \sum_{t=1}^T \zeta_{it} / \sqrt{T}$ is sufficiently close to $\max_{i \in G} \sum_{t=1}^T \widehat{\zeta}_{it} / \sqrt{T}$ and (2) the covariances of ζ_{it} and ζ_{jt} are well approximated by the bootstrap version. The first result is demonstrated in Lemma A7 that there exist $\zeta_1 > 0$ and $\zeta_2 > 0$ such that

$$P\left(\left|\max_{i \in G} \sum_{t=1}^T \widehat{\zeta}_{it} / \sqrt{T} - \max_{i \in G} \sum_{t=1}^T \zeta_{it} / \sqrt{T}\right| > \zeta_1\right) < \zeta_2,$$

where $\zeta_1 \sqrt{1 \vee \log(p/\zeta_1)} = o(1)$ and $\zeta_2 = o(1)$. The second result is shown in Lemma A6 that

$$\Delta = \max_{1 \leq i, j \leq p} |\widehat{\Omega}_f(k, k) \widehat{\sigma}_{ij} - \Omega_f(k, k) \sigma_{ij}| = o_P((\log p)^{-2}),$$

i.e., the maximum discrepancy between the empirical and population covariance matrices converges to zero.

Based on Theorem 1, for a given significance level $0 < \alpha < 1$, we define the test Φ_α by

$$\Phi_\alpha = I(M_{T,k} > c_{W_{T,k}}(\alpha)). \tag{6}$$

The hypothesis $H_{0,G}$ is rejected whenever $\Phi_\alpha = 1$.

Bootstrap is a commonly used resampling method and full theories about it can be found in [24]. There are many versions of bootstrap, for example, wild bootstrap, empirical bootstrap, and multiplier bootstrap, among others. As discussed in [25], other exchangeable bootstrap methods are asymptotically equivalent to the multiplier bootstrap. As our test statistic can be approximated by the maximum of sum of independent random vectors, we adopt the multiplier bootstrap method in [25] based on Gaussian approximation.

Alternatively, we propose the studentized statistic $M_{T,k}^* := \max_{i \in G} \sqrt{T} |\widehat{b}_{ik} - b_{ik}^{\text{null}}| / \sqrt{\widehat{\omega}_{ii}}$ for $H_{0,G}$, where $\widehat{\omega}_{ii} = \widehat{\Omega}_f(k, k) \widehat{\sigma}_{ii}$. Similarly to Section 2.1, we define the multiplier bootstrap statistic as

$$W_{T,k}^* = \max_{i \in G} T^{-1/2} \left| \sum_{t=1}^T \widehat{u}_{it} e_t \right| \sqrt{\widehat{\Omega}_f(k, k) / \widehat{\omega}_{ii}} = \max_{i \in G} T^{-1/2} \widehat{\sigma}_{ii}^{-1/2} \left| \sum_{t=1}^T \widehat{u}_{it} e_t \right|,$$

where $\{e_t\}_{t=1}^T \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ are independent of $\{\mathbf{y}_t, \mathbf{f}_t\}_{t=1}^T$. Then, the bootstrap critical value can be obtained via

$$c_{W_{T,k}}^*(\alpha) = \inf\{t \in \mathbb{R} : P(W_{T,k}^* \leq t | (\mathbf{Y}, \mathbf{F})) \geq 1 - \alpha\}.$$

Theorem 2 below justifies the validity of the bootstrap procedure for the studentized statistic.

Theorem 2. Under the assumptions in Theorem 1, we have

$$\sup_{\alpha \in (0,1)} \left| P\left(\max_{i \in G} \sqrt{T} |\widehat{b}_{ik} - b_{ik}| / \sqrt{\widehat{\omega}_{ii}} > c_{W_{T,k}}^*(\alpha)\right) - \alpha \right| = o(1).$$

Based on this result, for a given significance level $0 < \alpha < 1$, we define the test Φ_α^* by

$$\Phi_\alpha^* = I(M_{T,k}^* > c_{W_{T,k}}^*(\alpha)).$$

The hypothesis $H_{0,G}$ is rejected whenever $\Phi_\alpha^* = 1$.

For the studentized statistic, we can derive its asymptotic distribution. By Lemma 6 of the work in [15], for any $x \in \mathbb{R}$ and as $p \rightarrow \infty$, we have

$$P\left(\max_{1 \leq i \leq p} T|\widehat{b}_{ik} - b_{ik}|^2 / \widehat{\omega}_{ii} - 2 \log(p) + \log \log(p) \leq x\right) \rightarrow \exp\left\{-\frac{1}{\sqrt{\pi}} \exp\left(-\frac{x}{2}\right)\right\}.$$

However, the above alternative testing procedure may not perform well in practice, because it requires diverging p , and the convergence rate is typically slow.

In contrast to the extreme value approach, our testing procedure explicitly accounts for the effect of $|G|$ in the sense that the bootstrap critical value $c_{W_{T,k}}^*(\alpha)$ depends on G . Therefore, our approach is more robust to the change in $|G|$.

Next, we turn to the (asymptotic) power analysis of the above procedure. Denote by \mathbf{B}_k the k th column of \mathbf{B} . We focus on the case where $|G| \rightarrow \infty$ as $T \rightarrow \infty$ below. Define the separation set

$$\mathcal{U}_G(c) = \{(b_{1k}, \dots, b_{pk})^T : \max_{i \in G} |b_{ik} - b_{ik}^{\text{null}}| / \sqrt{\omega_{ii}} > c \sqrt{\log(|G|)/T}\}, \tag{7}$$

where $\omega_{ij} = \boldsymbol{\Omega}_f(k, k) \sigma_{ij}$. Let $\Theta = (\theta_{ij})_{i,j=1}^p$ with $\theta_{ij} = \omega_{ij} / \sqrt{\omega_{ii} \omega_{jj}} = \sigma_{ij} / \sqrt{\sigma_{ii} \sigma_{jj}}$, which is the correlation matrix of \mathbf{u}_t .

Assumption 5. Suppose $\max_{1 \leq i \neq j \leq p} |\theta_{ij}| \leq c_0 < 1$ for some constant c_0 .

Theorem 3. Under Assumptions 1–5, for any $\varepsilon_0 > 0$, we have

$$\inf_{\mathbf{B}_k \in \mathcal{U}_G(\sqrt{2} + \varepsilon_0)} P\left(\max_{i \in G} \sqrt{T} |\widehat{b}_{ik} - b_{ik}^{\text{null}}| / \sqrt{\widehat{\omega}_{ii}} > c_{W_{T,k}}^*(\alpha)\right) \rightarrow 1.$$

As long as one entry of $b_{ik} - b_{ik}^{\text{null}}$ has a magnitude larger than $(\sqrt{2} + \varepsilon_0) \sqrt{\log |G| / T}$, our bootstrap-assisted test can reject the null correctly. Therefore, with \mathbf{B} being sparse, our procedure performs well in detecting non-sparse alternatives. According to Section 3.2 of [26], the separation rate $(\sqrt{2} + \varepsilon_0) \sqrt{\log(|G|)/T}$ is minimax optimal under suitable assumptions.

2.2. Simultaneous Test for Multiple Factors

In this section, we test the elements of the loading matrix corresponding to different factors. The testing problem can be stated as follows,

$$H_{0,G^*} : b_{ik} = b_{ik}^{\text{null}} \quad \text{for all } (i, k) \in G^* \quad \text{versus} \quad H_{1,G^*} : b_{ik} \neq b_{ik}^{\text{null}} \quad \text{for some } (i, k) \in G^*,$$

where G^* is a subset of $M \equiv \{(i, j) : i = 1, \dots, p \text{ and } j = 1, \dots, K\}$. Define

$$\begin{aligned} \omega_{(i,k),(j,\ell)}^* &= \text{cov}(u_{it} \mathbf{f}_t' \boldsymbol{\Omega}_f \mathbf{v}_k, u_{jt} \mathbf{f}_t' \boldsymbol{\Omega}_f \mathbf{v}_\ell) = \sigma_{ij} \mathbf{v}_k' \boldsymbol{\Omega}_f \mathbf{v}_\ell, \\ \widehat{\omega}_{(i,k),(j,\ell)}^* &= (\widehat{\boldsymbol{\Omega}}_f \mathbf{v}_k)' \left(\frac{1}{T} \sum_{t=1}^T \widehat{u}_{it} \widehat{u}_{jt} \mathbf{f}_t \mathbf{f}_t'\right) (\widehat{\boldsymbol{\Omega}}_f \mathbf{v}_\ell). \end{aligned}$$

We propose the studentized test statistic

$$M_{T,G^*} = \max_{(i,k) \in G^*} \sqrt{T} |\widehat{b}_{ik} - b_{ik}| / \sqrt{\widehat{\omega}_{(i,k),(i,k)}^*}.$$

From the linear expansion in (5), the multiplier bootstrap statistic is defined as

$$W_{T,G^*} = \max_{(i,k) \in G^*} T^{-1/2} \left| \sum_{t=1}^T \hat{u}_{it} \mathbf{f}'_t \hat{\Omega}_f \mathbf{v}_k e_t \right| / \sqrt{\hat{\omega}_{(i,k),(i,k)}^*}$$

where $\{e_t\}_{t=1}^T \stackrel{i.i.d.}{\sim} N(0, 1)$ are independent of $\{\mathbf{y}_t, \mathbf{f}_t\}_{t=1}^T$. Then, the bootstrap critical value can be obtained via

$$c_{W_{T,G^*}}(\alpha) = \inf\{t \in \mathbb{R} : P(W_{T,G^*} \leq t | (\mathbf{Y}, \mathbf{F})) \geq 1 - \alpha\}.$$

Let $\gamma_3^{-1} = 4r_1^{-1} + 4r_2^{-1}$, $r_3^{-1} = 3r_1^{-1} + 9r_2^{-1}$ and $r = \max\{2/\gamma_3 - 1, 2/r_3 - 1, c_1 + 7\}$ for a constant $c_1 > 0$.

Theorem 4. Suppose $(\log p)^r = o(T)$, under Assumptions 1, 2, and 4, we have

$$\sup_{\alpha \in (0,1)} \left| P\left(\max_{(i,k) \in G^*} \sqrt{T} |\hat{b}_{ik} - b_{ik}| / \sqrt{\hat{\omega}_{(i,k),(i,k)}^*} > c_{W_{T,G^*}}(\alpha) \right) - \alpha \right| = o(1).$$

Based on Theorem 4, for a given significance level $0 < \alpha < 1$, we define the test $\Phi_\alpha(G^*)$ by

$$\Phi_\alpha(G^*) = I(M_{T,G^*} > c_{W_{T,G^*}}(\alpha)).$$

The hypothesis H_{0,G^*} is rejected whenever $\Phi_\alpha(G^*) = 1$.

Now we turn to the power analysis of the test $\Phi_\alpha(G^*)$. Similar to Section 2.1, we focus on the case where $|G^*| \rightarrow \infty$ as $T \rightarrow \infty$ and define the separation set

$$\mathcal{V}_{G^*}(c) = \{(b_{ik})_{i \leq p, k \leq K} : \max_{(i,k) \in G^*} |b_{ik} - b_{ik}^{\text{null}}| / \sqrt{\omega_{(i,k),(i,k)}^*} > c \sqrt{\log(|G^*|)/T}\},$$

Let $\theta_{(i,k),(j,\ell)}^* = \omega_{(i,k),(j,\ell)}^* / \sqrt{\omega_{(i,k),(i,k)}^* \omega_{(j,\ell),(j,\ell)}^*}$. We consider the following condition.

Assumption 6. Suppose $\max_{(i,k),(j,\ell)} |\theta_{(i,k),(j,\ell)}^*| \leq c_0^* < 1$ for some constant c_0^* .

The asymptotic power of the testing procedure is given as follows.

Theorem 5. Under the assumptions in Theorem 4 and Assumption 6, for any $\varepsilon_0 > 0$, we have

$$\inf_{\mathbf{B} \in \mathcal{V}_{G^*}(\sqrt{2} + \varepsilon_0)} P\left(\max_{(i,k) \in G^*} \sqrt{T} |\hat{b}_{ik} - b_{ik}^{\text{null}}| / \sqrt{\hat{\omega}_{(i,k),(i,k)}^*} > c_{W_{T,G^*}}(\alpha) \right) \rightarrow 1.$$

3. Multiple Testing with Strong FWER Control

In this section, we study the following multiple testing problem,

$$H_{0,i} : b_{ij} \leq b_{ij}^{\text{null}} \quad \text{versus} \quad H_{1,i} : b_{ij} > b_{ij}^{\text{null}} \quad \text{for all } i \in G.$$

For simplicity, we set $G = \{1, 2, \dots, p\}$ and let j be fixed. We combine the bootstrap-assisted procedure with the step-down method proposed by [17]. Our method can be seen as a special case in Section 5 of [25]. Note that this framework can cover the case of testing equalities ($H_{0,j} : b_{ij} = b_{ij}^{\text{null}}$) because equalities can be rewritten as pairs of inequalities.

We briefly illustrate the control of the FWER. Full details and theory can be found in [25]. Let Ω be the space for all data generating processes, and ω be the true process. Each null hypothesis $H_{0,i}$

is equivalent to $\omega \in \Omega_i$ for some $\Omega_i \subset \Omega$. For any $\eta \subset G$, denote by $\Omega^\eta = (\cap_{i \in \eta} \Omega_i) \cap (\cap_{i \notin \eta} \Omega_i^c)$ with $\Omega_i^c = \Omega \setminus \Omega_i$. The strong control of the FWER means that

$$\sup_{\eta \subset G} \sup_{\omega \in \Omega^\eta} P_\omega(\text{reject at least one hypothesis } H_{0,i}, i \in \omega) \leq \alpha + o(1), \tag{8}$$

where P_ω denotes the probability distribution under the data-generating process ω .

For $i = 1, \dots, p$, denote $t_{ij} = \sqrt{T}(\widehat{b}_{ij} - b_{ij}^{\text{null}})$. For a subset $\eta \subset G$, let $c_\eta(\alpha)$ be the bootstrapped estimate for the $(1 - \alpha)$ -quantile of $\max_{i \in \eta} t_{ij}$. The step-down procedure in [17] is described as follows. Define $\eta(1) = G$ at the first step and reject all $H_{0,i}$ satisfying $t_{ij} > c_{\eta(1)}(\alpha)$. If no $H_{0,i}$ is rejected, then stop the procedure. If some $H_{0,i}$ are rejected, let $\eta(2)$ be the set of indices for those hypotheses not being rejected at the first step. On step $\ell \geq 2$, let $\eta(\ell) \subset G$ be the subset of hypotheses that were not rejected at step $\ell - 1$. Reject all hypotheses $H_{0,i}$ for $i \in \eta(\ell)$ satisfying that $t_{ij} > c_{\eta(\ell)}(\alpha)$. If no hypothesis is rejected, then stop the procedure. Proceed in this way until the algorithm stops.

Romano and Wolf [17] proved the following result:

$$c_\eta(\alpha) \leq c_{\eta'}(\alpha), \quad \text{for } \eta \subset \eta' \tag{9}$$

$$\sup_{\eta \subset G} \sup_{\omega \in \Omega^\eta} P_\omega\left(\max_{i \in \eta} t_{ij} > c_\eta(\alpha)\right) \leq \alpha + o(1). \tag{10}$$

Therefore, we can show that the step-down method together with the multiplier bootstrap provide strong control of the FWER by verifying (9) and (10). The theoretical results are given in the proposition below. The proofs are similar to those of Theorem 1, which are omitted here.

Proposition 1. *Under the assumptions in Theorem 1, the step-down procedure with the bootstrap critical value $c_\eta(\alpha)$ satisfies (8).*

Our multiple testing method has the following two important features: (i) It can be applied to models with an increasing dimension; (ii) It takes into account the correlation amongst statistics and hence is asymptotically non-conservative.

In the simulation, we also consider Benjamini–Hochberg procedure [20] to control the false discovery rate (FDR), which is summarized as follows. For each of $H_{0,1}, \dots, H_{0,p}$, we calculate the p -values P_1, \dots, P_p based on the studentized test statistic. Let $P_{(1)} \leq \dots \leq P_{(p)}$ be the ordered p -values, and denote by $H_{0,(i)}$ the null hypothesis corresponding to $P_{(i)}$. Let $k = \max\{i : P_{(i)} \leq i\alpha/p\}$, and then reject all $H_{0,(i)}$ for $i = 1, \dots, k$.

4. Simulation Study

This section examines the performance of the proposed testing procedure by a simulation study. We fix the number of factors $K = 3$, the sample size $T \in \{200, 400\}$, and let the dimensionality p increase from 50 to 600. Throughout the simulation, we consider testing the first column of \mathbf{B} and repeat multiplier bootstrap 500 times.

Each row of \mathbf{B} is generated independently from $N(0, I_K)$, where I_K is $K \times K$ identity matrix. Let $\text{cov}(\mathbf{f}_t) = (\sigma_{ij}^f)_{K \times K}$ with $\sigma_{ij}^f = 0.6^{|i-j|}$. Here, we consider two models for the covariance structure Σ_u .

- (a) Model 1 (sparse): $\Omega_u = (\omega_{ij})_{1 \leq i, j \leq p}$ where $\omega_{ii} = 1$, $\omega_{ij} = 0.8$ for $2(k-1) + 1 \leq i \neq j \leq 2k$, where $k = 1, \dots, \lfloor p/2 \rfloor$ and $\omega_{ij} = 0$ otherwise. $\Sigma_u = \Omega_u^{-1}$.
- (b) Model 2 (non-sparse): $\Sigma_u = (\sigma_{ij})_{1 \leq i, j \leq p}$ where $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.5$ for $i \neq j$.

Under each model, $\{\mathbf{f}_t\}_{t=1}^T$ and $\{\mathbf{u}_t\}_{t=1}^T$ are generated independently from $N(0, \text{cov}(\mathbf{f}_t))$ and $N(0, \Sigma_u)$, respectively.

We calculate the empirical sizes of test for each column of \mathbf{B} under each model by considering hypothesis (3) with $G = \{1, 2, \dots, p\}$ and b_{ik}^{null} being the true value of b_{ik} . The results are summarized in Table 1. Here “NST”, “ST” denote the non-studentized, studentized Bootstrap-based test, respectively, and “EX” denotes the test using extreme value distribution. The estimated sizes of the three tests are reasonably close to the nominal level 0.05 for the values of p ranging from 50 to 600.

Table 1. Empirical sizes of tests, $\alpha = 0.05$, $T = 400$, and 500 replications.

	$p = 50$	$p = 100$	$p = 200$	$p = 400$	$p = 600$
Model 1					
NST	0.076	0.060	0.058	0.050	0.058
ST	0.074	0.064	0.064	0.058	0.078
EX	0.046	0.046	0.038	0.046	0.058
Model 2					
NST	0.050	0.052	0.056	0.060	0.038
ST	0.070	0.058	0.064	0.068	0.048
EX	0.038	0.030	0.024	0.018	0.016

For all $i \in G$, by varying $b_{ik} = b_{ik}^{\text{null}} + c/40$ with $c = \pm 0.8\ell$ and $\ell = 0, \dots, 10$, we plot the empirical powers of $M_{T,k}$ and $M_{T,k}^*$ in Figure 1. For ease of presentation, we only consider $p \in \{10, 200, 600\}$. The results for other dimensionality are similar in spirit, and are not presented here. For all tests, the significance level is fixed at $\alpha = 0.05$. From Figure 1, we can tell that the empirical rejection rate grows from the nominal level to one as c deviates away from zero. The difference between NST test and ST test is slight. For small p , the EX test does not perform well because this approach requires diverging p . Furthermore, for non-sparse error covariance matrix, our method performs better than the EX method. These numerical results confirm our theoretical analysis.

Next, we study the numerical performance of the step-down method in Section 3 and compare it with the Bonferroni–Holm procedure. Consider the following two-sided multiple testing problem; $H_{0,i} : b_{ij} = \tilde{b}_{ij}^{\text{null}}$ among all $i = 1, 2, \dots, p$ with $j = 1$. For Models 1 and 2, the first s_0 entries of $\{\tilde{b}_{ij}^{\text{null}}\}_{i=1}^p$ are $b_{ij}^{\text{null}} + 0.5$ and $b_{ij}^{\text{null}} + 0.35$, respectively, and the rest are equal to b_{ij}^{null} . We set $T \in \{200, 400\}$ and $p \in \{50, 200, 500, 600\}$.

We employ both the step-down method based on the studentized/non-studentized test statistic, and the Bonferroni–Holm procedure (based on the studentized test statistic) to control the FWER. We denote these three procedures by NST-FWER, ST-FWER, and BH-FWER, respectively. For comparison, we also consider using Benjamini–Hochberg procedure to control FDR. We denote this procedure by BH-FDR. Based on 500 replications, we calculate the average empirical FWER

$$\text{Average} \{ I\{ \text{At least one hypothesis } H_{0,i} \text{ is rejected}, i \in \{s_0 + 1, \dots, p\} \} \}$$

for methods NST-FWER, ST-FWER, and BH-FWER, the average empirical FDR

$$\text{Average} \left\{ \frac{\sum_{i \in S_0} I\{H_{0,i} \text{ is rejected}\}}{\sum_{i \in G} I\{H_{0,i} \text{ is rejected}\}} \right\}$$

for method BH-FDR, and the average empirical power

$$\text{Average} \left\{ \frac{\sum_{i \in S_0} I\{H_{0,i} \text{ is rejected}\}}{s_0} \right\}$$

for all the four methods, where $S_0 = \{1, 2, \dots, s_0\}$ and $G = \{1, \dots, p\}$. Under each model, we consider the case $s_0 = 3$ and $s_0 = 15$. Tables 2 and 3 report the empirical FWER, FDR, and the average power. From Tables 2 and 3, the proposed and Bonferroni–Holm procedures provide similar control on the FWER, and Benjamini–Hochberg procedure can control FDR. The empirical powers of the step-down method and Benjamini–Hochberg procedure are higher than that of the Bonferroni–Holm procedure. It is also seen that controlling the FDR is more powerful than controlling the FWER.

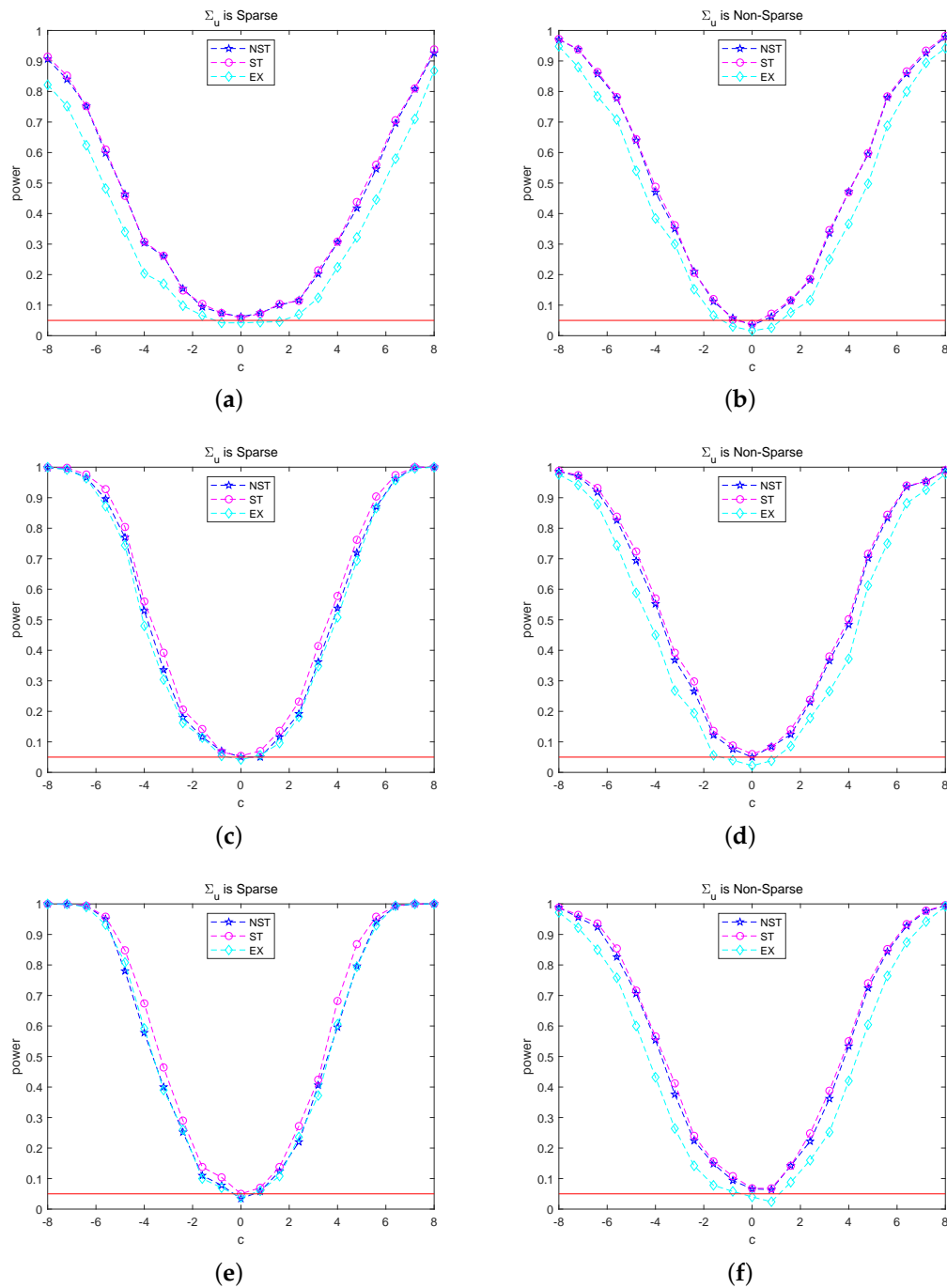


Figure 1. Empirical powers of the NST, ST, and EX methods. The figures in the left panels are based on Model 1, while those in the right panels are for Model 2. The red solid line corresponds to the nominal level. (a) $p = 10$, (b) $p = 10$, (c) $p = 200$, (d) $p = 200$, (e) $p = 600$, (f) $p = 600$.

Table 2. Empirical family-wise error rate (FWER) and false discovery rate (FDR) with power in the brackets of multiple testing based on Model 1, $\alpha = 0.05$, and 500 replications.

<i>T</i>	<i>s</i> ₀	Method	<i>p</i> = 50	<i>p</i> = 200	<i>p</i> = 500	<i>p</i> = 600
200	3	NST-FWER	0.058 (0.551)	0.062 (0.405)	0.048 (0.309)	0.056 (0.291)
		ST-FWER	0.074 (0.554)	0.082 (0.431)	0.086 (0.337)	0.090 (0.324)
		BH-FWER	0.054 (0.528)	0.070 (0.409)	0.074 (0.319)	0.068 (0.300)
		BH-FDR	0.061 (0.635)	0.064 (0.470)	0.086 (0.380)	0.069 (0.353)
	15	NST-FWER	0.056 (0.569)	0.050 (0.412)	0.040 (0.306)	0.046 (0.303)
		ST-FWER	0.066 (0.583)	0.086 (0.430)	0.074 (0.334)	0.084 (0.327)
		BH-FWER	0.060 (0.561)	0.066 (0.410)	0.056 (0.310)	0.068 (0.309)
		BH-FDR	0.043 (0.810)	0.064 (0.655)	0.06 (0.518)	0.061 (0.509)
400	3	NST-FWER	0.050 (0.935)	0.058 (0.889)	0.062 (0.839)	0.058 (0.808)
		ST-FWER	0.070 (0.937)	0.062 (0.885)	0.078 (0.842)	0.066 (0.813)
		BH-FWER	0.052 (0.931)	0.054 (0.873)	0.062 (0.834)	0.052 (0.795)
		BH-FDR	0.057 (0.957)	0.056 (0.924)	0.064 (0.889)	0.068 (0.863)
	15	NST-FWER	0.058 (0.947)	0.054 (0.881)	0.040 (0.819)	0.070 (0.815)
		ST-FWER	0.052 (0.946)	0.066 (0.881)	0.058 (0.825)	0.084 (0.882)
		BH-FWER	0.050 (0.942)	0.056 (0.871)	0.050 (0.809)	0.060 (0.806)
		BH-FDR	0.035 (0.989)	0.052 (0.968)	0.056 (0.946)	0.059 (0.941)

Table 3. Empirical FWER and FDR with power in the brackets of multiple testing based on Model 2, $\alpha = 0.05$, and 500 replications.

<i>T</i>	<i>s</i> ₀	Method	<i>p</i> = 50	<i>p</i> = 200	<i>p</i> = 500	<i>p</i> = 600
200	3	NST-FWER	0.044 (0.805)	0.052 (0.692)	0.066 (0.622)	0.056 (0.609)
		ST-FWER	0.058 (0.807)	0.066 (0.701)	0.084 (0.638)	0.066 (0.621)
		BH-FWER	0.030 (0.759)	0.042 (0.620)	0.024 (0.517)	0.024 (0.505)
		BH-FDR	0.039 (0.819)	0.046 (0.691)	0.038 (0.592)	0.030 (0.570)
	15	NST-FWER	0.042 (0.805)	0.060 (0.697)	0.058 (0.626)	0.048 (0.618)
		ST-FWER	0.050 (0.809)	0.080 (0.708)	0.080 (0.637)	0.072 (0.630)
		BH-FWER	0.028 (0.757)	0.042 (0.621)	0.034 (0.530)	0.038 (0.519)
		BH-FDR	0.035 (0.922)	0.044 (0.822)	0.046 (0.746)	0.040 (0.717)
400	3	NST-FWER	0.060 (0.989)	0.052 (0.985)	0.052 (0.971)	0.050 (0.970)
		ST-FWER	0.064 (0.989)	0.054 (0.985)	0.068 (0.970)	0.072 (0.973)
		BH-FWER	0.046 (0.983)	0.022 (0.975)	0.026 (0.951)	0.024 (0.945)
		BH-FDR	0.045 (0.995)	0.034 (0.990)	0.040 (0.972)	0.035 (0.973)
	15	NST-FWER	0.066 (0.992)	0.072 (0.986)	0.056 (0.975)	0.056 (0.975)
		ST-FWER	0.072 (0.992)	0.076 (0.986)	0.056 (0.975)	0.064 (0.975)
		BH-FWER	0.046 (0.988)	0.036 (0.973)	0.024 (0.952)	0.022 (0.950)
		BH-FDR	0.043 (0.999)	0.042 (0.998)	0.031 (0.993)	0.044 (0.992)

5. Real Data Analysis

This section conducts hypothesis testing for financial data from 1 January 2017 to 14 March 2018. The dataset consists of daily returns of 491 stocks from S&P 500 index. In addition, we collected Fama–French three factors [21] in the same period. In summary, the panel matrix is a 300 by 491 matrix **Y**, in addition to a factor matrix **F** of size 300 by 3. Here, 300 is the number of days and 491 is the number of stocks.

We first centralize and standardize the factor matrix **F** and **Y** is centralized as well. We consider testing the sparsity of each column of **B** and repeat the multiplier bootstrap 500 times. Simultaneous test of parameters corresponding to multiple factors is also considered. The hypotheses are

$$H_0 : b_{ik} = 0 \quad \text{for all } (i, k) \in s \quad \text{versus} \quad H_1 : b_{ik} \neq 0 \quad \text{for some } (i, k) \in s,$$

where $s = \{(i, k) : k \in s_* \subset \{1, 2, 3\}, |\widehat{b}_{ik}| \text{ are the smallest } \beta\% \text{ among } \{|\widehat{b}_{ik}|\}_{i \in \{1, \dots, p\}; k \in s_*}\}$, with $s_* = \{1\}, \{2\}, \{3\}$ or $\{2, 3\}$ and $\beta = 10, 30, 50, 70, 90$. The results are depicted in Table 4. For the first column of \mathbf{B} , it is therefore not reasonable to assume $b_{i1} = 0$. However, we can claim that the last two columns of \mathbf{B} are sparse. Hence, a sufficiently large number of stocks are not influenced by the last two factors.

Table 4. Results of sparse testing.

	$\beta = 10$	$\beta = 30$	$\beta = 50$	$\beta = 70$	$\beta = 90$
1st loading	R	R	R	R	R
2nd loading	A	A	A	A	A
3rd loading	A	A	A	A	R
2nd and 3rd loading	A	A	A	R	R

Note: “A” means accepting the null hypothesis; “R” denotes rejecting the null hypothesis.

Author Contributions: X.G. conceived and designed the experiments; Y.W. performed the experiments; Y.W. analyzed the data; X.G. contributed to analysis tools; X.G. and Y.W. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grants 12071452, 11601500, 11671374 and 11771418 and the Fundamental Research Funds for the Central Universities.

Acknowledgments: We thank the three referees for insightful comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Technical Details

We prove the main results in this section. First, we introduce some notations. Throughout this section, we denote by c, c', C, C', C_i constants that do not depend on p, T and may vary from place to place. Define $T_0 = \max_{i \in G} \sum_{t=1}^T \zeta_{it} / \sqrt{T}$ and $T_1 = \max_{i \in G} \sum_{t=1}^T \widehat{\zeta}_{it} / \sqrt{T}$. Let $\{z_t\}_{t=1}^T$ be a sequence of $N(0, \Omega_f(k, k)\Sigma_u)$ vectors. Define $c_{W_0}(\alpha) = \inf\{t \in \mathbb{R} : P(W_0 \leq t | (\mathbf{Y}, \mathbf{F})) \geq 1 - \alpha\}$ with $W_0 = \max_{i \in G} T^{-1/2} \sqrt{\widehat{\Omega}_f(k, k)} \sum_{t=1}^T \widehat{u}_{it} e_t$ and $c_{Z_0}(\alpha) = \inf\{t \in \mathbb{R} : P(Z_0 \leq t | (\mathbf{Y}, \mathbf{F})) \geq 1 - \alpha\}$ with $Z_0 = \max_{i \in G} \sum_{t=1}^T z_{it} / \sqrt{T}$. Denote by $\Sigma_f(m, n) = \mathbf{v}_m' \Sigma_f \mathbf{v}_n$ with $\Sigma_f = \Omega_f^{-1}$. We begin by presenting some useful lemmas that will be used in the proofs of the main results.

Lemma A1. Suppose that the random variables Z_1, Z_2 both satisfy the exponential-type tail condition: There exist $r_1, r_2 \in (0, 1)$ and $b_1, b_2 > 0$ such that $\forall s > 0$,

$$P(|Z_i| > s) \leq \exp\{1 - (s/b_i)^{r_i}\}, \quad i = 1, 2.$$

Then, for some r_3 and $b_3 > 0$, and any $s > 0$,

- (i) $P(|Z_1 Z_2| > s) \leq \exp\{1 - (s/b_3)^{r_3}\}$, where $r_3 \in (0, r)$ with $r = r_1 r_2 / (r_1 + r_2)$,
- (ii) $P(|Z| > s) \leq \exp\{1 - (s/b_3)^{r_3}\}$, where $Z = \max\{Z_1, Z_2\}$.

Proof of Lemma A1. The proof of the first claim can be found in the proof of Lemma A.2 in [4], thus we prove the second claim. For any $s > 0$, we have

$$\begin{aligned} P(|Z| > s) &\leq P(|Z_1| > s) + P(|Z_2| > s) \\ &= \exp\{1 - (s/b_1)^{r_1}\} + \exp\{1 - (s/b_2)^{r_2}\} \\ &\leq 2 \exp\{1 - (s/b)^r\}, \end{aligned}$$

where $b^r = \max\{b_1^{r_1}, b_2^{r_2}\}$. Pick up an $r_3 \in (0, r)$, and $b_3 > \max\{(r_3/r)^{1/r}b, (1 + \log 2)^{1/r}b\}$; then, it can be shown that $F(s) = (s/b)^r - (s/b_3)^{r_3}$ is increasing when $s > b_3$. Therefore, $F(s) > F(b_3) > \log 2$ when $s > b_3$, which implies when $s > b_3$,

$$P(|Z| > s) \leq 2 \exp\{1 - (s/b)^r\} \leq \exp\{1 - (s/b_3)^{r_3}\}.$$

When $s \leq b_3$,

$$P(|Z| > s) \leq 1 \leq \exp\{1 - (s/b_3)^{r_3}\}.$$

Then the proof is complete. \square

Lemma A2. Under Assumptions 1–4, we have

- (i) $\max_{i,j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| = O_p(\sqrt{\log p/T})$.
- (ii) $\max_{i \leq p, k \leq K} |(1/T) \sum_{t=1}^T f_{kt} u_{it}| = O_p(\sqrt{\log p/T})$.
- (iii) $\max_{i,j \leq K} |(1/T) \sum_{t=1}^T f_{it} f_{jt} - E f_{it} f_{jt}| = O_p(\sqrt{\log T/T})$.

Proof of Lemma A2. For a proof, see the proof of Lemma A.3 and Lemma B.1 in [4]. \square

Lemma A3. If a random variable X satisfies exponential-type tail: there exist $r > 0$ and $b > 0$ such that $\forall s > 0, P(|X| > s) \leq \exp\{1 - (s/b)^r\}$. Then $E(|X|) = O(1)$.

Proof of Lemma A3. Note that

$$\begin{aligned} E(|X|) &= \int_0^\infty P(|X| > x) dx \\ &\leq \int_0^\infty \exp\{1 - (x/b)^r\} dx \\ &\lesssim \int_0^\infty \exp(-x^r) dx := I \end{aligned}$$

It is not hard to check when $r \geq 1, I \leq 1 + e^{-1}$. When $r < 1, I = \alpha \Gamma(\alpha)$, where $\alpha = 1/r$ and $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx = O(1)$. Then, the proof is complete. \square

Lemma A4. Under the assumptions in Theorem 1, there exist constants $c, C > 0$ such that

$$\rho := \sup_{t \in \mathbb{R}} |P(T_0 \leq t) - P(Z_0 \leq t)| \leq CT^{-c}.$$

Proof of Lemma A4. Implied by Assumption 3, we have $(\log(pT))^7/T \leq C_1 T^{-c_1}$ for some constants $c_1, C_1 > 0$. We then apply Corollary 2.1 of [25] to the sequence ξ_{it} . What we should check is its Condition (E.2), that is, uniformly over i ,

$$c_0 \leq E(\xi_{it})^2 \leq C_0, \quad \max_{k=1,2} E(|\xi_{it}|^{k+2}/B^k) + E\{(\max_{1 \leq i \leq p} |\xi_{it}|/B)^4\} \leq 4,$$

where $c_0, C_0 > 0$ and B is some large enough constant. By Lemmas A1 and A3 we have $E(f_{it} f_{jt}) = O(1)$ uniformly for $i, j \leq K$. This implies $\|\Omega_f\|_\infty = O(1)$. Uniformly for $i \leq p$, we have

$$\xi_{it} = u_{it} \mathbf{f}'_t \Omega_f \mathbf{v}_k \leq u_{it} \|\mathbf{f}_t\|_\infty \|\Omega_f \mathbf{v}_k\|_1 := \gamma_{it}.$$

By Lemma A1, we know γ_{it} and $\max_{1 \leq i \leq p} |\gamma_{it}|$ is exponential-type tail. Then by Lemma A3 we have $E(|\xi_{it}|^4) \leq E(|\gamma_{it}|^4) = O(1)$ and $E(\max_{i \leq p} |\xi_{it}|^4) \leq E(\max_{i \leq p} |\gamma_{it}|^4) = O(1)$. Thus, we can find large enough B such that the above condition is satisfied. Then, the proof is complete. \square

Lemma A5. Under the assumptions in Theorem 1, there exists a sequence of positive numbers $\alpha_T \rightarrow \infty$ such that $\alpha_T/p = o(1)$ and $P(\alpha_T(\log p)^2 |\widehat{\Omega}_f(k, k) - \Omega_f(k, k)| > 1) \rightarrow 0$.

Proof of Lemma A5. By Lemma A2 (iii), we have $\|\widehat{\Omega}_f^{-1} - \Omega_f^{-1}\|_\infty = O_p(\sqrt{\log T/T})$. Since $\|\Omega_f\| = O(1)$ and $\|\widehat{\Omega}_f\|_\infty = O_p(1)$, we have

$$\begin{aligned} \|\widehat{\Omega}_f - \Omega_f\|_\infty &= \|\widehat{\Omega}_f(\Omega_f^{-1} - \widehat{\Omega}_f^{-1})\Omega_f\|_\infty \\ &\leq \|\widehat{\Omega}_f\|_\infty \|\widehat{\Omega}_f^{-1} - \Omega_f^{-1}\|_\infty \|\Omega_f\|_\infty \\ &= O_p(\sqrt{\log T/T}). \end{aligned}$$

On the other hand,

$$|\widehat{\Omega}_f(k, k) - \Omega_f(k, k)| \leq \|\widehat{\Omega}_f - \Omega_f\|_\infty = O_p(\sqrt{\log T/T}).$$

Choosing $\alpha_T = \sqrt{\log T}$, by Assumption 3, the proof is complete. \square

Lemma A6. Under the assumptions in Theorem 1, we have for every $\alpha \in (0, 1)$ and $\vartheta > 0$,

$$\begin{aligned} P(c_{W_0}(\alpha) \leq c_{Z_0}(\alpha + \pi(\vartheta))) &\geq 1 - P(\Delta > \vartheta), \\ P(c_{Z_0}(\alpha) \leq c_{W_0}(\alpha + \pi(\vartheta))) &\geq 1 - P(\Delta > \vartheta), \end{aligned}$$

Proof of Lemma A6. For $\vartheta > 0$, let $\pi(\vartheta) = C_2 \vartheta^{1/3} (1 \vee \log(p/\vartheta))^{2/3}$ with $C_2 > 0$. Recall that $\Delta = \max_{1 \leq i, j \leq p} |\widehat{\Omega}_f(k, k) \widehat{\sigma}_{ij} - \Omega_f(k, k) \sigma_{ij}|$. As $|\Omega_f(k, k)| = O(1)$ uniformly for $k \leq K$, by Lemma A2 (i), we have

$$\Delta = O_p(|\widehat{\Omega}_f(k, k) - \Omega_f(k, k)| + \sqrt{(\log p)/T}).$$

By Lemma A5 and Assumption 3, choosing $\vartheta = 1/(\alpha_T(\log p)^2)$, we have $P(\Delta > \vartheta) = o(1)$. By Lemma 3.1 of [25], on the event $\{(\mathbf{Y}, \mathbf{F}) : \Delta \leq \vartheta\}$, we have $|P(Z_0 \leq t) - P(P(W_0 \leq t | (\mathbf{Y}, \mathbf{F})))| \leq \pi(\vartheta)$ for all $t \in \mathbb{R}$, and so on this event

$$\begin{aligned} P(P(W_0 \leq c_{Z_0}(\alpha + \pi(\vartheta)) | (\mathbf{Y}, \mathbf{F}))) &\geq P(Z_0 \leq c_{Z_0}(\alpha + \pi(\vartheta))) - \pi(\vartheta) \\ &\geq \alpha + \pi(\vartheta) - \pi(\vartheta) = \alpha, \end{aligned}$$

implying the first claim. The second claim follows similarly. \square

Lemma A7. Under the assumptions in Theorem 1, there exist $\zeta_1, \zeta_2 > 0$ such that

$$P\left(\left| \max_{1 \leq i \leq p} \sum_{t=1}^T \widehat{\xi}_{it} / \sqrt{T} - \max_{1 \leq i \leq p} \sum_{t=1}^T \xi_{it} / \sqrt{T} \right| > \zeta_1\right) < \zeta_2,$$

where $\zeta_1 \sqrt{1 \vee \log(p/\zeta_1)} = o(1)$, $\zeta_2 = o(1)$.

Proof of Lemma A7. The arguments in the proof of Lemma A5 imply that

$$\|\widehat{\Omega}_f \mathbf{v}_k - \Omega_f \mathbf{v}_k\|_1 = O_p(\sqrt{\log T/T}).$$

By Lemma A2 (ii), uniformly for $i \leq p$, we have

$$\begin{aligned} \left| \sum_{t=1}^T \widehat{\zeta}_{it} / \sqrt{T} - \sum_{t=1}^T \zeta_{it} / \sqrt{T} \right| &= \left| \sum_{t=1}^T u_{it} \mathbf{f}'_t(\widehat{\Omega}_f \mathbf{v}_k - \Omega_f \mathbf{v}_k) / \sqrt{T} \right| \\ &\leq \|\widehat{\Omega}_f \mathbf{v}_k - \Omega_f \mathbf{v}_k\|_1 \max_{i \leq p, k \leq K} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T f_{kt} u_{it} \right| \\ &= O_p(\sqrt{\log p \log T / T}). \end{aligned}$$

Choosing $\zeta_1^2 = O(\sqrt{\log p \log T / T})$, we have

$$P\left(\max_{1 \leq i \leq p} \left| \sum_{t=1}^T \widehat{\zeta}_{it} / \sqrt{T} - \sum_{t=1}^T \zeta_{it} / \sqrt{T} \right| > \zeta_1\right) \leq \zeta_2, \quad \zeta_2 = o(1).$$

Note that

$$\left| \max_{1 \leq i \leq p} \sum_{t=1}^T \widehat{\zeta}_{it} / \sqrt{T} - \max_{1 \leq i \leq p} \sum_{t=1}^T \zeta_{it} / \sqrt{T} \right| \leq \max_{1 \leq i \leq p} \left| \sum_{t=1}^T \widehat{\zeta}_{it} / \sqrt{T} - \sum_{t=1}^T \zeta_{it} / \sqrt{T} \right|,$$

then the proof is complete. \square

Lemma A8. Under the assumptions in Theorem 4, we have

- (i) $\max_{i,j,m,n} |(1/T) \sum_{t=1}^T u_{it} u_{jt} f_{mt} f_{nt} - \sigma_{ij} \Sigma_f(m, n)| = O_p(\sqrt{\log p / T})$,
- (ii) $\max_{i,j,m,n} |(1/T) \sum_{t=1}^T u_{it} f_{jt} f_{mt} f_{nt}| = O_p(\sqrt{\log p / T})$,

Proof of Lemma A8. (i) By Assumption 1 and Lemma A1, $u_{it} f_{mt}$ satisfies the exponential tail condition, with parameter $2r_1 r_2 / (3r_1 + 3r_2)$ as shown in Lemma A1. Thus, $u_{it} u_{jt} f_{mt} f_{nt}$ satisfies the exponential tail condition, with parameter $r_1 r_2 / (4r_1 + 4r_2)$. It follows from $1.5(r_1^{-1} + r_2^{-1}) > 1$ that $\gamma_3 < 1$. Therefore, by the Bernstein’s inequality [23], there exist constants $C_i, i = 1, \dots, 5$, for any $s > 0$

$$\begin{aligned} \max_{i,j,m,n} P\left(\left| \frac{1}{T} \sum_{t=1}^T u_{it} u_{jt} f_{mt} f_{nt} - \sigma_{ij} \Sigma_f(m, n) \right| \geq s\right) &\leq T \exp\left(-\frac{(Ts)^{\gamma_3}}{C_1}\right) \\ &+ \exp\left(-\frac{T^2 s^2}{C_2(1 + TC_3)}\right) + \exp\left(-\frac{(Ts)^2}{C_4 T} \exp\left(\frac{(Ts)^{\gamma_3(1-\gamma_3)}}{C_5(\log Ts)^{\gamma_3}}\right)\right). \end{aligned} \tag{A1}$$

Using Bonferroni’s method, we have

$$\begin{aligned} P\left(\max_{i,j,m,n} \left| \frac{1}{T} \sum_{t=1}^T u_{it} u_{jt} f_{mt} f_{nt} - \sigma_{ij} \Sigma_f(m, n) \right| > s\right) \\ \leq (pK)^2 \max_{i,j,m,n} P\left(\left| \frac{1}{T} \sum_{t=1}^T u_{it} u_{jt} f_{mt} f_{nt} - \sigma_{ij} \Sigma_f(m, n) \right| > s\right). \end{aligned}$$

Let $s = C\sqrt{(\log p) / T}$ for some $C > 0$. It is not hard to check that when $(\log p)^{2/\gamma_3-1} = o(T)$ (by assumption), for large enough C ,

$$p^2 T \exp\left(-\frac{(Ts)^{\gamma_3}}{C_1}\right) + p^2 \exp\left(-\frac{(Ts)^2}{C_4 T} \exp\left(\frac{(Ts)^{\gamma_3(1-\gamma_3)}}{C_5(\log Ts)^{\gamma_3}}\right)\right) = o\left(\frac{1}{p^2}\right)$$

and

$$p^2 \exp\left(-\frac{T^2 s^2}{C_2(1+TC_3)}\right) = O\left(\frac{1}{p^2}\right).$$

As $K = O(1)$, this proves (i).

(ii) By Assumption 1 and Lemma A1, $u_{it}f_{jt}f_{mt}f_{nt}$ satisfies the exponential tail condition for the tail parameter $r_1r_2/(9r_1 + 3r_2)$. Therefore, again by the Bernstein’s inequality and the Bonferroni method on $u_{it}f_{jt}f_{mt}f_{nt}$ similar to (A1) with the parameter $r_3^{-1} = 3r_1^{-1} + 9r_2^{-1}$, it follows from $1.5(r_1^{-1} + r_2^{-1}) > 1$ that $r_3 < 1$. Thus, when $s = C\sqrt{\log p/T}$ for large enough C , as K is fixed, the term

$$pK^3 \exp\left(-\frac{T^2 s^2}{C_2(1+TC_3)}\right) \leq p^{-2},$$

and the rest terms on the right-hand side of the inequality, multiplied by pK^3 are of order $o(p^{-2})$. Hence when $(\log p)^{2/r_3-1} = o(T)$ (by assumption), we have

$$\max_{i,j,m,n} \left| \frac{1}{T} \sum_{t=1}^T u_{it}f_{jt}f_{mt}f_{nt} \right| = O_p(\sqrt{\log p/T}),$$

which completes the proof. \square

Lemma A9. Under the assumptions in Theorem 4, we have

- (i) $\max_{i,j,m,n} |(1/T) \sum_{t=1}^T (\hat{u}_{it}\hat{u}_{jt}f_{mt}f_{nt} - u_{it}u_{jt}f_{mt}f_{nt})| = O_p(\log p/T)$.
- (ii) $\max_{i,j,m,n} |(1/T) \sum_{t=1}^T \hat{u}_{it}\hat{u}_{jt}f_{mt}f_{nt} - \sigma_{ij}\Sigma_f(m, n)| = O_p(\sqrt{\log p/T})$.

Proof of Lemma A9. (i) By the triangular inequality, we have

$$\begin{aligned} & \left| \frac{1}{T} \sum_{t=1}^T (\hat{u}_{it}\hat{u}_{jt}f_{mt}f_{nt} - u_{it}u_{jt}f_{mt}f_{nt}) \right| \\ & \leq \underbrace{\left| \frac{1}{T} \sum_{t=1}^T (\hat{u}_{it}\hat{u}_{jt}f_{mt}f_{nt} - \hat{u}_{it}u_{jt}f_{mt}f_{nt}) \right|}_I + \underbrace{\left| \frac{1}{T} \sum_{t=1}^T (\hat{u}_{it}u_{jt}f_{mt}f_{nt} - u_{it}u_{jt}f_{mt}f_{nt}) \right|}_{II} \end{aligned}$$

For I , we have

$$\begin{aligned} I & = \left| \frac{1}{T} \sum_{t=1}^T \hat{u}_{it}(\hat{\mathbf{b}}_j - \mathbf{b}_j)' \mathbf{f}_t f_{mt} f_{nt} \right| \\ & \leq \underbrace{\left| \frac{1}{T} \sum_{t=1}^T (\hat{\mathbf{b}}_i - \mathbf{b}_i)' \mathbf{f}_t (\hat{\mathbf{b}}_j - \mathbf{b}_j)' \mathbf{f}_t f_{mt} f_{nt} \right|}_i + \underbrace{\left| \frac{1}{T} \sum_{t=1}^T u_{it}(\hat{\mathbf{b}}_j - \mathbf{b}_j)' \mathbf{f}_t f_{mt} f_{nt} \right|}_{ii} \end{aligned}$$

By Lemma 3.1 of [4], we have $\max_{i \leq p} \|\hat{\mathbf{b}}_i - \mathbf{b}_i\| = O_p(\sqrt{\log p/T})$. It is straightforward to see that

$$\max_{m,n \leq K} \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t \mathbf{f}_t' f_{mt} f_{nt} \right\|_\infty = O_p(1).$$

then we have,

$$i \leq \max_{i,m,n} \|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2 \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t \mathbf{f}_t' f_{mt} f_{nt} \right\|_\infty = O_p(\log p/T).$$

By Lemma A8 (ii), we have $\|(1/T) \sum_{t=1}^T u_{it} \mathbf{f}_t f_{mt} f_{nt}\|_\infty = O_p(\sqrt{\log p/T})$, which implies that

$$ii \leq \max_{j \leq p} \|\widehat{\mathbf{b}}_j - \mathbf{b}_j\| \left\| \frac{1}{T} \sum_{t=1}^T u_{it} \mathbf{f}_t f_{mt} f_{nt} \right\|_\infty = O_p(\log p/T)$$

Part II is similar to ii, thus we have

$$II = \left| \frac{1}{T} \sum_{t=1}^T (\widehat{\mathbf{b}}_i - \mathbf{b}_i)' \mathbf{f}_t u_{jt} f_{mt} f_{nt} \right| \leq O_p(\log p/T).$$

Then the proof is complete.

(ii) By the triangular inequality, we have

$$\begin{aligned} & \max_{i,j,m,n} \left| \frac{1}{T} \sum_{t=1}^T \widehat{u}_{it} \widehat{u}_{jt} f_{mt} f_{nt} - \sigma_{ij} \boldsymbol{\Sigma}_f(m,n) \right| \\ & \leq \max_{i,j,m,n} \left| \frac{1}{T} \sum_{t=1}^T (\widehat{u}_{it} \widehat{u}_{jt} f_{mt} f_{nt} - u_{it} u_{jt} f_{mt} f_{nt}) \right| + \max_{i,j,m,n} \left| \frac{1}{T} \sum_{t=1}^T u_{it} u_{jt} f_{mt} f_{nt} - \sigma_{ij} \boldsymbol{\Sigma}_f(m,n) \right| \\ & = O_p(\sqrt{\log p/T}), \end{aligned}$$

which proves the result. \square

Proof of Theorem 1. Without loss of generality, we set $G = \{1, 2, \dots, p\}$. First, we prove the following fact,

$$\sup_{\alpha \in (0,1)} |\mathbb{P}(T_1 > c_{W_0}(\alpha)) - \alpha| = o(1), \tag{A2}$$

For $\vartheta > 0$, let $\pi(\vartheta) := C_2 \vartheta^{1/3} (1 \vee \log(p/\vartheta))^{2/3}$ with $C_2 > 0$. In addition, Let $\kappa_1(\vartheta) := c_{Z_0}(\alpha - \zeta_2 - \pi(\vartheta))$ and $\kappa_2(\vartheta) := c_{Z_0}(\alpha + \zeta_2 + \pi(\vartheta))$. For every $\alpha \in (0, 1)$, note that

$$\begin{aligned} & \mathbb{P}(\{T_1 \leq c_{W_0}(\alpha)\} \ominus \{T_0 \leq c_{Z_0}(\alpha)\}) \\ & \lesssim_{(1)} \mathbb{P}(\kappa_1(\vartheta) - 2\zeta_1 < T_0 \leq \kappa_2(\vartheta) + 2\zeta_1) + \mathbb{P}(\Delta > \vartheta) + \zeta_2 \\ & \lesssim_{(2)} \mathbb{P}(\kappa_1(\vartheta) - 2\zeta_1 < Z_0 \leq \kappa_2(\vartheta) + 2\zeta_1) + \mathbb{P}(\Delta > \vartheta) + \rho + \zeta_2 \\ & \lesssim_{(3)} \pi(\vartheta) + \mathbb{P}(\Delta > \vartheta) + \rho + \zeta_1 \sqrt{1 \vee \log(p/\zeta_1)} + \zeta_2, \end{aligned}$$

where (1) follows from Lemmas A6 and A7, (2) follows from Lemma A4, and (3) follows from Lemma 2.1 in [25] and the fact that Z_0 has no point masses. Then, by the definition of ρ in Lemma A4, we have

$$\sup_{\alpha \in (0,1)} |\mathbb{P}(T_1 > c_{W_0}(\alpha)) - \alpha| \leq \rho_\ominus + \rho,$$

where $\rho_\ominus = \sup_{\alpha \in (0,1)} \mathbb{P}(\{T_1 \leq c_{W_0}(\alpha)\} \ominus \{T_0 \leq c_{Z_0}(\alpha)\})$. The right-hand side of the above inequality is $o(1)$, which has proved (A2). Since $\max_{i \in G} \sqrt{T} |\widehat{b}_{ik} - b_{ik}| = \sqrt{T} \max_{i \in G} \max\{\widehat{b}_{ik} - b_{ik}, b_{ik} - \widehat{b}_{ik}\}$, similar arguments imply that

$$\sup_{\alpha \in (0,1)} \left| \mathbb{P}\left(\max_{i \in G} \sqrt{T} |\widehat{b}_{ik} - b_{ik}| > c_{W_{T,k}}(\alpha)\right) - \alpha \right| = o(1),$$

which completes the proof. \square

Proof of Theorem 2. From the arguments in the proof of Lemma A6, we have

$$\Delta = O_p(|\widehat{\mathbf{\Omega}}_f(k, k) - \mathbf{\Omega}_f(k, k)| + \sqrt{\log p/T}),$$

which implies that $\max_{1 \leq i \leq p} |\widehat{\omega}_{ii} - \omega_{ii}| = O_p(|\widehat{\mathbf{\Omega}}_f(k, k) - \mathbf{\Omega}_f(k, k)| + \sqrt{\log p/T})$. We then have

$$P(\omega_{ii}/2 < \widehat{\omega}_{ii} < 2\omega_{ii} \text{ for all } 1 \leq i \leq p) \rightarrow 1. \tag{A3}$$

Define $\bar{T}_1 = \max_{i \in G} \sum_{t=1}^T \widehat{\xi}_{it} / \sqrt{T\widehat{\omega}_{ii}}$ and $\bar{T}_0 = \max_{i \in G} \sum_{t=1}^T \xi_{it} / \sqrt{T\omega_{ii}}$. Note that

$$\begin{aligned} & |\bar{T}_1 - \bar{T}_0| \\ & \leq \max_{1 \leq i \leq p} \left| \sum_{t=1}^T \widehat{\xi}_{it} / \sqrt{T\widehat{\omega}_{ii}} - \sum_{t=1}^T \xi_{it} / \sqrt{T\omega_{ii}} \right| \\ & \leq \max_{1 \leq i \leq p} \left| \sum_{t=1}^T \widehat{\xi}_{it} / \sqrt{T\widehat{\omega}_{ii}} - \sum_{t=1}^T \widehat{\xi}_{it} / \sqrt{T\omega_{ii}} \right| + \max_{1 \leq i \leq p} \left| \sum_{t=1}^T \widehat{\xi}_{it} / \sqrt{T\omega_{ii}} - \sum_{t=1}^T \xi_{it} / \sqrt{T\omega_{ii}} \right| \\ & \leq C' \max_{1 \leq i \leq p} \left| \sum_{t=1}^T \widehat{\xi}_{it} / \sqrt{T} \right| \max_{1 \leq i \leq p} |\sqrt{\omega_{ii}/\widehat{\omega}_{ii}} - 1| + C'' \max_{1 \leq i \leq p} \left| \sum_{t=1}^T (\widehat{\xi}_{it} - \xi_{it}) / \sqrt{T} \right| \\ & := I_1 + I_2, \end{aligned}$$

where $C', C'' > 0$.

On the event $\omega_{ii}/2 < \widehat{\omega}_{ii} < 2\omega_{ii}$ for all $1 \leq i \leq p$,

$$\begin{aligned} \max_{1 \leq i \leq p} |\sqrt{\omega_{ii}/\widehat{\omega}_{ii}} - 1| & \leq \max_{1 \leq i \leq p} |\sqrt{\omega_{ii}} - \sqrt{\widehat{\omega}_{ii}}| \max_{1 \leq i \leq p} \sqrt{2/\omega_{ii}} \\ & \leq \max_{1 \leq i \leq p} \left| \frac{\omega_{ii} - \widehat{\omega}_{ii}}{\sqrt{\omega_{ii}} + \sqrt{\widehat{\omega}_{ii}}} \right| \max_{1 \leq i \leq p} \sqrt{2/\omega_{ii}} \\ & \leq \max_{1 \leq i \leq p} |\omega_{ii} - \widehat{\omega}_{ii}| \max_{1 \leq i \leq p} 1/\omega_{ii} \\ & = O_p(|\widehat{\mathbf{\Omega}}_f(k, k) - \mathbf{\Omega}_f(k, k)| + \sqrt{(\log p)/T}). \end{aligned}$$

On the other hand,

$$\begin{aligned} \max_{1 \leq i \leq p} \left| \sum_{t=1}^T \widehat{\xi}_{it} / \sqrt{T} \right| & \leq \max_{1 \leq i \leq p} \left| \sum_{t=1}^T (\widehat{\xi}_{it} - \xi_{it}) / \sqrt{T} \right| + \max_{1 \leq i \leq p} \left| \sum_{t=1}^T \xi_{it} / \sqrt{T} \right| \\ & = O_p(\sqrt{\log p \log T/T} + \sqrt{\log p}) = O_p(\sqrt{\log p}). \end{aligned}$$

Therefore, on the above event, $I_1 \leq O_p(\sqrt{\log p} |\widehat{\mathbf{\Omega}}_f(k, k) - \mathbf{\Omega}_f(k, k)| + \log p / \sqrt{T})$. By Lemma A5, we can find ζ'_1 such that $P(I_1 > \zeta'_1) = o(1)$ and $\zeta'_1 \sqrt{1 \vee \log(p/\zeta'_1)} = o(1)$. Thus by Lemma A7 and (A3), we have

$$P(|\bar{T}_1 - \bar{T}_0| > \zeta_1) \leq P(I_1 + I_2 > \zeta_1) < \zeta_2,$$

for $\zeta_1 \sqrt{1 \vee \log(p/\zeta_1)} = o(1)$ and $\zeta_2 = o(1)$.

Let $\bar{\Delta} = \max_{1 \leq j, k \leq p} |\omega_{jk} / \sqrt{\omega_{jj}\omega_{kk}} - \widehat{\omega}_{jk} / \sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}|$. Note that

$$|\sqrt{\omega_{jj}\omega_{kk}} - \sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}| = \frac{|\omega_{jj}\omega_{kk} - \widehat{\omega}_{jj}\widehat{\omega}_{kk}|}{\sqrt{\omega_{jj}\omega_{kk}} + \sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}}.$$

On the event $\omega_{ii}/2 < \widehat{\omega}_{ii} < 2\omega_{ii}$ for all $1 \leq i \leq p$, we have

$$\frac{|\omega_{jj}\omega_{kk} - \widehat{\omega}_{jj}\widehat{\omega}_{kk}|}{\sqrt{\omega_{jj}\omega_{kk}} + \sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}} \leq \frac{|\omega_{jj}\omega_{kk} - \widehat{\omega}_{jj}\widehat{\omega}_{kk}|}{\sqrt{\omega_{jj}\omega_{kk}} + \sqrt{\omega_{jj}\omega_{kk}/4}} \leq (2/3)|\omega_{jj}\omega_{kk} - \widehat{\omega}_{jj}\widehat{\omega}_{kk}| \max_{1 \leq j \leq p} 1/\omega_{jj},$$

which implies that

$$\begin{aligned} \max_{1 \leq j, k \leq p} |\sqrt{\omega_{jj}\omega_{kk}/\widehat{\omega}_{jj}\widehat{\omega}_{kk}} - 1| &\leq \max_{1 \leq j, k \leq p} |\sqrt{\omega_{jj}\omega_{kk}} - \sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}| \max_{1 \leq j \leq p} 2/\omega_{jj} \\ &\leq (4/3) \max_{1 \leq j, k \leq p} |\omega_{jj}\omega_{kk} - \widehat{\omega}_{jj}\widehat{\omega}_{kk}| \max_{1 \leq j \leq p} 1/\omega_{jj}^2 \\ &= O_p(|\widehat{\Omega}_f(k, k) - \Omega_f(k, k)| + \sqrt{(\log p)/T}). \end{aligned}$$

Choosing $\vartheta = 1/(\alpha_T(\log p)^2)$, we can show that $P(\bar{\Delta} > \vartheta) = o(1)$. The rest of the proofs are similar to those in the proof of Theorem 1. We skip the details. \square

Proof of Theorem 3. Let $Z = (Z_1, \dots, Z_p) \stackrel{d}{\sim} N(0, \Theta)$. Following the arguments in the proof of Theorem 2, we can show that the distribution of $\max_{i \in G} \sqrt{T}|\widehat{b}_{ik} - b_{ik}|/\sqrt{\widehat{\omega}_{ii}}$ can be approximated by $\max_{i \in G} |Z_i|$. Under Assumption 5, by Lemma 6 of [15], we have for any $x \in \mathbb{R}$ and as $|G| \rightarrow \infty$,

$$P\left(\max_{i \in G} |Z_i|^2 - 2 \log(|G|) + \log \log(|G|) \leq x\right) \rightarrow F(x) := \exp\left\{-\frac{1}{\sqrt{\pi}} \exp\left(-\frac{x}{2}\right)\right\}.$$

It implies that

$$P\left(\max_{i \in G} T|\widehat{b}_{ik} - b_{ik}|^2/\widehat{\omega}_{ii} \leq 2 \log(|G|) - \log \log(|G|)/2\right) \rightarrow 1. \tag{A4}$$

The bootstrap consistency result implies that

$$|(c_{W_{T,k}}^*(\alpha))^2 - 2 \log(|G|) + \log \log(|G|) - q_\alpha| = o_p(1), \tag{A5}$$

where q_α is the $100(1 - \alpha)$ th quantile of $F(x)$. Consider any $i \in G$ such that $|b_{ik}^{\text{null}} - b_{ik}|/\sqrt{\omega_{ii}} > (\sqrt{2} + \varepsilon_0)\sqrt{\log |G|/T}$. Using the inequality $2a_1a_2 \leq \delta^{-1}a_1^2 + \delta a_2^2$ for any $\delta > 0$, we have

$$T|b_{ik}^{\text{null}} - b_{ik}|^2/\widehat{\omega}_{ii} \leq (1 + \delta^{-1})T|\widehat{b}_{ik} - b_{ik}|^2/\widehat{\omega}_{ii} + (1 + \delta)T|\widehat{b}_{ik} - b_{ik}^{\text{null}}|^2/\widehat{\omega}_{ii}, \tag{A6}$$

where $T|\widehat{b}_{ik} - b_{ik}|^2/\widehat{\omega}_{ii} = o_p(\log |G|)$ as i is fixed and $|G|$ grows. From the proof of Theorem 2, we know the difference between $T|b_{ik}^{\text{null}} - b_{ik}|^2/\widehat{\omega}_{ii}$ and $T|b_{ik}^{\text{null}} - b_{ik}|^2/\omega_{ii}$ is asymptotically negligible. Thus, by (A6) and the fact that $\mathbf{B}_k \in \mathcal{U}_G(\sqrt{2} + \varepsilon_0)$, we have

$$\max_{i \in G} T|\widehat{b}_{ik} - b_{ik}^{\text{null}}|^2/\widehat{\omega}_{ii} \geq \frac{1}{1 + \delta} \left\{ (\sqrt{2} + \varepsilon_0)^2 (\log |G|) - o_p(\log |G|) \right\}. \tag{A7}$$

The conclusion thus follows from (A7) and (A5) provided that δ is small enough. \square

Proof of Theorem 4. Without loss of generality, we set $G^* = M$. Define

$$\widehat{\Gamma} = \frac{1}{T} \sum_{t=1}^T \widehat{u}_{it} \widehat{u}_{jt} \mathbf{f}_t \mathbf{f}_t', \quad \text{and} \quad \Gamma = \sigma_{ij} \Sigma_f.$$

Let

$$\Delta^* := \max_{i, j, m, n} \left| (\widehat{\Omega}_f \mathbf{v}_m)' \widehat{\Gamma} (\widehat{\Omega}_f \mathbf{v}_n) - \sigma_{ij} \Omega_f(m, n) \right|$$

denote the maximum discrepancy between the empirical and population covariance matrices. By the triangular inequality, we have

$$\begin{aligned} & |(\widehat{\Omega}_f \mathbf{v}_m)' \widehat{\Gamma}(\widehat{\Omega}_f \mathbf{v}_n) - \sigma_{ij} \Omega_f(m, n)| \\ &= |(\widehat{\Omega}_f \mathbf{v}_m)' \widehat{\Gamma}(\widehat{\Omega}_f \mathbf{v}_n) - (\Omega_f \mathbf{v}_m)' \Gamma(\Omega_f \mathbf{v}_n)| \\ &\leq \underbrace{|(\widehat{\Omega}_f \mathbf{v}_m)' (\widehat{\Gamma} - \Gamma)(\widehat{\Omega}_f \mathbf{v}_n)|}_I + \underbrace{|(\widehat{\Omega}_f \mathbf{v}_m - \Omega_f \mathbf{v}_m)' \Gamma(\widehat{\Omega}_f \mathbf{v}_n)|}_II + \underbrace{|(\Omega_f \mathbf{v}_m)' \Gamma(\widehat{\Omega}_f - \Omega_f) \mathbf{v}_n|}_III. \end{aligned}$$

Note that $\|\widehat{\Omega}_f\|_\infty = O_p(1)$, by Lemma A9 (ii), we have

$$I \leq \|\widehat{\Omega}_f \mathbf{v}_m\|_\infty^2 \|\widehat{\Gamma} - \Gamma\|_\infty = O_p(\sqrt{\log p/T}).$$

By Lemma A2 (iii) and $\|\Gamma\|_\infty = O(1)$, we have

$$II \leq \|\widehat{\Omega}_f \mathbf{v}_m - \Omega_f \mathbf{v}_m\|_\infty \|\Gamma\|_\infty \|\widehat{\Omega}_f \mathbf{v}_n\|_\infty = O_p(\sqrt{\log T/T}).$$

Since $\|\Omega_f\|_\infty = O(1)$, we have

$$III \leq \|(\Omega_f \mathbf{v}_m)'\|_\infty \|\Gamma\|_\infty \|(\widehat{\Omega}_f - \Omega_f) \mathbf{v}_n\|_\infty = O_p(\sqrt{\log T/T}).$$

The above results hold uniformly for i, j, m, n , thus we have

$$\Delta^* = O_p(\sqrt{\log p/T} + \sqrt{\log T/T}).$$

The rest of the proofs is similar to those in the proof of Theorem 2. We skip the details. \square

Proof of Theorem 5. For a proof, see the proof of Theorem 3. \square

References

- Bai, J.; Liao, Y. Efficient estimation of approximate factor models via penalized maximum likelihood. *J. Econ.* **2016**, *191*, 1–18.
- Heinemann, A. Efficient estimation of factor models with time and cross-sectional dependence. *J. Appl. Econ.* **2017**, *32*, 1107–1122.
- Fan, J.; Fan, Y.; Lv, J. High dimensional covariance matrix estimation using a factor model. *J. Econ.* **2008**, *147*, 186–197.
- Fan, J.; Liao, Y.; Mincheva, M. High-dimensional covariance matrix estimation in approximate factor models. *Ann. Stat.* **2011**, *39*, 3320–3356.
- Fan, J.; Liao, Y.; Mincheva, M. Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2013**, *75*, 603–680.
- Fan, J.; Wang, W.; Zhong, Y. Robust covariance estimation for approximate factor models. *J. Econom.* **2019**, *208*, 5–22.
- Dickhaus, T.; Pauly, M. Simultaneous statistical inference in dynamic factor models. In *Time Series Analysis and Forecasting*; Springer: Berlin, Germany, 2016; pp. 27–45.
- Dickhaus, T.; Sirotko-Sibirskaya, N. Simultaneous statistical inference in dynamic factor models: Chi-square approximation and model-based bootstrap. *Comput. Stat. Data Anal.* **2019**, *129*, 30–46.
- Lucas, J.; Carvalho, C.; Wang, Q.; Bild, A.; Nevins, J.R.; West, M. Sparse statistical modelling in gene expression genomics. *Bayesian Inference Gene Expr. Proteom.* **2006**, *1*, 155–176.
- Carvalho, C.M.; Chang, J.; Lucas, J.E.; Nevins, J.R.; Wang, Q.; West, M. High-dimensional sparse factor modeling: applications in gene expression genomics. *J. Am. Stat. Assoc.* **2008**, *103*, 1438–1456.

11. Reis, R.; Watson, M.W. Relative goods' prices, pure inflation, and the Phillips correlation. *Am. Econ. J. Macroecon.* **2010**, *2*, 128–157.
12. Amengual, D.; Repetto, L. *Testing a Large Number of Hypotheses in Approximate Factor Models*; Technical Report; CEMFI: Madrid, Spain, 2014.
13. Candès, E.; Tao, T. The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Stat.* **2007**, *35*, 2313–2351.
14. Cai, T.; Liu, W.; Xia, Y. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Am. Stat. Assoc.* **2013**, *108*, 265–277. [[CrossRef](#)]
15. Cai, T.T.; Liu, W.; Xia, Y. Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. (Stat. Methodol.)* **2014**, *76*, 349–372.
16. Zhang, X.; Cheng, G. Simultaneous inference for high-dimensional linear models. *J. Am. Stat. Assoc.* **2017**, *112*, 757–768. [[CrossRef](#)]
17. Romano, J.P.; Wolf, M. Exact and approximate stepdown methods for multiple hypothesis testing. *J. Am. Stat. Assoc.* **2005**, *100*, 94–108.
18. Zhu, Y.; Yu, Z.; Cheng, G. High dimensional inference in partially linear models. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS), Naha, Okinawa, Japan, 16–18 April 2019; Volume 89, pp. 2760–2769.
19. Zhang, X.; Cheng, G. Bootstrapping high dimensional time series. *arXiv* **2014**, arXiv:1406.1037.
20. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. (Methodol.)* **1995**, *57*, 289–300. [[CrossRef](#)]
21. Fama, E.F.; French, K.R. The cross-section of expected stock returns. *J. Financ.* **1992**, *47*, 427–465. [[CrossRef](#)]
22. Bickel, P.; Levina, E. Some theory for Fisher's linear discriminant function, "naive Bayes", and some alternatives when there are many more variables than observations. *Bernoulli* **2004**, *10*, 989–1010. [[CrossRef](#)]
23. Merlevède, F.; Peligrad, M.; Rio, E. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Relat. Fields* **2011**, *151*, 435–474. [[CrossRef](#)]
24. Kosorok, M.R. *Introduction to Empirical Processes and Semiparametric Inference*; Springer: New York, NY, USA, 2008.
25. Chernozhukov, V.; Chetverikov, D.; Kato, K. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Stat.* **2013**, *41*, 2786–2819.
26. Ingster, Y.I.; Tsybakov, A.B.; Verzelen, N. Detection boundary in sparse regression. *Electron. J. Stat.* **2010**, *4*, 1476–1526. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).