Erasmus School of Economics

Power Enhancement of Sparse Alternatives

Bachelor Thesis BSc Econometrics and Operational Research

Author: Richard Lu
Student ID number: 511396

Supervisor: Nick Koning
Second assessor: Richard Paap

July 4, 2021

Abstract Fan, Liao, and Yao (2015) proposed a technique to boost the power of testing a high-dimensional vector H_0: $\boldsymbol{\theta} = \mathbf{0}$ against sparse alternatives where the null hypothesis is violated by only a few components. In this paper I improve their technique by dividing the vector $\boldsymbol{\theta}$ into two sub-vectors: $\boldsymbol{\theta_S}$ and $\boldsymbol{\theta_D}$, where $\boldsymbol{\theta_S}$ is a sparse vector containing the sparse alternatives and $\boldsymbol{\theta_D}$=$\boldsymbol{\theta}$-$\boldsymbol{\theta_S}$ is dense. Instead of empowering all the elements of $\boldsymbol{\theta}$ as Fan, Liao, and Yao (2015) did, only $\boldsymbol{\theta_S}$ is empowered in this research, which leads to a new power enhancement method. A new power enhancement component is formed through a screening technique, which screens out the elements of $\boldsymbol{\theta_S}$ that are bigger than a critical value. The proposed power enhancement is applied to testing the factor pricing models and validating the cross-sectional independence in panel data models. Simulation results show that the proposed power enhancement not only reduces size distortion under the null hypothesis, but also provides more power under sparse alternatives in comparison with the power enhancement method of Fan, Liao, and Yao (2015).

1 Introduction

The effect of dimensionality on power properties of tests has witnessed a lot of growing attention in recent years. Existing tests based on quadratic statistics are known to have low power against subsets of the parameter space in high dimensions(Kock and Preinerstorfer 2017). Fan, Liao, and Yao (2015) introduced a power enhancement principle, which is a technique to boost the power of testing a high-dimensional vector against sparse alternatives where the null hypothesis is violated by only a few components. In their paper they test the following high-dimensional structural parameter:

$$H_0 : \boldsymbol{\theta} = \mathbf{0}, \tag{1}$$

where N=dim($\boldsymbol{\theta}$) is allowed to grow faster than the sample size T. The power enhancement introduced by Fan, Liao, and Yao (2015) is defined on the whole parameter space. To improve their method I divide the vector $\boldsymbol{\theta}$ into two sub-vectors: $\boldsymbol{\theta_S}$ and $\boldsymbol{\theta_D}$. $\boldsymbol{\theta_S}$ is a sparse vector and contains the sparse alternatives, whereas $\boldsymbol{\theta_D}$=$\boldsymbol{\theta}$-$\boldsymbol{\theta_S}$ is dense. The method in this paper prevents wrong rejection of the null hypothesis because $\boldsymbol{\theta_D}$ is not included in the power enhancement component. Furthermore, it can enhance the power of the test under the alternative hypothesis and decrease the size distortion under the null hypothesis. This research not only is theoretically relevant, but also has great use for traders in practice. Namely, it can show whether a stock is mispriced or not based on testing the null hypothesis so that it potentially improves the discovery of mispriced stocks. For example, if one wants to test the null hypothesis for N stocks, only the θ's of the stocks that are suspected to be non-zero need to be empowered instead of all the θ's.

A typical example to test the power enhancement on is the factor pricing models in economics. Factor models play a fundamental role in the arbitrage theory and practice of capital asset pricing (Ross 1976). It uses multiple common factors to capture the systematic risk and explain financial market occurrences such as the co-movements of securities and equilibrium of asset prices. In this paper, the following factor model is used:

$$y_{it} = \theta_i + \boldsymbol{b_i} * \boldsymbol{f_t} + u_{it} \quad \text{for} \quad i = 1, \ldots, N \quad t = 1, \ldots, T, \tag{2}$$

where y_{it} denotes the excess return of the ith asset at time t, θ_i is the intercept of asset i, $\boldsymbol{f_t}$ is the K-dimensional observable factors, $\boldsymbol{b_i}$ is a vector of factor loadings, u_{it} represents the idiosyncratic error, N = dim(θ) and T is the sample size. I am interested in testing zero θ_i's through the null hypothesis as stated in equation (1).

Testing the validity of pricing models has always been essential to the asset pricing theory and practice. Jensen (1968) suggested a validity test based on standardized t-statistics using ordinary least squares regression for each asset. Gibbons, Ross, and Shanken (1989) proposed a multivariate F-test under the assumptions that the errors follow a normal distribution. Pesaran and Yamagata (2012) proposed a quadratic-form test statistic based on an adaptive thresholding estimator of the error covariance matrix. Fan, Liao, and Yao (2015) introduced the power enhancement test and applied it to the factor pricing model. The power enhancement test adds a power enhancement component $J_0 \geq 0$ to an asymptotically pivotal statistic constructed from the Wald test statistic, denoted by J_1. The proposed power enhancement statistic J_0 of Fan, Liao, and Yao (2015) is determined through the pivotal statistic J_1, and the power is improved via the contributions of sparse alternatives that survive the screening process. In this paper I propose a new power enhancement component J_n, of which the screening process is done only on the sparse vector $\boldsymbol{\theta_S}$ instead of on the whole vector $\boldsymbol{\theta}$.

In addition to studying the factor pricing model, another example to study is a cross-sectional independence in mixed effect panel data models:

$$y_{it} = \gamma_i + \boldsymbol{\zeta_i} * \boldsymbol{x_{it}} + \mu_i + u_{it} \quad \text{for} \quad i = 1, \dots, n \quad t = 1, \dots, T. \tag{3}$$

For this model, the cross-sectional independence is tested by the null hypothesis:

$$H_0 : \rho_{ij} = 0 \quad \text{for all} \quad i \neq j, \tag{4}$$

where ρ_{ij} denotes the correlation between u_{it} and u_{jt}. Therefore, the n $\times$ n covariance matrix Σ_u of u_{it} is diagonal under the null hypothesis. In the literature, most of the testing statistics for the mixed effect panel data models are based on the sum of squared residual correlations (Baltagi, Feng, and Kao 2012).

The purpose of this paper is to answer the following research question: "Does an empowerment only on the sparse alternatives lead to an improvement of the power enhancement testing?"

The remainder of this paper is organized as follows. Section 2 formulates the new power enhancement component. Section 3 explains how to select $\boldsymbol{\theta_S}$. Section 4 discusses applications of the power enhancement components. Section 5 describes the simulations using the power enhancement component in the paper of Fan, Liao, and Yao (2015) and the new power enhancement component in this research, respectively. Section 6 discusses the comparisons of the results. Section 7 provides the conclusions inferred by this research together with suggestions for further research.

2 Power enhancement components

The power enhancement technique considers the hypothesis testing problem of $H_0 : \boldsymbol{\theta} = \mathbf{0}$ against sparse alternatives. Literature has shown that traditional tests, such as the Wald test, have a low power. To enhance the power, Fan, Liao, and Yao (2015) introduced a power enhancement component which is zero under the null hypothesis with high probability and diverges quickly under sparse alternatives. Their power enhancement test has the form of $J = J_0 + J_1$, where J_1 is a test statistic that has a correct asymptotic size that may suffer from low powers under sparse alternatives. The power enhancement component J_0 augments the test and has to be chosen such that it satisfies the following power enhancement properties (a)-(c):

(a) J_0 is non-negative.

(b) No size distortion after adding J_0: under H_0, $\mathrm{P}(J_0 = 0 \,|\mathrm{H}_0) \to 1$.

(c) Power enhancement: J_0 diverges in probability under some specific regions of alternatives H_α.

J_0 is constructed by a screening procedure. The screening set S_0 screens out most of the estimation noises so that it contains only a few indices of the non-zero entries and is defined as follows:

$$S_0 = \{j : |\hat{\theta}_j| > \hat{v}_j^{1/2} * \delta_{N,T}, \quad j = 1, \ldots, N\}, \tag{5}$$

where $\hat{v}_j$ is a data-dependent normalizing constant that is taken as the estimated asymptotic variance of $\hat{\theta}_j$ and $\delta_{N,T}$ is the critical value that depends on (N,T). $\delta_{N,T}$ is chosen to be slightly larger than the noise level $\max_{j \le N} |\hat{\theta}_j - \theta_j|/\hat{v}_j^{1/2}$, specifically:

$$\inf_{\theta \in \Theta} P\left(\max_{j \le N} |\hat{\theta}_j - \theta_j|/\hat{v}_j^{1/2} < \delta_{N,T}\right) \to 1. \tag{6}$$

Fan, Liao, and Yao (2015) proposed $\delta_{N,T}$ for the factor model as follows:

$$\delta_{N,T} = log(logT)\sqrt{log(N)}. \tag{7}$$

The screening statistic J_0 is then defined as:

$$J_0 = \sqrt{N} \sum_{j \in S_0} \hat{\theta}_j^2 * \hat{v}_j^{-1}. \tag{8}$$

J_0 can then be added to another test statistic with an accurate asymptotic size J_1, so that the constructed power enhancement test takes the form $J = J_0 + J_1$. Fan, Liao, and Yao (2015) choose J_1 as the standardized Wald statistic:

$$J_1 = \frac{\hat{\theta}' \widehat{var}(\hat{\theta})^{-1} \hat{\theta} - N}{\sqrt{2N}}, \tag{9}$$

Consequently, J is N(0,1) asymptotically distributed under the null hypothesis. Because of equation (5) and (6), J_0 satisfies the non-negativeness and no-size-distortion properties. Under $H_0 : \boldsymbol{\theta} = \mathbf{0}$, it holds that:

$$P(J_0 = 0|H_0) = P(\hat{S} = \emptyset|H_0) = P\left(\max_{j \leq N} |\hat{\theta}_j|/\hat{v}_j^{1/2} < \delta_{NT}|H_0\right) \to 1. \tag{10}$$

In this paper, the vector $\boldsymbol{\theta}$ is divided into two sub-vectors: $\boldsymbol{\theta_S}$ and $\boldsymbol{\theta_D}$. $\boldsymbol{\theta_S}$ contains the sparse alternatives and $\boldsymbol{\theta_D}$ contains the remaining elements of $\boldsymbol{\theta}$, which is dense. Fan, Liao, and Yao (2015) empower all the elements of $\boldsymbol{\theta}$. I propose to empower only $\boldsymbol{\theta_S}$, which leads to a different screening set S_n and a different power enhancement component J_n:

$$S_n = \{j : |\hat{\theta}_j| > \hat{v}_j^{1/2} * \delta_{r,N,T}, \quad j = 1, \ldots, N \quad \hat{\theta}_j \in \boldsymbol{\theta_S}\}, \tag{11}$$

$$J_n = \sqrt{N} \sum_{j \in S_n} \hat{\theta}_j^2 * \hat{v}_j^{-1}, \tag{12}$$

where r is the number of elements in $\boldsymbol{\theta_S}$. J_n also satisfies the non-negativeness property (a) and the no size distortion property (b) since J_n is smaller than J_0 used in Fan, Liao, and Yao (2015). This J_n is added to the standardized Wald statistic J_1 in equation (9) so that the constructed power enhancement test takes the form:

$$J = J_1 + J_n. \tag{13}$$

For the factor pricing model, the threshold $\delta_{r,N,T}$ is defined by replacing N in $\delta_{N,T}$ in equation (7) with a linear combination of N and R :

$$\delta_{r,N,T} = log(logT)\sqrt{log(\alpha N + \beta r)} \quad with \quad \alpha + \beta = 1 \quad and \quad \alpha, \beta \geq 0. \tag{14}$$

The first extreme case for $\delta_{r,N,T}$ is when α=1 and β=0. In this case $\delta_{r,N,T}$=$\delta_{N,T}$, the same as Fan, Liao, and Yao (2015) used for the screening procedure for J_0. When $\delta_{r,N,T}$=$\delta_{N,T}$, J_1+J_n rejects the null hypothesis less than J_1+J_0, and therefore lowers the size distortion. By enhancing only $\boldsymbol{\theta_S}$ instead of the entire $\boldsymbol{\theta}$ vector, it is guaranteed that the null hypothesis is rejected less often for the same delta. Moreover, under the alternative, the θ_i's that are big enough to end up in the screening set in equation (5) are mainly in $\boldsymbol{\theta_S}$. The power enhancement on $\boldsymbol{\theta_S}$ by J_0 and J_n is the same. In very few cases, J_0 can be slightly larger than J_n when α=1 and β=0 for $\delta_{r,N,T}$, because J_0 empowers the whole vector $\boldsymbol{\theta}$ and therefore empowers some estimation errors outside the sparse alternatives when testing the alternative hypothesis. However, this difference is very small and can be ignored. The second extreme case is when α=0 and β=1. In this case $\delta_{r,N,T}$=$log(logT)\sqrt{log(r)}$ and under this $\delta_{r,N,T}$, J_1+J_n can reject the null hypothesis more frequently than J_1+J_0 when testing the alternative hypothesis and hence enhances the power.

However, this $\delta_{r,N,T}$ can lead to huge size distortion. In some of the cases when r<<N, $\delta_{r,N,T}$ can be much smaller than $\delta_{N,T}$ and J_n could empower some estimation errors. Therefore, there exists an optimum $\alpha \geq 0$ and $\beta \geq 0$ in equation (14) for J_n. With the optimum α and β, the proposed power enhancement not only reduces the size distortion under the null hypothesis, but also provides more power under the alternative hypothesis. In this research, I intuitively choose $\alpha = 0.25$ and $\beta = 0.75$.

3 Selection of $\boldsymbol{\theta_S}$

As mentioned in Section 2, in this paper the vector $\boldsymbol{\theta}$ is divided into two sub-vectors: $\boldsymbol{\theta_S}$ and $\boldsymbol{\theta_D}$. How to select the sub-vector $\boldsymbol{\theta_S}$ for both the factor pricing model and the cross-sectional independence model is discussed in this section.

For the factor pricing model, $\boldsymbol{\theta}$ is a vector of intercepts for all financial assets. Therefore the elements of $\boldsymbol{\theta_S}$ are the intercepts of the financial assets that might not be equal to zero.

For the cross-sectional model, $\boldsymbol{\theta_S}$ is a vector consisting of the correlations between stocks. Thus $\boldsymbol{\theta_S}$ consists of some entries of the sparse matrix Σ_u. There are many different ways to select $\boldsymbol{\theta_S}$ according to the applications. For example, $\boldsymbol{\theta_S}$ can be selected based on the regularity condition proposed in Fan, Liao, and Yao (2015). They use m_N and D_N to characterize the used sparse matrices Σ_u in the cross-sectional independence model:

$$m_N = \sum_{j=1}^{N} I\{(\Sigma_u)_{ij} \neq 0\}, \qquad D_N = \sum_{i \neq j} I\{(\Sigma_u)_{ij} \neq 0\}, \tag{15}$$

where m_N represents the maximum number of non-zeros in each row, and D_N represents the total number of nonzero off-diagonal entries. Suppose $N^{1/2}(logN)^{\gamma} \in$ O(T), where γ is a constant bigger than 2, and suppose $min_{(\Sigma_u)_{ij} \neq 0} |(\Sigma_u)_{ij}| >> \sqrt{(logN)/T}$, then one of the following two cases holds:

1.$D_N \in O(N^{1/2})$ and $m_N \in O\left(\frac{T}{N^{1/2} * log(N)^{\gamma}}\right)$.

2.$D_N \in O(N)$ and $m_N \in O(1)$.

In the first case, Σ_u is required to have no more than $O(N^{1/2})$ off-diagonal nonzero entries, but allows a diverging m_N, which represents the maximum number of non-zeros in each row. Moreover, there are only a small portion of firms whose individual shocks are correlated with many other firms. In this case, the elements of $\boldsymbol{\theta_S}$ consist of the correlations of those small portion of firms that are correlated with many other firms.

In the second case, Σ_u can have $O(N)$ off-diagonal nonzero entries, but m_N should be bounded.

This case is typical for firms whose individual shocks are correlated only within industries but not across industries. Those correlations are the elements for the sub-vector $\boldsymbol{\theta_S}$, which allows block-diagonal matrices with finite size of blocks or banded matrices with finite number of bands. Thus, for example, the off-diagonal entries of the block-diagonal matrices or the non-zero elements along a band can be chosen as elements for $\boldsymbol{\theta_S}$ in this case.

4 Application of the models

4.1 Factor pricing model

A factor pricing model is a financial model which uses multiple factors to analyze and explain asset prices:

$$y_{it} = \theta_i + \boldsymbol{b_i} * \boldsymbol{f_t} + u_{it} \quad \text{for} \quad i = 1, \ldots, N \quad t = 1, \ldots, T, \tag{16}$$

where N = dim(θ) is allowed to grow faster than the sample size T, y_{it} denotes the excess return of the ith asset at time t, θ_i is the intercept of asset i, $\boldsymbol{f_t}$ is the K-dimensional observable factors, $\boldsymbol{b_i}$ is a vector of factor loadings and u_{it} represents the idiosyncratic error. I am interested in testing whether the factor pricing model is consistent with empirical data through the following null hypothesis:

$$H_0 : \boldsymbol{\theta} = \mathbf{0}, \tag{17}$$

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)'$ is the vector of intercepts for all N financial assets.

To test this null hypothesis, I use two power enhancement tests. The first test has the form of $J = J_1 + J_0$ as in Fan, Liao, and Yao (2015) and the second test is the one proposed in this research with the form of $J = J_1 + J_n$ with J_1 as follows:

$$J_1 = \frac{a_{f,t} T \hat{\theta}' \hat{\Sigma}_u^{-1} \hat{\theta} - N}{\sqrt{2N}}, \tag{18}$$

where $a_{f,t} > 0$ is a constant and Σ_u is the N $\times$ N estimated covariance matrix of $u_t = (u_{1t}, ..., u_{N,t})$. Σ_u is estimated according to the threshold approach of Bickel and Levina (2008). The estimator of the covariance matrix is defined as:

$$\widehat{(\Sigma_u)}_{ij} = \begin{cases} s_{ii}, & if \quad i = j \\ h_{ij}(s_{ij}), & if \quad i \neq j \end{cases} \tag{19}$$

where $s_{ij} = \dfrac{1}{T} \sum_{t=1}^{T} \widehat{u_{it}} \widehat{u_{jt}}$ and $h_{ij}(s_{ij}) = s_{ij} I\{s_{ij} > C \left(s_{ii} s_{jj} \dfrac{logN}{T} \right)^{1/2}\}$ for some constant $C > 0$.

$a_{f,t}$ of equation (18) is defined as follows:

$$a_{f,t} = 1 - \bar{\boldsymbol{f}}' \boldsymbol{w}, \tag{20}$$

where $\bar{\boldsymbol{f}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}$, $\boldsymbol{w} = \left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\boldsymbol{f_t}'\right)\bar{\boldsymbol{f}}$. The OLS estimator of $\boldsymbol{\theta}$ can then be expressed as:

$$\hat{\theta}_j = \frac{1}{T * a_f, t}\sum_{t=1}^{T} y_{it}(1 - \boldsymbol{f_t}'\boldsymbol{w}). \tag{21}$$

When $\mathrm{cov}(\boldsymbol{f_t})$ is positive definite and there are no serial correlations, the conditional variance of $\hat{\theta}_j$ converges in probability to $v_j = var(u_{jt})/T * a_f$, with $a_f = 1 - E\boldsymbol{f_t}'(E\boldsymbol{f_t}\boldsymbol{f_t}')^{-1}E\boldsymbol{f_t}$. v_j can be estimated with the residuals of OLS estimator:

$$\hat{v}_j = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{jt}^2/T * a_{f,t}, \tag{22}$$

where $\hat{u}_{jt} = y_{jt} - \hat{\theta}_j - \hat{\boldsymbol{b}}_{\boldsymbol{j}}\boldsymbol{f_t}$. The power enhancement components J_0 and J_n augment the test J_1 and are constructed by a screening procedure as described in Section 2.

4.2 Cross-sectional independence model

Cross-sectional dependence is one of the most important diagnostics that a researcher should investigate before performing a panel data analysis. Hence, a study is also performed for the power enhancement components for cross-sectional independence in mixed effect panel data models:

$$y_{it} = \gamma_i + \boldsymbol{\zeta_i} * \boldsymbol{x_{it}} + \mu_i + u_{it} \quad \text{for} \quad i = 1, \ldots, n \quad t = 1, \ldots, T, \tag{23}$$

where y_{it} denotes the excess return of the ith asset at time t, $\boldsymbol{x_{it}}$ is the regressor, μ_i is the random effect and u_{it} is the idiosyncratic error. $\boldsymbol{x_{it}}$ could be correlated with the random effect μ_i but uncorrelated with u_{it}. The cross-sectional independence is tested by the null hypothesis:

$$H_0 : \rho_{ij} = 0 \quad \text{for all} \quad i \neq j, \tag{24}$$

where ρ_{ij} denotes the correlation between u_{it} and u_{jt}. This null hypothesis is equivalent to testing $H_0 : \boldsymbol{\theta} = \boldsymbol{0}$ with $\boldsymbol{\theta} = (\rho_{12}, ..., \rho_{1n}, \rho_{23}, ..., \rho_{2n}, ..., \rho_{n-1,n})$, a Nx1 matrix where N=n(n-1)/2. ρ_{ij} is estimated using the following: $\widetilde{y_{it}} = y_{it} - \sum_{t=1}^{T} y_{it}$, $\widetilde{x_{it}} = x_{it} - \sum_{t=1}^{T} x_{it}$ and $\widetilde{u_{it}} = u_{it} - \sum_{t=1}^{T} u_{it}$. Then $\widetilde{y_{it}} = \boldsymbol{\zeta_i} * \widetilde{\boldsymbol{x_{it}}} + \widetilde{u_{it}}$, so that $\widehat{\boldsymbol{\zeta_i}}$ can be estimated by OLS regression of $\widetilde{y_{it}}$ on $\widetilde{\boldsymbol{x_{it}}}$, which leads to the estimated residual $\widehat{u_{it}} = \widetilde{y_{it}} - \boldsymbol{\zeta_i} * \widetilde{\boldsymbol{x_{it}}}$. Using $\widehat{u_{it}}$, the estimations of $\widehat{\sigma_{ij}}$ and $\widehat{\rho_{ij}}$ are respectively as follows:

$$\widehat{\sigma_{ij}} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt}, \tag{25}$$

$$\widehat{\rho_{ij}} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_{ii}^{1/2}\hat{\sigma}_{ij}^{1/2}}. \tag{26}$$

The estimated asymptotic variance of $\widehat{\rho_{ij}}$ is given as:

$$\widehat{v_{ij}} = \left(1 - \widehat{\rho_{ij}}^2\right)^2/T. \tag{27}$$

Therefore, the screening set S_0 and J_0 are as follows:

$$S_0 = \{(i,j) : |\hat{\rho}_{ij}| > \hat{v}_{ij}^{1/2} * \delta_{N,T}, \quad i < j \leq n\}, \tag{28}$$

$$J_0 = \sqrt{N} \sum_{(i,j) \in S_0} \hat{\rho}_{ij}^2 * \hat{v}_j^{-1}. \tag{29}$$

In the tests for the cross-sectional independence model, Fan, Liao, and Yao (2015) used a different $\delta_{N,T} = 2.25log(N)(log(log(T)))^2$. The screening set S_n and the power enhancement component J_n for the cross-sectional independence model are modified accordingly as follows:

$$S_n = \{(i,j) : |\hat{\rho}_{ij}| > \hat{v}_{ij}^{1/2} * \delta_{N,T}, \quad i < j \leq n, \quad \hat{\rho}_{ij} \in \boldsymbol{\theta_S}\}, \tag{30}$$

$$J_n = \sqrt{N} \sum_{(i,j) \in S_n} \hat{\rho}_{ij}^2 * \hat{v}_{ij}^{-1}, \tag{31}$$

where $\delta_{r,N,T}$ is chosen according to $\delta_{N,T}$ for the cross-sectional independence model by replacing N with a linear combination of N and r as in the factor pricing model:
$\delta_{r,N,T} = 2.25log(logT)^2 log(\alpha N + \beta r)$ with α+β=1, α ≥0, β ≥0 and r=$|\boldsymbol{\theta_S}|$. Here $\boldsymbol{\theta_S}$ is the subset of $\hat{\rho}_{ij}$'s as described in Section 3. The quadratic statistic used in this research for this model is from Baltagi, Feng, and Kao (2012):

$$J_1 = \sqrt{\frac{1}{n(n-1)}} \sum_{i<j} (T\rho ij - 1) - \frac{n}{2(T-1)}. \tag{32}$$

5 Simulation

In this research, Monte Carlo simulations are used in order to examine the finite sample performance of the power enhancement tests. The simulations are replicating the simulations done in the paper of Fan, Liao, and Yao (2015) and can be obtained via running the code provided in the GitHub folder (GitHub 2021).

For the factor model in equation (16) $\{b_i\}_1^N$, $\{f_t\}_1^N$ and $\{u_t\}_1^N$ are simulated independently and respectively from $N_3(\mu_b, \Sigma_b)$, $N_3(\mu_f, \Sigma_f)$ and $N_N(0, \Sigma_u)$. $\Sigma_u = diag\{A_1, ..., A_{N/4}\}$ is a block-diagonal correlation matrix, where each diagonal block A_j is a 4 × 4 positive definite matrix, whose correlation matrix has off-diagonal entry ρ_j, generated from U(0, 0.5). The parameters are calibrated using daily returns of S&P 500's top 100 constituents for the period from July 1st, 2008 to June 29th, 2012 and can be found in Table 1.

Table 1: Parameters used to generate b_i and f_t

μ_B	Σ_b			μ_f	Σ_f		
0.9833	0.0921	-0.0178	0.0436	0.0260	3.2351	0.1783	0.7783
-0.1233	-0.0178	0.0862	-0.0211	0.0211	0.1783	0.5069	0.0102
0.0839	0.0436	-0.0211	0.7624	-0.0043	0.7783	0.0102	0.6586

The powers of the tests are evaluated under two specific alternatives:

$$\text{Sparse alternative}: H_\alpha^1 : \theta_i = \begin{cases} 0.3, & i \leq N/T \\ 0, & i > N/T \end{cases} \tag{33}$$

$$\text{Weak alternative}: H_\alpha^2 : \theta_i = \begin{cases} \sqrt{\frac{logN}{T}}, & i \leq N^{0.4} \\ 0, & i > N^{0.4} \end{cases} \tag{34}$$

For the factor model, $\boldsymbol{\theta_S}$ is chosen as the sub-vector containing the first 1.1*N/T elements of $\boldsymbol{\theta}$ = $(\theta_1, ..., \theta_n)'$ to include the sparse alternatives. For the weak alternative, $\boldsymbol{\theta_S}$ is chosen as the sub-vector containing the first $N^{0.5}$ elements of $\boldsymbol{\theta}$ to include the weak alternatives. For both alternatives, a little bit more θ_i's are added to the sub-vector $\boldsymbol{\theta_S}$ to make sure all the θ_i's which are not equal to zero are included under the assumption that sparse alternatives are not exactly known. Because of the different selections for the two alternatives, J_n will make use of different sub-vectors $\boldsymbol{\theta_S}$. Therefore, H_0^1 denotes the null hypothesis against the sparse alternative H_α^1 and H_0^2 denotes the null hypothesis against the weak alternative H_α^2.

For the cross-sectional independence model in equation (23), $x_{it} = 0.5$ is initialized at t = 1 for each i. μ_i is drawn from N(0, 0.25) for i = 1,...,n. The parameters γ and ζ are set to be -1 and 2, respectively and $\{u_t\}_1^N$ is generated from $N_N(0, \Sigma_u)$. Under the null hypothesis, Σ_u is set to be a diagonal matrix $\Sigma_{u,0} = diag\{\sigma_1, ..., \sigma_n\}$. The heteroskedastic errors are as follows:

$$\sigma_i^2 = \sigma^2(1 + k\bar{x}_i)^2, \tag{35}$$

where k=0.5, $\bar{x}_i$ is the average of x_{it} across t and σ^2 is scaled in order to fix the average of σ_i^2 at 1. Under the alternative hypothesis, Σ_u=$\Sigma_{u,0}^{1/2}\Sigma_{u,1}\Sigma_{u,0}^{1/2}$. I start with $\Sigma_{u,1}$=$diag\{\Sigma_1, ..., \Sigma_{n/4}\}$, where each Σ_i is defined as I_4 initially. Then, $\lfloor n^{0.3} \rfloor$ blocks are randomly chosen among them and made non-diagonal by setting $\Sigma_i(m, n) = \rho^{|m-n|}$ $(m, n \leq 4)$, with $\rho = 0.2$.

Under the alternative hypothesis, the matrix Σ_u of the cross-sectional independence model satisfies the second regularity condition described in Section 3. $\boldsymbol{\theta_S}$ is selected to include all the strictly upper triangular entries of each of the diagonal block matrices A_j from $\Sigma_u = diag\{A_1, ..., A_{n/4}\}$.

6 Numerical studies

6.1 Factor pricing model

Six testing methods are conducted and compared for the factor model: the standardized Wald test J_1, the thresholding test J_{thr} as in Fan (1996), their power enhancement versions $J_0 + J_1$ and $J_0 + J_{thr}$, and the power enhancement versions $J_n + J_1$ and $J_n + J_{thr}$ proposed in this paper. The relative frequency of the screening sets S_0 and S_n being empty for both power enhancement components, which approximates $P(S_0 = \emptyset)$ and $P(S_n = \emptyset)$, are also calculated. The thresholding test J_{thr} is defined as follows:

$$J_{thr} = \sigma_N^{-1} \left(\sum_{n=1}^{N} \hat{\theta}_j^2 \hat{v}_j^{-1} I\{|\hat{\theta}_j| \hat{v}_j^{-1} > t_N\} - \mu_N \right), \quad (36)$$

where $\sigma_N^2 = \sqrt{2/\pi} a^{-1} t_N^3 (1 + 3t_N^{-2})$, $\mu_N = \sqrt{2/\pi} a^{-1} t_N (1 + t_N^{-2})$, $t_N = \sqrt{2log(Na)}$ and $a = (logN)^{-2}$.

For each test, the relative frequency of rejection of the null hypothesis under H_0^1, H_0^2, H_α^1 and H_α^2 based on 2000 replications is calculated, with significance level q=0.05 for different pairs of (N,T). J_0 is calculated for $\delta_{N,T}$ and J_n is calculated for three different $\delta_{r,N,T}$'s: $\delta_{r,N,T}$=$\delta_{N,T}$, $\delta_{r,N,T}$=$log(logT)\sqrt{log(r)}$ and $\delta_{r,N,T}$=$log(logT)\sqrt{log(0.25N + 0.75r)}$.

Table 2 presents the empirical size and power of each testing method under the null hypothesis H_0^1 and H_0^2, where H_0^1 is the null hypothesis against H_α^1 and H_0^2 is the null hypothesis against H_α^2. Table 3 gives the empirical size and power of each testing method under the two alternative hypotheses H_α^1 and H_α^2 for $\delta_{r,N,T}$= $\delta_{N,T}$=$log(logT)\sqrt{log(N)}$.

Table 2: Size and power (%) of tests for the factor model under H_0^1 and H_α^2 for $\delta_{r,N,T} = log(logT)\sqrt{logN}$

(N,T)	J_1	J_1+J_0	J_1+J_n	J_{thr}	$J_{thr}+J_0$	$J_{thr}+J_n$	$P(S_0=\emptyset)$	$P(S_n=\emptyset)$
H_0^1(null hypothesis against$H_\alpha{}^1$)								
(500, 300)	6.2	6.9	6.3	7.7	8.2	7.7	99.3	100.0
(800, 300)	4.7	4.8	4.7	7.1	7.3	7.1	99.9	100.0
(1000, 300)	5.1	5.3	5.1	7.0	7.1	7.0	99.8	100.0
(1200, 300)	5.3	5.6	5.3	7.7	7.9	7.7	99.7	100.0
(500, 500)	6.2	6.3	6.2	6.9	7.0	6.9	99.9	100.0
(800, 500)	6.1	6.3	6.1	8.1	8.2	8.1	99.7	100.0
(1000, 500)	4.5	4.6	4.5	6.0	6.1	6.0	99.9	100.0
(1200, 500)	4.1	4.3	4.1	6.4	6.5	6.4	99.8	100.0
H_0^2(null hypothesis against$H_\alpha{}^2$)								
(500, 300)	5.6	6.1	5.6	6.1	6.5	6.1	99.5	100.0
(800, 300)	5.2	5.4	5.2	7.0	7.2	7.0	99.8	100.0
(1000, 300)	5.2	5.6	5.2	6.7	6.9	6.7	99.6	100.0
(1200, 300)	5.6	5.7	5.5	5.9	6.1	5.9	99.7	100.0
(500, 500)	4.6	4.7	4.6	5.6	5.8	5.6	99.8	100.0
(800, 500)	5.0	5.1	5.0	6.2	6.3	6.2	99.9	100.0
(1000, 500)	5.1	5.1	5.1	7.3	7.3	7.3	100.0	100.0
(1200, 500)	5.3	5.4	5.3	6.8	6.9	6.8	99.9	100.0

Under H_0^1, the sizes of J_1, J_1+J_0 and J_1+J_n are close to the significance level, while all of the three thresholding tests J_{thr}, $J_{thr}+J_0$ and $J_{thr}+J_n$ have significant size distortions. Adding J_0 gives a maximum of 0.7% size increase, while J_n gives a maximum of only 0.1% size increase. Furthermore, $P(S_0=\emptyset)$ is close to 1 for J_0, indicating that the power enhancement component screens off most of the estimation errors. For J_n, $P(S_n=\emptyset)$=1 for all pairs of (N,T), meaning that J_n screens off all the estimation errors every time. Similar results hold for H_0^2.

Table 3: Size and power (%) of tests for the factor model under the alternative hypotheses for $\delta_{r,N,T} = log(logT)\sqrt{logN}$

(N,T)	J_1	$J_1 + J_0$	$J_1 + J_n$	J_{thr}	$J_{thr} + J_0$	$J_{thr} + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$
H_α^1								
(500, 300)	45.1	94.1	94.1	64.0	94.5	94.5	8.6	8.7
(800, 300)	46.8	94.7	94.7	81.7	96.4	96.4	6.9	6.9
(1000, 300)	43.2	95.1	95.1	77.7	95.9	95.9	6.7	6.7
(1200, 300)	45.5	97.0	97.0	90.4	98.1	98.1	4.0	4.0
(500, 500)	45.5	98.45	98.4	52.4	98.4	98.4	2.1	2.2
(800, 500)	62.4	99.9	99.9	86.2	99.9	99.9	0.2	0.3
(1000, 500)	55.7	99.7	99.7	83.1	99.6	99.6	0.5	0.5
(1200, 500)	51.9	99.7	99.7	80.1	99.6	99.6	0.5	0.5
${H_\alpha}^2$								
(500, 300)	66.8	71.4	71.3	78.6	80.4	80.3	74.4	75.0
(800, 300)	65.6	70.6	70.55	82.2	83.8	83.8	74.8	74.9
(1000, 300)	68.3	73.3	73.3	87.1	88.4	88.4	74.5	74.8
(1200, 300)	67.1	71.2	71.2	88.0	88.8	88.0	74.3	74.3
(500, 500)	67.1	70.3	70.3	78.0	79.3	79.3	81.8	81.9
(800, 500)	73.0	75.8	75.8	82.9	83.9	82.9	81.8	81.9
(1000, 500)	76.1	78.5	78.5	86.4	86.9	86.9	82.6	82.7
(1200, 500)	77.7	80.3	80.25	88.2	88.8	88.8	83.45	83.5

The main findings of Table 3 are as follows:

1). Under H_α^1, the power of the thresholding test is much higher than that of the Wald test, as the Wald test accumulates too many estimation errors. Moreover, the power is significantly enhanced after J_0 and J_n are added. There is thus no significant difference in power between J_0 and J_n. Under H_α^1, $P(S_0 = \emptyset)$ and $P(S_n = \emptyset)$ are the same for almost all the pairs. The frequency of S_0 and S_n being empty is less than 9% for all pairs of (N,T) and sometimes even smaller than 1%. The 9% is because the screening procedure manages to capture the big θ's.

2). Under H_α^2, the thresholding test has higher power than $J_1 + J_0$ and $J_1 + J_n$, making both power enhancement components not substantial. The screening set of both statistics has a large chance of being empty, since the θ's are weak under this alternative.

Thus as expected, $J_1{+}J_n$ rejects the null hypothesis less than $J_1{+}J_0$ and $J_{thr}{+}J_n$ rejects the null hypothesis less than $J_{thr}{+}J_0$. Therefore, J_n lowers the size distortion with very little loss of power under the alternative hypotheses.

Table 4 presents the empirical size and power of each testing method under H_0^1 and H_0^2 and Table 5 displays the empirical size and power of each testing method under the two alternative hypothesis for $\delta_{r,N,T}$=$log(logT)\sqrt{log(r)}$.

Table 4: Size and power (%) of tests for the factor model under H_0^1 and H_α^1 for $\delta_{r,N,T}$ = $log(logT)\sqrt{log(r)}$

(N,T)	J_1	$J_1 + J_0$	$J_1 + J_n$	J_{thr}	$J_{thr} + J_0$	$J_{thr} + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$
H_0^1								
(500, 300)	5.1	5.5	19.0	6.7	6.9	20.7	99.6	85.1
(800, 300)	5.6	5.9	17.9	5.9	6.2	18.3	99.7	87.1
(1000, 300)	5.9	6.2	23.5	7.4	7.7	24.3	99.7	81.5
(1200, 300)	5.9	6.2	20.0	7.5	7.7	21.0	99.7	85.3
(500, 500)	5.2	5.4	99.3	7.7	7.9	99.5	99.7	0.0
(800, 500)	4.8	5.0	16.7	6.7	6.9	18.2	99.8	87.5
(1000, 500)	5.6	5.8	27.1	7.1	7.2	28.9	99.8	76.7
(1200, 500)	4.4	4.6	27.0	5.7	5.9	28.3	99.8	76.3
${H_0}^2$								
(500, 300)	6.5	7.0	18.6	6.9	7.4	19.1	99.4	86.6
(800, 300)	5.8	6.0	16.1	7.1	7.3	16.5	99.7	88.9
(1000, 300)	5.7	6.1	16.0	6.8	7.2	16.8	99.6	88.9
(1200, 300)	5.9	6.4	15.9	6.6	7.1	16.3	99.5	89.2
(500, 500)	5.4	5.7	13.1	7.8	8.1	14.8	99.7	91.6
(800, 500)	5.3	5.4	12.1	7.05	7.1	13.8	99.9	92.2
(1000, 500)	4.5	4.6	10.9	7.15	7.2	13.2	99.9	93.1
(1200, 500)	5.2	5.3	11.5	6.6	6.7	12.8	99.9	92.3

Under both H_0^1 and H_0^2, the sizes of J_1 and $J_1 + J_0$ are close to the significance level, while $J_1 + J_n$ has enormous size distortion because of the chosen $\delta_{r,N,T}$. Furthermore, all of the three thresholding tests J_{thr}, $J_{thr} + J_0$ and $J_{thr} + J_n$ have significant size distortions as well. Adding J_0 gives a maximum of 0.7% size increase, while J_n gives a big size increase for each pair (N,T), even more than 94% in the case of (500, 500). Furthermore, $P(S_0 = \emptyset)$ is close to one for J_0, indicating that the power enhancement component screens off most of the estimation errors. For J_n, $P(S = \emptyset)$ is equal to 80∼90% for most pairs of (N,T), except for (500, 500), meaning that J_n screens off majority of the estimation errors most of the time.

Table 5: Size and power (%) of tests for the factor model under the alternative hypotheses for $\delta_{r,N,T} = log(logT)\sqrt{log(r)}$

(N,T)	J_1	$J_1 + J_0$	$J_1 + J_n$	J_{thr}	$J_{thr} + J_0$	$J_{thr} + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$
H_α^1								
(500, 300)	44.8	94.0	100.0	63.4	94.7	100.0	8.4	0.0
(800, 300)	47.8	96.9	100.0	81.6	97.9	100.0	4.5	0.0
(1000, 300)	44.9	95.4	100.0	76.5	96.0	100.0	6.4	0.0
(1200, 300)	43.3	96.6	100.0	90.2	97.9	100.0	4.6	0.0
(500, 500)	47.1	98.5	100.0	51.5	98.0	100.0	2.3	0.0
(800, 500)	62.2	99.8	100.0	86.8	99.9	100.0	0.3	0.0
(1000, 500)	55.7	99.5	100.0	83.7	99.5	100.0	0.8	0.0
(1200, 500)	52.1	99.8	100.0	80.0	99.8	100.0	0.3	0.0
${H_\alpha}^2$								
(500, 300)	65.6	70.3	99.1	77.7	79.1	99.2	75.4	1.0
(800, 300)	66.0	70.9	99.8	82.1	83.4	99.7	74.2	0.3
(1000, 300)	66.8	70.9	99.7	85.4	86.3	99.7	74.5	0.4
(1200, 300)	69.5	74.8	99.8	87.2	88.5	99.8	73.9	0.3
(500, 500)	67.3	69.9	98.8	78.0	79.0	98.9	83.1	1.6
(800, 500)	71.1	73.9	99.4	82.5	83.2	99.3	83.1	0.9
(1000, 500)	76.6	78.7	99.7	86.5	87.0	99.6	83.7	0.7
(1200, 500)	78.4	80.1	99.4	86.6	87.3	99.3	83.2	1.0

The main findings of Table 5 are as follows:

1). Under H_α^1, the power is significantly enhanced after J_0 is added. The power is even more enhanced after J_n is added to both test statistics. J_1+J_n and J_{thr}+J_n even have a power of 100% under H_α^1, which means that the null hypothesis always gets rejected under this alternative. Therefore, $P(S_n = \emptyset)$=0.0 for all pairs of (N,T). Under H_α^1, $P(S_0 = \emptyset)$ is less then 9% for all pairs of (N,T) and sometimes even smaller than 1%.

2). Under H_α^2, the thresholding test has higher power than $J_1 + J_0$, making J_0 not substantial. The emptiness of the screening set S_0 is around 80% and therefore has a large chance of being empty, since the θ's are weak under this alternative. $J_1 + J_n$ almost has a power of 100% again for both the statistics and therefore the emptiness of its screening set S_n is close to zero.

Thus, under $\delta_{r,N,T} = log(logT)\sqrt{log(r)}$, J_n empowers the Wald statistic and threshold statistic much better than J_0. However, it suffers from huge size distortion.

Table 6 presents the empirical size and power of each testing method under H_0^1 and H_0^2 and Table 7 displays the empirical size and power of each testing method under the two alternative hypothesis for $\delta_{r,N,T} = log(logT)\sqrt{log(0.25N + 0.75r)}$.

Table 6: Size and power (%) of tests for the factor model under H_0^1 and H_α^2 for $\delta_{r,N,T} = log(logT)\sqrt{log(0.25N + 0.75r)}$

(N,T)	J_1	$J_1 + J_0$	$J_1 + J_n$	J_{thr}	$J_{thr} + J_0$	$J_{thr} + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$
H_0^1								
(500, 300)	4.9	5.4	4.9	6.7	7.3	6.7	99.3	100.0
(800, 300)	5.9	6.2	5.9	7.5	7.7	7.5	99.7	100.0
(1000, 300)	4.5	5.0	4.5	6.7	6.9	6.7	99.4	100.0
(1200, 300)	5.0	5.6	5.0	7.1	7.6	7.1	99.4	100.0
(500, 500)	5.3	5.5	5.3	6.0	6.2	6.0	99.7	100.0
(800, 500)	5.3	5.5	5.3	6.8	7.0	6.8	99.8	100.0
(1000, 500)	5.3	5.4	5.3	5.9	6.1	5.9	99.8	100.0
(1200, 500)	6.1	6.2	6.1	6.4	6.5	6.4	99.9	100.0
${H_0}^2$								
(500, 300)	5.2	5.6	5.3	7.2	7.6	7.4	99.5	99.8
(800, 300)	5.3	5.7	5.4	6.9	7.2	7.0	99.6	99.9
(1000, 300)	6.2	6.5	6.3	7.5	7.7	7.6	99.6	99.9
(1200, 300)	5.5	5.7	5.65	7.6	7.8	7.75	99.75	99.8
(500, 500)	6.1	6.3	6.1	6.7	6.8	6.7	99.8	100.0
(800, 500)	4.5	4.6	4.5	6.0	6.1	6.0	99.9	100.0
(1000, 500)	6.7	6.8	6.75	7.1	7.2	7.15	99.9	99.95
(1200, 500)	4.7	4.8	4.75	5.8	5.9	5.8	99.9	99.95

Under both H_0^1 and H_0^2, the sizes of J_1, $J_1 + J_0$ and $J_1 + J_n$ are close to the significance level, while all of the three thresholding tests J_{thr}, $J_{thr} + J_0$ and $J_{thr} + J_n$ have significant size distortions. Under H_0^1, adding J_0 gives a maximum of 0.6% size increase, while J_n results in 0% size increase. Furthermore, $P(S_0 = \emptyset)$ is close to one for J_0, indicating that the power enhancement component screens off most of the estimation errors. For J_n, $P(S_n = \emptyset)$=1 for all pairs of (N,T), meaning that J_n screens off all the estimation errors all the time. Under H_0^2, adding J_0 also gives a maximum of 0.6% size increase, while J_n results in a maximum of 0.2% size increase. For H_0^2, $P(S = \emptyset)$ is close to one for J_0 and J_n, indicating that both the power enhancement components screen off most of the estimation errors.

Table 7: Size and power (%) of tests for the factor model under the alternative hypotheses for $\delta_{r,N,T} = log(logT)\sqrt{log(0.25N + 0.75r)}$

(N,T)	J_1	$J_1 + J_0$	$J_1 + J_n$	J_{thr}	$J_{thr} + J_0$	$J_{thr} + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$
H_α^1								
(500, 300)	45.2	94.5	98.2	62.3	94.5	98.1	7.8	2.4
(800, 300)	47.5	95.7	99.3	81.7	96.7	99.3	6.0	1.1
(1000, 300)	43.2	94.4	98.9	76.5	95.2	99.0	7.6	1.4
(1200, 300)	41.6	96.7	99.3	91.2	98.3	99.6	4.6	0.9
(500, 500)	47.0	98.9	99.7	51.0	98.8	99.7	1.4	0.4
(800, 500)	60.8	99.9	100.0	86.3	99.9	100.0	0.2	0.0
(1000, 500)	55.9	99.7	99.9	83.8	99.8	99.95	0.6	0.2
(1200, 500)	50.6	99.8	99.95	79.9	99.7	99.9	0.6	0.2
${H_\alpha}^2$								
(500, 300)	65.6	70.2	78.7	77.7	78.8	83.4	74.6	44.5
(800, 300)	67.7	72.3	80.8	82.5	84.3	87.7	74.3	42.9
(1000, 300)	67.0	72.3	80.6	85.5	86.4	88.9	73.7	43.1
(1200, 300)	69.8	74.5	83.2	87.9	89.1	91.8	74.7	41.7
(500, 500)	65.7	68.8	76.1	77.0	78.1	81.2	83.0	54.4
(800, 500)	71.9	74.4	80.4	81.6	82.6	85.0	82.9	54.8
(1000, 500)	78.1	80.3	85.4	87.3	87.7	89.7	82.4	54.6
(1200, 500)	78.3	80.1	85.5	88.3	88.9	90.6	82.8	53.7

The main findings of Table 7 are as follows:

1). Under H_α^1, J_n enhances the power even more than J_0. The frequency of S_0 being empty is less than 8% and the frequency of S_n being empty is less than 2.5% for all pairs of (N,T).

2). Under H_α^2, the thresholding test has higher power than $J_1 + J_0$ and $J_1 + J_n$, making both power enhancement components not substantial. Both screening sets S_0 and S_n have a large chance of being empty, since the θ's are weak under this alternative. $P(S_n = \emptyset)$ is considerably smaller than $P(S_0 = \emptyset)$, meaning J_n enhances the tests more than J_0.

Thus, using $\delta_{r,N,T} = log(logT)\sqrt{log(0.25N + 0.75r)}$, $J_1 + J_n$ is not only smaller than $J_1 + J_0$ under both H_0^1 and H_0^2, meaning it has smaller size distortion, but also bigger than $J_1 + J_0$ under H_α^1 and H_α^2, meaning it rejects the null hypothesis more frequently under both alternatives. Therefore, J_n enhances the power of J_1 more compared to J_0. The same holds for the threshold statistic.

6.2 Cross-sectional independence model

The Monte Carlo simulations for the cross-sectional independence model are conducted for different pairs of (n,T) with significance level q=0.05 based on 2000 replications. For this model, J_0 is also calculated for $\delta_{N,T}$ and J_n is calculated for three different $\delta_{r,N,T}$'s: $\delta_{r,N,T} = 2.25log(logT)^2log(N)$, $\delta_{r,N,T} = 2.25log(logT)^2log(r)$ and $\delta_{r,N,T} = 2.25log(logT)^2log(0.25N + 0.75r)$. The quadratic test J_1 and the power enhancement test J_1+J_0 and J_1+J_n are performed. Besides, the frequency of the screening set being empty for both power enhancement test are also conducted. The results for $\delta_{r,N,T} = \delta_{N,T}$=$2.25log(logT)^2log(N)$ can be found in Table 8.

Table 8: Size and power (%) of tests for the cross-sectional independence model for $\delta_{r,N,T} = 2.25log(logT)^2log(N)$

(N,T)	J_1	$J_1 + J_0$	$J_1 + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$	J_1	$J_1 + J_0$	$J_1 + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$
H_0						H_α				
(200, 100)	6.0	6.0	6.0	100.0	100.0	25.6	93.6	93.55	7.05	7.1
(200, 200)	4.8	4.8	4.8	100.0	100.0	59.4	97.3	97.3	2.9	2.9
(200, 300)	5.5	5.5	5.5	100.0	100.0	78.8	98.6	98.6	1.8	1.8
(200, 500)	5.2	5.2	5.2	100.0	100.0	93.3	99.6	99.6	0.6	0.6
(400, 100)	4.8	4.9	4.8	99.9	100.0	18.6	98.1	98.1	2.1	2.1
(400, 200)	4.6	4.6	4.6	100.0	100.0	41.6	99.5	99.5	0.6	0.6
(400, 300)	5.3	5.3	5.3	100.0	100.0	65.8	99.95	99.95	0.1	0.1
(400, 500)	5.1	5.1	5.1	100.0	100.0	90.8	99.95	99.95	0.05	0.05
(600, 100)	4.6	4.6	4.6	99.95	100.0	12.8	97.4	97.4	2.8	2.8
(600, 200)	4.9	4.9	4.9	100.0	100.0	25.3	98.9	98.9	1.1	1.1
(600, 300)	5.5	5.5	5.5	100.0	100.0	42.5	99.9	99.9	0.2	0.2
(600, 500)	4.9	4.9	4.9	100.0	100.0	72.4	100.0	100.0	0.0	0.0
(800, 100)	5.7	5.7	5.7	100.0	100.0	11.4	98.5	98.5	1.7	1.7
(800, 200)	4.9	4.9	4.9	100.0	100.0	21.0	99.5	99.5	0.6	0.6
(800, 300)	4.6	4.6	4.6	100.0	100.0	35.1	99.9	99.9	0.2	0.2
(800, 500)	5.7	5.7	5.7	100.0	100.0	64.9	100.0	100.0	0.0	0.0

Under H_0, the sizes of J_1, J_1+J_0 and J_1+J_n are close to 5%. Under both the null and alternative hypothesis, the results using J_n and J_0 are almost identical and have no significant difference.

Table 9 shows the size and power of the bias-corrected quadratic test J_1 and those of the power enhancement tests J_1+J_0 and J_1+J_n for $\delta_{r,N,T} = 2.25log(logT)^2log(r)$.

Table 9: Size and power (%) of tests for the cross-sectional independence model for $\delta_{r,N,T} = 2.25log(logT)^2 log(r)$

(N,T)	J_1	$J_1 + J_0$	$J_1 + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$	J_1	$J_1 + J_0$	$J_1 + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$
H_0						H_α				
(200, 100)	4.8	5.0	5.0	99.8	99.8	27.3	94.1	97.5	7.1	3.0
(200, 200)	5.15	5.15	5.2	100.0	99.95	56.7	97.0	99.1	3.5	1.0
(200, 300)	5.2	5.2	5.2	100.0	100.0	77.5	98.9	99.9	1.6	2.0
(200, 500)	4.8	4.8	4.8	100.0	100.0	94.0	99.8	100.0	0.3	0.1
(400, 100)	5.3	5.3	5.6	100.0	99.8	19.1	98.4	99.5	1.7	0.6
(400, 200)	5.2	5.2	5.2	100.0	100.0	41.3	99.3	99.9	1.0	0.2
(400, 300)	5.1	5.1	5.1	100.0	100.0	64.5	99.8	100.0	0.4	0.0
(400, 500)	4.8	4.8	4.8	100.0	100.0	89.8	100.0	100.0	0.0	0.0
(600, 100)	5.4	5.4	5.6	100.0	99.9	12.3	97.8	99.4	2.4	0.7
(600, 200)	4.5	4.5	4.5	100.0	100.0	26.0	98.9	99.9	1.3	0.2
(600, 300)	6.0	6.0	6.0	100.0	100.0	40.9	99.7	100.0	0.3	0.0
(600, 500)	5.2	5.2	5.2	100.0	100.0	71.9	99.9	100.0	0.2	0.0
(800, 100)	4.75	4.8	4.9	99.9	99.9	10.7	98.3	99.6	1.9	0.5
(800, 200)	5.4	5.4	5.4	100.0	100.0	21.8	99.8	99.95	0.3	0.1
(800, 300)	5.1	5.1	5.1	100.0	100.0	35.4	99.8	99.9	0.3	0.1
(800, 500)	3.8	3.8	3.8	100.0	100.0	63.7	100.0	100.0	0.0	0.0

Under H_0, both the power enhancement tests have little distortion of the original size. However, the power of J_n is bigger than the power of J_0 under the alternative hypothesis and therefore J_n enhances J_1 more than J_0.

Table 10 exhibits the size and power of J_1 and the power enhanced tests $J_1{+}J_0$ and $J_1{+}J_n$ for $\delta_{r,N,T} = 2.25log(logT)^2 log(0.25N + 0.75r)$.

Table 10: Size and power (%) of tests for the cross-sectional independence model for $\delta_{r,N,T} = 2.25log(logT)^2log(0.25N + 0.75r)$

(N,T)	J_1	$J_1 + J_0$	$J_1 + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$	J_1	$J_1 + J_0$	$J_1 + J_n$	$P(S_0 = \emptyset)$	$P(S_n = \emptyset)$
H_0						H_α				
(200, 100)	4.4	4.5	4.4	99.9	100.0	26.0	92.6	93.9	7.8	6.5
(200, 200)	5.7	5.7	5.7	100.0	100.0	57.1	97.4	98.1	3.5	2.7
(200, 300)	4.5	4.5	4.5	100.0	100.0	77.8	99.1	99.5	1.4	0.9
(200, 500)	5.6	5.6	5.6	100.0	100.0	94.9	99.7	99.8	0.5	0.3
(400, 100)	4.4	4.5	4.4	99.9	100.0	19.1	98.2	98.6	2.1	1.6
(400, 200)	5.2	5.2	5.2	100.0	100.0	42.2	99.5	99.6	0.6	0.5
(400, 300)	6.0	6.0	6.0	100.0	100.0	66.1	99.8	99.9	0.2	0.1
(400, 500)	5.3	5.3	5.3	100.0	100.0	90.1	99.95	99.95	0.1	0.1
(600, 100)	5.5	5.6	5.5	99.9	100.0	11.8	97.8	98.4	2.4	1.8
(600, 200)	4.9	4.9	4.9	100.0	100.0	25.1	99.2	99.4	1.0	0.7
(600, 300)	5.0	5.0	5.0	100.0	100.0	43.1	99.8	99.9	0.3	0.2
(600, 500)	4.4	4.4	4.4	100.0	100.0	70.9	99.9	99.9	0.1	0.1
(800, 100)	5.95	6.0	5.95	99.95	100.0	12.2	98.7	99.0	1.5	1.1
(800, 200)	5.0	5.0	5.0	100.0	100.0	23.9	99.55	99.6	0.6	0.5
(800, 300)	4.4	4.4	4.4	100.0	100.0	36.9	99.9	99.9	0.1	0.1
(800, 500)	4.3	4.3	4.3	100.0	100.0	63.1	100.0	100.0	0.0	0.0

Under H_0, the sizes of J_1, J_1+J_0 and J_1+J_n are close to 5%. Both the power enhancement tests J_1+J_0 and J_1+J_n have little to zero distortion of the original size. The size distortion of J_n is even zero for all pairs. Moreover, the power of J_n is bigger than the power of J_0 under the alternative hypothesis. Therefore, J_n with $\delta_{r,N,T} = 2.25log(logT)^2log(0.25N + 0.75r)$ has more power than J_0.

7 Conclusion

In this paper, the vector $\boldsymbol{\theta}$ is divided into two sub-vectors: $\boldsymbol{\theta_S}$ and $\boldsymbol{\theta_D}$, where $\boldsymbol{\theta_S}$ is a sparse vector. Only $\boldsymbol{\theta_S}$ is empowered, which leads to a new power enhancement component J_n. This new power enhancement component is formed through a screening technique, which screens out the elements of $\boldsymbol{\theta_S}$ that are bigger than a critical value that depends on a threshold value $\delta_{r,N,T}$. When $\delta_{r,N,T}$=$\delta_{N,T}$, J_n does not suffer from size distortion. In fact, it improves the size

distortion without the loss of much power. Sometimes, J_1+J_n and $J_{thr}+J_n$ lose a little bit of power compared to J_1+J_0 and $J_{thr}+J_0$, because wrongly estimated θ_i's could get screened into J_0 and those wrongly estimated θ_i's are less likely to get screened into J_n since J_1 only empowers $\boldsymbol{\theta_S}$. In this research, I propose a new threshold value $\delta_{r,N,T}$ for power enhancement component J_n, depending on both N and r. Using the specific $\delta_{r,N,T}$ with parameters $\alpha = 0.25$ and $\beta = 0.75$, the proposed power enhancement not only has almost no size distortion under the null hypothesis, but also provides more power under sparse alternatives. To finally answer the research question, a conclusion is drawn that an empowerment only on the sparse alternatives leads to an improvement of the power enhancement testing.
The choice of the threshold value is very important for the power enhancement J_n. In this research, good thresholds are proposed for both the factor pricing model and the cross-sectional independence model by modifying the corresponding thresholds given by Fan, Liao, and Yao (2015). However, it still remains an open question: "What is the optimal threshold for J_n?" Answering this question in both theory and practice is a challenge for future research.

References

[1] Badi H. Baltagi, Qu Feng, and Chihwa Kao. "A Lagrange Multiplier test for cross-sectional dependence in a fixed effects panel data model". In: *Journal of Econometrics* 170.1 (2012), pp. 164–177.

[2] Peter J. Bickel and Elizaveta Levina. "Covariance Regularization by Thresholding". In: *The Annals of Statistics* 36.6 (2008), pp. 2577–2604.

[3] Jianqing Fan. "Test of Significance Based on Wavelet Thresholding and Neyman's Truncation". In: *Journal of the American Statistical Association* 91.434 (1996), pp. 674–688.

[4] Jianqing Fan, Yuan Liao, and Jiawei Yao. "Power Enhancement in High-dimensional Cross-sectional Tests". In: *Econometrica* 83.4 (2015), pp. 1497–1541.

[5] Michael R Gibbons, Stephen Ross, and Jay Shanken. "A Test of the Efficiency of a Given Portfolio". In: *Econometrica* 57.5 (1989), pp. 1121–52.

[6] GitHub. *Code Power Enhancements*. `https://github.com/Richard015/Code-Thesis`. 2021.

[7] Michael C. Jensen. "The Performance of Mutual Funds in The Period 1945–1964". In: *The Journal of Finance* 23.2 (1968), pp. 389–416.

[8] Anders Bredahl Kock and David Preinerstorfer. *Power in High-dimensional Testing Problems.* Working Papers Ecares 42. ULB – Universite Libre de Bruxelles, Nov. 2017.

[9] M. Hashem Pesaran and Takashi Yamagata. *Testing CAPM with a Large Number of Assets*. IZA Discussion Papers 6469. Institute of Labor Economics (IZA), Apr. 2012.

[10] Stephen Ross. “The arbitrage theory of capital asset pricing”. In: *Journal of Economic Theory* 13.3 (1976), pp. 341–360.