Testing Alphas in Conditional Time-Varying Factor Models with High Dimensional Assets

Shujie Ma, Wei Lan, Liangjun Su & Chih-Ling Tsai

Testing Alphas in Conditional Time-Varying Factor Models with High Dimensional Assets

Shujie Ma, Wei Lan, Liangjun Su, and Chih-Ling Tsai

Abstract

For conditional time-varying factor models with high dimensional assets, this article proposes a high dimensional alpha (HDA) test to assess whether there exist abnormal returns on securities (or portfolios) over the theoretical expected returns. To employ this test effectively, a constant coefficient test is also introduced. It examines the validity of constant alphas and factor loadings. Simulation studies and an empirical example are presented to illustrate the

Correspondence should be addressed to Wei Lan. Shujie Ma is Associate Professor, Department of Statistics, University of California at Riverside, Riverside, CA; E-mail: shujie.ma@ucr.edu. Wei Lan is Associate Professor, School of Statistics and Center of Statistical Research, Southwestern University of Finance and Economics, Chengdu, China; E-mail: lanwei@swufe.edu.cn. Liangjun Su is Lee Kong Chian Chair Professor, School of Economics, Singapore Management University, 90 Stamford Road, Singapore; E-mail: ljsu@smu.edu.sg. Chih-Ling Tsai is Distinguished Professor and Robert W. Glock Endowed Chair in Management, Graduate School of Management, University of California at Davis, Davis, CA; Email: cltsai@ucdavis.edu. Ma's research was partially supported by NSF grant DMS 1306972 and a Hellman Fellowship. Su gratefully acknowledges the Singapore Ministry of Education for an Academic Research Fund under grant number MOE2012-T2-2-021 and the funding support provided by the Lee Kong Chian Fund for Excellence.

finite sample performance and the usefulness of the proposed tests. Using the HDA test, the empirical example demonstrates that the FF three-factor model (Fama and French, 1993) is better than CAPM (Sharpe, 1964) in explaining the mean-variance efficiency of both the Chinese and US stock markets. Furthermore, our results suggest that the US stock market is more efficient in terms of mean-variance efficiency than the Chinese stock market.

Keywords: Conditional alpha test; High dimensional data; Mean-variance efficiency; Spline estimator; Time-varying coefficient

1 Introduction

Since the seminal works of Sharpe (1964) and Lintner (1965), the capital asset pricing model (CAPM) has played a fundamental role in modern finance. To measure investment performance, Jensen (1968) introduced the intercept term (i.e., 'Jensen's alpha or just 'alpha') into CAPM. Later, Fama and French (1993, 2015) extended the single-factor model, CAPM, to the three-factor and five-factor models, respectively. In these models, the excess return (the stock return minus the risk-free rate) for stock i at time t is denoted by R_{it}, and the risk premium on d-dimensional tradable systematic risks (d factors) at time t is denoted as $\mathbf{f}_t \in \mathbb{R}^d$. To incorporate the alpha into the general factor model, one can linearly relate the excess return of an asset (or a portfolio) to the factors (denoted by $\mathbf{f}_t$) through the intercept α_i and the factor loadings $\boldsymbol{\beta}_i \in \mathbb{R}^d$:

$$\mathbb{E}(R_{it}|\mathbf{f}_t) = \alpha_i + \boldsymbol{\beta}_i^\top \mathbf{f}_t, \tag{1}$$

where $i = 1, \cdots, N$ and $t = 1, \cdots, T$. It is worth noting that α_i should be zero for all N assets (or portfolios) in both CAPM and the Fama and French (FF) factor model.

To evaluate the marginal return associated with an additional strategy that is not explained by existing factors, many researchers employ the specification test for a factor model by testing

$$H_0 : \boldsymbol{\alpha} = 0 \quad \text{v.s.} \quad H_1 : \boldsymbol{\alpha} \neq 0,$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_N)^\top \in \mathbb{R}^N$ is a vector of intercepts involved in the factor model. For example, Gibbons, Ross and Shanken (1989, GRS hereafter) proposed an exact multivariate F-test for testing $\boldsymbol{\alpha} = 0$ under the joint normality assumption. Ever since, much effort has been devoted to this approach; see MacKinlay and Richardson (1991), Zhou (1993), and Beaulieu et al. (2007), to name just a few. However, the GRS test is only applicable when the number of assets (N) is smaller than the number of observations (T). In reality, N can be (much) larger than T. For example, in the Chinese stock market, there are about $N = 2,500$ stocks but only about $T = 1,500$ observations even for daily data, going back as far as 2007. In addition, a large T is likely to increase the possibility of structural changes in the factor loadings, which may adversely affect the performance of the GRS test (Pesaran and Yamagata, 2012).

To employ the GRS test, one needs to assume that the factor loadings are constant over time. This assumption can be quite restrictive in empirical finance. Much empirical evidence indicates that the factor loadings in the classical CAPM and the FF three-factor model vary substantially over time even at the portfolio level (see, e.g., Lewellen and Nagel, 2006; Ang and Chen, 2007). As a result, the GRS test can lead to inaccurate conclusions when the factor loadings are time-varying.

As discussed above, we have identified two limitations for using the GRS test; one is that the number of assets N is fixed, and the other is that the factor loadings are constant over time. To address the first one, Pesaran and Yamagata (2012) employed the thresholding covariance estimator of Fan et al. (2011) and proposed two novel Wald-type tests for testing the validity of CAPM. Accordingly, their tests are applicable to the $N > T$ case under certain conditions. However, the Wald-type tests often suffer from low power due to the accumulation of errors in estimating high-dimensional parameters. Thus, Fan et al. (2015) proposed a power enhancement screening procedure to strengthen the power under sparse alternatives (i.e., the null hypothesis is violated only by a very small number of components). Although the new tests of Pesaran and Yamagata (2012) and Fan et al. (2015) are not constrained by the limitation of $N < T$, they still require the factor loadings be constant.

To deal with the second limitation, Li and Yang (2011) and Ang and Kristensen (2012) considered conditional factor models and proposed nonparametric Wald-type tests to assess the significance of long-run conditional alphas (i.e., the average of alphas over a relatively long period) in the presence of time-varying factor loadings. However, their tests are only applicable when the number of assets N is fixed and the number of observations T tends to infinity. In practice, N can be very similar to or even greater than T. As a result, their tests can break down because either the covariance matrix of the estimators is not invertible or the sample covariance estimator is highly biased. Consequently, those tests can only resolve the second limitation, but not the first one.

To the best of our knowledge, there is no available test that can simultaneously address the aforementioned two limitations. The aim of this paper is to fill this gap in the literature. Specifically, we propose the High Dimensional Alpha (HDA) test for long-run conditional alphas while allowing for time-varying factor loadings. Similar to Fan et al. (2015), we consider large panels of assets and develop double asymptotics as both the number of assets N and the number of observations T tend to infinity. Moreover, the dimensionality N is allowed to grow faster than the sample size T. To allow for structural changes over the long run, we consider a time-varying factor model in which the factor loadings are assumed to be unknown smooth functions of time t. We estimate the factor loadings by linear combinations of spline basis functions.

Our HDA test circumvents the limitation of the Wald-type test because the latter is not applicable to the high-dimensional testing problem. Following the lead of Goeman et al. (2011), Lan et al. (2014), and Guo and Chen (2016), we could consider a score-type test. However, there are two major hurdles here. First, these authors considered the hypothesis test for a set of high-dimensional coefficients in parametric regression settings while we consider high-dimensional tests in semiparametric models because the factor loadings are modeled as a function of time in our setting. Second, the aforementioned works require that the dimension of nuisance parameters grow with the sample size at a slow rate, while the number of nuisance parameters in our model can be much larger than the time period T. Due to these major difficulties, it is challenging to apply the results of these existing studies. Instead, we propose a U-type statistic based on the residuals obtained from the null model, develop a bias-corrected estimator for the variance of the U-type statistic, and construct an asymptotically pivotal HDA test statistic. The detailed procedures for building up the test statistic associated with its theoretical property will be discussed in Section 2.3.

In practice, one may wonder whether the alphas and betas are varying over time before employing the HDA test. For this purpose, we subsequently propose a Constant Coefficients (CC) test to assess the constant alphas and factor loadings in the spirit of the generalized Chow (1960) F-test of Chen and Hong (2012) and Ang and Kristensen (2012). We show that this test statistic is asymptotically normally distributed under the null hypothesis of constant alphas and betas. When we fail to reject the null hypothesis in the CC test, we can apply an existing test (e.g., Fan et al.'s (2015) test) to examine the significance of the alphas. Otherwise, we should employ the HDA test to obtain a robust conclusion.

To assess the finite sample performance of our proposed tests, we conduct extensive simulations that demonstrate that both the HDA and CC tests perform well in terms of size and power. In empirical analysis, we study the market efficiency of both the Chinese and US stock markets via the conditional CAPM and conditional FF three-factor model. The CC test indicates that the alphas and betas are varying with time. Hence, we apply the HDA test to assess the validity of the two conditional pricing models. The results show that the FF three-factor model is better than CAPM in terms of explaining the variation of stock returns. In addition, based on the FF three-factor model, we often cannot reject the null hypothesis of market efficiency from the 53 and 300 rolling windows used for studying the Chinese and US stock markets, respectively. These results are more prominent for the US stock market, which suggests that the US stock market is more efficient than the Chinese stock market in terms of the mean-variance efficiency.

The rest of the paper is organized as follows. Section 2 presents the model and proposes the HDA test for assessing the market efficiency. Section 3 introduces the CC test for examining the constancy of factor loadings. Monte Carlo studies and empirical analyses for both the Chinese and US stock markets are given in Sections 4 and 5, respectively. Section 6 concludes. All technical details and some additional simulation and application results are relegated to the on-line supplemental materials.

2 Methodology

In this section, we present the conditional time-varying factor model and propose the HDA test for assessing the market efficiency.

2.1 Conditional Factor Model and Hypothesis

To explain the excess returns R_{it} of asset i at time t, we consider the following conditional factor model (Ang and Kristensen, 2012),

$$R_{it} = \alpha_{it} + \boldsymbol{\beta}_{it}^{\top}\mathbf{f}_t + e_{it} = \alpha_{it} + \sum_{j=1}^{d} \beta_{ijt} f_{jt} + e_{it}, \tag{2}$$

where $i = 1, \cdots, N$, $t = 1, \cdots, T$, α_{it} is the conditional alpha of asset i at time t, $\mathbf{f}_t = (f_{1t}, \cdots, f_{dt})^{\top}$ is a $d \times 1$ observable vector of common factors with fixed d, $\boldsymbol{\beta}_{it} = (\beta_{i1t}, \cdots, \beta_{idt})^{\top}$ is a $d \times 1$ vector of time-varying factor loadings, and e_{it} is the idiosyncratic error term. For the classic CAPM, the single common factor is the market risk, while for the FF three-factor model, the common factors are the market risk, SMB and HML, where SMB and HML measure the historic excess returns of small-cap stocks over big-cap stocks and of value stocks over growth stocks, respectively (see Fama and French, 1993). Furthermore, if α_{it} and $\boldsymbol{\beta}_{it}$ are not varying with time t, then the expected excess returns of model (2) are the same as that of model (1).

In model (2), the number of parameters is greater than the number of observations without additional assumptions on the model structure. To identify the parameters in model (2), we follow Li and Yang (2011) and assume that α_{it} and β_{ijt} are two smooth functions of time such that $\alpha_{it} = \alpha_i(t/T)$ and $\beta_{ijt} = \beta_{ij}(t/T)$. Our main interest is to test whether the average pricing error is equal to zero or not. To this end, we adopt the approach of Lewellen and Nagel (2006), Li and Yang (2011), and Ang and Kristensen (2012), and define the average alpha as $\delta_i^0 = T^{-1}\sum_{t=1}^{T} \alpha_{it}$ for $i = 1, \cdots, N$. Accordingly, we can rewrite model (2) as

$$R_{it} = \delta_i^0 + \delta_i(t/T) + \sum_{j=1}^{d} \beta_{ij}(t/T) f_{jt} + e_{it},$$

where $\delta_i(t/T) = \alpha_i(t/T) - T^{-1}\sum_{t=1}^{T} \alpha_i(t/T)$. Then the null and alternative hypotheses for testing the average alphas across the N assets are, respectively,

$$H_0 : \delta_i^0 = 0 \text{ for all } i = 1, \cdots, N \quad \text{v.s.} \quad H_1 : \delta_i^0 \neq 0 \text{ for some } i = 1, \cdots, N. \tag{3}$$

To construct the test statistic, we need to estimate $\delta_i(t/T)$ and $\beta_{ij}(t/T)$ under H_0.

Remark 1: It is worth mentioning that Jensen's alpha test is used for a similar purpose in a different context associated with the mean-variance efficiency. In fact, Jensen's alpha test is often used for testing the validity of CAPM (see, e.g., Jensen (1968) and Pesaran and Yamagata (2012)). On the other hand, Gibbons et al. (1989) pointed out that if a particular portfolio is mean-variance efficient (i.e., it minimizes variance for a given level of expected return), then the following first order condition must be satisfied for the given assets:

$$\mathbb{E}(R_{it}) = \alpha_{it} + \beta_i \mathbb{E}(r_{mt})$$

for some constant β_i and $\alpha_{it} = 0$, under the conditions that r_{mt} and the asset returns R_{it}'s are jointly normally distributed and linearly independent (see Gibbons et al., 1989). Here, r_{mt} is the excess return on the portfolio whose mean-variance efficiency is being tested. Accordingly, if the market portfolio exists, then testing the mean-variance efficiency is equivalent to testing $\alpha_{it} = 0$, which is essentially the same as the Jensen's alpha test.

2.2 Parameter Estimation

To proceed, we first introduce some notation. For any vector $\mathbf{v} = (v_1, ..., v_m)^\top$ / $\mathbb{R}^m$, let $\|\mathbf{v}\|$ be its L_2 norm and $\|\mathbf{v}\|_\infty = \max_{1\le i\le m} |v_i|$. In addition, let $\mathbf{1}_m$ be the $m \otimes 1$ vector of ones. For any positive numbers a_n and b_n, let $b_n \gg a_n$ denote $a_n^{-1} b_n = o(1)$, $a_n \approx b_n$ denote $\lim_{n\to\infty} a_n b_n^{-1} = 1$, and let $a_n \equiv b_n$ denote $\lim_{n\to\infty} a_n b_n^{-1} = c$ for some finite positive constant c. For an $m \otimes n$ matrix $\mathbf{A} = (a_{ij})$, let $tr\,(\mathbf{A})$ denote the trace of $\mathbf{A}$, $P_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$, and $M_{\mathbf{A}} = \mathbf{I}_m \quad P_{\mathbf{A}}$, where $\mathbf{I}_m$ is the $m \otimes m$ identity matrix. Moreover, denote $\|\mathbf{A}\|_\infty = \max_{1\le i\le m} \sum_{j=1}^n |a_{ij}|$ and $\|\mathbf{A}\| = \max_{\boldsymbol{\zeta}\in\mathbb{R}^n, \|\boldsymbol{\zeta}\|=1} \|\mathbf{A}\boldsymbol{\zeta}\|$. For any symmetric matrix $\mathbf{A}$ / $\mathbb{R}^{n\times n}$, let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ be the smallest and largest eigenvalues of $\mathbf{A}$, respectively. We use $(N, T)\ \infty \in$ to denote that N and T approach to infinity jointly. The operators $\overset{d}{\infty}$ and $\overset{p}{\infty}$ denote convergence in distribution and in probability, respectively, and plim denotes probability limit. Without further specification, the notations $o(\cdot)$, $o_p(\cdot)$, $O(\cdot)$ or $O_p(\cdot)$ hold as $(N, T)\ \infty \in$.

We employ the polynomial spline approach to estimating the unknown parameters $\alpha_i(t/T)$ and $\beta_{ij}(t/T)$. Let $0 = \xi_0 < \xi_1 < \cdots < \xi_n < 1 = \xi_{n+1}$ be a partition of $[0, 1]$ into subintervals

$I_\ell = [t_\ell, t_{\ell+1})$, $0 \geq \ell \geq n \quad 1$ and $I_n = [\xi_n, \xi_{n+1}]$ that satisfy

$$\max_{0\leq\ell\leq n} |\xi_{\ell+1} \quad \xi_\ell|| / \min_{0\leq\ell\leq n} |\xi_{\ell+1} \quad \xi_\ell|| \geq \tilde{m}$$

for some constant $0 < \tilde{m} < \in$, where $n \leq n(N,T)$ is the number of interior knots which satisfies $n \propto \in$ as $(N,T) \propto \in$ (see Su and Jin, 2012). For any t, define its location as $\ell(t)$ satisfying $\xi_{\ell(t)} \geq t/T < \xi_{\ell(t)+1}$. Consider the space of polynomial splines of order q on $[0,1]$, and then denote the normalized B spline basis of this space (de Boor, 2001, p.89) as $B(t/T) = \}B_1(t/T), \times\times\times, B_L(t/T)|^\top$, where $L = n + q$. To estimate $\delta_i(\times)$, we consider the centered spline basis functions, $B_\ell(t/T) = B_\ell(t/T) \quad T^{-1}\sum_{t=1}^{T} B_\ell(t/T)$, and denote $B(t/T) = \}B_1(t/T), \times\times\times, B_L(t/T)|^\top$. Then, the unknown functions $\delta_i(\times)$ and $\beta_{ij}(\times)$ can be well approximated by the B-spline functions (see Schumaker, 1981) such that

$$\delta_i(t/T) \ll \boldsymbol{\lambda}_{i0}^\top B(t/T) \quad \text{and} \quad \beta_{ij}(t/T) \ll \boldsymbol{\lambda}_{ij}^\top B(t/T),$$

where $\boldsymbol{\lambda}_{i0} \,/\, \mathbb{R}^{L\times 1}$ and $\boldsymbol{\lambda}_{ij} \,/\, \mathbb{R}^{L\times 1}$ are the coefficients of the B-spline functions. Under H_0, the estimators $\widetilde{\boldsymbol{\lambda}}_i = (\widetilde{\boldsymbol{\lambda}}_{ij}^\top, 0 \geq j \geq d)^\top$ can be obtained by minimizing

$$L_{NT}(\boldsymbol{\lambda}) = \widehat{\ }\sum_{i=1}^{N} \widehat{\ }\sum_{t=1}^{T} \Big\} R_{it} \quad \boldsymbol{\lambda}_{i0}^\top B(t/T) \quad \widehat{\ }\sum_{j=1}^{d} \boldsymbol{\lambda}_{ij}^\top B(t/T) f_{jt} \quad ^2,$$

where $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_i^\top, 1 \geq i \geq N)^\top$ and $\boldsymbol{\lambda}_i = (\boldsymbol{\lambda}_{ij}^\top, 0 \geq j \geq d)^\top$. Let $\mathbf{Z} = (\mathbf{Z}_1, \times\times\times, \mathbf{Z}_T)^\top$ where

$$\mathbf{Z}_t = \Big\} Z_{tk}, 1 \geq k \geq (1+d)L \quad ^\top = \Big\} B(t/T)^\top, \mathbf{f}_t^\top \asymp B(t/T)^\top \quad ^\top \,/\, \mathbb{R}^{(1+d)L\times 1}.$$

Then, we have

$$\widetilde{\boldsymbol{\lambda}}_i = (\widetilde{\boldsymbol{\lambda}}_{ij}^\top, 0 \geq j \geq d)^\top = (\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{R}_i,$$

where $\mathbf{R}_i = (R_{i1}, \times\times\times, R_{iT})^\top \,/\, \mathbb{R}^{T\times 1}$. Accordingly, the estimators of $\delta_i(t/T)$ and $\beta_{ij}(t/T)$ are $\widetilde{\delta}_i(t/T) = \widetilde{\boldsymbol{\lambda}}_{i0}^\top B(t/T)$ and $\widetilde{\beta}_{ij}(t/T) = \widetilde{\boldsymbol{\lambda}}_{ij}^\top B(t/T)$, respectively.

It is worth mentioning that the choice of basis functions does not affect the large-sample theories, according to our proofs. We choose B-spline basis functions because they are more computationally efficient and numerically stable in finite samples compared with other basis functions such as the truncated power series and trigonometric series (see Schumaker, 1981). Note that the above estimators depend on the number of interior knots, which is often unknown in practice. Thus, we follow the approach of Ma et al. (2014) and Ma and Song (2015) and employ the Bayesian information criterion (BIC) to select n by minimizing

$$\begin{aligned}\mathrm{BIC}\,(n) &= \log\left[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{R_{it}-\widetilde{\boldsymbol{\lambda}}_{i0}^{\top}B(t/T)-\sum_{j=1}^{d}\widetilde{\boldsymbol{\lambda}}_{ij}^{\top}B(t/T)f_{jt}\right\}^{2}\right]\\ &\quad+\frac{\log NT}{NT}(d+1)(n+q).\end{aligned}$$

2.3 High Dimensional Alpha (HDA) Test

Under the null hypothesis, the estimate of R_{it} is $\widetilde{R}_{it} = \widetilde{\delta}_i(t/T) + \widetilde{\boldsymbol{\beta}}_i(t/T)^{\top}\mathbf{f}_t$. Then, the resulting residuals are

$$\widetilde{e}_{it} = R_{it} - \widetilde{R}_{it} = R_{it} - \widetilde{\delta}_i(t/T) - \widetilde{\boldsymbol{\beta}}_i(t/T)^{\top}\mathbf{f}_t. \tag{4}$$

After simplification, we further have

$$\widetilde{e}_{it} = \delta_i^0 + \{\delta_i(t/T) - \widetilde{\delta}_i(t/T)\} + \{\boldsymbol{\beta}_i(t/T) - \widetilde{\boldsymbol{\beta}}_i(t/T)\}^{\top}\mathbf{f}_t + e_{it}.$$

Under H_0, $\delta_i^0 = 0$, and it can also be shown that $\delta_i(t/T) - \widetilde{\delta}_i(t/T) \xrightarrow{p} 0$ and $\beta_i(t/T) - \widetilde{\beta}_i(t/T) \xrightarrow{p} 0$ as $T \to \infty$. Accordingly, under H_0, $\widetilde{e}_{it} \xrightarrow{p} e_{it}$. This motivates us to consider the following statistic

$$\begin{aligned}J_{NT} &= N^{-1}\sum_{i=1}^{N}\left(T^{-1/2}\sum_{t=1}^{T}\widetilde{e}_{it}\right)^{2}\\ &= N^{-1}T^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{e}}_i^{\top}\mathbf{1}_T\mathbf{1}_T^{\top}\widetilde{\mathbf{e}}_i = N^{-1}T^{-1}\sum_{t,s=1}^{T}\widetilde{\mathbf{E}}_t^{\top}\widetilde{\mathbf{E}}_s,\end{aligned} \tag{5}$$

where $\widetilde{\mathbf{e}}_i = (\widetilde{e}_{i1}, \cdots, \widetilde{e}_{iT})^{\top}$ and $\widetilde{\mathbf{E}}_t = (\widetilde{e}_{1t}, \cdots, \widetilde{e}_{Nt})^{\top}$.

It is worth noting that, from simple mathematical derivation, we further have

$$J_{NT} = (\mathbf{1}_T^{\top}M_{\mathbf{Z}}\mathbf{1}_T)^2 T^{-1}N^{-1}G_{NT},$$

where $G_{NT} = \widetilde{\boldsymbol{\alpha}}^{\top}\widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{\alpha}} = (\widetilde{\alpha}_1, \cdots, \widetilde{\alpha}_N)^{\top}$, and $(\widetilde{\alpha}_i, \widetilde{\boldsymbol{\lambda}}_i) = \arg\min_{\alpha_i, \boldsymbol{\lambda}_i}\sum_{t=1}^{T}(R_{it} - \alpha_i - \boldsymbol{\lambda}_i^{\top}\mathbf{Z}_t)^2$. Thus, we can obtain J_{NT} through G_{NT}. It is also of interest to note that the GRS test statistic of Gibbons et al. (1989) for testing the efficiency of CAPM is proportional to $\widetilde{\boldsymbol{\alpha}}^{\top}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\alpha}}$, where the $N \times N$ matrix $\widetilde{\boldsymbol{\Sigma}}$ is the residual covariance matrix. Accordingly, the

GRS test statistic can be regarded as the weighted version of G_{NT} via the inverse of residual covariance matrix. However, the GRS test statistic is well defined only when $N < T - 1$ and the factor loadings are time-invariant. Accordingly, it is not applicable to our high-dimensional setting, in which N is larger than T, or when $\widetilde{\mathbf{\Sigma}}$ is not invertible. To resolve this issue, Pesaran and Yamagata (2012) proposed replacing $\hat{\mathbf{\Sigma}}$ with its diagonal version $D = \mathrm{diag}(\hat{\mathbf{\Sigma}})$, which is invertible even when N is larger than T. We name the resulting test statistic the PY test. It is worth noting that the PY test is designed for the models with time-invariant factor loadings, and it may not be applicable when the factor loadings are time varying; see simulation results in the supplementary materials.

To accommodate time varying factor loadings and to avoid using $\widetilde{\mathbf{\Sigma}}^{-1}$ in the high-dimensional setting, we propose to use a standardized version of J_{NT} as the test statistic. For this purpose, we need to calculate the mean and variance of J_{NT} under H_0, given below:

$$\begin{aligned} \mu_{NT}^0 &= N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}(e_{it}^2)\mathbb{E}(\eta_t^2), \text{ and} \\ \sigma_{NT}^2 &= 2N^{-2}T^{-2}tr(\mathbf{\Sigma}^2)\sum_{t\neq s}\mathbb{E}(\eta_t^2\eta_s^2), \end{aligned}$$

where

$$\eta_t = 1 - \mathbf{Z}_t^\top(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{1}_T, \tag{6}$$

$\mathbf{\Sigma} = \mathbb{E}\left(\mathbf{E}_t\mathbf{E}_t^\top\right)$, and $\mathbf{E}_t = (e_{1t}, \cdots, e_{Nt})^\top$. Let $\mu_{NT} = N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2\eta_t^2$ be an empirical approximation of the mean. Then, we centralize J_{NT} to yield

$$J_{NT}^* = J_{NT} - \mu_{NT}. \tag{7}$$

In our proposed test, we allow both N and T to be large. We need the conditions on the error vector $\mathbf{E}_t$ and its associated covariance matrix $\mathbf{\Sigma}$ given below.

(C1) (i) Assume that $\mathbf{E}_t = \mathbf{\Gamma}\mathbf{W}_t$ for $t = 1, \cdots, T$, where $\mathbf{\Gamma}$ is an $N \otimes v$ matrix for some $v \sim N$ and $\mathbf{W}_t = (w_{t1}, \cdots, w_{tv})^\top$ are v-variate independent and identically distributed random vectors satisfying $\mathbb{E}(\mathbf{W}_t) = 0$ and $\mathrm{Var}(\mathbf{W}_t) = \mathbf{I}_v$;
(ii) Assume that $\mathbb{E}(w_{tk}^4) = 3 + \Delta$ for some finite constant Δ and for any $1 \geq k \geq v$. In addition, assume

$$\mathbb{E}(w_{tk_1}^{\gamma_1}w_{tk_2}^{\gamma_2}\cdots w_{tk_u}^{\gamma_u}) = \mathbb{E}(w_{tk_1}^{\gamma_1})\mathbb{E}(w_{tk_2}^{\gamma_2})\cdots\mathbb{E}(w_{tk_u}^{\gamma_u})$$

for a positive integer u such that $\sum_{k=1}^{u} \gamma_k \geq 8$ and $k_1 \neq k_2 \neq \cdots \neq k_u$.

Condition (C1) is also used in Bai and Saranadasa (1996) and Chen and Qin (2010). Instead of assuming that the error terms are normally distributed, Condition (C1)(i) states that $\mathbf{E}_t$ can be expressed as a linear transformation of a v-variate $\mathbf{W}_t$ with mean 0 and variance matrix $\mathbf{I}_v$ that satisfies Condition (C1)(ii). As commented in Chen and Qin (2010), (C1)(i) is similar to factor models in multivariate analysis, but it allows $v \geq N$. Thus the rank and eigenvalues of $\boldsymbol{\Sigma}$ are not affected by the transformation. Simple calculations show that $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$.

(C2) (i) $tr\left(\boldsymbol{\Sigma}^4\right) = o\left\{tr^2\left(\boldsymbol{\Sigma}^2\right)\right\}$ as $N \to \infty$;
(ii) $T^{-2}\sum_{t=1}^{T} \mathbb{E}\left[\mathbf{E}_t^\top \boldsymbol{\Sigma}\mathbf{E}_t\mathbf{E}_t^\top \boldsymbol{\Sigma}\mathbf{E}_t\right] = o\left\{tr^2\left(\boldsymbol{\Sigma}^2\right)\right\}$.

Condition (C2)(i) is the same as Condition (3.7) given in Chen and Qin (2010), which is satisfied under various conditions on the eigenvalues of $\boldsymbol{\Sigma}$. If all eigenvalues are bounded, then (C2)(i) is trivially true. Note that our asymptotic results are established for $N \to \infty$, since we focus on studying the high-dimensional case with large N. For fixed N, theories can be derived with modifications of the proofs. As shown in the online Appendix, we have $T^{-2}\sum_{t=1}^{T} \mathbb{E}\left[\mathbf{E}_t^\top \mathbf{E}_t\mathbf{E}_t^\top \mathbf{E}_t\right] = T^{-1}\left\{tr(\boldsymbol{\Sigma})\right\}^2 + 2T^{-1}tr(\boldsymbol{\Sigma}^2)\{1+o(1)\}$. This result, together with the fact that $\|\boldsymbol{\Sigma}\|^2 \geq tr\left(\boldsymbol{\Sigma}^2\right)$, implies that

$$T^{-2}\sum_{t=1}^{T} \mathbb{E}\left[\mathbf{E}_t^\top \boldsymbol{\Sigma}\mathbf{E}_t\mathbf{E}_t^\top \boldsymbol{\Sigma}\mathbf{E}_t\right] \geq T^{-1}\left\{tr(\boldsymbol{\Sigma})\right\}^2\|\boldsymbol{\Sigma}\|^2 + 2T^{-1}tr^2\left[\boldsymbol{\Sigma}^2\right]\{1+o(1)\}.$$

Accordingly, if $T^{-1}\left\{tr(\boldsymbol{\Sigma})\right\}^2\|\boldsymbol{\Sigma}\|^2 = o\left\{tr^2\left(\boldsymbol{\Sigma}^2\right)\right\}$, then Condition (C2)(ii) holds. Let $\varsigma_1 \geq \cdots \geq \varsigma_N$ be the eigenvalues of $\boldsymbol{\Sigma}$. This condition is equivalent to $T^{-1}(\sum_{i=1}^{N}\varsigma_i)^2\varsigma_1^2 = o\{(\sum_{i=1}^{N}\varsigma_i^2)^2\}$ which implies (C2)(ii). This is trivially true when $\sum_{i=1}^{N}\varsigma_i\varsigma_1$ has the same order as $\sum_{i=1}^{N}\varsigma_i^2$.

(C3) (i) $TL^{-2r}N\left\{tr\left(\boldsymbol{\Sigma}^2\right)\right\}^{-1/2} = o(1)$, where $r > 3/2$ is the smooth order of the factor loading functions given in Assumption (A1) in the online Appendix;
(ii) $\left\{tr\left(\boldsymbol{\Sigma}^2\right)\right\}^{-1/2}\max_i \sum_{j=1}^{N}|\sigma_{ij}| = o(1)$, where σ_{ij} denotes the $(i,j)^{\text{th}}$ element of $\boldsymbol{\Sigma}$;
(iii) $T^{-1+\varrho}N^{1+\varrho}L\left\{tr\left(\boldsymbol{\Sigma}^2\right)\right\}^{-1/2} = O(1)$ for an arbitrarily small $\varrho > 0$.

Condition (C3) (i) and (iii) indicate that the number of spline basis functions L needs to satisfy $[TN\}tr\,(\boldsymbol{\Sigma}^2)|^{-1/2}]^{1/(2r)} \gg L \gg TN^{-1}\}tr\,(\boldsymbol{\Sigma}^2)|^{1/2}$ as $(N,T)\ \infty \in$. Furthermore, they imply that $N\}tr\,(\boldsymbol{\Sigma}^2)|^{-1/2} \gg T^{(2r-1)/(2r+1)}$. By assuming that $tr\,(\boldsymbol{\Sigma}^2) \equiv N^{1+a}$ for some $0 \geq a \geq 1$, we need $N^{1-a} \gg T^{2(2r-1)/(2r+1)}$ for $r > 3/2$. When $a = 1$, this is true for all N and T. When $0 \geq a < 1$, N is allowed to be larger than T since $2(2r \quad 1)/(2r+1) > 1$ for $r > 3/2$. When $tr\,(\boldsymbol{\Sigma}^2) \equiv N^{1+a}$, we require that $\max_i \sum_{j=1}^{N} \|\sigma_{ij}\| = o(N^{1/2+a/2})$ in order to satisfy (C3)(ii).

Now, let $\{_r$ denote the collection of all functions on $[0,1]$ such that the q^{th} order derivative satisfies the Hölder condition of order γ with $r \leq q + \gamma$. That is, there exists a constant $C_0 \ / \ (0, \in\ )$ such that for each $\phi \ / \ \{_r$,

$$\left(\phi^{(q)}(u_1) \quad \phi^{(q)}(u_2)\right(\geq C_0 \|u_1 \quad u_2\|^{\gamma}$$

for any $0 \geq u_1, u_2 \geq 1$. Let $\mathcal{H}_{NT,t} = \sigma\}\mathbf{f}, \}e_{it}, e_{i,t-1}, \times\times\times|_{i=1}^{N}|$ be the σ-algebra generated from $\}\mathbf{f}, \}e_{it}, e_{i,t-1}, \times\times\times|_{i=1}^{N}|$, where $\mathbf{f} = \}\mathbf{f}_1^{\top}, \times\times\times, \mathbf{f}_T^{\top}|^{\top}$. Denote $\mathbf{E}_{-t} = \}e_{i1}, \times\times\times, e_{it-1}, e_{i,t+1}, \times\times\times, e_{i,T}|_{i=1}^{N}$. To state the main results in this section, we add the following assumptions.

(A1) $\delta_i(\cdot) \ / \ \{_r$ and $\beta_{ij}(\cdot) \ / \ \{_r$ for some $r > 3/2$.

(A2) (i) There exist constants $0 < c_f \geq C_f < \in$ such that

$$c_f \geq \lambda_{\min}\}\mathbb{E}\}(1, \mathbf{f}_t^{\top})^{\top}(1, \mathbf{f}_t^{\top})|\,| \geq \lambda_{\max}\}\mathbb{E}\}(1, \mathbf{f}_t^{\top})^{\top}(1, \mathbf{f}_t^{\top})|\,| \geq C_f$$

holds uniformly for $t \ / \ [1,T]$; (ii) There exists a constant $0 < M < \in$ such that $\mathbb{E}\|\mathbf{f}_t\|^{4(2+\varkappa)} \geq M$ for some $\varkappa > 0$; (iii) The process $\}\mathbf{f}_t, t \sim 1|$ is strong mixing with mixing coefficient $\alpha(\cdot)$ satisfying $\sum_{k=0}^{\infty} \alpha(k)^{\varkappa/(2+\varkappa)} < \in$.

(A3) (i) $\mathbb{E}(e_{it}|\mathcal{H}_{NT,t-1}) = 0$ for each $i = 1, \times\times\times, N$; (ii) $\mathbb{E}\ \mathbf{E}_t\mathbf{E}_t^{\top}\ |\mathbf{E}_{-t}\left[\right. = \mathbb{E}(\mathbf{E}_t\mathbf{E}_t^{\top}) = \boldsymbol{\Sigma}$ for all $1 \geq t \geq T$, $\boldsymbol{\Sigma}$ is a positive definite matrix, and $\sigma_{ii} \ / \ (0, \in\ )$ for every $1 \geq i \geq N$; (iii) $\}\mathbf{f}_t|_{t=1}^{T}$ and $\}\mathbf{E}_t|_{t=1}^{T}$ are independent.

Assumption (A1) is the smoothness assumption on the unknown functions, which is commonly used in the nonparametric smoothing literature; see He and Shi (1996). Assumption

(A2)(i) is the same as Condition (C2) in Wang et al. (2008), and this assumption is a typical condition on the design matrix for regression. Following Fan et al. (2011) and Fan et al. (2015), we assume that the factors $\}\mathbf{f}_t|_{t=1}^{T}$ follow the strong mixing condition. Moreover, Assumptions (A2)(ii) and (iii) are weaker than Assumptions 3.2 and 3.3(ii) given in Fan et al. (2011). Assumption (A3)(i) is a typical assumption for a martingale difference sequence. Assumption (A3)(ii) ensures that the covariance of the error terms satisfies the homogeneity assumption. Assumption (A3)(iii) follows from Assumption 3.1(ii) of Fan et al. (2011).

We now state our first main result, which is about the asymptotic property of $\sigma_{NT}^{-1}J_{NT}^{*}$.

Theorem 1. *Suppose that Conditions (C1), (C2), (C3)(i)-(ii), and Assumptions (A1)-(A3) hold. Assuming* $L^3T^{-1} = o(1)$, *under the local alternative*

$$H_{1,NT} : \delta_i^0 \leq \delta_{i,NT}^0 = N^{-1/2}T^{-1/2}\}tr(\mathbf{\Sigma}^2)|^{1/4}c_i^0 \tag{8}$$

for any $i = 1, \times\times\times, N$, *where* $N^{-1}\sum_{i=1}^{N}(c_i^0)^2 \infty c_0 \; / \; [0, \in\;)$ *as* $N \infty \in$, *we have*

$$\sigma_{NT}^{-1}\Big\} J_{NT}^{*} \quad N^{-1}T^{-1}\widehat{\quad} \sum_{i=1}^{N} \delta_i^0 \Big[^2 \left(\mathbf{1}_T^\top M_{\mathbf{Z}}\mathbf{1}_T\right)^2 \quad \overset{d}{\infty} N(0,1),$$

as $(N,T) \infty \in$. *Moreover, there are some constants* $0 < c_M \geq C_M < \in$ *such that*

$$2c_M^2 N^{-2}tr \;\; \mathbf{\Sigma}^2\big[\; \}1 + o(1)| \; \geq \sigma_{NT}^2 \geq 2C_M^2 N^{-2}tr \;\; \mathbf{\Sigma}^2\big[\; \}1 + o(1)|\;.$$

The above theorem shows that, under H_0, $\sigma_{NT}^{-1}J_{NT}^{*}$ follows the standard normal distribution asymptotically (i.e., $\sigma_{NT}^{-1}J_{NT}^{*} \overset{d}{\infty} N(0,1)$). Under the local alternative (8), $\sigma_{NT}^{-1}J_{NT}^{*}$ has the asymptotic normal distribution with mean $\gamma^0 = \text{plim}_{(N,T)\to\infty}\sigma_{NT}^{-1}N^{-1}T^{-1}\sum_{i=1}^{N}\left(\delta_i^0\right)^2\left(\mathbf{1}_T^\top M_{\mathbf{Z}}\mathbf{1}_T\right)^2$ and variance 1. In addition, based on the result in Theorem 1, we have $\sigma_{NT}^2 \equiv N^{-2}tr\left(\mathbf{\Sigma}^2\right)$.

It is worth noting that $\sigma_{NT}^{-1}J_{NT}^{*}$ is usually unknown since it involves population parameters. Thus it cannot be used as a test statistic in practice. We therefore need to find consistent estimators of J_{NT}^{*} and σ_{NT}^2. In the proof of Theorem 2 below we show that J_{NT}^{*} can be consistently estimated by

$$\widetilde{J}_{NT}^{*} = J_{NT} \quad N^{-1}T^{-1}\widehat{\quad} \sum_{i=1}^{N} \widetilde{e}_{it}^2\eta_t^2$$

in the sense that $\widetilde{J}_{NT}^{*} \quad J_{NT}^{*} = o_p(\sigma_{NT})$. As for σ_{NT}^2, we need to estimate the unknown quantity $tr(\mathbf{\Sigma}^2)$. A natural estimate is given by $tr(\widetilde{\mathbf{\Sigma}}^2)$, where $\widetilde{\mathbf{\Sigma}} = T^{-1}\sum_{t=1}^{T}(\widetilde{\mathbf{E}}_t \quad \bar{\widetilde{\mathbf{E}}}_t)(\widetilde{\mathbf{E}}_t$

$\bar{\widetilde{\mathbf{E}}}_t)^\top$ and $\bar{\widetilde{\mathbf{E}}}_t = T^{-1}\sum_{t=1}^{T}\widetilde{\mathbf{E}}_t$. However, as demonstrated by Srivastava (2005), $tr(\widetilde{\mathbf{\Sigma}}^2)$ is not a consistent estimator of $tr(\mathbf{\Sigma}^2)$. To address this issue, we adopt the approach of Lan et al. (2014), and consider the following bias-corrected estimator

$$\widehat{tr(\mathbf{\Sigma}^2)} = T^2(T+(1+d)L \quad 1)^{-1}(T \quad (1+d)L)^{-1}\Big\}tr(\widetilde{\mathbf{\Sigma}}^2) \quad tr^2(\widetilde{\mathbf{\Sigma}})/(T \quad (1+d)L) \; .$$

Based on this estimator, we will demonstrate in the proof of the following theorem that σ^2_{NT} can be estimated consistently by

$$\widetilde{\sigma}^2_{NT} = 2N^{-2}T^{-2} \widehat{\ } \sum_{t\neq s} \eta_t^2\eta_s^2\widehat{tr(\mathbf{\Sigma}^2)}$$

in the sense that $\widetilde{\sigma}^2_{NT} \quad \sigma^2_{NT} = o_p(\sigma^2_{NT})$. Accordingly, we propose to use $\widetilde{\sigma}^{-1}_{NT}\widetilde{J}^*_{NT}$ as the test statistic, and the asymptotic distribution is given below.

Theorem 2. *Suppose that Conditions (C1)-(C3) and Assumptions (A1)-(A3) hold. Assume that $L^3T^{-1} = o(1)$ and $L^rT^{-3/2} = o(1)$. Then under the local alternative given in (8), we have $\sigma^{-1}_{NT}(\widetilde{J}^*_{NT} \quad J^*_{NT}) = o_p(1)$, $\widetilde{\sigma}^2_{NT} = \sigma^2_{NT}\}1 + o_p(1)|$, and*

$$\widetilde{\sigma}^{-1}_{NT}\Big\}\widetilde{J}^*_{NT} \quad N^{-1}T^{-1} \widehat{\ } \sum_{i=1}^{N} \delta_i^0 \big[^2 (\mathbf{1}_T^\top M_{\mathbf{Z}}\mathbf{1}_T)^2 \quad \overset{d}{\infty} N(0,1)$$

as $(N,T) \infty \in$.

Under H_0, the above theorem yields a test statistic $\widetilde{Z}_{NT} = \widetilde{\sigma}^{-1}_{NT}\widetilde{J}^*_{NT}$, which has $N(0,1)$ distribution asymptotically. This allows us to devise a test when N and T are large, so we name it the High Dimensional Alpha (HDA) test. Consequently, for any given significance level ν, we can reject the null hypothesis if $\widetilde{Z}_{NT} > z_{1-\nu}$, where $z_{1-\nu}$ denotes the ν-th upper quantile of a standard normal distribution. Furthermore, one can employ Theorem 2 to evaluate the power of the HDA test. The following remark presents the asymptotic power of the HDA test.

Remark 2: From the result in Theorem 2, we obtain that under the local alternative given in (8), $P(\widetilde{Z}_{NT} > z_{1-\nu}) \overset{p}{\infty} 1 \quad \Phi(z_{1-\nu} \quad \gamma^0)$ as $(N,T) \infty \in$, where $\Phi(\cdot)$ stands for the cumulative distribution of a standard normal distribution.

Remark 3: The above procedure for testing $\alpha_i = 0$ is also applicable for testing $\alpha_{it} = 0$ at each time t by establishing the asymptotic distribution of $\hat{e}_{it}$. As discussed by Lee (2001), however, the market efficiency should be considered a process rather than a destination. Hence, in some aspects, it is practically valuable to test the average α_{it} over relatively long periods.

3 Constant Coefficient Test

The proposed HDA test can be used to test alphas without assuming constant factor loadings. If the null hypothesis (of the alphas over a long period being consistently zero) is rejected, one would naturally ask whether the conditional alphas and factor loadings are homogeneous over time for each stock. In fact, testing the alphas under the homogeneity assumption on factor loadings has been extensively studied in the literature; see, e.g., Gibbons et al. (1989), MacKinlay and Richardson (1991), Zhou (1993), Beaulieu et al. (2007), Pesaran and Yamagata (2012), and Fan et al. (2015). In the varying coefficient scenario, one may consider the generalized likelihood ratio (GLR) approach proposed by Fan et al. (2001). Since the generalized F-test can be easily adapted to the B-spline-based estimation procedure by utilizing matrix projections, we borrow the idea from Chen and Hong (2012) and Ang and Kristensen (2012) of exploiting a generalized version of Chow's (1960) F-test. Accordingly, we derive a spline-based generalized F-test statistic, which has a simpler expression. Hence, it is easier to compute than the kernel-based test statistic given in Chen and Hong (2012) and Ang and Kristensen (2012). We also obtain the asymptotic distribution of the spline-based generalized F-test statistic in Proposition 1 below. It is worth noting that this test is only for testing each individual stock, and thus it is not a high dimensional testing problem.

For each stock i, define $\boldsymbol{\theta}_{it} = (\alpha_{it}, \boldsymbol{\beta}_{it}^\top)^\top \in \mathbb{R}^{d+1}$. Then, consider the following hypotheses:

$$H_0^{i,c} : \boldsymbol{\theta}_{i1} = \boldsymbol{\theta}_{i2} = \cdots = \boldsymbol{\theta}_{iT}, \quad \text{v.s.} \quad H_1^{i,c} : \boldsymbol{\theta}_{it_1} \neq \boldsymbol{\theta}_{it_2} \text{ for some } t_1 \neq t_2.$$

To test the null hypothesis, we estimate model (2) using the ordinary least squares method and the spline-based estimation method, respectively, under $H_0^{i,c}$ and $H_1^{i,c}$. Denote their corresponding residual sum of squares by $\mathrm{RSS}_0^{(i)}$ and $\mathrm{RSS}_1^{(i)}$. In addition, define $\mathbf{F}_t = (1, \mathbf{f}_t^\top)^\top \in \mathbb{R}^{d+1}$ and $\mathbf{F} = (\mathbf{F}_1, \cdots, \mathbf{F}_T)^\top \in \mathbb{R}^{T\times(d+1)}$. Let $\mathbf{Z}_t = \left\{\mathbf{F}_t^\top \otimes B(t/T)^\top\right\}^\top \in \mathbb{R}^{(1+d)L\times 1}$. After simple calculations, we have $\mathrm{RSS}_0^{(i)} = \mathbf{R}_i^\top M_{\mathbf{F}} \mathbf{R}_i$ and $\mathrm{RSS}_1^{(i)} = \mathbf{R}_i^\top M_{\mathbf{Z}} \mathbf{R}_i$. In the spirit of the generalized Chow (1960) F-test of Chen and Hong (2012) and Ang and Kristensen (2012), we propose the following test statistic

$$C_T^{(i)} = \left(\mathrm{RSS}_0^{(i)} - \mathrm{RSS}_1^{(i)}\right)/\mathrm{RSS}_0^{(i)} = \mathbf{R}_i^\top \left(M_{\mathbf{F}} - M_{\mathbf{Z}}\right)\mathbf{R}_i / \left(\mathbf{R}_i^\top M_{\mathbf{F}} \mathbf{R}_i\right).$$

The following proposition studies the theoretical property of $C_T^{(i)}$.

Proposition 1. *Suppose that Assumptions (A1)-(A3) hold, $L = o(T^{1/3})$, and $L^{-1} = o(1)$. For each $i = 1, \cdots, N$, under the null hypothesis of $H_0^{i,c}$, we have*

$$\left(2(L-1)(d+1)\right)^{-1/2}\left((T-d-1)C_T^{(i)} - (L-1)(d+1)\right) \xrightarrow{d} N(0,1)$$

as $(N,T) \to \infty$.

Define $\bar{C}_T^{(i)} = 2\left((L-1)(d+1)\right)^{-1/2}\left((T-d-1)C_T^{(i)} - (L-1)(d+1)\right)$. Then, by Proposition 1 we should reject the null hypothesis of $H_0^{i,c}$ for stock i if $\bar{C}_T^{(i)} > z_{1-\nu}$ for any given significance level ν. Since this test is useful for testing the constancy of coefficients across time, we name it the constant coefficient (CC) test.

To assess the market efficiency, we can first employ the CC test for testing the null hypothesis of $H_0^{i,c}$. If $H_0^{i,c}$ is rejected for some $i = 1, \cdots, N$, then one should use the HDA test for large N and the test of Li and Yang (2012) or Ang and Kristensen (2012) for small N. If the null hypothesis is not rejected for any $i = 1, \cdots, N$, it is more efficient to consider a GRS-type test since GRS tests are designed for homogeneous factor loadings.

Remark 4: The CC test can be modified to test the constancy of conditional alphas and the constancy of conditional betas separately (Ang and Kristensen, 2012; Li and Yang, 2011). Specifically, we can test the following two hypotheses individually,

$$H_0^{i,\alpha}: \alpha_{i1} = \alpha_{i2} = \cdots = \alpha_{iT}, \quad \text{v.s.} \quad H_1^{i,\alpha}: \alpha_{it_1} \neq \alpha_{it_2} \text{ for some } t_1 \neq t_2;$$

$$H_0^{i,\beta}: \boldsymbol{\beta}_{i1} = \boldsymbol{\beta}_{i2} = \cdots = \boldsymbol{\beta}_{iT}, \quad \text{v.s.} \quad H_1^{i,\alpha}: \boldsymbol{\beta}_{it_1} \neq \boldsymbol{\beta}_{it_2} \text{ for some } t_1 \neq t_2.$$

Applying the same techniques as those for deriving the CC test, we can obtain the corresponding test statistics. Their asymptotic normality can be established by following the same procedure as used in the proof of Proposition 1.

4 Simulation Studies

To evaluate the finite sample performance of the HDA and CC tests, we present three simulated examples that mimic the US stock market. We also conduct simulated experiments for mimicking the Chinese stock market. Since the simulation results yield similar findings, we relegate them to the supplementary materials to save space.

4.1 Three Examples

Example 1: One-factor model with time-varying coefficients. Following Li and Yang (2011), we generate the data from the conditional CAPM with the intercept alphas:

$$R_{it} = \alpha_{it} + \beta_{it} f_t + e_{it} \quad (i = 1, \cdots, N, t = 1, \cdots, T), \tag{9}$$

where f_t is the excess market return. We generate f_t by mimicking the US stock market data described in the next section. Specifically, we assume that f_t follows an AR(1)-GARCH(1,1) process,

$$f_t - 0.34 = 0.05(f_{t-1} - 0.34) + h_t^{1/2}\zeta_t,$$

where ζ_t follows a standard normal distribution, h_t is generated from the process

$$h_t = 0.32 + 0.67h_{t-1} + 0.13h_{t-1}\zeta_{t-1}^2,$$

and the above coefficients are obtained by fitting the model to the US stock market data.

We next consider factor loadings and alphas. Specifically, we borrow the setting from Su and Wang (2017) and set the conditional factor loadings to be $\beta_{it} = G(10t/T, 2, 2)$, so that β_{it} is a non-random smooth function of t/T for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, where $G(z, \kappa_1, \kappa_2) = [1+\exp\{-\kappa_1(z-\kappa_2)\}]^{-1}$ denotes the Logistic function with tuning parameter κ_1 and location parameter κ_2. In addition, the conditional alphas are set to be $\alpha_{it} = c_i t/T$ for $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Thus, under the null hypothesis, $c_i = 0$ for all i, which leads to the conditional CAPM. We lastly generate the error term $\mathbf{E}_t = (e_{1t}, \cdots, e_{Nt})^\top \in \mathbb{R}^N$. To examine the performance of HDA and CC for various error distributions, we generate the variable $\mathbf{E}_t$ via $\mathbf{E}_t = \mathbf{\Sigma}^{1/2}\mathbf{Z}_t^*$ for $t = 1, \cdots, T$, where each component of $\mathbf{Z}_t^*$ is independently simulated, respectively, from a standard normal distribution (N(0,1)), a standardized exponential distribution (exp(1)), and a mixture distribution $0.1N(0,9) + .9N(0,1)$. As for $\mathbf{\Sigma} = (\sigma_{j_1 j_2}) \in \mathbb{R}^{N\times N}$, we consider the following two settings: one is borrowed from Fan and Li (2001) with $\sigma_{j_1 j_2} = 0.5^{|j_1 - j_2|}$, which implies that $e_{j_1 t}$ and $e_{j_2 t}$ are approximately uncorrelated when the difference $|j_1 - j_2|$ is sufficiently large; the other is borrowed from Fan et al. (2015), where $\mathbf{\Sigma} = \mathrm{diag}(A_1, \cdots, A_{N/4})$ is a block-diagonal correlation matrix, and each diagonal block A_j for $j = 1, \cdots, N/4$ is a 4×4 positive definite matrix whose correlation matrix has equi-off-diagonal entry ρ_j generated from Uniform[0,0.5]. Since the two settings yield very similar patterns, we only present the results of the first setting here, while the results for the second setting are relegated to the supplementary materials.

The above process is simulated over the periods $t = -24, \cdots, 0, 1, \cdots, T$ with the initial values $R_{i,-25} = 0$, $h_{-25} = 1$, $z_{-25} = 0$ and $\sigma^2_{-25} = 1$. To offset the start-up effects, we drop the first 25 simulated observations and use $t = 1, \cdots, T$ in our studies.

Example 2: Three-factor model with time-varying coefficients. To study three factor effects on the tests, we consider the following Fama-French conditional factor model:

$$R_{it} = \alpha_{it} + \sum_{j=1}^{3} \beta_{ijt} f_{jt} + e_{it} \quad (i = 1, \cdots, N, t = 1, \cdots, T), \tag{10}$$

where f_{1t}, f_{2t} and f_{3t} represent the three factors, i.e., the market factor, SMB (small [size] minus big) and HML (high [value] minus low). To mimic the US stock market, these factors are correspondingly simulated from the following AR(1)-GARCH(1,1) processes,

$$\text{Market factor: } f_{1t} - 0.34 = 0.05(f_{1t-1} - 0.34) + h_{1t}^{1/2}\zeta_{1t},$$

$$\text{SMB factor: } f_{2t} - 0.04 = 0.07(f_{2t-1} - 0.04) + h_{2t}^{1/2}\zeta_{2t},$$

$$\text{HML factor: } f_{3t} - 0.06 = 0.04(f_{3t-1} - 0.06) + h_{3t}^{1/2}\zeta_{3t},$$

where ζ_{jt} ($j = 1, 2$ and 3) are simulated from a standard normal distribution, h_{jt} ($j = 1, 2$ and 3) are, respectively, generated through the following processes,

$$\text{Market factor: } h_{1t} = 0.32 + 0.67h_{1t-1} + 0.13h_{1t-1}\zeta^2_{1t-1},$$

$$\text{SMB:}\ \ h_{2t} = 0.33 + 0.51 h_{2t-1} + 0.03 h_{2t-1}\zeta^2_{2t-1},$$

$$\text{HML:}\ \ h_{3t} = 0.26 + 0.72 h_{3t-1} + 0.05 h_{3t-1}\zeta^2_{3t-1},$$

and the above coefficients are obtained by fitting the model to the US stock market data presented in Section 5.

The conditional factor loadings are $\beta_{ijt} = a_j G(10t/T, 2, 2) + b_j$ for $i = 1, \cdots, N$, $j = 1, 2, 3$, and $t = 1, \cdots, T$, where $(a_1, b_1) = (0.5, 0.5)$, $(a_2, b_2) = (0.1, 0.5)$, and $(a_3, b_3) = (0.2, 0.5)$. In addition, the conditional alphas are set to be $\alpha_{it} = c_i t/T$ for $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Thus, under the null hypothesis, $c_i = 0$ for all i, which leads to the conditional three-factor model. Finally, the error terms $\mathbf{E}_t$, initial values, and the simulated observations have the same settings as in Example 1.

Example 3: Three-factor model with random coefficients. In the above two examples, the factor loadings are set to be non-random smooth functions of t/T as described in Subsection 2.1. To assess the robustness of the proposed test for the random factor loadings, we consider the same model settings as in model (10) of Example 2, except that the conditional alphas α_{it} and the conditional factor loadings β_{ijt} are generated from the unobservable state variable z_t via $\alpha_{it} = c_i z_t$ and $\beta_{ijt} = a_j + b_j z_t$ for $i = 1, \cdots, N$, $j = 1, 2, 3$, and $t = 1, \cdots, T$. Furthermore, z_t follows an AR(1)-ARCH(1) process, $z_t = 0.8 z_{t-1} + u_t$, where $u_t = \sigma_t \epsilon_t$, ϵ_t follows a standard normal distribution, and $\sigma_t^2 = 0.1 + 0.6\sigma_{t-1}^2$ with $\sigma_0^2 = 1$.

4.2 Performance of the HDA Test

To evaluate the size performance of the HDA test, we set $c_i = 0$ for all i in the above three examples. Then, three different sample sizes ($T = 100$, 200, 500) and four different numbers of stocks ($N = 3$, 200, 500, 1,000) are considered. For each setting, all simulations are conducted via 1,000 realizations with nominal level $\alpha = 5\%$. The GRS test of Gibbons et al. (1989) and the tests of Pesaran and Yamagata (2012) are only applicable when the factor loadings are constant over time. Hence, we only compare our HDA test with the LY test from Li and Yang (2011).

Table 1 presents the sizes of the HDA test across three sample sizes and four different numbers of stocks for Examples 1–3, respectively. Here, the number of interior knots n is determined by the BIC criterion, as discussed in Subsection 2.2, and the order of B-splines is set at 3. The results in Table 1 indicate that HDA performs well regardless of $T = 100$, 200 or 500, $N = 3$, 200, 500 or 1,000, and the error distribution being normal, exponential, or a mixture. Hence, HDA is not only applicable to the case $N > T$, but also robust to various (N, T) specifications and error distributions. It is worth noting that the results of Example 3 yield a similar pattern to those in Examples 1–2. This implies that HDA is also robust to the specification of the factor loadings. In contrast, the LY test exhibits serious size distortion when N is relatively large. For example, the empirical sizes are equal to 1 for $N \sim 200$. This finding is not surprising since the LY test is not designed for $N > T$, and it performs well when $N = 3$. As a result, we only consider HDA in the evaluation of power.

To study the power of the HDA test, we consider the following two different types of alternative hypotheses for Examples 1–2. The first one is the dense alternative under which $\alpha_{it} = c_i t/T = ct/T$ for some constant c and $i = 1, \times\times\times, N$. The second one is the sparse alternative under which $\alpha_{it} = c_i t/T = ct/T$ for some constant c if $i \geq 20$; $\alpha_{it} = 0$, otherwise. This alternative setting is motivated by the empirical finding of Fan et al. (2015) that the market inefficiency is only induced by a very small portion of stocks. In both alternative settings, the signal strength c ranges from 0 to 0.2 with an increment of 0.01. For the sake of illustration, we only consider the normally distributed random errors with $N = 200$.

Figures 1 depicts the empirical powers of the HDA test over three sample sizes ($T = 100, 200, 500$), two different data generation processes (Examples 1 and 2), two types of alternatives (dense and sparse), and 20 (=0.2/0.01) signals of c. The results indicate that the empirical power of HDA steadily increases to 1 as the signal strength c gets larger. In addition, the power of HDA becomes large as the sample size T increases. In sum, the HDA test performs satisfactorily and comparably under both dense and sparse alternatives, and it is indeed consistent.

4.3 Performance of the CC Test

To study the finite sample performance of the CC test, without loss of generality, we consider the case of a single stock (i.e., $N = 1$). We adopt the settings given in Examples 1–2, except for $\beta_{it} = 1 + bz_t$ in Example 1 and $\beta_{ijt} = 1 + bz_t$ in Example 2 with $i = 1$. In addition, we set $b = 0$ for assessing the size of the test, and $b = 0.1$ and 0.3 for examining the power of the test. Table 2 reports the test results. From Table 2 we observe that the empirical sizes of the CC test are all around 0.05 regardless of the time length T and the error distribution. Furthermore, the empirical power of the CC test increases to 1 as b or T becomes larger. In sum, the CC test performs well in terms of both size and power.

5 Real Data Analysis

In this section, we employ the proposed tests to assess the mean-variance efficiency of both the Chinese and US stock markets. It is worth mentioning that our proposed HDA test is applicable even for $N \to T$, which allows us to target a large pool of stocks directly so that there is no need to group a large number of stocks into a small number of portfolios. The goal of this empirical study is two-fold: (i.) investigate the efficiency of the Chinese and US stock markets during the study period; (ii.) explore the differences between the Chinese and US stock markets based on the results of mean-variance efficiency. Since the US stock market is more mature than Chinese stock market, we first study the efficiency of the Chinese stock market and then compare the results with that of the US stock market.

5.1 Data Description

The Chinese stock market data are collected from the WIND database (one of the most authoritative databases in China), which contains securities in the Shanghai-Shenzhen 300 Index. We use this dataset because the 300 stocks in this index are the most frequently traded stocks in China; hence, these data are not significantly impacted by a survivorship bias (see Brown et al., 1995). After eliminating the stocks with missing observations to avoid analyzing an unbalanced panel, there remain $T = 153$ weekly observations for each of the $N = 292$ stocks from 11/25/2011 to 12/31/2014. As a result, each observation represents a particular firm's weekly excess return (the stock return minus the risk-free interest rate). The weekly return of the 1-year deposit is chosen to proxy the risk-free interest rate r_{ft}. Following the literature on the study of the Chinese stock market, we use the Shanghai Composite Index (the value-weighted return on all Shanghai A-share stocks) as the proxy for the market portfolio r_{mt}. Then, according to the definition given by Fama and French (1993), the factor SMB is the average return of the three smallest portfolios minus the average return on the three biggest portfolios, and the factor HML is the average return on the three highest value stock portfolios minus the average return on the three lowest value portfolios. Note that all the stocks used in this study are listed in the Shanghai and Shenzhen A-share stock market.

To make a comparison with the US stock market, we also collect data for securities in the Standard & Poor's 500 (S&P 500) index from the period 01/08/2010 to 08/25/2017. After eliminating the assets with missing observations, there remain $T = 399$ weekly observations for each of the $N = 442$ firms. The three factors are obtained from Ken French's data library web page. The one-month US treasury bill rate is chosen as the risk-free rate, and the value-weighted returns on all NYSE, AMEX, and NASDAQ stocks obtained from CRSP are used as a proxy for the market return.

Table 3 reports the descriptive statistics that include the mean, median, and standard deviation (SD) for the market factor, SMB and HML for both the Chinese and US stock markets. According to Table 3, we find that the returns on SMB and HML in the Chinese stock market are much larger than those in the US stock market. It is of interest to note that Aboody and Lev (2000) and Abosede and Oseni (2011), respectively, noticed that SMB and HML can be related to information asymmetry. In addition, Dai et al. (2013) indicated that the Chinese stock market suffers from more information asymmetry, such as insufficient information disclosure mechanisms and regulatory instruments, than stock markets in western countries. Hence, we conjecture that higher information asymmetry in Chinese market maybe leads to the larger observed returns on SMB and HML, but a thorough and rigorous study on this subject needs to be explored in future research and is not the focus of this paper.

5.2 Are Alphas and Betas Time-Varying?

Before assessing the market efficiency of the Chinese stock market (US stock market), it is reasonable to check whether the alphas and betas are time-varying. To this end, for each individual stock, we employ the CC test to examine the constancy of the alphas and factor loadings in both CAPM and the FF three-factor model. The results show that the p-value for testing the constancy of the alphas and factor loadings in each of the 292 (442) stocks is close to 0, regardless of the model. This strong evidence indicates that the alphas and betas are time-varying in both the Chinese and US stock markets, which suggests that the conditional time-varying factor model is more suitable than the traditional time-invariant factor model.

5.3 Mean-Variance Portfolio Efficiency

We first employ the HDA test to assess the market efficiency of the Chinese stock market based on $N = 292$ stocks with $T = 153$ corresponding observations for each stock recorded between 11/25/2011 and 12/31/2014. Specifically, we consider the following rolling window procedure with window length $h = 100$ to examine the dynamic movement of the market efficiency. Due to theoretical considerations, T cannot be too small. Accordingly, the rolling window h cannot be small either. Hence, we consider window length $h = 100$. For the sake of convenience, we also use $h = 100$ to study the US stock market.

For each $\tau \in \{1, \cdots, 153 - h\}$, we separately estimate CAPM and the FF three-factor model using the data from period τ to $\tau + h - 1$. As a result,

$$\begin{aligned} r_{it} - r_{ft} &= \hat{\alpha}_{it} + \hat{\beta}_{it}(r_{mt} - r_{ft}) + \hat{e}_{it} \quad (CAPM)\ ; \\ r_{it} - r_{ft} &= \hat{\alpha}_{it} + \hat{\beta}_{i1t}(r_{mt} - r_{ft}) + \hat{\beta}_{i2t} SMB_t + \hat{\beta}_{i3t} HML_t + \hat{e}_{it} \quad (FF) \end{aligned}$$

for $1 \geq t \geq \tau + h - 1$.

Based on the estimated residuals $\hat{e}_{it}$ obtained by separately fitting CAPM and the FF three-factor model to the data in each window, we calculate the HDA test statistics and their corresponding p-values. Here, the number of interior knots n is determined via BIC discussed in Subsection 2.2, and the order of B-splines is set at 3 for all estimation windows. For the sake of comparison, we also consider the PY test (Pesaran and Yamagata, 2012). The p-values across the 53 (300) windows obtained from the HDA and PY tests by testing the market efficiency of the Chinese (US) stock market based on CAPM and the FF three-factor model are, respectively, presented in Figures 2 and 3, while the descriptive statistics of these p-values are given in Table 4.

For the Chinese market, the left panel in Figure 2 depicts the p-values of the HDA and PY tests for CAPM across the 53 window periods, while the right panel is for FF. According to Figure 2 and Table 4, PY shows a similar pattern to HDA in explaining that the majority of the p-values from FF are larger than those from CAPM. However, these two tests can lead to very different conclusions in terms of market efficiency for some periods under our scrutiny. For example, the left panel in Figure 2 shows that, for 19 window periods (2, 4, 5, 7, 8, 9, 11, 13, 14, 16, 33, 34, 36, 40, 41, 43, 44, 49 and 50), the p-values obtained from HDA in the CAPM model are less than 5%; this indicates that the markets are inefficient over these window periods. In contrast, the p-values obtained from PY in the corresponding window periods are greater than 5%.

We next conduct the analysis for the US stock market data. Panels A and B in Figure 3 depict the p-values of the HDA and PY tests for CAPM and FF, respectively, across the 300 window periods. To highlight the difference between HDA and PY, Panels C and D present the p-values for the sub-window periods ranging from 101 to 150. Table 4 indicates that the averaged p-values obtained from HDA are smaller than those from PY; this can also be seen in Figure 3. In particular, panel D in Figure 3 suggests that, for 8 window periods (108, 110, 111, 112, 134, 135, 149 and 150), the p-values obtained from HDA in the FF model are less than 5%. Accordingly, the markets are inefficient in these window periods. On the other hand, the p-values obtained from PY in those corresponding periods are all greater than 5%.

It is also worth noting that the alphas and betas are time-varying, as confirmed by the

CC test in Section 5.2. This, together with the above discussion, implies that HDA is a better and more capable approach than PY to detecting mean-variance inefficiency in these two markets.

In addition to compare HDA and PY, we note from Figures 2 and 3 that most of the p-values from the FF three-factor model are larger than the 5% significance level, and they are also higher than those from CAPM. Hence, the FF three-factor model is better than CAPM in explaining the mean-variance efficiency of both the Chinese and US stock market. However, in the US stock market these findings are even more prominent than in the Chinese stock market. This suggests that the US stock market is more efficient than the Chinese stock market in terms of mean-variance efficiency. This finding is also confirmed by Table 4; the mean of the p-values from the FF three-factor model obtained for the US stock market is much larger than that for the Chinese stock market.

For robustness check of HDA, we further consider a short window of length $h = 60$ for both the Chinese and US stock market data as suggested by Pesaran and Yamagata (2012) and a relatively longer window of length $h = 200$ for the US stock market data; the results yield a similar pattern to that with $h = 100$. To save space, we present them in the supplementary materials. Finally, although we believe our findings make a contribution to the literature, one avenue of further research would be extending them to account for survival bias when there are missing observations (see Brown et al., 1995).

6 Conclusion

In this paper, we propose the HDA test to examine the market efficiency in conditional time-varying factor models. We also introduce the CC test to assess the constant alphas and factor loadings. Monte Carlo studies demonstrate that both tests perform satisfactorily and the numerical results also support theoretical findings. Moreover, the usefulness of these two tests is illustrated by two empirical examples.

To further broaden the usefulness of our proposed tests, we conclude this article by identifying the following possible research avenues. First, if the goal is to identify the significance of alphas for all possible assets, then one can apply our HDA test in a multiple testing procedure to control the false discovery rate (see, e.g., Benjamini and Hochberg, 1995; Fan et al., 2012). Second, our CC test can be extended to simultaneously test the significance of constant coefficients across all assets. Third, although we only consider CAPM and the Fama-French three-factor model in our applications, the proposed tests can be applied to other factor models with fixed d such as the Fama-French five-factor model (Fama and French, 2015). One can also extend out test to allow the number of factors d to increase with the number of time series observations, T, if necessary. Fourth, our HDA test is valid even for $N \to T$, but we do require that T tends to infinity in order to establish the desired asymptotic theory. For practical reasons, it is also useful to derive a test for finite T. Lastly, one can extend the proposed high dimensional testing procedure to Black's CAPM (Black, 1972), as suggested by an anonymous referee.

7 Supplementary Materials

The online supplementary appendix contains the proofs of the main results in the paper and some additional results for the simulation study and empirical analyses.

References

Aboody, D. and Lev, B. (2000). "Information asymmetry, R&D, and insider gains," *The Journal of Finance*, 55, 2747–2766.

Abosede, A. J. and Osezi, J. E. (2011). "Theoretical analysis of firm and market-specific proxies of information symmetry on equity prices in the stock markets, " *Australian Journal of Business and Management Research*, 1, 1–13.

Ang, A. and Kristensen, D. (2012). "Testing conditional factor models," *Review of Financial Studies*, 106, 132–156.

Ang, A. and Chen, J. (2007). "CAPM over the long run, 1928–2001," *Journal of Empirical Finance*, 14, 1–40.

Bai, Z. and Saranadasa, H. (1996). "Effect of high dimension: by an example of a two sample problem," *Statistica Sinica*, 6, 311–329.

Beaulieu, M. C., Dufour, J. M. and Khalaf, L. (2007). "Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors: an exact simulation-based approach," *Journal of Business and Economic Statistics*, 25, 398–410.

Benjamini, Y. and Hochberg, Y. (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing, " *Journal of the Royal Statistical Society, Series B*, 57, 289–300.

Black, F. (1972). "Capital market equilibrium with restricted borrowing, " *The Journal of Business*, 45, 444–455.

Brown, S., Goetzmann, W. and Ross, S. (1995). "Survival, " *The Journal of Finance*, 50, 853–873.

Chen, B. and Hong, Y. (2012). "Testing for smooth structural changes in time series models via nonparametric regression," *Econometrica,* 80, 1157–1183.

Chen, S. and Qin, Y. (2010). "A two-sample test for high-dimensional data with applications to gene-set testing," *Annals of Statistics,* 38, 808-835.

Chow, G. C. (1960). "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28, 591–605.

Dai, Y., Kong D. and Wang, L. (2013). "Information asymmetry, mutual funds and earnings management: Evidence from China," *China Journal of Accounting Research*, 6, 187–209.

de Boor, C. (2001). *A practical guide to splines, Vol. 27 of applied mathematical sciences*, Springer-Verlag, New York.

Fama, E. F. and French, K. R. (1993). "Common risk factors in the return on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

Fama, E. F. and French, K. R. (2015). "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

Fan, J. and Li, R. (2001). "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, 96, 1348–1360.

Fan, J., Liao, Y. and Mincheva, M. (2011). "High dimensional covariance matrix estimation in approximate factor model," *Annals of Statistics*, 39, 3320–3356.

Fan, J., Han, X. and Gu, W. (2012). "Estimating false discovery proportion under arbitrary covariance dependence, "*Journal of the American Statistical Association,* 107, 1019–1035.

Fan, J., Liao, Y. and Yao, J. (2015). "Power enhancement in high dimensional cross-sectional tests," *Econometrica*, 83, 1497–1541.

Fan, J., Zhang, C. and Zhang, J. (2001). "Generalized likelihood ratio statistics and Wilks phenomenon," *Annals of Statistics*, 29, 153–193.

Gibbons, M. R., Ross, S. A. and Shanken, J. (1989). "A test of the efficiency of a given portfolio," *Econometrica*, 57, 1121–1152.

Goeman, J. J., Van Houwelingen, H. C. and Finos, L. (2011). "Testing against a high-dimensional alternative in the generalized linear model: asymptotic type I error control," *Biometrika*, 98, 381–390.

Guo, B. and Chen, S. (2016). "Tests for high dimensional generalized linear models," *Journal of the Royal Statistical Society: Series B*, forthcoming.

He, X. and Shi, P. D. (1996). “Bivariate tensor-product B-splines in a partly linear model, ” *Journal of Multivariate Analysis*, 58, 162-181.

Jensen, M. C. (1968). “The performance of mutual funds in the period 1945-1964,” *Journal of Finance*, 23, 389–416.

Lan, W., Wang, H. and Tsai, C.-L. (2014). “Testing covariates in high-dimensional regression,” *Annals of the Institute of Statistical Mathematics*, 66, 279–301.

Lee, C. (2001). “Market efficiency and accounting research: a discussion of ‘capital market research in accounting’ by S.P. Kothari,” *Journal of Accounting and Economics*, 31, 233–253.

Lewellen, J. and Nagel, S. (2006). “The conditional CAPM does not explain asset pricing anomalies,” *Journal of Financial Economics*, 82, 289–314.

Li, Y. and Yang, L. (2011). “Testing conditional-factor models: A nonparametric approach,” *Journal of Empirical Finance*, 18, 972–992.

Lintner, J. (1965). “The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets,” *Review of Economics and Statistics*, 47, 13–37.

Ma, S., Liang, H. and Tsai, C. L. (2014). “Partially linear single index models for repeated measurements,” *Journal of Multivariate Analysis*, 130, 354–375.

Ma, S. and Song, P. X. K. (2015). “Varying index coefficient models,” *Journal of the American Statistical Association*, 110, 341–356.

MacKinlay A. C. and Richardson, M. P. (1991). “Using generalized method of moments to test mean–variance efficiency,” *Journal of Finance*, 46, 511–527.

Pesaran, M. H. and Yamagata, T. (2012) “Testing CAPM with a large number of assets,” *Working paper*, IZA, Germany.

Schumaker, L. L. (1981). *Spline Functions*, Wiley, New York.

Sharpe, W. F. (1964). “Capital asset prices: A theory of market equilibrium under conditions of risk,” *Journal of Finance*, 19, 425–442.

Srivastava, M. S. (2005). “Some tests concerning the covariance matrix in high dimensional data,” *Journal of Japan Statistical Association*, 35, 251–272.

Su, L. and Jin, S. (2012). “Sieve estimation of panel data models with cross section dependence,” *Journal of Econometrics*, 169, 34–47.

Su, L. and Wang, X. (2017). “On time-varying factor models: estimation and testing,” *Journal of Econometrics*, 198, 84-101.

Wang, L., Li, H. and Huang, J. (2008). “Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements,” *Journal of the American Statistical Association*, 103, 1556-1569.

Zhou, G. (1993). “Asset pricing test under alternative distributions,” *Journal of Finance*, 48, 1925–1942.

Table 1: The empirical sizes of the HDA and LY tests from Examples 1–3 for testing conditional alphas with a nominal level of 5%, where Normal Distribution, Exponential Distribution, and Mixture Distribution refer to the distribution from which the error term $\mathbf{E}_t$ is generated.

			Normal Distribution		Exponential Distribution		Mixture Distribution	
Example	N	T	HDA	LY-test	HDA	LY-test	HDA	LY-test
1	3	100	0.054	0.061	0.064	0.053	0.045	0.047
		200	0.056	0.052	0.044	0.054	0.048	0.052
		500	0.051	0.036	0.051	0.057	0.039	0.049
1	200	100	0.042	1	0.045	1	0.065	1
		200	0.047	1	0.066	1	0.039	1
		500	0.043	1	0.058	1	0.058	1
1	500	100	0.057	1	0.053	1	0.066	1
		200	0.058	1	0.045	1	0.050	1
		500	0.053	1	0.061	1	0.056	1
1	1000	100	0.052	1	0.064	1	0.050	1
		200	0.057	1	0.056	1	0.064	1
		500	0.051	1	0.052	1	0.053	1
2	3	100	0.055	0.049	0.044	0.061	0.062	0.048
		200	0.052	0.056	0.052	0.046	0.051	0.046
		500	0.061	0.058	0.053	0.055	0.057	0.058
2	200	100	0.065	1	0.062	1	0.045	1
		200	0.045	1	0.048	1	0.044	1
		500	0.042	1	0.037	1	0.035	1
2	500	100	0.048	1	0.065	1	0.063	1
		200	0.054	1	0.067	1	0.054	1
		500	0.051	1	0.035	1	0.046	1
2	1000	100	0.061	1	0.047	1	0.052	1
		200	0.051	1	0.053	1	0.056	1
		500	0.044	1	0.046	1	0.049	1
3	3	100	0.052	0.064	0.065	0.062	0.065	0.054
		200	0.046	0.052	0.056	0.049	0.045	0.054
		500	0.055	0.058	0.057	0.066	0.062	0.061
3	200	100	0.052	1	0.053	1	0.067	1
		200	0.047	1	0.058	1	0.066	1
		500	0.043	1	0.045	1	0.059	1
3	500	100	0.055	1	0.053	1	0.058	1
		200	0.045	1	0.062	1	0.054	1
		500	0.042	1	0.045	1	0.048	1
3	1000	100	0.061	1	0.065	1	0.056	1
		200	0.058	1	0.055	1	0.063	1
		500	0.045	1	0.040	1	0.049	1

Table 2: The empirical sizes and powers of the CC test from Examples 1–2 for testing the constancy with a nominal level of 5%, where Normal Distribution, Exponential Distribution, and Mixture Distribution refer to the distribution from which the error term $\mathbf{E}_t$ is generated.

		Normal Distribution			Exponential Distribution			Mixture Distribution		
Example	T	$b=0$	$b=0.1$	$b=0.3$	$b=0$	$b=0.1$	$b=0.3$	$b=0$	$b=0.1$	$b=0.3$
1	100	0.047	0.269	0.546	0.043	0.210	0.477	0.054	0.178	0.482
	200	0.042	0.354	0.704	0.053	0.336	0.711	0.048	0.365	0.703
	500	0.050	0.439	0.855	0.043	0.424	0.851	0.047	0.437	0.842
2	100	0.040	0.347	0.636	0.047	0.329	0.606	0.055	0.327	0.619
	200	0.042	0.494	0.779	0.052	0.448	0.802	0.053	0.447	0.776
	500	0.057	0.558	0.989	0.055	0.543	0.977	0.050	0.556	0.990

Table 3: Descriptive statistics for the Fama and French three-factor model in the Chinese and US stock markets.

Market	FF-Factors	Mean (%)	Median (%)	SD (%)
China	MF	0.24	0.29	2.37
	SMB	0.14	0.20	1.52
	HML	0.09	0.06	1.46
US	MF	0.34	0.37	1.53
	SMB	0.04	0.08	0.71
	HML	0.06	0.02	1.13

Table 4: Descriptive statistics for the p-values of HDA and PY obtained from FF three-factor model and CAPM in the Chinese and US stock markets.

		FF			CAPM		
Market	Tests	Mean(%)	Median(%)	SD(%)	Mean(%)	Median(%)	SD(%)
China	HDA	0.18	0.05	0.18	0.12	0.04	0.17
	PY	0.30	0.04	0.17	0.17	0.09	0.14
US	HDA	0.41	0.39	0.20	0.19	0.08	0.17
	PY	0.60	0.54	0.18	0.32	0.27	0.16

Figure 1: The empirical powers of the HDA test under $N = 200$ with three sample sizes $T = 100, 200$ and 500, where Panels A and C depict the power functions for Examples 1 and 2, respectively, with the dense alternative, while Panels B and D are the power functions for Examples 1 and 2, respectively, with the sparse alternative.

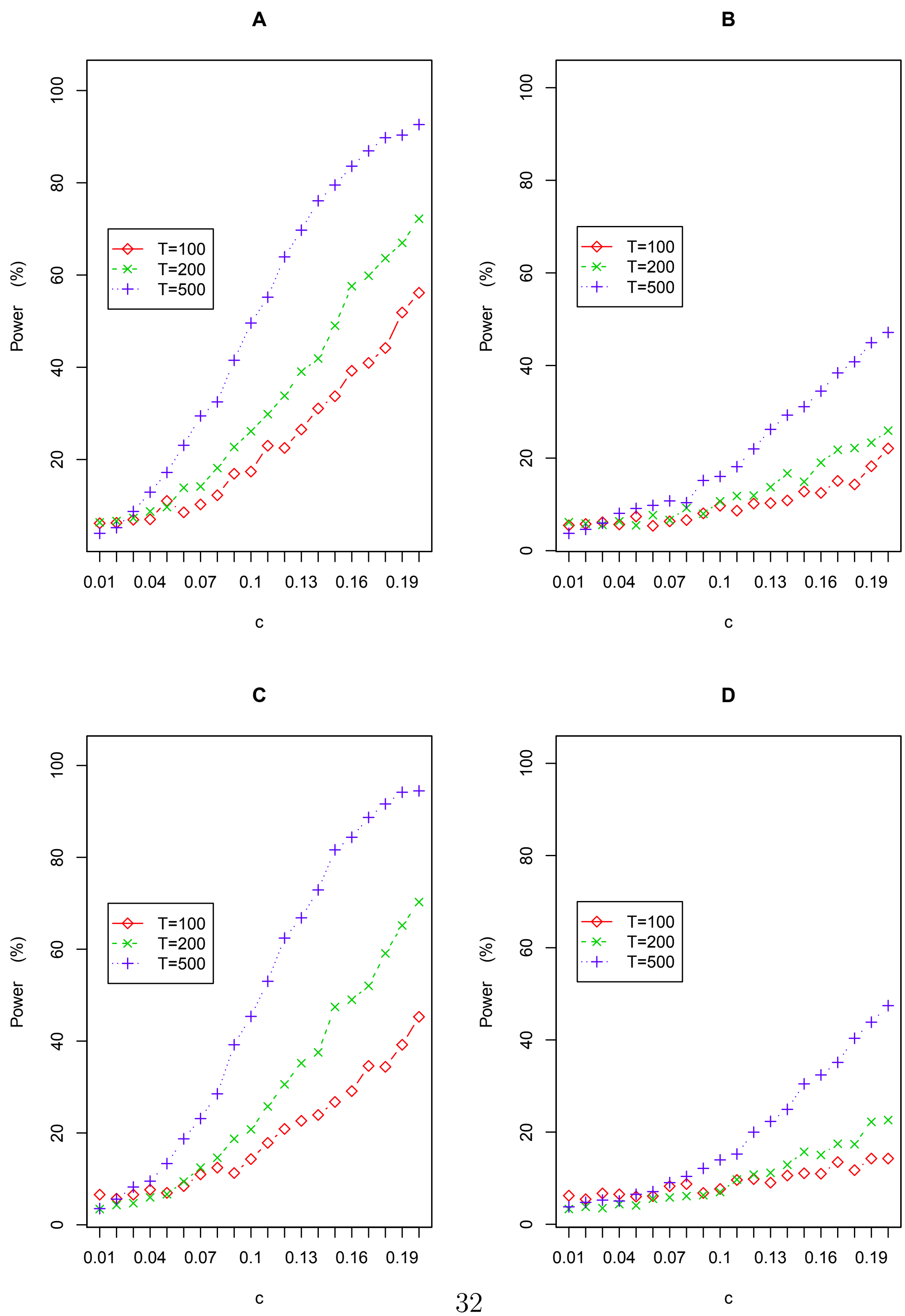

Figure 2: The dynamic movement of market efficiency in the Chinese stock market based on the p-values obtained from the HDA and PY tests by testing the conditional CAPM (left panel) and the conditional Fama-French three-factor model (right panel).

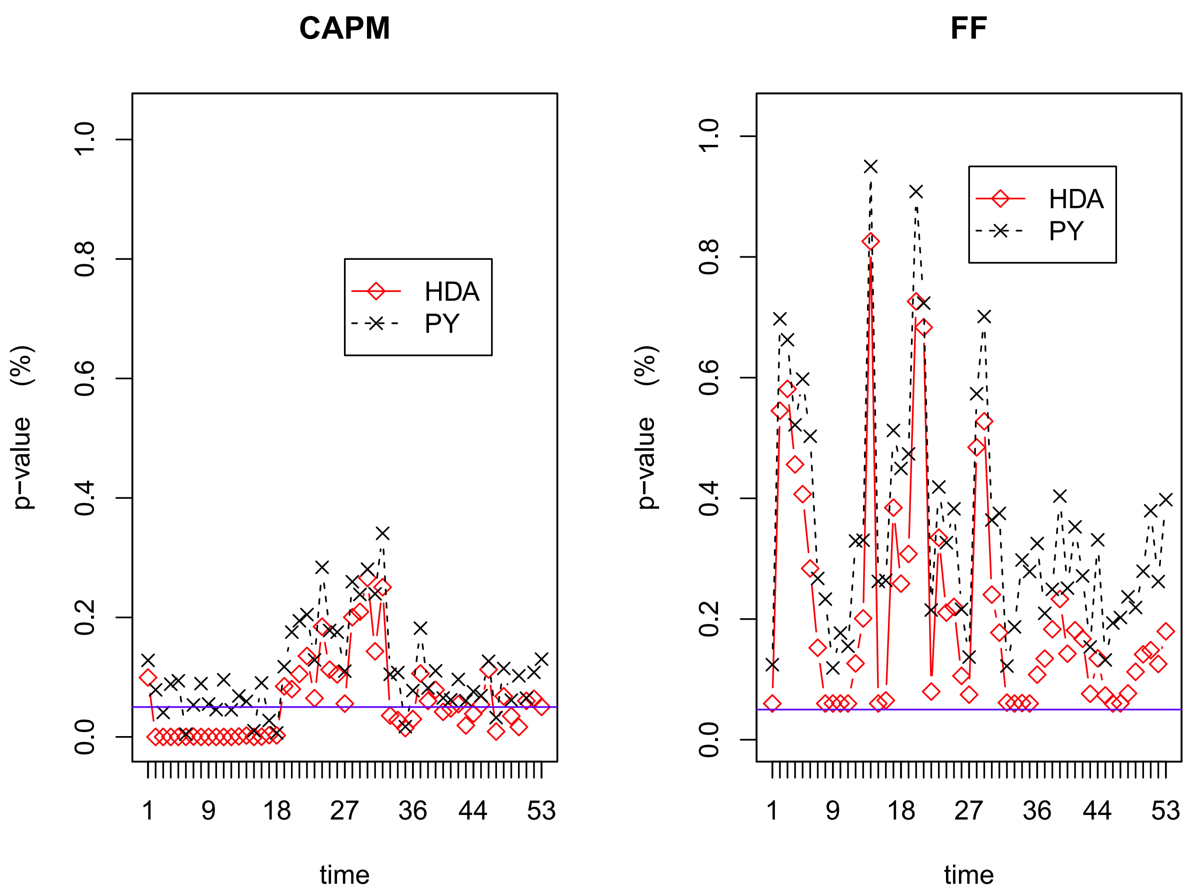

Figure 3: The dynamic movement of market efficiency in the US stock market based on the p-values obtained from the HDA and PY tests by testing the conditional CAPM (panels A and C) and the conditional Fama-French three-factor model (panels B and D). Note that panels A and B present the p-values across the whole study period, while panels C and D depict the p-values across the sub-window of periods 101 to 150.