Statistica Sinica Preprint No: SS-2022-0137	
Title	Maximum Conditional Alpha Test for Conditional Multi-Factor Models
Manuscript ID	SS-2022-0137
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0137
Complete List of Authors	Jun Zhang, Wei Lan, Xinyan Fan and Wen Chen
Corresponding Authors	Xinyan Fan
E-mails	1031820039@qq.com
Notice: Accepted version subject to English editing.	

Statistica Sinica

MAXIMUM CONDITIONAL ALPHA TEST FOR CONDITIONAL MULTI-FACTOR MODELS

Jun Zhang, Wei Lan, Xinyan Fan and Wen Chen

Southwestern University of Finance and Economics,

Renmin University of China and Nanjing Agricultural University

Abstract: In this paper, a novel test, called the maximum conditional alpha (MCA) test, which enhances the testing power for detecting alpha in linear multi-factor models, is proposed. This test is specifically designed for conditional multi-factor models with time-varying coefficients, where the number of test assets (N) exceeds the number of observations (T) and the alternative hypothesis is a sparse vector, meaning that only a few components violate the null hypothesis. By carefully studying the estimation error derived from the B-spline estimation, we rigorously demonstrate that the proposed test converges to a type-I extreme value distribution when $\min(T, N)$ tends to infinity, subject to mild conditions. Furthermore, the proposed MCA test was extended to incorporate latent factors within conditional multi-factor models. The small-sample properties of the

Corresponding author: Xinyan Fan, Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing 100086, China. E-mail: 1031820039@qq.com. Wen Chen, College of Finance, Nanjing Agricultural University, Nanjing 210095, China. E-mail: chenwenpku2008@163.com.

proposed MCA test were assessed via Monte Carlo simulations. Finally, the proposed method was applied to evaluate the efficiency of the U.S. stock market using the conditional Fama-French three-factor model. The results demonstrate that the MCA test outperforms existing tests in terms of statistical power.

Key words and phrases: B-Spline Estimator, Maximum Conditional Alpha Test, Sparse Alternative, Time-Varying Coefficient.

1. Introduction

Explaining the variations in average returns across different assets is a fundamental question in finance. The capital asset pricing model (CAPM), pioneered by Sharpe (1964) and Lintner (1965), has long been the cornerstone of asset pricing. However, CAPM has traditionally been the key framework for asset pricing. However, the CAPM has proven to be inadequate, leading to the development of alternative multi-factor models like the widely used three-factor model proposed by Fama and French Fama and French (1993). Typically, each factor in these multi-factor models has significant economic meaning and pricing ability.

In these models, denoting the excess return of asset i at time t as R_{it} and the $d \times 1$ observable vector of common factors as $\mathbf{f}_t = (f_{1t}, \ldots, f_{dt})^\top \in \mathbb{R}^d$, the linear multi-factor model with N test assets takes the following form:

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i^\top \mathbf{f}_t + \varepsilon_{it}, \tag{1.1}$$

for $i = 1, \cdots, N$, $t = 1, \cdots, T$, where $\boldsymbol{\beta}_i = (\beta_{i1}, \cdots, \beta_{id})^\top \in \mathbb{R}^d$ is a vector of factor loadings of asset i, and ε_{it} is the corresponding idiosyncratic error term. The intercept term α_i in (1.1) captures the excess return of the i-th asset. In financial investment, the focus lies on the value of α_i rather than R_{it}. According to the "mean-variance efficiency" theory, if the linear multi-factor model is correctly specified, the intercept α_i for any test asset i should be zero, indicating that all excess returns can be explained by the common factors. This leads to the question of how to assess the adequacy of a specific linear multi-factor model.

This question is essentially about testing whether the regression intercepts α_is are zeros. The hypotheses are as follows:

$$H_0 : \alpha_i = 0, \text{ for all } i = 1, \cdots, N, \text{ vs. } H_1 : \alpha_i \neq 0, \text{ for some } i = 1, \cdots, N. \tag{1.2}$$

The Gibbons-Ross-Shanken (GRS) test, introduced by Gibbons et al. (1989), provides a pioneering method to address this problem. The GRS test is a Wald-type test for hypotheses (1.2), formulated as follows:

$$\text{GRS} = \frac{T-N-d}{N} \frac{1}{1 + \bar{\mathbf{f}}^\top \widehat{\boldsymbol{\Sigma}}_{\mathbf{f}}^{-1} \bar{\mathbf{f}}} \widehat{\boldsymbol{\alpha}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\alpha}}, \tag{1.3}$$

where $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_1, \cdots, \widehat{\alpha}_N)^\top \in \mathbb{R}^N$ is the ordinary least-square (OLS) estimator of intercepts, $\widehat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix of $\widehat{\boldsymbol{\varepsilon}}_{(t)}$, $\widehat{\boldsymbol{\varepsilon}}_{(t)} = (\widehat{\varepsilon}_{1t}, \cdots, \widehat{\varepsilon}_{Nt})^\top \in \mathbb{R}^N$, $\widehat{\varepsilon}_{it}$ is the estimated residual for each $i \in \{1, \cdots, N\}$ and $t \in \{1, \cdots, T\}$, $\bar{\mathbf{f}} = T^{-1} \sum_{t=1}^{T} \mathbf{f}_t$ is the sample mean vector of factors, and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{f}} = T^{-1} \sum_{t=1}^{T} (\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})^\top$ is the estimated covariance matrix of $\mathbf{f}_t$. This approach has been widely studied (e.g., MacKinlay and Richardson, 1991; Zhou, 1993; Lan et al., 2018; Ma et al., 2020).

Although the GRS test is commonly used to address the above mentioned question, it has certain limitations. First, the GRS type test is only applicable when the number of assets N is fixed and much less than the number of observations T (e.g., Pesaran and Yamagata, 2017; Lan et al., 2018; Feng et al., 2022). This limitation arises from the inconsistency and nontrivial bias introduced by the sample covariance matrix estimator when $N > T$ (Bai and Silverstein, 2005). Second, the GRS test is not applicable to conditional time-varying multi-factor models. However, evidence shows that both the alphas (pricing errors) and betas (factor loadings) vary over time (e.g., Ghysels, 1998; Guo et al., 2017; Cooper and Maio, 2019). As Ghysels (1998) pointed out, possible misspecifications arise owing to the strong assumptions placed on the underlying probability distributions and the investors' attitude toward risk in the above mentioned time-invariant multi-factors models. Therefore, numerous studies have focused on estimating the CAPM model with time-varying alphas and betas. A partial list includes Ferson and Harvey (1999), Li and Yang (2011), Ang and Kristensen (2012) and Ma et al. (2020). In fact, the conditional time-varying multi-factor model is more suitable than the traditional time-invariant multi-factor models. For example, the empirical results in the appendix show that the null hypothesis of time-invariant alphas and betas is rejected in each of the 312 (334) portfolios at a significant level of 0.05, approximately 93.4%. Third, the GRS test suffers from low power when the alternative hypothesis H_1 in (1.2) is sparse, i.e., under H_1, only a small proportion of α_i's are far from zero. The main reason is that the sum-of-squares type statistics accumulate high-dimensional estimation errors under H_0, which leads to a large critical value that can dominate the signals in the sparse alternatives (e.g., see Fan et al., 2015 for detailed explanations). In practical applications, the alternative hypotheses are not always dense. For example, we present empirical evidence of sparse alternatives based on a total of 334 portfolios in the appendix, there are less than 10 significant nonzero-alpha in the time interval 1999-2020, approximately 3%, i.e., the signals in α_i spread out over a small number of test assets.

Some methods attempt to overcome the aforementioned three limitations since they are all empirically motivated. To resolve the first limitation, Pesaran and Yamagata (2017) developed two innovative Wald-type tests for alphas when $N > T$. Lan et al. (2018) proposed a novel technique based on random projections to project the N-dimensional test assets into a low-dimensional space of dimension $k \leq \min\{N, T\}$. However, these methods assume constant alphas and betas over time and are of the sum-of-squares type, limiting their applicability against sparse alternatives. To address the second limitation, Ma et al. (2020) proposed the high-dimensional alpha (HDA) test, allowing for time-varying alphas and betas and accommodating N significantly larger than T. However, the HDA test is constrained by sparse alternatives, resolving only the first two limitations. To tackle the third limitation, Fan et al. (2015) introduced a power enhancement screening procedure for increased test power against sparse alternatives, applicable to high-dimensional test assets. Feng et al. (2022) developed a max-of-squares-type test to assess the mean-variance efficiency. Testing based on the maximum is known to be highly effective when dealing with sparse alternatives and can also be applied when $N > T$. However, these tests assume constant alphas and betas. To summarize, while the tests proposed by Fan et al. (2015) and Feng et al. (2022) can address the first and third limitations, they are unable to overcome the second limitation. Consequently, none of the existing methods can overcome the three limitations simultaneously. Accordingly, we propose a novel max-of-squares type test, called the maximum conditional alpha (MCA) test, to address the three limitations simultaneously.

The main contributions of the proposed MCA test can be summarized as follows. First, it enables testing in a high-dimensional setting where $N > T$. Second, it accommodates time-varying alphas and betas in conditional multi-factor models. To estimate these time-varying factor models, we assume that both alphas and betas are unknown smooth functions of time t. We utilize the B-spline method to estimate the conditional alphas and betas. The detailed procedures for establishing the test statistics are discussed in Section 2.3. Third, MCA is a max-of-squares-type test, which can be applied when the alternative hypothesis is sparse. By carefully studying the estimation error generated from the B-spline estimation, we theoretically demonstrate that our test converges to the type-I extreme value distribution as $\min(T, N) \to \infty$ under proper conditions. Finally, we extend the MCA test to incorporate conditional multi-factor models with a latent structure and establish its theoretical properties. The advantages of the proposed test over existing methods are confirmed via simulations and empirical applications.

The remainder of this paper is organized as follows. Section 2 introduces methods for estimating the conditional multi-factor model and presents the proposed MCA test. The MCA null distribution and power properties are also discussed in Section 2. Section 3 reports the maximum conditional alpha test with latent factors. Section 4 presents the Monte Carlo simulations performed to examine the small-sample properties of the proposed MCA test. Finally, Section 5 provides the concluding remarks. All technical details are presented in the Supplementary Material.

2. Maximum Conditional Alpha Test

This section includes four subsections. In Section 2.1, we introduce the conditional multi-factor model and our hypotheses. In Section 2.2, we describe the B-spline method for estimating the conditional time-varying alphas and betas. In Section 2.3, we propose the MCA test. In Section 2.4, we study the asymptotic power of the proposed MCA test and theoretically compare the power of MCA with that of the HDA test proposed by Ma et al. (2020).

2.1 Conditional Multi-Factor Models and Hypotheses

We consider the following conditional multi-factor model:

$$R_{it} = \alpha_{it} + \boldsymbol{\beta}_{it}^{\top}\mathbf{f}_t + \varepsilon_{it} = \alpha_{it} + \sum_{j=1}^{d} \beta_{ijt} f_{jt} + \varepsilon_{it}, \qquad (2.4)$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, where R_{it} is the excess return of test asset i at time t, α_{it} is the conditional alpha for test asset i at time t, $\mathbf{f}_t = (f_{1t}, \cdots, f_{dt})^\top \in \mathbb{R}^d$ is the $d \times 1$ observable vector of common factors with fixed d, and $\boldsymbol{\beta}_{it} = (\beta_{i1t}, \cdots, \beta_{idt})^\top \in \mathbb{R}^d$ is the $d \times 1$ vector of conditional betas. In addition, $\boldsymbol{\varepsilon}_{(t)} = (\varepsilon_{1t}, \cdots, \varepsilon_{Nt})^\top \in \mathbb{R}^N$ is the idiosyncratic error term that is independent and identically distributed with mean zero and covariance matrix $\boldsymbol{\Sigma}$. To identify the parameters in model (2.4), following Cai (2007), Li and Yang (2011), and Ma et al. (2020), we assume that the conditional alphas and betas are generated from two smoothing functions of $\alpha_i(\cdot)$ and $\beta(\cdot)$, that is, $\alpha_{it} = \alpha_i(t/T)$, and $\beta_{ijt} = \beta_{ij}(t/T)$, respectively.

To assess the performance of a conditional multi-factor model, the natural null hypothesis is that the conditional alphas for any asset i at each time t are equal to zero; that is, $H_0 : \alpha_{it} = 0$ for any $i = 1, \cdots, N$ and $t = 1, \cdots, T$. However, Li and Yang (2011) pointed out that this null hypothesis is overly restrictive and can lead to easy rejection. Accordingly, following Lewellen and Nagel (2006), Li and Yang (2011), Ang and Kristensen (2012), and Ma et al. (2020), we test whether the average conditional alpha for asset i across all time periods is equal to zero. In other words, the null hypothesis can be stated as $H_0 : T^{-1}\sum_{t=1}^{T}\alpha_{it} = 0$ for any $i = 1, \cdots, N$. We denote the average conditional alpha and beta as $\alpha_{i,ACA} = T^{-1}\sum_{t=1}^{T}\alpha_{it}$

and $\beta_{ij,ACA} = T^{-1}\sum_{t=1}^{T}\beta_{ijt}$, respectively. Then, we can rewrite model (2.4) as

$$R_{it} = \alpha_{i,ACA} + \widetilde{\alpha}_i(t/T) + \sum_{j=1}^{d}\beta_{ij}\left(t/T\right)f_{jt} + \varepsilon_{it}, \tag{2.5}$$

where $\widetilde{\alpha}_i(t/T) = \alpha_i(t/T) - \alpha_{i,ACA}$. Then, the null and alternative hypotheses for testing the average conditional alphas across the N test assets are, respectively,

$$H_0 : \alpha_{i,ACA} = 0, \text{ for all } i = 1, \cdots, N, \text{ vs. } H_1 : \alpha_{i,ACA} \neq 0, \text{ for some } i = 1, \cdots, N. \tag{2.6}$$

To construct the test statistic for testing (2.6), it is necessary to estimate $\alpha_{i,ACA}$,$\widetilde{\alpha}_i(t/T)$ and $\beta_{ij}\left(t/T\right)$ involved in model (2.5), and we discuss their estimation procedure in detail in the next section.

2.2 B-spline Estimation

In this section, we employ the B-spline method to estimate the conditional alphas and betas involved in model (2.5). The B-spline basis functions are chosen because their computational efficiency and numerical stability in finite samples are higher than those of other basic functions, such as the truncated power series and the trigonometric series. Let $0 = \zeta_0 < \zeta_1 < \cdots < \zeta_n < 1 = \zeta_{n+1}$ be a partition of $[0, 1]$ divided into subintervals

$I_l = [\zeta_l, \zeta_{l+1}]$ for $0 \le l \le n-1$ and $I_n = [\zeta_n, \zeta_{n+1}]$ that satisfy

$$\max_{0\le l\le n} |\zeta_{l+1} - \zeta_l| / \min_{0\le l\le n} |\zeta_{l+1} - \zeta_l| \le \widetilde{m},$$

for some constant $0 < \widetilde{m} < \infty$, where $\{\zeta_l\}_{l=1}^n$ is a sequence of interior knots, ζ_0 and ζ_{n+1} are the two end points, and n is the number of interior knots. For any $t = 1, \cdots, T$, its location is defined as $l(t)$ and satisfies $\zeta_{l(t)} \le t/T < \zeta_{l(t)+1}$. Consider the space of polynomial splines of order q on [0, 1] as $\mathbf{B}(t/T) = \{B_1(t/T), \cdots, B_L(t/T)\}^\top \in \mathbb{R}^L$. Then, denote the normalized B-spline basis of this space as $\widetilde{B}_l(t/T) = B_l(t/T) - T^{-1}\sum_{t=1}^T B_l(t/T)$ and $\widetilde{\mathbf{B}}(t/T) = \left\{\widetilde{B}_1(t/T), \cdots, \widetilde{B}_L(t/T)\right\}^\top \in \mathbb{R}^L$ (e.g., De Boor, 1978; Schumaker, 1981). Here, L is related to the number of interior knots n through $L = n + q$, and q is the spline degree.

According to Ma and Song (2015), Guo et al. (2017), and Ma et al. (2020), the unknown functions $\widetilde{\alpha}_i(t/T)$ and $\beta_{ij}(t/T)$ can be well approximated by the B-spline functions as

$$\widetilde{\alpha}_i(t/T) \approx \boldsymbol{\gamma}_{i0}^\top \widetilde{\mathbf{B}}(t/T) \text{ and } \beta_{ij}(t/T) \approx \boldsymbol{\gamma}_{ij}^\top \mathbf{B}(t/T), \tag{2.7}$$

for any $j = 1, \cdots, d$, where $\boldsymbol{\gamma}_{i0} \in \mathbb{R}^L$ and $\boldsymbol{\gamma}_{ij} \in \mathbb{R}^L$ are the coefficients of the B-spline functions. Substituting expression (2.7) into model (2.5), we have

$$R_{it} \approx \alpha_{i,ACA} + \boldsymbol{\gamma}_{i0}^\top \widetilde{\mathbf{B}}(t/T) + \sum_{j=1}^d \boldsymbol{\gamma}_{ij}^\top \mathbf{B}(t/T) f_{jt} + \varepsilon_{it}, \tag{2.8}$$

and we can estimate $\alpha_{i,ACA}$, $\boldsymbol{\gamma}_{i0}$, and $\boldsymbol{\gamma}_{ij}$ by minimizing

$$\mathcal{L}\left(\alpha_{i,ACA}, \boldsymbol{\gamma}_{i0}, \boldsymbol{\gamma}_{ij}\right) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left\{R_{it} - \alpha_{i,ACA} - \boldsymbol{\gamma}_{i0}^{\top}\widetilde{\mathbf{B}}\left(t/T\right) - \sum_{j=1}^{d}\boldsymbol{\gamma}_{ij}^{\top}\mathbf{B}\left(t/T\right) f_{jt}\right\}^{2}.$$

Hereafter, we denote them as $\widehat{\alpha}_{i,ACA}$, $\widehat{\boldsymbol{\gamma}}_{i0}$, and $\widehat{\boldsymbol{\gamma}}_{ij}$, respectively.

To obtain the analytical expressions of $\widehat{\alpha}_{i,ACA}$, $\widehat{\boldsymbol{\gamma}}_{i0}$, and $\widehat{\boldsymbol{\gamma}}_{ij}$, we rewrite model (2.8) in matrix form as follows:

$$\mathbf{R}_i \approx \alpha_{i,ACA}\mathbf{1}_T + \mathbb{Z}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{2.9}$$

where $\mathbf{R}_i = \left(R_{i1}, \cdots, R_{iT}\right)^{\top} \in \mathbb{R}^T$, $\mathbb{Z} = \left(\mathbf{Z}_1, \cdots, \mathbf{Z}_T\right)^{\top} \in \mathbb{R}^{T\times(1+d)L}$ and $\mathbf{Z}_t = \left(Z_{tk}, 1 \le k \le (d+1)L\right)^{\top} = \left\{\widetilde{\mathbf{B}}\left(t/T\right)^{\top}, \mathbf{f}_t^{\top} \otimes \mathbf{B}\left(t/T\right)^{\top}\right\}^{\top} \in \mathbb{R}^{(1+d)L}$. In addition, $\boldsymbol{\gamma} = \left(\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_N\right) \in \mathbb{R}^{(1+d)L\times N}$ where $\boldsymbol{\gamma}_i = \left(\boldsymbol{\gamma}_{ij}^{\top}, 0 \le j \le d\right)^{\top} \in \mathbb{R}^{(1+d)L}$, and $\mathcal{E} = \left(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_N\right) \in \mathbb{R}^{T\times N}$ where $\boldsymbol{\varepsilon}_i = \left(\varepsilon_{i1}, \cdots, \varepsilon_{iT}\right)^{\top} \in \mathbb{R}^T$. Based on model (2.9), using the traditional OLS estimation method, we minimize $\mathcal{L}\left(\alpha_{i,ACA}, \boldsymbol{\gamma}_{i0}, \boldsymbol{\gamma}_{ij}\right)$ and obtain

$$\widehat{\alpha}_{i,ACA} \approx \left(\mathbf{1}_T^{\top}\mathbf{M}_{\mathbb{Z}}\mathbf{1}_T\right)^{-1}\mathbf{1}_T^{\top}\mathbf{M}_{\mathbb{Z}}\mathbf{R}_i \text{ and } \widehat{\boldsymbol{\alpha}}_{ACA}^{\top} \approx \left(\mathbf{1}_T^{\top}\mathbf{M}_{\mathbb{Z}}\mathbf{1}_T\right)^{-1}\mathbf{1}_T^{\top}\mathbf{M}_{\mathbb{Z}}\mathbf{R},$$

$$\widehat{\boldsymbol{\gamma}}_i = \left(\mathbb{Z}^{\top}\mathbf{M}_{\mathbf{1}_T}\mathbb{Z}\right)^{-1}\mathbb{Z}^{\top}\mathbf{M}_{\mathbf{1}_T}\mathbf{R}_i \text{ and } \widehat{\boldsymbol{\gamma}} = \left(\mathbb{Z}^{\top}\mathbf{M}_{\mathbf{1}_T}\mathbb{Z}\right)^{-1}\mathbb{Z}^{\top}\mathbf{M}_{\mathbf{1}_T}\mathbf{R},$$

where $\widehat{\boldsymbol{\alpha}}_{ACA} = \left(\widehat{\alpha}_{1,ACA}, \cdots, \widehat{\alpha}_{N,ACA}\right)^{\top} \in \mathbb{R}^N$, $\mathbf{R} = \left(\mathbf{R}_1, \cdots, \mathbf{R}_N\right) \in \mathbb{R}^{T\times N}$, $\mathbf{M}_{\mathbb{Z}} = \mathbf{I}_T - \mathbb{Z}\left(\mathbb{Z}^{\top}\mathbb{Z}\right)^{-1}\mathbb{Z}^{\top}$, and $\mathbf{M}_{\mathbf{1}_T} = \mathbf{I}_T - \mathbf{1}_T\left(\mathbf{1}_T^{\top}\mathbf{1}_T\right)^{-1}\mathbf{1}_T^{\top}$, with $\mathbf{1}_T = (1, \cdots, 1)^{\top} \in \mathbb{R}^T$ being a vector of dimension T whose entries are one, and $\mathbf{I}_T$ being a $T \times T$ identity matrix.

Furthermore, the estimators of $\widetilde{\alpha}_i(t/T)$ and $\beta_{ij}(t/T)$ are $\widehat{\widetilde{\alpha}}_i(t/T) = \widehat{\boldsymbol{\gamma}}_{i0}^{\top}\widetilde{\mathbf{B}}\,(t/T)$ and $\widehat{\beta}_{ij}(t/T) = \widehat{\boldsymbol{\gamma}}_{ij}^{\top}\mathbf{B}\,(t/T)$, respectively, where $\widehat{\boldsymbol{\gamma}}_{i0}$ and $\widehat{\boldsymbol{\gamma}}_{ij}$ are elements of $\widehat{\boldsymbol{\gamma}}_i$. The average conditional betas over the sample $\widehat{\beta}_{ij,ACA}$ are

$$\widehat{\beta}_{ij,ACA} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\beta}_{ijt} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{\gamma}}_{ij}^{\top}\mathbf{B}\,(t/T) = \widehat{\boldsymbol{\gamma}}_{ij}^{\top}\mathbf{B}_A,$$

for $i = 1, \cdots, N$, $j = 1, \cdots, d$. $\mathbf{B}_A = \frac{1}{T}\sum_{t=1}^{T}\mathbf{B}\,(t/T) \in \mathbb{R}^L$ is the average of the B-splines.

To ensure that our proposed method is useful in practice, it is crucial to find smoothing parameters that automatically adapt to the data both in theory and practice. The smoothing parameter L is related to the number of internal knots n through $L = n + q$, and q ($q \geq 1$) is the degree of the spline curve. Following Rice and Wu (2001), we use splines with equally spaced knots and a fixed degree q, and only choose the internal knots n. In this study, following Ma and Song (2015) and Ma et al. (2020), the order of B-splines q is set at 3 for all estimation windows and only the internal knots n are chosen. Specifically, we choose n by minimizing the Bayesian information criterion (BIC) as follows:

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

Then, the optimal number of interior knots is given by $\widehat{n} = \operatorname{argmin}_n \mathrm{BIC}(n)$.

2.3 Maximum Conditional Alpha (MCA) Test and Null Distribution

In this section, we propose the MCA test based on the estimators of conditional alphas and betas obtained from Section 2.2 and study the theoretical properties. To achieve this goal, we first introduce some notations. The operators $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and in distribution as $(N,T) \to \infty$, respectively. Let $\mathbf{D}$ denote the diagonal matrix of $\boldsymbol{\Sigma}$, and let $\Pi = (\varphi_{ij})_{N\times N} = \mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}$ denote the correlation matrix. For an $m_1 \times m_2$ matrix $\mathbf{A} = (a_{ij})$, let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote the minimum and maximum eigenvalues of matrix $\mathbf{A}$, respectively. Moreover, we denote $\|\mathbf{A}\|_\infty = \max_{1\le i\le m_1}\sum_{j=1}^{m_2}|a_{ij}|$. Let $\|B\|$ be the spectral norm of B if B is a matrix, and let ℓ_2 be the norm of B if B is a vector. For any two sequences of positive numbers $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$, we write $a_n = O(b_n)$ if there exists a constant C such that $a_n/b_n \le C$ for all n; we write $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$. Similarly, $a_n = O_p(b_n)$ if a_n/b_n is stochastically bounded, and $a_n = o_p(b_n)$, if $a_n/b_n \to_p 0$.

Next, we construct the test statistic. Note that for any $i = 1, \cdots, N$, the traditional t-test for testing $\alpha_{i,ACA} = 0$ is

$$\mathrm{T}_i = \frac{\widehat{\alpha}_{i,ACA}}{\sqrt{\widehat{Var\,(\widehat{\alpha}_{i,ACA})}}} = \frac{(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}}\mathbf{1}_T)^{1/2}\widehat{\alpha}_{i,ACA}}{\widehat{\sigma}_{ii}^{1/2}},$$

where $\widehat{Var(\widehat{\alpha}_{i,ACA})} = (\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T)^{-1} \widehat{\sigma}_{ii}$ is the variance estimate of $\widehat{\alpha}_{i,ACA}$, $\widehat{\sigma}_{ii}$ is the i-th diagonal element of the sample variance-covariance matrix $\widehat{\mathbf{\Sigma}} = T^{-1}\mathbf{R}^\top \mathbf{M}_{\widetilde{\mathbb{Z}}} \mathbf{R}$, $\widetilde{\mathbb{Z}} = (\mathbf{1}_T, \mathbb{Z}) \in \mathbb{R}^{T \times \{(1+d)L+1\}}$, and $\mathbf{M}_{\widetilde{\mathbb{Z}}} = \mathbf{I}_T - \widetilde{\mathbb{Z}} \left(\widetilde{\mathbb{Z}}^\top \widetilde{\mathbb{Z}} \right)^{-1} \widetilde{\mathbb{Z}}^\top$.

Under the null hypothesis of $\alpha_{i,ACA} = 0$, we expect T_i to be small. Further, we observe that the testing problem (2.6) is equivalent to testing $\alpha_{i,ACA} = 0$. Therefore, we expect T_i to be uniformly small across $i = 1, \cdots, N$, which motivates us to construct a max-of-squares type test as follows:

$$\mathrm{MCA} = \max_{1 \le i \le N} \mathrm{T}_i^2 = \max_{1 \le i \le N} \frac{(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T) \widehat{\alpha}_{i,ACA}^2}{\widehat{\sigma}_{ii}}.$$

Hereafter, we call this test the MCA test.

To derive the theoretical results of the proposed MCA test, we first provide some technical conditions. Let H_r denote the collection of all functions on $[0, 1]$ such that the q-th order derivative satisfies the *Hölder* condition of order ψ with $r = q + \psi$. That is, there exists a constant $C_0 \in (0, \infty)$ such that for each $\phi \in H_r$,

$$|\phi^{(q)}(u_1) - \phi^{(q)}(u_2)| \le C_0 |u_1 - u_2|^\psi,$$

for any $0 \le u_1, u_2 \le 1$. Subsequently, we assume the following:

Assumption (A.1) $\alpha_i(\cdot) \in H_r$ and $\beta_{ij}(\cdot) \in H_r$ for some $r > 3/2$.

Assumption (A.2) (i) $\boldsymbol{\varepsilon}_{(t)}$ is independent and identically distributed with mean zero and a covariance matrix $\mathbf{\Sigma} = (\sigma_{ij})_{N\times N}$; (ii) there exist two finite constants $\eta > 0$ and $\mathcal{K} > 0$ such that $\mathbb{E}\{\exp(\eta\varepsilon_{it}^2/\sigma_{ii})\} \le \mathcal{K}$ holds uniformly for $t \in \{1, \cdots, T\}$; (iii) $\{\boldsymbol{\varepsilon}_{(t)}\}_{t=1}^{T}$ and $\{\mathbf{f}_t\}_{t=1}^{T}$ are independent.

Assumption (A.3) (i) There exist constants $0 < c_f < C_f < \infty$ such that

$$c_f < \lambda_{\min}\left\{\mathbb{E}\left\{(1, \mathbf{f}_t^\top)^\top(1, \mathbf{f}_t^\top)\right\}\right\} \le \lambda_{\max}\left\{\mathbb{E}\left\{(1, \mathbf{f}_t^\top)^\top(1, \mathbf{f}_t^\top)\right\}\right\} < C_f,$$

holds uniformly for $t \in \{1, \cdots, T\}$; (ii) there exist finite positive constants a_1 and b_1 such that for any $s > 0$, $\max_{1\le j\le d} P(|f_{jt}| > s) \le \exp\{-(s/b_1)^{a_1}\}$; (iii) the process $\{\mathbf{f}_t, t \ge 1\}$ is strong mixing with mixing coefficient $\alpha(t)$. There exist positive constants a_2 with $3a_1^{-1} + a_2^{-1} > 1$ and C_α such that $\alpha(t) \le \exp(-C_\alpha t^{a_2})$.

Assumption (A.4) (i) There exists a finite positive constant c_0 such that $c_0^{-1} < \lambda_{\min}(\mathbf{\Sigma}) \le \lambda_{\max}(\mathbf{\Sigma}) < c_0$, and (ii) there exists a finite $r_1 > 0$ such that $\max_{1\le i<j\le N} |\varphi_{ij}| \le r_1 < 1$.

All assumptions are mild and reasonable. Assumption (A.1) corresponds to the smoothness assumption for the unknown functions, which is widely employed in the field of nonparametric smoothing (e.g., He and Shi, 1996). Assumption (A.2) (i) implies that ε_{it} have zero means and are serially uncorrelated such that $\mathbb{E}(\varepsilon_{it}\varepsilon_{js}) = 0$ for all i, j, and $t \neq s$. Assumption (A.2) (ii) assumes that ε_{it} has a sub-Gaussian-type tail. Assumption (A.2)

(iii) follows the similar assumptions (Assumption 3.1 (ii)) made in previous studies such as Fan et al. (2015) and Assumption (A3) (iii) of Ma et al. (2020). Assumption (A.3) (i) is a standard condition on the design matrix for regression models, commonly known as Condition (C2) in Wang et al. (2008). Assumptions (A.3) (ii) and (iii) are adapted from Assumptions 3.3 (ii) and 3.2 of Fan et al. (2011), ensuring weak correlation and satisfying the strong mixing condition among the factors $\{\mathbf{f}_1, \cdots, \mathbf{f}_T\}$ across different time periods. Assumption (A.4) combines Conditions 1 and 3 from Cai et al. (2014). Specifically, Assumption (A.4) (i), which focuses on eigenvalues, is a common assumption in high-dimensional settings. Assumption (A.4) (ii) is considered mild as it states that if $\max_{1\le i<j\le N} |\varphi_{ij}| = 1$, then $\boldsymbol{\Sigma}$ is singular.

Based on the aforementioned assumptions, we can state the following theorem.

Theorem 1. *Suppose that Assumptions (A.1)-(A.4) hold. Assume that $LT^{-1/3} = o(1)$, $L^{-r}T^{1/2}\log(N) = o(1)$, and $\log(N) = o(L^{2/3})$. Under the null hypothesis H_0, for any $x \in \mathbb{R}$ and $\min(N,T) \to \infty$, we have*

$$P\Big[MCA - 2\log(N) + \log\{\log(N)\} \le x\Big] \xrightarrow{d} \exp\left\{-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right\}.$$

Following Theorem 1, when $\min(N,T) \to \infty$, $\mathrm{MCA} - 2\log(N) + \log\{\log(N)\}$ converges to the type I extreme value distribution with the cumulative distribution function $\exp\left\{-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right\}$. Based on this limiting null distribution, with a pre-specified significance level λ, we reject the null hypothesis when

$$\text{MCA} - 2\log(N) + \log\{\log(N)\} \geq q_\lambda,$$

where $q_\lambda = -\log(\pi) - 2\log\left\{\log(1-\lambda)^{-1}\right\}$ is the $(1-\lambda)$-th quantile of the type I extreme value distribution.

2.4 Asymptotic Power

In this section, we evaluate the asymptotic power of the MCA test under the sparse alternatives. To characterize the signals of nonzero elements in $\boldsymbol{\alpha}_{ACA}$, we introduce the set

$$\mathcal{U}(c) = \left\{\boldsymbol{\alpha}_{ACA} : \max_{1\leq i\leq N} |\alpha_{i,ACA}/\sigma_{ii}^{1/2}| \geq c\sqrt{\log(N)/T}\right\}.$$

Then, we have the following result.

Theorem 2. *Suppose that Assumptions (A.1)-(A.4) hold. Assume that* $LT^{-1/3} = o(1)$, $L^{-r}T^{1/2}\log(N) = o(1)$, *and* $\log(N) = o(L^{2/3})$. *As* $\min(N,T) \to \infty$, *we have*

$$\inf_{\boldsymbol{\alpha}_{ACA}\in\mathcal{U}(2\sqrt{2}/\sqrt{c_m})} P(\Psi_\lambda = 1) \to 1,$$

where $\Psi_\lambda = I[MCA \geq 2\log(N) - \log\{\log(N)\} + q_\lambda]$, *and* c_m *is a constant defined in Lemma 4 in the Supplementary Material.*

Theorem 2 shows that the null hypothesis of (2.6) can be rejected with the probability approaching one if $\boldsymbol{\alpha}_{ACA}$ belongs to the class $\mathcal{U}(2\sqrt{2}/\sqrt{c_m})$ under the sparse alternative H_1. The asymptotic power of Ψ_λ approaches one.

In the literature, the HDA test recently proposed by Ma et al. (2020) is also suitable for conditional factor models to allow time-varying alphas and betas. To compare the powers of the MCA and HDA tests, we briefly review the HDA test. Denote $\widehat{e}_{it}$ as the t-th element of $\mathbf{M}_{\mathbb{Z}}\mathbf{R}_i$. Denote $\widehat{tr(\boldsymbol{\Sigma}^2)}$ as the bias-corrected estimator of $tr(\boldsymbol{\Sigma}^2)$ proposed by Lan et al. (2014). The HDA test uses a sum-type test based on the residuals obtained from the null model with the test statistic $\mathrm{T}_H = \widehat{J}^*_{NT}/\widehat{\sigma}_{NT}$, where $\widehat{J}^*_{NT} = J_{NT} - N^{-1}T^{-1}\sum_{i=1}^N\sum_{t=1}^T \widehat{e}_{it}^2 h_t^2$, $J_{NT} = N^{-1}T^{-1}\left(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}}\mathbf{1}_T\right)^2 \widehat{\boldsymbol{\alpha}}_{ACA}^\top \widehat{\boldsymbol{\alpha}}_{ACA}$, $h_t = 1 - \mathbf{Z}_t^\top\left(\mathbb{Z}^\top\mathbb{Z}\right)^{-1}\mathbb{Z}^\top\mathbf{1}_T$, and $\widehat{\sigma}^2_{NT} = 2N^{-2}T^{-2}\widehat{tr(\boldsymbol{\Sigma}^2)}\sum_{t\neq s}(h_t^2 h_s^2)$. To compare the two tests, we define a set of alternative hypotheses,

$$\mathcal{S}\left(k_N, \varpi\right) = \Big\{\boldsymbol{\alpha}_{ACA} \in \mathbb{R}^N : \sum_{i=1}^N I(\alpha_{i,ACA} \neq 0) = k_N, \sqrt{8\log(N)/(c_m T)}$$

$$\leq \max_{1\leq i\leq N} \alpha_{i,ACA}/\sigma_{ii}^{1/2} \leq \sqrt{8N^{\varpi}/(c_m T)}\Big\},$$

where we assume $k_N = N^p$ with $0 \leq p < 1/2$ and $\varpi < 1/2 - p$. Then, we have the following result.

Proposition 1. *Suppose that Assumptions (A.1)-(A.4) and Condition (C1)* *in Ma et al. (2020) hold. Assume that* $LT^{-1/3} = o(1)$, $L^r T^{-3/2} = o(1)$, $L^{-r}T^{1/2}N^{1/4} = o(1)$, *and* $T^{-1+\varrho}N^{1/2+\varrho}L = O(1)$ *for a small* $\varrho > 0$, *and* $N^{-1/2}\|\mathbf{\Sigma}\|_\infty = o(1)$. *If* $\boldsymbol{\alpha}_{ACA} \in \mathcal{S}(k_N, \varpi)$, *as* $\min(N,T) \to \infty$,

$$P(MCA \geq 2\log(N) - \log\{\log(N)\} + q_\lambda) \to 1, \text{ and } P\big(T_H > z_{1-\lambda}\big) \to \lambda,$$

where $z_{1-\lambda}$ *denotes the* λ*-th upper quantile of a standard normal distribution.*

Proposition 1 indicates that under the class of sparse alternatives $\mathcal{S}(k_N, \varpi)$, the HDA test would suffer from trivial power, while the MCA test has the full power. This result is expected since HDA is a sum-of-squares type test based on the sum of $\alpha_{i,ACA}^2$s, which are mainly designed for dense alternatives.

3. Maximum Conditional Alpha Test with Latent Factors

To establish the theoretical properties of the proposed MCA test, we assume that the eigenvalues of $\mathbf{\Sigma}$ are bounded from infinity. However, if the linear pricing model fails to sufficiently explain the asset returns, we can expect that $\lambda_{\max}(\mathbf{\Sigma}) \to \infty$. This is a structure that can be easily explained if the random noise $\boldsymbol{\varepsilon}_{(t)}$ admits a factor structure. That is,

$$\boldsymbol{\varepsilon}_{(t)} = \mathbf{\Lambda}^\top \mathbf{X}_t + \boldsymbol{\epsilon}_{(t)}, t = 1, \cdots, T, \tag{3.10}$$

where $\mathbf{X}_t = (X_{1t}, \cdots, X_{vt})^\top \in \mathbb{R}^v$ is the low dimension of v unknown latent factors with the identification restriction $\mathrm{Cov}(\mathbf{X}_t) = \mathbf{I}_v$; $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_N)^\top \in \mathbb{R}^{v\times N}$ are the unknown factor loadings, where $\boldsymbol{\lambda}_i = (\lambda_{i1}, \cdots, \lambda_{iv})^\top \in \mathbb{R}^v$, and $\boldsymbol{\epsilon}_{(t)} = (\epsilon_{1t}, \cdots, \epsilon_{Nt})^\top \in \mathbb{R}^N$ is the random noise independent of $\mathbf{f}_t$ and $\mathbf{X}_t$. For simplicity, we assume $\mathbf{X}_t$ follows a multivariate normal distribution and that there are no interaction effects between the explanatory variables $\mathbf{f}_t$ and the latent factors $\mathbf{X}_t$. To further model the test-specific variations, we assume that $\boldsymbol{\epsilon}_{(t)}$ is normally distributed, i.e., $\boldsymbol{\epsilon}_{(t)} \overset{d}{\sim} \mathbb{N}(\mathbf{0}, \Sigma_{\boldsymbol{\epsilon}})$ with $\Sigma_{\boldsymbol{\epsilon}} = \mathrm{diag}(\sigma_{\boldsymbol{\epsilon},11}, \cdots, \sigma_{\boldsymbol{\epsilon},NN})$. Under the multi-factor error structure of (3.10), model (2.9) can be further written as

$$\mathbf{R}_i \approx \alpha_{i,ACA}\mathbf{1}_T + \mathbb{Z}\boldsymbol{\gamma}_i + \mathbf{X}\boldsymbol{\lambda}_i + \boldsymbol{\epsilon}_i, \tag{3.11}$$

where $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_T)^\top \in \mathbb{R}^{T\times v}$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \cdots, \epsilon_{iT}) \in \mathbb{R}^T$, and let $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_N) \in \mathbb{R}^{T\times N}$. Then, we can obtain

$$\widehat{\alpha}_{i,ACA} - \alpha_{i,ACA} \approx \left(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T\right)^{-1} \mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{X}\boldsymbol{\lambda}_i + \left(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T\right)^{-1} \mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \boldsymbol{\epsilon}_i, \text{ and}$$

$$(\widehat{\boldsymbol{\alpha}}_{ACA} - \boldsymbol{\alpha}_{ACA})^\top \approx \left(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T\right)^{-1} \mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{X}\boldsymbol{\Lambda} + \left(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T\right)^{-1} \mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \boldsymbol{\epsilon}.$$

Note that $\left(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T\right)^{-1} \mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \boldsymbol{\epsilon}$ follows a multivariate normal distribution, that is,

$$\left(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T\right)^{-1/2} \mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \boldsymbol{\epsilon} \overset{d}{\sim} \mathbb{N}(0, \Sigma_{\boldsymbol{\epsilon}}).$$

To test $\alpha_{i,ACA} = 0$, we first need to remove the effects of $\mathbf{X}\boldsymbol{\lambda}_i$.

We estimate $\mathbf{X}$ and $\boldsymbol{\Lambda}$ using the principal component method (Wang, 2012). Specifically, we first extract the effect of observed factors $\mathbf{f}_t$ by regressing $\mathbf{R}_i$ on $\mathbf{f}_t$ using the B-spline method, obtaining the residual $\widehat{\mathcal{E}} = \mathbf{M}_{\mathbb{Z}}\mathbf{R}$. Next, we define $\widehat{\mu}_e$ as the e-th largest eigenvalue of $(TN)^{-1}\widehat{\mathcal{E}}\widehat{\mathcal{E}}^\top$ and $\widehat{\boldsymbol{\varrho}}_e$ as the corresponding eigenvector. Consequently, we set $\widehat{\mathbf{X}} = T^{1/2}\left(\widehat{\boldsymbol{\varrho}}_1, \cdots, \widehat{\boldsymbol{\varrho}}_{\widehat{v}}\right)$, and $\boldsymbol{\Lambda}$ can be estimated by $\widehat{\boldsymbol{\Lambda}} = \left(\widehat{\mathbf{X}}^\top\widehat{\mathbf{X}}\right)^{-1}\widehat{\mathbf{X}}^\top\widehat{\mathcal{E}}$. Based on the estimators $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\mathbf{X}}$, we can obtain the estimated random error as $\widehat{\boldsymbol{\epsilon}} = \mathbf{M}_{\widehat{\mathbf{X}}}\widehat{\mathcal{E}}$ with $\mathbf{M}_{\widehat{\mathbf{X}}} = \mathbf{I}_T - \widehat{\mathbf{X}}\left(\widehat{\mathbf{X}}^\top\widehat{\mathbf{X}}\right)^{-1}\widehat{\mathbf{X}}^\top$. Subsequently, $\sigma_{\boldsymbol{\epsilon},ii}$ can be estimated by $\widehat{\sigma}_{\boldsymbol{\epsilon},ii} = T^{-1}\widehat{\boldsymbol{\epsilon}}_i^\top\widehat{\boldsymbol{\epsilon}}_i$. Moreover, we define notations $\widetilde{\mu}_e, \widetilde{\boldsymbol{\varrho}}_e, \widetilde{\boldsymbol{\Lambda}}, \widetilde{\mathbf{X}}$, and $\widetilde{\sigma}_{\boldsymbol{\epsilon},ii}$ as the associated estimators based on extract error $\mathcal{E}$. In practice, following Wang (2012) and Ahn and Horenstein (2013), $\widehat{v}$ can be selected by maximizing the eigenvalue ratios as $\widehat{v} = \text{argmax}_{e\le\pi_{\max}}\widehat{\mu}_e/\widehat{\mu}_{e+1}$ with some pre-specified maximum possible order $\pi_{\max}$. As $\widehat{v} = v$ with probability approaching unity (Wang, 2012), we assume that $\widehat{v} = v$ in the following discussion.

Consequently, we define factor-adjusted test statistics as

$$\widetilde{\mathrm{T}}_i = \frac{\left(\mathbf{1}_T^\top\mathbf{M}_{\mathbb{Z}}\mathbf{1}_T\right)^{1/2}\widehat{\alpha}_{i,ACA} - \left(\mathbf{1}_T^\top\mathbf{M}_{\mathbb{Z}}\mathbf{1}_T\right)^{-1/2}\mathbf{1}_T^\top\mathbf{M}_{\mathbb{Z}}\widehat{\mathbf{X}}\widehat{\boldsymbol{\lambda}}_i}{\left(\widehat{\sigma}_{\epsilon,ii}\right)^{1/2}}.$$

Therefore, we expect $\widetilde{\mathrm{T}}_i$ to be uniformly small across $i = 1, \cdots, N$, which motivates us to construct an adjusted max-of-squares type test as follows:

$$\widetilde{\mathrm{MCA}} = \max_{1\le i\le N} \widetilde{\mathrm{T}}_i^2 = \frac{\left(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T\right)\left\{\widehat{\alpha}_{i,ACA} - \left(\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T\right)^{-1} \mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \widehat{\mathbf{X}} \widehat{\boldsymbol{\lambda}}_i\right\}^2}{\widehat{\sigma}_{\epsilon,ii}}.$$

To derive the theoretical results of our adjusted MCA test, we provide some additional technical conditions.

Assumption (A.5) Assume that latent factor $\mathbf{X}_t$ and $\boldsymbol{\epsilon}_{(t)}$ are independent and normally distributed for $t = 1, \ldots, T$. Additionally, $\mathbf{X}_t \stackrel{d}{\sim} \mathbb{N}(0, \mathbf{I}_v)$ and $\boldsymbol{\epsilon}_{(t)} \stackrel{d}{\sim} \mathbb{N}(0, \mathrm{diag}(\sigma_{\boldsymbol{\epsilon},11}, \cdots, \sigma_{\boldsymbol{\epsilon},NN}))$.

Assumption (A.6) There exists some positive definite matrix $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}$ of dimension v such that $N^{-1}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top = \boldsymbol{\Sigma}_\Lambda + O(N^{-1/2})$, with the eigenvalues of $\boldsymbol{\Sigma}_\Lambda$ being bounded from zero to infinity. Moreover, there exists some positive constant $\lambda_{\max}$ such that $\max_i \|\boldsymbol{\lambda}_i\|^2 \le \lambda_{\max}$. Assume that $\frac{1}{n}\sum_{i=1}^n \sigma_{\boldsymbol{\epsilon},ii} = \tilde{\sigma}_{\boldsymbol{\epsilon}_{ii}} + O(N^{-1/2})$.

Assumption (A.7) Let $\widetilde{\boldsymbol{\theta}} = (\sigma_{\boldsymbol{\epsilon},ii}\mathbf{1}_T^\top \mathbf{M}_{\mathbb{Z}} \mathbf{1}_T)^{-1/2}\mathbf{M}_{\mathbb{Z}}\mathbf{1}_T$. Assume for projection matrix $\mathbf{P}$ such that $\|\mathbf{P}_{\mathbf{X}} - \mathbf{P}\| = O_p(\sqrt{L/T})$, we have $|\widetilde{\boldsymbol{\theta}}^\top(\mathbf{P}_{\mathbf{X}} - \mathbf{P})\mathbf{b}| \le C_P\|\mathbf{P}_{\mathbf{X}} - \mathbf{P}\||\widetilde{\boldsymbol{\theta}}^\top \mathbf{P}_{\mathbf{X}}\mathbf{b}|$, for any random vector $\mathbf{b}$ and some constants $C_P < \infty$, with probability tending to 1.

Assumptions (A.5) and (A.6) can also be founded in Wang (2012). Assumption (A.7) is a technical condition to eliminate the influence of the factor estimation error. To illustrate this, let's consider an example where Assumption (A.7) holds. If $\mathbf{P} = c_p\mathbf{P}_{\mathbf{X}}$ for some constant matrix c_p, then we have $|\widetilde{\boldsymbol{\theta}}^{\top}(\mathbf{P}_{\mathbf{X}}-\mathbf{P})\mathbf{b}| = (1-c_p)|\widetilde{\boldsymbol{\theta}}^{\top}\mathbf{P}_{\mathbf{X}}\mathbf{b}| = \|(1-c_p)\mathbf{P}_{\mathbf{X}}\||\widetilde{\boldsymbol{\theta}}^{\top}\mathbf{P}_{\mathbf{X}}\mathbf{b}|$, and the Assumption (A.7) holds with $C_p = 1$.

Similarly, in the MCA test, we have the following theorem:

Theorem 3. *Suppose that Assumptions (A.1),(A.3), and (A.5) –(A.7) hold. Assume that* $LT^{-1/3} = o(1)$, $L^{-r}T^{1/2}\log(N) = o(1)$, *and* $\log(N) = o(L^{2/3})$. *Under the null hypothesis* H_0, *for any* $x \in \mathbb{R}$ *and* $\min(N,T) \to \infty$, *we have*

$$P\Big[\widetilde{MCA} - 2\log(N) + \log\{\log(N)\} \le x\Big] \xrightarrow{d} \exp\left\{-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right\}.$$

Compared with Theorem 1, Theorem 3 allows $\lambda_{\max}(\boldsymbol{\Sigma}) \to \infty$.

4. Simulation Studies

In this section, we employ Monte Carlo simulations to illustrate the finite sample performance of the MCA test and compare its power enhancement performance to that of the HDA test proposed in Ma et al. (2020). We consider two different simulation settings. For the first setting, the error term is cross-sectionally dependent, whereas for the second setting, we allow higher-order spatial auto-correlation error processes.

4.1 Two Simulation Settings

Example 1: We consider a modified version of the example studied in Section 4.1 of Ma et al. (2020). Response R_{it} are generated from the conditional three-factor model proposed by Fama and French (1993):

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

where f_{1t}, f_{2t} and f_{3t} represent the three factors, i.e., the MKT_t , SMB_t (small minus big), and HML_t (high minus low) factors, respectively. We assume that all these factors follow AR(1)-GARCH(1,1) processes, and the unknown coefficients for each factor are the same as those reported in Ma et al. (2020). The data generating process is as follows:

$$\mathrm{MKT}_t - 0.34 = 0.05(\mathrm{MKT}_{t-1} - 0.34) + g_{1t}^{1/2} \varsigma_{1t},$$

$$\mathrm{SMB}_t - 0.04 = 0.07(\mathrm{SMB}_{t-1} - 0.04) + g_{2t}^{1/2} \varsigma_{2t},$$

$$\mathrm{HML}_t - 0.06 = 0.04(\mathrm{HML}_{t-1} - 0.06) + g_{3t}^{1/2} \varsigma_{3t},$$

where ς_{jt} is independent and identically generated from a standard normal distribution for $j = 1, \cdots, 3$, and

$$g_{1t} = 0.32 + 0.67 g_{1t-1} + 0.13 g_{1t-1} \varsigma_{1t-1}^2,$$

$$g_{2t} = 0.33 + 0.51 g_{2t-1} + 0.03 g_{2t-1} \varsigma_{2t-1}^2,$$

$$g_{3t} = 0.26 + 0.72 g_{3t-1} + 0.05 g_{3t-1} \varsigma_{3t-1}^2.$$

The above processes are generated over the periods $t = -24, -23, \cdots, 0, 1, \cdots, T$ with the initial values $\text{MKT}_{-25} = 0$, $\text{SMB}_{-25} = 0$, $\text{HML}_{-25} = 0$, and $g_{j(-25)} = 1$, for $j = 1, \cdots, 3$. To offset the start-up effects, we discard the first 25 simulated observations and use $t = 1, \cdots, T$ in this studies. We consider that the conditional alphas and betas are driven by the unobservable state variable o_t. Let o_t follow the AR(1)-ARCH(1,1) process, that is, $o_t = 0.8o_{t-1} + u_t$, where $u_t = \delta_t \varrho_t$, $\varrho_t \sim N(0, 1)$, and $\delta_t^2 = 0.1 + 0.6u_{t-1}^2$ with $\delta_0^2 = 1$. The above processes are generated over the period $t = -24, -23, \cdots, 0, 1, \cdots, T$ with $o_{-25} = 0$ and $u_{-25} = 1$. Observations $t = 1, \cdots, T$ are used in the simulations. By definition, $\alpha_{it} = \alpha_i(t/T) = \alpha_{i,ACA} + \widetilde{\alpha}_i(t/T)$ in (2.5), and we set the conditional alphas $\alpha_{it} = c_i(1 + o_t)$, where $\alpha_{i,ACA} = c_i$ and $\widetilde{\alpha}_i(t/T) = c_i o_t$. Furthermore, the conditional betas are $\beta_{ijt} = \widetilde{a}_j + \widetilde{b}_j o_t$ for $i = 1, \cdots, N$, $t = 1, \cdots, T$, and $j = 1, \cdots, 3$, and we set $(\widetilde{a}_1, \widetilde{b}_1) = (0.5, 0.5)$, $(\widetilde{a}_2, \widetilde{b}_2) = (0.1, 0.5)$, and $(\widetilde{a}_3, \widetilde{b}_3) = (0.2, 0.5)$. Thus, under the null hypothesis, $c_i = 0$ for all $i = 1, \cdots, N$, which means that the conditional three-factor model holds.

Finally, we generate the idiosyncratic errors $\boldsymbol{\varepsilon}_{(t)} = (\varepsilon_{1t}, \cdots, \varepsilon_{Nt})^\top \in \mathbb{R}^N$, according to $\boldsymbol{\varepsilon}_{(t)} = \boldsymbol{\Sigma}^{1/2}\widetilde{\boldsymbol{\varepsilon}}_{(t)}$, $\widetilde{\boldsymbol{\varepsilon}}_{(t)} = (\widetilde{e}_{1t}, \cdots, \widetilde{e}_{Nt})^\top \in \mathbb{R}^N$, $\boldsymbol{\Sigma} = (\sigma_{ij}) \in \mathbb{R}^{N\times N}$, with $\sigma_{ij} = 0.5^{|i-j|}$ for $1 \le i, j \le N$, which means that the error term is approximately irrelevant when $|i - j|$ is sufficiently large. We consider three settings for $\widetilde{e}_{it}$: (1) Standardized normal distribution: $\widetilde{e}_{it} \sim N(0,1)$; (2) Standardized exponential distribution: $\widetilde{e}_{it} \sim \exp(1)$; (3) Mixture distribution: $\widetilde{e}_{it} \sim 0.9N(0,1) + 0.1N(0,9)$.

Example 2: The second experiment contains higher-order spatial autocorrelation error processes. The settings of this experiment are the same as those in Example 1, except $\mathbf{\Sigma}$. Following Fan et al. (2015), we start with $\mathbf{\Sigma}_1 = \text{diag}\left\{\Sigma_{1,1}, \cdots, \Sigma_{1,N/4}\right\}$ being a block-diagonal correlation matrix with 4×4 blocks located along the main diagonal, and each diagonal block $\Sigma_{1,j}$ for $j = 1, \cdots, N/4$ is assumed to be $\mathbf{I}_4$ initially. We then randomly choose $[N^{0.3}]$ blocks among them and make them nondiagonal by setting the (l,k)-th element of $\Sigma_{1,j}$ as $0.2^{|l-k|}$ for $1 \leq l, k \leq 4$. Here, $[\cdot]$ denotes the integer part of a real number. To allow for error cross-sectional heteroskedasticity, we set $\mathbf{\Sigma} = \mathbf{\Sigma}_0^{1/2}\mathbf{\Sigma}_1\mathbf{\Sigma}_0^{1/2}$, where $\mathbf{\Sigma}_0 = \text{diag}\{\sigma_1^2, \cdots, \sigma_N^2\}$ and $\sigma_i \sim U(1, 1.5)$ for $i = 1, \ldots, N$.

4.2 Simulation Results

To evaluate the empirical size and power of the proposed MCA test in the two examples mentioned above, we conduct simulations with varying numbers of observations (T = 120, 240, 360) and test assets (N = 50, 100, 200, 500). Further, the error terms are generated from the three aforementioned distributions (i.e., normal, exponential, and mixture distributions). For each setting, we use the BIC detailed in Section 2.2 to select the interior knots n, and the order of B-splines q is set to three for all the estimation windows. All simulations compare the actual rejection rates over 1,000 realizations with a nominal level of 5%. For comparison purposes, we also include the results of the HDA test. We set $c_i = 0$ for all $i = 1, \cdots, N$ to evaluate the empirical size of the MCA and HDA tests. To assess the empirical powers of the MCA and HDA tests, we set $c_i = c = \sqrt{\frac{2\log(N)}{Tp}}$ for $i \in \mathcal{S} \subseteq \{1, \cdots, N\}$ with $|\mathcal{S}| = [N^p]$, where each element in $\mathcal{S}$ is uniformly and randomly drawn from $\{1, \cdots, N\}$, and we consider three specific alternatives. The first case involves a dense alternative with $p = 1$. In this case, all $\alpha_{i,ACA}$s are non-zero, but their magnitudes are significantly small. The second case is the medium dense alternative, where we set two different signal strengths as $p = 0.8$ and 0.6. In this case, there are still some non-vanishing $\alpha_{I,ACA}$s with relatively small magnitudes. For example, when $N = 500$, the proportions of non-zero alphas are 8.3% and 28.9%, respectively. The third case is the sparse alternative, where we set two different signal strengths as $p = 0.4$ and 0.2. In this case, there are only a few non-zero $\alpha_{i,ACA}$s with relatively large magnitudes. For example, when $N = 500$, the proportions of non-zero alphas are 2.4% and 0.7%, respectively. In all

Table 1: Size and power of MCA and HDA tests from Examples 1 and 2 with normal distribution errors.

		size		power(dense)		power(medium dense)				power(sparse)			
				p=1		p=0.8		p=0.6		p=0.4		p=0.2	
T	N	MCA	HDA	MCA	HDA	MCA	HDA	MCA	HDA	MCA	HDA	MCA	HDA
						Example 1							
120	50	0.059	0.064	0.793	0.813	0.802	0.791	0.823	0.688	0.844	0.561	0.917	0.472
	100	0.052	0.059	0.869	0.877	0.856	0.823	0.89	0.685	0.866	0.416	0.947	0.436
	200	0.058	0.056	0.913	0.909	0.893	0.815	0.88	0.637	0.894	0.35	0.941	0.226
	500	0.061	0.06	0.946	0.942	0.927	0.849	0.923	0.582	0.924	0.192	0.943	0.096
240	50	0.047	0.044	0.765	0.84	0.729	0.708	0.779	0.513	0.849	0.493	0.934	0.363
	100	0.052	0.049	0.85	0.913	0.844	0.818	0.85	0.652	0.881	0.339	0.968	0.358
	200	0.053	0.057	0.886	0.924	0.892	0.864	0.893	0.577	0.91	0.23	0.963	0.147
	500	0.047	0.047	0.926	0.951	0.95	0.89	0.927	0.482	0.933	0.128	0.972	0.071
360	50	0.048	0.046	0.763	0.858	0.752	0.767	0.776	0.566	0.84	0.441	0.946	0.289
	100	0.05	0.045	0.84	0.895	0.821	0.795	0.848	0.599	0.891	0.285	0.975	0.276
	200	0.047	0.043	0.867	0.926	0.891	0.84	0.909	0.508	0.928	0.19	0.977	0.121
	500	0.051	0.05	0.924	0.967	0.937	0.889	0.941	0.446	0.954	0.137	0.993	0.066
						Example 2							
120	50	0.058	0.063	0.767	0.786	0.759	0.744	0.712	0.61	0.764	0.513	0.852	0.424
	100	0.059	0.064	0.808	0.833	0.844	0.829	0.836	0.688	0.834	0.44	0.896	0.421
	200	0.061	0.062	0.897	0.9	0.893	0.848	0.871	0.621	0.847	0.308	0.912	0.238
	500	0.062	0.06	0.938	0.944	0.922	0.863	0.915	0.575	0.913	0.222	0.902	0.104
240	50	0.056	0.059	0.713	0.809	0.762	0.819	0.709	0.58	0.773	0.486	0.877	0.365
	100	0.061	0.05	0.818	0.906	0.796	0.811	0.798	0.602	0.833	0.352	0.929	0.314
	200	0.059	0.041	0.863	0.912	0.86	0.821	0.846	0.527	0.874	0.226	0.921	0.153
	500	0.053	0.054	0.921	0.951	0.913	0.862	0.906	0.504	0.909	0.139	0.944	0.077
360	50	0.049	0.047	0.692	0.794	0.713	0.783	0.709	0.587	0.767	0.433	0.89	0.341
	100	0.052	0.042	0.795	0.892	0.795	0.837	0.796	0.558	0.836	0.284	0.944	0.29
	200	0.054	0.047	0.85	0.926	0.87	0.867	0.86	0.545	0.884	0.182	0.963	0.121
	500	0.055	0.048	0.93	0.948	0.92	0.881	0.913	0.418	0.915	0.117	0.959	0.06

of these cases, we set the remaining $c_i = 0$ for $i \notin \mathcal{S}$. Tables 1–3 present the sizes and powers of the MCA and HDA tests with normal, exponential, and mixture distribution errors, respectively.

Table 1 provides the MCA and HDA test results for Examples 1 and 2 with normal distribution errors, revealing the following findings. First, the sizes of the MCA and HDA tests are well, regardless of T and N for the two examples. Second, regarding the power performance, under the dense cases, the HDA tests slightly outperforms the MCA test. This result is expected as HDA is mainly designed for dense alternatives. However, in this setting, we still find that the performance of MCA is satisfactory and comparable with that of HDA. This is not the case for sparse alternatives. Specifically, under the medium dense cases, MCA is slightly superior to HDA. Under the sparse cases, MCA is greatly superior to HDA, and the HDA test performs poorly, especially when N and T are large. For example, when $N = 500$, $T = 360$, and $p = 0.2$, the power of the MCA test is 0.993 in Example 1, whereas that of the HDA test is only 0.066; the proposed MCA test power is approximately 15 times that of the HDA test. Moreover, when the error term follows exponential and mixture distributions, the simulation results reported in Tables 2 and 3, respectively, are quantitatively similar to the results in Table 1 for a normal distribution. These results imply that the

Table 2: Size and power of MCA and HDA tests from Examples 1 and 2 with exponential distribution errors.

		size		power(dense)		power(medium dense)				power(sparse)			
				p=1		p=0.8		p=0.6		p=0.4		p=0.2	
T	N	MCA	HDA	MCA	HDA	MCA	HDA	MCA	HDA	MCA	HDA	MCA	HDA
Example 1													
120	50	0.055	0.051	0.795	0.812	0.829	0.791	0.824	0.622	0.858	0.548	0.925	0.418
	100	0.053	0.059	0.859	0.859	0.878	0.833	0.853	0.671	0.875	0.435	0.944	0.46
	200	0.062	0.056	0.896	0.903	0.922	0.862	0.913	0.687	0.92	0.342	0.936	0.208
	500	0.061	0.064	0.927	0.936	0.931	0.86	0.916	0.567	0.924	0.225	0.949	0.102
240	50	0.053	0.05	0.749	0.829	0.793	0.796	0.8	0.632	0.872	0.458	0.934	0.351
	100	0.048	0.045	0.844	0.887	0.836	0.816	0.876	0.638	0.892	0.316	0.967	0.349
	200	0.051	0.046	0.869	0.917	0.905	0.866	0.889	0.507	0.924	0.228	0.968	0.136
	500	0.049	0.047	0.928	0.957	0.925	0.883	0.94	0.483	0.945	0.154	0.976	0.078
360	50	0.051	0.051	0.737	0.819	0.753	0.709	0.833	0.616	0.845	0.41	0.948	0.308
	100	0.045	0.049	0.832	0.9	0.809	0.766	0.861	0.609	0.906	0.286	0.972	0.253
	200	0.048	0.054	0.859	0.913	0.89	0.845	0.91	0.544	0.932	0.177	0.975	0.116
	500	0.042	0.045	0.937	0.968	0.927	0.882	0.93	0.438	0.958	0.116	0.987	0.079
Example 2													
120	50	0.06	0.054	0.797	0.826	0.778	0.752	0.755	0.586	0.838	0.55	0.868	0.405
	100	0.048	0.061	0.852	0.849	0.857	0.793	0.842	0.641	0.82	0.296	0.914	0.393
	200	0.053	0.051	0.901	0.911	0.898	0.832	0.897	0.604	0.873	0.325	0.918	0.209
	500	0.058	0.056	0.93	0.921	0.926	0.855	0.913	0.505	0.902	0.208	0.937	0.093
240	50	0.052	0.045	0.72	0.79	0.747	0.779	0.728	0.506	0.817	0.476	0.885	0.339
	100	0.046	0.05	0.812	0.873	0.832	0.817	0.841	0.621	0.852	0.295	0.931	0.325
	200	0.052	0.047	0.89	0.935	0.866	0.837	0.864	0.566	0.869	0.207	0.947	0.142
	500	0.05	0.053	0.915	0.946	0.927	0.867	0.916	0.47	0.92	0.13	0.954	0.07
360	50	0.051	0.044	0.719	0.825	0.742	0.765	0.743	0.614	0.782	0.43	0.896	0.291
	100	0.046	0.058	0.777	0.857	0.784	0.766	0.82	0.584	0.837	0.245	0.954	0.273
	200	0.053	0.047	0.864	0.915	0.849	0.845	0.86	0.504	0.899	0.161	0.957	0.109
	500	0.046	0.04	0.906	0.954	0.925	0.878	0.925	0.41	0.925	0.104	0.962	0.069

Table 3: Size and power of MCA and HDA tests from Examples 1 and 2 with mixture distribution errors.

		size		power(dense)		power(medium dense)				power(sparse)			
				p=1		p=0.8		p=0.6		p=0.4		p=0.2	
T	N	MCA	HDA	MCA	HDA	MCA	HDA	MCA	HDA	MCA	HDA	MCA	HDA
Example 1													
120	50	0.055	0.055	0.805	0.838	0.796	0.763	0.803	0.684	0.852	0.547	0.907	0.477
	100	0.051	0.051	0.856	0.874	0.863	0.828	0.88	0.685	0.894	0.446	0.942	0.464
	200	0.056	0.053	0.904	0.916	0.89	0.829	0.898	0.653	0.909	0.363	0.934	0.207
	500	0.055	0.065	0.934	0.943	0.938	0.862	0.936	0.584	0.926	0.212	0.946	0.088
240	50	0.054	0.05	0.768	0.846	0.82	0.834	0.785	0.568	0.863	0.478	0.942	0.38
	100	0.05	0.047	0.827	0.887	0.838	0.788	0.873	0.673	0.893	0.348	0.959	0.355
	200	0.049	0.056	0.89	0.937	0.899	0.868	0.912	0.576	0.918	0.26	0.974	0.161
	500	0.051	0.045	0.921	0.94	0.922	0.847	0.927	0.485	0.948	0.156	0.978	0.083
360	50	0.046	0.049	0.758	0.866	0.745	0.692	0.798	0.55	0.869	0.475	0.937	0.316
	100	0.049	0.05	0.815	0.865	0.862	0.835	0.874	0.65	0.907	0.304	0.971	0.296
	200	0.051	0.047	0.881	0.942	0.891	0.869	0.914	0.56	0.945	0.184	0.991	0.145
	500	0.05	0.051	0.933	0.959	0.918	0.866	0.942	0.427	0.961	0.132	0.984	0.072
Example 2													
120	50	0.045	0.053	0.787	0.819	0.773	0.77	0.765	0.643	0.796	0.557	0.86	0.469
	100	0.05	0.048	0.866	0.889	0.821	0.811	0.836	0.676	0.829	0.41	0.914	0.435
	200	0.056	0.062	0.901	0.917	0.903	0.852	0.857	0.654	0.88	0.38	0.915	0.229
	500	0.062	0.059	0.944	0.933	0.93	0.852	0.902	0.559	0.917	0.207	0.915	0.092
240	50	0.049	0.045	0.731	0.824	0.749	0.785	0.744	0.643	0.79	0.497	0.888	0.366
	100	0.055	0.045	0.837	0.904	0.841	0.839	0.81	0.588	0.852	0.352	0.953	0.321
	200	0.048	0.049	0.904	0.941	0.878	0.838	0.864	0.565	0.88	0.237	0.943	0.14
	500	0.052	0.053	0.927	0.95	0.927	0.867	0.905	0.45	0.921	0.141	0.946	0.075
360	50	0.052	0.055	0.692	0.791	0.703	0.744	0.712	0.506	0.789	0.44	0.905	0.363
	100	0.041	0.046	0.805	0.883	0.811	0.82	0.801	0.557	0.838	0.311	0.952	0.317
	200	0.052	0.051	0.874	0.93	0.875	0.832	0.869	0.529	0.887	0.179	0.959	0.118
	500	0.052	0.044	0.923	0.96	0.92	0.876	0.918	0.399	0.945	0.108	0.966	0.066

sizes of both the MCA and HDA tests are satisfactory; however, the MCA test is superior to the HDA test for sparse alternatives and comparable with the HDA test for dense alternatives. In summary, MCA is superior to HDA for the two examples presented here, and the results are robust to different error specifications.

Additionally, we conducted experiments to illustrate the finite sample performance of the proposed test under conditional multi-factor models with latent factors. The experimental results are summarized in Appendix S8, which show that the empirical size and power performances of the proposed adjusted MCA test are stable.

5. Conclusion

This paper addressed the problem of testing for the presence of alpha in conditional time-varying multi-factor models. We have introduced the MCA test as a novel approach, specifically designed for situations where the number of test assets N is significantly larger than the number of observations T and the alternative hypothesis is sparse. The theoretical analysis demonstrated that the convergence of the proposed test to the Type-I extreme value distribution as $\min(T, N) \to \infty$ under appropriate conditions. Furthermore, we extended the MCA test to incorporate unobservable latent factors in conditional pricing models. The simulation results and real data analysis confirmed the satisfactory finite sample performance of the proposed test.

The results and findings of this study suggest two possible avenues for future research. First, extending the proposed testing procedure to encompass nonparametric pricing models would enhance its applicability and provide a more comprehensive analysis of asset pricing. Second, addressing the issue of missing observations in the proposed method would improve its robustness and practical utility. This avenue could build upon existing work in the literature, such as the studies by Giglio et al. (2021) and Jin et al. (2021), which have discussed the implications of missing data in factor models. We believe that these efforts would considerably broaden the applicability of MCA.

Supplementary Material

The Supplementary Material consists of ten parts (Sections S1–S10). Section S1 provides eight useful lemmas, Section S2 provides the proof of Theorem 1, Section S3 provides the proof of Theorem 2, Section S4 provides the proof of Proposition 1, Section S5 provides the proof of Theorem 3, Section S6 describes the test portfolios, Section S7 presents the empirical evidence of the time-varying coefficients and sparse alternatives based on real data, Section S8 reports the simulation results of conditional multi-factor models with latent factors, Section S9 reports the simulation results of the MAX test proposed by Feng et al. (2022), and Section S10 provides the simulation results for a student-t distribution error.

Acknowledgements

Jun Zhang's research was supported by the Fundamental Research Funds for the Central Universities (JBK2307105). Wei Lans research was supported by the National Natural Science Foundation of China (NSFC, 12171395, 11931014, 71991472, 72333001), the National Key Research and Development Program of China (2022YFA1003702), the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics, and the Fundamental Research Funds for the Central Universities (JBK1806002). Xinyan Fans research was supported by the National Natural Science Foundation of China (NSFC, 12201626), the MOE Project of Key Research Institute of Humanities and Social Sciences (22JJD110001), and the Public Computing Cloud, Renmin University of China. Wen Chens research was supported by the National Natural Science Foundation of China (NSFC, 72003157). The authors are grateful to the Editor, the AE, and the anonymous referees for their insightful comments and constructive suggestions.

References

Ahmed, S., Bu, Z., and Tsvetanov, D. (2019). Best of the best: A comparison of factor models. *Journal of Financial and Quantitative Analysis*, 54(4):1713–1758.

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

Ang, A. and Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, 106(1):132–156.

Bai, Z. and Silverstein, J. W. (2005). *Spectral Analysis of Large Dimensional Random Matrices, 2nd Ed.* New York: Springer.

Cai, T. T., Liu, W., and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2): 349–372.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Cooper, I. and Maio, P. (2019). New evidence on conditional factor models. *Journal of Financial and Quantitative Analysis*, 54(5):1975–2016.

De Boor, C. (1978). *A practical guide to splines.* New York: springer-verlag.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics*, 39(6):3320.

Fan, J., Liao, Y., and Yao, J. (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, 83(4):1497–1541.

Feng, L., Lan, W., Liu, B., and Ma, Y. (2022). High-dimensional test for alpha in linear factor pricing models with sparse alternatives. *Journal of Econometrics*, 229(1):152–175.

Ferson, W. E. and Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. *The Journal of Finance*, 54(4):1325–1360.

Ghysels, E. (1998). On stable factor structures in the pricing of risk: do time-varying betas help or hurt? *The Journal of Finance*, 53(2):549–573.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57(5):1121–1152.

Giglio, S., Liao, Y., and Xiu, D. (2021). Thousands of alpha tests. *The Review of Financial Studies*, 34(7):3456–3496.

Guo, H., Wu, C., and Yu, Y. (2017). Time-varying beta and the value premium. *Journal of Financial and Quantitative Analysis*, 52(4):1551–1576.

He, X. and Shi, P. (1996). Bivariate tensor-product b-splines in a partly linear model. *Journal* *of Multivariate Analysis*, 58(2):162–181.

Jin, S., Miao, K., and Su, L. (2021). On factor models with random missing: Em estimation, inference, and cross validation. *Journal of Econometrics*, 222(1):745–777.

Lan, W., Ding, Y., Fang, Z., and Fang, K. (2016). Testing covariates in high dimension linear regression with latent factors. *Journal of Multivariate Analysis*, 144:25–37.

Lan, W., Feng, L., and Luo, R. (2018). Testing high-dimensional linear asset pricing models. *Journal of Financial Econometrics*, 16(2):191–210.

Lan, W., Wang, H., and Tsai, C.-L. (2014). Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics*, 66(2):279–301.

Lewellen, J. and Nagel, S. (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82(2):289–314.

Li, Y. and Yang, L. (2011). Testing conditional factor models: A nonparametric approach. *Journal of Empirical Finance*, 18(5):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(1):13–37.

Ma, S., Lan, W., Su, L., and Tsai, C.-L. (2020). Testing alphas in conditional time-varying factor models with high-dimensional assets. *Journal of Business & Economic Statistics*, 38(1):214–227.

Ma, S. and Song, P. X.-K. (2015). Varying index coefficient models. *Journal of the American*

Statistical Association, 110(509):341–356.

MacKinlay, A. C. and Richardson, M. P. (1991). Using generalized method of moments to test mean-variance efficiency. *The Journal of Finance*, 46(2):511–527.

Pesaran, M. H. and Yamagata, T. (2017). Testing for alpha in linear factor pricing models with a large number of securities. *Social Science Electronic Publishing*.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259.

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. *The Journal of Finance*, 19(3):425–442.

Wang, H. (2012). Factor profiled sure independence screening. *Biometrika*, 99(1):15–28.

Wang, L., Li, H., and Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association*, 103(484):1556–1569.

Zhou, G. (1993). Asset-pricing tests under alternative distributions. *The Journal of Finance*, 48(5):1927–1942.

Jun Zhang

School of Statistics and Center of Statistical Research, Southwestern University of Finance and Economics, Sichuan 611130, China

E-mail: junz@smail.swufe.edu.cn

Wei Lan

School of Statistics and Center of Statistical Research, Southwestern University of Finance and Economics, Sichuan 611130, China

E-mail: lanwei@swufe.edu

Xinyan Fan

School of Statistics and Center for Applied Statistics, Renmin University of China, Beijing 100086, China.

E-mail: 1031820039@qq.com

Wen Chen

College of Finance, Nanjing Agricultural University, Nanjing 210095, China.

E-mail: chenwenpku2008@163.com