Article:

Pesaran, M. Hashem and Yamagata, Takashi orcid.org/0000-0001-5949-8833 (2023) Testing for Alpha in Linear Factor Pricing Models with a Large Number of Securities. Journal of Financial Econometrics. ISSN: 1479-8409

https://doi.org/10.1093/jjfinec/nbad002

Journal of Financial Econometrics, 2023, 1–54
https://doi.org/10.1093/jjfinec/nbad002
Article

Testing for Alpha in Linear Factor Pricing Models with a Large Number of Securities*

M. Hashem Pesaran [1,2] and Takashi Yamagata[3,4]

[1]Department of Economics, University of Southern California, Los Angeles, USA, [2]Trinity College, Cambridge, UK, [3]Department of Economics and Related Studies, University of York, UK and [4]Institute of Social and Economic Research, Osaka University, Japan

Address correspondence to Takashi Yamagata, DERS, University of York, York YO10 5DD, UK or e-mail: takashi.yamagata@york.ac.uk.

Received September 23, 2021; revised January 5, 2023; editorial decision January 13, 2023; accepted January 16, 2023

Abstract

This article considers tests of alpha in linear factor pricing models when the number of securities, N, is much larger than the time dimension, T, of the individual return series. We focus on class of tests that are based on Student's t-tests of individual securities which have a number of advantages over the existing standardized Wald type tests, and propose a test procedure that allows for non-Gaussianity and general forms of weakly cross-correlated errors. It does not require estimation of an invertible error covariance matrix, it is much faster to implement, and is valid even if N is much larger than T. We also show that the proposed test can account for some limited degree of pricing errors allowed under Ross's arbitrage pricing theory condition. Monte Carlo evidence shows that the proposed test performs remarkably well even when $T=60$ and $N=5000$. The test is applied to monthly returns on securities in the S&P 500 at the end of each month in real time, using rolling windows of size 60. Statistically significant evidence against Sharpe–Lintner capital asset pricing model and Fama–French three and five factor models are found mainly during the period of Great Recession (2007M12–2009M06).

Key words: arbitrage asset pricing, CAPM, S&P 500 securities, testing for alpha, weak and spatial error cross-sectional dependence

JEL classification: C12, C15, C23, G11, G12

* This is a revised and updated version of the article entitled "Testing CAPM with a Large Number of Assets," initially released in April 2012 as IZA Discussion Papers No. 6469. We would like to thank two anonymous referees and the Editor, Dacheng Xiu, for valuable comments. We are grateful to Ron Smith, Natalia Bailey, and Jay Shanken and other participants at the American Finance Association Meeting in San Diego, in January 2013 for helpful comments. The first author wishes to acknowledge partial support from the ESRC Grant No. ES/I031626/1. The second author acknowledges partial support from the JSPS KAKENHI (grant numbers 20H01484, 20H05631, 21H00700, and 21H04397).

This article is concerned with testing for the presence of alpha in linear factor pricing models (LFPMs) such as the capital asset pricing model (CAPM) due to Sharpe (1964) and Lintner (1965), or the arbitrage pricing theory (APT) model due to Ross (1976), when factors are observed and the number of securities, N, is quite large relative to the time dimension, T, of the return series under consideration. There exists a large literature in empirical finance that tests various implications of Sharpe–Lintner model. Cross-sectional as well as time-series tests have been proposed and applied in many different contexts. Using time-series regressions, Jensen (1968) was the first to propose using standard t-statistics to test the null hypothesis that the intercept, α_i, in the ordinary least squares (OLS) regression of the excess return of a given security, i, on the excess return of the market portfolio is zero.[1]

However, when a large number of securities are under consideration, due to dependence of the errors across securities in the LFPM regressions, the individual t-statistics are correlated which makes controlling the overall size of the test problematic. Gibbons, Ross, and Shanken (1989, GRS) propose an exact multivariate version of the test which deals with this problem if the CAPM regression errors are Gaussian and $N < T$. This is the standard test used in the literature, but its application has been confined to testing the market efficiency of a relatively small number of portfolios, typically $20-30$, using monthly returns observed over relatively long time periods. The use of large T as a way of ensuring that $N < T$ is also likely to increase the possibility of structural breaks in the β's that could in turn adversely affect the performance of the GRS test.

Recently, there has been a growing body of finance literature which uses individual security returns rather than portfolio returns for the test of pricing errors. Ang, Liu, and Schwarz (2020) show that the smaller variation of beta estimates from creating portfolios may not lead to smaller variation of cross-section regression estimates. Cremers, Halling, and Weinbaum (2015) examine the pricing of both aggregate jump and volatility risk based on individual stocks rather than portfolios. Chordia, Goyal, and Shanken (2017) advocate the use of individual securities to investigate whether the source of expected return variation is from betas or security-specific characteristics.

Out of the two main assumptions that underlie the GRS test, the literature has focused on the implications of non-normal errors for the GRS test, and ways of allowing for non-normal errors when testing $\alpha_i = 0$. Affleck-Graves and Mcdonald (1989) were among the first to consider the robustness of the GRS test to non-normal errors who, using simulation techniques, find that the size and power of GRS test can be adversely affected if the departure from non-normality of the errors is serious, but conclude that the GRS test is "reasonably robust with respect to typical levels of nonnormality." (p. 889). More recently, Beaulieu, Dufour, and Khalaf (2007, BDK) and Gungor and Luger (2009) have proposed tests of $\alpha_i = 0$ that allow for non-normal errors, but retain the restriction $N < T$. BDK develop an exact test which is applicable to a wide class of non-Gaussian error distributions, and use Monte Carlo simulations to achieve the correct size for their test. GL propose two distribution-free nonparametric sign tests in the case of single factor models that allow the error distribution to be non-normal but require it to be cross-sectionally independent and conditionally symmetrically distributed around zero.

1 Cross-sectional tests of CAPM have been considered by Douglas (1967); Black, Jensen, and Scholes (1972); and Fama and MacBeth (1973), among others. An early review of the literature can be found in Jensen (1972), and more recently in Fama and French (2004).

Our primary focus in this article is on multivariate tests of $H_0 : \alpha_i = 0$, for $i = 1, 2, \ldots, N$, when $N > T$, while allowing for non-Gaussian and weakly cross-sectionally correlated errors. The latter condition is required for consistent estimation of the error covariance matrix, $\mathbf{V}$, when N is large relative to T. In the case of LFPM regressions with weakly cross-sectionally correlated errors, consistent estimation of $\mathbf{V}$ can be achieved by adaptive thresholding which sets to zero elements of the estimator of $\mathbf{V}$ that are below a given threshold. Alternatively, feasible estimators of $\mathbf{V}$ can be obtained by Bayesian or classical shrinkage procedures that scale down the off-diagonal elements of $\mathbf{V}$ relative to its diagonal elements.[2] Fan, Liao, and Mincheva (2011, 2013) consider consistent estimation of $\mathbf{V}$ in the context of an approximate factor model. They assume $\mathbf{V}$ is sparse and propose an adaptive thresholding estimator of $\mathbf{V}$, which they show to be positive definite with satisfactory small sample properties. Fan, Liao, and Yao (2015) consider a standardized Wald (SW) test based on the estimator of $\mathbf{V}$ proposed by Fan, Liao, and Mincheva (2013) and derive the conditions under which the SW test of H_0 can be asymptotically justified. Gungor and Luger (2016, GL) propose a simulation-based approach for testing pricing errors. They claim that their test procedure is robust against non-normality and cross-sectional dependence in the errors. Gagliardini, Ossola, and Scaillet (2016, GOS) develop two-pass regressions of individual stock returns, allowing time-varying risk premia, and propose a SW test. Lan, Feng, and Luo (2018) use random projection of the N security returns onto a smaller number of portfolios to circumvent the high-dimensional problem when testing for alphas, but require N and T to be of the same order of magnitude. Raponi, Robotti, and Zaffaroni (2019) propose a test of pricing error in cross-section regression for fixed number of time-series observations. They use a bias-corrected estimator of Shanken (1992) to standardize their test statistic. Ma et al. (2020) employ polynomial spline techniques to allow for time variations in factor loadings when testing for alphas. Feng et al. (2022) propose a max-of-square type test of alphas instead of the average used in the literature, and recommend using a combination of the two testing procedures. As noted by He et al. (2021), Bai and Saranadasa (1996, BS) consider yet another SW type test which requires N and T to be of the same order of magnitude.

In this article, we develop a test statistic that initially ignores the off-diagonal elements of $\mathbf{V}$ and base the test of H_0 on the average of the squared t-ratios for $\alpha_i = 0$, over $i = 1, 2, \ldots, N$. This idea was originally proposed in the working paper version of this article, independently of a similar approach subsequently followed by GOS. Despite the similarity of the two tests, as will be seen, our version of the test performs much better for all combinations of N and T considered in the literature, and delivers excellent size and power even if N is very large (around 5000), in contrast to other tests that tend to over reject as N is increased relative to T. We are also able to establish the asymptotic distribution of proposed test under much weaker conditions and without resorting to high level

2 There exists a large literature in statistics and econometrics on estimation of high-dimensional covariance matrices which use regularization techniques such as shrinkage, adaptive thresholding, or other dimension-reducing procedures that impose certain structures on the variance matrix such as sparsity, or factor structures. See, for example, Wong, Carter, and Kohn (2003); Ledoit and Wolf (2004); HuAng et al. (2006); BL; Fan, Fan, and Lv (2008); Cai and Liu (2011); Fan, Liao, and Mincheva (2011, 2013); and BPS.

assumptions.[3] We achieve this by making corrections to the numerator of the test statistic to ensure that the test is more accurately centered, and correct the denominator of the test statistic to allow for the effects of non-zero off-diagonal elements of the underlying error covariance matrix. The correction involves consistently estimating $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$, where $\mathbf{R} = (\rho_{ij})$ is the error correlation matrix. The estimation of $N^{-1}\mathrm{Tr}(\mathbf{R}^2) = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{ij}^2$ is subject to the curse of dimensionality which we address by using the multiple testing (MT) threshold estimator, $\tilde{\mathbf{R}}$, recently proposed by Bailey, Pesaran, and Smith (2019, BPS). We show that consistent estimation of $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ can be achieved under a more general specification of $\mathbf{R}$ when compared with tests that require a consistent estimator of the full matrix, $\mathbf{R}$. We are able to establish that the resultant test is applicable more generally and continues to be valid for a wider class of error covariances, and holds even if N rises faster than T. The proposed test is also corrected for small sample effects of non-Gaussian errors, which is of particular importance in finance. We refer to this test as Jensen's α test of LFPM and denote it by $\hat{J}_\alpha$. The test can also be viewed as a robust version of a SW test, in cases where the off-diagonal elements of $\mathbf{V}$ become relatively less important as $N \to \infty$. Further, the implementation of the $\hat{J}_\alpha$ test is computationally less demanding, since it does not involve estimation of an invertible high-dimensional error covariance matrix.

We note that the $\hat{J}_\alpha$ test is not the first one which is based on the standardized squared t-ratio for $\alpha_i = 0$. As discussed in He et al. (2021), Srivastava and Du (2008, SD) propose standardized squared t-ratio, using a different standardization from ours. As will be seen below, their standardization results in serious size distortion when N is larger than T (see the SD test discussed in Section 5). Also, Hwang and Satchell (2014) proposed a simulation-based test, using average of the squared t-ratios.

Our assumption regarding the sparsity of $\mathbf{V}$ advances on Chamberlain's (1983) approximate factor model formulation of the asset model, where it is assumed that the largest eigenvalue of $\mathbf{V}$ (or $\mathbf{R}$) is uniformly bounded in N (Chamberlain, 1983, p. 1307). We relax this assumption and allow the maximum column sum matrix norm of $\mathbf{R}$ to rise with N but at a rate slower than $\sqrt{N}$, while controlling the overall sparsity of $\mathbf{R}$ by requiring $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ to be bounded in N. In this way, we are able to allow for two types of cross-sectional error dependence: one due to the presence of weak common factors that are not sufficiently strong to be detectable using standard estimation techniques, such as principal components and another due to the error dependence that arises from interactive and spillover effects.

We establish that under the null hypothesis $H_0: \alpha_i = 0$, the $\hat{J}_\alpha$ test is asymptotically distributed as $N(0, 1)$ for T and $N \to \infty$ jointly, so long as $N/T^2 \to 0$, $m_N = \|\mathbf{R}\|_1 = O(N^{\delta_\rho})$, $0 \le \delta_\rho < 1/2$, and $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ is bounded in N. The test is also shown to have power against alternatives that rise in $N^{1/2}T$. We consider the implications of allowing for pricing errors on the asymptotic properties of the $\hat{J}_\alpha$ test and show that testing H_0 still allows for some very limited degree of non-zero pricing errors. The proofs are quite involved and in

3 Monte Carlo experiments reported by Feng et al. (2022) also show significant over-rejection of the null by the GOS test when $T = 50$ and $N = 500$. These authors do not report simulation results for larger values of N as they increase T to 100 and 200. It is therefore unclear if the over-rejection continues when N is also increased beyond 500 when $T = 100$. As we also note in the article, increasing T to avoid over-rejection increases the likelihood of breaks in factor loadings which could be another source of over-rejection.

some parts rather tedious. For the purpose of clarity, we provide statements of the main theorems with the associated assumptions in the article, but relegate the mathematical details to the Appendix.

Small sample properties of the $\hat{J}_\alpha$ test are investigated using Monte Carlo experiments designed specifically to match the distributional features of the residuals of Fama–French three factor regressions of individual securities in the Standard & Poor 500 (S&P 500) index. We consider the comparative test results for the following nine sample size combinations, $T \in \{60, 120, 240\}$ and $N = \{50, 100, 200\}$. The $\hat{J}_\alpha$ test performs well for all sample size combinations with empirical size very close to the chosen nominal value of 5%, and satisfactory power. Comparing the size and power of the $\hat{J}_\alpha$ test with the GRS test in the case of experiments with $N < T$ for which the GRS statistics can be computed, we find that the $\hat{J}_\alpha$ test has higher power than the GRS test in most experiments. This could be due to the non-normal errors adversely affecting the GRS test, as reported by Affleck-Graves and Mcdonald (1989, 1990). In addition, the $\hat{J}_\alpha$ test outperforms the test proposed by GOS as well as the SW test of Fan, Liao, and Yao (2015) and the SD test of Srivastava and Du (2008). The $\hat{J}_\alpha$ test also outperforms the simulation-based $F_{\max}$ test of Gungor and Luger (2016) and the BS test of Bai and Saranadasa (1996), which are shown to be substantially undersized across the various designs, and has lower power when compared with the $\hat{J}_\alpha$ test. Further, we carried out additional experiments using much larger values of N, namely $N = 500, 1000$, 2000, and 5000, while keeping T at 60, 120, and 240. We only considered the $\hat{J}_\alpha$ test for these experiments and found no major evidence of size distortions even for the experiments with $T = 60$ and $N = 5000$.

Encouraged by the satisfactory performance of the $\hat{J}_\alpha$ test even in cases where N is much larger than T, we applied the test to monthly returns on the securities in the S&P 500 index using rolling windows of size $T = 60$ months. The survivorship bias problem is minimized by considering the sample of securities included in the S&P 500 at the end of each month in real time. We report the $\hat{J}_\alpha$ test results for CAPM, three and five Fama–French factor models over the period September 1989 to April 2018, and the three sub-periods: (1) the Asian financial crisis (1997M07–1998M12), (2) the Dot-com bubble burst (2000M03–2002M10), and (3) the Great Recession (2007M12–2009M06) periods. We find that the $\hat{J}_\alpha$ test rejects $H_0 : \alpha_i = 0$, mainly during periods of major financial disruptions, particularly the period of Great Recession, with the GOS test rejecting the null for most periods, largely due to its tendency to over-reject when T is short relative to N.

The outline of the rest of the article is as follows. Section 1 sets out the LFPM, formulates the null hypothesis that underlies the tests for alphas which allow for pricing errors and weak latent or missing factors. Section 2 introduces the estimates of alpha and derives the GRS test as a point of departure for dealing with the case where $N > T$. Section 3 proposes the $\hat{J}_\alpha$ test for large N panels and derives its asymptotic distribution, and Section 4 summarizes the main theoretical results. Section 5 reports on small sample properties of $\hat{J}_\alpha$, GRS, GOS, SW, $F_{\max}$, BS and SD tests, using Monte Carlo techniques. Section 6 presents the empirical application. Section 7 concludes. The proofs of the main theorems are provided in the Appendix, and the lemmas which are used for the proofs, as well as the additional Monte Carlo evidence and the detailed discussion on data sources, are provided in the Supplementary Material.

Notations: We use K and c to denote finite and small positive constants. If $\{f_t\}_{t=1}^{\infty}$ is any real sequence and $\{g_t\}_{t=1}^{\infty}$ is a sequences of positive real numbers, then $f_t = O(g_t)$, if there exists a positive finite constant K such that $|f_t|/g_t \leq K$ for all t. $f_t = o(g_t)$ if $f_t/g_t \to 0$ as $t \to \infty$. If $\{f_t\}_{t=1}^{\infty}$ and $\{g_t\}_{t=1}^{\infty}$ are both positive sequences of real numbers, then $f_t = \ominus(g_t)$ if there exists $T_0 \geq 1$ and positive finite constants C_0 and C_1, such that $\inf_{t \geq T_0}(f_t/g_t) \geq C_0$ and $\sup_{t \geq T_0}(f_t/g_t) \leq C_1$. For a $N \times N$ matrix $\mathbf{A} = (a_{ij})$, the minimum and maximum eigenvalues of matrix $\mathbf{A}$ are denoted by $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$, respectively, its trace by $\text{Tr}(\mathbf{A})$, its maximum absolute column and row sum matrix norms by $\|\mathbf{A}\|_\infty = \sup_i \sum_{j=1}^{N} |a_{ij}|$, and, $\|\mathbf{A}\|_1 = \sup_j \sum_{i=1}^{N} |a_{ij}|$, respectively, its Frobenius and spectral norms by $\|\mathbf{A}\|_F = \sqrt{Tr(\mathbf{A}'\mathbf{A})}$, and $\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, respectively. For a $N \times 1$ dimensional vector, $\boldsymbol{\alpha}$, $\|\boldsymbol{\alpha}\| = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{1/2}$.

1 The LFPM and APT Restrictions

We base our test of alpha on the following statistical factor model:

$$R_{it} - r_t^f = \alpha_i + \boldsymbol{\beta}_i'\mathbf{f}_t + u_{it}, \tag{1}$$

where R_{it} is the return on security i during period t, r_t^f is the risk-free rate, $\mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{mt})'$ is the $m \times 1$ vector of observed factors, $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{im})'$ is the associated vector of risk factors with mean $\boldsymbol{\mu} = E(\mathbf{f}_t)$. Under the APT due to Ross (1976), the following restrictions are imposed:

$$E\left(R_{it} - r_t^f\right) = \mu_{ir} = \lambda_0 + \boldsymbol{\beta}_i'\boldsymbol{\lambda} + \varpi_i, \tag{2}$$

where λ_0 is the zero-beta expected excess return, $\boldsymbol{\lambda}$ is the $m \times 1$ vector of risk premia, and ϖ_i is the pricing error of security i such that

$$\sum_{i=1}^{N} \left(\mu_{ir} - \lambda_0 - \boldsymbol{\beta}_i'\boldsymbol{\lambda}\right)^2 = \sum_{i=1}^{N} \varpi_i^2 < K. \tag{3}$$

This latter condition is given by Equation (18) in Theorem II of Ross and ensures that under APT pricing errors are sparse. In this article, we consider a more general bound on the pricing errors and assume that

$$\sum_{i=1}^{N} \varpi_i^2 = O(N^{\delta_\varpi}), \tag{4}$$

where the exponent δ_ϖ measures the degrees of pervasiveness of pricing errors. Deviations from APT are measured in terms of δ_ϖ $(0 \leq \delta_\varpi < 1)$. Large values of δ_ϖ represent major departures from APT.

To motivate the alpha tests of $\alpha_i = 0$ in the statistical model, we note that under Equation (1),

$$E\left(R_{it} - r_t^f\right) = \alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\mu},$$

and for the statistical model to be compatible with the APT condition (2), we must have

$$\alpha_i = \lambda_0 + \boldsymbol{\beta}_i'(\boldsymbol{\lambda} - \boldsymbol{\mu}) + \varpi_i. \tag{5}$$

Therefore, testing the null hypothesis, $H_0: \alpha_i = 0$ for all i, can be viewed as tests of the joint hypothesis $\lambda_0 = 0$, $\boldsymbol{\lambda} = \boldsymbol{\mu}$, (referred to as "no spanning errors" here after), and testing $\varpi_i = 0$, for all i (referred to as "no pricing errors"). Under APT, the excess return regressions can be written as

$$y_{it} = \lambda_0 + \boldsymbol{\beta}_i'(\boldsymbol{\lambda} - \boldsymbol{\mu}) + \varpi_i + \boldsymbol{\beta}_i'\mathbf{f}_t + u_{it}, \text{ for } i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T, \tag{6}$$

where $y_{it} = R_{it} - r_t^f$ and ϖ_i satisfies Equation (4).[4] Under APT, the above model is often referred to as the LFPM, to be distinguished from the statistical linear factor model given by Equation (1). It is also worth noting that when testing H_0 it is still possible to allow for a limited degree of non-zero pricing errors, depending on the prevalence of the pricing errors and the relative expansion rates of N and T. See Remark 8 below.[5]

To ensure that the risk from common factors, $\mathbf{f}_t$, cannot be fully diversified we assume that at least one of the observed factors is strong, in the sense that

$$\sup_s \sum_{i=1}^{N} |\beta_{is}| = \ominus(N). \tag{7}$$

Our test does not require all the observed factors to be strong, and allows these factors to have different degrees of strength. In a recent paper, Bailey, Kapetanios, and Pesaran (2021) find that among over 140 factors proposed in the literature only the market factor can be regarded as strong. The other factors are estimated to be semi-strong, such that the sum of their loadings in absolute terms rises with N but at the rate of δ_β, where $1/2 < \delta_\beta < 1$. Also, there is no guarantee that all relevant factors are included in the asset pricing model, and to allow for possible missing (or latent) factors, we assume that

$$u_{it} = \boldsymbol{\gamma}_i'\mathbf{v}_t + \eta_{it}, \tag{8}$$

where $\mathbf{v}_t$ is a $k \times 1$ vector of latent common factors that are $\text{IID}(0, \mathbf{I}_k)$, $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{ik})'$ is the associated vector of factor loadings with bounded elements, $\sup_{i,s} |\gamma_{is}| < K$. The latent factors included in the error process must be weak such that

$$\sup_s \sum_{i=1}^{N} |\gamma_{is}| = O(N^{\delta_\gamma}), \text{ with } 0 \leq \delta_\gamma < 1/2. \tag{9}$$

The idiosyncratic errors, η_{it}, are also allowed to be weakly cross-correlated. Specifically, we assume that $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, \ldots, \eta_{Nt})' = \mathbf{Q}_\eta \boldsymbol{\varepsilon}_{\eta,t}$, where $\boldsymbol{\varepsilon}_{\eta,t} = (\varepsilon_{\eta,1t}, \varepsilon_{\eta,2t}, \ldots, \varepsilon_{\eta,Nt})'$, $\{\varepsilon_{\eta,it}\}$ are IID processes over i and t, with zero means , unit variances, $\gamma_{2,\varepsilon_\eta} = E\left(\varepsilon_{\eta,it}^4\right) - 3$, and $\sup_{i,t} E(|\varepsilon_{\eta,it}|^{8+c}) \leq K < \infty$, for some $c > 0$. We denote the correlation matrix of $\boldsymbol{\eta}_t$ by $\mathbf{R}_\eta = (\rho_{\eta,ij})$ and note that $\mathbf{R}_\eta = \mathbf{Q}_\eta \mathbf{Q}_\eta'$. To ensure that $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})'$ is weakly cross-correlated, we require that k, the number of weak factors, is finite, and $\|\mathbf{R}_\eta\|_1 \leq \|\mathbf{Q}_\eta\|_1 \|\mathbf{Q}_\eta\|_\infty \leq K$. The error specification in Equation (8) is quite general and allows for weak latent common factors as well as network and spatial error cross

4 Some researchers have focused on testing the restrictions $\boldsymbol{\lambda} - \boldsymbol{\mu} = 0$, allowing λ_0 to be unrestricted. See, for example, Shanken (1992).

5 Note that the GRS test is also based on the same null hypothesis, $H_0: \alpha_i = 0$, and assumes zero pricing errors.

dependence. We note that common factors cannot substitute for network dependence and allowing for both types of dependence in the errors is important.

2 Preliminaries and the GRS Test

It proves useful to stack the panel regressions in Equation (6) by time series as well as by cross-section observations. Stacking by time-series observations we have

$$\mathbf{y}_{i.} = \alpha_i \boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i.}, \tag{10}$$

where $\mathbf{y}_{i.} = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\boldsymbol{\tau}_T = (1, 1, \ldots, 1)'$, $\mathbf{F}' = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_T)$, and $\mathbf{u}_{i.} = (u_{i1}, u_{i2}, \ldots, u_{iT})'$. Stacking by cross-sectional observations we have

$$\mathbf{y}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \mathbf{u}_t, \tag{11}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)'$, $\mathbf{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_N)'$, and $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})'$.

For derivation of the exact GRS (Gibbons et al., 1989) test, we assume that $\mathbf{u}_t \sim \text{IIDN}(\mathbf{0}, \mathbf{V})$, namely errors, u_{it}, are Gaussian, have zero means, and are serially uncorrelated such that $E(u_{it}u_{jt'}) = 0$, for all i, j, and $t \neq t'$, with $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{V}$, where $\mathbf{V} = (\sigma_{ij})$ is an $N \times N$ symmetric positive definite matrix. A non-Gaussian version of this assumption will be considered below. Starting with Jensen's (1968) test of individual α_i's, we note that the OLS estimator of α_i is given by

$$\hat{\alpha}_i = \mathbf{y}_{i.}' \left(\frac{\mathbf{M}_F \boldsymbol{\tau}_T}{\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T} \right), \tag{12}$$

where $\mathbf{M}_F = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$, and is an efficient estimator despite the fact that $\mathbf{V}$ is not a diagonal matrix. This result follows since Equation (10) is a seemingly unrelated regression equation specification with the *same* set of regressors across all the N securities. It is also easily seen that

$$\hat{\alpha}_i = \left(\alpha_i \boldsymbol{\tau}_T' + \boldsymbol{\beta}_i'\mathbf{F}' + \mathbf{u}_{i.}'\right) \left(\frac{\mathbf{M}_F \boldsymbol{\tau}_T}{\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T} \right) = \alpha_i + \mathbf{u}_{i.}'\mathbf{c}, \text{ for } i = 1, 2, \ldots, N, \tag{13}$$

where

$$\mathbf{c} = \mathbf{M}_F \boldsymbol{\tau}_T / \boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T. \tag{14}$$

Stacking the N estimates in Equation (13), we have

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \begin{pmatrix} \mathbf{u}_{1.}'\mathbf{c} \\ \mathbf{u}_{2.}'\mathbf{c} \\ \vdots \\ \mathbf{u}_{N.}'\mathbf{c} \end{pmatrix},$$

where $\mathbf{u}_{i.}'\mathbf{c} = \sum_{t=1}^{T} u_{it}c_t$, and c_t is the tth element of $\mathbf{c}$. Hence,

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \sum_{t=1}^{T} \mathbf{u}_t c_t, \tag{15}$$

whereas before $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})'$. Therefore, under Gaussianity,

$$\hat{\boldsymbol{\alpha}} \sim N\left(\boldsymbol{\alpha}, \frac{1}{\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T} \mathbf{V}\right).$$

Also, in the case where $T \geq N + m + 1$, an unbiased and invertible estimator of $\mathbf{V}$ is given by $\left(\frac{T}{T-m-1}\right)\hat{\mathbf{V}}$, where $\hat{\mathbf{V}}$ is the sample covariance matrix estimator

$$\hat{\mathbf{V}} = T^{-1} \sum\nolimits_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t', \tag{16}$$

$\hat{\mathbf{u}}_t = (\hat{u}_{1t}, \hat{u}_{2t}, \ldots, \hat{u}_{Nt})'$, $\hat{u}_{it}$ is the OLS residual from the regression of y_{it} on an intercept and $\mathbf{f}_t$.

Under Gaussianity, $\hat{\mathbf{u}}_t$ has a multivariate normal distribution with zero means, $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{u}}_t$ are independently distributed, and hence using standard results from multivariate analysis it follows that (see, e.g., Theorem 5.2.2 in Anderson, 2003) the GRS statistic (see p. 1124 of GRS)

$$\text{GRS} = \hat{W}_0 = \frac{T-N-m}{N}\left(\frac{\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T}{T}\right) \hat{\boldsymbol{\alpha}}' \hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\alpha}}, \tag{17}$$

is distributed *exactly* as a non-central F distribution with $(T-N-m)$ and N degrees of freedom, and the non-centrality parameter $\mu_\alpha^2 = \frac{T-N-m}{N}\left(\frac{\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T}{T}\right) \boldsymbol{\alpha}' \mathbf{V}^{-1} \boldsymbol{\alpha}$, which is zero under $H_0 : \boldsymbol{\alpha} = \mathbf{0}$.[6]

As noted in the introduction, the single most important limiting feature of the GRS and other related tests proposed in the literature is the requirement that T must be larger than N. Due to this, in applications of the GRS test, individual securities are grouped into (sub)-portfolios and the GRS test is then typically applied to 20–30 portfolios over relatively long time periods. However, the market efficiency hypothesis implies that $\alpha_i = 0$ for all *individual* securities which form the market portfolio, and it is clearly desirable to develop tests which permit N to be much larger than T. This is even more so if we would like to minimize the adverse effects of possible time variations in the $\boldsymbol{\beta}_i$'s.

It is also worth bearing in mind that the GRS test does not impose any restrictions on $\mathbf{V}$, which is possible only because N is taken to be fixed as $T \to \infty$. Large T is required to take account of non-Gaussian errors. While in the context of the approximate factor models advanced in Chamberlain (1983), the errors are at most weakly correlated, which places restrictions on the off-diagonal elements of $\mathbf{V}$ and its inverse. In addition, such restrictions are also statistically important in order to estimate $\mathbf{V}$ and its inverse when $N > T$. The test developed in this article for a large number of individual securities is therefore clearly different from the GRS test, both theoretically and statistically. Furthermore, as we shall see below, a test that exploits restrictions implied by the weak cross-sectional correlation of the errors is likely to have much better power properties than the GRS test that does not make use of such restrictions. Finally, being a multivariate F-test, the power of the GRS test is

6 Noting that $(1 + \bar{\mathbf{f}}' \hat{\boldsymbol{\Omega}}^{-1} \bar{\mathbf{f}})^{-1} = T^{-1}(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T)$, where $\bar{\mathbf{f}} = T^{-1} \sum_{t=1}^{T} \mathbf{f}_t$ and $\hat{\boldsymbol{\Omega}} = T^{-1} \sum_{t=1}^{T} (\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})'$, it is easily seen that Equation (17) can be written as the widely used expression of the GRS statistic, $\frac{T-N-m}{N}(1 + \bar{\mathbf{f}}' \hat{\boldsymbol{\Omega}}^{-1} \bar{\mathbf{f}})^{-1} \hat{\boldsymbol{\alpha}}' \hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\alpha}}$. As discussed in GRS, $\hat{\boldsymbol{\alpha}}' \hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\alpha}}$ measures the *ex post* maximum pricing error.

primarily driven by the time dimension, T, while for the analysis of a large number of assets or portfolios we need tests that have the correct size and are powerful for large N.

3 Large *N* Tests of Alpha in LFPMs

To develop large N tests of $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, we consider the following version of the GRS statistic, as set out in Equation (17),

$$W_v = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{V}^{-1} \hat{\boldsymbol{\alpha}}, \tag{18}$$

where we have dropped the degrees of freedom adjustment term and replaced $\hat{\mathbf{V}}$ by its true value. Under $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, and assuming that the errors are Gaussian we have $W_v \sim \chi^2_N$. Since the mean and the variance of a χ^2_N random variable are N and $2N$, one could consider a SW test statistic defined by

$$\mathrm{SW}_v = \frac{\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{V}^{-1} \hat{\boldsymbol{\alpha}} - N}{\sqrt{2N}}. \tag{19}$$

Under Gaussianity and $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, $\mathrm{SW}_v \rightarrow_d N(0,1)$ as $N \rightarrow \infty$. To construct tests of large N panels, a suitable estimator of $\mathbf{V}$ is required. But as was noted in the introduction this is possible only if we are prepared to impose restrictions on the structure of $\mathbf{V}$. In the case of LFPM regressions where the errors are at most weakly cross-sectionally correlated, this can be achieved by adaptive thresholding which sets to zero elements of $\mathbf{V}$ that are sufficiently small, or by use of shrinkage type estimators that put a substantial amount of weight on the diagonal elements of the shrinkage estimator of $\mathbf{V}$. Fan, Liao, and Mincheva (2011, 2013) consider consistent estimation of $\mathbf{V}$ in the context of an approximate factor model. They assume $\mathbf{V}$ is sparse and propose an adaptive threshold estimator, denoted as $\hat{\mathbf{V}}_{\mathrm{POET}}$, which they show to be positive definite with satisfactory small sample properties. We refer to the feasible SW test statistic which replaces $\mathbf{V}$ with $\hat{\mathbf{V}}_{\mathrm{POET}}$ as $\mathrm{SW}_{\mathrm{POET}}$ test.[7]

3.1 A $\hat{J}_\alpha$ Test for Large *N* Securities

To overcome some of the above mentioned limitations of the plug-in procedures, we avoid using an estimator of $\mathbf{V}$ altogether and base our proposed test on diagonal elements of $\mathbf{V}$, namely the $N \times N$ diagonal matrix, $\mathbf{D} = \mathrm{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{NN})$, with $\sigma_{ii} = E\left(u_{it}^2\right)$, rather than the full covariance matrix. Specifically, we consider the statistic

$$W_d = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{D}^{-1} \hat{\boldsymbol{\alpha}} = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \sum_{i=1}^{N} \left(\frac{\hat{\alpha}_i^2}{\sigma_{ii}}\right), \tag{20}$$

7 Another candidate is the shrinkage estimator of $\mathbf{V}$ proposed by Ledoit and Wolf (2004), which we denote by $\hat{\mathbf{V}}_{\mathrm{LW}}$, and refer to the associated SW statistic as $\mathrm{SW}_{\mathrm{LW}}$. Such "plug-in" approaches are subject to two important shortcomings. First, even if $\mathbf{V}$ can be estimated consistently, the test might perform poorly in the case of non-Gaussian errors. Notice that the standardization of the Wald statistic is carried out assuming Gaussianity. Further, consistent estimation of $\mathbf{V}$ in the Frobenius norm sense still requires T to rise faster than N, and in practice threshold estimators of $\mathbf{V}$ are not guaranteed to be invertible in finite samples where $N \gg T$.

and its feasible counterpart given by

$$\hat{W}_d = (\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)\hat{\boldsymbol{\alpha}}'\hat{\mathbf{D}}_v^{-1}\hat{\boldsymbol{\alpha}} = \left(\frac{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}{v^{-1}T}\right)\sum_{i=1}^{N}\left(\frac{\hat{\alpha}_i^2}{\hat{\sigma}_{ii}}\right), \tag{21}$$

where $\hat{\sigma}_{ii} = \hat{\mathbf{u}}_{i.}'\hat{\mathbf{u}}_{i.}/T$. The degrees of freedom $v = T - m - 1$ are introduced to correct for small sample bias of the test.[8] The infeasible statistic, W_d, can also be written as

$$W_d = \sum_{i=1}^{N} z_i^2, \tag{22}$$

where

$$z_i^2 = \hat{\alpha}_i^2(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)/\sigma_{ii}. \tag{23}$$

It is then easily seen that

$$\hat{W}_d = \sum_{i=1}^{N} t_i^2, \tag{24}$$

where t_i denotes the standard t-ratio of α_i given by

$$t_i^2 = \frac{\hat{\alpha}_i^2(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)}{v^{-1}T\hat{\sigma}_{ii}}. \tag{25}$$

As with the panel testing strategy developed in Im, Pesaran, and Shin (2003), a standardized version of $\hat{W}_d$, defined by Equation (24), can now be considered:

$$\frac{N^{-1/2}[\hat{W}_d - E(\hat{W}_d)]}{\sqrt{\mathrm{Var}(\hat{W}_d)}}, \tag{26}$$

where

$$N^{-1}E(\hat{W}_d) = E\left(t_i^2\right), \tag{27}$$

$$N^{-1}\mathrm{Var}(\hat{W}_d) = N^{-1}\mathrm{Var}\left(\sum_{i=1}^{N} t_i^2\right) = N^{-1}\sum_{i=1}^{N}\mathrm{Var}\left(t_i^2\right) + \frac{2}{N}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\mathrm{Cov}\left(t_i^2, t_j^2\right). \tag{28}$$

Under Gaussianity, the individual t_i statistics are identically distributed as Student's t with v degrees of freedom, and we have (assuming $v = T - m - 1 > 4$)

$$E(t_i^2) = \frac{v}{v-2}, \quad \mathrm{Var}(t_i^2) = \left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4}. \tag{29}$$

Using Equations (27)–(29), the standardized statistic (26) can now be written as

$$J_\alpha\left(\theta_N^2\right) = \frac{N^{-1/2}[\hat{W}_d - E(\hat{W}_d)]}{\sqrt{\mathrm{Var}(\hat{W}_d)}} = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{v}{v-2}\right)}{\sqrt{\left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4}\left(1 + \theta_N^2\right)}}, \tag{30}$$

8 Only securities with $\hat{\sigma}_{ii} > 0$ are included in $\hat{W}_d$.

where

$$\theta_N^2 = N^{-1} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \mathrm{Corr}\left(t_i^2, t_j^2\right), \tag{31}$$

and

$$\mathrm{Corr}(t_i^2, t_j^2) = \mathrm{Cov}(t_i^2, t_j^2)/[\mathrm{Var}(t_i^2)\mathrm{Var}(t_j^2)]^{1/2}.$$

To make the J_α test operational, we need to provide a large N consistent estimator of θ_N^2. Second, we need to show that, despite the fact that J_α test is standardized assuming t_i has a standard t distribution, the test will continue to have satisfactory small sample performance even if such an assumption does not hold due to the non-Gaussianity of the underlying errors. More formally, in what follows we relax the Gaussianity assumption and assume that $\mathbf{u}_t = \mathbf{Q}\boldsymbol{\varepsilon}_t$, where $\mathbf{Q}$ is an $N \times N$ invertible matrix, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$, and $\{\varepsilon_{it}\}$ is an IID process over i and t, with means zero and unit variances, and for some $c > 0$, $E(|\varepsilon_{it}|^{8+c})$ exists, for all i and t. Then, $E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \mathbf{V} = (\sigma_{ij}) = \mathbf{QQ}'$ and $\mathbf{V}$ is an $N \times N$ symmetric positive definite matrix, with $\lambda_{\min}(\mathbf{V}) \geq c > 0$. We allow for cross-sectional error heteroskedasticity, but assume that the errors are homoskedastic over time. This assumption can be relaxed by replacing the assumption of error independence by a suitable martingale difference assumption. This extension will not be attempted in this article.[9]

3.2 Sparsity Conditions on Error Correlation Matrix

As noted already, we advance on the literature by allowing $\mathbf{V} = (\sigma_{ij})$ to be *approximately* sparse. Equivalently, we define sparsity in terms of the elements of the correlation matrix $\mathbf{R} = (\rho_{ij})$, where $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$. We consider the following two conditions:

$$m_N = \max_{1 \leq i \leq N} \sum_{j=1}^{N} |\rho_{ij}| = O(N^{\delta_\rho}), \text{ with } 0 \leq \delta_\rho < 1/2, \tag{32}$$

and

$$\mathrm{Tr}(\mathbf{R}^2) = \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{ij}^2 = O(N). \tag{33}$$

Under condition (32), m_N is allowed to rise with N, but at a slower rate than $\sqrt{N}$. For example, consider the case where condition (32) applies to the first p rows of $\mathbf{R}$ (with p fixed), and the rest of the $N - p$ rows of $\mathbf{R}$ are absolute summable, namely

$$\sum_{j=1}^{N} |\rho_{ij}| = O(N^{\delta_\rho}), \text{ for } i = 1, 2, \ldots, p,$$

$$\sum_{j=1}^{N} |\rho_{ij}| = O(1), \text{ for } i = p+1, p+2, \ldots, N.$$

9 We conducted an experiment with GARCH(1,1) errors and the evidence supports our claim. The results are reported in Table 5.

Then, since $|\rho_{ij}|^2 \leq |\rho_{ij}|$, it readily follows that

$$\begin{aligned}\mathrm{Tr}(\mathbf{R}^2) &= \sum_{i=1}^{p}\left(\sum_{j=1}^{N}\rho_{ij}^2\right) + \sum_{i=p+1}^{N}\sum_{j=1}^{N}\rho_{ij}^2 \\ &\leq \sum_{i=1}^{p}\left(\sum_{j=1}^{N}|\rho_{ij}|\right) + \sum_{i=p+1}^{N}\sum_{j=1}^{N}|\rho_{ij}| \\ &\leq O(pN^{\delta_\rho}) + (N-p)O(1) = O(N), \text{ for } 0 \leq \delta_\rho < 1/2.\end{aligned}$$

Another important case covered by our sparsity assumption is when u_{it} has the weak factor structure given by Equation (8), with the factor loadings, γ_i, satisfying Equation (9). Denoting the correlation matrix of the idiosyncratic errors, $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, \ldots, \eta_{Nt})'$ by $\mathbf{R}_\eta = (\rho_{\eta,ij})$, and assuming that

$$\|\mathbf{R}_\eta\|_\infty < K, \tag{34}$$

we have $\mathrm{Tr}(N^{-1}\mathbf{R}_\eta^2) = O(1)$. It is now easily seen that Conditions (32) and (33) are also satisfied under this set up. Denoting the correlation matrix of $\mathbf{u}_t$ by $\mathbf{R} = (\rho_{ij})$, we have

$$\rho_{ij} = \tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_j + \left(\frac{\sigma_{\eta,ii}\sigma_{\eta,jj}}{\sigma_{ii}\sigma_{jj}}\right)^{1/2}\rho_{\eta,ij}, \tag{35}$$

where $\tilde{\boldsymbol{\gamma}}_i = \boldsymbol{\gamma}_i/\sigma_{ii}^{1/2} = \boldsymbol{\gamma}_i/\left(\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i + \sigma_{\eta,ii}\right)^{1/2}$. Since $|\rho_{ij}| \leq \sum_{s=1}^{k}|\tilde{\gamma}_{is}||\tilde{\gamma}_{js}| + |\rho_{\eta,ij}|$, then (note that $\sigma_{\eta,ii} \leq \sigma_{ii} = \boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i + \sigma_{\eta,ii}$)

$$\begin{aligned}m_N = \|\mathbf{R}\|_\infty &= \max_i \sum_{j=1}^{N}\sum_{s=1}^{k}|\tilde{\gamma}_{is}||\tilde{\gamma}_{js}| + \max_i \sum_{j=1}^{N}|\rho_{\eta,ij}| \\ &\leq k\left(\sup_{i,s}|\tilde{\gamma}_{is}|\right)\left(\max_s \sum_{j=1}^{N}|\tilde{\gamma}_{js}|\right) + \|\mathbf{R}_\eta\|_\infty.\end{aligned}$$

Since $\sup_{i,s}|\tilde{\gamma}_{is}| \leq \sup_{i,s}|\gamma_{is}|$, and $\sup_s \sum_{j=1}^{N}|\tilde{\gamma}_{js}| \leq \sup_s \sum_{j=1}^{N}|\gamma_{js}| = O(N^{\delta_\gamma})$, and by assumption $\|\mathbf{R}_\eta\|_\infty < K$, Condition (32) is met if $\delta_\rho \leq \delta_\gamma$. Also, (noting that $\sup_{i,s}|\tilde{\gamma}_{is}| \leq 1$)

$$\begin{aligned}N^{-1}\mathrm{Tr}(\mathbf{R}^2) &\leq N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{s=1}^{k}|\tilde{\gamma}_{is}||\tilde{\gamma}_{js}| + |\rho_{\eta,ij}|\right)^2 \\ &\leq N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{s=1}^{k}|\tilde{\gamma}_{is}||\tilde{\gamma}_{js}|\right)^2 + 2N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{k}|\tilde{\gamma}_{is}||\tilde{\gamma}_{js}| + N^{-1}\mathrm{Tr}\left(\mathbf{R}_\eta^2\right) \\ &\leq N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{s=1}^{k}|\tilde{\gamma}_{is}||\tilde{\gamma}_{js}|\right)^2 + 2N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{k}|\tilde{\gamma}_{is}||\tilde{\gamma}_{js}| + N^{-1}\mathrm{Tr}\left(\mathbf{R}_\eta^2\right) \\ &\leq (k^2+2k)N^{-1}\left(\sup_s \sum_{i=1}^{N}|\gamma_{is}|\right)^2 + N^{-1}\mathrm{Tr}\left(\mathbf{R}_\eta^2\right).\end{aligned}$$

Hence,

$$N^{-1}\mathrm{Tr}(\mathbf{R}^2) = N^{-1}\mathrm{Tr}\left(\mathbf{R}_\eta^2\right) + O(N^{2\delta_\gamma - 1}),$$

and under Conditions (9) and (34), $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ is bounded in N if $0 \leq \delta_\gamma \leq 1/2$.

Remark 1: *Our assumption of approximate sparsity allows for a sufficiently high degree of cross error correlation, which is important for the analysis of financial data, where it is not guaranteed that inclusion of observed factors in the return regressions will totally eliminate weak error correlations due to spatial and/or within sector error correlations. It is important that both factor and spatial type error correlations, representing strong and weak forms of interdependencies are taken into account when testing for alpha. By allowing the error term to include weak factors, one only needs to focus on identification of strong and semi-strong factors to be included in* $\mathbf{f}_t$. *On this see also Bailey, Kapetanios, and Pesaran (2021).*

3.3 Non-Gaussianity

For the discussion of the effects of non-Gaussianity on the J_α test below, it is convenient to introduce the following scaled error:

$$\xi_{it} = u_{it}/\sqrt{\sigma_{ii}}, \tag{36}$$

so that for each i, ξ_{it} has zero mean and unit variance. In the case where the errors are non-Gaussian the skewness and excess kurtosis of u_{it} are given by $\gamma_{1,i} = E(\xi_{it}^3)$ and $\gamma_{2,i} = E(\xi_{it}^4) - 3$, respectively, and could differ across i. Note that under non-Gaussian errors, t_i is no longer Student t-distributed and $E(t_i^2)$ and $V(t_i^2)$ need not be the same across i, due to the heterogeneity of $\gamma_{1,i}$ and $\gamma_{2,i}$ over i. Using a slightly extended version of the Laplace approximation of moments of the ratio of quadratic forms by Lieberman (1994), we are able to derive the following approximations of $E(t_i^2)$ and $\mathrm{Var}(t_i^2)$[10]:

$$E\left(t_i^2\right) = \frac{v}{v-2} + O(T^{-3/2}), \tag{37}$$

and

$$\mathrm{Var}\left(t_i^2\right) = \left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{(v-4)} + O(T^{-1}). \tag{38}$$

Substituting Equations (37) and (38) into Equation (26), we have the following non-Gaussian version of $J_\alpha(\theta_N^2)$, defined by Equation (30):

$$J_\alpha\left(\theta_N^2\right) = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{v}{v-2}\right) + O\left(\sqrt{N/T^3}\right)}{\sqrt{\left[\left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{(v-4)} + O(T^{-1})\right]\left(1+\theta_N^2\right)}},$$

where θ_N^2 is defined by Equation (31). When the numerator of the J_α statistic is replaced by $N^{-1/2}\sum_{i=1}^{N}(t_i^2 - 1)$, which is the typical mean adjustment employed by Fan, Liao, and

10 See Lemma 21 in the Supplementary Material of the article.

Yao (2015) and GOS, then the order of the asymptotic error of the numerator of such test statistics becomes $\sqrt{N/T^2}$. This is one of the reasons why our proposed test performs better than the ones proposed in the literature, especially in cases where $N \gg T$, and there are significant departures from Gaussianity. The asymptotic error of using $\left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{(v-4)}$ for $\mathrm{Var}(t_i^2)$ under non-Gaussianity in the J_α test is $O(T^{-1})$, which is small for sufficiently large T.[11]

3.4 Allowing for Error Cross-Sectional Dependence

A second important difference between the J_α test and the other tests proposed in the literature is the inclusion of θ_N^2 in the denominator of the test statistic to take account of error correlations. Using Equation (31), we first note that as N and $T \to \infty$[12]

$$\theta_N^2 - (N-1)\rho_N^2 \to 0, \tag{39}$$

so long as $N/T^2 \to 0$, and $0 \le \delta_\gamma < 1/2$, where

$$\rho_N^2 = \frac{2}{N(N-1)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \rho_{ij}^2. \tag{40}$$

Here, ρ_N^2 is known as the average pair-wise squared correlation coefficient and plays a key role in tests of error cross-sectional correlations in panel regressions (see, e.g., Breusch and Pagan, 1980; Pesaran, Ullah, and Yamagata, 2008). To see the relationship between θ_N^2 and the sparsity of $\mathbf{V}$, we note that

$$N^{-1}\mathrm{Tr}(\mathbf{R}^2) = 1 + \frac{2}{N} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \rho_{ij}^2 = 1 + (N-1)\rho_N^2,$$

which in view of Equation (39) justifies replacing $1 + \theta_N^2$ by $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ for N and T sufficiently large so long as $N/T^2 \to 0$, and $0 \le \delta_\gamma < 1/2$. Therefore, ignoring θ_N^2 can lead to serious size-distortions even for large N and T panels when the errors are cross-correlated *and* $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ *does not* tend to zero, since the denominator of J_α will be under-estimated. The size distortion will be present even if we impose stronger sparsity conditions on $\mathbf{V}$, for example, by requiring m_N, defined by Equation (32), to be bounded in N. It is, therefore, important that θ_N^2 (or ρ_N^2) is replaced by a suitable estimator.

One possible way of estimating ρ_N^2 would be to use sample correlation coefficients, $\hat{\rho}_{ij}$, defined as

$$\hat{\rho}_{ij} = \hat{\sigma}_{ij} / \sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}, \tag{41}$$

where $\hat{\sigma}_{ij} = T^{-1}\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}$ and $\hat{u}_{it}$ is the residuals from the OLS regression of $\mathbf{y}_i$ on $\mathbf{G} = (\boldsymbol{\tau}_T, \mathbf{F})$. However, such an estimator is likely to perform poorly in cases where N is large relative to T, and some form of thresholding is required, as discussed in the literature on estimation of large covariance matrices.[13] Here, we consider the application of the MT approach to regularization of large covariance matrices proposed by BPS. However, BPS

11 Small sample evidence on the efficacy of using $N^{-1/2}\sum_{i=1}^{N}(t_i^2 - \frac{v}{v-2})$ over $N^{-1/2}\sum_{i=1}^{N}(t_i^2 - 1)$ is reported in Table 7.

12 For a proof of Equation (39), see Lemma 18 in the Supplementary Material.

13 See, for example, Cai and Liu (2011); Fan, Liao, and Mincheva (2013); BPS, among others.

establish their results for $y_{it} - \bar{y}_i$, while we need to apply the thresholding approach to $\hat{u}_{it}$. Second, BPS consider exact sparsity conditions on the error covariance matrix, while we allow for much more general sparsity conditions. We extend BPS's analysis to address both of these issues.[14]

The MT estimator of ρ_{ij}, denoted by $\tilde{\rho}_{ij}$, is given by

$$\tilde{\rho}_{ij} = \hat{\rho}_{ij} I[|\sqrt{v}\hat{\rho}_{ij}| > c_p(N)], \tag{42}$$

where $v = T - m - 1$,

$$c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2N^{\delta}}\right), \tag{43}$$

p is the nominal significance level for testing $\rho_{ij} = 0$ $(0 < p < 1)$, $T = c_d N^d$, where c_d, δ, and d are finite positive constants. Using Equation (42), the MT estimator of ρ_N^2 is given by

$$\tilde{\rho}^2_{N,T} = \frac{2}{N(N-1)} \sum\nolimits_{i=2}^{N} \sum\nolimits_{j=1}^{i-1} \tilde{\rho}^2_{ij}. \tag{44}$$

Under the sparsity conditions (32) and (33), it can be shown that $(N-1)(\tilde{\rho}^2_{N,T} - \rho^2_N) \to_p 0$ as well as in l_1-norm, so long as $N/T^2 \to 0$ (or equivalently if $d > 1/2$) as N and $T \to \infty$, jointly, and if

$$\delta > \frac{(2-d)}{(1-\epsilon)} \varphi_{\max}, \tag{45}$$

for some small $\epsilon > 0$, where $\varphi_{\max} \leq 1 + |\gamma_{2,\varepsilon_\eta}|$, where $\gamma_{2,\varepsilon_\eta} = E(\varepsilon^4_{\eta,it}) - 3$, $\varepsilon_{\eta,it}$ is the ith element of the $N \times 1$ error vector $\boldsymbol{\varepsilon}_{\eta,t} = \mathbf{Q}_\eta^{-1} \boldsymbol{\eta}_t$, with $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, \ldots, \eta_{Nt})'$.[15] The critical value function, $c_p(N)$, depends on the nominal level of significance, p, and the choice of δ, subject to Condition (45). The test results are unlikely to be sensitive to the choice of p, over the conventional values in the range of 1–10%. d determines the relative expansion rate of N and T. The value of φ depends on the degree of dependence of the errors even if they are uncorrelated. In the case where the errors, $\varepsilon_{\eta,it}$, are Gaussian $\gamma_{2,\varepsilon_\eta} = 0$ and $\varphi \leq 1$, and it is sufficient to set $\delta = 2 - d$. However, in the non-Gaussian case, and given the evidence provided by Longin and Solnik (2001) and Ang, Chen, and Xing (2006) on the degree of nonlinear dependence of asset returns, higher values of δ might be required. In simulations and empirical exercises to be reported below, we set $f(N) = N$, which is equivalent to setting $\delta = 1$.[16]

Accordingly, we propose the following feasible version of the J_α statistic

$$\hat{J}_\alpha = \frac{N^{-1/2} \sum\limits_{i=1}^{N} \left(t_i^2 - \frac{v}{v-2}\right)}{\left(\frac{v}{v-2}\right) \sqrt{\frac{2(v-1)}{(v-4)} [1 + (N-1)\tilde{\rho}^2_{N,T}]}}, \tag{46}$$

14 Other thresholding estimators of **V** proposed in the literature can also be used.

15 See Theorem 4 in Section 4 and its proof in the Appendix.

16 The robustness of the J_a test against non-Gaussianity is investigated and reported in Table 7. These results are generally supportive of setting $\delta = 1$.

where t_i is the t-ratio for testing $\alpha_i = 0$, defined by Equation (25), $v = T - m - 1$, and $\tilde{\rho}^2_{N,T}$ is given by Equation (44). The $\hat{J}_\alpha$ test is robust to non-Gaussian errors and allows for a relatively high degree of error cross-sectional dependence. In the next section, we provide a formal statement of the conditions under which $\hat{J}_\alpha$ tends to a normal distribution.

3.5 Survivorship Bias

When applying the $\hat{J}_\alpha$ test, it is important to minimize the effect of survivorship bias. To this end, the GRS type tests of alpha consider a relatively small number of portfolios over a relatively large time period to achieve sufficient power. By making use of portfolios rather than individual securities, the GRS test is less likely to suffer from survivorship bias. By comparison, tests such as the $\hat{J}_\alpha$ test can suffer from the survivorship bias due to the fact that they are applied to individual securities directly and obtain power from increases in N as well as from T. To deal with the survivorship bias, we propose that the $\hat{J}_\alpha$ test is applied recursively to securities that have been trading for at least T time periods (days or months) at any given time t. The set of securities included in the $\hat{J}_\alpha$ test varies over time and dynamically takes account of exit and entry of securities in the market. The number of securities, N_τ, used in the test at any point of time, τ, depends on the choice of T, and declines as T is increased. It is clearly important that a balance is struck between T and N_τ. Since the $\hat{J}_\alpha$ test is applicable even if N is much larger than T, and given that the power of the $\hat{J}_\alpha$ test rises both in N and T, then it is advisable to set T such that $\min_\tau(N_\tau)/T^2$ is sufficiently small. This procedure is followed in the empirical application discussed in Section 6, where we set $T = 60$ and end up with N_τ in the range $[464, 487]$, giving $\min_\tau(N_\tau)/T^2 = 0.12$.

3.6 Other Existing Tests

3.6.1 The GOS test

It might be helpful to compare our proposed test statistic $\hat{J}_\alpha$, given by Equation (46), with the one proposed by Gagliardini et al. (2016, pp. 1008–9):

$$\mathrm{GOS} = \frac{N^{-1/2} \sum_{i=1}^{N} \left(t_i^2 - 1\right)}{\sqrt{2[1 + (N-1)\hat{\rho}^2_{\mathrm{BL}}]}}, \tag{47}$$

where $\hat{\rho}^2_{\mathrm{BL}}$ is an estimator of ρ^2_N based on Bickel and Levina (2008, BL) threshold estimator of ρ_{ij}.[17] As noted in the introduction, the GOS statistic is closely related to the $\hat{J}_\alpha$ test statistic, and also differ from it in a number of important respects. First, GOS do not employ the degrees of freedom adjustment for the standardization of t_i^2, which we have shown will provide more accurate normal approximation especially when N is much larger than T. Despite the simplicity of the corrections, as can be seen from the Appendix and the Supplementary Material, the derivations and the proofs are not straightforward. Second, for the estimation of large variance–covariance matrix, the evidence in BPS shows that the MT estimator outperforms the BL estimator almost uniformly in their experiments, and our use of MT estimator of ρ^2_N turns out to yield much better results. Third, the BL estimation requires cross-validation, which can be computationally far more costly than the MT estimation. Finally, we derive limiting distribution of the $\hat{J}_\alpha$ test statistic under primitive

17 For more details, see Supplementary Section M1.1.

assumptions with fairly general error covariance structure, while GOS place the high level assumption of asymptotic normality of the test statistic (see their Assumption A.5) or only consider a restrictive error covariance structure (see their Appendix F).[18] We believe that our error specification is valid more generally in empirical asset pricing literature where not all factors can be identified and estimated, and in consequence one needs to allow for a much wider degree of error cross-correlations to take account of weak unobserved effects.

3.6.2 The GL $F_{\max}$

GL propose a resampling test based on $F_i = t_i^2$ test statistic for $\alpha_i = 0$, defined as

$$F_{\max} = \max_{1 \le i \le N} F_i. \tag{48}$$

They consider various versions of the test, and recommend the use of the maximum test which we will consider in our Monte Carlo exercise. The authors claim that their resampling test procedure is robust against non-normality and cross-sectional error dependence.[19] Their test effectively makes use of wild bootstrap resampling aimed at preserving the sample residual cross-sectional correlations, and deals with nuisance parameters by the introduction of a bounds testing procedure.

3.6.3 The BS and SD tests in He et al. (2021)

He et al. (2021) consider the following two test statistics. Based on BS, He et al. (2021) propose a SW type test which requires N and T to be of the same order of magnitude:

$$\mathrm{BS} = \frac{\left(T^{-1}\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}}/ - \mathrm{Tr}\left(\hat{\mathbf{V}}\right)/T}{c_1\{\mathrm{Tr}\left(\hat{\mathbf{V}}^2\right) - c_2[\mathrm{Tr}\left(\hat{\mathbf{V}}\right)]^2\}^{1/2}}, \tag{49}$$

where $c_1 = \frac{2(T-1)}{(T-2)(T-1)}$ and $c_2 = \frac{1}{T-1}$. Based on Srivastava and Du (2008), He et al. (2021) also propose a test statistic which is a standardized squared t-ratio, using different standardization from ours:

$$\mathrm{SD} = \frac{\left(T^{-1}\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)\hat{\boldsymbol{\alpha}}'\hat{\mathbf{D}}_v^{-1}\hat{\boldsymbol{\alpha}} - c_3}{\{c_4[\mathrm{Tr}\left(\hat{\mathbf{D}}_v^2\right) - c_5][1 + \mathrm{Tr}\left(\hat{\mathbf{D}}_v^2\right)/N^{3/2}]^2\}^{1/2}}, \tag{50}$$

where $c_3 = \frac{N(T-1)}{T(T-3)}$, $c_4 = \frac{2}{T^2}$, $c_5 = \frac{N^2}{T-1}$.

4 Summary of the Main Theoretical Results

In this section, we provide the list of assumptions and a formal statement of the theorems for the size and power of the proposed $\hat{J}_\alpha$. First, we state the assumptions required for establishing the results.

Assumption 1: The $m \times 1$ vector of common observed factors, $\mathbf{f}_t$, in the return regressions, Equation (6), are distributed independently of the errors, $u_{it'}$ for all i, t,

18 See Assumptions BD.1–3 in GOS.

19 We are grateful to Richard Luger for sharing the code to compute the resampling test.

and t'. The number of factors, m, is fixed, and at least one of the factors is strong, in the sense that

$$\sup_s \sum_{i=1}^{N} |\beta_{is}| = \ominus(N), \tag{51}$$

and the factors satisfy $\mathbf{f}_t'\mathbf{f}_t \leq K < \infty$, for all t. The $(m+1)\times(m+1)$ matrix $T^{-1}\mathbf{G}'\mathbf{G}$, with $\mathbf{G} = (\boldsymbol{\tau}_T, \mathbf{F})$, is a positive definite matrix for all T, and as $T \to \infty$, and $\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T > 0$, where $\mathbf{M}_F = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$.

Assumption 2: The errors, u_{it}, in Equation (6), have the following mixed weak-factor spatial representation

$$u_{it} = \boldsymbol{\gamma}_i'\mathbf{v}_t + \eta_{it}, \text{ for } i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T, \tag{52}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{ik})'$ is a $k \times 1$ vector of factor loadings, $\mathbf{v}_t = (v_{1t}, v_{2t}, \ldots, v_{kt})'$ is a $k \times 1$ vector of unobserved common factors, and η_{it} are the idiosyncratic components.

i. The unobserved factors $\mathbf{v}_t$ are serially independent and the k elements are independent of each other, such that $\mathbf{v}_t \sim \text{IID}\left(\mathbf{0}, \mathbf{I}_k\right)$, $\gamma_{2,v} = E\left(v_{st}^4\right) - 3$, and $\sup_{s,t} E\left(v_{st}^{8+c}\right) < K$, for some $c > 0$. The factor loadings, γ_{is} for $s = 1, 2, \ldots, k$, are bounded, $\sup_{i,s} |\gamma_{is}| < K$, and the factors, $\mathbf{v}_t$, are weak in the sense that

$$\sup_s \sum_{i=1}^{N} |\gamma_{is}| = O\left(N^{\delta_\gamma}\right), \text{ with } 0 \leq \delta_\gamma < 1/2. \tag{53}$$

ii. For any i and j, the T pairs of realizations, $\{(\eta_{i1}, \eta_{j1}), (\eta_{i2}, \eta_{j2}), \ldots, (\eta_{iT}, \eta_{jT})\}$, are independent draws from a common bivariate distribution with mean $E(\eta_{it}) = 0$, $\text{Var}(\eta_{it}) = \sigma_{\eta,ii}$, $0 < c < \sigma_{\eta,ii} \leq K$, and the covariance $E(\eta_{it}\eta_{jt}) = \sigma_{\eta,ij}$.

Writing the error factor specification, Equation (52), in matrix notation we have

$$\mathbf{u}_t = \boldsymbol{\Gamma}\mathbf{v}_t + \boldsymbol{\eta}_t, \tag{54}$$

where $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})'$, $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_N)'$, and $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, \ldots, \eta_{Nt})'$. Under Assumption 2, and denoting $E\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right) = \mathbf{V}_\eta = (\sigma_{\eta,ij})$, we have

$$E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \mathbf{V}_\eta = \mathbf{V} = (\sigma_{ij}), \text{ with } \sigma_{ij} = \boldsymbol{\gamma}_i'\boldsymbol{\gamma}_j + \sigma_{\eta,ij}. \tag{55}$$

Assumption 3: The covariance matrices $\mathbf{V}$ and $\mathbf{V}_\eta$ defined by Equation (55) are $N \times N$ symmetric, positive definite matrices with $\lambda_{\min}(\mathbf{V}) \geq \lambda_{\min}(\mathbf{V}_\eta) \geq c$,

$$\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})' = \mathbf{Q}^{-1}\mathbf{u}_t, \text{ and } \boldsymbol{\varepsilon}_{\eta,t} = (\varepsilon_{\eta,1t}, \varepsilon_{\eta,2t}, \ldots, \varepsilon_{\eta,Nt})' = \mathbf{Q}_\eta^{-1}\boldsymbol{\eta}_t, \tag{56}$$

where $\mathbf{Q}$ and $\mathbf{Q}_\eta$ are the Cholesky factors of $\mathbf{V}$ and $\mathbf{V}_\eta$, respectively. Matrix $\mathbf{Q}_\eta$ is row and column bounded in the sense that

$$\|\mathbf{Q}_\eta\|_\infty < K \text{ and } \|\mathbf{Q}_\eta\|_1 < K. \tag{57}$$

Here, $\{\varepsilon_{it}\}$ and $\{\varepsilon_{\eta,it}\}$ are IID processes over i and t, with zero means, unit variances, $\gamma_{2,\varepsilon_\eta} = E(\varepsilon_{\eta,it}^4) - 3$, $\sup_{i,t} E(|\varepsilon_{it}|^{8+c}) \leq K < \infty$, and $\sup_{i,t} E(|\varepsilon_{\eta,it}|^{8+c}) \leq K < \infty$, for some $c > 0$.

Remark 2: *The above assumptions allow the returns on individual securities to be strongly cross-sectionally correlated through the observed factors,* $\mathbf{f}_t$*, and allow for weak error cross-correlations once the effects of strong factors are removed.*

Remark 3: *Under Condition (57)*

$$\|\mathbf{V}_\eta\|_\infty \leq \|\mathbf{Q}_\eta \mathbf{Q}_\eta'\|_\infty \leq \|\mathbf{Q}_\eta\|_\infty \|\mathbf{Q}_\eta\|_1 < K = O(1), \tag{58}$$

nevertheless due to the weak factors we have

$$\|\mathbf{V}\|_\infty = \sup_j \sum_{i=1}^N |\sigma_{ij}| = O(N^{\delta_\gamma}),$$

and allow the overall error variance matrix, **V***, to be approximately sparse, in contrast to the literature that requires* $\|\mathbf{V}\|_\infty < K$*. The relaxation of the sparsity condition on* **V** *is particularly important in finance where security returns could be affected by weak unobserved factors.*

Remark 4: *The high-order moment conditions in Assumption 3 allow us to relax the Gaussianity assumption while at the same time ensuring that our test is applicable even if N is much larger than T.*

Remark 5: *Assumptions 2(ii) and 3 ensure that the sample cross-correlation coefficients of the residuals,* $\hat{\rho}_{ij}$*, have an Edgeworth expansion which is needed for consistent estimation of* ρ_N^2*, defined by Equation (40). For further details, see BPS.*

Our main theoretical results are set out in the following theorems. The proofs of these theorems are provided in the Appendix, and necessary lemmas for the proofs are given in the Supplementary Material.

Theorem 1: *Consider the return regression (6), and the statistic* $q_{NT} = N^{-1/2} \sum_{i=1}^N (z_i^2 - 1)$*, where* z_i^2 *is defined by Equation (23). Suppose that Assumptions 1–3 hold, and* $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ *is bounded in N, where* $\mathbf{R} = (\rho_{ij})$, $\rho_{ij} = E(\xi_{it}\xi_{jt})$*, and* $\xi_{it} = u_{it}/\sigma_{ii}^{1/2}$ *is the standardized error of the return regression Equation (6). Then, under* $H_0 : \alpha_i = 0$*, for all i,*

$$q_{NT} = N^{-1/2} \sum_{i=1}^N \left(z_i^2 - 1\right) \to_d N(0, 2\omega^2), \tag{59}$$

as $N \to \infty$ *and* $T \to \infty$*, jointly, where*

$$\omega^2 = \lim_{N\to\infty} N^{-1} Tr(\mathbf{R}^2) = 1 + \lim_{N\to\infty} (N-1)\rho_N^2,$$

and

$$\rho_N^2 = \frac{2}{N(N-1)} \sum_{i=2}^N \sum_{j=1}^{i-1} \rho_{ij}^2. \tag{60}$$

Theorem 2: *Consider the regression model (6), and the statistic S_{NT}, given below, where z_i^2 and t_i^2 are defined by Equations (23) and (25), respectively. Suppose that Assumptions 1–3 hold. Then, under the null hypothesis, $H_0: \alpha_i = 0$ for all i,*

$$S_{NT} = N^{-1/2}\sum_{i=1}^{N}\left(z_i^2 - t_i^2\right)\rightarrow_p 0,$$

as $N\to\infty$ and $T\to\infty$ jointly, so long as $N/T^2\to 0$, $0\le\delta_\gamma<1/2$, where δ_γ is defined by Equation (53).

Theorem 3: *Consider the regression model (6), and suppose that Assumptions 1–3 hold. Then, under $H_0: \alpha_i = 0$, for all i,*

$$J_\alpha(\rho_N^2) = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{v}{v-2}\right)}{\sqrt{\left(\frac{v}{v-2}\right)^2\frac{2(v-1)}{v-4}\left[1+(N-1)\rho_N^2\right]}}\rightarrow_d N(0,1), \tag{61}$$

so long as $N/T^2\to 0$, and $0\le\delta_\gamma<1/2$, as $N\to\infty$ and $T\to\infty$, jointly, where t_i, ρ_N^2, and δ_γ are defined by Equations (25), (60), and (53), respectively, with $v = T-m-1$.

Theorem 4: *Let*

$$\tilde{\rho}_{N,T}^2 = \frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\tilde{\rho}_{ij}^2, \tag{62}$$

where

$$\tilde{\rho}_{ij} = \hat{\rho}_{ij}I[|\sqrt{v}\hat{\rho}_{ij}| > c_p(N)], \tag{63}$$

$\rho_{ij} = E(\xi_{it}\xi_{jt})$, $\xi_{it} = u_{it}/\sigma_{ii}^{1/2}$, $v = T-m-1$, $\hat{\rho}_{ij}$ is defined by Equation (41)

$$c_p(N) = \Phi^{-1}\left(1-\frac{p}{2f(N)}\right), \tag{64}$$

p is the nominal p-value ($0<p<1$), $f(N) = N^\delta$, and $T = c_dN^d$, where c_d, δ, and d are finite positive constants. Suppose that Assumptions 1–3 hold and

$$\sum_{i,j=1}^{N}|\rho_{ij}| = O(N), \tag{65}$$

then, $(N-1)E|\tilde{\rho}_{N,T}^2-\rho_N^2|\to 0$, as N and $T\to\infty$, which implies $(N-1)(\tilde{\rho}_{N,T}^2-\rho_N^2)\rightarrow_p 0$, if $N/T^2 = \ominus(N^{1-2d})\to 0$, (or if $d>1/2$), and if $\delta>\frac{(2-d)}{(1-\epsilon)}\varphi_{\max}$, for some small $\epsilon>0$, where $\varphi_{\max}\le 1+|\gamma_{2,\varepsilon_\eta}|$ and $\gamma_{2,\varepsilon_\eta} = E(\varepsilon_{\eta,it}^4)-3$ (Assumption 3).

Theorem 5: *Consider the panel regression model (6) in asset returns and suppose that Assumptions 1–3 hold. Consider the test statistic*

$$\hat{J}_\alpha = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2-\frac{v}{v-2}\right)}{\left(\frac{v}{v-2}\right)\sqrt{\frac{2(v-1)}{(v-4)}\left[1+(N-1)\tilde{\rho}_{N,T}^2\right]}}, \tag{66}$$

where t_i *is given by* Equation (25), $v = T - m - 1$, $\tilde{\rho}^2_{N,T}$ *is defined by* Equation (62), *using the threshold* $c_p(N)$ *given by* Equation (64), *with* p $(0 < p < 1)$, $T = c_d N^d$, *where* c_d, δ, *and* d *are finite positive constants,* $\delta > \frac{(2-d)}{(1-\epsilon)}\varphi_{\max}$, *for some small* $\epsilon > 0$, *where* $\varphi_{\max} \leq 1 + |\gamma_{2,\varepsilon_\eta}|$ *and* $\gamma_{2,\varepsilon_\eta} = E(\varepsilon^4_{\eta,it}) - 3$. *Then, under* $H_0 : \alpha_i = 0$ *for all* i,

$$\hat{J}_\alpha \rightarrow_d N(0,1), \tag{67}$$

if $N/T^2 \rightarrow 0$, *as* N *and* $T \rightarrow \infty$, *jointly.*

To investigate the power properties of the $\hat{J}_\alpha$ test, we consider the local alternatives

$$H_{0a} : \alpha_i = \frac{\varsigma_i}{N^{1/4}T^{1/2}}, \text{ with } 0 < |\varsigma_i| < \infty, \text{ for all } i, \tag{68}$$

and establish the following theorem.

Theorem 6: *Consider the panel regression model* (6) *in asset returns, and suppose that conditions of Theorem 5 apply, and* $inf(\sigma_{ii}) > c > 0$. *Then, under the local alternatives,* $H_{0\alpha}$, *defined by* Equation (68),

$$\hat{J}_\alpha \rightarrow_d N\left(\phi^2/\sqrt{2}, 1\right), \tag{69}$$

where $\phi^2 = lim_{N\rightarrow\infty}\phi^2_N$ *and*

$$\phi^2_N = \frac{1}{N}\sum_{i=1}^{N}\varsigma_i^2/\sigma_{ii}. \tag{70}$$

Remark 7: *This theorem establishes that the* $\hat{J}_\alpha$ *test is consistent (in the sense that its power tends to unity), if* $\phi^2 > 0$. *It is also of interest that the power of the* $\hat{J}_\alpha$ *test increases uniformly with N and T, in contrast to the power of the GRS test that rises with T, only.*

Remark 8: *The above theorem also sheds light on the effects of allowing for pricing errors on the size and power of the* $\hat{J}_\alpha$ *test. It is clear that adding pricing errors* ϖ_i *to* α_i *in* Equation (68) *will increase* ϕ^2_N *and hence the power of the test. But this will be at the expense of size distortions since the null of the test is* $H_0 : \alpha_i = 0$ *while if we allow for the pricing errors the null will be* $H_0' : \alpha_i = \varpi_i$, *subject to the APT condition* $\sum_{i=1}^N \varpi_i^2 = O(N^{\delta_\varpi})$, *with* $\delta_\varpi = 0$. *(see* Equation (4)*). Under* H_0' *and the alternatives,* H_{0a} *in* Equation (68) *we have*

$$\sum_{i=1}^{N}\alpha_i^2 = \sum_{i=1}^{N}\varpi_i^2 = O(N^{\delta_\varpi}),$$

and

$$\sum_{i=1}^{N}\alpha_i^2 = \frac{1}{N^{1/2}T}\sum_{i=1}^{N}\varsigma_i^2.$$

These two conditions hold simultaneously if $N^{-1/2}T^{-1}\sum_{i=1}^{N}\varsigma_i^2 = O(N^{\delta_\varpi})$, *which in turn implies that*

$$\phi_N^2 = \frac{1}{N}\sum_{i=1}^{N}\varsigma_i^2/\sigma_{ii} \leq \inf_i(1/\sigma_{ii})\left(N^{-1}\sum_{i=1}^{N}\varsigma_i^2\right) = O\left(Tn^{\delta_\varpi - 1/2}\right).$$

Setting $T = \ominus(N^d)$, *we now have* $\phi_N^2 = O(N^{\delta_\varpi + d - 1/2})$ *and the* $\hat{J}_\alpha$ *test will have the correct size under* H_0' *if* $d < 1/2 - \delta_\varpi$. *Under Ross's APT condition where* $\delta_\varpi = 0$, *it is required that* $d < 1/2$. *But to allow for non-Gaussian errors and weak error cross-sectional dependence we require* $d > 1/2$ *so that* $N/T^2 \to 0$, *which is one of the conditions of Theorem 5. Hence, we would expect some size distortions if we allow pricing errors that satisfy the APT condition of Ross (1976). To avoid size distortions in the presence of pricing errors, we need to consider stronger restrictions on pricing errors so that they decline with N, for example,* $\varpi_i = O(N^{-\epsilon})$. *Under this specification, since* $\sum_{i=1}^{N}\varpi_i^2 = O(N^{1-2\epsilon})$, *then* $\delta_\varpi = 1 - 2\epsilon$, *and pricing errors can be accommodated in our analysis if* $\epsilon > d/2 + 1/4$. *Since Theorem 5 requires* $d > 1/2$, *then we must have* $\epsilon > 1/2$.

Remark 9: *Pricing errors cannot be allowed for in the case of the GRS test since it requires* $N < T$, *and with N fixed it is not possible to distinguish* α_i *from* ϖ_i *in the LFPM given by Equation (6).*

5 Small Sample Evidence Based on Monte Carlo Experiments

We examine the finite sample properties of the $\hat{J}_\alpha$ test by Monte Carlo experiments, and compare its performance to the existing tests, which are discussed in Section 3.6. Specifically, we consider the *GRS* test, the *GOS* test, and a feasible version of the SW test, as well as the distribution-free $F_{\max}$ test and the BS and SD tests, which are defined by Equations (3), (47), (19), (48), (49), and (50), respectively. Computational details of these tests are given in Section M1.1 of the Supplementary Material.

5.1 Monte Carlo Designs and Experiments

We consider the following data generating process (DGP):

$$r_{it} = \alpha_i + \sum_{l=1}^{3}\beta_i f_{lt} + \kappa\, u_{it}, \tag{71}$$

for $i = 1, 2, .., N$; $t = 1, 2, \ldots, T$, where $f_{\ell t}$ for $l = 1, 2, 3$ are the observed factors, and

$$u_{it} = \gamma_i v_t + \eta_{it}, \tag{72}$$

in which v_t is the missing factor and η_{it} is the idiosyncratic component of the return process defined below. The scalar coefficient κ is introduced so that the overall fit of the panel can be controlled to match the average fit of the return regressions defined by $R_{NT}^2 = N^{-1}\sum_{i=1}^{N} R_{iT}^2$, where R_{iT}^2 is the *R*-squared of the regression for r_{it}, computed for a given sample. We calibrate $\kappa = 6.5$ for $N = 500$ and $T = 120$ to match $R_{NT}^2 = 0.30$ for the model without omitted common component and spatial errors. The value of κ is fixed throughout the experiments.

The observed factors are calibrated to closely match the three Fama and French (1993, FF3) factors (market factor, HML, and SMB) and are generated as[20]

$$f_{\ell t} = \rho_{f,\ell} f_{\ell,t-1} + e_{\ell t},$$

where $(\rho_{f,1}, \rho_{f,2}, \rho_{f,3}) = (-0.1, 0.2, -0.2)$, $e_{\ell t} = \sqrt{h_{\ell t}}\xi_{\ell t}$ with $\xi_{\ell t} \sim \text{IIDN}(0, 1)$ follow GARCH(1,1) models:

$$h_{\ell t} = \omega_\ell(1 - \varrho_\ell - \varphi_\ell) + \varrho_\ell h_{\ell,t-1} + \varphi_\ell e^2_{\ell,t-1},$$

where $(\omega_1, \omega_2, \omega_3) = (20.25, 6.33, 5.98)$, $(\varrho_1, \varrho_2, \varrho_3) = (0.61, 0.70, -0.31)$, and $(\varphi_1, \varphi_2, \varphi_3) = (0.31, 0.21, 0.10)$.[21]

To calibrate the empirical FF3 model, we estimated it using S&P500 security level monthly excess return for 120 months ending on April 2018. We chose the series with the full sample period, which left 457 securities. The results are summarized in Table 1.

We generate the factor loadings as $\text{IIDU}(0.3, 1.8)$ for the market factor, $\text{IIDU}(-1.0, 1.0)$ for the HML factor, and $\text{IIDU}(-0.6, 0.9)$ for the SMB factor. In this way, we ensure that the means and standard deviations of the betas match their empirical counterparts and sufficient ranges of the estimates of $\beta's$ reported in Table 1 for the FF3 model are covered in the experiments.

The latent factor v_t is generated as IID(0, 1) and its loadings γ_i are generated to ensure a given factor strength denoted by the exponent δ_γ. We generate γ_i as

$$\begin{aligned} \gamma_i &\sim \text{IIDU}(0.7, 0.9), \text{ for } i = 1, 2, \dots, \lfloor N^{\delta_\gamma} \rfloor \\ \gamma_i &= 0, \text{ for } \lfloor N^{\delta_\gamma} \rfloor + 1, \lfloor N^{\delta_\gamma} \rfloor + 2, \dots, N, \end{aligned}$$

and to avoid systematic errors we then randomly reshuffle γ_i over i before assigning them to the individual returns, r_{it}. Our theoretical derivations suggest that the size of our proposed $\hat{J}_\alpha$ test should be under control so long as $\delta_\gamma < 1/2$. Accordingly, we consider the values of $\delta_\gamma = 0$, 1/4, and 1/2. Allowing for latent factors is important since in practice researchers cannot be sure that they have included all relevant risk factors in their models. The problem of missing (or latent) factors continues to apply even if we extend the list of observed factors as it is done in the recent literature. See, for example, Giglio and Xiu (2021) and the recent paper by Bailey, Kapetanios, and Pesaran (2021) who consider the estimation of factor strength.

In addition to allowing for latent factors, we also consider network (or spatial) type cross-sectional error dependence by generating the idiosyncratic errors $\varepsilon_{\eta,it}$ as

$$\begin{aligned} u_{it} &= \gamma_i v_t + \eta_{it}, \\ \text{Var}(u_{it}) &= \gamma_i^2 \text{Var}(v_t) + \text{Var}(\eta_{it}) \\ \eta_{it} &= \psi \sum_{j=1}^{N} w_{ij}\eta_{jt} + \sigma_{\eta i}\varepsilon_{\eta,it}, \text{ for } i = 1, 2, \dots, N, \end{aligned} \tag{73}$$

20 SMB stands for "small market capitalization minus big" and HML for "high book-to-market ratio minus low." See Fama and French (1993).

21 The estimates used in the generation of the factors and their volatilities are computed using monthly observations over the period May 2008–April 2018.

Table 1. Descriptive statistics of Fama–French three factor regression results

	Average β estimates for FF3 factors			Average skewness and excess kurtosis of the residuals	
	$\hat{\beta}_{\text{MKT}}$	$\hat{\beta}_{\text{HML}}$	$\hat{\beta}_{\text{SMB}}$	Skewness	Excess kurtosis
Mean	1.05	0.07	0.18	0.32	2.76
SD	0.43	0.57	0.45	0.87	5.61
Median	1.02	0.00	0.17	0.14	1.19
Min	0.19	−1.46	−1.95	−1.53	−0.53
Max	2.92	2.91	1.99	6.34	57.57

which can be solved for $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, \dots, \eta_{Nt})'$ as

$$\boldsymbol{\eta}_t = (\mathbf{I}_N - \psi\mathbf{W})^{-1}\mathbf{D}_\eta\boldsymbol{\varepsilon}_{\eta,t},$$

where $\boldsymbol{\varepsilon}_{\eta,t} = (\varepsilon_{\eta,1t}, \varepsilon_{\eta,2t}, \dots, \varepsilon_{\eta,Nt})'$, $\psi = \{0.0, 0.25\}$, $\mathbf{D}_\eta = \text{diag}(\sigma_{\eta 1}, \sigma_{\eta 2}, \dots, \sigma_{\eta N})'$. We adopt a rook form of $\mathbf{W} = (w_{ij})$, where all elements in $\mathbf{W}$ are zero except $w_{i+1,i} = w_{j-1,j} = 0.5$ for $i = 1, 2, \dots, n-2$ and $j = 3, 4, \dots, n$, with $w_{1,2} = w_{n,n-1} = 1$, and standardized such that $w_{ii} = 0$ and $\sum_{j=1}^{N} w_{ij} = 1$. Case of error cross-sectional independence arises for the parameter values $\psi = 0$ and $\delta_\gamma = 0$. We allow for error cross-sectional heteroskedasticity by generating $\sigma_{\eta i}^2$ as IID $(1 + \chi_{2,i}^2)/3$, and consider Gaussian (1) $\varepsilon_{\eta,it} \sim \text{IIDN}(0, 1)$, as well as non-Gaussian errors, (2) $\varepsilon_{\eta,it} \sim \text{IID}\frac{t_{\nu,it}}{[\nu/(\nu-2)]^{1/2}}$, where $t_{\nu,it}$ are independent draws from a t-distribution with ν degrees of freedom. In light of the properties of the empirical distribution of the FF3 regression residuals, for t distribution error, we choose $\nu = 8$, so that the value of excess kurtosis, 1.5, falls between the sample mean and sample median shown in Table 1.

All the N return series are generated from $t = -49, -48, \dots .0, 1, 2, \dots, T$, with $f_{\ell,-50} = 0$ and $h_{\ell,-50} = 1$ for $\ell = 1, 2, 3$. The first 50 observations are dropped to minimize the effects of the initial values and observations r_{it}, $\mathbf{f}_t = (f_{1t}, f_{2t}, f_{3t})'$, for $t = 1, 2, \dots, T$ are used in the MC experiments. Further details are provided in the Supplementary Material.

To estimate size of the tests, we set $\alpha_i = 0$ for all i. To investigate power, we consider alternatives based on Equation (5), setting $\lambda_0 = 0$, namely

$$\alpha_i = \boldsymbol{\beta}_i'(\boldsymbol{\lambda} - \boldsymbol{\mu}) + \varpi_i.$$

For the scenario called "Power 1," we set $\boldsymbol{\lambda} = \boldsymbol{\mu}$, and generated α_i as $\alpha_i = \varpi_i \sim \text{IIDN}(0, 1)$ for $i = 1, 2, \dots, N_\alpha$ with $N_\alpha = \lfloor N^{\delta_\alpha} \rfloor$; $\alpha_i = 0$ for $i = N_\alpha + 1, N_\alpha + 2, \dots, N$. We considered the values $\delta_\alpha = 0.7$. In another scenario called "Power 2," we assume there are no pricing errors and set $\varpi_i = 0$ for all i, but consider the case where $\boldsymbol{\lambda} - \boldsymbol{\mu} = c(2.92, -0.63, -9.96)'$, that match the estimates reported in Table 1 of GOS (p. 1011) for $c = 1$. To make the power of the tests for "Power 2" comparable for "Power 1," we set $c = 0.1$. We do not consider the case both $\boldsymbol{\lambda} \neq \boldsymbol{\mu}$ and $\varpi_i \neq 0$, as it is clear that in this case higher power will be achieved.

All combinations of $T = 60$, 120, 240 and $N = 50$, 100, 200 (and 500, 1000, 2000, 5000 for the $\hat{J}_\alpha$ test) are considered. All tests are conducted at the 5% significance level and all experiments are based on $R = 2000$ replications. To compute $\tilde{\rho}_{N,T}^2$ which enters the denominator of the $\hat{J}_\alpha$ statistic, given by Equation (46), we consider $p = \{0.05, 0.1\}$ and $\delta = \{1, 2\}$. The results are very insensitive to the choice of the values of (p, δ) and the case for $(p, \delta) = (0.05, 1)$ is reported. It is worth noting that that the choice of p when computing $\tilde{\rho}^2_{N,T}$ is not governed or affected by the choice of the nominal size of the $\hat{J}_\alpha$ test.

5.2 Size and Power

Table 2 reports the size and power of the $\hat{J}_\alpha$, GRS, GOS, SW, $F_{\max}$, BS, and SD tests in the case of normal errors, under various degrees of cross-sectional error correlations, as measured by the exponent, δ_γ.

First, consider Panel A of Table 2 which reports the size of the tests. The GRS test when applicable (namely when $T > N$) is an exact test and has the correct size. The empirical size of the $\hat{J}_\alpha$ test is also very close to the 5% nominal level for all combinations of N and T. Even when $N = 200$ and $\delta_\gamma = 0.5$, the size of the $\hat{J}_\alpha$ test lies in the range 5.9–6.4% for different values of T. In contrast, both GOS and SW tests grossly over-reject the null hypothesis, and the degree of the over-rejection becomes more serious as N increases for a given T. In line with the discussion in Section 3.4, the size distortion of these tests is mitigated when T increases. The $F_{\max}$ test severely under-rejects the null hypothesis, with the size ranging between 0.0% and 0.4%. Although less pronounced than the $F_{\max}$ test, the BS test is very conservative and the size steadily drops as T (and N) rises. Again, although less pronounced than the GOS and SW tests, the SD test tends to over-reject the null hypothesis and the degree of the over-rejection becomes more serious as N increases for a given T.

The power of the tests based on the "Power 1" design is reported in Panel B of Table 2. The power of $\hat{J}_\alpha$ test is substantially higher than that of the GRS test. This is in line with our discussion at the end of Section 1, and reflects the fact that GRS assumes an arbitrary degree of cross-sectional error correlations and thus relies on a large time dimension to achieve a reasonably high power. In contrast, the power of the $\hat{J}_\alpha$ test is driven largely by the cross-sectional dimension. The power comparison of the GOS, SW, and SD tests with the $\hat{J}_\alpha$ test seems inappropriate, given their large size-distortions. Having said this, it is perhaps remarkable that the power of the $\hat{J}_\alpha$ test is comparable to the unadjusted power of the GOS, SW_{POET}, and SW_{LW} tests. The power of the $F_{\max}$ and BS tests is uniformly lower than the power of the $\hat{J}_\alpha$ test, likely due to the conservative nature of these tests. The power of the tests based on the "Power 2" design is reported in Panel C of Table 2. The properties of the tests with the "Power 2" design reported in Panel C of Table 2 are qualitatively very similar to those of the "Power 1" design. A detailed discussion of Table 2 is therefore omitted.

We now consider the case in which the errors are non-normal. The size results are summarized in Table 3. The results show that the size of the $\hat{J}_\alpha$ test and the GRS test, as well as the $F_{\max}$, BS, and SD tests, is hardly affected by non-normality. The over-rejection of the GOS and SW tests tends to be somewhat magnified by non-normality.

Furthermore, the behavior of the test statistics is examined under the same DGP as that examined in Table 2, except that a spatial autoregressive component was incorporated into the error generation process. The results with such mixed factor-spatial errors are reported in Table 4. As can be seen, the size of the $\hat{J}_\alpha$ test and GRS test is well controlled, with a slight over-rejection for $T = 60$, which disappears when T is increased to 120. In contrast, the size distortion of GOS and SW seems to be amplified with this design. The size properties of the $F_{\max}$, BS, and SD tests remain similar to those in Table 2.

Table 2. Size and power of the $\hat{J}_\alpha$ and other tests with normal errors

Panel A: Size ($\alpha_i = 0$ for all i)

	(T, N)	$\delta_\gamma = 0$			$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$		
		50	100	200	50	100	200	50	100	200
$\hat{J}_\alpha$										
	60	6.4	5.6	4.7	6.1	6.1	6.1	5.5	6.8	5.9
	120	6.5	5.6	4.7	5.9	5.9	5.3	5.8	6.1	6.1
	240	4.9	5.8	5.2	5.7	5.8	4.7	6.0	6.2	6.4
GRS										
	60	5.0	–	–	4.1	–	–	5.3	–	–
	120	5.8	4.3	–	4.9	4.3	–	4.9	3.7	–
	240	4.3	4.9	4.5	4.8	5.4	4.9	5.9	4.6	5.1
GOS										
	60	17.4	23.5	30.3	17.3	22.5	31.5	16.9	23.8	29.9
	120	11.3	12.3	13.9	9.8	12.2	14.4	9.6	11.7	14.7
	240	7.2	8.9	9.3	7.4	8.4	8.6	7.7	8.4	9.6
SW										
	60	17.4	23.5	30.3	17.4	22.6	31.5	17.8	24.3	30.2
	120	11.3	12.3	13.9	10.0	12.2	14.4	22.9	19.6	16.0
	240	7.2	8.9	9.3	7.4	8.7	8.6	10.8	14.3	20.9
$F_{\max}$										
	60	0.4	0.2	0.1	0.1	0.0	0.2	0.4	0.2	0.0
	120	0.2	0.1	0.1	0.1	0.2	0.0	0.1	0.1	0.0
	240	0.1	0.2	0.2	0.1	0.1	0.2	0.1	0.1	0.1
BS										
	60	4.2	4.0	4.6	3.4	4.4	3.9	3.9	4.4	4.3
	120	3.4	2.9	2.7	2.7	2.9	2.4	2.9	3.5	3.5
	240	2.0	2.4	2.0	2.6	2.5	2.0	3.2	2.9	3.0
SD										
	60	10.9	12.0	13.2	10.2	12.1	13.5	9.3	11.2	11.9
	120	7.9	7.7	8.3	7.1	7.9	8.5	6.4	8.1	8.6
	240	5.0	6.7	6.7	5.7	6.3	5.8	5.9	6.7	7.3

Panel B: Power 1 ($\alpha_i = \varpi_i \sim N(0,1)$ for $i = 1, \ldots, \lfloor N^{0.7} \rfloor$ and $\alpha_i = 0$ for other i)

	(T, N)	$\delta_\gamma = 0$			$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$		
		50	100	200	50	100	200	50	100	200
$\hat{J}_\alpha$										
	60	70.3	81.7	90.8	64.6	78.1	86.9	53.4	66.0	77.0
	120	93.6	98.5	99.7	91.7	98.2	99.8	84.7	95.5	98.6
	240	99.5	99.9	100.0	99.4	100.0	100.0	98.8	99.9	100.0
GRS										
	60	14.7	–	–	13.4	–	–	14.5	–	–
	120	82.8	48.9	–	80.1	49.3	–	79.6	48.5	–
	240	99.0	99.8	95.5	99.0	99.8	95.6	99.0	99.7	95.4

(continued)

Table 2. (continued)

Panel B: Power 1 ($\alpha_i = \varpi_i \sim N(0,1)$ for $i = 1, \ldots, \lfloor N^{0.7} \rfloor$ and $\alpha_i = 0$ for other i)

	(T, N)	$\delta_\gamma = 0$			$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$		
		50	100	200	50	100	200	50	100	200
GOS										
	60	83.1	93.0	98.6	80.3	91.7	97.9	71.7	86.0	96.0
	120	95.1	99.2	99.9	94.5	99.1	100.0	89.2	97.6	99.5
	240	99.6	100.0	100.0	99.4	100.0	100.0	99.1	99.9	100.0
SW										
	60	83.1	93.0	98.6	80.4	91.7	97.9	72.7	86.5	96.1
	120	95.1	99.2	99.9	94.5	99.1	100.0	94.6	98.6	99.7
	240	99.6	100.0	100.0	99.4	100.0	100.0	99.6	100.0	100.0
$F_{\max}$										
	60	17.6	20.3	25.3	16.0	18.8	20.5	11.2	16.1	16.5
	120	53.2	65.8	76.0	50.0	63.6	72.7	38.2	50.3	65.0
	240	87.9	95.7	99.2	87.0	94.8	98.8	77.8	90.4	96.6
BS										
	60	39.8	49.4	63.1	38.0	49.4	58.8	28.9	39.7	48.9
	120	73.2	86.2	95.0	71.0	85.7	94.1	63.2	79.7	90.1
	240	96.3	99.4	100.0	95.5	99.6	100.0	92.8	98.6	99.9
SD										
	60	76.7	87.9	95.6	72.7	85.5	93.5	60.9	75.4	87.5
	120	94.4	98.8	99.8	93.0	98.7	99.9	86.3	96.4	99.1
	240	99.5	99.9	100.0	99.4	100.0	100.0	98.7	99.9	100.0

Panel C: Power 2 ($\alpha_i = \boldsymbol{\beta}_i'(\boldsymbol{\lambda} - \boldsymbol{\mu})$ with $(\boldsymbol{\lambda} - \boldsymbol{\mu}) = 0.1(2.92, -0.63, -9.96)'$)

	(T, N)	$\delta_\gamma = 0$			$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$		
		50	100	200	50	100	200	50	100	200
$\hat{J}_\alpha$										
	60	58.4	81.5	96.3	56.2	79.4	96.5	49.0	75.6	94.9
	120	94.4	99.7	100.0	93.0	99.6	100.0	90.0	99.4	100.0
	240	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0
GRS										
	60	11.8	–	–	12.3	–	–	12.4	–	–
	120	78.1	47.7	–	75.5	46.5	–	76.9	45.1	–
	240	99.9	100.0	99.3	99.8	100.0	99.0	99.8	100.0	99.1
GOS										
	60	76.2	94.8	99.7	75.0	94.0	99.9	72.0	93.2	99.7
	120	96.5	100.0	100.0	96.1	99.7	100.0	94.0	99.9	100.0
	240	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SW										
	60	77.4	93.8	99.9	78.1	93.7	99.8	75.6	92.5	99.8
	120	97.1	99.8	100.0	95.7	100.0	100.0	95.7	99.7	100.0
	240	100.0	100.0	100.0	99.9	100.0	100.0	99.9	100.0	100.0

(continued)

Table 2. (continued)

Panel C: Power 2 ($\alpha_i = \boldsymbol{\beta}_i'(\boldsymbol{\lambda} - \boldsymbol{\mu})$ with $(\boldsymbol{\lambda} - \boldsymbol{\mu}) = 0.1(2.92, -0.63, -9.96)'$)										
		$\delta_\gamma = 0$			$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$		
	(T, N)	50	100	200	50	100	200	50	100	200
$F_{\max}$										
	60	1.6	1.9	1.7	1.5	1.5	1.5	1.3	1.4	1.5
	120	7.6	9.0	10.3	6.6	7.6	9.1	7.5	7.7	8.9
	240	35.2	43.7	55.4	31.4	44.5	56.7	29.3	42.3	54.9
BS										
	60	25.8	44.6	70.9	23.4	42.1	69.4	18.7	33.8	57.7
	120	60.6	88.0	99.0	57.8	85.4	99.4	47.5	77.7	97.6
	240	96.3	100.0	100.0	95.2	100.0	100.0	91.9	99.6	100.0
SD										
	60	67.6	89.0	99.0	65.9	87.4	98.9	59.4	83.8	97.9
	120	95.1	99.8	100.0	94.5	99.7	100.0	90.8	99.8	100.0
	240	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0

Notes: This table summarizes the size and power of $\hat{J}_\alpha$, GRS, GOS, SW, $F_{\max}$, BS, and SD tests of $\alpha_i = 0$ for $i = 1, 2, \ldots, N$, in the case of three-factor models. The observations are generated as $y_{it} = \alpha_i + \sum_{\ell=1}^{3} \beta_{\ell i} f_{\ell t} + u_{it},\ i = 1, 2, .., N; t = 1, 2, \ldots, T,\ f_{\ell t} = \mu_{f\ell} + \rho_{f\ell} f_{\ell,t-1} + e_{\ell t}$, where $e_{\ell t} = \sqrt{h_{\ell t}}\,\xi_{\ell t},\ h_{\ell t} = \mu_{h\ell} + \rho_{1h\ell} h_{\ell,t-1} + \rho_{2h\ell} e_{\ell,t-1}^2,\ \xi_{\ell t} \sim \mathrm{IIDN}(0,1),\ t = -49, \ldots, T$ with $f_{\ell,-50} = 0$ and $h_{\ell,-50} = 0,\ \ell = 1, 2, 3$. The idiosyncratic errors are generated as $u_{it} = \gamma_i v_t + \sigma_{\eta i} \varepsilon_{\eta,it}$, where $\varepsilon_{\eta,it} \sim \mathrm{IIDN}(0,1)$, $v_t \sim \mathrm{IIDN}(0,1)$ and $\sigma_{\eta i}^2 \sim \mathrm{IID}(1 + \chi_{2,i}^2)/3$. The first $\lfloor N^{\delta_\gamma} \rfloor (< N)$ γ_i are generated as Uniform$(0.7, 0.9)$, and the remaining elements are set to 0. We consider the values $\delta_\gamma = 0, 1/4$, and 1/2. $\hat{J}_\alpha$ is the proposed test; GRS is the *F*-test due to Gibbons et al. (1989) which is distributed as $F_{N,T-N-m}$, which is applicable when $T > N + 4$. "–" signifies that the GRS statistic cannot be computed. GOS is the test proposed by Gagliardini et al. (2016) defined in Equation (47); SW is the test based on the POET estimator of Fan et al. (2013). $F_{\max}$ is proposed by GL, BS and SD are tests of He et al. (2021), which are defined in the Supplementary Material. Values of $\hat{J}_\alpha$, GOS, SW, BS, and SD are compared with a positive one-sided critical value of the standard normal distribution. All tests are conducted at the 5% significance level. Experiments are based on 2000 replications.

Since the autoregressive conditional heteroskedasticity is commonly found in security returns, the effect of cross-sectionally correlated errors with GARCH(1,1) processes is also investigated. The size properties of the tests are summarized in Table 5. The results are almost identical to those using unconditionally time-series homoskedastic (but cross-sectionally heteroskedastic) errors reported in Table 2. This is to be expected as the LFPM is a static model and unconditional homoskedastic GARCH errors do not affect our theoretical results.

The experimental results so far confirm that the finite sample performance of the $\hat{J}_\alpha$ test is superior to the other tests we have considered. In the light of these promising results, we further investigate the properties of *J*-alpha tests, in particular the sensitivity of the choice of the values for $\{\delta, p\}$ and the effectiveness of the standardization employed by the $\hat{J}_\alpha$.

First, we examine the sensitivity of the test to the choice of the value of $\{\delta, p\}$. As mentioned, the $\hat{J}_\alpha$ we have considered employs $\delta = 1$ and $p = 0.1$. To check whether this choice is appropriate, in the next experiment, we consider four combinations of $\{\delta, p\}$ using $\delta = 1, 2,\ p = 0.05, 0.01$. Table 6 summarizes the size and power results. As can be seen, the choice of p has little effect on the size and power characteristics. Meanwhile, the

Table 3. Size of the $\hat{J}_\alpha$ and other tests with non-normal errors

	Size: $\alpha_i = 0$ for all i								
	$\delta_\gamma = 0$			$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$		
(T, N)	50	100	200	50	100	200	50	100	200
$\hat{J}_\alpha$									
60	5.9	4.6	5.6	5.0	6.2	5.0	5.5	6.6	7.0
120	5.7	4.8	5.2	4.3	6.2	6.0	5.8	5.7	5.1
240	5.8	5.7	5.4	4.7	5.6	5.4	6.5	6.8	5.8
GRS									
60	5.0	–	–	4.5	–	–	5.4	–	–
120	4.9	5.1	–	4.8	4.7	–	3.6	5.1	–
240	5.5	4.7	4.2	3.7	5.0	4.7	5.4	5.6	5.0
GOS									
60	17.1	22.2	30.0	15.5	21.7	29.2	17.0	22.9	32.6
120	9.5	10.8	14.0	9.5	11.9	14.3	8.9	12.4	14.4
240	8.1	8.3	8.9	6.6	7.9	9.0	8.1	9.2	9.1
SW									
60	17.1	22.1	30.1	15.5	21.7	29.2	18.5	23.5	32.8
120	9.5	10.8	14.0	9.5	11.8	14.4	19.7	19.9	15.5
240	8.1	8.3	8.9	6.6	8.0	9.0	11.1	17.7	24.6
$F_{\max}$									
60	0.0	0.2	0.1	0.2	0.1	0.1	0.1	0.2	0.2
120	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.1	0.1
240	0.2	0.1	0.2	0.2	0.2	0.1	0.1	0.3	0.1
BS									
60	3.9	3.6	4.6	2.9	4.4	3.5	3.5	4.5	4.7
120	3.2	2.0	3.3	2.5	3.2	2.1	2.9	2.5	3.4
240	2.2	1.8	2.2	2.1	2.6	2.1	3.0	2.6	3.0
SD									
60	10.8	11.3	13.0	9.4	12.2	12.7	9.6	12.1	13.3
120	6.7	6.3	8.5	5.7	8.3	8.7	6.6	7.4	7.8
240	5.9	6.1	6.4	4.8	6.0	6.6	6.3	7.1	6.8

Notes: See the note to Table 2. The DGP is the same as in Table 2, except that $u_{it} = \gamma_i v_t + \sigma_{\eta i}\varepsilon_{\eta,it}$, where $\varepsilon_{\eta,it}$ is independently drawn from standardized student t-distribution with eight degrees of freedom.

performance of the test is slightly sensitive to the choice of δ, but this effect quickly disappears as T increases. These experimental results support the use of the $\hat{J}_\alpha$ test with $\delta = 1$ and $p = 0.1$.

Finally, an experiment was conducted to check the effectiveness of the standardization employed in the $\hat{J}_\alpha$. In particular, we check the effectiveness of the centering $t_i^2 - v/(v-2)$ employed by the $\hat{J}_\alpha$ test compared with $t_i^2 - 1$ employed by GOS, and the usefulness of estimating the cross-correlation of t_i^2 with the MT estimator $\tilde{\rho}_N$, respectively. For this purpose, two J-alpha test variants, $\tilde{J}_\alpha$ and $J_\alpha(0)$, are considered on top of the $\hat{J}_\alpha$ statistic. $\tilde{J}_\alpha$ is identical to $\hat{J}_\alpha$, but replaces $t_i^2 - v/(v-2)$ by $t_i^2 - 1$. The second statistic, $J_\alpha(0)$, sets $\tilde{\rho}_N$ equal to

Table 4. Size of the $\hat{J}_\alpha$ and other tests, spatially correlated errors

	Size: $\alpha_i = 0$ for all i									
		$\delta_\gamma = 0$			$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$		
	(T, N)	50	100	200	50	100	200	50	100	200
$\hat{J}_\alpha$										
	60	7.3	7.1	7.8	5.8	7.0	6.1	6.7	6.8	6.4
	120	6.1	6.5	6.1	6.0	5.2	5.7	6.5	6.2	6.6
	240	6.5	6.1	5.6	5.8	4.9	5.9	6.9	7.0	5.9
GRS										
	60	4.4	–	–	4.1	–	–	4.9	–	–
	120	5.5	5.4	–	4.4	5.2	–	5.4	5.5	–
	240	5.7	5.0	4.3	5.0	5.0	5.3	5.6	4.5	4.1
GOS										
	60	17.4	23.9	32.3	17.7	24.0	31.1	19.3	24.5	30.9
	120	11.4	13.8	16.5	11.0	12.6	15.2	10.9	11.5	16.9
	240	8.9	10.2	9.8	8.6	8.6	10.8	8.5	9.8	9.4
SW										
	60	17.5	23.9	32.2	17.8	24.1	31.2	20.5	25.5	31.0
	120	11.9	13.8	16.5	12.6	13.0	15.4	44.8	15.7	18.9
	240	17.7	12.8	11.3	15.8	14.3	12.9	20.3	44.9	26.5
$F_{\max}$										
	60	0.2	0.2	0.0	0.3	0.1	0.1	0.3	0.1	0.1
	120	0.1	0.1	0.1	0.2	0.1	0.0	0.0	0.2	0.1
	240	0.1	0.0	0.1	0.1	0.0	0.2	0.1	0.1	0.2
BS										
	60	4.0	4.2	3.8	3.8	3.6	3.5	4.0	4.4	3.6
	120	3.1	3.2	3.4	2.8	3.0	2.6	3.0	3.2	3.6
	240	2.7	3.0	2.4	2.9	2.4	2.4	3.0	3.4	2.5
SD										
	60	9.8	12.0	13.4	9.4	11.3	12.3	9.5	10.6	11.6
	120	6.8	7.7	7.9	6.4	6.9	7.7	7.4	7.0	8.0
	240	6.4	6.7	6.4	5.6	5.2	6.8	6.4	7.1	6.3

Notes: See the note to Table 2. The DGP is the same as in Table 2, except that $u_{it} = \gamma_i v_t + \eta_{it}$ with $\eta_{it} = \psi \sum_{j=1}^{N} w_{ij}\eta_{jt} + \sigma_{\eta i}\varepsilon_{\eta,it}$. We have chosen the value $\psi = 1/4$ and a rook form for $\mathbf{W} = (w_{ij})$, namely, all elements in $\mathbf{W}$ are zero except $w_{i+1,i} = w_{j-1,j} = 0.5$ for $i = 1, 2, \ldots, N-2$ and $j = 3, 4\ldots, N$, with $w_{1,2} = w_{N,N-1} = 1$.

zero (i.e., does not control for cross-correlation). In the present experiment, to investigate the behavior of the $\hat{J}_\alpha$ test in more challenging environments, N is considered with larger values, that is, $N = 500$, 1000, 2000 and 5000, while T is set to 60, 120, and 240 as before. The results are reported in Table 7, which reveal that the centering using $v/(v-2)$ as well as the control of error cross-correlations by the MT estimator play a very significant role in controlling the size of the test for large N (and large T as shown in Panel A of Table 2).

Table 5. Size of the $\hat{J}_\alpha$ and other tests, GARCH(1,1) errors

	Size: $\alpha_i = 0$ for all i								
	$\delta_\gamma = 0$			$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$		
(T, N)	50	100	200	50	100	200	50	100	200
$\hat{J}_\alpha$									
60	5.6	4.9	5.6	5.9	5.6	5.0	6.0	6.1	5.5
120	5.2	5.8	5.9	5.8	5.0	4.6	5.4	5.6	5.3
240	6.1	4.9	6.0	5.8	5.6	4.9	5.6	6.8	4.8
GRS									
60	3.9	–	–	4.8	–	–	4.6	–	–
120	3.7	4.8	–	5.3	5.6	–	4.9	4.9	–
240	4.5	5.0	5.8	4.8	5.4	5.5	5.0	5.4	5.3
GOS									
60	15.3	21.5	29.9	17.9	20.6	32.5	18.5	22.7	29.8
120	9.5	11.9	14.0	10.1	10.5	13.7	10.5	12.4	14.9
240	8.2	7.3	9.3	8.2	8.9	8.9	7.8	10.1	9.8
SW									
60	16.1	22.5	29.6	16.1	22.1	29.4	19.0	23.5	31.5
120	9.8	11.2	15.1	9.7	11.4	15.1	21.2	23.3	16.8
240	7.7	8.8	8.1	7.8	8.7	8.5	11.1	16.9	27.4
$F_{\max}$									
60	0.1	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.1
120	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.0	0.1
240	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1
BS									
60	4.0	4.0	3.8	3.7	3.3	4.3	4.0	3.8	4.0
120	2.9	3.3	3.9	2.8	2.8	2.9	3.0	3.1	2.3
240	2.6	1.6	2.0	2.7	2.6	2.3	2.7	2.6	2.4
SD									
60	8.7	10.8	12.5	9.9	11.2	13.4	10.3	10.8	11.3
120	6.6	8.2	9.1	7.2	7.2	7.7	6.4	7.4	7.3
240	6.3	5.5	6.8	5.9	6.6	6.6	5.6	7.3	6.1

Notes: See the note to Table 2. The DGP is the same as in Table 2, except that $u_{it} = \gamma_i v_t + \varepsilon_{\eta,it}$ with $\varepsilon_{\eta,it} = \sqrt{\omega_{it}}\zeta_{it}$ and $\zeta_{it} \sim \text{IIDN}(0,1)$, where $\omega_{it} = \sigma^2_{\eta i}(1 - \varrho - \varphi) + \varrho\omega_{i,t-1} + \varphi\varepsilon^2_{\eta,it-1}$. We set $\varrho = 0.2$ and $\varphi = 0.6$. First 50 time-series observations of $\varepsilon_{\eta,it}$ are discarded.

6 Empirical Application

6.1 Data Description

We consider the application of our proposed $\hat{J}_\alpha$ test to the securities in the S&P 500 index of large cap U.S. equities market. Since the index is primarily intended as a leading indicator of U.S. equities, the composition of the index is monitored by S&P to ensure the widest possible overall market representation while reducing the index turnover to a minimum. Changes to the composition of the index are governed by published guidelines.

Table 6. Size and power of the $\hat{J}_\alpha$ tests for $p = \{0.1, 0.05\}$ and $\delta = \{1, 2\}$ with normal errors

	$\delta_\gamma = 0$			$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$		
(T, N)	50	100	200	50	100	200	50	100	200
Size ($\alpha_i = 0$ for all i)									
$\hat{J}_\alpha(p = 0.1, \delta = 1)$									
60	6.4	5.6	4.7	6.1	6.1	6.1	5.5	6.8	5.9
120	6.5	5.6	4.7	5.9	5.9	5.3	5.8	6.1	6.1
240	4.9	5.8	5.2	5.7	5.8	4.7	6.0	6.2	6.4
$\hat{J}_\alpha(p = 0.1, \delta = 2)$									
60	6.6	5.7	5.0	6.2	6.1	6.2	6.0	7.6	6.8
120	6.6	5.6	4.7	6.0	5.9	5.3	6.0	6.5	6.5
240	5.0	5.9	5.3	5.7	5.8	4.8	6.0	6.3	6.4
$\hat{J}_\alpha(p = 0.05, \delta = 1)$									
60	6.4	5.6	4.8	6.1	6.1	6.1	5.6	6.9	5.9
120	6.5	5.6	4.7	5.9	5.9	5.3	5.9	6.2	6.2
240	4.9	5.9	5.3	5.7	5.8	4.8	6.0	6.2	6.4
$\hat{J}_\alpha(p = 0.05, \delta = 2)$									
60	6.6	5.7	5.0	6.2	6.1	6.2	6.1	7.6	6.9
120	6.6	5.6	4.7	6.0	5.9	5.3	6.0	6.6	6.5
240	5.0	5.9	5.3	5.7	5.8	4.8	6.0	6.3	6.4
Power 1 ($\alpha_i = \varpi_i \sim N(0,1)$ for $i = 1, \ldots, \lfloor N^{0.7} \rfloor$ and $\alpha_i = 0$ for other i)									
$\hat{J}_\alpha(p = 0.1, \delta = 1)$									
60	70.3	81.7	90.8	64.6	78.1	86.9	53.4	66.0	77.0
120	93.6	98.5	99.7	91.7	98.2	99.8	84.7	95.5	98.6
240	99.5	99.9	100.0	99.4	100.0	100.0	98.8	99.9	100.0
$\hat{J}_\alpha(p = 0.1, \delta = 2)$									
60	70.7	82.0	91.0	64.9	78.4	87.1	55.0	67.9	78.7
120	93.6	98.5	99.7	91.9	98.3	99.8	84.9	95.5	98.6
240	99.5	99.9	100.0	99.4	100.0	100.0	98.8	99.9	100.0
$\hat{J}_\alpha(p = 0.05, \delta = 1)$									
60	70.5	81.9	90.9	64.8	78.3	87.0	53.8	66.2	77.7
120	93.6	98.5	99.7	91.8	98.3	99.8	84.7	95.5	98.6
240	99.5	99.9	100.0	99.4	100.0	100.0	98.8	99.9	100.0
$\hat{J}_\alpha(p = 0.05, \delta = 2)$									
60	70.7	82.0	91.0	65.0	78.4	87.1	55.2	68.0	78.8
120	93.6	98.5	99.7	91.9	98.3	99.8	85.0	95.5	98.6
240	99.5	99.9	100.0	99.4	100.0	100.0	98.8	99.9	100.0

Notes: See the note to Table 2. The DGP is the same as in Table 2. The p and δ are for the MT estimator $\tilde{\rho}_{ij} = \hat{\rho}_{ij} I[|\sqrt{v}\hat{\rho}_{ij}| > c_p(N)]$, where $c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2N^\delta}\right)$.

In particular, a security is included if its market capitalization currently exceeds US$5.3 billion, is financially viable, and at least 50% of their equity is publicly floated. Companies that substantially violate one or more of the criteria for index inclusion, or are involved in merger, acquisition, or significant restructuring are replaced by other companies.

Table 7. Size of the $\hat{J}_\alpha$ tests, for very large N with normal and non-normal errors

	$\delta_\gamma = 0$				$\delta_\gamma = 1/4$				$\delta_\gamma = 1/2$			
(T, N)	500	1000	2000	5000	500	1000	2000	5000	500	1000	2000	5000
Panel A: Normal errors												
$\tilde{J}_\alpha$												
60	14.5	19.4	29.4	52.4	13.0	19.3	29.5	53.3	14.3	18.7	28.2	51.5
120	8.6	9.2	12.5	21.7	8.9	8.9	11.7	21.6	8.7	9.1	11.1	19.1
240	6.6	7.4	7.1	11.3	6.9	7.1	7.7	11.7	6.7	7.1	7.0	10.8
$J_\alpha(0)$												
60	6.9	5.3	4.3	5.2	5.5	5.7	5.2	5.1	7.5	7.4	6.9	7.8
120	5.1	4.4	4.9	5.0	5.7	4.5	5.3	4.6	7.1	6.1	5.8	7.2
240	5.0	5.0	4.2	5.2	5.1	5.1	4.1	5.0	6.9	6.6	6.0	7.1
$\hat{J}_\alpha$												
60	6.8	5.3	4.2	5.1	5.5	5.6	5.1	5.1	6.5	6.3	6.1	7.2
120	5.1	4.2	4.8	5.0	5.6	4.4	5.2	4.5	5.6	4.5	4.4	5.8
240	5.0	5.0	4.1	5.1	5.0	5.1	4.1	5.0	5.7	5.2	4.3	5.6
Panel B: Non-normal errors												
$\tilde{J}_\alpha$												
60	13.7	18.5	28.1	52.0	13.1	17.7	28.6	51.3	12.6	18.5	25.7	49.7
120	9.0	10.1	12.2	21.2	9.4	9.5	12.4	21.7	8.7	9.6	11.7	19.9
240	6.3	7.3	7.9	12.2	6.7	7.4	7.5	12.2	7.7	7.7	8.1	10.0
$J_\alpha(0)$												
60	5.6	5.0	4.1	4.1	4.9	4.6	4.0	4.6	7.3	5.9	6.1	5.8
120	5.7	5.4	4.8	4.7	5.3	4.8	5.1	4.9	7.7	6.2	5.7	6.0
240	5.2	5.4	4.7	5.4	5.3	4.8	4.5	5.3	7.7	7.2	6.0	6.5
$\hat{J}_\alpha$												
60	5.5	5.0	4.0	4.0	4.9	4.5	4.0	4.6	6.4	5.3	5.4	5.4
120	5.6	5.4	4.6	4.7	5.2	4.7	5.0	4.9	6.4	4.7	4.5	4.9
240	5.2	5.4	4.6	5.4	5.1	4.8	4.4	5.2	6.2	5.7	4.7	5.0

Notes: See the note to Table 2. The DGPs are the same as in Table 2 for normal errors and in Table 3 for non-normal errors. For the purpose of comparison to $\hat{J}_\alpha$, we also provide results for $\tilde{J}_\alpha$ test, which controls for error cross-correlations as the $\hat{J}_\alpha$ test but demean t_i^2 by 1 rather than $v/(v-2)$. The $J_\alpha(0)$ test is defined by Equation (61) with $\rho_N^2 = 0$, which does not control for error cross-correlations.

In order to take account for the change to the composition of the index over time, we compiled returns on all the 500 securities that constitute the S&P 500 index each month over the period January 1984 to April 2018. The monthly return of security i for month t is computed as $r_{it} = 100(P_{it} - P_{i,t-1})/P_{i,t-1} + \mathrm{DY}_{it}/12$, where P_{it} is the end of the month price of the security and DY_{it} is the percent per annum dividend yield on the security. Note that index i depends on the month in which the security i is a constituent of S&P 500, τ, say, which is suppressed for notational simplicity.

The time-series data on the safe rate of return, and the market factors are obtained from Ken French's data library web page. The one-month U.S. treasury bill rate is chosen as the risk-free rate (r_{ft}), the value-weighted return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) is used as a proxy for the market return (r_{mt}), the average return on the three small portfolios minus the average return on the three big portfolios (SMB_t), the average return on two value portfolios minus the average return on two growth portfolios (HML_t), the difference between the returns on diversified portfolios of stocks with robust and weak profitability (RMW_t), and the difference between the returns on diversified portfolios of the stocks of low and high investment firms (CMA_t). SMB, HML, RMW, and CMA are based on the stocks listed on the NYSE, AMEX, and NASDAQ. All data are measured in percent per month. See Section M1.3 in the Supplementary Material for further details.

6.2 Month End Test Results (September 1989–April 2018)

Encouraged by the satisfactory performance of the $\hat{J}_\alpha$ test, even in cases where N is much larger than T, we apply the $\hat{J}_\alpha$ test that allows for non-Gaussian and cross-correlated errors to all securities in the S&P 500 index at the end of each month spanning the period September 1989–April 2018.[22] In this way, we minimize the possibility of survivorship bias since the sample of securities considered at the end of each month is decided in real time. As far as the choice of T is concerned, to reduce the impact of possible time variations in betas, we select a relatively short time period of $T = 60$ months. Accordingly, we estimated the CAPM, Fama and French (1993) three factor (FF3), and Fama and French (2015) five factor (FF5) regressions. The estimated FF5 regression is

$$\begin{aligned} r_{i,\tau t} - r_{f,\tau t} &= \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}(r_{m,\tau t} - r_{f,\tau t}) + \hat{\beta}_{2,i\tau}\text{SMB}_{\tau t} + \hat{\beta}_{3i\tau}\text{HML}_{\tau t} \\ &\quad + \hat{\beta}_{4,i\tau}\text{RMW}_{\tau t} + \hat{\beta}_{5i\tau}\text{CMA}_{\tau t} + \hat{u}_{i,\tau t}, \end{aligned} \tag{74}$$

for $t = 1, 2, \ldots, 60$, $i = 1, 2, \ldots, N_\tau$, and the month ends, τ, from September 1989 to April 2018. The CAPM regression includes the first factor and the FF3 regression uses the first three factors in Equation (74) as regressors, respectively. All securities in the S&P 500 index are included except those with less than 60 months of observations and/or with five consecutive zeros in the middle of sample periods. See the Supplementary Material for discussions on the statistical properties of the regression residuals.

Table 8 reports the rejection frequencies of the $\hat{J}_\alpha$ and GOS tests based on the CAPM, FF3, and FF5 models over the month ends, for the full sample periods, and three market disruption periods: (1) the Asian financial crisis (1997M07–1998M12), (2) the Dot-com bubble burst (2000M03–2002M10), and (3) the Great Recession (2007M12–2009M06) periods. Depending on the factor model (CAPM, FF3, or FF5) and nominal size (5% or 1%) considered, the $\hat{J}_\alpha$ test rejects the null hypothesis $H_0 : \alpha_i = 0$, from 24% to 30% of the total number of tests carried out, which is much smaller than the rejection rates of the GOS test that lie between 39% and 72%. The high rejection rates and their wide range in the GOS test may be due to the tendency of this test to over-reject when T is relatively small, as documented by Monte Carlo experiments in Section 5.

As to be expected, rejection rates in the top panel of Table 8 (based on 5% level) are larger than those in the bottom panel (based on 1% level), but the differences are of second-order importance, particularly when compared with the choice of the underlying asset pricing models. Focusing on the test results based on the 5% level, we note wide

22 In all the empirical applications $T < N$ and the GRS test cannot be computed. We have also decided to exclude other tests discussed in the Monte Carlo Section on the grounds of their substantial size distortion of the null and/or low power.

Table 8. Empirical application: rejection frequencies of the $\hat{J}_\alpha$ and GOS tests

Test	$\hat{J}_\alpha$ test			GOS test		
Factor models	CAPM	FF3	FF5	CAPM	FF3	FF5
Significance level of 0.05						
Full sample period (1989M09–2018M04)	0.28	0.27	0.30	0.42	0.57	0.72
Three market disruption periods:						
(1) Asian financial crisis (1997M07–1998M12)	0.06	0.22	0.39	0.33	0.83	1.00
(2) The Dot-com Bubble Burst (2000M03–2002M10)	0.00	0.50	0.66	0.09	0.72	1.00
(3) The Great Recession (2007M12–2009M06)	0.84	0.95	0.74	1.00	1.00	0.95
Significance level of 0.01						
Full sample period (1989M09–2018M04)	0.24	0.27	0.24	0.39	0.49	0.62
Three market disruption periods:						
(1) Asian financial crisis (1997M07–1998M12)	0.00	0.11	0.28	0.28	0.83	0.67
(2) The Dot-com Bubble Burst (2000M03–2002M10)	0.00	0.25	0.56	0.03	0.59	1.00
(3) The Great Recession (2007M12–2009M06)	0.79	0.84	0.68	0.95	1.00	0.89

Notes: This table provides rejection frequencies of the $\hat{J}_\alpha$ and GOS tests with the significance levels of 0.05 and 0.01, applied to CAPM, FF3, and FF5 regressions of securities in the S&P 500 index using rolling $T = 60$ monthly estimation windows over the month ends during the full sample period and during the three market disruption periods.

variations in the test outcomes across models (CAPM, FF3, and FF5) particularly in the case of sub-samples representing the Asian Financial Crisis and the Dot-com Bubble. The test outcomes for these two sub-samples critically depend on the choice of the asset pricing model, although as for the full sample results the GOS test gives much larger rejection rates. Given the sensitivity of the test outcomes to the choice of the asset pricing model, no firm conclusions can be made in relation to these financial crises. The results based on the $\hat{J}_\alpha$ only lead to substantial rejections only in the case of Dot-com Bubble period and when we base the test on the FF5 model.

The situation is very different when we consider the Great Recession period, where we find substantial rejection of the null of market efficiency irrespective of the model choice. Using the $\hat{J}_\alpha$ there is no pattern to the rejection rates across the models, and using CAPM given a rejection rate of 84% when compared with 95% for FF3 and 74% for FF5. The GOS rejection rates are much higher (100% for CAPM and FF3 and 95% for FF5). Due to its over-rejection tendency, the GOS test seems to be less discriminatory when we compare the GOS rejection rates across the different sample periods. This is particularly so in the case of the GOS tests based on the FF5 model. Overall, both tests provide strong evidence of pricing errors during the Great Recession, but $\hat{J}_\alpha$ test appears to provide more sensible results than the GOS test in this application.

7 Conclusion

In this article, we propose a simple test of LFPMs, the $\hat{J}_\alpha$ test, when the number of securities, N, is large relative to the time dimension, T, of the return series. It is shown that the $\hat{J}_\alpha$ test is more robust against error cross-sectional correlation than the SW tests based on an adaptive thresholding estimator of $\mathbf{V}$, which is considered by Fan, Liao, and Yao (2015).

It allows N to be much larger than T, when compared with alternative tests proposed in the literature. The proposed test also allows for a wide class of error dependencies including mixed weak-factor spatial autoregressive processes, and is shown to be robust to random time-variations in betas.

Using Monte Carlo experiments, designed specifically to match the distributional features of the residuals of Fama–French three factor regressions of individual securities in the S&P 500 index, we show that the proposed $\hat{J}_\alpha$ test performs well even when N is much larger than T, and outperforms other existing tests such as the tests of GOS et al. (2015) and GL. Also, in cases where $N < T$ and the standard F-test due to GRS can be computed, we still find that the $\hat{J}_\alpha$ test has much higher power, especially when T is relatively small.

Application of the $\hat{J}_\alpha$ test to all securities in the S&P 500 index with 60 months of return data at the end of each month over the period September 1989–April 2018 clearly illustrates the utility of the proposed test. Statistically significant evidence against Sharpe–Lintner CAPM and Fama–French three and five factor models is found mainly during periods of financial crises and market disruptions.

Supplemental Data

Supplemental data are available at https://www.datahostingsite.com.

Appendix: Proofs of the Theorems

In this Appendix, we provide proofs of the theorems set out in Section 4 of the article. These proofs make use of lemmas which are provided, together with their proofs, in the Supplementary Material.

For further clarity and convenience, we summarize some repeatedly used notations below:

$$\mathbf{M}_G = (m_{tt'}) = \mathbf{I}_T - \mathbf{P}_G,\ \mathbf{P}_G = \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}',\ \mathbf{G} = (\boldsymbol{\tau}_T, \mathbf{F}),\ v = \mathrm{Tr}(\mathbf{M}_G) = T - m - 1, \tag{A.1}$$

$$\begin{aligned} &\mathbf{M}_F = (m_{F,tt'}) = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}',\ \mathbf{H}_F = \mathbf{h}\mathbf{h}' = (h_t h_{t'}) \\ &\text{with } \mathbf{h} = (h_t) = \mathbf{M}_F\boldsymbol{\tau}_T,\ w_T = \mathrm{Tr}(\mathbf{H}_F) = \mathbf{h}'\mathbf{h} = \boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T, \end{aligned} \tag{A.2}$$

where $\mathbf{F}$ is a $T \times m$ matrix and $\boldsymbol{\tau}_T = (1, 1, \ldots, 1)'$ is a $T \times 1$ vector of ones. Also, before providing a proof of Theorem 1, we state a theorem due to Kelejian and Prucha (2001, KP) which is used to establish it.

Lemma 1 (Central Limit Theorem for Linear Quadratic Forms): *Consider the following linear quadratic form*

$$Q_N = \boldsymbol{\varepsilon}' A \varepsilon + \mathbf{b}'\boldsymbol{\varepsilon} = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}\varepsilon_i\varepsilon_j + \sum_{i=1}^{N} b_i\varepsilon_i,$$

where $\{\varepsilon_i,\ i = 1, 2, \ldots, N\}$ are real-valued random variables, and a_{ij} and b_i denote real-valued coefficients of the quadratic and linear forms. Suppose the following assumptions hold: Assumption KP1: ε_i, for $i = 1, 2, \ldots, N$, have zero means and are independently distributed across i. Assumption KP2: $\mathbf{A}$ is symmetric and $\sup_i \sum_{j=1}^{N} |a_{ij}| < K$. Also,

$N^{-1}\sum_{i=1}^{N}|b_i|^{2+\varepsilon_0} < K$ *for some* $\varepsilon_0 > 0$. *Assumption KP3:* $\sup_i E|\varepsilon_i|^{4+\varepsilon_0} < K$ *for some* $\varepsilon_0 > 0$. *Then, assuming that* $N^{-1}\mathrm{Var}(Q_N) \geq c$ *for some* $c > 0$,

$$\frac{Q_N - E(Q_N)}{\sqrt{\mathrm{Var}(Q_N)}} \rightarrow_d N(0,1).$$

Proof: See KP (Theorem 1, p. 227). ∎

Proof of Theorem 1: Noting that $\mathbf{H}_F = \mathbf{h}\mathbf{h}'$, where $\mathbf{h} = (h_1, h_2, \ldots, h_T)' = \mathbf{M}_F\boldsymbol{\tau}_T$, we can write

$$z_i^2 = w_T^{-1}\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i$$

with $w_T = \boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T$. Then,

$$\sum_{i=1}^{N} z_i^2 = w_T^{-1}\sum_{i=1}^{N}\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i = w_T^{-1}\left(\sum_{t=1}^{T}\mathbf{u}_t h_t\right)'\mathbf{D}_\sigma^{-1}\left(\sum_{t=1}^{T}\mathbf{u}_t h_t\right),$$

where $\mathbf{D}_\sigma = \mathrm{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{NN})$. Using Equation (54)

$$\begin{aligned} N^{-1/2}\sum_{i=1}^{N} z_i^2 &= w_T^{-1}\sum_{i=1}^{N} N^{-1/2}\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i \\ &= w_T^{-1}\left[N^{-1/2}\sum_{t=1}^{T}(\boldsymbol{\Gamma}\mathbf{v}_t + \boldsymbol{\eta}_t)h_t\right]'\mathbf{D}_\sigma^{-1}\left[\sum_{t=1}^{T}(\boldsymbol{\Gamma}\mathbf{v}_t + \boldsymbol{\eta}_t)h_t\right] \\ &= a_{NT} + 2b_{NT} + c_{NT}, \end{aligned} \tag{A.3}$$

where

$$\begin{aligned} a_{NT} &= w_T^{-1}N^{-1/2}\left(\sum_{t=1}^{T} h_t\mathbf{v}_t'\boldsymbol{\Gamma}'\right)\mathbf{D}_\sigma^{-1}\left(\sum_{t=1}^{T} h_t\boldsymbol{\Gamma}\mathbf{v}_t\right), \\ b_{NT} &= w_T^{-1}N^{-1/2}\left(\sum_{t=1}^{T} h_t\mathbf{v}_t'\boldsymbol{\Gamma}'\right)\mathbf{D}_\sigma^{-1}\left(\sum_{t=1}^{T} h_t\boldsymbol{\eta}_t\right), \text{ and} \\ c_{NT} &= w_T^{-1}N^{-1/2}\left(\sum_{t=1}^{T} h_t\boldsymbol{\eta}_t'\right)\mathbf{D}_\sigma^{-1}\left(\sum_{t=1}^{T} h_t\boldsymbol{\eta}_t\right). \end{aligned} \tag{A.4}$$

Consider the first term, a_{NT}, and note that

$$\begin{aligned} a_{NT} &= w_T^{-1}N^{-1/2}\sum_{t=1}^{T}\sum_{r=1}^{T} h_t h_r\mathbf{v}_t'\boldsymbol{\Gamma}'\mathbf{D}_\sigma^{-1}\boldsymbol{\Gamma}\mathbf{v}_r \\ &= w_T^{-1}N^{-1/2}\sum_{t=1}^{T}\sum_{r=1}^{T} h_t h_r\left(\sum_{i=1}^{N}\tilde{\boldsymbol{\gamma}}_i'\mathbf{v}_t\mathbf{v}_r'\tilde{\boldsymbol{\gamma}}_i\right), \end{aligned}$$

where

$$\tilde{\boldsymbol{\gamma}}_i = \frac{\boldsymbol{\gamma}_i}{\sqrt{\sigma_{ii}}} = \frac{\boldsymbol{\gamma}_i}{\sqrt{\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i + \sigma_{\eta,ii}}}. \tag{A.5}$$

Equivalently, letting $\mathbf{d}_T = w_T^{-1/2}\sum_{1=1}^{T} h_t\mathbf{v}_t$, and noting that for any conformable real symmetric positive semi-definite matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathrm{Tr}(\mathbf{AB}) \leq \mathrm{Tr}(\mathbf{A})\lambda_{\max}(\mathbf{B})$ (this result is repeatedly used below), we have

$$\begin{aligned} a_{NT} = N^{-1/2}\sum_{i=1}^{N}\tilde{\boldsymbol{\gamma}}_i'\left[\left(w_T^{-1/2}\sum_{1=1}^{T} h_t\mathbf{v}_t\right)\left(w_T^{-1/2}\sum_{t=1}^{T} h_t\mathbf{v}_t\right)'\right] = N^{-1/2}\sum_{i=1}^{N}\tilde{\boldsymbol{\gamma}}_i'\mathbf{d}_T\mathbf{d}_T'\tilde{\boldsymbol{\gamma}}_i \\ \leq \left(N^{-1/2}\sum_{i=1}^{N}\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_i\right)\lambda_{\max}(\mathbf{d}_T\mathbf{d}_T') \leq \left(N^{-1/2}\sum_{i=1}^{N}\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_i\right)(\mathbf{d}_T'\mathbf{d}_T). \end{aligned}$$

But since h_t are given constants such that $\sum_{t=1}^{T} h_t^2 = w_T$, and by assumption $\mathbf{v}_t$ is $\text{IID}(0, \mathbf{I}_k)$, it then readily follows that $\mathbf{d}_T'\mathbf{d}_T \rightarrow_p 1$, and hence

$$a_{NT} = O_p\left(N^{-1/2}\sum_{i=1}^{N}\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_i\right).$$

Also, it is clear from Equation (A.5) that $|\tilde{\gamma}_{is}| \leq 1$ and $|\tilde{\gamma}_{is}| \leq |\gamma_{is}|$, and

$$N^{-1/2}\sum_{i=1}^{N}\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_i = N^{-1/2}\sum_{i=1}^{N}\sum_{s=1}^{k}\tilde{\gamma}_{is}^2 \leq N^{-1/2}\sum_{s=1}^{k}\left(\sum_{i=1}^{N}|\tilde{\gamma}_{is}|\right)$$
$$\leq N^{-1/2}\sum_{s=1}^{k}\left(\sum_{i=1}^{N}|\gamma_{is}|\right) \leq N^{-1/2}\sup_s \sum_{i=1}^{N}|\gamma_{is}|,$$

and hence by Assumption 2, $N^{-1/2}\sum_{i=1}^{N}\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_i = O(N^{\delta_\gamma - 1/2})$, and overall $a_{NT} = O_p(N^{\delta_\gamma - 1/2})$. Similarly,

$$\begin{aligned}
b_{NT} &= w_T^{-1}N^{-1/2}\left(\textstyle\sum_{t=1}^{T} h_t\mathbf{v}_t'\boldsymbol{\Gamma}'\right)\mathbf{D}_\sigma^{-1}\left(\textstyle\sum_{t=1}^{T} h_t\boldsymbol{\eta}_t\right)\\
&= w_T^{-1}N^{-1/2}\textstyle\sum_{t=1}^{T}\sum_{r=1}^{T} h_t h_r\mathbf{v}_t'\boldsymbol{\Gamma}'\mathbf{D}_\sigma^{-1}\boldsymbol{\eta}_r\\
&= w_T^{-1}N^{-1/2}\textstyle\sum_{t=1}^{T}\sum_{r=1}^{T} h_t h_r\displaystyle\sum_{i=1}^{N}\left(\frac{\eta_{ir}}{\sigma_{ii}^{1/2}}\right)\tilde{\boldsymbol{\gamma}}_i'\mathbf{v}_t\\
&= N^{-1/2}\left(w_T^{-1/2}\textstyle\sum_{t=1}^{T} h_t\mathbf{v}_t'\right)\left[w_T^{-1/2}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T} h_t\tilde{\boldsymbol{\gamma}}_i\left(\dfrac{\eta_{it}}{\sigma_{ii}^{1/2}}\right)\right]\\
&= N^{-1/2}\left[w_T^{-1/2}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{N} h_t(\mathbf{d}_T'\tilde{\boldsymbol{\gamma}}_i)\left(\dfrac{\eta_{it}}{\sigma_{ii}^{1/2}}\right)\right].
\end{aligned}$$

Since by Assumption, η_{it} and $\mathbf{v}_t$ (and hence $\mathbf{d}_T$) are independently distributed, it follows that $E(b_{NT}) = 0$. Consider now $\text{Var}(b_{NT})$, and note that for given values of $\boldsymbol{\gamma}_i$ we have (recall that η_{it} is independent over t and $\sum_{t=1}^{T} h_t^2 = w_T$)

$$\begin{aligned}
\text{Var}(b_{NT}) &= N^{-1}w_T^{-1}\textstyle\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} h_t h_r\left[\tilde{\boldsymbol{\gamma}}_i'E(\mathbf{d}_T\mathbf{d}_T')\tilde{\boldsymbol{\gamma}}_j\right]E\left(\dfrac{\eta_{it}\eta_{jr}}{\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}}\right)\\
&= N^{-1}w_T^{-1}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} h_t^2\left(\tilde{\boldsymbol{\gamma}}_i'E(\mathbf{d}_T\mathbf{d}_T')\tilde{\boldsymbol{\gamma}}_j\right)\left(\dfrac{\sigma_{\eta,ij}}{\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}}\right)\\
&= N^{-1}\textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\tilde{\boldsymbol{\gamma}}_i'E(\mathbf{d}_T\mathbf{d}_T')\tilde{\boldsymbol{\gamma}}_j\right)\left(\dfrac{\sigma_{\eta,ij}}{\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}}\right).
\end{aligned}$$

Also, $E(\mathbf{d}_T\mathbf{d}_T') = E\left[\left(w_T^{-1/2}\sum_{1=1}^{T} h_t\mathbf{v}_t\right)\left(w_T^{-1/2}\sum_{1=1}^{T} h_t\mathbf{v}_t'\right)\right] = \mathbf{I}_k$ and

$$\text{Var}(b_{NT}) = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_j\right)\left(\frac{\sigma_{\eta,ij}}{\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}}\right).$$

Further,

$$\left|\frac{\sigma_{\eta,ij}}{\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}}\right| = \frac{|\sigma_{\eta,ij}|}{\sqrt{\left(\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i + \sigma_{\eta,ii}\right)\left(\boldsymbol{\gamma}_j'\boldsymbol{\gamma}_j + \sigma_{\eta,jj}\right)}} = \frac{|\rho_{\eta,ij}|}{\sqrt{\left(\frac{\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i}{\sigma_{\eta,ii}} + 1\right)\left(\frac{\boldsymbol{\gamma}_j'\boldsymbol{\gamma}_j}{\sigma_{\eta,jj}} + 1\right)}} \leq |\rho_{\eta,ij}|.$$

Therefore (recalling that $\sup_{j,s}|\tilde{\gamma}_{js}| < K$ and $|\tilde{\gamma}_{is}| \leq |\gamma_{is}|$),

$$\begin{aligned}\mathrm{Var}(b_{NT}) &\leq N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_j||\rho_{\eta,ij}| \leq N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{k}|\tilde{\gamma}_{is}||\tilde{\gamma}_{js}||\rho_{\eta,ij}| \\ &\leq \sup_{j,s}|\tilde{\gamma}_{js}|\left[N^{-1}\sum_{s=1}^{k}\sum_{i=1}^{N}|\tilde{\gamma}_{is}|\left(\sum_{j=1}^{N}|\rho_{\eta,ij}|\right)\right] \\ &\leq KN^{-1}\sum_{s=1}^{k}\sum_{i=1}^{N}|\gamma_{is}|\left(\sum_{j=1}^{N}|\rho_{\eta,ij}|\right).\end{aligned}$$

But by Condition (57) in Assumption 3 and $\sigma_{\eta,ii} > c > 0$ imply $\sup_j \sum_{i=1}^{N}|\rho_{\eta,ij}| < K$ (also see Equation (58)) and by Equation (53), we have $\sup_s \sum_{i=1}^{N}|\gamma_{is}| = O(N^{\delta_\gamma})$. Then, it follows that $\mathrm{Var}(b_{NT}) = O(N^{\delta_\gamma - 1})$ and $b_{NT} = O(N^{\delta_\gamma/2-1/2})$. Therefore, b_{NT} is dominated by a_{NT} and using these results in Equation (A.3) we have

$$N^{-1/2}\sum_{i=1}^{N}z_i^2 = w_T^{-1}N^{-1/2}\left(\sum_{t=1}^{T}h_t\boldsymbol{\eta}_t'\right)\mathbf{D}_\sigma^{-1}\left(\sum_{t=1}^{T}h_t\boldsymbol{\eta}_t\right) + O_p(N^{\delta_\gamma-1/2}). \tag{A.6}$$

Now using Equation (56), we can express the above as

$$N^{-1/2}\sum_{i=1}^{N}z_i^2 = w_T^{-1}N^{-1/2}\left(\sum_{t=1}^{T}h_t\boldsymbol{\varepsilon}_{\eta,t}'\mathbf{Q}_\eta'\right)\mathbf{D}_\sigma^{-1}\left(\sum_{t=1}^{T}h_t\mathbf{Q}_\eta\boldsymbol{\varepsilon}_{\eta,t}\right) + O_p(N^{\delta_\gamma-1/2}).$$

where $\boldsymbol{\varepsilon}_{\eta,t} \sim \mathrm{IID}(0, \mathbf{I}_N)$. After some re-arrangement of the terms we now obtain

$$N^{-1/2}\sum_{i=1}^{N}\left(z_i^2 - 1\right) = N^{-1/2}w_T^{-1}\left(\sum_{t=1}^{T}h_t\boldsymbol{\varepsilon}_{\eta,t}'\right)\left(\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1}\mathbf{Q}_\eta\right)\left(\sum_{t=1}^{T}h_t\boldsymbol{\varepsilon}_{\eta,t}\right) + O_p(N^{\delta_\gamma-1/2})$$

$$q_{NT} = N^{-1/2}\left[\mathbf{x}_T'\mathbf{A}\mathbf{x}_T - \mathrm{Tr}(\mathbf{A})\right] + N^{-1/2}[\mathrm{Tr}(\mathbf{A}) - N] + O_p(N^{\delta_\gamma-1/2}), \tag{A.7}$$

where

$$\mathbf{x}_T = w_T^{-1/2}\sum_{t=1}^{T}h_t\boldsymbol{\varepsilon}_{\eta,t} \text{ and } \mathbf{A} = \mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1}\mathbf{Q}_\eta. \tag{A.8}$$

First consider the deterministic component of q_{NT}, and using Equation (55) and under Assumption 3, we have

$$\mathbf{R} = \tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Gamma}}' + \mathbf{D}_\sigma^{-1/2}\mathbf{Q}_\eta\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1/2}, \tag{A.9}$$

where $\tilde{\boldsymbol{\Gamma}} = (\tilde{\boldsymbol{\gamma}}_1, \tilde{\boldsymbol{\gamma}}_2, \ldots, \tilde{\boldsymbol{\gamma}}_N)'$. Then,

$$\mathrm{Tr}(\mathbf{R}) = N = \sum_{i=1}^{N}\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_i + \mathrm{Tr}(\mathbf{A}).$$

But, as before,

$$\begin{aligned}\mathrm{Tr}\big(\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Gamma}}'\big) &= \textstyle\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_i = \sum_{i=1}^{N}\sum_{s=1}^{k} \tilde{\gamma}_{is}^2 \\ &\leq \textstyle\sum_{s=1}^{k}\sum_{i=1}^{N} |\gamma_{is}| \leq k \sup_s \sum_{i=1}^{N} |\gamma_{is}| = O(N^{\delta_\gamma}).\end{aligned} \tag{A.10}$$

Hence,

$$N^{-1/2}[\mathrm{Tr}(\mathbf{A}) - N] = O\big(N^{\delta_\gamma - 1/2}\big),$$

and Equation (A.7) can be written as

$$q_{NT} = z_{NT} + O\big(N^{\delta_\gamma-1/2}\big) + O_p\big(N^{\delta_\gamma-1/2}\big), \tag{A.11}$$

where

$$z_{NT} = N^{-1/2}\mathbf{x}_T'\tilde{\mathbf{A}}\mathbf{x}_T, \text{ with } \tilde{\mathbf{A}} = \mathbf{A} - N^{-1}\mathrm{Tr}(\mathbf{A})\mathbf{I}_N. \tag{A.12}$$

We now apply the central limit theorem for linear quadratic forms due to KP to z_{NT}, which is reproduced for convenience as Lemma 1. We first establish the conditions required by KP's theorem (see Lemma 1). To this end, we first note that $E(\mathbf{x}_T) = 0$, and

$$\begin{aligned}\mathrm{Var}(\mathbf{x}_T) &= w_T^{-1}E\left[\left(\textstyle\sum_{t=1}^{T} h_t\boldsymbol{\varepsilon}_{\eta,t}\right)\left(\sum_{t=1}^{T} h_t\boldsymbol{\varepsilon}_{\eta,t}\right)'\right] \\ &= w_T^{-1}\textstyle\sum_{t=1}^{T} h_t^2 E\big(\boldsymbol{\varepsilon}_{\eta,t}\boldsymbol{\varepsilon}_{\eta,t}'\big) = \mathbf{I}_N.\end{aligned}$$

Denote the ith element of $\mathbf{x}_T$ by $x_{i,T}$ and note that it is given by $x_{i,T} = w_T^{-1/2}\sum_{t=1}^{T} h_t\varepsilon_{\eta,it} = w_T^{-1/2}\mathbf{h}'\boldsymbol{\varepsilon}_{\eta,i}$, where $\boldsymbol{\varepsilon}_{\eta,i} = (\varepsilon_{\eta,i1}\ \varepsilon_{\eta,i2},\dots,\varepsilon_{\eta,iT})'$, with an abuse of the notation. Then, $x_{i,T} = w_T^{-1/2}\boldsymbol{\varepsilon}_{\eta,i}'\mathbf{M}_F\boldsymbol{\tau}_T$ and $x_{i,T}^2 = w_T^{-1}\boldsymbol{\varepsilon}_{\eta,i}'\mathbf{H}_F\boldsymbol{\varepsilon}_{\eta,i}$; hence, for a given T, the elements of $\mathbf{x}_T$ have zero means, a unit variance, and are independently distributed as required by KP's theorem. Using results on the moments of quadratic forms, it is also easily established that $E(x_{i,T}^6) = w_T^{-3}E\big(\boldsymbol{\varepsilon}_{\eta,i}'\mathbf{H}_F\boldsymbol{\varepsilon}_{\eta,i}\big)^3 = 15 + O(v^{-1}) \leq K$ uniformly over i (see Lemma 11), and hence condition KP1 of the KP theorem (Lemma 1) is met. Consider now matrix $\tilde{\mathbf{A}}$ defined by Equation (A.12) and note that it is symmetric and we have

$$\|\tilde{\mathbf{A}}\|_\infty \leq \|\mathbf{A} - N^{-1}\mathrm{Tr}(\mathbf{A})\mathbf{I}_N\|_\infty \leq \|\mathbf{A}\|_\infty + N^{-1}\mathrm{Tr}(\mathbf{A})$$

and using Equation (A.8)

$$\begin{aligned}\|\tilde{\mathbf{A}}\|_\infty &\leq \|\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1}\mathbf{Q}_\eta\|_\infty + N^{-1}\mathrm{Tr}\Big(\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1}\mathbf{Q}_\eta\Big) \\ &\leq \left(\frac{1}{\min_i(\sigma_{ii})}\right)\|\mathbf{Q}_\eta\|_1\|\mathbf{Q}_\eta\|_\infty + N^{-1}\mathrm{Tr}\Big(\mathbf{Q}_\eta'\mathbf{Q}_\eta\Big)\lambda_{\max}\Big(\mathbf{D}_\sigma^{-1}\Big) \\ &\leq \left(\frac{1}{\min_i(\sigma_{ii})}\right)\Big[\|\mathbf{Q}_\eta\|_1\|\mathbf{Q}_\eta\|_\infty + N^{-1}\mathrm{Tr}\Big(\mathbf{Q}_\eta'\mathbf{Q}_\eta\Big)\Big].\end{aligned}$$

But under Condition (57) and noting that $\sigma_{ii} > c > 0$, then

$$\|\tilde{\mathbf{A}}\|_\infty = \sup_i \textstyle\sum_{j=1}^{N} |\tilde{a}_{ij}| < K,$$

and condition KP2 of Lemma 1 is met. To establish condition KP3, we note that

$$\mathrm{Tr}(\tilde{\mathbf{A}}) = 0,\ \mathrm{Tr}(\tilde{\mathbf{A}}^2) = \mathrm{Tr}(\mathbf{A}^2) - N^{-1}[\mathrm{Tr}(\mathbf{A})]^2.$$

Using Equation (A.9), let $\mathbf{B} = \mathbf{D}_\sigma^{-1/2}\mathbf{Q}_\eta\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1/2}$, and note that

$$\mathrm{Tr}(\mathbf{R}^2) = \mathrm{Tr}(\mathbf{B}^2) + \mathrm{Tr}\left[\left(\tilde{\mathbf{\Gamma}}'\tilde{\mathbf{\Gamma}}\right)^2\right] + 2\mathrm{Tr}\left(\tilde{\mathbf{\Gamma}}'B\tilde{\mathbf{\Gamma}}\right). \tag{A.13}$$

Also,

$$\mathrm{Tr}\left(\tilde{\mathbf{\Gamma}}'B\tilde{\mathbf{\Gamma}}\right) \leq \mathrm{Tr}\left(\tilde{\mathbf{\Gamma}}'\tilde{\mathbf{\Gamma}}\right)\lambda_{\max}(\mathbf{B}),$$

and in view of Equation (57), we have

$$\lambda_{\max}(\mathbf{B}) = \lambda_{\max}\left(\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1}\mathbf{Q}_\eta\right) \leq \|\left(\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1}\mathbf{Q}_\eta\right)\|_1 \leq \left(\frac{1}{\min_i(\sigma_{ii})}\right)\|\mathbf{Q}_\eta\|_1\|\mathbf{Q}_\eta\|_\infty < K,$$

and hence (using Equation (A.10)):

$$\mathrm{Tr}\left(\tilde{\mathbf{\Gamma}}'B\tilde{\mathbf{\Gamma}}\right) = O(N^{\delta_\gamma}). \tag{A.14}$$

Also (recalling that $|\tilde{\gamma}_{is}| \leq |\gamma_{is}|$),

$$\begin{aligned}
\mathrm{Tr}\left(\tilde{\mathbf{\Gamma}}'\tilde{\mathbf{\Gamma}}\right)^2 &= \mathrm{Tr}\left(\textstyle\sum_{i=1}^N \tilde{\boldsymbol{\gamma}}_i\tilde{\boldsymbol{\gamma}}_i'\right)^2 = \textstyle\sum_{i=1}^N\sum_{j=1}^N \mathrm{Tr}\left(\tilde{\boldsymbol{\gamma}}_i\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_j\tilde{\boldsymbol{\gamma}}_j'\right)\\
&= \textstyle\sum_{i=1}^N\sum_{j=1}^N \left(\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_j\right)^2 = \sum_{s=1}^k\sum_{s'=1}^k\sum_{i=1}^N\sum_{j=1}^N |\tilde{\gamma}_{is}\tilde{\gamma}_{js}\tilde{\gamma}_{is'}\tilde{\gamma}_{js'}|\\
&\leq \textstyle\sum_{s=1}^k\sum_{s'=1}^k\sum_{i=1}^N\sum_{j=1}^N |\gamma_{is}||\gamma_{js}||\gamma_{is'}||\gamma_{js'}|\\
&\leq k^2\left(\sup_i \textstyle\sum_{i=1}^N |\gamma_{is}|\right)^2 = O(N^{2\delta_\gamma}).
\end{aligned} \tag{A.15}$$

Hence, using Equations (A.14) and (A.15) in Equation (A.13), we have

$$\mathrm{Tr}(\mathbf{B}^2) = \mathrm{Tr}(\mathbf{R}^2) + O(N^{2\delta_\gamma}).$$

Also, in view of Equation (A.8)

$$\mathrm{Tr}(\mathbf{B}^2) = \mathrm{Tr}\left[\mathbf{D}_\sigma^{-1/2}\mathbf{Q}_\eta\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1/2}\mathbf{D}_\sigma^{-1/2}\mathbf{Q}_\eta\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1/2}\right] = \mathrm{Tr}\left[\left(\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1}\mathbf{Q}_\eta\right)^2\right] = \mathrm{Tr}(\mathbf{A}^2).$$

To summarize

$$\mathrm{Tr}\left(\mathbf{A}\right) = \sqrt{N} + O(N^{\delta_\gamma}) \text{ and } \mathrm{Tr}(\mathbf{A}^2) = \mathrm{Tr}\left(\mathbf{R}^2\right) + O(N^{2\delta_\gamma}),$$

which also yield (recall that $\delta_\gamma < 1/2$)

$$\begin{aligned}
&\mathrm{Tr}\left(\tilde{\mathbf{A}}^2\right) = \mathrm{Tr}(\mathbf{A}^2) - N^{-1}[\mathrm{Tr}(\mathbf{A})]^2\\
&= \mathrm{Tr}(\mathbf{R}^2) + O(N^{2\delta_\gamma}) - N^{-1}\left[\sqrt{N} + O(N^{\delta_\gamma})\right]^2\\
&= \mathrm{Tr}(\mathbf{R}^2) + O(N^{2\delta_\gamma}) + O(N^{2\delta_\gamma-1}) - 1.
\end{aligned}$$

Therefore,

$$N^{-1}\mathrm{Tr}\left(\tilde{\mathbf{A}}^2\right) = N^{-1}\mathrm{Tr}(\mathbf{R}^2) + O(N^{2\delta_\gamma-1}), \tag{A.16}$$

which is bounded in N under the assumptions that $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ is bounded in N and $0 \leq \delta_\gamma < 1/2$. Furthermore, it is readily seen that

$$N^{-1}\mathrm{Tr}(\mathbf{R}^2) = N^{-1}\sum_{i=1}^{N}\sum_{i=1}^{N}\rho_{ij}^2 = 1 + (N-1)\rho_N^2.$$

Finally, using Equation (A.12)

$$\mathrm{Var}(z_{NT}) = N^{-1}\mathrm{Var}\left(\mathbf{x}_T'\tilde{\mathbf{A}}\mathbf{x}_T\right) = N^{-1}E\left[\left(\mathbf{x}_T'\tilde{\mathbf{A}}\mathbf{x}_T\right)^2\right].$$

Consider

$$\begin{aligned}\left(\mathbf{x}_T'\tilde{\mathbf{A}}\mathbf{x}_T\right)^2 &= w_T^{-2}\left(\sum_{t=1}^{T}\sum_{t'=1}^{T}h_t h_{t'}\boldsymbol{\varepsilon}_{\eta,t}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t'}\right)^2 \\ &= w_T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{r=1}^{T}\sum_{r'=1}^{T}h_t h_{t'} h_r h_{r'}\left(\boldsymbol{\varepsilon}_{\eta,t}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t'}\right)\left(\boldsymbol{\varepsilon}_{\eta,r}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,r'}\right).\end{aligned}$$

Since, by assumption, $\boldsymbol{\varepsilon}_{\eta,t}$ are serially independent, then using the results on moments of the quadratic forms, we have

$$\begin{aligned}E\left[\left(\boldsymbol{\varepsilon}_{\eta,t}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t}\right)^2\right] &= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{i'=1}^{N}\sum_{j'=1}^{N}\tilde{a}_{ij}\tilde{a}_{i'j'}E(\varepsilon_{\eta,it}\varepsilon_{\eta,jt}\varepsilon_{\eta,i't}\varepsilon_{\eta,j't}) \\ &= \gamma_{2,\varepsilon_\eta}\sum_{i=1}^{N}\tilde{a}_{ii}^2 + \left(\sum_{i=1}^{N}\tilde{a}_{ii}\right)^2 + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{a}_{ij}\tilde{a}_{ji},\end{aligned}$$

where $\gamma_{2,\varepsilon_\eta} = E(\varepsilon_{\eta,it}^4) - 3$ and by assumption $|\gamma_{2,\varepsilon_\eta}| < K$. Also,

$$E\left[\left(\boldsymbol{\varepsilon}_{\eta,t}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t}\right)\left(\boldsymbol{\varepsilon}_{\eta,r}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,r}\right)\right] = [\mathrm{Tr}(\tilde{\mathbf{A}})]^2 \text{ for } t \neq r.$$

For $r = t \neq t' = r'$,

$$\begin{aligned}E\left[\left(\boldsymbol{\varepsilon}_{\eta,t}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t'}\right)\left(\boldsymbol{\varepsilon}_{\eta,t}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t'}\right)\right] &= E\left[\left(\boldsymbol{\varepsilon}_{\eta,t'}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t}\right)\left(\boldsymbol{\varepsilon}_{\eta,t}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t'}\right)\right] \\ &= E\left(\boldsymbol{\varepsilon}_{\eta,t'}'\tilde{\mathbf{A}}\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_\eta, t'\right) = \mathrm{Tr}(\tilde{\mathbf{A}}^2).\end{aligned}$$

Similarly, for $r' = t \neq t' = r$, we have $E\left[\left(\boldsymbol{\varepsilon}_{\eta,t}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t'}\right)\left(\boldsymbol{\varepsilon}_{\eta,t'}'\tilde{\mathbf{A}}\boldsymbol{\varepsilon}_{\eta,t}\right)\right] = \mathrm{Tr}(\tilde{\mathbf{A}}^2)$. Using these results

$$\begin{aligned}w_T^2 E\left[\left(\mathbf{x}_T'\tilde{\mathbf{A}}\mathbf{x}_T\right)^2\right] &= \left(\sum_{t=1}^{T}h_t^4\right)\left[\gamma_{2,\varepsilon_\eta}\sum_{i=1}^{N}\tilde{a}_{ii}^2 + \left(\sum_{i=1}^{N}\tilde{a}_{ii}\right)^2 + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{a}_{ij}\tilde{a}_{ji}\right] \\ &+ \left[\sum_{t=1}^{T}\sum_{r=1}^{T}h_t^2 h_r^2 - \left(\sum_{t=1}^{T}h_t^4\right)\right][\mathrm{Tr}(\tilde{\mathbf{A}})]^2 + 2\left[\sum_{t=1}^{T}\sum_{r=1}^{T}h_t^2 h_r^2 - \left(\sum_{t=1}^{T}h_t^4\right)\right]\mathrm{Tr}(\tilde{\mathbf{A}}^2).\end{aligned}$$

But $\left(\sum_{t=1}^{T}\sum_{r=1}^{T}h_t^2h_r^2\right)=\left(\sum_{t=1}^{T}h_t^2\right)^2$, $\sum_{i=1}^{N}\tilde{a}_{ii}=\mathrm{Tr}(\tilde{\mathbf{A}})=0$, $\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{a}_{ij}\tilde{a}_{ji}=\mathrm{Tr}(\tilde{\mathbf{A}}^2)$, and we have

$$\begin{aligned}\mathrm{Var}(z_{NT})&=N^{-1}E\left[\left(\mathbf{x}_T'\tilde{\mathbf{A}}\mathbf{x}_T\right)^2\right]\\&=\gamma_{2,\varepsilon_\eta}w_T^{-2}\left(N^{-1}\sum_{i=1}^{N}\tilde{a}_{ii}^2\right)\left(\sum_{t=1}^{T}h_t^4\right)+2w_T^{-2}\left(\sum_{t=1}^{T}h_t^2\right)^2N^{-1}\mathrm{Tr}(\tilde{\mathbf{A}}^2),\end{aligned}$$

and, further noting that $\sum_{t=1}^{T}h_t^2=w_T$, then

$$\mathrm{Var}(z_{NT})=2N^{-1}\mathrm{Tr}(\tilde{\mathbf{A}}^2)+\frac{\gamma_{2,\varepsilon_\eta}\left(\sum_{t=1}^{T}h_t^4\right)}{w_T^2}\left(N^{-1}\sum_{i=1}^{N}\tilde{a}_{ii}^2\right),$$

and using Equation (A.16)

$$\mathrm{Var}(z_{NT})=2N^{-1}\mathrm{Tr}(\mathbf{R}^2)+\frac{\gamma_{2,\varepsilon_\eta}\left(\sum_{t=1}^{T}h_t^4\right)}{w_T^2}\left(N^{-1}\sum_{i=1}^{N}\tilde{a}_{ii}^2\right)+O(N^{2\delta_\gamma-1}),$$

where by assumption $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ is bounded in N. Also, using Equation (S.15) in Lemma 8, $\sum_{t=1}^{T}h_t^4=O(T)$, and

$$\begin{aligned}&\frac{|\gamma_{2,\varepsilon_\eta}|\left(\sum_{t=1}^{T}h_t^4\right)}{w_T^2}\left(N^{-1}\sum_{i=1}^{N}\tilde{a}_{ii}^2\right)\le K\frac{\left(\sum_{t=1}^{T}h_t^4\right)}{w_T^2}\left(N^{-1}\mathrm{Tr}(\tilde{\mathbf{A}}^2)\right)\\&\le\frac{K}{T}\left[N^{-1}\mathrm{Tr}(\mathbf{R}^2)\right]+O(T^{-1}N^{2\delta_\gamma-1})=O(T^{-1})+O(T^{-1}N^{2\delta_\gamma-1}).\end{aligned}$$

Therefore,

$$\mathrm{Var}(z_{NT})=2N^{-1}\mathrm{Tr}(\mathbf{R}^2)+O(T^{-1})+O(N^{2\delta_\gamma-1}),\tag{A.17}$$

which is bounded for any N and T, so long as $N^{-1}\mathrm{Tr}(\mathbf{R}^2)$ is bounded in N and $0\le\delta_\gamma<1/2$. Also, using Equation (A.11), and under the same conditions, and as N and $T\to\infty$, in any order,

$$\lim_{N,T\to\infty}\mathrm{Var}(q_{NT})=2\omega^2>0,$$

as required. This result also ensures that condition KP3 of Lemma 1 is satisfied and therefore, we also have $q_{NT}\to_d N(0,2\omega^2)$, as N and $T\to\infty$, in any order. ■

Proof of Theorem 2: We have

$$S_{NT}=N^{-1/2}\sum_{i=1}^{N}\left[z_i^2\left(1-\frac{1}{\sigma_{ii}^{-1}\hat{\sigma}_{ii}}\right)\right],\tag{A.18}$$

where $z_i^2=\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i/w_T$, with $\boldsymbol{\xi}_i=\mathbf{u}_{i.}/\sigma_{ii}^{1/2}$ being the standardized error of the return equation (6) and $w_T=\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T$ and $\hat{\sigma}_{ii}=\hat{\mathbf{u}}_{i.}'\hat{\mathbf{u}}_{i.}/T$. Write $X_i=\sigma_{ii}^{-1}\tilde{\sigma}_{ii}$ and note that by assumption $\sigma_{ii} > 0$, and by construction only securities with $\hat{\sigma}_{ii} > c > 0$ are included in the $\hat{J}_\alpha$ test, so that

$$S_{NT} = N^{-1/2} \sum_{i=1}^{N} \left[z_i^2 \left(1 - \frac{1}{X_i} \right) \right], \tag{A.19}$$

where $X_i = \xi_i' \mathbf{M}_G \xi_i / v$, with $v = T - m - 1$ and $\mathbf{M}_G = (m_{tt'})$, defined by Equation (A.1). Also, by Equation (37), $E(t_i^2) = E(z_i^2 / X_i) = v/(v-2) + O(T^{-3/2})$ for each i, and by Lemma 11, $E(z_i^2) = E(\xi_i' \mathbf{H}_F \xi_i / w_T) = w_T^{-1} \mathrm{Tr}(\mathbf{H}_F) = 1$, for all i. Thus, we have

$$E(S_{NT}) = O\left(\sqrt{N/T^2} \right). \tag{A.20}$$

Next, for all $i = 1, 2, \ldots, N$, we have $X_i > 0$, and Equation (A.19) can be written as

$$\begin{aligned} S_{NT} &= N^{-1/2} \sum_{i=1}^{N} z_i^2 \left[(1 - X_i) + \frac{(1 - X_i)^2}{X_i} \right] \\ &= S_{1,NT} + S_{2,NT}, \end{aligned}$$

where

$$S_{1,NT} = N^{-1/2} \sum_{i=1}^{N} z_i^2 (1 - X_i), \tag{A.21}$$

and

$$S_{2,NT} = N^{-1/2} \sum_{i=1}^{N} \frac{z_i^2 (1 - X_i)^2}{X_i}. \tag{A.22}$$

But since $X_i > c > 0$ and $z_i^2 (1 - X_i)^2 \geq 0$, then

$$|S_{2,NT}| \leq c^{-1} N^{-1/2} \sum_{i=1}^{N} z_i^2 (1 - X_i)^2$$

and

$$E|S_{2,NT}| \leq c^{-1} N^{1/2} \sup_i E\left[z_i^2 (1 - X_i)^2 \right]. \tag{A.23}$$

But

$$\begin{aligned} E\left[z_i^2 (1 - X_i)^2 \right] &= E(z_i^2 X_i^2) - 2E(z_i^2 X_i) + E(z_i^2) \\ &= v^{-2} w_T^{-1} E\left[(\xi_i' \mathbf{H}_F \xi_i)(\xi_i' \mathbf{M}_G \xi_i)^2 \right] - 2 v^{-1} w_T^{-1} E\left[(\xi_i' \mathbf{H}_F \xi_i)(\xi_i' \mathbf{M}_G \xi_i) \right] + 1. \end{aligned}$$

Now using results from Lemma 11, we have

$$\begin{aligned} E\left[(\xi_i' \mathbf{H}_F \xi_i)(\xi_i' \mathbf{M}_G \xi_i) \right] &= v w_T + O(v), \\ E\left[(\xi_i' \mathbf{H}_F \xi_i)(\xi_i' \mathbf{M}_G \xi_i)^2 \right] &= v^2 w_T + O(v w_T), \end{aligned}$$

which yields

$$E\left[z_i^2(1-X_i)^2\right] = O(T^{-1}),\ \text{uniformly across } i. \tag{A.24}$$

Using this result in Equation (A.23), we obtain

$$E|S_{2,NT}| \le c^{-1}N^{1/2}\sup_i E\left[z_i^2(1-X_i)^2\right] = O\left(\frac{\sqrt{N}}{T}\right),$$

and by Markov inequality we have $S_{2,NT}\to_p 0$, so long as $N/T^2 \to 0$. Therefore, to establish $S_{NT}\to_p 0$, it is sufficient to show that $S_{1,NT}\to_p 0$. By Lemma 17, we have

$$N^{-1/2}\sum_{i=1}^{N} z_i^2(X_i-1) = N^{-1/2}\sum_{i=1}^{N} z_{\eta,i}^2(X_{\eta,i}-1) + O_p(N^{\delta_\gamma-1/2}).$$

where $z_{\eta,i}^2 = \boldsymbol{\eta}_i'\mathbf{H}_F\boldsymbol{\eta}_i/(w_T\sigma_{\eta,ii}) > 0$, $X_{\eta,i} = \boldsymbol{\eta}_i'\mathbf{M}_G\boldsymbol{\eta}_i/(v\sigma_{\eta,ii}) > 0$. Using results on the moments of quadratic forms, by Lemma 15, we have

$$N^{-1/2}\sum_{i=1}^{N} E\left[z_{\eta,i}^2(X_{\eta,i}-1)\right] = \frac{\sum_t h_t^2 m_{tt}}{v w_T}\gamma_{2,\varepsilon_\eta}N^{-1/2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\tilde{q}_{\eta,i\ell}^4,$$

where $\gamma_{2,\varepsilon_\eta} = E(\varepsilon_{\eta,it}^4) - 3$ (and $|\gamma_{2,\varepsilon_\eta}| < K$ by assumption), $\tilde{q}_{\eta,i\ell} = q_{\eta,i\ell}/\sigma_{\eta,ii}^{1/2}$ with $q_{\eta,i\ell}$ being such that $\mathbf{Q}_\eta = (q_{\eta,i\ell})$, $\mathbf{Q}_\eta$ defined by Equation (56). But as $0 \le m_{tt} \le 1$ ($\mathbf{M}_G = (m_{tt'})$) by Lemma 8, $v^{-1}w_T^{-1}\sum_{t=1}^T h_t^2 m_{tt} \le v^{-1}w_T^{-1}\sum_{t=1}^T h_t^2 = v^{-1}$ as $\sum_{t=1}^T h_t^2 = w_T$, and also that $0 \le \sum_{\ell=1}^N \tilde{q}_{\eta,i\ell}^4 \le 1$, as $\sum_{\ell=1}^N \tilde{q}_{\eta,i\ell}^2 = 1$ (since $\sum_{\ell=1}^N q_{\eta,i\ell}^2 = \sigma_{\eta,ii}$), and $|\gamma_{2,\varepsilon_\eta}| \le K$, we have

$$N^{-1/2}\sum_{i=1}^{N} E\left[z_{\eta,i}^2(X_{\eta,i}-1)\right] = O\left(\sqrt{N}/T\right).$$

Furthermore,

$$\begin{aligned}\operatorname{Var}\left[N^{-1/2}\sum_{i=1}^{N} z_{\eta,i}^2(X_{\eta,i}-1)\right] &= \frac{1}{N}\sum_i \operatorname{Var}\left[z_{\eta,i}^2(X_{\eta,i}-1)\right]\\ &+\frac{1}{N}\sum_{i\neq j}\operatorname{Cov}\left[z_{\eta,i}^2(X_{\eta,i}-1), z_{\eta,j}^2(X_{\eta,j}-1)\right].\end{aligned}$$

We first note that

$$\operatorname{Var}\left[z_{\eta,i}^2(X_{\eta,i}-1)\right] = E\left[z_{\eta,i}^4(X_{\eta,i}-1)^2\right] - \left\{E\left[z_{\eta,i}^2(X_{\eta,i}-1)\right]\right\}^2.$$

As has shown above,

$$E\left[z_{\eta,i}^2(X_{\eta,i}-1)\right] = O(T^{-1})$$

uniformly over i. Next, consider

$$E\left[z_{\eta,i}^4(X_{\eta,i}-1)^2\right] = E\left(z_{\eta,i}^4 X_{\eta,i}^2\right) - 2E\left(z_{\eta,i}^4 X_{\eta,i}\right) + E\left(z_{\eta,i}^4\right). \tag{A.25}$$

But, using results on the moments of quadratic forms, by Lemma 11, we have

$$E\left(z_{\eta,i}^4\right) = 3 + O(T^{-1}),\ E\left(z_{\eta,i}^4 X_{\eta,i}\right) = 3 + O(T^{-1}) \text{ and } E\left(z_{\eta,i}^4 X_{\eta,i}^2\right) = 3 + O(T^{-1}), \quad \text{(A.26)}$$

uniformly over i. Substituting Equation (A.26) into Equation (A.25), we have

$$E\left[z_{\eta,i}^4 (X_{\eta,i} - 1)^2\right] = O(T^{-1}),$$

therefore,

$$\mathrm{Var}\left[z_{\eta,i}^2 (X_{\eta,i} - 1)\right] = O(T^{-1})$$

uniformly over i. We conclude that

$$\frac{1}{N}\sum_i \mathrm{Var}\left[z_{\eta,i}^2 (X_{\eta,i} - 1)\right] = O(T^{-1}).$$

Secondly, by Lemma 16,

$$\frac{1}{N}\sum_{i\neq j} \mathrm{Cov}\left[z_{\eta,i}^2 (X_{\eta,i} - 1), z_{\eta,j}^2 (X_{\eta,j} - 1)\right] = O(T^{-1}) + O(N/T^2).$$

In sum, under Assumptions 1–3, $S_{NT} \rightarrow_p 0$, so long as $0 \leq \delta_\gamma < 1/2$, $N/T^2 \rightarrow 0$ as N and $T \rightarrow \infty$, jointly. ■

Proof of Theorem 3: Under Assumptions 1–3, using Theorem 2 we have

$$N^{-1/2}\sum_{i=1}^{N}\left(z_i^2 - t_i^2\right)/[2(1 + (N-1)\rho_N^2)]^{1/2} \rightarrow_p 0,$$

where z_i^2 is defined by Equation (22), so long as $(N-1)\rho_N^2 = O(1)$, $N/T^2 \rightarrow 0$, and $0 \leq \delta_\gamma < 1/2$, as N and $T \rightarrow \infty$, jointly. Under these conditions (by Lemma 4), it follows that $N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{v}{v-2}\right)/[2(1 + (N-1)\rho_N^2)]^{1/2}$ has the same limit distribution as $N^{-1/2}\sum_{i=1}^{N}\left(z_i^2 - 1\right)/[2(1 + (N-1)\rho_N^2)]^{1/2}$, which is shown to be standard normal by Theorem 1, and the desired result now follows, observing that $\lim_{T\rightarrow\infty}\left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4} = 2$. ■

Proof of Theorem 4: Let $\psi_{NT} = \frac{1}{N}\sum_{i,j=1}^{N}\left(\tilde{\rho}_{ij}^2 - \rho_{ij}^2\right)$, and note that

$$\psi_{NT} = \frac{1}{N}\sum\nolimits_{i,j=1}^{N}\left(\tilde{\rho}_{ij} + \rho_{ij}\right)\left(\tilde{\rho}_{ij} - \rho_{ij}\right),$$

and since $|\tilde{\rho}_{ij}| < 1$ and $|\rho_{ij}| < 1$, it also follows that

$$|\psi_{NT}| \leq \frac{2}{N}\sum\nolimits_{i,j=1}^{N}|\tilde{\rho}_{ij} - \rho_{ij}|. \quad \text{(A.27)}$$

Further, letting $I_{ij} = I\left[|\hat{\rho}_{ij}| > v^{-1/2}c_p(N)\right]$, we have

$$\tilde{\rho}_{ij} - \rho_{ij} = \hat{\rho}_{ij}I_{ij} - \rho_{ij} = \left[\hat{\rho}_{ij} - E\left(\hat{\rho}_{ij}\right)\right] \times I_{ij} + \left[E\left(\hat{\rho}_{ij}\right) - \rho_{ij}\right] \times I_{ij} - \rho_{ij}\left(1 - I_{ij}\right),$$

and hence

$$\frac{1}{2}E|\psi_{NT}| \leq \frac{1}{N}\sum\nolimits_{i,j=1}^{N} E\left(|\hat{\rho}_{ij} - E(\hat{\rho}_{ij})| \times I_{ij}\right) + \frac{1}{N}\sum\nolimits_{i,j=1}^{N} |E(\hat{\rho}_{ij}) - \rho_{ij}|E(I_{ij}) + \frac{1}{N}\sum\nolimits_{i,j=1}^{N} |\rho_{ij}|\left[1 - E(I_{ij})\right] = \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3. \tag{A.28}$$

Now using Equation (41), we note that

$$\hat{\rho}_{ij} = \frac{\mathbf{u}_{i.}'\mathbf{M}_G\mathbf{u}_{j.}}{\left(\mathbf{u}_{i.}'\mathbf{M}_G\mathbf{u}_{i.}\right)^{1/2}\left(\mathbf{u}_{j.}'\mathbf{M}_G\mathbf{u}_{j.}\right)^{1/2}},$$

where $\hat{\mathbf{u}}_{i.} = \mathbf{M}_G\mathbf{u}_{i.}$. Also, since $\mathbf{M}_G$ is an $(T \times T)$ idempotent matrix of rank $v = T - m - 1$, there exists an orthogonal $T \times T$ transformation matrix $\mathbf{L}$ $(\mathbf{LL}' = \mathbf{I}_T)$, defined by

$$\mathbf{LM}_G\mathbf{L}' = \begin{pmatrix} \mathbf{I}_v & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{A.29}$$

Hence, setting

$$\boldsymbol{\zeta}_{i.} = \sigma_{ii}^{-1/2}\mathbf{Lu}_{i.}, \tag{A.30}$$

$\hat{\rho}_{ij}$ can be written equivalently in terms of the first v elements of $\boldsymbol{\zeta}_{i.} = (\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{iT})'$ as (see Lemma 19)

$$\hat{\rho}_{ij} = \frac{\sum_{t=1}^{v} \zeta_{it}\zeta_{jt}}{\left(\sum_{t=1}^{v} \zeta_{it}^2\right)^{1/2}\left(\sum_{t=1}^{v} \zeta_{jt}^2\right)^{1/2}},$$

where $\zeta_{it} = \sum_{t'=1}^{T} l_{tt'}\xi_{it'}$ and $l_{tt'}$ is the (t, t') element of $\mathbf{L}$. Also, as shown in Lemma 19, for each i, ζ_{it}'s are independently distributed over t, and

$$E(\zeta_{it}) = 0,\ E\left(\zeta_{it}^2\right) = 1,\ E(\zeta_{it}\zeta_{jt}) = \rho_{ij}.$$

$$\kappa_{ij}(4,0) = E(\zeta_{it}^4) - 3,\ \kappa_{ij}(0,4) = E(\zeta_{it}^4) - 3,$$

$$\kappa_{ij}(3,1) = E(\zeta_{it}^3\zeta_{jt}) - 3\rho_{ij},\ \kappa_{ij}(1,3) = E(\zeta_{it}\zeta_{jt}^3) - 3\rho_{ij},$$

$$\kappa_{ij}(2,2) = E(\zeta_{it}^2\zeta_{jt}^2) - 2\rho_{ij}^2 - 1.$$

Furthermore, by Lemma 19

$$E(\hat{\rho}_{ij}) = \rho_{ij} + \frac{a_{ij}}{v} + O(v^{-2}), \tag{A.31}$$

$$\mathrm{Var}(\hat{\rho}_{ij}) = \frac{b_{ij}}{v} + O(v^{-2}), \tag{A.32}$$

where

$$a_{ij} = -\frac{1}{2}\rho_{ij}(1 - \rho_{ij}^2) + \frac{3}{8}\rho_{ij}\left[\kappa_{ij}(4,0) + \kappa_{ij}(0,4)\right] - \frac{1}{2}\left[\kappa_{ij}(3,1) + \kappa_{ij}(1,3)\right] + \frac{1}{4}\rho_{ij}\kappa_{ij}(2,2),$$

and

$$b_{ij} = (1-\rho_{ij}^2)^2 + \frac{1}{4}\rho_{ij}^2\left[\kappa_{ij}(4,0)+\kappa_{ij}(0,4)\right] - \rho_{ij}\left[\kappa_{ij}(3,1)+\kappa_{ij}(1,3)\right] + \frac{1}{2}\left(2+\rho_{ij}^2\right)\kappa_{ij}(2,2).$$

Hence, using Equation (A.31), $|E(\hat{\rho}_{ij})-\rho_{ij}| \leq \frac{1}{v}|a_{ij}| + O(T^{-2})$, and we have the following bound on the second term of Equation (A.28):

$$\mathcal{A}_2 = \frac{1}{N}\sum_{i,j=1}^{N}|E(\hat{\rho}_{ij})-\rho_{ij}|E(I_{ij}) \leq \frac{1}{vN}\sum_{i,j=1}^{N}|a_{ij}| + O(NT^{-2}).$$

Furthermore, since κ_{ij} are bounded, and by assumption $\sum_{i,j=1}^{N}|\rho_{ij}| = O(N)$, we have

$$\begin{aligned}
&\frac{1}{Nv}\sum_{i,j=1}^{N}|a_{ij}| \\
&\leq \frac{1}{2}\frac{1}{Nv}\sum_{i,j=1}^{N}|\rho_{ij}||1-\rho_{ij}^2| + \frac{3}{8}\frac{1}{Nv}\sum_{i,j=1}^{N}|\rho_{ij}||\kappa_{ij}(4,0)+\kappa_{ij}(0,4)| \\
&+\frac{1}{4}\frac{1}{Nv}\sum_{i,j=1}^{N}|\kappa_{ij}(3,1)+\kappa_{ij}(1,3)| + \frac{1}{2Nv}\sum_{i,j=1}^{N}|\rho_{ij}||\kappa_{ij}(2,2)|.
\end{aligned}$$

But

$$\frac{1}{Nv}\sum_{i,j=1}^{N}|\rho_{ij}||\kappa_{ij}(2,2)| \leq \sup_{ij}|\kappa_{ij}(2,2)|\frac{1}{Nv}\sum_{i,j=1}^{N}|\rho_{ij}| = O(v^{-1}),$$

and hence

$$\frac{1}{Nv}\sum_{i,j=1}^{N}|a_{ij}| \leq \frac{1}{4}\frac{1}{Nv}\sum_{i,j=1}^{N}|\kappa_{ij}(3,1)+\kappa_{ij}(1,3)| + O(v^{-1}). \tag{A.33}$$

Also,

$$\begin{aligned}
&\frac{1}{Nv}\sum_{i,j=1}^{N}|\kappa_{ij}(3,1)+\kappa_{ij}(1,3)| \\
&\leq \frac{1}{Nv}\sum_{i,j=1}^{N}|E(\zeta_{it}^3\zeta_{jt})+E(\zeta_{it}\zeta_{jt}^3)| + \frac{6}{Nv}\sum_{i,j=1}^{N}|\rho_{ij}| \\
&= \frac{1}{Nv}\sum_{i,j=1}^{N}|E(\zeta_{it}^3\zeta_{jt})+E(\zeta_{it}\zeta_{jt}^3)| + O(v^{-1}),
\end{aligned}$$

and as established in Lemma 20 (see (S.80) in the Supplementary Material), we have

$$\frac{1}{Nv}\sum_{i,j=1}^{N}|E(\zeta_{it}^3\zeta_{jt})+E(\zeta_{it}\zeta_{jt}^3)| = O(T^{-1}N^{2\delta_\gamma-1}) + O(T^{-1}),$$

which if used in Equation (A.33) yields

$$\frac{1}{Nv}\sum_{i,j=1}^{N}|a_{ij}| = O(v^{-1}N^{2\delta_\gamma-1}) + O(v^{-1}).$$

Overall, for the second term of Equation (A.28), we have

$$\mathcal{A}_2 = \frac{1}{N}\sum_{i,j=1}^{N}|E(\hat{\rho}_{ij})-\rho_{ij}|E(I_{ij}) = O(T^{-1}N^{2\delta_\gamma-1}) + O(v^{-1}) + O(Nv^{-2}),$$

and since by assumption $\delta_\gamma \leq 1/2$, and $N/T^2 \to 0$, as N and $T \to \infty$, then

$$\mathcal{A}_2 \to 0. \tag{A.34}$$

To deal with the first and the third terms of Equation (A.28), we need to distinguish between values of $|\rho_{ij}|$ that are strictly away from zero, namely those values that satisfy the condition $|\rho_{ij}| > \rho_{\min} > 0$, and those values that are *zero* or very close to zero. Note that for values of $|\rho_{ij}|$ sufficiently close to zero, in the sense that $|\rho_{ij}| \leq \kappa N^{-\phi_\rho}$, for some $\kappa > 0$ and $\phi_\rho > 1$, we have[23]

$$\mathcal{A}_3 \leq \frac{1}{N}\sum\nolimits_{i,j=1}^{N} |\rho_{ij}| \leq \kappa N^{1-\phi_\rho} \to 0, \text{ if } \phi_\rho > 1.$$

Therefore, without loss of generality, we only consider the case where $|\rho_{ij}| > \rho_{\min} > 0$, for all i and j. In this case, we have

$$\mathcal{A}_3 = \frac{1}{N}\sum\nolimits_{i,j=1,|\rho_{ij}|>\rho_{\min}}^{N} |\rho_{ij}| E\left(1 - I_{ij}\right) \leq \frac{1}{N}\sum\nolimits_{i,j=1,|\rho_{ij}|>\rho_{\min}}^{N} E\left(1 - I_{ij}\right). \tag{A.35}$$

Further, since $E\left(1 - I_{ij}\right) = \Pr\left[|\hat{\rho}_{ij}| \leq v^{-1/2} c_p(N)\right]$, then using result (A.7) in Lemma 4 of BPS (2017, supplement) we have (for some small $\epsilon > 0$)

$$\Pr\left[|\hat{\rho}_{ij}| \leq v^{-1/2} c_p(N) | \rho_{ij} \neq 0\right] \leq K e^{\frac{-(1-\epsilon)}{2} \frac{v\left(|\rho_{ij}| - \frac{c_p(N)}{\sqrt{v}}\right)^2}{b_{ij}}} [1 + o(1)].$$

Using this result in Equation (A.35) now yields

$$\mathcal{A}_3 \leq KN e^{\frac{-(1-\epsilon)}{2} \frac{v\left(\rho_{\min} - \frac{c_p(N)}{\sqrt{v}}\right)^2}{b_{\max}}} [1 + o(1)],$$

where $b_{\max} = \sup_{ij} b_{ij} < K$, which can be written equivalently as

$$\mathcal{A}_3 \leq K e^{\frac{-v(1-\epsilon)}{2} \frac{\left[\left(\rho_{\min} - \frac{c_p(N)}{\sqrt{v}}\right)^2 - \frac{2\ln(N)}{v(1-\epsilon)}\right]}{b_{\max}}} [1 + o(1)].$$

Noting that $c_p^2(N)/v$ and $\ln(N)/v$ have the same rate of convergence and both $\to 0$, as N and $T \to \infty$, it then follows that[24]

$$\mathcal{A}_3 \to 0, \text{ forsome } \rho_{\min} > 0. \tag{A.36}$$

Finally, consider the first term of Equation (A.28) and write it as

$$\mathcal{A}_1 = \frac{1}{N}\sum\nolimits_{i,j=1}^{N} E\left[|\hat{\rho}_{ij} - E(\hat{\rho}_{ij})| \times I_{ij}\right] = \frac{1}{N}\sum\nolimits_{i,j=1}^{N} \sqrt{\text{Var}(\hat{\rho}_{ij})} E\left(|z_{ij}| \times I_{ij}\right), \tag{A.37}$$

where $z_{ij} = \left[\hat{\rho}_{ij} - E(\hat{\rho}_{ij})\right] / \sqrt{\text{Var}\left(\hat{\rho}_{ij}\right)}$, and $\text{Var}\left(\hat{\rho}_{ij}\right)$ is given by Equation (A.32). Also, by Cauchy–Schwarz inequality (noting that $E\left(z_{ij}^2\right) = 1$)

23 Note that the sparsity condition given by Equation (65) can be violated if $\phi_\rho < 1$.

24 Note that since by assumption $T = c_d N^d$, with $d > 1/2$, then $\ln(N)/v = (T/(T - m - 1))$ $c_d^{-1} N^{-d} \ln(N) \to 0$, as $N \to \infty$. Recall that m, the number of factors, is fixed as $T \to \infty$.

$$E\left(|z_{ij}| \times I_{ij}\right) = E\left(|z_{ij}| I\left[|\hat{\rho}_{ij}| > v^{-1/2} c_p(N)\right]\right) \le \left[E\left(|z_{ij}|^2\right)\right]^{1/2} \left(E\{I\left[|\hat{\rho}_{ij}| > v^{-1/2} c_p(N)\right]\}\right)^{1/2}$$
$$\le \{\Pr\left[|\hat{\rho}_{ij}| > v^{-1/2} c_p(N)\right]\}^{1/2} \le 1.$$

Using this result and $\text{Var}(\hat{\rho}_{ij})$ from Equation (A.32) in Equation (A.37) and distinguishing between non-zero and near zero values of ρ_{ij}, we have

$$\mathcal{A}_1 = N^{-1} \sum_{i,j=1}^{N} E\left[|\hat{\rho}_{ij} - E(\hat{\rho}_{ij})| \times I_{ij}\right] \le$$
$$N^{-1}\left(\sqrt{\frac{b_{\max}}{v}} + O(v^{-1})\right) \sum_{i,j=1}^{N} \{\Pr\left[|\hat{\rho}_{ij}| > v^{-1/2} c_p(N)\right] \big| |\rho_{ij}| = 0\}^{1/2}$$
$$+ N^{-1}\left(\sqrt{\frac{b_{\max}}{v}} + O(v^{-1})\right) \sum_{i,j=1}^{N} \{\Pr\left[|\hat{\rho}_{ij}| > v^{-1/2} c_p(N) \big| |\rho_{ij}| > \rho_{\min}\right]\}^{1/2}$$
$$= \mathcal{A}_{11} + \mathcal{A}_{12}.$$

Under the sparsity conditions, Equations (32) and (33), the maximum number of non-zero $|\rho_{ij}|$ is given by m_N^2, and we have

$$\mathcal{A}_{12} \le \frac{1}{N}\left[\frac{\sqrt{b_{\max}}}{\sqrt{v}} + O(v^{-1})\right] m_N^2 = O\left(\frac{m_N^2}{N\sqrt{v}}\right), \tag{A.38}$$

where $m_N = O(N^{\delta_\rho})$. Hence, since by assumption $\delta_\rho < 1/2$, then it follows that $\mathcal{A}_{12} \to 0$, as N and $v \to \infty$. For $\mathcal{A}_{11}$, which relates to the near zero values of $|\rho_{ij}|$, making use of result (A.5) in Lemma 4 of BPS (2017, supplement) we have

$$\mathcal{A}_{11} \le K \frac{(N^2 - m_N^2)}{N}\left[\frac{\sqrt{b_{\max}}}{\sqrt{v}} + O(v^{-1})\right] \exp\left(\frac{-(1-\epsilon)}{4} \frac{c_p^2(N)}{\varphi_{\max}}\right)[1 + o(1)],$$

where $\varphi_{\max} = \max_{ij} \varphi_{ij} < K$. Then for $\mathcal{A}_1$ to tend to zero it is sufficient that (note that $N^{-1} m_N^2 \to 0$, since $\delta_\rho < 1/2$)

$$\frac{N}{\sqrt{v}} \exp\left(\frac{-(1-\epsilon)}{4} \frac{c_p^2(N)}{\varphi}\right) \to 0, \text{ as } N \text{ and } v \to \infty. \tag{A.39}$$

To obtain a sufficient condition for Equation (A.39) to hold, set $T = c_d N^d$ and note that (recall that $v = T - m - 1$ and $T/(T - m - 1) < K$, since m is fixed as $T \to \infty$)

$$\frac{N}{\sqrt{v}} \exp\left(\frac{-(1-\epsilon)}{4} \frac{c_p^2(N)}{\varphi}\right) \le \sqrt{\frac{T}{T - m - 1}} \exp\left(\frac{-(1-\epsilon)}{4} \frac{c_p^2(N)}{\varphi} + (1 - d/2)\log(N)\right)$$
$$= \sqrt{\frac{T}{T - m - 1}} \exp\left(-\log(N)\left[\frac{\frac{(1-\epsilon)}{4} \frac{c_p^2(N)}{\varphi} - (1 - d/2)\log(N)}{\log(N)}\right]\right).$$

But by result (b) of Lemma 2 of BPS (2017, supplement), $\lim_{N\to\infty} c_p^2(N)/\log(N) = 2\delta$, and Condition (A.39) is met if $\delta(1-\epsilon)/2\varphi_{\max} - (1 - d/2) > 0$, or equivalently if $\delta > \frac{(2-d)}{(1-\epsilon)} \varphi_{\max}$. Therefore, under this condition, $\mathcal{A}_{11} \to 0$, and together with Equation

(A.38) establishes that $\mathcal{A}_1 \to 0$. Therefore, using this result, Equations (A.34) and (A.36) in Equation (A.28) we have $E|\psi_{NT}| \to 0$, as required, and in turn implies $\psi_{NT} \to_p 0$, by Markov inequality. Finally, using (S.79) in the Supplementary Material established in Lemma 20, and setting $\gamma_i = 0$, for all i, and $\sigma_{\eta,ij} = 0$, for all $i \neq j$, to ensure that $\rho_{ij} = 0$, for all $i \neq j$, we have

$$\varphi_{ij} = E(\zeta_{it}^2 \zeta_{jt}^2 | \rho_{ij} = 0) = \gamma_{2,\varepsilon_\eta} \left(\sum_{r=1}^{T} l_{tr}^4 \right) \left(\sum_{\ell=1}^{N} \sigma_{ii}^{-1} \sigma_{jj}^{-1} q_{\eta,i\ell}^2 q_{\eta,j\ell}^2 \right) + \sigma_{ii}^{-1} \sigma_{jj}^{-1} \sigma_{\eta,ii} \sigma_{\eta,jj}.$$

where l_{tr} is the $(t,\ r)$ element of the $T \times T$ orthonormal matrix $\mathbf{L}$ defined by Equation (A.29), $q_{\eta,i\ell}$ is such that $\mathbf{Q}_\eta = (q_{\eta,i\ell})$, $\mathbf{Q}_\eta$ defined by Equation (56). Also, $|\sigma_{\eta,ii}/\sigma_{ii}| \leq 1$, $\sum_{r=1}^{T} l_{tr}^4 \leq \left(\sum_{r=1}^{T} l_{tr}^2 \right)^2 \leq 1$, $\sum_{\ell=1}^{N} \tilde{q}_{\eta,i\ell}^2 = \sum_{\ell=1}^{N} q_{\eta,i\ell}^2 / \sigma_{\eta,ii} = 1$, and

$$\left(\sum_{\ell=1}^{N} \sigma_{ii}^{-1} \sigma_{jj}^{-1} q_{\eta,i\ell}^2 q_{\eta,j\ell}^2 \right) = \left| \sum_{\ell=1}^{N} \tilde{q}_{\eta,i\ell}^2 \tilde{q}_{\eta,j\ell}^2 \right| \leq \left(\sum_{\ell=1}^{N} \tilde{q}_{\eta,i\ell}^4 \right)^{1/2} \left(\sum_{\ell=1}^{N} \tilde{q}_{\eta,j\ell}^4 \right)^{1/2} \leq 1.$$

Hence, $\sup_{ij} \varphi_{ij} \leq 1 + |\gamma_{2,\varepsilon_\eta}|$, as required. ■

Proof of Theorem 5: By Theorem 3, $J_\alpha(\rho_N^2) \to_d N(0,1)$ so long as $N/T^2 \to 0$, and $0 \leq \delta_\gamma < 1/2$, as $N \to \infty$ and $T \to \infty$, jointly, where $J_\alpha(\rho_N^2)$ and δ_γ are defined by Equations (61) and (53), respectively. Since Theorem 4 ensures that $\hat{J}_\alpha - J_\alpha(\rho_N^2) \to_p 0$, as $(N-1)(\tilde{\rho}_{N,T}^2 - \rho_N^2) \to_p 0$ when $d > 2/3$, as N and $T \to \infty$, and $\delta > \frac{(2-d)}{(1-\epsilon)} \varphi_{\max}$, for some small $\epsilon > 0$, where $\varphi_{\max} \leq 1 + |\gamma_{2,\varepsilon_\eta}|$, under these conditions, $\hat{J}_\alpha$ has the same limit distribution as $J_\alpha(\rho_N^2)$ (by Lemma 4), which establishes the result. ■

Proof of Theorem 6: The steps in the proof are similar to the ones in deriving the limiting distribution of $\hat{J}_\alpha$ under the null hypothesis. First, Lemma 22 provides the proof of the result, under Assumptions 1–3, and under the local alternatives (68), $N^{-1/2} \sum_{i=1}^{N} (z_{i,a}^2 - 1) \to_d N(\phi^2, 2\omega^2)$, as $N \to \infty$ and $T \to \infty$, jointly, where $z_{i,a}^2$ defined by (S.97) in the Supplementary Material, $\omega^2 = 1 + \lim_{N \to \infty} (N-1) \rho_N^2$, ρ_N^2 is defined by Equation (60). Also, by Lemma 23, we have $N^{-1/2} \sum_{i=1}^{N} (z_{i,a}^2 - t_i^2) = o_p(1)$. Finally, $\hat{J}_\alpha - J_\alpha = o_p(1)$, since the consistency result of the MT estimator $\tilde{\rho}_{N,T}^2$ given by Theorem 4 will not be affected by the introduction of local alternatives, as the MT estimator is obtained based on the regression residuals of the alternative model. This completes the proof of Theorem 6. ■

References

Affleck-Graves, J., and B. Mcdonald. 1989. Nonnormalities and Tests of Asset Pricing Theories. *The Journal of Finance* 44: 889–908.

Affleck-Graves, J., and B. Mcdonald. 1990. Multivariate Tests of Asset Pricing: The Comparative Power of Alternative Statistics. *Journal of Financial and Quantitative Analysis* 25: 163–185.

Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd edn. Wiley.

Ang, A., J. Chen, and Y. Xing. 2006. Downside Risk. *The Review of Financial Studies* 19: 1191–1239.

Ang, A., J. Liu, and K. Schwarz. 2020. Using Stocks or Portfolios in Tests of Factor Models. *Journal of Financial and Quantitative Analysis* 55: 709–750.

Bai, Z., and H. Saranadasa. 1996. Effect of High Dimension: By an Example of a Two Sample Problem. *Statistica Sinica* 6: 311–329.

Bailey, N., G. Kapetanios, and M. H. Pesaran. 2021. Measurement of Factor Strength: Theory and Practice. *Journal of Applied Econometrics* 36: 587–613.

Bailey, N., M. H. Pesaran, and L. V. Smith. 2019. A multiple Testing Approach to the Regularisation of Large Sample Correlation Matrices. *Journal of Econometrics* 208: 507–534.

Beaulieu, M.-C., J.-M. Dufour, and L. Khalaf. 2007. Multivariate Tests of Mean–Variance Efficiency with Possibly Non-Gaussian Errors. *Journal of Business & Economic Statistics* 25: 398–410.

Bickel, P. J., and E. Levina. 2008. Regularized Estimation of Large Covariance Matrices. *The Annals of Statistics* 36: 199–227.

Black, F., M. C. Jensen, and M. Scholes. 1972. "*The Capital Asset Pricing Model: Some Empirical Tests*." In M. C. Jensen (Ed.), *Studies in the Theory of Capital Markets*, pp. 79–121. New York: Praeger.

Breusch, T., and A. Pagan. 1980. The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics. *Review of Economic Studies* 47: 239–253.

Cai, T., and W. Liu. 2011. Adaptive Thresholding for Sparse Covariance Matrix Estimation. *Journal of the American Statistical Association* 106: 672–684.

Chamberlain, G. 1983. Funds, Factors, and Diversification in Arbitrage Pricing Models. *Econometrica* 51: 1305–1323.

Chordia, T., A. Goyal, and J. Shanken. 2017. *Cross-Sectional Asset Pricing with Individual Stocks: Betas versus Characteristics*. Columbia Business School.

Cremers, M., M. Halling, and D. Weinbaum. 2015. Aggregate Jump and Volatility Risk in the Cross-Section of Stock Returns. *The Journal of Finance* 70: 577–614.

Douglas, G. 1967. *Risk in Equity Markets: An Empirical Appraisal of Market Efficiency*. University Microfilms.

Fama, E. F., and K. R. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33: 3–56.

Fama, E. F., and K. R. French. 2004. The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives* 18: 25–46.

Fama, E. F., and K. R. French. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116: 1–22.

Fama, E. F., and J. D. MacBeth. 1973. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* 81: 607–636.

Fan, J., Y. Fan, and J. Lv. 2008. High Dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics* 147: 186–197.

Fan, J., Y. Liao, and M. Mincheva. 2011. High-Dimensional Covariance Matrix Estimation in Approximate Factor Models. *Annals of Statistics* 39: 3320–3356.

Fan, J., Y. Liao, and M. Mincheva. 2013. Large Covariance Estimation by Thresholding Principal Orthogonal Complements. *Journal of the Royal Statistical Society. Series B* 75: 603–680.

Fan, J., Y. Liao, and J. Yao. 2015. Power Enhancement in High-Dimensional Cross-Sectional Tests. *Econometrica: Journal of the Econometric Society* 83: 1497–1541.

Feng, L., W. Lan, B. Liu, and Y. Ma. 2022. High-Dimensional Test for Alpha in Linear Factor Pricing Models with Sparse Alternatives. *Journal of Econometrics* 229: 152–175.

Gagliardini, P., E. Ossola, and O. Scaillet. 2016. Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets. *Econometrica* 84: 985–1046.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A test of the Efficiency of a Given Portfolio. *Econometrica* 57: 1121–1152.

Giglio, S., and D. Xiu. 2021. Asset Pricing with Omitted Factors. *Journal of Political Economy* 129: 1947–1990.

Gungor, S., and R. Luger. 2009. Exact Distribution-Free Tests of Mean–Variance Efficiency. *Journal of Empirical Finance* 16: 816–829.

Gungor, S., and R. Luger. 2016. Multivariate Tests of Mean–Variance Efficiency and Spanning with a Large Number of Assets and Time-Varying Covariances. *Journal of Business & Economic Statistics* 34: 161–175.

He, A., D. Huang, M. Yuan, and G. Zhou. 2021. Tests of Asset Pricing Models with a Large Number of Assets. DOI:10.2139/ssrn.3143752.

Huang, J. Z., N. Liu, M. Pourahmadi, and L. Liu. 2006. Covariance Matrix Selection and Estimation via Penalised Normal Likelihood. *Biometrika* 93: 85–98.

Hwang, S., and S. E. Satchell. 2014. Testing Linear Factor Models on Individual Stocks Using the Average *F*-Test. *The European Journal of Finance* 20: 463–498.

Im, K. S., M. Pesaran, and Y. Shin. 2003. Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* 115: 53–74.

Jensen, M. 1972. *Studies in the Theory of Capital Markets*. Praeger.

Jensen, M. C. 1968. The Performance of Mutual Funds in the Period 1945–1964. *The Journal of Finance* 23: 389–416.

Kelejian, H. H., and I. R. Prucha. 2001. On the Asymptotic Distribution of the Moran I Test Statistic with Applications. *Journal of Econometrics* 140: 219–257.

Lan, W., L. Feng, and R. Luo. 2018. Testing High-Dimensional Linear Asset Pricing Models. *Journal of Financial Econometrics* 16: 191–210.

Ledoit, O., and M. Wolf. 2004. A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis* 88: 365–411.

Lieberman, O. 1994. A Laplace Approximation to the Moments of a Ratio of Quadratic Forms. *Biometrika* 81: 681–690.

Lintner, J. 1965. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics* 47: 13–37.

Longin, F., and B. Solnik. 2001. Extreme Correlation of International Equity Markets. *The Journal of Finance* 56: 649–676.

Ma, S., W. Lan, L. Su, and C.-L. Tsai. 2020. Testing Alphas in Conditional Time-Varying Factor Models with High-Dimensional Assets. *Journal of Business & Economic Statistics* 38: 214–227.

Pesaran, M. H., A. Ullah, and T. Yamagata. 2008. A Bias-Adjusted LM Test of Error Cross-Section Independence. *The Econometrics Journal* 11: 105–127.

Raponi, V., C. Robotti, and P. Zaffaroni. 2019. Testing Beta-Pricing Models Using Large Cross-Sections. *The Review of Financial Studies* 33: 2796–2842.

Ross, S. A. 1976. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* 13: 341–360.

Shanken, J. 1992. On the Estimation of Beta-Pricing Models. *The Review of Financial Studies* 5: 1–33.

Sharpe, W. F. 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance* 19: 425–442.

Srivastava, M. S., and M. Du. 2008. A Test for the Mean Vector with Fewer Observations than the Dimension. *Journal of Multivariate Analysis* 99: 386–402.

Wong, F., C. K. Carter, and R. Kohn. 2003. Efficient Estimation of Covariance Selection Models. *Biometrika* 90: 809–830.