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

M. Hashem Pesaran
Department of Economics & USC Dornsife INET,
University of Southern California, USA, and Trinity College, Cambridge

Takashi Yamagata
DERS, University of York and ISER, Osaka University

11 March 2017

Abstract

This paper proposes a novel test of zero pricing errors for the linear factor pricing model when the number of securities, N, can be large relative to the time dimension, T, of the return series. The test is based on Student t tests of individual securities and has a number of advantages over the existing standardised Wald type tests. It 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. Monte Carlo evidence shows that the proposed test performs remarkably well even when $T = 60$ and $N = 5,000$. 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 CAPM and Fama-French three factor models are found mainly during the recent financial crisis. Also we find a significant negative correlation between a twelve-months moving average p-values of the test and excess returns of long/short equity strategies (relative to the return on S&P 500) over the period November 1994 to June 2015, suggesting that abnormal profits are earned during episodes of market inefficiencies.

JEL Classification: C12, C15, C23, G11, G12
Keywords: CAPM, Testing for alpha, Weak and spatial error cross-sectional dependence, S&P 500 securities, Long/short equity strategy.

*The first author acknowledges partial support from the ESRC Grant No. ES/I031626/1.

1 Introduction

This paper is concerned with testing for the presence of alpha in Linear Factor Pricing Models (LFPM) 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 the number of securities, N, is quite large relative to the time dimension, T, of the return series under consideration. The Sharpe-Lintner CAPM model predicts that expected excess returns (measured relative to the risk-free rate) on any given security or a given portfolio of securities is proportional to the expected excess return on the market portfolio, with the constant of the proportionality, β, being security/portfolio specific.

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] The test can be applied to individual securities as well as to portfolios.

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 Shaken (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 $\beta's$ 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 (2016) 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. Chorida, Goyal and Shanken (2015) advocate the use of individual securities to investigate whether the source of expected return variation is from betas or security-specific characteristics.

It is clearly desirable to develop tests of market efficiency that can deal with a large number of securities over relatively short time periods so that the problem of time variations in $\beta's$ is somewhat mitigated. It is also important that such tests are reasonably robust to non-Gaussian errors, particularly as it is more likely that one would encounter non-normal errors in the case of LFPM regressions for individual securities as compared to regressions estimated on portfolios comprising a large number of securities.

Out of the two main assumptions that underlie the GRS test, the literature has focussed 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 amongst 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

[1] Cross sectional tests of CAPM have been considered by Douglas (1968), 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).

Khalaf (2007, BDK) and Gungor and Luger (2009, GL) 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. Gungor and Luger (2009) 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.[2]

Our primary focus in this paper is on development of multivariate tests of $H_0 : \alpha_i = 0$, for $i = 1, 2, ..., N$, when $N > T$, whilst 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.[3] Fan, Liao and Mincheva (2011, 2013) consider consistent estimation of $\mathbf{V}$ in the context 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) derive the conditions under which standardised Wald tests of H_0 can be asymptotically justified. Gagliardini, Ossola and Scaillet (2016) develop two-pass regressions of individual stock returns, allowing time-varying risk premia, and propose a standardised Wald test. Raponi, Robotti and Zaffaroni (2016) propose a test of pricing error in cross-section regression for fixed number of time series observations. They use a bias-corrected estimator of Shaken (1992) to standardise their test statistic. Gungor and Luger (2016) propose simulation based approach for testing pricing errors. They claim that their test procedure is robust against non-normality and cross-sectional dependence in errors. Amengual and Repetto (2014) consider the standardised F-type test statistic based on principal component estimation under both serial and cross-section correlation in errors.

In this paper we follow an alternative strategy where 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 t tests of $\alpha_i = 0$, over $i = 1, 2, ..., N$. We then correct the standardized version of this average statistic for the effects of non-zero off-diagonal elements. The correction involves consistently estimating $N^{-1}Tr\left(\mathbf{R}^2\right)$, where $\mathbf{R} = \left(\rho_{ij}\right)$ is the error correlation matrix. The estimation of $N^{-1}Tr\left(\mathbf{R}^2\right) = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{ij}^2$ is subject to the curse of dimensionality which we address by using the multiple testing threshold estimator, $\tilde{\mathbf{R}}$, recently proposed by Bailey, Pesaran and Smith (2016). We show that consistent estimation of $N^{-1}Tr\left(\mathbf{R}^2\right)$ can be achieved under more general specification of $\mathbf{R}$ as compared to tests that require 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 standardised

[2] Bossaerts, Plot and Zame (2007) provide a novel GMM test of CAPM which does not require large T, but is designed for the analysis of experimental data on a few risky assets held across a relatively large number of subjects. It is interesting to see if their approach can be adapted to the analysis of historical observations of the type considered in this paper.

[3] 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, Liu, Pourahmadi, and Liu (2006), Bickel and Levina (2008), Fan, Fan and Lv (2008), Cai and Liu (2011), Fan, Liao and Mincheva (2011, 2013), and Bailey, Pesaran and Smith (2016).

Wald test, in cases where the off-diagonal elements of $\mathbf{V}$ become relatively less important as $N \to \infty$. The implementation of the $\hat{J}_\alpha$ test is also computationally less demanding, since it does not involve estimation of an invertible high dimensional error covariance matrix.

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 $\sqrt{N}$, whilst controlling the overall sparsity of $\mathbf{R}$ by requiring $N^{-1}Tr(\mathbf{R}^2)$ 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 arise from interactive and spill-over effects.

We establish that under the null hypothesis of $\alpha_i = 0$, the $\hat{J}_\alpha$ test is asymptotically distributed as $N(0,1)$ for T and $N \to \infty$ jointly, so long as $N/T^2 \to 0$, $m_N = \|\mathbf{R}\|_1 = O\left(N^{\delta_\rho}\right)$, $0 \leq \delta_\rho < 1/2$, and $N^{-1}Tr\left(\mathbf{R}^2\right)$ is bounded in N. The test is also shown to have power against alternatives that rises in $N^{1/2}T$. The proofs are quite involved and in some parts rather tedious. For the purpose of clarity we provide statements of the main theorems with the associated assumptions in the paper, but relegate the mathematical details to an appendix.

Small sample properties of the $\hat{J}_\alpha$ test are investigated using Monte Carlo experiments designed specifically to match the correlations, volatilities, and other distributional features (skewness and kurtosis) 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 eight sample size combinations, $T = 60$ and 100, and $N = 50, 100, 200$ and 500. The $\hat{J}_\alpha$ test performs well for all sample size combinations with 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 $N = 50 < T = 60$, 100 for which the GRS statistics can be computed, we find that the $\hat{J}_\alpha$ test has a 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) and Affleck-Graves and McDonald (1990). In addition, the $\hat{J}_\alpha$ test outperforms the feasible versions of the standardised Wald tests, replacing $\mathbf{V}$ with the recently developed estimators of large dimensional variance-covariance matrix of Fan, Liao and Mincheva (2013, FLM) and Ledoit-Wolf (2004). The $\hat{J}_\alpha$ test also outperforms the simulation-based $F_{\max}$ test of Gungor and Luger (2016) that can be implemented when $N > T$. The $F_{\max}$ test is shown to be undersized substantially across the various designs, and have lower power uniformly as compared to the $\hat{J}_\alpha$ test. We also carried out additional experiments that allow for time variations in betas as well as errors with a mixture of weak factors and spatial autoregressive processes, using much larger values of N, namely $N = 1,000$, $2,000$ and $5,000$, whilst keeping T at 60 and 100. 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 $N = 5,000$.

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 Standard and Poor (S&P) 500 index using rolling windows of size 60 over the period September 1989 to June 2015. 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 statistics for a single-factor and a three Fama-French factor model over the period 1989-2015, and find statistically significant evidence against Sharpe-Lintner CAPM and Fama-French factor model only during the recent financial crisis.

Finally, we examine if there exists any relationship between the p-values of $\hat{J}_\alpha$ test and excess returns on long/short equity hedge funds (relative to the return on S&P 500). *A priori* one would expect a reverse relationship between market efficiency and excess return of an

investment strategy, with excess returns being low during periods of market efficiency (high p-values) and *vice versa*. In fact, we find a significant negative correlation between a twelve-months moving average p-values of $\hat{J}_\alpha$ test and excess returns of long/short equity strategies over the period November 1994 to June 2015, suggesting that abnormal profits are earned during episodes of market inefficiencies.

The outline of the rest of the paper is as follows. Section 2 sets out the panel data model for the analysis of LFPM, and the GRS test. Section 3 proposes the $\hat{J}_\alpha$ test for large N panels, derives its asymptotic distribution, and Section 4 summarises the main theoretical results. Section 5 reports on small sample properties of $\hat{J}_\alpha$, GRS, standardised Wald tests and Gungor and Luger (2016) simulation based $F_{\max}$ test, using Monte Carlo techniques. Section 6 presents the empirical application. Section 7 concludes. The proofs of main theorems are provided in Appendix A, and the lemmas which are used for the proofs, as well as the additional Monte Carlo evidence, are provided in an Online Supplement to this paper, that is available on request.

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 $t \to \infty$. For a $N \times N$ matrix $\mathbf{A} = (a_{ij})$, the minimum and maximum eigenvalues of matrix $\mathbf{A}$ is denoted by $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$, respectively, its trace by $Tr(\mathbf{A})$, 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, $\boldsymbol{\alpha}$, $\|\boldsymbol{\alpha}\| = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{1/2}$.

2 Some preliminaries and the GRS test

Under Arbitrage Pricing Theory (APT) of Ross (1976), we have

$$R_{it} = \nu_t + \boldsymbol{\beta}_i'\boldsymbol{\lambda} + \boldsymbol{\beta}_i'(\mathbf{f}_t - \boldsymbol{\mu}_f) + u_{it}, \text{ for } i = 1, 2, ..., N; t = 1, 2, ..., T, \tag{1}$$

where, R_{it} is return on security i during period t, $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$ is the $m \times 1$ vector of factors, $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, ..., \beta_{im})'$ is the associated vector of risk factors, and ν_t is zero-beta expected return which under APT should be equal to the risk-free rate, $\boldsymbol{\lambda}$ is the vector of expected cross-sectional risk premium and $\boldsymbol{\mu}_f = E(\mathbf{f}_t)$. Setting $\nu_t = r_t + \nu$, where r_t is the risk-free rate, the return regressions can be written as

$$y_{it} = \alpha_i + \boldsymbol{\beta}_i'\mathbf{f}_t + u_{it}, \text{ for } i = 1, 2, ..., N; t = 1, 2, ..., T, \tag{2}$$

where $y_{it} = R_{it} - r_t$, and

$$\alpha_i = \nu + \boldsymbol{\beta}_i'(\boldsymbol{\lambda} - \boldsymbol{\mu}_f). \tag{3}$$

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

$$\sup_s \sum_{i=1}^N |\beta_{is}| = O(N), \tag{4}$$

and allow for the presence of common unobserved weak factors in the error term u_{it}. Specifically we assume that

$$u_{it} = \boldsymbol{\gamma}_i'\mathbf{v}_t + \eta_{it}, \tag{5}$$

where $\mathbf{v}_t$ is a $k\times 1$ vector of unobserved common factors that are $IID(\mathbf{0},\mathbf{I}_k)$, $\boldsymbol{\gamma}_i=(\gamma_{i1},\gamma_{i2},...,\gamma_{ik})'$ is the associated vector of factor loadings with bounded elements, $\sup_{i,s}|\gamma_{is}|<K$. The factors included in the error process are weak in the sense that their effects are not pervasive and satisfy the condition

$$\sup_s \sum_{i=1}^{N}|\gamma_{is}| = O\left(N^{\delta_\gamma}\right), \text{ with } 0\leq \delta_\gamma < 1/2. \tag{6}$$

The idiosyncratic errors, η_{it}, are also allowed to be weakly cross correlated. Specifically, we assume that $\boldsymbol{\eta}_t=(\eta_{1t},\eta_{2t},....,\eta_{Nt})'=\mathbf{Q}_\eta\boldsymbol{\varepsilon}_{\eta,t}$, where $\boldsymbol{\varepsilon}_{\eta,t}=(\varepsilon_{\eta,1t},\varepsilon_{\eta,2t},....,\varepsilon_{\eta,Nt})'$, $\{\varepsilon_{\eta,it}\}$ are IID processes over i and t, with means zero, unit variances, $\gamma_{2,\varepsilon_\eta}=E\left(\varepsilon^4_{\eta,it}\right)-3$, 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 $\mathbf{R}_\eta=\left(\rho_{\eta,ij}\right)$, and note that $\mathbf{R}_\eta=\mathbf{Q}_\eta\mathbf{Q}'_\eta$. To ensure that $\mathbf{u}_t=(u_{1t},u_{2t},...,u_{Nt})'$ is weakly cross-correlated we require that k, the number of weak factors, is finite, and that $\|\mathbf{Q}_\eta\|_\infty\leq K$ and $\|\mathbf{Q}_\eta\|_1\leq K$. The error specification in (5) is quite general and allows for common factors as well as network and spatial error cross dependence, so long as the common factors are sufficiently weak.

Different tests of LFPM are proposed in the literature. Some researchers have focussed on testing $\nu=0$, which ensures that the zero-beta excess return is zero. Others have considered testing the restrictions $\boldsymbol{\lambda}=\boldsymbol{\mu}_f$, which require that the risk-premia on factors coincide with factor means.[4] In this paper we adopt a more direct approach and consider testing the joint hypotheses

$$H_0:\alpha_i=0,\ i=1,2,...,N, \tag{7}$$

allowing for the multiple testing nature of the null. In the context of the APT model, the test of $\alpha_i=0$ for all i can be interpreted as a test of the joint hypotheses that $\nu=0$, and $\boldsymbol{\lambda}=\boldsymbol{\mu}_f$.

It proves useful to stack the panel regressions in (2) by time series as well as by cross section observations. Stacking by time series observations we have

$$\mathbf{y}_{i.}=\alpha_i\boldsymbol{\tau}_T+\mathbf{F}\boldsymbol{\beta}_i+\mathbf{u}_{i.}, \tag{8}$$

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

$$\mathbf{y}_t=\boldsymbol{\alpha}+\mathbf{B}\mathbf{f}_t+\mathbf{u}_t, \tag{9}$$

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

For exact sample tests of LFPM, initially we assume that $\mathbf{u}_t\sim IIDN\left(\mathbf{0},\mathbf{V}\right)$, 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 $E\left(\mathbf{u}_t\mathbf{u}'_t\right)=\mathbf{V}$, 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 given by

$$\hat{\alpha}_i=\mathbf{y}'_{i.}\left(\frac{\mathbf{M}_F\boldsymbol{\tau}_T}{\boldsymbol{\tau}'_T\mathbf{M}_F\boldsymbol{\tau}_T}\right), \tag{10}$$

where $\mathbf{M}_F=\mathbf{I}_T-\mathbf{F}\left(\mathbf{F}'\mathbf{F}\right)^{-1}\mathbf{F}'$, is an efficient estimator despite the fact that $\mathbf{V}$ is not a diagonal matrix. This result follows since (8) is a seemingly unrelated regression equation (SURE) specification with the *same* set of regressors across all the N securities. It is also easily seen that for all $i=1,2,...,N$,

$$\hat{\alpha}_i=\left(\alpha_i\boldsymbol{\tau}'_T+\boldsymbol{\beta}'_i\mathbf{F}'+\mathbf{u}'_{i.}\right)\left(\frac{\mathbf{M}_F\boldsymbol{\tau}_T}{\boldsymbol{\tau}'_T\mathbf{M}_F\boldsymbol{\tau}_T}\right)=\alpha_i+\mathbf{u}'_{i.}\mathbf{c}, \tag{11}$$

[4] See, for example, Shanken (1992).

where

$$\mathbf{c} = \mathbf{M}_F \boldsymbol{\tau}_T / \boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T. \tag{12}$$

Writing the above set of estimates for all i in matrix notations, 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

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \sum_{t=1}^{T} \mathbf{u}_t c_t, \tag{13}$$

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

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

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

$$\hat{\mathbf{V}} = T^{-1} \sum\nolimits_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t', \tag{14}$$

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

Under the 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, for example, Theorem 5.2.2 in Anderson (2003)) the *GRS* statistic (see p.1124 of GRS)

$$GRS = \hat{W}_0 = \frac{T - N - m}{N} \left(\frac{\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T}{T}\right) \hat{\boldsymbol{\alpha}}' \hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\alpha}}, \tag{15}$$

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 $H_0 : \boldsymbol{\alpha} = \mathbf{0}$.[5]

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. To circumvent this limitation, 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, it is clearly desirable to develop tests of $\alpha_i = 0$, that can be applied to a very large number of *individual* securities over relatively short time periods (to minimize the adverse effects of structural change in β_i's) which inevitably lead to cases where $T < N$.

Even in cases where $N < T$, the power of the GRS test could be compromised since it assumes $\mathbf{V}$ to be unrestricted, whilst in the context of approximate factor model 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. As we shall see below, a test that exploits

[5] Noting that $(1 + \bar{\mathbf{f}}' \hat{\boldsymbol{\Omega}}^{-1} \bar{\mathbf{f}})^{-1} = T^{-1} (\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T)$, it is easily seen that (15) can be written as the widely used expression of *GRS* statistic, $\frac{T-N-m}{N}(1 + \bar{\mathbf{f}}' \hat{\boldsymbol{\Omega}}^{-1} \bar{\mathbf{f}})^{-1} \hat{\boldsymbol{\alpha}}' \hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\alpha}}$, where $\bar{\mathbf{f}} = T^{-1} \sum_{t=1}^{T} \mathbf{f}_t$, and $\hat{\boldsymbol{\Omega}} = T^{-1} \sum_{t=1}^{T} (\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})'$.

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. It is also important to bear in mind that being a multivariate F test, the power of the GRS test is primarily driven by the time dimension, T, whilst 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 LFMP models

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 (15),

$$W_v = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{V}^{-1} \hat{\boldsymbol{\alpha}}, \tag{16}$$

where we have dropped the degrees of freedom adjustment term and replaced $\hat{\mathbf{V}}$ by its true value. W_v can be regarded as a Wald test statistic, and under Gaussianity and $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, $W_v \sim \chi^2_N$. Since the mean and the variance of a χ^2_N random variable is N and $2N$, one could consider a standardised Wald test statistic

$$SW_v = \frac{\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{V}^{-1} \hat{\boldsymbol{\alpha}} - N}{\sqrt{2N}}. \tag{17}$$

Under Gaussianity and $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, $SW_v \to_d N(0,1)$ as $N \to \infty$. 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 some 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 thresholds estimator, $\hat{\mathbf{V}}_{POET}$, which they show to be positive definite with satisfactory small sample properties. We refer to the feasible standardized Wald test statistic replacing $\mathbf{V}$ with $\hat{\mathbf{V}}_{POET}$ as SW_{POET} test. Another candidate is the shrinkage estimator of $\mathbf{V}$ proposed by Ledoit-Wolf (2004), which we denote by $\hat{\mathbf{V}}_{LW}$, and refer to the associated standardised Wald statistic as SW_{LW}. Such "plug-in" approaches are subject to two important short comings. First, even if $\mathbf{V}$ can be estimated consistently, the test might perform poorly in the case of non-Gaussian errors. Notice that the standardisation 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 for finite samples where $N >> T$.

3.1 A $\hat{J}_\alpha$ test for large N securities

To avoid some of the above mentioned limitations of the plug-in procedure, 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, $\mathbf{D} = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})$, with $\sigma_{ii} = E\left(u_{it}^2\right)$, rather than the full covariance matrix. Specifically, we consider the statistic

$$W_d = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{D}^{-1} \hat{\boldsymbol{\alpha}} = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \sum_{i=1}^{N} \left(\frac{\hat{\alpha}_i^2}{\sigma_{ii}}\right), \tag{18}$$

and its feasible counterpart given by

$$\hat{W}_d = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \hat{\mathbf{D}}_v^{-1} \hat{\boldsymbol{\alpha}} = \left(\frac{\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T}{v^{-1} T}\right) \sum_{i=1}^{N} \left(\frac{\hat{\alpha}_i^2}{\hat{\sigma}_{ii}}\right), \tag{19}$$

where $\hat{\sigma}_{ii} = \hat{\mathbf{u}}_{i.}'\hat{\mathbf{u}}_{i.}/T$, and the degrees of freedom $v = T - m - 1$ is introduced to correct for small sample bias of the test.[6] The infeasible statistic, W_d, can also be written as

$$W_d = \sum_{i=1}^{N} z_i^2, \tag{20}$$

where

$$z_i^2 = \hat{\alpha}_i^2(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)/\sigma_{ii}. \tag{21}$$

It is then easily seen that

$$\hat{W}_d = \sum_{i=1}^{N} t_i^2, \tag{22}$$

where t_i denotes the standard t-ratio of α_i in the OLS regression of y_{it} on an intercept and $\mathbf{f}_t$, namely

$$t_i^2 = \frac{\hat{\alpha}_i^2(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)}{v^{-1}T\hat{\sigma}_{ii}}. \tag{23}$$

As with the panel testing strategy developed in Im et al. (2003), a standardized version of $\hat{W}_d$, defined by (19), can now be considered:

$$\frac{N^{-1/2}\left[\hat{W}_d - E\left(\hat{W}_d\right)\right]}{\sqrt{Var\left(\hat{W}_d\right)}}, \tag{24}$$

where

$$N^{-1}E\left(\hat{W}_d\right) = E\left(t_i^2\right), \tag{25}$$

$$N^{-1}Var\left(\hat{W}_d\right) = N^{-1}Var\left(\sum_{i=1}^{N} t_i^2\right) = N^{-1}\sum_{i=1}^{N} Var\left(t_i^2\right) + \frac{2}{N}\sum_{i=2}^{N}\sum_{j=1}^{i-1} Cov\left(t_i^2, t_j^2\right). \tag{26}$$

Under Gaussianity, the individual t_i statistics are identically distributed as Student t with v degrees of freedom, and we have (assuming $v = T - m - 1 > 4$)

$$E(t_i^2) = \frac{v}{v-2},\ Var(t_i^2) = \left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4}. \tag{27}$$

Using (25), (26) and (27), the standardized statistic (24) can now be written as

$$J_\alpha\left(\theta_N^2\right) = \frac{N^{-1/2}\left[\hat{W}_d - E\left(\hat{W}_d\right)\right]}{\sqrt{Var\left(\hat{W}_d\right)}} = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{v}{v-2}\right)}{\sqrt{\left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4}\left(1+\theta_N^2\right)}}, \tag{28}$$

where

$$\theta_N^2 = N^{-1}\sum_{i=2}^{N}\sum_{j=1}^{i-1} Corr\left(t_i^2, t_j^2\right), \tag{29}$$

and

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

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 standardised 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.

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

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, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$, 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 $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{V} = (\sigma_{ij}) = \mathbf{QQ}'$, 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 paper.

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 $\rho_{ij} = \sigma_{ij}/\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}$. We consider the following two conditions

$$m_N = \max_{1 \leq i \leq N} \sum_{j=1}^{N} |\rho_{ij}| = O(N^{\delta_\rho}), \text{ with } 0 \leq \delta_\rho < 1/2, \tag{30}$$

and

$$Tr(\mathbf{R}^2) = \sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{ij}^2 = O(N). \tag{31}$$

Under (30), m_N is allowed to rise with N, but at a slower rate than $N^{1/2}$. Strict sparsity requires m_N to be bounded in N, which is often assumed in the literature on consistent estimation of large covariance matrices. Conditions (30) and (31) allow for a general form of weak correlations across the errors, including the familiar spatial or local dependence, and is compatible with (30). For example, consider the case where condition (30) 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

$$\begin{aligned}
\sum_{j=1}^{N} |\rho_{ij}| &= O\left(N^{\delta_\rho}\right), \text{ for } i = 1, 2, ..., p, \\
\sum_{j=1}^{N} |\rho_{ij}| &= O(1), \text{ for } i = p+1, p+2, ..., N.
\end{aligned}$$

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

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

Another important case covered by our sparsity assumption is when u_{it} has the weak factor structure given by (5), with the factor loadings, $\boldsymbol{\gamma}_i$, satisfying (6). Denoting the correlation matrix of the idiosyncratic errors, $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, ..., \eta_{Nt})'$ by $\mathbf{R}_\eta = (\rho_{\eta,ij})$, and assuming that

$$\|\mathbf{R}_\eta\|_\infty < K, \tag{32}$$

we have $Tr\left(N^{-1}\mathbf{R}_{\eta}^{2}\right)=O(1)$. It is now easily seen that conditions (30) and (31) are also satisfied under this set up. Denoting the correlation matrix of $\mathbf{u}_t=(u_{1t},u_{2t},....,u_{Nt})'$ by $\mathbf{R}=\left(\rho_{ij}\right)$ we have

$$\rho_{ij}=\tilde{\boldsymbol{\gamma}}_i'\tilde{\boldsymbol{\gamma}}_j+\left(\frac{\sigma_{\eta,ii}\sigma_{\eta,jj}}{\sigma_{ii}\sigma_{jj}}\right)^{1/2}\rho_{\eta,ij}, \tag{33}$$

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 $\left|\rho_{ij}\right|\leq\sum_{s=1}^{k}\left|\tilde{\gamma}_{is}\right|\left|\tilde{\gamma}_{js}\right|+\left|\rho_{\eta,ij}\right|$, then

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

Since $\sup_{i,s}\left|\tilde{\gamma}_{is}\right|\leq\sup_{i,s}\left|\gamma_{is}\right|$, and $\sup_s\sum_{j=1}^{N}\left|\tilde{\gamma}_{js}\right|\leq\sup_s\sum_{j=1}^{N}\left|\gamma_{js}\right|=O(N^{\delta_\gamma})$, and by assumption $\|\mathbf{R}_{\eta}\|_{\infty}<K$, the condition (30) is met if $\delta_\rho\leq\delta_\gamma$. Also, (noting that $\sup_{i,s}\left|\tilde{\gamma}_{is}\right|\leq 1$)

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

Therefore, under conditions (6) and (32), $N^{-1}Tr\left(\mathbf{R}^2\right)$ is bounded in N if $0\leq\delta_\gamma<1/2$.

Remark 1 *Our assumption of approximate sparsity allows for a sufficiently high degree of cross error correlations, which is important for the analysis of financial data, where it is not guaranteed that inclusion of common 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 need to focus on identification of strong factors to be included in f_t, which can be achieved by using market factors or principal components of individual returns.*[7] *The error associated with the estimation of strong factors is likely to be negligible for N and T sufficiently large. In the present paper we abstract from such estimation errors and condition our analysis on given values of $\mathbf{f}_t$.*

[7] Note also that the consistency of the plug-in procedure proposed by Fan, Liao and Mincheva (2011, 2013) also requires that strong common factors are removed before estimation of the error covariance matrix, $\mathbf{V}$.

3.3 Non-Gaussianity

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

$$\xi_{it} = u_{it}/\sigma_{ii}^{1/2}, \tag{34}$$

so that for each i, ξ_{it} has mean zero a 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, that 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 Laplace approximation of moments of ratio of quadratic forms by Lieberman (1994), we are able to derive the following approximations of $E(t_i^2)$ and $Var(t_i^2)$:[8]

$$E\left(t_i^2\right) = \frac{v}{v-2} + O\left(v^{-3/2}\right), \tag{35}$$

and

$$Var\left(t_i^2\right) = \left(\frac{v}{v-2}\right)^2 \frac{2\left(v-1\right)}{\left(v-4\right)} + O\left(v^{-1}\right). \tag{36}$$

Substituting (35) and (36) into (24) we have the following non-Gaussian version of $J_\alpha\left(\theta_N^2\right)$, defined by (28):

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

where θ_N^2 is defined by (29). When the numerator of the J_α statistic is replaced by $N^{-1/2}\sum_{i=1}^N\left(t_i^2 - 1\right)$, which is the typical mean adjustment employed by Fan et al (2015) and Gagliardini et al. (2016), for example, then the order of the asymptotic error of the numerator such test statistics becomes $\sqrt{N/v^2}$ as compared to $\sqrt{N/v^3}$ obtained for the J_α test. This is one reason why our proposed test performs better than the ones proposed in the literature, especially in cases where $N >> T$. The asymptotic error of using $\left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{(v-4)}$ for $Var(t_i^2)$ under non-Gaussianity in the J_α test is $O(v^{-1})$, which is small for sufficiently large v.[9]

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. As it will be shown more formally below, the limiting property of θ_N^2 is governed by the sparsity of $\mathbf{V}$, and is given by[10]

$$\theta_N^2 - (N-1)\rho_N^2 \to 0, \tag{37}$$

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

$$\rho_N^2 = \frac{2}{N(N-1)}\sum_{i=2}^N\sum_{j=1}^{i-1}\rho_{ij}^2. \tag{38}$$

[8] See Lemma 21 in the Online Supplement to the paper, which is available upon request.

[9] Small sample evidence on the efficacy of using $N^{-1/2}\sum_{i=1}^N\left(t_i^2 - \frac{v}{v-2}\right)$ over $N^{-1/2}\sum_{i=1}^N\left(t_i^2 - 1\right)$ is reported in Table M3 of the Online Supplement, which is available upon request.

[10] (37) follows from Lemma 18 in the Online Supplement which is available on request.

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, for example, Breusch and Pagan (1980) and Pesaran, Ullah and Yamagata (2008). To see the relationship between θ_N^2 and the sparsity of $\mathbf{V}$, we note that

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

which in view of (37) justifies replacing $1+\theta_N^2$ by $N^{-1}Tr\left(\mathbf{R}^2\right)$ for N and T sufficiently large so long as $N/T^2 \to 0$, and $0 \leq \delta_\gamma < 1/2$. 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 $N^{-1}Tr\left(\mathbf{R}^2\right)$ 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 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

$$\hat{\rho}_{ij} = \hat{\sigma}_{ij}/\hat{\sigma}_{ii}^{1/2}\hat{\sigma}_{jj}^{1/2}, \tag{39}$$

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.[11] Here we consider the application of the multiple testing (MT) approach to regularisation of large covariance matrices recently proposed by Bailey Pesaran and Smith (2016, BPS). However, BPS establish their results for $y_{it} - \bar{y}_i$, whilst we need to apply the thresholding approach to $\hat{u}_{it}$. Second BPS consider exact sparsity conditions on the error covariance matrix, whilst we allow for a much more general sparsity conditions. We extend BPS's analysis to address both of these issues.[12] The multiple testing (MT) estimator of ρ_{ij}, denoted by $\tilde{\rho}_{ij}$, is given by

$$\tilde{\rho}_{ij} = \hat{\rho}_{ij} I\left[\left|\sqrt{v}\hat{\rho}_{ij}\right| > c_p(N)\right], \tag{40}$$

where $v = T - m - 1$,

$$c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right), \tag{41}$$

p is the nominal p-value $(0 < p < 1)$, and $f(N) = N^\delta$, $v = c_d N^d$, where c_d and δ are finite positive constants. Using (57), the multiple testing estimator of ρ_N^2 is given by

$$\tilde{\rho}_{N,T}^2 = \frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\tilde{\rho}_{ij}^2. \tag{42}$$

Under the sparsity conditions (30) and (31), it can be shown that $(N-1)\left(\tilde{\rho}_{N,T}^2 - \rho_N^2\right) \to_p 0$, so long as $N/T^2 \to 0$, as N and $T \to \infty$, jointly, and

$$\delta > (1 - 0.5d)\,\varphi, \tag{43}$$

where $\varphi \leq 1 + \left|\gamma_{2,\varepsilon_\eta}\right|$, where $\gamma_{2,\varepsilon_\eta} = E\left(\varepsilon_{\eta,it}^4\right) - 3$, $\varepsilon_{\eta,it}$ is the ith element of the $N \times 1$ error vector $\boldsymbol{\varepsilon}_{\eta,t} = \mathbf{Q}_\eta^{-1}\boldsymbol{\eta}_t$, with $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t}, ..., \eta_{Nt})'$.[13] The critical value function, $c_p(N)$,

[11] See, for example, Cai and Liu (2011), Fan et al. (2013), Bailey Pesaran and Smith (2016), among others.

[12] Other thresholding estimators of $\mathbf{V}$ proposed in the literature can also be used. The efficacy of using the estimator $\tilde{\rho}_{N,T}^2$ over other estimators in small samples is investigated and the results are summarised in Table M2 in the Online Supplement (available on request).

[13] See Theorem 4 in Section 4 and its proof in Appendix A.

depends on the nominal level of significance, p, and the choice of δ, subject to condition (43). The test results are unlikely to be sensitive to the choice of p, over the conventional values in the range of 1 to 10 per cent.[14] 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 $\delta > d/2$. 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, it is more prudent to set δ close to unity or even higher. In simulations and empirical exercises to be reported below we set $f(N) = N - 1$, which is equivalent to setting $\delta = 1$.[15]

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

$$\hat{J}_\alpha = \frac{N^{-1/2} \sum_{i=1}^{N} \left(t_i^2 - \frac{v}{v-2} \right)}{\left(\frac{v}{v-2} \right) \sqrt{\frac{2(v-1)}{(v-4)} \left[1 + (N-1)\tilde{\rho}_{N,T}^2 \right]}}, \tag{44}$$

where t_i is the t-ratio for testing $\alpha_i = 0$, defined by (23), $v = T - m - 1$, and $\tilde{\rho}_{N,T}^2$ is given by (42). The $\hat{J}_\alpha$ test is robust to non-Gaussian errors and allows for a relatively high degree of error cross-sectional dependence. In what follows we provide a formal statement of the conditions under which $\hat{J}_\alpha$ tends to a normal distribution.

3.5 Surviorship bias

Finally, it is important that the application of the $\hat{J}_\alpha$ test is not subject to the survivorship bias. The GRS type tests of alpha considers a relatively small number of portfolios over a relatively large time periods 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 surviorship 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 vary 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, say around 0.1. This procedure is followed in the empirical application discussed in Section 6 below, where we set $T = 60$ and end up with N_τ in the range $[464, 487]$, given $\min_\tau(N_\tau)/T^2 = 0.12$.

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 for establishing the results.

Assumption 1: The $m \times 1$ vector of common observed factors, $\mathbf{f}_t$, in the return regressions, (2), are distributed independently of the errors, $u_{it'}$ for all i, t and t'. The number of

[14] In the Monte Carlo experiments reported below, we set $p = 10\%$.

[15] The robustness of the J_a test against non-Gaussian and nonlinear error dependence is investigated and reported in Table 4 below. These results are generally supportive of setting $\delta = 1$.

factors, m, is fixed, and the factors can be strong in the sense that

$$\sup_s \sum_{i=1}^{N} |\beta_{is}| = O(N^{\delta_\beta}),\ 0 \leq \delta_\beta \leq 1. \tag{45}$$

and 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 $\mathbf{G} = (\mathbf{F}, \boldsymbol{\tau}_T)$, 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 $\mathbf{M}_F = \mathbf{I}_T - \mathbf{F}\left(\mathbf{F}'\mathbf{F}\right)^{-1}\mathbf{F}'$.

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

$$u_{it} = \boldsymbol{\gamma}_i'\mathbf{v}_t + \eta_{it},\ \text{for } i = 1, 2, ..., N; t = 1, 2, ..., T, \tag{46}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{ik})'$ is a $k \times 1$ vector of factor loadings, $\mathbf{v}_t = (v_{1t}, v_{2t}, ..., v_{kt})'$ 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 $\mathbf{v}_t \sim IID(\mathbf{0}, \mathbf{I}_k)$, $\gamma_{2,v} = E\left(v_{st}^4\right) - 3$, and $sup_{s,t}E\left(v_{st}^{8+c}\right) < K$, for some $c > 0$. The factor loadings, γ_{is} for $s = 1, 2, ..., k$, are bounded, $\sup_{i,s}|\gamma_{is}| < K$, and the factors, $\mathbf{v}_t$, are weak in the sense that

$$\sup_s \sum_{i=1}^{N} |\gamma_{is}| = O\left(N^{\delta_\gamma}\right), \text{ with } 0 \leq \delta_\gamma < 1/2. \tag{47}$$

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

Assumption 3: Denoting the standardized errors by $\xi_{it} = u_{it}/\sigma_{ii}^{1/2}$, with $\sigma_{ii} = \boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i + \sigma_{\eta,ii}$; then for any i and j, $\xi_{it}, \xi_{jt}, \xi_{it}^2, \xi_{jt}^2$, and $\xi_{it}\xi_{jt}$, for $t = 1, 2, ..., T$, are random draws from a common distribution which is absolutely continuous with non-zero density on subsets of $\mathbb{R}^5$.

Writing the error factor specification, (46), in matrix notation we have

$$\mathbf{u}_t = \boldsymbol{\Gamma}\mathbf{v}_t + \boldsymbol{\eta}_t, \tag{48}$$

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

$$E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \mathbf{V}_\eta = \mathbf{V} = (\sigma_{ij}),\ \text{with } \sigma_{ij} = \boldsymbol{\gamma}_i'\boldsymbol{\gamma}_j + \sigma_{\eta,ij}. \tag{49}$$

We now make the following further assumption.

Assumption 4: The covariance matrices $\mathbf{V}$ and $\mathbf{V}_\eta$ defined by (49) are $N \times N$ symmetric, positive definite matrices with $\lambda_{\min}(\mathbf{V}) \geq \lambda_{\min}(\mathbf{V}_\eta) \geq c$,

$$\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})' = \mathbf{Q}^{-1}\mathbf{u}_t, \text{ and } \boldsymbol{\varepsilon}_{\eta,t} = (\varepsilon_{\eta,1t}, \varepsilon_{\eta,2t}, ..., \varepsilon_{\eta,Nt})' = \mathbf{Q}_\eta^{-1}\boldsymbol{\eta}_t, \tag{50}$$

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

$$\|\mathbf{Q}_\eta\|_\infty < K, \text{ and } \|\mathbf{Q}_\eta\|_1 < K. \tag{51}$$

$\{\varepsilon_{it}\}$ and $\{\varepsilon_{\eta,it}\}$ are IID processes over i and t, with means zero, unit variances, $\gamma_{2,\varepsilon_\eta} = E\left(\varepsilon_{\eta,it}^4\right) - 3$, and $\sup_{i,t} E(|\varepsilon_{it}|^{8+c}) \leq K < \infty$, 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. Such residual interdependencies could arise due to spatial or other network type spill-over effects not captured by the observed common factors.*

Remark 3 *Under condition (51)*

$$\|\mathbf{V}_\eta\|_\infty \leq \left\|\mathbf{Q}_\eta \mathbf{Q}_\eta'\right\|_\infty \leq \|\mathbf{Q}_\eta\|_\infty \|\mathbf{Q}_\eta\|_1 < K = O(1), \tag{52}$$

nevertheless due to the weak factors we have

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

and allows the overall error variance matrix, $\mathbf{V}$*, to be approximately sparse, in contrast to the literature that requires* $\|\mathbf{V}\|_\infty < K$*. The relaxation of the sparsity condition on* $\mathbf{V}$ *is particularly important in finance where security returns could be affected by weak unobserved factors. Using principal components does not resolve the problem since, principal components provide consistent estimates of the factors (up to a rotation matrix) only if the factors are strong.*

Remark 4 *The high-order moment conditions in Assumption 4 allow us to relax the Gaussianity assumption whilst 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 (38). For further details see Bailey et al (2016).*

Our main theoretical results are set out in the following theorems. The proofs of these theorems are provided in Appendix A, and necessary lemmas for the proofs are given in the Online Supplement available upon request.

Theorem 1 *Consider the return regressions, (2), and the statistic* $\sum_{i=1}^N z_i^2$ *defined by (20). Suppose that Assumptions 1-4 hold, and* $N^{-1}Tr\left(\mathbf{R}^2\right)$ *is bounded in* N*, where* $\mathbf{R} = \left(\rho_{ij}\right)$*,* $\rho_{ij} = E(\xi_{it}\xi_{jt})$*, and* $\xi_{it} = u_{it}/\sigma_{ii}^{1/2}$ *is the standardized error of the return regression equation (2). Then, under* $H_0 : \alpha_i = 0$*, in (2) for all* i*,*

$$q_{NT} = N^{-1/2} \sum_{i=1}^N \left(z_i^2 - 1\right) \to_d N(0, 2\omega^2), \text{ as } N \to \infty \text{ and } T \to \infty, \text{ jointly,} \tag{53}$$

where

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

with

$$\rho_N^2 = \frac{2}{N(N-1)} \sum_{i=2}^N \sum_{j=1}^{i-1} \rho_{ij}^2. \tag{54}$$

Theorem 2 *Consider the regression model (2), and the statistics* $\sum_{i=1}^N z_i^2$ *and* $\sum_{i=1}^N t_i^2$*, which are defined by (20) and (22), respectively. Suppose that Assumptions 1-4 hold. Then, under the null hypothesis,* $H_0 : \alpha_i = 0$ *for all* i*,*

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

as $N \to \infty$ and $T \to \infty$ jointly, so long as $N/T^2 \to 0$, $0 \leq \delta_\gamma < 1/2$, where δ_γ is defined by (47).

Theorem 3 *Consider the regression model (2), and suppose that Assumptions 1-4 hold. Then, under $H_0 : \alpha_i = 0$, for all i,*

$$J_\alpha\left(\rho_N^2\right) = \frac{N^{-1/2}\sum_{i=1}^N \left(t_i^2 - \frac{v}{v-2}\right)}{\sqrt{\left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4}\left[1+(N-1)\rho_N^2\right]}} \to_d N(0,1), \tag{55}$$

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 t_i, ρ_N^2 and δ_γ are defined by (23), (54) and (47), respectively, with $v = T - m - 1$.

Theorem 4 *Let*

$$\tilde{\rho}_{N,T}^2 = \frac{2}{N(N-1)}\sum_{i=2}^N \sum_{j=1}^{i-1} \tilde{\rho}_{ij}^2, \tag{56}$$

where

$$\tilde{\rho}_{ij} = \hat{\rho}_{ij} I\left[\left|\sqrt{v}\hat{\rho}_{ij}\right| > c_p(N)\right], \tag{57}$$

$\rho_{ij} = E(\xi_{it}\xi_{jt})$, $\xi_{it} = u_{it}/\sigma_{ii}^{1/2}$ is the standardized error of the return regression equation (2), $v = T - m - 1$, $\hat{\rho}_{ij}$ is defined by (39)

$$c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right), \tag{58}$$

p is the nominal p-value ($0 < p < 1$), and $f(N) = N^\delta$ and $v = c_d N^d$, where c_d and δ are finite positive constants. Suppose that Assumptions 1-4 hold and

$$\sum_{i,j=1}^N \left|\rho_{ij}\right| = O(N).$$

Then

$$(N-1)\left(\tilde{\rho}_{N,T}^2 - \rho_N^2\right) \to_p 0,$$

if $N/v^2 = O\left(N^{1-2d}\right) \to 0$, (or if $d > 1/2$) as N and $v \to \infty$, and $\delta > (1-d/2)\varphi$, where $\varphi \leq 1 + \left|\gamma_{2,\varepsilon_\eta}\right|$, and $\gamma_{2,\varepsilon_\eta} = E\left(\varepsilon_{\eta,it}^4\right) - 3$ (Assumption 4).

Theorem 5 *Consider the panel regression model (2) in asset returns, and suppose that Assumptions 1-4 hold. Consider the statistic*

$$\hat{J}_\alpha = \frac{N^{-1/2}\sum_{i=1}^N \left(t_i^2 - \frac{v}{v-2}\right)}{\left(\frac{v}{v-2}\right)\sqrt{\frac{2(v-1)}{(v-4)}\left[1+(N-1)\tilde{\rho}_{N,T}^2\right]}}, \tag{59}$$

where t_i is given by (23), $v = T - m - 1$, $\tilde{\rho}_{N,T}^2$ is defined by (56), using the threshold $c_p(N)$ given by (58), with p ($0 < p < 1$), $f(N) = N^\delta$, $T = c_d N^d$, where c_d and δ are finite positive constants, $\delta > (1 - 0.5d)\varphi$, where $\varphi \leq 1 + \left|\gamma_{2,\varepsilon_\eta}\right|$, and $\gamma_{2,\varepsilon_\eta} = E\left(\varepsilon_{\eta,it}^4\right) - 3$. Then, under $H_0 : \alpha_i = 0$ for all i,

$$\hat{J}_\alpha \to_d N(0,1), \tag{60}$$

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

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

$$H_{0a} : \alpha_i = \frac{\varsigma_i}{N^{1/4}v^{1/2}}, \text{ with } 0 \leq |\varsigma_i| < \infty, \text{ for all } i. \tag{61}$$

Theorem 6 *Consider the panel regression model (2) in asset returns, and suppose that conditions of Theorem 5 apply. Then, under the local alternatives, $H_{0\alpha}$, defined by (61),*

$$\hat{J}_\alpha \to_d N\left(\phi^2/\sqrt{2}, 1\right), \tag{62}$$

where $\phi^2 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \varsigma_i^2/\sigma_{ii}$.

Remark 7 *This theorem establishes that the $\hat{J}_\alpha$ test is consistent (in the sense that its power tends to unity), if $\phi^2 > 0$, which is satisfied if $\lim_{N\to\infty}\left(N^{-1}\sum_{i=1}^N \varsigma_i^2\right) > 0$. It is also interesting to note 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. The $\hat{J}_\alpha$ test has power even if $\sum_{i=1}^N \alpha_i^2$ does not increase with N, so long as N increases sufficiently slowly as compared to T. To see this, let $\sum_{i=1}^N \alpha_i^2 = O\left(N^{\delta_\alpha}\right)$, and note that under the local alternatives, (61), and setting $T = O\left(N^d\right)$, we have $\sum_{i=1}^N \alpha_i^2 = \left(N^{-1}\sum_{i=1}^N \varsigma_i^2\right) N^{1/2-d} = O\left(N^{\delta_\alpha}\right)$, or $\left(N^{-1}\sum_{i=1}^N \varsigma_i^2\right) = O\left(N^{\delta_\alpha+d-1/2}\right)$. Hence, the proposed test will be consistent so long as $\delta_\alpha + d \geq 1/2$. The case of $\delta_\alpha = 0$ is of particular interest since it does not require all securities under consideration to have non-zero alphas for the test to have power.*

5 Small sample evidence based on Monte Carlo experiments

We examine the finite sample property of the $\hat{J}_\alpha$ test by Monte Carlo experiments, and compare its performance to a number of existing tests. For comparison, we consider the *GRS* test and the feasible versions of the standardised Wald tests, SW_{POET} and SW_{LW}, which are discussed in Section 3. We also consider the $F_{\max}$ test recently proposed by Gungor and Luger (2016, GL). They propose basing a test of $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ on simulated distribution of $F_{\max} = \max_{1\leq i\leq N} F_i$, where F_i is a standard F-statistic for testing $\alpha_i = 0$ in the OLS regression of y_{it} on an intercept and $\mathbf{f}_t$. The simulations are carried out by residual resampling allowing for cross-sectional correlations and cross-sectional heteroskedasticity using Wild bootstraps. GL employ a bounds testing approach to allow for unconsidered nuisance parameters, which could result in having inconclusive test outcomes.[16]

Computational details of the above tests are given in Section M1.2 of the Online Supplement available on request.

5.1 Monte Carlo designs and experiments

We consider the following data generating process (DGP)

$$r_{it} = \alpha_i + \sum_{\ell=1}^m \beta_{\ell i} f_{\ell t} + u_{it}, i = 1, 2, .., N; t = 1, 2, ..., T, \tag{63}$$

and calibrate its parameters to closely match the main features of the time series observations on individual returns and the three Fama-French factors (market factor, HML and SMB) used in the literature on tests of market efficiency.[17] The Monte Carlo (MC) design is also intended

[16] We also considered two distribution-free sign tests of $\alpha_i = 0$, proposed by Gungor and Luger (2009). These tests, referred to as SS and WS tests, are valid for single factor models with errors that are conditionally symmetric around zero, but they do allow for non-normal errors, are relatively easy to compute, and are applicable even when $N > T$. The results of these simulations are reported in Table M4 of the Online Supplement. These tests are also outperformed by the $\hat{J}_\alpha$ test.

[17] SMB stands for "small market capitalization minus big" and HML for "high book-to-market ratio minus low". See Fama and French (1993), and Appendix C for further details and data sources.

to match the models used for the empirical applications that follow. Accordingly, we set $m = 3$ and generate the factors as

$$\begin{aligned} f_{\ell t} &= 0.53 + 0.06 f_{\ell,t-1} + \sqrt{h_{\ell t}}\, \zeta_{\ell t}, \text{ for } \ell = 1, \text{ (Market factor)}, \\ f_{\ell t} &= 0.19 + 0.19 f_{\ell,t-1} + \sqrt{h_{\ell t}}\, \zeta_{\ell t}, \text{ for } \ell = 2, \ (HML), \\ f_{\ell t} &= 0.19 + 0.05 f_{\ell,t-1} + \sqrt{h_{\ell t}}\, \zeta_{\ell t}, \text{ for } \ell = 3, \ (SMB), \end{aligned}$$

where $\zeta_{\ell t} \sim IIDN(0,1)$ and[18]

$$\begin{aligned} h_{\ell t} &= 0.89 + 0.85 h_{\ell,t-1} + 0.11 \zeta^2_{\ell,t-1}, \text{ for } \ell = 1, \ Market, \\ h_{\ell t} &= 0.62 + 0.74 h_{\ell,t-1} + 0.19 \zeta^2_{\ell,t-1}, \text{ for } \ell = 2, HML, \\ h_{\ell t} &= 0.80 + 0.76 h_{\ell,t-1} + 0.15 \zeta^2_{\ell,t-1}, \text{ for } \ell = 3, SMB. \end{aligned}$$

The above processes are generated over the period $t = -49, -48, ...0, 1, 2, ..., T$ with $f_{\ell,-50} = 0$ and $h_{\ell,-50} = 1$ for $\ell = 1, 2, 3$. Observations $t = 1, 2, ..., T$ are used in the MC experiments.

To capture the main features of the individual asset returns and their cross correlations, we generate the idiosyncratic errors, $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$, according to $\mathbf{u}_t = \mathbf{Q}\boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$, and $\mathbf{Q} = \mathbf{D}^{1/2}\mathbf{P}$ with $\mathbf{D} = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})'$, $\sigma_{ii} = Var(u_{it})$, and $\mathbf{P}$ being a Cholesky factor of correlation matrix of $\mathbf{u}_t$, $\mathbf{R}$, which is an $N \times N$ matrix used to calibrate the cross correlation of returns. For each i, ε_{it} is generated such that u_{it} exhibits skewness and kurtosis which is typical of individual security returns. To this end, $\mathbf{R}$ is generated as

$$\mathbf{R} = \mathbf{I}_N + \mathbf{b}\mathbf{b}' - \check{\mathbf{B}}^2, \tag{64}$$

where $\mathbf{b} = (b_1, b_2, ..., b_N)'$, and $\check{\mathbf{B}} = diag(\mathbf{b})$. The correlation matrix $\mathbf{R}$ also arises from the single factor model, $u_{it} = \gamma_i v_t + \sigma^{1/2}_{\eta,ii}\eta_{it}$, with $v_t \sim IID(0,1)$, and $\eta_{it} \sim IID(0,1)$, and $b_i = \gamma_i/\sigma^{1/2}_{ii}$, where $\sigma_{ii} = \gamma_i^2 + \sigma_{\eta,ii}$. To generate different degrees of error cross-sectional dependence, we draw the first and the last N_γ $(< N)$ elements of $\mathbf{b}$ as $Uniform(0.7, 0.9)$, and set the remaining middle elements to 0. We set $N_\gamma = \lfloor N^{\delta_\gamma} \rfloor$, where $\lfloor A \rfloor$ is the largest integer part of A. Using δ_γ, our assumption $m_N = o(N^{1/2})$ can be expressed by $m_N = N^{\delta_\gamma}$ with $\delta_\gamma < 1/2$. In our experiments, we consider the values of exponents $\delta_\gamma = 1/4, 1/2$, and $3/5$. The case of no error cross-sectional dependence is obtained when $N_\gamma = 0$, and the error cross-sectional dependence is weak when $\delta_\gamma < 1/2$. The case of $\delta_\gamma = 3/5$ is included to see how the $\hat{J}_\alpha$ test performs when cross-sectional error correlations are higher than the threshold value of $1/2$ allowed by the theory. To save space, we omit reporting the results for the case where $\delta_\gamma = 0$ as they are qualitatively similar to the case with $\delta_\gamma = 1/4$. The present design focusses on the weak factor error correlations and assumes the idiosyncratic errors, η_{it}, are cross-sectionally uncorrelated. A more general design that allows for both forms of error correlations will be considered below.

Recently, Fan, Liao and Yao (2015; FLY) have derived the conditions under which the limiting normal distribution of SW_{POET} will be asymptotically justified. Under their assumptions the SW_{POET} test allows for $N > T$. However, FLY's assumptions are much more restrictive than ours.[19] For example, FLY do not cover cases where $1/4 < \delta_\gamma \leq 1/2$. When $\delta_\gamma = 1/4$, FLY require that $T = O\left(N \ln(N)^\kappa\right)$, for some $\kappa > 2$. Thus, when $\delta_\gamma = 1/4$, so long as T

[18] The estimates used in the generation of the factors and their volatilities are computed using monthly observations over the period April 1973 - September 2011.

[19] In addition to some regularity conditions, FLY require Assumption A.2. which defines their version of "sparseness": Suppose $N^{1/2}(\log N)^\kappa = o(T)$ for some $\gamma > 2$, and (i) $\min_{\sigma_{ij} \neq 0} |\sigma_{ij}| >> \sqrt{(\log N)/T}$; (ii) at least one of the following cases holds: (a) $D_N = 2\sum_{i=2}^N \sum_{j=1}^{i-1} I(\sigma_{ij} \neq 0) = O(N^{1/2})$ and $= O\left(\frac{T}{N^{1/2}(\ln N)^\kappa}\right)$ or; (b) $D_N = O(N)$ and $m_N^2 = O(1)$. Then they show that $SW_{POET} \to_d N(0,1)$,as $N, T \to \infty$ jointly (see Proposition 4.2 of FLY).

rises slightly faster than N, the SW_{POET} test is asymptotically justified. On the other hand, $\hat{J}_\alpha \to_d N(0,1)$ so long as $v = O\left(N^d\right)$ with $d > 2/3$ when $\delta_\gamma = 1/4$. Therefore, the $\hat{J}_\alpha$ test is expected to provide better finite sample approximation than the SW_{POET} test, especially when N is larger than T and/or when error cross-correlation is not very weak. The simulation results that follow seem to support these theoretical insights.[20]

To calibrate the variance, skewness and kurtosis of the simulated returns, we used estimated values of these measures based on residuals of Fama-French regressions for each security over the estimation windows τ =September 1989,..., September 2011, using sample of sizes of $T = 60$ months. Specifically, for each $i = 1, 2, ..., N_\tau$ we run the Fama-French regressions $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{\beta}_{2,i\tau} SMB_{t\tau} + \hat{\beta}_{3i} HML_{t\tau} + \hat{u}_{i,\tau t}$, $t = 1, 2, ..., 60$, at the end of each month τ =September 1989,..., September 2011, and computed $\hat{\sigma}_{ii,\tau} = \hat{m}_{2,i\tau}$, $\hat{\gamma}_{1,i,\tau} = \hat{m}_{3,i\tau}/\hat{m}_{2,i\tau}^{3/2}$ and $\hat{\gamma}_{2,i\tau} = \hat{m}_{4,i\tau}/\hat{m}_{2,i\tau}^2 - 3$ with $\hat{m}_{s,i\tau} = (60)^{-1}\sum_{t=1}^{60}\left(\hat{u}_{i,\tau t} - \overline{\hat{u}}_{i,\tau}\right)^s$, and $\overline{\hat{u}}_{i,\tau} = (60)^{-1}\sum_{t=1}^{60}\hat{u}_{i,\tau t}$. We ended up with 126,181 different values of $\hat{\sigma}_{ii,\tau}$, $\hat{\gamma}_{1,i,\tau}$ and $\hat{\gamma}_{2,i,\tau}$ estimated for around 476 securities over 265 different estimation windows. We discarded estimates that lied below the 2.5% and above the 97.5% quantiles to avoid the calibrated values being dominated by extreme outliers. The same procedure was applied to the estimated factor loadings, $\hat{\beta}_{\ell i.\tau}$. The means and medians of $\hat{\sigma}_{ii,\tau}$, $\hat{\gamma}_{1,i,\tau}$, $\hat{\gamma}_{2,i,\tau}$ and $\hat{\beta}_{\ell i,\tau}$ for $\ell = 1, 2, 3$, and their 2.5% and 97.5% quantiles are summarized in Table 1. As can be seen from these results there is a considerable degree of heterogeneity in estimates of the factor loadings and in the measures of deviations, skewness and kurtosis, across securities and sample periods. The details of the procedure to generate the non-normal and cross-correlated errors are described in Appendix B.

To estimate size of the tests, we set $\alpha_i = 0$ for all i. To investigate power, we generated α_i as $\alpha_i \sim IIDN(0,1)$ for $i = 1, 2, ..., N_\alpha$ with $N_\alpha = \lfloor N^{\lambda_\alpha} \rfloor$; $\alpha_i = 0$ for $i = N_\alpha + 1, N_\alpha + 2, ..., N$. We considered the values $\lambda_\alpha = 0.8, 0.9, 1.0$, but the power ended up to be very high even for $\lambda_\alpha = 0.8$. Therefore, we only report power estimates for $\lambda_\alpha = 0.80$.

All combinations of $T = 60, 100$ and $N = 50, 100, 200, 500$ (and $1,000$, $2,000$, $5,000$ for the $\hat{J}_\alpha$ test) are considered. All tests are conducted at a 5% significance level. Experiments are based on $R = 2,000$ replications.

5.2 Size and power

Table 2 reports the size and power of the GRS, $\hat{J}_\alpha$, SW_{POET}, SW_{LW} and $F_{\max}$ tests of Gungor and Luger (2016), in the case of models with three factors, under various degrees of cross-sectional error correlations, as measured by the exponent, δ_γ.

First, consider Panel A of Table 2 which deals with the case where the errors are normally distributed but cross-sectionally weakly dependent with $\delta_\gamma = 1/4$.[21] The GRS test when applicable (namely when $T > N$) being an exact test, 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 $N = 500$, the size of the $\hat{J}_\alpha$ test lies in the range 5.0% to 5.3% for different values of T. In contrast the SW_{POET} test grossly over-rejects the null hypothesis, and the degree of the over-rejection becomes more serious N increases for a given T. For example, when $T = 60$, increasing N from 50 to 500, the size of the SW_{POET} test rises from 18.3% to 53.1%. In line with the discussion in Section 3.4, the size distortion is mitigated when T increases. For $T = 60$ and $N = 50$ the size is 18.3% but it falls to 12.1% when $T = 100$ and $N = 50$. The size properties of the SW_{LW} test is very similar to those of the SW_{POET} test. The size of the $F_{\max}$ test tend to be substantially smaller than the nominal level for

[20] This may also explain why FLY test suffers from size-distortion as discussed by Bailey, Pesaran and Yamagata in Fan, Liao and Mincheva (2013), where N is allowed to increase with T fixed.

[21] In line with our theoretical findings (see Section 2), the results of cross-sectionally independent case (with $\delta_\gamma = 0$) is qualitatively similar to the case where $\delta_\gamma = 1/4$.

all combinations of N and T (this is in line with the reported results in Gungor and Luger, 2016). The rejection frequencies range between 0.1% to 0.2%. Furthermore, inconclusive test outcomes are observed more often, ranging between 2.7% and 4.6% of the outcomes.[22] The power of the $\hat{J}_\alpha$ tests is substantially higher than that of the GRS test. For example, for $T = 60$ and $N = 50$ the power of the GRS test is 15.0% as compared to 65.9% for the $\hat{J}_\alpha$ test, although both tests have similar sizes (4.6% for the GRS test and 7.4% for the $\hat{J}_\alpha$ test). This is in line with our discussion at the end of Section 2, 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. This can be seen clearly from the tabulated results. Keeping N fixed at 50, and increasing T from 60 to 100 increases the power of the GRS test from 15.0% to 69.2%, whilst the power of the $\hat{J}_\alpha$ test (for example) rises from 65.9% to 87.4%. It is interesting that even in this case (with T much larger than N) the $\hat{J}_\alpha$ test still has substantially higher power than the GRS test, with comparable type I errors. The power comparison of the SW_{POET} and SW_{LW} with other tests seem 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 SW_{POET} and SW_{LW} tests. The power of the $\hat{J}_\alpha$ test uniformly dominates that of the $F_{\max}$ test for all experiments. The low power of the $F_{\max}$ test is partially explained by the large proportions of inconclusive results. For $T = 60$, between 29.3% and 45.5% of inconclusive results are observed for different N. For $T = 100$, the proportion of inconclusive results tends to decline as N increases. For example, increasing N from 50 to 500 lowers the frequencies of inconclusive results of the $F_{\max}$ test from 39.0% to 29.1%.

Consider now the case where the errors are normally distributed and cross-sectionally relatively strongly dependent. First let us discuss the results when $\delta_\gamma = 1/2$. The $\hat{J}_\alpha$ test seems quite robust to cross-sectional error correlations, with its size falling in the range 5.1% to 6.6%. The size of the $\hat{J}_\alpha$ test for $N = 50$ and $T = 60$ is 6.4%, and its power is 53.6%, which still exceed the power of the GRS test, which is 20.7%. But, as expected, increasing T from 60 to 100 results in the power of the GRS test to rise to 84.9%, which marginally beat the power of the $\hat{J}_\alpha$ test at 82.3%. As discussed earlier, the SW_{POET} test is not justified asymptotically when $\delta_\gamma = 1/2$. For $N = 50$, increasing T from 60 to 100 does not improve the size distortion of these tests, with sizes of 21.5% and 23.3%, respectively.

When $\delta_\gamma = 3/5 > 1/2$, out of all the tests considered, only the GRS test is valid so long as $N < T$, and indeed has the correct size in such cases. However, interestingly, the size of the $\hat{J}_\alpha$ test is also close to its nominal level (at 5.5%-7.2%) even for such a high value of δ_γ. This seems to be due to the inclusion of $(N-1)\tilde{\rho}^2_{N,T}$ in the denominator of the $\hat{J}_\alpha$ statistic.

We now consider the empirically most relevant case where the errors are non-normal as well as being cross-sectionally correlated. The effects of non-normal errors on the tests are documented in Panel B of Table 2. Consider first the case where the errors are non-normal and cross-sectionally weakly correlated ($\delta_\gamma = 1/4$). We see that the size of the GRS test is hardly affected by the types of departures from Gaussianity observed in the regression residuals. The robustness of the GRS test to non-normal errors of the type encountered in practice has also been documented by Affleck-Graves and McDonald (1989). As to be expected from the theoretical discussions, the $\hat{J}_\alpha$ test is reasonably robust to non-Gaussian errors, and exhibit only a very mild tendency of over-rejecting the null hypothesis, even for relatively large N. For example, when $T = 60$, for $N = 50$, 100, 200, and 500, the sizes of the $\hat{J}_\alpha$ test are 6.5%, 6.9%, 5.9%, and 6.6%, respectively. The over-rejection of the SW_{POET} test tends to be somewhat magnified by non-normality. The effects of non-normality upon the size of the SW_{LW} is less

[22] The frequencies of inconclusive outcomes for the $F_{\max}$ test for different combinations of N and T are reported in Table M1 of the Online Supplement.

obvious. The size of the $F_{\max}$ test is again much smaller than the nominal level, but on average slightly higher than that under normal errors. For example, the average of the size of the $F_{\max}$ test for all the combinations of (N, T) is 0.14% under normal errors, but that under non-normal errors it is 0.25%. Also, on average the incidence of inconclusive outcomes for the $F_{\max}$ test is slightly higher under non-normal errors. For example, the average of the frequencies of the inconclusive outcomes for all the combinations of (N, T) is 3.7% under normal errors, but increases to 4.3% under non-normal. Under non-normal errors, the $\hat{J}_\alpha$ test continues to maintain its power superiority over the GRS and the $F_{\max}$ tests. When $\delta_\gamma = 1/2$ and $3/5$ the size of the $\hat{J}_\alpha$ test is reasonably controlled and lies in the range 6.0%-7.9%. The power comparisons discussed for the weakly cross-sectionally uncorrelated case ($\delta_\gamma = 1/4$) also carry over to the present set of experiments with the much higher degrees of error cross-sectional correlations ($\delta_\gamma = 1/2$ and $3/5$).

We also carried out additional experiments with much larger values of N, namely $N = 1,000$, $2,000$ and $5,000$, whilst keeping T at 60 and 100. We only considered the $\hat{J}_\alpha$ test for these experiments, as it is unlikely that other tests considered, given their relatively poor performance for values of $N \leq 500$, would perform better than the $\hat{J}_\alpha$ test. The results are summarised in Table 3. As can be seen, the size is satisfactorily controlled with good power properties, only showing moderate over-rejection under non-Gaussianity for $T = 60$, and for relatively strong error cross correlations. For example, for $N = 5,000$, when $T = 60$ with non-normal errors, the size of the $\hat{J}_\alpha$ test for $\delta_\gamma = 1/4, 1/2$ and $3/5$ are 7.8%, 9.5% and 9.3%, whereas, by increasing T to 100, for $N = 5,000$ the size of the test drops to 7.1%, 5.9% and 7.1%, respectively.

Finally, we investigated the robustness of the $\hat{J}_\alpha$ test against possible nonlinear dependence across security returns, discussed in the literature by Longin and Solnik (2001), and Ang, Chen and Xing (2006), among others. In the presence of nonlinear dependence, correlation of higher order moments of errors, such as $Corr\left(u_{it}^2, u_{jt}^2\right)$, could be non-zero even when u_{it} and u_{jt} are uncorrelated. Table 4 summarises the size and power of the $\hat{J}_\alpha$ test when the regression errors follow multivariate t distribution. Under this design N^{λ_c} securities' squared errors are cross-correlated, while the errors themselves are uncorrelated, which give rise to $\varphi \leq 2.5$. As can be seen, the $\hat{J}_\alpha$ test continues to perform well, giving the correct size and high power, across all of the MC designs.

5.3 Experiments with mixed factor-spatial error processes

So far we have considered error processes with a weak common factor structure but with cross-sectionally independent idiosyncratic errors. As we discussed in sub-section 3.2, our test, including estimators of the cross-correlation measure $(N-1)\,\rho_N^2$, continues to apply when the eigenvalues of variance matrix of idiosyncratic errors are bounded. Accordingly, we further investigate finite sample behaviour of the $\hat{J}_\alpha$ test under the DGPs identical to those considered for Table 2, except that spatial autoregressive component is incorporated into the error generating process. Specifically, the error correlation matrix is now given by $\mathbf{R} = \mathbf{D}_\sigma^{-1/2}\mathbf{V}\mathbf{D}_\sigma^{-1/2}$, where $\mathbf{D}_\sigma = diag\,(\sigma_{ii})$, $\mathbf{V} = (\sigma_{ij})$, $\mathbf{V} = \boldsymbol{\gamma}\boldsymbol{\gamma}' + (\mathbf{I}_N - \rho_\varepsilon \mathbf{W})^{-1}\,(\mathbf{I}_N - \rho_\varepsilon \mathbf{W}')^{-1}$ with $\boldsymbol{\gamma} = \left(\gamma_1, \gamma_2, ..., \gamma_{N_\gamma}, 0, 0, ..., 0\right)'$, γ_i for $i \leq N_\gamma = \lfloor N^{\delta_\gamma} \rfloor$ are drawn from uniform(0.7,0.9) distribution and $\gamma_i = 0$ for $i = N_\gamma + 1, N_\gamma + 2, ..., N$, ρ_ε is spatial coefficient such that $0 \leq |\rho_\varepsilon| < 1$, $\mathbf{W}' = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N)$ with $\boldsymbol{\tau}_N'\mathbf{w}_i = 1$ and its diagonal elements being all zero. Observe that when $N_\gamma = 0$, errors possess pure spatial autoregressive processes, and when $\rho_\varepsilon = 0$, the DGP becomes identical to that for the results reported earlier (in Tables 2 and 3). We have chosen the value $\rho_\varepsilon = 0.5, 0.8$ 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 $i = 1, 2, ..., N-2$ and $j = 3, 4..., N$, with $w_{1,2} = w_{N,N-1} = 1$. To investigate the importance of allowing for error correlations in the

construction of the J_α test we also consider a version the test that does not control for error cross-correlations. This version is denoted by $J_\alpha(0)$, and obtained by setting $\rho_N^2 = 0$ in $J_\alpha(\rho_N^2)$, defined by (55). Table 5 reports the results for $\rho_\varepsilon = 0.5$, both with and without the weak factor component. In the latter case $\boldsymbol{\gamma} = \mathbf{0}$, and error cross-correlations are only due to the spatial autregressive effects. As can be seen from the Panel A of the table, under Gaussianity, the size of the $\hat{J}_\alpha$ test is well controlled, with slight over-rejection when $T = 60$, which disappears when T is increased to 100. This result holds for all the values of N considered, including $N = 5,000$. In contrast, the $J_\alpha(0)$ test over-rejects the null hypothesis, around 10%, for all the combinations of N and T. This confirms that using the MT estimator of ρ_N^2 does a good job at correcting the bias of the J_α test for the spatial error correlations. The over-rejection of the test becomes more pronounced when the errors are non-Gaussian (see Panel B), but the size distortion becomes rather small for $T = 100$, even if $N > 1,000$. The results are very similar when the errors have a mixed spatial-factor models. When $\delta_\gamma = 1/4$ and $1/2$, there is no noticeable difference in the results from the case with $\boldsymbol{\gamma} = \mathbf{0}$ for both Gaussian and non-Gaussian errors. When $\delta_\gamma = 3/5$, as to be expected, we observe moderate size distortions, especially when $T = 60$ and $N \geq 1000$. The $\hat{J}_\alpha$ test continues to show good power performance for both types of error processes and for different values of δ_γ. As noted earlier, there is some loss of power δ_γ is increased. But the extent of the power loss is much smaller than those reported in Table 2.[23]

5.4 Experiments with time varying betas

We also investigated the robustness of the proposed test to random time variations in $\boldsymbol{\beta}_i$. In the case where betas are time-varying (2) can be written as

$$y_{it} = \alpha_{it} + \boldsymbol{\beta}_{it}'\mathbf{f}_t + u_{it}, \tag{65}$$

where $\alpha_{it} = \nu + \boldsymbol{\beta}_{it}'(\boldsymbol{\lambda} - \boldsymbol{\mu}_f)$. Suppose that time variations in $\boldsymbol{\beta}_{it}$ can be modelled by the following random coefficient model[24]

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_i + \boldsymbol{\upsilon}_{it}, \tag{66}$$

where $E(\boldsymbol{\beta}_{it}) = \boldsymbol{\beta}_i$, and $\boldsymbol{\upsilon}_{it} = (\upsilon_{1,it}, \upsilon_{2,it}, ..., \upsilon_{m,it})' \sim IID(\mathbf{0}, \boldsymbol{\Omega}_{\upsilon,ii})$ over i and t, and distributed independently of $u_{jt'}$ and $\mathbf{f}_s$ for all i, j, t, t', and s. Using (66) we now have

$$y_{it} = \alpha_i + \boldsymbol{\beta}_i'\mathbf{f}_t + \mathring{u}_{it}, \tag{67}$$

where $\mathring{u}_{it} = \boldsymbol{\upsilon}_{it}'\tilde{\mathbf{f}}_t + u_{it}$, and $\tilde{\mathbf{f}}_t = \mathbf{f}_t - \boldsymbol{\mu}_f + \boldsymbol{\lambda}$. Suppose that $\mathbf{f}_t$ is a stationary process with mean $\boldsymbol{\mu}_f$ and variance $\boldsymbol{\Omega}_f$. Then for each i, $\mathring{u}_{it}$ is serially independent with zero means and constant unconditional variances, namely

$$E(\mathring{u}_{it}) = 0,\ E(\mathring{u}_{it}\mathring{u}_{jt}) = \begin{cases} \mathring{\sigma}_{ii} = \sigma_{\upsilon,ii} + \sigma_{ii} \text{ for } i = j \\ \mathring{\sigma}_{ij} = \sigma_{ij} \text{ for } i \neq j, \end{cases}$$

where $\sigma_{\upsilon,ii} = E\left(\tilde{\mathbf{f}}_t'\boldsymbol{\upsilon}_{it}\boldsymbol{\upsilon}_{it}'\tilde{\mathbf{f}}_t\right) = \boldsymbol{\lambda}'\boldsymbol{\Omega}_{\upsilon,ii}\boldsymbol{\lambda} + Tr(\boldsymbol{\Omega}_f\boldsymbol{\Omega}_{\upsilon,ii})$. Hence,

$$Corr(\mathring{u}_{it}, \mathring{u}_{jt}) = \mathring{\rho}_{ij} = \frac{\rho_{ij}}{[1 + (\sigma_{\upsilon,ii}/\sigma_{ii})]^{1/2}[1 + (\sigma_{\upsilon,jj}/\sigma_{jj})]^{1/2}}, \text{ for } i \neq j, \tag{68}$$

[23] The results for $\rho_\varepsilon = 0.8$ are qualitatively similar to the results for $\rho_\varepsilon = 0.5$, which are summarised in Table M5 in the Online Supplement (available upon request).

[24] This set up is sufficiently general and accommodates a wide class of random coefficient models considered in the literature, but it rules out persistent and systematic time variations in betas. In practice, as with the empirical application discussed in Section 6 below, one can deal with such persistent time variations by considering tests of LFPM over relatively short time periods, which requires the test to apply in cases where N is much larger than T.

and it readily follows that $\left|\mathring{\rho}_{ij}\right| \leq \left|\rho_{ij}\right|$, and the presence of random variations in betas in fact reduces the degree of error cross sectional dependence. Therefore, the composite errors, $\mathring{\mathrm{u}}_{it}$, implied by the time-varying betas satisfy the sparsity conditions (30) and (31). However, the theoretical analysis become further complicated due to the fact that $\mathring{\mathrm{u}}_{it}$ are now conditionally heteroskedastic, namely $Var\left(\mathring{\mathrm{u}}_{it}\left|\tilde{\mathbf{f}}_t\right.\right) = \tilde{\mathbf{f}}_t'\mathbf{\Omega}_{\upsilon,ii}\tilde{\mathbf{f}}_t + \sigma_{ii}$. Nevertheless, our preliminary analysis suggests that the proposed test continues to be applicable in this case so long as $\mathbf{f}_t$ is stationary with bounded support and the in-sample mean of $\mathbf{f}_t$ is sufficiently small. A formal proof of this conjecture is beyond the scope of the present paper. But in support of our conjecture we provide additional Monte Carlo evidence in Table 6, where we present finite sample behaviour of the $\hat{J}_\alpha$ test under the DGPs identical to those considered for Table 5, except that betas are now generated to be time varying. Specifically, we generated betas as $\beta_{\ell it} = \beta_{\ell i} + \upsilon_{\ell it}$ with $\upsilon_{\ell it} \sim IIDN\left(0,1\right)$, and set $y_{it} = \alpha_i + \sum_{\ell=1}^{3}\beta_{\ell it}f_{\ell t} + u_{it}$, $i = 1,2,..,N; t = 1,2,...,T$. The results summarized in Table 6 are qualitatively similar to those in Table 5, suggesting that allowing for random time variations in betas do not adversely impact the small sample properties of the $\hat{J}_\alpha$ test, and if anything tend to correct the slight over-rejection of the test in the case of models with time-invariant betas, most likely due to the fact that random-variations in betas reduce the degree of error cross-correlations.

6 Empirical Application

6.1 Data description

We consider the application of our proposed $\hat{J}_\alpha$ test to the securities in the Standard & Poor 500 (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 Standard and Poor 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. 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.

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 June 2015. The monthly return of security i for month t is computed as $r_{it} = 100(P_{it} - P_{i,t-1})/P_{i,t-1} + DY_{it}/12$, where P_{it} is the end of the month price of the security and DY_{it} is the per cent per annum dividend yield on the security. Note that index i depends on the month of 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 US treasury bill rate is chosen as the risk-free rate (r_{ft}), the value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) is used as a proxy for market return (r_{mt}), the average return on the three small portfolios minus the average return on the three big portfolios (SMB_t), and the average return on two value portfolios minus the average return on two growth portfolio (HML_t). SMB and HML are based on the stocks listed on the NYSE, AMEX and NASDAQ. All data are measured in percent per month. See Appendix C for further details.

6.2 Month end test results (September 1989 - June 2015)

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 to June 2015.[25] 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 persistence or systematic time variations in betas, we select a relatively short time period of 60 months. Recall that the experimental results reported above show that our test is robust to random time variations in betas. Accordingly, we estimated the CAPM regressions

$$r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{u}_{i,\tau t}, \tag{69}$$

and the Fama-French (FF) three factor regressions,

$$r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{\beta}_{2,i\tau} SMB_{t\tau} + \hat{\beta}_{3i} HML_{t\tau} + \hat{u}_{i,\tau t}, \tag{70}$$

for $t = 1, 2, ..., 60$, $i = 1, 2, ..., N_\tau$, and the month ends, τ =September 1989,....,June 2015. All securities in the S&P 500 index are included except those with less than sixty months of observations and/or with five consecutive zeros in the middle of sample periods.

Table 7 reports summary statistics for p-values of the $\hat{J}_\alpha$ test, cross-sectional averages of measures of departure from non-normality and average pair-wise correlations of residuals from CAPM and FF regressions of securities in the S&P 500 index using five year estimation windows (sixty months) at the end of the months of September 1989 to June 2015. The results confirm important departures from normality in the residuals. The extent of the departures are particularly pronounced in the case of kurtosis measures where $\gamma_2 = 0$ is rejected in 26-29% of the samples under consideration. Three measures of average pair-wise correlations of residuals are reported in the last columns of the table, which indicate minor degrees of cross-sectional correlations. The residuals from FF regressions tend to be cross-sectionally less correlated than those of CAPM regressions. The p-values range from 0 to 1, with a mean and median of 0.52 and 0.63 for the CAPM model, and 0.46 and 0.50 for the FF model, suggesting important time variations in the degree of market efficiency.

Figure 1 provides plots of the evolution of p-values of the $\hat{J}_\alpha$ test based on CAPM and FF regressions at the end of the months of September 1989 to June 2015. The months at which the null of market efficiency is rejected at the 5% level based on both CAPM and FF regressions are August 1998, November 1998-February 1999, August 2007-March 2009 and November 2013-June 2015 (the last data point). The period around August 1998 and December 1998-February 1999 coincide with the Russian financial crisis (during August -September 1998) and the subsequent collapse of Long-Term Capital Management. The period August 2007-March 2009 matches the recent global financial crisis. November 2013-June 2015 corresponds to series of exogenous economic and financial shocks - unrest around Russian, started by the Ukraine crisis, then the negative oil price shock started around June 2014. In general, the $\hat{J}_\alpha$ test tends to result in rejection of the null of market efficiency, in the Sharpe-Lintner sense, during periods of major financial disruptions.

6.3 Long/short equity returns and p-values of the $\hat{J}_\alpha$ test

As the test results in Figure 1 clearly show important variations in the p-values of the $\hat{J}_\alpha$ test over time, it would be interesting to see if such variations are related to the performance of trading strategies. There are many trading strategies that are designed to exploit non-zero α's in selection of securities. A prominent example is the long/short equity strategy where securities are ordered by their predicted returns, from the most positive to the most negative.

[25] In all the empirical applications $T < N$, and the GRS test can not 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.

The investor then goes long on securities with positive predicted returns and goes short on securities with negative return predictions. There are many variations in the way that this strategy is implemented which need not concern us here. What we are interested in is to see if there are any relationships between the return on long/short (L/S) strategies and the evidence of market inefficiency as measured by estimated p-values. In time periods where $\alpha_i = 0$ for all i, the L/S strategy is unlikely to perform better than the market return, and could do even worse if one allows for transaction costs and management fees. But we would expect a higher return on the L/S strategies relative to the market if there are positive and negative alphas that the investor can identify and exploit. Therefore, *a priori* we would expect an inverse relationship between p-values and returns on L/S strategies relative to the market.

For return on L/S strategies we used Credit Suisse Long/Short Equity Hedge Fund Index that are available monthly from December 1993. This is a subset of the Credit Suisse Hedge Fund Index and provides the *aggregate* performance of long/short equity funds, and as such is not subject to a selection bias. We denote the monthly return on this index by r_{ht} and consider the relationship between $\tilde{r}_{ht} = r_{h\tau} - r_t$, where r_t is the return on S&P 500 index, and monthly p-values of the $\hat{J}_\alpha$ tests, which we denote by $\hat{\pi}_t$.[26] The p-values needed for this purpose are already reported in Figure 1. Given the considerable volatility of return data, in Figures 2 and 3 we plot twelve-month moving averages of returns and p-values computed as $\tilde{r}_{ht}(12) = \frac{1}{12}\sum_{j=0}^{11}\tilde{r}_{h,t-j}$, and $\hat{\pi}_t(12) = \frac{1}{12}\sum_{j=0}^{11}\hat{\pi}_{t-j}$, respectively. Figure 2 depicts the relationship for p-values computed using the CAPM regressions, and Figure 3 shows the relationship for the p-values computed using the FF regressions. There is a significant negative relationship between the p-values and the excess returns. The value of sample correlation between $\tilde{r}_{ht}(12)$ and CAPM p-value is -0.28 (s.e. 0.061), giving a t-ratio of -4.6, strongly rejecting the null of zero-correlation.[27] The value of sample correlation between $\tilde{r}_{ht}(12)$ and FF p-value is almost identical, giving -0.27 (s.e. 0.061) and a t-ratio of -4.4.

7 Conclusion

In this paper we propose a simple test of Linear Factor Pricing Models (LFPM), 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 standardised Wald tests based on an adaptive thresholding estimators of $\mathbf{V}$, which is considered by Fan, Liao and Yao (2015). It allows N to be much larger than T, as compared to 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 correlations, volatilities, and other distributional features of the residuals of Fama-French three factor regressions of individual securities in the Standard & Poor 500 index, we show that the proposed $\hat{J}_\alpha$ test performs well even when N is much larger than T, and outperform other existing tests such as the tests of Fan et al (2015) and Gungor and Luger (2016) test. 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 a 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 - June 2015 clearly illustrates the utility of the proposed test. Statistically significant evidence against Sharpe-Lintner CAPM and Fama-French three factor models is found during periods of financial crisis and market disruptions. Furthermore, a significant negative correlation is found between a twelve-month

[26] See Appendix C for further details and the source of the L/S equity hedge fund returns.

[27] The standard error of the sample correlation $\hat{\rho}_T$ is computed as $[(1-\hat{\rho}_T^2)/(T-2)]^{1/2}$.

moving average p-values of the $\hat{J}_\alpha$ test and excess returns of long/short equity strategies over the period November 1994 to June 2015.

Table 1: Summary statistics of the estimates used in the Monte Carlo simulations

This table reports the summary statistics for estimated $\beta's$, variance, skewness and kurtosis measures of residuals from Fama-French (FF) three factor regressions, estimated for all securities in the S&P 500 index with at least sixty months of return data using rolling estimation windows of sixty months, over the period September 1989 to September 2011. $\hat{\beta}_{i\tau}$ is estimated using the FF regressions: $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{\beta}_{2,i\tau} SMB_{t\tau} + \hat{\beta}_{3i} HML_{t\tau} + \hat{u}_{i,\tau t}$, for $i = 1, 2, ..., N_\tau$, and $t = 1, 2, ..., 60$, where N_τ denotes the number of securities available at the estimation windows $\tau =$ September 1989,..., September 2011. $\hat{\sigma}_{ii,\tau} = \hat{m}_{2,i\tau}$ $\hat{\gamma}_{1,i\tau} = \hat{m}_{3,i\tau}/\hat{m}_{2,i\tau}^{3/2}$ and $\hat{\gamma}_{2,i\tau} = \hat{m}_{4,i\tau}/\hat{m}_{2,i\tau}^{2} - 3$, which are computed using the FF residuals, where $\hat{m}_{s,i\tau} = (60)^{-1}\sum_{t=1}^{60}\left(\hat{u}_{i,\tau t} - \bar{\hat{u}}_{i,\tau}\right)^s$ and $\bar{\hat{u}}_{i,\tau} = (60)^{-1}\sum_{t=1}^{60}\hat{u}_{i,\tau t}$, for $s = 2, 3, 4$. All securities in the S&P 500 index are included except those with less than sixty months of observations and/or with five consecutive zeros in the middle of sample periods. Under normal errors we set $\gamma_{1,i} = \gamma_{2,i} = 0$.

	Mean	Median	2.5% Quantile	97.5% Quantile
$\hat{\sigma}_{ii,\tau}$	65.60	44.72	12.81	249.89
$\hat{\gamma}_{1,i,\tau}$	0.18	0.14	-0.89	1.46
$\hat{\gamma}_{2,i,\tau}$	1.00	0.38	-0.71	6.74
$\hat{\beta}_{1,i,\tau}$	1.10	0.51	0.24	2.26
$\hat{\beta}_{2,i,\tau}$	0.10	0.04	-0.91	1.47
$\hat{\beta}_{3,i,\tau}$	0.20	0.24	-1.55	1.72

Table 2: Size and power of GRS, $\hat{J}_\alpha$, SW_{POET}, SW_{LW} and $F_{\max}$ tests

This table summarises the size and power of GRS, $\hat{J}_\alpha$, SW_{POET}, SW_{LW} and $F_{\max}$ tests of $\alpha_i = 0$, for $i = 1, 2, ..., N$, in the case of three-factor models. The observations are generated as $y_{it} = \alpha_i + \sum_{\ell=1}^{3} \beta_{\ell i} f_{\ell t} + u_{it}$, $i = 1, 2, .., N; t = 1, 2, ..., T$, $f_{\ell t} = \mu_{f\ell} + \rho_{f\ell} f_{\ell,t-1} + \sqrt{h_{\ell t}}\ \zeta_{\ell t}$, $h_{\ell t} = \mu_{h\ell} + \rho_{1h\ell} h_{\ell,t-1} + \rho_{2h\ell} \zeta^2_{\ell,t-1}$, $\zeta_{\ell t} \sim IIDN(0,1)$, $t = -49, ..., T$ with $f_{\ell,-50} = 0$ and $h_{\ell,-50} = 0$, $\ell = 1, 2, 3$, $\mu_{f\ell} = 0.53, 0.19, 0.19$, $\rho_{f\ell} = 0.06, 0.19, 0.05$, $\mu_{h\ell} = 0.89, 0.62, 0.80$, $\rho_{1h\ell} = 0.85, 0.74, 0.76$, $\rho_{2h\ell} = 0.11, 0.19, 0.15$, for $\ell = 1, 2, 3$, respectively. For the size of the test, $\alpha_i = 0$ for all i, and for the power of the test, $\alpha_i \sim IIDN(0,1)$ for $i = 1, 2, ..., N_\alpha$ with $N_\alpha = \lfloor N^{\lambda_\alpha} \rfloor$, $\lambda_\alpha = 0.8$, otherwise $\alpha_i = 0$, where $\lfloor A \rfloor$ is the largest integer part of A. The idiosyncratic errors, $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$, are generated as $\mathbf{u}_t = \mathbf{Q}\boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$, and $\mathbf{Q} = \mathbf{D}^{1/2}\mathbf{P}$ with $\mathbf{D} = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})'$, $\sigma_{ii} = Var(\sigma_{ii})$, and $\mathbf{P}$ being a Cholesky factor of correlation matrix of $\mathbf{u}_t$, $\mathbf{R} = \mathbf{I}_N + \mathbf{b}\mathbf{b}' - \check{\mathbf{B}}^2$, which is an $N \times N$ matrix used to calibrate the cross correlation of returns, where $\mathbf{b} = (b_1, b_2, ..., b_N)'$, $\check{\mathbf{B}} = diag(\mathbf{b})$. The first and the last N_γ $(< N)$ elements of $\mathbf{b}$ are generated as $Uniform(0.7, 0.9)$, and the remaining middle elements are set to 0. We set $N_\gamma = \lfloor N^{\delta_\gamma} \rfloor$. We consider the values $\delta_\gamma = 1/4, 1/2$ and $3/5$. For the case of non-normal errors, u_{it} are generated following steps 1-4 of the procedure in Appendix B, using skewness and kurtosis measures, $\gamma_{1,i}$ and $\gamma_{2,i}$. σ_i^2, $\gamma_{1,i}$, $\gamma_{2,i}$ and $\beta_{\ell i}$ for $\ell = 1, 2, 3$, are randomly drawn from their respective empirical distributions, see Subsection 5.1 and Appendix B for details. GRS is the F test due to Gibbons et al. (1989) which is distributed as $F_{N,T-N-m}$, and is applicable when $T > N + m + 1$. N/A signifies that the GRS statistic can not be computed. $\hat{J}_\alpha$ is the propose large N test which is robust to non-Gaussian errors and cross-sectional correlations; SW_{POET} and SW_{LW} are the tests based on the POET estimator of Fan et al. (2013), $\hat{\boldsymbol{V}}^{-1}_{POET}$, and Ledoit-Wolf (2004) shrinkage estimator, $\hat{\boldsymbol{V}}^{-1}_{LW}$, as estimates of $\boldsymbol{V}^{-1}$ in (17). $F_{\max}$ is the bounds test of Gungor and Luger (2016), with frequencies of inconclusive test outcomes reported in Table M1 in the Online Supplement available on request. Values of the $\hat{J}_\alpha$, SW_{POET} and SW_{LW} test statistics are compared to a positive one-sided critical value of the standard normal distribution. All tests are conducted at the 5% significance level. Experiments are based on 2,000 replications.

Panel A: Normal Errors													
		$\delta_\gamma = 1/4$				$\delta_\gamma = 1/2$				$\delta_\gamma = 3/5$			
	(T,N)	50	100	200	500	50	100	200	500	50	100	200	500
Size: $\alpha_i = 0$ for all i													
GRS	60	4.6	N/A	N/A	N/A	5.3	N/A	N/A	N/A	5.4	N/A	N/A	N/A
	100	5.8	N/A	N/A	N/A	5.3	N/A	N/A	N/A	5.5	N/A	N/A	N/A
$\hat{J}_\alpha$	60	7.4	5.3	6.0	5.0	6.4	5.9	5.6	6.1	6.0	5.5	6.7	7.2
	100	6.6	5.2	5.5	5.3	6.1	6.6	5.1	5.3	6.7	6.3	5.6	5.8
SW_{POET}	60	18.3	26.2	34.0	53.1	21.5	25.0	30.4	48.6	21.4	23.1	30.6	45.2
	100	12.1	14.3	20.4	30.3	23.3	18.7	20.9	27.5	28.9	20.8	24.8	29.0
SW_{LW}	60	17.7	23.3	33.9	56.5	22.3	32.9	46.8	67.6	28.5	50.7	75.0	93.0
	100	12.7	16.7	21.6	31.3	16.9	24.1	37.3	50.8	21.6	43.9	76.1	94.0
$F_{\max}$	60	0.2	0.1	0.2	0.1	0.3	0.1	0.3	0.2	0.1	0.1	0.1	0.1
	100	0.2	0.1	0.1	0.1	0.0	0.2	0.1	0.1	0.2	0.2	0.1	0.2
Power: $\alpha_i \sim IIDN(0,1)$ for $i = 1, 2, ..., N_\alpha$ with $N_\alpha = \lfloor N^{\lambda_\alpha} \rfloor$, $\lambda_\alpha = 0.8$, otherwise $\alpha_i = 0$													
GRS	60	15.0	N/A	N/A	N/A	20.7	N/A	N/A	N/A	24.2	N/A	N/A	N/A
	100	69.2	N/A	N/A	N/A	84.9	N/A	N/A	N/A	87.6	N/A	N/A	N/A
$\hat{J}_\alpha$	60	65.9	80.2	93.2	98.8	53.6	67.2	84.1	96.4	42.2	53.9	66.3	82.1
	100	87.4	97.4	99.9	100.0	82.3	93.7	98.7	100.0	72.2	86.4	95.0	99.6
SW_{POET}	60	81.9	95.2	99.3	100.0	80.3	91.1	98.6	99.9	77.5	88.3	97.0	99.9
	100	93.5	99.3	100.0	100.0	97.3	98.9	99.9	100.0	95.8	97.9	99.8	100.0
SW_{LW}	60	68.8	82.7	93.5	99.7	77.4	89.7	96.7	99.7	86.1	96.4	99.9	100.0
	100	86.2	95.1	99.5	100.0	94.3	98.5	99.8	100.0	96.8	99.8	100.0	100.0
$F_{\max}$	60	11.5	12.5	17.6	22.2	12.5	15.1	16.6	22.7	11.6	13.8	17.8	24.3
	100	29.5	41.3	51.4	67.4	32.2	41.5	51.4	66.4	30.4	40.9	51.9	66.2

Table 2 —*Continued*

Panel B: Non-normal Errors													
		$\delta_\gamma = 1/4$				$\delta_\gamma = 1/2$				$\delta_\gamma = 3/5$			
	(T,N)	50	100	200	500	50	100	200	500	50	100	200	500
Size: $\alpha_i = 0$ for all i													
GRS	60	5.5	N/A	N/A	N/A	5.4	N/A	N/A	N/A	5.2	N/A	N/A	N/A
	100	4.4	N/A	N/A	N/A	5.4	N/A	N/A	N/A	5.3	N/A	N/A	N/A
$\hat{J}_\alpha$	60	6.5	6.9	5.9	6.6	6.0	6.9	6.5	6.3	6.3	7.9	6.4	7.6
	100	5.6	6.7	6.4	7.2	6.6	6.2	7.0	7.8	7.8	7.3	6.7	6.9
SW_{POET}	60	18.7	27.2	37.8	56.8	21.6	26.5	34.1	51.6	22.8	27.5	32.2	48.0
	100	11.7	17.2	21.6	33.4	30.7	20.5	22.8	31.7	30.6	21.3	23.8	31.2
SW_{LW}	60	17.5	23.2	33.2	56.0	21.2	34.8	47.2	69.3	27.9	49.2	77.2	93.4
	100	12.1	17.2	21.6	31.0	15.6	26.3	37.3	53.3	21.4	43.6	78.1	94.6
$F_{\max}$	60	0.3	0.2	0.4	0.2	0.2	0.4	0.1	0.1	0.2	0.2	0.2	0.1
	100	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.1
Power: $\alpha_i \sim IIDN(0,1)$ for $i = 1, 2, ..., N_\alpha$ with $N_\alpha = \lfloor N^{\lambda_\alpha} \rfloor$, $\lambda_\alpha = 0.8$, otherwise $\alpha_i = 0$													
GRS	60	15.7	N/A	N/A	N/A	19.0	N/A	N/A	N/A	23.1	N/A	N/A	N/A
	100	70.9	N/A	N/A	N/A	83.7	N/A	N/A	N/A	88.3	N/A	N/A	N/A
$\hat{J}_\alpha$	60	68.4	82.4	93.6	99.5	54.2	69.2	84.4	97.6	42.6	57.1	66.7	84.6
	100	88.7	96.7	99.8	100.0	82.2	93.3	99.0	100.0	73.4	86.0	95.3	99.7
SW_{POET}	60	83.8	95.2	99.4	100.0	80.3	92.1	98.7	99.9	74.7	89.1	97.6	100.0
	100	93.6	99.4	100.0	100.0	96.7	98.5	99.9	100.0	93.9	98.2	99.9	100.0
SW_{LW}	60	70.4	81.9	93.8	99.7	77.4	90.4	97.1	99.9	84.9	96.1	99.7	100.0
	100	87.0	94.8	99.0	99.9	93.6	98.6	99.8	100.0	97.3	99.7	100.0	100.0
$F_{\max}$	60	12.1	13.8	19.0	23.9	12.0	15.2	18.8	23.7	12.2	13.1	18.3	23.4
	100	31.8	41.4	51.6	67.7	30.9	40.2	53.0	68.5	30.3	40.6	51.8	64.8

Table 3: Size and power of the $\hat{J}_\alpha$ test for $N = 1,000, 2,000$ and $5,000$ in the case of models with three factors

This table summarises the size and power of the $\hat{J}_\alpha$ test in the case of models with three factors with focus on large values of N. The data is generated as described in the notes to Table 2.

	$\delta_\gamma = 1/4$			$\delta_\gamma = 1/2$			$\delta_\gamma = 3/5$		
(T, N)	$1,000$	$2,000$	$5,000$	$1,000$	$2,000$	$5,000$	$1,000$	$2,000$	$5,000$
Panel A: Normal Errors									
Size: $\alpha_i = 0$ for all i									
$T = 60$	5.9	5.3	6.3	5.9	6.2	6.3	6.5	7.0	8.1
$T = 100$	4.8	4.8	4.3	7.3	6.4	6.8	7.0	7.0	7.2
Power: $\alpha_i \sim IIDN(0,1)$ for $i = 1, 2, ..., N_\alpha$ with $N_\alpha = \lfloor N^{\lambda_\alpha} \rfloor$, $\lambda_\alpha = 0.8$, otherwise $\alpha_i = 0$									
$T = 60$	100.0	100.0	100.0	99.2	100.0	100.0	92.6	98.5	100.0
$T = 100$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Panel B: Non-normal Errors									
Size: $\alpha_i = 0$ for all i									
$T = 60$	6.3	7.6	7.8	7.7	8.4	9.5	7.5	8.6	9.3
$T = 100$	4.8	6.0	7.1	6.9	7.0	5.9	8.1	7.0	7.1
Power: $\alpha_i \sim IIDN(0,1)$ for $i = 1, 2, ..., N_\alpha$ with $N_\alpha = \lfloor N^{\lambda_\alpha} \rfloor$, $\lambda_\alpha = 0.8$, otherwise $\alpha_i = 0$									
$T = 60$	100.0	100.0	100.0	99.6	100.0	100.0	94.6	98.6	99.9
$T = 100$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 4: Size of the $\hat{J}_\alpha$ test **when** u_{it}^2 **and** u_{jt}^2 **are correlated for the pair** (i,j) **of** $\rho_{ij}=0$ **with multivariate t-distributed errors**

This table summarises the size and power of $\hat{J}_\alpha$ test when the errors follow multivariate t distribution with g degrees of freedom, and ε_{it}^2 and ε_{jt}^2 are correlated even when ε_{it} and ε_{jt} are uncorrelated. Specifically, the data is generated as described in the notes to Table 2 except that only the first $N_\gamma = \lfloor N^{\delta_\gamma} \rfloor$ errors are cross-correlated, where $\lfloor A \rfloor$ is the largest integer part of A, and $\varepsilon_{it} \sim IID\ t_g/\sqrt{(g/g-2)}$ for $i=1,2,...,N-N_c$ and all t with $N_c = \lfloor N^{\lambda_c} \rfloor$, $\varepsilon_{it} \sim \sqrt{(g-2)/\chi_{g,t}^2}\, z_{it}$, where $z_{it} \sim IIDN(0,1)$ and $\chi_{g,t}^2$ is a chi-squared random variate with g degrees of freedom, distributed independently of z_{it} for $i=N-N_c+1,...,N$ and all t. We set $g=8$, which yields $E\left(\varepsilon_{it}^4\right)-3=1.5$ so that $\varphi \leq 1+\left|\gamma_{2,\varepsilon}\right|=2.5$, and use $f(N)=N-1$ (or $\delta=1$). See also the notes to Table 2.

	$\delta_\gamma = 1/4$							$\delta_\gamma = 1/2$							$\delta_\gamma = 3/5$						
(T,N)	50	100	200	500	1,000	2,000	5,000	50	100	200	500	1,000	2,000	5,000	50	100	200	500	1,000	2,000	5,000
Panel A: $\lambda_c = 1/2$																					
Size: $\alpha_i = 0$ for all i																					
60	5.4	5.9	6.7	4.9	5.1	5.4	6.0	6.9	5.8	4.7	6.3	5.9	5.9	5.2	6.8	5.2	6.3	4.6	6.6	6.6	7.4
100	6.5	5.2	5.3	5.4	5.5	4.6	5.2	5.8	6.0	5.3	5.0	5.7	6.0	5.4	5.7	5.7	6.1	5.7	6.3	5.5	6.2
Power: $\alpha_i \sim IIDN(0,1)$ for $i=1,2,...,N_\alpha$ with $N_\alpha = \lfloor N^{\lambda_\alpha} \rfloor$, $\lambda_\alpha = 0.8$, otherwise $\alpha_i = 0$																					
60	66.3	81.8	92.7	99.6	100.0	100.0	100.0	58.8	75.0	89.2	98.3	99.9	100.0	100.0	52.7	67.0	80.3	93.1	98.7	99.8	100.0
100	89.3	97.1	99.7	100.0	100.0	100.0	100.0	85.9	96.5	99.7	100.0	100.0	100.0	100.0	81.4	92.2	98.1	100.0	100.0	100.0	100.0
Panel B: $\lambda_c = 3/5$																					
Size: $\alpha_i = 0$ for all i																					
60	5.6	5.6	6.1	4.7	4.9	4.8	5.5	6.2	5.9	5.5	5.2	5.4	5.4	5.7	6.2	6.5	6.4	5.6	6.5	6.1	6.8
100	5.4	5.2	5.9	5.9	4.9	4.7	5.2	6.1	5.4	6.1	5.4	6.2	5.3	5.6	5.8	6.0	5.6	5.7	6.0	6.0	6.3
Power: $\alpha_i \sim IIDN(0,1)$ for $i=1,2,...,N_\alpha$ with $N_\alpha = \lfloor N^{\lambda_\alpha} \rfloor$, $\lambda_\alpha = 0.8$, otherwise $\alpha_i = 0$																					
60	66.0	81.0	93.8	99.5	100.0	100.0	100.0	61.5	72.8	88.6	98.3	99.9	100.0	100.0	53.1	67.8	79.3	94.1	98.4	99.9	100.0
100	88.7	97.8	99.6	100.0	100.0	100.0	100.0	85.1	96.4	99.4	100.0	100.0	100.0	100.0	80.4	92.4	98.1	99.9	100.0	100.0	100.0
Panel C: $\lambda_c = 4/5$																					
Size: $\alpha_i = 0$ for all i																					
60	5.2	6.1	5.8	6.8	5.8	7.7	9.1	6.0	6.6	5.2	6.3	6.3	7.0	7.9	6.6	6.6	6.1	6.3	6.1	6.8	6.9
100	6.6	4.9	5.9	5.3	6.3	6.1	7.4	6.8	6.6	7.0	4.8	6.2	6.4	6.2	7.3	6.9	5.3	5.7	6.2	6.8	6.7
Power: $\alpha_i \sim IIDN(0,1)$ for $i=1,2,...,N_\alpha$ with $N_\alpha = \lfloor N^{\lambda_\alpha} \rfloor$, $\lambda_\alpha = 0.8$, otherwise $\alpha_i = 0$																					
60	67.1	81.3	91.6	98.8	99.9	100.0	100.0	60.1	74.5	88.1	97.4	99.8	100.0	100.0	53.3	66.0	78.9	93.8	98.2	99.7	100.0
100	88.6	97.3	99.7	100.0	100.0	100.0	100.0	84.6	95.9	99.4	100.0	100.0	100.0	100.0	80.8	92.3	98.1	100.0	100.0	100.0	100.0

Table 5: Size and power of $\hat{J}_\alpha$ test with mixed spatial-factor models with the value of spatial parameter $\rho_\varepsilon = 0.5$

Data is generated using the same set up as in Table 2, except that an spatial autoregressive component is added to the error generating process. Specifically, the error correlation matrix is given by $\mathbf{R} = \mathbf{D}_\sigma^{-1/2}\mathbf{V}\mathbf{D}_\sigma^{-1/2}$, where $\mathbf{D}_\sigma = diag\left(\sigma_{ii}\right)$, $\mathbf{V} = (\sigma_{ij})$, $\mathbf{V} = \boldsymbol{\gamma}\boldsymbol{\gamma}' + \left(\mathbf{I}_N - \rho_\varepsilon \mathbf{W}\right)^{-1}\left(\mathbf{I}_N - \rho_\varepsilon \mathbf{W}'\right)^{-1}$ with $\boldsymbol{\gamma} = \left(\gamma_1, \gamma_2, ..., \gamma_{N_\gamma}, 0, 0, ..., 0\right)'$, γ_i for $i \leq N_\gamma = \lfloor N^{\delta_\gamma} \rfloor$ are drawn from uniform(0.7,0.9) distribution and $\gamma_i = 0$ for $i = N_\gamma+1, N_\gamma+2, ..., N$, ρ_ε is spatial coefficient such that $0 \leq |\rho_\varepsilon| < 1$, $\mathbf{W}' = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N)$ with $\boldsymbol{\tau}_N'\mathbf{w}_i = 1$ and its diagonal elements being all zero. Observe that when $N_\gamma = 0$, errors possess pure spatial autoregressive processes, and when $\rho_\varepsilon = 0$ the DGP becomes identical to that for the results reported in Table 2. We have chosen the value $\rho_\varepsilon = 0.5$ 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 $i = 1, 2, ..., N-2$ and $j = 3, 4..., N$, with $w_{1,2} = w_{N,N-1} = 1$. For the purpose of comparison to $\hat{J}_\alpha$, we also provide results for $J_\alpha(0)$ test defined by (55) with $\rho_N^2 = 0$, which does not control for error cross-correlations, evaluated at. Panel A of the table reports size and power of $\hat{J}_\alpha$ and $J_\alpha(0)$ tests with normal errors, and Panel B reports size and power with non-normal errors. All tests are conducted at the 5% significance level. Experiments are based on 2,000 replications. See also the notes to Table 2.

Panel A: Normal Errors with $\rho_\varepsilon = 0.5$															
		Size							Power						
	(T,N)	50	100	200	500	1000	2000	5000	50	100	200	500	1000	2000	5000
Pure spatial models ($\boldsymbol{\gamma} = \mathbf{0}$)															
$\hat{J}_\alpha$	60	6.8	7.2	7.6	7.7	8.0	6.7	8.9	55.6	72.1	87.0	97.6	99.7	100.0	100.0
	100	6.8	6.8	6.1	5.9	5.8	5.8	5.1	82.0	94.4	99.0	100.0	100.0	100.0	100.0
$J_\alpha(0)$	60	10.1	10.5	10.5	11.1	10.8	8.9	10.6	63.9	78.4	91.4	98.3	99.8	100.0	100.0
	100	10.9	10.7	9.6	9.9	9.4	9.0	9.7	88.1	96.6	99.4	100.0	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 1/4$)															
$\hat{J}_\alpha$	60	5.9	5.6	6.2	6.3	6.5	7.0	7.9	57.6	70.0	86.0	97.8	99.5	100.0	100.0
	100	6.4	6.4	6.8	6.7	4.8	5.8	5.9	82.6	93.6	99.1	100.0	100.0	100.0	100.0
$J_\alpha(0)$	60	9.5	9.7	9.8	9.3	9.2	9.5	9.3	66.4	77.6	89.6	98.6	99.7	100.0	100.0
	100	10.5	12.1	10.9	10.4	8.9	9.6	9.9	87.5	96.2	99.6	100.0	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 1/2$)															
$\hat{J}_\alpha$	60	6.9	7.0	7.3	7.5	6.8	7.2	7.0	55.1	70.6	86.4	96.7	99.7	99.9	100.0
	100	6.3	6.5	6.7	7.1	5.4	6.9	6.2	82.3	93.9	99.1	100.0	100.0	100.0	100.0
$J_\alpha(0)$	60	10.9	11.1	10.5	10.7	10.4	10.0	9.1	65.1	79.2	90.7	98.0	99.8	100.0	100.0
	100	10.5	10.7	11.0	11.5	9.4	11.5	10.6	88.1	96.2	99.6	100.0	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 3/5$)															
$\hat{J}_\alpha$	60	6.8	7.5	6.2	8.4	8.8	9.7	9.8	53.5	71.1	84.9	96.4	99.7	100.0	100.0
	100	6.3	6.7	6.8	6.8	6.0	7.3	8.6	82.5	92.8	98.5	100.0	100.0	100.0	100.0
$J_\alpha(0)$	60	10.8	12.2	10.1	12.1	11.8	12.4	11.6	63.7	79.0	89.9	97.8	99.9	100.0	100.0
	100	11.0	11.3	11.2	11.0	10.8	11.2	12.2	89.3	96.0	99.4	100.0	100.0	100.0	100.0

Table 5 —*Continued*

Panel B: Non-normal Errors with $\rho_\varepsilon = 0.5$															
		Size							Power						
	(T,N)	50	100	200	500	1000	2000	5000	50	100	200	500	1000	2000	5000
Pure spatial models ($\boldsymbol{\gamma} = \mathbf{0}$)															
$\hat{J}_\alpha$	60	7.8	7.2	7.8	8.7	8.8	9.1	10.1	58.9	72.6	88.4	97.5	99.7	100.0	100.0
	100	7.2	6.8	6.9	6.4	6.3	7.2	7.3	82.2	93.1	99.1	99.9	100.0	100.0	100.0
$J_\alpha(0)$	60	11.9	11.1	11.7	12.2	11.8	12.0	12.4	68.0	79.0	92.3	98.6	99.9	100.0	100.0
	100	10.8	10.9	12.2	10.2	10.6	11.9	11.6	87.7	96.1	99.3	100.0	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 1/4$)															
$\hat{J}_\alpha$	60	7.5	6.8	8.1	7.3	8.2	8.6	10.1	57.9	72.4	87.4	97.8	99.5	100.0	100.0
	100	6.9	6.5	7.2	5.4	7.7	7.8	6.8	82.5	93.8	98.9	100.0	100.0	100.0	100.0
$J_\alpha(0)$	60	11.2	9.8	12.1	9.8	11.2	11.8	13.3	66.5	79.3	91.4	98.6	99.6	100.0	100.0
	100	10.6	10.9	12.0	9.5	11.8	11.5	11.1	86.9	96.2	99.4	100.0	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 1/2$)															
$\hat{J}_\alpha$	60	7.5	7.9	8.1	8.5	8.2	9.4	11.2	55.8	71.7	85.9	97.0	99.6	99.9	100.0
	100	7.9	7.1	8.2	6.7	6.5	7.6	7.3	80.0	94.2	98.7	100.0	100.0	100.0	100.0
$J_\alpha(0)$	60	11.4	12.3	12.5	12.0	11.8	13.0	13.5	65.5	79.6	90.8	98.2	99.8	100.0	100.0
	100	11.6	11.2	12.3	11.6	11.2	12.7	12.1	85.6	96.7	99.3	100.0	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 3/5$)															
$\hat{J}_\alpha$	60	7.0	7.0	7.5	8.3	10.3	9.5	12.5	53.9	71.5	85.6	96.4	99.5	100.0	100.0
	100	6.7	7.5	7.3	6.5	8.4	7.7	8.6	81.3	92.0	98.7	100.0	100.0	100.0	100.0
$J_\alpha(0)$	60	11.5	11.7	11.2	12.9	13.5	12.5	14.8	64.9	78.9	90.3	98.3	99.6	100.0	100.0
	100	12.0	12.2	13.1	11.0	13.7	12.8	13.5	87.8	96.1	99.3	100.0	100.0	100.0	100.0

Table 6: Size and power of $\hat{J}_\alpha$ test with time-varying beta and mixed spatial-factor model (spatial parameter $\rho_\varepsilon = 0.5$)

The data generating process is $y_{it} = \alpha_i + \sum_{\ell=1}^{3} \beta_{\ell it} f_{\ell t} + u_{it}$, $i = 1, 2, .., N; t = 1, 2, ..., T$, $\beta_{\ell it} = \beta_{\ell i} + \upsilon_{\ell it}$ with $\upsilon_{\ell it} \sim IIDN(0, 1)$, which are drawn independently over $\ell = 1, 2, 3$, i and t. See Table 5 and the notes to Table 2 for further details.

	Size							Power						
(T,N)	50	100	200	500	1000	2000	5000	50	100	200	500	1000	2000	5000
Panel A: Normal Errors														
Pure spatial models ($\boldsymbol{\gamma} = \mathbf{0}$)														
60	6.0	5.8	6.1	6.3	4.7	4.6	4.1	51.0	64.6	80.3	93.3	98.8	99.5	99.8
100	5.8	5.5	4.5	3.5	3.4	2.9	2.2	78.0	90.4	97.8	99.9	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 1/4$)														
60	5.4	5.2	5.5	4.6	4.0	4.6	3.1	50.1	64.0	78.5	93.5	98.5	99.8	99.8
100	5.8	5.4	5.0	4.9	2.9	2.9	2.1	77.0	89.8	98.0	99.9	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 1/2$)														
60	6.4	6.0	5.5	5.7	4.7	4.3	3.9	50.0	62.4	79.3	92.4	98.3	99.6	100.0
100	5.8	5.0	5.9	5.5	3.8	3.7	3.1	77.1	89.9	97.4	99.9	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 3/5$)														
60	6.1	6.8	5.5	6.2	5.6	6.0	5.5	47.3	63.7	77.1	91.8	98.3	99.6	100.0
100	5.8	5.4	5.8	4.8	4.2	4.7	3.7	77.6	88.7	97.1	99.9	100.0	100.0	100.0
Panel B: Non-normal Errors														
Pure spatial models ($\boldsymbol{\gamma} = \mathbf{0}$)														
60	7.1	6.2	6.1	6.4	5.2	5.8	4.7	52.7	65.6	80.3	94.1	98.4	99.6	100.0
100	5.9	5.7	5.8	4.4	3.8	3.9	2.8	77.7	90.4	97.9	99.9	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 1/4$)														
60	6.5	5.1	6.1	5.4	5.7	4.5	4.2	51.3	64.1	80.1	93.5	98.1	99.8	100.0
100	5.8	5.6	6.1	4.0	4.9	4.2	2.6	76.7	90.1	97.5	99.9	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 1/2$)														
60	6.5	6.5	6.9	6.7	5.3	5.8	5.3	48.8	64.0	78.3	91.9	97.9	99.5	100.0
100	6.6	6.3	5.9	4.8	4.2	4.6	3.0	73.8	90.5	97.2	99.8	100.0	100.0	100.0
Mixed spatial-factor models ($\delta_\gamma = 3/5$)														
60	6.2	6.9	5.7	5.7	8.2	6.2	6.2	47.3	64.2	77.6	92.4	97.4	99.1	99.9
100	6.2	6.8	6.3	4.8	5.6	4.7	3.8	76.2	88.5	96.8	100.0	100.0	100.0	100.0

Table 7: Summary Statistics of p-values, departure from non-normality and average pair-wise correlations of residuals

This table provides summary statistics for p-values of the $\hat{J}_\alpha$ tests applied to residuals from CAPM and FF regressions of securities in the S&P 500 index using rolling sixty months estimation windows over the period from September 1989 to June 2015. The table also reports cross-sectional averages of measures of departure from non-normality and average pair-wise correlations of the residuals. Results reported in panel A of the table refer to CAPM regression residuals, $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}(r_{m,\tau t} - r_{f,\tau t}) + \hat{u}_{i,\tau t}$, for $t = 1, 2, ..., 60$, and $i = 1, 2, ..., N_\tau$, and the months ending in τ =September 1989,..., June 2015. $\overline{\hat{\gamma}}_{\ell\tau} = N_\tau^{-1}\sum_{i=1}^{N_\tau}\hat{\gamma}_{\ell,i\tau}$ for $\ell = 1, 2$, $\hat{\gamma}_{1,i\tau} = \hat{m}_{3,i\tau}/\hat{m}_{2,i\tau}^{3/2}$ and $\hat{\gamma}_{2,i\tau} = \hat{m}_{4,i\tau}/\hat{m}_{2,i\tau}^2 - 3$ with $\hat{m}_{s,i\tau} = (60)^{-1}\sum_{t=1}^{60}\hat{u}_{i,\tau t}^s$. Skewness statistic for testing $\gamma_{1,i\tau} = 0$ is $SK_{i\tau} = T\hat{\gamma}_{1,i\tau}^2/3 \sim \chi_1^2$, and the Kurtosis statistic for testing $\gamma_{2,i\tau} = 0$ is $KR_{i\tau} = T\hat{\gamma}_{2,i\tau}^2/24 \sim \chi_1^2$. Jarque and Bera (1987) statistic for testing $\gamma_{1,i\tau} = \gamma_{2,i\tau} = 0$ is $SK_{i\tau} + KR_{i\tau} \sim \chi_2^2$. Rejection frequency refers to the proportion of normality tests rejected out of the N_τ tests carried at the end of each month, τ. $\overline{\hat{\rho}_\tau} = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{\tau,ij}$, $\hat{\rho}_{\tau\ N,T}^2 = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{\tau,ij}^2$ with $\hat{\rho}_{\tau,ij} = \hat{\mathbf{u}}_{i.\tau}'\hat{\mathbf{u}}_{j.\tau}/(\hat{\mathbf{u}}_{i.\tau}'\hat{\mathbf{u}}_{i.\tau})^{1/2}(\hat{\mathbf{u}}_{j.\tau}'\hat{\mathbf{u}}_{j.\tau})^{1/2}$, $\hat{\mathbf{u}}_{i.\tau} = (\hat{u}_{i,\tau 1}, \hat{u}_{i,\tau 2}..., \hat{u}_{i,\tau T})'$, and $\tilde{\rho}_{\tau\ N,T}^2$ is the MT estimator defined by (56). Results reported in panel B of the table refer to FF regression residuals: $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}(r_{m,\tau t} - r_{f,\tau t}) + \hat{\beta}_{2,i\tau}SMB_{t\tau} + \hat{\beta}_{3i}HML_{t\tau} + \hat{u}_{i,\tau t}$, for $t = 1, 2, ..., 60$, and $i = 1, 2, ..., N_\tau$, and the month ending in τ =September 1989,..., June 2015.

			Average skewness & excess kurtosis measures		Rejection frequency for normality tests at τ			Average pair-wise correlations		
	N_τ	p-value of $\hat{J}_\alpha$	$\overline{\hat{\gamma}}_{1\tau}$	$\overline{\hat{\gamma}}_{2\tau}$	$\gamma_{1,i\tau} = 0$	$\gamma_{2,i\tau} = 0$	$\gamma_{1,i\tau} = 0$ $\gamma_{2,i\tau} = 0$	$\overline{\hat{\rho}_\tau}$	$\hat{\rho}_{\tau\ N,T}^2$	$\tilde{\rho}_{\tau\ N,T}^2$
Panel A: CAPM regressions										
Mean	479	0.52	0.20	1.20	0.24	0.29	0.32	0.02	0.03	0.01
Median	480	0.63	0.19	1.16	0.24	0.28	0.31	0.01	0.03	0.01
Min	464	0.00	-0.01	0.38	0.13	0.12	0.15	0.01	0.02	0.00
Max	487	1.00	0.37	2.16	0.35	0.46	0.47	0.08	0.05	0.02
stand. dev.	5.9	0.38	0.09	0.46	0.06	0.09	0.08	0.03	0.01	0.00
Panel B: Fama-French regressions										
Mean	479	0.46	0.19	1.06	0.22	0.26	0.28	0.01	0.03	0.00
Median	480	0.50	0.20	1.02	0.23	0.25	0.28	0.01	0.03	0.00
Min	464	0.00	0.02	0.38	0.12	0.11	0.14	0.00	0.02	0.00
Max	487	0.98	0.34	1.91	0.31	0.40	0.42	0.03	0.03	0.01
stand. dev.	5.9	0.33	0.09	0.37	0.05	0.07	0.07	0.01	0.00	0.00

Figure 1: Plots of p-value of the $\hat{J}_\alpha$ test

This figure presents plots of the evolution of p-values of the $\hat{J}_\alpha$ test based on CAPM and FF regressions of securities in the S&P 500 index using five year estimation windows (sixty months) at the end of the months from September 1989 to June 2015. Reported plots are the p-values of the $\hat{J}_\alpha$ test, which are computed using CAPM regressions, $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{u}_{i,\tau t}$ and FF three factor regressions, $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{\beta}_{2,i\tau} SMB_{t\tau} + \hat{\beta}_{3i} HML_{t\tau} + \hat{u}_{i,\tau t}$, for $t = 1, 2, ..., 60$, and $i = 1, 2, ..., N_\tau$, of the month ends estimation windows τ =September 1989,..., June 2015.

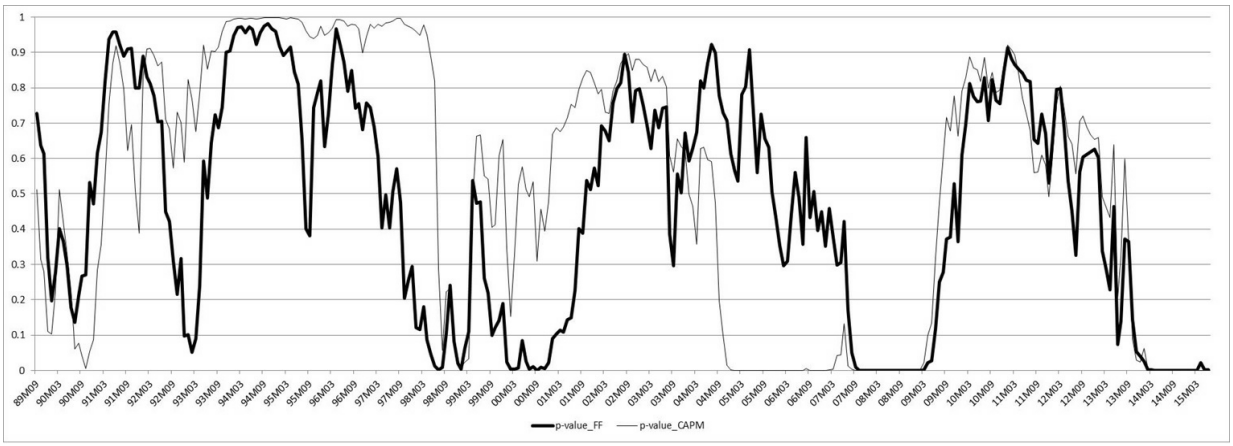


Figure 2: Plots of Hedge Fund Index relative to S&P 500 returns and p-values of the $\hat{J}_\alpha$ test **based on CAPM regressions**

This figure presents monthly rate of returns of Credit Suisse Core Long/Short Equity Hedge Fund Index relative to S&P 500 returns, and p-values of the $\hat{J}_\alpha$ test applied to CAPM regressions over the period November 1994 to June 2015. The long/short return variable, $\tilde{r}_{ht}(12)$, is computed as $\tilde{r}_{ht}(12) = \frac{1}{12}\sum_{j=0}^{11}\tilde{r}_{h,t-j}$, where $\tilde{r}_{ht} = r_{ht} - r_t$, r_{ht} is the return on Credit Suisse Core Long/Short Equity Hedge Fund Index, and r_t is the return on S&P 500 index. $\hat{\pi}_\tau(12) = \frac{1}{12}\sum_{j=0}^{11}\hat{\pi}_{\tau-j}$, where $\hat{\pi}_\tau$ is the p-values of the $\hat{J}_\alpha$ test at the end of month τ, computed using CAPM regressions estimated on rolling samples of sixty months. See the notes to Table 7 for details of CAPM regressions.

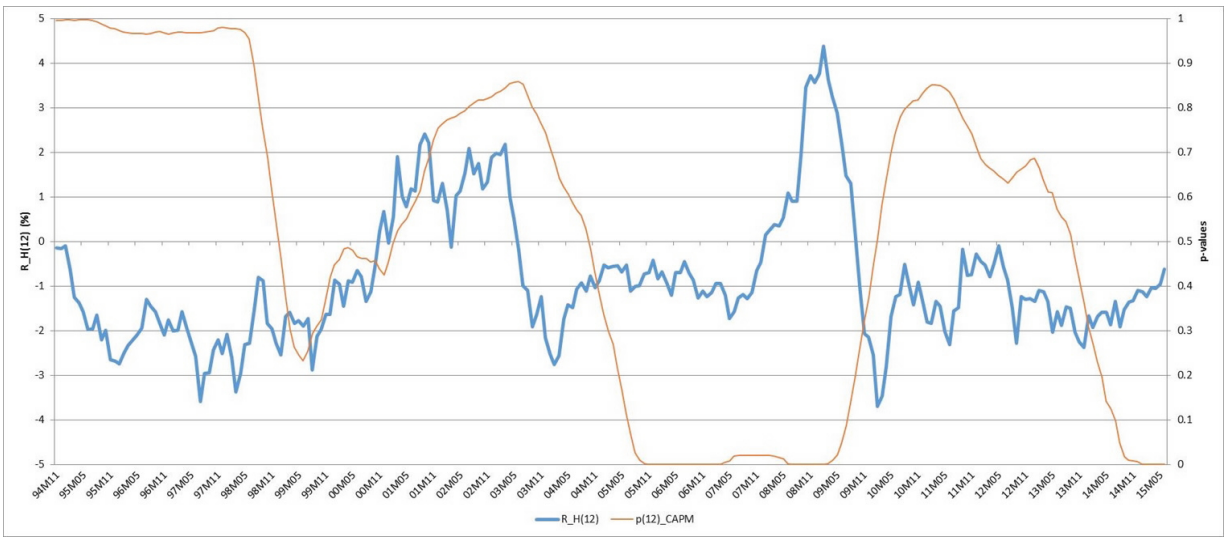
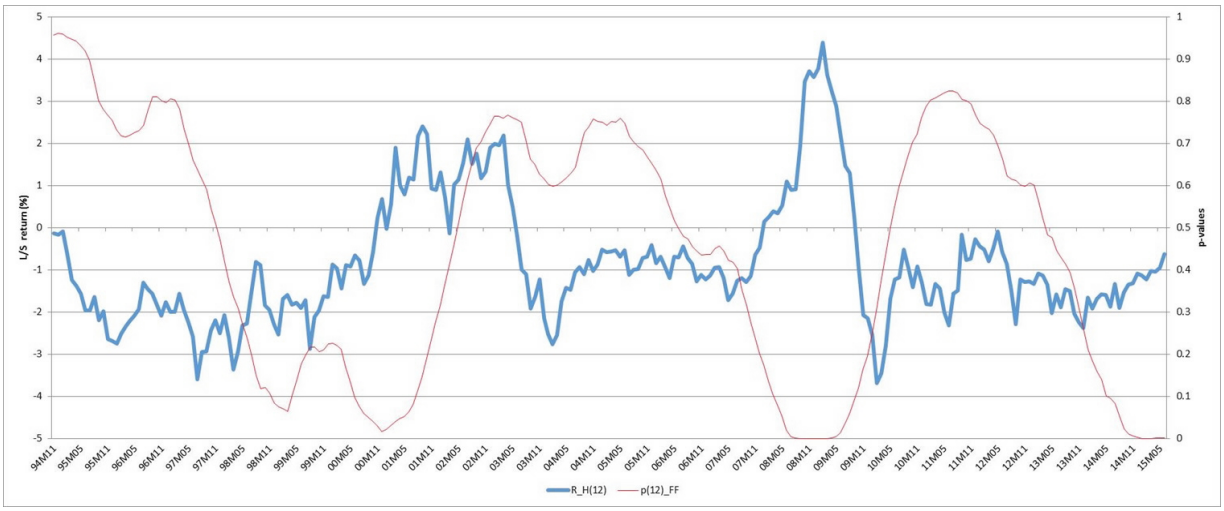


Figure 3: Plots of Hedge Fund Index relative to S&P 500 returns and p-values of the $\hat{J}_\alpha$ test **based on FF regressions**

This figure presents monthly rate of return of Credit Suisse Core Long/Short Equity Hedge Fund Index relative to S&P 500 return, and p-value of the $\hat{J}_\alpha$ test based on Fama-French regressions over the period November 1994 to June 2015. See the notes to Figure 2, and the notes to Table 7 for details of Fama-French regressions.



Appendices

Appendix A: Proofs of the theorems

In this appendix we provide proofs of the theorems set out in Section 4 of the paper. These proofs make use of Lemmas which are provided, together with their proofs, in an Online Supplement available on request.

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

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

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

where $\mathbf{F}$ is a $T \times m$ matrix, and $\boldsymbol{\tau}_T = (1, 1, ..., 1)'$ 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) which is used to establish it.

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

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

where $\{\varepsilon_i,\ i = 1, 2, ..., N\}$ 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 $i = 1, 2, ..., N$, 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 $N^{-1}Var\left(Q_N\right) \geq c$ for some $c > 0$,

$$\frac{Q_N - E\left(Q_N\right)}{\sqrt{Var\left(Q_N\right)}} \to_d N(0, 1).$$

Proof. See Kelejian and Prucha (2001, Theorem 1, p. 227). ■

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

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

where $\mathbf{D}_\sigma = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})$. Using (48)

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

where

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

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

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

where

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

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 $\mathbf{B}$, $Tr(\mathbf{AB}) \leq Tr(\mathbf{A})\lambda_{\max}(\mathbf{B})$ (this result is repeatedly used below), we have

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

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 $IID(\mathbf{0},\mathbf{I}_k)$, it then readily follows that $\mathbf{d}_T'\mathbf{d}_T \to_p 1$, and hence

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

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

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

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

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

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 $Var(b_{NT})$, and note that for given values of γ_i we have (recall that η_{it} is independent over t and $\sum_{t=1}^T h_t^2 = w_T$)

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

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

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

Further

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

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

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

But by condition (51) in Assumption 4 and $\sigma_{\eta,ii} > c > 0$ imply $\sup_j \sum_{i=1}^{N}\left|\rho_{\eta,ij}\right| < K$ (also see (52)), and by (47) we have $\sup_s \sum_{i=1}^{N}\left|\gamma_{is}\right| = O\left(N^{\delta_\gamma}\right)$. Then it follows that $Var\left(b_{NT}\right) = O\left(N^{\delta_\gamma - 1}\right)$, and $b_{NT} = O\left(N^{\delta_\gamma/2-1/2}\right)$. Therefore, b_{NT} is dominated by a_{NT} and using these results in (A.3) we have

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

Now using (50) we can express the above as

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

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

$$\begin{aligned} N^{-1/2}\sum_{i=1}^{N}\left(z_i^2 - 1\right) &= N^{-1/2}w_T^{-1}\left(\sum\nolimits_{t=1}^{T} h_t\boldsymbol{\varepsilon}_{\eta,t}'\right)\left(\mathbf{Q}_\eta'\mathbf{D}_\sigma^{-1}\mathbf{Q}_\eta\right)\left(\sum\nolimits_{t=1}^{T} h_t\boldsymbol{\varepsilon}_{\eta,t}\right) + O_p\left(N^{\delta_\gamma-1/2}\right) \\ q_{NT} &= N^{-1/2}\left[\mathbf{x}_T'\mathbf{A}\mathbf{x}_T - Tr\left(\mathbf{A}\right)\right] + N^{-1/2}\left[Tr(\mathbf{A}) - N\right] + O_p\left(N^{\delta_\gamma-1/2}\right). \end{aligned} \tag{A.7}$$

where

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

First consider the deterministic component of q_{NT}, and using (49) and under Assumption 4 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 $\tilde{\boldsymbol{\Gamma}} = (\tilde{\gamma}_1, \tilde{\gamma}_2, ..., \tilde{\gamma}_N)'$. Then

$$Tr\left(\mathbf{R}\right) = N = \sum\nolimits_{i=1}^{N}\tilde{\gamma}_i'\tilde{\gamma}_i + Tr\left(\mathbf{A}\right).$$

But, as before,

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

Hence

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

and (A.7) can be written as

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

where

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

We now apply the Central Limit Theorem for Linear Quadratic Forms due to Kelejian and Prucha (2001, KP) to z_{NT}, which is reproduced for convenience as Lemma 1 above. We first establish the conditions required by KP's theorem (see Lemma 1). To this end we first note that $E\left(\mathbf{x}_T\right) = \mathbf{0}$, and

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

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 $\boldsymbol{\varepsilon}_{\eta,i} = (\varepsilon_{\eta,i1}\ \varepsilon_{\eta,i2}, ..., \varepsilon_{\eta,iT})'$, 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 $E(x_{i,T}^6) = w_T^{-3}E\left(\boldsymbol{\varepsilon}'_{\eta,i}\mathbf{H}_F\boldsymbol{\varepsilon}_{\eta,i}\right)^3 = 15 + O(v^{-1}) \leq K$ 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 (A.12) and note that it is symmetric and we have

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

and using (A.8)

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

But under condition (51) and noting that $\sigma_{ii} > c > 0$, then

$$\left\|\tilde{\mathbf{A}}\right\|_\infty = \sup_i \sum\nolimits_{j=1}^N |\tilde{a}_{ij}| < K,$$

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

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

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

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

Also

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

and in view of (51) we have

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

and hence (using (A.10)):

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

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

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

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

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

Also in view of (A.8)

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

To summarize

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

which also yield

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

Therefore,

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

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

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

Finally, using (A.12)

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

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

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

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

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

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

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

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

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

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

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

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

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

and using (A.16)

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

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

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

Therefore

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

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

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

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 standardised error of the return equation (2) 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 $X_i=\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i/v$, with $v=T-m-1$ and $\mathbf{M}_G=(m_{tt'})$, defined by (A.1). Also, by (35), $E\left(t_i^2\right)=E\left(z_i^2/X_i\right)=v/\left(v-2\right)+O\left(v^{-3/2}\right)$ for each i, and by Lemma 11 $E\left(z_i^2\right)=E\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i/w_T\right)=w_T^{-1}Tr\left(\mathbf{H}_F\right)=1$, for all i. Thus, we have

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

Next, for all $i=1,2,...,N$ we have $X_i>0$, and (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

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

and

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

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

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

and

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

But

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

Now using results from Lemma 11 we have

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

which yields

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

Using this result in (A.23) we obtain

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

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

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

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

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

where $\gamma_{2,\varepsilon_\eta}=E(\varepsilon_{\eta,it}^4)-3$ (and $\left|\gamma_{2,\varepsilon_\eta}\right|<K$ 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 (50). But as $0\leq m_{tt}\leq 1$ ($\mathbf{M}_G=(m_{tt'})$) by Lemma 8, $v^{-1}w_T^{-1}\sum_{t=1}^T h_t^2 m_{tt}\leq v^{-1}w_T^{-1}\sum_{t=1}^T h_t^2=v^{-1}$ as $\sum_{t=1}^T h_t^2=w_T$, and also that $0\leq\sum_{\ell=1}^N\tilde{q}_{\eta,i\ell}^4\leq 1$, 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 $\left|\gamma_{2,\varepsilon_\eta}\right|\leq K$, we have

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

Furthermore,

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

We first note that

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

As has shown above,

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

uniformly over i. Next consider

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

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

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

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

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

therefore,

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

uniformly over i. We conclude that

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

Secondly, by Lemma 16,

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

In sum, under Assumptions 1-4, $S_{NT} \to_p 0$, so long as $0 \leq \delta_\gamma < 1/2$, $N/T^2 \to 0$ as N and $T \to \infty$, jointly. ■

Proof of Theorem 3. Under Assumptions 1-4, using Theorem 2 we have $N^{-1/2}\sum_{i=1}^N \left(z_i^2 - t_i^2\right)/[2\left(1+(N-1)\rho_N^2\right)]^{1/2} \to_p 0$ with z_i^2 defined by (20), so long as $(N-1)\rho_N^2 = O(1)$, $N/T^2 \to 0$, and $0 \leq \delta_\gamma < 1/2$, as N and $T \to \infty$, jointly. Under these conditions, (by Lemma 4) it implies that $N^{-1/2}\sum_{i=1}^N \left(t_i^2 - \frac{v}{v-2}\right)/[2\left(1+(N-1)\rho_N^2\right)]^{1/2}$ has the same limit distribution as $N^{-1/2}\sum_{i=1}^N \left(z_i^2 - 1\right)/[2\left(1+(N-1)\rho_N^2\right)]^{1/2}$, which is shown to be standard normal by Theorem 1 and the desired result now follows, observing that $\lim_{v\to\infty}\left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4} = 2$. ■

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

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

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

$$|\psi_{NT}| \leq \frac{2}{N}\sum_{i,j=1}^N \left|\tilde{\rho}_{ij} - \rho_{ij}\right|. \tag{A.27}$$

Further, letting $I_{ij} = I\left[\left|\sqrt{v}\hat{\rho}_{ij}\right| > c_p(N)\right]$, 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

$$\begin{aligned} E\left|\psi_{NT}\right| \quad \leq \quad & \frac{2}{N}\sum_{i,j=1}^N E\left(\left|\hat{\rho}_{ij} - E\left(\hat{\rho}_{ij}\right)\right| \times I_{ij}\right) + \frac{2}{N}\sum_{i,j=1}^N \left|E\left(\hat{\rho}_{ij}\right) - \rho_{ij}\right| E\left(I_{ij}\right) \\ & + \frac{2}{N}\sum_{i,j=1}^N \left|\rho_{ij}\right| E\left(1 - I_{ij}\right). \end{aligned} \tag{A.28}$$

Now using (39) 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 $\mathbf{L}$ $(\mathbf{L}\mathbf{L}' = \mathbf{I}_T)$, defined by

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

Hence, setting

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

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

$$\begin{aligned} E\left(\zeta_{it}\right) &= 0,\ E\left(\zeta_{it}^2\right) = 1,\ E\left(\zeta_{it}\zeta_{jt}\right) = \rho_{ij}, \\ \kappa_{ij}(4,0) &= E(\zeta_{it}^4) - 3,\ \kappa_{ij}(0,4) = E(\zeta_{it}^4) - 3, \\ \kappa_{ij}(3,1) &= E(\zeta_{it}^3\zeta_{jt}) - 3\rho_{ij},\ \kappa_{ij}(1,3) = E(\zeta_{it}\zeta_{jt}^3) - 3\rho_{ij}, \\ \kappa_{ij}(2,2) &= E(\zeta_{it}^2\zeta_{jt}^2) - 2\rho_{ij} - 1. \end{aligned}$$

Furthermore, by Lemma 19

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

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

where

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

and

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

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

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

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

$$\begin{aligned}
& \frac{1}{Nv}\sum_{i,j=1}^{N}\left|a_{ij}\right| \\
\leq\ & \frac{1}{2}\frac{1}{Nv}\sum_{i,j=1}^{N}\left|\rho_{ij}\right|\left|1-\rho_{ij}^2\right| + \frac{3}{8}\frac{1}{Nv}\sum_{i,j=1}^{N}\left|\rho_{ij}\right|\left|\kappa_{ij}(4,0)+\kappa_{ij}(0,4)\right| \\
& + \frac{1}{4}\frac{1}{Nv}\sum_{i,j=1}^{N}\left|\kappa_{ij}(3,1)+\kappa_{ij}(1,3)\right| + \frac{1}{2Nv}\sum_{i,j=1}^{N}\left|\rho_{ij}\right|\left|\kappa_{ij}(2,2)\right| \\
=\ & \frac{1}{4}\frac{1}{Nv}\sum_{i,j=1}^{N}\left|\kappa_{ij}(3,1)+\kappa_{ij}(1,3)\right| + O(v^{-1}). \qquad \text{(A.33)}
\end{aligned}$$

Also

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

and as established in Lemma 20 (see (S.80)) we have

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

which if used in (A.33) yields

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

and overall for the second term of (A.28) we have

$$\frac{2}{N}\sum_{i,j=1}^{N}\left|E\left(\hat{\rho}_{ij}\right)-\rho_{ij}\right| E\left(I_{ij}\right) = O(v^{-1}N^{2\delta_\gamma-1}) + O(v^{-1}) + O\left(Nv^{-2}\right), \qquad \text{(A.34)}$$

which tends to zero if $\delta_\gamma \leq 1/2$, and $N/v^2 \to 0$, as N and $v \to \infty$, jointly. To deal with the first and third terms of (A.28) we need to distinguish between values of $\left|\rho_{ij}\right|$ that are strictly away from zero, namely those values that satisfy the condition $\left|\rho_{ij}\right| > \rho_{\min} > 0$, and those values that are *zero* or very close to zero. Note that since by assumption $\sum_{i,j=1}^{N}\left|\rho_{ij}\right| = O(N)$, then it is not possible for all values of $\left|\rho_{ij}\right|$ to be strictly away from zero. To formalize the notation of $\left|\rho_{ij}\right|$ as being close to zero, we suppose that there exist integers N_0 and v_0 such that for all values of $N > N_0$ and $v > v_0$, $\left|\rho_{ij}\right| \leq v^{-1/2}c_p(N)$. The non-zero values are defined by $\left|\rho_{ij}\right| > v^{-1/2}c_p(N)$. In our analysis this is a natural categorization of $\left|\rho_{ij}\right|$, since $v^{-1/2}c_p(N) \to 0$, as N and $T \to 0$.[28] Given this categorization consider now the third term of (A.28) and note that

$$\begin{aligned}
\frac{2}{N}\sum_{i,j=1}^{N}\left|\rho_{ij}\right| E\left(1-I_{ij}\right) \leq\ & \frac{2}{N}T^{-1/2}c_p(N)\sum_{i,j=1}^{N}E\left[\left(1-I_{ij}\right)\Big|\left|\rho_{ij}\right| \leq T^{-1/2}c_p(N)\right] \\
& + \frac{2}{N}\sum_{i,j=1}^{N}E\left[\left(1-I_{ij}\right)\Big|\left|\rho_{ij}\right| > \rho_{\min} > T^{-1/2}c_p(N)\right]
\end{aligned}$$

Then following a similar line of proof as in Lemma 6 of BPS (2016, supplement) we have (for some small $\epsilon > 0$)

$$E\left[\left(1-I_{ij}\right)\Big|\left|\rho_{ij}\right| \leq v^{-1/2}c_p(N)\right] \leq Ke^{\frac{-(1-\epsilon)}{2}\frac{c_p^2(N)}{b_{ij}}}\left[1+o(1)\right],$$

and

$$\begin{aligned}
E\left[\left(1-I_{ij}\right)\Big|\left|\rho_{ij}\right| > v^{-1/2}c_p(N)\right] &= \Pr\left[\left|\sqrt{v}\hat{\rho}_{ij}\right| \leq c_p(N)\Big|\left|\rho_{ij}\right| > v^{-1/2}c_p(N)\right] \\
&\leq Ke^{\frac{-1}{2}\frac{v\left[\left|\rho_{ij}\right|-\frac{c_p(N)}{\sqrt{v}}\right]^2}{b_{ij}}}\left[1+o(1)\right].
\end{aligned}$$

[28] See result (a) in Lemma 3 of BPS (2016, supplement).

Hence,

$$\begin{aligned}
\frac{2}{N}\sum_{i,j=1}^{N}\left|\rho_{ij}\right|E\left(1-I_{ij}\right) &\leq \frac{2c_p(N)}{N\sqrt{v}}\sum_{i,j=1}^{N}Ke^{\frac{-(1-\epsilon)}{2}\frac{c_p^2(N)}{b_{ij}}}\left[1+o(1)\right] \\
&\quad +\frac{2K}{N}\sum_{i,j=1}^{N}e^{\frac{-1}{2}\frac{v\left[\left|\rho_{ij}\right|-\frac{c_p(N)}{\sqrt{v}}\right]^2}{b_{ij}}}\left[1+o(1)\right] \\
&\leq \frac{2c_p(N)N}{\sqrt{v}}Ke^{\frac{-(1-\epsilon)}{2}\frac{c_p^2(N)}{\max_{ij}(\varphi_{\iota j})}}\left[1+o(1)\right] \\
&\quad +KNe^{\frac{-1}{2}\frac{v\left[\rho_{\min}-\frac{c_p(N)}{\sqrt{v}}\right]^2}{\max_{ij}\left(b_{ij}\right)}}\left[1+o(1)\right],
\end{aligned}$$

where $\varphi_{ij}=b_{ij}(\rho_{ij}=0)=E(\zeta_{it}^2\zeta_{jt}^2\left|\rho_{ij}=0\right.)$. Finally, consider the first term of (A.28) and write it as

$$\frac{2}{N}\sum_{i,j=1}^{N}E\left[\left|\hat{\rho}_{ij}-E\left(\hat{\rho}_{ij}\right)\right|\times I_{ij}\right]=\frac{2}{N}\sum_{i,j=1}^{N}\sqrt{Var(\hat{\rho}_{ij})}E\left(\left|z_{ij}\right|\times I_{ij}\right),$$

where $z_{ij}=\left[\rho_{ij}-E(\hat{\rho}_{ij})\right]/\sqrt{Var\left(\hat{\rho}_{ij}\right)}$, and $Var\left(\hat{\rho}_{ij}\right)$ is given by (A.32). Also $E\left(\left|z_{ij}\right|\times I_{ij}\right)=E\left(\left|z_{ij}\right|\right)-E\left[\left|z_{ij}\right|\left(1-I_{ij}\right)\right]$, and using results in Lemma 4 of BPS (2016, supplement) we have

$$\begin{aligned}
&\frac{2}{N}\sum_{i,j=1}^{N}E\left[\left|\hat{\rho}_{ij}-E\left(\hat{\rho}_{ij}\right)\right|\times I_{ij}\right] \\
\leq\ &\frac{2}{N}\left[\frac{\sqrt{\max_{ij}b_{ij}}}{\sqrt{v}}+O\left(v^{-1}\right)\right]\times \\
&\sum_{i,j=1}^{N}\left[\phi\left(\frac{c_p(N)-\sqrt{v}\rho_{ij}}{b_{ij}+O\left(v^{-1}\right)}\right)+\phi\left(\frac{c_p(N)-\sqrt{v}\rho_{ij}}{b_{ij}+O\left(v^{-1}\right)}\right)\right]\left[1+o(1)\right].
\end{aligned}$$

Once again we need to distinguish between cases where $\left|\rho_{ij}\right|\leq T^{-1/2}c_p(N)$ and $\left|\rho_{ij}\right|>\rho_{\min}>0$. We do not require to know how many cases fall in one or the other category. Overall (noting that $c_p(N)/\sqrt{v}\to 0$) we have

$$\begin{aligned}
\frac{2}{N}\sum_{i,j=1}^{N}E\left[\left|\hat{\rho}_{ij}-E\left(\hat{\rho}_{ij}\right)\right|\times I_{ij}\right] &\leq KN\left[\frac{\sqrt{\max b_{ij}}}{\sqrt{v}}+O\left(v^{-1}\right)\right]\times \\
&\left[\exp\left(\frac{-1}{2}\frac{c_p^2(N)}{\max_{ij}\varphi_{ij}}\right)+\exp\left(\frac{-1}{2}\frac{v\rho_{\min}^2}{\max_{ij}b_{ij}}\right)\right]\left[1+o(1)\right].
\end{aligned}$$

Overall we require the following condition for $\psi_{NT}\to_p 0$:

$$\frac{Nc_p(N)}{\sqrt{v}}\exp\left(\frac{-1}{2}\frac{c_p^2(N)}{\max_{ij}\varphi_{ij}}\right)\to 0. \tag{A.35}$$

Note that since $\max_{ij}b_{ij}<K$, then

$$\frac{Nc_p(N)}{\sqrt{v}}\exp\left(\frac{-1}{2}\frac{v\rho_{\min}^2}{\max_{ij}b_{ij}}\right)\to 0,\text{ as } N \text{ and } v\to\infty.$$

A sufficient condition for (A.35) to hold is given by $\delta>(1-d/2)\max_{ij}\varphi_{ij}$. This follows since (with $v=N^d$)

$$\begin{aligned}
\frac{Nc_p(N)}{\sqrt{v}}\exp\left(\frac{-1}{2}\frac{c_p^2(N)}{\max_{ij}\varphi_{ij}}\right) &= \exp\left(\frac{-1}{2}\frac{c_p^2(N)}{\max_{ij}\varphi_{ij}}+(1-d/2)\log(N)+\log\left[c_p(N)\right]\right) \\
&= \exp\left(-\log(N)\left[\frac{\frac{1}{2}\frac{c_p^2(N)}{\max_{ij}\varphi_{ij}}-(1-d/2)\log(N)-\log\left[c_p(N)\right]}{\log(N)}\right]\right).
\end{aligned}$$

But $\lim_{N\to\infty}c_p^2(N)/\log(N)=2\delta$, and $\log\left[c_p(N)\right]/\log(N)\to 0$. Hence, condition (A.35) is met if $\left(\delta/\max_{ij}\varphi_{ij}\right)-(1-d/2)>0$, or equivalently if $\delta>(1-d/2)\varphi$, where $\varphi=\max_{ij}\varphi_{ij}$. But using (S.79) established in Lemma 20, and setting $\gamma_i=\mathbf{0}$, 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

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

where l_{tr} is the (t,r) element of the $T\times T$ orthonormal matrix $\mathbf{L}$ defined by (A.29), $q_{\eta,i\ell}$ is such that $\mathbf{Q}_\eta=(q_{\eta,i\ell})$, $\mathbf{Q}_\eta$ defined by (50). Also, $\left|\sigma_{\eta,ii}/\sigma_{ii}\right|\leq 1$, $\sum_{r=1}^{T}l_{tr}^4\leq\left(\sum_{r=1}^{T}l_{tr}^2\right)^2\leq 1$, $\sum_{\ell=1}^{N}\tilde{q}_{\eta,i\ell}^2=\sum_{\ell=1}^{N}q_{\eta,i\ell}^2/\sigma_{\eta,ii}=1$, and

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

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

Proof of Theorem 5. By Theorem 3, $J_\alpha\left(\rho_N^2\right) \to_d N(0,1)$ 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 (55) and (47), respectively. Since Theorem 4 ensures that $\hat{J}_\alpha - J_\alpha(\rho_N^2) \to_p 0$, as $(N-1)\left(\tilde{\rho}_{N,T}^2 - \rho_N^2\right) \to_p 0$ when $d > 2/3$, as N and $v \to \infty$, and $\delta > (1-d/2)\varphi$, where $\varphi \leq 1 + \left|\gamma_{2,\varepsilon_\eta}\right|$, under these conditions, $\hat{J}_\alpha$ has the same limit distribution as $J_\alpha\left(\rho_N^2\right)$ (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 1-4, and under the local alternatives (61), $N^{-1/2}\sum_{i=1}^N\left(z_{i,a}^2 - 1\right) \to_d N(\phi^2, 2\omega^2)$, as $N \to \infty$ and $T \to \infty$, jointly, where $z_{i,a}^2$ defined by (S.97), $\omega^2 = 1 + \lim_{N\to\infty}(N-1)\rho_N^2$, ρ_N^2 is defined by (54). Also, by Lemma 23 we have $N^{-1/2}\sum_{i=1}^N\left(z_{i,a}^2 - t_i^2\right) = o_p(1)$. 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. ■

Appendix B: Generating non-Gaussian errors

To generate non-normal correlated errors, $u_{it}^{(r)}$, with given skewness and kurtosis, we use the following procedure (see Section M1.1 in Online Supplement for full details). For each replication, r,

1. We generate N random draws $\sigma_{ii}^{(r)}$, $\gamma_{1,i}^{(r)}$ and $\gamma_{2,i}^{(r)}$, $i = 1, 2, ..., N$, as described in Section M1.1 and set
$$m_{3,i}^{(r)} = \left[\sigma_{ii}^{(r)}\right]^{3/2}\gamma_{1,i}^{(r)}, \text{ and } m_{4,i}^{(r)} = \left[\sigma_{ii}^{(r)}\right]^2\left(\gamma_{2,i}^{(r)} + 3\right).$$
2. We then set $m_{\varepsilon,1}^{(r)} = 0$ and $m_{\varepsilon,2}^{(r)} = 1$, and derive $m_{\varepsilon,3,i}^{(r)}$ and $m_{\varepsilon,4,i}^{(r)}$ as
$$\mathbf{m}_{\varepsilon,3}^{(r)} = \mathbf{Q}_{(3)}^{(r)-1}\mathbf{m}_3^{(r)}, \quad \boldsymbol{\kappa}_\varepsilon^{(r)} = \mathbf{Q}_{(4)}^{(r)-1}\boldsymbol{\kappa}^{(r)},$$
where , $\mathbf{m}_{\varepsilon,3}^{(r)} = (m_{\varepsilon,3,1}^{(r)}, m_{\varepsilon,3,2}^{(r)}, ..., m_{\varepsilon,3,N}^{(r)})$, $\mathbf{Q}_{(3)}^{(r)} = \mathbf{Q}^{(r)}\odot\mathbf{Q}^{(r)}\odot\mathbf{Q}^{(r)}$, $\mathbf{m}_3^{(r)} = (m_{3,1}^{(r)}, m_{3,2}^{(r)}, ..., m_{3,N}^{(r)})'$, $\boldsymbol{\kappa}_\varepsilon^{(r)} = (\kappa_{\varepsilon 1}^{(r)}, \kappa_{\varepsilon 2}^{(r)}, ..., \kappa_{\varepsilon N}^{(r)})'$, $\mathbf{Q}_{(4)}^{(r)} = \mathbf{Q}^{(r)}\odot\mathbf{Q}^{(r)}\odot\mathbf{Q}^{(r)}\odot\mathbf{Q}^{(r)}$, and $\boldsymbol{\kappa}^{(r)} = (\kappa_1^{(r)}, \kappa_2^{(r)}, ..., \kappa_{\mathrm{N}}^{(r)})'$ with $\kappa_{\varepsilon i}^{(r)} = m_{\varepsilon,4,i}^{(r)} - 3$ and $\kappa_i^{(r)} = m_{4,i}^{(r)} - 3\sigma_{ii}^{2(r)}$, $\mathbf{Q}^{(r)} = \mathbf{D}^{(r)1/2}\mathbf{P}^{(r)}$, with $\mathbf{D}^{(r)} = diag(\sigma_{11}^{(r)}, \sigma_{22}^{(r)}, ..., \sigma_{NN}^{(r)})'$ and $\mathbf{P}^{(r)}$ being a Cholesky factor of correlation matrix $\mathbf{R}^{(r)}$. The correlation matrix, $\mathbf{R}$, is defined by (64). The operator $\odot$ denotes the Hadamard or element-wise multiplication.
3. Following Fleishman (1978), we then generate ε_{it}, $t = 1, 2, ..., T$ as (suppressing the superscript r for notational convenience)
$$\varepsilon_{it} = a_i + b_i\upsilon_{it} + c_i\upsilon_{it}^2 + d_i\upsilon_{it}^3, \; i = 1, 2, ..., N,$$
where $\upsilon_{it} \sim IIDN(0,1)$ and the coefficients a_i, b_i, c_i and d_i are determined so that $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = 1$, $E(\varepsilon_{it}^3) = m_{\varepsilon,3,i}$ and $E(\varepsilon_{it}^4) - 3 = \kappa_{\varepsilon i}$. This involves solving the following system of equations for a_i, b_i, c_i and d_i:
$$a_i + c_i = 0,$$
$$b_i^2 + 6b_id_i + 2c_i^2 + 15d_i^2 = 1,$$
$$2c_i(b_i^2 + 24b_id_i + 105d_i^2 + 2) = m_{\varepsilon,3,i},$$
$$24[b_id_i + c_i^2(1 + b_i^2 + 28b_id_i) + d_i^2(12 + 48b_id_i + 141c_i^2 + 225d_i^2)] = \kappa_{\varepsilon i}.$$
4. Finally, we set $u_{it}^{(r)} = \sum_{j=1}^N q_{ij}^{(r)}\varepsilon_{jt}^{(r)}$, where $q_{ij}^{(r)}$ is the (i,j) element of $\mathbf{Q}^{(r)}$, and $\varepsilon_{jt}^{(r)}$ is the rth draw from the DGP in step 3 above.

Appendix C: Data sources and their descriptions

We downloaded price and dividend data on all 500 securities included in the S&P 500 index at close of each month from September 1989 to June 2015 (inclusive) using Datastream.[29] For example, the code LS&PCOMP1210 will give the 500 constituents of S&P 500 index as of December 2010.To construct our security return data, the security price (P) and dividend yield (DY) are obtained from Datastream, as specified the

[29] We could only download data for 499 securities on September 30, 2008, and it is confirmed on Standard & Poor's website that the S&P 500 index on this day was based on 499 securities.

table below. We adopted the following rules in selecting individual securities for inclusion in our analysis. At the end of each month under consideration, we downloaded historical return series on all 500 securities included in the S&P 500 index at the time. We then dropped all securities with less than 60 months of observations and/or with five consecutive zeros in the middle of sample periods.

Variable	**Description**	**Source (Code)**
P_{it}	Price of security i at the market close of the last day of the month (t), adjusted for subsequent capital actions.	Datastream (LS&PCOMP, P)
DY_{it}	Dividend per share as a percentage of the share price based on an anticipated annual dividend and excludes special or once-off dividends.	Datastream (LS&PCOMP, DY)
P_t	S&P 500 price index at close of the final day of the month (t).	Datastream (S&PCOMP, PI)
DY_t	'Dividend yield' on S&P 500 as a percentage of P_t.	Datastream (S&PCOMP, DY, up to Oct. 2012, S&PCOMZ, DY, Nov. 2012 onwards)
SMB_t	Average return in per cent on the three small portfolios minus the average return on the three big portfolios.	Ken French's data library (up to Jan. 2016)
HML_t	Average return in per cent on two value portfolios minus the average return on two growth portfolios.	Ken French's data library (up to Jan. 2016)
r_{it}	Monthly return of security i in month t in per cent, computed as $100(P_{it} - P_{i,t-1})/P_{i,t-1} + DY_{it}/12$.	Datastream
r_{ft}	One-month US treasury bill rate in per cent in month t as the risk-free asset return from Ibbotson Associates.	Ken French's data library (up to Jan. 2016)
r_{mt}	Value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) in per cent.	Ken French's data library (up to Jan. 2016)
r_t	Monthly return of S&P 500 portfolio at month t in per cent, computed as $100(P_t - P_{t-1})/P_{t-1} + DY_t/12$.	Datastream
r_{ht}	Monthly rate of return of Dow Jones Credit Suisse Core Long/Short Equity Hedge Fund (the end of the month)	Credit Suisse (ROR), up to May 2016 http://www.hedgeindex.com
$\tilde{r}_{ht}$	$r_{ht} - r_t$.	

References

Affleck-Graves, J., McDonald, B., 1989. Nonnormalities and tests of asset pricing theories. *Journal of Finance* 44, 889-908.

Affleck-Graves, J., McDonald, B., 1990. Multivariate tests of asset pricing: The comparative power of alternative statistics. *Journal of Financial and Quantitative Analysis* 25, 163-185.

Amengual, D., Repetto, L., 2014. Testing a large number of hypotheses in approximate factor models, cemfi wp/1410.

Anderson, T.W., 2003. *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Hoboken, NJ: John Wiley & Sons.

Ang, A., Chen, J., Xing, Y., 2006. Downside Risk. *Review of Financial Studies* 19, 1191-1239

Ang, A., Liu, J., Schwarz, K., 2016. Using stocks or portfolios in tests of factor models. The Wharton School, University of Pennsylvania.

Bailey, N., Pesaran, M. H., Smith, L.V., 2016. A Multiple Testing Approach to the Regularisation of Large Sample Correlation Matrices, QMUL Working Paper Series No. 764.

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

Bickel, P.J., Levina, E., 2008. Regularized estimation of large covariance matrices. *Annals of Statistics* 36, 199 227.

Black, F., Jensen, M.C., Scholes, M., 1972. The capital asset pricing model: Some empirical tests, in: *Studies in the Theory of Capital Markets*, Jensen, M.C., ed. New York: Praeger, pp.79–121.

Bossaerts, P., Plott, C., Zame, W.R., 2007. Prices and portfolio choices in financial markets: theory, econometrics, experiments. *Econometrica* 75, 993–1038.

Breusch, T.S., Pagan, A.R., 1980. The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47, 239-253.

Cai, T., Liu, W., 2011. Adaptive thresholding for sparse covariance matrix estimation. *Journal of American Statistical Association* 106, 672-84

Chamberlain, G., 1983. Funds, factors and diversification in Arbitrage pricing theory. *Econometrica* 51, 1305-1323.

Chorida, T., Goyal, A., Shanken, J.A., 2015. Cross-sectional asset pricing with individual stocks: Betas versus characteristics, Columbia Business School.

Cremers, M., Halling, M., Weinbaum, D., 2015. Aggregate jump and volatility risk in the cross-section of stock returns. *Journal of Finance* 70, 577–614.

Douglas, G.W., 1968. *Risk in the equity markets: An empirical appraisal of market efficiency.* Ann Arbor, Michigan: University Microfilms, Inc.

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

Fama, E.F., French, K.R., 2004. The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives* 18, 25–46.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81, 607–636.

Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147, 186-197.

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

Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. *Journal of Royal Statistical Society Series B 75*, 603-680.

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

Fleishman, A. I., 1978. A method for simulating non-normal distributions. *Psychometrika* 43, 521-532.

Gagliardini, P., Ossola, E., Scaillet, O., 2016. Time-varying risk premium in large cross-sectional equity datasets. *Econometrica* 84, 985–1046.

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

Gungor, S., Luger, R., 2009. Exact distribution-free tests of mean-variance efficiency. *Journal of Empirical Finance* 16, 816-829.

Gungor, S., Luger, R., 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.

Huang, J., Liu, N., Pourahmadi, M., Liu, L., 2006. Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* 93, 85–98.

Im, K.S., Pesaran, M.H., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115, 53–74.

Jarque, C.M., Bera, A.K., 1987. A test for normality of observations and regression residual. *International Statistical Review* 55, 163–172.

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

Jensen, M.C. (ed.), 1972. *Studies in the Theory of Capital Markets.* New York: Praeger.

Kelejian, H. H., and I. R. Prucha, 2001, On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics, 104, 219–257.

Ledoit, O., Wolf, M., 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. *Review of Economics and Statistics* 47, 13–37.

Longin, F., Solnik, B., 2001. Extreme correlation of international markets. *Journal of Finance* 56, 649-76.

Pesaran, M.H., Ullah, A., Yamagata, T., 2008. A bias-adjusted LM test of error cross-section independence. *Econometrics Journal* 11, 105-127.

Raponi, V., Robotti, C., Zaffaroni, P., 2016. Testing Beta-Pricing Models Using Large Cross-Sections. Imperial College London.

Ross, S.A., 1976. The arbitrage theory of capital asset pricing", *Journal of Economic Theory* 13, 341-360

Shanken, Jay, 1992. On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1-33.

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

Wong, F., Carter, C.K., Kohn, R., 2003. Efficient estimation of covariance selection models. *Biometrika* 90, 809-830.