No. 12/05

**Testing CAPM with a Large Number of Assets
By**

M Hashem Pesaran and Takashi Yamagata

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Testing CAPM with a Large Number of Assets[*]

M. Hashem Pesaran
Cambridge University and USC

Takashi Yamagata
University of York

March 12, 2012

Abstract

This paper is concerned with testing the time series implications of the capital asset pricing model (CAPM) due to Sharpe (1964) and Lintner (1965), when the number of securities, N, is large relative to the time dimension, T, of the return series. Two new tests of CAPM are proposed that exploit recent advances on the analysis of large panel data models, and are valid even if $N > T$. When the errors are Gaussian and cross sectionally independent, a test, denoted by $\hat{J}_{\alpha,1}$, is proposed which is $N(0,1)$ as $N \to \infty$, with T fixed. Even when the errors are non-Gaussian we are still able to show that $\hat{J}_{\alpha,1}$ tends to $N(0,1)$ so long as the errors are cross-sectionally independent and $N/T^3 \to 0$, with N and $T \to \infty$, jointly. In the case of cross sectionally correlated errors, using a threshold estimator of the average squares of pair-wise error correlations, a modified version of $\hat{J}_{\alpha,1}$, denoted by $\hat{J}_{\alpha,2}$, is proposed. Small sample properties of the tests are compared 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. Overall, the proposed tests perform best in terms of power, with empirical sizes very close to the chosen nominal value even in cases where N is much larger than T. The $\hat{J}_{\alpha,2}$ test (which allows for non-Gaussian and weakly cross correlated errors) is applied 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-September 2011. Statistically significant evidence against Sharpe-Lintner CAPM is found mainly during the recent financial crisis. Furthermore, a strong negative correlation is found between a twelve-month moving average p-values of the $\hat{J}_{\alpha,2}$ test and the returns of long/short equity strategies relative to the return on S&P 500 over the period December 2006 to September 2011, suggesting that abnormal profits are earned during episodes of market inefficiencies.

---

1 Introduction

This paper is concerned with testing the time series implications of the capital asset pricing model (CAPM) due to Sharpe (1964) and Lintner (1965), when the number of securities, N, under consideration is large relative to the time dimension, T, of the return series. The Sharpe-Lintner CAPM predicts that expected excess return (measured relative to the risk-free rate) on any given security or a given portfolio of securities is constant of proportionality 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 of the model 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 the individual test outcomes are difficult to interpret. Due to cross-sectional dependence of the errors in the CAPM regressions, the individual t-statistics are correlated which makes it difficult to control the overall size of the test. 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 normally distributed 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. But grouping of securities into portfolios can lead to loss of information and reduced power of the test, and under certain circumstances can also result in endogeneity problems ruled out in the multivariate GRS testing procedure. 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.

It is clearly desirable to develop tests of market efficiency that can deal with a large number of securities, ideally over short time periods so that the problem of time variations in $\beta's$ can be somewhat minimized. It is also important that such tests are reasonably robust to non-normal errors, particularly as it is more likely that one would encounter non-normal errors in the case of CAPM 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 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

---

[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 literature can be found in Jensen (1972), and more recently in Fama and French (2004).

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. In a more recent paper Gungor and Luger (2011) extend their analysis to models with more than one factor.[2]

Our primary focus in this paper is on development of multivariate tests of $\alpha_i = 0$, for $i = 1, 2, ..., N$, when $N > T$. We begin by assuming that the errors of CAPM regressions are Gaussian with a known variance-covariance matrix, $\mathbf{V}$, and use recent advances from the analysis of large panels to develop an exact T test which converges to a standard normal variate as $N \to \infty$.[3] We refer to this test as Jensen's α test of CAPM and denote it by $J_\alpha(\mathbf{V})$. To make the J_α test operational a suitable estimator of $\mathbf{V}$ is required. But when $N > T$ this is possible only if the off-diagonal elements of $\mathbf{V}$ are suitably restricted. In the case of CAPM regressions where the errors are at most weakly cross-sectionally correlated, this can be achieved by adaptive thresholding which sets to zero elements of the estimator of $\mathbf{V}$, say $\hat{\mathbf{V}}$, that are sufficiently small. Alternatively, feasible estimators of $\mathbf{V}$ can be obtained by Bayesian or classical shrinkage procedures that scale down the off-diagonal elements of $\hat{\mathbf{V}}$ relative to its diagonal elements.[4] Fan, Liao and Mincheva (2012) consider consistent estimation of $\mathbf{V}$ in the context an approximate factor model, identical to the one considered in this paper. They assume $\mathbf{V}$ is sparse and propose an adaptive thresholding estimator, $\check{\mathbf{V}}$, which they show to be positive definite with satisfactory small sample properties. However, as we show, using $\check{\mathbf{V}}$ instead of $\mathbf{V}$ in the construction of the $J_\alpha(\mathbf{V})$ test yields asymptotically valid test of CAPM only if $N \ln(N)/T \to 0$, that rules out cases where $N > T$. It seems also unlikely that the problem can be resolved by using other estimators of $\mathbf{V}$.

An alternative testing strategy that we follow in this paper is to initially ignore the off-diagonal elements of $\mathbf{V}$ and base the test of CAPM on $J_\alpha(\mathbf{D})$, where $\mathbf{D}$ is a diagonal matrix, having the same diagonal elements as $\mathbf{V}$, and then correct the resultant test statistic for the non-zero off-diagonal elements of $\mathbf{V}$. The test can be viewed as a robust version of $J_\alpha(\mathbf{V})$, in cases where the off-diagonal elements of $\mathbf{V}$ become relatively less important as $N \to \infty$. This condition prevails when the errors of CAPM are cross-sectionally weakly correlated, an assumption that underlies Chamberlain's (1983) approximate factor model formulation of the capital and arbitrage pricing models. More formally, it is established that a test based on $J_\alpha(\mathbf{D})$ will be asymptotically $N(0,1)$ for any fixed T and as $N \to \infty$, under the null hypothesis of $\alpha_i = 0$, for all i, if $N^{-1}Tr(\mathbf{R}^2)$ is bounded, where $\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{V}\mathbf{D}^{-1/2}$.

Based on these results, two new tests of CAPM are proposed that exploit recent advances on the analysis of large panel data models that are valid even if $N > T$. When the errors are Gaussian and cross-sectionally independent, a test, denoted by $\hat{J}_{\alpha,1}$, is proposed which is $N(0,1)$ as $N \to \infty$, with T fixed. Even when the errors are non-Gaussian we are still able to show that $\hat{J}_{\alpha,1}$ tends to $N(0,1)$, so long as the errors are cross-sectionally independent and $N/T^3 \to 0$, with N and $T \to \infty$, jointly. In the case of cross-sectionally

---

correlated errors, using a threshold estimator of the average squares of pair-wise error correlations, a modified version of $\hat{J}_{\alpha,1}$, denoted by $\hat{J}_{\alpha,2}$, is proposed. A distinguishing feature of our approach is that estimation of an invertible high dimensional error covariance matrix is not required.

Small sample properties of the $\hat{J}_{\alpha,1}$ and $\hat{J}_{\alpha,2}$ tests 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 test results for the following eight sample size combinations, $T = 60$ and 100, and $N = 50, 100, 200$ and 500. The $\hat{J}_{\alpha}$ tests perform remarkably well even for $N = 500$ and $T = 60$, with size very close to the chosen nominal value of 5%, and satisfactory power. Size and power of the GRS test are also compared with those of $\hat{J}_{\alpha}$ tests in the case of experiments with $N = 50 < T = 60, 100$ for which the GRS statistics can be computed. Interestingly, we find that the $\hat{J}_{\alpha}$ tests have 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) in their simulation exercise. Also see Affleck-Graves and McDonald (1990).

The $\hat{J}_{\alpha}$ tests also outperform the nonparametric sign tests of Gungor and Luger (2009) that can be implemented even if $N > T$ and are designed to deal with non-normal errors. The sign tests are shown to over-reject, for some designs quite substantially, and have lower power as compared to the $\hat{J}_{\alpha}$ tests.

Encouraged by the satisfactory performance of the $\hat{J}_{\alpha}$ tests, even in cases where N is much larger than T, we apply the $\hat{J}_{\alpha,2}$ test (which allows for both non-Gaussian and weakly cross correlated errors) to all securities in the S&P 500 index with 60 months of return data at the end of each month spanning the period September 1989 to September 2011. 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. For S&P 500 as a whole we report the $\hat{J}_{\alpha,2}$ test statistics for a single and a three Fama-French factor model over the period September 1989 to September 2011 and find statistically significant evidence against Sharpe-Lintner CAPM only during the recent financial crisis. But we find prolonged periods of low p-values for the $\hat{J}_{\alpha,2}$ test in non-crisis periods suggesting considerable time variations in the degree of market efficiency.

Finally, we examine if there exists any relationship between the p-values of the $\hat{J}_{\alpha,2}$ 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 strong negative correlation between a twelve-months moving average p-values of the $\hat{J}_{\alpha,2}$ test and excess returns of long/short equity strategies over the period December 2006 to September 2011, 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 CAPM, and discusses some of the limitations of the GRS test. Section 3 proposes the J_{α} test for large N panels, derives its asymptotic distribution. Section 4 proposes operational versions of the J_{α} test. Section 5 reports small sample properties of $\hat{J}_{\alpha}$, GRS and nonparametric sign tests. Section 6 presents the empirical applications. Some concluding remarks are provided in Section 7.

Throughout the paper, $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ will be used for the minimum and maximum eigenvalues of the $N \times N$ matrix $\mathbf{A} = (a_{ij})$, $Tr(\mathbf{A})$ for its trace, $\|\mathbf{A}\|_1 = \max_{1 \le j \le N} \left\{ \sum_{i=1}^{N} |a_{ij}| \right\}$ for its maximum absolute column sum norm, $\|\mathbf{A}\|_{\infty} = \max_{1 \le i \le N} \left\{ \sum_{j=1}^{N} |a_{ij}| \right\}$ for its

maximum absolute row sum norm, $\|\mathbf{A}\|_F = \sqrt{Tr(\mathbf{A}'\mathbf{A})}$ for its Frobenius norm, and $\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$ for its spectral norm. We also note that for a vector $\boldsymbol{\alpha}$, $\|\boldsymbol{\alpha}\| = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{1/2}$, K and ϵ denote finite and small positive constants.

2   The Panel Regression Model: Some Preliminaries and the GRS Test

The individual return regressions can be written in the form of the following panel regressions

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

where the $m \times 1$ vector of factors, $\mathbf{f}_t$, are observed. Stacking by time series observations we have

$$\mathbf{y}_{i.} = \alpha_i \boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i.}, \tag{2}$$

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

$$\mathbf{y}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \mathbf{u}_t, \tag{3}$$

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 $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$.

Assumption 1: The common factors, $\mathbf{f}_t$, are distributed independently of the errors, $u_{it'}$ for all i, t and t', $T^{-1}\mathbf{G}'\mathbf{G}$, 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 $\mathbf{M}_F = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$.

Assumption 2: $\mathbf{u}_t \sim IIDN(\mathbf{0}, \mathbf{V})$, where $\mathbf{V}$ is an $N \times N$ symmetric positive definite matrix.

Assumption 1 is standard in the literature on tests of CAPM. It is likely to be satisfied when i refers to individual securities. However, it is also often invoked when i refers to sub-market portfolios, although in such cases it is less likely to hold since the left and right hand side variables could end up with sizeable common components. Assumption 2 is needed for exact sample tests of CAPM, although as was pointed out in the Introduction, Beaulieu, Dufour and Khalaf (2007) have proposed a bootstrap procedure that allows exact test of CAPM under certain departures from error normality. Assumption 2 also implies that errors, u_{it}, are serially uncorrelated for each u_{it} and across the pair of errors u_{it} and u_{jt}, namely $E(u_{it}u_{jt'}) = 0$ for all i, j, and $t \neq t'$. A non-Gaussian version of Assumption 2 will be considered below.

Jensen's α test of CAPM are based on estimates of α_i which can be obtained efficiently by univariate OLS regressions, given by

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

This is an important result and follows since (2) 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

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

with

$$\mathbf{c} = \mathbf{M}_F\boldsymbol{\tau}_T/\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T. \tag{4}$$

Writing this in matrix notation 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

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \sum_{t=1}^{T} \mathbf{u}_t c_t, \tag{5}$$

where as before $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$. Hence, under Assumptions 1 and 2,

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

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

$$\hat{\mathbf{V}} = T^{-1} \sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}'_t, \tag{6}$$

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

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

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

$$\hat{W}_0 = \frac{T - N - m}{N} \left(\frac{\boldsymbol{\tau}'_T \mathbf{M}_F \boldsymbol{\tau}_T}{T}\right) \hat{\boldsymbol{\alpha}}' \hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\alpha}}, \tag{7}$$

which is the same as the GRS statistic (see p.1124 of GRS),

$$GRS = \hat{W}_0 = \frac{T - N - m}{N} \left(1 + \bar{\mathbf{f}}' \hat{\boldsymbol{\Omega}}^{-1} \bar{\mathbf{f}}\right)^{-1} \hat{\boldsymbol{\alpha}}' \hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\alpha}}, \tag{8}$$

where $\bar{\mathbf{f}} = T^{-1} \sum_{t=1}^{T} \mathbf{f}_t$, and $\hat{\boldsymbol{\Omega}} = T^{-1} \sum_{t=1}^{T} (\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})'$. It is easily seen that

$$\left(1 + \bar{\mathbf{f}}' \hat{\boldsymbol{\Omega}}^{-1} \bar{\mathbf{f}}\right)^{-1} = T^{-1} \left(\boldsymbol{\tau}'_T \mathbf{M}_F \boldsymbol{\tau}_T\right).$$

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. In applications where GRS is used the individual securities are grouped into portfolios and the GRS test is then typically applied to 20-30 portfolios over relatively long time periods. But grouping of individual securities is likely to result in loss of power. To see this suppose that N_p of the N securities are grouped using the portfolio weights $\mathbf{w}_p$ such that $\boldsymbol{\tau}'_N \mathbf{w}_p = 1$ for $p = 1, 2, ..., P$, and $\mathbf{w}'_p \mathbf{w}_s = 0$ for $p \neq s$, with $\sum_{p=1}^{P} N_p = N$. The GRS

test is then applied to the P portfolio excess returns defined by $\mathbf{w}_p'\mathbf{y}_t$, for $p = 1, 2, ..., P$, where P is a small fraction of N. The null hypothesis of market efficiency applied to these P portfolios is given by

$$H_0^p : \mathbf{w}_p'\boldsymbol{\alpha} = \mathbf{0}, \, p = 1, 2, ..., P$$

as compared to $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ used when all the underlying securities are considered individually. It is clear that $\boldsymbol{\alpha} = \mathbf{0}$ implies $\mathbf{w}_p'\boldsymbol{\alpha} = \mathbf{0}$, but not *vice versa*. Since α_i can take positive as well as negative values, failure to reject H_0^p does not necessarily imply that we shall also fail to reject H_0. It is quite possible that H_0 is rejected, but H_0^p is not.

A testing strategy based on portfolios rather than individual securities can also be subject to an endogeneity problem, if $N_p/N_m \rightarrow c_p$, for some p, where $c_p > 0$, and N_m is the total number of assets used to form the return on the market portfolio. In this case the excess return on the portfolio p with $c_p > 0$ will be a non-negligible component of the excess return on the market portfolio, and a regression of the former on the latter is likely to be subject to the endogeneity problem. In the case of individual securities where $N_p = 1$, the endogeneity bias will be of order $1/N_m$ which is likely to be negligible if, as Roll (1977) argues, the market portfolio is chosen to cover a sufficiently large number of assets, such that $1/N_m$ is sufficiently small.

The above difficulties can be somewhat mitigated if T and N_m can be chosen to be sufficiently large, since this allows the investigator to include a larger number of portfolios in the GRS test, without becoming unduly subject to the endogeneity problem. But due to exit and entry of firms in the market and the possibility of structural change in the parameters of the return regressions, the use of T large can present a new type of bias with unpredictable consequences for the test outcomes. Therefore, it is clearly desirable to develop tests of market efficiency that can be applied to a very large number of individual securities over relatively short time periods, 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 CAPM regressions the errors are at most weakly correlated, which places restrictions on the off-diagonal elements of $\mathbf{V}$ and its inverse.[5] As we shall see below, a test that exploits restrictions implied by the weak cross-sectional correlation of the errors is likely to have much better power properties than the GRS test that does not make use of such restrictions. It is also important to bear in mind that being a multivariate F test, the power of the GRS test is primarily driven by 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 J_α Tests for Large N Panels

Building on some of the preliminary derivations set out above, we now introduce the basic idea behind our test. To focus on the main issues involved, initially we assume that $\mathbf{V}$ is a known positive definite matrix and consider the following version of the GRS statistic, as set out in (7),

$$W_v = \left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)\hat{\boldsymbol{\alpha}}'\mathbf{V}^{-1}\hat{\boldsymbol{\alpha}}, \tag{9}$$

where we have dropped the degrees of freedom adjustment term and replaced $\hat{\mathbf{V}}$ by its true value. Consider now the asymptotic distribution of W_v for a fixed $T > m + 1$, with $N \rightarrow \infty$. Since $\mathbf{V}$ is a positive definite symmetric matrix it has the Cholesky decomposition $\mathbf{V} = \mathbf{Q}\mathbf{Q}'$, where $\mathbf{Q}$ is a non-singular $N \times N$ lower triangular matrix. The errors of CAPM

---

[5] A similar conclusion is reached under the approximate factor model advanced in Chamberlain (1983).

regressions can now be written as

$$\mathbf{u}_t = \mathbf{Q}\boldsymbol{\varepsilon}_t, \qquad (10)$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$. Using this decomposition and (5), we have

$$W_v = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\phi}}' \hat{\boldsymbol{\phi}},$$

$$\hat{\boldsymbol{\phi}} = \boldsymbol{\phi} + \sum_{t=1}^{T} c_t \boldsymbol{\varepsilon}_t. \qquad (11)$$

where $\hat{\boldsymbol{\phi}} = \mathbf{Q}^{-1}\hat{\boldsymbol{\alpha}}$, and $\boldsymbol{\phi} = \mathbf{Q}^{-1}\boldsymbol{\alpha}$. Under the null hypothesis $H_0 : \alpha_i = 0$ for all i,

$$W_v = \mathbf{x}_T' \mathbf{x}_T,$$

where

$$\mathbf{x}_T = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right)^{1/2} \sum_{t=1}^{T} c_t \boldsymbol{\varepsilon}_t, \qquad (12)$$

and recall that c_t is the tth element of $\mathbf{c}$ given by (4).

To derive the asymptotic distribution of W_v and related statistics, to be discussed below, we consider the following variant of Assumption 2, and introduce Assumption 3 which governs the degree of cross-sectional dependence in the errors, u_{it}:

Assumption 2^a: $\mathbf{u}_t \sim IID(\mathbf{0}, \mathbf{V})$, where $\mathbf{V}$ is an $N \times N$ symmetric positive definite matrix, such that $\mathbf{V} = \mathbf{Q}\mathbf{Q}'$, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})' = \mathbf{Q}^{-1}\mathbf{u}_t$. $\{\varepsilon_{it}\}$ is an IID process over i and t, with means zero and unit variances, and for some $\epsilon > 0$, $E(|\varepsilon_{it}|^{4+\epsilon})$ exists, for all i and t.

Assumption 3: $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ are non-singular lower triangular matrices with bounded absolute maximum column and row sum matrix norms.

Assumption 2^a allows the errors to be non-Gaussian. Assumption 3 ensures that the errors are weakly correlated in the sense defined by Chudik, Pesaran and Tosetti (2011). This is also in line with the assumption of approximate factor model analyzed by Chamberlain (1983). Note that under Assumption 3 $\|\mathbf{V}\|_1 = \|\mathbf{V}\|_\infty < K < \infty$, $0 < \lambda_{\min}(\mathbf{V}^{-1}) = \lambda_{\max}(\mathbf{V}) \leq \|\mathbf{V}\|_1 < K < \infty$, and $0 < \lambda_{\min}(\mathbf{V}) = \lambda_{\max}(\mathbf{V}^{-1}) \leq \|\mathbf{V}^{-1}\|_1 < K < \infty$. The above assumptions also ensure that $\sup_i(\sigma_i^2) < K < \infty$, and $\inf_i(\sigma_i^2) > 0$, for all i.[6]

As with the panel testing strategy developed in Im et al. (2003), a standardized version of W_v can now be considered. Using results in Appendix A, under Assumptions 1 and 2^a, we have[7]

$$
\begin{aligned}
E\left(W_v\right) &= N, \\
Var\left(W_v\right) &= 2N\left(1 + \frac{1}{2}\bar{\gamma}_{2,\varepsilon} q_T\right),
\end{aligned}
$$

where

$$q_T = \left(\sum_{t=1}^{T} c_t^4\right) \Big/ \left(\sum_{t=1}^{T} c_t^2\right)^2 = O_p(T^{-1}) \qquad (13)$$

$\bar{\gamma}_{2,\varepsilon} = E(\gamma_{2,\varepsilon i}) < \infty$, and $\gamma_{2,\varepsilon i} = E(\varepsilon_{it}^4) - 3$. Consider now the standardized test statistic

$$J_\alpha(\mathbf{V}) = \frac{\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{V}^{-1} \hat{\boldsymbol{\alpha}} - N}{\left(1 + \frac{1}{2}\bar{\gamma}_{2,\varepsilon} q_T\right)^{1/2} \sqrt{2N}},$$

---

[6] See, for example, Horn and Johnson (1985).

[7] The results can be obtained by setting $\mathbf{A} = \mathbf{I}_N$ in (48) and (50) in Appendix A.

and note that under $H_0 : \alpha_i = 0$, for all i,

$$J_\alpha(\mathbf{V}) = \frac{N^{-1/2} \left(\mathbf{x}_T' \mathbf{x}_T - N\right)}{\left(2 + \bar{\gamma}_{2,\varepsilon} q_T\right)^{1/2}},$$

which has mean zero and a unit variance for any N and $T > m + 1$. It is also easily established that the central limit theorem for linear-quadratic forms in Kelejian and Prucha (2001,KP) is applicable to $\mathbf{x}_T' \mathbf{x}_T$, and under Assumptions 1 and 2^a we have

$$J_\alpha(\mathbf{V}) \to_d N(0,1), \text{ as } N \to \infty, \text{ for any } T > m + 1.$$

It is interesting to note that this results does not require Assumption 3, so long as $\mathbf{V}$ is a positive definite matrix. The conditions of Theorem 1 of KP require that the ith element of $\mathbf{x}_T$ to be independent across i, and $N^{-1} Var(W_v) > K$ for some $K > 0$. Notice that in this application Assumption 2 of KP is trivially satisfied since in our application matrix $\mathbf{A}_n$ that enters KP's theorem is equal to $\mathbf{I}_N$. Note also that

$$N^{-1} Var(W_v) = 2 + \bar{\gamma}_{2,\varepsilon} q_T,$$

and under Assumption 2^a, we have $\left|\bar{\gamma}_{2,\varepsilon}\right| < \sup_i(\gamma_{2,\varepsilon i}) < \infty$. For a finite T, $q_T > 0$, and if it is further assumed that $\bar{\gamma}_{2,\varepsilon} > 0$, (which is likely to be met in most finance applications) then, $N^{-1} Var(W_v) > K > 0$, as required. Also noting that $E(\varepsilon_{it}^4) > 0$, then $\bar{\gamma}_{2,\varepsilon} > -3$, and since $q_T > 0$ and of order T^{-1}, then we would expect $2 + \bar{\gamma}_{2,\varepsilon} q_T > 0$, when T is relatively large.

The J_α test allows the degree of non-Gaussianity to vary across securities, with the test depending on the average excess kurtosis, $\bar{\gamma}_{2,\varepsilon}$, and not the individual kurtosis, $\gamma_{2,\varepsilon i}$. In the case where the errors are Gaussian, $\bar{\gamma}_{2,\varepsilon} = 0$ and the denominator of J_α statistic reduces to $\sqrt{2N}$. Also since $q_T = O_p(T^{-1})$, then as be expected, the dependence of J_α on $\bar{\gamma}_{2,\varepsilon}$ declines with T.

To obtain a better understanding of the asymptotic distribution of J_α, consider the Gaussian case and note that under Gaussian errors

$$J_\alpha(\mathbf{V}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{x_{iT}^2 - 1}{\sqrt{2}}\right)$$

where x_{iT} is the ith element of $\mathbf{x}_T$. In this case x_{iT}, being a linear combination of Gaussian errors, it is distributed as $N(0,1)$, independently of x_{jT} for all $i \neq j$. Then $z_{iT} = \frac{1}{\sqrt{2}}\left(x_{iT}^2 - 1\right)$ is a $(0,1)$ random variable, which is also distributed independently of z_{jT} for $i \neq j$. Hence, for a *fixed* $T > m + 1$ and as $N \to \infty$, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{x_{iT}^2 - 1}{\sqrt{2}}\right) \to_d N(0,1). \tag{14}$$

The above results are summarized in the following theorem.

Theorem 1 *Consider the CAPM regressions, (1), suppose Assumptions 1 and 2^a hold, and $\mathbf{V}$ is known. Then under $H_0 : \alpha_i = 0$ for all i, and as $N \to \infty$, for any $T > m + 1$*

$$J_\alpha(\mathbf{V}) = \frac{N^{-1/2} \left[\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{V}^{-1} \hat{\boldsymbol{\alpha}} - N\right]}{\left(2 + \bar{\gamma}_{2,\varepsilon} q_T\right)^{1/2}} \to_d N(0,1), \tag{15}$$

where $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_N)'$, $\hat{\alpha}_i$ is the OLS estimator of α_i in (1), $\bar{\gamma}_{2,\varepsilon} = E(\gamma_{2,\varepsilon i}) > 0$, $\gamma_{2,\varepsilon i} = E(\varepsilon_{it}^4) - 3$, $\boldsymbol{\varepsilon}_t = \mathbf{Q}^{-1} \mathbf{u}_t$, and $q_T = O_p(T^{-1})$ is defined by (13).

Using a similar line of reasoning, it is also relatively easy to derive the asymptotic distribution of J_α under the alternative hypothesis, $H_1 : \alpha_i \neq 0$, for some i. Using (11) we have

$$\hat{\boldsymbol{\alpha}}' \mathbf{V}^{-1} \hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\phi}}' \hat{\boldsymbol{\phi}} = \left(\boldsymbol{\phi} + \sum\nolimits_{t=1}^{T} c_t \boldsymbol{\varepsilon}_t \right)' \left(\boldsymbol{\phi} + \sum\nolimits_{t=1}^{T} c_t \boldsymbol{\varepsilon}_t \right),$$

and

$$J_\alpha(\mathbf{V}) = J_{1,\alpha}(\mathbf{V}) + \frac{N^{1/2} \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right) \boldsymbol{\phi}' \boldsymbol{\phi}}{\left(2 + \bar{\gamma}_{2,\varepsilon} q_T \right)^{1/2}},$$

where

$$J_{1,\alpha}(\mathbf{V}) = \frac{\mathbf{x}_T' \mathbf{x}_T + 2 \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right)^{1/2} \boldsymbol{\phi}' \mathbf{x}_T - N}{\left(1 + \frac{1}{2} \bar{\gamma}_{2,\varepsilon} q_T \right)^{1/2} \sqrt{2N}}.$$

Applying the KP central limit theorem for quadratic-linear forms to $J_{1,\alpha}(\mathbf{V})$, it is then easily established that for a fixed T and $N \to \infty$, $J_{1,\alpha}(\mathbf{V}) \to_d N(0, \omega_{1,\alpha}^2)$, where,[8]

$$\omega_{1,\alpha}^2 = 1 + \frac{4 \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right)}{\left(2 + \bar{\gamma}_{2,\varepsilon} q_T \right)} \left[\left(\frac{\boldsymbol{\phi}' \boldsymbol{\phi}}{N} \right) + \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right) \left(\sum\nolimits_{t=1}^{T} c_t^3 \right) \left(\frac{\sum_{i=1}^{N} \phi_i E \left(\varepsilon_{it}^3 \right)}{N} \right) \right],$$

which for a fixed T tends to a finite limit as $N \to \infty$. Hence, $J_\alpha(\mathbf{V})$ diverges under H_1 at the rate of $N^{-1/2} \boldsymbol{\alpha}' \mathbf{V}^{-1} \boldsymbol{\alpha}$. But, $N^{-1/2} \boldsymbol{\alpha}' \mathbf{V}^{-1} \boldsymbol{\alpha} \geq \left(N^{-1/2} \boldsymbol{\alpha}' \boldsymbol{\alpha} \right) \lambda_{\min}(\mathbf{V}^{-1})$, and since under Assumption 3, $0 < \lambda_{\min}(\mathbf{V}^{-1}) < K < \infty$, then $J_\alpha(\mathbf{V})$ test must have power against alternatives

$$H_1 : \sum\nolimits_{i=1}^{N} \alpha_i^2 = \boldsymbol{\alpha}' \boldsymbol{\alpha} = O(N^{\delta_\alpha}), \tag{16}$$

so long as $\delta_\alpha > 1/2$.[9] This specification requires that α_i is non-zero for a sufficiently large number of securities. It does not require that α_i to be non-zero for all securities under consideration.

3.1   A J_α test based on an adaptive thresholding estimator of $\mathbf{V}$

To make the J_α test operational 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 CAPM 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 $\mathbf{V}$. In their recent paper, Fan, Liao and Mincheva (2012) consider consistent estimation of $\mathbf{V}$ in the context of an approximate factor model, identical to the one considered in this paper. They assume $\mathbf{V}$ is sparse and propose an adaptive thresholding estimator, $\check{\mathbf{V}}$, which they show to be positive definite with satisfactory small sample properties. In particular, they assume that the rows of $\mathbf{V}$ have only a finite number of non-zero elements which is in line with our requirement that the maximum absolute row sum norm of $\mathbf{V}$ is bounded in N. Fan, Liao and Mincheva show that

$$\left\| \check{\mathbf{V}}^{-1} - \mathbf{V}^{-1} \right\| = O_p \left((m+1) r \sqrt{\frac{\ln(N)}{T}} \right), \tag{17}$$

---

[8] This result can be obtained using (53) with $\mathbf{A} = \mathbf{I}_N$ in Appendix A.

[9] In view of this result, the null hypothesis of Theorem 1 can be written more generally as $H_0 : \sum_{i=1}^{N} \alpha_i^2 = O(N^{\delta_\alpha})$, for $\delta_\alpha < 1/2$.

where m is the number of factors, and r is the maximum number of non-zero elements in the rows of $\mathbf{V}$. Using $\check{\mathbf{V}}^{-1}$ we have the following operational version of the J_α test (assuming $\bar{\gamma}_{2,\varepsilon}$ is known)

$$J_\alpha(\check{\mathbf{V}}) = \frac{\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \check{\mathbf{V}}^{-1} \hat{\boldsymbol{\alpha}} - N}{\left(1 + \frac{1}{2}\bar{\gamma}_{2,\varepsilon} q_T\right)^{1/2} \sqrt{2N}} = J_\alpha + \frac{\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \left(\check{\mathbf{V}}^{-1} - \mathbf{V}^{-1}\right) \hat{\boldsymbol{\alpha}}}{\left(1 + \frac{1}{2}\bar{\gamma}_{2,\varepsilon} q_T\right)^{1/2} \sqrt{2N}},$$

and $J_\alpha(\check{\mathbf{V}}) \to_d N(0,1)$, if (recalling that $\bar{\gamma}_{2,\varepsilon} q_T = O_p(T^{-1})$)

$$\left\| \frac{\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \left(\check{\mathbf{V}}^{-1} - \mathbf{V}^{-1}\right) \hat{\boldsymbol{\alpha}}}{\sqrt{N}} \right\| \to_p 0.$$

But

$$\left\| \frac{\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \left(\check{\mathbf{V}}^{-1} - \mathbf{V}^{-1}\right) \hat{\boldsymbol{\alpha}}}{\sqrt{N}} \right\| < \frac{1}{\sqrt{N}} \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \|\hat{\boldsymbol{\alpha}}\|^2 \left\| \check{\mathbf{V}}^{-1} - \mathbf{V}^{-1} \right\|,$$

and under $\boldsymbol{\alpha} = \mathbf{0}$ (and in view of (5))

$$\|\hat{\boldsymbol{\alpha}}\|^2 = \hat{\boldsymbol{\alpha}}' \hat{\boldsymbol{\alpha}} = \left(\sum_{t=1}^T \mathbf{u}_t c_t\right)' \left(\sum_{t=1}^T \mathbf{u}_t c_t\right) = Tr(\mathbf{V}) O_p \left(\sum_{t=1}^T c_t^2\right).$$

Using this result together with (17), and recalling that $\left(\sum_{t=1}^T c_t^2\right) \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) = 1$, and $Tr(\mathbf{V}) = O(N)$, we now have

$$\left\| \frac{\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \left(\check{\mathbf{V}}^{-1} - \mathbf{V}^{-1}\right) \hat{\boldsymbol{\alpha}}}{\sqrt{N}} \right\| < O_p \left((m+1)r\sqrt{\frac{N\ln(N)}{T}}\right).$$

Therefore, $J_\alpha(\check{\mathbf{V}}) \to_d N(0,1)$, if $N\ln(N)/T \to 0$. However, this is not good enough since our objective is to construct tests that are valid when $N > T$, whilst the condition $N\ln(N)/T \to 0$ requires just the reverse. It is not clear to us if the use of other available estimators of $\mathbf{V}$ can overcome the curse of testing $\boldsymbol{\alpha} = \mathbf{0}$, when $N > T$. Instead, we consider other formulations of the J_α test that do not require an estimate of $\mathbf{V}^{-1}$.

3.2 A J_α test based on the diagonal elements of V

In view of the above results and considering that in an approximate factor model we would expect the off-diagonal elements of $\mathbf{V}$ to become relatively less important as $N \to \infty$, in what follows we consider J_α type tests that are based on an $N \times N$ diagonal matrix, $\mathbf{D} = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_N^2)$ with $\sigma_i^2 = E\left(u_{it}^2\right)$, rather than the full covariance matrix.

Let $W_d = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{D}^{-1} \hat{\boldsymbol{\alpha}}$, and consider the standardized test statistic

$$J_\alpha(\mathbf{D}) = \frac{W_d - E(W_d)}{\sqrt{Var(W_d)}}.$$

Initially, assume that $\mathbf{D}$ is known and note that under $\boldsymbol{\alpha} = \mathbf{0}$,

$$\begin{aligned} W_d &= \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \left(\sum_{t=1}^T \mathbf{u}_t c_t\right)' \mathbf{D}^{-1} \left(\sum_{t=1}^T \mathbf{u}_t c_t\right) \\ &= \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \left(\sum_{t=1}^T \mathbf{Q}\boldsymbol{\varepsilon}_t c_t\right)' \mathbf{D}^{-1} \left(\sum_{t=1}^T \mathbf{Q}\boldsymbol{\varepsilon}_t c_t\right) \\ &= \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \left(\sum_{t=1}^T c_t \boldsymbol{\varepsilon}_t'\right) \mathbf{Q}' \mathbf{D}^{-1} \mathbf{Q} \left(\sum_{t=1}^T c_t \boldsymbol{\varepsilon}_t\right) \\ &= \mathbf{x}_T' \left(\mathbf{Q}' \mathbf{D}^{-1} \mathbf{Q}\right) \mathbf{x}_T. \end{aligned}$$

Therefore, setting $\mathbf{A} = \mathbf{Q}'\mathbf{D}^{-1}\mathbf{Q}$ in equations (48) and (50) of Appendix A we have

$$J_{\alpha}(\mathbf{D}) = \frac{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)\hat{\boldsymbol{\alpha}}'\mathbf{D}^{-1}\hat{\boldsymbol{\alpha}} - Tr(\mathbf{A})}{\sqrt{2Tr(\mathbf{A}^2) + \left(\sum_{i=1}^{N} a_{ii}^2 \gamma_{2,\varepsilon i}\right) q_T}}, \tag{18}$$

where a_{ii} is the ith diagonal element of $\mathbf{A}$, $\gamma_{2,\varepsilon i} = E(\varepsilon_{it}^4) - 3$, and q_T is given by (13). It is easily seen that $Tr(\mathbf{A}) = Tr(\mathbf{R}) = N$, and $Tr(\mathbf{A}^2) = Tr(\mathbf{R}^2)$, where $\mathbf{R} = (\rho_{ij})$ is the correlation matrix of the errors, $\mathbf{u}_t$.

To apply the Kelejian and Prucha's Theorem 1 to $J_{\alpha}(\mathbf{D})$ we note that under the null hypothesis, $\boldsymbol{\alpha} = \mathbf{0}$,

$$J_{\alpha}(\mathbf{D}) = \frac{\left(\mathbf{x}_T'\mathbf{A}\mathbf{x}_T - N\right)}{\sqrt{2Tr(\mathbf{A}^2) + \left(\sum_{i=1}^{N} a_{ii}^2 \gamma_{2,\varepsilon i}\right) q_T}},$$

where the elements of $\mathbf{x}_T$ are distributed independently across i, have zero means, and finite fourth-order moments. Therefore, Assumption 1 of KP's Theorem is satisfied. Also, Assumption 2 of KP is met since

$$\left\|\mathbf{Q}'\mathbf{D}^{-1}\mathbf{Q}\right\|_1 \leq \left\|\mathbf{Q}'\right\|_1 \left\|\mathbf{D}^{-1}\right\|_1 \left\|\mathbf{Q}\right\|_1 \leq \sup_i(\frac{1}{\sigma_i^2}) \left\|\mathbf{Q}\right\|_1 \left\|\mathbf{Q}\right\|_\infty,$$

which is bounded in N under our Assumption 3. Finally, KP require that $N^{-1}Var(W_d) \geq c$ for some $c > 0$. In the case of the present application

$$N^{-1}Var(W_d) = 2\left[N^{-1}Tr(\mathbf{A}^2)\right] + \left(N^{-1}\sum_{i=1}^{N} a_{ii}^2 \gamma_{2,\varepsilon i}\right) q_T, \tag{19}$$

where $q_T > 0$. But

$$N^{-1}Tr(\mathbf{A}^2) = N^{-1}Tr(\mathbf{R}\mathbf{R}') = N^{-1}\left\|\mathbf{R}\right\|_F^2$$

and since $\mathbf{R}$ is a full rank matrix, the following inequality holds between its Frobenius and spectral norms

$$\left\|\mathbf{R}\right\|^2 \leq \left\|\mathbf{R}\right\|_F^2 \leq N\left\|\mathbf{R}\right\|^2.$$

Hence

$$N^{-1}Tr(\mathbf{A}^2) = N^{-1}\left\|\mathbf{R}\right\|_F^2 \leq \left\|\mathbf{R}\right\|^2 = \left\|\mathbf{D}^{-1/2}\mathbf{Q}\mathbf{Q}'\mathbf{D}^{-1/2}\right\|^2 \leq \left\|\mathbf{D}\right\|_1^{-2} \left\|\mathbf{Q}\right\|_1^2 \left\|\mathbf{Q}\right\|_\infty^2,$$

which is bounded in N under Assumption 3. Furthermore,

$$Tr(\mathbf{R}^2) = \sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{ij}^2 = N\left[1 + (N-1)\overline{\rho^2}\right], \tag{20}$$

where

$$\overline{\rho^2} = \frac{2}{N(N-1)} \sum_{i=2}^{N}\sum_{j=1}^{i-1} \rho_{ij}^2 \geq 0. \tag{21}$$

Therefore, since under Assumption 3 $N^{-1}Tr(\mathbf{R}^2)$ is bounded in N, it also follows that $(N-1)\overline{\rho^2} < K < \infty$. Furthermore, since $N^{-1}\sum_{i=1}^{N} a_{ii}^2 \leq N^{-1}Tr(\mathbf{A}^2) = N^{-1}Tr(\mathbf{R}^2) = 1 + (N-1)\overline{\rho^2}$, we have

$$\left|N^{-1}\sum_{i=1}^{N} a_{ii}^2 \gamma_{2,\varepsilon i}\right| < N^{-1}\sum_{i=1}^{N} a_{ii}^2 \left|\gamma_{2,\varepsilon i}\right| < \sup_i(\left|\gamma_{2,\varepsilon i}\right|)N^{-1}\sum_{i=1}^{N} a_{ii}^2 < \sup_i(\left|\gamma_{2,\varepsilon i}\right|)\left[1 + (N-1)\overline{\rho^2}\right].$$

Using the above results in (19), it follows that $N^{-1}Var(W_d)$ is bounded in N, and we have $\lim_{N\to\infty} N^{-1}Var(W_d) \geq c > 0$, as required by KP's Theorem 1 if $\bar{\gamma}_{2,\varepsilon} > 0$.

Theorem 2 *Consider the CAPM regressions, (1), suppose Assumptions 1, 2^a, and 3 hold, and $\mathbf{V}$ is known. Then under $H_0 : \alpha_i = 0$ for all i, and as $N \to \infty$ for any $T > m + 1$*

$$J_\alpha(\mathbf{D}) = \frac{N^{-1/2} \left[(\boldsymbol{\tau}'_T \mathbf{M}_F \boldsymbol{\tau}_T) \, \hat{\boldsymbol{\alpha}}' \mathbf{D}^{-1} \hat{\boldsymbol{\alpha}} - N \right]}{\sqrt{2N^{-1} Tr(\mathbf{R}^2) + \left(N^{-1} \sum_{i=1}^{N} a_{ii}^2 \gamma_{2,\varepsilon i} \right) q_T}} \to_d N(0,1), \qquad (22)$$

where $\mathbf{V} = \mathbf{D}^{-1/2} \mathbf{R} \mathbf{D}^{-1/2}$, and as in Theorem 1 $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_N)'$, $\hat{\alpha}_i$ is the OLS estimator of α_i in (1), $\bar{\gamma}_{2,\varepsilon} = E(\gamma_{2,\varepsilon i}) > 0$, $\gamma_{2,\varepsilon i} = E(\varepsilon_{it}^4) - 3$, $\boldsymbol{\varepsilon}_t = \mathbf{Q}^{-1} \mathbf{u}_t$, and $q_T = O_p(T^{-1})$ is defined by (13).

Some of the conditions of the above theorem can be relaxed, but it serves our purpose of showing that an asymptotically valid test of $\boldsymbol{\alpha} = \mathbf{0}$ can be constructed in the large N case without needing to make use of the inverse of $\mathbf{V}$, so long as the cross correlation of the errors is not too strong, namely if $(N - 1)\overline{\rho^2}$ is bounded in N. But, this latter condition holds for a wide class of cross-sectionally weakly correlated errors.[10] To see this suppose that the errors have the following multi-factor structure

$$u_{it} = \boldsymbol{\varphi}'_i \mathbf{h}_t + \omega_i \varepsilon_{it},$$

where $\mathbf{h}_t$ is the $r \times 1$ vector of unobserved common factors, $\boldsymbol{\varphi}_i$ is the associated $r \times 1$ vector of factor loadings, $\varepsilon_{it} \sim iid(0,1)$ are the purely idiosyncratic components of u_{it}, and r is assumed to be fixed. To ensure that u_{it} are serially uncorrelated with mean zero and variance σ_i^2, without loss of generality, we further assume that $\mathbf{h}_t \sim IID(\mathbf{0}, \mathbf{I}_r)$, and note that $\sigma_i^2 = \boldsymbol{\varphi}'_i \boldsymbol{\varphi}_i + \omega_i^2$. The degree of cross-sectional dependence is measured by δ_b defined by (see Chudik, Pesaran and Tosetti, 2011, and Bailey, Kapetanious and Pesaran, 2012) $\sum_{i=1}^{N} \mathbf{b}'_i \mathbf{b}_i = O(N^{\delta_b})$, where $\mathbf{b}_i = \boldsymbol{\varphi}_i / \sigma_i$. The parameter δ_b is referred to as the exponent of cross-sectional dependence. In the standard factor literature it is assumed that $\delta_b = 1$, which corresponds to the case of strong cross-sectional dependence. It is now easily established that in fact $\lim_{N\to\infty}(N - 1)\overline{\rho^2} = 0$ for all values of $\delta_b < 1/2$, and we have the following general theorem which sets out empirically verifiable assumptions and conditions under which the $J_\alpha(\mathbf{D})$ test is applicable. In the borderline case of $\delta_b = 1/2$, $(N - 1)\overline{\rho^2}$ tends to a finite non-zero limit and in implementation of $J_\alpha(\mathbf{D})$ test knowledge of $\overline{\rho^2}$ is required. See Section 4.3.

Theorem 3 *Consider the CAPM regressions, (1), suppose Assumptions 1, 2^a and 3 hold, $E(u_{it}^2) = \sigma_i^2$, $i = 1, 2, ..., N$ are known, the errors, u_{it}, follow the multi-factor model*

$$u_{it} = \boldsymbol{\varphi}'_i \mathbf{h}_t + \omega_i \varepsilon_{it}, \qquad (23)$$

where $\varepsilon_{it} \sim iid(0,1)$, $\mathbf{h}_t$ is the $r \times 1$ vector of unobserved factors distributed as $\mathbf{h}_t \sim IID(\mathbf{0}, \mathbf{I}_r)$, $\boldsymbol{\varphi}_i$ is the $r \times 1$ vector of factor loadings, r is fixed, $\sigma_i^2 = \boldsymbol{\varphi}'_i \boldsymbol{\varphi}_i + \omega_i^2 < \infty$, for all i, and

$$\sum_{i=1}^{N} \mathbf{b}'_i \mathbf{b}_i = O(N^{\delta_b}), \text{ for some } \delta_b < 1/2, \qquad (24)$$

where $\mathbf{b}_i = \boldsymbol{\varphi}_i / \sigma_i$. Further assume that $\gamma_{2,\varepsilon i}$ and a_{ii} are independently distributed, where $\gamma_{2,\varepsilon i} = E(\varepsilon_{it}^4) - 3$, and a_{ii} is the ith diagonal element of $\mathbf{A} = \mathbf{Q}' \mathbf{D}^{-1} \mathbf{Q}$. Then under $H_0 : \alpha_i = 0$ for all i, and for any $T > m + 1$, as $N \to \infty$

$$J_\alpha(\mathbf{D}) = \frac{N^{-1/2} \left[(\boldsymbol{\tau}'_T \mathbf{M}_F \boldsymbol{\tau}_T) \, \hat{\boldsymbol{\alpha}}' \mathbf{D}^{-1} \hat{\boldsymbol{\alpha}} - N \right]}{\sqrt{2 + \bar{\gamma}_{2,\varepsilon} q_T}} \to_d N(0,1), \qquad (25)$$

---

[10] In cases where the errors of CAPM model are not weakly corrected but (semi-)strongly correlated, it can be argued that important factors are missing from the CAPM regressions.

where $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_N)'$, $\hat{\alpha}_i$ is the OLS estimator of α_i in (1), $\bar{\gamma}_{2,\varepsilon} = E(\gamma_{2,\varepsilon i})$, $\gamma_{2,\varepsilon i} = E(\varepsilon_{it}^4) - 3$, $\varepsilon_t = \mathbf{Q}^{-1}\mathbf{u}_t$, and $q_T = O_p(T^{-1})$ is defined by (13). When $\delta_b = 1/2$, $\lim_{N \to \infty}(N - 1)\overline{\rho^2}$ exists but does not vanish, and the term $(N-1)\overline{\rho^2}$ must be included in the denominator of $J_\alpha(\mathbf{D})$, as in (22). The $J_\alpha(\mathbf{D})$ test has power for a fixed $T > m + 1$, against alternatives H_1 defined by (16), with $\delta_\alpha > 1/2$.

A proof is provided in Appendix B.

Consider now the asymptotic power of $J_\alpha(\mathbf{D})$ test. Using the results in Appendix A it is easily seen that power function of $J_\alpha(\mathbf{D})$ test diverges at the rate of $N^{-1/2}\boldsymbol{\alpha}'\mathbf{D}^{-1}\boldsymbol{\alpha}$, and as with the $J_\alpha(\mathbf{V})$ test, it has power against alternatives defined by (16) for all values of the cross-sectional exponent, $\delta_\alpha > 1/2$. But, not surprisingly, $J_\alpha(\mathbf{V})$ test that makes use of the off-diagonal elements of the error covariance matrix will exhibit higher power, although the differential in the power function of the two tests is likely to diminish as $N \to \infty$, under the assumption that the errors, u_{it}, are cross-sectionally weakly correlated (with $\delta_b < 1/2$). The relative efficiency of the $J_\alpha(\mathbf{V})$ test over the $J_\alpha(\mathbf{D})$ test depends on the size of $\lambda_{\max}(\mathbf{R})$. The higher $\lambda_{\max}(\mathbf{R})$, the higher will be the power superiority of $J_\alpha(\mathbf{V})$ over the $J_\alpha(\mathbf{D})$ test. But, this gain in power is unlikely to be attainable in practice where a sufficiently accurate estimate of $\mathbf{V}^{-1}$ can not be obtained, as argued in Section 3.1.

The use of $J_\alpha(\mathbf{D})$ is also complicated by the fact that error variances σ_i^2, and $\bar{\gamma}_{2,\varepsilon}$ must be replaced by their estimators. In what follows we focus on alternative operational versions of the $J_\alpha(\mathbf{D})$ test.

4 Two Operational Versions of the J_α Test

In view of the above theoretical results we focus on the J_α test, by considering the following estimate of W_d

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

where $\hat{\mathbf{D}}_v = \left(v^{-1}T\right)diag(\hat{\sigma}_1^2, \hat{\sigma}_2^2, ..., \hat{\sigma}_N^2)$, $\hat{\sigma}_i^2 = \hat{\mathbf{u}}_{i.}'\hat{\mathbf{u}}_{i.}/T$, and the degrees of freedom $v = T - m - 1$ is introduced to correct for small sample bias of the test. It is then easily seen that

$$\hat{W}_d = \sum_{i=1}^{N} t_i^2, \tag{26}$$

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

$$t_i^2 = \frac{\hat{\alpha}_i^2(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)}{v^{-1}T\hat{\sigma}_i^2}. \tag{27}$$

As with the $J_\alpha(\mathbf{D})$ test statistic, which is defined by (18), a standardized version of $\hat{W}_d$ can now be considered

$$\hat{J}_\alpha = \frac{\hat{W}_d - E\left(\hat{W}_d\right)}{\sqrt{Var\left(\hat{W}_d\right)}}. \tag{28}$$

We examine the properties of $\hat{J}_\alpha$ under each of the four combinations of assumptions regarding the distribution of the errors and the strength of their cross-sectional correlations: (i) Gaussian cross-sectionally uncorrelated errors, (ii) (possibly) non-Gaussian and

cross-sectionally independent errors, (iii) Gaussian and (possibly) cross-sectionally correlated errors, (iv) non-Gaussian (possibly) and cross-sectionally correlated (possibly) errors. Firstly, we introduce a test, which we denote by $\hat{J}_{\alpha,1}$ under Gaussian cross-sectionally uncorrelated errors, which turns out to be robust to non-Gaussianity or weak cross-sectional error correlation, such that $\delta_b < 1/2$ (defined by (24)). Then, we introduce another variant, called $\hat{J}_{\alpha,2}$ test, which is expected to be robust even to moderately strong degrees of error cross-sectional correlations, namely when δ_b is close to or even slightly above $1/2$.

4.1 Gaussian and cross-sectionally uncorrelated errors

We begin with the basic case where the errors are Gaussian and uncorrelated, and then consider different ways that these assumptions could be relaxed. In this case where the errors, u_{it}, are normally distributed the individual t_i statistics are distributed as Student t with v degrees of freedom, and we have (assuming $v = T - m - 1 > 4$)

$$E(t_i^2) = \frac{v}{v-2}, \; E(t_i^4) = \frac{3v^2}{(v-2)(v-4)}$$

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

Therefore,

$$E\left(\hat{W}_d\right) = \sum_{i=1}^{N} E\left(t_i^2\right) = \frac{vN}{v-2},$$

and since the errors are cross-sectionally uncorrelated

$$Var\left(\hat{W}_d\right) = \frac{2N(v-1)}{(v-4)} \left(\frac{v}{v-2}\right)^2.$$

Thus we have the following exactly standardized test statistic

$$\hat{J}_{\alpha,1} = \frac{N^{-1/2} \sum_{i=1}^{N} \left(t_i^2 - \frac{v}{v-2}\right)}{\left(\frac{v}{v-2}\right) \sqrt{\frac{2(v-1)}{(v-4)}}}, \tag{29}$$

which is distributed as $N(0,1)$ under the $H_0 : \alpha_i = 0$, for all i, as $N \to \infty$, for any $T > m + 5$. The proof follows by application of the Lindeberg–Lévy central limit theorem directly to $N^{-1/2} \sum_{i=1}^{N} \left(t_i^2 - \frac{v}{v-2}\right)$, noting that t_i^2 has a finite second order moment since by assumption $v = T - m - 1 > 4$.

4.2 Non-Gaussian and cross-sectionally independent errors

Maintaining the error independence assumption, suppose now that the errors u_{it} are non-Gaussian with $\gamma_{1,i} = E(u_{it}^3)/\sigma_i^3$ and $\gamma_{2,i} = [E(u_{it}^4) - 3]/\sigma_i^4$.[11] 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. In order to deal with these problems, we need to slightly strengthen Assumptions 1 and 2^a, by further assuming that: (i) $\mathbf{f}_t'\mathbf{f}_t \leq K < \infty$ for all t, and; (ii) $E(|\varepsilon_{it}|^{8+\epsilon}) < K < \infty$ for some $\epsilon > 0$. Assumption (i) ensures that each element of $\mathbf{M}_F \boldsymbol{\tau}_T$ is bounded in absolute value. Assumption (ii) enables us to obtain higher order approximations for the mean and variance of t_i^2.

---

[11] Observe that under the error independence assumption, $\gamma_{2,i} = \gamma_{2,\varepsilon i}$.

Using a slightly extended version of Laplace approximation of moments of ratio of quadratic forms by Lieberman (1994), we can obtain $E(t_i^2)$ and $Var(t_i^2)$ up to $O(v^{-3/2})$, given $\mathbf{F}$ (see Proposition 16 proved in Appendix D):

$$E\left(t_i^2\right) = \frac{v}{v-2} + \gamma_{2,i}\frac{S_{0v}}{v^{3/2}} + \frac{W_{0,iv}}{v^2} \tag{30}$$

$$Var\left(t_i^2\right) = \left(\frac{v}{v-2}\right)^2 \frac{2\left(v-1\right)}{\left(v-4\right)} + \frac{\gamma_{1,i}^2 S_{1v} + \gamma_{2,i} S_{2v}}{v} + \frac{\gamma_{1,i}^2 S_{3v} + \gamma_{2,i} S_{4v}}{v^{3/2}} + \frac{W_{1,iv}}{v^2} \tag{31}$$

where

$$S_{0v} = \frac{v^{1/2} Tr\left(\mathbf{P}_G \odot \mathbf{H}_F\right)}{\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T},$$

$$S_{1v} = -\frac{8\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{H}_F\left(\mathbf{I}_T \odot \mathbf{M}_G\right)\boldsymbol{\tau}_T\right]}{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}, \; S_{2v} = vq_T - 3,$$

$$S_{3v} = -\frac{v^{1/2}12\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{M}_G\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\boldsymbol{\tau}_T\right]}{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}, \; S_{4v} = -\frac{v^{1/2}10Tr\left(\mathbf{P}_G \odot \mathbf{H}_F\right)}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T},$$

$\odot$ represents the Hadamard (or element-wise) product matrix operator, $\mathbf{H}_F = \mathbf{M}_F \boldsymbol{\tau}_T \boldsymbol{\tau}_T' \mathbf{M}_F$, $\mathbf{M}_G = \mathbf{I}_T - \mathbf{P}_G$, $\mathbf{P}_G = \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'$ with $\mathbf{G} = (\mathbf{F}, \boldsymbol{\tau}_T)$, S_{jv}, $j = 0, 1, 2, 3, 4$, $W_{0,iv}$ and $W_{1,iv}$ are $O(1)$ functions of $\mathbf{F}$, v and $\gamma_{\ell,i}$, $\ell = 1, 2, ..., 6$, which are at most $O(1)$, are given by $\gamma_{1,i} = E(\xi_{it}^3)$, $\gamma_{2,i} = E(\xi_{it}^4) - 3$, $\gamma_{3,i} = E(\xi_{it}^5) - 10\gamma_{1,i}$, $\gamma_{4,i} = E(\xi_{it}^6) - 10\gamma_{1,i}^2 - 15\gamma_{2,i} - 15$, $\gamma_{5,i} = E\left(\xi_{it}^7\right) - 21\gamma_{3,i} - 35\gamma_{2,i}\gamma_{1,i} - 105\gamma_{1,i}$, $\gamma_{6,i} = E\left(\xi_{it}^8\right) - 28\gamma_{4,i} - 56\gamma_{3,i}\gamma_{1,i} - 35\gamma_{2,i}^2 - 210\gamma_{2,i} - 280\gamma_{1,i}^2 - 105$, where $\xi_{it} = u_{it}/\sigma_i$. These results can now be used to establish the following theorem:

Theorem 4 *Consider the regression model (1), and suppose that Assumptions 1, 2^a and 3 hold. Further assume that: (i) $\mathbf{f}_t' \mathbf{f}_t \leq K < \infty$ for all t, and; (ii) $E(|\varepsilon_{it}|^{8+\epsilon}) < K < \infty$ for some $\epsilon > 0$. Consider the statistic, $\hat{J}_{\alpha,1}$, defined by (29). Then, under $H_0 : \alpha_i = 0$ for all i, $\hat{J}_{\alpha,1} \rightarrow_d N(0,1)$, if $N/T^3 \rightarrow 0$, as $N \rightarrow \infty$ and $T \rightarrow \infty$, jointly.*

Proof. Consider $\hat{J}_\alpha$ defined by (28), and note that under Assumption 1 it can be written as

$$\hat{J}_\alpha = \frac{N^{-1/2}\sum_{i=1}^{N}\left[t_i^2 - E(t_i^2)\right]}{\sqrt{N^{-1}\sum_{i=1}^{N}Var(t_i^2)}}.$$

Substituting for $E(t_i^2)$ and $Var(t_i^2)$ from (30) and (31), respectively, we have

$$\hat{J}_\alpha = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{v}{v-2}\right) + N^{1/2}\bar{\gamma}_2\frac{S_{0v}}{v^{3/2}} + N^{1/2}\frac{\bar{W}_{0v}}{v^2}}{\sqrt{\left(\frac{v}{v-2}\right)^2\frac{2(v-1)}{(v-4)} + \frac{\overline{\gamma_1^2}S_{1v} + \bar{\gamma}_2 S_{2v}}{v} + \frac{\overline{\gamma_1^2}S_{3v} + \bar{\gamma}_2 S_{4v}}{v^{3/2}} + \frac{\bar{W}_{1v}}{v^2}}},$$

where $\overline{\gamma_1^2} = N^{-1}\sum_{i=1}^{N}\gamma_{1,i}^2$, $\bar{\gamma}_2 = N^{-1}\sum_{i=1}^{N}\gamma_{2,i}$, $\bar{W}_{0v} = N^{-1}\sum_{i=1}^{N}W_{0,iv}$, and $\bar{W}_{1v} = N^{-1}\sum_{i=1}^{N}W_{1,iv}$. Under our assumptions, $\overline{\gamma_1^2}$ and $\bar{\gamma}_2$, being averages of bounded sequences, are $O(1)$, and as established in Appendix D, $\bar{W}_{0v}$ and $\bar{W}_{1v}$ are also $O_p(1)$. It is now easily seen $\hat{J}_\alpha = \hat{J}_{\alpha,1} + O_p\left(\frac{\sqrt{N}}{v^3}\right) + O_p\left(\frac{1}{v}\right)$, where $\hat{J}_{\alpha,1}$ is defined by (29). However, under Assumptions 1, 2^a and 3, $\hat{J}_\alpha \rightarrow_d N(0,1)$, as $N \rightarrow \infty$ with $T > m + 5$. Hence, by Lemma 4.7 in White (2001, p.67), $\hat{J}_{\alpha,1} \rightarrow_d N(0,1)$ if $\sqrt{N}/T^3 \rightarrow 0$, as N and $T \rightarrow \infty$, as required. ∎

Therefore, under non-Gaussian errors, the use of the $\hat{J}_{\alpha,1}$ test is asymptotically justified even when N is much larger than T, when T is not too small. Monte Carlo results, to be reported below, suggest that the test works well even if $T = 60$ and $N = 500$.

One could be tempted to adjust the $\hat{J}_{\alpha}$ test statistic by including the higher order terms given by (30) and (31). However, sample estimates of $\gamma_{1,i}^2$ and $\gamma_{2,i}$ needed to implement such second order corrections are themselves subject to estimation uncertainty and their use need not result in improvements over the $\hat{J}_{\alpha,1}$ test.

4.3 Gaussian and cross-sectionally correlated errors

To consider the case under cross-sectionally correlated errors, we assume that the correlation matrix $\mathbf{R}$ is sparse.[12] In terms of Assumption 3, it is sufficient to require that $\mathbf{Q}$ is sparse in addition to its inverse having bounded absolute maximum column and row sum matrix norms.

Suppose now that the errors are Gaussian but the errors are weakly cross-sectionally correlated such that $\delta_b < 1/2$. In this case the numerator of $\hat{J}_{\alpha,1}$ is still valid, but an exact expression for $Var\left(\hat{W}_d\right)$ now depends on the average of pair-wise correlation coefficients of t_i^2 and t_j^2, specifically

$$Var\left(\hat{W}_d\right) = \frac{2N(v-1)}{(v-4)}\left(\frac{v}{v-2}\right)^2\left[1 + \frac{2}{N}\sum_{i=2}^{N}\sum_{j=1}^{i-1}Corr(t_i^2, t_j^2)\right], \qquad (32)$$

where $Corr(t_i^2, t_j^2) = Cov(t_i^2, t_j^2)/\left[Var(t_i^2)Var(t_j^2)\right]^{1/2}$. The relevant J_{α} statistic for this case is given by

$$\hat{J}_{\alpha,1} = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{v}{v-2}\right)}{\left\{\frac{2N(v-1)}{(v-4)}\left(\frac{v}{v-2}\right)^2\left[1 + \frac{2}{N}\sum_{i=2}^{N}\sum_{j=1}^{i-1}Corr(t_i^2, t_j^2)\right]\right\}^{1/2}}. \qquad (33)$$

However, as shown in Appendix E, under Gaussianity and $\mathbf{R}$ being sparse, for all $i \neq j$ we have

$$Corr\left(t_i^2, t_j^2\right) = \begin{cases} \rho_{ij}^2 + O(v^{-1/2}), & \text{for } \rho_{ij} \neq 0, \\ 0, & \text{for } \rho_{ij} = 0, \end{cases}$$

where $\rho_{ij} = E(u_{it}u_{jt})/\sqrt{E(u_{it}^2)E(u_{jt}^2)}$ is the correlation coefficient of the errors, u_{it} and u_{jt}. Therefore, considering that by assumption $\mathbf{R}$ is sparse, then the maximum number of elements in each row of $\mathbf{R}$ must be bounded and we have

$$\frac{2}{N}\sum_{i=2}^{N}\sum_{j=1}^{i-1}Corr(t_i^2, t_j^2) = (N-1)\overline{\rho^2} + O\left(v^{-1/2}\right), \qquad (34)$$

where $\overline{\rho^2}$ is defined by (21). Hence, the J_{α} statistic in the present case can be written as

$$J_{\alpha} = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{v}{v-2}\right)}{\sqrt{Var\left(\hat{W}_d\right)}} = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{v}{v-2}\right)}{\sqrt{\left(\frac{v}{v-2}\right)^2\frac{2(v-1)}{(v-4)}\left[1 + (N-1)\overline{\rho^2} + O\left(v^{-1/2}\right)\right]}}. \qquad (35)$$

---

[12]Matrix $\mathbf{R} = (\rho_{ij})$ is said to be sparse if the maximum number of non-zero elements in its rows (columns) is bounded in N, namely if $\sum_{j=1}^{N} I\left(\rho_{ij} \neq 0\right) \leq K < \infty$, for all i, where $I(A)$ is an indicator function that takes the value of unity if A is true and zero otherwise.

The derivation of the asymptotic distribution of J_α is complicated by the fact that under error cross-sectional dependence the t_i^2 statistics are not independently distributed. But from Theorem 3 under Gaussian errors and when $\delta_b < 1/2$ (defined by (23) and (24)), we know that

$$J_\alpha(\mathbf{D}) = \frac{N^{-1/2}\left[\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)\hat{\boldsymbol{\alpha}}'\mathbf{D}^{-1}\hat{\boldsymbol{\alpha}} - N\right]}{\sqrt{2}} \to_d N(0,1). \qquad (36)$$

Also, when $\delta_b < 1/2$, then $(N-1)\overline{\rho^2} \to 0$, as $N \to \infty$, and

$$Var\left(\hat{W}_d\right) = \left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{(v-4)}\left[1 + (N-1)\overline{\rho^2} + O\left(v^{-1/2}\right)\right] \to 2, \text{ as } N \text{ and } T \to \infty, \text{ in any order.}$$

Hence, to establish the asymptotic normality of J_α it is sufficient to prove that Z_{NT} defined by

$$Z_{NT} = N^{-1/2}\sum_{i=1}^N \left(t_i^2 - \frac{v}{v-2}\right) - N^{-1/2}\left[\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)\hat{\boldsymbol{\alpha}}'\mathbf{D}^{-1}\hat{\boldsymbol{\alpha}} - N\right], \qquad (37)$$

also tends to 0 as N and $T \to \infty$. This result is established in Appendix F. The following corollary summarizes this finding:

Corollary 5 *Consider the regression model (1), and suppose that Assumptions 1, 2, and 3 hold, and $\mathbf{Q}$ defined in Assumption 3 is sparse. In addition, suppose that the errors, u_{it}, are sufficiently weakly correlated, such that $\delta_b < 1/2$, so that $(N-1)\overline{\rho^2} \to 0$, as $N \to \infty$. Then, under $H_0 : \alpha_i = 0$ for all i, $\hat{J}_{\alpha,1} \to_d N(0,1)$, as N and $T \to \infty$, jointly, where $\hat{J}_{\alpha,1}$ is defined by (29).*

Now consider the slightly stronger case of error cross-sectional dependence where $\delta_b \leq 1/2$. In this case, $(N-1)\overline{\rho^2}$ does not necessarily tend to zero but is merely bounded, as $N \to \infty$, thus, the $\hat{J}_{\alpha 1}$ test should be modified such that $(N-1)\overline{\rho^2}$ is replaced by a consistent estimator. Therefore, in order to accommodate the case $\delta_b = 1/2$ and also to improve the small sample performance of the test, particularly in cases where δ_b is very close to $1/2$, we propose to estimate $\frac{2}{N}\sum_{i=2}^N\sum_{j=1}^{i-1}Corr(t_i^2,t_j^2)$ by $(N-1)\widehat{\overline{\rho^2}}$, where

$$\widehat{\overline{\rho^2}} = \frac{2}{N(N-1)}\sum_{i=2}^N\sum_{j=1}^{i-1}\hat{\rho}_{ij}^2 I\left(v\hat{\rho}_{ij}^2 \geq \theta_N\right), \qquad (38)$$

$\hat{\rho}_{ij} = \hat{\mathbf{u}}_{i\cdot}'\hat{\mathbf{u}}_{j\cdot}/\sqrt{\left(\hat{\mathbf{u}}_{i\cdot}'\hat{\mathbf{u}}_{i\cdot}\right)\left(\hat{\mathbf{u}}_{j\cdot}'\hat{\mathbf{u}}_{j\cdot}\right)}$, and $I(A)$ is an indicator function which returns unity if A is true and zero otherwise. The threshold value θ_N is chosen such that $\Pr(\rho_{ij} \neq 0)$ decline steadily with N. This is because in the case of weakly correlated errors with $\mathbf{R}$ being sparse, there are only a finite number of non-zero elements per each row of matrix $\mathbf{R}$ as $N \to \infty$. Under the null hypothesis that errors are cross-sectionally independent, for each i, the $N-1$ separate tests of $\rho_{ij} = 0$, for $j = 1, 2, ..., N$, $j \neq i$, are also independent, and the size of the individual $N-1$ tests is given by $p_N = 1 - (1-p)^{1/(N-1)}$, or approximately $p_N = p/(N-1)$, where p is the overall size of the test.[13] Now noting that under $\rho_{ij} = 0$, and for T sufficiently large $v\hat{\rho}_{ij}^2 \sim \chi_1^2$, we obtain $\sqrt{\theta_N} = \Phi^{-1}\left(1 - \frac{p_N}{2}\right)$, where $\Phi^{-1}(.)$ represents the inverse of the cumulative distribution function of the standard normal variate. The

---

[13]The approximate formula also admits dependence across tests using Bonferroni's formula. We set $p = 10\%$ in the Monte Carlo experiments and the empirical analysis that follow.

test statistic for this case can then be written as

$$\hat{J}_{\alpha,2} = \frac{N^{-1/2} \sum_{i=1}^{N} \left(t_i^2 - \frac{v}{v-2} \right)}{\left(\frac{v}{v-2} \right) \sqrt{\frac{2(v-1)}{(v-4)} \left[1 + (N-1)\widehat{\overline{\rho^2}} \right]}}, \tag{39}$$

with $\widehat{\overline{\rho^2}}$ defined by (38).

By using a similar discussion for Corollary 5, from Theorem 2 under Gaussian errors and $\delta_b \leq 1/2$, we know that

$$J_{\alpha}(\mathbf{D}) = \frac{N^{-1/2} \left[\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right) \hat{\boldsymbol{\alpha}}' \mathbf{D}^{-1} \hat{\boldsymbol{\alpha}} - N \right]}{\sqrt{2 \left[1 + (N-1)\overline{\rho^2} \right]}} \rightarrow_d N(0,1). \tag{40}$$

Since it is shown that $Z_{NT} \rightarrow_p 0$ as N and $T \rightarrow \infty$ in any order when $\delta_b \leq 1/2$ with $\mathbf{R}$ being sparse in Appendix F, to establish conditions that $J_{\alpha}(\mathbf{D}) - \hat{J}_{\alpha,2}$ is asymptotically negligible, it is enough to verify conditions under which $plim_{N,T\rightarrow\infty}(N-1)\widehat{\overline{\rho^2}} = \lim_{N\rightarrow\infty}(N-1)\overline{\rho^2}$. In Appendix G we show that $(N-1)\overline{\rho^2}$ is consistently estimated by $(N-1)\widehat{\overline{\rho^2}}$, if $\mathbf{R}$ is sparse and $ln(N)/\sqrt{T} \rightarrow 0$, as N and $T \rightarrow \infty$. In terms of the factor specification in Theorem 3, $\hat{J}_{\alpha,2}$ test continues to be valid even if $\delta_b \leq 1/2$, which allows for a higher degree of error cross-sectional dependence as compared to $\delta_b < 1/2$ assumed in Theorem 3.

The following corollary covers the case where $\lim_{N\rightarrow\infty}(N-1)\overline{\rho^2}$ is finite but possibly non-zero:

Corollary 6 *Consider the regression model (1), and suppose that Assumptions 1, 2, and 3 hold, and $\mathbf{Q}$ defined in Assumption 3 is sparse. Then under $H_0 : \alpha_i = 0$ for all i, $\hat{J}_{\alpha,2} \rightarrow_d N(0,1)$, as N and $T \rightarrow \infty$, such that $ln(N)/\sqrt{T} \rightarrow 0$, where $\hat{J}_{\alpha,2}$ is defined by (39).*

In practice where the extent of error cross-correlations is unknown, $\hat{J}_{\alpha,2}$ is preferred to $\hat{J}_{\alpha,1}$. Small sample evidence in favour of this observation is provided in the Monte Carlo section.

4.4 Non-Gaussian and cross-sectionally correlated errors

When $\delta_b < 1/2$, combining the results of Theorem 4 and Corollary 5, then $\hat{J}_{\alpha,1}$ continues to be asymptotically $N(0,1)$ so long as $N/T^3 \rightarrow 0$, as $N \rightarrow \infty$ and $T \rightarrow \infty$, jointly, if it is further assumed that (i) $\mathbf{f}_t' \mathbf{f}_t \leq K < \infty$ for all t, and (ii) $E(|\varepsilon_{it}|^{8+\epsilon}) < K < \infty$ for some $\epsilon > 0$. When $\delta_b \leq 1/2$, combining the results obtained in sub-sections 4.2 and 4.3, $\hat{J}_{\alpha,2}$ is asymptotically a valid test so long as $N/T^3 \rightarrow 0$ and $ln(N)/\sqrt{T} \rightarrow 0$, as $N \rightarrow \infty$ and $T \rightarrow \infty$, jointly.

5 Monte Carlo Experiments

We examine the finite sample properties of the $\hat{J}_{\alpha}$ tests (namely $\hat{J}_{\alpha,1}$, and $\hat{J}_{\alpha,2}$) by Monte Carlo experiments, and compare their performance to existing tests whenever possible. For comparison, we consider the GRS test as well as two distribution-free sign tests of $\alpha_i = 0$, proposed by Gungor and Luger (2009). These tests, referred to as SS and WS tests, allow the error distribution to be non-normal but require it to be conditionally symmetric

around zero.[14] These tests are relatively easy to compute and are applicable even when $N > T$. However, they are constructed for models with a single factor and their validity is established only under $N < T$.

The SS test is based on the sign statistic

$$SS_N = \sum_{i=1}^{N} S_i^2, \tag{41}$$

where

$$S_i = \frac{\left[\sum_{t=1}^{\mathcal{T}} I\left(z_{it} > 0\right)\right] - \mathcal{T}/2}{\sqrt{\mathcal{T}/4}},$$

$I\left(A\right)$ is the indicator function as defined by (38),

$$z_{it} = \left(\frac{y_{i,t+\mathcal{T}}}{f_{t+\mathcal{T}}} - \frac{y_{it}}{f_t}\right)\left(\frac{f_t - f_{t+\mathcal{T}}}{f_t f_{t+\mathcal{T}}}\right), \, t = 1, 2, ..., \mathcal{T},$$

$\mathcal{T}$ is the nearest integer part of $T/2$. The WS test is based on the Wilcoxon signed rank statistic

$$WS_N = \sum_{i=1}^{N} W_i^2, \tag{42}$$

where

$$W_i = \frac{\left[\sum_{t=1}^{\mathcal{T}} I\left(z_{it} > 0\right) Rank(|z_{it}|)\right] - \mathcal{T}\left(\mathcal{T} + 1\right)/4}{\sqrt{\mathcal{T}\left(\mathcal{T} + 1\right)\left(2\mathcal{T} + 1\right)/24}},$$

$Rank(|z_{it}|)$ is the rank (natural number) of $|z_{it}|$ when $|z_{i1}|, |z_{i2}|, ..., |z_{i\mathcal{T}}|$ are placed in an ascending order of magnitude. Gungor and Luger (2009) show that under the null hypothesis, $\alpha_i = 0$ for all i, both S_i and W_i statistics have limiting (as $\mathcal{T} \to \infty$) standard normal distributions. Under the additional assumption that the errors in the CAPM regressions are cross-sectionally independent, conditional on the values of the single factor $(f_1, f_2, ..., f_T)$, SS_N and WS_N follow χ_N^2 distributions.[15]

5.1 Monte Carlo designs and experiments

We consider the following data generating process (DGP)

$$r_{it} = \alpha_i + \sum_{\ell=1}^{m} \beta_{\ell i} f_{\ell t} + u_{it}, i = 1, 2, .., N; t = 1, 2, ..., T, \tag{43}$$

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.[16] The Monte Carlo (MC) design is also intended to match the models used for the empirical applications that follow. Accordingly, we set $m = 3$ and generate the factors as

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

---

[14]See equation (13) in Gungor and Luger (2009) for the definition of SS and WS test statistics.

[15]Gungor and Luger (2011) extend their approach to accommodate multiple factors, but will not be considered here since even in the case of single factor models Gungor and Luger's proposed tests suffer from serious size distortions in the presence of non-normal errors and/or cross-correlated errors, as documented below.

[16]See Fama and French (1993). SMB stands for "small market capitalization minus big" and HML for "high book-to-market ratio minus low". See Subsection 6.1 and Appendix C for further details and data sources.

where $\zeta_{\ell t} \sim IIDN(0,1)$ and[17]

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

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 $\mathbf{D} = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_N^2)'$ 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, motivated by the error factor structure discussed in Subsection 3.2, we generate $\mathbf{R}$ as

$$
\mathbf{R} = \mathbf{I}_N + \mathbf{bb}' - \check{\mathbf{B}}^2, \tag{44}
$$

where $\mathbf{b} = (b_1, b_2, ..., b_N)'$ and $\check{\mathbf{B}} = diag(\mathbf{b})$. The choice of $\mathbf{R}$ is also related to the single error factor model, $u_{it} = \varphi_i h_t + \omega_i \varepsilon_{it}$, with $h_t \sim iidN(0,1)$, and $\varepsilon_{it} \sim iidN(0,1)$, where $Var(u_{it}^2) = \sigma_i^2 = \varphi_i^2 + \omega_i^2$, and letting $b_i = \varphi_i/\sigma_i$, then the correlation matrix of $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$ can be written as (44). To generate different degrees of cross-sectional dependence, we draw the first and the last N_b ($< N$) elements of $\mathbf{b}$ as $Uniform(0.7, 0.9)$, and set the remaining middle elements to 0. We set $N_b = \lfloor N^{\delta_b} \rfloor$, where $\lfloor A \rfloor$ is the largest integer part of A, and consider the exponents $\delta_b = 0.25, 0.50$, and 0.60. The case of no error cross-sectional dependence is obtained when $N_b = 0$, and the error cross-sectional dependence is weak when $\delta_b \leq 0.5$. The case of $\delta_b = 0.60$ is included to see how the two operational versions of the J_α test (namely $\hat{J}_{\alpha,1}$ and $\hat{J}_{\alpha,2}$) perform when cross-sectional error correlations are higher than the threshold value of 0.50 allowed by the theory. As discussed earlier, the J_α test is asymptotically valid when $N^{-1}Tr(\mathbf{R}^2) = O(1)$, and this condition is satisfied for $\delta_b \leq 1/2$, but not for $\delta_b > 1/2$.

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 $\tau = 1989M9, 1989M10, ..., 2011M9$, using sample of sizes of $T = 60$ months. Specifically, for each $i = 1, 2, ..., N_\tau$ we run the Fama-French regressions $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}(r_{m,\tau t} - r_{f,\tau t}) + \hat{\beta}_{2,i\tau}SMB_{t\tau} + \hat{\beta}_{3i}HML_{t\tau} + \hat{u}_{i,\tau t}$, $t = 1, 2, ..., 60$, at the end of each month $\tau = 1989M9$ to $2011M9$, and computed $\hat{\sigma}_{i,\tau}^2 = \hat{m}_{2,i\tau}$, $\hat{\gamma}_{1,i,\tau} = \hat{m}_{3,i\tau}/\hat{m}_{2,i\tau}^{3/2}$ and $\hat{\gamma}_{2,i\tau} = \hat{m}_{4,i\tau}/\hat{m}_{2,i\tau}^2 - 3$ with $\hat{m}_{s,i\tau} = (60)^{-1}\sum_{t=1}^{60}(\hat{u}_{i,\tau t} - \overline{\hat{u}}_{i,\tau})^s$, 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 $\hat{\sigma}_{i,\tau}^2$, $\hat{\gamma}_{1,i,\tau}$ 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 $\hat{\sigma}_{i,\tau}^2$, $\hat{\gamma}_{1,i,\tau}$, $\hat{\gamma}_{2,i,\tau}$ 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

---

[17]The estimates used in the generation of the factors and their volatilities are computed using monthly observations over the period 1973M4-2011M9.

of deviations, skewness and kurtosis, across securities and sample periods. Accordingly, in order to mimic as far as possible the main characteristics of observed security returns, for each replication, r, we generate $\sigma_i^{2(r)}$, $\gamma_{1,i}^{(r)}$, $\gamma_{2,i}^{(r)}$, $\left\{\beta_{\ell,i}^{(r)}, \text{ for } \ell = 1, 2, 3\right\}$, as random draws from their respective empirical distributions. For example, to generate $\sigma_i^{2(r)}$ over r and i, we first place the estimates $\hat{\sigma}_{i,\tau}^2$, for $i = 1, 2, ..., N_\tau$, and $\tau = 1, 2, ..., 265$, that lie in the 2.5% to 97.5% quantile range, into 10 bins and then randomly select a bin with probability equal to the proportion of the estimates in each bin, and then draw randomly a value for $\sigma_i^{(2r)}$ from the selected bin. This procedure is repeated over $i = 1, 2, ..., N$ and replications $r = 1, 2, ..., R$.

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

1. We generate N random draws $\sigma_i^{2(r)}$, $\gamma_{1,i}^{(r)}$ and $\gamma_{2,i}^{(r)}$, $i = 1, 2, ..., N$, as described above, and set
$$m_{3,i}^{(r)} = \sigma_i^{3(r)}\gamma_{1,i}^{(r)}, \text{ and } m_{4,i}^{(r)} = \sigma_i^{4(r)}\left(\gamma_{2,i}^{(r)} + 3\right).$$

2. We 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 $\boldsymbol{\kappa}^{(r)} = (\kappa_1^{(r)}, \kappa_2^{(r)}, ..., \kappa_N^{(r)})'$ with $\kappa_{\varepsilon i}^{(r)} = m_{\varepsilon,4,i}^{(r)} - 3$ and $\kappa_i^{(r)} = m_{4,i}^{(r)} - 3\sigma_i^{4(r)}$, $\mathbf{Q}^{(r)} = \mathbf{D}^{(r)1/2}\mathbf{P}^{(r)}$, with $\mathbf{D}^{(r)} = diag(\sigma_1^{2(r)}, \sigma_2^{2(r)}, ..., \sigma_i^{2(r)})'$ and $\mathbf{P}^{(r)}$ being a Cholesky factor of correlation matrix $\mathbf{R}^{(r)}$ which is generated as described above (see (44). 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)
$$\varepsilon_{it} = a_i + b_i v_{it} + c_i v_{it}^2 + d_i v_{it}^3, \ i = 1, 2, ..., N,$$
where $v_{it} \sim IIDN(0, 1)$ 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:
$$a_i + c_i = 0,$$
$$b_i^2 + 6b_i d_i + 2c_i^2 + 15d_i^2 = 1,$$
$$2c_i(b_i^2 + 24b_i d_i + 105d_i^2 + 2) = m_{\varepsilon,3,i},$$
$$24[b_i d_i + c_i^2(1 + b_i^2 + 28b_i d_i) + d_i^2(12 + 48b_i d_i + 141c_i^2 + 225d_i^2)] = \kappa_{\varepsilon i}.$$

4. Finally, we compute $u_{it}^{(r)} = \sum_{j=1}^N q_{ij}^{(r)}\varepsilon_{jt}^{(r)}$, where $q_{ij}^{(r)}$ is the (i, j) element of $\mathbf{Q}^{(r)}$, and $\varepsilon_{jt}^{(r)}$ is the rth draw from the above DGP.

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 $N_\alpha = \lfloor N^{\delta_\alpha}\rfloor$; $\alpha_i = 0$ for $i = N_\alpha + 1, N_\alpha +$

$2, ..., N$. We considered the values $\delta_\alpha = 0.8, 0.9, 1.0$, but the power ended up to be very high even for $\delta_\alpha = 0.8$. Therefore, we only report power estimates for $\delta_\alpha = 0.80$.

Having obtained a general calibrated design based on realized security return data, we consider four sets of experiments in order to investigate the effect of weak cross-sectional correlation and non-normality of the errors upon the performance of the tests:

(i) Cross sectionally independent normal case: $u_{it}^{(r)} = \sigma_i^{(r)} \varepsilon_{it}^{(r)}$ with $\varepsilon_{it}^{(r)} \sim IIDN(0,1)$ for all i and r.

(ii) Cross sectionally independent non-normal case: Use the procedure generating non-normal errors $u_{it}^{(r)}$ as specified in steps 1-4 above, with $\mathbf{Q}^{(r)}$ matrix replaced by $\mathbf{D}^{(r)1/2}$.

(iii) Cross sectionally correlated normal case: $u_{it}^{(r)} = \sum_{j=1}^{N} q_{ij}^{(r)} \varepsilon_{jt}^{(r)}$ with $\varepsilon_{it}^{(r)} \sim IIDN(0,1)$ for all i and r.

(iv) Cross sectionally correlated non-normal case: $u_{it}^{(r)}$ are generated following the steps 1-4 set out above.

All combinations of $T = 60, 100$ and $N = 50, 100, 200, 500$ are considered. All tests are conducted at a 5% significance level. Experiments are based on $R = 2000$ replications.

5.2 Test results

Table 2 reports the size and power of the SS and WS nonparametric tests of Gungor and Luger (2009), GRS and $\hat{J}_\alpha$ tests in the case of models with one factor, under various degrees of cross-sectional error correlations which are measured by the exponent, δ_b. Panel A shows the results under normality, and Panel B gives the results under non-normal errors.

With cross-sectionally independent and Gaussian errors, the size of the SS and WS tests is very close to the 5% nominal level for all combinations of N and T. The GRS test, being an exact test, has the correct size (in cases where $T > N$). The empirical size of $\hat{J}_{\alpha,1}$ and $\hat{J}_{\alpha,2}$ tests is also very close to the 5% nominal level for all combinations of N and T. Even when $N = 500$, the size of $\hat{J}_\alpha$ tests lie in the range 5.0% to 5.3% for different values of T. However, the power of the $\hat{J}_\alpha$ tests is substantially higher than that of the GRS, SS and WS tests. For example, for $T = 60$ and $N = 50$ the power of the GRS test is 22.1% as compared to 77.2% for the $\hat{J}_{\alpha,2}$ test, although both tests have similar sizes (5.3% for the GRS test and 6.1% for the $\hat{J}_{\alpha,2}$ 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 results in the power of the GRS test to rise from 22.1% to 77.2%, whilst the power of the $\hat{J}_{\alpha,2}$ test (for example) rises from 77.2% to 93.8%. 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 of Gungor-Luger tests is even smaller than that of the GRS test when $N = 50$. For example, for $N = 50$ and $T = 100$ the power of the SS and WS tests are 34.7% and 43.0%, respectively, as compared to 77.2% for the GRS test. Furthermore, whilst increasing N improves the power of nonparametric tests, the increase is much less than what is achieved by the $\hat{J}_\alpha$ tests. For example, keeping T fixed at 60 and increasing N from 50 to 200 raises the power of the WS test from 23.0% to 43.7%, while the power of the $\hat{J}_{\alpha,2}$ test rises from 77.2% to 97.8%. Perhaps this is not surprising considering that the nonparametric tests do not make full use of the available observations.

Consider now the case where the errors are normally distributed but cross-sectionally dependent. Recall that the strength of error cross-sectional correlation is measured by the exponent, δ_b. In line with our theoretical findings (see Section 2) for $\delta_b < 1/2$, in particular for $\delta_b = 1/4$, the sizes of $\hat{J}_{\alpha,1}$ and $\hat{J}_{\alpha,2}$ tests are very close to the nominal value of 5%, for all combinations of N and T. In particular, the size of the $\hat{J}_{\alpha,2}$ test for $N = 50$ and $T = 60$ is 6.0%, and its power is 77.9% which substantially exceeds the power of the GRS test at 21.0% for the same combination of the sample sizes. Similar results are also obtained when one considers the case where $N = 50$ and $T = 100$. Interestingly, the SS and WS tests have also correct size. When $\delta_b = 1/2$, the $\hat{J}_{\alpha,1}$ test is no longer valid asymptotically, and in line with this theoretical result, $\hat{J}_{\alpha,1}$ test is now over-sized with its empirical size ranging between 9.6% and 11.3%. In contrast, the $\hat{J}_{\alpha,2}$ test seems quite robust to cross-sectional error correlations, with its size falling in the range 5.4% to 6.9%. At this level of cross-sectional error correlations, the nonparametric tests also exhibit some size distortions. The sizes of SS and WS tests now lie in the range 7.0% to 8.5%, and 7.6% to 9.0%, respectively. The size of the $\hat{J}_{\alpha,2}$ test for $N = 50$ and $T = 60$ is 6.4%, and its power is 64.6%, which still exceed the power of the GRS test which for this sample size combination is 30.0%. But, as expected, increasing T from 60 to 100 results in the power of the GRS test to rise to 90.1%, which marginally beat the power of the $\hat{J}_{\alpha,2}$ test at 88.9%. When $\delta_b = 3/5 > 1/2$, 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 $\hat{J}_{\alpha,2}$ test is also close to its nominal level (at 6.2%-7.2%) even for such a high value of δ_b. This seems to be due to the inclusion of $(N-1)\widehat{\rho^2}$ in the denominator of the $\hat{J}_{\alpha,2}$ statistic, a term which is absent from the denominator of the $\hat{J}_{\alpha,1}$ statistic.

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 but cross-sectionally uncorrelated. The nonparametric tests exhibit a considerable size distortion, which is accentuated as T or N rises. For example, when $T = 60$, increasing N from 60 to 500 results in the size of the SS test to rise from 11.0% to 34.0%, and fixing $N = 500$ but increasing T from 60 to 100 causes the size of the SS test to rise from 11.0% to 64.6%. On the other hand, we see that the size of the GRS test is hardly affected by the types of departures from Gaussianity observed in the CAPM residuals. The robustness of the GRS test to non-normal CAPM 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 $\hat{J}_{\alpha,1}$ test and the $\hat{J}_{\alpha,2}$ test are 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 $\hat{J}_{\alpha,2}$ test are 6.5%, 6.6%, 6.1%, and 6.6%, respectively. Furthermore, the $\hat{J}_{\alpha,1}$ and $\hat{J}_{\alpha,2}$ tests continue to maintain their power superiority over the GRS test.

We now consider the empirically most relevant case where the errors are non-normal as well as being cross-sectionally correlated. When $\delta_b = 1/4$, there is no significant difference in the results from those reported above for the cross-sectionally uncorrelated case. When $\delta_b = 1/2$, the size of the $\hat{J}_{\alpha,1}$ test lies in the range 10.1%-11.8%, but the size of the $\hat{J}_{\alpha,2}$ test is reasonably controlled and lies in the range 5.5%-7.9%. The power comparisons discussed for the cross-sectionally uncorrelated case also carry over to the present more general set of experiments.

The Monte Carlo results for the Fama-French three-factor regressions are summarized in Table 3, and give empirical sizes and powers of the GRS test and the two $\hat{J}_\alpha$ tests. The SS and WS tests are only applicable to models with one factor and are therefore excluded from Table 3. As can be seen, the results are qualitatively very similar to those obtained

for the one factor model, and will not be discussed any further.

6   Empirical Application

6.1   Data description

We consider the application of our proposed $\hat{J}_\alpha$ tests 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 exceeds US\$ 4 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 September 2011. 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&P500, 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 (1989M9-2011M9)

Encouraged by the satisfactory performance of the J_α tests, even in cases where N is much larger than T, we apply the $\hat{J}_{\alpha,2}$ 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 September 2011.[18] 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 we selected a relatively small sample period of 60 months, primarily to reduce the impact of possible structural changes in β_i's, return volatilities and correlations on the test outcomes. Accordingly, we estimated the CAPM regressions

$$r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{u}_{i,\tau t}, \tag{45}$$

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

$$r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{\beta}_{2,i\tau}SMB_{t\tau} + \hat{\beta}_{3i}HML_{t\tau} + \hat{u}_{i,\tau t}, \tag{46}$$

---

[18]In all the empirical applications $T < N$, and the GRS test can not be computed. We have also decided to exclude the SS and WS tests discussed in the Monte Carlo Section on the grounds of their substantial over-rejection of the null particularly in the presence of non-normal errors.

for $t = 1, 2, ..., 60$, $i = 1, 2, ..., N_\tau$, and the month ends, $\tau = 1989M9, 1989M10, ..., 2011M9$. 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 4 reports summary statistics for p-values of $\hat{J}_{\alpha,2}$ 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 1989M09 to 2011M09. 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 25% 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 0.59 and 0.74 for the CAPM model, and 0.52 and 0.58 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 $\hat{J}_{\alpha,2}$ test based on CAPM and FF regressions at the end of the months 1989M09 to 2011M09. The months at which the null of market efficiency is rejected at the 5% level based on both CAPM and FF regressions are 1998M8, 1998M12-1999M2 and 2007M8-2009M3. The period around 1998M8 and 1998M12-1999M2 coincide with the Russian financial crisis (during August -September 1998) and the subsequent collapse of Long-Term Capital Management. The period 2007M8-2009M3 matches the recent global financial crisis. In general, the $\hat{J}_{\alpha,2}$ 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 $\hat{J}_\alpha$ test

The test results in Figure 1 clearly show important variations in the estimated p-values over time, and it would be interesting to see if such variations in degrees of market efficiency (as measured by p-values of the $\hat{J}_{a,2}$ test) 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. 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, (or more accurately when $\sum_{i=1}^{N} \alpha_i^2 = O(N^\delta)$ with $\delta < 1/2$), 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 Dow Jones Credit Suisse Core Long/Short Equity Hedge Fund Index that are available monthly from January 2005. This is a subset of the Dow Jones Core 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 $\hat{J}_{\alpha,2}$ tests, which we denote by $\hat{\pi}_t$.[19] The p-values needed for this purpose are already reported in Figure 1. Given the considerable volatility of return data, in Figures 2 & 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 strikingly negative relationship between the two variables as expected. The fit is much better when the FF p-values are used, yielding a correlation of around -0.84. The correlation between $\tilde{r}_{ht}(12)$ and the p-values computed using CAPM regressions is -0.74 which is still quite substantial.

7   Concluding Remarks

In this paper we propose a simple test of Sharpe-Lintner CAPM, the J_{α} test, when the number of securities, N, is large relative to the time dimension, T, of the return series. We considered two operational versions, $\hat{J}_{\alpha,1}$ and $\hat{J}_{\alpha,2}$ tests, both of which are robust to non-Gaussianity. The $\hat{J}_{\alpha,2}$ test is expected to be more robust against error cross-sectional correlation. 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 $\hat{J}_{\alpha,1}$ and $\hat{J}_{\alpha,2}$ tests perform well even when N is much larger than T. Also in cases where $N < T$ and the standard F test due to GRS can be computed, we still find that the $\hat{J}_{\alpha,1}$ and $\hat{J}_{\alpha,2}$ tests have a much higher power, especially when T is relatively small. We recommend the use of $\hat{J}_{\alpha,2}$ test (over $\hat{J}_{\alpha,1}$), as it is shown to be relatively robust to moderately strong degrees of error cross-sectional correlations. It is worth bearing in mind that under CAPM, we do not expect the errors to be strongly cross-correlated.

Application of the $\hat{J}_{\alpha,2}$ 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 - September 2011 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 surprisingly strong negative correlation is found between a twelve-month moving average p-values of the $\hat{J}_{\alpha,2}$ test and excess returns of long/short equity strategies over the period December 2006 to May 2011.

---

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

Appendix A: Mathematical Derivations and Proofs

Consider the quadratic form

$$W_{0,a} = \mathbf{x}_T' \mathbf{A} \mathbf{x}_T,$$

and the linear quadratic form

$$W_{1,\alpha} = \mathbf{x}_T' \mathbf{A} \mathbf{x}_T + 2 \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right)^{1/2} \boldsymbol{\phi}' \mathbf{A} \mathbf{x}_T,$$

where $\mathbf{x}_T = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right)^{1/2} \sum_{t=1}^T c_t \boldsymbol{\varepsilon}_t$, $\mathbf{c} = (c_1, c_2, ..., c_T)' = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right)^{-1} \mathbf{M}_F \boldsymbol{\tau}_T$, and $\mathbf{A} = (a_{ij})$ is a given symmetric with non-stochastic elements, a_{ij}, $\boldsymbol{\phi} = (\phi_1, \phi_2, ..., \phi_N)'$ is an $N \times 1$ vector of fixed constants. Elements of $\mathbf{A}$ and $\boldsymbol{\phi}$ could also vary with N, and strictly speaking they should be written as $a_{ij,N}$ and $\phi_{i,N}$. But to simplify the notations and without loss of generality we abstract from the subscript N.

Recall that

$$\sum_{t=1}^T c_t^2 = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right)^{-1} = O(v^{-1}) \tag{47}$$

$$
\begin{aligned}
E(\mathbf{x}_T) &= \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right)^{1/2} \sum_{t=1}^T c_t E(\boldsymbol{\varepsilon}_t) = \mathbf{0}, \\
Var(\mathbf{x}_T) &= E\left(\mathbf{x}_T \mathbf{x}_T' \right) = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right) E\left(\sum_{t=1}^T \sum_{t'=1}^T c_t c_{t'} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t'}' \right) = \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right) \left(\sum_{t=1}^T c_t^2 \right) \mathbf{I}_N = \mathbf{I}_N.
\end{aligned}
$$

$$
\begin{aligned}
E(W_{0,a}) &= E\left(\mathbf{x}_T' \mathbf{A} \mathbf{x}_T \right) = Tr\left[\mathbf{A} E\left(\mathbf{x}_T \mathbf{x}_T' \right) \right] = Tr(\mathbf{A}), \tag{48} \\
E(W_{1,\alpha}) &= Tr(\mathbf{A}). \tag{49}
\end{aligned}
$$

Also,

$$Var(W_{0,a}) = E\left[\left(\mathbf{x}_T' \mathbf{A} \mathbf{x}_T \right)^2 \right] - \left[Tr(\mathbf{A}) \right]^2 .$$

But

$$
\begin{aligned}
\left(\mathbf{x}_T' \mathbf{A} \mathbf{x}_T \right)^2 &= \left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right)^2 \left(\sum_{t=1}^T \sum_{t'=1}^T c_t c_{t'} \boldsymbol{\varepsilon}_t' \mathbf{A} \boldsymbol{\varepsilon}_{t'} \right)^2 \\
&= \sum_{t=1}^T \sum_{t'=1}^T \sum_{s=1}^T \sum_{s'=1}^T c_t c_{t'} c_s c_{s'} \left(\boldsymbol{\varepsilon}_t' \mathbf{A} \boldsymbol{\varepsilon}_{t'} \right) \left(\boldsymbol{\varepsilon}_s' \mathbf{A} \boldsymbol{\varepsilon}_{s'} \right).
\end{aligned}
$$

Since, by assumption, ε_{it} are *iid* with zero means then in the above sums only the terms associated with $t = t' = s = s'$; $t = t' \neq s = s'$; $s = t \neq t' = s'$, and $s' = t \neq t' = s$ are non-zero. Specifically, for $t = t' = s = s'$ we have

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

where $\gamma_{2,\varepsilon i} = E(\varepsilon_{it}^4) - 3$. For $t = t' \neq s = s'$

$$E\left[\left(\boldsymbol{\varepsilon}_t' \mathbf{A} \boldsymbol{\varepsilon}_t \right) \left(\boldsymbol{\varepsilon}_s' \mathbf{A} \boldsymbol{\varepsilon}_s \right) \right] = \left[Tr(\mathbf{A}) \right]^2 \text{ for } t \neq s.$$

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

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

Similarly, for $s' = t \neq t' = s$

$$E\left[\left(\boldsymbol{\varepsilon}_t' \mathbf{A} \boldsymbol{\varepsilon}_{t'} \right) \left(\boldsymbol{\varepsilon}_{t'}' \mathbf{A} \boldsymbol{\varepsilon}_t \right) \right] = Tr(\mathbf{A}^2).$$

Using the above results

$$
\begin{aligned}
\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T \right)^{-2} E\left[\left(\mathbf{x}_T' \mathbf{A} \mathbf{x}_T \right)^2 \right] &= \left(\sum_{t=1}^T c_t^4 \right) \left[\sum_{i=1}^N a_{ii}^2 \gamma_{2,\varepsilon i} + \left(\sum_{i=1}^N a_{ii} \right)^2 + 2 \sum_{i=1}^N \sum_{j=1}^N a_{ij} a_{ji} \right] \\
&+ \left[\sum_{t=1}^T \sum_{s=1}^T c_t^2 c_s^2 - \left(\sum_{t=1}^T c_t^4 \right) \right] \left[Tr(\mathbf{A}) \right]^2 + 2 \left[\sum_{t=1}^T \sum_{s=1}^T c_t^2 c_s^2 - \left(\sum_{t=1}^T c_t^4 \right) \right] Tr(\mathbf{A}^2).
\end{aligned}
$$

But $\left(\sum_{t=1}^{T}\sum_{s=1}^{T}c_t^2 c_s^2\right) = \left(\sum_{t=1}^{T}c_t^2\right)^2$, $\sum_{i=1}^{N}a_{ii} = Tr(\mathbf{A})$, $\sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij}a_{ji} = Tr(\mathbf{A}^2)$, and the above simplifies to

$$\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^{-2} E\left[\left(\mathbf{x}_T'\mathbf{A}\mathbf{x}_T\right)^2\right] = \left(\sum_{i=1}^{N}a_{ii}^2\gamma_{2,\varepsilon i}\right)\left(\sum_{t=1}^{T}c_t^4\right) + \left(\sum_{t=1}^{T}c_t^2\right)^2 \left[2Tr(\mathbf{A}^2) + [Tr(\mathbf{A})]^2\right],$$

and since $\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^{-1} = \left(\sum_{t=1}^{T}c_t^2\right)$

$$E\left[\left(\mathbf{x}_T'\mathbf{A}\mathbf{x}_T\right)^2\right] = \left(\sum_{i=1}^{N}a_{ii}^2\gamma_{2,\varepsilon i}\right)\frac{\left(\sum_{t=1}^{T}c_t^4\right)}{\left(\sum_{t=1}^{T}c_t^2\right)^2} + 2Tr(\mathbf{A}^2) + [Tr(\mathbf{A})]^2.$$

Hence

$$Var(W_{0,a}) = \left(\sum_{i=1}^{N}a_{ii}^2\gamma_{2,\varepsilon i}\right)\frac{\left(\sum_{t=1}^{T}c_t^4\right)}{\left(\sum_{t=1}^{T}c_t^2\right)^2} + 2Tr(\mathbf{A}^2). \tag{50}$$

Finally

$$Var\left(W_{1,a}\right) = Var(W_{0,a}) + 4\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)\left(\boldsymbol{\phi}'\mathbf{A}^2\boldsymbol{\phi}\right) + 4\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)E\left[\left(\mathbf{x}_T'\mathbf{A}\mathbf{x}_T\right)\left(\boldsymbol{\phi}'\mathbf{A}\mathbf{x}_T\right)\right], \tag{51}$$

but

$$\begin{aligned}
Cov\left(W_{0,a}, \boldsymbol{\phi}'\mathbf{A}\mathbf{x}_T\right) &= E\left[\left(\mathbf{x}_T'\mathbf{A}\mathbf{x}_T\right)\left(\boldsymbol{\phi}'\mathbf{A}\mathbf{x}_T\right)\right] \\
&= \left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^{3/2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}c_t c_{t'} c_s E\left(\boldsymbol{\varepsilon}_t'\mathbf{A}\boldsymbol{\varepsilon}_{t'}\boldsymbol{\phi}'\mathbf{A}\boldsymbol{\varepsilon}_s\right) \\
&= \left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^{3/2}\sum_{t=1}^{T}c_t^3 E\left(\boldsymbol{\varepsilon}_t'\mathbf{A}\boldsymbol{\varepsilon}_t\boldsymbol{\phi}'\mathbf{A}\boldsymbol{\varepsilon}_t\right) \\
&= \left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^{3/2}\left(\sum_{t=1}^{T}c_t^3\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}a_{ij}\phi_\ell a_{\ell k}E\left(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\right) \\
&= \left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^{3/2}\left(\sum_{t=1}^{T}c_t^3\right)\sum_{i=1}^{N}\sum_{\ell=1}^{N}a_{ii}a_{\ell i}\phi_\ell E\left(\varepsilon_{it}^3\right) \\
&= \left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^{3/2}\left(\sum_{t=1}^{T}c_t^3\right)\sum_{i=1}^{N}a_{ii}b_i E\left(\varepsilon_{it}^3\right), \tag{52}
\end{aligned}$$

where

$$b_i = \sum_{\ell=1}^{N}a_{\ell i}\phi_\ell,$$

and substituting (52) into (51) yields

$$Var(W_{1,a}) = Var(W_{0,a}) + 4\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)\left(\boldsymbol{\phi}'\mathbf{A}^2\boldsymbol{\phi}\right) + 4\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2\left(\sum_{t=1}^{T}c_t^3\right)\sum_{i=1}^{N}a_{ii}b_i E\left(\varepsilon_{it}^3\right). \tag{53}$$

Recall that $\mathbf{c} = \mathbf{M}_F\boldsymbol{\tau}_T/\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T$, and therefore $\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2\left(\sum_{t=1}^{T}c_t^3\right) = O_p(1)$.

Appendix B: Proof of Theorem 3

Using the factor model (23), the correlation coefficient of the errors u_{it} and u_{jt} is given by

$$\rho_{ij}^2 = \begin{cases} \frac{\left(\boldsymbol{\varphi}_i'\boldsymbol{\varphi}_j\right)^2}{\left(\omega_i^2 + \boldsymbol{\varphi}_i'\boldsymbol{\varphi}_i\right)\left(\omega_j^2 + \boldsymbol{\varphi}_j'\boldsymbol{\varphi}_j\right)} = \mathbf{b}_i'\mathbf{b}_j\mathbf{b}_j'\mathbf{b}_i, & \text{for all } i \neq j, \\ 1, & \text{for } i = j, \end{cases}$$

where $\mathbf{b}_i = \boldsymbol{\varphi}_i/\sigma_i$. Therefore,

$$N^{-1}Tr(\mathbf{R}^2) = 1 + N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{b}_i'\mathbf{b}_j\mathbf{b}_j'\mathbf{b}_i - N^{-1}\sum_{i=1}^{N}\left(\mathbf{b}_i'\mathbf{b}_i\right)^2.$$

But

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{b}_i'\mathbf{b}_j\mathbf{b}_j'\mathbf{b}_i \;=\; Tr\left[\sum_{i=1}^{N}\mathbf{b}_i'\left(\sum_{j=1}^{N}\mathbf{b}_j\mathbf{b}_j'\right)\mathbf{b}_i\right]=Tr\left[\sum_{i=1}^{N}\mathbf{b}_i'\mathbf{B}\mathbf{b}_i\right]$$

$$=\; Tr\left[\mathbf{B}\sum_{i=1}^{N}\mathbf{b}_i\mathbf{b}_i'\right]=Tr(\mathbf{B}^2),$$

where $\mathbf{B}=\sum_{i=1}^{N}\mathbf{b}_i\mathbf{b}_i'$. Using familiar matrix trace inequalities (see, for example, Magnus and Neudecker (1999, p.204))

$$Tr(\mathbf{B}^2)\leq\lambda_{\max}(\mathbf{B})Tr(\mathbf{B}).$$

Also since $\mathbf{B}$ is a positive semi-definite matrix with non-negative eigenvalues, $\lambda_{\max}(\mathbf{B})\leq Tr(\mathbf{B})$, and we have

$$Tr(\mathbf{B}^2)\leq[Tr(\mathbf{B})]^2.$$

Also,

$$\sum_{i=1}^{N}\left(\mathbf{b}_i'\mathbf{b}_i\right)^2\leq\left(\sum_{i=1}^{N}\mathbf{b}_i'\mathbf{b}_i\right)^2=[Tr(\mathbf{B})]^2.$$

Using these results we now have

$$N^{-1}Tr(\mathbf{R}^2)=1+N^{-1}Tr(\mathbf{B}^2)-N^{-1}\sum_{i=1}^{N}\left(\mathbf{b}_i'\mathbf{b}_i\right)^2\leq 1+\left[\frac{Tr(\mathbf{B})}{\sqrt{N}}\right]^2.$$

However, $N^{-1/2}Tr(\mathbf{B})=N^{-1/2}\sum_{i=1}^{N}\mathbf{b}_i'\mathbf{b}_i$, and for all values of $\delta_b<1/2$, (defined by (24)) we must have

$$\lim_{N\to\infty}N^{-1}Tr(\mathbf{R}^2)=1. \tag{54}$$

The equality holds since $N^{-1/2}Tr(\mathbf{B})\to 0$, simultaneously ensures that $N^{-1}Tr(\mathbf{B}^2)\to 0$ and $N^{-1}\sum_{i=1}^{N}\left(\mathbf{b}_i'\mathbf{b}_i\right)^2\to 0$.

Suppose now that $\gamma_{2,\varepsilon i}$ and a_{ii} are independently distributed and $E(\gamma_{2,\varepsilon i})=\bar{\gamma}_{2,\varepsilon}$, noting that ε_{it} are *iid* random variables we also have

$$N^{-1}\sum_{i=1}^{N}a_{ii}^2\gamma_{2,\varepsilon i}\to_p\bar{\gamma}_{2,\varepsilon}\lim_{N\to\infty}\left(N^{-1}\sum_{i=1}^{N}a_{ii}^2\right). \tag{55}$$

However, since $N^{-1}\sum_{i=1}^{N}a_{ii}^2\leq N^{-1}Tr(\mathbf{R}^2)$, then we also have $N^{-1}\sum_{i=1}^{N}a_{ii}^2\to 1$, as $N\to\infty$. Finally, using the results (54) and (55), the $J_\alpha(\mathbf{D})$ test statistic defined by (22) further simplifies to

$$J_\alpha(\mathbf{D})=\frac{N^{-1/2}\left[\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)\hat{\boldsymbol{\alpha}}'\mathbf{D}^{-1}\hat{\boldsymbol{\alpha}}-N\right]}{\sqrt{2+\bar{\gamma}_{2,\varepsilon}q_T}}\to_d N(0,1), \tag{56}$$

as required.

Appendix C: Data Description and Sources

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 September 2011 (inclusive) using Datastream.[20] For example, the code LS&PCOMP1210 will give the 500 constituents of S&P500 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 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

---

[20] 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.

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)
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
HML_t	Average return in per cent on two value portfolios minus the average return on two growth portfolios.	Ken French's data library
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
r_{mt}	Value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) in per cent.	Ken French's data library
r_t	Monthly return of S&P500 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) http://www.hedgeindex.com
$\tilde{r}_{ht}$	$r_{ht} - r_t$.	

Appendix D: Derivation of $E\left(t_i^2\right)$ and $Var\left(t_i^2\right)$ up to $O(v^{-3/2})$ under Non-Gaussian Errors

Recall the model

$$\mathbf{y}_{i.} = \alpha_i \boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i.}, \tag{57}$$

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 $\mathbf{u}_{i.} = (u_{i1}, u_{i2}, ..., u_{iT})'$. Define

$$\boldsymbol{\xi}_i = \mathbf{u}_{i.}/\sigma_i, \tag{58}$$

with $E(u_{it}^2) = \sigma_i^2$, so that $E(\xi_{it}) = 0$, $E(\xi_{it}^2) = 1$, $\gamma_{1,i} = E(\xi_{it}^3)$, $\gamma_{2,i} = E(\xi_{it}^4) - 3$ such that under normality, $\gamma_{\ell,i} = 0$ for all i and $\ell = 1,2$. Note that under Assumptions 2^a and 3, $|\gamma_{\ell,i}| < K$ for all i and $\ell = 1,2$.

The t-ratio for testing $\alpha_i = 0$ is given by

$$t_i^2 = \frac{v}{\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T} \left(\frac{\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i}{\boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i} \right), \tag{59}$$

where $\boldsymbol{\xi}_i \sim IID(\mathbf{0}, \mathbf{I}_T)$,

$$\mathbf{M}_F = \mathbf{I}_T - \mathbf{F}\left(\mathbf{F}'\mathbf{F}\right)^{-1}\mathbf{F}', \quad \mathbf{H}_F = \mathbf{M}_F \boldsymbol{\tau}_T \boldsymbol{\tau}_T' \mathbf{M}_F. \tag{60}$$

$$\mathbf{M}_G = \mathbf{I}_T - \mathbf{P}_G, \quad \mathbf{P}_G = \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}', \quad \mathbf{G} = (\mathbf{F}, \boldsymbol{\tau}_T), v = Tr(\mathbf{M}_G) = T - m - 1. \tag{61}$$

For notational convenience, without loss of generality, all the results in this section are obtained assuming that $\mathbf{F}$ is nonstochastic.

Initially, we state and proof a number of useful lemmas, before giving the proposition for the expression of $E\left(t_i^2\right)$ and $Var\left(t_i^2\right)$ up to $O(v^{-3/2})$ under non-normality:

Lemma 7 (Lieberman 1994) *Let $\boldsymbol{\Phi}$ be a $T \times T$ symmetric matrix and $\boldsymbol{\Gamma}$ a positive definite $T \times T$ matrix, and suppose that $\boldsymbol{\xi} \sim IID(\mathbf{0}, \mathbf{I}_T)$, where $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_T)'$. Denote the pth cumulant of $\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}$ by κ_p, and the $m+1$ order, $m+r$ degree generalized cumulant of $(\boldsymbol{\xi}'\boldsymbol{\Phi}\boldsymbol{\xi})^r(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})$ by κ_{rm}, and assume that the following conditions hold:*

- *Condition 1: For $p = 1,2,...,\kappa_p = O(T)$.*

- *Condition 2: For $r = 1,2,...,\kappa_{r0} = E(\boldsymbol{\xi}'\boldsymbol{\Phi}\boldsymbol{\xi})^r = O(T^r)$.*

- *Condition 3: For $r, m = 1,2,...,\kappa_{rm} = O(T^\ell)$, with $\ell \leq r$.*
  Then the Laplace approximate expansion for the rth moment of $\boldsymbol{\xi}'\boldsymbol{\Phi}\boldsymbol{\xi}/\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}$ is given by

$$E\left[\left(\frac{\boldsymbol{\xi}'\boldsymbol{\Phi}\boldsymbol{\xi}}{\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}}\right)^r\right] = \frac{E[(\boldsymbol{\xi}'\boldsymbol{\Phi}\boldsymbol{\xi})^r]}{[E(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})]^r} + \psi_{rT} + O(T^{-2}), \tag{62}$$

where

$$\psi_{rT} = \frac{r(r+1)}{2}\left[\frac{E\left[(\boldsymbol{\xi}'\boldsymbol{\Phi}\boldsymbol{\xi})^r\right]\kappa_2}{[E(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})]^{r+2}}\right] - r\left[\frac{\kappa_{r1}}{[E(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi})]^{r+1}}\right], \tag{63}$$

and

$$\kappa_{r1} = E[(\boldsymbol{\xi}'\boldsymbol{\Phi}\boldsymbol{\xi})^r\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}] - E[(\boldsymbol{\xi}'\boldsymbol{\Phi}\boldsymbol{\xi})^r]E(\boldsymbol{\xi}'\boldsymbol{\Gamma}\boldsymbol{\xi}). \tag{64}$$

Proof. See Lieberman (1994). ∎

Lemma 8 *(Moments of Products of Quadratic Forms under Nonnormality): Let* $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_T)' \sim IID(\mathbf{0}, \mathbf{I}_T)$, *with* $E(\xi_t^3) = \gamma_1$, $E(\xi_t^4) = \gamma_2 + 3$, $E(\xi_t^5) = \gamma_3 + 10\gamma_1$, $E(\xi_t^6) = \gamma_4 + 15\gamma_2 + 10\gamma_1^2 + 15$, *and suppose that* $\mathbf{A}_j$, $j = 1,2,3$, *are* $T \times T$ *real symmetric matrices, and* $\boldsymbol{\tau}_T$ *is a* $T \times 1$ *vector of ones. Then*

$$E\left(\boldsymbol{\xi}'\mathbf{A}_1\boldsymbol{\xi}\right) = Tr(\mathbf{A}_1) \tag{65}$$

$$E\left[\left(\boldsymbol{\xi}'\mathbf{A}_1\boldsymbol{\xi}\right)\left(\boldsymbol{\xi}'\mathbf{A}_2\boldsymbol{\xi}\right)\right] = \gamma_2 Tr\left[(\mathbf{A}_1 \odot \mathbf{A}_2)\right] + Tr\left(\mathbf{A}_1\right)Tr(\mathbf{A}_2) + 2Tr\left(\mathbf{A}_1\mathbf{A}_2\right) \tag{66}$$

$$E\left[\left(\boldsymbol{\xi}'\mathbf{A}_1\boldsymbol{\xi}\right)\left(\boldsymbol{\xi}'\mathbf{A}_2\boldsymbol{\xi}\right)\left(\boldsymbol{\xi}'\mathbf{A}_3\boldsymbol{\xi}\right)\right] = \gamma_4 Tr\left(\mathbf{A}_1 \odot \mathbf{A}_2 \odot \mathbf{A}_3\right) + \gamma_2 Tr\left(\mathbf{A}_1\right)Tr\left(\mathbf{A}_2 \odot \mathbf{A}_3\right) \tag{67}$$
$$+\gamma_2 Tr\left(\mathbf{A}_2\right)Tr\left(\mathbf{A}_1 \odot \mathbf{A}_3\right) + \gamma_2 Tr\left(\mathbf{A}_3\right)Tr\left(\mathbf{A}_1 \odot \mathbf{A}_2\right) + 4\gamma_2 Tr\left[\mathbf{A}_1 \odot (\mathbf{A}_2\mathbf{A}_3)\right]$$
$$+4\gamma_2 Tr\left[\mathbf{A}_2 \odot (\mathbf{A}_1\mathbf{A}_3)\right] + 4\gamma_2 Tr\left[\mathbf{A}_3 \odot (\mathbf{A}_1\mathbf{A}_2)\right] + 2\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{A}_1\right)\mathbf{A}_2\left(\mathbf{I}_T \odot \mathbf{A}_3\right)\boldsymbol{\tau}_T\right]$$
$$+2\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{A}_1\right)\mathbf{A}_3\left(\mathbf{I}_T \odot \mathbf{A}_2\right)\boldsymbol{\tau}_T\right] + 2\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{A}_2\right)\mathbf{A}_1\left(\mathbf{I}_T \odot \mathbf{A}_3\right)\boldsymbol{\tau}_T\right]$$
$$+4\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{A}_1 \odot \mathbf{A}_2 \odot \mathbf{A}_3\right)\boldsymbol{\tau}_T\right] + Tr\left(\mathbf{A}_1\right)Tr\left(\mathbf{A}_2\right)Tr\left(\mathbf{A}_3\right) + 2Tr\left(\mathbf{A}_1\right)Tr\left(\mathbf{A}_2\mathbf{A}_3\right)$$
$$+2Tr\left(\mathbf{A}_2\right)Tr\left(\mathbf{A}_1\mathbf{A}_3\right) + 2Tr\left(\mathbf{A}_3\right)Tr\left(\mathbf{A}_1\mathbf{A}_2\right) + 8Tr\left(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3\right),$$

where $\odot$ *represents the Hadamard (or element-wise) product matrix operator.*
Proof. *For (65) and (66), see Ullah (2004, Appendix A.5). Result (67) was provided to us through a private communication by Yong Bao. Also, see Bao and Ullah (2006).* ∎

Lemma 9 *Suppose that* $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_T)' \sim IID(\mathbf{0}, \mathbf{I}_T)$, $\gamma_1 = E(\xi_t^3)$, $\gamma_2 = E(\xi_t^4) - 3$ *and* $\gamma_4 = E\left(\xi_t^6\right) - 15\gamma_2 - 10\gamma_1^2 - 15$, *for all t. Then,*

$$E\left(\boldsymbol{\xi}'\mathbf{H}_F\boldsymbol{\xi}\right) = Tr(\mathbf{H}_F) = \boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T$$

$$E\left(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi}\right) = Tr(\mathbf{M}_G) = v$$

$$E\left[\left(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi}\right)^2\right] = \gamma_2 Tr\left(\mathbf{M}_G \odot \mathbf{M}_G\right) + v\left(v+2\right)$$

$$E\left[\left(\boldsymbol{\xi}'\mathbf{H}_F\boldsymbol{\xi}\right)\left(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi}\right)\right] = \gamma_2 Tr\left(\mathbf{M}_G \odot \mathbf{H}_F\right) + v(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)$$

$$E\left[\left(\boldsymbol{\xi}'\mathbf{H}_F\boldsymbol{\xi}\right)^2\right] = \gamma_2 Tr\left(\mathbf{H}_F \odot \mathbf{H}_F\right) + 3\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2$$

$$E\left[\left(\boldsymbol{\varepsilon}'\mathbf{H}_F\boldsymbol{\varepsilon}\right)^2\left(\boldsymbol{\varepsilon}'\mathbf{M}_G\boldsymbol{\varepsilon}\right)\right] = \gamma_4 Tr\left(\mathbf{H}_F \odot \mathbf{H}_F \odot \mathbf{M}_G\right) + 2\gamma_2 Tr\left(\mathbf{H}_F\right)Tr\left(\mathbf{H}_F \odot \mathbf{M}_G\right)$$
$$+\gamma_2 Tr\left(\mathbf{M}_G\right)Tr\left(\mathbf{H}_F \odot \mathbf{H}_F\right) + 4\gamma_2 Tr\left[\mathbf{M}_G \odot \mathbf{H}_F^2\right] + 4\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{H}_F\left(\mathbf{I}_T \odot \mathbf{M}_G\right)\boldsymbol{\tau}_T\right]$$
$$+2\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{M}_G\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\boldsymbol{\tau}_T\right] + 4\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{H}_F \odot \mathbf{H}_F \odot \mathbf{M}_G\right)\boldsymbol{\tau}_T\right] + 3\left[Tr\left(\mathbf{H}_F\right)\right]^2 Tr\left(\mathbf{M}_G\right)$$

where $\mathbf{M}_F$, $\mathbf{M}_G$, *and* $\mathbf{H}_F = \mathbf{M}_F\boldsymbol{\tau}_T\boldsymbol{\tau}_T'\mathbf{M}_F$, *are defined by (60) and (61).*
Proof. *These results are obtained noting that* $\mathbf{M}_G\mathbf{H}_F = \mathbf{0}$ *(since* $\mathbf{M}_F\mathbf{M}_G = \mathbf{M}_G$ *and* $\mathbf{M}_G\boldsymbol{\tau}_T = \mathbf{0}$*), and* $Tr(\mathbf{H}_F^2) = (\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)^2$, *and using the results established in Lemma 8.* ∎

Lemma 10 *Let* $\mathbf{A}$ *be a real symmetric* $T \times T$ *matrix with eigenvalues* $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_T$. *Then* $\lambda_1 \leq a_{tt} \leq \lambda_T$, *where* a_{tt} *is the* t^{th} *diagonal element of* $\mathbf{A}$.
Proof. *See Theorem 14 in Chapter11 of Magnus and Neudecker (1999, p.211-212).* ∎

Lemma 11 *Denote the* (t,s) *elements of* $\mathbf{M}_F$, $\mathbf{M}_G$, *and* $\mathbf{P}_G$ *(defied in (60) and (61)), by* $m_{F,ts}$, $m_{G,ts}$ *and* $p_{G,ts}$, *respectively, and suppose that* $\mathbf{f}_t'\mathbf{f}_t \leq \Delta_2 < \infty$ *for all t, where* Δ_2 *is a finite positive constant. Then for all t we have*

$$0 \leq m_{F,tt} = \sum_{s=1}^T m_{F,ts}^2 \leq 1, \tag{68}$$

$$0 \leq m_{G,tt} = \sum_{s=1}^T m_{G,ts}^2 \leq 1, \tag{69}$$

31

$$0 \leq p_{G,tt} = \sum\nolimits_{s=1}^{T} p_{G,ts}^2 \leq 1, \tag{70}$$

$$\sum\nolimits_{s=1}^{T} m_{F,ts} \leq \Delta_{M_F} < \infty, \tag{71}$$

$$\sum\nolimits_{s=1}^{T} m_{G,ts} \leq \Delta_{M_G} < \infty, \tag{72}$$

where Δ_{M_F} and Δ_{M_G} are finite positive constants.

Proof. (68), (69) and (70) follow immediately using Lemmas 10, since $\mathbf{M}_F$, $\mathbf{M}_G$ and $\mathbf{P}_G$ are idempotent and real symmetric matrices, eigenvalues of which are either one or zero. Next we note that

$$\mathbf{M}_F \boldsymbol{\tau}_T = \boldsymbol{\tau}_T - \mathbf{F}\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1}\frac{\mathbf{F}'\boldsymbol{\tau}_T}{T},$$

where by Assumption 1 all elements of $\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1}$ and $\frac{\mathbf{F}'\boldsymbol{\tau}_T}{T}$ are bounded. Let $\mathbf{z}_{F,T} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1}\frac{\mathbf{F}'\boldsymbol{\tau}_T}{T}$, and note that the m elements of $\mathbf{z}_{F,T}$, being the OLS estimates of the coefficients in the regression of 1 on $\mathbf{f}_t$ are bounded, and hence $\sum_{\ell=1}^{m}|z_{F,T,\ell}|^2 \leq \Delta_1 < \infty$ for all T, where Δ_1 is a positive constant. Then, the tth element of $\mathbf{M}_F \boldsymbol{\tau}_T$ can be written as

$$\sum\nolimits_{s=1}^{T} m_{F,ts} = 1 - \mathbf{f}_t' \mathbf{z}_{F,T} = 1 - \sum\nolimits_{\ell=1}^{m} f_{t,\ell} z_{F,T,\ell}.$$

Hence,

$$\left|\sum\nolimits_{s=1}^{T} m_{F,ts}\right| \leq 1 + \sum\nolimits_{\ell=1}^{m}|f_{t,\ell} z_{F,T,\ell}|,$$

and by the assumption $\sum_{\ell=1}^{m}|f_{t,\ell}|^2 \leq \Delta_2 < \infty$, for all t, and a positive constant, Δ_2, we have

$$\left|\sum\nolimits_{\ell=1}^{m} f_{t,\ell} z_{F,T,\ell}\right| \leq \sum\nolimits_{\ell=1}^{m}|f_{t,\ell} z_{F,T,\ell}| \leq \sum\nolimits_{\ell=1}^{m}|f_{t,\ell}|^2 \sum\nolimits_{\ell=1}^{m}|z_{F,T,\ell}|^2 \leq \Delta_1 \Delta_2 < \infty.$$

Therefore, $\sum_{s=1}^{T} m_{F,ts} \leq \Delta_{M_F} < \infty$ follows as required. Similarly, we have $\sum_{s=1}^{T} m_{G,ts} \leq \Delta_{M_G} < \infty$. ■

Lemma 12 Denote the (t,s) elements of $\mathbf{M}_F$, $\mathbf{M}_G$, and $\mathbf{P}_G$ (defied in (60) and (61)), by $m_{F,ts}$, $m_{G,ts}$ and $p_{G,ts}$, respectively, $v = T - m - 1$, and suppose that $\mathbf{f}_t'\mathbf{f}_t \leq \Delta_2 < \infty$ for all t, where Δ_2 is a finite positive constant. Then

$$\sum\nolimits_{t=1}^{T}\left(\sum\nolimits_{s=1}^{T} m_{F,ts}\right)^j = O(v), \; j = 1,2,3,4, \tag{73}$$

$$\sum\nolimits_{t=1}^{T} m_{G,tt}\left(\sum\nolimits_{s=1}^{T} m_{F,ts}\right)^j = O(v), \; j = 1,2,3,4, \tag{74}$$

$$\sum\nolimits_{t=1}^{T} p_{G,tt}\left(\sum\nolimits_{s=1}^{T} m_{F,ts}\right)^2 = O\left(v^{1/2}\right), \tag{75}$$

$$\sum\nolimits_{t=1}^{T}\left(\sum\nolimits_{s=1}^{T} m_{F,ts}\right)^2 \sum\nolimits_{\ell=1}^{T} m_{G,t\ell}\left(\sum\nolimits_{s=1}^{T} m_{F,\ell s}\right)^2 = O\left(v^{3/2}\right). \tag{76}$$

Proof. By (71) in Lemma 11, $\sum_{t=1}^{T}\left(\sum_{s=1}^{T} m_{F,ts}\right)^j \leq \sum_{t=1}^{T}\left|\sum_{s=1}^{T} m_{F,ts}\right|^j \leq v\left(T/v\right)\Delta_{M_F}^j = O(v)$, $j = 1,2,3,4$, which establishes (73). By (69) and (71) in Lemma 11, $\sum_{t=1}^{T} m_{G,tt}\left(\sum_{s=1}^{T} m_{F,ts}\right)^j \leq \sum_{t=1}^{T}|m_{G,tt}|\left|\sum_{s=1}^{T} m_{F,ts}\right|^j \leq \sum_{t=1}^{T}\left|\sum_{s=1}^{T} m_{F,ts}\right|^j = O(v)$ by (73), which establishes (74). By Cauchy-Schwarz inequality, and (71) in Lemma 11,

$$\sum\nolimits_{t=1}^{T} p_{G,tt}\left(\sum\nolimits_{s=1}^{T} m_{F,ts}\right)^2 \leq \sqrt{\sum\nolimits_{t=1}^{T} p_{G,tt}^2}\sqrt{\sum\nolimits_{t=1}^{T}\left(\sum\nolimits_{s=1}^{T} m_{F,ts}\right)^4}.$$

But, since $0 \leq p_{G,tt}^2 \leq p_{G,tt} \leq 1$, then $\sum_{t=1}^{T} p_{G,tt}^2 \leq \sum_{t=1}^{T} p_{G,tt} = m + 1$, and using (71), we have

$$\sum\nolimits_{t=1}^{T} p_{G,tt}\left(\sum\nolimits_{s=1}^{T} m_{F,ts}\right)^2 \leq \sqrt{(m+1)\, v\left(T/v\right)\Delta_{M_F}^4} = O(v^{1/2}),$$

which establishes (75). Finally,

$$\sum\nolimits_{t=1}^{T}\left(\sum\nolimits_{s=1}^{T} m_{F,ts}\right)^2 \sum\nolimits_{\ell=1}^{T} m_{G,t\ell}\left(\sum\nolimits_{s=1}^{T} m_{F,\ell s}\right)^2 \leq \sum\nolimits_{t=1}^{T}\left|\sum\nolimits_{s=1}^{T} m_{F,ts}\right|^2\left|\sum\nolimits_{\ell=1}^{T} m_{G,t\ell}\left(\sum\nolimits_{s=1}^{T} m_{F,\ell s}\right)^2\right|, \tag{77}$$

but, by Cauchy-Schwarz inequality, and (71) and (69) in Lemma 11, we have

$$\left|\sum\nolimits_{\ell=1}^{T} m_{G,t\ell}\left(\sum\nolimits_{s=1}^{T} m_{F,\ell s}\right)^2\right| \leq \sqrt{\sum\nolimits_{\ell=1}^{T} m_{G,t\ell}^2}\sqrt{\sum\nolimits_{\ell=1}^{T}\left|\sum\nolimits_{s=1}^{T} m_{F,\ell s}\right|^4} \leq \sqrt{m_{G,tt} T \Delta_{M_F}^4}, \tag{78}$$

32

and substituting (78) into (77) and using (69) yield

$$\sum_{t=1}^{T} \left(\sum_{s=1}^{T} m_{F,ts} \right)^2 \sum_{\ell=1}^{T} m_{G,t\ell} \left(\sum_{s=1}^{T} m_{F,\ell s} \right)^2 \leq \sum_{t=1}^{T} \left| \sum_{s=1}^{T} m_{F,ts} \right|^2 \sqrt{m_{G,tt} T \Delta_{M_F}^4} \leq \Delta_{M_F}^4 \sqrt{v^3 \left(T/v \right)^3} = O(v^{3/2}),$$

which establishes (76). ∎

Lemma 13 *Suppose that* $\mathbf{A}_j = (a_{j,ts})$, *for* $j = 1, 2, 3$ *are* $T \times T$ *real symmetric matrices. Then,*

$$Tr \left(\mathbf{A}_1 \odot \mathbf{A}_2 \right) = \sum_{t=1}^{T} a_{1,tt} a_{2,tt}, \tag{79}$$

$$Tr \left(\mathbf{A}_1 \odot \mathbf{A}_2 \odot \mathbf{A}_3 \right) = \sum_{t=1}^{T} a_{1,tt} a_{2,tt} a_{3,tt}, \tag{80}$$

$$\boldsymbol{\tau}_T' \left(\mathbf{I}_T \odot \mathbf{A}_1 \right) \mathbf{A}_2 \left(\mathbf{I}_T \odot \mathbf{A}_3 \right) \boldsymbol{\tau}_T = \sum_{t=1}^{T} \sum_{s=1}^{T} a_{1,tt} a_{2,ts} a_{3,ss}, \tag{81}$$

$$\boldsymbol{\tau}_T' \left(\mathbf{A}_1 \odot \mathbf{A}_2 \odot \mathbf{A}_3 \right) \boldsymbol{\tau}_T = \sum_{t=1}^{T} \sum_{s=1}^{T} a_{1,ts} a_{2,ts} a_{3,ts}. \tag{82}$$

Proof. *By direct computations.* ∎

Lemma 14 *Consider the matrices* $\mathbf{M}_G$, $\mathbf{P}_G$ *and* $\mathbf{H}_F$, *defined by (60) and (61), and* $v = Tr \left(\mathbf{M}_G \right)$. *Suppose* $\mathbf{f}_t' \mathbf{f}_t \leq K < \infty$ *for all t. Then,*

$$Tr \left(\mathbf{H}_F \odot \mathbf{H}_F \odot \mathbf{M}_G \right) = O \left(v \right), \tag{83}$$

$$Tr \left(\mathbf{H}_F \odot \mathbf{M}_G \right) = O \left(v \right), \tag{84}$$

$$Tr \left(\mathbf{H}_F \odot \mathbf{H}_F \right) = O \left(v \right), \tag{85}$$

$$Tr \left(\mathbf{P}_G \odot \mathbf{P}_F \right) \leq Tr \left(\mathbf{P}_F \right) = m + 1, \tag{86}$$

$$Tr \left(\mathbf{P}_G \odot \mathbf{H}_F \right) = O(v^{1/2}), \tag{87}$$

$$\boldsymbol{\tau}_T' \left(\mathbf{I}_T \odot \mathbf{H}_F \right) \mathbf{H}_F \left(\mathbf{I}_T \odot \mathbf{M}_G \right) \boldsymbol{\tau}_T = O \left(v^2 \right), \tag{88}$$

$$\boldsymbol{\tau}_T' \left(\mathbf{I}_T \odot \mathbf{H}_F \right) \mathbf{M}_G \left(\mathbf{I}_T \odot \mathbf{H}_F \right) \boldsymbol{\tau}_T = O(v^{3/2}), \tag{89}$$

and

$$\boldsymbol{\tau}_T' \left(\mathbf{H}_F \odot \mathbf{H}_F \odot \mathbf{M}_G \right) \boldsymbol{\tau}_T = O(v^{3/2}). \tag{90}$$

Proof. *Denote the* (t, s) *element of* $\mathbf{M}_F$, $\mathbf{M}_G$ *and* $\mathbf{P}_G$ *by* $m_{F,ts}$, $m_{G,ts}$ *and* $p_{G,ts}$, *respectively, and observe that the* (h, ℓ) *element of* $\mathbf{H}_F$ *is* $\left(\sum_{s=1}^{T} m_{F,hs} \right) \left(\sum_{s=1}^{T} m_{F,\ell s} \right)$. *Then using (80) in Lemma 13, and (74) in Lemma 12 yields*

$$Tr \left(\mathbf{H}_F \odot \mathbf{H}_F \odot \mathbf{M}_G \right) = \sum_{t=1}^{T} m_{G,tt} \left(\sum_{s=1}^{T} m_{F,ts} \right)^4 = O \left(v \right),$$

which establishes (83). Similarly, to establish (84) note that using (79) in Lemma 13, and (74) in Lemma 12 we have

$$Tr \left(\mathbf{H}_F \odot \mathbf{M}_G \right) = \sum_{t=1}^{T} m_{G,tt} \left(\sum_{s=1}^{T} m_{F,ts} \right)^2 = O \left(v \right).$$

Now using (79) in Lemma 13, then (73) in Lemma 12 we obtain

$$Tr \left(\mathbf{H}_F \odot \mathbf{H}_F \right) = \sum_{t=1}^{T} \left(\sum_{s=1}^{T} m_{F,ts} \right)^4 = O \left(v \right),$$

that establishes (85). Result (86) follows since $\sum_{t=1}^{T} p_{G,tt}^2 \leq \sum_{t=1}^{T} p_{G,tt} = m + 1$, *recalling that* $0 \leq p_{G,tt}^2 \leq p_{G,tt} \leq 1$ *by (70). Now using (79) in Lemma 13, and (75) in Lemma 12 we have*

$$Tr \left(\mathbf{P}_G \odot \mathbf{H}_F \right) = \sum_{t=1}^{T} p_{G,tt} \left(\sum_{s=1}^{T} m_{F,ts} \right)^2 = O(v^{1/2}),$$

which establishes (87). Further, using (81) in Lemma 13, (73) and (74) in Lemma 12 we have

$$\boldsymbol{\tau}_T' \left(\mathbf{I}_T \odot \mathbf{H}_F \right) \mathbf{H}_F \left(\mathbf{I}_T \odot \mathbf{M}_G \right) \boldsymbol{\tau}_T = \sum_{t=1}^{T} \left(\sum_{s=1}^{T} m_{F,ts} \right)^3 \sum_{\ell=1}^{T} m_{G,\ell\ell} \left(\sum_{s=1}^{T} m_{F,\ell s} \right) = O(v^2),$$

which establishes (88). Next, again using (81) in Lemma 13, and (76) in Lemma 12 we obtain

$$\boldsymbol{\tau}_T' \left(\mathbf{I}_T \odot \mathbf{H}_F \right) \mathbf{M}_G \left(\mathbf{I}_T \odot \mathbf{H}_F \right) \boldsymbol{\tau}_T = \sum_{t=1}^{T} \left(\sum_{s=1}^{T} m_{F,ts} \right)^2 \sum_{\ell=1}^{T} m_{G,t\ell} \left(\sum_{s=1}^{T} m_{F,\ell s} \right)^2 = O(v^{3/2}), \tag{91}$$

which establishes (89). Finally, using (82) in Lemma 13 and noting that the (h, ℓ) *element of* $\mathbf{H}_F$ *is* $\left(\sum_{s=1}^{T} m_{F,hs} \right) \left(\sum_{s=1}^{T} m_{F,\ell s} \right)$, *it is easily seen that* $\boldsymbol{\tau}_T' \left(\mathbf{H}_F \odot \mathbf{H}_F \odot \mathbf{M}_G \right) \boldsymbol{\tau}_T$ *is the same as the right hand side of (91), so that*

$$\boldsymbol{\tau}_T' \left(\mathbf{H}_F \odot \mathbf{H}_F \odot \mathbf{M}_G \right) \boldsymbol{\tau}_T = \boldsymbol{\tau}_T' \left(\mathbf{I}_T \odot \mathbf{H}_F \right) \mathbf{M}_G \left(\mathbf{I}_T \odot \mathbf{H}_F \right) \boldsymbol{\tau}_T = O(v^{3/2}) \tag{92}$$

which establishes (90). ∎

33

Lemma 15 *Suppose that* $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_T)' \sim IID(\mathbf{0}, \mathbf{I}_T)$, *and* $E(\xi_t^3) = \gamma_1$, $E(\xi_t^4) = \gamma_2 + 3$, $E(\xi_t^5) = \gamma_3 + 10\gamma_1$, *and* $E(\xi_t^6) = \gamma_4 + 10\gamma_1^2 + 15\gamma_2 + 15$, *for all* t. *Suppose that Assumptions 1 and 2^a hold and that:* (i) $\mathbf{f}_t'\mathbf{f}_t \leq K < \infty$ *for all* t, *where* Δ_2 *is finite constant, and*; (ii) $E(|\varepsilon_{it}|^{6+\epsilon}) < K < \infty$ *for some* $\epsilon > 0$ *exists. Then,*

$$\kappa_2 = E\left[(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi})^2\right] - \left[E(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi})\right]^2 = \gamma_2 Tr\left(\mathbf{M}_G \odot \mathbf{M}_G\right) + 2v = O(v), \tag{93}$$

$$\begin{aligned}
\kappa_{11} &= E[(\boldsymbol{\xi}'\mathbf{H}_F\boldsymbol{\xi})(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi})] - E(\boldsymbol{\xi}'\mathbf{H}_F\boldsymbol{\xi})E(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi}) \\
&= \gamma_2 Tr\left[(\mathbf{M}_G \odot \mathbf{H}_F)\right] = O(v),
\end{aligned} \tag{94}$$

$$\begin{aligned}
\kappa_{21} &= E\left[(\boldsymbol{\xi}'\mathbf{H}_F\boldsymbol{\xi})^2(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi})\right] - E[(\boldsymbol{\xi}'\mathbf{H}_F\boldsymbol{\xi})^2]E(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi}) \\
&= 6\gamma_2\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)Tr\left(\mathbf{M}_G \odot \mathbf{H}_F\right) + 4\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{H}_F\left(\mathbf{I}_T \odot \mathbf{M}_G\right)\boldsymbol{\tau}_T\right] \\
&\quad + 6\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{M}_G\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\boldsymbol{\tau}_T\right] + O(v) = O(v^2),
\end{aligned} \tag{95}$$

where $\mathbf{M}_G$, $\mathbf{H}_F$, *and* $\mathbf{M}_F$ *are defined by (60) and (61), and* $v = Tr(\mathbf{M}_G)$.

Proof. *The results (93) and (94) immediately follow using Lemmas 9 and 14. The result (95) follows using Lemmas 9 and 14, and the equality (92), noting that* $Tr\left(\mathbf{H}_F^2\right) = \left[Tr\left(\mathbf{H}_F\right)\right]^2$, *and* $Tr\left(\mathbf{M}_G \odot \mathbf{H}_F^2\right) = Tr\left(\mathbf{H}_F\right)Tr\left(\mathbf{M}_G \odot \mathbf{H}_F\right)$, *since* $\mathbf{H}_F^2 = Tr\left(\mathbf{H}_F\right)\mathbf{H}_F$. ∎

Proposition 16 *Consider the regression model (57), and suppose that Assumptions 1 and 2^a hold and that:* (i) $\mathbf{f}_t'\mathbf{f}_t \leq K < \infty$ *for all* t, *and*; (ii) $E(|\varepsilon_{it}|^{8+\epsilon}) < K < \infty$ *for some* $\epsilon > 0$ *exists. Then*

$$E\left(t_i^2\right) = \frac{v}{v-2} + \gamma_{2,i}\frac{S_{0v}}{v^{3/2}} + \frac{W_{0,iv}}{v^2} = \frac{v}{v-2} + O(v^{-3/2}), \tag{96}$$

$$Var\left(t_i^2\right) = \left(\frac{v}{v-2}\right)^2\frac{2(v-1)}{(v-4)} + \frac{\gamma_{1,i}^2 S_{1v} + \gamma_{2,i} S_{2v}}{v} + \frac{\gamma_{1,i}^2 S_{3v} + \gamma_{2,i} S_{4v}}{v^{3/2}} + \frac{W_{1,iv}}{v^2} \tag{97}$$

where t_i *is the t-ratio defined by (59),* $v = T - m - 1$,

$$S_{0v} = \frac{v^{1/2} Tr\left(\mathbf{P}_G \odot \mathbf{H}_F\right)}{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)}, \tag{98}$$

$$S_{1v} = -\frac{8\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{H}_F\left(\mathbf{I}_T \odot \mathbf{M}_G\right)\boldsymbol{\tau}_T\right]}{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}, \quad S_{2v} = vq_T - 3, \tag{99}$$

$$S_{3v} = -\frac{v^{1/2}12\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{M}_G\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\boldsymbol{\tau}_T\right]}{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}, \quad S_{4v} = -\frac{v^{1/2}10 Tr\left(\mathbf{P}_G \odot \mathbf{H}_F\right)}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}, \tag{100}$$

$\mathbf{G}$, $\mathbf{M}_F$, $\mathbf{M}_G$, $\mathbf{P}_G$, $\mathbf{H}_F$, *are defined by (60) and (61),* $S_{jv} = O(1)$ *for* $j = 0, 1, .., 4$, $\boldsymbol{\tau}_T = (1, 1, ..., 1)'$, $W_{0,iv}$ *and* $W_{1,iv}$ *are* $O(1)$ *functions of* $\mathbf{F}$, *and* v, *and* $\gamma_{\ell,i}$, $\ell = 1, 2, ..., 6$, *which are at most* $O(1)$, *are given by* $\gamma_{1,i} = E(\xi_{it}^3)$, $\gamma_{2,i} = E(\xi_{it}^4) - 3$, $\gamma_{3,i} = E(\xi_{it}^5) - 10\gamma_{1,i}$, $\gamma_{4,i} = E(\xi_{it}^6) - 10\gamma_{1,i}^2 - 15\gamma_{2,i} - 15$, $\gamma_{5,i} = E\left(\xi_{it}^7\right) - 21\gamma_{3,i} - 35\gamma_{2,i}\gamma_{1,i} - 105\gamma_{1,i}$, $\gamma_{6,i} = E\left(\xi_{it}^8\right) - 28\gamma_{4,i} - 56\gamma_{3,i}\gamma_{1,i} - 35\gamma_{2,i}^2 - 210\gamma_{2,i} - 280\gamma_{1,i}^2 - 105$, *where* $\xi_{it} = u_{it}/\sigma_i$, *and*

$$q_T = \left(\sum_{t=1}^{T} c_t^4\right) / \left(\sum_{t=1}^{T} c_t^2\right)^2, \tag{101}$$

with c_t *being the* t^{th} *element of* $\mathbf{c} = \mathbf{M}_F\boldsymbol{\tau}_T/\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T$.

Proof. *Using a slightly extended version of Laplace approximation of moments of the ratio of quadratic forms by Lieberman (1994), that allows $\boldsymbol{\Gamma}$ defined in Lemma 7 to be a positive semi-definite matrix, substituting $\boldsymbol{\Phi} = \mathbf{H}_F = \mathbf{M}_F\boldsymbol{\tau}_T\boldsymbol{\tau}_T'\mathbf{M}_F$ and $\boldsymbol{\Gamma} = \mathbf{M}_G$ into the Lieberman's formula, we have*

$$E\left(t_i^2\right) = \frac{v}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}\left[\frac{E\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i\right)}{E\left(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i\right)} + \psi_{i,1v}\right] + O(v^{-2}) \tag{102}$$

where

$$\psi_{i,1v} = \left[\frac{E(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i)\kappa_{i,2}}{[E(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i)]^3}\right] - \left[\frac{\kappa_{i,11}}{[E(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i)]^2}\right], \tag{103}$$

$$\kappa_{i,2} = E\left[(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i)^2\right] - \left[E(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i)\right]^2 \tag{104}$$

$$\kappa_{i,11} = E[(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i)(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i)] - E(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i)E(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i). \tag{105}$$

Using Lemma 9, it is easily seen that

$$\frac{v}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}\frac{E\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i\right)}{E\left(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i\right)} = 1$$

and hence

$$\frac{v\psi_{i,1v}}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T} = \frac{v}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}\left(\frac{E(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i)\kappa_{i,2}}{[E(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i)]^3} - \frac{\kappa_{i,11}}{[E(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i)]^2}\right)$$

$$= \frac{v}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}\left(\frac{(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)\left[\gamma_{2,i}Tr\left(\mathbf{M}_G\odot\mathbf{M}_G\right)+2v\right]}{v^3} - \frac{\gamma_{2,i}Tr\left(\mathbf{M}_G\odot\mathbf{H}_F\right)}{v^2}\right)$$

$$= \frac{2}{v} + \gamma_{2,i}K_v$$

where

$$K_v = \frac{1}{v}\left[\frac{Tr\left(\mathbf{M}_G\odot\mathbf{M}_G\right)}{v} - \frac{Tr\left(\mathbf{M}_G\odot\mathbf{H}_F\right)}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}\right]. \tag{106}$$

Noting that $\mathbf{M}_G = \mathbf{I}_T - \mathbf{P}_G$ with $\mathbf{P}_G = \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'$ with $\mathbf{G} = (\mathbf{F}, \boldsymbol{\tau}_T)$, we can write the first term in the square brackets of (106) as

$$\frac{Tr\left(\mathbf{M}_G\odot\mathbf{M}_G\right)}{v} = \frac{1}{v}Tr\left[(\mathbf{I}_T-\mathbf{P}_G)\odot(\mathbf{I}_T-\mathbf{P}_G)\right] \tag{107}$$

$$= \frac{1}{v}\left[T - 2Tr\left(\mathbf{P}_G\right) + Tr\left(\mathbf{P}_G\odot\mathbf{P}_G\right)\right] = 1 - \frac{Tr\left(\mathbf{P}_G\right)}{v} + \frac{Tr\left(\mathbf{P}_G\odot\mathbf{P}_G\right)}{v}.$$

Similarly, the second term in the square brackets of (106) can be written as

$$\frac{Tr\left(\mathbf{M}_G\odot\mathbf{H}_F\right)}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T} = \frac{1}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}Tr\left[(\mathbf{I}_T-\mathbf{P}_G)\odot\mathbf{H}_F\right] \tag{108}$$

$$= \frac{1}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}\left[Tr\left(\mathbf{H}_F\right) - Tr\left(\mathbf{P}_G\odot\mathbf{H}_F\right)\right] = 1 - \frac{Tr\left(\mathbf{P}_G\odot\mathbf{H}_F\right)}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}.$$

Substituting (107) and (108) into (106), then using $Tr\left(\mathbf{P}_G\odot\mathbf{P}_G\right) = O(1)$ and $Tr\left(\mathbf{P}_G\odot\mathbf{H}_F\right) = O(v^{1/2})$, which are established by (86) and (87) in Lemma 14, we have

$$K_v = \frac{1}{v^{3/2}}\frac{v^{1/2}Tr\left(\mathbf{P}_G\odot\mathbf{H}_F\right)}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T} + \frac{1}{v^2}Tr\left(\mathbf{P}_G\odot\mathbf{P}_G\right) - \frac{1}{v^2}Tr\left(\mathbf{P}_G\right) = \frac{S_{0v}}{v^{3/2}} + O(v^{-2}) \tag{109}$$

where

$$S_{0v} = \frac{v^{1/2}Tr\left(\mathbf{P}_G\odot\mathbf{H}_F\right)}{(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)}, \tag{110}$$

which is $O(1)$ by (87) and (47), *so that*

$$E\left(t_i^2\right) = 1 + \frac{2}{v} + \gamma_{2,i}\frac{S_{0v}}{v^{3/2}} + O(v^{-2}). \tag{111}$$

However, since

$$\frac{v}{v-2} - \left(1 + \frac{2}{v}\right) = \frac{4}{v\left(v-2\right)} = O(v^{-2})$$

and using Lemma *15 ensures that the three conditions in Lieberman's lemma are satisfied. Lemma 7 now implies that the last term can be rewritten as* $v^{-2}W_{0,iv}$, *where* $W_{0,iv} = O(1)$ *is a function of* $\mathbf{F}$, *and v. Then for given values of* $\gamma_{\ell,i}$, $\ell = 1,2,3,4$, *we have*

$$E\left(t_i^2\right) = \frac{v}{v-2} + \gamma_{2,i}\frac{S_{0v}}{v^{3/2}} + \frac{W_{0,iv}}{v^2}, \tag{112}$$

which establishes (96). To prove (97), we first note that

$$E\left(t_i^4\right) = \frac{v^2}{(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T)^2}E\left[\left(\frac{\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i}{\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i}\right)^2\right], \tag{113}$$

but *by Lammas 7 and 9 we have*

$$E\left[\left(\frac{\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i}{\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i}\right)^2\right] = \frac{E\left[(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i)^2\right]}{[E\left(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i\right)]^2} + \psi_{i,2v} + O(v^{-2})$$

$$= \frac{\gamma_{2,i}Tr\left(\mathbf{H}_F\odot\mathbf{H}_F\right)}{v^2} + \frac{3\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}{v^2} + \psi_{i,2v} + O(v^{-2}), \tag{114}$$

where

$$\psi_{i,2v} = 3\left[\frac{E\left[(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i)^2\right]\kappa_{i,2}}{[E(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i)]^4}\right] - 2\left[\frac{\kappa_{i,21}}{[E(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i)]^3}\right] = \frac{3E\left[(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i)^2\right]\kappa_{i,2}}{v^4} - \frac{2\kappa_{i,21}}{v^3}.$$

Using results in Lemmas 9 and 15, and noting that $Tr\left(\mathbf{H}_F \odot \mathbf{H}_F\right) = O(v)$ and $Tr\left(\mathbf{M}_G \odot \mathbf{M}_G\right) = O(v)$ (see Lemma 14 for a proof), it can be shown that

$$
\begin{aligned}
\frac{3E\left[\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i\right)^2\right]\kappa_{i,2}}{v^4} &= \frac{3\left[\gamma_{2,i}Tr\left(\mathbf{H}_F\odot\mathbf{H}_F\right)+3\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2\right]\left[\gamma_{2,i}Tr\left(\mathbf{M}_G\odot\mathbf{M}_G\right)+2v\right]}{v^4} \\
&= \frac{9\gamma_{2,i}\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2 Tr\left(\mathbf{M}_G\odot\mathbf{M}_G\right)}{v^4}+\frac{18\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}{v^3}+O(v^{-2}).
\end{aligned}
$$

Also

$$
\begin{aligned}
\frac{2\kappa_{i,21}}{v^3} &= \frac{12\gamma_{2,i}\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)Tr\left(\mathbf{M}_G\odot\mathbf{H}_F\right)}{v^3}+\frac{8\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T\odot\mathbf{H}_F\right)\mathbf{H}_F\left(\mathbf{I}_T\odot\mathbf{M}_G\right)\boldsymbol{\tau}_T\right]}{v^3} \\
&\quad+12\gamma_{1,i}^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T\odot\mathbf{H}_F\right)\mathbf{M}_G\left(\mathbf{I}_T\odot\mathbf{H}_F\right)\boldsymbol{\tau}_T\right]+O(v^{-2}),
\end{aligned}
$$

so that

$$
\begin{aligned}
\psi_{i,2T} &= \frac{18\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}{v^3}+\frac{9\gamma_{2,i}\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2 Tr\left(\mathbf{M}_G\odot\mathbf{M}_G\right)}{v^4} \\
&\quad-\frac{12\gamma_{2,i}\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)Tr\left(\mathbf{M}_G\odot\mathbf{H}_F\right)}{v^3}-\frac{8\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T\odot\mathbf{H}_F\right)\mathbf{H}_F\left(\mathbf{I}_T\odot\mathbf{M}_G\right)\boldsymbol{\tau}_T\right]}{v^3} \\
&\quad-\frac{12\gamma_{1,i}^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T\odot\mathbf{H}_F\right)\mathbf{M}_G\left(\mathbf{I}_T\odot\mathbf{H}_F\right)\boldsymbol{\tau}_T\right]}{v^3}+O(v^{-2}). \qquad (115)
\end{aligned}
$$

Substituting (114) and (115) into (113) yields

$$
\begin{aligned}
E\left(t_i^4\right) &= 3+\frac{18}{v}+\frac{\gamma_{2,i}Tr\left(\mathbf{H}_F\odot\mathbf{H}_F\right)}{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}+\frac{9\gamma_{2,i}Tr\left(\mathbf{M}_G\odot\mathbf{M}_G\right)}{v^2}-\frac{12\gamma_{2,i}Tr\left(\mathbf{M}_G\odot\mathbf{H}_F\right)}{v\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)} \qquad (116)\\
&\quad-\frac{8\gamma_1^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T\odot\mathbf{H}_F\right)\mathbf{H}_F\left(\mathbf{I}_T\odot\mathbf{M}_G\right)\boldsymbol{\tau}_T\right]}{v\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}-\frac{12\gamma_{1,i}^2\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T\odot\mathbf{H}_F\right)\mathbf{M}_G\left(\mathbf{I}_T\odot\mathbf{H}_F\right)\boldsymbol{\tau}_T\right]}{v\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}+O(v^{-2}).
\end{aligned}
$$

Using the result (111) we have

$$
\left[E\left(t_i^2\right)\right]^2 = \left(1+\frac{2}{v}\right)^2+\gamma_{2,i}\frac{2Tr\left(\mathbf{M}_G\odot\mathbf{M}_G\right)}{v^2}-\gamma_{2,i}\frac{2Tr\left(\mathbf{M}_G\odot\mathbf{H}_F\right)}{v\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)}+O(v^{-2}). \qquad (117)
$$

However

$$
\left[3+\frac{18}{v}-\left(1+\frac{2}{v}\right)^2\right]-\left(\frac{v}{v-2}\right)^2\frac{2\left(v-1\right)}{\left(v-4\right)}=-\frac{4\left(19v^3-70v^2+76v-16\right)}{v^2\left(v-2\right)^2\left(v-4\right)}=O(v^{-2}), \qquad (118)
$$

and also, noting that $\mathbf{H}_F = \mathbf{mm}'$ with $\mathbf{m} = \mathbf{M}_F\boldsymbol{\tau}_T = \left(\sum_{s=1}^{T}m_{F,1s},\sum_{s=1}^{T}m_{F,2s},...,\sum_{s=1}^{T}m_{F,Ts}\right)'$ and $\left(\mathbf{I}_T\odot\mathbf{H}_F\right) = diag\left(\mathbf{m}_{(2)}\right)$, $\mathbf{m}_{(2)} = \mathbf{m}\odot\mathbf{m}$, where $\odot$ is the Hadamard matrix operator (or element-wise operator), we have $Tr\left[\left(\mathbf{H}_F\odot\mathbf{H}_F\right)\right]=\mathbf{m}_{(2)}'\mathbf{m}_{(2)}$, so that

$$
\frac{\gamma_{2,i}Tr\left(\mathbf{H}_F\odot\mathbf{H}_F\right)}{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}=\gamma_{2,i}\frac{\mathbf{m}_{(2)}'\mathbf{m}_{(2)}}{\left(\mathbf{m}'\mathbf{m}\right)^2}. \qquad (119)
$$

Recalling that $\mathbf{c} = \mathbf{M}_F\boldsymbol{\tau}_T/\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T$, we can rewrite $\sum_{t=1}^{T}c_t^4 = \mathbf{m}_{(2)}'\mathbf{m}_{(2)}/\left(\mathbf{m}'\mathbf{m}\right)^4$ and $\sum_{t=1}^{T}c_t^2 = \mathbf{m}'\mathbf{m}/\left(\mathbf{m}'\mathbf{m}\right)^2$, so that q_T defined by (101) can be rewritten as

$$
q_T = \left(\sum_{t=1}^{T}c_t^4\right)/\left(\sum_{t=1}^{T}c_t^2\right)^2 = \frac{\mathbf{m}_{(2)}'\mathbf{m}_{(2)}}{\left(\mathbf{m}'\mathbf{m}\right)^2}=O(v^{-1}). \qquad (120)
$$

Using (116), (117), (118), (119) and (120), and noting that, by (107) and (108),

$$
\frac{7Tr\left(\mathbf{M}_G\odot\mathbf{M}_G\right)}{v}-\frac{10Tr\left(\mathbf{M}_G\odot\mathbf{H}_F\right)}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}=-3-\frac{10Tr\left(\mathbf{P}_G\odot\mathbf{H}_F\right)}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}+O(v^{-1}),
$$

we conclude that

$$
\begin{aligned}
Var\left(t_i^2\right) &= E\left(t_i^4\right)-\left[E\left(t_i^2\right)\right]^2 \\
&= \left(\frac{v}{v-2}\right)^2\frac{2\left(v-1\right)}{\left(v-4\right)}+\frac{\gamma_{1,i}^2 S_{1v}+\gamma_{2,i}S_{2v}}{v}+\frac{\gamma_{1,i}^2 S_{3v}+\gamma_{2,i}S_{4v}}{v^{3/2}}+O(v^{-2}), \qquad (121)
\end{aligned}
$$

36

where

$$S_{1v} = -\frac{8\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{H}_F\left(\mathbf{I}_T \odot \mathbf{M}_G\right)\boldsymbol{\tau}_T\right]}{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}, S_{2v} = vq_T - 3, \tag{122}$$

$$S_{3v} = -\frac{v^{1/2}12\left[\boldsymbol{\tau}_T'\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\mathbf{M}_G\left(\mathbf{I}_T \odot \mathbf{H}_F\right)\boldsymbol{\tau}_T\right]}{\left(\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T\right)^2}, S_{4v} = -\frac{v^{1/2}10Tr\left(\mathbf{P}_G \odot \mathbf{H}_F\right)}{\boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T}. \tag{123}$$

$S_{1v} = O(1)$ by (88) and (47), $S_{2v} = O(1)$ by (120) and (47), $S_{3v} = O(1)$ by (89) and (47), and $S_{4v} = O(1)$ by (87) and (47). *Lieberman's lemma implies that the last term can be rewritten as* $v^{-2}W_{1,iv}$, *where* $W_{1,iv}$ *is a function of* $\mathbf{F}$, v, *and* $\gamma_{\ell,i}$, *for* $\ell = 1, 2, ..., 6$. *It is easily seen that* $W_{1,iv} = O(1)$, *and the result (97) follows immediately.* ∎

Appendix E: Correlation between t_i^2 and t_j^2 when $Corr(u_{it}, u_{jt}) = \rho_{ij}$ under Gaussianity

We begin with the following lemma:

Lemma 17 *Suppose that* $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_T)' \sim IIDN(\mathbf{0}, \mathbf{I}_T)$. *Then,*

$$E\left\{\left[v/\left(\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi}\right)\right]^r\right\} = \frac{v^r}{(v-2)(v-4)...(v-2r)}, \tag{124}$$

so long as $v > 2r$, *where* $\mathbf{M}_G$ *is defined by (61),* $v = Tr(\mathbf{M}_G)$.
Proof. *The result follows noting that* $\boldsymbol{\xi}'\mathbf{M}_G\boldsymbol{\xi} \sim \chi_v^2$, *and then using established results for the moments of the inverse-χ^2 distribution. Also see Smith (1988) and Ullah (1974).* ∎

Now, using (59) and denoting, $w_T = \boldsymbol{\tau}_T'\mathbf{M}_F\boldsymbol{\tau}_T$, under the null we have

$$t_i^2 = \frac{\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i}{\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i}\frac{v}{w_T}, \text{ for } i = 1, 2, ..., N,$$

where $\mathbf{M}_G$, $\mathbf{H}_F$, and $\mathbf{M}_F$ are defined by (60) and (61) with $v = Tr(\mathbf{M}_G)$, $\boldsymbol{\xi}_i \sim N(\mathbf{0}, \mathbf{I}_T)$ with $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, ..., \xi_{iT})'$, $\xi_{it} = u_{it}/\sigma_i$ (defined by (58)). Consider t_i^2 and t_j^2 $(i \neq j)$ and denote the the correlation of the associated CAPM regression errors by $Corr(u_{it}, u_{jt}) = Corr(\xi_{it}, \xi_{jt}) = \rho_{ij}$. Then,

$$\boldsymbol{\xi}_j = \rho\boldsymbol{\xi}_i + \sqrt{1-\rho^2}\boldsymbol{\zeta} \text{, with } E\left(\boldsymbol{\zeta}\boldsymbol{\xi}_i'\right) = \mathbf{0}, \tag{125}$$

where $\boldsymbol{\zeta} \sim IIDN(\mathbf{0}, \mathbf{I}_T)$, and we have dropped the subscript ij from ρ_{ij} for notational convenience.

Note that $Corr\left(t_i^2, t_j^2\right) = Cov\left(t_i^2, t_j^2\right)/Var\left(t_i^2\right)$, and $Var\left(t_i^2\right) = Var\left(t_j^2\right) = \left(\frac{v}{v-2}\right)^2\frac{2(v-1)}{v-4}$. Now using Lemma 9, $Var\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i/w_T\right) = 2$, and it is easily seen that $\left[Var\left(t_i^2\right)\right]^{-1} - \left[Var\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i/w_T\right)\right]^{-1} = O(v^{-1})$. Therefore, since $\left|Cov\left(t_i^2, t_j^2\right)\right| \leq Var\left(t_j^2\right) < \infty$, we have

$$Corr\left(t_i^2, t_j^2\right) = \frac{Cov\left(t_i^2, t_j^2\right)}{Var\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i/w_T\right)} + O(v^{-1}).$$

Next, since $E\left(t_i^2\right) = E\left(t_j^2\right)$, then $Cov\left(t_i^2, t_j^2\right) = E\left(t_i^2t_j^2\right) - \left[E\left(t_i^2\right)\right]^2$. Also using Lemma 9 $E\left(t_i^2\right)^2 - \left[E\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i/w_T\right)\right]^2 = \left(\frac{v}{v-2}\right)^2 - 1 = O(v^{-1})$. Therefore

$$Corr\left(t_i^2, t_j^2\right) = \frac{E\left(t_i^2t_j^2\right) - \left[E\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i/w_T\right)\right]^2}{Var\left(\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i/w_T\right)} + O(v^{-1}). \tag{126}$$

Now

$$E\left(t_i^2t_j^2\right) = E\left[\left(\frac{\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i/w_T}{\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i/v}\right)\left(\frac{\boldsymbol{\xi}_j'\mathbf{H}_F\boldsymbol{\xi}_j/w_T}{\boldsymbol{\xi}_j'\mathbf{M}_G\boldsymbol{\xi}_j/v}\right)\right] = E\left[\frac{\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i}{w_T}\frac{\boldsymbol{\xi}_j'\mathbf{H}_F\boldsymbol{\xi}_j}{w_T}\right] + d_T \tag{127}$$

where

$$d_T = E\left\{\frac{\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i}{w_T}\frac{\boldsymbol{\xi}_j'\mathbf{H}_F\boldsymbol{\xi}_j}{w_T}\left[\frac{1}{\left(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i/v\right)\left(\boldsymbol{\xi}_j'\mathbf{M}_G\boldsymbol{\xi}_j/v\right)} - 1\right]\right\}.$$

But, by Cauchy-Schwarz inequality we have

$$d_T \leq E\left|\frac{\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i}{w_T}\frac{\boldsymbol{\xi}_j'\mathbf{H}_F\boldsymbol{\xi}_j}{w_T}\left[\frac{1}{\left(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i/v\right)\left(\boldsymbol{\xi}_j'\mathbf{M}_G\boldsymbol{\xi}_j/v\right)} - 1\right]\right|$$

$$\leq \sqrt{E\left|\frac{\boldsymbol{\xi}_i'\mathbf{H}_F\boldsymbol{\xi}_i}{w_T}\frac{\boldsymbol{\xi}_j'\mathbf{H}_F\boldsymbol{\xi}_j}{w_T}\right|^2}\sqrt{E\left|\frac{1}{\left(\boldsymbol{\xi}_i'\mathbf{M}_G\boldsymbol{\xi}_i/v\right)\left(\boldsymbol{\xi}_j'\mathbf{M}_G\boldsymbol{\xi}_j/v\right)} - 1\right|^2}. \tag{128}$$

Similarly,

$$E \left| \frac{\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i}{w_T} \frac{\boldsymbol{\xi}_j' \mathbf{H}_F \boldsymbol{\xi}_j}{w_T} \right|^2 \leq \sqrt{E \left| \frac{\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i}{w_T} \right|^4} \sqrt{E \left| \frac{\boldsymbol{\xi}_j' \mathbf{H}_F \boldsymbol{\xi}_j}{w_T} \right|^4},$$

and using results in Ullah (2004), we have $E \left| \frac{\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i}{w_T} \right|^2 = O(1)$, which also ensures that $E \left| \frac{\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i}{w_T} \frac{\boldsymbol{\xi}_j' \mathbf{H}_F \boldsymbol{\xi}_j}{w_T} \right|^2 = O(1)$, at most.

For notational simplicity let

$$X_i = \boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i / v \text{ and } X_j = \boldsymbol{\xi}_j' \mathbf{M}_G \boldsymbol{\xi}_j / v, \tag{129}$$

and write the second product of (128) as $E \left[\left(\frac{1}{X_i X_j} - 1 \right)^2 \right] > 0$. Also define the finite positive constant,

$$B_v = \sup E \left[\left(\frac{1}{X_i X_j} - 1 \right)^2 \right].$$

But noting that $\frac{1}{X_i} > 0$ and $\frac{1}{X_j} > 0$, and using the results that $\sup \{ f(z) + g(z) : z \in Z \} \leq \sup \{ f(z) : z \in Z \} + \sup \{ g(z) : z \in Z \}$, $\sup (aY) = a \sup Y$ for $a > 0$ and $\sup (bY) = b \inf Y$ for $b < 0$, we have

$$B_v = \sup \left[E \left(\frac{1}{X_i^2 X_j^2} \right) - 2 E \left(\frac{1}{X_i X_j} \right) + 1 \right] \leq \sup E \left(\frac{1}{X_i^2 X_j^2} \right) - 2 \inf E \left(\frac{1}{X_i X_j} \right) + 1. \tag{130}$$

First, by Cauchy-Schwarz inequality and noting that $E \left(\frac{1}{X_i^4} \right) = E \left(\frac{1}{X_j^4} \right)$, $E \left(\frac{1}{X_i^2 X_j^2} \right) \leq \sqrt{E \left(\frac{1}{X_i^4} \right)} \sqrt{E \left(\frac{1}{X_j^4} \right)} = E \left(\frac{1}{X_i^4} \right)$, or

$$\sup E \left(\frac{1}{X_i^2 X_j^2} \right) = E \left(\frac{1}{X_i^4} \right). \tag{131}$$

Next, noting that $E \left(\frac{1}{X_i} \right) = E \left(\frac{1}{X_j} \right)$ and $Var \left(\frac{1}{X_i} \right) = Var \left(\frac{1}{X_j} \right)$ we have

$$E \left(\frac{1}{X_i X_j} \right) = E \left(\frac{1}{X_i} \right) E \left(\frac{1}{X_j} \right) + Cov \left(\frac{1}{X_i}, \frac{1}{X_j} \right) = \left[E \left(\frac{1}{X_i} \right) \right]^2 + Var \left(\frac{1}{X_i} \right) \rho_{\left(\frac{1}{X_i}, \frac{1}{X_j} \right)},$$

where $\rho_{\left(\frac{1}{X_i}, \frac{1}{X_j} \right)}$ is the correlation coefficient between $\frac{1}{X_i}$ and $\frac{1}{X_j}$. Thus, $\inf E \left(\frac{1}{X_i X_j} \right) > 0$ is achieved when $\rho_{\left(\frac{1}{X_i}, \frac{1}{X_j} \right)} = -1$, since $E \left(\frac{1}{X_i} \right) > 0$ and $Var \left(\frac{1}{X_i} \right) > 0$:

$$\inf E \left(\frac{1}{X_i X_j} \right) = 2 \left[E \left(\frac{1}{X_i} \right) \right]^2 - E \left(\frac{1}{X_i^2} \right). \tag{132}$$

Therefore, substituting (131) and (132) into (130), then using (124) in Lemma 17, we obtain

$$B_v = \frac{v^4}{(v-2)(v-4)(v-6)(v-8)} - \frac{2v^2(v-6)}{(v-4)(v-2)^2} + 1 = O(v^{-1}),$$

and thus conclude that

$$E \left| \frac{1}{(\boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i / v)(\boldsymbol{\xi}_j' \mathbf{M}_G \boldsymbol{\xi}_j / v)} - 1 \right|^2 \leq B_v = O(v^{-1}). \tag{133}$$

Substituting (133) into (128), together with (127) and (126), we have

$$Corr \left(t_i^2, t_j^2 \right) = \frac{E \left((\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i / w_T)(\boldsymbol{\xi}_j' \mathbf{H}_F \boldsymbol{\xi}_j / w_T) \right) - \left[E (\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i / w_T) \right]^2}{Var (\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i / w_T)} + O(v^{-1/2}).$$

However, by (125),

$$(\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i)(\boldsymbol{\xi}_j' \mathbf{H}_F \boldsymbol{\xi}_j) = (\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i) \left(\rho_{ij}^2 \boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i + 2 \rho_{ij} \sqrt{1 - \rho_{ij}^2} \boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\zeta}_j + (1 - \rho_{ij}^2) \boldsymbol{\zeta}_j' \mathbf{H}_F \boldsymbol{\zeta}_j \right).$$

Under Gaussanity, noting that $E \left[(\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i / w_T)^2 \right] = Tr \left(\mathbf{H}_F^2 \right) + 2 \left[Tr (\mathbf{H}_F) \right]^2 = 3$ and $E \left[\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i / w_T \right] = 1$, we have

$$E \left((\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i / w_T)(\boldsymbol{\xi}_j' \mathbf{H}_F \boldsymbol{\xi}_j / w_T) \right) = 3 \rho_{ij}^2 + (1 - \rho_{ij}^2),$$

38

so that

$$Corr\left(t_i^2, t_j^2\right) = \frac{3\rho_{ij}^2 + \left(1 - \rho_{ij}^2\right) - 1}{2} + O(v^{-1/2}) = \rho_{ij}^2 + O(v^{-1/2}).$$

Also, it is clear that, when $\rho_{ij} = 0$, $Corr\left(t_i^2, t_j^2\right) = 0$ (i.e. no $O(v^{-1/2})$ term involves). Therefore

$$Corr\left(t_i^2, t_j^2\right) = \begin{cases} \rho_{ij}^2 + O(v^{-1/2}), & \text{for } i, j \text{ such that } \rho_{ij} \neq 0, \\ 0, & \text{for } i, j \text{ such that } \rho_{ij} = 0. \end{cases} \tag{134}$$

Since by assumption $\mathbf{R}$ is a sparse matrix, denoting the maximum number of non-zero elements in the rows of $\mathbf{R}$ by $r < K < \infty$, we have

$$N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} Corr(t_i^2, t_j^2) = 1 + \frac{2}{N} \sum_{i=2}^{N} \sum_{j=1}^{i-1} Corr(t_i^2, t_j^2) = 1 + (N-1)\overline{\rho^2} + O\left(\frac{2r}{v^{1/2}}\right)$$

$$= 1 + (N-1)\overline{\rho^2} + O\left(v^{-1/2}\right), \tag{135}$$

as required.

Appendix F: Proof for $Z_{NT} \rightarrow_p 0$ as N and $T \rightarrow \infty$ in no particular order under Gaussianity and cross-sectionally correlated errors

Recall that

$$Z_{NT} = N^{-1/2} \sum_{i=1}^{N} \left(t_i^2 - \frac{v}{v-2}\right) - N^{-1/2} \left[\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{D}^{-1} \hat{\boldsymbol{\alpha}} - N\right],$$

where

$$\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \mathbf{D}^{-1} \hat{\boldsymbol{\alpha}} = \sum_{i=1}^{N} t_i^2 \left(\frac{T \hat{\sigma}_i^2}{v \sigma_i^2}\right) = \sum_{i=1}^{N} t_i^2 \left(\frac{\boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i}{v}\right),$$

$$t_i^2 = \frac{\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i}{\boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i} \frac{v}{w_T},$$

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

$$Z_{NT} = N^{-1/2} \sum_{i=1}^{N} \left(t_i^2 - \frac{v}{v-2}\right) - N^{-1/2} \left[\sum_{i=1}^{N} t_i^2 \left(\frac{\boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i}{v}\right) - N\right]$$

$$= N^{-1/2} \sum_{i=1}^{N} \left[t_i^2 - t_i^2 \left(\frac{\boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i}{v}\right)\right] - \left[\frac{v\sqrt{N}}{v-2} - \sqrt{N}\right]. \tag{136}$$

We know that (by construction) $E(Z_{NT}) = 0$. Then it is sufficient to find conditions under which

$$Var\left(Z_{NT}\right) = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} cov\left(t_i^2 - t_i^2 \frac{\boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i}{v}, t_j^2 - t_j^2 \frac{\boldsymbol{\xi}_j' \mathbf{M}_G \boldsymbol{\xi}_j}{v}\right) \rightarrow 0.$$

For use below we note that

$$Var\left(Z_{NT}\right) = N^{-1} \sum_{i=1}^{N} var(t_i^2 - W_i) + 2N^{-1} \sum_{i=2}^{N} \sum_{j=1}^{i-1} cov\left(t_i^2 - W_i, t_j^2 - W_j\right), \tag{137}$$

where

$$W_i = t_i^2 \frac{\boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i}{v} = \boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i / w_T. \tag{138}$$

Firstly, consider $var(t_i^2 - W_i) = var(t_i^2) + var(W_i) - 2cov(t_i^2, W_i)$, and note that $E(W_i t_i^2)$ can be written equivalently as

$$E(W_i t_i^2) = E(W_i^2 / X_i),$$

where $X_i = \boldsymbol{\xi}_i' \mathbf{M}_G \boldsymbol{\xi}_i / v$, as defined by (129). But

$$E(W_i^2 / X_i) = E(W_i^2) + E\left[W_i^2 \left(\frac{1}{X_i} - 1\right)\right],$$

and by Cauchy-Schwarz inequality

$$E\left[W_i^2 \left(\frac{1}{X_i} - 1\right)\right] \leq \left[E(W_i^4)\right]^{1/2} \left(E\left(\frac{1}{X_i} - 1\right)^2\right)^{1/2},$$

and

$$E\left(\frac{1}{X_i} - 1\right)^2 = E\left(\frac{1}{X_i^2} + 1 - \frac{2}{X_i}\right).$$

Using the result (124) we have

$$E\left(\frac{1}{X_i^2} + 1 - \frac{2}{X_i}\right) = \frac{v^2}{(v-2)(v-4)} + 1 - \frac{2v}{v-2} = \frac{8 + 2v}{(v-2)(v-4)}.$$

It is also easily seen that $E(W_i^4) < K < \infty$, and

$$\left|E(W_i^2/X_i) - E(W_i^2)\right| = O\left(v^{-1/2}\right) \text{ for } v > 4. \tag{139}$$

Using (139), and recalling that $E(W_i^2) = 3$, $var(t_i^2) = \left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4}$, $var(W_i) = E(W_i^2) - 1 = 2$, $cov(t_i^2, W_i) = E(W_i t_i^2) - \frac{v}{v-2}$, we have

$$
\begin{aligned}
var(t_i^2 - W_i) &= var(t_i^2) + var(W_i) - 2cov(t_i^2, W_i) \\
&= \left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4} + 2 - 2\left[E(W_i t_i^2) - \frac{v}{v-2}\right] \\
&= \left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4} + 2 - 2\left[3 - \frac{v}{v-2}\right] - 2\left[E(W_i^2/X_i) - E(W_i^2)\right] \\
&= \frac{18v^2 - 64v + 64}{(v-2)^2(v-4)} + O\left(v^{-1/2}\right).
\end{aligned}
$$

Hence,

$$N^{-1} \sum_{i=1}^{N} var(t_i^2 - W_i) = O\left(v^{-1/2}\right), \text{ for all values of } N. \tag{140}$$

To derive the covariance terms we first note that (for $i \neq j$)

$$cov\left(t_i^2 - W_i, t_j^2 - W_j\right) = cov(t_i^2, t_j^2) - cov(t_i^2, W_j) - cov(W_i, t_j^2) + cov(W_i, W_j). \tag{141}$$

Set $\boldsymbol{\xi}_j = \rho_{ji}\boldsymbol{\xi}_i + \sqrt{1 - \rho_{ji}^2}\boldsymbol{\zeta}_{ji}$, where $\boldsymbol{\zeta}_{ji}$ is distributed independently of $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_{ji} \sim N(0, \mathbf{I}_T)$. Note that $\rho_{ij} = \rho_{ji}$. Then

$$cov(W_i, W_j) = E(W_i W_j) - 1,$$

and it is easily seen that

$$E(W_i W_j) = \rho_{ji}^2 E(W_i^2) + (1 - \rho_{ji}^2).$$

But from Lemma 9, $E(W_i^2) = 3$, and $cov(W_i, W_j) = 2\rho_{ji}^2$. Next

$$
\begin{aligned}
E\left(t_i^2 W_j\right) &= E\left[t_i^2 E\left(W_j | \boldsymbol{\xi}_i\right)\right] = \frac{\rho_{ij}^2 E\left(\boldsymbol{\xi}_i' \mathbf{H}_F \boldsymbol{\xi}_i t_i^2\right) + (1 - \rho_{ij}^2) w_T E(t_i^2)}{w_T} \\
&= \rho_{ij}^2 E\left(W_i t_i^2\right) + (1 - \rho_{ij}^2) E(t_i^2).
\end{aligned}
$$

Similarly,

$$E\left(t_j^2 W_i\right) = \rho_{ji}^2 E\left(W_j t_j^2\right) + (1 - \rho_{ji}^2) E(t_j^2) = E\left(t_i^2 W_j\right).$$

Recalling that $E(W_i) = 1$, $E(t_i^2) = v/(v-2)$, and using $E(t_i^2) - 1 = v/(v-2) - 1 = 2/(v-2)$ and (139), we have

$$
\begin{aligned}
E\left(W_i t_j^2\right) - E(W_i W_j) &= \rho_{ji}^2 E\left(W_j t_j^2\right) + (1 - \rho_{ji}^2) E(t_j^2) - 3\rho_{ji}^2 - (1 - \rho_{ji}^2) \\
&= O(v^{-1/2}),
\end{aligned}
$$

so that

$$cov(t_i^2, W_j) - cov(W_i, W_j) = O(v^{-1/2}).$$

Collecting terms we have

$$cov\left(t_i^2 - W_i, t_j^2 - W_j\right) = cov(t_i^2, t_j^2) - cov(W_i, W_j) + O(v^{-1/2}), \text{ for } i \neq j.$$

However, in Appendix E it is shown that $cov(t_i^2, t_j^2) - cov(W_i, W_j) = O(v^{-1/2})$. Hence, under Assumption 3 and assuming that $\mathbf{R}$ is sparse we have

$$
\begin{aligned}
\left|cov\left(t_i^2 - W_i, t_j^2 - W_j\right)\right| &= O\left(v^{-1/2}\right), \text{ if } \rho_{ij} \neq 0 \\
&= 0, \text{ if } \rho_{ij} = 0.
\end{aligned}
$$

If the maximum number of non-zero elements of the rows of $\mathbf{R}$ is denoted by r then

$$\left| N^{-1} \sum_{i=2}^{N} \sum_{j=1}^{i-1} cov\left(t_i^2 - W_i, t_j^2 - W_j\right) \right| = O\left(\frac{r}{\sqrt{v}}\right),$$

and for a fixed r we have

$$\left| N^{-1} \sum_{i=2}^{N} \sum_{j=1}^{i-1} cov\left(t_i^2 - W_i, t_j^2 - W_j\right) \right| \to 0, \text{ as } v \to \infty. \tag{142}$$

From (137), (140) and (142), and recalling $E\left(Z_{NT}\right) = 0$, we conclude that $Z_{NT} \to_p 0$ as N and $T \to \infty$, in no particular order.

Appendix G: Derivation of $plim_{N,T\to\infty}(N-1)\widehat{\overline{\rho^2}}$ under Gaussianity and cross-sectionally correlated errors

First of all we state a lemma:

Lemma 18 *Suppose that $x \sim N(\rho, \sigma^2)$, then*

$$\begin{aligned} E\left[x^2 I(x^2 \leq a^2)\right] &= (\sigma^2 + \rho^2)\left[\Phi\left(\frac{a-\rho}{\sigma}\right) - \Phi\left(\frac{-a-\rho}{\sigma}\right)\right] \\ &+ \sigma(\rho - a)\phi(\frac{a+\rho}{\sigma}) - \sigma(a+\rho)\phi(\frac{a-\rho}{\sigma}). \end{aligned} \tag{143}$$

Proof. Let $z = (x - \rho)/\sigma$, and note that

$$E\left[x^2 I(x^2 \leq a^2)\right] = \int_{-(a+\rho)/\sigma}^{(a-\rho)/\sigma} \left(\sigma^2 z^2 + \rho^2 + 2\rho\sigma z\right)\phi(z)dz,$$

where $\phi(z) = (2\pi)^{-1/2}\exp(-0.5z^2)$. But

$$\begin{aligned} \int_{-(a+\rho)/\sigma}^{(a-\rho)/\sigma} z^2 \phi(z)dz &= [-z\phi(z)]_{-(a+\rho)/\sigma}^{(a-\rho)/\sigma} + \Phi\left(\frac{a-\rho}{\sigma}\right) - \Phi\left(\frac{-a-\rho}{\sigma}\right) \\ &= \Phi\left(\frac{a-\rho}{\sigma}\right) - \Phi\left(\frac{-a-\rho}{\sigma}\right) - \left(\frac{a+\rho}{\sigma}\right)\phi(\frac{a+\rho}{\sigma}) - \left(\frac{a-\rho}{\sigma}\right)\phi(\frac{a-\rho}{\sigma}). \end{aligned}$$

$$\int_{-(a+\rho)/\sigma}^{(a-\rho)/\sigma} z\phi(z)dz = \phi(\frac{a+\rho}{\sigma}) - \phi(\frac{a-\rho}{\sigma})$$

Hence

$$E\left[x^2 I(x^2 \leq a^2)\right] = (\sigma^2 + \rho^2)\left[\Phi\left(\frac{a-\rho}{\sigma}\right) - \Phi\left(\frac{-a-\rho}{\sigma}\right)\right] + \sigma(\rho - a)\phi(\frac{a+\rho}{\sigma}) - \sigma(a+\rho)\phi(\frac{a-\rho}{\sigma}).$$

Noting that $I(x^2 \leq a^2) = 1 - I(x^2 > a^2)$, then

$$\begin{aligned} E\left[x^2 I(x^2 > a^2)\right] &= (\sigma^2 + \rho^2) - (\sigma^2 + \rho^2)\left[\Phi\left(\frac{a-\rho}{\sigma}\right) - \Phi\left(\frac{-a-\rho}{\sigma}\right)\right] \\ &- \sigma(\rho - a)\phi(\frac{a+\rho}{\sigma}) + \sigma(a+\rho)\phi(\frac{a-\rho}{\sigma}) \end{aligned}$$

or for future use

$$\begin{aligned} E\left[x^2 I(x^2 > a^2)\right] &= (\sigma^2 + \rho^2)\left[1 - \Phi\left(\frac{a-\rho}{\sigma}\right) + \Phi\left(\frac{-a-\rho}{\sigma}\right)\right] \\ &+ \sigma(a-\rho)\phi(\frac{a+\rho}{\sigma}) + \sigma(a+\rho)\phi(\frac{a-\rho}{\sigma}). \end{aligned} \tag{144}$$

∎

Consider now the estimator of the pair-wise error correlations in the denominator of $J_{\alpha,2}$:

$$(N-1)\widehat{\overline{\rho^2}} = \frac{2}{N}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\hat{\rho}_{ij}^2 I\left(v\hat{\rho}_{ij}^2 > \theta_N\right),$$

and write it as

$$(N-1)\widehat{\overline{\rho^2}} = \frac{2}{N}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right),$$

where (using T for v)

$$a = \frac{1}{\sqrt{T}} \Phi^{-1} \left(1 - \frac{p}{2(N-1)} \right).$$

Also note that as a first-order approximation $\hat{\rho}_{ij} \sim N(\rho_{ij}, \frac{1-\rho_{ij}^2}{T})$. It is clear that all elements in the summation for $\widehat{\rho^2}$ lie in the range $[0, 1]$, and hence are uniformly integrable, and the limit of $\widehat{\rho^2}$ as N and $T \to \infty$ is given by

$$p \lim_{N,T\to\infty} (N-1)\widehat{\rho^2} = \lim_{N,T\to\infty} \frac{2}{N} \sum_{i=2}^{N} \sum_{j=1}^{i-1} E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right)\right].$$

But the expression for $E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right)\right]$ depends on the value of ρ_{ij}. Under the assumption that $\mathbf{R}$ is sparse, there are only a finite number of non-zero elements in each row (column) of $\mathbf{R}$. Denote the number of non-zero off-diagonal elements of the ith row of $\mathbf{R}$ by r_i and note that the above probability limit can be written as

$$\lim_{N,T\to\infty} \frac{2}{N} \sum_{i=2}^{N} \sum_{j=1}^{i-1} E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right)\right]$$

$$= \lim_{N,T\to\infty} \frac{2}{N} \sum_{for\ all\rho_{ij}=0} E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right) | \rho_{ij} = 0\right]$$

$$+ \lim_{N,T\to\infty} \frac{2}{N} \sum_{for\ all\rho_{ij}\neq 0} E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right) | \rho_{ij} \neq 0\right]$$

$$= \lim_{N,T\to\infty} (N - \bar{r}) E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 \geq a^2\right) | \rho_{ij} = 0\right]$$

$$+ \lim_{N,T\to\infty} \frac{2}{N} \sum_{for\ all\rho_{ij}\neq 0} E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right) | \rho_{ij} \neq 0\right]$$

where $\bar{r} = \sum_{i=1}^{N} r_i/N$, is finite. Using Lemma 18 with $\rho = 0$, we have

$$E\left[x^2 I(x^2 \geq a^2) | \rho = 0\right] = \sigma^2 \left[1 - \Phi\left(\frac{a}{\sigma}\right) + \Phi\left(\frac{-a}{\sigma}\right)\right] + 2\sigma a \phi(\frac{a}{\sigma})$$

$$= 2\sigma^2 \left[1 - \Phi\left(\frac{a}{\sigma}\right)\right] + 2\sigma a \phi(\frac{a}{\sigma})$$

Also, when $\rho = 0$, then $\sigma^2 = 1/T$, and $a/\sigma = \Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)$. Further

$$\Phi\left[\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\right] = 1 - \frac{p}{2(N-1)}.$$

Therefore,

$$E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right) | \rho_{ij} = 0\right]$$

$$= \frac{2}{T}\left[1 - \Phi\left(\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\right)\right]$$

$$+ 2\frac{1}{T}\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\phi\left[\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\right]$$

$$= \frac{1}{T}\frac{p}{(N-1)} + \frac{2}{T}\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\phi\left[\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\right]$$

and

$$\lim_{N,T\to\infty} (N - \bar{r}) E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right) | \rho_{ij} = 0\right]$$

$$= \lim_{N,T\to\infty} \frac{1}{T}\frac{p(N-\bar{r})}{(N-1)} +$$

$$\lim_{N,T\to\infty} \frac{2(N-\bar{r})}{T}\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\phi\left[\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\right].$$

The first term clearly tends to zero as N and $T \to \infty$, in any order. The limiting behavior of the second term is more complicated. But using numerical techniques, it can be shown that (for a fixed p)

$$X_N = (N - \bar{r})\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\phi\left[\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)\right] = O(\ln(N)),$$

42

and the second term also tends to zero if $\log(N)/T \to 0$. (For $p = 0.05, \bar{r} = 1$ and $N = 20,000$, it is easily seen that $X_N = 0.58$.)

Consider now the term

$$\frac{2}{N} \sum_{for\ \text{all}\rho_{ij} \neq 0} E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right) |\rho_{ij} \neq 0\right]$$

and recall that the number of non-zero, distinct, off-diagonal elements of $\mathbf{R}$ is $N\bar{r}/2$, and hence

$$\frac{2}{N} \sum_{for\ \text{all}\rho_{ij} \neq 0} E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right) |\rho_{ij} \neq 0\right] \to \text{finite limit.}$$

To obtain this limit, using Lemma 18 we note that

$$
\begin{aligned}
&E\left[\hat{\rho}_{ij}^2 I\left(\hat{\rho}_{ij}^2 > a^2\right) |\rho_{ij} \neq 0\right] \\
=\ & \left(\frac{1 - \rho_{ij}^2}{T} + \rho_{ij}^2\right)\left[1 - \Phi\left(\frac{\sqrt{T}a - \sqrt{T}\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}}\right) + \Phi\left(\frac{-\sqrt{T}a - \sqrt{T}\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}}\right)\right] \\
& + \frac{\sqrt{1 - \rho_{ij}^2}}{\sqrt{T}}\left(a - \rho_{ij}\right)\phi\left(\frac{\sqrt{T}a + \sqrt{T}\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}}\right) + \frac{\sqrt{1 - \rho_{ij}^2}}{\sqrt{T}}\left(a + \rho_{ij}\right)\phi\left(\frac{\sqrt{T}a - \sqrt{T}\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}}\right),
\end{aligned}
$$

where as before $a = \frac{1}{\sqrt{T}}\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)$. For each $\rho_{ij} \neq 0$, since $\sqrt{T}a = \Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right) = O(\ln(N))$, then

$$\left(\frac{1 - \rho_{ij}^2}{T} + \rho_{ij}^2\right)\left[1 - \Phi\left(\frac{\sqrt{T}a - \sqrt{T}\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}}\right) + \Phi\left(\frac{-\sqrt{T}a - \sqrt{T}\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}}\right)\right] \to \rho_{ij}^2 \text{ as } N \text{ and } T \to \infty,$$

such that $\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right)/\sqrt{T} \to 0$. (Note that for $p = 0.05$ and $N = 20,000$, then $\Phi^{-1}\left(1 - \frac{p}{2(N-1)}\right) = 4.71$.)

Similarly, as N and $T \to \infty$ we have

$$\frac{\sqrt{1 - \rho_{ij}^2}}{\sqrt{T}}\left(a - \rho_{ij}\right)\phi\left(\frac{\sqrt{T}a + \sqrt{T}\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}}\right) + \frac{\sqrt{1 - \rho_{ij}^2}}{\sqrt{T}}\left(a + \rho_{ij}\right)\phi\left(\frac{\sqrt{T}a - \sqrt{T}\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}}\right) \to 0.$$

Collecting all the above results, we have

$$p \lim_{N,T \to \infty} (N-1)\widehat{\overline{\rho^2}} = \lim_{N \to \infty} \frac{2}{N} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \rho_{ij}^2,$$

which is finite, given sparse nature of the error correlation matrix, $\mathbf{R}$.

Appendix H: Simulating Multivariate Non-normal Random Variates

The objective is to generate N random variables u_i, $i = 1, 2, ..., N$ such that (in population) $E(u_i) = 0$, $E(u_i^2) = \sigma_i^2$, $E(u_i^3) = m_{3i}$, $E(u_i^4) = m_{4i}$ and $E(u_i u_j) = \rho_{ij}, i \neq j$ for $i, j = 1, 2, ..., N$.

The problem of generating multivariate non-normal random variables have been addressed in the literature by Vale and Maurelli (1983) and further discussed by Harwell and Serlin (1989) and Headrick and Sawilowsky (1999). Following Fleishman (1978), Vale and Maurelli (1983 VM) propose generating u_i as,

$$u_i = a_i + b_i\varepsilon_i + c_i\varepsilon_i^2 + d_i\varepsilon_i^3, \ i = 1, 2, ..., N,$$

where $\varepsilon_i \sim IIDN(0,1)$ and $E(\varepsilon_i\varepsilon_j) = \rho_{\varepsilon,ij}$. The unknown parameters $a_i, b_i, c_i, d_i, \rho_{\varepsilon,ij}$ are obtained using the following relationships (see equations (2)-(5) in VM)

$$a_i + c_i = 0, \tag{145}$$

$$b_i^2 + 6b_id_i + 2c_i^2 + 15d_i^2 = \sigma_i^2, \tag{146}$$

$$2c_i(b_i^2 + 24b_id_i + 105d_i^2 + 2) = m_{3i}, \tag{147}$$

$$24[b_id_i + c_i^2(1 + b_i^2 + 28b_id_i) + d_i^2(12 + 48b_id_i + 141c_i^2 + 225d_i^2)] = m_{4i}, \tag{148}$$

for $i = 1, 2, ..., N$, and (see equation (11) in VM)

$$\rho_{ij} = \rho_{\varepsilon,ij}(b_ib_j + 3b_id_j + 3d_ib_j + 9d_id_j) + \rho_{\varepsilon,ij}^2(2c_ic_j) + \rho_{\varepsilon,ij}^3(6d_id_j), \tag{149}$$

43

for $i \neq j = 1, 2, ..., N$.

The VM procedure is shown to work reasonably well for non-extreme values of skewness and kurtosis and when N is small. But even if one follows VM's two step procedure where the equations (145)-(148) are solved first, the procedure still requires solving a large number of cubic equations, and hoping that the solution of (149) for $\rho_{\varepsilon,ij}$ lies in the admissible range of $[-1, 1]$. No proof is provided that such a solution exists.

In what follows we propose a new more compact algorithm for generation of non-normal correlated random variables as a generalization of the standard Cholesky factor approach used routinely to generate correlated normal random variables. Let $\mathbf{u} = (u_1, u_2, ..., u_N)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_N)'$, and write each u_i as a linear combination of $\boldsymbol{\varepsilon}$

$$u_i = \sum_{j=1}^{N} q_{ij}\varepsilon_j, \text{for } i = 1, 2, ..., N,$$

or in matrix notation

$$\mathbf{u} = \mathbf{Q}\boldsymbol{\varepsilon},$$

where q_{ij} is the (i, j) element of $\mathbf{Q}$.

Our approach is to generate ε_j, $j = 1, 2, ..., N$, as independent draws from non-normal distributions with $E(\varepsilon_j) = 0, E(\varepsilon_j^2) = 1, E(\varepsilon_j^3) = m_{\varepsilon,3j}$ and $E(\varepsilon_j^4) = m_{\varepsilon,4j}$. Note also that ρ_{ij} is determined by $\mathbf{Q}$ and is given by the (i, j) element of $\mathbf{QQ}'$ scaled by $\sigma_i\sigma_j$ where $\sigma_i^2 = \sum_{j=1}^{N} q_{ij}^2$. For given values of ρ_{ij} and σ_i^2, $\mathbf{Q}$ can be obtained as the Cholesky factor of $E(\mathbf{uu}') = \mathbf{V}$. In such a case $\mathbf{Q}$ can be a lower or an upper triangular matrix with strictly positive diagonal elements. It is assumed that $\mathbf{V}$ is non-singular, and as a result $\mathbf{Q}$ will also be non-singular.

Consider now the problem of generating $\varepsilon_j's$ such that $E(u_i^3) = m_{i3}$ and $E(u_i^4) = m_{i4}$. To this end note that

$$m_{2i} \quad = \quad \sigma_i^2 = E(u_i^2) = \sum_{j=1}^{N} q_{ij}^2, \text{for } i = 1, 2, ..., N,$$

$$m_{3i} \quad = \quad E(u_i^3) = E\left[\sum_j \sum_{j'} \sum_s \sum_{s'} q_{ij} q_{ij'} q_{is} \varepsilon_j \varepsilon_{j'} \varepsilon_s\right] = \sum_{j=1}^{N} q_{ij}^3 m_{\varepsilon,3j}, \text{for } i = 1, 2, ..., N,$$

and

$$m_{4i} = E(u_i^4) = E\left[\sum_j \sum_{j'} \sum_s \sum_{s'} q_{ij} q_{ij'} q_{is} q_{is'} \varepsilon_j \varepsilon_{j'} \varepsilon_s \varepsilon_{s'}\right].$$

But since $\varepsilon_j's$ are independent draws with mean 0 and a unit variance we have

$$E\left(\varepsilon_j \varepsilon_{j'} \varepsilon_s \varepsilon_{s'}\right) \quad = \quad m_{\varepsilon,4j}, \text{ if } j = j' = s = s'$$
$$= \quad 1, \text{ if } j = j' \text{ and } s = s' \text{ or if } j = s \text{ and } j' = s' \text{ or if } j = s' \text{ and } j' = s$$
$$= \quad 0 \text{ otherwise.}$$

Hence, it readily follows that

$$m_{4i} = \sum_{j=1}^{N} q_{ij}^4 m_{\varepsilon,4j} + 3\sum_{j \neq s} q_{ij}^2 q_{is}^2. \tag{150}$$

But

$$\sum_{j \neq s} q_{ij}^2 q_{is}^2 = \sum_{j=1}^{N} \sum_{s=1}^{N} q_{ij}^2 q_{is}^2 - \sum_{j=1}^{N} q_{ij}^4 = \left(\sum_{j=1}^{N} q_{ij}^2\right)^2 - \sum_{j=1}^{N} q_{ij}^4 = \sigma_i^4 - \sum_{j=1}^{N} q_{ij}^4$$

Therefore, (150) can be written as

$$m_{4i} - 3\sigma_i^4 = \sum_{j=1}^{N} q_{ij}^4 \left(m_{\varepsilon,4j} - 3\right).$$

Let $\kappa_{\varepsilon j} = m_{\varepsilon,4j} - 3$ and $\kappa_i = m_{4i} - 3\sigma_i^4$, and write the above relations in matrix notation, namely

$$\boldsymbol{\kappa}_u = \mathbf{Q}_{(4)}\boldsymbol{\kappa}_\varepsilon,$$

where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, ..., \kappa_N)'$, $\boldsymbol{\kappa}_\varepsilon = (\kappa_{\varepsilon 1}, \kappa_{\varepsilon 2}, ..., \kappa_{\varepsilon N})'$ and $\mathbf{Q}_{(4)} = \mathbf{Q} \odot \mathbf{Q} \odot \mathbf{Q} \odot \mathbf{Q}$, where $\odot$ is the Hadamard matrix operator (or element-wise operator). Similarly, for the third moments we have

$$\mathbf{m}_3 = \mathbf{Q}_{(3)}\mathbf{m}_{\varepsilon,3},$$

where $\mathbf{m}_3 = (m_{3,1}, m_{3,2}, ..., m_{3,N})$, and $\mathbf{m}_{\varepsilon,3} = (m_{\varepsilon,3,1}, m_{\varepsilon,3,2}, ..., m_{\varepsilon,3,N})$. Since $\mathbf{Q}$ is a triangular matrix with strictly positive diagonal elements it follows that $\mathbf{Q}_{(3)}$ and $\mathbf{Q}_{(4)}$ are also non-singular and hence invertible. Thus

$$\mathbf{m}_{\varepsilon,3} = \mathbf{Q}_{(3)}^{-1}\mathbf{m}_3 \tag{151}$$

$$\boldsymbol{\kappa}_\varepsilon = \mathbf{Q}_{(4)}^{-1}\boldsymbol{\kappa}_u. \tag{152}$$

Denoting $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, ..., \sigma_N^2)'$ we also have $\boldsymbol{\sigma}^2 = \mathbf{Q}_{(2)}\boldsymbol{\tau}_N$.

Having computed $m_{\varepsilon,3i}$ and $m_{\varepsilon,4i}$ we can now generate ε_i as

$$\varepsilon_i = a_i + b_i \upsilon_i + c_i \upsilon_i^2 + d_i \upsilon_i^3, \ \ i = 1, 2, ..., N, \tag{153}$$

where $\upsilon_i \sim IIDN(0,1)$ and the coefficients a_i, b_i, c_i and d_i are determined so that $E(\varepsilon_i) = 0, E(\varepsilon_i^2) = 1$, $E(\varepsilon_i^3) = m_{\varepsilon,3i}$ and $E(\varepsilon_i^4) = m_{\varepsilon,4i}$, using Fleishman's formula

$$a_i + c_i = 0, \tag{154}$$

$$b_i^2 + 6b_i d_i + 2c_i^2 + 15d_i^2 = 1, \tag{155}$$

$$2c_i(b_i^2 + 24b_i d_i + 105d_i^2 + 2) = m_{\varepsilon,3i}, \tag{156}$$

$$24[b_i d_i + c_i^2(1 + b_i^2 + 28b_i d_i) + d_i^2(12 + 48b_i d_i + 141c_i^2 + 225d_i^2)] = \kappa_{\varepsilon i}. \tag{157}$$

References

Abadir, K.M., Distaso, W., Zikes, F., 2011. Design-free estimation of large variance matrices, Imperial College London, mimeo.

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.

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

Bailey, N., G. Kapetanios, and M.H. Pesaran, 2012. Exponent of Cross sectional Dependence: Estimation and Inference, CESifo Working Paper Series No. 3722 (http://ssrn.com/abstract=1998062).

Bao, Y., A. Ullah, 2006. Expectation of quadratic forms in normal and nonnormal variables with econometric applications, working paper, University of California, Riverside.

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

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

Chudik, A., Pesaran, M.H., Tosetti, E., 2011. Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal* 14, 45-90.

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., 2012. High dimensional covariance matrix estimation in approximate factor model. *Annals of Statistics*, forthcoming.

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

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., 2011. Testing linear factor pricing models with large cross-sections: A distribution-free approach. Georgia State University, mimeo.

Harwell, M.R., Serlin, R.C., 1989. A nonparametric test statistic for the general linear model. *Journal of Educational Statistics* 4, 351-371.

Headrick, T.C., Sawilowsky, S.S., 1999. Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika* 64, 25-35.

Horn, R.A. Johnson, C.R. 1985. *Matrix Analysis*, Cambridge University Press, Cambridge.

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.

Keleijan, H. H., Prucha, I.R. 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.

Magnus, J.R., Neudecker, H., 1999. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Revised Edition, Jonh Wiley & Sons.

Roll, R., 1977. A critique of the asset pricing theory's tests' Part I: On past and potential testability of the theory. *Journal of Financial Economics* 4, 129–76.

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

Smith, M.D. 1988. Convergent series expressions for inverse moments of quadratic forms in normal variables. *Australian Journal of Statistics* 30, 235-246.

Ullah, A., 1974. On the sampling distribution of improved estimators for coefficients in linear regression. *Journal of Econometrics 2*, 143-150.

Ullah, A., 2004. *Finite Sample Econometrics*, Oxford University Press.

Vale, D. C., Maurelli, V. A., 1983. Simulating multivariate nonnormal distributions. *Psychometrika* 48, 465-471.

White, H., 2001. *Asymptotic Theory for Econometricians*. Revised Edition, Academic Press.

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

Table 1: Summary statistics for estimated β, variance, skewness and kurtosis measures of residuals from CAPM regressions estimated for all securities in the S&P 500 index with at least sixty months of return data using five-year estimation windows (sixty months) at the end of each month September 1989-September 2011

	Mean	Median	2.5% Quantile	97.5% Quantile
$\hat{\sigma}^2_{i,\tau}$	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

Notes: $\hat{\beta}_{i\tau}$ 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 at the estimation windows $\tau = 1989M9, 1989M10, ..., 2011M9$. $\hat{\sigma}^2_{i,\tau} = \hat{m}_{2,i\tau}$ $\hat{\gamma}_{1,i\tau} = \hat{m}_{3,i\tau}/\hat{m}^{3/2}_{2,i\tau}$ and $\hat{\gamma}_{2,i\tau} = \hat{m}_{4,i\tau}/\hat{m}^2_{2,i\tau} - 3$, which are computed using the FF residuals, where $\hat{m}_{s,i\tau} = (60)^{-1}\sum_{t=1}^{60}\left(\hat{u}_{i,\tau t} - \bar{\bar{u}}_{i,\tau}\right)^s$ and $\bar{\bar{u}}_{i,\tau} = (60)^{-1}\sum_{t=1}^{60}\hat{u}_{i,\tau t}$, 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 have $\gamma_{1,i} = \gamma_{2,i} = 0$.

Table 2: Size and power of SS, WS, GRS and J_α tests in the case of models with a single factor

Table 2, Panel A: With Single Factor, Normal Errors																
N^{δ_b}		No Cross Correlation				$\delta_b = 1/4$				$\delta_b = 1/2$				$\delta_b = 3/5$		
(T, N)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
Size: $\alpha_i = 0$ for all i																
SS	60	4.2	4.8	4.7	4.6	4.3	5.2	4.3	5.1	7.0	7.7	8.5	7.8	9.1	9.7	12.6	12.4
	100	4.5	5.3	5.1	4.8	4.5	4.7	5.3	5.1	7.4	7.9	8.3	7.7	10.5	10.0	11.5	12.2
WS	60	4.3	5.0	4.6	4.8	4.3	4.8	4.4	4.6	7.6	8.2	9.0	8.6	9.8	9.9	13.1	13.2
	100	4.4	5.0	4.8	4.6	3.8	5.3	5.2	5.1	7.9	8.1	8.1	7.8	10.4	11.4	12.9	13.4
GRS	60	5.3	N/A	N/A	N/A	4.3	N/A	N/A	N/A	4.4	N/A	N/A	N/A	4.5	N/A	N/A	N/A
	100	5.0	N/A	N/A	N/A	5.0	N/A	N/A	N/A	4.5	N/A	N/A	N/A	4.5	N/A	N/A	N/A
$\hat{J}_{\alpha,1}$	60	6.2	6.0	5.6	5.1	6.1	6.6	5.6	5.8	9.6	10.2	11.0	10.1	14.0	13.6	16.3	15.9
	100	6.1	6.0	5.9	5.3	6.1	6.5	6.6	5.6	10.1	10.4	11.3	10.4	13.3	12.7	15.5	16.0
$\hat{J}_{\alpha,2}$	60	6.1	5.9	5.4	5.0	6.0	6.1	5.3	5.6	6.4	6.9	6.5	5.5	6.8	6.3	6.2	7.2
	100	6.0	5.9	5.8	5.2	5.9	6.1	6.3	5.5	6.8	6.4	6.8	5.4	6.8	6.2	7.1	6.9
Power: $\alpha_i \sim IIDN(0,1)$ for $i = 1, 2, ..., N_\alpha$ with $N_\alpha = \lfloor N^{\delta_\alpha}\rfloor$, otherwise $\alpha_i = 0$.																
SS	60	21.8	25.9	36.3	48.9	20.8	26.2	34.9	47.9	22.2	25.5	35.2	48.9	21.1	28.2	35.4	45.7
	100	34.7	45.2	60.7	81.4	36.6	47.0	62.8	80.7	35.1	45.6	59.9	77.9	35.3	44.5	56.8	72.6
WS	60	23.0	32.0	43.7	60.2	23.4	32.3	43.0	59.2	25.4	30.8	40.4	58.2	25.5	32.4	41.3	52.1
	100	43.0	55.2	73.0	91.2	44.3	58.7	74.0	90.3	42.0	55.3	70.9	87.6	41.5	51.9	67.2	83.3
GRS	60	22.1	N/A	N/A	N/A	21.0	N/A	N/A	N/A	30.0	N/A	N/A	N/A	35.7	N/A	N/A	N/A
	100	77.2	N/A	N/A	N/A	80.1	N/A	N/A	N/A	90.1	N/A	N/A	N/A	93.3	N/A	N/A	N/A
$\hat{J}_{\alpha,1}$	60	77.6	89.8	97.8	99.9	78.6	89.7	97.7	99.8	73.8	87.2	96.9	99.9	73.3	86.9	95.3	99.5
	100	93.9	99.3	100.0	100.0	95.1	98.9	100.0	100.0	94.0	98.3	100.0	100.0	93.5	98.3	99.9	100.0
$\hat{J}_{\alpha,2}$	60	77.2	89.7	97.8	99.9	77.9	89.2	97.7	99.8	64.6	80.3	93.2	99.7	54.6	68.1	81.4	94.5
	100	93.8	99.3	100.0	100.0	94.9	98.8	100.0	100.0	88.9	96.7	99.7	100.0	83.0	94.3	98.7	100.0

	N^{δ_b}	No Cross Correlation				$\delta_b = 1/4$				$\delta_b = 1/2$				$\delta_b = 3/5$			
	(T, N)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
								Size: $\alpha_i = 0$ for all i									
SS	60	11.0	12.7	19.9	34.0	10.3	13.8	19.9	33.4	11.8	14.0	18.5	33.4	11.8	17.4	22.8	32.2
	100	15.1	22.8	36.9	64.6	16.3	23.7	35.2	63.3	15.5	21.3	33.8	57.2	18.4	24.5	32.6	49.9
WS	60	9.2	11.6	15.1	25.1	8.3	11.5	16.5	24.9	12.7	12.7	16.9	26.8	13.1	16.5	19.1	28.7
	100	12.2	17.4	27.4	49.0	14.0	18.3	27.1	51.6	16.0	18.6	28.2	44.1	17.2	20.8	28.3	39.0
GRS	60	5.3	N/A	N/A	N/A	5.5	N/A	N/A	N/A	5.3	N/A	N/A	N/A	5.0	N/A	N/A	N/A
	100	6.1	N/A	N/A	N/A	4.9	N/A	N/A	N/A	6.2	N/A	N/A	N/A	5.6	N/A	N/A	N/A
$\hat{J}_{\alpha,1}$	60	6.5	6.6	6.2	6.7	7.2	6.6	6.4	6.5	11.8	10.2	10.2	11.0	13.8	13.3	15.9	17.6
	100	6.3	5.6	6.4	6.8	6.1	7.0	6.8	6.9	10.8	10.8	11.5	10.1	13.4	14.3	15.1	16.4
$\hat{J}_{\alpha,2}$	60	6.5	6.6	6.1	6.6	6.9	6.2	6.2	6.3	7.9	6.6	6.1	6.6	7.6	6.5	7.1	7.5
	100	6.2	5.5	6.4	6.5	5.9	6.4	6.7	6.6	7.1	6.8	6.5	5.5	6.7	6.5	6.7	5.8
							Power: $\alpha_i \sim IIDN(0,1)$ for $i = 1, 2, ..., N_\alpha$ with $N_\alpha = \lfloor N^{\delta_\alpha} \rfloor$, otherwise $\alpha_i = 0$.										
SS	60	31.2	43.6	61.7	85.0	31.8	43.5	57.7	83.2	30.6	42.1	57.0	79.8	29.2	41.0	54.8	74.1
	100	53.1	74.2	89.2	99.3	55.9	73.6	90.6	99.2	51.5	67.1	88.0	98.8	50.6	64.7	81.8	97.5
WS	60	33.1	46.9	64.1	87.0	33.3	46.2	62.6	87.1	32.2	44.6	61.2	81.5	32.3	43.3	55.8	76.1
	100	58.5	78.0	92.0	99.7	59.1	77.2	92.6	99.6	55.4	70.5	90.7	99.3	52.5	68.3	84.6	98.0
GRS	60	23.1	N/A	N/A	N/A	21.2	N/A	N/A	N/A	30.6	N/A	N/A	N/A	34.4	N/A	N/A	N/A
	100	77.2	N/A	N/A	N/A	81.3	N/A	N/A	N/A	89.0	N/A	N/A	N/A	93.1	N/A	N/A	N/A
$\hat{J}_{\alpha,1}$	60	76.8	90.9	98.1	100.0	76.7	90.6	97.1	100.0	75.0	88.4	97.3	100.0	73.5	85.4	93.8	99.6
	100	93.6	98.9	100.0	100.0	94.1	99.1	100.0	100.0	93.7	98.2	99.9	100.0	90.9	97.9	99.7	100.0
$\hat{J}_{\alpha,2}$	60	76.7	90.8	98.1	100.0	75.7	90.0	96.9	100.0	66.0	81.1	93.6	99.6	56.5	69.0	80.6	95.1
	100	93.5	98.9	100.0	100.0	93.9	99.0	100.0	100.0	89.0	96.4	99.8	100.0	81.4	93.1	98.2	100.0

Notes: The data is generated as $y_{it} = \alpha_i + \beta_{1i}f_{1t} + u_{it}$, $i = 1, 2, .., N$; $t = 1, 2, ..., T$, $f_{1t} = \mu_{f1} + \rho_{f1}f_{1,t-1} + \sqrt{h_{1t}}\,\zeta_{1t}$, $h_{1t} = \mu_{h1} + \rho_{1h1}h_{1,t-1} + \rho_{2h1}\zeta_{1,t-1}^2$, $\zeta_{1t} \sim IIDN(0,1)$, $t = -49, ..., 0, 1, ..., T$ with $f_{1,-50} = h_{1,-50} = 0$, $\mu_{f1} = 0.53$, $\rho_{f1} = 0.06$, $\mu_{h1} = 0.89$, $\rho_{1h1} = 0.85$, $\rho_{2h1} = 0.11$. 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 $\mathbf{D} = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_N^2)'$ 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. We consider the following design of generating $\mathbf{R}$: (i) No cross-sectional correlation, $\mathbf{R} = \mathbf{I}_N$; (ii) Cross sectionally correlated errors, $\mathbf{R} = \mathbf{I}_N + \mathbf{bb}' - \check{\mathbf{B}}^2$, where $\mathbf{b} = (b_1, b_2, ..., b_N)'$, $\check{\mathbf{B}} = diag(\mathbf{b})$, we draw the first and the last N_b $(< N)$ elements of $\mathbf{b}$ as $Uniform(0.7, 0.9)$, and set the remaining middle elements to 0. We set $N_b = \lfloor N^{\delta_b} \rfloor$, where $\lfloor A \rfloor$ is the largest integer part of A. We examine $\delta_b = 0.25, 0.50$ and 0.60. For non-normal case, u_{it} are generated following steps 1-4 of the procedure in Subsection 5.1, 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 for details. SS and WS are the signed and singed rank tests of Gungor and Luger (2009), which are distributed as χ_N^2 and applicable for one-factor model. 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. $\hat{J}_{\alpha,1}$ is the propose large N test which is robust to non-Gaussian errors. $\hat{J}_{\alpha,2}$ is the modified version of the $\hat{J}_{\alpha,1}$ test which is more robust to cross-sectional correlations; Values of J_α 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.

Table 3: Size and power of GRS and J_α tests in the case of models with three factors

|	N^{δ_b}	No Cross Correlation				$\delta_b = 1/4$				$\delta_b = 1/2$				$\delta_b = 3/5$			
Panel A: With Three Factors, Normal Errors																	
	(T,N)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
Size: $\alpha_i = 0$ for all i																	
GRS	60	6.0	N/A	N/A	N/A	5.1	N/A	N/A	N/A	4.0	N/A	N/A	N/A	5.0	N/A	N/A	N/A
	100	5.0	N/A	N/A	N/A	5.5	N/A	N/A	N/A	4.8	N/A	N/A	N/A	4.3	N/A	N/A	N/A
$\hat{J}_{\alpha,1}$	60	6.6	6.3	5.5	5.9	6.7	6.6	5.6	7.1	10.1	11.1	11.6	11.5	13.1	15.3	15.6	15.1
	100	6.0	5.8	5.4	5.3	6.6	6.7	6.3	5.8	10.9	11.8	11.0	9.9	12.9	14.1	16.2	16.1
$\hat{J}_{\alpha,2}$	60	6.5	6.2	5.4	5.8	6.1	5.9	5.4	6.9	6.1	6.9	6.9	6.6	6.7	6.5	7.0	6.4
	100	5.9	5.8	5.2	5.3	6.3	6.2	6.0	5.8	6.9	7.0	6.8	5.7	6.4	6.8	6.7	6.6
Power: $\alpha_i \sim IIDN(0,1)$ for $i = 1,2,...,N_\alpha$ with $N_\alpha = \lfloor N^{\delta_\alpha} \rfloor$, otherwise $\alpha_i = 0$.																	
GRS	60	14.8	N/A	N/A	N/A	15.2	N/A	N/A	N/A	20.3	N/A	N/A	N/A	24.3	N/A	N/A	N/A
	100	68.2	N/A	N/A	N/A	68.9	N/A	N/A	N/A	84.0	N/A	N/A	N/A	88.6	N/A	N/A	N/A
$\hat{J}_{\alpha,1}$	60	66.3	79.3	92.2	99.4	65.9	81.0	92.8	99.1	64.7	77.4	91.7	98.7	62.4	75.9	86.8	97.0
	100	88.8	96.7	99.9	100.0	88.8	96.9	99.7	100.0	87.6	96.5	99.6	100.0	85.3	96.1	99.3	100.0
$\hat{J}_{\alpha,2}$	60	66.0	79.2	92.0	99.3	65.2	80.2	92.5	99.0	53.9	66.3	85.4	97.4	45.0	54.8	65.9	83.8
	100	88.8	96.7	99.9	100.0	88.3	96.8	99.7	100.0	81.7	93.8	99.0	99.9	71.8	86.9	95.0	99.4
Panel B: With Three Factors, Non-normal Errors																	
	N^{δ_b}	No Cross Correlation				$\delta_b = 1/4$				$\delta_b = 1/2$				$\delta_b = 3/5$			
	(T,N)	50	100	200	500	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	4.8	N/A	N/A	N/A	5.2	N/A	N/A	N/A	5.5	N/A	N/A	N/A
	100	5.1	N/A	N/A	N/A	5.5	N/A	N/A	N/A	5.3	N/A	N/A	N/A	4.5	N/A	N/A	N/A
$\hat{J}_{\alpha,1}$	60	6.0	6.8	6.4	5.8	6.0	7.3	5.8	6.5	10.6	10.7	9.8	12.2	12.6	13.8	15.4	15.8
	100	6.5	6.8	5.9	5.4	7.3	6.9	5.8	6.5	9.9	9.9	11.5	10.7	12.5	14.6	16.9	15.6
$\hat{J}_{\alpha,2}$	60	5.8	6.8	6.3	5.7	5.6	6.7	5.3	6.1	7.0	6.0	6.5	7.0	6.4	7.3	7.1	7.2
	100	6.4	6.6	5.8	5.2	7.1	6.5	5.4	6.3	6.2	6.1	6.9	5.6	6.2	7.2	6.9	6.2
Power: $\alpha_i \sim IIDN(0,1)$ for $i = 1,2,...,N_\alpha$ with $N_\alpha = \lfloor N^{\delta_\alpha} \rfloor$, otherwise $\alpha_i = 0$.																	
GRS	60	14.5	N/A	N/A	N/A	14.5	N/A	N/A	N/A	20.3	N/A	N/A	N/A	23.9	N/A	N/A	N/A
	100	67.4	N/A	N/A	N/A	70.5	N/A	N/A	N/A	83.4	N/A	N/A	N/A	87.1	N/A	N/A	N/A
$\hat{J}_{\alpha,1}$	60	68.0	82.8	93.2	99.2	66.1	83.8	92.6	99.2	66.1	78.1	91.8	98.9	61.2	77.4	85.9	95.8
	100	90.0	97.4	99.8	100.0	88.4	97.0	99.8	100.0	87.0	96.9	99.6	100.0	85.5	95.4	99.2	100.0
$\hat{J}_{\alpha,2}$	60	67.8	82.5	93.2	99.2	65.3	82.9	92.5	99.2	53.9	67.3	84.6	96.8	43.0	56.4	64.4	81.9
	100	90.0	97.4	99.8	100.0	87.9	96.9	99.8	100.0	81.9	93.0	99.0	100.0	72.1	86.7	94.6	99.3

Notes: The data is generated as $y_{it} = \alpha_i + \sum_{\ell=1}^{3} \beta_{\ell i} f_{\ell t} + u_{it}$, $i = 1,2,..,N$; $t = 1,2,...,T$, $f_{\ell t} = \mu_{f\ell} + \rho_{f\ell} f_{\ell,t-1} + \sqrt{h_{\ell t}}\, \zeta_{\ell t}$, $h_{\ell t} = \mu_{h\ell} + \rho_{1h\ell} h_{\ell,t-1} + \rho_{2h\ell} \zeta_{\ell,t-1}^2$, $\zeta_{\ell t} \sim IIDN(0,1)$, $t = -49,...,T$ 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. The SS and WS tests are not included since they are not applicable when $m > 1$. See also notes to Table 2.

Table 4: Summary statistics for p-values of $\hat{J}_{a,2}$ 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 1989M09-2011M09

|	N_τ	p-value of $\hat{J}_{\alpha,2}$	Average skewness & excess kurtosis measures		Rejection frequency for normality tests at τ			Average pair-wise correlations		
			$\overline{\hat{\gamma}}_{1\tau}$	$\overline{\hat{\gamma}}_{2\tau}$	$\gamma_{1,i\tau}=0$	$\gamma_{2,i\tau}=0$	$\gamma_{1,i\tau}=0$ $\gamma_{2,i\tau}=0$	$\overline{\hat{\rho}_\tau}$	$\overline{\hat{\rho}_\tau^2}$	$\widehat{\overline{\rho_\tau^2}}$
			PANEL A: CAPM regressions							
Mean	476	0.59	0.18	1.12	0.24	0.28	0.31	0.03	0.03	0.01
Median	477	0.74	0.17	1.11	0.23	0.27	0.30	0.01	0.03	0.00
Min	462	0.00	-0.01	0.38	0.13	0.12	0.15	0.01	0.02	0.00
Max	487	1.00	0.35	2.06	0.34	0.46	0.47	0.09	0.05	0.02
stand. dev.	6.3	0.40	0.08	0.41	0.05	0.09	0.08	0.03	0.01	0.00
			PANEL B: Fama-French regressions							
Mean	476	0.52	0.18	1.00	0.22	0.25	0.28	0.01	0.03	0.00
Median	477	0.58	0.19	0.97	0.22	0.24	0.28	0.01	0.03	0.00
Min	462	0.00	0.02	0.38	0.12	0.11	0.14	0.00	0.02	0.00
Max	487	0.99	0.33	1.70	0.30	0.39	0.42	0.03	0.03	0.01
stand. dev.	6.3	0.35	0.08	0.33	0.05	0.07	0.07	0.01	0.00	0.00

Notes: Reported summary statistics in PANEL A are of the values which are computed using CAPM regression residuals, $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{u}_{i,\tau t}$, for $t = 1, 2, ..., 60$, and $i = 1, 2, ..., N_\tau$, and the month ends $\tau = 1989M09, 1989M10, ..., 2011M09$. $\overline{\hat{\gamma}}_{\ell\tau} = N_\tau^{-1} \sum_{i=1}^{N_\tau} \hat{\gamma}_{\ell,i\tau}$ 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. $\overline{\hat{\rho}_\tau} = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{\tau,ij}$, $\overline{\hat{\rho}_\tau^2} = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{\tau,ij}^2$ with $\hat{\rho}_{\tau,ij} = \hat{\mathbf{u}}_{i.\tau}'\hat{\mathbf{u}}_{j.\tau}/(\hat{\mathbf{u}}_{i.\tau}'\hat{\mathbf{u}}_{i.\tau})^{1/2}(\hat{\mathbf{u}}_{j.\tau}'\hat{\mathbf{u}}_{j.\tau})^{1/2}$, $\hat{\mathbf{u}}_{i.\tau} = \left(\hat{u}_{i,\tau 1}, \hat{u}_{i,\tau 2}..., \hat{u}_{i,\tau T}\right)'$, and $\widehat{\overline{\rho_\tau^2}} = \frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\hat{\rho}_{\tau,ij}^2 I\left(\hat{\rho}_{\tau,ij}^2 > \frac{\theta_N}{v}\right)$, θ_{N_τ} is chosen such that $\Pr\left(\chi_1^2 > \theta_{N_\tau}\right) = 0.1/(N_\tau - 1)$, $v = T - m - 1$. Reported summary statistics in PANEL B are of the values which are computed using FF regression residuals, $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{\beta}_{2,i\tau}SMB_{t\tau} + \hat{\beta}_{3i}HML_{t\tau} + \hat{u}_{i,\tau t}$, for $t = 1, 2, ..., 60$, and $i = 1, 2, ..., N_\tau$, and the month ends $\tau = 1989M09, 1989M10, ..., 2011M09$.

Figure 1: Plots of the evolution of p-values of $\hat{J}_{\alpha,2}$ 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 1989M09-2011M09
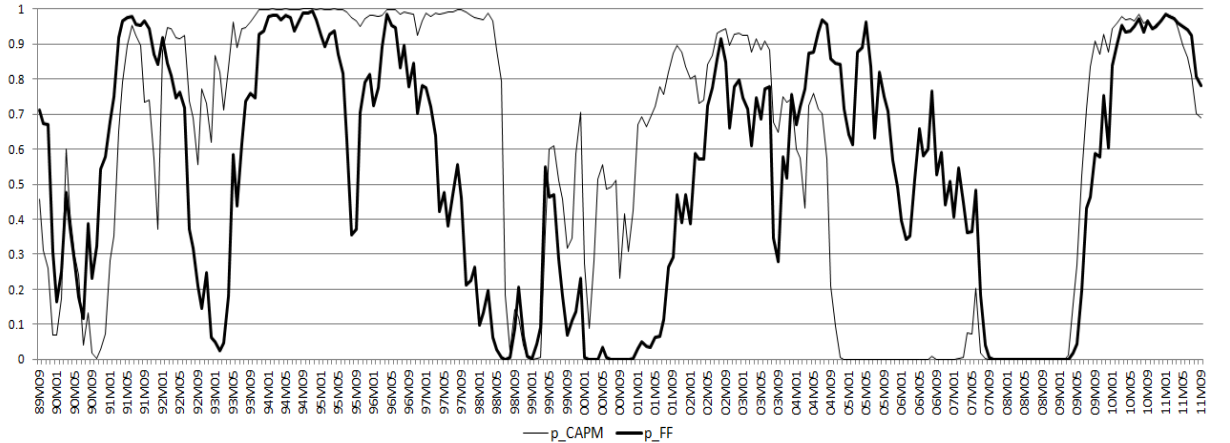

Notes: Reported plots are the p-values of $\hat{J}_{\alpha,2}$ test, which are computed using CAPM regressions, $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{u}_{i,\tau t}$ and FF three factor regressions, $r_{i,\tau t} - r_{f,\tau t} = \hat{\alpha}_{i\tau} + \hat{\beta}_{1,i\tau}\left(r_{m,\tau t} - r_{f,\tau t}\right) + \hat{\beta}_{2,i\tau}SMB_{t\tau} + \hat{\beta}_{3i}HML_{t\tau} + \hat{u}_{i,\tau t}$, for $t = 1, 2, ..., 60$, and $i = 1, 2, ..., N_\tau$, of the month ends estimation windows $\tau = 1989M09, 1989M10, ..., 2011M09$.

Figure 2: Monthly rate of returns of Dow Jones Credit Suisse Core Long/Short Equity Hedge Fund Index relative to S&P500 returns, and p-values of $\hat{J}_{\alpha,2}$ test based on CAPM regressions over the period December 2006 to September 2011
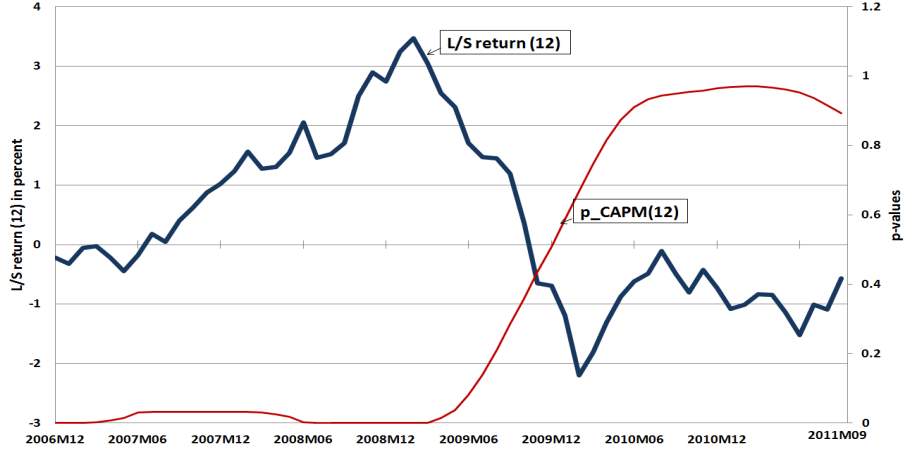
The long/short return variable, $\tilde{r}_{ht}(12)$, is computed as $\tilde{r}_{ht}(12) = \frac{1}{12}\sum_{j=0}^{11}\tilde{r}_{h,t-j}$, where $\tilde{r}_{ht} = r_{ht} - r_t$, r_{ht} is the return on Dow Jones Credit Suisse Core Long/Short Equity Hedge Fund Index, and r_t is the return on $S\&P$ 500 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 $\hat{J}_{\alpha,2}$ test at the end of month τ, computed using CAPM regressions estimated on samples of 60 months. See the notes to Table 4 for details of CAPM regressions.

Figure 3: Monthly rate of return of Dow Jones Credit Suisse Core Long/Short Equity Hedge Fund Index relative to S&P500 return, and p-values of $\hat{J}_{\alpha,2}$ test based on Fama-French regressions over the period December 2006 to September 2011
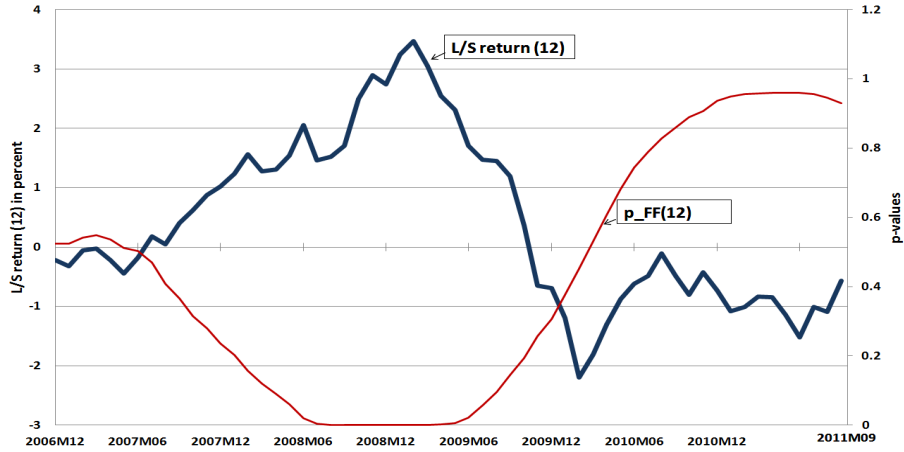
See the notes to Figure 2. See the notes to Table 4 for details of Fama-French regressions.