cemmap

centre for microdata methods and practice

An investigation into multivariate variance ratio statistics and their application to stock market predictability

Seok Young Hong
Oliver Linton
Hui Jun Zhang

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP13/15

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

An investigation into Multivariate Variance Ratio Statistics and their application to Stock Market Predictability*

Seok Young Hong[†], Oliver Linton[‡], and Hui Jun Zhang[§]

University of Cambridge

March 23, 2015

Abstract

We propose several multivariate variance ratio statistics. We derive the asymptotic distribution of the statistics and scalar functions thereof under the null hypothesis that returns are unpredictable after a constant mean adjustment (i.e., under the weak form Efficient Market Hypothesis). We do not impose the no leverage assumption of Lo and MacKinlay (1988) but our asymptotic standard errors are relatively simple and in particular do not require the selection of a bandwidth parameter. We extend the framework to allow for a time varying risk premium through common systematic factors. We show the limiting behaviour of the statistic under a multivariate fads model and under a moderately explosive bubble process: these alternative hypotheses give opposite predictions with regards to the long run value of the statistics. We apply the methodology to five weekly size-sorted CRSP portfolio returns from 1962 to 2013 in three subperiods. We find evidence of a reduction of linear predictability in the most recent

*We thank Greg Connor, Richard Davis, Tassos Magdalinos, Alexei Onatski, Andrew Patton, Hashem Pesaran, Peter Phillips, Mark Salmon, Allan Timmerman, and Wei Biao Wu for helpful comments and: CEMMAP, the Cambridge INET, and the Keynes fund for financial support. MATLAB code for calculating the statistics in this paper are available at http://www.oliverlinton.me.uk/research/software

[†]Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, United Kingdom, e-mail: syh30@cam.ac.uk

[‡]Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom, e-mail: obl20@cam.ac.uk

[§]Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom, e-mail: hjz20@cam.ac.uk

period, for small and medium cap stocks. The main findings are not substantially affected by allowing for a common factor time varying risk premium.

KEYWORDS: Bubbles; Fads; Martingale; Momentum; Predictability

JEL CLASSIFICATION: C10; C32; G10; G12

1 Introduction

Variance ratio tests (Lo and MacKinlay (1988) and Poterba and Summers (1988)) are widely used in empirical finance as a way of testing the weak form Efficient Markets Hypothesis (EMH) and to measure the degree and (cumulative) direction of departures from this hypothesis in financial time series. Indeed, this work has been extremely influential in understanding predictability in asset prices and in measuring market quality. A lot of empirical work followed immediately after the seminal contributions. Lo and MacKinlay (1988) presented evidence regarding predictability of the US stock market. They concluded that the Random Walk Hypothesis was soundly rejected by weekly US stock market returns. The graduate textbook Campbell, Lo, and MacKinlay (1997), henceforth CLM, presents variance ratios for weekly value weighted and equal weighted CRSP indexes and five size sorted portfolios over the period 1962-1994; they argue that the EMH is strongly rejected, although they find that the magnitude of the violation is less in the later subperiod 1978-1994. On the other hand, Cochrane (2001) writing only two years later argues that: "daily, weekly, and monthly stock returns are close to unpredictable". He emphasized the more recent work that had shown that low frequency returns are predictable from dividend price ratio and term premium variables. Regarding "medium frequency" settings, i.e., daily or weekly, most recent research has focussed on other markets, specifically: to major exchange rates, Liu and He (1991) and Luger (2003), to emerging market stock indexes, Chaudhuri and Wu (2003), and commodity markets, Peterson, Ma, and Ritchey (1992), and to carbon trading markets Montagnoli and de Vries (2010). Another recent direction for this methodology is in "high frequency" settings, i.e., intraday, where it has informed the debate on the evolution of market quality in the US. Castura, Litzenberger, Gorelick, and Dwivedi (2010) investigate trends in market efficiency in Russell 1000/2000 stocks over the period 1 January 2006 to 31 December 2009. Based on evidence from intraday variance ratios (they look at 10:1 second variance ratios as well as 60:10 and 600:60 second ratios) they argue that markets have become more efficient at the high frequency over time. Chordia, Roll, and Subrahmanian (2011) compared intraday variance ratios over the period 1993-2000 with the period 2000-2008 and found that the hourly to daily variance ratios of NYSE listed stocks came closer to the EMH predicted values on average in the second period.[1] Another use of these measures involves cross sectional or panel data regressions with variance ratios as dependent variables, see for example O'Hara and Ye (2009). In short, variance ratios are the *de facto* measure of predictability/market efficiency that is adopted universally by financial empiricists.

There have been some criticisms of the univariate variance ratio methodology as a test of uncorrelatedness. Specifically, it is not consistent against all (fixed of given order) alternatives unlike the Box-Pierce statistics. It is a linear functional of the autocorrelation function and so provides no new information relative to that. It seems like a redundant test. Faust (1992) argues that actually they form a class of tests optimal against certain alternatives. Specifically, he considers a more general class of univariate Filtered Variance Ratio tests. Let $r_t^{\phi} = \sum_{i=0}^{m} \phi_i r_{t-i}$ be a filtered return series for filter ϕ. He shows that each such test based on comparing $\mathrm{var}(r_t^{\phi})/\mathrm{var}(r_t)$ can be given a likelihood ratio interpretation and so is optimal against a certain alternative that is of the mean reverting type. The advantage of the variance ratio over the Box-Pierce statistic is that it gives some sense of the direction of predictability, which is lost in the Box-Pierce or other portmanteau tests. Hillman and Salmon (2007) have argued that the variance ratio (actually the related variogram) is better suited to irregularly spaced data and some kinds of nonstationarity than correlogram tests. There is a lot of work on improving the finite sample performance of both Box-Pierce statistics and variance ratio statistics, see for example Kim, Nelson, and Startz (1991) and Kan and Wang (2010). See Charles and Darné (2009) for a recent review of this methodology and its application.

We make several contributions. First, we develop a multivariate methodology. Many tests of the efficient markets hypothesis have been carried out using the univariate variance ratio approach, that is, conducted one asset at a time. This paper proposes a methodology for multivariate variance ratio tests. The rationale for the test is roughly the following. Suppose that the RW hypothesis is not rejected for asset i based on univariate variance ratio tests. Suppose however that returns on i are predicted by lags of some other variable. A univariate test could fail to detect this violation of the EMH, although a multivariate test could detect it. This generic argument about the efficacy of multivariate methods versus univariate is widely accepted. There is a lot of work on multivariate portmanteau statistics, i.e., generalizations of the Box-Pierce statistic to multivariate time series, see for example Chitturi (1974) and Hosking (1981). The variance ratio statistics convey directional information about cross-autocorrelations beyond that contained in the portmanteau statistics, that is, in the case of a violation of the hypothesis they give some sense of the direction of departure.

[1] See also Sheppard (2013) for some theoretical results using a continuous time framework.

The univariate variance ratios describe the behaviour of the asset variances, whereas the multivariate statistics also measure the behaviour of the cross correlations and their cumulative direction. This could be important for momentum based trading strategies, for example. It is also useful for judging the direction of price discovery.

Second, we propose an alternative distribution theory and standard errors (*heteroskedasticity and leverage consistent*) than are usually adopted. The limiting distribution established in Lo and MacKinlay (1988, Theorem 3) and repeated in CLM (and so used in most empirical studies) for the univariate variance ratio statistics is incorrect under their stated assumptions H1-H4 (i.e., RW3). The correct distribution would be much more complicated and would depend on a long run variance that may be hard to estimate well. Either one makes additional assumptions to ensure that the variance is as claimed, which is what we propose below, or one has to use more complicated inference methods based on long run variance estimation, Newey and West (1987), or self normalization, Lobato (2001). In fact, the omitted condition appears quite innocuous, so their essential approach seems correct. However, we think that the no-leverage assumption (Lo and MacKinlay's H4) is untenable, empirically. Although this latter condition is satisfied by GARCH volatility processes with symmetrically distributed innovations, it is not satisfied by volatility processes that allow for leverage effects such as the GJR GARCH process or the Nelson's EGARCH process, and it is not even satisfied by standard GARCH volatility processes where the innovation is asymmetric. The value of the restriction is that it simplifies the standard error calculation, although, as we show, the standard errors that allow for violations of this condition do not entail an inordinate increase in computation or complexity. Essentially, Lo and MacKinlay (1988) imposed an unnecessary assumption but fail to impose a necessary one. We propose modified assumptions that still preserve the possibility of simple inference methods but allow for leverage effects. Specifically, we establish the asymptotic distribution of our statistics under two sets of assumptions: (a) a stationary martingale difference hypothesis with fourth unconditional moments; (b) uncorrelatedness as in Lo and MacKinlay (1988) and with an additional uncorrelatedness condition on the products of returns but without the additional no-leverage condition. The asymptotic variance is different from that contained in Theorem 3 of Lo and MacKinlay (1988) (and used in much subsequent empirical work). Furthermore, extending the univariate framework of Chen and Deo (2006) we also derive the limiting distribution under the increasing horizon framework, and show that asymptotic normality can be obtained with a slower rate of convergence. We propose a simple analogue method for conducting inference that does not require the selection of a bandwidth parameter. We note that much of the evidence about predictability has been based on the Lo and MacKinlay (1988) standard errors, which we argue should be replaced by standard errors that rely on weaker more plausible assumptions. We show that in practice the standard errors can make a difference, especially when the time series is short (such as when stationarity is of concern).

Third, we also establish the asymptotic properties of our statistic under several plausible alternative models including a multivariate Muth (1960) fads model and the recently developed bubble process of Phillips and Yu (2010). These alternatives yield quite different predictions regarding the long run value of the variance ratio statistics.

Fourth, we apply our methods to five CRSP weekly size-sorted portfolio returns from 1962-2013 and the three subperiods 1962-1978, 1978-1994 and 1994-2013; the first two subperiods correspond to the data used in CLM. We show that the degree of inefficiency has reduced over the most recent period, and in some cases this improvement is statistically significant. Specifically, the univariate tests do not reject the null hypothesis for medium or large stocks in the most recent period. However, the multivariate tests do reject, albeit with a lower significance level. We also show that the degree of asymmetry in the dependence structure has reduced, although it is still significant. We extend our analysis to allow for a time varying risk premium, but find that the main empirical results are sustained. We further investigate the variance ratios at the long horizon. Simulation experiments indicate that our variance ratio tests are reliable, powerful against several alternatives.

In section 2 we introduce the multivariate ratio population statistics in various forms. In section 3 we introduce the estimators, while in section 4 we present the main central limit theorem and inference methods. In section 5 we consider a number of alternative hypotheses, while in section 6 we extend the analysis to allow for a time varying risk premium. In section 7 we briefly discuss the large dimensional case. In section 8 we present our application, while Section 9 concludes. The appendix contains the proofs of all results and a small simulation experiment.

2 Multivariate Variance Ratios

For expositional purposes we shall suppose in this section that we have a vector stationary ergodic discrete time series $X_t \in \mathbb{R}^d$; formal assumptions regarding the data are given below in section 3. Let $\widetilde{X}_t = X_t - \mu$, where $\mu = EX_t$ for all t. We are interested in testing the (weak form) Efficient Markets Hypothesis and quantifying departures from this hypothesis. This refers to whether past prices can be used to predict future prices (beyond some risk adjustment, which initially we assume to be constant and be denoted by μ). "Prices" are usually taken to mean just a sequence of past prices for the asset in question, but the spirit of this hypothesis should allow the past history of other assets not to matter either.

It seems natural in this context to assume that the risk adjusted return process satisfies

$$E(\widetilde{X}_t|\mathcal{F}_{t-1}) = 0, \tag{1}$$

where $\mathcal{F}_t$ denotes the past history of the prices of all the assets. This is a stronger assumption than that returns are uncorrelated with the past of all prices, i.e.,

$$E(\widetilde{X}_{it}\widetilde{X}_{jt-k}) = 0 \tag{2}$$

for all $i, j = 1, \ldots, d$ and for all $k \neq 0$, which itself is a stronger assumption[2] than that returns are uncorrelated with their own past, i.e.,

$$E(\widetilde{X}_{it}\widetilde{X}_{it-k}) = 0 \tag{3}$$

for all i and for all $k \neq 0$, which is what is adopted in Lo and MacKinlay (1988) (and referred to as RW3 in Campbell, Lo, and MacKinlay (1997) and in much subsequent work). RW3 has the advantage that if one rejects it, then one rejects the martingale hypothesis; on the other hand, if one does not reject RW3 then one can't conclude that the martingale hypothesis is valid.[3] Throughout we work with at least the multivariate uncorrelatedness hypothesis (2). We also develop a theory based on the stronger martingale difference assumption, because the additional regularity conditions can be stated very simply.

We next define the population versions of the multivariate variance ratios. Let $X_t(K) = X_t + X_{t-1} + \ldots + X_{t-K+1}$ for each K, and define the following population quantities:

$$\Sigma = \mathrm{var}(X_t) = E(\widetilde{X}_t\widetilde{X}_t^\intercal) \tag{4}$$

$$D = \mathrm{diag}\left\{E(\widetilde{X}_{1t}^2), \ldots, E(\widetilde{X}_{dt}^2)\right\} \tag{5}$$

$$\Sigma(K) = \mathrm{var}(X_t(K)) = E((X_t - KE(X_t))\,(X_t - KE(X_t))^\intercal) \tag{6}$$

$$\Gamma(j) = \mathrm{cov}(X_t, X_{t-j}) = E(\widetilde{X}_t\widetilde{X}_{t-j}^\intercal) \tag{7}$$

[2] This is not quite correct, since the martingale hypothesis only requires $E|X_t| < \infty$, whereas the uncorrelatedness hypothesis requires $EX_t^2 < \infty$ in order to be formulated.

[3] We note that there are many tests of the martingale hypothesis that make use of more information, Hong and Lee (2005) and Escanciano and Velasco (2006), and thereby obtain power against a larger class of alternatives.

$$R(j) = \Sigma^{-1/2}\Gamma(j)\Sigma^{-1/2} \tag{8}$$

$$R_L(j) = \Gamma(j)\Sigma^{-1} \quad ; \quad R_R(j) = \Sigma^{-1}\Gamma(j) \tag{9}$$

$$Rd(j) = D^{-1/2}\Gamma(j)D^{-1/2} \tag{10}$$

for $j = 0, \pm 1, \ldots$. Here, $A^{1/2}$ denotes a symmetric square root of a symmetric matrix A. We shall assume that Σ is strictly positive definite. Note that $Rd(j)$ is the usual definition of the cross-(auto)correlation matrix, while $R(j)$ is a multivariate correlation matrix.[4]

2.1 Two Sided Variance Ratios

Under condition (2), the variance covariance matrices obey the scaling law $\mathrm{var}(X_t(K)) = K\mathrm{var}(X_t)$, where K is some positive integer, from which we may obtain a number of different variance ratio statistics.

We define the two sided matrix normalized multivariate ratio (population) statistic as

$$\mathcal{VR}(K) = \mathrm{var}(X_t)^{-1/2}\mathrm{var}(X_t(K))\mathrm{var}(X_t)^{-1/2}/K. \tag{11}$$

Clearly, under the null hypothesis (2) we should have $\mathcal{VR}(K) = I_d$. Under the generic (stationary) alternative hypothesis we have

$$\mathcal{VR}(K) = I + \sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)(R(j) + R(j)^{\intercal}), \tag{12}$$

which is a symmetric matrix. The off-diagonal elements should be zero under the null hypothesis of no predictability. Both representations (11) and (12) can be used as the basis for estimation.

An alternative multivariate normalization is given by

$$\mathcal{VR}a(K) = \mathrm{var}(X_t(K))\mathrm{var}(X_t)^{-1}/K,$$

which can likewise generically be written

$$\mathcal{VR}a(K) = I + \sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)\left(R_L(j) + R_R(j)^{\intercal}\right). \tag{13}$$

[4]All three measures are invariant to common univariate affine transformations $X_{ti} \mapsto \alpha + \beta X_{ti}$ for any α, β; the quantity $\Gamma(j)$ is invariant under multivariate location and scale transformation, meaning $X_t \mapsto \Sigma^{-1/2}(X_t - \mu)$, while $\Gamma d(j)$ is invariant under univariate location and scale transformation $X_t \mapsto D^{-1/2}(X_t - \mu)$. The cross-autocorrelation matrix is invariant to marginalization (looking at submatrices), whereas $\Gamma(j)$, $\Gamma_L(j)$, and $\Gamma_R(j)$ are not.

This has a regression interpretation, see Chitturi (1974) and Wang (2003, p62). Note that $\mathcal{VR}(K) = I$ if and only if $\mathcal{VR}a(K) = I$.

A third quantity is the diagonally normalized variance ratio

$$\mathcal{VR}d(K) = D^{-1/2}\mathrm{var}(X_t(K))D^{-1/2}/K \tag{14}$$

$$= Rd(0) + \sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)(Rd(j) + Rd(j)^\intercal), \tag{15}$$

where $Rd(0) = D^{-1/2}\Gamma(0)D^{-1/2}$ is the $d \times d$ contemporaneous correlation matrix. Under the null hypothesis that the series is uncorrelated, we should have $\mathcal{VR}d(K) = Rd(0)$ the contemporaneous correlation matrix, whose off-diagonal elements are unrestricted by the null hypothesis. The diagonal elements of $\mathcal{VR}d(K)$ correspond to the univariate variance ratio statistics, while the off-diagonal elements provide information about the cumulative cross-dynamics between the assets. Note that if $\mathcal{VR}(K) = I$, then $\mathcal{VR}d(K)_{ii} = 1$ for all i, but not vice versa. This suggests that if one rejects a univariate test then one would reject the multivariate test but not necessarily vice versa. Specifically, suppose that X_t are iid but $X_{1t} = X_{2,t-1}$ then the univariate tests would fail but the multivariate one would not.

We also consider the two parameter family of variance ratio statistics

$$\mathcal{VR}(K, L) = \mathcal{VR}(L)^{-1/2}\mathcal{VR}(K)\mathcal{VR}(L)^{-1/2} \tag{16}$$

for some positive distinct integers K and L.[5] An alternative definition (that does not require computation of $\mathrm{var}(X_t)$) is

$$\mathcal{VR}^{\%}(K, L) = \mathrm{var}(X_t(L))^{-1/2}\mathrm{var}(X_t(K))\mathrm{var}(X_t(L))^{-1/2} \times L/K.$$

Under the null hypothesis (2), we have $\mathcal{VR}(K, L), \mathcal{VR}^{\%}(K, L) = I_d$ for all K, L. Likewise we can define two parameter versions the other statistics:

$$\mathcal{VR}a(K, L) = \frac{L}{K}\mathrm{var}(X_t(K))\mathrm{var}(X_t(L))^{-1} = \mathcal{VR}a(K) \times \mathcal{VR}a(L)^{-1},$$

which satisfies $\mathcal{VR}a(K, L) = I_d$ under the null hypothesis, and

$$\mathcal{VR}d(K, L) = \frac{L}{K}D_L^{-1/2}\mathrm{var}(X_t(K))D_L^{-1/2} = D_{\mathcal{VR}d(L)}^{-1/2}\mathcal{VR}d(K)D_{\mathcal{VR}d(L)}^{-1/2},$$

where D_L is the diagonal matrix of variance of sum of L period returns and $D_{\mathcal{VR}d(L)}$ is the diagonal matrix of $\mathcal{VR}d(L)$. Under the null hypothesis, we should have $\mathcal{VR}d(K, L) = Rd(0)$.

[5] Poterba and Summers (1988) considered this for the univariate case with monthly data and chose $L = 12$ throughout while $K = 1, 24, \ldots, 96$.

2.2 One Sided Variance Ratios

In the univariate case, the variance ratio process and the autocorrelation function contain the same information and one can recover the autocorrelation function from the variance ratio function. This is not so in the multivariate case because $\mathcal{VR}(K)$ and $\mathcal{VR}d(K)$ are both symmetric matrices whereas the autocorrelation function $Rd(j)$ is not necessarily symmetric. In fact, one can only recover $Rd(\cdot) + Rd(\cdot)^\intercal$ or $R(\cdot) + R(\cdot)^\intercal$ from the variance ratio functions $\mathcal{VR}d(\cdot)$ and $\mathcal{VR}(\cdot)$. This means that information about lead lag relations are eliminated. Instead we propose the following quantities:

$$\mathcal{VR}_+(K) = I + 2\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)R(j)$$

$$\mathcal{VR}d_+(K) = Rd(0) + 2\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)Rd(j),$$

and the negative counterparts $\mathcal{VR}_-(K) = \mathcal{VR}_+^\intercal(K)$ and $\mathcal{VR}d_-(K) = \mathcal{VR}d_+^\intercal(K)$, which have the property that: $\mathcal{VR}(K) = (\mathcal{VR}_+(K) + \mathcal{VR}_+^\intercal(K))/2$ and $\mathcal{VR}d(K) = (\mathcal{VR}d_+(K) + \mathcal{VR}d_+^\intercal(K))/2$. One can test the null hypothesis of lack of linear predictability based on the matrices $\mathcal{VR}d_+(K), \mathcal{VR}d_-(K)$ and one can compare the two statistics to quantify the asymmetry in lead lag effects.

2.3 Univariate Parameters of Interest

We discuss here some univariate parameters of interest both for statistical purposes and economic interpretatability.

2.3.1 Trace and Determinant

The determinant and trace are commonly used univariate functions of covariance matrices that feature in a lot of likelihood ratio testing literature, see for example Szroeter (1978).[6] The trace statistic is widely used to capture the average effect of many individual variance ratios, see for example Table 2.3 in Lo and MacKinlay (1999), and Castura et al. (2010). The Generalized Variance Ratio (Anderson, 2003) statistic would be

$$\det\left(\mathcal{VR}(K)\right) = \frac{\det\left(\Sigma(K)/K\right)}{\det\left(\Sigma\right)} = \frac{\det\left(\Sigma(K)\right)}{K^d \det\left(\Sigma\right)}.$$

[6] These quantities are both invariant to nonsingular linear transformations of the data, i.e., $X_t \mapsto a + AX_t$, where A is a nonsingular $d \times d$ matrix. Furthermore, for both these functions f, $f(\mathcal{VR}a(K)) = f(\mathcal{VR}(K))$.

Cho and White (2014) Lemma 1 says that $\mathcal{VR}(K) = I$ if and only if $\det(\mathcal{VR}(K)) = 1$ and $\operatorname{tr}(\mathcal{VR}(K)) = d$, so from a statistical point of view these quantities capture the meaning of the null hypothesis.[7]

2.3.2 Eigenvalues

Define the spectrum $\sigma(\mathcal{VR}(K)) = \{\lambda \in \mathbb{R} : \mathcal{VR}(K)x = \lambda x \text{ for some } x \in \mathbb{R}^d \backslash \{0\}\}$ of the variance ratio statistic and let $\lambda_{\max}(K), \lambda_{\min}(K)$ denote the largest and smallest elements of $\sigma(\mathcal{VR}(K))$. Under the null hypothesis, $\lambda_{\max}(K) = \lambda_{\min}(K) = 1$, but under the alternative hypothesis they can take any non-negative values. These quantities give univariate measures of the predictability obtainable within the series as we next show. Consider a portfolio of assets with fixed weights $w \in \mathbb{R}^d$. Denoting $vr_K(z_t)$ by the univariate variance ratio of the scalar series z_t, and letting $\widetilde{w} = \Sigma^{1/2}w$ and $Y_t = \Sigma^{-1/2}X_t$, we have

$$\begin{aligned} vr_K(w^\intercal X_t) &= vr_K(w^\intercal \Sigma^{1/2}\Sigma^{-1/2}X_t) \\ &= vr_K(\widetilde{w}^\intercal Y_t) \\ &= \frac{\widetilde{w}^\intercal \mathcal{VR}(K; Y_t)\widetilde{w}}{\widetilde{w}^\intercal \widetilde{w}} \\ &= \frac{\widetilde{w}^\intercal \mathcal{VR}(K; X_t)\widetilde{w}}{\widetilde{w}^\intercal \widetilde{w}} \\ &\leq \lambda_{\max}(\mathcal{VR}(K; X_t)). \end{aligned}$$

This follows because $\mathcal{VR}(K; X_t) = \mathcal{VR}(K; \Sigma^{-1/2}X_t) = \mathcal{VR}(K; Y_t)$. This says that the largest eigenvalue of the variance ratio matrix is an upper bound on the univariate variance ratio of any portfolio with fixed ex-post weights. Likewise, the smallest eigenvalue of the variance ratio matrix provides a lower bound on the variance ratio of any portfolio with fixed weights. The weights that achieve it are given by the corresponding eigenvectors of the variance ratio matrix. Compare with Lo and MacKinlay (1999, p258).

2.3.3 Global Minimum Variance

The variance ratio matrix can also tell us about other portfolios constructed from the underlying assets. The variance of the portfolio $w^\intercal X_t(K)$ is $w^\intercal \Sigma(K)w$. The global minimum variance portfolio

[7] The Gaussian likelihood ratio test for the equality of two matrices Σ_1 and Σ_2 can be based on the quantity $\det(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2})/(\det(I + \Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}))^2$, Anderson (2003, chapter 10).

weights are $w_{mv}(K) = \Sigma(K)^{-1}i/i^\intercal \Sigma(K)^{-1}i$, which results in global minimum variance $1/i^\intercal \Sigma(K)^{-1}i$. By plotting this as a function of K one sees the variation of the least risk portfolio by horizon. This comparison does not depend on the matrix Σ so if we consider the normalized returns $Y_t(K) = K^{-1/2}\Sigma^{-1/2}(X_t(K) - \mu)$ then the variance of $w^\intercal Y_t(K)$ is $w^\intercal \Sigma^{-1/2}\Sigma(K)\Sigma^{-1/2}w/K = w^\intercal \mathcal{VR}(K)w$ and the best portfolio is $w_{mv}(K) = \mathcal{VR}(K)^{-1}i/i^\intercal \mathcal{VR}(K)^{-1}i$ with resulting variance

$$GMV(K) = \frac{1}{i^\intercal \mathcal{VR}(K)^{-1}i}. \tag{17}$$

Under the null hypothesis this should be equal to $1/d$ for all K.

2.3.4 Off-Diagonal Elements

We are also interested in several other univariate parameters based on $\mathcal{VR}d_+(K)$. First, the diagonal elements of $\mathcal{VR}d_+(K)$ correspond to the univariate variance ratio statistics. Second, the off-diagonal elements of $\mathcal{VR}d_+(K)$ provide the information about the directional lead lag pattern between the assets. Third, the differences between two corresponding off-diagonal elements of $\mathcal{VR}d_+(K)$ indicate the asymmetry in the lead lag relationships between the assets. If one of the assets is a common factor portfolio, the corresponding off-diagonal elements of $\mathcal{VR}d_+(K)$ and $\mathcal{VR}d_-(K)$ give an idea of the dynamic comovement of the asset with the common factor portfolio, which could be used in cross-sectional regression analysis.

Another parameter of interest is the average of the off diagonal elements of $\mathcal{VR}d(K)$, which is

$$CS(K) = \frac{2}{d(d-1)}\sum_{i=1}^{d-1}\sum_{j=i+1}^{d}\mathcal{VR}d_{ij}(K) = \frac{1}{d(d-1)}\{i^\intercal \mathcal{VR}d(K)i - \mathrm{tr}(\mathcal{VR}d(K))\}, \tag{18}$$

see Solnik (1991) and Bailey, Kapetanios, and Pesaran (2012) who consider the case $K = 0$ and large d. Under the null hypothesis $CS(K) = CS(1)$ for all K. This measures in some average sense the cross dependence at different lags.

2.3.5 Dynamic Momentum/Contrarian Portfolio Profit

We consider a generalization of the Lo and MacKinlay (1990) type arbitrage portfolio contrarian strategies. Specifically, consider the following portfolio weights applied to the normalized investments $Z_t = D^{-1/2}(X_t - \mu)$

$$\widetilde{w}_{it}(K) = \pm\frac{2}{d(K-1)}\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)\left(Z_{i,t-j} - \overline{Z}_{t-j}\right) \tag{19}$$

where $\overline{Z}_s = \sum_{i=1}^d Z_{is}/d = i^\intercal Z_s/d$ so that $\sum_{i=1}^d \widetilde{w}_{it}(K) = 0$. This strategy considers the signals $Z_{i,t-1} - \overline{Z}_{t-1}, \ldots, Z_{i,t+1-K} - \overline{Z}_{t+1-K}$ and downweights them according to their lag: if the $\pm$ factor is positive, this can be considered a momentum strategy, while if it is negative, this can be considered a contrarian strategy. The expected profit of this strategy is

$$\begin{aligned}
\pi_\pm(K) &= E\widetilde{w}_t^\intercal(K)Z_t = \pm\frac{2}{d(K-1)}\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)E\left[\left(Z_{t-j}-\overline{Z}_{t-j}i\right)^\intercal Z_t\right] \\
&= \pm\frac{2}{d^2(K-1)}\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)E\left[i^\intercal Z_{t-j}Z_t^\intercal i\right] - \pm\frac{2}{d(K-1)}\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)E\left[Z_{t-j}^\intercal Z_t\right] \\
&= \pm\frac{2}{d^2(K-1)}i^\intercal\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)R(j)^\intercal i - \pm\frac{2}{d(K-1)}\mathrm{tr}\left(\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)R(j)\right) \\
&= \pm\frac{1}{d^2(K-1)}i^\intercal\mathcal{VR}d(K)i - \pm\frac{1}{d(K-1)}\mathrm{tr}\left(\mathcal{VR}d(K)\right) \pm \frac{1}{K-1}\left(1-\frac{1}{d^2}i^\intercal R(0)i\right) \\
&= \pm\frac{2}{d^2(K-1)}\sum_{i=1}^{d-1}\sum_{j=i+1}^{d}\left[\mathcal{VR}d_{ij}(K)-\rho_{ij}\right] \pm \frac{d-1}{d^2(K-1)}\mathrm{tr}\left(I-\mathcal{VR}d(K)\right).
\end{aligned}$$

Under the martingale hypothesis, $\pi_\pm(K) = 0$ for all K. This quantity weights diagonal departures and off diagonal departures similarly.

3 Estimation

Suppose that we observe the return vectors $\{X_t, t = 1, \ldots, T\}$ equally spaced in discrete time. We may estimate the variance ratios in several ways, for example by estimating the sample covariance matrix of the K frequency data and the original observations and then forming the ratio.[8] We can alternatively explicitly use the population connection with the autocorrelation matrix process in (12) for example.

We estimate the population quantities by sample averages:

$$\overline{X} = \frac{1}{T}\sum_{t=1}^T X_t \quad ; \quad \widehat{\Gamma}(j) = \frac{1}{T}\sum_{t=j+1}^T \left(X_t - \overline{X}\right)\left(X_{t-j} - \overline{X}\right)^\intercal,\ j = 0, 1, 2, \ldots$$

$$\widehat{\Sigma}(K) = \frac{1}{T}\sum_{t=K}^T \left(X_t(K) - K\overline{X}\right)\left(X_t(K) - K\overline{X}\right)^\intercal$$

[8] As pointed out by Hillman and Salmon (2007) with unequally spaced data, this approach can yield a "natural" variance ratio by classifying observations on the duration since the previous trade.

$$\widehat{\Sigma} = \widehat{\Gamma}(0) \quad ; \quad \widehat{D} = \mathrm{diag}[\widehat{\Gamma}(0)] \quad ; \quad \widehat{R}(j) = \widehat{\Sigma}^{-1/2}\widehat{\Gamma}(j)\widehat{\Sigma}^{-1/2};$$

$$\widehat{R}d(j) = \widehat{D}^{-1/2}\widehat{\Gamma}(j)\widehat{D}^{-1/2} \quad ; \quad \widehat{R}_L(j) = \widehat{\Gamma}(j)\widehat{\Sigma}^{-1} \quad ; \quad \widehat{R}_R(j) = \widehat{\Sigma}^{-1}\widehat{\Gamma}(j)$$

$$\widehat{\mathcal{VR}}(K) = I + \sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)(\widehat{R}(j) + \widehat{R}(j)^{\intercal})$$

$$\widehat{\mathcal{VR}}^{\&}(K) = \widehat{\Sigma}^{-1/2}\widehat{\Sigma}(K)\widehat{\Sigma}^{-1/2}/K$$

$$\widehat{\mathcal{VR}}_{+}(K) = I + 2\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)\widehat{R}(j),$$

and likewise for $\widehat{\mathcal{VR}}d(K)$, $\widehat{\mathcal{VR}}(K,L)$, $\widehat{\mathcal{VR}}d^{\&}(K)$, etc. Note that by construction $\widehat{\mathcal{VR}}(K)$ and $\widehat{\mathcal{VR}}^{\&}(K)$ are symmetric and positive semidefinite.

We may also calculate the univariate quantities by analogy. For example, define the estimated spectrum $\widehat{\sigma}(\widehat{\mathcal{VR}}(K)) = \{\lambda \in \mathbb{R} : \widehat{\mathcal{VR}}(K)x = \lambda x \text{ for some } x \in \mathbb{R}^d\backslash\{0\}\}$ of the variance ratio statistic and let $\widehat{\lambda}_{\max}(K), \widehat{\lambda}_{\min}(K)$ denote the largest (smallest) elements of $\widehat{\sigma}(\widehat{\mathcal{VR}}(K))$.

4 Asymptotic Theory and Inference

4.1 Regularity Conditions

We present two alternative sets of sampling assumptions, which we denote by A and MH*. Assumptions MH* are modified versions of the assumptions in Lo and MacKinlay (1988) adapted to the multivariate case and corrected for what appears to be an error; these conditions do not require stationarity although certain averages need to converge. Most treatments of variance ratios follow the Lo and MacKinlay (1988) assumption H, which includes a mixing condition and some further restriction on the structure of the higher moments (their condition H4), which purportedly implies that the sample autocorrelations are asymptotically independent.[9] In the multivariate context, their assumption H4 would be that

$$E[\widetilde{X}_{it}\widetilde{X}_{jt}\widetilde{X}_{kr}\widetilde{X}_{ls}] = 0 \text{ for all } i,j,k,l,t, \text{ and } r,s \text{ with } r < s < t. \tag{20}$$

This assumption rules out leverage type effects, e.g., $E[\widetilde{X}_{it}^2|\widetilde{X}_{ir}\widetilde{X}_{is}] \neq 0$, which may be important for some assets, see Nelson (1991). This assumption is not necessary for the distribution theory; imposing

[9] Some papers including Whang and Kim (2003) dispense with this latter assumption but maintain the mixing and moment assumption.

it (along with other conditions) would simplify the asymptotic variance to be single finite sums rather than double finite sums, but in practice this is not a big issue. We shall dispense with this assumption below, but we shall make a further assumption that appears to have been omitted by mistake from Lo and MacKinlay (1988). Namely, implicit in their analysis is that $\widetilde{X}_t\widetilde{X}_{t-j}$ is uncorrelated with $\widetilde{X}_s\widetilde{X}_{s-j}$ but this does not follow from $\widetilde{X}_t$ being an uncorrelated sequence (although it does follow if $\widetilde{X}_t$ were a martingale difference sequence).

Define for $j, k = 0, 1, 2, \ldots$:

$$\Xi_{jk} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left[\left(\widetilde{X}_{t-j}\widetilde{X}_{t-k}^{\intercal} \otimes \widetilde{X}_t\widetilde{X}_t^{\intercal}\right)\right] \quad ; \quad c_{j,K} = 2\left(1 - \frac{j}{K}\right)$$

$$\begin{aligned}
Q(K) &= \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_{j,K}c_{k,K}\left(\Sigma^{-1/2}\otimes\Sigma^{-1/2}\right)\Xi_{jk}\left(\Sigma^{-1/2}\otimes\Sigma^{-1/2}\right) \\
Qd(K) &= \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_{j,K}c_{k,K}\left(D^{-1/2}\otimes D^{-1/2}\right)\Xi_{jk}\left(D^{-1/2}\otimes D^{-1/2}\right) \\
Qa(K) &= \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_{j,K}c_{k,K}\left(\Sigma^{-1}\otimes I\right)\Xi_{jk}\left(\Sigma^{-1}\otimes I\right).
\end{aligned}$$

We shall assume that the matrices Σ, $Q(K)$, $Qd(K)$, and $Qa(K)$ are strictly positive definite. We consider the following sets of alternative assumptions:

Assumption A.

A1. *The process $\widetilde{X}_t$ is a stationary ergodic Martingale Difference sequence;*

A2. *The process $\widetilde{X}_t$ has finite fourth moments, i.e., for all i, j, k, l, $E[|\widetilde{X}_{it}\widetilde{X}_{jt}\widetilde{X}_{kt}\widetilde{X}_{lt}|] < \infty$.*

Assumption MH*.

MH1. (i) *For all t, $\widetilde{X}_t$ satisfies $E\widetilde{X}_t = 0$, $E\big[\widetilde{X}_t\widetilde{X}_{t-j}^{\intercal}\big] = 0$ for all $j \neq 0$;* (ii) *for all t, s with $s \neq t$ and all $j, k = 1, \ldots, K$, $E\big[\widetilde{X}_t\widetilde{X}_{t-j}^{\intercal} \otimes \widetilde{X}_s\widetilde{X}_{s-k}^{\intercal}\big] = 0$.*

MH2. *$\widetilde{X}_t$ is α-mixing with coefficient $\alpha(m)$ of size $r/(r-1)$, where $r > 1$, such that for all t and for any $j \geq 0$, there exists some $\delta > 0$ for which $\sup_t E|\widetilde{X}_{it}\widetilde{X}_{k,t-j}|^{2(r+\delta)} < \Delta < \infty$ for all $i, k = 1, \ldots, d$;*

MH3. *For all* j, k, *the following limits exist:* $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[\widetilde{X}_t\widetilde{X}_t^{\intercal}] =: \Sigma < \infty$ *and* $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left[\widetilde{X}_{t-j}\widetilde{X}_{t-k}^{\intercal} \otimes \widetilde{X}_t\widetilde{X}_t^{\intercal}\right] =: \Xi_{jk} < \infty$.

Chen and Deo (2006) work with martingale difference sequence but also assume a no leverage condition. In MH* we include the additional condition (ii) $E[\widetilde{X}_t\widetilde{X}_{t-j}^{\intercal} \otimes \widetilde{X}_s\widetilde{X}_{s-k}^{\intercal}] = 0$, for all $s \neq t$ and all $j, k = 1, \ldots, K$; this is not a consequence of (2) in general. Without this additional assumption the asymptotic variance of the variance ratio statistics are much more complicated and hard to estimate, involving the selection of a bandwidth parameter. Condition MH1(ii) is satisfied automatically under the martingale hypothesis, which itself is consistent with any kind of nonlinear multivariate ("semi-strong") GARCH process. In assumption A, we have assumed strict stationarity, whereas this is not required in MH* (although certain sums have to converge in MH3, which would rule out explosive nonstationarity).[10] In MH* we have assumed higher moments depending on the mixing decay rate, whereas for assumption A only four moments are required and no explicit mixing conditions are employed. It should be noted therefore that the conditions A and MH* are non-nested. We further note that under the assumption that returns are i.i.d. (referred to as RW1 in Campbell, Lo, and MacKinlay (1997)), the univariate version of the CLT's below are valid under only second moments, Brockwell and Davis (1991, Theorem 7.2.2), due to the self normalization present in the sample autocorrelations. For similar reasons, condition MH3 may not be strictly necessary in that mildly trending moments may still permit a CLT at the same rate due to the cancellation of numerator by denominator.

We remark that this theory is predicated on the existence of fourth moments, which may be problematic for some financial time series. Provided only the population variance exists, the matrix normalized variance ratio converges in probability to the identity, but may have a non-standard limiting distribution and a slower rate of convergence to it, Phillips and Solo (1992) and Mikosch and Stărică (2000).[11] Even if the population variance does not exist, the sample variance ratio may converge, due to the self-normalization, but one can expect a different scaling law. For example, if the return process is iid with a symmetric stable distribution with parameter $\alpha \in [1, 2]$, then the sample variances scale according to $K^{2/\alpha}$, that is, as $T \to \infty$, $\widehat{\mathcal{VR}}(K) \to K^{(2-\alpha)/\alpha}$ for all K. This is similar asymptotic behaviour to what is found under the bubble process of section 5.2 below when

[10] In the working paper version of this paper (HLZ) we extended conditions A to allow for a time varying mean (that has to be estimated) and a time varying variance (that does not have to be estimated).

[11] For stationary univariate linear processes, the sample autocorrelations can be root-T consistent and asymptotically normal under only second moment assumptions, Brockwell and Davis (1991, Theorem 7.2.2), but this result does not hold for nonlinear processes like GARCH.

$\alpha = 1$. Wright (2000) has proposed variance ratios based on signs and ranks that are robust to heavy tailed distributions, although require stronger assumptions elsewhere.

4.2 Finite/fixed horizon Limiting Distribution Theory

We next present our main results. In this subsection we consider the finite K framework.

THEOREM 1. *Suppose that either Assumption A or MH* holds. Then, as $T \to \infty$:*

$$\begin{aligned} \sqrt{T}\text{vec}\left(\widehat{\mathcal{VR}}_{+}(K) - I_d\right) &\implies N\big(0, Q(K)\big) \\ \sqrt{T}\text{vec}\left(\widehat{\mathcal{VR}d}_{+}(K) - \widehat{Rd}(0)\right) &\implies N\big(0, Qd(K)\big) \\ \sqrt{T}\text{vec}\left(\widehat{\mathcal{VR}a}_{+}(K) - I_d\right) &\implies N\big(0, Qa(K)\big). \end{aligned}$$

Asymptotic results for the corresponding two-sided statistics can be derived using the matrix transformation argument of Magnus and Neudecker (1980). In the paper it is shown that for any square matrix A, $\frac{1}{2}\text{vech}\left(A + A^{\intercal}\right) = L\frac{1}{2}\left(I + K\right)\text{vec}\left(A\right) = D_n^{+}\text{vec}\left(A\right)$ where L and K are the so-called elimination and commutation matrices, respectively, and D_n^{+} is the Moore-Penrose pseudoinverse of the duplication matrix. The reader is referred to their paper (Lemma 3.1 and 3.6) for precise definition of these matrices. It now follows that

$$\sqrt{T}\text{vech}\left(\widehat{\mathcal{VR}}(K) - I_d\right) \implies N\big(0, S(K)\big), \tag{21}$$

where $S(K) = D_n^{+}Q(K)D_n^{+\intercal}$. Likewise, $\sqrt{T}\text{vech}(\widehat{\mathcal{VR}d}(K) - \widehat{Rd}(0)) \implies N\big(0, Sd(K)\big)$ and $\sqrt{T}\text{vech}(\widehat{\mathcal{VR}a}(K) - I_d) \implies N\big(0, Sa(K)\big)$, where $Sd(K) = D_n^{+}Qd(K)D_n^{+\intercal}$ and $Sa(K) = D_n^{+}Qa(K)D_n^{+\intercal}$. We note that (under our conditions) the difference between $\widehat{\mathcal{VR}}^{\&}(K)$ and $\widehat{\mathcal{VR}}(K)$ for example is negligible, i.e., $O_p(T^{-1})$, see (51), so these statistics have exactly the same limiting distribution.

Limiting distributions for smooth functions of the variance ratio matrices can be obtained by the delta method. For any $f : \mathbb{R}^{d(d+1)/2} \to \mathbb{R}$ that is differentiable at $\theta_0(\equiv \text{vech}(I_d)$ or $\text{vech}(\widehat{\Gamma}d(0))$, respectively), we have

$$\sqrt{T}\left[f\left(\text{vech}\left(\widehat{VR}(K)\right)\right) - f\big(\text{vech}\left(I_d\right)\big)\right] \implies N\left(0, \nabla f(\theta_0)^{\intercal} S(K) \nabla f(\theta_0)\right), \tag{22}$$

where

$$\nabla f(\theta_0)^{\intercal} \equiv \left.\left(\frac{\partial f(y)}{\partial y_1}, \ldots, \frac{\partial f(y)}{\partial y_{d(d+1)/2}}\right)\right|_{y=\theta_0}$$

and likewise for the diagonally normalized statistic and the right normalized one. It is straightforward to obtain the asymptotic distributions of the CS, GMV, π, and other statistics; we collect the formulae below in the table.

Statistic	**Asymptotic Variance**[12]
$\det(VR(K)), \det(VRd(K))$	$\delta^\intercal S(K)\delta, \delta^\intercal Sd(K)\delta$
$\mathrm{tr}(VR(K)), \mathrm{tr}(VRd(K))$	$\delta^\intercal S(K)\delta, \delta^\intercal Sd(K)\delta$
$GMV(K), CS(K), \pi(K)$	$d^{-4}i^\intercal D_n^+ S(K) D_n^{+\intercal} i, \frac{1}{d^2(d-1)^2} b^\intercal Sd(K)b, c^\intercal Qd(K)c$
$VR_{ij}(K), VRd_{ij}(K)$	$e_{ij}^\intercal S(K)e_{ij}, e_{ij}^\intercal Sd(K)e_{ij}$

For the individual eigenvalues, we employ a different approach as they are not smooth functions of the variance ratio matrix under the null hypothesis. Specifically, Eaton and Tyler (1991, Theorem 3.2) show that if the random symmetric matrix $\sqrt{T}(\widehat{\mathcal{VR}}(K) - I_d)$ converges in distribution to a matrix random variable, denoted W, then with $i_d = (1, 1, \ldots, 1)^\intercal$

$$\sqrt{T}\left(\varphi(\widehat{\mathcal{VR}}(K)) - i_d\right) \Longrightarrow \varphi(W), \tag{23}$$

where $\varphi(\widehat{\mathcal{VR}}(K))$ and $\varphi(W)$ are $d \times 1$ vectors of ordered eigenvalues $\lambda_j \in \varphi(\widehat{\mathcal{VR}}(K))$ and $\lambda_j^* \in \varphi(W)$, respectively. Using the continuous mapping theorem (and/or the delta method) on (23), we may also derive asymptotics for the functions of univariate eigenvalues. For instance,

$$\sqrt{T}\left(\lambda_{\max} - 1\right), \sqrt{T}\left(\lambda_{\min} - 1\right) \Longrightarrow \left(\lambda_{\max}^*, \lambda_{\min}^*\right).$$

4.3 Standard Errors

From the expressions in Theorem 1 we can obtain pointwise confidence intervals for scalar functions of the matrices $\widehat{\mathcal{VR}}(K)$ or $\widehat{\mathcal{VR}}d(K) - \widehat{Rd}(0)$ or $\widehat{\mathcal{VR}a}(K)$. Specifically, let

$$\widehat{\Xi}_{jk} = \frac{1}{T}\sum_{t=\max\{j,k\}+1}^{T}\left(X_{t-j} - \overline{X}\right)\left(X_{t-k} - \overline{X}\right)^\intercal \otimes \left(X_t - \overline{X}\right)\left(X_t - \overline{X}\right)^\intercal \tag{24}$$

$$\widehat{Q}(K) = \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_{j,K}c_{k,K}\left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2}\right)\widehat{\Xi}_{jk}\left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2}\right) \tag{25}$$

[12] Here, i is a conformable column vector of ones, $\delta = \mathrm{vech}(I_d)$, $b := i - \delta$, and c is a column vector that has $(1-d)/(d^2(K-1))$ at $(l(d+1)+1)^{th}$ entries $(l = 0, \ldots, d-1)$, and $1/(d^2(K-1))$ at other entries. Also, e_{ij} is a column vector having ones at $d(j-1) + i - \{1 + \cdots + (j-1)\}$ entries and zeros otherwise.

and $\widehat{S}(K) = D_n^+ Q(K) D_n^{+\intercal}$. Similarly, we may define $\widehat{Q}a(K)$ and $\widehat{S}a(K)$, replacing $\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2}$ by $\widehat{\Sigma}^{-1} \otimes I$ in (25). For the diagonal statistic define $\widehat{Q}d(K)$ and $\widehat{S}d(K)$, replacing $\widehat{\Sigma}^{-1/2}$ by $\widehat{D}^{-1/2}$ in (25). Specifically, the standard error for $\widehat{\mathcal{VR}}d_{ii}(K)$ is

$$\widehat{Q}d_{ii}(K) = \frac{1}{\widehat{\sigma}_{ii}^2} \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_{j,K} c_{k,K} \widehat{\Xi}_{ii;jk} \tag{26}$$

$$\widehat{\Xi}_{ii;jk} = \frac{1}{T} \sum_{t=\max\{j,k\}+1}^{T} \left(X_{it-j} - \overline{X}_i\right) \left(X_{it-k} - \overline{X}_i\right) \left(X_{it} - \overline{X}_i\right)^2$$

$$\widehat{\sigma}_{ii} = \frac{1}{T} \sum_{t=1}^{T} \left(X_{it} - \overline{X}_i\right)^2 .$$

The standard errors for $\widehat{GMV}(K)$, $\widehat{\pi}(K)$, and other univariate quantities can be obtained from this.

Corollary 1. *Suppose that either Assumption A or MH* holds. Then (for each fixed K) the estimator $\widehat{Q}(K)$ is weakly consistent for $Q(K)$ (likewise, $\widehat{Q}d(K)$ and $\widehat{Q}a(K)$ are weakly consistent for $Qd(K)$ and $Qa(K)$), i.e., as $T \to \infty$,*

$$\widehat{Q}(K) \xrightarrow{P} Q(K).$$

It follows from this that Theorem 1 can be extended to include the feasible normalized test statistics.

Note that under the Lo and MacKinlay (1988) condition H4 (i.e. (20)) we have $\Xi_{jk} = 0$ for $j \neq k$, so that the asymptotic variance simplifies, a little. The commonly used standard error is in matrix notation

$$\widehat{Q}d_{LM}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 \left(\widehat{D}^{-1/2} \otimes \widehat{D}^{-1/2}\right) \widehat{\Xi}_{jj} \left(\widehat{D}^{-1/2} \otimes \widehat{D}^{-1/2}\right), \tag{27}$$

whose diagonal elements can be compared with (26).

In the iid case, we further have $\Xi_{jj} = \Sigma \otimes \Sigma$ and:

$$Q_{iid}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 I_{d^2} \ ; \ \widehat{Qd}_{iid}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 (\widehat{Rd}\,(0) \otimes \widehat{Rd}\,(0)) \ ; \ \widehat{Qa}_{iid}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 (\widehat{\Sigma}^{-1} \otimes \widehat{\Sigma}). \tag{28}$$

In the scalar case these are all nuisance parameter free. As we show in the application, the standard errors derived from (25), (27), and (28) can be quite different; generally speaking the standard errors from $\widehat{Q}(K)$ are larger than the standard errors from $\widehat{Q}_{LM}(K)$, which in turn are larger than the standard errors from the i.i.d special case $\widehat{Q}_{iid}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 I_{d^2}$.

Alternative inference methods such as self-normalization, or bootstrap and subsampling may give better results, although they are designed to accommodate the more general uncorrelatedness assumption that allows $E\big[\widetilde{X}_t\widetilde{X}_{t-j}^{\intercal} \otimes \widetilde{X}_s\widetilde{X}_{s-k}^{\intercal}\big] \neq 0$ for some $s \neq t$. The readers are directed to Lobato (2001) and Whang and Kim (2003) for description of these methods. In the Appendix we present a bias correction method based on asymptotic expansions, which may give better performance for long lags.

4.4 Two Parameter Statistics and Efficiency

For the two parameter variance ratio statistic $\widehat{\mathcal{VR}}(K, L)$ we obtain under the same conditions (either A or MH) that

$$\sqrt{T}\text{vech}\left(\widehat{\mathcal{VR}}(K, L) - I_d\right) \Longrightarrow N\left(0, S(K, L)\right), \tag{29}$$

where $S(K, L) = D^+Q(K, L)D^{+^\intercal}$ with

$$Q(K, L) = \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} \widetilde{c}_{j,K,L}\widetilde{c}_{k,K,L}\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\Xi_{jk}\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right),$$

$$\widetilde{c}_{j,K,L} = c_{j,K} - c_{j,L} = 2\left(\frac{K-L}{KL}j\right)1(j \leq L-1) + 2\left(1 - \frac{j}{K}\right)1(L \leq j \leq K-1).$$

Similar results hold for the other two parameter statistics. Note that under the iid case,

$$Q_{iid}(K, L) = \sum_{j=1}^{K-1} \widetilde{c}_{j,K,L}^2 I_{d^2}$$

We can compare the relative efficiency of the two parameter variance ratio estimator $\widehat{\mathcal{VR}}(LJ, L)$ relative to the one parameter variance ratio estimator $\widehat{\mathcal{VR}}(J)$, for any positive integers L, J. We show that the relative efficiency (when returns are iid) for the general $J, L \geq 2$ case is

$$\begin{aligned}
\frac{Q_{iid}(LJ, L)}{LQ_{iid}(J)} &= \frac{\sum_{j=1}^{JL-1}\widetilde{c}_{j,JL,L}^2}{L\sum_{j=1}^{J-1}c_{j,J}^2} \\
&= \frac{(2J-2)L^2 + 1}{L^2\left(2J-1\right)} \\
&= 1 - \frac{L^2 - 1}{L^2\left(2J-1\right)} > 2/3 \\
&< 1.
\end{aligned}$$

This gives quite modest improvements in efficiency.

4.5 Increasing horizon Limiting Distribution Theory

In this section we consider the case where K is allowed to increase. Richardson and Stock (1989) considered the framework in which $K = K(T)$ and $K/T \to \delta < 1$. Deo and Richardson (2003) point out that under this particular restriction variance ratio test is not consistent against several important mean reverting alternatives. Chen and Deo (2006) consider an alternative framework that allows K to increase in such a way that K/T tends to zero. They assumed a set of rather strong conditions on cross-moments (see their Assumption A3) that include a no-leverage condition, and a mixing-like condition (Assumption A6) that forces asymptotic independence of the process. We shall suppose that $K^2/T \to 0$ and otherwise impose weaker conditions that include Assumption A above.

Denote by $\overline{Q}(K) = K^{-1}Q(K)$, $\overline{Qd}(K) = K^{-1}Qd(K)$, and $\overline{Qa}(K) = K^{-1}Qa(K)$. Furthermore, define the matrix

$$Q(\infty) := \lim_{K\to\infty} \overline{Q}(K) = \lim_{K\to\infty} \frac{1}{K}\left[\sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_{j,K}c_{k,K}\left(\Sigma^{-1/2}\otimes\Sigma^{-1/2}\right)\Xi_{jk}\left(\Sigma^{-1/2}\otimes\Sigma^{-1/2}\right)\right]$$

and similarly $Qd(\infty) = \lim_{K\to\infty}\overline{Qd}(K)$, and $Qa(\infty) = \lim_{K\to\infty}\overline{Qa}(K)$.

Assumption T. *The horizon $K = K(T) \to \infty$ in such a way that $K^2/T \to 0$ as $T \to \infty$.*

Assumption C. *The following double sum is finite: $\sum_{p=-\infty}^{\infty}\sum_{q=-\infty}^{\infty}|\kappa(0,0,p,q)| < \infty$, where $\kappa(a,b,c,d)$ is the cumulant of 4th order between $(\widetilde{X}_a, \widetilde{X}_b, \widetilde{X}_c, \widetilde{X}_d)$.*

It can be shown that Assumption C (along with Assumption A) is sufficient to guarantee the existence and positive definiteness of the matrix limits $Q(\infty)$, $Qa(\infty)$ and $Qd(\infty)$. Assuming absolute summability of cumulants is rather common in the time series literature as it is rather a weak condition; it is not necessarily implied by ergodicity, but by a mild α-mixing condition (along with some higher moment condition) as is shown by Andrews (1991, Lemma 1). For example Assumption MH2 is sufficient for Assumption C.

We note that Assumption MH* itself alone is not suitable for deriving the limiting distribution when K is an increasing sequence, because the mixing property is (usually) not preserved under a measurable transformation of infinite dimension; $g : \mathbb{R}^\infty \to \mathbb{R}$, although one could work with near epoch dependence to obtain a similar result.

Theorem 2. *Suppose that Assumptions A, T and C hold. Then:*

$$\begin{aligned}
\overline{Q}(K)^{-1/2}\sqrt{\frac{T}{K}}\mathrm{vec}\left(\widehat{\mathcal{VR}}_{+}(K)-I_d\right) &\Longrightarrow N\left(0, I_{d^2}\right)\\
\overline{Qd}(K)^{-1/2}\sqrt{\frac{T}{K}}\mathrm{vec}\left(\widehat{\mathcal{VR}d}_{+}(K)-\widehat{Rd}(0)\right) &\Longrightarrow N\left(0, I_{d^2}\right)\\
\overline{Qa}(K)^{-1/2}\sqrt{\frac{T}{K}}\mathrm{vec}\left(\widehat{\mathcal{VR}a}_{+}(K)-I_d\right) &\Longrightarrow N\left(0, I_{d^2}\right).
\end{aligned}$$

As in the finite K framework, the 'two-sided' versions of the variance ratio statistics (and the univariate functions thereof) can be obtained by standard matrix transformation arguments. Now define

$$\begin{aligned}
\widehat{Q}(K) &= \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_{j,K}c_{k,K}\left(\widehat{\Sigma}^{-1/2}\otimes\widehat{\Sigma}^{-1/2}\right)\widehat{\Xi}_{jk}\left(\widehat{\Sigma}^{-1/2}\otimes\widehat{\Sigma}^{-1/2}\right)\\
\widehat{\Xi}_{jk} &= \frac{1}{T}\sum_{t=max\{j,k\}+1}^{T}\left(X_{t-j}-\overline{X}\right)\left(X_{t-k}-\overline{X}\right)^{T}\otimes\left(X_t-\overline{X}\right)\left(X_t-\overline{X}\right)^{T}.
\end{aligned}$$

Corollary 2. *Suppose that either Assumption A, and additionally T and C hold. Then, as $T\to\infty$,*

$$K^{-1}\widehat{Q}(K)-\overline{Q}(K)\xrightarrow{P}0.$$

This says that the inference methods we apply in the finite K case can be carried over to the increasing K case, at least where K is not too large relative to the square root of the sample size.

5 Alternative Hypotheses

There are many plausible alternative hypotheses to the null hypothesis (2). We look in detail at several alternative models in this section. In general they yield a prediction of the form

$$\Sigma_T(K)=K\Sigma+\Delta(K,T), \tag{30}$$

where $\Delta(K,T)$ is a symmetric matrix such that $\Sigma_T(K)>0$.

5.1 Local Alternatives

We first extend the arguments presented by Faust (1992) to the multivariate case and show that a trace test will be optimal against a certain class of alternatives. The type of mean reversion against which the test is best at detecting will be shown to be a special case of $VAR(K-1)$. The main idea is to find a statistic that is *asymptotically equivalent* to the likelihood ratio statistic, since in such a case the test based on that statistic will possess the same local large-sample optimality properties of LR tests, see Engle (1984). Below we show that the statistic based on $\mathrm{tr}(\widehat{VR}(K))$ (defined formally in section 8.1.3. below) is optimal (under normality) for testing the null hypothesis of no predictability/serial correlation, against the alternative hypothesis that *each marginal* process $\{X_{jt}\}_t$, $j = 1, \ldots, d$ belongs to what is called the $\phi-$*best* class proposed by Faust (1992). The $\phi-$best class is a particular class of $AR(K-1)$ models, and is defined as the set of those having AR polynomials $\rho_q(L)$ that satisfy

$$\rho_q(z)\rho_q(z^{-1}) = \alpha\left(1 + q\phi(z)\phi(z^{-1})\right) \tag{31}$$

for some constants q and $\alpha > 0$, and z inside the unit circle; the coefficients for the moving average filter $\phi(L)$ are $\phi_j = +1$ for all $j = 0, \ldots, K-1$. From the definition we see that under the alternative hypothesis, $\{X_t\}$ essentially belongs to a (particular) class of *vector* autoregressive process $VAR(K-1)$. We note that when $q = 0$ the process is a white noise (in weak sense of uncorrelatedness, although with joint normality this automatically implies independence). Denote by X the $T \times d$ matrix of sample observations. Then formally, the null and alternative hypotheses can be written as

$$\begin{aligned} &H_0 : X \sim \mathcal{N}_d^T\left(i\mu^\intercal, I_T \otimes \Sigma\right) && \text{[Uncorrelatedness]} \\ &H_1 : X \sim \mathcal{N}_d^T\left(i\mu^\intercal, \Sigma_{q*} \otimes \Sigma\right) && \text{[`}\phi-\text{best' temporal dependence]} \end{aligned}$$

where Σ_{q^*} refers to the variance-covariance matrix of the $\phi-$ best class process with the index of the process $q = q^* > 0$. The notation $\mathcal{N}_d^T$ stands for a matrix normal variable; each matrix (separated by the Kronecker product) in the variance represents the contribution from cross-sectional and temporal sides, respectively. So essentially, this is a one-sided test of the index q being zero *versus* q being a strictly positive constant. Examination of the local large-sample optimality is done by letting the index $q^* = q^*(T) = \delta/\sqrt{T}$ in the alternatives, where δ determines the direction to which the test departs from the null hypothesis.

Proposition FaustM. *Suppose that the data is normally distributed. Then, the trace test is locally most powerful (MP) invariant against alternatives in the $\phi-$best class of the form $q_T^* = \delta/\sqrt{T}$.*

It may be possible to characterize the class of alternatives against which other tests, such as the determinant test, are optimal.

The trace test, while optimal against the specific class above, may have zero power against some alternatives, as we next discuss. Suppose that $\Delta(K, T) = \Delta(K)/\sqrt{T}$, then

$$\sqrt{T}\left(\mathcal{VR}(K) - I\right) = \frac{1}{K}\Sigma^{-1/2}\Delta(K)\Sigma^{-1/2} \quad ; \quad \sqrt{T}\left(\mathcal{VR}d(K) - Rd(0)\right) = \frac{1}{K}D^{-1/2}\Delta(K)D^{-1/2}.$$

Provided $\Delta(K)$ is strictly definite, some tests based on these matrices will have positive power against this alternative. On the other hand, in some cases, the power may be zero. Specifically, suppose we take the trace test applied to the diagonally normalized variance ratio matrix, i.e., compare $\mathrm{tr}(\widehat{\mathcal{VR}d}(K)) - d$ (c.f. Castura et al. (2010)) with the critical values from its normal limit given above, then if $\Delta(K)$ is of the form

$$\Delta_{ij}(K) = \begin{cases} \delta(K) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

for some nonzero $\delta(K)$, then this test will have zero power (although note in this case the trace of the matrix normalized statistics will have power).

5.2 Multivariate Fads Model

We consider an alternative to the efficient market hypothesis (2), which allows for temporary misspricing through fads but assures that the rational price dominates in the long run. Consider the multivariate fads model for log prices:

$$p_t^* = \mu + p_{t-1}^* + \varepsilon_t \tag{32}$$

$$p_t = p_t^* + \eta_t, \tag{33}$$

where ε_t is iid with mean zero and variance matrix Ω_ε, while η_t is a stationary weakly dependent process with unconditional variance matrix Ω_η, and the two processes are mutually independent. It follows that the observed return satisfies

$$X_t = p_t - p_{t-1} = \mu + \varepsilon_t + \eta_t - \eta_{t-1}. \tag{34}$$

This is a multivariate generalization of the scalar Muth (1960) model, which was advocated in Poterba and Summers (1988). It allows actual prices p to deviate from fundamental prices p^* but only in the short run through the fad process η_t. This process is a plausible alternative to the efficient markets hypothesis. If η_t were i.i.d., then X_t would be (to second order) an MA(1) process, which is a structure implied by a number of market microstructure issues (Hasbrouck (2007)). In this case,

$$\mathcal{VR}(K) = I + (1 - \frac{1}{K})(R(1) + R(1)^\intercal) = I - 2(1 - \frac{1}{K})\left(\Omega_\varepsilon + 2\Omega_\eta\right)^{-1/2} \Omega_\eta \left(\Omega_\varepsilon + 2\Omega_\eta\right)^{-1/2},$$

and likewise for $\mathcal{VR}d(K)$.

In general, however, η_t might have any type of weak dependence structure. We next derive a restriction on the long run variance ratio statistic that should reflect the fads process. We do not restrict the fads process, and so can only obtain long run implications.

Consider the K period returns $X_t(K) = K\mu + p_t - p_{t-K} = \sum_{s=t-K}^{t} \varepsilon_s + \sum_{s=t-K}^{t}(\eta_s - \eta_{s-1}) = K\mu + \sum_{s=t-K}^{t} \varepsilon_s + \eta_t - \eta_{t-K}$. These have variance

$$\begin{aligned} \Sigma_K &= \operatorname{var}(X_t(K)) = \operatorname{var}\left(\sum_{s=t-K}^{t} \varepsilon_s\right) + \operatorname{var}\left(\eta_t - \eta_{t-K}\right) \\ &= KE\varepsilon_s\varepsilon_s^{\intercal} + E\left((\eta_t - \eta_{t-K})(\eta_t - \eta_{t-K})^{\intercal}\right) = K\Omega_\varepsilon + \Omega_\eta(K), \end{aligned}$$

where $\Omega_\eta(k) = \operatorname{var}\left(\eta_t - \eta_{t-k}\right) \geq 0$, $k = 1, 2, \ldots$. Therefore, $\mathcal{VR}(K) = \Sigma_1^{-1/2}\Sigma_K\Sigma_1^{-1/2}/K$ and $\mathcal{VR}d(K) = D_1^{-1/2}\Sigma_K D_1^{-1/2}/K$. The next result shows the behaviour of this variance ratio statistic in long horizons.

Theorem 3. *Suppose that the multivariate fads model (32)-(33) holds and suppose that* $\operatorname{cov}(\eta_{t+j}, \eta_t) \to 0$ *as* $j \to \infty$. *Then,* $\mathcal{VR}(\infty) = \lim_{K\to\infty} \mathcal{VR}(K) = I + \sum_{j=1}^{\infty}(R(j) + R(j)^{\intercal})$ *exists. Further suppose that* $\Omega_\eta(1) > 0$. *Then,*

$$\mathcal{VR}(\infty) < I_d$$

in the matrix partial order sense. Likewise, $\mathcal{VR}d(\infty) = \lim_{K\to\infty} \mathcal{VR}d(K)$ *exists, and*

$$\mathcal{VR}d(\infty) < Rd(0).$$

This result generalizes the existing results for the scalar fads process, which amount to $\mathcal{VR}d_{ii}(\infty) \leq Rd_{ii}(0)$ for $i = 1, \ldots, d$. In Theorem 3, we obtain stronger constraints on the off diagonal elements of $\mathcal{VR}d(\infty)$ and $\mathcal{VR}(\infty)$. Note that we also obtain $GMV(K) \to GMV(\infty) > 1/d$ as a corollary.

For the two parameter statistics we have:

$$\mathcal{VR}(\infty, L), \mathcal{VR}^{\%}(\infty, L) < I_d = \mathcal{VR}(\infty, \infty), \mathcal{VR}^{\%}(\infty, \infty) < \mathcal{VR}(K, \infty), \mathcal{VR}^{\%}(K, \infty). \quad (35)$$

Specifically, when both $K, L \to \infty$, the limit is the identity matrix. This says that if the fads model is assumed at very high frequency (consistent with intraday sampling), then the doubly long horizon statistic approaches the identity matrix. If microstructure were the cause of the misspricing, its effect would be washed out in long horizon weekly or even daily variance ratios.

We consider what happens to the long horizon sample variance ratio statistic under the fads model. We will consider the case where $K \to \infty$ as $T \to \infty$ such that $K/T \to 0$ (in contrast with the framework of Richardson and Stock (1989)). The consistency follows from the theory for the long run variance ratio, Parzen (1957), Andrews (1991), and Liu and Wu (2010). We adopt the framework of Liu and Wu (2010) and suppose that

$$X_t = R(\ldots, e_{t-1}, e_t),$$

where e_t are i.i.d random vectors of length $p \geq d$. This includes a wide range of linear and nonlinear processes for η_t, ε_t. Then define

$$\delta_t = E\left[\|(R(\ldots, e_0, \ldots, e_{t-1}, e_t) - R(\ldots, e_0', \ldots, e_{t-1}, e_t))\|\right],$$

where e_t' is an i.i.d. copy of e_t and $||.||$ denotes the Euclidean norm.

ASSUMPTION B. *The vector process X_t is stationary with finite fourth moments and weakly dependent in the sense that $\sum_{t=1}^{\infty} \delta_t < \infty$.*

THEOREM 4. *Suppose that the multivariate fads model (32)-(33) holds along with Assumption B, and suppose that $K \to \infty$ as $T \to \infty$ such that $K/T \to 0$. Then,*

$$\widehat{\mathcal{VR}}(K) \xrightarrow{P} \mathcal{VR}(\infty).$$

Likewise, $\widehat{\mathcal{VR}}d(K)$ consistently estimates $\mathcal{VR}d(\infty)$. More generally, we could obtain the limiting distribution of $\widehat{\mathcal{VR}}(K) - \mathcal{VR}(K)$ under either fixed K or K increasing asymptotics applying the methods of Liu and Wu (2010), but the limiting variance in either case is going to be very complicated.

5.3 Bubble Process

Several authors argue that the frequently observed excessive volatility in stock prices may be attributed to the presence of speculative bubbles. Blanchard and Watson (1982) and Flood and Hodrick (1986), inter alia, demonstrate in a theoretical framework that bubble components potentially generate excessive volatility. There is some debate about whether these constitute rational adjustment to fundamental pricing rules or arise from more behavioural reasons. Recently, Phillips and Yu (2010), and Phillips, Shi, and Yu (2012) have considered the following class of "bubble processes" for (log) prices p_t

$$p_t = \mu + p_{t-1} 1(t < \tau_e) + \delta_T 1(\tau_e \leq t \leq \tau_f) p_{t-1} + \left(\sum_{s=\tau_f+1}^{t} \varepsilon_s + p_{\tau_f}^*\right) 1(t > \tau_f) + \varepsilon_t 1(t \leq \tau_f), \quad (36)$$

where $p_{\tau_f}^*$ represents the restarting price after the bubble collapses at time τ_f, and $\delta_T = 1 + c/T^{\alpha}$ for $\alpha \in (0,1)$ and $c > 0$. The process is consistent with the efficient markets hypothesis during $[1, \tau_e]$ and

$[\tau_f, T]$ but has an explosive "irrational" moment in the middle. They propose econometric techniques to test for the presence of a bubble and indeed multiple bubbles. One can imagine this model also holding for a vector of asset prices caught up in the same bubble, so that ε_t is a vector of shocks, the indicator function is applied coordinatewise, and the coefficient δ_T is replaced by a diagonal matrix.

In the appendix we show that in the univariate bubble process with nontrivial bubble epoch (i.e., $(\tau_f - \tau_e)/T \to \tau_0 > 0$), that, as $T \to \infty$

$$\widehat{\mathcal{VR}}(K) \xrightarrow{P} K \tag{37}$$

for all K, so that the variance ratio statistic is greater than one for all K and gets larger with horizon. Essentially, the bubble period dominates all the sample statistics, and all return autocorrelations converge to one inside the bubble period, thereby making the ratio equal to the maximum it can achieve. In the multivariate case, the situation is more complicated, although Magdalinos (2014) has shown that in some special cases, $\lambda_{\max}(\widehat{\mathcal{VR}}(K)) \xrightarrow{P} K$.

In practice, rolling window versions of the variance ratio statistics can detect the bubble period in a similar way to the Phillips, Shi and Yu (2012) statistics (although they are not explicitly designed for this purpose and are not optimal for it). Our point here is just that these two different alternative models generate opposite predictions with regard to the variance ratio. We will check this empirically below.

5.4 Locally Stationary Alternatives

Suppose that $X_t = X_{t,T}$ can be approximated by a family of locally stationary processes $\{X_t(u), u \in [0,1]\}$, Dahlhaus (1997). For example, suppose that $X_t = \varepsilon_t + \Theta(t/T)\varepsilon_{t-1}$, where $\Theta(\cdot)$ is a matrix of smooth functions and ε_t is iid. This allows for zones of departure from the null hypothesis, say for $u \in U$, where U is a subinterval of $[0,1]$, e.g., $\Theta(u) \neq 0$ for $u \in U$. For example, during recessions the dependence structure may change and depart from efficient markets, but return to efficiency during normal times. This is consistent with the Adaptive Markets Hypothesis of Lo (2004, 2005) whereby the amount of inefficiency can change over time depending on " the number of competitors in the market, the magnitude of profit opportunities available, and the adaptability of the market participants".

Let $\widetilde{X}_t(u) = X_t(u) - EX_t(u)$ and:

$$\Sigma(u) = \mathrm{var}(X_t(u)) = E(\widetilde{X}_t(u)\widetilde{X}_t^{\intercal}(u))$$

$$D(u) = \text{diag}\left\{E(\widetilde{X}_{1t}^2(u)), \ldots, E(\widetilde{X}_{dt}^2(u))\right\}$$
$$\Gamma_u(j) = E(\widetilde{X}_t(u)\widetilde{X}_{t-j}^{\intercal}(u)).$$

The sample autocovariances converge, under some conditions, to the integrals of the autocovariances, e.g., $\widehat{\Gamma}(j) \to \int_0^1 \Gamma_u(j)du$. Then, define

$$\overline{R}(j) = \left(\int_0^1 \Sigma(u)du\right)^{-1/2} \int_0^1 \Gamma_u(j)du \left(\int_0^1 \Sigma(u)du\right)^{-1/2}.$$

It follows that under local stationarity

$$\widehat{\mathcal{VR}}(K) \xrightarrow{P} I + \sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)(\overline{R}(j) + \overline{R}(j)^{\intercal}).$$

The test will have power against some alternatives where $\Gamma_u(j) \neq 0$ for $u \in U$ and $\Gamma_u(j) = 0$ for $u \in U^c$.

5.5 Nonlinear Processes

In general, the class of statistics we consider will not have power against all nonlinear alternatives, Hong (2000). In that case, one may work with nonlinear transformations $Y_t = \tau(X_t)$ such as the quantile hit process, Han et al. (2014), and then calculate the "variance ratio" equivalent through (12)-(14). Wright (2000) has proposed variance ratios based on signs and ranks that have similar objectives.

6 Time Varying Risk Premium

It is now widely accepted that the risk premium is time varying, Mehra and Prescott (2008), in which case the tests discussed above are invalid in the sense that any rejection of the null hypothesis could be ascribed to omitting the risk premium. We investigate here how to adjust the variance ratio statistics and their critical values in this case. There are many papers that model the market risk premium and its evolution over time. One line of work specifies a parametric model for the vector of conditional means $\mu_t(\theta_0) = E(X_t|\mathcal{F}_{t-1})$, where the information set includes just past price information. For example, Engle, Lilien and Robins (1987) consider a multivariate ARCH model consistent with the conditional CAPM where the dynamic risk premium is related to the conditional covariance matrix of returns. This is appropriate for the medium frequency settings such as daily or weekly data, where only price data are available. Another line of work, associated with lower frequency macro and accounting data, involves specifying parametrically the stochastic discount function in terms of state variables like consumption. We note that generally the estimation of the risk premium parameters would affect the asymptotic distribution of the variance ratio statistics in a complicated way, and the details vary considerably according to the model adopted. We have considered two frameworks that allow for time varying risk premia but where the consequence for inference is not too onerous.[13]

We adopt a standard linear factor model for returns with constant betas but allow for time varying risk premia through the common factor dynamics. Specifically, we suppose that

$$X_{it} = \alpha_i + \beta_i^{\intercal} F_t + \varepsilon_{it}; \quad i = 1, \ldots, d, \tag{38}$$

where $\beta_i = (\beta_{1i}, \beta_{2i}, \ldots, \beta_{Pi})^{\intercal}$ is the vector of factor loadings for stock i, and $F_t = (f_{1t}, f_{2t}, \ldots, f_{Pt})^{\intercal}$ is the vector of common factors that may be observed or unobserved, and may be lagged.[14] We assume at least that the factors are uncorrelated with the idiosyncratic errors, which are themselves cross sectionally and temporally uncorrelated:

$$\begin{gathered} E(\varepsilon_{it}) = 0, \text{ for all } i, t; \quad \text{cov}(f_{jt}, \varepsilon_{is}) = 0, \text{ for all } j, i, t, s, \\ \text{cov}(\varepsilon_{it}, \varepsilon_{js}) = \begin{cases} \sigma_i^2 (< \overline{\sigma}^2 < \infty) & \text{if } i = j, t = s \\ 0 & \text{otherwise.} \end{cases} \end{gathered} \tag{39}$$

Then define the $P \times P$ time invariant covariance matrix of the factors, $\text{var}(F_t) = \Sigma_F$, and likewise the diagonal covariance matrix of the idiosyncratic errors $\text{var}(\varepsilon_t) = D_\varepsilon$. It follows that

$$\text{var}(X_t) = \beta \Sigma_F \beta^{\intercal} + D_\varepsilon,$$

$$\text{var}(X_t(K)) = \beta \text{var}(F_t(K)) \beta^{\intercal} + \text{var}(\varepsilon_t(K)),$$

where $F_t(K)$ and $\varepsilon_t(K)$ are defined similarly to $X_t(K)$. We allow the common risk factors F_t to have a time varying risk premium that we do not specify. In particular, they may be weakly dependent so

[13] In the the working paper version of this paper (HLZ, 2014) we considered an approach in which a deterministic nonparametric specification was adopted. We allowed for a slowly evolving risk premium that perhaps also varied according to the day of the week. We showed that provided the nonparametric trend functions were estimated suitably, that essentially the same standard errors could be used to conduct inference about the remaining predictability.

[14] Lo and MacKinlay (1999, chapter 9) consider a similar setting except they work with scalar lagged factors in a regression framework. Their purpose is to obtain maximally predictable portfolios based on the factor relation.

that $E\left(f_{jt}|\mathcal{F}_{t-1}\right)=\mu_j(\mathcal{F}_{t-1})$ for some unknown functions $\mu_j(.)$. This implies that the risk premium of asset i is of the form $E\left(X_{it}|\mathcal{F}_{t-1}\right)=\alpha_i+\beta_i^{\intercal}\mu(\mathcal{F}_{t-1})$ and varies over time in a potentially quite general way (that we will not model) except that it is only driven by the common risk factors. More usefully, it follows that

$$\mathrm{var}(X_t(K))/K=\beta\left(\Gamma_F(0)+\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)(\Gamma_F(j)+\Gamma_F(j)^{\intercal})\right)\beta^{\intercal}+D_{\varepsilon},$$

where $\Gamma_F(0)=\Sigma_F$ and $\Gamma_F(j)$ are the autocovariance matrices of the factor process. In this case, the variance ratio statistics we developed earlier would reject the null hypothesis, but in a rather specific way. We have

$$\begin{aligned}\mathcal{VR}(K)-I &= \mathrm{var}(X_t)^{-1/2}\left[\mathrm{var}(X_t(K))/K-\mathrm{var}(X_t)\right]\mathrm{var}(X_t)^{-1/2}\\ &= (\beta\Sigma_F\beta^{\intercal}+D_{\varepsilon})^{-1/2}\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)\beta(\Gamma_F(j)+\Gamma_F(j)^{\intercal})\beta^{\intercal}\left(\beta\Sigma_F\beta^{\intercal}+D_{\varepsilon}\right)^{-1/2}.\end{aligned}$$

This matrix is zero under the null of no predictability at all. Under the hypothesis that all the predictability is coming from the common factors, we should have that $\mathcal{VR}(K)-I$ is non zero but of rank less than or equal to P. This hypothesis could be tested under weak assumptions, specifically without specifying the factors, although it would require a complicated limit theory.

Instead, we shall suppose that the common factors are observed, e.g., the Fama French factors. We pursue an explicit regression method to obtain residuals $\widehat{\varepsilon}_{it}$ that can be tested for the hypothesis that the idiosyncratic error is uncorrelated or a martingale difference sequence

$$E\left(\varepsilon_{it}|\mathcal{F}_{t-1}\right)=0. \tag{40}$$

We estimate $\theta_i=(\alpha_i,\beta_i^{\intercal})^{\intercal}$ by the time series least squares estimator

$$\widehat{\theta}_i=(\widehat{\alpha}_i,\widehat{\beta}_i^{\intercal})^{\intercal}=\left(\sum_{s=1}^{T}G_sG_s^{\intercal}\right)^{-1}\sum_{s=1}^{T}G_sX_{is},$$

where $G_s=(1,F_s^{\intercal})^{\intercal}$. Then define the residuals $\widehat{\varepsilon}_{it}=X_{it}-\widehat{\alpha}_i-\widehat{\beta}_i^{\intercal}F_t$. We apply the variance ratio tests described above on these residuals *as if* we knew the thetas, and show that this is valid. Define $\widehat{\mathcal{VR}}_{\varepsilon+}(K)$, $\widehat{\mathcal{VR}d}_{\varepsilon+}(K)$, and $\widehat{\mathcal{VR}a}_{\varepsilon+}(K)$ as the variance ratio statistics computed with the OLS residuals. Furthermore, define $Q_{\varepsilon}(K)$, $Qd_{\varepsilon}(K)$, $Qa_{\varepsilon}(K)$ as above but with the vector of idiosyncratic errors ε_t replacing $\widetilde{X}_t$.

We now introduce new sets of assumptions required for the asymptotic theory, both of which are rather direct extensions of Assumptions A and MH* we had before.

ASSUMPTION AF.

A1. *F_t and ε_t are jointly stationary and ergodic, and are uncorrelated to each other both cross-sectionally and temporally. In particular ε_t* and *$R_t := \varepsilon_{t-j} \otimes F_t$ are Martingale Difference sequence with respect to past history of F* and ε;

A2. *The process ε_t has finite fourth moments, i.e., $\forall i, j, k, l$, $E[|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}|] < \infty$.*

A3. *The process F_t has finite second moments. $E[|F_t F_t^\intercal|] < \infty$.*

ASSUMPTION MHF*.

MH1. (i) *For all t, ε_t* satisfies *$E\varepsilon_t = 0, E[\varepsilon_t \varepsilon_t^\intercal] = 0$ for all $j \neq 0$;* (ii) *for all t, s with $s \neq t$ and all $j, k = 1, \ldots, K$, $E[\varepsilon_t \varepsilon_{t-j}^\intercal \otimes \varepsilon_s \varepsilon_{s-k}^\intercal] = 0$;*

MH2. *$Z_t := (F_t^\intercal, \varepsilon_t^\intercal)$ is α-mixing with coefficient $\alpha(m)$ of size $r/(r-1)$, where $r > 1$, such that for all t and for any $j \geq 0$, there exists some $\delta > 0$ for which $\sup_t E|Z_{it}Z_{k,t-j}|^{2(r+\delta)} < C < \infty$ for all $i, k = 1, \ldots, d$;*

MH3. *For all j, k, the following limits exist: $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[\varepsilon_t\varepsilon_t^\intercal] =: \Sigma < \infty$ and $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E[\varepsilon_{t-j}^\intercal \varepsilon_{t-k} \otimes \varepsilon_t \varepsilon_t^\intercal] =: \Xi_{jk} < \infty$*

THEOREM 5. *Suppose that either Assumption AF or MHF* holds. Then, as $T \to \infty$:*

$$\begin{aligned}
\sqrt{T}\mathrm{vec}\left(\widehat{\mathcal{VR}}_{\varepsilon+}(K) - I_d\right) &\implies N\left(0, Q_\varepsilon(K)\right) \\
\sqrt{T}\mathrm{vec}\left(\widehat{\mathcal{VR}d}_{\varepsilon+}(K) - \widehat{Rd}(0)\right) &\implies N\left(0, Qd_\varepsilon(K)\right) \\
\sqrt{T}\mathrm{vec}\left(\widehat{\mathcal{VR}a}_{\varepsilon+}(K) - I_d\right) &\implies N\left(0, Qa_\varepsilon(K)\right).
\end{aligned}$$

Let

$$\widehat{Q}_\varepsilon(K) = \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_{j,K}c_{k,K}\left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2}\right)\widehat{\Xi}_{jk}\left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2}\right).$$

COROLLARY 3. *Suppose that either Assumption AF or MHF* holds, then the estimator $\widehat{Q}_\varepsilon(K)$ is weakly consistent for $Q_\varepsilon(K)$ (likewise $\widehat{Qd}_\varepsilon(K)$ and $\widehat{Qa}_\varepsilon(K)$ are weakly consistent for $Qd_\varepsilon(K)$ and $Qa_\varepsilon(K)$); i.e.*

$$\widehat{Q}_\varepsilon(K) \xrightarrow{P} Q_\varepsilon(K). \tag{41}$$

In fact, we can allow the betas to vary slowly over time and to vary by day of the week or recession/boom categorization using the techniques developed in (HLZ, 2014). In practice, working with subperiods and weekly data goes some way to addressing these issues. In any case, there are some arguments that return predictability is primarily driven by time-series variation in risk premiums: Ferson and Korajczyk (1995), for example, argue that less than 1% of the predictable variation in returns to changing conditional betas.

7 Large Dimensional Data

We briefly consider some issues that arise when the dimensions d are large. In this case, the covariance matrices Σ and $\Sigma(K)$ may be ill conditioned, and so forming the ratio (11) may not be practically feasible or theoretically valid; likewise for any functions derived thereof such as the eigenvalues. The diagonal variance ratio matrix and simple univariate quantities derived from it like $CS(K)$ may fare better in this situation, since the marginal variances should be bounded away from zero. We remark that Castura, Litzenberger, Gorelick, and Dwivedi (2010) report the average variance ratio of the Russell 1000 and Russell 2000 stocks, which amounts to $\sum_{i=1}^{d} \widehat{\mathcal{VR}d}_{ii}(K)/d$. They do not report standard errors for this quantity, perhaps on the grounds that d is large (since $d = 3000$). However, when the individual stocks are contemporaneously correlated, which they typically are[15], the averaging will not reduce the order of magnitude of the standard error. Specifically, under the iid assumption, the correlation between $\widehat{\mathcal{VR}d}_{ii}(K)$ and $\widehat{\mathcal{VR}d}_{jj}(K)$ will be proportional to ρ_{ij}^2, where ρ_{ij} is the contemporaneous correlation between the returns on stock i and stock j. Under a factor model type assumption, it is straightforward to calculate the standard errors for $\sum_{i=1}^{d} \widehat{\mathcal{VR}d}_{ii}(K)/d$ in the large d, T case. However, for nonlinear functions of $\mathcal{VR}d(K)$ such as its eigenvalues, or for quantities derived from $\mathcal{VR}(K)$, the large d theory is more complicated.

An alternative strategy in the large d case may be to calculate scalar ratios from the matrix scaling law $\Sigma(K) = K\Sigma$. Specifically, if we calculate the ratio of the eigenvalues rather than the eigenvalues of the ratio, we may obtain better performance for moderate sized d by only looking at the upper ends of the marginal eigenvalue distributions. However, when d is comparable with T, one must use some sparsity structure or shrinkage method to obtain reasonable performance for complicated nonlinear functions of the covariance matrices.

[15] Although for very high frequency data, the correlation maybe quite small, Sheppard (2013).

8 Application

We apply our methodology to U.S. stock return data. In particular, we use weekly size-sorted equal-weighted portfolio returns from the Center for Research in Security Prices (CRSP) from 06/07/1962 to 27/12/2013.[16] Essentially the same data were used in Lo and MacKinlay (1988) and Campbell, Lo and Mackinlay (1997), which allows us to make comparison with their results, and to extend it to the more recent period.[17]

8.1 Short to Medium Horizon

8.1.1 Evidence on Linear Predictability

According to the results of Theorem 1 and Corollary 1, we report the test statistics

$$Zd(K) = \sqrt{T}\left(\theta_l^\intercal \widehat{Q}d(K)\theta_l\right)^{-1/2}\left[\theta_l^\intercal \mathrm{vec}\left(\widehat{VRd}_+(K) - \widehat{Rd}\,(0)\right)\right] \Longrightarrow N(0,1),$$

where θ_l is a $d^2 \times 1$ vector. $Zd_{LM}(K)$ and $Zd_{iid}(K)$ are defined similarly but using $\widehat{Qd}_{LM}(K)$ and $\widehat{Qd}_{iid}(K)$ respectively. In the following, we use $Zd(K)$, $Zd_{LM}(K)$ and $Zd_{iid}(K)$ statistics to test some specific linear function of vec $\left(\widehat{\mathcal{VR}d}_+(K) - \widehat{Rd}\,(0)\right)$ matrix.

We first test for the absence of serial correlation in each of three weekly size-sorted equal-weighted portfolio returns (smallest quintile, central quintile, and largest quintile). The null hypothesis is $[\mathcal{VR}d_+(K)]_{ll} = 1,\ l = 1, \ldots, d$ where $d = 3$ and $K = 2, 4, 8, 16$. We use $Zd(K)$, $Zd_{LM}(K)$ and $Zd_{iid}(K)$ statistics by setting θ_l as a vector that is 1 at the $((l-1)(d+1)+1)^{th}$ entry and 0 otherwise.

[16] The data are obtained from Kenneth French's Data Library. It was created by CMPT_ME_RETS using the 2013/12 CRSP database. It contains value- and equal-weighted returns for portfolios in five size quintiles. We compute weekly returns of portfolios by adding up Monday to Friday's daily returns.

[17] In general we compute variance ratio statistics over a given window, denoted W, that has a time span T_W. This allows the mean return or even the factor betas to vary with the window. In the working paper version (HLZ, 2014) we considered a framework where the window size was small relative to the whole available sample and so $T_W/T \to 0$. We invoked theory for kernel smoothing methods to give a theoretical treatment. We shall not pursue this here but our framework does allow for windows to vary but with sample sizes T_W proportional to the full sample size. We may also allow for "day of the week" effects quite simply in our existing framework. To be specific consider the two parameter statistic $\widehat{\mathcal{VR}}(K, L)$. We may compute this using returns computed with different starting points indexed by Δ with $\Delta = 1, \ldots, L$. In general then we obtain statistics that may be denoted $\widehat{\mathcal{VR}}_{W,\Delta}(K, L)$ with W denoting the particular window and Δ denoting the "day of the week". Implicitly we are allowing the mean return vector to vary with W, Δ, i.e., $\mu = \mu(W, \Delta)$, so that our procedures are robust to variation in the mean across subperiods and days of the week.

To compare with the results reported in Campbell, Lo and Mackinlay (1997, P71, Table 2.6), we divide the whole sample to three subsamples: 62:07:06-78:09:29 (848 weeks), 78:10:06-94:12:23 (847 weeks) and 94:12:30-13:12:27 (992 weeks). Table 1-A reports the results for the portfolio of small-size firms (smallest CRSP quintile), Table 1-B reports the results for the portfolio of medium-size firms (central CRSP quintile), and Table 1-C reports the results for the portfolio of large-size firms (largest CRSP quintile).

Table 1-A: Variance ratios for weekly small-size portfolio returns

		Lags			
Sample period	# of obs	$K=2$	$K=4$	$K=8$	$K=16$
62:07:06—78:09:29	848	**1.43**	**1.93**	**2.46**	**2.77**
		$(8.82)^*$	$(8.49)^*$	$(7.00)^*$	$(5.59)^*$
		$(8.82)^*$	$(10.81)^*$	$(11.00)^*$	$(9.33)^*$
		$(12.46)^*$	$(14.47)^*$	$(14.39)^*$	$(11.70)^*$
78:10:06—94:12:23	847	**1.43**	**1.98**	**2.65**	**3.19**
		$(6.20)^*$	$(7.07)^*$	$(7.37)^*$	$(6.48)^*$
		$(6.20)^*$	$(8.62)^*$	$(10.69)^*$	$(10.70)^*$
		$(12.52)^*$	$(15.25)^*$	$(16.26)^*$	$(14.45)^*$
94:12:30—13:12:27	992	**1.21**	**1.47**	**1.7**	**1.82**
		$(3.30)^*$	$(3.58)^*$	$(3.35)^*$	$(2.50)^*$
		$(3.30)^*$	$(4.13)^*$	$(4.15)^*$	$(3.44)^*$
		$(6.59)^*$	$(7.91)^*$	$(7.43)^*$	$(5.82)^*$

Table 1-B: Variance ratios for weekly medium-size portfolio returns

Sample period	# of obs	Lags			
		$K=2$	$K=4$	$K=8$	$K=16$
62:07:06—78:09:29	848	**1.25**	**1.54**	**1.79**	**1.91**
		$(5.41)^*$	$(5.55)^*$	$(4.35)^*$	$(3.22)^*$
		$(5.41)^*$	$(6.41)^*$	$(5.93)^*$	$(4.69)^*$
		$(7.37)^*$	$(8.42)^*$	$(7.78)^*$	$(6.05)^*$
78:10:06—94:12:23	847	**1.20**	**1.37**	**1.54**	**1.56**
		$(3.29)^*$	$(3.35)^*$	$(3.18)^*$	$(2.14)^*$
		$(3.29)^*$	$(3.72)^*$	$(3.90)^*$	$(2.93)^*$
		$(5.73)^*$	$(5.80)^*$	$(5.36)^*$	$(3.74)^*$
94:12:30—13:12:27	992	**0.99**	**1.05**	**1.02**	**0.89**
		(-0.02)	(0.38)	(0.10)	(-0.38)
		(-0.02)	(0.43)	(0.11)	(-0.48)
		(-0.04)	(0.78)	(0.20)	(-0.78)

Table 1-C: Variance ratios for weekly large-size portfolio returns

Sample period	# of obs	Lags			
		$K=2$	$K=4$	$K=8$	$K=16$
62:07:06—78:09:29	848	**1.05**	**1.15**	**1.21**	**1.19**
		(1.05)	(1.64)	(1.23)	(0.68)
		(1.05)	(1.54)	(1.32)	(0.84)
		(1.59)	$(2.33)^*$	$(2.06)^*$	(1.29)
78:10:06—94:12:23	847	**1.03**	**1.06**	**1.08**	**1.01**
		(0.63)	(0.61)	(0.54)	(0.03)
		(0.63)	(0.65)	(0.59)	(0.04)
		(0.95)	(0.91)	(0.75)	(0.04)
94:12:30—13:12:27	992	**0.93**	**0.94**	**0.89**	**0.81**
		(-0.99)	(-0.46)	(-0.53)	(-0.62)
		(-0.99)	(-0.52)	(-0.61)	(-0.77)
		$(-2.05)^*$	(-1.01)	(-1.14)	(-1.35)

Variance ratios reported in the main rows are the diagonal elements of $\widehat{VRd}_+(K)$. Test statistics ($Zd(K)$, $Zd_{LM}(K)$ and $Zd_{iid}(K)$) in parentheses marked with asterisks indicate that the variance ratios are statistically different from one at 5% level of significance.

The results for the earlier sample periods are broadly similar to those in Campbell, Lo and Mackinlay (1997, P71, Table 2.6) who compared the period 1962-1978 with the period 1978-1994 as well as the combined period 1962-1994. The variance ratios are greater than one and deviate further from one as the horizon lengthens. The departure from the random walk model is strongly statistically significant for the small and medium sized firms, but not so for the larger firms.

When we turn to the later period 1994-2013 we see that the variance ratios all reduce in magnitude. For the smallest stocks the statistics are still significantly greater than one and increase with horizon. However, they are much closer to one at all horizons and the statistical significance of the departures is substantially reduced. For medium sized firms, the variance ratios are reduced. They are in some cases below one and also no longer increasing with horizon. They are insignificantly different from one. For the largest firms, the ratios are all below one but are statistically inseparable from this value. One interpretation of these results is that the stock market (at the level of these portfolios) has become closer to efficient benchmark. This is consistent with the evidence presented in Castura, Litzenberger, Gorelick, and Dwivedi (2010) for high frequency stock returns. The biggest improvements seem to come in the most recent period, especially for the small stocks.

The test statistics change quite a lot depending on which covariance matrix $\widehat{Q}(K)$, $\widehat{Q}_{LM}(K)$ or $\widehat{Q}_{iid}(K)$ one uses, and in some cases this could affect ones conclusions, for instance, for large-size portfolio, test statistics based on $\widehat{Q}_{iid}(K)$ in some periods are statistically significant. Our sample size is relatively large, and for smaller samples, the differences could bite.

We may wish to test whether the variance ratio has "improved" significantly from one period (A) to the next (B). For this purpose we consider the statistic

$$\tau_{AB} = \left(\widehat{\mathcal{VR}d}_{+}^{A}(K) - \widehat{Rd}^{A}(0)\right) - \left(\widehat{\mathcal{VR}d}_{+}^{B}(K) - \widehat{Rd}^{B}(0)\right), \tag{42}$$

where $\widehat{\mathcal{VR}d}_{+}^{j}(K)$ and $\widehat{Rd}^{j}(0)$ denote the variance ratio statistic and the correlation matrix computed in period $j = A, B$. Under the martingale null hypothesis, the two subsample variance ratio statistics are asymptotically independent and the asymptotic variance of the $\sqrt{T}vec(\tau_{AB})$ is just the sum of the subperiod covariance matrices $Qd^{A}(K) + Qd^{B}(K)$. For example, we may consider the single element of statistic $[\widehat{\mathcal{VR}d}^{A}(K)]_{ll} - [\widehat{\mathcal{VR}d}^{B}(K)]_{ll}$ and compare it with the square root of the sum of the square of the associated standard errors to obtain a "test" of the hypothesis that the efficiency has improved across subperiods. For example, in Table 1-A, the change of the variance ratio for small stocks of 1.43 in the period 78:10:06-94:12:23 to 1.21 during 94:12:30-13:12:27 is statistically significant according to this calculation.

We have carried out this calculation using the Friday to Friday weekly returns as the base series, but we have also done it for other days of the week and for the two parameter statistic. Qualitatively the results are similar. Results are available from the authors upon request. We present here the two parameter statistics for comparison, i.e., $\widehat{\mathcal{VR}d}_{+}(5K,5)$ for $K = 2, 4, 8$, and 16 using daily returns of these three size sorted portfolios. We test the null of $\left[\widehat{\mathcal{VR}d}_{+}(K5,5) - \widehat{Rd}\,(0)\right]_{ll} = 0$. The test statistics $Zd(5K,5)$ are defined similarly as $Zd(K)$ but using $\widehat{\mathcal{VR}d}_{+}(5K,5)$ and $\widehat{Qd}\,(5K,5)$. Results are reported in Table 2.

Table 2: Two parameter variance ratios for daily size sorted portfolio returns

		Lags			
Sample period	# of obs	$K=2$	$K=4$	$K=8$	$K=16$
A. small-size portfolio					
62:07:02—78:09:29	4240	**1.34**	**1.77**	**2.24**	**2.51**
		(8.58)*	(8.46)*	(7.18)*	(5.74)*
78:10:02—94:12:23	4235	**1.39**	**1.88**	**2.48**	**2.97**
		(5.82)*	(6.30)*	(7.26)*	(7.36)*
94:12:26—13:12:27	4960	**1.20**	**1.45**	**1.69**	**1.80**
		(5.70)*	(6.12)*	(5.62)*	(4.30)*
B. medium-size portfolio					
62:07:02—78:09:29	4240	**1.21**	**1.45**	**1.68**	**1.79**
		(6.48)*	(6.04)*	(4.88)*	(3.82)
78:10:02—94:12:23	4235	**1.16**	**1.28**	**1.40**	**1.41**
		(2.86)*	(2.48)*	(2.52)*	(2.00)*
94:12:26—13:12:27	4960	**0.98**	**1.01**	**0.99**	**0.88**
		(−0.84)	(0.12)	(−0.06)	(−0.92)
C. large-size portfolio					
62:07:02—78:09:29	4240	**1.06**	**1.13**	**1.18**	**1.17**
		(2.06)*	(2.06)*	(1.66)	(1.04)
78:10:02—94:12:23	4235	**1.00**	**0.98**	**0.96**	**0.89**
		(0.12)	(−0.26)	(−0.40)	(−0.842)
94:12:26—13:12:27	4960	**0.91**	**0.88**	**0.86**	**0.80**
		(−3.12)*	(−2.04)*	(−1.54)	(−1.46)

Two parameter variance ratios reported in the main rows are the diagonal elements of $\widehat{VRd}_{+}(5K,5)$. Test statistics $Zd(5K, 5)$ in parentheses marked with asterisks indicate that the variance ratios are statistically different from one at 5% level of significance.

The results of two-parameter variance ratio test are similar to the single parameter ones, but the efficiency is improved by using two-parameter variance ratio tests. On the other hand the pooled two parameter statistic effectively imposes the same mean across each day of the week and so is less robust to such seasonal patterns. The results are similar to Table 1 although in some cases, the test statistics become marginally significant in the later period.

8.1.2 Lead Lag Relationships

We next test zero cross-autocorrelation (no lead-lag relationship) between returns of different size portfolios. Based on the multivariate ratio statistic $\mathcal{VR}d_+(K)$, we test the hypothesis that $[\mathcal{VR}d_+(K) - Rd(0)]_{lh} = 0$, for $l, h = 1, 2, 3,\ l \neq h$, using $Zd(K)$ statistics. Results are reported in Table 3.

Table 3: Lead-lag patterns between weekly size-sorted portfolio returns

Lags	Sample period	$\widehat{\mathcal{VR}d}_+(K) - \widehat{Rd}(0)$ From	To small	medium	large
$K = 2$	62:07:06—94:12:23	small		**0.20** (5.74)*	**0.04** (1.15)
		medium	**0.39** (9.61)*		**0.05** (1.47)
		large	**0.32** (8.21)*	**0.21** (5.42)*	
	94:12:30—13:12:27	small		−**0.02** (−0.33)	−**0.07** (−1.01)
		medium	**0.20** (3.32)*		−**0.05** (−0.83)
		large	**0.17** (2.74)*	−**0.01** (−0.08)	
$K = 4$	62:07:06—94:12:23	small		**0.406** (5.42)*	**0.08** (1.14)
		medium	**0.84** (10.39)*		**0.12** (1.756)
		large	**0.67** (9.03)*	**0.41** (5.75)*	
	94:12:30—13:12:27	small		−**0.00** (−0.00)	−**0.09** (−0.63)
		medium	**0.43** (3.54)*		−**0.05** (−0.38)
		large	**0.34** (2.93)*	**0.04** (0.38)	
$K = 8$	62:07:06—94:12:23	small		**0.57** (4.11)*	**0.10** (0.73)
		medium	**1.38** (10.21)*		0.18 (1.53)
		large	**1.07** (9.29)*	**0.59** (5.24)*	
	94:12:30—13:12:27	small		−**0.05** (−0.25)	−**0.16** (−0.72)
		medium	**0.60** (3.28)*		−**0.13** (−0.61)
		large	**0.51** (2.81)*	**0.05** (0.27)	
$K = 16$	62:07:06—94:12:23	small		**0.54** (2.39)*	−**0.03** (−0.11)
		medium	**1.77** (9.11)*		**0.13** (0.68)
		large	**1.36** (8.42)*	**0.64** (3.80)*	
	94:12:30—13:12:27	small		−**0.21** (−0.62)	−**0.28** (−0.83)
		medium	**0.67** (2.45)*		−**0.26** (−0.86)
		large	**0.61** (2.22)*	−**0.03** (−0.10)	

The off-diagonal elements of $\widehat{VRd}_+(k) - \widehat{Rd}(0)$ are reported. Test statistics marked with asterisks indicate that null hypothesis is rejected at 5% level of significance.

The results suggest there are strong lead-lag relationships, where medium and large firms lead and small firms lag for all horizons for both sample periods, although the evidence attenuates in the later period, especially at the longer horizon. Nevertheless, there is statistical significance at the 5% level in all such cases. The sign of these terms are all positive and increase with horizon. Also, the size of the coefficients decreases substantially in the later sample period. The evidence is weaker for cross-autocorrelation between current returns of medium sized firms and past returns of small and large ones. We do find that there is evidence of such relationships in the earlier sample period. However, in the later period none of these effects is significant. Finally, with regard to cross-autocorrelation between current returns of large firms and past returns of small and medium sized ones, in no period do we find evidence of this. These results may be interpreted as being consistent with the explanations given in Campbell, Lo and Mackinlay (1997). This is also inconsistent with the random walk hypothesis, but the declining statistical significance may be consistent with improvements in the efficiency of these markets. This test is related to the Granger noncausality test proposed in Pierce and Haugh (1977), where the series are prewhitened before testing zero cross-autocorrelation.

We also check if the lead-lag patterns are asymmetric. We test the null hypotheses that $[\mathcal{VR}d_+(K) - Rd(0)]_{lh} - [\mathcal{VR}d_+(K) - Rd(0)]_{hl} = 0$, for $l, h = 1, 2, 3,\ l > h$, using $Zd(K)$ statistics. Results are reported in Table 4.

Table 4: Asymmetry of lead-lag patterns

		$\left[\widehat{\mathcal{VR}d}_+(K) - \widehat{Rd}(0)\right]_{lh} - \left[\widehat{\mathcal{VR}d}_+(K) - \widehat{Rd}(0)\right]_{hl}$		
Lags	Sample period	$(S \to M) - (M \to S)$	$(S \to L) - (L \to S)$	$(M \to L) - (L \to M)$
$K = 2$	62:07:06—94:12:23	$-\mathbf{0.19}$ $(-8.75)^*$	$-\mathbf{0.28}$ $(-8.58)^*$	$-\mathbf{0.16}$ $(-8.10)^*$
	94:12:30—13:12:27	$-\mathbf{0.22}$ $(-6.62)^*$	$-\mathbf{0.23}$ $(-6.38)^*$	$-\mathbf{0.05}$ $(-2.31)^*$
$K = 4$	62:07:06—94:12:23	$-\mathbf{0.44}$ $(-9.63)^*$	$-\mathbf{0.59}$ $(-8.68)^*$	$-\mathbf{0.29}$ $(-7.46)^*$
	94:12:30—13:12:27	$-\mathbf{0.43}$ $(-7.15)^*$	$-\mathbf{0.43}$ $(-6.32)^*$	$-\mathbf{0.09}$ $(-2.37)^*$
$K = 8$	62:07:06—94:12:23	$-\mathbf{0.81}$ $(-10.58)^*$	$-\mathbf{0.97}$ $(-8.98)^*$	$-\mathbf{0.40}$ $(-7.02)^*$
	94:12:30—13:12:27	$-\mathbf{0.68}$ $(-7.19)^*$	$-\mathbf{0.67}$ $(-5.79)^*$	$-\mathbf{0.17}$ $(-3.00)^*$
$K = 16$	62:07:06—94:12:23	$-\mathbf{1.23}$ $(-10.16)^*$	$-\mathbf{1.38}$ $(-8.18)^*$	$-\mathbf{0.51}$ $(-6.05)^*$
	94:12:30—13:12:27	$-\mathbf{0.88}$ $(-6.26)^*$	$-\mathbf{0.89}$ $(-5.27)^*$	$-\mathbf{0.23}$ $(-3.03)^*$

S is portfolio of small firms, M is portfolio of medium firms, and L is portfolio of large firms. Test statistics marked with asterisks indicate that the lead-lag relationship is statistically asymmetric at 5% level of significance.

These results can be compared with Campbell, Lo and Mackinlay (1997, P71, Table 2.9) who look at the asymmetry of the cross-autocorrelation matrices. We find the same direction of asymmetry consistent with their results. The statistical significance does decline in the second period, but is still quite strong.

8.1.3 Multivariate Tests

The above univariate variance ratio tests gave evidence of predictability in smaller size portfolios. We next test for the absence of serial correlation for the vector of multiple size sorted portfolios, based on univariate parameters derived from $\mathcal{VR}(K)$ and $\mathcal{VR}d(K)$. Specifically, we consider $CS(K)$, $GMV(K)$, and $\pi(K)$, as well as the trace and determinant of these matrices. We consider the following test statistics whose distribution theories follow directly from Theorem 1 and Corollary 1:

$$Z_{CS}(K) = \sqrt{T}\left(\widehat{Q}_{CS}(K)\right)^{-1/2}\left[\widehat{CS}(K) - \widehat{CS}(1)\right] \Rightarrow N(0,1)$$

$$Z_{GMV}(K) = \sqrt{T}\left(\widehat{Q}_{GMV}(K)\right)^{-1/2}\left[\widehat{GMV}(K) - \frac{1}{3}\right] \Rightarrow N(0,1)$$

$$Z_{\pi}(K) = \sqrt{T}\left(\widehat{Q}_{\pi}(K)\right)^{-1/2}\widehat{\pi}(K) \Rightarrow N(0,1)$$

$$Z_{tr}(K) = \sqrt{T}\left(\widehat{Q}_{tr}(K)\right)^{-1/2}\left[\mathrm{tr}\left(\widehat{\mathcal{VR}}(K)\right) - 3\right] \Rightarrow N(0,1)$$

$$Z_{dt}(K) = \sqrt{T}\left(\widehat{Q}_{dt}(K)\right)^{-1/2}\left[\det(\widehat{\mathcal{VR}}(K)) - 1\right] \Rightarrow N(0,1)$$

$$Z_F(K) = T\mathrm{vech}\left(\widehat{\mathcal{VR}}(K) - I\right)^{\intercal}\widehat{S}(K)^{-1}\mathrm{vech}\left(\widehat{\mathcal{VR}}(K) - I\right) \Longrightarrow \chi^2(d(d+1)/2),$$

where: $\widehat{Q}_{CS}(K) = \theta_l^{\intercal}\widehat{Qd}(K)\,\theta_l$ and θ_l is a $d^2 \times 1$ vector that is 0 at the $((l-1)(d+1)+1)^{th}$ entries $(l = 1, \ldots, d)$ and 1 at the other entries; $\widehat{Q}_{GMV}(K) = \frac{i^{\intercal}\widehat{S}(K)i}{d^4}$; $\widehat{Q}_{\pi}(K) = c^{\intercal}\widehat{Qd}(K)c$, where c is a vector that is $(1-d)/(d^2(K-1))$ at $((l-1)(d+1)+1)^{th}$ entries $(l = 1, \ldots, d)$, and is $1/(d^2(K-1))$ at other entries; $\widehat{Q}_{tr}(K) = \delta^{\intercal}\widehat{S}(K)\,\delta = \widehat{Q}_{dt}(K)$, where $\delta = \mathrm{vech}(I_d)$.

Test results based on these statistics are reported in the following table.

Table 5: Multivariate variance ratio tests for weekly size sorted portfolio returns

	Lags			
	$K = 2$	$K = 4$	$K = 8$	$K = 16$
First period: 62:07:06-78:09:29				
$\widehat{CS}(K) - \widehat{CS}(1)$	**0.21**	**0.46**	**0.69**	**0.81**
	$(5.04)^*$	$(5.23)^*$	$(4.15)^*$	$(3.09)^*$
$\widehat{GMV}(K)$	**0.39**	**0.42**	**0.43**	**0.41**
	$(4.30)^*$	$(3.53)^*$	$(2.08)^*$	(1.01)
$\widehat{\pi}(K)$	**0.0209**	**0.0180**	**0.0124**	**0.0065**
	$(5.20)^*$	$(7.10)^*$	$(6.59)^*$	$(5.01)^*$
$\mathrm{tr}(\widehat{VR}(K))$	**3.61**	**4.16**	**5.22**	**5.44**
	$(6.59)^*$	$(7.79)^*$	$(6.89)^*$	$(4.90)^*$
$\det(\widehat{VR}(K))$	**1.62**	**2.67**	**3.61**	**3.57**
	$(6.72)^*$	$(8.95)^*$	$(8.10)^*$	$(5.15)^*$
$Z_F(K)$	**128.51***	**122.06***	**86.39***	**52.06***
Second period: 78:10:06-94:12:23				
$\widehat{CS}(K) - \widehat{CS}(1)$	**0.19**	**0.38**	**0.59**	**0.65**
	$(3.49)^*$	$(3.72)^*$	$(3.68)^*$	$(2.64)^*$
$\widehat{GMV}(K)$	**0.39**	**0.42**	**0.41**	**0.37**
	$(4.24)^*$	$(3.19)^*$	(1.87)	(0.49)
$\widehat{\pi}(K)$	**0.0210**	**0.0197**	**0.0162**	**0.0119**
	$(4.05)^*$	$(5.99)^*$	$(7.17)^*$	$(6.94)^*$
$\mathrm{tr}(\widehat{VR}(K))$	**3.46**	**4.27**	**5.33**	**6.45**
	$(5.08)^*$	$(7.31)^*$	$(8.06)^*$	$(7.57)^*$
$\det(\widehat{VR}(K))$	**1.37**	**1.94**	**2.48**	**2.82**
	$(4.03)^*$	$(5.38)^*$	$(5.11)^*$	$(3.99)^*$
$Z_F(K)$	**114.27***	**124.62***	**123.80***	**103.19***

Third period: 94:12:30-13:12:27

$\widehat{CS}(K) - \widehat{CS}(1)$	**0.04**	**0.11**	**0.14**	**0.08**
	(0.63)	(0.91)	(0.71)	(0.29)
$\widehat{GMV}(K)$	**0.34**	**0.35**	**0.33**	**0.27**
	(0.42)	(0.47)	(-0.14)	(-0.77)
$\widehat{\pi}(K)$	**0.0067**	**0.0090**	**0.0065**	**0.0039**
	$(2.19)^*$	$(3.89)^*$	$(3.36)^*$	$(2.53)^*$
$\mathrm{tr}(\widehat{VR}(K))$	**3.09**	**3.46**	**3.79**	**4.08**
	(0.87)	$(2.30)^*$	$(2.36)^*$	$(2.03)^*$
$\det(\widehat{VR}(K))$	**1.03**	**1.28**	**1.38**	**1.36**
	(0.31)	(1.39)	(1.12)	(0.69)
$Z_F(K)$	**67.28***	**73.23***	**61.90***	**48.20***

The estimates of statistics are reported in the main rows. Test statistics in parentheses marked with asterisks indicate statistically significant at 5% level. $Z_F(K)$ is marked with asterisks if it is larger than 12.592, the 5% critical value of $\chi^2(6)$.

We next check whether our results are driven by the choice of subsamples, which we have chosen to match the choices made by CLM for the purpose of replication and comparison. We carry out a rolling window analysis with a (trailing) window of 500 weeks from the beginning of the sample to the end. Below we show the time series of (standard normal) test statistics $Z_{CS}(K), Z_{GMV}(K)$ and $Z_\pi(K)$ for $K = 4$. This shows that for GMV and CS the sustained decline in statistical significance happened in the decade ending in 2008, although there was an earlier dip in significance in the decade ending in 1999. The profits measure π has shown a slower but equally sustained drop in statistical significance. There are some sudden jumps (both up and down) to the level of this statistic, which may be a cause for concern in practice.

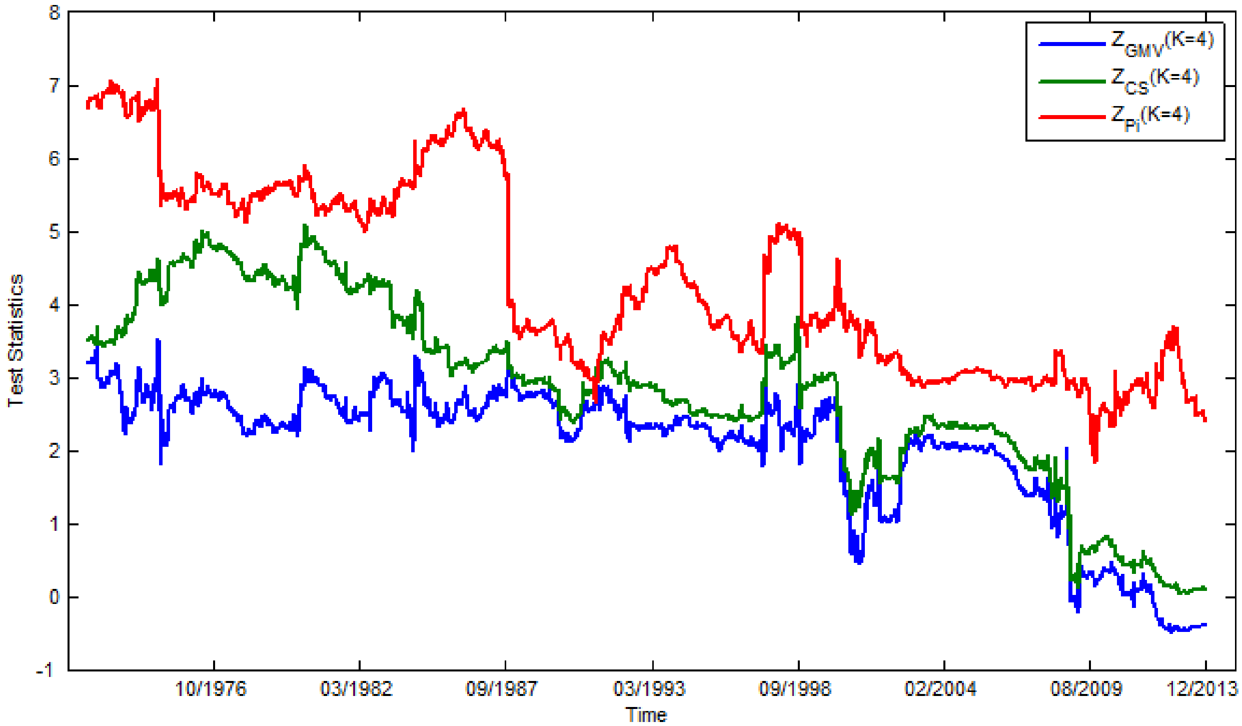

Figure 1: Trends of test statistics

8.1.4 Time Varying Risk Premium

In this section we consider whether the above results are explicable as coming from a time varying systematic risk factor. We use ten size sorted portfolios to run the following Fama and French's 3-factor regression model,

$$X_{lt} = \alpha_i + \beta_{1,l}(R_{mt} - R_{ft}) + \beta_{2,l}SMB_t + \beta_{3,l}HML_t + \varepsilon_{lt},$$

where X_{lt} is l^{th} size sorted portfolio returns, $R_{mt} - R_{ft}$ is market premium, SMB_t is small size premium, and HML_t is value premium. We then apply the OLS residuals to calculate the variance ratio statistics, based on which we test the predictability in residuals. The results of $Z_F(K)$ statistics are reported in the following table. To compare the predictability of stock returns before and after the factor model, we also report the $Z_F(K)$ statistics for the constant mean adjustment case.

Table 6: Tests based on $Z_F(K)$ statistics (10 size-portfolios)

			Lags			
Sample period	# of obs	Risk Premium	$K = 2$	$K = 4$	$K = 8$	$K = 16$
62:07:06-78:09:29	848	constant mean	207.95*	201.88*	156.35*	126.60*
		factor model	124.35*	138.12*	109.63*	89.16*
78:10:06-94:12:23	847	constant mean	223.17*	229.89*	227.92*	214.77*
		factor model	205.07*	237.23*	241.83*	234.15*
94:12:30-13:12:27	992	constant mean	127.04*	140.70*	128.43*	109.81*
		factor model	97.03*	116.47*	113.11*	96.84*

Test statistics is marked with asterisks if it is larger than 82.267, the 1% critical value of $\chi^2(55)$.

The results show that while the factor model reduces the level of the test statistic, it remains strongly significant, suggesting that the time series predictability in stock returns cannot be captured purely by a time varying risk premium in the common risk factors.

We also look at the quadratic form based on only the diagonal elements of $\widehat{\mathcal{VR}d}_+(K)$ and a quadratic form based on only the off-diagonal elements of $\widehat{\mathcal{VR}d}_+(K)$

$$ed(K) = \text{diag}\left(\widehat{\mathcal{VR}d}_+(K) - \widehat{Rd}\,(0)\right) \; ; \; eoff(K) = \text{offdiag}\left(\widehat{\mathcal{VR}d}_+(K) - \widehat{Rd}\,(0)\right),$$

where diag is the operator to select diagonal elements and offdiag is the operator to select all off-diagonal elements. Under the null, we have $ed(K) = 0$ and $eoff(K) = 0$. The test statistics are defined as

$$Z_{F_1}(K) = T \cdot ed(K)^\intercal \cdot \alpha_1^\intercal \widehat{Q}d(K)^{-1}\alpha_1 \cdot ed(K) \Longrightarrow \chi^2(d)$$

$$Z_{F_2}(K) = T \cdot eoff(K)^\intercal \cdot \alpha_2^\intercal \widehat{Q}d(K)^{-1}\alpha_2 \cdot eoff(K) \Longrightarrow \chi^2(d(d-1))$$

where α_1 is a $d^2 \times d$ matrix whose l^{th} column is 1 at the $((l-1)(d+1)+1)^{th}$ entry and 0 at the other entries. and α_2 is a $d^2 \times (d^2 - d)$ matrix which is obtained by deleting $((l-1)(d+1)+1)^{th}$ columns from I_{d^2} matrix. The results of $Z_{F_1}(K)$ and $Z_{F_2}(K)$ statistics based on constant mean and factor model adjustment are reported in the following table.

Table 7: Tests based on $Z_{F_1}(K)$ and $Z_{F_2}(K)$ statistics (10 size-portfolios)

Sample period	# of obs	Statistics	Risk Premium	Lags $K=2$	$K=4$	$K=8$	$K=16$
62:07:06-78:09:29	848	$Z_{F_1}(K)$	constant mean	113.90*	108.82*	85.21*	66.17*
			factor model	45.22*	47.80*	35.81*	22.31*
		$Z_{F_2}(K)$	constant mean	233.78*	245.55*	235.64*	214.51*
			factor model	141.52*	161.35*	129.92*	110.44*
78:10:06-94:12:23	847	$Z_{F_1}(K)$	constant mean	76.94*	85.98*	99.55*	93.00*
			factor model	108.97*	127.47*	143.15*	152.06*
		$Z_{F_2}(K)$	constant mean	204.97*	210.81*	244.07*	248.85*
			factor model	136.04*	174.99*	196.58*	194.68*
94:12:30-13:12:27	992	$Z_{F_1}(K)$	constant mean	67.15*	78.33*	65.38*	52.64*
			factor model	38.42*	48.88*	45.38*	37.20*
		$Z_{F_2}(K)$	constant mean	188.12*	222.24*	218.00*	195.87*
			factor model	138.13*	134.99*	139.21*	136.28*

Test statistics $Z_{F_1}(K)$ is marked with asterisks if it is larger than 23.209, the 1% critical value of $\chi^2(10)$. Test statistics $Z_{F_2}(K)$ is marked with asterisks if it is larger than 124.116, the 1% critical value of $\chi^2(90)$.

This shows that a lot of the power is coming from the off diagonal elements.

8.2 Long Horizon

We investigate the variance ratios at the long horizon. We again consider the three size-sorted CRSP portfolios. First, we evaluate the long run behaviour of the variance ratio statistics. In this case, we work with the bias-corrected estimators (defined in Appendix 10.1)

$$\widehat{\mathcal{VR}}^{bc}(K) = \widehat{\mathcal{VR}}(K)\left\{1 + \frac{K-1}{T}\right\} \quad ; \quad \widehat{\mathcal{VR}d}^{bc}(K) = \widehat{\mathcal{VR}d}(K)\left\{1 + \frac{K-1}{T}\right\}. \tag{43}$$

We show below the eigenvalues of $\widehat{\mathcal{VR}}^{bc}(K)$ for three weekly size-sorted CRSP portfolio returns against lags in three sub-samples: the red dashed lines are for eigenvalues of $\widehat{\mathcal{VR}}^{bc}(K)$ in the first sub-sample (62:07:06-78:09:29) and the green marked lines are for eigenvalues of $\widehat{\mathcal{VR}}^{bc}(K)$ in the second sub-sample (78:10:06-94:12:23), and the blue solid lines are for eigenvalues of $\widehat{\mathcal{VR}}^{bc}(K)$ in the third sub-sample (94:12:30-13:12:27).

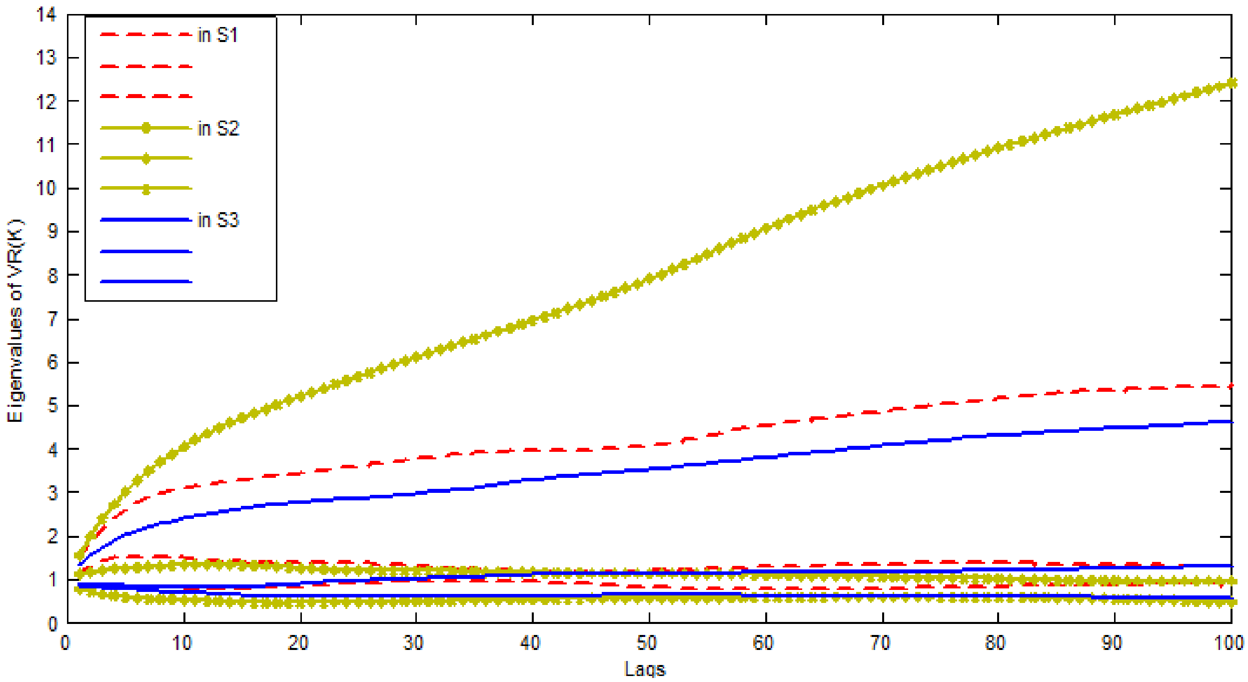

Figure 2: The eigenvalues of the variance ratio for weekly CRSP size sorted portfolio returns in three sub-samples as a function of lags.

We see that the largest eigenvalue increases steadily out to the two year horizon we consider in all three subperiods. In fact, the increase appears to be linear in lag, although the slope is far less than one. The last subperiod has the lowest values throughout, while surprisingly, the second period 1978-1994 seems to have the largest amount of potential linear predictability that could have been exploited during this period. The second and third eigenvalues are quite flat and close to one throughout. This evidence does not seem to be consistent with the fads model, or even the bubble process.

We next evaluate the long run behaviour of the $CS(K)$ statistics. Specifically, we consider two one sided statistics:

$$\widehat{CS}_{\pm}(K) = \frac{2}{d(d-1)} \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \left[\widehat{\mathcal{VR}d}_{\pm}^{bc}(K) \right]_{ij}$$

These statistics measure in some average sense the cross dependence for certain directions. We show below the $CS_{+}(K)$ and $CS\_(K)$ statistics for three weekly size-sorted CRSP portfolio returns against lag K in three sub-samples: the red solid line is for $CS_{+}(K)$ in the first sub-sample (62:07:06-78:09:29), the red dashed line is for $CS_{+}(K)$ in the second sub-sample (78:10:06-94:12:23), the red marked line is for $CS_+(K)$ in the third sub-sample (94:12:30-13:12:27); the blue solid line is for $CS_-(K)$ in the first sub-sample, the blue dashed line is for $CS_-(K)$ in the second sub-sample, and the blue marked line is for $CS_-(K)$ in the third sub-sample.

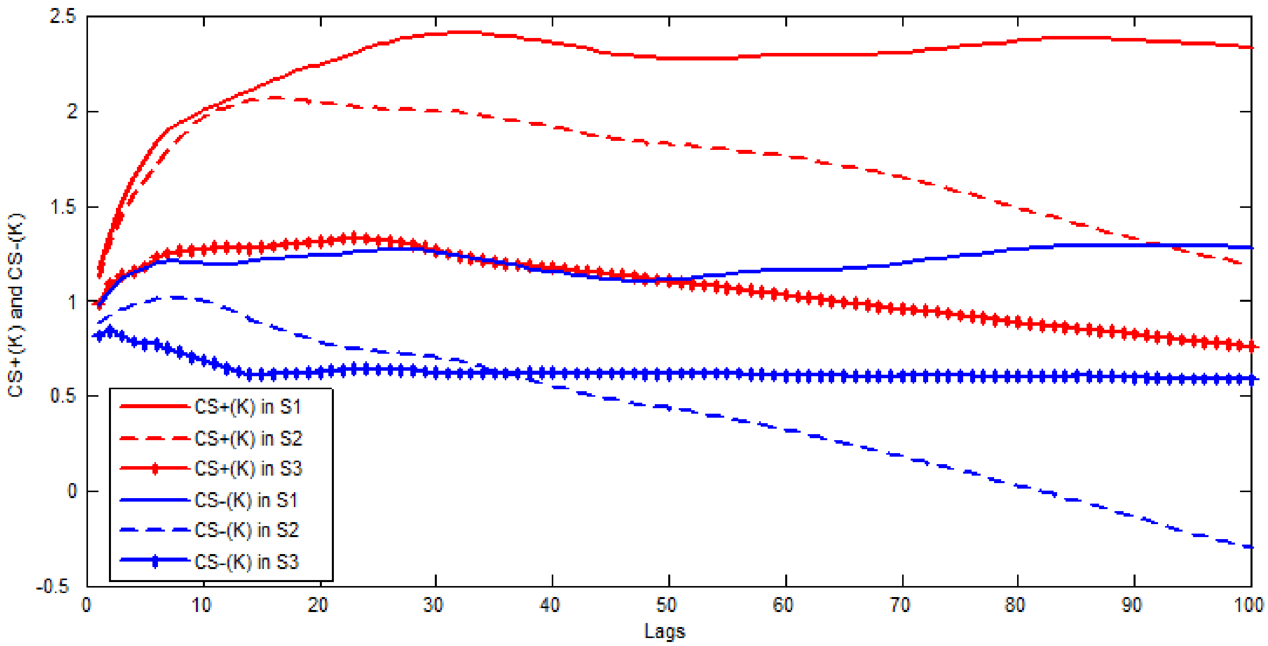

Figure 3: $CS_+(K)$ and $CS\_(K)$ statistics for weekly size sorted CRSP portfolio returns in three sub-samples as a function of lags.

In each subperiod, the $CS_+(K)$ measures all exceed the $CS_-(K)$ measures over all lags, which means that the average directional cross dependence from larger-size portfolios to smaller-size portfolios are stronger than those in the opposite directions, up to two years. The $CS_+(K)$ measures decrease in the recent period over the long horizon. Also the shape of the term structure is quite flat in the most recent period, whereas in the second period, and to a lesser extent in the first period, there seems to be a hump shaped curve suggesting this dependence reaches a maximum somewhere between 10 and 30 weeks. We can further detect that the average statistic, $CS(K) = [CS_+(K) + CS_-(K)]/2$, measuring the average cross dependence for both directions between three size-sorted CRSP portfolios, becomes weaker (more efficient) in recent periods along the long horizon.

We then examine the long run $GMV(K)$ statistics. We show below the $GMV(K)$ for three weekly size-sorted CRSP portfolio returns against lags in three sub-samples: the blue line is for $GMV(K)$ in the first sub-sample (62:07:06-78:09:29) and the green line is for $GMV(K)$ in the second sub-sample (78:10:06-94:12:23), and the red line is for $GMV(K)$ in the third sub-sample (94:12:30-13:12:27).

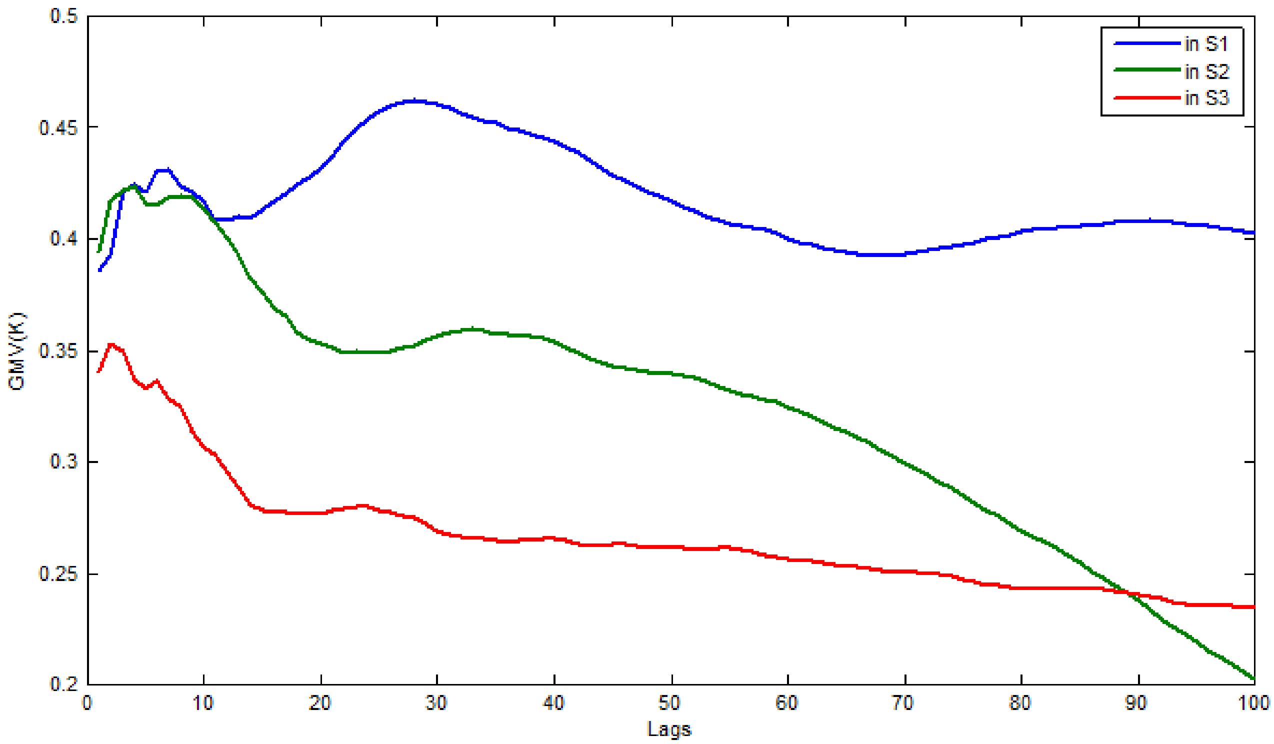

Figure 4: GMV(K) statistics for weekly size sorted CRSP portfolio returns in three sub-samples as a function of lags.

We lastly investigate the long run $\pi(K)$ statistics. We show below the $\pi(K)$ for three weekly size-sorted CRSP portfolio returns against lags in three sub-samples: the blue line is for $\pi(K)$ in the first sub-sample (62:07:06-78:09:29) and the green line is for $\pi(K)$ in the second sub-sample (78:10:06-94:12:23), and the red line is for $\pi(K)$ in the third sub-sample (94:12:30-13:12:27).

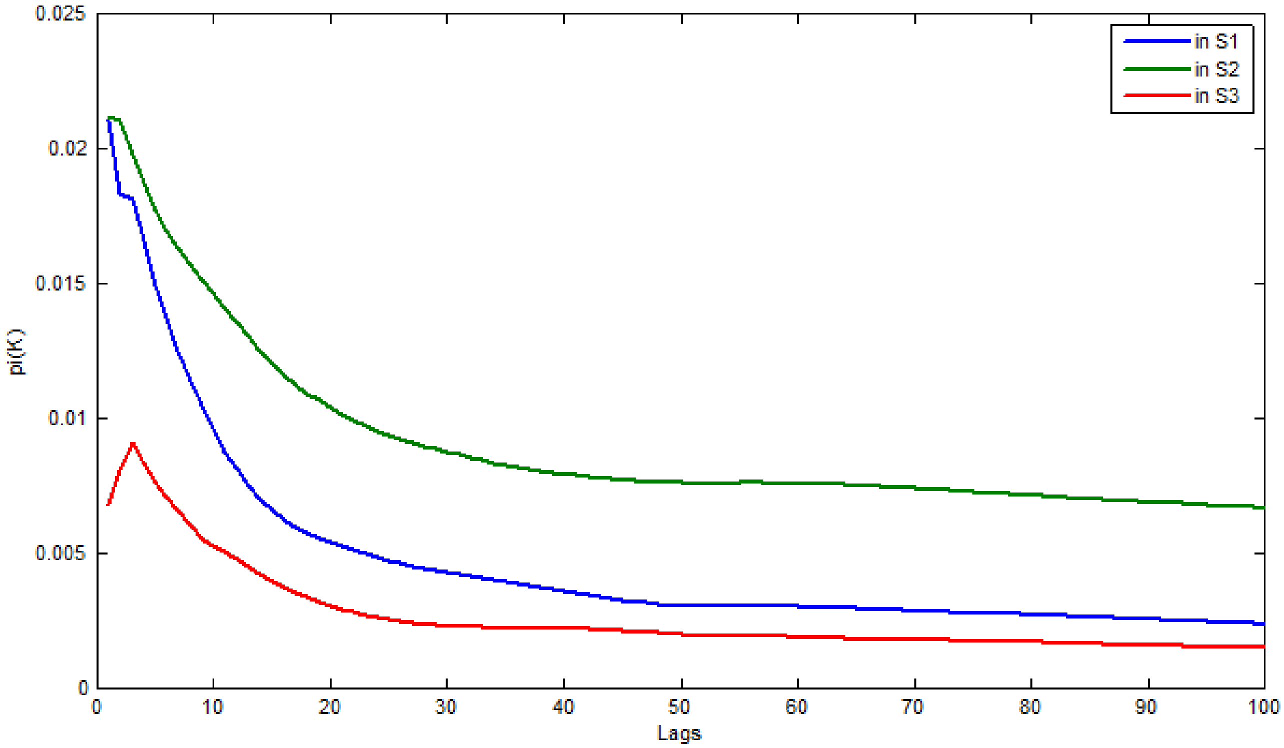

Figure 5: $\pi(K)$ statistics for weekly size sorted CRSP portfolio returns in three sub-samples as a function of lags.

9 Conclusions

The first methodological point we make is to propose confidence intervals that are consistent under the martingale hypothesis alone and do not require an additional no leverage/symmetric distribution assumption such as maintained in Lo and MacKinlay (1988), CLM, and in much subsequent work. Our confidence intervals are typically larger than those used elsewhere, and therefore reduce the significance of any associated test. We believe our theory is more credible with regard to the data generating process we expect for daily or even lower frequency stock returns. The second contribution is about embedding this theory in a multivariate framework. The multivariate variance ratios provides a basis for aggregating the the cross correlation behaviour of asset returns and providing tests of the multivariate null hypothesis. It implies many more restrictions on the data than the univariate ratios.

Our empirical work reports that the US size sorted stock portfolios seem to have come closer to the efficient markets prediction, although, especially for small caps, there remains some statistically significant linear predictability. Although many of the individual variance ratio statistics do not reject the null hypothesis with our standard errors, the joint tests of the multivariate hypothesis reject at the 1% level in all cases.

Typically, three competing explanations are advanced for the predictability in short horizon returns based on past prices (Boudoukh, Whitelaw, and Richardson (1994)): First, microstructure effects such as nonsynchronous trading and bid ask bounce. Second, time varying risk premia. Third, the irrational behaviour of market participants. It would seem that there is a lot of evidence that microstructure effects have reduced considerably over time. For example, it is hard to find even small cap stocks that do not trade now many times during a day. The microstructure explanation would imply that the long horizon daily or weekly variance ratios should return to unity, but this is not the case in our data even for the most recent period. We also provided a test of whether the autocorrelations could be explained by time varying risk premia inside a Fama French factor model. We found that this approach could not capture all the linear dependency in the data. Therefore, the first two explanations do not seem to be able to match the magnitude of the effects. On the other hand, the magnitude of the predictability has reduced in the most recent period. Furthermore, whether the found departures are exploitable is not clear. Timmerman (2008) investigates the forecasting performance of a number of linear and nonlinear models and says: "Most of the time the forecasting models perform rather poorly, but there is evidence of relatively short-lived periods with modest return predictability. The short duration of the episodes where return predictability appears to be present and the relatively weak degree of predictability even during such periods makes predicting returns an extraordinarily challenging task". Our (multivariate) evidence does not substantially contradict that, certainly using linear multivariate methods the amount of predictability we have found and its durability is limited and has reduced over time even through the recent financial crisis. The long horizon analysis suggests that the largest eigenvalue of the variance ratio matrix grows linearly with horizon, although the slope is far less than the unit slope predicted by the bubble process of section 5.3. Furthermore, the trajectory is flatter in the more recent period, again supporting the claim that market inefficiency has reduced.

10 Appendix

10.1 Simulation Study

10.1.1 Size

To investigate how our procedures work in practice, we perform a small simulation study for the $\widehat{\mathcal{VR}}(K)$ and $\widehat{\mathcal{VR}d}_{+}(K)$ statistics under two types of null hypothesis:

$$H_0^{(1)} : \text{i.i.d.}$$

$$H_0^{(2)} : \text{m.d.s.}$$

To simulate the null $H_0^{(1)}$, a sequence of T vector of X_t is drawn from a i.i.d normal distribution $N\left(0, I_d\right)$. We simulate the null $H_0^{(2)}$ by generating the data from a diagonal multivariate ARCH model,

$$X_t = H_t^{1/2} \varepsilon_t$$

$$H_t = \varpi + \alpha X_{t-1} X_{t-1}^{\intercal},$$

where $\varepsilon_t \sim i.i.d.N(0, I_d)$, $\varpi = I_d$ and $\alpha = 0.5$. All these simulations are based on 10000 replications, with sample size, $T = 1024$, dimension $d = 3$. The nominal size is chosen to be 5%.

We use the test statistics $Z_1^{(iid)}(K)$, $Z_1(K)$, $Z_2^{(iid)}(K)$ and $Z_2(K)$, in which $Z_1(K)$ and $Z_2(K)$ are as defined in the Application section. $Z_1^{(iid)}(K)$ and $Z_2^{(iid)}(K)$ are similarly defined except using $\widehat{S}_{iid}(K)$

$$\widehat{S}_{iid}(K) = D_n^{+} \widehat{Q}_{iid}(K) D_n^{+^{\intercal}}.$$

The empirical quantiles of $Z_1^{(iid)}(K)$, $Z_1(K)$, $Z_2^{(iid)}(K)$ and $Z_2(K)$ are obtained by simulating the quantiles of $\sum_{i=1}^{d} \lambda_i^{(W)}$ and $\prod_{i=1}^{d} \lambda_i^{(W)}$ respectively, where W is a $d \times d$ symmetric matrix such that $\text{vech}\left(W\right) \sim N(0, I_{d(d+1)/2})$.

Table 10-1: Empirical quantiles of $Z_1^{(iid)}(K)$, $Z_1(K)$, $Z_2^{(iid)}(K)$ and $Z_2(K)$

	d	0.025	0.975
$Z_1^{(iid)}(K), Z_1(K)$	3	-3.4047	3.3841
$Z_2^{(iid)}(K), Z_2(K)$	3	-7.9355	7.9863

Table 10-2 and Table 10-3 report the empirical size of nominal 5% variance ratio tests using $Z_1^{(iid)}(K)$, $Z_1(K)$, $Z_2^{(iid)}(K)$ and $Z_2(K)$ conducted under the null hypothesis: $H_0^{(1)}$: i.i.d and $H_0^{(2)}$: m.d.s. respectively.

Table 10-2: Empirical size of nominal 5% variance ratio tests of the null hypothesis $H_0^{(1)}$

			Size of 5 percent test			
Sample size	K	d	$Z_1^{(iid)}(K)$	$Z_1(K)$	$Z_2^{(iid)}(K)$	$Z_2(K)$
1024	2	3	0.0493	0.0481	0.0518	0.0517
1024	4	3	0.0504	0.0559	0.0517	0.0511
1024	8	3	0.0448	0.0511	0.0489	0.0525
1024	16	3	0.0470	0.0608	0.0487	0.0546

Table 10-3: Empirical size of nominal 5% variance ratio tests of the null hypothesis $H_0^{(2)}$

			Size of 5 percent test			
Sample size	K	d	$Z_1^{(iid)}(K)$	$Z_1(K)$	$Z_2^{(iid)}(K)$	$Z_2(K)$
1024	2	3	0.2697	0.0517	0.1842	0.0498
1024	4	3	0.2186	0.0523	0.1497	0.0515
1024	8	3	0.161	0.0561	0.1039	0.0501
1024	16	3	0.1177	0.0676	0.0767	0.0516

Table 10-2 shows that the empirical sizes of variance ratio tests using $Z_1^{(iid)}(K)$, $Z_1(K)$, $Z_2^{(iid)}(K)$ and $Z_2(K)$ are all close to the nominal value 5%. In Table 10-3, we see that under the null of m.d.s., the $Z_1^{(iid)}(K)$ and $Z_2^{(iid)}(K)$ are unreliable, for example, when $K = 2$, the empirical size of the 5% variance ratio test using $Z_1^{(iid)}(K)$ is 26.97%, using $Z_2^{(iid)}(K)$ is 18.42%. In this case, the empirical sizes of test using $Z_1(K)$ and $Z_2(K)$ are close to 5%.

Table 10-4 reports the empirical size of nominal 5% variance ratio tests using the $[Zd(K)]_{ii}$ statistic conducted under the null $H_0^{(2)}$. The results show that the $[Zd(K)]_{ii}$ statistic is reliable under the null of m.d.s.

Table 10-4: Empirical size of nominal 5% variance ratio tests [using the $[Zd(K)]_{ii}$ statistic] of the null hypothesis $H_0^{(2)}$

	Size of 5 percent test				
	Sample size	$K=2$	$K=4$	$K=8$	$K=16$
$[Zd(K)]_{11}$	1024	0.0415	0.0389	0.0401	0.0400
$[Zd(K)]_{22}$	1024	0.0462	0.0504	0.0498	0.0509
$[Zd(K)]_{33}$	1024	0.0490	0.0478	0.0523	0.0538

10.1.2 Power

Consider the following model:

$$p_t^* = \mu + p_{t-1}^* + \varepsilon_t$$

$$p_t = p_t^* + \eta_t$$

$$\eta_t = \beta \eta_{t-1} + \xi_t$$

where $\varepsilon_t \sim$ i.i.d.$(0, \Omega_\varepsilon)$, $\xi_t \sim$ i.i.d.$(0, \Omega_\xi)$. As shown in Fama and French (1998) for univariate case, if $\beta < 1$, we have $\widehat{\mathcal{VR}}(K) < I_d$. While Phillips, Wu and Yu (2009) suggested a bubble process which is a linear explosive process without collapsing, such as $\beta > 1$, for which we should have $\widehat{\mathcal{VR}}(K) > I_d$. We examine the power of the variance ratio tests using the $Z_1^{(iid)}(K)$ and $Z_2^{(iid)}(K)$ statistics against two alternative hypotheses:

$$H_1^{(1)} : \text{fads model with } \beta < 1$$

$$H_1^{(2)} : \text{explosive bubble without collapsing with } \beta > 1$$

Based on 10000 replications, we have the following results.

Table 10-5: Power of the variance ratio tests [using the $Z_1^{(iid)}(K)$ and $Z_2^{(iid)}(K)$ statistics]

			5 percent test			
Sample size	K	d	$\beta = 0.85$		$\beta = 1.01$	
1024	16	3	$Z_1^{(iid)}(K)$	$Z_2^{(iid)}(K)$	$Z_1^{(iid)}(K)$	$Z_2^{(iid)}(K)$
			0.9995	0.6349	1.0000	0.9971

Table 10-5 shows that the variance ratio tests using $Z_1^{(iid)}(K)$ and $Z_2^{(iid)}(K)$ are powerful against these alternatives.

10.2 Proof of Main Results

Proof of Theorem 1. We first present the proof under Assumption A. For $j = 1, \ldots, K$, it is straightforward to see that

$$\begin{aligned}
\sqrt{T} \cdot \mathrm{vec}\big(\widehat{\Gamma}(j)\big) &= \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \big(X_{t-j} - \overline{X}\big) \otimes \big(X_t - \overline{X}\big) \\
&= \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \Big(\widetilde{X}_{t-j} \otimes \widetilde{X}_t\Big) - \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \widetilde{X}_{t-j} \otimes \big(\overline{X} - \mu\big) \\
&\quad -\big(\overline{X} - \mu\big) \otimes \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \widetilde{X}_t + \frac{T-j}{\sqrt{T}} \big(\overline{X} - \mu\big) \otimes \big(\overline{X} - \mu\big) \qquad (44) \\
&= \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \Big(\widetilde{X}_{t-j} \otimes \widetilde{X}_t\Big) + o_p(1), \qquad (45)
\end{aligned}$$

where in (45) we made use of $\sum_{t=j+1}^{T} \widetilde{X}_t = O_p(\sqrt{T})$, a result implied by the CLT for stationary ergodic martingale difference. The CLT is justified by the fact that the difference $\big|\sqrt{T}^{-1}\big(\sum_{t=1}^{T} \widetilde{X}_t - \sum_{t=j+1}^{T} \widetilde{X}_t\big)\big| = o_p(1)$; similar arguments are implicitly used from hereafter. We shall also implicitly exploit the fact that condition A2 implies all moments less than four exists and finite by Jensen's inequality.

In the meantime, since $\widetilde{X}_t \widetilde{X}_t^{\intercal}$ is a measurable transformation of $\widetilde{X}_t$ it is again stationary ergodic, (although it no longer possesses a martingale difference structure anymore). Therefore, we may apply Birkhoff's ergodic theorem and continuous mapping theorem on $T^{-1} \sum_{t=1}^{T} \widetilde{X}_t \widetilde{X}_t^{\intercal}$, yielding $\widehat{\Sigma}^{-1/2} - \Sigma^{-1/2} = o_p(1)$. Consequently, for each j we have

$$\begin{aligned}
\mathrm{vec}(\widehat{R}(j)) &= \mathrm{vec}\left(\left[\widehat{\Sigma}^{-1/2} - \Sigma^{-1/2} + \Sigma^{-1/2}\right] \widehat{\Gamma}(j) \left[\widehat{\Sigma}^{-1/2} - \Sigma^{-1/2} + \Sigma^{-1/2}\right]\right) \\
&= \mathrm{vec}\left(\Sigma^{-1/2} \widehat{\Gamma}(j) \Sigma^{-1/2}\right) + T^{-1/2} O_p(1) \cdot o_p(1) + T^{-1/2} O_p(1) \cdot O(1) \\
&= \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right) \mathrm{vec}(\widehat{\Gamma}(j)) + o_p(1). \qquad (46)
\end{aligned}$$

because $\sqrt{T}^{-1} \sum_{t=j+1}^{T} (\widetilde{X}_{t-j} \otimes \widetilde{X}_t)$ is bounded in probability.

Next we observe that

$$
\begin{aligned}
\sqrt{T}\text{vec}\left(\widehat{\mathcal{VR}}_{+}(K) - I_d\right) &= \sqrt{T} \cdot \sum_{j=1}^{K-1} 2\left(1 - \frac{j}{K}\right) \cdot \text{vec}\left(\widehat{R}(j)\right) \\
&= \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right) \cdot \sum_{j=1}^{K-1} c_j \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \widetilde{X}_{t-j} \otimes \widetilde{X}_t + o_p(1) \\
&= \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right) \cdot \frac{1}{\sqrt{T}} \sum_{t=K}^{T} \left[\sum_{j=1}^{K-1} c_j \left(\widetilde{X}_{t-j} \otimes \widetilde{X}_t\right)\right] + o_p(1) \\
&=: \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right) \cdot \frac{1}{\sqrt{T}} \sum_{t=K}^{T} Z_t + o_p(1). \qquad (47)
\end{aligned}
$$

Now to establish the CLT on Z_t, take any constant vector $a = (a_1, \ldots, a_{d^2})^\intercal \in \mathbb{R}^{d^2}$, and note that $a^\intercal Z_t$ is a one-dimensional martingale difference sequence because we have $E[\widetilde{X}_{bt}\widetilde{X}_{c,t-j}|\mathcal{F}_{t-1}] = E[\widetilde{X}_{bt}|\mathcal{F}_{t-1}]\widetilde{X}_{c,t-j}$ a.s. for all $j \geq 1$ and $b, c = 1, \ldots, d$. Then, since the moment condition A2 ensure that

$$
E(a^\intercal Z_t)^2 = a^\intercal \text{var}(Z_t) a = a^\intercal \left[\sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_j c_k \Xi_{jk}\right] a < \infty,
$$

where $\Xi_{jk} = E[\widetilde{X}_{t-j} \otimes \widetilde{X}_t][\widetilde{X}_{t-k} \otimes \widetilde{X}_t]^\intercal$, the CLT for stationary ergodic martingale difference gives

$$
a^\intercal \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t\right) \Longrightarrow N\left(0, a^\intercal \left[\sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_j c_k \Xi_{jk}\right] a\right). \qquad (48)
$$

Hence by the Cramér-Wold device, continuous mapping and Slutsky's theorem we have

$$
\sqrt{T}\text{vec}\left(\widehat{\mathcal{VR}}_{+}(K) - I_d\right) \Longrightarrow N\left(0, \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_j c_k \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right) \Xi_{jk} \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right)\right),
$$

completing the proof.

Deriving the limiting distribution for the same statistic under Assumption MH* closely follows similar arguments. We note that the expansion for $\sqrt{T} \cdot \text{vec}(\widehat{\Gamma}(j))$ is still valid because the summations in the second, third and fourth terms in (44) still converges in probability to one in view of the CLT for mixing sequence, Herrndorf (1985, Theorem 0) whose regularity conditions are satisfied by MH1-MH3. As a consequence, we end up with (45) as before. Finally, condition MH2 and MH3 allow for the Law of Large Numbers for mixing variables, White (1984, Corollary 3.48), yielding (46) and (47) as before.

Now we are only left with verifying (48). Since any measurable transformation of $\widetilde{X}_t$ preserves the mixing property with the same rate specified in MH2, for any d^2-dimensional constant vector a Herrndorf's CLT we have

$$a^{\intercal}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\right) \Longrightarrow N\left(0, a^{\intercal}\left[\sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_j c_k \Xi_{jk}\right] a\right),$$

where $\Xi_{jk} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E[\widetilde{X}_{t-j}\otimes\widetilde{X}_t][\widetilde{X}_{t-k}\otimes\widetilde{X}_t]^{\intercal}$. The CLT above holds provided the following regularity conditions are ensured: $E(a^{\intercal}Z_{tj}) = 0$, $\sup_t E|a^{\intercal}Z_t|^{\beta} < \infty$ for some $\beta > 2$, and finally

$$\lim_{T\to\infty}\frac{1}{T}E\left(\sum_{t=1}^{T} a^{\intercal} Z_t\right)^2 = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\operatorname{var}\left(a^{\intercal} Z_t\right) = a^{\intercal}\left[\sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_j c_k \Xi_{jk}\right] a$$

is positive and finite.

The first condition is trivial by MH1, and the second and third conditions are satisfied by MH2 and MH3 along with positive definiteness of $Q(K)$, respectively. The rest of the arguments are exactly the same as before, completing the proof.

Similar arguments apply to the diagonally and one-sided normalized statistics. For $j = 1, \ldots, K-1$,

$$\begin{aligned}
\operatorname{vec}(\widehat{R}d(j)) &= \left(D^{-1/2}\otimes D^{-1/2}\right)\operatorname{vec}(\widehat{\Gamma}(j)) + o_p(T^{-1/2}) \\
\operatorname{var}\left(\sqrt{T}\operatorname{vec}(\widehat{R}d(j))\right) &= \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_j c_k\left(D^{-1/2}\otimes D^{-1/2}\right)\Xi_{jk}\left(D^{-1/2}\otimes D^{-1/2}\right),
\end{aligned}$$

and also

$$\begin{aligned}
\operatorname{vec}(\widehat{R}_L(j)) &= \left(\Sigma^{-1}\otimes I\right)\operatorname{vec}(\widehat{\Gamma}(j)) + o_p(T^{-1/2}) \\
\operatorname{var}\left(\sqrt{T}\operatorname{vec}(\widehat{R}_L(j))\right) &= \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_j c_k\left(\Sigma^{-1}\otimes I\right)\Xi_{jk}\left(\Sigma^{-1}\otimes I\right),
\end{aligned}$$

The entire proof is now complete. ■

Proof of Corollary 1. Because the proposed estimator $\widehat{\Sigma}$ for the covariance matrix is consistent under both sets of assumptions, it suffices to show consistency of $\widehat{\Xi}_{jk}$. Writing

$$\begin{aligned}\widehat{\Xi}_{jk} &= \frac{1}{T}\sum_{t=\max\{j,k\}+1}^{T}\left(X_{t-j}-\overline{X}\right)\left(X_{t-k}-\overline{X}\right)^{\intercal}\otimes\left(X_{t}-\overline{X}\right)\left(X_{t}-\overline{X}\right)^{\intercal} \\ &= \frac{1}{T}\sum_{t=\max\{j,k\}+1}^{T}\left[\widetilde{X}_{t-j}\widetilde{X}_{t-k}^{\intercal}\otimes\widetilde{X}_{t}\widetilde{X}_{t}^{\intercal}\right]+o_p(1) \\ &= \frac{1}{T}\sum_{t=\max\{j,k\}+1}^{T}\left[\left(\widetilde{X}_{t-j}\otimes\widetilde{X}_{t}\right)\left(\widetilde{X}_{t-k}\otimes\widetilde{X}_{t}\right)^{\intercal}\right]+o_p(1).\end{aligned}$$

we see that the desired result follows by applying either the Ergodic theorem or the Law of Large Numbers for mixing variables depending upon the set of assumption being imposed. The regularity conditions for each theorem are ensured by Assumption A2 and MH3, respectively. Note that these consistency results can be extended to almost sure sense in both cases, without requiring any further condition. ■

Proof of (29). We follow the similar approaches for the two parameter statistics. Under the null hypothesis, by the geometric series expansion we have

$$\begin{aligned}&\sqrt{T}\left(\widehat{\mathcal{VR}}_{+}(K,L)-I_d\right) \\ =\;& 2\sqrt{T}\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)\widehat{\Gamma}(j)-2\sqrt{T}\sum_{j^{\intercal}=1}^{L-1}\left(1-\frac{j}{L}\right)\widehat{\Gamma}(j)+o_p(1) \\ =\;& 2\sqrt{T}\sum_{j=1}^{K-1}\left[\left(1-\frac{j}{K}\right)-\left(1-\frac{j}{L}\right)1(j\le L)\right]\widehat{\Gamma}(j)+o_p(1) \\ =\;& \frac{K-L}{KL}\sum_{j=1}^{L-1}2j\sqrt{T}\widehat{\Gamma}(j)+2\sum_{j=L}^{K-1}\left(1-\frac{j}{K}\right)\sqrt{T}\widehat{\Gamma}(j)+o_p(1).\end{aligned}$$

Hence denoting

$$\widetilde{c}_{j,K,L}=c_{j,K}-c_{j,L}=2\left(\frac{K-L}{KL}j\right)1(j\le L-1)+2\left(1-\frac{j}{K}\right)1(L\le j\le K-1),$$

we have

$$\begin{aligned}\operatorname{var}\left(\sqrt{T}\operatorname{vec}\left(\widehat{\mathcal{VR}}_{+}(K)-I_d\right)\right) &= \operatorname{var}\left(\sqrt{T}\sum_{j=1}^{K-1}\widetilde{c}_{j,K,L}\cdot\operatorname{vec}\left(\widehat{\Gamma}(j)\right)\right) \\ &\to \sum_{j=1}^{K-1}\sum_{k=1}^{K-1}\widetilde{c}_{j,K,L}\widetilde{c}_{k,K,L}\Xi_{jk},\end{aligned}$$

so the proof is complete on employing the CLT. As before, the limiting distribution of the two sided statistic can be obtained by the transformation using the duplication matrix.

Finally, taking $K = LJ$ for positive integers J and L, we have

$$\begin{aligned}
\sum_{j=1}^{K-1} \widetilde{c}_{j,LJ,L}^{2} &= \left(\frac{JL-L}{JL^2}\right)^2 \sum_{j=1}^{L-1} j^2 + \sum_{j=L}^{JL-1} \left(1-\frac{j}{JL}\right)^2 \\
&= \left(\frac{J-1}{JL}\right)^2 \frac{L(2L-1)(L-1)}{6} + \frac{(J-1)(JL-L+1)(2JL-2L+1)}{6J^2L} \\
&= \frac{(J-1)(2JL^2-2L^2+1)}{6JL}.
\end{aligned}$$

whereas $L\sum_{j=1}^{J-1} c_{j,J}^2 = \frac{L(2J-1)(J-1)}{6J}$. Comparing both terms yield the relative efficiency as desired. ■

Proof of Theorem 2. The proof consists of two steps.

Step 1: From (44) we know that for each $j = 1, \dots, K$, $\sqrt{T} \cdot \mathrm{vec}\big(\widehat{\Gamma}(j)\big)$ can be decomposed into a main term plus three error terms. We show that the last three terms are 'asymptotically' negligible i.e. $o_p(1)$ *uniformly* over $j = 1, \dots, K$. It suffices to prove this for a single arbitrary component, as we shall do here (but without introducing extra notations for the sake of simplicity; For example, with a slight abuse of notation $\widetilde{X}_t$ is taken to mean $\widetilde{X}_{it}$ for some $i = 1, \dots, d$ and so on).

Consider the second term in (44):

$$\begin{aligned}
\max_{1\le j\le K} A_2 = \max_{1\le j\le K}\left[(\overline{X}-\mu)\cdot\left(\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_t\right)\right] &= \left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_t\right)\cdot\max_{1\le j\le K}\left(\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_t\right) \\
&= \frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{X}_t\right)\cdot\max_{1\le j\le K}\left(\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_t\right) \\
&= \frac{1}{\sqrt{T}}\cdot O_p(1)\cdot O_p(K) = O_p\left(\frac{K}{\sqrt{T}}\right) = o_p(1),
\end{aligned}$$

because the stochastic error

$$\left|\frac{1}{\sqrt{T}}\sum_{t=K+1}^{T}\widetilde{X}_t - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{X}_t\right| = \frac{1}{\sqrt{T}}\left|\sum_{t=1}^{K}\widetilde{X}_t\right| = \frac{\sqrt{K}}{\sqrt{T}}\left|\frac{1}{\sqrt{K}}\sum_{t=1}^{K}\widetilde{X}_t\right| = O_p\left(\sqrt{\frac{K}{T}}\right) = o_p(1)$$

ensures $\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_t$ to be bounded in probability for any $j = 1, \dots, K$. Similar argument applies to the first error term, yielding the same result.

As for the last error term A_3, because

$$\max_{1\le j\le K}\left[\frac{T-j}{\sqrt{T}}\cdot(\overline{X}-\mu)\cdot(\overline{X}-\mu)\right] = \left(\max_{1\le j\le K}\frac{T-j}{\sqrt{T}\sqrt{T}\sqrt{T}}\right)\cdot\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{X}_t\right)\cdot\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{X}_t\right)$$

where

$$\max_{1\le j\le K}\left|\frac{T-j}{\sqrt{T}\sqrt{T}\sqrt{T}}\right| = \max_{1\le j\le K}\left|\frac{1-j/T}{\sqrt{T}}\right| = \left|\frac{1-\frac{K}{T}}{\sqrt{T}}\right| \longrightarrow 0$$

we have asymptotic negligibility of the error terms as desired.

Step 2: The second step involves deriving the limiting distribution under this new asymptotics:

$$\begin{aligned}
\sqrt{\frac{T}{K}}\mathrm{vec}\left(\widehat{VR}(K)_+ - I_d\right) &= \frac{\sqrt{T}}{\sqrt{K}}\cdot\sum_{j=1}^{K-1}2\left(1-\frac{j}{K}\right)\cdot\mathrm{vec}\left(\widehat{R}(j)\right)\\
&= \left(\Sigma^{-1/2}\otimes\Sigma^{-1/2}\right)\cdot\frac{1}{\sqrt{K}}\sum_{j=1}^{K-1}c_j\left(\left[\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_{t-j}\otimes\widetilde{X}_t\right]+o_p(1)\right)\\
&= \left(\Sigma^{-1/2}\otimes\Sigma^{-1/2}\right)\cdot\frac{1}{\sqrt{KT}}\sum_{t=1}^{T}\left[\sum_{j=1}^{K-1}c_j(\widetilde{X}_{t-j}\otimes\widetilde{X}_t)\right]+o_p(1)\\
&=: \left(\Sigma^{-1/2}\otimes\Sigma^{-1/2}\right)\cdot\frac{1}{\sqrt{KT}}\sum_{t=1}^{T}Z_{tK}+o_p(1).
\end{aligned}$$

When two summations are 'swapped' in the above, we used to take the summation of t from K to T before, ignoring a finite number of terms because the difference was negligible in any case; however, here we shall instead take summation from 1 to T, and let any $\widetilde{X}_t$ with negative ts be zero.

We now denote by $\bar{Z}_{TK}$ and $\bar{Z}_T$ the average of Z_{tK} and $\lim_{K\to\infty}Z_{tK}$, respectively: i.e.

$$\bar{Z}_{TK} := \frac{1}{T}\sum_{t=1}^{T}Z_{tK}, \qquad \bar{Z}_T = \frac{1}{T}\sum_{t=1}^{T}Z_t = \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{j=1}^{\infty}\left(\widetilde{X}_{t-j}\otimes\widetilde{X}_t\right)\right]. \tag{49}$$

We know from Theorem 1 that as $T\to\infty$, *for each* $K = 1, 2, \ldots$

$$\frac{1}{\sqrt{T}}\frac{1}{\sqrt{K}}\sum_{t=1}^{T}Z_{tK} \equiv \sqrt{T}\bar{Z}_{TK} \Longrightarrow Y_K \sim N\left(0, \frac{1}{K}\sum_{j=1}^{K-1}\sum_{k=1}^{K-1}c_jc_kE\left[\widetilde{X}_{t-j}\widetilde{X}_{t-k}^{\intercal}\otimes\widetilde{X}_t\widetilde{X}_t^{\intercal}\right]\right).$$

Now, provided that $\lim_{K\to\infty}K^{-1}\sum_{j=1}^{K}\sum_{k=1}^{K}E\left[\widetilde{X}_{t-j}\widetilde{X}_{t-k}^{\intercal}\otimes\widetilde{X}_t\widetilde{X}_t^{\intercal}\right]<\infty$, as $K\to\infty$ we have

$$Y_K \Longrightarrow Y \sim N\left(0, \lim_{K\to\infty}\frac{1}{K}\sum_{j=1}^{K}\sum_{k=1}^{K}E\left[\widetilde{X}_{t-j}\widetilde{X}_{t-k}^{\intercal}\otimes\widetilde{X}_t\widetilde{X}_t^{\intercal}\right]\right)$$

Hence, if one can show that $\sqrt{T}\left(\bar{Z}_T - \bar{Z}_{TK}\right)$ is 'asymptotically negligible' in the sense that

$$\forall\varepsilon>0, \quad \lim_{K\to\infty}\limsup_{T\to\infty}P\left(\left|\sqrt{T}\left(\bar{Z}_T-\bar{Z}_{TK}\right)\right|>\varepsilon\right)=0$$

then by Proposition 6.3.9 of Brockwell and Davis (1991), it will follow that

$$\sqrt{T}\bar{Z}_T = \lim_{K\to\infty}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\sum_{j=1}^{K-1}\widetilde{X}_{t-j}\otimes\widetilde{X}_t\right]\right) \Longrightarrow Y\sim N\left(0, \lim_{K\to\infty}\frac{1}{K}\sum_{j=1}^{K}\sum_{k=1}^{K}E\left[\widetilde{X}_{t-j}\widetilde{X}_{t-k}^{\prime\intercal}\otimes\widetilde{X}_t\widetilde{X}_t^{\intercal}\right]\right).$$

as $T\to\infty$.

Now note that

$$\begin{aligned}
E\left(\left|\sqrt{T}\left(\bar{Z}_T-\bar{Z}_{TK}\right)\right|^2\right) &= E\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\sum_{j=K}^{\infty}c_j(\widetilde{X}_{t-j}\otimes\widetilde{X}_t)\right)\right|^2 = \frac{1}{T}\text{var}\left(\sum_{t=1}^{T}\left(\sum_{j=K}^{\infty}c_j(\widetilde{X}_{t-j}\otimes\widetilde{X}_t)\right)\right)\\
&=: \frac{1}{T}\text{var}\left(\sum_{t=1}^{T}R_t\right) = 2\sum_{s=1}^{T-1}\left(1-\frac{s}{T}\right)\cdot\text{cov}(R_t,R_{t-s})\\
&= 2\sum_{s=1}^{T-1}\left(1-\frac{s}{T}\right)\cdot E\left(\frac{1}{K}\sum_{j=K}^{\infty}\sum_{k=K}^{\infty}c_jc_k\left(\widetilde{X}_t\widetilde{X}_{t-s}\otimes\widetilde{X}_{t-j}\widetilde{X}_{t-s-k}\right)\right)\\
&\to 2\sum_{s=1}^{\infty}\left(\lim_{K\to\infty}\frac{1}{K}\sum_{j=K}^{K}\sum_{k=K}^{K}c_jc_kE\left(\widetilde{X}_t\widetilde{X}_{t-s}\otimes\widetilde{X}_{t-j}\widetilde{X}_{t-s-k}\right)\right)
\end{aligned}$$

as $T\to\infty$ (so then tends to zero as $K\to\infty$).

Hence by Markov's inequality we have

$$\lim_{K\to\infty}\limsup_{T\to\infty}P\left(\left|\sqrt{T}\left(\bar{Z}_T-\bar{Z}_{TK}\right)\right|>\varepsilon\right)\le\lim_{K\to\infty}\limsup_{T\to\infty}\frac{E\left|\sqrt{T}\left(\bar{Z}_T-\bar{Z}_{TK}\right)\right|^2}{\varepsilon^2}=0$$

as desired.

Finally, using the continuous mapping and Slutsky's theorem we end up with

$$\begin{aligned}
&\sqrt{T}\text{vec}\left(\widehat{VR}(K)_+ - I_d\right)\\
&\Longrightarrow N\left(0, \lim_{K\to\infty}\frac{1}{K}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}(\Sigma^{-1/2}\otimes\Sigma^{-1/2})E\left[\widetilde{X}_{t-j}\widetilde{X}_{t-k}^{\intercal}\otimes\widetilde{X}_t\widetilde{X}_t^{\intercal}\right](\Sigma^{-1/2}\otimes\Sigma^{-1/2})\right),
\end{aligned}$$

completing the proof. ■

Proof of Proposition FaustM. The proof proceeds by showing asymptotic equivalence of the trace (of the multivariate variance ratio) test and the likelihood ratio (LR) test under the null and alternative hypotheses. That is,

$$f\left(\text{tr}\left(\widehat{VR}(K)\right)\right)-LR\xrightarrow{P}0 \tag{50}$$

for some function f, in which case the tests based on two statistics will possess the same large sample properties.

Recall the alternative version of the estimator $\widehat{\mathcal{VR}}^{\&}(K)$. From the definitions it is not difficult to see that

$$\widehat{VR}(K) - \widehat{VR}^{\&}(K) = \frac{1}{K}\sum_{r=1}^{K-2}\left\{\widehat{\Sigma}^{-1/2}\left[(K-r)\frac{1}{T}\sum_{t=r+1}^{K-1}\left(X_t-\overline{X}\right)\left(X_{t-r}-\overline{X}\right)^{\mathsf{T}}\right]\widehat{\Sigma}^{-1/2}\right\}$$
$$+\frac{1}{K}\sum_{r=1}^{K-2}\left\{\widehat{\Sigma}^{-1/2}\left[(K-r)\frac{1}{T}\sum_{t=2}^{K-1}\left(X_{t-r}-\overline{X}\right)\left(X_t-\overline{X}\right)^{\mathsf{T}}\right]\widehat{\Sigma}^{-1/2}\right\}+o_p(1) \quad (51)$$

converges in probability to zero because each term in square brackets is $o_p(1)$ by Chebyshev's inequality, and $\widehat{\Sigma}^{-1/2} \xrightarrow{P} \Sigma^{-1/2}$. Now that we have $f(\mathrm{tr}(\widehat{VR}(K))) - f(\mathrm{tr}(\widehat{VR}^{\&}(K))) = o_p(1)$ due to linearity of trace, it remains to show that

$$f\left(\mathrm{tr}\left(\widehat{VR}^{\&}(K)\right)\right) - LR \xrightarrow{P} 0.$$

We denote by $\Phi \in \mathbb{R}^{(T-K+1)\times T}$ the 'coefficient matrix'

$$\Phi = \begin{pmatrix} \phi_{K-1} & \phi_{K-2} & \cdots & \phi_0 & 0 & 0 & \cdots & 0 \\ 0 & \phi_{K-1} & \cdots & \phi_1 & \phi_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \phi_1 & \phi_0 \end{pmatrix},$$

where $\phi_j = 1$ for all $j = 0, \ldots, K-1$. Denoting i by a conformable column vector of ones as before we can write

$$\begin{aligned}\widehat{\mathrm{var}}(X_t + \cdots + X_{t-K+1}) =: \widehat{\mathrm{var}}(X_t^{\phi}) &= \frac{1}{T}\sum_{t=1}^{T}\left(X_t^{\phi}-\overline{X}^{\phi}\right)\left(X_t^{\phi}-\overline{X}^{\phi}\right)^{\mathsf{T}} \\ &= \frac{1}{T}\left(\Phi X - \Phi i\overline{X}^{\mathsf{T}}\right)^{\mathsf{T}}\left(\Phi X - \Phi i\overline{X}^{\mathsf{T}}\right) \\ &= \frac{1}{T}\left(X - i\overline{X}^{\mathsf{T}}\right)^{\mathsf{T}}\Phi^{\mathsf{T}}\Phi\left(X - i\overline{X}^{\mathsf{T}}\right),\end{aligned}$$

from which it follows that

$$\begin{aligned}\widehat{VR}^{\&}(K) &:= \frac{1}{K}\widehat{\mathrm{var}}(X_t)^{-1/2}\widehat{\mathrm{var}}(X_t+\cdots+X_{t-K+1})\widehat{\mathrm{var}}(X_t)^{-1/2} = \frac{1}{K}\widehat{\Sigma}^{-1/2}\widehat{\Sigma}(K)\widehat{\Sigma}^{-1/2} \\ &= \frac{1}{K}\left[\left(X-i\overline{X}^{\mathsf{T}}\right)^{\mathsf{T}}\left(X-i\overline{X}^{\mathsf{T}}\right)\right]^{-1/2}\left(X-i\overline{X}^{\mathsf{T}}\right)^{\mathsf{T}}\Phi^{\mathsf{T}}\Phi\left(X-i\overline{X}^{\mathsf{T}}\right)\left[\left(X-i\overline{X}^{\mathsf{T}}\right)^{\mathsf{T}}\left(X-i\overline{X}^{\mathsf{T}}\right)\right]^{-1/2} \\ &=: \frac{1}{K}[(A^{\mathsf{T}}A)^{-1/2}]\cdot[A^{\mathsf{T}}\Phi^{\mathsf{T}}\Phi A]\cdot[(A^{\mathsf{T}}A)^{-1/2}]. \quad (52)\end{aligned}$$

It follows from the explicit expressions of the probability density for the matrix normal distributions that the rejection region based on the likelihood ratio statistic is given by

$$LR = \log\left(\frac{\det\Big[\left(X - i\widehat{\mu}_1^{\intercal}\right)^{\intercal}\Sigma_{q*}^{-1}\left(X - i\widehat{\mu}_1^{\intercal}\right)\Big]}{\det\Big[\left(X - i\overline{X}^{\intercal}\right)^{\intercal}\left(X - i\overline{X}^{\intercal}\right)\Big]}\right) < k$$

for some positive threshold constant k, where $\widehat{\mu}_1 \equiv \widetilde{X}$ is the maximum likelihood estimate of the mean $\mu = EX_t$ under the *alternative* hypotheses. Using a standard property of the logarithmic determinant we see that

$$\begin{aligned}
LR &= \log\left(\det\left\{\Big[\left(X - i\overline{X}^{\intercal}\right)^{\intercal}\left(X - i\overline{X}^{\intercal}\right)\Big]^{-1}\Big[\left(X - i\widetilde{X}^{\intercal}\right)^{\intercal}\Sigma_{q*}^{-1}\left(X - i\widetilde{X}^{\intercal}\right)\Big]\right\}\right) \\
&\leq \operatorname{tr}\left(\Big[\left(X - i\overline{X}^{\intercal}\right)^{\intercal}\left(X - i\overline{X}^{\intercal}\right)\Big]^{-1}\Big[\left(X - i\widetilde{X}^{\intercal}\right)^{\intercal}\Sigma_{q*}^{-1}\left(X - i\widetilde{X}^{\intercal}\right)\Big] - I\right) \\
&\leq \operatorname{tr}\left(\widehat{\Sigma}^{-1}\cdot\frac{1}{T}\Big[\left(X - i\widetilde{X}^{\intercal}\right)^{\intercal}\Sigma_{q*}^{-1}\left(X - i\widetilde{X}^{\intercal}\right)\Big]\right). \qquad (53)
\end{aligned}$$

Besides, it follows by the cyclic property of the trace operator that

$$\begin{aligned}
\operatorname{tr}\left(\widehat{VR}^{\&}(K)\right) &= \frac{1}{K}\operatorname{tr}\Big([(A^{\intercal}A)^{-1/2}]\cdot[A^{\intercal}\Phi^{\intercal}\Phi A]\cdot[(A^{\intercal}A)^{-1/2}]\Big) \\
&= \frac{1}{K}\operatorname{tr}\Big([(A^{\intercal}A)^{-1}]\cdot[A^{\intercal}\Phi^{\intercal}\Phi A]\Big) \\
&= \frac{1}{K}\operatorname{tr}\left(T\Big[\left(X - i\overline{X}^{\intercal}\right)^{\intercal}\left(X - i\overline{X}^{\intercal}\right)\Big]^{-1}\cdot\frac{1}{T}\Big[\left(X - i\overline{X}^{\intercal}\right)^{\intercal}\Phi^{\intercal}\Phi\left(X - i\overline{X}^{\intercal}\right)\Big]\right) \\
&= \frac{1}{K}\operatorname{tr}\left(\widehat{\Sigma}^{-1}\cdot\frac{1}{T}\Big[\left(X - i\widetilde{X}^{\intercal} + i\left(\widetilde{X}^{\intercal} - \overline{X}^{\intercal}\right)\right)^{\intercal}\Phi^{\intercal}\Phi\left(X - i\widetilde{X}^{\intercal} + i\left(\widetilde{X}^{\intercal} - \overline{X}^{\intercal}\right)\right)\Big]\right).
\end{aligned}$$

Now multiplying the last quantity by the horizon K, $q > 0$, adding $d = tr(I_d)$, and then lastly multiplying by a constant $\alpha > 0$ give

$$\begin{aligned}
&\operatorname{tr}\left(\widehat{\Sigma}^{-1}\cdot\frac{1}{T}\Big[\left(X - i\widetilde{X}^{\intercal} + i\left(\widetilde{X}^{\intercal} - \overline{X}^{\intercal}\right)\right)^{\intercal}\cdot\{\alpha(I + q\Phi^{\intercal}\Phi)\}\cdot\left(X - i\widetilde{X}^{\intercal} + i\left(\widetilde{X}^{\intercal} - \overline{X}^{\intercal}\right)\right)\Big]\right) \\
&= \operatorname{tr}\left(\widehat{\Sigma}^{-1}\cdot\frac{1}{T}\Big[\left(X - i\widetilde{X}^{\intercal} + i\left(\widetilde{X}^{\intercal} - \overline{X}^{\intercal}\right)\right)^{\intercal}\{\Sigma_q^{-1} + 0^*\}\left(X - i\widetilde{X}^{\intercal} + i\left(\widetilde{X}^{\intercal} - \overline{X}^{\intercal}\right)\right)\Big]\right) \qquad (54)
\end{aligned}$$

where 0^* is a matrix of zeros except for the $(K-1)\times(K-1)$ blocks in the northwest and southeast corners. The reader is directed to Faust (1992, Lemma 1) for the proof of the equivalence relationship $\alpha(I + q\Phi^{\intercal}\Phi) \equiv \Sigma_q^{-1} + 0^*$. Now replacing the sample estimator for the cross-sectional variance by its population version (with some negligible error), we see that the difference between (54) and (53) multiplied by $\sqrt{T}$ is given by

$$\begin{aligned}
&\sqrt{T}\cdot \operatorname{tr}\left(\Sigma^{-1}\cdot\frac{1}{T}\left[\left(i\left(\widetilde{X}^{\intercal}-\overline{X}^{\intercal}\right)\right)^{\intercal}\cdot\Sigma_q^{-1}\cdot\left(i\left(\widetilde{X}^{\intercal}-\overline{X}^{\intercal}\right)\right)\right]\right)+o_p(1)\\
&\quad=\operatorname{tr}\left(\Sigma^{-1}\cdot\left[\sqrt{T}\left(\widetilde{X}^{\intercal}-\overline{X}^{\intercal}\right)^{\intercal}\left\{\frac{i^{\intercal}\cdot\Sigma_q^{-1}\cdot i}{T}\right\}\left(\widetilde{X}^{\intercal}-\overline{X}^{\intercal}\right)\right]\right)+o_p(1)
\end{aligned}$$

because the trace is a linear mapping. It is trivial to show that the term inside $\{\cdot\}$ is bounded in probability. Furthermore, the proof of Proposition 2 in Faust (1992) suggests that the individual entries of the squared braket converges in probability to zero (hence so does the entire matrix by definition), yielding

$$\sqrt{T}\left|\alpha\left\{d+qK\cdot\operatorname{tr}\left(\widehat{\mathcal{VR}}^{\&}(K)\right)\right\}-LR\right|\xrightarrow{p}0. \tag{55}$$

This suggests that there exist some α and $q>0$ for which the trace test has the same large sample properties of the LR test against the $\phi-$best class alternatives. The proof is now complete because the sequence of the likelihood ratio test with $q^{*}=\delta/\sqrt{T}$ is locally most powerful (MP) invariant in view of Crowder (1976) and Engle (1984). ■

Proof of Theorem 3. Note that as $K\to\infty$, $\Omega_\eta(K)\to 2\Omega_\eta=2\operatorname{var}(\eta_t)$. It follows that as $K\to\infty$

$$\begin{aligned}
\mathcal{VR}(K) &= K^{-1}\Sigma_1^{-1/2}\Sigma_K\Sigma_1^{-1/2}=K^{-1}\Sigma_1^{-1/2}\left(K\Omega_\varepsilon+\Omega_\eta(K)\right)\Sigma_1^{-1/2}\\
&\longrightarrow \Sigma_1^{-1/2}\Omega_\varepsilon\Sigma_1^{-1/2}=\Sigma_1^{-1/2}\left[\Sigma_1-\Omega_\eta(1)\right]\Sigma_1^{-1/2}\\
&= I-\Sigma_1^{-1/2}\Omega_\eta(1)\Sigma_1^{-1/2}\le I,
\end{aligned}$$

since Σ_1 and $\Omega_\eta(1)$ are positive semidefinite. The strict inequality holds since $\Omega_\eta(1)$ is assumed strictly positive definite.

By similar arguments

$$\begin{aligned}
\mathcal{VR}d(K) &= K^{-1}D_1^{-1/2}\Sigma_K D_1^{-1/2}=K^{-1}D_1^{-1/2}\left(K\Omega_\varepsilon+\Omega_\eta(k)\right)D_1^{-1/2}\\
&\longrightarrow D_1^{-1/2}\Omega_\varepsilon D_1^{-1/2}=D_1^{-1/2}\left(\Sigma_1-\Omega_\eta(1)\right)D_1^{-1/2}\\
&= D_1^{-1/2}\Sigma_1 D_1^{-1/2}-D_1^{-1/2}\Omega_\eta(1)D_1^{-1/2}\\
&= Rd\,(0)-D_1^{-1/2}\Omega_\eta(1)D_1^{-1/2}\le Rd\,(0)
\end{aligned}$$

which is the instantaneous correlation matrix of the return process.

■

PROOF OF THEOREM 4. This follows from the multivariate extension of Theorem 1 of Liu and Wu (2010) applied to the frequency $\theta = 0$. The weighting scheme automatically satisfies their condition 1. See also Andrews (1991). ■

PROOF OF (37). For simplicity we suppose that $p_t = \delta_T p_{t-1} + \varepsilon_t$ with ε_t iid and $\delta_T = 1 + \frac{c}{k_T}$, where $k_T = T^\alpha, \alpha \in (0, 1/2)$ and some positive constant c. According to Phillips and Magdalinos (2007, Theorem 4.3) we have

$$\left((\delta_T^{-T}/k_T) \sum_{t=1}^{T} p_{t-1}\varepsilon_t, (\delta_T^{-2T}/k_T^2) \sum_{t=1}^{T} p_{t-1}^2 \right) \Longrightarrow (XY, Y^2),$$

where X, Y are iid copies of a $N(0, \sigma_\varepsilon^2/2c)$ distribution.

Since the observed return X_t is the difference of the log prices we have $X_t = p_t - p_{t-1} = \frac{c}{k_T} p_{t-1} + \varepsilon_t$, and consequently the sum of the squared return is

$$\begin{aligned} \sum_{t=1}^{T} X_t^2 &= \frac{c^2}{k_T^2} \sum_{t=1}^{T} p_{t-1}^2 + \frac{2c}{k_T} \sum_{t=1}^{T} p_{t-1}\varepsilon_{t-1} + \sum_{t=1}^{T} \varepsilon_{t-1}^2 \\ &\Rightarrow \frac{c^2}{k_T^2} k_T^2 \delta_T^{2T} Y^2 + \frac{2c}{k_T} k_T \delta_T^T XY + T\sigma_\varepsilon^2 + R \\ &= c^2 \delta_T^{2T} Y^2 + R, \end{aligned}$$

where R is a generic remainder term that contains smaller order terms. The first term dominates the others because $\delta_T^{2T} = (1 + \frac{c}{k_T})^{2T} \to \infty$ very fast. Therefore, we have

$$\delta_T^{-2T} \sum_{t=1}^{T} X_t^2 \Longrightarrow c^2 Y^2. \tag{56}$$

Likewise,

$$X_t(2) = p_t - p_{t-2} = (\delta_T^2 - 1)p_{t-2} + \varepsilon_t + \delta_T \varepsilon_{t-1} \simeq \frac{2c}{k_T} p_{t-2} + \varepsilon_t + \delta_T \varepsilon_{t-1},$$

by the Binomial approximation because $c/k_T = c/T^\alpha$ becomes negligible as T gets bigger. Therefore,

$$\delta_T^{-2T} \sum_{t=1}^{T} X_t(2)^2 \Longrightarrow 4c^2 Y^2.$$

Similarly for general K, as $T \to \infty$ we have:

$$X_t(K) = \left(\delta_T^K - 1\right) p_{t-K} + \sum_{j=0}^{K-1} \delta_T^j \varepsilon_{t-j}$$

$$\delta_T^{-2T} \sum_{t=1}^{T} X_t(K)^2 \Longrightarrow K^2 c^2 Y^2. \tag{57}$$

In fact, using Cramér-Wold device it can be shown that the convergence in (56) and (57) is joint. Therefore, by the continuous mapping theorem

$$\widehat{\mathcal{VR}}(K) \sim \frac{\sum_{t=1}^{T} X_t(K)^2}{K \sum_{t=1}^{T} X_t^2} \xrightarrow{P} K,$$

as required. ■

Proof of Theorem 5. As consistency of $\widehat{\theta}$ follows by standard arguments we shall only prove the main theorem under Assumption MHF. It is trivial to see that the same mixing rate and the moment condition of F_t applies to G_t. Hence the autocovariance is given by

$$\begin{aligned}
\sqrt{T} \cdot \operatorname{vec}\left(\widehat{\Gamma}(j)\right) &= \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \widehat{\varepsilon}_{t-j} \otimes \widehat{\varepsilon}_t = \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \left(X_{t-j} - \widehat{\theta}_{t-j}^{\intercal} G_t\right) \otimes \left(X_t - \widehat{\theta}^{\intercal} G_t\right) \\
&= \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \left[\varepsilon_{t-j} - \left(\widehat{\theta}^{\intercal} - \theta^{\intercal}\right) G_{t-j}\right] \otimes \left[\varepsilon_t - \left(\widehat{\theta}^{\intercal} - \theta^{\intercal}\right) G_t\right] \\
&= \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \left(\varepsilon_{t-j} \otimes \varepsilon_t\right) - \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \left[\left(\widehat{\theta}^{\intercal} - \theta^{\intercal}\right) G_{t-j} \otimes \varepsilon_t\right] \\
&\quad - \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \left[\varepsilon_{t-j} \otimes \left(\widehat{\theta}^{\intercal} - \theta^{\intercal}\right) G_t\right] + \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \left[\left(\widehat{\theta}^{\intercal} - \theta\right) G_{t-j} \otimes \left(\widehat{\theta}^{\intercal} - \theta^{\intercal}\right) G_t\right]. \\
&= \tau_1 + \tau_2 + \tau_3 + \tau_4.
\end{aligned} \tag{58}$$

Because F_t and ε_t are jointly mixing with the same mixing coefficient, standard arguments yield that any measurable transformation of Z_t is also mixing with same rate. Therefore it follows that $\tau_2 = O_p(1) \cdot o_p(1) = o_p(1) = \tau_3$ in view of Lemma 6 and the CLT for mixing variables. The last term $\tau_4 = o_p(1)$ due to similar arguments. Furthermore, since the Law of Large Numbers for mixing variables, cf. White (1984, Corollary 3.48) yields

$$\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t^{\intercal} = \widehat{\Sigma} \xrightarrow{P} \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\left[\varepsilon_t \varepsilon_t^{\intercal}\right] = \Sigma < \infty,$$

it follows that the variance ratio statistic can be written as

$$
\begin{aligned}
\sqrt{T}\text{vec}\left(\widehat{VR}_{\varepsilon+}(K) - I_d\right) &= \sqrt{T} \cdot \sum_{j=1}^{K-1} 2\left(1 - \frac{j}{K}\right) \cdot \text{vec}\left(\widehat{R}(j)\right) \\
&= \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right) \cdot \sum_{j=1}^{K-1} c_j \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \varepsilon_{t-j} \otimes \varepsilon_t + o_p(1) \\
&= \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right) \cdot \frac{1}{\sqrt{T}} \sum_{t=K}^{T} \left[\sum_{j=1}^{K-1} c_j \left(\varepsilon_{t-j} \otimes \varepsilon_t\right)\right] + o_p(1)
\end{aligned}
$$

the rest follows by mixing conditions on ε_t and the functional central limit theorem of Herrndorf (1985). The proof under Assumption AF can be done in a similar manner. ■

Proof of Corollary 3. It suffices to prove consistency of $\widehat{\Xi}_{jk}$ because consistency of sample covariance $\widehat{\Sigma}$ is trivial by a suitable law of large numbers. We have

$$
\begin{aligned}
\widehat{\Xi}_{jk} &= \frac{1}{T} \sum_{t=\max\{j,k\}+1}^{T} \left(\widehat{\varepsilon}_{t-j}\widehat{\varepsilon}_{t-k}^{\intercal}\right) \otimes \left(\widehat{\varepsilon}_t\widehat{\varepsilon}_t^{\intercal}\right) && (59) \\
&= \frac{1}{T} \sum_{t=\max\{j,k\}+1}^{T} \left[\varepsilon_{t-j} - \left(\widehat{\theta}^{\intercal} - \theta^{\intercal}\right) G_{t-j}\right] \left[\varepsilon_{t-k} - \left(\widehat{\theta}^{\intercal} - \theta^{\intercal}\right) G_{t-k}\right]^{\intercal} \\
&\qquad \otimes \left[\varepsilon_t - \left(\widehat{\theta}^{\intercal} - \theta^{\intercal}\right) G_t\right] \left[\varepsilon_t - \left(\widehat{\beta}^{\intercal} - \theta^{\intercal}\right) G_t\right]^{\intercal} \\
&= \frac{1}{T} \sum_{t=\max\{j,k\}+1}^{T} \left[\varepsilon_{t-j}\varepsilon_{t-k}^{\intercal}\right] \otimes \left[\varepsilon_t\varepsilon_t^{\intercal}\right] + o_p(1), && (60)
\end{aligned}
$$

and the law of large numbers for mixing variables yields the desired result. Same result holds under Assumption AF using the Ergodic theorem instead. ■

10.3 Bias Correction

We discuss the finite sample biases with a view to proposing a bias correction for the estimated variance ratios when the sample size is small and/or the lag length is large. We have

$$\begin{aligned} E\left[\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_{t-j}\otimes\left(\overline{X}-\mu\right)\right] &= E\left[\frac{1}{T\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_{t-j}\otimes\widetilde{X}_{t-j}\right]=\frac{T-j}{T\sqrt{T}}\sigma \\ E\left[(\overline{X}-\mu)\otimes\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_{t}\right] &= \frac{T-j}{T\sqrt{T}}\sigma \\ E\left[\frac{T-j}{\sqrt{T}}\left(\overline{X}-\mu\otimes\overline{X}-\mu\right)\right] &= \frac{T-j}{T\sqrt{T}}\sigma, \end{aligned}$$

where $\sigma=\mathrm{vec}(\Sigma)$. Therefore

$$E\widehat{v}_j=v_j-\frac{T-j}{T^2}\sigma+o(T^{-1}),$$

where $v_j=\mathrm{vec}\left(\Gamma(j)\right)$ and similarly $\widehat{v}_j=\mathrm{vec}\left(\widehat{\Gamma(j)}\right)$. Under the iid assumption (which allows us to ignore the denominator, see below) we have

$$\begin{aligned} E\left[\widehat{\mathcal{VR}}(K)\right] &= \mathcal{VR}(K)-\frac{2}{T}\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)\left(1-\frac{j}{T}\right)I_d+o(T^{-1}) \\ &= \mathcal{VR}(K)-\frac{K-1}{T}I_d+o(T^{-1}) \\ &= \mathcal{VR}(K)\left\{1-\frac{K-1}{T}\right\}+o(T^{-1}) \end{aligned}$$

under the null hypothesis. Likewise,

$$\begin{aligned} E\left[\widehat{\mathcal{VR}d}(K)\right] &= \mathcal{VR}d(K)-\frac{K-1}{T}Rd(0)+o(T^{-1}) \\ &= \mathcal{VR}d(K)\left\{1-\frac{K-1}{T}\right\}+o(T^{-1}). \end{aligned}$$

For the two parameter statistic, the bias adjustment is a bit more complicated:

$$E\left[\widehat{\mathcal{VR}}^*(K,L)\right]=\mathcal{VR}^*(K,L)-\frac{2}{T}\left[\frac{K-L}{KL}\sum_{j=1}^{L-1}j\left(1-\frac{j}{T}\right)+\sum_{j=L}^{K-1}\left(1-\frac{j}{K}\right)\left(1-\frac{j}{T}\right)\right]I_d+o(T^{-1}).$$

To do a full bias analysis of the variance ratio statistic under the martingale hypothesis, we need to take account of the denominator. By a Taylor expansion we have

$$\begin{aligned} \widehat{R}(j) &= \Sigma^{-1/2}\widehat{\Gamma}(j)\Sigma^{-1/2}-\frac{1}{2}\Sigma^{-1}\left(\widehat{\Sigma}-\Sigma\right)\Sigma^{-1}\widehat{\Gamma}(j)\Sigma^{-1/2} \\ &\quad -\frac{1}{2}\Sigma^{-1/2}\widehat{\Gamma}(j)\Sigma^{-1}\left(\widehat{\Sigma}-\Sigma\right)\Sigma^{-1}+o_p(T^{-1}), \end{aligned}$$

under the null hypothesis. To calculate the (approximate) expected value of the second and third terms, it suffices to replace $\sqrt{T}(\widehat{\Sigma}-\Sigma)$ and $\sqrt{T}\widehat{\Gamma}(j)$ with their limiting (joint) distributions. We have

$$\begin{aligned}\sqrt{T}\widehat{v}_j &= \frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\left(\widetilde{X}_{t-j}\otimes\widetilde{X}_t\right)+o_p(1)\\ \sqrt{T}(\widehat{v}_0-v_0) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\widetilde{X}_t\otimes\widetilde{X}_t\right)+o_p(1).\end{aligned}$$

Therefore,

$$\operatorname{acov}(\sqrt{T}\widehat{v}_j,\sqrt{T}(\widehat{v}_0-v_0))=E\left[\left(\widetilde{X}_{-j}\widetilde{X}_0^{\intercal}\otimes\widetilde{X}_0\widetilde{X}_0^{\intercal}\right)\right]+\sum_{s=1}^{\infty}E\left[\left(\widetilde{X}_{-j}\widetilde{X}_s^{\intercal}\otimes\widetilde{X}_0\widetilde{X}_s^{\intercal}\right)\right]. \qquad (61)$$

From this we can obtain a formula for $E[\Sigma^{-1}(\widehat{\Sigma}-\Sigma)\Sigma^{-1}\widehat{\Gamma}(j)\Sigma^{-1/2}]$ in terms of the right hand side of (61), but clearly it will be very complicated to use in practice. Under full independence we can ignore this term and just do a simple bias correction as described above.

References

[1] Anderson, T. W. (2003). An Introduction to Multivariate Statistical Analysis. New Jersey: Wiley & Sons.

[2] Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica, 59(3), 817-858.

[3] Bailey, N., Kapetanios, G. and Pesaran, H. (2012). Exponent of Cross-sectional Dependence: Estimation and Inference (January 2012). Working paper available at http://www.econ.cam.ac.uk/research/repec/cam/pdf/cwpe1206.pdf

[4] Blanchard, O. J. and Watson, M. W. (1982). Bubbles, rational expectations and financial markets. NBER Working Paper no. 945.

[5] Boudoukh, Whitelaw, and Richardson (1994). A tale of three schools: Insights on autocorrelations of short-horizon stock returns. Review of financial studies, 7(3), 539-573.

[6] Brockwell, P. J. and Davis, R. A. (1991). Time series: theory and methods. New York: Springer-Verlag.

[7] Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). The Econometrics of Financial Markets. Princeton, NJ: Princeton University Press.

[8] Castura, J., Litzenberger, R., Gorelick, R. and Dwivedi, Y. (2010). Market efficiency and microstructure evolution in US equity markets: A high-frequency perspective. Available at http://www.sec.gov/comments/s7-02-10/s70210-364.pdf

[9] Charles, A. and Darné, O. (2009). Variance-Ratio Tests of Random Walk: An Overview. Journal of Economic Surveys, 23(3), 503-527.

[10] Chaudhuri, K. and Wu, Y. (2003). Random walk versus breaking trend in stock prices: Evidence from emerging markets. Journal of Banking & Finance, 27(4), 575-592.

[11] Chen, W. W. and Deo, R. S. (2006). The variance ratio statistic at large horizons. Econometric Theory, 22(2), 206-234.

[12] Chitturi, R. V. (1974). Distribution of residual autocorrelations in multiple autoregressive schemes. Journal of the American Statistical Association, 69(348), 928-934.

[13] Cho, J. S. and White, H. (2014). Testing the Equality of Two Positive-Definite Matrices with Application to Information Matrix Testing (June 2014). Working paper available at ftp://165.132.78.176/repec/yon/wpaper/2014rwp-67.pdf.

[14] Chordia, T., Roll, R. and Subrahmanyam, A. (2011). Recent trends in trading activity and market quality. Journal of Financial Economics, 101(2), 243-263.

[15] Cochrane, J. H. (2001). Asset Pricing. Princeton, NJ: Princeton University Press.

[16] Crowder, M. J. (1976). Maximum likelihood estimation for dependent observations. Journal of the Royal Statistical Society: Series B, 38(1), 45-53.

[17] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. Annals of Statistics, 25(1), 1-37.

[18] Deo, R. and Richardson, M. (2003). On the asymptotic power of the variance ratio test. Econometric Theory, 19(2), 231-239.

[19] Eaton, M. L. and Tyler, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. Annals of Statistics, 19(1), 260-271.

[20] Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. Handbook of Econometrics, 2, 775-826.

[21] Engle, R. F., Lilien, D. M. and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. Econometrica, 391-407.

[22] Escanciano, J. C. and Velasco, C. (2006). Generalized spectral tests for the martingale difference hypothesis. Journal of Econometrics, 134(1), 151-185.

[23] Fama, E. F. and French, K. R. (1998). Value versus growth: The international evidence. The Journal of Finance, 53(6), 1975-1999.

[24] Faust, J. (1992). When are Variance Ratio Tests for Serial Dependence Optimal?c. Econometrica, 60(5), 1215-1226.

[25] Ferson, W. E. and Korajczyk, R. A. (1995). Do arbitrage pricing models explain the predictability of stock returns?. Journal of Business, 68(3). 309-349.

[26] Flood, R. P. and Hodrick, R. J. (1990). On testing for speculative bubbles. The Journal of Economic Perspectives, 4(2), 85-101.

[27] Han, H., Linton, O. B., Oka, T. and Whang, Y. J. (2014). The Cross-Quantilogram: Measuring Quantile Dependence and Testing Directional Predictability between Time Series (January 2014). Available at SSRN: http://ssrn.com/abstract=2338468 or http://dx.doi.org/10.2139/ssrn.2338468.

[28] Hasbrouck, J. (2007). Empirical market microstructure: The institutions, economics, and econometrics of securities trading. New York: Oxford University Press.

[29] Herrndorf, N. (1985). A functional central limit theorem for strongly mixing sequences of random variables. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 69(4), 541-550.

[30] Hillman, R. and Salmon, M. (2007). Intrinsic stationarity: investigating predictability in real-time forex transactions. Journal of Financial Forecasting, 1(1), 3-43.

[31] Hong, Y. (2000). Generalized spectral tests for serial dependence. Journal of the Royal Statistical Society: Series B, 62(3), 557-574.

[32] Hong, Y. and Lee, Y. J. (2005). Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. Review of Economic Studies, 72(2), 499-541.

[33] Hosking, J. R. M. (1981). Equivalent forms of the multivariate portmanteau statistic. Journal of the Royal Statistical Society: Series B, 43(2), 261-262.

[34] Kan, R. and Wang, X. (2010). On the distribution of the sample autocorrelation coefficients. Journal of Econometrics, 154(2), 101-121.

[35] Kim, M. J., Nelson, C. R. and Startz R. (1991). Mean reversion in stock prices? A reappraisal of the empirical evidence. Review of Economic Studies, 58, 515-528.

[36] Liu, C. Y. and He, J. (1991). A Variance-Ratio Test of Random Walks in Foreign Exchange Rates. The Journal of Finance, 46(2), 773-785.

[37] Liu, W. and Wu, W. B. (2010). Asymptotics of spectral density estimates. Econometric Theory, 26(4), 1218-1245.

[38] Lo, A. W. (2004). The adaptive markets hypothesis. The Journal of Portfolio Management, 30(5), 15-29.

[39] Lo, A. W. (2005). Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. Journal of Investment Consulting, 7(2), 21-44.

[40] Lo, A. W. and MacKinlay, A. C. (1999). A Non-Random Walk Down Wall Street. New Jersey: Princeton University Press.

[41] Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. Review of Financial Studies, 1(1), 41-66.

[42] Lo, A. W. and MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction?. Review of Financial studies, 3(2), 175-205.

[43] Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. Journal of the American Statistical Association, 96(455), 1066-1076.

[44] Luger, R. (2003). Exact non-parametric tests for a random walk with unknown drift under conditional heteroscedasticity. Journal of Econometrics, 115(2), 259-276.

[45] Magdalinos (2014). Personal communication.

[46] Magnus, J. R. and Neudecker, H. (1980). The elimination matrix: some lemmas and applications. SIAM Journal on Algebraic Discrete Methods, 1(4), 422-449.

[47] Mehra, R. and Prescott, E. C. (2008). Non-risk-based explanations of the equity premium. in R. Mehra, ed., Handbook of the Equity Risk Premium. Amsterdam: Elsevier.

[48] Mikosch, T. and Stărică, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH (1, 1) process. Annals of Statistics, 28(5), 1427-1451.

[49] Montagnoli, A. and de Vries, F. P. (2010). Carbon trading thickness and market efficiency. Energy Economics, 32(6), 1331-1336.

[50] Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. Journal of the American Statistical Association, 55(290), 299-306.

[51] Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. Econometrica, 59(2), 347-370.

[52] Newey, W. K. and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. International Economic Review, 28(3), 777-787.

[53] O'Hara, M. and Ye, M. (2009). Is fragmentation harming market quality?. Journal of Financial Economics, 100(3), 459-474.

[54] Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. Annals of Mathematical Statistics, 28(2), 329-348.

[55] Peterson, R. L., Ma, C. K. and Ritchey, R. J. (1992). Dependence in commodity prices. Journal of Futures Markets, 12(4), 429-446.

[56] Phillips, P. C. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. Journal of Econometrics, 136(1), 115-130.

[57] Phillips, P. C. B., Shi, S. P. and Yu, J (2012). Testing for multiple bubbles. Cowles Foundation Discussion Paper no. 1843.

[58] Phillips, P. C. B. and Solo, V. (1992). Asymptotics for Linear Processes. Annals of Statistics, 20(2), 971-1001.

[59] Phillips, P. C., Wu, Y, and Yu, J. (2009). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?. SMU Economics & Statistics Working Paper Series no. 19.

[60] Phillips, P. C. and Yu, J. (2010). Dating the timeline of financial bubbles during the subprime crisis. Cowles Foundation Discussion Paper no. 1770.

[61] Pierce, D. A. and Haugh, L. D. (1977). Causality in temporal systems: Characterization and a survey. Journal of Econometrics, 5(3), 265-293.

[62] Poterba, J. M. and Summers, L. H. (1988). Mean reversion in stock prices: Evidence and Implications. Journal of Financial Economics, 22(1), 27-59.

[63] Richardson, M. and Stock, J. H. (1989). Drawing inferences from statistics based on multiyear asset returns. Journal of Financial Economics, 25(2), 323-348.

[64] Sheppard, K. (2013). Measuring Market Speed (March 2013). Working paper available at http://www.kevinsheppard.com/images/f/fa/MeasuringMarketSpeed.pdf.

[65] Solnik, B. (1991). International investments. Reading, MA: Addison Wesley.

[66] Szroeter, J. (1978). Generalized variance-ratio tests for serial correlation in multivariate regression models. Journal of Econometrics, 8(1), 47-59.

[67] Timmerman, A. (2008). Elusive Return Predictability. International Journal of Forecasting, 24(1), 1-18.

[68] Wang, P. (2003). Econometric analysis of the real estate market and investment. London: Routledge.

[69] Whang, Y. J. and Kim, J. (2003). A multiple variance ratio test using subsampling. Economics Letters, 79(2), 225-230.

[70] White, H. (1984). Asymptotic theory for econometricians. New York: Academic Press.

[71] Wright, J. H. (2000). Alternative variance-ratio tests using ranks and signs. Journal of Business & Economic Statistics, 18(1), 1-9.