

Multivariate variance ratio statistics

Seok Young Hong
Oliver Linton
Hui Jun Zhang

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP29/14

Multivariate Variance Ratio Statistics*

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

University of Cambridge

June 20, 2014

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 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 smoothly varying risk premium in calendar time, and show that the limiting distribution is the same as in the constant mean adjustment case. 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 three 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 period, for small and medium cap stocks. We find similar results for the main UK stock indexes. The main findings are not substantially affected by allowing for a slowly varying risk premium.

KEYWORDS: Bubbles; Fads; Martingale; Momentum; Predictability

*We thank Alexei Onatski, Andrew Patton, Mark Salmon, and Wei Biao Wu for helpful comments and: CEMMAP, the Cambridge INET, and the Keynes fund for financial support.

[†]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

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 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. The methodology has been applied in many low frequency settings: to US stocks, Poterba and Summers (1988) and Lo and MacKinlay (1988), 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). It has also been applied more recently in high frequency settings, where it has informed the debate on the evolution of market quality. 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.

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 . Then consider tests based on comparing $\text{var}(r_t^\phi)/\text{var}(r_t)$. He shows that each such test 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 BP 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. See Charles and Darné (2009) for a recent review of this methodology and its application.

We make several contributions. First, we propose several multivariate variance ratio statistics. This allows for an across assets view of the EMH. The only papers concerning multivariate ratio tests we have found are by Szroeter (1978) and Cochrane and Sbordone (1988). 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), Hosking (1981), and Dufour and Roy (1986). 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. 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.

Second, we propose an alternative distribution theory and standard errors than are usually adopted. We point out that the limiting distribution established in Lo and MacKinlay (1988, Theorem 3) for the univariate variance ratio statistics is incorrect under their assumptions H1-H4 (i.e., RW3). The correct distribution is much more complicated and depends 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 or one has to use more complicated inference methods based on long run variance estimation, Newey and West (1987), or self normalization, Lobato (2001). Furthermore, 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, dispensing with this condition does 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) but without the additional no-leverage condition but with an additional uncorrelatedness condition on the products of returns. The asymptotic variance is different from that contained in Theorem 3 of Lo and MacKinlay (1988) (and used in much subsequent work). We propose a simple plug-in method for conducting inference that does not require the selection of a bandwidth parameter. We also establish the asymptotic properties of our statistic under several plausible alternative models including the 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.

We apply our methods to three CRSP weekly size-sorted portfolio returns from 1962-2013 and the three subperiods 1962-1978, 1978-1994 and 1994-2013. We show that the degree of inefficiency has reduced over the most recent period, and in some cases this improvement is statistically significant. We also show that the degree of asymmetry in the dependence structure has reduced, although it is still significant. We find similar results for some UK stock indexes. We extend our analysis to include a slowly varying risk premium and seasonal effects, but find that the main empirical results are unchanged. We further investigate the variance ratios at the long horizon. Simulation experiments indicate that our variance ratio tests are reliable, powerful against several alternatives, and useful in dating the origination and collapse of an explosive episode.

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 that has to be estimated from the data. In section 7 we present our application, while section 8 contains some simulation experiments. Section 9 concludes.

2 Multivariate Variance Ratios

For expositional purposes we shall suppose that we have a vector stationary ergodic discrete time series $X_t \subset \mathbb{R}^d$; formal assumptions regarding the data are given below in section 3. Let $\tilde{X}_t = X_t - \mu$, 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 we so far assume to be represented 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 excess return process satisfies

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

where \mathcal{F}_t denotes the past history of the prices of all the assets.¹ This is a stronger assumption than that returns are uncorrelated

$$E(\tilde{X}_t \tilde{X}_{t-j}^\top) = 0 \tag{2}$$

for all $j \neq 0$, which is what is adopted in Lo and MacKinlay (1988) as RW3 and referred to as such in much subsequent work.² RW3 has the advantage that if one rejects it, then one rejects the martingale hypothesis. We argue that in fact to test RW3 one needs to make much stronger additional assumptions that may not be warranted, or one has to employ more complicated inference methods. Lobato, Nankervis, and Savin (2001) do not make these additional assumptions and therefore have to estimate a long run variance consistently to modify their Box-Pierce statistics to achieve asymptotic chi-squared distribution. Lobato (2001) employs a self-normalization approach that leads to a non Gaussian limiting distribution for the sample autocorrelations and the Box-Pierce statistic that is correct under the uncorrelatedness hypothesis, plus some additional technical conditions. Lo and MacKinlay (1988) instead ruled out certain leverage effects to ensure what they thought would be simple standard errors. Unfortunately, they had neglected some important terms in their analysis, which means that their asymptotic distribution is of a more complicated form than they state, and indeed contains a long run variance. We provide corrected Lo and MacKinlay conditions that capture the spirit of their analysis and result in relatively simple limiting distributions. In particular, we drop their leverage hypothesis and replace it with an additional uncorrelatedness assumption that is more acceptable for applications.

We next define the population versions of the variance ratios. Define the following population quantities:

$$\Sigma = \text{var}(X_t) = E(\tilde{X}_t \tilde{X}_t^\top) \tag{3}$$

$$D = \text{diag} \left\{ E(\tilde{X}_{1t}^2), \dots, E(\tilde{X}_{dt}^2) \right\} \tag{4}$$

$$\Psi(j) = \text{cov}(X_t, X_{t-j}) = E(\tilde{X}_t \tilde{X}_{t-j}^\top) \tag{5}$$

¹We note that there are other 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.

²This 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.

$$\Gamma(j) = \Sigma^{-1/2}\Psi(j)\Sigma^{-1/2} \quad (6)$$

$$\Gamma_L(j) = \Psi(j)\Sigma^{-1} \quad ; \quad \Gamma_R(j) = \Sigma^{-1}\Psi(j) \quad (7)$$

$$\Gamma d(j) = D^{-1/2}\Psi(j)D^{-1/2} \quad (8)$$

for $j = 0, \pm 1, \dots$. 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 $\Gamma d(j)$ is the usual definition of the cross-(auto)correlation matrix, while $\Gamma(j)$ is a multivariate correlation matrix. 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 the 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.

2.1 Two Sided Variance Ratios

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

$$VR(K) = \text{var}(X_t)^{-1/2} \text{var}(X_t + X_{t-1} + \dots + X_{t+1-K}) \text{var}(X_t)^{-1/2} / K, \quad (9)$$

where K is some positive integer. Clearly, under the null hypothesis (2) we should have $VR(K) = I_d$. Under the generic (stationary) alternative hypothesis we have

$$VR(K) = I + \sum_{j=1}^{K-1} \left(1 - \frac{j}{K}\right) (\Gamma(j) + \Gamma(j)^\top), \quad (10)$$

which is a symmetric matrix. The off-diagonal elements should be zero under the null hypothesis of no predictability. Both representations (9) and (10) can be used as the basis for estimation.

An alternative multivariate normalization is given by

$$VRa(K) = \text{var}(X_t + X_{t-1} + \dots + X_{t+1-K}) \text{var}(X_t)^{-1} / K,$$

which can likewise generically be written

$$VRa(K) = I + \sum_{j=1}^{K-1} \left(1 - \frac{j}{K}\right) (\Gamma_L(j) + \Gamma_R(j)^\top). \quad (11)$$

This has a regression interpretation, see Chitturi (1974) and Wang (2003, p62).

A third quantity is the diagonally normalized variance ratio

$$VRd(K) = D^{-1/2} \text{var}(X_t + X_{t-1} + \dots + X_{t+1-K}) D^{-1/2} / K \quad (12)$$

$$= \Gamma d(0) + \sum_{j=1}^{K-1} \left(1 - \frac{j}{K}\right) (\Gamma d(j) + \Gamma d(j)^\top), \quad (13)$$

where $\Gamma d(0) = D^{-1/2} \Psi(0) D^{-1/2}$ is the $d \times d$ contemporaneous correlation matrix. Under the null hypothesis that the series is uncorrelated, we should have $VRd(K) = \Gamma d(0)$ the contemporaneous correlation matrix, whose off-diagonal elements are unrestricted by the null hypothesis. The diagonal elements of $VRd(K)$ correspond to the univariate variance ratio statistics, while the off-diagonal elements provide information about the cumulative cross-dynamics between the assets. For example, the typical off-diagonal element is

$$VRd_{ij}(K) = \Gamma d_{ij}(0) + \sum_{k=1}^{K-1} \left(1 - \frac{k}{K}\right) (\Gamma d_{ij}(k) + \Gamma d_{ji}(k)).$$

We may also consider the two parameter family of variance ratio statistics, as in Poterba and Summers (1988),

$$\begin{aligned} VR(K, L) &= \frac{L}{K} \text{var}(X_t + X_{t-1} + \dots + X_{t+1-L})^{-1/2} \text{var}(X_t + X_{t-1} + \dots + X_{t+1-K}) \times \\ &\quad \times \text{var}(X_t + X_{t-1} + \dots + X_{t+1-L})^{-1/2} \\ &= (\Sigma^{1/2} VR(L) \Sigma^{1/2})^{-1/2} (\Sigma^{1/2} VR(K) \Sigma^{1/2}) (\Sigma^{1/2} VR(L) \Sigma^{1/2})^{-1/2} \end{aligned}$$

for some positive integers K and L . Under the null hypothesis (2), we should have $VR(K, L) = I_d$. An alternative is

$$VR^*(K, L) = VR(L)^{-1/2} \times VR(K) \times VR(L)^{-1/2},$$

which satisfies $VR^*(K, L) = I_d$ under the null hypothesis.

For the two parameter version of the statistic $VRa(K, L)$, we might take

$$\begin{aligned} VRa(K, L) &= \frac{L}{K} \text{var}(X_t + X_{t-1} + \dots + X_{t+1-K}) \text{var}(X_t + X_{t-1} + \dots + X_{t+1-L})^{-1} \\ &= VRa(K) \times VRa(L)^{-1}, \end{aligned}$$

which satisfies $VRa(K, L) = I_d$ under the null hypothesis.

For $VRd(K, L)$ we also have several choices. Specifically,

$$\begin{aligned} VRd(K, L) &= \frac{L}{K} D_L^{-1/2} \text{var}(X_t + X_{t-1} + \dots + X_{t+1-K}) D_L^{-1/2} \\ &= D_{VRd(L)}^{-1/2} VRd(K) D_{VRd(L)}^{-1/2}, \end{aligned}$$

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

$$VRd^*(K, L) = VRd(L)^{-1/2} \times VRd(K) \times VRd(L)^{-1/2},$$

which satisfies $VRd^*(K, L) = I_d$ under the null hypothesis.

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 $VR(K)$ and $VRd(K)$ are both symmetric matrices whereas the autocorrelation function $\Gamma d(j)$ is not necessarily symmetric. In fact, one can only recover $\Gamma d(\cdot) + \Gamma d(\cdot)^\top$ or $\Gamma(\cdot) + \Gamma(\cdot)^\top$ from the variance ratio functions $VRd(\cdot)$ and $VR(\cdot)$. This means that information about lead lag relations are eliminated. Instead we propose the following quantities:

$$VR_+(K) = I + 2 \sum_{j=1}^{K-1} \left(1 - \frac{j}{K}\right) \Gamma(j)$$

$$VRd_+(K) = \Gamma d(0) + 2 \sum_{j=1}^{K-1} \left(1 - \frac{j}{K}\right) \Gamma d(j),$$

and the negative counterparts $VR_-(K) = VR_+^\top(K)$ and $VRd_-(K) = VRd_+^\top(K)$, which have the property that³:

$$VR(K) = (VR_+(K) + VR_+^\top(K))/2$$

$$VRd(K) = (VRd_+(K) + VRd_+^\top(K))/2.$$

One can test the null hypothesis of lack of linear predictability based on the matrices $VRd_+(K)$, $VRd_-(K)$ and one can compare the two statistics to quantify the asymmetry in lead lag effects.

For the two parameter statistics, we may consider the following quantities:

$$VR_+^*(K, L) = VR_+(L)^{-1/2} \times VR_+(K) \times VR_+(L)^{-1/2}$$

$$VRd_+^*(K, L) = D_{VRd(L)}^{-1/2} VRd_+(K) D_{VRd(L)}^{-1/2}.$$

³The variance ratio process $\{(VR_+(K), K = 2, 3, \dots)\}$ is a linear invertible functional of the autocorrelation process $\{\Gamma(j), j = \pm 1, \pm 2, \dots\}$. The spectral density matrix $\mathbf{f}(\lambda) = (2\pi)^{-1} \sum_{j=0}^{\infty} \exp(i\lambda j) (\Psi(j) + \Psi(j)^\top)$ is likewise a linear invertible functional of the autocovariance matrix. The covariogram (Cressie (1993)) is likewise a linear invertible functional of the autocovariances. These are just alternative ways of looking at the linear dependence of a series.

2.3 Univariate Parameters of Interest

We discuss some key univariate parameters of interest. 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). 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 , $f(VRa(K)) = f(VR(K))$.

Define the spectrum $\sigma(VR(K)) = \{\lambda \in \mathbb{R} : VR(K)x = \lambda x \text{ for some } x \in \mathbb{R}^d \setminus \{0\}\}$ of the variance ratio statistic and let $\lambda_{\max}(K), \lambda_{\min}(K)$ denote the largest and smallest elements of $\sigma(VR(K))$. 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 range of directional predictability within the series. We can give a further interpretation to these quantities. Consider a portfolio of assets with fixed weights $w \in \mathbb{R}^d$. Denoting $VR(K; w^\top X_t)$ by the univariate variance ratio of the portfolio $w^\top X_t$, while $\tilde{w} = \Sigma^{1/2}w$ and $Y_t = \Sigma^{-1/2}X_t$, we have (abusing the notation somewhat)

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

This follows because $VR(K; X_t) = VR(K; \Sigma^{-1/2} X_t) = VR(K; Y_t)$. This says that the largest eigenvalue of the variance ratio matrix is an upper bound on the 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. We may also be interested in the horizon K_{\max} for which this predictability is maximized.

We are also interested in several univariate parameters based on $VRd_+(K)$. First, the diagonal elements of $VRd_+(K)$ correspond to the univariate variance ratio statistics. Second, the off-diagonal elements of $VRd_+(K)$ provide the information about the directional lead lag pattern between the assets. Third, the differences between two corresponding off-diagonal elements of $VRd_+(K)$ 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 $VRd_+(K)$ and $VRd_-(K)$ 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 $VRd(K)$, which is

$$CS(K) = \frac{2}{d(d-1)} \sum_{i=1}^{d-1} \sum_{j=i+1}^d VRd_{ij}(K) = \frac{1}{d(d-1)} \{i^\top VRd(K)i - \text{tr}(VRd(K))\}, \quad (14)$$

see Bailey, Kapetanios, and Pesaran (2012) who consider the case $K = 0$ and large d . This measures in some average sense the cross dependence at different lags.⁴ It is also related to the expected profit of the Lo and MacKinlay (1990) portfolio momentum strategies (they chose weights $w_{it}(k) = -(1/d)(X_{i,t-k} - \bar{X}_{t-k})$, where \bar{X}_{t-k} is the equally weighted "market portfolio", and showed that the expected profit of this strategy $\pi(k) = \text{tr}(\Gamma(k))/d - i^\top \Gamma(k)i/d^2$, in the case where each asset has the same mean and variance).

3 Estimation

Suppose that we observe the return vectors $\{X_t, t = 1, \dots, T\}$ 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, $X_t(K) = X_t + X_{t-1} + \dots + X_{t+1-K}$, and the original observations and then forming the ratio.⁵ We can alternatively explicitly use the population connection with the autocorrelation matrix process in (10) for example.

We estimate the population quantities by sample averages:

$$\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t \quad ; \quad \hat{\Psi}(j) = \frac{1}{T} \sum_{t=j+1}^T (X_t - \bar{X})(X_{t-j} - \bar{X})^\top, \quad j = 0, 1, 2, \dots$$

$$\hat{\Sigma}(K) = \frac{1}{T} \sum_{t=K}^T (X_t(K) - K\bar{X})(X_t(K) - K\bar{X})^\top$$

$$\hat{\Sigma} = \hat{\Psi}(0) \quad ; \quad \hat{D} = \text{diag}[\hat{\Psi}(0)] \quad ; \quad \hat{\Gamma}(j) = \hat{\Sigma}^{-1/2} \hat{\Psi}(j) \hat{\Sigma}^{-1/2};$$

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

⁴We remark that Castura, Litzenberger, Gorelick, and Dwivedi (2010) report the average variance ratio of the Russell 1000 and Russell 2000 stocks, which amounts to reporting $\sum_{i=1}^d VRd_{ii}(K)/d$

⁵As 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. Theoretically, the two approaches can give similar inferences.

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

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

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

and likewise for $\widehat{VRd}(K)$, $\widehat{VR}(K, L)$, $\widehat{VRd}^{\&}(K)$, etc.

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

4 Asymptotic Theory and Inference

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 except 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.⁶ In the multivariate context, their assumption H4 would be that

$$E[\widetilde{X}_{it}\widetilde{X}_{jt}\widetilde{X}_{kr}\widetilde{X}_{ls}] = 0 \text{ for all } i, j, k, l, t, \text{ and } r, s \text{ with } r < s < t. \quad (15)$$

This assumption rules out leverage type effects, which may be important for some assets. This assumption is not necessary for the distribution theory; imposing it 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).

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

⁶Some papers including Whang and Kim (2003) dispense with this latter assumption but maintain the mixing and moment assumption.

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

$$Q(K) = \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_{j,K} c_{k,K} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \Xi_{jk} (\Sigma^{-1/2} \otimes \Sigma^{-1/2})$$

$$Qd(K) = \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_{j,K} c_{k,K} (D^{-1/2} \otimes D^{-1/2}) \Xi_{jk} (D^{-1/2} \otimes D^{-1/2})$$

$$Qa(K) = \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_{j,K} c_{k,K} (\Sigma^{-1} \otimes I) \Xi_{jk} (\Sigma^{-1} \otimes I)$$

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

ASSUMPTION A.

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

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

ASSUMPTION MH*.

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

MH2. \tilde{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 $\sup_t E|\tilde{X}_{it} \tilde{X}_{k,t-j}|^{2(r+\delta)} < \Delta < \infty$ for all $i, k = 1, \dots, d$;

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

In MH* we include the additional condition (ii) $E[\tilde{X}_t \tilde{X}_{t-j}^\top \otimes \tilde{X}_s \tilde{X}_{s-k}^\top] = 0$, for all $s \neq t$ and all $j, k = 1, \dots, K$; 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.⁷ 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). In section 6 below we will extend conditions A to allow for a time varying mean (that has to be estimated) and a time varying variance. 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 CLT's below are valid under only second moments, Brockwell and Davies (1991, Theorem 7.2.2), due to the self normalization present in the sample autocorrelations.

We next present our main result.

THEOREM 1. *Suppose that either Assumption A or MH* holds. Then,*

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

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 , $\frac{1}{2}\text{vech} (A + A^\top) = L\frac{1}{2} (I + K) \text{vec} (A) = D^+\text{vec} (A)$ where L and K are the so-called elimination and commutation matrices, respectively, and D^+ 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 the matrices. It now follows that

$$\sqrt{T}\text{vech} \left(\widehat{VR}(K) - I_d \right) \implies N(0, S(K)), \tag{16}$$

⁷In particular, the asymptotic variance of $\widehat{VR}_+(K)$, for example, becomes

$$\begin{aligned} Q^{LM}(K) &= \sum_{j=1}^{K-1} c_{j,K}^2 \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \Xi_{jj} \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \\ &\quad + \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_{j,K} c_{k,K} \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \Upsilon_{jk} \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \\ \Upsilon_{jk} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t \neq s} E \left[\left(\tilde{X}_{t-j} \tilde{X}_{s-k}^\top \otimes \tilde{X}_t \tilde{X}_s^\top \right) \right]. \end{aligned}$$

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

Limiting distributions for smooth functions of any of the above can be easily obtained via the delta method. However, as for the eigenvalues of the statistics, we need to employ a different approach as they are not smooth functions of the variance ratio matrix in general. Specifically, Eaton and Tyler (1991, Theorem 3.2) show that if the random symmetric matrix $\sqrt{T}(\widehat{VR}(K) - I_d)$ converges in distribution to a matrix random variable, denoted W , then with $i_d = (1, 1, \dots, 1)^\top$

$$\sqrt{T} \left(\varphi(\widehat{VR}(K)) - i_d \right) \implies \varphi(W), \quad (17)$$

where $\varphi(\widehat{VR}(K))$ and $\varphi(W)$ are $d \times 1$ vectors of ordered eigenvalues $\lambda_j \in \varphi(\widehat{VR}(K))$ and $\lambda_j^* \in \varphi(W)$, respectively. Using the continuous mapping theorem (and/or the delta method) on (11), we may also derive asymptotics for the functions of univariate eigenvalues. For instance,

$$\sqrt{T} \left(\sum_{j=1}^d \lambda_j - d \right) \implies \sum_{j=1}^d \lambda_j^*.$$

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

$$\widehat{\Xi}_{jk} = \frac{1}{T} \sum_{t=\max\{j,k\}+1}^T (X_{t-j} - \bar{X})(X_{t-k} - \bar{X})^\top \otimes (X_t - \bar{X})(X_t - \bar{X})^\top \quad (18)$$

$$\widehat{Q}(K) = \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_{j,K} c_{k,K} \left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2} \right) \widehat{\Xi}_{jk} \left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2} \right). \quad (19)$$

Similarly, we may define $\widehat{Qd}(K)$, replacing $\widehat{\Sigma}^{-1/2}$ by $\widehat{D}^{-1/2}$ in (19) and we may define $\widehat{Qa}(K)$, replacing $\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2}$ by $\widehat{\Sigma}^{-1} \otimes I$ in (19).

COROLLARY 1. *Suppose that either Assumption A or MH* holds, then the estimator $\widehat{Q}(K)$ is weakly consistent for $Q(K)$ (likewise, $\widehat{Qd}(K)$ and $\widehat{Qa}(K)$ are weakly consistent for $Qd(K)$ and $Qa(K)$), i.e.,*

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

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

multivariate generalization thereof) derived from

$$\widehat{Q}_{LM}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 \left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2} \right) \widehat{\Xi}_{jj} \left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2} \right) \quad (20)$$

reflects this structure. Similar results hold for $\widehat{Q}d_{LM}(K)$ and $\widehat{Q}a_{LM}(K)$. In the iid case, we further have $\Xi_{jj} = \Sigma \otimes \Sigma$ and: $Q_{iid}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 I_{d^2}$, $\widehat{Q}d_{iid}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 (\widehat{\Gamma}d(0) \otimes \widehat{\Gamma}d(0))$, and $\widehat{Q}a_{iid}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 (\widehat{\Sigma}^{-1} \otimes \widehat{\Sigma})$. In the scalar case these are all nuisance parameter free. As we show in the application, the standard errors 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 $E[\widetilde{X}_t \widetilde{X}_{t-j}^\top \otimes \widetilde{X}_s \widetilde{X}_{s-k}^\top] \neq 0$ 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.

Now we derive the asymptotic normality of the two parameter variance ratio statistics

$$\widehat{VR}_+^*(K, L) = \widehat{VR}(L)^{-1/2} \times \widehat{VR}(K) \times \widehat{VR}(L)^{-1/2}.$$

COROLLARY 2. *Suppose that Assumption A or MH* holds. Then,*

$$\sqrt{T} \text{vec} \left(\widehat{VR}_+^*(K, L) - I_d \right) \implies N(0, Q(K, L)),$$

where

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

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

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

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

We can compare the relative efficiency of the two parameter variance ratio estimator $\widehat{VR}_+^*(LJ, L)$ relative to the one parameter variance ratio estimator $\widehat{VR}_+(J)$, 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

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

This gives quite modest improvements in efficiency.

5 Alternative Hypotheses

There are many plausible alternative hypotheses to our null. We look in detail at several alternative models in this section.

5.1 Multivariate Fads Model

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

$$p_t^* = \mu + p_{t-1}^* + \varepsilon_t \tag{21}$$

$$p_t = p_t^* + \eta_t, \tag{22}$$

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

$$X_t = p_t - p_{t-1} = \varepsilon_t + \eta_t - \eta_{t-1}. \tag{23}$$

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 (2005)). In this case,

$$VR(K) = I + (1 - \frac{1}{K})(\Gamma(1) + \Gamma(1)^\top) = I - 2(1 - \frac{1}{K})(\Omega_\varepsilon + 2\Omega_\eta)^{-1/2} \Omega_\eta (\Omega_\varepsilon + 2\Omega_\eta)^{-1/2}.$$

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

$$X_t(K) = p_t - p_{t-K} = \sum_{s=t-K}^t \varepsilon_s + \sum_{s=t-K}^t (\eta_s - \eta_{s-1}) = \sum_{s=t-K}^t \varepsilon_s + \eta_t - \eta_{t-K}.$$

These have variance

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

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

THEOREM 2. *Suppose that the multivariate fads model (21)-(22) holds and suppose that $\text{cov}(\eta_{t+j}, \eta_t) \rightarrow 0$ as $j \rightarrow \infty$. Then, $VR(\infty) = \lim_{K \rightarrow \infty} VR(K) = I + \sum_{j=1}^{\infty} (\Gamma(j) + \Gamma(j)^\top)$ exists. Further suppose that $\Omega_\eta(1) > 0$. Then,*

$$VR(\infty) < I_d$$

in the matrix partial order sense. Likewise, $VRd(\infty) = \lim_{K \rightarrow \infty} VRd(K)$ exists, and

$$VRd(\infty) < \Gamma d(0).$$

This result generalizes the existing results for the scalar fads process, which amount to $VRd_{ii}(\infty) \leq \Gamma d_{ii}(0)$ for $i = 1, \dots, d$. In Theorem 2, we obtain stronger constraints on the off diagonal elements of $VRd(\infty)$ and $VR(\infty)$.

We consider what happens to the long horizon variance ratio statistic under the fads model. We will consider the case where $K \rightarrow \infty$ as $T \rightarrow \infty$ such that $K/T \rightarrow 0$ (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

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

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

$$\delta_t = E[\| (R(\dots, e_0, \dots, e_{t-1}, e_t) - R(\dots, e'_0, \dots, e_{t-1}, e_t)) \|^2],$$

where e'_t is an i.i.d. copy of e_t and $\|\cdot\|$ 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 3. *Suppose that the multivariate fads model (21)-(22) holds along with Assumption B. Then,*

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

Likewise, $\widehat{VRd}(K)$ consistently estimates $VRd(\infty)$. More generally, we could obtain the limiting distribution of $\widehat{VR}(K) - VR(K)$ 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.

5.2 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

$$p_t = p_{t-1}1(t < \tau_e) + \delta_T 1(\tau_e \leq t \leq \tau_f) p_{t-1} + \left(\sum_{s=\tau_f+1}^t \varepsilon_s + p_{\tau_f}^* \right) 1(t > \tau_f) + \varepsilon_t 1(t \leq \tau_f), \quad (24)$$

where $p_{\tau_f}^*$ represents the restarting price after the bubble collapses at time τ_f , and $\delta_T = 1 + c/T^\alpha$ for $\alpha \in (0, 1/2)$ 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., $(\tau_f - \tau_e)/T \rightarrow \tau_0 > 0$), that, as $T \rightarrow \infty$

$$\widehat{VR}(K) \xrightarrow{P} K \tag{25}$$

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 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.

5.3 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 $\tilde{X}_t(u) = X_t(u) - EX_t(u)$ and:

$$\Sigma(u) = \text{var}(X_t(u)) = E(\tilde{X}_t(u)\tilde{X}_t^\top(u))$$

$$D(u) = \text{diag} \left\{ E(\tilde{X}_{1t}^2(u)), \dots, E(\tilde{X}_{dt}^2(u)) \right\}$$

$$\Psi_u(j) = E(\tilde{X}_t(u)\tilde{X}_{t-j}^\top(u)).$$

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

$$\bar{\Gamma}(j) = \left(\int_0^1 \Sigma(u) du \right)^{-1/2} \int_0^1 \Psi_u(j) du \left(\int_0^1 \Sigma(u) du \right)^{-1/2}.$$

It follows that under local stationarity

$$\widehat{VR}(K) \xrightarrow{P} I + \sum_{j=1}^{K-1} \left(1 - \frac{j}{K} \right) (\bar{\Gamma}(j) + \bar{\Gamma}(j)^\top).$$

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

5.4 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 (10)-(12). Wright (2000) has proposed variance ratios based on signs and ranks that have similar objectives.

6 Time Varying Risk Premium and Calendar Time/Seasonal Effects

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 risk premium and its evolution over time. In general, one may have a parametric model for the vector of conditional means $\mu_t(\theta_0) = E(X_t | \mathcal{F}_{t-1})$. For example, Engle, Lilien and Robins (1987) consider a multivariate time series model consistent with the conditional CAPM where the dynamic risk premium is related

to the conditional covariance matrix of returns. We could estimate the parameters of a risk premium model and then compute the variance ratio statistics on the risk adjusted returns. We note that the details vary considerably according to the model adopted but generally the estimation of the risk premium parameters would affect the asymptotic distribution of the variance ratio statistics.

We focus on an alternative nonparametric framework, i.e., rolling windows. Specifically, suppose that $E(X_t|\mathcal{F}_{t-1}) = \mu_t$, where

$$\mu_t = \sum_{s=1}^{\tau} g_s(t/T)I_s(t), \quad (26)$$

where $g_s(\cdot)$ are continuously differentiable but unknown vector functions representing smooth trends that vary across $s = 1, \dots, \tau$ and $I_s(t) = 1(t \in J_s)$. We suppose that J_s form a mutually exclusive and exhaustive partition of the sample, i.e., $\{1, \dots, T\} = \cup_{s=1}^{\tau} J_s$ with $J_s \cap J_r = \emptyset$ for $r \neq s$. Furthermore, we shall suppose that the categories J_s are of the same order of magnitude, i.e., $\#J_s = T_s$ such that $T_s/T \rightarrow c_s$ for all $s = 1, \dots, \tau$ with τ fixed and $c_s \in (0, \infty)$. The trends capture the idea that the risk premium is slowly varying, like Dimson, Marsh, and Staunton (2008), but precisely how this is intermediated through the partition can represent a variety of phenomenon. We think of three main cases. In the first case, τ could be the known period of a common seasonal component and $I_s(t) = 1(t = k\tau + s \text{ for some } k)$ are then seasonal dummies, Vogt and Linton (2014). The second case is to classify observations according to how many calendar periods since a previous transaction price was observed, so that a regular Monday closing price would be three days since the last closing price was observed. This allows one to take account of public holidays like Easter and Christmas that vary over day of the week, as encountered in French and Roll (1986). These quasi seasonal effects could be consistent with a calendar time interpretation of the returns process and therefore also represent the rational part of the stock price variation. The final case is where the sets J_s are contiguous blocks of time in which case the model is capturing structural change (the change points are assumed to be known).

We suppose that for each J_s we can order the observation times $t_{s1} < t_{s2} < \dots < t_{sT_s}$. Define the set of time points $H_s(t, M, K) = \{t_{sj} : t_{sj^*-M}, \dots, t_{sj^*}, \text{ where } j^* = \arg \max_j t_j < t - K\} \cap J_s$ with cardinality $M_t \leq M$ (at interior points $M_t = M$) and then let:

$$\hat{\mu}_t = \sum_{s=1}^{\tau} \hat{g}_s(t/T)I_s(t) \quad (27)$$

$$\hat{g}_s(t/T) = \frac{1}{M_t} \sum_{t_{sj} \in H_s(t, M, K)} X_{t_{sj}}.$$

In other words, for each s we smooth over time using just the observations in J_s . At the beginning of the sample, we generally have fewer observations, which necessitates the edge adjustment used above, although in our application we actually have a presample that makes such edge adjustment unnecessary (so that $M_t = M$ for all t). In the purely periodic (seasonal) case (ignoring the edge effect) the notation can be simplified somewhat so that $\hat{\mu}_t = \sum_{m=1}^M X_{t-(m-\lfloor K/\tau \rfloor)\tau}/M$, where $\lfloor a \rfloor$ denotes the greatest integer strictly less than $a \in \mathbb{R}$. We could consider more general kernel based estimators, but have not done so here.

Consequently, the estimator for the autocovariance matrix and variance ratio are as follows:

$$\begin{aligned} \widehat{\Psi}(j) &= \frac{1}{T} \sum_{t=j+1}^T \left(X_t - \hat{\mu}_t \right) \left(X_{t-j} - \hat{\mu}_{t-j} \right)^\top, \quad j = 0, 1, 2, \dots, \\ \widehat{\Sigma} &= \widehat{\Psi}(0) \quad ; \quad \widehat{\Gamma}(j) = \widehat{\Sigma}^{-1/2} \widehat{\Psi}(j) \widehat{\Sigma}^{-1/2}; \\ \widehat{VR}_+(K) &= I + 2 \sum_{j=1}^{K-1} \left(1 - \frac{j}{K} \right) \widehat{\Gamma}(j). \end{aligned}$$

We next discuss the asymptotic properties of this modified variance ratio statistic. We require some additional assumptions

ASSUMPTION C.

C1. We suppose that for each J_s we can order the observation times: $t_{s1} < t_{s2} < \dots < t_{sT_s}$ such that

$$\max_{1 \leq j \leq T_s - 1} |t_{sj} - t_{s(j+1)}| \leq \frac{C}{T}$$

for some $C < \infty$;

C2. The functions g_s are continuously differentiable on $[0, 1]$, for all $s = 1, \dots, \tau$.

C3. $M = T^\alpha$ with $1/2 < \alpha < 3/4$.

C4. There is some deterministic family of covariance matrices Ω_t , with $0 < \inf_{t \geq 1} \lambda_{\min}(\Omega_t) \leq \sup_{t \geq 1} \lambda_{\max}(\Omega_t) < \infty$, such that

$$\widetilde{X}_t = \Omega_t^{-1/2} (X_t - \mu_t)$$

is stationary and ergodic (and a martingale difference sequence) and satisfies assumptions A.

Furthermore, for $j, k = 1, 2, \dots, K$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=K+1}^T \left(\Omega_{t-j}^{1/2} \otimes \Omega_t^{1/2} \right) \otimes \left(\Omega_{t-k}^{1/2} \otimes \Omega_t^{1/2} \right) = W_{jk} < \infty.$$

Assumption C1 means that the information accumulates in the usual way so that there are no "holes" in the categories J_s . Assumption C4 allows for unconditional heteroskedasticity of general form and of course conditional heteroskedasticity is also allowed in \tilde{X}_t . The rate condition C3 on M is tied to the specific implementation and the smoothness condition (one derivative) that we have adopted, and can be weakened under additional restrictions elsewhere. We note that we still only require four moments and do not assume mixing conditions. We have to establish uniform consistency of $\hat{g}_s(u)$ over $u \in [0, 1]$, and we use an exponential inequality for martingales of de la Peña (1999) to establish this under our weak conditions.

Define $Q^*(K)$ as:

$$Q^*(K) = \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_{j,K} c_{k,K} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \tilde{\Xi}_{jk} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}),$$

$$\tilde{\Xi}_{jk} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=\max\{j,k\}+1}^T \left(\Omega_{t-j}^{1/2} \otimes \Omega_t^{1/2} \right) E[\tilde{X}_{t-j} \tilde{X}_{t-k}^\top \otimes \tilde{X}_t \tilde{X}_t^\top] \left(\Omega_{t-k}^{1/2} \otimes \Omega_t^{1/2} \right).$$

Similarly define $\hat{Q}^*(K)$ as $\hat{Q}(K)$ in (19) but with

$$\hat{\Xi}_{jk} = \frac{1}{T} \sum_{t=\max\{j,k\}+1}^T (X_{t-j} - \hat{\mu}_{t-j}) (X_{t-k} - \hat{\mu}_{t-k})^\top \otimes (X_t - \hat{\mu}_{t-j}) (X_t - \hat{\mu}_{t-k})^\top.$$

We may similarly define $Qd^*(K)$, $Qa^*(K)$, and $\hat{Q}d^*(K)$ and $\hat{Q}a^*(K)$.

THEOREM 4. *Suppose that assumption C holds. Then,*

$$\sqrt{T} \text{vec} \left(\widehat{VR}_+(K) - I_d \right) \implies N(0, Q^*(K))$$

$$\hat{Q}^*(K) \xrightarrow{P} Q^*(K).$$

Likewise for $\hat{Q}d^*(K)$ and $\hat{Q}a^*(K)$. It follows that essentially the same critical values may be used for any of the test statistics considered above.

We note that this methodology is different from rolling window variance ratio or autocorrelation tests (for example, Lo (2005)). We are only using the rolling window to take care of slowly varying trends or periodic components; we estimate the short term predictability using the whole sample and compare it to a confidence interval obtained under the null hypothesis that precludes predictability. The full rolling window analysis could be analyzed under Theorem 1 but with a smaller sample size (at least for the "pointwise case").

7 Application

We apply our methodology to U.S. and U.K. 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 27/12/2013⁸, and the weekly stock returns for FTSE100 and FTSE250 from 13/01/1986 to 03/03/2014⁹. We investigate whether there has been a substantial change in the variance ratios at short-to-medium horizon over time. We analyze the effects of time-varying risk premium and seasonal effects. We also look at the variance ratios at the long horizon.

7.1 Short to Medium Horizon

According to the results of Theorem 1 and Corollary 1, we give the following testing statistics

$$[Zd(K)]_{ij} = \sqrt{T} \left(R_l^T \widehat{Q}d(K) R_l \right)^{-1/2} \left[R_l^T \text{vec} \left(\widehat{V}Rd_+(K) - \widehat{\Gamma}d(0) \right) \right] \implies N(0, 1)$$

where R_l is a $d^2 \times 1$ vector taking the value 1 at the l^{th} place and 0 at the other places, $1 \leq l \leq d^2$, $i, j = 1, \dots, d$. $[Zd_{LM}(K)]_{ij}$ and $[Zd_{iid}(K)]_{ij}$ are defined similarly but using $\widehat{Q}d_{LM}(K)$ and $\widehat{Q}d_{iid}(K)$ respectively. These statistics can be used to test the specific elements of $\widehat{V}Rd_+(K) - \widehat{\Gamma}d(0)$ matrix. For example, we can test $[VRd_+(K)]_{22} = 1$ using the statistic $[Zd(K)]_{22}$ by setting $l = 5$ and $i = j = 2$.

We first test for the absence of serial correlation in each of three weekly size-sorted equal-weighted portfolio returns (smallest quantile, central quantile, and largest quantile). We 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). Based on the multivariate variance ratio statistics $VRd_+(K)$, we test a series of hypotheses: $[VRd_+(K)]_{ii} = 1$ for $i = 1, 2, 3$, using the statistic $[Zd(K)]_{ii}$, $[Zd_{LM}(K)]_{ii}$ and $[Zd_{iid}(K)]_{ii}$ by setting $l = 1, 5, 9$. Table 1-A reports the results for the portfolio of small-size firms, Table 1-B reports the results for the portfolio of medium-size firms, and Table 1-C reports the results for the portfolio of large-size firms. We examine $K = 2, 4, 8, 16$ as in Campbell, Lo and Mackinlay (1997).

⁸The 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.

⁹The weekly price data are obtained from Yahoo Finance. The weekly returns are calculated from the close prices.

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

Sample period	# of obs	Lags			
		$K = 2$	$K = 4$	$K = 8$	$K = 16$
A. Portfolio of firms with market values in smallest CRSP quintile					
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			
		$K = 2$	$K = 4$	$K = 8$	$K = 16$
B. Portfolio of firms with market values in central CRSP quintile					
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			
		$K = 2$	$K = 4$	$K = 8$	$K = 16$
C. Portfolio of firms with market values in largest CRSP quintile					
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)
78:10:06—94:12:23	847	1.59	(2.33)*	(2.06)*	(1.29)
		1.03	1.06	1.08	1.01
		(0.63)	(0.61)	(0.54)	(0.03)
94:12:30—13:12:27	992	(0.63)	(0.65)	(0.59)	(0.04)
		(0.95)	(0.91)	(0.75)	(0.04)
		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 $\widehat{V}Rd_+(K)$. Test statistics ($[Zd(K)]_{ii}$, $[Zd_{LM}(K)]_{ii}$ and $[Zd_{iid}(K)]_{ii}$) 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. 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.

We then implement the procedure from section 6 using daily data and the day of the week dummy categorization we discussed there. We divide our data into five ($\tau = 5$) categories: Monday, Tuesday, Wednesday, Thursday, and Friday series. We take M to be 522 and calculate the time-varying risk premium, $\widehat{\mu}_t$ in the rolling window of 522 weeks (10 years). Below is shown the average common trend for each portfolio, which shows considerable time series variation, especially for the small-size portfolio.

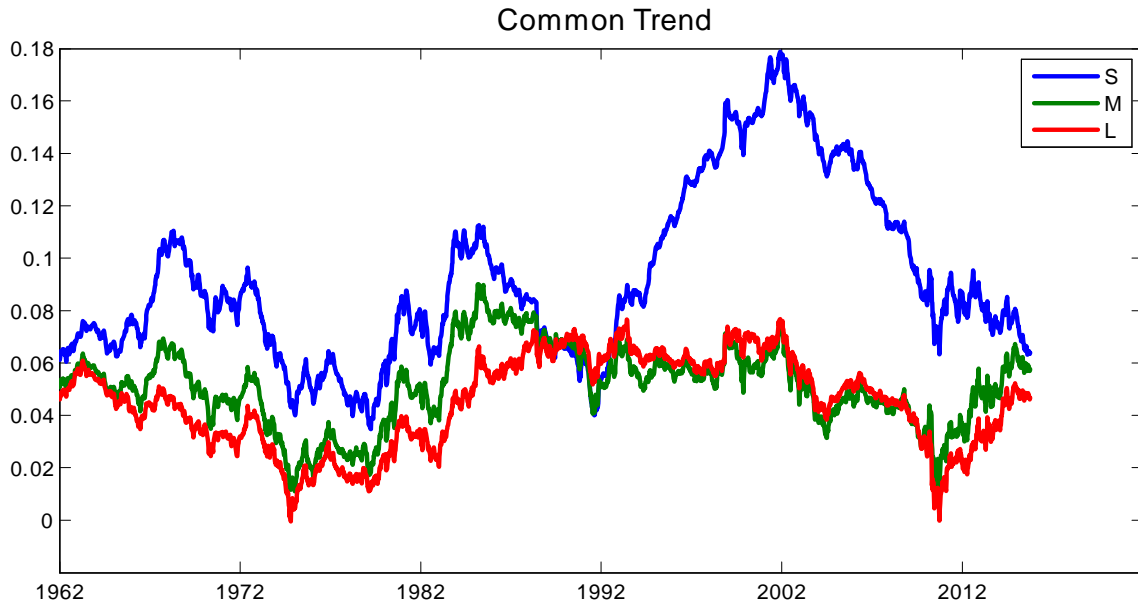


Figure 1: Average common trend for small-size, medium-size and large-size portfolios.

We then use the risk adjusted returns and carry out the variance ratio tests again in the same way as before, but only consider the $[Zd(K)]_{ii}$ statistics in two subsamples: 62:07:06-94:12:23 and 94:12:30-13:12:27. Remarkably, the results shown in Table 1-D do not change much compared with the results obtained by using the constant-mean adjusted returns (Table 1-A,B,C).

Table 1-D: Variance ratios (time-varying mean) for weekly size-sorted portfolio returns

Sample period	# of obs	Lags			
		$K = 2$	$K = 4$	$K = 8$	$K = 16$
A. Portfolio of firms with market values in smallest CRSP quantile					
62:07:06—94:12:23	1695	1.43 (10.86)*	1.95 (11.25)*	2.57 (10.22)*	2.98 (8.56)*
94:12:30—13:12:27	992	1.21 (3.34)*	1.47 (3.61)*	1.71 (3.39)*	1.83 (2.55)*
B. Portfolio of firms with market values in central CRSP quantile					
62:07:06—94:12:23	1695	1.23 (5.94)*	1.46 (6.20)*	1.68 (5.46)*	1.76 (3.92)*
94:12:30—13:12:27	992	1.00 (-0.01)	1.05 (0.40)	1.02 (0.12)	0.90 (-0.36)
C. Portfolio of firms with market values in largest CRSP quantile					
62:07:06—94:12:23	1695	1.04 (1.19)	1.10 (1.57)	1.15 (1.36)	1.11 (0.61)
94:12:30—13:12:27	992	0.94 (-0.99)	0.94 (-0.46)	0.89 (-0.53)	0.81 (-0.63)

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

We may wish to test whether the variance ratio has "improved" significantly from one period (A) to the next (B). We may consider, for example, the statistic

$$\tau_{AB} = \left(\widehat{VRd}_+^A(K) - \widehat{\Gamma d}^A(0) \right) - \left(\widehat{VRd}_+^B(K) - \widehat{\Gamma d}^B(0) \right), \quad (28)$$

where $\widehat{VRd}_+^j(K)$ and $\widehat{\Gamma d}^j(0)$ denotes 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 $\left[\widehat{VRd}_+^A(K) \right]_{ii} - \left[\widehat{VRd}_+^B(K) \right]_{ii}$ 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 then test zero cross-autocorrelation (no lead-lag relationship) between returns of different size portfolios. Based on the multivariate ratio statistic $VRd_+(K)$, we test the hypothesis that $[VRd_+(K) - \Gamma d(0)]_{ij} = 0$, for $i, j = 1, 2, 3$, $i \neq j$, using the statistic $[Zd(K)]_{ij}$ by setting $l = 2, 3, 4, 6, 7, 8$.

Table 2: Lead-lag patterns between weekly size-sorted portfolio returns

Lags	Sample period	$\widehat{VRd}_+(K) - \widehat{\Gamma d}(0)$		To		
		From	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 $\widehat{VRd}_+(k) - \widehat{\Gamma d}(0)$ 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 a series of hypotheses: $[VRd_+(K) - \Gamma d(0)]_{ij} - [VRd_+(K) - \Gamma d(0)]_{ji} = 0$, for $i, j = 1, 2, 3$, $i > j$. Results are reported in Table 3.

Table 3: Asymmetry of lead-lag patterns

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

We finally test for the absence of serial correlation for the vector of returns, based on eigenvalues of multivariate variance ratio statistic $VR(K)$. We consider the following two test statistics:

$$Z_1(K) = \sum_{i=1}^d \sqrt{T} \lambda_i^{(B)} \implies \sum_{i=1}^d \lambda_i^{(W)}$$

$$Z_2(K) = \prod_{i=1}^d \sqrt{T} \lambda_i^{(B)} \implies \prod_{i=1}^d \lambda_i^{(W)}$$

where $\lambda_{i=1,\dots,d}^{(B)}$ are ordered eigenvalues of matrix B , and $\lambda_{i=1,\dots,d}^{(W)}$ are ordered eigenvalues of matrix W . B and W are symmetric $d \times d$ matrix such that

$$\text{vech}(B) = \widehat{S}(K)^{-1/2} \text{vech}\left(\widehat{VR}(K) - I_d\right)$$

$$\text{vech}(W) \sim N(0, I_{d(d+1)/2})$$

In Table 4, we report the eigenvalues, test statistics and the associated p-values in two sub-samples.

Table 4: Tests based on eigenvalues

	$K = 2$	$K = 4$	$K = 8$	$K = 16$
62:07:06-94:12:23				
Eigenvalues	[0.84 1.14 1.52]	[0.77 1.29 2.21]	[0.70 1.38 3.04]	[0.66 1.40 3.64]
$Z_1(K)$	12.32* (0.00)	14.34* (0.00)	13.01* (0.00)	11.24* (0.00)
$Z_2(K)$	-318.57* (0.00)	-215.11* (0.00)	-129.87* (0.00)	-51.80* (0.00)
94:12:30-13:12:27				
Eigenvalues	[0.86 0.91 1.32]	[0.82 0.89 1.75]	[0.75 0.83 2.21]	[0.62 0.84 2.62]
$Z_1(K)$	2.90 (0.09)	3.43 (0.05)	2.85 (0.10)	2.24 (0.19)
$Z_2(K)$	27.46*(0.00)	31.05*(0.00)	24.83*(0.00)	17.42*(0.01)

Test statistics marked with asterisks indicate that the variance ratios are statistically different from I_d at 5% level of significance. p-values are reported in parentheses. The empirical quantiles of the statistics are obtained by simulation.

As before, we find the magnitude of the effect and its statistical significance has reduced in the later period.

We also examine the behavior of the variance ratio statistics on UK stocks. As Dimson, Marsh, and Staunton (2008) argue, the United States market has had relatively good performance over the long term compared with the most of the rest of the world. We look at weekly returns on the FTSE100 (Large cap) and FTSE250 (Mid Cap) indexes from 86:01:13 to 14:03:03. The results are shown below.

Table 5: Variance ratios for weekly returns in FTSE100 and FTSE250

Sample period	# of obs	Lags			
		$K = 2$	$K = 4$	$K = 8$	$K = 16$
A. FTSE100					
86:01:13—94:12:19	467	1.09 (1.07)	1.29 (1.77)	1.36 (1.52)	1.18 (0.56)
94:12:28—14:03:03	1002	0.92 (-1.67)	0.89 (-1.13)	0.8 (-1.29)	0.73 (-1.19)
B. FTSE250					
86:01:13—94:12:19	467	1.23 (2.36)*	1.62 (2.89)*	1.78 (2.63)*	1.54 (1.41)
94:12:28—14:03:03	1002	1.04 (0.73)	1.13 (1.06)	1.19 (0.98)	1.21 (0.72)

For FTSE100 (large cap), the variance ratios in both sample periods are insignificantly different from one, and we observe that the ratios are all below one in the later period, which is consistent with results for large-size CRSP portfolio. For FTSE250 (medium cap), the departure from the random walk model is statistically significant for the early sample period, but not so for the later sample period. We then test zero cross-autocorrelation (no lead-lag relationship) between FTSE100 and FTSE250 returns. The results are reported in Table 6.

Table 6: Lead-lag patterns between weekly returns in FTSE100 and FTSE250

Lags	Sample period	$\widehat{VRd}_+(K) - \widehat{\Gamma d}(0)$ From	To	
			FTSE100	FTSE250
K=2	86:01:13—94:12:19	FTSE100		0.27 (2.35)*
		FTSE250	0.06 (1.00)	
		Difference	0.19 (2.43)*	
	94:12:28—14:03:03	FTSE100		0.08 (1.47)
		FTSE250	-0.07 (-1.42)	
		Difference	0.15 (2.92)*	
K=4	86:01:13—94:12:19	FTSE100		0.7 (3.11)*
		FTSE250	0.21 (1.34)	
		Difference	0.49 (3.01)*	
	94:12:28—14:03:03	FTSE100		0.21 (2.01)*
		FTSE250	-0.12 (-1.02)	
		Difference	0.33 (3.30)*	
K=8	86:01:13—94:12:19	FTSE100		0.93 (3.17)*
		FTSE250	0.22 (0.86)	
		Difference	0.71 (3.49)*	
	94:12:28—14:03:03	FTSE100		0.26 (1.68)
		FTSE250	-0.2 (-0.99)	
		Difference	0.46 (3.04)*	
K=16	86:01:13—94:12:19	FTSE100		0.8 (2.1)*
		FTSE250	-0.07 (-0.2)	
		Difference	0.87 (3.4)*	
	94:12:28—14:03:03	FTSE100		0.3 (1.31)
		FTSE250	-0.29 (-0.95)	
		Difference	0.59 (2.97)*	

The results suggest there are strong lead-lag relationships, where FTSE100 leads and FTSE250 lags at all horizons for early sample periods, but only at horizon $k = 4$ for later sample period. With regard to cross-autocorrelation between current returns of FTSE100 and past returns of FTSE250, we find the values become negative in later sample period, but they are statistically inseparable from this value for all sample periods. We also find significant asymmetry of these lead-lag patterns.

7.2 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)

$$\widehat{VR}^{bc}(K) = \widehat{VR}(K) \left\{ 1 + \frac{K-1}{T} \right\} \quad ; \quad \widehat{VRd}^{bc}(K) = \widehat{VRd}(K) \left\{ 1 + \frac{K-1}{T} \right\}. \quad (29)$$

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

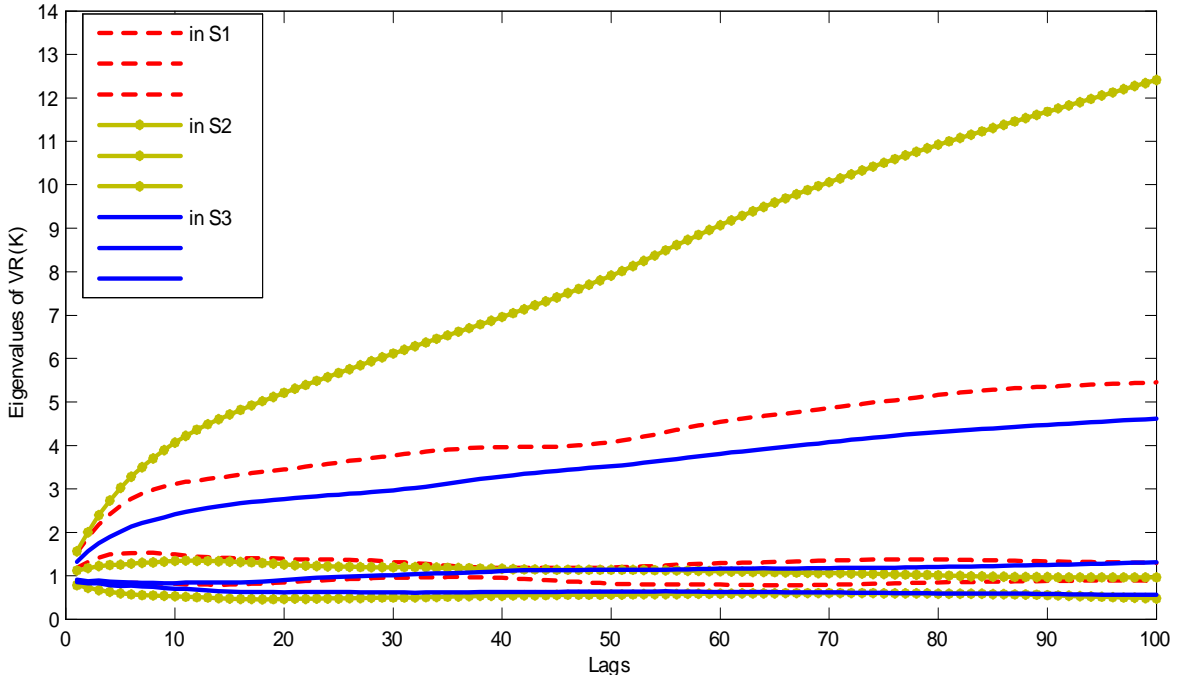


Figure 2: The eigenvalues of the variance ratio for weekly CRSP size-sorted portfolio returns in three sub-samples as a function of lags.

We see that the largest eigenvalue increases steadily out to the two year horizon we consider in all three subperiods, with the last subperiod having 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.

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

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

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.

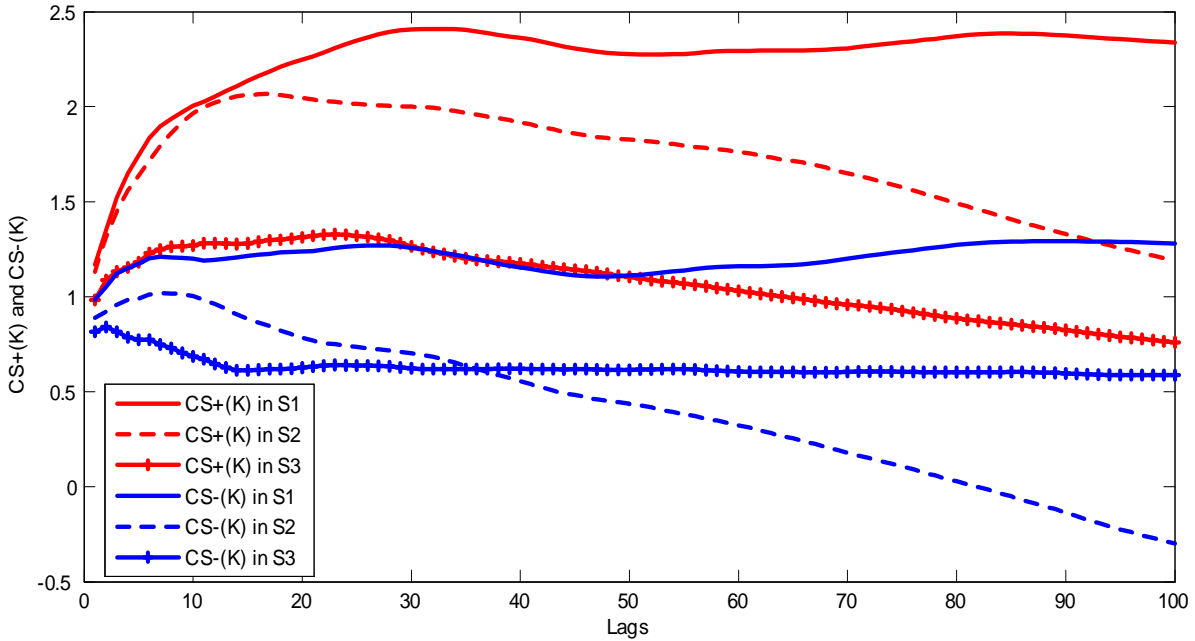


Figure 3: $CS_+(K)$ and $CS_-(K)$ statistics for weekly size-sorted CRSP portfolio returns in three sub-samples as a function of lags.

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.

8 Simulation Study

8.1 Size

To investigate how our procedures work in practice, we perform a small simulation study for the $\widehat{VR}(K)$ and $\widehat{VRd}_+(K)$ 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 $N(0, I_d)$. We simulate the null $H_0^{(2)}$ by generating the data from a diagonal multivariate ARCH model,

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

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

where $\varepsilon_t \sim \text{i.i.d.} N(0, I_d)$, $\varpi = I_d$ and $\alpha = 0.5I_d$. 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 $\widehat{S}_{iid}(K)$

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

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 $\text{vech}(W) \sim N(0, I_{d(d+1)/2})$.

Table 8-1: 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 8-2 and Table 8-3 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 8-2: Empirical size of nominal 5% variance ratio tests of the null hypothesis $H_0^{(1)}$

Sample size	K	d	Size of 5 percent test			
			$Z_1^{(iid)}(K)$	$Z_1(K)$	$Z_2^{(iid)}(K)$	$Z_2(K)$
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 8-3: Empirical size of nominal 5% variance ratio tests of the null hypothesis $H_0^{(2)}$

Sample size	K	d	Size of 5 percent test			
			$Z_1^{(iid)}(K)$	$Z_1(K)$	$Z_2^{(iid)}(K)$	$Z_2(K)$
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 8-2 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 8-3, 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 8-4 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 8-4: 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

8.2 Power

Consider the following model:

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

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

$$\eta_t = \beta\eta_t + \xi_t$$

where $\varepsilon_t \sim \text{i.i.d.}(0, \Omega_\varepsilon)$, $\xi_t \sim \text{i.i.d.}(0, \Omega_\xi)$. As shown in Fama and French (1998) for univariate case, if $\beta < 1$, we have $\widehat{VR}(K) < I_d$. 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 $\widehat{VR}(K) > I_d$. 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 8-5: 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	$\beta = 0.85$		$\beta = 1.01$	
			$Z_1^{(iid)}(K)$	$Z_2^{(iid)}(K)$	$Z_1^{(iid)}(K)$	$Z_2^{(iid)}(K)$
1024	16	3	0.9995	0.6349	1.0000	0.9971

Table 8-5 shows that the variance ratio tests using $Z_1^{(iid)}(K)$ and $Z_2^{(iid)}(K)$ are powerful against these alternatives.

8.3 Dating the Origination and Collapse of an Explosive Episode

We use a data generating mechanism that allows for the possibility of a single explosive episode as introduced in Section 5,

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

We assume $\varepsilon_t \sim i.i.d.N(0, \Omega_\varepsilon)$, all series catch up in the same bubble, where $\tau_e = [Tr_e]$, $\tau_f = [Tr_f]$, $p_{\tau_f}^* = p_{\tau_e} + p^*$. We simulate the data $\{X_t : t = 1, 2, \dots, \tau = [Tr]\}$ by setting: $d = 3$, $T = 1000$, $K = 16$, $\Omega_\varepsilon = I_d$, $\delta = 1.04$, $p^* = 0$, $r_e = 0.4$, $r_f = 0.6$. The minimum amount of data used for calculating the variance ratio test statistic is $\tau_0 = [Tr_0]$ with $r_0 = 0.3$. We date the origination of the explosive episode as $\hat{\tau}_e = [T\hat{r}_e]$ and the the collapse of the explosive episode as $\hat{\tau}_f = [T\hat{r}_f]$, where, for $i = 1, 2$ and $h = 1, 2$:

$$\hat{r}_e = \inf_{j \geq r_0} \{j : [Z_i^{(iid)}(K)]_j > C_{h,(0.975)}\} \quad ; \quad \hat{r}_f = \inf_{j \geq \hat{r}_e} \{j : C_{h,(0.025)} \leq [Z_i^{(iid)}(K)]_j \leq C_{h,(0.975)}\}.$$

Here, $C_{1,(0.025)} = -3.4047$, $C_{1,(0.975)} = 3.3841$, $C_{2,(0.025)} = -7.9355$, $C_{2,(0.975)} = 7.9863$ are simulated critical values of $Z_1^{(iid)}(K)$ and $Z_2^{(iid)}(K)$. Based on 1000 replications, we have the following results.

Table 8-6: Estimates of r_e and r_f based on the variance ratio tests

	$Z_1^{(iid)}(K)$		$Z_2^{(iid)}(K)$	
	\hat{r}_e	\hat{r}_f	\hat{r}_e	\hat{r}_f
Mean	0.4135	0.6000	0.4565	0.6192
Std	0.0122	0.0000	0.1091	0.0783
RMSE	0.0182	0.0000	0.1228	0.0806

Table 8-6 shows that the estimation of \hat{r}_e and \hat{r}_f based on $Z_1^{(iid)}(K)$ and $Z_2^{(iid)}(K)$ statistics are very close to their true values.

9 Conclusions

The multivariate variance ratio provides another way of seeing the cross correlation behaviour of asset returns. The long horizon properties depend on the alternative hypothesis considered and we

consider some cases where such characterization is possible. Our empirical work reports that the stock portfolios (especially the small cap ones) seem to have come closer to the efficient markets prediction, although, especially for small caps, there remains some linear predictability, although whether that is exploitable or not 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.

Our main practical point is to consider confidence intervals that are natural under the martingale hypothesis and do not require an additional no leverage/symmetric distribution assumption maintained in Lo and MacKinlay (1988) and in much subsequent work. These confidence intervals are larger but more credible with regard to the data generating process.

We remark that this theory is predicated on the existence of fourth moments, which may be problematic for some financial time series. Provided the population variance exists, the variance ratio converges in probability to one, but may have a non-standard limiting distribution and a slower rate of convergence to it, Mikosch and Starica (2000).¹⁰ 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 $T \rightarrow \infty$, $\widehat{VR}(K) \rightarrow K^{(2-\alpha)/\alpha}$ for all K . This is similar asymptotic behaviour to what is found under the bubble process of section 5.2 when $\alpha = 1$. Wright (2000) has proposed variance ratios based on signs and ranks that are robust to heavy tailed distributions.

10 Appendix

PROOF OF THEOREM 1. We first attempt the proof under assumption A.

¹⁰For stationary linear processes, the sample autocorrelations can be root-T consistent and asymptotically normal under only second moment assumptions, Brockwell and Davies (1991, Theorem 7.2.2), but this result does not hold for nonlinear processes like GARCH.

Due to uncorrelatedness of the martingale difference, for $j = 1, \dots, K$ we have

$$\begin{aligned}
\sqrt{T} \left(\text{vec}(\widehat{\Psi}(j)) - \text{vec}(\Psi(j)) \right) &= \frac{1}{\sqrt{T}} \sum_{t=j+1}^T (X_{t-j} - \bar{X}) \otimes (X_t - \bar{X}) \\
&= \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left(\tilde{X}_{t-j} \otimes \tilde{X}_t \right) - \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \tilde{X}_{t-j} \otimes (\bar{X} - \mu) \\
&\quad - (\bar{X} - \mu) \otimes \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \tilde{X}_t + \frac{T-j}{\sqrt{T}} (\bar{X} - \mu) \otimes (\bar{X} - \mu) \quad (30) \\
&= \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left(\tilde{X}_{t-j} \otimes \tilde{X}_t \right) + o_p(1), \quad (31)
\end{aligned}$$

where in (30) we made use of $\sum_{t=j+1}^T \tilde{X}_t = O_p(\sqrt{T})$, a result implied by the CLT for stationary ergodic martingale difference, e.g. Hayashi (2000). The CLT is justified by the fact that the difference $|\sqrt{T}^{-1}(\sum_{t=1}^T \tilde{X}_t - \sum_{t=j+1}^T \tilde{X}_t)| = o_p(1)$; 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 $\tilde{X}_t \tilde{X}_t^\top$ is a measurable transformation of \tilde{X}_t it is again stationary ergodic, (although it does not have a martingale difference structure anymore). Therefore, we may apply Birkhoff's ergodic theorem on $T^{-1} \sum_{t=1}^T \tilde{X}_t \tilde{X}_t^\top$, yielding $\widehat{\Sigma} - \Sigma = o_p(1)$, and then $\widehat{\Sigma}^{-1/2} - \Sigma^{-1/2} = o_p(1)$ by the continuous mapping theorem. Consequently, for each j we have

$$\begin{aligned}
\text{vec}(\widehat{\Gamma}(j)) &= \text{vec} \left(\left[\widehat{\Sigma}^{-1/2} - \Sigma^{-1/2} + \Sigma^{-1/2} \right] \widehat{\Psi}(j) \left[\widehat{\Sigma}^{-1/2} - \Sigma^{-1/2} + \Sigma^{-1/2} \right] \right) \\
&= \text{vec} \left(\Sigma^{-1/2} \widehat{\Psi}(j) \Sigma^{-1/2} \right) + T^{-1/2} O_p(1) \cdot o_p(1) \\
&= (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \text{vec}(\widehat{\Psi}(j)) + o_p(T^{-1/2}). \quad (32)
\end{aligned}$$

Next we observe that

$$\begin{aligned}
\sqrt{T} \text{vec} \left(\widehat{V} R_+(K) - I_d \right) &= \sqrt{T} \cdot \sum_{j=1}^{K-1} 2 \left(1 - \frac{j}{K} \right) \cdot \text{vec} \left(\widehat{\Gamma}(j) \right) \\
&= \sum_{j=1}^{K-1} c_j (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \cdot \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \widetilde{X}_{t-j} \otimes \widetilde{X}_t + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\sum_{j=1}^{K-1} c_j (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) (\widetilde{X}_{t-j} \otimes \widetilde{X}_t) \right] + o_p(1) \\
&=: \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{tj} + o_p(1). \tag{33}
\end{aligned}$$

Now for any constant vector $a = (a_1, \dots, a_{d^2})^\top \in \mathbb{R}^{d^2}$ we note that $a^\top Z_{tj}$ is a one-dimensional martingale difference sequence because we have $E[\widetilde{X}_t \widetilde{X}_{t-j}^\top | \mathcal{F}_{t-1}] = E[\widetilde{X}_t | \mathcal{F}_{t-1}] \widetilde{X}_{t-j}^\top$ a.s. for all $j \geq 1$. Consequently, since the moment condition A2 ensure that

$$E(a^\top Z_{tj})^2 = a^\top \text{var}(Z_{tj}) a = a^\top \left[\sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_j c_k (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \Xi_{jk} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \right] a < \infty,$$

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

$$a^\top \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{tj} \right) \implies N \left(0, a^\top \left[\sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_j c_k (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \Xi_{jk} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \right] a \right). \tag{34}$$

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

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

completing the proof.

Deriving the limiting distribution for the same statistic under assumption B closely follows similar arguments. Firstly, we note that the expansion for $\sqrt{T}(\text{vec}(\widehat{\Psi}(j)) - \text{vec}(\Psi(j)))$ is still valid due to uncorrelatedness ensured by assumption MH1. Moreover the summations in the second, third and fourth terms in (30) are still bounded in probability due to CLT for mixing sequence, Herrndorf (1985, Theorem 0) whose regularity conditions are satisfied by MH1-MH3. As a consequence, we end up with (31) as before. Finally, condition MH2 and MH3 allow for the Law of Large Numbers for mixing variables, White (1984, Corollary 3.48), yielding (32) and (33) as before.

Now we are only left with verifying (34). Since any measurable transformation of \tilde{X}_t preserves the mixing property with the same rate specified in MH2, for any d^2 -dimensional constant vector a Herndorf's CLT gives

$$a^\top \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T Z_{tj} \right) \Rightarrow N \left(0, a^\top \left[\sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_j c_k (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \Xi_{jk} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \right] a \right).$$

with $\Xi_{jk} = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E[\tilde{X}_{t-j} \otimes \tilde{X}_t][\tilde{X}_{t-k} \otimes \tilde{X}_t]^\top$, provided that the following regularity conditions are ensured: $E(a^\top Z_{tj}) = 0$, $\sup_t E|a^\top Z_{tj}|^\beta < \infty$ for some $\beta > 2$ and finally

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} E \left(\sum_{t=1}^T a^\top Z_{tj} \right)^2 &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \text{var} (a^\top Z_{tj}) \\ &= a^\top \left[\sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_j c_k (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \Xi_{jk} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \right] a \end{aligned}$$

is positive and finite.

The first condition is trivial by MH1, and the second and third conditions are satisfied by MH2 and MH3, respectively. The rest of the arguments are exactly the same as before, completing the proof.

Similar arguments apply to the other statistics. For $j = 1, \dots, K-1$,

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

and also

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

The entire proof is now complete. ■

PROOF OF COROLLARY 1. From the proof of Theorem 1, we know that whether Assumption A or MH* is used, the proposed estimator for the covariance matrix is consistent; i.e. $\hat{\Sigma} - \Sigma = o_p(1)$.

Therefore it suffices to show consistency of $\widehat{\Xi}_{jk}$. Writing

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

we see that Birkhoff's ergodic theorem, or the Law of Large Numbers for mixing variables can be used again to obtain the desired result. The regularity conditions for each theorem are ensured by Assumption A2 and MH3, respectively. Note that this consistency results can be extended to almost sure sense, without requiring any further condition. \blacksquare

PROOF OF COROLLARY 2. We follow the similar approaches for the two parameter statistics. Under the null hypothesis, by a geometric series expansion we have

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

Hence denoting

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

we have

$$\begin{aligned}
\text{var} \left(\sqrt{T} \text{vec} \left(\widehat{VR}_+(K) - I_d \right) \right) &= \text{var} \left(\sqrt{T} \sum_{j=1}^{K-1} \tilde{c}_{j,K,L} \cdot \text{vec} \left(\widehat{\Gamma}(j) \right) \right) \\
&\rightarrow \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} \tilde{c}_{j,K,L} \tilde{c}_{k,K,L} \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \Xi_{jk} \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right),
\end{aligned}$$

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

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

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. Note that as $K \rightarrow \infty$, $\Omega_\eta(K) \rightarrow 2\Omega_\eta = 2\text{var}(\eta_t)$. It follows that as $K \rightarrow \infty$

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

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

$$\begin{aligned}
VRd(K) &= K^{-1} D_1^{-1/2} \Sigma_K D_1^{-1/2} \\
&= K^{-1} D_1^{-1/2} (K\Omega_\varepsilon + \Omega_\eta(k)) D_1^{-1/2} \\
&\longrightarrow D_1^{-1/2} \Omega_\varepsilon D_1^{-1/2} \\
&= D_1^{-1/2} (\Sigma_1 - \Omega_\eta(1)) D_1^{-1/2} \\
&= D_1^{-1/2} \Sigma_1 D_1^{-1/2} - D_1^{-1/2} \Omega_\eta(1) D_1^{-1/2} \\
&= \Gamma d(0) - D_1^{-1/2} \Omega_\eta(1) D_1^{-1/2} \\
&\leq \Gamma d(0)
\end{aligned}$$

which is the instantaneous correlation matrix of the return process. ■

PROOF OF THEOREM 3. 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 THEOREM 4. We prove the result for seasonal (purely periodic) case and ignore the edge effect; this is just to keep the notation simple.

Given the conditional expectation

$$E(X_t | \mathcal{F}_{t-1}) = \mu_t = \sum_{s=1}^{\tau} g_s \left(\frac{t}{T} \right) I_s(t)$$

the backward looking rolling window estimator for the mean is

$$\hat{\mu}_t = \sum_{s=1}^{\tau} \hat{g}_s \left(\frac{t}{T} \right) I_s(t) = \frac{1}{M} \sum_{m=1}^M X_{t-(m+a)\tau}, \quad (35)$$

where $a = \lfloor K/\tau \rfloor$ in view of the periodic structure. That is, smoothing is done with the M most recent samples that belong to the same seasonal *class* with t . Due to the indicator this representation (35) holds for any t . Consequently, the estimator for the autocovariance is given by

$$\begin{aligned} \hat{\Psi}(j) &= \frac{1}{T} \sum_{t=j+1}^T (X_t - \hat{\mu}_t) (X_{t-j} - \hat{\mu}_{t-j})^\top, \quad j = 0, 1, 2, \dots \\ &= \frac{1}{T} \sum_{t=j+1}^T \left[X_t - \frac{1}{M} \sum_{m=1}^M X_{t-(m+a)\tau} \right] \left[X_{t-j} - \frac{1}{M} \sum_{m=1}^M X_{t-j-(m+a)\tau} \right]^\top. \end{aligned}$$

As in the global mean case, serial uncorrelatedness of the martingale difference gives

$$\begin{aligned} \sqrt{T} \left(\text{vec}(\hat{\Psi}(j)) - \text{vec}(\Psi(j)) \right) &= \frac{1}{\sqrt{T}} \sum_{t=j+1}^T (X_{t-j} - \hat{\mu}_{t-j}) \otimes (X_t - \hat{\mu}_t) \\ &= \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left[\tilde{X}_{t-j} + \mu_{t-j} - \hat{\mu}_{t-j} \right] \otimes \left[\tilde{X}_t + \mu_t - \hat{\mu}_t \right] \\ &= \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left(\tilde{X}_{t-j} \otimes \tilde{X}_t \right) - \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left(\hat{\mu}_{t-j} - \mu_{t-j} \right) \otimes \tilde{X}_t \\ &\quad - \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \tilde{X}_{t-j} \otimes \left(\hat{\mu}_t - \mu_t \right) + \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left(\hat{\mu}_{t-j} - \mu_{t-j} \right) \otimes \left(\hat{\mu}_t - \mu_t \right) \\ &= \tau_{T1} + \tau_{T2} + \tau_{T3} + \tau_{T4}. \end{aligned}$$

We first consider the last term. Writing $\bar{\mu}_t := 1/M \sum_{m=1}^M \mu((t - (m + a)\tau)/T)$ we have

$$\begin{aligned} \hat{\mu}_{t-j} - \mu_{t-j} &= \hat{\mu}_{t-j} - \bar{\mu}_{t-j} + \bar{\mu}_{t-j} - \mu_{t-j} \\ &= \left[\frac{1}{M} \sum_{m=1}^M \tilde{X}_{t-j-(m+a)\tau} \right] + \left[\frac{1}{M} \sum_{m=1}^M \mu\left(\frac{t-j-(m+a)\tau}{T}\right) - \mu\left(\frac{t-j}{T}\right) \right] \\ &= A_1 + A_2. \end{aligned}$$

Since continuous differentiability of $g_s(\cdot)$ over $[0, 1]$ implies that the function is Lipschitz on the same domain (by the Mean Value Theorem) we have

$$\begin{aligned} \max_{1 \leq t \leq T} \|\bar{\mu}_{t-j} - \mu_{t-j}\| &= \max_{1 \leq t \leq T} \left\| \frac{1}{M} \sum_{m=1}^M \mu\left(\frac{t-j-(m+a)\tau}{T}\right) - \mu\left(\frac{t-j}{T}\right) \right\| \\ &\leq \frac{C}{M} \sum_{m=1}^M \left| \frac{m+a}{T} \right| \leq \frac{C}{T} \cdot \left(\frac{M+1}{2} + a \right) = O\left(\frac{M}{T}\right). \end{aligned} \quad (36)$$

for some constant C . Next by the result to be shown later,

$$\begin{aligned} \max_{1 \leq t \leq T} \|\hat{\mu}_{t-j} - \bar{\mu}_{t-j}\| &= \max_{1 \leq t \leq T} \left\| \frac{1}{M} \sum_{m=1}^M \tilde{X}_{t-j-(m+a)\tau} \right\| \\ &\leq \max_{1 \leq t \leq T} \|\Omega_t^{-1/2}\| \times \max_{1 \leq t \leq T} \left\| \frac{1}{M} \sum_{m=1}^M \tilde{\tilde{X}}_{t-j-(m+a)\tau} \right\| = O_p\left(\sqrt{\frac{\log M}{M}}\right), \end{aligned} \quad (37)$$

we finally have

$$\begin{aligned} \|\tau_{T4}\| &= \left\| \frac{1}{\sqrt{T}} \sum_{t=j+1}^T (\hat{\mu}_{t-j} - \mu_{t-j}) \otimes (\hat{\mu}_t - \mu_t) \right\| \leq \frac{1}{\sqrt{T}} \sum_{t=1}^T \left\| (\hat{\mu}_{t-j} - \mu_{t-j}) \otimes (\hat{\mu}_t - \mu_t) \right\| \\ &= O_P\left(\frac{M^2}{T^{3/2}} + \frac{T^{1/2} \log M}{M} + \sqrt{\frac{M \cdot \log M}{T}}\right) = o_p(1), \end{aligned} \quad (38)$$

provided that $M = T^\alpha$ with $\alpha \in (1/2, 3/4)$.

As for the second term τ_{T2} we note that

$$\begin{aligned} \|\tau_{T2}\| &= \left\| \frac{1}{\sqrt{T}} \sum_{t=j+1}^T (\hat{\mu}_{t-j} - \mu_{t-j}) \otimes \tilde{X}_{t-j} \right\| \leq \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left\| (\hat{\mu}_{t-j} - \mu_{t-j}) \otimes \tilde{X}_{t-j} \right\| \\ &\leq \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left[\|\hat{\mu}_{t-j} - \mu_t\| \times \|\tilde{X}_{t-j}\| \right] \leq \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \|\tilde{X}_{t-j}\| \times \max_{1 \leq t \leq T} \|\hat{\mu}_{t-j} - \mu_{t-j}\| \\ &\leq \max_{1 \leq t \leq T} \|\Omega_t^{-1/2}\| \times \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \|\tilde{\tilde{X}}_{t-j}\| \times o_p(1) \leq \max_{1 \leq t \leq T} \|\Omega_t^{-1/2}\| \times \frac{1}{\sqrt{T}} \sum_{t=j+1}^T (\tilde{\tilde{X}}_{1,t-j} + \dots + \tilde{\tilde{X}}_{d,t-j}) \times o_p(1) \\ &= O(1) \times O_p(1) \times o_p(1) = o_p(1). \end{aligned}$$

where in the last inequality we used the fact that $T^{-1/2} \sum_{t=1+j}^T \tilde{X}_{i,t-j}$ is bounded in probability for any i by the CLT for stationary ergodic martingale difference.

The third term τ_{T3} can be shown to be $o_p(1)$ following similar arguments as above.

Finally, regarding the first term,

$$\tau_{T1} = \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left(\tilde{X}_{t-j} \otimes \tilde{X}_t \right) = \frac{1}{\sqrt{T}} \sum_{t=j+1}^T \left(\Omega_{t-j}^{1/2} \otimes \Omega_t^{1/2} \right) \left(\tilde{X}_{t-j} \otimes \tilde{X}_t \right),$$

we note that $\hat{\Sigma} - \Sigma = o_p(1)$, and hence $\sqrt{T} \text{vec}(\hat{\Gamma}(j)) = (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \sqrt{T} \text{vec}(\hat{\Psi}(j)) + o_p(1)$ as before.

Therefore we can write

$$\begin{aligned} \sqrt{T} \text{vec} \left(\widehat{V} R_+(K) - I_d \right) &= \sum_{j=1}^{K-1} c_j \cdot \sqrt{T} \text{vec} \left(\hat{\Gamma}(j) \right) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\sum_{j=1}^{K-1} c_j (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \left(\Omega_{t-j}^{1/2} \otimes \Omega_t^{1/2} \right) \left(\tilde{X}_{t-j} \otimes \tilde{X}_t \right) \right] + o_p(1) \\ &=: \frac{1}{\sqrt{T}} \sum_{t=1}^T R_{tj} + o_p(1). \end{aligned} \tag{39}$$

We observe that $(\tilde{X}_{t-j} \otimes \tilde{X}_t)$ is no longer stationary and ergodic, although it has a martingale structure. Therefore we apply the CLT for martingale difference arrays, Hall and Heyde (1980, Corollary 3.1).

For any constant vector $a = (a_1, \dots, a_{d^2})^\top \in \mathbb{R}^{d^2}$ we see that $a^\top R_{tj}$ is a one-dimensional martingale difference sequence. Therefore, by the CLT of Hall and Heyde,

$$a^\top \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T R_{tj} \right) \xrightarrow{d} N(0, \xi) \tag{40}$$

where

$$\frac{1}{T} \sum_{t=1}^T E \left[(a^\top R_{tj})^2 \mid \mathcal{F}_{t-1} \right] \xrightarrow{P} \xi$$

and

$$\begin{aligned}
\xi &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E[a^\top R_{tj}]^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \text{var}[a^\top R_{tj}] \\
&= a^\top \cdot \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \text{var} \left[\sum_{j=1}^{K-1} c_j (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \left(\Omega_{t-j}^{1/2} \otimes \Omega_t^{1/2} \right) \left(\tilde{X}_{t-j} \otimes \tilde{X}_t \right) \right] \cdot a \\
&= a^\top \cdot \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[\sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_j c_k (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \left(\Omega_{t-j}^{1/2} \otimes \Omega_t^{1/2} \right) \left(\tilde{X}_{t-j} \otimes \tilde{X}_t \right) \right. \\
&\quad \left. \times \left(\tilde{X}_{t-k} \otimes \tilde{X}_t \right)^\top \left(\Omega_{t-k}^{1/2} \otimes \Omega_t^{1/2} \right) (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \right] \cdot a \\
&= a^\top \cdot \left[\sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_j c_k (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \tilde{\Xi}_{jk} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \right] \cdot a \equiv a^\top Q^*(K) a \tag{41}
\end{aligned}$$

with

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left(\Omega_{t-j}^{1/2} \otimes \Omega_t^{1/2} \right) E \left[\left(\tilde{X}_{t-j} \otimes \tilde{X}_t \right) \left(\tilde{X}_{t-k} \otimes \tilde{X}_t \right)^\top \right] \left(\Omega_{t-k}^{1/2} \otimes \Omega_t^{1/2} \right) = \tilde{\Xi}_{jk}. \tag{42}$$

The conditional Lindeberg condition is satisfied because $\Omega(u)$ is bounded from above and below, element wise and eigenvalue-wise. Furthermore, ξ is ensured to be a positive constant because the limit in (42) converges to some asymptotic mean $\tilde{\Xi}_{jk}$ which is finite by condition C4.

The proof is now complete in view of (40), (41), Cramér-Wold device and Slutsky's theorem. Consistency of standard error is straightforward due to boundedness of Ω . ■

PROOF OF (37). We would like to show

$$\max_{1 \leq t \leq T} \left\| \frac{1}{M} \sum_{m=1}^M \tilde{X}_{t-j-(m-a)\tau} \right\| = O_p \left(\sqrt{\frac{\log M}{M}} \right)$$

It suffices to show componentwise convergence in probability. In other words, denoting $\tilde{X}_{i,t}$ by the i^{th} component of \tilde{X}_t , we shall prove that

$$\max_{1 \leq t \leq T} \left| \frac{1}{M} \sum_{m=j}^M \tilde{X}_{i,t-j-(m-a)\tau} \right| = O_p \left(\sqrt{\frac{\log M}{M}} \right), \quad \forall i = 1, \dots, d.$$

Note that any subsequence of a martingale is also a martingale; Motwani and Raghavan (1995, Theorem 4.12). Hence so is that of martingale difference sequence.

Now for any t we write $\tilde{X}_{i,t} = \tilde{X}_{i,t}^+ + \tilde{X}_{i,t}^-$ where

$$\begin{aligned}\tilde{X}_{i,t}^+ &= \tilde{X}_{i,t} \mathbf{1}\left(|\tilde{X}_{i,t}| \leq \sqrt{\frac{M}{\log M}}\right) - E\left[\tilde{X}_{i,t} \cdot \mathbf{1}\left(|\tilde{X}_{i,t}| \leq \sqrt{\frac{M}{\log M}}\right) \middle| \mathcal{F}_{t-1}\right] \\ \tilde{X}_{i,t}^- &= \tilde{X}_{i,t} \mathbf{1}\left(|\tilde{X}_{i,t}| > \sqrt{\frac{M}{\log M}}\right) - E\left[\tilde{X}_{i,t} \cdot \mathbf{1}\left(|\tilde{X}_{i,t}| > \sqrt{\frac{M}{\log M}}\right) \middle| \mathcal{F}_{t-1}\right].\end{aligned}$$

Then we have

$$\begin{aligned}\max_{1 \leq t \leq T} \left| \frac{1}{M} \sum_{m=1}^M \tilde{X}_{i,t-j-(m-a)\tau} \right| &\leq \max_{1 \leq t \leq T} \left| \frac{1}{M} \sum_{m=1}^M \tilde{X}_{i,t-j-(m-a)\tau}^+ \right| + \max_{1 \leq t \leq T} \left| \frac{1}{M} \sum_{m=1}^M \tilde{X}_{i,t-j-(m-a)\tau}^- \right| \\ &= A_1 + A_2.\end{aligned}\tag{43}$$

As for A_1 , we write $\sigma_{i,t}^2 := E[\tilde{X}_{i,t}^2 | \mathcal{F}_{t-1}]$, and note that $V_M^2 := \sum_{m=1}^M \sigma_{i,t-j-(m-a)\tau}^2 \leq M \cdot \sigma_M^2$, where $\sigma_M^2 := \max(\sigma_{i,t-j-(1-a)\tau}^2, \sigma_{i,t-j-(2-a)\tau}^2, \dots, \sigma_{i,t-j-(M-a)\tau}^2) < +\infty$. We can now apply the exponential inequality for martingale differences of de la Peña (1999, Theorem 1.2A):

$$\begin{aligned}P\left(\max_{1 \leq t \leq T} \left| \sum_{m=1}^M \tilde{X}_{i,t-j-(m-a)\tau}^+ \right| \geq \delta \sqrt{\frac{\log M}{M^{-1}}}\right) &= P\left(\max_{1 \leq t \leq T} \left| \sum_{m=1}^M \tilde{X}_{i,t-j-(m-a)\tau}^+ \right| \geq \delta \sqrt{\frac{\log M}{M^{-1}}}, V_M^2 \leq M \sigma_M^2\right) \\ &\leq 2T \exp\left[-\frac{\delta^2 M \log M}{2(M \sigma_M^2 + \sqrt{M/\log M} \cdot \delta \sqrt{M \log M})}\right] \\ &= 2T \exp\left[-\frac{\delta^2 \log M}{2(\sigma_M^2 + \delta)}\right] = 2T \frac{1}{M^{C\delta^2}} \longrightarrow 0\end{aligned}$$

for some constant $\delta > 0$, yielding

$$A_1 = \frac{1}{M} \max_{1 \leq t \leq T} \left| \sum_{m=1}^M \tilde{X}_{i,t-j-(m-a)\tau}^+ \right| = O_P\left(\sqrt{\frac{M}{\log M}}\right).\tag{44}$$

for any $i = 1, \dots, d$. As for the second term A_2 , we denote

$$\tilde{X}_{i,t-j-(m-a)\tau}^* := \tilde{X}_{i,t-j-(m-a)\tau} \mathbf{1}\left(|\tilde{X}_{i,t-j-(m-a)\tau}| > \sqrt{\frac{M}{\log M}}\right)\tag{45}$$

so that

$$\tilde{X}_{i,t-j-(m-a)\tau}^- = \tilde{X}_{i,t-j-(m-a)\tau}^* - E(\tilde{X}_{i,t-j-(m-a)\tau}^* | \mathcal{F}_{t-1}).$$

From (45) we see that, in order for

$$\sum_{m=1}^M \tilde{X}_{i,t-j-(m-a)\tau}^* \equiv \tilde{X}_{i,t-j-(1-a)\tau}^* + \tilde{X}_{i,t-j-(2-a)\tau}^* + \cdots + \tilde{X}_{i,t-j-(M-a)\tau}^*$$

to be positive (i.e. non-zero), then there must be at least one m such that $|\tilde{X}_{i,t-j-(m-a)\tau}^*| > \sqrt{M/\log M}$. Similarly, if $\max_{1 \leq t \leq T} |\sum_{m=1}^M \tilde{X}_{i,t-j-(m-a)\tau}^*| > 0$, then there should be at least one pair of (t, m) such that $|\tilde{X}_{i,t-j-(m-a)\tau}^*| > \sqrt{M/\log M}$. Hence for any $\delta > 0$, we have

$$\begin{aligned} & P\left(\max_{1 \leq t \leq T} \left| \sum_{m=1}^M \tilde{X}_{i,t-m}^* - E(\tilde{X}_{i,t-m}^* | \mathcal{F}_{t-1}) \right| > \delta \sqrt{M \log M}\right) \\ & \leq P\left(\max_{1 \leq t \leq T} \left| \sum_{m=1}^M \tilde{X}_{i,t-m}^* \right| > \delta \sqrt{M \log M}\right) \leq P\left(\max_{1 \leq t \leq T} \left| \sum_{m=1}^M \tilde{X}_{i,t-m}^* \right| > 0\right) \\ & = P\left(\max_{1 \leq t \leq T} \max_{1 \leq m \leq M} |\tilde{X}_{i,t-j-(m-a)\tau}^*| > \sqrt{\frac{M}{\log M}}\right) = P\left(\max_{1-M \leq s \leq T-1} |\tilde{X}_{i,s}^*| > \sqrt{\frac{M}{\log M}}\right) \\ & \leq (T + M - 1) \cdot P\left(|\tilde{X}_{i,s}^*| > \sqrt{\frac{M}{\log M}}\right) \leq (T + M - 1) \cdot \frac{(\log M)^{\kappa/2} E|\tilde{X}_{i,s}^*|^\kappa}{M^{\kappa/2}} \rightarrow 0 \end{aligned}$$

as $M \rightarrow \infty$ ($\kappa = 4$), implying that

$$A_2 = \frac{1}{M} \max_{1 \leq t \leq T} \left| \sum_{m=1}^M \tilde{X}_{i,t-m}^- \right| = o_P\left(\sqrt{\frac{M}{\log M}}\right). \quad (46)$$

for all $i = 1, \dots, d$. The proof is complete in view of (44) and (46). \blacksquare

PROOF OF (25). 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

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

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

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

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

$$\delta_T^{-2T} \sum_{t=1}^T X_t^2 \implies c^2 Y^2. \quad (47)$$

Likewise,

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

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

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

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

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

and

$$\delta_T^{-2T} \sum_{t=1}^T X_t(K)^2 \implies K^2 c^2 Y^2. \quad (48)$$

In fact, the convergence in (47) and (48) is joint. Therefore,

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

as required. ■

10.1 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

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

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

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

Under the iid assumption (which allows us to ignore the denominator, see below) we have we have

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

under the null hypothesis. Likewise,

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

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

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

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

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

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 $\sqrt{T}\widehat{\Psi}(j)$ 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,

$$\text{acov}(\sqrt{T}\widehat{v}_j, \sqrt{T}(\widehat{v}_0 - v_0)) = E \left[\left(\widetilde{X}_{-j} \widetilde{X}_0^\top \otimes \widetilde{X}_0 \widetilde{X}_0^\top \right) \right] + \sum_{s=1}^{\infty} E \left[\left(\widetilde{X}_{-j} \widetilde{X}_s^\top \otimes \widetilde{X}_0 \widetilde{X}_s^\top \right) \right]. \quad (49)$$

From this we can obtain a formula for $E[\Sigma^{-1}(\widehat{\Sigma} - \Sigma)\Sigma^{-1}\widehat{\Psi}(j)\Sigma^{-1/2}]$ in terms of the right hand side of (49), 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] Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3), 817-858.
- [2] Bailey, N., Kapetanios, G. and Pesaran, H. (2012). Exponent of cross-sectional dependence: Estimation and inference. Discussion Paper series no. 6318, Forschungsinstitut zur Zukunft der Arbeit.
- [3] Blanchard, O. J. and Watson, M. W. (1983). Bubbles, rational expectations and financial markets. in P. Wachtel, ed., *Crisis in the Economic and Financial Structure*. Lexington, VA: Lexington Books.
- [4] Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*. New York: Springer-Verlag.
- [5] Campbell, J.Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

- [6] 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>
- [7] Charles, A. and Darné, O. (2009). Variance-Ratio Tests of Random Walk: An Overview. *Journal of Economic Surveys*, 23(3), 503-527.
- [8] 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.
- [9] Chitturi, R. V. (1974). Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association*, 69(348), 928-934.
- [10] Chordia, T., Roll, R., and Subrahmanyam, A. (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics*, 101(2), 243-263.
- [11] Cochrane, J. H. and Sbordone, A. M. (1988). Multivariate estimates of the permanent components of GNP and stock prices. *Journal of Economic Dynamics and Control*, 12(2), 255-296.
- [12] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1), 1-37.
- [13] Dimson, E., Marsh, P. and Staunton, M. (2008). The worldwide equity premium: A smaller puzzle. in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Amsterdam: Elsevier.
- [14] Dufour, J. M. and Roy, R. (1986). Generalized portmanteau statistics and tests of randomness. *Communications in Statistics - Theory and Methods*, 15(10), 2953-2972.
- [15] Escanciano, J. C. and Velasco, C. (2006). Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics*, 134(1), 151-185.
- [16] 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.
- [17] 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.

- [18] Fama, E. F. and French, K. R. (1998). Value versus growth: The international evidence. *The Journal of Finance*, 53(6), 1975-1999.
- [19] Faust, J. (1992). When are variance ratio tests for serial dependence optimal?. *Econometrica*, 60(5), 1215-1226.
- [20] Flood, R. P. and Hodrick, R. J. (1990). On testing for speculative bubbles. *The Journal of Economic Perspectives*, 4(2), 85-101.
- [21] French, K. R. and Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of financial economics*, 17(1), 5-26.
- [22] Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. New York: Academic Press.
- [23] Han, H., Linton, O., 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>
- [24] Hasbrouck, J. (2006). *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*. New York: Oxford University Press.
- [25] Hayashi, F. (2000). *Econometrics*. Princeton, NJ: Princeton University Press.
- [26] 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.
- [27] Hillman, R. and Salmon, M. (2007). Intrinsic stationarity: Investigating predictability in real-time forex transactions. *Journal of Financial Forecasting*, 1, 3-43.
- [28] Hong, Y. (2000). Generalized spectral tests for serial dependence. *Journal of the Royal Statistical Society: Series B*, 62(3), 557-574.
- [29] Hong, Y. and Lee, Y. J. (2005). Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. *The Review of Economic Studies*, 72(2), 499-541.

- [30] Hosking, J. R. M. (1981). Equivalent forms of the multivariate portmanteau statistic. *Journal of the Royal Statistical Society. Series B*, 43(2), 261-262.
- [31] 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.
- [32] Liu, W. and Wu, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory*, 26(4), 1218-1245.
- [33] Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5), 15-29.
- [34] Lo, A. W. (2005). Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. *Journal of Investment Consulting*, 7(2), 21-44.
- [35] 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.
- [36] 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.
- [37] Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96(455), 1066-1076.
- [38] Lobato, I., Nankervis, J. C. and Savin, N. E. (2001). Testing for autocorrelation using a modified box-pierce Q test. *International Economic Review*, 42(1), 187-205.
- [39] Luger, R. (2003). Exact non-parametric tests for a random walk with unknown drift under conditional heteroscedasticity. *Journal of Econometrics*, 115(2), 259-276.
- [40] Magnus, J. R. and Neudecker, H. (1980). The elimination matrix: some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods*, 1(4), 422-449.
- [41] 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.
- [42] Mikosch, T. and Starica, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH (1, 1) process. *Annals of Statistics*, 24(4), 1427-1451.

- [43] Montagnoli, A. and De Vries, F. P. (2010). Carbon trading thickness and market efficiency. *Energy Economics*, 32(6), 1331-1336.
- [44] Motwani, R. and Raghavan, P. (1995). *Randomized Algorithms*. Cambridge: Cambridge University Press.
- [45] Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the american statistical association*, 55(290), 299-306.
- [46] Newey, W. K. and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28(3), 777-787.
- [47] Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics*, 28(2), 329-348.
- [48] de la Peña, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *Annals of Probability*, 27(1), 537-564.
- [49] Peterson, R. L., Ma, C. K. and Ritchey, R. J. (1992). Dependence in commodity prices. *Journal of Futures Markets*, 12(4), 429-446.
- [50] Phillips, P. C. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1), 115-130.
- [51] Phillips, P.C.B., Shi, S.P. and Jun Yu (2012). Testing for multiple bubbles. Cowles Foundation Discussion Paper no. 1843.
- [52] 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.
- [53] Phillips, P. C. and Yu, J. (2010). Dating the timeline of financial bubbles during the subprime crisis. Cowles Foundation Discussion Paper no. 1770.
- [54] Pierce, D. A. and Haugh, L. D. (1977). Causality in temporal systems: Characterization and a survey. *Journal of Econometrics*, 5(3), 265-293.
- [55] Poterba, J. M. and Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics*, 22(1), 27-59.

- [56] Richardson, M. and Stock, J. H. (1989). Drawing inferences from statistics based on multiyear asset returns. *Journal of Financial Economics*, 25(2), 323-348.
- [57] Szroeter, J. (1978). Generalized variance-ratio tests for serial correlation in multivariate regression models. *Journal of Econometrics*, 8(1), 47-59.
- [58] Timmerman, A. (2008). Elusive Return Predictability. *International Journal of Forecasting* 24, 1–18
- [59] Vogt, M. and Linton, O. (2014). Nonparametric estimation of a periodic sequence in the presence of a smooth trend. *Biometrika*, 101(1), 121-140.
- [60] Wang, P. (2003). *Econometric analysis of the real estate market and investment*. London: Routledge.
- [61] Whang, Y. J. and Kim, J. (2003). A multiple variance ratio test using subsampling. *Economics Letters*, 79(2), 225-230.
- [62] White, H. (1984). *Asymptotic theory for econometricians*. New York: Academic Press.
- [63] Wright, J. H. (2000). Alternative variance-ratio tests using ranks and signs. *Journal of Business & Economic Statistics*, 18(1), 1-9.