Multivariate Ratio Statistics

Seok Young Hong, Oliver Linton, and Hui Jun Zhang
University of Cambridge

National University of Singapore, May, 2014

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.

Under weak form EMH excess returns satisfy

$$E(\text{excess returns}_t | \mathcal{F}_{t-1}) = 0,$$

where $\mathcal{F}_t$ denotes the past history of prices. It follows that

$$\frac{\mathrm{var}\,(\text{low frequency returns})}{\mathrm{var}\,(\text{high frequency returns})} = \frac{\text{high frequency}}{\text{low frequency}}$$

Many applications/citations

Crime against statistics?

- 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 $\phi(L)$. Then consider tests based on comparing $\mathrm{var}(r_t^{\phi})/\mathrm{var}(r_t)$. 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.
- An advantage of the variance ratio over the Box-Pierce statistic is that it gives some sense of the direction of predictability (momentum or contrarian), 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.

Our Contribution

- We propose a number of different multivariate variance ratio statistics and univariate quantities derived thereof
- We give the asymptotic distribution under (stationary) martingale difference assumptions with finite fourth moments. We do not impose the "no leverage" assumption of Lo and MacKinlay (1987). Standard errors are thus a little more complicated but not much.
- We establish the limiting behaviour of the statistics under the "fads" type of alternative model
- We extend the framework to allow for deterministic nonparametric trend and seasonals and show the limiting behaviour is the same as in the stationary mds case.
- We apply the method to five CRSP size sorted portfolios over the period 1962-2014. We show that the variance ratios have come closer to the EMH prediction over time

Definitions

Suppose that we have a vector stationary ergodic discrete time series $\{X_t, t = 0, \pm 1, \dots\} \subset \mathbb{R}^d$. Let $\widetilde{X}_t = X_t - \mu$, where $\mu = EX_t$, and define the following population quantities for $j = 0, \pm 1, \dots.$:

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

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

$$\Psi(j) = E(\widetilde{X}_t \widetilde{X}_{t-j}^\intercal)$$

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

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

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

We define the multivariate ratio (population) statistic as

$$VR(K) = \mathrm{var}(X_t)^{-1/2}\mathrm{var}(X_t + X_{t+1} + \ldots + X_{t+K-1})\mathrm{var}(X_t)^{-1/2}/K.$$

Under the null hypothesis, we should have $VR(K) = I_d$.
Under the generic stationary alternative hypothesis we have

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

which is a symmetric matrix. The off-diagonal elements should be zero under the null hypothesis of no predictability.

An alternative definition is

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

Under the null hypothesis, we should have $VRa(K) = I_d$.
Under the generic stationary alternative hypothesis we have

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

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

We may instead look at the diagonally normalized variance ratio

$$VRd(K) = D^{-1/2}\mathrm{var}(X_t + X_{t+1} + \ldots + X_{t+K-1})D^{-1/2}/K$$

Under the null hypothesis, we should have $VRa(K) = I_d$.
Under the generic stationary alternative hypothesis we have

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

where $\Gamma d(0) = D^{-1/2}\Psi(0)D^{-1/2}$ is the $d \times d$ contemporaneous correlation matrix.

We may also consider a two parameter family of variance ratio statistics, as in Poterba and Summers (1988) - e.g., to calculate weekly variance ratios with daily data
First definition

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

for K, L. Under the null hypothesis, we should have $VR(K,L) = I_d$. An alternative definition 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.

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

$$\sigma(VR(K)) = \{\lambda \in \mathbb{R} : VR(K)x = \lambda x \text{ for some } x \in \mathbb{R}^d \backslash \{0\}\}$$

of the variance ratio statistic and let $\lambda_{\max}(K), \lambda_{\min}(K)$ denote the largest (smallest) elements of $\sigma(VR(K))$ (likewise $VRd(K)$).

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 multivariate directional predictability of the series. Consider a portfolio of assets with fixed weights $w \in \mathbb{R}^d$, we have (abusing the notation somewhat)

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

where $VR(K; w^\intercal X_t)$ denotes the univariate variance ratio of the portfolio $w^\intercal X_t$, while $\widetilde{w} = \Sigma^{1/2} w$ and $Y_t = \Sigma^{-1/2} X_t$.

Another parameter of interest is the average of the off diagonal elements of $VRd(K)$, which is

$$\begin{aligned} CS(K) &= \frac{2}{d(d-1)} \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} VRd_{ij}(K) \\ &= \frac{1}{d(d-1)} \{ i^\intercal VRd(K) i - \mathrm{tr}(VRd(K)) \}, \end{aligned}$$

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. It is also related to the expected profit of the Lo and MacKinlay (1990) portfolio momentum strategies (they chose weights $w_{it}(k) = -(1/d)(X_{i,t-k} - \overline{X}_{t-k})$, where $\overline{X}_{t-k}$ is the equally weighted "market portfolio", and showed that the expected profit of this strategy $\pi(k) = \mathrm{tr}(\Gamma(k))/d - i^\intercal \Gamma(k) i / d^2$, in the case where each asset has the same mean and variance).

One sided Statistics

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.

We propose the following quantities based on:

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

and the negative counterparts $VR_{-}(K) = VR^{\intercal}(K)$ and $VRd_{-}(K) = VRd^{\intercal}(K)$, which have the property that:

$$VR(K) = (VR_{+}(K) + VR_{+}^{\intercal}(K))/2.$$

Estimation

We estimate the population quantities by sample averages: for $j = 0, 1, 2, \ldots$

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

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

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

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

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

Lo and MacKinlay (1988)

ASSUMPTION H.

H1. *For all* t, $E\widetilde{X}_t = 0$ *and* $E[\widetilde{X}_t\widetilde{X}_{t-j}] = 0$

H2. $\widetilde{X}_t$ *is* $\alpha-$*mixing with coefficients* $\alpha(m)$ *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* $E[|\widetilde{X}_t\widetilde{X}_{t-j}|^{2(r+\delta)}] < \Delta < \infty$

H3. $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[\widetilde{X}_t^2] = \sigma^2 < \infty$

H4. *For all* t, $E[\widetilde{X}_t^2\widetilde{X}_{t-j}\widetilde{X}_{t-k}] = 0$ *for any* $j, k \neq 0$ *with* $j \neq k$

- Whang and Kim (2003) dispense with H4

Our Framework

ASSUMPTION A.

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

A2. *The process* $\widetilde{X}_t$ *has finite fourth moments, i.e., for all* i, j, k, l, $E[|X_{ti}X_{tj}X_{tk}X_{tl}|] < \infty$

A3. *Assume that*

$$\sum_{k=0}^{\infty} \left\| \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[\left(\widetilde{X}_{t-k}\widetilde{X}_t^{\intercal} \otimes \widetilde{X}_{t-k}\widetilde{X}_t^{\intercal}\right)\right] \right\| < \infty.$$

A4. *For* $j, k = 0, 1, 2, \ldots, K$

$$\sum_{\tau=0}^{\infty} \left\| \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[v_{t,\tau,j,k} v_{t,\tau,j,k}^{\intercal}\right] \right\| < \infty$$

$$v_{t,\tau,j,k} = \mathrm{vec}\left(\widetilde{X}_{t+\tau-j}\widetilde{X}_{t+\tau-k}^{\intercal} \otimes \widetilde{X}_{t+\tau}\widetilde{X}_{t+\tau}^{\intercal}\right).$$

Asymptotic Variance Matrices

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

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

Note that under H4, $\Xi_{jk} = 0$ if $j \neq k$ and only the $c_{j,K}^2$ terms remain.

We show below how our theory is robust to certain types of nonstationarity. We present the results for the one sided statistics.

THEOREM 1. *Suppose that Assumption A1-A3 holds or Assumption H1-H3 holds. Then,*

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

It follows from this that, for example,

$$\sqrt{T}\mathrm{vech}(\widehat{VR}(K) - I_d) \Longrightarrow N(0, S(K)),\ S(K) = D_n^+ N_n Q(K) N_n^\intercal D_n^{+\intercal},$$

where N_n and D_n^+ are the matrices of zeros and ones defined on pages 48 and 56 in Magnus (1988). Likewise for the other variance ratio statistics. The asymptotic distribution for smooth functions of the variance ratio matrix can easily be obtained via the delta method.

Note that under the Lo and MacKinlay (1988) condition H4 we have $\Xi_{jk} = 0$ unless $j = k$, so that the asymptotic variance simplifies, a little

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

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

$$Q(K) = \sum_{j=1}^{K-1} c_{j,K}^2 I_{d^2}.$$

The two parameter variance ratio $\widehat{VR}(K,L) = \widehat{VR}(L)^{-1/2} \times \widehat{VR}(K) \times \widehat{VR}(L)^{-1/2}$, satisfies

$$\sqrt{T}\mathrm{vec}\left(\widehat{VR}_{+}(K,L) - I_d\right) \Longrightarrow N\left(0, Q(K,L)\right),$$

where $Q(K,L)$ is as $Q(K)$ except with weights

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

We can compare the variance ratio estimator efficency in the case that $K = LJ$ for J, L positive integers. We show that the relative efficiency (when returns are iid) for the general J, L case is

$$\begin{aligned} \frac{\sum_{j=1}^{K-1} \widetilde{c}_{j,K}^2}{\sum_{j=1}^{K-1} c_{j,K}^2} &= \frac{(2J-2)L^2 + 1}{L^2(2J-1)} \\ &= 1 - \frac{L^2 - 1}{L^2(2J-1)} \\ &< 1 \end{aligned}$$

for any $L, J \geq 2$, i.e., the two parameter statistic using the highest frequency data is more efficient in this case.

Inference

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

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

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

Then (under the further summability condition A4), $\widehat{Q}(K) \rightarrow Q(K)$ in probability and

$$\widehat{Q}(K)^{-1} \sqrt{T} \mathrm{vec}\left(\widehat{VR}_{+}(K) - I_d\right) \Longrightarrow N\left(0, I_{d^2}\right)$$

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

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

reflects this structure. In the iid case, we further have $\Xi_{jj} = \Sigma \otimes \Sigma$ and

$$Q(K) = \sum_{j=1}^{K-1} c_{j,K}^2 I_{d^2}.$$

In that case the standard error is nuisance parameter free.

The eigenvalues are not smooth functions of the variance ratio matrix in general - at the null hyptohesis they are all equal to one - and so other methods need to be applied.
Specifically, Eaton and Tyler (1991) 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 $i_d = (1, 1, \ldots, 1)^\intercal$

$$\sqrt{T}\left(\sigma(\widehat{VR}(K)) - i_d\right) \Longrightarrow \sigma(W).$$

Alternative inference methods such as self normalization, Lobato (2001), or bootstrap and subsampling, Whang and Kim (2003), may give better results.

In the Appendix (section 7.1) we discuss a bias correction method based on asymptotic expansions, which may give better performance for long lags.

Multivariate Fads Model

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

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

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

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

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

This is a multivariate generalization of the Muth (1960) model. 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.

Consider the K period returns

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

These have variance

$$\begin{aligned} \Sigma_K &= \operatorname{var}(X_t(K)) = \operatorname{var}\left(\sum_{s=t}^{t+K} \varepsilon_s\right) + \operatorname{var}\left(\eta_{t+K} - \eta_{t-1}\right) \\ &= KE\varepsilon_s\varepsilon_s^{\intercal} + E\left((\eta_{t+K} - \eta_{t-1})(\eta_{t+K} - \eta_{t-1})^{\intercal}\right) = K\Omega_\varepsilon + \Omega_\eta(K), \end{aligned}$$

where $\Omega_\eta(k) = \operatorname{var}\left(\eta_{t+k} - \eta_{t-1}\right) \geq 0$, $k = 1, 2, \ldots.$

Let $\Gamma d_{\varepsilon}(0), \Gamma d(0)$ be the correlation matrices of ε_t and X_t respectively.
THEOREM 2. *Suppose that the multivariate fads model holds. Then,* $VR(\infty) = \lim_{K\to\infty} VR(K) = I + \sum_{j=1}^{\infty}(\Gamma(j) + \Gamma(j)^{\intercal})$ *exists and*

$$VR(\infty) \leq I_d$$

in the matrix partial order sense with strict inequality whenever $\Omega_{\eta} > 0$*. Likewise,* $VRd(\infty) = \lim_{K\to\infty} VRd(K)$ *exists and*

$$VRd(\infty) \leq \Gamma d_{\varepsilon}(0) \leq \Gamma d(0).$$

This says that in particular all eigenvalues of the long run variance ratio are less than or equal to one. We can test this hypothesis empirically. The corresponding result for the diagonalized variance ratio is not so useful as it involves the unknown contemporaneous correlation matrix, although we can say in the special case where all the correlation entries are positive, the eigenvalues of $\Gamma_d(0)$ lie between zero and one.

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

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

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

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

where e_t' is an i.i.d. copy of e_t and $||.||$ denotes the Euclidean norm.
Assumption B. *The vector process X_t is stationary with finite fourth moments and weakly dependent in the sense that $\sum_{t=1}^{\infty} \delta_t < \infty$.*
Theorem 3. *Suppose that the multivariate fads model holds along with Assumption B. Then,*

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

Bubble Process

Phillips and Yu (2010) and Phillips, Shi, and Yu (2012) considered the following class of "bubble processes", i.e., mildly explosive price regimes and martingale regimes:

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

where $p^*_{\tau_f}$ represents the restarting price after the bubble collapses at time τ_f, and $\delta_T > 1$. The process is consistent with efficient markets hypothesis during $[1, \tau_e]$ and $[\tau_f, T]$ but has an explosive irrational moment in the middle. They propose methods to test for the presence of a bubble using rolling window methods. One can imagine this process 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. Our simulations show that in this case the variance ratio statistics diverge to infinity (as horizon length K increases) for a long lasting bubble (that is, $(\tau_f - \tau_e)/T > 0$).

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). 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

$$\mu_t(\theta_0) = E(X_t|\mathcal{F}_{t-1})$$

see for example, Engle, Lilien and Robins (1987).
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. Specifically, suppose that $E(X_t|\mathcal{F}_{t-1}) = \mu_t$, where

$$\mu_t = \sum_{s=1}^{\tau} g_s(t/T)1(t \in J_s),$$

where $g_s(.)$ are continuously differentiable but unknown vector functions representing smooth trends that vary across $s = 1, \ldots, \tau$.
We take J_s such that

$$\{1, \ldots, T\} = \cup_{s=1}^{\tau} J_s \text{ with } J_s \cap J_r = \emptyset \text{ for } r \neq s$$

and $\#J_s = T_s$ such that $T_s/T \to c_s$ for all $s = 1, \ldots, \tau$ with τ fixed and $c_s \in (0, \infty)$.
The additive seasonal effect model in Vogt and Linton (2014, Biometrika) is a special case.

The trends capture the idea that the risk premium is slowly varying, like Dimson, Marsh, and Staunton (2008).

The second aspect may represent a seasonal effect that would not be represented inside a common trend function. τ could be the known period of a common seasonal component and $1(t \in J_s) = 1(t = k\tau + s$ for some $k)$ are then seasonal dummies, Vogt and Linton (2014). (Trading time hypothesis with day of the week effects (i.e., not EMH))

We allow additionally some irregularity in the sets J_s, to account for example for 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.

This model could also capture structural change instead by taking the intervals to be contiguous blocks of time.

We suppose that for each J_s we can order the observation times $t_{s1} < t_{s2} < \cdots < t_{sT_s}$ such that $\max_{1 \leq j \leq T_s - 1} \left| t_{sj} - t_{sj+1} \right| \leq C$ for some $C < \infty$, which means that the information accumulates in the usual way.

In this case we propose the following backward looking rolling window estimators of the mean μ_t. We let

$$\mathcal{N}_s(t, M, K) = \{M \text{ closest } r \text{ to } t \text{ in } J_s \text{ prior to } t - K\}$$

$$\widehat{g}_s(t/T) = \frac{1}{M} \sum_{r \in \mathcal{N}_s(t,M,K)} X_r \quad ; \quad \widehat{\mu}_t = \sum_{s=1}^{\tau} \widehat{g}_s(t/T) 1(t \in J_s)$$

In other words we smooth over time using just the observations in J_s that have the same seasonal affiliation. Then let

$$\widehat{\Psi}(j) = \frac{1}{T} \sum_{t=j+1}^{T} \left(X_t - \widehat{\mu}_{q(t,t-j)}\right) \left(X_{t-j} - \widehat{\mu}_{t-j}\right)^{\intercal},$$

and all the variance ratio statistics based on this.

We next discuss the asymptotic properties of this modified variance ratio statistic. We further suppose that there is some uniformly bounded deterministic family of covariance matrices Ω_t, such that

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

is stationary and ergodic (and a martingale difference sequence). This allows for periodic and trending components in the variance as well. Define $\widehat{Q}(K)$ is above but with

$$\begin{aligned} \widehat{\Xi}_{jk} &= \frac{1}{T} \sum_{t=\max\{j,k\}+1}^{T} \left(X_{t-j} - \widehat{\mu}_{t-\max\{j,k\}}\right) \left(X_{t-k} - \widehat{\mu}_{\max\{j,k\}}\right)^{\intercal} \\ &\quad \otimes \left(X_t - \widehat{\mu}_{t-\max\{j,k\}}\right) \left(X_t - \widehat{\mu}_{t-\max\{j,k\}}\right)^{\intercal}. \end{aligned}$$

Theorem 4. *Under some regularity conditions including that $M \to \infty$ and $M/T \to 0$, we have*

$$\widehat{Q}(K)^{-1/2} \sqrt{T} \mathrm{vec}\left(\widehat{VR}_+(K) - I_d\right) \Longrightarrow N(0, I_{d^2}).$$

We may further assume that

$$g_s(u) = a_s + b_s g(u)$$

for some common trend function g and seasonal coefficients with $\sum a_s = 0$ and $\sum b_s = 1$ in which case we can obtain the common function $g(.)$ by averaging over $s = 1, \ldots, \tau$

Then do time series regression to get a_s, b_s

Need the multiplicative scaling to be consistent with the calendar time hypothesis in which case $a_s = a$ and $b_M = 3 * b_{Tue}$ etc.

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").

Application

We apply our methodology to weekly size-sorted portfolio returns from the Center for Research in Security Prices (CRSP) from 06/07/1962 to 27/12/2013.
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 two subsamples: 06/07/1962-23/12/1994 (1695 weeks) and 30/12/1994-27/12/2013 (992 weeks). Based on the multivariate variance ratio statistics $VRd_+(K)$, we test a series of hypotheses: $[VRd_+(K)]_{ii} = 1$ for $i = 1, 2, 3$. Panel A of Table 1 reports the results for the portfolio of small-size firms, panel B reports the results for the portfolio of medium-size firms, and panel 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).

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 (10.72)*	1.95 (11.05)*	2.54 (9.96)*	2.92 (8.19)*
94:12:30—13:12:27	992	1.21 (3.30)*	1.47 (3.58)*	1.7 (3.35)*	1.82 (2.50)*
B. Portfolio of firms with market values in central CRSP quantile					
62:07:06—94:12:23	1695	1.23 (5.93)*	1.46 (6.18)*	1.68 (5.38)*	1.74 (3.81)*
94:12:30—13:12:27	992	0.99 (−0.02)	1.05 (0.38)	1.02 (0.10)	0.89 (−0.38)
C. Portfolio of firms with market values in largest CRSP quantile					
62:07:06—94:12:23	1695	1.04 (1.21)	1.11 (1.59)	1.15 (1.37)	1.12 (0.64)
94:12:30—13:12:27	992	0.93 (−0.99)	0.94 (−0.46)	0.89 (−0.53)	0.81 (−0.62)

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)

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 $i, j = 1, 2, 3, \ i \neq j$.

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 $VR(K)$. The null hypothesis is $H_0 : \lambda_i(VR(K)) = 1$. We report the maximum eigenvalue of $\widehat{VR}(K)$ and the simulated p-value.

		Lags			
Sample period	# of obs	$K = 2$	$K = 4$	$K = 8$	$K = 16$
62:07:06—94:12:23	1695	1.52	2.21	3.04	3.64
94:12:30—13:12:27	992	1.32	1.75	2.21	2.62

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

Leverage Effects

We also investigated the relevance of the "no leverage" assumption H4 of Lo and MacKinlay (1988). Specifically, we compare the matrices $\widehat{Q}(K)$ with $\widehat{Q}_{LM}(K)$ in the case that $K = 16$, using weekly returns in small-size and large-size CRSP portfolios

Sample period	$\widehat{Q}(K=16)$				$\widehat{Q}_{LM}(K=16)$			
62:07:06—94:12:23	98.75	0.84	9.58	-0.39	34.33	-1.53	3.82	0.36
	0.84	75.86	-0.39	19.55	-1.53	23.60	0.36	1.28
	9.58	-0.39	24.22	-2.30	3.82	0.36	26.73	-0.77
	-0.39	19.55	-2.30	39.29	0.36	1.28	-0.77	39.97
94:12:30—13:12:27	91.94	-1.67	1.57	29.79	40.37	4.80	3.96	10.91
	-1.67	98.25	29.79	13.76	4.80	35.21	10.91	8.82
	1.57	29.79	34.17	8.24	3.96	10.91	42.23	7.92
	29.79	13.76	8.24	56.68	10.91	8.82	7.92	43.07

Long Horizon/Individual Stocks

We look next at the long horizon, in particular out to two years. We shall focus on individual stocks. Specifically, we chose two large stocks and two small stocks from the CRSP universe and evaluate the long run behaviour of the variance ratio matrices. In this case, we work with weekly data from 2000-2014. We work with the bias-corrected estimator (defined in Appendix 8.1)

$$\widehat{VR}^{bc}(K) = \widehat{VR}(K) + \frac{K-1}{T} I_d.$$

We show below the two eigenvalues of both $\widehat{VR}(K)$ (solid lines) and $\widehat{VR}^{bc}(K)$ (dotted lines) against lag K.

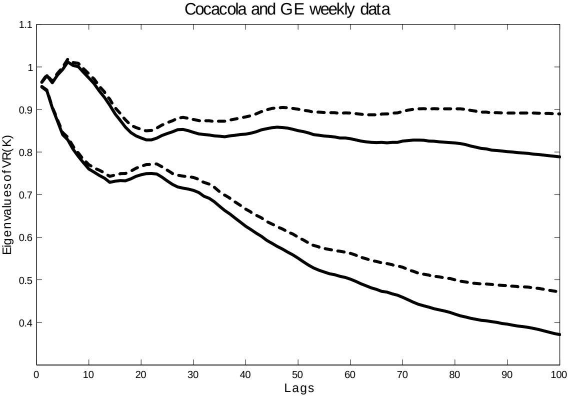


Figure 1. The eigenvalues of the variance ratio matrix for large stocks as a function of lag order.

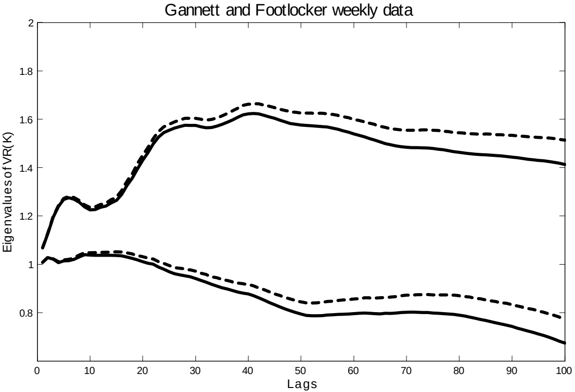
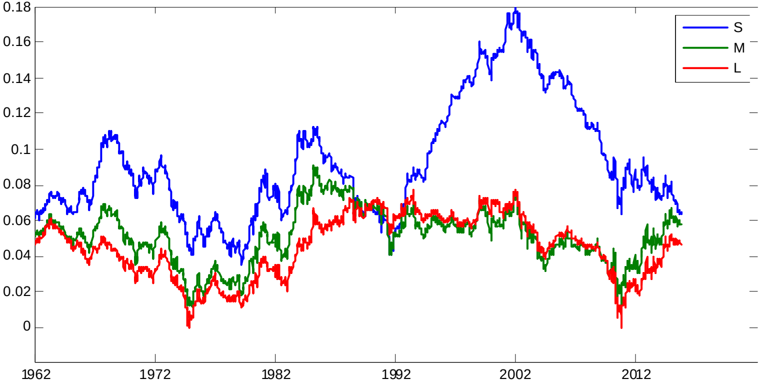


Figure 2. The eigenvalues of the variance ratio matrix for small stocks as a function of lag order.

The result for large stock returns (Figure 1) seems to support the fads interpretation rather than the explosive bubble process, while Figure 2 illustrates evidence of bubbles in small stock returns.



0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1962
1972
1982
1992
2002
2012
S
M
L

Conclusions

- Multivariate Ratio statistics convey more information than univariate ones
- MDS assumptions seem more appropriate than no leverage assumption. Standard errors not complicated.
- Methodology is robust to fitting a slowly varying time trend and seasonal effects
- Empirical results show that EMH violations have reduced for small stocks