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

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

University of Cambridge

November 7, 2016

Abstract

We propose several multivariate variance ratio statistics for "testing" the weak form Efficient Market Hypothesis and for measuring the direction and magnitude of departures from this hypothesis. 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. We propose asymptotic standard errors that are robust to departures from the "no leverage" assumption of Lo and MacKinlay (1988), but are relatively simple and in particular do not require the selection of a bandwidth parameter. 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 weekly returns for CRSP size-sorted portfolios 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, but we still reject the multivariate null hypothesis in the most recent period. The main findings are not substantially affected by allowing for a common factor time varying risk premium.

[†]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

It is fair to say that the profession is divided on the evidence regarding the Efficient Markets Hypothesis (henceforth for simplicity EMH). Authors like Fama (1970, 2013), Malkiel (2015), and Ross (2002) argue that rejections of the EMH are: small, not scalable, fleeting, statistically suspect, and not realizable profit opportunities. Furthermore, Fama has emphasized the joint hypothesis problem whereby one must measure abnormal returns relative to a market equilibrium return that provides compensation for bearing risk, so that statistical rejections of the hypothesis are potentially due instead to rejection of the assumed market equilibrium return. On the other hand, authors like Shiller (2013), Kahneman and Tversky (2000) and others have argued that market participants are irrational, behave irrationally, and that their interaction produces excess volatility in asset returns relative to fundamentals. Grossman and Stiglitz (1980) argue that even if market participants are fully rational, if information acquisition is costly, then prices cannot perfectly reflect the information which is available, since if it did, those who spent resources to obtain it would receive no compensation, leading to the conclusion that an informationally efficient market is impossible. King (2016) emphasizes the "radical uncertainty" issue, whereby the future states of the world are not perfectly knowable and formal concepts such as probability distributions and expectations that are core to EMH are of limited use. These are just some of the many varied and nuanced points of view on this subject. Our purpose is not to provide definitive evidence on this hypothesis one way or another, this is beyond our pay grade. Instead we focus on some methodological issues. As Robert Shiller says in his Nobel lecture: "Ultimately, the question in reconciling the apparently conflicting views comes down to that of constructing the right statistical tests." We contribute to this by investigating a celebrated class of tests of this hypothesis, which we think have been wrongly applied, and making some modest proposals to improve best practice with regard to their use. Halbert White made many fundamental contributions to the statistical underpinnings of hypothesis testing as applied in economics and finance, and his (2000) seminal contribution provides a formal insight as to how common applications of statistical methodology can lead to faulty conclusions regarding for example the EMH.

Variance ratio statistics (Lo and MacKinlay (1988) and Poterba and Summers (1988)) are widely used in empirical finance as a way of testing the EMH and to measure the degree and (cumulative) direction of departures from this hypothesis in financial time series. Indeed, this work has been extremely influential in understanding predictability in asset prices and in measuring market quality.

2

A key advantage of this methodology relative to say Box-Pierce statistics is that variance ratios give information about the direction of departures from the null hypothesis that can be interpreted in meaningful economic terms (i.e., momentum versus contrarian), so that the analysis does not reduce to yes/no decision-making on an uninformative test statistic. A lot of empirical work followed immediately after the seminal contributions. Lo and MacKinlay (1988) presented evidence regarding predictability of the US stock market. They concluded that the EMH was soundly rejected in weekly US stock market returns based on their standard errors. The graduate textbook Campbell, Lo, and MacKinlay (1997), henceforth CLM, presents variance ratios for weekly value weighted and equal weighted CRSP indexes and five size sorted portfolios over the period 1962-1994; they argue that the EMH is strongly rejected based on their standard errors, although they find that the magnitude of the violation is less in the later subperiod 1978-1994. On the other hand, Cochrane (2001, p388) writing only four years later argues that: "daily, weekly, and monthly stock returns are close to unpredictable".[1] One important recent direction for this methodology is in "high frequency" settings, i.e., intraday, where it has informed the debate on the evolution of "market quality" in the US stock market. 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.[2] One interpretation of these studies is that the computerized trading systems that now dominate equity markets have improved the functioning of those markets. Hasbrouck (2015) has recently used variance ratios to measure high frequency volatility in quoted prices, which also relates to this question. Finally, given that variance ratios are a standard measure of market quality, they are often used in cross sectional or panel data regressions as dependent variables, see for example O'Hara and Ye (2009). In short, variance ratios are the *de facto* measure of predictability/market efficiency that is adopted universally by financial empiricists. It is important therefore that this class of tests be given a firm foundation.

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) provides some intellectual

---

[1] He then emphasized the more recent work that had shown that low frequency returns (business cycle and longer) are predictable from dividend price ratio and term premium variables.

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

credibility: he shows that they can be given a likelihood ratio test interpretation and are optimal against certain alternatives of the mean reverting type. In that sense they are similar to the Durbin-Watson test. The advantage of the variance ratio over the Box-Pierce statistic is that it gives some sense of the direction of predictability, which is lost in the Box-Pierce or other portmanteau tests. Hillman and Salmon (2007) have argued that the variance ratio (actually the related variogram) is better suited to irregularly spaced data and some kinds of nonstationarity than correlogram tests. Finally, there is a lot of work on improving the finite sample performance (size and power) of both Box-Pierce statistics and variance ratio statistics, see for example Kim, Nelson, and Startz (1991) and Kan and Wang (2010). See Charles and Darné (2009) for a recent review of this methodology and its application.

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

Second, we propose an alternative distribution theory and standard errors (*heteroskedasticity and leverage consistent HLCM*) than are usually adopted (i.e., in the univariate case). The limiting distribution established in Lo and MacKinlay (1988, Theorem 3) and repeated in CLM (and so used in most empirical studies) for the univariate variance ratio statistics is incorrect under their stated assumptions H1-H4 (i.e., RW3).[3] The correct distribution would be much more complicated and would depend on a long run variance that may be hard to estimate well. Either one makes additional assumptions to ensure that the variance is as claimed, which is what we propose below, or one has to use more complicated inference methods based on long run variance estimation, Newey and West (1987), or self normalization, Lobato (2001). In fact, the omitted condition appears quite innocu-

---

[3]It makes use of the CLT developed by White and Domowitz (1984) and used by many others.

4

ous, so their essential approach seems correct. However, we think that the no-leverage assumption (Lo and MacKinlay's H4) is untenable, empirically. Although this latter condition is satisfied by GARCH volatility processes with symmetrically distributed innovations, it is not satisfied by volatility processes that allow for leverage effects such as the GJR GARCH process or the Nelson's EGARCH process, and it is not even satisfied by standard GARCH volatility processes where the innovation is asymmetric. The statistical value of the restriction is that it simplifies the standard error calculation, although, as we show, the standard errors that allow for violations of this condition do not entail an inordinate increase in computation or complexity. Essentially, Lo and MacKinlay (1988) imposed an unnecessary assumption but fail to impose a necessary one. We propose modified assumptions that still preserve the possibility of simple inference methods but allow for leverage effects. Specifically, we establish the asymptotic distribution of our statistics under two sets of assumptions: (a) a stationary martingale difference hypothesis with fourth unconditional moments; (b) uncorrelatedness as in Lo and MacKinlay (1988) and with an additional uncorrelatedness condition on the products of returns but without the additional no-leverage condition. The asymptotic variance is the same under our two different sets of assumptions but is different from that contained in Theorem 3 of Lo and MacKinlay (1988) (and used in much subsequent empirical work). We remark that their theory essentially imposes that the sample autocorrelations are asymptotically uncorrelated, which can lead to inappropriate standard errors and p-values. This fact has been long appreciated in the time series literature, see for example Dufour and Roy (1985); Francq, Roy, and Zakoian (2005) have provided a comprehensive theory for Box-Pierce statistics under stationarity and mixing conditions.

We propose a simple analogue method for conducting inference that does not require the selection of a bandwidth parameter. We note that the evidence about predictability of asset returns in a large number of papers has been based on the Lo and MacKinlay (1988) standard errors, which we argue should be replaced by standard errors that rely on weaker and more plausible assumptions.[4] We show that in practice the standard errors can make a difference, especially when the time series is short (such as when stationarity is of concern).

Third, we extend our null hypothesis limit theory to the long horizon and large dimension cases. We derive the null limiting distribution of the studentized statistics under the increasing horizon framework, and show that asymptotic normality holds albeit with a slower rate of convergence, extending the univariate results of Chen and Deo (2006). We also establish the same result for the average scalar variance ratio statistic in the case where the horizon is fixed but the dimensions of the vector time series increase with sample size.

Fourth, we also establish the asymptotic properties of our statistic under several plausible alternative models including a multivariate Muth (1960) fads model and the recently developed bubble

---

[4]At the current count there were 3756 google citations of that paper.

process of Phillips and Yu (2011). These alternatives yield quite different predictions regarding the long run value of the variance ratio statistics.

Finally, we apply our methods to weekly returns for CRSP size-sorted portfolios from 1962 to 2013 in three subperiods 1962-1978, 1978-1994 and 1994-2013; the first two subperiods correspond to the data used in CLM. We show that the degree of inefficiency has reduced over the most recent period, and in some cases this improvement is statistically significant. Specifically, the univariate tests do not reject the null hypothesis for medium or large stocks in the most recent period. However, the multivariate tests do reject, albeit with a lower significance level. We have also extended our analysis to allow for a time varying risk premium, but find that the main empirical results are sustained, and we omit these results here.[5] This evidence is presented based on our HLCM standard errors that are robust to leverage effects as well as heteroskedasticity. We also show that the degree of asymmetry in the dependence structure has reduced, although it is still statistically significant. We further investigate the variance ratios at the long horizon. Simulation experiments indicate that our variance ratio tests are reliable, and powerful against some alternatives.

There is a substantial literature on testing for nonlinear predictability using information beyond the simple autocorrelations, see for example Hong (2000), Hong and Lee (2005), Escanciano and Velasco (2006), and Phillips and Jin (2014). There is also a literature that emphasizes structural breaks and rolling window analysis, see for example Lo (2005) and Pesaran and Timmermann (2007). Finally, there is a large literature on "predictive regressions" using long horizons and covariates such as dividend price ratios, see Phillips (2015). Our methodology and application hopefully complements this vast body of research.

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 discuss the large dimensional case. We perform a small simulation study in section 7. In section 8 we present our application, while Section 9 concludes. The appendix contains the proofs of all results.

---

[5]In the working paper version of this paper, we extend the theory to allow for a time varying risk premium in two ways. One approach is to fit an observable common factor regression and compute our statistics from the residuals. The second approach is to fit explicitly a nonparametric trend model, which we also allow to vary across different "regimes" (such as days of the week), to each series, and then to compute our statistics from the residuals. We show that with minor additional conditions our distribution theory and inference method carry over to this case.

2 Multivariate Variance Ratios

For expositional purposes we shall suppose in this section that we have a vector stationary ergodic discrete time series $X_t \in \mathbb{R}^d$; formal assumptions regarding the data are given below in section 3. Let $\widetilde{X}_t = X_t - \mu$, where $\mu = E(X_t)$ for all t. We are interested in testing the (weak form) Efficient Markets Hypothesis and quantifying and signing departures from this hypothesis. According to Fama (1970), this is that "risk adjusted stock returns are unforecastable using past prices". "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. Regarding the risk adjustment, we shall assume that the risk premium is constant, unknown, and is denoted by μ. In the working paper version we have extended the theory to the case where μ_t is time varying and depends on multiple unknown quantities.

One interpretation of the EMH is to assume that the risk adjusted return process satisfies

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

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

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

for all $i, j = 1, \ldots, d$ and for all $k \neq 0$, which itself is a stronger assumption[6] than that returns are uncorrelated with their own past, i.e.,

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

for all i and for all $k \neq 0$, which is what is adopted in Lo and MacKinlay (1988) (and referred to as RW3 in Campbell, Lo, and MacKinlay (1997) and in much subsequent work). RW3 has the advantage that if one rejects it, then one rejects the martingale hypothesis; on the other hand, if one does not reject RW3 then one can't conclude that the martingale hypothesis is valid.[7] Throughout we work with at least the multivariate uncorrelatedness hypothesis (2). We also develop a theory based on the stronger martingale difference assumption (1), because the additional regularity conditions can be stated very simply.

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

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

---

[6]This is not quite correct, since the martingale hypothesis only requires $E|X_t| < \infty$, whereas the autocovariance of a stationary process requires $EX_t^2 < \infty$ in order to be well defined in general.

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

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

$$\Sigma(K) = \text{var}(X_t(K)) = E((X_t(K) - KE(X_t))(X_t(K) - KE(X_t))^\top) \tag{6}$$

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

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

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

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

for $j = 0, \pm 1, \ldots$. Here, $A^{1/2}$ denotes a symmetric square root of a symmetric matrix A. We shall assume that Σ is strictly positive definite.

2.1 Two Sided Variance Ratios

Under condition (2), the variance covariance matrices obey the scaling law $\text{var}(X_t(K)) = K\text{var}(X_t)$, where K is some positive integer, from which we may obtain a number of different variance ratio statistics. These will have different merits and drawbacks depending on the purpose to which the estimation/testing is directed.

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

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

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

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

which is a symmetric matrix. The off-diagonal elements should be zero under the null hypothesis of no predictability. Both representations (11) and (12) can be used as the basis for estimation.[8]

An alternative multivariate normalization is given by

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

which can likewise generically be written

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

---

[8]One can interpret the variance ratio matrix as a (scalar) affine transformation of the least squares closest value of $\mathcal{R}$ in an approximating model for the autocorrelations of the form: $\mathcal{R}(j) = \left(1 - \frac{j}{K}\right)\mathcal{R}$, $j = 1, \ldots, K$ and $\mathcal{R}(j) = 0$ for $j > K$.

This has a regression interpretation, see Chitturi (1974) and Wang (2003, p62). Note that $\mathcal{VR}(K) = I$ if and only if $\mathcal{VR}a(K) = I$. We shall not say anything further about this quantity $\mathcal{VR}a(K)$. Some discussion is given in the working paper version of the paper.

A third quantity is the diagonally normalized variance ratio

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

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

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

2.2 One Sided Variance Ratios

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

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

and the negative counterparts $\mathcal{VR}_-(K) = \mathcal{VR}_+^{\mathsf{T}}(K)$ and $\mathcal{VR}d_-(K) = \mathcal{VR}d_+^{\mathsf{T}}(K)$, which have the property that: $\mathcal{VR}(K) = (\mathcal{VR}_+(K) + \mathcal{VR}_+^{\mathsf{T}}(K))/2$ and $\mathcal{VR}d(K) = (\mathcal{VR}d_+(K) + \mathcal{VR}d_+^{\mathsf{T}}(K))/2$. One can compare the two statistics, $\mathcal{VR}d_+(K), \mathcal{VR}d_-(K)$, to quantify the asymmetry in lead lag effects.

2.3 Univariate Parameters of Interest

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

2.3.1 Trace and Determinant

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

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

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

2.3.2 Eigenvalues

Let $\lambda_1(K) = \lambda_{\max}(K) \geq \ldots \geq \lambda_d(K) = \lambda_{\min}(K)$ denote the eigenvalues of $\mathcal{VR}(K))$ arranged in decreasing order. Under the null hypothesis, $\lambda_j(K) = 1$, $j = 1, \ldots, d$, but under the alternative hypothesis they can take any non-negative values. These quantities give univariate measures of the predictability obtainable within the series as we next show. Consider a portfolio of assets with fixed weights $w \in \mathbb{R}^d$. Denoting $\mathcal{VR}_K(z_t)$ by the univariate variance ratio of the scalar series z_t, and letting $\widetilde{w} = \Sigma^{1/2}w$ and $Y_t = \Sigma^{-1/2}X_t$, we have

$$
\begin{aligned}
\mathcal{VR}_K(w^\intercal X_t) &= \mathcal{VR}_K(w^\intercal \Sigma^{1/2}\Sigma^{-1/2}X_t) = \mathcal{VR}_K(\widetilde{w}^\intercal Y_t) = \frac{\widetilde{w}^\intercal \mathcal{VR}(K; Y_t)\widetilde{w}}{\widetilde{w}^\intercal \widetilde{w}} \\
&= \frac{\widetilde{w}^\intercal \mathcal{VR}(K; X_t)\widetilde{w}}{\widetilde{w}^\intercal \widetilde{w}} \leq \lambda_1(\mathcal{VR}(K; X_t)).
\end{aligned}
$$

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

2.3.3 Global Minimum Variance

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

$$GMV(K) = \frac{1}{i^\intercal \mathcal{VR}(K)^{-1}i}. \tag{16}$$

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

2.3.4 Off-Diagonal Elements

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

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

$$CS(K) = \frac{2}{d(d-1)} \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \mathcal{VR}d_{ij}(K) = \frac{1}{d(d-1)}\{i^\intercal \mathcal{VR}d(K)i - \text{tr}(\mathcal{VR}d(K))\}, \tag{17}$$

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

2.3.5 Dynamic Momentum/Contrarian Portfolio Profit

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

$$\widetilde{w}_{it}(K) = \pm\frac{2}{d(K-1)} \sum_{j=1}^{K-1} \left(1 - \frac{j}{K}\right)\left(Z_{i,t-j} - \overline{Z}_{t-j}\right) \tag{18}$$

11

where $\overline{Z}_s = \sum_{i=1}^{d} Z_{is}/d = i^\mathsf{T} Z_s/d$ so that $\sum_{i=1}^{d} \widetilde{w}_{it}(K) = 0$. This strategy considers all the "signals": $Z_{i,t-1} - \overline{Z}_{t-1}, \ldots, Z_{i,t+1-K} - \overline{Z}_{t+1-K}$, and combines them with weights according to their lag. If the $\pm$ factor is positive, this can be considered a momentum strategy, while if it is negative, this can be considered a contrarian strategy. The total investment of the strategy at time t is $I_t(K) = \sum_{i=1}^{d} |\widetilde{w}_{it}(K)|/2$. The expected profit of this strategy is

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

Under the martingale hypothesis, $\pi_\pm(K) = 0$ for all K. This quantity weights diagonal departures and off diagonal departures similarly. If $\pi_\pm(K) > 0$, then the strategy should make money (in the absence of transaction costs).

3 Estimation of Variance Ratio Matrices

Suppose that we observe the return vectors $\{X_t, t = 1, \ldots, T\}$ equally spaced in discrete time. We may estimate the variance ratios in several ways, for example by estimating the sample covariance matrix of the K frequency data and the original observations and then forming the ratio.[9] We can alternatively explicitly use the population connection with the autocorrelation matrix process in (12) for example.

We estimate the population quantities by sample averages:

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

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

---

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

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

$$\widehat{\mathcal{VR}}(K) = I + \sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)(\widehat{R}(j) + \widehat{R}(j)^{\mathsf{T}}) \quad ; \quad \widehat{\mathcal{VR}}d(K) = I + \sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)(\widehat{R}d(j) + \widehat{R}d(j)^{\mathsf{T}})$$

$$\widehat{\mathcal{VR}}^{\&}(K) = \widehat{\Sigma}^{-1/2}\widehat{\Sigma}(K)\widehat{\Sigma}^{-1/2}/K \quad ; \quad \widehat{\mathcal{VR}}_+(K) = I + 2\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)\widehat{R}(j).$$

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

We may also calculate the univariate quantities by analogy. For example, define the estimated ordered eigenvalues $\widehat{\lambda}_1(K) \geq \ldots \geq \widehat{\lambda}_d(K)$ of $\widehat{\mathcal{VR}}(K)$.

4 Asymptotic Theory and Inference

4.1 Regularity Conditions

We present two alternative non-nested sets of sampling assumptions, which we denote by A and MH*. Assumptions A center on the martingale difference assumption and require stationarity and ergodicity. The theory makes use of arguments presented in Hall and Heyde (1980), and applied in Phillips and Guo (2001); see Escanciano and Lobato (2009) for a review of the literature surrounding martingale based testing. Assumptions MH* are modified versions of the assumptions in Lo and MacKinlay (1988) adapted to the multivariate case and corrected for what appears to be an error; these conditions do not require stationarity although certain averages need to converge. Most treatments of variance ratios employ 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.[10] In the multivariate context, their assumption H4 would be that

$$E[\widetilde{X}_{it}\widetilde{X}_{jt}\widetilde{X}_{kr}\widetilde{X}_{ls}] = 0 \text{ for all } i, j, k, l, t, \text{ and } r, s \text{ with } r < s < t. \tag{19}$$

This assumption rules out leverage type effects, e.g., $E[\widetilde{X}_{it}^2|\widetilde{X}_{ir}\widetilde{X}_{is}] \neq 0$, which may be important for some assets, see Nelson (1991). This assumption is not necessary for the distribution theory; imposing it (along with other conditions) would simplify the asymptotic variance to be single finite sums rather than double finite sums, but in practice this is not a big issue. We shall dispense with this assumption below, but we shall make a further assumption that appears to have been omitted by mistake from Lo and MacKinlay (1988). Namely, implicit in their analysis is that $\widetilde{X}_t\widetilde{X}_{t-j}$ is uncorrelated with

---

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

$\widetilde{X}_s \widetilde{X}_{s-j}$, but this does not follow from $\widetilde{X}_t$ being an uncorrelated sequence (although it does follow if $\widetilde{X}_t$ were a martingale difference sequence).

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

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

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

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

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

ASSUMPTION A.

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

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

ASSUMPTION MH*.

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

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

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

Chen and Deo (2006) work with martingale difference sequences but also assume a no leverage condition. Francq, Roy, and Zakoian (2005) assume both stationarity and mixing in their analysis of Box-Pierce statistics. In MH* we include the additional condition (ii) $E[\widetilde{X}_t \widetilde{X}_{t-j}^{\intercal} \otimes \widetilde{X}_s \widetilde{X}_{s-k}^{\intercal}] = 0$, for all $s \ne t$ 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, involving the selection of a bandwidth parameter. Condition MH1(ii) is satisfied

14

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 MH* we have assumed higher moments depending on the mixing decay rate, whereas for assumption A only four moments are required and no explicit mixing conditions are employed. It should be noted therefore that the conditions A and MH* are non-nested. We further note that under the assumption that returns are i.i.d. (referred to as RW1 in Campbell, Lo, and MacKinlay (1997)), the univariate version of the CLT's below are valid under only second moments, Brockwell and Davis (1991, Theorem 7.2.2), due to the self normalization present in the sample autocorrelations. For similar reasons, condition MH3 may not be strictly necessary in that mildly trending moments may still permit a CLT at the same rate due to the cancellation of numerator by denominator.

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

4.2 Finite/fixed horizon Limiting Distribution Theory

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

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

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

It follows that for any vector ω, $\omega^\intercal \text{vec}(\widehat{\mathcal{VR}}_+(K) - I_d)$ is asymptotically normal with mean zero

---

[11]For stationary univariate linear processes, the sample autocorrelations can be root-T consistent and asymptotically normal under only second moment assumptions, Brockwell and Davis (1991, Theorem 7.2.2), but this result does not hold for nonlinear processes like GARCH, nor for multivariate linear processes.

and variance $\omega^{\mathsf{T}} Q(K)\omega/T$. Limiting distributions for smooth functions of the variance ratio matrices can be obtained by the delta method.

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

$$\sqrt{T}\left(\varphi(\widehat{\mathcal{VR}}(K)) - i_d\right) \Longrightarrow \varphi(U), \tag{20}$$

where $\varphi(\widehat{\mathcal{VR}}(K))$ and $\varphi(U)$ are $d \times 1$ vectors of ordered eigenvalues $\widehat{\lambda}_j \in \varphi(\widehat{\mathcal{VR}}(K))$ and $\lambda_j^* \in \varphi(U)$, respectively. It follows for example that

$$\sqrt{T}\left(\widehat{\lambda}_{\max} - 1\right) \Longrightarrow \lambda_{\max}^*(U),$$

whose distribution can be computed by simulation.

4.3 Standard Errors and Test Statistics

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

$$\widehat{\Xi}_{jk} = \frac{1}{T}\sum_{t=\max\{j,k\}+1}^{T} \left(X_{t-j} - \overline{X}\right)\left(X_{t-k} - \overline{X}\right)^{\mathsf{T}} \otimes \left(X_t - \overline{X}\right)\left(X_t - \overline{X}\right)^{\mathsf{T}} \tag{21}$$

$$\widehat{Q}(K) = \sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_{j,K}c_{k,K}\left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2}\right)\widehat{\Xi}_{jk}\left(\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2}\right) \tag{22}$$

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

and $\widehat{S}(K) = D_n^+\widehat{Q}(K)D_n^{+\mathsf{T}}$ and $\widehat{S}d(K) = D_n^+\widehat{Q}d(K)D_n^{+\mathsf{T}}$, where D_n^+ is the Moore-Penrose pseudoinverse of the duplication matrix, Magnus and Neudecker (1980). Specifically, the asymptotic variance of $\widehat{\mathcal{VR}d}_{ii}(K)$ can be estimated by

$$\widehat{Q}d_{iiii}(K) = \frac{1}{\widehat{\sigma}_{ii}^2}\sum_{j=1}^{K-1}\sum_{k=1}^{K-1} c_{j,K}c_{k,K}\widehat{\Xi}_{jk;iiii} \tag{23}$$

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

16

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

Note that under the Lo and MacKinlay (1988) condition H4 (i.e. (19)) we have $\Xi_{jk} = 0$ for $j \neq k$, so that the asymptotic variance in Theorem 1 simplifies, a little. The commonly used asymptotic variance matrix is

$$\widehat{Q}d_{LM}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 \left(\widehat{D}^{-1/2} \otimes \widehat{D}^{-1/2} \right) \widehat{\Xi}_{jj} \left(\widehat{D}^{-1/2} \otimes \widehat{D}^{-1/2} \right), \tag{24}$$

whose diagonal elements can be compared with (23): they are the same except that $\widehat{\Xi}_{jk;iiii} = 0$ for $j \neq k$. In the iid case, we further have $\Xi_{jj} = \Sigma \otimes \Sigma$ and:

$$Q_{iid}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 I_{d^2} \; ; \; \widehat{Q}d_{iid}(K) = \sum_{j=1}^{K-1} c_{j,K}^2 (\widehat{Rd}(0) \otimes \widehat{Rd}(0)). \tag{25}$$

In the scalar case both these quantities are nuisance parameter free.

As we show in the application, the standard errors derived from (22), (24), and (25) can be quite different; although there is no necessary ordering, 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 $\widehat{Q}_{iid}(K)$.

The standard errors for univariate quantities of interest can be obtained from (22). Let $\tau_f = f(\text{vec}(\mathcal{VR}_+(K)))$ and $\tau d_f = f(\text{vec}(\mathcal{VR}d_+(K)))$ be scalar parameters of interest, where f is a continuously differentiable function with non-zero gradient, and let $e_f = \nabla f(\text{vec}(\mathcal{VR}_+(K))), ed_f = \nabla f(\text{vec}(\mathcal{VR}d_+(K))) \in \mathbb{R}^{d^2}$ denote the gradients of the functions at the true value. Let $Q_f(K) = e_f^\intercal Q(K) e_f$ and $Qd_f(K) = e_f^\intercal Qd(K) e_f$. Then $\widehat{Q}_f(K) = e_f^\intercal \widehat{Q}(K) e_f$ and $\widehat{Qd}_f(K) = e_f^\intercal \widehat{Qd}(K) e_f$ are consistent asymptotic variance estimators for $\widehat{\tau}_f$ and $\widehat{\tau d_f}$ respectively. For example, define the column vectors: b that is 0 at the $((l-1)(d+1)+1)^{th}$ entries $(l = 1, \ldots, d)$ and 1 otherwise; i is a conformable column vector of ones; c is a column vector that is $(1-d)/(d^2(K-1))$ at $((l-1)(d+1)+1)^{th}$ entries, and is $1/(d^2(K-1))$ at other entries; and $\delta = \text{vech}(I_d)$. Then, specifically, let:

$$\widehat{Q}_{CS}(K) = \frac{1}{d^2(d-1)^2} b^\intercal \widehat{Qd}(K) b$$

$$\widehat{Q}_{GMV}(K) = d^{-4} i^\intercal \widehat{S}(K) i,$$

$$\widehat{Q}_\pi(K) = c^\intercal \widehat{Qd}(K) c$$

$$\widehat{Q}_{tr}(K) = \delta^\intercal \widehat{S}(K) \delta = \widehat{Q}_{\det}(K).$$

We next define some test statistics. Let f be any continuously differentiable function with nonzero gradient (for example $CS, \det, GM, tr,$ or π), and let

$$Z_f(K) = \sqrt{T} \left(\widehat{Q}_f(K) \right)^{-1/2} \left[f(\mathrm{vec}(\widehat{\mathcal{VR}}_+(K))) - f(\mathrm{vec}(\mathcal{VR}_+(K))) \right] \tag{26}$$

$$Zd_f(K) = \sqrt{T} \left(\widehat{Q}d_f(K) \right)^{-1/2} \left[f(\mathrm{vec}(\widehat{\mathcal{VR}d}_+(K))) - f(\mathrm{vec}(\widehat{Rd}(0))) \right] \tag{27}$$

$$W_F(K) = T\mathrm{vech} \left(\widehat{\mathcal{VR}}(K) - I \right)^{\intercal} \widehat{S}(K)^{-1} \mathrm{vech} \left(\widehat{\mathcal{VR}}(K) - I \right) \tag{28}$$

$$Wd_F(K) = T\mathrm{vech} \left(\widehat{\mathcal{VR}d}(K) - \widehat{Rd}(0) \right)^{\intercal} \widehat{Sd}(K)^{-1} \mathrm{vech} \left(\widehat{\mathcal{VR}d}(K) - \widehat{Rd}(0) \right). \tag{29}$$

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

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

$$Z_f(K), Zd_f(K) \Longrightarrow N(0,1)$$

$$W_F(K), Wd_F(K) \Longrightarrow \chi^2(d(d+1)/2).$$

In the application we make use of a bias correction method based on asymptotic expansions (under the iid assumption), which may give better performance for long lags. A number of alternative inference methods such as self-normalization, or block bootstrap and subsampling have been suggested to accommodate the more general uncorrelatedness assumption that allows $E\left[\widetilde{X}_t \widetilde{X}_{t-j}^{\intercal} \otimes \widetilde{X}_s \widetilde{X}_{s-k}^{\intercal} \right] \neq 0$ for some $s \neq t$. The readers are directed to Lobato (2001) and Whang and Kim (2003) for description of these methods.

4.4   Increasing horizon Limiting Distribution Theory

It has been reported in the literature that inferences based on the asymptotic theory of the variance ratio statistic become unreliable in finite samples when the horizon K is large relative to the sample size T, see Lo and MacKinlay (1989). In view of this practical issue, Richardson and Stock (1989) considered the framework in which $K = K(T)$ and $K/T \to \delta < 1$, and showed that the limiting distribution is a function of Brownian motion. However, Deo and Richardson (2003) pointed out the inconsistency of the univariate variance ratio test under this particular restriction against some important mean reverting alternatives. Consequently, Chen and Deo (2006) studied an alternative setting where K is set to increase slower so that K/T tends to zero. Along with the ergodic martingale difference assumption, they imposed a set of strong conditions on cross-moments (Assumption A3) including the no-leverage condition, and some mixing-type conditions (Assumption A5 and A6) that imply asymptotic independence of the process.

In this section we investigate the increasing K asymptotics in the multivariate framework. Although a d-dimensional analogue of the conditions assumed in Chen and Deo (2006) can be adopted, we shall consider a different set of conditions including stationarity in Assumption A1 (but with a slightly higher moment condition). This is to be consistent with the previous fixed K theory, and to allow simple derivations under mild assumptions.

ASSUMPTION A′. *The process $\widetilde{X}_t$ is a stationary ergodic Martingale Difference sequence having finite $4 + \delta$ moments, i.e. $E|\widetilde{X}_{it}|^{4+\delta} \leq C < \infty$ for some $\delta > 0$ for all i.*

ASSUMPTION T. *The horizon $K \to \infty$ as $T \to \infty$ and $K/T \to 0$.*

ASSUMPTION S. *The following double sum is finite: $\sum_{a=-\infty}^{\infty} \sum_{b=-\infty}^{\infty} |\kappa_{pqrl}(a, b, 0, 0)| < \infty$ for all $p, q, r, l \leq d$, where $\kappa_{pqrl}(t_1, t_2, t_3, t_4)$ is the cumulant of 4th order between $(\widetilde{X}_{pt_1}, \widetilde{X}_{qt_2}, \widetilde{X}_{rt_3}, \widetilde{X}_{lt_4})$.*

Along with stationarity, Assumption S guarantees the existence and positive definiteness of the matrix limits $Q(\infty) = \lim_{K \to \infty} K^{-1}Q(K)$ and $Qd(\infty) = \lim_{K \to \infty} K^{-1}Qd(K)$, both of which will turn out to have simple forms as we will see below. Indeed, summability of the cumulants is a common assumption in the time series literature, see Rosenblatt (1985). The weak condition regulates the dependence structure of the process, and is implied by a mild α-mixing and moment condition (strictly higher than 4 as we shall assume below) as shown in Andrews (1991, Lemma 1), although it is stronger than ergodicity. For example, Assumption MH2 with size of mixing strengthened to $3r/(r-1)$ is sufficient for summability of cumulants under stationarity. Some relevant discussions can be found in the recent paper Shao and Wu (2007), where an alternative sufficient condition is given in terms of the notion of geometric moment contraction (GMC).

We derive the limiting distribution under the stationary ergodic martingale difference assumption. Note that one could alternatively work with mixing (Assumption MH*) or near epoch dependence, for example, to obtain a similar result, but we shall not proceed to this direction in this paper.

THEOREM 2. *Suppose that Assumptions A', T and S hold. Then, as $T \to \infty$:*

$$\sqrt{\frac{T}{K}}\text{vec}\left(\widehat{\mathcal{VR}}(K)_+ - I_d\right) \implies N\left(0, \frac{4}{3}J\right)$$

$$\sqrt{\frac{T}{K}}\text{vec}\left(\widehat{\mathcal{VR}}(K)_+ - I_d\right) \implies N\left(0, \frac{4}{3}Jd\right),$$

and

$$Z_f(K), Zd_f(K) \implies N(0,1)$$

$$W_F(K), Wd_F(K) \implies \chi^2(d(d+1)/2),$$

where J is the idendty matrix of dimension $d^2 \times d^2$, and $Jd = (D^{-1/2} \otimes D^{-1/2})(\Sigma \otimes \Sigma)(D^{-1/2} \otimes D^{-1/2})$ is the matrix whose diagonal entries are one.

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

5 Alternative Hypotheses

There are many plausible alternative hypotheses to the null hypothesis (2), and it is not possible in general to have power against all such departures. We can understand a little bit better the type of alternatives against which the variance ratio has power by looking at equation (12). We have $\mathcal{VR}(K) = I$ if and only if

$$\sum_{j=1}^{K-1} c_{j,K}\left(\mathcal{R}(j) + \mathcal{R}(j)^\intercal\right) = \mathbf{0}.$$

This says that the test will have power against alternatives for which the Bartlett weighted autocorrelations do not sum identically to zero. This seems like a reasonable class of alternative, because if the autocorrelations change sign enough that they cancel out, this hardly seems like a propitious setting to make excess returns from a trading strategy that treats these autocorrelations as signals. One wants not just departures from zero but some kind of reliable direction of dependence on which to bet. By contrast, the Box-Pierce statistic will also pick up highly oscillatory variation in the autocorrelations, which one might prefer to exclude from consideration.

We look in detail at several alternative models in this section. In general they yield a prediction of the form

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

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

5.1 Local Alternatives

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

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

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

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

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

THEOREM 3. *Suppose that the data are normally distributed. Then, the trace test is locally most powerful invariant against alternatives in the $\phi-$best class of the form $q_T^* = \delta/\sqrt{T}$.*

It may be possible to characterize the class of alternatives against which other tests, such as the determinant test, are optimal, but we leave this for future research.

The trace test, while optimal against the specific class above, may have zero power against some

alternatives, as we next discuss. Suppose that $\Delta(K, T) = \Delta(K)/\sqrt{T}$, then

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

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

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

for some nonzero $\delta(K)$, then this particular test will have zero power.

5.2 Multivariate Fads Model

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

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

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

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

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

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

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

and likewise for $\mathcal{VR}d(K)$. In general, however, η_t might have any type of weak dependence structure.

We next derive a restriction on the long run variance ratio statistic that reflects the presence of fads. We do not restrict the fads process, and so can only obtain long run implications.

THEOREM 4. *Suppose that the multivariate fads model (32)-(33) holds and suppose that* $\mathrm{cov}(\eta_{t+j}, \eta_t) \to 0$ *as* $j \to \infty$. *Then,* $\mathcal{VR}(\infty) = \lim_{K \to \infty} \mathcal{VR}(K) = I + \sum_{j=1}^{\infty} (R(j) + R(j)^{\intercal})$ *exists. Further suppose that* $\Omega_\eta(1) > 0$. *Then,*

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

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

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

This result generalizes the existing results for the scalar fads process, which amount to $\mathcal{VR}d_{ii}(\infty) \leq Rd_{ii}(0)$ for $i = 1, \ldots, d$. In Theorem 4, we obtain stronger constraints on the off diagonal elements of $\mathcal{VR}d(\infty)$ and $\mathcal{VR}(\infty)$. Note that we also obtain $GMV(K) \to GMV(\infty) > 1/d$ as a corollary.

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

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

where e_t are i.i.d random vectors of length $p \geq d$, and $\Psi : \mathbb{R}^p \times \mathbb{R}^p \times \cdots \to \mathbb{R}^d$. This includes a wide range of linear and nonlinear processes for η_t, ε_t. Then define

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

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

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

THEOREM 5. *Suppose that the multivariate fads model (32)-(33) holds along with Assumption B, and suppose that* $K \to \infty$ *as* $T \to \infty$ *such that* $K/T \to 0$. *Then,*

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

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

5.3 Bubble Process

Several authors argue that the frequently observed excessive volatility in stock prices may be attributed to the presence of speculative bubbles. Blanchard and Watson (1982) and Flood and Hodrick (1986), inter alia, demonstrate in a theoretical framework that bubble components potentially generate excessive volatility. There is some debate about whether these constitute rational adjustment to fundamental pricing rules or arise from more behavioural reasons. Recently, Phillips and Yu (2011), and Phillips, Shi, and Yu (2012) have considered the following class of "bubble processes" for (log) prices p_t

$$p_t = \mu + p_{t-1} 1\left(t < \tau_e\right) + \delta_T 1\left(\tau_e \leq t \leq \tau_f\right) p_{t-1} + \left(\sum_{s=\tau_f+1}^{t} \varepsilon_s + p_{\tau_f}^*\right) 1\left(t > \tau_f\right) + \varepsilon_t 1\left(t \leq \tau_f\right), \quad (35)$$

where $p_{\tau_f}^*$ represents the restarting price after the bubble collapses at time τ_f, and $\delta_T = 1 + c/T^\alpha$ for $\alpha \in (0,1)$ and $c > 0$. The process is consistent with the efficient markets hypothesis during $[1, \tau_e]$ and $[\tau_f, T]$ but has an explosive "irrational" moment in the middle. They propose econometric techniques to test for the presence of a bubble and indeed multiple bubbles. One can imagine this model also holding for a vector of asset prices caught up in the same bubble, so that ε_t is a vector of shocks, the indicator function is applied coordinatewise, and the coefficient δ_T is replaced by a diagonal matrix.

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

$$\widehat{\mathcal{VR}}(K) \xrightarrow{P} K \qquad (36)$$

for all K, so that the variance ratio statistic is greater than one for all K and gets larger with horizon. Essentially, the bubble period dominates all the sample statistics, and all return autocorrelations converge to one inside the bubble period, thereby making the ratio equal to the maximum it can achieve. In the multivariate case, Magdalinos (2014) has shown that in some special cases, $\lambda_{\max}(\widehat{\mathcal{VR}}(K)) \xrightarrow{P} K$. However, the multivariate case is more complicated because other eigenvalues may not behave in the same way.

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

5.4 Time Varying Expected Return

We briefly consider a simple statistical model for time varying expected return. This model could be consistent with rational pricing where the risk premium evolves slowly over time and has small variation relative to the shocks to risk adjusted returns. Specifically, suppose that observed returns are composed of a slowly varying risk premium μ_t and an iid shock ε_t, i.e.,

$$X_t = \mu + \mu_t + \varepsilon_t, \tag{37}$$

$$\mu_t = \mu_{t-1} + \eta_t, \tag{38}$$

where $\mu_0 = (0, 0, \ldots, 0)^\intercal$ and η_t is an iid mean zero shock that is "small" relative to ε_t. In this case observed returns are nonstationary so we must index populations by T. This specification is similar to that of equation 7.1.30 of CLM. We establish the following result.

THEOREM 6. *Suppose that the model (37)-(38) holds with η_t iid mean zero with $E\eta_t\eta_t^\intercal = \Sigma_\eta/T > 0$ and ε_t iid mean zero with $E\varepsilon_t\varepsilon_t^\intercal = \Sigma_\varepsilon > 0$. Then*

$$\lim_{K\to\infty} K^{-1} \lim_{T\to\infty} \mathcal{VR}_T(K) \to \left(\Sigma_\eta \frac{1}{2} + \Sigma_\varepsilon\right)^{-1/2} \frac{1}{2}\Sigma_\eta \left(\Sigma_\eta \frac{1}{2} + \Sigma_\varepsilon\right)^{-1/2} < I_d.$$

This model gives a similar prediction to the bubble model, except it says that all eigenvalues should grow linearly with the horizon with a slope less than one.

In the working paper version of this paper we consider several alternative approaches to capturing time varying expected returns including nonparametric mean model and linear factor models.

5.5 Locally Stationary Alternatives

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

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

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

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

$$\Gamma(j; u) = E(\widetilde{X}_t(u)\widetilde{X}_{t-j}^\intercal(u)).$$

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

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

It follows that under local stationarity

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

The test will have power against some alternatives where $\Gamma_u(j) \neq 0$ for $u \in U$ and $\Gamma_u(j) = 0$ for $u \in U^c$. The test will not detect alternatives where $\overline{\mathcal{VR}}(K) = I$ but $VR(K; u) = I + \sum_{j=1}^{K-1}(1 - \frac{j}{K})(R(j; u) + R(j; u)^\intercal) \neq 0$, where $R(j; u) = \Sigma(u)^{-1/2}\Gamma(j; u)\Sigma(u)^{-1/2}$.

5.6   Nonlinear Processes

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

6   Large Dimensional Data

We briefly consider some issues that arise when the dimensions d are large. In this case, the covariance matrices Σ and $\Sigma(K)$ may be ill conditioned, and so forming the ratio (11) may not be practically feasible or theoretically valid; likewise for any functions derived thereof such as the smallest eigenvalues. The diagonal variance ratio matrix and simple univariate quantities derived from it like $CS(K)$ may fare better in this situation, since the marginal variances should be bounded away from zero. We remark that Castura, Litzenberger, Gorelick, and Dwivedi (2010) report the average variance ratio of the Russell 1000 and Russell 2000 stocks, which amounts to $\sum_{i=1}^d \widehat{\mathcal{VR}d}_{ii}(K)/d$. They do not report standard errors for this quantity, perhaps on the grounds that d is large (since $d = 3000$). However, when the individual stocks are contemporaneously correlated, which they typically are[12], the averaging will not reduce the order of magnitude of the standard error. Specifically, under the iid

---

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

assumption, the correlation between $\widehat{\mathcal{VR}d}_{ii}(K)$ and $\widehat{\mathcal{VR}d}_{jj}(K)$ will be proportional to ρ_{ij}^2, where ρ_{ij} is the contemporaneous correlation between the returns on stock i and stock j. We show below how to calculate the standard errors for $\sum_{i=1}^{d} \widehat{\mathcal{VR}d}_{ii}(K)/d$ in the large d, T case. However, for nonlinear functions of $\mathcal{VR}d(K)$ such as its eigenvalues, or for quantities derived from $\mathcal{VR}(K)$, the large d theory is more complicated.

We present a simple result for the average trace statistic in the case where d grows but at a rate slower than T. We suppose that Assumption A′ holds for the d-dimensional vector process $\widetilde{X}_t$, and impose Assumption Sd below to ensure that the limiting variance is well-defined.

ASSUMPTION TD. *The dimension $d = d(T) \to \infty$ in such a way that $d/T \to 0$ as $T \to \infty$.*

ASSUMPTION SD. *The limit of the quadruple sum $qd(\infty)^* \in (0, \infty)$ exists, where*

$$qd(\infty)^* := \lim_{d \to \infty} \frac{1}{d^2} \sum_{i=1}^{d} \sum_{r=1}^{d} \left(\sum_{j=1}^{K-1} \sum_{k=1}^{K-1} \frac{c_{j,K} c_{k,K}}{\sigma_{ii} \sigma_{rr}} E\left[\widetilde{X}_{it} \widetilde{X}_{rt} \widetilde{X}_{i,t-j} \widetilde{X}_{r,t-k} \right] \right), \tag{39}$$

where σ_{ii} are the diagonal elements of Σ.

Under these conditions, we can derive the following asymptotic normality result:

THEOREM 7. *Suppose that Assumptions A′, S, Td, and Sd hold. Then:*

$$Zd_{tr}(K) \Longrightarrow N(0, 1).$$

We remark that the cross-sectional standard deviation of the individual variance ratios, a quantity that is often reported along with the average variance ratio, see for example CLM Table 2.7, is not necessarily related in any simple way to the true asymptotic standard deviation of the estimator that we report here.

An alternative strategy in the large d case may be to calculate scalar ratios from the matrix scaling law $\Sigma(K) = K\Sigma$. Specifically, we may look at quantities like $\lambda_{\max}(K)/K\lambda_{\max}(1)$ whose properties may follow from generalizations of results in Jin, Wang, Bai, Krishnan, and Harding (2014). However, when d is comparable with T, one must use some sparsity structure or shrinkage method to obtain reasonable performance for complicated nonlinear functions of the covariance matrices. Johnstone and Onatski (2015) develop a comprehensive theory for multivariate testing in large dimensional situations.

7 Simulation Study

We perform a small simulation study to assess the reliability of our multivariate variance ratio test statistics (the earlier version of this paper contains additional results not reported here for brevity). In particular we examine two multivariate variance ratio tests: the trace ($Z_{tr}(K)$) and the determinant ($Z_{\det}(K)$) tests.

We first simulate empirical size of nominal 5% multivariate variance ratio tests based on $Z_{tr}(K)$ and $Z_{\det}(K)$ statistics for the null hypothesis H_0 : $X_t = (X_{1,t}, X_{2,t})^\top$ is m.d.s. specified by the following bivariate constant conditional correlation (CCC)-GARCH (1,1) model:

$$X_{1,t} = \sqrt{h_{1,t}}\varepsilon_{1,t}, \ X_{2,t} = \sqrt{h_{2,t}}\varepsilon_{2,t}$$

$$h_{1,t} = 0.2 + 0.05X_{1,t-1}^2 + 0.9h_{1,t-1}$$

$$h_{2,t} = 0.1 + 0.08X_{2,t-1}^2 + 0.9h_{2,t-1}$$

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \rho = 0.5.$$

Based on 10000 replications, we have the following results.

Table 7-1: Empirical size of nominal 5% multivariate variance ratio tests
[using $Z_{tr}(K)$ and $Z_{\det}(K)$ statistics]

		Size of 5 percent test	
Sample size	K	$Z_{tr}(K)$	$Z_{\det}(K)$
1024	2	0.0488	0.0481
1024	4	0.0478	0.0455
1024	8	0.0467	0.0437
1024	16	0.0507	0.0422

Table 7-1 shows that the empirical sizes of variance ratio tests using $Z_{tr}(K)$ and $Z_{\det}(K)$ statistics are all close to the nominal value 5%. We then examine the power of multivariate variance ratio tests based on $Z_{tr}^{(iid)}(K)$ and $Z_{\det}^{(iid)}(K)$ statistics, at a 5% nominal level, against the alternative hypotheses H_1 : bivariate fads model for log prices, specified as (32) and (33) with $\mu = 0$ and $\eta_t = \beta\eta_{t-1} + \xi_t$, where $\varepsilon_t \sim i.i.d.N(0, \Omega_\varepsilon)$, $\xi_t \sim i.i.d.N(0, I_d)$, ε_t and η_t are mutually independent, $\beta = \begin{bmatrix} 0.95 & 0.02 \\ 0.05 & 0.9 \end{bmatrix}$. We consider three cases: $\Omega_\varepsilon = 2I_d$, $\Omega_\varepsilon = I_d$ and $\Omega_\varepsilon = \frac{1}{2}I_d$, so that the conditional variability of the random walk relative to the stationary component is two, one and one-half, respectively. We consider $Z_{tr}^{(iid)}(K)$ and $Z_{\det}^{(iid)}(K)$ statistics which are similarly defined as $Z_{tr}(K)$ and $Z_{\det}(K)$ but using $\widehat{Q}_{iid}(K)$. Based on 10000 replications, we have the following results.

Table 7-2: Power of multivariate variance ratio tests at a 5% nominal level [using $Z_{tr}^{(iid)}(K)$ and $Z_{\det}^{(iid)}(K)$ statistics]

Sample size	K	$\Omega_\varepsilon = \frac{1}{2}I_d$		$\Omega_\varepsilon = I_d$		$\Omega_\varepsilon = 2I_d$	
		$Z_{tr}^{(iid)}(K)$	$Z_{\det}^{(iid)}(K)$	$Z_{tr}^{(iid)}(K)$	$Z_{\det}^{(iid)}(K)$	$Z_{tr}^{(iid)}(K)$	$Z_{\det}^{(iid)}(K)$
1024	2	0.2021	0.1971	0.1357	0.1324	0.0844	0.0813
1024	4	0.3933	0.3806	0.2399	0.2273	0.1317	0.1216
1024	8	0.6334	0.6183	0.3932	0.3658	0.1980	0.1728
1024	16	0.8229	0.8009	0.5331	0.4716	0.2653	0.2061

Table 7-2 shows that the power of the tests increases with K. In addition, as the conditional variability of the random walk relative to the stationary component decreases, the power of tests increases, for example, when $\Omega_\varepsilon = \frac{1}{2}I_d$ and $K = 16$, the power of tests is very high which goes beyond 80%. Furthermore, we found the tests based on $Z_{tr}^{(iid)}(K)$ statistics are more powerful than those based on $Z_{\det}^{(iid)}(K)$ statistics across all cases.

8 Application

We apply our methodology to U.S. stock return data. In particular, we use weekly size-sorted equal-weighted portfolio returns from the Center for Research in Security Prices (CRSP) from 06/07/1962 to 27/12/2013.[13] Essentially we are using the extension of the same data that were used in Lo and MacKinlay (1988) and Campbell, Lo and Mackinlay (1997), which allows us to make comparison with their results, and to extend it to the more recent period. In the following parts, we first test the linear predictability for size-sorted CRSP portfolio returns at short to medium horizon; then we investigate the long-run behavior of variance ratio statistics.

8.1 Short to Medium Horizon

8.1.1 Evidence on Linear Predictability

Consider weekly returns for three size-sorted CRSP portfolios $X_t = (X_{1t}, X_{2t}, X_{3t})^\intercal$, where X_{1t} is for the portfolio of small-size firms (first quintile), X_{2t} is for the portfolio of medium-size firms (third quintile), and X_{3t} is for the portfolio of large-size firms (fifth quintile). $\widehat{\mathcal{VR}d}_+(K)$ and $\widehat{Rd}(0)$ of X_t can be estimated based on the method in Section 3.

---

[13]The data are obtained from Kenneth French's Data Library. It was created by CMPT_ME_RETS using the 2013/12 CRSP database. We compute weekly returns of portfolios by linearly adding up Monday to Friday's daily returns.

We first test the absence of serial correlation in each of these three weekly size-sorted portfolio returns. As we stated above, the diagonal elements of $\mathcal{VR}d_+(K)$ correspond to the univariate variance ratio statistics, for example, $[\mathcal{VR}d_+(K)]_{11}$ is the variance ratio of small-size portfolio returns. For each $i = 1, 2, 3$, we test the hypotheses of $H_0 : [\mathcal{VR}d_+(K)]_{ii} = 1$ against $H_1 : [\mathcal{VR}d_+(K)]_{ii} \neq 1$. To compare with the results reported in Campbell, Lo and Mackinlay (1997, P71, Table 2.6), we report $\left[\widehat{\mathcal{VR}d_+}(K)\right]_{ii}$ at $K = 2, 4, 8, 16$ and the corresponding $Zd(K)$, $Zd_{LM}(K)$ and $Zd_{iid}(K)$ statistics[14] in 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). Subsamples are considered to see if there has been changes in variance ratio over time. Table 8-1A reports the results for small-size portfolio, Table 8-1B reports the results for medium-size portfolio, and Table 8-1C reports the results for large-size portfolio.

Table 8-1A: Variance ratios for weekly small-size portfolio returns

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

Table 8-1B: Variance ratios for weekly medium-size portfolio returns

[14]For testing $[\mathcal{VR}d_+(K)]_{ii} = 1$, the $Zd(K)$, $Zd_{LM}(K)$ and $Zd_{iid}(K)$ statistics are calculated by setting e_f as a column vector that is 1 at the $d(i-1) + i$ entry and 0 otherwise.

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

Table 8-1C: Variance ratios for weekly large-size portfolio returns

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

$\left[\widehat{\mathcal{VR}d_+}(K)\right]_{ii}$ for $i = 1, 2, 3$ are reported in the main rows. Test statistics ($Zd(K)$, $Zd_{LM}(K)$ and $Zd_{iid}(K)$) in parentheses marked with asterisks indicate that the variance ratios are statistically different from one at 5% level of significance.

The results for the earlier sample periods are broadly similar to those in Campbell, Lo and Mackinlay (1997, P71, Table 2.6) who compared the period 1962-1978 with the period 1978-1994

as well as the combined period 1962-1994. The variance ratios are greater than one and deviate further from one as the horizon lengthens. The departure from the random walk model is strongly statistically significant for the small and medium sized firms, but not so for the larger firms.

When we turn to the later period 1994-2013 we see that the variance ratios all reduce in magnitude. For the smallest stocks the statistics are still significantly greater than one and increase with horizon. However, they are much closer to one at all horizons and the statistical significance of the departures is substantially reduced. For medium sized firms, the variance ratios are reduced. They are in some cases below one and also no longer increasing with horizon. They are insignificantly different from one. For the largest firms, the ratios are all below one but are statistically inseparable from this value. One interpretation of these results is that the stock market (at the level of these portfolios) has become closer to the 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 one's conclusions, for instance, for large-size portfolio, test statistics based on $\widehat{Q}_{iid}(K)$ in some periods are statistically significant. Our sample size is relatively large, and for smaller samples, the differences could matter a lot more.

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

$$\tau_{AB} = f\left(\widehat{\mathcal{VR}d}_+^A(K)\right) - f\left(\widehat{Rd}^A(0)\right) - f\left(\widehat{\mathcal{VR}d}_+^B(K)\right) + f\left(\widehat{Rd}^B(0)\right), \tag{40}$$

where $\widehat{\mathcal{VR}d}_+^j(K)$ and $\widehat{Rd}^j(0)$ denote the variance ratio statistic and the sample correlation matrix computed in period $j = A, B$, while f is some scalar valued smooth function such as the trace or determinant. Under the martingale null hypothesis (and assuming each subsample is large), the two subsample variance ratio statistics are asymptotically independent and the asymptotic variance of $\sqrt{T}vec(\tau_{AB})$ is just the sum of the subperiod covariance matrices $Qd_f^A(K) + Qd_f^B(K)$. For example, we may consider the single element of statistic $[\widehat{\mathcal{VR}d}_+^A(K)]_{ii} - [\widehat{\mathcal{VR}d}_+^B(K)]_{ii}$ 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 not improved across subperiods. For example, in Table 1-A, the change of the variance ratio for small stocks of 1.43 in the period 78:10:06-94:12:23 to 1.21 during 94:12:30-13:12:27 is statistically significant according to this calculation.

We have carried out this calculation using the Friday to Friday weekly returns as the base series, but we have also done it for other days of the week and for the two parameter statistic. Qualitatively the results are similar. Results are available from the authors upon request.

8.1.2 Lead Lag Relationships

In addition to the autocorrelation for each asset, the predictability can also come from the cross-autocorrelation (lead-lag relationship) between the assets. As we stated earlier, the off-diagonal elements of $\mathcal{VR}d_+(K) - Rd(0)$ provide information about the cumulative cross-dynamics between the assets. We test the hypothesis of $[\mathcal{VR}d_+(K) - Rd(0)]_{ij} = 0$, for $i, j = 1, 2, 3$, $i \neq j$, using the test statistics $Zd(K)$.[15] The results are reported in Table 8-2.[16]

Table 8-2: Lead-lag patterns between weekly size-sorted portfolio returns

---

[15]For testing $[\mathcal{VR}d_+(K) - Rd(0)]_{ij} = 0$, the $Zd(K)$ statistics are calculated by setting e_f as a column vector that is 1 at the $d(j-1) + i$ entry and 0 otherwise.

[16]In this examination, we divide the whole sample into two sub-samples: 62:07:06-94:12:23 and 94:12:30-13:12:27.

Lags	Sample period	From $\widehat{\mathcal{VR}d}_+(K) - \widehat{Rd}(0)$	small	To 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)	

Test statistics in parentheses 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.[17] 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/reduction in microstructure effects of these markets.

We also check if the lead-lag patterns are asymmetric. We test the null hypotheses that $[\mathcal{VR}d_+(K) - Rd(0)]_{ij} - [\mathcal{VR}d_+(K) - Rd(0)]_{ji} = 0$, for $i, j = 1, 2, 3$, $i > j$, using the test statistics $Zd(K)$.[18] The results are reported in Table 8-3.

Table 8-3: Asymmetry of lead-lag patterns

		$\left[\widehat{\mathcal{VR}d_+}(K) - \widehat{Rd}(0)\right]_{ij} - \left[\widehat{\mathcal{VR}d_+}(K) - \widehat{Rd}(0)\right]_{ji}$		
Lags	Sample period	$(S \to M) - (M \to S)$	$(S \to L) - (L \to S)$	$(M \to L) - (L \to M)$
$K = 2$	62:07:06—94:12:23	$-0.19\ (-8.75)^*$	$-0.28\ (-8.58)^*$	$-0.16\ (-8.10)^*$
	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 for small-size portfolio, M is for medium-size portfolio, and L is for large-size portfolio. 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.

---

[17]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.

[18]For testing $[\mathcal{VR}d_+(K) - Rd(0)]_{ij} - [\mathcal{VR}d_+(K) - Rd(0)]_{ji} = 0$, the $Zd(K)$ statistics are calculated by setting e_f as a column vector that is 1 at the $d(j-1) + i$ entry, -1 at the $d(i-1) + j$ entry and 0 otherwise.

8.1.3 Multivariate Tests

The above univariate variance ratio tests (Table 8-1A,B,C) provide evidence of linear predictability in returns for small and medium-size portfolios. We next test for the absence of serial correlation in the whole return vector of three size-sorted portfolios, based on univariate parameters derived from the variance ratio matrices $\mathcal{VR}(K)$ and $\mathcal{VR}d(K)$ of X_t. Specifically, we consider the trace and determinant of these matrices, as well as $CS(K)$, $GMV(K)$, and $\pi_+(K)$. Test results based on these statistics are reported in the following table.

Table 8-4: Multivariate variance ratio tests for weekly size-sorted portfolio returns

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

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

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

The estimates of statistics are reported in the main rows. Test statistics $[Zd_{CS}(K), Z_{GMV}(K), Zd_{\pi}(K), Z_{tr}(K), Z_{\det}(K)$ as defined in (26-27)] in parentheses marked with asterisks indicate statistically significant at 5% level. $W_F(K)$ [defined in (28)] is marked with asterisks if it is larger than 12.592, the 5% critical value of $\chi^2(6)$.

There are some differences of opinion between the measures in the most recent period. Specifically, the momentum profit measure is statistically significant at all horizons, and the trace statistic is significant at horizons $K = 4, 8$, and 16, while the other univariate quantities such as the determinant are not significantly different from their null values. In most cases, the univariate statistics are above their predicted values consistent with the earlier results. Although the momentum profit measure is significant in all three periods, the magnitude of the parameter has reduced substantially. The joint test of all the restrictions is strongly significant in all three periods and for all horizons.

We next check whether our results are driven by the choice of subsamples, which we have chosen to match the choices made by CLM for the purpose of replication and comparison. We carry out a rolling window analysis with a (trailing) window of 500 weeks from the beginning of the sample to the end. Below we show the time series of (standard normal) test statistics $Zd_{CS}(K), Z_{GMV}(K)$ and $Zd_{\pi}(K)$ for $K = 4$. This shows that for $\widehat{GMV}(K)$ and $\widehat{CS}(K)$ the sustained decline in statistical significance happened in the decade ending in 2008, although there was an earlier dip in significance in the decade ending in 1999. The profits measure $\widehat{\pi}(K)$ has shown a slower but equally sustained drop in statistical significance. There are some sudden jumps (both up and down) to the level of this statistic in particular, which may be a cause for concern in practice. The $\widehat{GMV}(K)$ statistic seems less affected by such movements.
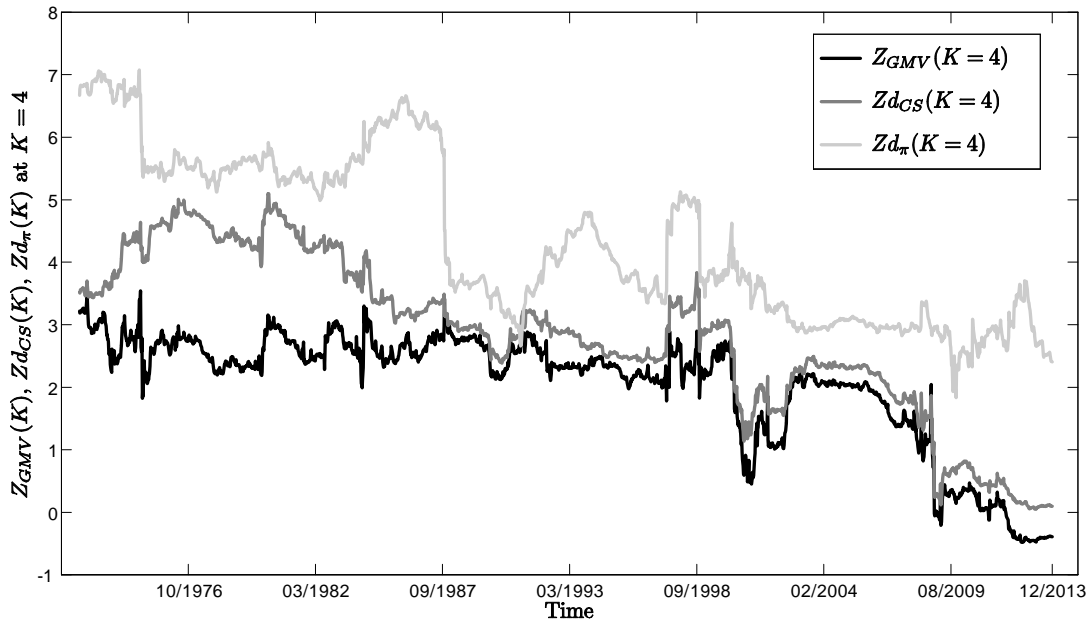
Figure 1: Trends of test statistics based on 10 year rolling windows.

8.2 Long Horizon

We further investigate the variance ratios at longer horizons. We still use the weekly returns for three size-sorted CRSP portfolios (first, third and fifth quintiles). Here, we work with the bias-corrected estimators

$$\widehat{\mathcal{VR}}^{bc}(K) = \widehat{\mathcal{VR}}(K)\left\{1 + \frac{K-1}{T}\right\} \quad ; \quad \widehat{\mathcal{VRd}}^{bc}(K) = \widehat{\mathcal{VRd}}(K)\left\{1 + \frac{K-1}{T}\right\}. \tag{41}$$

The ordered eigenvalues may suffer an even larger bias under the null hypothesis, because their limiting distribution is not centred at zero, and so we propose to modify the estimated eigenvalues by

$$\widehat{\lambda}_j^{bc}(K) = \widehat{\lambda}_j\left(\widehat{\mathcal{VR}}^{bc}(K)\right) - \frac{1}{\sqrt{T}}E\lambda_j^*(U_K),$$

where U_K is the limiting matrix distribution of $\sqrt{T}(\widehat{\mathcal{VR}}(K) - I)$. We calculate $E\lambda_j^*(U_K)$ by simulation.

First, we show below the three eigenvalues $\widehat{\lambda}_j^{bc}(K)$ of $\widehat{\mathcal{VR}}^{bc}(K)$ against K for the three sub-samples: the first panel is for $\widehat{\lambda}_j^{bc}(K)$ in the first sub-sample (62:07:06-78:09:29), the second panel shows $\widehat{\lambda}_j^{bc}(K)$ in the second sub-sample (78:10:06-94:12:23) and the third panel shows $\widehat{\lambda}_j^{bc}(K)$ in the third sub-sample (94:12:30-13:12:27). We also use the dashed lines to indicate the 95% pointwise confidence intervals of the largest eigenvalues for each period centred at the null hypothesis. We

39

show out to two years (100 lags), which is quite a long horizon relative to the sample size, and we urge caution in interpreting the results.
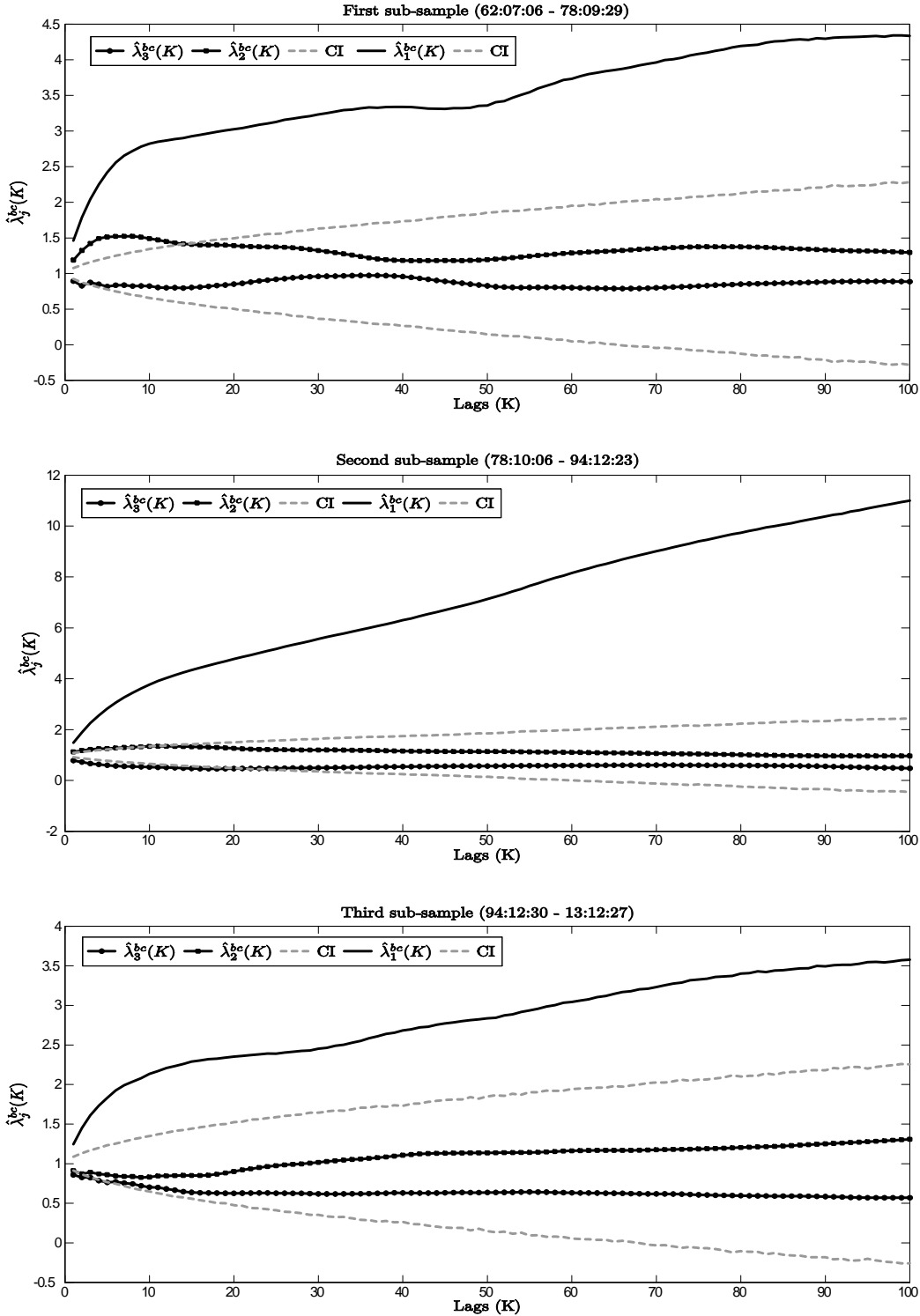

Figure 2: The bias corrected eigenvalues of the bias corrected variance ratio matrix in three sub-samples as a function of lags.

We see that the largest eigenvalue increases steadily out to the two year horizon we consider in all three subperiods. In fact, the increase appears to be linear in lag, although the slope is far less than one. The last subperiod has the lowest values throughout, while surprisingly, the second period 1978-1994 seems to have the largest amount of potential linear predictability that could have been exploited during this period. In all cases, the eigenvalues are statistically significant. The apparent increase in predictability at long horizons that this indicates is consistent with the results discussed in the predictive regression literature, see Phillips (2015), in which $X_t(K)$ is regressed on covariates such as (annual) dividend price ratio dated $t - 1$. The second and third eigenvalues are quite flat and close to one throughout. This evidence does not seem to be consistent with the fads model, or even the bubble process, although the confidence intervals are quite wide at the longer lags.

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

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

These statistics measure in some average sense the cross dependence for certain directions. We show below the $\widehat{CS}_{+}(K)$ and $\widehat{CS}_{-}(K)$ statistics for three weekly size-sorted CRSP portfolio returns against lag K in three sub-samples: the dark solid line is for $\widehat{CS}_{+}(K)$ in the first sub-sample (62:07:06-78:09:29), the dark dashed line is for $\widehat{CS}_{+}(K)$ in the second sub-sample (78:10:06-94:12:23), the dark marked line is for $\widehat{CS}_{+}(K)$ in the third sub-sample (94:12:30-13:12:27); the gray solid line is for $\widehat{CS}_{-}(K)$ in the first sub-sample, the gray dashed line is for $\widehat{CS}_{-}(K)$ in the second sub-sample, and the gray marked line is for $\widehat{CS}_{-}(K)$ in the third sub-sample.
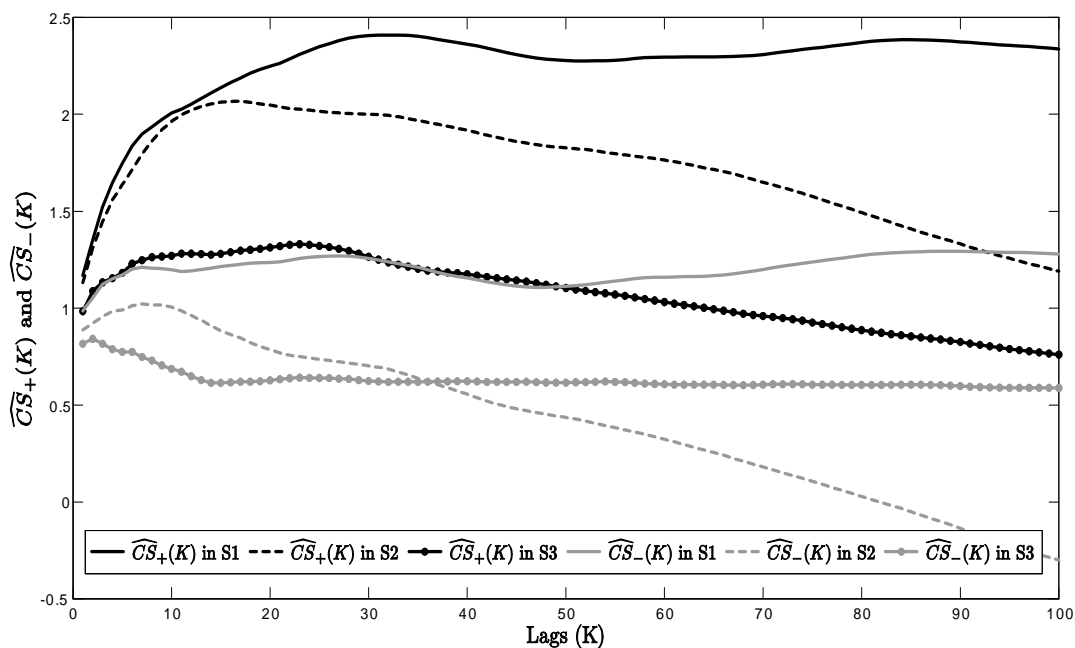
Figure 3: $\widehat{CS}_+(K)$ and $\widehat{CS}_-(K)$ statistics in three sub-samples as a function of lags.

In each subperiod, the $\widehat{CS}_+(K)$ measures all exceed the $\widehat{CS}_-(K)$ 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 $\widehat{CS}_+(K)$ 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, $\widehat{CS}(K) = \left[\widehat{CS}_+(K) + \widehat{CS}_-(K)\right]/2$, measuring the average cross dependence for both directions between three size-sorted CRSP portfolios, becomes weaker (more efficient) in recent periods at the long horizon.

We then examine the long run $GMV(K)$ statistics. We show below $\widehat{GMV}$ against K in the three sub-samples: the solid line is for $\widehat{GMV}(K)$ in the first sub-sample (62:07:06-78:09:29) and the dashed line is for $\widehat{GMV}(K)$ in the second sub-sample (78:10:06-94:12:23), and the marked line is for $\widehat{GMV}(K)$ in the third sub-sample (94:12:30-13:12:27). For readability we have omitted the confidence intervals, which are quite wide in this case and show that mostly this statistic is consistent with the null hypothesis in the most recent period. In this most recent period there is a quite steep fall off in the statistic out to about 3 months followed by a slower rate of decrease thereafter.
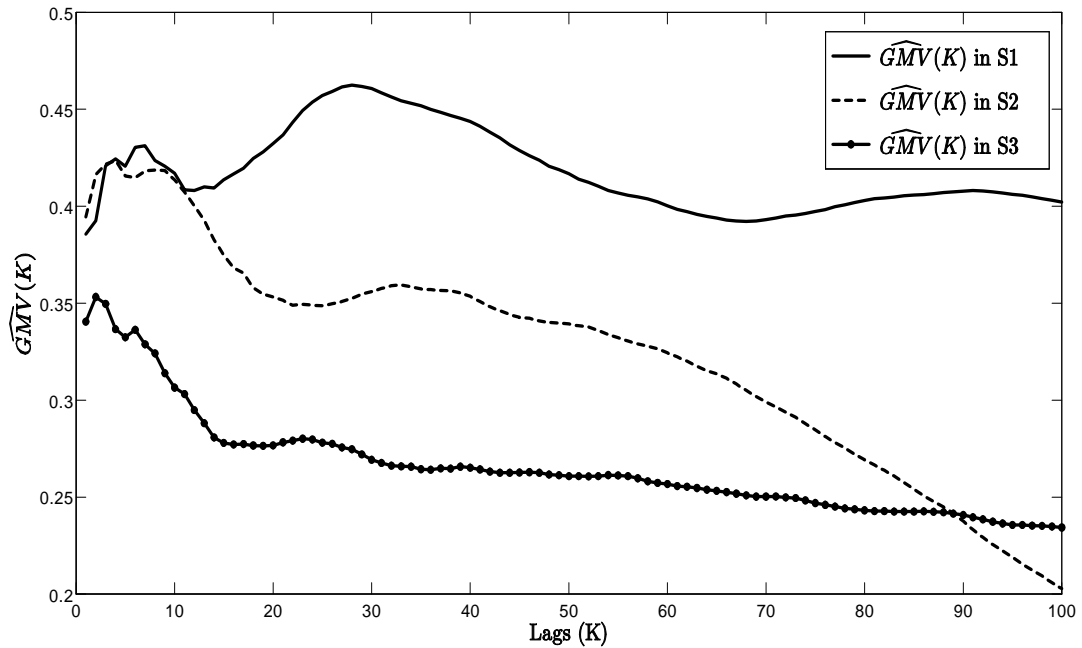
Figure 4: $\widehat{GMV}(K)$ statistics in three sub-samples as a function of lags.

We lastly investigate the $\pi(K)$ statistics. We show below π_{+} against K in three sub-samples: the solid line is for $\widehat{\pi}(K)$ in the first sub-sample (62:07:06-78:09:29) and the dashed line is for $\widehat{\pi}(K)$ in the second sub-sample (78:10:06-94:12:23), and the marked line is for $\widehat{\pi}(K)$ in the third sub-sample (94:12:30-13:12:27). Figure 5 shows that the profit measures $\widehat{\pi}(K)$ are positive across all horizons and subsamples we considered (and are also statistically significant for much of the time). We also see that the measures decrease with the lags till around 40 weeks, and then keep at a relatively stable level. In addition, we found the second period 1978-1994 has the largest measures that could have been exploited during this period.
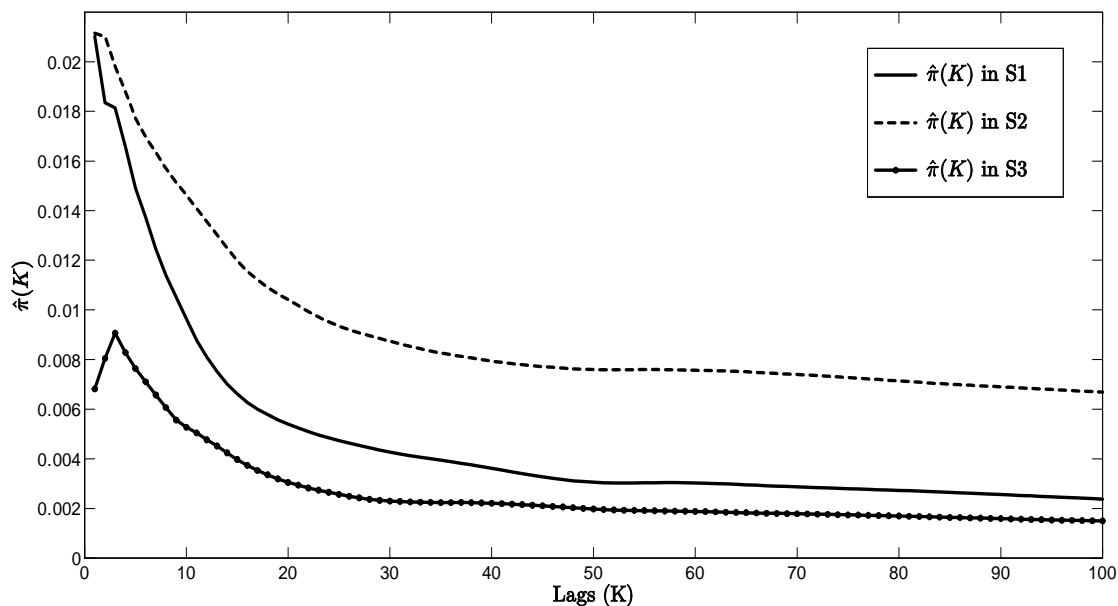
Figure 5: $\widehat{\pi}(K)$ statistics in three sub-samples as a function of lags.

9   Conclusions

The first methodological point we make is to propose confidence intervals that are consistent under uncorrelatedness conditions alone and do not require an additional no leverage/symmetric distribution assumption such as maintained in Lo and MacKinlay (1988), CLM, and in much subsequent work. Our confidence intervals are often (although not necessarily so) larger than those used elsewhere, and therefore reduce the significance of any associated test. We believe our theory is more credible with regard to the data generating process we expect for daily or even lower frequency stock returns. The second contribution is about embedding this theory in a multivariate framework. The multivariate variance ratios provides a basis for aggregating the cross correlation behaviour of asset returns and providing tests of the multivariate null hypothesis. It implies many more restrictions on the data than the univariate ratios. We present our theory for a single K and for sequences of K growing. One can also present result for the joint distribution of our test statistics over different horizons, which would provide some control against multiple testing. However, in practice, it is common to consider just a few horizons that have a specific practical meaning, and so there is no real danger of K-snooping here, although this does again add caution to prevent over interpretation.

Our empirical work reports that the US size sorted stock portfolios seem to have come closer to the efficient markets prediction, although there remains some statistically significant linear predictability at the 2 weeks to 16 weeks horizon. Although many of the individual variance ratio statistics do

44

not reject the null hypothesis with our standard errors, the joint tests of the multivariate hypothesis reject at the 1% level in all cases, meaning for all horizons. This is despite the fact that our standard errors are always larger than those of Lo and MacKinlay (1988), which are themselves a lot larger than those based on the iid assumption, which is the world where most applied studies still inhabit.

Typically, three competing explanations are advanced for the predictability in short horizon returns based on past prices (Boudoukh, Richardson, and Whitelaw (1994)): First, microstructure effects such as nonsynchronous trading and bid ask bounce. Second, time varying risk premia reflecting rational behaviour. Third, the irrational behaviour of market participants. It would seem that there is a lot of evidence that microstructure effects have reduced considerably over time. For example, it is hard to find even small cap stocks that do not trade now many times during a day. The microstructure explanation would imply that the long horizon daily or weekly variance ratios should return to unity, but this is not the case in our data even for the most recent period. There is also some evidence that the level (and perhaps therefore the local time variation) of the market risk premium has reduced in recent years, see for example Hertzberg (2010). In the working paper version of this paper we provided a test of whether the autocorrelations could be explained by time varying risk premia inside a Fama French factor model. We found that this approach could not capture all the linear dependency in the data even in the earlier periods, where the violations were strongest. Therefore, the first two explanations do not seem to be able to match the magnitude of the effects in the earlier periods, although both may make some contribution.[19] On the other hand, the magnitude of the predictability has reduced in the most recent period according to the statistical metrics we have presented here. The long horizon analysis suggests that the largest eigenvalue of the variance ratio matrix grows linearly with horizon, although the slope is far less than the unit slope predicted by the bubble process of section 5.3, which may in principle be consistent with very short bubbly episodes dominated by longer calmer periods.[20] Alternatively, this may be consistent with a very persistent time varying risk premium of the sort outlined in section 5.4. In any case,

---

[19]There is a literature that provides bounds on the implied magnitude of autocorrelations caused by specific microstructure imperfections such as nontrading and a similar literature that provides bounds on the implied magnitude of autocorrelations caused by rational time varying risk premia alone. In both cases strong assumptions are made, see for example Kirby (1998) and Anderson (2011), and Boudoukh, Richardson and Whitelaw (1994), and one "cause" is investigated at a time.

[20]Timmerman (2008) investigates the forecasting performance of a number of linear and nonlinear models and says: "Most of the time the forecasting models perform rather poorly, but there is evidence of relatively short-lived periods with modest return predictability. The short duration of the episodes where return predictability appears to be present and the relatively weak degree of predictability even during such periods makes predicting returns an extraordinarily challenging task". Our (multivariate) evidence does not substantially contradict that; certainly using linear multivariate methods the amount of predictability we have found and its durability is limited and has reduced over time even through the recent financial crisis.

the trajectory is flatter (and not statistically significant) in the more recent period, again supporting the claim that market inefficiency has reduced. Although the statistical magnitudes seem to have reduced, it is not clear whether the potential profit from exploiting linear predictability across the whole market has reduced, since the number of tradeable assets has increased and the transactions costs associated with any given trade seem to have reduced, Malkiel (2015).

10 Appendix

10.1 Proof of Main Results

PROOF OF THEOREM 1. We first present the proof under Assumption A. For each $j = 1, \ldots, K$,

$$
\sqrt{T} \cdot \text{vec}\big(\widehat{\Gamma}(j)\big) = \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \left(\widetilde{X}_{t-j} \otimes \widetilde{X}_t \right) - \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \widetilde{X}_{t-j} \otimes \left(\overline{X} - \mu \right)
$$

$$
- \left(\overline{X} - \mu \right) \otimes \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \widetilde{X}_t + \frac{T-j}{\sqrt{T}} \left(\overline{X} - \mu \right) \otimes \left(\overline{X} - \mu \right) \tag{42}
$$

$$
= \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \left(\widetilde{X}_{t-j} \otimes \widetilde{X}_t \right) + o_p(1) \tag{43}
$$

because $\sum_{t=j+1}^{T} \widetilde{X}_t = O_p(\sqrt{T})$ by the CLT for stationary ergodic martingale difference.

Since $\widetilde{X}_t \widetilde{X}_t^{\intercal}$ is stationary ergodic, the Ergodic theorem and continuous mapping on $T^{-1} \sum_{t=1}^{T} \widetilde{X}_t \widetilde{X}_t^{\intercal}$ yields $\widehat{\Sigma}^{-1/2} - \Sigma^{-1/2} = o_p(1)$. Consequently, we have

$$
\text{vec}(\widehat{R}(j)) = \text{vec}\left(\left[\widehat{\Sigma}^{-1/2} - \Sigma^{-1/2} + \Sigma^{-1/2} \right] \widehat{\Gamma}(j) \left[\widehat{\Sigma}^{-1/2} - \Sigma^{-1/2} + \Sigma^{-1/2} \right] \right)
$$

$$
= \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \text{vec}(\widehat{\Gamma}(j)) + o_p(1) \tag{44}
$$

for each j, and therefore

$$
\sqrt{T} \text{vec}\left(\widehat{\mathcal{VR}}_+(K) - I_d \right) = \sqrt{T} \cdot \sum_{j=1}^{K-1} 2 \left(1 - \frac{j}{K} \right) \cdot \text{vec}\left(\widehat{R}(j) \right)
$$

$$
= \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \cdot \sum_{j=1}^{K-1} c_j \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \widetilde{X}_{t-j} \otimes \widetilde{X}_t + o_p(1)
$$

$$
= \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \cdot \frac{1}{\sqrt{T}} \sum_{t=K}^{T} \left[\sum_{j=1}^{K-1} c_j \left(\widetilde{X}_{t-j} \otimes \widetilde{X}_t \right) \right] + o_p(1)
$$

$$
=: \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \cdot \frac{1}{\sqrt{T}} \sum_{t=K}^{T} Z_t + o_p(1). \tag{45}
$$

It now suffices to derive the limiting distribution of Z_t. Take any d^2-dimensional real constant vector $a = (a_1, \ldots, a_{d^2})^\intercal$, and note that $a^\intercal Z_t$ is a martingale difference sequence. Then, since by A2

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

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

$$a^\intercal \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T Z_t \right) \Longrightarrow N \left(0, a^\intercal \left[\sum_{j=1}^{K-1} \sum_{k=1}^{K-1} c_j c_k \Xi_{jk} \right] a \right). \tag{46}$$

completing the proof in view of the Cramér-Wold device, continuous mapping and Slutsky's theorem.

Similar arguments apply when we work with Assumption MH*. We note that the expansion (43) for $\sqrt{T} \cdot \operatorname{vec}(\widehat{\Gamma}(j))$ is still valid because the summations in the second, third and fourth terms in (42) still converge in probability to zero due to the CLT for mixing sequence, Herrndorf (1985, Theorem 0) whose regularity conditions are satisfied by MH1-MH3. Finally, condition MH2 and MH3 allow for the LLN for mixing variables, White (1984, Corollary 3.48), yielding (44) and (45) as before.

Now we are only left with verifying (46). For any d^2-dimensional constant vector a, $a^\intercal Z_t$ preserves the mixing property of $\widetilde{X}_t$ with the same rate, so by Herrndorf's CLT we have

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

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

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

is positive and finite. The first condition is trivial by MH1, and the second and third conditions are satisfied by MH2, MH3 and positive definiteness of $Q(K)$. The desired result readily follows. The arguments for the diagonally normalized is identical everywhere except that we have $\operatorname{vec}(\widehat{R}d(j)) = \left(D^{-1/2} \otimes D^{-1/2} \right) \operatorname{vec}(\widehat{\Gamma}(j)) + o_p(T^{-1/2})$ instead of (44). The entire proof is now complete. ∎

PROOF OF COROLLARY 1. It suffices to show consistency of $\widehat{\Xi}_{jk}$ for each j and k. Writing

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

we see that the desired result follows by applying either the Ergodic theorem or the Law of Large Numbers for mixing variables depending upon the set of assumption being imposed. The regularity conditions for each theorem are ensured by Assumption A2 and MH3, respectively. ∎

PROOF OF THEOREM 2. For later reference, we first derive the limiting distribution of

$$\frac{1}{\sqrt{T}}\sum_{t=K}^{T}\left[\frac{1}{\sqrt{K}}\sum_{j=1}^{K-1}c_j(\widetilde{X}_{t-j}\otimes\widetilde{X}_t)\right]=:\frac{1}{\sqrt{T}}\sum_{t=K}^{T}Z_{Tt}. \tag{47}$$

The asymptotic normality is established by applying the central limit theorem for triangular arrays of martingale difference in Pollard (1984, page 171) on $\{Z_{Tt},\mathcal{F}_{Tt}\}$, where $\mathcal{F}_{Tt}=\mathcal{F}_t=\sigma(\widetilde{X}_s;\ s\le t)$. Specifically, for some arbitrary non-zero constant vector $a=(a_1,a_2,\ldots,a_{d^2})^{\intercal}$ we check the following conditions:

(i) $T^{-1}\sum_t E((a^{\intercal}Z_{Tt})^2|\mathcal{F}_{t-1})\xrightarrow{P} a^{\intercal}\eta a>0$

(ii) $\forall\varepsilon>0,\ T^{-1}\sum_t E((a^{\intercal}Z_{Tt})^2 1\{|a^{\intercal}Z_{Tt}|>\varepsilon\sqrt{T}\}|\mathcal{F}_{t-1})\xrightarrow{P} 0$

under which (via the Cramér-Wold Theorem) it will follow that

$$\frac{1}{\sqrt{T}}\sum_{t=K}^{T}Z_{Tt}\Longrightarrow N\big(0,\eta\big). \tag{48}$$

As for the first condition (i), it suffices to show

$$\frac{1}{T}\sum_{t=K}^{T}E(a^{\intercal}Z_{Tt})^2\to a^{\intercal}\eta a \tag{49}$$

$$\frac{1}{T}\sum_{t=K}^{T}E\big((a^{\intercal}Z_{Tt})^2|\mathcal{F}_{t-1}\big)-E(a^{\intercal}Z_{Tt})^2\xrightarrow{P}0. \tag{50}$$

Denoting by $\widetilde{X}_{it}$ the ith element of the d-vector $\widetilde{X}_t$, the moment-cumulant relationship formula (see e.g. Hannan (1970, page 23)) suggests that for any $q,w,e,u=1,\ldots,d$

$$\frac{1}{K}\sum_{j=1}^{K-1}\sum_{r=1}^{K-1}c_j c_r E\big(\widetilde{X}_{q,t-j}\widetilde{X}_{w,t-r}\widetilde{X}_{et}\widetilde{X}_{ut}\big)$$

$$=\frac{1}{K}\sum_{j=1}^{K-1}\sum_{r=1}^{K-1}c_j c_r\Big[\gamma_{qw}(j-r)\gamma_{eu}(0)+\gamma_{qe}(j)\gamma_{wu}(r)+\gamma_{qu}(j)\gamma_{we}(r)+\kappa_{qweu}(-j,-r,0,0)\Big]$$

$$=\gamma_{qw}(0)\gamma_{eu}(0)\frac{1}{K}\sum_{p=1}^{K-1}4\Big(1-\frac{p}{K}\Big)^2+\frac{1}{K}\sum_{j=1}^{K-1}\sum_{r=1}^{K-1}\kappa_{qweu}(-j,-r,0,0)\ \to\ \frac{4}{3}\cdot\gamma_{qw}(0)\gamma_{eu}(0)$$

where $\gamma_{qw}(u) = E(\widetilde{X}_{qt}\widetilde{X}_{w,t-u})$. This is due to uncorrelatedness of $\widetilde{X}_t$ and double summability of the fourth order cumulant κ (Assumption S). Consequently,

$$
\begin{aligned}
\frac{1}{T}\sum_{t=K}^{T} E\big[(a^\intercal Z_{Tt})\big]^2 &= \frac{1}{T}\sum_{t=K}^{T} E\left[\frac{1}{\sqrt{K}}\sum_{j=1}^{K-1} a^\intercal c_j(\widetilde{X}_{t-j}\otimes\widetilde{X}_t)\right]^2 \\
&= \left(\frac{T-K}{T}\right) a^\intercal\left[\frac{1}{K}\sum_{j=1}^{K-1}\sum_{r=1}^{K-1} c_j c_r E\big(\widetilde{X}_{t-j}\widetilde{X}_{t-r}^\intercal\otimes\widetilde{X}_t\widetilde{X}_t^\intercal\big)\right] a \\
&= \frac{4}{3}a^\intercal(\Sigma\otimes\Sigma)a + O\left(\frac{K}{T}\right) \to a^\intercal\eta a
\end{aligned}
$$

which is strictly positive by the assumption that Σ is positive definite. This finite limit also implies (50) as a consequence of the ergodic theorem in view of the inherited stationary ergodicity through a measurable mapping, see for example Karlin and Taylor (1975, page 487-488).

It remains to check the conditional Lindeberg condition (ii). By stationarity, law of total expectations, the moment condition $E|\widetilde{X}_t|^{4+\delta} \le C$ (where δ is as in Assumption A'), Minkowski's inequality, and Burkholder's inequality for martingale difference (e.g. Gut (2005, page 506-507)) we have for any $\epsilon > 0$

$$
\begin{aligned}
P\left(\left|\frac{1}{T}\sum_{t=1}^{T} E\big((a^\intercal Z_{Tt})^2 1_{\{|a^\intercal Z_{Tt}|>\varepsilon\sqrt{T}\}}|\mathcal{F}_{t-1}\big)\right| > \epsilon\right) &\le \frac{1}{\epsilon}E\left|\frac{1}{T}\sum_{t=1}^{T} E\big((a^\intercal Z_{Tt})^2 1_{\{|a^\intercal Z_{Tt}|>\varepsilon\sqrt{T}\}}|\mathcal{F}_{t-1}\big)\right| \\
&\le \frac{1}{\epsilon}\left(\frac{1}{\varepsilon\sqrt{T}}\right)^{\frac{\delta}{2}} E\big((a^\intercal Z_{Tt})^{2+\frac{\delta}{2}}\big) \le \frac{1}{\epsilon}\left(\frac{1}{\varepsilon\sqrt{T}}\right)^{\frac{\delta}{2}} E\left(\sum_{i=1}^{d^2}\frac{1}{\sqrt{K}}\sum_{j=1}^{K-1} c_j a_i\widetilde{X}_{\lceil i/d\rceil,0}\widetilde{X}_{i-d(\lceil i/d\rceil-1),j}\right)^{2+\frac{\delta}{2}} \\
&\le \frac{1}{\epsilon}\left(\frac{1}{\varepsilon\sqrt{T}}\right)^{\frac{\delta}{2}}\left(\frac{1}{\sqrt{K}}\sum_{i=1}^{d^2}|a_i|\left[E\left|\sum_{j=1}^{K-1}\widetilde{X}_{i-d(\lceil i/d\rceil-1),j}\widetilde{X}_{\lceil i/d\rceil,0}\right|^{2+\delta/2}\right]^{1/(2+\delta/2)}\right)^{2+\frac{\delta}{2}} \\
&= \frac{C}{T^{\delta/4}}\left(\frac{1}{\sqrt{K}}\right)^{2+\delta/2}\max\{K^{1+\delta/4},K\} = O\left(\frac{1}{T^{\delta/4}}\right)
\end{aligned}
$$

where $\lceil\cdot\rceil$ is the ceiling function.

Now, starting from the decomposition (42) in the proof of Theorem 1, we can easily see using

Chebyshev's inequality, some results from Theorem 1 and elementary calculations that

$$
\sqrt{\frac{T}{K}}\mathrm{vec}\left(\widehat{\mathcal{VR}}(K)_+ - I_d\right) = \frac{\sqrt{T}}{\sqrt{K}} \cdot \sum_{j=1}^{K-1} 2\left(1 - \frac{j}{K}\right) \cdot \mathrm{vec}\left(\widehat{R}(j)\right)
$$

$$
= (\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2})\frac{1}{\sqrt{K}}\sum_{j=1}^{K-1} c_j \Bigg\{ \left[\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T} \widetilde{X}_{t-j} \otimes \widetilde{X}_t\right]
$$

$$
- \left[\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_{t-j} \otimes \left(\overline{X} - \mu\right)\right] - \left[\left(\overline{X} - \mu\right) \otimes \frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_t\right]
$$

$$
+ \left[\frac{T-j}{\sqrt{T}}\left(\overline{X} - \mu\right) \otimes \left(\overline{X} - \mu\right)\right]\Bigg\}
$$

$$
= (\widehat{\Sigma}^{-1/2} \otimes \widehat{\Sigma}^{-1/2})\frac{1}{\sqrt{T}}\sum_{t=K}^{T}\left[\frac{1}{\sqrt{K}}\sum_{j=1}^{K-1}c_j(\widetilde{X}_{t-j} \otimes \widetilde{X}_t)\right] + O_P\left(\sqrt{\frac{K}{T}}\right) + o_p\left(\sqrt{\frac{K}{T}}\right)
$$

$$
= (\Sigma^{-1/2} \otimes \Sigma^{-1/2})\frac{1}{\sqrt{T}}\sum_{t=K}^{T}Z_{Tt} + o_P(1) \tag{51}
$$

Now the desired asymptotic distribution holds in view of the results above and consistency of standard error via the ergodic theorem, completing the proof. ∎

PROOF OF THEOREM 3. The proof proceeds by showing asymptotic equivalence of the trace (of the multivariate variance ratio) test and the likelihood ratio (LR) test under the null and alternative hypotheses. That is,

$$
f\left(\mathrm{tr}\left(\widehat{\mathcal{VR}}(K)\right)\right) - LR \xrightarrow{P} 0 \tag{52}
$$

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

Recall the alternative estimator $\widehat{\mathcal{VR}}^{\&}(K)$. From the definitions it can be readily shown that

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

$$
+ \frac{1}{K}\sum_{r=1}^{K-2}\left\{\widehat{\Sigma}^{-1/2}\left[(K-r)\frac{1}{T}\sum_{t=2}^{K-1}\left(X_{t-r} - \overline{X}\right)\left(X_t - \overline{X}\right)^{\mathsf{T}}\right]\widehat{\Sigma}^{-1/2}\right\} + o_p(1) \tag{53}
$$

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

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

Let the 'coefficient matrix' Φ be the matrix of ones except for the $(T - K) \times (T - K)$ triangular blocks in the northeast and southwest corners where the entries are all zero. Then denoting by i a conformable column vector of ones, we have

$$\widehat{\Sigma}(K) = \frac{1}{T} \left(\Phi X - \Phi i \overline{X}^{\mathsf{T}} \right)^{\mathsf{T}} \left(\Phi X - \Phi i \overline{X}^{\mathsf{T}} \right) = \frac{1}{T} \left(X - i \overline{X}^{\mathsf{T}} \right)^{\mathsf{T}} \Phi^{\mathsf{T}} \Phi \left(X - i \overline{X}^{\mathsf{T}} \right)$$

from which it follows that

$$\widehat{\mathcal{VR}}^{\&}(K) = \frac{1}{K} [(A^{\mathsf{T}} A)^{-1/2}] \cdot [A^{\mathsf{T}} \Phi^{\mathsf{T}} \Phi A] \cdot [(A^{\mathsf{T}} A)^{-1/2}]$$

where $A := \left(X - i\overline{X}^{\mathsf{T}} \right)$.

The rejection region based on the likelihood ratio statistic is given by

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

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

$$
\begin{aligned}
LR &= \log \left(\det \left\{ \left[\left(X - i\overline{X}^{\mathsf{T}} \right)^{\mathsf{T}} \left(X - i\overline{X}^{\mathsf{T}} \right) \right]^{-1} \left[\left(X - i\widetilde{X}^{\mathsf{T}} \right)^{\mathsf{T}} \Sigma_{q*}^{-1} \left(X - i\widetilde{X}^{\mathsf{T}} \right) \right] \right\} \right) \\
&\leq \operatorname{tr} \left(\left[\left(X - i\overline{X}^{\mathsf{T}} \right)^{\mathsf{T}} \left(X - i\overline{X}^{\mathsf{T}} \right) \right]^{-1} \left[\left(X - i\widetilde{X}^{\mathsf{T}} \right)^{\mathsf{T}} \Sigma_{q*}^{-1} \left(X - i\widetilde{X}^{\mathsf{T}} \right) \right] - I \right) \\
&\leq \operatorname{tr} \left(\widehat{\Sigma}^{-1} \cdot \frac{1}{T} \left[\left(X - i\widetilde{X}^{\mathsf{T}} \right)^{\mathsf{T}} \Sigma_{q*}^{-1} \left(X - i\widetilde{X}^{\mathsf{T}} \right) \right] \right).
\end{aligned}
\tag{54}
$$

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

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

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

$$
\begin{aligned}
&\operatorname{tr} \left(\widehat{\Sigma}^{-1} \cdot \frac{1}{T} \left[\left(X - i\widetilde{X}^{\mathsf{T}} + i \left(\widetilde{X}^{\mathsf{T}} - \overline{X}^{\mathsf{T}} \right) \right)^{\mathsf{T}} \cdot \{ \alpha(I + q\Phi^{\mathsf{T}} \Phi) \} \cdot \left(X - i\widetilde{X}^{\mathsf{T}} + i \left(\widetilde{X}^{\mathsf{T}} - \overline{X}^{\mathsf{T}} \right) \right) \right] \right) \\
&= \operatorname{tr} \left(\widehat{\Sigma}^{-1} \cdot \frac{1}{T} \left[\left(X - i\widetilde{X}^{\mathsf{T}} + i \left(\widetilde{X}^{\mathsf{T}} - \overline{X}^{\mathsf{T}} \right) \right)^{\mathsf{T}} \{ \Sigma_q^{-1} + 0^* \} \left(X - i\widetilde{X}^{\mathsf{T}} + i \left(\widetilde{X}^{\mathsf{T}} - \overline{X}^{\mathsf{T}} \right) \right) \right] \right)
\end{aligned}
\tag{55}
$$

where 0^* is the matrix of zeros except for the $(K-1)\times(K-1)$ blocks in the northwest and southeast corners. The reader is directed to Faust (1992, Lemma 1) for the proof of the equivalence relationship $\alpha(I + q\Phi^\intercal\Phi) \equiv \Sigma_q^{-1} + 0^*$. Now replacing the sample estimator for the cross-sectional variance by its population version (with some negligible error), we see that the difference between (55) and (54) multiplied by $\sqrt{T}$ is given by

$$\sqrt{T}\cdot\text{tr}\left(\Sigma^{-1}\cdot\frac{1}{T}\left[\left(i\left(\widetilde{X}^\intercal - \overline{X}^\intercal\right)\right)^\intercal\cdot\Sigma_q^{-1}\cdot\left(i\left(\widetilde{X}^\intercal - \overline{X}^\intercal\right)\right)\right]\right) + o_p(1)$$

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

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

$$\sqrt{T}\left|\alpha\left\{d + qK\cdot\text{tr}\left(\widehat{\mathcal{VR}}^{\&}(K)\right)\right\} - LR\right| \xrightarrow{p} 0. \tag{56}$$

This suggests that there exist some α and q for which the trace test has the same large sample properties of the LR test against the $\phi-$best class alternatives. Since the sequence of the LR tests with $q^* = \delta/\sqrt{T}$ is locally most powerful invariant, e.g. Engle (1984), the proof is complete. ∎

PROOF OF THEOREM 4. Consider the K period returns $X_t(K) = K\mu + p_t - p_{t-K} = \sum_{s=t-K}^t \varepsilon_s + \sum_{s=t-K}^t(\eta_s - \eta_{s-1}) = K\mu + \sum_{s=t-K}^t \varepsilon_s + \eta_t - \eta_{t-K}$. These have variance

$$\Sigma_K = \text{var}(X_t(K)) = \text{var}\left(\sum_{s=t-K}^t \varepsilon_s\right) + \text{var}\left(\eta_t - \eta_{t-K}\right)$$

$$= KE\varepsilon_s\varepsilon_s^\intercal + E\left((\eta_t - \eta_{t-K})(\eta_t - \eta_{t-K})^\intercal\right) = K\Omega_\varepsilon + \Omega_\eta(K),$$

where $\Omega_\eta(k) = \text{var}\left(\eta_t - \eta_{t-k}\right) \geq 0$, $k = 1, 2, \ldots$. Therefore, $\mathcal{VR}(K) = \Sigma_1^{-1/2}\Sigma_K\Sigma_1^{-1/2}/K$ and $\mathcal{VR}d(K) = D_1^{-1/2}\Sigma_K D_1^{-1/2}/K$. Note that as $K \to \infty$, $\Omega_\eta(K) \to 2\Omega_\eta = 2\text{var}\left(\eta_t\right)$. It follows that as $K \to \infty$

$$\mathcal{VR}(K) = K^{-1}\Sigma_1^{-1/2}\Sigma_K\Sigma_1^{-1/2} = K^{-1}\Sigma_1^{-1/2}\left(K\Omega_\varepsilon + \Omega_\eta(K)\right)\Sigma_1^{-1/2}$$

$$\longrightarrow \Sigma_1^{-1/2}\Omega_\varepsilon\Sigma_1^{-1/2} = \Sigma_1^{-1/2}\left[\Sigma_1 - \Omega_\eta(1)\right]\Sigma_1^{-1/2}$$

$$= I - \Sigma_1^{-1/2}\Omega_\eta(1)\Sigma_1^{-1/2} \leq I,$$

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

strictly positive definite. By similar arguments,

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

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

PROOF OF THEOREM 5. 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 (36). For simplicity we suppose that $p_t = \delta_T p_{t-1} + \varepsilon_t$ with ε_t iid with variance σ_ε^2 and $\delta_T = 1 + \frac{c}{k_T}$, where $k_T = T^\alpha, \alpha \in (0, 1/2)$ and some positive constant c. According to Phillips and Magdalinos (2007, Theorem 4.3) we have

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

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

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

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

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

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

Likewise,

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

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

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

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

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

$$\delta_T^{-2T} \sum_{t=1}^{T} X_t(K)^2 \Longrightarrow K^2 c^2 Y^2. \tag{58}$$

In fact, using Cramér-Wold device it can be shown that the convergence in (57) and (58) is joint. Therefore, by the continuous mapping theorem

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

as required. ∎

PROOF OF THEOREM 6. From (37) and (38) it is straightforward to see that

$$X_{t+i} = (t+i)\mu + \mu_0 + \sum_{j=1}^{t+i} \eta_j + \varepsilon_{t+i}$$

$$\mathrm{var}(X_{t+i}) = \Sigma_\varepsilon + \frac{t+i}{T}\Sigma_\eta \quad \text{and} \quad \mathrm{cov}(X_{t+i}, X_{t+k}) = \frac{t+i}{T}\Sigma_\eta$$

for $i, k = 0, 1, \ldots, K-1$ and $i < k$. Consequently we have

$$\mathrm{var}(X_t + X_{t+1} + \cdots + X_{t+K-1}) = K^2 \frac{\Sigma_\eta}{T} \left(t + \sum_{j=1}^{K-1} \left(1 - \frac{j}{K}\right)^2 \right) + K\Sigma_\varepsilon$$

$$\mathcal{VR}_T(K) = \left(\Sigma_\eta \frac{T+1}{2T} + \Sigma_\varepsilon\right)^{-1/2} \left[K\Sigma_\eta \left(\frac{T+1}{2T} + \frac{1}{T}\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)^2\right) + \Sigma_\varepsilon\right] \left(\Sigma_\eta \frac{T+1}{2T} + \Sigma_\varepsilon\right)^{-1/2}$$

so that

$$\frac{1}{K} \lim_{T \to \infty} \mathcal{VR}_T(K) = \left(\Sigma_\eta \frac{1}{2} + \Sigma_\varepsilon\right)^{-1/2} \left[\Sigma_\eta \frac{1}{2} + \frac{1}{K}\Sigma_\varepsilon\right] \left(\Sigma_\eta \frac{1}{2} + \Sigma_\varepsilon\right)^{-1/2}$$

$$\longrightarrow \left(\Sigma_\eta \frac{1}{2} + \Sigma_\varepsilon\right)^{-1/2} \frac{1}{2}\Sigma_\eta \left(\Sigma_\eta \frac{1}{2} + \Sigma_\varepsilon\right)^{-1/2}$$

as $K \to \infty$, completing the proof. ∎

PROOF OF THEOREM 7. It is straightforward to see that

$$
\sqrt{T}\left[\frac{1}{d}\sum_{i=1}^{d}\left(\widehat{\mathcal{VR}d}_{+,ii}(K)-1\right)\right]=\sqrt{T}\cdot\frac{1}{d}\sum_{i=1}^{d}\left(2\sum_{j=1}^{K-1}\left(1-\frac{j}{K}\right)\frac{\widehat{\gamma}_{ii}(j)}{\widehat{\sigma}_{ii}}\right)
$$

$$
=\frac{1}{d}\sum_{i=1}^{d}\sum_{j=1}^{K-1}\frac{c_j}{\sigma_{ii}}\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{X}_{it}\widetilde{X}_{i,t-j}-\left[\max_{1\leq i\leq d}(\overline{X}_i-\mu_i)\right]\cdot\left\{\frac{1}{d}\sum_{i=1}^{d}\sum_{j=1}^{K-1}\frac{c_j}{\sigma_{ii}}\frac{1}{\sqrt{T}}\sum_{t=1+j}^{T}\widetilde{X}_{it}\right\}
$$

$$
-\left[\max_{1\leq i\leq d}(\overline{X}_i-\mu_i)\right]\cdot\left\{\frac{1}{d}\sum_{i=1}^{d}\sum_{j=1}^{K-1}\frac{c_j}{\sigma_{ii}}\frac{1}{\sqrt{T}}\sum_{t=1+j}^{T}\widetilde{X}_{i,t-j}\right\}+\frac{1}{d}\sum_{i=1}^{d}\sum_{j=1}^{K-1}\frac{c_j}{\sigma_{ii}}\frac{T-j-1}{\sqrt{T}}(\overline{X}_i-\mu_i)^2+o_p(1)
$$

$$
=\frac{1}{\sqrt{T}}\sum_{t=K}^{T}\left(\frac{1}{d}\sum_{i=1}^{d}\left[\sum_{j=1}^{K-1}\frac{c_j\widetilde{X}_{it}\widetilde{X}_{i,t-j}}{\sigma_{ii}}\right]\right)+o_p(1) \tag{59}
$$

because by uncorrelatedness of $\widetilde{X}_t$

$$
P\left(\left|\max_{1\leq i\leq d}(\overline{X}_i-\mu_i)\right|>\varepsilon\right)\leq\sum_{i=1}^{d}\frac{1}{\varepsilon^2}E\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{it}\right)^2
$$

$$
\leq\frac{d}{\varepsilon^2}\frac{1}{T^2}\sum_{t=1}^{T}\max_{1\leq i\leq d}E\left(\widetilde{X}_{it}^2\right)=O\left(\frac{d}{T}\right),
$$

and by the moment-cumulant relationship formula and Assumption S we have

$$
P\left(\max_{1\leq i\leq d}\left|\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{it}^2-E\left(\widetilde{X}_{it}^2\right)\right|>\varepsilon\right)\leq\frac{1}{\varepsilon^2}\sum_{i=1}^{d}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\text{Cov}\left(\widetilde{X}_{it}^2,\widetilde{X}_{is}^2\right)
$$

$$
\leq\frac{1}{\varepsilon^2}\sum_{i=1}^{d}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[2\left(\gamma_{ii}(t-s)\right)^2+\kappa_{iiii}(s,s,0,0)\right]=O\left(\frac{d}{T}\right),
$$

from which it follows that $\max_i|\widehat{\sigma}_{ii}-\sigma_{ii}|=o_p(1)$. The $\{.\}$ terms can be easily shown to be bounded in probability using Chebyshev's inequality and uncorrelatedness of $\widetilde{X}_t$.

It now suffices to derive the limiting distribution of (59). We will only briefly sketch the proof as the main arguments closely follow those of proof of Theorem 2. Since the asymptotic variance

$$
qd(\infty)^*=\lim_{d\to\infty}\frac{1}{d^2}\tau^{\mathsf{T}}Qd(K)\tau=\lim_{d\to\infty}\frac{1}{d^2}\sum_{i=1}^{d}\sum_{r=1}^{d}\left[\sum_{j=1}^{K-1}\sum_{k=1}^{K-1}\frac{c_jc_k}{\sigma_{ii}\sigma_{rr}}E(\widetilde{X}_{it}\widetilde{X}_{rt}\widetilde{X}_{i,t-j}\widetilde{X}_{r,t-k})\right], \tag{60}
$$

where $\tau=\text{vec}(I_d)$, is finite by Assumption Sd, we see that upon checking the required conditions the central limit theorem for martingale difference applies, yielding

$$
\frac{1}{\sqrt{T}}\sum_{t=K}^{T}\left(\frac{1}{d}\sum_{i=1}^{d}\left[\sum_{j=1}^{K-1}\frac{c_j\widetilde{X}_{it}\widetilde{X}_{i,t-j}}{\sigma_{ii}}\right]\right)\implies N\left(0,qd(\infty)^*\right).
$$

Note that the conditional Lindeberg condition can be shown to hold by repeatedly using Minkowski's inequality and by Assumption A′. The remaining consistency result follows by the ergodic theorem and Assumption Sd, completing the proof. ∎

References

[1] Anderson, R. M. (2011). Time-varying risk premia. Journal of Mathematical Economics, 47(3), 253-259.

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

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

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

[5] Blanchard, O. J. and Watson, M. W. (1982). Bubbles, rational expectations and financial markets. NBER Working Paper no. 945 available at http://www.nber.org/papers/w0945.

[6] Boudoukh, J., Richardson, M. P. and Whitelaw, R. E. (1994). A tale of three schools: Insights on autocorrelations of short-horizon stock returns. Review of Financial Studies, 7(3), 539-573.

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

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

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

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

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

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

[13] Cho, J. S. and White, H. (2014). Testing the Equality of Two Positive-Definite Matrices with Application to Information Matrix Testing. Advances in Econometrics: Essays in Honor of Peter C. B. Phillips. Vol. 33. Eds. Yoosoon Chang, Thomas B. Fomby, and Joon Y. Park. West Yorkshire, UK: Emerald Group Publishing Limited, forthcoming.

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

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

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

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

[18] Dufour, J.-M. and Roy, R. (1985). Some robust exact results on sample autocorrelations and test of randomness. Journal of Econometrics, 29(3), 257-273.

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

[20] Engle, R. F. (1984). Wald, Likelihood ratio, and Lagrange multiplier tests in Econometrics. in Handbook of Econometrics, Vol. II, ed. Z. Amsterdam: North Holland.

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

[22] Escanciano, J. C. and Lobato, I. N. (2009). Testing the martingale hypothesis. in Palgrave Handbook of Econometrics. Palgrave Macmillan UK.

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

[24] Fama, E. (1970). Effient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance 25(2), 383-417.

[25] Fama, E. (2013). Two pillars of asset pricing. Nobel prize lecture.

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

[27] Francq, C., Roy, R. and Zakoian, J,-M. (2005). Diagnostic Checking in ARMA Models with Uncorrelated Errors. Journal of the Americal Statistical Association, 100(470), 532-544.

[28] Grossman, S. and Stiglitz, J. (1980). On the Impossibility of Informationally Efficient Markets. American Economic Review, 70(3), 393-408.

[29] Gut, A. (2005). Probability: A Graduate Course. New York: Springer.

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

[31] Hannan, E. J. (1970). Multiple Time Series. New York: John Wiley and Sons.

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

[33] Hasbrouck, J. (2015). High Frequency Quoting: Short-Term Volatility in Bids and Offers. Working paper available at SSRN: http://ssrn.com/abstract=2237499. or http://dx.doi.org/10.2139/ssrn.2237499.

[34] 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.

[35] Hertzberg, D. (2010). The Great Stock Myth: Why the market's rate of return— and your nest egg—may never recover. The Atlantic. Working paper available at http://www.theatlantic.com/magazine/archive/2010/09/the-great-stock-myth/308178/.

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

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

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

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

[40] Jin, B., Wang, C., Bai, Z. D., Krishnan, N. K. and Harding, M. (2014) Limiting spectral distribution of a symmetrized auto-cross covariance matrix. Annals of Applied Probability, 24(3), 1199-1225.

[41] Johnstone, I.M. and A. Onatski (2015). Testing in high-dimensional spiked models. Available at https://arxiv.org/abs/1509.07269.

[42] Kahneman, D. and Tversky, A. (2000). Choices, Values and Frames. Cambridge: Cambridge University Press.

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

[44] Karlin, S. and Taylor, H. (1975). A first course in stochastic processes. San Diego: Academic Press.

[45] Kim, M. J., Nelson, C. R. and Startz R. (1991). Mean reversion in stock prices? A reappraisal of the empirical evidence. Review of Economic Studies, 58(3), 515-528.

[46] King, M. (2016). The End of Alchemy: Money, Banking and the Future of the Global Economy. New York: Norton & Company.

[47] Kirby, C. (1998). The restrictions on predictability implied by rational asset pricing models. Review of Financial Studies, 11(2), 343-382.

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

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

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

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

[52] 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.

[53] Lo, A. W. and MacKinlay, A. C. (1989). The size and power of the variance ratio test in finite samples: A Monte Carlo investigation. Journal of Econometrics, 40, 203-238.

[54] 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.

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

[56] Magdalinos (2014). Personal communication.

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

[58] Malkiel, B. G. (2015). A random walk down wall street. New York: Norton & Company.

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

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

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

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

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

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

[65] Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. Journal of Econometrics, 137(1), 134-161.

[66] Phillips, P. C. B. (2015). Halbert White Jr. Memorial JFEC Lecture: Pitfalls and Possibilities in Predictive Regression. Journal of Financial Econometrics, 13(3), 521-555.

[67] Phillips, P. C. B. and Jin, S. (2014). Testing the Martingale Hypothesis. Journal of Business & Economic Statistics, 32(4), 537-554.

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

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

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

[71] Phillips, P. C. B. and Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. Quantitative Economics, 2(3), 455-291.

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

[73] Pollard, D. (1984). Convergence of Stochastic Processes. New York: Springer-Verlag.

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

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

[76] Rosenblatt, M. (1985). Stationary Sequences and Random Fields, Boston: Birkhäuser.

[77] Ross, S. A. (2002). A Neoclassical Look at Behavioral Finance; Closed End Funds. (The Princeton Lectures in Finance III).

[78] Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. Annals of Statistics, 35(4), 1773-1801.

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

[80] Shiller, R. A. (2013). Speculative Asset prices. Nobel prize lecture.

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

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

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

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

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

[86] White, H., (2000), A Reality Check for Data Snooping, Econometrica 68, 1097-1126

[87] White, H., and I Domowitz (1984). Nonlinear Regression with Dependent Observations. Econometrica 52, 143-162.

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