



Variance ratio tests of random walk: An overview

Amélie Charles, Olivier Darné

► To cite this version:

Amélie Charles, Olivier Darné. Variance ratio tests of random walk: An overview. *Journal of Economic Surveys*, 2009, 23 (3), pp.503-527. 10.1111/j.1467-6419.2008.00570.x . hal-00771078

HAL Id: hal-00771078

<https://hal.science/hal-00771078v1>

Submitted on 20 Nov 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Variance ratio tests of random walk: An overview

Amélie Charles

Audencia Recherche, Audencia Group School of Management

Olivier Darné

LEMNA, Nantes University

Variance ratio tests of random walk: An overview

Abstract

This paper reviews the recent developments in the field of the variance-ratio tests of random walk and martingale hypothesis. In particular, we present the conventional individual and multiple VR tests as well as their improved modifications based on power- transformed statistics, rank and sign tests, subsampling and bootstrap methods, among others. We also re-examine the weak-form efficiency for five emerging equity markets in Latin America.

Keywords: Random walk hypothesis; Variance ratio tests; Stock market efficiency.

JEL Classification: G14; G15; C14.

1 Introduction

There exists a long tradition in the literature concerning the test of the random walk and martingale hypothesis, both in macroeconomics and finance. For instance, the random walk hypothesis [RWH] provides a mean to test the weak-form efficiency – and hence, non-predictability – of financial markets (Fama, 1970; 1991), and to measure the long-run effects of shocks on the path of real output in macroeconomics (Campbell and Mankiw, 1987; Cochrane, 1988; Cogley, 1990).

Given a time series $\{y_t\}_{t=1}^T$, the RWH correspond to $\phi = 1$ in the first-order autoregressive model

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t$$

where μ is an unknown drift parameter and the error terms ε_t are, in general, neither independent nor identically distributed (i.i.d.)¹.

Many statistical tests² were designed to test the RWH but a class of test, based on the variance-ratio [VR] methodology, has gained tremendous popularity in the recent years (see, e.g., Campbell and Mankiw, 1987; Cochrane, 1988; Lo and MacKinlay, 1988; Poterba and Summers, 1988). The VR methodology consists of testing the RWH against stationary alternatives, by exploiting the fact that the variance of random walk increments is linear in all sampling intervals, i.e., the sample variance of k -period return (or k -period differences), $y_t - y_{t-k}$, of the time series y_t , is k times the sample variance of one-period return (or the first difference), $y_t - y_{t-1}$. The VR at lag k is then defined as the ratio between $(1/k)$ th of the k -period return (or the k th difference) to the variance of the one-period return (or the first difference). Hence, for a random walk process, the variance computed at each individual lag interval k ($k = 2, 3, \dots$) should be equal to unity.

The use of the VR statistic can be advantageous when testing against several interesting alternatives to the random walk model, most notably those hypotheses associated with mean reversion. In fact, a number of authors (e.g., Lo and MacKinlay, 1989; Faust, 1992; Richardson and Smith, 1991) found that the VR statistic had optimal power against such alternative.

However, while the intuition behind the VR test is rather simple, conducting a statistical inference using the VR test is less straightforward. What makes thing complicated is that the VR test typically uses overlapping data in computing the

¹When the error terms are not an i.i.d. sequence, the random walk process is denominated *martingale process*, whereas the sequence $\{\varepsilon_t\}_{t=1}^T$ is the so-called martingale difference sequence (m.d.s.). Campbell et al. (1997) refers to the “random walk 3”.

²Daniel (2001) explores a wider range of possible test statistics.

variance of long-horizon returns. The use of overlapping was suggested by Lo and MacKinlay (1988) because it can potentially improve power of the VR test, but the use of overlapping data also adds to the difficulties of analyzing the exact distribution of the VR test statistic. However, virtually nothing is often known about the exact distribution of the VR test statistic that uses overlapping data, and not even its moments are known³. In practice, asymptotic distribution instead of exact distribution is often used for conducting statistical inference on the VR test, for fixed k and the sample size T increasing to infinity.

Lo and MacKinlay (1988) proposed two statistics for testing an individual VR estimate which are robust under homoscedasticity and heteroscedasticity. In practice, it is customary to examine the VR statistics for several k values. The null is rejected if it is rejected for some k value. However, as stressed by Chow and Denning (1993), this sequential procedure leads to an oversized testing strategy. In this context, multiple VR tests have been suggested such as multiple comparison tests (Chow and Denning, 1993) and Wald-type joint tests (Richardson and Smith, 1989; Cecchetti and Lam, 1994). Even though the individual Lo-MacKinlay and multiple VR tests are quite powerful testing for homoscedastic or heteroscedastic nulls, it is critical to note that these tests are asymptotic tests in that their sampling distributions are approximated by their limiting distributions. Indeed, the sampling distribution of the VR statistic can be far from normal in finite sample, showing severe bias and right skewness. These finite sample deficiencies may give rise to serious size distortions or low power, which can lead to misleading inferences. This is especially true when the sample size is not large enough to justify asymptotic approximations (Cecchetti and Lam, 1994). To circumvent this problem, some alternatives⁴ have been proposed, such as Chen and Deo (2006) with a power-transformed VR statistic, Wright (2000) with exact VR tests based on rank and sign, Whang and Kim (2003) with subsampling method, and Kim (2006) with bootstrap method, among others. Therefore, because of the important literature on the VR tests we propose an overview on this subject.

The rest of this paper is organized as follows. The individual and multiple VR tests are presented in Section 2 and 3, respectively. Section 4 discusses bootstrapping VR tests. An empirical illustration is proposed in Section 5. Section 6 concludes.

³Recently, Kan (2006) provided analytical formulas for the moments of the sample variance ratio under both the null and the alternatives. See also Shively (2002) for the case of mean.

⁴Hoque et al. (2007) proposed a comparison of several variance ratio tests.

2 Individual variance ratio tests

The VR test is often used (see Cochrane, 1988; Lo and MacKinlay, 1988; Poterba and Summers, 1988; among others) to test the hypothesis that a given time series or its first difference (or return), $x_t = y_t - y_{t-1}$, is a collection of i.i.d. observations or that it follows a martingale difference sequence. Define the VR of k -period return as

$$\begin{aligned} V(k) &= \frac{\text{Var}(x_t + x_{t-1} + \cdots + x_{t-k+1})/k}{\text{Var}(x_t)} \\ &= \frac{\text{Var}(y_t - y_{t-k})/k}{\text{Var}(y_t - y_{t-1})} = 1 + 2 \sum_{i=1}^{k-1} \left(\frac{(k-i)}{k} \right) \rho_i \end{aligned}$$

where ρ_i is the i -th lag autocorrelation coefficient of $\{x_t\}$. $V(k)$ is a particular linear combination of the first $(k-1)$ autocorrelation coefficients, with linearly declining weights. The central idea of the variance ratio test is based on the observation that when returns are uncorrelated over time, we should have $\text{Var}(x_t + \cdots + x_{t-k+1}) = k\text{Var}(x_t)$, i.e. $V(k) = 1$. One can therefore think of VR test as a specification test of $H_0 : \rho_1 = \cdots = \rho_k = 0$, i.e., returns are serially uncorrelated.

A test can be constructed by considering statistic based on an estimator of $V(k)$

$$VR(k) = \frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(1)}$$

where $\hat{\sigma}^2(1)$ is the unbiased estimator of the one-period return variance, using the one-period returns x_t , and is defined as

$$\begin{aligned} \hat{\sigma}^2(1) &= (T-1)^{-1} \sum_{t=1}^T (x_t - \hat{\mu})^2 \\ &= (T-1)^{-1} \sum_{t=1}^T (y_t - y_{t-1} - \hat{\mu})^2 \end{aligned} \tag{1}$$

with $\hat{\mu} = T^{-1} \sum_{t=1}^T x_t$ is the estimated mean. For the estimator of k -period return variance $\hat{\sigma}^2(k)$, using k -period returns $(x_t + \cdots + x_{t-k+1})$, there are many ways to do it. Due to limited sample size and the desire to improve the power of the test, this estimator is often performed using overlapping long-horizon returns (k -period), as

advocated by Lo and MacKinlay (1988)⁵, and it is defined as

$$\begin{aligned}\hat{\sigma}^2(k) &= m^{-1} \sum_{t=k}^T (x_t + x_{t-1} + \cdots + x_{t-k+1} - k\hat{\mu})^2 \\ &= m^{-1} \sum_{t=k}^T (y_t - y_{t-k} - k\hat{\mu})^2\end{aligned}$$

where $m = k(T - k + 1)(1 - kT^{-1})$. The value of m is chosen such as $\hat{\sigma}^2(k)$ is an unbiased estimator of the k -period return variance when σ_t^2 is constant over time.

Following Wright (2000), the VR statistic can be written as

$$VR(x; k) = \left\{ (Tk)^{-1} \sum_{t=k}^T (x_t + \cdots + x_{t-k+1} - k\hat{\mu})^2 \right\} \div \left\{ T^{-1} \sum_{t=1}^T (x_t - \hat{\mu})^2 \right\} \quad (2)$$

Moreover, Cochrane (1988) showed that the estimator of $V(k)$ can be interpreted in terms of the frequency domain. This estimator which uses the usual consistent estimators of variance is asymptotically equivalent to 2π the normalized spectral density estimator at the zero frequency which uses the Bartlett kernel. Formally, we have

$$VR_f = 2\pi \frac{f_{\Delta y}(0)}{\hat{\sigma}^2(1)} \quad (3)$$

where $f_{\Delta y}(0)$ represents the estimator of the spectrum evaluated at frequency 0 with

$$\begin{aligned}2\pi f_{\Delta y}(0) &= 2\pi \sum_{j=1}^m W_k(\lambda_j) I_{\Delta y}(\lambda_j) \\ I_{\Delta y} &= (2\pi T)^{-1} |d_y(\lambda_j)|^2 \\ d_y &= \sum_{t=1}^{T-1} [y_t - y_{t-1} - \hat{\mu}] e^{-i\lambda_j t} \\ W_k(\lambda_j) &= \sum_{|j| \leq k} (1 - |j|/k) \exp(-ij\lambda) = k^{-1} \left[\frac{\sin(k\lambda_j/2)}{\sin(\lambda_j/2)} \right]^2\end{aligned} \quad (4)$$

where $\lambda_j = 2\pi jT^{-1}$, $j = 1, \dots, T-1$, $I_{\Delta y}$ denotes the periodogram, d_y is the discrete Fourier transform, $\hat{\mu}$ is an estimate of the mean of Δy_t , and $W(\lambda_j)$ is the Bartlett window.

If the data-generating process of time series is a random walk, the expected value of $VR(x; k)$ should be equal to unity for all horizons k . If returns are positively

⁵Lo and MacKinlay (1988) and Campbell et al. (1997) argued that using overlapping data in estimating the variances allowed to obtain a more efficient estimator and hence a more powerful test.

(negatively) autocorrelated, the VR should be higher (lower) than unity. Time series (in level) is said to be mean-reverting if $VR(x; k)$ is significantly lower than unity at long horizons k . On the contrary, time series is mean averting, i.e. explosive, if $VR(x; k)$ is significantly higher than unity at long horizons (Poterba and Summers, 1988).

We describe the most popular individual VR tests developed by Lo and MacKinlay (1988) as well as some of its improvements.

2.1 Lo and MacKinlay (1988) tests

Lo and MacKinlay (1988) proposed the asymptotic distribution of $VR(x; k)$ by assuming that k is fixed when $T \rightarrow \infty$. They showed that if x_t is i.i.d., i.e. under the assumption of homoscedasticity, then under the null hypothesis that $V(k) = 1$, the test statistic $M_1(k)$ is given by

$$M_1(k) = \frac{VR(x; k) - 1}{\phi(k)^{1/2}} \quad (5)$$

which follows the standard normal distribution asymptotically. The asymptotic variance, $\phi(k)$, is given by

$$\phi(k) = \frac{2(2k-1)(k-1)}{3kT} \quad (6)$$

To accommodate x_t 's exhibiting conditional heteroscedasticity, Lo and MacKinlay (1988) proposed the heteroscedasticity⁶ robust test statistic $M_2(k)$

$$M_2(k) = \frac{VR(x; k) - 1}{\phi^*(k)^{1/2}} \quad (7)$$

which follows the standard normal distribution asymptotically under null hypothesis that $V(k) = 1$, where

$$\begin{aligned} \phi^*(k) &= \sum_{j=1}^{k-1} \left[\frac{2(k-j)}{k} \right]^2 \delta(j) \\ \delta(j) &= \left\{ \sum_{t=j+1}^T (x_t - \hat{\mu})^2 (x_{t-j} - \hat{\mu})^2 \right\} \div \left\{ \left[\sum_{t=1}^T (x_t - \hat{\mu})^2 \right]^2 \right\} \end{aligned}$$

The $M_2(k)$ test is applicable to x_t 's generated from a martingale difference time series (see Appendix for a discussion on the assumptions). The usual decision rule for the standard normal distribution is applied to both tests.

⁶It has been argued that misleading conclusions may be obtained with VR statistics when time-varying volatility is present in the data. See, for example, Kim, Nelson and Startz (1991, 1998) and Kim and Nelson (1998) who also proposed a solution based on a Bayesian approach and the use of a Gibbs sampler.

The finite-sample properties of the VR test were studied by Lo and MacKinlay (1989), who found that the two-sided test has size generally quite close to the nominal level, as long as the test is robustified against any conditional heteroscedasticity. The VR statistic has been found by several authors (e.g., Richardson and Smith, 1991; Faust, 1992) to be particularly powerful when testing against mean reverting alternatives to the random walk model, particularly when k is large.

2.2 Chen and Deo (2006) test

It is critical to note that the conventional VR tests, such as the Lo-MacKinlay test, are asymptotic tests in that their sampling distributions are approximated by their limiting distributions. Indeed, the practical use of the statistic has been impeded by the fact that the asymptotic theory provides a poor approximation to the small-sample distribution of the VR statistic. In general, the ability of the asymptotic distribution to approximate the finite sample distribution depends crucially on the value of the horizon k . More specifically, rather than being normally distributed (when standardized by $\sqrt{T}$) as the theory states, the statistics are severely biased and right skewed for large k (relative to T) (Lo and MacKinlay, 1989), which makes application of the statistic problematic. In other words, the finite-sample null distribution of the test statistic is quite asymmetric and non-normal.

A solution is provided in a series of theoretical papers such as those by Richardson and Stock (1989), Deo and Richardson (2003), Perron and Vodounou (2005) and Chen and Deo (2006)⁷. For example, to circumvent this problem, Richardson and Stock (1989) provided alternative asymptotic distribution of the VR statistic under the random walk null, assuming that both k and T increase to infinity but in such a way that k/T converges to a positive constant δ that is strictly less than 1⁸. Through Monte Carlo simulations, they demonstrated that this new distribution provides a far more robust approximation to the small-sample distribution of the VR statistic. Most current applications of the VR statistic cite the $k/T \rightarrow \delta > 0$ result as justification for using Monte Carlo distributions (i.e., set at $k = \delta T$) as representative of the VR statistic's sampling distribution. Perron and Vodounou (2005) also studied the VR statistic's

⁷Kan (2006) presented the exact distributions of the VR test with overlapping data. Moreover, Tse, Ng and Zhang (2004) suggested a modified VR statistic and proposed to approximate the small-sample distribution of this statistic using a beta distribution that matches the exact mean and the asymptotic variance.

⁸Richardson and Stock (1989) showed that the VR statistic, without any normalization, converges to a functional of Brownian motion.

properties under the Richardson and Stock (1989) framework, and characterized the maximal possible power by taking a continuous-time limit given a fixed data span T . However, Deo and Richardson (2003) argued that the VR statistic is inconsistent against an important class of mean reverting alternatives under this framework when the horizon k is increasing proportional to the sample size, i.e. $k/T \rightarrow \delta > 0$. Chen and Deo (2006) also showed that the $k/T \rightarrow \delta > 0$ asymptotic distribution cannot approximate the finite-sample distribution of the VR statistic when k/T is small and is sensitive to conditional heteroscedasticity.

Chen and Deo (2006) suggested a simple power transformation of the VR statistic that, when k is not too large⁹, provides a better approximation to the normal distribution in finite samples and is able to solve the well-known right skewness problem. They showed that the transformed VR statistic leads to significant gains in power against mean reverting alternatives. Furthermore, the distribution of the transformed VR statistic is shown, both theoretically and through simulations, to be robust to conditional heteroscedasticity¹⁰.

First, they defined the VR statistic based on the periodogram as

$$VR_p(k) = \frac{1}{(1 - k/T)} \frac{4\pi}{T\hat{\sigma}^2} \sum_{j=1}^{(T-1)/2} W_k(\lambda_j) I_{\Delta y}(\lambda_j) \quad (8)$$

where $I_{\Delta y}(\lambda_j)$ and $W_k(\lambda_j)$ are defined as in (4). This expression of the VR statistic is precisely the normalized discrete periodogram average estimate of the spectral density of a stationary process at the origin (Brockwell and Davis, 1996). To obtain their transformed VR statistic, noted $VR_p^\beta(k)$, they applied the following power transformation¹¹ to $VR_p(k)$

$$\beta = 1 - \frac{2}{3} \frac{\left(\sum_{j=1}^{[(T-1)/2]} W_k(\lambda_j) \right) \left(\sum_{j=1}^{[(T-1)/2]} W_k^3(\lambda_j) \right)}{\left(\sum_{j=1}^{[(T-1)/2]} W_k^2(\lambda_j) \right)^2} \quad (9)$$

2.3 Wright (2000) tests

As already noted, the Lo-MacKinlay tests, which are asymptotic tests whose sampling distribution is approximated based on its limiting distribution, are biased and right-skewed in finite samples. In this respect, Wright (2000) proposed a nonparametric

⁹Deo and Richardson (2003) advocated that large values of k should not be used when testing for the mean reversion using the VR statistics.

¹⁰To adjust for conditional heteroscedasticity, Chen and Deo (2006) proposed a modified version of the standard deviation of the transformed VR statistic (even for standard VR statistic).

¹¹See Chen and Deo (2004) for a discussion on power transformations.

alternative to conventional asymptotic VR tests using signs and ranks. Wright's (2000) tests have two advantages over Lo-MacKinlay test when sample size is relatively small: (i) as the rank (R_1 and R_2) and sign (S_1 and S_2) tests have exact sampling distribution, there is no need to resort to asymptotic distribution approximation, and (ii) the tests may be more powerful than the conventional VR tests against a wide range of models displaying serial correlation, including fractionally integrated alternatives. The tests based on ranks are exact under the i.i.d. assumption, whereas the tests based on signs are exact even under conditional heteroscedasticity. Moreover, Wright (2000) showed that ranks-based tests display low size distortions, under conditional heteroscedasticity.

Given T observations of first differences of a variable, $\{x_1, \dots, x_T\}$, and let $r(x)$ be the rank of x_t among $(x_1, \dots, x_T)$. Under the null hypothesis that x_t is generated from an i.i.d. sequence, $r(x)$ is a random permutation of the numbers of $1, \dots, T$ with equal probability. Wright (2000) suggested the R_1 and R_2 statistics, defined as

$$R_1(k) = \left(\frac{(Tk)^{-1} \sum_{t=k}^T (r_{1,t} + \dots + r_{1,t-k+1})^2}{T^{-1} \sum_{t=k}^T r_{1,t}^2} - 1 \right) \times \phi(k)^{-1/2} \quad (10)$$

$$R_2(k) = \left(\frac{(Tk)^{-1} \sum_{t=k}^T (r_{2,t} + \dots + r_{2,t-k+1})^2}{T^{-1} \sum_{t=k}^T r_{2,t}^2} - 1 \right) \times \phi(k)^{-1/2} \quad (11)$$

where the standardized ranks $r_{1,t}$ and $r_{2,t}$ are given by

$$r_{1,t} = \frac{r(x_t) - \frac{T+1}{2}}{\sqrt{(T-1)(T+1)/12}}$$

$$r_{2,t} = \Phi^{-1} \frac{r(x)}{T+1}$$

where $\phi(k)$ is defined in (6), and Φ^{-1} is the inverse of the standard normal cumulative distribution function. The R_1 and R_2 statistics follow the same exact sampling distribution. The critical values of these tests can be obtained by simulating their exact distributions.

The tests based on the signs of first differences are given by

$$S_1(k) = \left(\frac{(Tk)^{-1} \sum_{t=k}^T (s_t + \dots + s_{t-k+1})^2}{T^{-1} \sum_{t=k}^T s_t^2} - 1 \right) \times \phi(k)^{-1/2} \quad (12)$$

$$S_2(k) = \left(\frac{(Tk)^{-1} \sum_{t=k}^T (s_t(\bar{\mu}) + \dots + s_{t-k+1}(\bar{\mu}))^2}{T^{-1} \sum_{t=k}^T s_t(\bar{\mu})^2} - 1 \right) \times \phi(k)^{-1/2} \quad (13)$$

where $\phi(k)$ is defined in (6), $s_t = 2u(x_t, 0)$, $s_t(\bar{\mu}) = 2u(x_t, \bar{\mu})$, and

$$u(x_t, q) = \begin{cases} 0.5 & \text{if } x_t > q \\ -0.5 & \text{otherwise} \end{cases}$$

Similarly to R_1 and R_2 tests, the critical values of the S_1 and S_2 tests can be obtained by simulating its exact sampling distribution.

Note that S_1 assumes a zero drift value. In a recent work, Luger (2003) used ranks and signs to extend the Campbell and Dufour (1997) nonparametric approach to test for random walk with unknown drift. Belaire-Franch and Contreras (2004) and Belaire-Franch and Opong (2005a) applied the procedure described in Luger (2003) to compute S_2 .

2.4 Correction of size distortions

As pointed out by Wright (2000), using several k values in the Wright's tests would lead to an over rejection of the null hypothesis, as in Lo and MacKinlay's tests context (see Belaire-Franch and Opong, 2005). To overcome these test-size distortions, Belaire-Franch and Contreras (2004) and Belaire-Franch and Opong (2005) proposed different approaches to control the size of Wright's tests. They considered the application of individual VR tests as the application of different individual tests. Then, they applied different p -value adjustments for multiplicity, in line with Psaradakis (2000).

- They computed the Sidack-adjusted p -value for each test j as

$$\tilde{p}_{ij}^{(S)} = 1 - (1 - p_{ij})^m \quad i = 1, \dots, m$$

where p_{ij} is the p -value corresponding to the VR test j computed for an individual k value, and m is the number of k values.

- They also employed the Hochberg (1988) adjusted p -values which are obtained as

$$\tilde{p}_{ij}^{(H)} = \min\{[k - R(p_{ij}) + 1]p_{ij}, 1\}$$

Given a significance level α , the decision rule states that, using the VR test j , the null is rejected if $\tilde{p}_j^{(S)} = \min_{1 \leq i \leq m} \tilde{p}_{ij}^{(S)} \leq \alpha$ or $\tilde{p}_j^{(H)} = \min_{1 \leq i \leq m} \tilde{p}_{ij}^{(H)} \leq \alpha$.

However, these methods assume that the test statistics computed at different intervals are uncorrelated. In order to take into account possible correlations among the statistics, Belaire-Franch and Contreras (2004) and Belaire-Franch and Opong (2005) suggested to compute bootstrap-adjusted p -values as described in Psaradakis (2000). The goal of the procedure is to obtain an approximation to the null sampling distribution of $\min_{1 \leq i \leq m} p_{ij}$ by resampling with replacement from the original returns¹².

¹²Note that the p -value adjustments can be applied to other VR tests for a joint hypothesis.

2.5 Choi (1999) test

When implementing the VR tests, the choice of holding period k is important. However, this choice is usually rather arbitrary and *ad hoc*. To overcome this issue, Choi (1999) proposed a data-dependent procedure to determinate the optimal value of k . Choi (1999) suggested a VR test based in frequency domain since Cochrane (1988) showed that the estimator of $V(k)$ which uses the usual consistent estimators of variance is asymptotically equivalent to 2π the normalized spectral density estimator at the zero frequency which uses the Bartlett kernel. However, Choi (1999) employed rather the Quadratic Spectral [QS] kernel because this kernel is optimal in estimating the spectral density at the zero frequency (Andrews, 1991). The VR estimator is defined as

$$VR(k) = 1 + 2 \sum_{i=1}^{T-1} h(i/k) \hat{p}(i), \quad (14)$$

where $\hat{p}(i)$ is the autocorrelation function, and $h(x)$ is the QS window defined as

$$h(x) = \frac{25}{12\pi^2 x^2} \left[\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right]$$

The standardized statistic is

$$VR_f = \frac{VR(k) - 1}{(2)^{1/2} (T/k)^{-1/2}} \quad (15)$$

Under the null hypothesis the test statistic VR_f follows the standard normal distribution asymptotically¹³. Note that it is assumed that $T \rightarrow \infty$, $k \rightarrow \infty$ and $T/k \rightarrow \infty$.

Various methods for optimally selecting the truncation point for the spectral density at the zero frequency are available (Andrews, 1991; Andrews and Monahan, 1992; Newey and West, 1994; among others). Choi (1999) employed the Andrews's (1991) methods to select the truncation point optimally and compute the VR test. Note that the small sample properties of this automatic VR test under heteroscedasticity are unknown and have not investigated properly.

3 Multiple variance ratio tests

The Lo-MacKinlay test is an individual test where the null hypothesis is tested for an individual value of k . The question as to whether or not a time series is mean-reverting requires that the null hypothesis hold true for all values of k . In view of

¹³The estimator VR_f based on Bartlett window as in Cochrane (1988) also has a limiting normal distribution. However, Cogley (1990) showed that this estimator seems inappropriate because it is right skewed. Therefore, Cogley (1990) proposed to approximate this estimator by a multiple of chi-square variate, giving improvement over the normal.

this, it is necessary to conduct a joint test where a multiple comparison of VRs over a set of different time horizons is made. However, conducting separate individual tests for a number of k values may be misleading as it leads to over rejection of the null hypothesis of a joint test, above the nominal size. As stressed by Chow and Denning (1993), this sequential procedure leads to an oversized testing strategy. Thus, the weakness of Lo-MacKinlay's test is that it ignores the joint nature of testing for the RWH.

We present some multiple VR tests based on multiple comparison tests (Chow and Denning, 1993; Whang and Kim, 2003; Belaire-Franch and Contreras, 2004) and Wald-type tests (Richardson and Smith, 1991; Cecchetti and Lam, 1993; Chen and Deo, 2006) that combine the information contained in statistics at several horizons. These multiple VR tests consider the following joint null hypothesis $H_0: V(k_i) = 1$ for all $i = 1, \dots, m$, against the alternative $H_1: V(k_i) \neq 1$ for some k_i .

3.1 Chow and Denning (1993) tests

Chow and Denning (1993) proposed to using Hochberg's (1974) procedure for the multiple comparison of the set of VR estimates with unity, which allow us to examine a vector of individual VR tests while controlling for overall test size. For a set of m test statistics, the RWH is rejected if any one of the estimated VRs is significantly different from one.

To test the joint null hypothesis, Chow-Denning's (1993) test statistic is defined as

$$MV_1 = \sqrt{T} \max_{1 \leq i \leq m} |M_1(k_i)| \quad (16)$$

where $M_1(k_i)$ is defined in (5). This is based on the idea that the decision regarding the null hypothesis can be obtained from the maximum absolute value of the individual VR statistics. In order to control the size of the multiple VR test and because the limit distribution of these statistics is complex, they applied the Sidak (1967) probability inequality and give an upper bound to the critical values taken in the studentized maximum modulus [SMM] distribution. Indeed, the statistic follows the SMM distribution with m and T degrees of freedom, i.e. $SMM(\alpha, m, T)$, where m is the number of k values¹⁴. The null hypothesis is rejected at α level of significance if the MV_1 statistic is greater than the $[1 - (\alpha^*/2)]$ th percentile of the standard normal distribution where $\alpha^* = 1 - (1 - \alpha)^{1/m}$.

¹⁴The critical values of the test are tabulated in Hahn and Hendrickson (1971) and Stoline and Ury (1979). It should be noted that when T is large, the critical values of the test can be calculated from the limiting distribution of the statistic.

Similarly, the heteroscedasticity-robust version of the Chow-Denning test MV_2 can be written as

$$MV_2 = \sqrt{T} \max_{1 \leq i \leq m} |M_2(k_i)| \quad (17)$$

where $M_2(k_i)$ is defined in (7), and it has the same critical values as MV_1 . However, with finite-sample sizes it may be preferable to use critical values obtained by simulations as done by Chow and Denning themselves. Nevertheless, as pointed out by Fong et al. (1997), Hochberg's approach is valid only if the vector of test statistics is multivariate normal. This condition is satisfied by VRs if there is little overlap in the data, i.e. if k/T is small.

3.2 Whang and Kim (2003) test

Whang and Kim (2003) developed a multiple VR test which uses a subsampling technique of Politis, Romano and Wolf (1997), which is a data-intensive method of approximating the sampling distribution. It can show better properties than the conventional VR tests when the sample size is relatively small. The Monte Carlo experiment results reported in Whang and Kim (2003) confirmed that their new VR test show excellent power in small samples, coupled with little or no serious size distortions.

To test the joint null hypothesis, Whang and Kim (2003) considered the statistic

$$MV_T = \sqrt{T} g_N(x_1, \dots, x_T) \quad (18)$$

where

$$g_t(x_1, \dots, x_T) = \max_{1 \leq i \leq m} |M_r(k_i)|$$

with $M_r(k_i) = VR(x; k_i) - 1$, and $VR(x; k)$ is as defined in (2). The sampling distribution function for the MV_T statistic is written as

$$G_T(x) = P\left(\sqrt{T} g_t(x_1, \dots, x_T) \leq x\right)$$

Whang and Kim (2003) showed that the asymptotic null distribution of the statistic is that of a maximum of a multivariate normal vector with unknown covariance matrix, which is complicated to estimate. Therefore, they proposed to approximate the null distribution by means of the subsampling approach.

Consider a subsample $(x_t, \dots, x_{t-b+1})$ of size b for $t = 1, \dots, T - b + 1$. The statistic MV_T calculated from the subsample is denoted as $g_{T,b,t} = g_b(x_t, \dots, x_{t-b+1})$. Then, $G_T(x)$ is approximated by the distribution function obtained by the collection of $g_{T,b,t}$'s calculated from all individual subsamples. It can be written as

$$\hat{G}_{T,b}(x) = (T - b + 2)^{-1} \sum_{t=0}^{T-b+1} l\left(\sqrt{b}g_{T,b,t} \leq x\right)$$

where $l(\cdot)$ is the indicator function that takes 1 if the condition inside the bracket is satisfied and 0 otherwise.

The $100(1 - \alpha)\%$ critical value for the test can be calculated as the $(1 - \alpha)$ th percentile of $\hat{G}_{T,b}$, while the p -value of the test is estimated as $1 - \hat{G}_{T,b}(MV_T)$. The null hypothesis that $V(k_i) = 1$ ($i = 1, \dots, m$) is rejected at the level of significance α if the observed MV_T is greater than this critical value or if the p -value is less than α . To implement the subsampling technique, a choice of block length b should be made. Whang and Kim (2003) recommended that a number of block lengths from an equally spaced grid in the interval of $[2.5T^{0.3}, 3.5T^{0.6}]$ be taken. However, they found that the size and power properties of their test are not sensitive to the choice of the block length.

3.3 Belaire-Franch and Contreras (2004) tests

Recently, Belaire-Franch and Contreras (2004) proposed to substitute the standard VR tests by Wright's rank and sign-based tests, in the definition of Chow and Denning (1993) procedure to create a multiple rank and sign VR tests. The statistics are defined as

$$\begin{aligned} CD_{(R_1)} &= \max_{1 \leq i \leq m} |R_1(k_i)| \\ CD_{(R_2)} &= \max_{1 \leq i \leq m} |R_2(k_i)| \\ CD_{(S_1)} &= \max_{1 \leq i \leq m} |S_1(k_i)| \\ CD_{(S_2)} &= \max_{1 \leq i \leq m} |S_2(k_i)| \end{aligned}$$

The ranks-based procedures are exact under the i.i.d. assumption whereas the signs-based procedures are exact under both the i.i.d. and martingale difference sequence assumption. Belaire-Franch and Contreras (2004) showed that the ranks-based tests $CD_{(R_1)}$ and $CD_{(R_2)}$ are more powerful than their signs-based counterparts, $CD_{(S_1)}$ and $CD_{(S_2)}$ ¹⁵.

¹⁵Belaire-Franch and Contreras (2004) also suggested to substitute the Wright's rank and sign-based

Moreover, Colletaz (2005) and Kim and Shansuddin (2007) also proposed an extension to the Wright's VR methodology following Chow-Denning, but only for the rank (R_1) and sign (S_1) tests, respectively.

3.4 Wald-type tests

3.4.1 Richardson and Smith (1991) test

Richardson and Smith (1991) suggested a joint test based on the following Wald statistic

$$RS(k) = T (\mathbf{VR} - \mathbf{1}_k)' \Phi^{-1} (\mathbf{VR} - \mathbf{1}_k)$$

where $\mathbf{VR}$ is the $(k \times 1)$ vector of sample k variance ratios, $\mathbf{1}_k$ is the $(k \times 1)$ unit vector, and Φ is the covariance matrix of $\mathbf{VR}$. The joint $RS(k)$ statistic follows a χ^2 distribution with k degrees of freedom.

The usefulness of this test relies on the fact that, whenever the VR tests are computed over long lags with overlapping observations, the distribution of the VR test is non-normal; then, neither the Lo-MacKinlay test nor Chow-Denning procedure is valid for drawing inferences.

Moreover, Fong et al. (1997) argued that Richardson and Smith's (1991) joint VR test can be more powerful than Chow-Denning multiple comparison test for empirically relevant alternatives, and it displays low size distortion in the presence of heteroscedastic increments. However, their simulation results are based on an ARCH process with slope coefficient 0.1, which is "practically" an i.i.d. process. Therefore, the conclusion of Fong et al. (1997) could not hold under heteroscedasticity.

3.4.2 Cecchetti and Lam (1994) test

Cecchetti and Lam (1994) proposed a multivariate version of the VR statistic to test the RWH, in order to control the investment horizon. They suggested the following Wald statistic which incorporates the correlations between VR statistics at various horizons and weights them according to their variances

$$S(k) = [\mathbf{VR}(k) - E[\mathbf{VR}(k)]]' \Sigma^{-1}(k) [\mathbf{VR}(k) - E[\mathbf{VR}(k)]]$$

where E is the expectation operator, $\mathbf{VR}$ a column vector sequence of VR statistics $\mathbf{VR}(k) = [VR(2), \dots, VR(q)]$, and $\Sigma(k)$ is a measure of the covariance matrix of $\mathbf{VR}$.

tests in the definition of Richardson and Smith (1991) procedure but found that these tests were inferior to rank and sign-based $CD_{(\cdot)}$ tests.

The joint VR $S(k)$ statistic follows a χ^2 distribution with k degrees of freedom. Cecchetti and Lam (1994) studied the empirical distribution of $S(k)$ using Monte Carlo techniques. For each simulation, they computed a value for the statistic $S(k)$, using the mean vector and covariance matrix $\tilde{V}R(k)$ and $\tilde{\Sigma}(k)$ taken as the true population values, and tabulate the distribution¹⁶.

However, as suggested by Cecchetti and Lam (1994), the empirical distributions of the VR have large positive skewness, suggesting that inference based on the χ^2 distribution will be misleading.

3.4.3 Chen and Deo (2006) test

Chen and Deo (2006) also proposed a joint VR test based on their individual power transformed VR statistic. They define the following Wald statistic

$$QP(k) = (\mathbf{V}_{p,\beta}(k) - \mu_\beta)' \Sigma_\beta^{-1} (\mathbf{V}_{p,\beta}(k) - \mu_\beta)$$

where $\mathbf{V}_{p,\beta}$ a column vector sequence of VR statistics $\mathbf{V}_{p,\beta}(k) = [VR_p^\beta(2), \dots, VR_p^\beta(k)]$ with $VR_p^\beta(k)$ the power transformed VR as in (9), μ_β and $\Sigma(k)_\beta$ are a measure of the expectation and covariance matrix of $\mathbf{V}_{p,\beta}$, respectively. The joint VR $QP(k)$ statistic follows a χ^2 distribution with k degrees of freedom. Moreover, Chen and Deo (2006) showed from Monte Carlo simulations that their joint VR test displayed much higher power.

Note that the Chen-Deo (2006) test is a joint test with one-sided alternative ($H_1: V(k_i) < 1$, for some k_i) while the Richardson-Smith (1991) and Cecchetti-Lam (1994) tests are joint tests with two-sided alternative ($H_1: V(k_i) \neq 1$). Therefore, at α level of significance, the null hypothesis that $V(k_i) = 1$ is rejected if the test statistic is greater than the upper 2α critical value of a χ^2 distribution for the Chen-Deo (2006) test and than the upper α critical value of a χ^2 distribution for the Richardson-Smith (1991) and Cecchetti-Lam (1994) tests.

4 Bootstrapping variance ratio tests

As already noted, Wright (2000), based on ranks and signs, and Whang and Kim (2003), using the subsampling method, proposed the VR tests which do not rely

¹⁶Note that Cecchetti and Lam (1994) showed that the empirical distribution of the statistic $S(k)$ is numerically identical to the quadratic sum of the deviations of the first $(k-1)$ autocorrelations from their population values, weighted by their covariance matrix.

asymptotic approximations in order to overcome the difficulties due to using VR tests based on asymptotic approximations (severe bias and right skewness). As an alternative, some researchers proposed to employ a bootstrap method, which is a resampling method that approximates the sampling distribution of a test statistic (Efron, 1979), to the VR test statistic. The bootstrap is a distribution-free randomization technique, which can be used to estimate the sampling distribution of the VR statistic, when the distribution of the original population is unknown. We describe the two most used bootstrapping VR tests, i.e. those suggested by Kim (2006) in a theoretical framework and by Malliaropulos and Priestley (1999) in an empirical framework¹⁷.

4.1 Kim (2006) test

Kim (2006) used the wild bootstrap which is a resampling method that approximates the sampling distribution of the VR statistic, and is applicable to data with unknown forms of conditional and unconditional heteroscedasticity (see Mammen, 1993; Davidson and Flachaire, 2001).

Kim (2006) applied the wild bootstrap to Lo-MacKinlay, $M_2(k)$, and Chow-Denning, $MV_2(k_i)$, VR tests. The wild bootstrap test based on $MV_2(k_i)$ can be conducted in three stages as below

- (i) Form a bootstrap sample of T observations $X_t^* = \eta_t X_t$ ($t = 1, \dots, T$) where η_t is a random sequence with $E(\eta) = 0$ and $E(\eta^2) = 1$.
- (ii) Calculate $MV^* = MV_2(X^*; k_i)$, the $MV_2(X^*; k_i)$ statistic obtained from the bootstrap sample generated in stage (i).
- (iii) Repeat (i) and (ii) sufficiently many, say m , times to form a bootstrap distribution of the test statistic $\{MV_2(X^*; k_j; j)\}_{j=1}^m$.

The p -value of the test can be obtained as the proportion of $\{MV_2(X^*; k_j; j)\}_{j=1}^m$ greater than the sample value of $MV_2(k_i)$. The wild bootstrap version of $M_2(k)$ test can be implemented in a similar manner as a two-tailed test, where we obtain $M^* = M_2(X^*; k)$ in stage (ii) and $\{M_2(X^*; k_j; j)\}_{j=1}^m$ in stage (iii).

Conditionally on X_t , X_t^* is a serially uncorrelated sequence with zero mean and variance X_t^2 . As such, M^* and MV^* have the same asymptotic distributions as $M_2(k)$ and $MV_2(k_i)$, respectively. Since X_t^* is a serially uncorrelated sequence,

¹⁷Recently, Fleming et al. (2006) developed a bootstrap method for testing multiple inequality restrictions on VRs.

wild bootstrapping approximates the sampling distributions under the null hypothesis, which is a desirable property for a bootstrap test.

To implement the wild bootstrap test, a specific form of η_t should be chosen. Kim (2006) recommends using the standard normal distribution for η_t since he reports that other choices provided qualitatively similar small sample results. Note that the wild bootstrap is valid and the test statistics being bootstrapped are pivotal asymptotically under the condition that X_t follows a martingale difference sequence¹⁸.

Kim (2006) showed that the sub-sampling test of Whang and Kim (2003) displays small sample properties far inferior to the wild bootstrap test under a small sample size.

4.2 Malliaropulos and Priestley (1999) test

Malliaropulos and Priestley (1999) used a weighted bootstrap method proposed by Wu (1986) which is robust to the presence of heteroscedasticity, which is done by resampling normalized returns instead of actual returns. Basically, the returns are normalized by multiplying each observation of actual returns, for each one of the time series of returns, by a corresponding random factor and resample from these normalized returns¹⁹. The bootstrap scheme can be summarized with the following algorithm

- (i) For each t , draw a weighting factor z_t^* ($t = 1, \dots, T$) with replacement from the empirical distribution of normalized returns $z_t = (r_t - \bar{r})/\sigma(r)$, where $\bar{r} = T^{-1} \sum_{t=1}^T r_t$ is the mean and $\sigma(r) = \sqrt{T^{-1} \sum_{t=1}^T (r_t - \bar{r})^2}$ is the standard error of return.
- (ii) Form the bootstrap sample of T observations $\tilde{r}_t^* = z_t^* r_t$ ($t = 1, \dots, T$) by multiplying each observation of actual returns with its corresponding random weighting factor.
- (iii) Calculate the VR statistic $VR^*(k)$ from the pseudo data r_t^* for $k = 1, \dots, K$.
- (iv) Repeat steps (i) and (ii) M times, obtaining $VR^*(k; m)$ ($m = 1, \dots, M$) and calculate the relevant quantiles, mean, median and standard deviation of

¹⁸See MacKinnon (2002) for the advantages of bootstrapping asymptotical pivotal statistics. Note that there are other possible choices of two-point distributions for the wild bootstrap, which potentially outperform the standard normal distribution when the sample size is small (Davidson et al., 2007).

¹⁹See Malliaropulos and Priestley (1999) and Cajueiro and Tabak (2006) for a detailed discussion on the weighted bootstrap method proposed by Wu (1986).

the sampling distribution of $VR^*(k)$ under the null hypothesis of serially uncorrelated returns.

Using this procedure, resampling from normalized returns instead from actual returns, the weighted bootstrap method accounts for the possible non-constancy of the variance of returns. The strongest difficulty with resampling schemes, such as bootstrap, is that they may generate data that is less dependent than the original data. The main idea of the weighted bootstrap scheme is to overcome this difficulty²⁰.

Malliaropulos and Priestley (1999) and Cajueiro and Tabak (2006) used this bootstrap method to approximate the sampling distribution of the Lo-MacKinlay VR statistics as well as the Wald statistic of the Cecchetti and Lam (1994) test.

Note that Malliaropulos and Priestley (1999) bootstrapped the VR statistics, which not asymptotically pivotal, under heteroscedasticity in an empirical framework. Their bootstrap tests are not supported by any asymptotic theory or Monte Carlo evidence to evaluate their properties in contrast to the bootstrap tests proposed by Kim (2006).

5 Empirical applications

The VR tests have been widely used and their applications have often covered emerging markets: Asian markets (Kim and Shamsuddin, 2008; Hoque et al., 2007), Eastern European markets (Smith and Ryoo, 2003), African markets (Smith et al., 2002; Al-Khazali et al., 2007; Lagoarde-Segot and Lucey, 2008) and Latin American markets (Chaudhuri and Wu, 2003; Chang et al., 2004). In this section we propose an illustration by examining the RWH for five emerging markets in Latin American, including Argentina, Brazil, Chile, Ecuador and Mexico. We use daily market prices spanning 03 August 1993 to 22 May 2007. All data are collected from Thomson Financial Datastream.

There have been many studies that tested efficiency of Latin American stock markets. However, the results are overall mixed and scattered over studies that employ different sample periods, methods and data frequencies. Urrutia (1995), using the Lo-MacKinlay VR test, rejected the RWH for the Latin American emerging equity markets of Argentina, Brazil, Chile and Mexico, whereas the runs test indicated weak form efficiency over the period 1980:3–1988:12. In contrast, Ojah and Karemera (1999)

²⁰Cribari-Neto and Zarkos (1999) evaluated the performance of this bootstrap methodology by comparing the weighted with the unweighted bootstrap. Their results suggested that weighted bootstrap estimators perform very well, outperforming others estimators, even in the case of heteroscedastic errors and non-normality (fat tails).

found that the Latin American equity returns follow a random walk and were generally weak-form efficient. Grieb and Reyes (1999) reexamined the random walk properties of stocks traded in Brazil and Mexico over the period 1988:12–1995:6, using the Lo-MacKinlay VR tests, and concluded that the Mexican stock market exhibited mean aversion whereas the Brazilian stock market showed a tendency toward random walk. Karemera, Ojah and Cole (1999) also found that Brazil, Chile and Mexico did not follow the random walk under Lo-MacKinlay test, whereas Argentina did, over the period 1987:12–1997:5. However, this result changed when they Chow-Denning’s multiple VR test, showing that Argentina and Brazil followed a random walk. Chaudhuri and Wu (2003) investigated the efficiency for Argentina, Brazil, Chile, Colombia, Mexico and Venezuela over the period 1985:1–1997:2. Using Lo-MacKinlay VR test, they rejected the RWH only for Argentina and Brazil. Chang, Lima and Tabak (2004) rejected the RWH, using Wald-type test (Cecchetti and Lam, 1994), for Argentina, Brazil, Chile and Mexico over the period 1991:1–2004:1.

Table 1 presents summary statistics for the stock returns calculated as the first differences in the logs of the stock price indexes. The data are all leptokurtic as might be expected from daily stocks returns. Three series (Argentina, Brazil and Ecuador) are skewed. To check for nonlinear dependencies, we apply the Lagrange Multiplier test for autoregressive conditional heteroscedasticity (ARCH) on the residuals of the ARMA model, where the lag length is selected based on the Akaike and Schwarz information criterion. This particular specification of heteroscedasticity was motivated by the observation that in many financial time series, the magnitude of residuals appeared to be related to the magnitude of recent residuals. The LM(10) indicates clearly that all stocks show strong conditional heteroscedasticity.

Tables 2 and 3 report the results of various individual and multiple VR tests²¹, respectively, for the five Latin American markets. Since these stock returns exhibit conditional heteroscedasticity, we do not consider the Lo-MacKinlay $M_1(k)$ and Chow-Denning MV_1 tests. More precisely, for individual VR tests we apply the Lo-MacKinlay $M_2(k)$ test as well as the Wright’s R_1 , R_2 and S_2 tests. For multiple VR tests, we apply the Chow-Denning MV_2 test, the Richardson-Smith RS test, the

²¹Some computational resources for VR tests are available from the different authors. Kim (2007) provides various VR tests written in R (<http://www.r-project.org/>) available from <http://www-personal.buseco.monash.edu.au/jaekim/vrtest.htm>. Moreover, a Stata module for the Lo-MacKinlay test written by Baum (2006) is available from <http://fmwww.bc.edu/repec/bocode/l/lomackinlay.ado>.

Whang-Kim subsampling MV_T test, the Belaire-Contreras rank-based $CD(R_1)$ and $CD(R_2)$ tests as well as the Kim's bootstrap MV^* test²²

The holding periods (k) considered are (2, 5, 10, 30). As advocated by Deo and Richardson (2003), we use relatively short holding periods when testing for the mean reversion using VR tests. For the wild bootstrap test (MV^*), as suggested by Kim (2006)²³, the number of bootstrap replications m is set to 1000. As recommended by Whang and Kim (2003), we take a number of block lengths from an equally spaced grid in the interval of $[2.5T^{0.3}, 3.5T^{0.6}]$ for the subsampling test (MV_T).

Some convergence amongst the individual and multiple VR tests is observed for Chile and Mexico. Indeed, the individual and multiple VR tests reject the RWH for these two Latin American markets, indicating that these markets have not been weak-form efficient.

For Argentina, some divergence amongst the individual tests is observed. The statistics of individual tests do not provide the same results. Nevertheless, as shown by Wright (2000), the rank-based VR tests are more powerful than the conventional Lo-MacKinlay and the sign-based VR tests. Thus, in this context, it seems that Argentina does not follow a random walk. Furthermore, when the RWH is generally rejected under the multiple VR tests. Consequently, the exchange rate market of Argentina seems to be inefficiency.

Applying the individual VR tests shows that Brazil follows a random walk. However, this result changes when the multiple VR tests are employed (except for Chen-Deo $QP(k)$ test), showing that, as found by Chang et al. (2004), the Brazilian exchange rate market is not an efficient market. As already noted, conducting individual tests for a number of k values may be misleading as it leads to over rejection of the null hypothesis of a joint test, above the nominal size.

Finally, we found mixed results from the various VR tests for Ecuador. Indeed, the Lo-MacKinlay $M_2(k)$ test is not significant, whereas the rank and sign-based tests as well as Chen-Deo VR_p^β are significant. Furthermore, the Kim's bootstrap and the Whang-Kim's subsampling tests do not reject the RWH while the others multiple VR tests show that Ecuador follows a random walk. Consequently, it is impossible to conclude on the weak-form efficiency for Ecuadorian market.

²²We do not apply the S_1 test suggested by Wright (2001) since it assumes a zero drift which need not be satisfied in practice as well as the multiple signs-based tests developed by Belaire-Franch and Contreras (2004) since the rank-based tests are more powerful.

²³Following Kim (2006), we use the standard normal distribution for η_t to implement the wild bootstrap test. He reports that other choices provided qualitatively similar sample results.

6 Conclusion

This paper reviewed the recent developments in the field of the variance-ratio tests of random walk and martingale hypothesis. In particular, we presented the conventional individual and multiple VR tests as well as their improved modifications based on power- transformed statistics, rank and sign tests, subsampling and bootstrap methods, among others.

We also re-examined the weak-form efficiency for five emerging equity markets in Latin America. We found that Argentina, Brazil, Chile and Mexico follow reject the random walk hypothesis and, consequently, these four Latin American markets are not weak-form efficient. We do not conclude for Ecuador because the results are mixed. We did not deal with the possible presence of structural breaks, due to financial or economic events, which can affect the VR tests. We left this issue to further research which can be conducted even by applying VR tests using a moving subsample window (Yilmaz, 2003; Kim and Shansuddin, 2007) or by modifying VR tests to take into account structural changes (Lee and Kim, 2006).

Appendix: Assumptions on the VR tests

We present some of the main assumptions for the underlying time series which drive the VR tests.

The Lo-MacKinlay (1988), Chow-Denning (1993) and Kim (2006) tests are driven by Assumption H* of Lo and Mackinlay (1988) which assumes

H*1: For all t , $E(\varepsilon_t) = 0$, and $E(\varepsilon_t \varepsilon_{t-\tau}) = 0$ for any $\tau \neq 0$.

H*2: $\{\varepsilon_t\}$ is ϕ -mixing with coefficients $\phi(m)$ of size $r/(2r-1)$ or is α -mixing with coefficients $\alpha(m)$ of size $r/(r-1)$, where $r > 1$, such that for all t and for any $\tau \geq 0$, there exists some $\delta > 0$ for which

$$E|\varepsilon_t \varepsilon_{t-\tau}|^{2(r+\delta)} < \Delta < \infty$$

H*3: $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(\varepsilon_t^2) = \sigma_0^2 < \infty$

H*4: For all t , $E(\varepsilon_t \varepsilon_{t-j} \varepsilon_t \varepsilon_{t-i}) = 0$ for any nonzero j and i , where $j \neq i$

Assumption H*1 is the essential property of the random walk. Assumptions H*2 and H*3 are restrictions on the degree of dependence and heterogeneity which are allowed and yet still permit some form of law of large numbers and central limit theorem to obtain. This allows for a variety of forms of heteroscedasticity, including GARCH-type variances and variances with deterministic changes. Assumption H*4 implies that the sample autocorrelations of ε_t are asymptotically uncorrelated.

The Whang and Kim (2003) test is driven by a relaxed version of Assumption H*. Indeed, they do not impose the restriction that the sample autocorrelation of ε_t are asymptotically uncorrelated (Assumption H*4) by assuming

$$\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(\varepsilon_t^2) = \sigma_0^2$$

Their test is then robust to violations of this assumption.

The Wright (2000) tests are driven by Assumption A in which it is considered that $x_t = \mu + z_t$ and $z_t = \sigma_t \varepsilon_t$. Letting $I_t = \{x_t, x_{t-1}, x_{t-2}, \dots\}$, the assumptions are

A1: z_t is i.i.d.

A2: σ_t and ε_t are independent, conditional on I_{t-1} .

A3: $E(\varepsilon_t | I_{t-1}) = 0$ and $1(\varepsilon_t > 0)$ is an i.i.d. binomial variable that is 1 with probability $\frac{1}{2}$ and 0 otherwise.

Assumption A1 stipulates that the first-differences are i.i.d. while the combination of Assumptions A2 and A3 is sufficient, but not necessary, for x_t to be a martingale difference sequence. Moreover, Assumption A2 is satisfied in a GARCH model and also by a stochastic volatility model in which the innovations to volatility are independent of ε_t . Assumption A3 allows ε_t to be t -distributed with time-varying degrees of freedom. The rank-based tests of Wright (2000) and Belaire-Franch and Contreras (2004) on Assumption A1 while their sign-based tests are based on Assumptions A2 and A3.

Finally, the Chen and Deo's (2006) assumptions on the martingale difference sequence are also different

B1: $\{\varepsilon_t\}$ is ergodic and $E(\varepsilon_t|\mathfrak{F}_t) = 0$ for all t , where $\mathfrak{F}_t$ is a sigma field, ε_t is $\mathfrak{F}_t$ measurable, and $\mathfrak{F}_{t-1} \subset \mathfrak{F}_t$ for all t .

B2: $E(\varepsilon_t^2) = \sigma^2 < \infty$.

B3: For any integer q , $2 \leq q \leq 8$, and for q nonnegative integers s_i , $E\left(\prod_{i=1}^q \varepsilon_{t_i}^{s_i}\right) = 0$ when at least one s_i is exactly one and $\sum_{i=1}^q s_i \leq 8$.

B4: For any integer r , $2 \leq r \leq 4$, and for r nonnegative integers s_i , $E\left(\prod_{i=1}^r \varepsilon_{t_i}^{s_i}|\mathfrak{F}_t\right) = 0$ when at least one s_i is exactly one and $\sum_{i=1}^r s_i \leq 4$, for all $t < t_i$, $i = 1, 2, 3, 4$.

B5: $\lim_{T \rightarrow \infty} \text{Var}\left[E\left(\varepsilon_{t+T}^2 \varepsilon_{t+T+j}^2|\mathfrak{F}_t\right)\right] = 0$ uniformly in j for every $j > 0$.

B6: $\lim_{T \rightarrow \infty} E\left(\varepsilon_t^2 \varepsilon_{t-T}^2\right) = \sigma^4$.

Assumptions B1-B6 allow the innovations ε_t to be a martingale difference sequence with conditional heteroscedasticity. Chen and Deo (2006) showed that the stochastic volatility and GARCH models satisfy Assumptions B1-B6. Assumptions B3-B4 state that the series $\{\varepsilon_t\}$ shows product moment behavior similar to that of an independent white noise process. Assumptions B5-B6 state that ε_t and ε_{t-T} are roughly independent for large lags T .

References

- [1] Al-Khazali O.M., Ding D.K. and Pyun C.S. (2007), "A new variance ratio test of random walk in emerging markets: A revisit", *The Financial Review*, 42, 303-317.
- [2] Andrews D.W.K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation", *Econometrica*, 58, 817-858.
- [3] Andrews D.W.K. and Monahan C. (1992), "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator", *Econometrica*, 60, 953-966.
- [4] Belaire-Franch J. and Contreras D. (2004), "Ranks and signs-based multiple variance ratio tests", Working paper, Department of Economic Analysis, University of Valencia.
- [5] Belaire-Franch J. and Opong K.K. (2005), "Some evidence of random walk behaviour of Euro exchange rates using ranks and signs", *Journal of Banking and Finance*, 29, 1631-1643.
- [6] Brockwell P. and Davis R. (1996) *Time Series: Theory and Methods*, Springer.
- [7] Cajueiro D.O. and Tabak B.M. (2006), "Testing for predictability in equity returns for European transition markets", *Economic System*, 30, 56-78.
- [8] Campbell J.Y. and Mankiw N.G. (1987), "Permanent and transitory components in macroeconomic fluctuations", *American Economic Review*, 77, 111-117.
- [9] Campbell J.Y., Lo A.W. and MacKinlay A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.
- [10] Campbell B. and Dufour J-M. (1997), "Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter", *International Economic Review*, 38, 151-173.
- [11] Cecchetti S.G. and Lam P.S. (1994), "Variance ratio tests: Small-sample properties with an application to international output data", *Journal of Business and Economic Statistics*, 12, 177-186.
- [12] Chang E.J., Lima E.J.A. and Tabak B.M. (2004), "Testing predictability in emerging equity markets", *Emerging Markets Review*, 5, 295-316.

- [13] Chaudhuri K. and Wu Y. (2003), "Random walk versus breaking trend in stock prices: Evidence from emerging markets", *Journal of Banking and Finance*, 27, 575-592.
- [14] Chen W.W. and Deo R.S. (2004), "Power transformations to induce normality and their applications", *Journal of the Royal Statistical Society, Series B*, 66, 117-130.
- [15] Chen W.W. and Deo R.S. (2006), "The variance ratio statistic at large horizons", *Econometric Theory*, 22, 206-234.
- [16] Choi I. (1999), "Testing the random walk hypothesis for real exchange rates", *Journal of Applied Econometrics*, 14, 293-308.
- [17] Chow K.V. and Denning K.C. (1993), "A simple multiple variance ratio test", *Journal of Econometrics*, 58, 385-401.
- [18] Cochrane J.H. (1988), "How big is the random walk in GNP?", *Journal of Political Economy*, 96, 893-920.
- [19] Cogley T. (1990), "International evidence on the size of the random walk in output", *Journal of Political Economy*, 98, 501-518.
- [20] Colletaz G. (2005), "A simple multiple variance-ratio test based on ranks", Working paper, LEO, University of Orléans.
- [21] Cribati-Neto F. and Zarkos S. (1999), "Bootstrap methods for heteroskedastic regression models: Evidence on estimation and testing", *Econometric Review*, 18, 211-228.
- [22] Daniel K. (2001), "The power and size of mean reversion tests", *Journal of Empirical Finance*, 8, 493-535.
- [23] Davidson R. and Flachaire E. (2001), "The wild bootstrap, tamed at last", Working paper No 99, GREQAM, University of Marseille.
- [24] Davidson R., Monticini A. and Peel D. (2007), "Implementing the wild bootstrap using a two-point distribution", *Economics Letters*, 96, 309-315.
- [25] Deo R.S. and Richardson M. (2003), "On the asymptotic power of the variance ratio test", *Econometric Theory*, 19, 231-239.

- [26] Efron B. (1979), "Bootstrap methods: Another look at the jackknife", *Annals of Statistics*, 7, 1-26.
- [27] Fama E. (1970), "Efficient capital markets: A review of theory and empirical work", *Journal of Finance*, 25, 383-417.
- [28] Fama E. (1991), "Efficient capital markets: II", *Journal of Finance*, 46, 1575-1617.
- [29] Faust J. (1992), "When are variance ratio tests for serial dependence optimal?", *Econometrica*, 60, 1215-1226.
- [30] Fleming J., Kirby C. and Ostdiek B. (2006), "Bootstrap tests of multiple inequality restrictions on variance ratios", *Economics Letters*, 91, 343-348.
- [31] Fong W.M., Koh S.K. and Ouliaris S. (1997), "Joint variance-ratio tests of the martingale hypothesis for exchange rates", *Journal of Business and Economic Statistics*, 15, 51-59.
- [32] Grieb T. and Reyes M.G. (1999), "Random walk tests for Latin American equity indices and individual firms", *Journal of Financial Research*, 4, 371-383.
- [33] Hahn G.J. and Hendrickson R.W. (1971), "A table of percentage points of the distribution of the largest value of K Student T variates and its applications", *Biometrika*, 58, 323-332.
- [34] Hochberg Y. (1974), "Some conservative generalizations of the T-method in simultaneous inference", *Journal of Multivariate Analysis*, 4, 224-234.
- [35] Hochberg Y. (1988), "A sharper Bonferroni procedure for multiple tests of significance", *Biometrika*, 75, 800-802.
- [36] Hoque H.A.A.B., Kim J.H. and Pyun C.S. (2007), "A comparison of variance ratio tests of random walk: A case of Asian emerging stock markets", *International Review of Economics and Finance*, 16, 488-502.
- [37] Kan R. (2006), "Exact variance ratio test with overlapping data", Working paper, University of Toronto.
- [38] Karemera D., Ojah K. and Cole J.A. (1999), "Random walks and market efficiency tests: Evidence from emerging equity markets", *Review of Quantitative Finance and Accounting*, 13, 171-188.

- [39] Kim J.H. (2006), "Wild bootstrapping variance ratio tests", *Economics Letters*, 92, 38-43.
- [40] Kim C.J. and Nelson C.R. (1998), "Testing for mean-reversion in heteroskedastic data II: Autocorrelation tests based on Gibbs-sampling-augmented randomization", *Journal of Empirical Finance*, 5, 385-396.
- [41] Kim J.K., Nelson C.R. and Startz R. (1991), "Mean reversion in stock prices. A reappraisal of the empirical evidence", *Review of Economic Studies*, 58, 515-528.
- [42] Kim C.J., Nelson C.R. and Startz R. (1998), "Testing for mean reversion in heteroskedastic data based on Gibbs-sampling-augmented randomization", *Journal of Empirical Finance*, 5, 131-154.
- [43] Kim J.H. and Shamsuddin A. (2008), "Are Asian stock market efficient? Evidence from new multiple variance ratio tests", *Journal of Empirical Finance*, 15, 518-532.
- [44] Lee Y-S. and Kim T-H. (2006), "Spurious rejections by variance ratio tests in the presence of a break in drift", Working paper No 0603, Durham University.
- [45] Lo A.W. and MacKinlay A.C. (1988), "Stock market prices do not follow random walk: Evidence from a simple specification test", *The Review of Financial Studies*, 1, 41-66.
- [46] 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.
- [47] Lagoarde-Segot T. and Lucey B.M. (2007), "Efficiency in emerging markets - Evidence from the MENA region", *International Financial Markets, Institutions and Money*, 18, 94-105.
- [48] Luger R. (2003), "Exact non-parametric tests for a random walk with unknown drift under conditional heteroscedasticity", *Journal of Econometrics*, 115, 259-276.
- [49] MacKinnon J.G. (2002), "Bootstrap inference in econometrics", *Canadian Journal of Economics*, 35, 615-645.
- [50] Malliaropulos D. and Priestley R. (1999), "Mean reversion in Southeast Asian stock markets", *Journal of Empirical Finance*, 6 (4), 355-384.

- [51] Mammen E. (1993), "Bootstrap and wild bootstrap for high dimensional linear models", *Annals of Statistics*, 21, 255-285.
- [52] Newey W. and West K. (1994), "Automatic lag selection in covariance matrix estimation", *Review of Economic Studies*, 61, 631-653.
- [53] Ojah K. and Karemera D. (1999), "Random walks and market efficiency tests of Latin American emerging equity markets: A revisit", *The Financial Review*, 34, 57-72.
- [54] Perron P. and Vodounou C. (2005), "The variance ratio test: An analysis of size and power based on a continuous-time asymptotic framework", *Econometric Theory*, 21, 562-592.
- [55] Politis D.N., Romano J.P. and Wolf M. (1997), "Subsampling for heteroskedastic time series", *Journal of Econometrics*, 81, 281-317.
- [56] Poterba J.M. and Summers L.H. (1988), "Mean reversion in stock prices: Evidence and implications", *Journal of Financial Economics*, 22, 27-59.
- [57] Psaradakis Z. (2000), "p-Value adjustments for multiple tests for nonlinearity", *Studies in Nonlinear Dynamics and Econometrics*, 4, 95-100.
- [58] Richardson M. and Smith T. (1991), "Tests of financial models in the presence of overlapping observations", *Review of Financial Studies*, 4, 227-254.
- [59] Richardson M. and Stock J. (1989), "Drawing inferences from statistics based on multiyear asset returns", *Journal of Financial Economics*, 25, 323-348.
- [60] Shively P.A. (2002), "An exact invariant variance ratio test", *Economics Letters*, 75, 347-353.
- [61] Sidak Z. (1967), "Rectangular confidence regions for the means of multivariate normal distributions", *Journal of the American Statistical Association*, 62, 626-633.
- [62] Smith G. and Ryoo H-J. (2003), "Variance ratio tests of the random walk hypothesis for European emerging stock markets", *The European Journal of Finance*, 9, 290-300.
- [63] Smith G., Jefferis K. and Ryoo H. (2002), "African stock markets: Multiple variance ratio tests of random walks", *Applied Financial Economics*, 12, 475-484.

- [64] Stoline M.R. and Ury H.K. (1979), "Tables of the studentized maximum modulus distribution and an application to multiple comparisons among means", *Technometrics*, 21, 87-93.
- [65] Tse Y.K., Ng K.W. and Zhang X.B. (2004), "A small-sample overlapping variance ratio test", *Journal of Time Series Analysis*, 25, 127-135.
- [66] Urrutia J.L. (1995), "Tests of random walk and market efficiency for Latin American emerging markets", *Journal of Financial Research*, 18, 299-309.
- [67] Whang Y-J. and Kim J. (2003), "A multiple variance ratio test using subsampling", *Economics Letters*, 79, 225-230.
- [68] Wright J.H. (2000), "Alternative variance-ratio tests using ranks and signs", *Journal of Business and Economic Statistics*, 18, 1-9.
- [69] Wu C.F.J. (1986), "Jakknife, bootstrap and other resampling methods in regression analysis", *Annals of Statistics*, 14, 1261-1295.
- [70] Yilmaz K. (2003), "Martingale property of exchange rates and central bank intervention", *Journal of Business and Economic Statistics*, 21, 383-395.

Table 1: Summary statistics of stock returns

	Mean	SD	Skewness	Kurtosis	LM(10)
Argentina	0.000468	0.022095	-0.133875*	8.158586*	402.0338*
Brazil	0.001866	0.025078	0.550350*	13.58173*	365.7680*
Chile	0.000448	0.007172	-0.046965	8.218660*	555.2467*
Ecuador	0.000025	0.018418	1.125265*	44.79396*	225.6864*
Mexico	0.000792	0.016019	0.001551	9.235503*	380.2590*

* Means significant at 1% level, respectively. LM(10) indicates the lagrange multiplier test for conditional heteroscedasticity with 10 lags.

Table 2: Results of individual VR tests

VR tests	k	Argentina	Brazil	Chile	Ecuador	Mexico
$M_2(k)$	2	3.00*	2.22*	8.74*	-1.41	3.60*
	5	1.63	1.61	9.87*	-1.99*	2.05*
	10	1.13	1.13	10.11*	-1.17	1.56
	30	1.45	3.78*	10.66*	0.10	1.44
$R_1(k)$	2	3.34*	1.50	16.96*	1.33	7.31*
	5	2.45*	1.28	17.29*	5.11*	4.64*
	10	1.91*	-2.24*	16.91*	8.61*	3.53*
	30	2.93*	0.56	17.13*	14.55*	4.42*
$R_2(k)$	2	3.10*	1.61	15.87*	1.10	7.04*
	5	2.23*	-1.14	16.31*	4.14*	4.16*
	10	1.61	-2.30*	15.73*	7.73*	3.02*
	30	2.17*	0.31	15.34*	13.85*	3.57*
$S_2(k)$	2	2.09*	0.90	12.79*	7.34*	5.95*
	5	0.96	0.01	12.79*	14.98*	3.58*
	10	0.13	-0.01	12.92*	21.45*	2.37*
	30	1.15	0.41	13.15*	34.78*	2.68*
$VR_p^\beta(k)$	2	1.63**	0.71	0.71	0.31	2.83*
	5	1.45**	-0.25	-0.54	0.29	1.28
	10	0.95	-1.27	-1.46**	1.43**	0.63
	30	1.06	-0.88	-0.12	3.31*	0.60

* and ** Significant at the 5% and 10% level, respectively. We report the VR statistic for each test.

Table 3: Results of multiple VR tests

VR tests	Block length	Argentina	Brazil	Chile	Ecuador	Mexico
MV_2		3.00*	3.78*	10.66*	1.99	3.60*
RS		29.94*	83.91*	380.74*	46.94*	43.87*
$QP(k)$		3.81	6.17	9.63*	20.16*	9.40*
MV^*		0.01*	0.00*	0.00*	0.10	0.00*
$CD(R_1)$		3.69*	4.85*	18.14*	13.65*	7.84*
$CD(R_2)$		4.19*	6.42*	18.31*	11.63*	7.89*
MV_T	93	0.00*	0.00*	0.00*	0.01*	0.00*
	157	0.00*	0.00*	0.00*	0.05	0.00*
	221	0.00*	0.00*	0.00*	0.26	0.00*
	285	0.00*	0.00*	0.00*	0.32	0.00*
	349	0.00*	0.00*	0.00*	0.32	0.00*
	413	0.00*	0.00*	0.00*	0.28	0.00*

* Significant at the 5% level. The p-value are given for the MV^* (Kim, 2006) and MV_T (Belaire and Contreras, 2004) tests whereas the VR statistic is reported for the others tests.