

A Simple Multiple Variance-Ratio Test Based on Ranks^{*}

G. Colletaz

Laboratoire d'Economie d'Orléans, UFR droit-economie-gestion BP6739

45067 ORLEANS Cedex 2

tel : 02 38 41 70 36

email: gilbert.colletaz@univ-orleans.fr

Abstract

Using Chow and Denning's arguments applied to the individual hypothesis test methodology of Wright (2000) I propose a multiple variance-ratio test based on ranks to investigate the hypothesis of no serial correlation. This rank joint test can be exact if data are *i.i.d.*. Some Monte Carlo simulations show that its size distortions are small for observations obeying the martingale hypothesis while not being an *i.i.d.* process. Also, regarding size and power, it compares favorably with other popular tests.

1 Introduction

The random walk hypothesis is important in economics and, particularly in empirical finance and applied macroeconometrics, one is often interested in testing the absence of temporal dependence. A popular approach among practitioners is the variance-ratio analysis. This type of analysis exploits the fact that aggregation of data sampled at various frequencies verifies an interesting property under the *i.i.d.* hypothesis: $(1/k)$ th the variance of a sum of k consecutive observations is equal to the variance of the original series. A test for an individual variance-ratio, i.e. for a given value of k , was derived by Lo and Mc Kinlay (1988) and the extension to a joint test was carried out by Chow and Denning (1993). A drawback of this approach is that the distribution of the test statistic is quite complicated and only an upper bound for the critical

^{*} Expanded version of a talk presented at the XXXVèmes Journées de Statistique (Lyon, June 2003). I am grateful to V. Patilea for comments and suggestions

value is given. The use of an upper bound favors the null hypothesis so that the test might be too conservative¹. This may explain why for example, Gouriéroux and Jasiak (2001, p.28) agree that variance-ratio analysis is less efficient than tests based on empirical autocorrelations. Another drawback is that the asymptotic law is derived under a Gaussian setting. It can be shown that in the class of stable Paretian distributions this asymptotic distribution depends on the characteristic exponent and some Monte Carlo experiments found that the convergence can be extremely slow². However, in a recent study, Wright (2000) used a non parametric approach based on ranks or on signs to examine if an individual variance-ratio is unity. He also performed some Monte Carlo experiments which indicate that they may have greater power than their parametric counterparts. His procedure, though, is not designed to give a joint test for a given number of variance-ratios considered simultaneously which under the null must all be equal to one. In this case a proper test should use a multiple comparison in order to give an overall correct size.

In this paper, I propose a multiple variance-ratio test based on ranks that overcomes the preceding difficulties by merging Wright (2000) and Chow-Denning (1993) approaches. The logic behind this test is easy to understand: if when considering an individual hypothesis the nonparametric Wright's test improves over the parametric test of Lo and MacKinlay, and if according to Chow and Denning we also have an improvement with the use of a parametric multiple test over an individual test, then it can be useful to consider a nonparametric multiple test. Being non-parametric³ this joint test is exact under the *i.i.d.* hypothesis and we can easily approximate its critical values as much as we want. Moreover, Monte Carlo simulations indicate that size distortions are

¹ This issue is also addressed by Whang and Kim (2003) who use a subsampling procedure in order to approximate the asymptotic null distribution of modified versions of Chow and Denning's statistics. However with this approach one has to set the subsample size for a given sample size. As they notice, choosing the subsample size is difficult in practice and important as this selection may affect the properties of their test (for a method to choose this size in a different context see Delgado, Rodriguez-Poo and Wolf (2001)). Moreover their critical values are the asymptotic ones while here we can calculate exact critical values for any finite sample size under the *i.i.d.* hypothesis.

² On these points, see Tse and Zhang (2002). We can also note that Chow and Denning (1993) in their empirical applications do not use critical values associated with the asymptotic distribution but the ones obtained with Monte Carlo experiments.

³ Nonparametric methods and in particular rank tests have been used for many years in time series analysis. See for example the extensive bibliography of Dufour, Lepage and Zeidan (1982) and more recently Hallin and Puri (1992). In particular a rank test was constructed to investigate the presence of serial dependence by Dufour and Roy (1986) leading to a portmanteau statistic and by Breitung and Gouriéroux (1997) to test for the existence of a unit root based on rank counterpart of the Dickey-Fuller statistic, i.e. with a different approach than the one presented here.

small for independent but non identically distributed observations. It also has some power in rejecting the null for non independent observations generated by processes often used empirically for which it compares favorably with other popular tests. In these experiments for comparison purposes I consider the two statistics derived by Chow and Denning, the portmanteau or Q-statistic of Ljung and Box (1978) commonly used to test nullity of autocorrelations, and the Dm test recently proposed by Pena and Rodriguez (2002).

The paper is structured as follows. Section 2 briefly recalls the conventional variance-ratio tests. Section 3 sets out Wright's non parametric procedure and the construction of the joint test. Section 4 examines its size and power under various simulated alternatives. Section 5 concludes.

2 The Parametric Variance-Ratio Approach

The variance-ratio test, introduced by Lo and MacKinlay (1988) and Poterba and Summers (1988) is often used to test the hypothesis that a given time series or its first difference is a collection of independent and identically distributed observations (*i.i.d.*) or that it follows a martingale difference sequence (*m.d.s.*). This test uses the fact that the variance for an *i.i.d.* series increases linearly in each observation interval, that is, the variance of a k -sum is equal to k times the variance of the series, or equivalently that the variance-ratio is equal to one, i.e.,

$$VR(k) = \frac{Var(x_t + x_{t-1} + \cdots + x_{t+k+1})/k}{Var(x_t)} = 1$$

In order to test the *i.i.d.* hypothesis, Lo and MacKinlay (1988) consider statistics based on an estimator of $VR(k)$. For a series with T observations and $k \geq 1$,

$$\widehat{VR}(k) = \frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(1)}, \text{ where}$$

$$\hat{\sigma}^2(k) = \frac{1}{k(T-k+1)(1-k/T)} \sum_k^T (x_t + x_{t-1} + \cdots + x_{t+k+1} - k\hat{\mu})^2, \text{ and}$$

$$\hat{\mu} = \frac{1}{T} \sum_1^T x_t$$

They show that if x_t is *i.i.d.* and under some more weak assumptions then for $k \geq 2$,

$$Z_1(k) = \sqrt{T}(\widehat{VR}(k) - 1) / \sqrt{2(2k-1)(k-1)/3k} \longrightarrow_d N(0, 1)$$

When x_t exhibits heteroscedasticity, Lo and MacKinlay increase the robustness of the test by using White (1980) and White and Domowitz' (1984) arguments and propose the modified statistics⁴ Z_2 ,

$$Z_2(k) = \sqrt{T} \left(\widehat{VR}(k) - 1 \right) \left(\sum_{j=1}^{k-1} \left[\frac{2(k-j)}{k} \right]^2 \delta_j \right)^{-1/2},$$

where,

$$\delta_j = T \left\{ \sum_{t=j+1}^T (x_t - \hat{\mu})^2 (x_{t-j} - \hat{\mu})^2 \right\} / \left\{ \sum_{t=1}^T (x_t - \hat{\mu})^2 \right\}^2.$$

If x_t can be described by a martingale difference sequence, and again with some more assumptions⁵, then Z_2 is asymptotically standard normal.

As stressed by Chow and Denning (1993), these two statistics are appropriate to test an individual variance ratio, i.e. for a given value k . However, under the null hypothesis any variance ratio must be equal to one, so that a more powerful approach is a comparison of all selected variance-ratios with unity. Let k_i be any integer greater than one with $k_i \neq k_j$ for $i \neq j$, Chow and Denning formulate the null hypothesis as $H_0 : VR(k_i) = 1$ for $i = 1, 2, \dots, m$, and define their statistics as

$$Z_1^*(m) = \max_{1 \leq i \leq m} |Z_1(k_i)|,$$

$$Z_2^*(m) = \max_{1 \leq i \leq m} |Z_2(k_i)|$$

In order to control the size of the multiple variance ratio test and because the limit distribution of these statistics is complex, they apply the Sidak (1967) probability inequality, which improves over the Bonferroni inequality, and give an upper bound to the critical values taken in the Studentized Maximum Modulus distribution. The confidence interval of at least $100(1-\alpha)$ percent for these extreme statistics can be defined as $\pm SMM(\alpha, m, \infty)$ and asymptotic critical values can be calculated from the standard normal distribution as $\pm SMM(\alpha, m, \infty) = Z_{\alpha^+/2}$ where $\alpha^+/2 = 1 - (1-\alpha)^{1/m}$. However with finite sample sizes it may be preferable to use critical values obtained by simulations as done by Chow and Denning themselves.

⁴ It 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, 1998a, 1998b) who also propose a solution based on a Bayesian approach and the use of a Gibbs sampler.

⁵ See Lo and MacKinlay (1988) for details.

3 A Multiple Variance-Ratio rank Test extension of the Wright Procedure

Wright (2000) gives four alternatives based on ranks and signs to the parametric variance-ratio tests. Here, according to his own simulations⁶, we will only build upon the test that globally dominates the three others in terms of size or power. Let $r(x_t)$ be the rank of x_t among x_1, x_2, \dots, x_T and the corresponding standardized (zero-mean, unit variance) series r_{1t} given by:

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

He simply substitutes r_{1t} to x_t in the definition of the test statistic Z_1 so that the proposed test statistic is⁷:

$$R_1(k) = \left(\frac{\sum_{k+1}^T (r_{1t} + r_{1t-1} + \dots + r_{1t-k+1})^2}{k \sum_1^T r_{1t}^2} - 1 \right) \times \left(\frac{2(2k-1)(k-1)}{3kT} \right)^{-1/2}$$

By construction under the *i.i.d.* hypothesis $r(x_t)$ is a particular permutation of numbers $1, 2, \dots, T$ each having the same probability of realization, so that $R_1(k)$ has the same distribution as $R_1^*(k)$, where:

$$R_1^*(k) = \left(\frac{\sum_{k+1}^T (r_{1t}^* + r_{1t-1}^* + \dots + r_{1t-k+1}^*)^2}{k \sum_1^T r_{1t}^{*2}} - 1 \right) \times \left(\frac{2(2k-1)(k-1)}{3kT} \right)^{-1/2},$$

and r_{1t}^* is the standardized series obtained with any permutation of $1, 2, \dots, T$. Therefore the exact sampling distribution of $R_1(k)$ may be approximated with a bootstrap method to any desired degree of accuracy by considering the empirical distribution of $R_1^*(k)$; because it is free of nuisance parameters, it can be used to conduct an exact test.

Of course this property of equal probability is not true when there is some conditional heteroscedasticity in x_t even if the martingale independence hypothesis is valid. However, Monte Carlo simulations show that in this case the size distortions of the test are small.

⁶ See Wright (2000) and in particular tables 2 to 7 comparing the size and power of his statistics R_1, R_2, S_1, S_2 . Moreover the sign test depends upon a nuisance parameter namely the presence or the absence of a drift in the random walk.

⁷ Note that, like many others, Wright does not take into account the degree of freedom adjustment present in the consistent estimator of $\hat{\sigma}^2(k)$ derived by Lo and MacKinlay.

Table 1
Some Bootstrap Percentiles of the Exact Distribution ^a

percentiles	ZR			<ZR>			ZR			<ZR>		
	95	2.5	97.5	95	2.5	97.5	95	2.5	97.5	95	2.5	97.5
<i>m</i>	<i>T</i> = 100			<i>T</i> = 500			<i>T</i> = 1000					
2	1.98	-2.15	1.74	1.95	-2.04	1.85	1.97	-2.04	1.91			
3	2.12	-2.27	1.87	2.10	-2.18	2.00	2.13	-2.18	2.06			
4	2.18	-2.30	1.94	2.17	-2.24	2.08	2.20	-2.25	2.14			
5	2.20	-2.32	1.98	2.22	-2.29	2.14	2.25	-2.29	2.19			
10	2.25	-2.33	2.07	2.32	-2.37	2.25	2.35	-2.37	2.31			
20	2.26	-2.33	2.09	2.38	-2.41	2.35	2.41	-2.42	2.39			
40	2.26	-2.33	2.09	2.41	-2.41	2.40	2.46	-2.45	2.46			

a- Critical values are simulated with 50,000 replications in each case

This Wright procedure is in one way similar to that of Lo and MacKinlay as both consider only one variance-ratio at a time and we know that the test of a joint hypothesis is preferable if all selected variance-ratios are equal to unity under the null. Following the suggestions made by Chow and Denning, I propose an extension to the Wright rank variance-ratio methodology to create a multiple rank variance-ratio test. Given the null $H_0 : VR(k) = 1, k = 1, 2, \dots, m$, I consider the test statistic $|ZR(m)|$ given by ⁸:

$$|ZR(m)| = \max_{1 \leq k \leq m} |R_1(k)|$$

Under the *i.i.d.* hypothesis we can not only simulate the distribution of any $R_1(k)$ but also that of $|ZR(m)|$ to any desired degree of accuracy. Again, because there are no nuisance parameters this distribution can be used to construct an exact test. Table 1 gives the 5-percentile of the null distribution of $|ZR(m)|$ for some particular values of T and m ⁹. To take into account an asymmetry in the distribution of this statistic I also give the 2.5-percentile of $\min R_1(k)$ and the 97.5-percentile of $\max R_1(k)$ for $k = 1, \dots, m$. When considering these values I label the test $\langle ZR(m) \rangle$ ¹⁰. It can be seen in this table that asymmetry gradually declines with m but is sensitive for small sample sizes, here $T = 100$, up to $m = 40$.

⁸ In the formulation of the test I consider all variance-ratios corresponding to partial sums with maximal length m rather than an arbitrary chosen subset of these sums.

⁹ All simulations are done with Ox 3.3 and programs are available upon request.

¹⁰ It may be interesting to consider $\langle ZR(m) \rangle$ if the alternative hypothesis is stated in terms of "mean reversion" or "mean aversion" as defined by Kim, Nelson and Startz (1991)

4 Size and Power of the Test

To appreciate the size and power of the multiple variance-ratio rank test I carried out several experiments. Size was investigated under the martingale difference null hypothesis with two constructions that are of interest for real data. Firstly a stochastic volatility model of conditional heteroscedasticity previously used by Lo and MacKinlay (1989) and Wright (2000). Secondly I used series generated by a multi-fractal model, more precisely a random binomial cascade. This construction is able to reproduce the main features of financial prices: scale-consistency, varying volatility with long tails and long memory in the absolute value of returns while at the same time future returns are not predictable from past prices¹¹, i.e. it preserves the *m.d.s.* hypothesis. In the last sub-section, we analyze the power of the test under four hypotheses: stationary $AR(1)$, first differences of $ARMA(1, 1, 1)$, $ARFIMA(0, 1, 0)$, and absolute values of the preceding fractal data. With these experiments we hope to cover a wide range of real data characteristics, that is autocorrelated stationary variables, integrated series and long memory processes with volatility clustering. In addition to the Chow-Denning statistics based on their asymptotical critical values, Z_1^* and Z_2^* henceforth, or on their bootstrapped critical values¹², $Z_{1,BS}^*$ and $Z_{2,BS}^*$ henceforth, I also calculated the commonly used portmanteau test Q of Ljung and Box (1978) designed to investigate nullity of the first m autocorrelations of a time series¹³ and defined as:

$$Q(m) = T(T + 2) \sum_{i=1}^m \frac{r_i^2}{T - i},$$

where $r_i = \frac{\sum_{t=i+1}^T (x_t - \bar{x})(x_{t-i} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}$ and $\bar{x} = \frac{\sum_{t=1}^T x_t}{T}$

No degree of freedom adjustment is needed when x_t is observed and $Q(m)$ is distributed as a chi-square with $df = m$ under the null.

I also considered the statistic \hat{D}_m proposed by Pena and Rodriguez (2002) which, according to their simulations, can be up to 50% more powerful than

¹¹ See for example Mandelbrot, Fisher, Calvet (1997).

¹² Following Chow and Denning (1993), these bootstrap critical values are obtained with simulations under the *i.i.d.* Gaussian null and the heteroscedastic null. I used 50,000 replications for each sample sizes considered.

¹³ Variance ratios and autocorrelations are linked: Lo and MacKinlay (1988) show that their variance ratio is approximately a linear combination of autocorrelation coefficients similar to the Box-Pierce portmanteau statistic.

the Ljung and Box test¹⁴. Under the null hypothesis \hat{D}_m is distributed as a (weighted) sum of m random independent χ_1^2 and they approximate its distribution by a gamma with parameters¹⁵ $\alpha = \frac{3m(m+1)}{4(m+1)(2m+1)}$ and $\beta = \frac{3m(m+1)}{2(m+1)(2m+1)}$. Finally in order to appreciate the advantage of using the multiple rank tests I also compared their results with those obtained from the original Wright's test R_1 .

4.1 The test size

Our model of stochastic volatility, hereafter Model 1, is given by: $x_t = \exp(h_t/2)\epsilon_t$, where $h_t = .95h_{t-1} + \xi_t$ and ξ_t is *i.i.d.* $N(0, 1/10)$ independent from ϵ_t . Two definitions for ϵ_t were successively retained: *i.i.d.* normal, and *i.i.d.* standardized student with 3 df. This last configuration was used to examine the properties of the tests when applied to variables with fat tail distributions, a characteristic often observed on financial data¹⁶. In each case 5000 time series of sample sizes $T=100, 500$ and 1000 were generated. Table 2 reports the results of these Monte Carlo simulations. The probability of type I error was estimated by the percentage of rejections of the null hypothesis¹⁷ using a nominal size of 5%. It can be seen that main conclusions are unaffected by the type of residuals' distributions. Empirical sizes obtained with statistics $Q, Dm, Z_{1,BS}^*$ and to

¹⁴They consider the correlation matrix given by

$$\hat{R}_m = \begin{bmatrix} 1 & r_1 & \cdots & r_m \\ r_1 & 1 & \cdots & r_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_m & r_{m-1} & \cdots & 1 \end{bmatrix}$$

Under $H_0 : r_i = 0, i = 1, 2, \dots, m$, this matrix is an identity matrix and their proposed test statistic is $\hat{D}_m = T \left(1 - |\hat{R}_m|^{1/m}\right)$. Following their recommendations I used Ljung-Box corrected coefficients of autocorrelation $\tilde{r}_i = \sqrt{(T+2)/(T-i)}r_i$ in the construction of the matrix \hat{R}_m .

¹⁵Again, there is no need for degrees of freedom adjustment if data are observed. Pena and Rodriguez give the correction that must be made if data are estimated (e.g. empirical residuals of an *ARMA* filter). Note that this approximation by a gamma distribution is valid for reasonably small values of m (in their paper they used $m_{max} = 36$).

¹⁶This structure corresponds to the Model 2 of Wright (2000).

¹⁷In what follows, m is the number of variance ratios considered including $VR(1)$ which is unity by construction. Therefore tests $|ZR|, \langle ZR \rangle, Z_1^*$ and Z_2^* are informative only for $(m-1)$ variance-ratios. Accordingly, statistics Q and Dm test the nullity of $(m-1)$ autocorrelation coefficients r_1, r_2, \dots, r_{m-1} .

Table 2

Rejection probabilities with the stochastic volatility model^a.

	$T = 100$				$T = 500$				$T = 1000$			
	Distribution: standard normal											
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	7.3	6.9	6.9	6.9	7.6	7.6	7.6	7.7	7.7	7.7	7.6	7.4
$\langle ZR \rangle$	7.4	7.1	7.1	7.1	7.4	7.5	7.7	7.7	7.9	7.7	7.6	7.4
Z_1^*	8.8	7.3	6.3	5.7	13.8	11.5	10.1	8.4	16.6	13.5	11.1	9.4
Z_2^*	3.0	3.0	3.2	3.6	2.5	2.0	1.6	1.7	2.6	2.0	1.5	1.2
$Z_{1,BS}^*$	12.4	12.0	10.3	8.9	19.6	19.4	19.0	18.0	22.9	23.2	22.8	21.7
$Z_{2,BS}^*$	5.2	4.9	4.8	4.8	4.8	4.6	4.8	5.1	4.6	5.1	5.3	5.1
Q	16.0	18.4	17.4	12.7	29.1	36.7	40.1	35.7	34.2	44.7	50.0	45.8
Dm	15.2	16.8	15.0	7.8	26.4	34.3	39.6	37.8	31.1	41.6	50.5	51.7
R_1	6.7	6.1	5.8	5.5	6.8	6.5	6.3	6.2	7.2	7.4	6.6	6.4
	Distribution: standardized Student, df=3											
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	6.3	6.5	6.5	6.5	6.9	6.8	6.7	6.7	6.5	6.9	7.0	6.5
$\langle ZR \rangle$	6.6	6.7	6.8	6.8	6.8	6.9	6.7	6.7	6.5	6.8	7.0	6.6
Z_1^*	7.2	5.9	5.4	5.1	12.5	10.4	8.5	7.0	12.6	10.3	8.5	7.3
Z_2^*	3.5	3.2	3.6	4.2	2.4	1.9	1.5	1.7	2.2	1.3	1.1	1.0
$Z_{1,BS}^*$	10.4	10.0	8.8	8.0	17.7	17.7	17.5	16.4	17.6	18.0	18.4	18.1
$Z_{2,BS}^*$	5.4	5.4	5.4	5.7	4.7	5.4	5.3	5.3	4.5	4.1	4.1	4.3
Q	12.5	14.7	13.5	9.5	23.9	30.4	32.2	29.0	26.4	34.5	38.9	36.3
Dm	12.0	13.7	11.5	5.8	22.9	28.5	33.1	30.6	24.2	32.6	38.9	40.3
R_1	5.9	5.8	5.6	4.7	6.6	6.2	6.0	5.2	6.3	6.6	6.5	6.3

a- This table gives the simulated percentage size of the tests based on 5,000 replications of Model 1. Nominal size is 5%.

a lesser extent with Z_1^* largely overestimate the nominal one and these discrepancies tend to increase with the sample size. Of course these results imply that the power of these various statistics has an ambiguous interpretation. With R_1 , $|ZR|$ and $\langle ZR \rangle$ differences between nominal and empirical sizes are always positive but comparatively smaller, never over 7.9%. In contrast empirical sizes associated with Z_2^* have a strong tendency toward underestimation but the use of bootstrap critical values provides a clear improvement and $Z_{2,BS}^*$ does remarkably well in recovering the nominal size.

Model 2 is a binomial cascade model. The cascade begins by assigning uniform probability to the interval $[0, 1]$. In the first step, this interval is split into two subintervals of equal length, assigning a mass m_0 on $[0, 1/2]$ with probability p_1 and $(1 - m_0)$ on $[1/2, 1]$ with probability $(1 - p_1)$. This process is then repeated on each newly created interval, probability p_1 being chosen randomly at each step. The *cdf* of the resulting multifractal measure, $\theta(t)$, is used to define a random trading time thus allowing variations in volatility. The resulting price process is defined by $P_t = \exp(B_H[\theta(t)])$ where $B_H[t]$ is a fractional

Table 3

Rejection probabilities with differences of multifractal series^a.

m	$T = 100$				$T = 500$				$T = 1000$			
	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	7.6	7.3	7.3	7.3	8.8	8.9	8.8	8.7	7.9	8.4	8.5	8.4
$\langle ZR \rangle$	8.0	8.1	7.9	8.0	8.9	8.8	8.9	8.8	7.8	8.4	8.6	8.4
Z_1^*	12.4	9.2	7.7	6.2	22.7	19.4	16.2	13.6	26.4	23.5	20.6	18.0
Z_2^*	2.5	1.7	1.8	1.9	2.5	1.3	1.0	0.8	1.8	1.0	0.8	0.7
$Z_{1,BS}^*$	17.0	15.1	12.2	10.1	28.8	29.5	28.6	26.2	33.2	34.2	33.9	32.6
$Z_{2,BS}^*$	3.8	3.5	2.8	2.7	4.4	4.1	3.7	3.2	3.7	3.4	3.1	2.7
Q	22.5	24.6	23.3	13.0	41.6	53.5	61.1	64.3	47.7	60.3	70.6	76.4
Dm	23.0	24.9	22.5	12.1	39.7	50.5	58.9	64.1	45.2	57.4	68.5	75.4
R_1	6.5	5.9	5.9	5.2	8.0	7.3	6.9	6.3	7.6	7.4	7.3	7.1

a- This table gives the simulated percentage size of the tests based on 5,000 replications of Model 2. Nominal size is 5%.

Brownian Motion with self-affinity index H . In the simulations I used a standard Brownian Motion and retained the first differences of P_t as the working series¹⁸. By construction, these returns are nonautocorrelated but have long memory in their absolute values and conditional heteroscedasticity, i.e. they are not *i.i.d.* but the martingale difference sequence hypothesis is true.

Results obtained with these complex series are given in Table 3. They are globally similar to those derived under model 1. While Z_2^* exhibits some under-estimation, empirical sizes tend to be much higher than the nominal size of 5% with Q , Dm , Z_1^* and $Z_{1,BS}^*$. We also observe with $Z_{2,BS}^*$ some under-evaluation which increases with the number of variance ratios, m . For R_1 , $|ZR|$ and $\langle ZR \rangle$ differences between empirical and nominal sizes are constantly positive leading to an over-rejection rate of the true null hypothesis but the estimated rejection probability of R_1 is never greater than 8%, and the rejection probabilities of $|ZR|$ and $\langle ZR \rangle$ are never greater than 8.9%. Finally, as in the preceding experiment, we do not notice any major difference between the estimated sizes of $|ZR|$ and $\langle ZR \rangle$.

Results of these first experiments confirm the conclusion given by Wright (2000): the rank-based variance-ratio tests do not seem to suffer serious size distortion in the presence of conditional heteroscedasticity.

¹⁸ In the experiments, the mass m_0 was selected randomly with uniform probability on $[0.60, 0.75]$ for each of the 5000 simulated series.

4.2 The test power

In this section the results of several simulations using autocorrelated data with or without conditional heteroscedasticity are presented in order to illustrate the power of the new statistics. For this exercise only the two other non-size deficient statistics R_1 and $Z_{2,BS}^*$ are retained for comparison purposes¹⁹. Four models were successively considered. We expect that they are different enough to cover a wide range of characteristics present in real data:

- Model 3: data are generated by the following stationary first-order autoregressive process²⁰ $x_t = 0.10x_{t-1} + u_t$, with two variants:
 - homoscedastic residuals: u_t is *i.i.d.*, standard normal,
 - heteroscedastic residuals: $u_t = \exp(h_t/2)\epsilon_t$ where h_t defined as in Model 1.
- Model 4: $x_t = (1 - L)y_t$, where y_t is driven by an *ARIMA*(1, 1, 1) mean-reverting process considered by Summers (1986) and is the sum of a stationary *AR*(1), $w_t = \phi w_{t-1} + \epsilon_t$, and a random walk, $z_t = z_{t-1} + \tau_t$, where ϵ_t and τ_t are *i.i.d.* normal with variances of 1.0 and 0.5 respectively. It was used by Chow and Denning (1993) who noticed that when ϕ is close to one then autocorrelations are negative and small in the short horizon so that the mean-reversion only occurs over very long periods. Given this characteristic, I expect that the power of the tests increases with the number of variance-ratios considered, k . Parameter ϕ takes two values: 0.85 and 0.96.
- Model 5: x_t is generated by an *ARFIMA*(0, 1, 0), $x_t = (1 - L)^d u_t$, where u_t is *i.i.d.*, standard normal. The aim is to assess the capability of the tests to detect the long memory present in the data²¹. Moreover, parameter d was given two values $d = 0.1$ or $d = -0.1$ in order to examine the sensitivity of the results to the sign of these long term correlations.
- Model 6: x_t is given by the absolute values²² of fractal series considered in Model 2 above. We know that these series verify the *m.d.s.* hypothesis but have long memory in their higher moments so that the tests should reject the null hypothesis when they are considered in absolute terms.

¹⁹Note also that the other statistics which according to the preceding experiments were size-deficient do not have higher rates of rejection when the null is false. Complete results are available upon request.

²⁰This structure corresponds to Model 3 in Wright (2000).

²¹This is (without heteroscedasticity) Wright's Model 4. Data generation was carried out with the Ox instruction `diffpow`. For this model I discarded the first 5000 generated observations as the "burnin" while for other models I discard the first 50 simulated values.

²²Absolute values are often used to disentangle linear dependence and nonlinear dependence in financial time series and have also been taken as a measure of volatility, see for example Granger and Ding (1995).

Table 4
Rejection Probabilities with $AR(1)$ hypothesis, $\phi = 0.1^a$

	$T = 100$				$T = 500$				$T = 1000$			
Distribution: homoscedastic normal												
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	10.6	10.6	10.7	10.7	47.6	45.1	43.5	42.8	78.1	76.2	74.6	73.2
$\langle ZR \rangle$	14.7	14.0	13.7	13.7	51.1	47.9	44.8	43.1	79.9	77.3	75.3	73.1
$Z_{2,BS}^*$	17.0	16.4	14.8	12.7	55.8	52.4	49.8	46.4	83.9	82.0	80.4	78.2
R_1	8.7	6.2	5.4	3.0	33.1	19.6	11.5	7.6	58.4	35.7	20.0	12.0
Distribution: heteroscedastic normal												
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	11.9	12.2	12.3	12.3	47.1	44.9	43.7	42.9	75.2	73.1	71.6	70.0
$\langle ZR \rangle$	16.2	15.2	15.2	15.1	50.0	47.4	45.0	43.2	76.5	74.2	72.3	69.9
$Z_{2,BS}^*$	12.9	12.7	12.2	11.1	33.2	30.9	28.4	26.3	51.7	49.2	46.7	43.9
R_1	10.1	7.9	5.9	3.5	32.8	20.3	12.2	8.6	55.5	34.6	20.9	13.1
Distribution: heteroscedastic student, df=3												
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	15.1	15.2	15.3	15.3	62.1	59.6	57.8	57.1	89.4	87.8	86.6	85.7
$\langle ZR \rangle$	19.9	19.0	18.6	18.6	65.0	62.1	59.2	57.3	90.5	88.6	87.2	85.6
$Z_{2,BS}^*$	15.7	15.3	13.7	12.2	36.6	34.1	32.2	29.8	57.2	54.6	52.8	49.8
R_1	11.7	8.9	5.6	3.0	43.6	26.3	15.9	10.3	71.1	46.5	27.5	15.6

a- This table gives the simulated percentage power of the tests based on 5,000 replications of Model 3. Nominal size is 5%.

Results for Model 3 are given in Table 4. For a sample size of 100 the asymmetric statistic $\langle ZR \rangle$ is always more powerful than $|ZR|$ but this difference becomes practically negligible when the sample size is increased to $T = 500$ or $T = 1000$. Rates of rejection associated with the parametric test $Z_{2,BS}^*$ are one to five points higher than those obtained with the two multiple rank statistics when residuals are homoscedastic but are remarkably lower when heteroscedasticity is present and specially when sample sizes are large (for example the discrepancies are between 20 and 30 points when $T = 1000$). Clearly with this experiment and among the statistics which do not show strong deviations in size, $|ZR|$ and $\langle ZR \rangle$ must be preferred to $Z_{2,BS}^*$ when heteroscedasticity is suspected. It can be seen that Wright's test R_1 has by far the lowest rate of rejection among the four statistics. With this experiment there is a clear advantage to consider the multiple version of the ranks based test. Another point deserves some care: for a given sample size and for all statistics the rejection rate decreases with the number of variance-ratios considered. Of course such an evolution was expected as the simulated $AR(1)$ process is specially useful to modelize short term dependencies. However this reduction of power with m is also generally the lowest with our two multiple variance-ratios tests based on ranks.

Table 5
Rejection Probabilities with $ARIMA(1, 1, 1)$ hypothesis ^a

	$T = 100$				$T = 500$				$T = 1000$			
	$\phi = 0.85$											
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	9.4	9.1	8.9	8.9	36.9	49.3	55.1	54.2	65.5	84.8	92.0	92.2
$\langle ZR \rangle$	7.1	7.4	7.4	7.4	34.0	46.9	53.3	53.9	63.7	84.0	91.6	92.4
$Z_{2,BS}^*$	5.4	3.6	1.7	0.6	33.8	42.9	43.3	36.3	66.9	86.0	92.3	91.4
R_1	10.4	12.6	13.8	13.7	44.6	63.5	73.3	64.4	75.2	93.7	97.3	94.9
	$\phi = 0.96$											
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	5.4	5.6	5.5	5.5	6.5	7.1	7.2	6.9	8.8	10.8	14.3	16.1
$\langle ZR \rangle$	5.1	5.1	5.1	5.1	5.9	6.5	6.8	6.9	7.9	10.2	13.9	16.3
$Z_{2,BS}^*$	4.4	3.7	2.8	2.5	5.5	5.1	4.3	3.2	7.7	9.4	11.0	10.1
R_1	5.0	5.3	5.9	6.3	7.1	9.1	11.7	15.7	10.7	16.9	24.9	34.5

a- This table gives the simulated percentage power of the tests based on 5,000 replications of Model 4. Nominal size is 5%.

From this point of view, Model 4 is totally different as it implies small autocorrelations in the short term and a mean-reversion occurring only over a long period especially when the parameter ϕ of the AR component is close to one. Of course this feature is unfavorable to the multiple variance-ratios tests because rejection of the null hypothesis is hard to detect for small values of k , $k = 1, 2, \dots, m$. As can be seen in Table 5, this is precisely what happens: R_1 is the best test of the four considered even if the estimated power is generally growing with the number of variance-ratios for all tests. However, among the multiple tests, the rank based ones dominate the Chow and Denning statistic $Z_{2,BS}^*$ specially for small and medium sample sizes. Note that the estimated power declines substantially when ϕ is very close to unity but with Chow and Denning, we can doubt that any test having a good size will have a lot of power in such a case.

Results obtained with the fractionally integrated model 5 are reproduced in Table 6. We note that all tests are sensitive to the sign of fractional parameter d but that this sensitivity is the lowest for our two rank based statistics. In particular even if they are dominated by the Lo and Denning' statistic when the fractional parameter is positive the reverse is true and more pronounced when d is negative, the power of $Z_{2,BS}^*$ being very low when $T = 100$. As with preceding experiments, the two versions of our multiple rank test seem practically equivalent except when $T = 100$ but unfortunately their ranking depends on the sign of the fractional parameter ²³. We also note that the mul-

²³ note however that all the rejections are observed for $\langle ZR \rangle$ being greater than its upper bound when d is positive, and lower than its lower bound in the opposite

Table 6
Rejection Probabilities with $ARFIMA(0, d, 0)$ hypothesis^a

	$T = 100$				$T = 500$				$T = 1000$			
d = -0.10, distribution: standard normal												
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	14.8	16.3	16.7	16.7	69.1	70.2	70.6	70.7	93.9	95.7	94.6	94.6
$\langle ZR \rangle$	20.0	20.6	20.6	20.6	71.2	72.4	71.5	70.8	94.3	95.0	94.8	94.5
$Z_{2,BS}^*$	24.8	27.0	26.5	24.8	76.0	77.4	77.5	77.1	96.1	96.7	96.7	96.5
R_1	16.0	13.8	10.4	4.1	69.4	63.9	53.9	41.0	93.9	92.1	85.1	72.5
d = -0.10, distribution: student, df=3												
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	21.5	22.9	23.2	23.3	84.1	84.9	84.8	85.1	98.7	98.7	98.8	98.7
$\langle ZR \rangle$	26.5	27.0	26.9	26.9	85.6	86.0	85.4	85.2	98.8	98.8	98.8	98.7
$Z_{2,BS}^*$	25.8	28.8	28.1	26.7	75.9	78.3	78.4	78.0	95.3	96.5	96.5	96.2
R_1	21.9	19.3	13.1	4.9	83.7	79.5	69.3	54.4	98.8	97.9	94.4	85.8
d = 0.10, distribution: standard normal												
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	16.6	15.7	15.5	15.5	62.9	64.1	63.1	61.8	90.2	91.7	91.7	91.0
$\langle ZR \rangle$	12.9	13.1	13.1	13.1	59.9	62.0	61.7	61.5	89.3	90.9	91.5	91.2
$Z_{2,BS}^*$	9.8	6.9	3.7	1.9	61.0	59.7	55.6	49.8	91.8	92.5	92.0	90.5
R_1	16.7	15.6	13.6	12.7	65.6	61.3	51.9	39.4	92.1	91.2	84.6	73.2
d = 0.10, distribution: student, df=3												
m	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	19.7	18.9	18.7	18.6	72.2	73.8	72.1	71.5	95.1	96.3	96.4	96.1
$\langle ZR \rangle$	16.0	16.1	16.0	16.0	69.5	71.2	71.3	71.2	94.6	95.9	96.3	96.2
$Z_{2,BS}^*$	9.7	6.3	3.5	1.9	58.4	57.8	53.0	46.5	88.1	89.6	89.3	87.3
R_1	18.8	17.4	14.6	14.2	75.0	70.3	59.00	45.1	96.1	95.9	91.0	78.8

a- This table gives the simulated percentage power of the tests based on 5,000 replications of Model 5. Nominal size is 5%.

multiple rank tests generally dominate the Wright's test when sample sizes are small or medium, especially when d is negative and $m \geq 10$.

Finally, results associated with absolute values of fractal series already used in Model 2 are given in Table 7. Here all tests reject the *i.i.d.* or *m.d.s.* hypothesis with very low type II error for large sample sizes. However for small sample size and relatively large values of m multiple variance-ratio statistics are preferred over the individual variance-ratio test of Wright. Moreover we have seen that other statistics were notably size-deficient when the same data were not transformed²⁴ so that if one is interested in discriminating between linear and nonlinear dependence, an alternative that may be important in some cases,

case. It thus has some value to detect the sign of the dependence.

²⁴ See Table 3 above.

Table 7
Rejection Probabilities with absolute values of multifractal series^a

m	$T = 100$				$T = 500$				$T = 1000$			
	5	10	20	40	5	10	20	40	5	10	20	40
$ ZR $	56.6	61.0	62.0	62.0	99.1	99.6	99.7	99.7	100	100	100	100
$\langle ZR \rangle$	61.5	64.5	65.0	64.9	99.3	99.6	99.7	99.7	100	100	100	100
$Z_{2,BS}^*$	60.1	68.0	70.5	71.4	99.5	99.8	99.8	99.9	100	100	100	100
R_1	59.1	61.7	55.2	35.5	99.4	99.7	99.7	99.1	100	100	100	100

a- This table gives the simulated percentage power of the tests based on 5,000 replications of Model 2. Nominal size is 5%.

e.g. in risk management or portfolio selection, it may be worthwhile to consider these multiple variance-ratio tests.

5 Conclusion

In this research I examine the properties of a joint variance-ratio test based on ranks. This non parametric statistic lead to an exact test under the *i.i.d.* hypothesis but Monte Carlo simulations seem to indicate that its empirical size stays near the theoretical one when only a martingale sequence hypothesis is verified, and in particular when data are fat-tailed and have dependencies in their higher moments. For the simulated models considered, other tests commonly used suffer much more noticeable size distortions with the exception of the robust variance-ratio test proposed by Chow and Denning when used with bootstrap critical values. In particular our results do not support the generality of the view expressed by Gouriéroux and Jasiak (2001) about the superiority of tests based on empirical correlations. For the experiments illustrating a false null hypothesis, rates of rejection of the based rank tests are similar to and often higher than those of the deficient size statistics. Moreover in most of the cases considered here it clearly dominate the preceding robustified statistics which appeared to be much more dependent on some characteristics of the series, namely heteroscedasticity, deviations from normality and the sign of dependencies. Finally it is important to remember that our test supposes that the working series is an observed one. Some other non reported simulations using Monti's approach (1994) show that when this series is estimated then size and power are dramatically affected²⁵. Accordingly, it cannot be used to test the adequation of an empirical model by considering the properties of its

²⁵ Typically, Monti's approach consists in simulating various $ARMA(p, q)$ processes with $(p, q) \neq (1, 0)$ in order to detect linear dependencies in residuals of a misspecified $AR(1)$ filter.

estimated residuals²⁶.

References

- [1] Breitung, J. and C. Gourieroux, 1997, Rank tests for unit roots. *Journal of Econometrics* 81, 7-27.
- [2] Chow, K.V. and K.C. Denning, 1993, A simple multiple variance ratio test. *Journal of Econometrics* 58, 385-401.
- [3] Delgado, M., J.M. Rodriguez-Poo and M. Wolf, 2001, Subsampling inference in cube-root asymptotics with an application to Manski's maximum score estimator. *Economics Letters* 73, 241-250.
- [4] Dufour, J.-M., 1981, Rank test for serial dependence. *Journal of Time series Analysis*, 2, pp. 117-128.
- [5] Dufour, J.-M., Y. Lepage and H. Zeidan, 1982, Nonparametric testing for time series: A Bibliography. *Canadian Journal of Statistics*, 10, pp. 1-38.
- [6] Dufour, j.-M. and R. Roy, 1986, Generalized Portmanteau Statistics and tests of randomness. *Communications in statistics, Theory and Methods*, 15, pp. 2953-2972.
- [7] Gourieroux, C. and J. Jasiak, 2001, *Financial Econometrics - problems, Models and Methods*. Princeton University Press, Princeton and Oxford.
- [8] Granger, C.W.J. and Z. Ding, 1995, Some properties of absolute returns: An alternative measure of risk. *Annales d'Economie et de Statistique*, 40, 67-91.
- [9] Hallin, M. and M.L. Puri, 1992, Rank tests for time series analysis: A survey, in: D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M.S. Taqqu, (Eds.), *New Directions in Time Series Analysis, Part I*, vol. 45 of the IMA Volumes in Mathematics and its Applications, pp. 111-153. Springer-Verlag, New York.
- [10] Kim, J.K., C.R. Nelson and R. Startz, 1991, Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence. *Review of Economic Studies* 58, 515-528.
- [11] Kim, C.J., C.R. Nelson and R. Startz, 1998a, Testing for mean reversion in heteroskedastic data based on Gibbs-sampling-augmented randomization. *Journal of Empirical Finance* 5, 131-154.

²⁶ A possible explanation is that the addition of an estimation error to an observed series leads to more marked differences between the ranks of the true and the estimated series than between their respective autocorrelation sequences.

- [12] Kim, C.J. and C.R. Nelson, 1998b, Testing for mean-reversion in heteroscedastic data II: Autocorrelation tests based on Gibbs-sampling augmented randomization. *Journal of Empirical Finance* 5, 385-396.
- [13] Liu, C.Y. and J. He, 1991, A variance ratio test of random walks in foreign exchange rates. *Journal of Finance* 46, 773-785.
- [14] Lo, A.W. and A.C. MacKinlay, 1988, Stock prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies* 1, 41-66.
- [15] Lo, A. W. and A.C. MacKinlay, 1989, The size and power of the variance ratio test in finite samples. *Journal of econometrics* 40, 203-238.
- [16] Mandelbrot, B., A. Fisher and L. Calvet, 1997, A multifractal model of asset returns. Cowles Foundation Discussion Paper #1164.
- [17] Monti, A.C., 1994, A proposal for residual autocorrelation test in linear models. *Biometrika* 81, 776-780.
- [18] Poterba, J.M. and L.H. Summers, 1988, Mean reversion in stock prices: evidence and implications. *Journal of Financial Economics* 22, 27-59.
- [19] Sidak, Z., 1967, Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* 62, 626-633.
- [20] Summers, L.H., 1986, Does the stock market rationally reflect fundamentals values?. *Journal of Finance* 41, 591-600.
- [21] Tse, Y.K. and X.B. Zhang, 2002, The variance ratio test with stable paretian errors. *Journal of time Series Analysis* 23, 117-126.
- [22] Whang, Y.J., and J. Kim, 2003, A multiple variance ratio test using subsampling. *Economics Letters* 79, 225-230.
- [23] White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* 48, 817-838.
- [24] White, H. and I. Domowitz, 1984, Nonlinear regression with dependent observations. *Econometrica* 52, 143-161.
- [25] Wright, J.H., 2000, Alternative variance-ratio tests using ranks and signs. *Journal of Business & Economic Statistics* 18, 1-9.