

A Small-Sample Overlapping Variance-Ratio Test

Y.K. Tse

Department of Economics
National University of Singapore
ecstseyk@nus.edu.sg

K.W. Ng

Department of Statistics and Actuarial Science
University of Hong Kong
hrntnkwhk@hkucc.hku.hk

X.B. Zhang

Department of Economics
National University of Singapore
ecszxb@nus.edu.sg

March 2001

Abstract: The null distribution of the overlapping variance-ratio (OVR) test of the random-walk hypothesis is known to be downward biased and skewed to the right in small samples. As shown by Lo and MacKinlay (1989), the test under-rejects the null on the left tail seriously when the sample size is small. This unfortunate property adversely affects the applicability of the OVR test to macroeconomic time series, which usually have rather small samples. In this paper we propose a modified overlapping variance-ratio statistic and derive its exact mean under the normality assumption. We propose to approximate the small-sample distribution of the modified statistic using a Beta distribution that matches the (exact) mean and the (asymptotic) variance. A Monte Carlo experiment shows that the Beta approximation performs well in small samples.

Key Words: Beta distribution, Monte Carlo experiment, random-walk hypothesis, variance-ratio test

1 Introduction

Since the works of Cochrane (1988) and Lo and MacKinlay (1988, 1989) the variance-ratio (VR) statistic has been used widely as a test for the random-walk hypothesis. Campbell and Mankiw (1987a, 1987b, 1989), Cogley (1990) and Poterba and Summers (1988) used the VR statistic to measure the persistence in economic time series. Lo and MacKinlay (1989) demonstrated that the VR test is more powerful than either the Dickey-Fuller test or the Box-Pierce Q test for several interesting alternatives. Cecchetti and Lam (1994) summarized the advantages of using the VR test.

Lo and MacKinlay (1988) provided the asymptotic theory for the overlapping VR (OVR) statistic. Recently, Zhang and Tian (1999) derived the exact finite-sample distribution of the nonoverlapping VR (NVR) statistic, which follows a Beta distribution. As argued by Lo and MacKinlay (1989), the OVR test is expected to have higher power than the NVR test. This result was confirmed by their Monte Carlo experiment. However, Lo and MacKinlay (1989) found that the asymptotic OVR test seriously under-rejects the null on the lower tail and over-rejects the null on the upper tail in small and moderate samples. This result adversely affects the use of the OVR in practical applications, especially in macroeconomic studies. To overcome this problem, Lo and MacKinlay estimated the critical values of the OVR statistic for various sample sizes and differencing intervals. The use of this table, however, is rather limited as it is difficult to interpolate for other cases not found in the table. In this paper, we suggest an analytical method to calculate the approximate critical values of the OVR statistic. We propose a circulant OVR (COVR) statistic and derive its exact mean. The distribution of the COVR statistic is then approximated using a Beta distribution that matches the exact mean and the asymptotic variance. Our Monte Carlo experiment shows that the approximation works quite well when the differencing interval is up to about one sixth of the sample size.

The plan of this paper is as follows. In Section 2 we review the use of the OVR

statistic. We propose the COVR statistic in Section 3 and derive its exact mean. A method for approximating the distribution of the COVR statistic is suggested. In Section 4 we present some Monte Carlo results for the suggested test. Section 5 concludes the paper.

2 The OVR Statistic

Let $\{X_t\}$ denote a time series generated from the following equation

$$X_t = \mu + X_{t-1} + \varepsilon_t. \quad (1)$$

We denote H_0 as the null hypothesis that ε_t are independently and identically distributed (IID) as Gaussian variates with mean zero and variance σ^2 . That is, $H_0 : \varepsilon_t \sim \text{IID } N(0, \sigma^2)$. Assuming the data consist of $kq + 1$ observations X_0, X_1, \dots, X_{kq} , where k and q are arbitrary integers, we define (let $n = kq$)

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n (X_t - X_{t-1}) = \frac{1}{n} (X_n - X_0)$$

and

$$\hat{\sigma}^2 = \frac{1}{(n-1)} \sum_{t=1}^n (X_t - X_{t-1} - \hat{\mu})^2,$$

which is an unbiased estimate of σ^2 based on the one-period difference $\Delta X_t = X_t - X_{t-1}$. Alternatively we may consider the q -period difference of X_t , $X_t - X_{t-q}$, from which we obtain an estimate of the variance of the q -period difference as

$$\hat{\sigma}_q^2 = \frac{k}{(n-q+1)(k-1)} \sum_{t=q}^n (X_t - X_{t-q} - q\hat{\mu})^2.$$

Note that the formula above has taken account of the degrees-of-freedom adjustment suggested by Lo and MacKinlay (1989). Under H_0 , the variance of the q -period difference of X_t is $q\sigma^2$. Thus, an unbiased estimate of σ^2 can be obtained using $\hat{\sigma}_q^2/q$. A test for the random-walk hypothesis H_0 can be constructed by considering the centered ratio

$$M = \frac{\hat{\sigma}_q^2}{q\hat{\sigma}^2} - 1.$$

Lo and MacKinlay (1988) showed that under H_0 ,

$$\sqrt{n} M \xrightarrow{D} N\left(0, \frac{2(2q-1)(q-1)}{3q}\right),$$

where \xrightarrow{D} denotes convergence in distribution as $n \rightarrow \infty$ keeping q fixed. Denoting

$$V = \frac{2(2q-1)(q-1)}{3nq}, \quad (2)$$

the standardized OVR statistic is defined as

$$R = \frac{M}{\sqrt{V}}, \quad (3)$$

which is asymptotically distributed as $N(0, 1)$ under H_0 .

Lo and MacKinlay (1989) reported some results on the finite-sample distribution of R . They found that the statistic grossly under-rejects the null on the lower tail. Although the total empirical size of a two-sided test still approximates quite well to the nominal size,¹ there is a concern for loss in power. As the rejection of the null often comes from the lower tail, there is a serious loss in power when the lower-tail critical value is downward biased. This bias is a result of the null distribution being skewed to the right, which induces errors when the standard normal distribution is used as the asymptotic approximation. To provide a useful test in small samples Lo and MacKinlay estimated the critical values for various values of n and q using large-scale Monte Carlo runs. The use of the table of critical values is, however, limited by the cases simulated. In the next section we propose an analytical approximation to the critical values of a modified OVR statistic.

3 The COVR Statistic

Note that $\hat{\sigma}_q^2$ consists of overlapping differences. When the numerator of the variance ratio is based on the sum of k nonoverlapping q -period differences we obtain the (uncentered) NVR statistic R^* as

$$R^* = \frac{\sum_{t=1}^k (X_{qt} - X_{q(t-1)} - q\hat{\mu})^2}{q \sum_{t=1}^n (X_t - X_{t-1} - \hat{\mu})^2}.$$

¹The under-rejection in the lower tail is compensated by an over-rejection in the upper tail.

In a recent article, Zhang and Tian (1999) derived the finite-sample distribution of R^* .²

They showed that

$$R^* \sim \text{Beta} \left(\frac{k-1}{2}, \frac{k(q-1)}{2} \right).$$

If we denote

$$u_t = \frac{\Delta X_t - \hat{\mu}}{[\sum_{t=1}^n (\Delta X_t - \hat{\mu})^2]^{1/2}}, \quad t = 1, \dots, n,$$

and $r_t = \Delta X_t$, then R^* can be written as

$$\begin{aligned} R^* &= \frac{\sum_{t=1}^k (r_{q(t-1)+1} + \dots + r_{qt} - q\hat{\mu})^2}{q \sum_{t=1}^n (r_t - \hat{\mu})^2} \\ &= \frac{\sum_{t=1}^k [(r_{q(t-1)+1} - \hat{\mu}) + \dots + (r_{qt} - \hat{\mu})]^2}{q \sum_{t=1}^n (r_t - \hat{\mu})^2} \\ &= \frac{1}{q} [(u_1 + \dots + u_q)^2 + (u_{q+1} + \dots + u_{2q})^2 + \dots + (u_{(k-1)q+1} + \dots + u_{kq})^2]. \end{aligned}$$

Similarly, R can be written as

$$\begin{aligned} R &= \frac{1}{q} [(u_1 + \dots + u_q)^2 + (u_2 + \dots + u_{q+1})^2 + \dots + (u_{(k-1)q} + \dots + u_{kq-1})^2 \\ &\quad + (u_{(k-1)q+1} + \dots + u_{kq})^2]. \end{aligned}$$

Due to the end-point effect, R cannot be conveniently related to R^* . To exploit the result of the finite-sample distribution of R^* we construct a COVR statistic. Defining $r_{n+1} \equiv r_1, r_{n+2} \equiv r_2, \dots, r_{n+q-1} \equiv r_{q-1}$, we calculate the COVR statistic R_c as

$$R_c = \frac{\sum_{t=1}^n (r_t + \dots + r_{t+q-1} - q\hat{\mu})^2}{q \sum_{t=1}^n (r_t - \hat{\mu})^2},$$

which can be written as

$$\begin{aligned} R_c &= [\{(u_1 + \dots + u_q)^2 + (u_{q+1} + \dots + u_{2q})^2 + \dots + (u_{q(k-1)+1} + \dots + u_{kq})^2\} + \\ &\quad \{(u_2 + \dots + u_{q+1})^2 + (u_{q+2} + \dots + u_{2q+1})^2 + \dots + (u_{q(k-1)+2} + \dots + u_{kq+1})^2\} + \\ &\quad \dots + \{(u_q + \dots + u_{2q-1})^2 + (u_{2q} + \dots + u_{3q-1})^2 + \dots + (u_{kq} + \dots + u_{kq+q-1})^2\}]/q \\ &= [Q_1 + \dots + Q_q]/q, \end{aligned}$$

²Faust (1992) demonstrated how the OVR statistic can be written as a ratio of quadratic forms in normal variates. Thus, numerical methods (such as the Imhof method) can be used to calculate the distribution. The numerical methods are, however, computationally intensive, especially for sample sizes typically encountered in studies using the variance-ratio tests.

where Q_i , for $i = 1, \dots, q$, denote sequentially the terms in the curly brackets. By exchangeability argument, Q_i/q are identically distributed as $\text{Beta}((k-1)/2, k(q-1)/2)$ on H_0 , although they are not independent. Thus, we have the exact result

$$\mathbb{E}(R_c) = \frac{q(k-1)}{qk-1},$$

which implies

$$\mathbb{E} \left[\frac{qk-1}{q(k-1)} R_c \right] = 1.$$

Note that $(qk-1)/[q(k-1)]$ is the degrees-of-freedom correction for the asymptotic mean to be exact. With this correction, we apply Lo and MacKinlay's result so that when n is large,

$$\text{Var} \left[\frac{qk-1}{q(k-1)} R_c \right] \approx V,$$

where V is given in (2).

As R_c is the sum of q Beta variates each taking values between 0 and 1, we normalize R_c by a factor of q so that the normalized statistic lies within the support of a Beta variate. Our strategy is to approximate R_c/q using a Beta distribution that matches the exact mean and the asymptotic variance of R_c/q . Specifically, we denote

$$m = \frac{k-1}{qk-1}$$

and

$$\nu = \frac{(k-1)^2}{(qk-1)^2} V,$$

and approximate R_c/q by $\text{Beta}(\alpha, \beta)$ such that

$$m = \frac{\alpha}{\alpha + \beta}$$

and

$$\nu = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}.$$

Solving the simultaneous equations we obtain the parameters of the Beta distribution as

$$\alpha = \frac{m^2 - m^3 - m\nu}{\nu}$$

and

$$\beta = \frac{\alpha(1-m)}{m}.$$

Given a probability p , let ξ satisfies

$$\begin{aligned} p &= \Pr(R_c < \xi) \\ &= \Pr(R_c/q < \xi/q) \\ &\approx \Pr(\text{Beta}(\alpha, \beta) < \xi/q). \end{aligned}$$

Then, if η satisfies $\Pr(\text{Beta}(\alpha, \beta) < \eta) = p$, we can approximate ξ by $q\eta$.³

In the next section we examine the small-sample performance of the Beta approximation to the COVR statistic using Monte Carlo method.

4 Some Monte Carlo Results

We first examine the null distribution of R_c . We consider samples of size $n = 30, 60, 120, 180, 360$ and 720 . For each n , several values of q are considered such that n/q is an integer. We consider values of q up to one-sixth of n . Table 1 summarizes the empirical sizes of R_c corresponding to one-sided tests with nominal sizes of 2.5% and 5%, and two-sided tests with nominal sizes of 5% and 10%. The results are based on Monte Carlo runs of 100,000 samples. For the one-sided tests, there is under-rejection in the lower tail and over-rejection in the upper tail. Compared to the results of the OVR statistic as reported in Lo and MacKinlay (1989), there are marked improvements. At the nominal significance level of 5%, the approximation works very well for n as small as 30.

Next we consider the power of R_c when X_t are not generated from a random walk. Two models are considered. In Model 1, X_t are generated from the following equations:

$$X_t = Y_t + Z_t$$

³A GAUSS code to compute the percentiles of R_c based on the Beta approximation can be downloaded from the web site: <http://courses.nus.edu.sg/course/ecstseyk/tykhp.htm>.

$$\begin{aligned}
Y_t &= \theta Y_{t-1} + \varepsilon_t, & \varepsilon_t &\sim N(0, 1) \\
Z_t &= Z_{t-1} + \xi_t, & \xi_t &\sim N(0, \sigma^2).
\end{aligned}$$

Thus, X_t consists of an AR(1) component Y_t and a random-walk component Z_t , so that X_t follows an ARIMA(1, 1, 1) process. We consider $\theta = 0.95$ and 0.9 , and $\sigma^2 = 0.5$ and 1 . Note that σ^2 measures the size of the random-walk relative to the AR component. We expect the power of R_c to be large when θ and σ^2 are small. Model 2 is a simple AR(1) process in which $X_t = \phi X_{t-1} + \varepsilon_t$ with $\phi = 0.95$ and $\varepsilon_t \sim N(0, 1)$. In this model, without loss of generality the variance of the error has been taken to be unity. Table 2 reports the estimated power of two-sided R_c -test based on Monte Carlo runs with sample size of 500,000.

For the AR(1) process, the power is low for $n \leq 120$. When $n \geq 180$, the power of the test increases with q . Indeed, it can be seen that there is remarkable improvement in the power for using the COVR test with larger value of q when $n \geq 360$. For the ARIMA(1, 1, 1) model, the power decreases with q when $n \leq 60$. For $n \geq 180$, however, the power increases with q initially and then decreases. As expected, the power is higher for smaller θ and σ^2 .

5 Conclusions

We have proposed a small-sample approximation to the distribution of a circulant variance-ratio statistic for testing the random-walk hypothesis with Gaussian errors. The approximation is based on fitting a Beta distribution to the test statistic that matches its exact mean and asymptotic variance. The critical values of the test can be calculated analytically. The Monte Carlo experiment shows that the approximation works well in small samples.

References

- [1] Campbell, J.Y. and N.G. Mankiw, 1987a, “Are Output Variations Transitory?” *Quarterly Journal of Economics*, **102**, 857 – 880.
- [2] Campbell, J.Y. and N.G. Mankiw, 1987b, “Permanent and Transitory Components in Macroeconomic Fluctuations”, *American Economic Review*, **77**, 111 – 117.
- [3] Campbell, J.Y. and N.G. Mankiw, 1989, “International Evidence on the Persistence of Economic Fluctuations”, *Journal of Monetary Economics*, **23**, 297 – 318.
- [4] Cecchetti, S.G. and P.S. Lam, 1994, “Variance-Ratio Tests: Small-Sample Properties With an Application to International Output Data”, *Journal of Business and Economic Statistics*, **12**, 177 – 186.
- [5] Cochrane, J., 1988, “How Big Is the Random Walk in GNP?”, *Journal of Political Economy*, **96**, 893 – 920.
- [6] Cogley, T., 1990, “International Evidence on the Size of the Random Walk in Output”, *Journal of Political Economy*, **98**, 501 – 518.
- [7] Faust, J., 1992, “When Are Variance Ratio Tests for Serial Dependence Optimal?” *Econometrica*, **60**, 1215 – 1226.
- [8] Lo, A.W. and A.C. MacKinlay, 1988, “Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test”, *Review of Financial Studies*, **1**, 41 – 66.
- [9] Lo, A.W. and A.C. MacKinlay, 1989, “The Size and Power of the Variance Ratio Test in Finite Samples: A Monte Carlo Investigation”, *Journal of Econometrics*, **40**, 203 – 238.
- [10] Poterba, J.M. and L.H. Summers, 1988, “Mean Reversion in Stock Returns: Evidence and Implications”, *Journal of Financial Economics*, **22**, 27 – 60.

- [11] Zhang, Y. and G. Tian, 1999, “The Exact Distribution of Variance Ratio Statistics Used in Testing Financial Market Efficiency”, Beijing University, mimeo.

Table 1. Empirical Size (%) of the COVR Statistic

n	q	Lower Tail		Upper Tail		Two-sided	
		2.5%	5%	2.5%	5%	5%	10%
30	2	2.45	4.92	2.56	5.05	5.01	9.98
	3	2.04	4.50	2.89	5.46	4.93	9.96
	5	1.46	3.95	3.08	5.60	4.54	9.54
60	2	2.49	5.05	2.44	5.00	4.93	10.05
	3	2.18	4.67	2.74	5.22	4.92	9.90
	4	2.05	4.57	2.84	5.34	4.90	9.91
	5	1.97	4.52	3.00	5.52	4.97	10.04
	6	1.88	4.31	3.02	5.51	4.90	9.83
	10	1.43	3.79	3.20	5.77	4.62	9.56
120	2	2.49	4.97	2.48	4.94	4.97	9.92
	3	2.42	4.87	2.60	5.20	5.02	10.07
	4	2.19	4.74	2.76	5.40	4.95	10.14
	6	2.11	4.63	2.90	5.40	5.01	10.03
	10	1.94	4.42	2.98	5.51	4.93	9.92
	20	1.44	3.83	3.18	5.77	4.62	9.60
180	2	2.56	5.07	2.50	5.01	5.07	10.08
	4	2.42	4.95	2.69	5.16	5.11	10.11
	6	2.18	4.56	2.74	5.28	4.92	9.84
	10	2.09	4.60	2.91	5.27	5.01	9.87
	20	1.72	4.22	3.06	5.56	4.79	9.78
	30	1.37	3.71	3.18	5.71	4.55	9.42
360	2	2.47	5.02	2.47	4.99	4.95	10.01
	5	2.31	4.80	2.65	5.22	4.97	10.02
	10	2.21	4.67	2.78	5.32	4.99	9.99
	20	2.10	4.72	2.87	5.43	4.97	10.15
	40	1.79	4.22	3.03	5.65	4.82	9.87
	60	1.51	3.99	3.13	5.61	4.63	9.60
720	2	2.40	4.94	2.61	5.18	5.00	10.12
	5	2.39	4.91	2.58	5.08	4.97	9.98
	10	2.42	4.83	2.64	5.15	5.07	9.98
	30	2.18	4.70	2.81	5.41	4.99	10.11
	60	1.93	4.42	2.92	5.46	4.85	9.89
	120	1.48	3.87	3.27	5.78	4.75	9.65

Note: The figures are the estimated sizes of the COVR statistic when the critical values are computed using the Beta approximation. The estimates are based on Monte Carlo runs of 100,000 samples.

Table 2. Estimated Power (%) of the COVR Statistic

n	q	ARIMA(1; 1; 1) Process								AR(1) Process	
		$\mu = 0:95$				$\mu = 0:90$				$\hat{A} = 0:95$	
		$\frac{3}{4}^2 = 0:5$		$\frac{3}{4}^2 = 1:0$		$\frac{3}{4}^2 = 0:5$		$\frac{3}{4}^2 = 1:0$			
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
30	2	4.92	9.92	4.81	9.86	4.85	9.93	4.97	9.92	4.91	9.87
	3	4.59	9.70	4.64	9.83	4.46	9.38	4.45	9.26	4.53	9.40
	5	4.20	9.06	4.21	9.26	3.93	8.55	3.78	8.52	4.03	8.89
60	2	4.75	9.62	4.89	9.92	5.07	10.19	4.89	9.83	4.83	9.92
	3	4.67	9.55	4.72	9.60	4.73	9.98	4.60	9.52	4.68	9.66
	4	4.49	9.45	4.48	9.40	4.67	9.80	4.27	9.27	4.53	9.44
	5	4.29	9.22	4.33	9.16	4.40	9.57	4.27	9.23	4.29	9.28
	6	4.32	9.11	4.34	9.20	4.11	9.34	4.04	9.09	4.15	9.12
	10	3.66	8.36	3.72	8.35	3.33	8.48	3.38	8.16	3.27	8.01
120	2	5.12	10.12	4.99	9.97	5.80	11.02	5.20	10.49	5.24	10.38
	3	4.92	9.89	4.88	10.01	6.17	11.93	5.34	10.57	5.25	10.47
	4	4.76	9.97	4.66	9.50	6.30	12.37	5.23	10.71	5.35	10.72
	6	4.50	9.62	4.54	9.51	6.76	13.28	5.34	11.13	5.33	10.81
	10	4.43	9.37	4.15	9.06	6.87	13.84	5.06	10.84	5.13	10.85
	20	3.36	8.35	3.19	7.94	5.37	12.83	3.92	9.64	4.17	10.40
180	2	5.22	10.47	5.02	10.06	6.43	12.11	5.65	11.03	5.59	10.71
	4	5.18	10.28	5.05	10.19	8.04	14.77	6.13	12.10	6.18	12.11
	6	5.31	10.88	4.74	9.95	9.46	17.14	6.88	13.28	6.79	13.12
	10	5.26	10.94	4.64	9.85	10.34	19.53	7.34	14.22	7.57	14.77
	20	4.87	10.52	4.23	9.54	11.00	21.41	6.74	14.10	8.34	17.17
	30	4.15	9.97	3.42	8.46	8.97	19.75	5.50	12.60	7.62	17.25
360	2	5.87	11.07	5.37	10.64	8.66	15.45	6.98	12.92	6.77	12.69
	5	6.89	13.31	5.93	11.92	16.13	26.17	10.83	18.91	10.70	18.58
	10	8.66	15.80	6.51	12.60	24.43	37.43	13.93	23.89	16.26	26.62
	20	10.53	19.24	6.95	13.95	30.62	46.54	16.20	27.74	24.50	38.39
	40	10.35	20.81	6.61	14.13	27.31	44.56	13.25	24.88	32.27	50.73
	60	8.50	19.17	5.32	12.54	20.09	36.95	9.40	19.63	33.00	54.73
720	2	6.94	12.80	5.84	11.26	14.05	22.65	10.05	17.31	9.48	16.35
	5	10.75	18.45	7.76	14.26	31.87	44.84	19.21	29.93	19.54	30.26
	10	15.96	25.93	10.30	18.18	50.78	64.85	28.82	41.99	34.58	48.16
	30	27.43	42.27	14.88	25.36	70.19	83.00	36.34	51.56	70.94	83.11
	60	29.26	46.10	14.49	25.99	59.06	75.40	26.59	41.96	88.51	95.56
	120	19.32	35.87	9.11	19.38	33.56	52.59	14.14	26.98	92.46	98.30

Note: The ARIMA(1, 1) process is given by $X_t = Y_t + Z_t$, with $Y_t = \mu Y_{t-1} + \epsilon_t$; $\epsilon_t \sim N(0; 1)$ and $Z_t = Z_{t-1} + \eta_t$; $\eta_t \sim N(0; \frac{3}{4}^2)$: The AR(1) process is given by $X_t = \hat{A}X_{t-1} + \epsilon_t$ with $\hat{A} = 0:95$ and $\epsilon_t \sim N(0; 1)$: The estimates are based on Monte Carlo runs of 50,000 samples.