Journal of Econometrics 58 (1993) 385–401. North-Holland

A simple multiple variance ratio test

K. Victor Chow and Karen C. Denning*

West Virginia University, Morgantown, WV 26506-6025, USA

Received May 1991, final version received April 1992

Empirical applications of the variance ratio (VR) test frequently employ multiple VR estimates to examine the random walk hypothesis against stationary alternatives. Failing to control the joint test size for these estimates results in very large Type I errors. This manuscript extends the Lo and MacKinlay (1988) methodology and provides a simple modification for testing multiple variance ratios. Monte Carlo results indicate that the size of our test is close to its nominal size and that it is as reliable as the Dickey–Fuller (D-F) and the Phillips–Perron (P-P) unit root tests. For a stationary AR(1) alternative, our test is comparable to both the D-F and the P-P tests and seems to be more powerful than these tests against two unit root alternatives, an ARIMA(1,1,1) and an ARIMA(1,1,0).

1. Introduction

This manuscript suggests an extension to the Lo and MacKinlay (1988) (hereafter LOMAC) variance ratio methodology, for testing the random walk hypothesis against stationary alternatives. The LOMAC test exploits the fact that the variance of random walk increments is linear in any and all sampling intervals. If stocks prices are generated by a random walk (possibly with drift), then the variance ratio, $VR(q)$, of $(1/q)$th of the variance of q-holding-period return to that of one-holding-period return should be unity for all q, where q is any integer greater than one. Lo and MacKinlay demonstrate that this property holds asymptotically even when the disturbances of a random walk stochastic process are subject to some types of heteroscedasticity. LOMAC propose two statistics for testing an individual variance ratio estimate, the homoscedasticity-robust $Z_1(q)$ and the heteroscedasticity-robust $Z_2(q)$, based on sample estimates of $VR(q)$. [Details are provided in Lo and MacKinlay (1988).] The asymptotic sampling distribution of both $Z_1(q)$ and $Z_2(q)$ is normal.

Correspondence to: K. Victor Chow, College of Business and Economics, West Virginia University, P.O. Box 6025, Morgantown, WV 26506-6025, USA.

* The authors are grateful to an associate editor and two referees of the *Journal of Econometrics* for comments and suggestions.

Under the random walk hypothesis, the unity of $VR(q)$ holds for each $q = 2,3, \ldots$ Thus, it is necessary to examine $Z_1(q)$ and/or $Z_2(q)$ for several selected values of q and only fail to reject the random walk hypothesis if it is not rejected for all of the q selected. In this framework, the statistical problem of multiple comparisons arises naturally.[1] That is, a nominal 100α percent critical value is not appropriate for *each* q selected; instead an *overall* size of 100α percent is necessary, where α is the nominal size of the test. Consequently, failing to control test size for multiple comparisons causes an inappropriately large probability of Type I error. In simulation experiments, we find that this Type I error is approximately 6, 4, and 3 times as large as the nominal size for the 1, 5, and 10 percent tests, respectively.

Further, in empirical investigations using the variance ratio methodology, values of the aggregation parameter, q, are potentially selected differently. A natural tendency may be to focus on the extreme statistics.[2] Such a focus might reasonably be called data-snooping.[3] Failure to control the overall test size may create or exacerbate a data-snooping bias. For example, the extreme statistics may lie outside the individual $100(1 - \alpha)$ acceptance region but inside the overall $100(1 - \alpha)$ joint acceptance region.

This study proposes a simple testing procedure. Using the maximum absolute value of the LOMAC test statistics for a set of multiple variance ratio estimates, $Z_j^*(q) = \max_q |Z_j(q)|$ ($j = 1$ and/or 2) corresponding to a set of pre-defined q selections, the Studentized Maximum Modulus (SMM) critical values can be used to control overall test size and to further define a joint confidence interval for these $VR(q)$ estimates. Section 2 briefly reviews the LOMAC variance ratio testing methodology and presents our multiple comparison extension. To illustrate the superiority of our testing procedure and to detect possible Type I errors which may create data-snooping biases in previous empirical studies, we re-examine the random walk hypothesis for weekly and monthly returns of CRSP market indices in section 3. The results are generally different from those reported in both Lo and MacKinlay (1988) and Poterba and Summers (1988), indicating some of the previous rejections of the random walk hypothesis may be due to the Type I error, data-snooping biases, and/or heteroscedasticity. Section 4 reports the results of Monte Carlo simulation experiments aimed at

[1] In testing multiple variance ratios, when the null hypothesis is H_0: $VR(2) = VR(4) = \cdots = 1$, one joint test statistic such as the F statistic may be appropriate. However, when H_0 is rejected, further information concerning whether the individual variance ratios or all ratios are different from one is desirable. This can be tested simultaneously from a set of subhypotheses – H_{01}: $VR(2) = 1$, H_{02}: $VR(4) = 1$, H_{03}: $VR(8) = 1$, etc., by examining multiple Z test statistics. Miller (1966, 1977) and Savin (1984) provide an excellent survey of multiple hypothesis testing and simultaneous confidence interval procedures. For a comparison of the multiple Z test and the F test, also see Savin (1984).

[2] For example, Fama and French (1988) and Poterba and Summers (1988).

[3] We are grateful to an associate editor of this journal and an anonymous referee for coining the phrase 'data-snooping'.

documenting the size and power of our multiple variance ratio test. The test size is estimated under the null hypothesis of a random walk with independently and identically distributed (i.i.d.) Gaussian increments, and with uncorrelated, but heteroscedastic increments. The power is examined under three alternative stationary hypotheses: AR(1), ARIMA(1,1,1), and ARIMA(1,1,0). For comparison, we also report the size and power of alternative unit root tests including the homoscedasticity-robust Dickey–Fuller test and the heteroscedasticity-robust Phillips–Perron test.[4] Section 5 contains a brief summary.

2. A multiple variance ratio extension of the LOMAC procedure

The asymptotic sampling theory for an individual variance ratio statistic is fully developed by Lo and MacKinlay (1988). What follows is a brief summary of their main result. Let X_t denote a stochastic process satisfying the following recursive relation:

$$X_t = \mu + X_{t-1} + \varepsilon_t, \qquad \mathrm{E}[\varepsilon_t] = 0, \qquad \text{for all } t, \tag{1}$$

where μ is an arbitrary drift parameter. Without further restriction, the random disturbance term, ε_t, has a zero mean, $\mathrm{E}(\varepsilon_t) = 0$, and a zero autocovariance, $\mathrm{E}(\varepsilon_t \varepsilon_{t-h}) = 0$, for any nonzero h. Under the random walk hypothesis that the increments are uncorrelated, the variance of X_t increments must be linear in any observation interval. That is, consider a qth lag difference of X_t, where q is any integer greater than one. The ratio of $(1/q)$th of the variance of $(X_{t+q} - X_t)$ to the variance of $(X_{t+1} - X_t)$ is equal to one. As long as the increments are uncorrelated, this relationship holds asymptotically even in the presence of heteroscedasticity.

Suppose that one obtains $nq + 1$ sample observations, $(X_0, X_1, \ldots, X_{nq})$. The LOMAC consistent estimate of the variance ratio estimate minus one, $\bar{M}_r(q)$, is calculated as

$$\bar{M}_r(q) \equiv [\bar{\sigma}^2(q)/\bar{\sigma}^2(1)] - 1, \tag{2a}$$

where

$$\bar{\sigma}^2(1) = (nq-1)^{-1} \sum_{k=1}^{nq} (X_k - X_{k-1} - \bar{X})^2, \tag{2b}$$

$$\bar{\sigma}^2(q) = \mathscr{S}^{-1} \sum_{k=q}^{nq} (X_k - X_{k-q} - q\bar{X})^2, \tag{2c}$$

[4] Both a 'unit root' and 'uncorrelated increments' are required for a random walk process. The unit root tests of Dicky and Fuller (1979, 1981) assume uncorrelated increments, while Phillips (1987), Perron (1988), and Phillips and Perron (1988) modify the tests under the condition that increments are weakly dependent. We note that the random walk model is a proper subset of the unit root hypothesis. There are some important departures from the random walk process that unit root tests cannot detect. See Beveridge and Nelson (1981) for an illustration.

with $\mathscr{S} = q(nq - q + 1)(1 - q/nq)$, and where $\bar{X}$ is the sample mean of $(X_t - X_{t-1})$. This estimate is asymptotically equivalent to a weighted sum of serial autocorrelation coefficient estimates such that:

$$\bar{M}_r(q) \overset{\mathrm{a}}{=} \sum_{j=1}^{q-1} (2(q-j)/q)\bar{r}(j), \tag{3}$$

where $\bar{r}(j)$ is the estimator of jth autocorrelation coefficient.

LOMAC identify the asymptotic distribution of the estimator $\bar{M}_r(q)$, based on eq. (3), under two null hypotheses – H_1: homoscedastic increments random walk and H_2: heteroscedastic increments random walk – as follows:[5]

Under H_1:

$$\sqrt{nq}\bar{M}_r(q) \overset{\mathrm{a}}{\sim} \mathrm{N}(0, 2(2q-1)(q-1)/3q).$$

Under H_2:

$$\sqrt{nq}\bar{M}_r(q) \overset{\mathrm{a}}{\sim} \mathrm{N}(0, V(q)),$$

where $V(q)$ is the asymptotic variance of $\bar{M}_r(q)$ under H_2 and is a weighted sum of the variances of $\bar{r}(j)$, denoted by $\delta(j)$, for $j = 1, \dots, q-1$. The consistent estimate of $V(q)$ is calculated as

$$V(q) = \sum_{j=1}^{q-1} (2(q-j)/q)^2 \bar{\delta}(j), \tag{4a}$$

where

$$\bar{\delta}(j) = \frac{(nq) \sum_{k=j+1}^{nq} (X_k - X_{k-1} - \bar{X})^2 (X_{k-j} - X_{k-j-1} - \bar{X})^2}{\left[\sum_{k=1}^{nq} (X_k - X_{k-1} - \bar{X})^2\right]^2}. \tag{4b}$$

[5] LOMAC relax the i.i.d. Gaussian restriction and follow White's (1980) and White and Domowitz's (1984) use of mixing and moment conditions to derive the heteroscedasticity-consistent variance ratio estimates. Additional restrictions on the variances of the increments allow the law of large numbers and the central limit theorem to apply for their asymptotic results. For details, see Lo and MacKinlay (1988).

Consequently, the appropriate test statistics under H_1 and H_2 are

$$Z_1(q) \equiv \sqrt{nq}\bar{M}_r(q)[2(2q-1)(q-1)/3q]^{-1/2} \overset{a}{\sim} N(0, 1), \tag{5a}$$

$$Z_2(q) \equiv \sqrt{nq}\bar{M}_r(q)[V(q)]^{-1/2} \overset{a}{\sim} N(0, 1). \tag{5b}$$

The LOMAC approach is appropriate for testing a variance ratio corresponding to a specific aggregation value, q, by simply comparing the test statistic, $Z_1(q)$ or $Z_2(q)$, with the standard normal critical value. However, since the random walk hypothesis requires that the variance ratios for all the aggregation intervals, q, selected must equal one, a natural way to test the null hypothesis is the multiple comparison of all selected variance ratio estimates with unity. With multiple comparisons, it is necessary to control the test size.

To do so, consider a set of variance ratio estimates, $\{\bar{M}_r(q_i) \mid i = 1, 2, \dots, m\}$, corresponding to a set of pre-specified aggregation intervals, $\{q_i \mid i = 1, 2, \dots, m\}$. Let q_i be any integer greater than one with q_i not equal to q_j for $i \neq j$.[6] Under the random walk null hypothesis, we test a set of sub-hypotheses, H_{0i}: $M_r(q_i) = 0$ for $i = 1, 2, \dots, m$. Since any rejection of H_{0i} will lead to the rejection of the random walk null hypothesis, let the largest absolute value of the test statistics be

$$Z_1^*(q) = \max_{1 \le i \le m} |Z_1(q_i)|, \tag{6a}$$

$$Z_2^*(q) = \max_{1 \le i \le m} |Z_2(q_i)|. \tag{6b}$$

We apply the Sidak (1967) probability inequality and the results in Hochberg (1974) and Richmond (1982) to control the size of the multiple variance ratio test, and further define a joint confidence interval for a *set* of variance ratio statistics. To briefly elaborate without loosing generality, let $\{z_i \mid i = 1, \dots, m\}$ be a set of m standard normal variates. According to the well-known Bonferroni probability inequality, $\Pr[\max(|z_1|, \dots, |z_m|) \le Z_{\alpha'/2}] \ge (1-\alpha)$, where $Z_{\alpha'/2}$ is the upper $\alpha'/2$ point of the standard normal distribution and $\alpha' = \alpha/m$. Assuming that z_i are independent, Sidak (1967) provides a general inequality which gives a slightly sharper interval than the Bonferroni inequality such that $\Pr[\max(|z_1|, \dots, |z_m|) \le Z_{\alpha^+/2}] \ge (1-\alpha)$, where $\alpha^+ = 1-(1-\alpha)^{1/m}$. Hochberg

[6] Although the number and values of the aggregation value, q, chosen are arbitrary, Lo and MacKinlay (1989) suggest that one choose q to be no more than one-half the total sample size due to the unreliability of large-sample theory under the null when q is large relative to the sample size.

(1974), using Sidak's result, proves that even under the condition that z_i are correlated with an arbitrary correlation matrix Ω, the following inequality holds: $\Pr[\max(|z_1|, \ldots, |z_m|) \leq \mathrm{SMM}(\alpha; m; N)] \geq (1-\alpha)$, where $\mathrm{SMM}(\alpha; m; N)$ is the upper α point of the Studentized Maximum Modulus (SMM) distribution with parameter m and N (sample size) degrees of freedom. Asymptotically, when $N = \infty$, the Hochberg inequality is equivalent to the Sidak inequality in that $\mathrm{SMM}(\alpha; m; \infty) = Z_{\alpha^+/2}$. Richmond (1982) further shows that for each z_i, the $\Pr[|z_i| \leq \mathrm{SMM}(\alpha; m; N), i = 1, \ldots, m] \geq (1-\alpha)$, which provides the key to our following lemma.

Lemma 1. Let $Z_1 = (Z_1(q_1), \ldots, Z_1(q_m))'$ *and* $Z_2 = (Z_2(q_1), \ldots, Z_2(q_m))'$ *be vectors of LOMAC test statistics corresponding to a set of m pre-defined aggregation values,* $\{q_i \mid i = 1, 2, \ldots, m\}$, *such that* $q_1(=2) < q_2 < \cdots < q_m \leq N/2$, *where N is the sample size. The distributions of vectors* Z_1 *and* Z_2 *converge asymptotically to m-variate normal distributions with means zero and covariance matrixes* Ω_1 *and* Ω_2, *respectively. The confidence interval of at least* $100(1-\alpha)$ *percent for the extreme statistic,* $Z_1^*(q)$ *or* $Z_2^*(q)$, *can be defined as:*

Under H_1:

$$Z_1^*(q) \pm \mathrm{SMM}(\alpha; m; \infty). \tag{7a}$$

Under H_2:

$$Z_2^*(q) \pm \mathrm{SMM}(\alpha; m; \infty). \tag{7b}$$

$SMM(\alpha; m; \infty)$ *is the asymptotic critical value of the α-point of the studentized maximum modulus (SMM) distribution with parameter m and ∞ degrees of freedom. Consequently, the asymptotic joint confidence interval of at least* $100(1-\alpha)$ *percent for a set of variance ratio estimates is:*

Under H_1:

$$\sqrt{N}\bar{M}_r(q_i) \pm [(2(2q_i - 1)(q_i - 1)/3q_i)]^{1/2}\, SMM(\alpha; m; \infty). \tag{8a}$$

Under H_2:

$$\sqrt{N}\bar{M}_r(q_i) \pm [V(q_i)]^{1/2}\, \mathrm{SMM}(\alpha; m; \infty) \quad \textit{for} \quad i = 1, \ldots, m. \tag{8b}$$

The asymptotic SMM critical value can be calculated from the conventional standard normal distribution. That is, $SMM(\alpha, m; \infty) = Z_{\alpha^+/2}$, *defined above.*[7]

[7] The SMM table can be found in Hahn and Hendrickson (1971) and Stoline and Ury (1979).

Based on the results in (8) and (8b) and the structure of test statistics in (5a) and (5b), we can control the size of a multiple variance ratio test by simply comparing LOMAC test statisitcs with the SMM critical values.

3. Empirical illustrations

Our first empirical illustration uses the U.S. stock market indices example of Lo and MacKinlay (1988, table 1a, p. 51). In our table 1a, we report LOMAC variance ratio results for weekly returns of equally-weighted and value-weighted CRSP NYSE–AMEX indices over the period 1962 to 1985. To control the size of the multiple variance ratio test, we compare the LOMAC calculated statistics with the SMM critical value of 2.491 for the 5 percent level of significance. It appears that for the equally-weighted index, the test statistics are highly significant, and thus the random walk hypothesis is rejected. However, for the value-weighted index, none of the statistics are significant according to the SMM critical value, thus indicating that the random walk hypothesis cannot be rejected. When one incorrectly compares these statistics with the conventional standard normal critical value of 1.96 for a 5 percent test, three out of four heteroscedasticity-robust $Z_2(q)$ statistics for the value-weighted index indicate an (incorrect) rejection of the null hypothesis. This example highlights the biases which may result when one ignores the joint nature of the variance ratio test statistics and suggests that LOMAC (1988) conclusion concerning the random walk of the value-weighted indices may be inaccurate.

Table 1a

LOMAC variance ratios for weekly market indices, 1962–1985.[a]

Index		Sampling interval (q) in weeks			
		2	4	8	16
Equally-weighted		1.30	1.64	1.94	2.05
CRSP NYSE-AMEX	$Z_2(q)$	(7.51)[b]	(8.87)[b]	(8.48)[b]	(6.59)[b]
Value-weighted		1.08	1.16	1.22	1.22
CRSP NYSE-AMEX	$Z_2(q)$	(2.33)[c]	(2.31)[c]	(2.07)[c]	(1.38)

[a]As shown in Lo and MacKinlay (1988, p. 51), the variance ratios, $1 + M_r(q)$, are reported in main rows, with the heteroscedasticity-robust test statistic, $Z_2(q)$, given in parentheses immediately below each main row.

[b]Using a multiple comparison procedure for the four variance ratios, the corresponding variance ratios are statistically different from one at the 5 percent level when compared with the SMM critical value of 2.491.

[c]Inference error in which the test statistics are separately significant according to the standard normal critical value but are jointly insignificant.

The rejection of the random walk hypothesis in LOMAC (1988) is due primarily to short-horizon (weekly) return dependence. For long-horizon (e.g., four-week or monthly) returns, the random walk model is not rejected [see Lo and MacKinlay (1988, p. 53)]. Poterba and Summers (1988) also examine the long-horizon dependence of stock returns using a variance ratio approach. Employing Monte Carlo estimations, they reject the random walk hypothesis at the 8 percent level for value-weighted excess returns and at the 0.5 percent level for the equally-weighted excess returns. The Poterba and Summers variance ratio estimates are generally less than one and are reported as significant, particularly, for $q = 36$, 48, and 60 months. In addition, the variance ratio declines as the holding period, q, increases, indicating the autocorrelation becomes increasingly negative for longer holding-period returns. This result is interpreted by Poterba and Summers as support for Summers' (1986) conjecture of mean-reversion in long-horizon stock returns. Long-horizon dependence and/or mean-reversion in stock returns, suggested by both Fama and French (1988) and Poterba and Summers (1988), continues to be debated [e.g., see Kim, Nelson, and Startz (1990), Jegadeesh (1990, 1991), and Lo (1991)]. The controversy may be due in part to the methodology used in Poterba and Summers, which lacks a *formal* statistical inference procedure and/or a method to control the overall test size when examining multiple variance ratios. In addition, in some cases, focusing primarily only on long-horizon variance ratio statistics might create the data-snooping bias.

To re-examine part of the data used in Poterba and Summers (1988), we employ LOMAC variance ratio estimations and our multiple comparison testing procedure. We analyze monthly real and excess returns on both the equally-weighted and value-weighted CRSP indices for the 1926–1988 period. Following Poterba and Summers (1988), we calculate the excess monthly returns with the risk-free rate measured as the Treasury bill yield and real returns measured using the Consumer Price Index (CPI) inflation rate.[8] Table 1b reports our variance ratio results. It appears that for aggregation values of q for 24 months or more, none of the test statistics is significant at the 5 percent level. For $q = 2$ months, the homoscedasticity-robust $Z_1(q)$ statistics are significant according to the SMM test with a 5 percent critical value of 2.683. However, this may due to the existence of heteroscedasticity, since the heteroscedasticity-robust $Z_2(q)$ statistics are statistically insignificant. Clearly, an inference error may occur when one compares the calculated statistics $[Z_2(q)]$ to the standard normal critical value. After adjustment for heteroscedasticity, test size, and data-snooping, we fail to reject the random walk

[8] We obtain Treasury bill yield and CPI inflation rate from Ibbottson and Sinquefield (1989).

Table 1b
Variance ratios for monthly market indices in relation to the return variation over a one-month period, 1962–1988.[a]

Index		Sampling interval (q) in months						
		2	6	12	24	36	48	60
EW-Real		1.21	1.19	1.31	1.38	1.32	1.33	1.30
	$Z_1(q)$	(5.82)[b]	(2.12)[c]	(2.29)[c]	(1.91)	(1.29)	(1.16)	(0.93)
	$Z_2(q)$	(2.39)[c]	(0.75)	(0.90)	(0.77)	(0.54)	(0.51)	(0.43)
EW-Excs		1.20	1.20	1.24	1.22	1.10	1.04	0.94
	$Z_1(q)$	(5.36)[b]	(2.22)[c]	(1.76)	(1.12)	(0.40)	(0.15)	(− 0.19)
	$Z_2(q)$	(2.04)[c]	(1.06)	(0.94)	(0.61)	(0.22)	(0.09)	(− 0.11)
VW-Real		1.14	1.09	1.24	1.32	1.29	1.24	1.17
	$Z_1(q)$	(3.82)[b]	(0.99)	(1.75)	(1.59)	(1.19)	(0.85)	(0.52)
	$Z_2(q)$	(1.99)[c]	(0.48)	(0.90)	(0.83)	(0.64)	(0.47)	(0.30)
VW-Excs		1.12	1.09	1.17	1.12	1.00	0.89	0.81
	$Z_1(q)$	(3.20)[b]	(1.03)	(1.26)	(0.60)	(− 0.01)	(− 0.37)	(− 0.59)
	$Z_2(q)$	(1.97)[c]	(0.57)	(0.74)	(0.36)	(− 0.01)	(− 0.24)	(− 0.39)

[a]Calculations are based on the monthly observations of real and excess returns for equally-weighted and value-weighted CRSP NYSE and AMEX portfolios, denoted as EW-Real, EW-Excs, VW-Real, and VW-Excs, respectively. The variance ratios, $1 + M_r(q)$, are reported in the main rows, with the homoscedasticity- and heteroscedasticity-robust test statistics, $Z_1(q)$ and $Z_2(q)$, given in parentheses immediately below each main row.

[b]The corresponding variance ratios are statistically different from one at the 5 percent level when compared with the SMM critical value of 2.683.

[c]Inference error in which the test statistics are separately significant according to the standard normal critical value but are jointly insignificant.

hypothesis for all four monthly stock indices.[9] Since our inference procedure relies solely on asymptotic distribution theory, we examine the performance of our SMM multiple comparisons approach in finite samples in the ensuing section.

[9] We note that the variance ratio estimates reported in table 1b are different from those reported in Poterba and Summers (1988). This is due a difference in estimating procedure. We use the LOMAC approach, $VR(q) = \sigma^2(q)/\sigma^2(1)$, for the return variation over a one-month period, while Poterba and Summers' estimation is based on $VR^*(q) = \sigma^2(q)/\sigma^2(12)$ for the return variation over a twelve-month period. Since $VR^*(q) = VR(q)/VR(12)$, which is analogous to the Poterba and Summers' estimation, when we examine these estimates in table 1b, the estimates are generally less than one and consistent with Poterba and Summers' estimations. We use eq. (2c) to calculate the consistent estimate of $VR^*(q)$. Further, as shown in Poterba and Summers (1988),

$$[VR^*(q) - 1] \stackrel{a}{=} \sum_{j=1}^{q-1} (2(q-j)/q)r(j) - 2\sum_{j=1}^{11} (12-j)/12)r(j).$$

This allows us to apply the LOMAC sampling theory in section 2 and to determine the variance of the sample estimate of $[VR^*(q) - 1]$ and further calculate the test statistics. We test the $VR^*(q)$ for q up to 96 months and find none of the statistics is significant at the 5 percent level.

4. Size and power of the multiple variance ratio test

To demonstrate the potential errors which result when one performs multiple comparisons for various variance ratios without controlling the overall size of the test, and to analyse the size and power of the proposed testing procedure, we repeat the LOMAC (1989) simulation experiments. The size of the test is estimated under both the Gaussian i.i.d. null and the heteroscedastic null hypothesis.[10] We compare the power of the test against three alternatives of recent empirical interest: stationary AR(1), ARIMA(1,1,1), and ARIMA(1,1,0). For comparisons, we also report the empirical size and power of the Dickey–Fuller (D-F) and Phillips–Perron (P-P) unit root tests.[11]

4.1. The test size

Table 2 reports the results of simulation experiments conducted under the random walk null hypothesis of an i.i.d. Gaussian process and a process with heteroscedastic increments.[12] We estimate the probability of this Type I error (the test size) by calculating the percentage of rejections of the null hypothesis from 20,000 random samples generated from these pre-defined random walk populations. To examine the Type I error of the multiple variance ratio test without controlling the test size, we incorrectly test the largest absolute value of the statistics, $Z_1^*(q)$ or $Z_2^*(q)$, as if they were standard normal (SN) variates. This testing procedure, denoted as $Z_1^*(q)$-SN or $Z_2^*(q)$-SN, is equivalent to comparing each $Z_1(q)$ [or $Z_2(q)$] statistic to the conventional standard normal critical value. The probability of the Type I error for both the $Z_1^*(q)$-SN and $Z_2^*(q)$-SN test is quite large under both the i.i.d. Gaussian and heteroscedastic null. The

[10] Under the heteroscedasticity null hypothesis, the disturbance ε_t in eq. (1) satisfy the relation $\varepsilon_t \equiv \sigma_t v$, where v_t is i.i.d. N(0, 1), and σ_t follows $\ln\sigma_t^2 = \beta\ln\sigma_{t-1}^2 + \tau$, where τ is i.i.d. N(0, 1).

[11] In this paper, we use the D-F and P-P test in the regression model: $X^t = \mu + \beta(t - nq/2) + \alpha X_{t-1} + \varepsilon_t$. Based on the random walk model, we test the null hypothesis H_0: $(\mu,\beta,\alpha) = (\mu,0,1)$. Under the assumption that ε_t is i.i.d. normal, the appropriate test statistic, Φ_3, is provided by Dickey and Fuller (1981). Phillips (1987) and Phillips and Perron (1988) relax the Dickey–Fuller assumption and permit weakly dependent time series and heterogeneously distributed time series in the innovation sequence, $\{\varepsilon_t\}$. Under the null hypothesis, the modified test statistic, $Z(\Phi_3)$, is shown in detail by Perron (1988). Note that under Phillips and Perron's assumption, the limiting distribution of this statistic depends upon the ratio σ^2/σ_u^2, where $\sigma^2 = \lim_{T\to\infty} \mathrm{E}(S_T)^2/T\,(S_T = \sum_t \varepsilon_t)$, $\sigma_u^2 = \lim_{T\to\infty}[\mathrm{E}(\varepsilon_t^2)]/T$, and $T = nq$. The consistent estimator of σ^2 is obtained from the average sum of squared residuals under the null hypothesis. The estimate of σ_u^2 is calculated by a (truncated) linear weight function (or lag window) of the sample lagged autocovariances. We use the Parzen window and a lag truncation number, $l = 10$, as suggested by Perron (1988), for our estimations.

[12] To compare our simulation procedures to those of LOMAC (1989), we re-estimate the size of the separate test for each individual LOMAC statistic corresponding to the q value selected. The results are similar to those reported by LOMAC, indicating our experimental procedure is consistent with their's.

Table 2

Empirical sizes of the nominal 1, 5, and 10 percent multiple variance ratio tests for the random walk hypothesis with homoscedastic and heteroscedastic disturbances.[a]

Sample size		Homoscedasticity 1%	5%	10%		Heteroscedasticity 1%	5%	10%
				$m = 4$, $q = 2,4,8,16$				
32	$Z_1^*(q)$-SN	0.049	0.126	0.210	$Z_2^*(q)$-SN	0.066	0.157	0.264
32	$Z_1^*(q)$-SMM	0.026	0.056	0.082	$Z_2^*(q)$-SMM	0.039	0.073	0.105
32	D-F	0.012	0.057	0.091	P-P	0.046	0.078	0.113
				$m = 5$, $q = 2,4,8,16,32$				
64	$Z_1^*(q)$-SN	0.060	0.151	0.246	$Z_2^*(q)$-SN	0.064	0.153	0.262
64	$Z_1^*(q)$-SMM	0.030	0.060	0.088	$Z_2^*(q)$-SMM	0.034	0.064	0.092
64	D-F	0.013	0.047	0.092	P-P	0.020	0.049	0.092
				$m = 6$, $q = 2,4,8,16,32,64$				
128	$Z_1^*(q)$-SN	0.065	0.162	0.271	$Z_2^*(q)$-SN	0.072	0.168	0.273
128	$Z_1^*(q)$-SMM	0.032	0.060	0.084	$Z_2^*(q)$-SMM	0.038	0.067	0.091
128	D-F	0.009	0.048	0.095	P-P	0.010	0.049	0.093
				$m = 7$, $q = 2,4,8,16,32,64,128$				
256	$Z_1^*(q)$-SN	0.071	0.179	0.295	$Z_2^*(q)$-SN	0.079	0.188	0.307
256	$Z_1^*(q)$-SMM	0.030	0.061	0.088	$Z_2^*(q)$-SMM	0.038	0.069	0.095
256	D-F	0.008	0.051	0.101	P-P	0.009	0.050	0.099
				$m = 8$, $q = 2,4,8,16,32,64,128,256$				
512	$Z_1^*(q)$-SN	0.083	0.203	0.332	$Z_2^*(q)$-SN	0.086	0.206	0.330
512	$Z_1^*(q)$-SMM	0.035	0.066	0.094	$Z_2^*(q)$-SMM	0.039	0.070	0.097
513	D-F	0.009	0.049	0.099	P-P	0.008	0.050	0.101
				$m = 9$, $q = 2,4,8,16,32,64,128,256,512$				
1,024	$Z_1^*(q)$-SN	0.086	0.221	0.360	$Z_2^*(q)$-SN	0.083	0.214	0.348
1,024	$Z_1^*(q)$-SMM	0.034	0.064	0.091	$Z_2^*(q)$-SMM	0.034	0.064	0.090
1,024	D-F	0.010	0.051	0.101	P-P	0.008	0.048	0.099

[a]The simulation is carried out according to (1) ε_t is i.i.d. N(0, 1) (homoscedasticity), and (2) ε_t follows a specific form of homoscedasticity such that $\varepsilon_t = \sigma_t v$, $v \sim \mathrm{N}(0, 1)$, $\ln\sigma_t^2 = \frac{1}{2}\alpha\ln\sigma_{t-1}^2 + \tau$, $\tau \sim \mathrm{N}(0, 1)$. $Z_j^*(q)$ with $j = 1,2$ as the maximum absolute value of a set of statistics $Z_j(q)$ with $j = 1,2$. $Z_j^*(q)$-SN and $Z_j^*(q)$-SMM for $j = 1,2$ represent that $Z_j^*(q)$ are compared to the critical values from a standard normal and a SMM table, respectively. For comparison, the empirical size of the Dickey–Fuller Φ_3 test (D-F) and Phillips–Perron $Z(\Phi_3)$ test (P-P) are also reported. Each set of rows with a given sample size forms a separate and independent simulation experiment based on 20,000 replications.

empirical sizes of the 1, 5, and 10 percent $Z_1^*(q)$-SN ($Z_2^*(q)$-SN) tests range from 0.049(0.066),0.126 (0.157), 0.210 (0.264) to 0.086 (0.083), 0.221 (0.214), 0.360 (0.348), respectively. Perhaps the most striking result is that the error increases with the sample size for both the $Z_1^*(q)$-SN and the $Z_2^*(q)$-SN tests. For instance, for a 5 percent $Z_2^*(q)$-SN test the probability of the Type I error consistently increases from 15.7 percent to 21.4 percent as the sample size increases from 32 to 1,024. This highlights a potentially serious problem in empirical applications using this procedure in that the probability of incorrectly rejecting the random walk hypothesis is very large. The estimated probability of this Type I error is generally greater than 6, 3, and 2 times the nominal size for 1, 5, and 10 percent test, respectively.[13]

Using our multiple comparison procedure of simply comparing the test statistics instead to the asymptotic $\mathrm{SMM}(\alpha; m; \infty)$ critical values, denoted $Z_1^*(q)$-SMM and the $Z_2^*(q)$-SMM, it appears that the Type I error is significantly reduced and is much closer to the nominal size of the test. Table 2 shows that for finite samples the empirical size of the 1, 5, and 10 percent $Z_1^*(q)$-SMM and $Z_2^*(q)$-SMM tests are approximately 3, 6, and 9 percent, respectively. It appears that for a nominal size of 10 percent the SMM multiple comparison test is somewhat conservative, whereas the empirical sizes of the 1 and 5 percent tests are slightly greater than their nominal sizes. Our multiple variance ratio test, based on the asymptotic SMM approach, is by design conservative. However, Monte Carlo results at nominal sizes of 1 and 5 percent appear not to be conservatively sized. This may be due to the fact that under the null hypothesis, the finite sample SMM (empirical) distribution of the multiple variance ratio test statistics tends to be right-tail skewed instead of symmetric as is the asymptotic SMM distribution. Thus, using the asymptotic and theoretically symmetric SMM critical values may cause differences between the actual test size estimates and the nominal size. Nonetheless, the nominal size of the multiple comparison approach is much closer to being correct than is the testing of multiple variance ratios in a standard normal test as if they were separate. When comparing the multiple comparison SMM test to the D-F and the P-P tests, table 2 indicates that the SMM approach is generally as reliable as both of the alternative unit root tests, particularly at the 5 and 10 percent levels.

4.2. The test power

Tables 3a, 3b, and 3c report the empirical power of the multiple variance ratio test in finite samples.[14] Again, in these tables the empirical power of the D-F and P-P unit root test are also reported for comparison. Table 3a reports the power estimates for the AR(1) alternative with lag coefficients, $\phi = 0.85$ and 0.96. Overall, for small samples, both the unit root and the multiple variance ratio

[13] This result also highlights a potential for the data-snooping bias. For instance, when the most deviant variance ratio statistic is selected and compared to the standard normal critical value to test the random walk null hypothesis, this procedure may cause a bias in that the likelihood of obtaining a statistically significant statistic (the Type I error) is larger than it should be (the nominal size).

[14] Due to the asymmetry of the SMM distribution in finite samples, the critical values of the $Z_1^*(q)$-SMM and the $Z_2^*(q)$-SMM statistics are determined empirically by simulation under the i.i.d. Gaussian null and the heteroscedastic null. Further details are available upon request.

Table 3a

The power of the multiple variance ratio test for the AR(1) alternative.[a]

Sample size		$\phi = 0.85$			$\phi = 0.96$		
		1%	5%	10%	1%	5%	10%
		$m = 5$, $q = 2,4,8,16,32$					
64	$Z_1^*(q)$-SMM	0.018	0.079	0.148			
64	$Z_2^*(q)$-SMM	0.021	0.090	0.166			
64	D-F	0.024	0.070	0.133			
64	P-P	0.017	0.067	0.116			
		$m = 6$, $q = 2,4,8,16,32,64$					
128	$Z_1^*(q)$-SMM	0.043	0.192	0.344			
128	$Z_2^*(q)$-SMM	0.069	0.241	0.406			
128	D-F	0.110	0.344	0.506			
128	P-P	0.091	0.284	0.467			
		$m = 7$, $q = 2,4,8,16,32,64,128$					
256	$Z_1^*(q)$-SMM	0.497	0.652	0.867	0.011	0.060	0.013
256	$Z_2^*(q)$-SMM	0.517	0.729	0.911	0.015	0.075	0.149
256	D-F	0.782	0.975	0.994	0.015	0.084	0.152
256	P-P	0.734	0.961	0.989	0.014	0.085	0.160
		$m = 8$, $q = 2,4,8,16,32,64,128,256$					
512	$Z_1^*(q)$-SMM	0.874	0.998	0.999	0.023	0.141	0.277
512	$Z_2^*(q)$-SMM	0.983	0.999	1.000	0.056	0.220	0.379
512	D-F	1.000	1.000	1.000	0.127	0.385	0.552
512	P-P	1.000	1.000	1.000	0.130	0.386	0.557
		$m = 9$, $q = 2,4,8,16,32,64,128,256,512$					
1,024	$Z_1^*(q)$-SMM	1.000	1.000	1.000	0.153	0.574	0.804
1,024	$Z_2^*(q)$-SMM	1.000	1.000	1.000	0.232	0.683	0.879
1,024	D-F	1.000	1.000	1.000	0.875	0.985	0.997
1,024	P-P	1.000	1.000	1.000	0.858	0.979	0.998

[a]The specific form of the AR(1) alternative is: $X_t = \phi X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \mathrm{N}(0,1)$, and $\phi = 0.85$ and 0.96. The power of the Dickey–Fuller Φ_3 test (D-F) and Phillips–Perron $Z(\Phi_3)$ test (P-P) are also reported to compare to the power of the SMM test. Each set of rows for a given sample size forms a separate and independent simulation experiment based on 20,000 replications.

tests have low power which is close their respective test sizes when $\phi = 0.96$. The empirical power of the heteroscedasticity-robust $Z_2^*(q)$-SMM test appears somewhat greater than that of the $Z_1^*(q)$-SMM test, but lower than that of the unit root tests. However, as the sample size increases, the power of the SMM variance ratio test increases considerably and becomes close to that of the D-F and P-P tests. Specifically, the powers of all tests approach one for sample sizes ≥ 512 when $\phi = 0.85$. This suggests that for the AR(1) alternative our SMM multiple comparison test is comparable in power to these alternative unit root tests.

Table 3b
The power of the multiple variance ratio test for the ARIMA(1,1,1) alternative.[a]

Sample size		$\phi = 0.85$ 1%	5%	10%	$\phi = 0.96$ 1%	5%	10%
		$m = 5,\ q = 2,4,8,16,32$					
64	$Z_1^*(q)$-SMM	0.015	0.063	0.120			
64	$Z_2^*(q)$-SMM	0.017	0.074	0.136			
64	D-F	0.012	0.048	0.097			
64	P-P	0.009	0.016	0.043			
		$m = 6,\ q = 2,4,8,16,32,64$					
128	$Z_1^*(q)$-SMM	0.026	0.120	0.223			
128	$Z_2^*(q)$-SMM	0.041	0.151	0.263			
128	D-F	0.038	0.141	0.235			
128	P-P	0.018	0.108	0.201			
		$m = 7,\ q = 2,4,8,16,32,64,128$					
256	$Z_1^*(q)$-SMM	0.089	0.353	0.563	0.010	0.053	0.111
256	$Z_2^*(q)$-SMM	0.162	0.419	0.616	0.014	0.067	0.127
256	D-F	0.146	0.391	0.526	0.009	0.056	0.109
256	P-P	0.108	0.319	0.441	0.008	0.056	0.112
		$m = 8,\ q = 2,4,8,16,32,64,128,256$					
512	$Z_1^*(q)$-SMM	0.402	0.840	0.945	0.014	0.090	0.184
512	$Z_2^*(q)$-SMM	0.631	0.914	0.970	0.034	0.143	0.251
512	D-F	0.389	0.624	0.726	0.035	0.144	0.241
512	P-P	0.262	0.465	0.569	0.035	0.142	0.238
		$m = 9,\ q = 2,4,8,16,32,64,128,256,512$					
1,024	$Z_1^*(q)$-SMM	0.982	0.993	0.997	0.101	0.279	0.468
1,024	$Z_2^*(q)$-SMM	0.991	0.997	1.000	0.192	0.355	0.540
1,024	D-F	0.561	0.729	0.809	0.175	0.380	0.520
1,024	P-P	0.344	0.509	0.609	0.160	0.348	0.481

[a]The specific form of the ARIMA(1,1,1) alternative is: $X_t = Y_t + Z_t$, where $Y_t = 0.96Y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \mathrm{N}(0,1)$, and $Z_t = Z_{t-1} + \tau_t$, $\tau_t \sim \mathrm{N}(0, \frac{1}{2})$. The power of the Dickey–Fuller Φ_3 test (D-F) and Phillips–Perron $Z(\Phi_3)$ test (P-P) are also reported to compare to the power of the SMM test. Each set of rows for a given sample size forms a separate and independent simulation experiment based on 20,000 replications.

For the ARIMA(1,1,1) alternative, table 3b reports that both the $Z_1^*(q)$-SMM and the $Z_2^*(q)$-SMM test generally seem to dominate the Dickey–Fuller test.[15] Although, when $\phi = 0.96$, all the variance ratio and unit root tests have low power for samples size less than 256, the power increases significantly as the sample size increases and as the ϕ coefficient decreases. For example,

[15]The ARIMA (1, 1, 1) alternative, the mean-reverting process hypothesized by Summers (1986), is the sum of an stationary AR(1) process component, $Y_t = \phi Y_{t-1} + \varepsilon_t$, and a random walk component, $Z_t = Z_t + \tau_t$, in the $\{X_t\}$ series, where ε_t and τ_t are randomly distributed. When ϕ is close to one (e.g., $\phi = 0.96$), serial correlation in returns is negative and small in the short horizon. This would imply that mean-reversion occurs over very long periods of time. We expect that any existing unit root test has a low power against this alternative. [See Perron (1988) for illustrations.]

Table 3c
The power of the multiple variance ratio test for the ARIMA(1,1,0) alternative.[a]

Sample size		$k = 0.2$ 1%	5%	10%	$k = 0.4$ 1%	5%	10%
		$m = 4$, $q = 2,4,8,16$					
32	$Z_1^*(q)$-SMM	0.033	0.118	0.203	0.141	0.349	0.512
32	$Z_2^*(q)$-SMM	0.032	0.098	0.179	0.124	0.282	0.432
32	D-F	0.030	0.074	0.133	0.078	0.144	0.228
32	P-P	0.032	0.073	0.129	0.054	0.099	0.156
		$m = 5$, $q = 2,4,8,16,32$					
64	$Z_1^*(q)$-SMM	0.045	0.163	0.279	0.290	0.632	0.778
64	$Z_2^*(q)$-SMM	0.047	0.152	0.266	0.228	0.568	0.727
64	D-F	0.039	0.086	0.144	0.082	0.169	0.253
64	P-P	0.032	0.081	0.141	0.074	0.128	0.208
		$m = 6$, $q = 2,4,8,16,32,64$					
128	$Z_1^*(q)$-SMM	0.052	0.265	0.462	0.523	0.917	0.979
128	$Z_2^*(q)$-SMM	0.041	0.208	0.400	0.343	0.837	0.959
128	D-F	0.035	0.087	0.147	0.109	0.189	0.284
128	P-P	0.034	0.083	0.145	0.100	0.172	0.254
		$m = 7$, $q = 2,4,8,16,32,64,128$					
256	$Z_1^*(q)$-SMM	0.129	0.550	0.743	0.965	1.000	1.000
256	$Z_2^*(q)$-SMM	0.071	0.427	0.682	0.866	0.997	0.999
256	D-F	0.038	0.088	0.199	0.103	0.195	0.294
256	P-P	0.038	0.087	0.151	0.100	0.174	0.260
		$m = 8$, $q = 2,4,8,16,32,64,128,256$					
512	$Z_1^*(q)$-SMM	0.413	0.883	0.964	1.000	1.000	1.000
512	$Z_2^*(q)$-SMM	0.237	0.838	0.949	0.999	1.000	1.000
512	D-F	0.069	0.119	0.214	0.115	0.242	0.329
512	P-P	0.066	0.108	0.204	0.112	0.185	0.292
		$m = 9$, $q = 2,4,8,16,32,64,128,256,512$					
1,024	$Z_1^*(q)$-SMM	0.921	0.999	1.000	1.000	1.000	1.000
1,024	$Z_2^*(q)$-SMM	0.904	0.998	1.000	1.000	1.000	1.000
1,024	D-F	0.128	0.194	0.278	0.168	0.292	0.385
1,024	P-P	0.104	0.140	0.236	0.115	0.211	0.329

[a]The specific form of the ARIMA(1,1,0) is: $(X_t - X_{t-1}) = k(X_{t-1} - X_{t-2}) + v_t$, where $v_t \sim N(0,1)$ and $k = 0.2$ and 0.4. The power of the Dickey–Fuller Φ_3 test (D-F) and Phillips–Perron $Z(\Phi_3)$ test (P-P) are also reported to compare to the power of the SMM test. Each set of rows for a given sample size forms a separate and independent simulation experiment based on 20,000 replications.

the empirical power of a 5 percent $Z_2^*(q)$-SMM test when $\phi = 0.96$ increases from 14.3 percent to 35.5 percent when the sample increases from 512 observations to 1,024 observations. However, when $\phi = 0.85$, the empirical power increases from 91.4 percent to 99.7 percent. In general, our test seems to dominate the D-F and P-P unit root tests for this ARIMA(1,1,1) alternative. In addition, the heteroscedasticity-robust $Z_2^*(q)$-SMM test performs better than the i.i.d. Gaussian $Z_1^*(q)$-SMM test.

Table 3c presents power estimates for the ARIMA(1,1,0) alternative with recursive coefficients, $k = 0.2$ and 0.4. The power of the $Z_1^*(q)$-SMM test is slightly larger than that of the $Z_2^*(q)$-SMM test. As reported in LOMAC (1989), the separate variance ratio test seems to dominate the D-F unit root test for the ARIMA(1,1,0) alternative. Our results indicate that the multiple variance ratio test is also considerably more powerful than the alternative D-F or P-P unit root tests. The power of both the $Z_1^*(q)$-SMM and $Z_2^*(q)$-SMM approaches one for large samples (i.e., when $nq \geq 512$ for $k = 0.2$ and when $nq \geq 256$ for $k = 0.4$). Note that when $\{X_t\}$ is a log-price series, this ARIMA(1,1,0) process of $\{X_t\}$ is equivalent to an AR(1) process of the return series, $\{X_t - X_{t-1}\}$, with a positive k. This suggests that the autocorrelation of the return series is positive and large in the short horizon. Tables 1a and 1b, where the $VR(q)$ estimate for $q = 2$ is reported as greater than one, indicate that the first autocorrelation of stock returns is positive with an estimated value between 0.08 and 0.30. This tends to support the LOMAC (1989) conjecture that for weekly and monthly stock market indices returns, this ARIMA(1,1,0) process could be a more relevant alternative to the random walk null hypothesis than the AR(1) and/or the ARIMA(1,1,1) process. This being the case, for large samples with more than 256 observations, our proposed test can be a powerful tool to examine the plausibility of the random walk hypothesis of stock returns.

5. Conclusion

The variance ratio test exploits an important property of the random walk hypothesis – the variance of increments of a random walk is linear in any and all the sampling intervals (q). Empirical applications naturally employ different values of the aggregation parameter, q, and estimate multiple variance ratios. Examining multiple variance ratio estimates requires a multiple comparison statistical approach. We provide such an approach by suggesting a simple extension to the Lo and MacKinlay (1988) methodology.

We show that in a multiple comparison framework, without controlling test size, the probability of Type I error is generally greater than 6, 3, and 2 times the nominal size for the 1, 5, and 10 percent test, respectively. Alternatively, our simple testing procedure – treating the LOMAC test statistics as Studentized Maximum Modulus variates – is able to reduce the Type I error and control the size of a multiple variance ratio test. Monte Carlo evidence suggests that the power of our test is comparable to that of the Dickey–Fuller and the Phillips–Perron unit root tests against a stationary AR(1) alternative and is more powerful than these tests against the two unit root alternatives. It seems appropriate to end with a quote from Lo and MacKinlay (1989) who made path-breaking progress in developing asymptotic sampling theory for variance ratio testing: 'The simplicity, reliability, and flexibility of the (multiple) variance ratio test make it a valuable tool for statistical inference' concerning the stochastic process of security returns.

References

Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427–431.
Dickey, D.A. and W.A. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica 49, 1057–1072.
Fama, E. and K. French, 1988, Permanent and temporary components of stock prices, Journal of Political Economy 96, 246–273.
Hahn, G.R. and R.W. Hendrickson, 1971, A table of percentage points of the distribution of large absolute value of k Student t variates and its applications, Biometrika 58, 323–332.
Hochberg, Y., 1974, Some generalizations of T-method in simultaneous inference, Journal of Multivariate Analysis 4, 224–234.
Ibbottson, R. and R. Sinquefield, 1989, Stock, bonds, and inflation: The past and the future (The Financial Analysts Research Foundation, Charlottesville, VA).
Jegadeesh, N., 1990, Evidence of predictable behavior of security returns, Journal of Finance 45, 881–898.
Jegadeesh, N., 1991, Seasonality in stock price mean reversion: Evidence from the U.S. and U.K., Journal of Finance 46, 1427–1444.
Kim, M., C. Nelson, and R. Startz, 1991, Mean reversion in stock prices: A reappraisal of the empirical evidence, Review of Economic Studies 58, 515–528.
Lo, A., 1991, Long-term memory in stock market prices, Econometrica 59, 1279–1313.
Lo, A. 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.
Lo, A. and A.C. MacKinlay, 1989, The size and power of the variance ratio test in finite samples, Journal of Econometrics 40, 203–238.
Miller, R.G., 1966, Simultaneous statistical inference (McGraw-Hill, New York, NY).
Miller, R.G., 1977, Developments in multiple comparison, 1966–1976, Journal of the American Statistical Association 72, 779–788.
Perron, P., 1988, Trends and random walks in macroeconomic time series: Further evidence from a new approach, Journal of Economic Dynamics and Control 12, 297–332.
Phillips, P., 1987, Time series regression with a unit root, Econometrica 55, 277–301.
Phillips, P. and P. Perron, 1988, Testing for a unit root in time series regression, Biometrika 75, 335–346.
Poterba, J. and L. Summers, 1988, Mean reversion in stock returns: Evidence and implications, Journal of Financial Economics 22, 27–59.
Richmond, J., 1982, A general method for constructing simultaneous confidence intervals, Journal of the American Statistical Association 77, 455–460.
Savin, N.E., 1984, Multiple hypothesis testing, in: Z. Griliches and M.D. Intriligator, Handbook of economics, Vol. II (North-Holland, Amsterdam).
Sedak, Z., 1967, Rectangular confidence regions for the means of multivariate normal distributions, Journal of the American Statistical Association 62, 626–633.
Shiller, R.J. and P. Perron, 1985, Testing the random walk hypothesis: Power versus frequency of observations, Economics Letters 18, 381–386.
Stolin, M.R. and H.K. Ury, 1979, Tables of the studentized maximum modulus distribution and an application to multiple comparisons among means, Technometrics 21, 87–93.
Summers, L.H., 1986, Does the stock market rationally reflect fundamental values?, Journal of Finance 41, 591–600.
White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, Econometrica 48, 817–838.
White, H. and I. Domowitz, 1984, Nonlinear regression with dependent observations, Econometrica 52, 143–161.