A Review of Test for Randomness in Time Series Data

Alicia G. Strandberg[1], Boris Iglewicz[2]
[1]Department of Management and Operations, Villanova School of Business, Villanova University, Villanova, PA 19085
[2] Department of Statistics, The Fox School of Business, Temple University, Philadelphia, PA 19122

Abstract
We review and compare a new proposed test with well-known existing tests for detecting randomness in time series data, with special emphasis on stock market index data. By comparing popular variance ratio and traditional statistical tests plus a new proposed procedure, we have the most extensive simulation comparison of such tests. The investigated tests are compared over a diverse group of distributions, models, and stock market index applications. This study provides a useful guide to practical use of such testing for randomness procedures. We found that when evaluating common US stock market indices, the choice of data transformation can have a pronounced effect on test results.

Key Words: Data Transformation, Randomness, Test of Hypothesis, Stock Market Indices, Time Series, Variance Ratio

1. Introduction

In this paper we compare multiple tests for detecting randomness in time series data with special emphasis on stock market index data. Several methods exist for detecting randomness in time series data. We do not attempt to compare all possible and appropriate tests, instead, we focus on comparing the most commonly used reliable tests for this purpose. We have found that variance ratio tests, especially the ones considered in this paper, are the most commonly used methods for testing deviations from randomness in financial studies. Therefore, our comparison will include the popular variance ratio tests from Lo and MacKinlay (1988), Chow and Denning (1993) and Wright (2000) plus classical methods by Durbin and Watson (1950), and Ljung and Box (1978), along with a promising new method by Strandberg and Iglewicz (2013a). These tests will be compared over a diverse group of popular distributions, models, and applications to better understand the strengths and weaknesses of each considered test. By comparing popular variance ratio and traditional statistical tests plus a new proposed procedure, we have the most extensive simulation comparison of such tests.

In our comparisons differences between tests exists. For example, variance ratio tests are based on random walk models, thus not sensitive to changes in variance, while other considered tests are based on a random sample null hypothesis. Regardless, we will still consider these tests competitors. Varied sample sizes and distributional models will be considered in our comparisons. Since in practice most practitioners will not know the true distribution of their data, methods that perform well for a variety of distributions are preferred. This study will give researchers and practitioners added information about each considered test, so that they may make better informed decisions when deciding between methods.

2. Description of Tests

A brief overview of each considered test is given in this section. Please see the referenced sources for more detailed information.

A number of the tests considered in this section are of the variance ratio type. To determine lack of randomness of a times series, $Y_1, Y_2, ..., Y_N$, they compare the variance of the *k*-period return with *k* times the variance of the one-period return, that is $Var(Y_t - Y_{t-k}) = k\,Var(Y_t - Y_{t-1})$. Often test statistics are constructed based on a ratio estimator,

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

where $\hat{\sigma}^2(k)$ is the estimator of the *k*-period returns variance, defined as

$\hat{\sigma}^2(k) = m^{-1}\sum_{t=k}^{N}(Y_t - Y_{t-k} - k\hat{\mu})^2$ with $m = k(N-k+1)(1-kN^{-1})$ for $k \geq 1$, and

$\hat{\mu} = N^{-1}\sum_{t=1}^{N}\left(Y_t - Y_{t-1}\right)$, such that when $VR(k) = 1$ observations are serially uncorrelated. Likewise when $VR(k) \neq 1$some autocorrelations between observations exist.

Variance Ratio tests have the potential advantages of being robust under homoscedasticity and heteroscedasticity, and have reasonable power against a wide range of alternative hypotheses. However most variance ratio tests are based on an asymptotic normal distribution, but the sampling distribution of the test statistic can be skewed for small sample sizes. In addition, Variance ratio tests can be sensitive to the value chosen for *k.* Often low values, such as $k = 2$ are used as large values of *k* make the test statistic biased – since the variance ratio defined in (1) has a lower bound of zero. In practice, an individual test, designed to test only one value of *k*, may be repeated multiple times with different values of *k*. This can lead to an over rejection of the null hypothesis and an increased type I error. To address this issue, tests using multiple comparison procedures were created, see Chow and Denning (1993).

2.1 Lo and MacKinlay (1988)

Lo and MacKinlay (1988) developed a popular variance ratio test. Under the null hypothesis, the relation between observations is a random walk,

$$Y_t = \mu + Y_{t-1} + \varepsilon_t \tag{2}$$

where Y_t is the value of a return at time *t*, μ is an unknown arbitrary drift parameter and ε_t is a disturbance term with $Var(\varepsilon_t) = \sigma_\varepsilon^2$. Although, Lo and MacKinlay developed two variance ratio tests, M_1 and M_2, our simulation investigation, using the vrtest package of R (Kim 2010), showed that the simpler test, M_1, performs slightly better than M_2. As a result, our simulation results will only be reporting for M_1. Under the null hypothesis M_1 assumes ε_t in (2) are *i.i.d.* $N(0, \sigma_\varepsilon^2)$. The test statistic for M_1 is

$$M_1(k) = \frac{VR(k)-1}{\phi(k)^{1/2}} \tag{3}$$

where, $VR(k)$ as defined in (1), and $\phi(k) = \frac{2(2k-1)(k-1)}{3kN}$. When $VR(k) = 1$ observations are serially uncorrelated; likewise when $VR(k) \neq 1$some autocorrelations between observations exist (Lo and MacKinlay 1988, Charles and Darné 2009). Under the null hypothesis, M_1 follows asymptotically the standard normal distribution.

2.2 Wright (2000)

Wright (2000) created four nonparametric variance ratio tests using ranks and signs with exact sampling distributions. For each test Wright uses a martingale difference sequence rather than a random walk to define the null hypothesis, such that for a time series of asset returns, $Y_1, Y_2, ..., Y_N$,

$$Y_t = \mu + Z_t \tag{4}$$

where $Z_t = \sigma_t \varepsilon_t$. For the two rank-based tests, R_1 and R_2, Z_t is *i.i.d.* but with no additional distributional assumptions. R_1 and R_2 do not allow for conditional heteroscedasticity and are comparable to the assumptions of M_1. While two sign-based tests, S_1 and S_2, allow for conditional heteroscedasticity under the null hypothesis, they require conditional symmetry of ε_t from (4) given $\{Y_t, Y_{t-1}, ..., Y_2, Y_1\}$, with mean equal to zero (Wright 2000). Our investigation of R_1 and R_2 using the vrtest package of R (Kim 2010) showed that R_1 performs slightly better than R_2, so only simulation results for R_1 will be reported. Likewise in simulation studies S_2 was very conservative with low power and size distortion, see Wright (2000). Therefore only S_1 is considered here.

The rank-based test statistic, R_1, is defined

as $R_1 = \left(\dfrac{(Nk)^{-1} \sum_{t=k+1}^{N} \left(r_{1,t} + r_{1,t-1} + \ldots + r_{1,t-k} \right)^2}{N^{-1} \sum_{t=1}^{N} r_{1,t}^2} - 1 \right) \times \phi(k)^{-1/2}$, where $\phi(k)$ is defined in (3) and $r_{1,t} = \left(r(Y_t) - \dfrac{N+1}{2} \right) \Big/ \sqrt{\dfrac{(N-1)(N+1)}{12}}$. Here $r(Y_t)$ is the rank of Y_t for N observations among the series $Y_1, Y_2, ..., Y_N$, such that under the null hypothesis $r(Y)$ is a random permutation of the numbers 1, ..., N (Wright 2000). The sign-based test statistic, S_1, is defined as $S_1 = \left(\dfrac{(Nk)^{-1} \sum_{t=k+1}^{N} \left(s_t + s_{t-1} + \cdots + s_{t-k} \right)^2}{N^{-1} \sum_{t=1}^{N} s_t^2} - 1 \right) \times \phi(k)^{-1/2}$, where $\phi(k)$ is defined in (3), and $s_t = \begin{cases} 1 & \text{if } Y_t > 0 \\ -1 & \text{otherwise} \end{cases}$. S_1 is a driftless model, therefore in addition to the requirements of its null hypothesis, $\mu = 0$ in (4). The critical values of both tests can be obtained by simulating their exact distribution.

2.3 Chow and Denning (1993)

Chow and Denning (1993) apply theory from Hochberg's (1974) multiple comparison procedure to Lo and MacKinlay (1988) test statistics. This allows multiple values of *k* to be considered but results in only one test statistic properly adjusted for test size. They consider a set of *m* test statistics where *m* is the number of values of *k*. Consistent with other results, our simulations showed that the procedure based on M_1 performed better than the procedure based on M_2. Therefore, we will only report results for the multiple comparison test based on M_1 in our simulation summary. The test statistic for this multiple joint hypothesis is $MV_1 = \max_{1 \le i \le m} |M_1(k_i)|$, where $M_1(k)$ is defined in (3). To test the joint null hypothesis for a set of *m* test statistics, the null hypothesis is rejected if any of the estimated variance ratio tests is significantly different from one (Chow and Denning 1993).

2.4 Durbin and Watson (1950)

Durbin and Watson (1950) developed a very popular, traditional test for non-randomness in the residuals of an ordinary least squares regression equation using the test statistic,

$$d = \frac{\sum_{t=2}^{N} (e_t - e_{t-1})^2}{\sum_{t=1}^{N} e_t^2} \, , \tag{5}$$

where e_t is the t[th] residual and N is the number of observations. This test assumes *i.i.d.* $N(0, \sigma^2)$ errors. The test statistic, *d*, is asymptotically normal with mean of 2 and variance of 4/N (Harvey 1990). When data are positively (negatively) serially correlated, *d* will have a value that tends to zero (four). Although this test was designed for residuals of least squares regression, it can be used to test time series data such as $Y_1, Y_2, Y_3, ..., Y_N,$ by substituting $Y_t = e_t$ in (5), assuming the assumption of *i.i.d.* $N(0, \sigma^2)$ is valid.

2.5 Ljung and Box (1988)

The Ljung and Box (1988) test is another popular traditional statistics method. This test is used to determine if a specific ARIMA model fits. The test statistic is

$$Q = N(N+2)\sum_{k=1}^{h} \frac{r_k^2}{N-k} \tag{6}$$

where $r_k = \sum_{t=k+1}^{N} e_t e_{t-k} \Big/ \sum_{t=1}^{N} e_t^2$, e_t is the residual associated with the observation at time *t* and *h* is the number of lags being used. For large N, Q is distributed as χ_h^2 for ε_t *i.i.d.* $N(0, \sigma^2)$. This test can also be used to test time series data by replacing the residuals with corresponding time series data in (6), as long as the *i.i.d.* $N(0, \sigma^2)$ assumption holds.

2.6 Strandberg and Iglewicz (2013a)

Strandberg and Iglewicz (2013a) consider a time series, $Y_1, Y_2, ..., Y_N$ consisting of N observations, such that under the null hypothesis

$$Y_t = \mu_t + \varepsilon_t \tag{7}$$

where ε_t is *i.i.d.*$(0, \sigma^2)$, $Var[\varepsilon_t] = \sigma^2$ and $\mu_t = \mu$ for all *t*. No additional distributional assumptions are made on ε_t in (7). To test that these observations constitute a random sample, the set $\mathbf{Y} = \{Y_1, Y_2, …., Y_N\}$ is separated into **Y*** and **Y****, such that $\mathbf{Y^*} = \{Y_t : Y_t \in [Y_{0.025}, Y_{0.975}]\}$ where Y_p is the p[th] percentile, and **Y**** is the complement of **Y***, $\mathbf{Y^{**}} = \{Y_t : Y_t \notin [Y_{0.025}, Y_{0.975}]\}$, and $P(Y_t \in \mathbf{Y^{**}}) = \pi$.

The series of N observations is subdivided into M non-overlapping intervals each containing an equal number of K observations, where K is an integer that is small relative to N. For example, in our stock market example, complete weeks consisting of K = 5 days will be considered. Within each interval, counts of observations in **Y**** are measured such that W_j is the number of observations in interval *j* that are members of **Y****, *j* = 1, 2, ..., M and L denotes the number of the M groups where $W_j > 0$. Under the null hypothesis W_j is distributed as a binomial (K, π). The binomial test statistics is based on comparing the number of cases out of L, call it L_1, where W = 1, with $L \times P(W = 1 | W > 0)$, where $P(W = 1 | W > 0) = \frac{K\pi(1-\pi)^{K-1}}{1-(1-\pi)^K} = D$. The test statistic is

$$Z = \frac{\frac{L_1}{L} - D + H \times \frac{c}{L}}{\sqrt{\frac{D(1-D)}{L}}} \tag{8}$$

where $H = \begin{cases} 1 & if \ \frac{L_1}{L} \geq D \\ -1 & if \ \frac{L_1}{L} < D \end{cases}$ and $\frac{c}{L}$ is a useful correction factor. As suggested by Strandberg and Iglewicz, we use c = 0.5 in this study. For moderate to large values of N, Z ~ N(0, 1).

3. Simulation Results

Here, we consider random samples of N = 300, and 10,000 observations each generated from different distributional cases to verify appropriate test size. While a larger variety of distributional cases and sample sizes are given in Strandberg and Iglewicz (2013b), we consider only a few of these here. Null cases are considered first followed by alternative cases.

Using Tukey's g- and -h distributions – see Tukey (1977), Martinez and Iglewicz (1984), and Hoaglin (1985) – we approximate the standard normal distribution (Z), Student's t distribution with three degrees of freedom (t_3), and the chi-square distribution with four degrees of freedom (χ_4^2). The χ_4^2 distribution is standardized. Other null cases can be found in Strandberg and Iglewicz (2013b).

For each simulation we generated 10,000 random samples of N = 300, and 10,000 for the given distribution or model. The rejection rate is the percentage of tests out of 10,000 where the *p*-value < 0.05. Each individual variance ratio test is evaluated with *k* = 2, 5, and 10, likewise, each multiple variance ratio test is evaluated with *m* = 3, at *k* = 2, 5, and 10. The Ljung and Box test is evaluated at *h* = 2, and 5. A third value of $h \approx \ln(N)$ is also evaluated as suggested by Tsay (2001). Although the variance ratio tests are designed not to be sensitive to variance changes, we will still consider the studied tests comparable.

Simulation results for these distributional cases are listed in Table 1. Since these cases are expected to meet the required test size for the null hypothesis, we expect table entries to be close to 5.00%. Although some tests are slightly conservative when N = 300, in general most of the considered tests meet this requirement. There is one exception that should be noted, namely S_1, which exhibits high values for χ_4^2. Additionally, the Chow and Deming (1993) MV_1 test is overly conservative. In general, all other considered tests work reasonably well at meeting test-size requirements for null cases.

A fair number of alternative cases are considered in Strandberg and Iglewicz (2013b), with four selected in this paper. These are changing variance, changing mean, and two correlated cases C1 and C4. To create a model with constant mean and changing variance or constant variance and changing mean, let 30% of observations follow $f_1(y)$, then 40% follow $f_2(y)$, and the remaining 30% follow $f_3(y)$. A model with constant mean and changing variance is created by letting, $f_1(y)$ be N(0, 0.5625), $f_2(y)$ be N(0, 1), and $f_3(y)$ be N(0, 1.5625), while a model with constant variance and changing mean is created by letting $f_1(y)$ be N(-2, 1), $f_2(y)$ be N(0, 1), and $f_3(y)$ be N(2, 1).

Table 1 Simulation Rejection Rates Comparisons for Null Cases

	SI	DW		LB	MV1		M1	R1	S1
Distribution Cases from the g and h distribution when N = 300									
Z	4.21%	5.12%	*h* = 2	4.75%	3.02%	k=2	5.18%	4.83%	3.98%
			h = 5	5.13%		k=5	4.64%	5.32%	4.56%
			h ≈ ln(n)	5.11%		k=10	4.13%	5.10%	4.75%
t_3	4.06%	4.94%	*h* = 2	5.05%	3.52%	k=2	4.85%	4.78%	3.75%
			h = 5	4.89%		k=5	4.74%	4.94%	4.53%
			h ≈ ln(n)	5.14%		k=10	3.98%	5.33%	4.97%
X^2_4 Standardized	4.22%	4.78%	*h* = 2	4.76%	3.30%	k=2	4.90%	4.62%	7.94%
			h = 5	5.00%		k=5	4.61%	4.36%	18.11%
			h ≈ ln(n)	4.91%		k=10	3.91%	4.93%	31.22%
Distribution Cases from the g and h distribution when N = 10,000									
Z	5.29%	4.86%	*h* = 2	4.78%	3.73%	k=2	5.21%	5.04%	5.11%
			h = 5	4.55%		k=5	4.74%	5.00%	4.96%
			h ≈ ln(n)	5.21%		k=10	5.01%	5.24%	4.99%
t_3	4.96%	4.93%	*h* = 2	5.27%	3.44%	k=2	5.07%	4.80%	5.17%
			h = 5	5.15%		k=5	5.06%	5.44%	4.41%
			h ≈ ln(n)	5.24%		k=10	4.80%	5.68%	4.89%
X^2_4 Standardized	4.91%	4.92%	*h* = 2	5.03%	3.86%	k=2	4.96%	5.27%	92.93%
			h = 5	4.65%		k=5	5.15%	5.74%	100.00%
			h ≈ ln(n)	5.19%		k=10	4.84%	5.37%	100.00%

SI = Strandberg and Iglewicz (2013a) test, DW = Durbin and Watson (1950) test, LB = Ljung and Box (1978) test, MV1 = the Chow and Denning (1993) MV1 test, M1 = the Lo and MacKinlay (1988) M1 test, R1 and S1 are tests from Wright (2000)

Correlated cases are also considered. Here we will only discuss two. A total of four correlated cases are considered in Strandberg and Iglewicz (2013b). Each correlated model is motivated by stock market data and designed such that 90% of full trading weeks have observations that are *i.i.d.* N(0,1) while a correlation structure exists for the remaining 10% of weeks. Consider a typical member of the 10% correlated weeks. For the first case, Let Y^M be the Monday value coming from N(0, 4). The other generated values are Y^j, j = T, W, Th, F. Then $Y^T = \rho Y^M + \varepsilon$ where ε is N(0,1) and the remaining three days are from *i.i.d.* N(0, 1). This model will be labeled C1 since it is the first considered correlated model from Strandberg and Igelwicz (2013b). Next consider a correlation structure present over two weeks, such that Y^M in the first week, Y_1^M, has a value from N(0, 4). Then two days are correlated observations, such that $Y^j = \rho Y^M + \varepsilon$, where j is randomly chosen twice with replacement from T_1, W_1, Th_1, F_1, M_2, T_2, W_2, Th_2, F_2, while the remaining days are generated from *i.i.d.* N(0, 1). This model will be labeled C4 since it is the fourth considered correlated model from Strandberg and Igelwicz (2013b).

Results for alternative data models are included in Table 2. Simulations are generated as described in the null cases. As expected, greater power is generally seen in this table when N = 10,000. However, as can be seen, power can differ considerably across alternatives and considered methods. For example, when considering changing variances, only the Strandberg and Iglewicz (2013a) test (SI), is able to show power. It should be noted that this is not surprising, as the null hypothesis for variance ratio tests allows for changes in variance and conditional heteroscedasticity.

Table 2: Simulation Rejection Rates Comparisons for Alternative Cases

	SI	DW		LB	MV1		M1	R1	S1
Distribution Cases with N = 300									
Changing	13.81%	6.87%	*h* = 2	6.94%	4.87%	k=2	6.53%	5.55%	3.56%
Variance			*h* = 5	8.30%		k=5	6.24%	5.60%	4.63%
			h ≈ ln(n)	8.36%		k=10	5.59%	5.55%	4.72%
Changing	15.83%	100.00%	*h* = 2	100.00%	100.00%	k=2	100.00%	100.00%	100.00%
Mean			*h* = 5	100.00%		k=5	100.00%	100.00%	100.00%
			h ≈ ln(n)	100.00%		k=10	100.00%	100.00%	100.00%
C1	17.43%	22.81%	*h* = 2	15.09%	12.96%	k=2	18.40%	7.25%	3.98%
			h = 5	14.94%		k=5	11.47%	6.67%	4.33%
			h ≈ ln(n)	14.40%		k=10	7.80%	6.35%	4.98%
C4	8.18%	7.01%	*h* = 2	7.25%	5.54%	k=2	6.11%	5.20%	3.71%
			h = 5	7.72%		k=5	7.08%	5.31%	4.63%
			h ≈ ln(n)	8.24%		k=10	6.90%	5.83%	4.84%
Distribution Cases with N = 10,000									
Changing	95.32%	6.89%	*h* = 2	7.27%	5.52%	k=2	7.43%	5.35%	5.42%
Variance			*h* = 5	8.38%		k=5	7.14%	5.89%	5.21%
			h ≈ ln(n)	9.94%		k=10	6.98%	5.86%	5.00%
Changing	99.24%	100.00%	*h* = 2	100.00%	100.00%	k=2	100.00%	100.00%	100.00%
Mean			*h* = 5	100.00%		k=5	100.00%	100.00%	100.00%
			h ≈ ln(n)	100.00%		k=10	100.00%	100.00%	100.00%
C1	99.83%	100.00%	*h* = 2	100.00%	99.97%	k=2	100.00%	83.84%	27.70%
			h = 5	99.93%		k=5	99.26%	57.17%	17.25%
			h ≈ ln(n)	99.80%		k=10	86.99%	34.74%	11.08%
C4	65.96%	26.80%	*h* = 2	34.48%	64.71%	k=2	25.40%	8.94%	5.98%
			h = 5	47.92%		k=5	55.90%	16.53%	7.01%
			h ≈ ln(n)	52.78%		k=10	75.48%	25.03%	8.63%

SI = Strandberg and Iglewicz (2013a) test, DW = Durbin and Watson (1950) test, LB = Ljung and Box (1978) test, MV1 = the Chow and Denning (1993) MV1 test, M1 = the Lo and MacKinlay (1988) M1 test, R1 and S1 are tests from Wright (2000)

More consistent results are seen for the changing means case. Here, all tests do well with the exception of SI, which does poorly when N = 300 but gains power as N increases. Also for the first correlated case, C1, all considered tests show high power when N = 10,000, expect S_1 and to a lesser degree R_1. Differences are noticed with the other correlated case, C4. When N = 10,000 only SI maintains high power for C4 with MV_1 providing reasonable power.

In summary, when N is small, all considered procedures, excluding SI, have high power to detect changes in mean, while all tests have low power for other cases. As N becomes large, SI is the only test with high power for all considered alternative cases. In particular, if detecting changes in variance is a concern, the SI test should be used. In addition, for both correlated cases, especially C4, SI has high power and MV_1 does reasonably well.

4. Stock Market Analysis

Randomness of daily closing values of stock market indices are evaluated using the Dow Jones Industrial Average (DJIA), the Standards & Poor's 500 (S&P 500), and the National Association of Securities Dealers Automated Quotation System (Nasdaq). Data for full 5 day weeks were considered ending on December 18, 2009. These data were also analyzed by Strandberg and Iglewicz (2013a, 2013b). Using these data, three transformations are included - two common transformations and a newly proposed transformation by Strandberg and Iglewicz (2013a). The transformations are: lag 1 closing daily stock market index differences, $Y_t - Y_{t-1}$; daily percentage change defined as $100\left(\frac{Y_t(l)}{Y_{t-1}(l)} - 1\right)$ where $Y_t(l)$ is the daily closing price for index l on day t (Cizeau, Potters and Bouchaud 2001, Bandyopadhyay, Biswas and Mukherjee 2008, Mukherjee and Bandyopadhyay 2011), $l = 1, 2, 3$, for DJIA, S&P 500, and Nasdaq, respectively; the modified measure of percent change, MMPC, $100\left(\frac{Y_t(l)}{MA^{(l)}_{(t-2q:t-q-1)}} - 1\right)$ where MA is a delayed moving average of q observations such that $MA_{(t-2q:t-q-1)} = \frac{\sum_{j=t-2q}^{t-q-1} Y_t(l)}{q}$, were, as suggested by Strandberg and Iglewicz (2013a) , $q = 10$. It is reasonable to test whether these transformed observations constitute a random sample. Results are included in Table 3. To assist in the readability of Table 3, after each test statistics * is added if significance is at the 10% level, ** if significance is at the 5% level and *** if significance is at the 1% level. Extremely high test statistic values contain only *** with blanks for associated numbers

In Table 3, test results can differ depending on the transformation used. For example, S_1 is significance at 1% for all cases when $k = 2$, but can be non-significant at 10% for other values of k. The latter is not entirely surprising since S_1 is based on signs. Other individual variance ratio tests similarly show different results depending on the chosen value of k. R_1 and S_1 often show high rejection levels but can have conflicting conclusions depending on the chosen value of k, this is true to a lesser degree for M_1. This is also true for the Ljung and Box test with respect to the chosen value of h. The only test that shows consistent results among each transformation is the SI test. SI is significance at the 1% level for all indices and transformations. Notice that each of these stock market data sets consists of a fairly large number of observations. Consequently, the conclusions of this table could have been quite different if the data sets were considerably smaller.

Among transformations, only the third transformation, the MMPC, results in consistently rejecting the null hypothesis at the 1% significant level, while the other two transformations show inconsistent rejection levels among tests and indices.

In summary, when dealing with financial data, the test procedure used and choice of transformation can play a key role in resulting conclusions. The only test that does not show different results depending on the transformation used is Strandberg and Iglewicz (2013a). While the MMPC transformation is the only considered transformation with all test results significant at the 1% level.

Table 3: Tests and Transformation Comparisons for Stock Market Index Data

DJIA	SI	DW		LB	MV1		M1	R1	S1
$Y_t - Y_{t-1}$	-34.47***	2.11***	$h = 2$	***	8.12***	k=2	-7.44***	3.14***	5.04***
			$h = 5$	***		k=5	-8.12***	0.30	1.09
			$h \approx \ln(N)$	***		k=10	-6.67***	-0.26	0.70
$100((Y_{t/}\, Y_{t-1}) -1)$	-16.79***	1.99	$h = 2$	5.85*	0.70	k=2	0.70	6.76***	5.04***
			$h = 5$	10.64*		k=5	0.09	2.49**	1.09
			$h \approx \ln(N)$	24.62***		k=10	0.31	1.42	0.70
$100((Y_t /MA)-1)$	-36.64***	0.07***	$h = 2$	***	***	k=2	***	***	***
			$h = 5$	***		k=5	***	***	***
			$h \approx \ln(N)$	***		k=10	***	***	***
S&P 500									
$Y_t - Y_{t-1}$	-28.01***	2.11***	$h = 2$	69.06***	7.55***	k=2	-6.23***	2.21**	7.69***
			$h = 5$	73.17***		k=5	-7.45***	-0.84	3.88***
			$h \approx \ln(N)$	***		k=10	-7.55***	-1.76*	3.17***
$100((Y_{t/}\, Y_{t-1}) -1)$	-11.93***	1.94**	$h = 2$	29.16***	3.00***	k=2	3.00***	8.97***	7.69***
			$h = 5$	29.72***		k=5	0.00	4.44***	3.88***
			$h \approx \ln(N)$	59.78***		k=10	-1.11	2.34**	3.17***
$100((Y_t /MA)-1)$	-30.85***	0.08***	$h = 2$	***	***	k=2	***	***	97.00***
			$h = 5$	***		k=5	***	***	***
			$h \approx \ln(N)$	***		k=10	***	***	***
Nasdaq									
$Y_t - Y_{t-1}$	-18.71***	1.98	$h = 2$	1.27	2.62**	k=2	1.03	8.99***	12.97***
			$h = 5$	12.17**		k=5	0.60	7.79***	13.22***
			$h \approx \ln(N)$	***		k=10	-2.62***	7.29***	13.90***
$100((Y_{t/}\, Y_{t-1}) -1)$	-14.59***	1.87***	$h = 2$	32.55***	5.68***	k=2	5.68***	13.94***	12.97***
			$h = 5$	42.89***		k=5	5.39***	12.34***	13.22***
			$h \approx \ln(N)$	57.91***		k=10	3.48***	10.79***	13.90***
$100((Y_t /MA)-1)$	-26.34***	0.06***	$h = 2$	***	***	k=2	89.22***	89.19***	80.29***
			$h = 5$	***		k=5	***	155.39***	***
			$h \approx \ln(N)$	***		k=10	***	***	***

The test statistics show *** if significant at the 1% level, ** if significant at the 5% level and * if significant at the 10% level.
Test statistics greater than 100 or less than -100 are not shown since they are highly significant at the 1% level: SI = Strandberg and Iglewicz (2013a) test, DW = Durbin and Watson (1950) test, LB = Ljung and Box (1978) test, MV1 = the Chow and Denning (1993) MV1 test, M1 = the Lo and MacKinlay (1988) M1 test, R1 and S1 are tests from Wright (2000)

5. Conclusion

In this paper we summarize and compare popular tests for detecting randomness in time series data. Simulated distributions, alternative models and a practical stock market data

application are included in our study for all considered tests. When deciding among tests, consideration must not only be given to meeting size requirements for the null hypothesis, but also possible alternative hypotheses as power can differ considerably across alternatives. In our simulation study all tests met the size requirement over the varied considered null cases, except S_1 which exhibits high values for the χ_4^2 distribution. In the alternative hypotheses, all tests perform well for changes in mean, except SI, which has noticeably lower power when N = 300. When N is large, only SI has power to detect changes in variance. In the first correlated case, C1, all procedures did well for larger samples, except S_1 which has low power for this case. In the other correlated case, C4, only SI has high power when N = 10,000, with MV_1 showing respectable power.

For our stock market index data analysis we additionally review and compare three transformations and demonstrate that the choice of transformation can have a noticeable effect on test results. Only the MMPC transformation results in consistent rejection of the null hypothesis with high power for all tests and stock market indices, while the test by Strandberg and Iglewicz, SI, is the only test that is able to strongly reject the null hypothesis for all three studied indices and considered transformations. We note with interest that for most other considered tests, results can differ greatly depending on the transformation used.

In summary, this paper provides a simulation study that considers a number of null and alternative cases along with a useful analysis using stock market data to show the advantages and disadvantages of each studied test. These extensive comparisons and results are valuable, especially when dealing with financial data and tests of randomness.

References

Bandyopadhyay, U., Biswas, A., and Mukherjee, A. (2008). Controlling Type-I Error Rate in Monitoring Structural Changes Using Partially Sequential Procedures. *Communication in Statistics – Simulation and Computation* 37, 3:466-485.

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

Chen, W. W., and Deo, R. S. (2006), The Variance Ratio Statistic at Large Horizons, *Econometric Theory,* 22, 206-234.

Chow K.V. and Denning KC. (1993). A Simple Multiple Variance Ratio Test. *Journal of Econometrics,* 58:385-401.

Cizeau, P., Potters, M., and Bouchaud, J. (2001), Correlation Structure of Extreme Stock Returns, *Quantitative Finance,* 1, 217-222.

Durbin, J., and Watson, G. S. (1950), Testing for Serial Correlation in Least Squares Regression: I, *Biometrika,* 37, 409-428.

Harvey, A. C. (1990), *The Econometric Analysis of Time Series* (2nd ed.), Cambridge, MA: MIT Press.

Hochberg Y. (1974). Some Generalizations of T-method in Simultaneous Inference. *Journal of Multivariate Analysis,* 4:224-234.

Hoaglin, D. C. (1985), Summarizing Shape Numerically: The g-and-h Distributions, in *Exploring Data Tables, Trends, and Shapes: Robust and Exploratory Techniques*, Hoaglin, D. C., Mosteller, F., Tukey, J. W. (eds.), Hoboken, NJ: John Wiley & Sons, Inc., pp 461-513.

Kim, J.H. (2010), vrtest: Variance Ratio tests and other tests for Martingale Difference Hypothesis. R package version 0.95. http://CRAN.R-project.org/package=vrtest (accessed July 14, 2011).

Lo, A. W., and MacKinlay, A. C. (1988), Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test, *The Review of Financial Studies*, 1, 41-66.

Ljung G.M. and Box G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65:297-303.

Martinez, J., and Iglewicz, B. (1984), Some Properties of the Tukey g and h Family of Distributions, *Communications in Statistics-Theory and Methods*, 13, 353-369.

Mukherjee, A., and Bandyopadhyay, U. (2011). Some Partially Sequential Nonparametric Tests for Detecting Linear Trend. *Journal of Statistical Planning and Inference* 8: 2645-2655.

Strandberg, A.G., and Iglewicz, B. (2013a), A Nonparametric Test for Deviation from Randomness with Applications to Stock Market Index Data. *Communications in Statistics- Simulation and Computation*, 42, 689-697.

Strandberg, A.G., and Iglewicz, B. (2013b), Detecting Randomness: A Review of Existing Tests with New Comparisons. *Communications in Statistics- Simulation and Computation*, to Appear.

Tukey, J. W. (1977), Modern Techniques in Data analysis. NSF-sponsored regional research conference at Southeastern Massachusetts University, North Dartmouth, MA.

Tsay R.S. (2001). *Analysis of Financial Time Series*, Hoboken, NJ: John Wiley & Sons, Inc. pp 472.

Wright J.H. (2000). Alternative Variance-Ratio Tests Using Ranks and Signs. *Journal of Business & Economic Statistics*, 18:1-9.