A NONPARAMETRIC TEST FOR DEVIATION FROM RANDOMNESS

A Dissertation Submitted to the
Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

---

by
Alicia Graziosi Strandberg
January 2013

Dissertation Advisory Committee
Dr. Boris Iglewicz, Temple University, Statistics Department, Chair

Dr. Pallavi Chitturi, Temple University, Statistics Department
Dr. Yuexiao Dong, Temple University, Statistics Department
Dr. Inna Chervoneva, Thomas Jefferson University, (Outside Reader)

ABSTRACT

A NONPARAMETRIC TEST FOR DEVIATION FROM RANDOMNESS

Alicia Graziosi Strandberg

Doctor of Philosophy

Temple University, 2012

There are many existing tests used to determine if a series consists of a random sample. Often these tests have restrictive distributional assumptions, size distortions, or low power for key useful alternative situations. The interest of this dissertation lies in developing an alternative nonparametric test to determine whether a series consists of a random sample. The proposed test detects deviations from randomness, without a priori distributional assumption, when observations are not independent and identically distributed (*i.i.d.*), which is suitable for our motivating stock market index data. Departures from *i.i.d.* are tested by subdividing data into subintervals and then using a conditional probability measure within intervals as a binomial test. This nonparametric test is designed to detect deviations of neighboring observations from randomness when the data set consists of time series observations. Simulation results confirm correct test size for varied distributions and good power for detecting alternative cases. This test is compared to a number of other popular methods and shown to be a competitive alternative. Although the proposed test may be applicable to multiple areas, this dissertation is mostly interested in applications to stock market and regression data. The proposed test is effectively illustrated with the common three stock market index data sets using a newly created transformation, and shown to perform exceptionally well.

ACKNOWLEDGMENTS

I would like to express my deep appreciation and gratitude to my advisor, Dr. Boris Iglewicz for his guidance, patience, and dedication to my work. He is a great mentor and I am fortunate to have had the opportunity to work with him. I would also like to thank my committee members, Drs. Pallavi Chitturi, Yuexiao Dong, and Inna Chervoneva for their support, encouragement, and feedback. I have learned much though our conversations. All the professors whom I took classes with while obtaining my degree have generously shared their time and ideas, and for that I am grateful.

My family has been very supportive and lovingly while I pursed my degree. My parents, Bob and Roseann Graziosi, were my first teachers and taught me many valuable lessons. I credit them and thank them for instilling in me a strong work ethic. I would like to thank my husband, Jim Strandberg, for his encouragement, especially during times of great frustration. I truly appreciate all the sacrifices he has made so that I could complete this degree. Lastly I would like to thank my son, James Strandberg, for bring great joy to our life and giving me the final push I needed to finish my degree.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Consider the series $Y_1, Y_2, Y_3, \ldots, Y_N$. Interest centers on determining whether this series consists of a random sample. That is, are these random variables independent and identically distributed (*i.i.d.*). Deviations from the *i.i.d.* case can consist of independent observations that have differing location or scale or both. Alternatively, they may be dependent. If dependent, there are many ways that this can happen. The classic book on ARIMA models by Box and Jenkins (1976) studied a variety of ways to investigate and model dependency structures between time series observations.

There are many existing tests used to determine if a time series consists of a random sample. Often these tests have restrictive distributional assumptions, size distortions, or low power. The interest of this dissertation lies in developing an alternative nonparametric test to determine whether a series consists of a random sample. A competitive and technically relatively simple nonparametric test is proposed for testing departures from *i.i.d.* variables. The proposed relatively simple procedure is surprisingly shown to work well for varied symmetric and skewed distributions with reliance on an upper and lower percentile and straightforward conditional binomial probability, without a priori distributional assumption. The proposed test is then empirically compared with popular existing tests and shown to be nicely competitive. In our comparison varied sample sizes and distributional models will be considered. Existing research comparing similar tests often compares each test to a specific stock or index, see for example Charles and Darné (2009) and Al-Khazali, Ding and Pyun (2007). This dissertation adds to the discussion by comparing each considered test to a variety of known distributions and models, thus independently evaluating their effectiveness.

There are many tests that can be used to determining whether a series consists of a random sample. A well known classical test to determine departures of regression residuals from *i.i.d.* was introduced by Durbin and Watson (1950). Although this test is still popularly used, it has drawbacks consisting of its restricted conditions of normality and serial correlation. Ali and Sharma (1993) showed the Durbin and Watson test procedure does not meet the size requirement when the data are heavily skewed. Our independent investigation confirmed this finding. As an alternative to the Durbin and Watson test, one can also consider the Box and Pierce (1970) or Ljung and Box (1978) tests of the adequacy of a specific ARIMA model. However both tests assume the residuals of the ARIMA model are normally distributed. In addition there are a number of variance ratio tests found in the economic and finance literature, such as Lo and MacKinlay (1988), Wright (2000) and Chen and Deo (2006). These variance ratio tests are reviewed in Charles and Darné (2009). However variance ratio tests are often based on a random walk which assumes the conditional mean and variance are linear in time, a condition that may not always be reasonable. In addition, most variance ratio tests are based on an asymptotic normal distribution, yet the sampling distribution of the variance ratio test statistic is known to be skewed (Chen and Deo, 2006, Charles and Darné, 2009). The proposed test is nonparametric and designed to work well for time series data sets, specifically stock market data.

The proposed test will be evaluated in an applied application using stock market index data. These data are not normally distributed; rather they are nonstationary with typically an increasing long term trend. These data are unique because such data often consists of structural breakdowns where a relatively large number of time series observations follow a regular pattern for periods of time followed by highly irregular periods with changes in mean and variance (Chu, Stinchcombe and White, 1996 and Bandyopadhyay, Biswas, and Mukherjee, 2008). Sometimes the causes for such irregularities are easily understood, such as after 9/11 – the September 11, 2001 terrorist attacks caused global stock markets to decrease in value, when US markets reopened on September 17, 2001 stock prices were decreasing. For example, the Dow Jones Industrial Average, DJIA, experienced a drop of 14.3% during the week following its reopening. Although, for other times such insights are more difficult, as during October 1987 when stock markets worldwide quickly and dramatically decreased in value; as on October 19, 1987, when the DJIA dropped by 22.6%. Even after making a practical transformation the distributions of these stock market index data sets are often noticeably skewed.

The dissertation test is especially suitable for such situations as it is based on unusual rather than typical observations. This test subdivides data into subintervals with equal number of time series observations, such as trading weeks for stock market data, and then utilizes a conditional probability measure within intervals as a binomial test for departures from randomness. The proposed test assumes constant mean and variance but will not require a normality assumption or a priori knowledge of the distribution generating these data. Our motivating application involves large N consisting of daily closing stock market index data. Large sample data sets are becoming more common in other statistical areas, such as regression and linear and nonlinear models. We also apply the proposed test to the standardized residuals of ordinarily least square regression models. In this application the proposed test continues to perform well. Somewhat surprising for an asymptotic based test, the size requirements of the proposed procedure are shown to be met for sample –sizes as low as 100.

In addition to the proposed test, a new data transformation appropriate for stock market closing values is also introduced. This transformation is a modified measure of percent change (MMPC), similar to a stock return. In our stock index data analysis study, the proposed transformation, MMPC was the only considered transformation that led to consistently rejecting the null hypothesis with high power for all tests and stock market indices.

This dissertation is organized as follows. In Chapter 2 a literature review is given. Section 2.1 discusses the random walk model, Sections 2.2 to 2.5 explain relevant tests, then Section 2.6 details the g- and -h distribution and stock market data is discussed in Section 2.7. In Chapter 3 the proposed test is presented with some basic results along with a new proposed data transformation. Details about the proposed test and simulation result comparisons with the popular Lo and MacKinlay (1988) M_2 test are given in Sections 3.1 and 3.2, respectively. The proposed data transformation is further explained in Section 3.3, followed by a stock market data analysis in Section 3.4. In Chapter 4 a detailed comparison study is preformed that includes a number of existing tests. First, in Section 4.1, simulation results are discussed. Then Section 4.2 details a stock market data analysis. In Chapter 5 our proposed test is applied to the standardized residuals of ordinarily least square regression models. Finally, in Chapter 6 a summary of our results and future research is discussed.

CHAPTER 2
LITERATURE REVIEW

In this chapter, both statistical and financial research is briefly reviewed as it pertains to the proposed test and dissertation.

2.1 RANDOM WALK MODEL

When studying deviations from randomness in financial research, applications involving stock market data are often considered to determine the market's predictability. Most investors believe that if evidence of non-randomness exists, greater returns can be realized without taking on greater risks. Many researchers considered methods based on historical data to forecast future returns, this approach is known as a technical analysis (Fama, 1965, Malkiel, 2003). If stock market's returns fluctuate randomly, future values are unpredictable; hence future stock market returns are independent of past returns. Stock market returns that fluctuate randomly follow the efficient market hypothesis (EMH). According to Jensen (1978) "a market is efficient with respect to information set θ_t if it is impossible to make economic profits by trading on the basis of the information set θ_t." In a technical analysis, θ_t is the information contained in historical data of the market at time *t*. When the EMH is not rejected, investors are unable to earn above-average returns without incurring above-average risks. Therefore, a technical analysis cannot successfully predict future stock market returns since future returns are independent of past returns.

Evidence of the EMH can be studied using a random walk model (Working, 1934). The term random walk was first introduced by Pearson (1905). Earlier Brown (1828) defined Brownian motion, the random movement of tiny particles suspended in fluid, and Einstein (1905) used Brownian motion to develop a mathematical theory that provides powerful evidence for the existence of atoms (Ruppert, 2004). Even before Einstein, Bachelier's (1900) dissertation studied random processes to model financial speculation. Unfortunately his work was overlooked for over 50 years until rediscovered, translated into English, and published in a book of collected papers in *The Random Character of Stock Market Prices* (Cootner, 1964).

Consider the following random walk model

$$Y_t = \mu + Y_{t-1} + \varepsilon_t \qquad (1)$$

where Y_t is the value of an index at time t, μ is an unknown arbitrary drift parameter and ε_t is a disturbance term. Using Campbell, Lo, and MacKinlay's (1997) notation, the simplest version of a random walk, version RW1, is when ε_t is *i.i.d.*(0, σ_ε^2). The assumptions on ε_t in RW1 are relaxed in RW2, such that these ε_t are independent but not identically distributed. This allows for unconditional heteroscedasticity in the ε_t's, or time variation in volatility of stock market index returns (Campbell, Lo and MacKinlay, 1997). The RW3 model further relaxes the assumptions about the distribution of ε_t, such that dependent but uncorrelated increments are allowed. A time series defined in (1) where $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$ for all $k \neq 0$, but $Cov(\varepsilon_t^2, \varepsilon_{t-k}^2) \neq 0$ for some $k \neq 0$ has dependences with uncorrelated increments in the family of RW3 models. All random walk versions detailed above support the EMH and have conditional mean and variance at time t defined as $E[Y_t \mid Y_0] = Y_0 + \mu t$ and $Var[Y_t \mid Y_0] = \sigma^2 t$, respectively, where Y_0 is an initial value at $t = 0$. A random walk is often used to test the EMH.

Variance ratio tests of the random walk model are often used to test for evidence of predictability in financial markets – see Lo and MacKinlay (1988), Chow and Denning (1993), Wright (2000), and Chen and Deo (2006). In our research we have found variance ratio tests, specifically the ones considered in this dissertation, as being the most commonly used methods for testing deviations from randomness in financial studies. In addition to variance ratio tests, other methods can be used to test for lack of randomness including more traditional statistical tests such as Durbin and Watson (1950) and Ljung and Box (1978). In the remainder of this chapter, details about existing methods that influenced the proposed test are given. We consider these tests competitors of the proposed test even though the variance ratio tests are based on random walk models, while the other considered tests are based on a random sample null hypothesis. All of these tests will be compared to the proposed test in Chapter 4.

2.2 INDIVIDUAL VARIANCE RATIO TESTS

A number of the tests considered in this section are of the variance ratio type to determine lack of randomness of a series, $Y_1, Y_2, \ldots, Y_N$, by comparing the variance of the *k*-period return with *k* times the variance of the one-period return, that is $Var(Y_t - Y_{t-k}) = k\, Var(Y_t - Y_{t-1})$. The variance ratio of the *k*-period return is defined as $V(k) = \frac{Var(Y_t - Y_{t-k})/k}{Var(Y_t - Y_{t-1})} = 1 + 2\sum_{i=1}^{k-1}\frac{(k-i)}{k}\rho_i$ where ρ_i is the i^{th} lag autocorrelation coefficient or rather, a linear combination of the first $k - 1$ autocorrelation coefficients, with linearly declining weights (Charles and Darné, 2009). When $V(k) = 1$, $\rho_i = 0$ for all *i*, then returns are serially uncorrelated. Likewise when $V(k) \neq 1$, at least one $\rho_i \neq 0$. Then some autocorrelations between observations exit. Often test statistics are constructed based on an estimator,

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

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 $\hat{\mu}$, the estimated mean, defined as $\hat{\mu} = N^{-1}\sum_{t=1}^{N}(Y_t - Y_{t-1}) = N^{-1}(Y_N - Y_0)$. These tests have the potential advantages of being robust under homoscedasticity and heteroscedasticity, and have reasonable power against a wide range of alternative hypotheses. However some potential disadvantages exist. Variance ratio tests often use over lapping intervals when computing the long-horizon return, $\hat{\sigma}^2(k)$ in (2), as suggested by Lo and MacKinlay (1988). Although this can improve power, it also makes determining the exact distribution very difficult. Therefore asymptotic distributions are used. Most variance ratio tests are based on an asymptotic normal distribution, yet the sampling distribution of the test statistic can be skewed for small sample sizes. This has potential for leading to size distortions, low power and misleading inferences (Chen and Deo, 2006, Charles and Darné, 2009). Chen and Deo (2006) attempt to address this issue by suggesting a power transformation to better approximate the normal distribution, while Wright (2000) considers nonparametric tests with exact sampling distributions. These issues will be further clarified in our extensive simulation study summarized in chapter 4.

2.2.1 LO AND MACKINLAY (1988)

Lo and MacKinlay (1988) is the most referenced of the variance ratio tests found in the literature. They consider the relation between observations to be a random walk as defined in (1) with Var(ε_t) = σ_ε^2. They developed two variance ratio tests, M_1 and M_2. Under the null hypothesis both tests assume the random walk model in (1). M_1 assumes $\varepsilon_t \; i.i.d. \; N(0,\sigma_\varepsilon^2)$, while M_2 allows for changes in variance of ε_t as long as the increments are uncorrelated, $E(\varepsilon_t \, \varepsilon_{t-h}) = 0$ for any nonzero *h*, such as in a RW3 model. Since M_2 does not require a normal distributional assumption or *i.i.d.* observations for ε_t, it is considered a more robust test. Both tests are based on the same variance ratio estimator, *VR*(*k*) in (2), resulting in the following test statistics:

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

$$M_2 = \frac{VR(k)-1}{\phi^*(k)^{1/2}} \tag{4}$$

where $M_1 = M_2 = 0$ for $k = 1$, $\phi(k) = \frac{2(2k-1)(k-1)}{3kN}$ and

$$\phi^*(k) = \sum_{j=1}^{k-1}\left[\frac{2(k-j)}{k}\right]^2 \left(\frac{\sum_{t=j+1}^{N}(Y_t - Y_{t-1} - \hat{\mu})^2 (Y_{t-j} - Y_{t-j-1} - \hat{\mu})^2}{\left[\sum_{t=1}^{N}(Y_t - Y_{t-1} - \hat{\mu})^2\right]^2} \right)$$ for $k > 1$(Lo and MacKinlay, 1988, Charles and Darné, 2009). Under the null hypothesis, both M_1 and M_2 follow the standard normal distribution asymptotically.

2.2.2 WRIGHT (2000)

Wright (2000) created non-parametric variance ratio tests using ranks and signs with exact sampling distributions, therefore asymptotic distributions are not needed. For each test Wright uses a martingale difference sequence to define the null hypothesis, such that for a time series of asset returns, Y_1, Y_2, ..., Y_N,

$$Y_t = \mu + Z_t \quad (5)$$

where $Z_t = \sigma_t \varepsilon_t$. Under the null hypothesis for the rank-based tests, R_1 and R_2, Z_t is *i.i.d.* but with no additional distributional assumptions. These tests do not allow for any conditional heteroscedasticity and are comparable to the assumptions of M_1. The sign-based tests, S_1 and S_2, allow for conditional heteroscedasticity under the null hypothesis, but require conditional symmetry of ε_t given $\{Y_t, Y_{t-1}, Y_{t-2}, ..., Y_2, Y_1\}$, with mean equal to zero (Wright, 2000). This assumption allows S_1 and S_2 to be robust for many forms of conditional dependence, similar to the assumptions of M_2. In fact for both rank and sign-based tests, Wright claims his assumptions are neither stronger nor weaker than those of Lo and MacKinlay.

The rank-based test statistics, R_1 and R_2, are defined as

$$R_1 = \left(\frac{(Nk)^{-1} \sum_{t=k+1}^{N} \left(r_{1,t} + r_{1,t-1} + \ldots + r_{1,t-k} \right)^2}{N^{-1} \sum_{t=1}^{N} r_{1,t}^2} - 1 \right) \times \phi(k)^{-1/2} \text{ and}$$

$$R_2 = \left(\frac{(Nk)^{-1} \sum_{t=k+1}^{N} \left(r_{2,t} + r_{2,t-1} + \ldots + r_{2,t-k} \right)^2}{N^{-1} \sum_{t=1}^{N} r_{2,t}^2} - 1 \right) \times \phi(k)^{-1/2}$$

where $\phi(k)$ is defined in (3) and the standard ranks $r_{1,t}$ and $r_{2,t}$ are given as $r_{1,t} = \left(r(Y_t) - \frac{N+1}{2} \right) \Big/ \sqrt{\frac{(N-1)(N+1)}{12}}$ and $r_{2,t} = \Phi^{-1}\left(r(Y_t)/(N-1) \right)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution function and $r(Y_t)$ is the rank of Y_t for N observations among the series $Y_1, Y_2, Y_3, ..., Y_N$, such that under the null hypothesis $r(Y)$ is a random permutation of the numbers of 1, ..., N with equal probability (Wright, 2000). Both $r_{1,t}$ and $r_{2,t}$ are common transformations for ranks of a series. The critical values of these tests can be obtained by simulating their exact distributions.

When conditional heteroscedasticity exists, the exact distributions of R_1 and R_2 are unknown and a small amount of size distortion can exist. Although Wright shows both tests are still reasonable when testing alternatives with conditional heteroscedasticity, exact tests were developed using the signs of returns. Wright introduced two procedures, S_1 and S_2 analogous to R_1 and R_2 but found in simulation studies that S_2 was very conservative with low power and size distortion, see Wright (2000). Therefore only S_1 is considered here.

The sign-based test statistics, S_1, is defined as

$$S_1 = \left(\frac{(Nk)^{-1} \sum_{t=k+1}^{N} \left(s_t + s_{t-1} + \cdots + s_{t-k}\right)^2}{N^{-1} \sum_{t=1}^{N} s_t^2} - 1 \right) \times \phi(k)^{-1/2}$$

, where $\phi(k)$ is given in (3), and

$$s_t = \begin{cases} 1 & if\ Y_t > 0 \\ -1 & otherwise \end{cases}$$

. S_1 is a driftless model. Therefore, in addition to the requirements of its null hypothesis, $\mu = 0$ in (5). Similar to R_1 and R_2, the exact critical values of S_1 can be obtained through simulation.

2.3 MULTIPLE VARIANCE RATIO TESTS

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 (2) has a lower bound of zero. Therefore, when k is large most rejections of the test are due to the upper tail (Lo and MacKinlay, 1989). 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. When using multiple k values for individual tests the joint test size of the estimates must be controlled using multiple testing techniques. Chow and Denning (1993) and Chen and Deo (2006) developed multiple variance ratio tests that consider multiple k values and return one test statistic adjusted for the joint test size.

2.3.1CHOW AND DENNING (1993)

Chow and Denning (1993) developed two simple multiple variance ratio tests that consider multiple values of k and returns one test statistic properly adjusted for test size based on the Hochberg (1974) multiple comparison procedure. Their tests are an extension of the Lo and MacKinlay (1988) test statistics, such that they consider a set of m test statistics where m is the number of values of k. The two test statistics for the multiple joint hypothesis are $MV_1 = \max_{1 \le i \le m} |M_1|$ and $MV_2 = \max_{1 \le i \le m} |M_2|$, where M_1 and M_2 are defined in (3) and (4) respectively. To test the joint null hypothesis for a set of m test statistics, the null hypothesis is rejected if any one of the estimated variance ratio tests is significantly different from one. Using the Sidak (1967) probability inequality, the null hypothesis is rejected at the α level of significance if the variance ratio statistic is greater than the $[1 - (\alpha^*/2)]$ percentile where $\alpha^* = 1 - (1 - \alpha)^{1/m}$. Both test statistics, MV_1 and MV_2, follow the Studentized Maximum Modulus (SMM) distribution, SMM(α, m, N) with m and N degrees of freedom. As $N \to \infty$, SMM(α, m, N) $\to Z_{\alpha/2}$, that is, the asymptotic SMM critical value can be calculated from the standard normal distribution (Chow and Denning, 1993).

2.3.2 CHEN AND DEO (2006)

Chen and Deo (2006) consider power transformation of the variance ratio statistic when *k* is not too large compared to N, such as $k/N \leq 0.125$. This transformed test statistic helps resolve the skewness problem of the sampling distribution, and results in a better approximation to the normal distribution. Similar to M_2 and S_1 their individual variance ratio test is robust to conditional heteroscedasticity. Under the null hypotheses, Chen and Deo consider the random walk model in (1) where the disturbances, ε_t, are a martingale differences sequence with conditional heteroscedasticity such that ε_t and ε_{t-k} are roughly independent for large *k* lags. They then expand their individual test to consider multiple values of *k* by proposing a joint variance ratio test. Chen and Deo showed that their joint variance ratio test is able to obtain higher power than their individual tests. This joint test is labeled QP and used in our simulations and data analysis studies. We found this procedure far more complex and computer time consuming compared to other studied tests. Consequently, for our simulation study, using the vrtest package of R (Kim, 2010), we were able to simulate 100 replications per case with some difficulty versus 10,000 replication with ease for other tests. Inspection of the limited results did not provide evidence of special advantages of his procedure. As a result, we decided not to use these entries in the simulation study in Section 4.1, but do provide entries in the stock market data analysis in Section 4.2.

2.4 DURBIN AND WATSON (1950)

Durbin and Watson (1950) test for non-randomness in the residuals of an ordinary least squares regression equation with the test statistic,

$$d = \frac{\sum_{t=2}^{N}(e_t - e_{t-1})^2}{\sum_{t=1}^{N} e_t^2} \quad , \qquad (6)$$

where e_t is the t[th] residual and N is the number of observations. This test assumes *i.i.d.* N(0, σ^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 $Y_1, Y_2, Y_3, \ldots, Y_N,$ by substituting $Y_t = e_t$ in (6), assuming the assumption of *i.i.d.* N(0, σ^2) is valid.

2.5 LJUNG AND BOX (1978)

The Ljung and Box test is used to determine if a specific ARIMA model fits. The test statistic is

$$Q = N(N+2)\sum_{k=1}^{h}\frac{r_k^2}{N-k} \qquad (7)$$

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 e_t *i.i.d.* N(0, σ^2) . This test can be sensitive to the chosen value of *h* in (7).

2.6 THE G- AND -H DISTRIBUTIONS

In Section 3.2 and 4.1, random samples are simulated. For this purpose, Tukey's g- and -h family of distributions will be used for some of the cases (Tukey, 1977). The g- and -h distributional family discussed by Hoaglin (1985), Martinez and Iglewicz (1984), MacGillivray (1992) and others, can be used to create common distributional shapes. For Z a standard normal random variable, A, B, g and h specified constants, the g- and -h family of distributions is defined by the random variable $Y = A + BT_{g,h}(Z)$ where

$$T_{g,h}(Z) = \left(\frac{e^{gZ} - 1}{g}\right) e^{hZ^2/2}, \; -\infty < g < \infty, \text{ and } 0 \leq h. \qquad (8)$$

The parameters g and h allow this family to span over a wide region of skewness and kurtosis values, where g controls the skewness and h the kurtosis (Hoaglin, 1985). The parameters A and B account for location and scale, respectively.

Many known distributions can be approximated by the $T_{g,h}$(Z) family. For example, when g = h → 0 equation (8) simplifies to $T_{0,0}$(Z) = Z. This can easily be proven by considering the expansion of $\left(\frac{e^{gz} - 1}{g}\right) = Z + \frac{gZ^2}{2!} + \frac{g^2 Z^3}{3!} + \cdots$. When only h = 0, equation (8) yields a one parameter subfamily, the g-distributions $T_{g,0}(Z) = \left(\frac{e^{gZ} - 1}{g}\right)$, which are asymmetric with increasing skewness as |g| increases. Note that $T_{-g,0}(Z) = -T_{g,0}(-Z)$, therefore changing the sign of g does not change the level of skewness, only the direction (Hoaglin, 1985). When g is constant and positive, g-distributions are lognormal distributions.

When only g → 0, equation (8) yields the one parameter subfamily, the h-distributions, $T_{0,h}(Z) = Ze^{hZ^2/2}$, which are symmetric with increasing heavy tails, as h increases, as compared to the normal distribution (Martinez and Iglewicz, 1984). When h is restricted to only non negative real numbers, this provides a transformation that is monotonic for all Z. It is possible to use h < 0 in equation (8), but special consideration is needed because the monotonicity of $T_{g,h}$(Z) fails when $Z^2 > -1/h$ and the transformation is no longer one to one (Hoaglin, 1985). However for small negative values of h, say, $-0.08 \le h < 0$, the result is $Z^2 \ge 12.5$ or $|Z| > 3.5$, which may not be a serious constraint.

Due to its versatility, the g- and -h distribution is well-situated for simulation studies. Many known distributions can be approximated by the $T_{g,h}$(Z) family. A list of some popular distributions is included in Table 1 with corresponding g and h estimated values. Simulations in Sections 3.2 and 4.1 will utilize the g- and -h distributions guided by selections from Table 1.

Table 1: Values of Location, Scale, Skewness and Kurtosis for Selected g- and –h Distributions

Approximate Distribution	A	B	g→	h
Z	0.000	1.000	0.000	0.000
t with df=3	0.000	1.064	0.000	0.244
t with df=10	0.000	1.000	0.000	0.058
Heavy-Tailed Distribution	0.000	1.000	0.000	0.400
Chi Square with df=4	3.360	2.544	0.502	-0.046
Chi Square with df=6	5.348	3.237	0.406	-0.033
Lognormal	0.000	1.000	1.000	0.000

df = degrees of freedom, Martinez and Iglewicz (1984) Hoaglin (1985)

2.7 STOCK MARKET INDEX DATA

In an applied example, this dissertation will use stock market data obtained from three indices, 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). Stock indices are comprised of stocks from multiple companies. Therefore, pooling stocks in this manner helps ensure that investments are representative of the market rather than a few exceptional stocks. These specific stock indices were selected based on their popularity, longevity, and well established reputations. Also these data sets are large, not normally distributed, authentic and readily available, thus results can be reproduced by readers. Often these indices set benchmarks to track the performance of a collection of investments. Sizeable investments in retirement and other funds are modeled after these considered indices, thus making the study of these data especially relevant.

Yet each index is unique. The DJIA, the oldest index of the three and is available since October 1, 1928, consisting of 30 major American companies. By reputation it is considered the most conservative; that is, the least volatile. The S&P 500, started on January 3, 1950, consists of 500 leading companies in varied industries of the U.S. economy. With the addition of more diverse industries, this index is considered less conservative than the DJIA and a higher level of risk is perceived. The Nasdaq, the newest stock index of these three, available since February 5, 1971, incorporates over 3,000 newer, mostly high tech companies. Therefore, it has the reputation of being the least conservative of the three indices. It should be noted that only the Nasdaq does not have recorded closing values for September 26, 1973, October 7, 1974, and October 16, 1975. We only consider data for full 5 day weeks, with data ending on December 18, 2009.

CHAPTER 3
PROPOSED TEST AND SOME BASIC RESULTS

In this chapter, the proposed test is explained in detail and basic simulation results show its potential to be competitive with the Lo and MacKinlay M_2 test. A new proposed stock market data transformation is also explained and applied to both the proposed test and M_2. The work in this chapter follows closely the material of a paper accepted for publication in Communications in Statistics- Simulations and Computation (Strandberg and Iglewicz, 2012).

3.1 THE PROPOSED TEST

A procedure with minimal assumptions was an important consideration when developing the proposed test. Since data distributions are often unknown, developing a test that does not require a normality assumption or a priori knowledge of the distribution generating these data was highly desirable. Having a competitive yet technically relatively simple test was also important. Therefore the proposed test is based on a straightforward conditional binomial probability and test statistic that converges in distribution to the standard normal distribution. In addition, concentrating on observations outside upper and lower percentiles allows the proposed test to focus on tail observations, which are of special interest in the stock market investigations, rather than on the entire data set.

Let Y_1, Y_2, Y_3, …., Y_N be the considered series consisting of N observations and let $\mathbf{Y}$ = $\{Y_1, Y_2, Y_3 , …., Y_N\}$ be the set of these N observations. Notice that while we are dealing with a series, $\mathbf{Y}$ is a set. Consider a region R, such that $\mathbf{Y^*} =\{Y_t : Y_t \in R\}$ and $\mathbf{Y^{**}}$ is the complement of $\mathbf{Y^*}$. Assume that the null hypothesis, that these observations constitute a random sample, is true. Let N(**Y****) be the number of elements of **Y**** and $\Pr(Y_t \notin R) = \pi$. We can estimate π by $\hat{\pi}$ = N(**Y****)/N. Although we investigated a fair number of choices for the region R, we settled on the simple interval $[Y_{\{p\}}, Y_{\{1-p\}}]$, where $Y_{\{p\}}$ is the p^{th} sample percentile. Considerable simulations based on varied unimodal generating distributions showed that $[Y_{\{0.025\}}, Y_{\{0.975\}}]$ works well in terms of providing the appropriate significance level. Other chosen percentiles resulted in lower power and overly conservative test results, as seen in Table 2. Note that this choice of percentile is not directly related to the common significance level of $\alpha = 0.05$.

Next subdivide the N time series observations into M consecutive intervals each having an equal number of K observations, such that $M = \lfloor N/K \rfloor$ intervals, where $\lfloor \cdot \rfloor$ is the floor (largest integer) function and K is an integer that is small relative to N. If $N/K - M \neq 0$, then using M intervals leaves out $K(N/K - M) = b$ observations. We then recommend ignoring the first b observations in the time series. These b observations are considered an incomplete group of size less than K. Since N is large relative to K, having $b \neq 0$ should not be a problem. However, if it is believed critical time points are in the first few b observations, we would then suggest instead ignoring the last b observations. For our motivating example, we will consider complete weeks consisting of K = 5 days. Then N/K will always be an integer. For simplicity assume b = 0. For $i = 1, 2, \ldots, K$, $j = 1, \ldots, M$, denote $Y_i^{(j)} = Y_{K(j-1)+i}$ as the i^{th} observation in the j^{th} interval. For $j = 1, \ldots, M$, let $W_j = \sum_{i=1}^{K} I(Y_i^{(j)} \in \mathbf{Y^{**}})$, where $I(\cdot)$ is the indicator function.

Under the null hypothesis that $Y_1, \ldots, Y_N$ are *i.i.d.*, we have $W_1, \ldots, W_M$ *i.i.d.* Binomial (K, π). It follows that

$$P(W_j = 1 | W_j > 0) = \frac{K\pi(1-\pi)^{K-1}}{1-(1-\pi)^K}, \; j = 1, \ldots, M. \qquad (9)$$

We denote the above conditional probability in (9) by D. This suggests we consider all the intervals such that $W_j > 0$. Denote L = $\sum_{j=1}^{M} I(W_j > 0)$ and L_1 = $\sum_{j=1}^{M} I(W_j = 1)$, where I is the indicator function. Thus in L out of the M intervals, we have W_j greater than 0, or at least one observation within the jth interval belongs to the set **Y****. Similarly there are L_1 out of the M intervals, where exactly one observation belongs to the set **Y****.

Notice that the j^th interval with $W_j = 1$must also satisfy $W_j > 0$. Thus we can rewrite L_1 as $L_1 = \sum_{j=1}^{L} I(W_j = 1 | W_j > 0)$. Under the null hypothesis, it follows that $I(W_j = 1 | W_j > 0)$ is a Bernoulli random variable with success probability D = $P(W_j = 1 | W_j > 0)$. Consequently, L_1 follows Binominal (L, D). For large L, by central limit theorem, we know $\frac{(L_1/L) - D}{\sqrt{D(1-D)/L}} \xrightarrow{d} N(0,1)$ where "$\xrightarrow{d}$" means converge in distribution. For moderate L, the binomial approximation can be improved by using a correction factor with modified test statistic

$$Z = \frac{(L_1/L) - D + (cH/L)}{\sqrt{D(1-D)/L}} \qquad (10)$$

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 the useful correction factor needed to prevent the proposed test from being too conservative. Several different values of c were considered with results displayed in Table 3. This table can be used as a guide for choosing the constant c appropriate for a considered sample size. From this table we notice that c = 0.5 or 0.55 is typically a good choice, but smaller values of c are recommended for values of N in the interval $500 \leq N \leq 5000$.

3.2 COMPARISON OF THE PROPOSED TEST WITH THE LO AND MACKINLAY M_2 TEST

As an alternative to the proposed test, we considered the popular robust test of Lo and MacKinlay (1988), M_2, detailed in Section 2.2.1. We performed simulations using the common choice of $k = 2$. We additionally added a few entries for higher values of k for several correlated cases. For the proposed test we will be using K = 5 to mimic stock market data with full 5 day trading weeks, with the interval $[Y_{\{0.025\}}, Y_{\{0.975\}}]$ and c = 0.55 as these were shown to be reasonable values in Tables 2 and 3 respectfully. The value c = 0.55 is only used in this chapter and this is replaced with c = 0.50 in future chapters.

For our basic simulation study, we consider random samples generated from a variety of distributions. Some common distributional cases will be simulated using Tukey's g- and -h distributional family (Tukey, 1977) as detailed in Section 2.6. For our initial set of simulations using the proposed and M_2 tests, we will consider the g- and -h distributions detailed in Table 1. We will also consider mixtures of normal distributions. For theses mixture distributions, let

$$f(y) = (1 - p)f_1(y) + pf_2(y). \qquad (11)$$

Let $f_1(y)$ be N(0, 1) and $f_2(y)$ be a normal distribution with mean 0 and standard deviation 10, i.e. $N(0, 10^2)$ with $p = 0.01$ and 0.05. This model incorporates a small proportion of possible outliers but is still expected to meet the required test size.

In addition to the known distributions and models expected to preserve the test size, several alternative data models are also considered. We considered a model where 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 $N(0, s_1^2)$, $f_2(y)$ be $N(0, 1^2)$, and $f_3(y)$ be $N(0, s_2^2)$, while a model with constant variance and changing mean is created by letting $f_1(y)$ be $N(-t, 1^2)$, $f_2(y)$ be $N(0, 1^2)$, and $f_3(y)$ be $N(t, 1^2)$. Here we consider the cases s_1 = 0.80, s_2 = 1.20, then s_1 = 0.75, s_2 = 1.25, and t = 2 and 3. We consider another model where distributions of observations differ while the means and variances are constant and the observations are independent. The first model, D1, considers three distributions such that 30% of observations are $N(0,1^2)$, another 30% are χ^2_3, and the remaining 40% are from t_4. The second case, D2, considers four distributions such that 20% of observations are $N(0,1^2)$, 40% are χ^2_3, 30% are t_3, and the remaining 10% are χ^2_5. The χ^2 and t distributions are standardized before each test is preformed.

Next we consider three correlated cases, C1, C2, and C3. For these cases, motivated by stock market weekly data, observations are simulated from complete trading weeks where weeks with market closures or holidays are not included. Each correlated model is designed such that 90% of these weeks have observations that are *i.i.d.* $N(0,1^2)$ and a correlation structure exists for the remaining 10% of these weeks. Consider a typical member of the 10% correlated weeks. Let Y^M be the Monday value coming from $N(0, 10^2)$. The other generated values are Y^j, j = T, W, Th, F. For C1, $Y^T = \rho Y^M + \varepsilon$ where ε is $N(0,1^2)$ and the remaining three days are from *i.i.d.* $N(0, 1^2)$. For C2, $Y^F = \rho Y^M + \varepsilon$ and the remaining three days are from *i.i.d.* $N(0, 1^2)$. For C3, $Y^j = \rho Y^M + \varepsilon$, where j is randomly chosen from T, W, Th, F and the remaining three days are from *i.i.d.* $N(0, 1^2)$. In all three correlated cases we consider ρ = 0.90 and 0.80.

Table 4 summarizes the simulation results for the proposed test and also for the M_2 test, each based on simulated cases of 10,000 replications with 10,000 observations. The rejection rate is the percentage of tests out of 10,000 where p-value < 0.05. For the null cases, since the significance level is set at 5.00%, we expect rejection rates to be close to 5.00%. With 10,000 replications, the standard error of a 0.05 proportion of rejections, under the null hypothesis, is $\sqrt{(0.05\times 0.95)/10000}=0.0022$, therefore rejection rates in the range of $(0.05\pm 1.96\times 0.0022)\times 100=[4.57\%, 5.43\%]$ are expected. All considered null cases for the proposed test are within this range and therefore are reasonably close to the desired 5.00% rejection rate. For M_2 one considered case, the lognormal, is higher than the expected rejection rate range. For this heavily skewed lognormal case, M_2 is anti-conservative. In summary, all the null cases met the required $\alpha = 0.05$ for both SI and M_2, excluding the heavily skewed lognormal case where M_2 is noticeably anti-conservative.

When considering alternative cases, the proposed test provided high power, while M_2 was unable to do so for two situations, the case with changing variances and C2. Here, the latter case is most disturbing. For the correlated cases, similar results were realized between SI and M_2 only for C1 and C3 – although for C3 power is noticeably lower for M_2. For C2 only SI was able to show considerable power to correctly reject the null hypothesis, while rejection rates for M_2 where in the same range as the null cases. Only after increasing k in (4) from $k = 2$ to 5 does M_2 show reasonable power against the null hypothesis, however stronger evidence is still realized with SI. In this case, choosing the popular $k = 2$ option leads to total loss of power for M_2, while choosing $k = 5$ restores much of the lost power. In practice, the $k = 2$ option will very likely be chosen leading to reduction in power if the correlation is not between consecutive days. In summary, the proposed test works well irrespective of the correlation structure following a highly volatile day, without a priori knowledge of future dependencies, while procedures with set autocorrelation structures may not always be able to retain power in such cases.

When considering changing means, both SI and M_2, as expected, overwhelmingly rejected the null hypothesis. However when considering changing variances, only the proposed test was able to show considerable power to reject the null hypothesis, while rejection rates for

M_2 were in the same range as the null cases. This is not surprising, as the null hypothesis for M_2 consists of a random walk, thus tolerating changes in variance. This can be a liability of M_2 as more recent research suggests variance changes over time in the stock market (Ito, Lyons and Melvin, 1998). Consequently, the differing detection abilities of SI and M_2 for this case have practical implications for stock market data application and other possible uses. For two considered cases, D1 and D2 – where distributions differ but means and variances are the same – the powers are close to null case, however for D2 the power is somewhat elevated for SI. In summary, the proposed method works well for varied null and alternative situations, providing excellent power against a variety of alternatives to the null hypothesis.

Although this binomial based test is justified for large samples, surprising it meets size requirements and shows reasonable power for relatively small sample sizes, as seen in Table 5 when the proposed test is compared to M_2 with N = 300. Again here, for the null cases, we expect rejection rates to be close to 5.00%. In general the proposed test is slightly conservative for the considered null cases, while M_2 has a high rejection rate for the lognormal case.

When considering the alternative cases listed in Table 5 when N = 300, the proposed test is able to show greater power for all the correlated cases, C1, C2, and C3,while M_2 has very low power for C2 and C3 and only modest power for C1. SI is also able to show some power for cases with changing variances, while as expected M_2 considers these null cases. M_2 is able to show high power for alternative cases with changing means while at this small sample size, SI lacks power. Overall when N = 300 the proposed test does well, however some null cases are slightly conservative and SI exhibits noticeably lower power than M_2 in models with changing means; while for M_2 there is modest or no power for all alternative cases except models with changing means.

3.3 MODIFIED MEASURE OF PERCENT CHANGE TRANSFORMATION

We consider here the daily closing values of the DJIA, S&P 500, and Nasdaq as detailed in Section 2.7. Due to the nature of these data, it is more appropriate to not directly study the daily closing values, but rather a daily percentage changes. It is tempting to consider the commonly used daily percent change transformation given by

$$r_i(t) = 100\left(\frac{Y_t(i) - Y_{t-1}(i)}{Y_{t-1}(i)}\right) = 100\left(\frac{Y_t(i)}{Y_{t-1}(i)} - 1\right), \qquad (12)$$

where $Y_t(i)$ is the daily closing price for index i on day t (Cizeau, Potters and Bouchaud, 2001, Bandyopadhyay, Biswas and Mukherjee, 2008, Mukherjee and Bandyopadhyay, 2011), $i = 1, 2, 3$, for DJIA, S&P 500, and Nasdaq, respectively. The problem is that $r_i(t)$ and $r_i(t - 1)$ have $Y_{t-1}(i)$ in common, thus these transformed data are not independent. As an alternative we propose a modified measure of percent change,

$$\text{MMPC} = 100\left(\frac{Y_t(i) - MA^{(i)}_{(t-2q:t-q-1)}}{MA^{(i)}_{(t-2q:t-q-1)}}\right) = 100\left(\frac{Y_t(i)}{MA^{(i)}_{(t-2q:t-q-1)}} - 1\right) \qquad (13)$$

where MA is a delayed moving average of q observations such that $MA_{(t-2q:t-q-1)} = \dfrac{\sum_{j=t-2q}^{t-q-1} Y_t}{q}$. Data used in this moving average are q or more observations before $Y_t(i)$; we suggest setting $q = 10$ for large N, which provides a $q = 10$ washout period and a baseline moving average value replacing the previous observation. Consequently, it would be reasonable to test whether these transformed observations constitute a random sample. These considered data consist of full five day weeks only, excluding weeks containing holidays or having the market closed for other reasons. Due to the way we use moving averages, we lose 20 days, four weeks, at the beginning of the series. Since the first full week of DJIA data start on October 1, 1928, M = 3,462 weeks. Similarly, S&P 500 data start on January 9, 1950 with M = 2,581 weeks, and Nasdaq data start on February 8, 1971 with M = 1,696 weeks. The last considered observation for all three indices is December 18, 2009 (Yahoo! Finance 2010a,b,c).

3.4 BASIC STOCK MARKET DATA ANALYSIS

We use SI and M_2 to test the null hypothesis that these observations, transformed using the MMPC, constitute a random sample. Results are provided in Table 6. From Table 6 notice the very high absolute **Z** values for all considered cases resulting in *p*-values below 0.0001. (Exact *p*-values are not included in Table 6 because we do not claim that level of accuracy.) The null hypothesis is easily rejected for each of the three stock market indices and for both SI and M_2. Furthermore, **Z** values for SI are all negative, while **Z** values for M_2 are positive. The negative **Z** values from SI indicate that more than expected multiple unusual percent changes tend to occur within weeks containing at least one unusual percent change. The positive **Z** values for M_2 indicate that the variance of two time periods (since $k = 2$) is higher than expected. In summary, the proposed test worked nicely for each of these three data sets, as did the M_2, and showed conclusively that these modified measures of percent change do not constitute random samples.

Table 2: Percentage of Tests with *p*-value < 0.05 for Standard Normal Observations with Given Percentiles for Region R [$Y_{\{p\}}$, $Y_{\{1-p\}}$]

P	Percentage of Test with *p*-value < 0.05
0.005	4.52%
0.025	5.10%
0.050	3.55%
0.100	2.01%

For each percentile 10,000 trials of 10,000 Standard Normal observations was generated.

Table 3: Percentage of Tests Out of 10,000 Trials with *p*-values < 0.05 for Given Correction Factor, c, and Distribution

	c = 0.0	c = 0.1	c = 0.2	c = 0.3	c = 0.4	c = 0.5	c = 0.55	c = 0.6	c = 0.7
Standard Normal Distribution									
N = 100	1.97%	2.28%	2.27%	2.48%	3.51%	3.26%	3.42%	35.19%	36.73%
N = 300	3.99%	4.14%	4.16%	4.19%	3.99%	4.21%	4.25%	4.29%	17.44%
N = 500	3.39%	3.34%	3.50%	3.32%	3.35%	10.67%	10.77%	10.58%	11.39%
N = 1,000	2.07%	4.74%	5.65%	8.20%	8.40%	8.35%	8.46%	8.07%	8.15%
N = 2,000	2.82%	4.25%	5.52%	6.77%	6.61%	6.44%	7.21%	6.63%	6.75%
N = 5, 000	4.04%	3.83%	3.85%	4.13%	4.79%	5.40%	5.60%	5.56%	5.64%
N = 10,000	3.42%	4.21%	4.03%	4.64%	4.47%	4.98%	5.10%	5.37%	5.29%
t with 3 degrees of freedom									
N = 100	2.34%	2.30%	2.48%	2.01%	3.47%	3.49%	3.69%	35.41%	35.63%
N = 300	4.14%	4.01%	4.32%	4.02%	4.27%	4.06%	4.36%	4.20%	16.94%
N = 500	3.44%	3.69%	3.55%	3.37%	3.29%	10.66%	10.53%	10.34%	11.88%
N = 1,000	2.17%	4.84%	5.33%	8.17%	8.37%	8.09%	8.56%	8.01%	8.18%
N = 2,000	2.59%	4.34%	5.32%	6.94%	7.16%	7.00%	6.94%	6.82%	6.43%
N = 5, 000	4.07%	4.10%	4.14%	4.23%	4.60%	5.16%	5.16%	5.12%	6.33%
N = 10,000	3.61%	3.96%	4.18%	4.61%	4.25%	4.86%	4.67%	5.26%	5.37%
Chi Square with 4 degrees of freedom									
N = 100	2.14%	2.50%	2.49%	2.33%	3.67%	3.67%	3.36%	35.79%	35.58%
N = 300	4.15%	3.95%	4.04%	4.26%	3.77%	4.36%	4.53%	4.50%	16.71%
N = 500	3.74%	3.43%	3.84%	3.36%	3.52%	10.34%	10.56%	10.40%	12.27%
N = 1,000	2.24%	4.49%	5.65%	8.18%	8.05%	8.06%	8.29%	8.09%	8.32%
N = 2,000	2.46%	4.17%	5.83%	6.57%	6.94%	6.40%	6.46%	6.67%	6.59%
N = 5, 000	3.85%	4.02%	3.56%	3.82%	4.50%	4.92%	5.69%	5.43%	5.57%
N = 10,000	3.76%	3.86%	4.24%	4.46%	4.75%	5.28%	4.90%	4.77%	4.99%

Table 4: Simulation Rejection Rates of 10,000 Trials for Selected Distributions and Models when N = 10,000, c = 0.55 and k = 2

Null Distribution Cases	Rejection Rate	
	SI	M_2
Z	5.10%	4.68%
t with df=3	4.67%	5.29%
t with df=10	5.12%	5.05%
Wide-tailed distribution	4.61%	4.73%
Chi Square with df=4	4.90%	4.89%
Chi Square with df=6	4.87%	5.00%
Lognormal	4.93%	**5.81%**
Non g- and -h Null Distributions Cases		
Mixture Model with p = 0.01	5.12%	4.66%
Mixture Model with p = 0.05	5.09%	4.62%
Alternative Distribution Cases		
Changing Variance with $s_1 = 0.80$, $s_2 = 1.20$	76.70%	4.64%
Changing Variance with $s_1 = 0.75$, $s_2 = 1.25$	95.39%	5.28%
Changing Mean with $t = 2$	99.18%	100.00%
Changing Mean with $t = 3$	99.43%	100.00%
C1 - Monday and Tuesday Correlated in 10% of total weeks ρ=0.9	100.00%	100.00%
C1 - Monday and Tuesday Correlated in 10% of total weeks ρ=0.8	100.00%	100.00%
C2 - Monday and Friday Correlated in 10% of total weeks ρ=0.9	100.00%	5.26% k=2
		88.36% k=5
C2 - Monday and Friday Correlated in 10% of total weeks ρ=0.8	100.00%	5.17% k=2
		88.23% k=5
C3 - Monday and Random Day Correlated in 10% of total weeks ρ=0.9	100.00%	97.48%
C3 - Monday and Random Day Correlated in 10% of total weeks ρ=0.8	100.00%	97.43%
D1 -Changing Distribution with 3 Standardized Distributions	4.88%	5.10%
D2- Changing Distribution with 4 Standardized Distributions	10.08%	5.18%

Table 5: Simulation Rejection Rates of 10,000 Trials for Selected Distributions and Models when N = 300, c = 0.5 and k = 2

Null Distribution Cases	Rejection Rate SI	M_2
Z	4.21%	5.36%
t with df=3	4.06%	5.20%
Chi Square with df=4	4.36%	5.20%
Lognormal	4.21%	**7.78%**
Non g- and -h Null Distributions Cases		
Mixture Model with p = 0.01	3.86%	5.04%
Mixture Model with p = 0.05	4.42%	5.12%
Alternative Distribution Cases		
Changing Variance with $s_1 = 0.80$, $s_2 = 1.20$	10.52%	5.28%
Changing Variance with $s_1 = 0.75$, $s_2 = 1.25$	13.63%	5.08%
Changing Mean with $t = 2$	16.20%	100.00%
Changing Mean with $t = 3$	17.05%	100.00%
C1 - Monday and Tuesday Correlated in 10% of total weeks ρ=0.9	92.77%	14.91%
C1 - Monday and Tuesday Correlated in 10% of total weeks ρ=0.8	91.41%	15.05%
C2 - Monday and Friday Correlated in 10% of total weeks ρ=0.9	93.21%	6.33%
C2 - Monday and Friday Correlated in 10% of total weeks ρ=0.8	91.59%	5.50%
C3 - Monday and Random Day Correlated in 10% of total weeks ρ=0.9	92.99%	4.00%
C3 - Monday and Random Day Correlated in 10% of total weeks ρ=0.8	91.58%	3.84%

Table 6: **Z** Values for Stock Index Test Results

	DJIA	S&P 500	Nasdaq
SI (c =0.55)	-25.97	-26.37	-29.01
M_2 (k =2)	36.09	43.73	38.62

CHAPTER 4
DETAILED COMPARISON STUDY

In this chapter we review and compare a number of existing tests for detecting randomness in time series data, with emphasis on stock market index data. The investigated tests, detailed in Chapter 2, and the proposed test, are compared over a diverse group of distributions, models, and stock market applications. We will view these tests as competitive, even though the variance ratio tests are based on random walk models, thus not sensitive to changes in variance, while the other considered tests are based on a random sample null hypothesis. 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. In our stock market data analysis, we will show that the choice of data transformation can have a noticeable effect on test results.

4.1 SIMULATION RESULTS

In this section the basic simulation results discussed in Section 3.2 are now expanded and applied to all the tests detailed in Chapter 2. We will consider random samples of N = 100, 300, and 10,000 observations and will again consider the g - and - h distributions, and a mixture of normal distributions. Using the g- and -h distributional family we obtain or approximate five distributions: Z, t_3, lognormal, χ_4^2, and a heavy-tailed distribution with $g \rightarrow 0$ and $h = 0.40$. Unlike in Section 3.2, here the lognormal and χ_4^2 distributions are standardized before performing each test. For the mixture of normal distributions, let $f_1(y)$ be $N(0, 1^2)$ and $f_2(y)$ be $N(0, 10^2)$, with $p = 0.30$ in (11). Next we will consider a new mixture of normal distributions that represent a symmetric bimodal distribution. For this purpose, let $f_1(y)$ be $N(-1.5, 1^2)$, and $f_2(y)$ be $N(1.5, 1^2)$ with $p = 0.50$ in (11). Although these mixture models are not generating observations from the same distribution parameters, they are still expected to meet the required test size for all tests.

As in Section 3.2, in addition to the known distributions and models expected to preserve the test size, several alternative data models are considered. From section 3.2 we consider the following models: constant mean and changing variance ($s_1 = 0.75$ and $s_2 = 1.25$); constant variance and changing mean ($t = 2$); C1; C2; C3; D1; D2. A new correlated case, C4, is also considered. Here the correlation structure is present over two weeks, such that Y^M in the first week, Y_1^M, has a value from $N(0, 2^2)$. 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, the remaining days are generated from *i.i.d.* $N(0, 1^2)$. In all four correlated cases $\rho = 0.90$ with Y^M as $N(0, 2^2)$.

In addition to the models detailed in Section 3.2, two models by other authors are included. The first model, A1, is a stochastic volatility model of conditional heteroscedasticity considered by Wright (2000) and Lo and MacKinlay (1989). Here $Y_t = e^{h_t/2}\varepsilon_t$ where $h_1 = 1$, $h_t = 0.95h_{t-1} + \xi_t$ for $t > 1$, ξ_t ~*i.i.d.* $N(0, (1/10)^2)$, and ε_t ~*i.i.d.* $N(0, 1^2)$. In previous simulations, A1 has shown results similar to the null case for most of the variance ratio tests considered. The second model, A2, is an alternative mean reverting process from Chen and Deo (2006) and Lo and MacKinlay (1989), where $Y_1 = 1, Y_t = 0.92Y_{t-1} + \varepsilon_t$ for $t > 1$ and ε_t ~*i.i.d.* $N(0, 1^2)$. In previous simulations, most of the variance ratio tests considered show reasonable power against mean reverting processes – where, for example, stock prices return or revert eventually to the average price over time.

For each simulation we generated 10,000 random samples of N = 100, 300, and 10,000 for the given distribution or model. Each individual variance ratio test is evaluated with $k = 2, 5,$ and 10, likewise, each multiple variance ratio test is evaluated with $m = 3$ where the three values selected for k are 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.

Tables 7, 8 and 9 list the simulation results for the distributions and models expected to meet the required test size for the null hypothesis when N = 100, 300 and 10,000, respectively. For emphasis some noteworthy results are in bold. Since we set tests at 5% significance level, we expect these table entries to be close to 5.00%. In general most of the tests considered meet this requirement, with some exceptions. Of note are the very high entries for skewed distributional cases of the Wright (2000) S_1 procedure and the slightly anticonservative entries for the Lo and MacKinlay (1988) M_2 lognormal cases when N = 100. Additionally, the Chow and Deming (1993) MV_1 and MV_2 entries are a bit conservative. In summary, excluding the S_1 test for some skewed cases, the considered tests work reasonably well at meeting test-size requirements.

Tables 10, 11 and 12 summarize the simulation results for the alternative models considered when N = 100, 300 and 10,000, respectively. As before, some noteworthy results are in bold. As expected, greater power is generally seen as N increases, however power can differ considerably across alternatives. Only one test, the proposed test, is able to show considerable power against changing variances and the stochastic volatility model, A1, especially for large values of N. This is not surprising, as the null hypothesis for variance ratio tests allows for changes in variance and conditional heteroscedasticity. It should be noted that information regarding short term changes in variance can be informative in short term and options trading. Although the Durbin and Watson (1950) and Ljung and Box (1978) tests do show modestly elevated rejection rates for these variance changing cases, only SI shows high rejection rates, especially when N is large. For the changing means case, all test do well with the exception of SI, which does poorly when N =100. When considering differing distributions with constant means and variances and N large, most procedures consider these as null cases, with the exception of S_1, with large N, that provides reasonable power and R_1 that provides modest power. For the four considered correlated cases with larger values of N only SI maintains high power for each case, with MV_1 and MV_2 providing reasonable power. The exception is the serially correlated C1 case with large sample sizes, where all but S_1 provide high power. Overall, S_1 does poorly for these correlated cases. For the stochastic volatility (A1) and mean reverting process (A2) models, only SI, with N large, has reasonable power for the A1 case, while all considered tests have high power for A2.

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, the SI test is the only procedure with high power for all alternative cases, excluding the two consisting of random observations from different distributions. In particular, if detecting changes in variance is a concern, the SI test should be used. In addition to SI for correlated cases, MV_1 and MV_2 do reasonably well. However we do not find evidence that Wright's exact tests are more powerful than asymptotic variance ratio tests against models with serial correlations as suggested by Charles and Darné (2009).

4.2 STOCK MARKET DATA ANALYSIS

As an alternative comparison of the considered tests, we summarize an analysis of daily closing values of stock market indices using the DJIA, S&P 500 and Nasdaq as detailed in Sections 2.7 and 3.4. In this section we evaluate randomness of each index using three different transformations, which include two common transformations and our proposed transformation. Consider the following transformations: lag 1 closing daily stock market index differences, $Y_t - Y_{t-1}$; daily percentage change, as defined in (12); the proposed MMPC transformation defined in (13). Results are included in Table 13. To assist in the readability of Table 13, after each test statistics * is added if significance is at the 10% level, ** is added if significance is at the 5% level and *** is added if significance is at the 1% level. Extremely high test statistic values contain only *** with blanks for associated numbers.

In Table 13, test statistics differ between transformations and indices, with the exception of SI which is significant at the 1% level for all indices and transformations. Also, S_1 is significant at the 1% level for all cases when $k = 2$, but can be non significant at 10% for several other values of k. The latter is not entirely surprising since S_1 is based on signs. Among transformations, only the MMPC results in consistently rejecting the null hypothesis at the 1% significant level for all tests, while the other two transformations show inconsistent rejection levels among tests and indices. A smaller number of rejected cases are seen with *QP*. M_1 and M_2 have somewhat similar results with M_1 showing slightly more rejected tests and higher rejection levels. As expected, this is also true for MV_1 and MV_2. R_1, R_2, and S_1 often show rejection at 1%, but can have conflicting results depending on the chosen value of *k*.

In summary, it is interesting to observe that for most tests, results can differ depending on the transformation used and that test results based on the MMPC transformation are all significant at the 1% level. If early stock market studies used the third rather than the first transformation, it is possible that some non significant conclusions would have been declared as significant. The only test that does not show evidence of this concern in this study is the SI test as it rejects the null hypothesis at the 1% significance level for all transformations.

As seen in both our simulations and applied results, individual variance ratio tests can show very different results depending on the chosen value of *k*. This is also true for the Ljung and Box test with respect to the chosen value of *h*. The only transformation that shows consistent results among individual tests for each considered data set is the MMPC. All other tests and transformations had inconsistent results among the considered cases. In summary, when dealing with financial data, the test procedure used and choice of transformation can play a key role in resulting conclusions.

Table 7: Simulation Rejection Rates Comparisons for Null Cases N=100

	SI	DW		LB	MV1	MV2		M1	M2	R1	R2	S1
Distribution Cases from the g- and -h distribution												
Z	3.40%	5.21%	*h* = 2	4.89%	**2.82%**	**3.09%**	*k*=2	5.02%	5.78%	5.06%	5.36%	4.37%
			h = 5	4.92%			*k*=5	4.65%	4.97%	4.95%	4.83%	4.84%
			h ≈ ln(N)	5.35%			*k*=10	2.41%	3.04%	5.38%	5.21%	4.73%
t_3	3.40%	4.91%	*h* = 2	4.20%	**2.51%**	**3.43%**	*k*=2	4.45%	5.51%	5.23%	5.27%	4.95%
			h = 5	5.05%			*k*=5	3.94%	5.74%	5.23%	5.29%	5.14%
			h ≈ ln(N)	4.24%			*k*=10	2.20%	4.10%	4.67%	5.35%	4.59%
Lognormal Standardized	3.28%	5.23%	*h* = 2	3.82%	**2.64%**	**8.69%**	*k*=2	4.08%	**8.49%**	5.25%	5.16%	**29.90%**
			h = 5	4.07%			*k*=5	3.43%	**10.32%**	4.97%	5.11%	**60.87%**
			h ≈ ln(n)	4.24%			*k*=10	1.76%	**10.98%**	5.12%	5.32%	**73.74%**
Heavy-tailed Distribution	3.52%	4.64%	*h* = 2	4.00%	**2.95%**	5.03%	*k*=2	4.63%	5.19%	5.29%	4.88%	4.32%
			h = 5	4.17%			*k*=5	4.43%	6.62%	4.98%	4.68%	5.42%
			h ≈ ln(N)	4.39%			*k*=10	1.99%	7.00%	5.24%	5.38%	4.48%
X^2_4 Standardized	3.64%	4.68%	*h* = 2	4.60%	**2.40%**	**3.77%**	*k*=2	4.87%	6.06%	4.94%	5.00%	**6.68%**
			h = 5	4.76%			*k*=5	4.10%	5.37%	5.29%	4.93%	**10.73%**
			h ≈ ln(n)	4.43%			*k*=10	2.36%	3.98%	4.96%	5.36%	**14.83%**
Model Cases												
Mixture Model	4.95%	5.03%	*h* = 2	4.88%	**2.99%**	**2.79%**	*k*=2	5.18%	4.71%	4.99%	5.40%	4.34%
			h = 5	4.84%			*k*=5	4.07%	5.44%	5.20%	5.00%	4.80%
			h ≈ ln(N)	4.74%			*k*=10	2.34%	4.11%	5.13%	5.54%	4.55%
Bimodal	3.28%	4.89%	*h* = 2	4.67%	**2.82%**	**2.98%**	*k*=2	5.83%	4.89%	5.52%	4.90%	4.38%
			h = 5	5.39%			*k*=5	4.21%	4.65%	5.04%	5.13%	5.21%
			h ≈ ln(N)	5.47%			*k*=10	2.30%	2.98%	5.04%	5.34%	4.69%

SI = Strandberg and Iglewicz (2012) test with c = 0.5, DW = Durbin and Watson (1950) test, LB = Ljung and Box (1978) test,
MV1 and MV2 are tests from Chow and Denning (1993), M1 and M2 are test from Lo and MacKinlay (1988), R1, R2, and S1 are tests from Wright (2000)

Table 8: Simulation Rejection Rates Comparisons for Null Cases N=300

	SI	DW		LB	MV1	MV2		M1	M2	R1	R2	S1
Distribution Cases from the g and h distribution												
Z	4.21%	5.12%	*h* = 2	4.75%	**3.02%**	**3.34%**	k=2	5.18%	5.22%	4.83%	4.89%	3.98%
			h = 5	5.13%			k=5	4.64%	5.01%	5.32%	4.78%	4.56%
			h ≈ ln(n)	5.11%			k=10	4.13%	4.61%	5.10%	5.04%	4.75%
t_3	4.06%	4.94%	*h* = 2	5.05%	**3.52%**	**3.40%**	k=2	4.85%	4.79%	4.78%	4.85%	3.75%
			h = 5	4.89%			k=5	4.74%	4.94%	4.94%	4.81%	4.53%
			h ≈ ln(n)	5.14%			k=10	3.98%	5.10%	5.33%	5.29%	4.97%
Lognormal Standardized	4.17%	4.50%	*h* = 2	4.29%	**3.71%**	**8.56%**	k=2	4.11%	**8.41%**	4.51%	4.80%	**64.60%**
			h = 5	5.00%			k=5	4.16%	**9.11%**	4.60%	4.49%	**95.15%**
			h ≈ ln(n)	5.33%			k=10	3.70%	**8.41%**	4.78%	5.14%	**99.39%**
Heavy- Tailed Distribution	4.11%	5.08%	*h* = 2	4.27%	**3.62%**	4.01%	k=2	4.45%	4.60%	4.90%	4.93%	3.83%
			h = 5	4.78%			k=5	4.36%	5.68%	4.95%	4.80%	4.69%
			h ≈ ln(n)	5.11%			k=10	4.28%	6.12%	5.25%	4.92%	4.73%
X^2_4 Standardized	4.22%	4.78%	*h* = 2	4.76%	**3.30%**	4.09%	k=2	4.90%	5.87%	4.62%	4.46%	**7.94%**
			h = 5	5.00%			k=5	4.61%	5.06%	4.36%	4.80%	**18.11%**
			h ≈ ln(n)	4.91%			k=10	3.91%	4.54%	4.93%	5.16%	**31.22%**
Model Cases												
Mixture Model	4.23%	5.21%	*h* = 2	4.89%	**3.61%**	**3.04%**	k=2	5.30%	4.91%	4.87%	5.00%	3.72%
			h = 5	4.85%			k=5	4.82%	5.11%	5.22%	5.08%	4.66%
			h ≈ ln(n)	4.83%			k=10	4.39%	4.87%	5.33%	4.77%	4.97%
Bimodal	3.94%	5.10%	*h* = 2	5.06%	**3.65%**	**3.72%**	k=2	5.16%	5.25%	4.64%	4.56%	4.02%
			h = 5	5.25%			k=5	4.83%	4.83%	4.76%	4.92%	4.88%
			h ≈ ln(n)	5.11%			k=10	3.89%	4.89%	4.92%	5.29%	4.76%

SI = Strandberg and Iglewicz (2012) test with c = 0.5, 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)

Table 9: Simulation Rejection Rates Comparisons for Null Cases N=10,000

	SI	DW		LB	MV1	MV2		M1	M2	R1	R2	S1
Distribution Cases from the g- and -h distribution												
Z	5.29%	4.86%	*h* = 2	4.78%	**3.73%**	**3.56%**	*k*=2	5.21%	4.68%	5.04%	5.88%	5.11%
			h = 5	4.55%			*k*=5	4.74%	5.19%	5.00%	5.31%	4.96%
			h ≈ ln(N)	5.21%			*k*=10	5.01%	4.72%	5.24%	5.28%	4.99%
t_3	4.96%	4.93%	*h* = 2	5.27%	**3.44%**	**3.86%**	*k*=2	5.07%	5.29%	4.80%	5.14%	5.17%
			h = 5	5.15%			*k*=5	5.06%	4.95%	5.44%	4.96%	4.41%
			h ≈ ln(N)	5.24%			*k*=10	4.80%	4.97%	5.68%	5.44%	4.89%
Lognormal Standardized	4.91%	4.96%	*h* = 2	5.05%	**3.79%**	4.57%	*k*=2	5.00%	5.70%	4.52%	5.39%	**100.00%**
			h = 5	5.24%			*k*=5	4.64%	5.58%	5.67%	4.77%	**100.00%**
			h ≈ ln(n)	5.68%			*k*=10	4.90%	5.39%	5.58%	4.90%	**100.00%**
Heavy-tailed Distribution	4.94%	4.22%	*h* = 2	4.71%	**3.71%**	**3.38%**	*k*=2	4.16%	4.73%	5.33%	5.05%	5.08%
			h = 5	5.33%			*k*=5	4.60%	4.91%	5.09%	5.31%	5.05%
			h ≈ ln(N)	5.97%			*k*=10	4.22%	4.93%	5.50%	4.81%	4.70%
X^2_4 Standardized	4.91%	4.92%	*h* = 2	5.03%	**3.86%**	**3.94%**	*k*=2	4.96%	5.31%	5.27%	5.05%	**92.93%**
			h = 5	4.65%			*k*=5	5.15%	4.79%	5.74%	5.27%	**100.00%**
			h ≈ ln(n)	5.19%			*k*=10	4.84%	4.81%	5.37%	4.99%	**100.00%**
Model Cases												
Mixture Model	4.77%	5.24%	*h* = 2	4.96%	**3.94%**	**3.62%**	*k*=2	5.00%	4.97%	5.14%	5.49%	4.67%
			h = 5	4.88%			*k*=5	5.00%	4.80%	5.21%	5.23%	5.12%
			h ≈ ln(N)	5.16%			*k*=10	4.72%	5.20%	5.73%	5.39%	5.14%
Bimodal	4.99%	5.23%	*h* = 2	4.68%	**3.61%**	**3.86%**	*k*=2	5.02%	5.27%	4.95%	5.09%	5.29%
			h = 5	5.45%			*k*=5	4.73%	5.32%	5.42%	5.47%	4.82%
			h ≈ ln(N)	4.80%			*k*=10	5.28%	4.72%	5.61%	5.44%	4.61%

SI = Strandberg and Iglewicz (2012) test with c = 0.5, DW = Durbin and Watson (1950) test, LB = Ljung and Box (1978) test,
MV1 and MV2 are tests from Chow and Denning (1993), M1 and M2 are test from Lo and MacKinlay (1988), R1, R2, and S1 are tests from Wright (2000)

Table 10: Simulation Rejection Rates Comparisons for Alternative Cases N=100

	SI	DW		LB	MV1	MV2		M1	M2	R1	R2	S1
Changing Variance	**7.91%**	**6.61%**	$h = 2$	**6.60%**	**4.04%**	**2.51%**	k=2	**6.43%**	**5.48%**	**5.58%**	**6.34%**	**4.52%**
			$h = 5$	**8.15%**			k=5	**5.58%**	**4.27%**	**5.61%**	**5.82%**	**4.48%**
			$h \approx \ln(N)$	**7.77%**			k=10	**2.68%**	**2.52%**	**5.30%**	**5.96%**	**4.43%**
Changing Mean	**6.41%**	100.00%	$h = 2$	100.00%	100.00%	100.00%	k=2	100.00%	100.00%	100.00%	100.00%	100.00%
			$h = 5$	100.00%			k=5	100.00%	100.00%	100.00%	100.00%	100.00%
			$h \approx \ln(N)$	100.00%			k=10	100.00%	100.00%	100.00%	100.00%	100.00%
D1	3.07%	4.57%	$h = 2$	4.53%	2.83%	3.01%	k=2	4.87%	6.05%	5.86%	5.56%	4.49%
			$h = 5$	4.88%			k=5	4.17%	4.96%	5.89%	5.02%	6.72%
			$h \approx \ln(N)$	4.53%			k=10	2.39%	3.83%	5.83%	5.14%	6.90%
D2	2.96%	4.87%	$h = 2$	4.64%	2.65%	3.20%	k=2	5.15%	5.84%	5.62%	5.76%	5.37%
			$h = 5$	4.91%			k=5	3.99%	5.06%	5.80%	5.13%	7.53%
			$h \approx \ln(N)$	4.77%			k=10	2.50%	3.85%	5.58%	5.34%	8.55%
C1	6.89%	12.68%	$h = 2$	7.75%	4.69%	2.57%	k=2	8.87%	3.96%	5.63%	7.36%	4.06%
			$h = 5$	8.56%			k=5	5.60%	4.09%	6.36%	6.39%	4.94%
			$h \approx \ln(N)$	7.63%			k=10	3.05%	2.70%	5.50%	5.99%	4.67%
C2	6.43%	6.11%	$h = 2$	5.57%	3.16%	4.03%	k=2	6.22%	6.85%	5.42%	5.32%	4.45%
			$h = 5$	9.02%			k=5	4.52%	4.60%	5.34%	5.45%	5.26%
			$h \approx \ln(N)$	7.48%			k=10	2.56%	2.93%	5.03%	5.07%	4.71%
C3	6.17%	7.06%	$h = 2$	6.55%	3.62%	3.08%	k=2	5.74%	5.25%	5.99%	5.96%	4.41%
			$h = 5$	7.06%			k=5	4.54%	4.25%	5.52%	5.32%	5.14%
			$h \approx \ln(N)$	7.35%			k=10	2.73%	2.67%	5.01%	5.37%	4.52%
C4	3.61%	6.05%	$h = 2$	5.81%	3.19%	3.53%	k=2	5.83%	0.01%	5.29%	6.31%	4.34%
			$h = 5$	6.60%			k=5	5.43%	0.01%	5.54%	5.48%	4.84%
			$h \approx \ln(N)$	6.25%			k=10	2.76%	0.01%	5.44%	5.91%	4.20%
A1	9.07%	**6.87%**	$h = 2$	**6.59%**	**3.93%**	**2.71%**	k=2	**6.99%**	**5.87%**	**6.06%**	**6.27%**	**4.59%**
			$h = 5$	**7.20%**			k=5	**5.56%**	**4.42%**	**5.31%**	**5.57%**	**4.82%**
			$h \approx \ln(N)$	**6.59%**			k=10	**2.98%**	**2.27%**	**5.01%**	**5.71%**	**4.59%**
A2	47.63%	100.00%	$h = 2$	100.00%	100.00%	100.00%	k=2	100.00%	100.00%	100.00%	100.00%	100.00%
			$h = 5$	100.00%			k=5	100.00%	100.00%	100.00%	100.00%	100.00%
			$h \approx \ln(N)$	100.00%			k=10	99.98%	99.97%	100.00%	100.00%	99.98%

SI = Strandberg and Iglewicz (2012) test with c = 0.5, DW = Durbin and Watson (1950) test, LB = Ljung and Box (1978) test, MV1 and MV2 are tests from Chow and Denning (1993), M1 and M2 are test from Lo and MacKinlay (1988), R1, R2, and S1 are tests from Wright (2000)

Table 11: Simulation Rejection Rates Comparisons for Alternative Cases N=300

	SI	DW		LB	MV1	MV2		M1	M2	R1	R2	S1
Changing Variance	15.83%	**6.87%**	*h* = 2	**6.94%**	**4.87%**	**3.32%**	k=2	**6.53%**	**4.77%**	**5.55%**	**6.13%**	**3.56%**
			h = 5	**8.30%**			k=5	**6.24%**	**4.40%**	**5.60%**	**5.97%**	**4.63%**
			h ≈ ln(n)	**8.36%**			k=10	**5.59%**	**3.95%**	**5.55%**	**6.47%**	**4.72%**
Changing Mean	13.81%	100.00%	*h* = 2	100.00%	100.00%	100.00%	k=2	100.00%	100.00%	100.00%	100.00%	100.00%
			h = 5	100.00%			k=5	100.00%	100.00%	100.00%	100.00%	100.00%
			h ≈ ln(n)	100.00%			k=10	100.00%	100.00%	100.00%	100.00%	100.00%
D1	3.82%	4.44%	*h* = 2	4.66%	4.13%	3.91%	k=2	4.94%	5.18%	5.73%	5.69%	4.42%
			h = 5	5.27%			k=5	4.80%	5.17%	5.32%	4.82%	6.51%
			h ≈ ln(n)	4.86%			k=10	4.42%	4.63%	6.65%	5.18%	10.70%
D2	4.94%	4.43%	*h* = 2	5.06%	2.91%	3.87%	k=2	5.01%	5.39%	5.37%	4.96%	5.61%
			h = 5	4.90%			k=5	4.90%	4.89%	6.03%	5.31%	9.53%
			h ≈ ln(n)	4.83%			k=10	4.59%	4.93%	6.65%	5.22%	16.34%
C1	17.43%	22.81%	*h* = 2	15.09%	12.96%	5.66%	k=2	18.40%	19.20%	7.25%	12.29%	**3.98%**
			h = 5	14.94%			k=5	11.47%	10.51%	6.67%	8.59%	**4.33%**
			h ≈ ln(n)	14.40%			k=10	7.80%	7.69%	6.35%	7.25%	**4.98%**
C2	16.84%	6.28%	*h* = 2	6.42%	5.01%	4.59%	k=2	6.34%	6.48%	4.77%	5.31%	**3.78%**
			h = 5	15.47%			k=5	5.47%	5.70%	5.28%	5.44%	**4.40%**
			h ≈ ln(n)	15.32%			k=10	6.24%	6.42%	5.61%	6.49%	**5.41%**
C3	17.21%	7.50%	*h* = 2	7.84%	6.50%	4.81%	k=2	6.79%	7.08%	5.50%	5.86%	**3.63%**
			h = 5	9.92%			k=5	7.44%	7.24%	5.52%	7.03%	**4.63%**
			h ≈ ln(n)	9.75%			k=10	7.09%	6.99%	5.67%	7.22%	**5.05%**
C4	8.18%	7.01%	*h* = 2	7.25%	5.54%	4.67%	k=2	6.11%	6.14%	5.20%	5.20%	**3.71%**
			h = 5	7.72%			k=5	7.08%	7.48%	5.31%	6.46%	**4.63%**
			h ≈ ln(n)	8.24%			k=10	6.90%	6.67%	5.83%	6.59%	**4.84%**
M1	13.11%	**6.57%**	*h* = 2	**6.59%**	**4.48%**	**3.13%**	k=2	**6.95%**	**5.50%**	**5.11%**	**5.39%**	**3.75%**
			h = 5	**7.37%**			k=5	**5.90%**	**5.26%**	**5.27%**	**5.73%**	**4.46%**
			h ≈ ln(n)	**8.27%**			k=10	**5.69%**	**4.36%**	**5.41%**	**5.60%**	**5.04%**
M2	96.70%	100.00%	*h* = 2	100.00%	100.00%	100.00%	k=2	100.00%	100.00%	100.00%	100.00%	100.00%
			h = 5	100.00%			k=5	100.00%	100.00%	100.00%	100.00%	100.00%
			h ≈ ln(n)	100.00%			k=10	100.00%	100.00%	100.00%	100.00%	100.00%

SI = Strandberg and Iglewicz (2012) test with c = 0.5, 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)

Table 12: Simulation Rejection Rates Comparisons for Alternative Cases N=10,000

	SI	DW		LB	MV1	MV2		M1	M2	R1	R2	S1
Changing Variance	95.32%	**6.89%**	*h* = 2	**7.27%**	**5.52%**	**3.83%**	*k*=2	**7.43%**	**5.28%**	**5.35%**	**6.54%**	**5.42%**
			h = 5	**8.38%**			*k*=5	**7.14%**	**4.96%**	**5.89%**	**6.63%**	**5.21%**
			h ≈ ln(N)	**9.94%**			*k*=10	**6.98%**	**4.91%**	**5.86%**	**6.37%**	**5.00%**
Changing Mean	99.24%	100.00%	*h* = 2	100.00%	100.00%	100.00%	*k*=2	100.00%	100.00%	100.00%	100.00%	100.00%
			h = 5	100.00%			*k*=5	100.00%	100.00%	100.00%	100.00%	100.00%
			h ≈ ln(N)	100.00%			*k*=10	100.00%	100.00%	100.00%	100.00%	100.00%
D1	4.82%	4.98%	*h* = 2	4.93%	3.69%	3.86%	*k*=2	5.32%	5.10%	7.35%	5.72%	**29.50%**
			h = 5	4.96%			*k*=5	5.15%	4.77%	13.48%	5.65%	**70.97%**
			h ≈ ln(N)	4.93%			*k*=10	4.77%	5.04%	21.93%	5.31%	**93.83%**
D2	6.58%	4.65%	*h* = 2	4.88%	3.79%	3.80%	*k*=2	5.05%	5.18%	10.10%	5.57%	**58.54%**
			h = 5	5.00%			*k*=5	4.87%	5.19%	21.95%	6.00%	**96.50%**
			h ≈ ln(N)	4.77%			*k*=10	4.70%	4.99%	38.92%	6.16%	**99.91%**
C1	99.83%	100.00%	*h* = 2	100.00%	99.97%	99.95%	*k*=2	100.00%	100.00%	83.84%	99.91%	27.70%
			h = 5	99.93%			*k*=5	99.26%	98.63%	57.17%	95.49%	17.25%
			h ≈ ln(N)	99.80%			*k*=10	86.99%	83.72%	34.74%	73.57%	11.08%
C2	99.85%	6.73%	*h* = 2	7.55%	44.79%	39.32%	*k*=2	7.10%	5.07%	4.98%	6.55%	**5.07%**
			h = 5	99.92%			*k*=5	21.36%	19.75%	9.33%	17.43%	**6.28%**
			h ≈ ln(N)	99.81%			*k*=10	57.84%	51.64%	19.09%	42.71%	**8.32%**
C3	99.74%	37.63%	*h* = 2	50.89%	72.64%	66.32%	*k*=2	36.15%	30.65%	11.65%	26.04%	**5.92%**
			h = 5	70.04%			*k*=5	78.62%	74.63%	27.68%	63.60%	**9.74%**
			h ≈ ln(N)	61.78%			*k*=10	74.56%	70.36%	25.35%	58.20%	**9.27%**
C4	65.96%	26.80%	*h* = 2	34.48%	64.71%	60.61%	*k*=2	25.40%	21.18%	8.94%	18.36%	**5.98%**
			h = 5	47.92%			*k*=5	55.90%	52.45%	16.53%	42.32%	**7.01%**
			h ≈ ln(N)	52.78%			*k*=10	75.48%	73.22%	25.03%	60.24%	**8.63%**
A1	76.87%	**6.11%**	*h* = 2	**6.52%**	**5.27%**	**3.83%**	*k*=2	**6.12%**	**5.15%**	**5.66%**	**6.43%**	**5.21%**
			h = 5	**7.00%**			*k*=5	**6.49%**	**5.01%**	**5.25%**	**5.91%**	**4.93%**
			h ≈ ln(N)	**7.48%**			*k*=10	**6.22%**	**4.60%**	**5.66%**	**5.72%**	**4.91%**
A2	100.00%	100.00%	*h* = 2	100.00%	100.00%	100.00%	*k*=2	100.00%	100.00%	100.00%	100.00%	100.00%
			h = 5	100.00%			*k*=5	100.00%	100.00%	100.00%	100.00%	100.00%
			h ≈ ln(N)	100.00%			*k*=10	100.00%	100.00%	100.00%	100.00%	100.00%

SI = Strandberg and Iglewicz (2012) test with c = 0.5, DW = Durbin and Watson (1950) test, LB = Ljung and Box (1978) test, MV1 and MV2 are tests from Chow and Denning (1993), M1 and M2 are test from Lo and MacKinlay (1988), R1, R2, and S1 are tests from Wright (2000)

Table 13: Test Statistics for Stock Market Indices Application Results

DJIA	SI	DW		LB	*QP*	*MV1*	*MV2*		*M1*	*M2*	R1	R2	S1
Y_t - Y_{t-1}	-34.47***	2.11***	*h* = 2	***	6.39**	8.12***	2.26*	*k*=2	-7.44***	-2.23**	3.14***	0.42	5.04***
			h = 5	***				*k*=5	-8.12***	-2.26**	0.30	-1.99**	1.09
			h ≈ ln(N)	***				*k*=10	-6.67***	-1.82*	-0.26	-2.11**	0.70
100((Y_t/ Y_{t-1}) -1)	-16.79***	1.99	*h* = 2	5.85*	0.36	0.70	0.33	*k*=2	0.70	0.33	6.76***	5.03***	5.04***
			h = 5	10.64*				*k*=5	0.09	0.04	2.49**	2.16**	1.09
			h ≈ ln(N)	24.62***				*k*=10	0.31	0.14	1.42	1.45	0.70
100((Y_t /MA)-1)	-36.64***	0.07***	*h* = 2	***	***	***	87.92***	*k*=2	***	36.09***	***	***	***
			h = 5	***				*k*=5	***	64.09***	***	***	***
			h ≈ ln(N)	***				*k*=10	***	87.92***	***	***	***
S&P 500													
Y_t - Y_{t-1}	-28.01***	2.11***	*h* = 2	69.06***	8.03**	7.55***	2.46**	*k*=2	-6.23***	-2.13**	2.21**	-0.44	7.69***
			h = 5	73.17***				*k*=5	-7.45***	-2.44**	-0.84	-3.11***	3.88***
			h ≈ ln(N)	***				*k*=10	-7.55***	-2.46**	-1.76*	-3.99***	3.17***
100((Y_t/ Y_{t-1}) -1)	-11.93***	1.94**	*h* = 2	29.16***	4.27	3.00***	1.44	*k*=2	3.00***	1.44	8.97***	7.48***	7.69***
			h = 5	29.72***				*k*=5	0.00	0.00	4.44***	3.62***	3.88***
			h ≈ ln(N)	59.78***				*k*=10	-1.11	-0.52	2.34**	1.41	3.17***
100((Y_t /MA)-1)	-30.85***	0.08***	*h* = 2	***	***	***	***	*k*=2	***	43.73***	***	***	97.00***
			h = 5	***				*k*=5	***	77.73***	***	***	***
			h ≈ ln(N)	***				*k*=10	***	***	***	***	***
Nasdaq													
Y_t - Y_{t-1}	-18.71***	1.98	*h* = 2	1.27	3.52	2.62**	0.78	*k*=2	1.03	0.35	8.99***	6.38***	12.97***
			h = 5	12.17**				*k*=5	0.60	0.19	7.79***	4.95***	13.22***
			h ≈ ln(N)	***				*k*=10	-2.62***	-0.78	7.29***	4.10***	13.90***
100((Y_t/ Y_{t-1}) -1)	-14.59***	1.87***	*h* = 2	32.55***	10.49***	5.68***	3.10***	*k*=2	5.68***	3.10***	13.94***	11.32***	12.97***
			h = 5	42.89***				*k*=5	5.39***	2.80***	12.34***	9.92***	13.22***
			h ≈ ln(N)	57.91***				*k*=10	3.48***	1.81*	10.79***	8.23***	13.90***
100((Y_t /MA)-1)	-26.34***	0.06***	*h* = 2	***	***	***	95.76***	*k*=2	89.22***	38.62***	89.19***	89.38***	80.29***
			h = 5	***				*k*=5	***	68.98***	155.39***	156.20***	***
			h ≈ ln(N)	***				*k*=10	***	95.76***	***	***	***

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 (2012) test with c = 0.5, DW = Durbin and Watson (1950) test, LB = Ljung and Box (1978) test, QP = Chen and Deo (2006) test, MV1 and MV2 are tests from Chow and Denning (1993), M1 and M2 are test from Lo and MacKinlay (1988), R1, R2, and S1 are tests from Wright (2000)

CHAPTER 5
REGRESSION APPLICATIONS

This chapter is an investigative introduction to applying the proposed test to problems in simple linear least square regression. While the emphasis of this dissertation is on developing a competitive procedure for testing randomness in stock market index data, this newly developed test has the potential of far wider applications. In this chapter standardized residuals of small and large sample ordinary least square (OLS) regression models are tested for randomness of the error terms. When the assumption of *i.i.d.* errors is not met, the OLS regression model can result in loss of efficiency, biased estimators, and invalid inferences. Therefore tests such as the Durbin and Watson (1950) test – as detailed in Section 2.4 – and the Breusch and Pagan (1979) test are often used. In most statistics software packages, the Durbin and Watson test is the default test for testing correlation among error terms as the alternative to random errors; see for example SAS. The proposed test can also be used in this context.

We will here concentrate on the simple linear regression model when testing for independent errors of OLS models. Consider the simple linear regression model,

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{14}$$

where Y is the observed dependent variable, X is the independent variable, β_0 is the intercept term, β_1 is an unknown coefficient parameter, and ε is an unobservable disturbance (error) such that ε_t is assumed to be *i.i.d.* $t = 1, ..., N$, and X and ε are independent. Since the OLS regression model is sensitive to outliers and correlation among error terms, it is appropriate to test ε for deviations from randomness, namely *i.i.d.* Regression models where errors are independent and identically distributed from normal and non normal distributions, along with cases were dependent structures exist, will be considered in a simulated study. Each case is simulated using the proposed test, the Durbin and Watson (1950) test and the Breusch and Pagan (1979) test. Comparison between tests and considered cases are discussed in detail.

Breusch and Pagan (1979) test for the heteroscedasticity of the ε in an OLS regression model as given in (14). This test considers the null hypothesis of homosecedasticity such that $\sigma_t^2 = \sigma^2$ is constant for all ε_t. In this test, heteroscedasticity is determined if the estimated variance of ε is dependent on the values of the independent variable. This is tested by regressing the estimated squared residuals on the independent variables, namely $e_t^2 = \beta_1^* X_t + u_t$; note that the value of β_1^* will be different than the value of β_1 calculated in the regression model for Y. The null hypothesis is rejected if too much of the variance of ε is explained by X. Under the null hypothesis the test statistic follows a chi-squared distribution with p degrees of freedom, where p is the number of regressors in the model, here $p = 1$.

When generating simulated OLS regression models we start by assuming a bivariate normal model where X and Y are dependent observed values of random variables with Cor(X, Y) = ρ. Then $E(Y \mid X) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X)$ (Casella and Berger, 2002). For simplicity we will set $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_y$ and $Y = \rho X + \sqrt{1-\rho^2} \times \xi$ with ρ = 0.90, and ξ defined as *i.i.d.* $N(0, 1^2)$. This is our first simulated case, case 1. Next we then modify the distribution of only ξ and consider cases 2 thru 8. Cases 2 and 3 maintain the correlation structure ρ = 0.90 but are no longer bivariate normal, while cases 4 to 6 consider more complicated correlation structures by using ARIMA models; however in all cases 1 thru 8 X and ξ are independent. The following distributions and models for ξ are considered for cases 2 thru 8: (2) *i.i.d.* t_3; (3) *i.i.d.* χ^2_3 standardized; (4) ARIMA(0,1,1) which corresponds to serial correlation and is equivalent to simple exponential smoothing; (5) ARIMA(1,1,0) which is a differenced first-order autoregressive model; (6) ARIMA(1,1,1) which is a mixed model that includes features of both autoregressive and moving average models; (7) errors with changing variance such that 30% of the errors are distributed as N(0, 0.75^2), 40% as N(0, 1^2) and the remaining 30% as N(0, 1.25^2); (8) errors with changing mean such that 30% of the errors are distributed as N(-2, 1^2), 40% as N(0, 1^2) and the remaining 30% as N(2, 1^2).

In addition to these cases, four other cases are also considered. For cases 9, 10, 11, and 12, X and Y are dependent, however the correlation structure may not be ρ = 0.90. For cases 9 and 10, X ~ N(0, 1^2) but X and ξ are dependent, such that $Y = \rho X + \sqrt{1-\rho^2} \times \xi$ with ξ defined as $\xi_t = X_{t-2} + X_t^2$ for t =3, ..., N, and $\xi_t = X_t^2$ for t = 1, 2, in case 9 and $\xi_t = X_{t-2} + X_t^2 + X_t^3$ for t =3, ..., N, and $\xi_t = X_t^2 + X_t^3$ for t = 1, 2, in case 10. For cases 11 and 12, X and ξ are independent and an ARIMA(0,1,1) model is considered. For both cases $Y = \rho X + \sqrt{1-\rho^2} \times \xi$. Case 11 is similar to case 4 with ξ generated randomly from an ARIMA(0,1,1) model, but now X ~ t_3. In case 12, X is defined as an ARIMA(0,1,1) model, and $\xi \sim \chi_3^2$ that has been standardized. For all simulated cases, R software is used to randomly generate X and ξ, and then calculate Y; next, the full regression model defined in (14) is applied and the standardized residuals are tested. Note that when $\mu_X = \mu_Y = 0$, $\beta_0 = 0$ and the regression equation is through the origin. In our considered cases even when $\mu_X = \mu_Y = 0$ we still apply the full regression model in our simulations.

Tables 14, 15, and 16 shows the rejection rates for each considered case when testing the standardized residuals for N = 100, 300, 5,000. Each case was repeated 10,000 times and the rejection rate is the percentage of tests that rejected the null hypothesis. For the proposed test, K in (9) was set equal to 5, 8, and 10.

Table 14: Rejection Rates of Standardized Residuals N = 100

	DW	BP	SI K = 5	SI K = 8	SI K = 10
Case 1	4.65%	4.61%	3.52%	7.41%	10.70%
Case 2	4.51%	4.96%	3.70%	7.39%	9.77%
Case 3	4.67%	5.46%	3.46%	6.73%	9.44%
Case 4	100.00%	5.23%	69.03%	67.75%	71.84%
Case 5	100.00%	5.73%	71.51%	69.24%	74.57%
Case 6	100.00%	5.38%	74.18%	73.30%	77.90%
Case 7	6.24%	4.57%	8.63%	13.70%	19.82%
Case 8	100.00%	4.98%	6.05%	14.56%	22.46%
Case 9	3.91%	17.39%	15.38%	21.50%	23.89%
Case 10	4.33%	83.83%	6.86%	10.52%	13.91%
Case 11	100.00%	4.49%	68.75%	68.21%	71.86%
Case 12	4.81%	5.17%	5.11%	9.36%	12.33%

Table 15: Rejection Rates of Standardized Residuals N = 300

	DW	BP	SI K = 5	SI K = 8	SI K = 10
Case 1	4.64%	5.27%	4.02%	14.14%	9.26%
Case 2	4.94%	5.35%	3.94%	14.06%	10.20%
Case 3	4.59%	5.70%	4.24%	14.19%	9.58%
Case 4	100.00%	4.72%	99.75%	96.51%	89.70%
Case 5	100.00%	5.03%	99.80%	96.63%	90.09%
Case 6	100.00%	5.59%	99.79%	96.84%	91.47%
Case 7	6.36%	5.25%	13.43%	24.55%	20.60%
Case 8	100.00%	5.16%	16.68%	28.80%	29.18%
Case 9	7.13%	23.39%	8.92%	16.38%	11.32%
Case 10	4.38%	89.73%	5.58%	14.55%	10.05%
Case 11	100.00%	4.10%	99.80%	96.14%	89.76%
Case 12	4.88%	0.74%	10.07%	25.77%	28.04%

Table 16: Rejection Rates of Standardized Residuals N = 5,000

	DW	BP	SI K = 5	SI K = 8	SI K = 10
Case 1	5.08%	4.75%	5.02%	4.30%	3.56%
Case 2	4.69%	5.15%	5.90%	4.27%	3.73%
Case 3	5.27%	5.14%	4.91%	4.36%	3.72%
Case 4	100.00%	5.01%	100.00%	100.00%	100.00%
Case 5	100.00%	4.89%	100.00%	100.00%	100.00%
Case 6	100.00%	4.82%	100.00%	100.00%	100.00%
Case 7	6.52%	4.87%	74.74%	85.05%	87.28%
Case 8	100.00%	4.58%	87.10%	97.14%	98.90%
Case 9	43.42%	29.53%	23.49%	15.80%	18.27%
Case 10	12.17%	99.83%	9.41%	6.44%	6.32%
Case 11	100.00%	4.97%	100.00%	100.00%	100.00%
Case 12	4.27%	0.00%	78.99%	82.60%	71.50%

In general, results for the Durbin and Watson and Breusch and Pagan tests are consistent among sample sizes. For the Durbin and Watson test, cases 1, 2, 3, 7 and 12 have results similar to a null case, while cases 4, 5, 6, 8 and 11 show high power to reject the null hypothesis. These results are consistent with the simulation results seen in Chapter 4 – the Durbin and Watson test is able to detect the studied ARIMA models and changing means in standardized residuals. However this test is not able to detect changes in variance as rejection rates for case 7 are only slightly above the null rate of 5.00%. The Breusch and Pagan test considers every case, except cases 9, and 10 as a null case regardless of sample size with rejection rates reasonably close to 5.00%. This is not surprising since in these cases X and ε as defined in (14) are independent.

For the proposed test, differences exist among sample sizes. In general cases 1, 2, and 3 are considered null cases. When N is small; results are conservative for K = 5 and anticonservative for K = 8 and 10. In these cases, small sample sizes with larger K are not recommended. However when N = 5,000, test results for these cases are closer to the desired 5.00% rejection rate and become conservative as K increases. For cases 4, 5, and 6 the proposed test shows power to reject the null hypothesis for all considered values of K and as expected, power increases as N increases. For cases 7 and 8 weak power is seen when N is small although power does increase as K and N increases. When N = 5,000 high power is seen for cases 4 thru 8 with increasing power as K increases.

In cases 9 and 10 X and ε as defined in (14) are dependent. For case 9 the Breusch and Pagan and proposed test show modest power. Lower power is seen with the Durbin and Watson test when N is small, however when N = 5,000 the Durbin and Watson test has the highest power. For case 10, Breusch and Pagan is the only considered test to show high power for each sample size, while the proposed and the Durbin and Watson tests show modest power at best.

Although case 11 does not follow the bivariate normal model, when errors follow an ARIMA(0,1,1) model, as in this case and case 4, both the Durbin and Watson and proposed test show high power, similar to case 4, to detect the ARIMA model. As similar to the results in case 4, only the Breusch and Pagan test considers case 11 a null case. However in case 12 when the ARIMA(0,1,1) model is used to define X, none of the considered tests show consistent high power to reject the null hypothesis. In this case only the proposed test is able to show modest power that increases as N increase. The Durbin and Watson test considers this a null case while the Breusch and Pagan test grows extremely conservative as N increases. For both case 11 and 12, since X and ε are independent it is not surprising that the Breusch and Pagan test shows low rejection rates. For case 12 and large N, only the proposed test shows high power for rejecting the null hypothesis.

In summary, the Durbin and Watson test showed consistent results among the considered sample sizes. This test is able to detect ARIMA models and changes in means in the error term even when N is small. However this test is not recommended for detecting changes in variance or when X and ε are dependent. The Breusch and Pagan test considered every case, except cases 9 and 10, a null case regardless of sample size since there is no dependence among the estimated variance of the residuals and the independent variable. As X and ε are dependent for cases 9 and 10, higher power for the Breusch and Pagan procedure are observed. The proposed test is not recommended when N is small, especially when N is small and K is larger. When N is large, the proposed test is the only test that can detect changes in variance as in case 7 and ARIMA dependent variables with chi-square errors as in case 12. For other considered alternative cases – cases 4, 5, 6, 8 and 11 – the proposed test has comparable power to the Durbin and Watson test when N is large. As in cases 9 and 10 the proposed test shows modest power, it should not be recommended for these situations. Overall the proposed test is a competitor to the Durbin and Watson test for testing dependency of errors in simple linear regression, especially when N is large.

CHAPTER 6
CONCLUSION AND FUTURE RESEARCH

In this dissertation, a nonparametric test for determining whether a series constitutes a random sample is proposed. The method is unique because it is based on a simple test statistic that surprisingly works well for varied null and alternative models with very minimal assumptions. Although this is an asymptotic based test, it is also shown to be a valid procedure for modest sample sizes. Since our method is not based on autocorrelations or related measures it is shown to have high power for detecting correlated structures not only between consecutive observations but also over longer lags. Through extensive simulations, it is also shown that the proposed test meets the size requirements under a wide variety of null cases and has power to detect meaningful deviations from the null hypothesis in alternative cases.

In this dissertation we summarize and compare the proposed test with popular tests for detecting randomness in time series data, including comparisons in analyzing stock market index data, and for testing the error term in ordinary least square regression models. Our choice of tests, distributions, alternative distribution choices, and practical applications, makes this the most extensive comparison of such procedures. When deciding among tests, consideration must not only be given to meeting size requirements for the null hypothesis, but also the possible alternative hypotheses choices as power can differ considerably across alternatives. In our simulation study we considered a number of alternative cases to show the advantages and disadvantages of each of the studied tests.

For our stock market index data analysis we additionally review and compare multiple transformations and demonstrate that the choice of transformation can have a noticeable effect on test results. Of the studied tests, only the proposed test has power to detect changes in variance and the stochastic volatility model when N is large. In addition only the proposed test has high power for all four studied correlated cases when N is large, although respectable power is seen with the MV_1 and MV_2 tests.

In our application study, a modified formula for percent daily changes, MMPC, is introduced and the resulting transformed data are used to perform the tests. In this study, only the MMPC results in consistent rejection of the null hypothesis with high power for all tests and stock market indices, while the proposed test 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 considered tests, results can differ greatly depending on the transformation used. The only test that does not show evidence of this concern is the proposed test. Most individual variance ratio tests show reasonable results in these applications but can have conflicting conclusions depending on the chosen value of k. Other tests show inconsistent results. In summary, we believe that these comparisons and results will prove helpful, especially when dealing with financial data and tests of randomness.

In our regression application, when N is large, the proposed test is the only test that can detect changes in variance of the error term. In this study, the proposed test is a competitor to the Durbin and Watson test – the default test often used in statistical software packages – for testing dependency of errors in simple linear regression, especially when N is large. Future planned research will expand this study to test for randomness of the error terms in multiple regression models. Using a thoughtful selection of multiple regression models, the proposed test will be further compared to the traditional Durbin and Watson test.

Additional future research will include continued investigation of other application areas for the proposed test. Two areas that are of interest are missing data and monitoring structural changes. Consideration will be given to adjusting the proposed test to consider M intervals where data may be missing. The proposed test could then be applied to stock market data during weeks with missing observations due to market closures or holidays. This will give the test a broader range of applications. Also since the closing values for stock indices differ greatly, and increase over long periods of time while fluctuating considerably, they often consist of structural changes. Tests specifically designed for monitoring structural changes of more recent observations exists; see for example Andrews (1993) and Chu, Stinchcombe and White (1996). I believe the proposed testing approach has potential applications in this area.

REFERENCES

Ali, M. M., and Sharma, S. C. (1993), Robustness to Nonnormality of the Durbin-Watson Test for Autocorrelations, *Journal of Econometrics*, 57, 117-136.

Al-Khazali OM, Ding DK, and Pyun CS. (2007), A New Variance Ratio Test of Random Walk in Emerging Markets: A Revisit, *The Financial Review* 42, 303-317.

Andrews, D.W.K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61:821-856.

Bachelier, L. (1900), *Théorie de la Spéculation*, Paris: Gauthier-Villars.

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.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control* (rev. ed.), San Francisco: Holden-Day.

Box, G. E. P., and Pierce, D. A. (1970), Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models, *Journal of the American Statistical Association*, 65, 1509-1526.

Breusch, T. S., and Pagan, A. R., (1979), A Simple Test for Heteroscedasticity and Random Coefficient Variation, *Econometrica*, 47:1287-1294.

Brown, R. (1828), A Brief Account of Microscopical Observations Made on the Particles Contained in the Pollen of Plants, *London and Edinburgh Philosophical Magazine and Journal of Science*, 4, 161-173.

Campbell, J. Y., Lo, A. W, and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

Casella, G., and Berger, R. (2002), *Statistical Inference* (second ed.), Pacific Grove, CA: Duxbury.

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, K.C. (1993). A Simple Multiple Variance Ratio Test. *Journal of Econometrics* 58:385–401.

Chu, C. J., Stinchcombe, M., and White, H. (1996). Monitoring Structural Change, *Econometrica* 64:1045–1065.

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

Cootner, P. (ed.) (1964), *The Random Character of Stock Market Prices*, Cambridge, MA: MIT Press.

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

Einstein A. (1905), On the Motion of Small Particles Suspended in Liquids at Rest required by the Molecular Kinetic Theory of Heat, *Annalen der Physik*, 17, 549–560.

Fama, E. F. (1965), Random Walks in Stock Market Prices, *Financial Analysts Journal,* 2, 55-59.

French, K. R, Schwert, G. W., and Stambaugh, R. F. (1987), Expected Stock Returns and Volatility, *Journal of Financial Economics,* 19, 3-29.

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

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.

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

Iglewicz, B. (1983), Robust Scale Estimators and Confidence Interval for Location, in *Understanding Robust and Exploratory Data Analysis,* Hoaglin, D.C., Mosteller, F., Tukey, J.W. (eds.), Hoboken, NJ: John Wiley & Sons, Inc., pp 404-43.

Ito, T., Lyons, R. K., and Melvin, M. T. (1998), Is There Private Information in the FX Market? The Tokyo Experiment, *The Journal of Finance,* 53, 1111-1130.

Jensen, M. C. (1978), Some Anomalous Evidence Regarding Market Efficiency, *Journal of Finance Economics,* 6, 95-101.

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).

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

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.

Lo, A. W., and MacKinlay, A. C. (1989), The Size and Power of the Variance Ratio Test in Finite Samples: A Monte Carlo Investigation, *Journal of Econometrics*, 40, 203–238.

MacGillivray, H. L. (1992), Shape Properties of the g- and -h and Johnson Families, *Communications in Statistics-Theory and Methods*, 21, 1233-1250.

Malkiel, B. G. (2003), The Efficient Market Hypothesis and Its Critics, *The Journal of Economics Perspectives*, 17, 59-82.

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.

Pearson, K. (1905), The Problem of the Random Walk, *Nature*, 72, 294.

Ruppert, D. (2004), *Statistics and Finance an Introduction*, New York, NY: Springer.

Sidak, Z. (1967), Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association* 62, 626-633.

Strandberg, A.G., and Iglewicz, B. (2012), A Nonparametric Test for Deviation from Randomness with Applications to Stock Market Index Data. *Communications in Statistics- Simulation and Computation*, to Appear.

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

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Working, H. (1934), A Random-Difference Series for Use in the Analysis of Time Series, *Journal of the American Statistical Association*, 29, 11-24.

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

Yahoo! Finance, (2010a), *Dow Jones Industrial Average.* Retrieved June 2010, from http://finance.yahoo.com/q/hp?s=%5EDJI+Historical+Prices.

Yahoo! Finance, (2010b), *S&P 500 Index.* Retrieved June 2010, from http://finance.yahoo.com/q/hp?s=%5EGSPC+Historical+Prices.

Yahoo! Finance, (2010c), *NASDAQ Composite.* Retrieved June 2010, from http://finance.yahoo.com/q/hp?s=%5EIXIC+Historical+Prices.