The random walk of stock prices

implications of recent nonparametric tests

Dahl, Christian M.; Nielsen, Steen

Institut for Nationaløkonomi

Handelshøjskolen i København

Working paper 7-2001

THE RANDOM WALK OF STOCK PRICES: IMPLICATIONS OF RECENT NONPARAMETRIC TESTS

Christian M. Dahl **Steen Nielsen**

Department of Economics - Copenhagen Business School
Solbjerg Plads 3, DK-2000 Frederiksberg

The Random Walk of Stock Prices: Implications of Recent Nonparametric Tests*

Christian M. Dahl†

Steen Nielsen‡

August 4, 2001

JEL: G12. Key words: Random walk, nonparametric tests, stock returns.

ABSTRACT

This paper applies six recently developed nonparametric tests of serial independence to monthly US stock returns. Findings of previous studies based on the BDS test are supported since most of the new tests also reject the random walk hypothesis. Furthermore, power properties of the new tests are compared with those of the BDS test. The latter has much power against ARCH and GARCH alternatives whereas some of the more recent tests are superior against other alternatives. Finally, the power study of this paper shows, contrary to common belief, that ARCH and GARCH effects do not seem to explain rejection of the random walk.

---

*An early version of this paper was presented on a stock market workshop at Copenhagen Business School. Comments from Elroy Dimson and Richard Priestley are gratefully acknowledged. Gauss codes prepared for this paper are available upon request to the authors.

†Economics Department, Purdue University.

‡Address: Department of Economics, Copenhagen Business School, Solbjerg Plads 3, C5, 2000 Frederiksberg, Denmark. Phone: +45 3815 2575. Fax: +45 3815 2576. E-mail: Sn.eco@cbs.dk.

Abstract

This paper applies six recently developed nonparametric tests of serial independence to monthly US stock returns. Findings of previous studies based on the BDS test are supported since most of the new tests also reject the random walk hypothesis. Furthermore, power properties of the new tests are compared with those of the BDS test. The latter has much power against ARCH and GARCH alternatives whereas some of the more recent tests are superior against other alternatives. Finally, the power study of this paper shows, contrary to common belief, that ARCH and GARCH effects do not seem to explain rejection of the random walk.

I. Introduction

This paper reexamines the random walk theory of stock prices by means of recently developed nonparametric tests. In the early literature, the random walk received some support. This was due to the first generation of nonparametric tests, for example the Runs test which was applied to stock returns by Fama (1965). However, later research has shown that first generation tests lack power against relevant alternatives to the random walk which is also demonstrated below in the power simulations of section IV.

Subsequently, evidence has turned against the random walk theory. This is due to the nonparametric BDS test which was proposed in the late 1980s. A number of studies have documented that the BDS test rejects independence of stock returns, see f. ex. Hsieh (1991), Pagan (1996), and Scheinkman and LeBaron (1989). Furthermore, the BDS test is more powerful than the first generation tests. In particular, it has power against ARCH and GARCH alternatives which are popular descriptions of stock returns, see the survey in Pagan (1996).

Recent research has produced a number of alternative tests of serial independence. The main contribution of this paper is to apply a set of these second generation tests — including the BDS test — to a common, standard time series of stock returns. This serves to explore whether rejection of random walk for stock prices is sensitive to the choice of test procedure. Furthermore, power properties of the tests are compared. This analysis provides guidance on selecting the most powerful tests and as a byproduct we gain some insight into possible reasons for rejection of the random walk.

The present study focuses entirely on testing the strong version of the random walk theory, i.e., the hypothesis that stock returns are serially independent. Other papers deal with weaker versions such as the specification of returns as uncorrelated, see Pagan (1996) for a survey of this literature. Testing the random walk model is important for several reasons. First of all, theoretical work on portfolio selection and asset pricing is often based on the assumption that stock prices behave like a random walk. Secondly, rejection of the random walk has implications for investors who are trying to predict future stock returns. Thus, if returns are dependent, then in principle there is scope for market timing based on the historical behavior of stock prices.

The BDS test is documented in Brock et al. (1996). In the present paper, we study the BDS test and six other second generation tests: Ahmad and Li (1997), Hjellvik and Tjøstheim (1996) (two tests), Hong and White (2000), Robertson (1991), and Skaug and Tjøstheim (1993). For completeness, results on the Runs test are also included.

We suggest to group the second generation tests in two classes: One class contains tests based on distribution functions, and the other consists of tests based on density functions. The latter may be further divided into tests of difference between bivariate and product of marginal densities, entropy based tests, and tests involving conditional moments. Due to the very different nature of the tests, it is not obvious that they spawn similar inference.

The following section introduces the nonparametric tests applied in the paper. Section III presents the data and test results. Section IV contains power simulations. Finally, conclusions are offered in section V.

II. Nonparametric tests of serial independence

Our starting point is a time series of stock returns, R_t, $t = 1, 2, \ldots, T$. From this we construct log returns, $X_t = \ln(1 + R_t)$. The random walk hypothesis implies that X_t is independent of X_{t-1}. In principle, a random walk also implies that log return is independent of longer lags. However, for the sake of clarity we restrict attention to the first lag only.

Nonparametric tests of independence may be classified as based on either distribution or density functions. This paper includes one from the former class of tests and six from the latter. The following sections present the tests considered in this paper. Each test considers a particular measure of departure from H_0. This measure is in most cases normalized to obtain an asymptotically normal null distribution of the test statistic. However, large sample performance has been shown to be poor for many of the tests, see f.ex. Skaug and Tjøstheim (1996). Therefore, we have chosen to simulate critical values by Monte Carlo. This ensures correct size across tests. The poor large sample performance is also the reason for not going into details with normalizations in the present section.

A. A test based on distribution functions

Skaug and Tjøstheim (1993) consider an intuitive and easily computable test which is based on distribution functions. Consider the marginal distribution function, $G(x) = Pr(X_t \leq x)$, and the bivariate distribution function, $F(x,y) = Pr(X_{t-1} \leq x, X_t \leq y)$. Independence of X_t and X_{t-1} is equivalent to $F(x,y) = G(x)G(y)$ for almost all x,y. Thus, it is natural to study:

$$I^{ST} = \int \{F(x,y) - G(x)G(y)\}^2 dF(x,y) \tag{1}$$

since $I^{ST} = 0$ if and only if X_t and X_{t-1} are independent. I^{ST} is estimated by inserting the empirical distribution functions, $G_T(x) = \frac{1}{T}\sum_{t=1}^{T} 1(X_t \leq x)$ and $F_T(x,y) = \frac{1}{T-1}\sum_{t=2}^{T} 1(X_{t-1} \leq x)1(X_t \leq y)$ where $1(\cdot)$ is the indicator function. $G_T(x)$ is the frequency of log returns less than x and $F_T(x,y)$ is the frequency of pairs of lagged returns and returns less than x and y respectively. Hence, the resulting test statistic is:

$$I_T^{ST} = \frac{1}{T-1}\sum_{t=2}^{T}\{F_T(X_{t-1},X_t) - G_T(X_{t-1})G_T(X_t)\}^2 \tag{2}$$

The test is a measure of closeness of the empirical bivariate and product of marginal distributions, and the random walk hypothesis is rejected if $I^{ST} > 0$. In addition to its simplicity, this test is attractive because it does not require the researcher to choose any parameters. As we shall see in the sequel, the ST test is the only test considered here which has this feature.

B. Tests based on density functions

In this section, we consider a number of tests which are based on the assumption that marginal, $g(\cdot)$, and bivariate, $f(\cdot,\cdot)$, density functions exist. This assumption allows us to test hypotheses that are based on densities. For example, an obvious test of independence is to examine whether $f(x,y) = g(x)g(y)$ for almost all x,y. This is the approach of the tests mentioned in sections B.1 and B.2. An alternative approach is to verify if the conditional mean or variance of X_t given a certain value of X_{t-1} is equal to the unconditional mean or variance. This idea is pursued by the tests presented in section B.3.

Before proceeding to the presentation of tests, we briefly consider some issues in relation to density estimation. There is a huge literature on optimal estimation of densities, see f.ex. Pagan and Ullah (1999). This literature provides rules for choosing kernels and bandwidths. Since the focus of the present paper is a comparison of independence tests rather than density estimation, we pursue a unified approach to density estimation which may be applied to all the tests. Here is a brief discussion of the kernels and bandwidths used in this paper.

The role of the kernel is to provide a weighting of the observations when estimating a density. Suppose we wish to estimate the (marginal) density of returns at a point x. Then a straightforward procedure is to compute the frequency of observations, X_t, within a distance of $\frac{a}{2}$ from x. Alternatively, this may be expressed as the frequency of observations for which $-\frac{1}{2} < \frac{x-X_t}{a} < \frac{1}{2}$. This procedure is called the uniform kernel because each observation (within $\frac{a}{2}$ from x) receives the same weight. a is called the bandwidth. Better estimates of the density are obtained by attaching greater weight to observations that are closer to x and less weight to observations far from x. A number of different kernels are available for this purpose. One example is to assign weights according to the standard normal density. Thus, for each observation, the standard normal density is evaluated at $\frac{x-X_t}{a}$ and the density of returns at x is estimated as the sum across all observations relative to na. This is referred to as the standard normal kernel. When x itself belongs to the sample, it is customary to exclude that element (where $X_t = x$) from the sum in order to preclude effects of outliers on the estimate. This is called the leave-one-out method.

We apply the standard normal kernel to all densities except for two cases. The BDS test is originally developed with a uniform kernel and, hence, we follow that tradition. We also deviate from the standard normal kernel in our application of the Hong-White test because in that case the authors explicitly suggest using an alternative kernel, cf. section B.2. Furthermore, it should be noted that we use the leave-one-out approach for all tests.

Based on the literature on bandwidth selection, we choose to set bandwidth equal to $\sigma_X T^{-1/5}$ (except for the BDS test which treats bandwidth in a special way, cf. below). This standard choice meets the requirement that bandwidths vanish as sample size tends to infinity. Hence, observations not close to x are ignored as the sample expands. Furthermore, this bandwidth has been shown under certain conditions to balance the tradeoff between bias and variance of density estimates, cf. Pagan and Ullah (1999), p. 26.

The following subsections discuss density based tests. The tests are grouped into three categories: Difference between bivariate and product of marginals, entropy, and conditional moment tests.

B.1. Tests of difference between bivariate and product of marginals

This section describes the widely used BDS test discussed by Brock et al. (1996) and the test analyzed in Ahmad and Li (1997). Both tests deal with the null hypothesis $f(x,y) = g(x)g(y)$. The most obvious approach is to consider the mean departure from H_0:

$$I^{BDS} = \int \int \{f(x,y) - g(x)g(y)\} dF(x,y) \tag{3}$$

and reject the random walk hypothesis if $I^{BDS} \neq 0$. Under the null, this may be written as:

$$I^{BDS} = \int\int f(x,y)dF(x,y) - (\int g(y)dG(y))^2 \tag{4}$$

Pagan (1996) makes the useful observation that the BDS test described by Brock et al. (1996) is based on an estimate of (4):

$$I_T^{BDS} = \frac{1}{T-1}\sum_{t=2}^{T} f_T(X_{t-1}, X_t) - (\frac{1}{T}\sum_{t=1}^{T} g_T(X_t))^2 \tag{5}$$

where $f_T(\cdot,\cdot)$ and $g_T(\cdot)$ denote estimates of bivariate and marginal densities. Notice, that although in principle the standard normal kernel may be used, we follow Brock et al. (1996) by applying the uniform kernel.

The BDS test statistic involves a scaling of (5) (see Pagan, 1996) and normalization by the estimated standard deviation of I_T^{BDS}. In order to apply the test, a value must be chosen for the parameter ε which has the interpretation of a bandwidth. We let $\varepsilon = \sigma_X$, where σ_X is the sample standard deviation of log returns. This conforms to the choice of ε in Hsieh (1991). In contrast to the following tests, the asymptotics of BDS are developed under the assumption that the bandwidth does not vanish as sample size increases.

A related test considered by Ahmad and Li (1997) examines the following squared measure of deviation from the null hypothesis:

$$\begin{aligned} I^{AL} &= \int\int \{f(x,y) - g(x)g(y)\}^2 dxdy \\ &= \int\int \{f(x,y)dF(x,y) + \int g(x)dG(x)\int g(y)dG(y) - 2\int\int g(x)g(y)dF(x,y) \end{aligned} \tag{6}$$

I^{AL} is estimated by:

$$I_T^{AL} = \frac{1}{T-1}\sum_{t=2}^{T} f_T(X_{t-1},X_t) + \frac{1}{(T-1)^2}\sum_{t=2}^{T} g_T(X_{t-1})\sum_{t=2}^{T} g_T(X_t) - 2\frac{1}{T-1}\sum_{t=2}^{T} g_T(X_{t-1})g_T(X_t) \tag{7}$$

It is illuminating to compare the AL and BDS tests. Rewrite (6) as:

$$\begin{aligned} I^{AL} &= I^{BDS} - \int g(x)g(y)dF(x,y) + \int g(x)dG(x)\int g(y)dG(y) \\ &= I^{BDS} - \int\int g(x)g(y)\{f(x,y) - g(x)g(y)\}dxdy \end{aligned} \tag{8}$$

The extra term is zero under the null. In finite samples, however, I^{AL} generally differs from I^{BDS}. Thus, the two tests may produce conflicting inference beyond differences due to kernel and bandwidth discrepancies.

B.2. Entropy based tests

Entropy based tests consider the measure:

$$I = \int ln\Big(f(x,y)/g(x)g(y)\Big)dF(x,y) \tag{9}$$

This measure is nonnegative and equal to zero if and only if H_0 is true. It may be estimated the usual way by inserting density estimates. However, this approach does not have a limiting normal null distribution. Therefore, certain changes to (9) have been proposed.

Robinson (1991) introduces a split of the sample:

$$I_T^{Rob} = \frac{1}{T_\gamma - 1}\sum_{t=2}^{T} c_t(\gamma) ln\Big(f_T(X_{t-1},X_t)/g(X_t)^2\Big) \tag{10}$$

where $\left(c_t(\gamma), T_\gamma\right)$ equals $\left(1+\gamma, T+\gamma\right)$ when t is odd, and $\left(1-\gamma, T\right)$ when t is even, $\gamma \geq 0$. If the estimated bivariate density is nonpositive for some t, logarithms cannot be taken. Thus, such cases are excluded from the sum. We choose $\gamma = 1$ and the parameter δ which is used to normalize I_T^{Rob} is set equal to 0.

The strategy of Hong and White (2000) is to avoid the nuisance parameter, γ. They show that a limiting normal null distribution may be obtained from direct estimation of (9) by subtraction of a nonzero mean and proper scaling. They use a higher-order (socalled quartic) kernel rather than the standard normal kernel described above. We follow their suggestion and apply the same kernel to this test. Furthermore, this test requires a rescaling of the data onto $[0,1]$ and the use of a jack-knife (see Hong and White, 2000) kernel near the boundary of this interval.

B.3. Tests involving conditional moments

Hjellvik and Tjøstheim (1996) propose two tests that are based on mean and variance of returns conditional on lagged return. If the hypothesis of independence is correct, then the conditional moment equals the unconditional moment. Hence, this comparison forms the basis of the tests.

The conditional mean of X_t given $X_{t-1} = x$ is defined as:

$$M(x) = \int y f(x,y)/g(x) dy \tag{11}$$

In the application of this test, X_t is demeaned and, hence, $M(x) = 0$ if X_t is independent. Thus, we consider

$$I^{HTM} = \int M_T^2(x) g(x) w(x) dx \tag{12}$$

where $M_T(x)$ is a kernel based estimate of $M(x)$ and $w(\cdot)$ is a weight function to screen off extreme values (we follow the suggestion of Hjellvik and Tjøstheim and let $w(x)$ exclude all observations greater than 3). The null hypothesis is rejected if $I^{HTM} > 0$.

I^{HTM} is estimated by

$$I_T^{HTM} = \frac{1}{T-1}\sum_{t=2}^{T} M_T^2 X_t w(X_t) \tag{13}$$

Similarly, we may consider the conditional variance:

$$V(x) = \int y^2 f(x,y)/g(x)dy - M^2(x) \tag{14}$$

X_t is rescaled to have sample standard deviation equal to 1. Thus, the test statistic is based on the squared deviation of the conditional variance from 1:

$$I_T^{HTV} = \int \left\{V(x) - 1\right\}^2 g(x)w(x)dx \tag{15}$$

and the null is rejected if I_T^{HTV} is greater than 0.

III. Data and test results

We consider monthly total returns (i.e., including dividends) on the US large cap stock index from Ibbotson Associates (2000). This series is based on the S&P Composite which currently includes 500 large stocks; prior to 1957 it consisted of 90 large stocks. The sample period is January 1926 to December 1999. Thus, the total number of observations is $T = 888$.

Table I applies the seven tests described above plus the Runs test to the Ibbotson data set. Critical values for the tests at 5% significance level are also included in table I. In order to ensure the same significance level across tests, we have chosen to simulate critical values rather than rely on asymptotic distributions. The first step in this simulation is to estimate the sample mean, $\mu_{X,T}$, and standard deviation, $\sigma_{X,T}$, of log returns. Then 1000 samples with T IID observations are generated by the following model:

$$X_t = \mu_{X,T} + \sigma_{X,T}\varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0,1)$$

Finally, each test is applied to all 1000 samples and the critical value is taken to be the 95% quantile of the set of absolute test statistics.

Table I includes results on both log of raw returns, X_t, and log of excess returns, $\ln(1 + R_t - RF_t)$, where RF_t is total return on US T-Bills (from Ibbotson, 2000). Consider first the Runs test which was applied on stock returns by Fama (1965). This test counts the number of sequences of positive and negative returns (=Runs) in the sample. Properly normalized, this number is standard normal distributed, see Wallis and Roberts (1956). Table I shows that the Runs test does not reject the random walk hypothesis for the S&P Composite index. This is compatible with Fama's conclusion which is based on an analysis of individual stocks at time intervals up to 16 days.

On the other hand, later studies show that the BDS test rejects the random walk of stock prices, e.g. Hsieh (1991), Pagan (1996), and Scheinkman and LeBaron (1989). This result is also confirmed by table I since the BDS statistics for raw and excess returns are clearly greater than the BDS critical value. Most researchers prefer the BDS test to the Runs test because of its better power characteristics. We return to this issue in the following section.

Finally, table I contains new evidence based on the recently developed tests. As seen in the table, independence is rejected at the 5% level by the AL, HTV, HW, Rob, and ST tests whereas only the HTM test lends some support to the hypothesis by not rejecting in case of raw returns. Thus, the recent nonparametric tests of independence tend to uphold the BDS rejection. Hence, table I adds to the evidence against the random walk hypothesis.

In the following section, we explore power properties and draw inference from the observed rejection pattern.

IV. Power simulations

This section provides Monte Carlo simulations to examine power of the nonparametric tests of independence. We estimate a number of parametric models that are supposed to capture relevant aspects of stock return processes. Then the power of each test against these parametric alternatives is evaluated. Also, we compare power results with the rejection pattern described in the previous section to infer characteristics of the true data generating process.

Ten different models are considered: Autoregression with 1 lag (AR1), moving average with 1 lag (MA1), nonlinear moving average (NMA), threshold autoregression with 1 lag (TAR), autoregressive conditional heteroskedasticity with 1 lag (ARCH), generalized autoregressive conditional heteroskedasticity with 1 lag (GARCH). Furthermore, the analysis includes four versions of two-states Markov-switching: Different mean across states (MS1), different variance across states (MS2), different mean and variance across states (MS3), different mean, variance and autoregressive term across states (MS4).

All models are estimated by numerical maximum likelihood. Results are shown in table II. The estimated models are used to generate 1000 time series and compute rejection frequencies for each of the tests. The results of this work is presented in table III. We emphasize that power results to some extent depend on the parameters chosen. However, we believe that estimation of parameters provides a sensible, practical solution to that problem.

We notice that the BDS test is particularly powerful against ARCH and GARCH models. Since these models are often thought to be relevant for financial time series, the BDS test is a good choice in this context. BDS also has some power against regime switching models. Since the BDS test was the first test to question the original results of the Runs test, it is relevant to compare those two tests. Table III shows that the BDS test is clearly superior to the Runs test in the lower half of the table. This explains why failure to reject by the Runs test is not a very strong result.

HTV seems to be the most powerful alternative to BDS. It has more power against MS1 and also appears to be marginally preferable against the other regime-switching alternatives considered here. On the other hand, it is weaker when the alternative is ARCH or GARCH.

HTM does not fail completely against any of the alternatives. Hence, judging on the basis of table III, the failure of HTM to reject the random walk for raw returns must be due to chance.

The results of the Robinson test are disappointing. It has very low power against many alternatives including ARCH and GARCH. Thus, the fact that the Robinson test does reject H_0 is not likely to be explained by ARCH or GARCH effects in the data. In our view, this is one of the most interesting implications of the present study because of the popularity enjoyed by ARCH and GARCH models in recent years. In fact, rejection by BDS combined with its power against ARCH/GARCH led many researchers to believe that conditional heteroskedasticity effects are the cause of rejection, see e.g. Hsieh (1991) and Scheinkman and LeBaron (1989). A very important topic for future research is to identify aspects of the data that leads the Robinson test to reject. As a final point concerning the Robinson test, it should also be noted that the alternative entropy-based test by Hong and White clearly seems to offer an improvement of the Robinson test.

Finally, the tests generally have low succes against autoregressive and moving average alternatives, ie., AR1, MA1, NMA, and TAR. This is due to the small estimated coefficients of these models which on the other hand indicates that AR and MA effects are small in the sample. In simulations not reported here, we find that the models are powerful against AR1, MA1, and TAR when coefficients are greater and that BDS and HTV are capable of detecting NMA. It appears that ST has more power against autoregressive and moving average models than the other tests. This may suggest using ST as a supplement to other tests when analyzing financial data. However, ST should not be applied in isolation because of its inability to capture GARCH and regime-switching.

V. Conclusion

This paper extends the previous literature on the random walk of stock prices by comparing the originally used BDS test with six recently developed nonparametric tests. The rejection derived on the basis of the BDS test is confirmed by the new tests.

The BDS has relatively much power against ARCH and GARCH alternatives. Besides, it is shown that the HTV test is powerful against regime-switching alternatives and that the ST test performs well with autoregressive and moving average models. Hence, combined use of these three tests is suggested for analysis of stock returns.

Another important finding is that the random walk is rejected by the Robinson test which has almost no power against ARCH and GARCH. This implies that other aspects of the data are likely to explain why the random walk does not hold. Since much of the literature has focused on ARCH and GARCH, more research is needed to account for this result.

Finally, it may be appropriate to mention a few possible extensions. First, other parametric models than those considered in our power simulations may be relevant. Second, the analysis may be extended to independence at longer lag lengths. This would allow for an assessment of whether returns are also dependent beyond the one month horizon studied in this paper.

References

[1] Ahmad, I.A. and Q. Li (1997), Testing independence by nonparametric kernel method, *Statistics and Probability Letters* 34, 201-210.

[2] Brock, W.A., W.D. Dechert, J.A. Scheinkman and B. LeBaron (1996), A test for independence based on the correlation dimension, *Econometric Reviews* 15(3), 197-235.

[3] Fama, E.F. (1965), The behavior of stock-market prices, *Journal of Business* 38 (1), 34-105.

[4] Hjellvik, V. and D. Tjøstheim (1996), Nonparametric statistics for testing linearity and independence, *Nonparametric Statistics* 6, 223-251.

[5] Hong, Y. and H. White (2000), Asymptotic distribution theory for nonparametric entropy measures of serial dependence, working paper.

[6] Hsieh, D.A. (1991), Chaos and nonlinear dynamics: Application to financial markets, *Journal of Finance* 46 (5), 1839-1877.

[7] Ibbotson Associates (2000), *Stocks, Bonds, Bills, and Inflation 2000 Yearbook*.

[8] Pagan, A. (1996), The econometrics of financial markets, *Journal of Empirical Finance* 3, 15-102.

[9] Pagan, A. and A. Ullah (1999), *Nonparametric econometrics*, Cambridge:Cambridge University Press.

[10] Robertson, P.M. (1991), Consistent nonparametric entropy-based testing, *Review of Economic Studies* 58, 437-453.

[11] Scheinkman, J.A. and B. LeBaron (1989), Nonlinear dynamics and stock returns, *Journal of Business* 62 (3), 311-337.

[12] Skaug, H.J. and D. Tjøstheim (1993), A nonparametric test of serial independence based on the empirical distribution function, *Biometrika* 80 (3), 591-602.

[13] Skaug, H.J. and D. Tjøstheim (1996), Testing for serial independence using measures of distance beween densities, in P.M. Robinson and M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series - Vol II: Time Series Analysis*, New York: Springer, 1996.

[14] Wallis, W. and H. Roberts (1956), *Statistics: A new approach*, New York: Free Press.

Table I: Test statistics with monthly total returns

This table presents statistics of the Ahmad-Li, BDS, Hjellvik-Tjøstheim M, Hjellvik-Tjøstheim V, Hong-White, Robinson, Runs, and Skaug-Tjøstheim tests applied to monthly raw and excess returns, S&P Composite, 01.1926-12.1999. The final column reports simulated critical values at 5% significance level in 1000 samples of 888 observations. Simulation is based on IID process with empirical mean and variance (raw return). Rejection at the 5% level is marked by an asterisk. Probability values in parentheses.

	Raw return	Excess return	Critical value
AL	1.6*	1.7*	1.3
	(0.026)	(0.019)	
BDS	6.6*	6.7*	2.0
	(0.000)	(0.000)	
HTM	0.10	0.14*	0.13
	(0.615)	(0.006)	
HTV	0.60*	0.47*	0.22
	(0.000)	(0.000)	
HW	7.5*	9.3*	6.3
	(0.013)	(0.000)	
Rob	4.6*	3.8*	2.5
	(0.000)	(0.003)	
Runs	0.84	1.3	1.9
	(0.384)	(0.182)	
ST	0.060*	0.065*	0.057
	(0.041)	(0.030)	

Table II: Estimated parametric models

This table presents estimates of the following models applied to monthly raw returns, S&P Composite, 01.1926-12.1999: AR1, MA1, NMA, TAR, ARCH, GARCH plus four different two-states Markov-switching models: Different mean across states (MS1), different variance across states (MS2), different mean and variance across states (MS3), different mean, variance and autoregressive term across states (MS4). $\varepsilon_t \sim \mathrm{NID}(0,1)$.

AR1	$X_t = 0.0083 + 0.075X_t + 0.056\varepsilon_t$
MA1	$X_t = 0.0090 + 0.076\varepsilon_{t-1} + 0.056\varepsilon_t$
NMA	$X_t = 0.0090 - 0.10\varepsilon_{t-1}\varepsilon_{t-2} + 0.056\varepsilon_t$
TAR	$X_t = \begin{cases} 0.0076 + 0.071X_{t-1} + 0.055\varepsilon_t & \text{if } \mid X_t \mid \leq 0.25 \\ 0.12 + 0.077X_{t-1} + 0.055\varepsilon_t & \text{else} \end{cases}$
ARCH	$X_t = 0.0072 + 0.26X_{t-1} + u_t$ $u_t = h_t^{1/2}\varepsilon_t$ $h_t = 0.0021 + 0.41u_{t-1}^2$
GARCH	$X_t = 0.010 + 0.03X_{t-1} + u_t$ $u_t = h_t^{1/2}\varepsilon_t$ $h_t = 0.000064 + 0.87h_{t-1} + 0.12u_{t-1}^2$
MS1	$X_t \mid s_t = 1 \quad = 0.013 + 0.048\varepsilon_{1t}$ $X_t \mid s_t = 2 \quad = -0.019 + 0.048\varepsilon_{2t}$ $p_{11} = 0.98, p_{22} = 0.23, \pi_1 = 0.98$
MS2	$X_t \mid s_t = 1 \quad = 0.012 + 0.12\varepsilon_{1t}$ $X_t \mid s_t = 2 \quad = 0.012 + 0.038\varepsilon_{2t}$ $p_{11} = 0.94, p_{22} = 0.99, \pi_1 = 0.12$
MS3	$X_t \mid s_t = 1 \quad = -0.015 + 0.12\varepsilon_{1t}$ $X_t \mid s_t = 2 \quad = 0.012 + 0.038\varepsilon_{2t}$ $p_{11} = 0.93, p_{22} = 0.99, \pi_1 = 0.12$
MS4	$X_t \mid s_t = 1 \quad = -0.014 + 0.094X_{t-1} + 0.12\varepsilon_{1t}$ $X_t \mid s_t = 2 \quad = 0.013 - 0.0047X_{t-1} + 0.038\varepsilon_{2t}$ $p_{11} = 0.93, p_{22} = 0.99, \pi_1 = 0.12$

Table III: Size-corrected power of test statistics, 5% level

This table presents percentage rejection frequencies of the Ahmad-Li, BDS, Hjellvik-Tjøstheim M, Hjellvik-Tjøstheim V, Hong-White, Robinson, Runs, and Skaug-Tjøstheim tests when the true process is AR1, MA1, NMA, TAR, ARCH, GARCH, MS1, MS2, MS3, MS4. Results are based on 888 observations and 1000 replications.

	AL	BDS	HTM	HTV	HW	Rob	Runs	ST
AR1	8	8	15	5	6	4	25	55
MA1	7	7	12	5	7	4	25	52
NMA	5	7	7	5	4	6	5	5
TAR	7	7	14	6	5	5	22	51
ARCH	100	100	99	94	62	3	99	100
GARCH	43	99	77	86	43	5	7	32
MS1	3	6	32	18	6	4	4	7
MS2	22	95	90	100	78	11	4	14
MS3	19	97	91	100	83	9	5	22
MS4	19	96	82	100	82	10	6	31