Evaluation of linear trend tests using resampling techniques

Vafeiadis Thanasis, Efthimia Bora-Senta, Dimitris Kugiumtzis

Evaluation of linear trend tests using resampling techniques

Journal:	*Communications in Statistics - Simulation and Computation*
Manuscript ID:	LSSP-2007-0042.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	21-Nov-2007
Complete List of Authors:	Thanasis, Vafeiadis; Aristotle University of Thessaloniki, Mathematics Bora-Senta, Efthimia; Aristotle University of Thessaloniki, Mathematics Kugiumtzis, Dimitris; Aristotle University of Thessaloniki, Mathematical, Physical and Computational Sciences
Keywords:	Time series , Trend tests, Resampling techniques
Abstract:	Resampling techniques, of both bootstrap and surrogate data type, are used in the evaluation of linear trend tests. Monte Carlo simulations were done for several distributions of correlated and independent residuals. In particular for AR(1) residuals, the discrimination of strong autocorrelation from linear trend is investigated with respect to the sample size. The overall results show that resampling reduces the type I and II errors of the trend tests. Following the guidelines suggested by the simulation results, we could find significant linear trend in the data of land air temperature and sea surface temperature.

Evaluation of linear trend tests using resampling techniques

Vafeiadis Thanasis[1], Bora-Senta Efthimia[1], Kugiumtzis Dimitris[2]

[1] Division of Statistics and Operation Research, Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece

[2] Department of Mathematical, Physical and Computational Sciences, Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki, 54124 Greece

Abstract

A number of parametric and non-parametric linear trend tests for time series are evaluated in terms of test size and power, using also resampling techniques to form the empirical distribution of the test statistics under the null hypothesis of no linear trend. For resampling, both bootstrap and surrogate data are considered. Monte Carlo simulations were done for several types of residuals (uncorrelated and correlated with normal and non-normal distributions) and a range of small magnitudes of the trend coefficient. In particular for AR(1) **and ARMA(1,1)** residual process**es**, we investigate the discrimination of strong autocorrelation from linear trend with respect to the sample size. The correct test size is obtained for larger data sizes as autocorrelation increases and only when a randomization test that accounts for autocorrelation is used. The overall results show that the type I and II errors of the trend tests are reduced with the use of resampled data. Following the guidelines suggested by the simulation results, we could find significant linear trend in the data of land air temperature and sea surface temperature.

Key words: time series, linear trend tests, resampling techniques, surrogate data, bootstrap.

1 Introduction

The investigation of long-term trends in times series is an important issue in many applications. Long-term trends can be considered as stochastic trends attributed to power law autocorrelation decay, referred to as long term persistence (see e.g. Rybski et al, 2006), or as deterministic trends, which will be the focus of this work. The formal statistical approach for the latter is a test for the presence of a linear trend in the time series. Such tests have been used in many areas of climatology, such as global warming (Woodward and Gray, 1993; Cohn and Lins, 2005), in meteorology, such as rainfall (Bonaccorso et al, 2005) and temperature (Xu et al, 2002; Feidas et al, 2004), and in hydrology, such as stream flow (Wang et al, 2005; Yue et al 2002).

The standard decomposition of a time series Y_t, $t = 1, \ldots, n$, under the assumption of a linear trend reads

$$Y_t = a + bt + E_t, \qquad (1)$$

where a is a constant, b represents the magnitude of the trend and E_t is the residual. The null hypothesis for the trend test is $\mathrm{H}_0 : b = 0$. Rejection of H_0 establishes the presence of linear trend in the time series, provided that the model for the residuals is valid.

Many trend tests assume independent residuals, such as the rank-based non-parametric Mann-Kendall (MK) test (Mann, 1945; Kendall, 1975) and the parametric regression-based test (Woodward et al, 1993). For the latter, the statistic is simply the estimated trend coefficient $\hat{b}$ standardized with its standard error. When the residuals are short-term correlated, simple corrections in the estimation of the standard error of $\hat{b}$ are suggested making use of the autocovariance (Grenander, 1954) and the spectrum of E_t (Bloomfield

and Nychka, 1992). More involved schemes adjusting the solution for the trend coefficient in the presence of autocorrelation in the time series have been proposed in (Sun and Pantula, 1999; Roy et al, 2004).

Strong positive autocorrelation in the time series may form monotonic trend and give rise for false rejection for the trend test. This has been shown with Monte Carlo simulations and different estimators for the linear trend (Woodward and Gray, 1993; Sun and Pantula, 1999; Roy et al, 2004; Kim et al, 2003). On the other hand, a deterministic trend may alert the sample autocorrelation used in the statistic of the trend test (Fried and Imhoff, 2003).

Besides the correlation in the residuals, the distribution of the residuals, as well as the sample size, affect heavily the outcome of the test, depending also on the magnitude of the trend. In this work, we address all these factors for four standard tests. Moreover, we introduce randomization and bootstrap versions of the tests in an attempt to improve the performance of the tests. Particular emphasis is given on the limits of sample size that maintain small type I and II errors. The investigation is done using Monte Carlo simulations at different settings of time series length, magnitude of trend coefficient, as well as distribution and linear structure of the residual process. We considered also a real application and applied the tests to a time series of an index that combines land air temperature anomalies (Jones, 1994a) and sea surface temperature anomalies (Parker et al, 1995) on a 5o x 5o grid box basis, developed by the Climatic Research Unit (CRU) of University of East Anglia (http://www.cru.uea.ac.uk/).

The trend tests are presented in Section 2 and the simulation setup and results in Section 3. In Section 4, the application is presented and in Section 5 the conclusions are drawn.

2 Statistical testing of linear trend

In general, it is difficult to detect the small linear trend with eyeball judgment, and there is need for an accurate and sensitive trend test in order to assess the significance of such weak

linear trends that are often investigated in small time series, as for example in meteorology (Feidas et al, 2004). In the following, we briefly present four standard linear trend tests, three parametric and one non-parametric test.

2.1 Parametric trend tests

Under the assumption of independent and normally distributed residuals E_t with zero mean and variance σ^2, $E_t \sim \mathrm{N}(0,\sigma^2)$, from the regression of Y_t on time t, the least square estimator $\hat{b}$ is obtained as

$$\hat{b} = \frac{\sum_{t=1}^{n}(t-\bar{t})Y_t}{\sum_{t=1}^{n}(t-\bar{t})^2}, \qquad (2)$$

where $\bar{t}$ is the mean time.

The estimated standard error of $\hat{b}$ is given by

$$\hat{s}_1(\hat{b}) = \left[\frac{\sum_{t=1}^{n}(Y_t-\hat{a}-\hat{b}t)^2}{(n-2)\sum_{t=1}^{n}(t-\bar{t})^2}\right]^{1/2} = \left[\frac{12\sum_{t=1}^{n}(Y_t-\hat{a}-\hat{b}t)^2}{(n-2)n(n^2-1)}\right]^{1/2}, \qquad (3)$$

where $\hat{a} = \bar{Y} - \hat{b}\bar{t}$ and $\bar{Y}$ is the mean of the time series. Then the test statistic referred to as C1 is $t = \dfrac{\hat{b}}{\hat{s}_1(\hat{b})}$ and follows the Student distribution with n-2 degrees of freedom $t \sim t_{n-2}$ (Woodward and Gray, 1993).

When the residuals E_t are correlated, the estimated standard error of $\hat{b}$ is given by

$$\hat{s}_2(\hat{b}) = \left\{\frac{12}{n(n^2-1)}\left[\gamma_0 + \frac{24}{n(n^2-1)}\sum_{s=2}^{n}\sum_{t=1}^{s-1}(t-\bar{t})(s-\bar{t})\gamma_{s-t}\right]\right\}^{1/2}, \qquad (4)$$

where γ_k denotes the k-th order autocovariance of E_t (Grenander, 1954). Replacing in (4) γ_k with the respective estimate

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k} \hat{E}_{t+k}\hat{E}_t, \quad (5)$$

where E_t can be estimated by $\hat{E}_t = Y_t - \hat{a} - \hat{b}t$, except at $k=0$ where we use $n\hat{\gamma}_0 / n-2$ to estimate γ_0, the estimated standard error of $\hat{b}$, $\hat{s}_2(\hat{b})$, is derived. This is used to form the test statistic $t = \dfrac{\hat{b}}{\hat{s}_2(\hat{b})}$ referred to as C2 and it holds as before $t \sim t_{n-2}$.

In a different approach, the standard error of $\hat{b}$ is estimated from the power spectrum (Bloomfield and Nychka, 1992)

$$\hat{s}_3(\hat{b}) = \left[2\int_0^{0.5} W(f)S(f)df \right]^{1/2}, \quad (6)$$

where $W(f) = \left| \sum_{t=1}^{n} m_t e^{-2\pi i f t} \right|^2$ with $m_t = \dfrac{t-\bar{t}}{\sum_{t=1}^{n}(t-\bar{t})^2}$ and $S(f)$ denotes the sample spectrum of E_t given as

$$S(f_j) = \left(\frac{1}{2\pi}\right)\left(\hat{\gamma}_0 + 2\sum_{k=1}^{n-1}\hat{\gamma}_k \cos(f_j k)\right) \text{ where } f_j = \frac{2\pi j}{n}, j = 0,...,n/2.$$

The test with statistic $t = \dfrac{\hat{b}}{\hat{s}_3(\hat{b})} \sim t_{n-2}$ is denoted as C3.

2.2 Non-Parametric trend tests

Two non-parametric rank-based statistical tests, namely the Mann-Kendall (MK) test, also called Kendall's tau test due to Mann (1945) and Kendall (1975), and the Spearman's rho test (Lehmann, 1975; Sneyers, 1990) are used for detecting trend in time series data. Yue et al (2002) showed that these two tests have almost the same power to identify trends in time series data. In our study, we use the rank-based non-parametric Mann-Kendall (MK) test,

which seems to have been used more often in applications, such as stream flow (Yue et al, 2002; Yue and Pilon, 2004, Wang et al, 2005).

The null hypothesis for the MK test is that the time series Y_t, $t = 1, 2, ..., n$, is independent and identically distributed. The statistic S of Kendall's tau is

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}(Y_j - Y_i), \quad (7)$$

where

$$\text{sgn}(\theta) = \begin{cases} 1, & if\ \theta > 0 \\ 0, & if\ \theta = 0 \\ -1, & if\ \theta < 0 \end{cases}.$$

Mann (1945) and Kendall (1975) documented that when $n \geq 8$ the statistic S is approximately normally distributed with the mean and the variance as follows:

$$\text{E}(S) = 0$$

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{m=1}^{n} t_m m(m-1)(2m+5)}{18},$$

where t_m is the number of ties of extent m.

2.3 Randomization and bootstrap tests

The tests described above have all well defined asymptotic null distribution, i.e. distribution of the test statistic under H_0. However, departures from the nominal null distribution may occur, e.g. due to small sample size. Resampling techniques have been used to form the null distribution. Here we consider randomization and bootstrap tests.

A randomization test generates a randomly chosen subset of all possible permutations of the original sample consistent to H_0 (Fortin et al, 2002). The randomization tests used in this

study are adjusted to the correlation structure of the time series and the distribution of residuals E_t.

When the residuals are white noise, $E_t \sim WN(0,\sigma^2)$, the randomized or so-called surrogate data are generated by shuffling the Y_t time series. In case where E_t stem from a stochastic linear process, different surrogate data generating algorithms are called depending on whether the process is normal or not. When it is normal, the surrogate time series are generated by phase randomization making use of the Fourier transform and are referred to as FT surrogate data (Theiler et al, 1992). An FT surrogate time series is a normal time series with the same linear structure as Y_t, but contains no trend. In case E_t comes from a non-normal stochastic linear process, the more general algorithms of Improved Amplitude Adjusted Fourier Transform, IAAFT (Schreiber and Schmitz, 1996) and Statically Transformed Autoregressive Process, STAP (Kugiumtzis, 2002) are called. The IAAFT algorithm makes also use of the Fourier transform but in an iterative scheme that terminates when sufficient convergence of both power spectrum and marginal distribution is reached. The STAP algorithm generates the surrogate time series as statically transformed realizations of a normal (autoregressive) process so that both the original marginal distribution and linear structure are preserved. These algorithms were introduced to test nonlinear departures from the null hypothesis of linear stochastic process, but they can as well be used to test departures involving linear trend. The test using STAP surrogate data is more conservative than when using IAAFT surrogates and its power decreases faster with the decrease of the time series length (Kugiumtzis 2002). We employ the randomization test with both surrogate data types and compare their size and power on small sample sizes, typically encountered in trend investigation.

In addition to randomization tests we include bootstrap tests, adapted for each model assumption for E_t (Hinkley, 1988; Efron and Tibshirani, 1993). For white noise residuals, the standard bootstrap resampling is applied. When E_t is a realization of a linear stochastic process, there are a number of bootstrap approaches, such as the block, sieve, wild and local bootstraps, but we follow here the most standard "naïve" bootstrap approach, fitting an autoregressive model and drawing from the model residuals to generate the bootstrap time series (Buhlmann, 2002; Politis 2003).

3 Monte Carlo Simulations

3.1 Simulation setup

We generate Monte Carlo realizations for different stochastic processes with and without linear trend according to the model in (1). The length of the time series n varies as 2^k for $k = 4,5,6,7$ and the trend magnitude is monitored varying the linear trend coefficient as $b = -0.01(0.002)0.01$, where the no-trend scenario is for $b = 0$. For white noise residuals E_t, the normal, uniform and exponential distributions are considered. For correlated residuals E_t, **we consider** the first order autoregressive process **AR(1)**, $E_t = \varphi E_{t-1} + a_t$, **and the mixed process of first order autoregressive part and first order moving average part ARMA(1,1)** $E_t = \varphi E_{t-1} - \theta a_{t-1} + a_t$**,** where a_t follows normal, uniform and exponential distribution. In order to examine how the correlation in the residuals affects the detection of linear trend, we vary also the parameter φ (equal to one lag autocorrelation of residuals) as $\pm 0.95, \pm 0.8$ and ± 0.4. The combination of all the values of *n*, *b*, and φ (including zero) and the distribution types of input noise, gives a total of 4x11x7x3=924 cases. **For ARMA, the study is not exhaustive and is restricted to selected values of φ and θ.**

For each case, 1000 Monte Carlo realizations are generated and for each realization 199 surrogate and bootstrap data are generated by the appropriate algorithm. For the correlated residuals, both IAAFT and STAP surrogate data are generated along with the bootstrap data. The four tests are applied on each time series and the test decision is made on the basis of the analytic null distribution of the test statistic q. In addition, the null distribution is formed from the values of q computed on the resampled data, denoted as $q^1,...,q^{199}$, and the rejection of H_0 is deduced when q^0 computed on the original time series is not within the null distribution. In previous works on surrogate data, the rejection of H_0 is often determined from the significance S (provided that $q^1,...,q^M$, on the M surrogate data, are fairly normally distributed) denoted as

$$S = \frac{\left|q^0 - \bar{q}\right|}{s_q}, \tag{8}$$

where $\bar{q}$ is the average and s_q the standard deviation (SD) of $q^1,...,q^M$ (Kugiumtzis, 2000). Rank ordering has also been used, where for our case, H_0 is rejected, say, at significance level $\alpha = 0.01$ when q^0 is first or last in the ordered concatenated list $q^0, q^1,...,q^{199}$ and at $\alpha = 0.05$ when q^0 is at places 1 to 5 or 196 to 200.

3.2 Simulation results

All linear trend tests are performed using the test statistics C1, C2, C3 and MK and the test decision is made using the asymptotic approach and the randomization and bootstrap approach. The probability of rejecting H_0 is estimated by the relative frequency of rejections in the ensemble of 1000 time series.

The significance S in (8) gives better resolution in the p-value than the rank ordering. However, normality tests on the statistics from a sample of 199 surrogates showed departures

from normality. Therefore, the surrogate and bootstrap test results below are for rank ordering.

One would expect that all the tests perform well in the case of independent residuals. However, even when the residual series is normal white noise, the asymptotic test shows larger size when using C2 and small power when using C3, given with highlighted values in Table 1. For C2 this is a significant drawback that persists for other distributions of E_t (e.g. uniform noise in Table 1) and questions the detection of trend with this method (e.g. note the higher probability of rejection for b=±0.002 as compared to C1 and MK). The shortcomings of C2 and C3 are recovered with the use of the randomization approach. Indeed randomization tests attain always the correct size of the test and the same level of power as the asymptotic approach. The results on non-zero trend coefficients suggest that C1 and MK, constructed under the assumption of independent residuals, have somehow larger power than the C2 and C3 statistics for any white noise distribution. The distribution of E_t seems to affect the significance of the linear trend, e.g. the power of all tests is increased when the distribution changes from normal to uniform (see Table 1). The results of the bootstrap tests are the same as for the respective randomization tests.

(Here should be placed Table 1)

For correlated residuals, the degree of correlation, monitored in the simulations with the coefficient φ of the AR(1) model for E_t, in combination with the time series length have major effect on the size and power of all tests. On the other hand, the distribution of E_t (actually we determine the distribution of the input noise a_t of AR(1) in the simulations) does not seem to have significant effect on the test accuracy.

The simulations showed that when consecutive residuals are anti-correlated (φ negative **in AR(1)**) the null distribution of the t statistic of C2 tends to be wider than the respective nominal distribution, whereas for C1, C3 and MK tests it is narrower, as shown in Fig. 1. For

example, in the absence of trend and for $n=128$, the estimated variance of the C2 statistic when $\varphi=-0.8$ is 1.74 and when $\varphi=-0.2$ is 2.8, which are both far from the nominal unit variance. For the other tests, the estimated variance is much smaller than the nominal unit variance. This is observed for all three types of noise.

(Here should be placed Figure 1)

Thus the asymptotic approach tends to give larger test size for C2 test and smaller power for the other statistics. This is shown in Fig. 2a and 2d for **AR(1) residuals with** $\varphi=-0.8$ and $\varphi=-0.4$, respectively, where the data size is n=128 and a_t follows normal distribution. The power of all tests increases with the decrease of anticorrelation (φ closer to zero) and the increase of the magnitude of b, similarly for upward and downward trend. C2 has the largest power but spuriously given the large test size for b=0, and that is because $\hat{s}_2(\hat{b})$ is a poorly behaved estimator (Woodward et al, 1993). The other three test statistics perform similarly having insignificant power for small b (see Fig. 2a). The respective randomization tests using FT surrogates give better results: they eliminate the type I error of the asymptotic tests for C2, with a loss of power (see Fig. 2b and 2e). The randomization tests using C2 and C3 tend to have more symmetric increase of power than C1 and MK (for positive and negative b). When bootstrap data are used, the power of all tests is further improved and all four tests perform similarly, as shown in Fig. 2c and 2f.

(Here should be placed Figure 2)

Similar results are obtained when E_t is an AR(1) process with uniform or exponential input white noise. The test results for the two noise distributions are shown in Fig. 3, for $\varphi=-0.4$. Note that C2 and C3 attain larger power when resampling techniques are used, especially when the input noise is uniform.

(Here should be placed Figure 3)

STAP gives similar results to IAAFT for the randomization test in these non-normal AR(1) processes. Both randomization and bootstrap tests eliminate type I error for all statistics, but bootstrap tests obtain somehow larger power than the randomization tests.

The actual null distribution of the test statistics deviates from the nominal null distribution also when φ **in AR(1) residuals** is positive but in a different way and at a larger degree (compare Fig. 4 to Fig. 1). For example, for $b=0$, normal white noise and $n=128$, as φ coefficient increases from 0.2 to 0.8, the variance jumps from 1.49 to 9.44 for C1, from 3.15 to 4.4 for C2, from 0.29 to 0.68 for C3 from 4.37 to 7.67 for MK test. This explains the very large test size we found when using C1, C2 and MK with the asymptotic approach. All test statistics, except for C3, have much wider empirical distribution as shown in Fig. 4 for normal and uniform noise.

(Here should be placed Figure 4)

The Monte Carlo simulations showed that the empirical test size gets larger for the asymptotic tests as φ **in AR(1) residuals** increases away from 0.4. The same problem was found also for the tests using resampling techniques but at a lesser amount. Among all test statistics, only C3 performs properly for large positive autocorrelation. In Fig. 5a, the test results using C3 are shown for $\varphi=0.8$, $n=128$ and the uniform input noise. There is still a small type I error, especially for IAAFT and STAP randomization tests. On the other hand, the bootstrap test eliminates the type I error at the cost of smaller power compared to IAAFT and STAP. For exponential input noise all tests do not have any significant power (see Fig. 5b) and the same holds for normal input noise (not shown here).

(Here should be placed Figure 5)

The power increases fast for larger data sets, as shown in Fig. 6 for C3, bootstrap approach and $n = 256$, i.e. double than the sample size in Fig. 5. For strong positive correlations as for $\varphi = 0.8$ used in Fig. 5 and Fig. 6, the increase of the power with the sample size is lower for normal and exponential input white noise.

(Here should be placed Figure 6)

The test results when the residuals are from an ARMA(1,1) turn out to be similar to the results shown above for AR(1) residuals, at least for the corresponding values of φ that we tested for. For examples, as shown in Fig. 7a the performance of the bootstrap tests with C2 and C3 for ARMA(1,1) residuals with $\varphi = -0.8$, $\theta = 0.8$ and normal input noise are substantially the same as the respective results for AR(1) shown in Fig.2c, whereas C1 and MK show less power for ARMA(1,1). This difference with C1 and MK gets larger when the randomization test is used instead (compare Fig. 7b with Fig. 2b). For all statistics the test for the same ARMA residual process improves both in terms of significance and power when the input noise is uniform, as we observed for the AR process (see Fig. 7c). Other values of φ and θ gave results similar to the corresponding AR(1) residual process. For example the results for $\varphi = -0.4$ show the same test performance for C2, C3 and difference for C1,MK for the ARMA and AR case as discussed earlier for $\varphi = -0.8$ (compare Fig. 7d to Fig. 2f). For positive values of φ in the ARMA(1,1) residual process, the power of the test (bootstrap and randomization) decreases in the same way as for the AR(1) residual process.

(Here should be placed Figure 7)

It is of practical interest to investigate the dependence of sample size on the strength of positive autocorrelation (positive φ) under the condition of maintaining the correct test size.

For this, we made Monte Carlo simulations for **AR(1) residual processes with** $\varphi = 0.4(0.05)0.95,0.97,0.99$ and we found the smallest n that preserves the actual size of the test within the limit of 0.06 for $a = 0.05$. The results for C3, bootstrap approach and the three distributions of input white noise are shown in Fig. **8**. It is clear that the demand on more data points increases faster as φ approaches 1, i.e. the random walk scenario that regards fully stochastic trend. It turns out that for uniform input white noise the correct application of the test (using C3 and bootstrap approach) does not require as long time series as for normal and exponential white noise.

(Here should be placed Figure 8)

According to the simulation results, C3 with the use of bootstrap test is the most suitable trend test to identify the presence of small linear trend in time series data under varying conditions of autocorrelation and amplitude distribution of the time series.

4 Application

We applied the asymptotic and resampling tests to time series of an index that combines land air temperature anomalies (Jones, 1994a) and sea surface temperature anomalies (Parker et al, 1995) on a 5o x 5o grid box basis, developed by the Climatic Research Unit (CRU) of University of East Anglia (http://www.cru.uea.ac.uk/).. We consider the time series for each month from January to December in the period from 1856 to 2005 ($n = 150$, the whole sample), referred to as period 1, and from 1961 to 2005 ($n = 45$), referred to as period 2. In addition, the records of the mean annual values for the two periods are analyzed. These time series show weak linear trends and we want to investigate whether these trends are significant. For example, as shown in Fig. **9**, the index of period 2 for January shows a long steep upward trend starting at around 1970 suggesting significance of the trend, whereas for

July the trend can be seen in a smaller part of the same period and is thus of questionable significance.

(Here should be placed Figure 9)

According to Akaike (AIC) and Swartz (BIC) criteria the order of the AR model of the residuals, for the most time series was 1 (and mostly for $n = 150$). The estimated coefficient of AR(1) was about 0.5 for most of the time series, but the distribution of the residuals did not appear to have the same form across the time series. For example, the Kolmogorov – Smirnov test for normality gave rejection for most of the time series. According to our simulation results for correlated residuals at the order of $\varphi \simeq 0.5$, the minimum sample size for the appropriate use of the trend test is at 100 to 150 (see Fig. **8**). In table 2, the standardized coefficients (s.c.) for the magnitude of the trend for all the time series are shown. The s.c. for period 2 are larger than for period 1 for all months except November (11).

(Here should be placed Table 2)

All asymptotic tests give significant linear trend for all months for period 2 and only for autumn and winter months for period 1, as shown in Fig. **10** using C3. This result cannot be trusted due to the presence of positively correlated residuals and according to the simulation results for sample sizes at the level of period 1 and 2. On the other hand, when the bootstrap and randomization tests were used, significant linear trend was found only for the winter and partly spring months. As shown for C3 and bootstrap in Fig. **10**, significant trend at $\alpha = 0.05$ was found for months September to May for period 1 and for months January to April for period 2. However, the test results for period 2 should be treated with caution as for such a small sample size the presence of positive autocorrelation in the residuals (here it is about 0.5) may be the cause of the statistically significant trend (see also Fig. **8**). Note in particular that the linear trend for January of period 2 was found significant with all approaches,

whereas for July only the asymptotic approach found significant trend (see also Fig. **10**). According to the simulation results, we should thus trust the bootstrap test for July of period 2.

(Here should be placed Figure 10)

As for the mean annual time series (at point 13 of the horizontal axis of Fig. **10**), C3 asymptotic and bootstrap tests give significant linear trend for period 1, whereas for period 2 the linear trend was found significant for the asymptotic but not for the bootstrap test.

The overall results from the test of C3 and bootstrap approach suggest that there is a trend during the winter and spring months, better expressed in the long record (1856 – 2005).

5 Conclusion

Monte Carlo simulations were made on four test statistics for asymptotic, randomization and bootstrap test of linear trend under different settings of time series length, residual distribution and autocorrelation. The comparative results showed clear superiority of the randomization and bootstrap test over the asymptotic test and revealed differences and limitations in the performance of the test statistics.

For correlated residuals, the C3 test statistic, using spectrum-based estimation of the variance of the slope coefficient, gives the smallest size of the asymptotic test and when resampling techniques are used the test size decreases to the nominal level. Further, it attains high power compared to the other test statistics. However, when the residuals are white noise, the power of the test using C3 is smaller than when using a test statistic formed under the assumption of white noise residual.

The asymptotic test gives generally large type I error and the use of resampling techniques recovers the correct test size in most of the settings considered in the study. For correlated residuals, suitable surrogate data generation techniques have been used for the randomization

test and the residual-based bootstrap for the bootstrap test. The simulation results showed that the bootstrap test turns out to attain higher power than the randomization test.

The overall simulation results suggest the use of the C3 statistic in a bootstrap test. Even this test cannot distinguish linear trend from strong positive autocorrelation depending on the time series length. We found that under the condition of retaining the correct test size, the time series length has a functional dependence on the positive autocorrelation (for values larger than about 0.4) that varies with the input white noise distribution. These functional relations can serve as a guide for the limits of implementation of the test in real-world applications.

We applied the asymptotic and resampling tests with the four test statistics to time series of an index of land air and sea surface temperature anomalies at different periods, for all 12 months separately and for the annual average. For some months, a linear trend was found for some statistics using the asymptotic test (and sometimes even the resampling test) whereas it was not found when using C3 and the bootstrap test, indicating spurious detection of trend. However, consistent detection of trend could be obtained in the winter and spring months, especially when considering the whole record that allows for a proper implementation of the test, given also the relatively small positive autocorrelation of the residuals.

We believe that this work shed some light on the performance of standard tests for linear trend and showed the need of resampling techniques in the implementation of the tests. There are other tests for linear trend not considered in this work and it would be interesting to include them in a future comparative work.

References

Bloomfield P. and Nychka D. (1992). Climate Spectra and Detecting Climate Change. Climatic Change 21:275-287

Bonaccorso B., Cancelliere A. and Rossi G. (2005). Detecting trends of extreme rainfall series in Sicily. Advance in Geosciences 2:7-11

Bowerman - O' Connell (2000). *Forecasting and Time Series. An applied approach*. Third Edition.Duxbury Press.

Brockwell J. P., Richard A. D. (2002). *Introduction to Time Series and Forecasting*. Second Edition. Springer.

Buhlmann P. (2002). Bootstrap for time series. Statistical Science 17:52-72

Chandler R. (2002). Trend Analysis For The Environmental Science-A Review, ESSG Meeting, March 2002

Cohn A.T. and Lins F.H. (2005). Nature's style: Naturally trendy. Geophysical Research Letters 32:L23402

Cryer D. J. (1986) *Time Series Analysis*. PWS-KENT Publishing Company- Boston.

Efron B., Tibishirani J. R.(1993). *An Introduction to the Bootstrap.* Chapman and Hall/CRC

Feidas H., Makrogiannis T. and Bora-Senta E. (2004). Trend analysis of Air Temperature Time Series Data In Greece Determined By Ground and Satellite Data. Theoretical and Applied Climatology 79:185 - 208

Folland, C.K., Rayner, N.A., Brown, S.J., Smith, T.M., Shen, S.S.P., Parker, D.E., Macadam, I., Jones, P.D., Jones, R.N., Nicholls, N. and Sexton, D.M.H., (2001). Global temperature change and its uncertainties since 1861. Geophysical Research Letters 28:2621-2624.

Fortin M.-J., Jasquez G. & Shipley B. (2002). Computer-intensive methods. Encyclopedia of Environmetrics 1:399-402

Fried R. and Imhoff M. (2004). On the online detection of monotonic trends in time series. Biometrical Journal 46:90-102

Grenander U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. The Annals of Mathematical Statistics 25:252-272

Hinkley V. D. (1988). Bootstrap Methods. Journal of Royal Statistical Society. Series B (Methodological) 50:321-337

Jones P. D. (1994). Hemispheric surface air temperature variations: a reanalysis and an update to 1993. Journal of Climate 71794-1802

Kim T. H., Pfaffenzeller S., Rayner T. and Newbold P. (2003). Testing for linear trend with application to relative primary commodity prices. Journal of Time Series Analysis 24:539-551

Kugiumtzis D. (2000). Surrogate Data Test on Time Series. In: A. Sool, L. Cao, *Nonlinear Deterministic Modeling and Forecasting of Economics and Financial Time Series,* Kluwer Academic Publishers

Kugiumtzis D. (2002). Statically transformed autoregressive process and surrogate data test for nonlinearity. Physical Review 66:025201(R)

Mann B. H. (1945). Nonparametric Tests Against Trend. Econometrica 13:245-259 (No.3)

Nordgaad A., Grimvall A. (2006). A resampling technique for estimating the power of non-parametric trend tests. Environmetrics, 17:257-267

NIST/SEMATECH e-Handbook of Statistical Methods. www.nist.gov/stat.handbook

Parker DE, Folland CK, Jackson M (1995). Marine surface temperature observed variations and data requirements. Climatic Change 31:559-600

Politis, D. N. (2003). The Impact of Bootstrap Methods on Time Series Analysis. Statistical Science 18:219-230

Roy A., Falk B. and Fuller A. W (2004). Testing for trend in the presence of autoregressive error. Journal of American Statistical Association 99:1082-1091

Rybski D., Bunde A., Havlin S., and Storch H. (2006). Long-term persistence in climate and the detection problem. Geophysical Research Letters 33L06718

Schreiber T. and Schmitz A (1996). Improved Surrogate Data for Nonlinearity Tests. Physical Review Letters 77:635- 638

Sun H. and Pantula G. S. (1999). Testing for trend in correlated data. Statistics & Probability Letters 41:87-95

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. (1992).Testing for nonlinearity in time series: The method of surrogate data. Physica 58:77-94.

Vandaele W. (1983). *Applied Time Series and Box-Jenkins Models*. INC, Harcourt Brace and Company. Academic Press

Wang W., Van Gelder P.H.A.J.M. and Vrijling J.K. (2005). Trend and stationary analysis for streamflow processes of rivers in Western Europe in the 20th century. IWA International Conference on Water Economics, Statistics and Finance, Rethymno, Greece, 8-10 July 2005

Woodward A. Wayne and H. L. Gray (1993). Global Warming and the Problem of Testing for Trend in Time Series Data . American Meteorological Society 6:953-962

Woodward A. Wayne, Bottone Steven and H. L. Gray (1997). Improved Tests for Trend in Time Series Data. Journal of Agricultural, Biological and Environmental Statistics 2:403-416

Xu. X. Z, Takeuchi K. and Ishidaira H. (2002). Long term trends of annual temperature and precipitation time series in Japan. Journal of Hydroscience and Hydraulic Engineering 20:11-26

Yue S., Pilon P. (2004). A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. Journal of Hydrological Science 49:21-37

Yue S., Pilon P. and Cavadias G. (2002). Power of Mann-Kendall and Spearman's rho tests for detecting monotonic trend in hydrological series. Journal of Hydrology 259:254-271

Yue S., Pilon P. Phinney B. and Cavadias G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes 16:1807-1829

Tables and Figures

Table 1.

		E_t ~ N(0,1)				E_t ~ U[-1/2,1/2]			
b	**Test (α=0.05)**	**C1**	**C2**	**C3**	**MK**	**C1**	**C2**	**C3**	**MK**
b=-0.004	Asymptotic	0.375	0.645	0.101	0.377	1.000	1.000	0.999	0.972
	Randomization	0.328	0.270	0.235	0.307	1.000	0.990	0.994	0.999
b=-0.002	Asymptotic	0.133	0.343	**0.029**	0.117	0.820	0.931	0.414	0.777
	Randomization	0.104	0.104	0.108	0.108	0.780	0.642	0.667	0.739
b=0.0	Asymptotic	0.061	**0.206**	0.013	0.057	0.048	**0.241**	0.002	0.044
	Randomization	0.049	0.046	0.061	0.049	0.048	0.053	0.051	0.050
b=0.002	Asymptotic	0.130	0.380	**0.031**	0.119	0.800	0.933	0.388	0.762
	Randomization	0.140	0.110	0.113	0.129	0.792	0.688	0.712	0.757
b=0.004	Asymptotic	0.401	0.655	0.099	0.377	1.000	1.000	1.000	0.961
	Randomization	0.389	0.328	0.199	0.386	1.000	0.991	0.995	1.000

Figure 1.

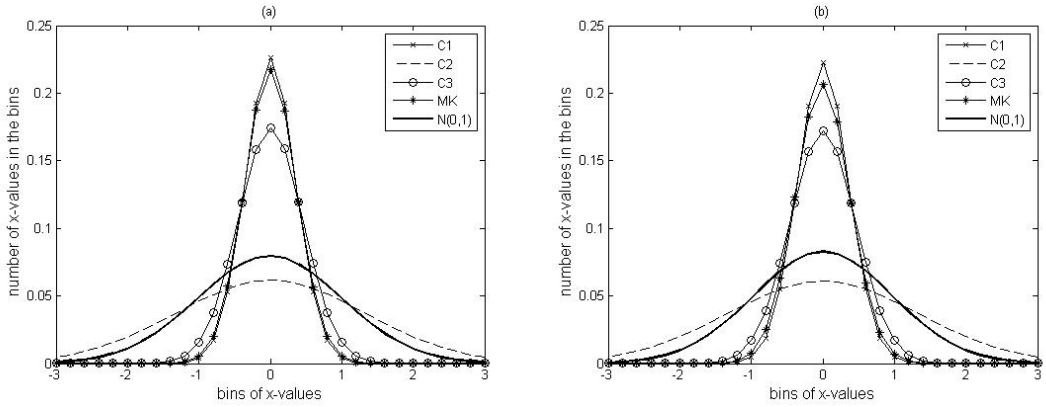


Figure 2.

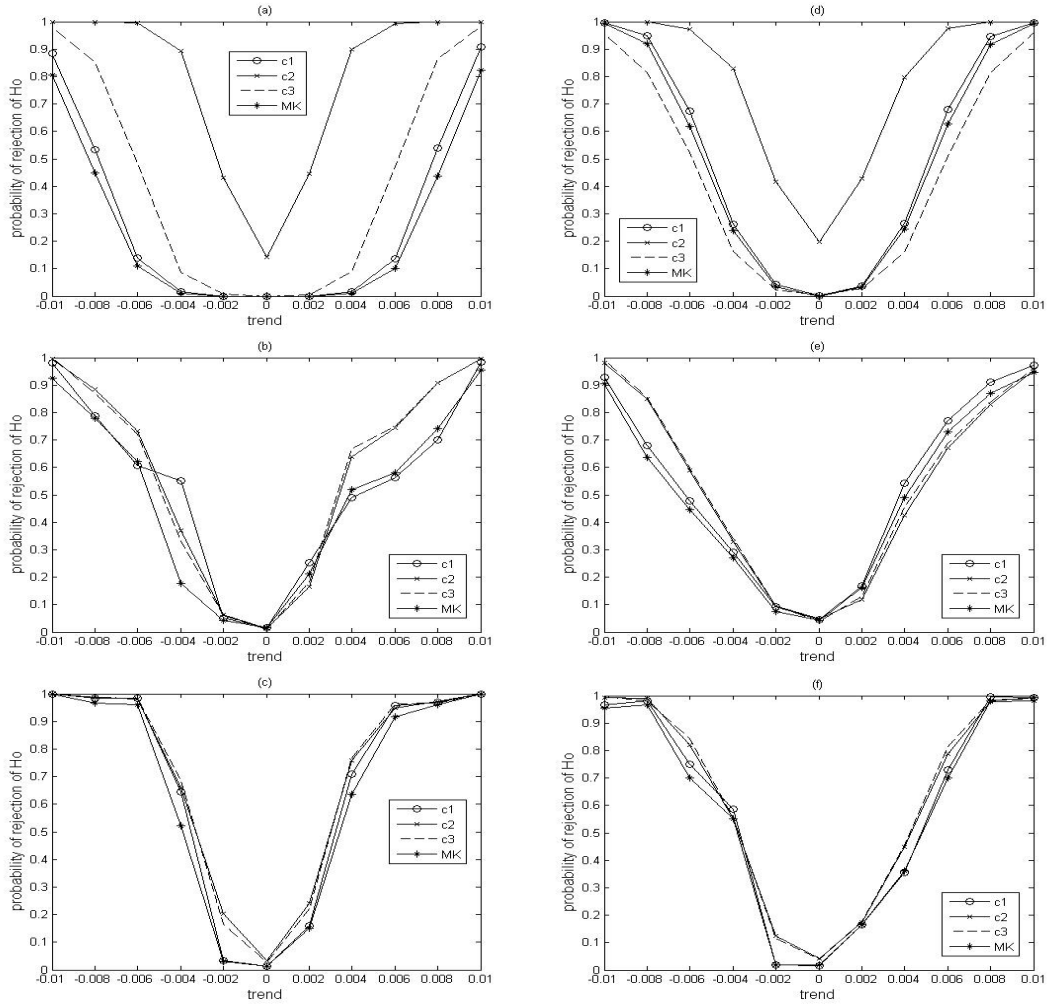


Figure 3.

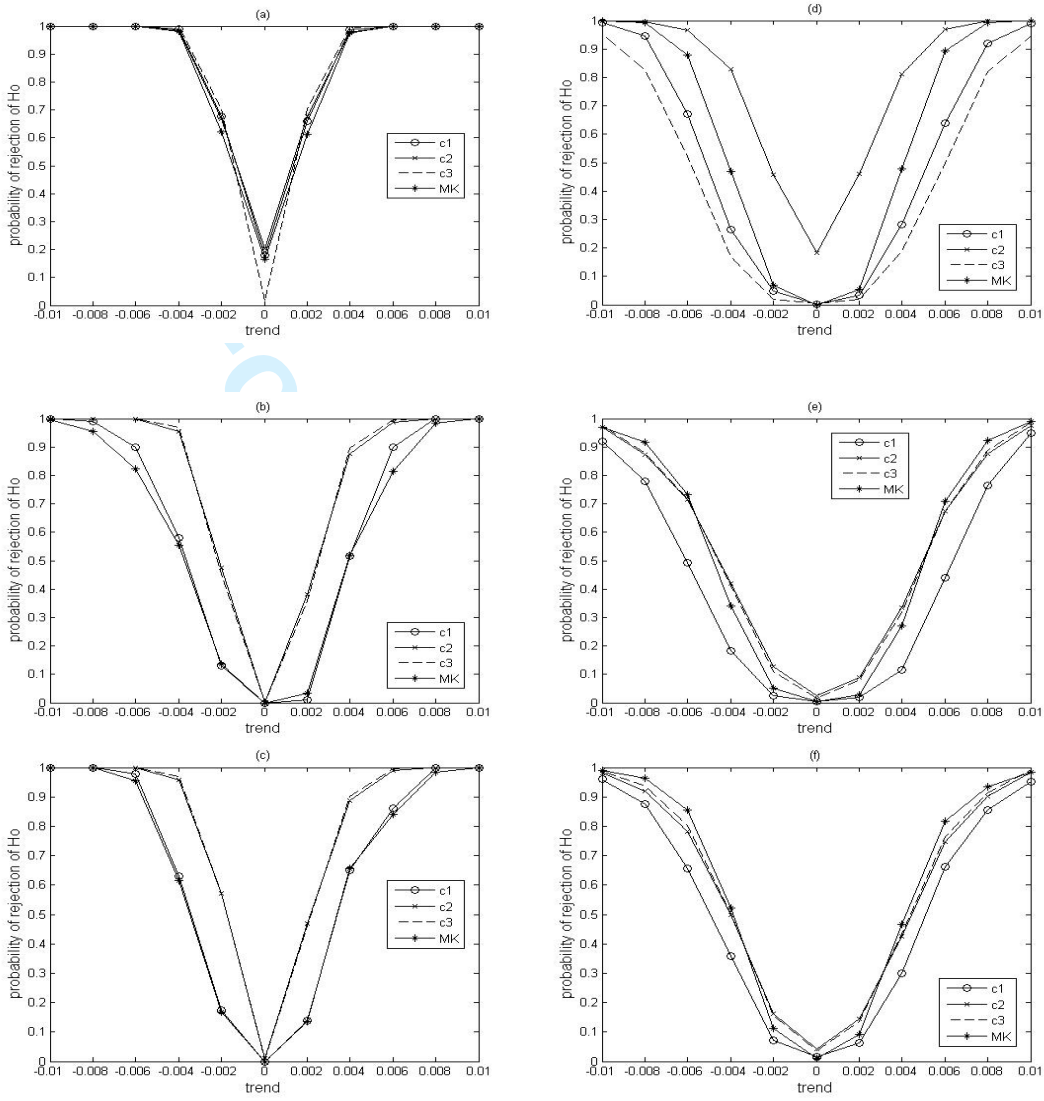


Figure 4.

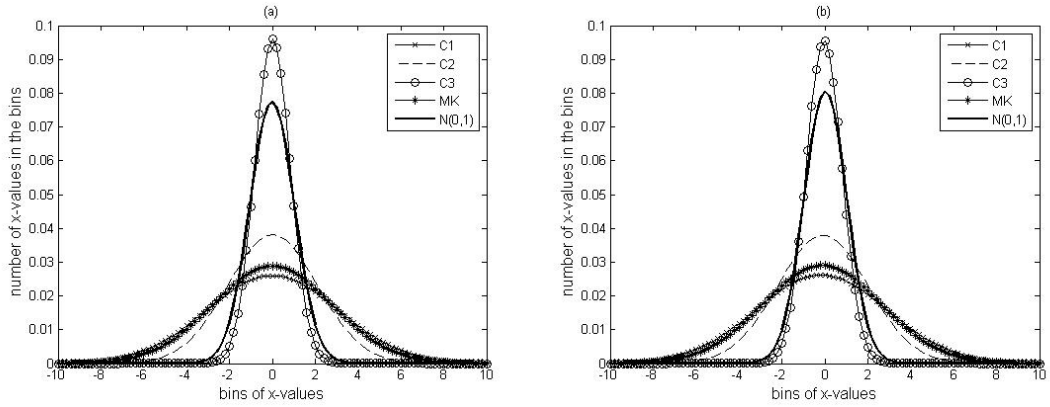


Figure 5.

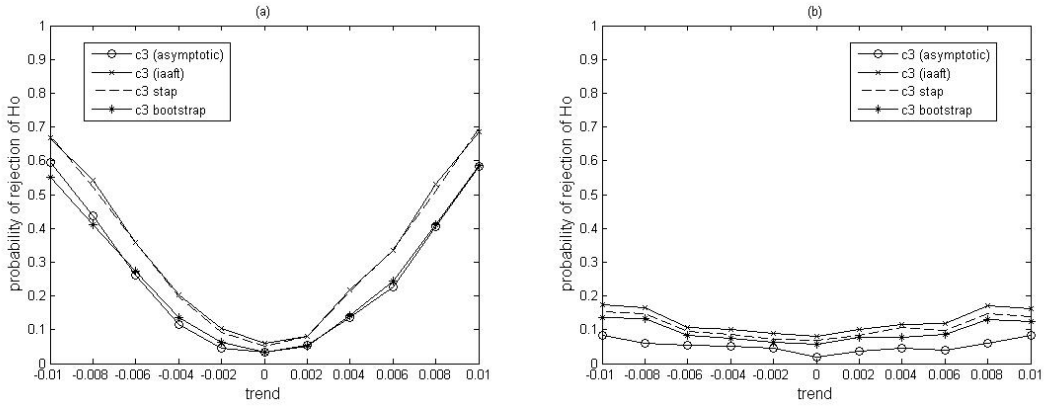


Figure 6.

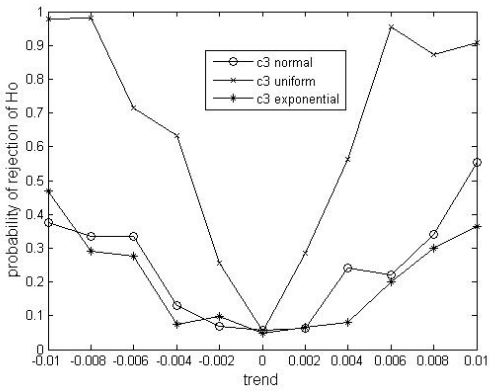


Figure 7.

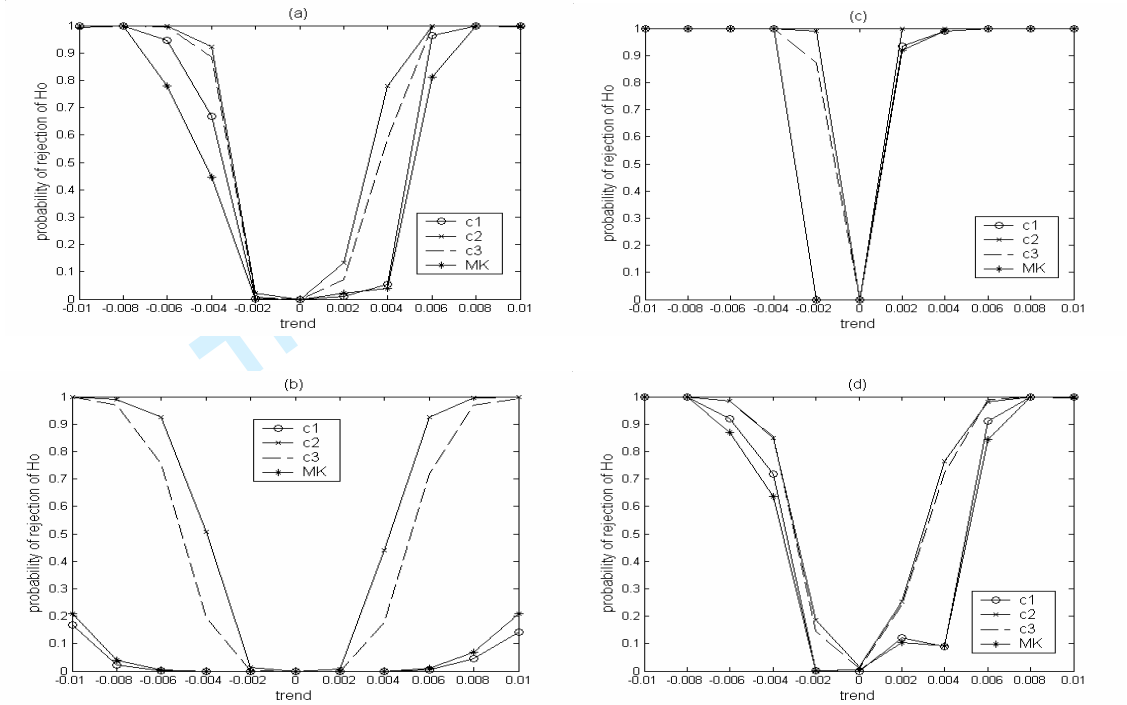


Figure 8.

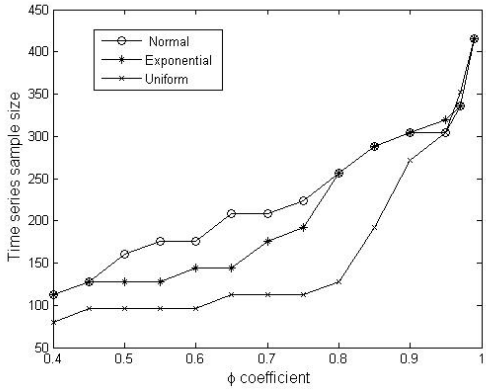


Figure 9.

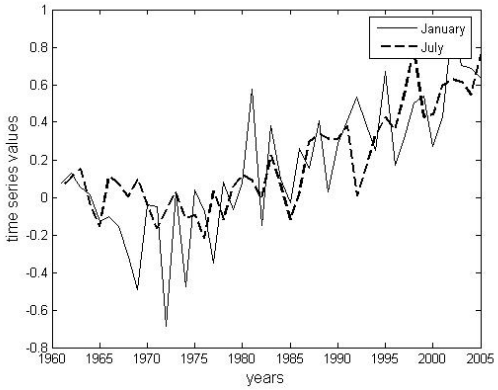
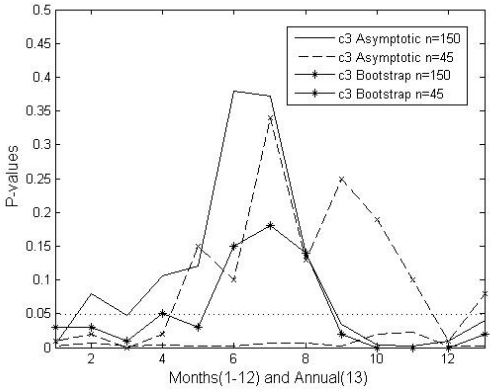


Table 2.

Period	Months												Annual
	1	2	3	4	5	6	7	8	9	10	11	12	
1	0.530	0.559	0.665	0.653	0.621	0.490	0.468	0.601	0.646	0.704	0.746	0.610	0.755
2	0.746	0.660	0.719	0.816	0.830	0.826	0.796	0.783	0.794	0.705	0.649	0.715	0.833

Figure 10.



Tables and Figures Captions

Table 1. Probability of rejecting H_0 when E_t is normal or uniform white noise (n=128) for all test statistics, asymptotic and randomization approach and five trend coefficients, given in the first column.

Figure 1. Estimated null distribution from 1000 realizations of C1, C2, C3 and MK trend statistics together with the standard normal distribution when a_t is normal white noise in (a) and uniform white noise in (b), where n=128 and φ=-0.8.

Figure 2. Probability of rejecting H_0 for the four trend statistics where E_t is generated by AR(1) with φ=-0.8 and normal input white noise. The asymptotic test is used in (a), the randomization test in (b), and the bootstrap test in (c). The same for φ=-0.4 in (d), (e) and (f), respectively.

Figure 3. Probability of rejecting H_0 for the four trend statistics where E_t is generated by AR(1) with φ=-0.4 and uniform input white noise. The asymptotic test is used in (a), the randomization (IAAFT) test in (b), and the bootstrap test in (c). The same is shown for the exponential input white noise in (d), (e) and (f), respectively.

Figure 4. Estimated null distributions from 1000 realizations of C1, C2, C3 and MK trend statistics together with standard normal distribution when a_t is normal white noise (a) and uniform white noise in (b) where n=128 and φ=0.8.

Figure 5. Simulation results for uniform (a) and exponential (b) white noise when φ=0.8 and n=128.

Figure 6. Bootstrap randomization test of C3, when φ=0.8 and n=256 for normal, uniform and exponential input white noise.

Figure 7. Probability of rejecting H_0 for the four trend statistics where E_t is generated by ARMA(1,1). (a) φ=-0.8, θ=0.8, normal input white noise and bootstrap test. (b) φ=-0.8, θ=0.8, normal input white noise and randomization test. (c) φ=-0.8, θ=0.8, uniform input white noise and bootstrap test. (d) φ=-0.4, θ=0.4, normal input white noise and bootstrap test.

Figure 8. Sample size as a function of φ for which the test using C3 and bootstrap approach attains a size less than 0.06 for a=0.05. The results are shown for different distributions of input white noise as shown in the legend.

Figure 9. The temperature time series of period 2 for January and July.

Table 2. The standardized coefficient for the linear trend estimated for all time series.

Figure 10. P-values of C3 statistic for periods 1856-2005 (n=150) and 1961-2005 (n=45) and for asymptotic and bootstrap test as given in the legend.

Tables and Figures Captions

Table 1. Probability of rejecting H_0 when E_t is normal or uniform white noise (n=128) for all test statistics, asymptotic and randomization approach and five trend coefficients, given in the first column.

Figure 1. Estimated null distribution from 1000 realizations of C1, C2, C3 and MK trend statistics together with the standard normal distribution when a_t is normal white noise in (a) and uniform white noise in (b), where n=128 and φ=-0.8.

Figure 2. Probability of rejecting H_0 for the four trend statistics where E_t is generated by AR(1) with φ=-0.8 and normal input white noise. The asymptotic test is used in (a), the randomization test in (b), and the bootstrap test in (c). The same for φ=-0.4 in (d), (e) and (f), respectively.

Figure 3. Probability of rejecting H_0 for the four trend statistics where E_t is generated by AR(1) with φ=-0.4 and uniform input white noise. The asymptotic test is used in (a), the randomization (IAAFT) test in (b), and the bootstrap test in (c). The same is shown for the exponential input white noise in (d), (e) and (f), respectively.

Figure 4. Estimated null distributions from 1000 realizations of C1, C2, C3 and MK trend statistics together with standard normal distribution when a_t is normal white noise (a) and uniform white noise in (b) where n=128 and φ=0.8.

Figure 5. Simulation results for uniform (a) and exponential (b) white noise when φ=0.8 and n=128.

Figure 6. Bootstrap randomization test of C3, when φ=0.8 and n=256 for normal, uniform and exponential input white noise.

Figure 7. Probability of rejecting H_0 for the four trend statistics where E_t is generated by ARMA(1,1). (a) φ=-0.8, θ=0.8, normal input white noise and bootstrap test. (b) φ=-0.8, θ=0.8, normal input white noise and randomization test. (c) φ=-0.8, θ=0.8, uniform input white noise and bootstrap test. (d) φ=-0.4, θ=0.4, normal input white noise and bootstrap test.

Figure 8. Sample size as a function of φ for which the test using C3 and bootstrap approach attains a size less than 0.06 for a=0.05. The results are shown for different distributions of input white noise as shown in the legend.

Figure 9. The temperature time series of period 2 for January and July.

Table 2. The standardized coefficient for the linear trend estimated for all time series.

Figure 10. P-values of C3 statistic for periods 1856-2005 (n=150) and 1961-2005 (n=45) and for asymptotic and bootstrap test as given in the legend.

Evaluation of linear trend tests using resampling techniques

Vafeiadis Thanasis [1], Bora-Senta Efthimia [1], Kugiumtzis Dimitris [2]

[1] Division of Statistics and Operation Research, Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece

[2] Department of Mathematical, Physical and Computational Sciences, Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki, 54124 Greece

Abstract

A number of parametric and non-parametric linear trend tests for time series are evaluated in terms of test size and power, using also resampling techniques to form the empirical distribution of the test statistics under the null hypothesis of no linear trend. For resampling, both bootstrap and surrogate data are considered. Monte Carlo simulations were done for several types of residuals (uncorrelated and correlated with normal and non-normal distributions) and a range of small magnitudes of the trend coefficient. In particular for AR(1) and ARMA(1,1) residual processes, we investigate the discrimination of strong autocorrelation from linear trend with respect to the sample size. The correct test size is obtained for larger data sizes as autocorrelation increases and only when a randomization test that accounts for autocorrelation is used. The overall results show that the type I and II errors of the trend tests are reduced with the use of resampled data. Following the guidelines suggested by the simulation results, we could find significant linear trend in the data of land air temperature and sea surface temperature.

Key words: time series, linear trend tests, resampling techniques, surrogate data, bootstrap.

1 Introduction

The investigation of long-term trends in times series is an important issue in many applications. Long-term trends can be considered as stochastic trends attributed to power law autocorrelation decay, referred to as long term persistence (see e.g. Rybski et al, 2006), or as deterministic trends, which will be the focus of this work. The formal statistical approach for the latter is a test for the presence of a linear trend in the time series. Such tests have been used in many areas of climatology, such as global warming (Woodward and Gray, 1993; Cohn and Lins, 2005), in meteorology, such as rainfall (Bonaccorso et al, 2005) and temperature (Xu et al, 2002; Feidas et al, 2004), and in hydrology, such as stream flow (Wang et al, 2005; Yue et al 2002).

The standard decomposition of a time series Y_t, $t = 1,\ldots,n$, under the assumption of a linear trend reads

$$Y_t = a + bt + E_t, \qquad (1)$$

where a is a constant, b represents the magnitude of the trend and E_t is the residual. The null hypothesis for the trend test is $\mathrm{H}_0 : b = 0$. Rejection of H_0 establishes the presence of linear trend in the time series, provided that the model for the residuals is valid.

Many trend tests assume independent residuals, such as the rank-based non-parametric Mann-Kendall (MK) test (Mann, 1945; Kendall, 1975) and the parametric regression-based test (Woodward et al, 1993). For the latter, the statistic is simply the estimated trend coefficient $\hat{b}$ standardized with its standard error. When the residuals are short-term correlated, simple corrections in the estimation of the standard error of $\hat{b}$ are suggested making use of the autocovariance (Grenander, 1954) and the spectrum of E_t (Bloomfield

and Nychka, 1992). More involved schemes adjusting the solution for the trend coefficient in the presence of autocorrelation in the time series have been proposed in (Sun and Pantula, 1999; Roy et al, 2004).

Strong positive autocorrelation in the time series may form monotonic trend and give rise for false rejection for the trend test. This has been shown with Monte Carlo simulations and different estimators for the linear trend (Woodward and Gray, 1993; Sun and Pantula, 1999; Roy et al, 2004; Kim et al, 2003). On the other hand, a deterministic trend may alert the sample autocorrelation used in the statistic of the trend test (Fried and Imhoff, 2003).

Besides the correlation in the residuals, the distribution of the residuals, as well as the sample size, affect heavily the outcome of the test, depending also on the magnitude of the trend. In this work, we address all these factors for four standard tests. Moreover, we introduce randomization and bootstrap versions of the tests in an attempt to improve the performance of the tests. Particular emphasis is given on the limits of sample size that maintain small type I and II errors. The investigation is done using Monte Carlo simulations at different settings of time series length, magnitude of trend coefficient, as well as distribution and linear structure of the residual process. We considered also a real application and applied the tests to a time series of an index that combines land air temperature anomalies (Jones, 1994a) and sea surface temperature anomalies (Parker et al, 1995) on a 5o x 5o grid box basis, developed by the Climatic Research Unit (CRU) of University of East Anglia (http://www.cru.uea.ac.uk/).

The trend tests are presented in Section 2 and the simulation setup and results in Section 3. In Section 4, the application is presented and in Section 5 the conclusions are drawn.

2 Statistical testing of linear trend

In general, it is difficult to detect the small linear trend with eyeball judgment, and there is need for an accurate and sensitive trend test in order to assess the significance of such weak

linear trends that are often investigated in small time series, as for example in meteorology (Feidas et al, 2004). In the following, we briefly present four standard linear trend tests, three parametric and one non-parametric test.

2.1 Parametric trend tests

Under the assumption of independent and normally distributed residuals E_t with zero mean and variance σ^2, $E_t \sim N(0,\sigma^2)$, from the regression of Y_t on time t, the least square estimator $\hat{b}$ is obtained as

$$\hat{b} = \frac{\sum_{t=1}^{n}(t-\bar{t})Y_t}{\sum_{t=1}^{n}(t-\bar{t})^2}, \qquad (2)$$

where $\bar{t}$ is the mean time.

The estimated standard error of $\hat{b}$ is given by

$$\hat{s}_1(\hat{b}) = \left[\frac{\sum_{t=1}^{n}(Y_t-\hat{a}-\hat{b}t)^2}{(n-2)\sum_{t=1}^{n}(t-\bar{t})^2}\right]^{1/2} = \left[\frac{12\sum_{t=1}^{n}(Y_t-\hat{a}-\hat{b}t)^2}{(n-2)n(n^2-1)}\right]^{1/2}, \qquad (3)$$

where $\hat{a} = \bar{Y} - \hat{b}\bar{t}$ and $\bar{Y}$ is the mean of the time series. Then the test statistic referred to as C1 is $t = \frac{\hat{b}}{\hat{s}_1(\hat{b})}$ and follows the Student distribution with n-2 degrees of freedom $t \sim t_{n-2}$ (Woodward and Gray, 1993).

When the residuals E_t are correlated, the estimated standard error of $\hat{b}$ is given by

$$\hat{s}_2(\hat{b}) = \left\{\frac{12}{n(n^2-1)}\left[\gamma_0 + \frac{24}{n(n^2-1)}\sum_{s=2}^{n}\sum_{t=1}^{s-1}(t-\bar{t})(s-\bar{t})\gamma_{s-t}\right]\right\}^{1/2}, \qquad (4)$$

where γ_k denotes the k-th order autocovariance of E_t (Grenander, 1954). Replacing in (4) γ_k with the respective estimate

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k} \hat{E}_{t+k}\hat{E}_t, \quad (5)$$

where E_t can be estimated by $\hat{E}_t = Y_t - \hat{a} - \hat{b}t$, except at $k=0$ where we use $n\hat{\gamma}_0 / n-2$ to estimate γ_0, the estimated standard error of $\hat{b}$, $\hat{s}_2(\hat{b})$, is derived. This is used to form the test statistic $t = \dfrac{\hat{b}}{\hat{s}_2(\hat{b})}$ referred to as C2 and it holds as before $t \sim t_{n-2}$.

In a different approach, the standard error of $\hat{b}$ is estimated from the power spectrum (Bloomfield and Nychka, 1992)

$$\hat{s}_3(\hat{b}) = \left[2\int_0^{0.5} W(f)S(f)df \right]^{1/2}, \quad (6)$$

where $W(f) = \left| \sum_{t=1}^{n} m_t e^{-2\pi i f t} \right|^2$ with $m_t = \dfrac{t-\bar{t}}{\sum_{t=1}^{n}(t-\bar{t})^2}$ and $S(f)$ denotes the sample spectrum of E_t given as

$$S(f_j) = \left(\frac{1}{2\pi}\right)\left(\hat{\gamma}_0 + 2\sum_{k=1}^{n-1}\hat{\gamma}_k \cos(f_j k)\right) \text{ where } f_j = \frac{2\pi j}{n}, j = 0,...,n/2.$$

The test with statistic $t = \dfrac{\hat{b}}{\hat{s}_3(\hat{b})} \sim t_{n-2}$ is denoted as C3.

2.2 Non-Parametric trend tests

Two non-parametric rank-based statistical tests, namely the Mann-Kendall (MK) test, also called Kendall's tau test due to Mann (1945) and Kendall (1975), and the Spearman's rho test (Lehmann, 1975; Sneyers, 1990) are used for detecting trend in time series data. Yue et al (2002) showed that these two tests have almost the same power to identify trends in time series data. In our study, we use the rank-based non-parametric Mann-Kendall (MK) test,

which seems to have been used more often in applications, such as stream flow (Yue et al, 2002; Yue and Pilon, 2004, Wang et al, 2005).

The null hypothesis for the MK test is that the time series Y_t, $t = 1, 2, ..., n$, is independent and identically distributed. The statistic S of Kendall's tau is

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}(Y_j - Y_i), \quad (7)$$

where

$$\text{sgn}(\theta) = \begin{cases} 1, & if\ \theta > 0 \\ 0, & if\ \theta = 0 \\ -1, & if\ \theta < 0 \end{cases}.$$

Mann (1945) and Kendall (1975) documented that when $n \geq 8$ the statistic S is approximately normally distributed with the mean and the variance as follows:

$$\text{E}(S) = 0$$

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{m=1}^{n} t_m m(m-1)(2m+5)}{18},$$

where t_m is the number of ties of extent m.

2.3 Randomization and bootstrap tests

The tests described above have all well defined asymptotic null distribution, i.e. distribution of the test statistic under H_0. However, departures from the nominal null distribution may occur, e.g. due to small sample size. Resampling techniques have been used to form the null distribution. Here we consider randomization and bootstrap tests.

A randomization test generates a randomly chosen subset of all possible permutations of the original sample consistent to H_0 (Fortin et al, 2002). The randomization tests used in this

study are adjusted to the correlation structure of the time series and the distribution of residuals E_t.

When the residuals are white noise, $E_t \sim WN(0,\sigma^2)$, the randomized or so-called surrogate data are generated by shuffling the Y_t time series. In case where E_t stem from a stochastic linear process, different surrogate data generating algorithms are called depending on whether the process is normal or not. When it is normal, the surrogate time series are generated by phase randomization making use of the Fourier transform and are referred to as FT surrogate data (Theiler et al, 1992). An FT surrogate time series is a normal time series with the same linear structure as Y_t, but contains no trend. In case E_t comes from a non-normal stochastic linear process, the more general algorithms of Improved Amplitude Adjusted Fourier Transform, IAAFT (Schreiber and Schmitz, 1996) and Statically Transformed Autoregressive Process, STAP (Kugiumtzis, 2002) are called. The IAAFT algorithm makes also use of the Fourier transform but in an iterative scheme that terminates when sufficient convergence of both power spectrum and marginal distribution is reached. The STAP algorithm generates the surrogate time series as statically transformed realizations of a normal (autoregressive) process so that both the original marginal distribution and linear structure are preserved. These algorithms were introduced to test nonlinear departures from the null hypothesis of linear stochastic process, but they can as well be used to test departures involving linear trend. The test using STAP surrogate data is more conservative than when using IAAFT surrogates and its power decreases faster with the decrease of the time series length (Kugiumtzis 2002). We employ the randomization test with both surrogate data types and compare their size and power on small sample sizes, typically encountered in trend investigation.

In addition to randomization tests we include bootstrap tests, adapted for each model assumption for E_t (Hinkley, 1988; Efron and Tibshirani, 1993). For white noise residuals, the standard bootstrap resampling is applied. When E_t is a realization of a linear stochastic process, there are a number of bootstrap approaches, such as the block, sieve, wild and local bootstraps, but we follow here the most standard "naïve" bootstrap approach, fitting an autoregressive model and drawing from the model residuals to generate the bootstrap time series (Buhlmann, 2002; Politis 2003).

3 Monte Carlo Simulations

3.1 Simulation setup

We generate Monte Carlo realizations for different stochastic processes with and without linear trend according to the model in (1). The length of the time series n varies as 2^k for $k = 4,5,6,7$ and the trend magnitude is monitored varying the linear trend coefficient as $b = -0.01(0.002)0.01$, where the no-trend scenario is for $b = 0$. For white noise residuals E_t, the normal, uniform and exponential distributions are considered. For correlated residuals E_t, we consider the first order autoregressive process AR(1), $E_t = \varphi E_{t-1} + a_t$, and the mixed process of first order autoregressive part and first order moving average part ARMA(1,1) $E_t = \varphi E_{t-1} - \theta a_{t-1} + a_t$, where a_t follows normal, uniform and exponential distribution. In order to examine how the correlation in the residuals affects the detection of linear trend, we vary also the parameter φ (equal to one lag autocorrelation of residuals) as $\pm 0.95, \pm 0.8$ and ± 0.4. The combination of all the values of n, b, and φ (including zero) and the distribution types of input noise, gives a total of 4x11x7x3=924 cases. For ARMA, the study is not exhaustive and is restricted to selected values of φ and θ.

For each case, 1000 Monte Carlo realizations are generated and for each realization 199 surrogate and bootstrap data are generated by the appropriate algorithm. For the correlated residuals, both IAAFT and STAP surrogate data are generated along with the bootstrap data. The four tests are applied on each time series and the test decision is made on the basis of the analytic null distribution of the test statistic q. In addition, the null distribution is formed from the values of q computed on the resampled data, denoted as $q^1,...,q^{199}$, and the rejection of H_0 is deduced when q^0 computed on the original time series is not within the null distribution. In previous works on surrogate data, the rejection of H_0 is often determined from the significance S (provided that $q^1,...,q^M$, on the M surrogate data, are fairly normally distributed) denoted as

$$S = \frac{|q^0 - \bar{q}|}{s_q}, \qquad (8)$$

where $\bar{q}$ is the average and s_q the standard deviation (SD) of $q^1,...,q^M$ (Kugiumtzis, 2000). Rank ordering has also been used, where for our case, H_0 is rejected, say, at significance level $\alpha = 0.01$ when q^0 is first or last in the ordered concatenated list $q^0, q^1,...,q^{199}$ and at $\alpha = 0.05$ when q^0 is at places 1 to 5 or 196 to 200.

3.2 Simulation results

All linear trend tests are performed using the test statistics C1, C2, C3 and MK and the test decision is made using the asymptotic approach and the randomization and bootstrap approach. The probability of rejecting H_0 is estimated by the relative frequency of rejections in the ensemble of 1000 time series.

The significance S in (8) gives better resolution in the p-value than the rank ordering. However, normality tests on the statistics from a sample of 199 surrogates showed departures

from normality. Therefore, the surrogate and bootstrap test results below are for rank ordering.

One would expect that all the tests perform well in the case of independent residuals. However, even when the residual series is normal white noise, the asymptotic test shows larger size when using C2 and small power when using C3, given with highlighted values in Table 1. For C2 this is a significant drawback that persists for other distributions of E_t (e.g. uniform noise in Table 1) and questions the detection of trend with this method (e.g. note the higher probability of rejection for b=±0.002 as compared to C1 and MK). The shortcomings of C2 and C3 are recovered with the use of the randomization approach. Indeed randomization tests attain always the correct size of the test and the same level of power as the asymptotic approach. The results on non-zero trend coefficients suggest that C1 and MK, constructed under the assumption of independent residuals, have somehow larger power than the C2 and C3 statistics for any white noise distribution. The distribution of E_t seems to affect the significance of the linear trend, e.g. the power of all tests is increased when the distribution changes from normal to uniform (see Table 1). The results of the bootstrap tests are the same as for the respective randomization tests.

(Here should be placed Table 1)

For correlated residuals, the degree of correlation, monitored in the simulations with the coefficient φ of the AR(1) model for E_t, in combination with the time series length have major effect on the size and power of all tests. On the other hand, the distribution of E_t (actually we determine the distribution of the input noise a_t of AR(1) in the simulations) does not seem to have significant effect on the test accuracy.

The simulations showed that when consecutive residuals are anti-correlated (φ negative in AR(1)) the null distribution of the t statistic of C2 tends to be wider than the respective nominal distribution, whereas for C1, C3 and MK tests it is narrower, as shown in Fig. 1. For

example, in the absence of trend and for $n = 128$, the estimated variance of the C2 statistic when $\varphi = -0.8$ is 1.74 and when $\varphi = -0.2$ is 2.8, which are both far from the nominal unit variance. For the other tests, the estimated variance is much smaller than the nominal unit variance. This is observed for all three types of noise.

(Here should be placed Figure 1)

Thus the asymptotic approach tends to give larger test size for C2 test and smaller power for the other statistics. This is shown in Fig. 2a and 2d for AR(1) residuals with $\varphi = -0.8$ and $\varphi = -0.4$, respectively, where the data size is n=128 and a_t follows normal distribution. The power of all tests increases with the decrease of anticorrelation (φ closer to zero) and the increase of the magnitude of b, similarly for upward and downward trend. C2 has the largest power but spuriously given the large test size for b=0, and that is because $\hat{s}_2(\hat{b})$ is a poorly behaved estimator (Woodward et al, 1993). The other three test statistics perform similarly having insignificant power for small b (see Fig. 2a). The respective randomization tests using FT surrogates give better results: they eliminate the type I error of the asymptotic tests for C2, with a loss of power (see Fig. 2b and 2e). The randomization tests using C2 and C3 tend to have more symmetric increase of power than C1 and MK (for positive and negative b). When bootstrap data are used, the power of all tests is further improved and all four tests perform similarly, as shown in Fig. 2c and 2f.

(Here should be placed Figure 2)

Similar results are obtained when E_t is an AR(1) process with uniform or exponential input white noise. The test results for the two noise distributions are shown in Fig. 3, for $\varphi = -0.4$. Note that C2 and C3 attain larger power when resampling techniques are used, especially when the input noise is uniform.

(Here should be placed Figure 3)

STAP gives similar results to IAAFT for the randomization test in these non-normal AR(1) processes. Both randomization and bootstrap tests eliminate type I error for all statistics, but bootstrap tests obtain somehow larger power than the randomization tests.

The actual null distribution of the test statistics deviates from the nominal null distribution also when φ in AR(1) residuals is positive but in a different way and at a larger degree (compare Fig. 4 to Fig. 1). For example, for $b=0$, normal white noise and $n=128$, as φ coefficient increases from 0.2 to 0.8, the variance jumps from 1.49 to 9.44 for C1, from 3.15 to 4.4 for C2, from 0.29 to 0.68 for C3 from 4.37 to 7.67 for MK test. This explains the very large test size we found when using C1, C2 and MK with the asymptotic approach. All test statistics, except for C3, have much wider empirical distribution as shown in Fig. 4 for normal and uniform noise.

(Here should be placed Figure 4)

The Monte Carlo simulations showed that the empirical test size gets larger for the asymptotic tests as φ in AR(1) residuals increases away from 0.4. The same problem was found also for the tests using resampling techniques but at a lesser amount. Among all test statistics, only C3 performs properly for large positive autocorrelation. In Fig. 5a, the test results using C3 are shown for $\varphi=0.8$, $n=128$ and the uniform input noise. There is still a small type I error, especially for IAAFT and STAP randomization tests. On the other hand, the bootstrap test eliminates the type I error at the cost of smaller power compared to IAAFT and STAP. For exponential input noise all tests do not have any significant power (see Fig. 5b) and the same holds for normal input noise (not shown here).

(Here should be placed Figure 5)

The power increases fast for larger data sets, as shown in Fig. 6 for C3, bootstrap approach and $n = 256$, i.e. double than the sample size in Fig. 5. For strong positive correlations as for $\varphi = 0.8$ used in Fig. 5 and Fig. 6, the increase of the power with the sample size is lower for normal and exponential input white noise.

(Here should be placed Figure 6)

The test results when the residuals are from an ARMA(1,1) turn out to be similar to the results shown above for AR(1) residuals, at least for the corresponding values of φ that we tested for. For examples, as shown in Fig. 7a the performance of the bootstrap tests with C2 and C3 for ARMA(1,1) residuals with $\varphi = -0.8$, $\theta = 0.8$ and normal input noise are substantially the same as the respective results for AR(1) shown in Fig.2c, whereas C1 and MK show less power for ARMA(1,1). This difference with C1 and MK gets larger when the randomization test is used instead (compare Fig. 7b with Fig. 2b). For all statistics the test for the same ARMA residual process improves both in terms of significance and power when the input noise is uniform, as we observed for the AR process (see Fig. 7c). Other values of φ and θ gave results similar to the corresponding AR(1) residual process. For example the results for $\varphi = -0.4$ show the same test performance for C2, C3 and difference for C1,MK for the ARMA and AR case as discussed earlier for $\varphi = -0.8$ (compare Fig. 7d to Fig. 2f). For positive values of φ in the ARMA(1,1) residual process, the power of the test (bootstrap and randomization) decreases in the same way as for the AR(1) residual process.

(Here should be placed Figure 7)

It is of practical interest to investigate the dependence of sample size on the strength of positive autocorrelation (positive φ) under the condition of maintaining the correct test size. For this, we made Monte Carlo simulations for AR(1) residual processes with

$\varphi = 0.4(0.05)0.95,0.97,0.99$ and we found the smallest n that preserves the actual size of the test within the limit of 0.06 for $a = 0.05$. The results for C3, bootstrap approach and the three distributions of input white noise are shown in Fig. 8. It is clear that the demand on more data points increases faster as φ approaches 1, i.e. the random walk scenario that regards fully stochastic trend. It turns out that for uniform input white noise the correct application of the test (using C3 and bootstrap approach) does not require as long time series as for normal and exponential white noise.

(Here should be placed Figure 8)

According to the simulation results, C3 with the use of bootstrap test is the most suitable trend test to identify the presence of small linear trend in time series data under varying conditions of autocorrelation and amplitude distribution of the time series.

4 Application

We applied the asymptotic and resampling tests to time series of an index that combines land air temperature anomalies (Jones, 1994a) and sea surface temperature anomalies (Parker et al, 1995) on a 5o x 5o grid box basis, developed by the Climatic Research Unit (CRU) of University of East Anglia (http://www.cru.uea.ac.uk/).. We consider the time series for each month from January to December in the period from 1856 to 2005 ($n = 150$, the whole sample), referred to as period 1, and from 1961 to 2005 ($n = 45$), referred to as period 2. In addition, the records of the mean annual values for the two periods are analyzed. These time series show weak linear trends and we want to investigate whether these trends are significant. For example, as shown in Fig. 9, the index of period 2 for January shows a long steep upward trend starting at around 1970 suggesting significance of the trend, whereas for

July the trend can be seen in a smaller part of the same period and is thus of questionable significance.

(Here should be placed Figure 9)

According to Akaike (AIC) and Swartz (BIC) criteria the order of the AR model of the residuals, for the most time series was 1 (and mostly for $n = 150$). The estimated coefficient of AR(1) was about 0.5 for most of the time series, but the distribution of the residuals did not appear to have the same form across the time series. For example, the Kolmogorov – Smirnov test for normality gave rejection for most of the time series. According to our simulation results for correlated residuals at the order of $\varphi \approx 0.5$, the minimum sample size for the appropriate use of the trend test is at 100 to 150 (see Fig. 8). In table 2, the standardized coefficients (s.c.) for the magnitude of the trend for all the time series are shown. The s.c. for period 2 are larger than for period 1 for all months except November (11).

(Here should be placed Table 2)

All asymptotic tests give significant linear trend for all months for period 2 and only for autumn and winter months for period 1, as shown in Fig. 10 using C3. This result cannot be trusted due to the presence of positively correlated residuals and according to the simulation results for sample sizes at the level of period 1 and 2. On the other hand, when the bootstrap and randomization tests were used, significant linear trend was found only for the winter and partly spring months. As shown for C3 and bootstrap in Fig. 10, significant trend at $\alpha = 0.05$ was found for months September to May for period 1 and for months January to April for period 2. However, the test results for period 2 should be treated with caution as for such a small sample size the presence of positive autocorrelation in the residuals (here it is about 0.5) may be the cause of the statistically significant trend (see also Fig. 8). Note in particular that the linear trend for January of period 2 was found significant with all approaches,

whereas for July only the asymptotic approach found significant trend (see also Fig. 10). According to the simulation results, we should thus trust the bootstrap test for July of period 2.

(Here should be placed Figure 10)

As for the mean annual time series (at point 13 of the horizontal axis of Fig. 10), C3 asymptotic and bootstrap tests give significant linear trend for period 1, whereas for period 2 the linear trend was found significant for the asymptotic but not for the bootstrap test.

The overall results from the test of C3 and bootstrap approach suggest that there is a trend during the winter and spring months, better expressed in the long record (1856 – 2005).

5 Conclusion

Monte Carlo simulations were made on four test statistics for asymptotic, randomization and bootstrap test of linear trend under different settings of time series length, residual distribution and autocorrelation. The comparative results showed clear superiority of the randomization and bootstrap test over the asymptotic test and revealed differences and limitations in the performance of the test statistics.

For correlated residuals, the C3 test statistic, using spectrum-based estimation of the variance of the slope coefficient, gives the smallest size of the asymptotic test and when resampling techniques are used the test size decreases to the nominal level. Further, it attains high power compared to the other test statistics. However, when the residuals are white noise, the power of the test using C3 is smaller than when using a test statistic formed under the assumption of white noise residual.

The asymptotic test gives generally large type I error and the use of resampling techniques recovers the correct test size in most of the settings considered in the study. For correlated residuals, suitable surrogate data generation techniques have been used for the randomization

test and the residual-based bootstrap for the bootstrap test. The simulation results showed that the bootstrap test turns out to attain higher power than the randomization test.

The overall simulation results suggest the use of the C3 statistic in a bootstrap test. Even this test cannot distinguish linear trend from strong positive autocorrelation depending on the time series length. We found that under the condition of retaining the correct test size, the time series length has a functional dependence on the positive autocorrelation (for values larger than about 0.4) that varies with the input white noise distribution. These functional relations can serve as a guide for the limits of implementation of the test in real-world applications.

We applied the asymptotic and resampling tests with the four test statistics to time series of an index of land air and sea surface temperature anomalies at different periods, for all 12 months separately and for the annual average. For some months, a linear trend was found for some statistics using the asymptotic test (and sometimes even the resampling test) whereas it was not found when using C3 and the bootstrap test, indicating spurious detection of trend. However, consistent detection of trend could be obtained in the winter and spring months, especially when considering the whole record that allows for a proper implementation of the test, given also the relatively small positive autocorrelation of the residuals.

We believe that this work shed some light on the performance of standard tests for linear trend and showed the need of resampling techniques in the implementation of the tests. There are other tests for linear trend not considered in this work and it would be interesting to include them in a future comparative work.

References

Bloomfield P. and Nychka D. (1992). Climate Spectra and Detecting Climate Change. Climatic Change 21:275-287

Bonaccorso B., Cancelliere A. and Rossi G. (2005). Detecting trends of extreme rainfall series in Sicily. Advance in Geosciences 2:7-11

Bowerman - O' Connell (2000). *Forecasting and Time Series. An applied approach*. Third Edition.Duxbury Press.

Brockwell J. P., Richard A. D. (2002). *Introduction to Time Series and Forecasting*. Second Edition. Springer.

Buhlmann P. (2002). Bootstrap for time series. Statistical Science 17:52-72

Chandler R. (2002). Trend Analysis For The Environmental Science-A Review, ESSG Meeting, March 2002

Cohn A.T. and Lins F.H. (2005). Nature's style: Naturally trendy. Geophysical Research Letters 32:L23402

Cryer D. J. (1986) *Time Series Analysis*. PWS-KENT Publishing Company- Boston.

Efron B., Tibishirani J. R.(1993). *An Introduction to the Bootstrap.* Chapman and Hall/CRC

Feidas H., Makrogiannis T. and Bora-Senta E. (2004). Trend analysis of Air Temperature Time Series Data In Greece Determined By Ground and Satellite Data. Theoretical and Applied Climatology 79:185 - 208

Folland, C.K., Rayner, N.A., Brown, S.J., Smith, T.M., Shen, S.S.P., Parker, D.E., Macadam, I., Jones, P.D., Jones, R.N., Nicholls, N. and Sexton, D.M.H., (2001). Global temperature change and its uncertainties since 1861. Geophysical Research Letters 28:2621-2624.

Fortin M.-J., Jasquez G. & Shipley B. (2002). Computer-intensive methods. Encyclopedia of Environmetrics 1:399-402

Fried R. and Imhoff M. (2004). On the online detection of monotonic trends in time series. Biometrical Journal 46:90-102

Grenander U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. The Annals of Mathematical Statistics 25:252-272

Hinkley V. D. (1988). Bootstrap Methods. Journal of Royal Statistical Society. Series B (Methodological) 50:321-337

Jones P. D. (1994). Hemispheric surface air temperature variations: a reanalysis and an update to 1993. Journal of Climate 71794-1802

Kim T. H., Pfaffenzeller S., Rayner T. and Newbold P. (2003). Testing for linear trend with application to relative primary commodity prices. Journal of Time Series Analysis 24:539-551

Kugiumtzis D. (2000). Surrogate Data Test on Time Series. In: A. Sool, L. Cao, *Nonlinear Deterministic Modeling and Forecasting of Economics and Financial Time Series,* Kluwer Academic Publishers

Kugiumtzis D. (2002). Statically transformed autoregressive process and surrogate data test for nonlinearity. Physical Review 66:025201(R)

Mann B. H. (1945). Nonparametric Tests Against Trend. Econometrica 13:245-259 (No.3)

Nordgaad A., Grimvall A. (2006). A resampling technique for estimating the power of non-parametric trend tests. Environmetrics, 17:257-267

NIST/SEMATECH e-Handbook of Statistical Methods. www.nist.gov/stat.handbook

Parker DE, Folland CK, Jackson M (1995). Marine surface temperature observed variations and data requirements. Climatic Change 31:559-600

Politis, D. N. (2003). The Impact of Bootstrap Methods on Time Series Analysis. Statistical Science 18:219-230

Roy A., Falk B. and Fuller A. W (2004). Testing for trend in the presence of autoregressive error. Journal of American Statistical Association 99:1082-1091

Rybski D., Bunde A., Havlin S., and Storch H. (2006). Long-term persistence in climate and the detection problem. Geophysical Research Letters 33L06718

Schreiber T. and Schmitz A (1996). Improved Surrogate Data for Nonlinearity Tests. Physical Review Letters 77:635- 638

Sun H. and Pantula G. S. (1999). Testing for trend in correlated data. Statistics & Probability Letters 41:87-95

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. (1992).Testing for nonlinearity in time series: The method of surrogate data. Physica 58:77-94.

Vandaele W. (1983). *Applied Time Series and Box-Jenkins Models*. INC, Harcourt Brace and Company. Academic Press

Wang W., Van Gelder P.H.A.J.M. and Vrijling J.K. (2005). Trend and stationary analysis for streamflow processes of rivers in Western Europe in the 20th century. IWA International Conference on Water Economics, Statistics and Finance, Rethymno, Greece, 8-10 July 2005

Woodward A. Wayne and H. L. Gray (1993). Global Warming and the Problem of Testing for Trend in Time Series Data . American Meteorological Society 6:953-962

Woodward A. Wayne, Bottone Steven and H. L. Gray (1997). Improved Tests for Trend in Time Series Data. Journal of Agricultural, Biological and Environmental Statistics 2:403-416

Xu. X. Z, Takeuchi K. and Ishidaira H. (2002). Long term trends of annual temperature and precipitation time series in Japan. Journal of Hydroscience and Hydraulic Engineering 20:11-26

Yue S., Pilon P. (2004). A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. Journal of Hydrological Science 49:21-37

Yue S., Pilon P. and Cavadias G. (2002). Power of Mann-Kendall and Spearman's rho tests for detecting monotonic trend in hydrological series. Journal of Hydrology 259:254-271

Yue S., Pilon P. Phinney B. and Cavadias G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes 16:1807-1829

Tables and Figures

Table 1.

		E_t ~ N(0,1)				E_t ~ U[-1/2,1/2]			
b	**Test (α=0.05)**	**C1**	**C2**	**C3**	**MK**	**C1**	**C2**	**C3**	**MK**
b=-0.004	Asymptotic	0.375	0.645	0.101	0.377	1.000	1.000	0.999	0.972
	Randomization	0.328	0.270	0.235	0.307	1.000	0.990	0.994	0.999
b=-0.002	Asymptotic	0.133	0.343	**0.029**	0.117	0.820	0.931	0.414	0.777
	Randomization	0.104	0.104	0.108	0.108	0.780	0.642	0.667	0.739
b=0.0	Asymptotic	0.061	**0.206**	0.013	0.057	0.048	**0.241**	0.002	0.044
	Randomization	0.049	0.046	0.061	0.049	0.048	0.053	0.051	0.050
b=0.002	Asymptotic	0.130	0.380	**0.031**	0.119	0.800	0.933	0.388	0.762
	Randomization	0.140	0.110	0.113	0.129	0.792	0.688	0.712	0.757
b=0.004	Asymptotic	0.401	0.655	0.099	0.377	1.000	1.000	1.000	0.961
	Randomization	0.389	0.328	0.199	0.386	1.000	0.991	0.995	1.000

Figure 1.

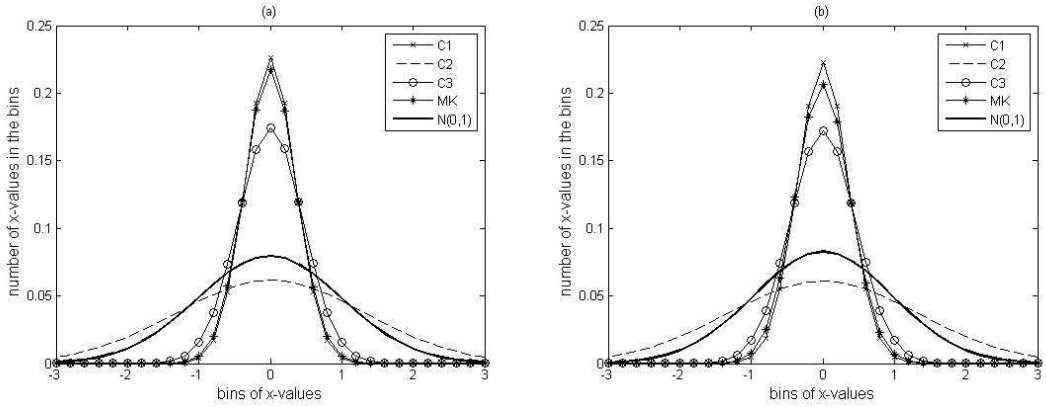


Figure 2.

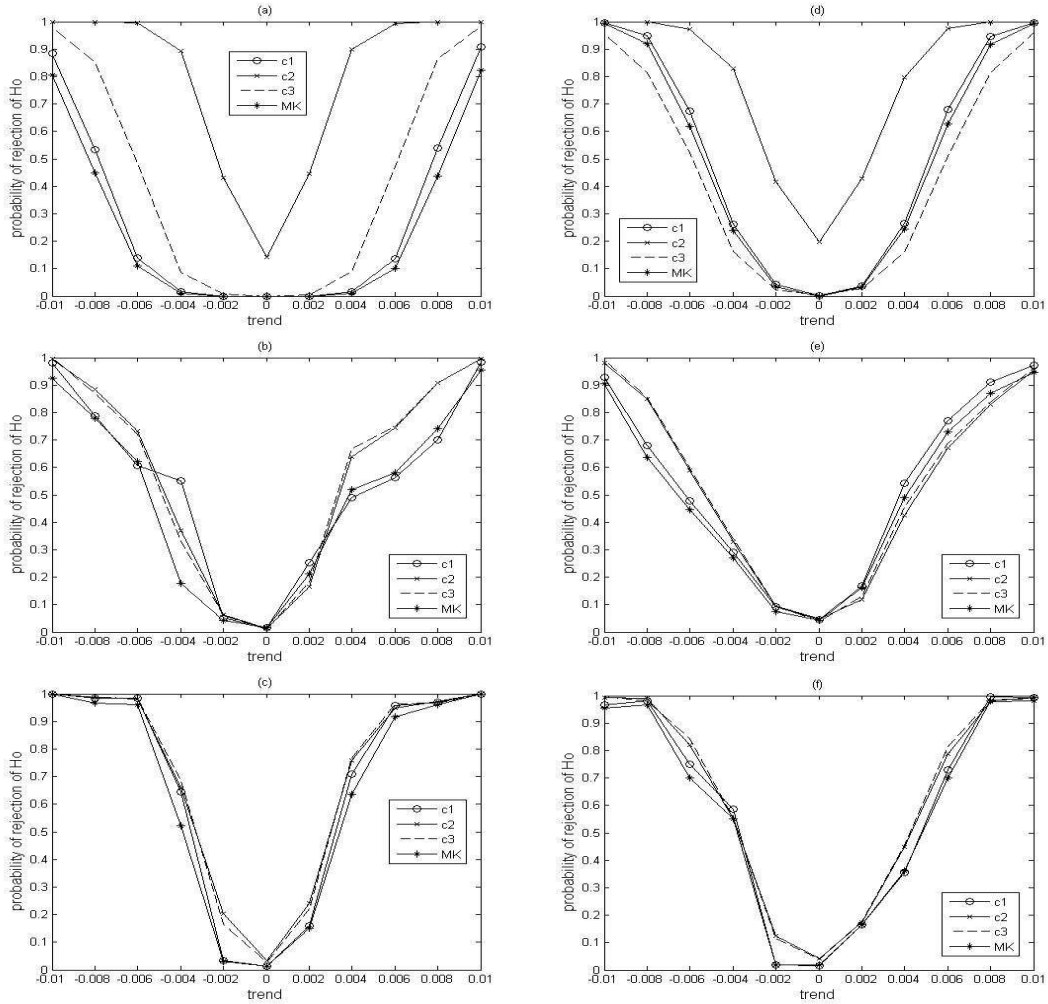


Figure 3.

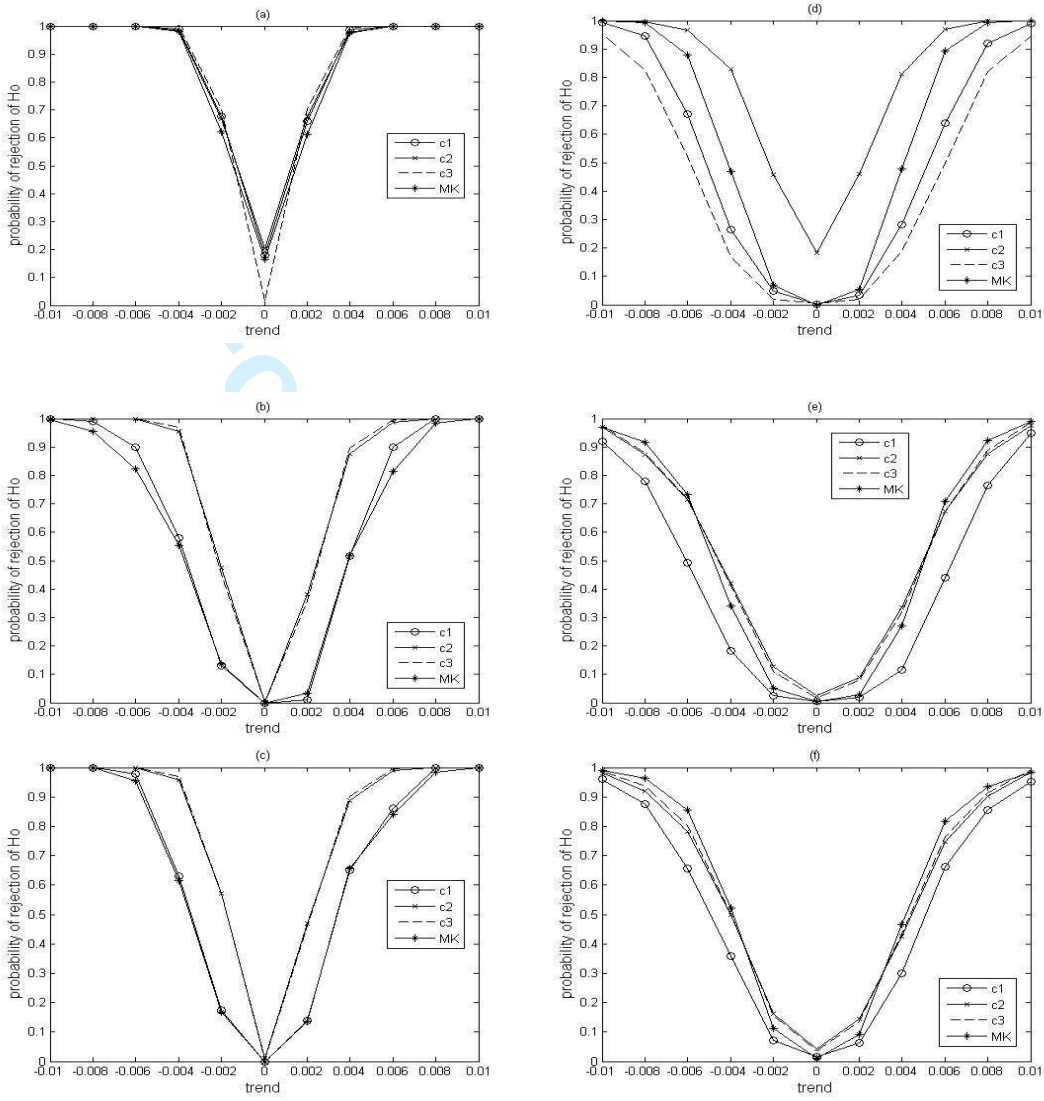


Figure 4.

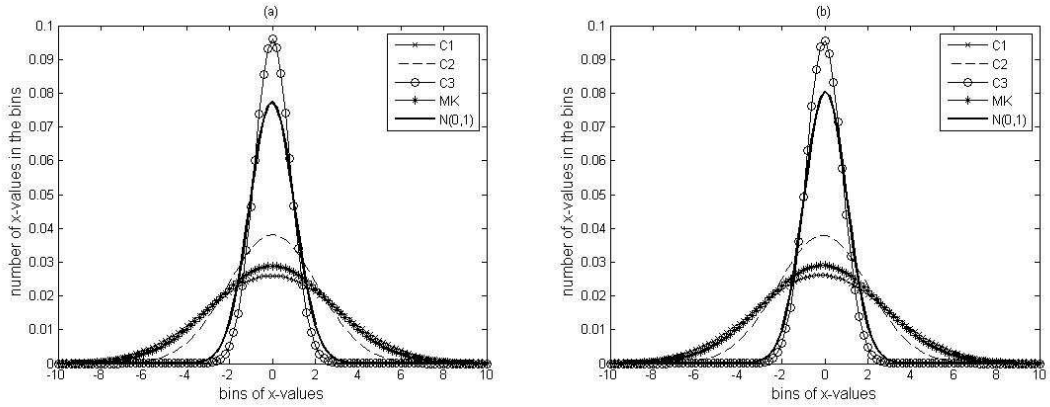


Figure 5.

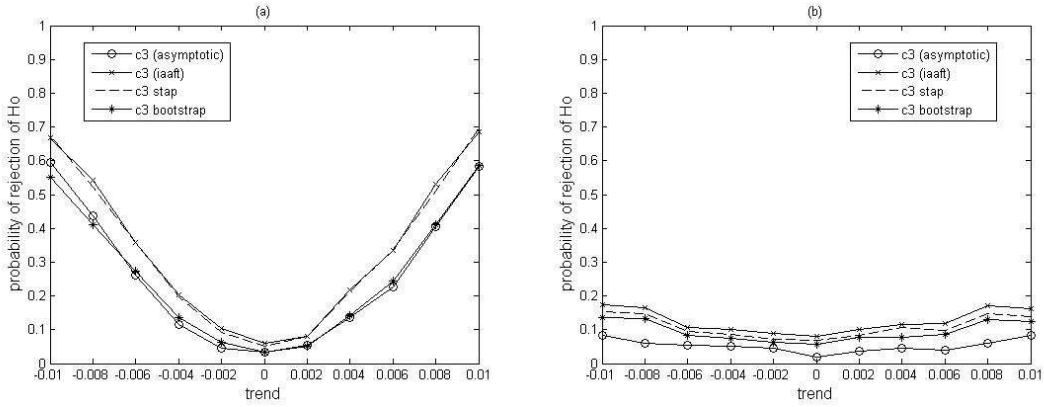


Figure 6.

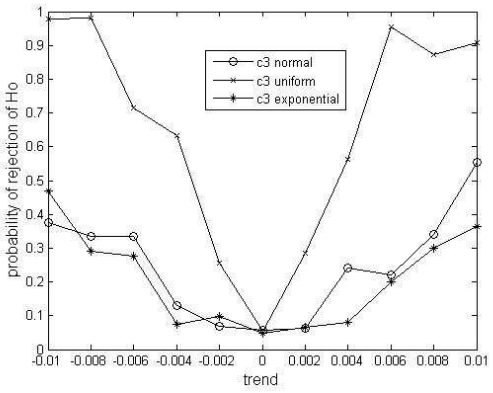


Figure 7.

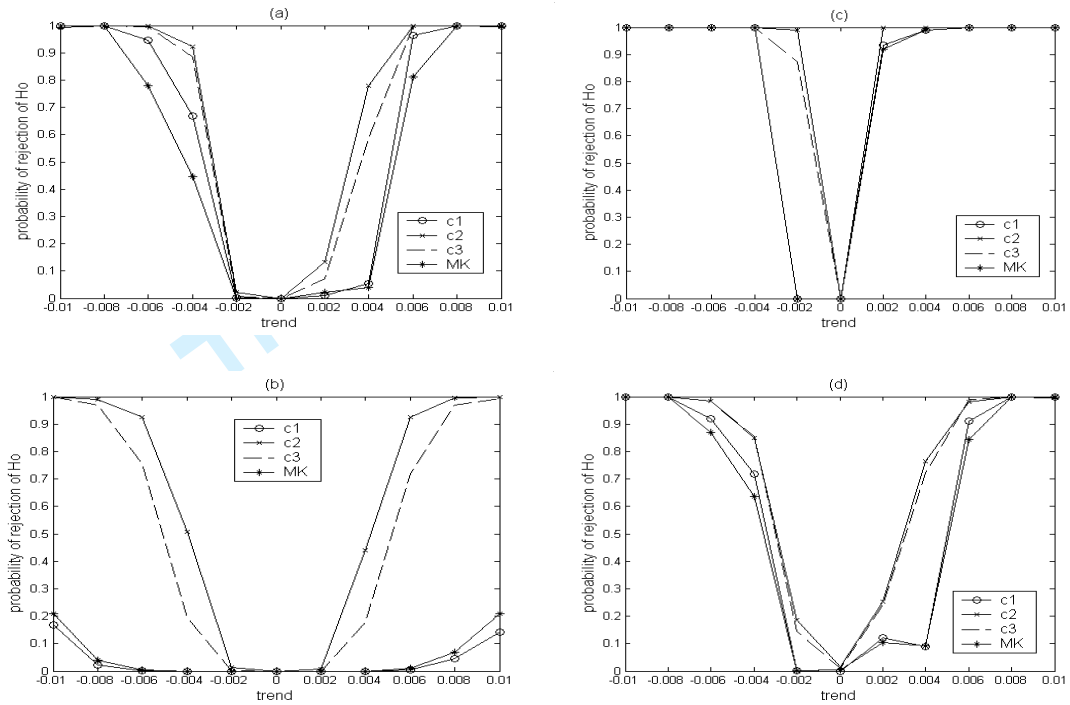


Figure 8.

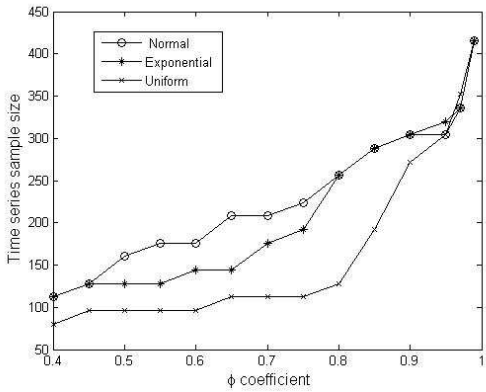


Figure 9.

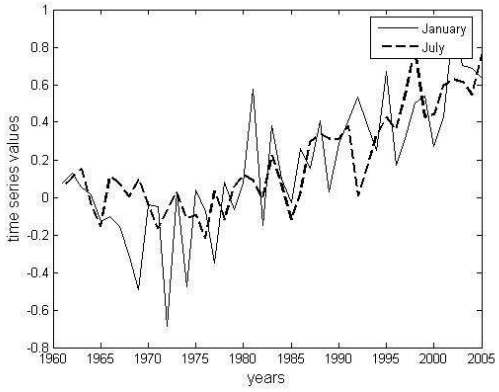


Table 2.

Period	Months												Annual
	1	2	3	4	5	6	7	8	9	10	11	12	
1	0.530	0.559	0.665	0.653	0.621	0.490	0.468	0.601	0.646	0.704	0.746	0.610	0.755
2	0.746	0.660	0.719	0.816	0.830	0.826	0.796	0.783	0.794	0.705	0.649	0.715	0.833

Figure 10.