Detecting Nonlinearity in Time Series: Surrogate and Bootstrap Approaches

Melvin J. Hinich[*]   Eduardo M. Mendes[†]

Lewi Stone[‡]

[*], hinich@mail.la.utexas.edu

[†]Federal University of Minas Gerais - Brazil, emmendes@cpdee.ufmg.br

[‡], lewi@lanina.tau.ac.il

Detecting Nonlinearity in Time Series:
Surrogate and Bootstrap Approaches[*]

Melvin J. Hinich, Eduardo M. Mendes, and Lewi Stone

Abstract

Detecting nonlinearity in financial time series is a key point when the main interest is to understand the generating process. One of the main tests for testing linearity in time series is the Hinich Bispectrum Nonlinearity Test (HINBIN). Although this test has been succesfully applied to a vast number of time series, further improvement in the size power of the test is possible. A new method that combines the bispectrum and the surrogate method and bootstrap is then presented for detecting nonlinearity, gaussianity and time reversibility. Simulated and real data examples are given to demonstrate the efficacy of the new tests.

---

1   Introduction

All economic systems are inherently stochastically nonlinear since preferences are assume to be convex and the economy is subject to various exogenous random shocks. The stochastic nonlinearity is usually lost in the highly aggregated time series that are generated by governmental agencies. Financial time series such as high frequency stock returns and currency exchange rates are nonlinear. M.J. Hinich and D.M. Patterson (1989) were the first financial data analysts to present evidence of nonlinearity in daily data from the New York Stock Exchange using the Hinich Bispectrum Nonlinearity Test (HINBIN). The original HINBIN, the related Hinich test of Gaussianity and the Hinich-Rothman test of time reversibility were based upon asymptotic properties of the normalized bispectrum.

Since the introduction Efron's boostrapping in the late seventies (B. Efron, 1979), much attention has been attracted in both theoretical and applied sides of Statistics. The bootstrap approach attempts to retrieve more information from sample data so as to solve problems that are not easily solved by some traditional methods. It is known that most test statistics do not have a known finite sample distribution. One either uses asymptotic theory to compute a critical value or some form of resampling known in statistics as bootstrapping. In practice, one cannot determine the validity of critical values determined by asymptotic theory since the rate of convergence of the central limit theorems used in the theory is a function of unknown parameters. Bootstrapping (also known as the resampling technique) is presented as a way out but standard bootstraps do not fit into time series problems since they are formulated on the assumption that time dependence can be ignored. In cases where there is linear dependence, alternative bootstrap methods are available in the literature. See, for instance, (R. M. Vogel and A. L. Shallcross, 1996; P. Hall et al., 1995; Hidalgo, 2003) for the moving blocks bootstrap. The surrogate method of Theiler et al. J. Theiler et al. (1992) is a type of bootstrapping that takes advantage of the statistical properties of Gaussian time series and has the potential to be a very useful tool since it is appropriate for data that is time-dependent.

The surrogate data method (J. Theiler et al., 1992) tests whether an observed time-series is consistent with the null hypothesis of a linear gaussian process (LGP). This is implemented by first generating a set of surrogate data sets whose spectra are identical with the observed time series and which are by construction LGP. It is then possible to test whether the observed time-series has statistical properties that are significantly different from the "random" surrogate data sets. If so, the LGP null hypothesis is rejected. This technique is recognized as a powerful method (D. Prichard and J. Theiler, 1994) and has formed the basis of a large number of studies with the aim of detecting nonlinearity in physical (J. Theiler and D. Pichard, 1997) and biological time series typically including ECG, EEG, neural, epidemiological and climate signals (Y.-J. Lee et al., 2001; F.X. Witkowski et al., 1995) as well as for detecting unstable fixed points (P. So et al., 1996).

Although the surrogate method has been widely used in the literature as pointed out before, it has been shown that the surrogate method has major drawbacks and can often fail to maintain reasonable significance levels when testing null hypothe-

ses based on even the simplest of test statistics (M. J Hinich et al., 2002). The LGP null hypothesis is particularly restrictive when used to test against the alternative hypothesis of nonlinearity. In the real world most linear processes are nearly always nongaussian. Hence for this important and large class of *linear* nongaussian processes, tests based on the surrogate method can routinely reject the linear null hypothesis even though the time-series is purely linear.

In this paper, alternative bootstrap methods are introduced for detecting nonlinearity in time series data, and examine diagnostic tools that test for three important characteristics, namely: i) linearity, ii) gaussianity and iii) time-reversibility. The tests make it possible to discriminate between those linear processes, which are gaussian, and those, which are nongaussian. Detection of time-irreversibility provides complementary information, since all stationary Gaussian processes are time-reversible.

This paper is divided as follows. The background material is given in Sec. 2. The new tests for nonlinearity, gaussianity and time reversibility are introduced in Sec. 3. Examples using simulated and real data sets are given in Sec. 4. Sec. 5 summarizes the results presented in this work.

2　Background Material

Before proceeding to the description of the proposed tests, it is important to define what is meant by a stochastic *linear* process. A random sampled process $\{x(t_n)\}$ is linear if it is of the form $x(t_n) = \sum_{k=-\infty}^{\infty} h(t_{n-k})\varepsilon(t_k)$ where $\{\varepsilon(t_n)\}$ is a sequence of independent and identically distributed random variables, $\sum_{k=-\infty}^{\infty} |h(t_k)| < \infty$ and $t_n = n\tau$ for a fixed sampling rate τ^{-1}. Using signal processing terminology $\{x(t_n)\}$ is the output of a stable linear filter whose impulse response is $\{h(t_n)\}$ and whose input is the *pure white noise* process $\{\varepsilon(t_n)\}$.

We restrict the null hypothesis of linearity to the class of linear processes that can be whitened using a least squares fit of an autoregressive $AR(p)$ model where the maximum lag p is much smaller than the sample size N of the data to be analyzed. This class includes $AR(p)$ processes where $p << N$ but the restriction is inherently empirical since the user will fit a model using one of the various methods to yield white residuals based on some whiteness criterion. The residuals do not have to mimic a pure noise process as is required for the M.J. Hinich (1996) bicorrelation test for third order dependence since the shuffle bootstrap will destroy any higher order dependence in the residuals.

The signal's bispectrum is

$$B(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} c_{xxx}(\tau_1, \tau_2) \exp\left[-i2\pi(f_1\tau_1 + f_2\tau_2)\right] \qquad (1)$$

where $B(\omega_1, \omega_2)$ is the bispectrum of the signal (M. J Hinich and C. S. Clay, 1968) and $c_{xxx}(\tau_1, \tau_2)$ is the bicorrelation.

The bispectrum is computed using conventional nonparametric methods. When computing the bispectrum we take advantage of two properties: 1) Bispectrum val-

ues are approximately normally distributed (M. J. Hinich, 1982), and 2) Bispectral estimators are approximately independent across frequencies (M. J. Hinich and H. Messer, 1995).

Suppose that we have a sample $\{x(1),...,x(N)\}$ that we partition into $P = [N/L]$ non-overlapping frames of length L where the last frame is dropped if it has less than L observations. The pth frame is $\{x_p(1),...,x_p(L)\} = \{x((p-1)L+1),...,x(pL)\}$. The discrete Fourier transform of the pth frame is $X_p(k) = \sum_{t=1}^{L} x_p(t)\exp\left(-i2\pi\frac{kt}{L}\right)$ and the periodogram of the mth frame is $\frac{1}{L}|X_p(k)|^2 = \frac{1}{L}X_p(k)X_p(-k)$. Because $N \simeq LP$ the frame-averaged estimate of the spectrum at frequency $\omega_k = \frac{2\pi k}{L}$ is

$$\hat{S}(\omega_k) = \frac{1}{N}\sum_{p=1}^{P}|X_p(k)|^2 \tag{2}$$

Then $E\left[\hat{S}_x(f_k)\right] = S(\omega_k) + O\left(\frac{1}{L}\right)$ where the error term of order $1/L$ is due to the frame windowing of the spectrum, and the variance of the estimate for large values of L and P is $\frac{1}{P}S^2(\omega_k)$.

Similarly, the frame-averaged estimate of the bispectrum at frequencies $(\omega_{k_1}, f_{\omega_2})$ is

$$\hat{B}(\omega_{k_1}, \omega_{k_2}) = \frac{1}{N}\sum_{p=1}^{P} X_p(k_1)X_p(k_2)X_p(-k_1-k_2) \tag{3}$$

with $E\left[\hat{B}(\omega_{k_1}, \omega_{k_2})\right] = B(\omega_{k_1}, \omega_{k_2}) + O\left(\frac{1}{L}\right)$ and variance for large L and P expressed as $\frac{L}{P}S(\omega_{k_1})S(\omega_{k_2})S(\omega_{k_1} + \omega_{k_2})$.

The ideas briefly laid here are the base for the well-known Hinich test for gaussianity and linearity of stationary time-series (M. J. Hinich, 1982) and the tests proposed in this work.

3 Linearity, Gaussianity and Time-Reversibility tests

Specific statistical properties of an estimate of the bispectrum are now discussed in order to understand the logic behind the tests (M. J. Hinich, 1982) of linearity and gaussianity and the Hinich-Rothman test for time reversibility (The Fortran program written by Hinich, available, upon request, finds K for whatever band is selected. In (R.A. Ashley et al., 1986), $q = 0.8$ is used but a more robust test uses the $q = 0.9^{th}$ quantile based upon numerous tests of the method on various real and artificial data.).

Let $\{x(t_n)\}$ denote a zero mean strictly stationary random process that is bandlimited and sampled at a rate sufficient to avoid aliasing with $t_n = n\tau$. To simplify notation let $\tau = 1$. The bicorrelation of the process is $c_{xxx}(\tau_1, \tau_2) = Ex(n)x(n+\tau_1)x(n+\tau_2)$ and its bispectrum is the two-dimensional Fourier transform $B_x(\omega_1, \omega_2) = \sum_{n_1=-\infty}^{\infty}\sum_{n_2=-\infty}^{\infty} c_{xxx}(\tau_1, \tau_2)\exp\left[-i(\omega_1\tau_1 + \omega_2\tau_2)\right]$. For further details see (B. Efron, 1979; M. J. Hinich, 1982; M. J. Hinich and H. Messer, 1995).

For linear processes, it follows that $B_x(\omega_1, \omega_2) = \mu_{3\varepsilon} H(\omega_1) H(\omega_2) H(-\omega_1 - \omega_2)$ where $H(\omega)$ is the Fourier transform of h(k), and $\mu_{3\varepsilon} = E\varepsilon^3(n)$ is the skewness of $\varepsilon(n)$.

In what follows a detailed description of the proposed tests for linearity, gaussianity and time-reversibility is given.

3.1   Testing Linearity

Under the linear null hypothesis, as long as the sample size $N >> p$, all the serial correlation in the data $\{x_k\}$ will be removed by the initial AR(p) fit. The critical *null hypothesis* for the linearity test is that the residuals $\{e_k\}$ obtained from the AR(p) fit are *independently distributed*. Thus statistically significant sample bicovariances due to nonlinearities will falsify the null hypothesis.

Let $\hat{B}_e(\omega_1, \omega_2)$ denote the estimate of the bispectrum of the residuals at bifrequency (ω_1, ω_2) using a resolution bandwidth of Δ. Using Theorem 5.3.1 of (D. Brillinger, 1975) it can be shown that the real and imaginary parts $N^{\frac{1}{2}}\Delta[\hat{B}(\omega_1, \omega_2) - B(\omega_1, \omega_2)]$ are independently distributed and gaussian with mean zero and variance $\sigma_e^2/2$ as N goes to infinity. Thus the large sample distribution of the *normalized skewness function* defined by $V(\omega_1, \omega_2) = 2N\Delta^2\sigma_e^{-6}[B(\omega_1, \omega_2)]^2$ is $\chi_2^2(\lambda)$, a chi squared with two degree-of-freedom and non-centrality parameter $\lambda = 2N\Delta^2\sigma_e^{-6}[\mu_{3e}]^2$ for each bifrequency for the null hypothesis of independence. This parameter is estimated by $\hat{\lambda}$, the average skewness $V(\omega_1, \omega_2)$ for all bifrequencies in the bispectrum's principal domain.

Let $F(\upsilon|\lambda)$ denote the cumulative distribution function of a $\chi_2^2(\lambda)$ and let $U(\omega_1, \omega_2) = F[V(\omega_1, \omega_2)|\lambda]$. The normalized skewness values are then transformed into uniform (0,1) variates under the null hypothesis by this transformation. Then the modified Hinich test for linearity (independence of the residuals) is to compute the qth quantile of the sorted U statistics for all K bifrequencies in the principal domain, where the user selects q. If the whole bandwidth up to the folding frequency is used then there are approximately K=$\frac{1}{16\Delta^2}$ bifrequencies in the principal domain (R.A. Ashley et al., 1986). The qth quantile is approximately gaussian with mean q and variance σ^2=q(1-q)/K under the null hypothesis, and we use the q=0.9th quantile (R.A. Ashley et al., 1986).

Using these estimates of the mean and variance, the asymptotic gaussian distribution the 5% critical value for the one tailed test of linearity is easily found to be $0.9 + 0.492/\sqrt{K}$. If the 0.9th quantile is larger than this value the null hypothesis of linearity is rejected at the 5% size level. Thus under the null hypothesis 5% of such statistics would be larger than the above value. Since the gaussian distribution is only a large sample approximation whose accuracy for a given N is unknown, simulations are needed to determine how well the approximation works. Unpublished simulations run by Hinich shows that the test is conservative, that is its false rejection rate for the nominal 5% level is around 2%. This makes the test less powerful than a test that has a true 5% level. To improve the power of the test to detect nonlinearity, some sort of bootstrap is required.

The bootstrap method using surrogates will be described next.

3.1.1 A Surrogate based Test for Nonlinearity

Although the surrogate method was shown to have major drawbacks (M. J Hinich et al., 2002), it is nevertheless possible to make use of surrogate based approaches for detecting nonlinearity. In particular we will make use of the Hinich test for non-linearity, which although already proven to be an effective test, has been shown to be conservative (Douglas M. Patterson and Richard A. Ashley, 2000). A surrogate-based approach has the advantage of providing a more exacting test.

We first appeal to the fact that a large subset of linear processes are contained in the set of stable and invertible AR(p) processes of the form

$$\sum_{k=0}^{p} \beta(k) x(n-k) = \varepsilon(n) \tag{4}$$

where $\beta(0) = 1$. Consider a sample $(x(1), x(2), \ldots, x(N))$ from an AR(p) process. Note that the residuals $(e(1), e(2), \ldots, e(N))$ obtained after fitting an AR(p) model (via the Yule Walker equations) to such a sample are approximately *iid* and will be close approximations to the unobserved pure noise input $\{\varepsilon(n)\}$ when $N \gg p$ (T. W. Anderson, 1971).

The test proposed here requires the following steps:

i) The time series $(x(1), x(2), \ldots, x(N))$ is initially "whitened" by fitting an AR(p) model to the data and separating out the residuals of the fit $(e(1), e(2), \ldots, e(N))$.

ii) A set of M surrogates or bootstraps of the residuals $\{e(k)\}$ are created to yield the surrogate residuals $(e'(1), e'(2), .., e'(N))$. (Alternatively, if surrogates of the original time-series are required, they can be constructed by driving the AR model found in i) with the surrogate residuals $\{e'(1), e'(2), .., e'(N)\}$.

iii) The surrogates allow determination of the 5% critical value for the given test statistic, whether it tests for linearity, gaussianity, or time reversibility etc.

3.2 Testing Gaussianity

If the density of the noise variates $\{\varepsilon(n)\}$ is symmetric about its mean (zero) then the skewness is zero, and its bispectrum will not be statistically significant from zero. The Hinich test statistic (M. J. Hinich, 1982) to test for input symmetry is the sum over the $V(\omega_1, \omega_2)$ for the K bifrequencies. Since the bispectral estimates are approximately independent across the bifrequency grid, this sum will be approximately distributed as a $\chi^2_{2M}(0)$. The non-central parameter is zero for the null hypothesis since the skewness is zero. The null hypothesis is rejected if the sum is greater than a one tailed threshold that is determined by the size probability required by the user who employs the above large sample chi squared distribution for

the sum. Note that a gaussian density is symmetric and thus the Hinich test for gaussianity is really a test for the more general hypothesis of noise density symmetry.

3.3 Testing Time Reversibility

If the purely random process is time reversible then its bicorrelation function will have the symmetry $E\varepsilon(n)\varepsilon(n+\tau_1)\varepsilon(n+\tau_2) = E\varepsilon(n)\varepsilon(n-\tau_1)\varepsilon(n-\tau_2)$ for every τ_1 and τ_2. This implies that the imaginary part of its bispectrum is zero. The Hinich-Rothman test statistic (M.J. Hinich and P. Rothman, 1998) is the sum of $R(\omega_1,\omega_2) = 2N\Delta^2\sigma_e^{-6}|ImB_e(\omega_1,\omega_2)|^2$, which are distributed as a $\chi_M^2(0)$ under the null hypothesis of time reversibility. Thus the test is similar to the Hinich test for noise density symmetry but with M degrees-of-freedom.

4 Applications

The above tests have been successfully applied to a variety of different nonlinear models, and have also been used to test biological, environmental and economic time series. The next three examples will illustrate the application of the proposed tests.

4.1 Gaussian, Uniform and Double Exponential Innovations

Using simulations, we first check the size of the nonlinearity test using three bootstraps a) the Theiler surrogate; b) the temporal shuffle (also called resampling without replacement) and c) Efron's bootstrap (resampling with replacement) for gaussian, uniform and double exponential innovations.

    In the analysis that follows, we make use of an initial set of S=4,000 random (gaussian, uniform and double exponential) 'control' time series $(e(1),e(2),\ldots,e(N))$ (with N=100 here). To check the sizes when testing the mean μ (or any other statistic) we proceed by examining each of the S=4000 'control' time series in turn as follows:

  i) Estimate the mean μ_c of the 'control' time-series.

 ii) Construct M surrogate time-series (e.g., via the method of Theiler et al., the shuffle or the Efron bootstrap).

iii) Determine the distribution of the M means μ of the M surrogate time series.

iv) Calculate the 5% critical value $\mu_{.05}$, for which 5% of the surrogates have a mean value μ that is greater than $\mu_{.05}$.

 v) Determine whether the control time-series has a mean larger than the 5% critical value i.e., whether $\mu_c > \mu_{.05}$.

	Theiler			Bootstrap			Shuffle		
	NL	G	TR	NL	G	TR	NL	G	TR
Gaussian	5.0	4.8	4.6	3.9	1.6	4.4	5.3	5.0	5.2
Uniform	1.5	1.2	0.7	4.0	3.3	5.4	5.0	4.9	5.3
Exp.	12.1	27.1	15.9	2.8	0.7	4.1	5.0	4.9	5.2

Table 1: Sizes for nonlinearity (**NL**), gaussianity (**G**) and time reversibility (**TR**) tests.

Repeating steps $(i - v)$ for each of the S=4,000 random control time series, the size α may be calculated by determining the proportion of times for which $\mu_c > \mu_{.05}$. If the bootstrap is operating correctly the size should be α=5%.

Surprisingly, as Table 1 shows, the only bootstrap that is successful is the temporal shuffle, which maintains sizes reasonably close to the expected 5% level in all cases. The same proves to be true when the exercise is repeated for the tests of gaussianity and time reversibility (TR). Table 1 confirms that the Theiler method provides the correct size for gaussian distributions only.

The same procedure was repeated for S=500 random control time series and the results closely follow the ones presented in Table 1. This indicates that the the results presented here are consistent and do not depend on the sample size.

Note that an AR model is not fitted in this case to avoid the contamination of the simulations with the nuisance parameters of various AR models. The tests are only simulating what can be retrieved from the residuals when N is much larger than p and when p is good enough to get white residuals.

4.2   Henon Map

Consider first the example from Theiler where four independent realizations of the Henon map (the x-coordinate) are added yielding a time series of N=1,000 points. The superimposed Henon data is fitted with a recursive AR procedure that finds the model that minimizes the sum of squared residuals. If for example we start with p=10, the routine fits an AR(10) and then finds the t-values for the lag parameter estimates. If the t-value's probability value of say lag 2 is greater than a preset threshold then that lag is removed from the next fit. The procedure continues until either all lags from one to ten are not significant or the remaining lags are significant with respect to the threshold. The best fit found was an AR(6) model with lags 1,3,4,5, and 6 (R^2=0.28).

The one tailed 0.9^{th} quantile test for nonlinearity based on asymptotic theory had a probability level of p=0.03, and thus was significant at the 3% level. However, this is an asymptotic result and thus open to interpretation for finite data sets, particularly when there is border-line significance as found in this example. We thus repeated the test on 500 'shuffle' bootstraps of the observed data. Not one of the bootstrapped test statistics had a probability level greater than p=0.03. Hence the bootstrapped test found the data to exhibit significant nonlinearity. Similarly, the test for gaussianity and time reversibility were both rejected with p<0.0001.
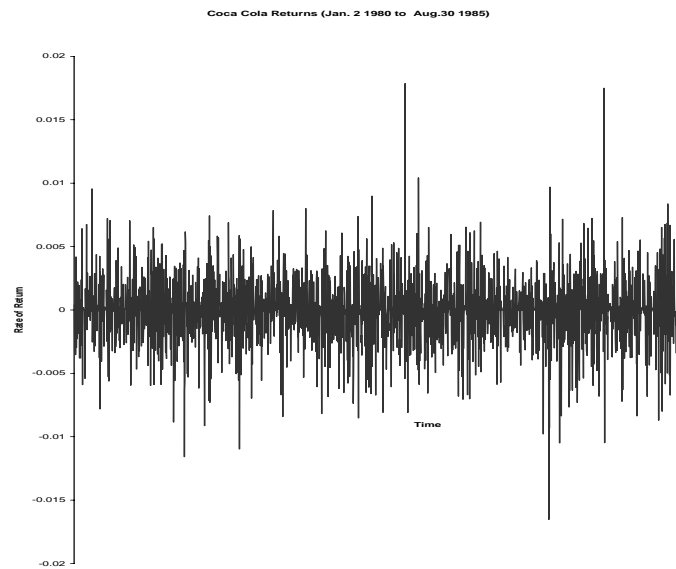
Figure 1: Coca Cola daily rate returns Jan 2, 1980 to Aug. 30, 1985

4.3 Coke Data

Fig 1 displays a series of within day rates of return of Coca Cola from January 2, 1980 to August 30, 1985. These rates of return were constructed from the actual traded prices by a method that obtains unaliased ten minute aggregates for each trading day. The details of the sampling method used are in (T. Schreiber, 1998). There are 36 of these ten minute return aggregates for each trading day yielding N=51,622 data points. Fig. 2 shows all significant U-values in the principal domain of the bispectrum at the appropriate bispectral frequencies. The 0.9^{th} quantile test for nonlinearity based on asymptotic theory had a probability level of p=0.002, and thus is significant at the 0.2% level. With such high significance, there is no need for a bootstrap test, but in any case the latter detected unequivocally nonlinearity in the time series. Similarly, the test for gaussianity and time reversibility were both rejected with p<0.0001.

It could be argued that the Coke data is dominated by a GARCH-type model and therefore the results shown above are not interesting. In order to show otherwise the Coke data was clipped. The sampling interval is 10 minutes and there are 36 observations per trading day. The data is clipped to 1/-1 where $r(t) = 1$ if the return is positive and -1 otherwise. It has been noted that changing the clips to be around the mean doesn't change the results.

Tables 2 and 3 show the parameters, amplitudes and periods for the AR(4 and 5) fitted to the clipped coke data. Note that the four AR coefficients are highly statistically significant for the whole sample. This implies that the returns can not
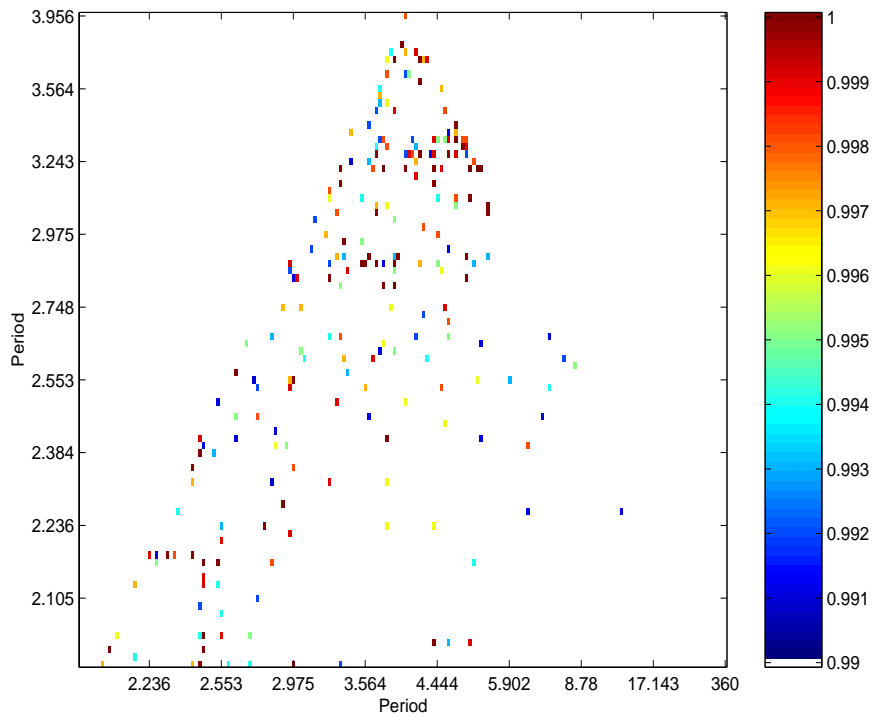
Figure 2: Raw data - Bifrequencies whose probability values are greater than 0.99 are marked in the principal domain of the bispectrum. The rest are set to zero to remove noise clutter

	AR(1)	AR(2)	AR(3)	AR(4)
Parameters	-0.14	-0.06	-0.06	-0.03
T values	-31.70	-12.42	-12.59	-7.43

Table 2: AR(4) parameters / t values. Adjusted R Square = 0.023 Std Error of AR Fit = 0.214E-02. p Value Threshold for the Iterative AR Prewhitening Method = 0.010

	AR(1)	AR(2)	AR(3)	Ar(5)
Amplitudes	0.39	0.39	0.46	0.46
Periods	2.52	2.52	6.26	6.26

Table 3: Amplitudes & Periods of the Roots of the AR Polynomial in units of 10min. A real root is given a period = 0.
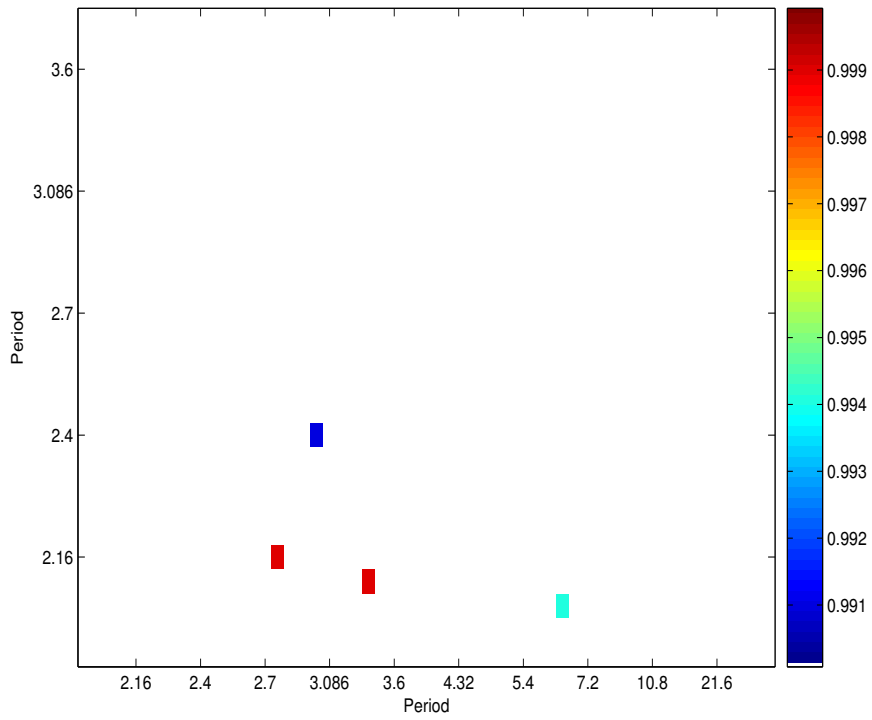
Figure 3: Clipped data - Bifrequencies whose probability values are greater than 0.99 are marked in the principal domain of the bispectrum. The rest are set to zero to remove noise clutter
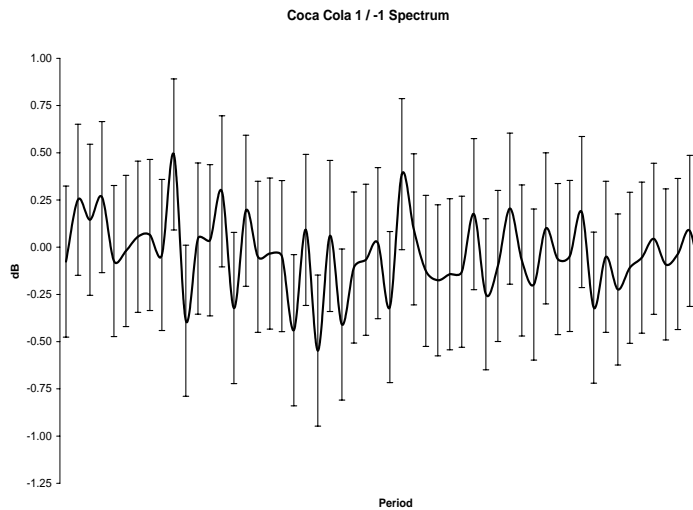

Figure 4: Spectrum for Coke clipped data

be ARCH or GARCH. To corrobate these results bispectrum values were calculated and shown to be highly significant. See Figures 3 and 4 for the bispectrum values and spectrum of the coke clipped data.

5 Conclusions

It should be observed that no comparison was made between the proposed tests and the ones available in the literature (BDS, NEGM, Kaplan Test etc.). As rightly stated in (W. Barnett et al., 1997) the available can not be completely compared as their null are not enterily compatible. However many of the available tests can be used jointly. In this paper, new tests for nonlinearity, gaussianity and time reserversibility using surrogate and bootstrap methods have been proposed. It has been shown that to improve the power of Hinich's earlier tests to detect nonlinearity, gaussianity and reversibility some sort of bootstrap was required. In particular, the use of shuffle bootstrap in conjuction with Hinich's tests was demonstrated to hold proper sizes. Two simulated and one real data examples have been given to illustrate the power of the tests.

In the example using the Coke Data, the results presented in this paper are in accordance with (C. Brookes and M. J Hinich, 1998), where the authors tested the validity of specifying a GARCH error structure for financial time series data in the context of a set of ten daily Sterling exchange rates. The results demonstrate that there are statistical structures present in the data that cannot be captured by a GARCH model, or any of its variants. The nonlinear structure of the data was unequivocally detected using the proposed tests.

Finally, the main result in this paper is the introduction of a shuffle bootstrap to set the critical values for the tests and to compare the shuffle bootstrap with two other bootstrap methods.

References

M.J. Hinich and D.M. Patterson. Evidence of Nonlinearity in the Trade-by-Trade Stock Market Return Generating Process. In and K. Shell W. Barnett, J. Geweke, editor, *Economic Complexity, Chaos, Sunspots, Bubbles, and Nonlinearity*, volume 3, chapter 16, pages 383–409. Cambridge University Press, 1989.

B. Efron. Computers and the theory of statistics. *SIAM Rev.*, 21:460–480, 1979.

R. M. Vogel and A. L. Shallcross. The moving blocks bootstrap versus parametric time series models. *Water Resources Research*, 32(6):1875–1882, 1996.

P. Hall, J. L. Horowitz, and B. -Y. Jing. On blocking rules for the bootstrap with dependent data. *Biometrika*, 82:561–574, 1995.

J. Hidalgo. A alternative bootstrap to moving blocks for time series regression models. Discussion Paper, No. EM/03/452 452, London School of Economics and Political Science, 2003.

J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J.D. Farmer. Testing for nonlineaity in time series: the method of surrogate data. *Physica D*, 85:77, 1992.

D. Prichard and J. Theiler. Generating Surrogate Data for Time Series with Several Simultaneously measured variables. *Phys. Rev. Lett*, 73:991–954, 1994.

J. Theiler and D. Pichard. Using "Surrogate Surrogate Data" to Calibrate the Actual Rate of False Positives in Tests for Nonlinearity in Time Series. *Fields Institute Communications*, 11:99–113, 1997.

Y.-J. Lee, Y.-S. Zhu, Y.-H. Xu, M.-F. Shen, H.-X. Zhang, and N. V. Thakor. Detection of Non-linearity in the EEG of Schizophrenic Patients. *Clinical Neurophysiology*, 112:1288–1294, 2001.

F.X. Witkowski, K.M. Kavanagh, P.A. Penkoske, R. Plonsey, M.L. Spano, W.L. Ditto, and D.T. Kaplan. Evidence for Determinism in Ventricular-fibrillation. *Phys. Rev. Lett*, 75:1230–1233, 1995.

P. So, E. Ott, S. J. Schiff, D. T. Kaplan, T. Sauer, and C. Grebogi. Detecting Unstbale Periodic Orbits in Chaotic Experimental Data. *Phys. Rev. Lett*, 76:4705–4708, 1996.

M. J Hinich, L. Stone, and E. M. A. M. Mendes. Surrogate data method: False alarms, confidence intervals and validity. *Submitted to Technometrics*, 2002.

M.J. Hinich. testing for Dependence in the Input to a Linear Time Series Model. *Journal of Nonparametric Statistics*, 6:205–221, 1996.

M. J Hinich and C. S. Clay. Application of discrete fourier transform in the estimation of power spectra, coherence and bispectra of geophysical data. *Reviews of Geophysics*, 6(3):347–363, 1968.

M. J. Hinich. Testing for Gaussianity and Linearity of a Stationary Time Series. *Journal of Time Series Analysis*, 3(3):169–176, 1982.

M. J. Hinich and H. Messer. On the Principal Domain of the Discrete Bispectrum of a Stationary Signal. *IEEE Trans. on Signal Processing*, 43(9):2130–2134, 1995.

R.A. Ashley, D.M. Patterson, and M.J. Hinich. A Diagnostic Test for Nonlinear Serial Dependence in Time Series Fitting Errors. *Journal of Time Series Analysis*, 7:165–178, 1986.

D. Brillinger. *Time Series, Data Analysis and Theory*. Holt, Rinehart, and Winston, New York, 1975.

Douglas M. Patterson and Richard A. Ashley. *A Nonlinear Time Series Workshop : A toolkit for detecting and identifying nonlinear serial dependence*. Boston: Kluwer Academic Publishers, 2000.

T. W. Anderson. *The Statistical Analysis of Time Series*. John Wiley and Sons, New York, 1971.

M.J. Hinich and P. Rothman. A Frequency Domain Test of Time Reversibility. *Macroeconomic Dynamics*, 2:72–88, 1998.

T. Schreiber. Constrained Randomization of Time Series. *Phys. Rev. Lett*, 80:2105–2108, 1998.

W. Barnett, A. R. Gallant, M. J. Hinich, J. A. Jungeilges, D. T. Kaplan, and M. J. Jensen. A single–blind controlled competition among tests for nonlinearity and chaos. *Journal of Econometrics*, 82:157–192, 1997.

C. Brookes and M. J Hinich. Episodic nonstationarity in exchange rates. *Applied Economics Letters*, 5(11):719–722, 1998.