Article Submitted to Journal of the American Statistical Association

Detecting Nonlinearity in Time Series: Surrogate and Bootstrap Approaches

MELVIN J.HINICH*[1], L. STONE†[2] AND EDUARDO M. A. M. MENDES‡[3]

[1]*Applied Research Laboratories, University of Texas at Austin, Austin TX 78713-8029 USA*

[2]*Biomathematics Unit, Porter Super center for Ecological and Environmental Studies, Tel Aviv University, 69978, Israel.*

[3]*Departamento de Engenharia Eletrônica, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Belo Horizonte, MG, Brasil, 31.270-901, Tel: +55 (31) 3449 48662, Fax: +55 (31) 3449 4850*

Abstract

In Physics literature, the method of surrogate data has been widely used for testing for the presence of nonlinearity in time series whereas in the Statistical literature, the bootstrap method is the choice for establishing confidence intervals. A new method that combines the bispectrum and the surrogate method and bootstrap is then presented for detecting nonlinearity, gaussianity and time reversibility.

Keywords: Bootstrap, Surrogate data method, random processes

1. Introduction

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 threshold or some form of resampling known in statistics as bootstrapping. In practice, one cannot determine the validity of thresholds 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 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. 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

*E-mail: hinich@mail.la.utexas.edu

†E-mail: lewi@lanina.tau.ac.il

‡E-mail: emmendes@cpdee.ufmg.br

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 [Ying-Jie 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 hypotheses based on even the simplest of test statistics [Melvin 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 new tests for nonlinearity, gaussianity and time reversibility are introduced in Sec. 3, 4 and 5, respectively. Examples using real data sets are given in Sec. 6. Sec. 7 summarizes the results presented in this work.

2. A Surrogate based Test for Nonlinearity

Although the surrogate method was shown to have major drawbacks [Melvin 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 nonlinearity, 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

$$\sum_{k=0}^{p} \beta(k)\, x(n-k) = \varepsilon(n) \tag{1}$$

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 threshold 5% false alarm rate for the given test statistic, whether it tests for linearity, gaussianity, or time reversibility etc.

A brief review of the statistical properties of an estimate of the bispectrum is required 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 τ =1. The bicorrelation of the process is $c_{xxx}(m_1, m_2) = Ex(n)x(n+m_1)x(n+m_2)$ and its bispectrum is the two-dimensional Fourier transform $B_x(\omega_1, \omega_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} c_{xxx}(m_1, m_2)\exp[-i(\omega_1 m_1 + \omega_2 m_2)]$. For further details see [B. Efron, 1979, M. J. Hinich, 1982, M. J. Hinich and H. Messer, 1995].

The process is called *linear* if it is the output of a linear filtering operation $x(n) = \sum_{k=0}^{\infty} h(k)\varepsilon(n-k)$ whose input $\{\varepsilon(n)\}$ is a sequence of independent and identically distributed zero mean random variables (called purely random noise). It then 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 $\varepsilon(n)$.

3. 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 $\hat{B}_e(w_1, w_2)$ denote the estimate of the bispectrum of the residuals at bifrequency (w_1, w_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 $N^{\frac{1}{2}}\Delta[\hat{B}(w_1, w_2) - B(w_1, w_2)]$ 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 $V(w_1, w_2) = 2N\Delta^2\sigma_e^{-6}[B(w_1, w_2)]^2$ 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 $U(w_1, w_2) = F[V(w_1, w_2)|\lambda]$. 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% threshold 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 threshold the null hypothesis of linearity is rejected at the 5% false alarm level. Thus under the null hypothesis 5% of such statistics would be larger than the above threshold. 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.

Using simulations, we first check the false alarm rate of the nonlinearity test using three bootstraps a) the Theiler surrogate; b) the temporal shuffle and c) Efron's bootstrap for gaussian, uniform and exponential innovations.

In the analysis that follows, we make use of an initial set of S=4,000 random (gaussian, uniform and exponential) 'control' time series (e_1, e_2,. . . .,e_N) (with N=100 here). To check the false alarm rates 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 described below).

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

iv) Calculate the 5% threshold level $\mu_{.05}$, 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% threshold i.e., whether $\mu_c > \mu_{.05.}$

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

	Theiler			**Bootstrap**			**Shuffle**		
	Nonlin	**Gauss**	**TR**	**Nonlin**	**Gauss**	**TR**	**Nonlin**	**Gauss**	**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**
Exponential	**12.1**	**27.1**	**15.9**	**2.8**	**0.7**	**4.1**	**5.0**	**4.9**	**5.2**

Table 1: False alarm rates of nonlinearity, gaussianity and time reversibility tests.

Surprisingly, as Table 1 shows, the only bootstrap that is successful is the temporal shuffle, which maintains false alarm rates 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) to be described shortly. Table 1 confirms that the Theiler method provides the correct false alarm rate for gaussian distributions only.

4. 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 false alarm 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.

5. Testing Time Reversibility

If the purely random process is time reversible then its bicorrelation function will have the symmetry $E\varepsilon(n)\varepsilon(n+m_1)\varepsilon(n+m_2) = E\varepsilon(n)\varepsilon(n-m_1)\varepsilon(n-m_2)$ for every $m_1$andm_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 $R(\omega_1, \omega_2) = 2N\Delta^2\sigma_e^{-6}\left|ImB_e(\omega_1, \omega_2)\right|^2$, which are distributed as a $\chi^2_M(0)$ under the null hypothesis of time reversibility. Thus the H-R test is similar to the Hinich test for noise density symmetry but with M degrees-of-freedom.

6. 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 two examples will illustrate the application of the proposed tests.

6.1. 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.

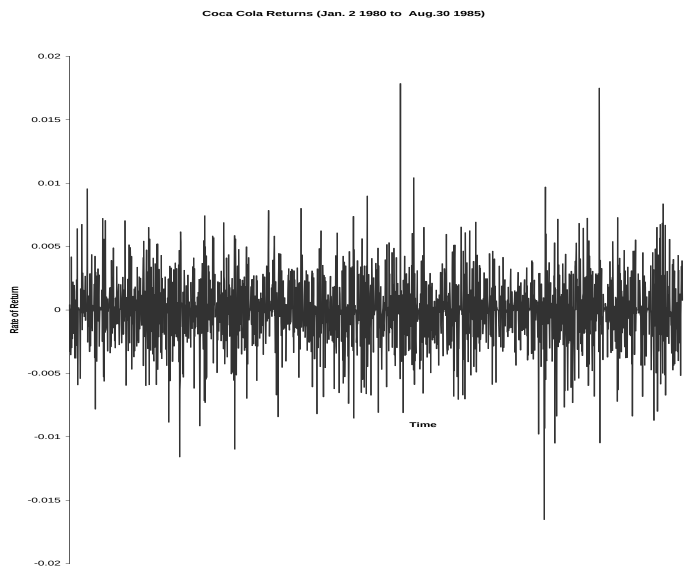

Figure 1: Coca Cola daily rate returns Jan 2, 1980 to Aug. 30, 1985

6.2. 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 plots 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.

7. Conclusions

In this paper, new tests for nonlinearity, gaussianity and time reserversibility using surrogate and bootstrap methods has been proposed. Real data examples have been given to illustrate the power of the tests.

References

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

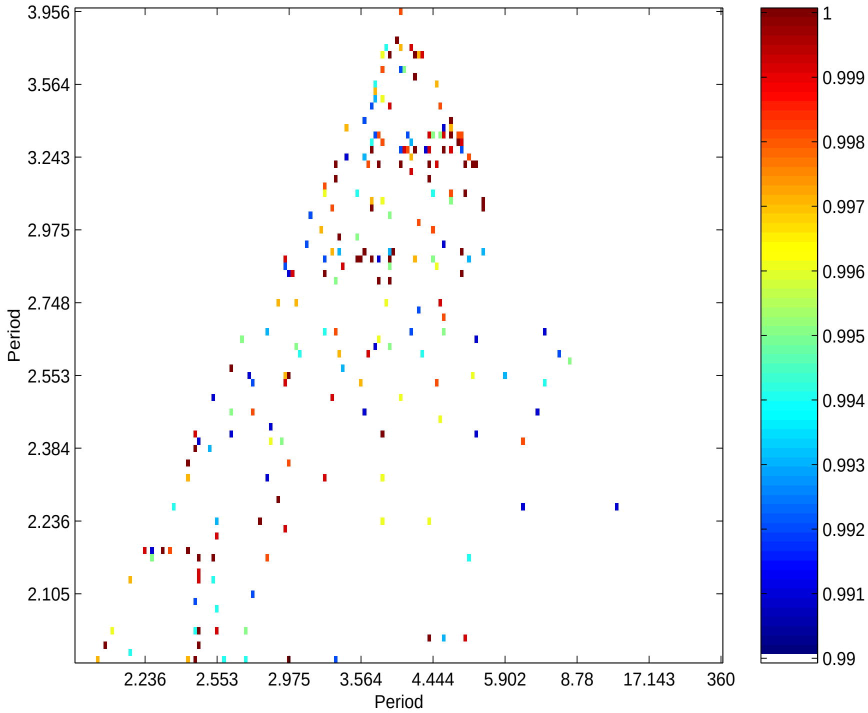

Figure 2: Bifrequencies whose probability values are greater than 0.99 are marked in the principal domain of the bispectrum. The rest are set to zero to remove noise clutter

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.

Ying-Jie Lee, Yi-Sheng Zhu, Yu-Hong Xu, Min-Fen Shen, Hong-Xuan 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.

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

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

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.

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.

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

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.