2005 International Symposium on Nonlinear
Theory and its Applications (NOLTA2005)
Bruges, Belgium, October 18-21, 2005

Nonparametric Rank Test for Nonlinearity Detection

Xiaodong Luo, Jie Zhang and Michael Small

Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
Email: enxdluo@eie.polyu.edu.hk, enzhangjie@eie.polyu.edu.hk, ensmall@eie.polyu.edu.hk

Abstract—
In this report we propose a rank test scheme to detect the potential nonlinearity in a scalar time series. Our scheme is based on the fact that, for a stationary linear stochastic process with jointly symmetric innovations, it proves that its ordinary least square (OLS) prediction error of linear autoregressive (AR) predictors are symmetric about zero. With this knowledge, a discriminating statistic, namely the Wilcoxon signed rank statistic, can be derived from the prediction error. The advantage of this statistic is that it has a known null distribution, thus we can perform statistical inference of the underlying system with exact confidence level. In addition, the exactness of the null distribution of the rank statistic does not depend on the sample size, which is usually not possessed by many other discriminating statistics such as the correlation dimension. To illustrate the discriminating power of the test scheme, we examine several examples through our methods.

1. Introduction

It has become popular to apply nonlinear statistics such as the correlation dimension and the Lyapunov exponent [1] for identification of the underlying dynamical systems in many fields. However, as it was reported in [2], these statistics could have finite and predictable values even for a linear stochastic process with simple autocorrelation, therefore, if one intends to identify the determinism of the underlying system by simply examining if its nonlinear statistics are convergent, it is quite possible that one will mistake a linear stochastic process for a nonlinear deterministic system. This observation suggests that one instead examines the basic properties, e.g. nonlinearity, determinism, continuity and so on, of the underlying system in order to apply nonlinear analysis methods with more safety. Various methods based on this viewpoint has been developed to investigate the distinction between (stationary) linear stochastic and nonlinear deterministic systems (for example, [4, 5]).

In general, these methods focus on exploring the difference of some characteristic behaviors or properties between stochastic and deterministic systems [16]. As a measurement of the reliability of the test results, the confidence level of the inference will be always preferred. However, it is often difficult for us to find out the exact confidence level, especially for practical situations wherein only a scalar time series from an unknown source is available. A remedy for this case, as proposed in [4], is to first assign a null hypothesis to the underlying system (usually assume that the underlying system is linear stochastic), and then apply the bootstrap method to produce a set of surrogate data, which should have the same statistic distribution as the original time series under the null hypothesis. One can calculate the statistic values of both the original time series and the surrogates, and thus obtain the empirical null distribution of the statistic in test. One then determines whether to reject the null or not. However, since the exact knowledge of the statistic distribution is often not available, one will resort to certain discriminating criterion to help make the decision more objectively and determine the corresponding confidence level (if to reject).

In the past research works, there are usually two classes of discriminating criteria: One is parametric. It assumes that the statistic follows a Gaussian distribution, and the distribution parameters, i.e. the mean and the variance, would be estimated from the finite samples. One can determine whether to reject the null by examining that if the statistic of the original time series follows the statistic distribution of the surrogates. The corresponding confidence level of inference can be calculated from the estimated statistic distribution; The other one, proposed in [6], is nonparametric. The main idea is that, suppose the statistic of the original time series is ϕ_0 and the surrogate values are $\{\phi_i\}_{i=1}^N$ given N surrogate realizations, then if the statistic of both the original time series and the surrogates follows the same distribution, the probability is $1/(N+1)$ for ϕ_0 to be the smallest or largest among all of the values $\{\phi_0, \phi_1 ... \phi_N\}$. Thus if N is large, when one finds that ϕ_0 is smaller or larger than all of the values in $\{\phi_i\}_{i=1}^N$, it is quite possible that ϕ_0 instead follows a different distribution from that of $\{\phi_i\}_{i=1}^N$. Hence the criterion rejects the null hypothesis whenever the original statistic ϕ_0 is the smallest or largest among $\{\phi_0, \phi_1 ... \phi_N\}$, the false rejection rate is considered as $1/(N+1)$ for one-sided tests and $2/(N+1)$ for two-sided ones.

Although the above two criteria are heuristic, they are often questionable in practice. For example, for the first criterion, the normality assumption may be seriously violated. For the second, let us discuss it with slightly more details. Suppose that the statistic values ϕ_0 and $\{\phi_i\}_{i=1}^N$ follow the same distribution and let the range of the surrogate values $\{\phi_i\}_{i=1}^N$ be Ψ, while the support of the null distribution be

Ω, and $\Delta = \Omega \backslash \Psi$ be the complement set of Ψ given Ω. Then according to the rule, one rejects the null hypothesis whenever the original statistic $\phi_0 \in \Delta$, and the actual false rejection rate is the probability $\Pr(\phi_0 \in \Delta | \{\phi_i\}_{i=1}^N)$, which usually cannot be simply estimated according to the number of surrogate realizations only.

In this report we propose a new test scheme to detect the potential nonlinearity of a scalar time series. The focus of this scheme is that we adopt a new measure, namely the Wilcoxon signed rank statistic, which could be derived from the linear prediction error of the time series and proves to be a Wilcoxon variate under weak conditions. As an advantage, this test statistic has a known null distribution, thus inference with exact confidence level becomes possible based on the knowledge of the statistic distribution. In addition, the exactness does not depend on the sample size, as a good property not shared by many nonlinear discriminating statistics.

The remaining of this report goes as follows: In Sec. 2 we will mainly introduce our test scheme, including our null hypothesis, the discriminating statistic and the procedures to determine the confidence level of inference. In Sec. 3 we illustrate the test power of our scheme through several examples. Finally we summarize the whole report.

2. A New Test Scheme

As the first step to introduce our test scheme, we need to specify the null hypothesis. Since we aim to detect if there exists nonlinearity of the underlying system given a stationary time series $\{x_i\}$, we assume that *the time series $\{x_i\}$ is from a linear stochastic process with independent jointly symmetric innovations*, which is the null hypothesis adopted in the whole report. For clarity, let us explain the null in more detail. According to the theorem of Wold decomposition [7, p. 65], under the condition of stationarity, a purely stochastic process $\{x_i\}$ can be expressed in an infinite moving average ($MA(\infty)$) process. For convenience of our later discussion, we adopt the p-th order autoregressive ($AR(p)$) process instead to model the stationary time series with the concrete form of

$$x_i = a_0 + \sum_{j=1}^{p} a_j x_{i-j} + \epsilon_i, \tag{1}$$

where $\{\epsilon_i\}$ denote the innovation terms, which are considered to be independent of $\{x_i\}$, mutually independent of each other and have a distribution with joint symmetry. By "joint symmetry" of a stochastic process $\{\epsilon_i\}$ we mean that there exists some constant μ so that $\{\epsilon_i - \mu\}$ and $\{\mu - \epsilon_i\}$ have the same joint distributions, i.e. the probability density function (PDF) $f(\epsilon_1 - \mu, \epsilon_2 - \mu, ...) = f(\mu - \epsilon_1, \mu - \epsilon_2, ...)$ [8]. So it is easy to see that, the linear autocorrelated Gaussian processes examined in [4] are consistent with our null hypothesis. However, the coverage of our null could be extended to a wider range, e.g. the stationary processes with independent (not necessarily identical) and jointly symmetric innovations.

After the selection of the null hypothesis, we then choose the discriminating statistic. To derive our statistic, let us first consider the problem of predicting k-step ahead value of a linear stochastic process $\{x_i\}$ with jointly symmetric innovations $\{\epsilon_i\}$. Let $\hat{x}_i^k$ be the prediction at time i ($\hat{x}_i^k = x_{i+k}$ if $k \leqslant 0$), and $e_i^k = x_{i+k} - \hat{x}_i^k$ denote the corresponding prediction error. In [8] Dufour proves that, when we use an ordinary least square (OLS) linear predictor $\hat{x}_i^k = a_{i,0} + \sum_{j=1}^{p'} a_{i,j} \hat{x}_i^{k-j}$ [17] to predict the k-step ahead value of a linear stochastic processes $\{x_i\}$ with jointly symmetric innovations, even if we misspecify the fitting order of the predictor (either lower or higher) and thus adopt inaccurate estimated parameters in prediction, the distributions of the prediction error e_i^k will be symmetric about zero, i.e. e_i^k and $-e_i^k$ share the same distribution. This fact immediately implies that the probability $\Pr(e_i^k > 0) = \Pr(-e_i^k > 0) = \Pr(e_i^k < 0) = 1/2$ when the distribution of e_i^k is continuous so that the probability $\Pr(e_i^k = 0) = 0$ in the sense of Lebesgue measure (see [9] and [10, Lemma 10.1.24] for more details).

Let $\{I_i\}_{i=1}^m$ be an indicator series with m data points so that $I_i(e_i^k) = 1$ if $e_i^k > 0$ and $I_i(e_i^k) = -1$ if $e_i^k < 0$. Clearly I_i is a Bernoulli variate uniformly distributed on $\{-1, 1\}$, i.e. $\Pr(I_i = -1) = \Pr(I_i = 1) = 1/2$. With this knowledge, we can derive a class of linear signed rank statistics

$$SR_m = \sum_{i=1}^{m} I_i(e_i^k) \times S_i(rank(|e_i^k|)) \tag{2}$$

to test our null hypothesis, where $\{S_i(\cdot)\}$ is the set of scores of the series $\{|e_i^k|\}_{i=1}^m$ with $rank(|e_i^k|)$ denoting the rank (in the ascending order) of the absolute value $|e_i^k|$ among $\{|e_i^k|\}_{i=1}^m$ [10, p. 252]. Here we choose $S_i(rank(|e_i^k|)) = rank(|e_i^k|)$ so that

$$SR_m = \sum_{i=1}^{m} i \times I_i(e_i^k) \tag{3}$$

is the widely used Wilcoxon signed rank statistic, which is well approximated by the normal distribution $N(0, m(m+1)(2m+1)/6)$ for even small numbers, say, $m = 15$ [11, chaper 2].

Since we know the distribution of the test statistic, based on the realization value of SR_m, we can determine whether to reject the null hypothesis with an exact confidence level. Take two-sided test as an example, if we want the type-I error (the false rejection rate of a correct null) to be less than α, then we first find two critical values n_u and n_l such that n_u is the largest integer satisfying $\Pr(SR_m > n_u) < \alpha/2$ and n_l is the smallest integer satisfying $\Pr(SR_m < n_l) < \alpha/2$. If for a time series in test, its statistic $SR_m > n_u$ or $SR_m < n_l$, then we reject the null hypothesis. The false rejection rate is α, or in other words, the confidence level to reject the null hypothesis is $1 - \alpha$. The procedures to perform one-sided tests are similar, except that we need to locate only one critical value, n_u or n_l, which instead satisfies $\Pr(SR_m > n_u) < \alpha$ or $\Pr(SR_m < n_l) < \alpha$ separately for right or left side test.

Table 1: Numbers of rejections of the null hypothesis (out of 1000 replica) for data generation processes (DGPs) with different fitting orders.

DGP	Rejections for the fitting order of				
	6	7	8	9	10
AR(6)	0	0	0	0	0
ARMA(1,1)	43	41	43	48	48
Henon	1000	999	999	1000	1000
Rössler	1000	1000	1000	1000	1000

Table 2: Rejections of the null hypothesis for 1000 temporal-shift surrogates of DGPs with different fitting orders.

DGP	Rejections for the fitting order of				
	6	7	8	9	10
AR(6)	0	0	0	0	0
ARMA(1,1)	0	0	0	0	0
Henon	1000	1000	1000	1000	1000
Rössler	1000	1000	1000	1000	1000

Table 3: Rejections of the null hypothesis for 1000 constrained-realization surrogates of DGPs with different fitting orders.

DGP	Rejections for the fitting order of				
	6	7	8	9	10
AR(6)	0	0	0	0	0
ARMA(1,1)	0	0	0	0	0
Henon	1000	1000	1000	945	766
Rössler	1000	1000	1000	1000	1000

We will apply the above idea to test our null hypothesis for nonlinearity detection. The whole procedures go as follows: For each time series $\{x_i\}$ in test, we first predict its one-step ahead values $\{\hat{x}_i^1\}$ via the OLS linear predictor and calculate the prediction error. Suppose the error series $\{e_i^1\}_{i=1}^m$ has m data points, then we use the normal distribution $N(0, m(m+1)(2m+1)/6)$ to find the critical values (n_u and n_l) for two-sided test at the nominal confidence level of 95%. If the calculated Wilcoxon statistic $SR_m \notin [n_l, n_u]$, then we can reject the null hypothesis with the false rejection rate no more than 5%.

3. Examples

In the following we will demonstrate through several examples the power of our scheme for the null hypothesis test. The first example is an $AR(6)$ process $x_i = \sum_{j=1}^{6} a_j x_{i-j} + \epsilon_i$ with coefficients $(a_1, ..., a_6) = (0.6,\ 0,\ 0.5,\ 0,\ -0.6,\ 0.3)$, where innovations $\{\epsilon_i\}$ are uniformly distributed on interval $[0,\ 0.1]$ (symmetric about 0.05). The second is an $ARMA(1,1)$ process, i.e. $x_i = a_1 x_{i-1} + \epsilon_i - b_1 \epsilon_{i-1}$ with parameters $a_1 = b_1 = 0.5$, where innovation terms $\{\epsilon_i\}$ follow the normal distribution $N(0,1)$. The third data generation process (DGP) is the Hénon map [12] $H(x,y) = (y + 1 - \alpha x^2, 0.3x)$, where parameter α is uniformly drawn from the interval $[1.35,\ 1.4]$. We will take out the first coordinate x for test. The final case is the Rössler system [13] with continuous description equations of $(\dot{x}, \dot{y}, \dot{z}) = (-y - z, x + 0.15y, 0.2 + xz - cz)$, where parameter c is uniformly drawn from the interval $[9.5,\ 10]$. The sampling time is 0.1 time units, and the observations for calculation are taken from the second coordinate y.

Since usually one does not know the true order of an underlying process, a fitting order has to be specified for prediction. As aforementioned, if the underlying process of the test time series is linear stochastic with jointly symmetric innovations, misidentification of the fitting order would also lead to the symmetric prediction error. Thus we need not seek the optimal fitting order of prediction for our purpose. Instead, we could simply choose several fitting orders for all of the DGPs, say, those starting from 6 to 10 [18]. To indicate the power of our test scheme, for all of the DGPs in examination, we produce 1000 realizations with 2000 data points for each, and predict the one-step ahead values for the last 1000 data points. With the prediction error, we calculate the Wilcoxon signed rank statistic to determine whether to reject the null or not. We record the rejection numbers of our null hypothesis and show them in Table 1, from which we see that our test scheme indicates remarkable power. For the linear stochastic processes of $AR(6)$ and $ARMA(1,1)$, the rejection rates are less than 5%, as we expect. While for the two nonlinear systems, their rejections are far beyond the expected rates, thus we are almost sure that they cannot be linear stochastic processes.

In practical situations, one often has only a scalar time series on hand. Therefore, for the reliability of the test, we suggest that one first uses the bootstrap method, such as [4] and [14], to generate a number of surrogates, and then calculates the test statistic of the surrogates and determines whether to reject the null hypothesis or not. If the actual rejection rate is higher than the nominal one (i.e. the rate when the original time series is consistent with our null hypothesis), then we can safely reject the null hypothesis. For illustration, let us examine the previous examples again. For each example, we generate only one sample, and use the bootstrap method to generate 1000 surrogates of the sample. In our test scheme, we adopt the temporal-shift algorithm in [14] to generate surrogates since it need not conduct the Fourier transform and thus avoids some of the shortcomings [19]. The main idea of the algorithm is that, if a time series $\{x_i\}$ is consistent with our null hypothesis, i.e. it may have a form of $x_i = a_0 + \sum_{j=1}^{p} a_j x_{i-j} + \epsilon_i$, then we see that, for any coefficients β and γ, the surrogates $\{y_i^\tau = \beta x_i + \gamma x_{i+\tau}\}$ also follow linear stochastic forms with the constants of $(\beta+\gamma)a_0$ and the innovations of $\{\beta\epsilon_i + \gamma\epsilon_{i+\tau}\}$ respectively. A linear stochastic time series $\{x_i\}$ will always produce linear stochastic surrogates $\{y_i^\tau = \beta x_i + \gamma x_{i+\tau}\}$,

therefore in principle we cannot reject the null hypothesis test via the Wilcoxon statistic. In our calculations, we let parameter β be uniformly drawn from the interval [0.6, 0.8] and parameter $\gamma = (1 - \beta^2)^{1/2}$ to produce surrogates. The results of null hypothesis test are presented in Table 2, from which we see that, for the two linear stochastic processes, we cannot reject the null hypothesis; while for the two nonlinear systems, the null hypothesis is rejected at very high rates as expected. For comparison, we also generate 1000 surrogates through the constrained-realization method [4]. The test results are listed in Table 3, from which we could obtain the same conclusion.

4. Conclusion

To summarize, we have proposed an exact nonparametric inference scheme to detect the potential nonlinearity in a scalar time series. The exactness of our inference comes from the knowledge of the exact distribution of the adopted discriminating statistic, i.e., the Wilcoxon signed rank statistic, which indicates remarkable test power through the several examples examined in this communication. One advantages of this statistic is that it possesses the known null distribution. Thus it is easy for us to find the exact confidence interval for the inference of the underlying process. Furthermore, the exactness of the statistic distribution does not rely on the size of the sample time series in test. Comparatively, for many nonlinear discriminating statistics adopted in the literature, e.g. the correlation dimension, computations with too short samples may cause serious distortions.

Before the end of our report, we would like to mention a problem about the Wilcoxon signed rank statistic, that is, the sensitivity of this statistic to nonlinearity may dramatically decrease for certain nonlinear systems. For example, when we apply our test to the Logistic map, although we can still reject the null hypothesis, the rejections, for 1000 tests at the nominal rejection rate of 5%, decrease to $144, 133, 147, 152$ and 166 at the fitting orders of $6, 7, 8, 9$ and 10 respectively, quite low rates compared to the nonlinear cases in Table 1. Indeed, the fact is that it is difficult to find universal scores in Eq.(2) to obtain the most powerful signed rank statistic for all systems (see [10, Theorem 10.1.19.]).

Acknowledgments

This research is supported by a Hong Kong University Grants Council Competitive Earmarked Research Grant (CERG) number PolyU 5235/03E.

References

[1] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge Univ. Press, Cambridge, 1997).

[2] A.R. Osborne and A. Provenzale, *Physica D* 35, 357 (1989).

[3] T. Schreiber and A. Schmitz, *Phys. Rev. E* 55, 5443 (1997).

[4] J. Theiler, S. Eubank, A. Longtin, B. Gaidrikian and J.D. Farmer, *Physica D* 58, 77 (1992); J. Theiler and D. Prichard, *Physica D* 94, 221 (1996).

[5] D.T. Kaplan and L.Glass, *Phys. Rev. Lett.* 68, 427 (1992); D. T. Kaplan, *Physica D* 73, 38 (1994).

[6] J. Theiler and D. Prichard, *Fields Inst. Commun.* 11, 99 (1997).

[7] M. Pourahmadi, *Foundations of Time Series Analysis and Prediction Theory* (Wiley, New York, 2001).

[8] J.M. Dufour, *Econometrica* 52, 209 (1984).

[9] R. Luger, *Journal of Econometrics* 115, 259 (2003).

[10] R.H. Randles and D.A. Wolfe, *Introduction to the Theory of Nonparametric Statistics* (Wiley, New York, 1979).

[11] J.S. Maritz, *Distribution-free Statistical Methods* (Chapman &Hall, 1995).

[12] M. Hénon, *Comm. Math. Phys.* 50, 69 (1976).

[13] O.E. Roessler, *Phys. Lett. A* 57, 397 (1976).

[14] T. Nakamura, X. Luo and M. Small, *submitted*.

[15] A. Galka, *Topics in Nonlinear Time Series Analysis with Implications for EEG Analysis* (World Scientific, 2000).

[16] [3] provides a good study of various discriminating statistics

[17] Here p' is the specified fitting order, and $\{a_{i,j} : j = 0, 1, ..., p'\}$ are the estimated parameters at time i via the criterion of ordinary least squares (OLS), which intends to minimize the forward squared error of the prediction.

[18] This choice is arbitrary. But note that usually very large fitting orders are not preferred for parsimony.

[19] An extensive review of the conventional surrogate algorithms can be found in, for example, [15, chapter 11].