ASYMPTOTICALLY INVARIANT GAUSSIANITY TEST FOR CAUSAL INVERTIBLE TIME SERIES

Roxana Ojeda, Jean-François Cardoso, Eric Moulines

Ecole Nationale Supérieure des Télécommunications (ENST), Département Signal
46, rue Barrault, 75634 Paris Cedex 13, France
Fax: 33 1 45 88 79 35; Email: cardoso@sig.enst.fr

ABSTRACT

This paper introduces a Gaussianity test for causal invertible time series. It is based on a quadratic form in differences between sample means and expected values of certain finite memory nonlinear functions of the estimated innovation sequence. The test has, by construction, an interesting property: under reasonable assumptions on the regularity of the stationary process, it is *asymptotically invariant* with respect to the spectral density of the process. Monte-Carlo experiments are included to illustrate the proposed approach.

1. INTRODUCTION

Because of the central role played by the normal distribution in statistical modeling, it is of importance to be able to test that a time series is Gaussian. Several contributions have considered the problem of testing the Gaussianity of a time series. Frequency domain approaches are mainly based on the property that higher-order cumulants of jointly Gaussian variables are equal to zero [1, 2]. The advantage of this approach is that no specific parametric model is assumed but it requires large sample size in order to obtain reliable estimates of the higher-order spectra cumulants. In the time domain, one typically finds goodness-of-fit tests developed for independent and identically distributed (iid) time series [3] and tests based on quadratic forms in differences between sample means and expected values of certain non-linear functions of the sample [4, 5, 6].

This paper revisits the latter approach applied to causal invertible sequences under reasonable regularity assumptions. We study the convergence of functionals of the (normalized) innovation sequence when estimated by a linear predictor of order k computed from T samples. We extend the results of [7] on the asymptotic rates of the prediction order [7] to obtain limiting distributions of such functionals. By this approach, tests can be developed which are asymptotically *invariant* with respect to the spectral density of the process. This asymptotic invariance constitutes the main theme of the present contribution.

2. PRINCIPLE OF AN INVARIANT NORMALITY TEST

We consider times series $\{X_t\}_{t\in\mathbb{Z}}$ in the form

$$X_t = \sigma \sum_{k=0}^{\infty} \psi_k Z_{t-k}, \quad \psi_0 = 1, \quad \sigma > 0, \tag{1}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of independently and identically distributed normal variables with $E\{Z_t\} = 0$, $E\{{Z_t}^2\} = 1$. For a Gaussian time series, and for any positive integer q, the $q \times 1$ vector

$$\mathbf{Z}_t = [Z_t, Z_{t-1}, \dots, Z_{t-q+1}]^T$$

of normalized innovations is distributed to a $\mathcal{N}(0, I_q)$ normal distribution. Hence, one may construct a Gaussianity test by combining an estimate of the normalized innovation sequence and a test for the *simple* hypothesis that $\mathbf{Z}_t$ has distribution $\mathcal{N}(0, I_q)$. These two issues are considered separately.

Regarding the normality of $\mathbf{Z}_t$, we base our test on functions $g : \mathbb{R}^q \mapsto \mathbb{R}^l$ verifying:

$$\mathbf{Z}_t \sim \mathcal{N}(0, I_q) \quad \Rightarrow \quad E\{g(\mathbf{Z}_t)\} = 0.$$

Typically, one may pick up an arbitrary function from $\mathbb{R}^q$ to $\mathbb{R}^l$ and subtract its mean value under the $\mathcal{N}(0, I_q)$ normal distribution in order to satisfy the above condition.

The proposed approach consists in testing an empirical version of $Eg(\mathbf{Z}_t) = 0$ using normalized estimated residuals. Denote $\hat{\pi}(k)$ a $k \times 1$ vector of linear prediction coefficients. The innovation sequence is estimated by

$$\begin{aligned} \hat{Y}_t &= X_t + \hat{\pi}(k)^T \mathbf{X}_{t-1}(k), \qquad (2) \\ \mathbf{X}_t(k) &= [X_t, X_{t-1}, \dots, X_{t-k+1}]^T. \end{aligned}$$

A normalized innovation sequence $\{\hat{Z}_t\}$ is then obtained by

$$\begin{aligned} \hat{Z}_t &= \hat{\sigma}^{-1} \hat{Y}_t, \qquad (3) \\ \text{where} \quad \hat{\sigma}^2 &= \frac{1}{T-L} \sum_{t=L}^{T-1} \hat{Y}_t^2, \quad L = k + q, \end{aligned}$$

from which a vector sequence

$$\hat{\mathbf{Z}}_t = [\hat{Z}_t, \hat{Z}_{t-1}, \dots, \hat{Z}_{t-q+1}]^T$$

is formed and the $l \times 1$ vector U_T is computed:

$$U_T = \frac{1}{\sqrt{T-L}} \sum_{t=L}^{T-1} g(\hat{\mathbf{Z}}_t), \tag{4}$$

which is an (appropriately normalized) empirical version of $Eg(\mathbf{Z}_t)$. The proposed test statistic, denoted η_T, is constructed in three steps:

1. Compute a k-th order linear predictor $\hat{\pi}(k)$.
2. Estimate and normalize the innovation sequence according to (2) and (3).

Copyright 1997 IEEE

3. Compute the test statistic $\eta_T = U_T^T \Gamma_g^{-1} U_T$ with Γ_g the asymptotic covariance matrix of U_T.

The strength of the residual based approach is that, under reasonable regularity conditions, the asymptotic distribution of U_T does *not* depend on the correlation of $\{X_t\}$ but only on the choice of a particular function $g(\cdot)$, hence the qualification of 'asymptotic invariance'. It follows in particular that for a given function g, matrix Γ_g can be computed once for all, independently of the data set.

To implement such a test, one must specify a procedure for estimating the prediction coefficients, the prediction order k and finally a particular function g. We provide below the detail of the estimation of the linear predictor.

The best (in the mean square sense) linear predictor of order k, denoted $\bar{\pi}(k)$, minimizes the prediction error variance of $E\{(X_{t+1} + \mathbf{X}_t(k)^T \pi(k))^2\}$. Denote the autocovariance coefficients by $r(\tau) = E\{X_0 X_\tau\}$ and denote $R(k)$ the covariance matrix of vector $\mathbf{X}_t(k)$. The vector of coefficients $\bar{\pi}(k)$ is then given by:

$$R(k)\bar{\pi}(k) = -\ E\{X_{t+1}\mathbf{X}_t(k)\}. \qquad (5)$$

The estimated kth order prediction coefficients is denoted $\hat{\pi}(k)$ and is obtained as the solution of $\hat{R}(k)\hat{\pi}(k) = -\hat{r}(k)$, where

$$\hat{R}(k) = (T-k)^{-1} \sum_{t=k}^{T-1} \mathbf{X}_t(k)\mathbf{X}_t(k)^T, \qquad (6)$$

$$\hat{r}(k) = (T-k)^{-1} \sum_{t=k}^{T-1} \mathbf{X}_t(k) X_{t+1}.$$

3. ASYMPTOTIC INVARIANCE

3.1. Assumptions

Our working assumptions will be the following

H 1 *$\Psi(z) = \sum_{k=0}^{\infty} \psi_k z^k$ is bounded and bounded away from zero on the disc $|z| \le 1$.*

As shown by Akutowicz [8], this is equivalent to assuming that $\{Z_t\}$ has a causal representation

$$Z_t = \sigma^{-1} \sum_{k=0}^{\infty} \pi_k X_{t-k}, \qquad \pi_0 = 1, \qquad (7)$$

where $\Pi(z) = \sum_{k=0}^{\infty} \pi_k z^k = 1/\Psi(z)$ is bounded and bounded away from zero on $|z| \le 1$. An important quantity regarding convergence rates is

$$\beta_k = \sum_{l=k+1} \pi_l^2. \qquad (8)$$

H 2 *With probability* 1, *$\hat{R}(k) > 0$ for $T > k$.*

H 3 *The fourth-order moment of Z_t is bounded,* i.e. *$E\{Z_t^4\} < \infty$.*

H 4 *One can find k such that $k = o(\sqrt{T})$ and $\sqrt{T}\ \beta_k = o(1)$.*

This means that convergence results are obtained by assuming some spectral regularity (so that β_k decreases fast enough) and letting the prediction order k increases with the sample size T.

H 5 *Function g is twice continuously differentiable and its second derivatives are uniformly bounded : $|\partial^2 g_u / \partial z_i \partial z_j| \le M < \infty$ for $0 \le i, j \le q-1$, $0 \le u \le l-1$.*

H 6 *The fourth order moments of $g_u(\mathbf{Z}_t)$ $(0 \le u \le l-1)$ and of the partial derivatives of g evaluated at $\mathbf{Z}_t$ are upper bounded: $E\{g_u(\mathbf{Z}_t)^4\} < \infty$ and $E\{(\partial g_u(\mathbf{Z}_t)/\partial z_i)^4\} < \infty$.*

3.2. Main results

The convergence of $\hat{\pi}$ to π can be characterized by defining a $(q-1) \times 1$ vector $\mathbf{s} = [\mathbf{s}_1, \ldots, \mathbf{s}_{q-1}]^T$:

$$\mathbf{s}_i \stackrel{def}{=} \sum_{j=1}^{q-1} \psi_{i-j}(\hat{\pi}_j - \pi_j), \quad i = 1, \ldots, q-1,$$

which admits the following invariant asymptotic equivalent:

Theorem 1 *If* **H1-4** *hold, then, for all $q > 1$:*

$$\mathbf{s} = -\frac{1}{T} \sum_{t=1}^{T} \mathbf{Z}_t(q-1) Z_{t+1} + O_P(\beta_k + kT^{-1}) \qquad (9)$$

where $\mathbf{Z}_t(q-1) = [Z_t, Z_{t-1}, \ldots, Z_{t-q+2}]^T$.

In a next step, we look for an invariant asymptotic equivalent of U_T. Consider the Taylor expansion of g at $\mathbf{Z}_t$:

$$g(\hat{\mathbf{Z}}_t) = g(\mathbf{Z}_t) + Dg(\mathbf{Z}_t)(\hat{\mathbf{Z}}_t - \mathbf{Z}_t) + o_P(\|\hat{\mathbf{Z}}_t - \mathbf{Z}_t\|), \quad (10)$$

and define

$$A_n \stackrel{def}{=} E\{Dg(\mathbf{Z}_t)\mathbf{Z}_{t-n}\} \qquad (11)$$

where the expectation is under the $\mathcal{N}(0, I_q)$ distribution. The modified function:

$$\tilde{g}(\mathbf{Z}_t) = g(\mathbf{Z}_t) - \mathbf{a}^T \mathbf{z}_t,$$
$$\mathbf{a} = [\frac{1}{2}A_0, A_1, \ldots, A_{q-1}]^T,$$
$$\mathbf{z}_t = [Z_t^2 - 1, Z_t Z_{t+1}, Z_{t-1} Z_{t+1}, \ldots, Z_{t-q+2} Z_{t+1}]^T,$$

allows to write an asymptotic equivalent of U_T as stated in this theorem:

Theorem 2 (Asymptotic invariance). *Under* **H1-6**,

$$U_T = \frac{1}{\sqrt{T-L}} \sum_{t=L}^{T-1} \tilde{g}(\mathbf{Z}_t) + o_P(1).$$

This establishes the asymptotic invariance of the test statistic U_T under the null hypothesis that Z_t is a Gaussian time series since this asymptotic equivalent of U_T is a function of the process $\{Z_t\}$ only. It does not depend on the system since the coefficients A_n do not depend on it either but only on function g.

The term $(\frac{1}{2}A_0(Z_t^2 - 1))$ stems from the estimation of σ^2. The other terms of $\mathbf{a}^T \mathbf{z}_t$ appear when the function g depends on $q > 1$ values of the estimated innovation.

Finally, we can establish an asymptotic normality:

Theorem 3 (Asymptotic normality). *Under assumptions* **H1-6**, *the $(l \times 1)$ random vector U_T is asymptotically Gaussian with zero mean and covariance matrix Γ_g: $U_T \to \mathcal{N}(0, \Gamma_g)$ with*

$$\Gamma_g = \sum_{\tau=-q+1}^{q-1} \mathrm{Cov}(\tilde{g}(Z_\tau), \tilde{g}(Z_0)). \qquad (12)$$

Copyright 1997 IEEE

Again, the covariance matrix does not depend on the system but only on function g. The proof of theorems 1 and 2 is in [9]. Theorem 3 is obtained by a central limit theorem for m-dependent processes [10].

4. NUMERICAL EXPERIMENTS

We compute the test statistic $\eta_T = U_T^T \Gamma_g^{-1} U_T$ based on the function $g : \mathbb{R}^q \to \mathbb{R}^l, q = 2, l = 8$ given by:

$$g_i(\hat{\mathbf{Z}}_t) = cos(\lambda_i \hat{\mathbf{Z}}_t) - exp(-\frac{\lambda_i \lambda_i^T}{2}), \quad i = 1, \ldots, l,$$

where λ_i is a $(1 \times q)$ vector of arbitrary coefficients. We define by λ, a $(l \times q)$ matrix which lines are given by λ_i. Function g verifies **H5-6** and we find

$$\tilde{g}_i(\mathbf{Z}_t) = g_i(\mathbf{Z}_t, \lambda_i) - 0.5 A_{0,i}(Z_t^2 - 1) - A_{1,i} Z_t Z_{t+1},$$

where $A_{0,i} = E\{Dg_i(\mathbf{Z}_t)\mathbf{Z}_t\}$ and $A_{1,i} = E\{Dg_i(\mathbf{Z}_t)\mathbf{Z}_{t-1}\}$. The asymptotic covariance matrix of U_T is given by:

$$\begin{aligned} \{\Gamma_g\}_{i,j} &= 0.5\ c_{i,j} \exp^{-0.5(\lambda_i \lambda_i^T + \lambda_j \lambda_j^T)}, \qquad (13) \\ c_{i,j} &= \exp^{-(\lambda_{j,1}\lambda_{i,1} + \lambda_{j,2}\lambda_{i,2})} \\ &+ \exp^{(\lambda_{j,1}\lambda_{i,1} + \lambda_{j,2}\lambda_{i,2})} + \exp^{-(\lambda_{i,1}\lambda_{j,2})} \\ &+ \exp^{(\lambda_{i,1}\lambda_{j,2})} + \exp^{-(\lambda_{j,1}\lambda_{i,2})} \\ &+ \exp^{(\lambda_{j,1}\lambda_{i,2})} - 6 - 2\lambda_{j,1}\lambda_{j,2}\lambda_{i,1}\lambda_{i,2} \\ &- \lambda_i \lambda_i^T \lambda_j \lambda_j^T, \quad i,j = 1, \ldots, l. \end{aligned}$$

Since U_T is a zero-mean asymptotic Gaussian random variable with covariance matrix Γ_g, η_T has a χ^2 distribution of l degrees of freedom.

For each experience, the size of the test is set to 0.05, hence the threshold for a $\chi^2(8)$ distribution is 15.507. We realize N independent trials for different values of T and $\lambda_1 = [1\ 1]$, $\lambda_2 = [1\ 2]$, $\lambda_3 = [1\ 3]$, $\lambda_4 = [2\ 1]$, $\lambda_5 = [2\ 2]$, $\lambda_6 = [2\ 3]$, $\lambda_7 = [3\ 1]$, $\lambda_8 = [3\ 2]$.

4.1. Size of the test

In this experience, we analyse the empirical distribution of the test under the null hypothesis of Gaussianity. The processes under consideration are:

- (M1): a zero-mean Gaussian AR(4) process with poles at 0.9 $exp(\pm j\pi/3)$ and 0.95 $exp(\pm 3j\pi/4)$.
- (M2): a zero-mean Gaussian MA(2) process with zeroes at 0.9 $exp(\pm j\pi/2)$.
- (M3): a zero-mean Gaussian ARMA(2,2) process with poles at 0.9 $exp(\pm 2j\pi/3)$ and zeroes at 0.8 $exp(\pm j\pi/2)$.

Fig. (1) shows an excellent match between the empirical null distribution of the test and the theoretical one. To calculate the empirical distribution of the test, we use the MA model and k=40. Fig. (2) shows the empirical distribution of the test for M2 when k varies. This plot suggests that the null distribution of the test varies 'weakly' with the prediction order when it is taken large enough. Fig. (3) compares the empirical distribution for models M1, M2 and M3. It shows the invariant property of the test under the null hypothesis. In the next experience, we consider a AR(1) process given by:

- (M4): $X_t = aX_{t-1} + Z_t$

where Z_t is a sequence of iid unit-variance zero-mean Gaussian variables. We analyse the empirical size of the test for different values of a.

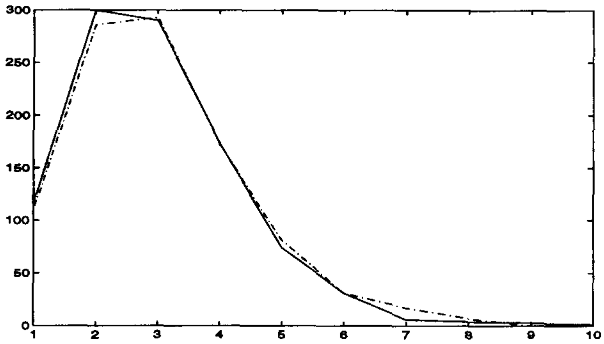

Figure 1. Comparison between theoretical '–' and empirical null '-.' distributions of the test; N=1000; k=40; T=1024.

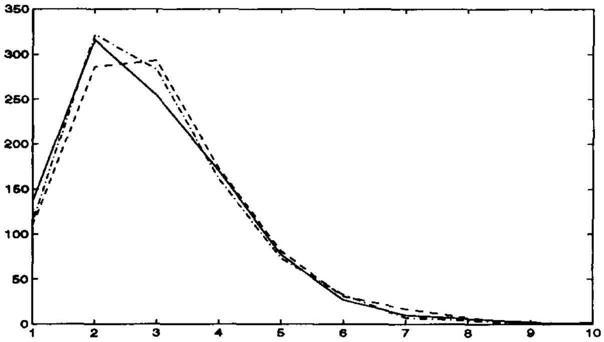

Figure 2. Empirical distribution of the test for M2; k=30 '–', k=35 '-.', k=40 '- -'; N=1000; T=1024.

a	T	$k = 40$	bispectrum	[5]
0.9	256	6.8	42	7
0.9	1024	4.8	21	4
0.5	256	6.0	10	5
0.5	1024	5.6	5	4
0	256	5.6	7	6
0	1024	5.2	5	5
-0.5	256	5.6	6	4
-0.5	1024	5.2	6	5
-0.9	256	5.2	49	5
-0.9	1024	4.8	28	5

Table 1: empirical size (%) of the test for M4; N=250.

Table 1 shows that invariance with respect to the spectral density is reached even with $T = 256$ samples. We compare the results with the bispectrum test (published in [5]) and with [5]. The bispectrum test is inaccurate when the autoregressive coefficient a becomes close to the unit circle, even when the sample size is $T = 1024$. On the other hand, invariant test results are comparable to [5] but our test offers a much easier implementation.

4.2. Power of the test

Non-linear time series.

In this section, we consider four non-linear time series:

- (M5) bilinear: $X_t = 0.7 * X_{t-2}Z_{t-1} + Z_t$,
- (M6) non-linear MA: $X_t = 0.8 * Z_{t-2}Z_{t-1} + Z_t$,
- (M7) extended non-linear MA: $X_t = 0.8 * Z_{t-1} + Z_{t-2} \sum_{j=2}^{20} (0.8)^{j-2} Z_{t-j}$,

Copyright 1997 IEEE

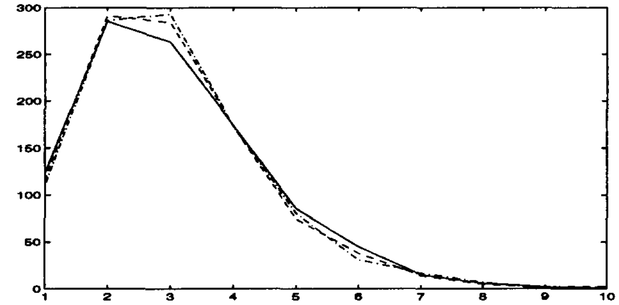

Figure 3. Empirical distribution of the test for M1 '–', M2 '-.' and M3' - -'; N=1000; k=40; T=1024.

- (M8) threshold AR: $X_t = -0.5 * X_{t-1} 1_{X_{t-1}<1} + 0.4 X_{t-1} 1_{X_{t-1}>1} + Z_t$,

where Z_t are iid standardized random Gaussian variables. In fig. (4), we plot the probability of detection versus the probability of false alarm. We compare the results for the nonlinear models with respect to the null hypothesis (M1). The test has no difficulty detecting these nonlinear models.

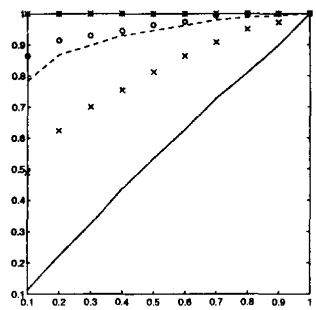

Figure 4. Detection probability against false alarm probability. M1: '–', M5: 'o', M6: '- -', M7: '*', M8: 'x'; N=500; k=40; T=512.

Non-Gaussian linear process.

We analyse an AR(1) non-Gaussian process, with $a = 0.95$ and the following excitation:

- (M9): Z_t is a sequence of iid discrete random variables taking two values $\{+1, -1\}$ with equal probability.

For this process, we obtain an excellent power of the test.

models	T	$k = 40$	[5]
M9	256	100	97
M9	512	100	100
M9	1024	100	100

Table 2: power of the test (%) against non-Gaussian linear process; N=500.

Non-additive contamination.

For this example, we generate two AR(1) models with the same parameter $a = -0.9$ (M10) but with excitations iid $N(0,1)$ or iid $N(0,3)$. The first model is select with probability 0.9.
As in the previous example, the contamination is easily detected.

models	T	$k = 40$	[5]
M10	256	99.6	98
M10	512	100	100
M10	1024	100	100

Table 3: power of the test (%) against non-additive contamination; N=500.

5. CONCLUSION

This article introduced a time domain Gaussianity test for causal invertible linear time series. The test is asymptotically invariant with respect to the spectral density of the process under the linear hypothesis. The principal results were given without demonstrations. The proof of theorems will be made available at the conference. Several numerical experiments have shown the good performance of the test but further characterization is in progress. We plan to extend this approach to the more challenging problem of testing the linearity of a times series.

REFERENCES

[1] T. S. Rao and M. M. Gabr, "A test for linearity of stationary time series," *Journal of Time Series Analysis*, vol. 1, no. 1, pp. 145–158, 1980.

[2] M. J. Hinich, "Testing for Gaussianity and linearity of a stationary time series," *Journal of Time Series Analysis*, vol. 3, no. 3, pp. 169–176, 1982.

[3] R. D. E. Pearson and K. Bowman, "Tests for departure from normality: comparison of powers," *Biometrika*, vol. 64, no. 2, pp. 231–246, 1977.

[4] T. Epps, "Testing that a stationary time series is Gaussian," *Ann. of Stat.*, vol. 15, pp. 185–188, 1987.

[5] E. Moulines and K. Choukri, "Time-domain procedures for testing that a stationary time-series is Gaussian," *IEEE Trans. on Signal Processing*, vol. 44, no. 8, pp. 2010-2025, 1996.

[6] A. Zoubir, "Gaussiantity test for zero-skewed real and complex data," in *Proc. of IEEE Signal Process. Workshop on Higher-Order Statistics* (IEEE, ed.), pp. 327–331, 1993.

[7] K. N. Berk, "Consistent autoregressive spectral estimates," *The Annals of Statistics*, vol. 2, no. 3, pp. 489–502, 1974.

[8] E. J. Akutowicz, "On an explicit formula in linear least squares prediction," *Math. Scand.*, vol. 5, 1957.

[9] R. Ojeda, J.-F. Cardoso and E. Moulines, "Gaussianity test for causal invertible linear time series using the innovation sequence," 1996. (In preparation.)

[10] P. J. Brockwell and R. A. Davis, *Time Series: theory and methods.* Springer-Verlag, 1991.

Copyright 1997 IEEE