Testing for Gaussianity and Linearity of a Stationary Time Series

Melvin J. Hinich
Virginia Tech
October, 1981

Abstract

Stable autoregressive (AR) and autoregressive moving average (ARMA) processes belong to the class of stationary linear time series. A linear time series \( \{x(t) = \sum_{i=1}^{\infty} h(t-i) \varepsilon(s)\} \) is Gaussian if the distribution of the independent innovations \( \{\varepsilon(t)\} \) is normal. Assuming that \( E\varepsilon(t) = 0 \), some of the third order cumulants \( c_{xxx}^{(m,n)} = E[x(t)x(t+m)x(t+n)] \) will be non-zero if the \( \varepsilon(t) \) are not normal and \( E\varepsilon^3(t) \neq 0 \). If the relationship between \( \{x(t)\} \) and \( \{\varepsilon(t)\} \) is non-linear, then \( \{x(t)\} \) is non-Gaussian even if the \( \varepsilon(t) \) are normal. This paper presents a simple estimator of the bispectrum, the Fourier transform of \( \{c_{xxx}^{(m,n)}\} \). This sample bispectrum is used to construct a statistic to test whether the bispectrum of \( \{x(t)\} \) is non-zero. A rejection of the null hypothesis implies a rejection of the hypothesis that \( \{x(t)\} \) is Gaussian. A related test statistic is then presented for testing the hypothesis that \( \{x(t)\} \) is linear. The asymptotic properties of the sample bispectrum are incorporated in these test statistics. The tests are consistent as the sample size \( N \rightarrow \infty \).
Testing for Gaussianity and Linearity of a Stationary Time Series

Melvin J. Hinich
Virginia Tech

1. Introduction

Time series data from a variety of sources are often analyzed under the explicit or implicit assumption that they are generated by autoregressive (AR) or autoregressive moving average (ARMA) processes. These processes are finite parameter linear stationary stochastic processes. The \(t\)th element of a causal linear processes \(\{x(t)\}\) is of the form

\[
x(t) = \sum_{s=0}^{\infty} h(s)e(t-s),
\]

where the \(e(t)\) are independent identically distributed random innovations with \(Ee(t) = 0\). In filter theory terminology, the stationary process \(\{e(t)\}\) is the input to a time invariant linear filter whose impulse response is \(\{h(t): t = 0, 1, \ldots\}\). If \(\sum_{t=0}^{\infty} h^2(t) < \infty\), the covariance function of the stationary output process \(\{x(t)\}\) is finite. If the input is Gaussian, then the output is Gaussian and its covariance function completely determines the joint distributions of the process.

But suppose that the \(e(t)\) are not normal and \(\mu_3 = Ee^3(t) \neq 0\). Then the third order cumulant \(c_{xxx}(m,n) = Ex(t)x(t+m)x(t+n) \neq 0\) for many values of \(m\) and \(n\). The same is true if \(\{x(t)\}\) is generated by a nonlinear filtering operation satisfies a Volterra functional expansion (Brillinger, Sec. 2.10, 1975). Nonlinear models are beginning to play a
role in applied time series. An overview of nonlinear models is given by Priestley (1980).

Using an estimator of the bispectrum of \{x(t)\} Subba Rao and Gabr (1980) present tests for whether the process is Gaussian and whether it is linear. They do not use the asymptotic variance of the sample bispectrum in their multivariate procedures, which is reasonable for smallish sample sizes. This paper presents a modification of their approach that makes heavy use of the large sample properties of the sample bispectrum. Let us begin with a brief review of the bispectrum of a stationary zero-mean process.

2. The Bispectrum

The bispectrum \(B(\omega_1, \omega_2)\) gives a measure of the multiplicative nonlinear interaction of frequency components in \(\{x(t)\}\) (Hasselman et al., 1963). For a real stationary time series with \(E[x(t)] = 0\), the bispectrum is defined as follows:

\[
B_x(\omega_1, \omega_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{xxx}(m,n)\exp[-i(\omega_1 m + \omega_2 n)], \quad (2.1)
\]

assuming that \(|c_{xxx}(m,n)|\) is summable. Given the symmetries of \(B_x(\omega_1, \omega_2)\), its principle domain is the triangular set \(\Omega = \{0 < \omega_1 < \pi, \omega_2 < \omega_1, 2\omega_1 + \omega_2 < 2\pi\}\) (Van Ness 1966). The area of this triangle is 2/3 of the area of the triangle \(\{0 < \omega_1 < \pi, \omega_2 < \omega_1\}\).

If \(\{x(t)\}\) is linear, then

\[
B_x(\omega_1, \omega_2) = \mu_3 H(\omega_1) H(\omega_2) H^*(\omega_1 + \omega_2) \quad (2.2)
\]
where \( H(\omega) = \sum_{t=0}^{\infty} h(t) \exp(-i\omega t) \) is the filter transfer function and the star denotes complex conjugate (Brillinger, 1965). Thus if \( \mu_3 = \text{E} x^3(t) \neq 0 \), then \( B_x(\omega_1, \omega_2) \neq 0 \).

The finite Fourier transform of a sample \( \{x(0), x(1), \ldots, x(N-1)\} \) of the process can be used to construct a consistent estimator of the bispectrum.\(^1\) Let \( \omega_n = 2\pi n/N \) for \( n = 0, 1, \ldots, N-1 \). For each pair of integers \( j \) and \( k \), define

\[
F(j,k) = \frac{1}{N} X(\omega_j) X(\omega_k) X^*(\omega_{j+k}),
\]

(2.3)

where

\[
X(\omega_j) = \sum_{t=0}^{N-1} x(t) \exp(-i\omega_j t).
\]

Since \( X(\omega_j+N) = X(\omega_j) \) and \( X(\omega_{N-j}) = X^*(\omega_j) \), the principal domain of \( F(j,k) \) is the triangular grid \( D = \{0 < j < N/2, 0 < k < j, 2j + k < N\} \) (let \( N \) be even).

Set \( X(0) = 0 \), which is equivalent to subtracting the sample mean from the data. Thus \( F(j,0) = F(0,k) \equiv 0 \). Given a summability condition for the cumulants of \( \{x(t)\} \), it follows from expression (4.3.15) in Brillinger (1975) that

\[
E[F(j,k)] = B_x(\omega_j, \omega_k) + O(N^{-1}),
\]

(2.4)

and the complex variance is

---

\(^1\)Subba Rao–Gabr use the standard windowed sample covariance method.
\[ E|F(j,k) - B_x(\omega_j, \omega_k)|^2 = NS_x(\omega_j)S_x(\omega_k)S_x(\omega_j+k) \]
\[ [1 + \delta(j-k) + \delta(N-2j-k) + 4\delta(N-3j)\delta(N-3k)] + O(1), \]

where \( S_x(\omega) \) is the spectrum of \( \{x(t)\} \) and \( \delta(k) = 0 \) unless \( k = 0 \) when \( \delta(0) = 1 \). The complex covariance between \( F(j,k) \) and \( F(j',k') \) is \( O(N^{-2}) \) for \( j \neq j' \) or \( k \neq k' \) (it is \( O(N^{-1}) \) for the "exclusive or" cases).

There are many ways to average the \( F(j,k) \) to obtain a consistent estimate of the bispectrum on a lattice of points in the set \( D \). For the lattice \( L = \{(2m-1)M/2,(2n-1)M/2:m,n = 1,2,... \text{ and } M = N^c \text{ for } 1/2 < c < 1\} \) in \( D \), a simple approach is to average the \( F(j,k) \) in a square of \( M^2 \) points centered at \( ((2m-1)M/2, (2n-1)M/2) \) if all the \( (j,k) \) are in the domain (Fig. 1). The estimator is then

\[ \hat{B}_x(m,n) = M^{-2} \sum_{j=(m-1)M}^{nM-1} \sum_{k=(n-1)M}^{nM-1} F(j,k) \]

If a square has points outside the set \( D \), those points are not included in the average.

If \( B_x(\omega_1, \omega_2) \) is slowly varying over a square of width \( 2\pi M/N \) and \( S_x(\omega) \) is slowly varying over a \( 2\pi M/N \) band, it follows from (2.4) that

\[ E[\hat{B}_x(m,n)] = B_x(2\pi(2m-1)/M, 2\pi(2n-1)/M) + O(M/N). \]

From (2.5), the complex variance is

\[ \text{Var} \hat{B}_x(m,n) = E|\hat{B}_x(m,n)|^2 - |\hat{B}_x(m,n)|^2 \]
\[ = NM^{-4}O(S_x(2\pi(2m-1)/M) \in S_x(2\pi(2n-1)/M) + O(M/N)) \]
where \( Q \) is the number of \((j, k)\) in the square that are in \( D \) but not on the boundaries \( j = k \) or \( 2j + k = N \), plus twice the number on these boundaries.

If the square is within \( D \), \( Q = M^2 \). For any square, \( Q \geq H^2/8 \) since the smallest set in the domain (a triangle) is for the square centered at \((0, 0)\). Since \( M = N^c \) for \( 1/2 < c < 1 \), \( M/N \to 0 \) and

\[
NM^{-4}Q < NM^{-2} = N^{1-2c} \to 0 \quad \text{as } N \to \infty.
\]

This implies that \( B \) is a consistent estimator of the bispectrum at \((\omega_1, \omega_2)\) as \( N \to \infty \) for the sequence \( \{m(N) = [\omega_1 N^{1-c}], n(N) = [\omega_2 N^{1-c}]\} \), where \([x]\) denotes the integer part of \( x \).

The asymptotic distribution of each estimator is complex normal since the estimator is asymptotically equal to the one analyzed by Van Ness. Applying Theorem 4.4.2 in Brillinger (1975), the estimators are asymptotically independent. Thus from (2.8), the distribution of

\[
X_{m,n} = (N^{1-4c}Q)^{-1/2} \left\{ S_x(2\pi(2m-1)M/2N)S_x(2\pi(2n-1)M/2N) \right. \\
\left. \quad S_x(2\pi(m+n-1)M/N) \right\}^{-1/2} B_x(m,n)
\]

is complex normal with unit variance. Consequently, \( 2|X_{m,n}|^2 \) is approximately chi-square with two degrees of freedom and noncentrality parameter

\[
\lambda_{m,n} = 2(N^{1-4c}Q)^{-1} \gamma_x(m,n) > 2N^2c^{-1} \gamma_x(m,n)
\]

where

\[
\gamma_x(m,n) = S^{-1}(\frac{2\pi (2m-1)M}{2N}) S^{-1}(\frac{2\pi (2n-1)M}{2N}) S^{-1}(\frac{2\pi (m+n-1)M}{N}) \left| B_x(\frac{2\pi (2m-1)M}{2N}, \frac{2\pi (2n-1)M}{2N}) \right|^2.
\]
Moreover, the statistic $S = 2\sum_{(m,n) \in L} |X_{m,n}|^2$ is approximately $\chi^2_P(\lambda)$ where 
\[ \lambda = \sum_{(m,n) \in L} \lambda_{m,n} \quad \text{and} \quad P = \text{the number of } (m,n) \text{ in } L. \] 
Since $n = 1, \ldots, [N/2M]$, $P$ is approximately $N^2/12M^2$.

3. Testing for Gaussianity

The statistic $S$ is basically the Subba Rao-Gabr test statistic with the asymptotic variance-covariance matrix instead of their sample estimate. Under the null hypothesis, $B(\omega_1, \omega_2) = 0$, and thus $S$ is approximately $\chi^2_{2P}(0)$ for large $N$. This suggests the following test statistic: 
\[ \hat{S} = 2\sum_{m,n} |\hat{X}_{m,n}|^2 \] 
where $\hat{X}_{m,n}$ is given by (2.9) with $S_x(\omega)$ replaced by an estimate $\hat{S}_x(\omega)$ for each lattice frequency in the domain. If the spectrum is estimated by averaging $M$ adjacent periodogram ordinates (Fuller, Sec. 7.2, 1976), then $\hat{S}_x(\omega) = S_x(\omega)[1 + (M/N)Y]$ where $Y$ is (approximately) a standard normal variate. It then follows from (2.9) that $\hat{S} = S + O_P(M/N)$, and thus the distribution of $S$ is also approximately $\chi^2_{2P}(\lambda)$ for large $N$. An approximate $\alpha$-level test of the null hypothesis that $B(\omega_1, \omega_2) = 0$ is to reject it if $\hat{S} > t_\alpha$ where $\alpha = Pr(\chi^2_{2P} > t_\alpha)$.

If the null hypothesis is rejected, then the Gaussian assumption must be rejected. If not, then the process may be non-Gaussian but the data is consistent with a zero bispectrum.

I will now show that the test is consistent as $N \to \infty$. For simplicity, suppose that all the squares are in $D$ so that $Q = N^2$. Set
\[ \alpha = (2\pi)^{-2} \int_\Omega [S_x(\omega_1)S_x(\omega_2)S_x(\omega_1 + \omega_2)]^{-1} |B_x(\omega_1, \omega_2)|^2 d\omega_1, d\omega_2; \] \[ (3.1) \] 
where $\Omega$ is the principal domain. From (2.10),
\[ p^{-1} \sum_{(m,n) \in \mathcal{L}} \lambda_{m,n} = 2N^2c^{-1} + O(N^{c-1}) \] 
\[ (3.2) \]

using the integral approximation of a sum.

This approximation of \( \lambda/P \) will now be used to obtain the large sample property of the power of the chi-square test. For large \( N \) and thus large \( P \),
\[ (2P)^{-1/2} \chi^2_{2P}(\lambda) \] is approximately normal with mean \( 1+(2P)^{-1/2} \lambda \) and variance \( (2P)^{-2}(4P+4\lambda) \). Thus the large sample power of the test is a monotonically increasing function of the "signal-to-noise" ratio
\[ \frac{P + \lambda/2}{(P+\lambda)^{1/2}} = [P(\lambda/4P)]^{1/2} \]
\[ = (aN/24)^{1/2} \] 
\[ (3.3) \]
since \( P = N^{2(1-c)}/12 \). Thus the test is consistent. Moreover, the null hypothesis will be correctly rejected with probability near one if
\[ (aN/24)^{1/2} > 4. \] If \( a = 1 \), for example, then a sample size of \( N > 384 \) is needed to achieve this somewhat conservative bound for high power.

4. Testing for Linearity

If \( \{x(t)\} \) is a linear process, then \( S_x(\omega) = |H(\omega)|^2 \sigma^2 \) where
\[ \sigma^2 = \varepsilon^2(t). \] Thus from (2.2),
\[ Y_x(m,n) \equiv \sigma^2 \mu_j \] 
\[ (4.1) \]
for all \( (m,n) \in \mathcal{L} \). From (2.10),
\[ \hat{\chi} = p^{-1} \sum_{(m,n) \in \mathcal{L}} (|x_{m,n}|^2 - 1) \quad (4.2) \]

is a consistent estimator of \( N^{2c-1} \gamma_x \) under the null hypothesis that \( \{x(t)\} \) is linear. It then follows from the results in Section 2 that

\[ \hat{S} = n^{-1/2} \sum_{(m,n) \in \mathcal{L}} (|x_{m,n}|^2 - 1 - \hat{\chi}) \quad (4.3) \]

is approximately normal \( N(0, \gamma_x/6) \) under the null hypothesis. This statistic with \( \hat{S}_x \) instead of \( S_x \) can be used to test for linearity. The consistency of this test follows from the large sample analysis used to show the consistency of the \( S \) test.

5. Conclusion

Simple tests for Gaussianity and linearity of a time series have been presented. The large sample variance and covariance of the asymptotically normal bispectrum estimator are used to simplify the Subba Rao–Gabr test statistics. The asymptotic properties of the statistic for testing for a zero bispectrum have been presented. The power of this test is high when \( aN \) is large.
References


The Lattice in the Principal Domain

Figure 1
**REPORT DOCUMENTATION PAGE**

<table>
<thead>
<tr>
<th>Element</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report Number</td>
<td>Technical Report 23</td>
</tr>
<tr>
<td>Title</td>
<td>Testing for Gaussianity and Linearity of a Stationary Time Series</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Melvin J. Hinich</td>
</tr>
<tr>
<td>Performing Organization Name and Address</td>
<td>Virginia Tech Department of Economics Blacksburg, VA 24061</td>
</tr>
<tr>
<td>Controlling Office Name and Address</td>
<td></td>
</tr>
<tr>
<td>Report Date</td>
<td>October 1981</td>
</tr>
<tr>
<td>Number of Pages</td>
<td>10</td>
</tr>
<tr>
<td>Security Class (of this report)</td>
<td>Unclassified</td>
</tr>
<tr>
<td>Distribution Statement (of this Report)</td>
<td>Distribution of this document is unlimited</td>
</tr>
<tr>
<td>Keywords</td>
<td>Bispectrum, Gaussian, Linear Process, Time Series, Asymptotic Variance</td>
</tr>
<tr>
<td>Abstract</td>
<td></td>
</tr>
</tbody>
</table>

**DD Form 1473 Edition of 1 Nov 65 IS OBSOLETE**

S N 0102 LF 014 6601
END
DATE
FILMED

19-82

DTIC