Modelling Non-linear and Non-stationary Time Series

Chapter 4: Cumulants and polyspectra

Henrik Madsen

Advanced Time Series Analysis

September 2016

Contents

Wolds' Theorem

Wold's Theorem: Let X_t be a zero mean second order stationary process. Then X_t can be expressed in the form

$$X_t = U_t + V_t$$

where

- $\{U_t\}$ and $\{V_t\}$ are uncorrelated processes.
- $\{U_t\}$ is non-deterministic with a one-sided representation $U_t = \sum_{i=0}^{\infty} a_i \eta_{t-i}$ where $a_0 = 1$ and $\sum_{i=1}^{\infty} a_i^2 < \infty$. Besides $\{\eta_t\}$ is an uncorrelated sequence and uncorrelated with $\{V_t\}$.
- The sequences $\{a_i\}$ and $\{\eta_t\}$ are uniquely determined.
- $\{V_t\}$ is deterministic.

Gaussianity and Linearity

Therefore any Guassian process conforms to a linear process. Note that the converse is not necessarily true, i.e. not every linear process is Gaussian.

Nevertheless, the theorem reveals a strong connection between linearity and Gaussianity.

Joint Cumulants

Joint Cumulants: The k'th order joint cumulant for random variables $X_1, X_2, ..., X_k$ is defined by

$$C\{X_1, X_2, ..., X_k\} = \sum_{\nu} (-1)^{p-1}(p-1)! \mu_{\nu_1} \mu_{\nu_2} ... \mu_{\nu_p}$$

where μ_{ν_i} is the mean of the products of X's corresponding to the partition ν_i.

Remark:

$$C\{X_1, X_2\} = cov\{X_1, X_2\}$$

Moment and Cumulant Generating Functions

Moment Generating Function: The moment generating function for the random variables $X_1, X_2, ..., X_k$ is defined by

$$M(\theta_1, ..., \theta_k) = E\{\exp(\theta_1 X_1 + ... + \theta_k X_k)\}$$

Cumulant Generating Function: The cumulant generating function $K(\theta_1, ..., \theta_k)$ is defined by

$$K(\theta_1, ..., \theta_k) = \log M(\theta_1, ..., \theta_k)$$

Some tedious (but straight forward) algebra shows that

$$\frac{\partial}{\partial \theta_1} \cdots \frac{\partial}{\partial \theta_k} M(\theta_1, ..., \theta_k)|_{\theta_i = 0} = E\{X_1 ... X_k\}$$

and

$$\frac{\partial}{\partial \theta_1} \cdots \frac{\partial}{\partial \theta_k} K(\theta_1, ..., \theta_k)|_{\theta_i = 0} = C\{X_1 ... X_k\}$$

Joint Cumulants for a Stationary Process

- Consider a **stationary stochastic process** $\{X_t\}$ and define

$$C(\tau_1, \tau_2, \dots, \tau_{k-1}) = C\{X_t, X_{t+\tau_1}, X_{t+\tau_2}, \dots, X_{t+\tau_{k-1}}\}$$

- This is the generalization of the autocovariance function.

Poly-spectrum for a Stationary Process

Now the k-th order poly-spectrum for the process $\{X_t\}$ denoted by $f_X(\omega_1, \omega_2, ..., \omega_{k-1})$ is defined by the Fourier transform of $C\{X_t, X_{t+\tau_1}, X_{t+\tau_2}, ..., X_{t+\tau_{k-1}}\}$, i.e.

$$f_X(\omega_1, \omega_2, ..., \omega_{k-1}) =$$
$$\left(\frac{1}{2\pi}\right)^{k-1} \sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_{k-1}=-\infty}^{\infty}$$
$$C(\tau_1, \tau_2, ..., \tau_{k-1}) \exp\{-i(\omega_1\tau_1 + \omega_2\tau_2 + ... + \omega_{k-1}\tau_{k-1})\}$$

Poly-Spectrum as a Measure of Non-Gaussianity

Consider a stationary Gaussian stochastic process $\{X_t\}$. Then all of its poly-spectra of order higher than 2 vanish.

Proof: Assume that the random varibles $X_1, X_2, ..., X_k$, $k \geq 2$, have a Gaussian joint distribution with mean $\underline{m}$ and covariance Σ. The characteristic function of the random varibles is found to be

$$\exp\{i\underline{m}^T\underline{t} - \frac{1}{2}\underline{t}^T\Sigma\underline{t}\}$$

Thus the moment generating function is given by

$$\exp\{\underline{m}^T\underline{\theta} + \frac{1}{2}\underline{\theta}^T\Sigma\underline{\theta}\}$$

This shows that the cumulant generating function is quadratic in θ and its derivatives with order higher than two vanish. This completes the proof.

Bispectrum
Bispectrum and Gaussianity

The third order poly-spectrum $f(\omega_1, \omega_2)$ of a stationary stochastic process is called bispectrum.

The conclusion is that the bispectrum of a stationary Gaussian process is identically zero. This provides a basis for **test for Gaussianity**.

Bispectra and Linearity

Suppose that the two stationary processes $\{X_t\}$ and $\{Y_t\}$ are related through

$$X_t = \sum_{i=0}^{\infty} h_i Y_{t-i}$$

Then

$$f_X(\omega_1, \omega_2) = H(-\omega_1 - \omega_2) H(\omega_1) H(\omega_2) f_Y(\omega_1, \omega_2)$$

where $H(\omega)$ is the transfer function of the filter with impulse response $\{h_i\}$.

Proof: Follows from the definition of bispectrum.

Bispectra and Linearity (cont.)

Now assume that a stationary process $\{X_t\}$ has the linear representation

$$X_t = \sum_{i=-\infty}^{\infty} a_i \epsilon_{t-i}$$

where $\{\epsilon_t\}$ is a sequence of independent, zero mean, identically distributed random variables. Denoting the third moment of ϵ_t by μ_3, we have

$$(\frac{\mu_3}{2\pi})^2 = \frac{f_X(\omega_1, \omega_2)}{H(\omega_1)H(\omega_2)H(-\omega_1 - \omega_2)}$$

and denoting the second moment of ϵ_t by μ_2,

$$f_X(\omega) = \frac{\mu_2}{2\pi}|H(\omega)|^2$$

Bispectra and Linearity (cont.)

- The result is that under the linearity hypothesis, the quantity

$$\frac{|f_X(\omega_1, \omega_2)|^2}{|f_X(-\omega_1 - \omega_2) f_X(\omega_1) f_X(\omega_2)|}$$

  is a constant for all ω_1 and ω_2.
- This provides a basis for tests in linearity. Furthermore, if the process is Gaussian, this constant is identically zero for all frequency pairs.
- Therefore, the estimation of the bispectrum from observed data is a fundamental step towards testing in linearity and Gaussianity.

Example: Bispectrum of Non-Gaussian ARMA(1,1) Proces

- Consider the ARMA(1,1) process

$$X_t + aX_{t-1} = \epsilon_t + b\epsilon_{t-1} \tag{1}$$

  where $\{\epsilon_t\}$ is a sequence of identically distributed independent $\Gamma(k, \beta)$-distributed random variables.
- The frequency response function is easily found as (see Madsen (2008))

$$H(\omega) = \frac{1 + be^{-i\omega}}{1 + ae^{-i\omega}} \tag{2}$$

Non-Gaussian ARMA(1,1) (cont.)

When we include the above in equation on a previous slide we get

$$f_X(\omega_1, \omega_2) = f_\epsilon(\omega_1, \omega_2)\frac{P(b, \omega_1, \omega_2)}{P(a, \omega_1, \omega_2)} \tag{3}$$

where the real part is given by

$$\begin{aligned} \mathrm{Re}[P(x, \omega_1, \omega_2)] &= 1 + x^3 + (x^2 + x)\cos(\omega_1) + & (4) \\ &\quad (x^2 + x)\cos(\omega_2) + (x^2 + x)\cos(\omega_1 + \omega_2) & (5) \end{aligned}$$

and the imaginary part by

$$\begin{aligned} \mathrm{Im}[P(x, \omega_1, \omega_2)] &= (x^2 - x)\sin(\omega_1) + (x^2 - x)\sin(\omega_2) + \\ &\quad (x - x^2)\sin(\omega_1 + \omega_2). \end{aligned}$$

Non-Gaussian ARMA(1,1) (cont.)

As $\{\epsilon_t\}$ is a sequence of independent random variables it follows directly from the definition of the bispectrum that

$$f_\epsilon(\omega_1, \omega_2) = \frac{\mu_3}{(2\pi)^2}, \tag{6}$$

where $\mu_3 = m(0,0)$. Since ϵ_t is Gamma distributed

$$f_t(\omega_1, \omega_2) = \frac{2k\beta^3}{(2\pi)^2} \tag{7}$$

since $\mu_3 = 2k\beta^3$ for $X \sim \Gamma(k, \beta)$.
The theoretical bispectrum for the ARMA(1,1) process given by $a = 0.5$ and $b = -0.3$ is shown on the next slides.

Real part of bispectrum

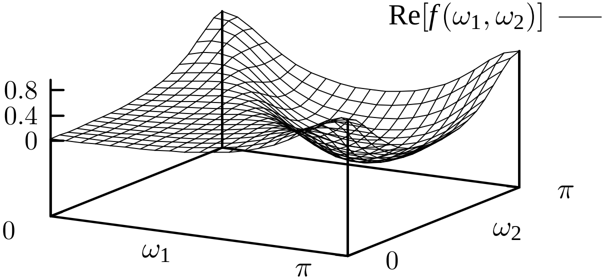


Figure : Real part of the theoretical bispectrum for an ARMA(1,1) process with Gamma distributed noise

Imaginary part of bispectrum

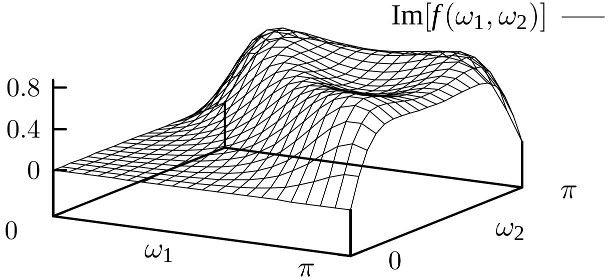


Figure : Imaginary part of the theoretical bispectrum for an ARMA(1,1) process with Gamma distributed noise

Estimation of third order order cumulant

Let $X_1, X_2, ..., X_N$ be realizations from a stationary (at least up to third order) process $\{X_t\}$ with third central moment $C(\tau_1, \tau_2)$. The natural estimate of $C(\tau_1, \tau_2)$ will be

$$\hat{C}(\tau_1, \tau_2) = \frac{1}{N} \sum_{t=1}^{N-\gamma}$$
$$(X_t - \bar{X})(X_{t+\tau_1} - \bar{X})(X_{t+\tau_2} - \bar{X})$$
$$\tau_1, \tau_2 > 0$$
$$\gamma = \max(0, \tau_1, \tau_2)$$

Estimation of third order periodogram
Raw Bispectrum

Then from the definition of bispectrum, we introduce the estimate $I(\omega_1, \omega_2)$:

$$I(\omega_1, \omega_2) = \frac{1}{N(2\pi)^2}\{\textstyle\sum_1^N (X_t - \bar{X}) \exp(-i\omega_1 t)\}$$
$$\times\{\textstyle\sum_1^N (X_t - \bar{X}) \exp(-i\omega_2 t)\}$$
$$\times\{\textstyle\sum_1^N (X_t - \bar{X}) \exp(-i(\omega_1 + \omega_2)t)\}$$

This estimate is often called the third order periodogram. Exactly as in the periodogram case, this estimate is asymptotically unbiased but not consistent. Thus, it must be smoothed by choosing some suitable window.

Consistent estimates of bispectra (Summary)

- See Madsen (2008) Table 7.2 on page 199. Choose a window and set $M = 1$ to obtain $\lambda_0(\cdot)$.
- Compute the three dimensional window $\lambda_0(\tau_1, \tau_2) = \lambda_0(\tau_1)\lambda_0(\tau_2)\lambda_0(\tau_1 - \tau_2)$. Choose M, noting that
  - If it is possible, M should be less than the square root of the sample size, N.
  - It should be smaller than the value of M used in estimating the spectral density.
  - Once for some selection of kernel X, the optimal bandwidth M_X is found the M value for others, say, Parzen window M_{Par} is calculated according to

$$M_{Par} = \frac{M_{R,Parzen}}{M_{R,X}} M_X \tag{8}$$

    where $M_{R,\cdot}$ denotes the value of M_R for the corresponding window selection.

Example: Estimation of Bispectra

- Consider a time series of length $N = 500$ which is generated by the ARMA(1,1) process with Gamma distributed noise considered in the previous theoretical example.
- Figure 4 and 6 show the estimated bispectrum with $M = 10$ and $M = 18$, respectively. In both cases a Parzen window is used.

Example: Estimation of Bispectra

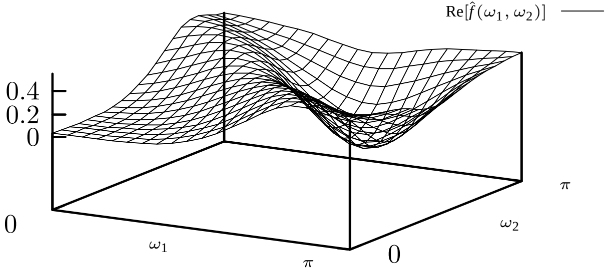


Figure : Real part of estimated bispectrum, Parzen window, $N = 500$, $M = 10$

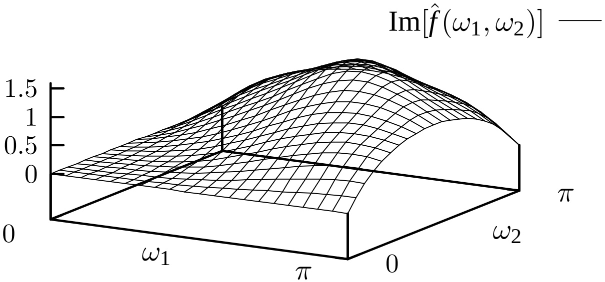


Figure : Imaginary part of estimated bispectrum, Parzen window, $N = 500$, $M = 10$

Example: Estimation of Bispectra

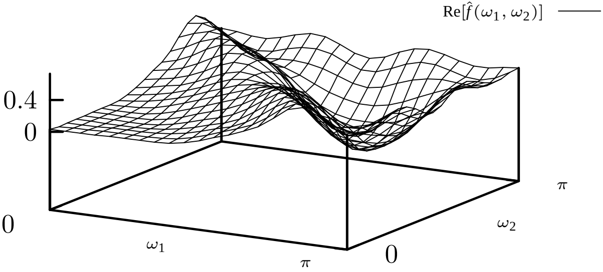


Figure : Real part of estimated bispectrum, Parzen window, $N = 500$, $M = 18$

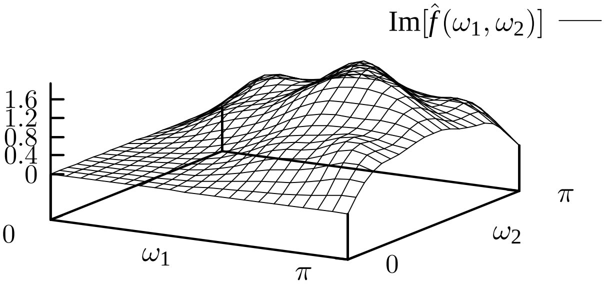


Figure : Imaginary part of estimated bispectrum, Parzen window, $N = 500$, $M = 18$

Parametric estimates

This concludes our discussion of *non-parametric methods* for estimating the bispectrum of a process. If the data generating mechanism is known *a priori*, eg. a bilinear model, it is also possible to explicitly compute the bispectrum for that model and evaluate the result for some estimated parameter values. This is a *parametric method*.

Tests using the bispectrum for Gaussianity

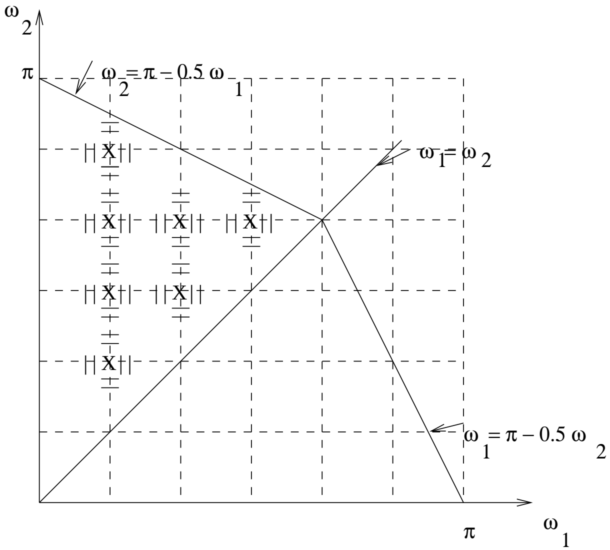
Here we have to test the null hypothesis H_0: $f(\omega_i, \omega_j) = 0$ for all ω_i, ω_j. However, as a result of the following symmetry relations:

$$
\begin{aligned}
f(\omega_1, \omega_2) &= f(\omega_2, \omega_1) && (9)\\
&= f(\omega_1, -\omega_1 - \omega_2) && (10)\\
&= f(-\omega_1 - \omega_2, \omega_2) && (11)\\
&= f^*(\omega_2, \omega_1) && (12)
\end{aligned}
$$

it will be enough to evaluate the bispectrum at the points confined to the area with boundaries:

$$
\begin{cases}
\omega_2 - \omega_1 = 0 \\
\omega_2 + \frac{1}{2}\omega_1 = \pi \\
\omega_1 = 0
\end{cases}
$$

Tests for Gaussianity



Tests for Gaussianity

Our aim is to estimate a (complex) $P \times 1$ vector $\boldsymbol{\eta}$ where the i'th element is the bispectrum of the i'th node. In order to estimate $\boldsymbol{\eta}$, we estimate the bispectrum for the points corresponding to fine grid (see Figure 7). These estimates are then gathered in n vectors (each vector having P elements), each denoted by $\boldsymbol{\xi}_{(i)}$, $i = 1, \ldots, n$. Now estimate:

$$\hat{\boldsymbol{\eta}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_{(i)} \quad (13)$$

$$\boldsymbol{A} = \sum_{i=1}^{n} [\xi_{(i)} - \hat{\eta}][\xi_{(i)} - \hat{\eta}]^T \quad (14)$$

$$\hat{\boldsymbol{\Sigma}}_{\xi} = \frac{\boldsymbol{A}}{n} \quad (15)$$

Test for Gaussianity

Under the null hypothesis that $f(\omega_i, \omega_j) = 0$ for all ω_i, ω_j, the statistic F_1 defined by

$$F_1 = \frac{2(n-p)}{2p} T^2 \quad (16)$$

$$T^2 = n\hat{\boldsymbol{\eta}}^T \boldsymbol{A}^{-1} \hat{\boldsymbol{\eta}} \quad (17)$$

is distributed as a central F distribution with $(2P, 2(n-p))$ degrees of freedom. The latter is the complex generalization of Hotelling's T^2 test (see **??**).

Test for Linearity

This test is very similar to the previous one, with the following modifications. The vector $\boldsymbol{\xi}_{(i)}$ is modified to $\boldsymbol{Y}_{(i)}$. The elements of the vector are now

$$\frac{|f(\omega_i,\omega_j)|^2}{f(\omega_i)f(\omega_j)f(\omega_i+\omega_j)} \tag{18}$$

instead of $f(\omega_i,\omega_j)$. Denote the counter-part of the vector $\boldsymbol{\eta}$ by $\boldsymbol{Z}$. Now estimate:

$$\hat{\boldsymbol{Z}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Y}_{(i)} \tag{19}$$

$$\boldsymbol{S} = \sum_{i=1}^{n}(\boldsymbol{Y}_{(i)}-\hat{\boldsymbol{Z}})(Y_{(i)}-\hat{\boldsymbol{Z}})^T \tag{20}$$

$$\hat{\boldsymbol{\Sigma}}_Y = \frac{\boldsymbol{S}}{n} \tag{21}$$

Test for Linearity

Define the statistic F_2 by

$$
\begin{aligned}
F_2 &= \frac{n-Q}{Q} T^2 && (22)\\
T^2 &= n\boldsymbol{\beta}^T \boldsymbol{S}_B^{-1} \boldsymbol{\beta} && (23)\\
\boldsymbol{\beta} &= \boldsymbol{B}\hat{\boldsymbol{Z}} && (24)\\
\boldsymbol{S}_B &= \boldsymbol{B}\hat{\boldsymbol{\Sigma}}_Y \boldsymbol{B}^T && (25)
\end{aligned}
$$

where $Q = P - 1$ and $\boldsymbol{B}$ is a $Q \times P$ matrix defined by

$$
\begin{bmatrix}
1 & -1 & 0 & \dots & 0 \\
0 & 1 & -1 & \dots & 0 \\
 & & \vdots & & \\
0 & 0 & \dots & 1 & -1
\end{bmatrix}
$$

Under the null hypothesis, the statistic F_2 is F distributed with $(Q, n-Q)$ degrees of freedom.