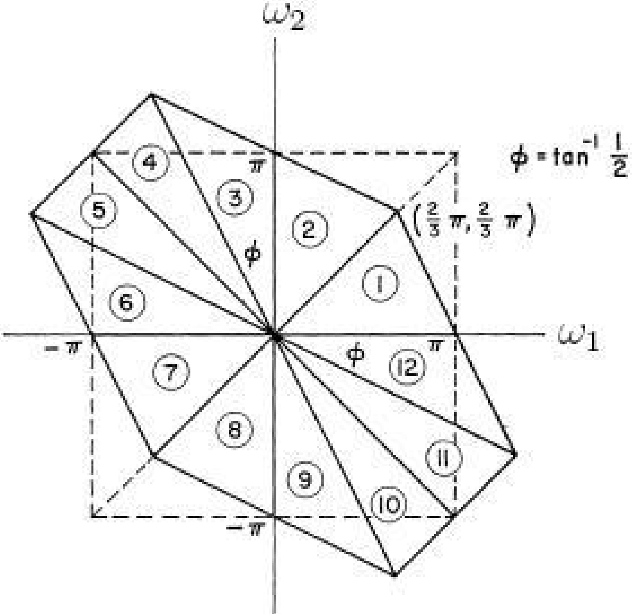
A Bootstrapped Test of Linearity



Linear Time Series

Definition (linear time series)

A time series is **linear** if it has the representation

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \qquad (\star)$$

where $\varepsilon_t \overset{\text{iid}}{\sim} (0, \sigma^2)$ and $\sum |\psi_j| < \infty$. (So X_t is $MA(\infty)$ with **iid** residuals.)

Example

- ARMA processes are linear: $\phi(B)X_t = \theta(B)\varepsilon_t$
- I(d) process with $\frac{-1}{2} < d < \frac{1}{2}$ is linear
- GARCH(α,β) process is not linear

Wold Decomposition

Wold Decomposition (1938)

Any zero-mean covariance-stationary process X_t can be represented in the form

$$X_t = \underbrace{\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}}_{\text{linearly indeterministic}} + \underbrace{\kappa_t}_{\text{linearly deterministic}}$$

where $\sum_{j=1}^{\infty} \psi_j^2 < \infty$ and $\boldsymbol{\varepsilon_t}$ **is white noise**.

Example

$$X_t = \underbrace{A \sin\left(\frac{2\pi t}{12} + \lambda\right)}_{\text{linearly deterministic}} + \underbrace{\left[\frac{\theta(B)}{\phi(B)}\right] \varepsilon_t}_{\text{ARMA}}$$

where A is constant, $\lambda \sim \text{unif}[-\pi, \pi]$, and $\left[\frac{\theta(B)}{\phi(B)}\right] \varepsilon_t$ is a causal and invertible ARMA process.

Spectral Domain

Let X_t be covariance-stationary with mean μ and autocovariance function

$$\gamma(h) = \operatorname{cov}(X_t, X_{t+h})$$

Up to the first two moments, all information about X_t is contained in $\gamma(h)$.

Since $\gamma(h)$ is non-negative definite, Herglotz's theorem (1911) guarantees the unique existence of a right-continuous, non-decreasing, bounded function on $[-\pi, \pi]$ with $F(-\pi) = 0$ such that

$$\gamma(h) = \int_{(-\pi,\pi]} e^{-ih\omega}\, dF(\omega) \qquad \forall h = 0, \pm 1, \ldots$$

Theorem (Lebesgue Decomposition Theorem)

F can has a decomposition into an absolutely continuous component (with density) and a singular component; i.e.,

$$F = F_{ac} + F_s$$

Spectral Density

Definition (spectral density)

If the spectral distribution function F has the representation

$$F(\omega) = \int_{-\pi}^{\omega} f(\lambda)\,d\lambda \qquad -\pi \leq \omega \leq \pi$$

then f is called the spectral density.

Wold decomposition: f exists if $\kappa_t = 0$

Lebesgue decomposition: f exists if $F_s = 0$

Fourier transform theory

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\omega} f(\omega)\,d\omega \quad \Longleftrightarrow \quad f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \gamma(h)$$

Central Limit Theorems

Theorem (CLT for linear time series)

Let X_t be a **linear time series** *with the representation*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$$

where ε_t are iid$(0, \sigma^2)$ and $\sum |\psi_j| < \infty$, then

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sum_{h=-\infty}^{\infty} \gamma(h)\right)$$

Generalizations:

- α-mixing

  $$\alpha(s) \equiv \sup \left\{ |P(A \cap B) - P(A)P(B)| : -\infty < t < \infty, A \in X_{-\infty}^{t}, B \in X_{t+s}^{\infty} \right\}$$

- martingale difference sequences

  $$\mathbb{E}[\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots] = 0$$

The Asymptotic Variance

To construct a confidence interval for $\bar{X}$, we need to estimate $\sum_{h=-\infty}^{\infty} \gamma(h)$.

Recall:

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h)$$

When $\lambda = 0$,

$$f(0) = \sum_{h=-\infty}^{\infty} \gamma(h)$$

So we need an estimate of the spectral density at $\lambda = 0$.

Autocovariance Estimation

Let $\{x_t\}_{t=1}^{n}$ be the observation of a covariance-stationary time series. The autocovariance function

$$\gamma(h) = \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)]$$

is naturally estimated by

$$\begin{aligned}\hat{\gamma}(h) &= \begin{cases} \frac{1}{n}\sum_{t=1}^{n-|h|}(x_t - \bar{x})(x_{t+|h|} - \bar{x}) & |h| < n \\ 0 & |h| \geq n \end{cases} \\ &= \frac{1}{n}\sum_{t=1}^{n-|h|}(x_t - \bar{x})(x_{t+|h|} - \bar{x})\end{aligned}$$

And also by the “unbiased” version

$$\hat{\hat{\gamma}}(h) = \left(\frac{n}{n-|h|}\right)\hat{\gamma}(h) = \frac{1}{n-|h|}\sum_{t=1}^{n-|h|}(x_t - \bar{x})(x_{t+|h|} - \bar{x})$$

Both estimates of $\gamma(h)$ are optimal in the sense that they are $\sqrt{n}$–consistent.

Spectral Density Estimation

The spectral density

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega}.$$

cannot be consistently estimated with

$$\tilde{f}(\omega) = I(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \widehat{\gamma}(h) e^{-ih\omega}.$$

Why? $\hat{\gamma}(h)$ is consistent for *fixed* h, but it is not suitable for values close to n.

The fix? Truncate the sum, or more generally, weight the values of $\hat{\gamma}(h)$ with a lag-window function λ.

Bispectrum

- $X_1, \ldots, X_n$ second-order stationary time series means

$$\gamma(h) = \operatorname{cov}(X_t, X_{t+h}) \text{ is free of } t.$$

- Third-order stationary means

$$\gamma(\tau_1, \tau_2) = \mathbb{E}\left[(X_t - \mu)(X_{t+\tau_1} - \mu)(X_{t+\tau_2} - \mu)\right] \text{ is free of t .}$$

- Fourth and higher-order stationary means the *joint cumulant* is free of t.

The spectral density, or spectrum, is the Fourier transform of $C(j)$, i.e.

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega}.$$

Likewise, the bispectral density, or bispectrum, is

$$f(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \gamma(\tau_1, \tau_2) e^{-i\tau_1\omega_1 - i\tau_2\omega_2}.$$

Spectral and Bispectral Estimation

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega} \Leftrightarrow \widehat{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \lambda\left(b\,h\right) \widehat{\gamma}(h) e^{-ih\omega}$$

$$f(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \gamma(\tau_1, \tau_2) e^{-i\tau_1\omega_1 - i\tau_2\omega_2}.$$

$$\Updownarrow$$

$$\widehat{f}(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \lambda\left(b\,\tau_1, b\,\tau_2\right) \widehat{\gamma}(\tau_1, \tau_2) e^{-i\tau_1\omega_1 - i\tau_2\omega_2}$$

Definition (Lag-window Order)

λ has order p if $\lambda(\boldsymbol{\tau}) = 1 + a\|\boldsymbol{\tau}\|^p + O(\|\boldsymbol{\tau}\|^p)$ as $\|\boldsymbol{\tau}\| \to 0$.

Lag-Window Examples

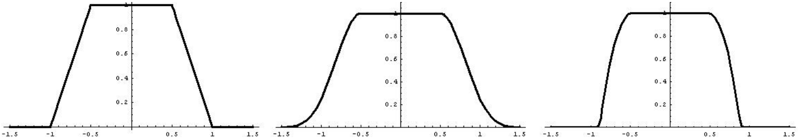


Figure: One Dimensional Flat-top Lag-Windows: $\lambda(bh)$

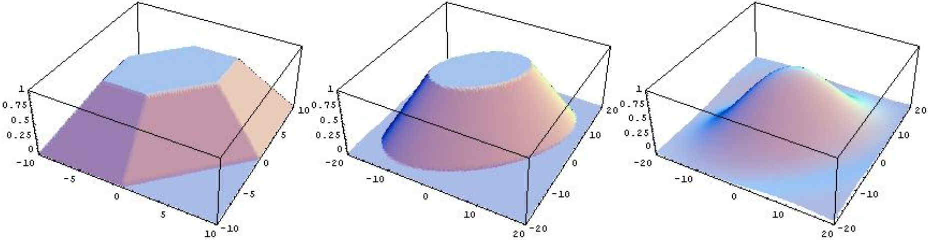


Figure: Two Dimensional Lag-Windows: $\lambda\left(b\,\tau_1, b\,\tau_2\right)$

Spectrum of a Linear Time Series

Let X_t be a linear time series with representation

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$$

and define

$$H(\omega) = \sum_{u=-\infty}^{\infty} a_u e^{-iu\omega}.$$

Then the spectral density of X_t is

$$f(\omega) = \frac{\sigma^2}{2\pi} |H(\omega)|^2$$

and bispectral density is

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

Normalized Bispectrum

Definition (normalized bispectrum)

$$\begin{aligned}\phi(\omega_1,\omega_2) &\triangleq \frac{|f(\omega_1,\omega_2)|^2}{f(\omega_1)f(\omega_2)f(\omega_1+\omega_2)}\\ &\overset{\text{linearity}}{=} \frac{\left(\frac{\mu_3'}{(2\pi)^2}H(-\omega_1-\omega_2)H(\omega_1)H(\omega_2)\right)^2}{\left(\frac{\sigma^2}{2\pi}|H(\omega_1)|^2\right)\left(\frac{\sigma^2}{2\pi}|H(\omega_2)|^2\right)\left(\frac{\sigma^2}{2\pi}|H(\omega_1+\omega_2)|^2\right)}\\ &= \frac{(\mu_3')^2}{2\pi\sigma^6}\\ &\overset{\text{Gaussianity}}{=} 0\end{aligned}$$

This motivates many statistical tests:

Gaussianity	test if the normalized bispectrum is $\equiv 0$
linearity	test if the normalized bispectrum is constant

Symmetries of the Bispectrum

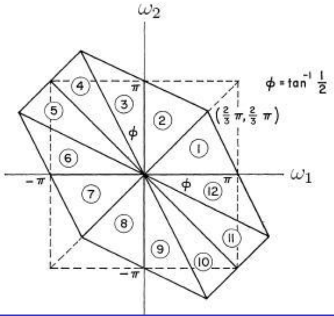
Symmetries of the ACF

$$\gamma(\tau_1,\tau_2)=\gamma(\tau_2,\tau_1)=\gamma(-\tau_1,\tau_2-\tau_1)=\gamma(\tau_1-\tau_2,-\tau_2)$$

induce symmetries in the bispectrum (overline represents complex conjugation)

$$f(\omega_1,\omega_2)=f(\omega_2,\omega_1)=f(\omega_1,-\omega_1-\omega_2)=\overline{f(-\omega_1,-\omega_2)}$$

These symmetries along with being doubly-periodic allow the bispectrum to be completely specified in any one of the twelve regions



A Group Representation

Define $\psi : \mathbb{R}^3 \to \mathbb{R}^2$ given by $\psi(a,b,c) \mapsto (b-a, c-a)$. Take $\sigma = (12)$, then

$$(x,y) \longrightarrow (0,x,y) \xrightarrow{\sigma} (x,0,y) \xrightarrow{\psi} (-x, y-x) \longrightarrow \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix}$$

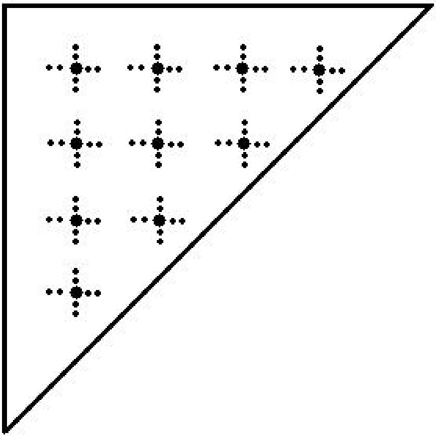
$$\begin{array}{llll}
e \longleftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \longleftrightarrow \gamma(x,y) & (13) \longleftrightarrow \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} \longleftrightarrow \gamma(x-y,-y) \\
(12) \longleftrightarrow \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix} \longleftrightarrow \gamma(-x,y-x) & (123) \longleftrightarrow \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \longleftrightarrow \gamma(-y,x-y) \\
(23) \longleftrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \longleftrightarrow \gamma(y,x) & (132) \longleftrightarrow \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \longleftrightarrow \gamma(y-x,-x)
\end{array}$$

Suppose $\sigma = (12)$ and $\tau = (13)$, then $\gamma = (132) = \sigma\tau$ and

$$\rho(\gamma) = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} = \rho(\sigma)\rho(\tau)$$

Estimating in a Sector

Lag-window estimates of the normalized bispectrum are asymptotically normal. One of the twelve sectors is considered. A coarse grid of p points and a fine grid of n points at each coarse point is estimated.



Subba-Rao-Gabr Gaussianity Test

The vector of estimates over the course points are asymptotically p-variate complex normal

$$\boldsymbol{x}_i \sim \mathcal{N}_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right) \qquad (i = 1, \ldots, n)$$

Testing Gaussianity ($\boldsymbol{\mu} \equiv \mathbf{0}$) is done with Hotelling T^2 statistic (multivariate t-test)

$$T^2 = n\bar{\boldsymbol{x}}^{\dagger}\hat{\boldsymbol{S}}^{-1}\bar{\boldsymbol{x}}$$

$\boldsymbol{S}$ where † is the adjoint operation (complex transpose) and $\hat{\boldsymbol{S}}$ is the (complex) sample variance-covariance matrix.

Under the null,

$$T^2 \sim \frac{2p}{2(n-p)} F_{2p,2(n-p)}$$

Subba-Rao-Gabr Linearity Test

Now we wish to test $\mu_1 = \cdots = \mu_p$ (without knowing the value of μ_1).
Let $\boldsymbol{C}$ be a $(p-1) \times p$ contrast matrix (i.e. has rank $p-1$ and $\boldsymbol{C}\mathbf{1}_p = 0$). E.g.,

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}$$

or the “Helmert matrix"

$$\boldsymbol{C} = \begin{pmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{-1}{\sqrt{1\cdot 2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{-2}{\sqrt{2\cdot 3}} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{p(p-1)}} & \frac{1}{\sqrt{p(p-1)}} & \frac{1}{\sqrt{p(p-1)}} & \cdots & \frac{-(p-1)}{\sqrt{p(p-1)}} \end{pmatrix}$$

Therefore we see

$$\boldsymbol{C}\boldsymbol{x} \sim \mathcal{N}_{2(p-1)}\left(\boldsymbol{C}\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'\right) \Longrightarrow \tilde{T}^2 = n(\boldsymbol{C}\boldsymbol{x})^*(\boldsymbol{C}\boldsymbol{S}\boldsymbol{C})^{-1}(\boldsymbol{C}\boldsymbol{x})$$

Two new approaches

Asymptotic distributions drive these linearity tests. Resampling methods come into play here.

- Use subsampling distributions instead of asymptotics
- A sieve bootstrap

Sieve Bootstrap

Every linear time series either has an AR(∞) representation or can be closely approximated by an AR(∞) process of the form

$$X_t = \mu + \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t$$

This motivates the following sieve bootstrap linearity test.

1. Fit an AR(p) model to the data ($p \to \infty$ with n).
2. Under H_0, the fitted residuals are iid. The residuals are centered and resampled.
3. New data is generated from the AR(p) recursion and bootstrapped residuals.
4. The bootstrapped data is used to generates the null distribution of the linearity test statistic.
5. Critical values are computed from bootstrap distribution.

Thank You