



Table of Contents

1. Some Basic Concepts	1 – 1
1.1 Introduction	1 – 1
1.2 Linearity and Gaussianity: Preliminaries, Reversibility	1 – 2
1.3 Initial Data Analysis	1 – 8
1.4 Some Book References	1 – 9
2. Some Univariate NL Models	2 – 1
2.1 Volterra Series Expansion	2 – 1
2.2 State-dependent Models: BL, ExpARMA, RCAR, NLMA	2 – 3
2.3 Threshold Models: SETARMA, Multiple thresholds, asARMA, NeSETARMA	2 – 10
2.4 Smooth Transition (ST) Models: LSTAR, ESTAR	2 – 15
2.5 Nonlinear non-Gaussian Models	2 – 17
2.6 Artificial Neural Network Models	2 – 18
2.7 Markov Switching Models	2 – 21
2.8 Some Software References	2 – 23
3. Probabilistic Properties	3 – 1
3.1 Strict Stationarity	3 – 1
3.2 Invertibility	3 – 7
3.3 Some References	3 – 11

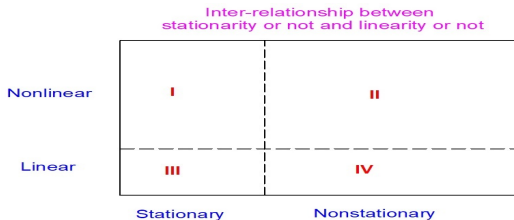


1. Some Basic Concepts

1.1 Introduction: Why nonlinear (nL) time series models?

Real world time series processes exhibit multifarious nonlinear and nonstationary dynamic behavior including aperiodic (asymmetric) patterns or cycles, intermittencies, and other transient behavior.

Additional features include symptoms such as non-Gaussianity, multi-modality, and time-irreversibility.



“Classical”, i.e. stationary and linear (III).



Preliminaries:

Def: The stochastic process $\{Y_t, t \in \mathbb{Z}\}$ defined on a probability space is called a **discrete-time time series**.

Def: $\{Y_t\}$ is **stationary** or **weakly stationary (WS)** if

1. $\mathbb{E}|Y_t|^2 < \infty, \forall t.$
2. $\mathbb{E}(Y_t) = \mu, \forall t.$
3. $\text{Cov}(Y_t, Y_{t+\ell}) = \gamma_Y(\ell)$ depends on $\ell \in \mathbb{Z}$ only.

Def: $\{Y_t\}$ is **strictly stationary (SS)** if $(Y_1, \dots, Y_T) \stackrel{D}{\sim} (Y_{1+\ell}, \dots, Y_{T+\ell}) \forall T \geq 1$ and $\ell \in \mathbb{Z}.$

Remarks:

1. $\text{SS} + (\mathbb{E}|Y_t|^2 < \infty) \implies \text{WS}.$
2. $\text{WS} \not\Rightarrow \text{SS}$ (think of an example)



1.2 Linearity and Gaussianity

Def: $\{Y_t\}$ is a **Gaussian time series** if

(Y_m, \dots, Y_n) is multivariate normal (or Gaussian)

for all integers $m < n$, i.e., all finite dimensional distributions are normal.

Remarks:

1. A Gaussian time series is completely determined by the mean function $\mu_t = \mathbb{E}(Y_t)$ and the covariance functions $\gamma(s, t) = \text{Cov}(Y_s, Y_t)$.
2. A Gaussian time series is **stationary** (SS or WS) if and only if $\mu_t = \mu$ and $\gamma(s, t) = \gamma(t - s)$ depends only on the time lag $t - s$.

Def: $\{Y_t\}$ is a **linear process** with mean 0 if for all $t \in \mathbb{Z}$

$$Y_t = \sum_{i=-\infty}^{\infty} \psi_i \varepsilon_{t-i}, \text{ where } \sum_{i=-\infty}^{\infty} \psi_i^2 < \infty, \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2). \quad (*)$$

Remarks:

1. The sequence $\{\varepsilon_t\}$ is called **strict white noise (WN)** as opposed to **weak WN**, which is a stationary sequence of uncorrelated random variables.
2. The requirement that $\{\varepsilon_t\}$ is i.i.d. is more restrictive than that $\{\varepsilon_t\}$ is serially uncorrelated. Independence implies that 3rd and higher-order non-contemporaneous moments of $\{\varepsilon_t\}$ are 0. When $\{\varepsilon_t\}$ is assumed to be Gaussian distributed, the two concepts of WN coincide.



Def: $\{Y_t\}$ is said to be **linear causal** if $\psi_i = 0$ for $i < 0$, i.e., if

$$Y_t = \varepsilon_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}, \text{ where } \sum_{i=1}^{\infty} \psi_i^2 < \infty, \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2). \quad (**)$$

Remarks:

1. The $MA(\infty)$ representation in $(**)$ should not be confused with the **Wold** decomposition theorem. In $(**)$ $\{\varepsilon_t\}$ is only assumed to be i.i.d. and **not** weakly WN as in the **Wold** representation.
2. If $\{\varepsilon_t\}$ is also Gaussian, $(**)$ becomes

$$Y_t = \varepsilon_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}, \text{ where } \sum_{i=1}^{\infty} \psi_i^2 < \infty, \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2).$$

Then the best **mean square error** predictor of $\{Y_t\}$ equals the best linear predictor.

Def: $\{Y_t\}$ is said to be **nonlinear** (nL) if neither $(*)$ nor $(**)$ hold.

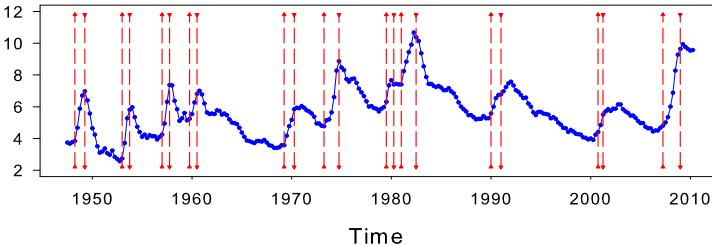


Figure: Quarterly U.S. unemployment rate (in %) (252 observations); *red triangle up* = business cycle peak, *red triangle down* = business cycle trough.

Example: The series displays step increases that end in sharp peaks and alternate with much more gradual and longer declines that end in mild troughs. Time series that exhibit such **strong asymmetric behavior** cannot be adequately modeled by linear time series models with normally distributed errors.



Reversibility:

Def: A strictly stationary stochastic process $\{Y_t\}$ is *time-reversible* (TR) if $(Y_1, Y_2, \dots, Y_T) \stackrel{D}{\sim} (Y_T, Y_{T-1}, \dots, Y_1)$.

Result: i.i.d. sequences $\{Y_t\}$ are TR.

Thm: (Breidt and Davies, *JTSA*, 1992) Consider the linear time series $\{Y_t\}$

$$Y_t = \sum_{i=-\infty}^{\infty} \psi_i \varepsilon_{t-i}, \quad \{\varepsilon_t\} \sim \text{i.i.d.},$$

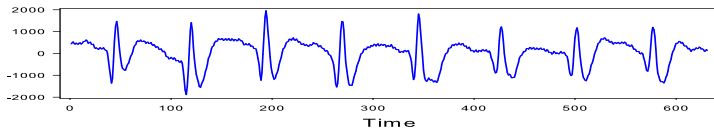
where $\psi(z) \neq z^r \psi(z^{-1})$ for any integer r . Assume either

- (a) ε_0 has mean 0 and finite variance, and $\{Y_t\}$ has a spectral density almost everywhere; or
- (b) $1/\psi(z) = \pi(z) = \sum \pi_j z^j$, the series converging in some annulus $D = \{z: d < |z| < d^{-1}\}$ with $d < 1$, and $\pi(B)Y_t = \sum \pi_j Y_{t-j} = \varepsilon_t$.

Then $\{Y_t\}$ is TR if and only if ε_0 is Gaussian.

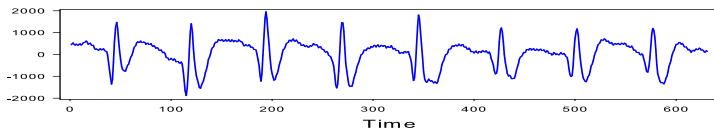


Example: Does the following time series look time-reversible?





Example: Does the following time series look time-reversible?



Application: Plot $\{Y_t\}$ with the time axis reversed. If the mirror image is not similar to the original plot there is “evidence” against TR. Formal test statistics follow soon.

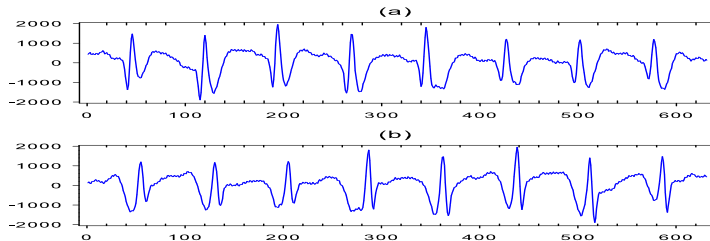


Figure: (a) EEG recordings of electrical potentials (activity) of the brain, and (b) the reversed plot.



1.3 Initial Data Analysis

Normality of independent data:

- Jarque–Bera (*Int. Stat. Review*, 1987);
- Lin–Mudholkar (*Biometrika*, 1980).

Weakly dependent data (Lobato and Velasco, *Econometric Theory*, 2004).

Robust JB test statistic (Bai and Ng, *JBES*, 2005).

What about the sample ACF and PACF?

Using asymptotic Bartlett 95% confidence bands $\pm 1.96/\sqrt{T}$ for nL time series can lead to **spurious** results (Berlinet and Francq, *JTSA*, 1997).

Some alternatives:

- Generalizations of the sample ACF and PACF (nonparametric); Nielsen and Madsen (*CSDA*, 2004). Matlab and R codes at <http://www2.imm.dtu.dk/courses/02427>.
- Kendall's (partial) tau test statistic; Harvill and Ray (*Comm. Stats.: Theory & Methods*, 1999).

Some additional tools:

- (Partial) mutual information coefficient; Granger and Lin (*JTSA*, 1994).
- Recurrence and directed scatter plots (analysis of deterministic chaos).



1.4 Some Books References:

- De Gooijer, J.G. (2017). *Elements of Nonlinear Time Series Analysis and Forecasting*. Springer-Verlag, New York.
<http://doi.org/10.1007/978-3-319-43252-6>.
- Douc, R., Moulines, E., and Stoffer, D.S. (2014). *Nonlinear Time Series: Theory, Methods, and Applications with R Examples*. Chapman & Hall/CRC, London.
<http://doi.org/10.1201/b16331>.
- Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York. <http://doi.org/10.1007/b97702>.
- Franses, P.H. and van Dijk, D. (2000). *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press, Cambridge, UK.
<http://doi.org/10.1017/cbo9780511754067>.
- Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall/CRC.
<http://doi.org/10.1201/9781420011210>.
- Teräsvirta, T., Tjøstheim, D., and Granger, C.W.J. (2010). *Modelling Nonlinear Economic Time Series*. Oxford University Press, New York.
<http://doi.org/10.1093/acprof:oso/9780199587148.001.0001>.
- Tsay, R.S. and Chen, R. (2018). *Nonlinear Time Series Analysis*. Wiley.
<http://doi.org/10.1002/9781119514312>.



2. Some Univariate nL Models

2.1 Volterra Series Expansion

Let $h(\cdot)$ denote a suitably smooth real-valued function. Then a general form for modeling $\{Y_t\}$ can be expressed as

$$h(Y_t, Y_{t-1}, Y_{t-2}, \dots) = \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2), \quad (*)$$

which is independent of future observations and due to its generality may be considered as an nL model. Find $h(\cdot)$ such that (*) is **causally invertible**, i.e.

$$Y_t = \tilde{h}(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots). \quad (**)$$

Assume that $\tilde{h}(\cdot)$ is a sufficiently well-behaved function so that we can expand (**) in a Taylor series about some fixed time point – say $\mathbf{0} = (0, 0, \dots)'$. Then we can write Y_t in a so-called **discrete-time Volterra series**, i.e.,

$$Y_t = \mu + \sum_{u=0}^{\infty} g_u \varepsilon_{t-u} + \sum_{u,v=0}^{\infty} g_{uv} \varepsilon_{t-u} \varepsilon_{t-v} + \sum_{u,v,w=0}^{\infty} g_{uvw} \varepsilon_{t-u} \varepsilon_{t-v} \varepsilon_{t-w} + \dots,$$

where

$$\mu = g(\mathbf{0}), \quad g_{u_1} = \left(\frac{\partial \tilde{h}}{\partial \varepsilon_{t-u_1}} \right)_{\mathbf{0}}, \dots, g_{u_1, \dots, u_n} = \left(\frac{\partial^n \tilde{h}}{\partial \varepsilon_{t-u_1} \dots \partial \varepsilon_{t-u_n}} \right)_{\mathbf{0}}.$$



Similarly, we have the **dual Volterra series**

$$\varepsilon_t = \mu' + \sum_{u=0}^{\infty} g'_u Y_{t-u} + \sum_{u,v=0}^{\infty} g'_{uv} Y_{t-u} Y_{t-v} + \sum_{u,v,w=0}^{\infty} g'_{uvw} Y_{t-u} Y_{t-v} Y_{t-w} + \dots,$$

where $\{g'_u\}$, $\{g'_{uv}\}$, $\{g'_{uvw}\}$, ... are defined in a similar way as above. A more parsimonious representation is

$$\begin{aligned} \mu' + \sum_{u=0}^p g'_u Y_{t-u} + \sum_{u,v=0}^p g'_{uv} Y_{t-u} Y_{t-v} + \sum_{u,v,w=0}^p g'_{uvw} Y_{t-u} Y_{t-v} Y_{t-w} + \dots = \\ \mu + \sum_{u=0}^q g_u \varepsilon_{t-u} + \sum_{u,v=0}^q g_{uv} \varepsilon_{t-u} \varepsilon_{t-v} + \sum_{u,v,w=0}^q g_{uvw} \varepsilon_{t-u} \varepsilon_{t-v} \varepsilon_{t-w} + \dots, \end{aligned}$$

which can be expressed more generally as,

$$h^*(Y_t, \dots, Y_{t-p}) = g^*(\varepsilon_t, \dots, \varepsilon_{t-q}).$$

A further generalization, assuming $h^*(\cdot)$ is invertible, is given by

$$Y_t = G(Y_{t-1}, \dots, Y_{t-p}, \varepsilon_t, \dots, \varepsilon_{t-q}). \quad (***)$$

Note (***) treats $\{\varepsilon_t\}$ as an **observable input**; therefore, the input-output relationships are expressed in terms of a finite # of past inputs and outputs.

When $\{\varepsilon_t\}$ is **unobservable** and instead is taken as a rv, redefine $G(\cdot)$ as

$$Y_t = \tilde{G}(Y_{t-1}, \dots, Y_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}) + \varepsilon_t.$$



2.2 State-dependent Models

Assume $\tilde{G}(\cdot)$ is a sufficiently well-behaved function; then, we may proceed by expanding the right-hand side of (***) in a Taylor series about $(0, 0, \dots, 0)'$. Retaining only the 1st term in the expansion, gives

$$Y_t = \mu(\mathbf{s}_{t-1}) + \sum_{i=1}^p f_i(\mathbf{s}_{t-1}) Y_{t-i} + \sum_{j=1}^q g_j(\mathbf{s}_{t-1}) \varepsilon_{t-j} + \varepsilon_t,$$

where

$$\begin{aligned} \mathbf{s}_t &= (Y_t, \dots, Y_{t-p+1}, \varepsilon_t, \dots, \varepsilon_{t-q+1})', \\ \mu(\mathbf{s}_{t-1}) &= \tilde{G}(Y_{t-1}, \dots, Y_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}), \\ f_i(\mathbf{s}_{t-1}) &= \left(\frac{\partial \tilde{G}}{\partial Y_{t-i}} \right)_{\mathbf{s}_{t-1}}, \quad g_j(\mathbf{s}_{t-1}) = \left(\frac{\partial \tilde{G}}{\partial \varepsilon_{t-j}} \right)_{\mathbf{s}_{t-1}}. \end{aligned}$$

In ARMA-like notation, the **state-dependent model** (SDM) is given by

$$Y_t = \mu(\mathbf{s}_{t-1}) + \sum_{i=1}^p \phi_i(\mathbf{s}_{t-1}) Y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j(\mathbf{s}_{t-1}) \varepsilon_{t-j}. \quad (***)$$

Due to the characterization of the SDM as a locally linear ARMA model 'identifiability' like conditions need to be imposed on (***) .



Bilinear (BL) model:

A general BL model of order (p, q, P, Q) is defined as

$$Y_t = \phi_0 + \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{u=1}^P \sum_{v=1}^Q \psi_{uv} Y_{t-u} \varepsilon_{t-v}.$$

The model is linear in the Y_t 's and also in the ε_t 's separately but not in both. So, provided $\psi_{uv} \neq 0$, the ARMA(p, q) model is **nested** within the BL model.

Complete BL model:

$$Y_t = \varepsilon_t + \sum_{u=1}^P \sum_{v=1}^Q \psi_{uv} Y_{t-u} \varepsilon_{t-v}.$$

Three special cases:

- **Diagonal:** $\psi_{uv} = 0 \forall u \neq v$.
- **Superdiagonal:** $\psi_{uv} = 0 \forall u > v$. Here the multiplicative terms with non-zero coefficients are such that the input variable ε_{t-v} occurs after Y_{t-u} so that these terms are independent. This fact makes analysis somewhat **easier**.
- **Subdiagonal:** $\psi_{uv} = 0 \forall u < v$. The variable Y_{t-u} occurs strictly after ε_{t-v} , making analysis more **difficult**.



Example: Consider the following specifications

$$Y_t = \phi Y_{t-1} + \varepsilon_t \quad (\text{linear AR(1)})$$

$$Y_t = \phi Y_{t-1} + \varepsilon_t + \psi Y_{t-2} \varepsilon_{t-1} \quad (\text{subdiagonal})$$

$$Y_t = \phi Y_{t-1} + \varepsilon_t + \psi Y_{t-1} \varepsilon_{t-1} \quad (\text{diagonal})$$

$$Y_t = \phi Y_{t-1} + \varepsilon_t + \psi Y_{t-1} \varepsilon_{t-2} \quad (\text{superdiagonal})$$

with $\phi = 0.99$ and $\psi = -0.5$, and where $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

Compare each BL model through the effect of a **one-unit shock** on Y_t at time $t = 1$, i.e. $\varepsilon_1 = 1$, and $\varepsilon_2 = \varepsilon_3 = \dots = 0$, given the history ω_{t-1} .

Measure: The difference between the conditional expectation with/without the shock, called **generalized impulse response function (GIRF)**. We have

$$\text{GIRF}_Y(t, 1, \omega_{t-1}) = \mathbb{E}[Y_t | \varepsilon_1 = 1, \varepsilon_2 = 0, \varepsilon_3 = 0, \dots] - \mathbb{E}[Y_t | \varepsilon_1 = 0, \varepsilon_2 = 0, \dots].$$

Iterating each BL model, we get

$$\text{GIRF}^{(\text{sub})} = \phi^{t-1}, \quad \text{GIRF}^{(\text{diag})} = \phi^{t-1}, \quad \text{GIRF}^{(\text{super})} = \phi^{t-2}(\phi + \psi), \quad (t \geq 2).$$

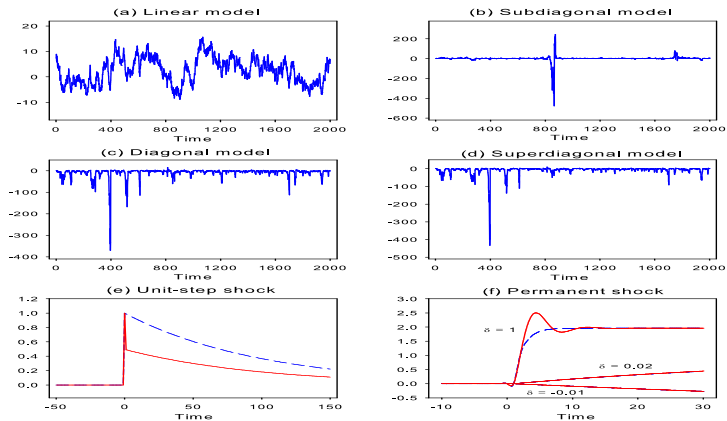


Figure: (a) – (d) Realizations of 4 time series processes ($T = 2,000$); (e) GIRFs for both diagonal and subdiagonal BL models (blue medium dashed line), and superdiagonal BL model (red solid line) for a unit-shock at $t = 1$; (f) GIRFs for both diagonal and superdiagonal BL models (blue medium dashed line) and subdiagonal model (red solid lines) for a permanent shock δ of magnitude $-0.01, 0.02$, and 1 at time $t = 1$.



Exponential ARMA (ExpARMA) model:

$$Y_t = \phi_0 + \sum_{i=1}^p \{\phi_i + \xi_i \exp(-\gamma Y_{t-d}^2)\} Y_{t-i} + \sum_{j=1}^q \{\theta_j + \tau_j \exp(-\gamma Y_{t-d}^2)\} \varepsilon_{t-j} + \varepsilon_t,$$

where $\gamma > 0$ denotes a **scaling factor**.

ExpARMA changes smoothly between 2 extreme linear models, since for large $|Y_{t-d}|$, the coefficients are almost ϕ_i 's and θ_j 's. For small values of $|Y_{t-d}|$, they are $\phi_i + \xi_i$ and $\theta_j + \tau_j$ and the exponential function changes smoothly between these 2 extreme values.

A sufficient condition for **strict stationarity** is that all the roots of the associated characteristic equation

$$z^p - c_1 z^{p-1} - \dots - c_p = 0$$

are inside the unit circle, where $c_i = \max\{|\phi_i|, |\phi_i + \xi_i|\}$ ($i = 1, \dots, p$).

Hence, the characteristic roots are **amplitude-dependent**, instead of constant.



Random Coefficient AR (RCAR) model:

Let $\mu(\mathbf{S}_{t-1}) = \mu$ as constant, $\theta_j(\mathbf{S}_{t-1}) = 0 \forall j$, and $\phi_i(\mathbf{S}_{t-1}) = \{\phi_i + \beta_{i,t}\}$. Then the SDM) reduces to,

$$Y_t = \mu + \sum_{i=1}^p \{\phi_i + \beta_{i,t}\} Y_{t-i} + \varepsilon_t,$$

where $\{\mathbf{B}_t = (\beta_{1,t}, \dots, \beta_{p,t})'\}$ is a sequence of i.i.d. random vectors with $\mathbb{E}(\mathbf{B}_t) = \mathbf{0}$ and $\text{Cov}(\mathbf{B}_t) = \Sigma_\beta$, and $\{\mathbf{B}_t\}$ is independent of $\{\varepsilon_t\}$.

If $p = 1$, a necessary and sufficient condition for **2nd-order stationarity** is that $\phi^2 + \sigma_\beta^2 < 1$. By introducing random coefficients to an ARMA model, we can generalize the RCAR model.

By assuming the coefficients $\beta_{i,t}$ are **not** independent but follow an arbitrary strictly stationary stochastic process (say an MA process) defined on the same probability space as $\{\varepsilon_t\}$, one obtains the so-called **doubly stochastic** model; **Tjøstheim** (JTSA, 1989).



Nonlinear MA (NLMA) model:

$$\begin{aligned}
 Y_t &= \varepsilon_t + \sum_{j=1}^q \theta_{j,i_1}(\mathbf{S}_{t-1}) \varepsilon_{t-j i_1} \\
 &= \varepsilon_t + \sum_{i_1=0}^Q \beta_{i_1} \varepsilon_{t-i_1} + \sum_{i_1=0}^Q \sum_{i_2=0}^Q \beta_{i_1, i_2} \varepsilon_{t-i_2} \varepsilon_{t-2i_1} + \cdots + \sum_{i_1=0}^Q \sum_{i_2=1}^Q \cdots \sum_{i_q=0}^Q \beta_{i_1, i_2, \dots, i_q} \varepsilon_{t-i_2} \varepsilon_{t-i_3} \cdots \varepsilon_{t-i_q},
 \end{aligned}$$

where η is the highest order of summations.

Example:

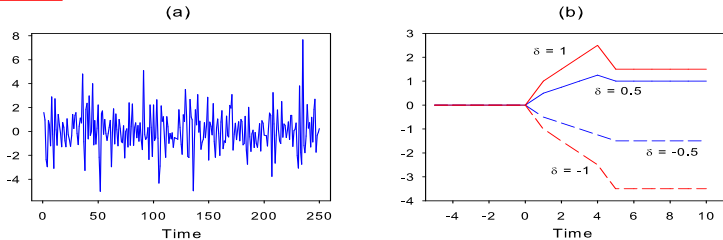


Figure: (a) A realization of the NLMA model $Y_t = \varepsilon_t + \beta(\varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3}) - \varepsilon_t \varepsilon_{t-4}$ with $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, $\beta = 0.5$ ($T = 250$); (b) 4 permanent step response functions.



2.3 Threshold Models

Self-exciting threshold ARMA (SETARMA) model:

$$Y_t = \sum_{i=1}^k \left(\phi_0^{(i)} + \sum_{u=1}^{p_i} \phi_u^{(i)} Y_{t-u} + \varepsilon_t^{(i)} + \sum_{v=1}^{q_i} \theta_v^{(i)} \varepsilon_{t-v}^{(i)} \right) I(Y_{t-d} \in \mathbb{R}^{(i)}),$$

where $\varepsilon_t^{(i)} = \sigma_i^2 \varepsilon_t$, and $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, 1)$.

Example:

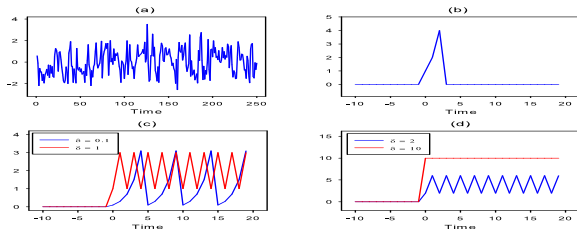


Figure: (a) A realization of $Y_t = 2Y_{t-1} + \varepsilon_t$ if $|Y_{t-1}| \leq 2$ and $Y_t = \varepsilon_t$ otherwise ($T = 250$), and $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, 1)$; (b) GIRF for a one-unit shock at time $t = 1$ ($Y_0 = 0$); (c) Permanent step responses for $\delta = 0.1$ and $\delta = 1$; (d) Permanent step responses for $\delta = 2$ and $\delta = 10$. **Note:** $\{Y_t\}$ switches between a *locally nonstationary* and a *locally stationary* process. *Globally*, it is stationary.



Multivariate Thresholds:

Control the dynamics of a SETARMA model by introducing multivariate thresholds with **unknown** linear relationships between the thresholds.

Example: Consider the SETAR(2; 1, 1)₂ model

$$Y_t = 0.5 + 0.9Y_{t-1} - 1.8Y_{t-1}I(\omega'_1\tilde{\mathbf{X}}_{t-1} \leq 0) - I(\omega'_2\tilde{\mathbf{X}}_{t-1} \leq 0) + \varepsilon_t, \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

where $\omega_1 = (1, -1)'$, $\omega_2 = (0, 1)'$, and $\tilde{\mathbf{X}}_{t-1} = (Y_{t-1}, Y_{t-2})'$. Rewriting the model in 4 separate regimes gives

$$Y_t = \begin{cases} -0.5 - 0.9Y_{t-1} + \varepsilon_t, & \text{I: if } Y_{t-1} - Y_{t-2} \leq 0 \text{ and } Y_{t-2} \leq 0, \\ -0.5 + 0.9Y_{t-1} + \varepsilon_t, & \text{II: if } Y_{t-1} - Y_{t-2} > 0 \text{ and } Y_{t-2} \leq 0, \\ 0.5 - 0.9Y_{t-1} + \varepsilon_t, & \text{III: if } Y_{t-1} - Y_{t-2} \leq 0 \text{ and } Y_{t-2} > 0, \\ 0.5 + 0.9Y_{t-1} + \varepsilon_t, & \text{IV: if } Y_{t-1} - Y_{t-2} > 0 \text{ and } Y_{t-2} > 0. \end{cases}$$

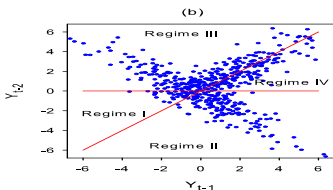
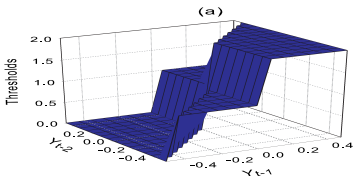


Figure: (a) Threshold boundaries; (b) Scatter plot of Y_{t-2} versus Y_{t-1} with 2 separating hyperplanes (red solid lines); $T = 500$.



Asymmetric ARMA (asARMA) model:

$$Y_t = \phi_0 + \sum_{i=1}^p \phi_i^+ Y_{t-i}^+ + \sum_{i=1}^p \phi_i^- Y_{t-i}^- + \varepsilon_t + \sum_{j=1}^q \theta_j^+ \varepsilon_{t-j}^+ + \sum_{j=1}^q \theta_j^- \varepsilon_{t-j}^-.$$

Here Y_t^\pm and ε_t^\pm denote the asymmetric component processes, defined as

$$Y_t^- = Y_t I(\varepsilon_t \leq 0), \quad Y_t^+ = Y_t I(\varepsilon_t > 0), \quad \varepsilon_t^- = \varepsilon_t I(\varepsilon_t \leq 0), \quad \varepsilon_t^+ = \varepsilon_t I(\varepsilon_t > 0),$$

with $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2)$.

If $p \neq 0$ and $q = 0$, the model reduces to an **asymmetric AR(p) (asAR)**. It is an **asymmetric MA(q) (asMA)** model for $p = 0$ and $q \neq 0$. **Note:** The model has 4 filters, 2 for positive innovations and 2 for negative innovations.

An alternative form is

$$Y_t = \sum_{i=1}^p (\phi_i^- + \alpha_i I(\varepsilon_{t-i} > 0)) Y_{t-i} + \varepsilon_t + \sum_{j=1}^q (\theta_j^- + \beta_j I(\varepsilon_{t-j} > 0)) \varepsilon_{t-j},$$

where $\alpha_i = \phi_i^+ - \phi_i^-$ ($i = 1, \dots, p$), and $\beta_j = \theta_j^+ - \theta_j^-$ ($j = 1, \dots, q$).



Nested SETARMA (NeSETARMA) model (Astatkie et al., *IJF*, 1997):

Suppose that a strictly stationary process $\{Y_t\}$ (output) has 2 input variables $\{X_t\}$ and $\{Z_t\}$. Moreover, assume that the regime-switching is conditional on the values of the delayed observable variables Y_t and X_t .

Then the complete dynamic system is divided in **2 subsystems**, or **stages**. Each stage consists of regimes, with the **2nd stage regimes** nested within those of the **1st stage**. The regimes are formed in such a way that there is a linear relationship between Y_t and its lagged values, and a linear relationship between Y_t and lagged values of X_t .

If Y_t is used as regime-switching variable in the **1st stage**, then X_t will be used in the **2nd stage** and the resulting model is called an **output-input** NeSETARMA model.

On the other hand, if X_t is used in the **1st stage** and Y_t in the **2nd**, then the model is called an **input-output** NeSETARMA model. The (possibly lagged) relationship between Y_t and Z_t may be linear or quadratic.



Nested SETARMA (NeSETARMA) model (Cont'd)

Some notation:

- Let $k_1 \geq 1$ be the # of **1st-stage** regimes formed by partitioning the values of Y_{t-d_1} into non-overlapping intervals with $d_1 \in \mathbb{Z}^+$ the **1st-stage delay**.
- Let $\mathbb{R}^{(i)} = (r_{i-1}, r_i]$ denote the i th ($i = 1, \dots, k_1$) interval with $r_0 = -\infty$ and $r_{k_1} = \infty$. The parameters r_1, \dots, r_{k_1-1} are the **1st-stage thresholds**.
- Let $\ell_{i,2} \geq 1$ ($i = 1, \dots, k_1$) be the # of **2nd-stage** regimes formed by using X_{t-d_2} as a threshold variable with $d_2 \in \mathbb{Z}^+$ the **2nd-stage delay**.
- Let $\mathbb{R}^{(i,j)} = (r_{i,j-1}, r_{i,j}]$ ($i = 1, \dots, k_1; j = 1, \dots, \ell_{i,2}$) denote the j th **2nd-stage** regime within the i th **1st-stage** regime with $r_{i,0} = -\infty$ and $r_{i,\ell_{i,2}} = \infty$. The set $\{r_{i,1}, \dots, r_{i,\ell_{i,2}-1}\}$ represents the **2nd-stage thresholds**.

The **NeSETARMA** model is defined as

$$Y_t = \sum_{i=1}^{k_1} \left\{ \sum_{j=1}^{\ell_{i,2}} \left(\phi_0^{(i,j)} + \sum_s \phi_s^{(i,j)} Y_{t-s} + \sum_u \xi_u^{(i,j)} X_{t-u} \right. \right. \\ \left. \left. + \sum_v \eta_v^{(i,j)} Z_{t-v} + \varepsilon_t^{(i,j)} + \sum_w \theta_w^{(i,j)} \varepsilon_{t-w} \right) I(X_{t-d_2} \in \mathbb{R}^{(i,j)}) \right\} I(Y_{t-d_1} \in \mathbb{R}^{(i)}),$$

where $\{\varepsilon_t^{(i,j)}\} \stackrel{\text{i.i.d.}}{\sim} (0, 1)$. Clearly, the model consists of $\sum_{i=1}^{k_1} \ell_{i,2}$ regimes.



2.4 Smooth Transition (ST) models

A (two-regime) **smooth transition autoregressive (STAR)** model of order $(2; p, p)$ is defined as

$$\begin{aligned} Y_t &= \left\{ \phi_0^{(1)} + \sum_{u=1}^p \phi_u^{(1)} Y_{t-u} \right\} (1 - G(z_t)) + \left\{ \phi_0^{(2)} + \sum_{u=1}^p \phi_u^{(2)} Y_{t-u} \right\} G(z_t) + \varepsilon_t, \\ &= \phi_0 + \sum_{u=1}^p \phi_u Y_{t-u} + \left\{ \xi_0 + \sum_{u=1}^p \xi_u Y_{t-u} \right\} G(z_t) + \varepsilon_t, \end{aligned}$$

where $\phi_u = \phi_u^{(1)}$ and $\xi_u = \phi_u^{(2)} - \phi_u^{(1)}$ ($u = 0, 1, \dots, p$).

The **transition (or activation-level) function** $G(\cdot)$ allows the conditional expectation of the model to change smoothly.

Assume $G(z_t) \equiv G(Y_{t-d}; \gamma, c) = \Phi(\gamma\{Y_{t-d} - c\})$, where $\Phi(\cdot)$ is the CDF of the $\mathcal{N}(0, 1)$ distribution. $d \geq 1$ is the **delay parameter**, c is a **location value**, indicating when the transition is occurring, whereas $\gamma > 0$ is a **slope parameter**.

When γ increases, the transition is completed in a short period of time, and $\Phi(\gamma\{Y_{t-d} - c\})$ approaches $I(Y_{t-d} - c)$. Then the model reduces to a SETAR(2; p, p). When $\gamma \approx 0$ it may be well approximated by an AR(p).



Two plausible transition functions:

- Logistic function:

$$G(Y_{t-d}; \gamma, c) = \frac{1}{1 + \exp\{-\gamma(Y_{t-d} - c)\}}, \quad \gamma > 0,$$

and the resulting model is the **logistic STAR (LSTAR)** model.

- Exponential function:

$$G(Y_{t-d}; \gamma, c) = 1 - \exp\{-\gamma(Y_{t-d} - c)^2\}, \quad \gamma > 0,$$

giving rise to the **exponential STAR (ESTAR)** model.

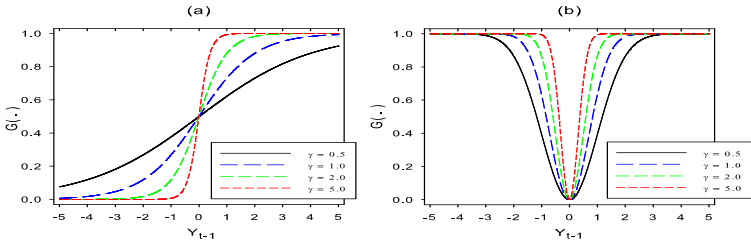


Figure: Effects of various values of the smoothness parameter γ on (a) the logistic transition function, and (b) the exponential transition function. Both functions with $c = 0$ and $d = 1$.



2.5 Nonlinear non-Gaussian Models

Example: NEAR(1)

$$Y_t = \varepsilon_t + \begin{cases} \beta Y_{t-1} & \text{with prob. } \alpha, \\ 0 & \text{with prob. } (1 - \alpha), \end{cases}$$

$$= \beta J_t Y_{t-1} + \varepsilon_t,$$

where

$$\varepsilon_t = \begin{cases} E_t & \text{with prob. } \rho_1 = (1 - \beta)/(1 - (1 - \alpha)\beta) \\ (1 - \alpha)\beta E_t & \text{with prob. } 1 - \rho_1 = \alpha\beta/(1 - (1 - \alpha)\beta) \end{cases}$$

$$J_t = \begin{cases} 0 & \text{with prob. } (1 - \alpha) \\ 1 & \text{with prob. } \alpha, \end{cases}$$

where $\{E_t\}$ is a sequence of i.i.d. unit mean exponential random variables which ensures that $f_Y(y) = \exp(-y)$ ($0 \leq y < \infty$).

The parameters α and β are allowed to take values over the domain defined by $0 \leq \alpha, \beta \leq 1$ with $\alpha = \beta \neq 1$. The innovation process is not allowed to take on negative values, i.e. $\mathbb{P}(E_t \leq 0) = 0$.

Two extreme cases:

- $\alpha = 1, 0 \leq \beta \leq 1$: the **exponential AR (EAR)** model; Lawrance and Lewis (*JRSS-B*, 1980).
- $\beta = 1, 0 \leq \alpha < 1$: the **transposed EAR (TEAR)** model; Lawrance and Lewis (*Adv. Appl. Prob.*, 1981).



2.6 Artificial Neural Network Models

Example: A single hidden layer feed-forward AR-NN(2; 0, 1) model

$$Y_t = 0.15 + \xi \tanh(Y_{t-1} - 1) - \xi \tanh(Y_{t-1} - 1.5) + \varepsilon_t,$$

where $\tanh(x) = (\exp(2x) - 1)/(\exp(2x) + 1)$, and $Y_0 = 0.1$.

Note:

- For approximately $1 \leq \xi \leq 3.4$ the model converges to a stable fixed point.
- For approximately $3.4 < \xi < 4.5$ we see a local stable oscillation of period 2.
- The oscillation period is doubled for $4.5 < \xi < 5.8$.
- At about $\xi = 5.8$ the plot hints at deterministic chaos, i.e. the model loses predictability.

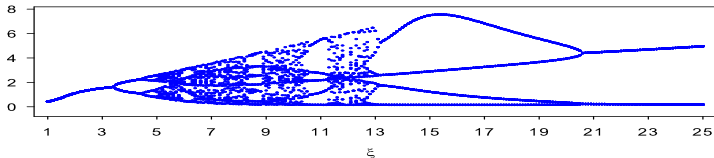


Figure: Skeleton of the AR-NN(2; 0, 1) model for 25 iterations of $\{Y_t\}$ for each value of $\xi = 1, 1.1, \dots, 24.9, 25$.



Neuro-coefficient STAR model:

NCTAR model of order p with q activation-level functions, $\text{NCTAR}(k; p, \dots, \rho)_q$, is defined as

$$Y_t = \phi_0 + \phi' \mathbf{X}_{t-1} + \sum_{j=1}^k (\xi_{0j} + \xi_j' \mathbf{X}_{t-1}) G(\tilde{\mathbf{X}}_{t-1}; \tilde{\omega}_j, c_j) + \varepsilon_t,$$

where

$$G(\tilde{\mathbf{X}}_{t-1}; \tilde{\omega}_j, c_j) = (1 + \exp(-[\tilde{\omega}_j' \tilde{\mathbf{X}}_{t-1} - c_j]))^{-1},$$

with

$$\begin{aligned} \mathbf{X}_{t-1} &= (Y_{t-1}, \dots, Y_{t-p})', & \tilde{\mathbf{X}}_{t-1} &= (Y_{t-1}, \dots, Y_{t-q})' \\ \tilde{\omega}_j &= (\tilde{\omega}_{1j}, \dots, \tilde{\omega}_{qj})', & \xi_j &= (\xi_{1j}, \dots, \xi_{pj})', \quad (j = 1, \dots, k), \end{aligned}$$

and k the # of nodes in the hidden layer.

Some parameter restrictions need to be imposed, to guarantee identifiability.

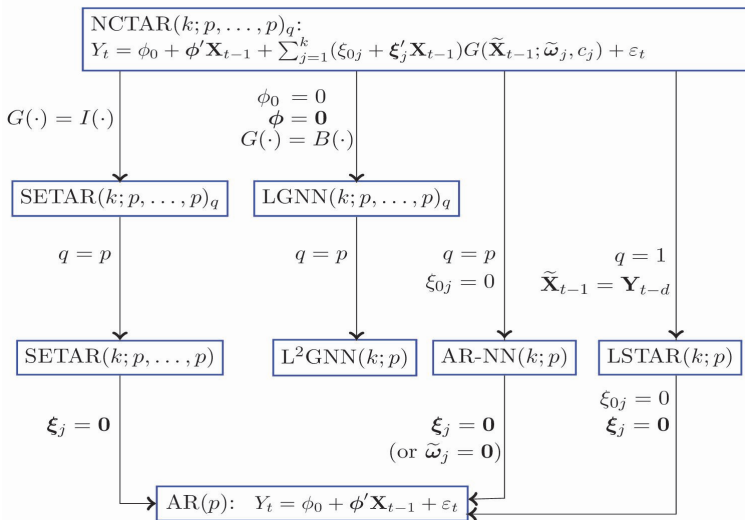


Figure: Flow diagram of various relationships between (non)linear AR models.



2.7 Markov Switching (MS) Models

A Markov chain $\{S_t\}$ is a discrete stochastic process $S_t \in \{1, \dots, k\}$, satisfying

$$\mathbb{P}(S_t = j | S_{t-1} = i, S_{t-2} = r, \dots) = \mathbb{P}(S_t = j | S_{t-1} = i) = p_{ij},$$

$$\sum_{j=1}^k p_{ij} = 1, \quad p_{ij} \geq 0, \quad \forall i, j \in \{1, \dots, k\}.$$

Loosely speaking, a Markov chain (process) is called **irreducible** if any state j can be reached from state i in a few steps, and it is termed **aperiodic** if the # of steps it needs to return to a state has no period. The chain is **ergodic** if it is irreducible and aperiodic.

Any Markov chain has a stationary distribution $\{\pi_j = \mathbb{P}(S_t = j)\}_{j=1}^k$. For an ergodic Markov chain, $\pi_j = \lim_{n \rightarrow \infty} \mathbb{P}(S_n = j | S_1 = i)$ (independent of i).



Markov switching ARMA model:

Assume $\{Y_t\}$ is influenced by a hidden discrete stochastic Markov process $\{S_t\}$. Then a **Markov-switching** ARMA (MS-ARMA) is defined as

$$Y_t = \sum_{i=1}^k \delta_{ti} \left(\phi_0^{(i)} + \sum_{u=1}^{p_i} \phi_u^{(i)} Y_{t-u} + \varepsilon_t^{(i)} + \sum_{v=1}^{q_i} \theta_v^{(i)} \varepsilon_{t-v}^{(i)} \right),$$

where

$$\delta_{ti} = \begin{cases} 1 & \text{if } S_t = i, \\ 0 & \text{otherwise,} \end{cases}$$

with $\varepsilon_t^{(i)} = \sigma_i^2 \varepsilon_t$, and $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, 1)$, independent of $\{S_t\}$. So, S_t denotes the **regime** or **state** prevailing at time t , one of k possible cases. When $k = 1$ there is only one state and $\{Y_t\}$ degenerates to an ARMA process.

Adding exogenous variables, such as trends, is a straightforward extension. Another extension is to allow for **GARCH errors**. Multivariate modeling, including modeling **cointegrated** processes, is also an option.



2.8 Some Software References

- The R-tsDyn package contains a host of functions for modeling univariate/multivariate threshold- and smooth transition models.
- Chap. 18 in the book by Zivot (2006) covers some popular nL time series models/methods. S-Plus script files, using the S-Plus FinMetrics module, are available at <http://faculty.washington.edu/ezivot/MFTS2ndEditionScripts.htm>. R scripts: <http://faculty.washington.edu/ezivot/MFTSR.htm>.
- The website <http://sites.google.com/site/marcelocmedeiros/Home/codes> offers a set of MATLAB codes to estimate logistic smooth transition regression models with and without long memory.
- The MATLAB toolbox NNSYSID contains a number of m-files for training and evaluation of multi-layer perceptron type neural networks; see http://personales.upv.es/ccarrasc/doc/1998-1999/trabajo_vivi/Programas/nnsysid.html.
- MS_Regress is a MATLAB package for estimating Markov regime switching models; see <http://msperlin.github.io/>.
- GAUSS code for estimating MS models is available from James D. Hamilton's website at <http://econweb.ucsd.edu/~jhamilton/software.htm>.



3. Probabilistic Properties

3.1 Strict Stationarity

In a multivariate setting, a **stochastic recurrence equation (SRE)** is defined as

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t, \quad t \in \mathbb{Z},$$

where $\mathbf{Y}_t = (Y_t, \dots, Y_{t-m+1})'$ and \mathbf{B}_t are random vectors in \mathbb{R}^m , \mathbf{A}_t are random $m \times m$ matrices, and $\{(\mathbf{A}_t, \mathbf{B}_t), t \in \mathbb{Z}\}$ is an i.i.d. sequence.

Clearly, the SRE is the defining equation of a vector AR(1) process with random coefficient matrix \mathbf{A}_t .

Hence, it is also called a **generalized (multivariate) random coefficient AR** process or RCA for short. The process is Markovian with transition probability equal to the distribution of $\mathbf{A}_t \mathbf{y} + \mathbf{B}_t$. The SRE embeds many nL DGPs.



Let $|\cdot|$ be any vector norm in \mathbb{R}^m , and $\|\cdot\|$ the corresponding matrix norm, i.e. for any $m \times m$ matrix \mathbf{A} , $\|\mathbf{A}\| = \sup_{|\mathbf{x}|=1} |\mathbf{A}\mathbf{x}|$. For an i.i.d. sequence of $m \times m$ matrices $\{\mathbf{A}_n, n \in \mathbb{Z}\}$, with $\mathbb{E}(\log^+ \|\mathbf{A}_1\|) < \infty$, we define the associated **top Lyapunov exponent** $\gamma(\cdot)$ by

$$\gamma(\mathbf{A}) = \inf_{n \in \mathbb{N}} \frac{1}{n} \mathbb{E}(\log \|\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_n\|) \stackrel{\text{a.s.}}{=} \lim_{n \rightarrow \infty} \frac{1}{n} \log \|\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_n\|.$$

By recursive substitution of the lagged values of \mathbf{Y}_t , the SRE model can be rewritten as

$$\mathbf{Y}_t = \left(\prod_{i=0}^s \mathbf{A}_{t-i} \right) \mathbf{Y}_{t-s-1} + \sum_{i=0}^s \left(\prod_{j=0}^{i-1} \mathbf{A}_{t-j} \right) \mathbf{B}_{t-i}, \quad \forall s \in \mathbb{N},$$

with the usual convention $\prod_{j=0}^{-1} \mathbf{A}_{t-j} = \mathbf{I}_m$. If $\lim_{s \rightarrow \infty} \left(\prod_{i=0}^s \mathbf{A}_{t-i} \right) \mathbf{Y}_{t-s-1} \stackrel{\text{a.s.}}{=} \mathbf{0}_m$ holds, then it is reasonable to hope that the model has a solution process $\{\mathbf{Y}_t\}$ that is stationary.



Suppose $\gamma(\mathbf{A}) < 0$. Then, under some mild conditions, the series

$$\mathbf{Y}_t = \mathbf{B}_t + \sum_{s=1}^{\infty} \mathbf{A}_t \mathbf{A}_{t-1} \cdots \mathbf{A}_{t-s+1} \mathbf{B}_{t-s}, \quad (*)$$

converges a.s., and the process $\{\mathbf{Y}_t\}$ is a **non-anticipative** stationary solution to SRE.

Here, **non-anticipative** (or causal) means that $\{\mathbf{Y}_t\}$ is independent of $\{(\mathbf{A}_{t+h}, \mathbf{B}_{t+h}), h \in \mathbb{N}\}$ for each t . Further, the condition $\gamma(\mathbf{A}) < 0$ is sufficient when $\{(\mathbf{A}_t, \mathbf{B}_t)\}$ is strictly stationary and ergodic.

Assume $m = 1$. Then, $\{Y_t\}$ as in (*) is the unique strictly stationary solution of SRE provided $-\infty \leq \mathbb{E}(\log |A_1|) < 0$ and $\mathbb{E}(\log^+ |B_1|) < \infty$. These two conditions are easy to check, and $\gamma(\mathbf{A}) = \mathbb{E}(\log |A_1|)$ can be obtained explicitly.



Example: (Evaluating the Top Lyapunov Exponent)

Consider the stochastic process

$$Y_t = \varepsilon_t + \beta_1 Y_{t-1} \varepsilon_{t-1} + \beta_2 Y_{t-2} \varepsilon_{t-2}^2, \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2).$$

This can be written in the form of an SRE with

$$\mathbf{Y}_t = \begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix}, \quad \mathbf{A}_t = \begin{pmatrix} \beta_1 \varepsilon_{t-1} & \beta_2 \varepsilon_{t-2}^2 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{B}_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}.$$

When $\beta_2 = 0$ (i.e., $m = 1$), the strict stationarity condition based on the top Lyapunov exponent takes the simple form

$$\gamma(\mathbf{A}) = \mathbb{E}(\log |\beta_1 \varepsilon_t|) = \log |\beta_1| + \mathbb{E}(\log |\varepsilon_t|) < 0.$$

If $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$, the condition reduces to $\sigma_\varepsilon |\beta_1| < \sqrt{2} \exp(C/2) = 1.8874 \dots$, where C is **Euler's constant**.

When $m > 1$, closed form expressions for $\gamma(\mathbf{A})$ are hard to obtain, and one has to resort to MC simulations.



Second-order Stationarity:

The **spectral radius** $\rho(\mathbf{M})$ of a square matrix \mathbf{M} is defined as

$$\rho(\mathbf{M}) = \sup\{|\lambda| : \lambda \text{ is eigenvalue of } \mathbf{M}\}.$$

Then, provided $\mathbb{E}\|\mathbf{B}_t\|^2 < \infty$, it can be deduced that

$$\rho(\mathbb{E}(\mathbf{A}_t \otimes \mathbf{A}_t)) < 1$$

is a **necessary and sufficient condition** for the moments of order two to exist.

This condition has a similar implication as that the characteristic polynomial associated with a linear AR process has no roots on and within the unit circle.

If, in addition \mathbf{A}_t has finite moments of order $2m$ ($m > 1$), then a necessary and sufficient condition ensuring finiteness of higher-order moments is $\rho[\mathbb{E}\{(\mathbf{A}_t)^{\otimes 2m}\}] < 1$, where $\mathbf{M}^{\otimes m} = \mathbf{M} \otimes \dots \otimes \mathbf{M}$ (m factors).



Ergodicity: It is **hard** to obtain a necessary and sufficient condition for ergodicity or geometrical ergodicity (ensures strict stationarity) for a given class of nL DGPs.

Example: Consider the SETAR(2; 1, 1) model

$$Y_t = \begin{cases} \phi_1 Y_{t-1} + \varepsilon_t & \text{if } Y_{t-d} \leq 0, \\ \phi_2 Y_{t-1} + \varepsilon_t & \text{if } Y_{t-d} > 0, \end{cases}$$

where $\{\varepsilon_j\}$ are i.i.d. random variables with absolutely continuous marginal distribution and positive pdf over the real line \mathbb{R}^1 , and $\mathbb{E}|\varepsilon_t| < \infty$.

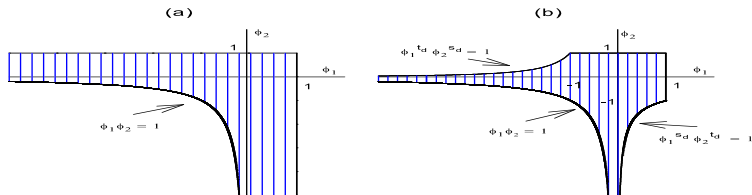


Figure: Ergodic region of a SETAR(2; 1, 1) model; (a) $d = 1$, and (b) general d , where s_d and t_d are nonnegative integers depending on d , and s_d and t_d are odd and even numbers, respectively.



3.1 Invertibility

Suppose $\{Y_t\}$ follows the stationary and ergodic NLARMA(p, q) model

$$Y_t = g(Y_{t-1}, \dots, Y_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}; \boldsymbol{\theta}) + \varepsilon_t, \quad (*)$$

where $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2)$, and $g(\cdot; \boldsymbol{\theta})$ is a known real-valued function for a known parameter vector $\boldsymbol{\theta}$.

Granger–Andersen (global) invertibility: Suppose that q initial values, say $\bar{\varepsilon}_j$ ($j = -q + 1, \dots, 0$), of (*) are given and that all Y_t are known. Let $\{\hat{\varepsilon}_t\}$ be a sequence of **innovations** generated by

$$\hat{\varepsilon}_t = Y_t - g(Y_{t-1}, \dots, Y_{t-p}, \hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-q}; \boldsymbol{\theta}),$$

where $\hat{\varepsilon}_i = \bar{\varepsilon}_i$ for $i \leq 0$. Define the **reconstruction errors** as $\mathbf{e}_t = \varepsilon_t - \hat{\varepsilon}_t$. Then the NLARMA model is said to be **invertible**, if

$$\mathbb{E}|\mathbf{e}_t|^r \rightarrow 0 \quad \text{as } t \rightarrow \infty \quad (r = 1, 2, \dots).$$

If (*) involves **estimated parameters**, the condition becomes

$$\mathbb{E}|\mathbf{e}_t|^r \rightarrow c \quad \text{as } t \rightarrow \infty,$$

where $c < \infty$ is some constant. It is appropriate to take $r = 2$.



A **ready-to-use method** to check invertibility.

Empirical invertibility of an NLARMA(p, q) model:

- (i) Generate a random sample of i.i.d. innovations $\{\tilde{\varepsilon}_t\}_{t=T+1}^N$ from the known distribution function (e.g., normal) of the residual series $\{\hat{\varepsilon}_t\}_{t=1}^T$, where N is some large value, say $N = 1,000$.
- (ii) Replace ε_t by $\tilde{\varepsilon}_t$ for $t = T + 1, \dots, N$ and use past values Y_{t-k} ($k = 0, \dots, p$), and $\hat{\varepsilon}_{t-k}$ ($k = 0, \dots, q$), to generate a new set of observations $\{\tilde{Y}_t\}_{t=T+1}^N$.
- (iii) Calculate $\{\hat{\theta}_t = \tilde{Y}_t - \hat{Y}_t\}_{t=T+1}^N$, where \hat{Y}_t are the out-of-sample fitted values. Estimate $\mathbb{E}(e_t^2)$ by $(\tau - T)^{-1} \sum_{t=T+1}^{\tau} \hat{\theta}_t^2$. If for all values of $\tau = T + 1, \dots, N$, this sequence does not exceed a pre-fixed value the process $\{Y_t\}$ is said to be **empirically (globally) invertible**, otherwise it suggests **non-invertibility**.



Example: Consider the asMA(1) model

$$Y_t = \varepsilon_t + \beta^+ \varepsilon_{t-1}^+ + \beta^- \varepsilon_{t-1}^-, \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),$$

where $\varepsilon_t^+ = I(\varepsilon_t \geq 0)\varepsilon_t$ and $\varepsilon_t^- = I(\varepsilon_t < 0)\varepsilon_t$.

Below is a graph of the empirical invertibility region for a simulated time series of size $T = 100$, using 1,000 MC replications, and $N = 1,000$.

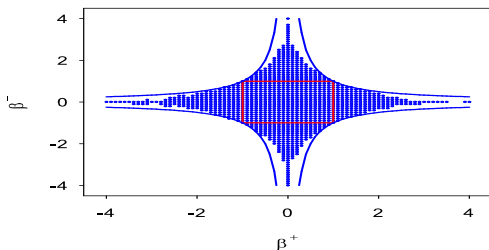


Figure: Contour plot of the invertibility region for the asMA(1) model for $T = 100$ (blue dots) and the function $|\beta^-| = 1/|\beta^+|$ (blue solid lines), and Wecker's (JASA, 1984) invertibility region $|\beta^+| < 1$ and $|\beta^-| < 1$ (red solid lines).



Global and Local Invertibility:

When $q > 1$, a strictly stationary and ergodic SETMA($k; q, \dots, q$) process

$Y_t = \varepsilon_t + \sum_{i=1}^k \left(\sum_{j=1}^q \psi_j^{(i)} \varepsilon_{t-j} \right) I(Y_{t-d} \in \mathbb{R}^{(i)})$ is **globally invertible** if

$\prod_{i=1}^k \rho(\Psi^{(i)})^{p_i} < 1$ with

$$\Psi^{(i)} = \begin{pmatrix} \psi_1^{(i)} & \dots & \psi_q^{(i)} \\ \mathbf{I}_{q-1} & \mathbf{0}_{(q-1) \times 1} & \mathbf{0} \end{pmatrix} \text{ and } p_i = \mathbb{E}[I(Y_{t-d} \in \mathbb{R}^{(i)})] \quad (0 < p_i < 1).$$

Example: SETMA(2; 2, 2) with $d = 1$, $\psi_1^{(1)} = 1.4$, $\psi_2^{(1)} = -0.7$, $\psi_1^{(2)} = 1.5$, $\psi_2^{(2)} = -0.5$, and $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

The corresponding 2×2 **companion matrices** $\Psi^{(i)}$ ($i = 1, 2$) have eigenvalues $\lambda_{1,2}^{(1)} = 0.7 \pm 0.4583i$ and $\lambda_{1,2}^{(2)} = 0.75 \pm 0.25i$, respectively.

Notes:

1. The MA process in the first ($Y_{t-1} \leq 0$) regime is **(locally) invertible**.
2. When $Y_{t-1} > 0$, the MA process is **not invertible** with one root on the unit circle and one root less than one.
3. $\{Y_t\}$ is **globally invertible**. With $\rho(\Psi^{(1)}) = |0.7 \pm 0.4583i| = 0.8367$ and $\rho(\Psi^{(2)}) = |0.75 \pm 0.25i| = 1$, we have $\rho(\Psi^{(1)})^{1-p_1} \times \rho(\Psi^{(2)})^{p_1} = (0.8367)^{0.4984} \times (1)^{0.5016} < 1$, where $\hat{p}_1 = 0.5016$ is an estimate of $p_1 = \mathbb{E}(Y_{t-1} \leq 0)$.



3.3 Some References

Threshold models:

- **Chen, C.W.S. et al.** (2011). A review of threshold time series models in finance. *Statistics and Its Interface*, 4, 167–181. <http://doi.org/10.4310/sii.2011.v4.n2.a12>.
- **Hansen, B.E.** (2011). Threshold autoregression in economics. *Statistics and Its Interface*, 4, 123–127. <http://doi.org/10.4310/sii.2011.v4.n2.a4>.

Stationarity and ergodicity for models with nonlinearities in **both the conditional mean and the conditional variance**:

- **Chen, M. and Chen, G.** (2000). Geometric ergodicity of nonlinear autoregressive models with changing conditional variances. *Can. J. Statistics*, 28, 605–613. <http://doi.org/10.2307/3315968>.
- **Ferrante, M., et al.** (2003). Geometric ergodicity, regularity of the invariant distribution and inference for a threshold bilinear Markov process. *Statistica Sinica*, 13, 367–384.
- **Liebscher, E.** (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *J. Time Series Analysis*, 26, 669–689. <http://doi.org/10.1111/j.1467-9892.2005.00412.x>.
- **Meitz, M. and Saikkonen, P.** (2008). Stability of nonlinear AR-GARCH models. *J. Time Series Analysis*, 29, 453–475. <http://doi.org/10.2139/ssrn.945085>.
- **Meitz, M. and Saikkonen, P.** (2010). A note on the geometric ergodicity of a nonlinear AR-ARCH model. *Statistics & Probability Letters*, 80(7-8), 631–638. <http://doi.org/10.1016/j.spl.2009.12.020>.



Ph.D. workshop

Nonlinear Time Series Analysis and Forecasting

Part II: Wednesday 15:45 – 17:30

Jan G. De Gooijer

Amsterdam School of Economics
University of Amsterdam
e-mail: j.g.degooijer@uva.nl

10th Nordic Econometric Meeting, Stockholm
May 22-23, 2019

More information: <http://www.jandegooijer.nl>



Table of Contents

1. Frequency-Domain Tests	1 – 1
1.1 Introduction	1 – 1
1.2 Subba Rao–Gabr Test Statistics	1 – 12
1.3 Hinich's Tests Statistics	1 – 14
1.4 Related Test Statistics	1 – 21
1.5 Which Test to Use?	1 – 23
1.6 Some Software References	1 – 24
1.7 Some References	1 – 25
2. Time-Domain Tests	2 – 1
2.1 Lagrange Multiplier (LM) Test Statistics: BL, ExpAR, STAR	2 – 1
2.2 Likelihood Ratio (LR) Test Statistics: Nested SETAR	2 – 4
2.3 Wald (W) Test Statistic: asMA	2 – 6
2.4 Tests Based on 2nd-order Volterra Expansion	2 – 7
2.5 Arranged Autoregression Tests	2 – 9
2.6 Some Software References	2 – 14
2.7 Some References	2 – 15
3. Model Estimation, Selection, and Checking	3 – 1
3.1 Model Estimation	3 – 1
3.2 Model Selection	3 – 8
3.3 Diagnostic Checking	3 – 9
3.4 Some Software References	3 – 13
3.5 Some References	3 – 14



1. Frequency-Domain Tests

1.1 Introduction

Bispectrum:

Suppose $\{Y_t\}_{t=1}^T$ is a time series arising from a real-valued 3rd-order strictly stationary stochastic process $\{Y_t\}$ with mean 0. The inherent strength of dependence can be measured by the ACVF, $\gamma_Y(\ell) = \mathbb{E}(Y_t Y_{t+\ell})$ ($\ell \in \mathbb{Z}$).

However, we **cannot** base tests for **nonlinearity** (nL) and **non-Gaussianity** (nG) on 2nd moments.

A direct approach is to consider moments of higher order. A useful function is the **3rd-order cumulant**, defined as $\gamma_Y(\ell_1, \ell_2) = \mathbb{E}(Y_t Y_{t+\ell_1} Y_{t+\ell_2})$, ($\ell_1, \ell_2 \in \mathbb{Z}$).

Both functions are **time invariant** and unaffected by permutations in their arguments, which creates the symmetries

$$\begin{aligned}\gamma_Y(\ell) &= \gamma_Y(-\ell), \\ \gamma_Y(\ell_1, \ell_2) &= \gamma_Y(\ell_2, \ell_1) = \gamma_Y(-\ell_1, \ell_2 - \ell_1) = \gamma_Y(\ell_1 - \ell_2, -\ell_2).\end{aligned}$$



If $\sum_{\ell=-\infty}^{\infty} |\gamma_Y(\ell)| < \infty$, then we define the *spectral density* function (*spectrum*) of $\{Y_t\}$ to be the discrete-time *Fourier transform* (FT) of the ACVF,

$$f_Y(\omega) = \sum_{\ell=-\infty}^{\infty} \gamma_Y(\ell) \exp(-2\pi i \omega \ell), \quad \omega \in [0, 1],$$

where ω denotes the frequency.

If, in addition, $\sum_{\ell_1, \ell_2=-\infty}^{\infty} |\gamma_Y(\ell_1, \ell_2)| < \infty$, then the *bispectral density* function (*bispectrum*) exists and is defined as the bivariate, or double, FT of the 3rd-order cumulant function,

$$f_Y(\omega_1, \omega_2) = \sum_{\ell_1, \ell_2=-\infty}^{\infty} \gamma_Y(\ell_1, \ell_2) \exp\{-2\pi i(\omega_1 \ell_1 + \omega_2 \ell_2)\},$$

$$(\omega_1, \omega_2) \in [0, 1]^2.$$



We have the following relations,

$$f_Y(\omega) = f_Y(-\omega),$$

$$f_Y(\omega_1, \omega_2) = f_Y(\omega_2, \omega_1) = f_Y(\omega_1, -\omega_1 - \omega_2) = f_Y(-\omega_1 - \omega_2, \omega_2).$$

The 3rd-order cumulant and the bispectrum are **mathematically equivalent**, as are the spectrum and the ACVF.

Clearly, $f_Y(\omega)$ is symmetric about 0.5. Due to the periodicity of the FT, the bispectrum in the entire plane can be determined from the values inside **one** of the **12 sectors** shown in the Figure on the next page. Therefore, it is sufficient to consider only frequencies in the 1st triangular region, which we define as the **principal domain**

$$\mathcal{D} = \{(\omega_1, \omega_2) : \omega_1 = \omega_2, \omega_1 = 0, \omega_1 = (1 - \omega_2)/2\}.$$

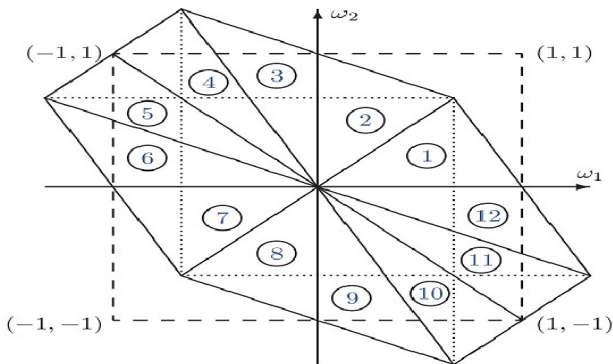


Figure: Values of $f_Y(\omega_1, \omega_2)$ defined over the entire plane, as completely specified by the values over any one of the 12 labeled sectors.



Some properties:

- If $\{X_t\}$ and $\{Y_t\}$ are two statistically independent processes and $Z_t = X_t + Y_t$, then $f_Z(\omega_1, \omega_2) = f_X(\omega_1, \omega_2) + f_Y(\omega_1, \omega_2)$.
- If $\{X_t\}$ is Gaussian and i.i.d., then $\gamma_X(l_1, l_2) = 0, \forall(l_1, l_2)$, and $f_X(\omega_1, \omega_2) = 0, \forall(\omega_1, \omega_2)$, so $f_Z(\omega_1, \omega_2) = f_Y(\omega_1, \omega_2)$, i.e. **symmetric noise is suppressed** in the bispectrum.
- The **imaginary part** of $f_Y(\omega_1, \omega_2)$ is **0 for a TR** process. Then, the 3rd-order cumulant function has the additional property that $\gamma_Y(l_1, l_2) = \gamma_Y(-l_1, -l_2)$, and hence

$$\begin{aligned} \Im\{f_Y(\omega_1, \omega_2)\} &= \sum_{l_1, l_2=-\infty}^{\infty} \gamma_Y(l_1, l_2) \sin 2\pi(\omega_1 l_1 + \omega_2 l_2) \\ &= \sum_{l_1, l_2=0}^{\infty} \gamma_Y(l_1, l_2) \{\sin 2\pi(\omega_1 l_1 + \omega_2 l_2) + \sin 2\pi(-\omega_1 l_1 - \omega_2 l_2)\} \\ &= 0, \end{aligned}$$

using the identity $\sin A + \sin B = 2 \sin \frac{A+B}{2} \cos \frac{A-B}{2}$.



Assume that the process $\{Y_t\}$ is *linear causal*, i.e.,

$$Y_t = \varepsilon_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}, \text{ where } \sum_{i=1}^{\infty} \psi_i^2 < \infty, \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2),$$

with $\psi_i = 0$ for $i < 0$. Then

$$\gamma_Y(\ell_1, \ell_2) = \mathbb{E}(\varepsilon_t^3) \sum_{j=0}^{\infty} \psi_j \psi_{j+\ell_1} \psi_{j+\ell_2}.$$

Hence

$$f_Y(\omega_1, \omega_2) = \mathbb{E}(\varepsilon_t^3) H(\omega_1) H(\omega_2) H^*(\omega_1 + \omega_2),$$

where $H(\omega) = \sum_{j=0}^{\infty} \psi_j \exp(-2\pi i \omega j)$ is known as the *transfer function* and $H^*(\omega) = H(-\omega)$ its *complex conjugate*.



For **linear process** the spectral density function reduces to

$$f_Y(\omega) = \sigma_\varepsilon^2 |H(\omega)|^2.$$

Hence, defining the **normalized bispectrum**,

$$B_Y(\omega_1, \omega_2) = \frac{f_Y(\omega_1, \omega_2)}{\sqrt{f_Y(\omega_1)f_Y(\omega_2)f_Y(\omega_1 + \omega_2)}}, \quad (\omega_1, \omega_2) \in \mathcal{D}.$$

The square modulus of the normalized bispectrum, called **frequency bicoherence** is simply

$$|B_Y(\omega_1, \omega_2)|^2 = \frac{\{\mathbb{E}(\varepsilon_t^3)\}^2}{\sigma_\varepsilon^6} \equiv \frac{\mu_{3,\varepsilon}^2}{\sigma_\varepsilon^6}, \quad (\omega_1, \omega_2) \in \mathcal{D},$$

where $\mu_{3,\varepsilon} = \mathbb{E}(\varepsilon_t^3)$.



The above property may be viewed as the **basis** of several frequency-domain tests for Gaussianity and linearity.

If $\{Y_t\}$ is **linear**, and the distribution of $\{\varepsilon_t\}$ is **symmetric**, then $\mu_{3,\varepsilon} = 0$ and so $|B_Y(\omega_1, \omega_2)|^2 \equiv 0 \forall (\omega_1, \omega_2) \in \mathcal{D}$. However, this is also true for linear Gaussian time series processes. Thus the skewness function is a **constant** if $\{Y_t\}$ is **linear** and that constant is **0** if $\{Y_t\}$ is **Gaussian**. Consequently, the null hypotheses of interest are, respectively,

$$\mathbb{H}_0^{(1)}: \quad f_Y(\omega_1, \omega_2) = 0, \quad \forall (\omega_1, \omega_2) \in \mathcal{D}; \quad \text{and}$$

$$\mathbb{H}_0^{(2)}: \quad |B_Y(\omega_1, \omega_2)|^2 = \text{constant}, \quad \forall (\omega_1, \omega_2) \in \mathcal{D}.$$



Consistent estimates of the spectrum and bispectrum can be obtained through various techniques.

- Nonparametric methods (dominate the literature): direct vs. indirect.
- Parametric or model-based methods (e.g. AR modeling).
- Criterion-based methods (e.g. maximum entropy algorithm).

Nonparametric:

The (*sample periodogram*), as a natural estimator of $f_Y(\omega)$, is defined as the discrete FT of the sample ACVF, i.e.

$$I_T(\omega) = \sum_{\ell=-(T-1)}^{T-1} \hat{\gamma}_Y(\ell) \exp\{-2\pi i\omega\ell\}, \quad \omega \in [0, \frac{1}{2}],$$

where $\hat{\gamma}_Y(\ell) = T^{-1} \sum_{t=1}^{T-\ell} Y_t Y_{t+\ell}$. However, $I_T(\omega)$ is **not** a consistent estimator of $f_Y(\omega)$. Similarly, the 3rd-order periodogram (**direct method**) is an **inconsistent** estimator of $f_Y(\omega_1, \omega_2)$.



Consistent estimators are defined as

$$\hat{f}_Y(\omega) = \sum_{\ell=-M}^M \lambda\left(\frac{\ell}{M}\right) \hat{\gamma}_Y(\ell) \exp(-2\pi i \omega \ell), \quad \omega \in [0, \frac{1}{2}],$$

$$\hat{f}_Y(\omega_1, \omega_2) = \sum_{\ell_1, \ell_2=-M}^M \lambda\left(\frac{\ell_1}{M}, \frac{\ell_2}{M}\right) \hat{\gamma}_Y(\ell_1, \ell_2) \exp\{-2\pi i(\omega_1 \ell_1 + \omega_2 \ell_2)\},$$

$$(\omega_1, \omega_2) \in \mathcal{D}.$$

Here $\hat{\gamma}_Y(\ell_1, \ell_2) = T^{-1} \sum_{t=1}^{T-\beta} Y_t Y_{t+\ell_1} Y_{t+\ell_2}$, with $\beta = \max\{0, \ell_1, \ell_2\}$, $(\ell_1, \ell_2 = 0, 1, \dots, T-1)$ and $1 \leq M \ll T$ ($M = \text{truncation point}$).

The function $\lambda(\cdot)$ is a *lag window* (e.g. **Parzen** or **Bartlett**), satisfying $\lambda(0) = 1$ and the symmetry conditions.



Example:

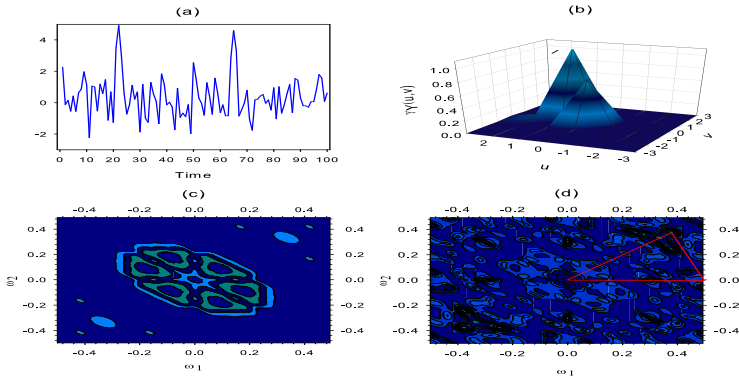


Figure: (a) A realization of the BL(0, 0, 1, 1) (diagonal) process $Y_t = 0.4Y_{t-1}\varepsilon_{t-1} + \varepsilon_t$ with $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$; (b) Three-dimensional plot of $\gamma_Y(u, v)$; (c) Contour plot of the bicoherence estimates of the BL process in (a); (d) Contour plot of the bicoherence estimates of a series generated by $Y_t = 0.4Y_{t-1} + \varepsilon_t$. Superimposed is a plot of the principal domain \mathcal{D} ; $T = 100$.



1.2 Subba Rao – Gabr Test Statistics

Testing for Gaussianity (Broad outline of steps):

- Obtain an estimate of $f_Y(\omega_1, \omega_2)$ on a set of lattice (designated) frequencies in the principle domain \mathcal{D} .
- **Tests** these quantities for **constancy** by estimating $|B_Y(\omega_1, \omega_2)|^2$.

Estimates are constructed at the N points in the fine grid $(\omega_{j_p}, \omega_{k_q})$ such that they do not overlap. Several parameters are needed for this step.

- The resulting test statistic is a complex analog of **Hotelling's** T^2 statistic for testing a multivariate mean vector.

Under $\mathbb{H}_0^{(1)}$, and as $T \rightarrow \infty$, the test statistic follows an F distribution with d.f. depending on N and $P = \sum_{i=1}^{\lfloor 2K/3 \rfloor} (K - \lfloor i/2 \rfloor - 1 - i)$ with $K \ll T$.



If $\mathbb{H}_0^{(1)}$ (Gaussianity) is **rejected**, consider testing $\mathbb{H}_0^{(2)}$ (Linearity).

Testing for linearity (Broad outline of steps):

- Obtain estimates of $|B_Y(\omega_{j_p}, \omega_{k_q})|$ at the N points in the fine grid $(\omega_{j_p}, \omega_{k_q})$.
- Place them in a $P \times N$ matrix. Average the values in the columns of this matrix to obtain a random sample of N estimates of the $P \times 1$ mean vector $\mathbf{Z} = (Z_1, \dots, Z_P)'$. These estimates are asymptotically normally distributed.
- If $\mathbb{H}_0^{(2)}$ is “true” then all the elements of \mathbf{Z} are identical. Using results from multivariate statistical analysis, a test statistic can be defined.
- Under $\mathbb{H}_0^{(2)}$, and as $T \rightarrow \infty$, the test statistic follows an F distribution with d.f. depending on N and P .

Disadvantage: Great skill is needed in applying **both** test statistics (i.e. linearity and Gaussianity) because of the **large number of parameters** involved. In particular, this applies to choice of the **lag window** $\lambda(\cdot)$, **truncation point** M , and the **placing of the grids**.



1.3 Hinich's Test Statistics

Hinich (JTSA, 1982) modifies the Subba Rao–Gabr tests to use **all** the bispectrum Fourier frequency gridpoints. The test statistics are based on a consistent estimator of the bispectrum at frequency pair (ω_m, ω_n) obtained by smoothing the 3rd-order periodogram over adjacent frequency pairs.

Let $\omega_j = (j - 1)/T$ ($j = 1, \dots, [T/2] + 1$). For each pair (j, k) ($j, k \in \mathbb{Z}$), define the complex random variable

$$F_Y(\omega_j, \omega_k) = Y(\omega_j)Y(\omega_k)Y^*(\omega_{j+k})/T,$$

where $Y(\omega_j) = \sum_{t=1}^T Y_t \exp\{-2\pi i \omega_j(t - 1)\}$. The **principal domain** of $F_Y(\omega_j, \omega_k)$ is the triangular set

$$\Delta = \{(j, k): 0 < j \leq T/2, 0 < k \leq j, 2j + k \leq T\},$$

assuming T is even.



Procedure: Obtain a consistent estimate of $f_Y(\omega_1, \omega_2)$ by averaging $F_Y(\omega_j, \omega_k)$ in a square of M^2 points, where the centers of the squares are defined by a lattice \mathcal{L} of points such that $\mathcal{L} \in \Delta$. Then

$$\hat{f}_Y(\omega_m, \omega_n) = \frac{1}{M^2} \sum_{j=(m-1)M}^{mM-1} \sum_{k=(n-1)M}^{nM-1} F_Y(\omega_j, \omega_k),$$

with $M = \lfloor T^c \rfloor$ ($\frac{1}{2} < c < 1$). An estimator of $f_Y(\omega_m)$ is given by

$$\hat{f}_Y(\omega_m) = \frac{1}{M} \sum_{j=(m-1)M}^{mM-1} |Y(\omega_j)|^2.$$



Hinich (JTSA, 1982) shows that if $\{Y_t\}$ is **linear**, then for the boxes entirely within the triangular set Δ , the distribution of the statistic

$$\widehat{B}_Y(\omega_m, \omega_n) = \frac{\widehat{f}_Y(\omega_m, \omega_n)}{\{T^{1-4c} Q_{m,n} \widehat{f}_Y(\delta_m) \widehat{f}_Y(\delta_n) \widehat{f}_Y(\delta_{m+n})\}^{1/2}}$$

is complex normal with unit variance. $Q_{m,n}$ is the # of (j, k) in the squares that are in Δ but **not** on the boundaries $j = k$ or $(2j + k) = T$, plus twice the # on these boundaries, and $\delta_x = (2x - 1)M/2T$.

The estimates $2|\widehat{B}_Y(\omega_m, \omega_n)|^2$ are asymptotically independent random variables having the noncentral $\chi_2^2(\lambda_{m,n})$ distribution, with two d.f. and noncentrality parameter

$$\lambda_{m,n} = 2(T^{1-4c} Q_{m,n})^{-1} |B_Y(\omega_m, \omega_n)|^2 \geq 2T^{2c-1} |B_Y(\omega_m, \omega_n)|^2.$$

c (in practice c is slightly $> 1/2$) controls the bias-variance trade-off of $\widehat{B}_Y(\cdot, \cdot)$: $c = 1/2$ **small bias**, and $c = 1$ **small variance**.



Testing for linearity:

If $\{Y_t\}$ follows a zero-mean stationary **linear** (L) process, then

$$\lambda_{m,n} = 2T^{2c-1} \frac{\mu_{3,\varepsilon}^2}{\sigma_\varepsilon^6} \equiv \lambda_0 \quad (\text{a constant}).$$

One way to proceed is to use the asymptotic normality of the **interquartile range** (IQR_M) of the $2|\widehat{B}_Y(\omega_m, \omega_n)|$'s entirely within Δ .

Let $q_{0.75} - q_{0.25}$ denote the IQR from a $\chi_2^2(\lambda_0)$ distribution. Then, under $\mathbb{H}_0^{(2)}$ (**Linearity**), the approximate distribution of IQR_M , as deduced from the theory of order statistics, is given by

$$Z_{\text{IQR}}^L = \frac{\text{IQR}_M - (q_{0.75} - q_{0.25})}{\sigma_0} \xrightarrow{D} \mathcal{N}(0, 1), \text{ as } T \rightarrow \infty,$$

where

$$\sigma_0^2 = \frac{3[f_{\chi_2^2(\lambda_0)}(q_{0.25})]^{-2} - 2[f_{\chi_2^2(\lambda_0)}(q_{0.25})f_{\chi_2^2(\lambda_0)}(q_{0.75})]^{-1} + 3[f_{\chi_2^2(\lambda_0)}(q_{0.75})]^{-2}}{16P},$$

and $f_{\chi_2^2(\lambda_0)}(\cdot)$ is the density function of a $\chi_2^2(\lambda_0)$ random variable.

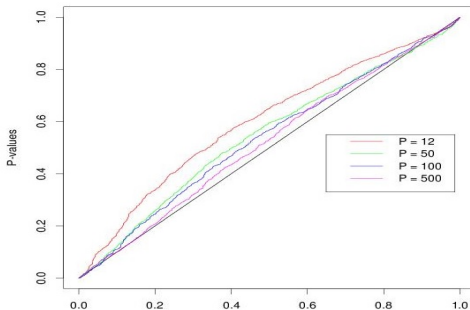


Figure: Sorted p -values of *Hinich's* test for linear time series together with the line $y = x$ (black solid line) for 1,000 replications. $P = 12, 50, 100$ and 500 correspond to time series of lengths $T = 144, 600, 1,200,$ and $6,000$ when using $M = T^{1/2}$.

It is clear that, even for large P , the p -values of *Hinich's* test statistic do **not** follow a $U(0, 1)$ distribution.



Testing for Gaussianity:

If the error process $\{\varepsilon_t\}$ in the linear DGP is **Gaussian (G)**, then $\lambda_0 \equiv 0$. In that case the following test statistic may be used

$$T^G = 2 \sum_{(m,n) \in \mathcal{L}} |\hat{B}_Y(\omega_m, \omega_n)|^2,$$

which is asymptotically distributed as a central χ_{2P}^2 variate under $\mathbb{H}_0^{(2)}$ (**Linearity**), with $P \approx T^2/(12M^2)$.

$P = 100$ requires a series of $T = 1,000$ when $M = \lfloor T^c \rfloor$ ($\frac{1}{2} < c < 1$).

Disadvantage: Power linearity test is low.



1.4 Related Tests Statistics

- Goodness-of-fit test statistics (Anderson–Darling (AD) and Cramér–von Mises (CvM)). Jahan and Harvill (*Comm. Stat.: Theory & Methods*, 2008) approximate the noncentral $\chi_2^2(\cdot)$ by a normal distribution.

Disadvantage: Sensitive to user-specified parameters.

- Maximal standardized *interdecile* (IDR) fractile statistic (Rusticelli et al., *Econometric Reviews*, 2009) defined as

$$MD_{IDR}^L = \max_{M^L \leq M \leq M^H} \{IDR_M\},$$

where

$$IDR_M = \frac{\{f_{\chi_2^2(\lambda_{m,n})}(q_{0.9}) - f_{\chi_2^2(\lambda_{m,n})}(q_{0.1})\} - \{f_{\chi_2^2(\hat{\lambda}_0)}(q_{0.9}) - f_{\chi_2^2(\hat{\lambda}_0)}(q_{0.1})\}}{\hat{\sigma}_0}.$$

Disadvantage: The test statistic relies on the asymptotic normality of the bispectrum.



1.4 Related Test Statistics (Contd.)

- Bootstrapped-based tests (Berg et al., *J. Stat. Plann. Inference*, 2010) for L+nG, L+S (S=symmetry), and G. Uses a 'kernelized' form of Hinich's test statistic for linearity.

Disadvantage: Choice of M , P , and order p of a prewhitening $AR(p)$ process.

Advantage: No asymptotic distributions are utilized.

- MSFE-based linearity test (Terdik and Máth, *JTSA*, 1998). If the process is nG, the one-step ahead quadratic forecast has a smaller asymptotic MSFE than the one-step ahead linear forecast.

Disadvantages: (i) low power, (ii) only examines 2nd-order features in departure from the null hypothesis, (iii) depends on the order p of the fitted AR process.



1.5 Which Test to Use?

- Bootstrapped test statistics give generally **better power** results than **Hinich's** G and L tests. The "classical" **Hinich** L test gives **poor** answers for very short series (< 100 observations).
- The maximal IDR test Z_{IDR}^L has the **largest** power improvement over the **Hinich** L test.
- The power of the two goodness-of-fit tests (AD and CvM) for testing Gaussianity is **comparable** with the **Hinich** G test statistic, but often higher.
- There is **no** frequency-domain test statistic which uniformly outperforms all other tests for all DGPs and sample size considered in the literature.

[**Tentative recommendation:** Use model-based bootstrap method.]



1.6 Some Software References (go to <http://www.jandegooijer.nl>)

- Subba Rao–Gabr (1984, Appendices): F77.
- Hinich's test statistics: <http://www.la.utexas.edu/hinich/> (F77). <http://ashleymac.econ.vt.edu> (a user-friendly executable version). See also the MATLAB toolbox HOSA-GLSTAT.
- AD and CvM test statistics: goodnessfit.m.
- BS form of Hinich's linearity test: R-code.
- Terdik-Máth MSFE linearity test: TerM.m.



1.7 Some References

- **Berg, A., Paparoditis, E., and Politis, D.N. (2010).** A bootstrap test for time series linearity. *J. Stat. Planning & Inference*, 140, 3841–3857.
<http://doi.org/10.1016/j.jspi.2010.04.047>.
- **Hinich, M.J. (1982).** Testing for Gaussianity and linearity of stationary time series. *J. Time Series Analysis*, 3, 169–176.
<http://doi.org/10.1111/j.1467-9892.1982.tb00339.x>.
- **Jahan, N. and Harvill, J.L. (2008).** Bispectral-based goodness-of-fit tests of Gaussianity and linearity of stationary time series. *Comm. Stat.: Theory & Methods*, 37, 3216–3227. <http://doi.org/10.1080/03610920802133319>.
- **Rusticelli, E., Ashley, R.A., Dagum, E.B., and Patterson, D.G. (2009).** A new bispectral test for nonlinear serial dependence. *Econometric Reviews*, 28, 279–293. <http://doi.org/10.1080/07474930802388090>.
- **Subba Rao, T. and Gabr, M.M. (1984).** *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer-Verlag, New York.
<http://doi.org/10.1007/978-1-4684-6318-7>.
- **Terdik, G. and Máth, J. (1998).** A new test of linearity of time series based on the bispectrum. *J. Time Series Analysis*, 19, 737–753.
<http://doi.org/10.1111/1467-9892.00120>.



2. Time-Domain Tests

2.1 Lagrange Multiplier (LM) Test Statistics

Consider the nL process

$$Y_t = g(Y_{t-1}, \dots, Y_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}; \boldsymbol{\theta}) + \varepsilon_t,$$

where $g(\cdot)$ is a sufficiently well-behaved function on \mathbb{R} . Assume the model nests a **linear (L)** time series model. This implies that

$$\underbrace{\boldsymbol{\theta}}_{\nu \times 1} = \left(\underbrace{\boldsymbol{\theta}'_1}_{\nu_1 \times 1(L)}, \underbrace{\boldsymbol{\theta}'_2}_{\nu_2 \times 1(nL)} \right)',$$

with $\mathbb{H}_0: \boldsymbol{\theta}_2 = \mathbf{0}$. Define $\mathbf{z}_t(\boldsymbol{\theta}) = \partial \varepsilon_t(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ and denote $\widehat{\mathbf{z}}_t = \mathbf{z}_t(\widehat{\boldsymbol{\theta}})$ and $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\theta}})$ with $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}'_1, \mathbf{0}')'$. Partitioning $\widehat{\mathbf{z}}_t$ conformably to $\boldsymbol{\theta}$ yields $\widehat{\mathbf{z}}_t = (\widehat{\mathbf{z}}'_{1,t}, \widehat{\mathbf{z}}'_{2,t})'$. The LM test statistic is given by

$$\text{LM}_T = \widehat{\sigma}_\varepsilon^{-2} \left(\sum_{t=1}^T \widehat{\mathbf{z}}_{2,t} \widehat{\varepsilon}_t \right)' \left(\widehat{\boldsymbol{\Sigma}}_{22} - \widehat{\boldsymbol{\Sigma}}_{21} \widehat{\boldsymbol{\Sigma}}_{11}^{-1} \widehat{\boldsymbol{\Sigma}}_{12} \right)^{-1} \left(\sum_{t=1}^T \widehat{\mathbf{z}}_{2,t} \widehat{\varepsilon}_t \right),$$

where

$$\widehat{\boldsymbol{\Sigma}}_{21} = \widehat{\boldsymbol{\Sigma}}'_{12} = \sum_t \widehat{\mathbf{z}}_{2,t} \widehat{\mathbf{z}}'_{1,t}, \quad \widehat{\boldsymbol{\Sigma}}_{ii} = \sum_t \widehat{\mathbf{z}}_{i,t} \widehat{\mathbf{z}}'_{i,t}, \quad (i = 1, 2), \quad \text{and} \quad \widehat{\sigma}_\varepsilon^2 = T^{-1} \sum_t \widehat{\varepsilon}_t^2.$$



Under \mathbb{H}_0 : $\text{LM}_T \xrightarrow{D} \chi_{\nu_2}^2$ as $T \rightarrow \infty$.

Auxiliary regression:

$$\hat{\varepsilon}_t = \hat{\mathbf{z}}'_{1,t} \boldsymbol{\beta}_1 + \hat{\mathbf{z}}'_{2,t} \boldsymbol{\beta}_2 + \eta_t,$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are artificial parameter vectors of dimension ν_1 and ν_2 respectively, and $\{\eta_t\}$ is an **artificial error process**.

Let SSE be the residual SSs in the linear regression, and SSE_0 the residual SSs under the null hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}$. Then, applying standard LS regression theory, we have

$$\text{LM}_T = T \left(\frac{\text{SSE}_0 - \text{SSE}}{\text{SSE}_0} \right).$$

Use the above formulation as a **first step** to derive various variants of LM test statistics. These variants depend on the form of the vector $\hat{\mathbf{z}}_{2,t}$, which is determined by the **type of nonlinearity investigated**.



Table: Asymptotic distribution of some LM-type linearity test statistics.

Model	\mathbb{H}_0	$\text{LM}_T \xrightarrow{D}$
BL: $Y_t = \phi_0 + \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t$ $+ \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{u=1}^p \sum_{v=1}^Q \psi_{uv} Y_{t-u} \varepsilon_{t-v}$	$\psi_{uv} = 0$	$\chi_{PQ-r(r+1)/2}^2$ $[r = \max\{0, \min(P - q, Q - p) - 1\}]$
ExpAR: $Y_t = \phi_0 + \sum_{i=1}^p \{\phi_i + \xi_i \exp(-\gamma Y_{t-d}^2)\} Y_{t-i} + \varepsilon_t$	$\gamma = 0$	χ_p^2
STAR-1 [1st-order Taylor approximation of $G(Y_{t-d}; \gamma, c)$] $Y_t = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{j=1}^p \sum_{j=i}^p \beta_{ij} Y_{t-i} Y_{t-j}$ $+ \eta_t$	$\beta_{ij} = 0$	$\chi_{\frac{1}{2}p(p+1)}^2$
STAR-3 [3rd-order Taylor approximation of $G(Y_{t-d}; \gamma, c)$] $Y_t = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{i=1}^p \sum_{j=i}^p \beta_{ij} Y_{t-i} Y_{t-j}$ $+ \sum_{i,j=1}^p \psi_{ij} Y_{t-i} Y_{t-j}^2 + \sum_{i,j=1}^p \kappa_{ij} Y_{t-i} Y_{t-j}^3$ $+ \eta_t$	$\beta_{ij} = 0,$ $\psi_{ij} = 0, \kappa_{ij} = 0$	$\chi_{\frac{1}{2}(p(p+1)+2p^2)}^2$
STAR-1 [augmented] $Y_t = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{i=1}^p \sum_{j=i}^p \phi_{ij} Y_{t-i} Y_{t-j}$ $+ \sum_{i=1}^p \psi_i Y_{t-i}^3 + \eta_t$	$\phi_{ij} = 0,$ $\psi_i = 0$	$\chi_{\frac{1}{2}p(p+1)+p}^2$

F test statistics for nL-MA type models, NCTAR and AR-NN are available.



2.2 Likelihood Ratio (LR) Test Statistics

$$\text{SETAR: } Y_t = \phi_0^{(1)} + \sum_{i=1}^p \phi_i^{(1)} Y_{t-i} + \{\phi_0^{(2)} + \sum_{i=1}^p \phi_i^{(2)} Y_{t-i}\} I(Y_{t-d} \leq r) + \varepsilon_t,$$

with p, d known ($1 \leq d \leq p$), $r \in \tilde{\mathbb{R}} = [\underline{r}, \bar{r}]$, with \underline{r} and \bar{r} finite constants.

Let $\phi_i = (\phi_0^{(i)}, \dots, \phi_p^{(i)})'$ ($i = 1, 2$), and $\theta = (\phi_1', \phi_2')'$. Parameter space: $\Theta = \Theta_{\phi_1} \times \Theta_{\phi_2}$, where Θ_{ϕ_1} and Θ_{ϕ_2} are compact subsets of \mathbb{R}^{p+1} . Suppose the true parameter vector $\theta_0 = (\phi_{10}', \phi_{20}')'$, is an interior point of Θ .

Hypotheses of interest are

$$\mathbb{H}_0: \phi_{20} = \mathbf{0}, \quad \mathbb{H}_1: \phi_{20} \neq \mathbf{0} \text{ for some } r \in \tilde{\mathbb{R}}.$$

Observe that SETAR = linear regression model with the $p + 1$ vector of added variables $\mathbf{X}_t' I(Y_{t-d} \leq r)$ with $\mathbf{X}_t = (1, Y_{t-1}, \dots, Y_{t-p})'$. Then, with $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2)$, a LR-type test statistic is given by

$$F_T = T \left(\frac{\sup_{r \in \tilde{\mathbb{R}}} \{ \text{SSE}_0 - \text{SSE}_1(\hat{\phi}_{2T}(r), r) \}}{\inf_{r \in \tilde{\mathbb{R}}} \text{SSE}_1(\hat{\phi}_{2T}(r), r)} \right),$$

where SSE_0 and SSE_1 are the SSs of residuals under \mathbb{H}_0 and \mathbb{H}_1 ,

$\hat{\phi}_{1T} = \arg \min_{\phi_1 \in \Theta_{\phi_1}} \sum_t \hat{\varepsilon}_t^2(\phi_1)$, $\hat{\phi}_{2T} = \arg \min_{\phi_2 \in \Theta} \sum_t \hat{\varepsilon}_t^2(\phi_2, r)$,

$\hat{\varepsilon}_t(\phi_1) = \hat{\varepsilon}_t(\theta, -\infty)$, and $\hat{\varepsilon}_t(\phi_2, r)$ is based on the iterative SETAR equation.



For **nested SETARs**: generalize the F test statistic to a SETAR($k; p, \dots, p$) model ($k \geq 2$). Let $\mathbf{X}_t = (1, Y_{t-1}, \dots, Y_{t-p})'$ be a $(p+1) \times 1$ vector. Rewrite the k -regime SETAR model as

$$Y_t = \phi_1' \mathbf{X}_t I_t^{(1)}(\mathbf{r}, d) + \dots + \phi_k' \mathbf{X}_t I_t^{(k)}(\mathbf{r}, d) + \varepsilon_t, \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2),$$

where $\mathbf{r} = (r_1, \dots, r_{k-1})'$, $r_0 = -\infty$, $r_k = \infty$, and $I_t^{(i)}(\mathbf{r}, d) = I(r_{i-1} < Y_{t-d} \leq r_i)$ ($i = 1, \dots, k$).

When $k = 1$: SETAR($1; p$) (\equiv AR(p)) with 0 thresholds, being the **most restrictive** within the class of k -regime SETARs.

The models within this class are strictly **nested**. This simply means that the i -regime SETAR being tested, the \mathbb{H}_0 is a special case of the alternative SETAR($j; p, \dots, p$) model ($i < j; i = 1, \dots, k$) against which it is being tested.

Again, use an F test statistic and bootstrap its distribution.



2.3 Wald (W) Test Statistic: asMA

$$Y_t = \mu + \varepsilon_t + \sum_{j=1}^q \theta_j^+ \varepsilon_{t-j} + \sum_{j=1}^q \delta_j I(\varepsilon_{t-j} \leq 0) \varepsilon_{t-j}, \quad \delta_j = \theta_j^- - \theta_j^+.$$

The \mathbb{H}_0 of **symmetry** is equivalent to testing the restriction $\boldsymbol{\theta}^+ = \boldsymbol{\theta}^-$, where $\boldsymbol{\theta}^+ = (\theta_1^+, \dots, \theta_q^+)'$, and $\boldsymbol{\theta}^- = (\theta_1^-, \dots, \theta_q^-)'$.

Let $\boldsymbol{\theta} = (\mu, (\boldsymbol{\theta}^+)', (\boldsymbol{\theta}^-)')'$, and \mathbf{R} is a $q \times (2q + 1)$ **restriction matrix** such that $\mathbf{R}\boldsymbol{\theta} = \mathbf{r}$, and \mathbf{r} is a $(2q + 1)$ -vector. From the partition $\mathbf{R} = (\mathbf{R}_1 : \mathbf{R}_2)$, where $\mathbf{R}_1 = \mathbf{0}_{q \times 1}$, the problem becomes one of testing the null hypothesis

$$\mathbb{H}_0: \mathbf{R}_2 \boldsymbol{\theta} = \mathbf{0} \quad \text{against} \quad \mathbb{H}_1: \mathbf{R}_2 \boldsymbol{\theta} \neq \mathbf{0}.$$

W test statistic:

$$W_T = (\mathbf{R}_2 \hat{\boldsymbol{\theta}})' \left[\mathbf{R} \hat{\mathcal{H}}_T^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{R}' \right]^{-1} \mathbf{R}_2 \hat{\boldsymbol{\theta}},$$

where $\hat{\mathcal{H}}$ is the **empirical Hessian**, and $\hat{\boldsymbol{\theta}}$ is an unrestricted estimate of $\boldsymbol{\theta}$. Under \mathbb{H}_0 , and as $T \rightarrow \infty$, W_T has an asymptotic χ_q^2 distribution.



2.4 Tests Based on 2nd-order Volterra Expansion

Linearity (\mathbb{H}_0) versus the alternative hypothesis (\mathbb{H}_1):

$$Y_t = \mu + \varepsilon_t + \sum_{u=-\infty}^{\infty} \psi_u \varepsilon_{t-u} + \sum_{u,v=-\infty}^{\infty} \psi_{uv} \varepsilon_{t-u} \varepsilon_{t-v}, \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2).$$

Tukey's (T) nonadditivity test (Keenan, *Biometrika*, 1985):

- (i) Choose $p \in [4, 8]$. Regress Y_t on $\{1, Y_{t-1}, \dots, Y_{t-p}\}$; compute the fitted values $\{\hat{Y}_t\}$, the residuals $\{\hat{\varepsilon}_t\}_{t=p+1}^T$, and $\text{SSE} = \sum_t \hat{\varepsilon}_t^2$.
- (ii) Regress $\{\hat{Y}_t^2\}$ on $\{1, Y_{t-1}, \dots, Y_{t-p}\}$; compute the residuals $\{\hat{\xi}_t\}_{t=p+1}^T$.
- (iii) Regress $\hat{\varepsilon}_t$ on $\hat{\xi}_t$.
- (iv) From the regression in (iii) calculate the test statistic

$$F_T^{(T)} = \frac{\hat{\eta}^2}{(\text{SSE} - \hat{\eta}^2)/(T - 2p - 2)},$$

where $\hat{\eta} = \hat{\eta}_0 \left(\sum_t \hat{\xi}_t^2 \right)^{1/2}$ with $\hat{\eta}_0$ the regression coefficient in step (ii).

Under \mathbb{H}_0 , $F_T^{(T)} \xrightarrow{D} F_{\nu_1, \nu_2}$ with $\nu_1 = 1$ and $\nu_2 = (T - p) - (p + 1) - 1$.

The estimated size can be improved by using $T - p$ instead of $T - 2p - 2$.



Advantage: Tukey's nonadditivity test $F_T^{(T)}$ is easy to implement involving little subject choice of parameters. Reasonably powerful in finite samples.

Original F test ($F_T^{(O)}$):

This F test statistic is a direct modification of the *original* (O) Tukey nonadditivity-type test statistic. The test considers the residuals of regressions that include the individual nL terms and quadratic terms up to 3rd-order $\{Y_{t-1}^2, Y_{t-1}Y_{t-2}, \dots, Y_{t-1}Y_{t-p}, Y_{t-2}^2, Y_{t-2}Y_{t-3}, \dots, Y_{t-p}^3\}$ while $F_T^{(T)}$ considers the residuals of regressions on only the **squared** terms.

Advantage: $F_T^{(O)}$ is **more powerful** than $F_T^{(T)}$ for BL-type nonlinearity.

Augmented F test ($F_T^{(A)}$):

Extends the $F_T^{(O)}$ test statistic by including the regression of the **cubic** terms $\{Y_t^3\}$ on $(1, Y_{t-1}, \dots, Y_{t-p})$ in the set of regressions in steps (ii) – (iv).



2.5 Arranged Autoregression Tests

Introduction

Consider a SETAR(2; p, p) process. Given $\{Y_t\}_{t=1}^T$, Y_{t-d} can assume the values $\{Y_i\}_{i=h}^{T-d}$, where $h = \max\{1, p + 1 - d\}$.

Let τ_j be the time index of the j th **smallest** observation among $\{Y_i\}_{i=h}^{T-d}$. Assume that the recursive ARs begin with a minimum # of start-up values, say $n_{\min} > p + 1$. Denote the resulting ordered time series by $\{Y_{\tau_j}\}_{j=n_{\min}+1}^{T-d-h+1}$. Then we can write the SETAR as

$$Y_{\tau_j+d} = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^p \phi_i^{(1)} Y_{\tau_j+d-i} + \varepsilon_{\tau_j+d}, & (j = n_{\min}+1, \dots, s), \\ \phi_0^{(2)} + \sum_{i=1}^p \phi_i^{(2)} Y_{\tau_j+d-i} + \varepsilon_{\tau_j+d}, & (j = s+1, \dots, T-d-h+1), \end{cases}$$

where s satisfies $Y_{\tau_s} < r \leq Y_{\tau_{s+1}}$.

Estimate the parameters sequentially through **recursive LS** (Ertel and Fowlkes, JASA, 1976; Tsay, JASA, 1989) giving standardized predictive residuals. [Alternatively, use the **Kalman filter**].



CUSUM test for SETAR(2; p, p):

- (i) Choose the AR order p , the lag d , and $n_{\min} > p + 1$ of start-up values.
- (ii) For $n_{\min} \leq r \leq T - p$, find the recursive LS estimates; compute the standardized predictive residuals e_{τ_j+d} ($j = n_{\min} + 1, \dots, T - d - h + 1$; $h = \max\{1, p + 1 - d\}$).
- (iii) Compute $Z_j = \sum_{i=n_{\min}+1}^j \hat{e}_i$, ($j = n_{\min} + 1, \dots, T - d - h + 1$), and the associated CUSUM test statistic

$$Q_T = \max_{n_{\min}+1 \leq j \leq T-d-h+1} |Z_j| / \sqrt{T^*},$$

where $T^* = T - d - h + 1 - n_{\min}$. Under mild conditions on the noise process $\{\varepsilon_t\}$, the limiting distribution of Q_T is given by

$$\mathbb{P}((Q_T / \sqrt{T^*}) \leq \alpha) \equiv \sum_{j=-\infty}^{\infty} (-1)^j [\Phi(\alpha(2j+1)) - \Phi(\alpha(2j-1))],$$

where $\Phi(\cdot)$ is the normal distribution function, and α the nominal significance level.

Advantage: Easy to implement since it does **not** require estimation of the SETAR model under \mathbb{H}_a .



F test for SETAR:

- (i) Perform the arranged AR, and calculate $\hat{e}_{\tau_{j+1}+d}$.
- (ii) Compute a 2nd regression with the predictive residuals on Y_{τ_j+d} , i.e.

$$\hat{e}_{\tau_j+d} = \beta_0 + \sum_{i=1}^p \beta_i Y_{\tau_j+d-i} + \omega_{\tau_j+d}, \quad (j = n_{\min} + 1, \dots, T - d - h + 1).$$

- (iii) Compute the associated test statistic

$$F_T^* = \frac{[\sum_t \hat{e}_t^2 - \sum_t \hat{\omega}_t^2]/(p+1)}{\sum_t \hat{\omega}_t^2/(T-d-n_{\min}-p-h)},$$

where $\hat{\omega}_t$ is the LS residual of the regression in step (ii). Under the null hypothesis of linearity, and as $T \rightarrow \infty$,

$$F_T^* \xrightarrow{D} F_{\nu_1, \nu_2},$$

with d.f. $\nu_1 = p + 1$ and $\nu_2 = T - d - n_{\min} - p - h$. Furthermore, $(p + 1)F_T^*$ is asymptotically a χ_ν^2 random variable with $\nu = p + 1$ d.f.

Advantage: More power than CUSUM.



New (N) F test for BL, STAR, and ExpAR:

- (i) For a given delay d , fit recursively an arranged $AR(p)$ to $\{Y_t\}_{t=1}^T$ and calculate the standardized predictive residuals $\{\hat{\epsilon}_t\}_{t=n_{\min}+1}^T$.
- (ii) Calculate $SSE_0 = \sum_t \hat{\epsilon}_t^2$.
- (iii) Regress $\hat{\epsilon}_t$ on $\{1, Y_{t-1}, \dots, Y_{t-p}\}$, $\{Y_{t-i}\hat{\epsilon}_{t-i}, \hat{\epsilon}_{t-i}\hat{\epsilon}_{t-i-1}\}$ ($i = 1, \dots, p$), and $\{Y_{t-1} \exp(-\gamma Y_{t-1}), \Phi(z_{t-d}), Y_{t-1}\Phi(Y_{t-d})\}$, where $z_t = (Y_{t-d} - \bar{Y}_d)/s_d$ with \bar{Y}_d, s_d are the sample mean and standard deviation of the Y_{t-d} , respectively. Calculate the residual sum of squares from this regression, $SSE_1 = \sum_t \hat{\omega}_t^2$.
- (iv) The associated test statistic is given by

$$F_T^{(N)} = \frac{(SSE_1 - SSE_0)/[3(p+1)]}{SSE_0/[T - n_{\min} - 3(p+1)]}.$$

Under \mathbb{H}_0 , and as $T \rightarrow \infty$,

$$F_T^{(N)} \xrightarrow{D} F_{\nu_1, \nu_2},$$

with $\nu_1 = 3(p+1)$ and $\nu_2 = T - n_{\min} - 3(p+1)$ d.f.



Summary of size and power studies (De Gooijer, 2017, Appendix 5B):

“Positive points”

- A test statistics is effective at identifying the type of nL it is designed to detect.
- The form of the nL functional relationship in the state-dependent model is of less importance with tests based on the LR, LM, and W testing principles. Finding the **correct dimension** (order) of the state vector is more likely to be the **key factor** (Pitarakis, *Studies in Nonlinear Dynamics & Econometrics*, 2006).

“Negative points”

- There is **no** test statistic that can be used as an overall tool against any type of nL. Nevertheless, all LM-type tests seem to have reasonable size and power properties.
- The presence and size of an **intercept** in a nL model seems to have a **considerable influence** on the size and power of a test when T is **not** large. Centering data is **not** recommended since then the asymptotic null distributions are **no** longer valid.



2.6 Some Software References

- http://www.estima.com/procs_perl/mainproclistwrapper.shtml has RATS code for LM-type testing of STAR models. Also, there are codes for the arranged AR test statistic, the $F_T^{(0)}$, the F test statistic of Hansen, and the Hinich L and G test statistics.
- Hansen's webpage at <http://www.ssc.wisc.edu/~bhansen/> offers MATLAB, GAUSS, and R code (and data) to replicate some of the empirical work reported in his papers on SETAR model selection and estimation (e.g., *Econometrica*, 1996, 2000).

Based on papers written by Hansen and his co-workers, the R-tsDyn package has a host of test statistics for various forms of SETAR nonlinearity, including the BS version of the test statistics for nested SETARs.

- The function `lin.test` in the R-nlts package computes the $F_T^{(0)}$ test statistic for AR(p) processes up to order $p = 5$.



2.7 Some References

- Brännäs, K., De Gooijer, J.G., and Teräsvirta, T. (1998). Testing linearity against nonlinear moving average models. *Comm. Stat.: Theory & Methods*, 27, 2025–2035. <https://doi.org/10.1080/03610929808832207>.
- Gao, Z., Ling, S., and Tong, H. (2018), Tests for TAR models vs. STAR models – A separate family of hypotheses approach. *Statistica Sinica*. http://www3.stat.sinica.edu.tw/ss_newpaper/SS-2016-0497_na.pdf.
- Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988a). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75, 491–499. <https://doi.org/10.2307/2336599>.
- Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988b). Testing linearity in univariate time series. *Scandinavian J. Stat.*, 15, 161–175.
- Saikkonen, P. and Luukkonen, R. (1988). Lagrange multiplier tests for testing non-linearities in time series models. *Scandinavian J. Stat.*, 15, 55–68.
- Saikkonen, P. and Luukkonen, R. (1991). Power properties of a time series linearity test against some simple bilinear alternatives. *Statistica Sinica*, 2, 453–464.
- Vávra, M. (2010). Testing for non-linearity and asymmetry in time series. Ph.D. thesis, Birbeck college, University of London, UK. Available at: <http://bbktheses.da.ulcc.ac.uk/97/1/final%20Marian%20Vavra.pdf>.



3. Model Estimation, Selection, and Checking

3.1 Model Estimation: Introduction

Model:

$$Y_t = g(Y_{t-1}, \dots, Y_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}; \theta_g) + \eta_t,$$

where

$$\eta_t = h(Y_{t-1}, \dots, Y_{t-u}, \varepsilon_{t-1}, \dots, \varepsilon_{t-v}; \theta_h)^{1/2} \varepsilon_t.$$

Here $\{Y_t\}$ is a strictly stationary and ergodic univariate stochastic process; $g(\cdot; \theta_g)$ and $h(\cdot; \theta_h)$ are two real-valued measurable (**known**) functions on \mathbb{R}^{p+q} and \mathbb{R}^{u+v} ($u \leq p$), respectively; and $\theta = (\theta'_g, \theta'_h)'$ is a vector of unknown parameters that we wish to estimate, and we have available a set of observations $\{Y_t\}_{t=1}^T$ with which to do so.

Further, we assume that $h(\cdot; \theta)$ is a non-negative function of past Y_t 's and ε_t 's.



Quasi ML estimator:

Let $p^* = p \vee u$, $q^* = q \vee v$, $\mathbf{Y}_0 = (Y_0, \dots, Y_{1-p^*})'$ be the initial starting values of $\{Y_t\}$, and $\varepsilon_0 = (\varepsilon_0, \dots, \varepsilon_{1-q^*})'$ be the starting innovations. In addition, let $\theta_0 = (\theta'_{0,g}, \theta'_{0,h})'$ denote the true value of the parameter vector θ , and $\mathbf{Y}_t = (Y_1, \dots, Y_t)'$. Assume that θ_0 belongs to $\Theta = \Theta_{\theta_g} \times \Theta_{\theta_h} \subset \mathbb{R}^{p+q} \times \mathbb{R}^{u+v}$.

Conditional mean and variance of $\{Y_t\}$ given \mathbf{Y}_{t-1} and Θ :

$$\mathbb{E}(Y_t | \mathbf{Y}_{t-1}, \Theta) = g(Y_{t-1}, \dots, Y_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}; \theta_{0g}) \equiv \mu_t(\theta_{0,g})$$

$$\text{Var}(Y_t | \mathbf{Y}_{t-1}, \Theta) = h(Y_{t-1}, \dots, Y_{t-u}, \varepsilon_{t-1}, \dots, \varepsilon_{t-v}; \theta_{0h}) \varepsilon_t \equiv \sigma_t^2(\theta_{0,h}).$$

Given \mathbf{Y}_0 , and the density function $f_\varepsilon(\cdot)$, the (conditional) likelihood function evaluated at $\theta \in \Theta$ is equal to

$$L_T(\theta) = \prod_{t=1}^T \frac{1}{\sigma_t(\theta_h)} f_\varepsilon\left(\frac{Y_t - \mu_t(\theta_g)}{\sigma_t(\theta_h)}\right),$$

assuming $\sigma_t(\theta_h) \neq 0$.



Treat $f_\varepsilon(\cdot)$ and \mathbf{Y}_0 as unknown parameters (computationally demanding). As an alternative replace, for instance, $f_\varepsilon(\cdot)$ by the $\mathcal{N}(0, 1)$ density, and approximate $\mu_t(\boldsymbol{\theta}_g)$ and $\sigma_t(\boldsymbol{\theta}_h)$ by

$$\tilde{\mu}_t(\boldsymbol{\theta}_g) = g(Y_{t-1}, \dots, Y_1, 0, \dots; \boldsymbol{\theta}_g), \quad \tilde{\sigma}_t(\boldsymbol{\theta}_h) = h^2(Y_{t-1}, \dots, Y_1, 0, \dots; \boldsymbol{\theta}_h).$$

Then the **quasi ML** (QML) estimator $\hat{\boldsymbol{\theta}}_T$ of $\boldsymbol{\theta}_0$ is given by

$$\hat{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} \left\{ \frac{1}{T} \sum_{t=1}^T \left(\frac{Y_t - \tilde{\mu}_t(\boldsymbol{\theta}_g)}{\tilde{\sigma}_t(\boldsymbol{\theta}_h)} \right)^2 + \log \tilde{\sigma}_t(\boldsymbol{\theta}_h) \right\}.$$

If $\tilde{\sigma}_t(\boldsymbol{\theta}_h) \equiv \sigma_0^2 > 0$, the QML estimator coincides with the “classical” nonlinear least squares estimator.

The QML estimator is strongly consistent, and asymptotic normally distributed (Tjøstheim, *Stoch. Processes and their Applications*, 1986).



Conditional LS (CLS) estimator: (Ne)SETARMA, BL, STAR, multiple-regime STAR, subset SETARMA (a genetic algorithm to reduce the # of parameters).

Example: CLS-based estimation of a subdiagonal BL model

$$Y_t = \tau Y_{t-2} \varepsilon_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2).$$

$\{Y_t\}$ is a stationary, ergodic and causal process if $\sigma_Y^2 = \sigma_\varepsilon^2 / (1 - \tau^2 \sigma_\varepsilon^2)$ exists, i.e., if $\tau^2 \sigma_\varepsilon^2 < 1$. In that case it can be shown that $\{Y_t\}$ has the unique representation

$$Y_t = \varepsilon_t + \sum_{k=1}^{\infty} \tau^k \varepsilon_{t-2k} \prod_{j=0}^{k-1} \varepsilon_{t-2j-1},$$

in L_2 sense.

The CLS-based estimator of τ is given by

$$\hat{\tau} = \frac{\hat{\beta}_2(1)}{\hat{\beta}_1(0)} = \frac{\sum_{t=3}^T Y_t Y_{t-1} Y_{t-2}}{\hat{\sigma}_\varepsilon^2 \sum_{t=3}^T Y_{t-2}^2}.$$



Example: (Cont'd)

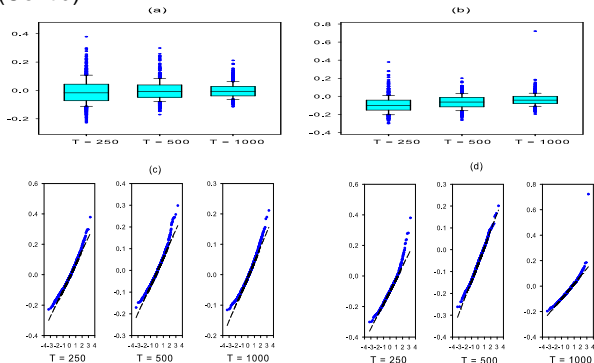


Figure: Boxplots and Q-Q plots of $\sqrt{T}(\hat{\tau} - \tau)$ for $\tau = 0.3$ (panels (a) and (c)), and $\tau = 0.5$ (panels (b) and (d)); 1,000 MC replications, and $\sigma_{\varepsilon}^2 = 1$.

Notes:

- For increasing values of τ , the nonlinearity becomes **more prominent** and as a consequence CLS estimation becomes **more difficult**.
- All distributions tend to have **negative means and medians**. Reduces with increasing values of T .



Iteratively weighted LS, IWLS (Mak, JRSS B, 1993):

Let θ be an $m \times 1$ parameter vector. Assume that θ_0 generating \mathbf{y} , a $T \times 1$ random vector of observations with density function $f(\mathbf{y}; \theta)$, belongs to an open parameter space $\Theta \subseteq \mathbb{R}^m$.

The ML estimate $\hat{\theta}$ of θ_0 follows from solving $G(\mathbf{y}, \theta) \equiv \partial \log f(\mathbf{y}; \theta) / \partial \theta = 0$.

For any $\theta, \tilde{\theta} \in \Theta$, let $g(\tilde{\theta}, \theta) = \mathbb{E}\{f(\mathbf{y}; \theta) | \tilde{\theta}\}$. Then:

- (i) Fisher's information matrix is given by $\partial g(\tilde{\theta}, \theta) / \partial \tilde{\theta} |_{\tilde{\theta}=\theta}$.
- (ii) If $\theta^{(0)}$ is a given starting value, and define in the $(u+1)$ th iteration $\theta^{(u+1)}$ ($u \geq 0$) as a root of the equation, $g(\tilde{\theta}, \theta^{(u)}) = G(\mathbf{y}, \theta^{(u)})$, then $\theta^{(u)} \rightarrow \hat{\theta}$ as $u \rightarrow \infty$. Furthermore, $|\theta^{(u)} - \hat{\theta}| = \mathcal{O}_p(T^{-u/2})$.

(ii) implies that if $g(\tilde{\theta}, \theta) = G(\mathbf{y}, \theta)$ can be solved explicitly for $\tilde{\theta}$, the algorithm provides sufficient numerical accuracy in a few iterations.

Alternatively, use the linearization

$$G(\mathbf{y}, \theta) \simeq g(\theta, \theta) + \left(\frac{\partial g(\tilde{\theta}, \theta)}{\partial \tilde{\theta}} \Big|_{\tilde{\theta}=\theta} \right)' (\tilde{\theta} - \theta) = \left(\frac{\partial g(\tilde{\theta}, \theta)}{\partial \tilde{\theta}} \Big|_{\tilde{\theta}=\theta} \right)' (\tilde{\theta} - \theta).$$

Hence,

$$\tilde{\theta} \approx \theta + \left\{ \left(\frac{\partial g(\tilde{\theta}, \theta)}{\partial \tilde{\theta}} \Big|_{\tilde{\theta}=\theta} \right)' \right\}^{-1} G(\mathbf{y}, \theta),$$

and at the $(u+1)$ th step: $\hat{\theta}^{(u+1)} = \hat{\theta}^{(u)} + \left\{ \left(\frac{\partial g(\tilde{\theta}, \theta)}{\partial \tilde{\theta}} \Big|_{\tilde{\theta}=\hat{\theta}^{(u)}} \right)' \right\}^{-1} G(\mathbf{y}, \hat{\theta}^{(u)})$.



Example: Daily Hong Kong Hang Seng Index, 2010.

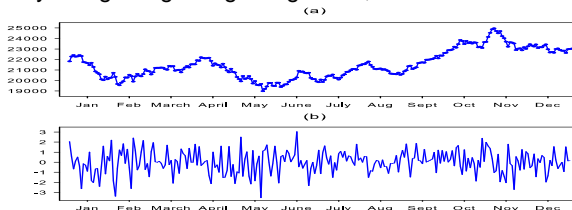


Figure: Time plots of (a) the daily closing prices P_t , and (b) $Y_t = 100(\log P_t - \log P_{t-1})^{252}$.

SETAR(3; 1, 5, 6)–TARCH(3; 1, 1, 3):

$$\text{with } Y_t = \begin{cases} 0.13 + 0.07Y_{t-1} + \varepsilon_t^{(1)} & \text{if } Y_{t-1} \leq 0.16, \\ -0.47 + 0.69Y_{t-1} + 0.02Y_{t-2} + 0.19Y_{t-3} \\ -0.37Y_{t-4} + 0.41Y_{t-5} + \varepsilon_t^{(2)} & \text{if } 0.16 < Y_{t-1} \leq 1.03, \\ 0.61 - 0.39Y_{t-1} + 0.10Y_{t-2} + 0.08Y_{t-3} \\ + 0.09Y_{t-4} - 0.16Y_{t-5} + 0.23Y_{t-6} + \varepsilon_t^{(3)} & \text{if } Y_{t-1} > 1.03, \end{cases}$$

$$\sigma_t^2 = \begin{cases} 1.29 + 0.02\varepsilon_{t-1}^2 & \text{if } Y_{t-1} \leq 0.16, \\ 0.91 + 0.73\varepsilon_{t-1}^2 & \text{if } 0.16 < Y_{t-1} \leq 1.03, \\ 0.24 + 0.02\varepsilon_{t-1}^2 + 0.07\varepsilon_{t-2}^2 + 0.13\varepsilon_{t-3}^2 & \text{if } Y_{t-1} > 1.03, \end{cases}$$

where $\varepsilon_t^{(i)} = \sigma_t^2 \varepsilon_t$ ($i = 1, 2, 3$) $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. The sample variances of $\{\varepsilon_t^{(i)}\}$ are 1.31 ($T = 138$), 1.18 ($T = 58$), and 57 ($T = 49$), respectively.



3.2 Model Selection: AIC-type model selection rules

General family of model selection criteria:

$$\min_{\theta_m \in \Theta} \{ -2 \log f_m(\mathbf{y}; \hat{\theta}_{T,m}) + p_m C(T, p_m) \},$$

where

$$C(T, p_m) = \begin{cases} \frac{2}{\frac{T-p_m}{T-p_m-1}} + 2T \log \left\{ \frac{T}{T-p_m} \right\} & \text{for AIC [Akaike, 1969, 1970, 1973, 1974, 1977],} \\ \frac{2T p_m}{T-p_m-1} + 2T \log \left\{ \frac{T}{T-p_m} \right\} & \text{for AIC}_C \text{ [Hurwich and Tsai, Biometrika, 1989],} \\ \frac{2T p_m}{T-p_m-1} + 2T \log \left\{ \frac{T}{T-p_m} \right\} & \text{for AIC}_U \text{ [McQuarrie et al., Stats.Prob. Letters, 1997],} \\ \log T & \text{for BIC [Schwarz, Annals Stats., 1978],} \\ \nu + 1 & \text{for GIC with } \nu \in [2, 5] \text{ [Bhansali and Downham, Biometrika, 1977].} \end{cases}$$

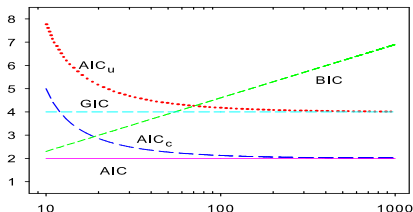


Figure: Penalty functions $C(T, p_m)$ of AIC (pink solid line), AIC_C with $p_m = 5$ (blue long dashed line), AIC_U with $p_m = 5$ (red dotted line), BIC (green short dashed line), and GIC with $\nu = 3$ (cyan medium dashed line).



3.3 Diagnostic Checking: Pearson residuals

$$\hat{\varepsilon}_t \equiv \hat{\varepsilon}_t(\hat{\boldsymbol{\theta}}_T) = (Y_t - \mathbb{E}(Y_t | \mathcal{F}_{t-1}, \hat{\boldsymbol{\theta}}_T)) / \sqrt{\text{Var}(Y_t | \mathcal{F}_{t-1}, \hat{\boldsymbol{\theta}}_T)}.$$

McLeod–Li test statistic has **high power** in detecting ARCH structures. The test has **little power** in detecting other type of nL dependencies in the residuals.

Chen (*J. Appl. Econometrics*, 2008) presents a **general framework** for testing **Pearson residuals** from the p th-order NLAR model with conditional heteroskedasticity. The model is given by

$$Y_t = g(\mathbf{Y}_{t-1}; \boldsymbol{\theta}) + \eta_t, \quad \eta_t = h(\mathbf{Y}_{t-1}; \boldsymbol{\theta})^{1/2} \varepsilon_t,$$

where $\mathbf{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})'$, and $\boldsymbol{\theta} \in \Theta$ denotes a parameter vector in a compact parameter space Θ . Here, $g(\cdot; \boldsymbol{\theta})$ and $h(\cdot; \boldsymbol{\theta})$ are twice continuously differentiable functions, and $\{\varepsilon_t\}$ is an i.i.d. WN process with moments $\mu_{1,\varepsilon} = 0$, $\mu_{2,\varepsilon} = 1$, and $\mu_{4,\varepsilon} < \infty$, where $\mu_{r,\varepsilon} = \mathbb{E}(\varepsilon_t^r)$.

Using residual autocorrelations, the null hypothesis is given by

$$\mathbb{H}_0: \{\varepsilon_t\} \text{ is an i.i.d. sequence for some } \boldsymbol{\theta}_0 \in \Theta.$$



3.3 Diagnostic Checking: Quantile residuals

Let $f(\mathbf{y}; \boldsymbol{\theta}_{0,m})$ be the true pdf of the observations $\{Y_t\}_{t=1}^T$, $\boldsymbol{\theta}_{0,m} \in \Theta \subset \mathbb{R}^m$, and $\mathbf{y} = (Y_1, \dots, Y_T)'$. For each $f: \Theta \times \mathbb{R}^T \rightarrow \mathbb{R}^+$, we can write

$$f(\mathbf{y}; \boldsymbol{\theta}_m) = \prod_{t=1}^T f_{t-1}(Y_t; \boldsymbol{\theta}_m),$$

where $f_{t-1}(Y_t; \boldsymbol{\theta}_m) \equiv f(Y_t; \boldsymbol{\theta}_m | \mathcal{F}_{t-1})$ is the conditional density function of $\{Y_t\}$ given \mathcal{F}_{t-1} , the σ -algebra generated by $\{\mathbf{Y}_0, Y_1, \dots, Y_{t-1}\}$, $\boldsymbol{\theta}_m \in \mathbb{R}^m$, and where \mathbf{Y}_0 represents the initial model values. Then, the *theoretical quantile residual* is defined by

$$R_{t,\boldsymbol{\theta}_m} = \Phi^{-1}(F_{t-1}(Y_t; \boldsymbol{\theta}_m)),$$

where $\Phi^{-1}(\cdot)$ is the inverse CDF of the $\mathcal{N}(0, 1)$ distribution, and $F_{t-1}(Y_t; \boldsymbol{\theta}_m) = \int_{-\infty}^{Y_t} f_{t-1}(u; \boldsymbol{\theta}_m) du$ is the conditional CDF of $\{Y_t\}$ (or *probability integral transform*, PIT). The corresponding *sample quantile residual* is

$$r_{t,\hat{\boldsymbol{\theta}}_T} = \Phi^{-1}(F_{t-1}(Y_t; \hat{\boldsymbol{\theta}}_T)),$$

where $\hat{\boldsymbol{\theta}}_T$ is a QML estimate of $\boldsymbol{\theta}_{0,m}$.



Steps suggested by Kalliovirta (*Econometrics J.*, 2012):

1. Transform R_{t,θ_0} by a continuously differentiable function $g: \mathbb{R}^d \rightarrow \mathbb{R}^n$ such that $\mathbb{E}(g(\mathbf{R}_{t,\theta_0})) = \mathbf{0}$, where $\mathbf{R}_{t,\theta_0} = (R_{t,\theta_0}, \dots, R_{t-d+1,\theta_0})'$, and d and n are the dimensions of the domain and range of g .
2. Conditional on \mathbf{Y}_0 , and assuming that $f_{t-1}(Y_t; \theta_m)$ exist, the log-likelihood function of the sample is given by $\ell_T(\mathbf{y}, \theta) = \sum_{t=1}^T \ell_t(Y_t, \theta) = \sum_{t=1}^T \log f_{t-1}(Y_t; \theta)$.
3. Then, under some fairly standard regularity conditions, it follows that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T g(\mathbf{R}_{t,\hat{\theta}_T}) \xrightarrow{D} \mathcal{N}_d(\mathbf{0}, \Omega),$$

where

$$\Omega = \mathbf{G}\mathcal{I}(\theta_0)^{-1}\mathbf{G}' + \Psi\mathcal{I}(\theta_0)^{-1}\mathbf{G}' + \mathbf{G}\mathcal{I}(\theta_0)^{-1}\Psi' + \mathbf{H},$$

with $\mathbf{G} = \mathbb{E}(\partial g(\mathbf{R}_{t,\theta_0})/\partial \theta')$, $\mathbf{H} = \mathbb{E}(g(\mathbf{R}_{t,\theta_0})g(\mathbf{R}_{t,\theta_0})')$, and where $\mathcal{I}(\theta_0)$ denotes the expected information matrix evaluated at θ_0 , and Ψ is a constant matrix.

The first 3 terms in the asymptotic covariance matrix Ω represent **model uncertainty** due to the effect of parameter estimation.



General test statistic:

$$S_{T,d} = \frac{1}{T-d+1} \sum_{t=1}^{T-d+1} \mathbf{g}(\mathbf{r}_{t,\hat{\theta}_T})' \hat{\Omega}_T^{-1} \sum_{t=1}^{T-d+1} \mathbf{g}(\mathbf{r}_{t,\hat{\theta}_T}),$$

where $\mathbf{r}_{t,\hat{\theta}_T} = (r_{t,\hat{\theta}_T}, \dots, r_{t-d+1,\hat{\theta}_T})'$. Under \mathbb{H}_0 , and as $T \rightarrow \infty$, the test statistic has an asymptotic χ_n^2 distribution, with $n = K_1, K_2$, or 3 depending on \mathbb{H}_0 ; see the table below.

Table: Three diagnostic test statistics based on univariate quantile residuals, as special cases of the general test statistic $S_{T,d}$.

Null hypothesis \mathbb{H}_0	Transformation function g	Test statistic
$\rho_{R_{t,\theta_0}}(\ell) = 0, \forall t,$ $(\ell = 1, \dots, K_1; K_1 \ll T)$ (Autocorrelation)	$g: \mathbb{R}^{K_1+1} \rightarrow \mathbb{R}^{K_1}$ $\mathbf{g}(\mathbf{r}_{t,\theta}) =$ $(r_{t,\theta} r_{t+1,\theta}, \dots, r_{t,\theta} r_{t+K_1,\theta})'$	$A_{T,K_1} = S_{T,d}$ with $d = K_1 + 1$
$\rho_{R_{t,\theta_0}^2}(\ell) = 0, \forall t,$ $(\ell = 1, \dots, K_2; K_2 \ll T)$ (Heteroskedasticity)	$g: \mathbb{R}^{K_2+1} \rightarrow \mathbb{R}^{K_2}$ $\mathbf{g}(\mathbf{r}_{t,\theta}) =$ $((r_{t,\theta}^2 - 1)r_{t+1,\theta}^2, \dots, (r_{t,\theta}^2 - 1)r_{t+K_2,\theta}^2)'$	$H_{T,K_2} = S_{T,d}$ with $d = K_2 + 1$
$\mathbb{E}(R_{t,\theta_0}^2 - 1, R_{t,\theta_0}^3,$ $R_{t,\theta_0}^4 - 3)' = \mathbf{0}, \forall t$ (Normality)	$g: \mathbb{R} \rightarrow \mathbb{R}^3$ $\mathbf{g}(r_{t,\theta}) = (r_{t,\theta}^2 - 1, r_{t,\theta}^3, r_{t,\theta}^4 - 3)'$	$N_T = S_{T,d}$ with $d = 1$



3.4 Some Software References

- RSTAR is a package for smooth transition AR modeling and forecasting; see http://www.researchgate.net/publication/293486017_RSTAR_A_Package_for_Smooth_Transition_Autoregressive_STAR_Modeling_Using_R.
- Smooth transition regression (STR) models can be specified, estimated and checked in the freely available, and menu-driven, computer package JMulti.
- An EViews add-in for STR analysis is available at <http://forums.eviews.com/viewtopic.php?f=23&t=11597&sid=e01abc77f3732bfcdebcbf2bce8dd1888>.
- Another option is the Ox-STR2 package at <http://www.doornik.com/download.html>, and based on Timo Teräsvirta's GAUSS code.



3.5 Some References

Model Estimation:

- Amano, T. (2009). Asymptotic efficiency of estimating function estimators for nonlinear time series models. *J. Japan Statistical Society*, 39, 209–231. <http://doi.org/10.14490/jjss.39.209>.
- Coakley, J., Fuertes, A-M., and Pérez, M-T. (2003). Numerical issues in threshold autoregressive modeling of time series. *J. Economic Dynamics & Control*, 27, 2219–2242. [http://doi.org/10.1016/s0165-1889\(02\)00123-9](http://doi.org/10.1016/s0165-1889(02)00123-9).
- Li, D. and Tong, H. (2016). Nested sub-sample search algorithm for estimation of threshold models. *Statistica Sinica*, 26, 1543-1554. <http://doi.org/10.5705/ss.2013.394t>.

Model Selection:

- Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection, *Statistics Surveys*, 4, 40-79. <http://doi.org/10.1214/09-ss054>.
- Hamaker, E.L. (2009). Using information criteria to determine the number of regimes in threshold autoregressive models. *J. Mathematical Psychology*, 53, 518–529. <http://doi.org/10.1016/j.jmp.2009.07.006>.
- Ding, J., Tarokh, V., and Yang, Y. (2017). Bridging AIC and BIC: A new criterion for autoregression. *IEEE Trans. Inform. Theory*, 64, 4024–4043. <http://ieeexplore.ieee.org/document/7953690/>.

Diagnostic Checking:

- Kalliovirta, L. and Saikkonen, P. (2010). Reliable residuals for multivariate nonlinear time series models; <http://blogs.helsinki.fi/lkvaisan/files/2010/08/ReliableResiduals.pdf>.



Ph.D. workshop

Nonlinear Time Series Analysis and Forecasting

Part III: Thursday 10:45 – 12:00

Jan G. De Gooijer

Amsterdam School of Economics
University of Amsterdam
e-mail: j.g.degooijer@uva.nl

10th Nordic Econometric Meeting, Stockholm
May 22-23, 2019

More information: <http://www.jandegooijer.nl>



Table of Contents

1. Tests for Serial Independence	1 – 1
1.1 Null hypothesis	1 – 1
1.2 Distance Measures and Dependence Functionals	1 – 3
1.3 Kernel-based Tests	1 – 13
1.4 High-Dimensional Tests	1 – 16
1.5 Distribution-based Test	1 – 21
1.6 Application	1 – 21
1.7 Some Software References	1 – 22
2. Time-Reversibility	2 – 1
2.1 Time-Domain Tests	2 – 1
2.2 Frequency-Domain Tests	2 – 4
2.3 Other Nonparametric Tests	2 – 6
2.4 Some Software References	2 – 8
3. Forecasting	3 – 1
3.1 Exact LS Methods	3 – 1
3.2 Approximate Methods	3 – 6
3.3 Forecast Intervals/Regions	3 – 15
3.4 Forecast Evaluation	3 – 16
3.5 Forecast Combination	3 – 19
3.6 Some Software References	3 – 21



1. Tests for Serial Independence

1.1 Null hypothesis (4 specifications)

Consider a time series process in \mathbb{R}^m , which at lag ℓ , is given by

$$\mathbf{Y}_t^{(\ell)} = (Y_{1,t}, \dots, Y_{m,t})' = (Y_t, Y_{t-\ell}, \dots, Y_{t-(m-1)\ell})', \quad (m \in \mathbb{Z}^+, \ell \in \mathbb{Z}),$$

Specifications:

- (1) If $\{Y_t\}$ is i.i.d., the joint pdf will be equal to the product of the individual marginals, and the null hypothesis of interest is

$$\mathbb{H}_0: f_m(\mathbf{y}) = f(y_1) \times \dots \times f(y_m), \quad \forall \mathbf{y} \in \mathbb{R}^m.$$

- (2) Moreover, if $\{\mathbf{Y}_t^{(\ell)}\}$ admits a continuous distribution function $F_m(\mathbf{y})$, the above null hypothesis can also be formulated as follows

$$\mathbb{H}_0: F_m(\mathbf{y}) = F(y_1) \times \dots \times F(y_m), \quad \forall \mathbf{y} \in \mathbb{R}^m,$$

where $F(y_i)$ is the marginal distribution of $\{Y_{t-(i-1)\ell}\}$ ($i = 1, \dots, m$).



- (3) Let $\phi_\ell(\mathbf{u}) = \mathbb{E}\{\exp(i(\sum_{k=1}^m u_k Y_{t-(k-1)|\ell|}))\}$ be the **joint characteristic function** where $\mathbf{u} = (u_1, \dots, u_m)' \in \mathbb{R}^m$. Define $D_\ell(\mathbf{u}) = \phi_\ell(\mathbf{u}) - \prod_{k=1}^m \phi(u_k)$ ($\ell = 0, \pm 1, \dots$). This expression is 0 $\forall \mathbf{u} \in \mathbb{R}^m$, if and only if there is **no** serial dependence of order $m - 1$ or, equivalently,

$$\mathbb{H}_0: D_\ell(\mathbf{u}) = 0, \quad \forall \mathbf{u} \in \mathbb{R}^m.$$

- (4) Consider $F_m(\mathbf{y}): \mathbb{R}^m \rightarrow [0, 1]$, with marginal distributions $F(y_i)$ which are assumed to be absolutely continuous. According to **Sklar's theorem** there exists an m -copula function $\mathbb{C}(\cdot)$ of $\{\mathbf{Y}_t^{(\ell)}\}$, such that $\forall \mathbf{y} \in \mathbb{R}^m$, $F_m(\mathbf{y}) = \mathbb{C}(F(y_1), \dots, F(y_m))$. The corresponding joint pdf is

$$f_m(\mathbf{y}) = c(F(y_1), \dots, F(y_m)) \prod_{i=1}^m f(y_i),$$

where $c(\mathbf{u})$, the density of the copula $\mathbb{C}(\mathbf{u})$, is given by

$$c(\mathbf{u}) = \frac{\partial^m \mathbb{C}(\mathbf{u})}{\partial u_1 \times \dots \times \partial u_m} = \frac{f_m(\mathbf{u})}{\prod_{i=1}^m f(u_i)}, \quad \mathbf{u} \in [0, 1]^m.$$

Hence, testing serial independence corresponds to the null hypothesis

$$\mathbb{H}_0: c(\mathbf{u}) = 1.$$



1.2 Distance Measures and Dependence Functionals

(I) Correlation integral

Consider the so-called *correlation integral* as a measure of spatial correlation in $\{\mathbf{Y}_t^{(\ell)}\}$ with $\ell = 1$, which we denote by $\{\mathbf{Y}_t\}$. This measure of distance is characterized by

$$C_{m,Y}(h) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} I(\|\mathbf{y} - \mathbf{x}\| \leq h) d\mu_m(\mathbf{y}) d\mu_m(\mathbf{x}),$$

where h is a **bandwidth**, depending on T , $\|\cdot\|$ a norm, and $\mu_m(\cdot)$ a **probability measure**.

If $\{\mathbf{Y}_t\}$ clusters in any dimension, then $C_{m,Y}(h)$ will take on relatively **large** values. If, however, $\{\mathbf{Y}_t\}$ is i.i.d. the correlation integral factorizes, i.e.

$$C_{m,Y}(h) = \{C_{1,Y}(h)\}^m,$$

and this equality can be used as a basis for a test statistic (i.e. **BDS test statistic**) of serial independence; see below.

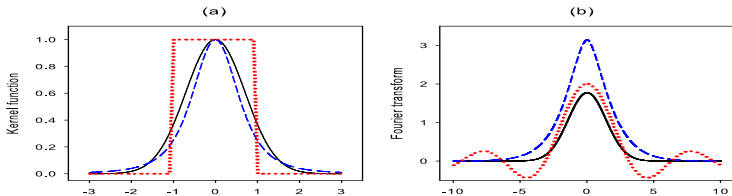


Figure: Three kernel functions (left panel) and their associated FTs (right panel): Gaussian (black solid line), squared Cauchy (blue medium dashed line), and uniform (red dotted line).

Example:

- (i) (Gaussian kernel) $\mathcal{K}(x) = e^{-x^2}$ and its FT $\tilde{\mathcal{K}}(\omega) = \sqrt{\pi}e^{-\omega^2/4}$.
- (ii) (Squared Cauchy kernel) $\mathcal{K}(x) = 1/(1+x^2)^2$ and its FT $\tilde{\mathcal{K}}(\omega) = \pi(|\omega|+1)e^{-|\omega|}$.
- (iii) (Uniform kernel) $\mathcal{K}(x) = I(|x| \leq 1)$ and its FT $\tilde{\mathcal{K}}(\omega) = (2/\omega) \sin(\omega)$.

Notes:

- The FTs of (i) and (ii) are everywhere positive. Hence, the Gaussian product kernel is positive definite and defines a quadratic form suitable for detecting any differences between a pair of distributions. Similarly, (ii) corresponds, after normalizing, to a density function.
- (iii) is not a positive definite kernel, as its FT takes negative values for certain frequencies ω .



Within the context of **serial correlation** tests, the L_2 -norm can be used as a distance measure. **Specifically**, given the m -dimensional process $\{\mathbf{Y}_t\}$, a quadratic (Q) form measuring the serial dependence is given by

$$\Delta^Q(m) = \|\mu_m^{(1)} - \mu_m^{(2)}\|^2 = (\mu_m^{(1)}, \mu_m^{(1)}) - 2(\mu_m^{(1)}, \mu_m^{(2)}) + (\mu_m^{(2)}, \mu_m^{(2)}),$$

where

$$(\mu_m^{(i)}, \mu_m^{(j)}) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} \mathcal{K}_h(\mathbf{y} - \mathbf{x}) d\mu_m^{(i)}(\mathbf{y}) d\mu_m^{(j)}(\mathbf{x}), \quad (i, j = 1, 2),$$

with $\mathcal{K}_h(\cdot)$ a nonnegative definite, spherically symmetric m -variate kernel function, and $h > 0$ a **bandwidth parameter**.

Use kernels that **factorize** as $\mathcal{K}_h(\mathbf{z}) = \prod_{i=1}^m K(z_i)/h$. Here, $K(\cdot)$ is a **one-dimensional kernel** function, which is symmetric around 0.

The functional $(\mu_m^{(1)}, \mu_m^{(1)}) - (\mu_m^{(2)}, \mu_m^{(2)})$ with the “naive” or **identity kernel** function $K_h(z) = I(|z| < h)$ corresponds to $C_{m, \gamma}(h)$.



(III) Density-based measures

Consider [specification \(1\)](#) on page 1.1. Take $m = 2$. That is, for a strictly stationary time series $\{Y_t\}$ with marginal density function $f(\cdot)$ and joint pdf $f_\ell(\cdot, \cdot)$ of $(Y_t, Y_{t-\ell})'$ ($\ell \in \mathbb{Z}$), we measure the degree of dependence by $\Delta(\ell) \equiv \Delta(f_\ell(x, y), f(x)f(y))$.

$\Delta(\cdot)$ has the following **basic properties**: (i) **nonnegativity**, (ii) **maximal information**, and (iii) **invariance** under continuous monotonic increasing transformations.

For divergence measures **not** satisfying (iii), one can obtain scale and location invariance by simply standardizing $\{Y_t\}$, assuming that the 2nd moments exist.

Or retain invariance under **continuous monotonic transformations** by transforming the data to any given marginal density function (e.g. take ranks or transform to a standard normal marginal).



The functionals considered below are all of the type

$$\Delta(\ell) = \int_{S^2} B\{f_\ell(x, y), f(x), f(y)\} f_\ell(x, y) dx dy,$$

where $B(\cdot, \cdot, \cdot)$ is a real-valued function, and the integrals are taken over the support, say S^2 , of $(Y_t, Y_{t-\ell})'$.

Divergence measures:

- Generalized Kolmogorov (K):

$$\Delta_q^K(\ell) = \left\{ \int_{S^2} |f_\ell(x, y) - f(x)f(y)|^q dx dy \right\}^{1/q}, \quad (q > 0),$$

which for $q = 1$ is the L_1 -norm. $\Delta_q^K(\cdot)$ satisfies properties (i) – (ii), but **not** (iii).

- Csiszár (C) (1967):

$$\Delta^C(\ell) = \int_{S^2} \phi \left\{ \frac{f_\ell(x, y)}{f(x)f(y)} \right\} f_\ell(x, y) dx dy,$$

where $\phi(\cdot)$ is some strictly convex function on $[0, \infty)$. Thus, $B\{z_1, z_2, z_2\} \equiv \phi(z_1/z_2 z_3)$.

- Rényi (R) (1961):

$$\Delta_q^R(\ell) = \frac{1}{q-1} \log \int_{S^2} \left\{ f_\ell(x, y) \right\}^{q-1} \left\{ f(x)f(y) \right\}^q dx dy, \quad (0 < q < 1).$$

- Tsallis (T) (1998):

$$\Delta_q^T(\ell) = \begin{cases} \frac{1}{1-q} \int_{S^2} \left\{ 1 - \left(\frac{f(x)f(y)}{f_\ell(x, y)} \right)^{1-q} \right\} f_\ell(x, y) dx dy & (q \neq 1), \\ \int_{S^2} \log \left(\frac{f_\ell(x, y)}{f(x)f(y)} \right) f_\ell(x, y) dx dy & (q = 1). \end{cases}$$

Both Rényi's (R) measure and Tsallis' (T) measure **satisfy** basic properties (i) – (iii).



Other possible candidates for measuring statistical (serial) dependence include the **difference functional** of Skaug and Tjøstheim (1993) which is given by

$$\Delta^*(\ell) = \int_{S^2} \{f_\ell(\mathbf{x}, \mathbf{y}) - f(\mathbf{x})f(\mathbf{y})\} f_\ell(\mathbf{x}, \mathbf{y}) d\mathbf{x}d\mathbf{y},$$

and the **Hellinger (H) distance** which, with $B\{z_1, z_2, z_3\} = (1 - (z_1/z_2 z_3)^{-1/2})^2$, is defined as

$$\begin{aligned} \Delta^H(\ell) &= \int_{S^2} \left\{ f_\ell^{1/2}(\mathbf{x}, \mathbf{y}) - \left(f(\mathbf{x})f(\mathbf{y}) \right)^{1/2} \right\}^2 d\mathbf{x}d\mathbf{y} \\ &= 2 - 2 \int_{S^2} \left(\frac{f(\mathbf{x})f(\mathbf{y})}{f_\ell(\mathbf{x}, \mathbf{y})} \right)^{1/2} f_\ell(\mathbf{x}, \mathbf{y}) d\mathbf{x}d\mathbf{y}. \end{aligned}$$

It is easy to see that $\Delta^{(H)}(\ell)$ is **symmetric**, and hence it can serve as a distance measure contrary to other divergences.

Various relations exist between the above divergence measures.



(IV) Distribution-based measures

Two useful types of functionals are

$$C_q(\ell) = \int_{S^2} \Delta_q^{\text{CR}}(\ell) dw_\ell(x, y), \quad \text{and} \quad C_q^{\text{max}}(\ell) = \sup_{S^2} [\Delta_q^{\text{CR}}(\ell) w_\ell(x, y)],$$

where $w_\ell(\cdot, \cdot)$ is a positive weight function and $\Delta_q^{\text{CR}}(\cdot)$ is the so-called **Cressie–Read (CR)** (*JRSS B*, 1984) divergence measure:

$$\Delta_q^{\text{CR}}(\ell) = \frac{2}{q+1} \left\{ F(x)F(y) \left(\frac{F(x)F(y)}{F_\ell(x, y)} \right)^q + \left(1 - F(x)F(y) \right) \left(\frac{1 - F(x)F(y)}{1 - F_\ell(x, y)} \right)^q - 1 \right\}.$$

Using $q = 1$ and $w_\ell(x, y) = F_\ell(x, y)(1 - F_\ell(x, y))dF_\ell(x, y)$ in $C_q(\cdot)$ gives the **Cramér–von Mises (CvM)** functional

$$\Delta^{\text{CvM}}(\ell) = \int_{S^2} \left\{ F_\ell(x, y) - F(x)F(y) \right\}^2 dF_\ell(x, y).$$

This measure satisfies the basic properties of nonnegativity and maximal information, but is **not** invariant under continuous monotonic increasing transformations.



Setting $q = 1$ and $w_\ell(x, y) = F_\ell(x, y)(1 - F_\ell(x, y))$ in $C_q^{\max}(\cdot)$, in $C_q(\cdot)$ gives

$$\left(\Delta^{\text{KS}}(\ell)\right)^2 = \left(\sup_{S^2} |F_\ell(x, y) - F(x)F(y)|\right)^2,$$

where $\Delta^{\text{KS}}(\cdot)$ is the **Kolmogorov–Smirnov (KS)** divergence measure. This measure satisfies the basic properties (i) – (iii).

Setting $q = 1$ and $w_\ell(x, y) = dF_\ell(x, y)$ in $C_q(\cdot)$ gives the **Anderson–Darling (AD)** functional

$$\Delta^{\text{AD}}(\ell) = \int_{S^2} \left(F(x)F(y) - F_\ell(x, y)\right)^2 F_\ell^{-1}(x, y) \left(1 - F_\ell(x, y)\right)^{-1} dF_\ell(x, y).$$

Multiple-lag testing procedure. One simple procedure is to form M linear combinations of single-lag two-dimensional test functionals $\Delta(\ell)$.

However, **pairwise** (serial) independence for all combinations of paired random variables does **not** imply joint (serial) independence in general. Hence, methods for the detection of serial dependence in $m > 2$ dimensions are needed; see p. 1–15 below.



(V) Copula-based measures

An m -dimensional copula-based version of $\Delta_q^T(\cdot)$ (T =Tsallis, c =copula) is defined as

$$\Delta_{m,q}^{T,c}(\ell) = \begin{cases} \frac{1}{1-q} \int_{[0,1]^m} \left\{ 1 - \left(\frac{1}{c(\mathbf{u})} \right)^{1-q} \right\} c(\mathbf{u}) d\mathbf{u} & (q \neq 1), \\ \int_{[0,1]^m} c(\mathbf{u}) \log[c(\mathbf{u})] d\mathbf{u} & (q = 1), \end{cases}$$

where $c(\mathbf{u})$ is the **copula density** of $\{\mathbf{Y}_t^{(\ell)}\}$.

It can be shown that $\Delta_{m,q}^{T,c}(\ell) \geq 0$ and $\Delta_{m,q}^{T,c}(\ell) = 0$ if and only if $\{\mathbf{Y}_t^{(\ell)}\}$ is serially independent. Equivalently, $\Delta_{m,q}^{T,c}(\mathbb{C}) = 0$ if and only if $\mathbb{C}(\mathbf{u}) = \Pi(\mathbf{u})$, where $\Pi(\mathbf{u}) \equiv \prod_{i=1}^m u_i$ being the **independence copula** ($m \geq 2$).

Other m -variate copula-based measures can be obtained in a similar manner as we previously applied to introduce the 4 major density-based measures.



1.3 Kernel-Based Tests

Nonparametric testing methods for which $f(\cdot)$ and $f_{\ell}(\cdot, \cdot)$ are assumed to be **unknown** under the \mathbb{H}_0 of serial independence.

Research questions:

- What is the most appropriate technique to estimate the densities?
- Which divergence measure should we adopt?
- Should we compute the functional estimates directly, or can we approximate the integration by a summation?
- Is there a need to include a trimming (weighting) function in the test functional, that is, screening off outliers by bounding the set of observations to some compact set?
- What is the most appropriate method of computing p -values: a BS approach or an MC permutation (random shuffle) approach of the data at hand?

Bagnato et al. (*Meth. & Computing in Appl. Probability*, 2014) present an exhaustive MC simulation comparison of 10 nonparametric serial independence tests, both single-lag and multiple-lag test procedures, using a wide class of L and nL models.

Conclusion: The integrated estimator of the KL functional ($J^{KL} \equiv \Delta_1^T$) combined with Gaussian kernel density estimation, provides the **best** performance in terms of empirical size and power. Also, a permutation-based approach is to be **preferred** over BS, and trimming functions are **not** needed.



Table: Single-lag ($m = 2$) tests. $S_T(\ell) \equiv \{t \in \mathbb{N} : \ell < t \leq T, \hat{f}_2(Y_t, Y_{t-\ell}) > 0, \hat{f}(Y_t) > 0, \hat{f}(Y_{t-\ell}) > 0\}$.

Reference	Divergence measure	Test statistic
Density functions		
Chan and Tran (1992)	Δ_1^K	$\hat{\Delta}_{T,1}^{CT}(\ell) = \sum_{t \in S_T(\ell)} \hat{f}(Y_t, Y_{t-\ell}) - \hat{f}(Y_t)\hat{f}(Y_{t-\ell}) $
Robinson (1991) ⁽⁴⁾	$J^{KL} \equiv \Delta_1^T$	$\hat{\Delta}_{T,\gamma}^R(\ell) = \frac{1}{T-\ell} \sum_{t \in S_T(\ell)} C_t(\gamma) \log \left(\frac{\hat{f}(Y_t, Y_{t-\ell})}{\hat{f}(Y_t)\hat{f}(Y_{t-\ell})} \right)$
Skaug and Tjøstheim (1993a)	Δ^H	$\hat{\Delta}_T^{ST1}(\ell) = \frac{1}{T-\ell} \sum_{t \in S_T(\ell)} 2 \left\{ 1 - \sqrt{\frac{\hat{f}(Y_t, Y_{t-\ell})}{\hat{f}(Y_t)\hat{f}(Y_{t-\ell})}} \right\} w_t(\ell)$
Skaug and Tjøstheim (1996)	Δ^*	$\hat{\Delta}_T^{ST2}(\ell) = \frac{1}{T-\ell} \sum_{t \in S_T(\ell)} \{\hat{f}(Y_t, Y_{t-\ell}) - \hat{f}(Y_t)\hat{f}(Y_{t-\ell})\} w_t(\ell)$
Granger and Lin (1994)	$1 - e^{-2J^{KL}}$	$\hat{\Delta}_T^{GL}(\ell) = 1 - \exp \left[\frac{-2}{T-\ell} \sum_{t \in S_T(\ell)} \log \left(\frac{\hat{f}(Y_t, Y_{t-\ell})}{\hat{f}(Y_t)\hat{f}(Y_{t-\ell})} \right) \right]$
Hong (2005)	$J^{KL} \equiv \Delta_1^T$	$\hat{\Delta}_T^{HW}(\ell) = \frac{1}{T-\ell} \sum_{t \in S_T(\ell)} \log \left(\frac{\hat{f}(Y_t, Y_{t-\ell})}{\hat{f}(Y_t)\hat{f}(Y_{t-\ell})} \right)$
Fernandes and Néri (2010)	$\Delta_{q \in \{\frac{1}{2}, 1, 2, 4\}}^T$	$\hat{\Delta}_{T,q}^{FN}(\ell) = \frac{1}{(1-q)(T-\ell)} \sum_{t \in S_T(\ell)} \left\{ 1 - \left(\frac{\hat{f}(Y_t)\hat{f}(Y_{t-\ell})}{\hat{f}(Y_t, Y_{t-\ell})} \right)^{1-q} \right\} w_t(\ell)$
Distribution functions		
Skaug and Tjøstheim (1993b)	Δ^{CvM}	$\hat{\Delta}_T^{ST3}(\ell) = \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} \{\hat{F}(Y_t, Y_{t+\ell}) - \hat{F}(\infty, Y_t)\hat{F}(Y_{t+\ell}, \infty)\}^2$

$C_t(\gamma) = 1 - \gamma$ if t is odd, $C_t(\ell) = 1 + \ell\gamma$ if $t = 1 \pmod{\ell + 1}$, and $C_t(\ell) = 1 - \gamma$ otherwise, with $\gamma \in (0, 1)$.



Multiple-lag test statistics:

Using the CvM functional, **Hong** (*JRSS B*, 1998) proposed

$$\hat{Q}^{H_1}(M) = \sum_{\ell=1}^M (T - \ell) \hat{\Delta}_T^{\text{ST}_3}(\ell).$$

A sensible generalization is to include a symmetric continuous window kernel $\lambda(\cdot)$ with $\lambda(0) = 1$. This ensures that the asymptotic bias of the test statistic vanishes.

Under \mathbb{H}_0 , $\{(T - \ell)\hat{\Delta}_T^{\text{ST}_3}(\ell); \ell = 1, \dots, T - 1\}$ can be viewed as an asymptotically i.i.d. sequence with mean $1/6^2$ and variance $2/90^2$. These results suggest the test statistic

$$\hat{Q}^{H_2}(M) = \frac{\sum_{\ell=1}^{T-1} \lambda^2(\ell/M) \{(T - \ell)\hat{\Delta}_T^{\text{ST}_3}(\ell) - 1/6^2\}}{\sqrt{2 \sum_{\ell=1}^{T-2} \lambda^4(\ell/M)/90^2}},$$

with the **Daniell lag window** $\lambda(u) = \sin(\pi u)/\pi u$, which is optimal over a class of window kernels.

It can be shown that $\hat{Q}^{H_2}(M) \xrightarrow{D} \mathcal{N}(0, 1)$ under the \mathbb{H}_0 of serial independence.



1.4 High-Dimensional Tests

BDS-test statistic:

Assume that the m -dimensional process $\{\mathbf{Y}_t\}$ admits a common continuous joint pdf $f_m(\mathbf{y})$ for $\mathbf{y} = (y_1, \dots, y_m)'$. Hence, $C_{m,\gamma}(h)$ (see p. 1–3) can be rewritten as $\mathbb{E}[I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq h)]$. An estimator $\widehat{C}_{m,\gamma}(h)$ of $C_{m,\gamma}(h)$ is given by

$$\widehat{C}_{m,\gamma}(h) = \binom{N}{2}^{-1} \sum_{1 \leq i < j \leq N} I(\|\mathbf{Y}_i - \mathbf{Y}_j\| < h),$$

where $N = T - m + 1$ is the # of vectors obtained from a time series $\{\mathbf{Y}_t\}_{t=1}^T$. Now, given the divergence measure $C_{m,\gamma}(h) - \{C_{1,\gamma}(h)\}^m$, a test statistic for serial independence is defined as

$$S_{m,\gamma}(h) = \sqrt{N} \frac{\widehat{C}_{m,\gamma}(h) - \{\widehat{C}_{1,\gamma}(h)\}^m}{\widehat{\sigma}_{m,\gamma}(h)},$$

where $\widehat{\sigma}_{m,\gamma}^2(h)$ is a consistent estimator of the variance of $\sqrt{N}(C_{m,\gamma}(h) - \{C_{1,\gamma}(h)\}^m)$. Under \mathbb{H}_0 and by exploiting the asymptotic theory for **U-statistics**, it can be shown that, as $T \rightarrow \infty$,

$$S_{m,\gamma}(h) \xrightarrow{D} \mathcal{N}(0, 1), \quad \forall h \in (0, \infty).$$



BDS test statistic (residuals, e) is defined as

$$S_{m,e}(h) = \sqrt{T} \frac{\widehat{C}_{m,e}(h) - \{\widehat{C}_{1,e}(h)\}^m}{\widehat{\sigma}_{m,e}(h)},$$

where in this case the sample correlation integral is given by

$$\widehat{C}_{m,e}(h) = \binom{T-m+1}{2}^{-1} \sum_{t=m+1}^T \sum_{s=m}^{t-1} \prod_{j=0}^{m-1} I(|e_{t-j} - e_{s-j}| < h).$$

Under \mathbb{H}_0 , $S_{m,e}(h)$ is asymptotically standard normal distributed.

Some problems:

- There is **arbitrariness** in the choice of h (**bandwidth**), which may affect both the power and size of the test. In fact, some choices of h may render the BDS test statistic **inconsistent** against certain alternatives. Thus, the probability of rejecting \mathbb{H}_0 does **not** always approach 1, as $T \rightarrow \infty$.
- Another problem is that the BDS test statistic has **high rates of Type I error**, especially for non-Gaussian data.



Rank-based analogue of the BDS test statistic:

Let $\tilde{e}_t = \text{rank}(e_t)/(T + 1)$ denote the normalized ranks of the time series $\{e_t\}_{t=1}^T$. Write $\tilde{\mathbf{W}}_t = (\tilde{W}_{1,t}, \dots, \tilde{W}_{m,t})' = (\tilde{e}_t, \dots, \tilde{e}_{t-m+1})'$. Then a **rank-based version** of the BDS statistic may be defined as

$$S_{m, \tilde{W}}(h) = \sqrt{T} \frac{\hat{C}_{m, \tilde{W}}(h) - \{\hat{C}_{1, \tilde{W}}(h)\}^m}{\hat{\sigma}_{m, \tilde{W}}(h)}.$$

Under the \mathbb{H}_0 of no serial dependence, it follows that $S_{m, \tilde{W}}(h) \xrightarrow{D} \mathcal{N}(0, 1)$, $\forall h \in (0, \infty)$, as $T \rightarrow \infty$.

Finite-sample performances of $S_{m, \tilde{W}}(h)$ depends on the **choice of h** . A common way to get around this problem is to integrate out h with regard to some empirical process using various continuous functionals.

Adopting **direct integration (D)**, the **KS** and **CvM** functionals, and 2 empirical processes, **Genest et al.** (JASA, 2007) propose 6 **rank-based BDS test statistics**; see the next Table. Moreover, they show that under \mathbb{H}_0 , all 6 test statistics converge in distribution to centered Gaussian variables.



Table: Rank-based BDS test statistics of serial independence using 3 functionals (direct integration (D), Kolmogorov–Smirnov (KS), and Cramér–von Mises (CvM)), and 2 empirical processes.

Functional	Empirical processes ^{(1) (2)}	
	$\tilde{\mathbb{D}}_T(\mathbf{u}) = \sqrt{T} \{ \tilde{B}_T(\mathbf{u}) - \prod_{k=1}^m \tilde{G}_T(u_k) \}$	$\tilde{\mathbb{B}}_T^*(\mathbf{u}) = 2\sqrt{T} \{ \tilde{B}_T^*(\mathbf{u}) - \tilde{B}_T(\mathbf{u}) \}$
D	$\tilde{T}_{m, \tilde{W}} = \int_0^1 \tilde{\mathbb{D}}_T(h, \dots, h) d\tilde{G}(h)$	$\tilde{T}_{m, \tilde{W}}^* = \int_0^1 \tilde{\mathbb{B}}_T^*(h, \dots, h) d\tilde{G}(h)$
KS	$\tilde{M}_{m, \tilde{W}} = \max_{i \in \{1, \dots, T\}} \left \tilde{\mathbb{D}}_T\left(\frac{i}{T+1}, \dots, \frac{i}{T+1}\right) \right $	$\tilde{M}_{m, \tilde{W}}^* = \max_{i \in \{1, \dots, T\}} \left \tilde{\mathbb{B}}_T^*\left(\frac{i}{T+1}, \dots, \frac{i}{T+1}\right) \right $
CvM	$\tilde{T}_{m, \tilde{W}} = \int_{[0,1]^m} \tilde{\mathbb{D}}_T(\mathbf{u}) ^2 d\tilde{B}(\mathbf{u})$	$\tilde{T}_{m, \tilde{W}}^* = \int_{[0,1]^m} \tilde{\mathbb{B}}_T^*(\mathbf{u}) ^2 d\tilde{B}(\mathbf{u})$

(1) $\tilde{B}_T(\mathbf{u}) = \binom{T}{2}^{-1} \sum_{1 \leq i \leq j \leq T} \prod_{k=1}^m I(|\tilde{W}_{k,i} - \tilde{W}_{k,j}| \leq u_k)$ with $\mathbf{u} = (u_1, \dots, u_m)' \in [0, 1]^m$;

$\tilde{G}_T(h) = \tilde{B}_T(h, 1, \dots, 1)$ with $h \in (0, 1]$.

(2) $\tilde{B}_T^*(\mathbf{u}) = T^{-1} \sum_{i=1}^T \prod_{k=1}^m \{ \tilde{F}(\tilde{w}_{k,i} + u_k) - \tilde{F}(\tilde{w}_{k,i} - u_k) \}$, where $\tilde{F}(\cdot)$ is the distribution of a $U(0, 1)$ random variable; $\tilde{B}_T^*(\mathbf{u}) = \prod_{k=1}^m \tilde{G}(u_k)$ with $\tilde{G}(\cdot)$ a Beta(1,2) distribution.

Based on simulation results: $\tilde{T}_{m, \tilde{W}}$ is the **best** performing rank-based BDS test. But **computationally demanding**.



Example:

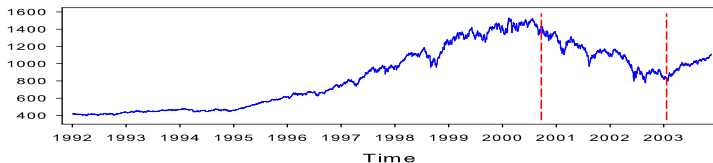


Figure: S&P 500 daily stock price index P_t (3,102 observations) with 2 subperiods, denoted by vertical red medium dashed lines, from 11/2000 – 2/2003 ($T = 608$) and 3/2003 – 12/2003 ($T = 218$).

Table: Bootstrap p -values of 7 test statistics for serial independence applied to daily S&P 500 stock returns; $B = 1,000$. Blue-typed numbers indicate rejection of \mathbb{H}_0 at the 5% nominal significance level.

Period	m	BDS	Rank-based BDS test statistics					
		$S_{m,T}$	$\tilde{I}_{m,\tilde{R}}^*$	$\tilde{M}_{m,\tilde{R}}^*$	$\tilde{T}_{m,\tilde{R}}^*$	$\tilde{I}_{m,\tilde{R}}$	$\tilde{M}_{m,\tilde{R}}$	$\tilde{T}_{m,\tilde{R}}$
11/2000 – 02/2003	2	0.21	0.07	0.14	0.08	0.57	0.53	0.91
	4	0.29	0.00	0.02	0.00	0.30	0.59	0.09
	6	0.36	0.00	0.02	0.00	0.30	0.58	0.01
03/2003 – 12/2003	8	0.43	0.00	0.02	0.00	0.29	0.76	0.00
	2	0.21	0.91	0.31	0.89	0.33	0.22	0.00
	4	0.30	0.91	0.49	0.80	0.10	0.85	0.00
	6	0.36	0.41	0.34	0.48	0.12	0.88	0.00
	8	0.46	0.13	0.15	0.15	0.31	0.75	0.00

$$\tilde{R}_t = \text{rank}(R_t)/(T + 1) \text{ with } R_t = \log(P_t/P_{t-1}).$$



1.6 Application

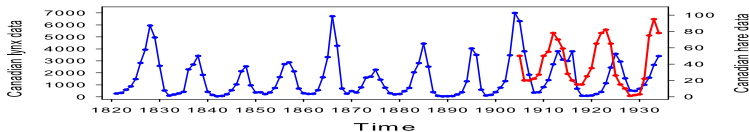


Figure: Yearly Canadian lynx data for (blue solid line), and yearly Canadian snowshoe hare data (in 1,000 (red solid line)). Snowshoe hare (prey) constitute a major part of the lynx's (predator) diet.

Table: BS p -values of 8 tests applied to the residuals of 5 models fitted to the log Canadian lynx data; $T = 114$, $B = 1,000$. Blue-typed numbers indicate rejection of \mathbb{H}_0 at a 5% significance level.

Model	m	BDS	Rank-based BDS test statistics						
		$S_{m,T}$	$\tilde{I}_{m,T}^*$	$\tilde{M}_{m,T}^*$	$\tilde{T}_{m,T}^*$	$\tilde{I}_{m,T}$	$\tilde{M}_{m,T}$	$\tilde{T}_{m,T}$	$\hat{\Delta}_{m,T}^{DP}$
AR(2)	2	0.25	0.07	0.55	0.04	0.67	0.54	0.01	0.23
	4	0.31	0.01	0.38	0.01	0.40	0.12	0.01	0.29
	6	0.43	0.01	0.62	0.01	0.56	0.04	0.02	0.50
SETAR(2; 7, 2)	2	0.26	0.33	0.64	0.34	0.59	0.81	0.21	0.25
	4	0.34	0.15	0.67	0.13	0.94	0.28	0.09	0.44
	6	0.44	0.25	0.58	0.21	0.63	0.15	0.08	0.60
SETAR(3; 1, 7, 2)	2	0.25	0.66	0.38	0.63	0.98	0.56	0.13	0.50
	4	0.32	0.40	0.27	0.32	0.92	0.27	0.15	0.52
	6	0.41	0.44	0.17	0.41	0.62	0.15	0.14	0.38
ExpAR(2)	2	0.25	0.12	0.32	0.01	0.12	0.15	0.01	0.04
	4	0.33	0.14	0.39	0.01	0.14	0.68	0.02	0.15
	6	0.43	0.38	0.55	0.00	0.38	0.32	0.04	0.33
LSTAR(11)	2	0.25	0.02	0.41	0.03	0.23	0.91	0.26	0.37
	4	0.32	0.01	0.20	0.01	0.19	0.99	0.24	0.09
	6	0.42	0.04	0.18	0.04	0.08	0.95	0.30	0.15

○○
 ○○○○○○○○○○
 ○○○
 ○○○○

○
 ●

○○○
 ○○
 ○○
 ○

○○○○
 ○○○○○○○○
 ○
 ○○
 ○○
 ○

1.7 Some Software References

- The entire R code for replicating the MC simulation study of **Bagnato et al. (2014)** is available at: <http://www.jandegooijer.nl>.
- A fast MATLAB code for computing BDS test statistic is based on: http://papers.ssrn.com/paper.taf?abstract_id=151669, is available at:
<http://econpapers.repec.org/software/bocbocode/t891501.htm>.
 Also BDS-C+, and BDS-MATLAB source codes are available at the address:
<http://people.brandeis.edu/~blebaron/>.
- The C+ code of the $\widehat{\Delta}_{m,T}^{\text{DP}}$ (DP=Diks-Panchenko test based on quadratic forms) test statistic can be downloaded from **Cees Diks'** web page located at <http://cendef.uva.nl/people>.



2. Time-Reversibility

2.1 Time-domain Test: Bicovariance-based

Since TR implies the equivalence of various distributions, it also implies the equality of subsets of moments from the joint distribution of $\{Y_t\}$, where they exist. Use **higher-order moments**. Assume $\{Y_t\}$ has mean 0. Then a sufficient (**not** necessary) condition for TR is

$$\mathbb{E}(Y_t^i Y_{t-\ell}^j) = \mathbb{E}(Y_t^j Y_{t-\ell}^i), \quad \forall (i, j) \in \mathbb{N} \text{ and } \forall \ell \in \mathbb{Z}.$$

For $i = 1, j = 2$, define the **symmetric-bicovariance function** as follows

$$\psi_Y(\ell) = \gamma_Y^{(2,1)}(\ell) - \gamma_Y^{(1,2)}(\ell),$$

where $\gamma_Y^{(i,j)}(\ell) = \mathbb{E}(Y_t^i Y_{t-\ell}^j)$. If $\{Y_t\}$ is TR, then $\psi_Y(\ell) = 0 \forall \ell \in \mathbb{Z}$. The TR test statistic (**Ramsey – Rothman, J. Money Credit & Banking, 1996**) is based on

$$\hat{\psi}_Y(\ell) = \hat{\gamma}_Y^{(2,1)}(\ell) - \hat{\gamma}_Y^{(1,2)}(\ell), \quad (\ell \in \mathbb{Z}),$$

where $\hat{\gamma}_Y^{(i,j)}(\ell) = (T - \ell)^{-1} \sum_{t=\ell+1}^T Y_t^i Y_{t-\ell}^j$ with $(i, j) = (1, 2)$. Its variance is given by

$$\text{Var}\{\hat{\psi}_Y(\ell)\} = 2(\mu_{4,Y} \mu_{2,Y} - \mu_{3,Y}^2) / (T - \ell) - 2\mu_{2,Y}^3 (T - 2\ell) / (T - \ell)^2.$$



Replacing $\mu_{3,Y}$ and $\mu_{4,Y}$ by their sample counterparts leads to $\widehat{\text{Var}}\{\widehat{\psi}_Y(\ell)\}$. Then the TR test statistic is defined by

$$\text{TR}(\ell) = \widehat{\psi}_Y(\ell) / \sqrt{\widehat{\text{Var}}\{\widehat{\psi}_Y(\ell)\}}.$$

Under $\mathbb{H}_0: \psi_Y(\ell) = 0$, it can be shown that $\text{TR}(\ell) \xrightarrow{D} \mathcal{N}(0, 1)$ as $T \rightarrow \infty$.

Algorithm:

Stage 1: Type I (DGP=nL, with symmetric $f_\varepsilon(\cdot)$), and Type II time-irreversibility (DGP = Linear and $\{\varepsilon_t\}$ is i.i.d. with a symmetric distribution): Rejection of \mathbb{H}_0 is consistent with both Type I and type II time-irreversibility.

Stage 2: Distinguish between Type I and Type II time-irreversibility.

Comment: ARMA prewhitening of $\{Y_t\}$ in Stage 1 may destroy TR since it induces a phase shift in the series (Hinich et al. *J. Macroeconomics*, 2006).

TR test based on the characteristic function:

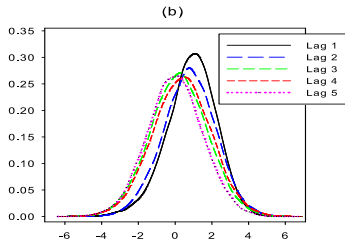
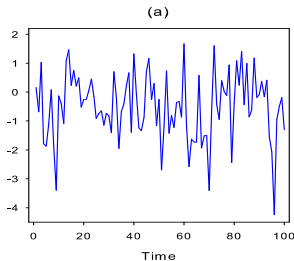
- Chen et al. (*J. Econometrics*, 2000): Unconditional symmetry.
- Chen and Kuan (*J. Appl. Econometrics*, 2002): Replace $\{Y_t\}$ by some residual series.



Example: A simple way to explore $\{Y_t\}_{t=1}^T$ for TR is to detect asymmetries in plots of the sample distributions of $W_t(\ell) = Y_t - Y_{t-\ell}$ ($\ell = 1, 2, \dots$). Consider the stationary SETAR(2; 1, 1) process

$$Y_t = \begin{cases} 0.5Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} \leq 0, \\ -0.4Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} > 0, \end{cases}$$

where $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. Fig. (a) shows a plot of a typical subset of length $T = 100$ of a simulated time series of 10,000 observations. Fig. (b) displays the kernel smoothed densities of $\{W_t(\ell)\}_{t=1}^{10,000}$ ($\ell = 1, \dots, 5$), using a normal kernel. Observe that the distributions are **not** symmetric about the origin, indicating the SETAR process is **time-irreversible**.





2.2 Frequency-domain: Bispectrum-based

Hinich and Rothman (*Macroeconomic Dynamics*, 1998): Under the \mathbb{H}_0 of TR, $\Im\{f_Y(\omega_1, \omega_2)\} = 0 \forall (\omega_1, \omega_2) \in \mathcal{D}$ where \mathcal{D} is the principal domain (see Part II, p. 1–3). This result can be used to define a frequency-domain TR test statistic based on the **imaginary part** of the **normalized estimated bispectrum** $\widehat{B}_Y(\omega_1, \omega_2)$, say $\Im\{\widehat{B}_Y(\omega_1, \omega_2)\}$.

Trispectrum-based test:

Triple FT of the 4th-order cumulant function of a stationary time series $\{Y_t\}$, i.e.,

$$f_Y(\omega_1, \omega_2, \omega_3) = \sum_{l_1, l_2, l_3 = -\infty}^{\infty} \gamma_Y(l_1, l_2, l_3) \exp\{-2\pi i(\omega_1 l_1 + \omega_2 l_2 + \omega_3 l_3)\},$$

where $(\omega_1, \omega_2, \omega_3) \in [0, 1]^3$ are normalized frequencies, and the 3rd-order cumulant function is defined as $\gamma_Y(l_1, l_2, l_3) = \mathbb{E}(Y_t Y_{t+l_1} Y_{t+l_2} Y_{t+l_3})$. Owing to symmetry relations, the trispectrum need to be calculated only in a **subset** of the complete $(\omega_1, \omega_2, \omega_3)$ -space (similar to sector ① in the figure on p. 1–4 of Part II).



Steps:

- (i) Divide $\{Y_t\}_{t=1}^T$ into K non-overlapping stretches, or frames, of length N so that $K = \lfloor T/N \rfloor$. Define the discrete Fourier frequencies $\omega_j = j/N$ ($j = 1, \dots, N$).
- (ii) Calculate the discrete FT $Y_k(\omega_j) = \sum_{t=1}^N Y_{t+(k-1)N} \exp\{-2\pi i \omega_j (t + (k-1)N)\}$, and the periodogram of the k th frame ($k = 1, \dots, K$).
- (iii) Compute

$$\widehat{f}_Y(\omega_{j_1}, \omega_{j_2}, \omega_{j_3}) = \frac{1}{T} \sum_{k=1}^K Y_k(\omega_{j_1}) Y_k(\omega_{j_2}) Y_k(\omega_{j_3}) Y_k(-\omega_{j_1} - \omega_{j_2} - \omega_{j_3}).$$

Then the **normalized estimated trispectrum** is

$$\widehat{T}_Y(\omega_{j_1}, \omega_{j_2}, \omega_{j_3}) = \frac{\widehat{f}_Y(\omega_{j_1}, \omega_{j_2}, \omega_{j_3})}{\sqrt{\widehat{f}_Y(\omega_{j_1}) \widehat{f}_Y(\omega_{j_2}) \widehat{f}_Y(\omega_{j_3}) \widehat{f}_Y(\omega_{j_1} + \omega_{j_2} + \omega_{j_3})}}.$$

- (iv) Compute the TR test statistic

$$S_{TR}^* = 2T^{2c-1} \sum_{\omega_{j_1}, \omega_{j_2}, \omega_{j_3} \in \Omega} |\Im\{\widehat{T}_Y(\omega_{j_1}, \omega_{j_2}, \omega_{j_3})\}|^2, \quad \left(\frac{1}{2} < c < 1\right).$$

Under \mathbb{H}_0 : $\Im\{T_Y(\omega_{j_1}, \omega_{j_2}, \omega_{j_3})\} = 0$, and as $T \rightarrow \infty$,

$$S_{TR}^* \xrightarrow{D} \chi_{M^*}^2$$

with M^* the # of frequency triples in Ω . This number is automatically computed in the available software code. **Disadvantage:** S_{TR}^* is applicable if the one-dimensional marginal distribution of $\{Y_t\}$ has a finite 8th moment. **Computationally demanding.**



2.3 Other Nonparametric Tests:

- **Copula-based test for Markov chains:** Beare and Seo (*Econometric Theory*, 2014).
- **Kernel-based test:** Diks et al. (*Physics Letters A*, 1995).
- **Sign test:** Psaradakis (*JTSA*, 2008). The projection of the m -dimensional delay vectors on $(Y_t, Y_{t-\ell})$ ($\ell = 1, \dots, m-1$) can be evaluated by exploiting the fact that for a TR process $\{X_t(\ell) = Y_t - Y_{t-\ell}\}$, we have

$$\mathbb{P}(X_0(\ell) > 0) = \mathbb{P}(X_0(\ell) < 0) = \frac{1}{2}, \quad (\ell = 1, \dots, m-1).$$

The object of interest is thus the probability $\pi(\ell) \equiv \mathbb{P}(X_0(\ell) > 0)$, which may be thought of as a simple measure of deviation from 0 of the one-dimensional distribution of $\{X_t(\ell)\}$. A point estimator of $\pi(\ell)$ is

$$\hat{\pi}(\ell) = \frac{1}{T-\ell} \sum_{t=\ell+1}^T I(X_t(\ell) > 0), \quad (\ell = 1, \dots, m-1).$$

For each fixed $\ell \in \mathbb{N}$, as $T \rightarrow \infty$,

$$\sqrt{T-\ell}(\hat{\pi}(\ell) - \pi(\ell)) \xrightarrow{D} \mathcal{N}(0, \sigma_X^2(\ell)),$$

where

$$\sigma_X^2(\ell) = \pi(\ell)(1 - \pi(\ell)) + 2\pi(\ell) \sum_{t=1}^{\infty} \{\mathbb{P}(X_t(\ell) > 0) | X_0(\ell) > 0\} - \pi(\ell).$$



Example: Consider the stochastic process

$$Y_t = 1 - 1.4Y_{t-\ell}^2 + 0.3Y_{t-2\ell-1} + \varepsilon_t, \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} U(-0.01, 0.01).$$

A “**clothed**”, or randomized, version of the time-delayed deterministic (*its skeleton*) **Hénon map**. Time series generated by the Hénon map are **known** to be **time-irreversible**. Compute

$$\mathcal{R}(m) = \frac{1}{m-1} \sum_{\ell=1}^{m-1} \left| \frac{1}{2} - \hat{\pi}(\ell) \right| \times 100 \quad (m = 2, \dots, 15).$$

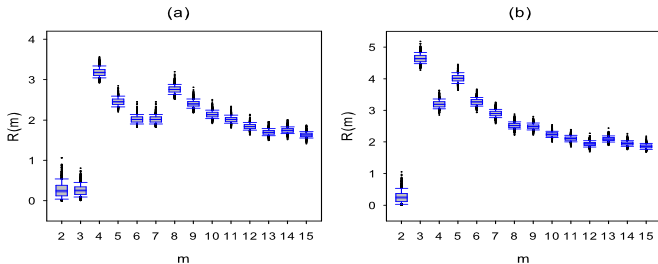


Figure: Boxplots of $\mathcal{R}(m)$ based on 1,000 MC replications of series of length $T = 5,000$ generated from the time-delayed Hénon map with (a) $\ell = 1$ and (b) $\ell = 2$. **Note:** time-irreversibility.

○○
 ○○○○○○○○○○
 ○○○
 ○○○○

○
 ○

○○○
 ○○
 ○○
 ●

○○○○
 ○○○○○○○○
 ○
 ○○○
 ○○
 ○

2.4 Some Software References

- F77 code to calculate the first and second stage of the **Ramsey–Rothman** TR test statistic can be found at the website <http://www.jandegooijer.nl>.

Also, a GAUSS program for running the **Chen–Chou–Kuan** TR test statistic is available. In addition, the above website has MATLAB code for computing the copula-based TR test statistic for Markov chains.

- C source code and a Linux/Windows executable of the kernel-based TR test statistic can be downloaded from **Cees Diks**' web page, located at <http://cendef.uva.nl/profile/d/i/c.g.h.diks/c.g.h.diks.html>.



3. Forecasting

3.1 Exact LS Methods: Introduction

For a strictly stationary stochastic process $\{Y_t\}$ the least squares (LS), or MMSE forecast of Y_{t+H} ($H = 1, 2, \dots$), is given by $\mathbb{E}(Y_{t+H} | Y_s, -\infty < s \leq t)$ when this exists.

For a p th-order Markov process, the MMSE forecast of Y_{t+H} equals $Y_{t+H|t}^{\text{LS}} = \mathbb{E}(Y_{t+H} | \mathbf{X}_t)$, where $\mathbf{X}_t = (Y_t, Y_{t-1}, \dots, Y_{t-p+1})'$. Calculation of $Y_{t+H|t}^{\text{LS}}$ requires knowledge of the conditional pdf of $\{Y_t\}$. For an NLAR(p) model

$$Y_t = \mu(\mathbf{X}_{t-1}; \boldsymbol{\theta}) + \varepsilon_t,$$

where $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_\varepsilon^2)$ such that ε_t is independent of \mathbf{X}_{t-1} , $\boldsymbol{\theta}$ is a finite-dimensional vector of unknown parameters, and $\mu: \mathbb{R}^p \rightarrow \mathbb{R}$. The one-step ahead LS forecast at time t equals

$$Y_{t+1|t}^{\text{LS}} = \mathbb{E}(Y_{t+1} | \mathbf{X}_t) = \mathbb{E}\{\mu(\mathbf{X}_t; \boldsymbol{\theta}) + \varepsilon_{t+1} | \mathbf{X}_t\} = \mu(\mathbf{X}_t; \boldsymbol{\theta}).$$

So, for $H = 1$, the **conditional mean** is **independent** of the **distribution** of ε_{t+1} which is an important property for both linear and NLAR models.



When $H \geq 2$, however, this is true **only for linear models**. For example, the **two-step ahead LS** forecast for an NLAR(p) model is given by

$$\begin{aligned} Y_{t+2|t}^{\text{LS}} &= \mathbb{E}(Y_{t+2} | \mathbf{X}_t) = \mathbb{E}\{\mu(\mathbf{X}_{t+1}; \boldsymbol{\theta}) + \varepsilon_{t+2} | \mathbf{X}_t\} \\ &= \mathbb{E}\{\mu(\mu(\mathbf{X}_t; \boldsymbol{\theta}) + \varepsilon_{t+1}) | \mathbf{X}_t\} = \int_{-\infty}^{\infty} \mu(\mu(\mathbf{X}_t; \boldsymbol{\theta}) + \varepsilon) dF(\varepsilon), \end{aligned}$$

where $F(\cdot)$ is the distribution function of $\{\varepsilon_t\}$.

Thus, the 2nd term on the right-hand side depends on $F(\cdot)$, and **cannot** further be reduced as with the expression of the one-step ahead forecast.

The reason is that, in general, the conditional expectation of a nonlinear function is **not** equal to the function evaluated at the expected value of its argument.



NLAR(p) model

Let $g(\cdot)$ be the pdf of $\{\varepsilon_t\}$. Using the **Chapman–Kolmogorov relation**, the conditional pdf of Y_{t+H} given $\mathbf{X}_t = \mathbf{x}_t$ can be written as

$$f(y_{t+H}|\mathbf{x}_t) = \int_{-\infty}^{\infty} f(y_{t+H}|\mathbf{x}_{t+1})f(y_{t+1}|\mathbf{x}_t)d\mathbf{x}_{t+1},$$

where

$$f(y_{t+1}|\mathbf{x}_t) = g(y_{t+1} - \mu(\mathbf{x}_t; \boldsymbol{\theta})).$$

Alternatively, this equation can be obtained by considering the joint pdf of $Y_{t+H}, Y_{t+H-1}, \dots, Y_{t+1}$ conditional on $\mathbf{X}_t = \mathbf{x}$ and integrating out the unwanted variables. Introducing $f_H(\cdot) \equiv f_{Y_{t+H}|Y_t}(\cdot|\mathbf{x})$, we have

$$f_H(\mathbf{x}) = \int_{-\infty}^{\infty} f_{H-1}(x)g(z - \mu(\mathbf{x}; \boldsymbol{\theta}))dz.$$

Given $f_H(\cdot)$ at step $H = 1$, the conditional mean for $H \geq 2$ is given by

$$Y_{t+H|t} = \int_{-\infty}^{\infty} f_{H-1}(Y_{t+1})g(Y_{t+1} - \mu(\mathbf{X}_t; \boldsymbol{\theta}))dY_{t+1}.$$

Computing $Y_{t+H|t}$ requires $p + 1$ numerical integrations; **Pemberton** (*JTSA*, 1987), **Al-Qassem and Lane** (*JTSA*, 1989), and **Cai** (*JoF*, 2005).



Example:

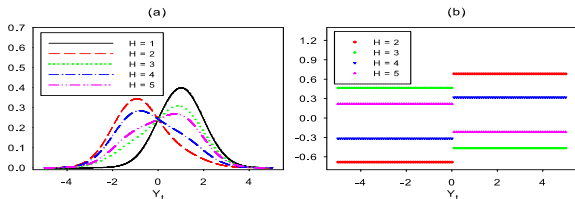


Figure: (a) Forecast density $f(y_{t+H}|\mathbf{x}_t)$ ($H = 1, \dots, 5$) for a SETAR(2; 0, 0) model; (b) Conditional mean $\mathbb{E}(Y_{t+H}|\mathbf{X}_t)$ ($H = 2, \dots, 5; \alpha = 1$).

Consider the SETAR(2; 0, 0) model

$$Y_t = \begin{cases} \alpha + \varepsilon_t & \text{if } Y_{t-1} \leq 0, \\ -\alpha + \varepsilon_t & \text{if } Y_{t-1} > 0, \end{cases}$$

where $\{\varepsilon_i\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. Let $\varphi(\cdot)$ be the pdf and $\Phi(\cdot)$ the CDF of $\mathcal{N}(0, 1)$. Then

$$f(y_t) = \{\varphi(y_t + \alpha) + \varphi(y_t - \alpha)\}/2.$$

The exact LS conditional pdf of Y_{t+H} ($H = 1, 2, \dots$) given $\mathbf{X}_t = \mathbf{x}$ has the form

$$f(y_{t+H}|\mathbf{x}) = w_1^{(H)}(\beta)\varphi(Y_{t+H} - I(Y_t \leq 0)\alpha) + w_2^{(H)}(\beta)\varphi(Y_{t+H} + I(Y_t > 0)\alpha),$$

where $w_1^{(H)}(\beta) = (1 - \beta^{H-1})/2$, $w_2^{(H)}(\beta) = 1 - w_1^{(H)}(\beta)$, and $\beta = 1 - 2\Phi(\alpha)$.



Two-regime SETARMA:

$$Y_t = Y_t^{(1)} I_{t-d} + Y_t^{(2)} (1 - I_{t-d}),$$

where in the i th regime by $Y_t^{(i)} \sim \text{ARMA}(p_i, q_i)$ ($i = 1, 2$).

Assume the joint process $\{(Y_t^{(1)}, Y_t^{(2)}, I_{t-d})\}$ is strictly stationary, invertible, and ergodic. The **exact** H -step ahead ($H \geq 2$) LS forecast is given by

$$Y_{t+H|t}^{\text{LS}} = Y_{t+H|t}^{(1)} \mathbb{E}(I_{t+H-d} | \mathcal{F}_t) + Y_{t+H|t}^{(2)} (1 - \mathbb{E}(I_{t+H-d} | \mathcal{F}_t)),$$

where $Y_{t+H|t}^{(i)}$ is the ARMA forecast in regime i , and $\mathcal{F}_t = \{Y_t, Y_{t-1}, \dots\}$ denotes the information set up to time t .

Depending on the case $H \leq d$ (**easiest**) or $H > d$, expressions for the variance of the LS forecast errors can be derived; **Amendola et al.** (*Physics & Chemistry of the Earth*, 2006).



3.2 Approximate Methods: Monte Carlo (MC)

Given a $H = 1$ forecast at time t , the **MC** method is a simple recursive simulation method to approximate the expectation of Y_{t+H} ($H \geq 2$) conditional upon \mathcal{F}_t . The $H = 2$ forecast can be constructed as

$$Y_{t+2|t}^{\text{MC}} = \frac{1}{N} \sum_{i=1}^N Y_{t+2|t}^{\text{MC}_i} \quad \text{where} \quad Y_{t+2|t}^{\text{MC}_i} = \mu((Y_{t+1|t}; \boldsymbol{\theta}) + \varepsilon_{2,i}),$$

with $\{\varepsilon_{2,i}\}_{i=1}^N$ a set of pseudo-random numbers drawn from the presumed distribution of $\{\varepsilon_{t+1}\}$, and with N large. The **H -step ahead forecast** is given by

$$Y_{t+H|t}^{\text{MC}} = \frac{1}{N} \sum_{i=1}^N Y_{t+H|t}^{\text{MC}_i} \quad (H > 2),$$

where

$$Y_{t+H|t}^{\text{MC}_i} = \mu((Y_{t+H-1|t}^{\text{MC}_i}; \boldsymbol{\theta}) + \varepsilon_{H,i}) = \mu(\mu(\cdots (\mu(Y_{t+1|t}; \boldsymbol{\theta}) + \varepsilon_{2,i}) + \cdots) + \varepsilon_{H,i}),$$

with $\varepsilon_{j,i}$ ($j = 2, \dots, H; i = 1, \dots, N$) independent pseudo-random numbers drawn from some pre-specified distribution of $\{\varepsilon_{t+H}\}$.



Bootstrap (BS):

Forecasts obtained from the BS method are similar to the MC simulation method except that the $e_{j,i}^*$ are drawn randomly (**with replacement**) from the within-sample residuals e_i ($i = 2, 3, \dots, T$), assuming a set of T historical data is available to obtain some consistent estimate of θ .

In this case the **H -step ahead** ($H \geq 2$) forecast is given by

$$Y_{t+H|t}^{\text{BS}} = \frac{1}{T-1} \sum_{i=2}^T Y_{t+H|t}^{\text{BS}_i}$$

where

$$\begin{aligned} Y_{t+H|t}^{\text{BS}_i} &= \mu((Y_{t+H-1|t}^{\text{BS}_i}; \theta) + e_{H,i}^*) \\ &= \mu(\mu(\dots(\mu(Y_{t+1|t}; \theta) + e_{2,i}^*) + \dots) + e_{H,i}^*). \end{aligned}$$

Advantage of the BS method over the MC method is that **no** assumptions are made about the distribution of the innovation process.



Deterministic, Naive, or Skeleton (SK):

“Switch off the white noise” in the MC method. Thus, the two-step ahead forecast is given by

$$Y_{t+2|t}^{\text{SK}} = \mu(Y_{t+1|t}; \theta).$$

This approach leads to **biased predictions** since $Y_{t+2|t}^{\text{SK}} \neq \mathbb{E}(Y_{t+2|t})$. By induction, the **H -step ahead** forecast can be computed as

$$Y_{t+H|t}^{\text{SK}} = \mu(\mu(\cdots \mu(Y_{t+1|t}^{\text{SK}}; \theta))).$$

Advantage: The SK method is computationally **inexpensive**. However, unlike most approximate forecasting methods, the SK forecasts **do not** necessarily converge to the mean of the process. (**Disadvantage**)

Moreover, as σ_ε^2 increases there is the possibility that the deterministic component of the model ceases to dictate the behavior of the process and the noise part starts to be dominant, causing for instance switches between different limit/oscillation points. (**Disadvantage**)



Empirical least squares (ELS): Guo et al. (*Stat. Sinica*, 1999)

Assume an NLAR(1) model is known and correctly specified for the DGP, but the innovation distribution is **unspecified**.

Use $\hat{F}_T(x) = (T-1)^{-1} \sum_{i=2}^T I(e_i < x)$ as an estimate of the innovation distribution. Then the two-step ahead ($H=2$) ELS forecast can be defined as

$$Y_{t+2|t}^{\text{ELS}} = \frac{1}{T-1} \sum_{i=2}^T \mu(\mu(Y_{t+1|t}; \boldsymbol{\theta}) + e_i).$$

The ELS method can be readily extended to the case $H > 2$. For instance, the **exact three-step ahead LS forecast** is given by

$$Y_{t+3|t}^{\text{LS}} = \int_{-\infty}^{\infty} \mu(\mu(\mu(Y_{t+1|t}; \boldsymbol{\theta}) + \varepsilon) + \varepsilon') dF(\varepsilon) dF(\varepsilon').$$

Thus, as a three-stage ELS forecast, we may take

$$Y_{t+3|t}^{\text{ELS}} = \frac{1}{(T-1)(T-2)} \sum_{2 \leq i \neq j \leq T} \mu(\mu(\mu(Y_{t+1|t}; \boldsymbol{\theta}) + e_i) + e_j).$$

The ELS method can be generalized to NLARs with ARCH errors.



Normal forecasting error (NFE):

An alternative to the H -step ahead ($H \geq 2$) exact LS predictor is to assume as an approximation that all $(H - 1)$ forecast errors $e_{t+H-1|t}$ ($H \geq 2$) are **normally distributed** with mean 0 and variance $\sigma_{e,H-1}^2 \equiv \text{Var}(e_{t+H-1|t})$.

- **Al-Quassem and Lane** (*JTSA*, 1989) for ExpAR(1) models.
- **De Gooijer and De Bruin** (*Stats. Prob. Letters*, 1998) for SETAR(2;1,1) models.

The normality assumption avoids the use of numerical methods. However, as $\mu(\cdot; \theta)$ is an nL function the multi-step ahead forecast errors $e_{t+H-1|t}$ will **not** equal the linear innovations, **nor** will they follow an i.i.d. Gaussian process. (**Disadvantage**)



Linearization (LN):

Taylor's expansion up to order two of $\mu(\cdot; \theta)$ about the point $Y_{t+H-1|t}$ gives

$$\mu(Y_{t+H-1}; \theta) \simeq \mu(Y_{t+H-1|t}; \theta) + e_{t+H-1|t} \mu^{(1)}(Y_{t+H-1|t}; \theta) + \frac{1}{2} e_{t+H-1|t}^2 \mu^{(2)}(Y_{t+H-1|t}; \theta),$$

where $\mu^{(i)}(\cdot; \theta)$, ($i = 1, 2$) denotes the i th derivative of $\mu(\cdot; \theta)$ with respect to $Y_{t+H-1|t}$, and $e_{t+H-1|t}$ is the $(H - 1)$ -step ahead forecast error ($H \geq 2$).

Assume $\{e_{t+H-1|t}\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{e,H-1}^2)$. Then, substituting the above expression in the NLAR(1) model and taking the conditional expectation of the resulting specification, gives the **H -step ahead LN forecast**, i.e.

$$Y_{t+H|t}^{\text{LN}} \simeq \mu(Y_{t+H-1|t}; \theta) + \frac{1}{2} \sigma_{e,H-1}^2 \mu^{(2)}(Y_{t+H-1|t}; \theta).$$

Disadvantage: Forecasts obtained from this method can be quite different from the exact prediction method or from the NFE method for moderate or large σ_{ε}^2 (mainly $\geq 10^{-2}$).



Example:

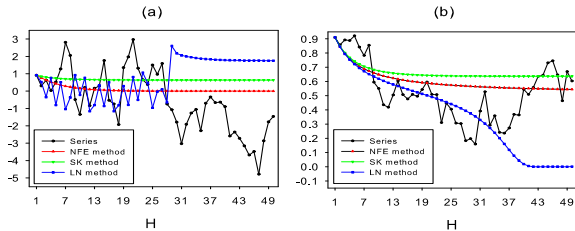


Figure: Forecasts from an ExpAR(1) model with the NFE, SK, and LN methods; (a) $\sigma_\varepsilon^2 = 1$, and (b) $\sigma_\varepsilon^2 = 0.01$; $\{\varepsilon_t\} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $Y_0 = 1$.

ExpAR(1):

$$\mu(X; \boldsymbol{\theta}) = \{\phi + \xi \exp(-\gamma X^2)\}X,$$

where $\boldsymbol{\theta} = (\phi, \xi, \gamma)' = (0.8, 0.3, 1)'$. Using expressions for the partial derivatives of $\mu(\cdot; \boldsymbol{\theta})$ with respect to X , we get

$$Y_{t+H|t}^{\text{LN}} = \phi + \xi f_{H-1} \exp(-\gamma(Y_{t+H-1|t})^2) Y_{t+H-1|t},$$

where $f_{H-1} = 1 + \gamma \sigma_{\varepsilon, H-1}^2 (2\gamma(Y_{t+H-1|t})^2 - 3)$.



The specified ExpAR(1) process has two limit points at ± 0.6368 .

Notes:

- Fig. (a): it is clear that NFE forecasts go to a limit point 0, SK forecasts go to the upper limit point 0.6368, while the series of LN forecast are **unstable** up to about $H = 30$, then stabilize to a point far off the upper limit point.
- Fig. (b) shows that for $H \leq 5$ there is **hardly any noticeable difference** between the three forecasting methods, provided σ_ε^2 is small. For $H \geq 30$ the LN method may go to the **“wrong”** limit point.



Dynamic estimation (DE):

The DE method is based on the **in-sample relationship** between Y_t and Y_{t+H} , **ignoring** contributions of intermediate values, to produce H -step ahead forecasts. That is, we replace the NLAR(1) model by the following specification

$$Y_{t+H} = \mu(Y_t; \theta_H^*) + \varepsilon_{t+H}^*,$$

where θ_H^* is a vector of parameters depending upon H . These parameters can be estimated by minimizing the sum of squares of ε_{T+H}^* over θ_H^* for the sample period $t = 1, \dots, T$. So that, given $\hat{\theta}_H^*$, the corresponding H -step ahead DE forecast can be written as

$$Y_{t+H|t}^{\text{DE}} = \mu(Y_t; \hat{\theta}_H^*).$$

Comments:

- In a linear setting, there are **no** gains in terms of increased forecast accuracy using DE over the traditional minimization of in-sample sum of squares of one-step ahead errors when the model is correctly specified. When a nL model is correctly specified, DE may result in **better** out-of-sample forecasts due to its **simplicity**.
- An obvious **drawback** is that the nonlinear model needs to be estimated for each H .
- **Clements and Smith** (*IJF*, 1997) conclude that the DE method is **worse** than the BS, MC and NFE methods for SETAR(2; 1, 1) models with Gaussian disturbances and 0 intercepts.



3.3 Forecast Intervals/Regions

Forecast intervals/regions are **more informative** than point predictions as they indicate the likely range of forecast outcomes. As such, they are a measure of the inherent model accuracy.

The conditional distribution of Y_{T+H} given $\mathcal{F}_t = \{Y_t, Y_{t-1}, \dots\}$ can be used to construct a forecast interval/region for Y_{T+H} . Given $\mathbf{X}_t = \mathbf{x}$, $Q_\alpha \equiv Q_\alpha(\mathbf{x}) \subset \mathbb{R}$ is such an interval with coverage probability $1 - \alpha$ ($\alpha \in [0, 1]$). That is $\mathbb{P}\{Y_{T+H} \in Q_\alpha(\mathbf{x}) | \mathbf{X}_{T-H-p+1} = \mathbf{x}\} = 1 - \alpha$, assuming the DGP is strictly stationary and p -Markovian. The set Q_α is called **forecast region** (FR). When Q_α is a connected set, it is called **forecast interval** (FI).

Obviously an FR/FI can be constructed in an infinite number of ways. For instance, a natural FI for the conditional median of Y_{T+H} is the **conditional percentile interval** (CPI) given by

$$\text{CPI}_{1-\alpha} = [\hat{\xi}_{\alpha/2}(\mathbf{x}), \hat{\xi}_{1-\alpha/2}(\mathbf{x})],$$

where $\hat{\xi}_\alpha(\mathbf{x})$ is the α th conditional percentile of $\xi_\alpha(\mathbf{x}) = \inf\{z: F(z|\mathbf{x}) \geq \alpha\}$ and α the quantile level.



3.4 Forecast Evaluation

Forecast schemes:

- Fixed:** Let $\{Y_t\}_{t=1}^{T+H}$ be the sample of observations, where $H \equiv H_{\max} \geq 1$ denotes the longest forecast horizon of interest. Assume the available data set is divided into **in-sample** and **out-sample** portions, with R the total # of in-sample observations and P the # of H -step ahead forecasts. This setup implies that P out-of-sample forecasts depend on the **same** parameter vector estimated on the 1st R observations. So, the forecast scheme is based on a **single, fixed**, estimation sample.
- Recursive:** The 1st forecast is based on a model with parameter vector estimated using $\{Y_t\}_{t=1}^R$, the 2nd on a parameter vector estimated using $\{Y_t\}_{t=1}^{R+1}, \dots$, the last on a parameter vector estimated using $\{Y_t\}_{t=1}^{R+P-1}$, where $T \equiv R + P - 1$.
- Rolling:** The sequence of parameter estimates is always generated from a **fixed, but rolling**, sample of size R : The 1st forecast is based on parameter estimates obtained from the set of observations $\{Y_t\}_{t=1}^R$, the next on parameter estimates obtained from $\{Y_t\}_{t=2}^{R+1}$, and so on.



Point forecast evaluation:

- Diebold–Mariano (DM):** Based on the \mathbb{H}_0 that 2 forecasts are the same in terms of forecasting accuracy, for some arbitrary loss function $L(e_{i,t+H|t})$ where $e_{i,t+H|t} = Y_{t+H} - Y_{i,t+H|t}$ is the H -step ahead forecast error with $Y_{i,t+H|t}$ the forecasts from model i ($i = 1, 2$). The H -step ahead **loss differential** is defined as

$$d_t = L(e_{i,t+H|t}) - L(e_{j,t+H|t}), \quad (i, j = 1, 2; i \neq j).$$

So

$$\mathbb{H}_0: \quad \mu_d \equiv \mathbb{E}(d_t) = 0.$$

- Modified DM (MDM):** For a small # of forecasts, use MDM. The modification follows from replacing an approximation of $\text{Var}(\bar{d})$, with $\bar{d} = P^{-1} \sum_{t=R+H}^{R+P+H-1} d_t$, by its exact variance.

Advantage: DM is “**model-free**”, i.e., the forecast models are assumed to be correctly specified, but unknown, and $L(\cdot)$ does not rests on additional, conditioning, information. The competing forecasts $Y_{1,t+H|t}$ and $Y_{2,t+H|t}$ are obtained from **non-nested** models.

Disadvantage: With **nested** models the limiting distribution of DM is **non-standard**.



Interval evaluation: Christoffersen (*Int. Econ. Review*, 1998). LR-based test statistics.

Density evaluation: Consider a set of P one-step ahead forecast densities for the future value of a process $\{Y_t\}$, denoted by $\{\hat{f}_t(Y_t|\mathcal{F}_{t-1})\}_{t=1}^P$, made at time t with $f_1(Y_1|\mathcal{F}_0) \equiv f(y_1)$. The PIT, denoted by U_t , is defined as

$$U_t \equiv \int_{-\infty}^{Y_t} \hat{f}_t(u|\mathcal{F}_{t-1}) du, \quad (t = 1, \dots, P).$$

Under the \mathbb{H}_0 that the model forecasting density corresponds to the true conditional density, given by the DGP which is denoted by $f_t(\cdot|\mathcal{F}_{t-1})$, i.e. $\hat{f}_t(\cdot|\mathcal{F}_{t-1}) = f_t(\cdot|\mathcal{F}_{t-1})$, $\{U_t\} \stackrel{\text{i.i.d.}}{\sim} U(0, 1)$.

Testing the **uniformity part:**

- Use a nonparametric goodness-of-fit test like KS, AD or CvM.
- Plot the CDF of the U_t and visually compare it with a line at an angle of 45° representing the cumulative $U(0, 1)$ distribution.

Testing the **independence part:**

- Use an LM-type test statistic for serial correlation in the sequences $\{(U_t - \bar{U})^u\}_{t=1}^P$ ($u = 1, 2$), where \bar{U} is the sample mean of U_t .



3.5 Forecast Combination

If the individual forecasts are unbiased then common practice is to obtain a weighted average of forecasts, with weights $w_i \geq 0$ and $\sum_{i=1}^n w_i = 1$. The w_i follow from minimizing some loss function, usually the MSFE.

Equal-weighting (ew) often outperforms estimated optimal forecast combinations, i.e.

$$Y_{t+H|t}^{\text{ew}} = \frac{1}{n} \sum_{i=1}^n Y_{i,t+H|t}.$$

Interval forecasts: FIs are frequently **too narrow**, i.e. too many observations are in the tails of the forecast distribution. One most likely reason is that forecast errors are nG distributed because the underlying DGP is nL.

A simple method to construct non-symmetrical FIs is to **combine (C)** the H -step ahead conditional quantile predictor $\{\hat{\xi}_{i,q}(\mathbf{x})\}_{i=1}^n$ ($q \in (0, 1)$) obtained from n different models with weights $w_{i,q}(\mathbf{x})$ based upon within-sample estimation, i.e.,

$$\hat{\xi}_q^{\text{C}}(\mathbf{x}) = \sum_{i=1}^n w_{i,q}(\mathbf{x}) \hat{\xi}_{i,q}(\mathbf{x}).$$



Density forecasts: Let $\{\hat{f}_{i,t}(Y_t|\mathcal{F}_{i,t-1})\}_{t=1}^P$ ($i = 1, \dots, n$) denote n sequences of P individual **one-step ahead** forecast densities at some time t . Then, assuming the density forecasts are continuous, the combined (C) density forecast is defined as

$$\hat{f}_t^C(Y_t) = \sum_{i=1}^n w_i \hat{f}_{i,t}(Y_t|\mathcal{F}_{i,t-1}), \quad (t = 1, \dots, P),$$

with $w_i \geq 0$, $\sum_{i=1}^n w_i = 1$, and $\mathcal{F}_{i,t-1}$ represents the i th information set.

The **key issue** is to find w_i . Most simply, various authors (see, e.g., **Hendry and Clements**, *Econometrics J.*, 2004) advocate the use of **equal weights** $w_i = 1/n$.

A related topic is finding the set of weights that minimize the Kullback–Leibler (KL) divergence between the combined density forecast and the true, but unknown, conditional density $f_t(\cdot|\mathcal{F}_{t-1})$; see, e.g., **Bao et al.** (*JoF*, 2007) and **Hall and Mitchell** (*IJF*, 2007).

○○
 ○○○○○○○○○○
 ○○○
 ○○○○

○
 ○

○○○
 ○○
 ○○
 ○

○○○○○
 ○○○○○○○○
 ○
 ○○○
 ○○
 ●

3.6 Some Software References

- The R-BootPR package provides a way to obtain BS bias-corrected coefficients for forecasting linear AR models. The code can easily be adapted to SETAR-type models.
- The R-hdrcde package contains computer code for the calculation and plotting of HDRs. GAUSS and MATLAB codes for computing the conditional mean, median, mode, SCMI (shortest conditional modal interval) and HDR (highest density region) are available at:
<http://www.jandegooijer.nl>.
- The MATLAB function dmtest retrieves the DM test statistic (under quadratic loss) using the **Newey–West** estimator for the covariance matrix of the loss differential. The R-forecast package contains the function dm.test.



Ph.D. workshop Nonlinear Time Series Analysis and Forecasting Supplement

Jan G. De Gooijer

Amsterdam School of Economics
University of Amsterdam
e-mail: j.g.degooijer@uva.nl

10th Nordic Econometric Meeting, Stockholm
May 22-23, 2019

More information: <http://www.jandegooijer.nl>



Table of Contents

1. Vector Parametric Models and Methods	1 – 1
1.1 General Multivariate	1 – 1
1.2 BL model	1 – 2
1.3 General threshold ARMA (TARMA) model	1 – 5
1.4 Vector STAR	1 – 6
1.5 Other Vector nL Models	1 – 8
1.6 Model Selection, Diagnostic Checking, and Forecasting	1 – 9
1.7 Some Software References	1 – 11
2. Vector Semi- and Nonparametric Methods	2 – 1
2.1 Some Methods	2 – 1
2.2 Nonparametric Causality Testing	2 – 2
2.3 Some Software References	2 – 12

1. Vector Parametric Models and Methods

1.1 General Multivariate

Consider an m -dimensional process $\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{m,t})'$. Let $\mathbf{g}(\cdot) = (g_1(\cdot), \dots, g_m(\cdot))'$ denote a smooth vector function on \mathbb{R}^m , and $\boldsymbol{\theta}$ a vector of unknown parameters. Then denote a **general nL vector time series model** by

$$\mathbf{Y}_t = \mathbf{g}(\mathbf{Y}_{t-1}, \dots, \mathbf{Y}_{t-p}, \boldsymbol{\varepsilon}_{t-1}, \dots, \boldsymbol{\varepsilon}_{t-q}; \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{m,t})' \stackrel{\text{i.i.d.}}{\sim} (\mathbf{0}_m, \boldsymbol{\Sigma}_\varepsilon)$, independent of \mathbf{Y}_t .

Volterra series expansion, i th ($i=1, \dots, m$) component:

$$Y_{i,t} = \mu_i + \varepsilon_{i,t} + \sum_{u=1}^m \sum_{k=1}^{\infty} b_{i,u,k} \varepsilon_{u,t-k} + \sum_{u,v=1}^m \sum_{k,\ell=1}^{\infty} b_{i,u,v,k,\ell} \varepsilon_{v,t-k} \varepsilon_{u,t-\ell} + \dots$$

Multivariate SDM of order (p, q) , i th component:

$$Y_{i,t} = \mu_i(\mathbf{S}_{t-1}) + \sum_{j=1}^p \phi_{i,j}(\mathbf{S}_{t-1}) Y_{i,t-j} + \varepsilon_{i,t} + \sum_{\ell=1}^q \theta_{i,\ell}(\mathbf{S}_{t-1}) \varepsilon_{i,t-\ell},$$

where $\mathbf{S}_t = (\mathbf{Y}'_t, \dots, \mathbf{Y}'_{t-p+1}, \boldsymbol{\varepsilon}'_t, \dots, \boldsymbol{\varepsilon}'_{t-q+1})'$.



1.2 BL model

$$\mathbf{Y}_t = \sum_{j=1}^p \Phi^j \mathbf{Y}_{t-j} + \boldsymbol{\varepsilon}_t + \sum_{j=1}^q \Theta^j \boldsymbol{\varepsilon}_{t-j} + \sum_{u=1}^P \sum_{v=1}^Q \Psi^{uv} \{\boldsymbol{\varepsilon}_{t-v} \otimes \mathbf{Y}_{t-u}\}.$$

Here, $\Phi^j = \{\phi_{i,u}^j, 1 \leq j \leq p\}$ and $\Theta^j = \{\theta_{i,u}^j, 1 \leq j \leq q\}$ are $m \times m$ matrices, and Ψ^{uv} ($1 \leq u \leq P; 1 \leq v \leq Q$) is an $m \times m^2$ matrix with the i th row obtained by vectorizing the $m \times m$ matrix $\psi_i^{uv} = \{\psi_{i,k,\ell}^{uv}, 1 \leq k, \ell \leq m\}$, where k is the row index and ℓ th column index, i.e.

$$\Psi^{uv} = ((\text{vec}(\psi_1^{uv}))', \dots, (\text{vec}(\psi_m^{uv}))')'.$$

Note: The BL model involves $PQm^2 + m(p+q)$ parameters, making it **too general** to be of use in practice. As for the univariate BL model, special cases include the

- **Superdiagonal:** $\psi_{i,k,\ell}^{uv} = 0, \forall u > v.$
- **Subdiagonal:** $\psi_{i,k,\ell}^{uv} = 0, \forall u < v.$
- **Diagonal:** $\psi_{i,k,\ell}^{uv} = 0, \forall u \neq v.$



Example: Consider a **bivariate** ($m = 2$) BL model with $p = P = Q = 1$ and a single lag in the noise term, say at lag $q = 1$. Then the SDM is given by

$$\mathbf{S}_t = \mathbf{F}\boldsymbol{\varepsilon}_t + \mathbf{A}\mathbf{S}_{t-1} + \mathbf{C}_1[\boldsymbol{\varepsilon}_{t-1} \otimes \mathbf{I}_4]\mathbf{S}_{t-1},$$

where

$$\mathbf{F} = \begin{pmatrix} 0.5 & 0 \\ 0 & -0.7 \\ 0.5 & 0 \\ 0 & -0.7 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0.2 & 0.3 & 0 & 0 \\ 0.1 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{C}_1 = \begin{pmatrix} 0.2 & -0.1 & 0 & 0 & 0.1 & 0.3 & 0 & 0 \\ 0.4 & -0.3 & 0 & 0 & -0.3 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The **second order stationarity condition** (see Part I, p. 3–5) is given by $\rho\{(\mathbf{A} \otimes \mathbf{A}) + (\mathbf{C}_1 \otimes \mathbf{C}_1)\mathbf{H}\} < 1$ with $\mathbf{H} = \mathbb{E}[\{\boldsymbol{\varepsilon}_t \otimes \mathbf{I}_4\} \otimes \{\boldsymbol{\varepsilon}_t \otimes \mathbf{I}_4\}]$. The **global invertibility condition** is $\|\boldsymbol{\Psi}^{11}\| \sqrt{\mathbb{E}\|\mathbf{Y}_t\|^2} < 1$, where

$$\boldsymbol{\Psi}^{11} = \begin{pmatrix} 0.2 & -0.1 & 0.1 & 0.3 \\ 0.4 & -0.3 & -0.3 & 0.4 \end{pmatrix}.$$

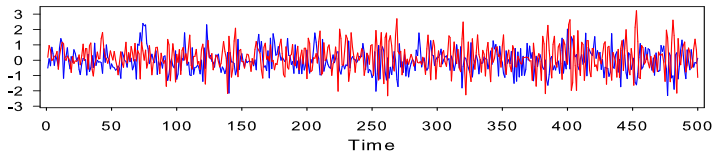


Example: (Cont'd): Replace the expression for the invertibility condition by the approximation $\|\Psi^{11}\|(1,000)^{-1} \sum_{t=1}^{1,000} \|\mathbf{Y}_t\|^2$.

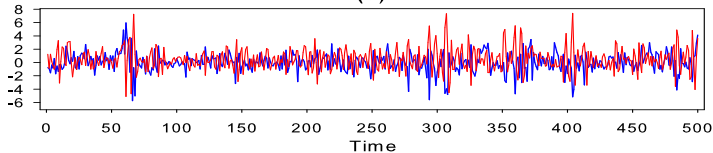
Notes:

- When $\Sigma_\varepsilon = \mathbf{I}_2$, the stationarity condition equals 0.57, and the values of the approximate invertibility condition are in the range (0.71, 1.19) with an average of 0.89. Fig. (a) shows a typical realization. $Y_{1,t}(Y_{2,t}) =$ blue (red) solid line.
- When $\Sigma_\varepsilon = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}$ the stationarity condition equals 0.70. The values of the approximate invertibility condition are in the range (1.26, 1.62), so indicating that the process is **non-invertible**; Fig. (b).

(a)



(b)





1.3 General threshold ARMA (TARMA) model

Let $\{\mathbf{X}_t\}$ denote a weakly stationary m -variate continuous process in \mathbb{R}^m . Assume that \mathbb{R}^m can be partitioned into $k > 1$ non-overlapping subspaces \mathbb{R}_i^m , i.e. $\mathbb{R}_i^m \cap \mathbb{R}_{i'}^m = \emptyset \forall i \neq i' (i, i' = 1, \dots, k)$ determined by the values of $\{\mathbf{X}_{t-d}\}$, where $d > 0$. Then, for an m -dimensional strictly stationary time series process $\{\mathbf{Y}_t\}$, a **VTARMA**($k; p, \dots, p, q, \dots, q$) is defined as

$$\mathbf{Y}_t = \sum_{i=1}^k \left(\Phi_0^{(i)} + \sum_{u=1}^p \Phi_u^{(i)} \mathbf{Y}_{t-u} + \varepsilon_t^{(i)} + \sum_{v=1}^q \Psi_v^{(i)} \varepsilon_{t-v}^{(i)} \right) I((\boldsymbol{\omega}^{(i)})' \mathbf{X}_{t-d} \in \mathbb{R}_i^m),$$

where $\Phi_0^{(i)}$ are $m \times 1$ constant vectors, $\Phi_u^{(i)}$ and $\Psi_v^{(i)}$ are $m \times m$ matrix parameters, and $\boldsymbol{\omega}^{(i)} = (\omega_1^{(i)}, \dots, \omega_m^{(i)})'$ is a pre-specified m -dimensional vector.

When $\boldsymbol{\omega}^{(i)} = (1, 0, \dots, 0)'$, the threshold variable is simply $X_{1,t-d}$ (**Tsay**, JASA, 1998). The error process in the i th regime satisfies $\varepsilon_t^{(i)} = (\boldsymbol{\Sigma}_\varepsilon^{(i)})^{1/2} \boldsymbol{\varepsilon}_t$, where $(\boldsymbol{\Sigma}_\varepsilon^{(i)})^{1/2}$ are symmetric positive definite matrices and $\{\boldsymbol{\varepsilon}_t\} \stackrel{\text{i.i.d.}}{\sim} (\mathbf{0}_m, \mathbf{I}_m)$.

$\{\mathbf{X}_t\}$ can include **lagged values** of $\{\mathbf{Y}_t\}$, or lagged values of an exogenous variable. The order $(p, \dots, p, q, \dots, q)$ can be **different** in each regime. Also, the threshold regimes may include **lagged exogenous** variables.



1.4 Vector STAR

Let $\mathbf{Z}_t = (1, \mathbf{Y}'_{t-1}, \dots, \mathbf{Y}'_{t-p})'$ be an $(mp + 1) \times 1$ vector. Then an m -dimensional k -regime VSTAR model of order $(k; p, \dots, p)$ is defined as

$$\begin{aligned} \mathbf{Y}_t &= \sum_{i=1}^k \left\{ (\Phi_0^{(i)} + \sum_{u=1}^p \Phi_u^{(i)} \mathbf{Y}_{t-u}) (\mathbf{G}_t^{(i-1)} - \mathbf{G}_t^{(i)}) \right\} + \varepsilon_t \\ &= \left\{ \sum_{i=1}^k (\mathbf{G}_t^{(i-1)} - \mathbf{G}_t^{(i)}) (\Phi^{(i)})' \right\} \mathbf{Z}_t + \varepsilon_t, \end{aligned}$$

where $\Phi^{(i)}$ is an $(mp + 1) \times m$ matrix given by

$$\Phi^{(i)} = ((\Phi_0^{(i)})', (\Phi_1^{(i)})', \dots, (\Phi_p^{(i)})')',$$

and $\mathbf{G}_t^{(i)} \equiv \mathbf{G}(\mathbf{X}_t^{(i)}; \gamma^{(i)}, \mathbf{c}^{(i)})$ is a **diagonal matrix** of transition functions

$$\mathbf{G}_t^{(i)} = \text{diag}\{G(X_{1,t}^{(i)}; \gamma_1^{(i)}, \mathbf{c}_1^{(i)}), \dots, G(X_{m,t}^{(i)}; \gamma_m^{(i)}, \mathbf{c}_m^{(i)})\}, \quad (i = 1, \dots, k-1),$$

with $\mathbf{G}_t^{(0)} = \mathbf{I}_m$, $\mathbf{G}_t^{(k)} = \mathbf{0}$, $\gamma^{(i)} = (\gamma_1^{(i)}, \dots, \gamma_m^{(i)})'$ (**slope**), $\mathbf{c}^{(i)} = (\mathbf{c}_1^{(i)}, \dots, \mathbf{c}_m^{(i)})'$ (**location**), $\gamma_j^{(i)} > 0$, $\forall i, j$, and $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (\mathbf{0}_m, \Sigma_\varepsilon)$, independent of \mathbf{Y}_t .

The variable $\mathbf{X}_t^{(i)} = (X_{1,t}^{(i)}, \dots, X_{m,t}^{(i)})'$ ($i = 1, \dots, k-1$) can be a **lagged** variable of one of the components of $\{\mathbf{Y}_t\}$; a **linear** combination of the m series; a **weakly stationary** exogenous variable; or a **deterministic** time trend.



Special case: When $k = 2$, we have

$$\begin{aligned} \mathbf{Y}_t &= \{(\mathbf{I}_m - \mathbf{G}_t^{(1)})(\Phi^{(1)})' + \mathbf{G}_t^{(1)}(\Phi^{(2)})'\} \mathbf{Z}_t + \varepsilon_t \\ &= \Phi_0 + \sum_{u=1}^p \Phi_u \mathbf{Y}_{t-u} + \left\{ \tilde{\Phi}_0 + \sum_{u=1}^p \tilde{\Phi}_u \mathbf{Y}_{t-u} \right\} \mathbf{G}(\mathbf{X}_t; \boldsymbol{\gamma}, \mathbf{c}) + \varepsilon_t, \end{aligned}$$

where $\Phi_u = \Phi_u^{(1)}$, $\tilde{\Phi}_u = \Phi_u^{(2)} - \Phi_u^{(1)}$ ($u = 1, \dots, p$), $\Phi_0 = \Phi_0^{(1)}$, $\tilde{\Phi}_0 = \Phi_0^{(2)} - \Phi_0^{(1)}$, and with $\mathbf{X}_t \equiv \mathbf{X}_t^{(1)}$, $\boldsymbol{\gamma} \equiv \boldsymbol{\gamma}^{(1)}$, and $\mathbf{c} \equiv \mathbf{c}^{(1)}$.

From the 1st expression we see that each location parameter $c_j^{(1)}$ ($j = 1, \dots, m$) represents the **inflection point** in which the transition function has value 1/2, i.e. the process is halfway through the transition from $\mathbf{G}_t^{(1)}$ to $\mathbf{G}_t^{(2)}$.

When the diagonal elements of $\mathbf{G}_t^{(j)}$ are logistic functions, the VSTAR becomes the so-called **logistic VSTAR (LVSTAR)** model. See **Hubrich and Teräsvirta** (ftp://ftp.econ.au.dk/creates/rp/13/rp13_18.pdf, 2013) for a survey.



1.5 Some Other Vector nL Models:

- Vector RCAR (Nicholls and Quinn, *J. Multiv. Anal.*, 1981).
- Vector smooth transition error correction.
- Threshold vector error correction.
- VSETAR with multivariate thresholds.
- LVTAR with nonlinear cointegration Li and He (Report, 2012); http://ideas.repec.org/p/hhs/oruesi/2012_007.html.
- LVSTAR with common nonlinear features (Anderson and Vahid, *J. Econometrics*, 1998).
- Vector contemporaneous-threshold STAR (Dueker et al. *J. Econometrics*, 2011).

Time-domain linearity tests:

- Nonadditivity-type test, Tukey's nonadditivity-type test, and original F test (Harvill and Ray, *Biometrika*, 1999; Psaradakis and Vávra, *Economics Letters*, 2014); Extensions of the tests given in Part II, pp. 2–7 – 2–8.
- LR test statistic for VSETAR (Liu, *Studies in Math. Sciences*, 2011), extending the LR test in Part II, p. 2–4.
- LM-type test statistic for VSTAR (Teräsvirta and Yang, http://www.uclouvain.be/cps/ucl/doc/core/documents/coredp2014_62web.pdf, 2014).



1.6 Model Selection, Diagnostic Checking, and Forecasting

Example: (Model Selection)

Consider a strictly stationary m -dimensional time series $\{\mathbf{Y}_t\}$ and a v -dimensional exogenous variable \mathbf{Z}_t which follow a **VTARX model** of the form

$$\mathbf{Y}_t = \sum_{i=1}^k \left(\Phi_0^{(i)} + \sum_{u=1}^{p_i} \Phi_u^{(i)} \mathbf{Y}_{t-u} + \sum_{r=1}^{q_i} \Psi_r^{(i)} \mathbf{Z}_{t-r} + \varepsilon_t^{(i)} \right) I(X_{t-d} \in \mathbb{R}_i),$$

where $\{X_t\}$ and the # of regimes k are known. For a fixed delay d , the # of data points in regime i equals $T_i = \sum_{t=h+1}^T (I_{t-d}^{(i-1)} - I_{t-d}^{(i)})$, where $h = \max(p_1, \dots, p_k, d)$, and $T =$ total # of observations.

With $p = (p_1, \dots, p_k)$ and $q = (q_1, \dots, q_k)$, the **multivariate** versions of **AIC** and **BIC** (see Part II, p. 3–7) are defined as

$$\text{AIC}(p, q, d, k) = \sum_{i=1}^k (T_i \log |\widehat{\Sigma}_\varepsilon^{(i)}| + 2m(mp_i + vq_i + 1)),$$

$$\text{BIC}(p, q, d, k) = \sum_{i=1}^k (T_i \log |\widehat{\Sigma}_\varepsilon^{(i)}| + \log(T_i)m(mp_i + vq_i + 1)),$$

where $\widehat{\Sigma}_\varepsilon^{(i)}$ is an estimate of the residual covariance matrix in each regime i .



Diagnostic Checking:

- Two portmanteau-type tests statistics (Chabot–Hallé and Duchesne, *Stat. Prob. Letters*, 2008).
- Test statistics based on quantile residuals (Kalliovirta and Saikkonen, <http://blogs.helsinki.fi/lkvaisan/files/2010/08/ReliableResiduals.pdf>, 2010).

Forecasting:

Consider a vector NLAR(p) model. The one-step ($H = 1$) ahead LS forecast of the m -dimensional time series process $\{\mathbf{Y}_t\}$ at time t is given by

$$\mathbf{Y}_{t+1|t}^{\text{LS}} = \mathbb{E}(\mathbf{Y}_{t+1}|\mathcal{F}_t) = \mathbb{E}\{g(\mathbf{Y}_t; \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_{t+1}|\mathcal{F}_t\} \neq g(\mathbf{Y}_t; \boldsymbol{\theta}),$$

since $\mathbb{E}(\boldsymbol{\varepsilon}_{t+1}|\mathcal{F}_t) \neq \mathbf{0}$. When $H = 2$, the LS forecast is given by

$$\begin{aligned} \mathbf{Y}_{t+2|t}^{\text{LS}} &= \mathbb{E}(\mathbf{Y}_{t+2}|\mathcal{F}_t) = \mathbb{E}\{g(\mathbf{Y}_{t+1}; \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_{t+2}|\mathcal{F}_t\} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(g^*(\mathbf{Y}_t; \boldsymbol{\theta}) + \boldsymbol{\eta}_{t+1}) + \boldsymbol{\varepsilon}_{t+2}|\mathcal{F}_t) dF(\boldsymbol{\eta}, \boldsymbol{\varepsilon}), \end{aligned}$$

where $\boldsymbol{\eta}_t$ and $g^*(\cdot)$ are defined in a similar way as $\boldsymbol{\varepsilon}_t$ and $g(\cdot)$ respectively, and $F(\cdot)$ is the joint distribution function of the dependent processes $\{\boldsymbol{\eta}_t\}$ and $\{\boldsymbol{\varepsilon}_t\}$. Use MC simulation or BS procedures.



1.7 Some Software References

- MATHEMATICA source code for testing and estimating bivariate TAR models can be downloaded from <http://www.math.sk/bacigal/homepage/>.
- The website http://repec.wirtschaft.uni-giessen.de/~repec/RePEc/jns/Datenarchiv/v233y2013i1/y233y2013i1p3_21/ provides access to C source code and executable files for multivariate threshold VSETAR analysis using GAs.
- **Yang, Y.** (2012, Appendix) provides a collection of R functions for the specification and evaluation of VSTAR models; see http://pure.au.dk/portal/files/45638557/Yukai_Yang_PhD_Thesis.pdf.

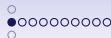


2. Vector Semi- and Nonparametric Methods

2.1 Some Methods

- **Nonparametric methods:** Conditional quantiles, kernel-based forecasting, k -nearest neighbors.
- **Semiparametric methods:** PolyMARS (De Gooijer and Ray, CSDA, 2003), projection-pursuit regression, vector functional-coefficient AR model (Harvill and Ray, *IJF*, 2005), frequency-domain tests (Wong, *JTSA*, 1997).
- **Lag selection:** Kendall's tau and Kendall's partial tau test statistics (Harvill and Ray, *Comm. Stat.: Theory & Methods*, 2000).

See De Gooijer (2017, Chapter 12) for details and examples.



2.2 Nonparametric Causality Testing

General framework:

Assume $\{(X_t, Y_t)\}$ is a strictly stationary bivariate ($m = 2$) time series process. $\{X_t\}$ is a strictly **Granger cause** of $\{Y_t\}$ if past and current values of X_t contain **additional information** on future values of $\{Y_t\}$ that is **not** contained in the past and current Y_t -values alone.

More formally, let $\mathcal{F}_{X,t}$ and $\mathcal{F}_{Y,t}$ denote the information sets consisting of past observations of X_t and Y_t up to and including time t . Then $\{X_t\}$ is a **Granger cause** of $\{Y_t\}$ if, for some $H \geq 1$ (set, e.g., $H = 1$),

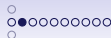
$$(Y_{t+1}, \dots, Y_{t+H})' | (\mathcal{F}_{X,t}, \mathcal{F}_{Y,t}) \stackrel{D}{\not\sim} (Y_{t+1}, \dots, Y_{t+H})' | \mathcal{F}_{Y,t}.$$

For finite orders, define the delay vectors

$$\mathbf{X}_t = (X_t, \dots, X_{t-\ell_X+1})' \text{ and } \mathbf{Y}_t = (Y_t, \dots, Y_{t-\ell_Y+1})', \quad (\ell_X, \ell_Y \geq 1).$$

The null hypothesis of interest is given by

$$\mathbb{H}_0: Y_{t+1} | (\mathbf{X}_t, \mathbf{Y}_t) \sim Y_{t+1} | \mathbf{Y}_t.$$



General framework: (Cont'd)

\mathbb{H}_0 comes down to a statement about the **invariant distribution** of the $d_W = (\ell_X + \ell_Y + 1)$ -dimensional vector $\mathbf{W}_t = (\mathbf{X}'_t, \mathbf{Y}'_t, Z_t)'$ where $Z_t = Y_{t+1}$.

Simplified notation: $\mathbf{W} = (\mathbf{X}', \mathbf{Y}', Z)'$.

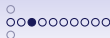
Under \mathbb{H}_0 , the conditional distribution of Z given $(\mathbf{X}', \mathbf{Y}')' = (\mathbf{x}', \mathbf{y}')'$ is the same as that of Z given $\mathbf{Y} = \mathbf{y}$. Then \mathbb{H}_0 can be restated in terms of ratios of joint distributions. **Specifically**, the joint pdf $f_{X,Y,Z}(\mathbf{x}, \mathbf{y}, z)$ and its marginals must satisfy the relationship

$$\frac{f_{X,Y,Z}(\mathbf{x}, \mathbf{y}, z)}{f_{X,Y}(\mathbf{x}, \mathbf{y})} = \frac{f_{Y,Z}(\mathbf{y}, z)}{f_Y(\mathbf{y})},$$

or equivalently

$$\frac{f_{X,Y,Z}(\mathbf{x}, \mathbf{y}, z)}{f_Y(\mathbf{y})} = \frac{f_{X,Y}(\mathbf{x}, \mathbf{y})}{f_Y(\mathbf{y})} \frac{f_{Y,Z}(\mathbf{y}, z)}{f_Y(\mathbf{y})},$$

for each vector $(\mathbf{x}', \mathbf{y}', z)'$ in the support of \mathbf{W} .



A bivariate nL causality test statistic is given by:

$$Q_{T,W}(h) = \frac{\widehat{C}_{X,Y,Z}(h)}{\widehat{C}_Y(h)} - \frac{\widehat{C}_{X,Y}(h)}{\widehat{C}_Y(h)} \frac{\widehat{C}_{Y,Z}(h)}{\widehat{C}_Y(h)},$$

where

$$\widehat{C}_W(h) = \binom{T}{2}^{-1} \sum_{1 \leq i < j \leq T} I(\|\mathbf{W}_i - \mathbf{W}_j\| < h).$$

Since the correlation integral is a U-statistic it can be shown (Hiemstra and Jones (HJ), *J. Finance*, 1994) that, under \mathbb{H}_0 ,

$$\sqrt{T} Q_{T,W}(h) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_W^2(h)), \quad \text{as } T \rightarrow \infty.$$

Disadvantage: For a given nominal size, the actual rejection rate of $Q_{T,W}$ may tend to one as T increases, i.e. the HJ test statistic **over-rejects** \mathbb{H}_0 ; Diks and Panchenko (*Studies in Nonlinear Dynamics & Econometrics*, 2005).



A modified bivariate causality test statistic:

Diks and Panchenko (*J. Econ. Dynamics & Control*, 2006) rewrite the null hypothesis as

$$\mathbb{H}_0: \mathbb{E} \left[\left(\frac{f_{X,Y,Z}(\mathbf{X}, \mathbf{Y}, Z)}{f_Y(\mathbf{Y})} - \frac{f_{X,Y}(\mathbf{X}, \mathbf{Y})}{f_Y(\mathbf{Y})} \frac{f_{Y,Z}(\mathbf{Y}, Z)}{f_Y(\mathbf{Y})} \right) g(\mathbf{X}, \mathbf{Y}, Z) \right] = 0.$$

Here $g(\mathbf{x}, \mathbf{y}, z)$ is a positive weight function which for convenience is set at $g(\mathbf{x}, \mathbf{y}, z) = f_Y^2(\mathbf{y})$, giving more stable results than alternative weight functions.

The corresponding dependence functional is given by

$$\Delta \equiv \mathbb{E}[f_{X,Y,Z}(\mathbf{X}, \mathbf{Y}, Z)f_Y(\mathbf{Y}) - f_{X,Y}(\mathbf{X}, \mathbf{Y})f_{Y,Z}(\mathbf{Y}, Z)] = 0.$$

Under \mathbb{H}_0 the term within square brackets **vanishes**, so that the expectation is zero.



Test statistic:

Let $\hat{f}_W(W_i)$ denote a local density estimator of a d_W -variate random vector \mathbf{W} at W_i defined by

$$\hat{f}_W(W_i) = \frac{(2h)^{-d_W}}{T-1} \sum_{j, j \neq i} I_{ij}^{(W)},$$

where $I_{ij}^{(W)} = I(\|\mathbf{W}_i - \mathbf{W}_j\| < h)$. Given this estimator, the proposed **nonparametric Granger causality** (bivariate) test statistic is given by

$$Q_{T,W}^*(h) = \frac{T-1}{T(T-2)} \sum_i \hat{f}_Y^2(Y_i) \{ \hat{f}_{X,Z|Y}(X_i, Z_i | Y_i) - \hat{f}_{X|Y}(X_i | Y_i) \hat{f}_{Z|Y}(Z_i | Y_i) \}.$$

Assume that $h = cT^{-\beta}$ ($c > 0$, $\beta > 0$), and setting $d_X = d_Y = d_Z = 1$, it can be shown that, as $T \rightarrow \infty$, we have

$$\sqrt{T} \frac{Q_{T,W}^*(h) - \Delta}{\sigma_W(h)} \xrightarrow{D} \mathcal{N}(0, 1), \text{ iff } \frac{1}{2\nu} < \beta < \frac{1}{d_X + d_Y + d_Z},$$

where ν is the order of the density estimation kernel and where

$$\sigma_W^2(h) = 9 \text{Var}(r_0(W_i)), \quad \text{with } r_0(w) = \lim_{h \rightarrow 0} \mathbb{E}(K(w_1, W_2, W_3)),$$

and W_i ($i = 1, 2, 3$) are i.i.d. random variables according to \mathbf{W} .



Example:

Climate variables $Y_{1,t}$ (carbon isotope), $Y_{2,t}$ (oxygen isotope), and $Y_{3,t}$ (insolation). The absence of an arrow from node i to node j means that a $Y_{i,t}$ is a non-Granger cause of $Y_{j,t}$ (H_0 is not rejected).

Note: Very strong nL causal relationship from $Y_{2,t} \rightarrow Y_{1,t}$ and $Y_{2,t} \leftrightarrow Y_{1,t}$. $Y_{3,t}$ is an important driving force for global warming either directly or, mediated by $Y_{2,t}$, indirectly.

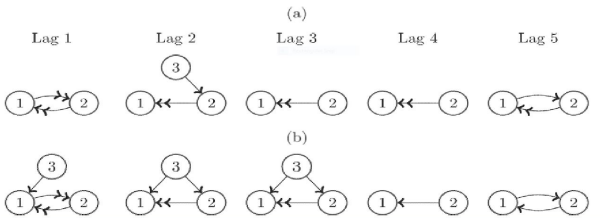


Figure: Nonparametric causality testing: (a) $Q_{T,W}(h)$ (Hiemstra-Jones (HJ) test statistic) and (b) $Q_{T,W}^*(h)$ (modified HJ test statistic). The single arrow symbol marks a p -value in the range 1%–5%, and the double arrow symbol marks a p -value $< 1\%$; $h = 1.5$, $T = 216$.



Some applications:

- **Bekiros and Diks** (*J. Macroeconomics*, 2008): L and nL causal linkages among 6 currencies.
- **De Gooijer and Sivarajasingham** (*Physica A*, 2008): Apply both parametric and nonparametric Granger causality tests to determine linkages between international stock markets.
- **Francis et al.** (*J. Emp. Finance*, 2010): Use both L and nL causality tests to examine the relationship between the returns on large and small firms.

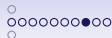
Evaluation:

Diks, C. and Fang, H. (2017). Transfer entropy for nonparametric Granger causality detection: An evaluation of different resampling methods. *Entropy*, 19, 372. <http://doi.org/10.3390/e19070372>.

A modified DP test statistic:

Fang, H. (2018). *Multivariate Density Forecast Evaluation and Nonparametric Granger Causality Testing*. Ph.D. thesis, University of Amsterdam.

<http://dare.uva.nl/search?identifier=cc5207dd-b36c-4494-aa68-f3c0783942c0>.



A multivariate causality test statistic:

(Diks and Wolski, *J. Appl. Econometrics*, 2016).

Consider the strictly stationary multivariate time series process $\{(X_t, Y_t, Q_t)\}$, where $\{X_t\}$ and $\{Y_t\}$ are univariate time series processes, and $\{Q_t\}$ is a **univariate or multivariate** time series process. Then $\{X_t\}$ is a **Granger cause** of $\{Y_t\}$ if, for some $H \geq 1$,

$$(Y_{t+1}, \dots, Y_{t+H})' | (\mathcal{F}_{X,t}, \mathcal{F}_{Y,t}, \mathcal{F}_{Q,t}) \stackrel{D}{\not\sim} (Y_{t+1}, \dots, Y_{t+H})' | \mathcal{F}_{Y,t} \mathcal{F}_{Q,t},$$

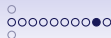
where $\mathcal{F}_{X,t}$, $\mathcal{F}_{Y,t}$, and $\mathcal{F}_{Q,t}$ are the corresponding information sets.

The assumption that both $\{X_t\}$ and $\{Y_t\}$ are scalar-valued time series processes makes it possible to determine whether the causal relationship between these 2 processes is **direct** or **mediated** by other variables.

The **multivariate analogue** of the null hypothesis is given by

$$\mathbb{H}_0: Y_{t+1} | (\mathbf{X}_t, \mathbf{Y}_t, \mathbf{Q}_t) \sim Y_{t+1} | (\mathbf{Y}_t, \mathbf{Q}_t),$$

where $\mathbf{Q}_t = (Q_t, \dots, Q_{t-\ell_Q+1})'$.



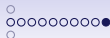
A multivariate causality test statistic: (Cont'd)

Assume $l_X = l_Y = l_Q = 1$. Thus, the dimensionality of $\mathbf{W}_t = (\mathbf{X}'_t, \mathbf{Y}'_t, \mathbf{Q}'_t, Z_t)'$, where $Z_t = Y_{t+1}$, is a number $d_W \geq 4$. The asymptotic normality condition becomes $1/(2\nu) < \beta < 1/d_W$. So, for a standard $\nu = 2$ kernel and $d_W \geq 4$, there is **no** feasible β -region which would endow the test statistic $Q_{T,W}^*(h)$ with asymptotic normality.

“**Solution**”: Improve the precision of the density estimator by **reducing** the kernel estimator **bias** using **data-sharpening** (Hall and Minotte, *JRSS-B*, 2002). The sharpened (s) form of the plug-in density estimator is given by

$$\hat{f}_W^s(W_i) = \frac{h^{-d_W}}{T-1} \sum_{j, j \neq i} K\left(\frac{W_i - \psi_p(W_j)}{h}\right),$$

where $\psi_p(\cdot)$ is a so-called **sharpening function**, with p the order of bias reduction.



A multivariate causality test statistic: (Cont'd)

The **sharpened form** of the causality test statistic is given by

$$Q_{T,w}^s(h) = \frac{T-1}{T(T-2)} \sum_i (\widehat{f}_{X,YZ}^s(X_i, Y_i, Z_i) \widehat{f}_Y^s(Y_i) - \widehat{f}_{X,Y}^s(X_i, Y_i) \widehat{f}_{Y,Z}^s(Y_i, Z_i)).$$

Under certain mixing conditions it can be shown that, as $T \rightarrow \infty$,

$$\sqrt{T} \frac{Q_{T,w}^s(h) - \Delta}{S_T} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \text{ iff } \frac{1}{2p} < \beta < \frac{1}{d_w},$$

where S_T^2 is a consistent estimator of the asymptotic variance of $\sqrt{T} (Q_{T,w}^s(h) - \Delta)$.



2.3 Some Software References

- PolyMARS (or PMARS) is available in the R-polspline package. The R-fRegression package has an option for computing a PMARS model as a part of the function regFit.
- The function ppr in the R-stat package, and the function ppreg in S-Plus both allow for PPR model fitting with multivariate responses.
- R codes for performing the HJ (hj.r) and the DP (dp.r) nonparametric test statistics are available at the website <http://www.jandegooijer.nl>. C source code, and an executable file, for computing both test statistics can be downloaded from

<http://www1.fee.uva.nl/cendef/upload/6/hjt2.zip>.

Alternatively, a windows version and C source code are available at <http://research.economics.unsw.edu.au/vpanchenko/#software>. C source code for the multivariate nonlinear nonparametric Granger causality test is available at

<http://qed.econ.queensu.ca/jae/datasets/diks001/>.