Nonlinear time series

Based on the book by FAN/YAO: *Nonlinear Time Series*

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

October 27, 2009

Outline

What is a nonlinear time series?

Formal definition: a *nonlinear process* is any stochastic process that is not linear. To this aim, a *linear process* must be defined. Realizations of time-series processes are called time series but the word is also often applied to the generating processes.

Intuitive definition: nonlinear time series are generated by nonlinear dynamic equations. They display features that cannot be modelled by linear processes: time-changing variance, asymmetric cycles, higher-moment structures, thresholds and breaks.

Definition of a linear process

Definition

A stochastic process $(X_t, t \in \mathbf{Z})$ is said to be a *linear process* if for every $t \in \mathbf{Z}$

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},$$

where $a_0 = 1$, $(\varepsilon_t, t \in \mathbf{Z})$ is *iid* with $\mathrm{E}\varepsilon_t = 0$, $\mathrm{E}\varepsilon_t^2 < \infty$, and $\sum_{j=0}^{\infty} |a_j| < \infty$.

For comparison: the Wold theorem

Theorem (Wold's Theorem)

Any covariance-stationary process (X_t) *has a unique representation as the sum of a purely deterministic component and an infinite sum of white-noise terms, in symbols*

$$X_t = \delta_t + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},$$

with $a_0 = 1$, $\sum_{j=0}^{\infty} a_j^2 < \infty$, *and the terms* ε_t *defined as the linear innovations* $X_t - \mathrm{E}^*(X_t|\mathcal{H}_{t-1})$, *where* E^* *denotes the linear expectation or projection on the space* $\mathcal{H}_{t-1}$ *that is generated by the observations* $X_s, s \leq t-1$.

Linear processes and the Wold representation

- There are many covariance-stationary processes that are not linear: either the innovations are not independent (though white noise) or the absolute coefficients do not converge.
- If it is just the absolute coefficients, the processes are *long memory*. Those are not really nonlinear, and we will not handle them here.
- Many nonlinear processes have a Wold representation that looks linear. It is correct but it just describes the autocovariance structure. It is an incomplete representation.
- If $\mathrm{E}\varepsilon_t^2 < \infty$ is violated, (X_t) becomes an infinite-variance linear process, conditions on coefficient series must be adjusted. Outside the scope here.

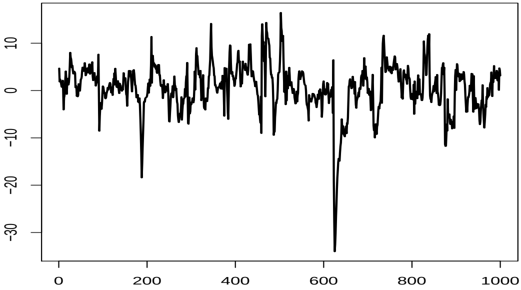
A simple example: an AR-ARCH model

The dynamic generating law

$$\begin{aligned} X_t &= 0.9X_{t-1} + u_t, && (1)\\ u_t &= h_t^{0.5}\varepsilon_t, && (2)\\ h_t &= 1 + 0.9u_{t-1}^2, \quad \varepsilon_t \sim NID(0,1), && (3) \end{aligned}$$

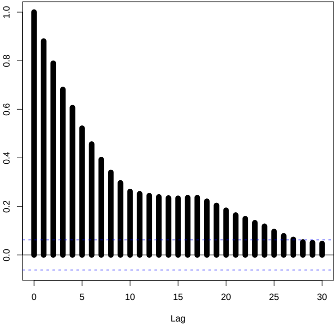
defines a stable AR-ARCH process. It *is* an AR(1) model with Wold innovations u_t, which are white noise but not independent.

A time series plot of 1000 observations



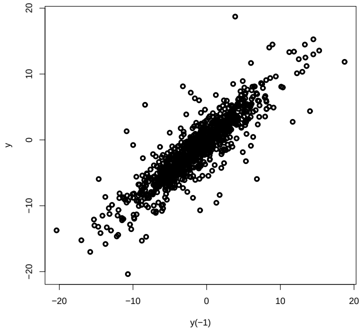
Impression: not too nonlinear, but outlier patches point to fat-tailed distributions: variable has no finite kurtosis.

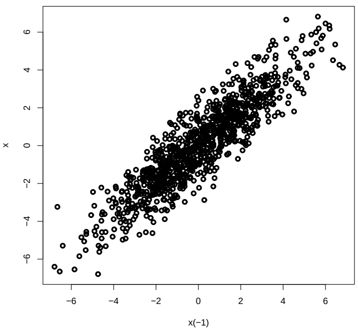
Correlograms



Impression: AR-ARCH on the left inconspicuous, fairly identical to a standard AR(1) correlogram on the right.

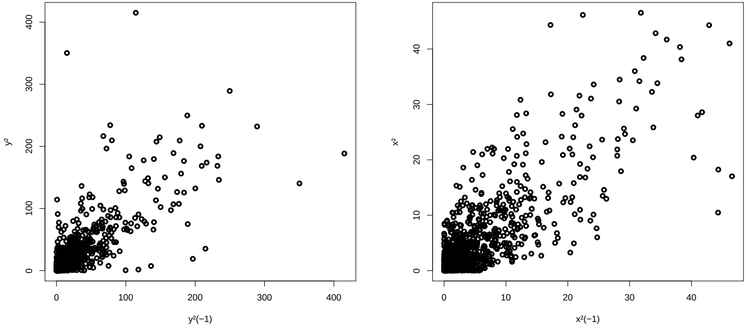
Plots versus lags





Impression: a bit more dispersed than the standard AR(1) plot on the right: leptokurtosis. Basically linear (as should be).

Plots of squares versus lagged squares



Impression: even this device, suggested by Andrew A. Weiss allows no reliable discrimination to the standard AR case on the right.

I. Characteristics of Time Series

The meaning of this section:

This section corresponds to Section 2 of the book by FAN & YAO and is meant to review the basic concepts of (mostly linear) time-series analysis.

Stationarity

Definition

A time series $(X_t, t \in \mathbb{Z})$ is (weakly, covariance) *stationary* if (a) $\mathrm{E}X_t^2 < \infty$, (b) $\mathrm{E}X_t = \mu \ \forall t$, and (c) $\mathrm{cov}(X_t, X_{t+k})$ is independent of $t \ \forall k$.

Remark. For $k = 0$, this definition yields time-constant finite variance.

Definition

A time series $(X_t, t \in \mathbb{Z})$ is *strictly stationary* if $(X_1, \ldots, X_n)$ and $(X_{1+k}, \ldots, X_{n+k})$ have the same distribution for any $n \geq 1$ and $n, k \in \mathbb{Z}$.

Remark. For nonlinear time series, often strict stationarity is the more 'natural' concept.

ARMA processes

The ARMA(p, q) model with $p, q \in \mathbb{N}$

$$X_t = b_1 X_{t-1} + \ldots + b_p X_{t-p} + \varepsilon_t + a_1 \varepsilon_{t-1} + \ldots + a_q \varepsilon_{t-q},$$

for (ε_t) white noise, is technically rewritten as

$$b(B) X_t = a(B) \varepsilon_t,$$

with polynomials $a(z)$ and $b(z)$ and the backshift operator B defined by $B^k X_t = X_{t-k}$.

This is the most popular linear time-series model. Often, the expression *ARMA process* is reserved for stable polynomials.

Stationarity of ARMA processes

Theorem

The process defined by the ARMA(p,q) model is stationary if $b(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$, assuming that $a(z)$ and $b(z)$ have no common factors.

Remark. Note that $t \in \mathbb{Z}$. If $t \in \mathbb{N}$, the ARMA process can be 'started' from arbitrary conditions, and the usual distinction of 'stable' and 'stationary' applies. Clearly, here a pure MA process $(p = 0)$ is always stationary.

Causal time series

Definition

A time series (X_t) is *causal* if for all t

$$X_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j}, \sum_{j=0}^{\infty} |d_j| < \infty,$$

with white noise (ε_t).

Remark. A causal process is always stationary. $X_t = 2X_{t-1} + \varepsilon_t$ violates the ARMA stability conditions and nonetheless has a stationary non-causal solution. These are at odds with intuition and will be excluded.

Stationary Gaussian processes

A process (X_t) is called *Gaussian* if all its finite-dimensional marginal distributions are normal. For Gaussian processes, the Wold representation holds with *iid* innovations.

- The purely nondeterministic part of a Gaussian process is *linear*.
- A Gaussian MA(q) process is *q–dependent*, i.e. X_t and X_{t+q+k} are independent for all $k \geq 1$.
- For a Gaussian AR process, X_t is independent of $X_{t-k}, k > p$ given $X_{t-1}, \ldots, X_{t-p}$.

Autocovariance and autocorrelation

Definition

Let (X_t) be stationary. The *autocovariance function* (ACVF) of (X_t) is

$$\gamma(k) = \mathrm{cov}(X_{t+k}, X_t), \quad k \in \mathbb{Z}.$$

The *autocorrelation function* (ACF) of (X_t) is

$$\rho(k) = \gamma(k)/\gamma(0) = \mathrm{corr}(X_{t+k}, X_t), \quad k \in \mathbb{Z}.$$

It follows that $\gamma(-k) = \gamma(k)$ and $\rho(-k) = \rho(k)$ (even functions).

Characterization of the ACVF

Theorem

A function $\gamma(.) : \mathbb{Z} \to \mathbb{R}$ is the ACVF of a stationary time series if and only if it is even and nonnegative definite in the sense that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \gamma(i-j) \geq 0$$

for all integer $n \geq 1$ and arbitrary real $a_1, \ldots, a_n$.

Remark. One direction is easy to show, as it uses the properties of a covariance matrix of $X_t, \ldots, X_{t-n+1}$. The reverse direction is hard to prove and needs a Theorem by Kolmogorov.

ACF of stationary ARMA processes

Any ARMA process has an MA(∞) representation

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},$$

with $\sum_{j=0}^{\infty} |a_j| < \infty$. It follows that

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \sum_{j=0}^{\infty} a_j a_{j+k}, \quad \rho(k) = \frac{\sum_{j=0}^{\infty} a_j a_{j+k}}{\sum_{j=0}^{\infty} a_j^2}.$$

Clearly, for pure MA(q) processes, these expressions become 0 for $k > q$.

Properties of the ACF for ARMA processes

Proposition

1. *For causal ARMA processes,* $\rho(k) \to 0$ *like* c^k *for* $|c| < 1$ *as* $k \to \infty$ *(exponential);*
2. *for MA(q) processes,* $\rho(k) = 0$ *for* $k > q$.

This proposition does not warrant a clear distinction between AR and general ARMA processes.

Estimating the ACF

The sample ACF (the *correlogram*) is defined by $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$, where

$$\hat{\gamma}(k) = \frac{1}{T}\sum_{t=1}^{T-k}(X_t - \bar{X}_T)(X_{t+k} - \bar{X}_T),$$

for small k, for example $k \leq T/4$ or $k \leq 2\sqrt{T}$, with $\bar{X}_T = T^{-1}\sum_{t=1}^{T} X_t$.

Statistical properties of the mean estimate

Theorem

Let (X_t) is a linear stationary process defined by $X_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, with (ε_t) iid$(0, \sigma^2)$ and $\sum_{j=0}^{\infty} |a_j| < \infty$. If $\sum_{j=0}^{\infty} a_j \neq 0$, $\sqrt{T}(\bar{X}_T - \mu) \Rightarrow N(0, \nu_1^2)$, where

$$\nu_1^2 = \sum_{j=-\infty}^{\infty} \gamma(j) = \sigma^2 \left(\sum_{j=0}^{\infty} a_j \right)^2 .$$

Remark. The variance of the mean estimate depends on the spectrum at 0 and may be called the long-run variance.

The long-run variance

A stationary time-series process $X_t = \sum a_j \varepsilon_{t-j}$ has the variance

$$\mathrm{var} X_t = \gamma(0) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} a_j^2,$$

which usually differs from the *long-run variance*

$$\sigma_\varepsilon^2 (\sum_{j=0}^{\infty} a_j)^2.$$

The long-run variance may also be seen as the spectrum at 0, $\sum_{j=-\infty}^{\infty} \gamma(j)$, or as the limit variance

$$\lim_{n\to\infty} n^{-1} \mathrm{var} \sum_{t=0}^{n} X_t.$$

For white noise (X_t), variance and long-run variance coincide.

Statistical properties of the variance estimate

Theorem

Let (X_t) is a linear stationary process defined by $X_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, with (ε_t) iid$(0, \sigma^2)$ and $\sum_{j=0}^{\infty} |a_j| < \infty$. If $\mathrm{E}\varepsilon_t^4 < \infty$, $\sqrt{T}\{\hat{\gamma}(0) - \gamma(0)\} \Rightarrow N(0, \nu_2^2)$, where

$$\nu_2^2 = 2\sigma^2(1 + 2\sum_{j=1}^{\infty} \rho(j)^2).$$

Remark. This is reminiscent of the portmanteau Q.

Statistical properties of the correlogram

Theorem

Let (X_t) is a linear stationary process defined by $X_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, with (ε_t) iid$(0, \sigma^2)$ and $\sum_{j=0}^{\infty} |a_j| < \infty$. If $\mathrm{E}\varepsilon_t^4 < \infty$, $\sqrt{T}\{\hat{\varrho}(k) - \varrho(k)\} \Rightarrow N(0, \mathbf{W})$, where

$$\begin{aligned} w_{ij} &= \sum_{t=-\infty}^{\infty} \{\rho(t+i)\rho(t+j) + \rho(t-i)\rho(t+j) + 2\rho(i)\rho(j)\rho(t)^2 \\ &\quad -2\rho(i)\rho(t)\rho(t+j) - 2\rho(j)\rho(t)\rho(t+i)\}, \end{aligned}$$

where $\varrho(k) = (\rho(1), \ldots, \rho(k))'$.

Remark. This is Bartlett's formula, impressive but not immediately useful. In simple cases, it allows determining confidence bands for the correlogram. White noise yields $\sqrt{T}\hat{\rho}_k \Rightarrow N(0,1)$ for $k \neq 0$.

Partial autocorrelation function

Definition

(X_t) is a stationary process with $\mathrm{E}X_t = 0$. The *partial autocorrelation function* (PACF) $\pi : \mathbb{N} \to [-1, 1]$ is defined by $\pi(1) = \rho(1)$ and

$$\pi(k) = \mathrm{corr}(R_{1|2,\ldots,k}, R_{k+1|2,\ldots,k}),$$

for $k \geq 2$, where $R_{j|2,\ldots,k}$ denotes residuals from regressing X_j on $X_2, \ldots, X_k$ by least squares.

Remark. For non-Gaussian processes, the thus defined PACF does not necessarily correspond to partial correlations, if partial correlations are defined via conditional expectations.

Properties of the PACF

Proposition

1. *For any stationary process, $\pi(k)$ is a function of the ACVF values $\gamma(1), \ldots, \gamma(k)$;*
2. *For an AR(p) process, $\pi(k) = 0$ for $k > p$, i.e. the PACF 'cuts off at p'.*

Remark. Mathematically, the PACF does not provide any new information on top of the ACVF. The sample PACF facilitates visual pattern recognition and may indicate AR models and their lag order.

Mixing

White noise is usually insufficient for ergodic theorems—such as laws of large numbers (LLN) or central limit theorems (CLT). For linear processes with *iid* innovations, some moment conditions will suffice. For nonlinear processes, more is needed.

Generally, mixing conditions guarantee that X_t and X_{t+h} are more or less independent for large h. The metaphor is mixing drinks: a drop of a liquid will not remain close to its origin. If we mix two separate glasses, the outcome will be stationary but not mixing.

Five mixing coefficients

$$
\begin{aligned}
\alpha(n) &= \sup_{A\in\mathcal{F}_{-\infty}^{0},B\in\mathcal{F}_{n}^{\infty}} |P(A)P(B)-P(AB)|,\\
\beta(n) &= \mathrm{E}\left\{\sup_{B\in\mathcal{F}_{n}^{\infty}} |P(B)-P(B|X_0,X_{-1},X_{-2},\ldots)|\right\},\\
\rho(n) &= \sup_{X\in\mathcal{L}^2(\mathcal{F}_{-\infty}^{0}),Y\in\mathcal{L}^2(\mathcal{F}_{n}^{\infty})} |\mathrm{corr}(X,Y)|,\\
\phi(n) &= \sup_{A\in\mathcal{F}_{-\infty}^{0},B\in\mathcal{F}_{n}^{\infty},P(A)>0} |P(B)-P(B|A)|,\\
\psi(n) &= \sup_{A\in\mathcal{F}_{-\infty}^{0},B\in\mathcal{F}_{n}^{\infty},P(A)P(B)>0} |1-P(B|A)/P(B)|.
\end{aligned}
$$

Here, $\mathcal{F}_k^l$ is the σ–algebra generated by $X_t, k \leq t \leq l$, and $\mathcal{L}^2$ are square integrable functions. Good mixing coefficients converge to 0 as $n \to \infty$.

Mixing processes

Definition

A process (X_t) is α–*mixing* ('strong mixing') if $\alpha(n) \to 0$ as $n \to \infty$, and similarly for β–*mixing* etc.

A basic property is

$$\alpha(k) \leq \frac{1}{4}\rho(k) \leq \frac{1}{2}\sqrt{\phi(k)},$$

and generally ϕ–mixing implies ρ–mixing (and β–mixing), and ρ–mixing (or β–mixing) implies α–mixing. ψ–mixing implies ϕ–mixing and thus all others. Even for simple examples, the mixing coefficients cannot be evaluated directly.

Some properties of mixing processes

1. If (X_t) is mixing (any definition) and $m(.)$ is a measurable function, then $(m(X_t))$ is again mixing. The property is 'hereditary'.
2. If (X_t) is a linear ARMA process and ε_t has a density, it is β–mixing with $\beta(n) \to 0$ exponentially.
3. A strictly stationary process on $\mathbb{Z}$ is mixing iff its restriction to $\mathbb{N}$ is mixing (any definition).
4. Finite-dependent (such as strict MA) or independent processes are mixing.
5. Deterministic processes are not mixing.

A LLN for α–mixing processes

Proposition

Assume (X_t) is strictly stationary and α–mixing, and $\mathrm{E}|X_t| < \infty$. Then, as $n \to \infty$, $S_n/n \to \mathrm{E}X_t$ a.s., where $S_n = \sum_{t=1}^{n} X_t$.

A comparable LLN with slightly stronger conditions guarantees convergence of $n^{-1}\mathrm{var}S_t$ to the long-run variance.

A CLT for α–mixing processes

Theorem

Assume (X_t) is strictly stationary with $\mathrm{E}X_t = 0$, the long-run variance σ^2 is positive, and one of the following conditions hold

1. $\mathrm{E}|X_t|^\delta < \infty$ *and* $\sum_{j=0}^{\infty} \alpha(j)^{1-2/\delta} < \infty$ *for some* $\delta > 2$;
2. $P(|X_t| < c) = 1$ *for some* $c > 0$ *and* $\sum_{j=1}^{\infty} \alpha(j) < \infty$.

Then,

$$S_n/\sqrt{n} \Rightarrow N(0, \sigma^2),$$

with $S_n = \sum_{t=1}^{n} X_t$.

Remark. The theorem cares for processes with bounded support (case 2) as well as for some quite fat-tailed ones (case 1 for $\delta = 2 + \epsilon$), which however need fast decrease of their mixing coefficients.