# 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

April 23, 2013

# 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* $\varepsilon_t$ *defined as the linear innovations* $X_t - \mathrm{E}^*(X_t|\mathcal{H}_{t-1})$, *where* $\mathrm{E}^*$ *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.

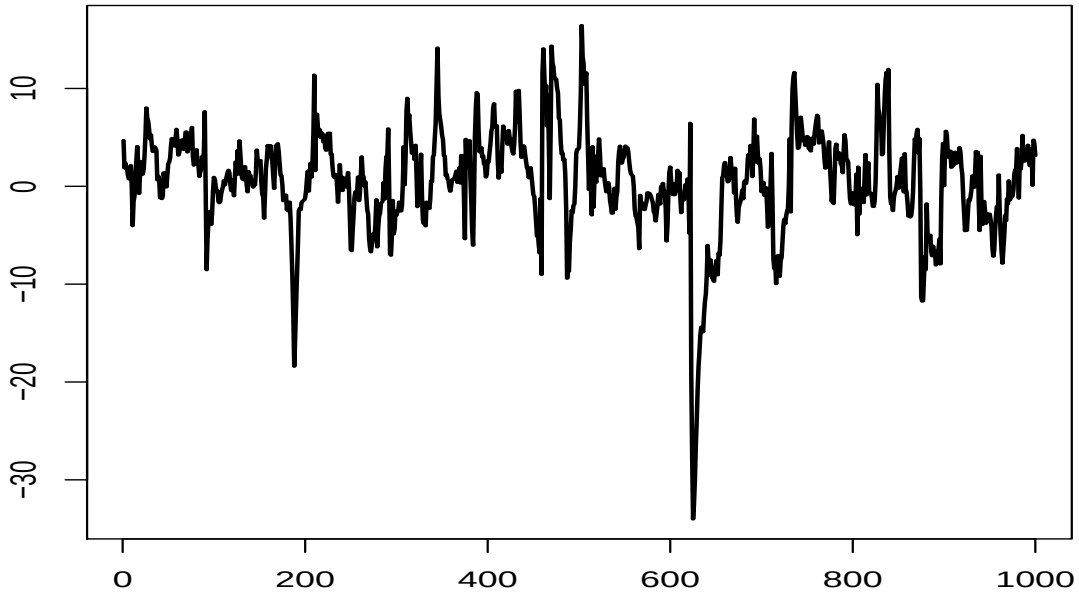
# A simple example: an AR-ARCH model

The dynamic generating law

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

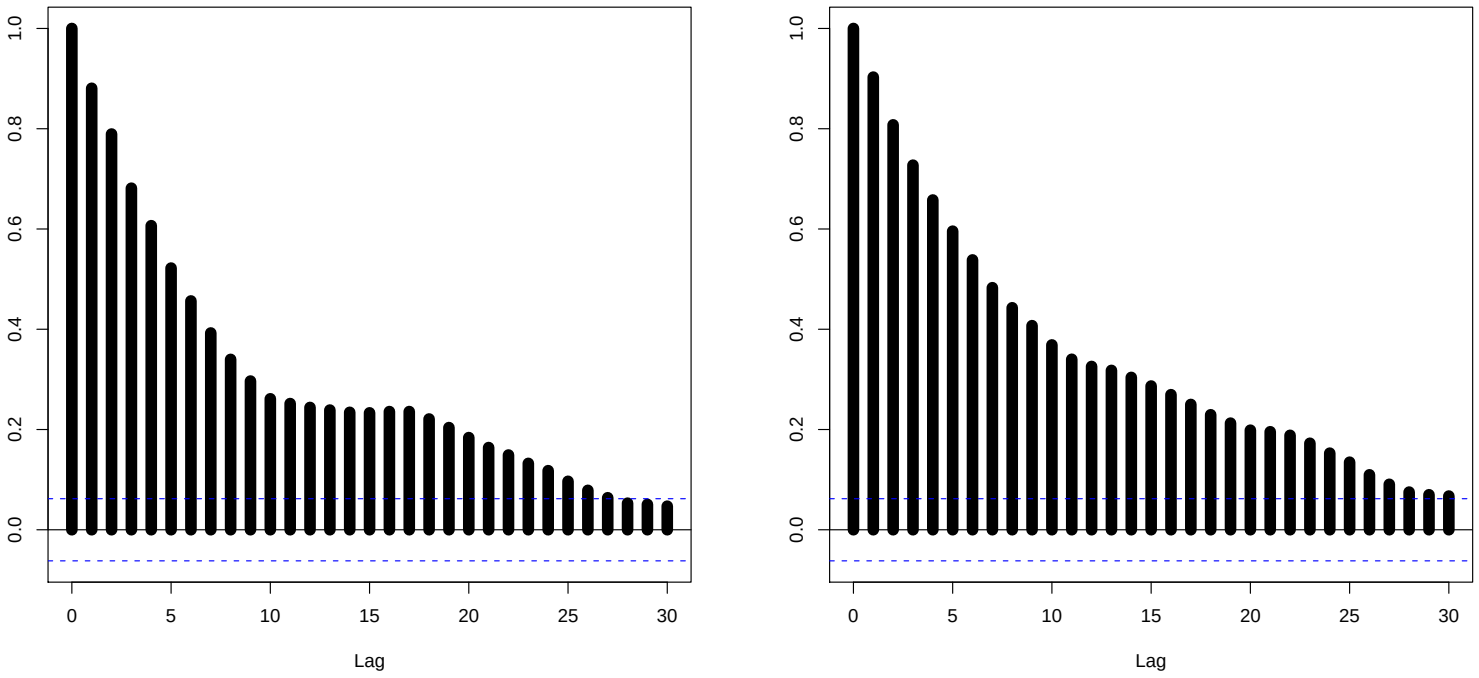
defines a stable AR-ARCH process. It *is* an AR(1) model with Wold innovations $u_t$, which are white noise but not independent.

# A time series plot of 1000 observations



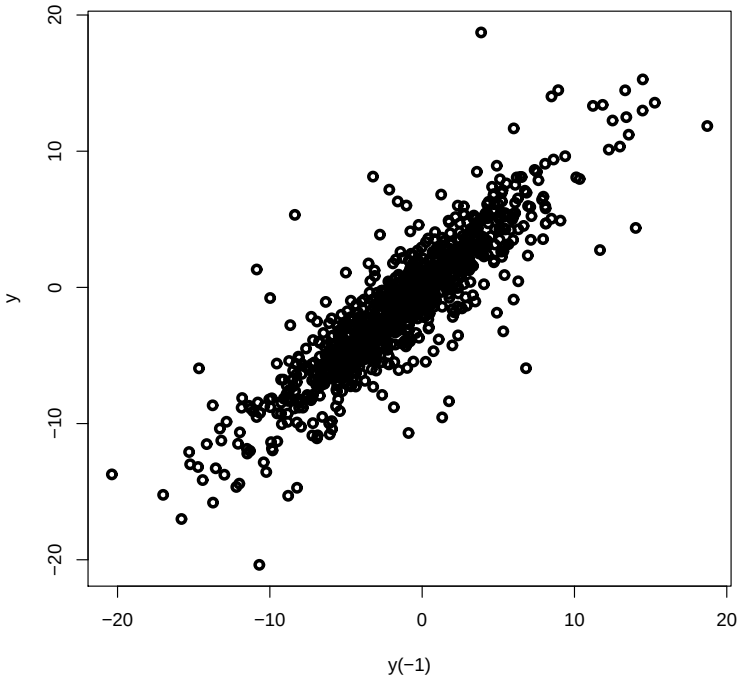

Impression: not too nonlinear, but outlier patches point to fat-tailed distributions: variable has no finite kurtosis.

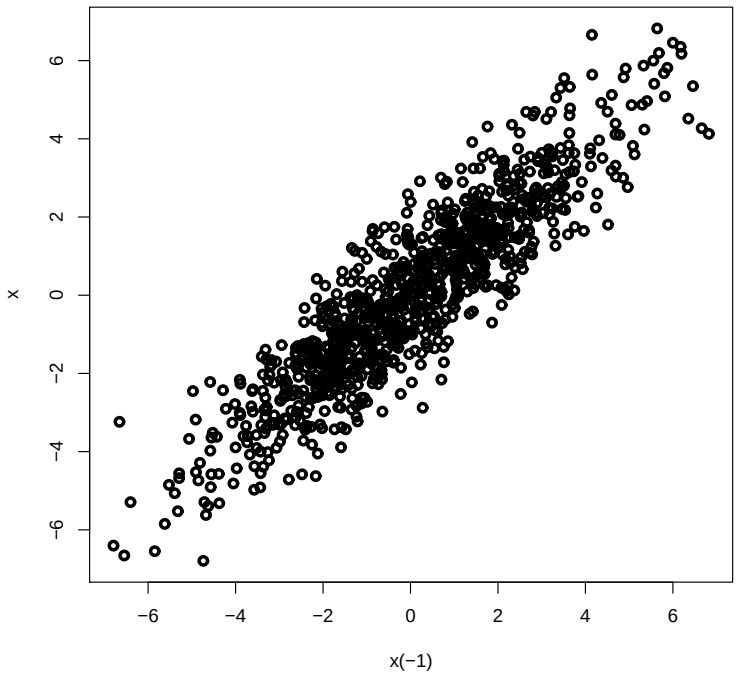
# Correlograms




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

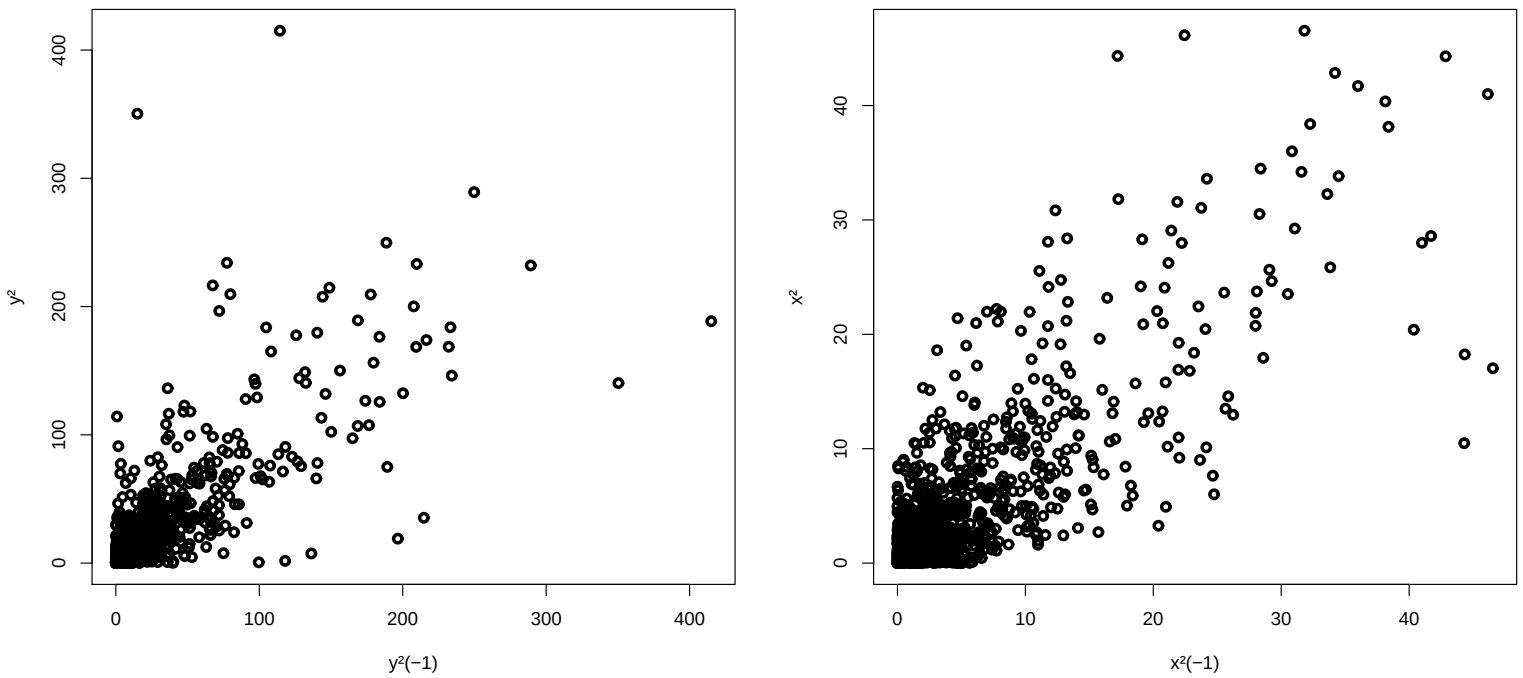
# Plots versus lags




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 $(\varepsilon_t)$ 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 $(\varepsilon_t)$.

**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($\infty$) representation

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

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

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

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* $(\varepsilon_t)$ *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* $(\varepsilon_t)$ *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* $(\varepsilon_t)$ *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

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

Here, $\mathcal{F}^l_k$ is the $\sigma$–algebra generated by $X_t, k \le t \le l$, 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 $\alpha$–*mixing* ('strong mixing') if $\alpha(n) \to 0$ as $n \to \infty$, and similarly for $\beta$–*mixing* etc.

A basic property is

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

and generally $\phi$–mixing implies $\rho$–mixing (and $\beta$–mixing), and $\rho$–mixing (or $\beta$–mixing) implies $\alpha$–mixing. $\psi$–mixing implies $\phi$–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 $\varepsilon_t$ has a density, it is $\beta$–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 $\alpha$–mixing processes

## Proposition

*Assume $(X_t)$ is strictly stationary and $\alpha$–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 $n^{-1}\mathrm{var}S_n$ to the long-run variance.

# A CLT for $\alpha$–mixing processes

## Theorem

*Assume $(X_t)$ is strictly stationary with $\mathrm{E}X_t = 0$, the long-run variance $\sigma^2$ 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.

# II. Threshold models

Threshold autoregressions (TAR) have been introduced by Howell Tong as SETAR (*self-exciting threshold autoregressions*). Building on the simple autoregression, they cover the feature that a variable behaves differently in expansions and contractions. They are used in economics and in many other sciences.

# The definition of TAR

## Definition

A threshold autoregressive (TAR) model with $k \geq 2$ is defined by

$$X_t = \sum_{j=1}^{k} \{b_{j0} + b_{j1} X_{t-1} + \ldots + b_{j,p_j} X_{t-p_j} + \sigma_j \varepsilon_t\} I(X_{t-d} \in A_j),$$

where $(\varepsilon_t)$ is $iid(0,1)$, $d$ is the delay, $A_j$ form a partition of $\mathbb{R}$, and lag orders $p_j$ may change across the sets $A_j$.

# Stability of TAR models

**Sufficiency**. Clearly, the TAR model is stable if $\sigma_1 = \ldots = \sigma_k$ and all autoregressive submodels are stable.

**Necessity**. No general criterion for the stability of TAR models exists. Such criteria have been elaborated for simple cases, such as $p_1 = \ldots = p_k = 1$. Apparently, it suffices that all submodels on unbounded sets $A_j$ are stable. However, there exist simple cases of stable TAR with all submodels unstable.

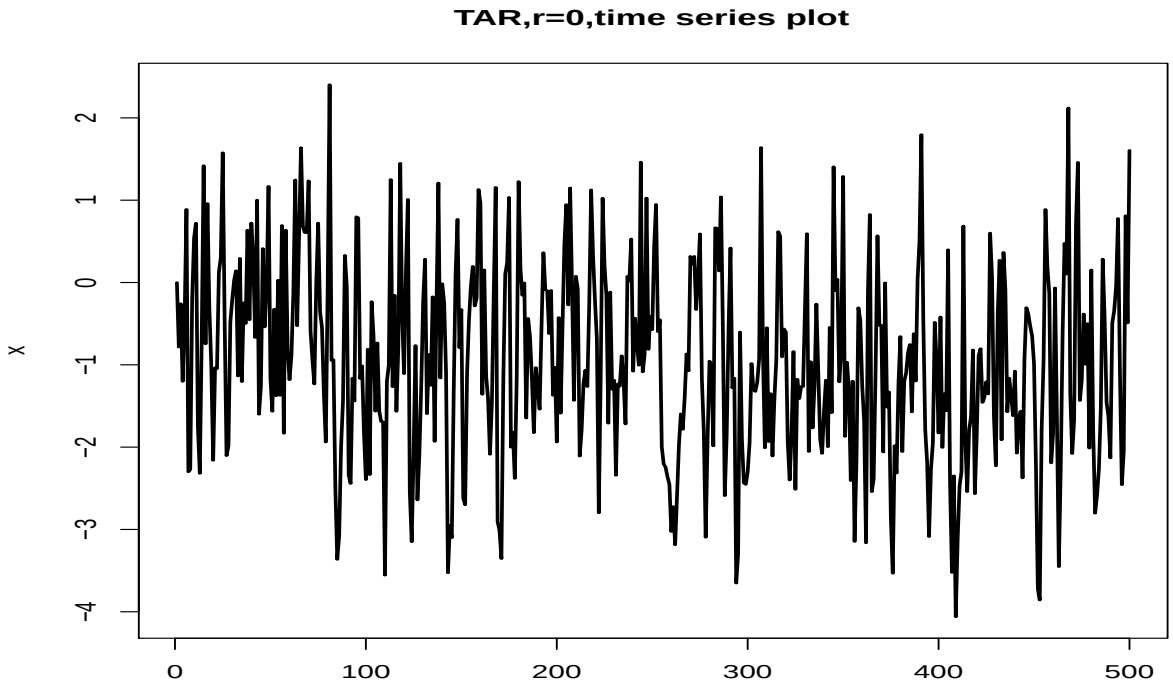
# An example for a TAR model

Consider the two-regime model

$$X_t = \begin{cases} -0.7X_{t-1} + \varepsilon_t, & X_{t-1} \geq r, \\ 0.7X_{t-1} + \varepsilon_t, & X_{t-1} < r, \end{cases}$$

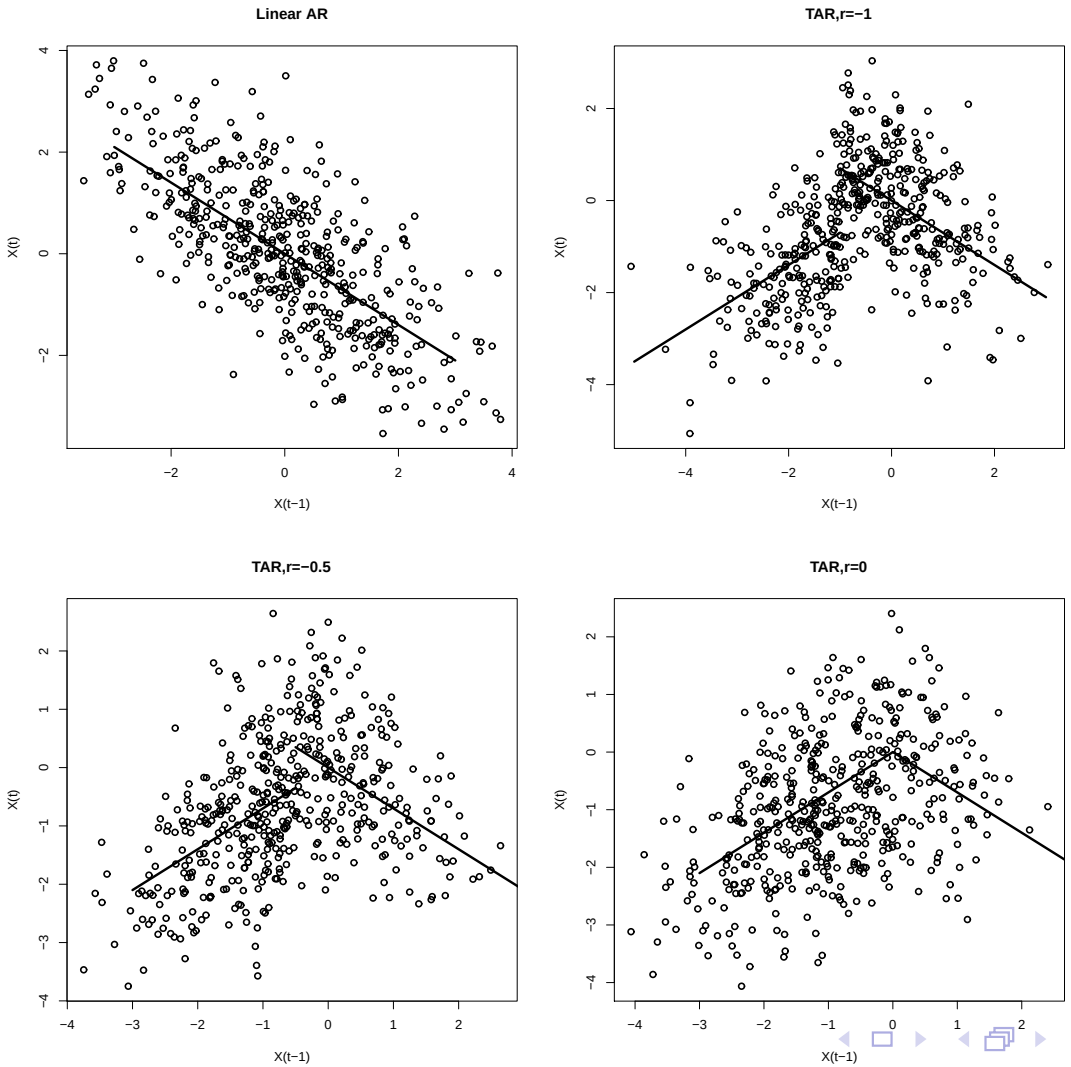
with $r$ varied in $\{-\infty, -1, -0.5, 0\}$. Note that $r = -\infty$ yields the linear AR model. We generate trajectories of 500 starting from zero.

# A time series plot for $r = 0$

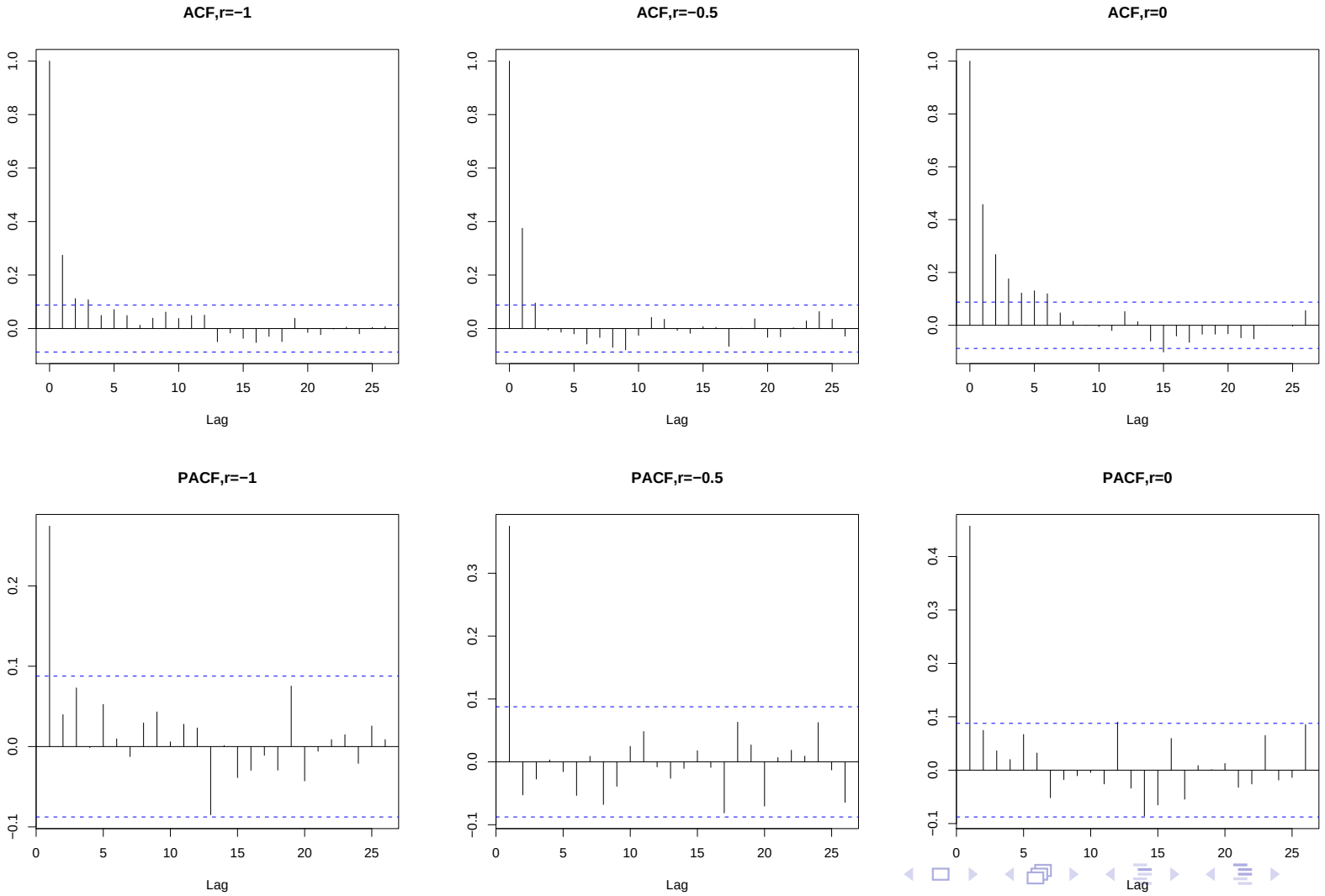



Impression: nonlinear feature not recognizable.

# Plots versus lags



# Sample ACF and PACF



# Estimation of TAR models

**Main idea**. Within each 'regime', the model is linear. Given the delay $d$ and $A_j, j = 1, \ldots, k$, 'profile' maximum likelihood is well approximated by least squares.

Therefore, one may start by $k = 2$, $d = 1$, and perform a 'grid search' over thresholds in an inner region of the sample, 50% or 90%. Optimize over $p$ using AIC. Then, increase $d$ and maybe $k$, and optimize again.

# The AIC to be minimized for TAR

$$\mathrm{AIC}(\{p_j\}) = \sum_{j=1}^{k} [T_j \log\{\hat{\sigma}_j^2(p_j)\} + 2(p_j + 1)],$$

with $\{p_j\}$ indicating the $k$ lag orders and $\hat{\sigma}_j^2$ estimated by least squares from the $T_j$ observations that belong to the regime $j$. Higher $k$ and $p_j$ increase the penalty term but not $d$. Long delays are not plausible.

# The consistency of the estimator

**Theorem**

*Assume $(X_t)$ is TAR with $k = 2$, $p_1 = p_2 = p$, and is ergodic (mixing) and strictly stationary with finite variance, and that all densities have support $\mathbb{R}$. Then, all ML estimators for coefficients, variances, threshold, and delay are strongly consistent.*

**Remark**. The theorem assumes a real threshold between regimes $r$, not general sets $A_1$ and $A_2$. For other, more complex models, the asymptotic properties are largely unknown.

# Asymptotic normality of the estimator

**Theorem**

*With the assumptions of the consistency theorem and*

1. *geometric ergodicity;*
2. $\varepsilon_t$ *has a positive and continuous density, and fourth moments of* $\varepsilon_t$ *and of* $X_t$ *exist;*
3. *the autoregressive function is discontinuous across regimes;*

*it follows that* $T(\hat{r} - r) = O_p(1)$ *(superconsistency), and the coefficient estimates are asymptotically independent of* $\hat{r}$ *and normally distributed.*

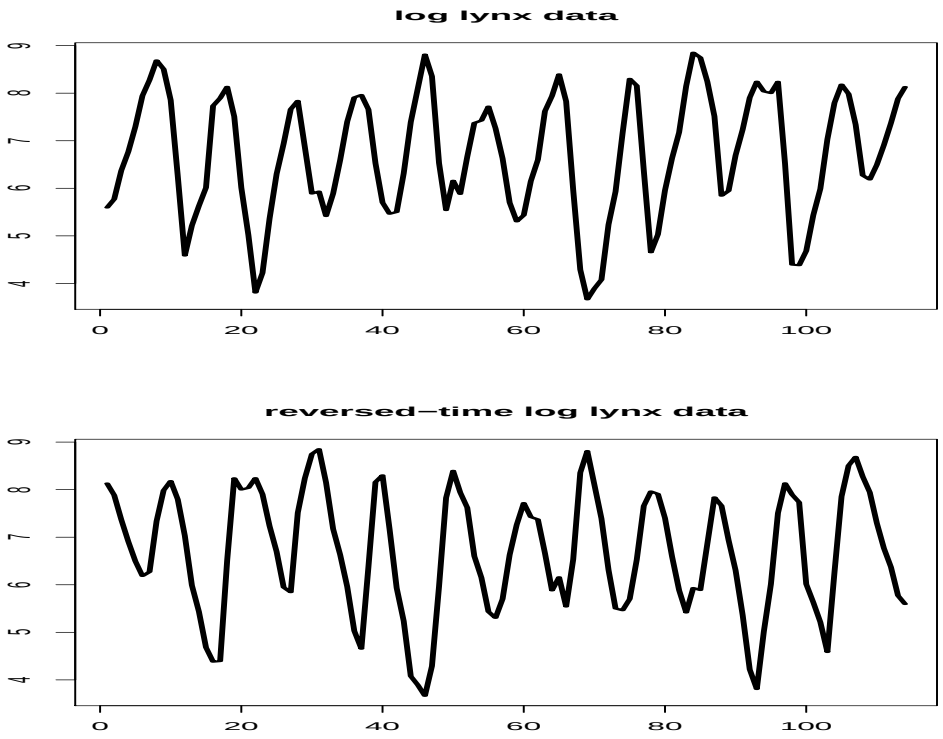
# Testing for TAR

If the likelihood for the linear AR model and the likelihood for a homoskedastic TAR model are available, the $F$–type test statistic

$$S_T = \{T - \max(p, d)\}(\hat{\sigma}^2 - \hat{\sigma}_0^2)/\hat{\sigma}^2,$$

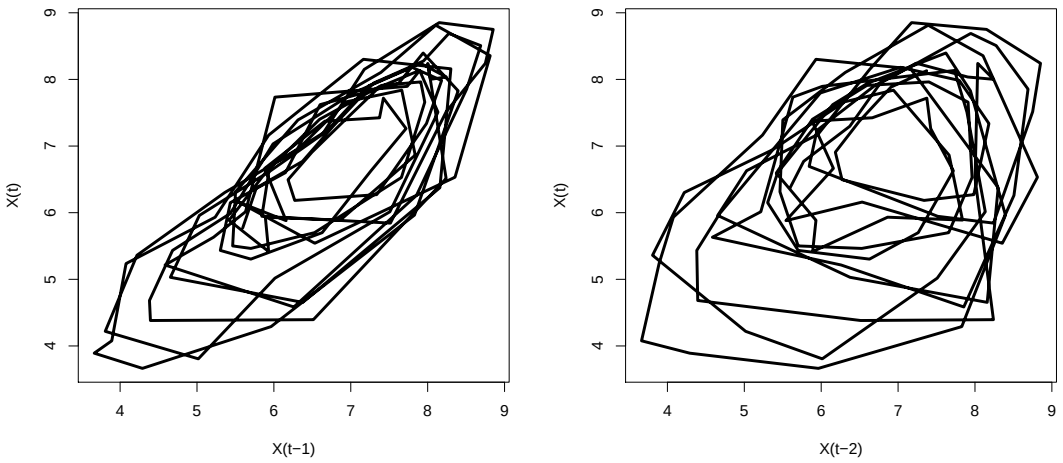
with $\hat{\sigma}_0^2$ estimated under the linear null, can be used for a significance test. The asymptotic distribution is nonstandard and was tabulated by Chan. Alternatively, one may bootstrap it to fit the observed sample under the null.

# An example: the famous lynx data




**Impression**: The asymmetry of the cycle (time irreversibility) raises doubts on a linear, Gaussian model.

# Plots versus lags for the lynx data




**Impression**: The hole in the center is in conflict with a linear, Gaussian model.

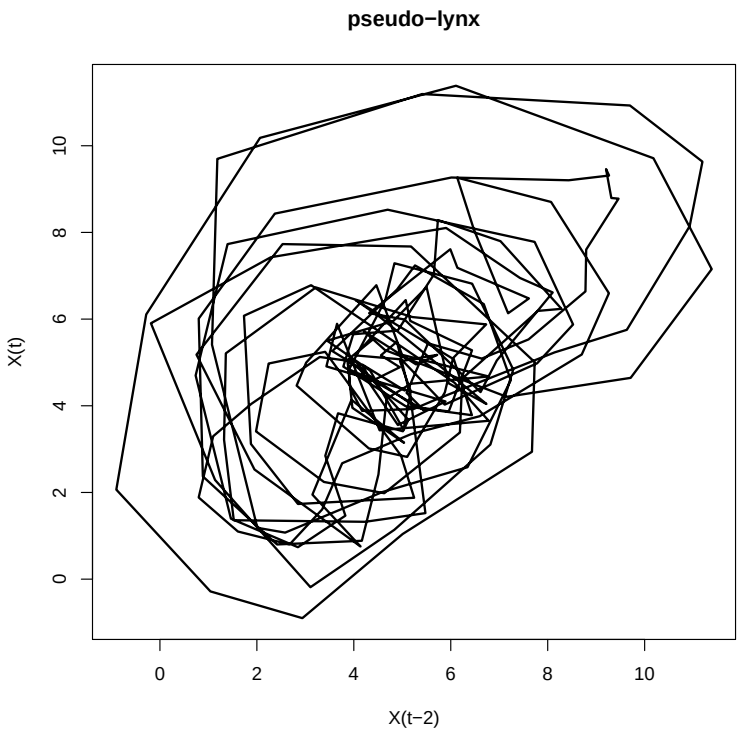
# A TAR model for the lynx data

The TAR test rejects its null. Some trial and error and optimization led HOWELL TONG to consider the TAR model

$$X_t = \begin{cases} 0.62 + 1.25X_{t-1} - 0.43X_{t-2} + \varepsilon_t, & X_{t-2} \leq 3.25, \\ 2.25 + 1.52X_{t-1} - 1.24X_{t-2} + \varepsilon_t, & X_{t-2} > 3.25, \end{cases}$$

with delay 2 years, for the regimes of few predators (and much prey) and of too many predators. It can be shown that this means that lynx starve fast and proliferate slowly: the asymmetry of the cycles.

# Plot against lag 2 for the Tong lynx model




This plot was generated with a burn-in of 300 observations and a sample of 200, using normal errors.

**Impression**: Nice cycles, but the empty center visible in the data is not reproduced. More complex models may be needed.

# III. ARCH and GARCH models

ARCH models were introduced by Robert F. Engle in 1982 to model time-changing volatility (variance) in a time-homogeneous model. The model, at first introduced for monthly inflation, proved to be extremely successful for daily finance data. The model is rarely used outside economics.

# Definition of the ARCH model

## Definition

An *autoregressive conditionally heteroskedastic* (ARCH) model with integer order $p \geq 1$ is defined by

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = c_0 + b_1 X_{t-1}^2 + \ldots + b_p X_{t-p}^2,$$

with $c_0 > 0$, $b_j \geq 0$, $(\varepsilon_t)$ *iid*(0,1), and $\varepsilon_t$ independent of $X_s, s < t$. A process $(X_t)$ obeying these equations is then an ARCH($p$) process.

**Note**. There is a mean and a variance equation. The variance equation does not have any additional error and exactly determines the unobserved local variance $\sigma_t^2$.

# Stability of the ARCH model

**Theorem**

*The ARCH model has a strictly and covariance stationary solution iff* $\sum_{j=1}^{p} b_j < 1$. *Then,* $\mathrm{E}X_t = 0$ *and*

$$\mathrm{E}X_t^2 = \frac{c_0}{1 - \sum_{j=1}^{p} b_j}.$$

**Note**. This condition is not necessary for a strictly stationary solution with infinite variance. For Gaussian $\varepsilon_t$, the sum may even be as large as 3.

# ARCH processes with finite kurtosis

## Theorem

*Assume $(X_t)$ is ARCH(p) with $\mathrm{E}\varepsilon_t^4 < \infty$ and*

$$\sum_{j=1}^{p} b_j < \frac{1}{\sqrt{\mathrm{E}\varepsilon_t^4}}.$$

*Then, the strictly stationary solution for $(X_t)$ has finite fourth moment.*

**Remark**. The upper bound is fairly restrictive. For Gaussian $\varepsilon_t$, it yields $1/\sqrt{3} = 0.577$. Even then, ARCH processes have considerable leptokurtosis, which may correspond well to financial time series data.

# The AR representation for ARCH squares

It is easy to derive that, for an ARCH($p$) process $(X_t)$, $X_t^2$ can be written as

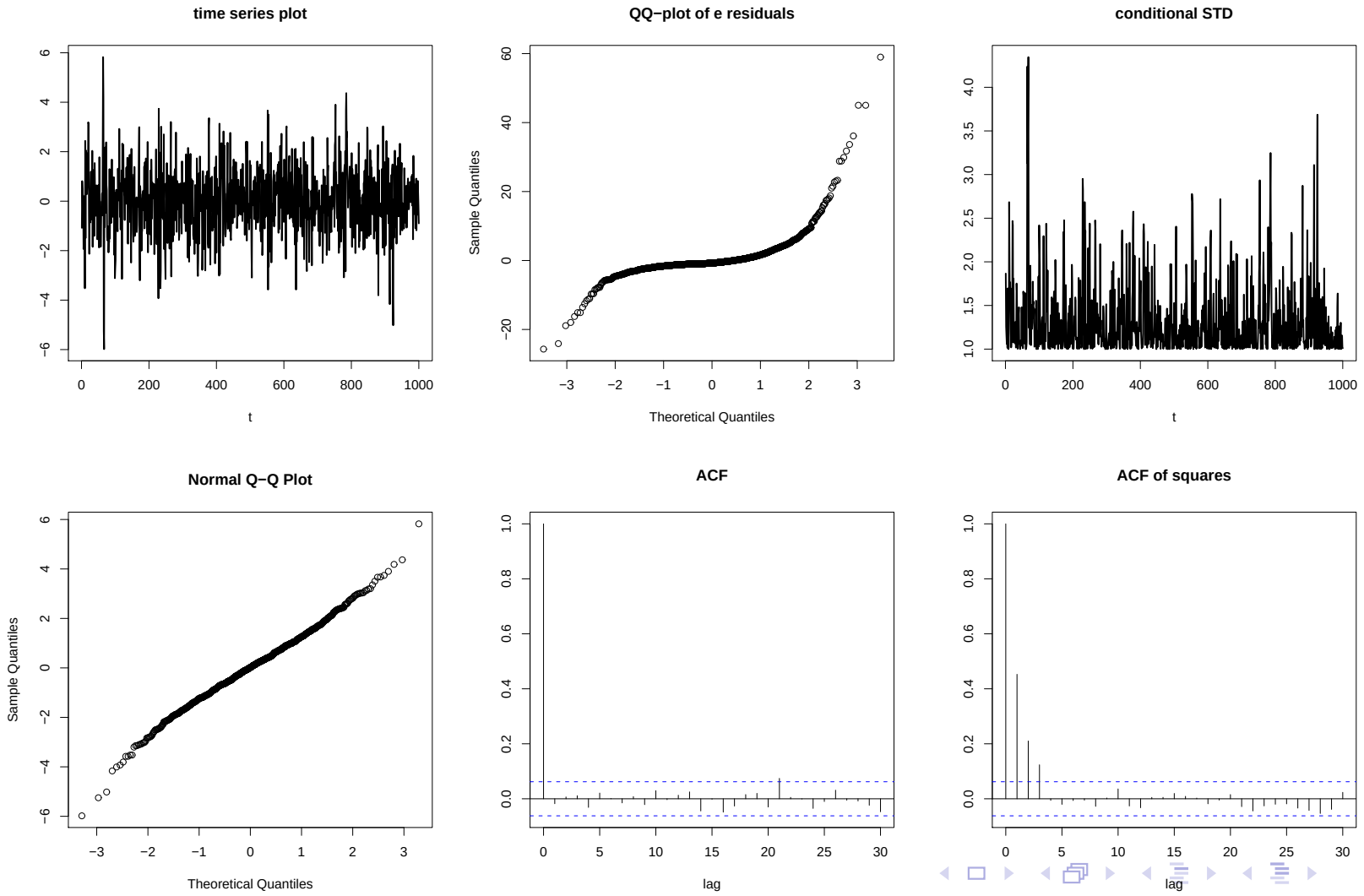
$$X_t^2 = c_0 + b_1 X_{t-1}^2 + \ldots + b_p X_{t-p}^2 + e_t,$$

with $(e_t)$ white noise and MDS. However, the errors

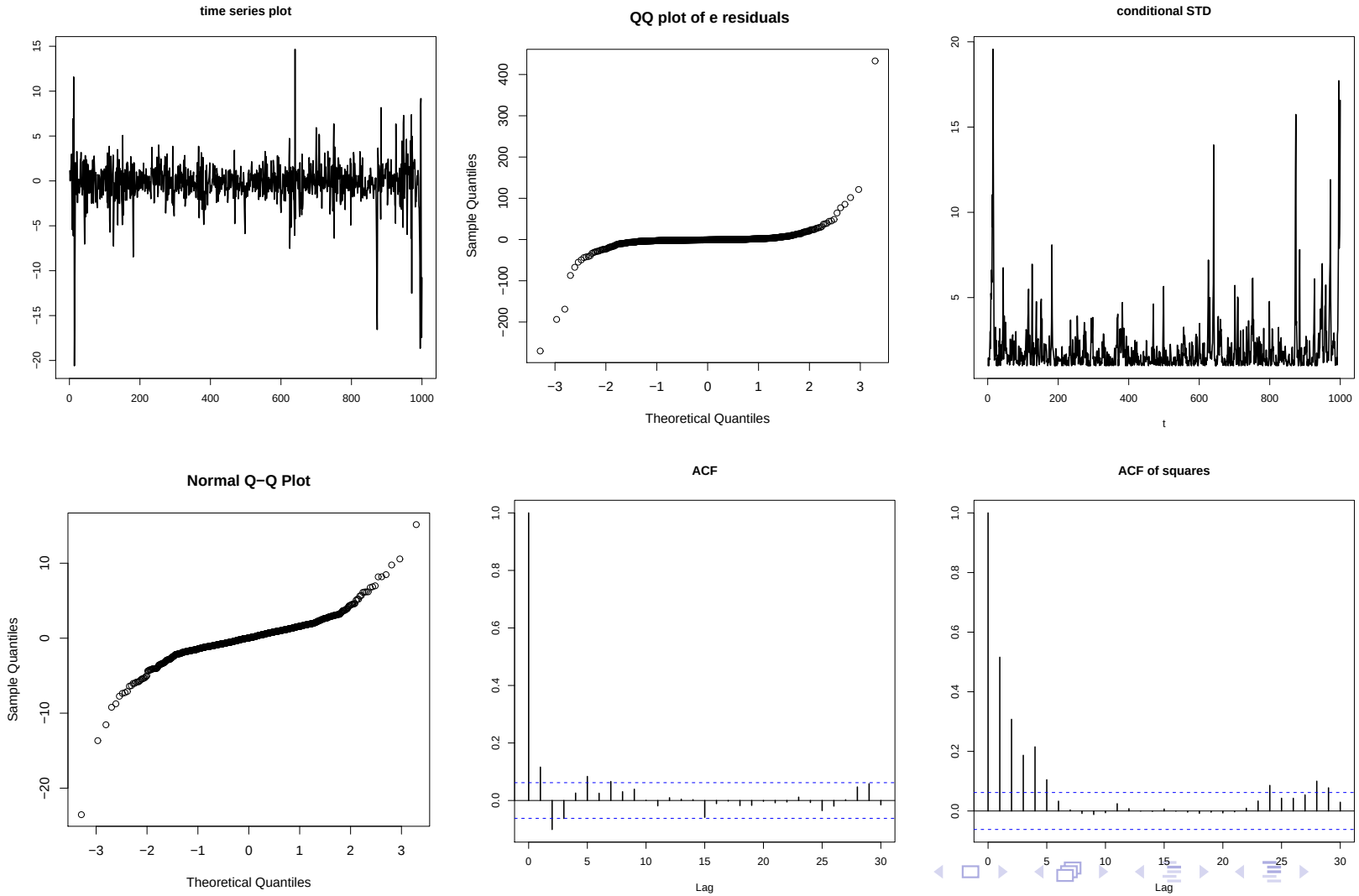
$$e_t = (\varepsilon_t^2 - 1)\{c_0 + \sum_{j=1}^{p} b_j X_{t-j}^2\}$$

are not *iid*. For Gaussian $\varepsilon_t$, $e_t$ has a non-normal skew distribution. Least-squares estimation of the AR model is inefficient.

# Generated ARCH(1) with $b_1 = 0.5$



# Generated ARCH(1) with $b_1 = 0.9$



# Observations regarding the plots

- These simple ARCH processes are white noise. The correlogram shows values close to 0.
- If $b_1 = 0.9$, fourth moments are infinite, and the convergence of the correlogram is problematic.
- A similar remark holds for the correlogram of squares. It should reflect the AR representation of the ARCH process but second moments of squares do not exist if $b_1 = 0.9$.
- The correlogram of squares shows the feature of *volatility clustering*, and so does the time plot of $\sigma_t$.
- The QQ plots show the considerable *leptokurtosis*. For $b_1 = 0.9$, kurtosis is undefined ('infinite').

# Properties of ARCH processes

## Proposition

*Assume $(X_t)$ is a strictly stationary ARCH(p) process with $c_0 > 0$ and $\sum b_j < 1$. Then,*

1. *$(X_t)$ is white noise with variance $c_0/(1 - \sum b_j)$;*
2. *if fourth moments are finite, $(X_t^2)$ is a (causal) AR(p) process with non-negative ACF;*
3. *if fourth moments are finite, the kurtosis of $X_t$ exceeds the kurtosis of $\varepsilon_t$.*

Financial time series often have high kurtosis. Part of it can be modelled by the ARCH structure, part of it by a leptokurtic $\varepsilon_t$.

# GARCH processes

In 1986, Tim Bollerslev found that large ARCH orders are needed to fit observed financial time series. He suggested a more parsimonious representation with geometric decay of ARCH coefficients. These GARCH processes, even as GARCH(1,1) were very successful. Steve Taylor made the same discovery but overlooked that the correspondence to the ARMA model is not perfect and that his MACH models do not work.

# The definition of GARCH

## Definition

A generalized autoregressive conditional heteroskedastic (GARCH) model with orders $p \geq 1$ and $q \geq 0$ is defined by

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = c_0 + \sum_{j=1}^{p} b_j X_{t-j}^2 + \sum_{j=1}^{q} a_j \sigma_{t-j}^2,$$

with $c_0 > 0$, $b_j \geq 0$, $a_j \geq 0$, $(\varepsilon_t)$ $iid(0,1)$, and $(\varepsilon_t)$ independent of $X_s, s < t$. A thus defined stochastic process is called a GARCH$(p, q)$ process.

**Remark**. GARCH$(0, q)$ for $q > 0$ does not work, as it would define a non-stochastic linear difference equation for $\sigma_t^2$. Some coefficients might be permitted to be negative, but conditions are restrictive for this case.

# Stability of the GARCH model

**Theorem**

*The GARCH$(p,q)$ model has a unique strictly and covariance stationary solution iff*

$$\sum_{j=1}^{p} b_j + \sum_{j=1}^{q} a_j < 1.$$

*Then,* $\mathrm{E}X_t = 0$, $(X_t)$ *is white noise, and*

$$\mathrm{var}X_t = \frac{c_0}{1 - \sum_{j=1}^{p} b_j - \sum_{j=1}^{q} a_j}.$$

$\mathrm{E}X_t^4 < \infty$ *if* $\mathrm{E}\varepsilon_t^4 < \infty$ *and*

$$\sqrt{\mathrm{E}\varepsilon_t^4}\frac{\sum_{j=1}^{p} b_j}{1 - \sum_{j=1}^{q} a_j} < 1.$$

# The ARMA representation of the GARCH model

## Proposition

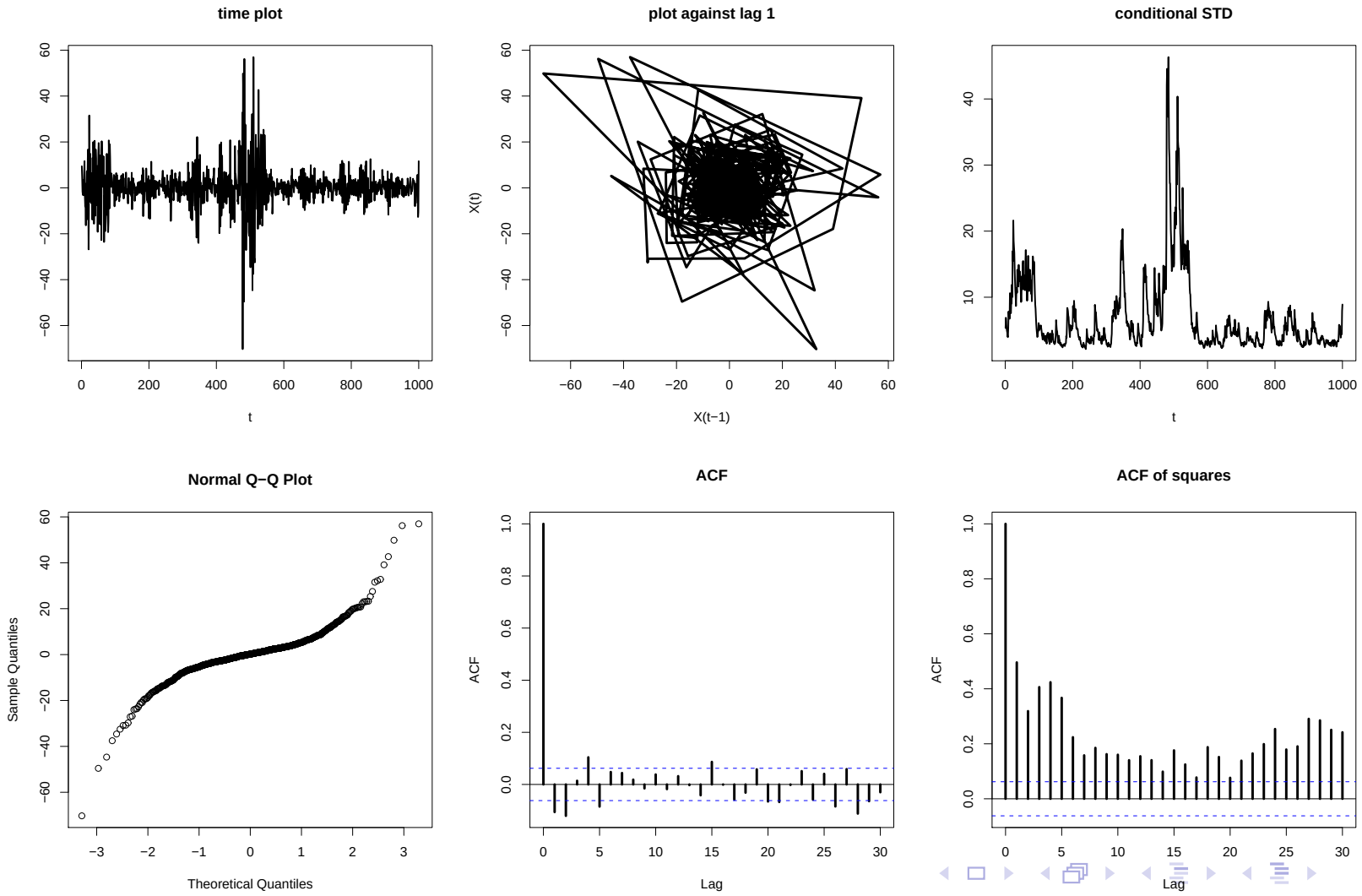
*If $(X_t)$ is a strictly and covariance stationary GARCH$(p,q)$ process with finite fourth moments, $(X_t^2)$ will be a causal and invertible ARMA$(\max(p,q),q)$ process. Its kurtosis exceeds the kurtosis of $\varepsilon_t$.*

In detail,

$$X_t^2 = c_0 + \sum_{j=1}^{\max(p,q)} (b_j + a_j) X_{t-j}^2 + e_t - \sum_{j=1}^{q} a_j e_{t-j}$$

for $e_t = X_t^2 - \sigma_t^2$.

# Generated GARCH(1,1) with $b_1 = 0.3$ and $a_1 = 0.7$



# Comments on the IGARCH graphs

1. The model does not fulfill the stability conditions, nonetheless the model with $a_1 + b_1 = 1$ has a strictly stationary solution;
2. in analogy to integrated processes, such processes are called IGARCH (integrated GARCH) but note that the correspondence is not perfect, as integrated processes are non-stationary;
3. volatility clustering is more conspicuous for GARCH processes than for pure ARCH.

# Estimation of GARCH models

1. Least squares (for ARCH only, considered by ENGLE, 1982);
2. Conditional maximum likelihood;
3. other estimators (Whittle, ML, robust methods).

# Least-squares regression of ARCH models

ENGLE was interested in the regression with ARCH errors

$$\begin{aligned} X_t &= \mu + \sum_{j=1}^{P} \phi_j X_{t-j} + u_t, \\ u_t &= \sigma_t \varepsilon_t, \quad \sigma_t^2 = c_0 + b_1 u_{t-1}^2 + \ldots + b_p u_{t-p}^2. \end{aligned}$$

He considered estimating the mean equation by OLS (consistent but inefficient), then the variance equation for squared OLS residuals by OLS (inefficient), then the mean equation by weighted LS, etc.

The procedure is unreliable and mainly used for obtaining starting values for ML–based estimation.

# Conditional ML for ARCH models

It is relatively straight forward to show that the log-likelihood $\ell(X_1, \ldots, X_T | c_0, b_1, \ldots, b_p)$ can be represented as

$$\begin{aligned} \ell(X_{p+1}, \ldots, X_T | X_1, \ldots, X_p, c_0, b_1, \ldots, b_p) &\propto -\sum_{t=p+1}^{T} (\log \sigma_t^2 + \frac{X_t^2}{\sigma_t^2}), \\ \sigma_t^2 &= c_0 + \sum_{j=1}^{p} b_j X_{t-j}^2. \end{aligned}$$

Most estimation algorithms maximize this likelihood by brute force, possibly restricting the area of admissible parameter values (non-negativity, stability constraints).

# Properties of the GARCH estimator

**Theorem**

*Assume* $\mathrm{E}\varepsilon_t^4 < \infty$, *denote* $\theta = (c_0, b_1, \ldots, b_p, a_1, \ldots, a_q)'$ *and* $\hat{\theta}$ *for the corresponding CML estimator. Then,*

$$\frac{\sqrt{T}}{\sqrt{\mathrm{E}(\varepsilon_t^4) - 1}}(\hat{\theta} - \theta) \Rightarrow N(0, \mathbf{M}^{-1}),$$

*with* $\mathbf{M}$ *a specific matrix essentially containing an ACF of squares.*

If the fourth-moment condition is violated, $T$–consistency may hold under some additional conditions.

# Testing for ARCH

ARCH testing is more common than estimating ARCH models. The ARCH test serves as a routine specification check.

ENGLE suggested the $TR^2$ approximation of the LM test. This statistic is asymptotically $\chi^2_p$ distributed under the null of no ARCH.

The $TR^2$ statistic is simply calculated by regressing $X_t^2$ on $X_{t-1}^2, \ldots, X_{t-p}^2$ and keeping the $R^2$, preferably using $T - p$ in lieu of $T$.

# Testing for ARCH with a nontrivial mean equation

The ARCH-LM test can easily be implemented if the mean equation looks like

$$X_t = \mu + u_t,$$

where mean-corrected $X_t$ is used in the auxiliary regression, or even with the AR mean equation

$$X_t = \mu + \sum_{j=1}^{P} \Phi_j X_{t-j} + u_t,$$

where the residuals $u_t$ can be used. In any case, the asymptotic chi-square distribution for $TR^2$ will hold.

# Other ARCH models

Many ARCH variants are popular in empirical finance. For example, the exponential GARCH (EGARCH) model by Daniel Nelson

$$\begin{aligned} X_t &= \varepsilon_t \exp(h_t/2), \\ h_t &= \gamma_0 + \gamma_1 h_{t-1} + \omega \varepsilon_{t-1} + \lambda(|\varepsilon_{t-1}| - \mathrm{E}|\varepsilon_{t-1}|), \end{aligned}$$

or the ARCH-in-mean (ARCH-M) model

$$X_t = \theta_0 + \theta_1 \sigma_t^2 + \varepsilon_t \sigma_t, \quad \sigma_t^2 = c_0 + b(X_{t-1} - \theta_0 - \theta_1 \sigma_{t-1}^2)^2.$$

# IV. Bilinear models

Bilinear models are maybe the most natural generalization of linear time-series models. They were introduced by Clive Granger and A.P. Andersen (1978) in a monograph and have been used in various sciences to model time series with rare outbursts, not so much in economics.

# Definition of the bilinear model

## Definition

Assume $(\varepsilon_t)$ is $iid(0, \sigma^2)$, then

$$X_t = \sum_{j=1}^{p} b_j X_{t-j} + \varepsilon_t + \sum_{k=1}^{q} a_k \varepsilon_{t-k} + \sum_{j=1}^{P} \sum_{k=1}^{Q} c_{jk} X_{t-j} \varepsilon_{t-k}$$

defines the bilinear process $(X_t)$ with order $(p, q, P, Q)$.

# The simplest bilinear model

The BL(1,0,1,1) model

$$X_t = bX_{t-1} + \varepsilon_t + cX_{t-1}\varepsilon_{t-1},$$

with $(\varepsilon_t)$ $iid(0, \sigma^2)$ and $\mathrm{E}\varepsilon_t^4 < \infty$ can be shown to have a strictly stationary solution with $\mathrm{E}X_t^2 < \infty$ for $b^2 + c^2\sigma^2 < 1$. Then, the closed form

$$X_t = \varepsilon_t + \sum_{j=1}^{\infty} \left\{ \prod_{k=1}^{j} (b + c\varepsilon_{t-k}) \right\} \varepsilon_{t-j}$$

is derived by continuous substitution.

# The mean of the BL(1,0,1,1) process

Direct evaluation of the closed form yields

$$\mathrm{E}X_t = \sum_{j=1}^{\infty} b^{j-1} c \mathrm{E}(\varepsilon_{t-j}^2) = \frac{\sigma^2 c}{1-b}.$$

The parameters $b, c, \sigma^2$ all increase the mean. There is also a closed but complicated expression for the variance.

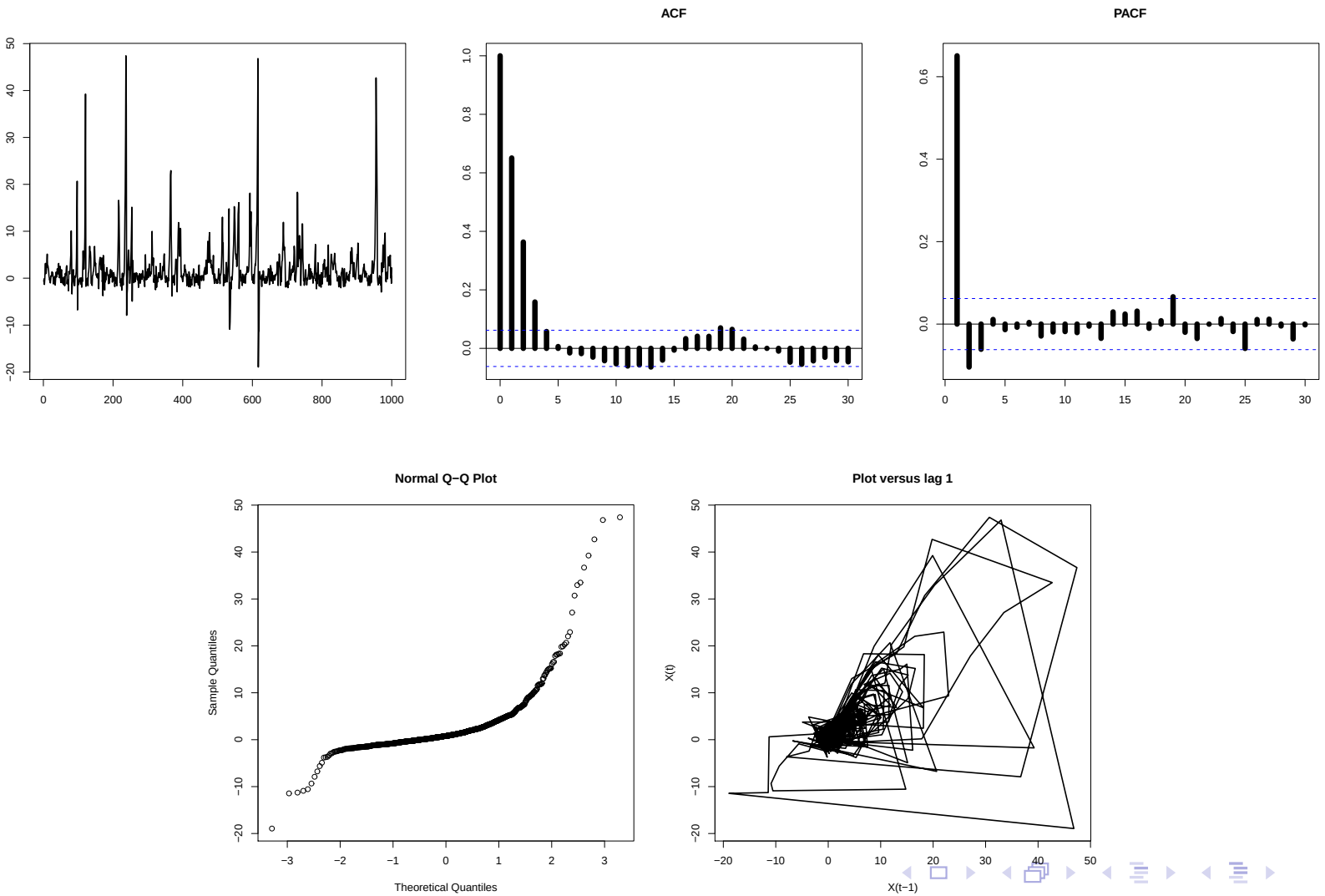
# The ARMA representation of the BL(1,0,1,1) model

Denoting $\mathrm{E}X_t$ by $\mu_x$, it is straight forward to derive

$$X_t - \mu_x = b(X_{t-1} - \mu_x) + \varepsilon_t + c(X_{t-1}\varepsilon_{t-1} - \sigma^2),$$

such that it is easily shown that the ACF is identical to an ARMA(1,1) process. Again, there is a correct Wold-type representation but it pays to model the nonlinear remainder.

# Generated BL(1,0,1,1) with $b = 0.75$ and $c = 0.6$



# Stability of the BL(1,0,1,1) model

The model can be re-written as

$$X_t = (b + c\varepsilon_{t-1})X_{t-1} + \varepsilon_t,$$

i.e. as a *random-coefficient* model $X_t = A_t X_{t-1} + \varepsilon_t$. Iterated substitution yields

$$X_t = (\prod_{j=0}^{k-1} A_{t-j})X_{t-k} + \sum_{j=1}^{k-1}(\prod_{l=0}^{j-1} A_{t-l})\varepsilon_{t-j} + \varepsilon_t.$$

Convergence of the product term to 0 is necessary for stability but not sufficient. A sufficient condition is that the 'upper Lyapunov exponent' of the sequence $A_t$ is negative, a condition that involves $b, c$ as well as the distribution of $\varepsilon_t$.

In principle, the stability of higher-order BL models can be evaluated similarly.