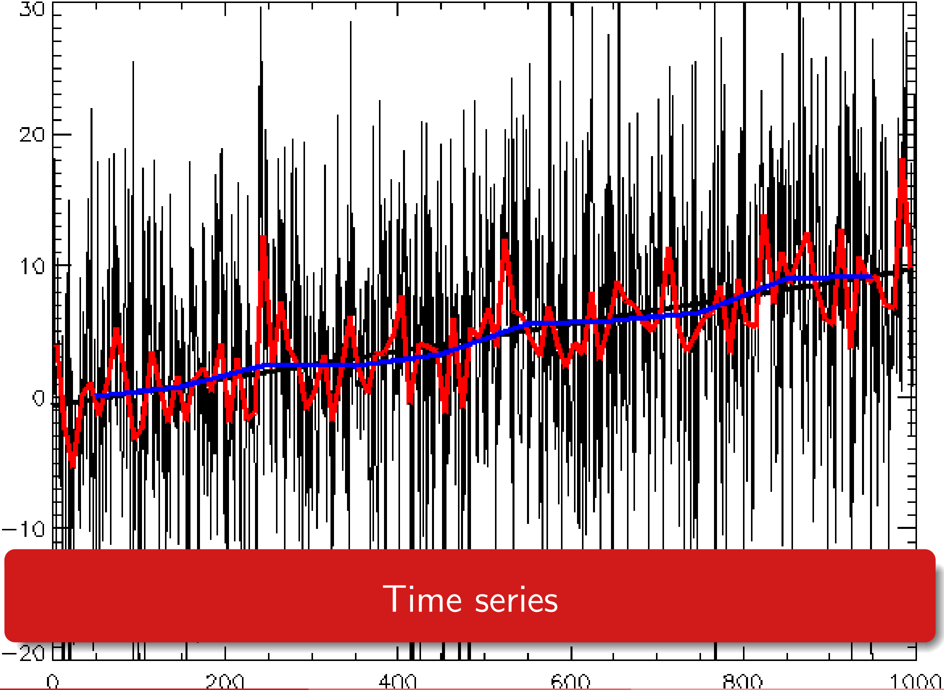


Time series

Preliminary definitions

Time series definition [1]

Informal definition

A time-series is a set of observation x_t each one being recorder at a specific time t.

Formal definition

A time series model for the observed data x_t is a specification of the joint distribution (or possible only the means covariance) of a sequence of random variable X_t of which x_t is postulated to be a realization

A binary process

Consider the sequence of iid random variables, with $P[X_t = 1] = p$ and $P[X_t = -1] = 1 - p$

Random walk

The random walk is obtained by cumulatively summing iid random variables. Thus a random walk with zero mean is obtained by defining

Stationarity, autocovariance and autocorrelation[1]

Mean Function

Let X_t be a time series with $E(x_t^2 < \infty$ The mean function of X_t is $\mu_X(t) = E(X_t)$. The covariance function of X_t is $\gamma_X(r,s) = Cov(X_r, X_S) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))] \quad \forall r, s$

Weakly stationary TS

X_t is weakly stationary if i) $\mu_X(t)$ is independent from time t and ii) $\gamma_X(t+h)$ is indipendent of t $\forall h$

Autocovariance function

At lag h the auto-covariance function is defined as $\gamma_X(h) = Cov(X_{t+h}, X_t)$

Autocorrelation function

At lag h the autocorrelation function is defined as

$$\rho(h)_X = \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_{t+h}, X_t)$$

Stationarity, autocovariance and autocorrelation [1]

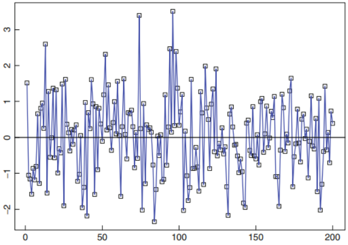


Figure 1-12
200 simulated values of IID N(0,1) noise

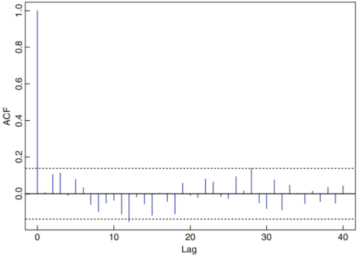
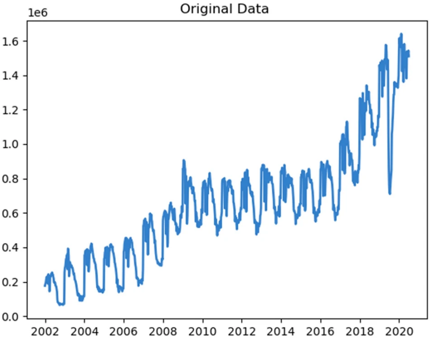
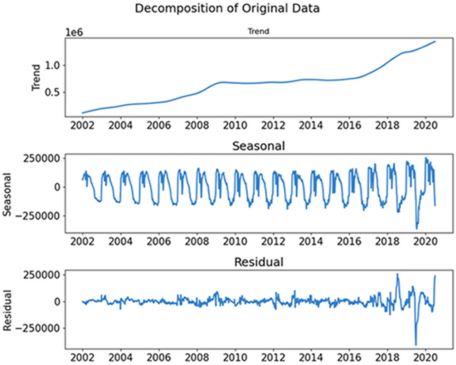


Figure 1-13
The sample autocorrelation function for the data of Figure 1-12 showing the bounds $\pm 1.96/\sqrt{n}$

Stationarity, autocovariance and autocorrelation [2]



(a)



(b)

Definition [1]

Linear process

Time series X_t is a linear process if it has the representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

for all t, where $Z_t \approx \mathrm{WN}(0, \sigma^2)$ and ψ_j is a sequence of costants with $\sum_{j=-\infty}^{\infty} \psi_j < \infty$. In terms of backward shift operator B $BX_t = X_{t-1}$ we have $X_t = \phi(B)Z_t$ Therefore the previous definition can be rewritten as $X_t = \phi(B)Z_t$ in which $\phi(B)$ can be thought as a linear filter that when applied to the white noise input series Z_t produces the output X_t

Definition [1]

Linear process

The time series X_t is an ARMA(1,1) process if its stationary and satisfiy (for every t)

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$$

where $Z_t \approx \mathrm{WN}(0, \phi^2)$ and $\phi + \theta \neq 0$ or in terms of filters ϕ and θ

$$\phi(B) X_t = \theta(B) Z_t$$

Causality [1]

Remarks

- A stationary solution of the ARMA(1,1) equation exists if and only if $\phi \neq \pm 1$
- If $|\phi| < 1$, then the unique stationary solution is given by $X_t = Z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}$. In this case we say that X_t is causal or a causal function of Z_t or a causal function of Z_t since X_t can be expressed in terms of the current and past values Z_s, $s \leq t$
- If $|\phi| > 1$, then the unique stationary solution is given by $X_t = -\theta\phi^{-1} Z_t - (\phi + \theta) \sum_{j=1}^{\infty} \phi^{-j-1} Z_{t-j}$. The solution is non-causal, since X_t is then a functional of Z_s $s \geq t$

Wold decomposition [1]

Prediction operator based on the infinite past $X_t, -\infty < t < n$

$\tilde{P}_n X_{n+h} = \lim_{m \to -\infty} P_{m,n} X_{n+h}$

Wold decomposition $X_t, -\infty < t < n$

X_t is a non-deterministic stationary time series, then

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t$$

where V_t is deterministic

Definition [1]

ARMA (p,q)

X_t is an ARMA(p,q) process if X_t is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - ... - \phi_1 X_{t-q} - = Z_t + \theta Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$$

where $Z_t \approx WN(0, \sigma^2)$ and the polynomials $(1 - \phi_1 z - ... - \phi_p z^p)$ and $(1 + \phi_1 z + ... + \phi_p z^p)$ have no common factor.

Definition [1]

Causality

An ARMA(p,q) process X_t is causal, or a causal function of Z_t if there exist constants ϕ_j such that $\sum_{j=0}^{\infty} |\phi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \sum \phi_j Z_{t-j} \quad \forall \quad t$$

this is equivalent to the condition

$$\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p \neq \quad |z| < 1$$

Spectral Analysis

Spectral Analysis [1]

Spectral density

Given a zero mean stationary time series X_t with autocovariance function $\gamma()$ satfying $\sum_{h=-\infty} |\gamma(h)| < \infty$, the spectral density of X_t is the function defined by

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

with the condition that

$$f(\lambda) \geq 0 \quad \forall \lambda$$

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda \forall h$$

Spectral Analysis [1]

Spectral Representation of The ACVF

A function $\gamma()$ defined on the integers is the ACVF of a stationary time series if and only if there exists a right-continuous, non-decresing, bounded function F on $[-\pi, \pi]$ with $F(-\pi) = 0$ such that:

$$\gamma(h) = \int_{(-\pi,\pi]} e^{ih\lambda} dF(\lambda)$$

for all integers h. F is a generalized distribution function that is called the spectral distribution function of $\gamma()$

Spectral Analysis [1]

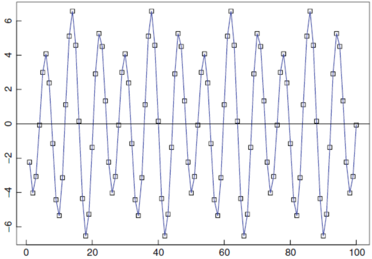


Figure 4-1
A sample path of size 100 from the time series in Example 4.1.2

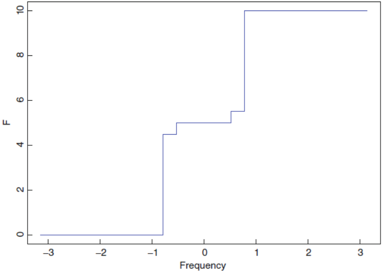


Figure 4-2
The spectral distribution function $F(\lambda)$, $-\pi \leq \lambda \leq \pi$, of the time series in Example 4.1.2

Time-Invariant Linear Filters [1]

Linear process

The process Y_t is the output of a linear filter $C = \{c_{t,k}, t, k = 0 \pm 1, ...\}$ applied to an input process X_t if $Y_t = \sum_{k=-\infty}^{\infty} c_{t,k} X_k \quad t = 0, \pm 1, ...$

Time invariant

The filter is said to be time-invariant if the weights $c_{t,t-k}$ are independent of t e.g. $c_{t,t-k} = \phi_k$

$$Y_t = \sum_{k=-\infty}^{\infty} \phi_{t,k} X_{t-k}$$

$$Y_{t-s} = \sum_{k=-\infty}^{\infty} \phi_{t,k} X_{t-s-k}$$

The TLF ϕ is to be causal if $\phi_j = 0$ for $j < 0$

Time-Invariant Linear Filters [1]

Transfer function

Let X_t be a stationary time series with mean zero and spectral density $f_x(\lambda)$. Suppose that $\Phi = \{\phi_j, j = 0, \pm 1, ...\}$ is an absolutely summable TLF. Then the time series

$$Y_t = \sum_{j=-\infty}^{\infty} \phi_j X_{t-j}$$

is stationary with mean zero and spectral density

$$f_Y(\lambda) = |\Psi(e^{-i\lambda})|^2 f_X(\lambda) = \Psi(e^{-i\lambda})\Psi(e^{i\lambda}) f_X(\lambda)$$

where $\Psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \phi_j e^{-ij\lambda}$ where $\Psi(e^{-i})$ is called the transfer function of the filter, and the squared modulus $|\Psi(e^{-i})|$ is referred to as the power transfer function of the filter.

Time-Invariant Linear Filters [1]

Simple moving average

$$Y_t = \frac{1}{2q+1} \sum_{|j|<q} X_{t-j}$$

where $\psi = (2q+1)^{-1}, j = -q, ..., q$ and ψ_j

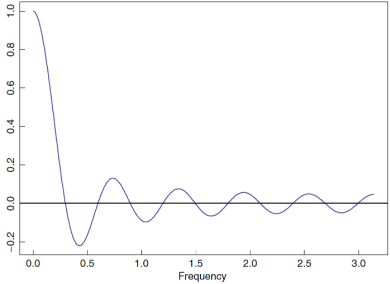


Figure 4-12
The transfer function $D_{10}(\lambda)$ for the simple moving-average filter

Time-Invariant Linear Filters [1]

Gibbs phenomenon

The poor approximation in the neighbourhood of cut-off frequency (ω_c)

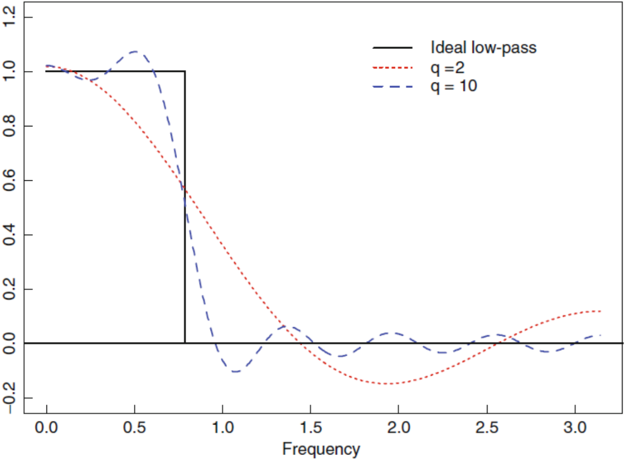


Figure 4-13
The transfer function for the ideal low-pass filter and truncated Fourier approximations $\Psi^{(q)}$ for $q = 2, 10$

Estimation

Esimation [1]

Gaussian Likehood

$$L(\hat{\boldsymbol{\phi}}, \hat{\theta}, \boldsymbol{\sigma}^2) = \frac{1}{\sqrt{(2\pi\sigma)^n r_0 \ldots r_{n-1}}} \exp(-\frac{1}{2\sigma^2} \sum \frac{(X_j - \hat{X}_j)^2}{r_{j-1}})$$

Maximum Likehood Estimators

Differentiating $\ln L(\phi, \theta\, \sigma^2)$ partially with respect to σ^2 and noting that $\hat{X}_j$ and r_j are independent of σ

$$\hat{\sigma}^2 = n^{-1} S(\hat{\boldsymbol{\phi}}, \hat{\theta})$$

$$S(\hat{\boldsymbol{\phi}}, \hat{\theta}) = \sum_{j=1}^{n} (X_j - \hat{X}_j)^2 / r_{j-1}$$

where $\hat{\boldsymbol{\phi}}$,$\hat{\theta}$ are the values of $\boldsymbol{\phi}$,θ that minimize:

$$\mathcal{L}(\phi, \theta) = \ln(n^{-1} S(\phi, \theta)) + n^{-1} \sum_{j=1}^{n} \ln r_{j-1}$$

Estimation [1]

Akaike information criterion

$$AICC = -2\ln L(\hat{\phi}, \hat{\theta}, S((\phi_p, \theta_q)/n) + 2(p+q+1)n/(n-p-q-2)$$

Residual

$$\hat{W}_t = \left(X_t - \hat{X}_t \phi, \theta\right) / \left(r_{t-1}(\hat{\phi}, \hat{\theta})\right)^{1/2} \quad t = 1, ..., n$$

This should have properties similar to the white noise sequence, thus one can define the rescaled residuals

$$\hat{R}_t = \hat{W}_t / \hat{\sigma}$$

$$\hat{\sigma} = \sqrt{\left(\sum_{t=1}^{n} \hat{W}_t^2\right) / n}$$

Residuals [1]

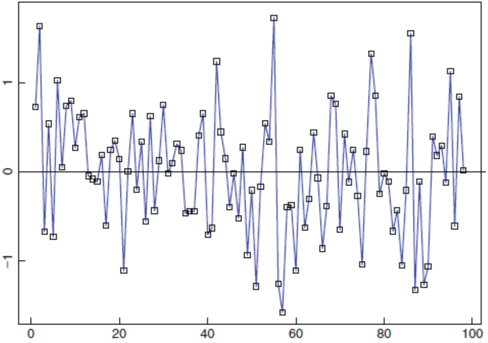
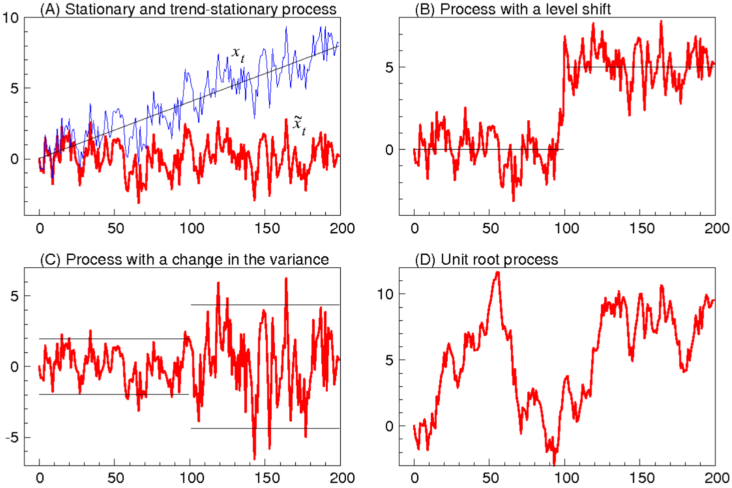


Figure 5-5
The rescaled residuals after fitting the ARMA(1,1) model of Example 5.2.5 to the lake data

Nonstationary time series

Residuals [**ur**, 1]



Estimation [1]

Akaike information criterion

If d is a non-negative integer, then X_t is an ARIMA(p,d,q) process if

$$Y_t = (1 - B)^d X_t$$

is causal ARMA(p,q) process $BX_t = X_{t-1}$. Note thaat $\nabla X_t = X_t - X_{t-1}$

Bibliography I

[1] Peter J Brockwell e Richard A Davis. *Introduction to time series and forecasting*. Springer, 2002.

[2] Jianhua Hao e Fangai Liu. "Improving long-term multivariate time series forecasting with a seasonal-trend decomposition-based 2-dimensional temporal convolution dense network". In: *Scientific Reports* 14.1 (2024), p. 1689.