Time Series Analysis

Selected Topics

Lecture Notes

Alois Pichler[1]

Summer 2023

Version as of May 16, 2023

[1]Technische Universität Chemnitz, Faculty of Mathematics
https://www.tu-chemnitz.de/mathematik/fima/

Preface and Acknowledgment

The purpose of these lecture notes is to facilitate the content of the lecture and the course. From experience it is helpful and recommended to attend and follow the lectures in presence. The lecture notes do not cover the lectures completely.

Initial literature on the subject includes Box et al. (2013). Brockwell and Davis (1987) properly describe the mathematics of time series.

Important references for this lecture include Brockwell and Davis (2002) and Shumway and Stoffer (2000). Härdle et al. (1997) and Fan and Yao (2003) discuss nonparametric time series. Time series for financial applications can be found in Andersen et al. (2009); Brooks (2014) and Franke et al. (2004). Some content (including problems) follows these references very closely.

Please report mistakes, errors, violations of copyright, improvements or necessary completions.

Further description of the course:
https://www.tu-chemnitz.de/mathematik/studium/module/2013/M22.pdf

Additional material: kick-starting time series in R by Salima Abdalla,
https://www.tu-chemnitz.de/mathematik/fima/public/ZeitreihenAbdalla.pdf

Contents

Bibliography

T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. V. Mikosch. *Handbook of Financial Time Series*. Springer, 2009. doi:10.1007/978-3-540-71297-8. 3

G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis*. John Wiley & Sons, Inc., 2013. doi:10.1002/9781118619193. 3, 9

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer New York, 2nd edition, 1987. doi:10.1007/978-1-4419-0320-4. URL `https://books.google.de/books?id=ZW_ThhYQiXIC`. 3, 53, 65

P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer Texts in Statistics, 3rd edition, 2002. doi:10.1007/b97391. 3

C. Brooks. *Introductory econometrics for finance*. Cambridge University Press, 2014. doi:10.1002/ijfe.223. 3

J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics, 2003. doi:10.1007/b97702. 3

J. Franke, W. Härdle, and C. Hafner. *Einführung in die Statistik der Finanzmärkte*. Springer Berlin Heidelberg, 2004. doi:10.1007/978-3-642-17049-2. 3

G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 4th edition, 2013. ISBN 1421407949. 27

W. Härdle, H. Lütkepohl, and R. Chen. A review of nonparametric time series analysis. *International Statistical Review*, 65(1):49–72, 1997. doi:10.2307/1403432. 3

R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications*. Springer Texts in Statistics. Springer, 2000. doi:10.1007/978-3-319-52452-8. URL `http://www.stat.pitt.edu/stoffer/tsa4/tsa4.htm`. 3, 52

Preliminaries, Notations, ... 1

The fundamental cause of the trouble is that in the modern world the stupid are cocksure while the intelligent are full of doubt.

Bertrand Russell, 1872–1970

1.1 NOTATION AND CONVENTION

Time series analysis is a subarea of (mathematical) statistics.

Definition 1.1. A stochastic process on a probability space $(\Omega, \mathcal{A}, P)$ is a family of random variables $(X_t)_{t\in T}$.

Typical index sets for time series include $T = \mathbb{N}$ and $T = \mathbb{Z}$.

By convention, the time-series $X = (X_t)_{t\in T}$ is a row vector (mainly, because C/ C++ and NumPy (Python) use row-major (lexicographical) order; Julia, Matlab and R are column-major).

1.2 BOX–JENKINS MODELING

The Box–Jenkin modeling approach is a three-step ((ii)–(iv) below) modeling approach (cf. Box et al. (2013)[1]):

(i) Data preparation
(ii) Model identification and model selection
(iii) Parameter estimation
(iv) Model checking
(v) Forecasting

The law of parsimony, aka. Occam's razor.[2]

Example 1.2 (Classical decomposition)**.** A typical result of the Box–Jenkins modeling is the decomposition (the classical decomposition)

$$X_t = \underbrace{m_t}_{\text{trend}} + \underbrace{k_t}_{\text{economic cycle}} + \underbrace{s_t}_{\text{season}} + \underbrace{f(u_t)}_{\text{nonlinear control}} + \underbrace{Z_t,}_{\text{residual, unexplained}}$$

[1]See also https://robjhyndman.com/papers/BoxJenkins.pdf for a nice overview.
[2]William of Ockham, 1287–1347

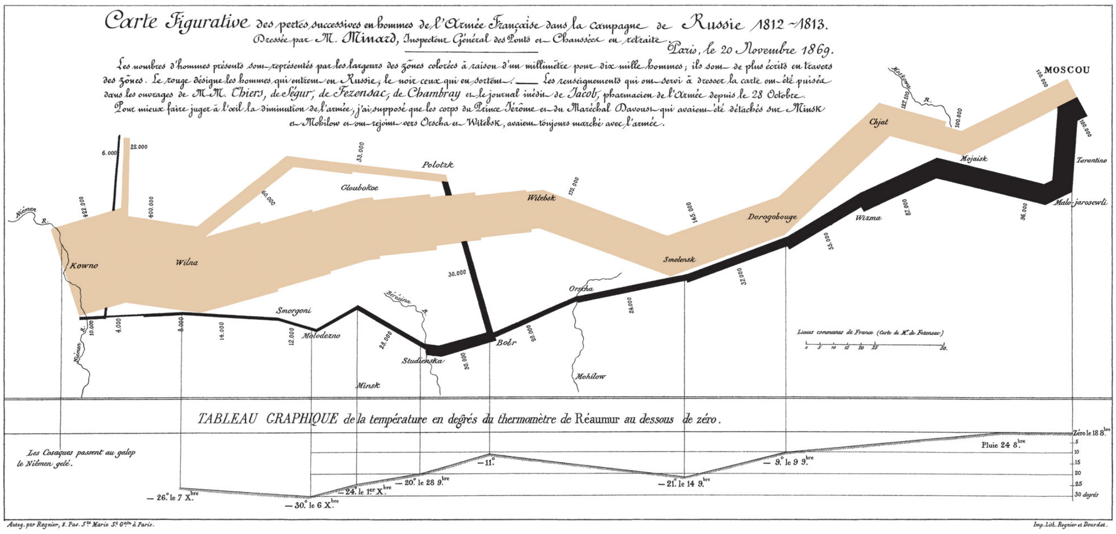


Figure 1.1: Charles Minard's map of Napoleon's Russian campaign of 1812, https://en.wikipedia.org/wiki/Charles_Joseph_Minard

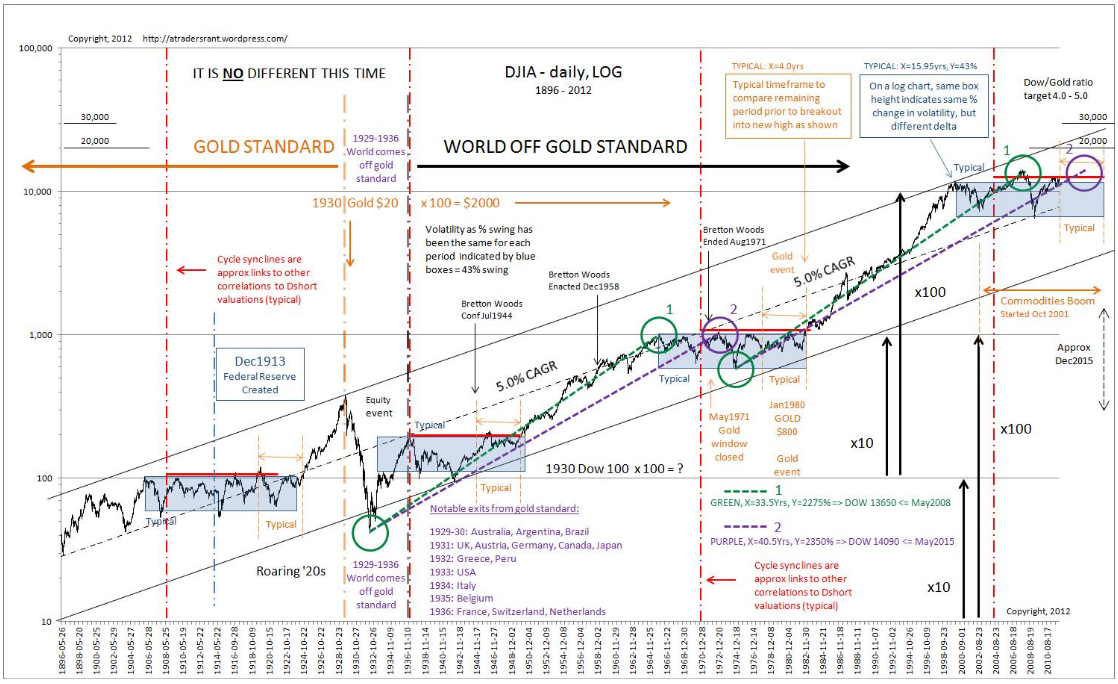


Figure 1.2: Dow Jones Insdustrial Average, historic chart. Source: http://allstarcharts.com/110-years-of-the-dow-jones-industrial-average-volatility-is-normal/

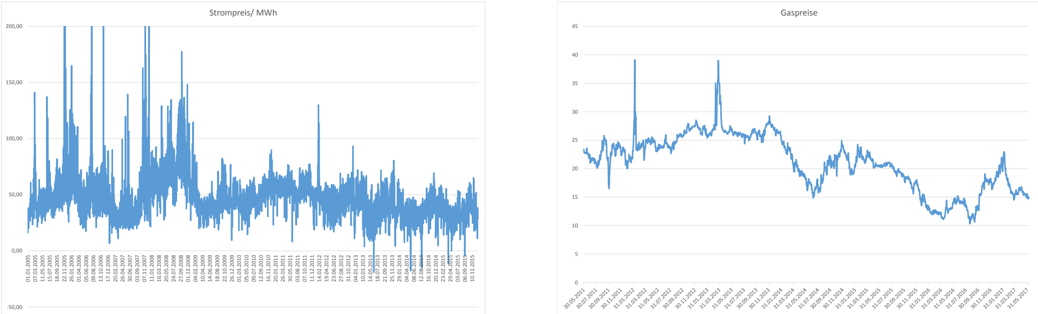


Figure 1.3: Prices for electricity and natural gas

where m_t is a trend component (k_t another, short-term trend, regime), s_t a seasonal component ($f(u_t)$ a control) and Z_t an unexplained error, or noise. For an example consider Figure 1.4a.

1.3 TIMESTAMP

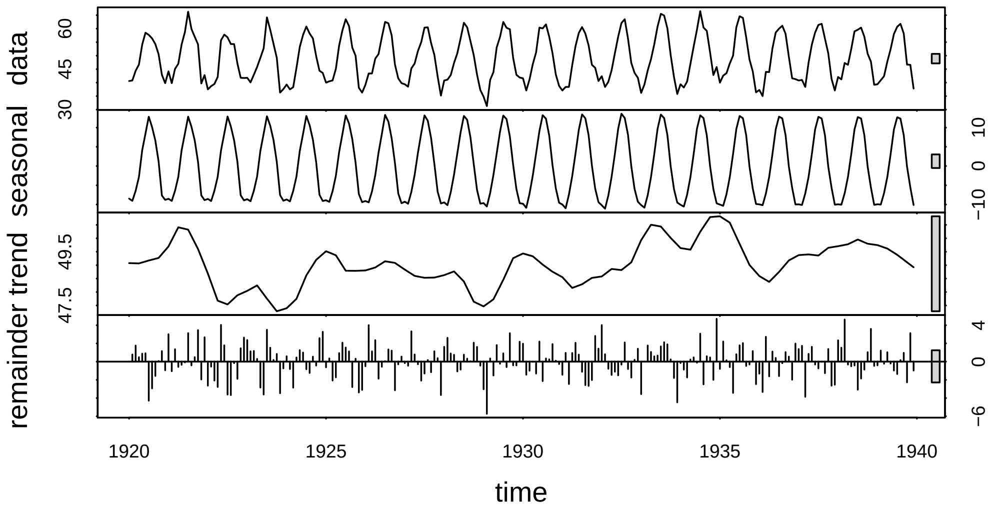
The timestamp is an index which can be identified with a float number. In Excel, e.g., Jan 1$^{\text{st}}$, 1900 = 1,00 or 44000,35 = June 18$^{\text{th}}$, 2020, 8:24. The Astronomers' time stamp is 2018-05-27 22:50:55.338162 + 02:00 = 2458266.3686960433, for example.

Python's datetime and panda's timestamp start with 1900 as well. Unix time is the number of seconds since Jan 1$^{\text{st}}$, 1970, 00:00 UTC, without leap seconds.

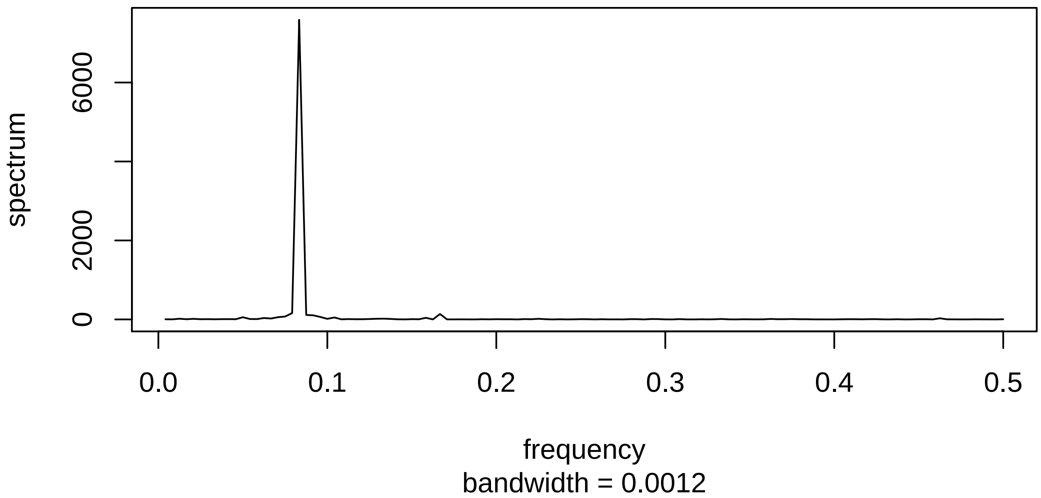
Note, that this approach allows algebra on dates. $t+1$ is the next instant of time day (day, say, or second, year) based on the implementation; $t_2 - t_1$ is the term between different dates, measured in base time units (seconds, in Unix, e.g.).

Of course, including the time stamp t_i in the time series X_{t_i} one can consider the new time series $(t_i, X_{t_i})_{i\in\mathbb{N}}$, indexed by $\mathbb{N}$, say.

As an example for a time series with non equidistant timestams see Figure 1.5.



(a) Decomposition



(b) The periodogram of the nottem data exposes the frequency $f = \frac{1}{12} \approx 0.083$

Figure 1.4: The *nottem* data from R

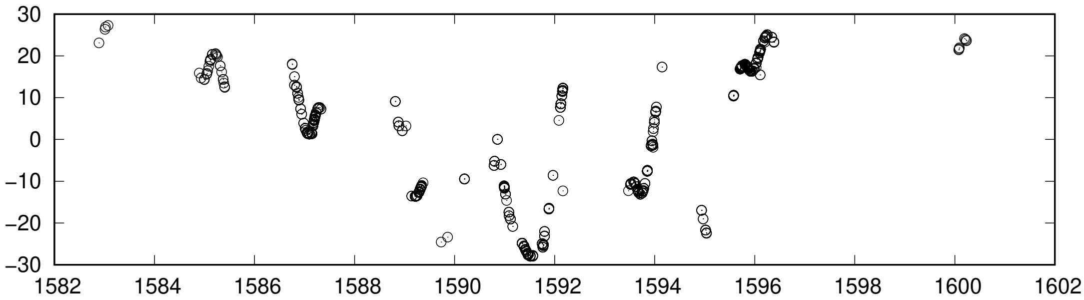


Figure 1.5: Declination of mars as measured by Tycho Brahe (1546–1601). Note the time stamp

2 The trend

All shall be well and all shall be well
and all manner of thing shall be well.

Julian of Norwich, 1342–1416

2.1 FILTERS

Filters are employed to increase the signal-to-noise ratio without greatly distorting the signal.

Definition 2.1. A *Filter* is a map, mapping a time series to another time series

$$(X_t)_{t\in\mathbb{Z}} \mapsto (m_t)_{t\in\mathbb{Z}}\,.$$

A general *linear filter* has the form

$$m_t = \sum_{j\in\mathbb{Z}} a_j X_{t+j}. \tag{2.1}$$

In what follows we discuss *low-pass* filters, aka. high-cut filter: a low-pass filter is a filter that passes signals with a frequency lower than a certain cutoff frequency and attenuates signals with frequencies higher than the cutoff frequency.

Note, that we may rewrite (2.1) formally as matrix product, $m = AX$, or

$$\begin{pmatrix} \vdots \\ m_{-1} \\ m_0 \\ m_1 \\ \vdots \end{pmatrix} = \begin{pmatrix} \ddots & \ddots & \ddots & & \\ \ddots & a_0 & a_1 & \ddots & \\ \ddots & a_{-1} & a_0 & a_1 & \ddots \\ & \ddots & a_{-1} & a_0 & \ddots \\ & & \ddots & \ddots & \ddots \end{pmatrix} \cdot \begin{pmatrix} \vdots \\ X_{-1} \\ X_0 \\ X_1 \\ \vdots \end{pmatrix}$$

on $\mathbb{R}^{\mathbb{Z}}$.

2.2 THE LEAST SQUARES FILTER

Cf. linear models in math. statistics,
https://www.tu-chemnitz.de/mathematik/fima/public/mathematischeStatistik.pdf.

2.3 POLYNOMIAL FITTING—SAVITZKY–GOLAY FILTER

The data points X_t are observed at $t+z$ with $z \in \{z_i : i = 1, \dots, m\} \subset \mathbb{Z}$ and approximated/ fitted with a function

$$m_\beta(z) = \beta_1 \cdot g_1(z) + \beta_2 \cdot g_2(z) + \dots + \beta_k \cdot g_k(z) = g(z)^\top \beta \tag{2.2}$$

with $g(z) = \big(g_0(z), \dots, g_k(z)\big)^\top$. For $g_j(z) = z^{j-1}$, the function m_β is a polynomial.

The coefficients $\beta = (\beta_1, \dots, \beta_k)$ are chosen to minimize

$$\sum_{i=1}^m w_i \left(X_{t+z_i} - \sum_{j=1}^k \beta_j\, g_j(z_i) \right)^2 = \sum_{i=1}^m w_i \big(X_{t+z_i} - m_\beta(z_i) \big)^2.$$

Set

$$G := \big(g_j(z_i)\big)_{i=1:m}^{j=0:k} \in \mathbb{R}^{(k+1)\times m}.$$

Differentiating with respect to β_ℓ , $\ell = 1, \dots, k$, gives the first order conditions

$$0 = \sum_{i=1}^m w_i 2 \left(X_{t+z_i} - \sum_{j=1}^k \beta_j\, g_j(z_i) \right) g_\ell(z_i) = 2 \sum_{i=1}^m g_\ell(z_i) w_i X_{t+z_i} - 2 \sum_{i=1}^m g_\ell(z_i) w_i \sum_{j=0}^k g_j(z_i)\, \beta_j \Bigg),$$

i.e., the normal equations

$$G^\top W X = G^\top W G \beta \tag{2.3}$$

with solution $\beta = \left(G^\top W G\right)^{-1} G^\top W X$ (or $\beta = \left(G^\top G\right)^{-1} G^\top X$, if $W = \mathbb{1}$). Note that only $z = 0$ is important to evaluate the polynomial (2.2), i.e., $m_\beta(0) \approx X_t$. That is,

$$X_t \approx g(0)^\top \beta = g(0)^\top \left(G^\top W G\right)^{-1} G^\top W X.$$

Remark 2.2. The formula (2.2) can be employed to predict $X_t \approx m_\beta(0)$ or to extrapolate the smoothed data by simply evaluating $X_{t+\Delta} = m_\beta(\Delta)$ at $z = \Delta$ appropriately.

Remark 2.3. The idea can be extended and used to higher dimensional data as well.

Example 2.4 (Savitzky–Golay filter)**.** For $m = 5$ and polynomials of degree $k = 3$ ($g(z) = (1, z, z^2, \dots, z^k)$) with $z \in \left\{-\frac{m-1}{2}, \dots, 0, \dots \frac{m-1}{2}\right\}$ (m odd) we obtain $G = \begin{pmatrix} 1 & -2 & 4 & -8 \\ 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{pmatrix}$ and $\left(G^\top G\right)^{-1} G^\top = \begin{pmatrix} -\frac{3}{35} & \frac{12}{35} & \frac{17}{35} & \frac{12}{35} & -\frac{3}{35} \\ \frac{1}{12} & -\frac{2}{3} & 0 & \frac{2}{3} & -\frac{1}{12} \\ \frac{1}{7} & -\frac{1}{14} & -\frac{1}{7} & -\frac{1}{14} & \frac{1}{7} \\ -\frac{1}{12} & \frac{1}{6} & 0 & -\frac{1}{6} & \frac{1}{12} \end{pmatrix}$. The regression polynomial, evaluated at $z = 0$, is the linear filter

$$m_t = \frac{1}{35} \left(-3\, X_{t-2} + 12\, X_{t-1} + 17\, X_t + 12\, X_{t+1} - 3\, X_{t+2}\right).$$

Example 2.5. For $z_i \in \{0, -1, -2, -3, -4\}$ and $k = 3$, the filter is

$$m_t = \frac{1}{70} \left(69\, X_t + 4\, X_{t-1} - 6\, X_{t-2} + 4\, X_{t-3} - X_{t-4}\right).$$

2.3.1 Spencer filter

The Spencer 15-point moving average (MA) filter has the weights

$$(a_{-7},\dots,a_7)=\frac{1}{320}\left(-3,-6,-5,3,21,46,67,74,67,46,21,3,-5,-6,-3\right).$$

Which polynomials are not distorted by the Spencer filter?

2.3.2 The moving average filter

The Savitzky–Golay filter with $k=0$ is given by $G=\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$ and $\left(G^\top G\right)^{-1}G^\top=\frac{1}{m}\left(1,\dots,1\right)$.

Here, the regression thus is $m_i=\frac{1}{m}\sum_{i=-\frac{m-1}{2}}^{\frac{m-1}{2}}X_i$ or

$$m_t=\frac{1}{2q+1}\sum_{j=-q}^{q}X_{t+j},\tag{2.4}$$

the moving average filter.

Remark 2.6. The filter (2.4) is also optimal for $k=1$.

Weights. The Savitzky–Golay filter with $k=0$ and weights $w=(w_1,\dots,w_m)$ (cf. (2.4)) is

$$m_t=\sum_{i\in W}\frac{w_i}{\sum_{i\in W}w_i}X_{t+i}.$$

2.4 DIFFERENCING

Definition 2.7. The *(backward) difference* operator is

$$\nabla X_t:=X_t-X_{t-1}=(\mathbb{1}-B)X_t,$$

where B is the *backshift*,[1]

$$BX_t=X_{t-1}.\tag{2.5}$$

Powers of this operator $\nabla^0:=\mathbb{1}$ and $\nabla^{j+1}:=\nabla\nabla^j$ are obvious. For example, $\nabla^2X_t=X_t-2X_{t-1}+X_{t-2}$, etc.

[1]The backward shift operator is occasionally called *lag operator* and denoted L.

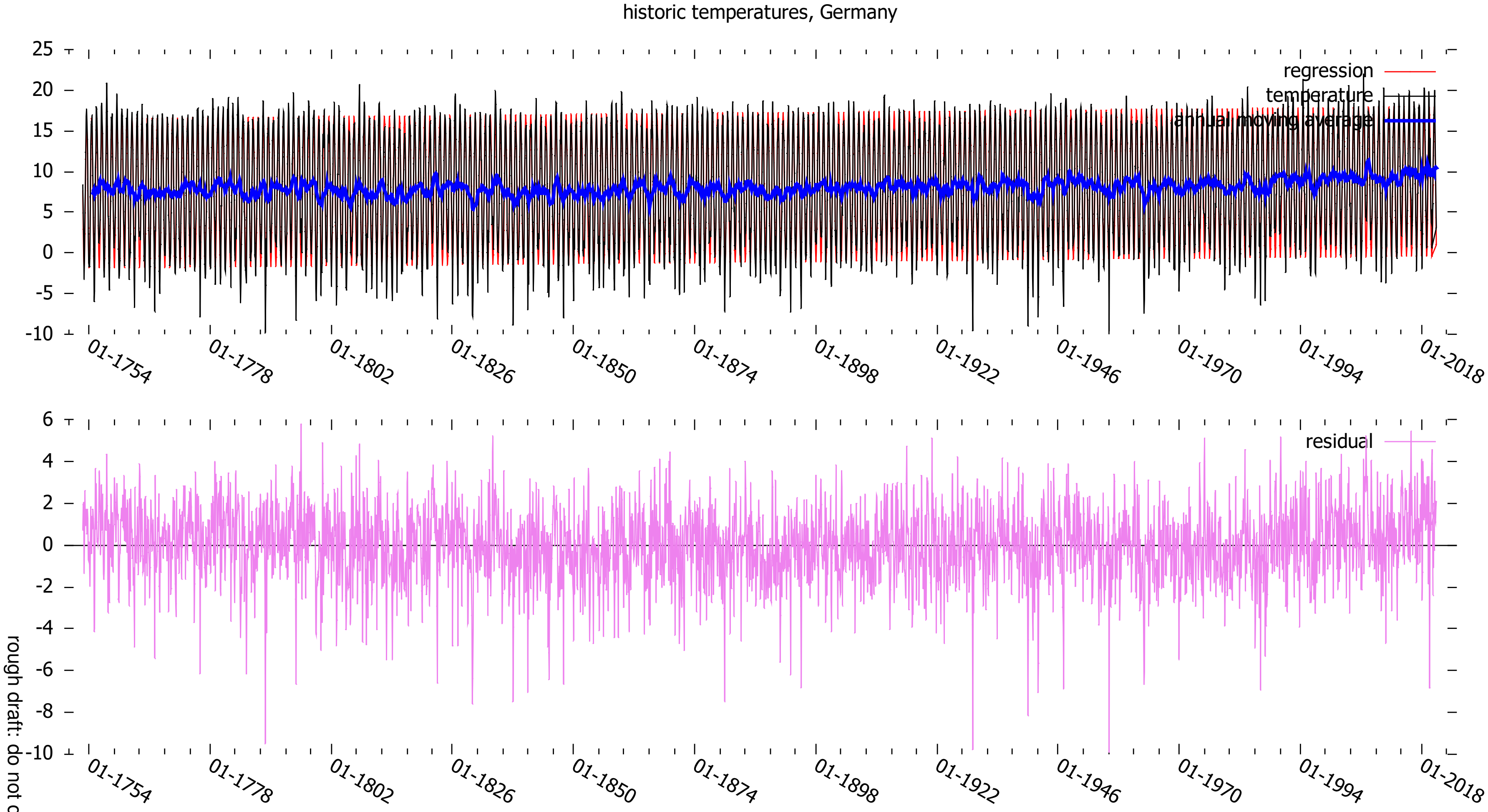


Figure 2.1: Time series of historic temperatures, Germany: temperature increases by $0.546°$ per century

Remark 2.8. As a matrix, mapping $(X_t)_{t\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$ to itself, the backshift is

$$B = \begin{pmatrix} \ddots & \ddots & \ddots & \ddots & \\ \ddots & 0 & 0 & 0 & \ddots \\ \ddots & 1 & 0 & 0 & \ddots \\ \ddots & 0 & 1 & 0 & \ddots \\ & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Definition 2.9. The *(forward) difference* is

$$\Delta X_t := X_{t+1} - X_t = (S - \mathbb{1})X_t, \tag{2.6}$$

where $S := B^{-1} = B^*$ is the (forward) shift.

Remark 2.10. The operators S and B are adjoint ($S = B^*$ and $B = S^*$) with respect to the inner product $\langle X \mid Y\rangle = \sum_{t\in\mathbb{Z}} X_t Y_t$, as $\langle X \mid SY\rangle = \sum_t X_t Y_{t+1} = \sum_t X_{t-1} Y_t = \langle BX \mid Y\rangle$.

Example 2.11 (Polynomial trend)**.** Suppose that $X_t = \underbrace{a + b\,t}_{\text{trend } m_t} + \underbrace{Z_t}_{\text{noise}}$, then $\nabla X_t = b + \nabla Z_t$ has constant trend and $\nabla^2 X_t = \nabla^2 Z_t$. More generally, for $X_t = \sum_{i=0}^k a_j\, t^j + Z_t$, then $\nabla^k X_t = k!a_k + \nabla^k Z_t$ and $\nabla^{k+1} X_t = \nabla^{k+1} Z_t$ completely removes the polynomial trend.

Definition 2.12. The operator

$$\nabla_\ell := \mathbb{1} - B^\ell \tag{2.7}$$

is called *lag-ℓ difference* operator.

Remark 2.13. Note that $\nabla_\ell = \mathbb{1} - B^\ell \neq (\mathbb{1} - B)^\ell = \nabla^\ell$ ($\ell > 1$).

2.5 LOG AND DIFFERENCING THE LOG

Consider and differentiate the transformed time series $\log X_t$. Note, that this filter is not linear.

2.6 THE SEASONAL COMPONENT

2.6.1 Lag-ℓ difference

To deseasonalize, one may also consider the filter $\nabla_\ell := \mathbb{1} - B^\ell$, cf. (2.7). For period d, applying ∇_d to the model $X_t = m_t + s_t + Z_t$ gives the new series $\nabla_d X_t = m_t - m_{t-d} + 0 + Z_t - Z_{t-d}$ with seasonal component s_t removed.

A further option is

$$m_t := \frac{1}{2}\left(X_t + X_{t-d/2}\right) = \frac{1}{2}\left(\mathbb{1} + B^{d/2}\right) X_t = \left(\mathbb{1} - {}^1\!/_2 \nabla_{d/2}\right) X_t; \tag{2.8}$$

indeed, $\frac{1}{2}\left(\mathbb{1}+B^{d/2}\right)X = \frac{1}{2}\left(m_t + m_{t-d/2}\right) + \frac{1}{2}\underbrace{\left(s_t + s_{t-d/2}\right)}_{0} + \frac{1}{2}\left(Z_t + Z_{t-d/2}\right)$ has the seasonal component with period d removed as well.

2.6.2 Non-integer periods

A generalization for periods $d \in \mathbb{R}$ which are not necessarily integers is the operator

$$\nabla_d := \left(1-(d-\lfloor d\rfloor)\right)\nabla_{\lfloor d\rfloor}X + \left(d-\lfloor d\rfloor\right)\nabla_{\lfloor d\rfloor+1}, \tag{2.9}$$

so that the formula (2.8) remains applicable (cf. (2.11)); equivalently,

$$B^d := (1-d+\lfloor d\rfloor)\, B^{\lfloor d\rfloor} + (d-\lfloor d\rfloor)\, B^{\lfloor d\rfloor+1}. \tag{2.10}$$

2.6.3 Moving average

The seasonal component can be removed by averaging. If the period is $d = 2q+1$, then the moving average filter (2.4) can do the job; for $d = 2q$, a useful filter to deseasonalize is

$$m_t = \frac{1}{2q}\left(\frac{1}{2}X_{t-q} + X_{t-q+1} + \cdots + X_{t+q-1} + \frac{1}{2}X_{t+q}\right).$$

Another variant is

$$m_t = \frac{1}{d}\left(X_t + X_{t-1} + \cdots + X_{t-\lfloor d\rfloor+1} + (d-\lfloor d\rfloor)\cdot X_{t-\lfloor d\rfloor}\right) \tag{2.11}$$

for a non-integer period $d > 0$.

2.7 EXPONENTIAL MOVING AVERAGE (EMA)

A.k.a. exponential smoothing. The smoothing operation is given recursively by

$$\begin{aligned} m_t &= \alpha X_t + (1-\alpha)m_{t-1} \\ &= m_{t-1} + \alpha\left(X_t - m_{t-1}\right) \end{aligned} \tag{2.12}$$

and $m_0 = X_0$, where $\alpha \in [0,1]$ is a model parameter called *exponential weight*. The parameter is often $\alpha = \frac{1}{d}$ or $\alpha = \frac{2}{d+1}$, where d is a sample period comparable to the period of the moving average. An explicit formula is

$$m_t = \sum_{i=1}^{t} \alpha(1-\alpha)^{t-i}X_i + (1-\alpha)^t X_0. \tag{2.13}$$

2.8 PROBLEMS

Exercise 2.1. *Show that a linear filter* (a_j) *passes every polynomial of degree* k *without distortion, i.e.,* $m_t = \sum_j a_j m_{t-1}$ *for all* $m_t = \sum_{i=0}^k c_i t^i$*, iff* $\sum_j a_j = 1$ *and* $\sum_j j^r a_j = 0$ *for* $r = 1, \dots, k$*.*

Exercise 2.2. *Show that the Spencer filter does not distort polynomials up to degree* 3*.*

Exercise 2.3. *The filter with binomial weights is* $a_j = \frac{1}{2^q}\binom{q}{j+q/2}$, $j = -\frac{q}{2}, \dots, \frac{q}{2}$*. Investigate its properties.*

Exercise 2.4. *Show that the backward difference operator satisfies* $\nabla^j X_t = \sum_{i=0}^j (-1)^i \binom{j}{i} X_{t-i}$*. Give the corresponding formula for the forward difference operator?*

Exercise 2.5. *Show that* (2.9) *and* (2.10) *are equivalent.*

Exercise 2.6 (Newton's backward difference formula)**.** *Show that*

$$X_t = X_0 + \frac{t}{1}\nabla_0^1 + \frac{t(t+1)}{2!}\nabla_0^2 + \frac{t(t+1)(t+2)}{3!}\nabla_0^3 + \dots$$

and compare the formula with the Taylor series expansion.

Exercise 2.7 (Newton's forward difference formula)**.** *Show that*

$$X_t = X_0 + \frac{t}{1}\Delta_0^1 + \frac{t(t-1)}{2!}\Delta_0^2 + \frac{t(t-1)(t-2)}{3!}\Delta_0^3 + \dots$$

and compare the formula with the Taylor series expansion.

Exercise 2.8. *Implement and visualize the filters* (2.8) *and* (2.11) *for the time series Example* (3.7)*.*

Exercise 2.9. *Implement the exponential smoothing filter* (2.12) *in Exercise 3.3.*

Exercise 2.10. *Argue why the filter* $\frac{1}{2}\left(\mathbb{1} + B^{d+d/2}\right) X_t$ *removes seasonality of period* d *as well.*

Exercise 2.11. *Remove the seasonality of the time series* $X_t = \sin(2\pi\xi_0 t + \varphi) + Z_t$ *(*ξ_0 *and* φ *deterministic), where* Z_t *are iid.*

Exercise 2.12. *Remove all seasonalities of the time series* $X_t = A_1 \sin(2\pi\xi_1 t + \varphi_1) + A_2 \sin(2\pi\xi_2 t + \varphi_2) + Z_t$*.*

Exercise 2.13. *Verify the exponential moving average* (2.13)*; show as well that the weights sum to* 1*.*

Stationarity 3

Things never happen the same way twice.

C. S. Lewis, 1889–1936

In what follows we assume that the trend and seasonalities are already removed.

Definition 3.1. Let $X_t \in \mathbb{R}^d$ be a stochastic process.

(i) mean function of a stochastic process is $\mu(t) := \mathbb{E}\, X_t$ ($\mu\colon T \to \mathbb{R}^d$).
(ii) The variance function is $\sigma^2(t) := \operatorname{var} X_t = \mathbb{E}\,\big(X_t - \mu(t)\big)\big(X_t - \mu(t)\big)^\top$ ($\sigma^2\colon T \to \mathbb{R}^{d\times d}$);
(iii) The *autocovariance function* is the Pearson covariance $\gamma(t,t') := \operatorname{cov}\big(X_t, X_{t'}\big)$ ($\gamma\colon T\times T \to \mathbb{R}^{d\times d}$).
(iv) The *autocorrelation function* is the Pearson correlation $\rho(t,t') := \frac{\operatorname{cov}\left(X_t, X_{t'}\right)}{\sqrt{\operatorname{var} X_t \cdot \operatorname{var} X_{t'}}}$.

Proposition 3.2. *We have that*

$$2\,\gamma(t,t') = (\mu(t) - \mu(t'))^2 + \operatorname{var} X_t + \operatorname{var} X_{t'} - \mathbb{E}\,(X_{t'} - X_t)^2 .$$

Proof. Indeed,

$$\begin{aligned}
\mathbb{E}\,(X_{t'} - X_t)^2 &= \mathbb{E}\Big(X_{t'} - \mu(t') - \big(X_t - \mu(t)\big) + \big(\mu(t') - \mu(t)\big)\Big)^2 \\
&= \mathbb{E}\big(X_{t'} - \mu(t')\big)^2 + \mathbb{E}\big(X_t - \mu(t)\big)^2 + \big(\mu(t') - \mu(t)\big)^2 \\
&\qquad - 2\cdot \mathbb{E}\big(X_{t'} - \mu(t')\big)\big(X_t - \mu(t)\big) \\
&\qquad + 2\cdot \Big(\mathbb{E}\big(X_{t'} - \mu(t')\big) - \mathbb{E}\big(X_t - \mu(t)\big)\Big)\cdot\big(\mu(t') - \mu(t)\big) \\
&= \operatorname{var} X_{t'} + \operatorname{var} X_t + \big(\mu(t) - \mu(t')\big)^2 - 2\,\gamma(t,t'),
\end{aligned}$$

from which the assertion follows. □

Definition 3.3. A stochastic process X_t is *weakly* or *wide-sense stationary* or *covariance stationary* if

(i) $\mathbb{E}\, X_t = \mu_X(t) = \mu_X(t+\tau) =: \mu$ for all $\tau \in T$ ($\mu_X\colon T \to \mathbb{R}^d$),
(ii) $\operatorname{var} X_t < \infty$ for all $t \in T$ and
(iii) $\operatorname{cov}(X_t, X_{t'}) = \mathbb{E}\,\big(X_t - \mu_X(t)\big)\big(X_{t'} - \mu_X(t')\big) =: \gamma_X(t,t') = \gamma_X(|t-t'|)$ for $\gamma_X\colon \mathbb{Z} \to \mathbb{R}$.

Proposition 3.4. *Suppose the process is weakly stationary. Then*

$$\gamma(h) = \operatorname{var} X_t - \frac{1}{2}\,\mathbb{E}\,(X_{t+h} - X_t)^2 .$$

Proof. The assertion is immediate from Proposition 3.2. □

Remark 3.5 (Variogram). A spatial analogue of the (temporal) covariance used in geostatistics is the variogram (semivariogram; not to be confused with covariance; kriging). It is defined as $\gamma(x,y)=\frac{1}{2}\,\mathbb{E}\,(Z(x)-Z(y))^2$, where $Z(\cdot)$ is a random field.

Definition 3.6 (Strict stationarity). A stochastic process X_t is *stationary* (strictly stationary), if the cumulative distribution functions satisfy

$$F_X(x_{t_1+\tau},\dots,x_{t_k+\tau})=F_X(x_{t_1},\dots,x_{t_k})$$

for all $t_1<\dots<t_k\in T$ and $\tau\ge 0$.

A process is a *Gaussian process* if $(X_{t_1},\dots X_{t_n})$ is multivariate normal for every n-tuple $(t_1,\dots t_n)$.

Remark 3.7. The augmented Dickey–Fuller test (ADF test) is the most prominent test to test stationarity.

Definition 3.8. Let X_t be a weakly stationary process. The covariance function is the even function

$$\gamma(\tau)\coloneqq\operatorname{cov}\left(X_{t+\tau},X_t\right).$$

The autocorrelation function (aka. serial correlation or lagged correlation) is

$$\rho(\tau)\coloneqq\frac{\gamma(\tau)}{\sqrt{\operatorname{var}X_t}\cdot\sqrt{\operatorname{var}X_{t+\tau}}}.$$

Remark 3.9. Note, that $\gamma(\tau)=\gamma(-\tau)$, that $\gamma(0)=\operatorname{var}X_t$ and $\rho(0)=1$.

Remark 3.10 (Z-transform). For a weakly stationary process X_t with $\mu_X\coloneqq\mathbb{E}\,X_t$ set $\sigma_X^2\coloneqq\gamma_X(0)=\operatorname{var}X_t$. Then the time series $X_t'\coloneqq\frac{X_t-\mu_X}{\sigma_X}$ is zero mean ($\mathbb{E}\,X_t'=0$) and variance $\sigma_{X'}^2\coloneqq\operatorname{var}X_t'=1$. The covariance is $\gamma_X(t)=\sigma_X^2\cdot\rho_{X'}(t)$ so that is enough to consider the correlation ρ in what follows.

Proposition 3.11. *The covariance function is non-negative definite, i.e.,*

$$\sum_{i,j=1}^{n}a_i\,\gamma(i-j)\,a_j\ge 0 \tag{3.1}$$

for all $n\ge 1$ and all $a_1,\dots,a_n$.

Proof. Consider the random vector $Z\coloneqq(X_1-\mathbb{E}\,X_1,\dots,X_n-\mathbb{E}\,X_n)$. It holds that

$$0\le\operatorname{var}\left(a^\top Z\right)=\mathbb{E}\left(a^\top Z\right)\left(a^\top Z\right)^\top=\mathbb{E}\,a^\top ZZ^\top a=a^\top\,\mathbb{E}\left(ZZ^\top\right)a=\sum_{i,j}a_i\,\gamma(i-j)\,a_j$$

and thus (3.1). □

Definition 3.12 (*White noise* or *white independent noise*)**.** The time series X_t with uncorrelated (but not necessarily independent) components is called *white noise* and often denoted w_t. We shall write

$$w_t \sim \left(\mu_w, \sigma_w^2\right).$$

The autocovariance function of the white noise is the covariance function of iid noise is

$$\gamma(t+\tau, t) = \gamma(\tau) = \begin{cases} \sigma_w^2 & \text{if } \tau = 0, \\ 0 & \text{else.} \end{cases} \tag{3.2}$$

Definition 3.13 (iid noise)**.** The time series $X_1, X_2, \dots$ for X_i iid with mean $\mathbb{E}\, X_i = 0$ is called iid noise.

It holds that $P\big(X_1 \le x_1, \dots X_n \le x_n\big) = P(X_1 \le x_1) \cdot \ldots \cdot P(X_n \le x_n)$ and thus $P\big(X_{n+\ell} \le x \mid X_1, \dots X_n\big) = P(X_{n+\ell} \le x)$ and thus has no value for predicting the time series. The autocovariance function (provided that $\operatorname{var} X_j < \infty$) is (3.2).

Definition 3.14 (Gaussian)**.** Terms as Gaussian white noise or Gaussian iid noise are evident.

Example 3.15 (Periodic time series)**.** Consider the periodic time series

$$X_t = A\cos(2\pi\xi_0 t) + B\sin(2\pi\xi_0 t) \tag{3.3}$$

for A, B uncorrelated, mean zero, variance σ^2 and angular frequency ξ_0 fixed. Then[1]

$$\begin{aligned}
\gamma(\tau) &= \operatorname{cov}(X_t, X_{t+\tau}) = \mathbb{E}\, X_{t+\tau} X_t \\
&= \mathbb{E}\big(A\cos 2\pi\xi_0 t + B\sin 2\pi\xi_0 t\big)\big(A\cos 2\pi\xi_0(t+\tau) + B\sin 2\pi\xi_0(t+\tau)\big) \\
&= \mathbb{E}\, A^2 \cos 2\pi\xi_0 t \cdot \cos 2\pi\xi_0(t+\tau) + B^2 \sin 2\pi\xi_0 t \cdot \sin 2\pi\xi_0(t+\tau) \\
&= \sigma^2 \cos 2\pi\xi_0\big(t - (t+\tau)\big) = \sigma^2 \cos 2\pi\xi_0 \tau.
\end{aligned}$$

Example 3.16 (Cf. Proposition 4.6 below)**.** Consider $X_t \coloneqq Z_t + \theta\, Z_{t-1}$ with Z_t uncorrelated, zero-mean and variance σ_Z^2. Then

$$\gamma(\ell) = \begin{cases} (1+\theta^2)\,\sigma_Z^2 & \text{if } \ell = 0, \\ \theta\,\sigma_Z^2 & \text{if } \ell = \pm 1, \\ 0 & \text{else} \end{cases}$$

and X_t is weakly stationary, as $m_t = 0$.

[1] Recall the trigonometric identities

$$\begin{aligned}
\sin(\alpha \pm \beta) &= \sin\alpha\cos\beta \pm \cos\alpha\sin\beta \quad \text{and} \\
\cos(\alpha \pm \beta) &= \cos\alpha\cos\beta \mp \sin\alpha\sin\beta.
\end{aligned}$$

Example 3.17 (Random walk)**.** For X_t, $t = 1, 2, \dots$ uncorrelated, zero-mean and variance σ^2 define $S_t := X_1 + X_2 + \dots + X_t$. Then

$$\operatorname{cov}(S_{t+h}, S_t) = \operatorname{cov}\left(\sum_{i=1}^{t+h} X_i, \sum_{j=1}^{t} X_j\right) = \sum_{i=1}^{t+h}\sum_{j=1}^{t} \operatorname{cov}\left(X_i, X_j\right) = \sum_{i,j=1}^{t} \operatorname{cov}\left(X_i, X_j\right) = t \cdot \sigma^2,$$

which depends on t but not on h. S_t thus is *not* stationary.

3.1 LINEAR PROCESS WITH GIVEN AUTOCOVARIANCE

We are interested in a weakly stationary time series $X_0, X_1, \dots$ so that $\operatorname{var} X_k = \sigma^2$ and $\operatorname{cov}(X_k, X_\ell) = \gamma_{k-\ell}$.

Proposition 3.18 (Yule–Walker)**.** *Suppose that Z_t, $t = 0, \dots$, are uncorrelated, zero mean $\mathbb{E}\, Z_t = 0$ with variance $\operatorname{var} Z_t = 1$ (not necessarily iid, white noise, e.g.). Then, for $\gamma(\cdot)$ positive (cf. (3.1)), the time series*

$$X_t = \phi_{tt} \cdot X_0 + \dots + \phi_{t1} \cdot X_{t-1} + \psi_t \cdot Z_t = \sum_{i=0}^{t-1} \phi_{tt-i}\, X_i + \psi_t\, Z_t \tag{3.4}$$

has the acf $\operatorname{cov}(X_k, X_\ell) = \gamma_{k-\ell}$, where the coefficients satisfy

$$\underbrace{\begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{t-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{t-1} & \dots & \gamma_1 & \gamma_0 \end{pmatrix}}_{\Gamma_t} \underbrace{\begin{pmatrix} \phi_{t1} \\ \phi_{t2} \\ \vdots \\ \phi_{tt} \end{pmatrix}}_{\Phi_t} = \underbrace{\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_t \end{pmatrix}}_{r_t} \tag{3.5}$$

and

$$\psi_t^2 := \sigma^2 - r_t^\top \Phi_t > 0. \tag{3.6}$$

Remark 3.19. The matrix Γ_t is a Toeplitz matrix. Note the reverse order in (3.4).

Corollary 3.20. *The function $\gamma(\cdot)$ is the acf of a time series iff $\gamma(\cdot)$ is positive.*

Corollary 3.21 (Cf. Proposition 3.11)**.** *The matrix*

$$\Gamma_t := \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{t-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{t-1} & \dots & \gamma_1 & \gamma_0 \end{pmatrix} = \sigma^2 \cdot \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{t-1} \\ \rho_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_1 \\ \rho_{t-1} & \dots & \rho_1 & 1 \end{pmatrix}$$

is positive definite.

Definition 3.22 (Yule–Walker)**.** Equations (3.5) are the *Yule–Walker* equations.

Proof of Proposition 3.18. By construction, Z_t is independent from X_i, $i = 0, \dots, t-1$, so we deduce from (3.4) that

$$\begin{aligned}\gamma_k = \mathbb{E}\, X_t X_{t-k} &= \mathbb{E}\left(\sum_{i=1}^t \phi_{ti} X_{t-i} + \psi_t Z_t\right) \cdot X_{t-k},\\ &= \sum_{i=1}^t \gamma_{k-i}\,\phi_{ti}, \qquad k = 1, \dots t,\end{aligned}$$

i.e., $r_t = \Gamma_t\,\Phi_t$. It follows that X_t has the desired covariance structure if the coefficients in (3.4) are $\Phi_t = \Gamma_t^{-1}\, r_t$.

Further,

$$\operatorname{var} X_t = \sum_{i,j=1}^t \phi_{ti}\,\phi_{tj}\ \mathbb{E}\, X_{t-i} X_{t-j} + \psi_t^2 = \sum_{i,j=1}^t \phi_{ti}\,\gamma_{i-j}\,\phi_{tj} + \psi_t^2$$

and we thus find $\Phi_t^\top \Gamma_t \Phi_t + \psi_t^2 = \sigma^2$ to obtain $\operatorname{var} X_t = \sigma^2$, i.e., (3.6) by employing (3.5). □

Proposition 3.23 (Durbin, cf. the Levinson Algorithm in Golub and Van Loan (2013))**.** *The solution of the Yule–Walker equations can be updated recursively as*

$$\alpha_{t+1} = \frac{\gamma_{t+1} - r_t^\top J_t \Phi_t}{\psi_t^2}, \tag{3.7}$$

$$\Phi_{t+1} = \begin{pmatrix}\Phi_t - \alpha_{t+1} J_t \Phi_t \\ \alpha_{t+1}\end{pmatrix} \quad \text{and} \tag{3.8}$$

$$\psi_{t+1}^2 = \psi_t^2 \left(1 - \alpha_{t+1}^2\right), \tag{3.9}$$

where $J_t := \begin{pmatrix} \dots & 0 & 1 \\ \iddots & \iddots & 0 \\ 1 & \iddots & \end{pmatrix}$ *is the t-by-t exchange matrix.*

Remark 3.24*.* The initial conditions and first solutions are

- $t = 0$: $\Phi_0 := r_0 := ()$, $\psi_0 := \sigma^2$, $\alpha_1 = \frac{\gamma_1}{\sigma} = \rho_1$ (cf. (3.7)) and thus $X_0 = \sigma\, Z_0$;
- $t = 1$: $\Phi_1 = r_1 = (\rho_1)$, $\psi_1^2 = \sigma^2(1-\rho_1^2)$ and thus $X_1 = \rho_1 \cdot X_0 + \sigma\sqrt{1-\rho_1^2} \cdot Z_1$;
- $t = 2$: $\Phi_2 = \frac{1}{1-\rho_1^2}\begin{pmatrix}\rho_1 - \rho_1\rho_2 \\ \rho_2 - \rho_1^2\end{pmatrix}$ and thus

$$X_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} X_0 + \frac{\rho_1 - \rho_1\rho_2}{1-\rho_1^2} X_1 + \sigma\sqrt{\frac{1 - 2\rho_1^2 + \rho_2^2}{1-\rho_1^2}} Z_2. \tag{3.10}$$

Note that the necessary memory allocation for the update is $t + O(1)$ and the time to compute the update is $t + O(1)$. So the total cost to compute Φ_t and α_t are $t^2/2 + O(t)$ instead of $O(t^3)$ when inverting (3.5) directly.

Proof of Proposition 3.23. Note that $J_t\Gamma_t = \Gamma_t J_t$ (cf. Exercise 3.12). The Yule–Walker equations (3.5) for $t+1$ read $\underbrace{\begin{pmatrix}\Gamma_t & J_t r_t\\ r_t^\top J_t & \sigma^2\end{pmatrix}}_{\Gamma_{t+1}}\begin{pmatrix}z_t\\ \alpha_{t+1}\end{pmatrix} = \begin{pmatrix}r_t\\ \gamma_{t+1}\end{pmatrix}$ and it follows that

$$z_t = \Gamma_t^{-1}\left(r_t - \alpha_{t+1}J_t r_t\right) = \Phi_t - \alpha_{t+1}J_t\Gamma_t^{-1}r_t = \Phi_t - \alpha_{t+1}J_t\Phi_t$$

and

$$\sigma^2\,\alpha_{t+1} = \gamma_{t+1} - r_t^\top J_t z_t = \gamma_{t+1} - r_t^\top J_t\left(\Phi_t - \alpha_{t+1}J_t\Phi_t\right) = \gamma_{t+1} - r_t^\top J_t\Phi_t + \alpha_{t+1}r_t^\top\Phi_t$$

and thus $\alpha_{t+1} = \frac{\gamma_{t+1} - r_t^\top J_t\Phi_t}{\sigma^2 - r_t^\top\Phi_t}$, so (3.7) and (3.8) with (3.6). Next,

$$\begin{aligned}\psi_{t+1}^2 &\underset{(3.6)}{=} \sigma^2 - r_{t+1}^\top\Phi_{t+1} = \sigma^2 - \begin{pmatrix}r_t\\ \gamma_{t+1}\end{pmatrix}^\top\begin{pmatrix}\Phi_t - \alpha_{t+1}J_t\Phi_t\\ \alpha_{t+1}\end{pmatrix}\\ &= \sigma^2 - r_t^\top\Phi_t + \alpha_{t+1}r_t^\top J_t\Phi_t - \alpha_{t+1}\gamma_{t+1} = \psi_t^2 - \alpha_{t+1}\left(\gamma_{t+1} - r_t^\top J_t\Phi_t\right)\\ &\underset{(3.7)}{=} \psi_t^2 - \alpha_{t+1}\alpha_{t+1}\psi_t^2 = \psi_t^2\left(1 - \alpha_{t+1}^2\right)\end{aligned}$$

and thus (3.9).

Finally recall that the matrix Γ_{t+1} is positive definite. It follows for $\begin{pmatrix}-J_t\phi_t\\ 1\end{pmatrix}$ that

$$0 \le \begin{pmatrix}-J_t\phi_t\\ 1\end{pmatrix}^\top\underbrace{\begin{pmatrix}\Gamma_t & J_t r_t\\ r_t^\top J_t & \sigma^2\end{pmatrix}}_{\Gamma_{t+1}}\begin{pmatrix}-J_t\Phi_t\\ 1\end{pmatrix} = \begin{pmatrix}-\Phi_t^\top J_t & 1\end{pmatrix}\begin{pmatrix}0\\ \sigma^2 - r_t^\top\Phi_t\end{pmatrix} = \sigma^2 - r_t^\top\Phi_t = \psi_t^2$$

and thus $\psi_t > 0$ is well defined. □

Definition 3.25 (Partial autocorrelation)**.** The *partial autocorrelation at lag* ℓ (or order ℓ) of the stationary time series X_t is

$$\alpha(\ell) = \operatorname{corr}\left(X_{t+\ell}, X_t \mid X_{t+1}, \dots X_{t+\ell-1}\right) = \operatorname{corr}\left(X_t, X_{t-\ell} \mid X_{t-\ell+1}, \dots X_{t-1}\right)$$

(conditioning on the intervening variables).

The partial autocorrelations are often called *reflection coefficients* (particularly in signal processing).

Remark 3.26. Apparently, $\alpha(1) = \rho_1$.

Proposition 3.27. *For a time series with mean* 0 *it holds that*

$$\alpha(\ell) = \Phi_{\ell\ell} = \alpha_\ell,$$

where $\Phi_\ell = R_\ell^{-1} r_\ell$ *(the solution of the Yule–Walker equation).*

Proof. Indeed, this follows with (3.4) from

$$X_t = \underbrace{\phi_{\ell 1} \cdot X_{t-1} + \dots + \phi_{\ell\ell-1} \cdot X_{t-\ell+1}}_{\text{conditioned}} + \phi_{\ell\ell} \cdot X_{t-\ell} + \psi_\ell \cdot Z_\ell.$$

□

From Remark 3.24 it follows that $\alpha(0) = 1$, $\alpha(1) = \rho_1$, $\alpha(2) = \frac{\rho_2 - \rho_1\rho_1}{1-\rho_1^2}$ and $\alpha(3) = \frac{\rho_1^3 - \rho_1^2\rho_3 - \rho_1\rho_2(2-\rho_2) + \rho_3}{(1-\rho_2)(1-2\rho_1^2+\rho_2)}$, etc.

3.2 PROBLEMS

Exercise 3.1. *Simulate and visualize the time series* (3.3)*.*

Exercise 3.2. *Visualize samples of the time series from Example 3.16.*

Exercise 3.3 (Constant acf)**.** *Let* Z_i *be independent with* $\mathbb{E}\, Z_i = 0$ *and* $\operatorname{var} Z_i =: \sigma^2$*,* $i = 0, 1, \dots$*. Define* $X_0 := Z_0$ *and recursively*

$$X_i := \rho_i \cdot \frac{1}{i} \sum_{j=0}^{i-1} X_j + \sqrt{1 - \rho_i \cdot \rho} \cdot Z_i$$

with $\rho_i = \frac{i\rho}{1+(i-1)\rho}$*. Simulate and visualize the time series* X_i*,* $i = 0, 1, \dots$*.*

Exercise 3.4. *Consider the time series* X_i *given in Exercise 3.3. Show that* $\mathbb{E}\, X_i = 0$*,* $\operatorname{var} X_i = \sigma^2$ *for all* $i \in \{0, 1, \dots\}$ *and* $\operatorname{corr}(X_i, X_j) = \rho$ *whenever* $i \neq j$ *(Hint: show the result for* $i = 0$*,* $i = 1$ *first and use induction on* i*; as a side result,* $\operatorname{var}\left(\frac{1}{i}\sum_{j=0}^{i-1} X_j\right) = \frac{i + i(i-1)\rho}{i^2}$*.)*

Exercise 3.5. *Suppose that* $\operatorname{corr}(X_i, X_j) \le \rho$ *for* $0 \le i, j \le n$*,* $i \neq j$*. Show that* $n \ge -\frac{1}{\rho}$*. Discuss the consequences for the time series in Example 3.3 and show as well that* $\rho_0 = 0 \le \rho = \rho_1 \le \rho_i \le \rho_{i+1} \xrightarrow[i\to\infty]{} 1$*.*

Exercise 3.6. *Discuss Exercise 3.3 for Gaussian random variables.*

Exercise 3.7. *Simulate a time series with autocovariance function* $\ell \mapsto \begin{cases} 1 & \text{if } \ell = 0, \\ 0.9 & \text{if } \ell = \pm 1, \\ 0.7 & \text{if } \ell = \pm 2 \end{cases}$*?*

Exercise 3.8. *Is there a time series with autocovariance function* $\ell \mapsto \begin{cases} 1 & \text{if } \ell = 0, \\ 0.9 & \text{if } \ell = \pm 1, \\ 0.6 & \text{if } \ell = \pm 2 \end{cases}$ *?*

Exercise 3.9. *Show that* $\ell \mapsto \begin{cases} 1 & \text{if } \ell = 0, \\ \rho & \text{if } \ell = \pm 1, \\ 0 & \text{else} \end{cases}$ *is an autocovariance function of a time series iff* $|\rho| \le \frac{1}{2}$. *(Hint: choose* $a = (1, -1, 1, -1, \dots)$ *in* (3.1)*).*

Exercise 3.10. *Verify* (3.10) *explicitly.*

Exercise 3.11. *Verify the Woodbury matrix identity.*

Exercise 3.12. *Verify the update* (3.6) *(use that* R_n *and* R_n^{-1} *are persymmetric matrices, i.e.,* $R_n^{-1} J_n = J_n R_n^{-1}$*).*

Exercise 3.13. *Implement the algorithm* (3.4) *and run tests for your choice of* ρ_ℓ*, where*
$\sum_{\ell \in \mathbb{Z}} |\rho_\ell| < \infty$ *(i.e.,* $(\rho_\ell)_\ell \in \ell_1$*, the space of absolutely summable sequencs)*
$\rho_\ell \xrightarrow[\ell \to \infty]{} 0$ *but* $\sum_{\ell \in \mathbb{Z}} |\rho_\ell| = \infty$ *and*
$\liminf_{\ell \to \infty} \rho_\ell > 0$.

Exercise 3.14. *Set* $\overline{R}_t := \begin{pmatrix} R_t & 0 \\ 0 & 1 \end{pmatrix}$*. With* $U := \begin{pmatrix} \rho_t & 0 \\ \vdots & \vdots \\ \rho_1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} J_t r_t & 0 \\ 0 & 1 \end{pmatrix}$ *and* $V := \begin{pmatrix} 0 & \dots & 0 & 1 \\ \rho_t & \dots & \rho_1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ r_t^\top J_t & 0 \end{pmatrix}$*, then* $R_{t+1} = \overline{R}_t + U \cdot V$*. By employing the Woodbury matrix identity (rank two update, aka. Sherman–Morrison–Woodbury formula, Exercise 3.11)*

$$R_{t+1}^{-1} = \overline{R}_t^{-1} - \overline{R}_t^{-1} U \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + V \overline{R}_t^{-1} U \right)^{-1} V \overline{R}_t^{-1}.$$

We have (use Exercise 3.12) $\overline{R}_t^{-1} U = \begin{pmatrix} R_t^{-1} J_t r_t & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} J_t \Phi_t & 0 \\ 0 & 1 \end{pmatrix}$*, thus* $V \overline{R}_t^{-1} U = \begin{pmatrix} 0 & 1 \\ r_t^\top \Phi_t & 0 \end{pmatrix}$ *and* $\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + V \overline{R}_t^{-1} U \right)^{-1} = \frac{1}{1 - r_t^\top \Phi_t} \begin{pmatrix} 1 & -1 \\ -r_t^\top \Phi_t & 1 \end{pmatrix}$*. It follows that*

$$\begin{aligned}
\Phi_{t+1} = R_{t+1}^{-1} r_{t+1} &= \begin{pmatrix} \Phi_t \\ \rho_{t+1} \end{pmatrix} - \begin{pmatrix} J_t \Phi_t & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{1 - r_t^\top \Phi_t} \begin{pmatrix} 1 & -1 \\ -r_t^\top \Phi_t & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ r_t^\top J_t & 0 \end{pmatrix} \begin{pmatrix} \Phi_t \\ \rho_{t+1} \end{pmatrix} \\
&= \begin{pmatrix} \Phi_t \\ \rho_{t+1} \end{pmatrix} - \frac{1}{1 - r_t^\top \Phi_t} \begin{pmatrix} J_t \Phi_t & -J_t \Phi_t \\ -r_t^\top \Phi_t & 1 \end{pmatrix} \begin{pmatrix} \rho_{t+1} \\ r_t^\top J_t \Phi_t \end{pmatrix} \\
&= \begin{pmatrix} \Phi_t \\ \rho_{t+1} \end{pmatrix} - \frac{1}{1 - r_t^\top \Phi_t} \begin{pmatrix} \left(\rho_{t+1} - r_t^\top J_t \Phi_t \right) J_t \Phi_t \\ r_t^\top J_t \Phi_t - \rho_{t+1} r_t^\top \Phi_t \end{pmatrix},
\end{aligned}$$

a restatement of (3.8).

Exercise 3.15. *In the setting of Example 3.16 set* $m(\lambda) \coloneqq \mathbb{E}\exp(\lambda Z_i)$. *Express the joint moment generating function* $\mathbb{E}\exp\left(\sum_{i=1}^n \lambda_i X_i\right)$ *in terms of the function* $m(\cdot)$. *Deduce that* (X_t) *is stationary.*

Exercise 3.16. *Which of the following processes is weakly, which is strictly stationary for iid.* Z_t, $t \in \mathbb{Z}$*?*

- $X_t = a + bZ_t + cZ_{t-1}$,
- $X_t = a + bZ_0$,
- $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$,
- $X_t = Z_0 \cos(ct)$,
- $X_t = Z_t \cos(ct) + Z_{t-1} \sin(ct)$,
- $X_t = Z_t Z_{t-1}$.

Exercise 3.17. *For* Y_t *iid define* $X_t \coloneqq a + bt + Y_t$ *and* $W_t \coloneqq \frac{1}{2q+1}\sum_{j=-q}^{q} X_{t+j}$. *Is* W_t *starionary? Compute* $\operatorname{cov}(W_{t+\ell}, W_t)$.

Exercise 3.18. *Suppose that* (X_t) *and* (Y_t) *are each stationary and independent. Compute the acf. of the process* $X_t + Y_t$.

4 Parametric models

4.1 ARMA

ARMA (Autoregressive-Moving Average) provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the autoregression and the second for the moving average. The general ARMA model was described in the 1951 thesis of Whittle,[1] Hypothesis testing in time series analysis, and it was popularized in the 1970 book by Box[2] and Jenkins.[3] (Wikipedia)

Definition 4.1 (ARMA)**.** The process X_t is an ARMA(p,q) process if the recursion

$$\underbrace{X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p}}_{\text{auto regressive, AR}} + \underbrace{Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}}_{\text{moving average, MA}} \tag{4.1}$$

is valid for the innovation $Z_t \sim \mathcal{N}(0, \sigma_Z^2)$, a white noise process. The parameters are ϕ_i, $i = 1, \dots, p$, and θ_j, $j = 1, \dots q$. For convenience, we set $\theta_0 \coloneqq 1$. The *lag orders* are p and q.

Definition 4.2. With an ARMA(p,q) model we associate the polynomials

$$\begin{aligned}\phi(z) &\coloneqq 1 - \phi_1 z - \dots - \phi_p z^p \text{ (AR polynomial) and}\\ \theta(z) &\coloneqq 1 + \theta_1 z + \dots + \theta_q z^q \text{ (MA polynomial).}\end{aligned}$$

Employing the backshift operator B (cf. (2.5)) the ARMA(p,q) time series X_t solves the equation

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

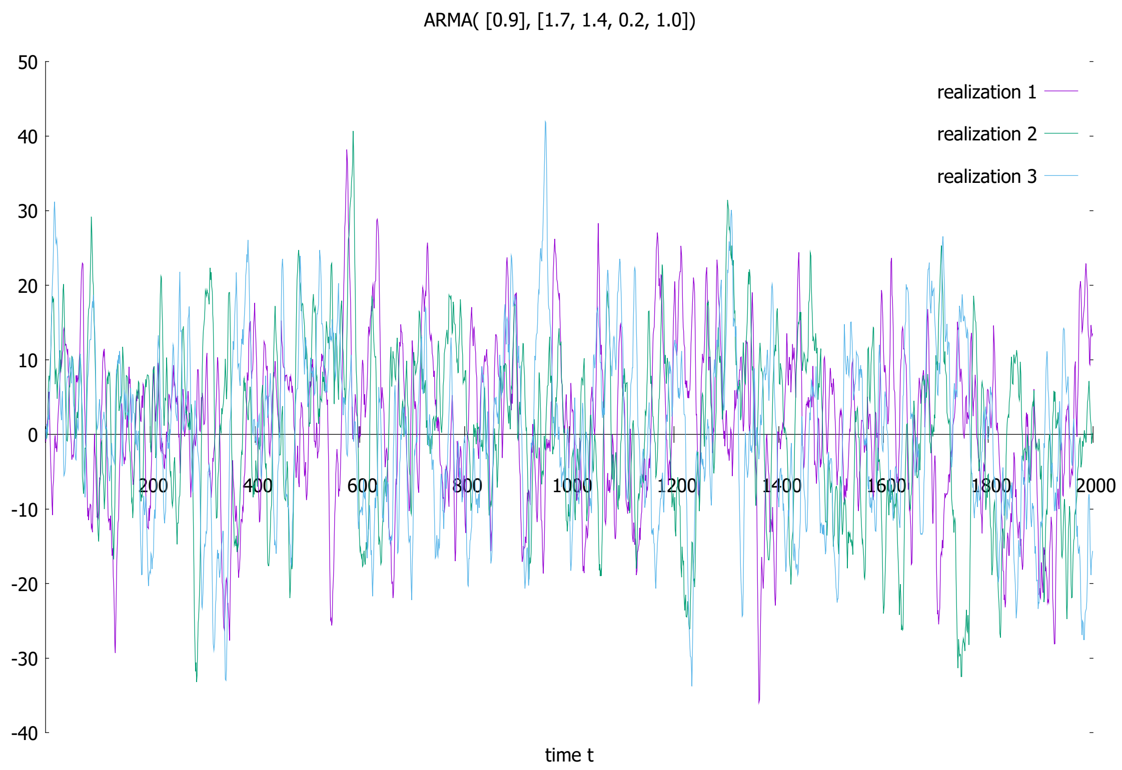
Remark 4.3 (Expectation)*.* Taking expectations in (4.1) reveals that

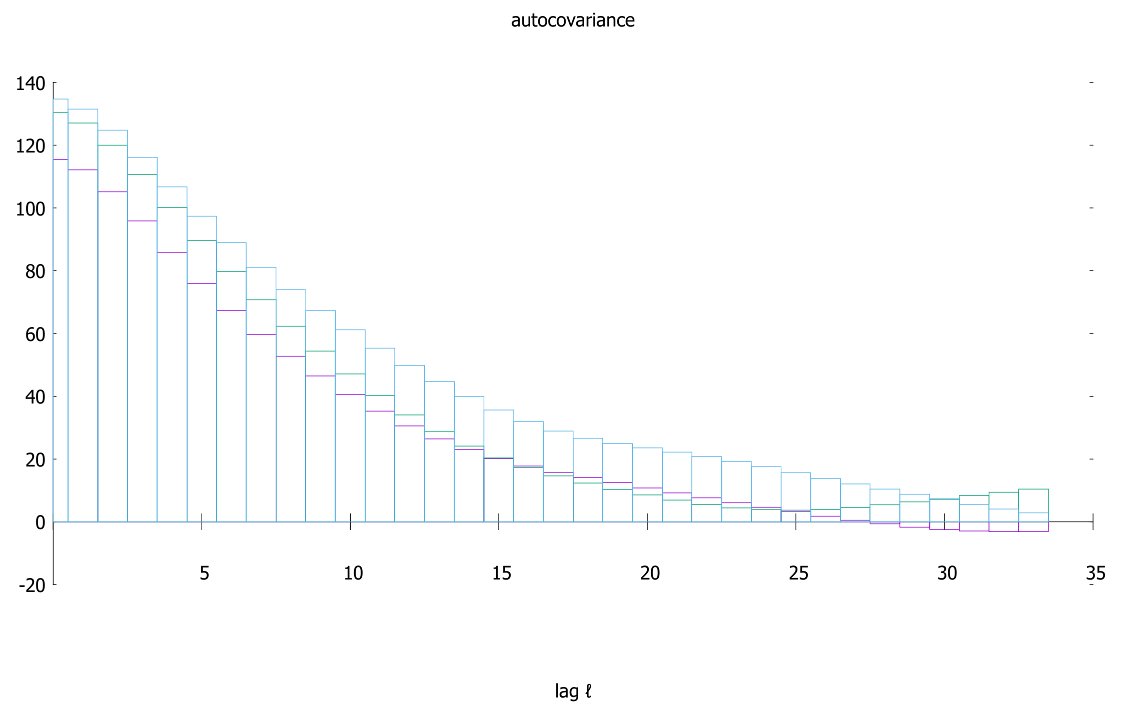
$$\mathbb{E}\, X_t = \frac{\theta(1)}{\phi(1)} \mathbb{E}\, Z_t.$$

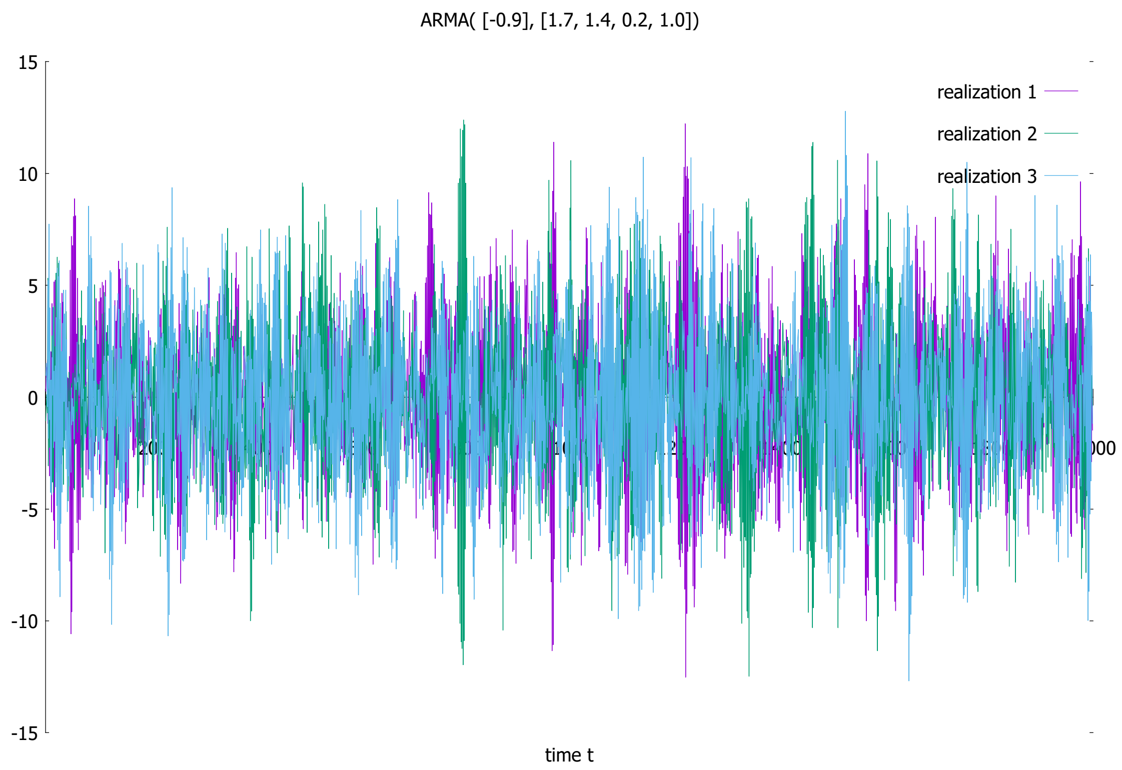
Remark 4.4 (Normalizing, standardizing)*.* Suppose that the stationary time series $\tilde{X}_t$ satisfies the more general equations

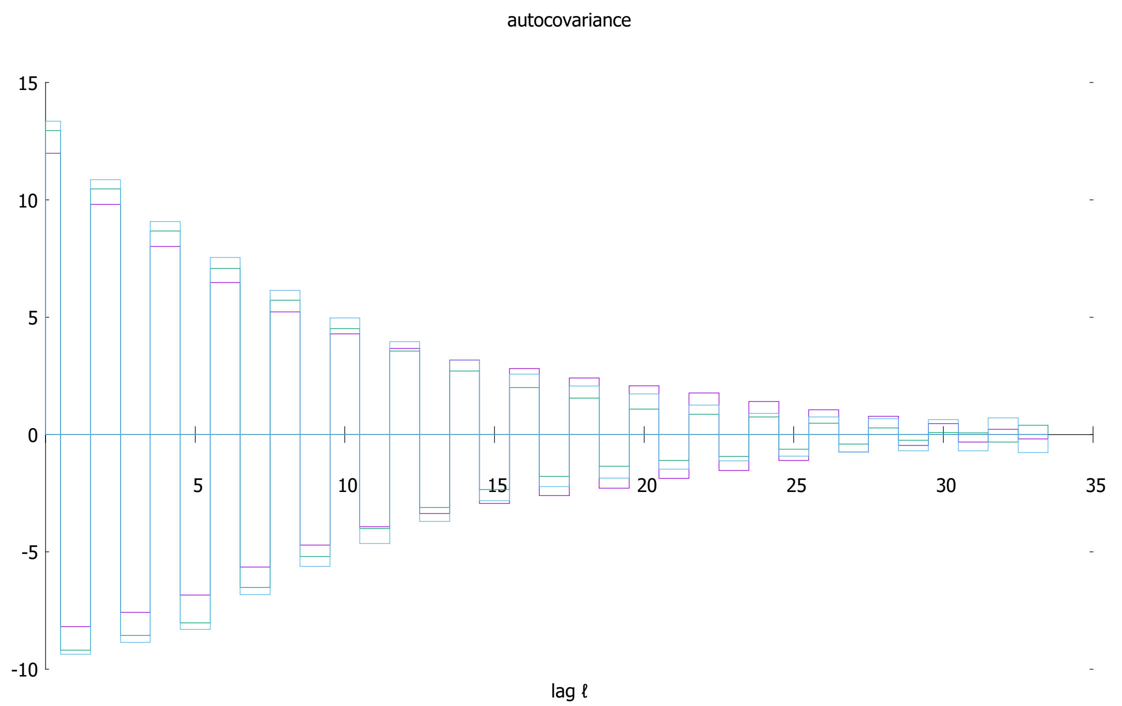
$$\underbrace{\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \dots + \phi_p \tilde{X}_{t-p}}_{\phi(B)\tilde{X}_t} + \nu + \underbrace{\tilde{\theta}_0 \tilde{Z}_t + \tilde{\theta}_1 \tilde{Z}_{t-1} + \dots + \tilde{\theta}_q \tilde{Z}_{t-q}}_{\tilde{\theta}(B)\tilde{Z}_t}, \tag{4.2}$$

[1] Peter Whittle, 1927–2021
[2] George E. P. Box, 1919–2013
[3] Gwilym Jenkins, 1932–1982


ARMA([0.9], [1.7, 1.4, 0.2, 1.0])
realization 1
realization 2
realization 3
time t
50
40
30
20
10
0
-10
-20
-30
-40
200
400
600
800
1000
1200
1400
1600
1800
2000


autocovariance
lag ℓ
140
120
100
80
60
40
20
0
-20
5
10
15
20
25
30
35


ARMA([-0.9], [1.7, 1.4, 0.2, 1.0])
15
10
5
0
-5
-10
-15
realization 1
realization 2
realization 3
time t


autocovariance
15
10
5
0
-5
-10
5
10
15
20
25
30
35
lag ℓ

i.e., $\phi(B)\tilde{X} = \nu + \theta(B)\tilde{Z}$. Then the expectation is (with $\mu_{\tilde{X}} := \mathbb{E}\,\tilde{X}_t$ and $\mu_{\tilde{Z}} := \mathbb{E}\,\tilde{Z}_t$)

$$\mu_{\tilde{X}} = \phi_1\mu_{\tilde{X}} + \cdots + \phi_p\mu_{\tilde{X}} + \nu + \tilde{\theta}_0\mu_{\tilde{Z}} + \cdots + \tilde{\theta}_q\mu_{\tilde{Z}},$$

that is

$$\mathbb{E}\,\tilde{X}_t = \frac{\nu + \tilde{\theta}(1)\cdot\mathbb{E}\,\tilde{Z}_t}{\phi(1)}.$$

Remark 4.5 (Transformation[4]). The transformed time series $X_t := \frac{\tilde{X}_t - \delta_X}{\sigma_X}$ and $Z_t := \frac{\tilde{Z}_t - \delta_Z}{\sigma_Z}$ satisfy

$$\begin{aligned}\delta_X + \sigma_X X_t = {} & \phi_1(\delta_X + \sigma_X X_{t-1}) + \cdots + \phi_p(\delta_X + \sigma_X X_{t-p}) \\ & + \nu \\ & + \tilde{\theta}_0(\delta_Z + \sigma_Z Z_t) + \cdots + \tilde{\theta}_q(\delta_Z + \sigma_Z Z_{t-q}),\end{aligned}$$

or

$$\begin{aligned}X_t = {} & \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} \\ & + \frac{\nu}{\sigma_X} - \frac{\delta_X}{\sigma_X}(1 - \phi_1 - \cdots - \phi_p) + \frac{\delta_Z}{\sigma_X}(\tilde{\theta}_0 + \cdots + \tilde{\theta}_q) \\ & + \frac{\sigma_Z}{\sigma_X}(\tilde{\theta}_0 Z_t + \cdots + \tilde{\theta}_q Z_{t-q}),\end{aligned}$$

that is

$$\phi(B)X = \underbrace{\frac{\nu - \delta_X\phi(1) + \delta_Z\tilde{\theta}(1)}{\sigma_X}}_{=:c} + \underbrace{\frac{\sigma_Z}{\sigma_X}\tilde{\theta}(B)}_{=:\theta(B)}Z$$

with $\theta_i := \frac{\sigma_Z}{\sigma_X}\theta_i'$.

The special choices

- $\delta_Z := \mu_{\tilde{Z}} = \mathbb{E}\,\tilde{Z}_t$ and $\sigma_Z := \operatorname{var}\tilde{Z}_t$ to obtain $Z_t \sim (0, 1)$ (a standard white noise, cf. Definition 3.12),
- $\delta_X := \frac{\nu + \delta_Z\tilde{\theta}(1)}{\phi(1)}$ to have $c = 0$;
- $\sigma_X := \tilde{\theta}_0\,\sigma_Z$ to have $\theta_0 = 1$

reveal the standard ARMA(p, q) representation (4.1).

[4]In German also Z-*Transformation*

4.2 MOVING AVERAGE, MA

The moving average process MA(q) is a special ARMA process MA(q) = ARMA$(0, q)$ with $\phi(\cdot) = 1$ (i.e., $p = 0$, or $\phi_1 = \cdots = \phi_p = 0$ in (4.1)).

Proposition 4.6. *The covariance function of an MA(q) process is*

$$\operatorname{cov}(X_t, X_{t+\tau}) = \begin{cases} \sigma_Z^2 \cdot \sum_{j=0}^{q-|\tau|} \theta_j\, \theta_{j+|\tau|} & |\tau| \le q, \\ 0 & \tau > q. \end{cases} \tag{4.3}$$

Proof. For the expected value we have

$$\mathbb{E}\, X_t = \sum_{j=0}^{q} \theta_j\, \mathbb{E}\, Z_{t-j} = 0.$$

The covariance is

$$\operatorname{cov}\left(X_t, X_{t+\tau}\right) = \mathbb{E}\left(\sum_{j=0}^{q} \theta_j Z_{t-j}\right)\left(\sum_{k=0}^{q} \theta_k Z_{t+\tau-k}\right) = \sum_{j,k=0}^{q} \theta_j\, \theta_k\, \underbrace{\mathbb{E}\, Z_{t-j} Z_{t+\tau-k}}_{\sigma_Z^2 \cdot \delta_{j-k+\tau}}$$

from which the assertion is immediate. □

Remark 4.7. Estimating the MA parameters is a nontrivial task which can be accomplished by nonlinear curve fitting.

Remark 4.8. Note that the autocovariance function $\gamma(\tau)$ stops abruptly, as $\gamma(\tau) = 0$ for $\tau > q$.

Example 4.9. Cf. Example 3.16.

4.3 AUTOREGRESSIVE AR

The autoregressive process is the special AR(p) = ARMA$(p, 0)$ process (i.e., $q = 0$ or $\theta_1 = \cdots = \theta_q = 0$ in (4.1)),

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t. \tag{4.4}$$

Proposition 4.10 (Yule-Walker equations)**.** *The covariance function of an AR(p) process satisfies the recursive equations*

$$\gamma(0) = \sum_{j=1}^{p} \phi_j\, \gamma(j) + \sigma_Z^2, \qquad \text{for } \tau = 0, \tag{4.5}$$

$$\gamma(\tau) = \sum_{j=1}^{p} \phi_j\, \gamma(\tau - j) \qquad \text{for } \tau > 0. \tag{4.6}$$

i.e., $\gamma_0 = \sum_{j=1}^{p} \gamma_j \phi_j + \operatorname{var} Z$ *and (cf.* (3.5)*)*

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{p-1} & \dots & \gamma_1 & \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{pmatrix}. \tag{4.7}$$

Proof. With (4.4) we have

$$\begin{aligned} \gamma(\tau) = \operatorname{cov}\left(X_t, X_{t+\tau}\right) &= \mathbb{E}\, X_t \cdot \left(\sum_{j=1}^{p} \phi_j X_{t+\tau-j} + Z_{t+\tau} \right) \\ &= \sum_{j=1}^{p} \phi_j\, \gamma(\tau - j) + \operatorname{cov}\left(X_t, Z_{t+\tau}\right). \end{aligned}$$

Now note that X_t depends on $\dots, Z_{t-1}, Z_t$ and thus $\operatorname{cov}\left(Z_{t+\tau}, X_t\right) = \begin{cases} \operatorname{var} Z_t & \text{if } \tau = 0, \\ 0 & \text{if } \tau > 0. \end{cases}$
Hence the result. □

Example 4.11. Consider the AR(1) process $X_t = \phi_1 X_{t-1} + Z_t$. The equations (4.5)–(4.6) with $\phi = (\phi_1, 0, 0, \dots)^\top$ read

$$\begin{aligned} \gamma_0 &= \gamma_1 \cdot \phi_1 + \sigma_Z^2, \\ \gamma_1 &= \gamma_0 \cdot \phi_1, \\ \gamma_2 &= \gamma_1 \cdot \phi_1, \text{ etc.} \end{aligned} \tag{4.8}$$

It follows that $\gamma_\ell = \gamma_0 \cdot \phi_1^\ell$ and with (4.8) thus the general solution $\gamma_\ell = \frac{\sigma_Z^2\, \phi_1^{|\ell|}}{1-\phi_1^2}$.

Remark 4.12*.* Suppose that z is a root of the polynomial $\phi(\cdot)$, i.e., $\phi(z) = 0$. Then $1 = \sum_{j=1}^{p} \phi_j z^j$ or $z^{-\tau} = \sum_{j=1}^{p} \phi_j z^{-(\tau-j)}$, i.e., $\gamma(\ell) := z^{-\ell}$ solves (4.6). By linearity, the autocovariance function of an AR(p) process has the general form

$$\gamma(\tau) = \sum_{k=1}^{p} \frac{c_k}{z_k^{|\tau|}} \tag{4.9}$$

for some constants c_k, where z_k are the roots (zeros) of the polynomial $\phi(\cdot)$, $\phi(z_k) = 0$, $k = 1, \dots, p$. The constants c_k are determined by the initial conditions (4.6).

Proposition 4.13. *The general form of the autocovariance function is given by* (4.9)*.*

Remark 4.14*.* In contrast to the MA process, the autocovariance function $\gamma(\cdot)$ does not terminate abruptly (cf. Remark 4.8).

Remark 4.15*.* If X is an AR(p) process, then the autocorrelation is $\alpha(\ell) = 0$ for $\ell > p$. Table 4.1 outlines the behavior further. Notice also that $\alpha(p) = \phi_p$.

Remark 4.16*.* Generalized Yule–Walker equations

(i) The roots z_k determine decay of the covariance function. Note, that X_t cannot explode if $|z_k| > 1$ for all $k = 1, \dots, p$, i.e., $\phi(z) \neq 0$ for $|z| \leq 1$.

(ii) The roots z_k and thus the decay do not depend on the moving average operator, $\theta_1, \dots \theta_q$.

(iii) The constants c_k need to be determined by the initial conditions in (4.6).

Remark 4.17*.* If $\phi_1, \dots, \phi_p$ and $\sigma_Z^2 = \operatorname{var} Z_t$ are known (p+1 parameters), then $\gamma(0), \dots, \gamma(p)$ can be computed from (4.6). For $\tau > p$, the correlations can be computed recursively from (4.6).

Alternatively, if $\gamma(0), \dots, \gamma(p)$ are known or estimated, then (4.6) can be used to compute $\phi_1, \dots, \phi_p$ and $\operatorname{var} Z$.

Remark 4.18*.* The Yule–Walker equations provide a way to estimate the parameters $\phi_1, \dots, \phi_p$ by replacing $\gamma_0, \dots, \gamma_p$ by their estimates $\hat{\gamma}_0, \dots, \hat{\gamma}_p$.

4.4 STATIONARY ARMA PROCESSES

Proposition 4.19 (Linear transformation)**.** *Suppose that Y_t is stationary (but not necessarily iid.) and $\sum_{j\in\mathbb{Z}} |\psi_j| < \infty$. Then $X_t = \sum_{j\in\mathbb{Z}} \psi_j Y_{t-j}$ is well-defined, stationary and*

$$\gamma_X(\ell) = \sum_{j,k\in\mathbb{Z}} \psi_j\, \psi_k\, \gamma_Y(\ell - j + k).$$

Proof. The expectation is $\mathbb{E}\, X_t = \sum_{j\in\mathbb{Z}} \psi_j\, \mathbb{E}\, Y_{t-j} = \mu_Y \cdot \sum_j \psi_j < \infty$. For the autocovariance, we have that

$$\begin{aligned}
\gamma_X(\ell) &= \mathbb{E}\, X_{t+\ell} \cdot X_t - \mathbb{E}\, X_{t+\ell} \cdot \mathbb{E}\, X_t \\
&= \lim_{n\to\infty} \sum_{j=-n}^{n} \mathbb{E}\, \psi_j Y_{t+\ell-j} \sum_{k=-n}^{n} \psi_k Y_{t-k} - \sum_{j,k=-n}^{n} \psi_j \psi_k\, \mathbb{E}\, Y_{t+\ell-k}\, \mathbb{E}\, Y_{t-k} \\
&= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \left(\mathbb{E}\, Y_{t+\ell-j} \cdot Y_{t-k} - \mathbb{E}\, Y_{t+\ell-j} \cdot \mathbb{E}\, Y_{t-k}\right) \\
& \sum_{j,k=-\infty}^{\infty} \psi_j\, \psi_k\, \gamma_Y(\ell - j + k)
\end{aligned}$$

and thus the assertion. ◻

Definition 4.20 (Causal process)**.** The ARMA(p, q) process X_t is *causal* if there are constants ψ_j such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \psi_j \cdot Z_{t-j}. \tag{4.10}$$

As above, we shall also associate the function $\psi(z) := \sum_{j=0}^{\infty} z^j$ and write $X_t = \psi(B) Z_t$.

Example 4.21. Recall the process $X_t = \phi_1 X_{t-1} + Z_t$ from Example 4.11. It holds that

$$\begin{aligned} X_t &= Z_t + \phi_1 X_{t-1} \\ &= Z_t + \phi_1 Z_{t-1} + \phi_1^2 X_{t-2} \\ &\dots \\ &= \sum_{j=0}^{\infty} \phi_1^j Z_{t-j} \end{aligned}$$

so that this AR(1) can be seen as a MA(∞) process.

Theorem 4.22. *The covariance function of a causal time series X_t is*

$$\gamma_X(h) = \sigma_Z^2 \cdot \sum_{j=0}^{\infty} \psi_{j+|h|} \psi_j. \tag{4.11}$$

Proof. This is a consequence of (4.10) and (4.3). □

Definition 4.23. The function $G(z) := \psi(z) \cdot \psi\left(z^{-1}\right)$ is the covariance generating function.

Theorem 4.24. *It holds that $\sum_{h\in\mathbb{Z}} \gamma_X(h)\, z^h = \sigma_X^2\, G(z)$.*

Proof. Indeed,

$$\begin{aligned} G(z) = \psi\left(z^{-1}\right) \cdot \psi(z) &= \sum_{j=0} \sum_{k=0} \psi_j \psi_k x^{k-j} = \sum_{h\in\mathbb{Z}} z^h \sum_{k-j=h} \psi_j \psi_k \\ &= \sum_{h\in\mathbb{Z}} z^h \sum_{j=0} \psi_{j+h} \psi_j = \sum_{h\in\mathbb{Z}} z^h \gamma_X(h) \end{aligned}$$

with (4.11), hence the assertion. □

Theorem 4.25. *For a causal ARMA(p,q) process it holds that*

$$2\gamma(\tau) = \sum_{j=1}^{p} \gamma(\tau - j)\, \phi_j + \sigma_Z^2 \cdot \sum_{k=\tau}^{q} \theta_k\, \psi_{k-\tau} \qquad \text{for } \tau \le q, \tag{4.12}$$

$$\gamma(\tau) = \sum_{j=1}^{p} \gamma(\tau - j)\, \phi_j \qquad \text{for } \tau > q. \tag{4.13}$$

Proof. Indeed,

$$\begin{aligned} \gamma(\tau) = \operatorname{cov}\left(X_{t-\tau}, X_t\right) &= \mathbb{E}\, X_{t-\tau} \cdot \left(\sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{k=0}^{q} \theta_k Z_{t-k} \right) \\ &= \sum_{j=1}^{p} \phi_j\, \mathbb{E}\, X_{t-\tau} X_{t-j} + \sum_{k=0}^{q} \theta_k\, \mathbb{E}\, X_{t-\tau} Z_{t-k} \\ &= \sum_{j=1}^{p} \phi_j \gamma(\tau - j) + \sum_{k=0}^{q} \theta_k\, \mathbb{E}\, X_{t-\tau} Z_{t-k}. \end{aligned}$$

Now note that $X_{t-\tau}$ depends on $\dots, Z_{t-\tau-1}, Z_{t-\tau}$. Hence (4.13) for $\tau > q$.

Recall next the causal representation (4.10) so that further

$$\begin{aligned}\gamma(\tau) &= \sum_{j=1}^{p} \phi_j \gamma(\tau - j) + \sum_{k=0}^{q} \theta_k \sum_{j=0}^{\infty} \psi_j \, \mathbb{E}\, Z_{t-\tau-j} Z_{t-k} \\ &= \sum_{j=1}^{p} \phi_j \gamma(\tau - j) + \sigma_Z^2 \sum_{k=\tau}^{q} \theta_k \psi_{k-\tau}\end{aligned}$$

and thus (4.12). □

Remark 4.26. By (4.13), Remark 4.12 applies to ARMA(p,q) as well.

Theorem 4.27. *Let X_t be an ARMA(p,q) process (where $\theta(\cdot)$ and $\phi(\cdot)$ do not have common zeros).*

X_t is causal iff $\phi(z) \neq 0$ for $|z| \leq 1$. The coefficients are given by the generating function

$$\psi(z) := \frac{\theta(z)}{\phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j =: \psi(z). \tag{4.14}$$

Proof. It holds that $\phi(z) \neq 0$ for $|z| \leq 1$ and ϕ is a polynomial. There is hence $\varepsilon > 0$ so that $\xi(z) := \frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \xi_j z^j$ for $|z| < 1 + \varepsilon$. Consequently, $\xi_j \left(1 + \frac{\varepsilon}{2}\right)^j \xrightarrow[j\to\infty]{} 0$ and there exists $K > 0$ so that $|\xi_j| < \frac{K}{(1+\varepsilon/2)^j}$. In particular, $\sum_{j=0} |\xi_j| < \infty$ and $(\xi_j)_{j=0}^{\infty} \in \ell_1$. By Proposition 4.19 we may apply $\xi(B)$ to $\phi(B)X_t = \theta(B)Z_t$ and get $X_t = \underbrace{\xi(B)\theta(B)}_{\psi(B)} Z_t$ with ψ as in (4.14).

As for the contrary, assume that the ARMA(p,q) process X_t is causal, then $X_t = \sum_{j=0} \psi_j Z_{t-j}$ for some ψ_j with $\sum_{j=0}^{\infty} |\psi_j| < \infty$. It holds $\theta(B)Z_t = \phi(B)X_t = \underbrace{\phi(B)\psi(B)}_{\eta(B)} Z_t$ with $\eta(z) := \phi(z)\psi(z)$, which converges for $|z| \leq 1$, that is

$$\sum_{j=0}^{q} \theta_j Z_{t-j} = \sum_{j=0}^{\infty} \eta_j Z_{t-j}.$$

Take the inner product with Z_{t-k} on each side gives $\theta_k = \eta_k$ and thus $\theta(z) = \eta(z) = \phi(z)\psi(z)$ for $|z| \leq 1$. It follows that $\phi(z) \neq 0$ for $|z| \leq 1$, as $\theta(z)$ and $\phi(z)$ have no common zeros and as $\psi(z) < \infty$ for all $|z| \leq 1$. □

Corollary 4.28 (Corollary to Theorem 4.27 and Theorem 4.24)**.** *The covariance generating function of the general ARMA(p,q) is*

$$\sum_{h\in\mathbb{Z}} \gamma_X(h)\, z^h = \sigma_X^2 \cdot \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}.$$

	AR(p)	ARMA(p,q)	MA(q)
autocorrelation $\gamma(h)$	geometric decay	geometric after q	cuts off at q
partial autocorrelation $\alpha(h)$	cuts off at p	geometric after p	geometric decay

Table 4.1: Autocorrelation and partial autocorrelation

Definition 4.29. A ARMA(p,q) process X_t is invertible, if there are constants π_j so that

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \qquad t \in \mathbb{Z}.$$

Theorem 4.30. *X_t is invertible, iff $\theta(z) \neq 0$ for $|z| \leq 1$, cf. Theorem 4.27 with $\pi(z) = \sum_{j=0} \pi_j z^j = \frac{\phi(z)}{\psi(z)}$.*

Remark 4.31. Since an invertible moving average can be represented as infinite regression, the partial autocorrelations of a moving average process decay geometrically (cf. Table 4.1).

4.5 SEASONAL ARMA

These models are often given by

$$\phi_s(B^s)\phi(B)X_t = c + \theta_s(B^s)\theta(B)Z_t,$$

where the polynomials ϕ_s and θ_s model the seasonal components (cf. (4.2)).

4.6 ARMAX

ARMAX models have an additional exogenous variable,

$$\phi(B)X_t = c + \theta(B)Z_t + e(B)Y_t,$$

where Y_t is an exogenous time series.

4.7 ARIMA

A time series X_t is ARIMA(p,d,q) if $\Delta^d X_t$ is ARMA(p,q) (for the forward difference operator Δ see (2.6)).

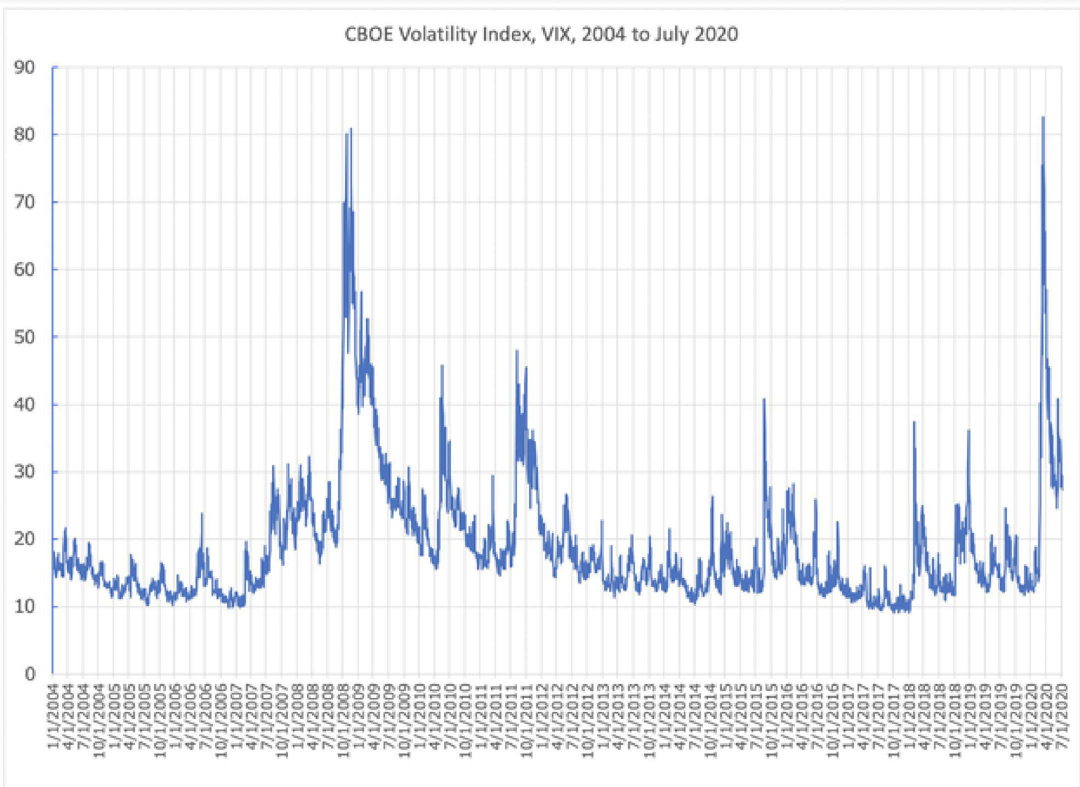


Figure 4.1: VIX, http://www.cboe.com or https://en.wikipedia.org/wiki/VIX

4.8 GARCH

ARCH (autoregressive conditional heteroscedasticity) models have been developed by Engle.[5] The ARCH(p) series satisfy the recursive equations

$$\begin{aligned} x_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_p x_{t-p}^2 \end{aligned}$$

with parameters $\alpha_1, \dots \alpha_p$.

GARCH(p,q) (generalized ARCH) follow the recursion

$$\begin{aligned} x_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_p x_{t-p}^2 \\ &\quad + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 \end{aligned}$$

with additional parameters $\beta_1, \dots \beta_q$.

Remark 4.32. Note, that $\gamma(\tau) = 0$ for $\tau > 0$.

4.9 VAR

The vector autoregression (VAR) is

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \tag{4.15}$$

where $\phi_0 \in \mathbb{R}^d$ and $\phi_j \in \mathbb{R}^{d\times d}$ $(j > 0)$. Further, the error is assumed to satisfy

(i) $\mathbb{E}\, Z_t = 0$,

(ii) $\mathbb{E}\, Z_t Z_t^\top = \Sigma$ and

(iii) $\mathbb{E}\, Z_t Z_{t-k}^\top = 0$.

4.10 MODEL SELECTION

Occam's razor.

Which parametric model should one choose to characterize a time series? ARMA(1,2) or ARMA(2,1)? Or is ARMA(3,0) a better choice? Will ARMA(3,1) be better compared to ARMA(2,1)?

To select a model among others, the following criteria can be employed.

In what follows, k is the number of parameters, L is the likelihood function and n is the number of observations.

[5]Robert F. Engle (1942), Nobel Memorial Price in Economic Sciences 2003

Figure 4.2: Robert Engle, 1942. Nobel Memorial Price 1942 in Economic Sciences

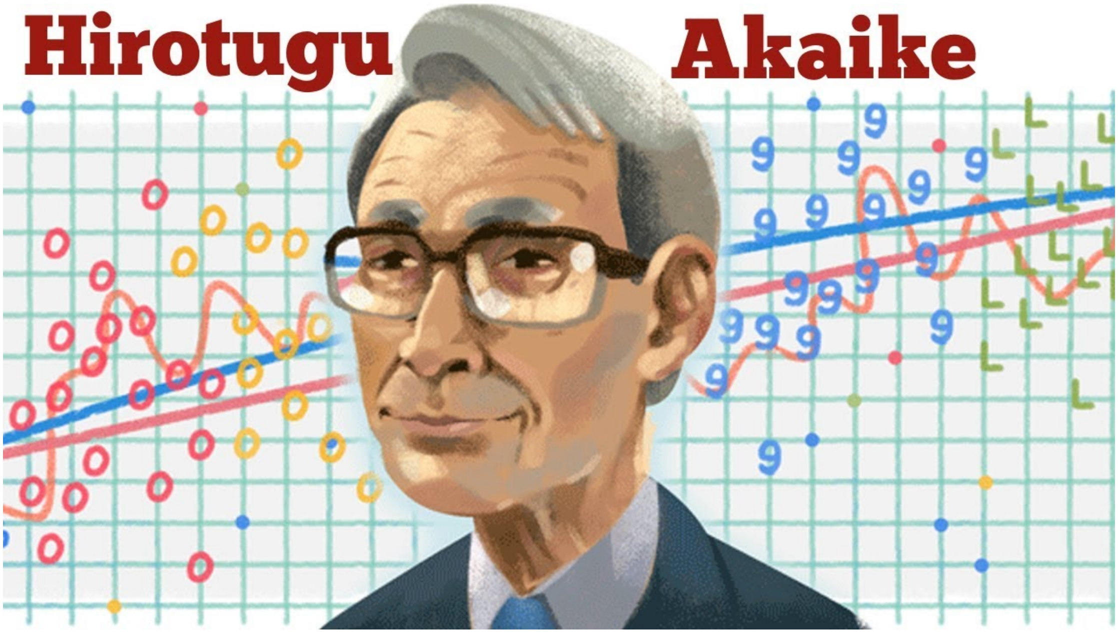


Figure 4.3: Hirotugu Akaike, 1927–2009, Japanese

4.10.1 Ordinary least squares

For the AR(p) model with $X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t$ we consider the regression model, where X_t is the endogenous variable, $X_{t-1}, \dots, X_{t-p}$ are the regressors and Z_t is the error term. In matrix representation (and notation),

$$\begin{pmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_T \end{pmatrix} = \begin{pmatrix} 1 & X_p & X_{p-1} & \dots & X_1 \\ 1 & X_{p+1} & X_p & \dots & X_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{T-1} & X_{T-2} & \dots & X_{T-p} \end{pmatrix} \begin{pmatrix} c \\ \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} + \begin{pmatrix} Z_{p+1} \\ Z_{p+2} \\ \vdots \\ Z_T \end{pmatrix},$$

i.e., $X = \mathbf{X}\beta + Z$.

The ordinary least squares estimator for $\beta = \left(c, \phi_1, \dots \phi_p\right)^\top$ is given by $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X} X$ (cf. the normal equations (2.3)). We may estimate σ^2 via the OLS residuals $\hat{\varepsilon} := X - \mathbf{X}\hat{\beta}$ by $\hat{\sigma}^2 = \frac{\hat{\varepsilon}^\top \hat{\varepsilon}}{T-p}$.

4.10.2 Maximum likelihood

4.10.3 Akaike information criterion

Hirotugu Akaike

AIC$(p,q) = \log \hat{\sigma}^2_{p,q} + (p+q)\frac{2}{T}$

Schwarz information criterion:

SIC$(p,q) = \log \hat{\sigma}^2_{p,q} + (p+q)\frac{\log T}{T}$
Hannan-Quinn information criterion:
SIC$(p,q) = \log \hat{\sigma}^2_{p,q} + (p+q)\frac{2\log\log T}{T}$

4.10.4 Bayesian information criterion

Bayesian information (BIC) criterion or Schwarz criterion is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function and it is closely related to the Akaike information criterion (AIC).

BIC $= k\ln n - 2\ln\hat{L}$, where n is the number of data points.

4.11 PROBLEMS

Exercise 4.1. *Which type of parametric process is the time series $X_t = Z_t - 2Z_{t-1} + Z_{t-2}$.*
- *Plot some paths,*
- *the autocorrelation and*
- *the partial autocorrelation function.*
- *Compare with theoretical results elaborated in this chapter.*

Exercise 4.2. *As Exercise 4.1, for the time series $X_t = 0.9X_{t-1} + Z_t$.*

Exercise 4.3. *As Exercise 4.1, for the time series $(1-\eta_1 B)(1-\eta_2 B)X_t = Z_t$ with*
- $\eta_1 = 1/2$, $\eta_2 = 1/5$,
- $\eta_1 = 90\%$, $\eta_2 = 50\%$,
- $\eta_1 = -90\%$, $\eta_2 = 50\%$ *and*
- $\eta_{1,1} = \frac{3}{8}\left(1 \pm i\sqrt{3}\right)$.

Exercise 4.4. *Show that the ARMA$(2,q)$ time series*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \sum_{i=1}^{q} \theta_i Z_{t-i}$$

(for $\phi_1, \phi_2 \in \mathbb{R}$) is stationary iff $\phi_2 \in (-1,1)$ and $\phi_1 \in (\phi_2 - 1,\, 1-\phi_2)$.

Exercise 4.5. *Consider the ARMA process*

$$X_t = X_{t-1} - \frac{1}{4}X_{t-2} + Z_t + Z_{t-1}$$

and show that $X_t = \sum_{k=0}^{\infty}(1+3k)\,2^{-k} Z_{t-k}$. Further, the autocovariance is $\gamma_k = 2^{-k}\left(\frac{32}{3} + 8k\right)$.

Exercise 4.6. *Consider $X_t = 90\% X_{t-1} + Z_t$. Show that the partial autocorrelation function is* $\alpha(t) = \begin{cases} 90\% & \text{for } t = \pm 1, \\ 0 & \text{else.} \end{cases}$

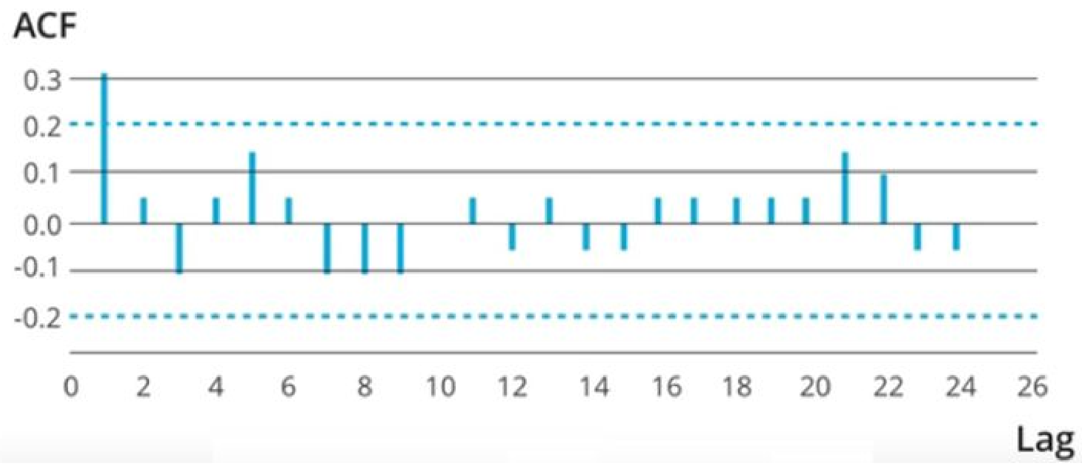


Figure 4.4: acf

Figure 4.5: acf and partial acf

Exercise 4.7. *Simulate some paths of a GARCH series.*

Exercise 4.8. *Simulate some paths of an ARIMA series.*

Exercise 4.9. *Describe and simulate some paths of an ARMAX series.*

Exercise 4.10 (From https://www.analyticsvidhya.com)**.** *Looking at the below ACF plot on Figure 4.4, would you suggest to apply AR or MA in ARIMA modeling technique?*

Exercise 4.11 (From https://www.analyticsvidhya.com)**.** *How many AR and MA terms should be included for the time series by looking at the above acf and pacf plots in Firgure 4.5?*

Estimators

5

Eternity is a very long time, especially towards the end.

Woody Allen, 1952

5.1 ESTIMATION OF MEAN AND VARIANCE

In what follows we shall assume that X_t is weakly stationary with mean μ and autocovariance function $\gamma(\cdot)$. Set

$$\hat{\mu}_n \coloneqq \overline{X}_n \coloneqq \frac{1}{n}\sum_{t=1}^{n} X_t.$$

Theorem 5.1. *Let X_t be weakly stationary.*

(i) It holds that $\mathbb{E}\,\overline{X}_n = \mu$ and

(ii) $\operatorname{var}\overline{X}_n = \frac{1}{n}\sum_{\ell=-n}^{n}\left(1-\frac{|\ell|}{n}\right)\gamma(\ell)$.

(iii) Suppose that $\gamma(\ell)\xrightarrow[\ell\to\infty]{}0$, then $\operatorname{var}\overline{X}_n \xrightarrow[n\to\infty]{} 0$.

(iv) Suppose that $\sum_{\ell=-\infty}^{\infty}|\gamma_\ell|<\infty$, then $n\cdot\operatorname{var}\overline{X}_n \xrightarrow[n\to\infty]{} \sum_{\ell=-\infty}^{\infty}\gamma_\ell$.

The estimator $\hat{\mu}_n \coloneqq \overline{X}_n$ is an unbiased and consistent estimator for μ.

Proof. As the time series X_t is weakly stationary it holds that $\mathbb{E}\,X_t = \mu$ and hence, by linearity, $\mathbb{E}\,\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\,X_t = \mu$.

For the variance we have

$$\begin{aligned}\operatorname{var}\overline{X}_n &= \mathbb{E}\,\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\cdot\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu) = \frac{1}{n^2}\sum_{i,j=1}^{n}\mathbb{E}(X_i-\mu)(X_j-\mu)\\ &= \frac{1}{n^2}\sum_{i,j=1}^{n}\gamma(i-j) = \frac{1}{n^2}\left(n\gamma(0)+2\sum_{\ell=1}^{n-1}(n-\ell)\gamma(\ell)\right) = \frac{1}{n}\sum_{\ell=-n}^{n}\left(1-\frac{|\ell|}{n}\right)\gamma(\ell).\end{aligned}$$

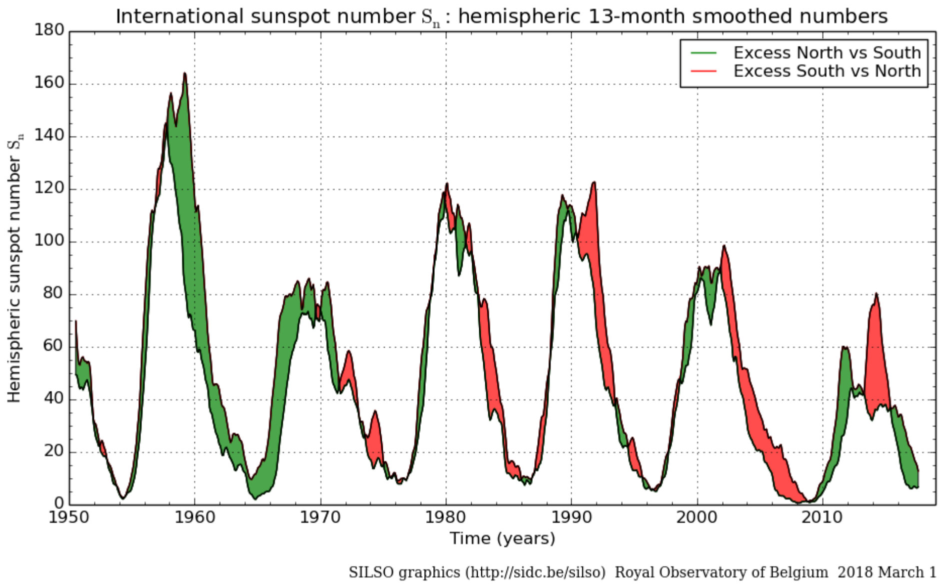


Figure 5.1: Sunspot numbers, http://www.sidc.be/silso/

As for (iii) choose $N \in \mathbb{N}$ large enough so that $|\gamma(n)| < \varepsilon$ for all $n \geq N$. Then

$$\operatorname{var} \overline{X}_n = \left| \frac{1}{n^2} \sum_{i,j=1}^{n} \gamma(i-j) \right| \leq \frac{1}{n^2} \sum_{i,j=1}^{n} |\gamma(i-j)|$$
$$\leq \frac{(2N+1)n\,\gamma(0) + (n-N)^2 \varepsilon}{n^2} \xrightarrow[n\to\infty]{} \varepsilon.$$

The assertion follows, as $\varepsilon > 0$ was chosen arbitrarily.

Finally, we have that

$$\lim_{n\to\infty} n \cdot \operatorname{var} \overline{X}_n = \lim_{n\to\infty} \sum_{\ell=-n}^{n} \left(1 - \frac{|\ell|}{n}\right) \gamma(\ell) = \sum_{\ell=-n}^{n} \gamma(\ell)$$

and thus the assertion (iv). □

Remark 5.2. It holds that $n \operatorname{var} \overline{X}_n \xrightarrow[n\to\infty]{} \sum_{\ell=-\infty}^{\infty} \gamma(\ell) = \sigma_X^2 \cdot \sum_{\ell=-\infty}^{\infty} \rho(\ell)$ and thus

$$\operatorname{var} \overline{X}_n \approx \frac{\sigma_X^2}{n/\tau},$$

where $\tau := \sum_{\ell=-\infty}^{\infty} \rho(\ell)$. The effect of the correlation (compared to the uncorrelated case) corresponds to a reduction of the sample size from n to n/τ.

Corollary 5.3. *It holds that*

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \sim \mathcal{N}\left(0, \sum_{\ell=-n}^{n} \left(1 - \frac{|\ell|}{n}\right) \gamma(\ell)\right). \tag{5.1}$$

5.2 ESTIMATION OF AUTOCOVARIANCE

Definition 5.4 (Sample autocovariance function, empirical autocovariance)**.** The *sample autocovariance function* for some data $X_1, \dots, X_n$ for $\ell \in \mathbb{Z}$ is

$$\hat{\gamma}_\ell := \frac{1}{n} \sum_{t=1}^{n-|\ell|} \left(X_{t+|\ell|} - \overline{X}_n\right)\left(X_t - \overline{X}_n\right). \tag{5.2}$$

The sample autocorrelation is $\hat{\rho}_\ell := \frac{\hat{\gamma}_\ell}{\hat{\gamma}_0}$.

Definition 5.5 (Sample partial autocorrelation)**.** The *sample partial autocorrelation* function of a stationary time series X_t is defined as the autocorrelation (see Definition 3.25), but based on the sample covariance $\hat{\rho}$ instead of the covariance ρ.

Remark 5.6 (Bessel Correction)*.* See Exercise 5.1 for the denominator n instead of $n - |\ell|$ or $n - |\ell| - 1$ in (5.2).

Remark 5.7*.* Note that $\overline{X}_n$ includes all samples $X_1, \dots, X_n$ although the first, nor the second factor in the product (5.2) involve all.

Proposition 5.8 (Non-negative definiteness)**.** *The matrix*

$$\hat{\Gamma}_n := \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \dots & \hat{\gamma}_{n-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\gamma}_1 \\ \hat{\gamma}_{n-1} & \dots & \hat{\gamma}_1 & \hat{\gamma}_0 \end{pmatrix}$$

is positive semi-definite. This is important for forecasting.

Proof. Define $M := \begin{pmatrix} 0 & 0 & \tilde{X}_1 & \tilde{X}_2 & \dots & \tilde{X}_n \\ 0 & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & 0 \\ \tilde{X}_1 & \tilde{X}_2 & \dots & \tilde{X}_n & 0 & 0 \end{pmatrix}$ with $\tilde{X}_i := X_i - \overline{X}_n$ and observe that $\hat{\Gamma}_n = \frac{1}{n} M M^\top$, thus

$$a^\top \hat{\Gamma}_n a = \frac{1}{n} a^\top M M^\top a = \frac{1}{n}\left(M^\top a\right)^\top M^\top a = \frac{1}{n}\left\|M^\top a\right\|^2 \ge 0$$

for every $a \in \mathbb{R}^n$ and thus the assertion. □

Theorem 5.9. *Let X_t be stationary and $\ell \in \mathbb{Z}$ be fixed. Then*

(i) $\mathbb{E}\,\hat{\gamma}_\ell \xrightarrow[n\to\infty]{} \gamma(\ell)$*, if* $\gamma(n) \xrightarrow[n\to\infty]{} 0$*, i.e.,* $\hat{\gamma}_\ell$ *is biased, but asymptotically consistent.*

(ii) $\operatorname{cov}\big(\tilde{\gamma}(k), \tilde{\gamma}(\ell)\big) = \frac{1}{n}\sum_{u=-n}^{n}\left(1 - \frac{|u|}{n}\right) V_u$*, where* $\mathbb{E}\, w_t^4 = \eta\sigma^4$ *and*[1] $V_u = \gamma(u)\gamma(u+k-\ell) + \gamma(u+k)\gamma(u-\ell) + (\eta-3)\sigma^4 \sum_{i\in\mathbb{Z}} \psi_{i+u+k}\psi_{i+k}\psi_{i+\ell}\psi_i$

[1]Bartlett's formula; Peter Bartlett, 1942

Lemma 5.10. *For any $\xi, \eta \in \mathbb{R}$ it holds that*

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \xi)(Y_i - \eta) - \frac{1}{n}\sum_{i=1}^{n}(X_i - \xi)\frac{1}{n}\sum_{j=1}^{n}(Y_j - \eta). \tag{5.3}$$

Proof. Indeed,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n) &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \xi - (\overline{X}_n - \xi)\right)\left(Y_i - \eta - (\overline{Y}_n - \eta)\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \xi\right)\left(Y_i - \eta\right) - \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \xi\right)\left(\overline{Y}_n - \eta\right)\\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\left(\overline{X}_n - \xi\right)\left(Y_i - \eta\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\overline{X}_n - \xi\right)\left(\overline{Y}_n - \eta\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \xi\right)\left(Y_i - \eta\right) - \left(\overline{X}_n - \xi\right)\left(\overline{Y}_n - \eta\right)\end{aligned}$$

and thus the assertion. □

Proof of (i). We replace $X_t \leftarrow X_{t+\ell}$, $Y_t \leftarrow X_t$ and $\xi = \eta = \mu$ in (5.3). Then

$$\hat{\gamma}_\ell = \frac{1}{n}\sum_{i=1}^{n-\ell}(X_{i+\ell} - \mu)(X_i - \mu) - \frac{1}{n}\sum_{i=1}^{n-\ell}(X_{i+\ell} - \mu)\cdot\frac{1}{n}\sum_{j=1}^{n-\ell}(X_j - \mu).$$

It follows that

$$\mathbb{E}\,\hat{\gamma}_\ell = \frac{1}{n}\sum_{i=1}^{n-\ell}\gamma(\ell) - \frac{1}{n^2}\sum_{i,j=1}^{n-\ell}\gamma(i+\ell-j) = \frac{n-\ell}{n}\gamma(\ell) - \frac{1}{n^2}\sum_{i,j=1}^{n-\ell}\gamma(i+\ell-j). \tag{5.4}$$

Again, let $N > \ell$ be large enough so that $|\gamma(n)| < \varepsilon$ for $n \geq N$. Then

$$\left|\frac{1}{n^2}\sum_{i,j=1}^{n-\ell}\gamma(i+\ell-j)\right| \leq \frac{1}{n^2}\sum_{i,j=1}^{n-\ell}|\gamma(i+\ell-j)| \leq \frac{(2N+1)n + \varepsilon n^2}{n^2} \xrightarrow[n\to\infty]{} \varepsilon.$$

The assertion follows from (5.4).

For the (rather messy) proof of (ii) we refer to Shumway and Stoffer (2000, (A.50)). □

5.3 PROBLEMS

Exercise 5.1. *It is occasionally proposed to scale (5.2) with $\frac{1}{n-\ell}$ instead of $\frac{1}{n}$. Then the matrix $\hat{\Gamma}_n$ is not positive semi-definite any longer. Give a counterexample.*

Exercise 5.2. *Verify* (5.1) *by simulations.*

Exercise 5.3. *Use Example 3.13 and investigate* (5.1) *by simulations.*

Exercise 5.4. *Give a histogram for* (5.2) *by simulation and compare with the result in Theorem 5.9 (i).*

Exercise 5.5. *The time series X_i in Exercise 3.3 has constant acf. Do the results of Theorem 5.1 still hold true? As well, investigate the results by simulations.*

Exercise 5.6. *Use the Levinson Algorithm (Proposition 3.23) to simulate a time series with $\gamma(\ell) \to 0$, but $\sum_{\ell\in\mathbb{Z}} \gamma(\ell) = \infty$. Investigate the results of Theorem 5.1 by simulations.*

Exercise 5.7 (Brockwell and Davis (1987, Problem 7.3))**.** *Show that the sample autocovariance $\hat{\gamma}$ of a time series $(x_1, \dots, x_n)$ satisfies $\sum_{\ell<n} \hat{\gamma}(\ell) = 0$.*

Fourier transform in sequence spaces 6

6.1 DEFINITIONS AND PROPERTIES

Definition 6.1. A series $(x_t)_{t\in\mathbb{Z}}$ is absolutely p-summable if $\|x\|_p := \left(\sum_{t\in\mathbb{Z}} |x_t|^p\right)^{1/p} < \infty$. We set $\ell_p(\mathbb{Z},\mathbb{C}) := \left\{(x_t)_{t\in\mathbb{Z}} : x_t \in \mathbb{C},\ \|x\|_p < \infty\right\}$.

Remark 6.2. The theory here can be developed for $t \in \mathbb{N}$, i.e., $\ell_1(\mathbb{N},\mathbb{R})$, as well.

Lemma 6.3. *For $x, y \in \ell_1$ it holds that*

- $x + y := (x_t + y_t)_{t=-\infty}^{\infty} \in \ell_1$ *and*
- $x \cdot y := (x_t \cdot y_t)_{t\in\mathbb{Z}} \in \ell_1$.

Proof. We have

(i) $\|x + y\|_1 \le \|x\|_1 + \|y\|_1 < \infty$ and
(ii) $\|x \cdot y\|_1 \le \sum_{t\in\mathbb{Z}} |x_t\, y_t| \le \sup_{t\in\mathbb{Z}} |x_t| \cdot \sum_{t\in\mathbb{Z}} |y_t| \le \|x\|_1 \cdot \|y\|_1 < \infty$ by Hölder's inequality.

□

Definition 6.4 (Fourier transform)**.** For $x \in \ell_1$, the function

$$\begin{aligned} \hat{x} \colon \mathbb{R} &\to \mathbb{C} \\ \nu \mapsto \hat{x}(\nu) &:= \sum_{t\in\mathbb{Z}} e^{-2\pi i\nu t} x_t \end{aligned}$$

is the Fourier transform of x, often also denoted by $F_x := \hat{x}$. The mapping

$$\begin{aligned} \mathcal{F} \colon \ell_1 &\to C(\mathbb{R},\mathbb{C}) \\ x \mapsto \mathcal{F}(x) &:= \hat{x}(\cdot) \colon \mathbb{R} \to \mathbb{C} \end{aligned}$$

is the *Fourier transform*. Note, that $\mathcal{F}$ maps sequences (ℓ_1) to functions ($C(\mathbb{R})$).

Remark 6.5. Note that $\hat{x}(\nu + 1) = \hat{x}(\nu)$. For this reason it is enough to restrict $\hat{x}$ to $[0, 1]$.

Definition 6.6 (Fourier cosine and sine transform)**.** The Fourier sine and cosine transform are

$$\mathcal{F}_c(x)(\nu) := \hat{x}^c(\nu) := \sum_{t\in\mathbb{Z}} x_t \cdot \cos(2\pi\nu t) \text{ and} \tag{6.1}$$

$$\mathcal{F}_s(x)(\nu) := \hat{x}^s(\nu) := \sum_{t\in\mathbb{Z}} x_t \cdot \sin(2\pi\nu t). \tag{6.2}$$

Remark 6.7. It follows from Euler's formula $e^{i\varphi} = \cos\varphi + i\sin\varphi$ that $\hat{x}(\cdot) = \mathcal{F}(x)(\cdot) = \hat{x}^c(\cdot) - i\,\hat{x}^s(\cdot)$.

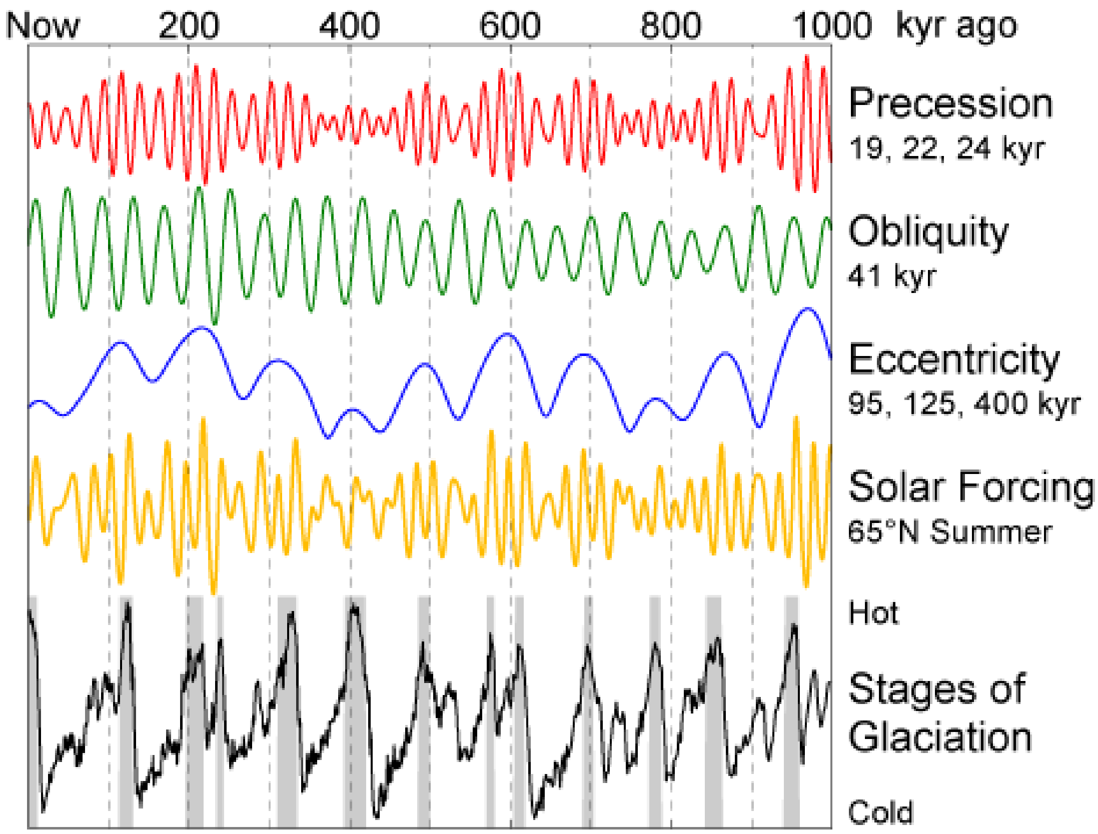


Figure 6.1: Milankovitch cycles, https://en.wikipedia.org/wiki/Milankovitch_cycles

Proposition 6.8. *The Fourier transform is well-defined and for all $x \in \ell_1$ and $\nu \in \mathbb{R}$ it holds that*

(i) $\hat{x}$ is uniformly bounded,

$$\|\hat{x}\|_\infty \le \|x\|_1 , \tag{6.3}$$

i.e., $|\hat{x}(\nu)| \le \|x\|_1$ for every $\nu \in \mathbb{R}$,

(ii) $\hat{x}(0) = \sum_{t\in\mathbb{Z}} x_t$,

(iii) $\hat{x}(\nu) = \hat{x}(\nu+1)$, i.e., the period is 1, and

(iv) $\hat{x}(-\nu) = \overline{\hat{x}(\nu)}$.

Further, the Fourier transform is linear, it holds that

(v) $\widehat{\alpha\, x + \beta\, y} = \alpha\,\hat{x} + \beta\,\hat{y}$.

Proof. Define the partial sum $F_n(\nu) := \sum_{t=-n}^n x_t\, e^{-2\pi i \nu t}$ and observe that for $m < n$,

$$|F_n(\nu) - F_m(\nu)| \le \sum_{m<|t|\le n} \left|x_t e^{-2\pi i\nu t}\right| \le \sum_{|t|>m} |x_t| \xrightarrow[m\to\infty]{} 0.$$

Note that convergence is uniform in $n > m$ and $\nu \in \mathbb{R}$. As $C(\mathbb{R})$ is closed under uniform limits it follows that the limit $F := \lim F_n$ is continuous, i.e., $F \in C(\mathbb{R})$. The remaining statements are obvious. □

Theorem 6.9. *For $x \in \ell_2(\mathbb{Z};\mathbb{C})$ it holds that*

$$\int_0^1 |\hat{x}(\nu)|^2 \,\mathrm{d}\nu = \sum_{t\in\mathbb{Z}} |x_t|^2 ,$$

i.e., $\|\hat{x}\|_{L^2([0,1])} = \|x\|_{\ell_2}$.

Proof. This is a consequence of the following more general statement. □

Theorem 6.10. *It holds that*

$$\int_0^1 \hat{x}(\nu)\overline{\hat{y}(\nu)}\,\mathrm{d}\nu = \sum_{t\in\mathbb{Z}} x_t \overline{y_t}.$$

Proof. Notice first the integral representation of Kronecker's delta,

$$\int_0^1 e^{2\pi i\nu(t-t')}\,\mathrm{d}\nu = \begin{cases} \int_0^1 1\,\mathrm{d}\nu & \text{if } t = t', \\ \frac{e^{2\pi i\nu(t-t')}}{2\pi i(t-t')}\Big|_{\nu=0}^1 & \text{if } t-t' \in \mathbb{Z}\setminus\{0\} \end{cases} = \begin{cases} 1 & \text{if } t = t', \\ 0 & \text{if } t-t' \in \mathbb{Z}\setminus\{0\} \end{cases} = \delta_{t,t'}. \tag{6.4}$$

Thus

$$\begin{aligned}\int_0^1 \hat{x}(\nu)\cdot\overline{\hat{y}(\nu)}\,\mathrm{d}\nu &= \int_0^1 \sum_{t\in\mathbb{Z}} e^{-2\pi i\nu t} x_t \cdot \sum_{t'\in\mathbb{Z}} e^{2\pi i\nu t'}\overline{y_{t'}}\,\mathrm{d}\nu \\ &= \sum_{t,t'\in\mathbb{Z}} x_t\cdot\overline{y_{t'}} \int_0^1 e^{2\pi i\nu(t'-t)}\,\mathrm{d}\nu = \sum_{t\in\mathbb{Z}} x_t\cdot\overline{y_t},\end{aligned}$$

the statement.

Parseval's theorem follows by choosing $x = y$. □

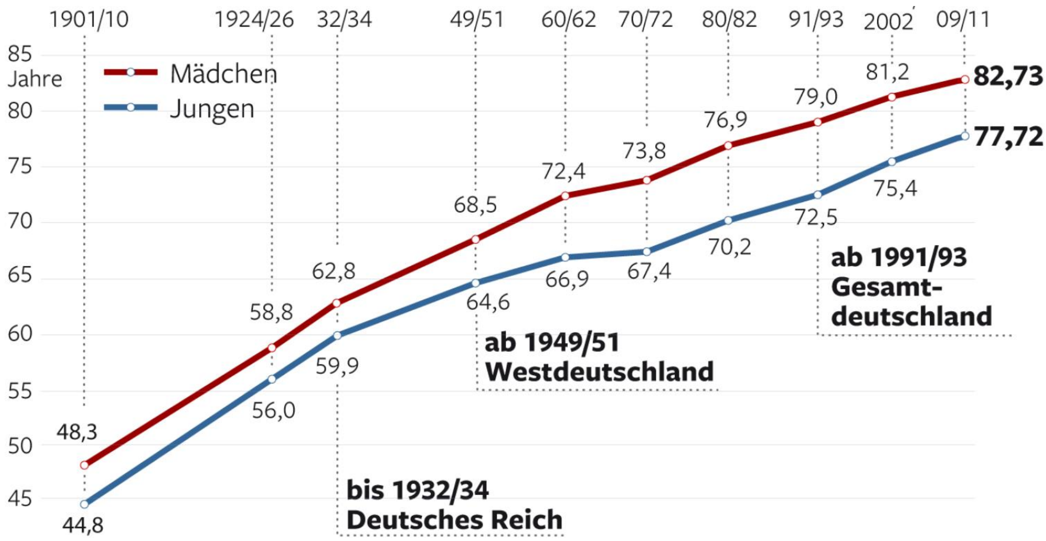


Figure 6.2: future lifetime, https://www.welt.de/article149577156/

Corollary 6.11. *It holds that*

$$\sum_{t\in\mathbb{Z}} |x_t|^2 = \int_0^1 \hat{x}^c(\nu)^2 + \hat{x}^s(\nu)^2\,\mathrm{d}\nu,$$

where $\hat{x}^c$ ($\hat{x}^s$, resp.) is the Fourier cosine (Fourier sine, resp.) transform, cf. (6.1)*.*

6.2 INVERSION

Proposition 6.12 (Inversion of the Fourier transform)**.** *For $x \in \ell_1$ it holds that*

$$x_t = \int_0^1 e^{2\pi i\nu t}\,\hat{x}(\nu)\,\mathrm{d}\nu, \qquad t\in\mathbb{Z}. \tag{6.5}$$

Remark 6.13*.* The inverse Fourier transform for $x \in L^1$ is occasionally denoted $\check{x}_t = \int_0^1 e^{2\pi i\nu t} x(\nu)\,\mathrm{d}\nu$.

Proof. Recall that $\hat{x}(\nu) = \sum_{t'\in\mathbb{Z}} x_{t'} e^{-2\pi i\nu t'}$, hence

$$\int_0^1 \hat{x}(\nu) e^{2\pi i\nu t}\,\mathrm{d}\nu = \int_0^1 \sum_{t'\in\mathbb{Z}} x_{t'} e^{-2\pi i\nu t'}\cdot e^{2\pi i\nu t}\,\mathrm{d}\nu = \sum_{t'\in\mathbb{Z}} x_{t'} \int_0^1 e^{2\pi i\nu(t-t')}\,\mathrm{d}\nu = x_t,$$

where we have use (6.4). Thus the result. ◻

Corollary 6.14. *It holds that $\int_0^1 \hat{x}(\nu)\,\mathrm{d}\nu = x_0$ and $\hat{x}(0) = \sum_{t\in\mathbb{Z}} x_t$.*

6.3 CONVOLUTION

Definition 6.15. The convolution of $x, y \in \ell_1$ is the sequence

$$x * y := \left(\sum_{\tau \in \mathbb{Z}} x_{t-\tau} \cdot y_\tau \right)_{t \in \mathbb{Z}}. \tag{6.6}$$

Remark 6.16. Note, that $(x * y)_t = \sum_{\tau \in \mathbb{Z}} x_{t-\tau} \cdot y_\tau = \sum_{\tau \in \mathbb{Z}} x_\tau \cdot y_{t-\tau}$.

Lemma 6.17. *For $x, y \in \ell_1$ it holds that*

$$\|x * y\|_1 \le \|x\|_1 \cdot \|y\|_1$$

and thus $x * y \in \ell_1$.

Proof. By the triangular inequality, $\|x * y\|_1 \le \sum_{t \in \mathbb{Z}} \sum_{\tau \in \mathbb{Z}} |x_{t-\tau} y_\tau| = \sum_{t \in \mathbb{Z}} \sum_{\tau \in \mathbb{Z}} |x_t||y_\tau| = \|x\|_1 \cdot \|y\|_1 < \infty$. □

Proposition 6.18 (Convolution theorem)**.** *It holds that*

(i) $\widehat{x * y} = \hat{x} \cdot \hat{y}$, *i.e.,* $\widehat{x * y}(\nu) = \hat{x}(\nu) \cdot \hat{y}(\nu)$ *and*

(ii) $\widehat{x \cdot y} = \hat{x} * \hat{y}$, *i.e.,* $\widehat{x \cdot y}(\nu) = (\hat{x} * \hat{y})(\nu)$, *where* $(f * g)(\nu) := \int_0^1 f(\nu') g(\nu - \nu')\, \mathrm{d}\nu'$ *(cf. (6.6)) is the convolution of the functions $f, g \in L^2$.*

Proof. It holds that

$$\begin{aligned}
\widehat{x * y}(\nu) &= \sum_{t \in \mathbb{Z}} (x * y)_t \cdot e^{-2\pi i \nu t} = \sum_{t \in \mathbb{Z}} \sum_{\tau \in \mathbb{Z}} x_\tau y_{t-\tau} e^{-2\pi i \nu t} \\
&= \sum_{\tau \in \mathbb{Z}} \sum_{t \in \mathbb{Z}} x_\tau y_t e^{-2\pi i \nu (t+\tau)} \\
&= \sum_{\tau \in \mathbb{Z}} x_\tau e^{-2\pi i \nu \tau} \cdot \sum_{t \in \mathbb{Z}} y_t e^{-2\pi i \nu t} = \hat{x}(\nu) \cdot \hat{y}(\nu).
\end{aligned}$$

Further,

$$\begin{aligned}
\widehat{x \cdot y}(\nu) &= \sum_{t \in \mathbb{Z}} x_t \cdot y_t\, e^{-2\pi i \nu t} \underset{(6.5)}{=} \sum_{t \in \mathbb{Z}} \int_0^1 \hat{x}(\nu')\, e^{2\pi i \nu' t}\, \mathrm{d}\nu' \cdot y_t\, e^{-2\pi i \nu t} \\
&= \int_0^1 \hat{x}(\nu') \cdot \sum_{t \in \mathbb{Z}} e^{-2\pi i (\nu - \nu') t} \cdot y_t\, \mathrm{d}\nu' \\
&= \int_0^1 \hat{x}(\nu') \hat{y}(\nu - \nu')\, \mathrm{d}\nu' = \big(\hat{x} * \hat{y}\big)(\nu).
\end{aligned} \tag{6.7}$$

The integral and sum in (6.7) can be interchanged by the monotone convergence theorem (Lebesgue's theorem) as the integrand is uniformly bounded by

$$\left|\hat{x}(\nu') \cdot \sum_{|t|\leq n} e^{-2\pi i(\nu-\nu')t} \cdot y_t\right| = \left|\hat{x}(\nu') \cdot \sum_{t\leq n} e^{-2\pi i(\nu-\nu')t} \cdot y_t\right|$$
$$\leq \|\hat{x}\|_\infty \, \|y\|_1 \leq \|x\|_1 \, \|y\|_1 \,,$$

by (6.3). □

7 Spectral analysis

Spectral analysis is the analysis of the time series in the frequency domain.

Definition 7.1. The *temporal frequency* f, the *period* T and the *angular frequency* ω are related by $\omega = 2\pi f$ and $f = {}^1\!/_T$. Tabular 7.1 compares temporal and spatial frequency terms.

Remark 7.2. For an *amplitude* A and a *phase shift* φ we have from the angle addition theorems that[1]

$$A \cdot \sin\left(\frac{2\pi t}{T} + \varphi\right) = A_s \cdot \cos\left(\frac{2\pi t}{T}\right) + A_c \cdot \sin\left(\frac{2\pi t}{T}\right),$$

where

$$A_c := A \cdot \cos\varphi \text{ and } A_s := A \cdot \sin\varphi;$$

note as well the inverse relation

$$A = \sqrt{A_s^2 + A_c^2} \text{ and } \tan\varphi = \frac{A_s}{A_c}$$

and consequently

$$\begin{aligned}&\text{span}\,\{t \mapsto A\sin(\omega t + \varphi) \colon A \in \mathbb{R}, \varphi \in [0, 2\pi)\}\\ &\quad = \text{span}\,\{t \mapsto A_c \sin(\omega t),\ t \mapsto A_s \sin(\omega t) \colon A_c, A_s \in \mathbb{R}\}\end{aligned}$$

7.1 SPECTRAL DENSITY

Remark 7.3. The process (3.3) is random, but Exercise 7.4 demonstrates that X_t is perfectly predictable from its past (deterministic).

[1]Cf. Footnote 1 (page 25)

	temporal	spatial	SI unit
period	T period	λ wavelength	m
linear frequency	$f = 1/T = \nu$	$\xi = \nu = 1/\lambda$ (wavenumber, repetency)	hertz= s^{-1}
angular frequency	$\omega = 2\pi f$	$k = 2\pi\xi$ (angular wavenumber, Kreiszahl)	radiant/ s
speed	$c = \lambda f$		m/s

Table 7.1: Frequencies

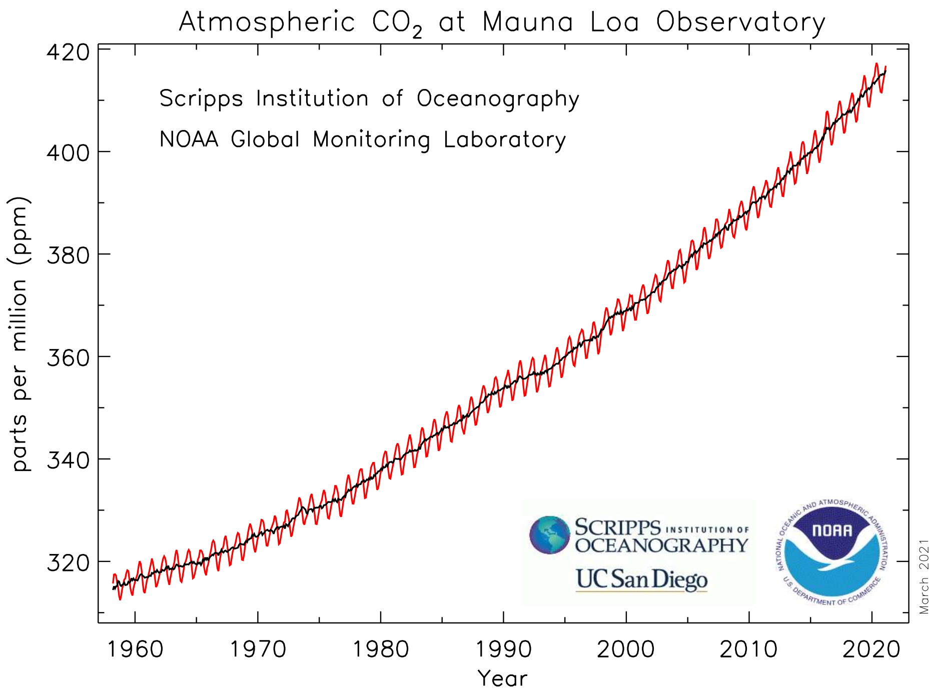


Figure 7.1: Keeling Curve, CO_2 at Mauna Loa, https://www.esrl.noaa.gov/gmd/ccgg/trends/; see also https://www.youtube.com/watch?v=gbxEsG8g6BA

Example 7.4 (Cf. Example 3.15)**.** Consider a time series

$$X_t = \sum_{j=1} A_j \cos 2\pi\nu_j t + B_j \sin 2\pi\nu_j t \tag{7.1}$$

with zero mean, uncorrelated A_j, B_j and $\operatorname{var} A_j = \operatorname{var} B_j = \sigma_j^2$, i.e., $A_j\, B_j \sim (0, \sigma_j^2)$. Then

$$\gamma(\tau) = \sum_{j=1} \sigma_j^2 \cos 2\pi\nu_j\tau. \tag{7.2}$$

Note, that the frequencies ν_j are explicit frequencies in the autocovariance function $\gamma(\cdot)$.

Define the measure

$$\mu(\cdot) \coloneqq \sum_{j=1} \frac{\sigma_j^2}{2}\left(\delta_{\nu_j}(\cdot) + \delta_{1-\nu_j}(\cdot)\right)$$

then, by (7.2),

$$\int_0^1 e^{2\pi i\tau\nu}\mu(\mathrm{d}\nu) = \sum_{j=1} \frac{\sigma_j^2}{2}\left(e^{2\pi i\tau\nu_j} + e^{2\pi i\tau(1-\nu_j)}\right) = \sum_{j=1} \sigma_j^2 \cos 2\pi\nu_j\tau = \gamma(\tau)$$

for $\tau \in \mathbb{Z}$.

The density of $\mu(\cdot)$ is the spectral density.

Definition 7.5 (Spectral density)**.** A function f is the *spectral density* of a stationary time series X_t with autocovariance function $\gamma(\cdot)$ if

(i) $f(\nu) \geq 0$ for all $\nu \in \mathbb{R}$ and

(ii) $\gamma(\tau) = \int_0^1 e^{2\pi i\tau\nu} f(\nu)\,\mathrm{d}\nu$ for all integers $\tau \in \mathbb{Z}$.

Remark 7.6. The inversion of the Fourier transform (6.12) suggests the notation $f(\cdot) = \hat{\gamma}(\cdot)$: this should not be mixed with the sample autocovariance, also denoted by $\hat{\gamma}$. A distinction is always clear by the differing argument: we write $\hat{\gamma}_\ell$ for the acf depending on the lag $\ell \in \mathbb{Z}$ (cf. (5.2)), but $\hat{\gamma}(\nu)$ for the spectral density depending on a frequency $\nu \in [0, 1]$.

Suppose that X_t is a zero mean stationary time series with autocovariance function $\gamma(\cdot)$ satisfying $\sum_{\ell\in\mathbb{Z}} |\gamma(\tau)| < \infty$. From (6.12) (the inversion of the Fourier transform), the *spectral density* of the time series is the Fourier transform of the autocovariance function,

$$f(\nu) = \hat{\gamma}(\nu) = \sum_{\tau\in\mathbb{Z}} e^{-2\pi i\nu\tau}\gamma(\tau), \qquad \nu \in \mathbb{R}. \tag{7.3}$$

Occasionally, the spectral density is $\frac{1}{2\pi}\sum_{\tau\in\mathbb{Z}} e^{-i\nu\tau}\gamma(\tau)$ instead of (7.3)

Example 7.7 (White noise)**.** Let $f(\nu) = \sigma^2$ be constant. Then, by (6.4),

$$\gamma(\tau) = \begin{cases} \sigma^2 & \text{if } \tau = 0, \\ 0 & \text{else} \end{cases} \tag{7.4}$$

and thus $\hat{\gamma}(\nu) = \sigma^2$ is constant. This is the spectral density of the white noise process $X_t = \sigma^2 \varepsilon_t$ for some iid, zero mean and variance 1 error ε.

Note, that the fact that $\hat{\gamma}(\cdot) = \text{constant}$ explains the term *white noise*.

Remark 7.8. Recall that $\gamma(\cdot)$ is even, i.e., $\gamma(\tau) = \gamma(-\tau)$. Hence, by (7.3),

$$\begin{aligned} f(\nu) = \hat{\gamma}(\nu) &= \gamma(0) + 2 \sum_{\tau=1}^{\infty} \gamma(\tau) \cos(2\pi\nu\tau) \\ &= \sum_{\tau \in \mathbb{Z}} \gamma(\tau) \cos(2\pi\nu\tau), \qquad \nu \in \mathbb{R}, \end{aligned}$$

is even as well.

Remark 7.9. Recall from Theorem 5.1 (iv) that

$$n \operatorname{var} \overline{X}_n \xrightarrow[n\to\infty]{} \sum_{\ell=-\infty}^{\infty} \gamma_\ell = \hat{\gamma}(0).$$

Proposition 7.10 (Properties of the spectral density)**.** *It holds that*

- *(i)* $\hat{\gamma}(\cdot)$ *is even, i.e.,* $\hat{\gamma}(\nu) = \hat{\gamma}(-\nu)$ *with period* 1, $\hat{\gamma}(\cdot + 1) = \hat{\gamma}(\cdot)$,
- *(ii)* $\hat{\gamma}(\nu) \ge 0$ *for all* $\nu \in \mathbb{R}$ *and*
- *(iii) for* $\tau \in \mathbb{Z}$,

$$\gamma(\tau) = \int_0^1 e^{2\pi i \tau \nu} \cdot \hat{\gamma}(\nu)\, \mathrm{d}\nu = \int_0^1 \cos(2\pi\tau\nu) \cdot \hat{\gamma}(\nu)\, \mathrm{d}\nu. \tag{7.5}$$

Remark 7.11. It follows from (7.5) that $\operatorname{var} X_t = \gamma(0) = \int_0^1 \hat{\gamma}(\nu)\, \mathrm{d}\nu$. The spectral density $\hat{\gamma}(\cdot)$ restricted to $[0, 1]$ (or $[-1/2, 1/2]$) thus is indeed a density up to scaling by $\operatorname{var} X_t$. Replacing the autocovariance by the autocorrelation in (7.3) removes this gap.

Proof. (i) is obvious from the definition and (7.5) follows from Proposition 6.12. To see (ii) define

$$\begin{aligned} f_n(\nu) &\coloneqq \frac{1}{n} \mathbb{E} \left| \sum_{t=1}^{n} X_t\, e^{-2\pi i t \nu} \right|^2 = \frac{1}{n} \mathbb{E} \sum_{t=1}^{n} X_t\, e^{-2\pi i t \nu} \sum_{s=0}^{n} X_s\, e^{2\pi i s \nu} \\ &= \frac{1}{n} \sum_{s,t=1}^{n} e^{-2\pi i \nu (t-s)}\, \gamma(s-t) = \sum_{\ell=0}^{n-1} \frac{n - |\ell|}{n} e^{-2\pi i \nu \ell}\, \gamma(\ell). \end{aligned} \tag{7.6}$$

The assertion follows with $n \to \infty$ as $f_n(\cdot) \ge 0$ and $f_n(\nu) \to \hat{\gamma}(\nu)$. ◻

Remark 7.12 (Periodic time series). Consider the spectral density $\hat{\gamma}(\nu) = \sum_{j=1} \sigma_j^2 \, \delta_{\nu_j}(\nu)$. Using the property (7.5) we obtain that

$$\gamma(\tau) = \sum_{j=1} \sigma_j^2 \cos 2\pi\tau\nu_j = (7.2).$$

However, $\hat{\gamma}(\cdot)$ is a distribution and not a classical function and so the periodic time series (7.1) does *not* have a spectral density.

However, define $F_{\hat{\gamma}}(\nu) := \sum_{j=1} F_j(\nu)$ with $F_j(\nu) := \begin{cases} 0 & \text{if } \nu < \nu_j, \\ \sigma_j^2/2 & \text{if } \nu_j \le \nu < 1-\nu_j, \\ \sigma_j^2 & \text{else, i.e., } 1-\nu_j \le \nu. \end{cases}$ Then $\gamma(\tau) = \int_0^1 e^{2\pi i\tau\nu}\, \mathrm{d}F_{\hat{\gamma}}(\nu)$.

Definition 7.13. The representation

$$\gamma(\tau) = \int_0^1 e^{2\pi i\tau\nu}\, \mathrm{d}F(\nu) \tag{7.7}$$

is the *spectral representation* of the autocovariance function $\gamma(\cdot)$. The integrand $F(\cdot)$ is the *spectral distribution function*.

If $F(\nu) = \int_0^\nu \hat{\gamma}(\nu')\, \mathrm{d}\nu'$, then $\hat{\gamma}$ is the spectral density.

Definition 7.14. The time series has a continuous spectrum, if it has a spectral density, and a discrete spectrum otherwise.

Theorem 7.15. *A function $\gamma\colon \mathbb{Z} \to \mathbb{R}$ is an autocovariance function, iff it can be written in the form (7.7) for some nondecreasing function $F(\cdot)$.*

Proof. (cf. Brockwell and Davis (1987)) We show first that γ is nonnegative if it has the representation (7.7). Indeed,

$$\begin{aligned}\sum_{s,t=1}^n a_s\gamma(s-t)a_t &= \sum_{s,t=1}^n a_s \int_0^1 e^{2\pi i(s-t)\nu}\, \mathrm{d}F(\nu)a_t \\ &= \int_0^1 \left|\sum_{s,t=1}^n a_s e^{2\pi i s\nu}\right|^2 \mathrm{d}F(\nu) \ge 0.\end{aligned}$$

Conversely, if γ is nonnegative definite, then $f_n(\nu) := \frac{1}{n}\sum_{s,t=1}^n e^{-2\pi i s\nu}\gamma(s-t)e^{2\pi i t\nu}$ and $F_n(\nu) := \int_0^\nu f_n(\nu')\, \mathrm{d}\nu'$ is a (generalized) cdf, which is nondecreasing, as $f_n(\nu) \ge 0$, as $\gamma(\cdot)$ is nonnegative. We have

$$\begin{aligned}\int_0^1 e^{2\pi i\tau\nu}\, \mathrm{d}F_n(\nu) &= \int_0^1 e^{2\pi i\tau\nu}\frac{1}{n}\sum_{s,t=1}^n e^{-2\pi i s\nu}\gamma(s-t)e^{2\pi i t\nu}\, \mathrm{d}\nu \\ &= \int_0^1 e^{2\pi i\tau\nu}\sum_{|k|\le n}\left(1-\frac{|k|}{n}\right)\gamma(k)e^{-2\pi i k\nu}\, \mathrm{d}\nu \qquad (s-t=k) \\ &= \begin{cases}\left(1-\frac{|\tau|}{n}\right)\gamma(\tau) & \text{if } \tau \le n \\ 0 & \text{else.}\end{cases}\end{aligned}$$

The assertion follows from Helly's selection theorem by letting $n \to \infty$ (note that $F_n(1) = \int_0^1 f_n(\nu)\,\mathrm{d}\nu = \gamma(0) < \infty$). □

7.2 THE SPECTRUM OF AN ARMA PROCESS

Theorem 7.16 (Linear transformation)**.** *Suppose that X_t is a covariance stationary process with acf γ_X and $\sum_{j\in\mathbb{Z}} |\gamma_X(j)| < \infty$. Define $Y_t := \sum_{j=0}^\infty \psi_j X_{t-j}$ with $\sum_{j=0}^\infty \psi_j^2 < \infty$. Then Y_t is covariance stationary with spectral density*

$$f_Y(\nu) = \left|\sum_{j=0} \psi_j\, e^{-2\pi i\nu j}\right|^2 \cdot f_X(\nu),$$

where f_X (f_Y, resp.) is the spectral density of X (Y, resp.).

Proof. Recall that (Proposition 4.19)

$$\begin{aligned}\gamma_Y(h) &= \operatorname{cov}(Y_t, Y_{t-h})\\ &= \sum_{j=0}\sum_{k=0} \psi_j\psi_k \operatorname{cov}\left(X_{t-j}, X_{t-h-k}\right)\\ &= \sum_{j=0}\sum_{k=0} \psi_j\psi_k\gamma_X(h+k-j).\end{aligned}$$

Next,

$$\begin{aligned}f_Y(\nu) &= \sum_{h\in\mathbb{Z}} e^{-2\pi i\nu h}\gamma_Y(h)\\ &= \sum_{h\in\mathbb{Z}} e^{-2\pi i\nu h} \sum_{j=0}\sum_{k=0} \psi_j\,\psi_k\,\gamma_X(h+k-j)\\ &= \sum_{j=0} \psi_j\, e^{-2\pi i\nu j} \sum_{k=0} \psi_k\, e^{2\pi i\nu k} \sum_{h\in\mathbb{Z}} e^{-2\pi i\nu(h+k-j)}\,\gamma_X(h+k-j)\\ &= \left(\sum_{j=0} \psi_j\, e^{-2\pi i\nu j}\right)\left(\sum_{k=0} \psi_k\, e^{2\pi i\nu k}\right) f_X(\nu)\\ &= \left|\sum_{j=0} \psi_j\, e^{-2\pi i\nu j}\right|^2 f_X(\nu),\end{aligned}$$

the assertion. □

Remark 7.17 (AR(∞) spectral density). The spectral density of an AR(∞) time series $X_t = \psi(B)W_t$ is (cf. (7.4))

$$f_X(\nu) = \sigma_w^2 \left|\psi\left(e^{-2\pi i\nu}\right)\right|^2.$$

Corollary 7.18 (ARMA spectral density)**.** *The spectral density of an ARMA time series* $\phi(B)X_t = \theta(B)W_t$ *is (for* $\psi(\cdot) = \frac{\theta(\cdot)}{\phi(\cdot)}$ *see* (4.14)*)*

$$f_X(\nu) = \sigma_w^2 \cdot \left|\psi\left(e^{-2\pi i\nu}\right)\right|^2 = \sigma_w^2 \cdot \left|\frac{\theta\left(e^{-2\pi i\nu}\right)}{\phi\left(e^{-2\pi i\nu}\right)}\right|^2. \tag{7.8}$$

Definition 7.19. The spectrum (7.8) is called a *rational sprectrum*.

Remark 7.20*.* By (7.8), the spectrum of an invertible process (cf. Theorem 4.30) is

$$f^{\text{inverse}}(\nu) = \frac{\sigma_w^4}{f_X(\nu)},$$

which explains (again, finally) the name inverse process.

7.3 DISCRETE FOURIER TRANSFORM

Definition 7.21. For $x, y \in \mathbb{R}^n$ we shall write $\langle y, x\rangle := \sum_{i=1}^n \overline{y_i} x_i$. We set

$$e_k := \frac{1}{\sqrt{n}}\begin{pmatrix} e^{2\pi i k\cdot 0/n} \\ e^{2\pi i k\cdot 1/n} \\ \vdots \\ e^{2\pi i k\cdot (n-1)/n}\end{pmatrix}, \quad k = 1, \dots, n$$

(these are *not* the unit vectors).

Remark 7.22*.* The vectors $e_k = e_{k+n}$ are orthonormal, i.e.,

$$\begin{aligned}\langle e_k, e_\ell\rangle = \overline{e}_k^\top e_\ell &= \frac{1}{n}\sum_{j=0}^{n-1} \overline{e^{2\pi i k\cdot j/n}} e^{2\pi i \ell\cdot j/n} \\ &= \frac{1}{n}\sum_{j=0}^{n-1} e^{2\pi i j(\ell-k)/n} = \begin{cases} 1 & \text{if } k = \ell, \\ \frac{e^{2\pi i n(\ell-k)/n}-1}{e^{2\pi i(\ell-k)/n}-1} = 0 & \text{else} \end{cases} = \delta_{k,\ell}.\end{aligned}$$

It follows that

$$X = \sum_{k=1}^n \langle e_k, X\rangle \cdot e_k = \sum_{k=0}^{n-1} \hat{X}_k \cdot e_k$$

for every $X \in \mathbb{C}^n$, where $\hat{X}_k := \langle e_k, X\rangle = \sum_{j=1}^n e^{-j\cdot 2\pi i k/n} X_j$.

Proposition 7.23 (Parseval)**.** *It holds that*

$$\|X\|^2 = \sum_{k=1}^n |\langle e_k, X\rangle|^2, \text{ i.e., } \sum_{i=1}^n X_i^2 = \sum_{k=0}^{n-1} \hat{X}_k^2. \tag{7.9}$$

Proof. Indeed,

$$\begin{aligned}\|X\|^2 &= \left\langle \sum_{k=1}^n \langle e_k, X\rangle \cdot e_k, \sum_{\ell=1}^n \langle e_\ell, X\rangle \cdot e_\ell \right\rangle \\ &= \sum_{k,\ell=1}^n \overline{\langle e_k, X\rangle} \langle e_\ell, X\rangle \langle e_k, e_\ell\rangle = \sum_{k=1}^n |\langle e_k, X\rangle|^2 = \sum_{k=1}^n \left|\hat{X}_k\right|^2 = \left\|\hat{X}\right\|^2,\end{aligned}$$

the assertion. □

7.4 PERIODOGRAM

In this section we shall assume that the time series is mean adjusted, i.e., $\overline{X}_n = \frac{1}{n}\sum_{t=1}^n X_t = 0$. We are interested in an estimator for the spectral density $\hat{\gamma}(\cdot)$ (cf. (7.3)).

Definition 7.24. The preriodogram[2] of the sample $X_1, \dots, X_n$ is the function (cf. (7.6))

$$I_n(\nu) := \frac{1}{n}\left|\sum_{t=1}^n e^{-2\pi i t\nu} X_t\right|^2 \tag{7.10}$$

Remark 7.25*.* Note, that $I_n({}^k\!/\!{}_n) = |\langle e_k, X\rangle|^2$ and thus $\|X\|^2 = \sum_{k=1}^n I_n({}^k\!/\!{}_n)$ by (7.9).

Remark 7.26 (Discrete Fourier sine and cosine transform)*.* It holds that

$$\begin{aligned}\hat{X}_k = \langle e_k, X\rangle &= \frac{1}{\sqrt{n}}\sum_{j=1}^n e^{-2\pi i j\cdot k/n} X_j \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^n X_j \cos\frac{2\pi i j k}{n} - i\frac{1}{\sqrt{n}}\sum_{j=1}^n X_j \sin\frac{2\pi i j k}{n} \\ &=: \hat{X}_k^c - i\hat{X}_k^s.\end{aligned}$$

More generally,

$$I_n(\nu) = \left(\frac{1}{\sqrt{n}}\sum_{t=1}^n X_t \cos 2\pi t\nu\right)^2 + \left(\frac{1}{\sqrt{n}}\sum_{t=1}^n X_t \sin 2\pi t\nu\right)^2.$$

Proposition 7.27. *For $k \neq 0$ it holds that*

$$I_n({}^k\!/\!{}_n) = \sum_{|\tau|<n} \hat{\gamma}_X(\tau) e^{-2\pi i k\tau/n}, \tag{7.11}$$

where $\hat{\gamma}_X$ is the sample autocovariance function (5.2) *(not to be confused with the Fourier transform $\hat{\gamma}$ here).*

[2] Stichprobenspektrum, Periodogramm, Germ.

Corollary 7.28 (Proposition 7.27 for $k = 0$)**.** *For a mean adjusted time series it holds that* $I_n(k/n) = \sum_{|\tau|<n} \hat{\gamma}(\tau) e^{-2\pi i k\tau/n}$ *for all* $k \in \{-n, \dots, n\}$*, i.e.,* including $k = 0$.

Proof. Expanding (7.10) gives $I_n(\nu) = \frac{1}{n}\sum_{s,t=1}^n e^{-2\pi i(t-s)\nu} X_s X_t$. Note that

$$\frac{1}{n}\sum_{s,t=1}^n e^{-2\pi i(t-s)k/n} = \frac{1}{n}\sum_{s=1}^n e^{2\pi i s k/n} \cdot \sum_{t=1}^n e^{-2\pi i t k/n} = 0$$

provided that $k \neq 0$. Hence

$$\begin{aligned} I_n(k/n) &= \frac{1}{n}\sum_{s,t=1}^n e^{-2\pi i(t-s)k/n}\left(X_s - \overline{X}_n\right)\left(X_t - \overline{X}_n\right) \\ &= \sum_{\tau<n} e^{-2\pi i\tau k/n} \frac{1}{n}\sum_{t-s=\tau}\left(X_{t-\tau} - \overline{X}_n\right)\left(X_t - \overline{X}_n\right) \\ &= \sum_{|\tau|<n} e^{-2\pi i\tau k/n}\hat{\gamma}(\tau), \end{aligned}$$

the result. □

Fact. *Although Proposition 7.27 suggests that (replace* $k/n \leftarrow \nu$*)*

$$I_n(\nu) \xrightarrow[n\to\infty]{} \hat{\gamma}(\nu) = \sum_{\tau\in\mathbb{Z}} e^{-2\pi i\nu\tau}\gamma(\tau),$$

the periodogram (7.11) *is* not *a* consistent estimator *of the spectral density* $\hat{\gamma}$*.*

Example 7.29. Figure 1.4b displays the periodogram of the nottem data, which exhibit the monthly frequency with $f = \frac{1}{12} \approx 0,0833$.

7.5 DIFFICULTIES IN READING THE PERIODOGRAM

7.5.1 Leakage

The periodogram I_n is continuous for n finite. Hence, frequencies close to a peak frequency ν_0 are too high (leakage[3]). When increasing the length of the time series, then the peak frequencies get sharper. The resolution, in general, is approximately $1/n$ (where n is the length of the time series observed).

7.5.2 Aliasing

Consider the time series

$$X_t := \sin(2\pi f t + \varphi) \text{ and } \tilde{X}_t := -\sin(2\pi(k-f)t - \varphi).$$

[3] Durchsickern, Germ.

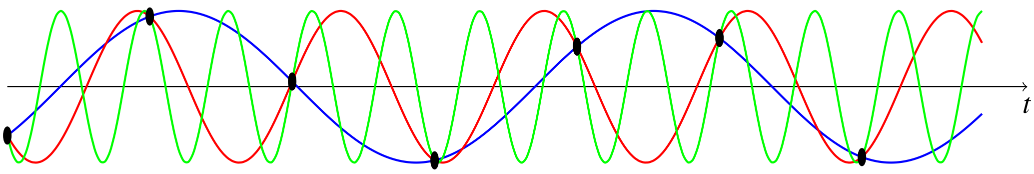


Figure 7.2: What is the true frequency for the points observed?

Note, that $\tilde{X}_t = \sin(2\pi f t + \varphi - 2\pi k) = X_t$ for all $t \in \mathbb{Z}$! However, their true frequencies (which are f and $1 - f$) differ; they cannot be detected (aliasing[4]).

Further, note that

$$I_n(\nu) = I_n(k + \nu) = I_n(k - \nu)$$

for every $k \in \mathbb{Z}$. A peak at ν in the peridogram indicates a frequency in $\{k + \nu, k - \nu \colon k \in \mathbb{Z}\}$.

A higher sampling frequency is necessary to decide on the true frequency.

Example 7.30. Table 7.2 gives different periods for a peak frequency at $\nu = 0.11$.

$0.083 \hat{=} 12.0$	$1.08 \hat{=} 0.92$	$2.08 \hat{=} 0.48$	$3.08 \hat{=} 0.32$	$\dots$
	$0.92 \hat{=} 1.09$	$1.92 \hat{=} 0.52$	$2.92 \hat{=} 0.34$	$\dots$

Table 7.2: Aliasing. A peak at $\nu = 0.11$ may indicate different periods

Definition 7.31. The largest frequency, which can be detected in a signal, is called *Nyquist frequency.*[5] For time series, the Nyquist frequency is $\nu_{\text{Nyquist}} = \frac{1}{2}$ (i.e., the period 2, see Figure 1.4b).

7.5.3 Overtones

The time series ($k \in \mathbb{Z}$)

$$X_t = \sin(2\pi k f t + \varphi)$$

has frequency kf (period $\frac{1}{k\nu}$), but f (period $\frac{1}{f}$) is a valid frequency too (overtones[6]).

7.6 PROBLEMS

Exercise 7.1. *Consider the time series* $X_{i+1} = \rho_i \overline{X}_i + \sqrt{1 - \rho_i \rho}\, Y_{i+1}$.

Exercise 7.2 (AR(1))**.** *Consider the process* $X_t = \phi_1 X_{t-1} + Z_t$ *with* $\operatorname{var} Z_t = \sigma^2$*. Show that*

$$\gamma(\tau) = \frac{\sigma^2 \phi_1^{|\tau|}}{1 - \phi_1^2} \text{ and } \hat{\gamma}(\nu) = \frac{\sigma^2}{1 - 2\phi_1 \cos 2\pi\nu + \phi_1^2}.$$

[4]Maskierung, Germ.
[5]Harry Nyquist, 1889–1976, Swedish engineer
[6]Oberschwingungen, Germ.

Plot trajectories of the time series for $\phi_1 = 0.9$ *and* $\phi_1 = -0.9$ *and the spectral density. Discuss the properties for various signs of* ϕ_1*:*

$\phi_1 > 0$*, positive autocorrelation, spectrum is dominated by low frequency components—smooth in time domain;*

$\phi_1 < 0$*, negaitve autocorrelation, spectrum is dominated by high frequency components—rough in time domain.*

Exercise 7.3 (MA(1))**.** *Consider the process* $X_t = Z_t + \theta_1 Z_{t-1}$*. Recall, that*

$$\gamma(\tau) = \begin{cases} \sigma^2(1-\theta_1^2) & \text{if } \tau = 0, \\ \sigma^2\theta_1 & \text{if } \tau = 1, \\ 0 & \text{else} \end{cases} \quad \text{and} \quad \hat{\gamma}(\nu) = \sigma^2\left(1 + \theta_1^2 + 2\theta_1 \cos 2\pi\nu\right).$$

Plot trajectories of the time series for $\theta_1 = 0.9$ *and* $\theta_1 = -0.9$ *and the spectral density. Discuss the properties for various signs of* θ_1*:*

$\theta_1 > 0$*, positive autocorrelation, spectrum is dominated by low frequency components—smooth in time domain;*

$\theta_1 < 0$*, negaitve autocorrelation, spectrum is dominated by high frequency components—rough in time domain.*

Exercise 7.4. *Show that the time series* (3.3) *is perfectly predictable, it holds that* $X_t = 2\cos(2\pi\nu_0) \cdot X_{t-1} - X_{t-2}$.

Exercise 7.5. *Give the recursion for* $X_t = e^{-\beta t}\left(A\cos(2\pi\nu_0 t) + B\sin(2\pi\nu_0 t)\right)$*, similarly to Exercise 7.4.*

8 Singular spectrum analysis, SSA

See https://en.wikipedia.org/wiki/Singular_spectrum_analysis,
Zhigljavsky, Anatoly
earth temperature: http://earth-temperature.com
Caterpillar-SSA: http://www.gistatgroup.com/
Forecasting Hyndman: https://www.otexts.org/fpp

9 Wold decomposition

Definition 9.1 (Linear process)**.** The time series X_t is a linear process if

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \text{ and } \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \tag{9.1}$$

where Z_t is a white noise (cf. Definition 3.12).

Proposition 9.2 (Cf. Proposition 4.19)**.** *The autocovariance function of the linear process is*

$$\gamma(\ell) = \sigma_Z^2 \cdot \sum_{j=-\infty}^{\infty} \psi_{j+\ell} \cdot \psi_j. \tag{9.2}$$

Definition 9.3 (Cf. Definition 4.20)**.** A linear process is *causal* if $\psi_j = 0$ for every $j < 0$ in the representation (9.1).

Proposition 9.4 (Cf. Theorem 4.22)**.** *The autocovariance function of the causal linear process is*

$$\gamma(\ell) = \sigma_Z^2 \cdot \sum_{j=0}^{\infty} \psi_{j+\ell} \psi_j.$$

Suppose that X_t is stationary. Then $Z_t := X_t - \mathbb{E}(X_t \mid X_{t-1}, X_{t-2}, \dots)$ is a white noise with variance $\sigma^2 := \mathbb{E} Z_t^2 = \mathbb{E} X_t Z_t$ and $\mathbb{E} X_t Z_u = 0$ whenever $t < u$.

Proof. For $t < u$ it holds that $\mathbb{E}(Z_t \cdot X_u \mid X_{u-1}, \dots) = Z_t \cdot \mathbb{E}(X_u \mid X_{u-1}, \dots)$. Hence

$$\begin{aligned}
\mathbb{E} Z_t Z_u &= \mathbb{E} Z_t \cdot (X_u - \mathbb{E}(X_u \mid X_{u-1}, \dots)) \\
&= \mathbb{E} Z_t X_u - \mathbb{E} Z_t \cdot \mathbb{E}(X_u \mid X_{u-1}, \dots) \\
&= \mathbb{E} Z_t X_u - \mathbb{E}\,\mathbb{E}(Z_t \cdot X_u \mid X_{u-1}, \dots) \\
&= \mathbb{E} Z_t X_u - \mathbb{E} Z_t X_u = 0.
\end{aligned}$$

Further note that the distribution of Z_t does not depend on t and hence $\sigma^2 := \operatorname{var} Z_t$ is well-defined, the variance of the white noise. To see the assertion $\mathbb{E} X_t Z_u = 0$ replace Z_t by X_t in the latter display.

Finally

$$\begin{aligned}
\mathbb{E} X_t Z_t - \mathbb{E} Z_t Z_t &= \mathbb{E}(X_t - Z_t) \cdot Z_t \\
&= \mathbb{E}\left[\mathbb{E}(X_t \mid X_{t-1}, \dots) \cdot \left(X_t - \mathbb{E}(X_t \mid X_{t-1}, \dots)\right)\right] = 0
\end{aligned}$$

by the projection property of the conditional expectation. □

Theorem 9.5. *Every covariance-stationary time series X_t has the representation*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + \eta_t, \tag{9.3}$$

where

(i) *Z_t is a white noise with variance σ_Z^2,*

(ii) *$\psi_0 = 1$ and $\sum_{j=1}^{\infty} |\psi_j|^2 < \infty$ and*

(iii) *η_t is deterministic, or perfectly predictable from its past, i.e., $\mathbb{E}\, \eta_t Z_s = 0$ for all (sic!) $s, t \in \mathbb{Z}$.*

Remark 9.6. See Exercise 7.4 below for a perfectly predictable process.

Proof. We demonstrate the statement only for stationary processes. Define

$$Z_t := X_t - \mathbb{E}\left(X_t \mid X_{t-1}, X_{t-2}, \dots\right)$$

We have seen in Proposition 9.4 that Z_t is a white noise and we may set $\sigma_Z^2 := \operatorname{var} Z_t$. Now we may set

$$\psi_j := \frac{1}{\sigma_Z^2} \mathbb{E}\, X_t Z_{t-j}$$

and

$$\eta_t := X_t - \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

The coefficient ψ_j is well-defined, as the time series is stationary.

Note that $\frac{1}{\sigma_Z} Z_t$ is an orthonormal subset of L^2 and by Bessel's inequality thus $\infty > \|X_t\|^2 \ge \sum_{j=0}^{\infty} \left|\left\langle \frac{Z_{t-j}}{\sigma_Z}, X_t \right\rangle\right|^2 = \sum_{j=0}^{\infty} |\psi_j|^2$. Further, by Proposition 9.4,

$$\psi_0 = \frac{\mathbb{E}\, X_t Z_t}{\mathbb{E}\, Z_t^2} = 1 \tag{9.4}$$

and thus (ii). As Z_j are orthogonal we have Proposition 9.4 that

$$\mathbb{E}\left(X_t \mid Z_j \colon j \in \mathbb{Z}\right) = \sum_{j \in \mathbb{Z}} \frac{Z_j}{\sigma_Z} \mathbb{E}\, \frac{Z_j}{\sigma_Z} X_t = \sum_{j=-\infty}^{t} Z_j \,\mathbb{E}\, \frac{Z_j}{\sigma_Z^2} X_t = \sum_{j=0}^{\infty} Z_{t-j} \,\mathbb{E}\, \frac{Z_{t-j}}{\sigma_Z^2} X_t = \sum_{j=0}^{\infty} Z_{t-j} \psi_j.$$

Finally note that

$$\begin{aligned} X_t &= X_t - \mathbb{E}\left(X_t \mid Z_j \colon j \in \mathbb{Z}\right) + \mathbb{E}\left(X_t \mid Z_j \colon j \in \mathbb{Z}\right) \\ &= X_t - \sum_{j=0}^{\infty} \psi_j Z_{t-j} + \sum_{j=0}^{\infty} \psi_j Z_{t-j} = \eta_t + \sum_{j=0}^{\infty} \psi_j Z_{t-j}. \end{aligned} \tag{9.5}$$

Finally note that $\mathbb{E}\, \eta_t Z_u = 0$ whenever $u > t$ by Proposition 9.4. Then we have $\mathbb{E}\, \eta_t Z_t = 0$ by (9.4) and for $u < t$ we get the result from (9.5). □

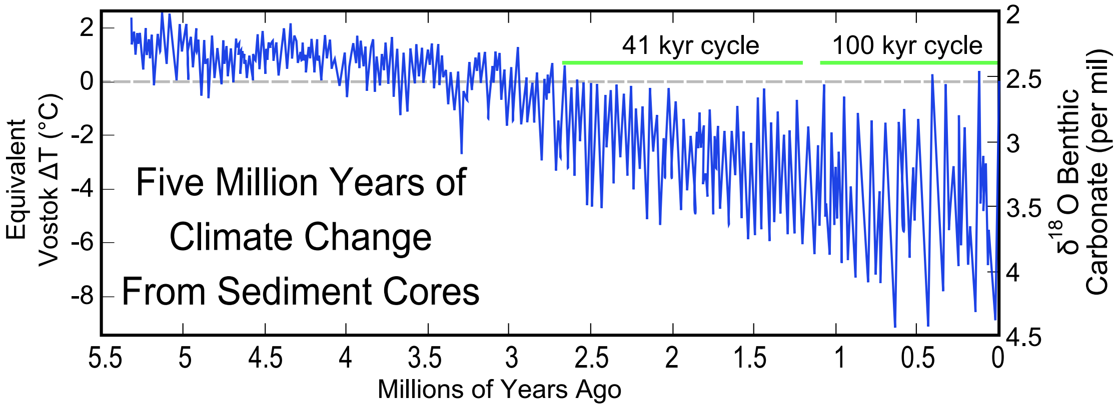


Figure 9.1: Climate history. Source: https://en.wikipedia.org/wiki/Geologic_temperature_record

Remark 9.7 (Properties). The following hold true for the Wold decomposition

(i) $\mathbb{E}\, X_t = \eta_t$, from (9.3);

(ii) $\operatorname{cov}(X_t, X_{t+\ell}) = \gamma(\ell) = \sigma_Z^2 \cdot \sum_{j=0}^{\infty} \psi_{j+\ell}\, \psi_j$ from (9.2) and in particular

(iii) $\operatorname{var} X_t = \gamma(0) = \sigma_Z^2 \cdot \sum_{j=0}^{\infty} \psi_{j+\ell}^2$.

10 Nonparametric forecasts

10.1 THE COMPOSITION METHOD

Suppose a random variable X has a density function of the particular form $f_X(\cdot) = \sum_{i=1} p_i f_i(\cdot)$, where $p_i \ge 0$ and $\sum_{i=1} p_i = 1$. To get a sample of X with density $f_X(\cdot)$ one may, first, sample a random i^* with $P(i^* = i) = p_i$ (for example, sample a uniform $U \in [0,1]$ and find i^* such that $\sum_{i=1}^{i^*-1} p_i \le U \le \sum_{i=1}^{i^*} p_i$); second, get a sample X from $f_{i^*}(\cdot)$. The variable X then has density $f_X(\cdot)$. In symbols, $X_i \sim f_i(\cdot)$ and $X_{i^*} \sim f_X(\cdot)$.

Example 10.1. The usual kernel density estimator for $f_X(x)$ based on observations X_i, $i = 1, \dots, n$, is $\hat{f}(x) := \sum_{i=1}^n \frac{1}{n} k_h(x - X_i)$. Here, the weights are simply $p_i = \frac{1}{n}$ and $f_i(x) = k_h(x - X_i)$, where $k_h(x) := \frac{1}{h} k\left(\frac{x}{h}\right)$ is the scaled kernel. Samples from $f_i(\cdot) = k_h(\cdot - X_i)$ are $X_i + h \cdot K$, where K is a sample based on the (unscaled) kernel with density $k(\cdot)$. In symbols, $K \sim k(\cdot)$, $X_i + hK \sim f_i(\cdot)$ and $X_{i^*} \sim \hat{f}(\cdot)$.

Example 10.2 (Conditional density $f(\cdot|y)$ for y fixed)**.** The density estimator for $f(x|y)$ based on observations (X_i, Y_i), $i = 1, \dots, n$, is $\hat{f}(x|y) = \sum_{i=1}^n \underbrace{\frac{k_h(y - Y_i)}{\sum_{j=1}^n k_h(y - Y_j)}}_{p_i(y)} \cdot k_h(x - X_i)$.

Here, the weights are $p_i(y) = \frac{k_h(y-Y_i)}{\sum_{j=1}^n k_h(y-Y_j)}$ and the functions $f_i(\cdot) = k_h(\cdot - X_i)$ are as above. Samples from $f_i(\cdot)$, in particular, are $X_i + h \cdot K$ (as above). In symbols, $K \sim k(\cdot)$, $X_i \sim f_i(\cdot)$ and $X_{i^*} \sim \hat{f}(\cdot|y)$.

Example 10.3 (Markovian time series)**.** Suppose the transition probability of a discrete-time Markovian time series has a density, $P(X_{t+1} \in \mathrm{d}x | X_t = y) = f(x|y)\,\mathrm{d}x$. A typical observation for such models is a trajectory $(X_0, X_1, X_2, \dots, X_n)$ and every X_{t+1} is a realization based (conditioned) on the previous observation $y = X_t$ with density $f(\cdot|X_t)$.

To estimate the transition density $f(x|y)$ based on the previous Example 10.2 we consider the paired observations (X_i, X_{i-1}), $i = 1, \dots, n$, i.e., we set $Y_i := X_{i-1}$. This gives the explicit estimator

$$\hat{f}(x \mid y) = \sum_{i=2}^n \underbrace{\frac{k_h(y - X_{i-1})}{\sum_{j=2}^n k_h(y - X_{j-1})}}_{p_i(y)} \cdot k_h(x - X_i) \tag{10.1}$$

for $f(x|y)$. The estimator $\hat{f}(x|y)$ is based on the observed trajectory $(X_0, X_1, X_2, \dots, X_n)$.

To sample a new time series $(x_0, x_1, x_2 \dots, x_t, x_{t+1}, \dots)$ based on the observation $(X_0, X_1, X_2, \dots, X_n)$ we pick an (arbitrary, but reasonable) start value x_0. Next, generate

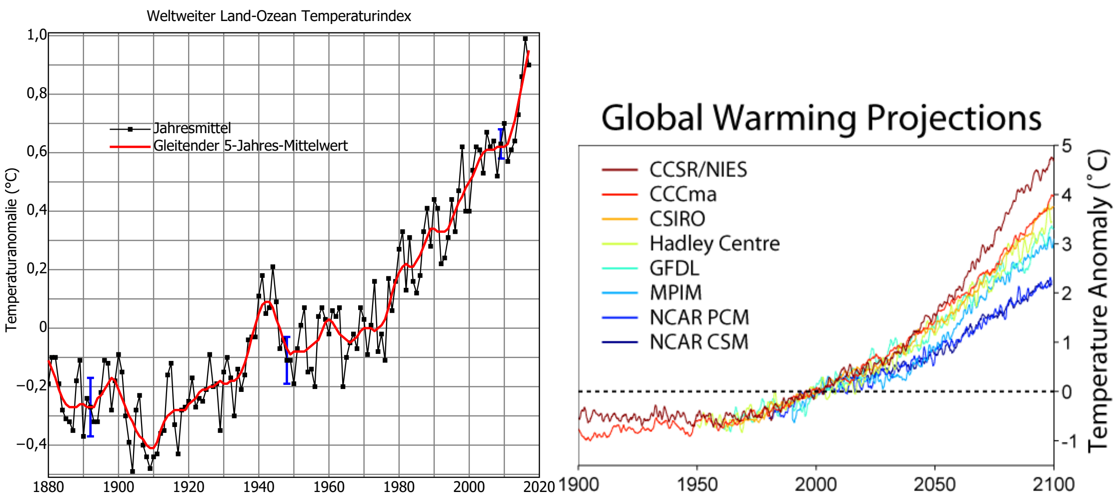


Figure 10.1: Global warming precition, https://en.wikipedia.org/wiki/Global_warming

a sample x_1 with $x_1 \sim \hat{f}(\cdot|x_0)$ by setting $y = x_0$ in (10.1) and by applying the procedure described in Example 10.2 with $p_i(y) = \frac{k_h(y-X_{i-1})}{\sum_{j=1}^n k_h(y-X_{j-1})}$.

In general, suppose the new series generated is $(x_0, x_1, \dots x_t)$. The series is continued by generating $x_{t+1} \sim \hat{f}(\cdot|x_t)$, where $y = x_t$ in (10.1) (x_t is the previously generated sample, i.e., the last entry in the new series). Once x_{t+1} is found, we may restart with $(x_0, x_1, \dots x_t, x_{t+1})$, etc.

Example 10.4 (Time series with fixed lag $\ell \in \mathbb{N}$)**.** Here, the distribution of the next x_{t+1} depends on the historic ℓ values $x_{t-\ell+1}, \dots, x_t$, i.e., $x_{t+1} \sim f(\cdot \mid x_{t-\ell+1}, \dots x_t)$. To estimate the density as above we may employ the density estimator

$$\hat{f}(\cdot \mid y_{-\ell}, \dots y_{-1}) := \sum_{i=\ell+1}^n \underbrace{\frac{k_h(y_{-\ell} - X_{i-\ell}) \cdot \dots k_h(y_{-1} - X_{i-1})}{\sum_{j=\ell+1}^n k_h(y_{-\ell} - X_{j-\ell}) \cdot \dots k_h(y_{-1} - X_{j-1})}}_{p_i(y_{-\ell}, \dots, y_{-1})} \cdot k_h(\cdot - X_i). \quad (10.2)$$

To sample a new time series $(x_0, x_1, x_2 \dots, x_t, x_{t+1}, \dots)$ based on the observation $(X_0, X_1, X_2, \dots, X_n)$ pick an (arbitrary, but reasonable) start sequence $(x_{1-\ell}, \dots, x_0)$. Next, generate a sample x_1 with $x_1 \sim \hat{f}(\cdot \mid x_{1-\ell}, \dots, , x_0)$ by using (10.2), then $x_2 \sim \hat{f}(\cdot \mid x_{2-\ell}, \dots, x_0, x_1)$; in general $x_{t+1} \sim \hat{f}(\cdot \mid x_{t-\ell+1}, \dots, x_t)$.

Notice as well that the vector $(X_i, \dots, X_{i-\ell+1})_{i=\ell}$ is Markovian and Example 10.3 is the special case with lag $\ell = 1$.

10.2 DIEBOLD-MARIANO TEST

In empirical applications it is often the case that two or more time series models are available for forecasting a particular variable of interest. The Diebold-Mariano test addresses the question if they are equally good.

10.3 IMPLEMENTATIONS IN JULIA AND R

Julia implementation of the nonparametric forecast (10.2) to reproduce Figure **??**.

```
using CSV, DataFrames, Distributions, Gnuplot
kernel= Logistic(0., 0.5) # Logistic with bandwidth

function Kernel(x, y)
  SigmoidKernel(x,y; lag= 7.5)
end

df= CSV.read("C:/Users/Alois/Dropbox/Julia/StochasticProcess/nottem.csv", DataFrame)

lag= 4; simulations= 20*12
times= [df.time; 1940:1/12:1940+ (simulations- 1)/ 12]
temp=  copy(times); temp[1:length(df.time)].= df.temperature
n= length(df.time); weight= Vector{Float64}(undef, n-lag)
for k= 1:simulations
  for i= 1:n-lag
    weight[i]= prod(pdf(kernel, temp[n+k-lag:n+k-1]- temp[i:i+lag-1]))
  end
  U= rand(); iStar= lag+ findfirst(x-> U* sum(weight)<x, cumsum(weight))
  temp[n+k]= temp[iStar] + rand(kernel)
end

@gp "reset;■set■title■'nottem';■set■border■3"
@gp :- df.time df.temperature "ls■-1■title■'temperature'■with■linespoints"
@gp :- times[n:end] temp[n:end] "ls■-1■lt■rgb'blue'■title■'simulation■conditional■pdf'■with■lir

condExp= RKHSTS(df.temperature; lag= lag, \lambda=.3, kernel= Kernel)  # new realization
for k=1:simulations
  temp[n+k]= condExp(temp[n+k-lag:n+k-1])+ 2.1*randn()
end
```

Implementation in R of the nonparametric forecast (10.2).

```
temp<- read.csv("~/../Dropbox/Lehre/Vorlesungen/Zeitreihen/HistoricTSTemperatureGermany.csv", sep= ";", de
temp$date<- as.Date(temp$date, "%m/%d/%Y")
simulations<- 100              # forecasts to simulate
lags<-          4      # lags used in simulation
n<- length(temp$temperature)  # length of time series
(bandwidth<- sd(temp$temperature)/ (n^ (1/(lags+ 4)))/ 12)
kernel<- function(t, h){    #exp(-1/2*(t/h)^2)           #Gaussian kernel
                            1/ (exp(t/h)+exp(-t/h))^2 }   #Logistic kernel

tempSimulation<- vector(length= simulations+ lags)
(tempSimulation[1:lags]<- tail(temp$temperature, lags))  # most recent observations
weight<- vector(length= n)
for(k in ((lags+1):(lags+simulations))){ # simulation count
  for(i in ((lags+1):n)){                # run next simulation step
    weight[i]<- prod(kernel(    tempSimulation[(k-lags):(k-1)]
                            - temp$temperature[(i-lags):(i-1)], bandwidth))}
  u<- runif(1, min= 0, max= 1)           # composition method
  iStar<- min(which(cumsum(weight) > u* sum(weight), arr.ind= TRUE))
  tempSimulation[k]= temp$temperature[iStar]        # sample next forecast
              + bandwidth* rlogis(1, location= 0, scale= 1)}

```

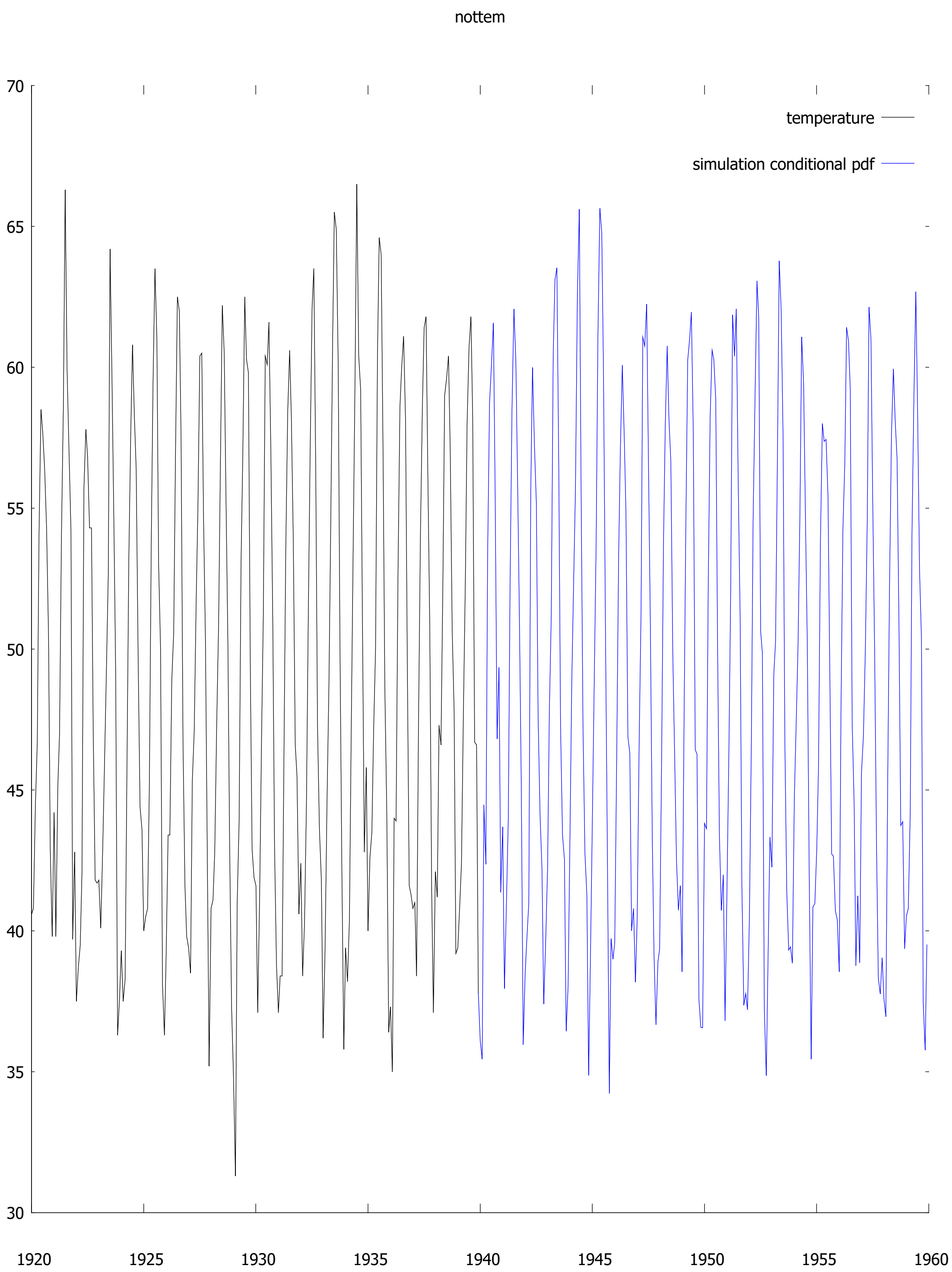


Figure 10.2: nottem time series with nonparametric forecast, computed with methods from Section 10.3; cf. also Figure 1.4

```r
# fix output
temp$simulation<- NA                                            # append new column
tmp<- seq(max(temp$date), by= 'month', length= simulations+1) # new months
ntemp<- nrow(temp)                                              # total number of rows
temp[(ntemp+1):(ntemp+ simulations),]$date<- tmp[-1]           # append new months and
temp[(ntemp+1):(ntemp+ simulations),]$simulation<- tempSimulation[-(1:lags)] # simulations

plot(temperature ~ date, temp[(n-150):(n+simulations),],
                        type= 'l', xlab='year/ forecast (blue)', ylab= 'termperature/ C')
lines(simulation ~ date, temp, col= 'blue', type= 'l')
```

Index