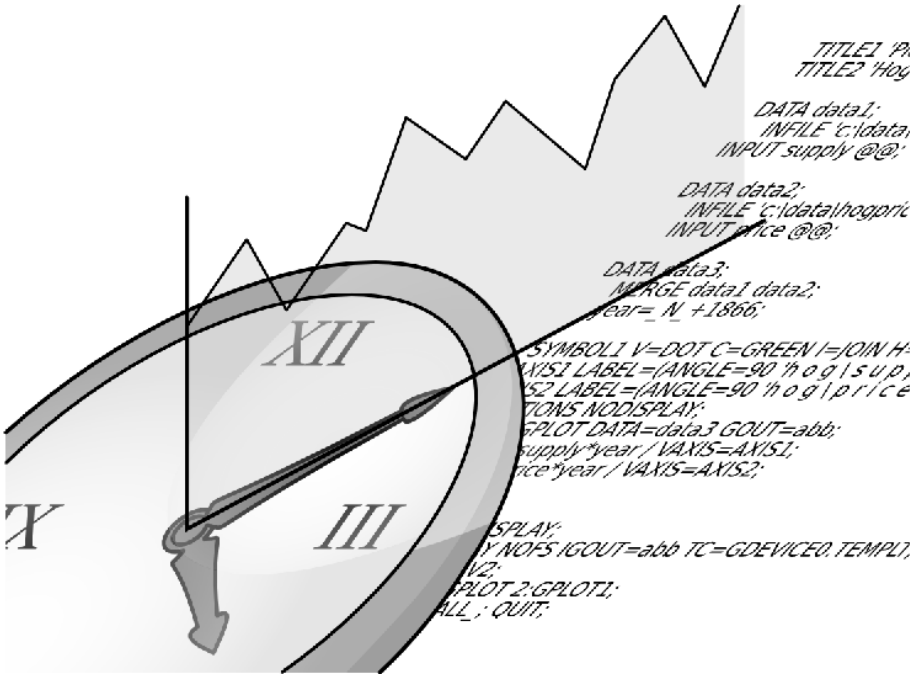
A First Course on Time Series Analysis

Examples with SAS



Chair of Statistics
University of Würzburg

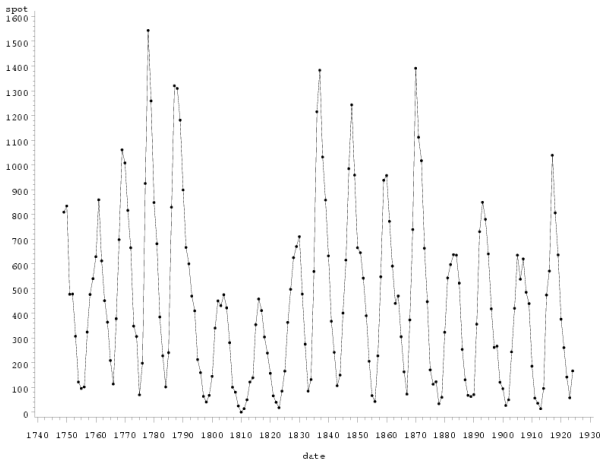
Version: 2005.March.01

Editors	Michael Falk, Frank Marohn, René Michel, Daniel Hofmann, Maria Macke
Programs	Bernward Tewes, René Michel, Daniel Hofmann

Preface

The analysis of real data by means of statistical methods with the aid of a software package common in industry and administration usually is not an integral part of mathematics studies, but it will certainly be part of a future professional work.

The practical need for an investigation of time series data is exemplified by the following plot, which displays the yearly sunspot numbers between 1749 and 1924. These data are also known as the Wolf or Wölfer (a student of Wolf) Data. For a discussion of these data and further literature we refer to Wei (1990), Example 5.2.5.



The present book links up elements from time series analysis with a selection of statistical procedures used in general practice including the statistical software package SAS (Statistical Analysis System). Consequently this book addresses students of statistics as well as students of other branches such as economics, demography and engineering, where lectures on statistics belong to their academic training. But it is also intended for the practician who, beyond the use of statistical tools, is interested in their mathematical background. Numerous problems illustrate the applicability of the presented statistical procedures, where SAS gives the solutions. The programs used are explicitly listed and explained. No previous experience is expected neither in SAS nor in a special computer system so that a short training period is guaranteed.

This book is meant for a two semester course (lecture, seminar or practical training) where the first two chapters can be dealt with in the first semester. They

provide the principal components of the analysis of a time series in the time domain. Chapters 3, 4 and 5 deal with its analysis in the frequency domain and can be worked through in the second term. In order to understand the mathematical background some terms are useful such as convergence in distribution, stochastic convergence, maximum likelihood estimator as well as a basic knowledge of the test theory, so that work on the book can start after an introductory lecture on stochastics. Each chapter includes exercises. An exhaustive treatment is recommended.

Due to the vast field a selection of the subjects was necessary. Chapter 1 contains elements of an exploratory time series analysis, including the fit of models (logistic, Mitscherlich, Gompertz curve) to a series of data, linear filters for seasonal and trend adjustments (difference filters, Census $X-11$ Program) and exponential filters for monitoring a system. Autocovariances and autocorrelations as well as variance stabilizing techniques (Box–Cox transformations) are introduced. Chapter 2 provides an account of mathematical models of stationary sequences of random variables (white noise, moving averages, autoregressive processes, ARIMA models, cointegrated sequences, ARCH- and GARCH-processes, state-space models) together with their mathematical background (existence of stationary processes, covariance generating function, inverse and causal filters, stationarity condition, Yule–Walker equations, partial autocorrelation). The Box–Jenkins program for the specification of ARMA-models is discussed in detail (AIC, BIC and HQC information criterion). Gaussian processes and maximum likelihod estimation in Gaussian models are introduced as well as least squares estimators as a nonparametric alternative. The diagnostic check includes the Box–Ljung test. Many models of time series can be embedded in state-space models, which are introduced at the end of Chapter 2. The Kalman filter as a unified prediction technique closes the analysis of a time series in the time domain. The analysis of a series of data in the frequency domain starts in Chapter 3 (harmonic waves, Fourier frequencies, periodogram, Fourier transform and its inverse). The proof of the fact that the periodogram is the Fourier transform of the empirical autocovariance function is given. This links the analysis in the time domain with the analysis in the frequency domain. Chapter 4 gives an account of the analysis of the spectrum of the stationary process (spectral distribution function, spectral density, Herglotz's theorem). The effects of a linear filter are studied (transfer and power transfer function, low pass and high pass filters, filter design) and the spectral densities of ARMA-processes are computed. Some basic elements of a statistical analysis of a series of data in the frequency domain are provided in Chapter 5. The problem of testing for a white noise is dealt with (Fisher's κ-statistic, Bartlett–Kolmogorov–Smirnov test) together with the estimation of the spectral density (periodogram, discrete spectral average estimator, kernel estimator, confidence intervals).

This book is consecutively subdivided in a statistical part and an SAS-specific part. For better clearness the SAS-specific part, including the diagrams generated

with SAS, always starts with a computer symbol, representing the beginning of a session at the computer, and ends with a printer symbol for the end of this session.

This SAS-specific part is again divided in a diagram created with SAS, the program, which generated the diagram, and explanations to this program. In order to achieve a further differentiation between SAS-commands and individual nomenclature, SAS-specific commands were written in CAPITAL LETTERS, whereas individual notations were written in lower-case letters.

Contents

Chapter 1

Elements of Exploratory Time Series Analysis

A time series is a sequence of observations that are arranged according to the time of their outcome. The annual crop yield of sugar-beets and their price per ton for example is recorded in agriculture. The newspapers' business sections report daily stock prices, weekly interest rates, monthly rates of unemployment and annual turnovers. Meteorology records hourly wind speeds, daily maximum and minimum temperatures and annual rainfall. Geophysics is continuously observing the shaking or trembling of the earth in order to predict possibly impending earthquakes. An electroencephalogram traces brain waves made by an electroencephalograph in order to detect a cerebral disease, an electrocardiogram traces heart waves. The social sciences survey annual death and birth rates, the number of accidents in the home and various forms of criminal activities. Parameters in a manufacturing process are permanently monitored in order to carry out an on-line inspection in quality assurance.

There are, obviously, numerous reasons to record and to analyze the data of a time series. Among these is the wish to gain a better understanding of the data generating mechanism, the prediction of future values or the optimal control of a system. The characteristic property of a time series is the fact that the data are *not generated independently*, their dispersion varies in time, they are often governed by a trend and they have cyclic components. Statistical procedures that suppose *independent* and *identically distributed* data are, therefore, excluded from the analysis of time series. This requires proper methods that are summarized under *time series analysis*.

1.1 The Additive Model for a Time Series

The additive model for a given time series $y_1, \ldots, y_n$ is the assumption that these data are realizations of random variables Y_t that are themselves sums of four components

$$Y_t = T_t + Z_t + S_t + R_t, \qquad t = 1, \ldots, n. \tag{1.1}$$

where T_t is a (monotone) function of t, called *trend*, and Z_t reflects some nonrandom long term cyclic influence. Think of the famous business cycle usually consisting of recession, recovery, growth, and decline. S_t describes some nonrandom short term cyclic influence like a seasonal component whereas R_t is a random variable grasping all the deviations from the ideal non-stochastic model $y_t = T_t + Z_t + S_t$. The variables T_t and Z_t are often summarized as

$$G_t = T_t + Z_t, \tag{1.2}$$

describing the long term behavior of the time series. We suppose in the following that the expectation $\mathrm{E}(R_t)$ of the error variable exists and equals zero, reflecting the assumption that the random deviations above or below the nonrandom model balance each other on the average. Note that $\mathrm{E}(R_t) = 0$ can always be achieved by appropriately modifying one or more of the nonrandom components.

Example 1.1.1. (Unemployed1 Data). The following data y_t, $t = 1, \ldots, 51$, are the monthly numbers of unemployed workers in the building trade in Germany from July 1975 to September 1979.

MONTH	T	UNEMPLYD
July	1	60572
August	2	52461
September	3	47357
October	4	48320
November	5	60219
December	6	84418
January	7	119916
February	8	124350
March	9	87309
April	10	57035
May	11	39903
June	12	34053
July	13	29905

August	14	28068
September	15	26634
October	16	29259
November	17	38942
December	18	65036
January	19	110728
February	20	108931
March	21	71517
April	22	54428
May	23	42911
June	24	37123
July	25	33044
August	26	30755
September	27	28742
October	28	31968
November	29	41427
December	30	63685
January	31	99189
February	32	104240
March	33	75304
April	34	43622
May	35	33990
June	36	26819
July	37	25291
August	38	24538
September	39	22685
October	40	23945
November	41	28245
December	42	47017
January	43	90920
February	44	89340
March	45	47792
April	46	28448
May	47	19139
June	48	16728
July	49	16523
August	50	16622
September	51	15499

Figure 1.1.1. Listing of Unemployed1 Data.

```
***     Program 1_1_1   ***;
TITLE1 'Listing';
TITLE2 'Unemployed1 Data';

DATA data1;
```

```
 INFILE 'c:\data\unemployed1.txt';
 INPUT month $ t unemplyd;

PROC PRINT DATA = data1 NOOBS;
RUN;QUIT;
```

This program consists of two main parts, a `DATA` and a `PROC` step.

The `DATA` step started with the `DATA` statement creates a temporary dataset named `data1`. The purpose of `INFILE` is to link the `DATA` step to a raw dataset outside the program. The pathname of this dataset depends on the operating system; we will use the syntax of MS-DOS, which is most commonly known. `INPUT` tells SAS how to read the data. Three variables are defined here, where the first one contains character values. This is determined by the `$` sign behind the variable name. For each variable one value per line is read from the source into the computer's memory.

The statement `PROC procedurename DATA=filename;` invokes a procedure that is linked to the data from `filename`. Without the option `DATA=filename` the most recently created file is used.

The `PRINT` procedure lists the data; it comes with numerous options that allow control of the variables to be printed out, 'dress up' of the display etc. The SAS internal observation number (`OBS`) is printed by default, `NOOBS` suppresses the column of observation numbers on each line of output. An optional `VAR` statement determines the order (from left to right) in which variables are displayed. If not specified (like here), all variables in the data set will be printed in the order they were defined to SAS. Entering `RUN;` at any point of the program tells SAS that a unit of work (`DATA` step or `PROC`) ended. SAS then stops reading the program and begins to execute the unit. The `QUIT;` statement at the end terminates the processing of SAS.

A line starting with an asterisk `*` and ending with a semicolon `;` is ignored. These comment statements may occur at any point of the program except within raw data or another statement.

The `TITLE` statement generates a title. Its printing is actually suppressed here and in the following.

The following plot of the Unemployed1 Data shows a seasonal component and a downward trend. The period from July 1975 to September 1979 might be too short to indicate a possibly underlying long term business cycle.

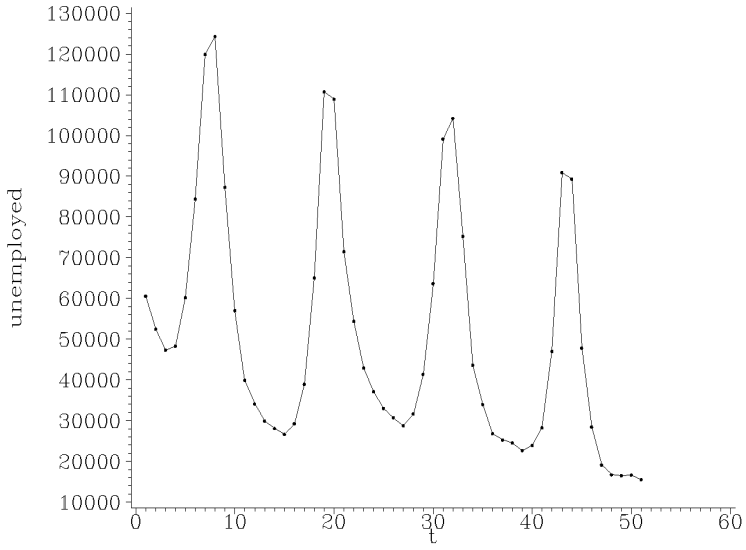


Figure 1.1.2. Plot of Unemployed1 Data.

```
***     Program 1_1_2   ***;
TITLE1 'Plot';
TITLE2 'Unemployed1 Data';

DATA data1;
INFILE 'c:\data\unemployed1.txt';
INPUT month $ t unemplyd;

AXIS1 LABEL=(ANGLE=90 'unemployed');
AXIS2 LABEL=('t');
SYMBOL1 V=DOT C=GREEN I=JOIN H=0.4 W=1;
PROC GPLOT DATA=data1;
    PLOT unemplyd*t / VAXIS=AXIS1 HAXIS=AXIS2;
RUN; QUIT;
```

Variables can be plotted by using the `GPLOT` procedure, where the graphical output is controlled by numerous options.

The `AXIS` statements with the `LABEL` options control labelling of the vertical and horizontal axes. `ANGLE=90` causes a rotation of the label of 90° so that it parallels the (vertical) axis in this example.

The `SYMBOL` statement defines the manner in which the data are displayed. `V=DOT`

`C=GREEN I=JOIN H=0.4 W=1` tell SAS to plot green dots of height 0.4 and to join them with a line of width 1. The `PLOT` statement in the `GPLOT` procedure is of the form `PLOT y-variable*x-variable / options;`, where the options here define the horizontal and the vertical axes.

Models with a Nonlinear Trend

In the additive model $Y_t = T_t + R_t$, where the nonstochastic component is only the trend T_t reflecting the growth of a system, and assuming $\mathrm{E}(R_t) = 0$, we have

$$\mathrm{E}(Y_t) = T_t =: f(t).$$

A common assumption is that the function f depends on several (unknown) *parameters* $\beta_1, \ldots, \beta_p$, i.e.,

$$f(t) = f(t; \beta_1, \ldots, \beta_p). \tag{1.3}$$

However, the *type* of the function f is known. The parameters $\beta_1, \ldots, \beta_p$ are then to be estimated from the set of realizations y_t of the random variables Y_t. A common approach is a *least squares estimate* $\hat{\beta}_1, \ldots, \hat{\beta}_p$ satisfying

$$\sum_t \Big(y_t - f(t; \hat{\beta}_1, \ldots, \hat{\beta}_p)\Big)^2 = \min_{\beta_1,\ldots,\beta_p} \sum_t \Big(y_t - f(t; \beta_1, \ldots, \beta_p)\Big)^2, \tag{1.4}$$

whose computation, if it exists at all, is a numerical problem. The value $\hat{y}_t := f(t; \hat{\beta}_1, \ldots, \hat{\beta}_p)$ can serve as a *prediction* of a future y_t. The *observed* differences $y_t - \hat{y}_t$ are called *residuals*. They contain information about the goodness of the fit of our model to the data. In the following we list several popular examples of trend functions.

The Logistic Function

The function

$$f_{\log}(t) := f_{\log}(t; \beta_1, \beta_2, \beta_3) := \frac{\beta_3}{1 + \beta_2 \exp(-\beta_1 t)}, \qquad t \in \mathbb{R}, \tag{1.5}$$

with $\beta_1, \beta_2, \beta_3 \in \mathbb{R} \setminus \{0\}$ is the widely used *logistic function*.

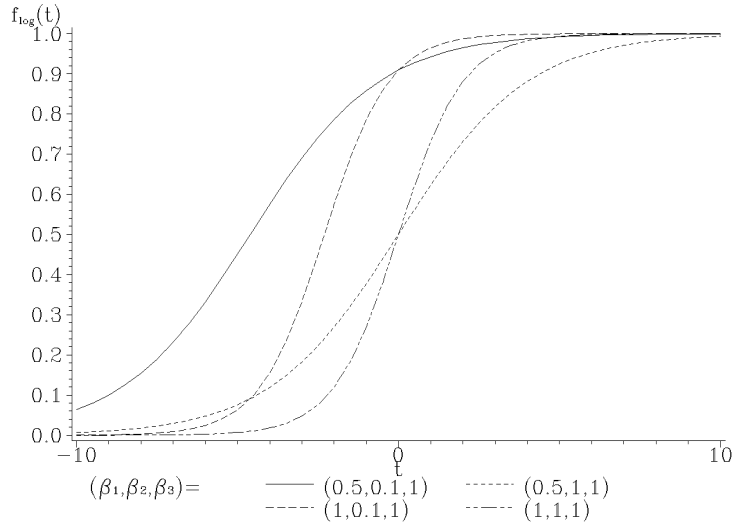


Figure 1.1.3. The logistic function $f_{\log}$ with different values of $\beta_1, \beta_2, \beta_3$.

```
***     Program 1_1_3   ***;
TITLE1 'Plots of the Logistic Function';

DATA data1;
 beta3=1;
 DO beta1= 0.5, 1;
   DO beta2=0.1, 1;
     DO t=-10 TO 10 BY 0.5;
        s = COMPRESS('(' || beta1 || ',' || beta2 ||
                     ',' || beta3 || ')');
        f_log=beta3/(1+beta2*EXP(-beta1*t));
    OUTPUT;
   END; END; END;


SYMBOL1 C=GREEN V=NONE I=JOIN L=1;
SYMBOL2 C=GREEN V=NONE I=JOIN L=2;
SYMBOL3 C=GREEN V=NONE I=JOIN L=3;
SYMBOL4 C=GREEN V=NONE I=JOIN L=33;
AXIS1 LABEL=(H=2 'f' H=1 'log' H=2 '(t)');
AXIS2 LABEL=('t');
```

```
LEGEND1 LABEL=(F=CGREEK H=2 '(b' H=1 '1' H=2 ',
         b' H=1 '2' H=2 ',b' H=1 '3' H=2 ')=');
PROC GPLOT DATA=data1;
    PLOT f_log*t=s / VAXIS=AXIS1 HAXIS=AXIS2
        LEGEND=LEGEND1;
RUN; QUIT;
```

A function is plotted by computing its values at numerous grid points and then joining them. The computation is done in the `DATA` step, where the data file `data1` is generated. It contains the values of `f_log`, computed at the grid $t = -10, -9.5, \ldots, 10$ and indexed by the vector `s` of the different choices of parameters. This is done by nested `DO` loops. The operator `||` merges two strings and `COMPRESS` removes the empty space in the string. `OUTPUT` then stores the values of interest of `f_log`, `t` and `s` (and the other variables) in the data set `data1`. The four functions are plotted by the `GPLOT` procedure by adding `=s` in the `PLOT` statement. This also automatically generates a legend, which is customized by the `LEGEND1` statement. Here the label is modified by using a greek font (`F=CGREEK`) and generating smaller letters of height 1 for the indices, while assuming a normal height of 2 (`H=1` and `H=2`). The last feature is also used in the axis statement. For each value of `s` SAS takes a new `SYMBOL` statement. They generate lines of different line types (`L=1,2, 3, 33`).

We obviously have $\lim_{t\to\infty} f_{\log}(t) = \beta_3$, if $\beta_1 > 0$. The value β_3 often resembles the maximum impregnation or growth of a system. Note that

$$\begin{aligned}
\frac{1}{f_{\log}(t)} &= \frac{1+\beta_2\exp(-\beta_1 t)}{\beta_3} \\
&= \frac{1-\exp(-\beta_1)}{\beta_3} + \exp(-\beta_1)\frac{1+\beta_2\exp(-\beta_1(t-1))}{\beta_3} \\
&= \frac{1-\exp(-\beta_1)}{\beta_3} + \exp(-\beta_1)\frac{1}{f_{\log}(t-1)} \\
&= a + \frac{b}{f_{\log}(t-1)}. \qquad (1.6)
\end{aligned}$$

This means that there is a linear relationship among $1/f_{\log}(t)$. This can serve as a basis for estimating the parameters $\beta_1, \beta_2, \beta_3$ by an appropriate linear least squares approach, see Exercises 2 and 3. In the following example we fit the logistic trend model (1.5) to the population growth of the area of North Rhine-Westphalia (NRW), which is a federal state of Germany.

Example 1.1.2. (Population1 Data). The following table shows the population sizes y_t in millions of the area of North-Rhine-Westphalia in 5 years steps from 1935 to 1980 as well as their predicted values $\hat{y}_t$, obtained from a least squares estimation as described in (1.4) for a logistic model.

Year	t	Population sizes y_t (in millions)	Predicted values $\hat{y}_t$ (in millions)
1935	1	11.772	10.930
1940	2	12.059	11.827
1945	3	11.200	12.709
1950	4	12.926	13.565
1955	5	14.442	14.384
1960	6	15.694	15.158
1965	7	16.661	15.881
1970	8	16.914	16.548
1975	9	17.176	17.158
1980	10	17.044	17.710

Table 1.1.1. Population1 Data.

As a prediction of the population size at time t we obtain in the logistic model

$$\begin{aligned}\hat{y}_t &:= \frac{\hat{\beta}_3}{1+\hat{\beta}_2\exp(-\hat{\beta}_1 t)}\\ &= \frac{21.5016}{1+1.1436\exp(-0.1675\,t)}\end{aligned}$$

with the estimated saturation size $\hat{\beta}_3 = 21.5016$. The following plot shows the data and the fitted logistic curve.

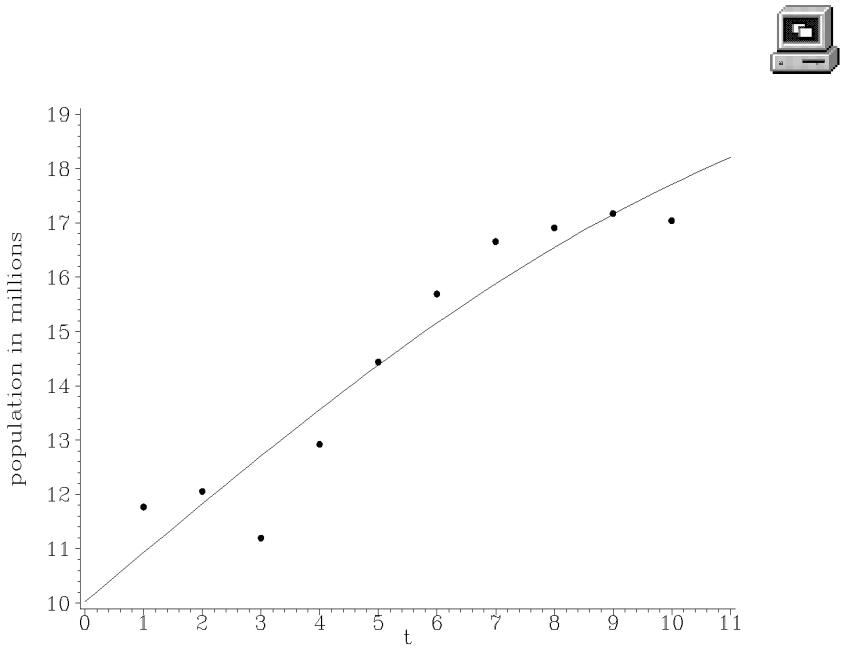


Figure 1.1.4. NRW population sizes and fitted logistic function.

```
***     Program 1_1_4   ***;
TITLE1 'Population sizes and logistic fit';
TITLE2 'Population1 Data';

DATA data1;
INFILE 'c:\data\population1.txt';
INPUT year t pop;

PROC NLIN DATA=data1 OUTEST=estimate;
    MODEL pop=beta3/(1+beta2*EXP(-beta1*t));
    PARAMETERS beta1=1 beta2=1 beta3=20;
RUN;

DATA data2;
SET estimate(WHERE=(_TYPE_='FINAL'));
DO t1=0 TO 11 BY 0.2;
  f_log=beta3/(1+beta2*EXP(-beta1*t1));
  OUTPUT;
END;

DATA data3;
```

```
    MERGE data1 data2;

AXIS1 LABEL=(ANGLE=90 'population in millions');
AXIS2 LABEL=('t');
SYMBOL1 V=DOT C=GREEN I=NONE;
SYMBOL2 V=NONE C=GREEN I=JOIN W=1;
PROC GPLOT DATA=data3;
    PLOT pop*t=1 f_log*t1=2 / OVERLAY VAXIS=AXIS1
                 HAXIS=AXIS2;
RUN; QUIT;
```

The procedure `NLIN` fits nonlinear regression models by least squares. The `OUTEST` option names the data set to contain the parameter estimates produced by `NLIN`. The `MODEL` statement defines the prediction equation by declaring the dependent variable and defining an expression that evaluates predicted values. A `PARAMETERS` statement must follow the `PROC NLIN` statement. Each `parameter=value` expression specifies the starting values of the parameter. Using the final estimates of `PROC NLIN` by the `SET` statement in combination with the `WHERE` data set option, the second data step generates the fitted logistic function values. The options in the `GPLOT` statement cause the data points and the predicted function to be shown in one plot, after they were stored together in a new data set `data3` merging `data1` and `data2` with the `MERGE` statement.

The Mitscherlich Function

The *Mitscherlich function* is typically used for modelling the long term growth of a system:

$$f_M(t) := f_M(t; \beta_1, \beta_2, \beta_3) := \beta_1 + \beta_2 \exp(\beta_3 t), \qquad t \geq 0, \tag{1.7}$$

where $\beta_1, \beta_2 \in \mathbb{R}$ and $\beta_3 < 0$. Since β_3 is negative we have $\lim_{t\to\infty} f_M(t) = \beta_1$ and thus the parameter β_1 is the saturation value of the system. The (initial) value of the system at the time $t = 0$ is $f_M(0) = \beta_1 + \beta_2$.

The Gompertz Curve

A further quite common function for modelling the increase or decrease of a system is the *Gompertz curve*

$$f_G(t) := f_G(t; \beta_1, \beta_2, \beta_3) := \exp(\beta_1 + \beta_2 \beta_3^t), \qquad t \geq 0, \tag{1.8}$$

where $\beta_1, \beta_2 \in \mathbb{R}$ and $\beta_3 \in (0, 1)$.

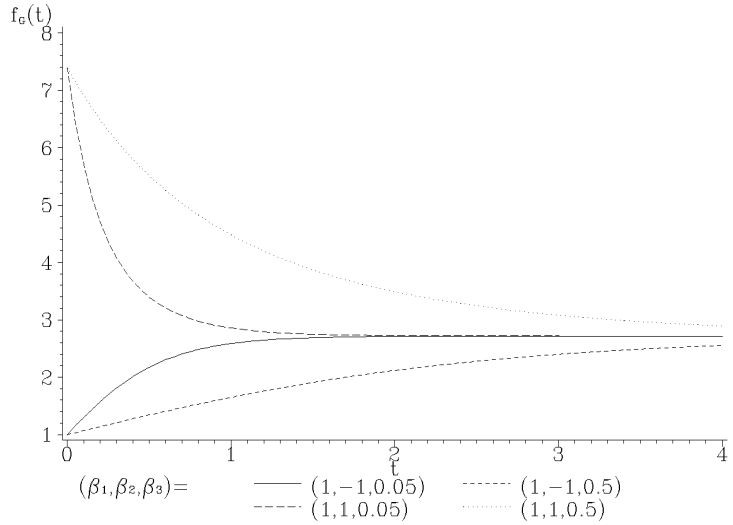


Figure 1.1.5. Gompertz curves with different parameters.

```
***     Program 1_1_5   ***;
TITLE1 'Gompertz curves';

DATA data1;
 DO beta1=1;
   DO beta2=-1, 1;
     DO beta3=0.05, 0.5;
       DO t=0 TO 4 BY 0.05;
         s = COMPRESS('(' || beta1 || ',' || beta2 ||
                          ',' || beta3 || ')');
         f_g=EXP(beta1+beta2*beta3**t);
     OUTPUT;
   END; END; END; END;

SYMBOL1 C=GREEN V=NONE I=JOIN L=1;
SYMBOL2 C=GREEN V=NONE I=JOIN L=2;
SYMBOL3 C=GREEN V=NONE I=JOIN L=3;
SYMBOL4 C=GREEN V=NONE I=JOIN L=33;
AXIS1 LABEL=(H=2 'f' H=1 'G' H=2 '(t)');
```

```
AXIS2 LABEL=('t');
LEGEND1 LABEL=(F=CGREEK H=2 '(b' H=1 '1' H=2
                ',b' H=1 '2' H=2 ',b' H=1 '3' H=2 ')=');
PROC GPLOT DATA=data1;
     PLOT f_g*t=s / VAXIS=AXIS1
                    HAXIS=AXIS2 LEGEND=LEGEND1;
RUN; QUIT;
```

We obviously have

$$\log(f_G(t)) = \beta_1 + \beta_2\beta_3^t = \beta_1 + \beta_2 \exp(\log(\beta_3)t),$$

and thus $\log(f_G)$ is a Mitscherlich function with parameters $\beta_1, \beta_2, \log(\beta_3)$. The saturation size obviously is $\exp(\beta_1)$.

The Allometric Function

The *allometric function*

$$f_a(t) := f_a(t; \beta_1, \beta_2) = \beta_2 t^{\beta_1}, \qquad t \geq 0, \tag{1.9}$$

with $\beta_1 \in \mathbb{R}$, $\beta_2 > 0$, is a common trend function in biometry and economics. It can be viewed as a particular *Cobb–Douglas function*, which is a popular econometric model to describe the output produced by a system depending on an input. Since

$$\log(f_a(t)) = \log(\beta_2) + \beta_1 \log(t), \qquad t > 0,$$

is a linear function of $\log(t)$, with slope β_1 and intercept $\log(\beta_2)$, we can assume a linear regression model for the logarithmic data $\log(y_t)$

$$\log(y_t) = \log(\beta_2) + \beta_1 \log(t) + \varepsilon_t, \qquad t \geq 1,$$

where ε_t are the error variables.

Example 1.1.3. (Income Data). The following table shows the (accumulated) annual average increases of gross and net incomes in thousands DM (deutsche mark) in Germany, starting in 1960.

Year	t	Gross income x_t	Net income y_t
1960	0	0	0
1961	1	0.627	0.486
1962	2	1.247	0.973
1963	3	1.702	1.323
1964	4	2.408	1.867
1965	5	3.188	2.568
1966	6	3.866	3.022
1967	7	4.201	3.259
1968	8	4.840	3.663
1969	9	5.855	4.321
1970	10	7.625	5.482

Table 1.1.2. Income Data.

We assume that the increase of the net income y_t is an allometric function of the time t and obtain

$$\log(y_t) = \log(\beta_2) + \beta_1 \log(t) + \varepsilon_t. \tag{1.10}$$

The least squares estimates of β_1 and $\log(\beta_2)$ in the above linear regression model are (see, for example, Theorem 3.2.2 in Falk et al. (2002))

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{10}(\log(t) - \overline{\log(t)})(\log(y_t) - \overline{\log(y)})}{\sum_{t=1}^{10}(\log(t) - \overline{\log(t)})^2} = 1.019,$$

where $\overline{\log(t)} := 10^{-1}\sum_{t=1}^{10}\log(t) = 1.5104$, $\overline{\log(y)} := 10^{-1}\sum_{t=1}^{10}\log(y_t) = 0.7849$, and hence

$$\widehat{\log(\beta_2)} = \overline{\log(y)} - \hat{\beta}_1\overline{\log(t)} = -0.7549$$

We estimate β_2 therefore by

$$\hat{\beta}_2 = \exp(-0.7549) = 0.4700.$$

The predicted value $\hat{y}_t$ corresponds to the time t

$$\hat{y}_t = 0.47t^{1.019}. \tag{1.11}$$

The following table lists the residuals $y_t - \hat{y}_t$ by which one can judge the goodness of fit of the model (1.11).

t	$y_t - \hat{y}_t$
1	0.0159
2	0.0201
3	-0.1176
4	-0.0646
5	0.1430
6	0.1017
7	-0.1583
8	-0.2526
9	-0.0942
10	0.5662

Table 1.1.3. Residuals of Income Data.

A popular measure for assessing the fit is the *squared multiple correlation coefficient* or R^2*-value*

$$R^2 := 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{1.12}$$

where $\bar{y} := n^{-1}\sum_{t=1}^{n} y_t$ is the average of the observations y_t (cf Section 3.3 in Falk et al. (2002)). In the linear regression model with $\hat{y}_t$ based on the least squares estimates of the parameters, R^2 is necessarily between zero and one with the implications $R^2 = 1$ iff[1] $\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 = 0$ (see Exercise 4). A value of R^2 close to 1 is in favor of the fitted model. The model (1.10) has R^2 equal to .9934, whereas (1.11) has $R^2 = .9789$. Note, however, that the initial model (1.9) is not linear and $\hat{\beta}_2$ is not the least squares estimates, in which case R^2 is no longer necessarily between zero and one and has therefore to be viewed with care as a crude measure of fit.

The annual average gross income in 1960 was 6148 DM and the corresponding net income was 5178 DM. The actual average gross and net incomes were therefore $\tilde{x}_t := x_t + 6.148$ and $\tilde{y}_t := y_t + 5.178$ with the estimated model based on the above predicted values $\hat{y}_t$

$$\hat{\tilde{y}}_t = \hat{y}_t + 5.178 = 0.47t^{1.019} + 5.178.$$

Note that the residuals $\tilde{y}_t - \hat{\tilde{y}}_t = y_t - \hat{y}_t$ are not influenced by adding the constant 5.178 to y_t. The above models might help judging the average tax payer's situation between 1960 and 1970 and to predict his future one. It is apparent from the residuals in Table 1.1.3 that the net income y_t is an almost perfect multiple of t for t between 1 and 9, whereas the large increase y_{10} in 1970 seems to be an outlier. Actually, in 1969 the German government had changed and in 1970 a long strike in Germany caused an enormous increase in the income of civil servants.

[1] if and only if

1.2 Linear Filtering of Time Series

In the following we consider the additive model (1.1) and assume that there is no long term cyclic component. Nevertheless, we allow a trend, in which case the smooth nonrandom component G_t equals the trend function T_t. Our model is, therefore, the decomposition

$$Y_t = T_t + S_t + R_t, \qquad t = 1, 2, \ldots \tag{1.13}$$

with $\mathrm{E}(R_t) = 0$. Given realizations y_t, $t = 1, 2, \ldots, n$, of this time series, the aim of this section is the derivation of estimators $\hat{T}_t, \hat{S}_t$ of the nonrandom functions T_t and S_t and to remove them from the time series by considering $y_t - \hat{T}_t$ or $y_t - \hat{S}_t$ instead. These series are referred to as the *trend* or *seasonally adjusted* time series. The data y_t are decomposed in smooth parts and irregular parts that fluctuate around zero.

Linear Filters

Let $a_{-r}, a_{-r+1}, \ldots, a_s$ be arbitrary real numbers, where $r, s \geq 0$, $r + s + 1 \leq n$. The linear transformation

$$Y_t^* := \sum_{u=-r}^{s} a_u Y_{t-u}, \qquad t = s+1, \ldots, n-r,$$

is referred to as a *linear filter* with *weights* $a_{-r}, \ldots, a_s$. The Y_t are called *input* and the Y_t^* are called *output*.

Obviously, there are less output data than input data, if $(r, s) \neq (0, 0)$. A positive value $s > 0$ or $r > 0$ causes a truncation at the beginning or at the end of the time series; see Example 1.2.1 below. For convenience, we call the vector of weights $(a_u) = (a_{-r}, \ldots, a_s)^T$ a (linear) filter.

A filter (a_u), whose weights sum up to one, $\sum_{u=-r}^{s} a_u = 1$, is called *moving average*. The particular cases $a_u = 1/(2s+1)$, $u = -s, \ldots, s$, with an odd number of equal weights, or $a_u = 1/(2s)$, $u = -s+1, \ldots, s-1$, $a_{-s} = a_s = 1/(4s)$, aiming at an even number of weights, are *simple moving averages of order* $2s+1$ and $2s$, respectively.

Filtering a time series aims at smoothing the irregular part of a time series, thus detecting trends or seasonal components, which might otherwise be covered by fluctuations. While for example a *digital* speedometer in a car can provide its instantaneous velocity, thereby showing considerably large fluctuations, an *analog* instrument that comes with a hand and a built-in smoothing filter, reduces these fluctuations but takes a while to adjust. The latter instrument is much more

comfortable to read and its information, reflecting a trend, is sufficient in most cases.

To compute the output of a simple moving average of order $2s+1$, the following obvious equation is useful:

$$Y_{t+1}^* = Y_t^* + \frac{1}{2s+1}(Y_{t+s+1} - Y_{t-s}).$$

This filter is a particular example of a *low-pass filter*, which preserves the slowly varying trend component of a series but removes from it the rapidly fluctuating or high frequency component. There is a trade-off between the two requirements that the irregular fluctuation should be reduced by a filter, thus leading, for example, to a large choice of s in a simple moving average, and that the long term variation in the data should not be distorted by oversmoothing, i.e., by a too large choice of s. If we assume, for example, a time series $Y_t = T_t + R_t$ without seasonal component, a simple moving average of order $2s+1$ leads to

$$Y_t^* = \frac{1}{2s+1}\sum_{u=-s}^{s} Y_{t-u} = \frac{1}{2s+1}\sum_{u=-s}^{s} T_{t-u} + \frac{1}{2s+1}\sum_{u=-s}^{s} R_{t-u} =: T_t^* + R_t^*,$$

where by some law of large numbers argument

$$R_t^* \sim \mathrm{E}(R_t) = 0,$$

if s is large. But T_t^* might then no longer reflect T_t. A small choice of s, however, has the effect that R_t^* is not yet close to its expectation.

Seasonal Adjustment

A simple moving average of a time series $Y_t = T_t + S_t + R_t$ now decomposes as

$$Y_t^* = T_t^* + S_t^* + R_t^*,$$

where S_t^* is the pertaining moving average of the seasonal components. Suppose, moreover, that S_t is a p-periodic function, i.e.,

$$S_t = S_{t+p}, \qquad t = 1, \ldots, n-p.$$

Take for instance monthly average temperatures Y_t measured at fixed points, in which case it is reasonable to assume a periodic seasonal component S_t with period $p = 12$ months. A simple moving average of order p then yields a *constant* value $S_t^* = S$, $t = p, p+1, \ldots, n-p$. By adding this constant S to the trend function T_t and putting $T_t' := T_t + S$, we can assume in the following that $S = 0$. Thus we obtain for the differences

$$D_t := Y_t - Y_t^* \sim S_t + R_t$$

and, hence, averaging these differences yields

$$\bar{D}_t := \frac{1}{n_t}\sum_{j=0}^{n_t-1} D_{t+jp} \sim S_t, \qquad t = 1, \ldots, p,$$
$$\bar{D}_t := \bar{D}_{t-p} \qquad \text{for } t > p,$$

where n_t is the number of periods available for the computation of $\bar{D}_t$. Thus,

$$\hat{S}_t := \bar{D}_t - \frac{1}{p}\sum_{j=1}^{p} \bar{D}_j \sim S_t - \frac{1}{p}\sum_{j=1}^{p} S_j = S_t \tag{1.14}$$

is an estimator of $S_t = S_{t+p} = S_{t+2p} = \ldots$ satisfying

$$\frac{1}{p}\sum_{j=0}^{p-1} \hat{S}_{t+j} = 0 = \frac{1}{p}\sum_{j=0}^{p-1} S_{t+j}.$$

The differences $Y_t - \hat{S}_t$ with a seasonal component close to zero are then the seasonally adjusted time series.

Example 1.2.1. For the 51 Unemployed1 Data in Example 1.1.1 it is obviously reasonable to assume a periodic seasonal component with $p = 12$ months. A simple moving average of order 12

$$Y_t^* = \frac{1}{12}\Big(\frac{1}{2}Y_{t-6} + \sum_{u=-5}^{5} Y_{t-u} + \frac{1}{2}Y_{t+6}\Big), \qquad t = 7, \ldots, 45,$$

then has a constant seasonal component, which we assume to be zero by adding this constant to the trend function. The following table contains the values of D_t, $\bar{D}_t$ and the estimates $\hat{S}_t$ of S_t.

	d_t (rounded values)					
Month	1976	1977	1978	1979	$\bar{d}_t$ (rounded)	$\hat{s}_t$(rounded)
January	53201	56974	48469	52611	52814	53136
February	59929	54934	54102	51727	55173	55495
March	24768	17320	25678	10808	19643	19966
April	-3848	42	-5429	–	-3079	-2756
May	-19300	-11680	-14189	–	-15056	-14734
June	-23455	-17516	-20116	–	-20362	-20040
July	-26413	-21058	-20605	–	-22692	-22370
August	-27225	-22670	-20393	–	-23429	-23107
September	-27358	-24646	-20478	–	-24161	-23839
October	-23967	-21397	-17440	–	-20935	-20612
November	-14300	-10846	-11889	–	-12345	-12023
December	11540	12213	7923	–	10559	10881

Table 1.2.1. Table of d_t, $\bar{d}_t$ and of estimates $\hat{s}_t$ of the seasonal component S_t in the Unemployed1 Data.

We obtain for these data

$$\hat{s}_t = \bar{d}_t - \frac{1}{12}\sum_{j=1}^{12} \bar{d}_j = \bar{d}_t + \frac{3867}{12} = \bar{d}_t + 322.25.$$

The Census X-11 Program

In the fifties of the 20th century the U.S. Bureau of the Census has developed a program for seasonal adjustment of economic time series, called the *Census X-11 Program*. It is based on monthly observations and assumes an additive model

$$Y_t = T_t + S_t + R_t$$

as in (1.13) with a seasonal component S_t of period $p = 12$. We give a brief summary of this program following Wallis (1974), which results in a moving average with symmetric weights. The census procedure is discussed in Shiskin and Eisenpress (1957); a complete description is given by Shiskin, Young and Musgrave (1967). A theoretical justification based on stochastic models is provided by Cleveland and Tiao (1976).

The X-11 Program essentially works as the seasonal adjustment described above, but it adds iterations and various moving averages. The different steps of this program are

(1) Compute a simple moving average Y_t^* of order 12 to leave essentially a trend $Y_t^* \sim T_t$.

(2) The difference

$$D_t := Y_t - Y_t^* \sim S_t + R_t$$

then leaves approximately the seasonal plus irregular component.

(3) Apply a moving average of order 5 to *each month* separately by computing

$$\bar{D}_t^{(1)} := \frac{1}{9}\Big(D_{t-24}^{(1)} + 2D_{t-12}^{(1)} + 3D_t^{(1)} + 2D_{t+12}^{(1)} + D_{t+24}^{(1)}\Big) \sim S_t,$$

which gives an estimate of the seasonal component S_t. Note that the moving average with weights $(1, 2, 3, 2, 1)/9$ is a simple moving average of length 3 of simple moving averages of length 3.

(4) The $\bar{D}_t^{(1)}$ are adjusted to approximately sum up to 0 over any 12-months period by putting

$$\hat{S}_t^{(1)} := \bar{D}_t^{(1)} - \frac{1}{12}\Big(\frac{1}{2}\bar{D}_{t-6}^{(1)} + \bar{D}_{t-5}^{(1)} + \cdots + \bar{D}_{t+5}^{(1)} + \frac{1}{2}\bar{D}_{t+6}^{(1)}\Big).$$

(5) The differences

$$Y_t^{(1)} := Y_t - \hat{S}_t^{(1)} \sim T_t + R_t$$

then are the preliminary seasonally adjusted series, quite in the manner as before.

(6) The adjusted data $Y_t^{(1)}$ are further smoothed by a *Henderson* moving average Y_t^{**} of order 9, 13, or 23.

(7) The differences

$$D_t^{(2)} := Y_t - Y_t^{**} \sim S_t + R_t$$

then leave a second estimate of the sum of the seasonal and irregular components.

(8) A moving average of order 7 is applied to each month separately

$$\bar{D}_t^{(2)} := \sum_{u=-3}^{3} a_u D_{t-12u}^{(2)},$$

where the weights a_u come from a simple moving average of order 3 applied to a simple moving average of order 5 of the original data, i.e., the vector of weights is $(1, 2, 3, 3, 3, 2, 1)/15$. This gives a second estimate of the seasonal component S_t.

(9) Step (4) is repeated yielding approximately centered estimates $\hat{S}_t^{(2)}$ of the seasonal components.

(10) The differences

$$Y_t^{(2)} := Y_t - \hat{S}_t^{(2)}$$

then finally give the seasonally adjusted series.

Depending on the length of the Henderson moving average used in step (6), $Y_t^{(2)}$ is a moving average of length 165, 169 or 179 of the original data. Observe that this leads to averages at time t of the past and future seven years, roughly, where seven years is a typical length of business cycles observed in economics (Juglar cycle)[2].

The U.S. Bureau of Census has recently released an extended version of the X-11 Program called *Census X-12-ARIMA*. It is implemented in SAS version 8.1 and higher as PROC X12; we refer to the SAS online documentation for details.

We will see in Example 4.2.4 that linear filters may cause unexpected effects and so, it is not clear a priori how the seasonal adjustment filter described above behaves.

[2] `http://www.drfurfero.com/books/231book/ch05j.html`

Moreover, end-corrections are necessary, which cover the important problem of adjusting current observations. This can be done by some extrapolation.

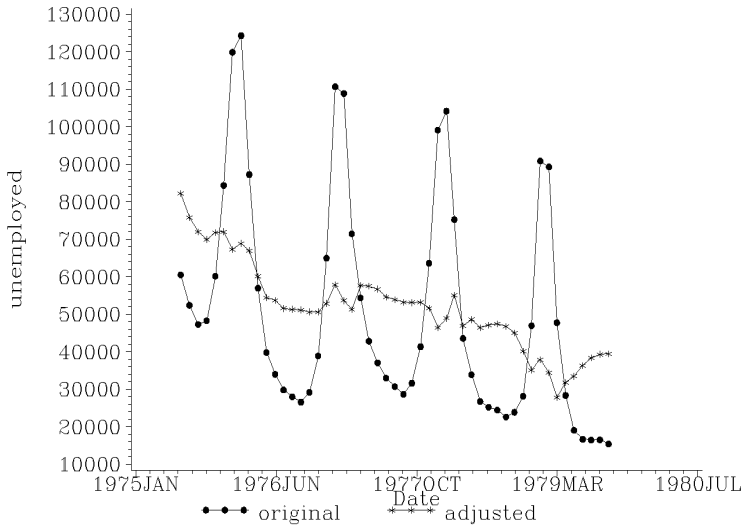


Figure 1.2.1. Plot of the Unemployed1 Data y_t and of $y_t^{(2)}$, seasonally adjusted by the X-11 procedure.

```
***     Program 1_2_1    ***;
TITLE1 'Original and X11 seasonal adjusted data';
TITLE2 'Unemployed1 Data';

DATA data1;
INFILE 'c:\data\unemployed1.txt';
INPUT month $ t upd;
date=INTNX('month','01jul75'd, _N_-1);
FORMAT date yymon.;

PROC X11 DATA=data1;
MONTHLY DATE=date ADDITIVE;
VAR upd;
OUTPUT OUT=data2 B1=upd D11=updx11;
```

```
AXIS1 LABEL=(ANGLE=90 'unemployed');
AXIS2 LABEL=('Date') ;
SYMBOL1 V=DOT C=GREEN I=JOIN H=1 W=1;
SYMBOL2 V=STAR C=GREEN I=JOIN H=1 W=1;
LEGEND1 LABEL=NONE VALUE=('original' 'adjusted');
PROC GPLOT DATA=data2;
  PLOT upd*date=1  updx11*date=2
      / OVERLAY VAXIS=AXIS1 HAXIS=AXIS2 LEGEND=LEGEND1;
RUN; QUIT;
```

In the data, step values for the variables `month`, `t` and `upd` are read from an external file, where `month` is defined as a character variable by the succeeding `$` in the `INPUT` statement. By means of the function `INTNX`, a new variable in a date format is generated containing monthly data starting from the 1st of July 1975. The temporarily created variable `_N_`, which counts the number of cases, is used to determine the distance from the starting value. The `FORMAT` statement attributes the format `yymon` to this variable, consisting of four digits for the year and three for the month.

The SAS procedure `X11` applies the Census X–11 Program to the data. The `MONTHLY` statement selects an algorithm for monthly data, `DATE` defines the date variable and `ADDITIVE` selects an additive model (default: multiplicative model). The results for this analysis for the variable `upd` (unemployed) are stored in a data set named `data2`, containing the original data in the variable `upd` and the final results of the X–11 Program in `updx11`.

The last part of this SAS program consists of statements for generating the plot. Two `AXIS` and two `SYMBOL` statements are used to customize the graphic containing two plots, the original data and the by `X11` seasonally adjusted data. A `LEGEND` statement defines the text that explains the symbols.

Best Local Polynomial Fit

A simple moving average works well for a locally almost linear time series, but it may have problems to reflect a more twisted shape. This suggests fitting higher order local polynomials. Consider $2k+1$ consecutive data $y_{t-k}, \ldots, y_t, \ldots, y_{t+k}$ from a time series. A *local polynomial estimator* of order $p < 2k+1$ is the minimizer $\beta_0, \ldots, \beta_p$ satisfying

$$\sum_{u=-k}^{k} (y_{t+u} - \beta_0 - \beta_1 u - \cdots - \beta_p u^p)^2 = \min . \tag{1.15}$$

If we differentiate the left hand side with respect to each β_j and set the derivatives equal to zero, we see that the minimizers satisfy the $p+1$ linear equations

$$\beta_0 \sum_{u=-k}^{k} u^j + \beta_1 \sum_{u=-k}^{k} u^{j+1} + \cdots + \beta_p \sum_{u=-k}^{k} u^{j+p} = \sum_{u=-k}^{k} u^j y_{t+u}$$

for $j = 0, \ldots, p$. These $p+1$ equations, which are called *normal equations*, can be written in matrix form as

$$\boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\beta} = \boldsymbol{X}^T \boldsymbol{y} \tag{1.16}$$

where

$$\boldsymbol{X} = \begin{pmatrix} 1 & -k & (-k)^2 & \ldots & (-k)^p \\ 1 & -k+1 & (-k+1)^2 & \ldots & (-k+1)^p \\ \vdots & & & \ddots & \vdots \\ 1 & k & k^2 & \ldots & k^p \end{pmatrix} \tag{1.17}$$

is the *design matrix*, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^T$ and $\boldsymbol{y} = (y_{t-k}, \ldots, y_{t+k})^T$. The rank of $\boldsymbol{X}^T \boldsymbol{X}$ equals that of $\boldsymbol{X}$, since their null spaces coincide (Exercise 11). Thus, the matrix $\boldsymbol{X}^T \boldsymbol{X}$ is invertible iff the columns of $\boldsymbol{X}$ are linearly independent. But this is an immediate consequence of the fact that a polynomial of degree p has at most p different roots (Exercise 12). The normal equations (1.16) have, therefore, the unique solution

$$\boldsymbol{\beta} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}. \tag{1.18}$$

The linear prediction of y_{t+u}, based on $u, u^2, \ldots, u^p$, is

$$\hat{y}_{t+u} = (1, u, \ldots, u^p)\boldsymbol{\beta} = \sum_{j=0}^{p} \beta_j u^j.$$

Choosing $u = 0$ we obtain in particular that $\beta_0 = \hat{y}_t$ is a predictor of the central observation y_t among $y_{t-k}, \ldots, y_{t+k}$. The local polynomial approach consists now in replacing y_t by the intercept β_0.

Though it seems as if this local polynomial fit requires a great deal of computational effort by calculating β_0 for each y_t, it turns out that it is actually a moving average. First observe that we can write by (1.18)

$$\beta_0 = \sum_{u=-k}^{k} c_u y_{t+u}$$

with some $c_u \in \mathbb{R}$ which *do not depend* on the values y_u of the time series and hence, (c_u) is a linear filter. Next we show that the c_u sum up to 1. Choose to this end $y_{t+u} = 1$ for $u = -k, \ldots, k$. Then $\beta_0 = 1$, $\beta_1 = \cdots = \beta_p = 0$ is an obvious solution of the minimization problem (1.15). Since this solution is unique, we obtain

$$1 = \beta_0 = \sum_{u=-k}^{k} c_u$$

and thus, (c_u) is a moving average. As can be seen in Exercise 13 it actually has symmetric weights. We summarize our considerations in the following result.

Theorem 1.2.2. *Fitting locally by least squares a polynomial of degree p to $2k+1 > p$ consecutive data points $y_{t-k}, \ldots, y_{t+k}$ and predicting y_t by the resulting intercept β_0, leads to a moving average (c_u) of order $2k+1$, given by the first row of the matrix $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$.*

Example 1.2.3. Fitting locally a polynomial of degree 2 to five consecutive data points leads to the moving average (Exercise 13)

$$(c_u) = \frac{1}{35}(-3, 12, 17, 12, -3)^T.$$

An extensive discussion of local polynomial fit is in Kendall and Ord (1993), Sections 3.2–3.13. For a book-length treatment of local polynomial estimation we refer to Fan and Gijbels (1996). An outline of various aspects such as the choice of the degree of the polynomial and further background material is given in Section 5.2 of Simonoff (1996).

Difference Filter

We have already seen that we can remove a periodic seasonal component from a time series by utilizing an appropriate linear filter. We will next show that also a polynomial trend function can be removed by a suitable linear filter.

Lemma 1.2.4. *For a polynomial $f(t) := c_0 + c_1 t + \cdots + c_p t^p$ of degree p, the difference*

$$\Delta f(t) := f(t) - f(t-1)$$

is a polynomial of degree at most $p-1$.

Proof. The assertion is an immediate consequence of the binomial expansion

$$(t-1)^p = \sum_{k=0}^{p} \binom{p}{k} t^k (-1)^{p-k} = t^p - pt^{p-1} + \cdots + (-1)^p.$$

□

The preceding lemma shows that differencing reduces the degree of a polynomial. Hence,

$$\Delta^2 f(t) := \Delta f(t) - \Delta f(t-1) = \Delta(\Delta f(t))$$

is a polynomial of degree not greater than $p-2$, and

$$\Delta^q f(t) := \Delta(\Delta^{q-1} f(t)), \qquad 1 \le q \le p,$$

is a polynomial of degree at most $p-q$. The function $\Delta^p f(t)$ is therefore a constant. The linear filter

$$\Delta Y_t = Y_t - Y_{t-1}$$

with weights $a_0 = 1, a_1 = -1$ is the first order *difference filter*. The recursively defined filter

$$\Delta^p Y_t = \Delta(\Delta^{p-1} Y_t), \qquad t = p, \ldots, n,$$

is the difference filter of order p.

The difference filter of second order has, for example, weights $a_0 = 1, a_1 = -2, a_2 = 1$

$$\begin{aligned}\Delta^2 Y_t &= \Delta Y_t - \Delta Y_{t-1} \\ &= Y_t - Y_{t-1} - Y_{t-1} + Y_{t-2} = Y_t - 2Y_{t-1} + Y_{t-2}.\end{aligned}$$

If a time series Y_t has a polynomial trend $T_t = \sum_{k=0}^{p} c_k t^k$ for some constants c_k, then the difference filter $\Delta^p Y_t$ of order p removes this trend up to a constant. Time series in economics often have a trend function that can be removed by a first or second order difference filter.

Example 1.2.5. (Electricity Data). The following plots show the total annual output of electricity production in Germany between 1955 and 1979 in millions of kilowatt-hours as well as their first and second order differences. While the original data show an increasing trend, the second order differences fluctuate around zero having no more trend, but there is now an increasing variability visible in the data.

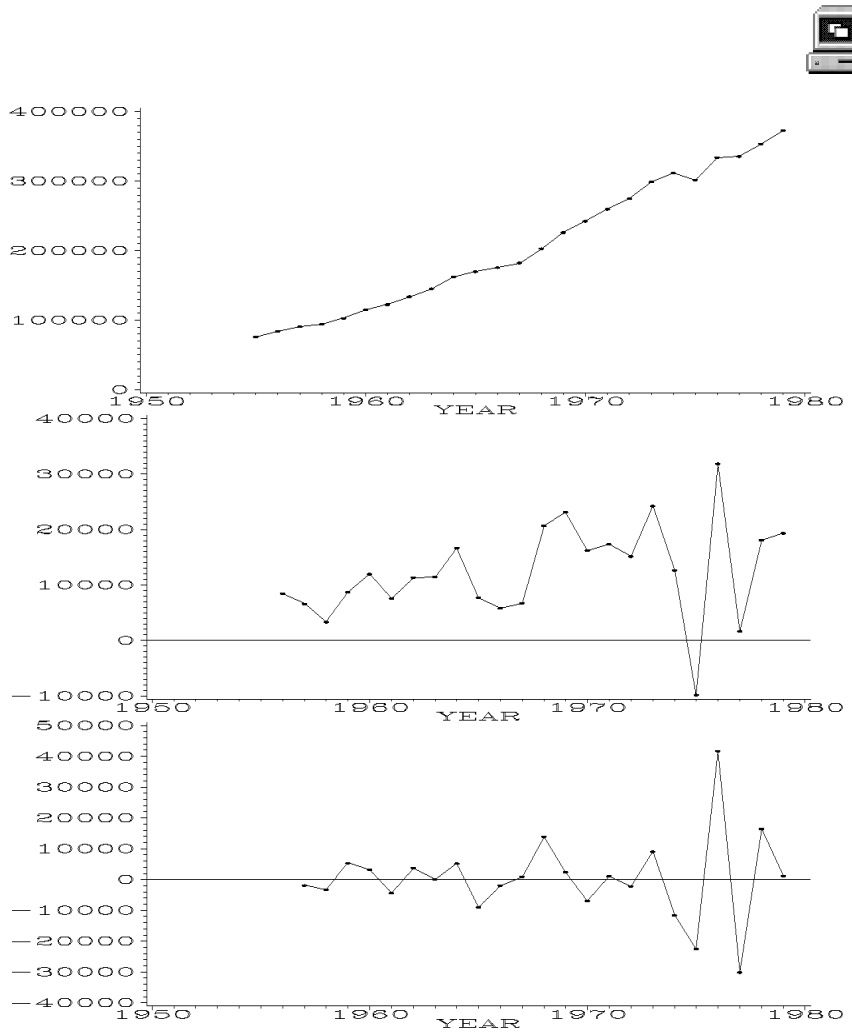


Figure 1.2.2. Annual electricity output, first and second order differences

```
***     Program 1_2_2   ***;
TITLE1 'First and second order differences';
TITLE2 'Electricity Data';

DATA data1(KEEP=year sum delta1 delta2);
  INFILE 'c:\data\electric.txt';
  INPUT year t jan feb mar apr may jun
    jul aug sep oct nov dec;
```

```
  sum=jan+feb+mar+apr+may+jun+
    jul+aug+sep+oct+nov+dec;
  delta1=DIF(sum);
  delta2=DIF(delta1);


AXIS1 LABEL=NONE;
SYMBOL1 V=DOT C=GREEN I=JOIN H=0.5 W=1;
GOPTIONS NODISPLAY;
PROC GPLOT DATA=data1 GOUT=fig;
  PLOT sum*year    / VAXIS=AXIS1 HAXIS=AXIS2;
  PLOT delta1*year / VAXIS=AXIS1 VREF=0;
  PLOT delta2*year / VAXIS=AXIS1 VREF=0;
RUN;

GOPTIONS DISPLAY;
PROC GREPLAY NOFS IGOUT=fig TC=SASHELP.TEMPLT;
  TEMPLATE=V3;
  TREPLAY 1:GPLOT 2:GPLOT1 3:GPLOT2;
RUN; DELETE _ALL_; QUIT;
```

In the first data step, the raw data are read from a file. Because the electric production is stored in different variables for each month of a year, the sum must be evaluated to get the annual output. Using the `DIF` function, the resulting variables `delta1` and `delta2` contain the first and second order differences of the original annual sums.

To display the three plots of `sum`, `delta1` and `delta2` against the variable `year` within one graphic, they are first plotted using the procedure `GPLOT`. Here the option `GOUT=fig` stores the plots in a graphics catalog named `fig`, while `GOPTIONS NODISPLAY` causes no output of this procedure. After changing the `GOPTIONS` back to `DISPLAY`, the procedure `GREPLAY` is invoked. The option `NOFS` (no full-screen) suppresses the opening of a `GREPLAY` window. The subsequent two line mode statements are read instead. The option `IGOUT` determines the input graphics catalog, while `TC=SASHELP.TEMPLT` causes SAS to take the standard template catalog. The `TEMPLATE` statement selects a template from this catalog, which puts three graphics one below the other. The `TREPLAY` statement connects the defined areas and the plots of the the graphics catalog. `GPLOT`, `GPLOT1` and `GPLOT2` are the graphical outputs in the chronological order of the `GPLOT` procedure. The `DELETE` statement after `RUN` deletes all entries in the input graphics catalog.

Note that SAS by default prints borders, in order to separate the different plots. Here these border lines are suppressed by defining `WHITE` as the border color.

For a time series $Y_t = T_t + S_t + R_t$ with a periodic seasonal component $S_t =$

$S_{t+p} = S_{t+2p} = \dots$ the difference

$$Y_t^* := Y_t - Y_{t-p}$$

obviously removes the seasonal component. An additional differencing of proper length can moreover remove a polynomial trend, too. Note that the order of seasonal and trend adjusting makes no difference.

Exponential Smoother

Let $Y_0, \dots, Y_n$ be a time series and let $\alpha \in [0,1]$ be a constant. The linear filter

$$Y_t^* = \alpha Y_t + (1-\alpha) Y_{t-1}^*, \qquad t \geq 1,$$

with $Y_0^* = Y_0$ is called *exponential smoother.*

Lemma 1.2.6. *For an exponential smoother with constant $\alpha \in [0,1]$ we have*

$$Y_t^* = \alpha \sum_{j=0}^{t-1} (1-\alpha)^j Y_{t-j} + (1-\alpha)^t Y_0, \qquad t = 1, 2, \dots, n.$$

Proof. The assertion follows from induction. We have for $t = 1$ by definition $Y_1^* = \alpha Y_1 + (1-\alpha) Y_0$. If the assertion holds for t, we obtain for $t+1$

$$\begin{aligned} Y_{t+1}^* &= \alpha Y_{t+1} + (1-\alpha) Y_t^* \\ &= \alpha Y_{t+1} + (1-\alpha) \Big(\alpha \sum_{j=0}^{t-1} (1-\alpha)^j Y_{t-j} + (1-\alpha)^t Y_0 \Big) \\ &= \alpha \sum_{j=0}^{t} (1-\alpha)^j Y_{t+1-j} + (1-\alpha)^{t+1} Y_0. \end{aligned}$$

□

The parameter α determines the smoothness of the filtered time series. A value of α close to 1 puts most of the weight on the actual observation Y_t, resulting in a highly fluctuating series Y_t^*. On the other hand, an α close to 0 reduces the influence of Y_t and puts most of the weight to the past observations, yielding a smooth series Y_t^*. An exponential smoother is typically used for monitoring a system. Take, for example, a car having an analog speedometer with a hand. It is more convenient for the driver if the movements of this hand are smoothed, which can be achieved by α close to zero. But this, on the other hand, has the effect that an essential alteration of the speed can be read from the speedometer only with a certain delay.

Corollary 1.2.7. *(i) Suppose that the random variables $Y_0, \dots, Y_n$ have common expectation μ and common variance $\sigma^2 > 0$. Then we have for the exponentially smoothed variables with smoothing parameter $\alpha \in (0,1)$*

$$\begin{aligned} \mathrm{E}(Y_t^*) &= \alpha \sum_{j=0}^{t-1} (1-\alpha)^j \mu + \mu(1-\alpha)^t \\ &= \mu(1-(1-\alpha)^t) + \mu(1-\alpha)^t = \mu. \end{aligned} \tag{1.19}$$

If the Y_t are in addition uncorrelated, then

$$\begin{aligned} \mathrm{E}((Y_t^* - \mu)^2) &= \alpha^2 \sum_{j=0}^{t-1} (1-\alpha)^{2j} \sigma^2 + (1-\alpha)^{2t} \sigma^2 \\ &= \sigma^2 \alpha^2 \frac{1-(1-\alpha)^{2t}}{1-(1-\alpha)^2} + (1-\alpha)^{2t} \sigma^2 \\ &\longrightarrow_{t\to\infty} \frac{\sigma^2 \alpha}{2-\alpha} < \sigma^2. \end{aligned} \tag{1.20}$$

(ii) Suppose that the random variables $Y_0, Y_1, \dots$ satisfy $\mathrm{E}(Y_t) = \mu$ for $0 \le t \le N-1$, and $\mathrm{E}(Y_t) = \lambda$ for $t \ge N$. Then we have for $t \ge N$

$$\begin{aligned} \mathrm{E}(Y_t^*) &= \alpha \sum_{j=0}^{t-N} (1-\alpha)^j \lambda + \alpha \sum_{j=t-N+1}^{t-1} (1-\alpha)^j \mu + (1-\alpha)^t \mu \\ &= \lambda(1-(1-\alpha)^{t-N+1}) + \mu\Big((1-\alpha)^{t-N+1}(1-(1-\alpha)^{N-1}) + (1-\alpha)^t\Big) \\ &\longrightarrow_{t\to\infty} \lambda. \end{aligned} \tag{1.21}$$

The preceding result quantifies the influence of the parameter α on the expectation and on the variance i.e., the smoothness of the filtered series Y_t^*, where we assume for the sake of a simple computation of the variance that the Y_t are uncorrelated. If the variables Y_t have common expectation μ, then this expectation carries over to Y_t^*. After a *change point* N, where the expectation of Y_t changes for $t \ge N$ from μ to $\lambda \ne \mu$, the filtered variables Y_t^* are, however, biased. This bias, which will vanish as t increases, is due to the still inherent influence of past observations Y_t, $t < N$. The influence of these variables on the current expectation can be reduced by switching to a larger value of α. The price for the gain in correctness of the expectation is, however, a higher variability of Y_t^* (see Exercise 16).

An exponential smoother is often also used to make forecasts, explicitly by predicting Y_{t+1} through Y_t^*. The forecast error $Y_{t+1} - Y_t^* =: e_{t+1}$ then satisfies the equation $Y_{t+1}^* = \alpha e_{t+1} + Y_t^*$.

1.3 Autocovariances and Autocorrelations

Autocovariances and autocorrelations are measures of dependence between variables in a time series. Suppose that $Y_1, \ldots, Y_n$ are square integrable random variables with the property that the covariance $\mathrm{Cov}(Y_{t+k}, Y_t) = \mathrm{E}((Y_{t+k} - \mathrm{E}(Y_{t+k}))(Y_t - \mathrm{E}(Y_t)))$ of observations with lag k does not depend on t. Then

$$\gamma(k) := \mathrm{Cov}(Y_{k+1}, Y_1) = \mathrm{Cov}(Y_{k+2}, Y_2) = \ldots$$

is called *autocovariance function* and

$$\rho(k) := \frac{\gamma(k)}{\gamma(0)}, \qquad k = 0, 1, \ldots$$

is called *autocorrelation function*.

Let $y_1, \ldots, y_n$ be realizations of a time series $Y_1, \ldots, Y_n$. The empirical counterpart of the autocovariance function is

$$c(k) := \frac{1}{n} \sum_{t=1}^{n-k} (y_{t+k} - \bar{y})(y_t - \bar{y}) \quad \text{with } bary = \frac{1}{n} \sum_{t=1}^{n} y_t$$

and the empirical autocorrelation is defined by

$$r(k) := \frac{c(k)}{c(0)} = \frac{\sum_{t=1}^{n-k} (y_{t+k} - \bar{y})(y_t - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}.$$

See Exercise 8 (ii) in Chapter 2 for the particular role of the factor $1/n$ in place of $1/(n-k)$ in the definition of $c(k)$. The graph of the function $r(k)$, $k = 0, 1, \ldots, n-1$, is called *correlogram*. It is based on the assumption of equal expectations and should, therefore, be used for a trend adjusted series. The following plot is the correlogram of the first order differences of the Sunspot Data. The description can be found on page 176. It shows high and decreasing correlations at regular intervals.

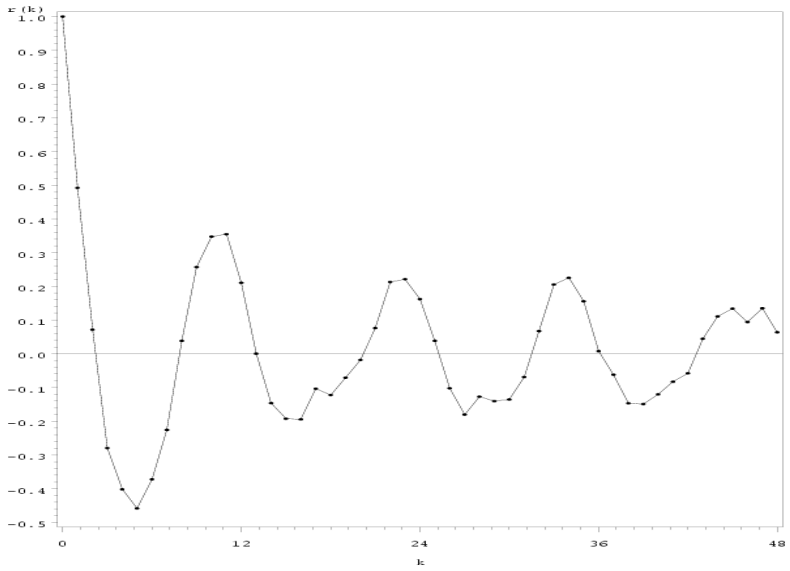


Figure 1.3.1. Correlogram of the first order differences of the Sunspot Data.

```
***     Program 1_3_1   ***;
TITLE1 'Correlogram of first order differences';
TITLE2 'Sunspot Data';

DATA data1;
  INFILE 'c:\data\sunspot.txt';
  INPUT spot @@;
  date=1748+_N_;
  diff1=DIF(spot);

PROC ARIMA DATA=data1;
  IDENTIFY VAR=diff1 NLAG=49 OUTCOV=corr NOPRINT;

AXIS1 LABEL=('r(k)');
AXIS2 LABEL=('k') ORDER=(0 12 24 36 48) MINOR=(N=11);
SYMBOL1 V=DOT C=GREEN I=JOIN H=0.5 W=1;
PROC GPLOT DATA=corr;
  PLOT CORR*LAG / VAXIS=AXIS1 HAXIS=AXIS2 VREF=0;
```

```
RUN; QUIT;
```

In the data step, the raw data are read into the variable `spot`. The specification `@@` suppresses the automatic line feed of the `INPUT` statement after every entry in each row. The variable `date` and the first order differences of the variable of interest `spot` are calculated.

The following procedure `ARIMA` is a crucial one in time series analysis. Here we just need the autocorrelation of `delta`, which will be calculated up to a lag of 49 (`NLAG=49`) by the `IDENTIFY` statement. The option `OUTCOV=corr` causes SAS to create a data set `corr` containing among others the variables `LAG` and `CORR`. These two are used in the following `GPLOT` procedure to obtain a plot of the autocorrelation function. The `ORDER` option in the `AXIS2` statement specifies the values to appear on the horizontal axis as well as their order, and the `MINOR` option determines the number of minor tick marks between two major ticks. `VREF=0` generates a horizontal reference line through the value 0 on the vertical axis.

The autocovariance function γ obviously satisfies $\gamma(0) \geq 0$ and, by the Cauchy-Schwarz inequality

$$\begin{aligned} |\gamma(k)| &= |\operatorname{E}((Y_{t+k} - \operatorname{E}(Y_{t+k}))(Y_t - \operatorname{E}(Y_t)))| \\ &\leq \operatorname{E}(|Y_{t+k} - \operatorname{E}(Y_{t+k})||Y_t - \operatorname{E}(Y_t)|) \\ &\leq \operatorname{Var}(Y_{t+k})^{1/2} \operatorname{Var}(Y_t)^{1/2} \\ &= \gamma(0) \quad \text{for } k \geq 0. \end{aligned}$$

Thus we obtain for the autocovariance function the inequality

$$|\rho(k)| \leq 1 = \rho(0).$$

Variance Stabilizing Transformation

The scatterplot of the points (t, y_t) sometimes shows a variation of the data y_t depending on their height.

Example 1.3.1. (Airline Data). Figure 1.3.2, which displays monthly totals in thousands of international airline passengers from January 1949 to December 1960, exemplifies the above mentioned dependence. These Airline Data are taken from Box and Jenkins (1976); a discussion can be found in Section 9.2 of Brockwell and Davis (1991).

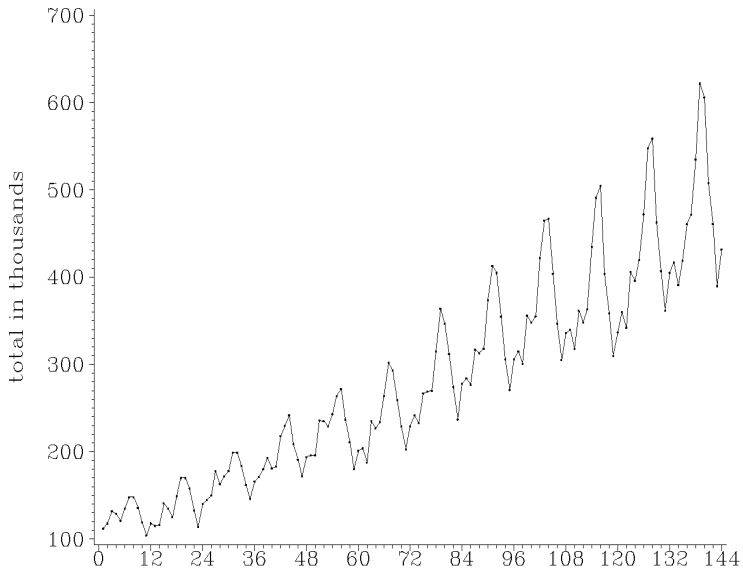


Figure 1.3.2. Monthly totals in thousands of international airline passengers from January 1949 to December 1960.

```
***     Program 1_3_2   ***;
 TITLE1 'Monthly totals from January 49 to December 60';
 TITLE2 'Airline Data';

DATA data1;
  INFILE 'c:\data\airline.txt';
  INPUT y;
  t=_N_;


AXIS1 LABEL=NONE ORDER=(0 12 24 36 48 60 72 84
    96 108 120 132 144)  MINOR=(N=5);
AXIS2 LABEL=(ANGLE=90 'total in thousands');
SYMBOL1 V=DOT C=GREEN I=JOIN H=0.2;
PROC GPLOT DATA=data1;
  PLOT y*t / HAXIS=AXIS1 VAXIS=AXIS2;
RUN; QUIT;
```

In the first data step, the monthly passenger totals are read into the variable `y`. To get a time variable `t`, the temporarily created SAS variable `_N_` is used; it counts the observations. The passenger totals are plotted against `t` with a line joining the data points, which are symbolized by small dots. On the horizontal axis a label is suppressed.

The variation of the data y_t obviously increases with their height. The logtransformed data $x_t = \log(y_t)$, displayed in the following figure, however, show no dependence of variability from height.

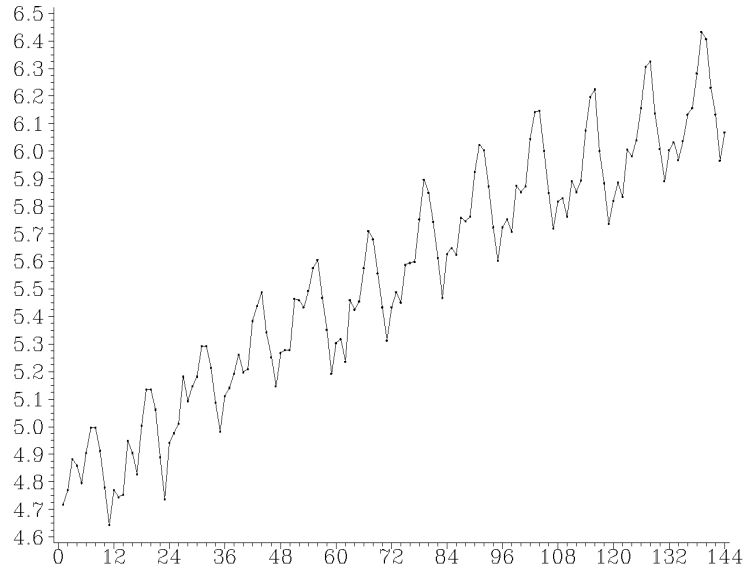


Figure 1.3.3. Logarithm of Airline Data $x_t = \log(y_t)$.

```
*** Program 1_3_3   ***;
TITLE1 'Logarithmic transformation';
TITLE2 'Airline Data';

DATA data1;
  INFILE 'c\data\airline.txt';
  INPUT y;
```

```
  t=_N_;
  x=LOG(y);

AXIS1 LABEL=NONE ORDER=(0 12 24 36 48 60 72 84
                        96 108 120 132 144) MINOR=(N=5);
AXIS2 LABEL=NONE;
SYMBOL1 V=DOT C=GREEN I=JOIN H=0.2;
PROC GPLOT DATA=data1;
  PLOT x*t / HAXIS=AXIS1 VAXIS=AXIS2;
RUN; QUIT;
```

The plot of the log-transformed data is done in the same manner as for the original data in Program 1_3_2. The only differences are the log-transformation by means of the `LOG` function and the suppressed label on the vertical axis.

The fact that taking the logarithm of data often reduces their variability, can be illustrated as follows. Suppose, for example, that the data were generated by random variables, which are of the form $Y_t = \sigma_t Z_t$, where $\sigma_t > 0$ is a scale factor depending on t, and Z_t, $t \in \mathbb{Z}$, are independent copies of a positive random variable Z with variance 1. The variance of Y_t is in this case σ_t^2, whereas the variance of $\log(Y_t) = \log(\sigma_t) + \log(Z_t)$ is a constant, namely the variance of $\log(Z)$, if it exists.

A transformation of the data, which reduces the dependence of the variability on their height, is called *variance stabilizing*. The logarithm is a particular case of the general Box–Cox (1964) transformation T_λ of a time series (Y_t), where the parameter $\lambda \geq 0$ is chosen by the statistician:

$$T_\lambda(Y_t) := \begin{cases} (Y_t^\lambda - 1)/\lambda, & Y_t \geq 0,\ \lambda > 0 \\ \log(Y_t), & Y_t > 0,\ \lambda = 0. \end{cases}$$

Note that $\lim_{\lambda \searrow 0} T_\lambda(Y_t) = T_0(Y_t) = \log(Y_t)$ if $Y_t > 0$ (Exercise 19). Popular choices of the parameter λ are 0 and 1/2. A variance stabilizing transformation of the data, if necessary, usually precedes any further data manipulation such as trend or seasonal adjustment.

Exercises

1. Plot the Mitscherlich function for different values of $\beta_1, \beta_2, \beta_3$ using `PROC GPLOT`.

2. Put in the logistic trend model (1.5) $z_t := 1/y_t \sim 1/\,\mathrm{E}(Y_t) = 1/f_{\log}(t)$, $t = 1, \ldots, n$. Then we have the linear regression model $z_t = a + bz_{t-1} + \varepsilon_t$, where ε_t is the error variable. Compute the least squares estimates $\hat{a}, \hat{b}$ of a, b and motivate the estimates $\hat{\beta}_1 := -\log(\hat{b})$, $\hat{\beta}_3 := (1 - \exp(-\hat{\beta}_1))/\hat{a}$ as well as

$$\hat{\beta}_2 := \exp\Big(\frac{n+1}{2}\hat{\beta}_1 + \frac{1}{n}\sum_{t=1}^{n}\log\Big(\frac{\hat{\beta}_3}{y_t} - 1\Big)\Big),$$

proposed by Tintner (1958); see also the next exercise.

3. The estimate $\hat{\beta}_2$ defined above suffers from the drawback that all observations y_t have to be strictly less than the estimate $\hat{\beta}_3$. Motivate the following substitute of $\hat{\beta}_2$

$$\tilde{\beta}_2 = \Big(\sum_{t=1}^{n}\frac{\hat{\beta}_3 - y_t}{y_t}\exp\big(-\hat{\beta}_1 t\big)\Big)\Big/\sum_{t=1}^{n}\exp\big(-2\hat{\beta}_1 t\big)$$

as an estimate of the parameter β_2 in the logistic trend model (1.5).

4. Show that in a linear regression model $y_t = \beta_1 x_t + \beta_2$, $t = 1, \ldots, n$, the squared multiple correlation coefficient R^2 based on the least squares estimates $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{y}_t := \hat{\beta}_1 x_t + \hat{\beta}_2$ is necessarily between zero and one with $R^2 = 1$ if and only if $\hat{y}_t = y_t, t = 0, \ldots, n$ (see (1.12)).

5. (Population2 Data) The following table lists total population numbers of North Rhine-Westphalia between 1961 and 1979. Suppose a logistic trend for these data and compute the estimators $\hat{\beta}_1, \hat{\beta}_3$ using `PROC REG`. Since some observations exceed $\hat{\beta}_3$, use $\tilde{\beta}_2$ from Exercise 3 and do an ex post-analysis.

Year	t	Total Population in millions
1961	1	15.920
1963	2	16.280
1965	3	16.661
1967	4	16.835
1969	5	17.044
1971	6	17.091
1973	7	17.223
1975	8	17.176
1977	9	17.052
1979	10	17.002

Use `PROC NLIN` and do an ex post-analysis. Compare these two procedures by their residual sums of squares.

6. (Income Data) Suppose an allometric trend function for the income data in Example 1.1.3 and do a regression analysis. Plot the data y_t versus $\hat{\beta}_2 t^{\hat{\beta}_1}$. To this end compute the R^2-coefficient. Estimate the parameters also with `PROC NLIN` and compare the results.

7. (Unemployed2 Data) The following table lists total numbers of unemployed (in thousands) in West Germany between 1950 and 1993. Compare a logistic trend function with an allometric one. Which one gives the better fit?

Year	Unemployed
1950	1869
1960	271
1970	149
1975	1074
1980	889
1985	2304
1988	2242
1989	2038
1990	1883
1991	1689
1992	1808
1993	2270

8. Give an update equation for a simple moving average of (even) order 2s.

9. (Public Expenditures Data) The following table lists West Germany's public expenditures (in billion D-Marks) between 1961 and 1990. Compute simple moving averages of order 3 and 5 to estimate a possible trend. Plot the original data as well as the filtered ones and compare the curves.

Year	Public Expenditures	Year	Public Expenditures
1961	113,4	1976	546,2
1962	129,6	1977	582,7
1963	140,4	1978	620,8
1964	153,2	1979	669,8
1965	170,2	1980	722,4
1966	181,6	1981	766,2
1967	193,6	1982	796,0
1968	211,1	1983	816,4
1969	233,3	1984	849,0
1970	264,1	1985	875,5
1971	304,3	1986	912,3
1972	341,0	1987	949,6
1973	386,5	1988	991,1
1974	444,8	1989	1018,9
1975	509,1	1990	1118,1

10. (Unemployed Females Data) Use `PROC X11` to analyze the monthly unemployed females between ages 16 and 19 in the United States from January 1961 to December 1985 (in thousands).

11. Show that the rank of a matrix $\boldsymbol{A}$ equals the rank of $\boldsymbol{A}^T\boldsymbol{A}$.

12. The $p+1$ columns of the design matrix $\boldsymbol{X}$ in (1.17) are linear independent.

13. Let (c_u) be the moving average derived by the best local polynomial fit. Show that

(i) fitting locally a polynomial of degree 2 to five consecutive data points leads to

$$(c_u) = \frac{1}{35}(-3, 12, 17, 12, -3)^T,$$

(ii) the inverse matrix $\boldsymbol{A}^{-1}$ of an invertible $m \times m$-matrix $\boldsymbol{A} = (a_{ij})_{1\le i,j\le m}$ with the property that $a_{ij} = 0$, if $i + j$ is odd, shares this property,

(iii) (c_u) is symmetric, i.e., $c_{-u} = c_u$.

14. (Unemployed1 Data) Compute a seasonal and trend adjusted time series for the Unemployed1 Data in the building trade. To this end compute seasonal differences and first order differences. Compare the results with those of `PROC X11`.

15. Use the SAS function `RANNOR` to generate a time series $Y_t = b_0+b_1t+\varepsilon_t$, $t = 1, \ldots, 100$, where $b_0, b_1 \neq 0$ and the ε_t are independent normal random variables with mean μ and variance σ_1^2 if $t \le 69$ but variance $\sigma_2^2 \neq \sigma_1^2$ if $t \ge 70$. Plot the exponentially filtered variables Y_t^* for different values of the smoothing parameter $\alpha \in (0, 1)$ and compare the results.

16. Compute under the assumptions of Corollary 1.2.7 the variance of an exponentially filtered variable Y_t^* after a change point $t = N$ with $\sigma^2 := \mathrm{E}(Y_t - \mu)^2$ for $t < N$ and $\tau^2 := \mathrm{E}(Y_t - \lambda)^2$ for $t \ge N$. What is the limit for $t \to \infty$?

17. (Bankruptcy Data) The following table lists the percentages to annual bancruptcies among all US companies between 1867 and 1932:

1.33	0.94	0.79	0.83	0.61	0.77	0.93	0.97	1.20	1.33
1.36	1.55	0.95	0.59	0.61	0.83	1.06	1.21	1.16	1.01
0.97	1.02	1.04	0.98	1.07	0.88	1.28	1.25	1.09	1.31
1.26	1.10	0.81	0.92	0.90	0.93	0.94	0.92	0.85	0.77
0.83	1.08	0.87	0.84	0.88	0.99	0.99	1.10	1.32	1.00
0.80	0.58	0.38	0.49	1.02	1.19	0.94	1.01	1.00	1.01
1.07	1.08	1.04	1.21	1.33	1.53				

Compute and plot the empirical autocovariance function and the empirical autocorrelation function using the SAS procedures `PROC ARIMA` and `PROC GPLOT`.

18. Verify that the empirical correlation $r(k)$ at lag k for the trend $y_t = t$, $t = 1, \ldots, n$ is given by

$$r(k) = 1 - 3\frac{k}{n} + 2\frac{k(k^2-1)}{n(n^2-1)}, \quad k = 0, \ldots, n.$$

Plot the correlogram for different values of n. This example shows, that the correlogram has no interpretation for non-stationary processes (see Exercise 17).

19. Show that

$$\lim_{\lambda \downarrow 0} T_\lambda(Y_t) = T_0(Y_t) = \log(Y_t), \qquad Y_t > 0$$

for the Box-Cox transformation T_λ.

Chapter 2

Models of Time Series

Each time series $Y_1, \ldots, Y_n$ can be viewed as a clipping from a sequence of random variables $\ldots, Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2, \ldots$ In the following we will introduce several models for such a *stochastic process* Y_t with index set $\mathbb{Z}$.

2.1 Linear Filters and Stochastic Processes

For mathematical convenience we will consider *complex valued* random variables Y, whose range is the set of complex numbers $\mathbb{C} = \{u + iv : u, v \in \mathbb{R}\}$, where $i = \sqrt{-1}$. Therefore, we can decompose Y as $Y = Y_{(1)} + iY_{(2)}$, where $Y_{(1)} = \mathrm{Re}(Y)$ is the *real part* of Y and $Y_{(2)} = \mathrm{Im}(Y)$ is its *imaginary part*. The random variable Y is called *integrable* if the real valued random variables $Y_{(1)}, Y_{(2)}$ both have finite expectations, and in this case we define the expectation of Y by

$$\mathrm{E}(Y) := \mathrm{E}(Y_{(1)}) + i\,\mathrm{E}(Y_{(2)}) \in \mathbb{C}.$$

This expectation has, up to monotonicity, the usual properties such as $\mathrm{E}(aY + bZ) = a\,\mathrm{E}(Y) + b\,\mathrm{E}(Z)$ of its real counterpart. Here a and b are complex numbers and Z is a further integrable complex valued random variable. In addition we have $\overline{\mathrm{E}(Y)} = \mathrm{E}(\bar{Y})$, where $\bar{a} = u - iv$ denotes the *conjugate complex number* of $a = u + iv$. Since $|a|^2 := u^2 + v^2 = a\bar{a} = \bar{a}a$, we define the *variance* of Y by

$$\mathrm{Var}(Y) := \mathrm{E}((Y - \mathrm{E}(Y))(\overline{Y - \mathrm{E}(Y)})) \geq 0.$$

The complex random variable Y is called *square integrable* if this number is finite. To carry the equation $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$ for a real random variable X over

to complex ones, we define the *covariance* of complex square integrable random variables Y, Z by

$$\operatorname{Cov}(Y,Z) := \operatorname{E}((Y - \operatorname{E}(Y))(\overline{Z - \operatorname{E}(Z)})).$$

Note that the covariance $\operatorname{Cov}(Y,Z)$ is no longer symmetric with respect to Y and Z, as it is for real valued random variables, but it satisfies $\operatorname{Cov}(Y,Z) = \overline{\operatorname{Cov}(Z,Y)}$.

The following lemma implies that the Cauchy–Schwarz inequality carries over to complex valued random variables.

Lemma 2.1.1. *For any integrable complex valued random variable $Y = Y_{(1)} + iY_{(2)}$ we have*

$$|\operatorname{E}(Y)| \le \operatorname{E}(|Y|) \le \operatorname{E}(|Y_{(1)}|) + \operatorname{E}(|Y_{(2)}|).$$

Proof. We write $\operatorname{E}(Y)$ in polar coordinates $\operatorname{E}(Y) = re^{i\vartheta}$, where $r = |\operatorname{E}(Y)|$ and $\vartheta \in [0, 2\pi)$. Observe that $\operatorname{Re}(e^{-i\vartheta}Y) = \operatorname{Re}\Big((\cos(\vartheta) - i\sin(\vartheta))(Y_{(1)} + iY_{(2)})\Big) = \cos(\vartheta)Y_{(1)} + \sin(\vartheta)Y_{(2)} \le (\cos^2(\vartheta) + \sin^2(\vartheta))^{1/2}(Y_{(1)}^2 + Y_{(2)}^2)^{1/2} = |Y|$ by the Cauchy–Schwarz inequality for real numbers. Thus we obtain

$$\begin{aligned} |\operatorname{E}(Y)| = r &= \operatorname{E}(e^{-i\vartheta}Y) \\ &= \operatorname{E}\Big(\operatorname{Re}(e^{-i\vartheta}Y)\Big) \le \operatorname{E}(|Y|). \end{aligned}$$

The second inequality of the lemma follows from $|Y| = (Y_{(1)}^2 + Y_{(2)}^2)^{1/2} \le |Y_{(1)}| + |Y_{(2)}|$. □

The next result is a consequence of the preceding lemma and the Cauchy–Schwarz inequality for real valued random variables.

Corollary 2.1.2. *For any square integrable complex valued random variable we have*

$$|\operatorname{E}(YZ)| \le \operatorname{E}(|Y||Z|) \le \operatorname{E}(|Y|^2)^{1/2} \operatorname{E}(|Z|^2)^{1/2}$$

and thus,

$$|\operatorname{Cov}(Y,Z)| \le \operatorname{Var}(Y)^{1/2} \operatorname{Var}(Z)^{1/2}.$$

Stationary Processes

A stochastic process $(Y_t)_{t\in\mathbb{Z}}$ of square integrable complex valued random variables is said to be *(weakly) stationary* if for any $t_1, t_2, k \in \mathbb{Z}$

$$\operatorname{E}(Y_{t_1}) = \operatorname{E}(Y_{t_1+k}) \text{ and } \operatorname{E}(Y_{t_1}\overline{Y}_{t_2}) = \operatorname{E}(Y_{t_1+k}\overline{Y}_{t_2+k}).$$

The random variables of a stationary process $(Y_t)_{t\in\mathbb{Z}}$ have identical means and variances. The autocovariance function satisfies moreover for $s, t \in \mathbb{Z}$

$$\begin{aligned}\gamma(t,s) &:= \mathrm{Cov}(Y_t, Y_s) &&= \mathrm{Cov}(Y_{t-s}, Y_0) &&=: \gamma(t-s)\\ &= \overline{\mathrm{Cov}(Y_0, Y_{t-s})} &&= \overline{\mathrm{Cov}(Y_{s-t}, Y_0)} &&= \overline{\gamma(s-t)},\end{aligned}$$

and thus, the autocovariance function of a stationary process can be viewed as a function of a single argument satisfying $\gamma(t) = \overline{\gamma(-t)}, t \in \mathbb{Z}$.

A stationary process $(\varepsilon_t)_{t\in\mathbb{Z}}$ of square integrable and uncorrelated real valued random variables is called *white noise* i.e., $\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$ and there exist $\mu \in \mathbb{R},\ \sigma \geq 0$ such that

$$\mathrm{E}(\varepsilon_t) = \mu,\ \mathrm{E}((\varepsilon_t - \mu)^2) = \sigma^2, \qquad t \in \mathbb{Z}.$$

In Section 1.2 we defined linear filters of a time series, which were based on a finite number of real valued weights. In the following we consider linear filters with an infinite number of complex valued weights.

Suppose that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise and let $(a_t)_{t\in\mathbb{Z}}$ be a sequence of complex numbers satisfying $\sum_{t=-\infty}^{\infty} |a_t| := \sum_{t\geq 0} |a_t| + \sum_{t\geq 1} |a_{-t}| < \infty$. Then $(a_t)_{t\in\mathbb{Z}}$ is said to be an *absolutely summable (linear) filter* and

$$Y_t := \sum_{u=-\infty}^{\infty} a_u \varepsilon_{t-u} := \sum_{u\geq 0} a_u \varepsilon_{t-u} + \sum_{u\geq 1} a_{-u} \varepsilon_{t+u}, \qquad t \in \mathbb{Z},$$

is called a *general linear process.*

Existence of General Linear Processes

We will show that $\sum_{u=-\infty}^{\infty} |a_u \varepsilon_{t-u}| < \infty$ with probability one for $t \in \mathbb{Z}$ and, thus, $Y_t = \sum_{u=-\infty}^{\infty} a_u \varepsilon_{t-u}$ is well defined. Denote by $L_2 := L_2(\Omega, \mathcal{A}, \mathcal{P})$ the set of all complex valued square integrable random variables, defined on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$, and put $||Y||_2 := \mathrm{E}(|Y|^2)^{1/2}$, which is the L_2-pseudonorm on L_2.

Lemma 2.1.3. *Let $X_n,\ n \in \mathbb{N}$, be a sequence in L_2 such that $||X_{n+1} - X_n||_2 \leq 2^{-n}$ for each $n \in \mathbb{N}$. Then there exists $X \in L_2$ such that $\lim_{n\to\infty} X_n = X$ with probability one.*

Proof. Write $X_n = \sum_{k\leq n}(X_k - X_{k-1})$, where $X_0 := 0$. By the monotone conver-

gence theorem, the Cauchy–Schwarz inequality and Corollary 2.1.2 we have

$$\begin{aligned}\mathrm{E}\Big(\sum_{k\geq 1}|X_k - X_{k-1}|\Big) &= \sum_{k\geq 1}\mathrm{E}(|X_k - X_{k-1}|) \leq \sum_{k\geq 1}||X_k - X_{k-1}||_2 \\ &\leq ||X_1||_2 + \sum_{k\geq 1} 2^{-k} < \infty.\end{aligned}$$

This implies that $\sum_{k\geq 1}|X_k - X_{k-1}| < \infty$ with probability one and hence, the limit $\lim_{n\to\infty}\sum_{k\leq n}(X_k - X_{k-1}) = \lim_{n\to\infty} X_n = X$ exists in $\mathbb{C}$ almost surely. Finally, we check that $X \in L_2$:

$$\begin{aligned}\mathrm{E}(|X|^2) &= \mathrm{E}(\lim_{n\to\infty}|X_n|^2) \\ &\leq \mathrm{E}\Big(\lim_{n\to\infty}\Big(\sum_{k\leq n}|X_k - X_{k-1}|\Big)^2\Big) \\ &= \lim_{n\to\infty}\mathrm{E}\Big(\Big(\sum_{k\leq n}|X_k - X_{k-1}|\Big)^2\Big) \\ &= \lim_{n\to\infty}\sum_{k,j\leq n}\mathrm{E}(|X_k - X_{k-1}|\,|X_j - X_{j-1}|) \\ &\leq \limsup_{n\to\infty}\sum_{k,j\leq n}||X_k - X_{k-1}||_2\,||X_j - X_{j-1}||_2 \\ &= \limsup_{n\to\infty}\Big(\sum_{k\leq n}||X_k - X_{k-1}||_2\Big)^2 \\ &= \Big(\sum_{k\geq 1}||X_k - X_{k-1}||_2\Big) < \infty.\end{aligned}$$

□

Theorem 2.1.4. *The space $(L_2, ||\cdot||_2)$ is complete i.e., suppose that $X_n \in L_2$, $n \in \mathbb{N}$, has the property that for arbitrary $\varepsilon > 0$ one can find an integer $N(\varepsilon) \in \mathbb{N}$ such that $||X_n - X_m||_2 < \varepsilon$ if $n,\ m \geq N(\varepsilon)$. Then there exists a random variable $X \in L_2$ such that $\lim_{n\to\infty}||X - X_n||_2 = 0$.*

Proof. We can obviously find integers $n_1 < n_2 < \ldots$ such that

$$||X_n - X_m||_2 \leq 2^{-k} \qquad \text{if } n, m \geq n_k.$$

By Lemma 2.1.3 there exists a random variable $X \in L_2$ such that $\lim_{k\to\infty} X_{n_k} = X$ with probability one. Fatou's lemma implies

$$\begin{aligned}||X_n - X||_2^2 &= \mathrm{E}(|X_n - X|^2) \\ &= \mathrm{E}\Big(\liminf_{k\to\infty}|X_n - X_{n_k}|^2\Big) \leq \liminf_{k\to\infty}||X_n - X_{n_k}||_2^2.\end{aligned}$$

The right-hand side of this inequality becomes arbitrarily small if we choose n large enough, and thus we have $\lim_{n\to\infty} ||X_n - X||_2^2 = 0$. □

The following result implies in particular that a general linear process is well defined.

Theorem 2.1.5. *Suppose that $(Z_t)_{t\in\mathbb{Z}}$ is a complex valued stochastic process such that $\sup_t \mathrm{E}(|Z_t|) < \infty$ and let $(a_t)_{t\in\mathbb{Z}}$ be an absolutely summable filter. Then we have $\sum_{u\in\mathbb{Z}} |a_u Z_{t-u}| < \infty$ with probability one for $t \in \mathbb{Z}$ and, thus, $Y_t := \sum_{u\in\mathbb{Z}} a_u Z_{t-u}$ exists almost surely in $\mathbb{C}$. We have moreover $\mathrm{E}(|Y_t|) < \infty$, $t \in \mathbb{Z}$, and*

(i) $\mathrm{E}(Y_t) = \lim_{n\to\infty} \sum_{u=-n}^{n} a_u \mathrm{E}(Z_{t-u})$, $t \in \mathbb{Z}$,

(ii) $\mathrm{E}(|Y_t - \sum_{u=-n}^{n} a_u Z_{t-u}|) \overset{n\to\infty}{\longrightarrow} 0$.

If, in addition, $\sup_t \mathrm{E}(|Z_t|^2) < \infty$, then we have $\mathrm{E}(|Y_t|^2) < \infty$, $t \in \mathbb{Z}$, and

(iii) $||Y_t - \sum_{u=-n}^{n} a_u Z_{t-u}||_2 \overset{n\to\infty}{\longrightarrow} 0$.

Proof. The monotone convergence theorem implies

$$\mathrm{E}\Big(\sum_{u\in\mathbb{Z}} |a_u|\,|Z_{t-u}|\Big) \le \lim_{n\to\infty} \Big(\sum_{u=-n}^{n} |a_u|\Big) \sup_{t\in\mathbb{Z}} \mathrm{E}(|Z_{t-u}|) < \infty$$

and, thus, we have $\sum_{u\in\mathbb{Z}} |a_u||Z_{t-u}| < \infty$ with probability one as well as $\mathrm{E}(|Y_t|) \le \mathrm{E}(\sum_{u\in\mathbb{Z}} |a_u||Z_{t-u}|) < \infty$, $t \in \mathbb{Z}$. Put $X_n(t) := \sum_{u=-n}^{n} a_u Z_{t-u}$. Then we have $|Y_t - X_n(t)| \to_{n\to\infty} 0$ almost surely. By the inequality $|Y_t - X_n(t)| \le 2\sum_{u\in\mathbb{Z}} |a_u||Z_{t-u}|$, $n \in \mathbb{N}$, the dominated convergence theorem implies (ii) and therefore (i):

$$\begin{aligned} |\,\mathrm{E}(Y_t) - \sum_{u=-n}^{n} a_u \mathrm{E}(Z_{t-u})| \;&=\; |\,\mathrm{E}(Y_t) - \mathrm{E}(X_n(t))| \\ &\le\; \mathrm{E}(|Y_t - X_n(t)|) \longrightarrow_{n\to\infty} 0. \end{aligned}$$

Put $K := \sup_t \mathrm{E}(|Z_t|^2) < \infty$. The Cauchy–Schwarz inequality implies for $m, n \in \mathbb{N}$

and $\varepsilon > 0$

$$\begin{aligned}
\mathrm{E}(|X_{n+m}(t) - X_n(t)|^2) &= \mathrm{E}\left(\left|\sum_{|u|=n+1}^{n+m} a_u Z_{t-u}\right|^2\right) \\
&= \sum_{|u|=n+1}^{n+m} \sum_{|w|=n+1}^{n+m} a_u \bar{a}_w \,\mathrm{E}(Z_{t-u}\bar{Z}_{t-w}) \\
&\le K^2 \left(\sum_{|u|=n+1}^{n+m} |a_u|\right)^2 \le K^2 \left(\sum_{|u|\ge n} |a_u|\right)^2 < \varepsilon
\end{aligned}$$

if n is chosen sufficiently large. Theorem 2.1.4 now implies the existence of a random variable $X(t) \in L_2$ with $\lim_{n\to\infty} ||X_n(t) - X(t)||_2 = 0$. For the proof of (iii) it remains to show that $X(t) = Y_t$ almost surely. Markov's inequality implies

$$P\{|Y_t - X_n(t)| \ge \varepsilon\} \le \varepsilon^{-1} \,\mathrm{E}(|Y_t - X_n(t)|) \longrightarrow_{n\to\infty} 0$$

by (ii), and Chebyshev's inequality yields

$$P\{|X(t) - X_n(t)| \ge \varepsilon\} \le \varepsilon^{-2} ||X(t) - X_n(t)||_2 \longrightarrow_{n\to\infty} 0$$

for arbitrary $\varepsilon > 0$. This implies

$$\begin{aligned}
P\{|Y_t - X(t)| \ge \varepsilon\} &\le P\{|Y_t - X_n(t)| + |X_n(t) - X(t)| \ge \varepsilon\} \\
&\le P\{|Y_t - X_n(t)| \ge \varepsilon/2\} + P\{|X(t) - X_n(t)| \ge \varepsilon/2\} \longrightarrow_{n\to\infty} 0
\end{aligned}$$

and thus $Y_t = X(t)$ almost surely, which completes the proof of Theorem 2.1.5. □

Theorem 2.1.6. *Suppose that $(Z_t)_{t\in\mathbb{Z}}$ is a stationary process with mean $\mu_Z := \mathrm{E}(Z_0)$ and autocovariance function γ_Z and let (a_t) be an absolutely summable filter. Then $Y_t = \sum_u a_u Z_{t-u}$, $t \in \mathbb{Z}$, is also stationary with*

$$\mu_Y = \mathrm{E}(Y_0) = \Big(\sum_u a_u\Big)\mu_Z$$

and autocovariance function

$$\gamma_Y(t) = \sum_u \sum_w a_u \bar{a}_w \gamma_Z(t + w - u).$$

Proof. Note that

$$\begin{aligned}
\mathrm{E}(|Z_t|^2) &= \mathrm{E}\left(|Z_t - \mu_Z + \mu_Z|^2\right) \\
&= \mathrm{E}\left((Z_t - \mu_Z + \mu_Z)(\overline{Z_t - \mu_z} + \overline{\mu_z})\right) \\
&= \mathrm{E}\left(|Z_t - \mu_Z|^2\right) + |\mu_Z|^2 \\
&= \gamma_Z(0) + |\mu_Z|^2
\end{aligned}$$

and, thus,

$$\sup_{t\in\mathbb{Z}} \mathrm{E}(|Z_t|^2) < \infty.$$

We can, therefore, now apply Theorem 2.1.5. Part (i) of Theorem 2.1.5 immediately implies $\mathrm{E}(Y_t) = (\sum_u a_u)\mu_Z$ and part (iii) implies for $t, s \in \mathbb{Z}$

$$\begin{aligned}
\mathrm{E}((Y_t - \mu_Y)(\overline{Y_s - \mu_Y})) &= \lim_{n\to\infty} \mathrm{Cov}\Big(\sum_{u=-n}^{n} a_u Z_{t-u}, \sum_{w=-n}^{n} a_w Z_{s-w}\Big) \\
&= \lim_{n\to\infty} \sum_{u=-n}^{n}\sum_{w=-n}^{n} a_u \bar{a}_w \ \mathrm{Cov}(Z_{t-u}, Z_{s-w}) \\
&= \lim_{n\to\infty} \sum_{u=-n}^{n}\sum_{w=-n}^{n} a_u \bar{a}_w \gamma_Z(t-s+w-u) \\
&= \sum_u \sum_w a_u \bar{a}_w \gamma_Z(t-s+w-u).
\end{aligned}$$

The covariance of Y_t and Y_s depends, therefore, only on the difference $t - s$. Note that $|\gamma_Z(t)| \le \gamma_Z(0) < \infty$ and thus, $\sum_u \sum_w |a_u \bar{a}_w \gamma_Z(t-s+w-u)| \le \gamma_Z(0)(\sum_u |a_u|)^2 < \infty$, i.e., (Y_t) is a stationary process. □

The Covariance Generating Function

The *covariance generating function* of a stationary process with autocovariance function γ is defined as the double series

$$G(z) := \sum_{t\in\mathbb{Z}} \gamma(t) z^t = \sum_{t\ge 0} \gamma(t) z^t + \sum_{t\ge 1} \gamma(-t) z^{-t},$$

known as a *Laurent* series in complex analysis. We assume that there exists a real number $r > 1$ such that $G(z)$ is defined for all $z \in \mathbb{C}$ in the annulus $1/r < |z| < r$. The covariance generating function will help us to compute the autocovariances of filtered processes.

Since the coefficients of a Laurent series are uniquely determined (see e.g. Chapter V, 1.11 in Conway (1975)), the covariance generating function of a stationary process is a constant function if and only if this process is a white noise i.e., $\gamma(t) = 0$ for $t \neq 0$.

Theorem 2.1.7. *Suppose that $Y_t = \sum_u a_u \varepsilon_{t-u}$, $t \in \mathbb{Z}$, is a general linear process with $\sum_u |a_u||z^u| < \infty$, if $r^{-1} < |z| < r$ for some $r > 1$. Put $\sigma^2 := \mathrm{Var}(\varepsilon_0)$. The process (Y_t) then has the covariance generating function*

$$G(z) = \sigma^2 \Big(\sum_u a_u z^u\Big)\Big(\sum_u \bar{a}_u z^{-u}\Big), \qquad r^{-1} < |z| < r.$$

Proof. Theorem 2.1.6 implies for $t \in \mathbb{Z}$

$$\begin{aligned}\operatorname{Cov}(Y_t, Y_0) &= \sum_u \sum_w a_u \bar{a}_w \gamma_\varepsilon(t + w - u) \\ &= \sigma^2 \sum_u a_u \bar{a}_{u-t}.\end{aligned}$$

This implies

$$\begin{aligned}G(z) &= \sigma^2 \sum_t \sum_u a_u \bar{a}_{u-t} z^t \\ &= \sigma^2 \Big(\sum_u |a_u|^2 + \sum_{t \geq 1} \sum_u a_u \bar{a}_{u-t} z^t + \sum_{t \leq -1} \sum_u a_u \bar{a}_{u-t} z^t \Big) \\ &= \sigma^2 \Big(\sum_u |a_u|^2 + \sum_u \sum_{t \leq u-1} a_u \bar{a}_t z^{u-t} + \sum_u \sum_{t \geq u+1} a_u \bar{a}_t z^{u-t} \Big) \\ &= \sigma^2 \sum_u \sum_t a_u \bar{a}_t z^{u-t} = \sigma^2 \Big(\sum_u a_u z^u \Big) \Big(\sum_t \bar{a}_t z^{-t} \Big).\end{aligned}$$

□

Example 2.1.8. Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be a white noise with $\operatorname{Var}(\varepsilon_0) =: \sigma^2 > 0$. The covariance generating function of the simple moving average $Y_t = \sum_u a_u \varepsilon_{t-u}$ with $a_{-1} = a_0 = a_1 = 1/3$ and $a_u = 0$ elsewhere is then given by

$$\begin{aligned}G(z) &= \frac{\sigma^2}{9}(z^{-1} + z^0 + z^1)(z^1 + z^0 + z^{-1}) \\ &= \frac{\sigma^2}{9}(z^{-2} + 2z^{-1} + 3z^0 + 2z^1 + z^2), \quad z \in \mathbb{R}.\end{aligned}$$

Then the autocovariances are just the coefficients in the above series

$$\gamma(0) = \frac{\sigma^2}{3},\ \gamma(1) = \gamma(-1) = \frac{2\sigma^2}{9},\ \gamma(2) = \gamma(-2) = \frac{\sigma^2}{9},\ \gamma(k) = 0 \text{ elsewhere.}$$

This explains the name covariance generating function.

The Characteristic Polynomial

Let (a_u) be an absolutely summable filter. The Laurent series

$$A(z) := \sum_{u \in \mathbb{Z}} a_u z^u$$

is called *characteristic polynomial* of (a_u). We know from complex analysis that $A(z)$ exists either for all z in some annulus $r < |z| < R$ or almost nowhere. In the

first case the coefficients a_u are uniquely determined by the function $A(z)$ (see e.g. Chapter V, 1.11 in Conway (1975)).

If, for example, (a_u) is absolutely summable with $a_u = 0$ for $u \geq 1$, then $A(z)$ exists for all complex z such that $|z| \geq 1$. If $a_u = 0$ for all large $|u|$, then $A(z)$ exists for all $z \neq 0$.

Inverse Filters

Let now (a_u) and (b_u) be absolutely summable filters and denote by $Y_t := \sum_u a_u Z_{t-u}$, the filtered stationary sequence, where $(Z_u)_{u\in\mathbb{Z}}$ is a stationary process. Filtering $(Y_t)_{t\in\mathbb{Z}}$ by means of (b_u) leads to

$$\sum_w b_w Y_{t-w} = \sum_w \sum_u b_w a_u Z_{t-w-u} = \sum_v (\sum_{u+w=v} b_w a_u) Z_{t-v},$$

where $c_v := \sum_{u+w=v} b_w a_u$, $v \in \mathbb{Z}$, is an absolutely summable filter:

$$\sum_v |c_v| \leq \sum_v \sum_{u+w=v} |b_w a_u| = (\sum_u |a_u|)(\sum_w |b_w|) < \infty.$$

We call (c_v) the *product filter* of (a_u) and (b_u).

Lemma 2.1.9. *Let (a_u) and (b_u) be absolutely summable filters with characteristic polynomials $A_1(z)$ and $A_2(z)$, which both exist on some annulus $r < |z| < R$. The product filter $(c_v) = (\sum_{u+w=v} b_w a_u)$ then has the characteristic polynomial*

$$A(z) = A_1(z)A_2(z).$$

Proof. By repeating the above arguments we obtain

$$A(z) = \sum_v \Big(\sum_{u+w=v} b_w a_u \Big) z^v = A_1(z)\, A_2(z).$$

□

Suppose now that (a_u) and (b_u) are absolutely summable filters with characteristic polynomials $A_1(z)$ and $A_2(z)$, which both exist on some annulus $r < z < R$, where they satisfy $A_1(z)A_2(z) = 1$. Since $1 = \sum_v c_v z^v$ if $c_0 = 1$ and $c_v = 0$ elsewhere, the uniquely determined coefficients of the characteristic polynomial of the product filter of (a_u) and (b_u) are given by

$$\sum_{u+w=v} b_w a_u = \begin{cases} 1 & \text{if } v = 0 \\ 0 & \text{if } v \neq 0. \end{cases}$$

In this case we obtain for a stationary process (Z_t) that almost surely

$$Y_t = \sum_u a_u Z_{t-u} \quad \text{and} \quad \sum_w b_w Y_{t-w} = Z_t, \quad t \in \mathbb{Z}. \tag{2.1}$$

The filter (b_u) is, therefore, called the *inverse filter* of (a_u).

Causal Filters

An absolutely summable filter $(a_u)_{u\in\mathbb{Z}}$ is called *causal* if $a_u = 0$ for $u < 0$.

Lemma 2.1.10. *Let $a \in \mathbb{C}$. The filter (a_u) with $a_0 = 1$, $a_1 = -a$ and $a_u = 0$ elsewhere has an absolutely summable and causal inverse filter $(b_u)_{u\geq 0}$ if and only if $|a| < 1$. In this case we have $b_u = a^u$, $u \geq 0$.*

Proof. The characteristic polynomial of (a_u) is $A_1(z) = 1 - az$, $z \in \mathbb{C}$. Since the characteristic polynomial $A_2(z)$ of an inverse filter satisfies $A_1(z)A_2(z) = 1$ on some annulus, we have $A_2(z) = 1/(1 - az)$. Observe now that

$$\frac{1}{1-az} = \sum_{u\geq 0} a^u z^u, \qquad \text{if } |z| < 1/|a|.$$

As a consequence, if $|a| < 1$, then $A_2(z) = \sum_{u\geq 0} a^u z^u$ exists for all $|z| < 1$ and the inverse causal filter $(a^u)_{u\geq 0}$ is absolutely summable, i.e., $\sum_{u\geq 0} |a^u| < \infty$. If $|a| \geq 1$, then $A_2(z) = \sum_{u\geq 0} a^u z^u$ exists for all $|z| < 1/|a|$, but $\sum_{u\geq 0} |a|^u = \infty$, which completes the proof. □

Theorem 2.1.11. *Let $a_1, a_2, \ldots, a_p \in \mathbb{C}$, $a_p \neq 0$. The filter (a_u) with coefficients $a_0 = 1, a_1, \ldots, a_p$ and $a_u = 0$ elsewhere has an absolutely summable and causal inverse filter if the p roots $z_1, \ldots, z_p \in \mathbb{C}$ of $A(z) = 1 + a_1 z + a_2 z^2 + \ldots + a_p z^p = 0$ are outside of the unit circle i.e., $|z_i| > 1$ for $1 \leq i \leq p$.*

Proof. We know from the Fundamental Theorem of Algebra that the equation $A(z) = 0$ has exactly p roots $z_1, \ldots, z_p \in \mathbb{C}$ (see e.g. Chapter IV, 3.5 in Conway (1975)), which are all different from zero, since $A(0) = 1$. Hence we can write (see e.g. Chapter IV, 3.6 in Conway (1975))

$$\begin{aligned} A(z) &= a_p(z - z_1)\cdots(z - z_p) \\ &= c\Big(1 - \frac{z}{z_1}\Big)\Big(1 - \frac{z}{z_2}\Big)\cdots\Big(1 - \frac{z}{z_p}\Big), \end{aligned}$$

where $c := a_p(-1)^p z_1 \cdots z_p$. In case of $|z_i| > 1$ we can write for $|z| < |z_i|$

$$\frac{1}{1 - \frac{z}{z_i}} = \sum_{u\geq 0} \Big(\frac{1}{z_i}\Big)^u z^u,$$

where the coefficients $(1/z_i)^u,\ u \geq 0$, are absolutely summable. In case of $|z_i| < 1$, we have for $|z| > |z_i|$

$$\frac{1}{1-\frac{z}{z_i}} = -\frac{1}{\frac{z}{z_i}}\,\frac{1}{1-\frac{z_i}{z}} = -\frac{z_i}{z}\sum_{u\geq 0} z_i^u z^{-u} = -\sum_{u\leq -1}\Big(\frac{1}{z_i}\Big)^u z^u,$$

where the filter with coefficients $-(1/z_i)^u,\ u \leq -1$, is not a causal one. In case of $|z_i| = 1$, we have for $|z| < 1$

$$\frac{1}{1-\frac{z}{z_i}} = \sum_{u\geq 0}\Big(\frac{1}{z_i}\Big)^u z^u,$$

where the coefficients $(1/z_i)^u,\ u \geq 0$, are not absolutely summable. Since the coefficients of a Laurent series are uniquely determined, the factor $1 - z/z_i$ has an inverse $1/(1-z/z_i) = \sum_{u\geq 0} b_u z^u$ on some annulus with $\sum_{u\geq 0} |b_u| < \infty$ if $|z_i| > 1$. A small analysis implies that this argument carries over to the product

$$\frac{1}{A(z)} = \frac{1}{c\Big(1-\frac{z}{z_1}\Big)\cdots\Big(1-\frac{z}{z_p}\Big)}$$

which has an expansion $1/A(z) = \sum_{u\geq 0} b_u z^u$ on some annulus with $\sum_{u\geq 0} |b_u| < \infty$ if each factor has such an expansion, and thus, the proof is complete. □

Remark 2.1.12. Note that the roots $z_1, \ldots, z_p$ of $A(z) = 1 + a_1 z + \ldots + a_p z^p$ are complex valued and thus, the coefficients b_u of the inverse causal filter will, in general, be complex valued as well. The preceding proof shows, however, that if a_p and each z_i are real numbers, then the coefficients b_u, $u \geq 0$, are real as well.

The preceding proof shows, moreover, that a filter (a_u) with complex coefficients $a_0, a_1, \ldots, a_p \in \mathbb{C}$ and $a_u = 0$ elsewhere has an absolutely summable inverse filter if *no* root $z \in \mathbb{C}$ of the equation $A(z) = a_0 + a_1 z + \ldots + a_p z^p = 0$ has length 1 i.e., $|z| \neq 1$ for each root. The additional condition $|z| > 1$ for each root then implies that the inverse filter is a causal one.

Example 2.1.13. The filter with coefficients $a_0 = 1, a_1 = -0.7$ and $a_2 = 0.1$ has the characteristic polynomial $A(z) = 1 - 0.7z + 0.1z^2 = 0.1(z-2)(z-5)$, with $z_1 = 2, z_2 = 5$ being the roots of $A(z) = 0$. Theorem 2.1.11 implies the existence of an absolutely summable inverse causal filter, whose coefficients can be obtained

by expanding $1/A(z)$ as a power series of z:

$$\begin{aligned}\frac{1}{A(z)} &= \frac{1}{\left(1-\frac{z}{2}\right)\left(1-\frac{z}{5}\right)} = \sum_{u\geq 0}\left(\frac{1}{2}\right)^u z^u \sum_{w\geq 0}\left(\frac{1}{5}\right)^w z^w \\ &= \sum_{v\geq 0}\sum_{u+w=v}\left(\frac{1}{2}\right)^u\left(\frac{1}{5}\right)^w z^v \\ &= \sum_{v\geq 0}\sum_{w=0}^{v}\left(\frac{1}{2}\right)^{v-w}\left(\frac{1}{5}\right)^w z^v \\ &= \sum_{v\geq 0}\left(\frac{1}{2}\right)^v \frac{1-\left(\frac{2}{5}\right)^{v+1}}{1-\frac{2}{5}} z^v = \sum_{v\geq 0}\frac{10}{3}\left(\left(\frac{1}{2}\right)^{v+1}-\left(\frac{1}{5}\right)^{v+1}\right)z^v.\end{aligned}$$

The preceding expansion implies that $b_v := (10/3)(2^{-(v+1)} - 5^{-(v+1)}),\ v \geq 0$, are the coefficients of the inverse causal filter.

2.2 Moving Averages and Autoregressive Processes

Let $a_1, \ldots, a_q \in \mathbb{R}$ with $a_q \neq 0$ and let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise. The process

$$Y_t := \varepsilon_t + a_1\varepsilon_{t-1} + \cdots + a_q\varepsilon_{t-q}$$

is said to be a *moving average* of order q, denoted by $MA(q)$. Put $a_0 = 1$. Theorem 2.1.6 and 2.1.7 imply that a moving average $Y_t = \sum_{u=0}^{q} a_u\varepsilon_{t-u}$ is a stationary process with covariance generating function

$$\begin{aligned}G(z) &= \sigma^2\Big(\sum_{u=0}^{q} a_u z^u\Big)\Big(\sum_{w=0}^{q} a_w z^{-w}\Big) \\ &= \sigma^2\sum_{u=0}^{q}\sum_{w=0}^{q} a_u a_w z^{u-w} \\ &= \sigma^2\sum_{v=-q}^{q}\sum_{u-w=v} a_u a_w z^{u-w} \\ &= \sigma^2\sum_{v=-q}^{q}\Big(\sum_{w=0}^{q-v} a_{v+w}a_w\Big)z^v, \qquad z\in\mathbb{C},\end{aligned}$$

where $\sigma^2 = \mathrm{Var}(\varepsilon_0)$. The coefficients of this expansion provide the autocovariance function $\gamma(v) = \mathrm{Cov}(Y_0, Y_v), v \in \mathbb{Z}$, which cuts off after lag q.

Lemma 2.2.1. *Suppose that* $Y_t = \sum_{u=0}^{q} a_u \varepsilon_{t-u}$, $t \in \mathbb{Z}$, *is a* $MA(q)$*-process. Put* $\mu := E(\varepsilon_0)$ *and* $\sigma^2 := Var(\varepsilon_0)$. *Then we have*

(i) $\mathrm{E}(Y_t) = \mu \sum_{u=0}^{q} a_u$,

(ii) $$\gamma(v) = \mathrm{Cov}(Y_v, Y_0) = \begin{cases} 0 & , \quad v > q, \\ \sigma^2 \sum\limits_{w=0}^{q-v} a_{v+w} a_w & , \quad 0 \le v \le q, \end{cases}$$
$\gamma(-v) = \gamma(v)$,

(iii) $\mathrm{Var}(Y_0) = \gamma(0) = \sigma^2 \sum_{w=0}^{q} a_w^2$,

(iv) $$\rho(v) = \frac{\gamma(v)}{\gamma(0)} = \begin{cases} 0 & , \quad v > q, \\ \Big(\sum\limits_{w=0}^{q-v} a_{v+w} a_w \Big) \Big/ \Big(\sum_{w=0}^{q} a_w^2 \Big) & , \quad 0 < v \le q, \\ 1 & , \quad v = 0, \end{cases}$$
$\rho(-v) = \rho(v)$.

Example 2.2.2. The $MA(1)$-process $Y_t = \varepsilon_t + a\varepsilon_{t-1}$ with $a \neq 0$ has the autocorrelation function

$$\rho(v) = \begin{cases} 1 & , \ v = 0 \\ a/(1+a^2) & , \ v = \pm 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Since $a/(1+a^2) = (1/a)/(1+(1/a)^2)$, the autocorrelation functions of the two $MA(1)$-processes with parameters a and $1/a$ coincide. We have, moreover, $|\rho(1)| \le 1/2$ for an arbitrary $MA(1)$-process and thus, a large value of the empirical autocorrelation function $r(1)$, which exceeds $1/2$ essentially, might indicate that an $MA(1)$-model for a given data set is not a correct assumption.

Invertible Processes

The $MA(q)$-process $Y_t = \sum_{u=0}^{q} a_u \varepsilon_{t-u}$, with $a_0 = 1$ and $a_q \neq 0$, is said to be *invertible* if all q roots $z_1, \ldots, z_q \in \mathbb{C}$ of $A(z) = \sum_{u=0}^{q} a_u z^u = 0$ are outside of the unit circle i.e., if $|z_i| > 1$ for $1 \le i \le q$.

Theorem 2.1.11 and representation (2.1) imply that the white noise process (ε_t), pertaining to an invertible $MA(q)$-process $Y_t = \sum_{u=0}^{q} a_u \varepsilon_{t-u}$, can be obtained by means of an absolutely summable and causal filter $(b_u)_{u \ge 0}$ via

$$\varepsilon_t = \sum_{u \ge 0} b_u Y_{t-u}, \qquad t \in \mathbb{Z},$$

with probability one. In particular the $MA(1)$-process $Y_t = \varepsilon_t - a\varepsilon_{t-1}$ is invertible iff $|a| < 1$, and in this case we have by Lemma 2.1.10 with probability one

$$\varepsilon_t = \sum_{u \geq 0} a^u Y_{t-u}, \qquad t \in \mathbb{Z}.$$

Autoregressive Processes

A real valued stochastic process (Y_t) is said to be an *autoregressive process of order* p, denoted by $AR(p)$ if there exist $a_1, \ldots, a_p \in \mathbb{R}$ with $a_p \neq 0$, and a white noise (ε_t) such that

$$Y_t = a_1 Y_{t-1} + \ldots + a_p Y_{t-p} + \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{2.2}$$

The value of an $AR(p)$-process at time t is, therefore, regressed on its own past p values plus a random shock.

The Stationarity Condition

While by Theorem 2.1.6 $MA(q)$-processes are automatically stationary, this is not true for $AR(p)$-processes (see Exercise 26). The following result provides a sufficient condition on the constants $a_1, \ldots, a_p$ implying the existence of a uniquely determined stationary solution (Y_t) of (2.2).

Theorem 2.2.3. *The $AR(p)$-equation (2.2) with given constants $a_1, \ldots, a_p$ and white noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ has a stationary solution $(Y_t)_{t\in\mathbb{Z}}$ if all p roots of the equation $1 - a_1 z - a_2 z^2 - \ldots - a_p z^p = 0$ are outside of the unit circle. In this case, the stationary solution is almost surely uniquely determined by*

$$Y_t := \sum_{u \geq 0} b_u \varepsilon_{t-u}, \qquad t \in \mathbb{Z},$$

where $(b_u)_{u\geq 0}$ is the absolutely summable inverse causal filter of $c_0 = 1$, $c_u = -a_u$, $u = 1, \ldots, p$ and $c_u = 0$ elsewhere.

Proof. The existence of an absolutely summable causal filter follows from Theorem 2.1.11. The stationarity of $Y_t = \sum_{u\geq 0} b_u \varepsilon_{t-u}$ is a consequence of Theorem 2.1.6, and its uniqueness follows from

$$\varepsilon_t = Y_t - a_1 Y_{t-1} - \ldots - a_p Y_{t-p}, \qquad t \in \mathbb{Z},$$

and equation (2.1). □

The condition that all roots of the characteristic equation of an $AR(p)$-process $Y_t = \sum_{u=1}^{p} a_u Y_{t-u} + \varepsilon_t$ are outside of the unit circle i.e.,

$$1 - a_1 z - a_2 z^2 - \ldots - a_p z^p \neq 0 \text{ for } |z| \leq 1, \tag{2.3}$$

will be referred to in the following as the *stationarity condition* for an $AR(p)$-process.

Note that a stationary solution (Y_t) of (2.1) exists in general if no root z_i of the characteristic equation lies on the unit sphere. If there are solutions in the unit circle, then the stationary solution is noncausal, i.e., Y_t is correlated with future values of ε_s, $s > t$. This is frequently regarded as unnatural.

Example 2.2.4. The $AR(1)$-process $Y_t = aY_{t-1} + \varepsilon_t,\ t \in \mathbb{Z}$, with $a \neq 0$ has the characteristic equation $1 - az = 0$ with the obvious solution $z_1 = 1/a$. The process (Y_t), therefore, satisfies the stationarity condition iff $|z_1| > 1$ i.e., iff $|a| < 1$. In this case we obtain from Lemma 2.1.10 that the absolutely summable inverse causal filter of $a_0 = 1,\ a_1 = -a$ and $a_u = 0$ elsewhere is given by $b_u = a^u,\ u \geq 0$, and thus, with probability one

$$Y_t = \sum_{u \geq 0} b_u \varepsilon_{t-u} = \sum_{u \geq 0} a^u \varepsilon_{t-u}.$$

Denote by σ^2 the variance of ε_0. From Theorem 2.1.6 we obtain the autocovariance function of (Y_t)

$$\begin{aligned} \gamma(s) &= \sum_u \sum_w b_u b_w \operatorname{Cov}(\varepsilon_0, \varepsilon_{s+w-u}) \\ &= \sum_{u \geq 0} b_u b_{u-s} \operatorname{Cov}(\varepsilon_0, \varepsilon_0) \\ &= \sigma^2 a^s \sum_{u \geq s} a^{2(u-s)} = \sigma^2 \frac{a^s}{1-a^2}, \qquad s = 0, 1, 2, \ldots \end{aligned}$$

and $\gamma(-s) = \gamma(s)$. In particular we obtain $\gamma(0) = \sigma^2/(1-a^2)$ and thus, the autocorrelation function of (Y_t) is given by

$$\rho(s) = a^{|s|}, \qquad s \in \mathbb{Z}.$$

The autocorrelation function of an $AR(1)$-process $Y_t = aY_{t-1} + \varepsilon_t$ with $|a| < 1$ therefore decreases at an exponential rate. Its sign is alternating if $a \in (-1, 0)$.

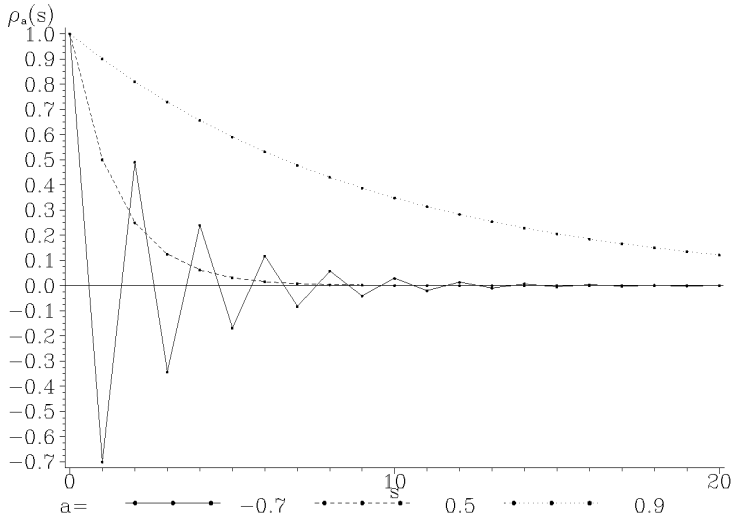


Figure 2.2.1. Autocorrelation functions of $AR(1)$-processes $Y_t = aY_{t-1} + \varepsilon_t$ with different values of a.

```
***  Program 2_2_1   ***;
TITLE1 'Autocorrelation functions of AR(1)-processes;

DATA data1;
  DO a=-0.7, 0.5, 0.9;
    DO s=0 TO 20;
      rho=a**s;
      OUTPUT;
    END;
  END;


SYMBOL1 C=GREEN V=DOT I=JOIN H=0.3 L=1;
SYMBOL2 C=GREEN V=DOT I=JOIN H=0.3 L=2;
SYMBOL3 C=GREEN V=DOT I=JOIN H=0.3 L=33;
AXIS1 LABEL=('s');
AXIS2 LABEL=(F=CGREEK 'r' F=COMPLEX H=1 'a' H=2 '(s)');
LEGEND1 LABEL=('a=') SHAPE=SYMBOL(10,0.6);
```

```
PROC GPLOT DATA=data1;
  PLOT rho*s=a / HAXIS=AXIS1 VAXIS=AXIS2
        LEGEND=LEGEND1 VREF=0;
RUN; QUIT;
```

The data step evaluates `rho` for three different values of `a` and the range of `s` from 0 to 20 using two loops. The plot is generated by the procedure `GPLOT`. The `LABEL` option in the `AXIS2` statement uses, in addition to the greek font `CGREEK`, the font `COMPLEX` assuming this to be the default text font (`GOPTION FTEXT=COMPLEX`). The `SHAPE` option `SHAPE=SYMBOL(10,0.6)` in the `LEGEND` statement defines width and height of the symbols presented in the legend.

The following figure illustrates the significance of the stationarity condition $|a| < 1$ of an $AR(1)$-process. Realizations $Y_t = aY_{t-1} + \varepsilon_t,\ t = 1, \ldots, 10$, are displayed for $a = 0.5$ and $a = 1.5$, where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{10}$ are independent standard normal in each case and Y_0 is assumed to be zero. While for $a = 0.5$ the sample path follows the constant zero closely, which is the expectation of each Y_t, the observations Y_t decrease rapidly in case of $a = 1.5$.

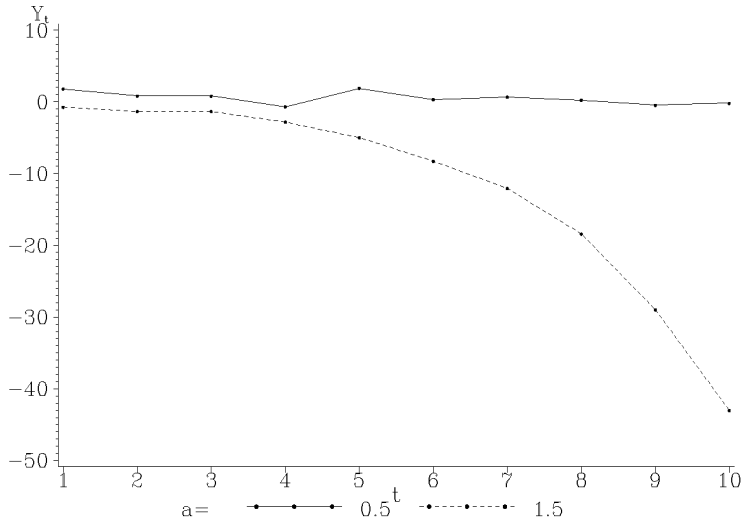


Figure 2.2.2. Realizations $Y_t = 0.5Y_{t-1} + \varepsilon_t$ and $Y_t = 1.5Y_{t-1} + \varepsilon_t,\ t = 1, \ldots, 10$, with ε_t independent standard normal and $Y_0 = 0$.

```
***    Program 2_2_2   ***;
TITLE1 'Realizations of AR(1)-processes;


DATA data1;
  DO a=0.5, 1.5;
    t=0; y=0; OUTPUT;
    DO t=1 TO 10;
      y=a*y+RANNOR(1);
      OUTPUT;
    END;
  END;


SYMBOL1 C=GREEN V=DOT I=JOIN H=0.4 L=1;
SYMBOL2 C=GREEN V=DOT I=JOIN H=0.4 L=2;
AXIS1 LABEL=('t') MINOR=NONE;
AXIS2 LABEL=('Y' H=1 't');
LEGEND1 LABEL=('a=') SHAPE=SYMBOL(10,0.6);
PROC GPLOT DATA=data1(WHERE=(t>0));
  PLOT y*t=a / HAXIS=AXIS1 VAXIS=AXIS2 LEGEND=LEGEND1;
RUN; QUIT;
```

The data are generated within two loops, the first one over the two values for `a`. The variable `y` is initialized with the value 0 corresponding to `t=0`. The realizations for `t=1, ..., 10` are created within the second loop over `t` and with the help of the function `RANNOR` which returns pseudo random numbers distributed as standard normal. The argument `1` is the initial seed to produce a stream of random numbers. A positive value of this seed always produces the same series of random numbers, a negative value generates a different series each time the program is submitted. A value of `y` is calculated as the sum of `a` times the actual value of `y` and the random number and stored in a new observation. The resulting data set has 22 observations and 3 variables (`a`, `t` and `y`).

In the plot created by `PROC GPLOT` the initial observations are dropped using the `WHERE` data set option. Only observations fulfilling the condition `t>0` are read into the data set used here. To suppress minor tick marks between the integers `0,1, ...,10` the option `MINOR` in the `AXIS1` statement is set to `NONE`.

The Yule–Walker Equations

The Yule–Walker equations entail the recursive computation of the autocorrelation function ρ of an $AR(p)$-process satisfying the stationarity condition (2.3).

Lemma 2.2.5. *Let $Y_t = \sum_{u=1}^{p} a_u Y_{t-u} + \varepsilon_t$ be an AR(p)-process, which satisfies the stationarity condition (2.3). Its autocorrelation function ρ then satisfies for $s = 1, 2, \ldots$ the recursion*

$$\rho(s) = \sum_{u=1}^{p} a_u \rho(s-u), \tag{2.4}$$

known as Yule–Walker equations.

Proof. With $\mu := \mathrm{E}(Y_0)$ we have for $t \in \mathbb{Z}$

$$Y_t - \mu = \sum_{u=1}^{p} a_u (Y_{t-u} - \mu) + \varepsilon_t - \mu\Big(1 - \sum_{u=1}^{p} a_u\Big), \tag{$\star$}$$

and taking expectations of $(\star)$ gives $\mu(1 - \sum_{u=1}^{p} a_u) = \mathrm{E}(\varepsilon_0) =: \nu$ due to the stationarity of (Y_t). By multiplying equation $(\star)$ with $Y_{t-s} - \mu$ for $s > 0$ and taking expectations again we obtain

$$\begin{aligned}
\gamma(s) &= \mathrm{E}((Y_t - \mu)(Y_{t-s} - \mu)) \\
&= \sum_{u=1}^{p} a_u \,\mathrm{E}((Y_{t-u} - \mu)(Y_{t-s} - \mu)) + \mathrm{E}((\varepsilon_t - \nu)(Y_{t-s} - \mu)) \\
&= \sum_{u=1}^{p} a_u \gamma(s-u).
\end{aligned}$$

for the autocovariance function γ of (Y_t). The final equation follows from the fact that Y_{t-s} and ε_t are uncorrelated for $s > 0$. This is a consequence of Theorem 2.2.3, by which almost surely $Y_{t-s} = \sum_{u \geq 0} b_u \varepsilon_{t-s-u}$ with an absolutely summable causal filter (b_u) and thus, $\mathrm{Cov}(Y_{t-s}, \varepsilon_t) = \sum_{u \geq 0} b_u \,\mathrm{Cov}(\varepsilon_{t-s-u}, \varepsilon_t) = 0$, see Theorem 2.1.5. Dividing the above equation by $\gamma(0)$ now yields the assertion. □

Since $\rho(-s) = \rho(s)$, equations (2.4) can be represented as

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(p) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & & \rho(p-2) \\ \rho(2) & \rho(1) & 1 & & \rho(p-3) \\ \vdots & & & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{pmatrix} \tag{2.5}$$

This matrix equation offers an estimator of the coefficients $a_1, \ldots, a_p$ by replacing the autocorrelations $\rho(j)$ by their empirical counterparts $r(j)$, $1 \leq j \leq p$.

Equation (2.5) then formally becomes $\boldsymbol{r} = \boldsymbol{R}\boldsymbol{a}$, where $\boldsymbol{r} = (r(1), \ldots, r(p))^T$, $\boldsymbol{a} = (a_1, \ldots, a_p)^T$ and

$$\boldsymbol{R} := \begin{pmatrix} 1 & r(1) & r(2) & \cdots & r(p-1) \\ r(1) & 1 & r(1) & \cdots & r(p-2) \\ \vdots & & & & \vdots \\ r(p-1) & r(p-2) & r(p-3) & \cdots & 1 \end{pmatrix}.$$

If the $p \times p$-matrix $\boldsymbol{R}$ is invertible, we can rewrite the formal equation $\boldsymbol{r} = \boldsymbol{R}\boldsymbol{a}$ as $\boldsymbol{R}^{-1}\boldsymbol{r} = \boldsymbol{a}$, which motivates the estimator

$$\hat{\boldsymbol{a}} := \boldsymbol{R}^{-1}\boldsymbol{r} \tag{2.6}$$

of the vector $\boldsymbol{a} = (a_1, \ldots, a_p)^T$ of the coefficients.

The Partial Autocorrelation Coefficients

We have seen that the autocorrelation function $\rho(k)$ of an $MA(q)$-process vanishes for $k > q$, see Lemma 2.2.1. This is not true for an $AR(p)$-process, whereas the partial autocorrelation coefficients will share this property. Note that the correlation matrix

$$\begin{aligned} \boldsymbol{P}_k &:= \Big(\operatorname{Corr}(Y_i, Y_j) \Big)_{1 \le i,j \le k} \\ &= \begin{pmatrix} 1 & \rho(1) & \rho(2) & \ldots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & & \rho(p-2) \\ \rho(2) & \rho(1) & 1 & & \rho(p-3) \\ \vdots & & & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & \cdots & 1 \end{pmatrix} \end{aligned} \tag{2.7}$$

is positive semidefinite for any $k \ge 1$. If we suppose that $\boldsymbol{P}_k$ is positive definite, then it is invertible, and the equation

$$\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(k) \end{pmatrix} = \boldsymbol{P}_k \begin{pmatrix} a_{k1} \\ \vdots \\ a_{kk} \end{pmatrix} \tag{2.8}$$

has the unique solution

$$\boldsymbol{a}_k := \begin{pmatrix} a_{k1} \\ \vdots \\ a_{kk} \end{pmatrix} = \boldsymbol{P}_k^{-1} \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(k) \end{pmatrix}.$$

The number a_{kk} is called *partial autocorrelation coefficient at lag k*, denoted by $\alpha(k)$, $k \ge 1$. Observe that for $k \ge p$ the vector $(a_1, \ldots, a_p, 0, \ldots, 0) \in \mathbb{R}^k$, with

$k-p$ zeros added to the vector of coefficients $(a_1, \ldots, a_p)$, is by the Yule–Walker equations (2.4) a solution of the equation (2.8). Thus we have $\alpha(p) = a_p$, $\alpha(k) = 0$ for $k > p$. Note that the coefficient $\alpha(k)$ also occurs as the coefficient of Y_{n-k} in the best linear one-step forecast $\sum_{u=0}^{k} c_u Y_{n-u}$ of Y_{n+1}, see equation (2.17) in Section 2.3.

If the empirical counterpart $\boldsymbol{R}_k$ of $\boldsymbol{P}_k$ is invertible as well, then

$$\hat{\boldsymbol{a}}_k := \boldsymbol{R}_k^{-1} \boldsymbol{r}_k,$$

with $\boldsymbol{r}_k := (r(1), \ldots, r(k))^T$ being an obvious estimate of $\boldsymbol{a}_k$. The k-th component

$$\hat{\alpha}(k) := \hat{a}_{kk} \tag{2.9}$$

of $\hat{\boldsymbol{a}}_k = (\hat{a}_{k1}, \ldots, \hat{a}_{kk})$ is the *empirical partial autocorrelation coefficient at lag* k. It can be utilized to estimate the order p of an $AR(p)$-process, since $\hat{\alpha}(p) \approx \alpha(p) = a_p$ is different from zero, whereas $\hat{\alpha}(k) \approx \alpha(k) = 0$ for $k > p$ should be close to zero.

Example 2.2.6. The Yule–Walker equations (2.4) for an $AR(2)$-process $Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t$ are for $s = 1, 2$

$$\rho(1) = a_1 + a_2 \rho(1), \qquad \rho(2) = a_1 \rho(1) + a_2$$

with the solutions

$$\rho(1) = \frac{a_1}{1 - a_2}, \qquad \rho(2) = \frac{a_1^2}{1 - a_2} + a_2.$$

and thus, the partial autocorrelation coefficients are

$$\alpha(1) = \rho(1),\ \alpha(2) = a_2,\ \alpha(j) = 0, \quad j \geq 3.$$

The recursion (2.4) entails the computation of $\rho(s)$ for an arbitrary s from the two values $\rho(1)$ and $\rho(2)$.

The following figure displays realizations of the $AR(2)$-process $Y_t = 0.6 Y_{t-1} - 0.3 Y_{t-2} + \varepsilon_t$ for $1 \leq t \leq 200$, conditional on $Y_{-1} = Y_0 = 0$. The random shocks ε_t are iid standard normal. The corresponding partial autocorrelation function is shown in Figure 2.2.4.

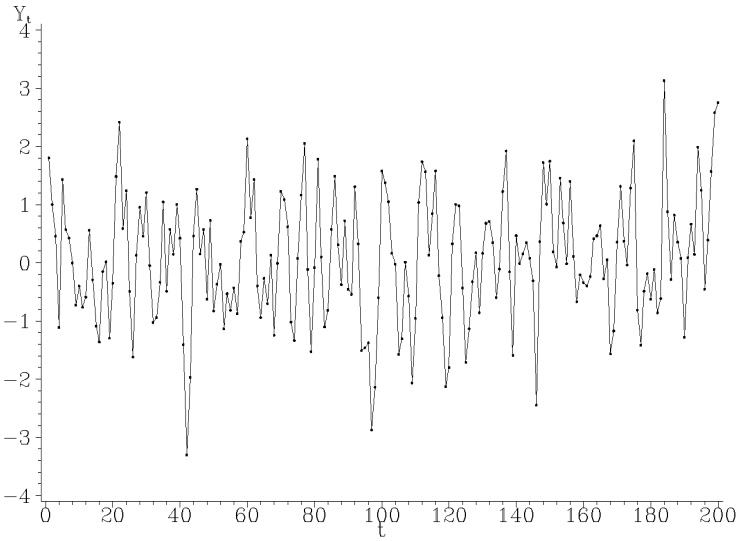


Figure 2.2.3. Realization of the $AR(2)$-process $Y_t = 0.6Y_{t-1} - 0.3Y_{t-2} + \varepsilon_t$, conditional on $Y_{-1} = Y_0 = 0$. The ε_t, $1 \le t \le 200$, are iid standard normal.

```
***     Program 2_2_3    ***;
TITLE1 'Realisation of an AR(2)-process;

DATA data1;
  t=-1; y=0; OUTPUT;
  t=0; y1=y; y=0; OUTPUT;
  DO t=1 TO 200;
     y2=y1;
     y1=y;
     y=0.6*y1-0.3*y2+RANNOR(1);
     OUTPUT;
  END;

SYMBOL1 C=GREEN V=DOT I=JOIN H=0.3;
AXIS1 LABEL=('t');
AXIS2 LABEL=('Y' H=1 't');
PROC GPLOT DATA=data1(WHERE=(t>0));
  PLOT y*t / HAXIS=AXIS1 VAXIS=AXIS2;
```

```
RUN; QUIT;
```

The two initial values of `y` are defined and stored in an observation by the `OUTPUT` statement. The second observation contains an additional value `y1` for y_{t-1}. Within the loop the values `y2` (for y_{t-2}), `y1` and `y` are updated one after the other. The data set used by `PROC GPLOT` again just contains the observations with $t > 0$.

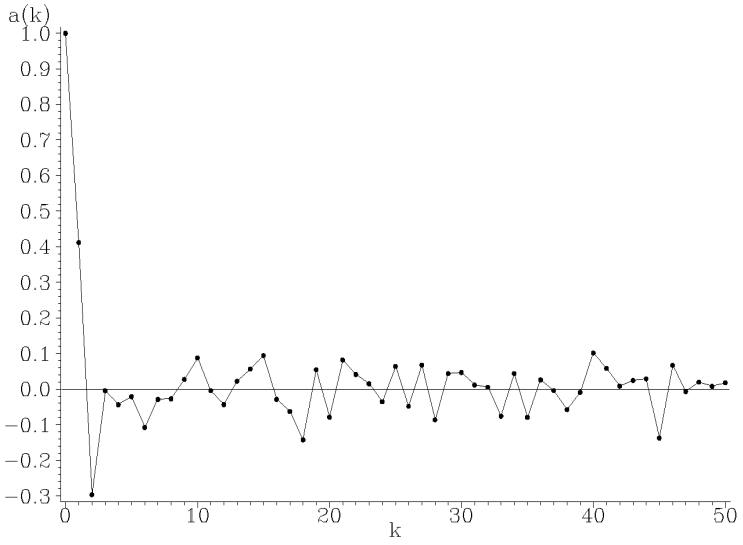


Figure 2.2.4. Empirical partial autocorrelation function of the $AR(2)$-data in Figure 2.2.3.

```
***     Program 2_2_4   ***;
TITLE1 'Empirical partial autocorrelation function';
TITLE2 'of simulated AR(2)-process data';
* Note that this program requires
     data1 generated by program 2_2_3;

PROC ARIMA DATA=data1(WHERE=(t>0));
  IDENTIFY VAR=y NLAG=50 OUTCOV=corr NOPRINT;

SYMBOL1 C=GREEN V=DOT I=JOIN H=0.7;
AXIS1 LABEL=('k');
AXIS2 LABEL=('a(k)');
PROC GPLOT DATA=corr;
  PLOT PARTCORR*LAG / HAXIS=AXIS1 VAXIS=AXIS2 VREF=0;
```

```
RUN; QUIT;
```

This program requires to be submitted to SAS for execution within a joint session with Program 2_2_3, because it uses the temporary data step `data1` generated there. Otherwise you have to add the block of statements to this program concerning the data step.

Like in Program 1_3_1 the procedure `ARIMA` with the `IDENTIFY` statement is used to create a data set. Here we are interested in the variable `PARTCORR` containing the values of the empirical partial autocorrelation function from the simulated AR(2)-process data. This variable is plotted against the lag stored in variable `LAG`.

ARMA-Processes

Moving averages $MA(q)$ and autoregressive $AR(p)$-processes are special cases of so called autoregressive moving averages. Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise, $p, q \geq 0$ integers and $a_0, \ldots, a_p, b_0, \ldots, b_q \in \mathbb{R}$. A real valued stochastic process $(Y_t)_{t\in\mathbb{Z}}$ is said to be an *autoregressive moving average process of order* p, q, denoted by $ARMA(p, q)$, if it satisfies the equation

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q}. \tag{2.10}$$

An $ARMA(p, 0)$-process with $p \geq 1$ is obviously an $AR(p)$-process, whereas an $ARMA(0, q)$-process with $q \geq 1$ is a moving average $MA(q)$. The polynomials

$$A(z) := 1 - a_1 z - \ldots - a_p z^p \tag{2.11}$$

and

$$B(z) := 1 + b_1 z + \ldots + b_q z^q, \tag{2.12}$$

are the characteristic polynomials of the autoregressive part and of the moving average part of an $ARMA(p, q)$-process (Y_t), which we can represent in the form

$$Y_t - a_1 Y_{t-1} - \ldots - a_p Y_{t-p} = \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q}.$$

Denote by Z_t the right-hand side of the above equation i.e., $Z_t := \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q}$. This is a $MA(q)$-process and, therefore, stationary by Theorem 2.1.6. If all p roots of the equation $A(z) = 1 - a_1 z - \ldots - a_p z^p = 0$ are outside of the unit circle, then we deduce from Theorem 2.1.11 that the filter $c_0 = 1$, $c_u = -a_u$, $u = 1, \ldots, p$, $c_u = 0$ elsewhere, has an absolutely summable causal inverse filter $(d_u)_{u\geq 0}$. Consequently we obtain from the equation $Z_t = Y_t - a_1 Y_{t-1} - \ldots - a_p Y_{t-p}$

and (2.1) that with $b_0 = 1$, $b_w = 0$ if $w > q$

$$\begin{aligned} Y_t = \sum_{u\geq 0} d_u Z_{t-u} &= \sum_{u\geq 0} d_u(\varepsilon_{t-u} + b_1\varepsilon_{t-1-u} + \ldots + b_q\varepsilon_{t-q-u}) \\ &= \sum_{u\geq 0}\sum_{w\geq 0} d_u b_w \varepsilon_{t-w-u} = \sum_{v\geq 0}\Big(\sum_{u+w=v} d_u b_w\Big)\varepsilon_{t-v} \\ &= \sum_{v\geq 0}\Big(\sum_{w=0}^{\min(v,q)} b_w d_{v-w}\Big)\varepsilon_{t-v} =: \sum_{v\geq 0}\alpha_v\varepsilon_{t-v} \end{aligned}$$

is the almost surely uniquely determined stationary solution of the $ARMA(p,q)$-equation (2.10) for a given white noise (ε_t) .

The condition that all p roots of the characteristic equation $A(z) = 1 - a_1 z - a_2 z^2 - \ldots - a_p z^p = 0$ of the $ARMA(p,q)$-process (Y_t) are outside of the unit circle will again be referred to in the following as the *stationarity condition* (2.3). In this case, the process $Y_t = \sum_{v\geq 0}\alpha_v\varepsilon_{t-v}$, $t \in \mathbb{Z}$, is the almost surely uniquely determined stationary solution of the $ARMA(p,q)$-equation (2.10), which is called *causal*.

The $MA(q)$-process $Z_t = \varepsilon_t + b_1\varepsilon_{t-1} + \ldots + b_q\varepsilon_{t-q}$ is by definition invertible if all q roots of the polynomial $B(z) = 1 + b_1 z + \ldots + b_q z^q$ are outside of the unit circle. Theorem 2.1.11 and equation (2.1) imply in this case the existence of an absolutely summable causal filter $(g_u)_{u\geq 0}$ such that with $a_0 = -1$

$$\begin{aligned} \varepsilon_t &= \sum_{u\geq 0} g_u Z_{t-u} = \sum_{u\geq 0} g_u(Y_{t-u} - a_1 Y_{t-1-u} - \ldots - a_p Y_{t-p-u}) \\ &= -\sum_{v\geq 0}\Big(\sum_{w=0}^{\min(v,p)} a_w g_{v-w}\Big)Y_{t-v}. \end{aligned}$$

In this case the $ARMA(p,q)$-process (Y_t) is said to be *invertible*.

The Autocovariance Function of an ARMA-Process

In order to deduce the autocovariance function of an $ARMA(p,q)$-process (Y_t), which satisfies the stationarity condition (2.3), we compute at first the absolutely summable coefficients $\alpha_v = \sum_{w=0}^{\min(q,v)} b_w d_{v-w}$, $v \geq 0$, in the above representation $Y_t = \sum_{v\geq 0}\alpha_v\varepsilon_{t-v}$. The characteristic polynomial $D(z)$ of the absolutely summable causal filter $(d_u)_{u\geq 0}$ coincides by Lemma 2.1.9 for $0 < |z| < 1$ with $1/A(z)$, where $A(z)$ is given in (2.11). Thus we obtain with $B(z)$ as given in (2.12) for $0 < |z| < 1$

$$
\begin{aligned}
A(z)(B(z)D(z)) &= B(z) \\
&\Leftrightarrow \Big(-\sum_{u=0}^{p} a_u z^u\Big)\Big(\sum_{v\geq 0} \alpha_v z^v\Big) = \sum_{w=0}^{q} b_w z^w \\
&\Leftrightarrow \sum_{w\geq 0}\Big(-\sum_{u+v=w} a_u\alpha_v\Big) z^w \quad = \sum_{w\geq 0} b_w z^w \\
&\Leftrightarrow \sum_{w\geq 0}\Big(-\sum_{u=0}^{w} a_u\alpha_{w-u}\Big) z^w \quad = \sum_{w\geq 0} b_w z^w \\
&\Leftrightarrow \begin{cases} \alpha_0 = 1 & \\ \alpha_w - \sum_{u=1}^{w} a_u\alpha_{w-u} = b_w & \text{for } 1\leq w\leq p \\ \alpha_w - \sum_{u=1}^{p} a_u\alpha_{w-u} = 0 & \text{for } w > p. \end{cases} \qquad (2.13)
\end{aligned}
$$

Example 2.2.7. For the $ARMA(1,1)$-process $Y_t - aY_{t-1} = \varepsilon_t + b\varepsilon_{t-1}$ with $|a| < 1$ we obtain from (2.13)

$$\alpha_0 = 1,\ \alpha_1 - a = b,\ \alpha_w - a\alpha_{w-1} = 0. \qquad w \geq 2,$$

This implies $\alpha_0 = 1$, $\alpha_w = a^{w-1}(b+a)$, $w \geq 1$, and, hence,

$$Y_t = \varepsilon_t + (b+a)\sum_{w\geq 1} a^{w-1}\varepsilon_{t-w}.$$

Theorem 2.2.8. *Suppose that $Y_t = \sum_{u=1}^{p} a_u Y_{t-u} + \sum_{v=0}^{q} b_v\varepsilon_{t-v}$, $b_0 := 1$, is an $ARMA(p,q)$-process, which satisfies the stationarity condition* (2.3)*. Its autocovariance function γ then satisfies the recursion*

$$
\begin{aligned}
\gamma(s) - \sum_{u=1}^{p} a_u\gamma(s-u) &= \sigma^2 \sum_{v=s}^{q} b_v\alpha_{v-s}, \qquad 0\leq s\leq q, \\
\gamma(s) - \sum_{u=1}^{p} a_u\gamma(s-u) &= 0, \qquad s \geq q+1, \qquad (2.14)
\end{aligned}
$$

where α_v, $v \geq 0$, are the coefficients in the representation $Y_t = \sum_{v\geq 0}\alpha_v\varepsilon_{t-v}$, which we computed in (2.13) *and σ^2 is the variance of ε_0.*

By the preceding result the autocorrelation function ρ of the $ARMA(p,q)$-process (Y_t) satisfies

$$\rho(s) = \sum_{u=1}^{p} a_u\rho(s-u), \qquad s \geq q+1,$$

which coincides with the autocorrelation function of the stationary $AR(p)$-process $X_t = \sum_{u=1}^{p} a_u X_{t-u} + \varepsilon_t$, c.f. Lemma 2.2.5.

Proof of Theorem 2.2.8. Put $\mu := \mathrm{E}(Y_0)$ and $\nu := \mathrm{E}(\varepsilon_0)$. Then we have

$$Y_t - \mu = \sum_{u=1}^{p} a_u (Y_{t-u} - \mu) + \sum_{v=0}^{q} b_v (\varepsilon_{t-v} - \nu), \qquad t \in \mathbb{Z}.$$

Recall that $Y_t = \sum_{u=1}^{p} a_u Y_{t-u} + \sum_{v=0}^{q} b_v \varepsilon_{t-v}$, $t \in \mathbb{Z}$. Taking expectations on both sides we obtain $\mu = \sum_{u=1}^{p} a_u \mu + \sum_{v=0}^{q} b_v \nu$, which yields now the equation displayed above. Multiplying both sides with $Y_{t-s} - \mu$, $s \geq 0$, and taking expectations, we obtain

$$\mathrm{Cov}(Y_{t-s}, Y_t) = \sum_{u=1}^{p} a_u \,\mathrm{Cov}(Y_{t-s}, Y_{t-u}) + \sum_{v=0}^{q} b_v \,\mathrm{Cov}(Y_{t-s}, \varepsilon_{t-v}),$$

which implies

$$\gamma(s) - \sum_{u=1}^{p} a_u \gamma(s-u) = \sum_{v=0}^{q} b_v \,\mathrm{Cov}(Y_{t-s}, \varepsilon_{t-v}).$$

From the representation $Y_{t-s} = \sum_{w \geq 0} \alpha_w \varepsilon_{t-s-w}$ and Theorem 2.1.5 we obtain

$$\mathrm{Cov}(Y_{t-s}, \varepsilon_{t-v}) = \sum_{w \geq 0} \alpha_w \,\mathrm{Cov}(\varepsilon_{t-s-w}, \varepsilon_{t-v}) = \begin{cases} 0 & \text{if } v < s \\ \sigma^2 \alpha_{v-s} & \text{if } v \geq s. \end{cases}$$

This implies

$$\gamma(s) - \sum_{u=1}^{p} a_u \gamma(s-u) = \sum_{v=s}^{q} b_v \,\mathrm{Cov}(Y_{t-s}, \varepsilon_{t-v}) = \begin{cases} \sigma^2 \sum_{v=s}^{q} b_v \alpha_{v-s} & \text{if } s \leq q \\ 0 & \text{if } s > q, \end{cases}$$

which is the assertion. □

Example 2.2.9. For the $ARMA(1,1)$-process $Y_t - aY_{t-1} = \varepsilon_t + b\varepsilon_{t-1}$ with $|a| < 1$ we obtain from Example 2.2.7 and Theorem 2.2.8 with $\sigma^2 = \mathrm{Var}(\varepsilon_0)$

$$\gamma(0) - a\gamma(1) = \sigma^2(1 + b(b+a)),\ \gamma(1) - a\gamma(0) = \sigma^2 b,$$

and thus

$$\gamma(0) = \sigma^2 \frac{1 + 2ab + b^2}{1 - a^2}, \quad \gamma(1) = \sigma^2 \frac{(1+ab)(a+b)}{1-a^2}.$$

For $s \geq 2$ we obtain from (2.14)

$$\gamma(s) = a\gamma(s-1) = \cdots = a^{s-1}\gamma(1).$$

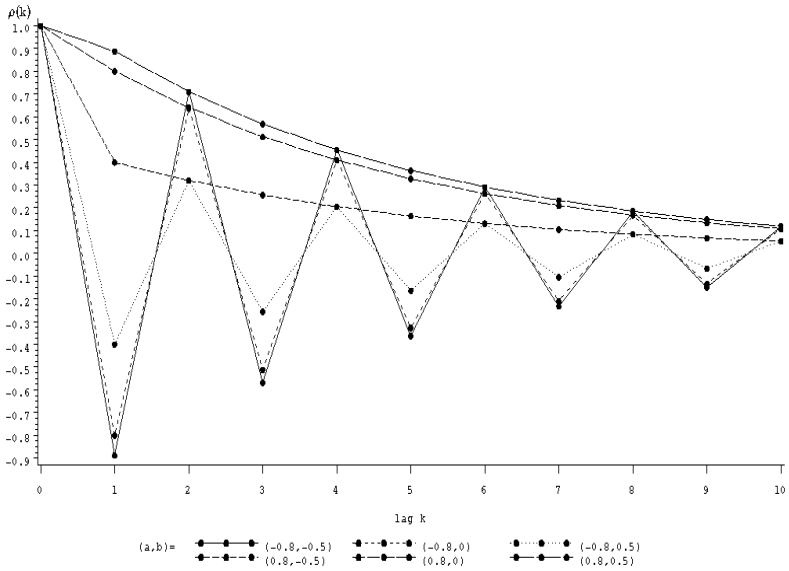


Figure 2.2.5. Autocorrelation functions of $ARMA(1,1)$-processes with $a = 0.8/-0.8$, $b = 0.5/0/-0.5$ and $\sigma^2 = 1$.

```
***     Program 2_2_5   ***;
TITLE1 'Autocorrelation functions
            of ARMA(1,1)-processes';

DATA data1;
 DO a=-0.8, 0.8;
   DO b=-0.5, 0, 0.5;
     s=0; rho=1;
     q=COMPRESS('(' || a || ',' || b || ')'); OUTPUT;
     s=1; rho=(1+a*b)*(a+b)/(1+2*a*b+b*b);
     q=COMPRESS('(' || a || ',' || b || ')'); OUTPUT;
     DO s=2 TO 10;
        rho=a*rho;
        q=COMPRESS('(' || a || ',' || b || ')');
        OUTPUT;
     END;
   END;
 END;
```

```
SYMBOL1 C=RED V=DOT I=JOIN H=0.7 L=1;
SYMBOL2 C=YELLOW V=DOT I=JOIN H=0.7 L=2;
SYMBOL3 C=BLUE V=DOT I=JOIN H=0.7 L=33;
SYMBOL4 C=RED V=DOT I=JOIN H=0.7 L=3;
SYMBOL5 C=YELLOW V=DOT I=JOIN H=0.7 L=4;
SYMBOL6 C=BLUE V=DOT I=JOIN H=0.7 L=5;
AXIS1 LABEL=(F=CGREEK 'r' F=COMPLEX '(k)');
AXIS2 LABEL=('lag k') MINOR=NONE;
LEGEND1 LABEL=('(a,b)=') SHAPE=SYMBOL(10,0.8);
PROC GPLOT DATA=data1;
  PLOT rho*s=q / VAXIS=AXIS1 HAXIS=AXIS2 LEGEND=LEGEND1;
RUN; QUIT;
```

In the data step the values of the autocorrelation function belonging to an $ARMA(1,1)$ process are calculated for two different values of `a`, the coefficient of the $AR(1)$-part, and three different values of `b`, the coefficient of the $MA(1)$-part. Pure $AR(1)$–processes result for the value `b=0`. For the arguments (lags) `s=0` and `s=1` the computation is done directly, for the rest up to `s=10` a loop is used for a recursive computation. For the `COMPRESS` statement see Program 1_1_3.

The second part of the program uses `PROC GPLOT` to plot the autocorrelation function, using known statements and options to customize the output.

ARIMA-Processes

Suppose that the time series (Y_t) has a polynomial trend of degree d. Then we can eliminate this trend by considering the process $(\Delta^d Y_t)$, obtained by d times differencing as described in Section 1.2. If the filtered process $(\Delta^d Y_d)$ is an $ARMA(p,q)$-process satisfying the stationarity condition (2.3), the original process (Y_t) is said to be an *autoregressive integrated moving average of order* p,d,q, denoted by $ARIMA(p,d,q)$. In this case constants $a_1,\ldots,a_p, b_0 = 1, b_1,\ldots,b_q \in \mathbb{R}$ exist such that

$$\Delta^d Y_t = \sum_{u=1}^{p} a_u \Delta^d Y_{t-u} + \sum_{w=0}^{q} b_w \varepsilon_{t-w}, \qquad t \in \mathbb{Z},$$

where (ε_t) is a white noise.

Example 2.2.10. An $ARIMA(1,1,1)$-process (Y_t) satisfies

$$\Delta Y_t = a\Delta Y_{t-1} + \varepsilon_t + b\varepsilon_{t-1}, \qquad t \in \mathbb{Z},$$

where $|a| < 1$, $b \neq 0$ and (ε_t) is a white noise, i.e.,

$$Y_t - Y_{t-1} = a(Y_{t-1} - Y_{t-2}) + \varepsilon_t + b\varepsilon_{t-1}, \qquad t \in \mathbb{Z}.$$

This implies $Y_t = (a+1)Y_{t-1} - aY_{t-2} + \varepsilon_t + b\varepsilon_{t-1}$.

A *random walk* $X_t = X_{t-1} + \varepsilon_t$ is obviously an $ARIMA(0,1,0)$-process.

Consider $Y_t = S_t + R_t$, $t \in \mathbb{Z}$, where the random component (R_t) is a stationary process and the seasonal component (S_t) is periodic of length s, i.e., $S_t = S_{t+s} = S_{t+2s} = \ldots$ for $t \in \mathbb{Z}$. Then the process (Y_t) is in general not stationary, but $Y_t^* := Y_t - Y_{t-s}$ is. If this seasonally adjusted process (Y_t^*) is an $ARMA(p,q)$-process satisfying the stationarity condition (2.3), then the original process (Y_t) is called a *seasonal $ARMA(p,q)$-process with period length s*, denoted by $SARMA_s(p,q)$. One frequently encounters a time series with a trend as well as a periodic seasonal component. A stochastic process (Y_t) with the property that $(\Delta^d(Y_t - Y_{t-s}))$ is an $ARMA(p,q)$-process is, therefore, called a $SARIMA_s(p,d,q)$-process. This is a quite common assumption in practice.

Cointegration

In the sequel we will frequently use the notation that a time series (Y_t) is $I(d)$, $d = 0, 1$, if the sequence of differences $(\Delta^d Y_t)$ of order d is a stationary process. By the difference $\Delta^0 Y_t$ of order zero we denote the undifferenced process Y_t, $t \in \mathbb{Z}$.

Suppose that the two time series (Y_t) and (Z_t) satisfy

$$Y_t = aW_t + \varepsilon_t, \quad Z_t = W_t + \delta_t, \qquad t \in \mathbb{Z},$$

for some real number $a \neq 0$, where (W_t) is $I(1)$, and (ε_t), (δ_t) are uncorrelated white noise processes, i.e., $\mathrm{Cov}(\varepsilon_t, \delta_s) = 0$, $t, s \in \mathbb{Z}$.

Then (Y_t) and (Z_t) are both $I(1)$, but

$$X_t := Y_t - aZ_t = \varepsilon_t - a\delta_t, \qquad t \in \mathbb{Z},$$

is $I(0)$.

The fact that the combination of two nonstationary series yields a stationary process arises from a common component (W_t), which is $I(1)$. More generally, two $I(1)$ series (Y_t), (Z_t) are said to be *cointegrated*, if there exist constants μ, α_1, α_2 with α_1, α_2 different from 0, such that the process

$$X_t = \mu + \alpha_1 Y_t + \alpha_2 Z_t, \qquad t \in \mathbb{Z}$$

is $I(0)$. Without loss of generality, we can choose $\alpha_1 = 1$ in this case.

Such cointegrated time series are often encountered in macroeconomics (Granger (1981), Engle and Granger (1987)). Consider, for example, prices for the same commodity in different parts of a country. Principles of supply and demand, along with the possibility of arbitrage, means that, while the process may fluctuate more-or-less randomly, the distance between them will, in equilibrium, be relatively constant (typically about zero).

The link between cointegration and error correction can vividly be described by the humorous tale of the drunkard and his dog, c.f. Murray (1994). In the same way a drunkard seems to follow a random walk an unleashed dog wanders aimlessly. We can, therefore, model their ways by random walks

$$\begin{aligned} Y_t &= Y_{t-1} + \varepsilon_t \quad \text{and} \\ Z_t &= Z_{t-1} + \delta_t, \end{aligned}$$

where the individual single steps (ε_t), (δ_t) of man and dog are uncorrelated white noise processes. Random walks are not stationary, since their variances increase, and so both processes (Y_t) and (Z_t) are not stationary.

And if the dog belongs to the drunkard? We assume the dog to be unleashed and thus, the distance $Y_t - Z_t$ between the drunk and his dog is a random variable. It seems reasonable to assume that these distances form a stationary process, i.e., that (Y_t) and (Z_t) are cointegrated with constants $\alpha_1 = 1$ and $\alpha_2 = -1$.

We model the cointegrated walks above more tritely by assuming the existence of constants $c, d \in \mathbb{R}$ such that

$$\begin{aligned} Y_t - Y_{t-1} &= \varepsilon_t + c(Y_{t-1} - Z_{t-1}) \quad \text{and} \\ Z_t - Z_{t-1} &= \delta_t + d(Y_{t-1} - Z_{t-1}). \end{aligned}$$

The additional terms on the right-hand side of these equations are the *error correction terms*.

Cointegration requires that both variables in question be $I(1)$, but that a linear combination of them be $I(0)$. This means that the first step is to figure out if the series themselves are $I(1)$, typically by using *unit root tests*. If one or both are not $I(1)$, cointegration is not an option.

Whether two processes (Y_t) and (Z_t) are cointegrated can be tested by means of a linear regression approach. This is based on the *cointegration regression*

$$Y_t = \beta_0 + \beta_1 Z_t + \varepsilon_t,$$

where (ε_t) is a stationary process and β_0, $\beta_1 \in \mathbb{R}$ are the cointegration constants.

One can use the ordinary least squares estimates $\hat{\beta}_0$, $\hat{\beta}_1$ of the target parameters β_0, β_1, which satisfy

$$\sum_{t=1}^{n} \left(Y_t - \hat{\beta}_0 - \hat{\beta}_1 Z_t \right)^2 = \min_{\beta_0, \beta_1 \in \mathbb{R}} \sum_{t=1}^{n} \left(Y_t - \beta_0 - \beta_1 Z_t \right)^2,$$

and one checks, whether the estimated residuals

$$\hat{\varepsilon}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 Z_t$$

are generated by a stationary process.

A general strategy for examining cointegrated series can now be summarized as follows:

(i) Determine that the two series are $I(1)$.

(ii) Compute $\hat{\varepsilon}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 Z_t$ using ordinary least squares.

(iii) Examine $\hat{\varepsilon}_t$ for stationarity, using

- the Durbin–Watson test
- standard unit root tests such as Dickey–Fuller or augmented Dickey–Fuller.

Example 2.2.11. (Hog Data) Quenouille's (1957) Hog Data list the annual hog supply and hog prices in the U.S. between 1867 and 1948. Do they provide a typical example of cointegrated series? A discussion can be found in Box and Tiao (1977).

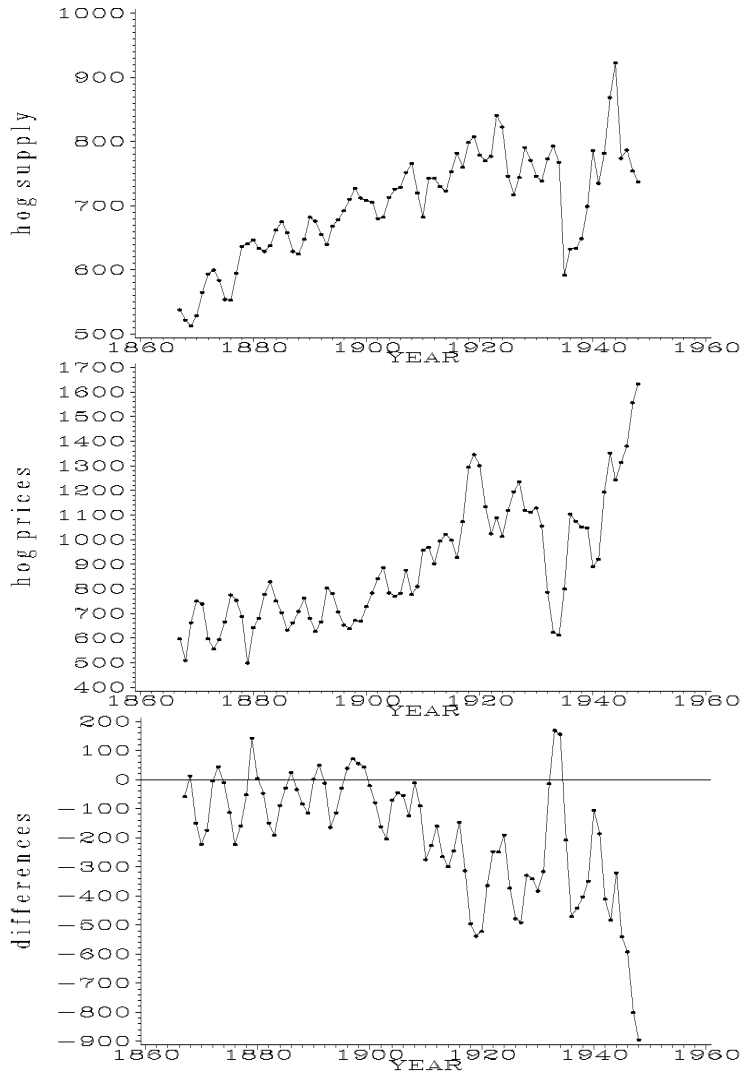


Figure 2.2.6. Hog Data: hog supply and hog prices.

```
***     Program 2_2_6   ***;
TITLE1 'Hog supply, hog prices and differences';
TITLE2 'Hog Data (1867-1948)';


DATA data1;
  INFILE 'c:\data\hogsuppl.txt';
  INPUT supply @@;

DATA data2;
  INFILE 'c:\data\hogprice.txt';
  INPUT price @@;


DATA data3;
  MERGE data1 data2;
  year=_N_+1866;
  diff=supply-price;


SYMBOL1 V=DOT C=GREEN I=JOIN H=0.5 W=1;
AXIS1 LABEL=(ANGLE=90 'h o g \ s u p p l y');
AXIS2 LABEL=(ANGLE=90 'h o g \ p r i c e s');
AXIS3 LABEL=(ANGLE=90 'd i f f e r e n c e s');

GOPTIONS NODISPLAY;
PROC GPLOT DATA=data3 GOUT=abb;
  PLOT supply*year / VAXIS=AXIS1;
  PLOT price*year / VAXIS=AXIS2;
  PLOT diff*year / VAXIS=AXIS3 VREF=0;
RUN;

GOPTIONS DISPLAY;
PROC GREPLAY NOFS IGOUT=abb TC=SASHELP.TEMPLT;
  TEMPLATE=V3;
  TREPLAY 1:GPLOT 2:GPLOT1 3:GPLOT2;
RUN; DELETE _ALL_; QUIT;
```

The `supply` data and the `price` data read in from two external files are merged in `data3`. `Year` is an additional variable with values $1867, 1868, \ldots, 1932$. By `PROC GPLOT` hog `supply`, hog `prices` and their differences `diff` are plotted in three different plots stored in the graphics catalog `abb`. The horizontal line at the zero level is plotted by the option `VREF=0`. The plots are put into a common graphic using `PROC GREPLAY` and the template `V3`. Note that the labels of the vertical axes are spaced out as SAS sets their characters too close otherwise.

Hog supply (=: y_t) and hog price (=: z_t) obviously increase in time t and do, therefore, not seem to be realizations of stationary processes; nevertheless, as they behave similarly, a linear combination of both might be stationary. In this case, hog supply and hog price would be cointegrated.

This phenomenon can easily be explained as follows. A high price z_t at time t is a good reason for farmers to breed more hogs, thus leading to a large supply y_{t+1} in the next year $t+1$. This makes the price z_{t+1} fall with the effect that farmers will reduce their supply of hogs in the following year $t+2$. However, when hogs are in short supply, their price z_{t+2} will rise etc. There is obviously some *error correction mechanism* inherent in these two processes, and the observed cointegration helps us to detect its existence.

```
The AUTOREG Procedure

Dependent Variable    supply


               Ordinary Least Squares Estimates

SSE                 338324.258    DFE                       80
MSE                       4229    Root MSE            65.03117
SBC                 924.172704    AIC                919.359266
Regress R-Square        0.3902    Total R-Square         0.3902
Durbin-Watson           0.5839


        Phillips-Ouliaris
       Cointegration Test

Lags           Rho           Tau

   1      -28.9109       -4.0142


                                    Standard                  Approx
Variable        DF     Estimate        Error    t Value    Pr > |t|

Intercept        1     515.7978      26.6398      19.36      <.0001
```

```
price           1       0.2059       0.0288       7.15       <.0001
```

Figure 2.2.7. Phillips–Ouliaris test for cointegration of Hog Data. (Using `PROC AUTOREG` with `STATIONARITY=(PHILLIPS)` option.)

```
***     Program 2_2_7   ***;
TITLE1 'Testing for cointegration';
TITLE2 'Hog Data (1867-1948)';


PROC AUTOREG DATA=data3;
  MODEL supply=price / STATIONARITY=(PHILLIPS);
RUN; QUIT;
```

The procedure `AUTOREG` (for autoregressive models) uses `data3` from Program 2_2_6. In the `MODEL` statement a regression from supply on price is defined and the option `STATIONARITY=(PHILLIPS)` makes SAS calculate the statistics of the Phillips–Ouliaris test for cointegration of order 1.

The output of the above program contains some characteristics of the regression, the Phillips-Ouliaris test statistics and the regression coefficients with their t-ratios. The Phillips-Ouliaris test statistics need some further explanation.

The hypothesis of the Phillips–Ouliaris cointegration test is *no cointegration*. Unfortunately SAS does not provide the p-value, but only the values of the test statistics denoted by `RHO` and `TAU`. Tables of critical values of these test statistics can be found in Phillips and Ouliaris (1990). Note that in the original paper the two test statistics are denoted by $\hat{Z}_\alpha$ and $\hat{Z}_t$. The hypothesis is to be rejected if `RHO` or `TAU` are below the critical value for the desired type I level error α. For this one has to differentiate between the following cases.

(1) If the estimated cointegrating regression does not contain any intercept, i.e. $\beta_0 = 0$, and none of the explanatory variables has a nonzero trend component, then use the following table for critical values of `RHO` and `TAU`. This is the so-called *standard* case.

α	0.15	0.125	0.1	0.075	0.05	0.025	0.01
RHO	-10.74	-11.57	-12.54	-13.81	-15.64	-18.88	-22.83
TAU	-2.26	-2.35	-2.45	-2.58	-2.76	-3.05	-3.39

(2) If the estimated cointegrating regression contains an intercept, i.e. $\beta_0 \neq 0$, and none of the explanatory variables has a nonzero trend component, then use the following table for critical values of `RHO` and `TAU`. This case is referred to as *demeaned*.

α	0.15	0.125	0.1	0.075	0.05	0.025	0.01
`RHO`	-14.91	-15.93	-17.04	-18.48	-20.49	-23.81	-28.32
`TAU`	-2.86	-2.96	-3.07	-3.20	-3.37	-3.64	-3.96

(3) If the estimated cointegrating regression contains an intercept, i.e. $\beta_0 \neq 0$, and at least one of the explanatory variables has a nonzero trend component, then use the following table for critical values of `RHO` and `TAU`. This case is said to be *demeaned and detrended*.

α	0.15	0.125	0.1	0.075	0.05	0.025	0.01
`RHO`	-20.79	-21.81	-23.19	-24.75	-27.09	-30.84	-35.42
`TAU`	-3.33	-3.42	-3.52	-3.65	-3.80	-4.07	-4.36

In our example with an arbitrary β_0 and a visible trend in the investigated time series, the `RHO`-value is -28.9109 and the `TAU`-value -4.0142. Both are smaller than the critical values of -27.09 and -3.80 in the above table of the demeaned and detrended case and thus, lead to a rejection of the null hypothesis of no cointegration at the 5%-level.

For further information on cointegration we refer to Chapter 19 of the time series book by Hamilton (1994).

ARCH and GARCH-Processes

In particular the monitoring of stock prices gave rise to the idea that the *volatility* of a time series (Y_t) might not be a constant but rather a random variable, which depends on preceding realizations. The following approach to model such a change in volatility is due to Engle (1982).

We assume the multiplicative model

$$Y_t = \sigma_t Z_t, \qquad t \in \mathbb{Z},$$

where the Z_t are independent and identically distributed random variables with

$$\mathrm{E}(Z_t) = 0 \text{ and } \mathrm{E}(Z_t^2) = 1, \qquad t \in \mathbb{Z}.$$

The *scale* σ_t is supposed to be a function of the past p values of the series:

$$\sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j Y_{t-j}^2, \qquad t \in \mathbb{Z},$$

where $p \in \{0, 1, \ldots\}$ and $a_0 > 0$, $a_j \geq 0$, $1 \leq j \leq p-1$, $a_p > 0$ are constants.

The particular choice $p = 0$ yields obviously a white noise model for (Y_t). Common choices for the distribution of Z_t are the standard normal distribution or the (standardized) t-distribution, which in the nonstandardized form has the density

$$f_m(x) := \frac{\Gamma((m+1)/2)}{\Gamma(m/2)\sqrt{\pi m}} \left(1 + \frac{x^2}{m}\right)^{-(m+1)/2}, \qquad x \in \mathbb{R}.$$

The number $m \geq 1$ is the *degree of freedom* of the t-distribution. The scale σ_t in the above model is determined by the past observations $Y_{t-1}, \ldots, Y_{t-p}$, and the *innovation* on this scale is then provided by Z_t. We assume moreover that the process (Y_t) is a causal one in the sense that Z_t and Y_s, $s < t$, are independent. Some autoregressive structure is, therefore, inherent in the process (Y_t). Conditional on $Y_{t-j} = y_{t-j}$, $1 \leq j \leq p$, the variance of Y_t is $a_0 + \sum_{j=1}^p a_j y_{t-j}^2$ and, thus, the conditional variances of the process will generally be different. The process $Y_t = \sigma_t Z_t$ is, therefore, called an *autoregressive* and *conditional heteroscedastic* process of order p, *ARCH*(p)-process for short.

If, in addition, the causal process (Y_t) is stationary, then we obviously have

$$\mathrm{E}(Y_t) = \mathrm{E}(\sigma_t)\,\mathrm{E}(Z_t) = 0$$

and

$$\begin{aligned} \sigma^2 := \mathrm{E}(Y_t^2) &= \mathrm{E}(\sigma_t^2)\,\mathrm{E}(Z_t^2) \\ &= a_0 + \sum_{j=1}^p a_j\,\mathrm{E}(Y_{t-j}^2) = a_0 + \sigma^2 \sum_{j=1}^p a_j, \end{aligned}$$

which yields

$$\sigma^2 = \frac{a_0}{1 - \sum_{j=1}^p a_j}.$$

A necessary condition for the stationarity of the process (Y_t) is, therefore, the inequality $\sum_{j=1}^p a_j < 1$. Note, moreover, that the preceding arguments immediately imply that the Y_t and Y_s are uncorrelated for different values of s and t, but they are not independent.

The following lemma is crucial. It embeds the *ARCH*(p)-processes to a certain extent into the class of *AR*(p)-processes, so that our above tools for the analysis of autoregressive processes can be applied here as well.

Lemma 2.2.12. *Let (Y_t) be a stationary and causal ARCH(p)-process with constants $a_0, a_1, \ldots, a_p$. If the process of squared random variables (Y_t^2) is a stationary one, then it is an AR(p)-process:*

$$Y_t^2 = a_1 Y_{t-1}^2 + \ldots + a_p Y_{t-p}^2 + \varepsilon_t,$$

where (ε_t) *is a white noise with* $\mathrm{E}(\varepsilon_t) = a_0$, $t \in \mathbb{Z}$.

Proof. From the assumption that (Y_t) is an $ARCH(p)$-process we obtain

$$\varepsilon_t := Y_t^2 - \sum_{j=1}^{p} a_j Y_{t-j}^2 = \sigma_t^2 Z_t^2 - \sigma_t^2 + a_0 = a_0 + \sigma_t^2 (Z_t^2 - 1), \qquad t \in \mathbb{Z}.$$

This implies $\mathrm{E}(\varepsilon_t) = a_0$ and

$$\begin{aligned} \mathrm{E}((\varepsilon_t - a_0)^2) &= \mathrm{E}(\sigma_t^4)\,\mathrm{E}((Z_t^2 - 1)^2) \\ &= \mathrm{E}\left((a_0 + \sum_{j=1}^{p} a_j Y_{t-j}^2)^2\right) \mathrm{E}((Z_t^2 - 1)^2) =: c, \end{aligned}$$

independent of t by the stationarity of (Y_t^2). For $h \in \mathbb{N}$ the causality of (Y_t) finally implies

$$\begin{aligned} \mathrm{E}((\varepsilon_t - a_0)(\varepsilon_{t+h} - a_0)) &= \mathrm{E}(\sigma_t^2 \sigma_{t+h}^2 (Z_t^2 - 1)(Z_{t+h}^2 - 1)) \\ &= \mathrm{E}(\sigma_t^2 \sigma_{t+h}^2 (Z_t^2 - 1))\,\mathrm{E}(Z_{t+h}^2 - 1) = 0 \end{aligned}$$

i.e., (ε_t) is a white noise with $\mathrm{E}(\varepsilon_t) = a_0$. □

The process (Y_t^2) satisfies, therefore, the stationarity condition (2.3) if all p roots of the equation $1 - \sum_{j=1}^{p} a_j z^j = 0$ are outside of the unit circle. Hence, we can estimate the order p using an estimate as in (2.9) of the partial autocorrelation function of (Y_t^2). The Yule–Walker equations provide us, for example, with an estimate of the coefficients $a_1, \ldots, a_p$, which then can be utilized to estimate the expectation a_0 of the error ε_t.

Note that conditional on $Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p}$, the distribution of $Y_t = \sigma_t Z_t$ is a normal one if the Z_t are normally distributed. In this case it is possible to write down explicitly the joint density of the vector $(Y_{p+1}, \ldots, Y_n)$, conditional on $Y_1 = y_1, \ldots, Y_p = y_p$ (Exercise 36). A numerical maximization of this density with respect to $a_0, a_1, \ldots, a_p$ then leads to a *maximum likelihood estimate* of the vector of constants; see also Section 2.3.

A *generalized ARCH*-process, $GARCH(p, q)$ for short (Bollerslev (1986)), adds an autoregressive structure to the scale σ_t by assuming the representation

$$\sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j Y_{t-j}^2 + \sum_{k=1}^{q} b_k \sigma_{t-k}^2,$$

where the constants b_k are nonnegative. The set of parameters a_j, b_k can again be estimated by conditional maximum likelihood as before if a parametric model for the distribution of the innovations Z_t is assumed.

Example 2.2.13. (Hongkong Data). The daily Hang Seng closing index was recorded between July 16th, 1981 and September 30th, 1983, leading to a total amount of 552 observations p_t. The daily *log returns* are defined as

$$y_t := \log(p_t) - \log(p_{t-1}),$$

where we now have a total of $n = 551$ observations. The expansion $\log(1+\varepsilon) \approx \varepsilon$ implies that

$$y_t = \log\left(1 + \frac{p_t - p_{t-1}}{p_{t-1}}\right) \approx \frac{p_t - p_{t-1}}{p_{t-1}},$$

provided that p_{t-1} and p_t are close to each other. In this case we can interpret the return as the difference of indices on subsequent days, relative to the initial one.

We use an $ARCH(3)$ model for the generation of y_t, which seems to be a plausible choice by the partial autocorrelations plot. If one assumes t-distributed innovations Z_t, SAS estimates the distribution's degrees of freedom and displays the reciprocal in the `TDFI`-line, here $m = 1/0.1780 = 5.61$ degrees of freedom. Following we obtain the estimates $a_0 = 0.000214$, $a_1 = 0.147593$, $a_2 = 0.278166$ and $a_3 = 0.157807$. The SAS output also contains some general regression model information from an ordinary least squares estimation approach, some specific information for the (G)ARCH approach and as mentioned above the estimates for the ARCH model parameters in combination with t ratios and approximated p-values. The following plots show the returns of the Hang Seng index, their squares, the pertaining partial autocorrelation function and the parameter estimates.

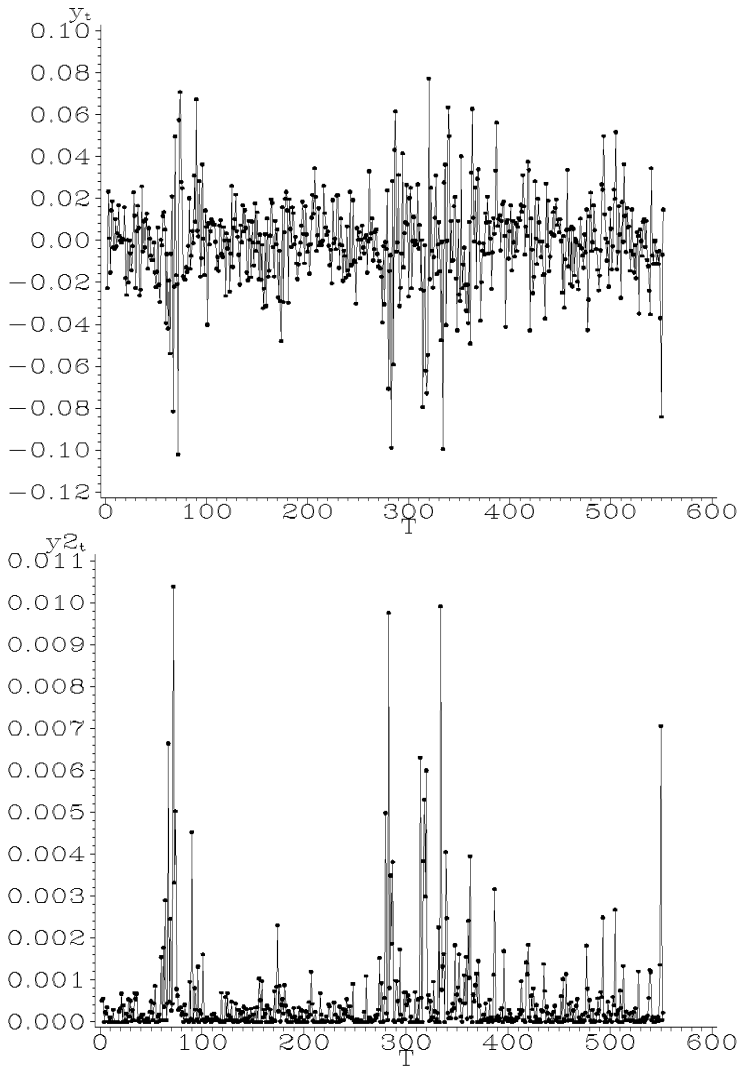
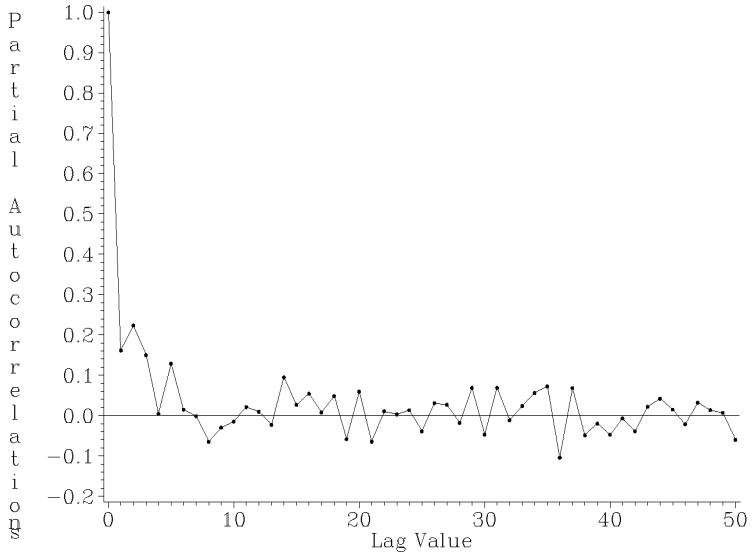


Figure 2.2.8. Log returns of Hang Seng index and their squares.

```
***     Program 2_2_8   ***;
TITLE1 'Daily log returns and their squares';
TITLE2 'Hongkong Data ';

DATA data1;
  INFILE 'c:\data\hongkong.txt';
  INPUT p;
  t=_N_;
  y=DIF(LOG(p));
  y2=y**2;

SYMBOL1 C=RED V=DOT H=0.5 I=JOIN L=1;
AXIS1 LABEL=('y' H=1 't') ORDER=(-.12 TO .10 BY .02);
AXIS2 LABEL=('y2' H=1 't');
GOPTIONS NODISPLAY;
PROC GPLOT DATA=data1 GOUT=abb;
  PLOT y*t / VAXIS=AXIS1;
  PLOT y2*t / VAXIS=AXIS2;
RUN;

GOPTIONS DISPLAY;
PROC GREPLAY NOFS IGOUT=abb TC=SASHELP.TEMPLT;
  TEMPLATE=V2;
  TREPLAY 1:GPLOT 2:GPLOT1;
RUN; DELETE _ALL_; QUIT;
```

In the `DATA` step the observed values of the Hang Seng closing index are read into the variable `p` from an external file. The time index variable `t` uses the SAS-variable `_N_`, and the log transformed and differenced values of the index are stored in the variable `y`, their squared values in `y2`.

After defining different axis labels, two plots are generated by two `PLOT` statements in `PROC GPLOT`, but they are not displayed. By means of `PROC GREPLAY` the plots are merged vertically in one graphic.



Autoreg Procedure

Dependent Variable = Y

Ordinary Least Squares Estimates

SSE	0.265971	DFE	551
MSE	0.000483	Root MSE	0.021971
SBC	-2643.82	AIC	-2643.82
Reg Rsq	0.0000	Total Rsq	0.0000
Durbin-Watson	1.8540		

NOTE: No intercept term is used. R-squares are redefined.

GARCH Estimates

SSE	0.265971	OBS	551
MSE	0.000483	UVAR	0.000515
Log L	1706.532	Total Rsq	0.0000
SBC	-3381.5	AIC	-3403.06
Normality Test	119.7698	Prob>Chi-Sq	0.0001

Variable	DF	B Value	Std Error	t Ratio	Approx Prob

ARCH0	1	0.000214	0.000039	5.444	0.0001
ARCH1	1	0.147593	0.0667	2.213	0.0269
ARCH2	1	0.278166	0.0846	3.287	0.0010
ARCH3	1	0.157807	0.0608	2.594	0.0095
TDFI	1	0.178074	0.0465	3.833	0.0001

Figure 2.2.9. Partial autocorrelations of squares of log returns of Hang Seng index and parameter estimates in the $ARCH(3)$ model for stock returns.

```
***     Program 2_2_9   ***;
TITLE1 'ARCH(3)-model';
TITLE2 'Hongkong Data';
* Note that this program needs data1
         generated by program 2_2_8;

PROC ARIMA DATA=data1;
  IDENTIFY VAR=y2 NLAG=50 OUTCOV=data2;

SYMBOL1 C=RED V=DOT H=0.5 I=JOIN;
PROC GPLOT DATA=data2;
  PLOT partcorr*lag / VREF=0;
RUN;

PROC AUTOREG DATA=data1;
  MODEL y = / NOINT GARCH=(q=3) DIST=T;
RUN;
```

To identify the order of a possibly underlying ARCH process for the daily log returns of the Hang Seng closing index, the empirical partial autocorrelations of their squared values, which are stored in the variable `y2` of the data set `data1` in Program 2_2_8, are calculated by means of `PROC ARIMA` and the `IDENTIFY` statement. The subsequent procedure `GPLOT` displays these partial autocorrelations. A horizontal reference line helps to decide whether a value is substantially different from 0.

`PROC AUTOREG` is used to analyze the ARCH(3) model for the daily log returns. The `MODEL` statement specifies the dependent variable `y`. The option `NOINT` suppresses an intercept parameter, `GARCH=(q=3)` selects the ARCH(3) model and `DIST=T` determines a t distribution for the innovations Z_t in the model equation. Note that, in contrast to our notation, SAS uses the letter `q` for the ARCH model order.

2.3 Specification of ARMA-Models: The Box–Jenkins Program

The aim of this section is to fit a time series model $(Y_t)_{t\in\mathbb{Z}}$ to a given set of data $y_1, \ldots, y_n$ collected in time t. We suppose that the data $y_1, \ldots, y_n$ are (possibly) variance-stabilized as well as trend or seasonally adjusted. We assume that they were generated by clipping $Y_1, \ldots, Y_n$ from an $ARMA(p,q)$-process $(Y_t)_{t\in\mathbb{Z}}$, which we will fit to the data in the following. As noted in Section 2.2, we could also fit the model $Y_t = \sum_{v\geq 0} \alpha_v \varepsilon_{t-v}$ to the data, where (ε_t) is a white noise. But then we would have to determine infinitely many parameters α_v, $v \geq 0$. By the *principle of parsimony* it seems, however, reasonable to fit only the finite number of parameters of an $ARMA(p,q)$-process.

The Box–Jenkins program consists of four steps:

1. **Order selection:** Choice of the parameters p and q.

2. **Estimation of coefficients:** The coefficients $a_1, \ldots, a_p$ and $b_1, \ldots, b_q$ are estimated.

3. **Diagnostic check:** The fit of the $ARMA(p,q)$-model with the estimated coefficients is checked.

4. **Forecasting:** The prediction of future values of the original process.

The four steps are discussed in the following.

Order Selection

The order q of a moving average $MA(q)$-process can be estimated by means of the empirical autocorrelation function $r(k)$ i.e., by the correlogram. Lemma 2.2.1 (iv) shows that the autocorrelation function $\rho(k)$ vanishes for $k \geq q+1$. This suggests to choose the order q such that $r(q)$ is clearly different from zero, whereas $r(k)$ for $k \geq q+1$ is quite close to zero. This, however, is obviously a rather vague selection rule.

The order p of an $AR(p)$-process can be estimated in an analogous way using the empirical partial autocorrelation function $\hat{\alpha}(k)$, $k \geq 1$, as defined in (2.9). Since $\hat{\alpha}(p)$ should be close to the p-th coefficient a_p of the $AR(p)$-process, which is different from zero, whereas $\hat{\alpha}(k) \approx 0$ for $k > p$, the above rule can be applied again with r replaced by $\hat{\alpha}$.

The choice of the orders p and q of an $ARMA(p,q)$-process is a bit more challenging. In this case one commonly takes the pair (p,q), minimizing some function, which is based on an estimate $\hat{\sigma}^2_{p,q}$ of the variance of ε_0. Popular functions are *Akaike's Information Criterion*

$$AIC(p,q) := \log(\hat{\sigma}^2_{p,q}) + 2\frac{p+q+1}{n+1},$$

the *Bayesian Information Criterion*

$$BIC(p,q) := \log(\hat{\sigma}^2_{p,q}) + \frac{(p+q)\log(n+1)}{n+1}$$

and the *Hannan-Quinn Criterion*

$$HQ(p,q) := \log(\hat{\sigma}^2_{p,q}) + \frac{2(p+q)c\log(\log(n+1))}{n+1} \quad \text{with } c > 1.$$

AIC and *BIC* are discussed in Section 9.3 of Brockwell and Davis (1991) for Gaussian processes (Y_t), where the joint distribution of an arbitrary vector $(Y_{t_1}, \ldots, Y_{t_k})$ with $t_1 < \ldots < t_k$ is multivariate normal, see below. For the *HQ*-criterion we refer to Hannan and Quinn (1979). Note that the variance estimate $\hat{\sigma}^2_{p,q}$ will in general become arbitrarily small as $p+q$ increases. The additive terms in the above criteria serve, therefore, as penalties for large values, thus helping to prevent *overfitting* of the data by choosing p and q too large.

Estimation of Coefficients

Suppose we fixed the order p and q of an $ARMA(p,q)$-process $(Y_t)_{t\in\mathbb{Z}}$, with $Y_1, \ldots, Y_n$ now modelling the data $y_1, \ldots, y_n$. In the next step we will derive estimators of the constants $a_1, \ldots, a_p$, $b_1, \ldots, b_q$ in the model

$$Y_t = a_1 Y_{t-1} + \ldots + a_p Y_{t-p} + \varepsilon_t + b_1\varepsilon_{t-1} + \ldots + b_q\varepsilon_{t-q}, \qquad t \in \mathbb{Z}.$$

The Gaussian Model: Maximum Likelihood Estimator

We assume first that (Y_t) is a *Gaussian process* and thus, the joint distribution of $(Y_1, \ldots, Y_n)$ is a n-dimensional normal distribution

$$P\{Y_i \le s_i,\ i = 1, \ldots, n\} = \int_{-\infty}^{s_1} \ldots \int_{-\infty}^{s_n} \boldsymbol{\varphi}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(x_1, \ldots, x_n)\, dx_n \ldots dx_1$$

for arbitrary $s_1, \ldots, s_n \in \mathbb{R}$. Here

$$\begin{aligned}&\boldsymbol{\varphi}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(x_1, \ldots, x_n)\\ &= \frac{1}{(2\pi)^{n/2}(\det \boldsymbol{\Sigma})^{1/2}} \exp\Big(-\frac{1}{2}((x_1, \ldots, x_n) - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}((x_1, \ldots, x_n) - \boldsymbol{\mu})^T\Big)\end{aligned}$$

for arbitrary $x_1, \dots, x_n \in \mathbb{R}$ is the density of the n-dimensional normal distribution with mean vector $\boldsymbol{\mu} = (\mu, \dots, \mu)^T \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma} = (\gamma(i-j))_{1 \le i,j \le n}$ denoted by $\boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mu = \mathrm{E}(Y_0)$ and γ is the autocovariance function of the stationary process (Y_t).

The number $\boldsymbol{\varphi_{\mu,\Sigma}}(x_1, \dots, x_n)$ reflects the probability that the random vector $(Y_1, \dots, Y_n)$ realizes close to $(x_1, \dots, x_n)$. Precisely, we have for $\varepsilon \downarrow 0$

$$\begin{aligned} &P\{Y_i \in [x_i - \varepsilon, x_i + \varepsilon],\ i = 1, \dots, n\} \\ &= \int_{x_1-\varepsilon}^{x_1+\varepsilon} \dots \int_{x_n-\varepsilon}^{x_n+\varepsilon} \boldsymbol{\varphi_{\mu,\Sigma}}(z_1, \dots, z_n)\, dz_n \dots dz_1 \approx 2^n \varepsilon^n \boldsymbol{\varphi_{\mu,\Sigma}}(x_1, \dots, x_n). \end{aligned}$$

The *likelihood principle* is the fact that a random variable tends to attain its most likely value and thus, if the vector $(Y_1, \dots, Y_n)$ actually attained the value $(y_1, \dots, y_n)$, the unknown underlying mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ ought to be such that $\boldsymbol{\varphi_{\mu,\Sigma}}(y_1, \dots, y_n)$ is maximized. The computation of these parameters leads to the *maximum likelihood estimator* of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

We assume that the process (Y_t) satisfies the stationarity condition (2.3), in which case $Y_t = \sum_{v \ge 0} \alpha_v \varepsilon_{t-v},\ t \in \mathbb{Z}$, is invertible, where (ε_t) is a white noise and the coefficients α_v depend only on $a_1, \dots, a_p$ and $b_1, \dots, b_q$. Consequently we have for $s \ge 0$

$$\gamma(s) = \mathrm{Cov}(Y_0, Y_s) = \sum_{v \ge 0} \sum_{w \ge 0} \alpha_v \alpha_w \,\mathrm{Cov}(\varepsilon_{-v}, \varepsilon_{s-w}) = \sigma^2 \sum_{v \ge 0} \alpha_v \alpha_{s+v}.$$

The matrix

$$\boldsymbol{\Sigma'} := \sigma^{-2} \boldsymbol{\Sigma},$$

therefore, depends only on $a_1, \dots, a_p$ and $b_1, \dots, b_q$. We can write now the density $\boldsymbol{\varphi_{\mu,\Sigma}}(x_1, \dots, x_n)$ as a function of $\boldsymbol{\vartheta} := (\sigma^2, \mu, a_1, \dots, a_p, b_1, \dots, b_q) \in \mathbb{R}^{p+q+2}$ and $(x_1, \dots, x_n) \in \mathbb{R}^n$

$$\begin{aligned} p(x_1, \dots, x_n | \boldsymbol{\vartheta}) &:= \boldsymbol{\varphi_{\mu,\Sigma}}(x_1, \dots, x_n) \\ &= (2\pi\sigma^2)^{-n/2} (\det \boldsymbol{\Sigma'})^{-1/2} \exp\Big(-\frac{1}{2\sigma^2} Q(\boldsymbol{\vartheta} | x_1, \dots, x_n) \Big), \end{aligned}$$

where

$$Q(\boldsymbol{\vartheta} | x_1, \dots, x_n) := ((x_1, \dots, x_n) - \boldsymbol{\mu}) \boldsymbol{\Sigma'}^{-1} ((x_1, \dots, x_n) - \boldsymbol{\mu})^T$$

is a quadratic function. The *likelihood function* pertaining to the outcome $(y_1, \dots, y_n)$ is

$$L(\boldsymbol{\vartheta} | y_1, \dots, y_n) := p(y_1, \dots, y_n | \boldsymbol{\vartheta}).$$

A parameter $\hat{\boldsymbol{\vartheta}}$ maximizing the likelihood function

$$L(\hat{\boldsymbol{\vartheta}} | y_1, \dots, y_n) = \sup_{\boldsymbol{\vartheta}} L(\boldsymbol{\vartheta} | y_1, \dots, y_n),$$

is then a maximum likelihood estimator of $\boldsymbol{\vartheta}$.

Due to the strict monotonicity of the logarithm, maximizing the likelihood function is in general equivalent to the maximization of the *loglikelihood function*

$$l(\boldsymbol{\vartheta}|y_1,\ldots,y_n) = \log L(\boldsymbol{\vartheta}|y_1,\ldots,y_n).$$

The maximum likelihood estimator $\hat{\boldsymbol{\vartheta}}$ therefore satisfies

$$\begin{aligned} l(\hat{\boldsymbol{\vartheta}}|y_1,\ldots,y_n) &= \sup_{\boldsymbol{\vartheta}} l(\boldsymbol{\vartheta}|y_1,\ldots,y_n) \\ &= \sup_{\boldsymbol{\vartheta}} \Big(-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log(\det \boldsymbol{\Sigma}') - \frac{1}{2\sigma^2} Q(\boldsymbol{\vartheta}|y_1,\ldots,y_n) \Big). \end{aligned}$$

The computation of a maximizer is a numerical and usually computer intensive problem.

Example 2.3.1. The $AR(1)$-process $Y_t = aY_{t-1} + \varepsilon_t$ with $|a| < 1$ has by Example 2.2.4 the autocovariance function

$$\gamma(s) = \sigma^2 \frac{a^s}{1-a^2}, \qquad s \geq 0,$$

and thus,

$$\boldsymbol{\Sigma}' = \frac{1}{1-a^2} \begin{pmatrix} 1 & a & a^2 & \ldots & a^{n-1} \\ a & 1 & a & & a^{n-2} \\ \vdots & & & \ddots & \vdots \\ a^{n-1} & a^{n-2} & a^{n-3} & \ldots & 1 \end{pmatrix}.$$

The inverse matrix is

$$\boldsymbol{\Sigma}'^{-1} = \begin{pmatrix} 1 & -a & 0 & 0 & \ldots & 0 \\ -a & 1+a^2 & -a & 0 & & 0 \\ 0 & -a & 1+a^2 & -a & & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \ldots & & -a & 1+a^2 & -a \\ 0 & 0 & \cdots & & 0 & -a \quad 1 \end{pmatrix}.$$

Check that the determinant of $\boldsymbol{\Sigma}'^{-1}$ is $\det(\boldsymbol{\Sigma}'^{-1}) = 1 - a^2 = 1/\det(\boldsymbol{\Sigma}')$, see Exercise 40. If (Y_t) is a Gaussian process, then the likelihood function of $\boldsymbol{\vartheta} = (\sigma^2, \mu, a)$ is given by

$$L(\boldsymbol{\vartheta}|y_1,\ldots,y_n) = (2\pi\sigma^2)^{-n/2}(1-a^2)^{1/2} \exp\Big(-\frac{1}{2\sigma^2} Q(\boldsymbol{\vartheta}|y_1,\ldots,y_n) \Big),$$

where

$$\begin{aligned} & Q(\boldsymbol{\vartheta}|y_1,\ldots,y_n) \\ &= ((y_1,\ldots,y_n)-\boldsymbol{\mu})\boldsymbol{\Sigma'}^{-1}((y_1,\ldots,y_n)-\boldsymbol{\mu})^T \\ &= (y_1-\mu)^2+(y_n-\mu)^2+(1+a^2)\sum_{i=2}^{n-1}(y_i-\mu)^2-2a\sum_{i=1}^{n-1}(y_i-\mu)(y_{i+1}-\mu). \end{aligned}$$

A Nonparametric Approach: Least Squares

If $\mathrm{E}(\varepsilon_t)=0$, then

$$\hat{Y}_t = a_1Y_{t-1}+\ldots+a_pY_{t-p}+b_1\varepsilon_{t-1}+\ldots+b_q\varepsilon_{t-q}$$

would obviously be a reasonable one-step forecast of the $ARMA(p,q)$-process

$$Y_t = a_1Y_{t-1}+\ldots+a_pY_{t-p}+\varepsilon_t+b_1\varepsilon_{t-1}+\ldots+b_q\varepsilon_{t-q},$$

based on $Y_{t-1},\ldots,Y_{t-p}$ and $\varepsilon_{t-1},\ldots,\varepsilon_{t-q}$. The prediction error is given by the residual

$$Y_t-\hat{Y}_t=\varepsilon_t.$$

Suppose that $\hat{\varepsilon}_t$ is an estimator of ε_t, $t\le n$, which depends on the choice of constants $a_1,\ldots,a_p,b_1,\ldots,b_q$ and satisfies the recursion

$$\hat{\varepsilon}_t = y_t-a_1y_{t-1}-\ldots-a_py_{t-p}-b_1\hat{\varepsilon}_{t-1}-\ldots-b_q\hat{\varepsilon}_{t-q}.$$

The function

$$\begin{aligned} & S^2(a_1,\ldots,a_p,\,b_1,\ldots,b_q) \\ &= \sum_{t=-\infty}^{n}\hat{\varepsilon}_t^2 \\ &= \sum_{t=-\infty}^{n}(y_t-a_1y_{t-1}-\ldots-a_py_{t-p}-b_1\hat{\varepsilon}_{t-1}-\ldots-b_q\hat{\varepsilon}_{t-q})^2 \end{aligned}$$

is the *residual sum of squares* and the *least squares approach* suggests to estimate the underlying set of constants by minimizers $a_1,\ldots,a_p,b_1,\ldots,b_q$ of S^2. Note that the residuals $\hat{\varepsilon}_t$ and the constants are nested.

We have no observation y_t available for $t\le 0$. But from the assumption $\mathrm{E}(\varepsilon_t)=0$ and thus $\mathrm{E}(Y_t)=0$, it is reasonable to *backforecast* y_t by zero and to put $\hat{\varepsilon}_t:=0$ for $t\le 0$, leading to

$$S^2(a_1,\ldots,a_p,\,b_1,\ldots,b_q)=\sum_{t=1}^{n}\hat{\varepsilon}_t^2.$$

The estimated residuals $\hat{\varepsilon}_t$ can then be computed from the recursion

$$\begin{aligned}
\hat{\varepsilon}_1 &= y_1 \\
\hat{\varepsilon}_2 &= y_2 - a_1 y_1 - b_1 \hat{\varepsilon}_1 \\
\hat{\varepsilon}_3 &= y_3 - a_1 y_2 - a_2 y_1 - b_1 \hat{\varepsilon}_2 - b_2 \hat{\varepsilon}_1 \\
&\vdots \\
\hat{\varepsilon}_j &= y_j - a_1 y_{j-1} - \ldots - a_p y_{j-p} - b_1 \hat{\varepsilon}_{j-1} - \ldots - b_q \hat{\varepsilon}_{j-q},
\end{aligned}$$

where j now runs from $\max\{p, q\}$ to n.

The coefficients $a_1, \ldots, a_p$ of a pure $AR(p)$-process can be estimated directly, using the Yule-Walker equations as described in (2.6).

Diagnostic Check

Suppose that the orders p and q as well as the constants $a_1, \ldots, a_p, b_1, \ldots, b_q$ have been chosen in order to model an $ARMA(p, q)$-process underlying the data. The *Portmanteau-test* of Box and Pierce (1970) checks, whether the estimated residuals $\hat{\varepsilon}_t$, $t = 1, \ldots, n$, behave approximately like realizations from a white noise process. To this end one considers the pertaining empirical autocorrelation function

$$\hat{r}_\varepsilon(k) := \frac{\sum_{j=1}^{n-k} (\hat{\varepsilon}_j - \bar{\varepsilon})(\hat{\varepsilon}_{j+k} - \bar{\varepsilon})}{\sum_{j=1}^{n} (\hat{\varepsilon}_j - \bar{\varepsilon})^2}, \qquad k = 1, \ldots, n-1,$$

where $\bar{\varepsilon} = n^{-1} \sum_{j=1}^{n} \hat{\varepsilon}_j$, and checks, whether the values $\hat{r}_\varepsilon(k)$ are sufficiently close to zero. This decision is based on

$$Q(K) := n \sum_{k=1}^{K} \hat{r}_\varepsilon^2(k),$$

which follows asymptotically for $n \to \infty$ a χ^2-distribution with $K-p-q$ degrees of freedom if (Y_t) is actually an $ARMA(p, q)$-process (see e.g. Section 9.4 in Brockwell and Davis (1991)). The parameter K must be chosen such that the sample size $n - k$ in $\hat{r}_\varepsilon(k)$ is large enough to give a stable estimate of the autocorrelation function. The $ARMA(p, q)$-model is rejected if the *p-value* $1 - \chi^2_{K-p-q}(Q(K))$ is too small, since in this case the value $Q(K)$ is unexpectedly large. By χ^2_{K-p-q} we denote the distribution function of the χ^2-distribution with $K - p - q$ degrees of freedom. To accelerate the convergence to the χ^2_{K-p-q} distribution under the null hypothesis of an $ARMA(p, q)$-process, one often replaces the Box–Pierce statistic $Q(K)$ by the Box–Ljung statistic (Ljung and Box (1978))

$$Q^*(K) := n \sum_{k=1}^{K} \left(\left(\frac{n+2}{n-k} \right)^{1/2} \hat{r}_\varepsilon(k) \right)^2 = n(n+2) \sum_{k=1}^{K} \frac{1}{n-k} \hat{r}_\varepsilon^2(k)$$

with weighted empirical autocorrelations.

Forecasting

We want to determine weights $c_0^*, \ldots, c_{n-1}^* \in \mathbb{R}$ such that for $h \in \mathbb{N}$

$$\mathrm{E}\left(\left(Y_{n+h} - \sum_{u=0}^{n-1} c_u^* Y_{n-u}\right)^2\right) = \min_{c_0,\ldots,c_{n-1}\in\mathbb{R}} \mathrm{E}\left(\left(Y_{n+h} - \sum_{u=0}^{n-1} c_u Y_{n-u}\right)^2\right).$$

Then $\hat{Y}_{n+h} := \sum_{u=0}^{n-1} c_u^* Y_{n-u}$ with minimum mean squared error is said to be a best (linear) h-step forecast of Y_{n+h}, based on $Y_1, \ldots, Y_n$. The following result provides a sufficient condition for the optimality of weights.

Lemma 2.3.2. *Let (Y_t) be an arbitrary stochastic process with finite second moments. If the weights $c_0^*, \ldots, c_{n-1}^*$ have the property that*

$$\mathrm{E}\left(Y_i\left(Y_{n+h} - \sum_{u=0}^{n-1} c_u^* Y_{n-u}\right)\right) = 0, \qquad i = 1, \ldots, n, \tag{2.15}$$

then $\hat{Y}_{n+h} := \sum_{u=0}^{n-1} c_u^* Y_{n-u}$ is a best h-step forecast of Y_{n+h}.

Proof. Let $\tilde{Y}_{n+h} := \sum_{u=0}^{n-1} c_u Y_{n-u}$ be an arbitrary forecast, based on $Y_1, \ldots, Y_n$. Then we have

$$\begin{aligned}
&\mathrm{E}((Y_{n+h} - \tilde{Y}_{n+h})^2) \\
&\quad = \mathrm{E}((Y_{n+h} - \hat{Y}_{n+h} + \hat{Y}_{n+h} - \tilde{Y}_{n+h})^2) \\
&\quad = \mathrm{E}((Y_{n+h} - \hat{Y}_{n+h})^2) + 2\sum_{u=0}^{n-1}(c_u^* - c_u)\,\mathrm{E}(Y_{n-u}(Y_{n+h} - \hat{Y}_{n+h})) \\
&\qquad\qquad + \mathrm{E}((\hat{Y}_{n+h} - \tilde{Y}_{n+h})^2) \\
&\quad = \mathrm{E}((Y_{n+h} - \hat{Y}_{n+h})^2) + \mathrm{E}((\hat{Y}_{n+h} - \tilde{Y}_{n+h})^2) \\
&\quad \geq \mathrm{E}((Y_{n+h} - \hat{Y}_{n+h})^2).
\end{aligned}$$

□

Suppose that (Y_t) is a stationary process with mean zero and autocorrelation function ρ. The equations (2.15) are then of Yule-Walker type

$$\rho(h+s) = \sum_{u=0}^{n-1} c_u^* \rho(s-u), \qquad s = 0, 1, \ldots, n-1,$$

or, in matrix language

$$\begin{pmatrix} \rho(h) \\ \rho(h+1) \\ \vdots \\ \rho(h+n-1) \end{pmatrix} = \boldsymbol{P}_n \begin{pmatrix} c_0^* \\ c_1^* \\ \vdots \\ c_{n-1}^* \end{pmatrix} \tag{2.16}$$

with the matrix $\boldsymbol{P}_n$ as defined in (2.7). If this matrix is invertible, then

$$\begin{pmatrix} c_0^* \\ \vdots \\ c_{n-1}^* \end{pmatrix} := \boldsymbol{P}_n^{-1} \begin{pmatrix} \rho(h) \\ \vdots \\ \rho(h+n-1) \end{pmatrix} \tag{2.17}$$

is the uniquely determined solution of (2.16).

If we put $h = 1$, then equation (2.17) implies that c_{n-1}^* equals the partial autocorrelation coefficient $\alpha(n)$. In this case, $\alpha(n)$ is the coefficient of Y_1 in the best linear one-step forecast $\hat{Y}_{n+1} = \sum_{u=0}^{n-1} c_u^* Y_{n-u}$ of Y_{n+1}.

Example 2.3.3. Consider the $MA(1)$-process $Y_t = \varepsilon_t + a\varepsilon_{t-1}$ with $\mathrm{E}(\varepsilon_0) = 0$. Its autocorrelation function is by Example 2.2.2 given by $\rho(0) = 1,\ \rho(1) = a/(1+a^2),\ \rho(u) = 0$ for $u \geq 2$. The matrix $\boldsymbol{P}_n$ equals therefore

$$\boldsymbol{P}_n = \begin{pmatrix} 1 & \frac{a}{1+a^2} & 0 & 0 & \cdots & 0 \\ \frac{a}{1+a^2} & 1 & \frac{a}{1+a^2} & 0 & & 0 \\ 0 & \frac{a}{1+a^2} & 1 & \frac{a}{1+a^2} & & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \cdots & & & 1 & \frac{a}{1+a^2} \\ 0 & 0 & \cdots & & \frac{a}{1+a^2} & 1 \end{pmatrix}.$$

Check that the matrix $\boldsymbol{P}_n = (\mathrm{Corr}(Y_i, Y_j))_{1 \leq i,j \leq n}$ is positive definite, $\boldsymbol{x}^T \boldsymbol{P}_n \boldsymbol{x} > 0$ for any $\boldsymbol{x} \in \mathbb{R}^n$ unless $\boldsymbol{x} = \boldsymbol{0}$ (Exercise 40), and thus, $\boldsymbol{P}_n$ is invertible. The best forecast of Y_{n+1} is by (2.17), therefore, $\sum_{u=0}^{n-1} c_u^* Y_{n-u}$ with

$$\begin{pmatrix} c_0^* \\ \vdots \\ c_{n-1}^* \end{pmatrix} = \boldsymbol{P}_n^{-1} \begin{pmatrix} \frac{a}{1+a^2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

which is $a/(1+a^2)$ times the first column of $\boldsymbol{P}_n^{-1}$. The best forecast of Y_{n+h} for $h \geq 2$ is by (2.17) the constant 0. Note that Y_{n+h} is for $h \geq 2$ uncorrelated with $Y_1, \ldots, Y_n$ and thus not really predictable by $Y_1, \ldots, Y_n$.

Theorem 2.3.4. *Suppose that* $Y_t = \sum_{u=1}^{p} a_u Y_{t-u} + \varepsilon_t,\ t \in \mathbb{Z}$, *is a stationary* $AR(p)$*-process, which satisfies the stationarity condition* (2.3) *and has zero mean* $\mathrm{E}(Y_0) = 0$. *Let* $n \geq p$. *The best one-step forecast is*

$$\hat{Y}_{n+1} = a_1 Y_n + a_2 Y_{n-1} + \ldots + a_p Y_{n+1-p}$$

and the best two-step forecast is

$$\hat{Y}_{n+2} = a_1 \hat{Y}_{n+1} + a_2 Y_n + \ldots + a_p Y_{n+2-p}.$$

The best h*-step forecast for arbitrary* $h \geq 2$ *is recursively given by*

$$\hat{Y}_{n+h} = a_1 \hat{Y}_{n+h-1} + \ldots + a_{h-1} \hat{Y}_{n+1} + a_h Y_n + \ldots + a_p Y_{n+h-p}.$$

Proof. Since (Y_t) satisfies the stationarity condition (2.3), it is invertible by Theorem 2.2.3 i.e., there exists an absolutely summable causal filter $(b_u)_{u\geq 0}$ such that $Y_t = \sum_{u\geq 0} b_u \varepsilon_{t-u},\ t \in \mathbb{Z}$, almost surely. This implies in particular $\mathrm{E}(Y_t \varepsilon_{t+h}) = \sum_{u\geq 0} b_u \mathrm{E}(\varepsilon_{t-u}\varepsilon_{t+h}) = 0$ for any $h \geq 1$, cf. Theorem 2.1.5. Hence we obtain for $i = 1, \ldots, n$

$$\mathrm{E}((Y_{n+1} - \hat{Y}_{n+1})Y_i) = \mathrm{E}(\varepsilon_{n+1} Y_i) = 0$$

from which the assertion for $h = 1$ follows by Lemma 2.3.2. The case of an arbitrary $h \geq 2$ is now a consequence of the recursion

$$\begin{aligned}
\mathrm{E}((Y_{n+h} &- \hat{Y}_{n+h})Y_i) \\
&= \mathrm{E}\left(\left(\varepsilon_{n+h} + \sum_{u=1}^{\min(h-1,p)} a_u Y_{n+h-u} - \sum_{u=1}^{\min(h-1,p)} a_u \hat{Y}_{n+h-u}\right) Y_i\right) \\
&= \sum_{u=1}^{\min(h-1,p)} a_u \,\mathrm{E}\left(\left(Y_{n+h-u} - \hat{Y}_{n+h-u}\right) Y_i\right) = 0, \qquad i = 1, \ldots, n,
\end{aligned}$$

and Lemma 2.3.2. □

A repetition of the arguments in the preceding proof implies the following result, which shows that for an $ARMA(p,q)$-process the forecast of Y_{n+h} for $h > q$ is controlled only by the AR-part of the process.

Theorem 2.3.5. *Suppose that* $Y_t = \sum_{u=1}^{p} a_u Y_{t-u} + \varepsilon_t + \sum_{v=1}^{q} b_v \varepsilon_{t-v},\ t \in \mathbb{Z}$, *is an* $ARMA(p,q)$*-process, which satisfies the stationarity condition (2.3) and has zero mean, precisely* $\mathrm{E}(\varepsilon_0) = 0$. *Suppose that* $n + q - p \geq 0$. *The best* h*-step forecast of* Y_{n+h} *for* $h > q$ *satisfies the recursion*

$$\hat{Y}_{n+h} = \sum_{u=1}^{p} a_u \hat{Y}_{n+h-u}.$$

Example 2.3.6. We illustrate the best forecast of the $ARMA(1,1)$-process

$$Y_t = 0.4Y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}, \qquad t \in \mathbb{Z},$$

with $\mathrm{E}(Y_t) = \mathrm{E}(\varepsilon_t) = 0$. First we need the optimal 1-step forecast $\widehat{Y}_i$ for $i = 1, \ldots, n$. These are defined by putting unknown values of Y_t with an index $t \leq 0$ equal to their expected value, which is zero. We, thus, obtain

$$\begin{array}{lll}
\widehat{Y}_1 := 0, & & \hat{\varepsilon}_1 := Y_1 - \widehat{Y}_1 = Y_1, \\
\widehat{Y}_2 := 0.4Y_1 + 0 - 0.6\hat{\varepsilon}_1 & & \\
\quad = -0.2Y_1, & & \hat{\varepsilon}_2 := Y_2 - \widehat{Y}_2 = Y_2 + 0.2Y_1, \\
\widehat{Y}_3 := 0.4Y_2 + 0 - 0.6\hat{\varepsilon}_2 & & \\
\quad = 0.4Y_2 - 0.6(Y_2 + 0.2Y_1) & & \\
\quad = -0.2Y_2 - 0.12Y_1, & & \hat{\varepsilon}_3 := Y_3 - \widehat{Y}_3, \\
\quad \vdots & & \quad \vdots
\end{array}$$

until $\widehat{Y}_i$ and $\hat{\varepsilon}_i$ are defined for $i = 1, \ldots, n$. The actual forecast is then given by

$$\begin{array}{lllll}
\widehat{Y}_{n+1} & = & 0.4Y_n + 0 - 0.6\hat{\varepsilon}_n & = & 0.4Y_n - 0.6(Y_n - \widehat{Y}_n), \\
\widehat{Y}_{n+2} & = & 0.4\widehat{Y}_{n+1} + 0 + 0, & & \\
\quad \vdots & & & & \\
\widehat{Y}_{n+h} & = & 0.4\widehat{Y}_{n+h-1} \quad = \quad \ldots & = & 0.4^{h-1}\widehat{Y}_{n+1} \to_{h\to\infty} 0,
\end{array}$$

where ε_t with index $t \geq n+1$ is replaced by zero, since it is uncorrelated with Y_i, $i \leq n$.

In practice one replaces the usually unknown coefficients a_u, b_v in the above forecasts by their estimated values.

2.4 State-Space Models

In state-space models we have, in general, a nonobservable target process (X_t) and an observable process (Y_t). They are linked by the assumption that (Y_t) is a linear function of (X_t) with an added noise, where the linear function may vary in time. The aim is the derivation of best linear estimates of X_t, based on $(Y_s)_{s \leq t}$.

Many models of time series such as $ARMA(p,q)$-processes can be embedded in state-space models if we allow in the following sequences of random *vectors* $\boldsymbol{X}_t \in \mathbb{R}^k$ and $\boldsymbol{Y}_t \in \mathbb{R}^m$.

A multivariate *state-space model* is now defined by the *state equation*

$$\boldsymbol{X}_{t+1} = \boldsymbol{A}_t \boldsymbol{X}_t + \boldsymbol{B}_t \boldsymbol{\varepsilon}_{t+1} \in \mathbb{R}^k, \tag{2.18}$$

describing the time-dependent behavior of the *state* $\boldsymbol{X}_t \in \mathbb{R}^k$, and the *observation equation*

$$\boldsymbol{Y}_t = \boldsymbol{C}_t \boldsymbol{X}_t + \boldsymbol{\eta}_t \in \mathbb{R}^m. \tag{2.19}$$

We assume that $(\boldsymbol{A}_t)$, $(\boldsymbol{B}_t)$ and $(\boldsymbol{C}_t)$ are sequences of known matrices, $(\boldsymbol{\varepsilon}_t)$ and $(\boldsymbol{\eta}_t)$ are uncorrelated sequences of white noises with mean vectors $\mathbf{0}$ and known covariance matrices $\mathrm{Cov}(\boldsymbol{\varepsilon}_t) = \mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T) =: \boldsymbol{Q}_t$, $\mathrm{Cov}(\boldsymbol{\eta}_t) = \mathrm{E}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^T) =: \boldsymbol{R}_t$.

We suppose further that $\boldsymbol{X}_0$ and $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\eta}_t$, $t \geq 1$, are uncorrelated, where two random vectors $\boldsymbol{W} \in \mathbb{R}^p$ and $\boldsymbol{V} \in \mathbb{R}^q$ are said to be uncorrelated if their components are i.e., if the matrix of their covariances vanishes

$$\mathrm{E}((\boldsymbol{W} - \mathrm{E}(\boldsymbol{W})(\boldsymbol{V} - \mathrm{E}(\boldsymbol{V}))^T) = \mathbf{0}.$$

By $\mathrm{E}(\boldsymbol{W})$ we denote the vector of the componentwise expectations of $\boldsymbol{W}$. We say that a time series $(\boldsymbol{Y}_t)$ has a *state-space representation* if it satisfies the representations (2.18) and (2.19).

Example 2.4.1. Let (η_t) be a white noise in $\mathbb{R}$ and put

$$Y_t := \mu_t + \eta_t$$

with linear trend $\mu_t = a + bt$. This simple model can be represented as a state-space model as follows. Define the state vector $\boldsymbol{X}_t$ as

$$\boldsymbol{X}_t := \begin{pmatrix} \mu_t \\ 1 \end{pmatrix},$$

and put

$$\boldsymbol{A} := \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$$

From the recursion $\mu_{t+1} = \mu_t + b$ we then obtain the state equation

$$\boldsymbol{X}_{t+1} = \begin{pmatrix} \mu_{t+1} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_t \\ 1 \end{pmatrix} = \boldsymbol{A}\boldsymbol{X}_t,$$

and with

$$\boldsymbol{C} := (1, 0)$$

the observation equation

$$Y_t = (1, 0) \begin{pmatrix} \mu_t \\ 1 \end{pmatrix} + \eta_t = \boldsymbol{C}\boldsymbol{X}_t + \eta_t.$$

Note that the state $\boldsymbol{X}_t$ is nonstochastic i.e., $\boldsymbol{B}_t = \mathbf{0}$. This model is moreover time-invariant, since the matrices $\boldsymbol{A}$, $\boldsymbol{B} := \boldsymbol{B}_t$ and $\boldsymbol{C}$ do not depend on t.

Example 2.4.2. An $AR(p)$-process

$$Y_t = a_1 Y_{t-1} + \ldots + a_p Y_{t-p} + \varepsilon_t$$

with a white noise (ε_t) has a state-space representation with state vector

$$\boldsymbol{X}_t = (Y_t, Y_{t-1}, \ldots, Y_{t-p+1})^T.$$

If we define the $p \times p$-matrix $\boldsymbol{A}$ by

$$\boldsymbol{A} := \begin{pmatrix} a_1 & a_2 & \ldots & a_{p-1} & a_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

and the $p \times 1$-matrices $\boldsymbol{B}$, $\boldsymbol{C}^T$ by

$$\boldsymbol{B} := (1, 0, \ldots, 0)^T =: \boldsymbol{C}^T,$$

then we have the state equation

$$\boldsymbol{X}_{t+1} = \boldsymbol{A}\boldsymbol{X}_t + \boldsymbol{B}\varepsilon_{t+1}$$

and the observation equation

$$Y_t = \boldsymbol{C}\boldsymbol{X}_t.$$

Example 2.4.3. For the $MA(q)$-process

$$Y_t = \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q}$$

we define the non observable state

$$\boldsymbol{X}_t := (\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q})^T \in \mathbb{R}^{q+1}.$$

With the $(q+1) \times (q+1)$-matrix

$$\boldsymbol{A} := \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & & 0 & 0 \\ \vdots & & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix},$$

the $(q+1) \times 1$-matrix

$$\boldsymbol{B} := (1, 0, \ldots, 0)^T$$

and the $1 \times (q+1)$-matrix

$$\boldsymbol{C} := (1, b_1, \ldots, b_q)$$

we obtain the state equation

$$\boldsymbol{X}_{t+1} = \boldsymbol{A}\boldsymbol{X}_t + \boldsymbol{B}\varepsilon_{t+1}$$

and the observation equation

$$\boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t.$$

Example 2.4.4. Combining the above results for $AR(p)$ and $MA(q)$-processes, we obtain a state-space representation of $ARMA(p,q)$-processes

$$Y_t = a_1 Y_{t-1} + \ldots + a_p Y_{t-p} + \varepsilon_t + b_1\varepsilon_{t-1} + \ldots + b_q\varepsilon_{t-q}.$$

In this case the state vector can be chosen as

$$\boldsymbol{X}_t := (Y_t, Y_{t-1}, \ldots, Y_{t-p+1}, \varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q+1})^T \in \mathbb{R}^{p+q}.$$

We define the $(p+q)\times(p+q)$-matrix

$$\boldsymbol{A} := \begin{pmatrix} a_1 & a_2 & \ldots & a_{p-1} & a_p & b_1 & b_2 & \ldots & b_{q-1} & b_q \\ 1 & 0 & & 0 & 0 & 0 & & \ldots & & 0 \\ \vdots & \ddots & & & \vdots & & & & & \vdots \\ 0 & 0 & & 1 & 0 & & & & & 0 \\ 0 & & \ldots & & 0 & & & \ldots & & 0 \\ 0 & & & & 0 & 1 & 0 & & & 0 \\ 0 & & & & 0 & 0 & 1 & & & 0 \\ \vdots & & & & \vdots & & & \ddots & & \vdots \\ 0 & & \ldots & 0 & 0 & 0 & & \ldots & 1 & 0 \end{pmatrix},$$

the $(p+q)\times 1$-matrix

$$\boldsymbol{B} := (1, 0, \ldots, 0, 1, 0, \ldots, 0)^T$$

with the entry 1 at the first and $(p+1)$-th position and the $1\times(p+q)$-matrix

$$\boldsymbol{C} := (1, 0, \ldots, 0).$$

Then we have the state equation

$$\boldsymbol{X}_{t+1} = \boldsymbol{A}\boldsymbol{X}_t + \boldsymbol{B}\varepsilon_{t+1}$$

and the observation equation

$$Y_t = \boldsymbol{C}\boldsymbol{X}_t.$$

The Kalman-Filter

The key problem in the state-space model (2.18), (2.19) is the estimation of the nonobservable state $\boldsymbol{X}_t$. It is possible to derive an estimate of $\boldsymbol{X}_t$ recursively from

an estimate of $\boldsymbol{X}_{t-1}$ together with the last observation $\boldsymbol{Y}_t$, known as the *Kalman recursions* (Kalman 1960). We obtain in this way a unified prediction technique for each time series model that has a state-space representation.

We want to compute the best linear prediction

$$\hat{\boldsymbol{X}}_t := \boldsymbol{D}_1\boldsymbol{Y}_1 + \ldots + \boldsymbol{D}_t\boldsymbol{Y}_t \tag{2.20}$$

of $\boldsymbol{X}_t$, based on $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t$ i.e., the $k \times m$-matrices $\boldsymbol{D}_1, \ldots, \boldsymbol{D}_t$ are such that the mean squared error is minimized

$$\begin{aligned}
&\mathrm{E}((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)^T(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)) \\
&= \mathrm{E}((\boldsymbol{X}_t - \sum_{j=1}^{t} \boldsymbol{D}_j\boldsymbol{Y}_j)^T(\boldsymbol{X}_t - \sum_{j=1}^{t} \boldsymbol{D}_j\boldsymbol{Y}_j)) \\
&= \min_{k\times m-\text{matrices } \boldsymbol{D}'_1,\ldots,\boldsymbol{D}'_t} \mathrm{E}((\boldsymbol{X}_t - \sum_{j=1}^{t} \boldsymbol{D}'_j\boldsymbol{Y}_j)^T(\boldsymbol{X}_t - \sum_{j=1}^{t} \boldsymbol{D}'_j\boldsymbol{Y}_j)).
\end{aligned} \tag{2.21}$$

By repeating the arguments in the proof of Lemma 2.3.2 we will prove the following result. It states that $\hat{\boldsymbol{X}}_t$ is a best linear prediction of $\boldsymbol{X}_t$ based on $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t$ if each component of the vector $\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t$ is orthogonal to each component of the vectors $\boldsymbol{Y}_s$, $1 \le s \le t$, with respect to the inner product $\mathrm{E}(XY)$ of two random variable X and Y.

Lemma 2.4.5. *If the estimate $\hat{\boldsymbol{X}}_t$ defined in (2.20) satisfies*

$$\mathrm{E}((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)\boldsymbol{Y}_s^T) = \boldsymbol{0}, \qquad 1 \le s \le t, \tag{2.22}$$

then it minimizes the mean squared error (2.21).

Note that $\mathrm{E}((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)\boldsymbol{Y}_s^T)$ is a $k \times m$-matrix, which is generated by multiplying each component of $\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t \in \mathbb{R}^k$ with each component of $\boldsymbol{Y}_s \in \mathbb{R}^m$.

Proof. Let $\boldsymbol{X}'_t = \sum_{j=1}^{t} \boldsymbol{D}'_j\boldsymbol{Y}_j \in \mathbb{R}^k$ be an arbitrary linear combination of $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t$.

Then we have

$$\begin{aligned}
&\mathrm{E}((\boldsymbol{X}_t - \boldsymbol{X}'_t)^T(\boldsymbol{X}_t - \boldsymbol{X}'_t)) \\
&= \mathrm{E}\left(\left(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t + \sum_{j=1}^{t}(\boldsymbol{D}_j - \boldsymbol{D}'_j)\boldsymbol{Y}_j\right)^T\left(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t + \sum_{j=1}^{t}(\boldsymbol{D}_j - \boldsymbol{D}'_j)\boldsymbol{Y}_j\right)\right) \\
&= \mathrm{E}((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)^T(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)) + 2\sum_{j=1}^{t}\mathrm{E}((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)^T(\boldsymbol{D}_j - \boldsymbol{D}'_j)\boldsymbol{Y}_j) \\
&\qquad + \mathrm{E}\left(\left(\sum_{j=1}^{t}(\boldsymbol{D}_j - \boldsymbol{D}'_j)\boldsymbol{Y}_j\right)^T\sum_{j=1}^{t}(\boldsymbol{D}_j - \boldsymbol{D}'_j)\boldsymbol{Y}_j\right) \\
&\geq \mathrm{E}((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)^T(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)),
\end{aligned}$$

since in the second-to-last line the final term is nonnegative and the second one vanishes by the property (2.22). □

Let now $\hat{\boldsymbol{X}}_{t-1}$ be the best linear prediction of $\boldsymbol{X}_{t-1}$ based on $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1}$. Then

$$\tilde{\boldsymbol{X}}_t := \boldsymbol{A}_{t-1}\hat{\boldsymbol{X}}_{t-1} \tag{2.23}$$

is the best linear prediction of $\boldsymbol{X}_t$ based on $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1}$, which is easy to see. We simply replaced $\boldsymbol{\varepsilon}_t$ in the state equation by its expectation $\mathbf{0}$. Note that $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{Y}_s$ are uncorrelated if $s < t$ i.e., $\mathrm{E}((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)\boldsymbol{Y}_s^T) = \mathbf{0}$ for $1 \leq s \leq t-1$.

From this we obtain that

$$\tilde{\boldsymbol{Y}}_t := \boldsymbol{C}_t\tilde{\boldsymbol{X}}_t$$

is the best linear prediction of $\boldsymbol{Y}_t$ based on $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1}$, since $\mathrm{E}((\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)\boldsymbol{Y}_s^T) = \mathrm{E}((\boldsymbol{C}_t(\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t) + \boldsymbol{\eta}_t)\boldsymbol{Y}_s^T) = \mathbf{0},\ 1 \leq s \leq t-1$; note that $\boldsymbol{\eta}_t$ and $\boldsymbol{Y}_s$ are uncorrelated if $s < t$.

Define now by

$$\boldsymbol{\Delta}_t := \mathrm{E}((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)^T) \quad \text{and} \quad \tilde{\boldsymbol{\Delta}}_t := \mathrm{E}((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)(\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)^T).$$

the covariance matrices of the approximation errors. Then we have

$$\begin{aligned}
\tilde{\boldsymbol{\Delta}}_t &= \mathrm{E}((\boldsymbol{A}_{t-1}(\boldsymbol{X}_{t-1} - \hat{\boldsymbol{X}}_{t-1}) + \boldsymbol{B}_{t-1}\boldsymbol{\varepsilon}_t)(\boldsymbol{A}_{t-1}(\boldsymbol{X}_{t-1} - \hat{\boldsymbol{X}}_{t-1}) + \boldsymbol{B}_{t-1}\boldsymbol{\varepsilon}_t)^T) \\
&= \mathrm{E}(\boldsymbol{A}_{t-1}(\boldsymbol{X}_{t-1} - \hat{\boldsymbol{X}}_{t-1})(\boldsymbol{A}_{t-1}(\boldsymbol{X}_{t-1} - \hat{\boldsymbol{X}}_{t-1}))^T) \\
&\qquad + \mathrm{E}((\boldsymbol{B}_{t-1}\boldsymbol{\varepsilon}_t)(\boldsymbol{B}_{t-1}\boldsymbol{\varepsilon}_t)^T) \\
&= \boldsymbol{A}_{t-1}\boldsymbol{\Delta}_{t-1}\boldsymbol{A}_{t-1}^T + \boldsymbol{B}_{t-1}\boldsymbol{Q}_t\boldsymbol{B}_{t-1}^T,
\end{aligned}$$

since $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{X}_{t-1} - \hat{\boldsymbol{X}}_{t-1}$ are obviously uncorrelated. In complete analogy one shows that

$$\mathrm{E}((\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)^T) = \boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T + \boldsymbol{R}_t.$$

Suppose that we have observed $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1}$, and that we have predicted $\boldsymbol{X}_t$ by $\tilde{\boldsymbol{X}}_t = \boldsymbol{A}_{t-1}\hat{\boldsymbol{X}}_{t-1}$. Assume that we now also observe $\boldsymbol{Y}_t$. How can we use this additional information to improve the prediction $\tilde{\boldsymbol{X}}_t$ of $\boldsymbol{X}_t$? To this end we add a matrix $\boldsymbol{K}_t$ such that we obtain the best prediction $\hat{\boldsymbol{X}}_t$ based on $\boldsymbol{Y}_1, \ldots \boldsymbol{Y}_t$:

$$\tilde{\boldsymbol{X}}_t + \boldsymbol{K}_t(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t) = \hat{\boldsymbol{X}}_t \tag{2.24}$$

i.e., we have to choose the matrix $\boldsymbol{K}_t$ according to Lemma 2.4.5 such that $\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t$ and $\boldsymbol{Y}_s$ are uncorrelated for $s = 1, \ldots, t$. In this case, the matrix $\boldsymbol{K}_t$ is called the *Kalman gain*.

Lemma 2.4.6. *The matrix $\boldsymbol{K}_t$ in (2.24) is a solution of the equation*

$$\boldsymbol{K}_t(\boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T + \boldsymbol{R}_t) = \tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T. \tag{2.25}$$

Proof. The matrix $\boldsymbol{K}_t$ has to be chosen such that $\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t$ and $\boldsymbol{Y}_s$ are uncorrelated for $s = 1, \ldots, t$, i.e.,

$$\mathbf{0} = \mathrm{E}((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)\boldsymbol{Y}_s^T) = \mathrm{E}((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t - \boldsymbol{K}_t(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t))\boldsymbol{Y}_s^T), \qquad s \le t.$$

Note that an arbitrary $k \times m$-matrix $\boldsymbol{K}_t$ satisfies

$$\begin{aligned}
&\mathrm{E}\left((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t - \boldsymbol{K}_t(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t))\boldsymbol{Y}_s^T\right) \\
&\quad = \mathrm{E}((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)\boldsymbol{Y}_s^T) - \boldsymbol{K}_t\,\mathrm{E}((\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)\boldsymbol{Y}_s^T) = \mathbf{0}, \qquad s \le t-1.
\end{aligned}$$

In order to fulfill the above condition, the matrix $\boldsymbol{K}_t$ needs to satisfy only

$$\begin{aligned}
\mathbf{0} &= \mathrm{E}((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)\boldsymbol{Y}_t^T) - \boldsymbol{K}_t\,\mathrm{E}((\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)\boldsymbol{Y}_t^T) \\
&= \mathrm{E}((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)^T) - \boldsymbol{K}_t\,\mathrm{E}((\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)^T) \\
&= \mathrm{E}((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)(\boldsymbol{C}_t(\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t) + \boldsymbol{\eta}_t)^T) - \boldsymbol{K}_t\,\mathrm{E}((\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)^T) \\
&= \mathrm{E}((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)(\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)^T)\boldsymbol{C}_t^T - \boldsymbol{K}_t\,\mathrm{E}((\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)^T) \\
&= \tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T - \boldsymbol{K}_t(\boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T + \boldsymbol{R}_t).
\end{aligned}$$

But this is the assertion of Lemma 2.4.6. Note that $\tilde{\boldsymbol{Y}}_t$ is a linear combination of $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1}$ and that $\boldsymbol{\eta}_t$ and $\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t$ as well as $\boldsymbol{\eta}_t$ and $\boldsymbol{Y}_s$ are uncorrelated for $s \le t-1$. □

If the matrix $\boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T + \boldsymbol{R}_t$ is invertible, then

$$\boldsymbol{K}_t := \tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T(\boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T + \boldsymbol{R}_t)^{-1}$$

is the uniquely determined Kalman gain. We have, moreover, for a Kalman gain

$$\begin{aligned}
\boldsymbol{\Delta}_t &= \mathrm{E}((\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)^T) \\
&= \mathrm{E}\left((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t - \boldsymbol{K}_t(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t))(\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t - \boldsymbol{K}_t(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t))^T\right) \\
&= \tilde{\boldsymbol{\Delta}}_t + \boldsymbol{K}_t \mathrm{E}((\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)^T)\boldsymbol{K}_t^T \\
&\qquad - \mathrm{E}((\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)^T)\boldsymbol{K}_t^T - \boldsymbol{K}_t \mathrm{E}((\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t)(\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)^T) \\
&= \tilde{\boldsymbol{\Delta}}_t + \boldsymbol{K}_t(\boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T + \boldsymbol{R}_t)\boldsymbol{K}_t^T \\
&\qquad - \tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T\boldsymbol{K}_t^T - \boldsymbol{K}_t\boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t \\
&= \tilde{\boldsymbol{\Delta}}_t - \boldsymbol{K}_t\boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t
\end{aligned}$$

by (2.25) and the arguments in the proof of Lemma 2.4.6.

The recursion in the discrete *Kalman filter* is done in two steps: From $\hat{\boldsymbol{X}}_{t-1}$ and $\boldsymbol{\Delta}_{t-1}$ one computes in the *prediction step* first

$$\begin{aligned}
\tilde{\boldsymbol{X}}_t &= \boldsymbol{A}_{t-1}\hat{\boldsymbol{X}}_{t-1}, \\
\tilde{\boldsymbol{Y}}_t &= \boldsymbol{C}_t\tilde{\boldsymbol{X}}_t, \\
\tilde{\boldsymbol{\Delta}}_t &= \boldsymbol{A}_{t-1}\boldsymbol{\Delta}_{t-1}\boldsymbol{A}_{t-1}^T + \boldsymbol{B}_{t-1}\boldsymbol{Q}_t\boldsymbol{B}_{t-1}^T.
\end{aligned} \tag{2.26}$$

In the *updating step* one then computes $\boldsymbol{K}_t$ and the updated values $\hat{\boldsymbol{X}}_t$, $\boldsymbol{\Delta}_t$

$$\begin{aligned}
\boldsymbol{K}_t &= \tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T(\boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t\boldsymbol{C}_t^T + \boldsymbol{R}_t)^{-1}, \\
\hat{\boldsymbol{X}}_t &= \tilde{\boldsymbol{X}}_t + \boldsymbol{K}_t(\boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t), \\
\boldsymbol{\Delta}_t &= \tilde{\boldsymbol{\Delta}}_t - \boldsymbol{K}_t\boldsymbol{C}_t\tilde{\boldsymbol{\Delta}}_t.
\end{aligned} \tag{2.27}$$

An obvious problem is the choice of the initial values $\tilde{\boldsymbol{X}}_1$ and $\tilde{\boldsymbol{\Delta}}_1$. One frequently puts $\tilde{\boldsymbol{X}}_1 = \boldsymbol{0}$ and $\tilde{\boldsymbol{\Delta}}_1$ as the diagonal matrix with constant entries $\sigma^2 > 0$. The number σ^2 reflects the degree of uncertainty about the underlying model. Simulations as well as theoretical results show, however, that the estimates $\hat{\boldsymbol{X}}_t$ are often not affected by the initial values $\tilde{\boldsymbol{X}}_1$ and $\tilde{\boldsymbol{\Delta}}_1$ if t is large, see for instance Example 2.4.7 below.

If in addition we require that the state-space model (2.18), (2.19) is completely determined by some parametrization $\boldsymbol{\vartheta}$ of the distribution of $(\boldsymbol{Y}_t)$ and $(\boldsymbol{X}_t)$, then we can estimate the matrices of the Kalman filter in (2.26) and (2.27) under suitable conditions by a maximum likelihood estimate of $\boldsymbol{\vartheta}$; see e.g. Section 8.5 of Brockwell and Davis (2002) or Section 4.5 in Janacek and Swift (1993).

By iterating the 1-step prediction $\tilde{\boldsymbol{X}}_t = \boldsymbol{A}_{t-1}\hat{\boldsymbol{X}}_{t-1}$ of $\hat{\boldsymbol{X}}_t$ in (2.23) h times, we obtain the h-step prediction of the Kalman filter

$$\tilde{\boldsymbol{X}}_{t+h} := \boldsymbol{A}_{t+h-1}\tilde{\boldsymbol{X}}_{t+h-1}, \qquad h \geq 1,$$

with the initial value $\tilde{\boldsymbol{X}}_{t+0} := \hat{\boldsymbol{X}}_t$. The pertaining h-step prediction of $\boldsymbol{Y}_{t+h}$ is then

$$\tilde{\boldsymbol{Y}}_{t+h} := \boldsymbol{C}_{t+h}\tilde{\boldsymbol{X}}_{t+h}, \qquad h \geq 1.$$

2.4.7 Example. Let (η_t) be a white noise in $\mathbb{R}$ with $\mathrm{E}(\eta_t) = 0$, $\mathrm{E}(\eta_t^2) = \sigma^2 > 0$ and put for some $\mu \in \mathbb{R}$

$$Y_t := \mu + \eta_t, \qquad t \in \mathbb{Z}.$$

This process can be represented as a state-space model by putting $X_t := \mu$, with state equation $X_{t+1} = X_t$ and observation equation $Y_t = X_t + \eta_t$ i.e., $A_t = 1 = C_t$ and $B_t = 0$. The prediction step (2.26) of the Kalman filter is now given by

$$\tilde{X}_t = \hat{X}_{t-1},\ \tilde{Y}_t = \tilde{X}_t,\ \tilde{\Delta}_t = \Delta_{t-1}.$$

Note that all these values are in $\mathbb{R}$. The h-step predictions $\tilde{X}_{t+h}$, $\tilde{Y}_{t+h}$ are, therefore, given by $\tilde{X}_t$. The update step (2.27) of the Kalman filter is

$$\begin{aligned} K_t &= \frac{\Delta_{t-1}}{\Delta_{t-1} + \sigma^2} \\ \hat{X}_t &= \hat{X}_{t-1} + K_t(Y_t - \hat{X}_{t-1}) \\ \Delta_t &= \Delta_{t-1} - K_t\Delta_{t-1} = \Delta_{t-1}\frac{\sigma^2}{\Delta_{t-1} + \sigma^2}. \end{aligned}$$

Note that $\Delta_t = \mathrm{E}((X_t - \hat{X}_t)^2) \geq 0$ and thus,

$$0 \leq \Delta_t = \Delta_{t-1}\frac{\sigma^2}{\Delta_{t-1} + \sigma^2} \leq \Delta_{t-1}$$

is a decreasing and bounded sequence. Its limit $\Delta := \lim_{t\to\infty} \Delta_t$ consequently exists and satisfies

$$\Delta = \Delta\frac{\sigma^2}{\Delta + \sigma^2}$$

i.e., $\Delta = 0$. This means that the mean squared error $\mathrm{E}((X_t - \hat{X}_t)^2) = \mathrm{E}((\mu - \hat{X}_t)^2)$ vanishes asymptotically, no matter how the initial values $\tilde{X}_1$ and $\tilde{\Delta}_1$ are chosen. Further we have $\lim_{t\to\infty} K_t = 0$, which means that additional observations Y_t do not contribute to $\hat{X}_t$ if t is large. Finally, we obtain for the mean squared error of the h-step prediction $\tilde{Y}_{t+h}$ of Y_{t+h}

$$\begin{aligned} \mathrm{E}((Y_{t+h} - \tilde{Y}_{t+h})^2) &= \mathrm{E}((\mu + \eta_{t+h} - \hat{X}_t)^2) \\ &= \mathrm{E}((\mu - \hat{X}_t)^2) + \mathrm{E}(\eta_{t+h}^2) \to_{t\to\infty} \sigma^2. \end{aligned}$$

Example 2.4.7. The following figure displays the Airline Data from Example 1.3.1 together with 12-step forecasts based on the Kalman filter. The original

data y_t, $t = 1, \ldots, 144$ were log-transformed $x_t = \log(y_t)$ to stabilize the variance; first order differences $\Delta x_t = x_t - x_{t-1}$ were used to eliminate the trend and, finally, $z_t = \Delta x_t - \Delta x_{t-12}$ were computed to remove the seasonal component of 12 months. The Kalman filter was applied to forecast z_t, $t = 145, \ldots, 156$, and the results were transformed in the reverse order of the preceding steps to predict the initial values y_t, $t = 145, \ldots, 156$.

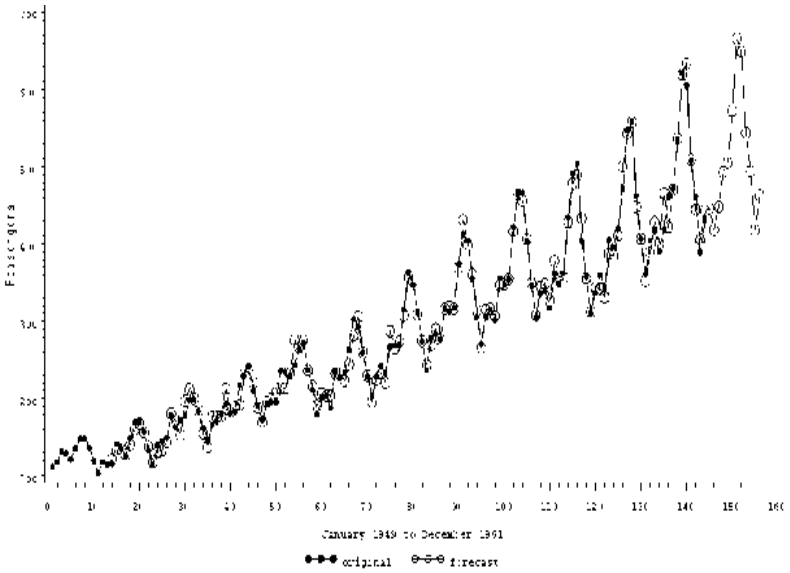


Figure 2.4.1. Airline Data and predicted values using the Kalman filter.

```
***     Program 2_4_1    ***;
TITLE1 'Original and Forecasted Data';
TITLE2 'Airline Data';

DATA data1;
  INFILE 'c:\data\airline.txt';
  INPUT y;
  yl=LOG(y); t=_N_;

PROC STATESPACE DATA=data1 OUT=data2 LEAD=12;
 VAR yl(1,12); ID t;

DATA data3;
```

```
   SET data2; yhat=EXP(FOR1);

DATA data4(KEEP=t y yhat);
   MERGE data1 data3;
   BY t;

LEGEND1 LABEL=('') VALUE=('original' 'forecast');
SYMBOL1 C=BLACK V=DOT H=0.7 I=JOIN L=1;
SYMBOL2 C=BLACK V=CIRCLE H=1.5 I=JOIN L=1;
AXIS1 LABEL=(ANGLE=90 'Passengers');
AXIS2 LABEL=('January 1949 to December 1961');
PROC GPLOT DATA=data4;
 PLOT y*t=1 yhat*t=2 / OVERLAY VAXIS=AXIS1
        HAXIS=AXIS2 LEGEND=LEGEND1;
RUN; QUIT;
```

In the first data step the Airline Data are read into `data1`. Their logarithm is computed and stored in the variable `yl`. The variable `t` contains the observation number.

The statement `VAR yl(1,12)` of the `PROC STATESPACE` procedure tells SAS to use first order differences of the initial data to remove their trend and to adjust them to a seasonal component of 12 months. The data are identified by the time index set to `t`. The results are stored in the data set `data2` with forecasts of 12 months after the end of the input data. This is invoked by `LEAD=12`.

`data3` contains the exponentially transformed forecasts, thereby inverting the log-transformation in the first data step.

Finally, the two data sets are merged and displayed in one plot.

Exercises

1. Suppose that the complex random variables Y and Z are square integrable. Show that

$$\mathrm{Cov}(aY + b, Z) = a\,\mathrm{Cov}(Y, Z), \quad a, b \in \mathbb{C}.$$

2. Give an example of a stochastic process (Y_t) such that for arbitrary $t_1, t_2 \in \mathbb{Z}$ and $k \neq 0$

$$\mathrm{E}(Y_{t_1}) \neq \mathrm{E}(Y_{t_1+k}) \quad \text{but} \quad \mathrm{Cov}(Y_{t_1}, Y_{t_2}) = \mathrm{Cov}(Y_{t_1+k}, Y_{t_2+k}).$$

3. (i) Let $(X_t), (Y_t)$ be stationary processes such that $\mathrm{Cov}(X_t, Y_s) = 0$ for $t, s \in \mathbb{Z}$. Show that for arbitrary $a, b \in \mathbb{C}$ the linear combinations $(aX_t + bY_t)$ yield a stationary process.

(ii) Suppose that the decomposition $Z_t = X_t + Y_t$, $t \in \mathbb{Z}$ holds. Show that stationarity of (Z_t) does not necessarily imply stationarity of (X_t).

4. (i) Show that the process $Y_t = Xe^{iat}$, $a \in \mathbb{R}$, is stationary, where X is a complex valued random variable with mean zero and finite variance.

(ii) Show that the random variable $Y = be^{iU}$ has mean zero, where U is a uniformly distributed random variable on $(0, 2\pi)$ and $b \in \mathbb{C}$.

5. Let Z_1, Z_2 be independent and normal $N(\mu_i, \sigma_i^2)$, $i = 1, 2$, distributed random variables and choose $\lambda \in \mathbb{R}$. For which means $\mu_1, \mu_2 \in \mathbb{R}$ and variances $\sigma_1^2, \sigma_2^2 > 0$ is the *cosinoid process*

$$Y_t = Z_1 \cos(2\pi\lambda t) + Z_2 \sin(2\pi\lambda t), \quad t \in \mathbb{Z}$$

stationary?

6. Show that the autocovariance function $\gamma : \mathbb{Z} \to \mathbb{C}$ of a complex-valued stationary process $(Y_t)_{t\in\mathbb{Z}}$, which is defined by

$$\gamma(h) = \mathrm{E}(Y_{t+h}\bar{Y}_t) - \mathrm{E}(Y_{t+h})\,\mathrm{E}(\bar{Y}_t), \quad h \in \mathbb{Z},$$

has the following properties: $\gamma(0) \geq 0$, $|\gamma(h)| \leq \gamma(0)$, $\gamma(h) = \overline{\gamma(-h)}$, i.e., γ is a Hermitian function, and $\sum_{1\leq r,s\leq n} z_r\gamma(r-s)\bar{z}_s \geq 0$ for $z_1, \ldots, z_n \in \mathbb{C}$, $n \in \mathbb{N}$, i.e., γ is a positive semidefinite function.

7. Suppose that Y_t, $t = 1, \ldots, n$, is a stationary process with mean μ. Then $\hat{\mu}_n := n^{-1}\sum_{t=1}^n Y_t$ is an unbiased estimator of μ. Express the mean square error $\mathrm{E}(\hat{\mu}_n - \mu)^2$ in terms of the autocovariance function γ and show that $\mathrm{E}(\hat{\mu}_n - \mu)^2 \to 0$ if $\gamma(n) \to 0$, $n \to \infty$.

8. Suppose that $(Y_t)_{t\in\mathbb{Z}}$ is a stationary process and denote by

$$c(k) := \begin{cases} \frac{1}{n}\sum_{t=1}^{n-|k|}(Y_t - \bar{Y})(Y_{t+|k|} - \bar{Y}), & |k| = 0, \ldots, n-1, \\ 0, & |k| \geq n. \end{cases}$$

the empirical autocovariance function at lag k, $k \in \mathbb{Z}$.

(i) Show that $c(k)$ is a biased estimator of $\gamma(k)$ (even if the factor n^{-1} is replaced by $(n-k)^{-1}$) i.e., $\mathrm{E}(c(k)) \neq \gamma(k)$.

(ii) Show that the k-dimensional empirical covariance matrix

$$\boldsymbol{C}_k := \begin{pmatrix} c(0) & c(1) & \ldots & c(k-1) \\ c(1) & c(0) & & c(k-2) \\ \vdots & & \ddots & \vdots \\ c(k-1) & c(k-2) & \ldots & c(0) \end{pmatrix}$$

is positive semidefinite. (If the factor n^{-1} in the definition of $c(j)$ is replaced by $(n-j)^{-1}$, $j = 1, \ldots, k$, the resulting covariance matrix may not be positive

semidefinite.) Hint: Consider $k \geq n$ and write $\boldsymbol{C}_k = n^{-1}\boldsymbol{A}\boldsymbol{A}^T$ with a suitable $k \times 2k$-matrix $\boldsymbol{A}$. Show further that $\boldsymbol{C}_m$ is positive semidefinite if $\boldsymbol{C}_k$ is positive semidefinite for $k > m$.

(iii) If $c(0) > 0$, then $\boldsymbol{C}_k$ is nonsingular, i.e., $\boldsymbol{C}_k$ is positive definite.

9. Suppose that (Y_t) is a stationary process with autocovariance function γ_Y. Express the autocovariance function of the difference filter of first order $\Delta Y_t = Y_t - Y_{t-1}$ in terms of γ_Y. Find it when $\gamma_Y(k) = \lambda^{|k|}$.

10. Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary process with mean zero. If its autocovariance function satisfies $\gamma(\tau) = 0$ for some $\tau > 0$, then γ is periodic with length τ, i.e., $\gamma(t+\tau) = \gamma(t)$, $t \in \mathbb{Z}$.

11. Let (Y_t) be a stochastic process such that for $t \in \mathbb{Z}$

$$P\{Y_t = 1\} = p_t = 1 - P\{Y_t = -1\}, \quad 0 < p_t < 1.$$

Suppose in addition that (Y_t) is a *Markov process*, i.e., for any $t \in \mathbb{Z}$, $k \geq 1$

$$P(Y_t = y_0 | Y_{t-1} = y_1, \ldots, Y_{t-k} = y_k) = P(Y_t = y_0 | Y_{t-1} = y_1).$$

(i) Is $(Y_t)_{t\in\mathbb{N}}$ a stationary process?

(ii) Compute the autocovariance function in case $P(Y_t = 1 | Y_{t-1} = 1) = \lambda$ and $p_t = 1/2$.

12. Let $(\varepsilon_t)_t$ be a white noise process with independent $\varepsilon_t \sim N(0,1)$ and define

$$\tilde{\varepsilon}_t = \begin{cases} \varepsilon_t, & \text{if } t \text{ is even,} \\ (\varepsilon_{t-1}^2 - 1)/\sqrt{2}, & \text{if } t \text{ is odd.} \end{cases}$$

Show that $(\tilde{\varepsilon}_t)_t$ is a white noise process with $\mathrm{E}(\tilde{\varepsilon}_t) = 0$ and $\mathrm{Var}(\tilde{\varepsilon}_t) = 1$, where the $\tilde{\varepsilon}_t$ are neither independent nor identically distributed. Plot the path of $(\varepsilon_t)_t$ and $(\tilde{\varepsilon}_t)_t$ for $t = 1, \ldots, 100$ and compare!

13. Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise. The process $Y_t = \sum_{s=1}^{t} \varepsilon_s$ is said to be a *random walk*. Plot the path of a random walk with normal $N(\mu, \sigma^2)$ distributed ε_t for each of the cases $\mu < 0, \mu = 0$ and $\mu > 0$.

14. Let (a_u) be absolutely summable filters and let (Z_t) be a stochastic process with $\sup_{t\in\mathbb{Z}} \mathrm{E}(Z_t^2) < \infty$. Put for $t \in \mathbb{Z}$

$$X_t = \sum_u a_u Z_{t-u}, \quad Y_t = \sum_v b_v Z_{t-v}.$$

Then we have

$$\mathrm{E}(X_t Y_t) = \sum_u \sum_v a_u b_v \, \mathrm{E}(Z_{t-u} Z_{t-v}).$$

Hint: Use the general inequality $|xy| \leq (x^2 + y^2)/2$.

15. Let $Y_t = aY_{t-1} + \varepsilon_t$, $t \in \mathbb{Z}$ be an $AR(1)$-process with $|a| > 1$. Compute the autocorrelation function of this process.

16. Compute the orders p and the coefficients a_u of the process $Y_t = \sum_{u=0}^{p} a_u \varepsilon_{t-u}$ with $\mathrm{Var}(\varepsilon_0) = 1$ and autocovariance function $\gamma(1) = 2, \gamma(2) = 1, \gamma(3) = -1$ and $\gamma(t) = 0$ for $t \geq 4$. Is this process invertible?

17. The autocorrelation function ρ of an arbitrary $MA(q)$-process satisfies

$$-\frac{1}{2} \leq \sum_{v=1}^{q} \rho(v) \leq \frac{1}{2} q.$$

Give examples of $MA(q)$-processes, where the lower bound and the upper bound are attained, i.e., these bounds are sharp.

18. Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary stochastic process with $\mathrm{E}(Y_t) = 0$, $t \in \mathbb{Z}$, and

$$\rho(t) = \begin{cases} 1 & \text{if } t = 0 \\ \rho(1) & \text{if } t = 1 \\ 0 & \text{if } t > 1, \end{cases}$$

where $|\rho(1)| < 1/2$. Then there exists $a \in (-1, 1)$ and a white noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ such that

$$Y_t = \varepsilon_t + a\varepsilon_{t-1}.$$

Hint: Example 2.2.2.

19. Find two $MA(1)$-processes with the same autocovariance functions.

20. Suppose that $Y_t = \varepsilon_t + a\varepsilon_{t-1}$ is a noninvertible $MA(1)$-process, where $|a| > 1$. Define the new process

$$\tilde{\varepsilon}_t = \sum_{j=0}^{\infty} (-a)^{-j} Y_{t-j}$$

and show that $(\tilde{\varepsilon}_t)$ is a white noise. Show that $\mathrm{Var}(\tilde{\varepsilon}_t) = a^2 \mathrm{Var}(\varepsilon_t)$ and (Y_t) has the *invertible* representation

$$Y_t = \tilde{\varepsilon}_t + a^{-1}\tilde{\varepsilon}_{t-1}.$$

21. Plot the autocorrelation functions of $MA(p)$-processes for different values of p.

22. Generate and plot $AR(3)$-processes (Y_t), $t = 1, \ldots, 500$ where the roots of the characteristic polynomial have the following properties:

(i) all roots are outside the unit disk,

(ii) all roots are inside the unit disk,

(iii) all roots are on the unit circle,

(iv) two roots are outside, one root inside the unit disk,

(v) one root is outside, one root is inside the unit disk and one root is on the unit circle,

(vi) all roots are outside the unit disk but close to the unit circle.

23. Show that the $AR(2)$-process $Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t$ for $a_1 = 1/3$ and $a_2 = 2/9$ has the autocorrelation function

$$\rho(k) = \frac{16}{21}\Big(\frac{2}{3}\Big)^{|k|} + \frac{5}{21}\Big(-\frac{1}{3}\Big)^{|k|}, \quad k \in \mathbb{Z}$$

and for $a_1 = a_2 = 1/12$ the autocorrelation function

$$\rho(k) = \frac{45}{77}\Big(\frac{1}{3}\Big)^{|k|} + \frac{32}{77}\Big(-\frac{1}{4}\Big)^{|k|}, \quad k \in \mathbb{Z}.$$

24. Let (ε_t) be a white noise with $\mathrm{E}(\varepsilon_0) = \mu, \mathrm{Var}(\varepsilon_0) = \sigma^2$ and put

$$Y_t = \varepsilon_t - Y_{t-1}, \quad t \in \mathbb{N}, Y_0 = 0.$$

Show that

$$\mathrm{Corr}(Y_s, Y_t) = (-1)^{s+t} \min\{s,t\}/\sqrt{st}.$$

25. An $AR(2)$-process $Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t$ satisfies the stationarity condition (2.3), if the pair (a_1, a_2) is in the triangle

$$\Delta := \Big\{(\alpha,\beta) \in \mathbb{R}^2 : -1 < \beta < 1, \alpha + \beta < 1 \text{ and } \beta - \alpha < 1\Big\}.$$

Hint: Use the fact that necessarily $\rho(1) \in (-1, 1)$.

26. (i) Let (Y_t) denote the unique stationary solution of the autoregressive equations

$$Y_t = aY_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z},$$

with $|a| > 1$. Then (Y_t) is given by the expression $Y_t = -\sum_{j=1}^{\infty} a^{-j}\varepsilon_{t+j}$ (see the proof of Lemma 2.1.10). Define the new process

$$\tilde{\varepsilon}_t = Y_t - \frac{1}{a}Y_{t-1},$$

and show that $(\tilde{\varepsilon}_t)$ is a white noise with $\mathrm{Var}(\tilde{\varepsilon}_t) = \mathrm{Var}(\varepsilon_t)/a^2$. These calculations show that (Y_t) is the (unique stationary) solution of the causal AR-equations

$$Y_t = \frac{1}{a}Y_{t-1} + \tilde{\varepsilon}_t, \quad t \in \mathbb{Z}.$$

Thus, every $AR(1)$-process with $|a| > 1$ can be represented as an $AR(1)$-process with $|a| < 1$ and a new white noise.

(ii) Show that for $|a| = 1$ the above autoregressive equations have *no stationary* solutions. A stationary solution exists if the white noise process is degenerated, i.e., $\mathrm{E}(\varepsilon_t^2) = 0$.

27. (i) Consider the process

$$\tilde{Y}_t := \begin{cases} \varepsilon_1 & \text{for } t = 1 \\ aY_{t-1} + \varepsilon_t & \text{for } t > 1, \end{cases}$$

i.e., $\tilde{Y}_t, t \geq 1$, equals the $AR(1)$-process $Y_t = aY_{t-1} + \varepsilon_t$, conditional on $Y_0 = 0$. Compute $\mathrm{E}(\tilde{Y}_t), \mathrm{Var}(\tilde{Y}_t)$ and $\mathrm{Cov}(Y_t, Y_{t+s})$. Is there something like asymptotic stationarity for $t \to \infty$?

(ii) Choose $a \in (-1, 1)$, $a \neq 0$, and compute the correlation matrix of $Y_1, \ldots, Y_{10}$.

28. Use the `IML` function `ARMASIM` to simulate the stationary $AR(2)$-process

$$Y_t = -0.3Y_{t-1} + 0.3Y_{t-2} + \varepsilon_t.$$

Estimate the parameters $a_1 = -0.3$ and $a_2 = 0.3$ by means of the Yule–Walker equations using the SAS procedure `PROC ARIMA`.

29. Show that the value at lag 2 of the partial autocorrelation function of the $MA(1)$-process

$$Y_t = \varepsilon_t + a\varepsilon_{t-1}, \quad t \in \mathbb{Z}$$

is

$$\alpha(2) = -\frac{a^2}{1 + a^2 + a^4}.$$

30. (Unemployed1 Data) Plot the empirical autocorrelations and partial autocorrelations of the trend and seasonally adjusted Unemployed1 Data from the building trade, introduced in Example 1.1.1. Apply the Box–Jenkins program. Is a fit of a pure $MA(q)$- or $AR(p)$-process reasonable?

31. Plot the autocorrelation functions of $ARMA(p, q)$-processes for different values of p, q using the `IML` function `ARMACOV`. Plot also their empirical counterparts.

32. Compute the autocovariance function of an $ARMA(1, 2)$-process.

33. Derive the least squares normal equations for an $AR(p)$-process and compare them with the Yule–Walker equations.

34. Show that the density of the t-distribution with m degrees of freedom converges to the density of the standard normal distribution as m tends to infinity. Hint: Apply the dominated convergence theorem (Lebesgue).

35. Let $(Y_t)_t$ be a stationary and causal $ARCH(1)$-process with $|a_1| < 1$.

(i) Show that $Y_t^2 = a_0 \sum_{j=0}^{\infty} a_1^j Z_t^2 Z_{t-1}^2 \cdot \cdots \cdot Z_{t-j}^2$ with probability one.

(ii) Show that $\mathrm{E}(Y_t^2) = a_0/(1-a_1)$.

(iii) Evaluate $\mathrm{E}(Y_t^4)$ and deduce that $\mathrm{E}(Z_1^4)a_1^2 < 1$ is a sufficient condition for $\mathrm{E}(Y_t^4) < \infty$.

Hint: Theorem 2.1.5.

36. Determine the joint density of $Y_{p+1}, \ldots, Y_n$ for an $ARCH(p)$-process Y_t with normal distributed Z_t given that $Y_1 = y_1, \ldots, Y_p = y_p$. Hint: Recall that the joint density $f_{X,Y}$ of a random vector (X, Y) can be written in the form $f_{X,Y}(x, y) = f_{Y|X}(y|x) f_X(x)$, where $f_{Y|X}(y|x) := f_{X,Y}(x, y)/f_X(x)$ if $f_X(x) > 0$, and $f_{Y|X}(y|x) := f_Y(y)$, else, is the (conditional) density of Y given $X = x$ and f_X, f_Y is the (marginal) density of X, Y.

37. Generate an $ARCH(1)$-process $(Y_t)_t$ with $a_0 = 1$ and $a_1 = 0.5$. Plot $(Y_t)_t$ as well as $(Y_t^2)_t$ and its partial autocorrelation function. What is the value of the partial autocorrelation coefficient at lag 1 and lag 2? Use `PROC ARIMA` to estimate the parameter of the $AR(1)$-process $(Y_t^2)_t$ and apply the Box–Ljung test.

38. (Hong Kong Data) Fit an $GARCH(p, q)$-model to the daily Hang Seng closing index of Hong Kong stock prices from July 16, 1981, to September 31, 1983. Consider in particular the cases $p = q = 2$ and $p = 3, q = 2$.

39. (Zurich Data) The daily value of the Zurich stock index was recorded between January 1st, 1988 and December 31st, 1988. Use a difference filter of first order to remove a possible trend. Plot the (trend-adjusted) data, their squares, the pertaining partial autocorrelation function and parameter estimates. Can the squared process be considered as an $AR(1)$-process?

40. (i) Show that the matrix $\boldsymbol{\Sigma'}^{-1}$ in Example 2.3.1 has the determinant $1 - a^2$.

(ii) Show that the matrix $\boldsymbol{P}_n$ in Example 2.3.3 has the determinant $(1 + a^2 + a^4 + \cdots + a^{2n})/(1 + a^2)^n$.

41. (Car Data) Apply the Box–Jenkins program to the Car Data.

42. Consider the two state-space models

$$\begin{aligned} \boldsymbol{X}_{t+1} &= \boldsymbol{A}_t \boldsymbol{X}_t + \boldsymbol{B}_t \boldsymbol{\varepsilon}_{t+1} \\ \boldsymbol{Y}_t &= \boldsymbol{C}_t \boldsymbol{X}_t + \boldsymbol{\eta}_t \end{aligned}$$

and

$$\begin{aligned} \tilde{\boldsymbol{X}}_{t+1} &= \tilde{\boldsymbol{A}}_t \tilde{\boldsymbol{X}}_t + \tilde{\boldsymbol{B}}_t \tilde{\boldsymbol{\varepsilon}}_{t+1} \\ \tilde{\boldsymbol{Y}}_t &= \tilde{\boldsymbol{C}}_t \tilde{\boldsymbol{X}}_t + \tilde{\boldsymbol{\eta}}_t, \end{aligned}$$

where $(\boldsymbol{\varepsilon}_t^T, \boldsymbol{\eta}_t^T, \tilde{\boldsymbol{\varepsilon}}_t^T, \tilde{\boldsymbol{\eta}}_t^T)^T$ is a white noise. Derive a state-space representation for $(\boldsymbol{Y}_t^T, \tilde{\boldsymbol{Y}}_t^T)^T$.

43. Find the state-space representation of an $ARIMA(p,d,q)$-process $(Y_t)_t$. Hint: $Y_t = \Delta^d Y_t - \sum_{j=1}^{d}(-1)^j \binom{d}{d} Y_{t-j}$ and consider the state vector $\boldsymbol{Z}_t := (\boldsymbol{X}_t, \boldsymbol{Y}_{t-1})^T$, where $\boldsymbol{X}_t \in \mathbb{R}^{p+q}$ is the state vector of the $ARMA(p,q)$-process $\Delta^d Y_t$ and $\boldsymbol{Y}_{t-1} := (Y_{t-d}, \ldots, Y_{t-1})^T$.

44. Assume that the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ in the state-space model (2.18) are independent of t and that all eigenvalues of $\boldsymbol{A}$ are in the interior of the unit circle $\{z \in \mathbb{C} : |z| \leq 1\}$. Show that the unique stationary solution of equation (2.18) is given by the infinite series $\boldsymbol{X}_t = \sum_{j=0}^{\infty} \boldsymbol{A}^j \boldsymbol{B} \boldsymbol{\varepsilon}_{t-j+1}$. Hint: The condition on the eigenvalues is equivalent to $\det(I_r - \boldsymbol{A}z) \neq 0$ for $|z| \leq 1$. Show that there exists some $\varepsilon > 0$ such that $(I_r - \boldsymbol{A}z)^{-1}$ has the power series representation $\sum_{j=0}^{\infty} \boldsymbol{A}^j z^j$ in the region $|z| < 1 + \varepsilon$.

45. Apply `PROC STATESPACE` to the simulated data of the $AR(2)$-process in Exercise 26.

46. (Gas Data) Apply `PROC STATESPACE` to the gas data. Can they be stationary? Compute the one-step predictors and plot them together with the actual data.

Chapter 3

The Frequency Domain Approach of a Time Series

The preceding sections focussed on the analysis of a time series in the *time domain*, mainly by modelling and fitting an $ARMA(p,q)$-process to stationary sequences of observations. Another approach towards the modelling and analysis of time series is via the *frequency domain*: A series is often the sum of a whole variety of cyclic components, from which we had already added to our model (1.2) a long term cyclic one or a short term seasonal one. In the following we show that a time series can be completely decomposed into cyclic components. Such cyclic components can be described by their *periods* and *frequencies*. The period is the interval of time required for one cycle to complete. The frequency of a cycle is its number of occurrences during a fixed time unit; in electronic media, for example, frequencies are commonly measured in *hertz*, which is the number of cycles per second, abbreviated by *Hz*. The analysis of a time series in the frequency domain aims at the detection of such cycles and the computation of their frequencies.

Note that in this chapter the results are formulated for any data $y_1, \ldots, y_n$, which need for mathematical reasons not to be generated by a stationary process. Nevertheless it is reasonable to apply the results only to realizations of stationary processes, since the empirical autocovariance function occurring below has no interpretation for non-stationary processes, see Exercise 18 in Chapter 1.

3.1 Least Squares Approach with Known Frequencies

A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is said to be *periodic* with *period* $P > 0$ if $f(t + P) = f(t)$ for any $t \in \mathbb{R}$. A smallest period is called a *fundamental* one. The reciprocal value $\lambda = 1/P$ of a fundamental period is the fundamental frequency. An arbitrary (time) interval of length L consequently shows $L\lambda$ cycles of a periodic function f with fundamental frequency λ. Popular examples of periodic functions are sine and cosine, which both have the fundamental period $P = 2\pi$. Their fundamental frequency, therefore, is $\lambda = 1/(2\pi)$. The predominant family of periodic functions within time series analysis are the harmonic components

$$m(t) := A\cos(2\pi\lambda t) + B\sin(2\pi\lambda t), \qquad A, B \in \mathbb{R},\ \lambda > 0,$$

which have period $1/\lambda$ and frequency λ. A linear combination of harmonic components

$$g(t) := \mu + \sum_{k=1}^{r} \big(A_k \cos(2\pi\lambda_k t) + B_k \sin(2\pi\lambda_k t)\big), \qquad \mu \in \mathbb{R},$$

will be named a *harmonic wave of length* r.

Example 3.1.1. (Star Data). To analyze physical properties of a pulsating star, the intensity of light emitted by this pulsar was recorded at midnight during 600 consecutive nights. The data are taken from Newton (1988). It turns out that a harmonic wave of length two fits the data quite well. The following figure displays the first 160 data y_t and the sum of two harmonic components with period 24 and 29, respectively, plus a constant term $\mu = 17.07$ fitted to these data, i.e.,

$$\begin{aligned}\tilde{y}_t = 17.07 &- 1.86\cos(2\pi(1/24)t) + 6.82\sin(2\pi(1/24)t) \\ &+ 6.09\cos(2\pi(1/29)t) + 8.01\sin(2\pi(1/29)t).\end{aligned}$$

The derivation of these particular frequencies and coefficients will be the content of this section and the following ones. For easier access we begin with the case of known frequencies but unknown constants.

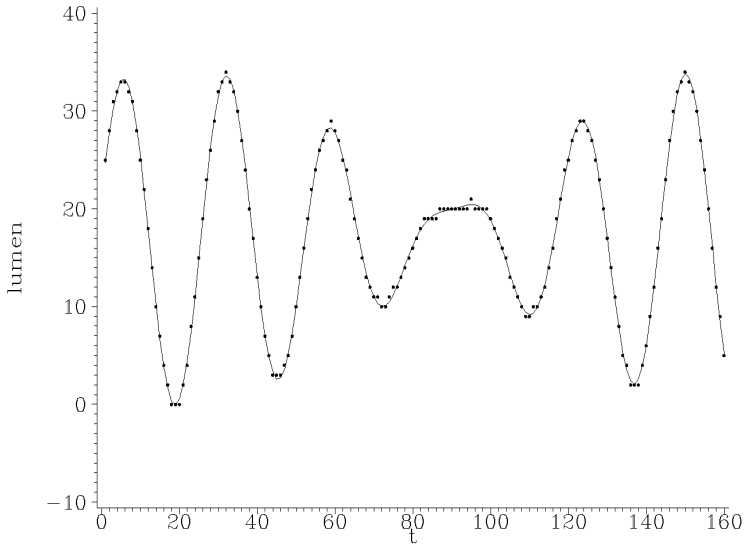


Figure 3.1.1. Intensity of light emitted by a pulsating star and a fitted harmonic wave.

```
Model: MODEL1
Dependent Variable: LUMEN

                        Analysis of Variance

                          Sum of          Mean
Source           DF      Squares        Square      F Value       Prob>F

Model             4        48400         12100      49297.2       <.0001
Error           595    146.04384       0.24545
C Total         599        48546

    Root MSE       0.49543     R-square       0.9970
    Dep Mean      17.09667     Adj R-sq       0.9970
    C.V.           2.89782

                        Parameter Estimates

                  Parameter      Standard
```

Variable	DF	Estimate	Error	t Value	Prob > |T|
Intercept	1	17.06903	0.02023	843.78	<.0001
sin24	1	6.81736	0.02867	237.81	<.0001
cos24	1	-1.85779	0.02865	-64.85	<.0001
sin29	1	8.01416	0.02868	279.47	<.0001
cos29	1	6.08905	0.02865	212.57	<.0001

```
***     Program 3_1_1   ***;
TITLE1 'Harmonic wave';
TITLE2 'Star Data';

DATA data1;
  INFILE 'c:\data\star.txt';
  INPUT lumen @@;
  t=_N_;
  pi=CONSTANT('PI');
  sin24=SIN(2*pi*t/24);
  cos24=COS(2*pi*t/24);
  sin29=SIN(2*pi*t/29);
  cos29=COS(2*pi*t/29);

PROC REG DATA=data1;
  MODEL lumen=sin24 cos24 sin29 cos29;
  OUTPUT OUT=regdat P=predi;

SYMBOL1 C=GREEN V=DOT I=NONE H=.4;
SYMBOL2 C=RED V=NONE I=JOIN;
AXIS1 LABEL=(ANGLE=90 'lumen');
AXIS2 LABEL=('t');
PROC GPLOT DATA=regdat(OBS=160);
  PLOT lumen*t=1 predi*t=2 / OVERLAY VAXIS=AXIS1
                                     HAXIS=AXIS2;
RUN; QUIT;
```

The number π is generated by the SAS function `CONSTANT` with the argument `'PI'`. It is then stored in the variable `pi`. This is used to define the variables `cos24`, `sin24`, `cos29` and `sin29` for the harmonic components. The other variables here are `lumen` read from an external file and `t` generated by `_N_`.

The `PROC REG` statement causes SAS to make a regression from the independent variable `lumen` defined on the left side of the `MODEL` statement on the harmonic components which are on the right side. A temporary data file named `regdat` is generated by the `OUTPUT` statement. It contains the original variables of the source data step and the values predicted by the regression for `lumen` in the variable `predi`.

The last part of the program creates a plot of the observed `lumen` values and a curve of the predicted values restricted on the first 160 observations.

The output of Program 3_1_1 is the standard text output of a regression with an ANOVA table and parameter estimates. For further information on how to read the output, we refer to Chapter 3 of Falk et al. (2002).

In a first step we will fit a harmonic component with fixed frequency λ to mean value adjusted data $y_t - \bar{y},\ t = 1, \ldots, n$. To this end, we put with arbitrary A, $B \in \mathbb{R}$

$$m(t) = Am_1(t) + Bm_2(t),$$

where

$$m_1(t) := \cos(2\pi\lambda t), \quad m_2(t) = \sin(2\pi\lambda t).$$

In order to get a proper fit uniformly over all t, it is reasonable to choose the constants A and B as minimizers of the *residual sum of squares*

$$R(A, B) := \sum_{t=1}^{n} (y_t - \bar{y} - m(t))^2.$$

Taking partial derivatives of the function R with respect to A and B and equating them to zero, we obtain that the minimizing pair A, B has to satisfy the *normal equations*

$$\begin{aligned} Ac_{11} + Bc_{12} &= \sum_{t=1}^{n} (y_t - \bar{y}) \cos(2\pi\lambda t) \\ Ac_{21} + Bc_{22} &= \sum_{t=1}^{n} (y_t - \bar{y}) \sin(2\pi\lambda t), \end{aligned}$$

where

$$c_{ij} = \sum_{t=1}^{n} m_i(t) m_j(t).$$

If $c_{11}c_{22} - c_{12}c_{21} \neq 0$, the uniquely determined pair of solutions A, B of these equations is

$$\begin{aligned} A = A(\lambda) &= n \frac{c_{22}C(\lambda) - c_{12}S(\lambda)}{c_{11}c_{22} - c_{12}c_{21}} \\ B = B(\lambda) &= n \frac{c_{21}C(\lambda) - c_{11}S(\lambda)}{c_{12}c_{21} - c_{11}c_{22}}, \end{aligned}$$

where

$$C(\lambda) := \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})\cos(2\pi\lambda t),$$
$$S(\lambda) := \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})\sin(2\pi\lambda t) \tag{3.1}$$

are the *empirical (cross-) covariances* of $(y_t)_{1\le t\le n}$ and $(\cos(2\pi\lambda t))_{1\le t\le n}$ and of $(y_t)_{1\le t\le n}$ and $(\sin(2\pi\lambda t))_{1\le t\le n}$, respectively. As we will see, these cross-covariances $C(\lambda)$ and $S(\lambda)$ are fundamental to the analysis of a time series in the frequency domain.

The solutions A and B become particularly simple in the case of *Fourier frequencies* $\lambda = k/n,\ k = 0, 1, 2, \ldots, [n/2]$, where $[x]$ denotes the greatest integer less than or equal to $x \in \mathbb{R}$. If $k \neq 0$ and $k \neq n/2$ in the case of an even sample size n, then we obtain from (3.2) below that $c_{12} = c_{21} = 0$ and $c_{11} = c_{22} = n/2$ and thus

$$A = 2C(\lambda),\ B = 2S(\lambda).$$

Harmonic Waves with Fourier Frequencies

Next we will fit harmonic waves to data $y_1, \ldots, y_n$, where we restrict ourselves to Fourier frequencies, which facilitates the computations. The following equations are crucial. For arbitrary $0 \le k, m \le [n/2]$ we have (Exercise 3)

$$\sum_{t=1}^{n}\cos\left(2\pi\frac{k}{n}t\right)\cos\left(2\pi\frac{m}{n}t\right) = \begin{cases} n, & k = m = 0 \text{ or } n/2, \quad \text{if } n \text{ is even} \\ n/2, & k = m \neq 0 \text{ and } \neq n/2, \quad \text{if } n \text{ is even} \\ 0, & k \neq m \end{cases}$$
$$\sum_{t=1}^{n}\sin\left(2\pi\frac{k}{n}t\right)\sin\left(2\pi\frac{m}{n}t\right) = \begin{cases} 0, & k = m = 0 \text{ or } n/2, \quad \text{if } n \text{ is even} \\ n/2, & k = m \neq 0 \text{ and } \neq n/2, \quad \text{if } n \text{ is even} \\ 0, & k \neq m \end{cases}$$
$$\sum_{t=1}^{n}\cos\left(2\pi\frac{k}{n}t\right)\sin\left(2\pi\frac{m}{n}t\right) = 0. \tag{3.2}$$

The above equations imply that the $2[n/2] + 1$ vectors in $\mathbb{R}^n$

$$(\sin(2\pi(k/n)t))_{1\le t\le n}, \qquad k = 1, \ldots, [n/2],$$

and

$$(\cos(2\pi(k/n)t))_{1\le t\le n}, \qquad k = 0, \ldots, [n/2],$$

span the space $\mathbb{R}^n$. Note that by (3.2) in the case of n odd the above $2[n/2]+1 = n$ vectors are linearly independent, whereas in the case of an even sample size n the

vector $(\sin(2\pi(k/n)t))_{1\le t\le n}$ with $k = n/2$ is the null vector $(0,\dots,0)$ and the remaining n vectors are again linearly independent. As a consequence we obtain that for a given set of data $y_1,\dots,y_n$, there exist in any case uniquely determined coefficients A_k and B_k, $k = 0,\dots,[n/2]$, with $B_0 := 0$ such that

$$y_t = \sum_{k=0}^{[n/2]} \left(A_k \cos\left(2\pi\frac{k}{n}t\right) + B_k \sin\left(2\pi\frac{k}{n}t\right)\right), \qquad t = 1,\dots,n. \tag{3.3}$$

Next we determine these coefficients A_k, B_k. They are obviously minimizers of the residual sum of squares

$$R := \sum_{t=1}^{n} \left(y_t - \sum_{k=0}^{[n/2]} \left(\alpha_k \cos\left(2\pi\frac{k}{n}t\right) + \beta_k \sin\left(2\pi\frac{k}{n}t\right)\right)\right)^2$$

with respect to α_k, β_k. Taking partial derivatives, equating them to zero and taking into account the linear independence of the above vectors, we obtain that these minimizers solve the normal equations

$$\begin{aligned}
\sum_{t=1}^{n} y_t \cos\left(2\pi\frac{k}{n}t\right) &= \alpha_k \sum_{t=1}^{n} \cos^2\left(2\pi\frac{k}{n}t\right), \qquad k = 0,\dots,[n/2] \\
\sum_{t=1}^{n} y_t \sin\left(2\pi\frac{k}{n}t\right) &= \beta_k \sum_{t=1}^{n} \sin^2\left(2\pi\frac{k}{n}t\right), \qquad k = 1,\dots,[(n-1)/2].
\end{aligned}$$

The solutions of this system are by (3.2) given by

$$\begin{aligned}
A_k &= \begin{cases} \frac{2}{n}\sum_{t=1}^{n} y_t \cos\left(2\pi\frac{k}{n}t\right), & k = 1,\dots,[(n-1)/2] \\ \frac{1}{n}\sum_{t=1}^{n} y_t \cos\left(2\pi\frac{k}{n}t\right), & k = 0 \text{ and } k = n/2, \text{ if } n \text{ is even} \end{cases} \\
B_k &= \frac{2}{n}\sum_{t=1}^{n} y_t \sin\left(2\pi\frac{k}{n}t\right), \quad k = 1,\dots,[(n-1)/2].
\end{aligned} \tag{3.4}$$

One can substitute A_k, B_k in R to obtain directly that $R = 0$ in this case. A popular equivalent formulation of (3.3) is

$$y_t = \frac{\tilde{A}_0}{2} + \sum_{k=1}^{[n/2]} \left(A_k \cos\left(2\pi\frac{k}{n}t\right) + B_k \sin\left(2\pi\frac{k}{n}t\right)\right), \qquad t = 1,\dots,n, \tag{3.5}$$

with A_k, B_k as in (3.4) for $k = 1,\dots,[n/2]$, $B_{n/2} = 0$ for an even n, and

$$\tilde{A}_0 = 2A_0 = \frac{2}{n}\sum_{t=1}^{n} y_t = 2\bar{y}.$$

Up to the factor 2, the coefficients $A_k,\ B_k$ coincide with the empirical covariances $C(k/n)$ and $S(k/n),\ k = 1, \ldots, [(n-1)/2]$, defined in (3.1). This follows from the equations (Exercise 2)

$$\sum_{t=1}^{n} \cos\Big(2\pi \frac{k}{n} t\Big) = \sum_{t=1}^{n} \sin\Big(2\pi \frac{k}{n} t\Big) = 0, \qquad k = 1, \ldots, [n/2]. \tag{3.6}$$

3.2 The Periodogram

In the preceding section we exactly fitted a harmonic wave with Fourier frequencies $\lambda_k = k/n,\ k = 0, \ldots, [n/2]$, to a given series $y_1, \ldots, y_n$. Example 3.1.1 shows that a harmonic wave including only two frequencies already fits the Star Data quite well. There is a general tendency that a time series is governed by different frequencies $\lambda_1, \ldots, \lambda_r$ with $r < [n/2]$, which on their part can be approximated by Fourier frequencies $k_1/n, \ldots, k_r/n$ if n is sufficiently large. The question *which* frequencies actually govern a time series leads to the *intensity* of a frequency λ. This number reflects the influence of the harmonic component with frequency λ on the series. The intensity of the Fourier frequency $\lambda = k/n,\ 1 \le k \le [n/2]$, is defined via its residual sum of squares. We have by (3.2), (3.6) and the normal equations

$$\begin{aligned}\sum_{t=1}^{n} &\left(y_t - \bar{y} - A_k \cos\Big(2\pi \frac{k}{n} t\Big) - B_k \sin\Big(2\pi \frac{k}{n} t\Big)\right)^2 \\ &= \sum_{t=1}^{n} (y_t - \bar{y})^2 - \frac{n}{2}\big(A_k^2 + B_k^2\big), \qquad k = 1, \ldots, [(n-1)/2],\end{aligned}$$

and

$$\sum_{t=1}^{n} (y_t - \bar{y})^2 = \frac{n}{2} \sum_{k=1}^{[n/2]} \big(A_k^2 + B_k^2\big).$$

The number $(n/2)(A_k^2 + B_k^2) = 2n(C^2(k/n) + S^2(k/n))$ is therefore the contribution of the harmonic component with Fourier frequency $k/n,\ k = 1, \ldots, [(n-1)/2]$, to the total variation $\sum_{t=1}^{n} (y_t - \bar{y})^2$. It is called the *intensity of the frequency* k/n. Further insight is gained from the Fourier analysis in Theorem 3.2.4. For general frequencies $\lambda \in \mathbb{R}$ we define its intensity now by

$$\begin{aligned} I(\lambda) &= n\big(C(\lambda)^2 + S(\lambda)^2\big) \\ &= \frac{1}{n}\left(\Big(\sum_{t=1}^{n} (y_t - \bar{y}) \cos(2\pi\lambda t)\Big)^2 + \Big(\sum_{t=1}^{n} (y_t - \bar{y}) \sin(2\pi\lambda t)\Big)^2\right). \end{aligned} \tag{3.7}$$

This function is called the *periodogram*. The following Theorem implies in particular that it is sufficient to define the periodogram on the interval $[0, 1]$. For Fourier frequencies we obtain from (3.4) and (3.6)

$$I(k/n) = \frac{n}{4}\big(A_k^2 + B_k^2\big), \qquad k = 1, \ldots, [(n-1)/2].$$

Theorem 3.2.1. *We have*

(i) $I(0) = 0$,

(ii) I *is an even function, i.e.,* $I(\lambda) = I(-\lambda)$ *for any* $\lambda \in \mathbb{R}$,

(iii) I *has the period 1.*

Proof. Part (i) follows from $\sin(0) = 0$ and $\cos(0) = 1$, while (ii) is a consequence of $\cos(-x) = \cos(x)$, $\sin(-x) = -\sin(x)$, $x \in \mathbb{R}$. Part (iii) follows from the fact that 2π is the fundamental period of sin and cos. □

Theorem 3.2.1 implies that the function $I(\lambda)$ is completely determined by its values on $[0, 0.5]$. The periodogram is often defined on the scale $[0, 2\pi]$ instead of $[0, 1]$ by putting $I^*(\omega) := 2I(\omega/(2\pi))$; this version is, for example, used in SAS. In view of Theorem 3.2.4 below we prefer $I(\lambda)$, however.

The following figure displays the periodogram of the Star Data from Example 3.1.1. It has two obvious peaks at the Fourier frequencies $21/600 = 0.035 \approx 1/28.57$ and $25/600 = 1/24 \approx 0.04167$. This indicates that essentially two cycles with period 24 and 28 or 29 are inherent in the data. A least squares approach for the determination of the coefficients A_i, B_i, $i = 1, 2$ with frequencies $1/24$ and $1/29$ as done in Program 3_1_1 then leads to the coefficients in Example 3.1.1.

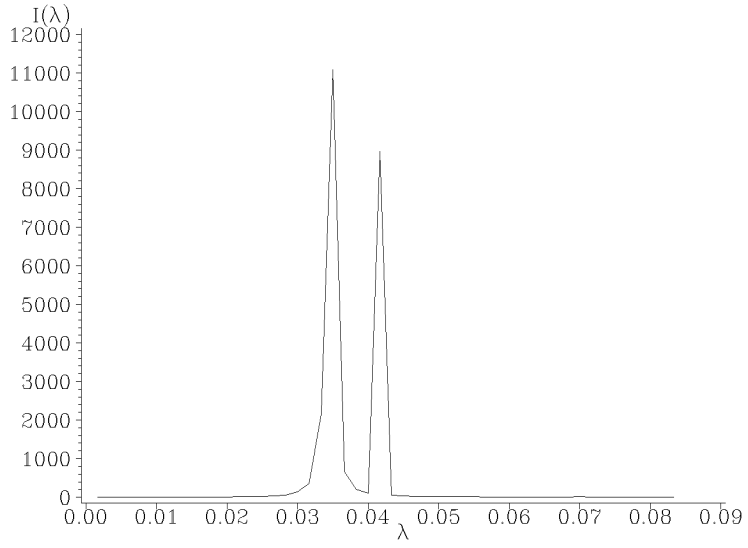


Figure 3.2.1. Periodogram of the Star Data.

		COS_01		
		34.1933		
PERIOD	COS_01	SIN_01	P	LAMBDA
28.5714	-0.91071	8.54977	11089.19	0.035000
24.0000	-0.06291	7.73396	8972.71	0.041667
30.0000	0.42338	-3.76062	2148.22	0.033333
27.2727	-0.16333	2.09324	661.25	0.036667
31.5789	0.20493	-1.52404	354.71	0.031667
26.0870	-0.05822	1.18946	212.73	0.038333

Table 3.2.1. Print of the constant $\tilde{A}_0 = 2A_0 = 2\bar{y}$ and of the six Fourier frequencies $\lambda = k/n$ with largest $I(k/n)$-values, their inverses and the Fourier coefficients pertaining to the Star Data.

```
***     Program 3_2_1   ***;
TITLE1 'Periodogram';
TITLE2 'Star Data';

DATA data1;
  INFILE 'c:\data\star.txt';
  INPUT lumen @@;

PROC SPECTRA DATA=data1 COEF P OUT=data2;
  VAR lumen;

DATA data3;
  SET data2(FIRSTOBS=2);
  p=P_01/2;
  lambda=FREQ/(2*CONSTANT('PI'));
  DROP P_01 FREQ;

SYMBOL1 V=NONE C=GREEN I=JOIN;
AXIS1 LABEL=('I(' F=CGREEK 'l)');
AXIS2 LABEL=(F=CGREEK 'l');
PROC GPLOT DATA=data3(OBS=50);
  PLOT p*lambda=1 / VAXIS=AXIS1 HAXIS=AXIS2;

PROC SORT DATA=data3 OUT=data4;
  BY DESCENDING p;

PROC PRINT DATA=data2(OBS=1) NOOBS;
  VAR COS_01;
PROC PRINT DATA=data4(OBS=6) NOOBS;
RUN; QUIT;
```

The first step is to read the star data from an external file into a data set. Using the SAS procedure `SPECTRA` with the options `P` (periodogram), `COEF` (Fourier coefficients), `OUT=data2` and the `VAR` statement specifying the variable of interest, an output data set is generated. It contains periodogram data `P_01` evaluated at the Fourier frequencies, a `FREQ` variable for this frequencies, the pertaining period in the variable `PERIOD` and the variables `COS_01` and `SIN_01` with the coefficients for the harmonic waves. Because SAS uses different definitions for the frequencies and the periodogram, here in `data3` new variables `lambda` (dividing `FREQ` by 2π to eliminate an additional factor 2π) and `p` (dividing `P_01` by 2) are created and the no more needed ones are dropped. By means of the data set option `FIRSTOBS=2` the first observation of `data2` is excluded from the resulting data set.

The following `PROC GPLOT` just takes the first 50 observations of `data3` into account. This means a restriction of `lambda` up to $50/600 = 1/12$, the part with the greatest peaks in the periodogram.

The procedure `SORT` generates a new data set

`data4` containing the same observations as the input data set `data3`, but they are sorted in descending order of the periodogram values. The two `PROC PRINT` statements at the end make SAS print to datasets `data2` and `data4`.

The first part of the output is the coefficient $\tilde{A}_0$ which is equal to two times the mean of the `lumen` data. The results for the Fourier frequencies with the six greatest periodogram values constitute the second part of the output. Note that the meaning of `COS_01` and `SIN_01` are slightly different from the definitions of A_k and B_k in (3.4), because SAS lets the index run from 0 to $n-1$ instead of 1 to n.

The Fourier Transform

From Euler's equation $e^{iz} = \cos(z) + i\sin(z),\ z \in \mathbb{R}$, we obtain for $\lambda \in \mathbb{R}$

$$D(\lambda) := C(\lambda) - iS(\lambda) = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})e^{-i2\pi\lambda t}.$$

The periodogram is a function of $D(\lambda)$, since $I(\lambda) = n|D(\lambda)|^2$. Unlike the periodogram, the number $D(\lambda)$ contains the complete information about $C(\lambda)$ and $S(\lambda)$, since both values can be recovered from the complex number $D(\lambda)$, being its real and negative imaginary part. In the following we view the data $y_1, \ldots, y_n$ again as a clipping from an infinite series y_t, $t \in \mathbb{Z}$. Let $a := (a_t)_{t\in\mathbb{Z}}$ be an absolutely summable sequence of real numbers. For such a sequence a the complex valued function

$$f_a(\lambda) = \sum_{t\in\mathbb{Z}} a_t e^{-i2\pi\lambda t}, \qquad \lambda \in \mathbb{R},$$

is said to be its *Fourier transform.* It links the empirical autocovariance function to the periodogram as it will turn out in Theorem 3.2.3 that the latter is the Fourier transform of the first. Note that $\sum_{t\in\mathbb{Z}} |a_t e^{-i2\pi\lambda t}| = \sum_{t\in\mathbb{Z}} |a_t| < \infty$, since $|e^{ix}| = 1$ for any $x \in \mathbb{R}$. The Fourier transform of $a_t = (y_t - \bar{y})/n,\ t = 1, \ldots, n$, and $a_t = 0$ elsewhere is then given by $D(\lambda)$. The following elementary properties of the Fourier transform are immediate consequences of the arguments in the proof of Theorem 3.2.1. In particular we obtain that the Fourier transform is already determined by its values on $[0, 0.5]$.

Theorem 3.2.2. *We have*

(i) $f_a(0) = \sum_{t\in\mathbb{Z}} a_t$,

(ii) $f_a(-\lambda)$ *and* $f_a(\lambda)$ *are conjugate complex numbers i.e.,* $f_a(-\lambda) = \overline{f_a(\lambda)}$,

(iii) f_a *has the period 1.*

Autocorrelation Function and Periodogram

Information about cycles that are inherent in given data, can also be deduced from the empirical autocorrelation function. The following figure displays the autocorrelation function of the Bankruptcy Data, introduced in Exercise 17 of Chapter 1.

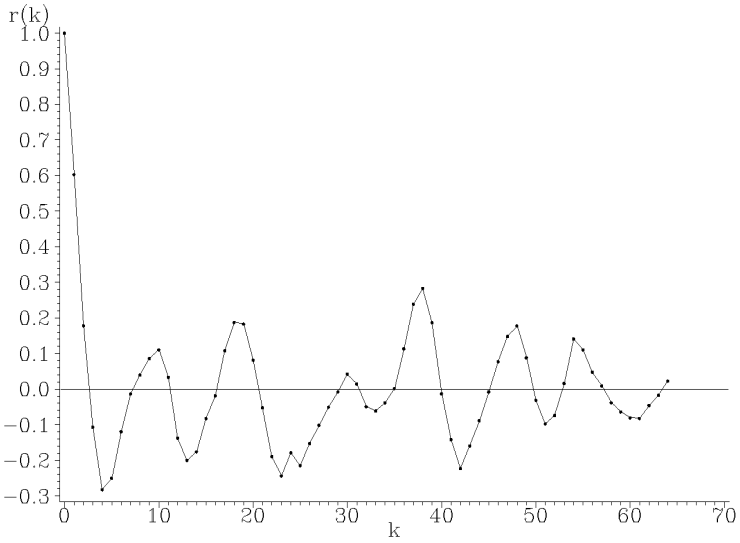


Figure 3.2.2. Autocorrelation function of the Bankruptcy Data.

```
***     Program 3_2_2   ***
TITLE1 'Correlogram';
TITLE2 'Bankruptcy Data';

DATA data1;
  INFILE 'c:\data\bankrupt.txt';
  INPUT year bankrupt;

PROC ARIMA DATA=data1;
```

```
  IDENTIFY VAR=bankrupt NLAG=64 OUTCOV=corr NOPRINT;

AXIS1 LABEL=('r(k)');
AXIS2 LABEL=('k');
SYMBOL1 V=DOT C=GREEN I=JOIN H=0.4 W=1;
PROC GPLOT DATA=corr;
  PLOT CORR*LAG / VAXIS=AXIS1 HAXIS=AXIS2 VREF=0;
RUN; QUIT;
```

After reading the data from an external file into a data step, the procedure `ARIMA` calculates the empirical autocorrelation function and stores them into a new data set. The correlogram is generated using `PROC GPLOT`.

The next figure displays the periodogram of the Bankruptcy Data.

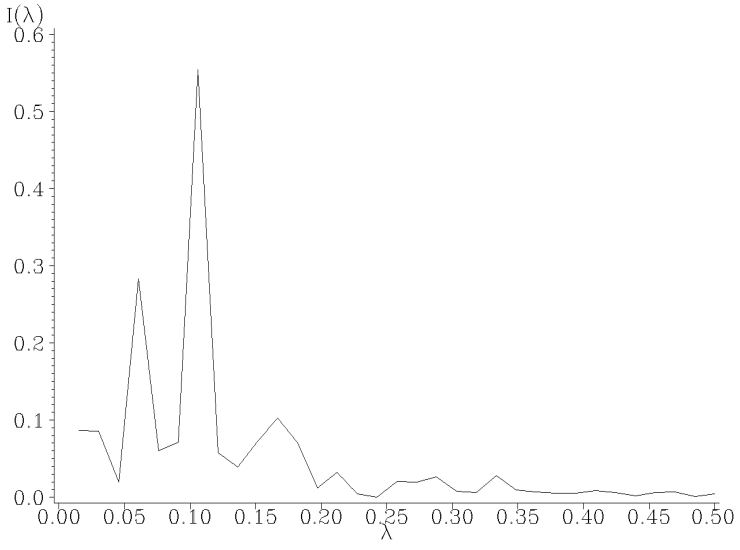


Figure 3.2.3. Periodogram of the Bankruptcy Data.

```
***     Program 3_2_3   ***;
TITLE1 'Periodogram';
TITLE2 'Bankruptcy Data';

DATA data1;
  INFILE 'c:\data\bankrupt.txt';
  INPUT year bankrupt;

PROC SPECTRA DATA=data1 P OUT=data2;
  VAR bankrupt;

DATA data3;
  SET data2(FIRSTOBS=2);
  p=P_01/2;
  lambda=FREQ/(2*CONSTANT('PI'));

SYMBOL1 V=NONE C=GREEN I=JOIN;
AXIS1 LABEL=('I' F=CGREEK '(l)') ;
AXIS2 ORDER=(0 TO 0.5 BY 0.05) LABEL=(F=CGREEK 'l');
PROC GPLOT DATA=data3;
  PLOT p*lambda / VAXIS=AXIS1 HAXIS=AXIS2;
RUN; QUIT;
```

This program again first reads the data and then starts a spectral analysis by `PROC SPECTRA`. Due to the reasons mentioned in the comments to Program 3_2_1 there are some transformations of the periodogram and the frequency values generated by `PROC SPECTRA` done in `data3`. The graph results from the statements in `PROC GPLOT`.

The autocorrelation function of the Bankruptcy Data has extreme values at about multiples of 9 years, which indicates a period of length 9. This is underlined by the periodogram in Figure 3.2.3, which has a peak at $\lambda = 0.11$, corresponding to a period of $1/0.11 \sim 9$ years as well. As mentioned above, there is actually a close relationship between the *empirical* autocovariance function and the periodogram. The corresponding result for the *theoretical* autocovariances is given in Chapter 4.

Theorem 3.2.3. *Denote by c the empirical autocovariance function of* $y_1, \ldots, y_n$, *i.e.,* $c(k) = n^{-1} \sum_{j=1}^{n-k} (y_j - \bar{y})(y_{j+k} - \bar{y})$, $k = 0, \ldots, n-1$, *where* $\bar{y} := n^{-1} \sum_{j=1}^{n} y_j$.

Then we have with $c(-k) := c(k)$

$$\begin{aligned} I(\lambda) &= c(0) + 2\sum_{k=1}^{n-1} c(k)\cos(2\pi\lambda k) \\ &= \sum_{k=-(n-1)}^{n-1} c(k)e^{-i2\pi\lambda k}. \end{aligned}$$

Proof. From the equation $\cos(x_1)\cos(x_2) + \sin(x_1)\sin(x_2) = \cos(x_1 - x_2)$ for $x_1, x_2 \in \mathbb{R}$ we obtain

$$\begin{aligned} I(\lambda) &= \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n}(y_s - \bar{y})(y_t - \bar{y}) \\ &\quad \times \big(\cos(2\pi\lambda s)\cos(2\pi\lambda t) + \sin(2\pi\lambda s)\sin(2\pi\lambda t)\big) \\ &= \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n} a_{st}, \end{aligned}$$

where $a_{st} := (y_s - \bar{y})(y_t - \bar{y})\cos(2\pi\lambda(s-t))$. Since $a_{st} = a_{ts}$ and $\cos(0) = 1$ we have moreover

$$\begin{aligned} I(\lambda) &= \frac{1}{n}\sum_{t=1}^{n} a_{tt} + \frac{2}{n}\sum_{k=1}^{n-1}\sum_{j=1}^{n-k} a_{jj+k} \\ &= \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2 + 2\sum_{k=1}^{n-1}\Big(\frac{1}{n}\sum_{j=1}^{n-k}(y_j - \bar{y})(y_{j+k} - \bar{y})\Big)\cos(2\pi\lambda k) \\ &= c(0) + 2\sum_{k=1}^{n-1} c(k)\cos(2\pi\lambda k). \end{aligned}$$

The complex representation of the periodogram is then obvious:

$$\begin{aligned} \sum_{k=-(n-1)}^{n-1} c(k)e^{-i2\pi\lambda k} &= c(0) + \sum_{k=1}^{n-1} c(k)\big(e^{i2\pi\lambda k} + e^{-i2\pi\lambda k}\big) \\ &= c(0) + \sum_{k=1}^{n-1} c(k)2\cos(2\pi\lambda k) = I(\lambda). \end{aligned}$$

□

Inverse Fourier Transform

The empirical autocovariance function can be recovered from the periodogram, which is the content of our next result. Since the periodogram is the Fourier transform of the empirical autocovariance function, this result is a special case of the inverse Fourier transform in Theorem 3.2.5 below.

Theorem 3.2.4. *The periodogram*

$$I(\lambda) = \sum_{k=-(n-1)}^{n-1} c(k) e^{-i2\pi\lambda k}, \qquad \lambda \in \mathbb{R},$$

satisfies the inverse formula

$$c(k) = \int_0^1 I(\lambda) e^{i2\pi\lambda k}\, d\lambda, \qquad |k| \leq n-1.$$

In particular for $k = 0$ we obtain

$$c(0) = \int_0^1 I(\lambda)\, d\lambda.$$

The sample variance $c(0) = n^{-1} \sum_{j=1}^n (y_j - \bar{y})^2$ equals, therefore, the area under the curve $I(\lambda),\ 0 \leq \lambda \leq 1$. The integral $\int_{\lambda_1}^{\lambda_2} I(\lambda)\, d\lambda$ can be interpreted as that portion of the total variance $c(0)$, which is contributed by the harmonic waves with frequencies $\lambda \in [\lambda_1, \lambda_2]$, where $0 \leq \lambda_1 < \lambda_2 \leq 1$. The periodogram consequently shows the distribution of the total variance among the frequencies $\lambda \in [0, 1]$. A peak of the periodogram at a frequency λ_0 implies, therefore, that a large part of the total variation $c(0)$ of the data can be explained by the harmonic wave with that frequency λ_0.

The following result is the inverse formula for general Fourier transforms.

Theorem 3.2.5. *For an absolutely summable sequence $a := (a_t)_{t\in\mathbb{Z}}$ with Fourier transform $f_a(\lambda) = \sum_{t\in\mathbb{Z}} a_t e^{-i2\pi\lambda t},\ \lambda \in \mathbb{R}$, we have*

$$a_t = \int_0^1 f_a(\lambda) e^{i2\pi\lambda t}\, d\lambda, \qquad t \in \mathbb{Z}.$$

Proof. The dominated convergence theorem implies

$$\begin{aligned} \int_0^1 f_a(\lambda) e^{i2\pi\lambda t}\, d\lambda &= \int_0^1 \Big(\sum_{s\in\mathbb{Z}} a_s e^{-i2\pi\lambda s} \Big) e^{i2\pi\lambda t}\, d\lambda \\ &= \sum_{s\in\mathbb{Z}} a_s \int_0^1 e^{i2\pi\lambda(t-s)}\, d\lambda = a_t, \end{aligned}$$

since

$$\int_0^1 e^{i2\pi\lambda(t-s)}\,d\lambda = \begin{cases} 1, & \text{if } s = t \\ 0, & \text{if } s \neq t. \end{cases}$$

□

The inverse Fourier transformation shows that the complete sequence $(a_t)_{t\in\mathbb{Z}}$ can be recovered from its Fourier transform. This implies in particular that the Fourier transforms of absolutely summable sequences are uniquely determined. The analysis of a time series in the frequency domain is, therefore, equivalent to its analysis in the time domain, which is based on an evaluation of its autocovariance function.

Aliasing

Suppose that we observe a *continuous time* process $(Z_t)_{t\in\mathbb{R}}$ only through its values at $k\Delta$, $k \in \mathbb{Z}$, where $\Delta > 0$ is the sampling interval, i.e., we actually observe $(Y_k)_{k\in\mathbb{Z}} = (Z_{k\Delta})_{k\in\mathbb{Z}}$. Take, for example, $Z_t := \cos(2\pi(9/11)t)$, $t \in \mathbb{R}$. The following figure shows that at $k \in \mathbb{Z}$, i.e., $\Delta = 1$, the observations Z_k coincide with X_k, where $X_t := \cos(2\pi(2/11)t)$, $t \in \mathbb{R}$. With the sampling interval $\Delta = 1$, the observations Z_k with high frequency 9/11 can, therefore, not be distinguished from the X_k, which have low frequency 2/11.

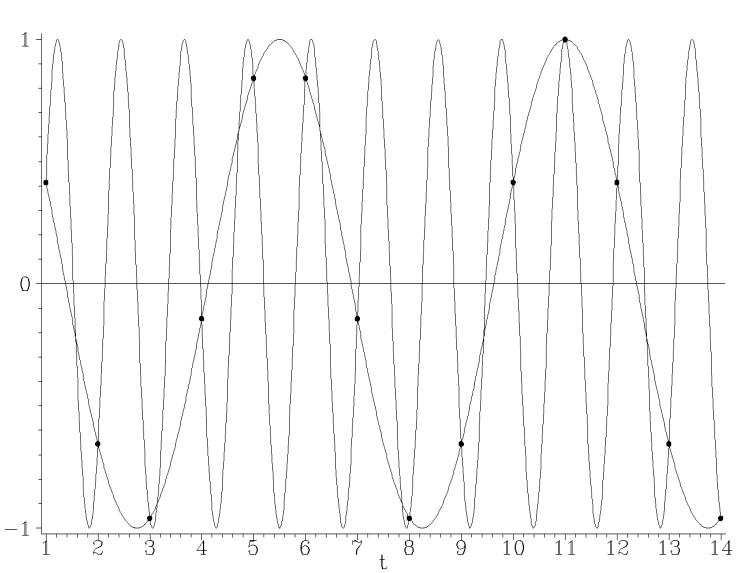


Figure 3.2.4. Aliasing of $\cos(2\pi(9/11)k)$ and $\cos(2\pi(2/11)k)$.

```
***     Program 3_2_4   ***;
TITLE1 'Aliasing';

DATA data1;
  DO t=1 TO 14 BY .01;
    y1=COS(2*CONSTANT('PI')*2/11*t);
    y2=COS(2*CONSTANT('PI')*9/11*t);
    OUTPUT;
  END;

DATA data2;
  DO t0=1 TO 14;
    y0=COS(2*CONSTANT('PI')*2/11*t0);
    OUTPUT;
  END;

DATA data3;
  MERGE data1 data2;

SYMBOL1 V=DOT  C=GREEN I=NONE H=.8;
SYMBOL2 V=NONE C=RED   I=JOIN;
AXIS1 LABEL=NONE;
AXIS2 LABEL=('t');
PROC GPLOT DATA=data3;
  PLOT y0*t0=1 y1*t=2 y2*t=2 / OVERLAY VAXIS=AXIS1
                               HAXIS=AXIS2 VREF=0;
RUN; QUIT;
```

In the first data step a tight grid for the cosine waves with frequencies 2/11 and 9/11 is generated. In the second data step the values of the cosine wave with frequency 2/11 are generated just for integer values of `t` symbolizing the observation points.

After merging the two data sets the two waves are plotted using the `JOIN` option in the `SYMBOL` statement while the values at the observation points are displayed in the same graph by dot symbols.

This phenomenon that a high frequency component takes on the values of a lower one, is called *aliasing*. It is caused by the choice of the sampling interval Δ, which is 1 in Figure 3.2.4. If a series is not just a constant, then the shortest observable period is 2Δ. The highest observable frequency λ^* with period $1/\lambda^*$, therefore, satisfies $1/\lambda^* \geq 2\Delta$, i.e., $\lambda^* \leq 1/(2\Delta)$. This frequency $1/(2\Delta)$ is known as the

Nyquist frequency. The sampling interval Δ should, therefore, be chosen small enough, so that $1/(2\Delta)$ is above the frequencies under study. If, for example, a process is a harmonic wave with frequencies $\lambda_1, \ldots, \lambda_p$, then Δ should be chosen such that $\lambda_i \leq 1/(2\Delta)$, $1 \leq i \leq p$, in order to visualize p different frequencies. For the periodic curves in Figure 3.2.4 this means to choose $9/11 \leq 1/(2\Delta)$ or $\Delta \leq 11/18$.

Exercises

1. Let $y(t) = A\cos(2\pi\lambda t) + B\sin(2\pi\lambda t)$ be a harmonic component. Show that y can be written as $y(t) = \alpha\cos(2\pi\lambda t - \varphi)$, where α is the amplitiude, i.e., the maximum departure of the wave from zero and φ is the phase displacement.

2. Show that

$$\sum_{t=1}^{n} \cos(2\pi\lambda t) = \begin{cases} n, & \lambda \in \mathbb{Z} \\ \cos(\pi\lambda(n+1))\frac{\sin(\pi\lambda n)}{\sin(\pi\lambda)}, & \lambda \notin \mathbb{Z} \end{cases}$$

$$\sum_{t=1}^{n} \sin(2\pi\lambda t) = \begin{cases} 0, & \lambda \in \mathbb{Z} \\ \sin(\pi\lambda(n+1))\frac{\sin(\pi\lambda n)}{\sin(\pi\lambda)}, & \lambda \notin \mathbb{Z}. \end{cases}$$

Hint: Compute $\sum_{t=1}^{n} e^{i2\pi\lambda t}$, where $e^{i\varphi} = \cos(\varphi) + i\sin(\varphi)$ is the complex valued exponential function.

3. Verify the equations (3.2). Hint: Exercise 2.

4. Suppose that the time series $(y_t)_t$ satisfies the additive model with seasonal component

$$s(t) = \sum_{k=1}^{s} A_k \cos\left(2\pi\frac{k}{s}t\right) + \sum_{k=1}^{s} B_k \sin\left(2\pi\frac{k}{s}t\right).$$

Show that $s(t)$ is eliminated by the seasonal differencing $\Delta_s y_t = y_t - y_{t-s}$.

5. Fit a harmonic component with frequency λ to a time series $y_1, \ldots, y_N$, where $\lambda \in \mathbb{Z}$ and $\lambda - 0.5 \in \mathbb{Z}$. Compute the least squares estimator and the pertaining residual sum of squares.

6. Put $y_t = t$, $t = 1, \ldots, n$. Show that

$$I(k/n) = \frac{n}{4\sin^2(\pi k/n)}, \quad k = 1, \ldots, [(n-1)/2].$$

Hint: Use the equations

$$\sum_{t=1}^{n-1} t\sin(\theta t) = \frac{\sin(n\theta)}{4\sin^2(\theta/2)} - \frac{n\cos((n-0.5)\theta)}{2\sin(\theta/2)}$$

$$\sum_{t=1}^{n-1} t\cos(\theta t) = \frac{n\sin((n-0.5)\theta)}{2\sin(\theta/2)} - \frac{1-\cos(n\theta)}{4\sin^2(\theta/2)}.$$

7. (Unemployed1 Data) Plot the periodogram of the first order differences of the numbers of unemployed in the building trade as introduced in Example 1.1.1.

8. (Airline Data) Plot the periodogram of the variance stabilized and trend adjusted Airline Data, introduced in Example 1.3.1. Add a seasonal adjustment and compare the periodograms.

9. The contribution of the autocovariance $c(k)$, $k \geq 1$, to the periodogram can be illustrated by plotting the functions $\pm\cos(2\pi\lambda k)$, $\lambda \in [0.5]$.

(i) Which conclusion about the intensities of large or small frequencies can be drawn from a positive value $c(1) > 0$ or a negative one $c(1) < 0$?

(ii) Which effect has an increase of $|c(2)|$ if all other parameters remain unaltered?

(iii) What can you say about the effect of an increase of $c(k)$ on the periodogram at the values $40, 1/k, 2/k, 3/k, \ldots$ and the intermediate values $1/2k, 3/(2k), 5/(2k)$? Illustrate the effect at a time series with seasonal component $k = 12$.

10. Establish a version of the inverse Fourier transform in real terms.

11. Let $\boldsymbol{a} = (a_t)_{t\in\mathbb{Z}}$ and $\boldsymbol{b} = (b_t)_{t\in\mathbb{Z}}$ be absolute summable sequences.

(i) Show that for $\alpha\boldsymbol{a} + \beta\boldsymbol{b} := (\alpha a_t + \beta b_t)_{t\in\mathbb{Z}}$, $\alpha, \beta \in \mathbb{R}$,

$$f_{\alpha\boldsymbol{a}+\beta\boldsymbol{b}}(\lambda) = \alpha f_{\boldsymbol{a}}(\lambda) + \beta f_{\boldsymbol{b}}(\lambda).$$

(ii) For $\boldsymbol{ab} := (a_t b_t)_{t\in\mathbb{Z}}$ we have

$$f_{\boldsymbol{ab}}(\lambda) = f_{\boldsymbol{a}} * f_{\boldsymbol{b}}(\lambda) := \int_0^1 f_{\boldsymbol{a}}(\mu) f_{\boldsymbol{b}}(\lambda - \mu)\, d\mu.$$

(iii) Show that for $\boldsymbol{a} * \boldsymbol{b} := (\sum_{s\in\mathbb{Z}} a_s b_{t-s})_{t\in\mathbb{Z}}$ (convolution)

$$f_{\boldsymbol{a}*\boldsymbol{b}}(\lambda) = f_{\boldsymbol{a}}(\lambda) f_{\boldsymbol{b}}(\lambda).$$

12. (Fast Fourier Transform (FFT)) The Fourier transform of a finite sequence a_0, $a_1, \ldots, a_{N-1}$ can be represented under suitable conditions as the composition of Fourier transforms. Put

$$f(s/N) = \sum_{t=0}^{N-1} a_t e^{-i2\pi st/N}, \quad s = 0, \ldots, N-1,$$

which is the Fourier transform of length N. Suppose that $N = KM$ with $K, M \in \mathbb{N}$. Show that f can be represented as Fourier transform of length K, computed for a Fourier transform of length M.
Hint: Each $t, s \in \{0, \ldots, N-1\}$ can uniquely be written as

$$\begin{aligned} t &= t_0 + t_1 K, \quad t_0 \in \{0, \ldots, K-1\}, \quad t_1 \in \{0, \ldots, M-1\} \\ s &= s_0 + s_1 M, \quad s_0 \in \{0, \ldots, M-1\}, \quad s_1 \in \{0, \ldots, K-1\}. \end{aligned}$$

Sum over t_0 and t_1.

13. (Star Data) Suppose that the Star Data are only observed weekly (i.e., keep only every seventh observation). Is an aliasing effect observable?

Chapter 4

The Spectrum of a Stationary Process

In this chapter we investigate the spectrum of a real valued stationary process, which is the Fourier transform of its (theoretical) autocovariance function. Its empirical counterpart, the periodogram, was investigated in the preceding sections, cf. Theorem 3.2.3.

Let $(Y_t)_{t\in\mathbb{Z}}$ be a (real valued) stationary process with absolutely summable autocovariance function $\gamma(t)$, $t \in \mathbb{Z}$. Its Fourier transform

$$f(\lambda) := \sum_{t\in\mathbb{Z}} \gamma(t) e^{-i2\pi\lambda t} = \gamma(0) + 2\sum_{t\in\mathbb{N}} \gamma(t)\cos(2\pi\lambda t), \quad \lambda \in \mathbb{R},$$

is called *spectral density* or *spectrum* of the process $(Y_t)_{t\in\mathbb{Z}}$. By the inverse Fourier transform in Theorem 3.2.5 we have

$$\gamma(t) = \int_0^1 f(\lambda) e^{i2\pi\lambda t}\, d\lambda = \int_0^1 f(\lambda)\cos(2\pi\lambda t)\, d\lambda.$$

For $t = 0$ we obtain

$$\gamma(0) = \int_0^1 f(\lambda)\, d\lambda,$$

which shows that the spectrum is a decomposition of the variance $\gamma(0)$. In Section 4.3 we will in particular compute the spectrum of an $ARMA$-process. As a preparatory step we investigate properties of spectra for arbitrary absolutely summable filters.

4.1 Characterizations of Autocovariance Functions

Recall that the autocovariance function $\gamma : \mathbb{Z} \to \mathbb{R}$ of a stationary process $(Y_t)_{t\in\mathbb{Z}}$ is given by

$$\gamma(h) = \mathrm{E}(Y_{t+h}Y_t) - \mathrm{E}(Y_{t+h})\,\mathrm{E}(Y_t), \quad h \in \mathbb{Z},$$

with the properties

$$\gamma(0) \geq 0,\ |\gamma(h)| \leq \gamma(0),\ \gamma(h) = \gamma(-h), \quad h \in \mathbb{Z}. \tag{4.1}$$

The following result characterizes an autocovariance function in terms of positive semidefiniteness.

Theorem 4.1.1. *A symmetric function $K : \mathbb{Z} \to \mathbb{R}$ is the autocovariance function of a stationary process $(Y_t)_{t\in\mathbb{Z}}$ iff K is a positive semidefinite function, i.e., $K(-n) = K(n)$ and*

$$\sum_{1\leq r,s\leq n} x_r K(r-s) x_s \geq 0 \tag{4.2}$$

for arbitrary $n \in \mathbb{N}$ and $x_1, \ldots, x_n \in \mathbb{R}$.

Proof. It is easy to see that (4.2) is a necessary condition for K to be the autocovariance function of a stationary process, see Exercise 19. It remains to show that (4.2) is sufficient, i.e., we will construct a stationary process, whose autocovariance function is K.

We will define a family of finite-dimensional normal distributions, which satisfies the *consistency condition* of *Kolmogorov's theorem*, cf. Theorem 1.2.1 in Brockwell and Davies (1991). This result implies the existence of a process $(V_t)_{t\in\mathbb{Z}}$, whose finite dimensional distributions coincide with the given family.

Define the $n \times n$-matrix

$$K^{(n)} := \big(K(r-s)\big)_{1\leq r,s\leq n},$$

which is positive semidefinite. Consequently there exists an n-dimensional normal distributed random vector $(V_1, \ldots, V_n)$ with mean vector zero and covariance matrix $K^{(n)}$. Define now for each $n \in \mathbb{N}$ and $t \in \mathbb{Z}$ a distribution function on $\mathbb{R}^n$ by

$$F_{t+1,\ldots,t+n}(v_1, \ldots, v_n) := P\{V_1 \leq v_1, \ldots, V_n \leq v_n\}.$$

This defines a family of distribution functions indexed by consecutive integers. Let now $t_1 < \cdots < t_m$ be arbitrary integers. Choose $t \in \mathbb{Z}$ and $n \in \mathbb{N}$ such that $t_i = t + n_i$, where $1 \leq n_1 < \cdots < n_m \leq n$. We define now

$$F_{t_1,\ldots,t_m}((v_i)_{1\leq i\leq m}) := P\{V_{n_i} \leq v_i,\ 1 \leq i \leq m\}.$$

Note that $F_{t_1,\ldots,t_m}$ does not depend on the special choice of t and n and thus, we have defined a family of distribution functions indexed by $t_1 < \cdots < t_m$ on $\mathbb{R}^m$ for each $m \in \mathbb{N}$, which obviously satisfies the consistency condition of *Kolmogorov's theorem*. This result implies the existence of a process $(V_t)_{t\in\mathbb{Z}}$, whose finite dimensional distribution at $t_1 < \cdots < t_m$ has distribution function $F_{t_1,\ldots,t_m}$. This process has, therefore, mean vector zero and covariances $\mathrm{E}(V_{t+h}V_t) = K(h)$, $h \in \mathbb{Z}$. □

Spectral Distribution Function and Spectral Density

The preceding result provides a characterization of an autocovariance function in terms of positive semidefiniteness. The following characterization of positive semidefinite functions is known as *Herglotz's* theorem. We use in the following the notation $\int_0^1 g(\lambda)\,dF(\lambda)$ in place of $\int_{(0,1]} g(\lambda)\,dF(\lambda)$.

Theorem 4.1.2. *A symmetric function $\gamma : \mathbb{Z} \to \mathbb{R}$ is positive semidefinite iff it can be represented as an integral*

$$\gamma(h) = \int_0^1 e^{i2\pi\lambda h}\,dF(\lambda) = \int_0^1 \cos(2\pi\lambda h)\,dF(\lambda), \quad h \in \mathbb{Z}, \tag{4.3}$$

where F is a real valued measure generating function on $[0,1]$ with $F(0) = 0$. The function F is uniquely determined.

The uniquely determined function F, which is a right-continuous, increasing and bounded function, is called the *spectral distribution function* of γ. If F has a derivative f and, thus, $F(\lambda) = F(\lambda) - F(0) = \int_0^\lambda f(x)\,dx$ for $0 \le \lambda \le 1$, then f is called the *spectral density* of γ. Note that the property $\sum_{h\ge 0} |\gamma(h)| < \infty$ already implies the existence of a spectral density of γ, cf. Theorem 3.2.5.

Recall that $\gamma(0) = \int_0^1 dF(\lambda) = F(1)$ and thus, the autocorrelation function $\rho(h) = \gamma(h)/\gamma(0)$ has the above integral representation, but with F replaced by the *distribution function* $F/\gamma(0)$.

Proof of Theorem 4.1.2. We establish first the uniqueness of F. Let G be another measure generating function with $G(\lambda) = 0$ for $\lambda \le 0$ and constant for $\lambda \ge 1$ such that

$$\gamma(h) = \int_0^1 e^{i2\pi\lambda h}\,dF(\lambda) = \int_0^1 e^{i2\pi\lambda h}\,dG(\lambda), \quad h \in \mathbb{Z}.$$

Let now ψ be a continuous function on $[0,1]$. From calculus we know (cf. Section 4.24 in Rudin (1974)) that we can find for arbitrary $\varepsilon > 0$ a trigonometric

polynomial $p_\varepsilon(\lambda) = \sum_{h=-N}^{N} a_h e^{i2\pi\lambda h}$, $0 \le \lambda \le 1$, such that

$$\sup_{0\le\lambda\le 1} |\psi(\lambda) - p_\varepsilon(\lambda)| \le \varepsilon.$$

As a consequence we obtain that

$$\begin{aligned}\int_0^1 \psi(\lambda)\, dF(\lambda) &= \int_0^1 p_\varepsilon(\lambda)\, dF(\lambda) + r_1(\varepsilon)\\ &= \int_0^1 p_\varepsilon(\lambda)\, dG(\lambda) + r_1(\varepsilon)\\ &= \int_0^1 \psi(\lambda)\, dG(\lambda) + r_2(\varepsilon),\end{aligned}$$

where $r_i(\varepsilon) \to 0$ as $\varepsilon \to 0$, $i = 1, 2$, and, thus,

$$\int_0^1 \psi(\lambda)\, dF(\lambda) = \int_0^1 \psi(\lambda)\, dG(\lambda).$$

Since ψ was an arbitrary continuous function, this in turn together with $F(0) = G(0) = 0$ implies $F = G$.

Suppose now that γ has the representation (4.3). We have for arbitrary $x_i \in \mathbb{R}$, $i = 1, \dots, n$

$$\begin{aligned}\sum_{1\le r,s\le n} x_r \gamma(r-s) x_s &= \int_0^1 \sum_{1\le r,s\le n} x_r x_s e^{i2\pi\lambda(r-s)}\, dF(\lambda)\\ &= \int_0^1 \Big| \sum_{r=1}^{n} x_r e^{i2\pi\lambda r} \Big|^2 dF(\lambda) \ge 0,\end{aligned}$$

i.e. γ is positive semidefinite.

Suppose conversely that $\gamma : \mathbb{Z} \to \mathbb{R}$ is a positive semidefinite function. This implies that for $0 \le \lambda \le 1$ and $N \in \mathbb{N}$ (Exercise 2)

$$\begin{aligned}f_N(\lambda) :&= \frac{1}{N} \sum_{1\le r,s\le N} e^{-i2\pi\lambda r} \gamma(r-s) e^{i2\pi\lambda s}\\ &= \frac{1}{N} \sum_{|m|<N} (N - |m|)\gamma(m) e^{-i2\pi\lambda m} \ge 0.\end{aligned}$$

Put now

$$F_N(\lambda) := \int_0^\lambda f_N(x)\, dx, \qquad 0 \le \lambda \le 1.$$

Then we have for each $h \in \mathbb{Z}$

$$\int_0^1 e^{i2\pi\lambda h}\, dF_N(\lambda) = \sum_{|m|<N} \Big(1 - \frac{|m|}{N}\Big)\gamma(m) \int_0^1 e^{i2\pi\lambda(h-m)}\, d\lambda$$
$$= \begin{cases} \Big(1 - \frac{|h|}{N}\Big)\gamma(h), & \text{if } |h| < N \\ 0, & \text{if } |h| \geq N. \end{cases} \tag{4.4}$$

Since $F_N(1) = \gamma(0) < \infty$ for any $N \in \mathbb{N}$, we can apply *Helly's selection theorem* (cf. Billingsley (1968), page 226ff) to deduce the existence of a measure generating function $\tilde{F}$ and a subsequence $(F_{N_k})_k$ such that F_{N_k} converges weakly to $\tilde{F}$ i.e.,

$$\int_0^1 g(\lambda)\, dF_{N_k}(\lambda) \longrightarrow_{k\to\infty} \int_0^1 g(\lambda)\, d\tilde{F}(\lambda)$$

for every continuous and bounded function $g : [0,1] \to \mathbb{R}$ (cf. Theorem 2.1 in Billingsley (1968)). Put now $F(\lambda) := \tilde{F}(\lambda) - \tilde{F}(0)$. Then F is a measure generating function with $F(0) = 0$ and

$$\int_0^1 g(\lambda)\, d\tilde{F}(\lambda) = \int_0^1 g(\lambda)\, dF(\lambda).$$

If we replace N in (4.4) by N_k and let k tend to infinity, we now obtain representation (4.3). □

Example 4.1.3. A white noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ has the autocovariance function

$$\gamma(h) = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \in \mathbb{Z} \setminus \{0\}. \end{cases}$$

Since

$$\int_0^1 \sigma^2 e^{i2\pi\lambda h}\, d\lambda = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \in \mathbb{Z} \setminus \{0\}, \end{cases}$$

the process (ε_t) has by Theorem 4.1.2 the constant spectral density $f(\lambda) = \sigma^2$, $0 \leq \lambda \leq 1$. This is the name giving property of the white noise process: As the white light is characteristically perceived to belong to objects that reflect nearly all incident energy throughout the visible spectrum, a white noise process weighs all possible frequencies equally.

Corollary 4.1.4. *A symmetric function $\gamma : \mathbb{Z} \to \mathbb{R}$ is the autocovariance function of a stationary process $(Y_t)_{t\in\mathbb{Z}}$, iff it satisfies one of the following two (equivalent) conditions:*

(i) $\gamma(h) = \int_0^1 e^{i2\pi\lambda h}\, dF(\lambda)$, $h \in \mathbb{Z}$, *where F is a measure generating function on $[0,1]$ with $F(0) = 0$.*

(ii) $\sum_{1\le r,s\le n} x_r\gamma(r-s)x_s \ge 0$ *for each* $n \in \mathbb{N}$ *and* $x_1, \ldots, x_n \in \mathbb{R}$.

Proof. Theorem 4.1.2 shows that (i) and (ii) are equivalent. The assertion is then a consequence of Theorem 4.1.1. □

Corollary 4.1.5. *A symmetric function* $\gamma : \mathbb{Z} \to \mathbb{R}$ *with* $\sum_{t\in\mathbb{Z}} |\gamma(t)| < \infty$ *is the autocovariance function of a stationary process iff*

$$f(\lambda) := \sum_{t\in\mathbb{Z}} \gamma(t)e^{-i2\pi\lambda t} \ge 0, \quad \lambda \in [0,1].$$

The function f *is in this case the spectral density of* γ.

Proof. Suppose first that γ is an autocovariance function. Since γ is in this case positive semidefinite by Theorem 4.1.1, and $\sum_{t\in\mathbb{Z}} |\gamma(t)| < \infty$ by assumption, we have (Exercise 2)

$$\begin{aligned} 0 \le f_N(\lambda) :&= \frac{1}{N} \sum_{1\le r,s\le N} e^{-i2\pi\lambda r}\gamma(r-s)e^{i2\pi\lambda s} \\ &= \sum_{|t|<N} \Big(1 - \frac{|t|}{N}\Big)\gamma(t)e^{-i2\pi\lambda t} \to f(\lambda) \quad \text{as } N \to \infty, \end{aligned}$$

see Exercise 8. The function f is consequently nonnegative. The inverse Fourier transform in Theorem 3.2.5 implies $\gamma(t) = \int_0^1 f(\lambda)e^{i2\pi\lambda t}\,d\lambda$, $t \in \mathbb{Z}$ i.e., f is the spectral density of γ.

Suppose on the other hand that $f(\lambda) = \sum_{t\in\mathbb{Z}} \gamma(t)e^{i2\pi\lambda t} \ge 0$, $0 \le \lambda \le 1$. The inverse Fourier transform implies $\gamma(t) = \int_0^1 f(\lambda)e^{i2\pi\lambda t}\,d\lambda = \int_0^1 e^{i2\pi\lambda t}\,dF(\lambda)$, where $F(\lambda) = \int_0^\lambda f(x)\,dx$, $0 \le \lambda \le 1$. Thus we have established representation (4.3), which implies that γ is positive semidefinite, and, consequently, γ is by Corollary 4.1.4 the autocovariance function of a stationary process. □

Example 4.1.6. Choose a number $\rho \in \mathbb{R}$. The function

$$\gamma(h) = \begin{cases} 1, & \text{if } h = 0 \\ \rho, & \text{if } h \in \{-1, 1\} \\ 0, & \text{elsewhere} \end{cases}$$

is the autocovariance function of a stationary process iff $|\rho| \le 0.5$. This follows

from

$$\begin{aligned} f(\lambda) &= \sum_{t\in\mathbb{Z}} \gamma(t) e^{i2\pi\lambda t} \\ &= \gamma(0) + \gamma(1)e^{i2\pi\lambda} + \gamma(-1)e^{-i2\pi\lambda} \\ &= 1 + 2\rho\cos(2\pi\lambda) \geq 0 \end{aligned}$$

for $\lambda \in [0,1]$ iff $|\rho| \leq 0.5$. Note that the function γ is the autocorrelation function of an $MA(1)$-process, cf. Example 2.2.2.

The spectral distribution function of a stationary process satisfies (Exercise 10)

$$F(0.5+\lambda) - F(0.5^-) = F(0.5) - F((0.5-\lambda)^-), \qquad 0 \leq \lambda < 0.5,$$

where $F(x^-) := \lim_{\varepsilon\downarrow 0} F(x-\varepsilon)$ is the left-hand limit of F at $x \in (0,1]$. If F has a derivative f, we obtain from the above symmetry $f(0.5+\lambda) = f(0.5-\lambda)$ or, equivalently, $f(1-\lambda) = f(\lambda)$ and, hence,

$$\gamma(h) = \int_0^1 \cos(2\pi\lambda h)\, dF(\lambda) = 2\int_0^{0.5} \cos(2\pi\lambda h) f(\lambda)\, d\lambda.$$

The autocovariance function of a stationary process is, therefore, determined by the values $f(\lambda)$, $0 \leq \lambda \leq 0.5$, if the spectral density exists. Recall, moreover, that the smallest nonconstant period P_0 visible through observations evaluated at time points $t = 1, 2, \ldots$ is $P_0 = 2$ i.e., the largest observable frequency is the Nyquist frequency $\lambda_0 = 1/P_0 = 0.5$, cf. the end of Section 3.2. Hence, the spectral density $f(\lambda)$ matters only for $\lambda \in [0, 0.5]$.

Remark 4.1.7. The preceding discussion shows that a function $f : [0,1] \to \mathbb{R}$ is the spectral density of a stationary process iff f satisfies the following three conditions

(i) $f(\lambda) \geq 0$,

(ii) $f(\lambda) = f(1-\lambda)$,

(iii) $\int_0^1 f(\lambda)\, d\lambda < \infty$.

4.2 Linear Filters and Frequencies

The application of a linear filter to a stationary time series has a quite complex effect on its autocovariance function, see Theorem 2.1.6. Its effect on the spectral density, if it exists, turns, however, out to be quite simple. We use in the following again the notation $\int_0^\lambda g(x)\, dF(x)$ in place of $\int_{(0,\lambda]} g(x)\, dF(x)$.

Theorem 4.2.1. *Let $(Z_t)_{t\in\mathbb{Z}}$ be a stationary process with spectral distribution function F_Z and let $(a_t)_{t\in\mathbb{Z}}$ be an absolutely summable filter with Fourier transform f_a. The linear filtered process $Y_t := \sum_{u\in\mathbb{Z}} a_u Z_{t-u}$, $t \in \mathbb{Z}$, then has the spectral distribution function*

$$F_Y(\lambda) := \int_0^\lambda |f_a(x)|^2 \, dF_Z(x), \quad 0 \le \lambda \le 1. \tag{4.5}$$

If in addition $(Z_t)_{t\in\mathbb{Z}}$ has a spectral density f_Z, then

$$f_Y(\lambda) := |f_a(\lambda)|^2 f_Z(\lambda), \qquad 0 \le \lambda \le 1, \tag{4.6}$$

is the spectral density of $(Y_t)_{t\in\mathbb{Z}}$.

Proof. Theorem 2.1.6 yields that $(Y_t)_{t\in\mathbb{Z}}$ is stationary with autocovariance function

$$\gamma_Y(t) = \sum_{u\in\mathbb{Z}} \sum_{w\in\mathbb{Z}} a_u a_w \gamma_Z(t-u+w), \quad t \in \mathbb{Z},$$

where γ_Z is the autocovariance function of (Z_t). Its spectral representation (4.3) implies

$$\begin{aligned}
\gamma_Y(t) &= \sum_{u\in\mathbb{Z}} \sum_{w\in\mathbb{Z}} a_u a_w \int_0^1 e^{i2\pi\lambda(t-u+w)} \, dF_Z(\lambda) \\
&= \int_0^1 \Big(\sum_{u\in\mathbb{Z}} a_u e^{-i2\pi\lambda u} \Big) \Big(\sum_{w\in\mathbb{Z}} a_w e^{i2\pi\lambda w} \Big) e^{i2\pi\lambda t} \, dF_Z(\lambda) \\
&= \int_0^1 |f_a(\lambda)|^2 e^{i2\pi\lambda t} \, dF_Z(\lambda) \\
&= \int_0^1 e^{i2\pi\lambda t} \, dF_Y(\lambda).
\end{aligned}$$

Theorem 4.1.2 now implies that F_Y is the uniquely determined spectral distribution function of $(Y_t)_{t\in\mathbb{Z}}$. The second to last equality yields in addition the spectral density (4.6). □

Transfer Function and Power Transfer Function

Since the spectral density is a measure of intensity of a frequency λ inherent in a stationary process (see the discussion of the periodogram in Section 3.2), the effect (4.6) of applying a linear filter (a_t) with Fourier transform f_a can easily be interpreted. While the intensity of λ is diminished by the filter (a_t) iff $|f_a(\lambda)| < 1$, its intensity is amplified iff $|f_a(\lambda)| > 1$. The Fourier transform f_a of (a_t) is, therefore, also called *transfer function* and the function $g_a(\lambda) := |f_a(\lambda)|^2$ is referred to as the *gain* or *power transfer function* of the filter $(a_t)_{t\in\mathbb{Z}}$.

Example 4.2.2. The simple moving average of length three

$$a_u = \begin{cases} 1/3, & u \in \{-1, 0, 1\} \\ 0 & \text{elsewhere} \end{cases}$$

has the transfer function

$$f_a(\lambda) = \frac{1}{3} + \frac{2}{3}\cos(2\pi\lambda)$$

and the power transfer function

$$g_a(\lambda) = \begin{cases} 1, & \lambda = 0 \\ \left(\frac{\sin(3\pi\lambda)}{3\sin(\pi\lambda)}\right)^2, & \lambda \in (0, 0.5] \end{cases}$$

(see Exercise 13 and Theorem 3.2.2). This power transfer function is plotted in Figure 4.2.1 below. It shows that frequencies λ close to zero i.e., those corresponding to a large period, remain essentially unaltered. Frequencies λ close to 0.5, which correspond to a short period, are, however, damped by the approximate factor 0.1, when the moving average (a_u) is applied to a process. The frequency $\lambda = 1/3$ is completely eliminated, since $g_a(1/3) = 0$.

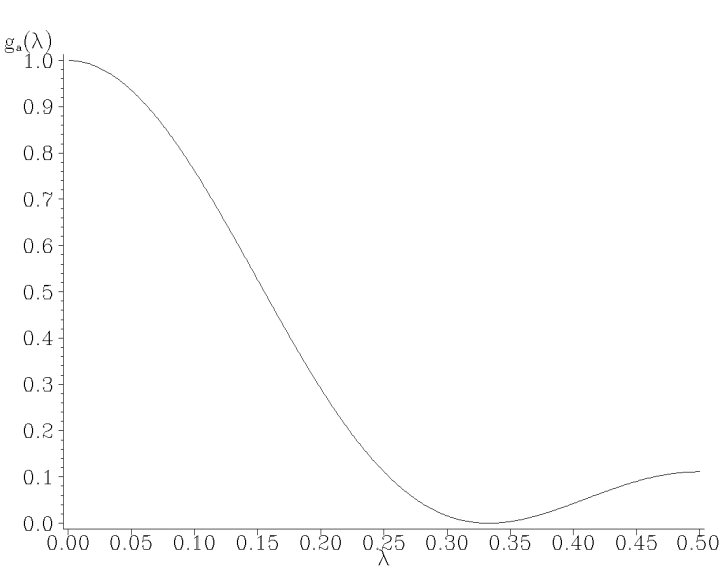


Figure 4.2.1. Power transfer function of the simple moving average of length three.

```
***     Program 4_2_1   ***;
TITLE1 'Power transfer function';
TITLE2 'of the simple moving average of length 3';

DATA data1;
  DO lambda=.001 TO .5 BY .001;
    g=(SIN(3*CONSTANT('PI')*lambda)/
        (3*SIN(CONSTANT('PI')*lambda)))**2;
    OUTPUT;
 END;

AXIS1 LABEL=('g' H=1 'a' H=2 '(' F=CGREEK 'l)');
AXIS2 LABEL=(F=CGREEK 'l');
SYMBOL1 V=NONE C=GREEN I=JOIN;
PROC GPLOT DATA=data1;
  PLOT g*lambda / VAXIS=AXIS1 HAXIS=AXIS2;
RUN; QUIT;
```

Example 4.2.3. The first order difference filter

$$a_u = \begin{cases} 1, & u = 0 \\ -1, & u = 1 \\ 0 & \text{elsewhere} \end{cases}$$

has the transfer function

$$f_a(\lambda) = 1 - e^{-i2\pi\lambda}.$$

Since

$$f_a(\lambda) = e^{-i\pi\lambda}\left(e^{i\pi\lambda} - e^{-i\pi\lambda}\right) = ie^{-i\pi\lambda}2\sin(\pi\lambda),$$

its power transfer function is

$$g_a(\lambda) = 4\sin^2(\pi\lambda).$$

The first order difference filter, therefore, damps frequencies close to zero but amplifies those close to 0.5.

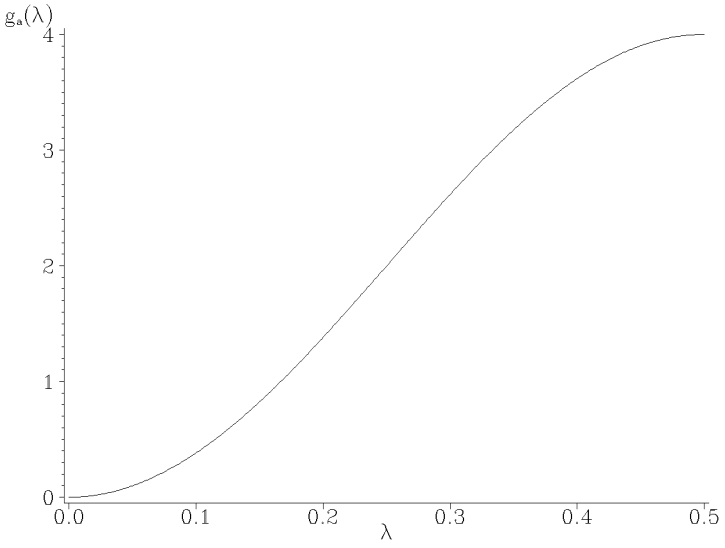


Figure 4.2.2. Power transfer function of the first order difference filter.

Example 4.2.4. The preceding example immediately carries over to the seasonal difference filter of arbitrary length $s \geq 0$ i.e.,

$$a_u^{(s)} = \begin{cases} 1, & u = 0 \\ -1, & u = s \\ 0 & \text{elsewhere,} \end{cases}$$

which has the transfer function

$$f_{a^{(s)}}(\lambda) = 1 - e^{-i2\pi\lambda s}$$

and the power transfer function

$$g_{a^{(s)}}(\lambda) = 4\sin^2(\pi\lambda s).$$

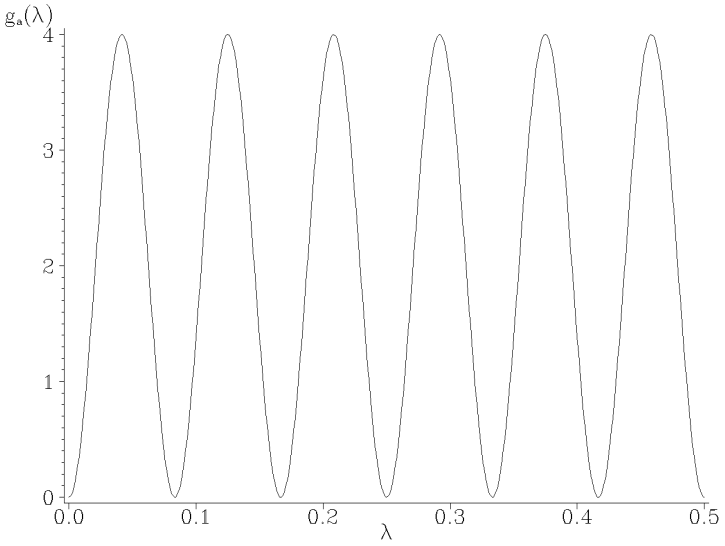


Figure 4.2.3. Power transfer function of the seasonal difference filter of order 12.

Since $\sin^2(x) = 0$ iff $x = k\pi$ and $\sin^2(x) = 1$ iff $x = (2k+1)\pi/2$, $k \in \mathbb{Z}$, the power transfer function $g_{a^{(s)}}(\lambda)$ satisfies for $k \in \mathbb{Z}$

$$g_{a^{(s)}}(\lambda) = \begin{cases} 0, & \text{iff } \lambda = k/s \\ 4 & \text{iff } \lambda = (2k+1)/(2s). \end{cases}$$

This implies, for example, in the case of $s = 12$ that those frequencies, which are multiples of $1/12 = 0.0833$, are eliminated, whereas the midpoint frequencies $k/12 + 1/24$ are amplified. This means that the seasonal difference filter on the one hand does what we would like it to do, namely to eliminate the frequency $1/12$, but on the other hand it generates unwanted side effects by eliminating also multiples of $1/12$ and by amplifying midpoint frequencies. This observation gives rise to the problem, whether one can construct linear filters that have prescribed properties.

Least Squares Based Filter Design

A *low pass filter* aims at eliminating high frequencies, a *high pass filter* aims at eliminating small frequencies and a *band pass filter* allows only frequencies in a certain band $[\lambda_0 - \Delta, \lambda_0 + \Delta]$ to pass through. They consequently should have the ideal power transfer functions

$$g_{low}(\lambda) = \begin{cases} 1, & \lambda \in [0, \lambda_0] \\ 0, & \lambda \in (\lambda_0, 0.5] \end{cases}$$

$$g_{high}(\lambda) = \begin{cases} 0, & \lambda \in [0, \lambda_0) \\ 1, & \lambda \in [\lambda_0, 0.5] \end{cases}$$

$$g_{band}(\lambda) = \begin{cases} 1, & \lambda \in [\lambda_0 - \Delta, \lambda_0 + \Delta] \\ 0 & \text{elsewhere,} \end{cases}$$

where λ_0 is the cut off frequency in the first two cases and $[\lambda_0 - \Delta, \lambda_0 + \Delta]$ is the cut off interval with bandwidth $2\Delta > 0$ in the final one. Therefore, the question naturally arises, whether there actually exist filters, which have a prescribed power transfer function. One possible approach for fitting a linear filter with weights a_u to a given transfer function f is offered by utilizing least squares. Since only filters of finite length matter in applications, one chooses a transfer function

$$f_a(\lambda) = \sum_{u=r}^{s} a_u e^{-i2\pi\lambda u}$$

with fixed integers r, s and fits this function f_a to f by minimizing the integrated squared error

$$\int_0^{0.5} |f(\lambda) - f_a(\lambda)|^2 \, d\lambda$$

in $(a_u)_{r \le u \le s} \in \mathbb{R}^{s-r+1}$. This is achieved for the choice (Exercise 16)

$$a_u = 2 \operatorname{Re}\Big(\int_0^{0.5} f(\lambda) e^{i2\pi\lambda u} \, d\lambda \Big), \quad u = r, \dots, s,$$

which is formally the real part of the inverse Fourier transform of f.

Example 4.2.5. For the low pass filter with cut off frequency $0 < \lambda_0 < 0.5$ and ideal transfer function

$$f(\lambda) = 1_{[0,\lambda_0]}(\lambda)$$

we obtain the weights

$$a_u = 2 \int_0^{\lambda_0} \cos(2\pi\lambda u) \, d\lambda = \begin{cases} 2\lambda_0, & u = 0 \\ \frac{1}{\pi u} \sin(2\pi\lambda_0 u), & u \neq 0. \end{cases}$$

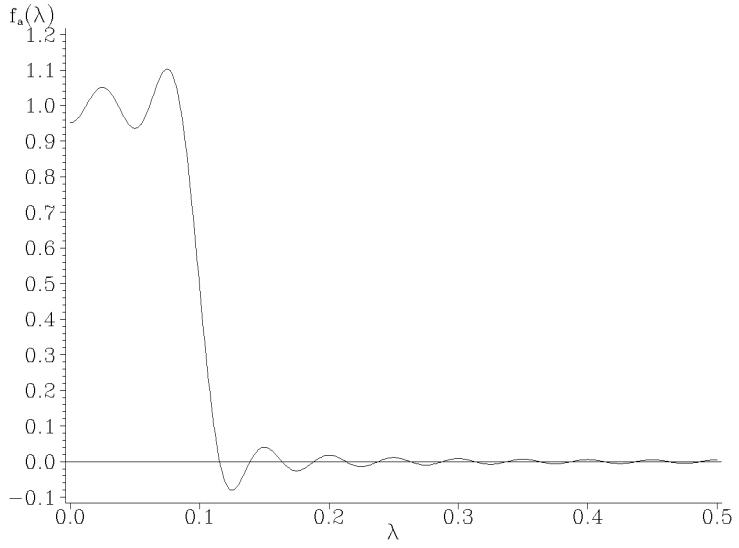


Figure 4.2.4. Transfer function of least squares fitted low pass filter with cut off frequency $\lambda_0 = 1/10$ and $r = -20$, $s = 20$.

```
***     Program 4_2_4   ***;
TITLE1 'Transfer function';
TITLE2 'of least squares fitted low low pass filter';

DATA data1;
 DO lambda=0 TO .5 BY .001;
   f=2*1/10;
   DO u=1 TO 20;
     f=f+2*1/(CONSTANT('PI')*u)*SIN(2*CONSTANT('PI')*
                1/10*u)*COS(2*CONSTANT('PI')*lambda*u);
   END;
   OUTPUT;
 END;

AXIS1 LABEL=('f' H=1 'a' H=2 F=CGREEK '(l)');
AXIS2 LABEL=(F=CGREEK 'l');
SYMBOL1 V=NONE C=GREEN I=JOIN L=1;
PROC GPLOT DATA=data1;
  PLOT f*lambda / VAXIS=AXIS1 HAXIS=AXIS2 VREF=0;
```

```
RUN; QUIT;
```

The programs in Section 4.2 are just made for the purpose of generating graphics, which demonstrate the shape of power transfer functions or, in case of Program 4_2_4, of a transfer function. They all consist of two parts, a `DATA` step and a `PROC` step.

In the `DATA` step values of the power transfer function are calculated and stored in a variable `g` by a `DO` loop over `lambda` from 0 to 0.5 with a small increment. In Program 4_2_4 it is necessary to use a second `DO` loop within the first one to calculate the sum used in the definition of the transfer function `f`.

Two `AXIS` statements defining the axis labels and a `SYMBOL` statement precede the procedure `PLOT`, which generates the plot of `g` or `f` versus `lambda`.

The transfer function in Figure 4.2.4, which approximates the ideal transfer function $1_{[0,0.1]}$, shows higher oscillations near the cut off point $\lambda_0 = 0.1$. This is known as *Gibbs' phenomenon* and requires further smoothing of the fitted transfer function if these oscillations are to be damped (cf. Section 6.4 of Bloomfield (1976)).

4.3 Spectral Densities of ARMA-Processes

Theorem 4.2.1 enables us to compute the spectral density of an *ARMA*-process.

Theorem 4.3.1. *Suppose that*

$$Y_t = a_1 Y_{t-1} + \cdots + a_p Y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}, \quad t \in \mathbb{Z},$$

is a stationary ARMA(p,q)*-process, where* (ε_t) *is a white noise with variance* σ^2*. Put*

$$A(z) := 1 - a_1 z - a_2 z^2 - \cdots - a_p z^p,$$
$$B(z) := 1 + b_1 z + b_2 z^2 + \cdots + b_q z^q$$

and suppose that the process (Y_t) *satisfies the stationarity condition (2.3), i.e., the roots of the equation* $A(z) = 0$ *are outside of the unit circle. The process* (Y_t) *then has the spectral density*

$$f_Y(\lambda) = \sigma^2 \frac{|B(e^{-i2\pi\lambda})|^2}{|A(e^{-i2\pi\lambda})|^2} = \sigma^2 \frac{|1 + \sum_{v=1}^{q} b_v e^{-i2\pi\lambda v}|^2}{|1 - \sum_{u=1}^{p} a_u e^{-i2\pi\lambda u}|^2}. \tag{4.7}$$

Proof. Since the process (Y_t) is supposed to satisfy the stationarity condition (2.3) it is causal, i.e., $Y_t = \sum_{v\geq 0} \alpha_v \varepsilon_{t-v}$, $t \in \mathbb{Z}$, for some absolutely summable constants α_v, $v \geq 0$, see Section 2.2. The white noise process (ε_t) has by Example 4.1.3 the spectral density $f_\varepsilon(\lambda) = \sigma^2$ and, thus, (Y_t) has by Theorem 4.2.1 a spectral density f_Y. The application of Theorem 4.2.1 to the process

$$X_t := Y_t - a_1 Y_{t-1} - \cdots - a_p Y_{t-p} = \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}$$

then implies that (X_t) has the spectral density

$$f_X(\lambda) = |A(e^{-i2\pi\lambda})|^2 f_Y(\lambda) = |B(e^{-i2\pi\lambda})|^2 f_\varepsilon(\lambda).$$

Since the roots of $A(z) = 0$ are assumed to be outside of the unit circle, we have $|A(e^{-i2\pi\lambda})| \neq 0$ and, thus, the assertion of Theorem 4.3.1 follows. □

The preceding result with $a_1 = \cdots = a_p = 0$ implies that an $MA(q)$-process has the spectral density

$$f_Y(\lambda) = \sigma^2 |B(e^{-i2\pi\lambda})|^2.$$

With $b_1 = \cdots = b_q = 0$, Theorem 4.3.1 implies that a stationary $AR(p)$-process, which satisfies the stationarity condition (2.3), has the spectral density

$$f_Y(\lambda) = \sigma^2 \frac{1}{|A(e^{-i2\pi\lambda})|^2}.$$

Example 4.3.2. The stationary $ARMA(1,1)$-process

$$Y_t = aY_{t-1} + \varepsilon_t + b\varepsilon_{t-1}$$

with $|a| < 1$ has the spectral density

$$f_Y(\lambda) = \sigma^2 \frac{1 + 2b\cos(2\pi\lambda) + b^2}{1 - 2a\cos(2\pi\lambda) + a^2}.$$

The $MA(1)$-process, in which case $a = 0$, has, consequently the spectral density

$$f_Y(\lambda) = \sigma^2 (1 + 2b\cos(2\pi\lambda) + b^2),$$

and the stationary $AR(1)$-process with $|a| < 1$, for which $b = 0$, has the spectral density

$$f_Y(\lambda) = \sigma^2 \frac{1}{1 - 2a\cos(2\pi\lambda) + a^2}.$$

The following figures display spectral densities of $ARMA(1,1)$-processes for various a and b with $\sigma^2 = 1$.

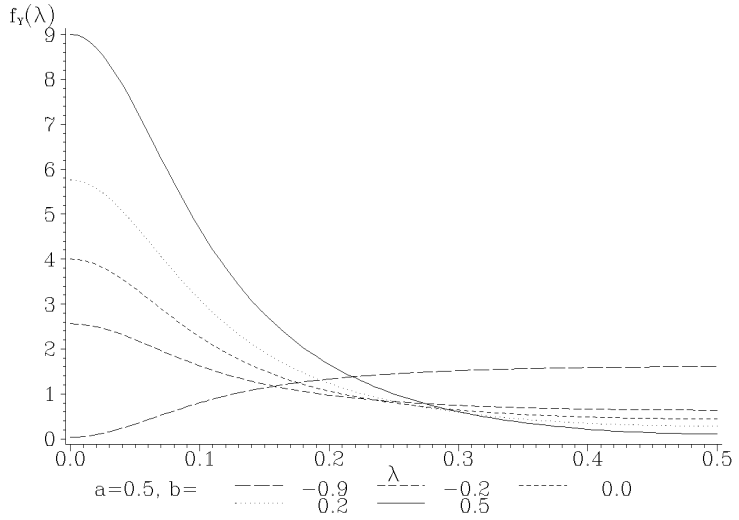


Figure 4.3.1. Spectral densities of $ARMA(1,1)$-processes $Y_t = aY_{t-1} + \varepsilon_t + b\varepsilon_{t-1}$ with fixed a and various b; $\sigma^2 = 1$.

```
***     Program 4_3_1***;
TITLE1 'Spectral densities of ARMA(1,1)-processes';

DATA data1;
 a=.5;
 DO b=-.9, -.2, 0, .2, .5;
   DO lambda=0 TO .5 BY .005;
     f=(1+2*b*COS(2*CONSTANT('PI')*lambda)+b*b)
            /(1-2*a*COS(2*CONSTANT('PI')*lambda)+a*a);
     OUTPUT;
   END;
 END;

AXIS1 LABEL=('f' H=1 'Y' H=2 F=CGREEK '(l)');
AXIS2 LABEL=(F=CGREEK 'l');
SYMBOL1 V=NONE C=GREEN I=JOIN L=4;
SYMBOL2 V=NONE C=GREEN I=JOIN L=3;
SYMBOL3 V=NONE C=GREEN I=JOIN L=2;
SYMBOL4 V=NONE C=GREEN I=JOIN L=33;
```

```
SYMBOL5 V=NONE C=GREEN I=JOIN L=1;
LEGEND1 LABEL=('a=0.5, b=');
PROC GPLOT DATA=data1;
  PLOT f*lambda=b / VAXIS=AXIS1 HAXIS=AXIS2
                    LEGEND=LEGEND1;
RUN; QUIT;
```

Like in the preceding section the programs here just generate graphics. In the `DATA` step some loops over the varying parameter and over `lambda` are used to calculate the values of the spectral densities of the corresponding processes. Here `SYMBOL` statements are necessary and a `LABEL` statement to distinguish the different curves generated by `PROC GPLOT`.

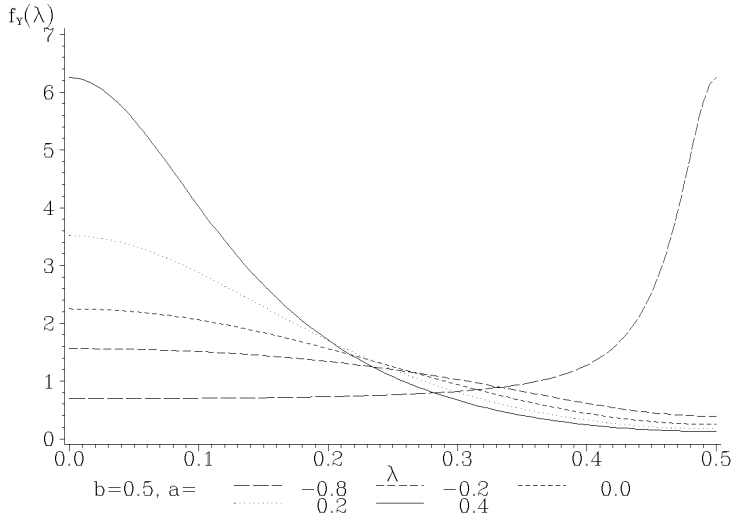


Figure 4.3.2. Spectral densities of $ARMA(1,1)$-processes with parameter b fixed and various a.

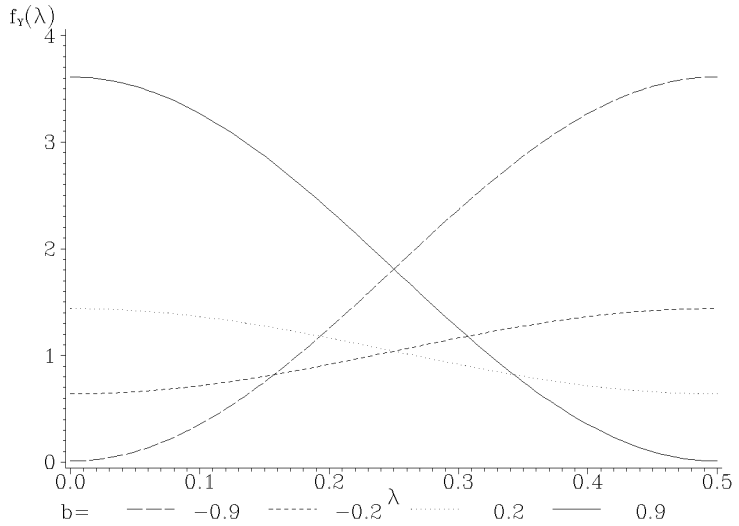


Figure 4.3.3. Spectral densities of $MA(1)$-processes $Y_t = \varepsilon_t + b\varepsilon_{t-1}$ for various b.

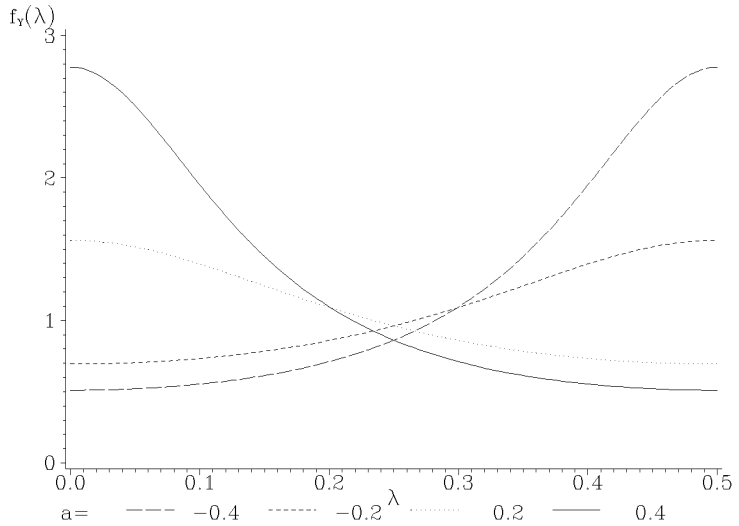


Figure 4.3.4. Spectral densities of $AR(1)$-processes $Y_t = aY_{t-1} + \varepsilon_t$ for various a.

Exercises

1. Formulate and prove Theorem 4.1.1 for Hermitian functions K and complex-valued stationary processes. Hint for the sufficiency part (for the necessary part see Exercise 1): Let K_1 be the real part and K_2 be the imaginary part of K. Consider the real-valued $2n \times 2n$-matrices

$$M^{(n)} = \frac{1}{2}\begin{pmatrix} K_1^{(n)} & K_2^{(n)} \\ -K_2^{(n)} & K_1^{(n)} \end{pmatrix}, \quad K_l^{(n)} = \big(K_l(r-s)\big)_{1\le r,s\le n}, \quad l = 1, 2.$$

Then $M^{(n)}$ is a positive semidefinite matrix (check that $z^T K \bar{z} = (x,y)^T M^{(n)}(x,y)$, $z = x+iy$, $x, y \in \mathbb{R}^n$). Proceed as in the proof of Theorem 4.1.1: Let $(V_1, \ldots, V_n, W_1, \ldots, W_n)$ be a $2n$-dimensional normal distributed random vector with mean vector zero and covariance matrix $M^{(n)}$ and define for $n \in \mathbb{N}$ the family of finite dimensional distributions

$$F_{t+1,\ldots,t+n}(v_1, w_1, \ldots, v_n, w_n) := P\{V_1 \le v_1, W_1 \le w_1, \ldots, V_n \le v_n, W_n \le w_n\},$$

$t \in \mathbb{Z}$. By Kolmogorov's theorem there exists a bivariate Gaussian process $(V_t, W_t)_{t\in\mathbb{Z}}$ with mean vector zero and covariances

$$\begin{aligned} \mathrm{E}(V_{t+h}V_t) = \mathrm{E}(W_{t+h}W_t) &= \frac{1}{2}K_1(h) \\ \mathrm{E}(V_{t+h}W_t) = -\mathrm{E}(W_{t+h}V_t) &= \frac{1}{2}K_2(h). \end{aligned}$$

Conclude by showing that the complex-valued process $Y_t := V_t - iW_t$, $t \in \mathbb{Z}$, has the autocovariance function K.

2. Suppose that $\boldsymbol{A}$ is a real positive semidefinite $n \times n$-matrix i.e., $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \ge 0$ for $\boldsymbol{x} \in \mathbb{R}^n$. Show that $\boldsymbol{A}$ is also positive semidefinite for complex numbers i.e., $\boldsymbol{z}^T\boldsymbol{A}\bar{\boldsymbol{z}} \ge 0$ for $\boldsymbol{z} \in \mathbb{C}^n$.

3. Use (4.3) to show that for $0 < a < 0.5$

$$\gamma(h) = \begin{cases} \frac{\sin(2\pi a h)}{2\pi h}, & h \in \mathbb{Z} \setminus \{0\} \\ a, & h = 0 \end{cases}$$

is the autocovariance function of a stationary process. Compute its spectral density.

4. Compute the autocovariance function of a stationary process with spectral density

$$f(\lambda) = \frac{0.5 - |\lambda - 0.5|}{0.5^2}, \quad 0 \le \lambda \le 1.$$

5. Suppose that F and G are measure generating functions defined on some interval $[a, b]$ with $F(a) = G(a) = 0$ and

$$\int_{[a,b]} \psi(x)\, F(dx) = \int_{[a,b]} \psi(x)\, G(dx)$$

for every continuous function $\psi : [a,b] \to \mathbb{R}$. Show that F=G. Hint: Approximate the indicator function $1_{[a,t]}(x)$, $x \in [a,b]$, by continuous functions.

6. Generate a white noise process and plot its periodogram.

7. A real valued stationary process $(Y_t)_{t\in\mathbb{Z}}$ is supposed to have the spectral density $f(\lambda) = a + b\lambda$, $\lambda \in [0, 0.5]$. Which conditions must be satisfied by a and b? Compute the autocovariance function of $(Y_t)_{t\in\mathbb{Z}}$.

8. (Césaro convergence) Show that $\lim_{N\to\infty} \sum_{t=1}^{N} a_t = S$ implies $\lim_{N\to\infty} \sum_{t=1}^{N-1}(1 - t/N)a_t = S$. Hint: $\sum_{t=1}^{N-1}(1 - t/N)a_t = (1/N)\sum_{s=1}^{N-1}\sum_{t=1}^{s} a_t$.

9. Suppose that $(Y_t)_{t\in\mathbb{Z}}$ and $(Z_t)_{t\in\mathbb{Z}}$ are stationary processes such that Y_r and Z_s are uncorrelated for arbitrary $r, s \in \mathbb{Z}$. Denote by F_Y and F_Z the pertaining spectral distribution functions and put $X_t := Y_t + Z_t$, $t \in \mathbb{Z}$. Show that the process (X_t) is also stationary and compute its spectral distribution function.

10. Let (Y_t) be a real valued stationary process with spectral distribution function F. Show that for any function $g : [-0.5, 0.5] \to \mathbb{C}$ with $\int_0^1 |g(\lambda - 0.5)|^2 \, dF(\lambda) < \infty$

$$\int_0^1 g(\lambda - 0.5)\, dF(\lambda) = \int_0^1 g(0.5 - \lambda)\, dF(\lambda).$$

In particular we have

$$F(0.5 + \lambda) - F(0.5^-) = F(0.5) - F((0.5 - \lambda)^-).$$

Hint: Verify the equality first for $g(x) = \exp(i2\pi hx)$, $h \in \mathbb{Z}$, and then use the fact that, on compact sets, the trigonometric polynomials are uniformly dense in the space of continuous functions, which in turn form a dense subset in the space of square integrable functions. Finally, consider the function $g(x) = 1_{[0,\xi]}(x)$, $0 \le \xi \le 0.5$ (cf. the hint in Exercise 5).

11. Let (X_t) and (Y_t) be stationary processes with mean zero and absolute summable covariance functions. If their spectral densities f_X and f_Y satisfy $f_X(\lambda) \le f_Y(\lambda)$ for $0 \le \lambda \le 1$, show that

(i) $\Gamma_{n,Y} - \Gamma_{n,X}$ is a positive semidefinite matrix, where $\Gamma_{n,X}$ and $\Gamma_{n,Y}$ are the covariance matrices of $(X_1, \ldots, X_n)^T$ and $(Y_1, \ldots, Y_n)^T$ respectively, and

(ii) $\mathrm{Var}(a^T(X_1, \ldots, X_n)) \le \mathrm{Var}(a^T(Y_1, \ldots, Y_n))$ for all $a = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$.

12. Compute the gain function of the filter

$$a_u = \begin{cases} 1/4, & u \in \{-1, 1\} \\ 1/2, & u = 0 \\ 0 & \text{elsewhere.} \end{cases}$$

13. The simple moving average

$$a_u = \begin{cases} 1/(2q+1), & u \le |q| \\ 0 & \text{elsewhere} \end{cases}$$

has the gain function

$$g_a(\lambda) = \begin{cases} 1, & \lambda = 0 \\ \left(\frac{\sin((2q+1)\pi\lambda)}{(2q+1)\sin(\pi\lambda)} \right)^2, & \lambda \in (0, 0.5]. \end{cases}$$

Is this filter for large q a low pass filter? Plot its power transfer functions for $q = 5/10/20$. Hint: Exercise 2 in Chapter 3.

14. Compute the gain function of the exponential smoothing filter

$$a_u = \begin{cases} \alpha(1-\alpha)^u, & u \ge 0 \\ 0, & u < 0, \end{cases}$$

where $0 < \alpha < 1$. Plot this function for various α. What is the effect of $\alpha \to 0$ or $\alpha \to 1$?

15. Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary process, $(a_u)_{u\in\mathbb{Z}}$ an absolutely summable filter and put $Y_t := \sum_{u\in\mathbb{Z}} a_u X_{t-u}$, $t \in \mathbb{Z}$. If $(b_w)_{w\in\mathbb{Z}}$ is another absolutely summable filter, then the process $Z_t = \sum_{w\in\mathbb{Z}} b_w Y_{t-w}$ has the spectral distribution function

$$F_Z(\lambda) = \int_0^\lambda |F_a(\mu)|^2 |F_b(\mu)|^2 \, dF_X(\mu)$$

(cf. Exercise 11(iii) in Chapter 3).

16. Show that the function

$$D(a_r, \ldots, a_s) = \int_0^{0.5} |f(\lambda) - f_a(\lambda)|^2 \, d\lambda$$

with $f_a(\lambda) = \sum_{u=r}^s a_u e^{-i2\pi\lambda u}$, $a_u \in \mathbb{R}$, is minimized for

$$a_u := 2 \operatorname{Re}\Big(\int_0^{0.5} f(\lambda) e^{i2\pi\lambda u} \, d\lambda \Big), \quad u = r, \ldots, s.$$

Hint: Put $f(\lambda) = f_1(\lambda) + i f_2(\lambda)$ and differentiate with respect to a_u.

17. Compute in analogy to Example 4.2.5 the transfer functions of the least squares high pass and band pass filter. Plot these functions. Is Gibbs' phenomenon observable?

18. An $AR(2)$-process

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t$$

satisfying the stationarity condition (2.3) (cf. Exercise 23 in Chapter 2) has the spectral density

$$f_Y(\lambda) = \frac{\sigma^2}{1 + a_1^2 + a_2^2 + 2(a_1 a_2 - a_1)\cos(2\pi\lambda) - 2a_2\cos(4\pi\lambda)}.$$

Plot this function for various choices of a_1, a_2.

19. Show that (4.2) is a necessary condition for K to be the autocovariance function of a stationary process.

Chapter 5

Statistical Analysis in the Frequency Domain

We will deal in this chapter with the problem of testing for a white noise and the estimation of the spectral density. The empirical counterpart of the spectral density, the periodogram, will be basic for both problems, though it will turn out that it is not a consistent estimate. Consistency requires extra smoothing of the periodogram via a linear filter.

5.1 Testing for a White Noise

Our initial step in a statistical analysis of a time series in the frequency domain is to test, whether the data are generated by a white noise $(\varepsilon_t)_{t\in\mathbb{Z}}$. We start with the model

$$Y_t = \mu + A\cos(2\pi\lambda t) + B\sin(2\pi\lambda t) + \varepsilon_t,$$

where we assume that the ε_t are independent and normal distributed with mean zero and variance σ^2. We will test the nullhypothesis

$$A = B = 0$$

against the alternative

$$A \neq 0 \quad \text{or} \quad B \neq 0,$$

where the frequency λ, the variance $\sigma^2 > 0$ and the intercept $\mu \in \mathbb{R}$ are unknown. Since the periodogram is used for the detection of highly intensive frequencies

inherent in the data, it seems plausible to apply it to the preceding testing problem as well. Note that $(Y_t)_{t\in\mathbb{Z}}$ is a stationary process only under the nullhypothesis $A = B = 0$.

The Distribution of the Periodogram

In the following we will derive the tests by Fisher and Bartlett–Kolmogorov–Smirnov for the above testing problem. In a preparatory step we compute the distribution of the periodogram.

Lemma 5.1.1. *Let $\varepsilon_1, \dots, \varepsilon_n$ be independent and identically normal distributed random variables with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. Denote by $\bar{\varepsilon} := n^{-1}\sum_{t=1}^n \varepsilon_t$ the sample mean of $\varepsilon_1, \dots, \varepsilon_n$ and by*

$$C_\varepsilon\left(\frac{k}{n}\right) = \frac{1}{n}\sum_{t=1}^n (\varepsilon_t - \bar{\varepsilon})\cos\left(2\pi\frac{k}{n}t\right),$$
$$S_\varepsilon\left(\frac{k}{n}\right) = \frac{1}{n}\sum_{t=1}^n (\varepsilon_t - \bar{\varepsilon})\sin\left(2\pi\frac{k}{n}t\right)$$

the cross covariances with Fourier frequencies k/n, $1 \le k \le [(n-1)/2]$, cf. (3.1). Then the $2[(n-1)/2]$ random variables

$$C_\varepsilon(k/n),\ S_\varepsilon(k/n), \quad 1 \le k \le [(n-1)/2],$$

are independent and identically $N(0, \sigma^2/(2n))$-distributed.

Proof. Note that with $m := [(n-1)/2]$ we have

$$\begin{aligned}\boldsymbol{v} &:= \big(C_\varepsilon(1/n), S_\varepsilon(1/n), \dots, C_\varepsilon(m/n), S_\varepsilon(m/n)\big)^T \\ &= \boldsymbol{A} n^{-1}(\varepsilon_t - \bar{\varepsilon})_{1\le t\le n} \\ &= \boldsymbol{A}(\boldsymbol{I}_n - n^{-1}\boldsymbol{E}_n)(\varepsilon_t)_{1\le t\le n},\end{aligned}$$

where the $2m \times n$-matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} := \frac{1}{n}\begin{pmatrix} \cos\left(2\pi\frac{1}{n}\right) & \cos\left(2\pi\frac{1}{n}2\right) & \dots & \cos\left(2\pi\frac{1}{n}n\right) \\ \sin\left(2\pi\frac{1}{n}\right) & \sin\left(2\pi\frac{1}{n}2\right) & \dots & \sin\left(2\pi\frac{1}{n}n\right) \\ \vdots & & & \vdots \\ \cos\left(2\pi\frac{m}{n}\right) & \cos\left(2\pi\frac{m}{n}2\right) & \dots & \cos\left(2\pi\frac{m}{n}n\right) \\ \sin\left(2\pi\frac{m}{n}\right) & \sin\left(2\pi\frac{m}{n}2\right) & \dots & \sin\left(2\pi\frac{m}{n}n\right) \end{pmatrix}.$$

$\boldsymbol{I}_n$ is the $n \times n$-unity matrix and $\boldsymbol{E}_n$ is the $n \times n$-matrix with each entry being 1. The vector $\boldsymbol{v}$ is, therefore, normal distributed with mean vector zero and covariance matrix

$$\begin{aligned}
&\sigma^2 \boldsymbol{A}(\boldsymbol{I}_n - n^{-1}\boldsymbol{E}_n)(\boldsymbol{I}_n - n^{-1}\boldsymbol{E}_n)^T \boldsymbol{A}^T \\
&= \sigma^2 \boldsymbol{A}(\boldsymbol{I}_n - n^{-1}\boldsymbol{E}_n)\boldsymbol{A}^T \\
&= \sigma^2 \boldsymbol{A}\boldsymbol{A}^T \\
&= \frac{\sigma^2}{2n}\boldsymbol{I}_n,
\end{aligned}$$

which is a consequence of (3.6) and the orthogonality properties (3.2); see e.g. Definition 2.1.2 in Falk et al. (2002). □

Corollary 5.1.2. *Let $\varepsilon_1, \ldots, \varepsilon_n$ be as in the preceding lemma and let*

$$I_\varepsilon(k/n) = n\Big\{C_\varepsilon^2(k/n) + S_\varepsilon^2(k/n)\Big\}$$

be the pertaining periodogram, evaluated at the Fourier frequencies k/n, $1 \le k \le [(n-1)/2]$. The random variables $I_\varepsilon(k/n)/\sigma^2$ are independent and identically standard exponential distributed i.e.,

$$P\{I_\varepsilon(k/n)/\sigma^2 \le x\} = \begin{cases} 1 - \exp(-x), & x > 0 \\ 0, & x \le 0. \end{cases}$$

Proof. Lemma 5.1.1 implies that

$$\sqrt{\frac{2n}{\sigma^2}} C_\varepsilon(k/n),\ \sqrt{\frac{2n}{\sigma^2}} S_\varepsilon(k/n)$$

are independent standard normal random variables and, thus,

$$\frac{2I_\varepsilon(k/n)}{\sigma^2} = \left(\sqrt{\frac{2n}{\sigma^2}} C_\varepsilon(k/n)\right)^2 + \left(\sqrt{\frac{2n}{\sigma^2}} S_\varepsilon(k/n)\right)^2$$

is χ^2-distributed with two degrees of freedom. Since this distribution has the distribution function $1 - \exp(-x/2)$, $x \ge 0$, the assertion follows; see e.g. Theorem 2.1.7 in Falk et al. (2002). □

Denote by $U_{1:m} \le U_{2:m} \le \cdots \le U_{m:m}$ the ordered values pertaining to independent and uniformly on $(0,1)$ distributed random variables $U_1, \ldots, U_m$. It is a well known result in the theory of order statistics that the distribution of the vector $(U_{j:m})_{1 \le j \le m}$ coincides with that of $((Z_1 + \cdots + Z_j)/(Z_1 + \cdots + Z_{m+1}))_{1 \le j \le m}$,

where $Z_1, \ldots, Z_{m+1}$ are independent and identically exponential distributed random variables; see, for example, Theorem 1.6.7 in Reiss (1989). The following result, which will be basic for our further considerations, is, therefore, an immediate consequence of Corollary 5.1.2; see also Exercise 3. By $=_D$ we denote equality in distribution.

Theorem 5.1.3. *Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent $N(\mu, \sigma^2)$-distributed random variables and denote by*

$$S_j := \frac{\sum_{k=1}^{j} I_\varepsilon(k/n)}{\sum_{k=1}^{m} I_\varepsilon(k/n)}, \quad j = 1, \ldots, m := [(n-1)/2],$$

the cumulated periodogram. Note that $S_m = 1$. Then we have

$$\big(S_1, \ldots, S_{m-1}\big) =_D \big(U_{1:m-1}, \ldots, U_{m-1:m-1}\big).$$

The vector $(S_1, \ldots, S_{m-1})$ has, therefore, the Lebesgue-density

$$f(s_1, \ldots, s_{m-1}) = \begin{cases} (m-1)!, & \text{if } 0 < s_1 < \cdots < s_{m-1} < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

The following consequence of the preceding result is obvious.

Corollary 5.1.4. *The empirical distribution function of $S_1, \ldots, S_{m-1}$ is distributed like that of $U_1, \ldots, U_{m-1}$, i.e.,*

$$\hat{F}_{m-1}(x) := \frac{1}{m-1} \sum_{j=1}^{m-1} 1_{(0,x]}(S_j) =_D \frac{1}{m-1} \sum_{j=1}^{m-1} 1_{(0,x]}(U_j), \quad x \in [0,1].$$

Corollary 5.1.5. *Put $S_0 := 0$ and*

$$M_m := \max_{1 \le j \le m} (S_j - S_{j-1}) = \frac{\max_{1 \le j \le m} I_\varepsilon(j/n)}{\sum_{k=1}^{m} I_\varepsilon(k/n)}.$$

The maximum spacing M_m has the distribution function

$$G_m(x) := P\{M_m \le x\} = \sum_{j=0}^{m} (-1)^j \binom{m}{j} (\max\{0, 1 - jx\})^{m-1}, \quad x > 0.$$

Proof. Put

$$V_j := S_j - S_{j-1}, \quad j = 1, \ldots, m.$$

By Theorem 5.1.3 the vector $(V_1, \ldots, V_m)$ is distributed like the length of the m consecutive intervals into which $[0,1]$ is partitioned by the $m-1$ random points $U_1, \ldots, U_{m-1}$:

$$(V_1, \ldots, V_m) =_D (U_{1:m-1}, U_{2:m-1} - U_{1:m-1}, \ldots, 1 - U_{m-1:m-1}).$$

The probability that M_m is less than or equal to x equals the probability that all spacings V_j are less than or equal to x, and this is provided by the *covering theorem* as stated in Theorem 3 in Section I.9 of Feller (1971). □

Fisher's Test

The preceding results suggest to test the hypothesis $Y_t = \varepsilon_t$ with ε_t independent and $N(\mu, \sigma^2)$-distributed, by testing for the uniform distribution on $[0, 1]$. Precisely, we will reject this hypothesis if *Fisher's* κ-statistic

$$\kappa_m := \frac{\max_{1 \le j \le m} I(j/n)}{(1/m) \sum_{k=1}^{m} I(k/n)} = m M_m$$

is significantly large, i.e., if one of the values $I(j/n)$ is significantly larger than the average over all. The hypothesis is, therefore, rejected at error level α if

$$\kappa_m > c_\alpha \quad \text{with } 1 - G_m(c_\alpha) = \alpha.$$

This is Fisher's test for hidden periodicities. Common values are $\alpha = 0.01$ and $= 0.05$. The following table, taken from Fuller (1976), lists several critical values c_α.

m	$c_{0.05}$	$c_{0.01}$	m	$c_{0.05}$	$c_{0.01}$
10	4.450	5.358	150	7.832	9.372
15	5.019	6.103	200	8.147	9.707
20	5.408	6.594	250	8.389	9.960
25	5.701	6.955	300	8.584	10.164
30	5.935	7.237	350	8.748	10.334
40	6.295	7.663	400	8.889	10.480
50	6.567	7.977	500	9.123	10.721
60	6.785	8.225	600	9.313	10.916
70	6.967	8.428	700	9.473	11.079
80	7.122	8.601	800	9.612	11.220
90	7.258	8.750	900	9.733	11.344
100	7.378	8.882	1000	9.842	11.454

Table 5.1.1. Critical values c_α of Fisher's test for hidden periodicities.

Note that these quantiles can be approximated by corresponding quantiles of a Gumbel distribution if m is large (Exercise 12).

The Bartlett–Kolmogorov–Smirnov Test

Denote again by S_j the cumulated periodogram as in Theorem 5.1.3. If actually $Y_t = \varepsilon_t$ with ε_t independent and identically $N(\mu, \sigma^2)$-distributed, then we know

from Corollary 5.1.4 that the empirical distribution function $\hat{F}_{m-1}$ of $S_1, \ldots, S_{m-1}$ behaves stochastically exactly like that of $m-1$ independent and uniformly on $(0,1)$ distributed random variables. Therefore, with the Kolmogorov–Smirnov statistic

$$\Delta_{m-1} := \sup_{x \in [0,1]} |\hat{F}_{m-1}(x) - x|$$

we can measure the maximum difference between the empirical distribution function and the theoretical one $F(x) = x$, $x \in [0,1]$. The following rule is quite common. For $m > 30$ i.e., $n > 62$, the hypothesis $Y_t = \varepsilon_t$ with ε_t being independent and $N(\mu, \sigma^2)$-distributed is rejected if $\Delta_{m-1} > c_\alpha / \sqrt{m-1}$, where $c_{0.05} = 1.36$ and $c_{0.01} = 1.63$ are the critical values for the levels $\alpha = 0.05$ and $\alpha = 0.01$.

This Bartlett-Kolmogorov-Smirnov test can also be carried out visually by plotting for $x \in [0,1]$ the sample distribution function $\hat{F}_{m-1}(x)$ and the band

$$y = x \pm \frac{c_\alpha}{\sqrt{m-1}}.$$

The hypothesis $Y_t = \varepsilon_t$ is rejected if $\hat{F}_{m-1}(x)$ is for some $x \in [0,1]$ outside of this *confidence band*.

Example 5.1.6. (Airline Data). We want to test, whether the variance stabilized, trend eliminated and seasonally adjusted Airline Data from Example 1.3.1 were generated from a white noise $(\varepsilon_t)_{t \in \mathbb{Z}}$, where ε_t are independent and identically normal distributed. The Fisher test statistic has the value $\kappa_m = 6.573$ and does, therefore, not reject the hypothesis at the levels $\alpha = 0.05$ and $\alpha = 0.01$, where $m = 65$. The Bartlett–Kolmogorov–Smirnov test, however, rejects this hypothesis at both levels, since $\Delta_{64} = 0.2319 > 1.36/\sqrt{64} = 0.17$ and also $\Delta_{64} > 1.63/\sqrt{64} = 0.20375$.

```
                    SPECTRA Procedure

----- Test for White Noise for variable DLOGNUM -----

Fisher's Kappa: M*MAX(P(*))/SUM(P(*))
Parameters:     M         =        65
                MAX(P(*)) =     0.028
                SUM(P(*)) =     0.275
Test Statistic: Kappa     =    6.5730
```

```
Bartlett's Kolmogorov-Smirnov Statistic:
          Maximum absolute difference of the standardized
          partial sums of the periodogram and the CDF of a
          uniform(0,1) random variable.

Test Statistic =    0.2319
```

Figure 5.1.1. Fisher's κ and the Bartlett-Kolmogorov-Smirnov test with $m = 65$ for testing a white noise generation of the adjusted Airline Data.

```
***     Program 5_1_1    ***;
TITLE1 'Tests for white noise';
TITLE2 'for the trend und seasonal';
TITLE3 'adjusted Airline Data';

DATA data1;
  INFILE 'c:\data\airline.txt';
  INPUT num @@;
  dlognum=DIF12(DIF(LOG(num)));

PROC SPECTRA DATA=data1 P WHITETEST OUT=data2;
  VAR dlognum;
RUN; QUIT;
```

In the `DATA` step the raw data of the airline passengers are read into the variable `num`. A log-transformation, building the fist order difference for trend elimination and the 12th order difference for elimination of a seasonal component lead to the variable `dlognum`, which is supposed to be generated by a stationary process.

Then `PROC SPECTRA` is applied to this variable, whereby the options `P` and `OUT=data2` generate a data set containing the periodogram data. The option `WHITETEST` causes SAS to carry out the two tests for a white noise, Fisher's test and the Bartlett-Kolmogorov-Smirnov test. SAS only provides the values of the test statistics but no decision. One has to compare these values with the critical values from Table 5.1.1 and the approximative ones $c_\alpha/\sqrt{m-1}$.

The following figure visualizes the rejection at both levels by the Bartlett-Kolmogorov-Smirnov test.

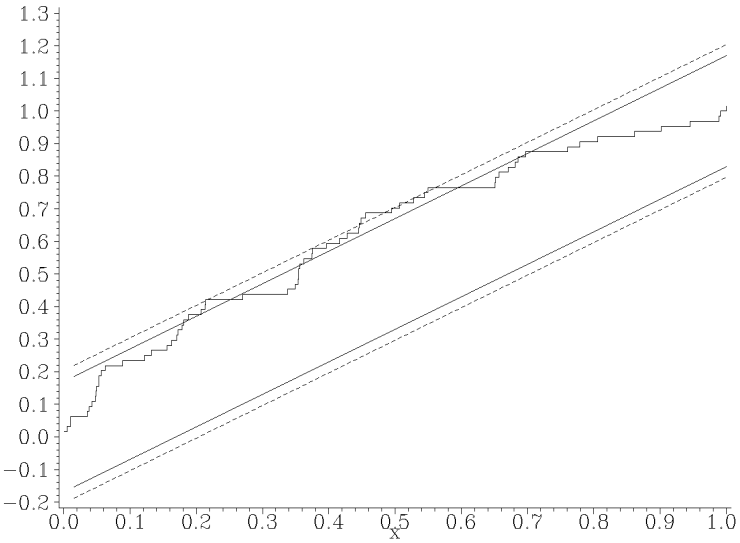


Figure 5.1.2. Bartlett-Kolmogorov-Smirnov test with $m = 65$ testing for a white noise generation of the adjusted Airline Data. Solid line/broken line = confidence bands for $\hat{F}_{m-1}(x)$, $x \in [0,1]$, at levels $\alpha = 0.05/0.01$.

```
***     Program 5_1_2   ***;
TITLE1 'Visualisation of the test for white noise';
TITLE2 'for the trend und seasonal adjusted';
TITLE3 'Airline Data';
* Note that this program needs data2
        generated by program 5_1_1;

PROC MEANS DATA=data2(FIRSTOBS=2) NOPRINT;
  VAR P_01;
  OUTPUT OUT=data3 SUM=psum;

DATA data4;
  SET data2(FIRSTOBS=2);
  IF _N_=1 THEN SET data3;
  RETAIN s 0;
  s=s+P_01/psum;
  fm=_N_/(_FREQ_-1);
  yu_01=fm+1.63/SQRT(_FREQ_-1);
```

```
  yl_01=fm-1.63/SQRT(_FREQ_-1);
  yu_05=fm+1.36/SQRT(_FREQ_-1);
  yl_05=fm-1.36/SQRT(_FREQ_-1);

SYMBOL1 V=NONE I=STEPJ C=GREEN;
SYMBOL2 V=NONE I=JOIN C=RED L=2;
SYMBOL3 V=NONE I=JOIN C=RED L=1;
AXIS1 LABEL=('x') ORDER=(.0 TO 1.0 BY .1);
AXIS2 LABEL=NONE;
PROC GPLOT DATA=data4;
  PLOT fm*s=1 yu_01*fm=2 yl_01*fm=2 yu_05*fm=3
       yl_05*fm=3 / OVERLAY HAXIS=AXIS1 VAXIS=AXIS2;
RUN; QUIT;
```

This program uses the data set `data2` created by Program 5_1_1, where the first observation belonging to the frequency 0 is dropped. `PROC MEANS` calculates the sum (keyword `SUM`) of the SAS periodogram variable `P_0` and stores it in the variable `psum` of the data set `data3`. The `NOPRINT` option suppresses the printing of the output.

The next `DATA` step combines every observation of `data2` with this sum by means of the `IF` statement. Furthermore a variable `s` is initialized with the value 0 by the `RETAIN` statement and then the portion of each periodogram value from the sum is cumulated. The variable `fm` contains the values of the empirical distribution function calculated by means of the automatically generated variable `_N_` containing the number of observation and the variable `_FREQ_`, which was created by `PROC MEANS` and contains the number `m`. The values of the upper and lower band are stored in the `y` variables.

The last part of this program contains `SYMBOL` and `AXIS` statements and `PROC GPLOT` to visualize the Bartlett-Kolmogorov-Smirnov statistic. The empirical distribution of the cumulated periodogram is represented as a step function due to the `I=STEPJ` option in the `SYMBOL1` statement.

5.2 Estimating Spectral Densities

We suppose in the following that $(Y_t)_{t\in\mathbb{Z}}$ is a stationary real valued process with mean μ and absolutely summable autocovariance function γ. According to Corollary 4.1.5, the process (Y_t) has the continuous spectral density

$$f(\lambda) = \sum_{h\in\mathbb{Z}} \gamma(h) e^{-i2\pi\lambda h}.$$

In the preceding section we computed the distribution of the empirical counterpart of a spectral density, the periodogram, in the particular case, when (Y_t) is a

Gaussian white noise. In this section we will investigate the limit behavior of the periodogram for arbitrary independent random variables (Y_t).

Asymptotic Properties of the Periodogram

In order to establish asymptotic properties of the periodogram, its following modification is quite useful. For Fourier frequencies k/n, $0 \le k \le [n/2]$ we put

$$\begin{aligned} I_n(k/n) &= \frac{1}{n}\Big|\sum_{t=1}^{n} Y_t e^{-i2\pi(k/n)t}\Big|^2 \\ &= \frac{1}{n}\left\{\Big(\sum_{t=1}^{n} Y_t \cos\Big(2\pi\frac{k}{n}t\Big)\Big)^2 + \Big(\sum_{t=1}^{n} Y_t \sin\Big(2\pi\frac{k}{n}t\Big)\Big)^2\right\}. \end{aligned} \tag{5.1}$$

Up to $k = 0$, this coincides by (3.6) with the definition of the periodogram as given in (3.7). From Theorem 3.2.3 we obtain the representation

$$I_n(k/n) = \begin{cases} n\bar{Y}_n^2, & k = 0 \\ \sum_{|h|<n} c(h)e^{-i2\pi(k/n)h}, & k = 1, \ldots, [n/2] \end{cases} \tag{5.2}$$

with $\bar{Y}_n := n^{-1}\sum_{t=1}^{n} Y_t$ and the sample autocovariance function

$$c(h) = \frac{1}{n}\sum_{t=1}^{n-|h|} \big(Y_t - \bar{Y}_n\big)\big(Y_{t+|h|} - \bar{Y}_n\big).$$

By representation (5.1) and the equations (3.6), the value $I_n(k/n)$ does not change for $k \neq 0$ if we replace the sample mean $\bar{Y}_n$ in $c(h)$ by the theoretical mean μ. This leads to the equivalent representation of the periodogram for $k \neq 0$

$$\begin{aligned} I_n(k/n) &= \sum_{|h|<n} \frac{1}{n}\Big(\sum_{t=1}^{n-|h|} (Y_t - \mu)(Y_{t+|h|} - \mu)\Big)e^{-i2\pi(k/n)h} \\ &= \frac{1}{n}\sum_{t=1}^{n}(Y_t - \mu)^2 + 2\sum_{h=1}^{n-1}\frac{1}{n}\Big(\sum_{t=1}^{n-|h|}(Y_t - \mu)(Y_{t+|h|} - \mu)\Big)\cos\Big(2\pi\frac{k}{n}h\Big). \end{aligned} \tag{5.3}$$

We define now the periodogram for $\lambda \in [0, 0.5]$ as a piecewise constant function

$$I_n(\lambda) = I_n(k/n) \quad \text{if} \quad \frac{k}{n} - \frac{1}{2n} < \lambda \le \frac{k}{n} + \frac{1}{2n}. \tag{5.4}$$

The following result shows that the periodogram $I_n(\lambda)$ is for $\lambda \neq 0$ an asymptotically unbiased estimator of the spectral density $f(\lambda)$.

Theorem 5.2.1. *Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary process with absolutely summable autocovariance function γ. Then we have with $\mu = \mathrm{E}(Y_t)$*

$$\begin{aligned}\mathrm{E}(I_n(0)) - n\mu^2 &\longrightarrow_{n\to\infty} f(0),\\ \mathrm{E}(I_n(\lambda)) &\longrightarrow_{n\to\infty} f(\lambda), \qquad \lambda \neq 0.\end{aligned}$$

If $\mu = 0$, then the convergence $\mathrm{E}(I_n(\lambda)) \longrightarrow_{n\to\infty} f(\lambda)$ holds uniformly on $[0, 0.5]$.

Proof. By representation (5.2) and the Césaro convergence result (Exercise 8 in Chapter 4) we have

$$\begin{aligned}\mathrm{E}(I_n(0)) - n\mu^2 &= \frac{1}{n}\Big(\sum_{t=1}^{n}\sum_{s=1}^{n}\mathrm{E}(Y_tY_s)\Big) - n\mu^2\\ &= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathrm{Cov}(Y_t, Y_s)\\ &= \sum_{|h|<n}\Big(1 - \frac{|h|}{n}\Big)\gamma(h) \longrightarrow_{n\to\infty} \sum_{h\in\mathbb{Z}}\gamma(h) = f(0).\end{aligned}$$

Define now for $\lambda \in [0, 0.5]$ the auxiliary function

$$g_n(\lambda) := \frac{k}{n}, \quad \text{if } \frac{k}{n} - \frac{1}{2n} < \lambda \leq \frac{k}{n} + \frac{1}{2n},\ k \in \mathbb{Z}. \tag{5.5}$$

Then we obviously have

$$I_n(\lambda) = I_n(g_n(\lambda)). \tag{5.6}$$

Choose now $\lambda \in (0, 0.5]$. Since $g_n(\lambda) \longrightarrow \lambda$ as $n \to \infty$, it follows that $g_n(\lambda) > 0$ for n large enough. By (5.3) and (5.6) we obtain for such n

$$\begin{aligned}\mathrm{E}(I_n(\lambda)) &= \sum_{|h|<n}\frac{1}{n}\sum_{t=1}^{n-|h|}\mathrm{E}\Big((Y_t - \mu)(Y_{t+|h|} - \mu)\Big)e^{-i2\pi g_n(\lambda)h}\\ &= \sum_{|h|<n}\left(1 - \frac{|h|}{n}\right)\gamma(h)e^{-i2\pi g_n(\lambda)h}.\end{aligned}$$

Since $\sum_{h\in\mathbb{Z}}|\gamma(h)| < \infty$, the series $\sum_{|h|<n}\gamma(h)\exp(-i2\pi\lambda h)$ converges to $f(\lambda)$ uniformly for $0 \leq \lambda \leq 0.5$. Kronecker's Lemma (Exercise 8) implies moreover

$$\Big|\sum_{|h|<n}\frac{|h|}{n}\gamma(h)e^{-i2\pi\lambda h}\Big| \leq \sum_{|h|<n}\frac{|h|}{n}|\gamma(h)| \longrightarrow_{n\to\infty} 0,$$

and, thus, the series

$$f_n(\lambda) := \sum_{|h|<n} \Big(1 - \frac{|h|}{n}\Big)\gamma(h)e^{-i2\pi\lambda h}$$

converges to $f(\lambda)$ uniformly in λ as well. From $g_n(\lambda) \longrightarrow_{n\to\infty} \lambda$ and the continuity of f we obtain for $\lambda \in (0, 0.5]$

$$\begin{aligned} |\,\mathrm{E}(I_n(\lambda)) - f(\lambda)| &= |f_n(g_n(\lambda)) - f(\lambda)| \\ &\le |f_n(g_n(\lambda)) - f(g_n(\lambda))| + |f(g_n(\lambda)) - f(\lambda)| \longrightarrow_{n\to\infty} 0. \end{aligned}$$

Note that $|g_n(\lambda) - \lambda| \le 1/(2n)$. The uniform convergence in case of $\mu = 0$ then follows from the uniform convergence of $g_n(\lambda)$ to λ and the uniform continuity of f on the compact interval $[0, 0.5]$. □

In the following result we compute the asymptotic distribution of the periodogram for independent and identically distributed random variables with zero mean, which are not necessarily Gaussian ones. The Gaussian case was already established in Corollary 5.1.2.

Theorem 5.2.2. *Let $Z_1, \ldots, Z_n$ be independent and identically distributed random variables with mean $\mathrm{E}(Z_t) = 0$ and variance $\mathrm{E}(Z_t^2) = \sigma^2 < \infty$. Denote by $I_n(\lambda)$ the pertaining periodogram as defined in* (5.6).

(i) The random vector $(I_n(\lambda_1), \ldots, I_n(\lambda_r))$ with $0 < \lambda_1 < \cdots < \lambda_r < 0.5$ converges in distribution for $n \to \infty$ to the distribution of r independent and identically exponential distributed random variables with mean σ^2.

(ii) If $\mathrm{E}(Z_t^4) = \eta\sigma^4 < \infty$, then we have for $k = 0, \ldots, [n/2]$

$$\mathrm{Var}(I_n(k/n)) = \begin{cases} 2\sigma^4 + n^{-1}(\eta - 3)\sigma^4, & k = 0 \text{ or } k = n/2, \text{ if } n \text{ even} \\ \sigma^4 + n^{-1}(\eta - 3)\sigma^4 & \text{elsewhere} \end{cases} \tag{5.7}$$

and

$$\mathrm{Cov}(I_n(j/n), I_n(k/n)) = n^{-1}(\eta - 3)\sigma^4, \quad j \ne k. \tag{5.8}$$

For $N(0, \sigma^2)$-distributed random variables Z_t we have $\eta = 3$ (Exercise 9) and, thus, $I_n(k/n)$ and $I_n(j/n)$ are for $j \ne k$ uncorrelated. Actually, we established in Corollary 5.1.2 that they are independent in this case.

Proof. Put for $\lambda \in (0, 0.5)$

$$A_n(\lambda) := A_n(g_n(\lambda)) := \sqrt{2/n} \sum_{t=1}^{n} Z_t \cos(2\pi g_n(\lambda) t),$$
$$B_n(\lambda) := B_n(g_n(\lambda)) := \sqrt{2/n} \sum_{t=1}^{n} Z_t \sin(2\pi g_n(\lambda) t),$$

with g_n defined in (5.5). Since

$$I_n(\lambda) = \frac{1}{2}\Big\{A_n^2(\lambda) + B_n^2(\lambda)\Big\},$$

it suffices, by repeating the arguments in the proof of Corollary 5.1.2, to show that

$$\big(A_n(\lambda_1), B_n(\lambda_1), \ldots, A_n(\lambda_r), B_n(\lambda_r)\big) \longrightarrow_D \boldsymbol{N(0,}\sigma^2\boldsymbol{I}_{2r}\boldsymbol{)},$$

where $\boldsymbol{I}_{2r}$ denotes the $2r \times 2r$-unity matrix and $\longrightarrow_D$ convergence in distribution. Since $g_n(\lambda) \longrightarrow_{n\to\infty} \lambda$, we have $g_n(\lambda) > 0$ for $\lambda \in (0, 0.5)$ if n is large enough. The independence of Z_t together with the definition of g_n and the orthogonality equations in (3.2) imply

$$\begin{aligned}\mathrm{Var}(A_n(\lambda)) &= \mathrm{Var}(A_n(g_n(\lambda))) \\ &= \sigma^2 \frac{2}{n} \sum_{t=1}^{n} \cos^2(2\pi g_n(\lambda) t) = \sigma^2.\end{aligned}$$

For $\varepsilon > 0$ we have

$$\begin{aligned}&\frac{1}{n} \sum_{t=1}^{n} \mathrm{E}\left(Z_t^2 \cos^2(2\pi g_n(\lambda) t) 1_{\{|Z_t \cos(2\pi g_n(\lambda) t)| > \varepsilon\sqrt{n\sigma^2}\}}\right) \\ &\quad \le \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}\left(Z_t^2 1_{\{|Z_t| > \varepsilon\sqrt{n\sigma^2}\}}\right) = \mathrm{E}\left(Z_1^2 1_{\{|Z_1| > \varepsilon\sqrt{n\sigma^2}\}}\right) \longrightarrow_{n\to\infty} 0\end{aligned}$$

i.e., the triangular array $(2/n)^{1/2} Z_t \cos(2\pi g_n(\lambda) t)$, $1 \le t \le n$, $n \in \mathbb{N}$, satisfies the *Lindeberg condition* implying $A_n(\lambda) \longrightarrow_D N(0, \sigma^2)$; see, for example, Theorem 7.2 in Billingsley (1968). Similarly one shows that $B_n(\lambda) \longrightarrow_D N(0, \sigma^2)$ as well. Since the random vector $(A_n(\lambda_1), B_n(\lambda_1), \ldots, A_n(\lambda_r), B_n(\lambda_r))$ has by (3.2) the covariance matrix $\sigma^2 \boldsymbol{I}_{2r}$, its asymptotic joint normality follows easily by applying the *Cramér-Wold device*, cf. Theorem 7.7 in Billingsley (1968), and proceeding as before. This proves part (i) of Theorem 5.2.2.

From the definition of the periodogram in (5.1) we conclude as in the proof of Theorem 3.2.3

$$I_n(k/n) = \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} Z_s Z_t e^{-i2\pi(k/n)(s-t)}$$

and, thus, we obtain

$$\begin{aligned}&\mathrm{E}(I_n(j/n)I_n(k/n))\\&= \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{u=1}^{n}\sum_{v=1}^{n}\mathrm{E}(Z_sZ_tZ_uZ_v)e^{-i2\pi(j/n)(s-t)}e^{-i2\pi(k/n)(u-v)}.\end{aligned}$$

We have

$$\mathrm{E}(Z_sZ_tZ_uZ_v) = \begin{cases}\eta\sigma^4, & s=t=u=v\\ \sigma^4, & s=t\neq u=v,\ s=u\neq t=v,\ s=v\neq t=u\\ 0 & \text{elsewhere}\end{cases}$$

and

$$e^{-i2\pi(j/n)(s-t)}e^{-i2\pi(k/n)(u-v)} = \begin{cases}1, & s=t, u=v\\ e^{-i2\pi((j+k)/n)s}e^{i2\pi((j+k)/n)t}, & s=u, t=v\\ e^{-i2\pi((j-k)/n)s}e^{i2\pi((j-k)/n)t}, & s=v, t=u.\end{cases}$$

This implies

$$\begin{aligned}&\mathrm{E}(I_n(j/n)I_n(k/n))\\&= \frac{\eta\sigma^4}{n} + \frac{\sigma^4}{n^2}\Big\{n(n-1) + \Big|\sum_{t=1}^{n}e^{-i2\pi((j+k)/n)t}\Big|^2 + \Big|\sum_{t=1}^{n}e^{-i2\pi((j-k)/n)t}\Big|^2 - 2n\Big\}\\&= \frac{(\eta-3)\sigma^4}{n} + \sigma^4\Big\{1 + \frac{1}{n^2}\Big|\sum_{t=1}^{n}e^{i2\pi((j+k)/n)t}\Big|^2 + \frac{1}{n^2}\Big|\sum_{t=1}^{n}e^{i2\pi((j-k)/n)t}\Big|^2\Big\}.\end{aligned}$$

From $\mathrm{E}(I_n(k/n)) = n^{-1}\sum_{t=1}^{n}\mathrm{E}(Z_t^2) = \sigma^2$ we finally obtain

$$\begin{aligned}&\mathrm{Cov}(I_n(j/n), I_n(k/n))\\&= \frac{(\eta-3)\sigma^4}{n} + \frac{\sigma^4}{n^2}\Big\{\Big|\sum_{t=1}^{n}e^{i2\pi((j+k)/n)t}\Big|^2 + \Big|\sum_{t=1}^{n}e^{i2\pi((j-k)/n)t}\Big|^2\Big\},\end{aligned}$$

from which (5.7) and (5.8) follow by using (3.6)). □

Remark 5.2.3. Theorem 5.2.2 can be generalized to filtered processes $Y_t = \sum_{u\in\mathbb{Z}}a_uZ_{t-u}$, with $(Z_t)_{t\in\mathbb{Z}}$ as in Theorem 5.2.2. In this case one has to replace σ^2, which equals by Example 4.1.3 the constant spectral density $f_Z(\lambda)$, in (i) by the spectral density $f_Y(\lambda_i)$, $1\le i\le r$. If in addition $\sum_{u\in\mathbb{Z}}|a_u||u|^{1/2} < \infty$, then we have in (ii) the expansions

$$\mathrm{Var}(I_n(k/n)) = \begin{cases}2f_Y^2(k/n) + O(n^{-1/2}), & k=0 \text{ or } k=n/2, \text{ if } n \text{ is even}\\ f_Y^2(k/n) + O(n^{-1/2}) & \text{elsewhere,}\end{cases}$$

and

$$\mathrm{Cov}(I_n(j/n), I_n(k/n)) = O(n^{-1}), \quad j \neq k,$$

where I_n is the periodogram pertaining to $Y_1, \ldots, Y_n$. The above terms $O(n^{-1/2})$ and $O(n^{-1})$ are uniformly bounded in k and j by a constant C. We omit the highly technical proof and refer to Section 10.3 of Brockwell and Davis (1991).

Recall that the class of processes $Y_t = \sum_{u \in \mathbb{Z}} a_u Z_{t-u}$ is a fairly rich one, which contains in particular *ARMA*-processes, see Section 2.2 and Remark 2.1.12.

Discrete Spectral Average Estimator

The preceding results show that the periodogram is not a consistent estimator of the spectral density. The law of large numbers together with the above remark motivates, however, that consistency can be achieved for a smoothed version of the periodogram such as a simple moving average

$$\sum_{|j| \leq m} \frac{1}{2m+1} I_n\Big(\frac{k+j}{n}\Big),$$

which puts equal weights $1/(2m+1)$ on adjacent values. Dropping the condition of equal weights, we define a general linear smoother by the linear filter

$$\hat{f}_n\Big(\frac{k}{n}\Big) := \sum_{|j| \leq m} a_{jn} I_n\Big(\frac{k+j}{n}\Big). \tag{5.9}$$

The sequence $m = m(n)$, defining the adjacent points of k/n, has to satisfy

$$m \longrightarrow_{n\to\infty} \infty \ \text{ and } \ m/n \longrightarrow_{n\to\infty} 0, \tag{5.10}$$

and the weights a_{jn} have the properties

$$\begin{aligned}
&\text{(a)} \ a_{jn} \geq 0,\\
&\text{(b)} \ a_{jn} = a_{-jn},\\
&\text{(c)} \ \textstyle\sum_{|j| \leq m} a_{jn} = 1,\\
&\text{(d)} \ \textstyle\sum_{|j| \leq m} a_{jn}^2 \longrightarrow_{n\to\infty} 0.
\end{aligned} \tag{5.11}$$

For the simple moving average we have, for example

$$a_{jn} = \begin{cases} 1/(2m+1), & |j| \leq m \\ 0 & \text{elsewhere} \end{cases}$$

and $\sum_{|j|\le m} a_{jn}^2 = 1/(2m+1) \longrightarrow_{n\to\infty} 0$. For $\lambda \in [0, 0.5]$ we put $I_n(0.5+\lambda) := I_n(0.5-\lambda)$, which defines the periodogram also on $[0.5, 1]$. If $(k+j)/n$ is outside of the interval $[0,1]$, then $I_n((k+j)/n)$ is understood as the periodic extension of I_n with period 1. This also applies to the spectral density f. The estimator

$$\hat{f}_n(\lambda) := \hat{f}_n(g_n(\lambda)),$$

with g_n as defined in (5.5), is called the *discrete spectral average estimator*. The following result states its consistency for linear processes.

Theorem 5.2.4. *Let $Y_t = \sum_{u\in\mathbb{Z}} b_u Z_{t-u}$, $t \in \mathbb{Z}$, where Z_t are iid with $\mathrm{E}(Z_t) = 0$, $\mathrm{E}(Z_t^4) < \infty$ and $\sum_{u\in\mathbb{Z}} |b_u||u|^{1/2} < \infty$. Then we have for $0 \le \mu, \lambda \le 0.5$*

(i) $\lim_{n\to\infty} \mathrm{E}\left(\hat{f}_n(\lambda)\right) = f(\lambda)$,

(ii) $$\lim_{n\to\infty} \frac{\mathrm{Cov}\left(\hat{f}_n(\lambda), \hat{f}_n(\mu)\right)}{\left(\sum_{|j|\le m} a_{jn}^2\right)} = \begin{cases} 2f^2(\lambda), & \lambda = \mu = 0 \text{ or } 0.5 \\ f^2(\lambda), & 0 < \lambda = \mu < 0.5 \\ 0, & \lambda \neq \mu. \end{cases}$$

Condition (d) in (5.11) on the weights together with (ii) in the preceding result entails that $\mathrm{Var}(\hat{f}_n(\lambda)) \longrightarrow_{n\to\infty} 0$ for any $\lambda \in [0, 0.5]$. Together with (i) we, therefore, obtain that the mean squared error of $\hat{f}_n(\lambda)$ vanishes asymptotically:

$$\mathrm{MSE}(\hat{f}_n(\lambda)) = \mathrm{E}\left\{\left(\hat{f}_n(\lambda) - f(\lambda)\right)^2\right\} = \mathrm{Var}(\hat{f}_n(\lambda)) + \mathrm{Bias}^2(\hat{f}_n(\lambda)) \longrightarrow_{n\to\infty} 0.$$

Proof. By the definition of the spectral density estimator in (5.9) we have

$$\begin{aligned} |\,\mathrm{E}(\hat{f}_n(\lambda)) - f(\lambda)| &= \Big| \sum_{|j|\le m} a_{jn} \Big\{ \mathrm{E}\left(I_n(g_n(\lambda) + j/n)\right) - f(\lambda) \Big\} \Big| \\ &= \Big| \sum_{|j|\le m} a_{jn} \Big\{ \mathrm{E}\left(I_n(g_n(\lambda) + j/n)\right) - f(g_n(\lambda) + j/n) \\ &\qquad + f(g_n(\lambda) + j/n) - f(\lambda) \Big\} \Big|, \end{aligned}$$

where (5.10) together with the uniform convergence of $g_n(\lambda)$ to λ implies

$$\max_{|j|\le m} |g_n(\lambda) + j/n - \lambda| \longrightarrow_{n\to\infty} 0.$$

Choose $\varepsilon > 0$. The spectral density f of the process (Y_t) is continuous (Exercise 16), and hence we have

$$\max_{|j|\le m} |f(g_n(\lambda) + j/n) - f(\lambda)| < \varepsilon/2$$

if n is sufficiently large. From Theorem 5.2.1 we know that in the case $\mathrm{E}(Y_t) = 0$

$$\max_{|j|\le m} |\,\mathrm{E}(I_n(g_n(\lambda)+j/n)) - f(g_n(\lambda)+j/n)| < \varepsilon/2$$

if n is large. Condition (c) in (5.11) together with the triangular inequality implies $|\,\mathrm{E}(\hat{f}_n(\lambda)) - f(\lambda)| < \varepsilon$ for large n. This implies part (i).

From the definition of $\hat{f}_n$ we obtain

$$\begin{aligned}&\mathrm{Cov}(\hat{f}_n(\lambda), \hat{f}_n(\mu))\\ &= \sum_{|j|\le m}\sum_{|k|\le m} a_{jn}a_{kn}\,\mathrm{Cov}\Big(I_n(g_n(\lambda)+j/n), I_n(g_n(\mu)+k/n)\Big).\end{aligned}$$

If $\lambda \neq \mu$ and n sufficiently large we have $g_n(\lambda)+j/n \neq g_n(\mu)+k/n$ for arbitrary $|j|, |k| \le m$. According to Remark 5.2.3 there exists a universal constant $C_1 > 0$ such that

$$\begin{aligned}|\,\mathrm{Cov}(\hat{f}_n(\lambda), \hat{f}_n(\mu))| &= \Big|\sum_{|j|\le m}\sum_{|k|\le m} a_{jn}a_{kn}O(n^{-1})\Big|\\ &\le C_1 n^{-1}\Big(\sum_{|j|\le m} a_{jn}\Big)^2\\ &\le C_1\frac{2m+1}{n}\sum_{|j|\le m} a_{jn}^2,\end{aligned}$$

where the final inequality is an application of the Cauchy–Schwarz inequality. Since $m/n \longrightarrow_{n\to\infty} 0$, we have established (ii) in the case $\lambda \neq \mu$. Suppose now $0 < \lambda = \mu < 0.5$. Utilizing again Remark 5.2.3 we have

$$\begin{aligned}\mathrm{Var}(\hat{f}_n(\lambda)) &= \sum_{|j|\le m} a_{jn}^2\Big\{f^2(g_n(\lambda)+j/n) + O(n^{-1/2})\Big\}\\ &\quad + \sum_{|j|\le m}\sum_{|k|\le m} a_{jn}a_{kn}O(n^{-1}) + o(n^{-1})\\ &=: S_1(\lambda) + S_2(\lambda) + o(n^{-1}).\end{aligned}$$

Repeating the arguments in the proof of part (i) one shows that

$$S_1(\lambda) = \Big(\sum_{|j|\le m} a_{jn}^2\Big)f^2(\lambda) + o\Big(\sum_{|j|\le m} a_{jn}^2\Big).$$

Furthermore, with a suitable constant $C_2 > 0$, we have

$$|S_2(\lambda)| \le C_2\frac{1}{n}\Big(\sum_{|j|\le m} a_{jn}\Big)^2 \le C_2\frac{2m+1}{n}\sum_{|j|\le m} a_{jn}^2.$$

Thus we established the assertion of part (ii) also in the case $0 < \lambda = \mu < 0.5$. The remaining cases $\lambda = \mu = 0$ and $\lambda = \mu = 0.5$ are shown in a similar way (Exercise 17). □

The preceding result requires zero mean variables Y_t. This might, however, be too restrictive in practice. Due to (3.6), the periodograms of $(Y_t)_{1\le t\le n}$, $(Y_t - \mu)_{1\le t\le n}$ and $(Y_t - \bar{Y})_{1\le t\le n}$ coincide at Fourier frequencies different from zero. At frequency $\lambda = 0$, however, they will differ in general. To estimate $f(0)$ consistently also in the case $\mu = E(Y_t) \neq 0$, one puts

$$\hat{f}_n(0) := a_{0n} I_n(1/n) + 2 \sum_{j=1}^{m} a_{jn} I_n\big((1+j)/n\big). \tag{5.12}$$

Each time the value $I_n(0)$ occurs in the moving average (5.9), it is replaced by $\hat{f}_n(0)$. Since the resulting estimator of the spectral density involves only Fourier frequencies different from zero, we can assume without loss of generality that the underlying variables Y_t have zero mean.

Example 5.2.5. (Sunspot Data). We want to estimate the spectral density underlying the Sunspot Data. These data, also known as the Wolf or Wölfer (a student of Wolf) Data, are the yearly sunspot numbers between 1749 and 1924. For a discussion of these data and further literature we refer to Wei (1990), Example 6.2. Figure 5.2.1 shows the pertaining periodogram and Figure 5.2.2 displays the discrete spectral average estimator with weights $a_{0n} = a_{1n} = a_{2n} = 3/21$, $a_{3n} = 2/21$ and $a_{4n} = 1/21$, $n = 176$. These weights pertain to a simple moving average of length 3 of a simple moving average of length 7. The smoothed version joins the two peaks close to the frequency $\lambda = 0.1$ visible in the periodogram. The observation that a periodogram has the tendency to split a peak is known as the *leakage phenomenon*.

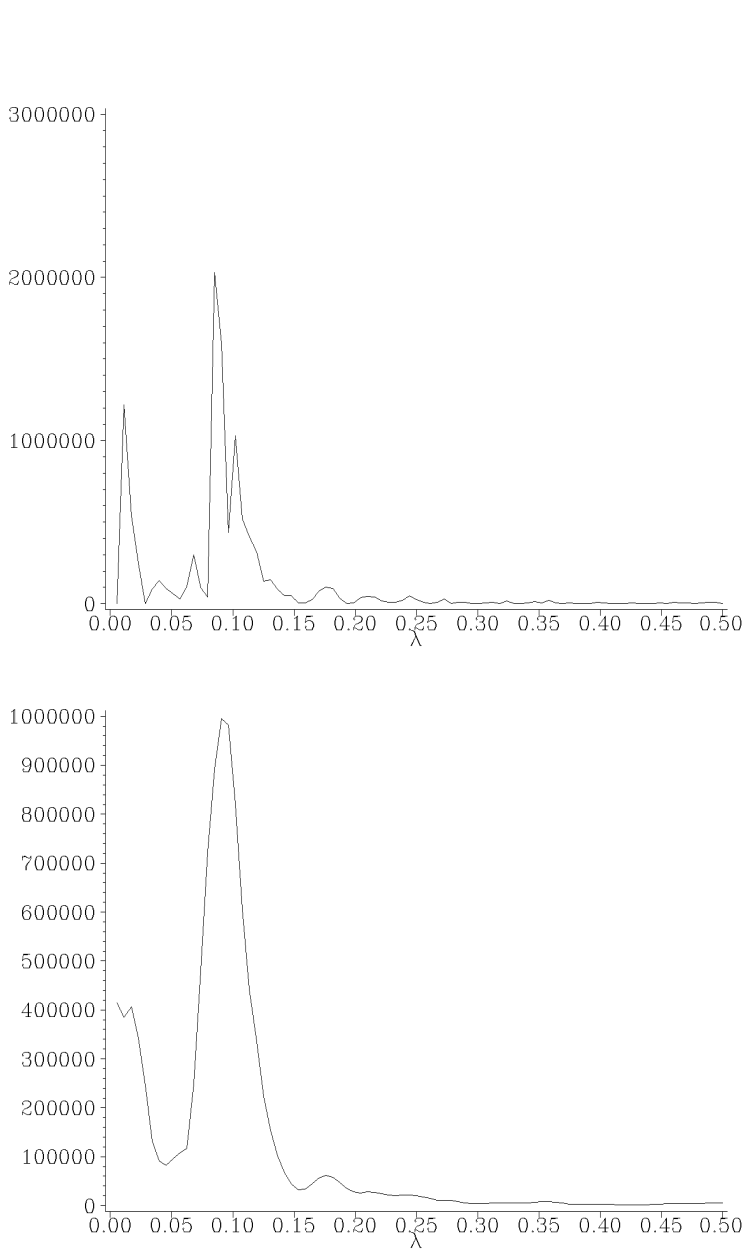


Figure 5.2.1. Periodogram and discrete spectral average estimate for Sunspot Data.

```
***     Program 5_2_1   ***;
TITLE1 'Periodogram and spectral density estimate';
TITLE2 'Woelfer Sunspot Data';

DATA data1;
  INFILE 'c:\data\sunspot.txt';
  INPUT num @@;

PROC SPECTRA DATA=data1 P S OUT=data2;
  VAR num;
  WEIGHTS 1 2 3 3 3 2 1;

DATA data3;
  SET data2(FIRSTOBS=2);
  lambda=FREQ/(2*CONSTANT('PI'));
  p=P\_01/2;
  s=S\_01/2*4*CONSTANT('PI');

SYMBOL1 I=JOIN C=RED V=NONE L=1;
AXIS1 LABEL=(F=CGREEK 'l') ORDER=(0 TO .5 BY .05);
AXIS2 LABEL=NONE;
PROC GPLOT DATA=data3;
  PLOT p*lambda / HAXIS=AXIS1 VAXIS=AXIS2;
  PLOT s*lambda / HAXIS=AXIS1 VAXIS=AXIS2;
RUN; QUIT;
```

In the `DATA` step the data of the sunspots are read into the variable `num`.

Then `PROC SPECTRA` is applied to this variable, whereby the options `P` (see Program 5_1_1) and `S` generate a data set stored in `data2` containing the periodogram data and the estimation of the spectral density which SAS computes with the weights given in the `WEIGHTS` statement. Note that SAS automatically normalizes these weights.

In following `DATA` step the slightly different definition of the periodogram by SAS is being adjusted to the definition used here (see Program 3_2_1). Both plots are then printed with `PROC GPLOT`.

A mathematically convenient way to generate weights a_{jn}, which satisfy the conditions (5.11), is the use of a kernel function. Let $K : [-1, 1] \to [0, \infty)$ be a symmetric function i.e., $K(-x) = K(x),\ x \in [-1, 1]$, which satisfies $\int_{-1}^{1} K^2(x)\,dx < \infty$. Let now $m = m(n) \longrightarrow_{n\to\infty} \infty$ be an arbitrary sequence of integers with

$m/n \longrightarrow_{n\to\infty} 0$ and put

$$a_{jn} := \frac{K(j/m)}{\sum_{i=-m}^{m} K(i/m)}, \quad -m \le j \le m. \tag{5.13}$$

These weights satisfy the conditions (5.11) (Exercise 18).

Take for example $K(x) := 1 - |x|$, $-1 \le x \le 1$. Then we obtain

$$a_{jn} = \frac{m - |j|}{m^2}, \quad -m \le j \le m.$$

Example 5.2.6. (i) The *truncated kernel* is defined by

$$K_T(x) = \begin{cases} 1, & |x| \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

(ii) The *Bartlett* or *triangular kernel* is given by

$$K_B(x) := \begin{cases} 1 - |x|, & |x| \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

(iii) The *Blackman–Tukey kernel* (1959) is defined by

$$K_{BT}(x) = \begin{cases} 1 - 2a + 2a\cos(x), & |x| \le 1 \\ 0 & \text{elsewhere,} \end{cases}$$

where $0 < a \le 1/4$. The particular choice $a = 1/4$ yields the *Tukey–Hanning kernel*.

(iv) The *Parzen kernel* (1957) is given by

$$K_P(x) := \begin{cases} 1 - 6|x|^2 + 6|x|^3, & |x| < 1/2 \\ 2(1 - |x|)^3, & 1/2 \le |x| \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

We refer to Andrews (1991) for a discussion of these kernels.

Example 5.2.7. We consider realizations of the $MA(1)$-process $Y_t = \varepsilon_t - 0.6\varepsilon_{t-1}$ with ε_t independent and standard normal for $t = 1, \ldots, n = 160$. Example 4.3.2 implies that the process (Y_t) has the spectral density $f(\lambda) = 1 - 1.2\cos(2\pi\lambda) + 0.36$. We estimate $f(\lambda)$ by means of the Tukey–Hanning kernel.

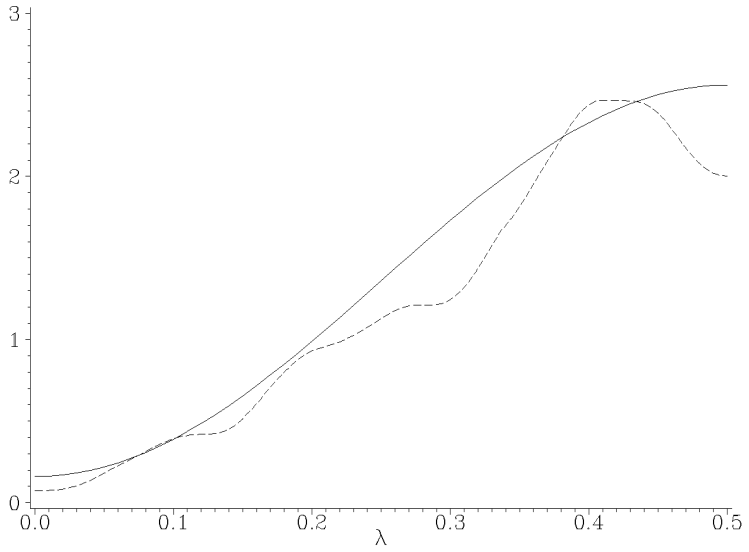


Figure 5.2.2. Discrete spectral average estimator (broken line) with Blackman–Tukey kernel with parameters $r = 10$, $a = 1/4$ and underlying spectral density $f(\lambda) = 1 - 1.2\cos(2\pi\lambda) + 0.36$ (solid line).

```
***     Program 5_2_2   ***;
TITLE1 'Spectral density and Blackman-Tukey estimator';
TITLE2 'of MA(1)-process';

DATA data1;
  DO t=0 TO 160;
    e=RANNOR(1);
    y=e-.6*LAG(e);
    OUTPUT;
  END;

PROC SPECTRA DATA=data1(FIRSTOBS=2) S OUT=data2;
  VAR y;
  WEIGHTS TUKEY 10 0;
RUN;

DATA data3;
  SET data2;
```

```
  lambda=FREQ/(2*CONSTANT('PI'));
  s=S_01/2*4*CONSTANT('PI');

DATA data4;
  DO l=0 TO .5 BY .01;
    f=1-1.2*COS(2*CONSTANT('PI')*l)+.36;
    OUTPUT;
  END;

DATA data5;
  MERGE data3(KEEP=s lambda) data4;

AXIS1 LABEL=NONE;
AXIS2 LABEL=(F=CGREEK 'l') ORDER=(0 TO .5 BY .1);
SYMBOL1 I=JOIN C=BLUE V=NONE L=1;
SYMBOL2 I=JOIN C=RED V=NONE L=3;
PROC GPLOT DATA=data5;
  PLOT f*l=1 s*lambda=2
             / OVERLAY VAXIS=AXIS1 HAXIS=AXIS2;
RUN; QUIT;
```

In the first `DATA` step the realizations of an $MA(1)$-process with the given parameters are created. Thereby the function `RANNOR`, which generates standard normally distributed data, and `LAG`, which accesses the value of `e` of the preceding loop, are used.

As in Program 5_2_1 `PROC SPECTRA` computes the estimator of the spectral density (after dropping the first observation) by the option `S` and stores them in `data2`. The weights used here come from the Tukey–Hanning kernel with a specified bandwidth of m=10. The second number after the `TUKEY` option can be used to refine the choice of the bandwidth. Since this is not needed here it is set to `0`.

The next `DATA` step adjusts the different definitions of the periodogram used here and by SAS (see Program 3_2_1). The following `DATA` step generates the values of the underlying spectral density. These are merged with the values of the estimated spectral density and then displayed by `PROC GPLOT`.

Confidence Intervals for Spectral Densities

The random variables $I_n((k+j)/n)/f((k+j)/n)$, $0 < k+j < n/2$, will by Remark 5.2.3 for large n approximately behave like independent and standard exponential distributed random variables X_j. This suggests that the distribution of the discrete

spectral average estimator

$$\hat{f}(k/n) = \sum_{|j|\le m} a_{jn} I_n\Big(\frac{k+j}{n}\Big)$$

can be approximated by that of the weighted sum $\sum_{|j|\le m} a_{jn} X_j f((k+j)/n)$. Tukey (1949) showed that the distribution of this weighted sum can in turn be approximated by that of cY with a suitably chosen $c > 0$, where Y follows a *gamma distribution* with parameters $p := \nu/2 > 0$ and $b = 1/2$ i.e.,

$$P\{Y \le t\} = \frac{b^p}{\Gamma(p)} \int_0^t x^{p-1} \exp(-bx)\, dx, \quad t \ge 0,$$

where $\Gamma(p) := \int_0^\infty x^{p-1} \exp(-x)\, dx$ denotes the gamma function. The parameters ν and c are determined by the method of moments as follows: ν and c are chosen such that cY has mean $f(k/n)$ and its variance equals the leading term of the variance expansion of $\hat{f}(k/n)$ in Theorem 5.2.4 (Exercise 21):

$$\begin{aligned} \mathrm{E}(cY) &= c\nu = f(k/n), \\ \mathrm{Var}(cY) &= 2c^2\nu = f^2(k/n) \sum_{|j|\le m} a_{jn}^2. \end{aligned}$$

The solutions are obviously

$$c = \frac{f(k/n)}{2} \sum_{|j|\le m} a_{jn}^2$$

and

$$\nu = \frac{2}{\sum_{|j|\le m} a_{jn}^2}.$$

Note that the gamma distribution with parameters $p = \nu/2$ and $b = 1/2$ equals the χ^2-distribution with ν degrees of freedom if ν is an integer. The number ν is, therefore, called the *equivalent degree of freedom*. Observe that $\nu/f(k/n) = 1/c$; the random variable $\nu\hat{f}(k/n)/f(k/n) = \hat{f}(k/n)/c$ now approximately follows a $\chi^2(\nu)$-distribution with the convention that $\chi^2(\nu)$ is the gamma distribution with parameters $p = \nu/2$ and $b = 1/2$ if ν is not an integer. The interval

$$\left(\frac{\nu\hat{f}(k/n)}{\chi^2_{1-\alpha/2}(\nu)}, \frac{\nu\hat{f}(k/n)}{\chi^2_{\alpha/2}(\nu)} \right) \tag{5.14}$$

is a confidence interval for $f(k/n)$ of approximate level $1-\alpha$, $\alpha \in (0,1)$. By $\chi^2_q(\nu)$ we denote the q-quantile of the $\chi^2(\nu)$-distribution i.e., $P\{Y \le \chi^2_q(\nu)\} = q$,

$0 < q < 1$. Taking logarithms in (5.14), we obtain the confidence interval for $\log(f(k/n))$

$$C_{\nu,\alpha}(k/n) := \Big(\log(\hat{f}(k/n)) + \log(\nu) - \log(\chi^2_{1-\alpha/2}(\nu)), \\ \log(\hat{f}(k/n)) + \log(\nu) - \log(\chi^2_{\alpha/2}(\nu)) \Big).$$

This interval has constant length $\log(\chi^2_{1-\alpha/2}\ (\nu)/\chi^2_{\alpha/2}(\nu))$. Note that $C_{\nu,\alpha}(k/n)$ is a level $(1-\alpha)$-confidence interval only for $\log(f(\lambda))$ at a fixed Fourier frequency $\lambda = k/n$, with $0 < k < [n/2]$, but not *simultaneously* for $\lambda \in (0, 0.5)$.

Example 5.2.8. In continuation of Example 5.2.7 we want to estimate the spectral density $f(\lambda) = 1 - 1.2\cos(2\pi\lambda) + 0.36$ of the $MA(1)$-process $Y_t = \varepsilon_t - 0.6\varepsilon_{t-1}$ using the discrete spectral average estimator $\hat{f}_n(\lambda)$ with the weights 1, 3, 6, 9, 12, 15, 18, 20, 21, 21, 21, 20, 18, 15, 12, 9, 6, 3, 1, each divided by 231. These weights are generated by iterating simple moving averages of lengths 3, 7 and 11. Figure 5.2.3 displays the logarithms of the estimates, of the true spectral density and the pertaining confidence intervals.

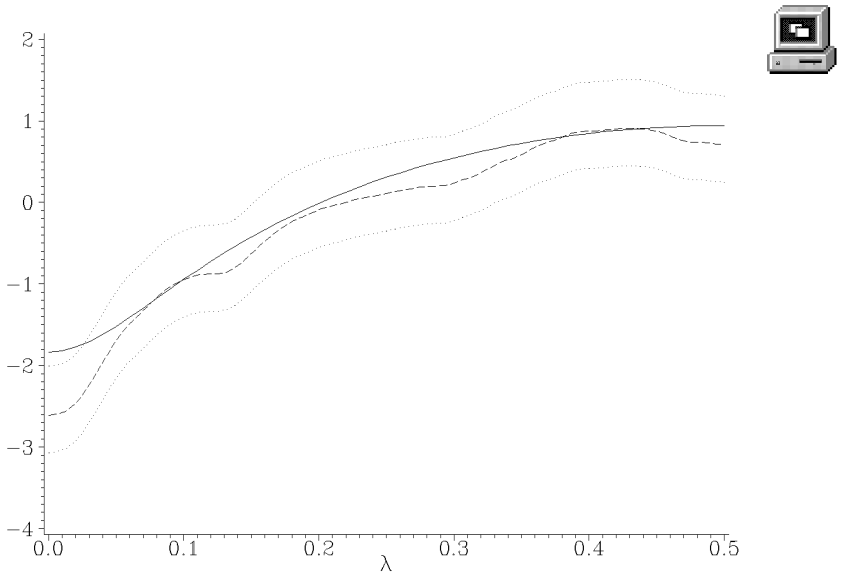


Figure 5.2.3. Logarithms of discrete spectral average estimates (broken line), of spectral density $f(\lambda) = 1 - 1.2\cos(2\pi\lambda) + 0.36$ (solid line) of $MA(1)$-process $Y_t = \varepsilon_t - 0.6\varepsilon_{t-1}$, $t = 1, \ldots, n = 160$, and confidence intervals of level $1 - \alpha = 0.95$ for $\log(f(k/n))$.

```
***     Program 5_2_3   ***;
TITLE1 'Logarithms of spectral density,;
TITLE2 'of their estimates and confidence intervals';
TITLE3 'of MA(1)-process';

DATA data1;
  DO t=0 TO 160;
    e=RANNOR(1);
    y=e-.6*LAG(e);
    OUTPUT;
  END;

PROC SPECTRA DATA=data1(FIRSTOBS=2) S OUT=data2;
  VAR y; WEIGHTS 1 3 6 9 12 15 18 20 21 21 21
                          20 18 15 12 9 6 3 1;
  RUN;

DATA data3; SET data2;
  lambda=FREQ/(2*CONSTANT('PI'));
  log_s_01=LOG(S_01/2*4*CONSTANT('PI'));
  nu=2/(3763/53361);
  c1=log_s_01+LOG(nu)-LOG(CINV(.975,nu));
  c2=log_s_01+LOG(nu)-LOG(CINV(.025,nu));

DATA data4;
  DO l=0 TO .5 BY 0.01;
    log_f=LOG((1-1.2*COS(2*CONSTANT('PI')*l)+.36));
    OUTPUT;
  END;

DATA data5;
  MERGE data3(KEEP=log_s_01 lambda c1 c2) data4;

AXIS1 LABEL=NONE;
AXIS2 LABEL=(F=CGREEK 'l') ORDER=(0 TO .5 BY .1);
SYMBOL1 I=JOIN C=BLUE V=NONE L=1;
SYMBOL2 I=JOIN C=RED V=NONE L=2;
SYMBOL3 I=JOIN C=GREEN V=NONE L=33;
PROC GPLOT DATA=data5;
  PLOT log_f*l=1 log_s_01*lambda=2 c1*lambda=3
       c2*lambda=3 / OVERLAY VAXIS=AXIS1 HAXIS=AXIS2;
RUN; QUIT;
```

This programm starts identically to Program 5_2_2 with the generation of an $MA(1)$-process and of the computation the spectral density estimator. Only this time the weights are directly given to SAS.

In the next `DATA` step the usual adjustment of the frequencies is done. This is followed by the computation of ν according to its definition. The logarithm of the confidence intervals is calculated with the help of the function `CINV` which returns quantiles of a χ^2-distribution with ν degrees of freedom.

The rest of the program which displays the logarithm of the estimated spectral density, of the underlying density and of the confidence intervals is analogous to Program 5_2_2.

Exercises

1. For independent random variables X, Y having continuous distribution functions it follows that $P\{X = Y\} = 0$. Hint: Fubini's theorem.

2. Let $X_1, \ldots, X_n$ be iid random variables with values in $\mathbb{R}$ and distribution function F. Denote by $X_{1:n} \le \cdots \le X_{n:n}$ the pertaining order statistics. Then we have

$$P\{X_{k:n} \le t\} = \sum_{j=k}^{n} \binom{n}{j} F(t)^j (1 - F(t))^{n-j}, \quad t \in \mathbb{R}.$$

The maximum $X_{n:n}$ has in particular the distribution function F^n, and the minimum $X_{1:n}$ has distribution function $1-(1-F)^n$. Hint: $\{X_{k:n} \le t\} = \big\{ \sum_{j=1}^{n} 1_{(-\infty,t]}(X_j) \ge k \big\}$.

3. Suppose in addition to the conditions in the preceding exercise that F has a (Lebesgue) density f. The ordered vector $(X_{1:n}, \ldots, X_{n:n})$ then has a density

$$f_n(x_1, \ldots, x_n) = n! \prod_{j=1}^{n} f(x_j), \quad x_1 < \cdots < x_n,$$

and zero elsewhere. Hint: Denote by S_n the group of permutations of $\{1, \ldots, n\}$ i.e., $(\tau(1), \ldots, (\tau(n))$ with $\tau \in S_n$ is a permutation of $(1, \ldots, n)$. Put for $\tau \in S_n$ the set $B_\tau := \{X_{\tau(1)} < \cdots < X_{\tau(n)}\}$. These sets are disjoint and we have $P(\sum_{\tau \in S_n} B_\tau) = 1$ since $P\{X_j = X_k\} = 0$ for $i \ne j$ (cf. Exercise 1).

4. (i) Let X and Y be independent, standard normal distributed random variables. Show that the vector $(X, Z)^T := (X, \rho X + \sqrt{1 - \rho^2} Y)^T$, $-1 < \rho < 1$, is normal distributed with mean vector $(0, 0)$ and covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, and that X and Z are independent if and only if they are uncorrelated (i.e., $\rho = 0$).

(i) Suppose that X and Y are normal distributed and uncorrelated. Does this imply the independence of X and Y? Hint: Let X $N(0, 1)$-distributed and define the

random variable $Y = VX$ with V independent of X and $P\{V = -1\} = 1/2 = P\{V = 1\}$.

5. Generate 100 independent and standard normal random variables ε_t and plot the periodogram. Is the hypothesis that the observations were generated by a white noise rejected at level $\alpha = 0.05(0.01)$? Visualize the Bartlett-Kolmogorov-Smirnov test by plotting the empirical distribution function of the cumulated periodograms S_j, $1 \leq j \leq 48$, together with the pertaining bands for $\alpha = 0.05$ and $\alpha = 0.01$

$$y = x \pm \frac{c_\alpha}{\sqrt{m-1}}, \quad x \in (0,1).$$

6. Generate the values

$$Y_t = \cos\left(2\pi \frac{1}{6} t\right) + \varepsilon_t, \quad t = 1, \ldots, 300,$$

where ε_t are independent and standard normal. Plot the data and the periodogram. Is the hypothesis $Y_t = \varepsilon_t$ rejected at level $\alpha = 0.01$?

7. (Share Data) Test the hypothesis that the share data were generated by independent and identically normal distributed random variables and plot the periodogramm. Plot also the original data.

8. (Kronecker's lemma) Let $(a_j)_{j \geq 0}$ be an absolute summable complexed valued filter. Show that $\lim_{n\to\infty} \sum_{j=0}^{n} (j/n)|a_j| = 0$.

9. The normal distribution $N(0, \sigma^2)$ satisfies

(i) $\int x^{2k+1}\, dN(0,\sigma^2)(x) = 0, \quad k \in \mathbb{N} \cup \{0\}$.
(ii) $\int x^{2k}\, dN(0,\sigma^2)(x) = 1 \cdot 3 \cdot \cdots \cdot (2k-1)\sigma^{2k}, \quad k \in \mathbb{N}$.
(iii) $\int |x|^{2k+1}\, dN(0,\sigma^2)(x) = \frac{2^{k+1}}{\sqrt{2\pi}} k! \sigma^{2k+1}, \quad k \in \mathbb{N} \cup \{0\}$.

10. Show that a $\chi^2(\nu)$-distributed random variable satisfies $\mathrm{E}(Y) = \nu$ and $\mathrm{Var}(Y) = 2\nu$. Hint: Exercise 9.

11. (Slutzky's lemma) Let X, X_n, $n \in \mathbb{N}$, be random variables in $\mathbb{R}$ with distribution functions F_X and F_{X_n}, respectively. Suppose that X_n converges in distribution to X (denoted by $X_n \to_{\mathcal{D}} X$) i.e., $F_{X_n}(t) \to F_X(t)$ for every continuity point of F_X as $n \to \infty$. Let Y_n, $n \in \mathbb{N}$, be another sequence of random variables which converges stochastically to some constant $c \in \mathbb{R}$, i.e., $\lim_{n\to\infty} P\{|Y_n - c| > \varepsilon\} = 0$ for arbitrary $\varepsilon > 0$. This implies

(i) $X_n + Y_n \to_{\mathcal{D}} X + c$.
(ii) $X_n Y_n \to_{\mathcal{D}} cX$.

(iii) $X_n/Y_n \to_{\mathcal{D}} X/c$, if $c \neq 0$.

This entails in particular that stochastic convergence implies convergence in distribution. The reverse implication is not true in general. Give an example.

12. Show that the distribution function F_m of Fisher's test statistic κ_m satisfies under the condition of independent and identically normal observations ε_t

$$F_m(x + \ln(m)) = P\{\kappa_m \leq x + \ln(m)\} \to_{m\to\infty} \exp(-e^{-x}) =: G(x), \quad x \in \mathbb{R}.$$

The limiting distribution G is known as the *Gumbel* distribution. Hence we have $P\{\kappa_m > x\} = 1 - F_m(x) \approx 1 - \exp(-m\, e^{-x})$. Hint: Exercise 2 and Slutzky's lemma.

13. Which effect has an outlier on the periodogram? Check this for the simple model $(Y_t)_{t,\ldots,n}$ ($t_0 \in \{1, \ldots, n\}$)

$$Y_t = \begin{cases} \varepsilon_t, & t \neq t_0 \\ \varepsilon_t + c, & t = t_0, \end{cases}$$

where the ε_t are independent and identically normal $N(0, \sigma^2)$-distributed and $c \neq 0$ is an arbitrary constant. Show to this end

$$\begin{aligned} \mathrm{E}(I_Y(k/n)) &= \mathrm{E}(I_\varepsilon(k/n)) + c^2/n \\ \mathrm{Var}(I_Y(k/n)) &= \mathrm{Var}(I_\varepsilon(k/n)) + 2c^2\sigma^2/n, \quad k = 1, \ldots, [(n-1)/2]. \end{aligned}$$

14. Suppose that $U_1, \ldots, U_n$ are uniformly distributed on $(0, 1)$ and let $\hat{F}_n$ denote the pertaining empirical distribution function. Show that

$$\sup_{0\leq x\leq 1} |\hat{F}_n(x) - x| = \max_{1\leq k\leq n} \left\{ \max\left\{ U_{k:n} - \frac{(k-1)}{n}, \frac{k}{n} - U_{k:n} \right\} \right\}.$$

15. (Monte Carlo Simulation) For m large we have under the hypothesis

$$P\{\sqrt{m-1}\Delta_{m-1} > c_\alpha\} \approx \alpha.$$

For different values of $m\,(> 30)$ generate 1000 times the test statistic $\sqrt{m-1}\Delta_{m-1}$ based on independent random variables and check, how often this statistic exceeds the critical values $c_{0.05} = 1.36$ and $c_{0.01} = 1.63$. Hint: Exercise 14.

16. In the situation of Theorem 5.2.4 show that the spectral density f of $(Y_t)_t$ is continuous.

17. Complete the proof of Theorem 5.2.4 (ii) for the remaining cases $\lambda = \mu = 0$ and $\lambda = \mu = 0.5$.

18. Verify that the weights (5.13) defined via a kernel function satisfy the conditions (5.11).

19. Use the `IML` function `ARMASIM` to simulate the process

$$Y_t = 0.3Y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}, \quad 1 \le t \le 150,$$

where ε_t are independent and standard normal. Plot the periodogram and estimates of the log spectral density together with confidence intervals. Compare the estimates with the log spectral density of $(Y_t)_{t\in\mathbb{Z}}$.

20. Compute the distribution of the periodogram $I_\varepsilon(1/2)$ for independent and identically normal $N(0,\sigma^2)$-distributed random variables $\varepsilon_1, \ldots, \varepsilon_n$ in case of an even sample size n.

21. Suppose that Y follows a gamma distribution with parameters p and b. Calculate the mean and the variance of Y.

22. Compute the length of the confidence interval $C_{\nu,\alpha}(k/n)$ for fixed α (preferably $\alpha = 0.05$) but for various ν. For the calculation of ν use the weights generated by the kernel $K(x) = 1 - |x|$, $-1 \le x \le 1$ (see equation (5.13).

23. Show that for $y \in \mathbb{R}$

$$\frac{1}{n}\Big|\sum_{t=0}^{n-1} e^{i2\pi yt}\Big|^2 = \sum_{|t|<n}\Big(1 - \frac{|t|}{n}\Big)e^{i2\pi yt} = K_n(y),$$

where

$$K_n(y) = \begin{cases} n, & y \in \mathbb{Z} \\ \frac{1}{n}\Big(\frac{\sin(\pi y n)}{\sin(\pi y)}\Big)^2, & y \notin \mathbb{Z}, \end{cases}$$

is the Fejer kernel of order n. Verify that it has the properties

(i) $K_n(y) \ge 0$,

(ii) the Fejer kernel is a periodic function of period length one,

(iii) $K_n(y) = K_n(-y)$,

(iv) $\int_{-0.5}^{0.5} K_n(y)\,dy = 1$,

(v) $\int_{-\delta}^{\delta} K_n(y)\,dy \to_{n\to\infty} 1, \quad \delta > 0.$

24. (Nile Data) Between 715 and 1284 the river Nile had its lowest annual minimum levels. These data are among the longest time series in hydrology. Can the trend removed Nile Data be considered as being generated by a white noise, or are these hidden periodicities? Estimate the spectral density in this case. Use discrete spectral estimators as well as lag window spectral density estimators. Compare with the spectral density of an $AR(1)$-process.

References

Andrews, D.W.K. (1991). Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York.

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction.* John Wiley, New York.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307–327.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations (with discussion). *J.R. Stat. Sob. B.* **26**, 211–252.

Box, G.E.P. and Jenkins, G.M. (1976). *Times Series Analysis: Forecasting and Control.* Rev. ed. Holden–Day, San Francisco.

Box, G.E.P. and Pierce, D.A. (1970). Distribution of residual autocorrelation in autoregressive-integrated moving average time series models. *J. Amex. Statist. Assoc.* **65**, 1509–1526.

Box, G.E.P. and Tiao, G.C. (1977). A canonical analysis of multiple time series. *Biometrika* **64**, 355–365.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* 2nd ed. Springer Series in Statistics, Springer, New York.

Brockwell, P.J. and Davis, R.A. (2002). *Introduction to Time Series and Forecasting.* 2nd ed. Springer Texts in Statistics, Springer, New York.

Cleveland, W.P. and Tiao, G.C. (1976). Decomposition of seasonal time series: a model for the census X–11 program. *J. Amex. Statist. Assoc.* **71**, 581–587.

Conway, J.B. (1975). *Functions of One Complex Variable,* 2nd. printing. Springer, New York.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* **50**, 987–1007.

Engle, R.F. and Granger, C.W.J. (1987). Co–integration and error correction: representation, estimation and testing. *Econometrica* **55**, 251–276.

Falk, M., Marohn, F. and Tewes, B. (2002). *Foundations of Statistical Analyses and Applications with SAS.* Birkhäuser, Basel.

Fan, J. and Gijbels, I. (1996). *Local Polynominal Modeling and Its Application — Theory and Methodologies.* Chapman and Hall, New York.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications.* Vol. II, 2nd. ed. John Wiley, New York.

Fuller, W.A. (1976). *Introduction to Statistical Time Series.* John Wiley, New York.

Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *J. Econometrics* **16**, 121–130.

Hamilton, J.D. (1994). *Time Series Analysis.* Princeton University Press, Princeton.

Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an autoregression. *J.R. Statist. Soc. B* **41**, 190–195.

Janacek, G. and Swift, L. (1993). *Time Series. Forecasting, Simulation, Applications.* Ellis Horwood, New York.

Kalman, R.E. (1960). A new approach to linear filtering. *Trans. ASME J. of Basic Engineering* **82**, 35–45.

Kendall, M.G. and Ord, J.K. (1993). *Time Series.* 3rd ed. Arnold, Sevenoaks.

Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303.

Murray, M.P. (1994). A drunk and her dog: an illustration of cointegration and error correction. *The American Statistician* **48**, 37–39.

Newton, H.J. (1988). *Timeslab: A Time Series Analysis Laboratory.* Brooks/Cole, Pacific Grove.

Parzen, E. (1957). On consistent estimates of the spectrum of the stationary time series. *Ann. Math. Statist.* **28**, 329–348.

Phillips, P.C. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica* **58**, 165–193.

Phillips, P.C. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* **75**, 335–346.

Quenouille, M.H. (1957). *The Analysis of Multiple Time Series.* Griffin, London.

Reiss, R.D. (1989). *Approximate Distributions of Order Statistics. With Applications to Nonparametric Statistics.* Springer Series in Statistics, Springer, New York.

Rudin, W. (1974). *Real and Complex Analysis,* 2nd. ed. McGraw-Hill, New York.

SAS Institute Inc. (1992). *SAS Procedures Guide, Version 6, Third Edition.* SAS Institute Inc. Cary, N.C.

Shiskin, J. and Eisenpress, H. (1957). Seasonal adjustment by electronic computer methods. *J. Amer. Statist. Assoc.* **52**, 415–449.

Shiskin, J., Young, A.H. and Musgrave, J.C. (1967). The X–11 variant of census method II seasonal adjustment program. Technical paper no. **15**, Bureau of the Census, U.S. Dept. of Commerce.

Simonoff, J.S. (1966). *Smoothing Methods in Statistics.* Springer Series in Statistics, Springer, New York.

Tintner, G. (1958). Eine neue Methode für die Schätzung der logistischen Funktion. *Metrika* **1**, 154–157.

Tukey, J. (1949). The sampling theory of power spectrum estimates. Proc. Symp. on Applications of Autocorrelation Analysis to Physical Problems, NAVEXOS-P-735, Office of Naval Research, Washington, 47–67.

Wallis, K.F. (1974). Seasonal adjustment and relations between variables. *J. Amer. Statist. Assoc.* **69**, 18–31.

Wei, W.W.S. (1990). *Time Series Analysis. Unvariate and Multivariate Methods.* Addison–Wesley, New York.

Index

SAS-Index

Appendix A

GNU Free Documentation License

Version 1.2, November 2002

59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The **"Document"**, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as **"you"**. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A **"Modified Version"** of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A **"Secondary Section"** is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The **"Invariant Sections"** are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The **"Cover Texts"** are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A **"Transparent"** copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of

formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called **"Opaque"**.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The **"Title Page"** means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section **"Entitled XYZ"** means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as **"Acknowledgements"**, **"Dedications"**, **"Endorsements"**, or **"History"**.) To **"Preserve the Title"** of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five

of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright ©YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.