

Zeitreihen und Prognosen

SS 2018

Irina Penner

penner@htw-berlin.de

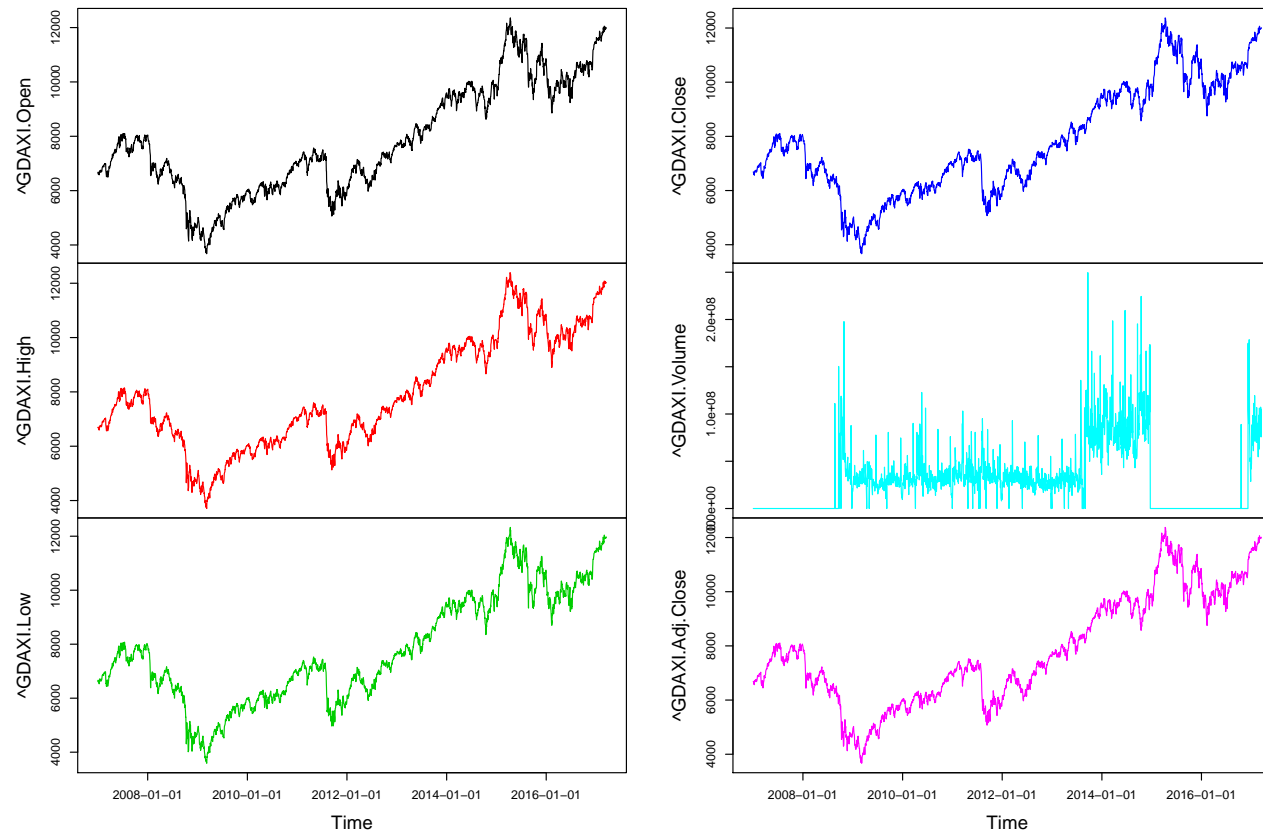
Worum geht es?

Eine **Zeitreihe** ist eine zeitlich geordnete Folge von Beobachtungen eines quantitativen Merkmals:

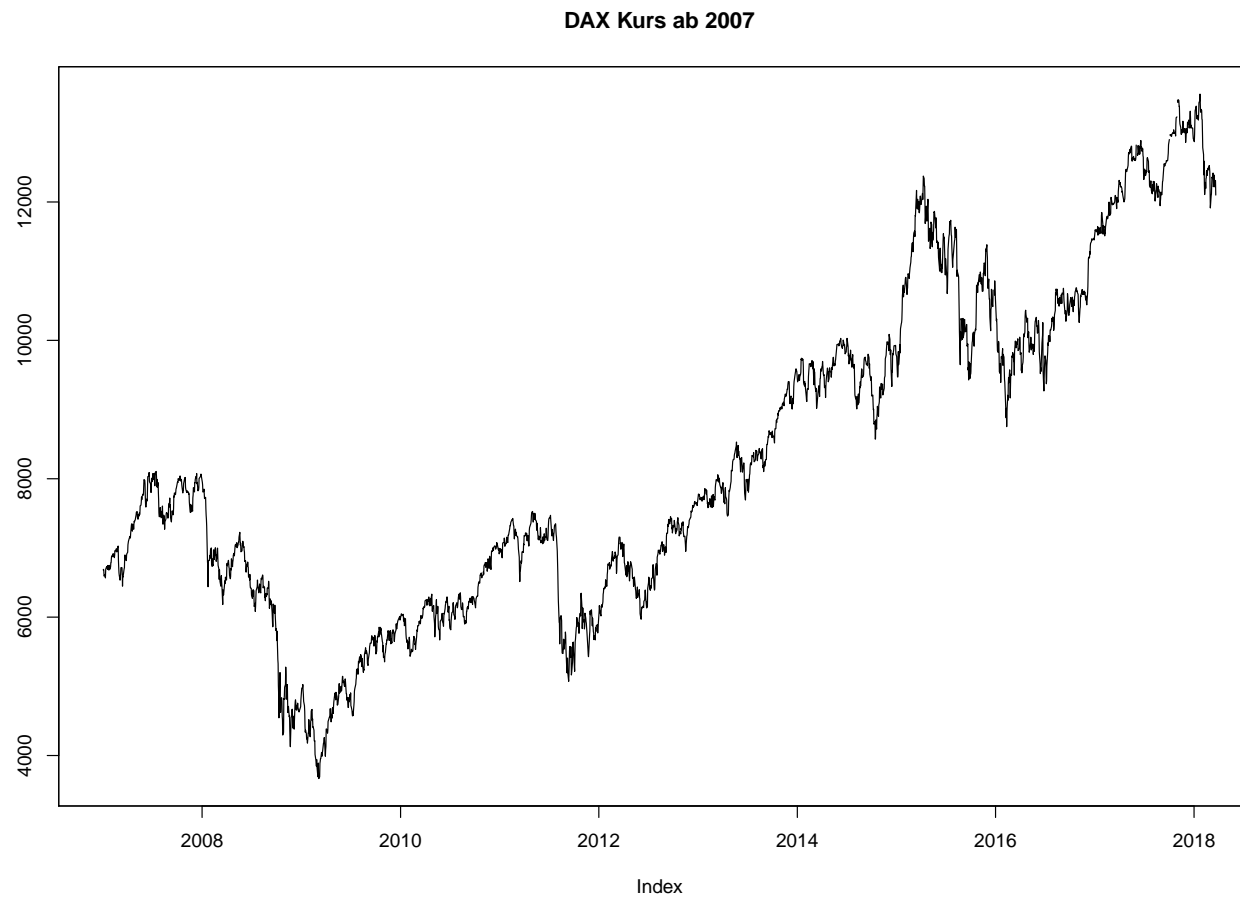
$$(x_t)_{t \in \mathbb{T}}, \quad x_t = \text{Beobachtung zur Zeit } t$$

Die **Zeitindexmenge** \mathbb{T} kann dabei diskret ($\mathbb{T} \subseteq \mathbb{N}$) oder stetig ($\mathbb{T} \subseteq \mathbb{R}$) sein, man spricht dann von einer diskreten oder einer stetigen Zeitreihe.

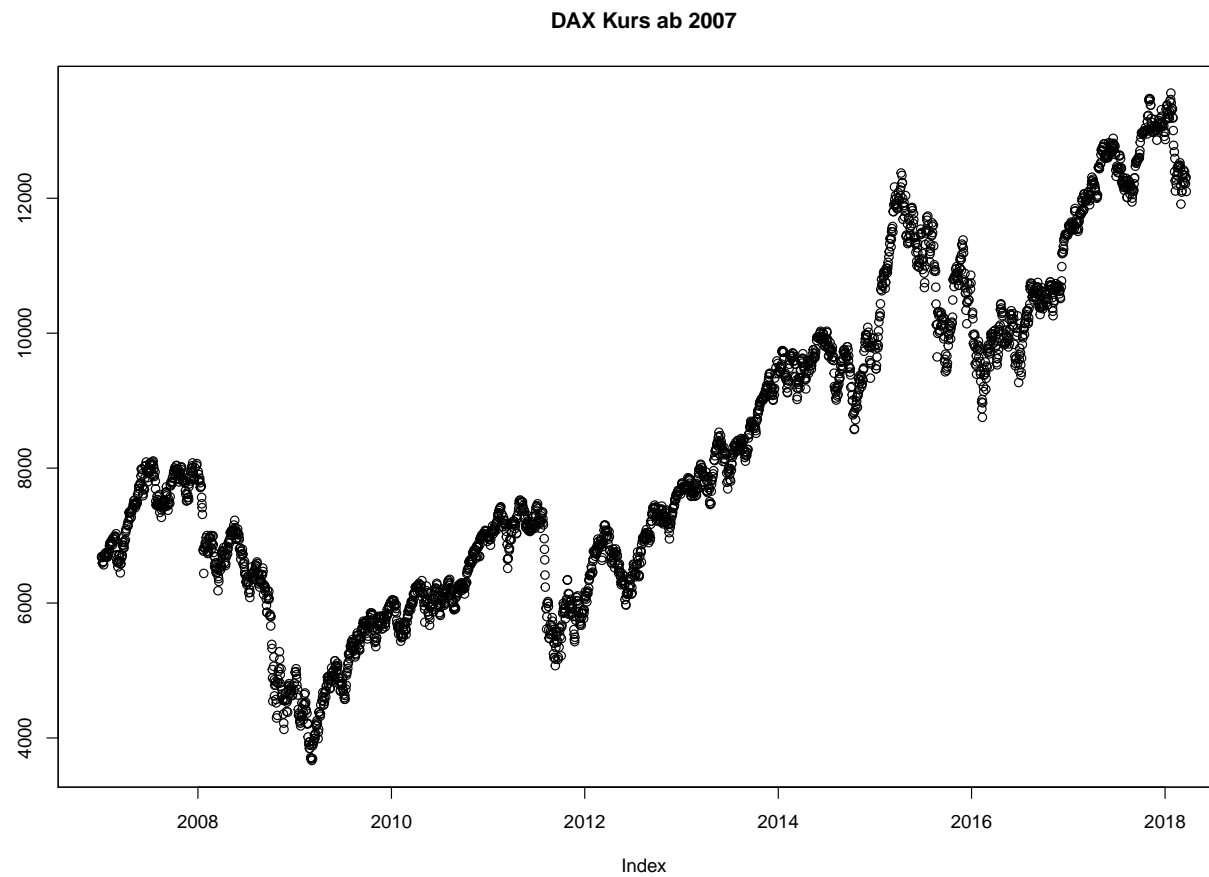
Einige Beispiele



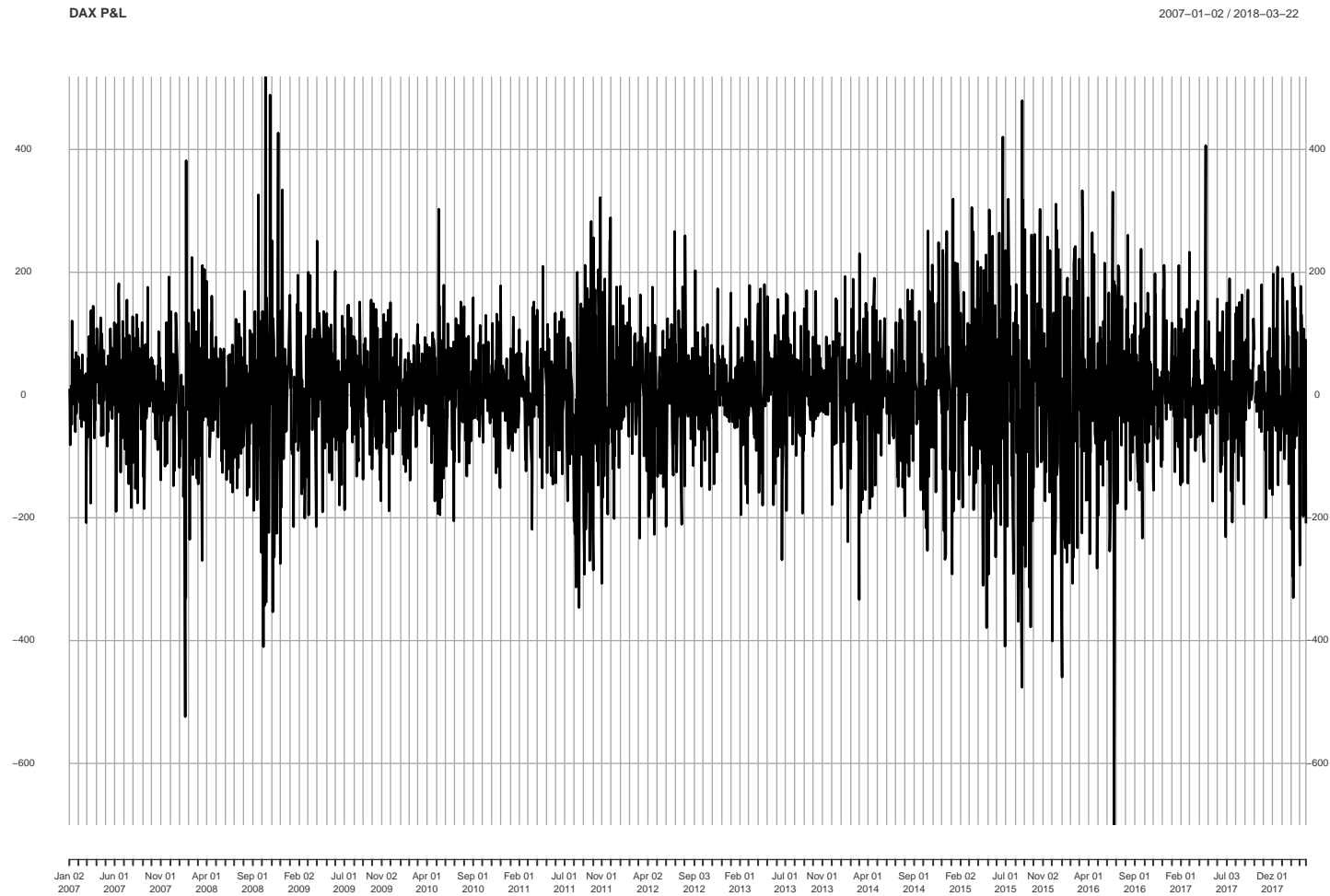
Einige Beispiele



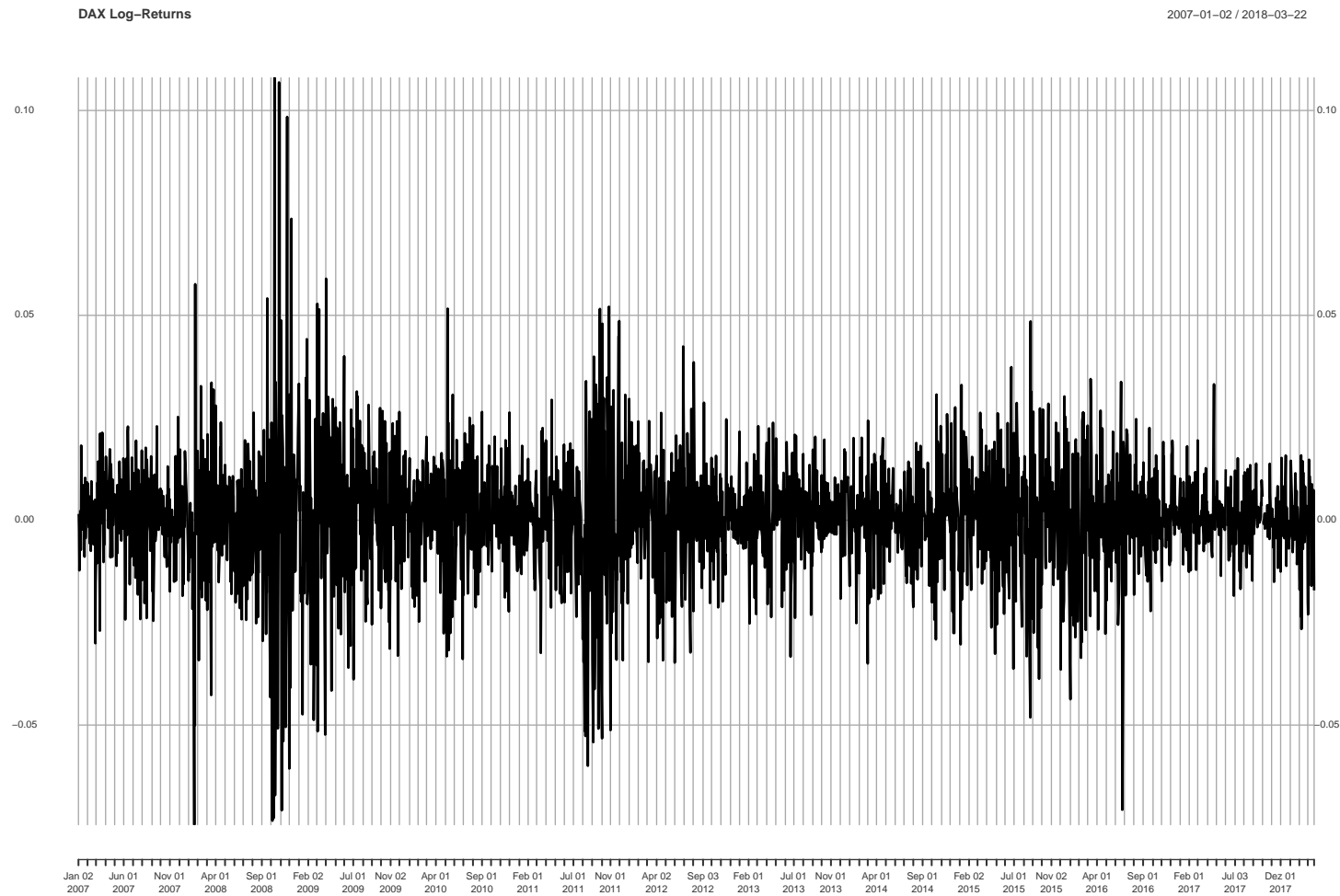
Einige Beispiele



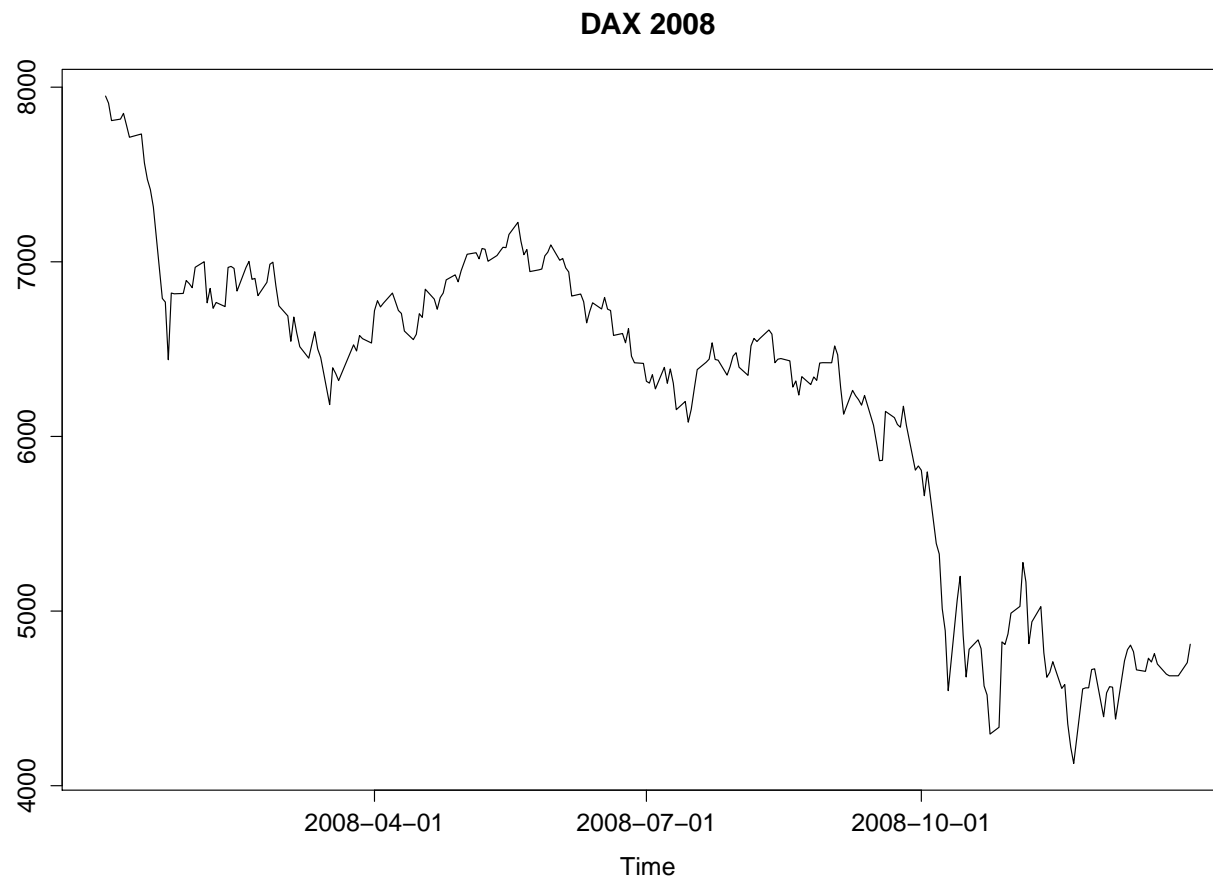
Einige Beispiele



Einige Beispiele



Einige Beispiele



Einige Beispiele

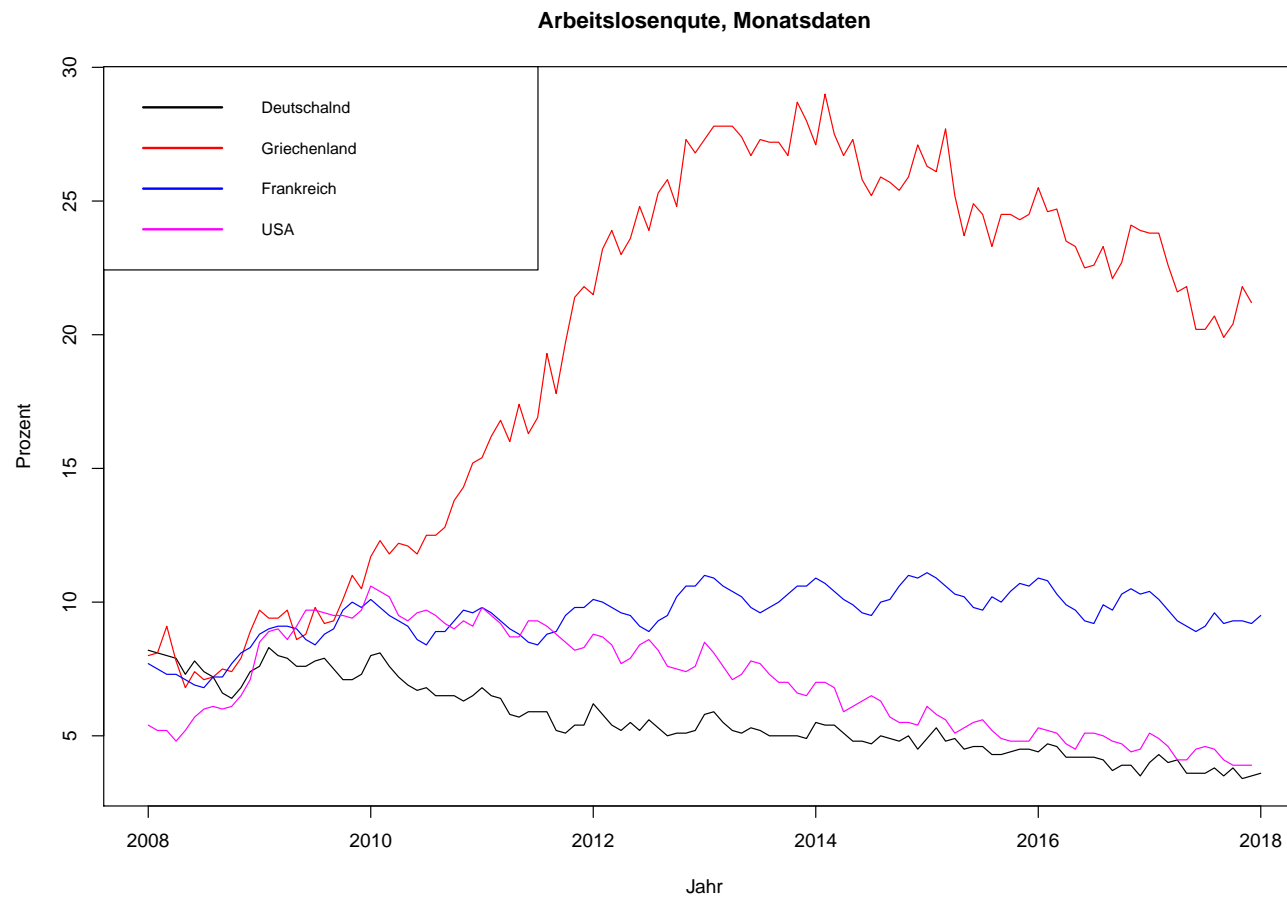
```
library(quantmod)
getSymbols("^GDAXI", src = "yahoo")
View(GDAXI)
head(GDAXI)
plot(as.zoo(GDAXI), main="DAX ab 2007")

DAX07=GDAXI[,6]
plot(DAX07, main="DAX Kurs ab 2007", ylab="")
plot(DAX07, main="DAX Kurs ab 2007", ylab="", type="p")

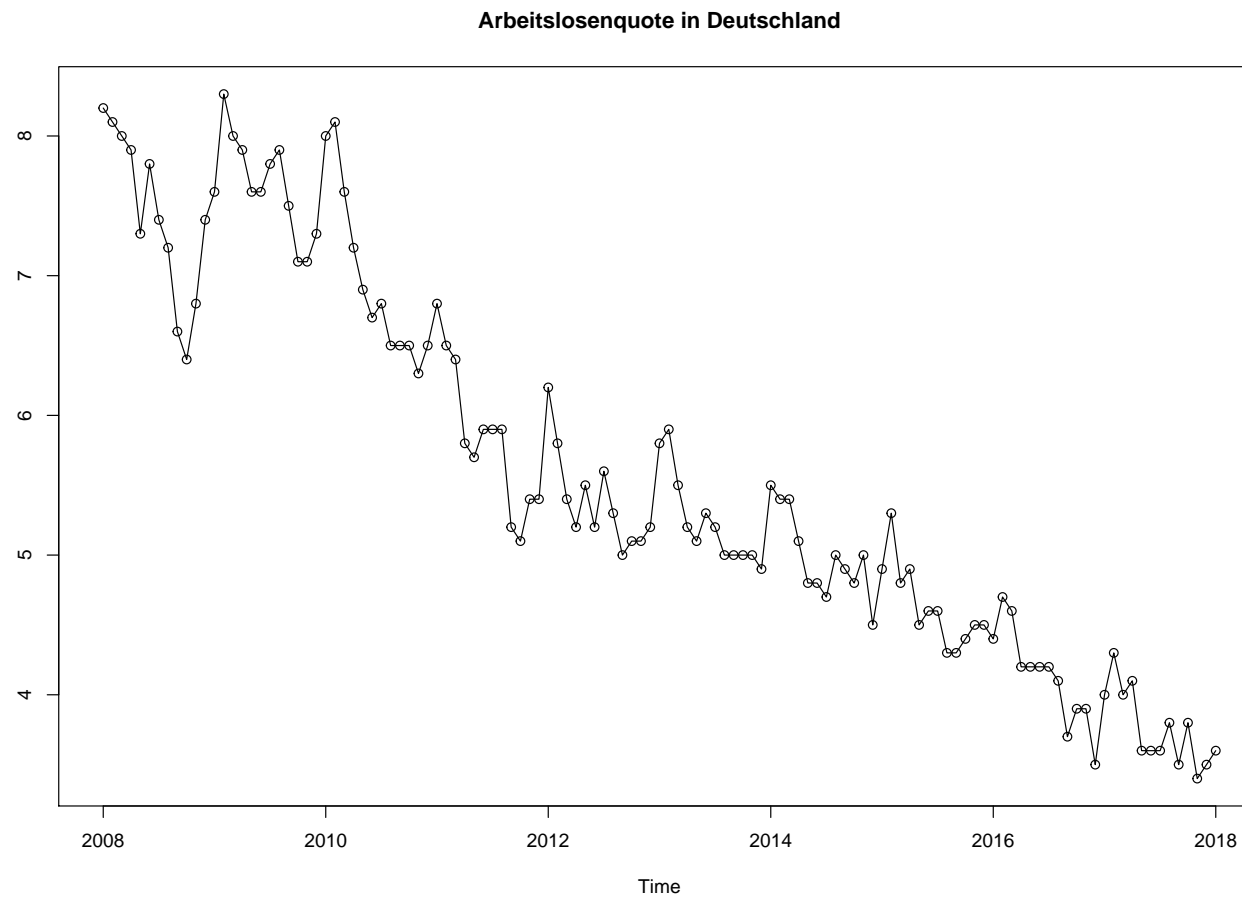
PnL=diff(DAX07)
plot(PnL, main="DAX P&L", ylab="")
lretDAX=diff(log(DAX07))
plot(lretDAX, main="DAX Log>Returns", ylab="")

DAX08=window(DAX07, start="2008-01-01", end="2008-12-31")
plot(DAX08, ylab="", main="DAX 2008")
```

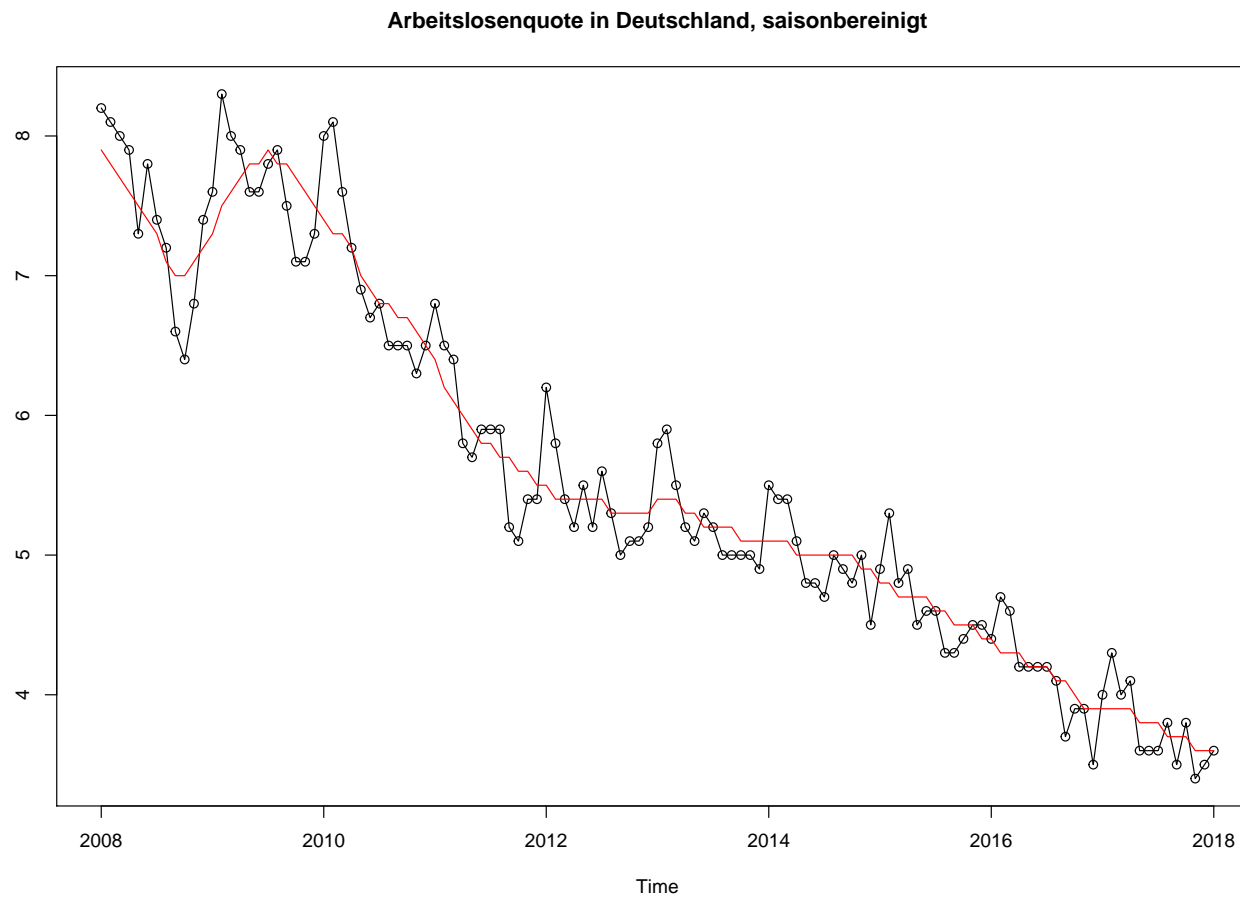
Einige Beispiele



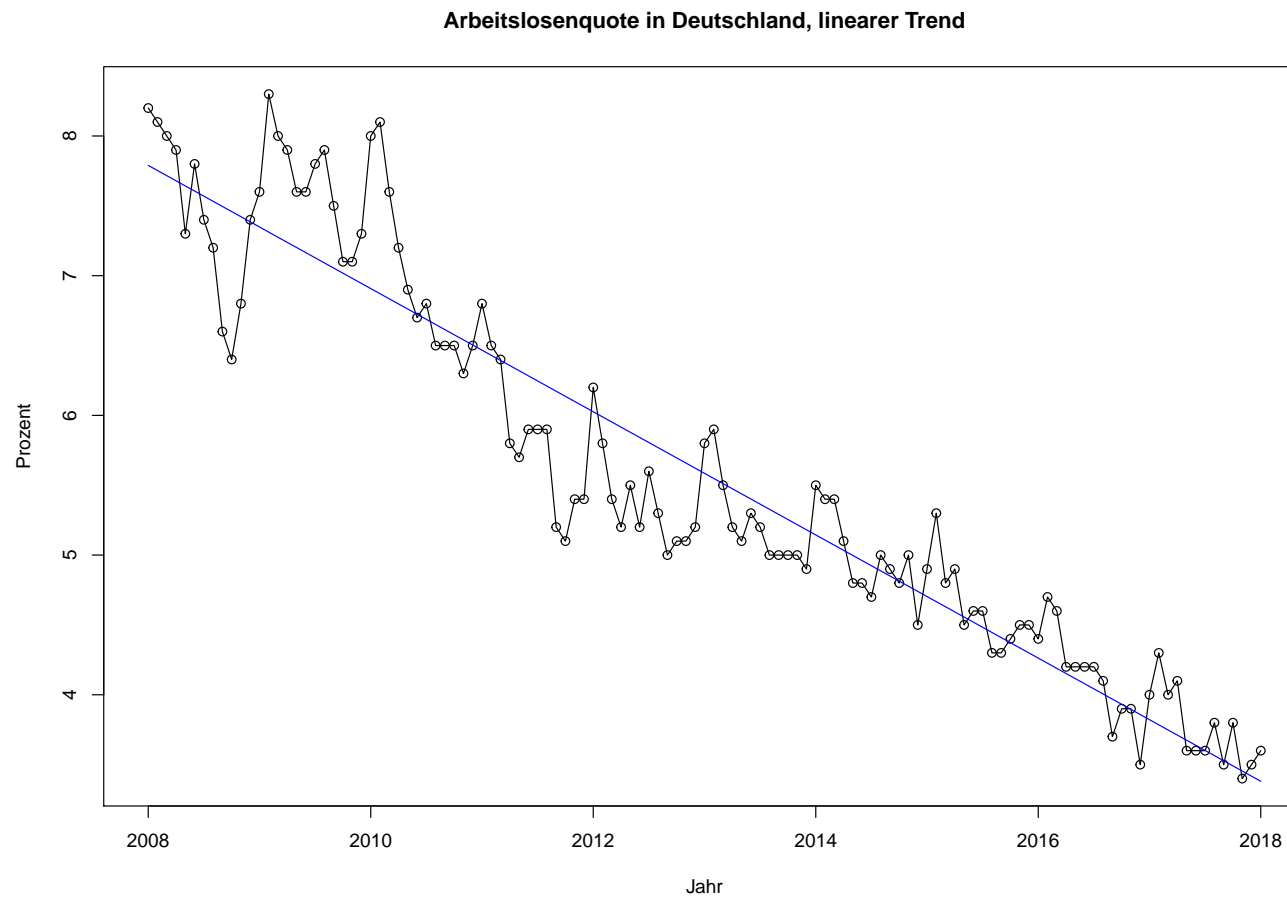
Einige Beispiele



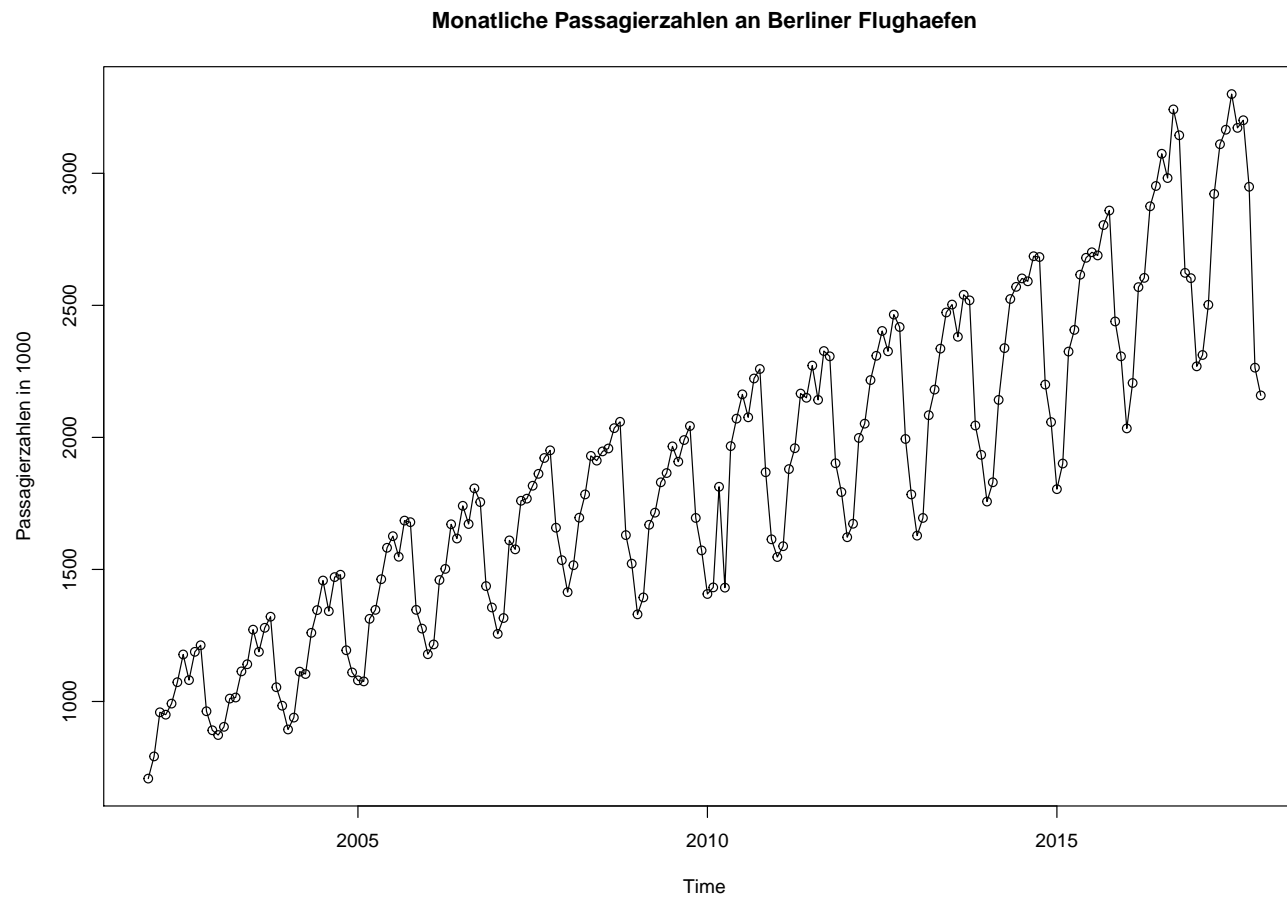
Einige Beispiele



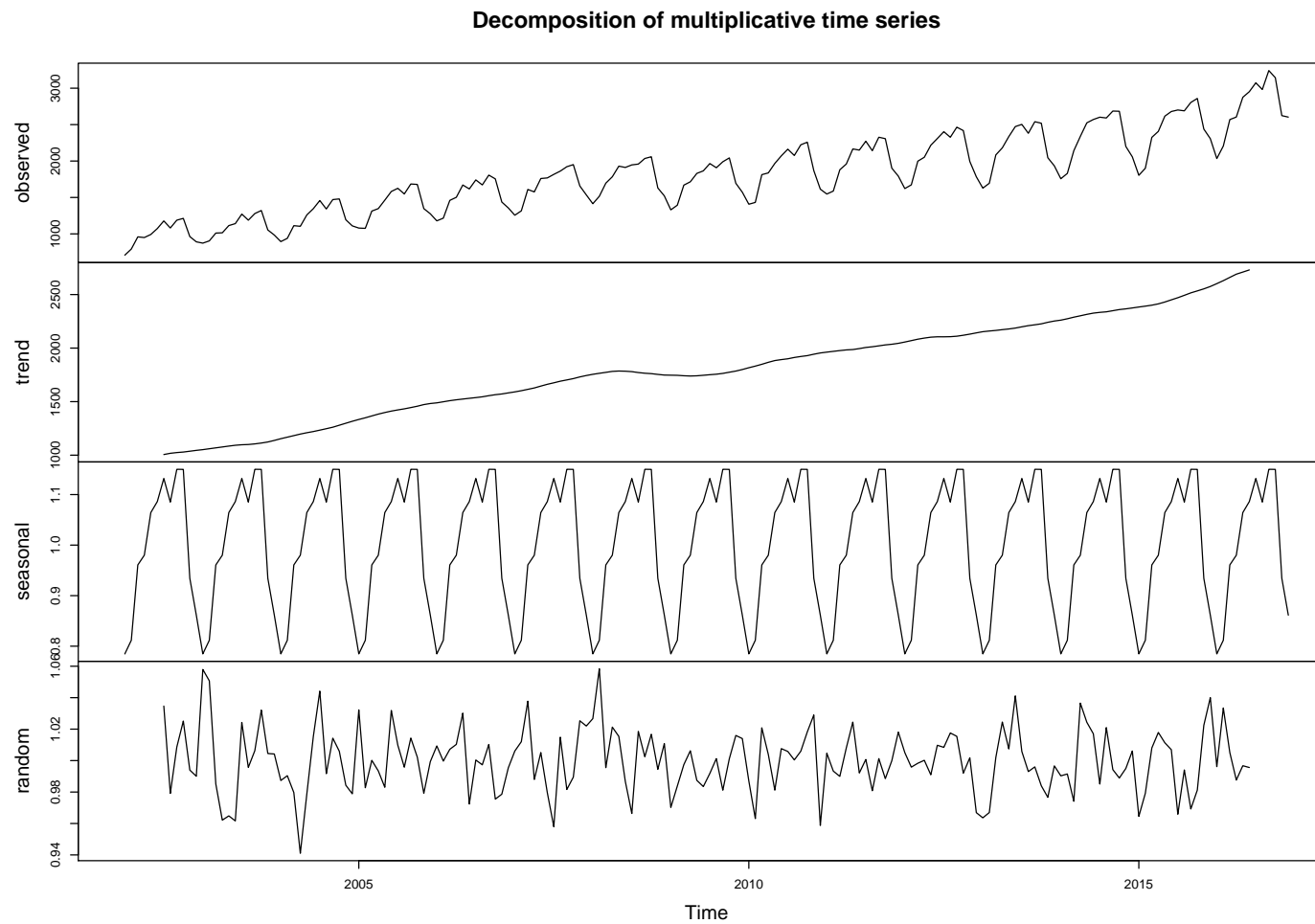
Einige Beispiele



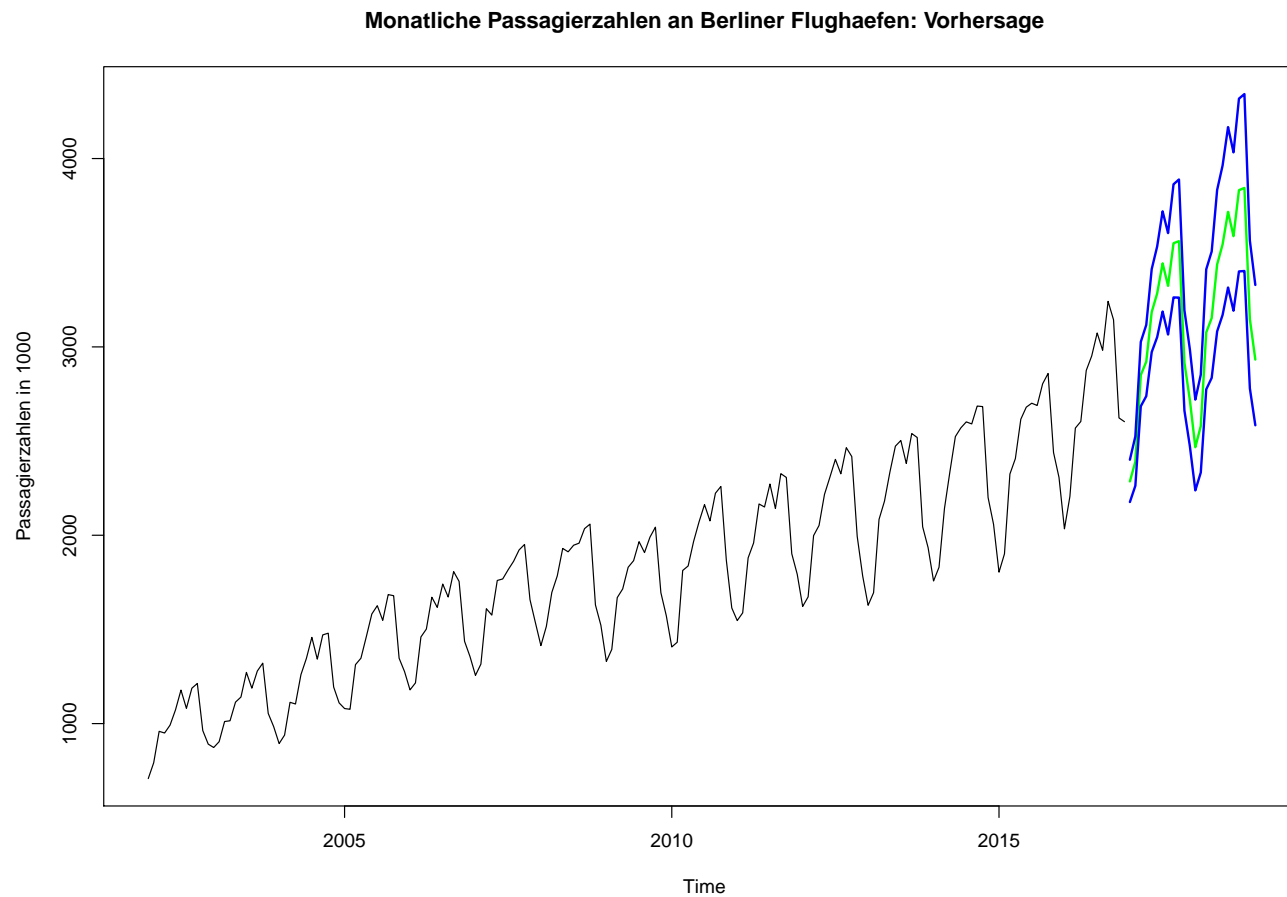
Einige Beispiele



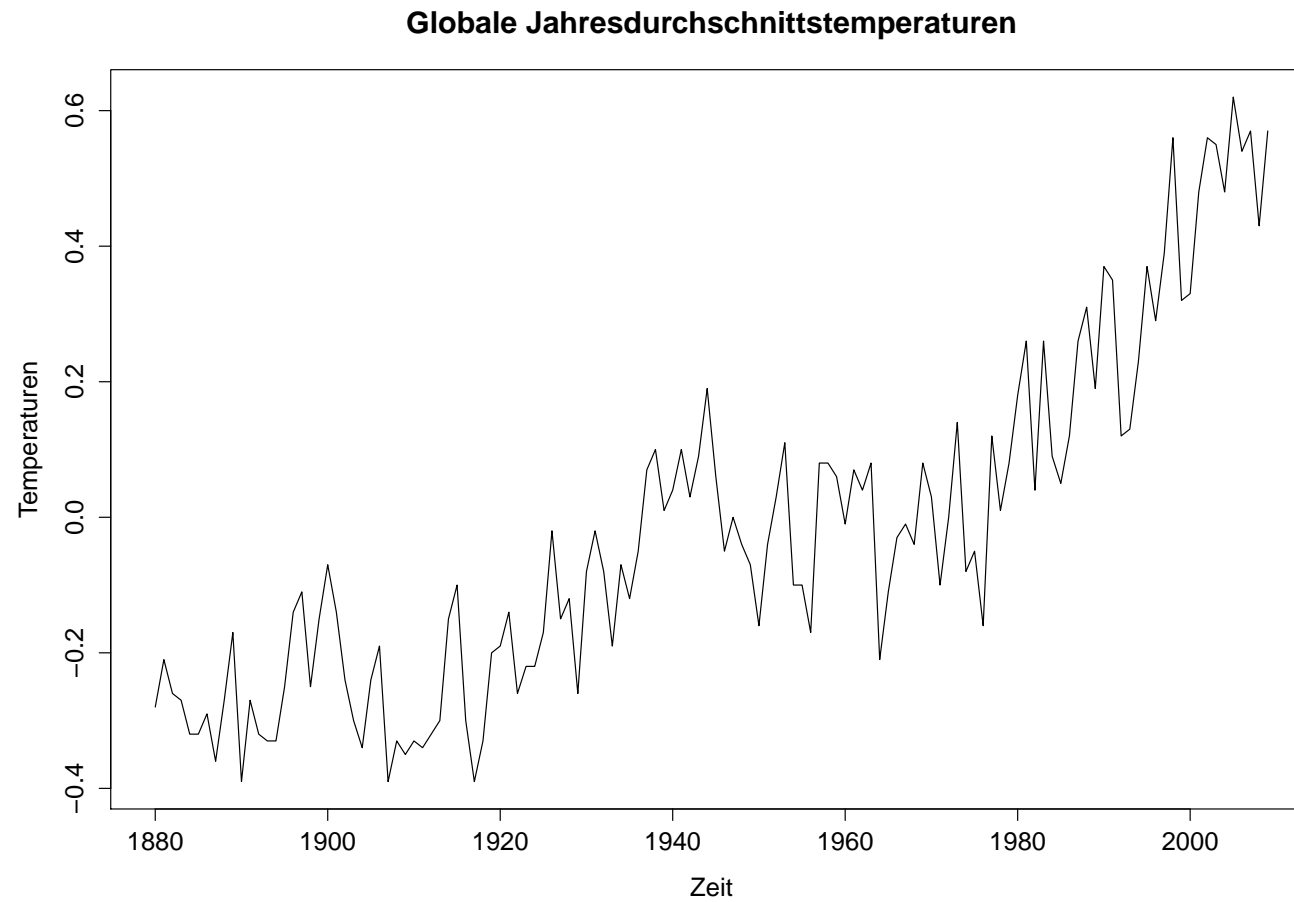
Einige Beispiele



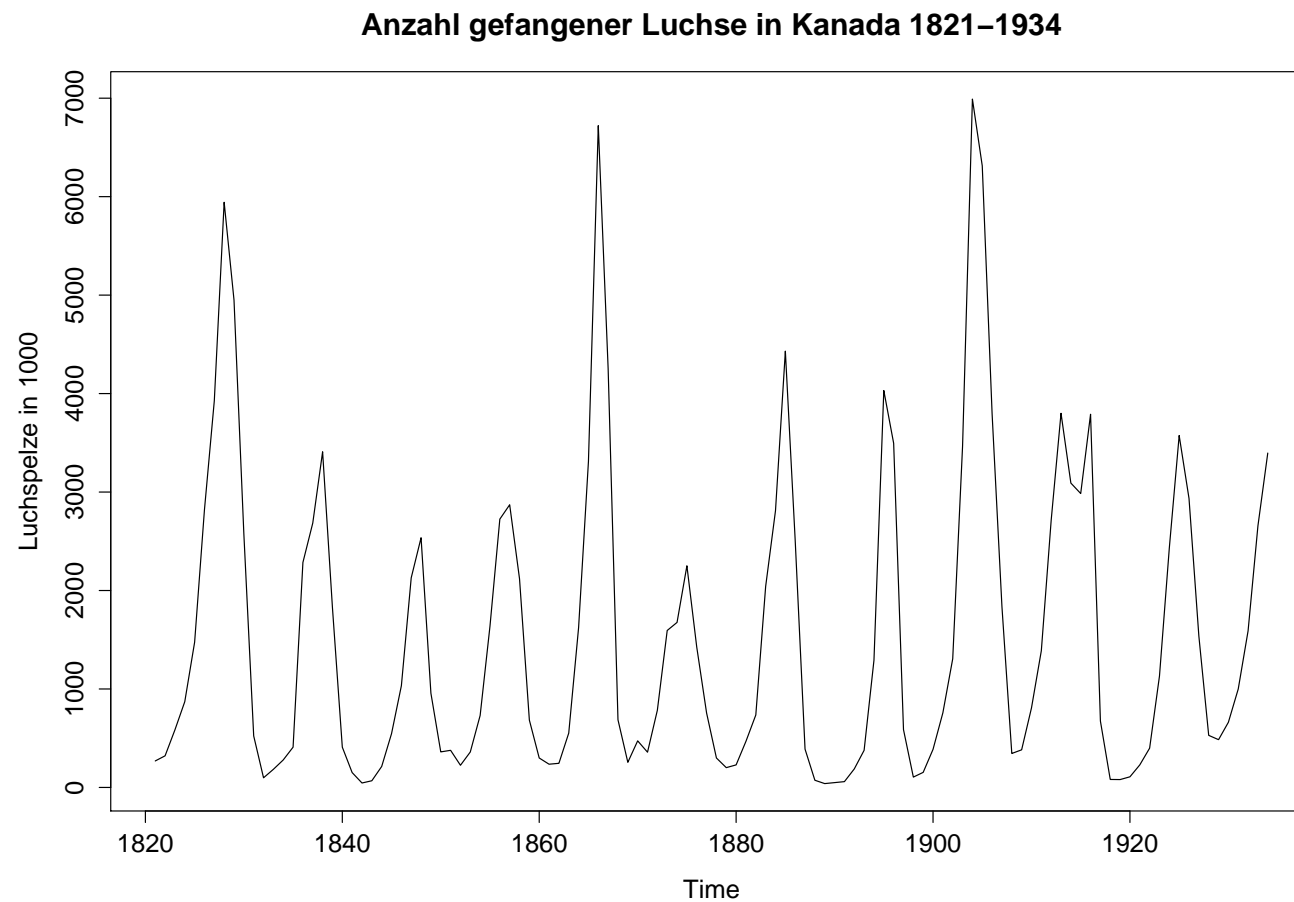
Einige Beispiele



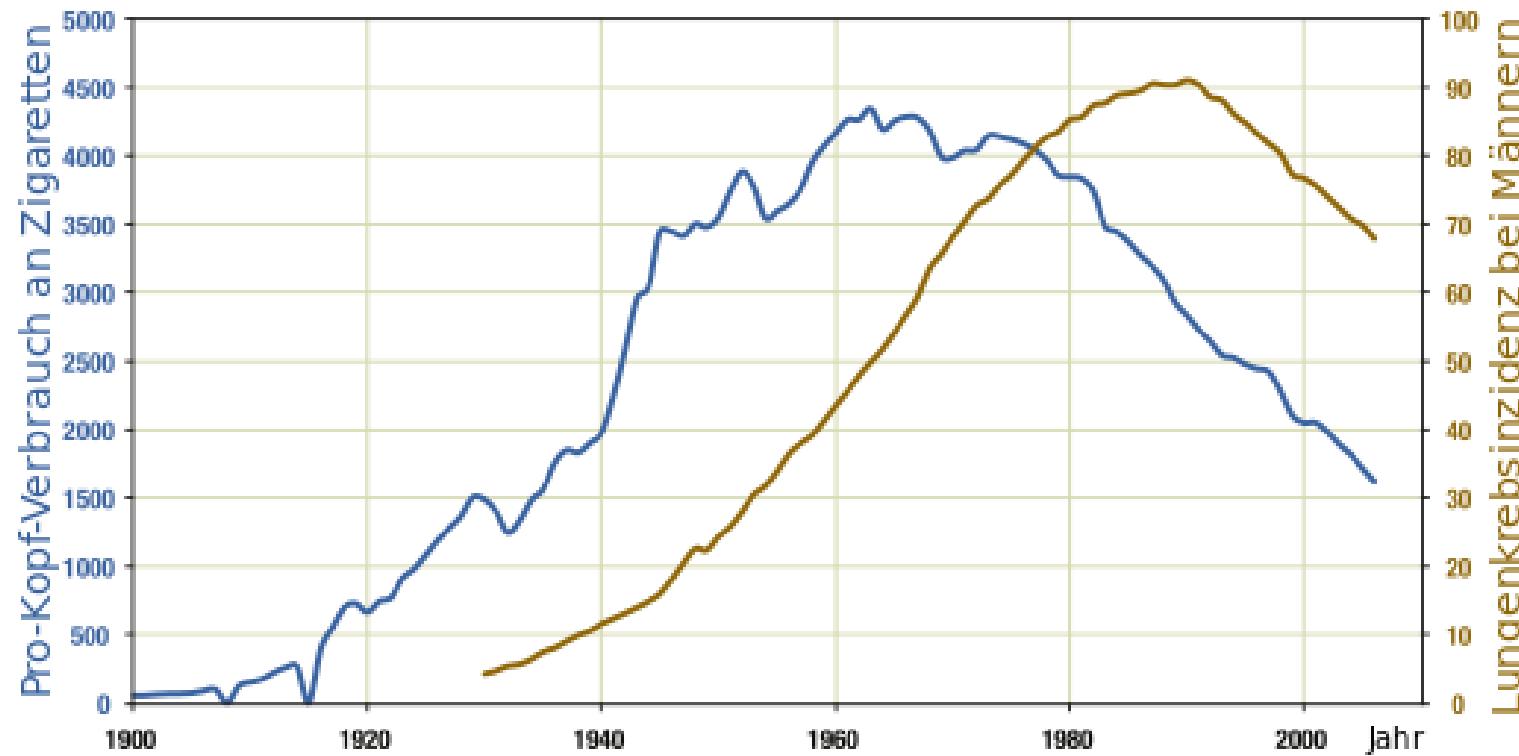
Einige Beispiele



Einige Beispiele



Einige Beispiele



Gliederung der Vorlesung

1) Einführung und Grundlagen

2) Klassische Zeitreihenanalyse

- a) Trendbestimmung
- b) Glättung von Zeitreihen durch Filter
- c) Saisonbereinigung
- d) Exponentielles Glätten, Prognosen

3) Modellierung von Reihen durch stochastische Prozesse

- a) Autoregressive (AR) Modelle
- b) Moving Average (MA) Modelle
- c) ARMA Modelle
- d) ARIMA Modelle

Literatur

- R.H. Shumway & D.S. Stoffer: “Time Series Analysis and Its Applications: With R Examples”
frei verfügbar unter
<http://www.stat.pitt.edu/stoffer/tsa4/tsaEZ.pdf>
dort auch ein R-Tutorium
http://www.stat.pitt.edu/stoffer/tsa4/R_toot.htm
oder als Anhang im Buch
- R. Schlittgen & B.H.J. Streitberg: “Zeitreihenanalyse”
- R. Schlittgen: “Angewandte Zeitreihenanalyse mit R”
- P.J. Brockwell und R.A. Davis: “Time series: theory and methods”
- W. Stier: “Methoden der Zeitreihenanalyse”

R

- R: <http://www.r-project.org>
- R-Studio: <http://www.rstudio.com>
- J. Groß & B. Peters: “R Reader”, <https://cran.r-project.org/doc/contrib/Grosz+Peters-R-Reader.pdf>
- R-Pakete:

```
getOption("defaultPackages")    #immer dabei
installed.packages()[,1:3]       #Ueberblick ueber installierte Pakete
update.packages(ask=FALSE)       #Updaten (solle man regelmaessig machen)

install.packages("astsa")
install.packages("fImport")
library(astsa)                  #laden
require(fImport)                #dasselbe wie library
library(help=astsa)              #Dokumentation
```

Datenquellen

- Viele Datasets mitgeliefert in R und Paket “astsa”, z.B.
 - EuStockMarkets: Tägliche Schlusskurse von DAX, SMI, CAC und FTSE (1991-1998)
 - Globale Temperatur (gtemp), Luchse (lynx) usw...

Übersicht mit `data()`

abrufen direkt mit Namen, z.B. `plot(EuStockMarkets)`

- Börsenkurse auf Yahoo Finance:
<https://de.finance.yahoo.com/> oder Onvista
<http://www.onvista.de/>
(als .csv speichern oder direkt einlesen mit Paket “quantmod”)

Datenquellen

- Eurostat <http://ec.europa.eu/eurostat/de/data/database>
- Statistisches Bundesamt, GENESIS-Datenbank
<https://www-genesis.destatis.de/genesis/online/logon>
- Amt für Statistik Berlin Brandenburg
<https://www.statistik-berlin-brandenburg.de/Statistiken/inhalt-statistiken.asp>
- St. Louis FED: <http://research.stlouisfed.org/fred2/>
- EZB, Bundesbank, OECD, IWF, . . .

(als .csv oder .txt speichern und einlesen mit [read.csv](#), [read.csv2](#) oder [read.table](#))

Materialien, Klausur

Materialien zur Vorlesung → **moodle** Die Lernplattform der HTW Berlin

Es gibt eine 90-minütige Abschlussklausur. Zeit und Ort werden bekanntgegeben. Zugelassen sind:

- ein beidseitig beschriebenes DIN A4 Blatt
- ein nicht-programmierbarer Taschenrechner.



Aufbau: Vorlesung und Übung; Hausaufgaben. Es gibt Aufgaben* mit 3-5 Punkten (meistens Arbeiten mit R). Durch Vorführen dieser Aufgaben kann man maximal 20% der Klausurpunkte sicherstellen.

Stochastische Prozesse

Ein **stochastischer Prozess** ist eine Folge $(X_t)_{t \in \mathbb{T}}$ von Zufallsvariablen.

Der Index \mathbb{T} (bei uns meistens $\mathbb{T} \subseteq \mathbb{N}$, aber auch $\mathbb{T} \subseteq \mathbb{Z}$ oder $\mathbb{T} \subseteq \mathbb{R}$ möglich) wird als Zeit aufgefasst.

Die Folge x_1, \dots, x_N von **Realisierungen (Zeitpfaden, Trajektorien)** eines Abschnitts von (X_t) definiert eine **Zeitreihe**:

$x_t := X_t(\omega) = \text{Beobachtung zur Zeit } t \text{ beim Szenario } \omega, \quad t = 1, \dots, N.$

Ein stochastischer Prozess beschreibt also das zugrundeliegende **Modell** einer Zeitreihe.

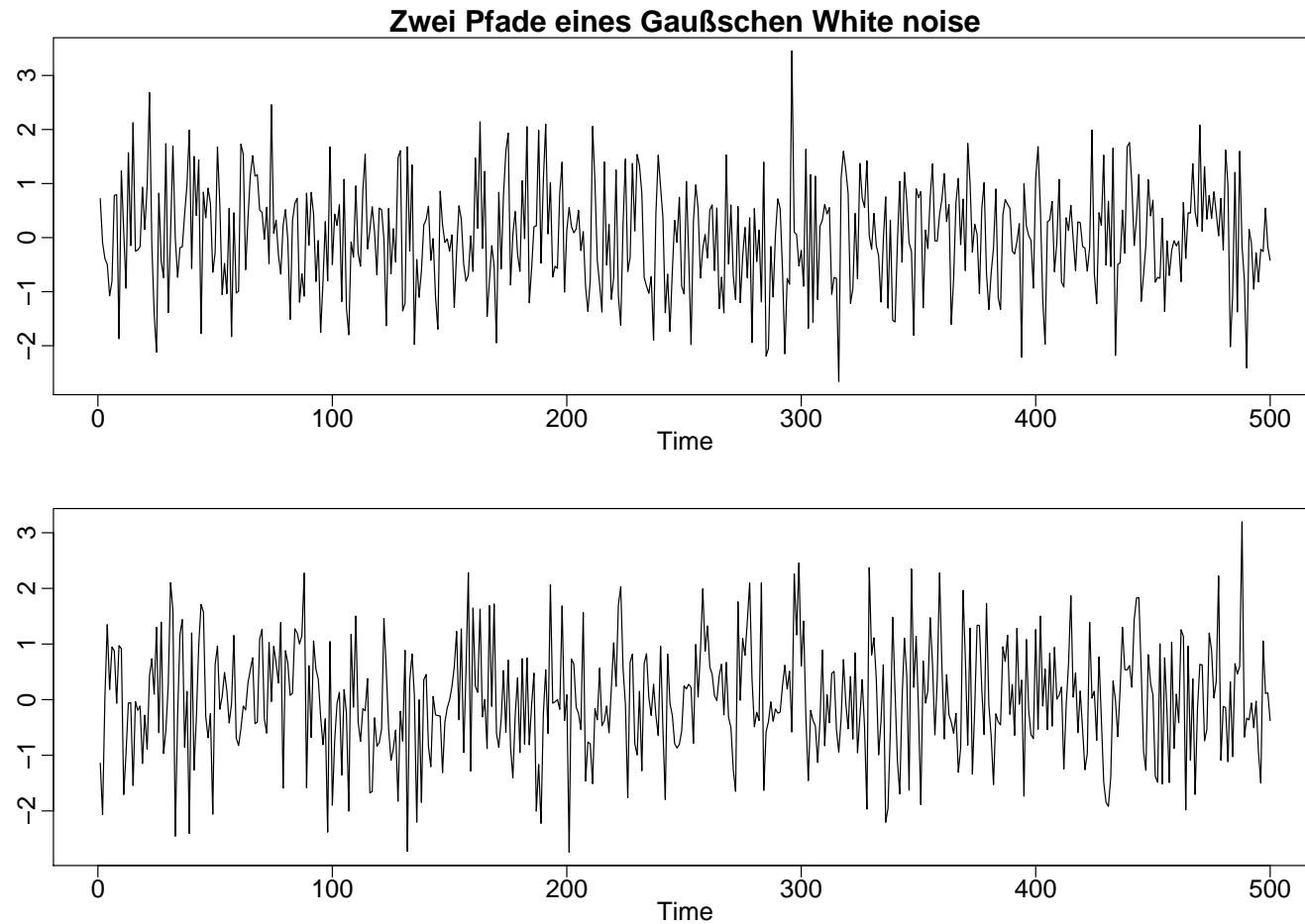
Beispiele stochastischer Prozesse

- Ein stochastischer Prozess (X_t) bei dem die Variablen X_{t_1}, \dots, X_{t_k} für alle möglichen Indizes t_1, \dots, t_k gemeinsam (multivariat) normalverteilt sind, heißt ein **Gaußprozess**.
- Ein Folge (ε_t) von **unabhängigen identisch verteilten (iid)** Zufallsvariablen mit endlichem Erwartungswert und endlicher Varianz heißt ein (striktes) **White Noise** (weißes Rauschen, reiner Zufallsprozess). Oft wird angenommen, dass die Zufallsvariablen zentriert sind, d.h. $\mu = \mathbb{E}[\varepsilon_t] = 0$. Ist das nicht der Fall, so definiert $\tilde{\varepsilon}_t := \varepsilon_t - \mu$ ($t \in \mathbb{T}$) ein White Noise mit Erwartungswert 0.
- Manchmal verlangt man auch nur, dass die Zufallsvariablen (ε_t) **unkorreliert** sind, den **gleichen** endlichen **Erwartungswert** haben, und die **gleiche** endliche **Varianz** (\rightarrow **schwaches White Noise**).

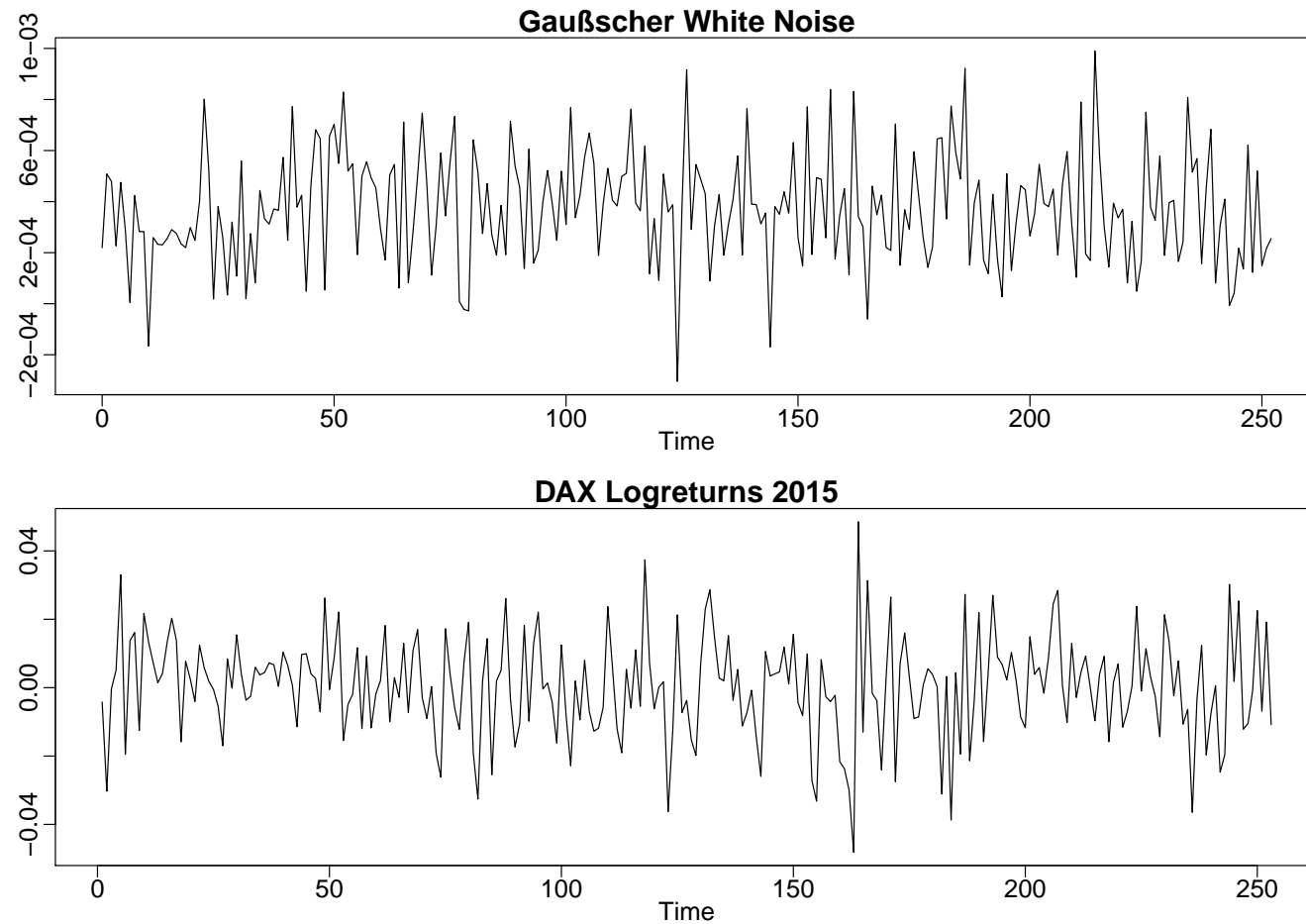
Beispiele stochastischer Prozesse

- Notation für ein White Noise (ε_t) mit Erwartungswert μ und Varianz σ^2 : $(\varepsilon_t) \sim \text{WN}(\mu, \sigma^2)$.
- Ein White Noise Prozess (ε_t) bei dem die Zufallsvariablen normalverteilt sind ($\varepsilon_t \sim N(\mu, \sigma^2)$) heißt ein **Gaußscher White Noise Prozess**.
- Ein striktes (iid) White Noise ist automatisch auch ein schwaches, die Umkehrung gilt nicht – außer beim Gaußschen White Noise.

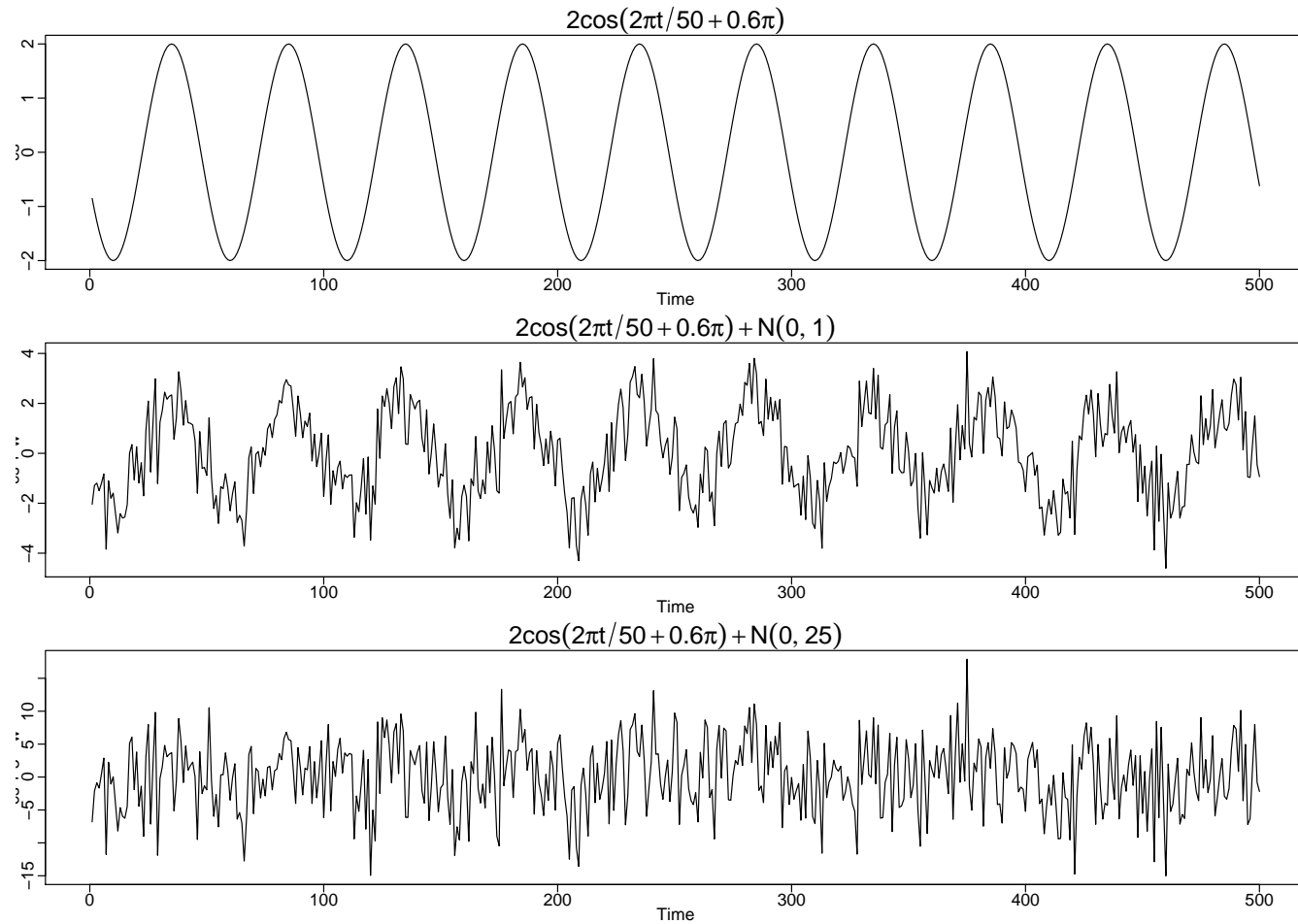
White Noise



White Noise



Signal plus Rauschen



White Noise

```
wn1 = rnorm(500,0,1) # simuliert 500 N(0,1)-Variablen
wn2=rnorm(500,0,1)
par(mfrow=c(2,1), mar=c(3,2,1,0)+.5, mgp=c(1.6,.6,0))
plot.ts(wn1, main="Zwei Pfade eines Gaußschen White noise", ylab="")
plot.ts(wn2, ylab="")

DAX15=window(lretDAX, start="2015-01-01", end="2016-01-01")
wn = rnorm(length(DAX15), mean(DAX15), sd(DAX15))
par(mfrow=c(2,1), mar=c(3,2,1,0)+.5, mgp=c(1.6,.6,0))
plot.ts(wn, start= 0, frequency = 1), main="Gaußscher White Noise", ylab="")
plot.ts(DAX15, ylab="", main="DAX Logreturns 2015")

#periodisches signal plus rauschen
signal = 2*cos(2*pi*1:500/50 + .6*pi)
wn = rnorm(500,0,1)
par(mfrow=c(3,1), mar=c(3,2,2,1), cex.main=1.5)
plot.ts(signal, main=expression(2*cos(2*pi*t/50+.6*pi)))
plot.ts(signal+wn, main=expression(2*cos(2*pi*t/50+.6*pi) + N(0,1)))
plot.ts(signal+5*wn, main=expression(2*cos(2*pi*t/50+.6*pi) + N(0,25)))
```


Beispiele stochastischer Prozesse

- Ein stochastischer Prozess

$$S_t = \sum_{n=1}^t \varepsilon_n, \quad t = 1, 2, \dots$$

wobei (ε_t) einen White Noise Prozess bezeichnet, heißt **Random Walk**

- Ist $\mathbb{E}[\varepsilon_t] = 0$, so heißt (S_t) **Random Walk ohne Drift**. (Manchmal auch nur Random Walk.)
- Ist $\mathbb{E}[\varepsilon_t] = \mu \neq 0$, so heißt (S_t) **Random Walk mit Drift**.

Beispiele stochastischer Prozesse

Sei $\mathbb{E}[\varepsilon_t] = \mu$, und (S_t) der zugehörige **Random Walk** (μ ist beliebig).

- (S_t) kann auch als

$$S_t = \mu t + \sum_{n=1}^t \tilde{\varepsilon}_n, \quad t = 1, 2, \dots$$

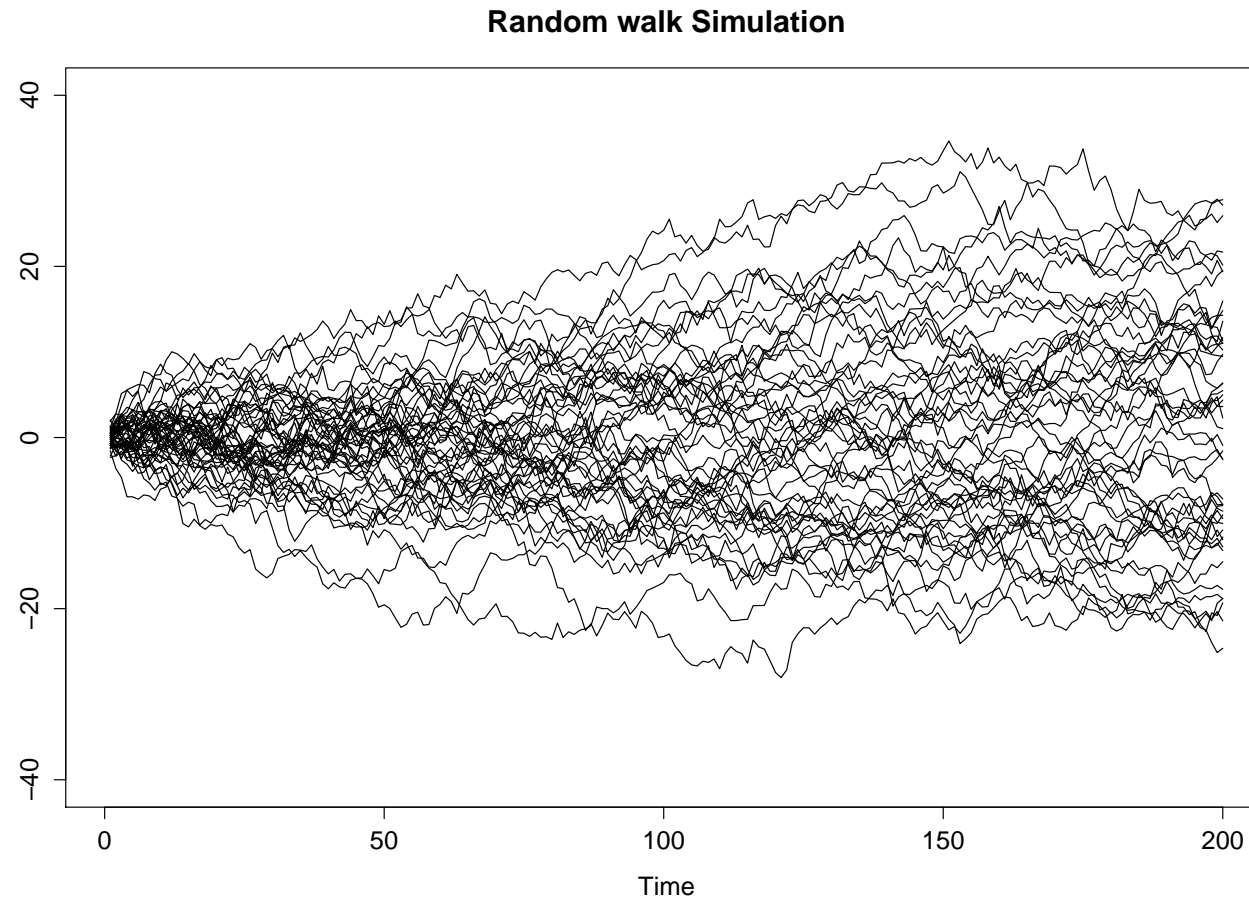
geschrieben werden, wobei $(\tilde{\varepsilon}_t)$ ein White Noise Prozess mit $\mathbb{E}[\tilde{\varepsilon}_t] = 0$ ist. (Linearer Trend + kumuliertes Rauschen)

- Eine äquivalente Definition von (S_t) kann durch die rekursive Gleichung geschehen:

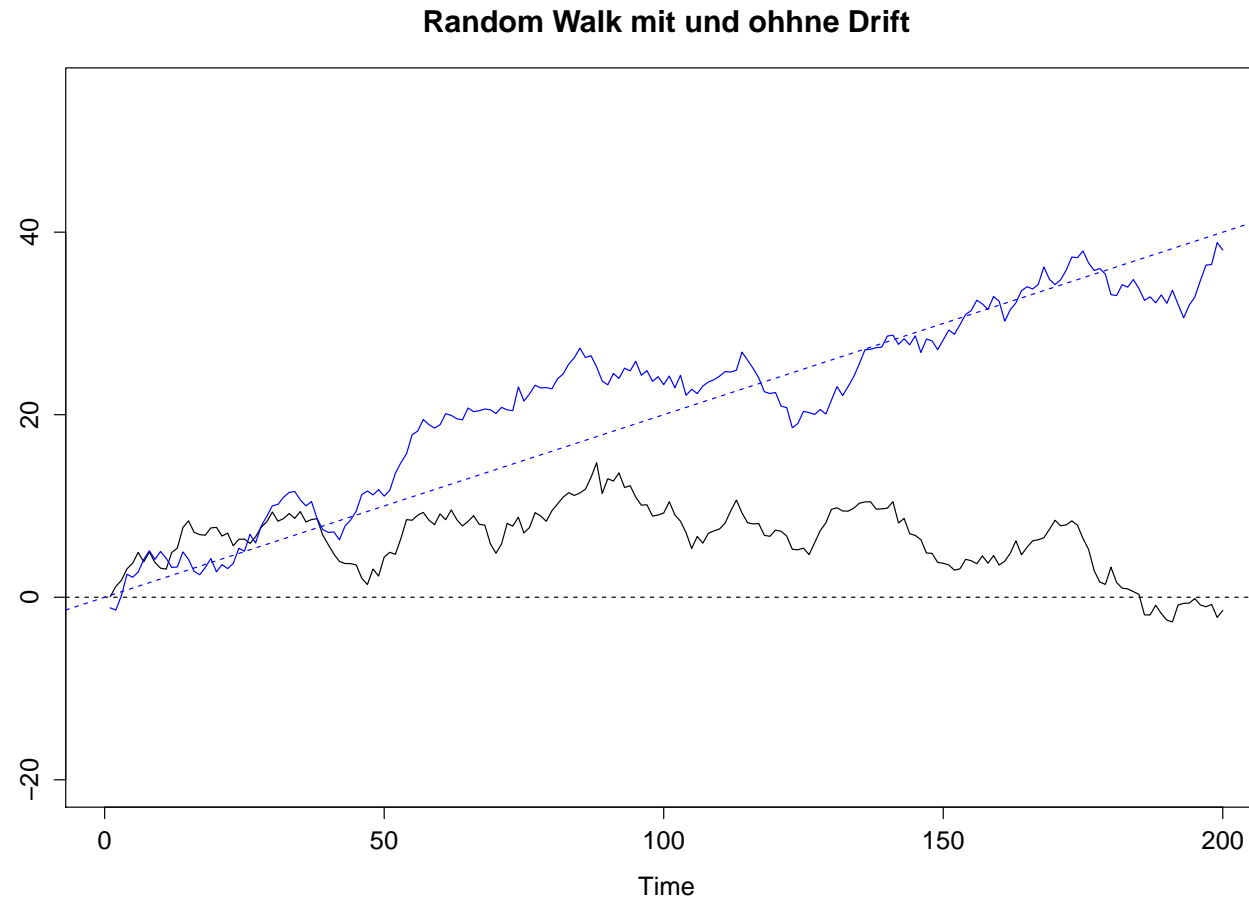
$$S_0 := 0, \quad S_t = \mu + S_{t-1} + \tilde{\varepsilon}_t, \quad t = 1, 2, \dots$$

Damit ist Random Walk ein **autoregressiver Prozess erster Ordnung**, ein **AR(1)** Prozess. (**Übung**)

Random Walk



Random Walk



Random Walk

```
#random walk
w = rnorm(200,0,1)
x = cumsum(w)
plot.ts(x, ylim=c(-40,40), main="Random walk Simulation", ylab="")
i=1
while(i<50){w = rnorm(200,0,1); x = cumsum(w); lines(x); i=i+1}
#set.seed(154) # wuerde zufallsgenerator festlegen
wn1 = rnorm(200,0,1); rw1 = cumsum(wn1)
wd = rnorm(200,0,1) +.2; rw2 = cumsum(wd)
plot.ts(rw1, ylim=c(-20,55), main="Random Walk mit und ohhne Drift", ylab='')
lines(rw2, col=4)
abline(h=0, lty=2)
abline(a=0, b=.2, col=4, lty=2)
```

Wichtige Kennzahlen

Sei (X_t) ein stochastischer Prozess.

- Die **Erwartungswert-** und die **Varianzfunktion** von (X_t) sind gegeben durch

$$\mu_t = \mathbb{E}[X_t], \quad \sigma_t^2 = \text{var}[X_t] = \mathbb{E}[(X_t - \mu_t)^2], \quad t = 1, 2, \dots$$

- Die **Autokovarianzfunktion** von (X_t) ist definiert als

$$\gamma(t, s) = \text{cov}[X_t, X_s] = E[(X_t - \mu_t)(X_s - \mu_s)], \quad t, s = 1, 2, \dots$$

Es gilt $\gamma(t, s) = \gamma(s, t)$ und $\gamma(t, t) = \sigma_t^2$.

- Die **Autokorrelationsfunktion** (ACF) von (X_t) ist gegeben durch

$$\rho(t, s) = \frac{\gamma(t, s)}{\sqrt{\gamma(t, t)\gamma(s, s)}} = \frac{\gamma(t, s)}{\sqrt{\sigma_t^2 \sigma_s^2}}, \quad t, s = 1, 2, \dots$$

Es gilt $-1 \leq \rho(t, s) \leq 1$, und $|\rho(t, s)| = 1 \Leftrightarrow X_t = a + bX_s$.

Wichtige Kennzahlen

Ein stochastischer Prozess (X_t) heißt (schwach) **stationär**, wenn

- seine Erwartungswertfunktion nicht von t abhängt, d.h.

$$\mathbb{E}[X_t] = \mu \quad \text{für alle} \quad t = 1, 2, \dots$$

- seine Autokovarianzfunktion nur von der Differenz (dem **Lag**) $\tau = |t - s|$ abhängt, d.h.

$$\gamma(t, s) = \gamma(|t - s|) = \gamma(\tau), \quad \text{für alle} \quad t, s = 1, 2, \dots$$

Das bedeutet insbesondere, dass $\sigma_t^2 = \gamma(t, t) = \gamma(0) = \sigma^2$ für alle t , d.h. die Varianzfunktion ist auch unabhängig von t .

Für stationäre Prozesse benutzen wir die Notation

$$\gamma(\tau) := \gamma(t + \tau, t) = \gamma(-\tau) \quad \text{und} \quad \rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \frac{\gamma(\tau)}{\sigma^2} = \rho(-\tau).$$

Wichtige Kennzahlen

Ein stochastischer Prozess (X_t) heißt **streng stationär**, wenn seine endlich-dimensionalen Verteilungen in jedem Zeitausschnitt gleich sind, d.h. die Verteilung von jedem Vektor

$$(X_{t_1}, \dots, X_{t_k})$$

stimmt überein mit der Verteilung von den “verschobenen” Vektor

$$(X_{t_1+\tau}, \dots, X_{t_k+\tau})$$

für alle $k = 1, 2, \dots$, alle Zeitpunkte $t_1, \dots, t_k \subseteq \mathbb{T}$ und alle Lags $\tau = 0, \pm 1, \pm 2, \dots$

Jeder streng stationäre Prozess ist auch schwach stationär. Für Gaußsche Prozesse (und nur für sie!) gilt auch die Umkehrung: Jeder Gaußscher stationäre Prozess ist auch streng stationär.

Kurze Erinnerung: Kovarianzen

$$\mu := \mathbb{E}[X], \quad \sigma^2 = \text{var}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mu^2$$

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \quad \text{var}(aX + b) = a^2 \text{var}(X), \quad \text{var}(-X) = \text{var}(X)$$

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$$

$$\text{cov}(aX, Z) = a \cdot \text{cov}(X, Z),$$

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y).$$

Sind X und Z unabhängig (z.B. wenn $Z = c = \text{const}$), dann gilt

$$\text{cov}(X, Z) = 0.$$

Bedeutung der Autokorrelation

Sei (X_t) ein stationärer Prozess mit Erwartungswert μ , Varianz σ^2 und der Autokorrelationsfunktion $\rho(\tau)$. Angenommen, wir kennen X_t und suchen damit den **besten linearen Schätzer** für $X_{t+\tau}$, also eine Funktion

$$aX_t + b \quad \text{so dass} \quad \mathbb{E} [(X_{t+\tau} - aX_t - b)^2] \xrightarrow{!} \min$$

Dann gilt:

$$a = \rho(\tau), \quad b = \mu(1 - \rho(\tau)) \quad (1)$$

und der mittlere quadratische Schätzfehler ist

$$\mathbb{E} [(X_{t+\tau} - aX_t - b)^2] = \sigma^2(1 - \rho(\tau)^2).$$

Insbesondere ist $X_{t+\tau}$ linear abhängig von X_t , falls $|\rho(\tau)| = 1$. Ist der Prozess (X_t) Gaußsch, so ist der Schätzer (1) sogar der beste Schätzer unter allen (nicht nur linearen).

Wichtige Kennzahlen: Beispiele

Sei $(\varepsilon_t) \sim \text{WN}(\mu, \sigma^2)$. Dann gilt:

- $\mathbb{E}[\varepsilon_t] = \mu$ für alle $t = 1, 2, \dots$
- $\gamma(t + \tau, t) = \begin{cases} \sigma^2 & \text{falls } \tau = 0 \\ 0 & \text{sonst} \end{cases}$ für alle $t = 1, 2, \dots$

Damit ist White Noise ein stationärer Prozess mit $\mathbb{E}[\varepsilon_t] = \mu$,

$$\gamma(\tau) = \begin{cases} \sigma^2 & \text{falls } \tau = 0 \\ 0 & \text{sonst} \end{cases} \quad \text{und} \quad \rho(\tau) = \begin{cases} 1 & \text{falls } \tau = 0 \\ 0 & \text{sonst.} \end{cases}$$

Wichtige Kennzahlen: Beispiele

Sei (S_t) ein Random Walk ohne Drift, d.h.

$$S_t = \sum_{k=1}^t \varepsilon_k, \quad t = 1, 2, \dots$$

mit $(\varepsilon_t) \sim \text{WN}(0, \sigma^2)$. Dann gilt: $\mathbb{E}[S_t] = 0$ für alle t aber

$$\gamma(t + \tau, t) = \text{cov} \left(\sum_{k=1}^{t+\tau} \varepsilon_k, \sum_{j=1}^t \varepsilon_j \right) = t\sigma^2.$$

Damit ist Random Walk ohne Drift **kein stationärer** Prozess. Insbesondere gilt: $\sigma_t^2 = \gamma(t, t) = t\sigma^2$, d.h. die **Varianz wächst** mit der Zeit.

Der **differenzierte** Prozess

$$\nabla S_t := S_t - S_{t-1} = \varepsilon_t, \quad t = 2, 3, \dots$$

ist aber wieder stationär.

Wichtige Kennzahlen: Beispiele

Sei (X_t) ein Signal plus Rauschen, z.B.

$$X_t = 2 \cos \left(\frac{2\pi t}{50} + 0,6\pi \right) + \varepsilon_k, \quad t = 1, 2, \dots$$

mit $(\varepsilon_t) \sim \text{WN}(0, \sigma^2)$. Dann:

- $\mathbb{E}[X_t] = 2 \cos \left(\frac{2\pi t}{50} + 0,6\pi \right)$ für alle $t = 1, 2, \dots$
- $\gamma(t + \tau, t) = \text{cov}(\varepsilon_{t+\tau}, \varepsilon_t) = \begin{cases} \sigma^2 & \text{falls } \tau = 0 \\ 0 & \text{sonst} \end{cases}$ für alle t

Damit ist (X_t) **kein stationärer** Prozess. Aber der Prozess

$$Y_t := X_t - \mathbb{E}[X_t] = \varepsilon_t, \quad t = 2, 3, \dots$$

ist stationär.

Beispiele stochastischer Prozesse

Sei $(\varepsilon_t) \sim \text{WN}(0, \sigma^2)$. Der Prozess

$$X_t := \mu + \sum_{n=-\infty}^{\infty} \psi_n \varepsilon_{t-n} \quad \text{wobei} \quad \sum_{n=-\infty}^{\infty} |\psi_n| < \infty \quad (2)$$

heißt ein **linearer Prozess**. Ein linearer Prozess hängt von der Vergangenheit ($n > 0$), Gegenwart ($n = 0$) und der Zukunft $n < 0$ ab. In unseren Beispielen wird $\psi_n = 0$ für $n < 0$ sein (keine Abhängigkeit von der Zukunft). Z.B. für

$$\psi_n = \begin{cases} 1 & \text{falls } n = 0 \\ 0 & \text{sonst} \end{cases}$$

erhalten wir White Noise mit Erwartungswert μ und Varianz σ^2 .

Beispiele stochastischer Prozesse

Sei $(\varepsilon_t) \sim \text{WN}(0, \sigma^2)$. Der Prozess

$$X_t := \mu + \sum_{n=-\infty}^{\infty} \psi_n \varepsilon_{t-n} \quad \text{wobei} \quad \sum_{n=-\infty}^{\infty} |\psi_n| < \infty$$

heißt ein **linearer Prozess**. Für $\mu = 0$ und

$$\psi_n = \begin{cases} 1 & \text{falls } n = 0 \\ \theta & \text{falls } n = 1 \\ 0 & \text{sonst} \end{cases}$$

erhalten wir

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

– ein sog. **Moving Average** Prozess erster Ordnung, ein **MA(1)** Prozess.

Beispiele stochastischer Prozesse

Sei $(\varepsilon_t) \sim \text{WN}(0, \sigma^2)$. Der Prozess

$$X_t := \mu + \sum_{n=-\infty}^{\infty} \psi_n \varepsilon_{t-n} \quad \text{wobei} \quad \sum_{n=-\infty}^{\infty} |\psi_n| < \infty$$

heißt ein **linearer Prozess**.

Satz 1 Jeder lineare Prozess ist **stationär** mit

$$\mathbb{E}[X_t] = \mu \quad \text{für alle } t \quad \text{und} \quad \gamma(\tau) = \sigma^2 \sum_{n=-\infty}^{\infty} \psi_{n+\tau} \psi_n \quad \text{für alle } \tau \geq 0.$$

(Gilt auch schon wenn (ε_t) nur ein schwaches White Noise ist.)

Wichtige Kennzahlen empirisch

Sei $(x_t)_{t=1,\dots,N}$ eine Zeitreihe.

- Das **arithmetische Mittel**, die **empirische Varianz** und die **empirische Standardabweichung** von $(x_t)_{t=1,\dots,N}$ sind gegeben durch

$$\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t, \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2, \quad \hat{\sigma} = \sqrt{\hat{\sigma}^2}.$$

- Die **empirische Autokovarianzfunktion** von $(x_t)_{t=1,\dots,N}$ ist definiert als

$$\hat{\gamma}(\tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x}), \quad \tau = 0, 1, \dots, N-1$$

Es gilt $\hat{\gamma}(0) = \hat{\sigma}^2$.

Notation: $\hat{\gamma}(-\tau) := \hat{\gamma}(\tau)$.

Wichtige Kennzahlen empirisch

- Die **empirische Autokorrelationsfunktion** (ACF) einer Zeitreihe $(x_t)_{t=1,\dots,N}$ ist entsprechend

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)} = \frac{\hat{\gamma}(\tau)}{\hat{\sigma}^2} \quad \tau = 0, 1, \dots, N - 1.$$

Es gilt $-1 \leq \hat{\rho}(\tau) \leq 1$. Die empirische Autokorrelationsfunktion $\hat{\rho}(\tau)$ entspricht im Wesentlichen dem **Korrelationskoeffizienten** zwischen den $N - \tau$ Beobachtungspaaren

$$(x_1, x_{1+\tau}) \quad (x_2, x_{2+\tau}) \quad \dots \quad (x_{N-\tau}, x_N),$$

misst also die lineare Abhängigkeit zwischen den Beobachtungen mit dem Zeitabstand $\tau > 0$.

Je größer der Abstand τ , desto weniger Paare stehen zur Verfügung, daher wird $\hat{\rho}(\tau)$ meist nur für $\tau \leq \frac{N}{4}$ betrachtet.

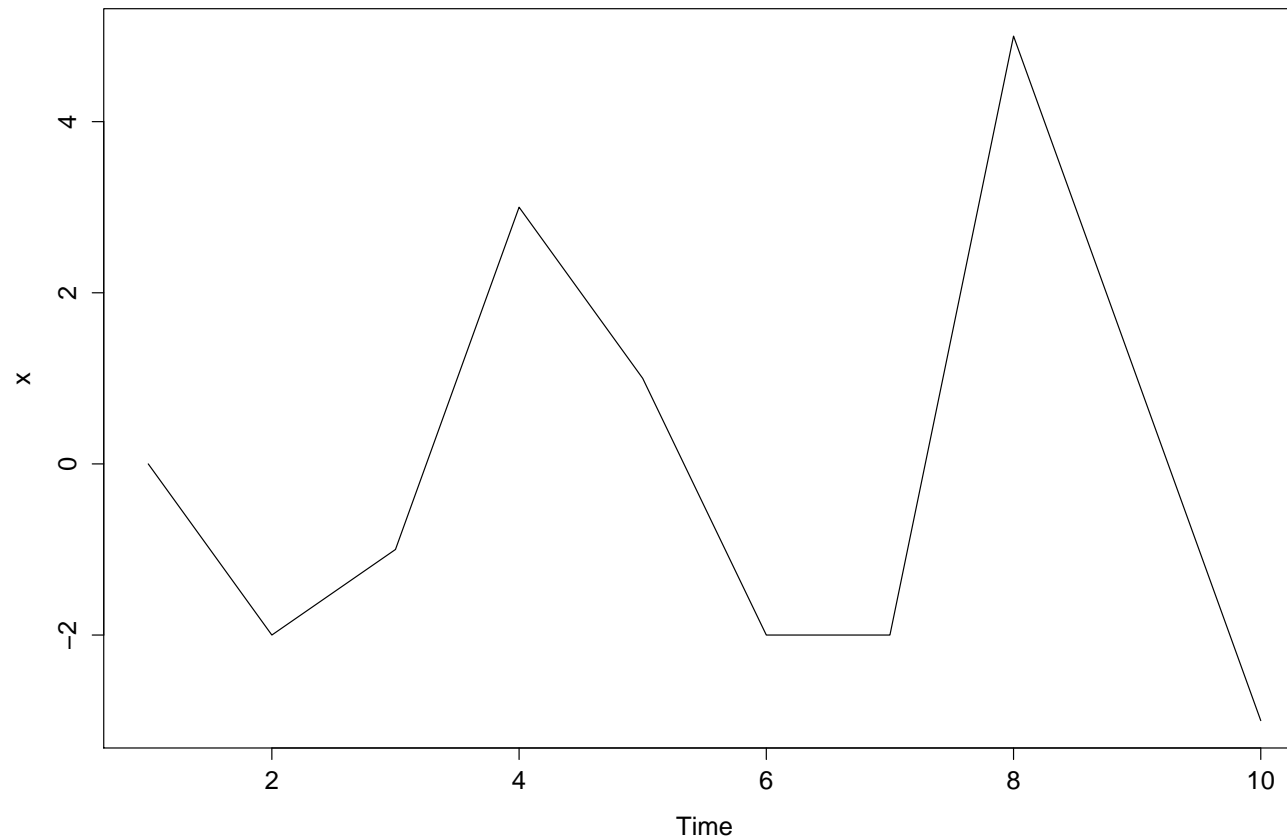
Wichtige Kennzahlen empirisch

Z.B. für eine (hypothetische) Zeitreihe der Länge 10:

t	x_t	(x_t, x_t)	(x_t, x_{t+1})	(x_t, x_{t+2})	(x_t, x_{t+3})
1	0	(0,0)	(0,-2)	(0,-1)	(0,3)
2	-2	(-2,-2)	(-2,-1)	(-2,3)	(-2,1)
3	-1	(-1,-1)	(-1,3)	(-1,1)	(-1,-2)
4	3	(3,3)	(3,1)	(3,-2)	(3,-2)
5	1	(1,1)	(1,-2)	(1,-2)	(1,5)
6	-2	(-2,-2)	(-2,-2)	(-2,5)	(-2,1)
7	-2	(-2,-2)	(-2,5)	(-2,1)	(-2,-3)
8	5	(5,5)	(5,1)	(5,-3)	
9	1	(1,1)	(1,-3)		
10	-3	(-3,-3)			

$$\rightarrow \bar{x} = 0 \quad \hat{\sigma}^2 = 5,8 \quad \hat{\rho}(1) = -0,07 \quad \hat{\rho}(2) = -0,72 \quad \hat{\rho}(3) = 0,05$$

Die hypothetische Reihe



$$\bar{x} = 0 \quad \hat{\sigma}^2 = 5,8 \quad \hat{\rho}(1) = -0,07 \quad \hat{\rho}(2) = -0,72 \quad \hat{\rho}(3) = 0,05$$

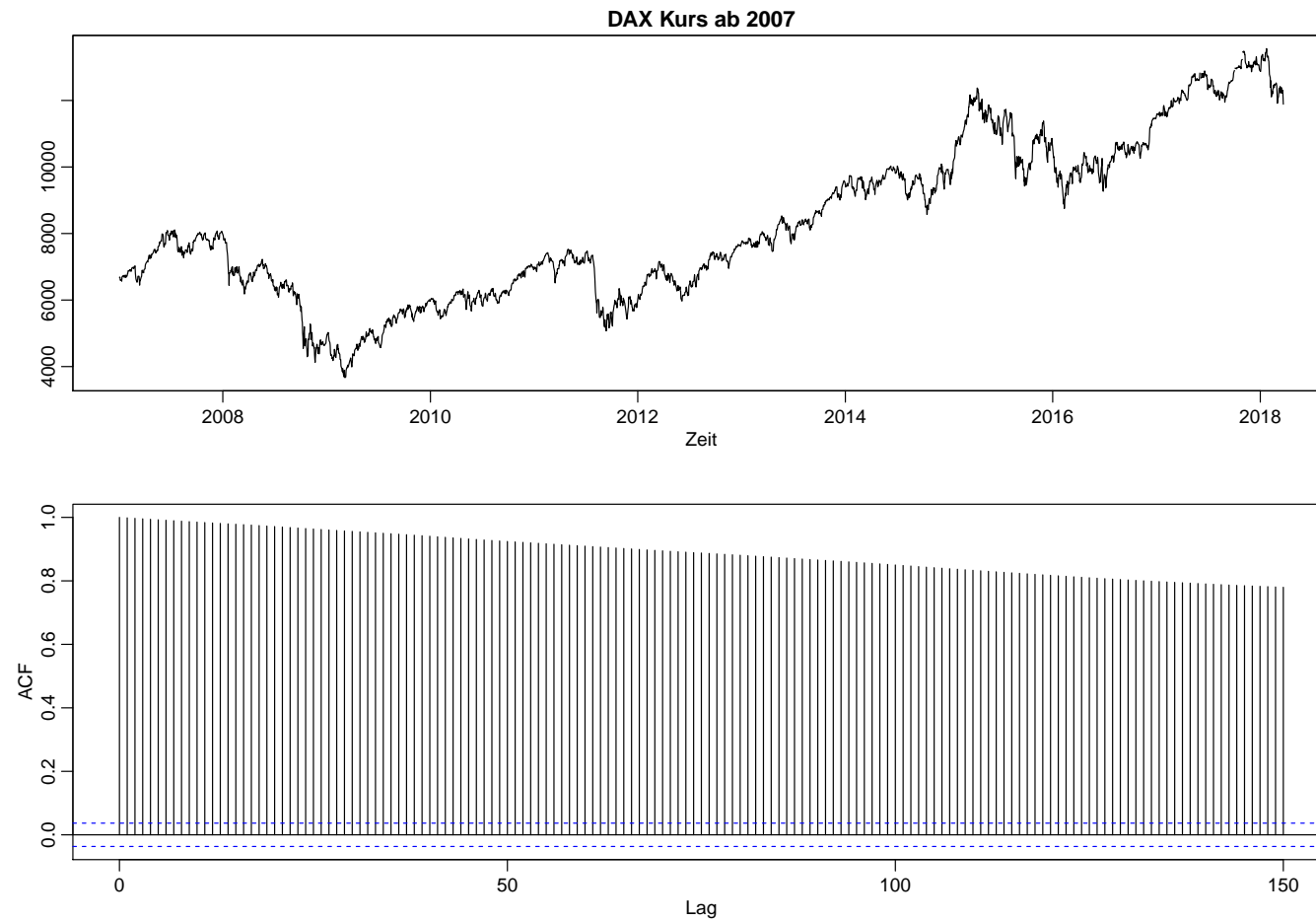
Die hypothetische Reihe

```
x=c(0,-2,-1,3,1,-2,-2,5,1,-3)
mean(x)
sum(x^2)/10
(x[-1]); (x[-10])
crossprod(x[-10],x[-1])/sum(x^2)
crossprod(x[-10][-9],x[-1][-1])/sum(x^2)
crossprod(x[-10][-9][-8],x[-1][-1][-1])/sum(x^2)
plot.ts(x)
#oder
(x=as.ts(x))
(lag(x,1)); (lag(x,-1))
(ts.intersect(x, lag(x,-1)))
crossprod(ts.intersect(x, lag(x,-1)))[1,2]/sum(x^2)
crossprod(ts.intersect(x, lag(x,-2)))[1,2]/sum(x^2)
crossprod(ts.intersect(x, lag(x,-3)))[1,2]/sum(x^2)
```

Korrelogramm

Der Graph der empirischen Autokorrelationsfunktion einer Zeitreihe wird als **Korrelogramm** bezeichnet. Daraus kann man (mit etwas Übung) viele in der Zeitreihe vorhandene Strukturen erkennen.

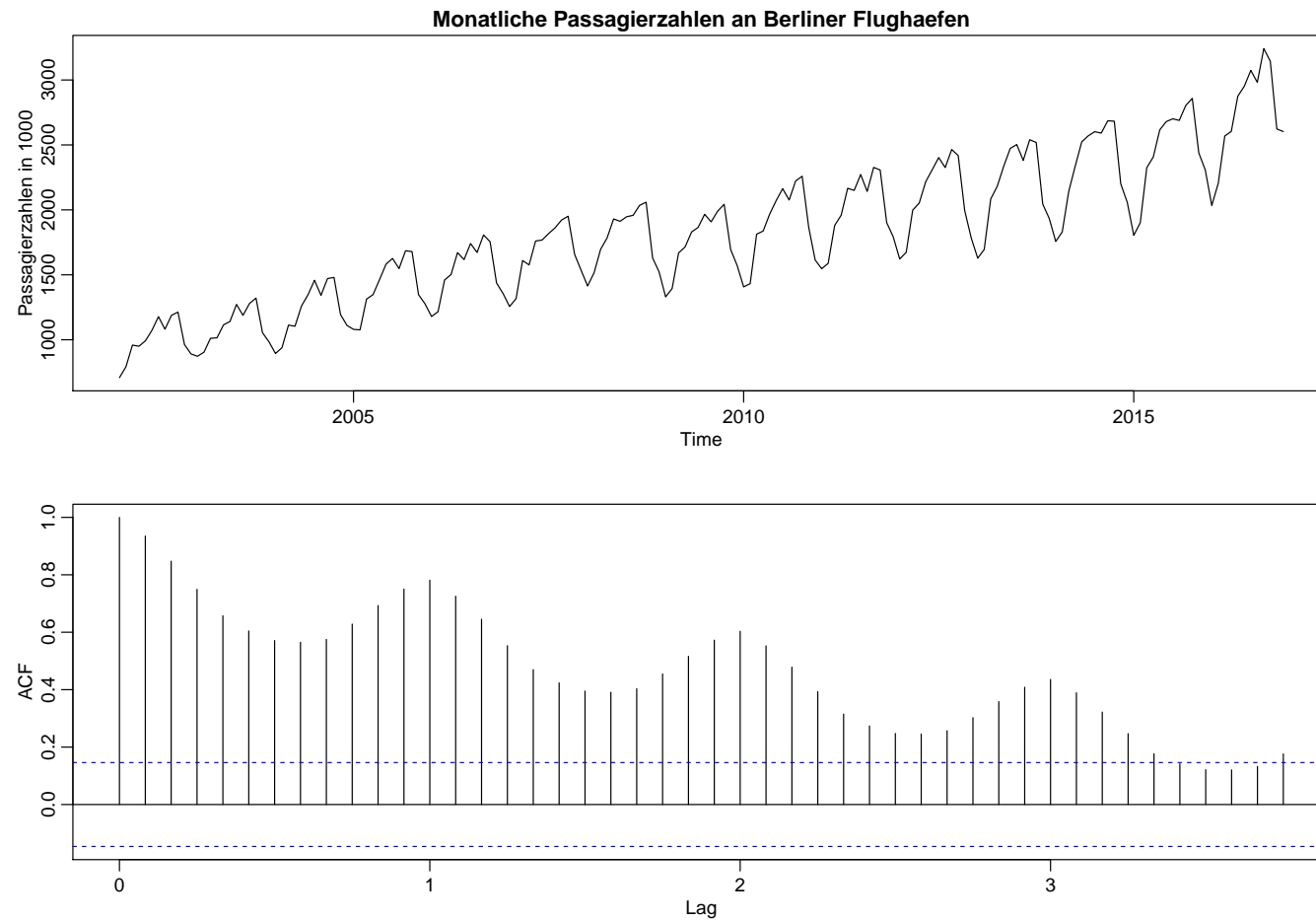
Korrelogramm: Beispiele



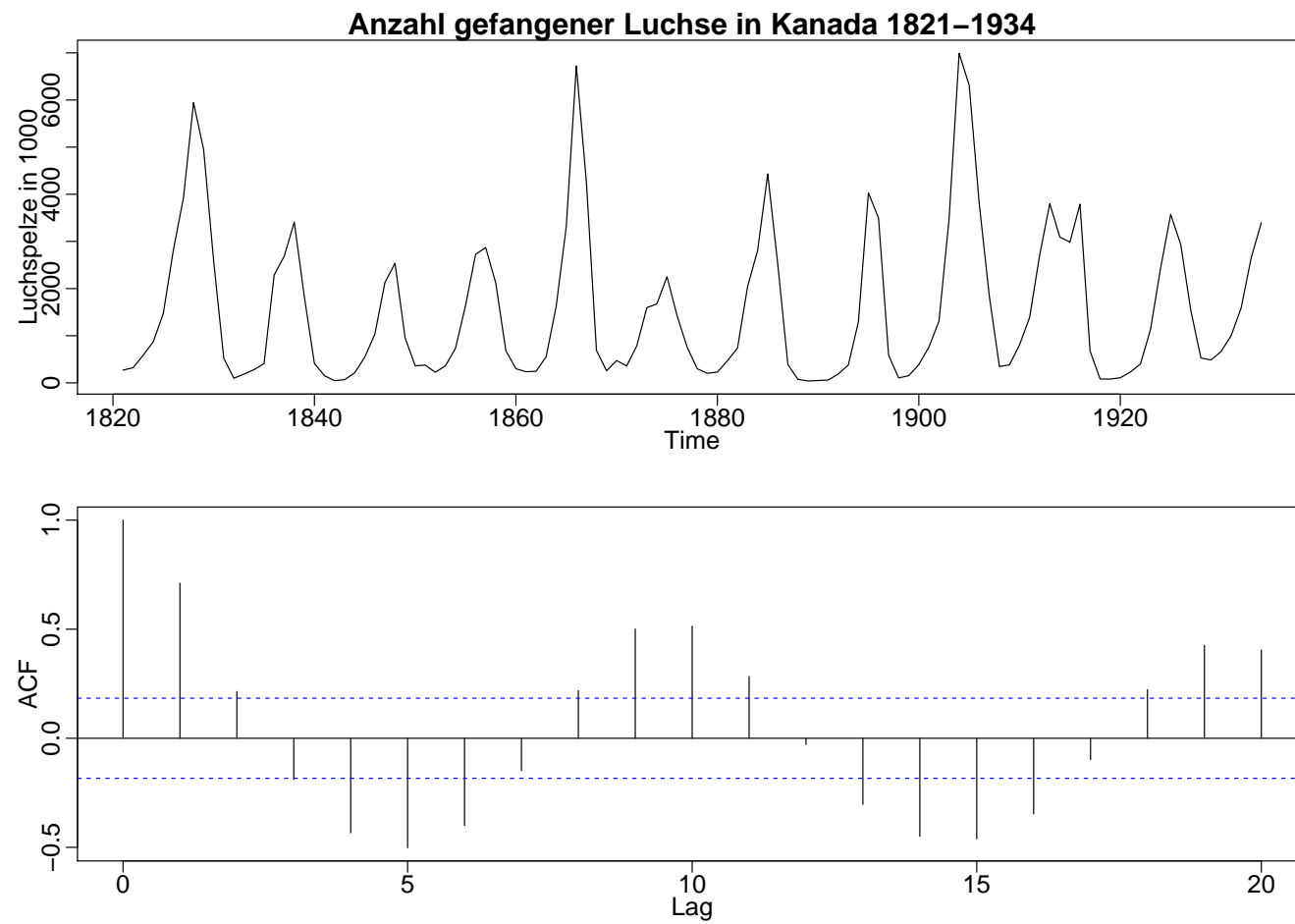
Korrelogramm: Beispiele

```
par(mfrow=c(2,1), mar=c(3,2,1,0)+.5, mgp=c(1.6,.6,0)) # set up the graphics
plot(DAX07, main="DAX Kurs ab 2007", ylab="", xlab="Zeit")
acf(DAX07, 150)
```

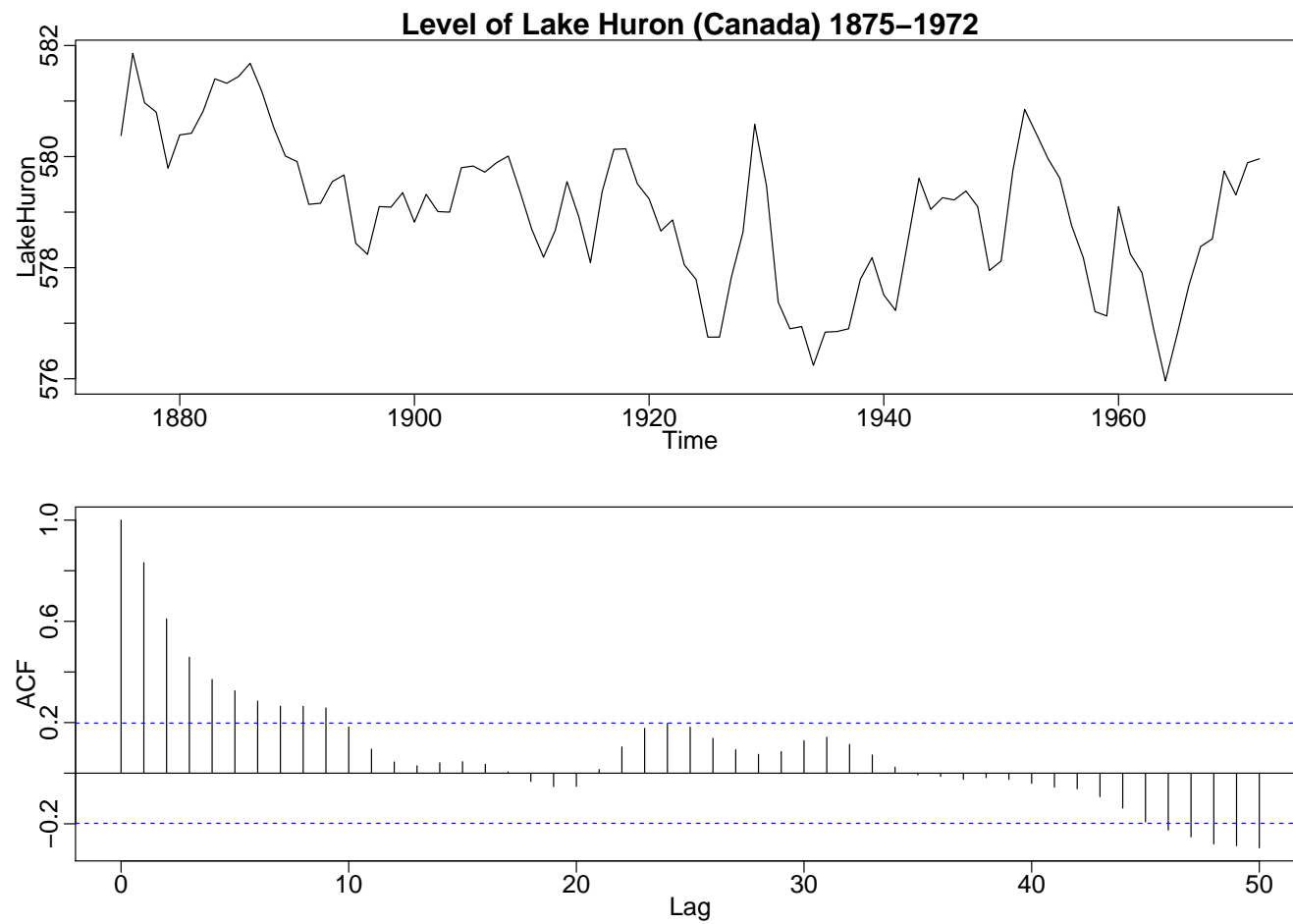

Korrelogramm: Beispiele



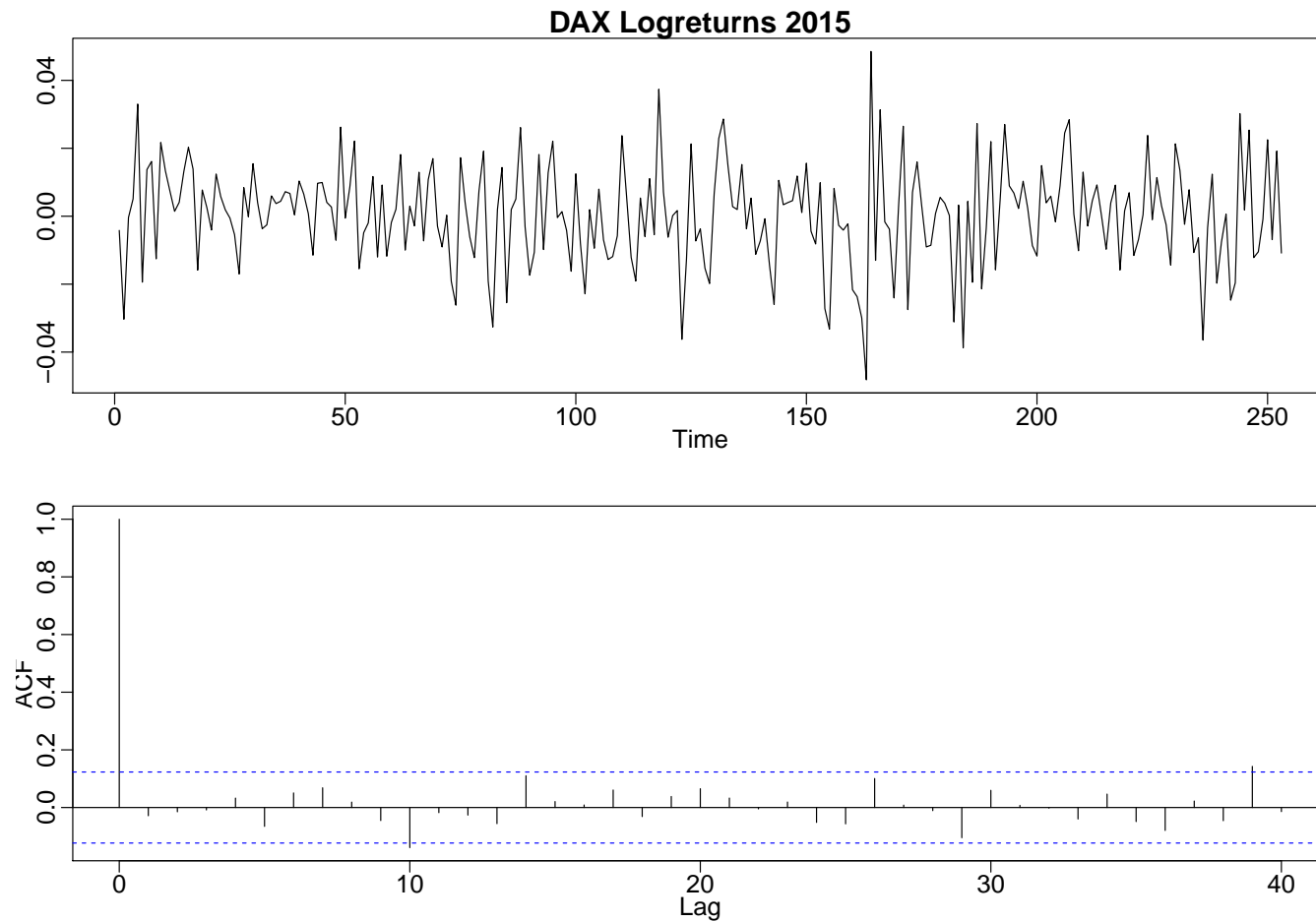
Korrelogramm: Beispiele



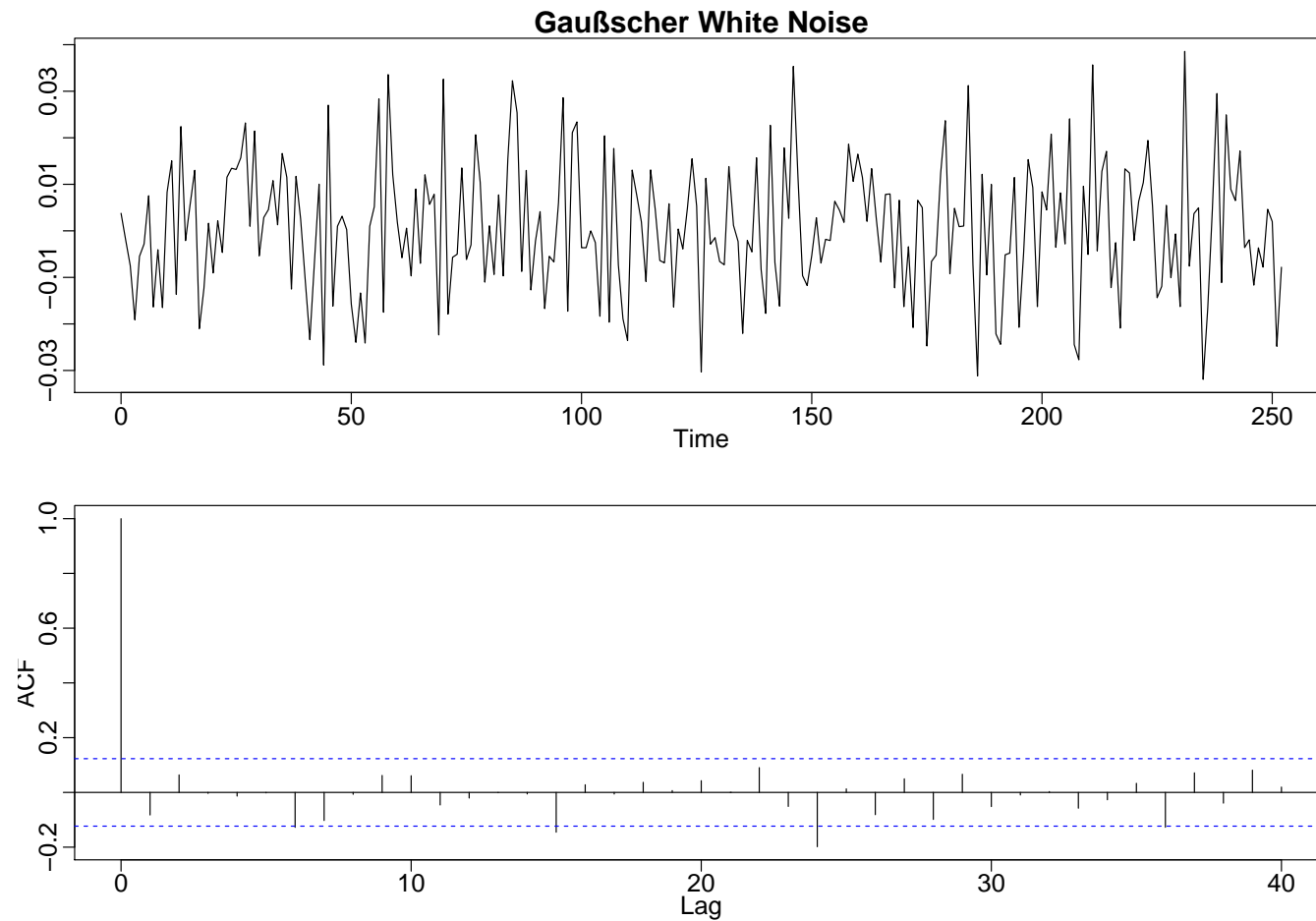
Korrelogramm: Beispiele



Korrelogramm: Beispiele



Korrelogramm: Beispiele



Identifikation von White Noise

Satz 2 Ist $(\varepsilon_t)_{t=1,\dots,N}$ ein (iid) White Noise mit $\mathbb{E}[\varepsilon_t] = 0$ und $\text{var}(\varepsilon_t) = \sigma^2$, so sind die empirischen ACFs $\hat{\rho}(1), \dots, \hat{\rho}(\tau)$ (τ fest) ungefähr iid $N(0, \frac{1}{N})$ für N groß.

Das heißt für White Noise, ungefähr 95% der empirischen ACFs sollten innerhalb des 95%-Konfidenzintervalls

$$0 \pm z_{0,975} \frac{1}{\sqrt{N}} = \pm 1,96/\sqrt{N}$$

liegen. Das Intervall wird im R-Korrelogramm standardmäßig eingezeichnet.

Dies bietet uns eine erste **Identifikationsmöglichkeit** für White Noise.

(cf. “Time series: theory and methods” by Brockwell and Davis, Example 7.2.1)

Identifikation von White Noise

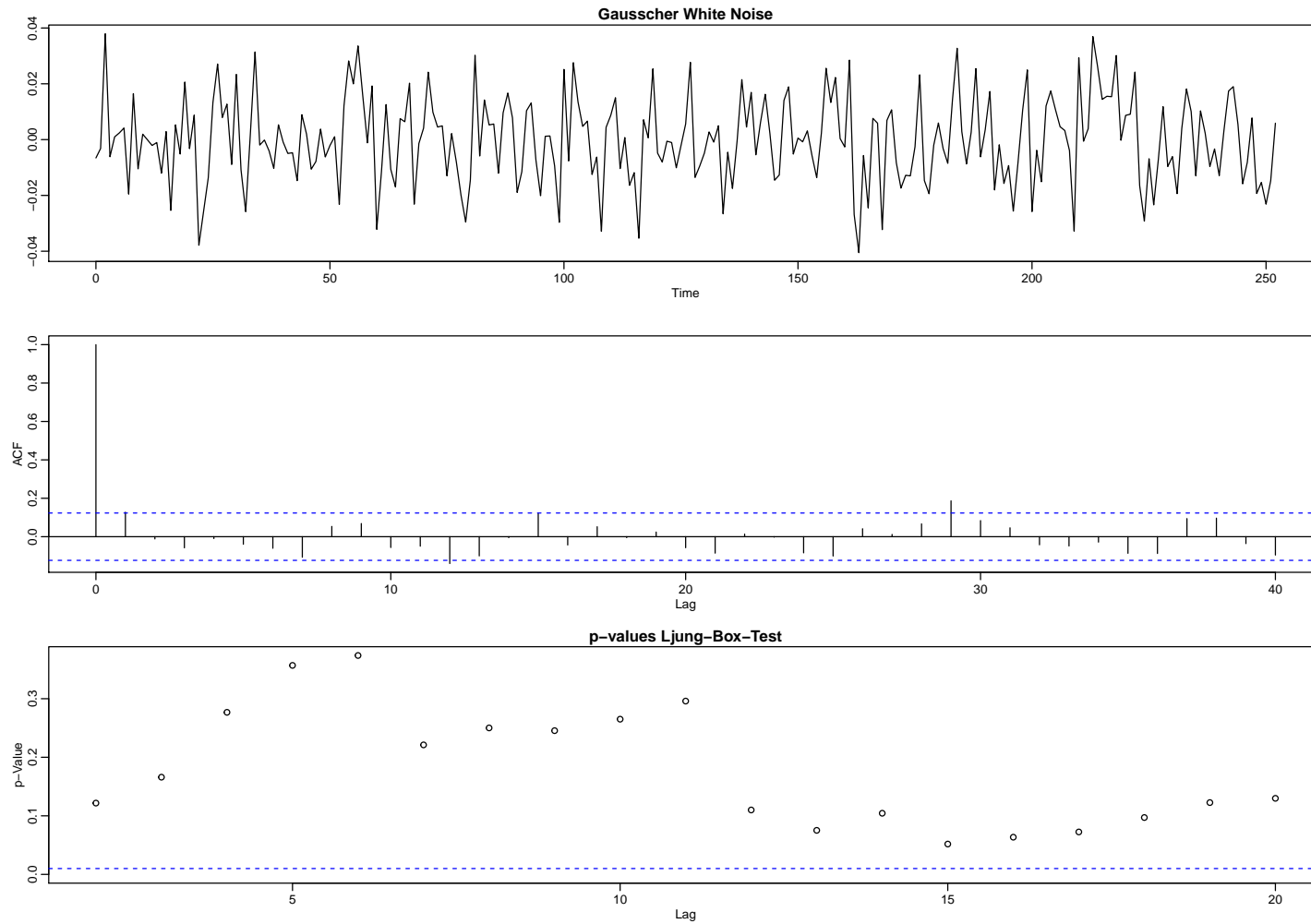
Noch ein Test zu Identifikation von White Noise (genauer: Test auf Unkorreliertheit) ist der **Ljung-Box-Pierce-Test** (Portmanteau-Test). Er basiert darauf, dass die Test-Statistik

$$Q = N(N + 2) \sum_{j=1}^k \frac{\hat{\rho}_j^2}{N - j}$$

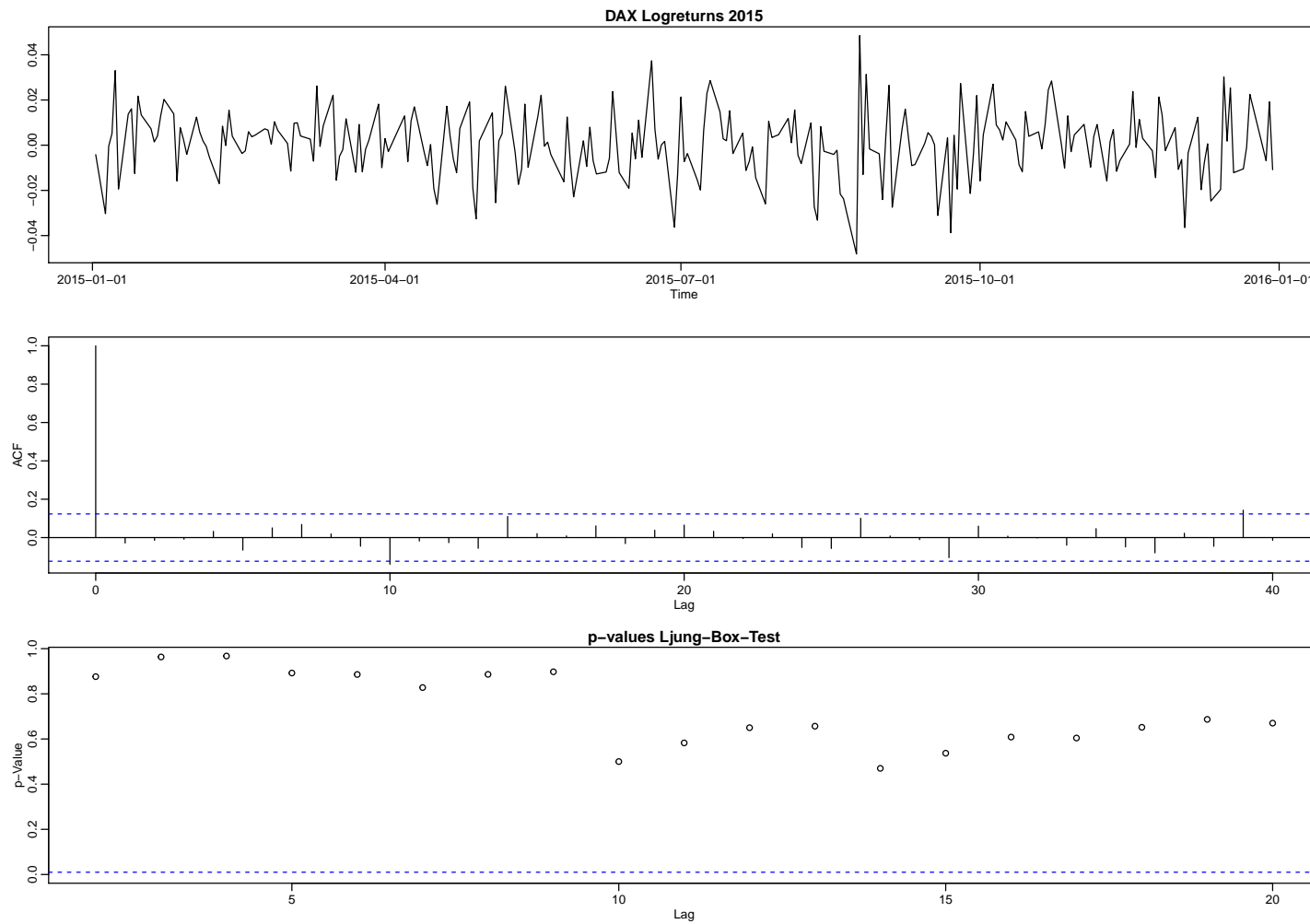
unter der Nullhypothese $\rho(1) = \dots = \rho(k) = 0$ asymptotisch χ^2 -verteilt mit k Freiheitsgraden ist. Man kann dabei verschiedene k -Werte wählen, typisch 6,12,18,20...

In R: `Box.test(input, lag=k, type="Ljung-Box")`

Identifikation von White Noise

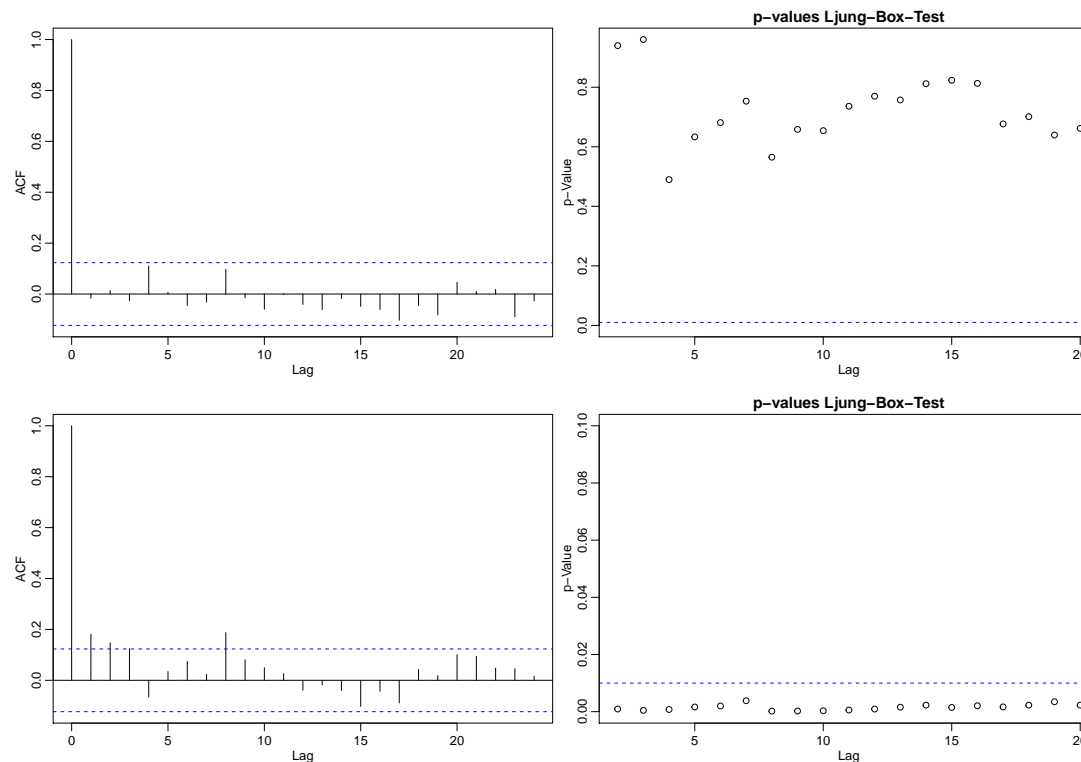


Identifikation von White Noise



Identifikation von White Noise

Aber **Vorsicht**: Hier ACF der **Quadrate** von White Noise (oben) und von Dax-Logreturns (unten):



Kein iid White Noise! → ARCH/GARCH-Modelle für Aktienkurse.

Schätzung von Parametern (kurzer Einblick)

Sind die empirischen Größen \bar{x} , $\hat{\gamma}(\tau)$ Schätzer für die “wahren” Parameter $\mathbb{E}[X_t]$, $\gamma(\tau)$?

Eine **notwendige** Bedingung dafür ist **Stationarität**. Außerdem braucht man noch, dass die Kovarianzen $\gamma(\tau)$ mit wachsendem Lag τ schnell genug abklingen. (X_t) heißt dann **mittelwertergodisch**.

Satz 3 *Ist (X_t) stationär mit Erwartungswert μ und der Autokovarianzfunktion $\gamma(\cdot)$, dann gilt für $n \rightarrow \infty$*

$$\mathbb{E} [(\bar{X}_n - \mu)^2] \rightarrow 0 \quad \text{falls} \quad \gamma(n) \rightarrow 0$$

$$n\mathbb{E} [(\bar{X}_n - \mu)^2] \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h) \quad \text{falls} \quad \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty.$$

(Es gilt $\sum_{h=-\infty}^{\infty} \gamma(h) = \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h)$.)

Schätzung von Parametern (kurzer Einblick)

Die Voraussetzungen von Satz 3 sind insbesondere erfüllt, wenn (X_t) ein **linearer** Prozess ist (mit iid Noise). In diesem Fall gilt

$$\sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2,$$

und man kann sogar zeigen, dass \bar{X}_n dann asymptotisch $N(\mu, \frac{1}{n} \sum_{h=-\infty}^{\infty} \gamma(h))$ - verteilt ist.

Dasselbe gilt, wenn (X_t) ein **Gaußscher** Prozess mit $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ ist.

Das erlaubt Bestimmung von Konfidenzintervallen für μ .

Schätzung von Parametern (kurzer Einblick)

Satz 4 Ist (X_t) ein *linearer* Prozess mit iid Noise (ε_t) und

$$\mathbb{E}[\varepsilon_t^4] < \infty \quad \text{oder} \quad \sum_{j=-\infty}^{\infty} \psi_j^2 |j| < \infty,$$

so sind die ACFs $\hat{\rho}_n(1), \dots, \hat{\rho}_n(\tau)$ (τ fest aber beliebig) asymptotisch gemeinsam normalverteilt mit Erwartungswert $\rho(1), \dots, \rho(\tau)$ und der Kovarianzmatrix $\frac{1}{n}W$, wobei

$$w_{ij} = \sum_{k=1}^{\infty} \{ \rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k) \} \\ \times \{ \rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k) \}.$$

(*Formel von Barlett.*)

Dasselbe gilt für einen *Gaußschen* Prozess mit $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$.

Ziel

Wir wollen also möglichst mit **stationären** Prozessen arbeiten, noch besser mit **linearen**. Aber Prozesse mit

- Trend
- saisonalen Schwankungen

sind **nicht stationär**. Daher versucht man, sie stationär zu machen
→ Trendbereinigung und Saisonbereinigung.

Gliederung der Vorlesung

1) Einführung und Grundlagen

2) Klassische Zeitreihenanalyse

a) Trendbestimmung

b) Glättung von Zeitreihen durch Filter

c) Saisonbereinigung

d) Exponentielles Glätten, Prognosen

3) Modellierung von Reihen durch stochastische Prozesse

a) Autoregressive (AR) Modelle

b) Moving Average (MA) Modelle

c) ARMA Modelle

d) ARIMA Modelle

Das klassische Komponentenmodell

Komponentenmodelle für Zeitreihen gehen aus von einer Zerlegung der Form

$$x_t = m_t + s_t + u_t \quad (\text{additives Modell})$$

$$x_t = m_t \cdot s_t + u_t \quad (\text{quasimultiplikatives Modell})$$

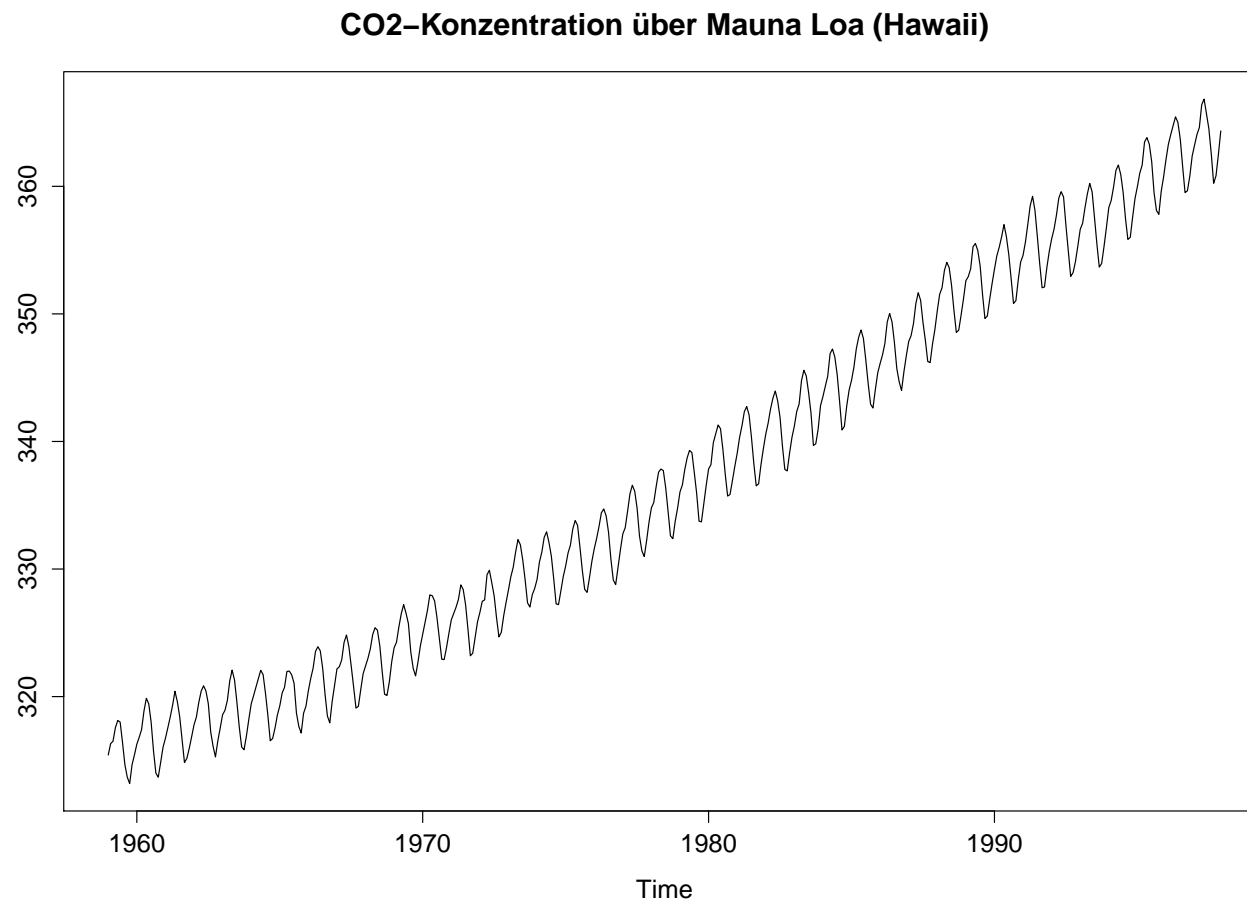
$$x_t = m_t \cdot s_t \cdot u_t \quad (\text{multiplikatives Modell})$$

Dabei bezeichnen

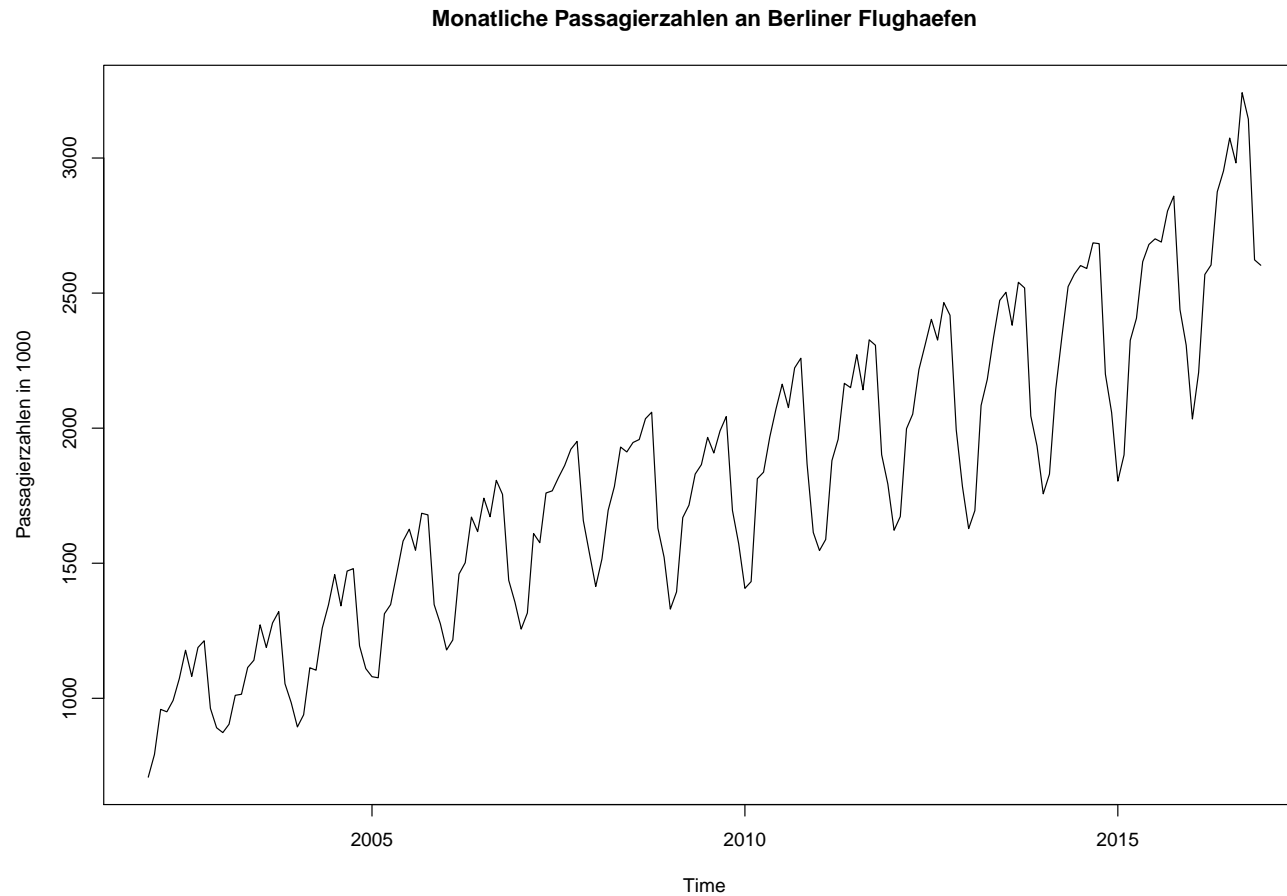
- (m_t) den **Trend**: eine langfristige systemische Veränderung des mittleren Niveaus der Zeitreihe
- (s_t) die **Saison**: eine regelmäßige zyklische Schwankung mit Periode p , d.h. $s_{t+p} = s_t$ und $\sum_{k=1}^p s_k = 0$.
- (u_t) den **Rest**: nicht zu erklärende Einflüsse oder Störungen

Manchmal noch Konjunkturkomponente, Kalenderkomponente...

Das klassische Komponentenmodell

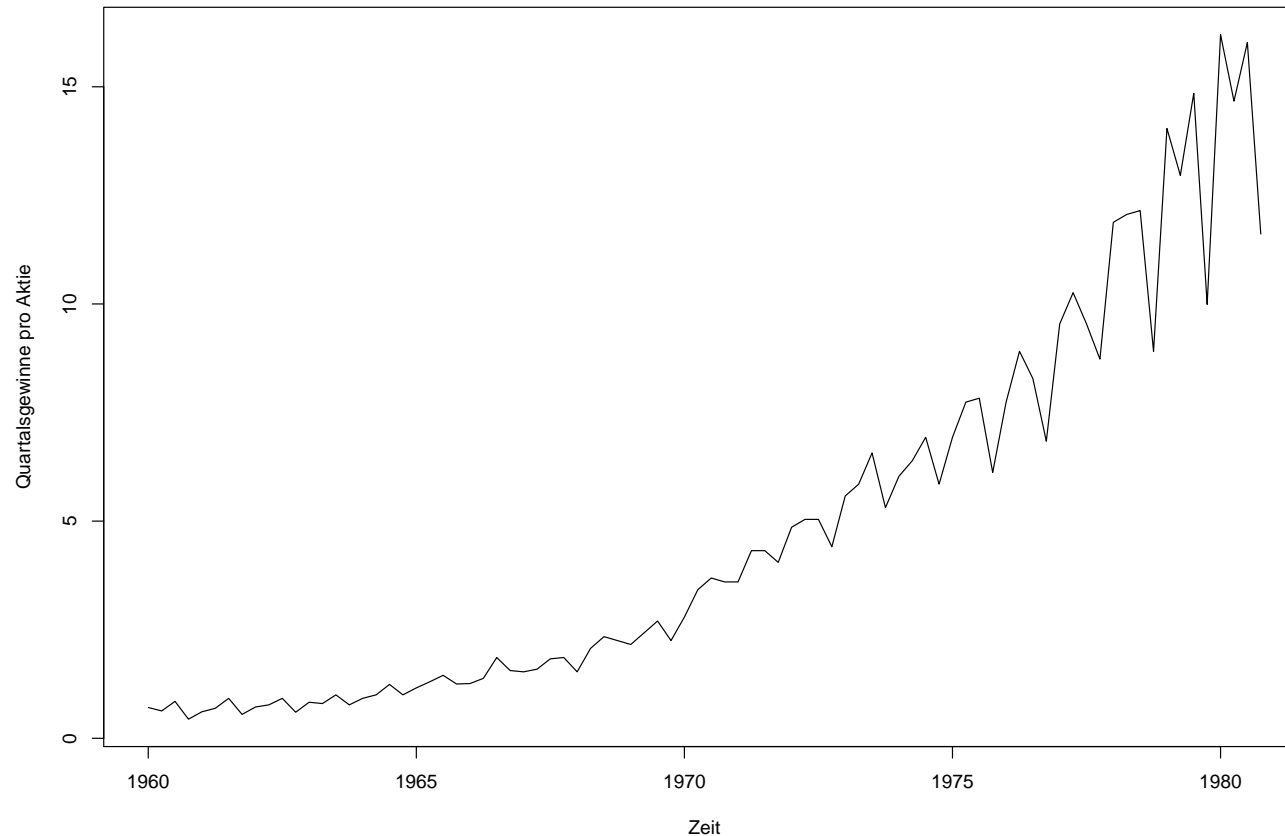


Das klassische Komponentenmodell



Das klassische Komponentenmodell

Quartalsgewinne pro Aktie von Johnson & Johnson



Transformation von Zeitreihen

Komponentenmodelle für Zeitreihen gehen aus von einer Zerlegung der Form

$$x_t = m_t + s_t + u_t \quad (\text{additives Modell})$$

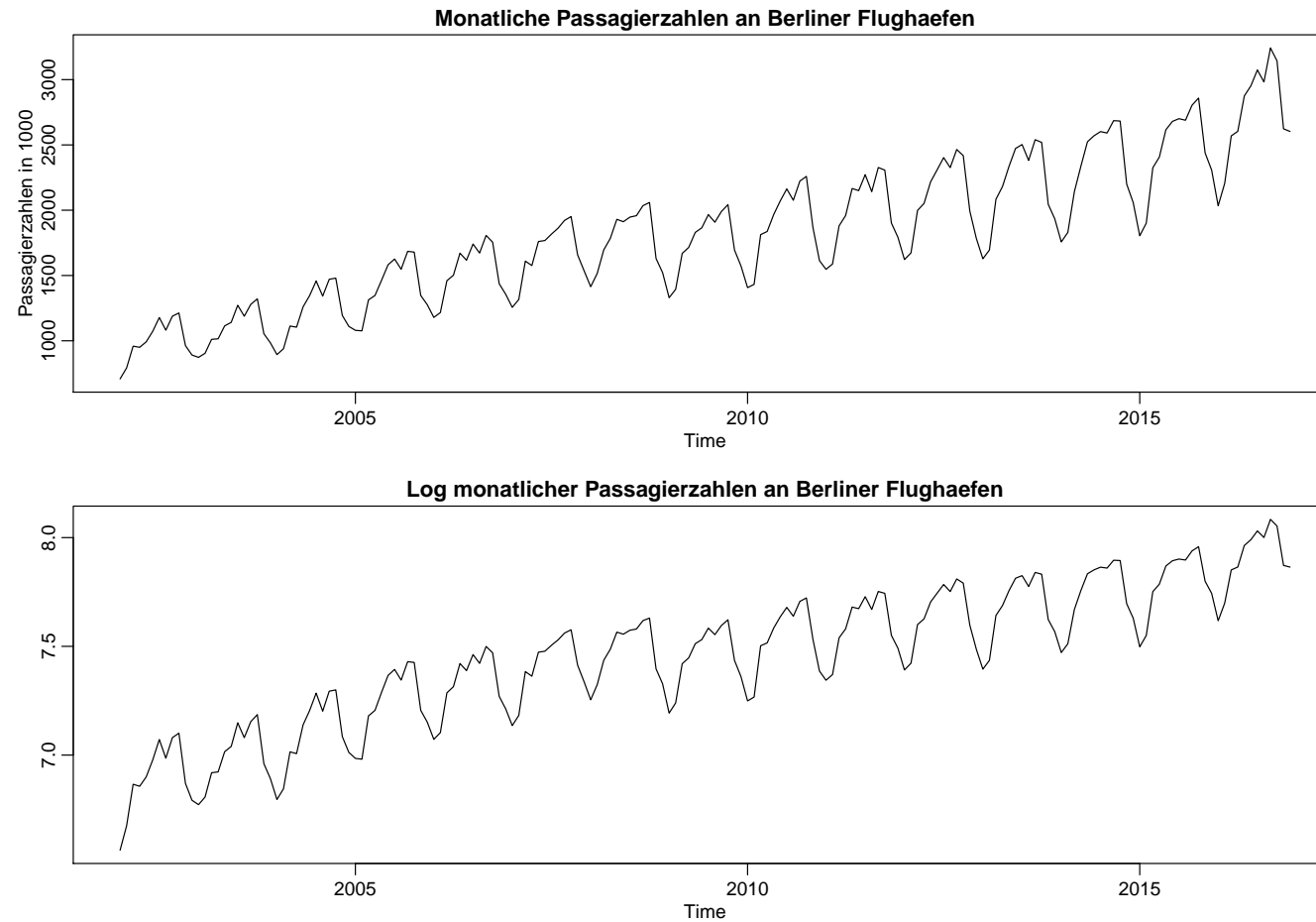
$$x_t = m_t \cdot s_t + u_t \quad (\text{quasimultiplikatives Modell})$$

$$x_t = m_t \cdot s_t \cdot u_t \quad (\text{multiplikatives Modell})$$

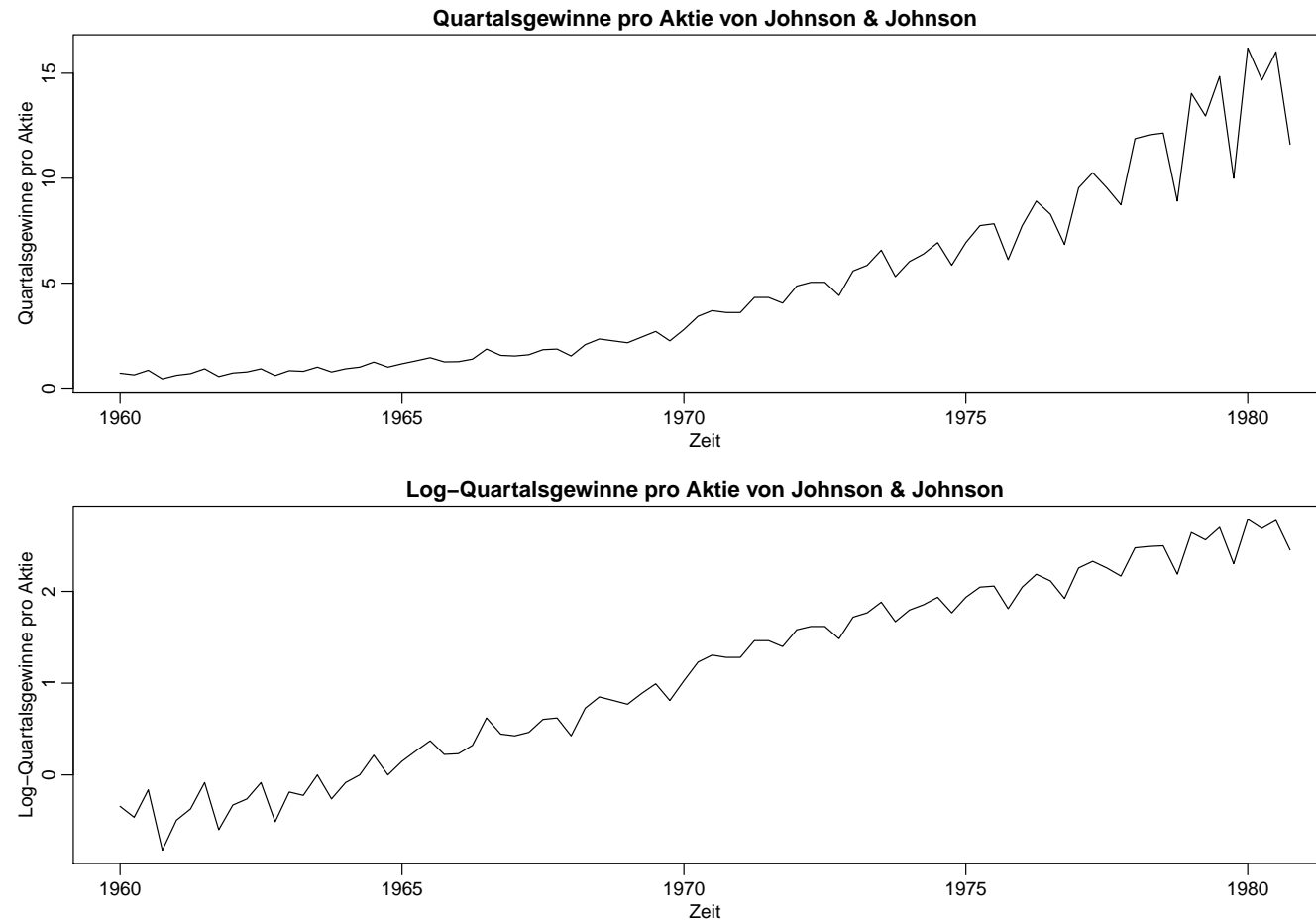
Ein multiplikatives Modell kann durch **Logarithmieren** in ein additives umgewandelt werden:

$$\log(x_t) = \log(m_t) + \log(s_t) + \log(u_t).$$

Transformation von Zeitreihen

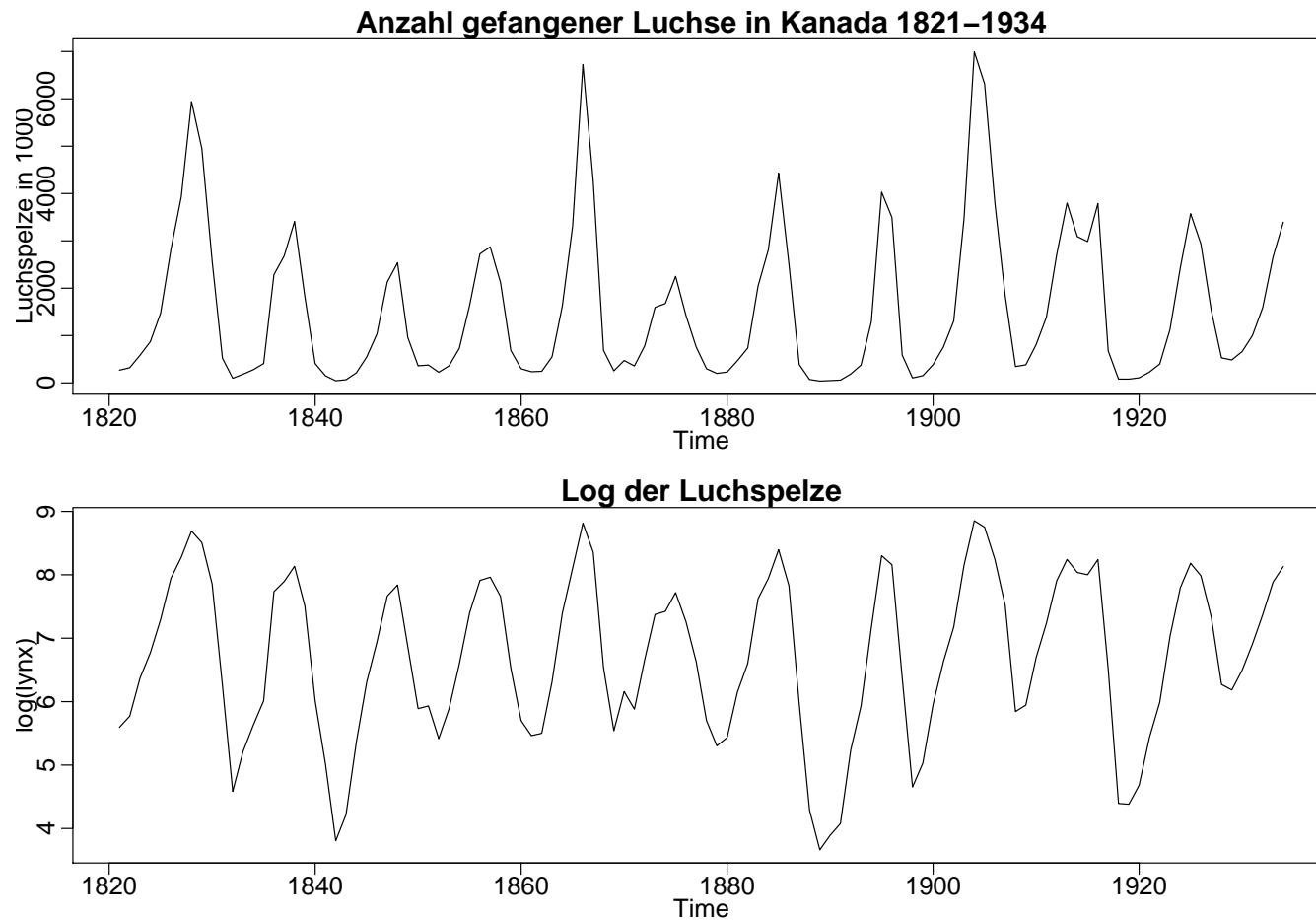


Transformation von Zeitreihen



Transformation von Zeitreihen

Logarithmieren kann auch eine Reihe **symmetrischer** machen:



Transformation von Zeitreihen

Übliche (instantane) Transformationen:

$$y_t = \begin{cases} \frac{(x_t + c)^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(x_t + c) & \lambda = 0 \end{cases}$$

Dabei ist c eine Konstante die sicherstellen soll, dass $x_t + c > 0$ (oft ist $c = 0$ ausreichend).

Es gibt Verfahren, um das “richtige” λ zu finden, oft reicht ausprobieren.

Trendbestimmung

basiert im wesentlichen auf zwei unterschiedlichen Ansätzen:

- **globale Trendbestimmung**: für die gesamte Reihe wird ein **lineares** bzw. ein **nichtlineares Regressionsmodell** unterstellt, die Parameter werden aus allen Werten mit Hilfe der **Kleinsten-Quadrate-Methode (LSE)** geschätzt.
- **lokale Trendbestimmung**: Der Trend (glatte Komponente) wird **stückweise** an die Reihe angepasst (Zerlegung in **Stützbereiche**.)
→ **Filter**.

Globale lineare Trendbestimmung

Wir gehen zunächst von einem reinen Trendmodell aus, also

$$x_t = m_t + u_t.$$

Die einfachste Trendfunktion ist eine Gerade:

$$m_t = \beta_0 + \beta_1 t.$$

Methode der kleinsten Quadrate

$$\sum_{t=1}^N u_t^2 = \sum_{t=1}^N (x_t - \beta_0 - \beta_1 t)^2 \stackrel{!}{=} \min_{\beta_0, \beta_1}$$

ergibt

$$\beta_1 = \frac{\overline{tx} - \bar{t}\bar{x}}{\overline{t^2} - (\bar{t})^2}, \quad \beta_0 = \bar{x} - \beta_1 \bar{t}.$$

Globale lineare Trendbestimmung

Für $t = 1, \dots, N$ gilt

$$\overline{tx} = \frac{1}{N} \sum_{t=1}^N tx_t, \quad \bar{t} = \frac{1}{N} \sum_{t=1}^N t = \frac{N+1}{2},$$

$$\overline{t^2} = \frac{1}{N} \sum_{t=1}^N t^2 = \frac{(N+1)(2N+1)}{6},$$

also erhält man

$$\beta_1 = \frac{12}{N(N^2 - 1)} \left(\sum_{t=1}^N tx_t - \frac{N(N+1)}{2} \bar{x} \right), \quad \beta_0 = \bar{x} - \beta_1 \frac{N+1}{2}.$$

Globale lineare Trendbestimmung

Variante: Zeitindex so wählen, dass $\bar{t} = 0$, z.B.

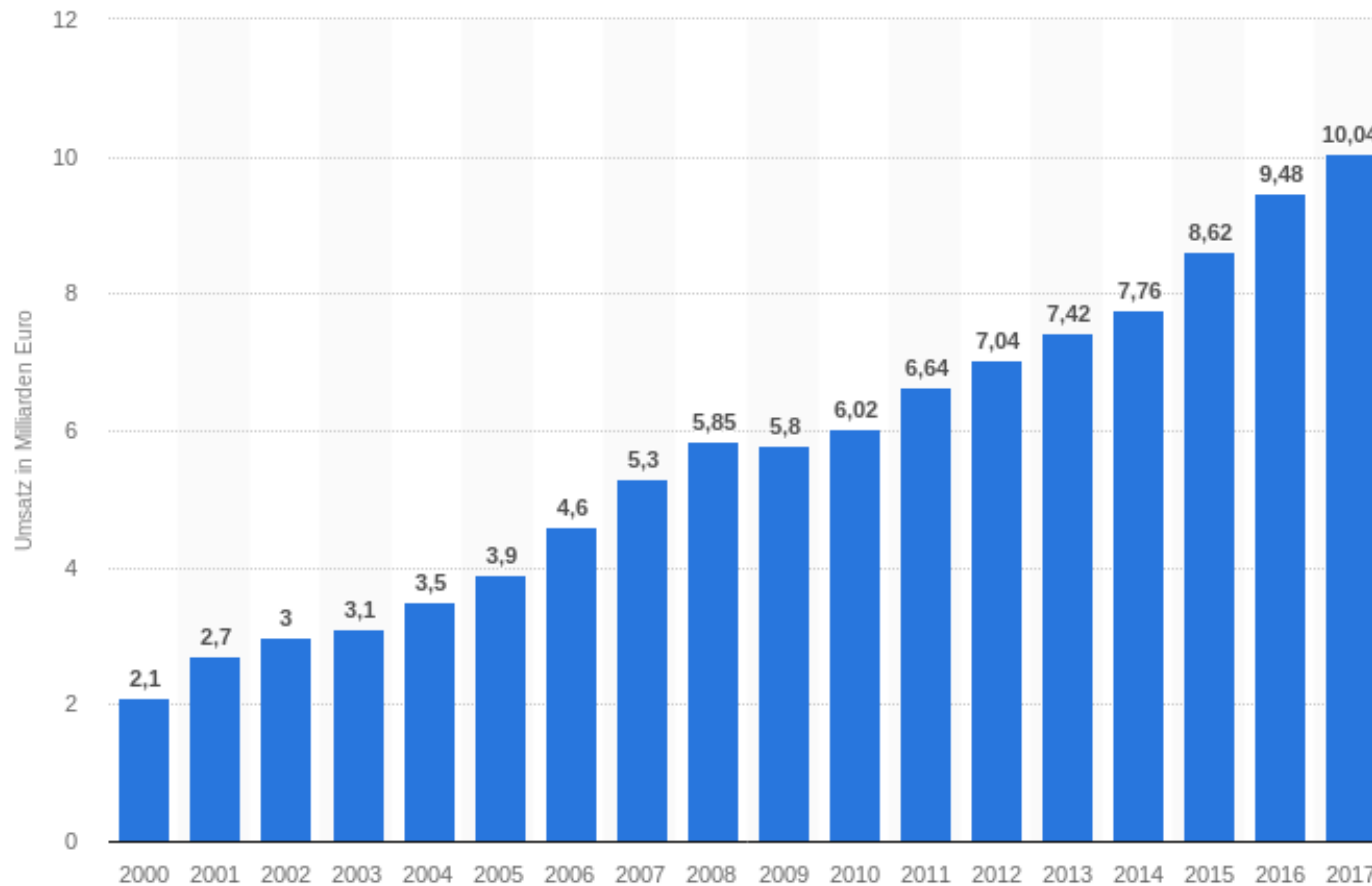
- Falls N ungerade: $-0.5(N-1), \dots, -1, 0, 1, \dots, 0.5(N-1)$
- Falls N gerade: $-0.5(N-1), \dots, -0.5, 0.5, \dots, 0.5(N-1)$.

Dann

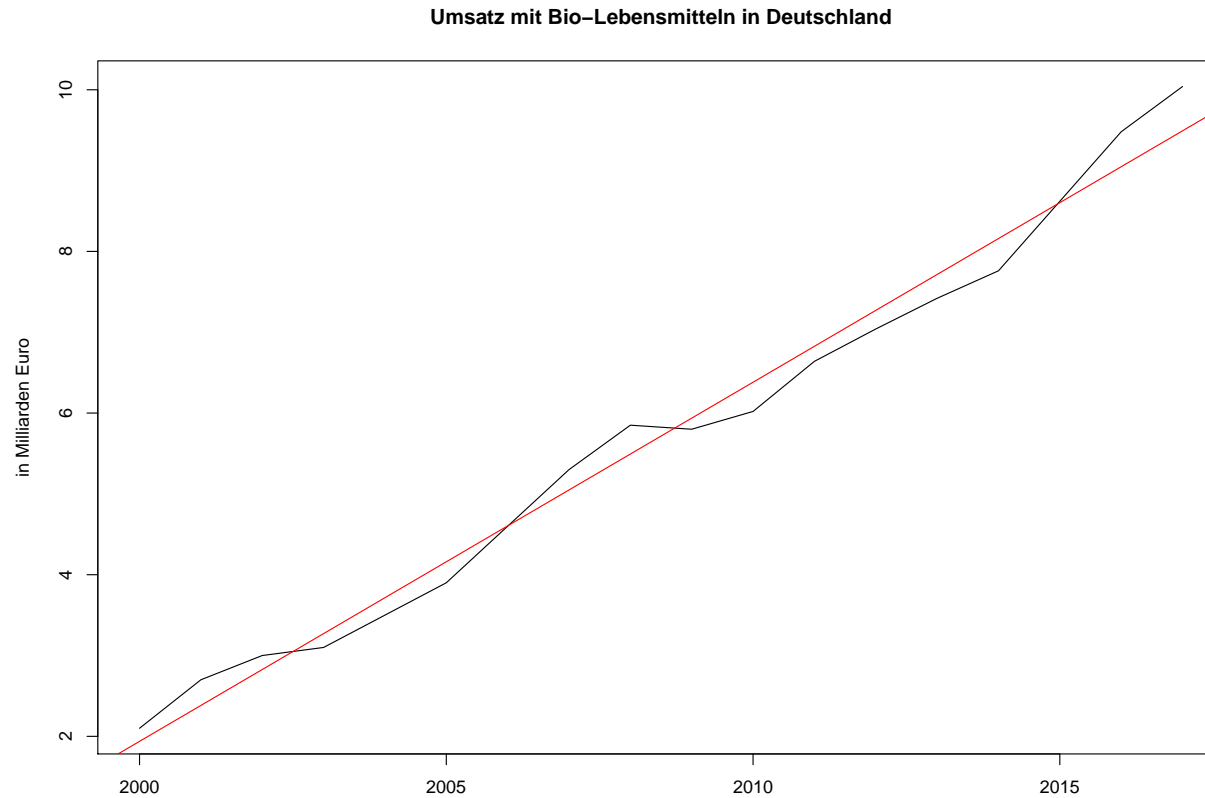
$$\beta_1 = \frac{\overline{tx}}{\overline{t^2}}, \quad \beta_0 = \bar{x}. \quad (3)$$

Globale lineare Trendbestimmung: Beispiel

Umsatz mit Bio-Lebensmitteln in Deutschland (in Milliarden Euro)



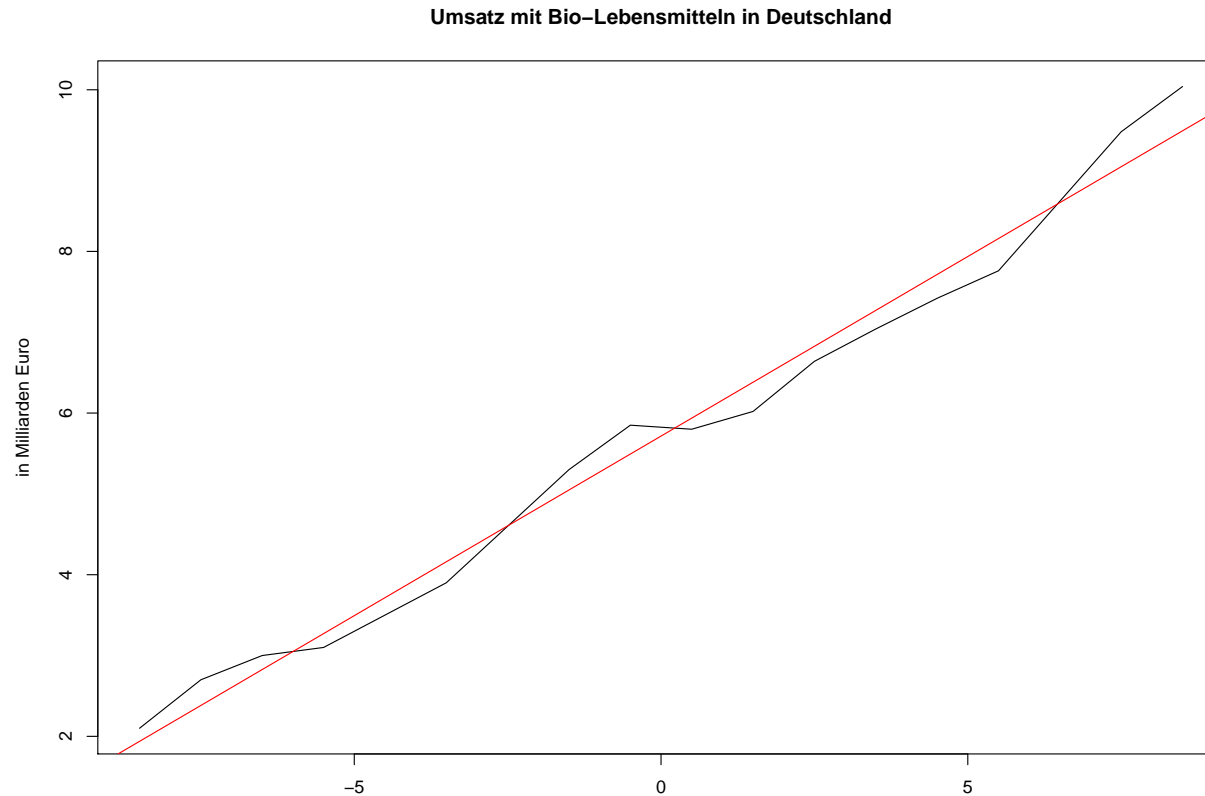
Globale lineare Trendbestimmung: Beispiel



$$t = 2000, \dots, 2017 : \quad \beta_1 = 0.44, \quad \beta_0 = -886.75.$$

Vorhersage 2018?

Globale lineare Trendbestimmung: Beispiel



$$t = -8.5, \dots, -0.5, 0.5, \dots, 8.5 : \quad \beta_1 = 0.44, \quad \beta_0 = 5.715.$$

Vorhersage 2018?

Globale lineare Trendbestimmung: Beispiel

```
x=c(2.1,2.7,3,3.1,3.5,3.9,4.6,5.3,5.85,5.8,6.02,6.64, 7.04, 7.42, 7.76,
    8.62,9.48, 10.04)
t=c(2000:2017)
mean(t); mean(x); mean(t*x)
(beta1=(mean(t*x)-mean(t)*mean(x))/(mean(t^2)-mean(t)^2))
(beta0=mean(x)-beta1*mean(t))
plot(t,x, main="Umsatz mit Bio-Lebensmitteln in Deutschland",
     ylab="in Milliarden Euro", type="l", xlab="")
abline(a=beta0, b=beta1, col=2)
#vorhersage:
beta0+2018*beta1
.4443447*2018-886.7513

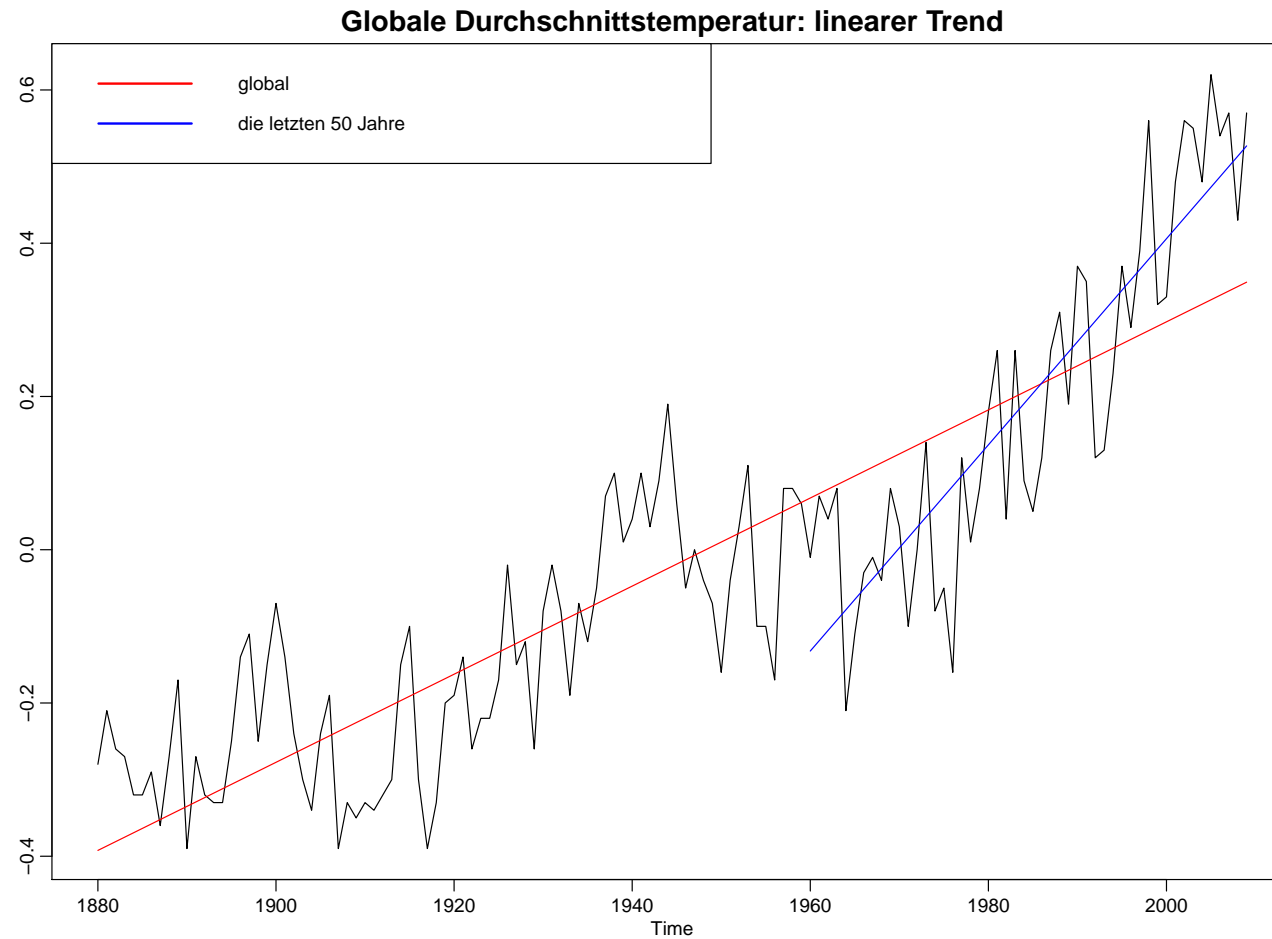
#oder Zeit 1:18
tn=c(1:length(x)); N=length(x)
(bet1=12/N/(N^2-1)*(sum(tn*x)-N*(N+1)/2*mean(x)))
(bet0=mean(x)-bet1*(N+1)/2)
plot(tn,x, main="Umsatz mit Bio-Lebensmitteln in Deutschland",
     ylab="in Milliarden Euro", type="l", xlab="")
abline(a=bet0, b=bet1, col=2)
bet0+19*bet1
```


Globale lineare Trendbestimmung: Beispiel

```
#oder Zeitindex zentrieren
(tz=seq(from=-.5*(N-1), length.out = N))
mean(tz*x); mean(tz^2)
(beta1z=mean(tz*x)/mean(tz^2)); (beta0z=mean(x))
plot(tz, x, main="Umsatz mit Bio-Lebensmitteln in Deutschland",
      ylab="in Milliarden Euro", type="l", xlab="")
abline(a=beta0z, b=beta1z, col=2)
#vorhersage:
beta0z+9.5*beta1z

#direkt mit lm
fit=lm(x~t)
fit$coefficients
summary(fit)
plot(t,x, main="Umsatz mit Bio-Lebensmitteln in Deutschland",
      ylab="in Milliarden Euro", type="l", xlab="")
abline(fit, col=2)
```

Globale lineare Trendbestimmung: Beispiel



Globale lineare Trendbestimmung: Beispiel

```
> fit1=lm(gtemp~time(gtemp))
> fit1$coefficients
(Intercept)  time(gtemp)
-11.19949646   0.00574846
> p=window(gtemp, start=1960)
> length(p)
[1] 50
> fitp=lm(p~time(p))
> fitp$coefficients
(Intercept)      time(p)
-26.49924706   0.01345258
```

https://de.wikipedia.org/wiki/Globale_Erw%C3%A4rmung

Globale nichtlineare Trendbestimmung

Wir bleiben beim Trendmodell

$$x_t = m_t + u_t.$$

Man kann auch eine allgemeinere **nichtlineare** Trendfunktion anpassen:

$$m_t = \beta_0 z_{t0} + \beta_1 z_{t1} + \beta_2 z_{t2} + \cdots + \beta_n z_{tn}$$

Wie auch bei der linearen Regression, liefert Methode der kleinsten Quadrate

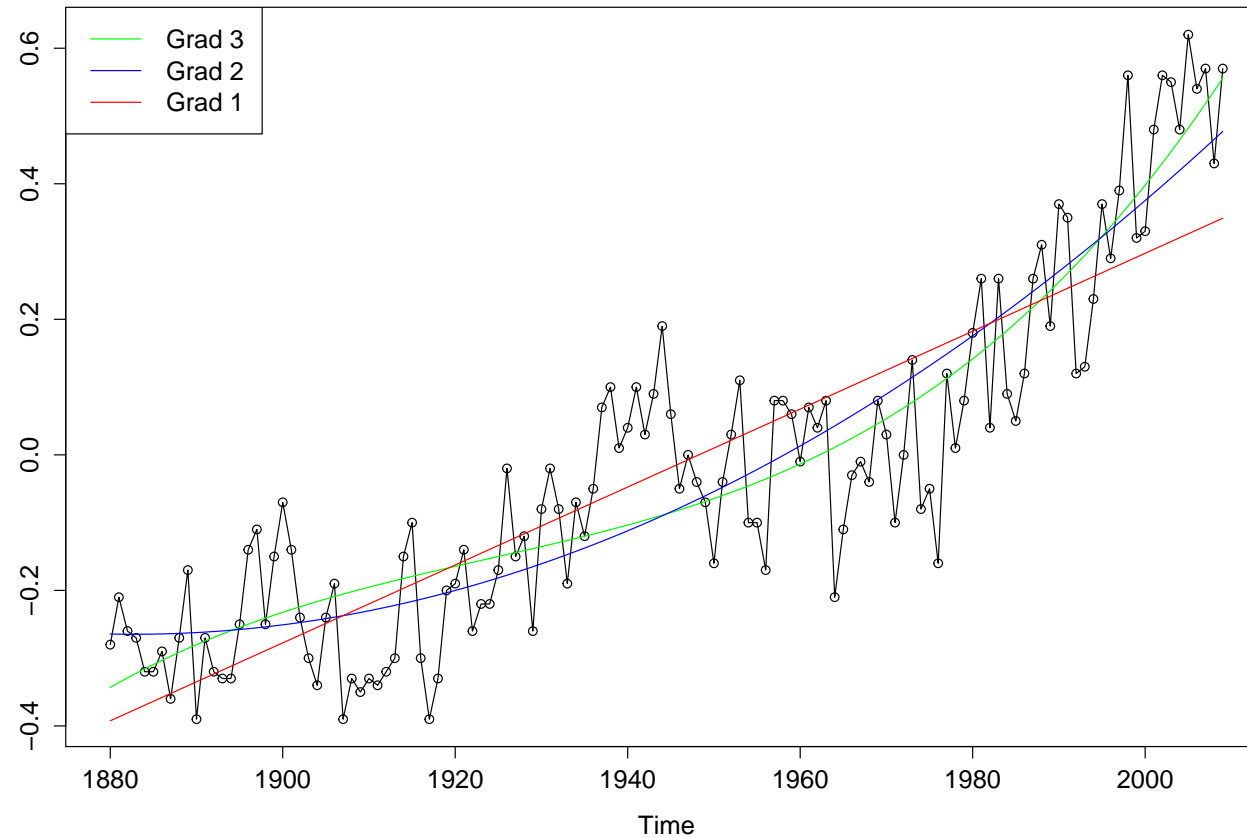
$$Q := \sum_{t=1}^N u_t^2 = \sum_{t=1}^N (x_t - \beta_0 z_{t0} - \beta_1 z_{t1} - \cdots - \beta_n z_{tn})^2 \stackrel{!}{=} \min_{\beta_0, \dots, \beta_n}$$

Gleichungen für β_0, \dots, β_n .

Typisches Beispiel: **polynomiale Regression**: $z_{ti} = t^i, \quad i = 0, \dots, n.$

Globale Trendbestimmung: Beispiel

Globale Durchschnittstemperatur: polynomiale Trends



Globale Trendbestimmung: Beispiel

```
t=c(1:length(gtemp))
t2=t^2; t3=t^3
(fit2=lm(gtemp~t+t2))
(fit3=lm(gtemp~t+t2+t3))
#plotten
plot(gtemp, type="o", main="Globale Durchschnittstemperatur: polynomiale Trends", ylab="")
linal2=ts(fit2$fitted.values, frequency=1, start=1880)
linal3=ts(fit3$fitted.values, frequency=1, start=1880)
lines(linal3, col="green4")
lines(linal2, col="blue")
lines(ts(fit1$fitted.values, frequency = 1, start=1880), col="red")
colors <- c("green4","blue","red")
legend("topleft", c(paste("Grad 3"),
                    paste("Grad 2"),
                    paste("Grad 1")),
      lwd = 1, cex=1, col=colors, pt.lwd = 1)
```

Güte der Anpassung

Sei $m_t = \beta_0 z_{t0} + \beta_1 z_{t1} + \beta_2 z_{t2} + \cdots + \beta_n z_{tn}$ der angepasste Trend mit $n + 1$ Parametern. Die **Residuen** sind dann der unerklärte Rest:

$$u_t = x_t - m_t.$$

Die **Quadratsumme der Residuen (RSS)**, der **Residual Standard Error \hat{s}**

$$RSS = \sum_{t=0}^N u_t^2, \quad \hat{s} = \sqrt{\frac{RSS}{N - (n + 1)}}$$

und der **Bestimmtheitsmaß R^2**

$$R^2 = 1 - \frac{RSS}{\sum_{t=0}^N (x_t - \bar{x})^2} = \frac{\sum_{t=0}^N (m_t - \bar{x})^2}{\sum_{t=0}^N (x_t - \bar{x})^2}$$

sind Maße für die Güte der Anpassung.

Adjusted R^2 berücksichtigt noch die Anzahl der Parameter:

$$R_{adj}^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - 1 - (n + 1)}.$$

Güte der Anpassung

Sind die Residuen (u_t) unabhängig und normalverteilt, so erhält man Tests für die Signifikanz einzelner Parameter (t -Test: $H_0 : \beta_i = 0$ gegen $H_1 : \beta_i \neq 0$) und der Gesamtregression (F -Test: $H_0 : \beta_i = 0$ für alle $i = 1, \dots, n$ gegen $H_1 : \beta_i \neq 0$ für mind. ein i).

```
> summary(fit1)
```

```
Call:
```

```
lm(formula = gtemp ~ time(gtemp))
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-0.31946	-0.09722	0.00084	0.08245	0.29383

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.120e+01	5.689e-01	-19.69	<2e-16 ***
time(gtemp)	5.749e-03	2.925e-04	19.65	<2e-16 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1251 on 128 degrees of freedom
```

```
Multiple R-squared:  0.7511,    Adjusted R-squared:  0.7492
```

```
F-statistic: 386.3 on 1 and 128 DF,  p-value: < 2.2e-16
```

Güte der Anpassung

```
> summary(fit2)
```

Call:

```
lm(formula = gtemp ~ t + t2)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.300105	-0.080650	0.004134	0.074619	0.280003

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.640e-01	2.960e-02	-8.920	4.32e-15 ***
t	-3.469e-04	1.043e-03	-0.333	0.74
t2	4.653e-05	7.714e-06	6.032	1.65e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1108 on 127 degrees of freedom

Multiple R-squared: 0.8065, Adjusted R-squared: 0.8035

F-statistic: 264.7 on 2 and 127 DF, p-value: < 2.2e-16

```
> summary(fit3)
```

Call:

```
lm(formula = gtemp ~ t + t2 + t3)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.263583	-0.076365	0.005983	0.078910	0.279050

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.505e-01	3.850e-02	-9.103	1.66e-15 ***
t	7.423e-03	2.536e-03	2.927	0.00406 **
t2	-1.012e-04	4.488e-05	-2.254	0.02591 *
t3	7.517e-07	2.253e-07	3.337	0.00111 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1066 on 126 degrees of freedom

Multiple R-squared: 0.8222, Adjusted R-squared: 0.818

F-statistic: 194.3 on 3 and 126 DF, p-value: < 2.2e-16

Ein Koeffizient in fit2 ist nicht signifikant, man kann den also weglassen:

Güte der Anpassung

```
> fit2b=lm(gtemp~t2)
> summary(fit2b)
```

Call:

```
lm(formula = gtemp ~ t2)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.301812	-0.081320	0.003791	0.073300	0.276515

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.726e-01	1.456e-02	-18.73	<2e-16 ***
t2	4.404e-05	1.908e-06	23.09	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1104 on 128 degrees of freedom

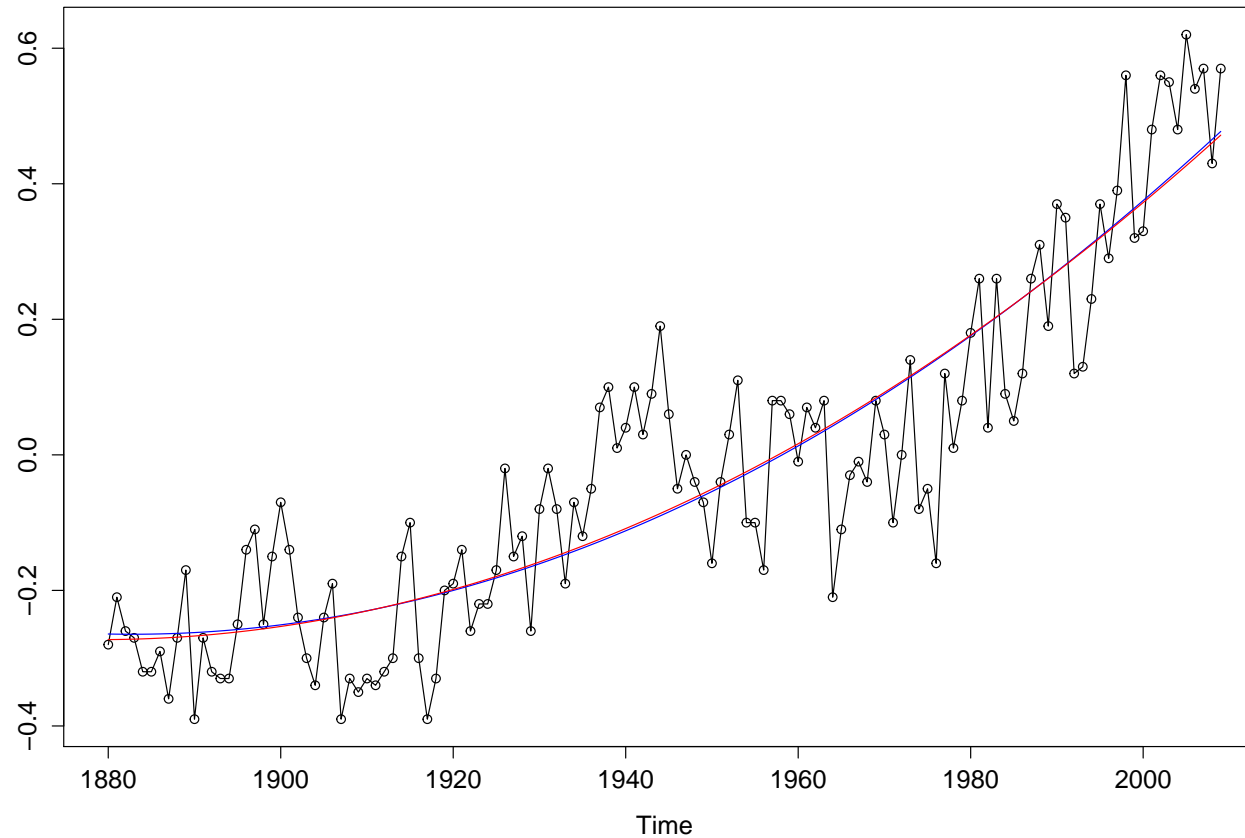
Multiple R-squared: 0.8064, Adjusted R-squared: 0.8048

F-statistic: 533 on 1 and 128 DF, p-value: < 2.2e-16

```
> plot(gtemp, type="o", main="Globale Durchschnittstemperatur: quadratischer Trend", ylab="")
> lines(linal2, col="blue")
> lines(ts(fit2b$fitted.values, frequency = 1, start=1880), col=2)
```

Globale Trendbestimmung: Beispiel

Globale Durchschnittstemperatur: quadratischer Trend



Güte der Anpassung

Man möchte ein möglichst **sparsames** Modell: eine Balance zwischen Schätzungsfehler und Anzahl der Parameter. Dazu nutzt man **Akaikes Informationskriterium (AIC)**

$$\text{AIC}(n+1) = \log \left(\frac{RSS}{N} \right) + \frac{2(n+1)}{N} \quad \text{bzw.}$$

$$\text{AIC}(n+1) = N \log \left(\frac{RSS}{N} \right) + 2(n+1)$$

und **Bayesscher Informationskriterium (BIC)**

$$\text{BIC}(n+1) = \log \left(\frac{RSS}{N} \right) + \frac{(n+1) \log(N)}{N}$$

$$\text{BIC}(n+1) = N \log \left(\frac{RSS}{N} \right) + (n+1) \log(N) \quad \text{bzw.}$$

Je kleiner AIC und BIC, desto besser das Modell.

Güte der Anpassung

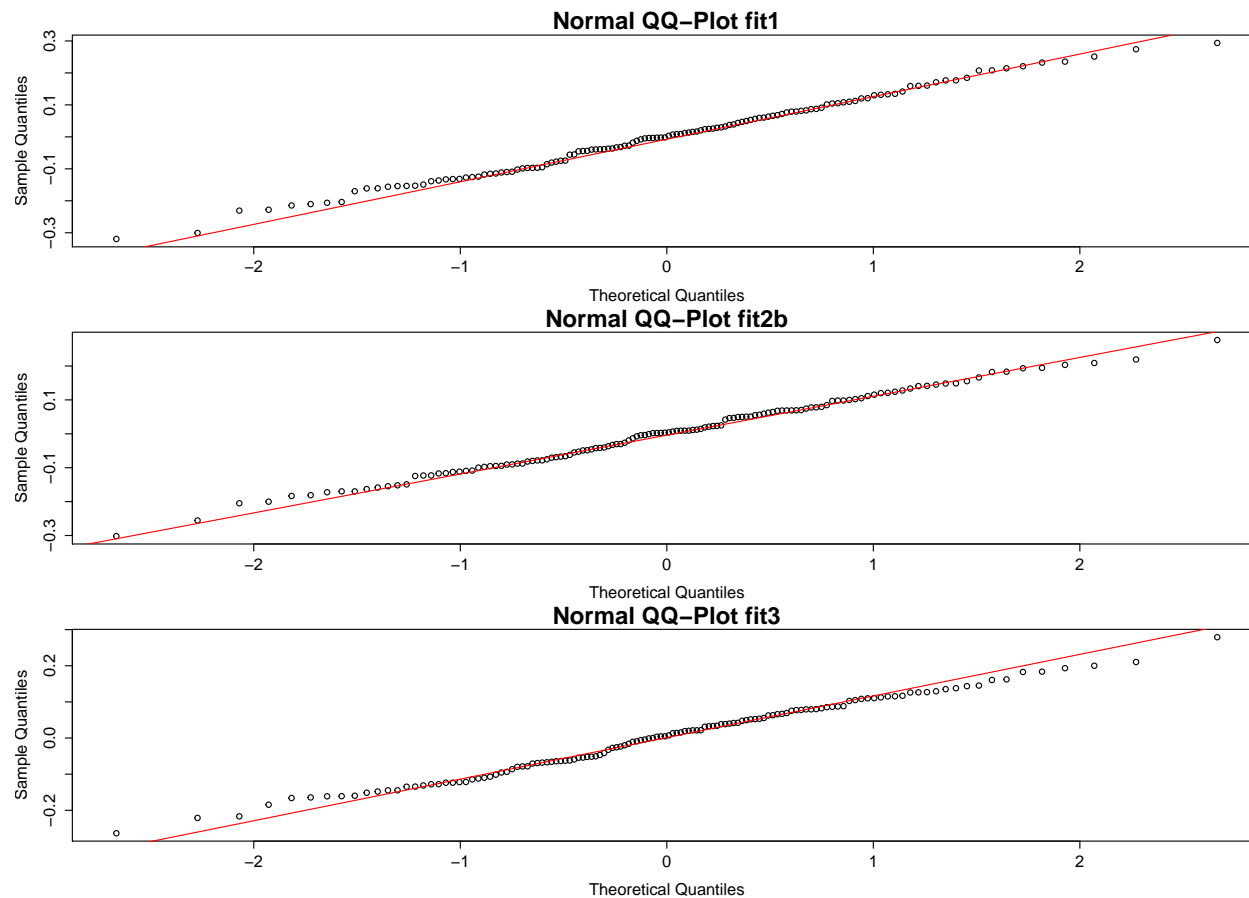
```
> AIC(fit1, fit2, fit2b, fit3)
      df      AIC
fit1   3 -167.4368
fit2   4 -198.1861
fit2b  3 -200.0730
fit3   5 -207.1945
> BIC(fit1, fit2, fit2b, fit3)
      df      BIC
fit1   3 -158.8342
fit2   4 -186.7160
fit2b  3 -191.4704
fit3   5 -192.8568
```

In R besteht die Möglichkeit, Modelle mit besserem AIC sukzessive mittels “add1” oder “drop1” zu wählen (s. Help dazu...)

Güte der Anpassung

Sind die Residuen (u_t) unabhängig und normalverteilt? Leider selten...

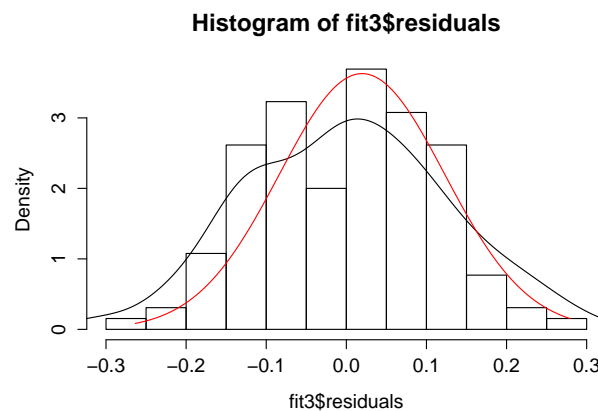
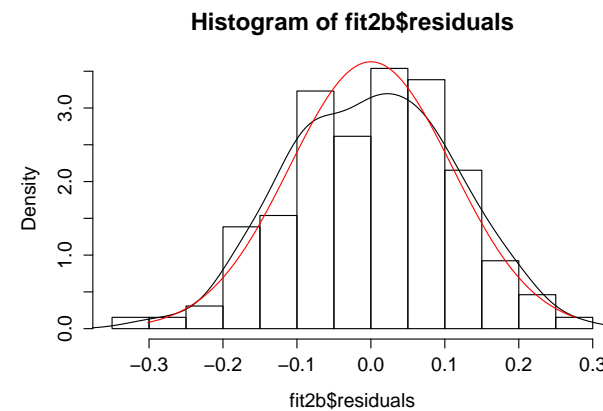
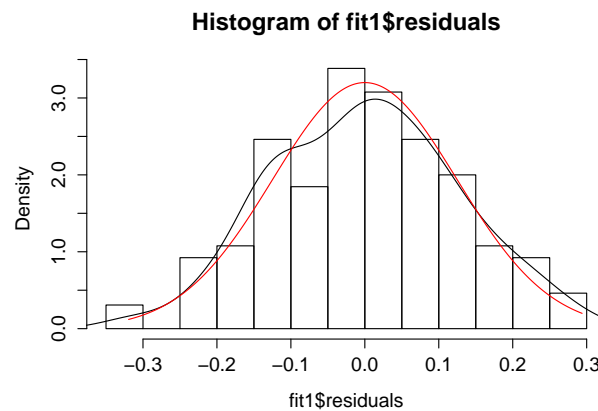
Überprüfung der Normalverteilungsannahme durch **QQ-Plots**:



Güte der Anpassung

Sind die Residuen (u_t) unabhängig und normalverteilt? Leider selten...

Überprüfung der Normalverteilungsannahme durch **Histogramm**:



Güte der Anpassung

```
par(mfrow=c(3,1), mar=c(4,4,2,1))
qqnorm(fit1$residuals, main="Normal QQ-Plot fit1") # qq-plots
qqline(fit1$residuals, col=2)
qqnorm(fit2b$residuals, main="Normal QQ-Plot fit2b")
qqline(fit2b$residuals, col=2)
qqnorm(fit3$residuals, main="Normal QQ-Plot fit3")
qqline(fit3$residuals, col=2)

par(mfrow=c(2,2))
hist(fit1$residuals, prob=TRUE) # histogram
lines(density(fit1$residuals))
dn=dnorm(x=seq(min(fit1$residuals),max(fit1$residuals),length.out=500), mean(fit1$residuals), sd(fit1$residuals))
lines(x=seq(min(fit1$residuals),max(fit1$residuals),length.out=500), dn, col=2)

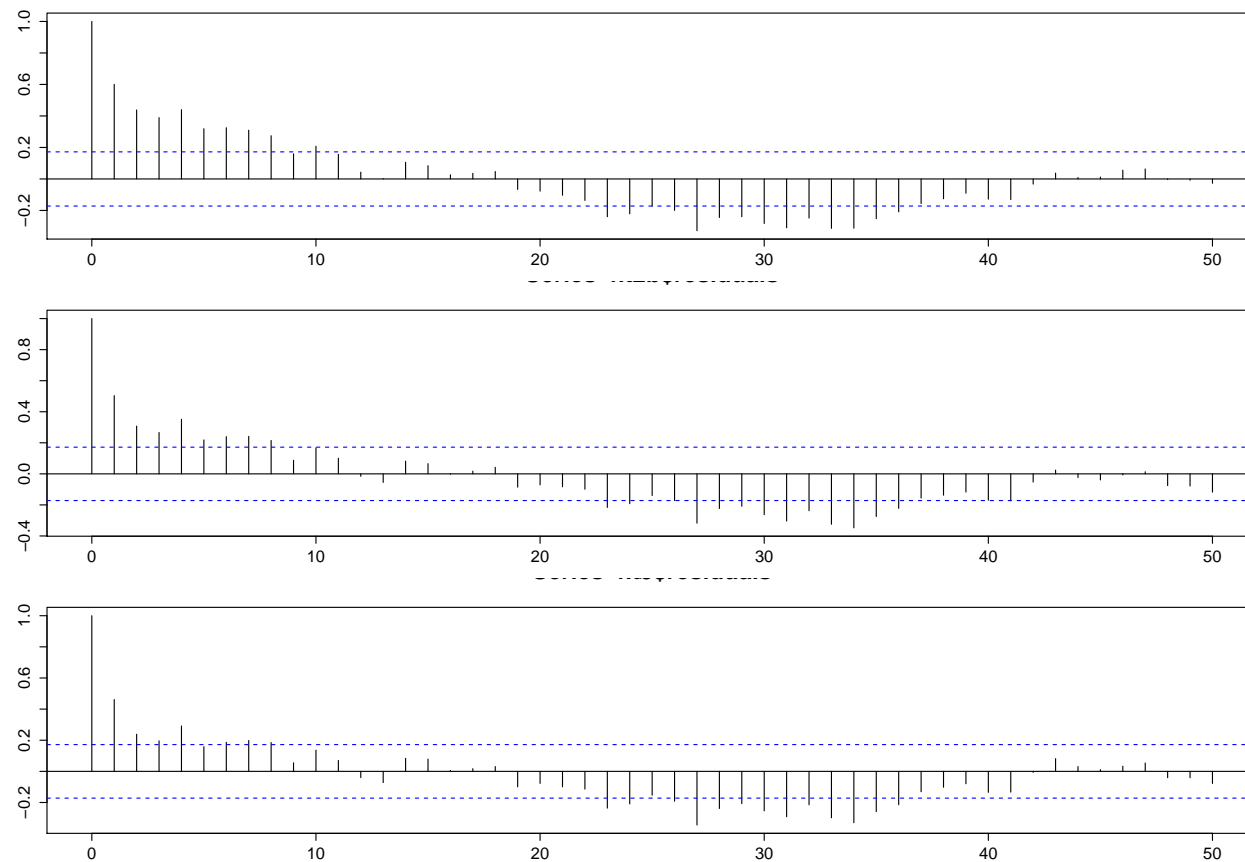
hist(fit2b$residuals, prob=TRUE)
lines(density(fit2b$residuals))
dn2=dnorm(x=seq(min(fit2b$residuals),max(fit2b$residuals),length.out=500), mean(fit2b$residuals), sd(fit2b$residuals))
lines(x=seq(min(fit2b$residuals),max(fit2b$residuals),length.out=500), dn2, col=2)

hist(fit3$residuals, prob=TRUE)
lines(density(fit3$residuals))
dn3=dnorm(x=seq(min(fit3$residuals),max(fit3$residuals),length.out=500), mean(fit3$residuals), sd(fit3$residuals))
lines(x=seq(min(fit3$residuals),max(fit3$residuals),length.out=500), dn3, col=2)
```

Güte der Anpassung

Die **Unabhängigkeit** ist kritisch...

Überprüfung der Unabhängigkeitsannahme durch **Korrelogramm**:



```
> par(mfrow=c(3,1), mar=c(3,2.5,2,1), cex.main=1.5)
> acf(fit1$residuals, 40)
> acf(fit2b$residuals, 40)
> acf(fit3$residuals, 40)
> Box.test(fit1$residuals, lag=20, type="Ljung-Box")
```

Box-Ljung test

```
data: fit1$residuals
X-squared = 191.71, df = 20, p-value < 2.2e-16
```

```
> Box.test(fit2b$residuals, lag=20, type="Ljung-Box")
```

Box-Ljung test

```
data: fit2b$residuals
X-squared = 112.92, df = 20, p-value = 5.773e-15
```

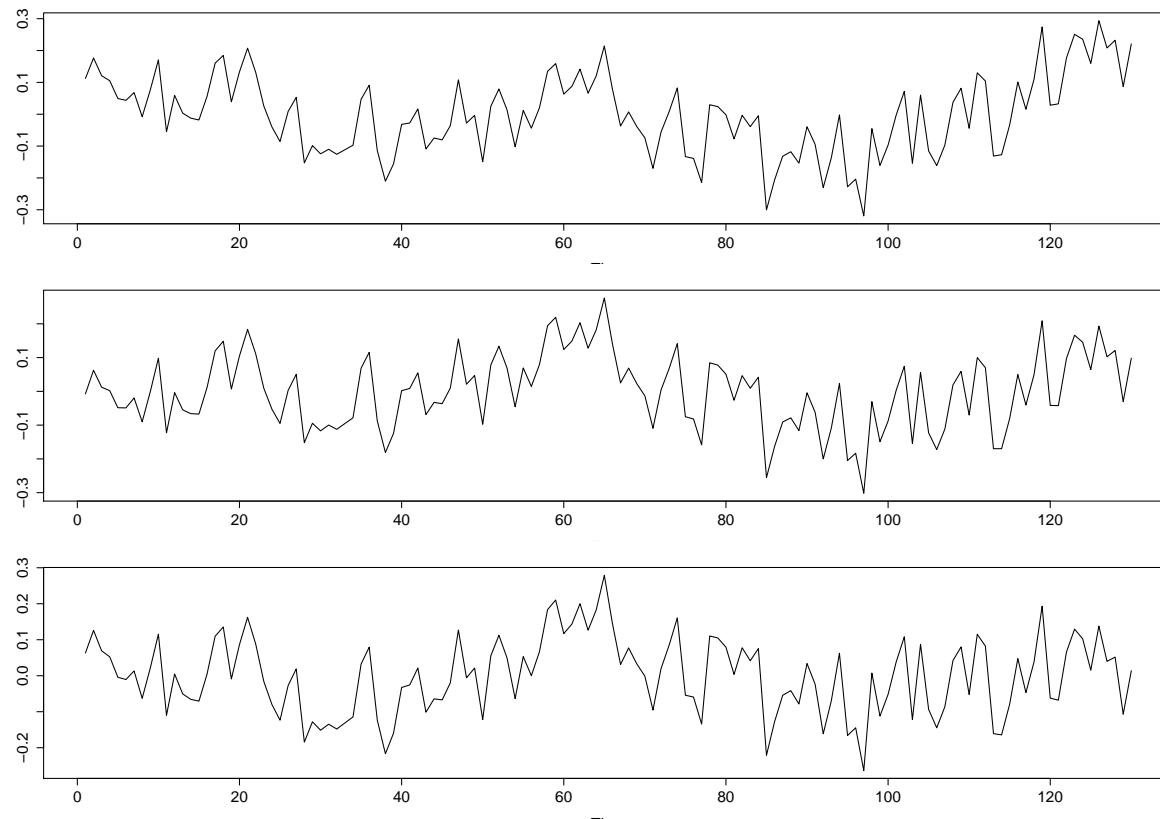
```
> Box.test(fit3$residuals, lag=20, type="Ljung-Box")
```

Box-Ljung test

```
data: fit3$residuals
X-squared = 80.803, df = 20, p-value = 2.867e-09
```

Güte der Anpassung

Zumindest sehen die “trendbereinigten” Zeitreihen $u_t = x_t - m_t$ schon etwas **stationärer** aus:



Güte der Anpassung

Zumindest sehen die “trendbereinigten” Zeitreihen $u_t = x_t - m_t$ schon etwas **stationärer** aus

→ Man könnte versuchen, die Residuen bzw. die Reihe selbst durch ARIMA Prozesse zu modellieren. Bessere Stationarität erhält man durch Differenzenbildung – später mehr dazu.

Gliederung der Vorlesung

- 1) Einführung und Grundlagen
- 2) Klassische Zeitreihenanalyse
 - a) Trendbestimmung
 - b) Glättung von Zeitreihen durch Filter (lokale Trendbestimmung)
 - c) Saisonbereinigung
 - d) Exponentielles Glätten, Prognosen
- 3) Modellierung von Reihen durch stochastische Prozesse
 - a) Autoregressive (AR) Modelle
 - b) Moving Average (MA) Modelle
 - c) ARMA Modelle
 - d) ARIMA Modelle

Lokale Trendbestimmung

Man legt um jeden Beobachtungswert x_t einen **Stützbereich** der Länge $M < N$ und ersetzt den Wert x_t durch die Schätzung aus diesem Bereich. Die einfachste Methode ist **lokale Mittelwerte** als Schätzung zu nehmen → **einfache gleitende Durchschnitte (moving averages)**:

$$v_t = \frac{1}{2n+1} \sum_{j=-n}^n x_{t+j} \quad \text{für } t = n+1, \dots, N-n$$

hier der **ungeraden** Ordnung $M = 2n + 1$. Z.B. für

$$n = 1, M = 3 : \quad v_t = \frac{1}{3}(x_{t-1} + x_t + x_{t+1})$$

$$n = 2, M = 5 : \quad v_t = \frac{1}{5}(x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2})$$

Gleitende Durchschnitte

Soll M **gerade** sein ($M = 2n$), so nimmt man noch einen Punkt hinzu, und die Randwerte gehen nur mit dem halben Gewicht ein:

$$v_t = \frac{1}{2n} \left(\frac{x_{t-n}}{2} + \sum_{j=-n+1}^{n-1} x_{t+j} + \frac{x_{t+n}}{2} \right) \quad \text{für } t = n+1, \dots, N-n$$

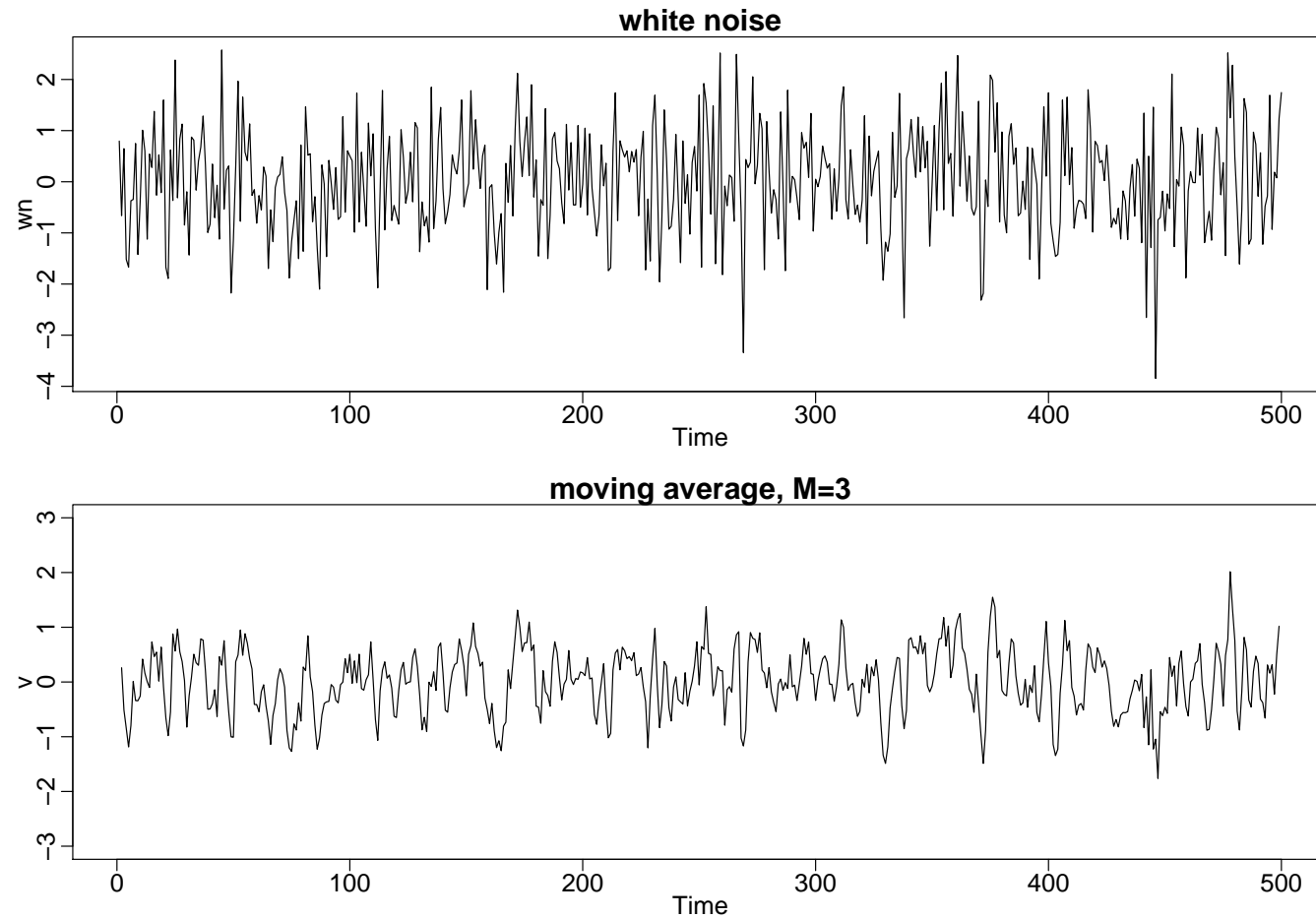
Z.B. für

$$n = 2, M = 4: \quad v_t = \frac{1}{4} \left(\frac{1}{2}x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2}x_{t+2} \right).$$

Gleitende Durchschnitte

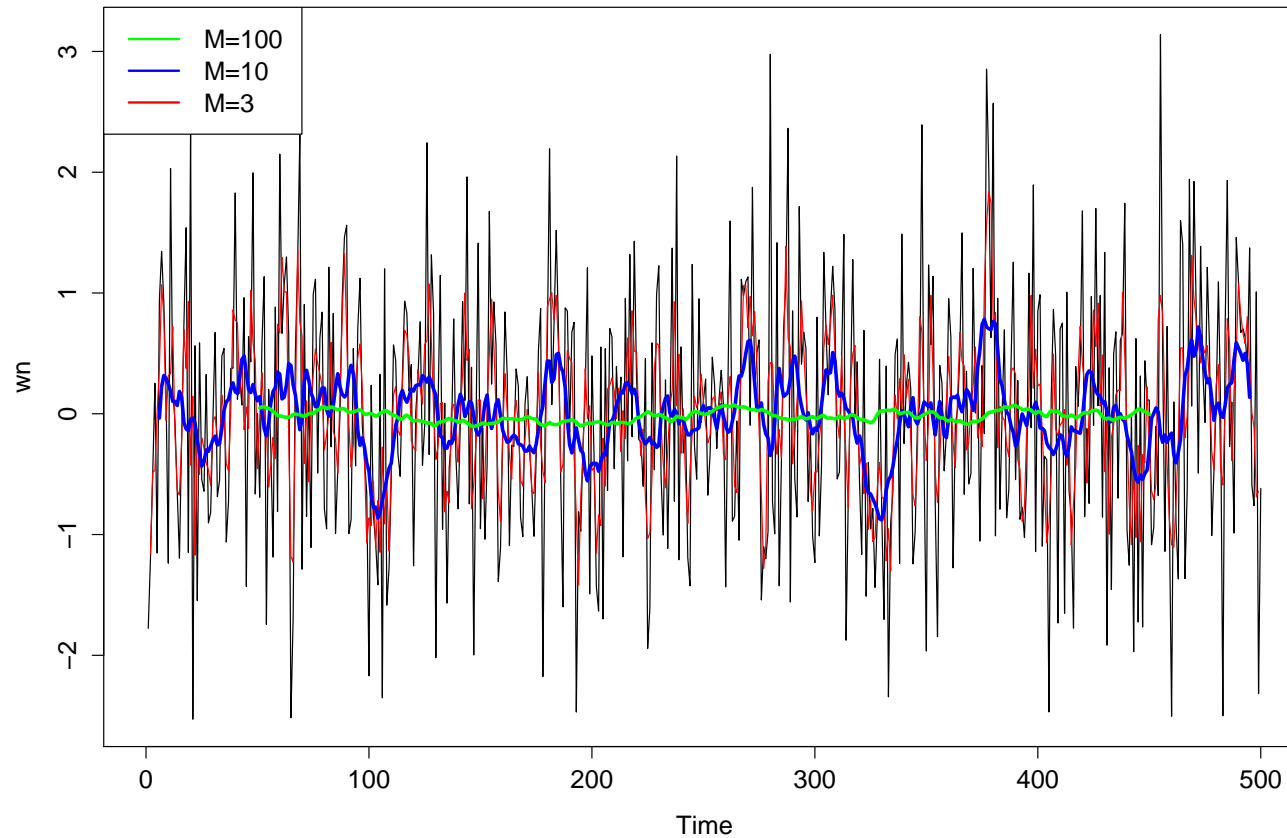
- Einfache gleitende Durchschnitte **glätten** die Zeitreihe.
- Die geglättete Zeitreihe (v_t) besitzt nur noch $N - 2n$ Werte: An den Rändern fallen jeweils n Werte weg.
- Je größer der Stützbereich M , desto stärker die Glättung und kürzer die geglättete Reihe.

Gleitende Durchschnitte: Beispiel



Gleitende Durchschnitte: Beispiel

Glättung von White Noise



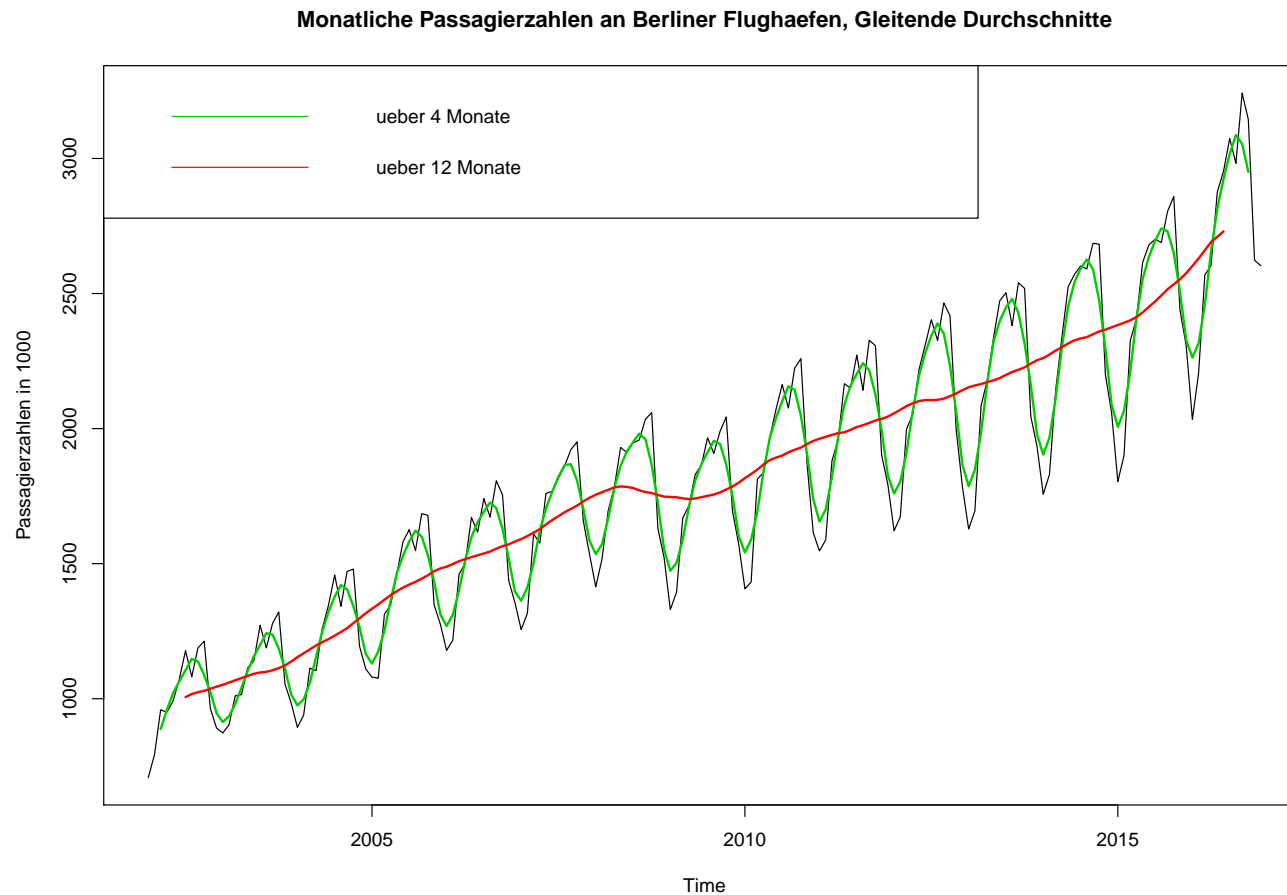
Gleitende Durchschnitte: Beispiel

```
#moving average
wn = rnorm(500,0,1) # 500 N(0,1) simulationen
v = filter(wn, sides=2, filter=rep(1/3,3)) # moving average
par(mfrow=c(2,1))
plot.ts(wn, main="white noise")
plot.ts(v, ylim=c(-3,3), main="moving average, M=3")

#moving average als Funktion
MA=function(mydata, n){d=n%%2
if(d==1){wg=rep(1/n,n)}else{wg=c(.5, rep(1,n-1), .5)/n}
ans= filter(mydata, sides=2, filter=wg)
return(ans)
}
v3 = MA(wn, 3)
v10 = MA(wn, 10)
v100 = MA(wn, 100)

plot.ts(wn, main="Glättung von White Noise")
lines(v3, col="red", lwd=2)
lines(v10, col="blue", lwd=3)
lines(v100, col="green", lwd=3)
colors <- c("green","blue","red")
legend("topleft", c(paste("M=100"),
                    paste("M=10"),
                    paste("M=3")),
      lwd = 3, cex=1, col=colors, pt.lwd = 1)
```

Gleitende Durchschnitte: Beispiel



→ **Saisoneinflüsse** werden hier eliminiert durch $M = 12$!

Gleitende Durchschnitte

Anstelle der einfachen können auch **gewichtete Mittelwerte**

$$v_t = \sum_{j=-n}^n a_j x_{t+j} \quad \text{für } t = n+1, \dots, N-n$$

betrachtet werden, bei denen die Gewichte (a_j) nicht alle gleich groß sind. Der zugehörige gleitende Durchschnitt wird mit

$$(a_{-n}, \dots, a_n)$$

notiert. Es gilt immer $\sum_{j=-n}^n a_j = 1$.

Z.B. der *symmetrische gleitende 15-ner Durchschnitt von Spencer*:

$$\frac{1}{320}(-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3)$$

→ Glättung von Sterbetafeln.

Lineare Filter

Gleitende Durchschnitte sind ein Spezialfall von **linearen Filtern**.

Definition 5 Ein linearer Filter $(a_j)_{j=-n,\dots,s}$ ist eine Folge von Koeffizienten, so dass die Zeitreihe (x_t) in einer Zeitreihe (y_t) mittels

$$y_t = \sum_{j=-n}^s a_j x_{t+j}, \quad t = n+1, \dots, N-s$$

transformiert wird. (y_t) heißt **Output** des Filters, (x_t) **Input**.

Ein Filter heißt **symmetrisch**, falls $s = n$ und $a_j = a_{-j}$ für alle j gilt.

Z.B. für $a_j = \frac{1}{2n+1}$ für alle $j = -n, \dots, 0, \dots, n$ erhalten wir den einfachen gleitenden Durchschnitt.

Lemma 6 Ist der Input (X_t) ein stationärer Prozess und $(a_j)_{j=-n,\dots,s}$ ein linearer Filter, so ist der Output des Filters (Y_t) wieder stationär.

Differenzenfilter

Eine Möglichkeit der Trendbereinigung ist die **Differenzenbildung**:

- Die ersten Differenzen sind definiert als

$$\nabla x_t = x_t - x_{t-1} =: (1 - B)x_t,$$

wobei B den **Backshift-Operator** $B(x_t) := x_{t-1}$ bezeichnet.

- Differenzen 2. Ordnung erhält man durch zweimalige Anwendung:

$$\nabla^2 x_t = \nabla(\nabla x_t) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2}$$

oder mit dem Backshift-Operator

$$\nabla^2 x_t = (1 - B)^2 x_t = (1 - 2B + B^2)x_t = x_t - 2x_{t-1} + x_{t-2}.$$

- Entsprechend sind Differenzen höherer Ordnung definiert:

$$\nabla^d = (1 - B)^d.$$

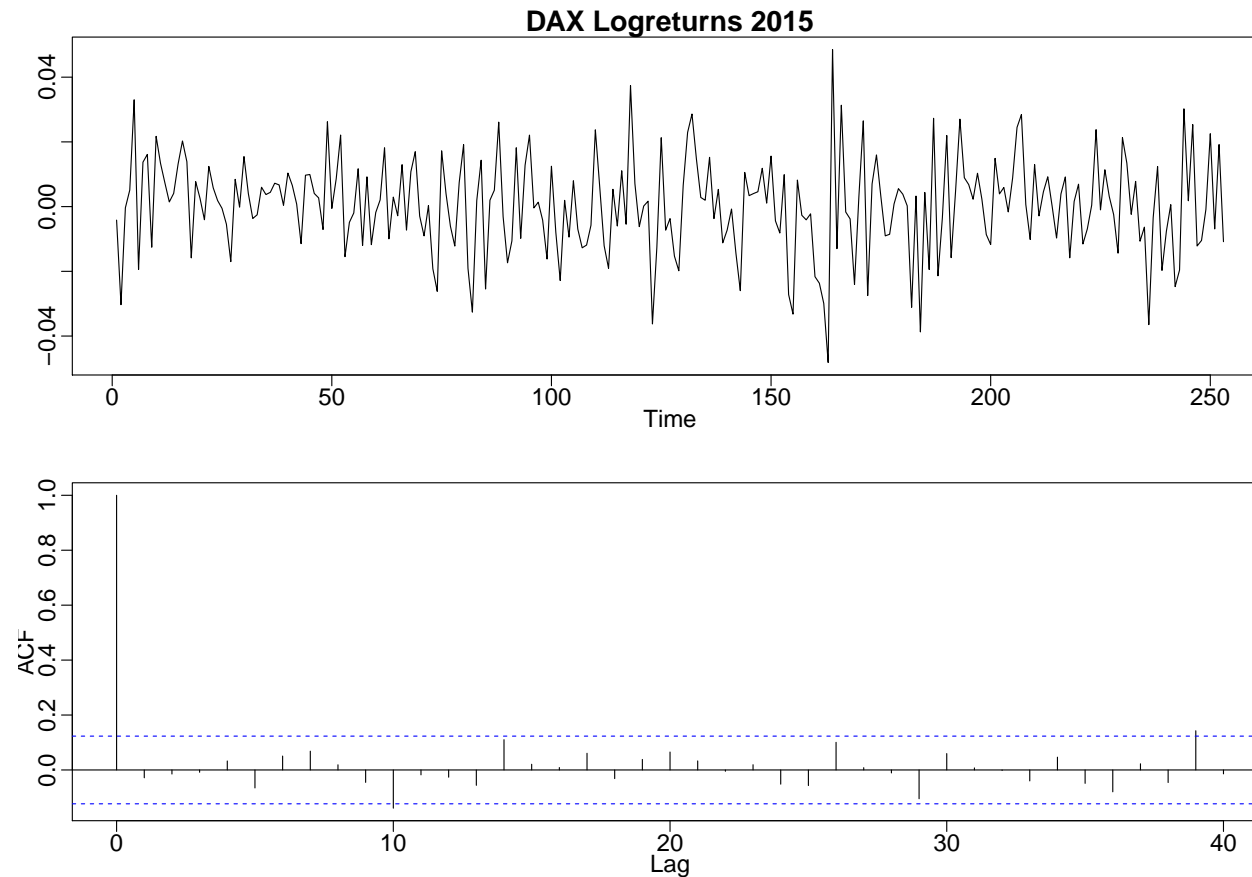
Differenzenfilter

- Differenzenbildung ist ein linearer (asymmetrischer) Filter mit den Koeffizienten ($a_{-1} = -1, a_0 = 1$).
- Differenzenbildung führt oft zu stationären Prozessen. Ein Prozess (X_t) heißt integriert vom Grad d (Typ $I(d)$) falls $\nabla^d X_t$ stationär ist. (Wir haben bereits gesehen, dass Random Walk vom Typ $I(1)$ ist.)
- Ist $x_t = \beta_0 + \beta_1 t + u_t$, wobei die Residuen (u_t) stationär sind, so führt die Differenzenbildung 1. Ordnung zu einem stationären Prozess:

$$\nabla x_t = \beta_1 + \nabla u_t.$$

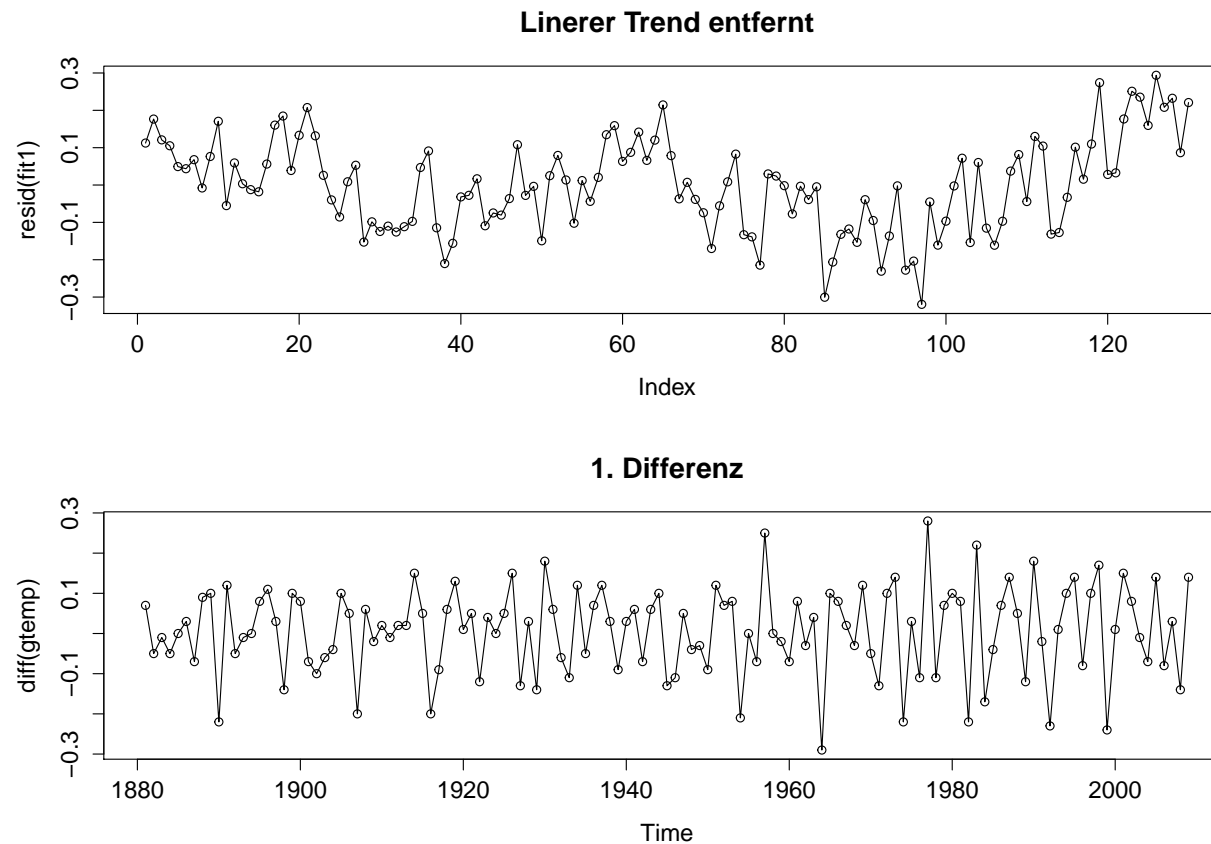
- Analog führt bei den polynomialen Trends vom Grad d der Differenzenfilter d -ter Ordnung zu Stationarität. (In den Anwendungen ist selten mehr als 2 mal Differenzenbildung nötig.)

Differenzenfilter: Beispiele



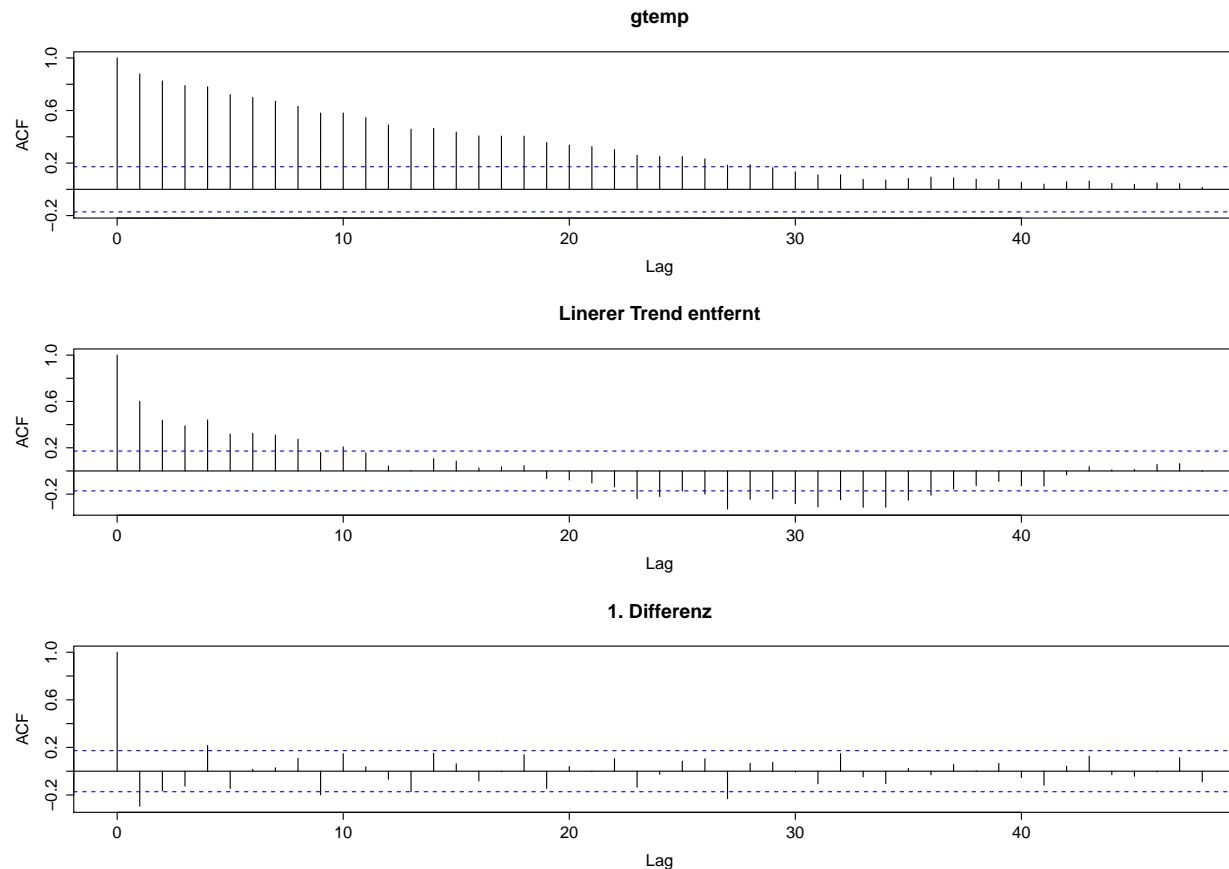
Hier zuerst logarithmieren, dann Differenzenbildung → sieht stationär aus.

Differenzenfilter: Beispiele



Trendbereinigung: Linearen Trend abziehen vs. Differenzenbildung von “gtemp”.

Differenzenfilter: Beispiele



Nach Differenzenbildung **fast** wie White Noise, also kann man globale Temperatur ungefähr als Random Walk mit Drift sehen.

Differenzenfilter: Beispiele

```
> mean(diff(gtemp))  
[1] 0.006589147  
> sd(diff(gtemp))/sqrt(length(diff(gtemp))) # = 0.00966 (SE)  
[1] 0.009658972  
> t=c(1:length(gtemp))  
> (fit0=lm(gtemp~t))
```

Call:

```
lm(formula = gtemp ~ t)
```

Coefficients:

(Intercept)	t
-0.398140	0.005748

→ Globale Erwärmung von etwa 0,006 bzw. 0,007 Grad pro Jahr.

https://de.wikipedia.org/wiki/Globale_Erw%C3%A4rmung

Differenzenfilter: Beispiele

Saisonale Effekte können durch die **saisonale Differenz** (1. Ordnung oder höher) eliminiert werden:

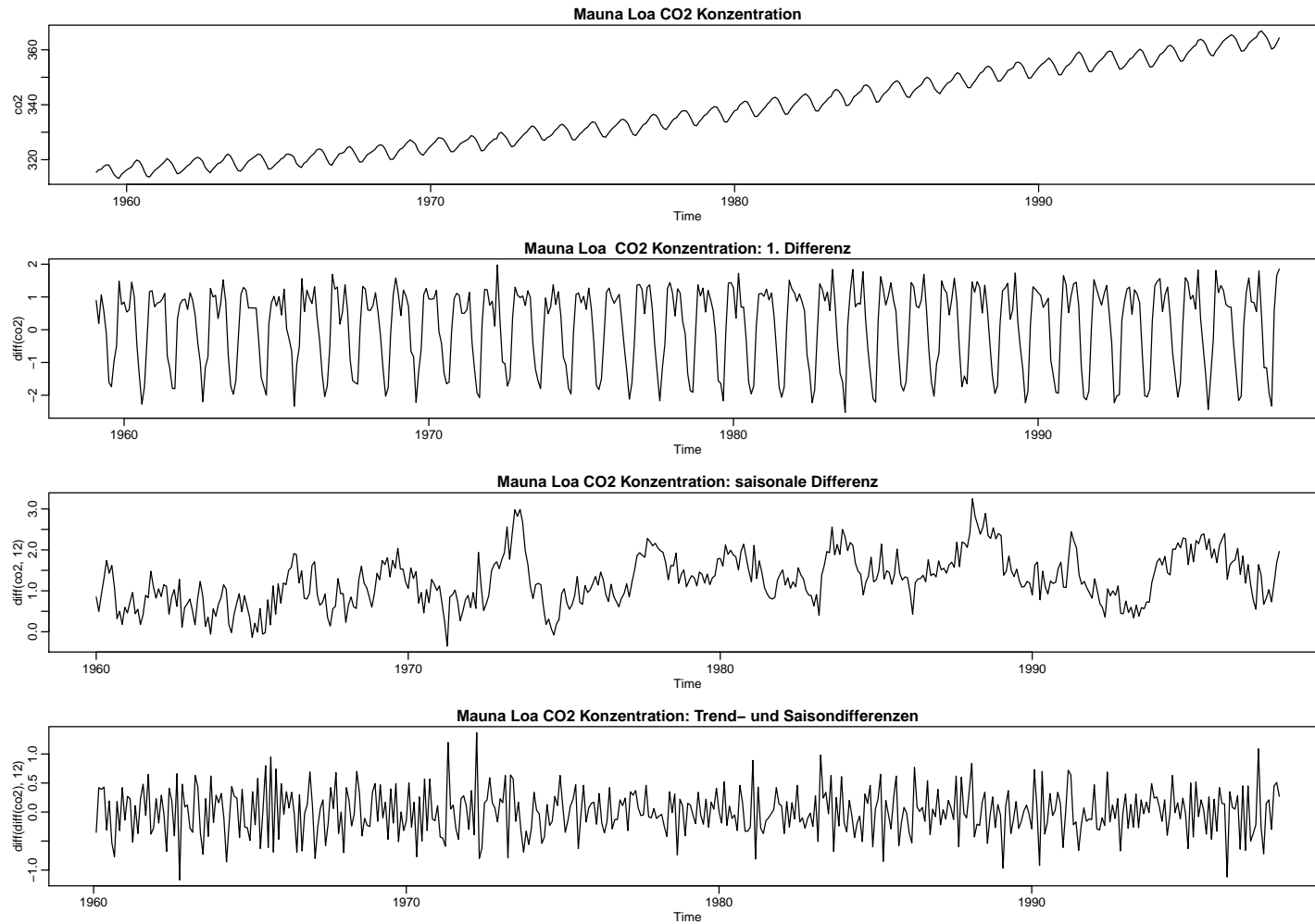
$$\nabla_s x_t = x_t - x_{t-s} = (1 - B^s)(x_t),$$

wobei s die Saisonlänge ist, z.B. 12 (Monate).

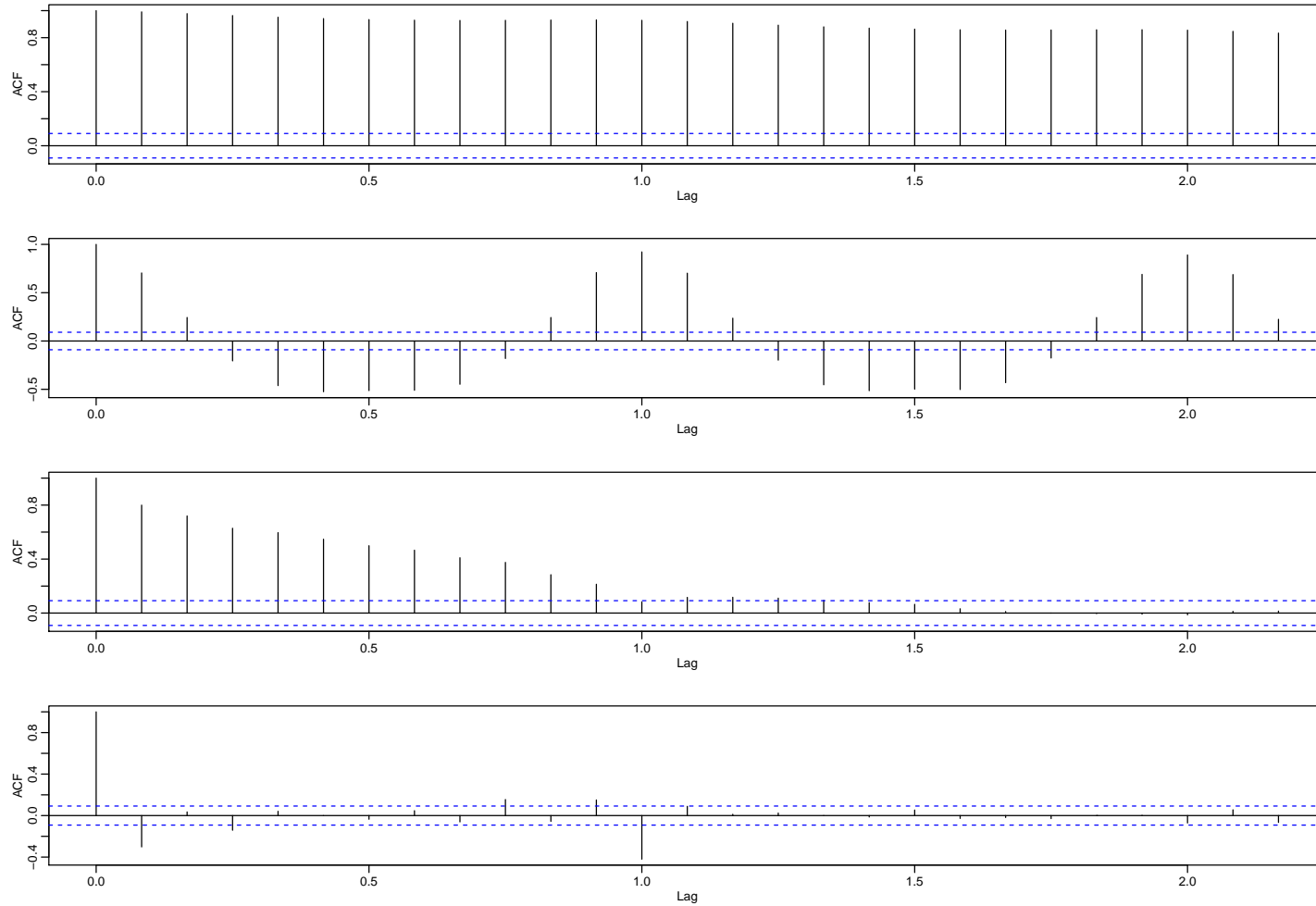
Oft ist es nötig beides, **Trend und Saison** zu bereinigen. Dann wendet man den Differenzenfilter, bis der Trend bereinigt ist, und die saisonalen Differenzen, bis die Saison bereinigt ist:

$$\nabla_s^k \nabla^d(x_t) = (1 - B^s)^k (1 - B)^d(x_t)$$

Differenzenfilter: Beispiele



Differenzenfilter: Beispiele



Differenzenfilter: Beispiele

```
frequency(co2)
par(mfrow=c(4,1), mar=c(3,2,1,0)+.5, mgp=c(1.6,.6,0))
plot(co2, main="Mauna Loa CO2 Konzentration")
plot(diff(co2), main="Mauna Loa CO2 Konzentration: 1. Differenz")
plot(diff(diff(co2),12), main="Mauna Loa CO2 Konzentration: saisonale Differenz")
plot(diff(diff(diff(co2),12),12), main="Mauna Loa CO2 Konzentration: Trend- und Saisondifferenzen")
par(mfrow=c(2,1), mar=c(3,2,1,0)+.5, mgp=c(1.6,.6,0))
acf(diff(co2,12))
acf(diff(diff(co2,12)))
```

Filter: Faltung

Faltung entspricht Hintereinanderausführung von Filtern

$$(x_t) \rightarrow \boxed{\text{Filter 1}} \rightarrow (y_t) \rightarrow \boxed{\text{Filter 2}} \rightarrow (z_t)$$

Sind beide Filter linear, der erste mit den Koeffizienten (a_j) , der zweite mit (b_i) , so ist die Faltung wieder ein linearer Filter mit den Koeffizienten

$$c_r = \sum_i b_i a_{r-i}$$

Schreibweise: $(c_r) = (a_j) * (b_i)$. Z.B. erhalten wir für $\nabla \nabla_{12}(x_t)$:

$$\begin{aligned} (a_{-12}, \dots, a_0) &= (-1, 0, \dots, 0, 1), & (b_{-1}, b_0) &= (-1, 1) \\ \rightarrow (c_{-13}, c_{-12}, \dots, c_0) &= (1, -1, 0, \dots, 0, -1, 1) \end{aligned}$$

Oder mit dem Backshift-Operator:

$$\nabla \nabla_{12} = (1 - B)(1 - B^{12}) = 1 - B^{12} - B + B^{13}$$