



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Master-Modul 85-011: Zeitreihenanalyse und Prognoserechnung

Sommersemester 2019
Hamburg

Dr. Cristina Sattarhoff
Institut für Statistik
Moorweidenstraße 18
20148 Hamburg

Kapitel 0:
Organisation und Hinweise

Inhaltsübersicht

Organisation und Hinweise

Materialien

Die Materialien zur Vorlesung und Übung können unter

<https://www.bwl.uni-hamburg.de/statistik/studium/sose2019/zeitreihenanalyse.html>

heruntergeladen werden.

Die Master-Vorlesung

Zeitreihenanalyse und Prognoserechnung

besteht aus 3 SWS Vorlesung und 1 SWS Übung.

Die Vorlesung findet donnerstags 14:15 – 16:45 Uhr mit einer 15-minütigen Pause im Hörsaal Wiwi B1 statt.

Die ersten zwei Übungen finden mittwochs 10:15 – 11:00 Uhr im Hörsaal Wiwi B1 statt. Ab dem 24. April werden die Übungen mittwochs 9:15 – 10:00 und bzw. 10:15 – 11:00 Uhr im Computerraum Wiwi 2043/2047 durchgeführt werden.

Die Vorlesung basiert auf den Lehrbüchern:

- Schlittgen, R. und Streitberg, B. H. J. (1999), *Zeitreihenanalyse*, 8. Auflage, Oldenbourg Verlag, München.
- Schlittgen, R. (2015), *Angewandte Zeitreihenanalyse mit R*, 3. Auflage, de Gruyter Oldenbourg Verlag, Berlin / München / Boston.

Literatur

Weiterführende Literatur

- Shumway, R. H. und Stoffer, D. S. (2011), *Time Series Analysis and Its Applications*, 3. Auflage, Springer Verlag, New York.
- Pfaff, B. (2008), *Analysis of Integrated and Cointegrated Time Series with R*, 2. Auflage, Springer Verlag, New York.
- Brockwell, P. J. und Davis, R. A. (2006), *Time Series: theory and methods*, 2. Auflage, Springer Verlag, New York.
- van der Vaart, A.W. (2013), *Time Series*, Skript, Universiteit Leiden, Leiden, abrufbar auf der Internetseite <http://www.math.leidenuniv.nl/~avdvaart/timeseries/index.html> unter Lecture Notes.
- Tsay, R. S. (2010), *Analysis of financial time series*, 3. Auflage, Wiley Verlag, Hoboken.

Am Ende der Vorlesung wird es eine 90-minütige Klausur geben. Erlaubte Hilfsmittel:

- handschriftlich selbst verfasste Notizen im Umfang von 10 DIN-A4-Seiten
- nichtprogrammierbarer, zweizeiliger Taschenrechner

Klausurtermine	Tag	Datum	Uhrzeit	Hörsaal
Semestertermin	Di	9.07.2019	18:15 - 19:45	Audimax 2
Ferientermin	Do	19.09.2019	14:15 - 15:45	Wiwi A

Vorwissen im Umfang der Pflichtmodule Mathematik und Statistik des B.Sc.-Studiengangs BWL wird vorausgesetzt. Eine kurze Wiederholung finden Sie in den Anhängen A, B und C aus Schlittgen und Streitberg (1999).

Folgende Begriffe bzw. Sachverhalte sollten grundsätzlich bekannt sein:

- Zufallsvariable und ihre Verteilung
- Momente einer Zufallsvariablen (Erwartungswert, Varianz, Kovarianz, Korrelation)
- Zufallsvektor und seine multivariate Verteilung
- statistischer Test, Konfidenzintervalle
- trigonometrische Funktionen

Sie erreichen mich unter:

Dr. Cristina Sattarhoff

Raum	0039.2 (Moorweidenstraße 18)
E-Mail	cristina.sattarhoff@uni-hamburg.de
Telefon	040-42838-1519
Sprechstunde	nach Vereinbarung

Kapitel 1:
Stochastische Prozesse

Inhaltsübersicht

Stochastische Prozesse

- Darstellung von Zeitreihen

- Schätzung der theoretischen Momente

 - Stationarität

 - Asymptotische Verteilungen

 - Ergodizität

Kapitel 1:

Stochastische Prozesse

1.1 Darstellung von Zeitreihen

Darstellung von Zeitreihen

Fallbeispiel Anzahlen von Luchspelzen

Der Datensatz enthält die jährlichen Anzahlen der Luchspelze, die im Zeitraum 1821–1934 in Kanada im McKenzie River District gefangen wurden.

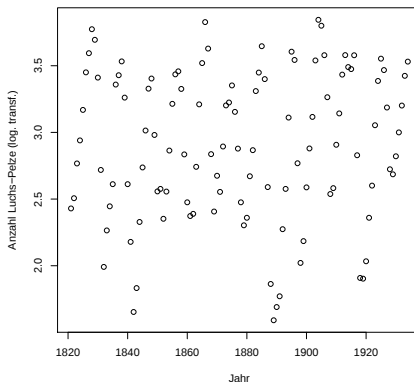


Abb. 1.1: Die Zeitreihe *Luchs* (Darstellung mit Punkten)

Darstellung von Zeitreihen

Fallbeispiel Anzahlen von Luchspelzen

Der Datensatz enthält die jährlichen Anzahlen der Luchspelze, die im Zeitraum 1821–1934 in Kanada im McKenzie River District gefangen wurden.

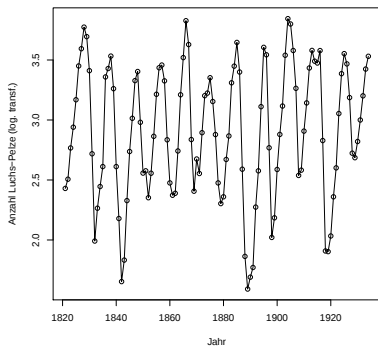


Abb. 1.2: Die Zeitreihe *Luchs* (Darstellung mit Punkten und Linien)

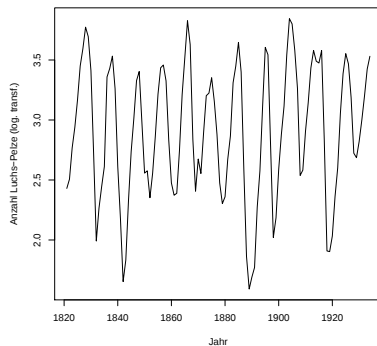


Abb. 1.3: Die Zeitreihe *Luchs* (Darstellung mit Linien)

Darstellung von Zeitreihen

Fallbeispiel Globaltemperatur

Der Datensatz stammt aus Shumway und Stoffer (2011).

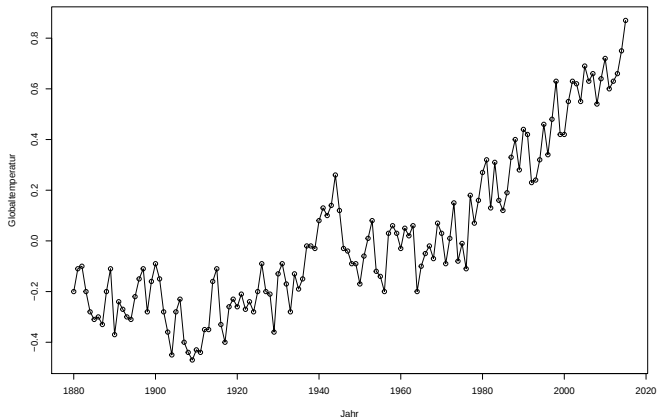


Abb. 1.4: Die Zeitreihe *globtemp*

Darstellung von Zeitreihen

Fallbeispiel Steuern

Der Datensatz enthält die in DE monatlich gezahlten Gemeinschaftssteuern in Mio. Euro, Jan. 1994 – Jan. 2000.

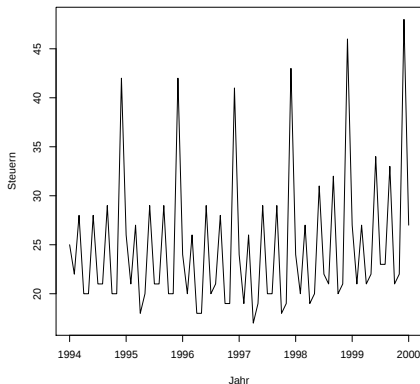


Abb. 1.5: Die Zeitreihe *steuern*

Definition (Stochastischer Prozess)

Ein **stochastischer Prozess** ist eine Folge $(Y_t)_{t \in T}$ von Zufallsvariablen Y_t . Der Index $t \in T$ wird als Zeit aufgefasst. Für die Indexmenge T wählen wir i.d.R. $T \subseteq \mathbb{N}, \mathbb{N}_0$ oder $\mathbb{Z}$.

Definition (Zeitreihe)

Eine **Zeitreihe** ist eine Folge $y_1, \dots, y_N$ von Realisationen eines Ausschnittes von $(Y_t)_{t \in T}$. Man nennt diese auch Zeitpfad oder Trajektorie des Prozesses.

Wenn es auf die Indexmenge T nicht ankommt, schreiben wir für einen stochastischen Prozess einfach (Y_t) , bzw. für eine Zeitreihe (y_t) .

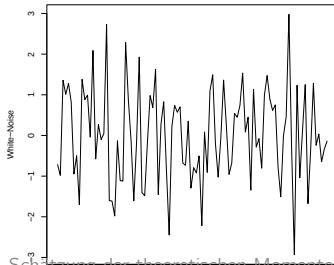
Definition (White-Noise-Prozess)

Ein **White-Noise-Prozess** ist eine Folge $(\varepsilon_t)_{t \in T}$ von unabhängigen und identisch verteilten Zufallsvariablen ε_t .

Im Allgemeinen wird ein White-Noise-Prozess als stochastischer Prozess $(\varepsilon_t)_{t \in T}$ mit $\mathbb{E}[\varepsilon_t] = 0$ und Autokovarianz-Funktion

$$\gamma_\varepsilon(h) = \begin{cases} \sigma^2, & \text{für } h = 0 \\ 0, & \text{für } h \neq 0 \end{cases}.$$

definiert.



Definition (Random-Walk-Prozess)

$(\varepsilon_t)_{t \in \mathbb{N}}$ sei ein White-Noise-Prozess. Ein **Random-Walk-Prozess** $(Y_t)_{t \in \mathbb{N}}$ ist definiert durch die Rekursion

$$Y_t := \begin{cases} \varepsilon_1, & t = 1 \\ Y_{t-1} + \varepsilon_t, & t = 2, 3, \dots \end{cases}$$

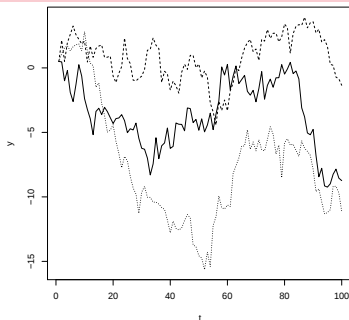


Abb. 1.7: Drei Realisationen eines Random-Walk-Prozesses

Empirische Momente

Sei $(y_t)_{t=1,2,\dots,N}$ eine Zeitreihe.

- Das arithmetische Mittel $\bar{y}$ beschreibt die zentrale Lage der Werte von (y_t)

$$\bar{y} := \frac{1}{N} \sum_{t=1}^N y_t.$$

- Die empirische Standardabweichung s , bzw. die empirische Varianz s^2 beschreiben die Stärke der Schwankung der Zeitreihe um $\bar{y}$

$$s^2 := \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y})^2.$$

- Für zwei Zeitreihen $(y_t)_{t=1,2,\dots,N}$ und $(x_t)_{t=1,2,\dots,N}$ gibt die empirische Kovarianz die Stärke des linearen Zusammenhangs an

$$c_{XY} := \frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})(y_t - \bar{y}).$$

- Der empirische Korrelationskoeffizient ist

$$r_{XY} := \frac{c_{XY}}{s_X s_Y}.$$

Empirische Momente

Die ACF

Die empirische Autokovarianzfunktion $(c_t)_{\tau \geq 0}$ mit

$$c_\tau := \frac{1}{N} \sum_{t=1}^{N-\tau} (y_t - \bar{y}) (y_{t+\tau} - \bar{y})$$

Die empirische Autokorrelationsfunktion, kurz ACF, $(r_\tau)_{\tau \geq 0}$ mit

$$r_\tau := \frac{c_\tau}{c_0}.$$

τ bezeichnet den Zeitabstand oder den Lag, $\tau \geq 0$. Für $\tau < 0$ wird $c_\tau = c_{-\tau}$ und $r_\tau = r_{-\tau}$ gesetzt. Der Graph der empirischen ACF heißt Korrelogramm.

Empirische Momente

Die ACF

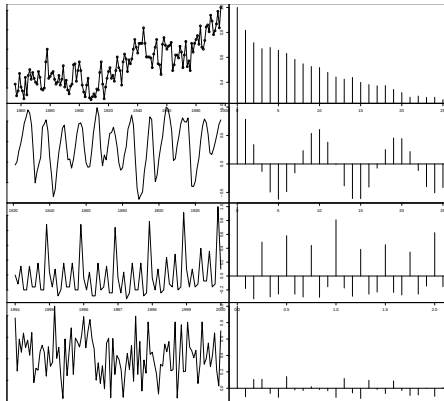


Abb. 1.8: Zeitreihen und ihre ACFs

Theoretische Momente

Seien Y und X Zufallsvariablen und quadratintegrierbar. Dann sind folgende Größen definiert:

- Erwartungswert

$$\mathbb{E}[Y]$$

- Varianz

$$\text{Var}[Y] := \mathbb{E}[(Y - \mathbb{E}[Y])^2]$$

- Kovarianz

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

- Korrelation

$$\text{Corr}[Y, X] := \frac{\text{Cov}[Y, X]}{\sqrt{\text{Var}[Y] \text{Var}[X]}}$$

Theoretische Momente

Die ACF

Sei (Y_t) ein *stationärer* stochastischer Prozess.

Die theoretische Autokovarianzfunktion $(\gamma_\tau)_{\tau \in \mathbb{Z}}$ mit

$$\gamma_\tau := \text{Cov}[Y_t, Y_{t+\tau}]$$

Die theoretische Autokorrelationsfunktion, kurz ACF, $(\rho_\tau)_{\tau \in \mathbb{Z}}$ mit

$$\rho_\tau := \frac{\gamma_\tau}{\gamma_0}$$

Es gelten die Beziehungen $\gamma_\tau = \gamma_{-\tau}$ bzw. $\rho_\tau = \rho_{-\tau}$.

Kapitel 1:

Stochastische Prozesse

1.2 Schätzung der theoretischen Momente

Schätzung der theoretischen Momente

Schätzfunktion

Sei Y eine Zufallsvariable mit dem unbekannten Verteilungsparameter β . Es liege eine Zufallsstichprobe $y_1, \dots, y_N$ bzw. eine Zeitreihe, $(y_1, \dots, y_N)$, vor. N bezeichnet die Stichprobengröße oder die Zeitreihenlänge.

Sei $\hat{\beta}_N$ eine Schätzfunktion (ein Schätzer) für β .

Die Schätzfunktion $\hat{\beta}_N$ heißt **konsistent**, falls

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[\left(\hat{\beta}_N - \beta \right)^2 \right] = 0.$$

Schätzung der theoretischen Momente

Schätzfunktion

► Klein-Stichproben-Eigenschaften:

- **Erwartungstreue (Unverzerrtheit):** $\hat{\beta}_N$ ist erwartungstreu (unverzerrt) für β , falls $\mathbb{E}[\hat{\beta}_N] = \beta$ gilt.
- **Minimale Varianz:** $\hat{\beta}_N$ heißt varianzminimal, falls $\text{Var}[\hat{\beta}_N] \leq \text{Var}[\tilde{\beta}_N]$ für alle anderen Schätzer $\tilde{\beta}_N$ von β gilt.
- **Mittlerer quadratischer Fehler (mean squared error, MSE):**
 $\text{MSE}[\hat{\beta}_N] = \mathbb{E}\left[\left(\hat{\beta}_N - \beta\right)^2\right]$. Der MSE eines erwartungstreuen Schätzers ist gleich der Varianz des Schätzers.

► Asymptotische Eigenschaften:

- **Asymptotische Erwartungstreue:** $\hat{\beta}_N$ heißt asymptotisch erwartungstreu, falls $\lim_{N \rightarrow \infty} \mathbb{E}[\hat{\beta}_N] = \beta$ gilt.

- **Konsistenz (Konvergenz im quadratischen Mittel):** $\hat{\beta}_N$ heißt konsistent, falls

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[\left(\hat{\beta}_N - \beta \right)^2 \right] = 0.$$

- **Schwache Konsistenz (Konvergenz nach Wahrscheinlichkeit):** $\hat{\beta}_N$ heißt schwach konsistent, falls für jedes $\varepsilon > 0$ gilt

$$\lim_{N \rightarrow \infty} \mathbb{P} \left[\left| \hat{\beta}_N - \beta \right| > \varepsilon \right] = 0.$$

Schätzung der theoretischen Momente

Annahmen

In den BA-Vorlesungen zur Statistik haben Sie bereits die Schätzung der theoretischen Momente einer Zufallsvariablen Y auf der Grundlage von unabhängigen Versuchswiederholungen (Zufallsstichproben) behandelt. Bspw. Schätzung des Erwartungswerts von Y mit Hilfe des arithmetischen Mittels $\bar{Y}_N$.

Seien $Y_1, Y_2, \dots, Y_N$ die Stichprobenvariablen. Wir benötigen zwei Annahmen:

1. Die Stichprobenvariablen sind identisch verteilt.

und

2. Die Stichprobenvariablen sind unabhängig voneinander.

Schätzung der theoretischen Momente

Annahmen

Wie können wir die theoretischen Momente für *abhängige* Versuchsergebnisse, nämlich Zeitreihen, schätzen?

Wir benötigen zwei Annahmen:

1. Stationarität

und

2. Ergodizität

Stationarität

Ein stochastischer Prozess $(Y_t)_{t \in T}$ heißt

- mittelwertstationär, wenn $\mathbb{E}[Y_t]$ konstant ist:

$$\mathbb{E}[Y_t] = \mu \quad \text{für alle } t \in T$$

- varianzstationär, wenn $\text{Var}[Y_t]$ konstant ist:

$$\text{Var}[Y_t] = \sigma^2 \quad \text{für alle } t \in T$$

- kovarianzstationär, wenn $\text{Cov}[Y_{t+\tau}, Y_t]$ nur vom Lag τ abhängt:

$$\text{Cov}[Y_{t+\tau}, Y_t] = \gamma_\tau \quad \text{für alle } t \in T, \tau \geq 0$$

- schwach stationär, wenn er mittelwert- und kovarianzstationär ist
- streng stationär, wenn für jede endliche Auswahl von Zeitpunkten $t_1, t_2, \dots, t_k$ die Zufallsvariablen $Y_{t_1}, Y_{t_2}, \dots, Y_{t_k}$ identisch multivariat verteilt sind.

Für das Weitere reicht die Forderung der schwachen Stationarität, die wir auch einfach *Stationarität* nennen werden.

Sei (Y_t) ein stationärer Prozess. Wir betrachten

- das arithmetische Mittel $\bar{Y}_N := \frac{1}{N} \sum_{t=1}^N Y_t$ als Schätzfunktion für den Erwartungswert $\mu = \mathbb{E}[Y_t]$
- die empirische Autokovarianz $\hat{\gamma}_N(\tau) := c_\tau = \frac{1}{N} \sum_{t=1}^{N-\tau} (Y_t - \bar{Y}_N) (Y_{t+\tau} - \bar{Y}_N)$ als Schätzfunktion für die theoretische Autokovarianz $\gamma_\tau = \mathbb{E}[(Y_t - \mu)(Y_{t+\tau} - \mu)]$
- die empirische Autokorrelation $\hat{\rho}_N(\tau) := r_\tau = \frac{\hat{\gamma}_N(\tau)}{\hat{\gamma}_N(0)}$ als Schätzfunktion für die theoretische Autokorrelation $\rho_\tau = \frac{\gamma_\tau}{\gamma_0}$.

Asymptotische Verteilungen

Das arithmetische Mittel

Sei (Y_t) ein stationärer Prozess mit Erwartungswert $\mu = \mathbb{E}[Y_t]$ und Autokovarianzfunktion $\gamma_\tau = \mathbb{E}[(Y_t - \mu)(Y_{t+\tau} - \mu)]$. Die Schätzfunktion $\bar{Y}_N$ hat die folgenden Eigenschaften:

- $\bar{Y}_N$ ist erwartungstreu: $\mathbb{E}[\bar{Y}_N] = \mu$
- $\text{Var}[\bar{Y}_N] = \text{MSE}[\bar{Y}_N] = \frac{1}{N} \left(\gamma_0 + 2 \sum_{\tau=1}^{N-1} \frac{N-\tau}{N} \gamma_\tau \right)$
- $\bar{Y}_N$ ist konsistent, wenn (γ_τ) absolut summierbar ist.

Zur Erinnerung: Eine Folge $(y_t)_{t \in \mathbb{Z}}$ heißt absolut summierbar, wenn die Folge der Partialsummen

$$s_t = |y_{-t}| + |y_{-t+1}| + \dots + |y_0| + |y_1| + \dots + |y_t|$$

für $t \rightarrow \infty$ gegen einen endlichen Grenzwert konvergiert, d.h. wenn

$$\lim_{t \rightarrow \infty} \sum_{u=-t}^t |y_u| < \infty.$$

Asymptotische Verteilungen

Das arithmetische Mittel eines Normalprozesses

Ein stochastischer Prozess (Y_t) heißt **Normalprozess** oder **Gauß-Prozess**, wenn für jede endliche Auswahl von Zeitpunkten $t_1, t_2, \dots, t_k$ die Zufallsvariablen $Y_{t_1}, Y_{t_2}, \dots, Y_{t_k}$ multivariat normalverteilt sind.

Theorem

Für stationäre Gauß-Prozesse (Y_t) mit absolut summierbarer Autokovarianzfunktion (γ_τ) gilt

$$\sqrt{N} (\bar{Y}_N - \mu) \xrightarrow{d} N \left(0, \gamma_0 + 2 \sum_{\tau=1}^{\infty} \gamma_\tau \right).$$

Asymptotische Verteilungen

Das arithmetische Mittel eines allgemeinen linearen Prozesses

Ein allgemeiner linearer Prozess ist ein stochastischer Prozess

$$Y_t := \sum_{u=-\infty}^{\infty} a_u \varepsilon_{t-u}$$

mit White-Noise $(\varepsilon_t)_{t \in \mathbb{Z}}$, mit Erwartungswert μ und Varianz σ^2 sowie (a_u) eine absolut summierbare Folge von Gewichten.

Theorem

(Y_t) sei ein allgemeiner linearer Prozess

$$Y_t := \mu + \sum_{u=-\infty}^{\infty} c_u \varepsilon_{t-u}$$

mit $\mathbb{E}[\varepsilon_t] = 0$, $\text{Var}[\varepsilon_t] = \sigma^2 < \infty$ und (c_u) absolut summierbar. Dann gilt

$$\sqrt{N} (\bar{Y}_N - \mu) \xrightarrow{d} N \left(0, \gamma_0 + 2 \sum_{\tau=1}^{\infty} \gamma_{\tau} \right).$$

Asymptotische Verteilungen

Die empirische Autokovarianzfunktion

Theorem

(Y_t) sei ein stationärer Normalprozess mit abs. summierbarer Autokovarianzfunktion (γ_τ) .

alternativ:

(Y_t) sei ein allgemeiner linearer Prozess $Y_t - \mu = \sum_{u=-\infty}^{\infty} \beta_u \varepsilon_{t-u}$ mit $\mathbb{E}[\varepsilon_t^4] < \infty$ und (β_u) absolut summierbar. Die Größen

$$\sqrt{N} (\hat{\gamma}_N(0) - \gamma_0), \dots, \sqrt{N} (\hat{\gamma}_N(k) - \gamma_k)$$

sind für jeweils festes k gemeinsam asymptotisch normalverteilt mit Erwartungswertvektor $\mathbf{0}$ und Kovarianzen

$$N \operatorname{Cov}[\hat{\gamma}_N(\tau), \hat{\gamma}_N(\kappa)] = \sum_{u=-\infty}^{\infty} (\gamma_u \gamma_{u+\tau+\kappa} + \gamma_{u-\kappa} \gamma_{u+\tau}).$$

Insbesondere gilt $\lim_{N \rightarrow \infty} \mathbb{E}[\hat{\gamma}_N(\tau)] = \gamma_\tau$ und $\lim_{N \rightarrow \infty} \operatorname{Var}[\hat{\gamma}_N(\tau)] = 0$.

Asymptotische Verteilungen

Die empirische Autokorrelationsfunktion

Theorem

(Y_t) sei ein stationärer Normalprozess mit abs. summierbarer Autokovarianzfunktion (γ_τ) .

alternativ:

(Y_t) sei ein allgemeiner linearer Prozess $Y_t - \mu = \sum_{u=-\infty}^{\infty} \beta_u \varepsilon_{t-u}$ mit $\mathbb{E}[\varepsilon_t^2] < \infty$ und $\sum |u| \beta_u^2 < \infty$. Für die empirische Autokorrelationsfunktion

$\hat{\rho}_N(\tau)$ gilt

$$\lim_{N \rightarrow \infty} \mathbb{E}[\hat{\rho}_N(\tau)] = \rho_\tau, \quad \lim_{N \rightarrow \infty} \text{Var}[\hat{\rho}_N(\tau)] = 0,$$

und die Größen

$$\sqrt{N}(\hat{\rho}_N(1) - \rho_1), \dots, \sqrt{N}(\hat{\rho}_N(k) - \rho_k)$$

sind für jeweils festes k gemeinsam asymptotisch normalverteilt.

Asymptotische Verteilungen

White-Noise

Für einen White-Noise-Prozess (ε_t) mit $\mathbb{E}[\varepsilon_t] = \mu$ und $\text{Var}[\varepsilon_t] = \sigma^2$ und für Lags $\tau, \kappa \geq 1$ gilt

$$\lim_{N \rightarrow \infty} \mathbb{E}[\hat{\rho}_N(\tau)] = 0 \quad \text{und} \quad \lim_{N \rightarrow \infty} \text{Cov}[\hat{\rho}_N(\tau), \hat{\rho}_N(\kappa)] = 0.$$

Approximation für großes N :

$$\mathbb{E}[\hat{\rho}_N(\tau)] \approx -\frac{1}{N} \quad \text{und} \quad \text{Cov}[\hat{\rho}_N(\tau), \hat{\rho}_N(\kappa)] \approx \begin{cases} \frac{1}{N} & \tau = \kappa \\ 0 & \tau \neq \kappa \end{cases}$$

und

$$\hat{\rho}_N(\tau) \approx N \left(-\frac{1}{N}, \frac{1}{N} \right)$$

Asymptotische Verteilungen

White-Noise – Schwankungsintervall für die empirische Autokorrelation

Für hinreichend großes N erhalten wir das approximative Schwankungsintervall für die empirische Autokorrelation $\hat{\rho}_N(\tau)$ mit $\tau > 0$

$$-\frac{1}{N} \pm 1.96 \frac{1}{\sqrt{N}}.$$

Dabei ist 1.96 rund das 0.975-Quantil der Standardnormalverteilung.

- Realisationen von $\hat{\rho}_N(\tau)$ außerhalb des Schwankungsintervalls können in lediglich 5% der Fälle vorkommen, wenn der datengenerierende Prozess White-Noise ist (Irrtumswahrscheinlichkeit 5%).
- Für das Beispiel auf der nächsten Folie entspricht das $0.05 * 25 = 1.25$ Fällen. Daher ist diese empirische ACF, trotz dem auffallenden Wert für $\tau = 7$, mit einem White-Noise-Prozess vereinbar.

Asymptotische Verteilungen

White-Noise – Schwankungsintervall für die empirische Autokorrelation

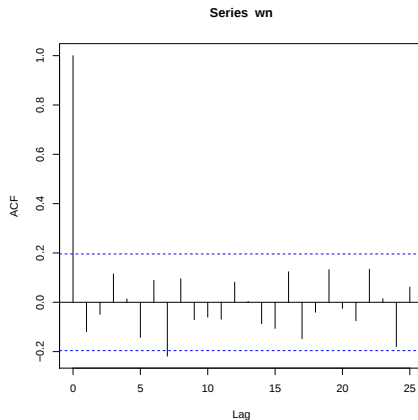


Abb. 1.9: Die ACF für die Realisation eines White-Noise-Prozesses

Asymptotische Verteilungen

White-Noise – Schwankungsintervall für die empirische Autokorrelation

```
wn <- rnorm(100,0,1) #100 Zufallszahlen von der Verteilung  
N(0,1)  
wn <- ts(wn, start=1, frequency=1)  
acf(wn, lag.max=25)
```

Ergodizität

(Y_t) sei ein stationärer Prozess mit dem Erwartungswert μ und der Autokovarianzfunktion (γ_τ) . (Y_t) heißt

- mittelwertergodisch, wenn

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[(\bar{Y}_N - \mu)^2 \right] = 0;$$

- kovarianzergodisch, wenn für alle τ gilt

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{N} \sum_{t=1}^N (Y_t - \mu) (Y_{t+\tau} - \mu) - \gamma_\tau \right)^2 \right] = 0.$$

Theorem

$(Y_t)_{t \in \mathbb{Z}}$ sei ein stationärer Prozess mit Erwartungswert μ und Autokovarianzfunktion (γ_τ) . Ist (γ_τ) absolut summierbar, so ist (Y_t) mittelwertergodisch.

Das bedeutet, dass μ durch $\bar{Y}_N$ konsistent geschätzt werden kann, wenn die Abhängigkeit von weit auseinanderliegenden Zeitpunkten genügend klein ist.

Theorem

Sei (Y_t) ein stationärer Normalprozess mit absolut summierbarer Autokovarianzfunktion (γ_τ) .

alternativ:

Sei (Y_t) ein allgemeiner linearer Prozess $Y_t - \mu = \sum_{u=-\infty}^{\infty} \beta_u \varepsilon_{t-u}$ mit $\mathbb{E}[\varepsilon_t^4] < \infty$ und $\sum_u |\beta_u| < \infty$. Dann ist (Y_t) kovarianzergodisch.

Der Satz besagt, unter welchen Bedingungen γ_τ durch $\hat{\gamma}_N(\tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} (Y_t - \bar{Y}_N)(Y_{t+\tau} - \bar{Y}_N)$ konsistent geschätzt werden kann.

- Schlittgen und Streitberg (1999), Zeitreihenanalyse: Kapitel 5.1, 5.2, 5.3, 5.4
- Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 1, 2.1

- Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 13

Literaturhinweise

Wiederholung der statistischen Grundlagen

- Ihre BA-Vorlesungen zur Statistik
- Schlittgen und Streitberg (1999), Zeitreihenanalyse: Anhänge C.1, C.2, C.3, C.6

Kapitel 2:
Das klassische Komponentenmodell

Inhaltsübersicht

Das klassische Komponentenmodell

- Die Trendkomponente

 - Polynomialer Trend

 - Die glatte Komponente

- Die Saisonkomponente

Das klassische Komponentenmodell

Überprüfung der Stationaritätsannahme

Stationäre Prozesse weisen eine approximative Konstanz der interessierenden empirischen Momente auf.

Stationarität	Nicht-Stationarität
approx. konstantes Niveau approx. konstante Streuung approx. konstante Abhängigkeitsstruktur	Trend die Varianz steigt mit dem Niveau an Saison

Die Stationaritätsannahme kann überprüft werden

- durch visuelle Inspektion anhand des Zeitreihenplots
- rechnerisch – die Zeitreihe in gleich lange Segmente zerlegen und die approx. Konstanz der empirischen Maßzahlen für die Segmente überprüfen

Das klassische Komponentenmodell

Wir beschreiben die Daten (y_t) mit dem Modell:

$$Y_t := T_t + S_t + \varepsilon_t.$$

- Die Trendkomponente T_t erfasst die langfristige systematische Veränderung des mittleren Niveaus der Zeitreihe. Diese kann als polynomialer Trend oder als glatte Komponente modelliert werden.
- Die Saisonkomponente S_t erfasst jahreszeitlich bedingte Schwankungen, die sich relativ unverändert jedes Jahr wiederholen.
- Der Rest ε_t enthält die nicht zu erklärenden Einflüsse oder Störungen.

Anmerkung: Das klassische Komponentenmodell ist kein eindeutig definiertes Modell sondern ein Konzept. Daher große Flexibilität in der Modellierung der einzelnen Komponenten.

Ziel: Bestimmung einer trend- und saisonbereinigten Reihe, die mit Hilfe eines stationären und ergodischen Prozesses modelliert werden kann.

Das klassische Komponentenmodell

Fallbeispiel Anzahlen von Luchspelzen – Überprüfung der Stationaritätsannahme

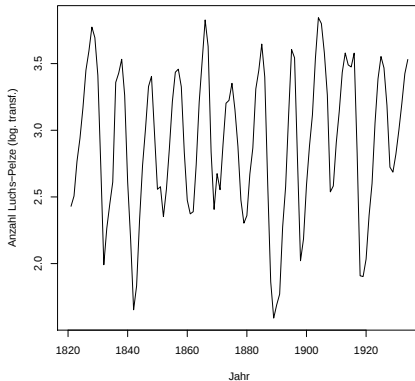


Abb. 2.1: Die Zeitreihe *Luchs*

Das klassische Komponentenmodell

Fallbeispiel Anzahlen von Luchspelzen – Überprüfung der Stationaritätsannahme

	Segment-Mittelwerte	Segment-Standardabweichungen
[1,]	2.888403	0.5478154
[2,]	2.823095	0.5815392
[3,]	2.999493	0.5456087

Das klassische Komponentenmodell

Fallbeispiel Anzahlen von Luchspelzen – Überprüfung der Stationaritätsannahme

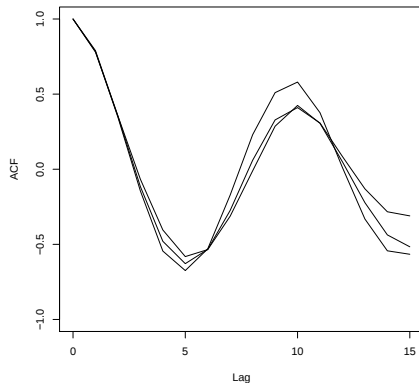


Abb. 2.2: Die ACFs für 3 Segmente der Zeitreihe *Luchs*

Das klassische Komponentenmodell

Fallbeispiel Anzahlen von Luchspelzen – Code

```
y <- scan("C:/ZRA/Luchs.dat")  
#Die Varianz steigt mit dem Niveau an.  
#Durch die logarithmische Transformation wird die Varianz  
stabilisiert.  
y <- ts(log10(y), start=1821, frequency=1)  
tsp(y)  
ts.plot(y, xlab="Jahr", ylab="Anzahl Luchs-Pelze  
(log.transf.)")  
  
source("C:/ZRA/tsutil.r")  
statcheck(y,3)
```

Das klassische Komponentenmodell

Fallbeispiel Verkehrsunfälle – Überprüfung der Stationaritätsannahme

Die monatlichen Anzahlen der Verkehrsunfälle mit Personenschäden in Deutschland, Jan. 1979 – Dez. 2017.

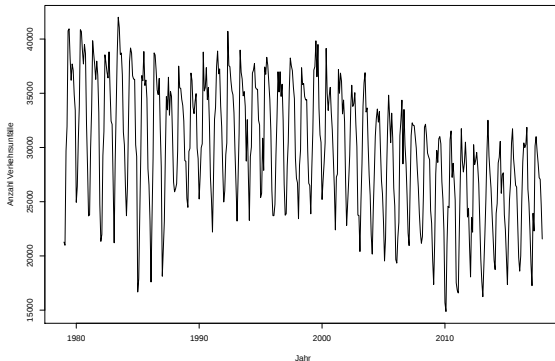


Abb. 2.3: Die Zeitreihe *Unfall*

Das klassische Komponentenmodell

Fallbeispiel Verkehrsunfälle – Überprüfung der Stationaritätsannahme

	Segment-Mittelwerte	Segment-Standardabweichungen
[1,]	33470.22	6094.125
[2,]	31941.89	6659.370
[3,]	31028.96	4676.784
[4,]	32141.57	4789.325
[5,]	32059.17	4979.130
[6,]	32197.15	4458.553
[7,]	29166.13	4832.825
[8,]	26986.78	4535.740
[9,]	24538.76	4835.763
[10,]	25223.02	4197.525

Das klassische Komponentenmodell

Fallbeispiel Verkehrsunfälle – Überprüfung der Stationaritätsannahme

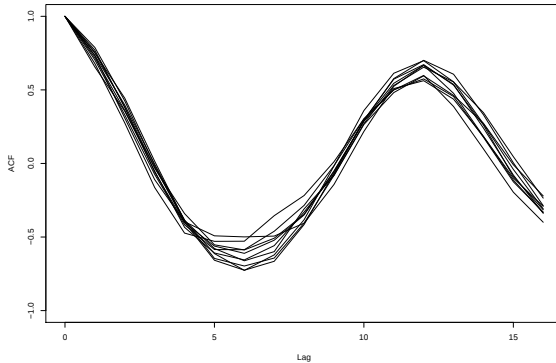


Abb. 2.4: Die ACFs für 10 Segmente der Zeitreihe *Unfall*

Das klassische Komponentenmodell

Fallbeispiel Verkehrsunfälle – Code

```
Unfall <- scan("C:/ZRA/Unfall.dat")
Unfall.ts <- ts(Unfall, start=c(1979,1), frequency=12)
ts.plot(Unfall.ts, xlab="Jahr", ylab="Anzahl Verkehrsunfälle")
source("C:/ZRA/tsutil.r")
statcheck(Unfall.ts, 10)
```

Kapitel 2:

Das klassische Komponentenmodell

2.1 Die Trendkomponente

Die Trendkomponente

Polynomialer Trend

Wir beschreiben die Daten (y_t) mit dem Modell:

$$Y_t := T_t + \varepsilon_t \quad (2.1)$$

mit dem Trend

$$T_t := \beta_0 + \beta_1 t$$

oder allgemein, mit einem polynomialen Trend der Ordnung m

$$T_t := \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_m t^m.$$

- Das Modell (2.1) ist ein lineares Modell.
- Schätzung der Modellparameter mit Hilfe der Methode der kleinsten Quadrate. Dazu noch mehr in Kapitel 'Regressionsmodelle für Zeitreihen'.

Die Trendkomponente

Fallbeispiel Verkehrsunfälle

$$Y_t := \beta_0 + \beta_1 t + \varepsilon_t$$

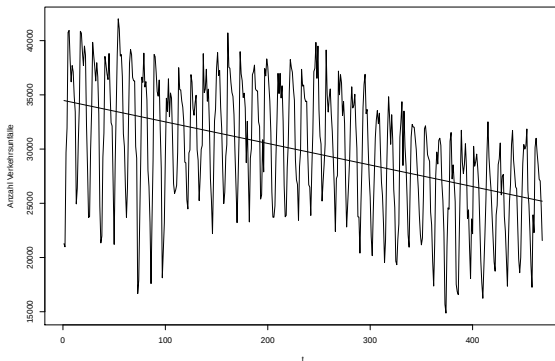


Abb. 2.5: Die Zeitreihe *Unfall* mit polynomialem Trend 1. Ordnung

Die Trendkomponente

Fallbeispiel Verkehrsunfälle

$$Y_t := \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \varepsilon_t$$

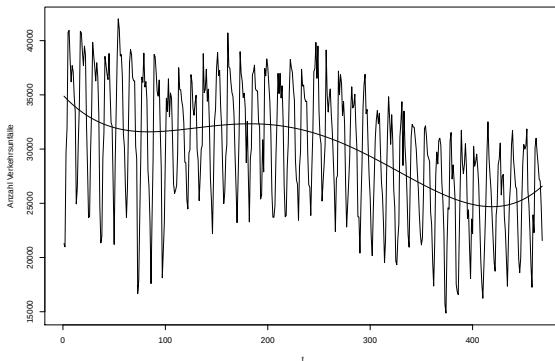


Abb. 2.6: Die Zeitreihe *Unfall* mit polynomialem Trend 4. Ordnung

Die Trendkomponente

Fallbeispiel Verkehrsunfälle – Code

```
n <- length(Unfall)
t <- c(1:n)
out1 <- lm(Unfall ~1 + t)
plot(Unfall, xlab="t", ylab="Anzahl Verkehrsunfälle", type="l")
lines(t, out1$fitted.values)

t2 <- t^2
t3 <- t^3
t4 <- t^4
out2 <- lm(Unfall ~1 + t + t2 + t3 + t4)
plot(Unfall, xlab="t", ylab="Anzahl Verkehrsunfälle", type="l")
lines(t, out2$fitted.values)
```

Die Trendkomponente

Die glatte Komponente

Die glatte Komponente kann mit dem Verfahren der gleitenden Durchschnitte, wie folgt, bestimmt werden:

1. Die Zeitreihe in gleich lange Segmente zerlegen
2. Das arithmetische Mittel für jedes Segment bestimmen
3. Die arithmetischen Mittel den jeweiligen zeitlichen Mittelpunkten der Segmente zuordnen
4. Wir erhalten eine Zeitreihe aus (gleitenden) arithmetischen Mitteln. Diese ist die glatte Komponente der Ausgangsreihe.

Die Trendkomponente

Fallbeispiel Verkehrsunfälle

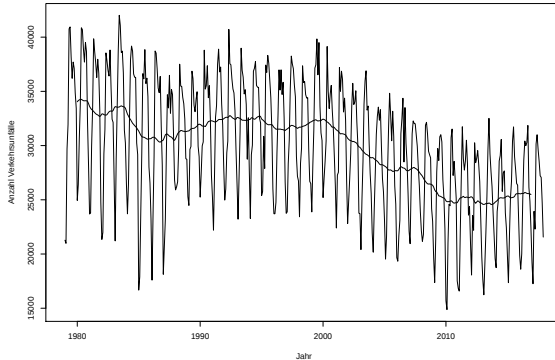


Abb. 2.7: Die Zeitreihe *Unfall* mit gleitenden 24er Durchschnitten

Die Trendkomponente

Fallbeispiel Verkehrsunfälle – Code

```
source("C:/ZRA/tsutil.r")  
glatt <- runmean(Unfall.ts, 24)  
ts.plot(Unfall.ts, xlab="Jahr", ylab="Anzahl Verkehrsunfälle")  
lines(glatt)
```

Kapitel 2:

Das klassische Komponentenmodell

2.2 Die Saisonkomponente

Die Saisonkomponente

Sei s die Saisonperiode ($s = 4$ für Quartalsdaten, $s = 12$ für Monatsdaten).
Wir betrachten die Dummy-Variablen $S_{i,t}$, $i = 1, 2, \dots, s$

$$S_{i,t} := \begin{cases} 1 & t \text{ findet in der Jahresperiode } i \text{ statt} \\ 0 & \text{sonst} \end{cases}.$$

Für eine zentrierte Reihe (y_t) , die keinen Trend aufweist, verwenden wir das Modell:

$$Y_t := S_t + \varepsilon_t. \quad (2.2)$$

Alternativ für eine Reihe (y_t) mit polynomialem Trend:

$$Y_t := \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_p t^p + S_t + \varepsilon_t. \quad (2.3)$$

Die Saisonkomponente

Modellierung der Saisonkomponente (1)

$$S_t := \beta_1 S_{1,t} + \beta_2 S_{2,t} + \dots + \beta_{s-1} S_{s-1,t}$$

Die Regressoren für Quartalsdaten ($s = 4$) sind:

t	$S_{1,t}$	$S_{2,t}$	$S_{3,t}$
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0
5	1	0	0
6	0	1	0
7	0	0	1
8	0	0	0
9	1	0	0
10	0	1	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Die Saisonkomponente

Modellierung der Saisonkomponente (1)

Wir schätzen das Modell (2.2) bzw. (2.3) mit der Methode der kleinsten Quadrate.

Die geschätzte Saison für Quartalsdaten ist

$$\hat{S}_t = \hat{\beta}_1 S_{1,t} + \hat{\beta}_2 S_{2,t} + \hat{\beta}_3 S_{3,t}.$$

Die saisonbereinigte Reihe ist

$$y_{t,b} = y_t - \hat{S}_t + \left(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 \right) S_{4,t}.$$

Die Saisonkomponente

Modellierung der Saisonkomponente (2)

Die Regressoren für Quartalsdaten sind:

t	$S_{1,t}^*$	$S_{2,t}^*$	$S_{3,t}^*$
1	1	0	0
2	0	1	0
3	0	0	1
4	-1	-1	-1
5	1	0	0
6	0	1	0
7	0	0	1
8	-1	-1	-1
9	1	0	0
10	0	1	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Die Saisonkomponente

Modellierung der Saisonkomponente (2)

Wir schätzen das Modell (2.2) bzw. (2.3) mit der Methode der kleinsten Quadrate.

Die geschätzte Saison für Quartalsdaten ist

$$\hat{S}_t = \hat{\beta}_1 S_{1,t}^* + \hat{\beta}_2 S_{2,t}^* + \hat{\beta}_3 S_{3,t}^*.$$

Die saisonbereinigte Reihe ist

$$y_{t,b} = y_t - \hat{S}_t.$$

Die Saisonkomponente

Fallbeispiel Steuern

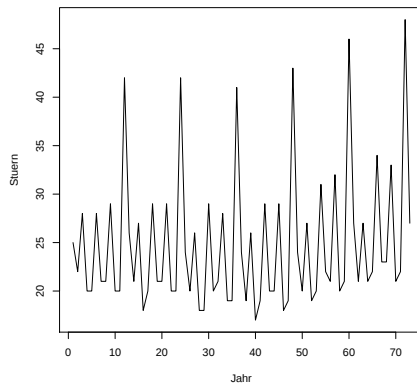


Abb. 2.8: Die Zeitreihe *Steuern*

Die Saisonkomponente

Fallbeispiel Steuern

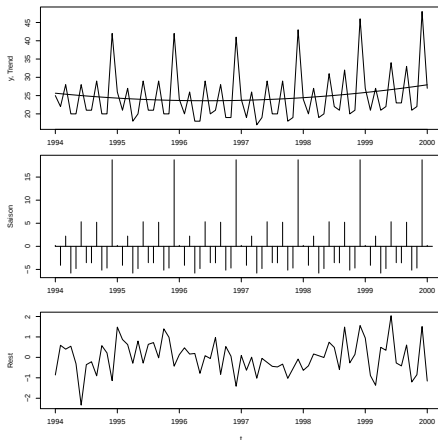


Abb. 2.9: Zerlegung der Zeitreihe *Steuern* mit polynomialem Trend 2. Ordnung und Saison

Die Saisonkomponente

Fallbeispiel Steuern – Code

```
steuern <- scan("C:/ZRA/steuern.dat")  
y <- ts(steuern, start=c(1994,1), frequency=12)  
ts.plot(steuern, xlab="Jahr", ylab="Steuern")  
source("C:/ZRA/tsutil.r")  
simpledecomp(y,trend=2, saison=12)
```

Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 2.2, 2.3, 2.4

Kapitel 3: Lineare Modelle

Inhaltsübersicht

Lineare Modelle

Lineare Filter

- MA-Prozesse

- AR-Prozesse

- ARMA-Prozesse

Invertierbare lineare Filter

- Stationäre AR-Prozesse

- Invertierbare MA-Prozesse

- Stationäre und invertierbare ARMA-Prozesse

Anpassung linearer Modelle

- Anpassung autoregressiver Modelle

- Anpassung von Moving-Average-Modellen

- Anpassung von Autoregressiven-Moving-Average-Modellen

Saisonale ARIMA-Modelle

Differenzen- vs. trendstationäre lineare Modelle

Kapitel 3: Lineare Modelle

3.1 Lineare Filter

Definition (Backshift-Operator B)

Die Transformation B mit

$$BY_t = Y_{t-1}.$$

bezeichnen wir als **Backshift-Operator**.

Die Transformation B verschiebt den stochastischen Prozess um eine Zeiteinheit. Hintereinanderausführungen von B schreiben wir als Potenzen; allgemein ist

$$B^2 Y_t = B(BY_t) = BY_{t-1} = Y_{t-2},$$

und damit rekursiv

$$B^p Y_t = Y_{t-p}.$$

Definition (Linearer Filter)

Eine lineare Transformation eines stochastischen Prozesses (X_t) in einen anderen (Y_t) gemäß

$$Y_t = \sum_{u=-r}^s a_u X_{t-u}$$

wird als **linearer Filter** bezeichnet. Ein Filter wird durch seine Gewichte

$$(a_u)_{u=-r, \dots, s}$$

eindeutig definiert. Alternativ kann er als gewichtete Summe von Shiftoperatoren dargestellt werden

$$A(B) = \sum_{u=-r}^s a_u B^u.$$

Dann schreiben wir für Y_t

$$Y_t = A(B) X_t.$$

Lineare Filter

Wir bezeichnen die Anwendung eines Filters auf einen stochastischen Prozess (X_t) als Filtration des Prozesses, (X_t) als Input und den gefilterten Prozess (Y_t) als Output des Filters.

Definition (*kausaler* Filter)

Ein linearer Filter mit der Gewichtsfolge (a_u) heißt *kausal*, wenn $a_u = 0$ für $u < 0$ ist, d.h. wenn

$$A(B) = \sum_{u=0}^{\infty} a_u B^u.$$

Definition (*absolut summierbarer* Filter)

Ein Filter heißt *absolut summierbar*, wenn seine Gewichtsfolge (a_u) absolut summierbar ist.

Theorem (Unendliche lineare Filter)

Sei $(X_t)_{t \in \mathbb{Z}}$ ein stochastischer Prozess mit der Eigenschaft

$$\mathbb{E}[X_t^2] \leq K \quad \text{für alle } t \in \mathbb{Z},$$

wobei K eine endliche Konstante ist. Sei (a_u) ein absolut summierbarer Filter. Dann existiert ein stochastischer Prozess $(Y_t)_{t \in \mathbb{Z}}$, so dass für alle $t \in \mathbb{Z}$

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(Y_t - \sum_{u=-n}^n a_u X_{t-u} \right)^2 \right] = 0, \quad (3.1)$$

$$\mathbb{E}[Y_t^2] < \infty \quad \text{und} \quad \mathbb{E}[Y_t] = \lim_{n \rightarrow \infty} \sum_{u=-n}^n a_u \mathbb{E}[X_{t-u}].$$

Weiter benutzen wir für den Limes (3.1) die Notation:

$$Y_t = \sum_{u=-\infty}^{\infty} a_u X_{t-u}.$$

Unendliche lineare Filter mit absolut summierbarer Folge von Gewichten können unbekümmert im Sinne des vorigen Satzes gehandhabt werden.

- Bspw. ist die Vertauschung von Summation und Erwartungswertoperator erlaubt.
- Ist der Input (X_t) stationär, so ist auch der Output (Y_t) stationär.

Lineare Filter

Allgemeine lineare Prozesse

In Kapitel 1 haben wir den allgemeinen linearen Prozess

$$Y_t = \sum_{u=-\infty}^{\infty} a_u \varepsilon_{t-u}$$

mit $(\varepsilon_t)_{t \in \mathbb{Z}}$ White-Noise und (a_u) absolut summierbar eingeführt. Der allgemeine lineare Prozess wird durch Filtration des White-Noise-Prozesses $(\varepsilon_t)_{t \in \mathbb{Z}}$

$$Y_t = A(B) \varepsilon_t$$

mit dem unendlichen linearen Filter

$$A(B) = \sum_{u=-\infty}^{\infty} a_u B^u$$

erzeugt.

Definition (Moving-Average-Prozess der Ordnung q , MA[q])

Ein stochastischer Prozess (Y_t) ,

$$Y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q}$$

heißt **Moving-Average-Prozess der Ordnung q** , kurz MA[q]-Prozess. Dabei ist (ε_t) ein White-Noise-Prozess. Im Fall $q = \infty$ sprechen wir von einem unendlichen Moving-Average-Prozess (MA[∞]).

Ein MA-Prozess entsteht durch Filtration des White-Noise-Prozesses (ε_t)

$$Y_t = \beta(B) \varepsilon_t$$

mit dem kausalen linearen Filter

$$\beta(B) = 1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_q B^q.$$

Lineare Filter

MA-Prozesse – Theoretische Momente

Sei $\mathbb{E}[\varepsilon_t] = \mu_\varepsilon$, $\text{Var}[\varepsilon_t] = \sigma_\varepsilon^2$ und $\beta_0 = -1$.

MA[q]-Prozesse (Y_t) sind *stationäre* Prozesse mit den Momenten:

1. $\mu_Y = \mathbb{E}[Y_t] = \mu_\varepsilon \sum_{u=0}^q (-\beta_u)$
2. $\gamma_Y(\tau) = \sum_{u=0}^q \sum_{v=0}^q \beta_u \beta_v \gamma_\varepsilon(\tau + u - v) =$
$$\begin{cases} 0, & \tau > q \\ \sigma_\varepsilon^2 \sum_{u=0}^{q-\tau} \beta_u \beta_{u+\tau}, & 0 \leq \tau \leq q \\ \gamma_Y(-\tau), & \tau < 0 \end{cases}$$

Lineare Filter

MA-Prozesse – Theoretische Momente

$$3. \sigma_Y^2 = \text{Var}[Y_t] = \sigma_\varepsilon^2 \sum_{u=0}^q \beta_u^2$$

$$4. \rho_Y(\tau) = \frac{\gamma_Y(\tau)}{\gamma_Y(0)} = \begin{cases} 0, & \tau > q \\ \sum_{u=0}^{q-\tau} \beta_u \beta_{u+\tau} / \sum_{u=0}^q \beta_u^2, & 0 < \tau \leq q \\ 1, & \tau = 0 \\ \rho_Y(-\tau), & \tau < 0 \end{cases}$$

Die ACF eines MA[q]-Prozesses verschwindet für Lags $\tau > q$.

Definition (Autoregressiver Prozess der Ordnung p , $\text{AR}[p]$)

Ein stochastischer Prozess (Y_t) ,

$$Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t$$

heißt **Autoregressiver Prozess der Ordnung p** , kurz $\text{AR}[p]$ -Prozess. Dabei ist (ε_t) ein White-Noise-Prozess. Im Fall $p = \infty$ heißt (Y_t) ein unendlicher AR-Prozess ($\text{AR}[\infty]$).

Alternative Darstellung:

$$\varepsilon_t = Y_t - \alpha_1 Y_{t-1} - \dots - \alpha_p Y_{t-p}.$$

Der White-Noise-Prozess (ε_t) ist der Output des kausalen linearen Filters

$$\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p$$

mit dem Input dem AR-Prozess (Y_t)

$$\varepsilon_t = \alpha(B) Y_t.$$

Definition (Autoregressiver-Moving-Average-Prozess der Ordnung p, q , ARMA[p, q])

Ein stochastischer Prozess (Y_t) ,

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q}$$

mit (ε_t) ein White-Noise-Prozess heißt **Autoregressiver Moving-Average-Prozess der Ordnung p, q** , kurz ARMA[p, q]-Prozess.

Alternative Darstellung:

$$Y_t - \alpha_1 Y_{t-1} - \alpha_2 Y_{t-2} - \dots - \alpha_p Y_{t-p} = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q},$$

oder kurz:

$$\alpha(B) Y_t = \beta(B) \varepsilon_t$$

mit den kausalen linearen Filtern

$$\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p$$

$$\beta(B) = 1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_q B^q$$

Kapitel 3: Lineare Modelle

3.2 Invertierbare lineare Filter

Problem

Gegeben sei der lineare Filter $A(B)$,

$$A(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p. \quad (3.2)$$

Bestimmen Sie einen zu $A(B)$ inversen linearen Filter $A^{-1}(B)$ mit Gewichten (c_u) mit den Eigenschaften

- $A^{-1}(B)$ ist kausal

und

- $A^{-1}(B)$ ist absolut summierbar, falls $A^{-1}(B)$ ein unendlicher Filter ist, so dass für jeden stochastischen Prozess (X_t) gilt

$$Y_t = \sum_{u=0}^p a_u X_{t-u} \Leftrightarrow X_t = \sum_u c_u Y_{t-u}.$$

Invertierbare lineare Filter

Die Funktion

$$A(z) = 1 - a_1 z - a_2 z^2 - \dots - a_p z^p, \quad z \in \mathbb{C}$$

heißt das charakteristische Polynom des Filters

$$A(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p.$$

Jeder lineare Filter $A(B)$ ist eindeutig durch sein charakteristisches Polynom $A(z)$ bestimmt. Daher werden wir das weiter oben beschriebene Problem für $A(z)$ anstatt für $A(B)$ lösen.

D.h. wir suchen ein zu $A(z)$ inverses charakteristisches Polynom

$$A^{-1}(z) = c_0 + c_1 z + c_2 z^2 + \dots$$

mit absolut summierbarer Koeffizientenfolge (c_u) , falls $A^{-1}(z)$ ein unendliches Polynom ist.

Theorem

Für den linearen Filter

$$A(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p$$

existiert dann und nur dann ein kausaler, absolut summierbarer, inverser Filter

$$A^{-1}(B) = c_0 + c_1 B + c_2 B^2 + \dots,$$

wenn alle Nullstellen des charakteristischen Polynoms

$$A(z) = 1 - a_1 z - a_2 z^2 - \dots - a_p z^p$$

außerhalb des Einheitskreises liegen.

Invertierbare lineare Filter

Eindeutige Beschreibung mit Hilfe der ACF

Hinweis: Ein stochastischer Prozess Y_t

$$Y_t = A(B) X_t$$

mit invertierbarem linearem Filter

$$A(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p$$

und kausalem, absolut summierbarem, inversem Filter $A^{-1}(B)$ ist eindeutig durch seine Autokorrelationsfunktion definiert.

Theorem

Ein $AR[p]$ -Prozess $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t$ kann dann und nur dann als $MA[\infty]$ -Prozess

$$Y_t = \sum_{u=0}^{\infty} \beta_u \varepsilon_{t-u}$$

mit absolut summierbarer Koeffizientenfolge (β_u) dargestellt werden, wenn alle Nullstellen des charakteristischen Polynoms

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0$$

außerhalb des Einheitskreises liegen. Speziell ist (Y_t) dann stationär.

Stationäre AR-Prozesse

Stationaritätsbedingung

- Stationaritätsbedingung für AR[1]-Prozesse $Y_t = \alpha Y_{t-1} + \varepsilon_t$

$$|\alpha| < 1$$

- Stationaritätsbedingungen für AR[2]-Prozesse $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t$

$$\begin{cases} \alpha_1 + \alpha_2 < 1 \\ \alpha_2 - \alpha_1 < 1 \\ \alpha_2 > -1 \end{cases}$$

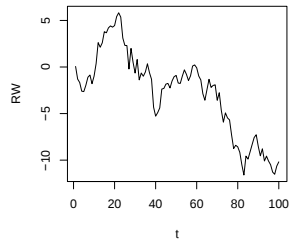
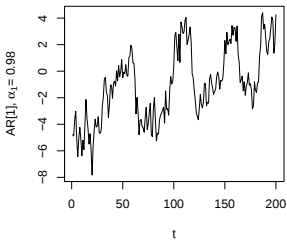
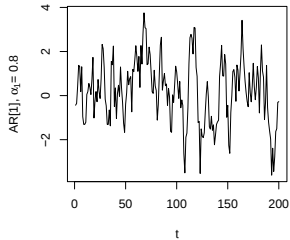
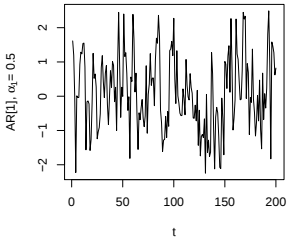


Abb. 3.1: Realisationen von AR-Prozessen und einem Random-Walk

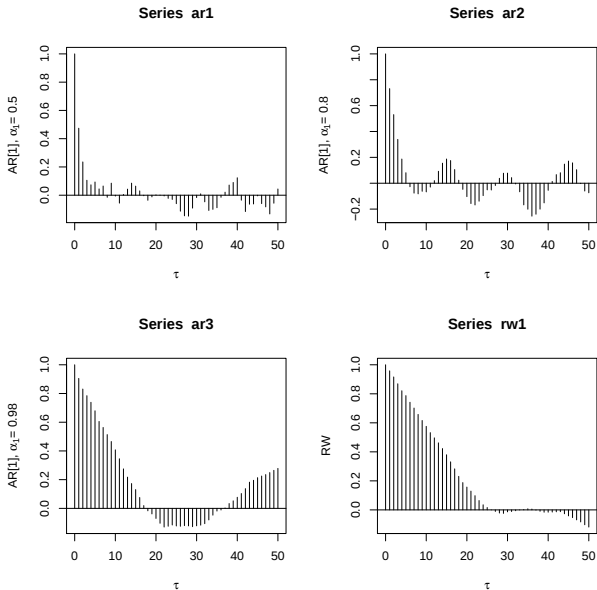


Abb. 3.2: Die empirischen ACF

Stationäre AR-Prozesse

Eindeutige Beschreibung mit Hilfe der ACF

Stationäre AR-Prozesse sind eindeutig durch ihre ACF definiert. Berechnung der

ACF für einen stationären AR-Prozess Y_t mit $\mathbb{E}[Y_t] = 0$:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} \dots + \alpha_p Y_{t-p} + \varepsilon_t$$

$$\Downarrow$$

$$Y_t Y_{t-\tau} = \alpha_1 Y_{t-1} Y_{t-\tau} + \alpha_2 Y_{t-2} Y_{t-\tau} \dots + \alpha_p Y_{t-p} Y_{t-\tau} + \varepsilon_t Y_{t-\tau}$$

$$\Downarrow$$

$$\mathbb{E}[Y_t Y_{t-\tau}] = \alpha_1 \mathbb{E}[Y_{t-1} Y_{t-\tau}] + \alpha_2 \mathbb{E}[Y_{t-2} Y_{t-\tau}] \dots + \alpha_p \mathbb{E}[Y_{t-p} Y_{t-\tau}] + \mathbb{E}[\varepsilon_t Y_{t-\tau}]$$

$$\Downarrow$$

$$\gamma(\tau) = \alpha_1 \gamma(\tau - 1) + \alpha_2 \gamma(\tau - 2) \dots + \alpha_p \gamma(\tau - p)$$

$$\Downarrow$$

$$\rho(\tau) = \alpha_1 \rho(\tau - 1) + \alpha_2 \rho(\tau - 2) + \dots + \alpha_p \rho(\tau - p)$$

Stationäre AR-Prozesse

Eindeutige Beschreibung mit Hilfe der ACF

Das System der so genannten Yule-Walker-Gleichungen:

$$\rho(\tau) = \alpha_1 \rho(\tau - 1) + \alpha_2 \rho(\tau - 2) + \dots + \alpha_p \rho(\tau - p), \quad 1 \leq \tau \leq p \quad (3.3)$$

bzw. in der Matrixschreibweise:

$$\begin{pmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix}.$$

Dieses System enthält p Gleichungen in p Unbekannten ρ_τ , $1 \leq \tau \leq p$.

Es ist eindeutig lösbar für stationäre AR-Prozesse.

Die Yule-Walker-Gleichungen

Das empirische Analogon zu den Yule-Walker-Gleichungen:

$$\begin{pmatrix} 1 & r_1 & \dots & r_{p-1} \\ r_1 & 1 & \dots & r_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & \dots & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{pmatrix}$$

Hier sind $\hat{\alpha}_i$, $1 \leq i \leq p$, die Unbekannten. Die Werte r_τ stellen die empirischen Autokorrelationen dar. Durch Lösen des Gleichungssystems erhalten wir die Parameterschätzungen $\hat{\alpha}_i$. Diese sind die so genannten Yule-Walker-Schätzer des AR-Modells.

Die Levinson-Durbin-Rekursion

Schätzung eines AR[3]-Modells

Gegeben seien die Yule-Walker-Gleichungen für den AR[2]-Prozess:

$$\begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$$

mit bekannten Lösungen $\hat{\alpha}_1$ und $\hat{\alpha}_2$. Wir berechnen die Lösungen $\hat{\theta}_1$, $\hat{\theta}_2$ und $\hat{\theta}_3$ der Yule-Walker-Gleichungen für den AR[3]-Prozess:

$$\begin{pmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}.$$

Die Levinson-Durbin-Rekursion

Schätzung eines AR[3]-Modells

Wir erhalten die Lösungen:

$$\left\{ \begin{array}{l} \hat{\theta}_1 = \hat{\alpha}_1 - \hat{\alpha}_2 \hat{\theta}_3 \\ \hat{\theta}_2 = \hat{\alpha}_2 - \hat{\alpha}_1 \hat{\theta}_3 \\ \hat{\theta}_3 = \frac{r_3 - r_2 \hat{\alpha}_1 - r_1 \hat{\alpha}_2}{1 - r_2 \hat{\alpha}_2 - r_1 \hat{\alpha}_1} \end{array} \right.$$

Theorem (Die Levinson-Durbin-Rekursion)

Gegeben seien

- *die Startwerte: $\hat{\alpha}_1(1) = r_1$, $\hat{Q}(1) = 1 - r_1^2$*
- *die Lösungen $\hat{\alpha}_1(p-1)$, $\hat{\alpha}_2(p-1)$, $\dots$, $\hat{\alpha}_{p-1}(p-1)$ für die Yule-Walker-Gleichungen der Ordnung $p-1$*
- *der Wert $\hat{Q}(p-1) = 1 - \hat{\alpha}_1(p-1)r_1 - \dots - \hat{\alpha}_{p-1}(p-1)r_{p-1}$.*

Dann ergeben sich die Lösungen der Yule-Walker-Gleichungen der Ordnung p

$$\hat{\alpha}_p(p) = \Delta(p) / \hat{Q}(p-1)$$

$$\hat{\alpha}_i(p) = \hat{\alpha}_i(p-1) - \hat{\alpha}_p(p) \hat{\alpha}_{p-i}(p-1), \quad i = 1, \dots, p-1$$

wobei $\hat{Q}(p) = \hat{Q}(p-1) (1 - \hat{\alpha}_p^2(p))$ und

$$\Delta(p) = r_p - (\hat{\alpha}_1(p-1)r_{p-1} - \dots - \hat{\alpha}_{p-1}(p-1)r_1).$$

Die Levinson-Durbin-Rekursion

Bei der Anwendung der Levinson-Durbin-Rekursion für die Berechnung der Yule-Walker-Schätzer $\hat{\alpha}_i$, $1 \leq i \leq p$ eines $\text{AR}[p]$ -Modells werden in einem Durchgang auch

- die Yule-Walker-Schätzer für alle AR-Modelle niedrigerer Ordnung 1 bis $p - 1$

und

- die empirische partielle Autokorrelationsfunktion ($\hat{\pi}_\tau$) des $\text{AR}[p]$ -Modells berechnet.

Die empirischen partiellen Autokorrelationen $\hat{\pi}_p$ sind gerade die Koeffizienten $\hat{\alpha}_p(p)$, die sich bei der Levinson-Durbin-Rekursion ergeben.

Die partielle Autokorrelationsfunktion

Die partielle Autokorrelation zweier Zeitreihenvariablen Y_t und $Y_{t-\tau}$ ist die Autokorrelation von Y_t und $Y_{t-\tau}$ unter Konstanthaltung / Ausschaltung der dazwischen liegenden Zufallsvariablen $Y_{t-\tau+1}, Y_{t-\tau+2}, \dots, Y_{t-1}$. Die

Konstanthaltung / Ausschaltung der dazwischen liegenden Variablen erfolgt, indem man die mit Hilfe der auszuschaltenden Variablen $Y_{t-\tau+1}, Y_{t-\tau+2}, \dots, Y_{t-1}$ erstellten Prognosen $\hat{Y}_t$ und $\hat{Y}_{t-\tau}$ jeweils abzieht. Die partielle Autokorrelation ist dann die Korrelation zwischen $Y_t - \hat{Y}_t$ und $Y_{t-\tau} - \hat{Y}_{t-\tau}$.

Die partielle Autokorrelationsfunktion

Definition (Die partielle Autokorrelationsfunktion (PACF))

Sei (Y_t) ein stationärer stochastischer Prozess. Die **partielle Autokorrelationsfunktion** (π_τ) , kurz PACF, ist die Folge der Korrelationen von $Y_t - \hat{Y}_t$ und $Y_{t-\tau} - \hat{Y}_{t-\tau}$ wobei $\hat{Y}_t$ und $\hat{Y}_{t-\tau}$ die optimalen linearen Prognosen von Y_t und $Y_{t-\tau}$ durch $Y_{t-\tau+1}, Y_{t-\tau+2}, \dots, Y_{t-1}$ sind. Wir definieren $\pi_0 = 1$ und $\pi_1 = \rho_1$.

Die partielle Autokorrelationsfunktion

Wir betrachten einen AR[1]-Prozess (Y_t)

$$\begin{array}{ccccccccc} Y_{t-2} & \xrightarrow{\alpha} & Y_{t-1} & \xrightarrow{\alpha} & Y_t & \xrightarrow{\alpha} & Y_{t+1} & \xrightarrow{\alpha} & Y_{t+2} & \xrightarrow{\alpha} & \dots \\ \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\ \varepsilon_{t-2} & & \varepsilon_{t-1} & & \varepsilon_t & & \varepsilon_{t+1} & & \varepsilon_{t+2} & & \end{array}$$

Die PACF ist

$$\pi_\tau = \begin{cases} 1, & \tau = 0 \\ \rho_1, & \tau = 1 \\ 0, & \tau > 1 \end{cases}$$

Die partielle Autokorrelationsfunktion

Theorem

(Y_t) ist dann und nur dann ein $AR[p]$ -Prozess $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t$ mit $\alpha_p \neq 0$, wenn $\pi_p \neq 0$ und $\pi_\tau = 0$ für $\tau > p$ gilt.

Die PACF eines $AR[p]$ -Prozesses verschwindet für Lags $\tau > p$. Sie hat die folgende approximative Verteilung für hinreichend großes N :

$$\hat{\pi}_N(\tau) \approx N\left(0, \frac{1}{N}\right) \text{ für } \tau > p$$

Die partielle Autokorrelationsfunktion

Schwankungsintervall

Sei (Y_t) ein $\text{AR}[p]$ -Prozess. Für hinreichend großes N erhalten wir das approximative Schwankungsintervall für die empirische partielle Autokorrelation $\hat{\pi}_N(\tau)$ mit $\tau > p$

$$\pm 1.96 \frac{1}{\sqrt{N}}.$$

Dabei ist 1.96 rund das 0.975-Quantil der Standardnormalverteilung.

- Realisationen von $\hat{\pi}_N(\tau)$ außerhalb des Schwankungsintervalls können in lediglich 5% der Fälle vorkommen, wenn der datengenerierende Prozess (Y_t) ein $\text{AR}[p]$ -Prozess ist (Irrtumswahrscheinlichkeit 5%).

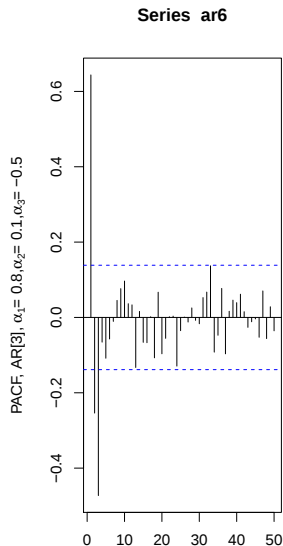
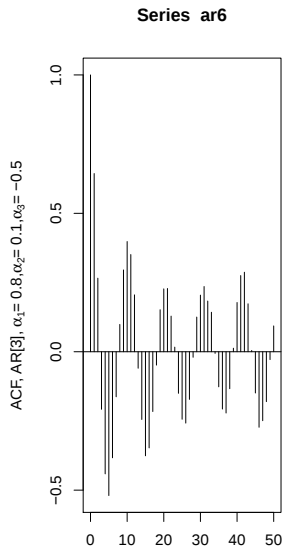
Unter *Spezifikation* oder *Identifikation* eines AR-Modells verstehen wir die Bestimmung der Ordnung p des Modells. 2 Ansätze:

1. Der klassische Box-Jenkins-Ansatz besteht in der Interpretation der empirischen ACF und PACF.
2. Wir passen AR-Modelle verschiedener Ordnungen an die Daten an und wählen das geeignete Modell mit Hilfe eines Informationskriteriums aus.

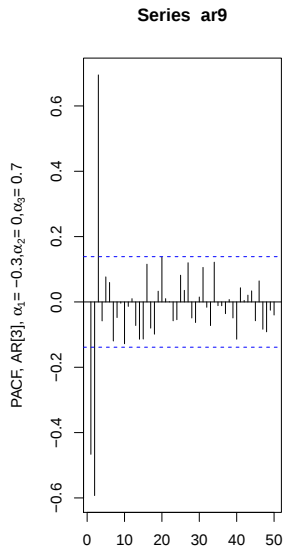
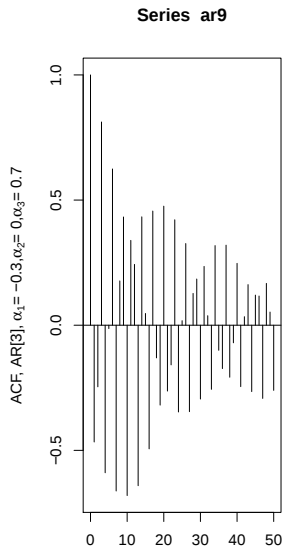
- Verhaltensmuster der ACF: Die ACF $\rho(\tau)$ eines stationären AR[p]-Prozesses klingt exponentiell ab und konvergiert gegen Null für $\tau \rightarrow \infty$.
- Verhaltensmuster der PACF: Die PACF $\pi(\tau)$ eines stationären AR[p]-Prozesses verschwindet für Lags $\tau > p$.

Für genügend lange Reihen können diese Verhaltensmuster anhand der empirischen Kennfunktionen überprüft werden.

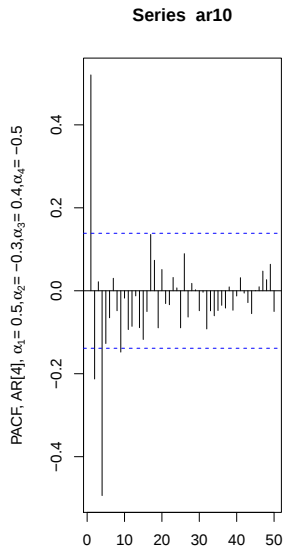
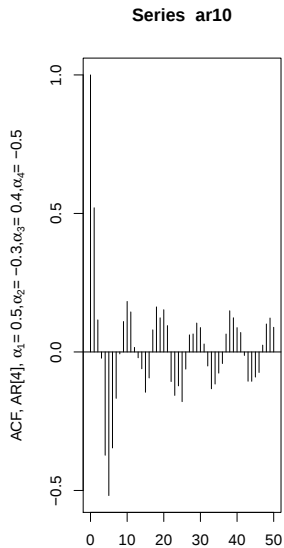
Simulationen



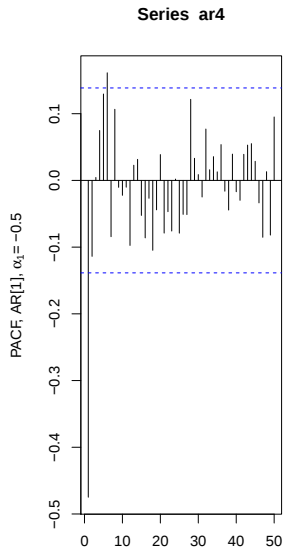
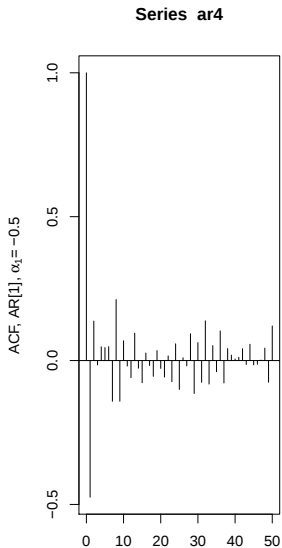
Simulationen



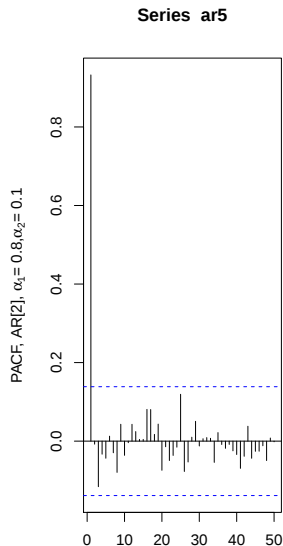
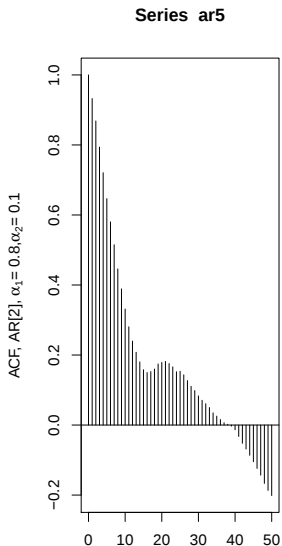
Simulationen



Simulationen



Simulationen



Definition

Ein $\text{MA}[q]$ -Prozess $Y_t = \beta(B) \varepsilon_t$ heißt **invertierbar**, wenn alle Nullstellen des charakteristischen Polynoms

$$\beta(z) = 1 - \beta_1 z - \beta_2 z^2 - \dots - \beta_q z^q$$

außerhalb des Einheitskreises liegen.

Invertierbare $\text{MA}[q]$ -Prozesse können in der Form $\varepsilon_t = \beta^{-1}(B) Y_t$ geschrieben werden, wobei

$$\beta^{-1}(B) = c_0 + c_1 B + c_2 B^2 + \dots$$

ein kausaler linearer Filter mit absolut summierbarer Koeffizientenfolge ist. Dies entspricht einer Darstellung des invertierbaren MA-Prozesses als unendlicher AR-Prozess.

Eindeutige Beschreibung mit Hilfe der ACF

Invertierbare MA-Prozesse sind eindeutig durch ihre ACF definiert. Die ACF ist

$$\rho_Y(\tau) = \frac{\gamma_Y(\tau)}{\gamma_Y(0)} = \begin{cases} 0, & \tau > q \\ \sum_{u=0}^{q-\tau} \beta_u \beta_{u+\tau} / \sum_{u=0}^q \beta_u^2, & 0 < \tau \leq q \\ 1, & \tau = 0 \\ \rho_Y(-\tau), & \tau < 0 \end{cases}$$

Die ACF eines MA[q]-Prozesses verschwindet für Lags $\tau > q$.

Eindeutige Beschreibung mit Hilfe der ACF

Das Gleichungssystem

$$\rho(\tau) = \sum_{u=0}^{q-\tau} \beta_u \beta_{u+\tau} \bigg/ \sum_{u=0}^q \beta_u^2, \quad 0 < \tau \leq q,$$

d.h.:

$$\left\{ \begin{array}{l} \rho_1 = \frac{\beta_0 \beta_1 + \beta_1 \beta_2 + \beta_2 \beta_3 + \dots + \beta_{q-1} \beta_q}{\beta_0^2 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2} \\ \rho_2 = \frac{\beta_0 \beta_2 + \beta_1 \beta_3 + \beta_2 \beta_4 + \dots + \beta_{q-2} \beta_q}{\beta_0^2 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2} \\ \vdots \\ \rho_{q-1} = \frac{\beta_0 \beta_{q-1} + \beta_1 \beta_q}{\beta_0^2 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2} \\ \rho_q = \frac{\beta_0 \beta_q}{\beta_0^2 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2} \end{array} \right.$$

hat q Gleichungen in q Unbekannten. Es ist eindeutig lösbar für invertierbare MA-Prozesse.

Eindeutige Beschreibung mit Hilfe der ACF

Für die empirische Autokorrelation $\hat{\rho}_N(\tau)$ von einem $\text{MA}[q]$ -Prozess, $\tau > q$, gilt für hinreichend großes N :

$$\mathbb{E}[\hat{\rho}_N(\tau)] \doteq 0 \quad \text{und} \quad \text{Var}[\hat{\rho}_N(\tau)] \approx \frac{1}{N} (1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_q^2)$$

und

$$\hat{\rho}_N(\tau) \approx N \left(0, \frac{1}{N} (1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_q^2) \right).$$

Eindeutige Beschreibung mit Hilfe der ACF

Schwankungsintervall

Für hinreichend großes N erhalten wir das approximative Schwankungsintervall für die empirische Autokorrelation $\hat{\rho}_N(\tau)$ von einem $\text{MA}[q]$ -Prozess, $\tau > q$

$$\pm 1.96 \sqrt{(1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_q^2) / N}.$$

Dabei ist 1.96 rund das 0.975-Quantil der Standardnormalverteilung.

In praktischen Anwendungen verwenden wir die so genannten Bartlett-Grenzen

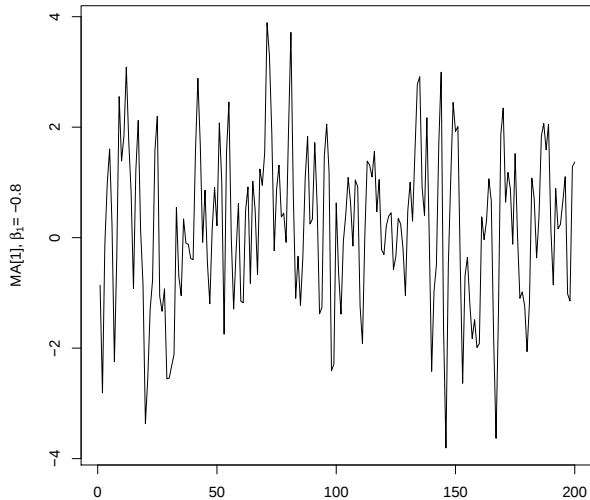
$$\pm 1.96 \sqrt{(1 + 2\hat{\rho}_1^2 + 2\hat{\rho}_2^2 + \dots + 2\hat{\rho}_q^2) / N}.$$

- Realisationen von $\hat{\rho}_N(\tau)$ außerhalb des Schwankungsintervalls können in lediglich 5% der Fälle vorkommen, wenn der datengenerierende Prozess ein $\text{MA}[q]$ -Prozess ist (Irrtumswahrscheinlichkeit 5%).

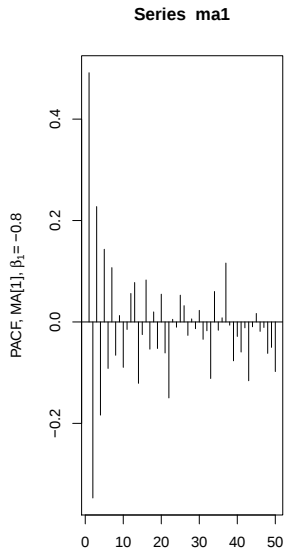
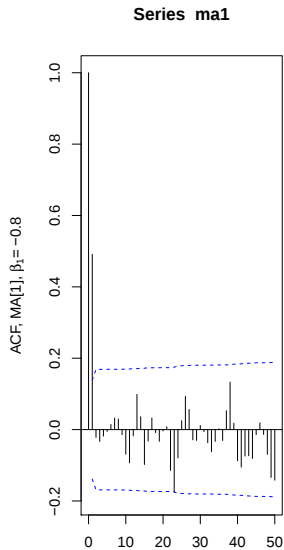
- Verhaltensmuster der ACF: Die ACF $\rho(\tau)$ eines $\text{MA}[q]$ -Prozesses verschwindet für Lags $\tau > q$.
- Verhaltensmuster der PACF: Die PACF $\pi(\tau)$ eines $\text{MA}[q]$ -Prozesses klingt für $\tau > q$ exponentiell ab und konvergiert gegen Null für $\tau \rightarrow \infty$.

Für genügend lange Reihen können diese Verhaltensmuster anhand der empirischen Kennfunktionen überprüft werden.

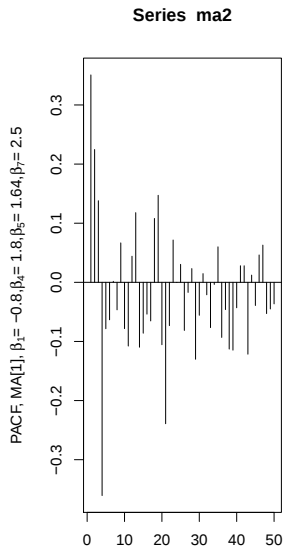
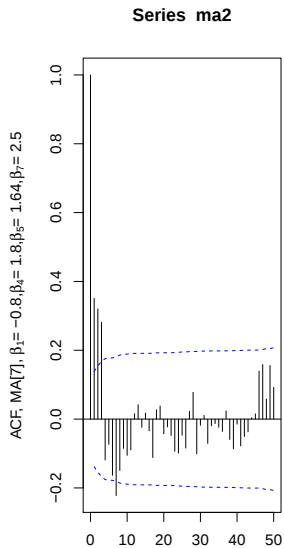
Simulationen



Simulationen



Simulationen



Theorem

1. Ein ARMA-Prozess $\alpha(B) Y_t = \beta(B) \varepsilon_t$ kann als $MA[\infty]$ -Prozess dargestellt werden

$$Y_t = \alpha(B)^{-1} \beta(B) \varepsilon_t,$$

wenn alle Nullstellen des charakteristischen Polynoms des AR-Teils

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0$$

außerhalb des Einheitskreises liegen.

Speziell ist (Y_t) dann ein stationärer ARMA-Prozess.

Theorem

2. Ein ARMA-Prozess $\alpha(B) Y_t = \beta(B) \varepsilon_t$ kann als $AR[\infty]$ -Prozess dargestellt werden

$$\beta(B)^{-1} \alpha(B) Y_t = \varepsilon_t$$

wenn alle Nullstellen des charakteristischen Polynoms des MA-Teils

$$1 - \beta_1 z - \beta_2 z^2 - \dots - \beta_q z^q = 0$$

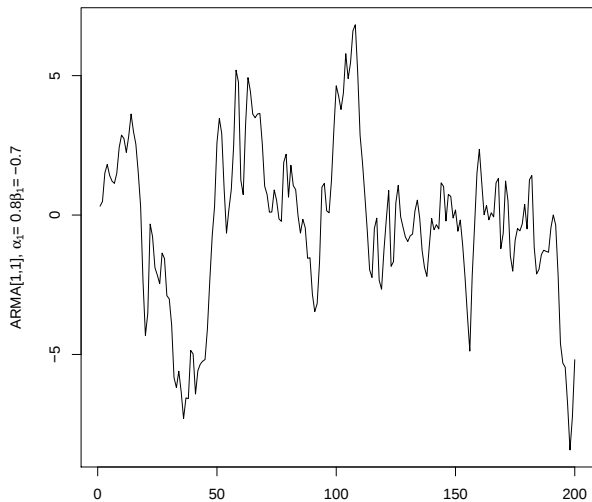
außerhalb des Einheitskreises liegen.

Wir sprechen von einem invertierbaren ARMA-Prozess (Y_t) .

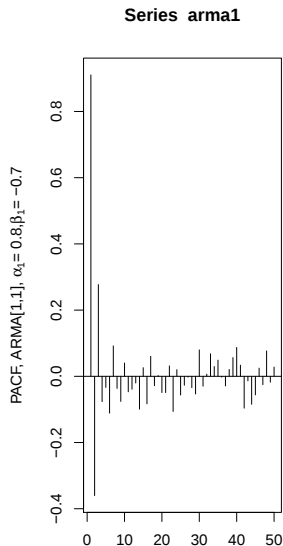
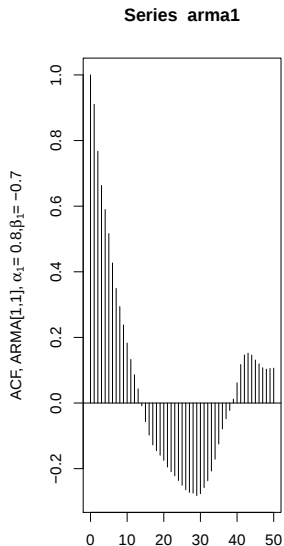
- Verhaltensmuster der ACF: Die ACF $\rho(\tau)$ eines stationären ARMA[p, q]-Prozesses klingt exponentiell ab und konvergiert gegen Null für $\tau \rightarrow \infty$.
- Verhaltensmuster der PACF: Die PACF $\pi(\tau)$ eines stationären ARMA[p, q]-Prozesses klingt exponentiell ab und konvergiert gegen Null für $\tau \rightarrow \infty$.

Die Spezifikation gemischter ARMA[p, q]-Modelle mit Hilfe der ACF und PACF ist eine schwierige Aufgabe.

Simulationen



Simulationen



Ergodizität von ARMA-Prozessen

Wiederholung aus Kapitel 1

Theorem

$(Y_t)_{t \in \mathbb{Z}}$ sei ein stationärer Prozess mit Erwartungswert μ und Autokovarianzfunktion (γ_τ) . Ist (γ_τ) absolut summierbar, so ist (Y_t) mittelwertergodisch.

Das bedeutet, dass μ durch $\bar{Y}_N$ konsistent geschätzt werden kann, wenn die Abhängigkeit von weit auseinanderliegenden Zeitpunkten genügend klein ist.

Ergodizität von ARMA-Prozessen

Wiederholung aus Kapitel 1

Theorem

Sei (Y_t) ein stationärer Normalprozess mit absolut summierbarer Autokovarianzfunktion (γ_τ) .

alternativ:

Sei (Y_t) ein allgemeiner linearer Prozess $Y_t - \mu = \sum_{u=-\infty}^{\infty} \beta_u \varepsilon_{t-u}$ mit $\mathbb{E}[\varepsilon_t^4] < \infty$ und $\sum |\beta_u| < \infty$. Dann ist (Y_t) kovarianzergodisch.

Der Satz besagt, unter welchen Bedingungen γ_τ durch $\hat{\gamma}_N(\tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} (Y_t - \bar{Y}_N) (Y_{t+\tau} - \bar{Y}_N)$ konsistent geschätzt werden kann.

- Stationäre ARMA[p, q]-Prozesse sind mittelwertergodisch.
- Stationäre ARMA[p, q]-Prozesse mit $\mathbb{E}[\varepsilon_t^4] < \infty$ sind kovarianzergodisch.

Hinweis: Diese Aussagen gelten auch für MA-Prozesse und für stationäre AR-Prozesse.

Schlittgen und Streitberg (1999), Zeitreihenanalyse: Kapitel 2.3.2, 2.3.3, 2.3.4, 2.3.5, 4.1.2, 6.2.3

Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 3

Literaturhinweise

Wiederholung von mathematischen Grundlagen

Schlittgen und Streitberg (1999), Zeitreihenanalyse: Anhänge B.1, B.2

Kapitel 3: Lineare Modelle

3.3 Anpassung linearer Modelle

Vorgehensweise in der Anpassung von linearen Zeitreihenmodellen

1. **Datentransformation** im Rahmen des klassischen Komponentenmodells: Erzeugen einer trend- und saisonbereinigten Reihe, die mit Hilfe eines stationären und ergodischen Prozesses modelliert werden kann.
2. **Modellspezifikation**: Bestimmung der Ordnung (p, q) des Modells. Zwei Ansätze: auf der Grundlage der ACF und PACF oder auf der Grundlage von Informationskriterien.
3. **Modellschätzung**: Schätzung der Modellparameter bei gegebener Ordnung
4. **Modelldiagnose**: Beurteilung der Anpassungsgüte
5. **Modellanwendung**: Interpretation, Prognose

In Kapitel 1 haben wir den Trend einer Zeitreihe mit Hilfe von

- polynomialen Trendfunktionen

oder

- der glatten Komponente

bestimmt und bereinigt. Alternativ kann die Trendbereinigung mit Hilfe des

Differenzenfilters durchgeführt werden.

Definition (Differenzenfilter Δ)

Wir nennen den linearen Filter Δ ,

$$\Delta Y_t = (1 - B) Y_t = Y_t - Y_{t-1}, \quad t = 2, 3, \dots, N.$$

Differenzenfilter 1. Ordnung.

Hintereinanderausführungen von Δ schreiben wir als Potenzen

$$\Delta^2 Y_t = (1 - B)^2 Y_t = (1 - B) ((1 - B) Y_t) = \Delta (\Delta Y_t)$$

und allgemein für $d \in \mathbb{N}$

$$\Delta^d Y_t = \Delta^{d-1} (\Delta Y_t).$$

Definition (Integrierter Prozess)

Ein Prozess (Y_t) heißt **integriert** vom Grade d , kurz vom Typ $I(d)$, wenn $\Delta^d Y_t$ stationär ist.

Definition (ARIMA $[p, d, q]$ -Prozess)

Ein stochastischer Prozess (Y_t) heißt

Autoregressiver-Integrierter-Moving-Average-Prozess der Ordnung $[p, d, q]$, kurz ARIMA $[p, d, q]$ -Prozess, wenn der nach d -maligem Differenzieren entstandene Prozess $\Delta^d Y_t$ einem stationären ARMA $[p, q]$ -Prozess genügt.

Schätzung von AR-Modellen

Das zentrierte AR-Modell:

$$Y_t - \bar{y} = \alpha_1 (Y_{t-1} - \bar{y}) + \dots + \alpha_p (Y_{t-p} - \bar{y}) + \varepsilon_t \quad (3.4)$$

Das geschätzte Modell:

$$\widehat{Y_t - \bar{y}} = \hat{\alpha}_1 (Y_{t-1} - \bar{y}) + \dots + \hat{\alpha}_p (Y_{t-p} - \bar{y})$$

Die Residuen:

$$\hat{\varepsilon}_t = Y_t - \bar{y} - \widehat{Y_t - \bar{y}}$$

Wir minimieren die Residuenquadratsumme:

$$\sum_{t=p+1}^N \hat{\varepsilon}_t^2 \stackrel{!}{=} \min_{\hat{\alpha}_1, \dots, \hat{\alpha}_p}$$

Schätzung von AR-Modellen

Die Normalgleichungen:

$$\hat{\alpha}_1 \sum_{t=p+1}^N (y_{t-1} - \bar{y})^2 + \dots + \hat{\alpha}_p \sum_{t=p+1}^N (y_{t-p} - \bar{y})(y_{t-1} - \bar{y}) = \sum_{t=p+1}^N (y_t - \bar{y})(y_{t-1} - \bar{y})$$

$$\hat{\alpha}_1 \sum_{t=p+1}^N (y_{t-1} - \bar{y})(y_{t-2} - \bar{y}) + \dots + \hat{\alpha}_p \sum_{t=p+1}^N (y_{t-p} - \bar{y})(y_{t-2} - \bar{y}) = \sum_{t=p+1}^N (y_t - \bar{y})(y_{t-2} - \bar{y})$$

$\vdots$

$$\hat{\alpha}_1 \sum_{t=p+1}^N (y_{t-1} - \bar{y})(y_{t-p} - \bar{y}) + \dots + \hat{\alpha}_p \sum_{t=p+1}^N (y_{t-p} - \bar{y})^2 = \sum_{t=p+1}^N (y_t - \bar{y})(y_{t-p} - \bar{y})$$

Schätzung von AR-Modellen

Drei Schätzverfahren durch Lösung der Normalgleichungen:

1. CLS (Conditional Least Squares)

Die Summationen in den Normalgleichungen fangen bei $t = 1$ anstatt $t = p + 1$ an. Für die Terme $y_0, y_{-1}, \dots, y_{-(p-1)}$ wird der Erwartungswert Null eingesetzt.

2. ULS (Unconditional Least Squares)

Das AR-Modell wird zuerst mit einem alternativen Verfahren geschätzt und es werden die Prognosen in die Vergangenheit $\hat{y}_0, \hat{y}_{-1}, \dots, \hat{y}_{-(p-1)}$ der ersten p Reihenwerte berechnet. Dann werden die Normalgleichungen gelöst, wobei die Summationen bei $t = 1$ anstatt $t = p + 1$ anfangen. Für die Terme $y_0, y_{-1}, \dots, y_{-(p-1)}$ werden ihre Prognosen eingesetzt.

3. Yule-Walker

Die Normalgleichungen werden mit Hilfe der empirischen Autokorrelationen r_τ geschätzt. Wir erhalten das empirische Analogon der Yule-Walker-Gleichungen:

$$\begin{pmatrix} 1 & r_1 & \dots & r_{p-1} \\ r_1 & 1 & \dots & r_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & \dots & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{pmatrix}$$

Durch Lösen des Gleichungssystems erhalten wir die Parameterschätzungen $\hat{\alpha}_i$, $1 \leq i \leq p$.

Schätzung von AR-Modellen

Modellannahmen im Rahmen des linearen Regressionsmodells (3.4)

- (i) $\mathbb{E}[\varepsilon_t] = 0, \quad t = 1, 2, \dots, N.$
- (ii) $\mathbb{E}[\varepsilon_s \varepsilon_t] = \begin{cases} \sigma_\varepsilon^2, & \text{falls } s = t \\ 0, & \text{falls } s \neq t \end{cases}$
- (iii) Die Regressoren $Y_{t-1} - \bar{y}, \dots, Y_{t-p} - \bar{y}$ sind nichtstochastisch und die Matrix $(Y_{t-1} - \bar{y}, \dots, Y_{t-p} - \bar{y})$ hat den Rang p .
- (iv) Es gibt über $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p$ und σ^2 keine Vorinformation.

Schätzung von AR-Modellen

Modellannahmen im Rahmen des linearen Regressionsmodells (3.4)

Annahmeverletzung für Zeitreihendaten:

- die Regressoren sind *stochastisch*. Daher besondere (asymptotische) Betrachtung der Schätzungsgüte.

Schätzung von AR-Modellen

Die ML-Methode

4. Maximum Likelihood (ML)

Die beobachtete Zeitreihe $(y_t) = (y_1, y_2, \dots, y_N)$ ist eine Realisation des zugrundeliegenden AR-Prozesses (Y_t) .

Die multivariate Wahrscheinlichkeitsdichtefunktion $f((y_t))$ fasst die Wahrscheinlichkeiten zusammen, mit welchen sich beliebige Reihen (y_t) oder Reihenabschnitte $(y_{i_1}, y_{i_2}, \dots, y_{i_k})$, $1 \leq i_1, i_2, \dots, i_k \leq N$ realisieren können. Sei θ der unbekannte Parametervektor des AR-Prozesses (Y_t) ,

$$\theta = (\sigma^2, \alpha_1, \dots, \alpha_p).$$

Für jede Ausprägung θ_i , $i = 1, 2, \dots$ von θ betrachten wir die multivariate Wahrscheinlichkeitsdichtefunktion $f((y_t) | \theta_i)$ bedingt auf den Parameter θ_i .

Schätzung von AR-Modellen

Die ML-Methode

- Wir können $f((y_t) | \theta_i)$ als Funktion des AR-Prozesses (Y_t) ausdrücken. Dann sprechen wir, wie oben, von *einer bedingten Wahrscheinlichkeitsdichtefunktion*.
- Oder als Funktion des Parametervektors θ . Diese ist dann die so genannte *Likelihood-Funktion*

$$L(\theta | (y_t)) := f((y_t) | \theta).$$

Der ML-Schätzer $\hat{\theta}$ ist der Parametervektor, für den die Likelihood-Funktion ihr Maximum annimmt:

$$L(\theta | (y_t)) \stackrel{!}{=} \max_{\theta}$$

Alternativ kann $\hat{\theta}$ durch Maximierung der so genannten *Loglikelihood-Funktion* berechnet werden

$$l(\theta | (y_t)) := \ln(L(\theta | (y_t))) \stackrel{!}{=} \max_{\theta}$$

Dabei gehen wir von einem *Gauß-Prozess* (Y_t) aus.

Schätzung von AR-Modellen

Schätzungsgüte

Theorem

Sei

$$Y_t - \mu = \alpha_1 (Y_{t-1} - \mu) + \dots + \alpha_p (Y_{t-p} - \mu) + \varepsilon_t$$

ein stationärer AR[p]-Prozess mit $\mathbb{E}[\varepsilon_t] = 0$ und $\text{Var}[\varepsilon_t] = \sigma^2$. Dann gilt für die nach einer der vier Methoden berechneten Schätzer $\hat{\alpha}_i$, $i = 1, \dots, p$:

- $\hat{\alpha}_i$ konvergiert fast sicher (mit Wahrscheinlichkeit 1) gegen den wahren Parameterwert α_i wenn $N \rightarrow \infty$

$$\mathbb{P}\left(\lim_{N \rightarrow \infty} \hat{\alpha}_i(N) = \alpha_i\right) = 1$$

Theorem

- die Zufallsgrößen

$$\sqrt{N}(\hat{\alpha}_1 - \alpha_1), \dots, \sqrt{N}(\hat{\alpha}_p - \alpha_p)$$

sind asymptotisch multivariat normalverteilt mit Erwartungswert $\mathbf{0}$ und Kovarianzmatrix $\sigma^2 \sum_p^{-1}$ mit $\sum_p$ die Kovarianzmatrix von p aufeinander folgenden Variablen des Prozesses (Y_t)

$$\sum_p = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ & & \ddots & \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{pmatrix}.$$

Unter *Spezifikation* oder *Identifikation* eines AR-Modells verstehen wir die Bestimmung der Ordnung p des Modells. 2 Ansätze:

1. Der klassische Box-Jenkins-Ansatz besteht in der Interpretation der empirischen ACF und PACF.
2. Wir passen AR-Modelle verschiedener Ordnungen an die Daten an und wählen das geeignete Modell mit Hilfe eines Informationskriteriums. Wir wählen das Modell mit dem *kleinsten* Wert des betrachteten Informationskriteriums aus.

Modellspezifikation

Interpretation der empirischen ACF und PACF

Wir unterscheiden zwischen drei Verhaltensmustern der empirischen Kennfunktionen ACF und PACF:

- die Kennfunktion klingt exponentiell ab (siehe die ACF eines AR-, die PACF eines MA- sowie beide die ACF und die PACF eines gemischten ARMA-Modells). Dieses Verhalten ist dadurch gekennzeichnet, dass die Werte schnell abfallen, jedoch sehr langsam gegen Null konvergieren.
- die Kennfunktion klingt langsam von $+1$ herkommend ab (Hinweis auf eine Zeitreihe mit Trend);
- die Kennfunktion bricht abrupt ab (siehe die ACF eines White-Noise- oder MA-Prozesses und die PACF eines AR-Prozesses).

- AIC (Akaike's Information Criterion):

$$AIC := \ln\left(\frac{RSS}{N}\right) + \frac{2p}{N}$$

- AICC (AIC with correction):

$$AICC := \ln\left(\frac{RSS}{N}\right) + \frac{2p}{N - p - 2}$$

- BIC (Bayesian Information Criterion):

$$BIC := \ln\left(\frac{RSS}{N}\right) + \frac{p \ln N}{N}$$

Modellspezifikation

Überprüfung der Parametersignifikanz

- Asymptotische Verteilung der Parameterschätzer $\hat{\alpha}_i$, $1 \leq i \leq p$ in einem AR[p]-Modell

$$\sqrt{N}(\hat{\alpha}_i - \alpha_i) \xrightarrow{d} N(0, \sigma_i^2).$$

Die Varianz σ_i^2 kann mit Hilfe des Satzes über die Schätzungsgüte bestimmt werden.

- Approximative Verteilung für N hinreichend groß

$$\hat{\alpha}_i \dot{\approx} N(\alpha_i, \sigma_{\hat{\alpha}_i}^2).$$

Modellspezifikation

Überprüfung der Parametersignifikanz

Testhypothesen:

$$H_0 : \alpha_i = 0 \quad \text{gegen} \quad H_1 : \alpha_i \neq 0$$

Teststatistik:

$$z = \frac{\hat{\alpha}_i}{\hat{\sigma}_{\hat{\alpha}_i}} \overset{\cdot}{\approx} N(0, 1) | H_0$$

Testentscheidung:

- Bestimme den kritischen Wert κ_α zu gegebenem $\alpha \in (0, 1)$:

$$\kappa_\alpha = z_{1-\frac{\alpha}{2}}.$$

- Lehne H_0 mit Irrtumswahrscheinlichkeit α ab, falls

$$|z| > \kappa_\alpha.$$

- **Untersuchung der Abhängigkeitsstruktur** anhand der empirischen ACF und PACF der Residuen oder mit Hilfe von statistischen Tests (bspw. der Ljung-Box-Pierce-Test)
- **Überprüfung der Normalverteilungsannahme im Falle von ML-Schätzungen** mit dem QQ-Diagramm oder mit Hilfe von statistischen Tests

Modelldiagnose - Untersuchung der Residuen

Der Ljung-Box-Pierce-Test

Sei $(\tilde{r}_\tau)$ die empirische ACF der Residuen $(\hat{\varepsilon}_t)$.

Testhypothesen:

$H_0 : (\varepsilon_t)$ White-Noise gegen $H_1 : (\varepsilon_t)$ kein White-Noise

Teststatistik:

$$Q = N(N+2) \sum_{\tau=1}^k \frac{\tilde{r}_\tau^2}{N-\tau} \approx \chi_{k-p}^2 | H_0 .$$

Erfahrungsgemäß sollte Q für verschiedene Werte von $k = 6, 12, 18, 24, \dots$ geprüft werden, wobei $k > p$ gelten soll.

Testentscheidung:

- Bestimme den kritischen Wert κ_α zu gegebenem $\alpha \in (0, 1)$:

$$\kappa_\alpha = \chi_{1-\alpha; k-p}.$$

- Lehne H_0 mit Irrtumswahrscheinlichkeit α ab, falls

$$Q > \kappa_\alpha .$$

Schätzung von MA-Modellen

Sei (Y_t) ein $\text{MA}[q]$ -Prozess

$$Y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q}$$

mit (ε_t) White-Noise. Y_t ist eine *nichtlineare* Funktion der unbekannten Parameter $\beta_1, \dots, \beta_q$. Bspw. für das $\text{MA}[1]$ -Modell

$$Y_t = - \sum_{u=1}^{t-1} \beta^u Y_{t-u} + \varepsilon_t - \beta^t \varepsilon_0.$$

Schätzung von MA-Modellen

Schätzung eines MA[1]-Modells

Das MA[1]-Modell:

$$Y_t = - \sum_{u=1}^{t-1} \beta^u Y_{t-u} + \varepsilon_t - \beta^t \varepsilon_0.$$

Das geschätzte MA[1]-Modell:

$$\hat{Y}_t = - \sum_{u=1}^{t-1} \hat{\beta}^u Y_{t-u}$$

Die Residuen:

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t = Y_t + \sum_{u=1}^{t-1} \hat{\beta}^u Y_{t-u}$$

Wir minimieren die Residuenquadratsumme:

$$\sum_t \hat{\varepsilon}_t^2 \stackrel{!}{=} \min_{\hat{\beta}}$$

Schätzung von MA-Modellen

Zwei Schätzverfahren basieren auf der Minimierung der Residuenquadratsumme:

1. CLS (Conditional Least Squares) – Für die Terme $\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-(q-1)}$ wird der Erwartungswert Null eingesetzt.
2. ULS (Unconditional Least Squares) – Das MA-Modell wird zuerst mit dem CLS-Verfahren geschätzt und es werden die Prognosen in die Vergangenheit $\hat{\varepsilon}_0, \hat{\varepsilon}_{-1}, \dots, \hat{\varepsilon}_{-(q-1)}$ berechnet. Dann wird die Residuenquadratsumme minimiert, wobei für die Terme $\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-(q-1)}$ ihre Prognosen eingesetzt werden.

Der CLS- und die ULS-Schätzer sind *nichtlineare* Schätzfunktionen, die mit Hilfe von *numerischen* Optimierungsverfahren berechnet werden. Alternativ:

3. ML (Maximum Likelihood)

Schätzung von ARMA-Modellen

1. CLS (Conditional Least Squares) – Für die Terme $y_0, y_{-1}, \dots, y_{-(p-1)}, \varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-(q-1)}$ wird jeweils 0 eingesetzt.
2. ULS (Unconditional Least Squares) – Das ARMA-Modell wird zuerst mit dem CLS-Verfahren geschätzt und es werden die Prognosen in die Vergangenheit $\hat{y}_0, \hat{y}_{-1}, \dots, \hat{y}_{-(p-1)}, \hat{\varepsilon}_0, \hat{\varepsilon}_{-1}, \dots, \hat{\varepsilon}_{-(q-1)}$ berechnet. Dann wird die Residuenquadratsumme minimiert, wobei für die Terme $y_0, y_{-1}, \dots, y_{-(p-1)}, \varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-(q-1)}$ ihre Prognosen eingesetzt werden.

Der CLS- und die ULS-Schätzer sind *nichtlineare* Schätzfunktion, die mit Hilfe von *numerischen* Optimierungsverfahren berechnet werden. Alternativ:

3. ML (Maximum Likelihood)

Theorem (Asymptotische Eigenschaften der Parameterschätzer eines ARMA-Prozesses)

Sei (Y_t) ein stationärer und invertierbarer ARMA $[p, q]$ -Prozess $\alpha(B)Y_t = \beta(B)\varepsilon_t$ mit $\mathbb{E}[\varepsilon_t] = 0$ und $\text{Var}[\varepsilon_t] = \sigma^2$. Ferner sei unterstellt, dass die charakteristischen Polynome der Filter $\alpha(B)$ und $\beta(B)$ keine gemeinsame Nullstellen haben. Dann gilt für die nach einer der drei Methoden berechneten Schätzer $\hat{\alpha} = (\hat{\alpha}_1 \quad \hat{\alpha}_2 \quad \dots \quad \hat{\alpha}_p)$ und $\hat{\beta} = (\hat{\beta}_1 \quad \hat{\beta}_2 \quad \dots \quad \hat{\beta}_q)$:

- $\hat{\alpha}$ und $\hat{\beta}$ konvergieren fast sicher (mit Wahrscheinlichkeit 1) gegen die wahren Parametervektoren α und β wenn $N \rightarrow \infty$

$$\mathbb{P}\left(\lim_{N \rightarrow \infty} \hat{\alpha}(N) = \alpha\right) = 1 \quad \text{und} \quad \mathbb{P}\left(\lim_{N \rightarrow \infty} \hat{\beta}(N) = \beta\right) = 1$$

Theorem (Asymptotische Eigenschaften der Parameterschätzer eines ARMA-Prozesses)

- die Zufallsgrößen $\sqrt{N}(\hat{\alpha} - \alpha)$, $\sqrt{N}(\hat{\beta} - \beta)$ sind asymptotisch gemeinsam multivariat normalverteilt mit Erwartungswertvektoren $\mathbf{0}$ und Kovarianzmatrix

$$\mathbf{V} = \begin{bmatrix} \phi_{aa} & -\phi_{ab} \\ -\phi'_{ab} & \phi_{bb} \end{bmatrix}^{-1}.$$

Seien (a_t) und (b_t) die Folgen der Gewichte der inversen Filter $\alpha^{-1}(B)$ und $\beta^{-1}(B)$. Die Kovarianzmatrix $\mathbf{V}$ wird mit Hilfe der unendlichen Reihen

$$\phi_{ab}(\tau) = \sum_{t=0}^{\infty} a_t b_{t+\tau}$$

und analog dazu $\phi_{aa}(\tau)$ und $\phi_{bb}(\tau)$ für $\tau \geq 0$ definiert.

Theorem (Asymptotische Eigenschaften der Parameterschätzer eines ARMA-Prozesses)

Die Matrix $\mathbf{V}$ hat die Partitionen:

$$\Phi_{aa} = \begin{pmatrix} \phi_{aa}(0) & \phi_{aa}(1) & \dots & \phi_{aa}(p-1) \\ \phi_{aa}(1) & \phi_{aa}(0) & & \vdots \\ \vdots & & \ddots & \phi_{aa}(1) \\ \phi_{aa}(p-1) & \dots & \phi_{aa}(1) & \phi_{aa}(0) \end{pmatrix}$$

$$\Phi_{bb} = \begin{pmatrix} \phi_{bb}(0) & \phi_{bb}(1) & \dots & \phi_{bb}(q-1) \\ \phi_{bb}(1) & \phi_{bb}(0) & & \vdots \\ \vdots & & \ddots & \phi_{bb}(1) \\ \phi_{bb}(q-1) & \dots & \phi_{bb}(1) & \phi_{bb}(0) \end{pmatrix}$$

Theorem (Asymptotische Eigenschaften der Parameterschätzer eines ARMA-Prozesses)

$$\phi_{ab} = \begin{pmatrix} \phi_{ab}(0) & \phi_{ab}(1) & \dots & \phi_{ab}(q-1) \\ \phi_{ab}(1) & \phi_{ab}(0) & & \phi_{ab}(q-2) \\ \vdots & & \ddots & \vdots \\ \phi_{ab}(p-1) & \phi_{ab}(p-2) & \dots & \phi_{ab}(0) \end{pmatrix}$$

Schlittgen und Streitberg (1999), Zeitreihenanalyse: Kapitel 6.1.1, 6.1.7
Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 3

Kapitel 3: Lineare Modelle

3.4 Saisonale ARIMA-Modelle

Im Rahmen der Datentransformation kann

- der Trend durch das d -fache Bilden von einfachen Differenzen

$$\Delta^d Y_t = (1 - B)^d Y_t$$

- die Saisonkomponente durch das D -fache Bilden von saisonalen Differenzen

$$\Delta_s Y_t = Y_t - Y_{t-s} = Y_t - B^s Y_t = (1 - B^s) Y_t$$

$$\Delta_s^D Y_t = (1 - B^s)^D Y_t$$

eliminiert werden.

Saisonale ARIMA-Modelle

Saisonale Differenzen

Einen Hinweis auf das Vorliegen einer Saisonkomponente liefern die ACF und PACF, wenn sie signifikante Werte für kleine Lags sowie für die Lags um die Saisonperiode s , $2s$,... aufweisen. Gleichzeitig sind die dazwischen liegenden Lags nicht signifikant von Null verschieden.

Saisonale ARIMA-Modelle

Saisonale Differenzen

Oft ist es nötig, beide Trend und Saisonkomponente zu bereinigen

$$(1 - B)^d (1 - B^s)^D Y_t.$$

Wir entscheiden über die Anzahl d und D der Differenzenbildungen mit Hilfe der Methode der *variatischen Differenzen*. Wir wählen das Paar (d, D) mit der kleinsten empirischen Varianz der transformierten Zeitreihe

$(1 - B)^d (1 - B^s)^D y_t$ aus.

Definition (SARIMA $[p, d, q] \times [P, D, Q]_s$ -Prozess)

Ein stochastischer Prozess (Y_t) heißt **saisonaler Autoregressiver-Integrierter-Moving-Average-Prozess der Ordnungen $[p, d, q]$ und $[P, D, Q]_s$** , kurz SARIMA $[p, d, q] \times [P, D, Q]_s$ -Prozess, wenn er der Gleichung

$$\alpha(B) \eta(B^s) (1 - B)^d (1 - B^s)^D Y_t = \beta(B) \theta(B^s) \varepsilon_t$$

mit (ε_t) ein White-Noise-Prozess genügt.

Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Abschnitt *Saisonale ARIMA-Modelle* in Kapitel 3.4 und Abschnitt *Lineare Filterung von Zeitreihen* in Kapitel 2.5

Lineare Modelle

Differenzen- vs. trendstationäre lineare Modelle

Kapitel 3: Lineare Modelle

3.5 Differenzen- vs. trendstationäre lineare Modelle

Differenzen- vs. trendstationäre lineare Modelle

Wir unterscheiden zwischen

- einem *deterministischen* Trend, z.B. dem polynomialen Trend. Ein lineares Modell mit deterministischem Trend bezeichnen wir als *trendstationär*.

und

- einem *stochastischen* Trend, welcher im Rahmen eines instationären linearen Modells der Klasse ARIMA entsteht. Ein lineares Modell mit stochastischem Trend bezeichnen wir als *differenzenstationär*.

Differenzen- vs. trendstationäre lineare Modelle

Sei (ε_t) ein White-Noise-Prozess und (y_t) die zur Verfügung stehende Zeitreihe.
Wir betrachten das lineare Modell

$$Y_t = \theta_0 + \theta_1 t + Z_t \quad \text{mit} \quad (1 - \alpha B) Z_t = \varepsilon_t.$$

1. $|\alpha| < 1$, (Z_t) ist stationär, (Y_t) ist trendstationär.
Die fehlerhafte Behandlung der Zeitreihe (y_t) als Realisation eines differenzenstationären Prozesses hat vertretbare Konsequenzen für die Modellschätzung und für das Testen.
2. $\alpha = 1$, (Z_t) ist nicht stationär, (Y_t) ist (zumindest) differenzenstationär.
Die fehlerhafte Behandlung der Zeitreihe (y_t) als Realisation eines trendstationären und nicht differenzenstationären Prozesses hat gravierende Konsequenzen für die Modellschätzung und für das Testen.
- (3.) $|\alpha| > 1$, ökonomisch nicht relevant.

Differenzen- vs. trendstationäre lineare Modelle

Der DF-Test

Der Einheitswurzeltest (Unit-Root-Test) von Dickey und Fuller erfolgt im Modell

$$\Delta Y_t = \mu + \beta t + \pi Y_{t-1} + \varepsilon_t$$

mit

$$\begin{cases} \mu = \theta_0 (1 - \alpha) + \alpha \theta_1 \\ \beta = \theta_1 (1 - \alpha) \\ \pi = \alpha - 1 \end{cases} .$$

Differenzen- vs. trendstationäre lineare Modelle

Der ADF-Test

Der erweiterte (*augmented*) Einheitswurzeltest von Dickey und Fuller erfolgt im Modell

$$\Delta Y_t = \mu + \beta t + \pi Y_{t-1} + \sum_{j=1}^k \gamma_j \Delta Y_{t-j} + \varepsilon_t$$

mit

$$\begin{cases} \mu = \theta_0 (1 - \alpha) + \alpha \theta_1 \\ \beta = \theta_1 (1 - \alpha) \\ \pi = \alpha - 1 \end{cases} .$$

Differenzen- vs. trendstationäre lineare Modelle

Der ADF-Test

Modell: $\Delta Y_t = \mu + \beta t + \pi Y_{t-1} + \sum_{j=1}^k \gamma_j \Delta Y_{t-j} + \varepsilon_t$ **Testhypothesen:**

$$H_0 : \beta = \pi = 0 \quad \text{gegen} \quad H_1 : \beta \neq 0 \vee \pi \neq 0$$

Teststatistik: Die Teststatistik ist die übliche F -Teststatistik. Ihre Verteilung unter H_0 wurde von Dickey und Fuller (1981) tabelliert. **Testentscheidung**

(unter Berücksichtigung der jeweiligen Fehlerwahrscheinlichkeiten):

- H_0 beibehalten: die Hypothese eines differenzenstationären und nicht-trendstationären Modells kann nicht abgelehnt werden.
- H_0 ablehnen: die Hypothese eines differenzenstationären Modells wird abgelehnt. In diesem Fall können wir mit dem Ausgangsmodell wieder arbeiten und die Trendstationarität des Modells mit einem Standardtestverfahren überprüfen.

Differenzen- vs. trendstationäre lineare Modelle

Der ADF-Test

Modell: $Y_t = \theta_0 + \theta_1 t + Z_t$, mit $\alpha(B) Z_t = \varepsilon_t$, Z_t stationär. **Testhypothesen:**

$$H_0 : \theta_1 = 0 \quad \text{gegen} \quad H_1 : \theta_1 \neq 0$$

Teststatistik:

$$z = \frac{\hat{\theta}_1}{\hat{\sigma}_{\hat{\theta}_1}} \sim t_{N-2} | H_0$$

Testentscheidung (unter Berücksichtigung der jeweiligen Fehlerwahrscheinlichkeiten):

- H_0 beibehalten: die Hypothese eines nicht-trendstationären Modells kann nicht abgelehnt werden. **Insgesamt:** die Hypothese eines *stationären Modells* kann nicht abgelehnt werden.
- H_0 ablehnen: die Hypothese eines nicht-trendstationären Modells wird abgelehnt. **Insgesamt:** Hinweis auf ein *trendstationäres und nicht-differenzenstationäres Modell*.

Achtung! Der Störterm (Z_t) ist autokorreliert. Dazu noch mehr in Kapitel „Regressionsmodelle für Zeitreihen“.

Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 4

Literaturhinweise

Weiterführende Literatur

Kapitel 3.1, 3.2, 5.1 in

Pfaff, B. (2008), *Analysis of integrated and cointegrated time series with R*, 2. Auflage, Springer Verlag, New York

Kapitel 4: Prognose

Inhaltsübersicht

Prognose

- Die optimale Prognose aus der endlichen Vergangenheit

- Die optimale lineare Prognose aus der unendlichen Vergangenheit

 - Der Box-Jenkins-Ansatz

- Die exponentielle Glättung

- Auswertung von Prognoseergebnissen

Prognose

Die optimale Prognose aus der endlichen Vergangenheit

Kapitel 4: Prognose

4.1 Die optimale Prognose aus der endlichen Vergangenheit

Die optimale Prognose aus der endlichen Vergangenheit

Sei (Y_t) ein stochastischer Prozess mit $\mathbb{E}[Y_t^2] < \infty$ und $(y_t)_{1 \leq t \leq N}$ die zur Verfügung stehende Zeitreihe.

- Gesucht ist eine Prognose für den Wert y_{N+h} auf der Grundlage der Zeitreihe $(y_t)_{1 \leq t \leq N}$.
- Für einen positiven Prognosehorizont $h > 0$ liegt y_{N+h} in der Zukunft, für einen negativen Prognosehorizont $h < 0$ – in der Vergangenheit.
- Eine Prognose wird mit Hilfe einer Prognosefunktion $\hat{Y}_{N,h}$ der Zeitreihenvariablen des Prozesses

$$\hat{Y}_{N,h}(Y_1, Y_2, \dots, Y_N)$$

berechnet.

Die optimale Prognose aus der endlichen Vergangenheit

Prognosen sind *Schätzungen*, Prognosefunktionen – *Schätzfunktionen* für eine Realisation y_{N+h} des Prozesses.

Die optimale Prognose aus der endlichen Vergangenheit

Wir betrachten den Prognosefehler

$$Y_{N+h} - \hat{Y}_{N,h}.$$

Gesucht ist die optimale/beste Prognosefunktion für y_{N+h} , d.h. die Prognosefunktion, welche den mittleren quadratischen Prognosefehler $MSE[\hat{Y}_{N,h}]$ minimiert

$$MSE[\hat{Y}_{N,h}] = \mathbb{E}\left[\left(Y_{N+h} - \hat{Y}_{N,h}\right)^2\right].$$

Die optimale Prognose aus der endlichen Vergangenheit

Theorem (Optimale Prognose)

Sei (Y_t) ein stochastischer Prozess mit $\mathbb{E}[Y_t^2] < \infty$. Die optimale Prognose von y_{N+h} mit Hilfe einer Prognosefunktion der Zeitreihenvariablen $Y_1, Y_2, \dots, Y_N$ ist der bedingte Erwartungswert

$$\hat{Y}_{N+h}(Y_1, Y_2, \dots, Y_N) = \mathbb{E}[Y_{N+h} | Y_1 = y_1, Y_2 = y_2, \dots, Y_N = y_N].$$

Die optimale lineare Prognose aus der endlichen Vergangenheit

Theorem (Optimale lineare Prognose)

Sei (Y_t) ein stationärer stochastischer Prozess mit Autokorrelationsfunktion (ρ_τ) . Die optimale Prognose von y_{N+h} mit Hilfe einer linearen Prognosefunktion

$$\hat{Y}_{N,h}(Y_1, Y_2, \dots, Y_N) = \sum_{u=0}^{N-1} a_u Y_{N-u}$$

benutzt die Gewichte $(a_u)_{0 \leq u \leq N-1}$, welche das System der Yule-Walker-Gleichungen löst

$$\begin{pmatrix} 1 & \rho_1 & \dots & \rho_{N-1} \\ \rho_1 & 1 & \dots & \rho_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N-1} & \rho_{N-2} & \dots & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{pmatrix} = \begin{pmatrix} \rho_h \\ \rho_{h+1} \\ \vdots \\ \rho_{h+N-1} \end{pmatrix}.$$

Die optimale lineare Prognose aus der endlichen Vergangenheit

Die optimale lineare Prognose $\hat{Y}_{N,h}$ aus der endlichen Vergangenheit $\{y_1, y_2, \dots, y_N\}$ kann mit Hilfe der R-Funktion *TrenchForecasts* aus dem Paket *ltsa* (Linear Time Series Analysis) implementiert werden.

- Dabei werden die Autokovarianzen mit Lags $\tau \in \{0, 1, 2, \dots, h + N - 1\}$ benötigt. Werden die $h + N$ Werte der Autokovarianzfunktion aus der Zeitreihe $\{y_1, y_2, \dots, y_N\}$ geschätzt, so können lediglich Prognosen mit negativem Prognosehorizont $h < 0$ erstellt werden. Für Prognosen in die Zukunft muss die Autokovarianzfunktion bekannt sein oder mit einem alternativen Verfahren geschätzt werden.

Für ARIMA-Modelle kann die optimale lineare Prognose $\hat{Y}_{N,h}$ aus der endlichen Vergangenheit auch mit Hilfe der Funktion *predict* berechnet werden.

Prognose

Die optimale lineare Prognose aus der unendlichen Vergangenheit

Kapitel 4: Prognose

4.2 Die optimale lineare Prognose aus der unendlichen Vergangenheit

Die optimale lineare Prognose aus der unendlichen Vergangenheit

Minimierung des mittleren quadratischen Prognosefehlers

Wir betrachten im Folgenden lineare Prognosen aus der unendlichen Vergangenheit

$$\hat{Y}_{t,h}(Y_t, Y_{t-1}, \dots) = \sum_{u=0}^{\infty} a_u Y_{t-u}.$$

Der Prognosefehler ist

$$Y_{t+h} - \hat{Y}_{t,h}.$$

Gesucht ist die optimale/beste Prognosefunktion für y_{t+h} , d.h. die Prognosefunktion, welche den mittleren quadratischen Prognosefehler $\text{MSE}[\hat{Y}_{t,h}]$ minimiert

$$\text{MSE}[\hat{Y}_{t,h}] = \mathbb{E} \left[\left(Y_{t+h} - \hat{Y}_{t,h} \right)^2 \right].$$

Die optimale lineare Prognose aus der unendlichen Vergangenheit

Minimierung des mittleren quadratischen Prognosefehlers

Diese Minimierungsaufgabe führt zu einem unendlichen System von Yule-Walker-Gleichungen mit den Unbekannten (a_u)

$$\begin{pmatrix} 1 & \rho_1 & \dots \\ \rho_1 & 1 & \dots \\ \vdots & \vdots & \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \end{pmatrix} = \begin{pmatrix} \rho_h \\ \rho_{h+1} \\ \vdots \end{pmatrix}.$$

Die optimale lineare Prognose aus der unendlichen Vergangenheit

Prognose für MA[∞]-Prozesse

Sei (Y_t) ein MA[∞]-Prozess

$$Y_t = - \sum_{u=0}^{\infty} \beta_u \varepsilon_{t-u},$$

wobei $\beta_0 = -1$ gesetzt wurde.

Die optimale lineare Prognose für den Wert y_{t+h} ist

$$\hat{Y}_{t,h} = - \sum_{u=0}^{\infty} \beta_{h+u} \varepsilon_{t-u}.$$

Die optimale lineare Prognose aus der unendlichen Vergangenheit

Unkorreliertheit zwischen dem Prognosefehler und den Zeitreihenvariablen

Alternativ kann die Folge (a_u) mit Hilfe des folgenden Satzes bestimmt werden.

Theorem

Sei (Y_t) ein rein nicht-deterministischer Prozess. Die lineare Prognose aus der unendlichen Vergangenheit

$$\hat{Y}_{t,h}(Y_t, Y_{t-1}, \dots) = \sum_{u=0}^{\infty} a_u Y_{t-u}$$

ist dann und nur dann optimal, wenn der Prognosefehler $Y_{t+h} - \hat{Y}_{t,h}$ unkorreliert mit allen Zeitreihenvariablen $Y_t, Y_{t-1}, \dots$ ist.

Die optimale lineare Prognose aus der unendlichen Vergangenheit

Unkorreliertheit zwischen dem Prognosefehler und den Zeitreihenvariablen

Wir betrachten den Prognosefehler zum Zeitpunkt $t + h$

$$Y_{t+h} - \hat{Y}_{t,h}$$

und den mittleren quadratischen Prognosefehler

$$\text{MSE}[\hat{Y}_{t,h}] = \mathbb{E}\left[\left(Y_{t+h} - \hat{Y}_{t,h}\right)^2\right].$$

Sei $\Delta(h)$ der mittlere quadratische Prognosefehler für die optimale lineare Prognose

$$\Delta(h) := \min_{(a_u)} \text{MSE}[\hat{Y}_{t,h}].$$

Die optimale lineare Prognose aus der unendlichen Vergangenheit

Unkorreliertheit zwischen dem Prognosefehler und den Zeitreihenvariablen

Definition (deterministischer Prozess)

Ein stationärer Prozess (Y_t) heißt **deterministisch** (singulär, exakt vorhersagbar), wenn gilt

$$\Delta(h) = 0 \quad \forall h > 0.$$

Definition (rein nicht-deterministischer Prozess)

Ein stationärer Prozess (Y_t) heißt rein nicht-deterministisch, wenn gilt

$$\lim_{h \rightarrow \infty} \Delta(h) = \gamma_0.$$

Ein stationärer Prozess ist dann und nur dann rein nicht-deterministisch, wenn er sich als $\text{MA}[\infty]$ -Prozess

$$Y_t = \sum_{u=0}^{\infty} \beta_u \varepsilon_{t-u}$$

mit quadratisch summierbaren Gewichten $\sum_{u=0}^{\infty} \beta_u^2 < \infty$ darstellen lässt

Die optimale lineare Prognose aus der unendlichen Vergangenheit

Die optimale lineare Prognose aus der unendlichen Vergangenheit

Unkorreliertheit zwischen dem Prognosefehler und den Zeitreihenvariablen

Theorem (Wold'scher Zerlegungssatz)

Jeder stationäre Prozess (Y_t) kann eindeutig als Summe zweier unkorrelierter stationärer Prozesse, nämlich eines deterministischen und eines rein nicht-deterministischen Prozesses dargestellt werden.

Weiter werden wir uns nur mit der Prognose rein nicht-deterministischer Prozesse beschäftigen.

Theorem

Sei (Y_t) ein rein nicht-deterministischer Prozess. Die lineare Prognose aus der unendlichen Vergangenheit

$$\hat{Y}_{t,h}(Y_t, Y_{t-1}, \dots) = \sum_{u=0}^{\infty} a_u Y_{t-u}$$

ist dann und nur dann optimal, wenn der Prognosefehler $Y_{t+h} - \hat{Y}_{t,h}$ unkorreliert mit allen Zeitreihenvariablen $Y_t, Y_{t-1}, \dots$ ist.

Der Box-Jenkins-Ansatz

Der Box-Jenkins-Ansatz ist ein *rekursives* Berechnungsverfahren der optimalen linearen Prognose aus der unendlichen Vergangenheit für allgemeine ARMA-Prozesse.

Der Box-Jenkins-Ansatz

Prognose für MA[q]-Prozesse

Sei (Y_t) ein MA[q]-Prozess

$$Y_t = - \sum_{u=0}^q \beta_u \varepsilon_{t-u},$$

wobei $\beta_0 = -1$ gesetzt wurde. Wir können (Y_t) als MA[∞]-Prozess mit Parametern $\beta_u = 0$ für $u > q$ darstellen.

Dann lautet die optimale lineare Prognose für den Wert y_{t+h}

$$\hat{Y}_{t,h} = \begin{cases} - \sum_{u=0}^{q-h} \beta_{h+u} \varepsilon_{t-u}, & h \leq q \\ 0, & h > q \end{cases}$$

Der Box-Jenkins-Ansatz

Prognose für MA[q]-Prozesse

Die Störungen $(\varepsilon_u)_{t+h-q \leq u \leq t}$ entsprechen den Prognosefehlern der 1-Schritt-Prognosen

$$\begin{aligned}\varepsilon_t &= Y_t - \hat{Y}_{t-1,1} \\ \varepsilon_{t-1} &= Y_{t-1} - \hat{Y}_{t-2,1} \\ \varepsilon_{t-2} &= Y_{t-2} - \hat{Y}_{t-3,1} \\ &\vdots \\ \varepsilon_{t-h-q} &= Y_{t+h-q} - \hat{Y}_{t+h-q-1,1}\end{aligned}$$

Sie werden *rekursiv* berechnet.

Der Box-Jenkins-Ansatz

Prognose für MA[q]-Prozesse

Wir schreiben die Definitionsgleichung für Y_{t+h} fort

$$Y_{t+h} = \varepsilon_{t+h} - \beta_1 \varepsilon_{t+h-1} - \dots - \beta_h \varepsilon_t - \beta_{h+1} \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t+h-q}.$$

Die Prognose $\hat{Y}_{t,h}$ wird aus Y_{t+h} , wie folgt, berechnet

$$\hat{Y}_{t,h} = \varepsilon_{t+h} - \beta_1 \varepsilon_{t+h-1} - \dots - \beta_h \varepsilon_t - \beta_{h+1} \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t+h-q}.$$

- Die Störungen $\varepsilon_{t+h}, \varepsilon_{t+h-1}, \dots, \varepsilon_{t+1}$ werden mit ihrem Erwartungswert von Null ersetzt.
- Die Störungen $\varepsilon_t, \varepsilon_{t-1}, \dots$ entsprechen den Prognosefehlern $Y_t - \hat{Y}_{t-1,1}, Y_{t-1} - \hat{Y}_{t-2,1}, \dots$ der 1-Schritt-Prognosen.

Der Box-Jenkins-Ansatz

Prognose für AR[p]-Prozesse

Sei (Y_t) ein AR[p]-Prozess

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t.$$

Die optimalen linearen Prognosen sind

$$\begin{aligned}\hat{Y}_{t,1} &= \alpha_1 Y_t + \alpha_2 Y_{t-1} + \dots + \alpha_p Y_{t+1-p} \\ \hat{Y}_{t,2} &= \alpha_1 \hat{Y}_{t,1} + \alpha_2 Y_t + \alpha_3 Y_{t-1} + \dots + \alpha_p Y_{t+2-p}\end{aligned}$$

Für einen beliebigen Prognosehorizont h

$$\begin{aligned}\hat{Y}_{t,h} &= \alpha_1 \hat{Y}_{t,h-1} + \alpha_2 \hat{Y}_{t,h-2} + \dots + \alpha_{h-1} \hat{Y}_{t,1} + \\ &\quad \alpha_h Y_t + \alpha_{h+1} Y_{t-1} + \dots + \alpha_p Y_{t+h-p}\end{aligned}$$

Der Box-Jenkins-Ansatz

Prognose für AR[p]-Prozesse

Wir schreiben die Definitionsgleichung für Y_{t+h} fort

$$Y_{t+h} = \alpha_1 Y_{t+h-1} + \alpha_2 Y_{t+h-2} + \dots + \alpha_p Y_{t+h-p} + \varepsilon_{t+h}.$$

Die Prognose $\hat{Y}_{t,h}$ wird aus Y_{t+h} , wie folgt, berechnet

$$\hat{Y}_{t,h} = \alpha_1 Y_{t+h-1} + \alpha_2 Y_{t+h-2} + \dots + \alpha_p Y_{t+h-p} + \varepsilon_{t+h}.$$

- Die unbekannten Variablen $Y_{t+h-1}, \dots, Y_{t+1}$ werden mit den optimalen linearen Prognosen $\hat{Y}_{t,h-1}, \dots, \hat{Y}_{t,1}$ ersetzt.
- Für die Variablen $Y_t, Y_{t-1}, \dots$ setzen wir ihre Beobachtungen ein.
- Die Störung ε_{t+h} wird mit ihrem Erwartungswert von Null ersetzt.

Der Box-Jenkins-Ansatz

Prognose für ARMA $[p, q]$ -Prozesse

Sei (Y_t) ein ARMA $[p, q]$ -Prozess

$$Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q}.$$

Wir schreiben die Definitionsgleichung für Y_{t+h} fort

$$Y_{t+h} = \alpha_1 Y_{t+h-1} + \dots + \alpha_p Y_{t+h-p} + \varepsilon_{t+h} - \beta_1 \varepsilon_{t+h-1} - \dots - \beta_q \varepsilon_{t+h-q}.$$

Der Box-Jenkins-Ansatz

Prognose für ARMA[p, q]-Prozesse

Die Prognose $\hat{Y}_{t,h}$ wird aus Y_{t+h} , wie folgt, berechnet

$$\hat{Y}_{t,h} = \alpha_1 Y_{t+h-1} + \dots + \alpha_p Y_{t+h-p} + \varepsilon_{t+h} - \beta_1 \varepsilon_{t+h-1} - \dots - \beta_q \varepsilon_{t+h-q}.$$

- Die unbekannten Variablen $Y_{t+h-1}, \dots, Y_{t+1}$ werden mit den optimalen linearen Prognosen $\hat{Y}_{t,h-1}, \dots, \hat{Y}_{t,1}$ ersetzt.
- Für die Variablen $Y_t, Y_{t-1}, \dots$ setzen wir ihre Beobachtungen ein.
- Die Störungen $\varepsilon_{t+h}, \varepsilon_{t+h-1}, \dots, \varepsilon_{t+1}$ werden mit ihrem Erwartungswert von Null ersetzt.
- Die Störungen $\varepsilon_t, \varepsilon_{t-1}, \dots$ entsprechen den Prognosefehlern $Y_t - \hat{Y}_{t-1,1}, Y_{t-1} - \hat{Y}_{t-2,1}, \dots$ der 1-Schritt-Prognosen.

Der Box-Jenkins-Ansatz

Prognose für ARMA[p, q]-Prozesse

Wir schreiben den ARMA[p, q]-Prozess (Y_t) , wie folgt, auf:

$$Y_t = c(B) \varepsilon_t$$

mit

$$c(B) = \alpha^{-1}(B) \beta(B).$$

Die Varianz des Prognosefehlers ist

$$\text{Var}\left[Y_{t+h} - \hat{Y}_{t,h}\right] = \sigma_\varepsilon^2 \sum_{u=0}^{h-1} c_u^2$$

mit (c_u) die Gewichte des Filters $c(B)$.

Der Box-Jenkins-Ansatz

Prognose für ARMA $[p, q]$ -Prozesse

Die optimale lineare Prognose eines ARMA $[p, q]$ -Modells hat die folgenden Eigenschaften:

$$\lim_{h \rightarrow \infty} \text{Var} \left[Y_{t+h} - \hat{Y}_{t,h} \right] = \text{Var}[Y_t]$$

$$\lim_{h \rightarrow \infty} \hat{Y}_{t,h} = \mathbb{E}[Y_t] = 0$$

Daher ist sie *unbrauchbar für großen Prognosehorizont h* .

Der Box-Jenkins-Ansatz

Prognose für ARMA $[p, q]$ -Prozesse

Sei (Y_t) ein Gaußscher ARMA $[p, q]$ -Prozess. Dann gilt

$$Y_{t+h} - \hat{Y}_{t,h} \sim N \left(0, \sigma_\varepsilon^2 \sum_{u=0}^{h-1} c_u^2 \right).$$

Prognoseintervall für Y_{t+h} :

$$\hat{Y}_{t,h} \pm z_{1-\alpha/2} \sigma_\varepsilon \sqrt{\sum_{u=0}^{h-1} c_u^2}$$

mit $z_{1-\alpha/2}$ das $1 - \alpha/2$ -Quantil der Standardnormalverteilung.

Der Box-Jenkins-Ansatz

Prognose für angepasste ARMA[p, q]-Prozesse

Der Box-Jenkins-Ansatz ist ein rekursives Berechnungsverfahren für die optimale lineare Prognose $\hat{Y}_{t,h}$, wenn

- die Parameter $\alpha = (\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_p)$ und $\beta = (\beta_1 \quad \beta_2 \quad \dots \quad \beta_q)$ bekannt sind,
- eine unendliche Zeitreihe $y_{-\infty}, \dots, y_t$ zur Verfügung steht (unendliche Vergangenheit).

Der Box-Jenkins-Ansatz

Prognose für angepasste ARMA[p, q]-Prozesse

In praktischen Anwendungen verwenden wir eine angepasste Prognosefunktion auf der Grundlage von $\hat{Y}_{t,h}$. Dabei setzen wir die folgenden Werte ein:

- die Parameterschätzungen $\hat{\alpha} = (\hat{\alpha}_1 \quad \hat{\alpha}_2 \quad \dots \quad \hat{\alpha}_p)$ und $\hat{\beta} = (\hat{\beta}_1 \quad \hat{\beta}_2 \quad \dots \quad \hat{\beta}_q)$ nach einer der Methoden CLS, ULS oder ML,
- geeignete Startwerte $\tilde{\varepsilon}_0, \tilde{\varepsilon}_{-1}, \dots, \tilde{Y}_0, \tilde{Y}_{-1}, \dots$
- Schätzungen (Approximationen) $\tilde{\varepsilon}_t, \dots, \tilde{\varepsilon}_1$ für die Störungen $\varepsilon_t, \dots, \varepsilon_1$.

Diese Prognosefunktion bezeichnen wir im Folgenden mit $\tilde{Y}_{t,h}$.

Der Box-Jenkins-Ansatz

Prognose für angepasste ARMA[p, q]-Prozesse

Theorem

Sei (Y_t) ein stationärer und invertierbarer ARMA[p, q]-Prozess, $\alpha(B)Y_t = \beta(B)\varepsilon_t$, mit White-Noise-Prozess (ε_t) . Sei $\tilde{Y}_{t,h}$ die entsprechend angepasste Prognosefunktion auf der Grundlage der optimalen linearen Prognose aus der unendlichen Vergangenheit $\hat{Y}_{t,h}$. Dann gilt für $\varepsilon > 0$ beliebig

$$\lim_{N \rightarrow \infty} \mathbb{P} \left[\sqrt{N} \left| \tilde{Y}_{t,h} - \hat{Y}_{t,h} \right| > \varepsilon \right] = 0.$$

- Die Anwendung der Prognosefunktion $\tilde{Y}_{t,h}$ anstatt von $\hat{Y}_{t,h}$ hat einen vernachlässigbaren Effekt auf die Prognosegüte.

Schlittgen und Streitberg (1999), Zeitreihenanalyse: Kapitel 4.1.1, 4.2.1, 4.3.1
Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 5.2

Kapitel 4: Prognose

4.3 Die exponentielle Glättung

Die exponentielle Glättung

Die exponentielle Glättung ist ein Extrapolationsverfahren für die Berechnung von Prognosen. Es basiert auf das Bilden rekursiver Fortschreibungen der Form

$$\hat{y}_{N,1} = (1 - \alpha)\hat{y}_{N-1,1} + \alpha y_N.$$

Durch rekursives Einsetzen der Prognosen erhalten wir die Darstellung als exponentiell gewichtete Summe von Zeitreihenwerten

$$\hat{y}_{N,1} = (1 - \alpha)^{N-1}y_1 + \alpha \sum_{u=0}^{N-2} (1 - \alpha)^u y_{N-u}.$$

Dabei wurde das Verfahren mit $\hat{y}_{1,1} = y_1$ initialisiert.

Die exponentielle Glättung

Alternative Darstellung:

$$\hat{y}_{N,1} = \hat{y}_{N-1,1} + \alpha(y_N - \hat{y}_{N-1,1})$$

Eine neue Prognose wird durch Aufdatierung der letzten Prognose erstellt.
Dabei erfolgt eine Korrektur um den letzten Prognosefehler.

Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 5.1

Kapitel 4: Prognose

4.4 Auswertung von Prognoseergebnissen

Auswertung von Prognoseergebnissen

Prognoseverfahren:

- optimale lineare Prognose $\hat{Y}_{N,h}$ aus der endlichen Vergangenheit $\{y_1, y_2, \dots, y_N\}$. Berechnung durch Lösen der Yule-Walker-Gleichungen mit Hilfe des Levinson-Durbin-Algorithmus.
- optimale lineare Prognose $\hat{Y}_{t,h}$ aus der unendlichen Vergangenheit $\{y_t, y_{t-1}, y_{t-2}, \dots\}$. Speziell für ARIMA-Modelle kann $\hat{Y}_{t,h}$ mit Hilfe des Box-Jenkins-Ansatzes berechnet werden.

Auswertung von Prognoseergebnissen

Gegeben sei eine Zeitreihe (y_t) der Länge N , $t \in \{1, 2, \dots, N\}$. Wir unterscheiden zwischen

- **in-sample Prognose (innerhalb der Stichprobe):** wir schätzen das Zeitreihenmodell (ZR-Modell) aus der Zeitreihe (y_t) und erstellen Prognosen ebenfalls für (y_t) .
- **out-of-sample Prognose (außerhalb der Stichprobe):** wir schätzen das ZR-Modell aus der Zeitreihe (y_t) . Die Prognosen werden für einen unterschiedlichen, zukünftigen Zeitraum $s \in \{N + 1, N + 2, \dots, N + H\}$ erstellt.

Auswertung von Prognoseergebnissen

Gegeben sei eine Zeitreihe (y_t) der Länge N . Wir erstellen Prognosen $(\hat{y}_s)$ für den Prognosezeitraum $s \in \{N+1, N+2, \dots, N+H\}$. Wir untersuchen die Prognosegüte von $(\hat{y}_s)$ ex-post, d.h. durch Vergleiche zwischen den Prognosen $(\hat{y}_s)$ und den tatsächlichen Realisationen (y_s) .

Fragestellung 1: Wir berechnen $(\hat{y}_s)$. Wir möchten die Prognosegüte von $(\hat{y}_s)$ beurteilen.

Fragestellung 2: Wir erstellen Prognosen $(\hat{y}_s^A)$ und $(\hat{y}_s^B)$ mit zwei unterschiedlichen ZR-Modellen oder zwei unterschiedlichen Prognoseverfahren A und B . Wir möchten bestimmen, welches Modell/Verfahren die beste Prognose geliefert hat.

Auswertung von Prognoseergebnissen

Empirische Maßzahlen

Wir berechnen

- den empirischen mittleren absoluten Prognosefehler

$$\widehat{MAE}(\hat{y}_s) = \frac{1}{H} \sum_{s=N+1}^{N+H} |y_s - \hat{y}_s|$$

- den empirischen mittleren quadratischen Prognosefehler

$$\widehat{MSE}(\hat{y}_s) = \frac{1}{H} \sum_{s=N+1}^{N+H} (y_s - \hat{y}_s)^2.$$

Niedrige Werte dieser empirischen Maßzahlen sprechen für eine gute Prognose ($\hat{y}_s$).

Auswertung von Prognoseergebnissen

Mincer-Zarnowitz-Regression

Wir führen eine einfache lineare Regression der Realisationen (y_s) auf die Prognosen $(\hat{y}_s)$

$$y_s = \beta_0 + \beta_1 \hat{y}_s + u_s, \quad s = N + 1, N + 2, \dots, N + H$$

durch.

- Eine gute Prognose $(\hat{y}_s)$ ist durch ein hohes Bestimmtheitsmaß R^2 gekennzeichnet.
- Auch können wir einen F -Test mit den Testhypothesen:

$$H_0 : \beta_0 = 0 \wedge \beta_1 = 1 \quad \text{gegen} \quad H_1 : \beta_0 \neq 0 \vee \beta_1 \neq 1$$

durchführen. Die Beibehaltung der H_0 -Hypothese weist auf eine gute Prognose $(\hat{y}_s)$ hin.

Achtung! Der Störterm (u_s) ist nicht White-Noise. Dazu noch mehr in Kapitel „Regressionsmodelle für Zeitreihen“.

Auswertung von Prognoseergebnissen

Wir betrachten die Prognosen $(\hat{y}_s^A)$ und $(\hat{y}_s^B)$ mit zwei unterschiedlichen ZR-Modellen/Prognoseverfahren A und B . Wir können die beiden Prognosen bzgl.

- der empirischen Maßzahlen
- der Güte der jeweiligen Mincer-Zarnowitz-Regressionen

miteinander vergleichen.

Diese Resultate werden auf der Grundlage einer Realisation (y_s) ermittelt und sind dem Zufall überlassen. Neue Realisationen (y_s) führen möglicherweise zu unterschiedlichen Resultaten. Wir brauchen Methoden, womit wir einen statistisch signifikanten Unterschied zwischen $(\hat{y}_s^A)$ und $(\hat{y}_s^B)$ belegen können.

Auswertung von Prognoseergebnissen

Diebold-Mariano-Test

Wir möchten die Prognosen $(\hat{y}_s^A)$ und $(\hat{y}_s^B)$ mit zwei ZR-Modellen/Prognoseverfahren A und B bzgl. ihrer Genauigkeit vergleichen. Wir betrachten

- den Abstand zwischen ihren quadrierten Prognosefehlern zum Zeitpunkt s

$$d_s = (y_s - \hat{y}_s^A)^2 - (y_s - \hat{y}_s^B)^2$$

- den Abstand zwischen ihren empirischen MSE -Werten

$$\begin{aligned}\bar{d} &= \widehat{MSE}(\hat{y}_s^A) - \widehat{MSE}(\hat{y}_s^B) \\ &= \frac{1}{H} \sum_{s=N+1}^{N+H} (y_s - \hat{y}_s^A)^2 - \frac{1}{H} \sum_{s=N+1}^{N+H} (y_s - \hat{y}_s^B)^2.\end{aligned}$$

Auswertung von Prognoseergebnissen

Diebold-Mariano-Test

Theorem (Prognosen mit gleicher Prognosegenauigkeit)

Zwei Prognosen $(\hat{y}_s^A)$ und $(\hat{y}_s^B)$ haben die gleiche Genauigkeit dann und nur dann wenn

$$E[d_s] = 0$$

für alle Prognosezeitpunkte s gilt.

Auswertung von Prognoseergebnissen

Diebold-Mariano-Test

Testhypothesen:

$$H_0 : E[d_s] = 0 \quad \forall s \quad \text{gegen} \quad H_1 : \exists s \quad E[d_s] \neq 0$$

Teststatistik:

$$\frac{\sqrt{H} \bar{d}}{\sqrt{\hat{\omega}}} \xrightarrow{d} N(0, 1) | H_0$$

mit $\hat{\omega}$ einem asymptotischen konsistenten Schätzer der Größe $\text{Var}[\sqrt{H} \bar{d}]$.

Testentscheidung:

- Bestimme den kritischen Wert κ_α zu gegebenem $\alpha \in (0, 1)$:

$$\kappa_\alpha = z_{1-\frac{\alpha}{2}}.$$

- Lehne H_0 mit Irrtumswahrscheinlichkeit α ab, falls

$$|z| > \kappa_\alpha.$$

Dann haben die Prognosen $(\hat{y}_s^A)$ und $(\hat{y}_s^B)$ unterschiedliche Prognosegenauigkeit (Irrtumswahrscheinlichkeit α).

Die optimale Prognose aus der endlichen Vergangenheit Die optimale lineare Prognose aus der unendlichen Vergangenheit

Kapitel 5:
Statistische Methoden im Frequenzbereich

Statistische Methoden im Frequenzbereich

- Das Periodogramm

- Fouriertransformationen für Zeitreihen

- Das Spektrum

- Lineare Filter im Frequenzbereich

- Spektralschätzung

 - Das Periodogramm als Spektralschätzer

 - Direkte Spektralschätzer

 - Indirekte Spektralschätzer

 - Parametrische Spektralschätzer

Kapitel 5:

Statistische Methoden im Frequenzbereich

5.1 Das Periodogramm

Periodizitäten in Zeitreihen

Der Datensatz enthält die in DE monatlich gezahlten Gemeinschaftssteuern in Mio. Euro, Jan. 1994 – Jan. 2000.

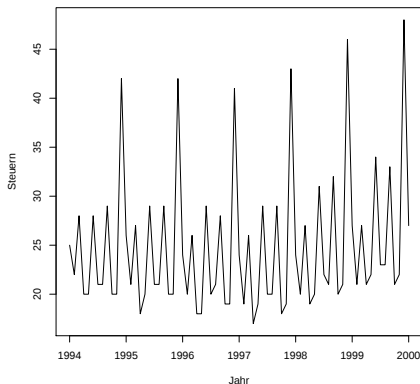


Abb. 5.1: Die Zeitreihe *Steuern*

Periodizitäten in Zeitreihen

Die Frequenz

Die Zeitreihe *Steuern* weist zwei Arten von periodischen Schwankungen auf:

- Quartalsschwankungen mit der Periode $P = 3$
- jährliche Schwankungen mit der Periode $P = 12$.

Diese Periodizitäten haben jeweils die Frequenzen

- $\lambda = 1/3$, d.h. in einem Monat läuft ein drittel Quartalszyklus ab
- $\lambda = 1/12$, d.h. in einem Monat läuft ein zwölftel Jahreszyklus ab.

Periodizitäten in Zeitreihen

Die Frequenz

Definition (Frequenz)

Sei $f(t)$ eine periodische Funktion mit der Periode P , d.h. $f(t + P) = f(t)$. Die **Frequenz** λ von $f(t)$ ist der Kehrwert von P

$$\lambda = \frac{1}{P}.$$

Eine periodische Zeitreihe (y_t) besteht aus sich nahezu identisch wiederkehrenden Teilsegmenten (Zyklen). Die Periode gibt die Länge eines Zyklus an, die Frequenz besagt, wieviele Zyklen in einer Zeiteinheit ablaufen.

Periodizitäten in Zeitreihen

Die Frequenz

- Viele der folgenden Resultate erlauben eine besonders einfache Interpretation für die so genannten *Fourierfrequenzen*:

$$\lambda = \frac{k}{N}, \quad k \in \mathbb{Z}.$$

- Das Periodogramm $I(\lambda)$ ist eine Funktion der Frequenz λ . Mit dem Periodogramm können die Periodizitäten in einer Zeitreihe aufgedeckt werden. Ein Peak im Periodogramm bei einer Frequenz λ^* ist ein Hinweis auf eine mögliche Periodizität mit Frequenz λ^* .

Das Periodogramm

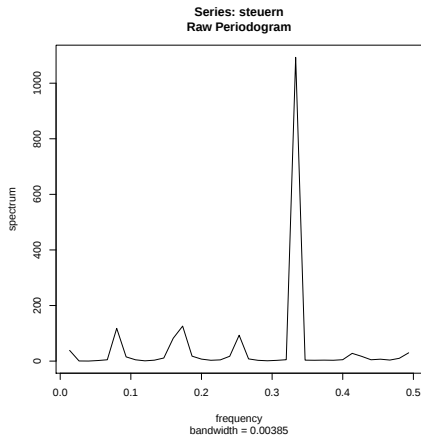


Abb. 5.2: Das Periodogramm der Reihe *Steuern*

Definition (Periodogramm)

Das **Periodogramm** oder Stichprobenspektrum der Zeitreihe (y_t) ist eine Funktion der Frequenz λ

$$I(\lambda) = N \left(C(\lambda)^2 + S(\lambda)^2 \right)$$

mit $C(\lambda)$ die empirische Kovarianz der Zeitreihe (y_t) mit der Cosinuswelle $(\cos(2\pi\lambda t))$

$$C(\lambda) = \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) \cos(2\pi\lambda t) .$$

und $S(\lambda)$ die empirische Kovarianz der Zeitreihe (y_t) mit der Sinuswelle $(\sin(2\pi\lambda t))$

$$S(\lambda) = \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) \sin(2\pi\lambda t) .$$

Das Periodogramm

Die Kovarianzen $C(\lambda)$ und $S(\lambda)$ wurden mit der folgenden Formel für die Kovarianz ermittelt:

$$c_{YX} = \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) x_t.$$

Die Gültigkeit dieser Formel ist leicht zu prüfen:

$$\begin{aligned} c_{YX} &= \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) (x_t - \bar{x}) \\ &= \frac{1}{N} \sum_{t=1}^N y_t x_t - \bar{y} \bar{x} \\ &= \frac{1}{N} \sum_{t=1}^N y_t x_t - \frac{1}{N} \sum_{t=1}^N \bar{y} x_t \\ &= \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) x_t. \end{aligned}$$

Das Periodogramm

Eigenschaften des Periodogramms

1. $I(0) = 0$
2. $I(-\lambda) = I(\lambda)$, d.h. das Periodogramm ist eine gerade Funktion von λ .
3. $I(\lambda + 1) = I(\lambda)$, d.h. das Periodogramm ist periodisch mit $P = 1$.
4. $I(\lambda) \geq 0$

Aufgrund der Eigenschaften 2. und 3. ist es ausreichend, das Periodogramm im Intervall $[0, 0.5]$ zu betrachten.

Das Periodogramm

Das Periodogramm $I(\lambda)$ einer Zeitreihe (y_t) bemisst den Erklärungsgehalt einer harmonischen Schwingung mit der Frequenz λ für die zentrierte Zeitreihe $(y_t - \bar{y})$.

- Als *harmonische Schwingung* bezeichnen wir die Funktion

$$f(t) = a \cos(2\pi\lambda t) + b \sin(2\pi\lambda t)$$

mit $a, b \in \mathbb{R}$. Harmonische Schwingungen sind periodisch mit der Frequenz λ , bzw. mit der Periode $P = 1/\lambda$:

$$\begin{aligned} f\left(t + \frac{1}{\lambda}\right) &= a_1 \cos\left(2\pi\lambda\left(t + \frac{1}{\lambda}\right)\right) + a_2 \sin\left(2\pi\lambda\left(t + \frac{1}{\lambda}\right)\right) \\ &= a_1 \cos(2\pi\lambda t + 2\pi) + a_2 \sin(2\pi\lambda t + 2\pi) \\ &= a_1 \cos(2\pi\lambda t) + a_2 \sin(2\pi\lambda t). \end{aligned}$$

Das Periodogramm

Das Regressionsmodell mit einer harmonischen Schwingung

Wir betrachten das lineare Regressionsmodell

$$y_t - \bar{y} = a \cos(2\pi\lambda t) + b \sin(2\pi\lambda t) + \varepsilon_t. \quad (5.1)$$

In diesem Modell wird die zentrierte Reihe $(y_t - \bar{y})$ mit Hilfe lediglich einer harmonischen Schwingung der Frequenz λ erklärt. Das geschätzte Modell:

$$\widehat{y_t - \bar{y}} = \hat{a} \cos(2\pi\lambda t) + \hat{b} \sin(2\pi\lambda t).$$

Die Residuen:

$$\hat{\varepsilon}_t = y_t - \bar{y} - \hat{a} \cos(2\pi\lambda t) - \hat{b} \sin(2\pi\lambda t).$$

Das Periodogramm

Das Regressionsmodell mit einer harmonischen Schwingung

Für Fourierfrequenzen $\lambda = k / N$ hat die OLS-Schätzung im Modell (5.1) eine besonders einfache Darstellung:

$$\hat{a} = 2C(\lambda), \quad \hat{b} = 2S(\lambda)$$

und

$$\begin{aligned} RSS &= \sum_{t=1}^N \hat{\varepsilon}_t^2 \\ &= \sum_{t=1}^N (y_t - \bar{y})^2 - 2N \left(C(\lambda)^2 + S(\lambda)^2 \right). \end{aligned}$$

Der Wert $N \left(C(\lambda)^2 + S(\lambda)^2 \right)$ in der Residuenquadratsumme entspricht dem Wert des Periodogramms an der Stelle λ , $I(\lambda)$. Daher: je höher der Wert des Periodogramms $I(\lambda)$ an der Stelle λ , umso besser die Anpassung im Modell (5.1). (Für Nichtfourierfrequenzen erlaubt $I(\lambda)$ eine ähnliche Interpretation.)

Das Periodogramm

Das Regressionsmodell mit mehreren harmonischen Schwingungen

Wir betrachten das lineare Regressionsmodell

$$y_t - \bar{y} = \sum_{i=1}^r (a_i \cos(2\pi \lambda_i t) + b_i \sin(2\pi \lambda_i t)) + \varepsilon_t. \quad (5.2)$$

In diesem Modell wird die zentrierte Reihe $(y_t - \bar{y})$ mit Hilfe von r harmonischen Schwingungen der Frequenzen λ_i , $i = 1, \dots, r$ erklärt. Das geschätzte Modell:

$$\widehat{y_t - \bar{y}} = \sum_{i=1}^r \left(\hat{a}_i \cos(2\pi \lambda_i t) + \hat{b}_i \sin(2\pi \lambda_i t) \right).$$

Die Residuen:

$$\hat{\varepsilon}_t = y_t - \bar{y} - \sum_{i=1}^r \left(\hat{a}_i \cos(2\pi \lambda_i t) + \hat{b}_i \sin(2\pi \lambda_i t) \right).$$

Das Periodogramm

Das Regressionsmodell mit mehreren harmonischen Schwingungen

Für Fourierfrequenzen λ_i hat die OLS-Schätzung im Modell (5.2) eine besonders einfache Darstellung:

$$\hat{a}_i = 2C(\lambda_i), \quad \hat{b}_i = 2S(\lambda_i)$$

und

$$\begin{aligned} RSS &= \sum_{t=1}^N \hat{\varepsilon}_t^2 \\ &= \sum_{t=1}^N (y_t - \bar{y})^2 - 2N \sum_{i=1}^r \left(C(\lambda_i)^2 + S(\lambda_i)^2 \right) \\ &= \sum_{t=1}^N (y_t - \bar{y})^2 - 2 [I(\lambda_1) + I(\lambda_2) + \dots + I(\lambda_r)] \end{aligned}$$

Das Periodogramm

Das Regressionsmodell mit mehreren harmonischen Schwingungen

Tatsächlich lässt sich die Zeitreihe *Steuern* mit Hilfe der harmonischen Schwingungen mit den Frequenzen $\lambda_1 = 1/12$, $\lambda_2 = 1/6$, $\lambda_3 = 1/4$ und $\lambda_4 = 1/3$ gut approximieren. Das geschätzte Modell ist

$$\widehat{y_t - \bar{y}} = 1.55 \cos(2\pi\lambda_1 t) - 2.09 \sin(2\pi\lambda_1 t) + \\ 0.99 \cos(2\pi\lambda_2 t) + 2.14 \sin(2\pi\lambda_2 t) + \\ 0.61 \cos(2\pi\lambda_3 t) + 2.35 \sin(2\pi\lambda_3 t) + \\ 7.89 \cos(2\pi\lambda_4 t) + 0.57 \sin(2\pi\lambda_4 t).$$

Das Periodogramm

Das Regressionsmodell mit mehreren harmonischen Schwingungen

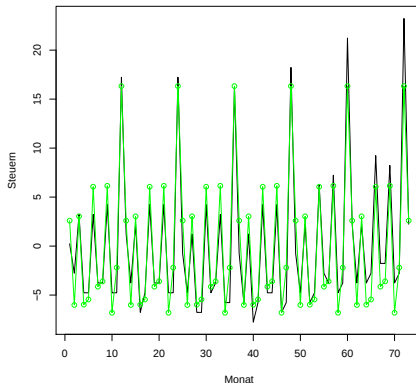


Abb. 5.3: Die harmonische Approximation für die Reihe *Steuern*

Das Periodogramm

Das Regressionsmodell mit mehreren harmonischen Schwingungen

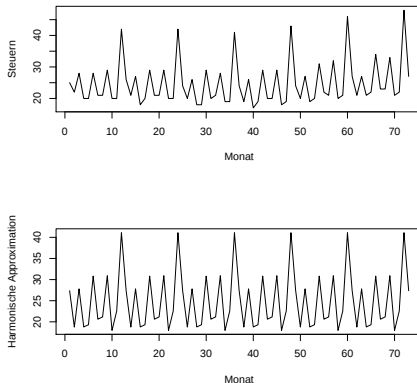


Abb. 5.4: Die harmonische Approximation für die Reihe *Steuern*

Das Periodogramm

Das Regressionsmodell mit mehreren harmonischen Schwingungen

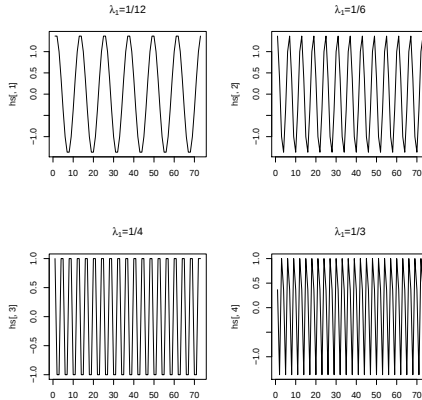


Abb. 5.5: Die Regressoren für die Approximation der Reihe *Steuern*

Probleme bei der Interpretation des Periodogramms

Im Folgenden betrachten wir drei Phänomene, welche die Interpretation des Periodogramms erschweren. Diese Phänomene bezeichnen wir als

1. Aliasing (Maskierung)
2. Leakage (Durchsickern)
3. Harmonics (Oberschwingungen)

Probleme bei der Interpretation des Periodogramms

Vorläufige Anmerkungen

- Wir bezeichnen eine Schwingung (oder allg. eine Periodizität) als
 - *kurzwellig*, wenn sie eine *kurze* Periode (kurze "Welle") und somit eine *hohe* Frequenz hat,
 - *langwellig*, wenn sie eine *lange* Periode (lange "Welle") und somit eine *niedrige* Frequenz hat.
- Sei (y_t) eine Zeitreihe mit einer Periodizität der Frequenz λ . Durch Anwendung der Periodogrammeigenschaften erhalten wir

$$I(\lambda) = I(k + \lambda), \quad \forall k \in \mathbb{Z}$$

und

$$I(\lambda) = I(-\lambda) = I(k - \lambda), \quad \forall k \in \mathbb{Z}.$$

Dementsprechend haben λ und ihre so genannten *Aliases*

$$k + \lambda, k - \lambda, \quad \forall k \in \mathbb{Z}$$

denselben Wert $I(\lambda)$ im Periodogramm. Diese Eigenschaft kann zu einer *Maskierung (Aliasing)* der tatsächlichen Periodizität der Frequenz λ führen.

Probleme bei der Interpretation des Periodogramms

Aliasing

- *Aliasing (Maskierung)*: Eine Zeitreihe kann nicht kontinuierlich erfasst werden, sondern sie wird an diskreten Zeitpunkten abgetastet. Daher können tatsächliche Periodizitäten des zugrundeliegenden stochastischen Prozesses oft nicht beobachtet werden. Stattdessen erscheinen sie als längerwellige Periodizitäten. Bspw. sei (y_t) eine Zeitreihe, die zu den Zeitpunkten $t = 1, 2, 3, \dots, N$ beobachtet wurde. In diesem Fall könnte eine nicht ganzzahlige Periodizität des zugrundeliegenden Prozesses mit Hilfe der Reihe (y_t) nicht aufgedeckt werden.

Probleme bei der Interpretation des Periodogramms

Aliasing

- Beispiel: Sei (Y_t) ein kontinuierlicher stochastischer Prozess mit einer Periodizität der Periode $P = 7/6$, bzw. $\lambda = 6/7$. Diese Periodizität kann in einer Reihe (y_t) mit $t = 1, 2, 3, \dots, N$ nicht erfasst werden. Sie hat jedoch die Aliases:

$$k + \lambda = \left\{ \dots, -\frac{15}{7}, -\frac{8}{7}, -\frac{1}{7}, \frac{6}{7}, \frac{13}{7}, \frac{20}{7}, \frac{27}{7}, \dots \right\}$$

$$k - \lambda = \left\{ \dots, -\frac{27}{7}, -\frac{20}{7}, -\frac{13}{7}, -\frac{6}{7}, \frac{1}{7}, \frac{8}{7}, \frac{15}{7}, \dots \right\}.$$

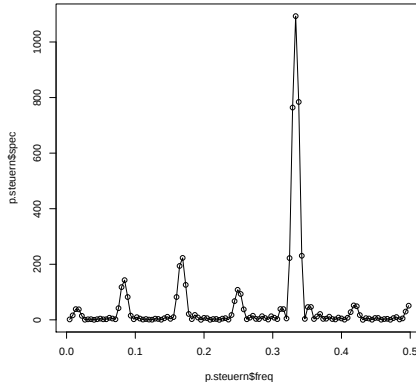
Daher wird uns die Periodizität der Frequenz $\lambda = 6/7$ als langwellige Periodizität mit $\lambda = 1/7$, bzw. $P = 7$ erscheinen.

- *Hauptalias (principal alias)* nennen wir den Alias im Intervall $[0, 0.5]$. Im vorigen Beispiel ist $\lambda = 1/7$ der Hauptalias.

Probleme bei der Interpretation des Periodogramms

Leakage

- *Leakage (Durchsickern)*: bei Vorliegen einer Periodizität der Frequenz λ in der Zeitreihe (y_t) ist der Wert des Periodogramms nicht nur an der Stelle λ , sondern auch in einer Umgebung von λ erhöht.



Probleme bei der Interpretation des Periodogramms

Oberschwingungen

- *Das Auftauchen von Oberschwingungen (harmonics)*: Sei (y_t) eine Zeitreihe mit *einer* Periodizität der Frequenz λ . Das Periodogramm von (y_t) wird einen Peak an der Stelle λ aufweisen, und zusätzlich auch an den Stellen $2\lambda, 3\lambda, 4\lambda, \dots$. Dabei müssen die Frequenzen $2\lambda, 3\lambda, 4\lambda, \dots$ keine eigentlichen Periodizitäten von (y_t) darstellen.

Probleme bei der Interpretation des Periodogramms

Oberschwingungen

- Die harmonischen Schwingungen mit den Frequenzen $2\lambda, 3\lambda, 4\lambda, \dots$ werden als *Oberschwingungen* der harmonischen Schwingung mit der Frequenz λ bezeichnet.
- Sei (y_t) eine Zeitreihe mit einer Periodizität der Frequenz λ , bzw. mit der Periode $P = 1/\lambda$. Das Auftauchen von Oberschwingungen im Periodogramm von (y_t) ist darauf zurückzuführen, dass die harmonische Schwingung mit der Frequenz λ und ihre Oberschwingungen eine gemeinsame Periodizität mit der Periode $P = 1/\lambda$ haben. Daher werden sowohl die harmonische Schwingung der Frequenz λ als auch ihre Oberschwingungen einen Beitrag zu der Erklärung der Reihe (y_t) im linearen Modell leisten.

Probleme bei der Interpretation des Periodogramms

Oberschwingungen

Der Datensatz enthält die monatliche Bierproduktion in Australien, im Zeitraum Januar 1956 – August 1995.

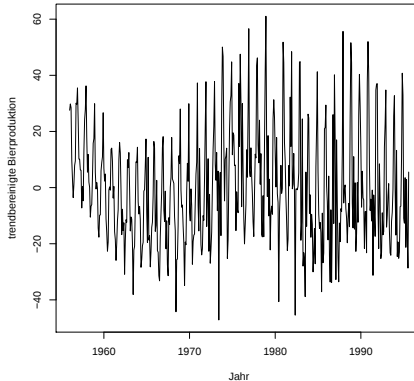


Abb. 5.7: Die Zeitreihe *Bier*

Probleme bei der Interpretation des Periodogramms

Oberschwingungen

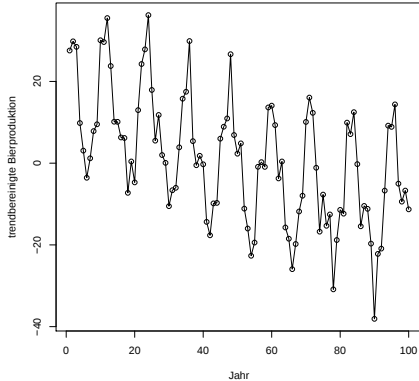


Abb. 5.8: Die Periodizität der Reihe *Bier*

Probleme bei der Interpretation des Periodogramms

Oberschwingungen

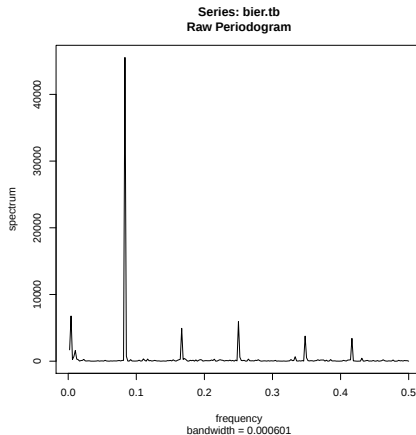


Abb. 5.9: Das Auftauchen von Oberschwingungen dargestellt an der Reihe *Bier*

Kapitel 5:

Statistische Methoden im Frequenzbereich

5.2 Fouriertransformationen für Zeitreihen

Fouriertransformation

Definition (Fouriertransformierte)

Sei (a_t) eine absolut summierbare Folge. Die **Fouriertransformierte** von (a_t) ist die Funktion der Frequenz $\lambda \in \mathbb{R}$

$$F(\lambda) = \sum_{t=-\infty}^{+\infty} a_t e^{i2\pi\lambda t}. \quad (5.3)$$

Die Transformation von (a_t) in $F(\lambda)$ heißt Fouriertransformation.

Theorem (Inverse Fouriertransformierte)

Sei (a_t) eine absolut summierbare Folge mit der Fouriertransformierten $F(\lambda)$. Es gilt die folgende Umkehrung der Fouriertransformation für jedes $t \in \mathbb{Z}$

$$a_t = \int_{-0.5}^{0.5} F(\lambda) e^{-i2\pi\lambda t} d\lambda.$$

Die Transformation von $F(\lambda)$ in (a_t) heißt inverse Fouriertransformation.

Das Amplitudengramm (Fouriertransformationen für Zeitreihen) Das Operandum (diskrete Fall) im Frequenzbereich

- Die Darstellung einer Folge im „Zeitbereich“ anhand (a_t) ist äquivalent zu ihrer Darstellung im „Frequenzbereich“ anhand $F(\lambda)$.
- Die Fouriertransformierte $F(\lambda)$ kann als „Bild“ der Folge (a_t) im „Frequenzbereich“ interpretiert werden.

Hinweis: Die Fouriertransformation kann mit dem Algorithmus der schnellen Fouriertransformation (Fast Fourier Transform, FFT) zeiteffizient durchgeführt werden.

Fouriertransformation

Eigenschaften der Fouriertransformierten

1. $F(0) = \sum_{t=-\infty}^{+\infty} a_t$, d.h. $F(0)$ ist die Summe der Folgeglieder.
2. $F(-\lambda) = \overline{F(\lambda)}$, d.h. $F(-\lambda)$ ist die komplex Konjugierte von $F(\lambda)$.
3. $F(\lambda + 1) = F(\lambda)$, d.h. $F(\lambda)$ ist periodisch mit der Periode 1.

Aufgrund der Eigenschaften 2. und 3. ist es ausreichend, die Fouriertransformierte $F(\lambda)$ im Intervall $[0, 0.5]$ zu betrachten.

In der Zeitreihenanalyse sind die folgenden Fouriertransformierten wichtig:

- die Fouriertransformierte $F(\lambda)$ einer Zeitreihe (y_t)

und

- die Fouriertransformierte $I(\lambda)$ der empirischen Autokovarianzfunktion (c_τ) .

Theorem (Zusammenhang zwischen $c(\tau)$ und $I(\lambda)$)

Sei (y_t) eine Zeitreihe mit der empirischen Autokovarianzfunktion (c_τ) . Zwischen der Autokovarianzfunktion (c_τ) und dem Periodogramm $I(\lambda)$ besteht der folgende Zusammenhang

$$I(\lambda) = c_0 + 2 \sum_{\tau=1}^{N-1} c_\tau \cos(2\pi\lambda\tau)$$

oder, äquivalent dazu, in der komplexen Darstellung

$$I(\lambda) = \sum_{\tau=-(N-1)}^{N-1} c_\tau e^{i2\pi\lambda\tau}.$$

Dieser Satz besagt, dass das Periodogramm $I(\lambda)$ die Fouriertransformierte der Autokovarianzfunktion (c_τ) ist.

Theorem (Parseval'sches Theorem)

Sei (y_t) eine Zeitreihe mit der empirischen Autokovarianzfunktion (c_τ) und dem Periodogramm $I(\lambda)$. Mit Hilfe der Umkehrung der Fouriertransformation erhalten wir

$$c_\tau = \int_{-0,5}^{0,5} I(\lambda) e^{-i2\pi\lambda\tau} d\lambda.$$

Speziell für $\tau = 0$ erhalten wir die empirische Varianz der Zeitreihe (y_t)

$$c_0 = \int_{-0,5}^{0,5} I(\lambda) d\lambda. \quad (5.4)$$

Gleichung (5.4) entspricht einer Verteilung der empirischen Varianz c_0 der Reihe (y_t) auf Schwingungskomponenten mit Frequenzen $-0.5 \leq \lambda \leq 0.5$.

Schlittgen und Streitberg (1999), Zeitreihenanalyse: Kapitel 1.6

Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 6.1, 6.2, 6.3

Kapitel 5:

Statistische Methoden im Frequenzbereich

5.3 Das Spektrum

Die Spektraldichtefunktion

Das theoretische Gegenstück zum Periodogramm $I(\lambda)$ ist die Spektraldichtefunktion $f(\lambda)$.

Definition

Sei (Y_t) ein stationärer stochastischer Prozess mit absolut summierbarer Autokovarianzfunktion (γ_τ) . Die Spektraldichtefunktion oder kurz das Spektrum $f(\lambda)$ von (Y_t) ist die Fouriertransformierte der Autokovarianzfunktion

$$f(\lambda) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau e^{i2\pi\lambda\tau} = \gamma_0 + 2 \sum_{\tau=1}^{\infty} \gamma_\tau \cos(2\pi\lambda\tau).$$

Beispiel: Das Spektrum eines White-Noise-Prozesses (ε_t) mit Varianz $\text{Var}[\varepsilon_t] = \sigma^2$ ist konstant

$$f(\lambda) = \sigma^2.$$

Die Spektraldichtefunktion

Eigenschaften der Spektraldichtefunktion

1. $f(0) = \sum_{\tau=-\infty}^{\infty} \gamma_{\tau}$
2. $f(-\lambda) = \overline{f(\lambda)}$
3. $f(\lambda + 1) = f(\lambda)$
4. $f(\lambda) \geq 0$
5. $f(\lambda)$ stetig

Aufgrund der Eigenschaften 2. und 3. ist es ausreichend, das Spektrum im Intervall $[0, 0.5]$ zu betrachten.

Theorem (Parseval'sches Theorem für die theoretische Autokovarianzfunktion)

Sei (Y_t) ein stationärer stochastischer Prozess mit der Autokovarianzfunktion (γ_τ) und der Spektraldichtefunktion $f(\lambda)$. Mit Hilfe der Umkehrung der Fouriertransformation erhalten wir

$$\gamma_\tau = \int_{-0,5}^{0,5} f(\lambda) e^{-i2\pi\lambda\tau} d\lambda.$$

Speziell für $\tau = 0$ erhalten wir die Prozessvarianz γ_0

$$\gamma_0 = \int_{-0,5}^{0,5} f(\lambda) d\lambda. \tag{5.5}$$

Gleichung (5.5) entspricht einer Zerlegung der Prozessvarianz γ_0 in Frequenzanteile.

Kapitel 5:

Statistische Methoden im Frequenzbereich

5.4 Lineare Filter im Frequenzbereich

Theorem

Sei (X_t) ein stationärer Prozess mit absolut summierbarer Autokovarianzfunktion $\gamma_X(\tau)$ und Spektraldichtefunktion $f_X(\lambda)$. Sei (Y_t) der Output eines absolut summierbaren linearen Filters (a_u) angewendet auf (X_t)

$$Y_t = \sum_u a_u X_{t-u}.$$

Dann hat (Y_t) das Spektrum

$$f_Y(\lambda) = |F_a(\lambda)|^2 f_X(\lambda)$$

mit $F_a(\lambda)$ die Fouriertransformierte der Folge (a_u) .

Die Fouriertransformierte $F_a(\lambda)$ eines linearen Filters (a_u) heißt Transferfunktion. $|F_a(\lambda)|$ ist die so genannte Gainfunktion des Filters:

$$G_a(\lambda) = |F_a(\lambda)|.$$

Spektrum autoregressiver Prozesse

Sei (ε_t) ein White-Noise-Prozess mit Varianz $\text{Var}[\varepsilon_t] = \sigma^2$ und (Y_t) ein AR $[p]$ -Prozess

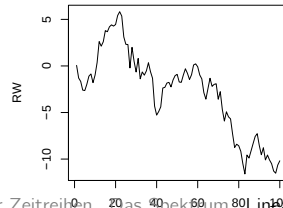
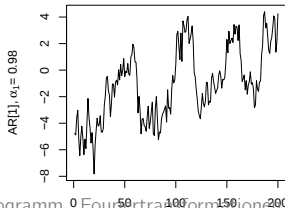
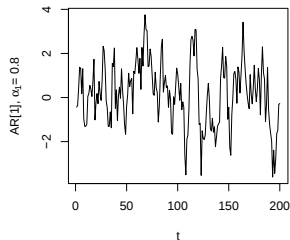
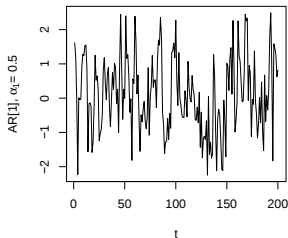
$$\alpha(B) Y_t = \varepsilon_t, \quad \alpha(B) = 1 - \alpha_1 B - \dots - \alpha_p B^p.$$

Die Spektraldichtefunktion des AR $[p]$ -Prozesses ist

$$f_Y(\lambda) = \frac{\sigma^2}{\left| 1 - \sum_{u=1}^p \alpha_u e^{i2\pi\lambda u} \right|^2}.$$

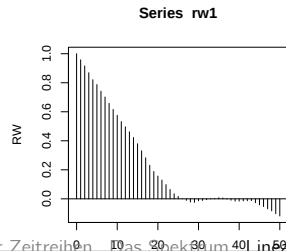
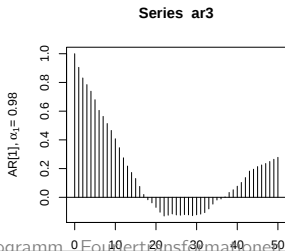
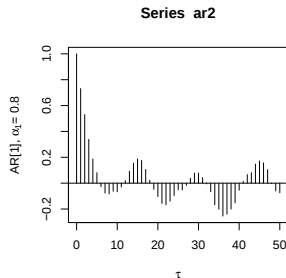
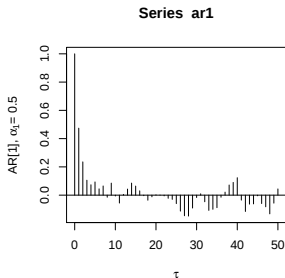
Spektrum autoregressiver Prozesse

Der Übergang von einem stationären AR[1]-Modell zu einem Random-Walk



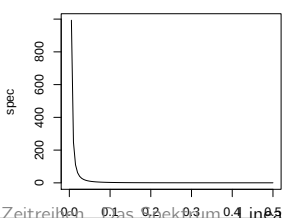
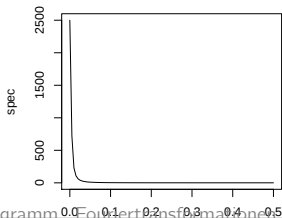
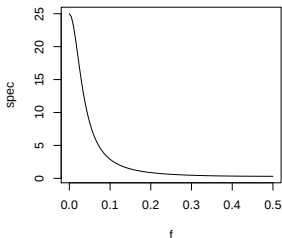
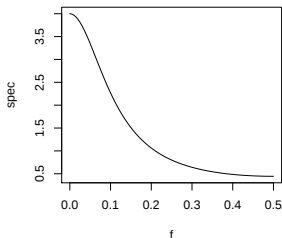
Spektrum autoregressiver Prozesse

Der Übergang von einem stationären AR[1]-Modell zu einem Random-Walk

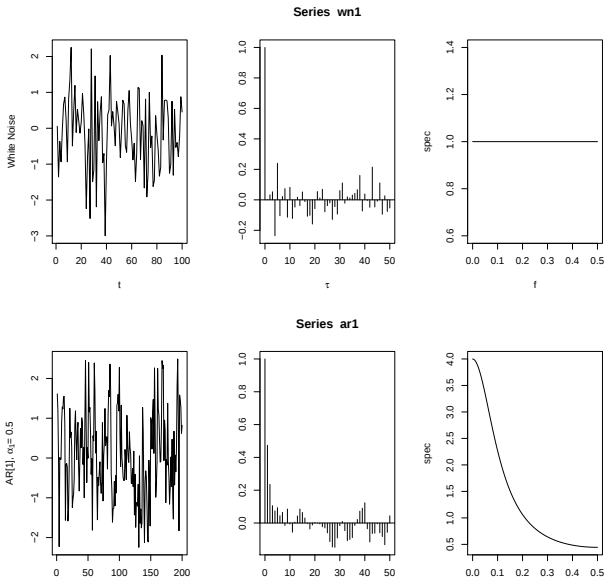


Spektrum autoregressiver Prozesse

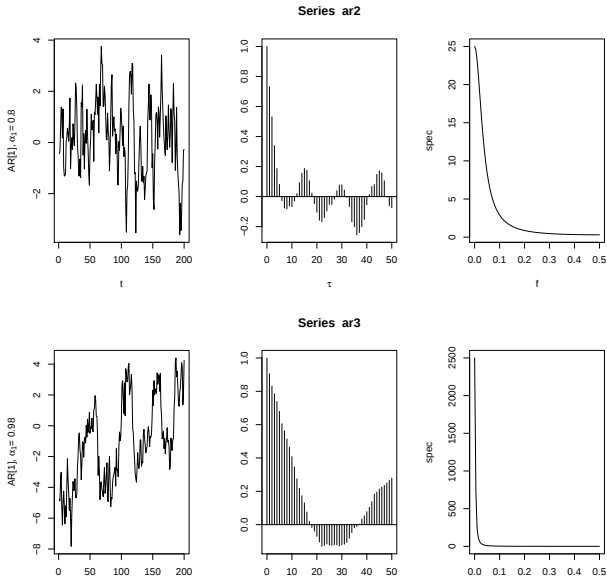
Der Übergang von einem stationären AR[1]-Modell zu einem Random-Walk



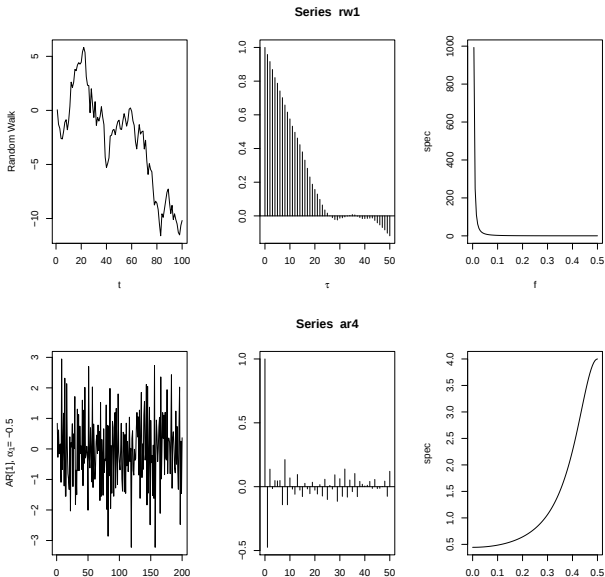
Spektrum autoregressiver Prozesse



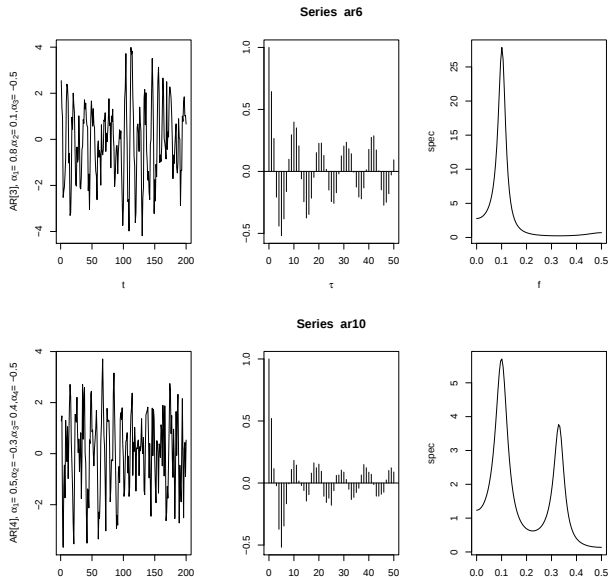
Spektrum autoregressiver Prozesse



Spektrum autoregressiver Prozesse



Spektrum autoregressiver Prozesse



Spektrum MA[q]-Prozesse

Sei (ε_t) ein White-Noise-Prozess mit Varianz $\text{Var}[\varepsilon_t] = \sigma^2$ und (Y_t) ein MA[q]-Prozess

$$Y_t = \beta(B) \varepsilon_t, \quad \beta(B) = 1 - \beta_1 B - \dots - \beta_q B^q.$$

Die Spektraldichtefunktion des MA[q]-Prozesses lautet

$$f_Y(\lambda) = \left| 1 - \sum_{u=1}^q \beta_u e^{i2\pi\lambda u} \right|^2 \sigma^2.$$

Spektrum ARMA[p, q]-Prozesse

Sei (ε_t) ein White-Noise-Prozess mit Varianz $\text{Var}[\varepsilon_t] = \sigma^2$ und (Y_t) ein ARMA[p, q]-Prozess

$$\alpha(B) Y_t = \beta(B) \varepsilon_t,$$

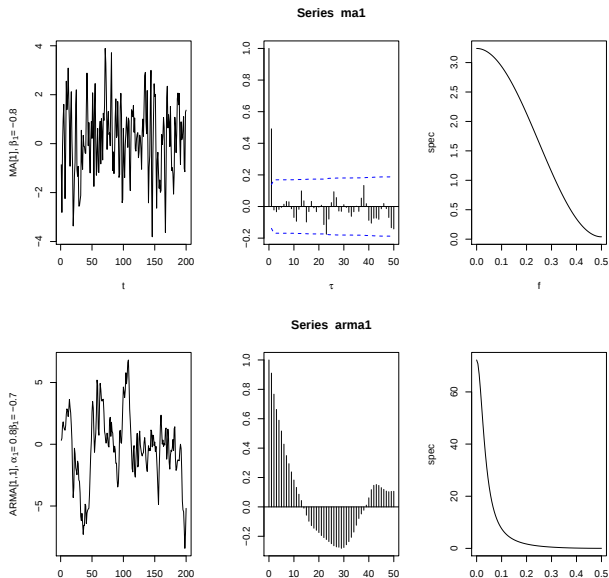
mit

$$\alpha(B) = 1 - \alpha_1 B - \dots - \alpha_p B^p, \quad \beta(B) = 1 - \beta_1 B - \dots - \beta_q B^q.$$

Die Spektraldichtefunktion des ARMA[p, q]-Prozesses lautet

$$f_Y(\lambda) = \frac{\left| 1 - \sum_{u=1}^q \beta_u e^{i2\pi\lambda u} \right|^2}{\left| 1 - \sum_{u=1}^p \alpha_u e^{i2\pi\lambda u} \right|^2} \sigma^2.$$

Spektrum MA[q]- und ARMA[p, q]-Prozesse



Der Differenzenfilter im Frequenzbereich

Sei (Y_t) ein stochastischer Prozess und (ΔY_t) seine ersten Differenzen

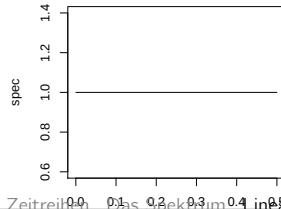
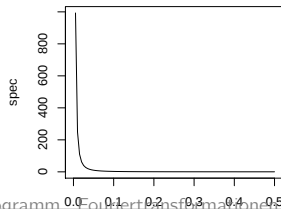
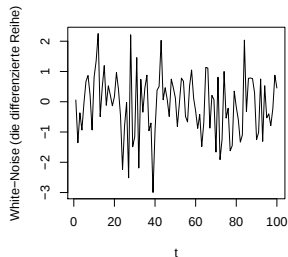
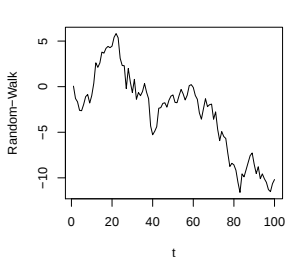
$$\Delta Y_t = Y_t - Y_{t-1}, \quad t = 2, 3, \dots, N.$$

Die Spektraldichtefunktion für die ersten Differenzen ist

$$f_{\Delta Y}(\lambda) = 4 \sin^2(\pi \lambda) f_Y(\lambda).$$

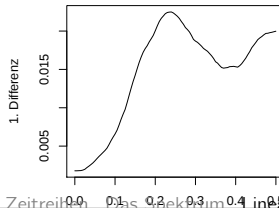
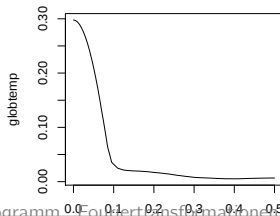
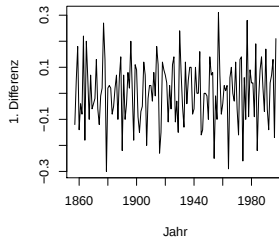
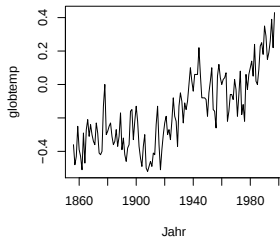
Der Differenzenfilter im Frequenzbereich

Die Auswirkung auf das Random-Walk-Modell



Der Differenzenfilter im Frequenzbereich

Die Auswirkung auf die Reihe *globtemp*



Kapitel 5:

Statistische Methoden im Frequenzbereich

5.5 Spektralschätzung

Gesucht ist ein konsistenter Schätzer für die Spektraldichtefunktion $f(\lambda)$. **Zur**

Erinnerung:

- Sei $\hat{\beta}_N$ ein allgemeiner Schätzer für einen Parameter β . Der Schätzer $\hat{\beta}_N$ heißt **konsistent**, falls

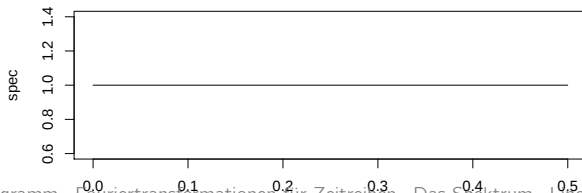
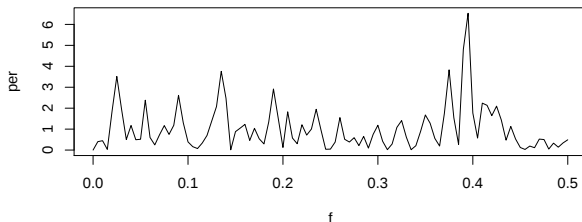
$$\lim_{N \rightarrow \infty} \mathbb{E} \left[\left(\hat{\beta}_N - \beta \right)^2 \right] = 0.$$

- Sei $\hat{\beta}_N$ ein asymptotisch erwartungstreuer Schätzer für β . Der Schätzer $\hat{\beta}_N$ ist dann und nur dann konsistent wenn gilt

$$\lim_{N \rightarrow \infty} \text{Var} \left[\hat{\beta}_N \right] = 0.$$

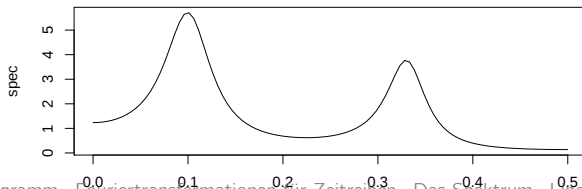
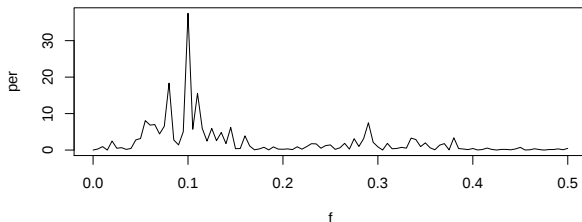
Nicht konsistente Spektralschätzung mit dem Periodogramm

Das Periodogramm zeigt ein sehr unruhiges Bild.



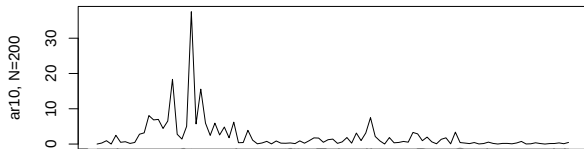
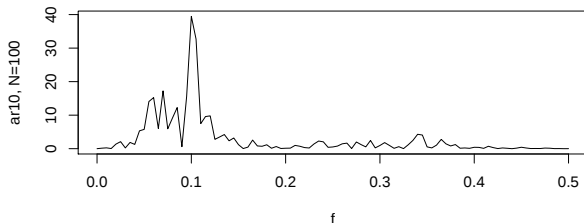
Nicht konsistente Spektralschätzung mit dem Periodogramm

Die Größenordnung und die Lage von Hauptpeaks werden falsch geschätzt.



Nicht konsistente Spektralschätzung mit dem Periodogramm

Die Schätzung durch das Periodogramm ist nicht besser, wenn N größer gewählt wird.



Theorem

Sei (Y_t) ein stochastischer Prozess mit dem Spektrum $f(\lambda)$ und $I(\lambda) = I_N(\lambda)$ das von einer Realisation (y_t) von (Y_t) der Länge N berechnete Periodogramm. Das Periodogramm $I(\lambda)$ ist asymptotisch erwartungstreu:

$$\lim_{N \rightarrow \infty} \mathbb{E}[I(\lambda)] = f(\lambda) \quad \text{für } \lambda \in (0, 0.5].$$

Das Periodogramm $I(\lambda)$ hat die asymptotische Varianz

$$\lim_{N \rightarrow \infty} \text{Var}[I(\lambda)] = \begin{cases} f^2(\lambda) & 0 < \lambda < 0.5 \\ 2f^2(\lambda) & \lambda = 0.5 \end{cases}.$$

Es gilt

$$\lim_{N \rightarrow \infty} \text{Cov}[I(\lambda), I(\alpha)] = 0 \quad \text{für } \lambda \neq \alpha.$$

Nicht konsistente Spektralschätzung mit dem Periodogramm

- Das Periodogramm $I(\lambda)$ ist asymptotisch erwartungstreu. Daher wäre das über mehrere Realisationen des zugrundeliegenden Prozesses gemittelte Periodogramm $\overline{I(\lambda)}$ ein zuverlässiger Schätzer der Spektraldichtefunktion.
- Die Varianz verschwindet für $N \rightarrow \infty$ nicht. Daher ist das Periodogramm $I(\lambda)$ kein konsistenter Schätzer für die Spektraldichtefunktion.
- Die Streuung des Periodogramms ist von der Größenordnung des Spektrums selbst. Daher sind die Schätzfehler teilweise erheblich.
- Das Spektrum $f(\lambda)$ ist eine stetige Funktion. Jedoch zeigt das Periodogramm $I(\lambda)$ ein sehr unruhiges Bild, wodurch falsche Peaks vorgetäuscht werden können. Das liegt an der Unkorreliertheit der Periodogrammmordinaten.

Insgesamt ist eine Beurteilung über die Periodizitäten in einer Zeitreihe mit Hilfe des Periodogramms unzuverlässig.

Theorem

Sei (Y_t) ein stationärer stochastischer Prozess, der sich als $MA[\infty]$ -Prozess $Y_t = \sum_{u=0}^{\infty} \beta_u \varepsilon_{t-u}$ darstellen lässt, wobei $\sum_{u=0}^{\infty} (1 + |u|) |\beta_u| < \infty$ gilt. Seien $f(\lambda)$ das Spektrum von (Y_t) und $I(\lambda) = I_N(\lambda)$ das von einer Realisation (y_t) von (Y_t) der Länge N berechnete Periodogramm. Wir betrachten die Fourierfrequenzen $\lambda = \lambda_k = k/N$, $k \in \mathbb{Z}$ im Intervall $0 \leq \lambda < 0.5$.

- Das Periodogramm $I(\lambda)$ ist asymptotisch exponentiell verteilt mit dem Erwartungswert $f(\lambda)$

$$I(\lambda) \xrightarrow{d} \text{Exp}[f(\lambda)].$$

- Für Fourierfrequenzen $\lambda \neq \alpha$ sind $I(\lambda)$ und $I(\alpha)$ stochastisch unabhängig.

Direkte Spektralschätzer

Für eine zuverlässige Schätzung der Spektraldichtefunktion wird das Periodogramm mit Hilfe eines *gleitenden Durchschnitts* geglättet.

Definition

Sei (y_t) eine Zeitreihe mit dem Periodogramm $I(\lambda)$. Wir bezeichnen als direkten Spektralschätzer $\hat{f}(\lambda)$ der Schätzer

$$\hat{f}(\lambda) = \sum_u \omega_u I\left(\lambda - \frac{u}{N}\right).$$

Die Folge der Gewichte $(\omega_u)_{u \in \mathbb{Z}}$ heißt diskretes Periodogrammfenster.

■ Beispiel: das Daniell-Fenster

$$\omega_u = \begin{cases} \frac{1}{2q+1}, & \text{für } -q \leq u \leq q \\ 0, & \text{sonst} \end{cases}.$$

Direkte Spektralschätzer

Asymptotische Eigenschaften

Theorem

Sei (Y_t) ein stationärer stochastischer Prozess, der sich als $MA[\infty]$ -Prozess $Y_t = \sum_{u=0}^{\infty} \beta_u \varepsilon_{t-u}$ darstellen lässt, wobei $\sum_{u=0}^{\infty} (1 + |u|) |\beta_u| < \infty$ gilt. Seien $f(\lambda)$ das Spektrum von (Y_t) und $\hat{f}(\lambda) = \hat{f}_N(\lambda)$,

$$\hat{f}(\lambda) = \sum_{u=-q}^q \omega_u I\left(\lambda - \frac{u}{N}\right)$$

ein von einer Realisation (y_t) von (Y_t) der Länge N berechneter direkter Spektralschätzer mit dem diskreten Periodogrammfenster (ω_u) , $\omega_u = 0$ für $|u| > q$. Wir betrachten die Fourierfrequenzen $\lambda = \lambda_k = k/N$, $k \in \mathbb{Z}$ im Intervall $0 \leq \lambda < 0.5$. Dann $\hat{f}(\lambda)$ hat

- den asymptotischen Erwartungswert

$$\lim_{N \rightarrow \infty} \mathbb{E}[\hat{f}(\lambda)] = f(\lambda) \sum \omega_u$$

Theorem

- *die asymptotische Varianz*

$$\lim_{N \rightarrow \infty} \text{Var} \left[\hat{f}(\lambda) \right] = f(\lambda)^2 \sum_u \omega_u^2.$$

- $\hat{f}(\lambda)$ ist die gewichtete Summe von asymptotisch unabhängigen Periodogrammwerten $I\left(\lambda - \frac{u}{N}\right)$ mit

$$I\left(\lambda - \frac{u}{N}\right) \xrightarrow{d} \text{Exp}[f(\lambda)].$$

Direkte Spektralschätzer

Asymptotische Eigenschaften

- Durch die Wahl der Gewichte (ω_u) können *asymptotisch erwartungstreu* und *konsistente* Schätzer $\hat{f}(\lambda)$ konstruiert werden. Es gibt insgesamt $b = 2q + 1$ Gewichte $(\omega_u)_{-q \leq u \leq q}$ ungleich Null, die bestimmt werden müssen.
- Direkte Spektralschätzer $\hat{f}(\lambda)$ sind asymptotisch erwartungstreu, wenn gilt

$$\sum_{u=-q}^q \omega_u = 1.$$

Beispiel: der Daniell-Schätzer.

Direkte Spektralschätzer

Asymptotische Eigenschaften

- Um konsistente direkte Spektralschätzer $\hat{f}(\lambda)$ zu erhalten, ist es wichtig, dass die Anzahl $b = 2q + 1$ der Gewichte ω_u ungleich Null eine Funktion des Umfangs N ist, $b = b_N$. Die Größe $b = b_N$ ist die so genannte Bandbreite des Periodogrammfensters. Direkte Spektralschätzer $\hat{f}(\lambda)$ sind asymptotisch erwartungstreu und konsistent für Periodogrammfenster (ω_u) mit

$$\lim_{N \rightarrow \infty} b_N = \infty \quad \text{und} \quad \lim_{N \rightarrow \infty} \frac{b_N}{N} = 0.$$

Direkte Spektralschätzer

Asymptotische Eigenschaften

Bspw. sind die direkten Spektralschätzer $\hat{f}(\lambda)$ mit einem der folgenden Periodogrammfenster und entsprechend gewählter Bandbreite $b_N = 2q_N + 1$ konsistent

- das Daniell-Fenster

$$\omega_u = \begin{cases} \frac{1}{2q_N+1} & \text{für } -q_N \leq u \leq q_N \\ 0 & \text{sonst} \end{cases}$$

- das Bartlett-Priestley-Fenster

$$\omega_u = \begin{cases} \frac{3q_N}{4q_N^2-1} \left(1 - \left(\frac{u}{q_N} \right)^2 \right) & \text{für } -q_N \leq u \leq q_N \\ 0 & \text{sonst} \end{cases}$$

Hinweis: Die beiden direkten Spektralschätzer $\hat{f}(\lambda)$ haben dieselben asymptotischen Eigenschaften. Jedoch unterscheiden sie sich in ihren Eigenschaften in *endlichen* Stichproben.

Direkte Spektralschätzer

Eigenschaften für endliche Stichproben

Definition

Sei $\hat{f}(\lambda)$ ein direkter Spektralschätzer mit dem Periodogrammfenster (ω_u) mit $\sum_u \omega_u = 1$. Wir bezeichnen als äquivalente Bandbreite (EBW) die Größe

$$EBW \left[\hat{f}(\lambda) \right] = \frac{1}{N \sum_u \omega_u^2}.$$

Interpretation: Die EBW gibt den Abstand zweier Peaks im Spektrum $f(\lambda)$ an, welche vom Schätzer $\hat{f}(\lambda)$ noch getrennt werden können.

- Schätzer mit geringer EBW haben eine hohe Auflösungsfähigkeit und somit einen geringen Bias in endlichen Stichproben. Die EBW ist ein Maß für den Bias in endlichen Stichproben.

Direkte Spektralschätzer

Eigenschaften für endliche Stichproben

Für festes N gilt

$$\text{Bandbreite} \cdot \text{Varianz} = \text{Konstante}$$

oder genauer dargestellt

$$EBW \left[\hat{f}(\lambda) \right] \cdot \text{Var} \left[\hat{f}(\lambda) \right] \approx \frac{f^2(\lambda)}{N}.$$

Das Zielkonflikt der Spektralschätzung: die Bandbreite (und somit der Bias) und die Varianz können nicht gleichzeitig minimiert werden.

- Nachteil des Daniell-Schätzers: er hat eine sehr geringe Varianz, dafür einen hohen Bias.
- Der direkte Bartlett-Priestley-Schätzer ist aufgrund seines ausgewogenen Verhältnisses zwischen Bias und Varianz zu bevorzugen.

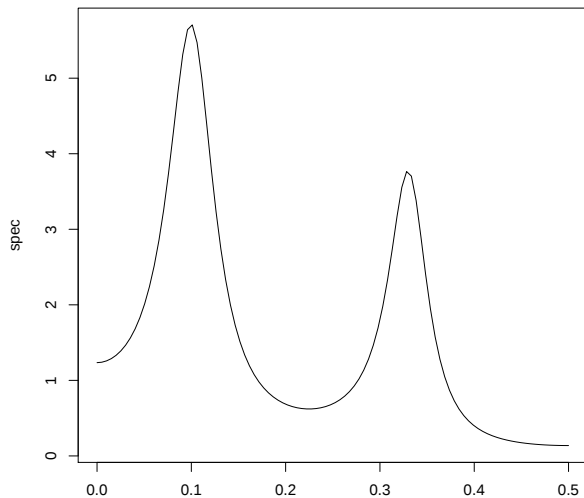
Direkte Spektralschätzer

Eigenschaften für endliche Stichproben

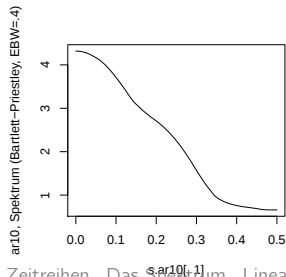
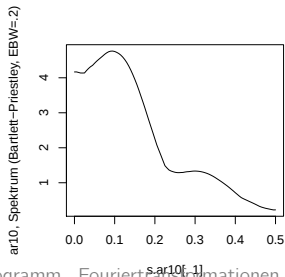
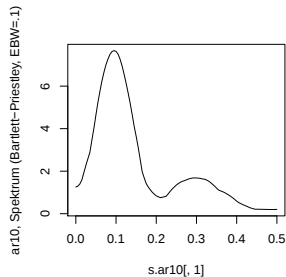
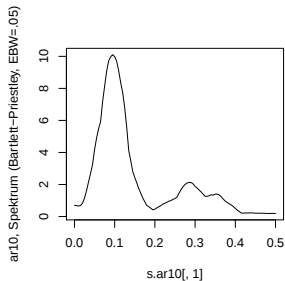
Schätzprozedur:

- Wir wählen ein diskretes Periodogrammfenster (ω_u) mit einem ausgewogenen Verhältnis zwischen Bias und Varianz in endlichen Stichproben aus, z.B. das Bartlett-Priestley-Periodogrammfenster.
- Wir schätzen das Spektrum $f(\lambda)$ für verschiedene EBW-Werte.
- Wir bestimmen den EBW-Wert in Abhängigkeit von den gewünschten Auflösung und Glattheit des geschätzten Spektrums $\hat{f}(\lambda)$.

Direkte Spektralschätzer



Direkte Spektralschätzer



Schlittgen und Streitberg (1999), Zeitreihenanalyse: Kapitel 3.1, 3.2, 3.3, 7.1, 7.4.1, 7.4.2

Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 6.4, 6.5

Literaturhinweise

Wiederholung trigonometrische Funktionen und komplexe Zahlen

Schlittgen und Streitberg (1999), Zeitreihenanalyse: Anhänge A, B

Direkte Spektralschätzer

Konfidenzintervall

Laut Satz (61) ist $\hat{f}(\lambda)$ die gewichtete Summe von asymptotisch exponentialverteilten und stochastisch unabhängigen Zufallsvariablen. Wir können eine Linearkombination von exponentialverteilten und stochastisch unabhängigen Zufallsvariablen mit Hilfe des s -fachen einer χ^2 -verteilten Zufallsvariablen nach dem *Verfahren von Satterthwaite* approximieren. Wir erhalten die Approximation

$$\lim_{N \rightarrow \infty} \hat{f}(\lambda) \approx sZ \quad \text{mit} \quad Z \sim \chi_r^2$$

mit

$$s = \frac{f(\lambda)}{2} \cdot \frac{\sum_u \omega_u^2}{\sum_u \omega_u}, \quad r = 2 \cdot \frac{(\sum_u \omega_u)^2}{\sum_u \omega_u^2},$$

bzw. für einen asymptotisch erwartungstreuen direkten Spektralschätzer $\hat{f}(\lambda)$

$$s = \frac{f(\lambda)}{2} \cdot \sum_u \omega_u^2, \quad r = \frac{2}{\sum_u \omega_u^2}.$$

Direkte Spektralschätzer

Konfidenzintervall

Sei (Y_t) ein stationärer stochastischer Prozess, der sich als $\text{MA}[\infty]$ -Prozess $Y_t = \sum_{u=0}^{\infty} \beta_u \varepsilon_{t-u}$ darstellen lässt, wobei $\sum_{u=0}^{\infty} (1 + |u|) |\beta_u| < \infty$ gilt. Seien $f(\lambda)$ das Spektrum von (Y_t) und $\hat{f}(\lambda)$ ein asymptotisch erwartungstreuer direkter Spektralschätzer für $f(\lambda)$. Sei $\lambda = k/N$, $k \in \mathbb{Z}$ eine Fourierfrequenz im Intervall $0 \leq \lambda < 0.5$. Wir konstruieren ein $(1 - \alpha) \times 100\%$ -Konfidenzintervall für $f(\lambda)$ wie folgt

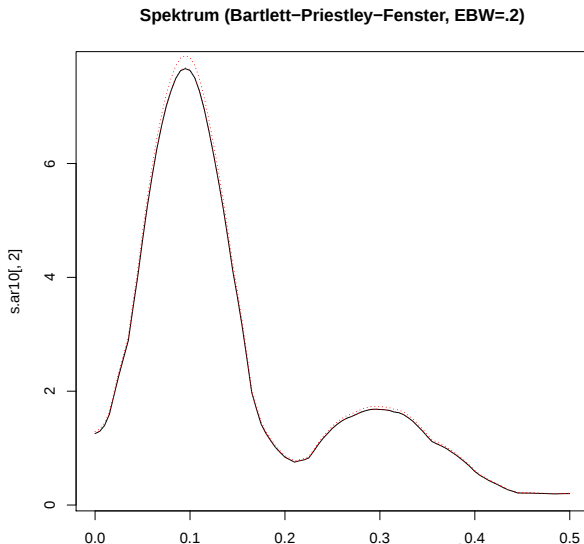
$$\left[\lim_{N \rightarrow \infty} \hat{f}(\lambda) \cdot \frac{r}{\chi_{1-\alpha/2, r}^2}, \quad \lim_{N \rightarrow \infty} \hat{f}(\lambda) \cdot \frac{r}{\chi_{\alpha/2, r}^2} \right]$$

mit

$$r = \frac{2}{\sum_u \omega_u^2}.$$

Direkte Spektralschätzer

Konfidenzintervall



Seien (Y_t) ein stationärer stochastischer Prozess mit der Spektraldichtefunktion $f(\lambda)$ und (y_t) eine Realisation von (Y_t) . Laut Satz (53) besteht der folgende Zusammenhang zwischen dem Periodogramm $I(\lambda)$ und der empirischen Autokovarianzfunktion (c_τ)

$$\begin{aligned} I(\lambda) &= c_0 + 2 \sum_{\tau=1}^{N-1} c_\tau \cos(2\pi\lambda\tau) \\ &= \sum_{\tau=-(N-1)}^{N-1} c_\tau \cos(2\pi\lambda\tau). \end{aligned} \tag{5.6}$$

- Ein direkter Spektralschätzer ist ein gewichteter Mittelwert des Periodogramms $I(\lambda)$

$$\hat{f}(\lambda) = \sum_u \omega_u I\left(\lambda - \frac{u}{N}\right)$$

mit Gewichten (diskretem Periodogrammfenster) (ω_u) .

- Alternativ können *indirekte Spektralschätzer* durch die Gewichtung der Autokovarianzen (c_τ) in (5.6) konstruiert werden

$$\hat{f}(\lambda) = \sum_{\tau=-(N-1)}^{N-1} \kappa_\tau c_\tau \cos(2\pi\lambda\tau).$$

Die Gewichte (κ_τ) bilden das so genannte *Lagfenster*.

Indirekte Spektralschätzer

Der Newey-West-Schätzer

Wir betrachten das lineare Regressionsmodell

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{u} \quad (5.7)$$

mit autokorrelierten Störungen

$$\text{Cov}[\mathbf{u}] = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{N-1} \\ \gamma_1 & \gamma_0 & & \gamma_{N-2} \\ \vdots & & \ddots & \vdots \\ \gamma_1 & \gamma_2 & \cdots & \gamma_0 \end{pmatrix}$$

Sei $\hat{\boldsymbol{\theta}}_N$ der OLS-Schätzer im Modell (5.7). Die asymptotische Kovarianzmatrix von $\hat{\boldsymbol{\theta}}_N$

$$\lim_{N \rightarrow \infty} \text{Cov}[\hat{\boldsymbol{\theta}}_N] = \lim_{N \rightarrow \infty} \frac{1}{N} \left(\frac{1}{N} \mathbf{X}'\mathbf{X} \right)^{-1} \frac{1}{N} \mathbf{X}'\mathbb{E}[\mathbf{u}\mathbf{u}'] \mathbf{X} \left(\frac{1}{N} \mathbf{X}'\mathbf{X} \right)^{-1}$$

kann mit Hilfe des Newey-West-Schätzers für die Größe

$$\lim_{N \rightarrow \infty} J_N = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{X}'\mathbb{E}[\mathbf{u}\mathbf{u}'] \mathbf{X}$$

Indirekte Spektralschätzer

Der Newey-West-Schätzer

Es gilt

$$J_N = \Gamma_0 + \sum_{\tau=1}^{N-1} \frac{N-\tau}{N} (\Gamma_\tau + \Gamma_{-\tau})$$

mit Γ_τ die Autokovarianzmatrix vom Lag τ für den multivariaten stochastischen Prozess $(\mathbf{x}'_t u_t)_t$

$$\Gamma_\tau = \mathbb{E} \left[\mathbf{x}'_t u_t (\mathbf{x}'_{t+\tau} u_{t+\tau})' \right].$$

Dann existiert der Grenzwert

$$\lim_{N \rightarrow \infty} J_N = \sum_{\tau=-\infty}^{\infty} \Gamma_\tau = f(\mathbf{0}),$$

mit $f(\mathbf{0})$ die spektrale Dichtematrix von $(\mathbf{x}'_t u_t)_t$ an der Stelle Null.

Indirekte Spektralschätzer

Der Newey-West-Schätzer

Der Newey-West-Schätzer ist der indirekte Spektralschätzer

$$\hat{f}(\mathbf{0}) = \hat{\Gamma}_0 + \sum_{\tau=1}^{N-1} \left(\hat{\Gamma}_{\tau} + \hat{\Gamma}'_{\tau} \right) \kappa_{\tau}$$

mit Bartlett-Lagfenster

$$\kappa_{\tau,N} = \begin{cases} 1 - \frac{\tau}{b_N+1}, & 1 \leq \tau \leq b_N + 1 \\ 0, & \tau > b_N + 1 \end{cases}$$

und Bandbreite des Lagfensters $b_N \geq 0$.

Schätzprozedur:

- Anpassung eines ARMA[p, q]-Modells

$$\alpha(B) Y_t = \beta(B) \varepsilon_t$$

an die Realisation (y_t) von (Y_t) und Schätzung der Modellparameter $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ und der Varianz des Störterms σ^2 .

- Den parametrischen Spektralschätzer $\hat{f}(\lambda)$ erhalten wir durch das Einsetzen der Schätzwerte $\hat{\alpha}_1, \dots, \hat{\alpha}_p, \hat{\beta}_1, \dots, \hat{\beta}_q$ und $\hat{\sigma}^2$ in die theoretische Spektraldichtefunktion für den ARMA[p, q]-Prozess

$$\hat{f}(\lambda) = \frac{\left| 1 - \sum_{u=1}^q \hat{\beta}_u e^{i2\pi\lambda u} \right|^2}{\left| 1 - \sum_{u=1}^p \hat{\alpha}_u e^{i2\pi\lambda u} \right|^2} \hat{\sigma}^2.$$

Schlittgen und Streitberg (1999), Zeitreihenanalyse: Kapitel 7.2, 7.4.2
Schlittgen (2015), Angewandte Zeitreihenanalyse mit R: Kapitel 6.5