



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Master-Modul 85-011: Zeitreihenanalyse

Sommersemester 2018
Hamburg

Dr. Cristina Sattarhoff

Inhaltsverzeichnis

Inhaltsverzeichnis	i		
0 Organisation und Hinweise	ii		
1 Stochastische Prozesse	1	2 Das klassische Komponentenmodell	16
1.1 Darstellung von Zeitreihen . . .	1	2.1 Die Trendkomponente	20
1.2 Schätzung der theoretischen Momente	8	2.1.1 Polynomialer Trend . . .	20
1.2.1 Stationarität	10	2.1.2 Die glatte Komponente	22
		2.2 Die Saisonkomponente	24
		1.2.2 Asymptotische Verteilungen	11
		1.2.3 Ergodizität	14

0 Organisation und Hinweise

Materialien

1

Die Materialien zur Vorlesung und Übung können unter

<https://www.bwl.uni-hamburg.de/statistik/studium/sose2018/zeitreihenanalyse.html>

heruntergeladen werden.

Vorlesungszeiten

2

Die Master-Vorlesung

Zeitreihenanalyse und Prognoserechnung

besteht aus 3 SWS Vorlesung und 1 SWS Übung.

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

Die ersten drei Übungen finden mittwochs 9:15 – 10:00 im Hörsaal Wiwi B1 statt. Ab dem 2. Mai werden die Übungen mittwochs 9:15 – 10:00 und bzw. 10:15 – 11:00 Uhr im Computerraum Wiwi 2043/2047 durchgeführt werden.

Literatur

3

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

4

- 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.

Klausur

5

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	Do	19.07.2018	14:15 - 15:45	Wiwi A
Ferientermin	Fr	21.09.2018	14:15 - 15:45	Wiwi A

Vorwissen

6

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

Kontakt

7

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

1 Stochastische Prozesse

1.1 Darstellung von Zeitreihen

Darstellung von Zeitreihen ■ Fallbeispiel Anzahlen von Luchspelzen

1.1

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

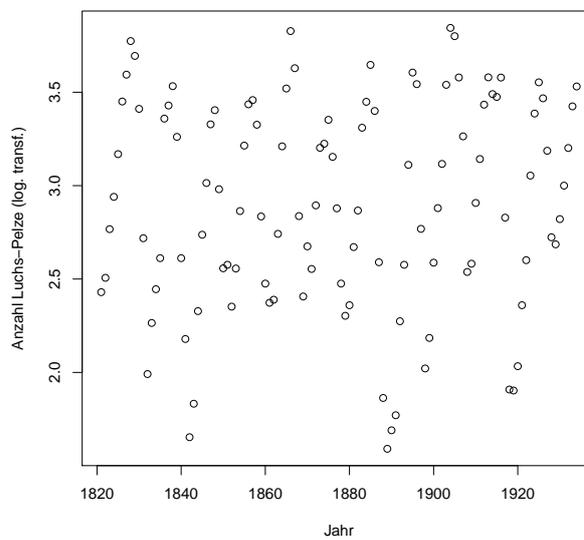


Abb. 1.1: Die Zeitreihe *Luchs* (Darstellung mit Punkten)

Darstellung von Zeitreihen ■ Fallbeispiel Anzahlen von Luchspelzen

1.2

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

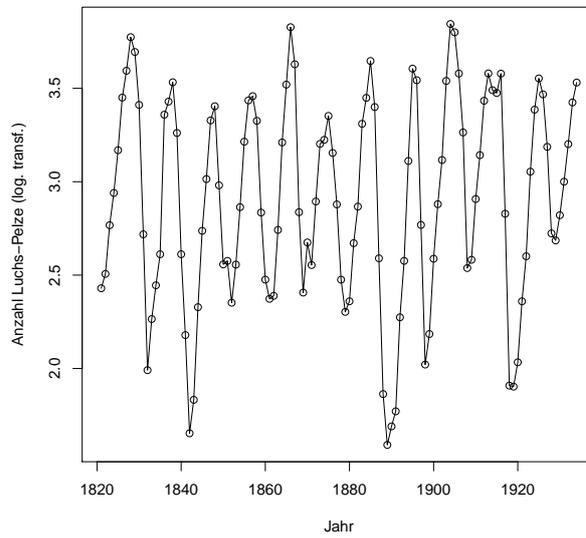


Abb. 1.2: Die Zeitreihe *Luchs* (Darstellung mit Punkten und Linien)

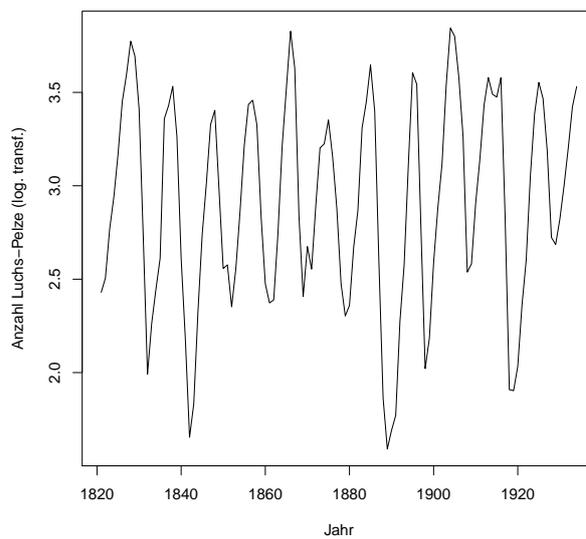


Abb. 1.3: Die Zeitreihe *Luchs* (Darstellung mit Linien)

Darstellung von Zeitreihen ■ Fallbeispiel Globaltemperatur

1.3

Der Datensatz stammt aus Shumway und Stoffer (2011).

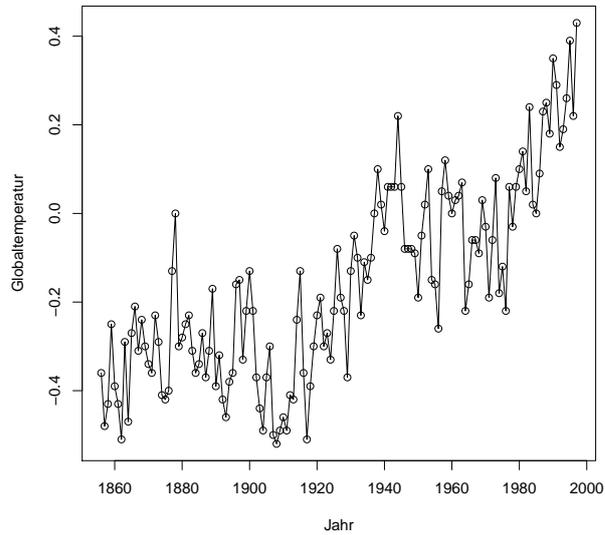


Abb. 1.4: Die Zeitreihe *globtemp*

Darstellung von Zeitreihen ■ Fallbeispiel Steuern

1.4

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

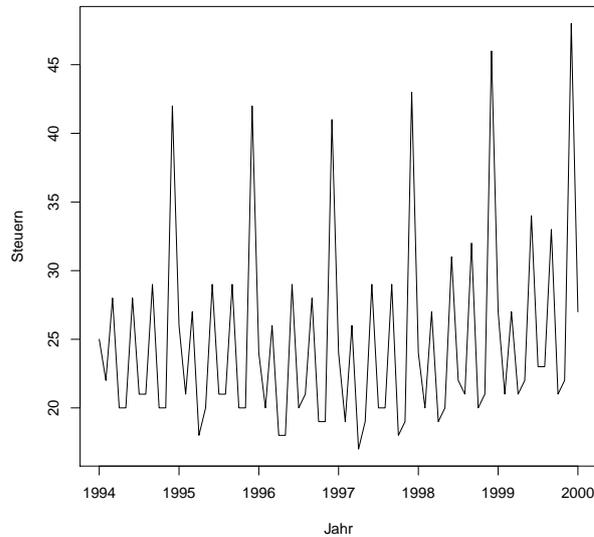


Abb. 1.5: Die Zeitreihe *steuern*

Darstellung von Zeitreihen

1.5

Definition 1 (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 2 (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) .

Zur Erinnerung: Eine Folge $(a_n)_{n \in \mathbb{N}}$ ist eine Abbildung $\mathbb{N} \rightarrow \mathbb{R}$. Jedem $n \in \mathbb{N}$ wird also ein $a_n \in \mathbb{R}$ zugeordnet.

Darstellung von Zeitreihen

1.6

Definition 3 (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.

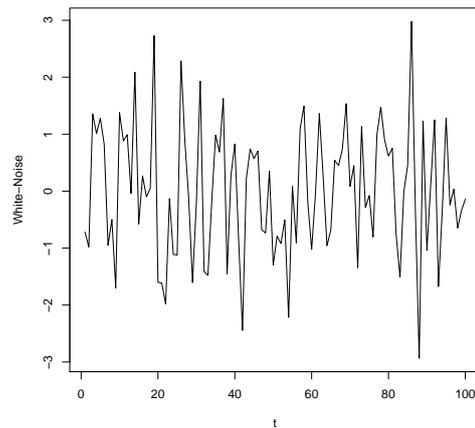


Abb. 1.6: Eine Realisation eines White-Noise-Prozesses

Darstellung von Zeitreihen

1.7

Definition 4 (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

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

Wir sprechen von einem Random-Walk ohne Drift, wenn $\mathbb{E}[\varepsilon_t] = 0$ gilt, und von einem Random-Walk mit Drift sonst.

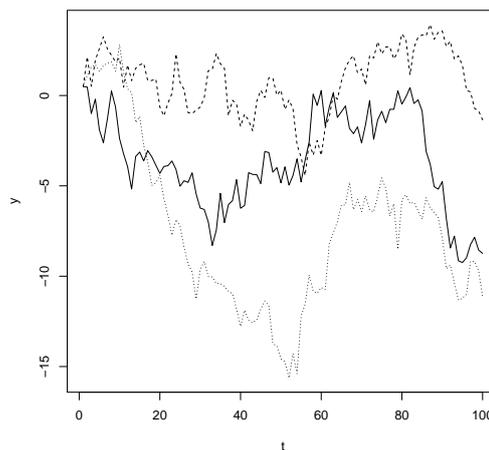


Abb. 1.7: Drei Realisationen eines Random-Walk-Prozesses

Empirische Momente

1.8

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

1.9

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

1.10

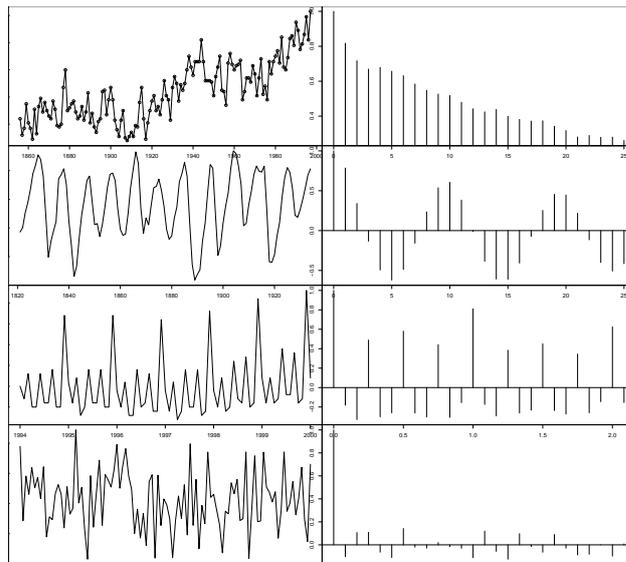


Abb. 1.8: Zeitreihen und ihre ACFs

Theoretische Momente

1.11

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

1.12

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. $\rho_\tau = \rho_{-\tau}$.

1.2 Schätzung der theoretischen Momente

Schätzung der theoretischen Momente ■ Schätzfunktion

1.13

Sei Y eine Zufallsvariable mit dem unbekanntem 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

1.14

► Klein-Stichproben-Eigenschaften: ¹

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

Schätzung der theoretischen Momente ■ Schätzfunktion

1.15

► Asymptotische Eigenschaften: ²

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

¹Siehe Anhang A in Stahlecker, P., Kröh, P. und M. Breitig (2018), *Eine Einführung in das lineare Modell*, Skript, Version 2.0, Universität Hamburg, Hamburg.

²Siehe Anhang A in Stahlecker, P., Kröh, P. und M. Breitig (2018), *Eine Einführung in das lineare Modell*, Skript, Version 2.0, Universität Hamburg, Hamburg.

- **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

1.16

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

1.17

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

Wir benötigen zwei Annahmen:

1. Stationarität

und

2. Ergodizität

1.2.1 Stationarität

Stationarität

1.18

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.

Stationarität

1.19

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

1.2.2 Asymptotische Verteilungen

Asymptotische Verteilungen ■ Das arithmetische Mittel

1.20

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

1.21

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 5. 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

1.22

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 6. (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

1.23

Theorem 7. (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

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

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

Asymptotische Verteilungen ■ Die empirische Autokorrelationsfunktion

1.24

Theorem 8. (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

1.25

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

1.26

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 $0.05 \cdot 25 = 1.25$ 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

1.27

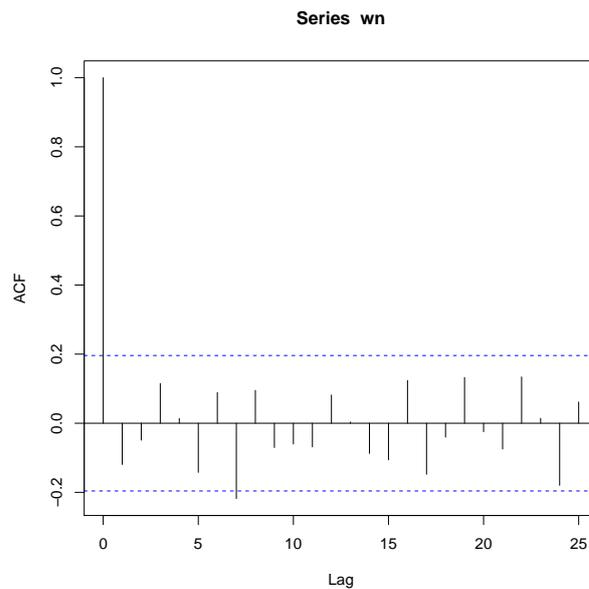


Abb. 1.9: Die ACF für die Realisation eines White-Noise-Prozesses

Asymptotische Verteilungen ■ White-Noise – Schwankungsintervall für die empirische Autokorrelation

1.28

```
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)
```

1.2.3 Ergodizität

Ergodizität

1.29

(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

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

Ergodizität

1.30

Theorem 9. $(Y_t)_{t \in \mathbb{Z}}$ 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

1.31

Theorem 10. 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.

Literaturhinweise

1.32

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

Literaturhinweise ■ Einführung in R

1.33

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

Literaturhinweise ■ Wiederholung der statistischen Grundlagen

1.34

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

2 Das klassische Komponentenmodell

Das klassische Komponentenmodell ■ Überprüfung der Stationaritätsannahme

2.1

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

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

2.2

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

2.3

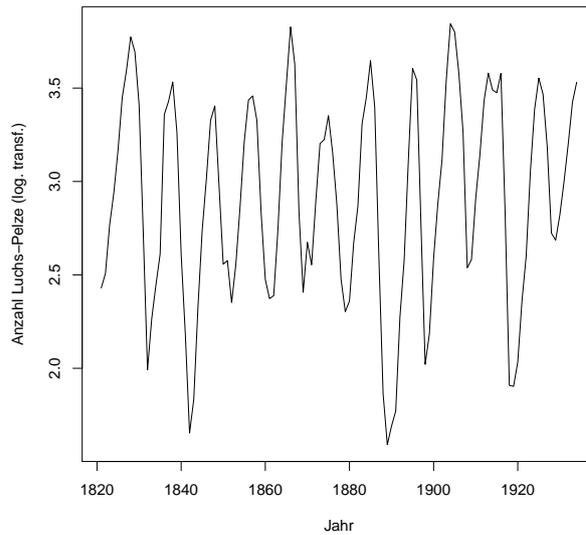


Abb. 2.1: Die Zeitreihe *Luchs*

Das klassische Komponentenmodell ■ Fallbeispiel Anzahlen von Luchspelzen – Überprüfung der Stationaritätsannahme

2.4

	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

2.5

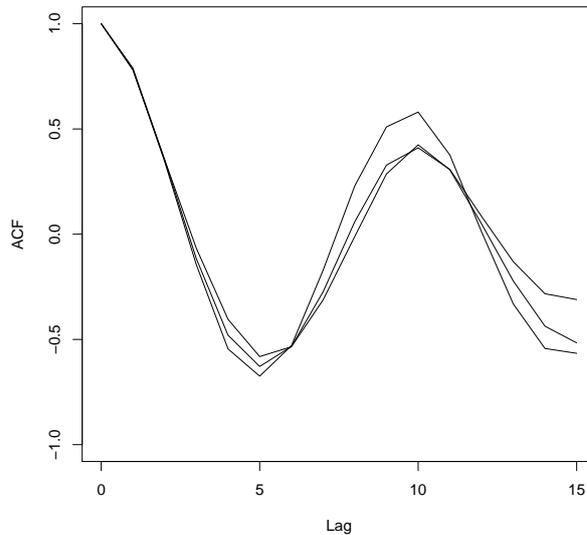


Abb. 2.2: Die ACFs für 3 Segmente der Zeitreihe *Luchs*

Das klassische Komponentenmodell ■ Fallbeispiel Anzahlen von Luchspelzen – Code

2.6

```
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="Änzahl Luchs-Pelze (log.transf.)")
source("C:/ZRA/tsutil.r")
statcheck(y,3)
```

Das klassische Komponentenmodell ■ Fallbeispiel Verkehrsunfälle – Überprüfung der Stationaritätsannahme

2.7

Die monatlichen Anzahlen der Verkehrsunfälle mit Personenschäden in Deutschland, Jan. 1974 – Jun. 1997.

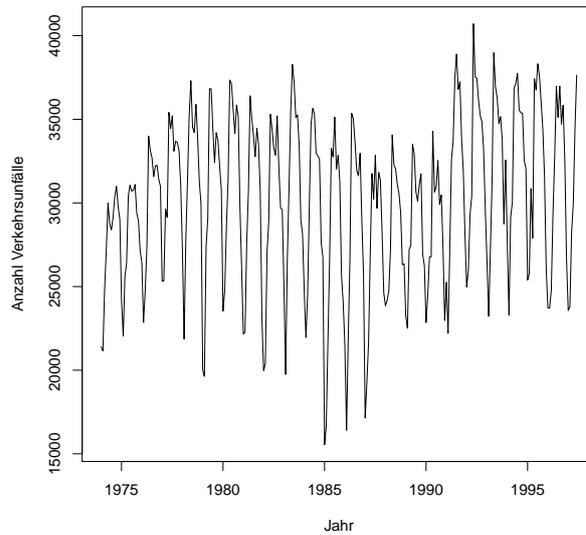


Abb. 2.3: Die Zeitreihe *Unfall*

Das klassische Komponentenmodell ■ Fallbeispiel Verkehrsunfälle – Überprüfung der Stationaritätsannahme

2.8

	Segment-Mittelwerte	Segment-Standardabweichungen
[1,]	27505.46	3106.377
[2,]	31666.79	3540.312
[3,]	31347.71	4993.230
[4,]	29414.14	5209.583
[5,]	30185.11	6017.300
[6,]	27903.86	5390.663
[7,]	28095.46	3482.299
[8,]	31676.32	5010.978
[9,]	32535.93	4297.104
[10,]	30950.68	5146.038

Das klassische Komponentenmodell ■ Fallbeispiel Verkehrsunfälle – Überprüfung der Stationaritätsannahme

2.9

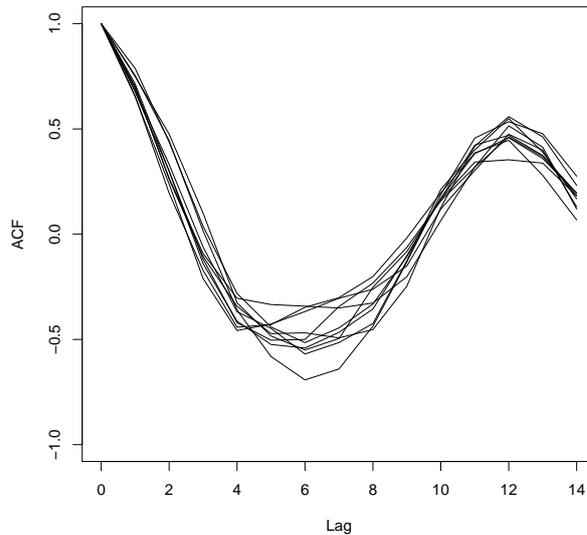


Abb. 2.4: Die ACFs für 10 Segmente der Zeitreihe *Unfall*

Das klassische Komponentenmodell ■ Fallbeispiel Verkehrsunfälle – Code

2.10

```
Unfall <- scan("C:/ZRA/Unfall.dat")
Unfall.ts <- ts(Unfall, start=c(1974,1), frequency=12)
ts.plot(Unfall.ts, xlab="Jahr", ylab="Anzahl Verkehrsunfälle")
source("C:/ZRA/tsutil.r")
statcheck(Unfall.ts, 10)
```

2.1 Die Trendkomponente

2.1.1 Polynomialer Trend

Die Trendkomponente ■ Polynomialer Trend

2.11

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

2.12

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

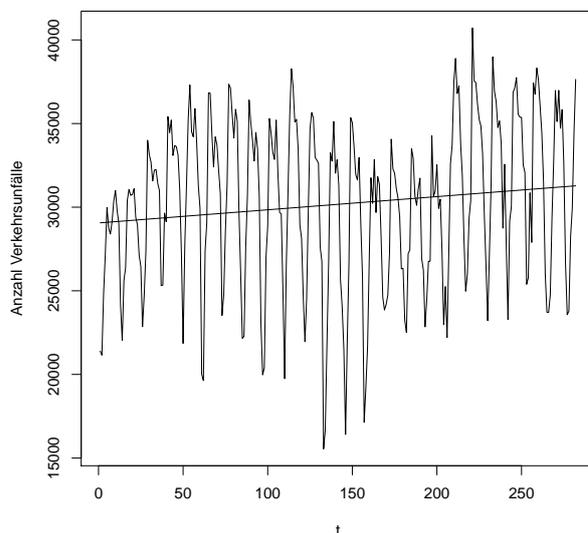


Abb. 2.5: Die Zeitreihe *Unfall* mit polynomialem Trend 1. Ordnung

Die Trendkomponente ■ Fallbeispiel Verkehrsunfälle

2.13

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

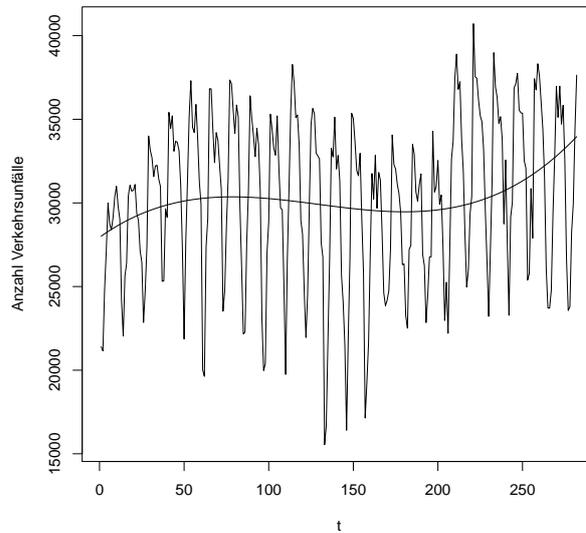


Abb. 2.6: Die Zeitreihe *Unfall* mit polynomialem Trend 3. Ordnung

Die Trendkomponente ■ Fallbeispiel Verkehrsunfälle – Code

2.14

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

2.1.2 Die glatte Komponente

Die Trendkomponente ■ Die glatte Komponente

2.15

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

2.16

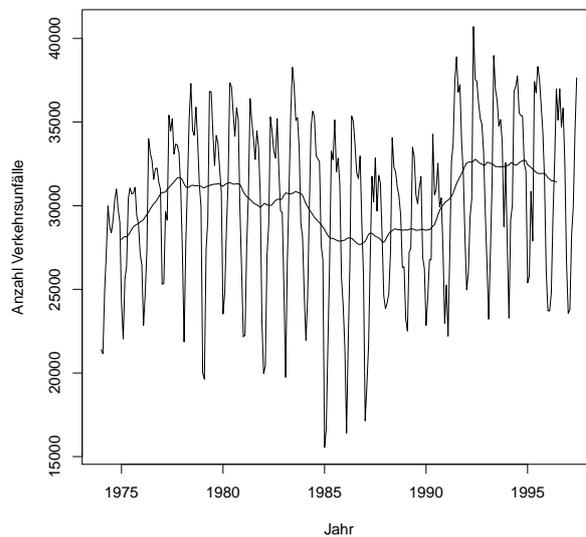


Abb. 2.7: Die Zeitreihe *Unfall* mit gleitenden 24er Durchschnitten

Die Trendkomponente ■ Fallbeispiel Verkehrsunfälle – Code

2.17

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

2.2 Die Saisonkomponente

Die Saisonkomponente

2.18

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)

2.19

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

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)

2.20

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

Die geschätzte Saison für Quartalsdaten ist

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

Die saisonbereinigte Reihe ist

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

Die Saisonkomponente ■ Modellierung der Saisonkomponente (2)

2.21

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)

2.22

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

Die geschätzte Saison für Quartalsdaten ist

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

Die saisonbereinigte Reihe ist

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

Die Saisonkomponente ■ Fallbeispiel Steuern

2.23

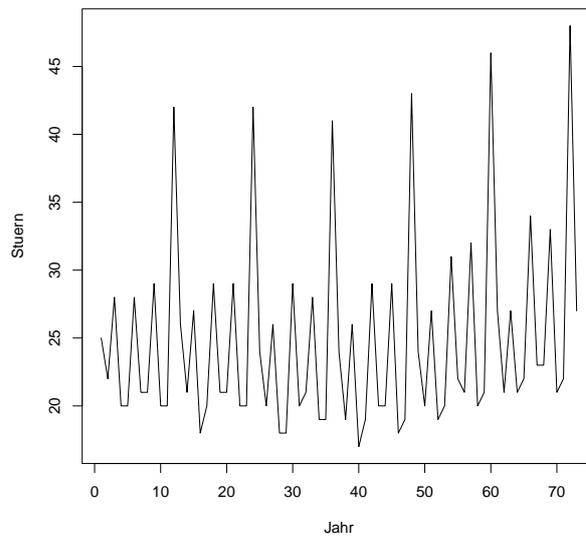


Abb. 2.8: Die Zeitreihe *Steuern*

Die Saisonkomponente ■ Fallbeispiel Steuern

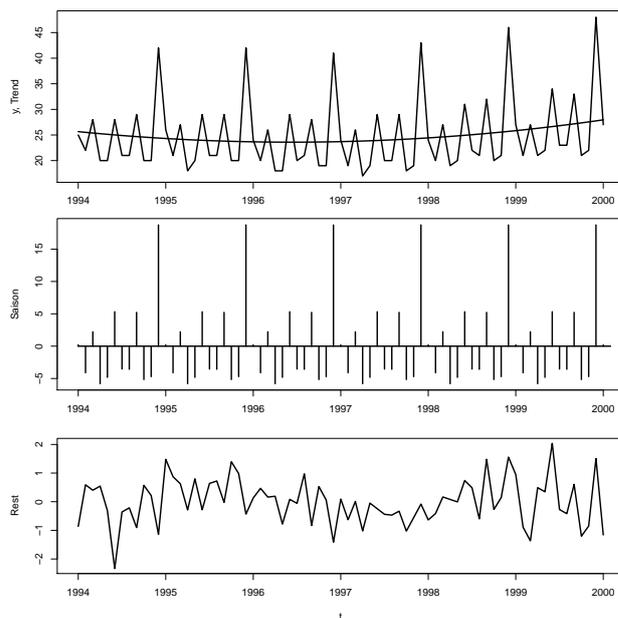


Abb. 2.9: Zerlegung der Zeitreihe *Steuern* mit polynomialem Trend 2. Ordnung und Saison

Die Saisonkomponente ■ Fallbeispiel Steuern – Code

2.25

```
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)
```

Literaturhinweise

2.26

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