

Rainer Schlittgen, Cristina Sattarhoff

Angewandte Zeitreihenanalyse mit R

Lehr- und Handbücher der Statistik

Herausgegeben von
Prof. Dr. Rainer Schlittgen

Rainer Schlittgen, Cristina Sattarhoff

Angewandte Zeitreihenanalyse mit R

4., erweiterte und aktualisierte Auflage

DE GRUYTER
OLDENBOURG

ISBN 978-3-11-069409-3
e-ISBN (PDF) 978-3-11-069439-0
e-ISBN (EPUB) 978-3-11-069466-6

Library of Congress Control Number: 2020940952

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

© 2020 Walter de Gruyter GmbH, Berlin/Boston
Satz: le-tex publishing services GmbH, Leipzig
Druck und Bindung: CPI books GmbH, Leck

www.degruyter.com

Vorwort zur vierten Auflage

Die in den bisherigen Auflagen entwickelte Konzeption des Textes hat ganz offensichtlich breite Zustimmung gefunden. Dementsprechend wurden für diese Auflage keine strukturellen Änderungen vorgenommen. Mehr noch, um den Anwendungs- und Lehrbezug aufrecht zu erhalten, wurde Frau Sattarhoff als Koautorin gewonnen. Sie ist als Dozentin mit dem Forschungsschwerpunkt Zeitreihenanalyse tätig. Ihr sind die neuen Ergänzungen und Erweiterungen sowie viele Klarstellungen und Fehlerkorrekturen zu verdanken. Die hervorstechendste Änderung ist die Umgestaltung der begleitenden R-Funktionen, die zusammen mit den Zeitreihen in dem R-Paket `tsapp` zusammengefasst wurden. Dies hat als Konsequenz, dass die Zeitreihen in dem Paket englische Bezeichnungen haben. In der im Anhang zusammengestellten Übersicht der Zeitreihen sind diese Bezeichnungen in Klammern den deutschen angefügt. Für alle Beispiele im Buch wird davon ausgegangen, dass das Paket in der R-Konsole mit `library(tsapp)` geladen wurde.

Die R-Codes der Beispiele, die Zeitreihen im ASCII-Format und das R-Paket `tsapp` stehen als Begleitmaterial auf der Web-Seite des Verlages zur Verfügung. Die Web-Adresse ist:

<http://www.degruyter.com/view/title/575978>

Dort finden Sie dies unter Zusatzinformationen.

Das Paket ist zudem auf dem Comprehensive R Archive Network (CRAN) zu finden.

Natürlich wurden die bekannt gewordenen Fehler korrigiert. Allen, die uns diesbezüglich Hinweise gegeben haben, danken wir herzlich.

Berlin, Rainer Schlittgen

Hamburg, Cristina Sattarhoff

Vorwort zur zweiten Auflage

Das Buch geht zurück auf eine Reihe von Kursen über Zeitreihenanalyse, die ich für Forscher aus verschiedenen Anwendungsbereichen gehalten habe. Von den Teilnehmern wurde immer wieder nach einem Text gefragt, der sie schnell in die Lage versetzen würde, selber Zeitreihenanalysen durchzuführen; er sollte das Gebiet ohne ‚formalen Ballast‘ darstellen. Dies hatte ich in der ersten Auflage zu realisieren versucht; diese Zielvorstellung prägt auch die vorliegende zweite Auflage. Hierfür wurde der Text gründlich überarbeitet, aktualisiert und gänzlich neu gestaltet. Das Methodenspektrum für die Behandlung von Fragestellungen aus den Wirtschaftswissenschaften wurde wesentlich stärker berücksichtigt. Einmal hat sich in der kontinuierlichen Lehrpraxis an einer wirtschaftswissenschaftlichen Fakultät gezeigt, dass auch Studierende immer mehr an einem nicht-technischen Überblick über das Gebiet interessiert sind als an einer formal-methodisch ausgerichteten Darstellung. Dann ist die Anwendergruppe aus den Wirtschaftswissenschaften in den letzten Jahren stark gewachsen, wie meine Beratungsangebote für Ökonomen gezeigt haben.

Auch wenn der formale Rahmen für ein Verständnis der Zeitreihenanalyse recht eng gehalten werden kann, ist es wichtig, bei der Anwendung der Verfahren über einen ausreichenden Hintergrund zu verfügen; man muss verstehen, was man tut. Daraus wird hier auch die jeweilige Grundlage der Methoden angesprochen, jedoch unter möglichster Vermeidung von mathematischen Ausführungen und ohne Beweise. An mehreren Stellen habe ich die Grundlagen anhand von Spezialfällen illustriert, um die wesentlichen Aspekte zu verdeutlichen.

In den Text wurde Vieles eingearbeitet, was sich in meiner langjährigen Praxis in diesem Gebiet als wichtig oder zumindest als günstig erwiesen hat. Meine Erfahrungen finden ihren Niederschlag in zahlreichen eingestreuten Bemerkungen, die für eigene Auswertungen bedeutsam sein können. Dem Anwendungsgesichtspunkt wird weiter Rechnung getragen durch die in dieser Auflage realisierte Ausrichtung der Umsetzung von Zeitreihenanalysen mit dem Programm R. R stellt eine Fülle von Funktionen zur Analyse von Zeitreihen bereit; auch die Umsetzung eigener Auswertungswünsche ist mit vertretbarem Aufwand möglich. Diese statistische Programmierumgebung kann vom R-Project for Statistical Computing kostenfrei unter den Bedingungen der General Public License bezogen werden. Das Arbeiten mit R findet in der Weise seinen Niederschlag im Text, dass zu den Anwendungsbeispielen jeweils am Beispieldende der zugehörige R-Code angegeben wird. Er wird auch kommentiert, am Anfang ausführlicher, gegen Ende des Buches etwas weniger ausführlich. So lassen sich die Beispiele direkt reproduzieren. Bei den Codes habe ich allerdings die Teile weggelassen, die nur der Verschönerung des Ausdruckes oder der Grafik dienen. Weiter wird im letzten Kapitel eine kurze Einführung in R gegeben. Zudem sind alle in den Beispielen vorkommenden R-Funktionen in einer Tabelle aufgelistet. So-

VIII — Vorwort zur zweiten Auflage

fern sie aus einem der zahlreichen Pakete stammen, die vom CRAN heruntergeladen werden können, ist dies mit angegeben.

Auch wenn viel Mühe darauf verwandt wurde, Fehler zu eliminieren, ist es erfahrungsgemäß unmöglich, alle zu bereinigen. Daher freue ich mich über jede entsprechende Rückmeldung. Eine jeweils aktuelle Errata-Liste der bereits gefundenen Fehler wird ebenfalls auf der Web-Seite des Verlages für dieses Buch unter den Zusatzinformationen veröffentlicht.

Allen, die mich bei der Erstellung des Buches unterstützt haben, möchte ich herzlich danken.

Rainer Schlittgen

Inhalt

Vorwort zur vierten Auflage — V

Vorwort zur zweiten Auflage — VII

1 Fragestellungen und Datensituation — 1

2 Grundlagen und einfache Methoden — 9

- 2.1 Stationäre Zeitreihen — 9
- 2.2 Das Komponentenmodell — 26
- 2.3 Deterministische Trends — 27
- 2.4 Saisonbereinigung — 43
- 2.5 Transformationen — 52
- 2.6 Einfache Extrapolationsverfahren — 59
- 2.7 R-Funktionen — 65

3 Lineare Zeitreihenmodelle — 69

- 3.1 Autoregressive Modelle — 69
- 3.2 MA-Modelle — 87
- 3.3 ARMA-Modelle — 92
- 3.4 ARIMA-Modelle — 100
- 3.5 R-Funktionen — 109

4 Differenzen- und Trendinstationarität — 111

- 4.1 Instationaritätstypen und ihre Implikationen — 111
- 4.2 Einheitswurzeltests — 114
- 4.3 R-Funktionen — 118

5 Prognosen mit univariaten Zeitreihenmodellen — 119

- 5.1 Verfahren der exponentiellen Glättung — 119
- 5.2 Prognosen mit ARIMA-Modellen — 122
- 5.3 Trendextrapolation mit ARMA-Störungen — 127
- 5.4 Zur Auswahl eines Prognosemodells — 129
- 5.5 R-Funktionen — 132

6 Periodizitäten in Zeitreihen — 133

- 6.1 Periodizitäten — 133
- 6.2 Periodische Trends — 136
- 6.3 Das Periodogramm — 138
- 6.4 Spektren — 148

6.5	Spektralschätzung — 160
6.6	R-Funktionen — 171
7	Prozesse mit langem Gedächtnis — 173
7.1	Einführung der Prozesse — 173
7.2	Bestimmung des fraktionellen Exponenten — 177
7.3	Prognosen mit ARFIMA-Modellen — 184
7.4	R-Funktionen — 186
8	Mehrdimensionale Zeitreihen — 187
8.1	Kenngrößen mehrdimensionaler Zeitreihen — 187
8.2	Mehrdimensionale lineare Zeitreihenmodelle — 199
8.3	R-Funktionen — 222
9	Regressionsmodelle für Zeitreihen — 225
9.1	Regression mit autokorrelierten Störungen — 225
9.2	Interventionsanalysen — 232
9.3	Transferfunktionsmodelle — 239
9.4	R-Funktionen — 246
10	Zustandsraummodelle und Kalman-Filter — 247
10.1	Zustandsraummodelle — 247
10.2	Kalman-Filter — 252
10.3	R-Funktionen — 261
11	Nichtlineare Modelle — 263
11.1	Nichtlinearität in Zeitreihen — 263
11.2	Markov-switching Modelle — 265
11.3	Threshold-Modelle — 273
11.4	Bedingt heteroskedastische Modelle — 275
11.5	R-Funktionen — 287
12	Spezielle Probleme — 289
12.1	Fehlende Werte — 289
12.2	Nicht gleichabständige Beobachtungen — 292
12.3	Ausreißer und robuste Verfahren — 295
12.4	R-Funktionen — 308
13	Einführung in R — 309
13.1	Erste Schritte — 309
13.2	Datentypen und Objekte — 312
13.3	Operatoren und Funktionen — 318

13.4	Bibliotheken und Programmierung — 327
13.5	Einlesen und Exportieren von Daten — 329
13.6	Grafik — 330

Die aufgerufenen R-Funktionen — 333

Die Zeitreihen — 335

Literatur — 337

Abkürzungen und Symbole — 345

Sachindex — 347

1 Fragestellungen und Datensituation

Zeitreihen sind zeitlich geordnete Beobachtungswerte des jeweils gleichen Sachverhaltes. Beispiele sind etwa

1. Monatliche Auftragseingänge für eine Fertigungsstätte eines Unternehmens (in Währungseinheiten);
2. Tägliche Niederschlagsmenge (in mm);
3. Konzentration eines Hormons im Blut (mg/ml);
4. Sauerstoffgehalt (mg/l) in der Leine bei Neustadt;
5. Das kontinuierlich abgenommene EEG (Mikrovolt).

In den ersten beiden Fällen liegen Beobachtungen y_t zu gleichen Zeitabständen Δt vor. Konkret ist Δt ein Monat bei 1.; bei 2. gilt $\Delta t = 1$ Tag. Bei 3. und 4. werden die Beobachtungen ebenfalls in dieser Weise vorgenommen. Sehr oft wird bei 3. ein zeitlicher Abstand von 15 Minuten gewählt, beim vierten Beispiel sind 30 Minutenintervalle adäquat. Wir sprechen dann von Beobachtungen in diskreter Zeit und nummerieren die Zeitpunkte einfach durch: $t = 1, 2, 3, \dots, N$. Diese Situation ist sicherlich die einfachste. Sie bildet die Grundlage für das Folgende und wird im Mittelpunkt der Diskussion stehen.

Im fünften Beispiel wird die interessierende Variable stetig beobachtet. Zeitreihen mit stetigem (Zeit-) Parameter werden für die Auswertung i. d. R. diskret abgetastet (mit kleinem Δt) und dann so analysiert, als ob sie von vornherein diskret beobachtet worden wären. Wir gehen daher auf Zeitreihen mit stetigem Zeitparameter hier nur insoweit ein, als die Frage der Wahl von Δt angesprochen wird.

Falls Beobachtungen ausfallen, resultieren dann Zeitreihen mit Lücken oder allgemeiner mit nicht gleichabständigen Beobachtungen. Wie weit sich die Methoden für die Standardsituation zur Erfassung auch dieser Situation modifizieren lassen, wird an geeigneter Stelle besprochen.

Zu Beginn der Beschäftigung mit einer Zeitreihe y_1, \dots, y_n sollte stets eine Problemstellung formuliert sein. Daraus ergibt sich auch, welche der Methoden zur Analyse von Zeitreihen anzuwenden ist.

Beispiel 1.1 (Stromspitzenbedarf)

Im Zusammenhang mit der Diskussion über den Bau eines Kraftwerkes war die Prognose des jährlichen Stromspitzenbedarfs, auch Jahreshöchstlast genannt, von Berlin (West) Ende der 70er Jahre von Bedeutung. Die Befürworter argumentierten mit einem sehr starken Anstieg des Bedarfs, der sich in Zukunft noch stärker ausprägen würde. Gegner hielten dagegen, dass sich der Trend viel moderater darstellen würde und Mehrbedarf über alternative Ansätze befriedigen lassen würde.

Beispiel 1.2 (ifo-Geschäftsklima)

Konjunkturzyklen in marktwirtschaftlichen Systemen sind Schwankungen des Auslastungsgrades des gesamtwirtschaftlichen Produktionspotentials. Ein wichtiger Frühindikator für die deutsche Konjunktur ist das „ifo Geschäftsklima für die gewerbliche Wirtschaft“. Anhand dieses Indikators möchte man etwa möglichst früh eine Abschwungsphase erkennen. Dann hat man allerdings saisonale Einflüsse auszuschalten, Einflüsse, die durch jahreszeitliche Bedingungen gegeben sind und sich mehr oder weniger deutlich Jahr für Jahr in gleicher Weise zeigen.

Beispiel 1.3 (Schweinezyklus)

Unter dem Begriff Schweinezyklus versteht man eine periodische Schwankung auf der Angebotsseite. Der Begriff, der fester Bestandteil der Preistheorie geworden ist, kommt daher, dass das Phänomen zuerst auf dem Markt für Schweinefleisch beobachtet wurde, siehe Hanau (1928). Die Periodizität wird nach den entsprechenden Vorstellungen dadurch erzeugt, dass es bei hohen Preisen zu verstärkten Investitionen kommt. Diese wirken sich erst verzögert auf das Angebot aus. Dann führen sie aber zu einem Überangebot und Preisverfall. In der Folge kommt es zu einer Reduzierung der Produktion, die sich ebenfalls erst zeitverzögert auswirkt. Insgesamt ergeben die Zeitverzögerungen im Regelmechanismus zwischen Angebot, Nachfrage und Preis eine instabile Marktsituation, die das Angebot regelmäßig schwanken lässt.

Weitere Bereiche, bei denen solche regelmäßigen Schwankungen zu beobachten sind, sind die Lehrerausbildung und die Ausbildung von Maschinenbau-Ingenieuren.

Beispiel 1.4 (Sekretions-Rhythmus von Hormonen)

Praktisch alle Lebensprozesse weisen rhythmische Schwankungen auf. Bereits bei Einzellern bestehen Zellteilungs- und Wachstumsrhythmen. Bei höher entwickelten Organismen überlagern sich eine Vielfalt von rhythmischen Vorgängen und machen häufig eine Analyse sehr schwierig. Nach ihrer Zykluslänge werden Rhythmen im biologischen Bereich als infradian (Zykluslänge > 24 Stunden; Beispiel: Menstruationszyklus), circadian (Zykluslänge ≈ 24 Stunden; Beispiel: Schlaf/Wach-Rhythmus) und ultradian (Zykluslänge < 24 Stunden; Beispiel: Schlafphasen-Rhythmus von ≈ 90–105 Minuten) eingeteilt. Im Bereich der hormonellen Steuerung des Säugetierorganismus – einschließlich des Menschen – finden sich viele derartige rhythmische Vorgänge. So erzeugt die Nebennierenrinde in den frühen Morgenstunden einen relativ hohen und in den Abendstunden einen nur sehr niedrigen Blutspiegel des wichtigen Hormons Cortisol. Diesem circadianen Rhythmus sind Schwankungen kleinerer Amplitude überlagert, die durch kurzdauernde Sekretionsschübe des Cortisols verursacht werden. Diese führen oft zu extremen Blutspiegelwerten.

Allgemein wird der Sekretions-Rhythmus vieler Hormone durch äußere oder innere Einflüsse moduliert. Dies kann durch den Licht/Dunkel-Wechsel, durch den Schlaf/Wach-Rhythmus oder sogar durch den Schlaf-Phasen-Rhythmus geschehen. Diese können sich dann circadianen oder kurzfristigen episodischen Schwankungen überlagern.

Hormonale Zeitreihen bestehen aus kurzabständigen, regelmäßigen Messungen der Hormonkonzentration; die Abstände betragen i. d. R. 10 oder 15 Minuten. Ein typisches Muster ist ein recht periodisches Verhalten: steilen Anstiegen folgen langsamere Abfälle des Hormonspiegels. Der Anstieg resultiert dabei von der Ausschüttung des Hormons in das Blutplasma; der langsamere Abfall kommt durch biochemische Abbau-Prozesse zustande. Die Frage ist dann, ob neben dem ausgeprägten circadianen Rhythmus auch andere deutlich hervorstechende ultradiane Rhythmen bei der Hormonausschüttung zu erkennen sind.

Beispiel 1.5 (Jahresringe von Bäumen)

Das Wachstum von Bäumen wird anhand von Jahresringen untersucht; je mehr ein Baum pro Jahr wächst, desto breiter ist der zugehörige Jahresring. Die Jahresringe sind ja gut an einer Schnittfläche zu erkennen und zu messen.

Das Interesse an den Jahresringen resultiert unter anderem aus der Möglichkeit der Datierung anhand von Jahresringen; das ist ein ganzer Wissenschaftsbereich, die Dendrochronologie. So kann das Alter eines Hauses anhand der Untersuchung von Balken, die zum Bau verwendet wurden, ebenso wie die Entstehungszeit eines Bildes, das auf Holz gemalt ist, eingegrenzt werden. Eine aktuelle Anwendung ist etwa ein wieder entdecktes Gemälde Shakespeares, von dem man wissen wollte, ob es zu seinen Lebzeiten gemalt wurde. Zu den verschiedenen Untersuchungsmethoden zur Altersbestimmung gehörte auch die Analyse der Jahresringe. Diese ergab, dass das Bild frühestens 1597 hergestellt wurde.

Die Stärke eines Jahresringes Y_t wird dabei von klimatischen Größen abhängen. Dazu gehören die jährlichen Durchschnittstemperaturen T_t und der jährliche Niederschlag N_t . Die Frage ist dann, in welcher Weise die Abhängigkeit zu erfassen ist.

Beispiel 1.6 (Werbeausgaben und Umsatz)

Eine der wichtigsten Probleme der Marktforschung ist die Bestimmung der besten Werbestrategie. Eine zentrale Frage ist dabei der Einfluss des Werbeaufwandes X auf den Umsatz Y . Wird dieses über die Zeit betrachtet, so sind Auswirkungen nicht nur instantan zu berücksichtigen, sondern es sind auch Vergangenheitswerte des Werbeaufwandes einzubeziehen. Zudem kann u. U. eine Feedback-Beziehung existieren, so dass ein Einfluss des Umsatzes auf die (nachfolgenden) Werbeausgaben zu berücksichtigen ist.

Beispiel 1.7 (Black-Out in New York)

Am 9. November 1965 gingen in New York um 17.27 Uhr die Lichter aus – und blieben es die ganze Nacht. Am 10. August 1966 meldete die New York Times eine erhöhte Geburtenzahl. Es stellt sich die Frage, ob an der Meldung ‚etwas dran‘ war. Hatte dieser Black-Out tatsächlich neun Monate später ein Ansteigen der Anzahlen der Geburten zur Folge, wie immer wieder in der Presse behauptet wird? Dann müsste sich ein Interventionseffekt in der Zeitreihe der wöchentlichen Geburten nachweisen lassen. Gesucht ist also ein Modell für die normale Zeitreihe und eine Möglichkeit, den Black-Out-Effekt zu testen.

Die Fragen, die in den Beispielen angeschnitten werden, gehören zu dem Katalog von Fragestellungen, die im Rahmen der Zeitreihenanalyse untersucht werden:

- Weist die Zeitreihe einen Trend auf, d. h. ein langfristiges Ansteigen oder Abfallen des Niveaus der Werte? Welche Gestalt hat er?
- Gibt es eine Saison, ein durch die Jahreszeit bedingtes, regelmäßig wiederkehrendes Muster? Welche Gestalt hat es?
- Gibt es sonstige zyklische Schwankungen? Welche Länge haben die Zyklen bzw. wie oft kehren die Schwankungen wieder? Wie stark sind sie?
- Weist die Zeitreihe systematische Veränderungen im Zeitverlauf auf? Wenn solche vermutet werden, wie lassen sie sich dann nachweisen?
- Sind einzelne Werte als Ausreißer anzusehen, d. h. sind sie so extrem, dass sie nicht mit dem sonstigen Verlauf der Reihe in Einklang zu bringen sind?
- Wie wird sich die Größe, die durch die Zeitreihe dargestellt ist, weiterentwickeln?
- Gibt es Zusammenhänge zwischen verschiedenen Zeitreihen?

Um solche Fragen, zu beantworten, ist es unabdingbar, Annahmen über das Zustandekommen der jeweiligen Reihe(n) zu machen, m. a. Worten ein geeignetes Modell zu unterstellen. Den Vorstellungen der Statistik entsprechend werden als erstes die beobachteten Werte als Realisationen von Zufallsvariablen aufgefasst. Das bedeutet, dass man eine Folge von Zufallsvariablen Y_1, \dots, Y_N zu betrachten hat. Schon wegen der Prognosen, aber auch aus theoretischen Gründen ist es sinnvoll, einen größeren Zeithorizont zuzulassen, als den, in dem die vorliegenden Werte beobachtet wurden.

Definition 1.8 (Stochastischer Prozess, Zeitreihe)

Ein *stochastischer Prozess* ist eine Folge (Y_t) von Zufallsvariablen. Der Index t , $t \in \mathbb{N}, \mathbb{N}_0$, oder \mathbb{Z} wird i. d. R. als Zeit aufgefasst. Eine *Zeitreihe* ist eine Folge y_1, \dots, y_N von Realisationen eines Ausschnittes von (Y_t) . Man nennt diese auch *Zeitpfad* oder *Trajektorie* des Prozesses. Oft wird (Y_t) oder Y_1, \dots, Y_N selbst als Zeitreihe bezeichnet.

Als grundlegende Annahme unterstellt man in der Statistik üblicherweise, dass die den Beobachtungen zugrundeliegenden Zufallsvariablen Y_t eine Folge von unabhängigen, identisch verteilten Zufallsvariablen darstellen. Solche Folgen sind auch

in der Zeitreihenanalyse von großer Bedeutung; sie sind zumindest Bausteine für interessantere Prozesse.

Definition 1.9 (White-Noise-Prozess)

Ein *White-Noise-Prozess* ist eine Folge von unabhängigen identisch verteilten Zufallsvariablen. Derartige Prozesse werden im Folgenden meist mit (ε_t) bezeichnet.

White-Noise-Prozesse können beliebige (Rand-) Verteilungen haben. In diesem Text wird i. d. R. unterstellt, dass Erwartungswert und Varianz wohldefiniert sind. Außerdem werden solche Prozesse meist bei null zentriert unterstellt.

Wie bei der klassischen Statistik basieren viele Methoden und Aussagen in der Zeitreihenanalyse auf Verfahren, welche die Normalverteilung voraussetzen. Demgemäß bildet die Voraussetzung eines Normalprozesses die zweite, für weite Teile der Praxis wesentliche Annahme in der Zeitreihenanalyse.

Definition 1.10 (Normalprozess/Gauß-Prozess)

Ein stochastischer Prozess (Y_t) , bei dem für alle möglichen Indizes t_1, t_2, \dots, t_k die Variablen $Y_{t_1}, Y_{t_2}, \dots, Y_{t_k}$ multivariat normalverteilt sind, heißt *Normalprozess* oder *Gauß-Prozess*.

Eine besondere Bedeutung kommt in der Zeitreihenanalyse den *Gaußschen White-Noise-Prozessen* zu, also den Folgen (ε_t) von unabhängigen, identisch normalverteilten Zufallsvariablen $\varepsilon_t \sim \mathcal{N}(\mu_\varepsilon, \sigma_\varepsilon^2)$.

Solche Prozesse sind zwar in der Praxis bei zeitlich geordneten Beobachtungsdaten kaum anzutreffen; sie spielen jedoch eine große Rolle beim Modellieren. Wird nämlich ein Modell für eine Zeitreihe gesucht, so orientiert man sich wesentlich daran, ob die nicht erklärten Reste oder Störungen als Realisation eines White-Noise-Prozesses angesehen werden können. Ist dies erreicht, so kann man davon ausgehen, dass alle systematischen Komponenten der Zeitreihe im Modell erfasst sind. Die Normalverteilung erleichtert dabei das Überprüfen der Relevanz der Parameter des Modells bzw. der Einschätzung der Variabilität von Schätzungen und Prognosen.

Definition 1.11 (Random-Walk)

Sei $(\varepsilon_t)_{t \in \mathbb{N}_0}$ ein White-Noise-Prozess. Der stochastische Prozess $Y_t = \mu + \sum_{i=0}^t \varepsilon_i$ heißt *Random-Walk*.

Ist $E(\varepsilon_t) = 0$, so liegt ein Random-Walk *ohne Drift* vor. Bei $E(\varepsilon_t) \neq 0$ spricht man von einem Random-Walk *mit Drift*.

Allgemein dient der Random-Walk als Modell für einen *Irrfahrtsprozess*. Können speziell die zur Definition eines Random-Walk benötigten ε_t nur die Werte -1 und 1 annehmen, so kann er als Auszahlungsmodell für Glücksspiele dienen. Nach t Spielen hat sich ein Gewinn bzw. Verlust von Y_t Geldeinheiten angesammelt. Mit stetig verteilten ε_t bilden Random-Walks vielfach Modelle für Aktienkursreihen.

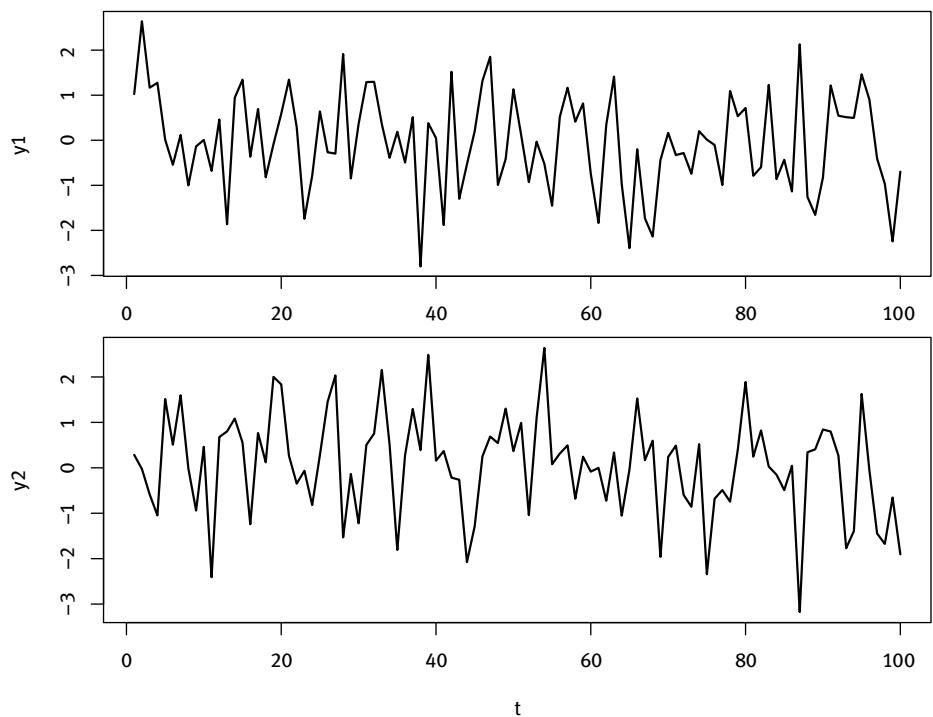


Abb. 1.1: Zwei Realisationen eines Gaußschen White-Noise-Prozesses

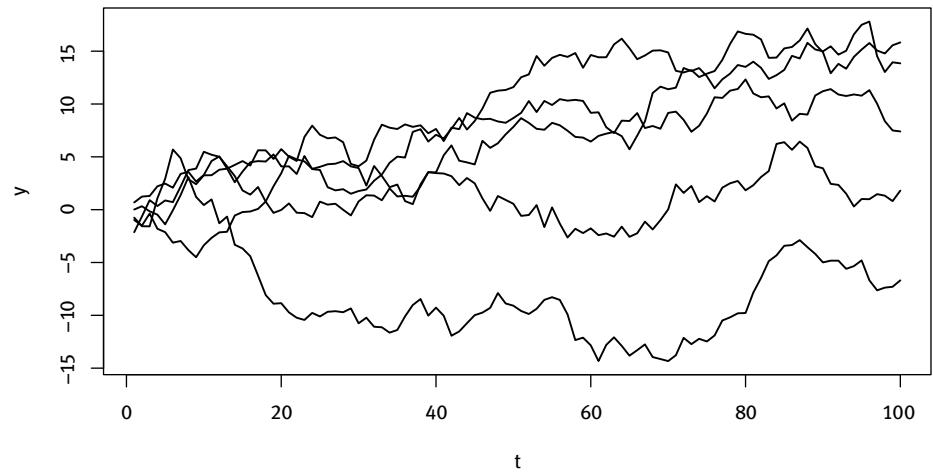


Abb. 1.2: Fünf Realisationen eines Random-Walk ohne Drift

In den meisten zeitreihenanalytischen Anwendungen wird es schwierig sein, den zugrundeliegenden Zufallsvorgang substanzial zu spezifizieren. Ein stochastischer Prozess dient dann als *Modellvorstellung*, in die bestimmte Annahmen über den Mechanismus eingehen, der die beobachtete Zeitreihe produziert hat.

Anders als bei der klassischen Statistik liegt in der Zeitreihensituation streng genommen nur eine Beobachtung, genauer eine Beobachtungssequenz, vor. Zu jedem Zeitpunkt t wird nur eine Beobachtung der jeweiligen Zufallsvariablen Y_t gemacht; für verschiedene Zeitpunkte t, s sind die zugehörigen Y_t, Y_s ja unterschiedliche Zufallsvariablen. Damit es dennoch möglich ist, daraus Schlussfolgerungen zu ziehen, sind geeignete Annahmen zu treffen.

Die zentrale Forderung ist die *Stationarität*. Diese sichert, dass sich die stochastischen Charakteristika über die Zeit nicht ändern. Gegebenenfalls ist die Stationarität ‚herzustellen‘. Dies ist ein zentrales Thema des folgenden Kapitels. Zusätzlich wird noch die *Ergodizität* benötigt. Grob gesprochen ergibt diese Forderung, dass Charakteristika der Verteilungen je endlich vieler Zufallsvariablen Y_{t_1}, \dots, Y_{t_k} wie Erwartungswerte, Varianzen und Kovarianzen sich durch geeignete, aus dem zeitlichen Verlauf bestimmte Größen konsistent schätzen lassen.

2 Grundlagen und einfache Methoden

2.1 Stationäre Zeitreihen

Darstellung von Zeitreihen

Die Analyse von Zeitreihen hat das Ziel, Strukturen und Regelmäßigkeiten in Zeitreihen aufzuspüren und zur Beschreibung oder Modellierung auszunutzen. Darauf basieren ja, wie im ersten Kapitel angeführt, die weiteren Schritte. Dies wird etwa an folgendem Beispiel deutlich.

Beispiel 2.1 (Regen in LA)

Am Ende eines verregneten Januars tröstete die Los Angeles Times die sonnenverwöhnten Südkalifornier: „Wenn Sie anfangen, das Ende des Regens herbeizusehen, vertrauen Sie auf die Zyklen der Natur. Wie die historische Darstellung zeigt, ist Los Angeles unbeständig, was den Regen betrifft. Im Juni mögen dann einige Leute sich nach diesen feuchten Januartagen sehnen.“ (LA Times vom 28.1.1997) Diesem Trost wurde eine verschönerte Version der Abbildung 2.1 der jährlichen Niederschlagsmengen beigefügt. Die zugehörige Zeitreihe wird mit REGEN bezeichnet.

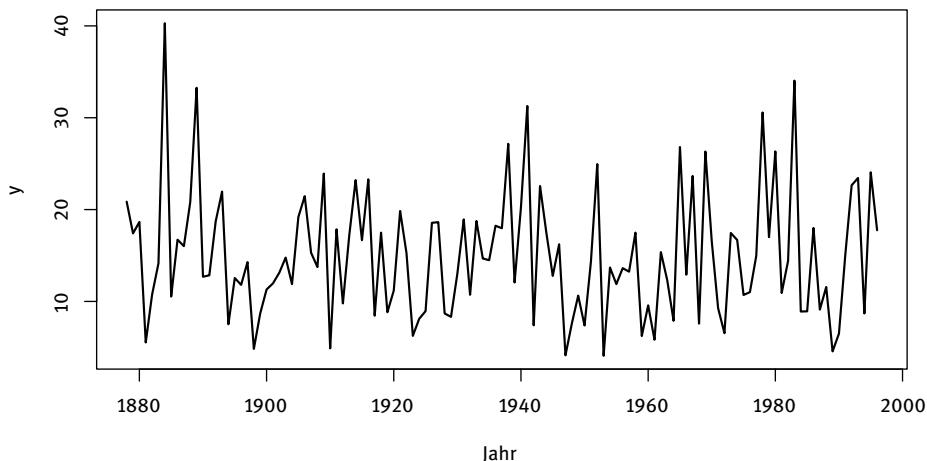


Abb. 2.1: Jährliche Niederschlagsmenge (Inch) in Los Angeles

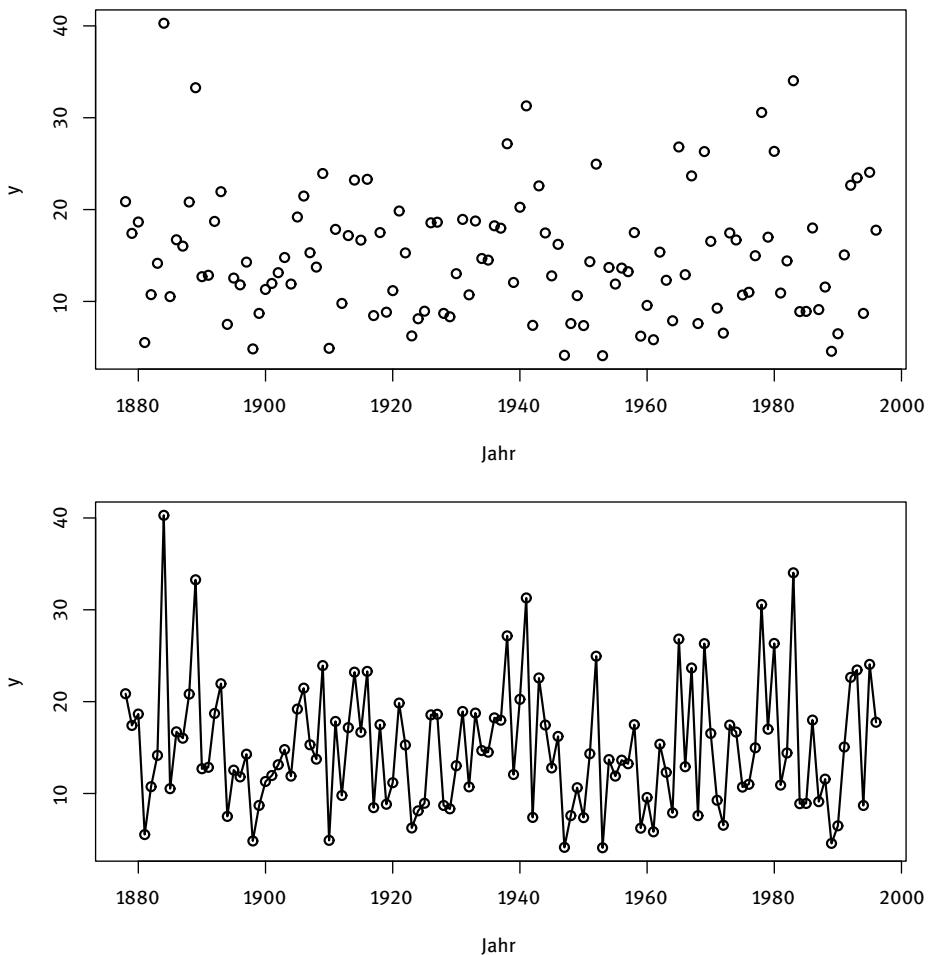


Abb. 2.2: Weitere Darstellungen der Zeitreihe REGEN

R-Code:

```

1 data(RAINFALL)
2 y <- RAINFALL
3 plot(y)
4 plot(y,type="p")
5 plot(y,type="o")

```

Die Datensätze im R-Paket tsapp sind alle Zeitreihen, d. h. sie sind mit dem Zeitreihenattribut versehen. Daher wird in der in Zeile 3 erzeugten Grafik ein Linienplot erstellt. Andere Grafiksymbole sind direkt anzufordern. "p" steht für Punkt und "o" für Punkt und Verbindungslien.

Wie in diesem Beispiel ist es *stets der erste Schritt* bei der Analyse einer Zeitreihe, sich eine grafische Darstellung, einen *Plot*, der Reihe anzusehen. Dabei ergibt die Darstellung der linearen Verbindungslien meist einen besseren Eindruck der Gesamtstruktur der Reihe als eine Darstellung einzelner Punkte. Manchmal ist eine Überlagerung von Punkten und Verbindungslien günstig.

Anhand eines solchen Plots können wichtige Charakteristika der Reihe erkannt werden; dazu zählen etwa das Vorliegen eines Trends, einer Saison, die Existenz von Ausreißern, d. h. von einzelnen Werten, die stark aus dem sonstigen Verlauf der Werte herausfallen, und auch das Vorhandensein von Lücken. Man hat herausgefunden, dass die Möglichkeit Strukturen zu erkennen, stark von dem *Aspektverhältnis*, dem Quotienten „Breite/Höhe“ der grafischen Darstellung der Zeitreihe abhängt, vgl. Cleveland (1994). Günstig ist es, wenn dieses so gewählt wird, dass die durchschnittliche Orientierung der Anstiege einen Winkel von 45° und die der Abstiege -45° aufweist. Damit wird die Diskriminierungsmöglichkeit aufeinanderfolgender Streckenzüge maximiert. Um dies zu erreichen, wird das Verhältnis von Breite und Höhe $\alpha = b/h$ über die Minimierung des Zielkriteriums

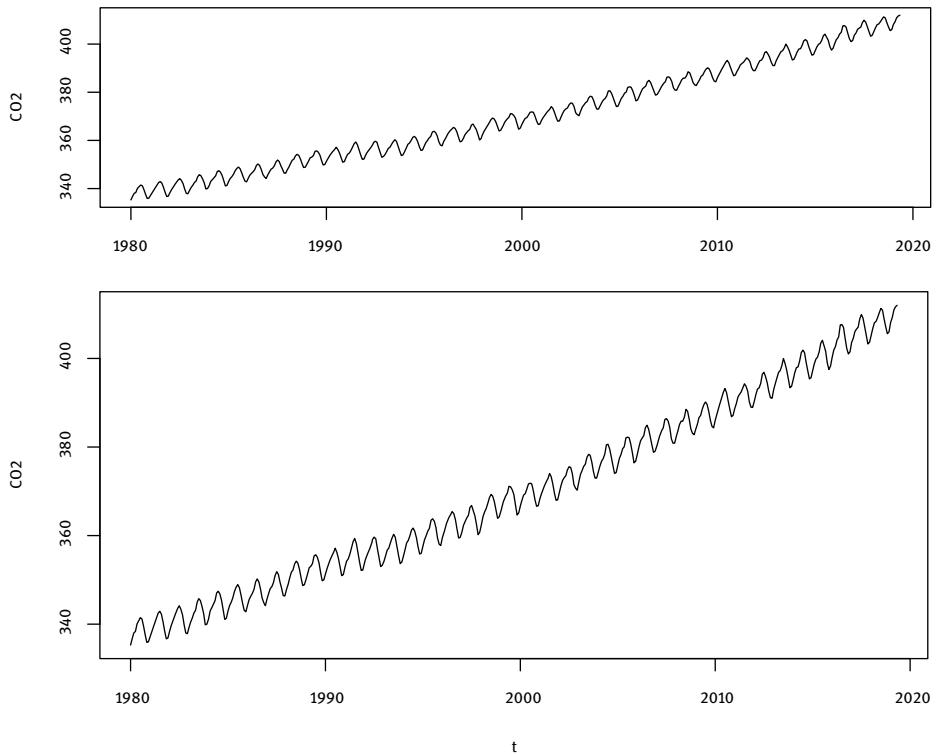
$$Z(\alpha) = \left(\frac{1}{n} \sum_{t=2}^n \left| \tan^{-1} \left(\frac{s_t}{\alpha} \right) \right| - 1 \right)^2. \quad (2.1)$$

bestimmt. Hierbei ist s_t die Steigung eines Liniensegments, einfach $s_t = y_t - y_{t-1}$, wenn wie üblich die Zeitskala fortlaufend durchnummieriert ist. Das erlaubt vor allem schnellere Oszillationen in den Daten zu sehen. Allerdings werden Strukturen, die sich langsamer entwickeln, oft schwerer erkennbar. Noch besser erscheint, entweder nur die Anstiege oder nur die Abstiege an der 45° -Orientierung auszurichten.

Beispiel 2.2 (CO₂-Konzentration über Mauna Loa)

Das Observatorium in Mauna Loa ermittelt regelmäßig die CO₂-Konzentration (ppmv) über Hawaii. Die CO₂-Entwicklung in der Atmosphäre gilt ja als eine wesentliche Ursache der globalen Erwärmung. Die als MAUNALOA bezeichnete Reihe sind monatliche Durchschnitte von März 1958 bis Mai 2019. Die Daten stammen von Keeling et al. (2001). Der Abschnitt von Januar 1980 bis Mai 2019 ist in der Abbildung 2.3 dargestellt. Ein modifiziertes Zielkriterium, bei dem nur die Anstiege eingehen, führt zu dem Aspektverhältnis von 4.4. Die obere Grafik ist entsprechend skaliert. Man erkennt deutlich, dass die Anstiege an der 45° -Linie ausgerichtet sind; die Abstiege sind dagegen steiler.

Eine übliche Darstellung würde die Ordinate viel weiter auseinanderziehen, etwa so wie im unteren Teil der Abbildung 2.3 ($\alpha = 2$). Hier ist wiederum besser zu erkennen, dass der längerfristige Anstieg zwischen 1990 und 1995 etwas langsamer verlief als in den angrenzenden Zeitintervallen.

**Abb. 2.3:** CO₂-Konzentration über Mauna Loa**R-Code:**

```

1 data(MAUNALOA)
2 y <- window(MAUNALOA,start=c(1980,1),end= c(2019,5))
3 a <- aspectratio(y)
4 windows(width=8*a, height=8)
5 plot(y,ylab="CO2",xlab="")
6 windows(width=8*a, height=8*a)
7 plot(y,ylab="CO2",xlab="")

```

Die Reihe wurde verkürzt, da sonst die Charakteristiken nicht gut zu erkennen sind. Zudem kommen im ersten Teil NA's vor. Mit der in Zeile 3 aufgerufenen Funktion `aspectratio`, die im Paket `tsapp` implementiert ist, wird das (modifizierte) Aspektverhältnis bestimmt. Die Funktion `windows` dimensioniert dann die Grafikausgabe (im Windows-Betriebssystem) in der gewünschten Weise.

Beispiel 2.3 (Regen in LA – Fortsetzung von Seite 9)

Gefragt wird nun, welche Strukturen aus dem Plot der REGEN-Reihe zu ersehen sind.

Zuerst sieht man, dass das Niveau der jährlichen Niederschläge über die dargestellte Zeit nicht steigt oder fällt. Auch die Streuung der Werte ist weitgehend konstant. Andere der angegebenen Charakteristika wie Lücken, Ausreißer oder saisonales Verhalten sind ebenfalls nicht zu erkennen.

Man kann nun noch versuchen, weiter gehende Regelmäßigkeiten eventuell aus der Abfolge von hohen und niedrigen Niederschlagsmengen zu erkennen. Folgten etwa jedes Mal auf sieben regenreiche Jahre stets sieben trockene, so ließe sich das zu langfristigen Prognosen ausnutzen.

Autokovarianz- und Autokorrelationsfunktion

Das Niveau einer Zeitreihe wird in aller Regel mit dem einfachen arithmetischen Mittel \bar{y} , die Streuung mit der Standardabweichung s bzw. mit der Varianz s^2 erfasst:

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N y_t, \quad s^2 = \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y})^2, \quad s = \sqrt{s^2}. \quad (2.2)$$

Die im Beispiel 2.3 als letztes genannte Struktur lässt sich mittels der Korrelation erfassen. Zunächst ist die empirische Kovarianz für einen zweidimensionalen Datensatz $(x_i, y_i)', i = 1, \dots, N$ definiert als

$$c_{XY} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}). \quad (2.3)$$

Daraus gewinnt man den empirischen Korrelationskoeffizienten durch Standardisierung:

$$r_{XY} = \frac{c_{XY}}{\sqrt{s_X^2} \sqrt{s_Y^2}}. \quad (2.4)$$

Für die Zeitreihensituation wird aus den Werten y_1, \dots, y_N für jeden zeitlich konstanten Abstand, jedes *Lag* τ , ein künstlicher zweidimensionaler Datensatz gebildet: $(y_{1+\tau}, y_1), \dots, (y_N, y_{N-\tau})$. Daraus wird dann die Kovarianz bzw. der Korrelationskoeffizient berechnet. Für jedes Lag τ ist eine Kovarianz bzw. Korrelation zu bestimmen; so ergibt sich eine ganze Funktion. Unterscheiden sich Mittelwert und Varianz der verschiedenen Segmente nicht relevant, so können jeweils die globalen Werte \bar{y} und s_Y^2 eingesetzt werden. Zudem wird zur Normierung generell der Faktor $1/N$ verwendet; dies hat theoretische Vorteile.

Definition 2.4 (Empirische Autokovarianz- und Autokorrelationsfunktion)

Die *empirische Autokovarianzfunktion* (c_τ) einer Zeitreihe (y_t) sowie ihre *empirische Autokorrelationsfunktion*, kurz ACF, (r_τ) sind definiert durch

$$c_\tau = \frac{1}{N} \sum_{t=1}^{N-\tau} (y_{t+\tau} - \bar{y})(y_t - \bar{y}), \quad (\tau \geq 0) \quad (2.5)$$

$$r_\tau = \frac{c_\tau}{c_0} \quad (\tau \geq 0). \quad (2.6)$$

Für $\tau < 0$ wird $c_\tau = c_{-\tau}$ und $r_\tau = r_{-\tau}$ gesetzt.

Wegen $c_0 = s_Y^2$ entspricht r_τ im Wesentlichen dem Korrelationskoeffizienten. Somit misst r_τ für $\tau > 0$ die lineare Abhängigkeit zwischen den entsprechend weit auseinander liegenden Zeitpunkten. Da sich die beide Funktionen c_τ und r_τ nur um einen Faktor unterscheiden, wird auch für die Autokovarianzfunktion das Kürzel ACF verwendet.

Beispiel 2.5 (Berechnung von Kovarianzen für verschiedene Lags)

Für eine Reihe der Länge $N = 10$ soll die Bestimmung der Kovarianzen beispielhaft illustriert werden. Die Reihe ist bereits *zentriert*, d. h. es ist der Mittelwert von allen einzelnen Zeitreihenwerten abgezogen worden. Deutlich wird insbesondere, dass mit größer werdendem Lag immer weniger Summanden in die Berechnung der zugehörigen Kovarianz eingehen.

t	y_t	(y_t, y_t)	(y_{t+1}, y_t)	(y_{t+2}, y_t)	(y_{t+3}, y_t)
1	0	(0, 0)	(-2, 0)	(-1, 0)	(3, 0)
2	-2	(-2,-2)	(-1,-2)	(3,-2)	(1,-2)
3	-1	(-1,-1)	(3,-1)	(1,-1)	(-2,-1)
4	3	(3, 3)	(1, 3)	(-2, 3)	(-2, 3)
5	1	(1, 1)	(-2, 1)	(-2, 1)	(5, 1)
6	-2	(-2,-2)	(-2,-2)	(5,-2)	(1,-2)
7	-2	(-2,-2)	(5,-2)	(1,-2)	(-3,-2)
8	5	(5, 5)	(1, 5)	(-3, 5)	
9	1	(1, 1)	(-3, 1)		
10	-3	(-3,-3)			
Kovarianz zum Lag τ :		5.8	-0.4	-4.2	0.3

Beispiel 2.6 (Zeitreihen und Autokorrelationsfunktionen)

Welches Erscheinungsbild eine empirische ACF haben wird, kann mit ein wenig Übung vorab zumindest in etwa gesagt werden. In einer Zeitreihe vorhandene Strukturen führen zu Strukturen der ACF. Um Sie als Leserin/den Leser zu animieren, dies selbst zu erkunden, sind die Zuordnungen der Autokorrelationsfunktionen zu den Reihen in der folgenden Abbildung durcheinander gebracht worden. Versuchen Sie die richtigen Entsprechungen selbst zu bestimmen, bevor Sie weiterlesen.

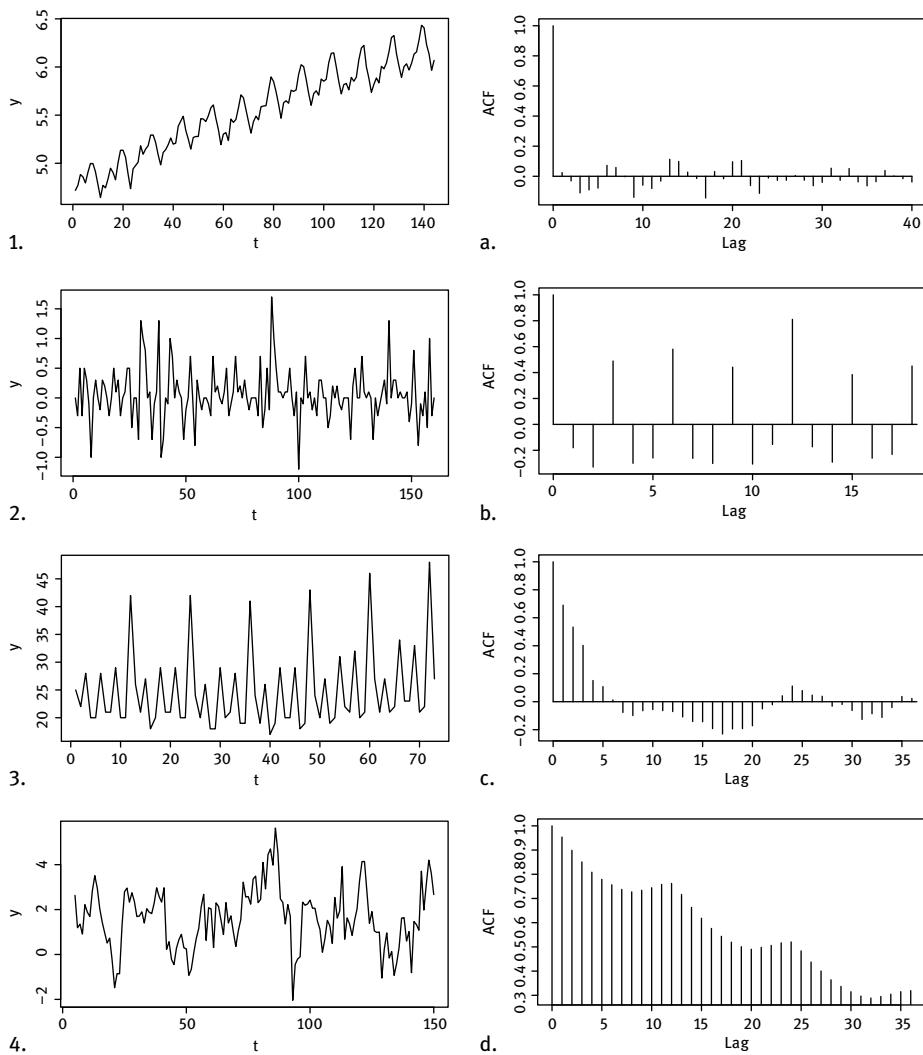


Abb. 2.4: Zeitreihen und Autokorrelationsfunktionen

Die Reihe 1. weist ein ansteigendes Niveau und eine Saison auf. Das ansteigende Niveau bewirkt, dass zeitlich näher beieinander liegende Werte auch größtmäßig dichter beieinander liegen. Die ACF wird also nur langsam abnehmen. Zudem sind die um jeweils eine Saisonperiode voneinander weg liegenden Zeitreihenwerte in der Tendenz alle gleichmäßig erhöht bzw. niedriger. Folglich wird dies auch in der ACF zu sehen sein. Die ACF „d.“ zeigt dieses Verhalten.

Die Zeitreihe 2. zeigt keine erkennbare Regelmäßigkeit. Große und kleine Werte wechseln sich irregulär ab, die extremen Werte zeigen keine regelmäßigen Abstände. Daher ist die ACF „a.“ passend.

Bei der Reihe 3. sieht man keine wesentliche Veränderung im Niveau, aber eine deutliches zyklisches Verhalten. Durch einfaches Abzählen sieht man, dass eine Zykluslänge drei beträgt. Dann gibt es höhere Gipfel, die jeweils zwölf Zeiteinheiten voneinander weg liegen. Also sollte – entsprechend der Diskussion zur Reihe 1. – die ACF bei den Lags drei, sechs usw. erhöht sein und zusätzlich bei zwölf noch größer sein. Das passt auf „b.“.

Bei der Reihe 4. gehen die Veränderungen nicht so abrupt vor sich wie bei 2., zeitlich benachbarte Werte sind noch näher beieinander. Sonst ist keine regelmäßige Struktur zu sehen. Hierzu korrespondiert die ACF „c.“.

Stationarität

Die Bestimmung der Kovarianzen bzw. Korrelationen gemäß (2.5) und (2.6) ist natürlich nur dann sinnvoll, wenn sich die Abhängigkeitsstruktur über die Zeit nicht ändert. Zusammen mit der bei der Ableitung unterstellten Konstanz von Mittelwerten und Varianzen aus zeitlichen Segmenten bedeutet dies für die zugrunde liegende Folge von Zufallsvariablen, dass alle Y_t den gleichen Erwartungswert μ sowie die gleiche Varianz σ^2 haben und dass auch die Korrelationen $\text{Corr}(Y_t, Y_{t+\tau})$ nur von dem Lag τ und nicht von dem Zeitpunkt t abhängen.

Definition 2.7 (Stationarität)

Ein stochastischer Prozess (Y_t) heißt *mittelwertstationär*, wenn $E(Y_t) = \mu$, *varianzstationär*, wenn $\text{Var}(Y_t) = \sigma^2$, und *kovarianzvarianzstationär*, wenn $\text{Cov}(Y_{t+\tau}, Y_t) = \gamma_\tau$, jeweils für alle t, τ . Er heißt *(schwach) stationär*, wenn alle drei Bedingungen erfüllt sind.

Die *schwache Stationarität* stellt nur Anforderungen an die ersten und zweiten Momente. Die *strenge Stationarität* verlangt dagegen die Gleichheit der gemeinsamen Verteilungen jedes Ausschnittes von Prozessvariablen. Dabei wird die Existenz der Momente nicht unterstellt. Somit ist die beiden Konzepte zwar überlappend, aber keines der beiden Konzepte ist echt in dem anderen enthalten. Für das Weitere reicht die Forderung der schwachen Stationarität, so dass etwas lax einfach von Stationarität gesprochen wird.

Definition 2.8 (Theoretische Autokovarianz- und Autokorrelationsfunktion)

Bei einem stationären stochastischen Prozess (Y_t) wird die Funktion

$$\gamma_\tau = \text{Cov}(Y_{t+\tau}, Y_t) \quad (\tau = 0, \pm 1, \pm 2, \dots) \quad (2.7)$$

des Lags τ als *(theoretische) Autokovarianzfunktion* bezeichnet. Sie stimmt für $\tau = 0$ mit der Varianz $\text{Var}(Y_t)$ überein. Die *(theoretische) Autokorrelationsfunktion*, kurz *ACF*, ist

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} \quad (\tau = 0, \pm 1, \pm 2, \dots). \quad (2.8)$$

Für die ACF gilt die Beziehung $\gamma_\tau = \gamma_{-\tau}$; damit ist auch $\rho_\tau = \rho_{-\tau}$.

Wie bei den empirischen Größen wird auch die Autokovarianzfunktion mitunter als ACF bezeichnet. Da sich die Funktionen (γ_τ) und (ρ_τ) nur um den Faktor $1/\gamma_0$ unterscheiden, ist die gelegentliche Verwendung des Akronyms ACF auch für die Autokovarianzfunktion unproblematisch.

Soll im folgenden die Zugehörigkeit der eingeführten Größen zu einem speziellen Prozess, etwa (Y_t) , herausgehoben werden, so wird $\gamma_Y(\tau)$ für γ_τ , $\rho_Y(\tau)$ für ρ_τ und μ_Y sowie σ_Y^2 für μ bzw. σ^2 geschrieben.

Beispiel 2.9 (White-Noise- und Random Walk-Prozesse)

White-Noise-Prozesse (ε_t) sind stationär, sofern Erwartungswert und Varianz existieren; es ist ja $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ für $t \neq s$. Random-Walk-Prozesse sind dagegen nicht stationär. Random-Walks ohne Drift sind varianzinstationär, die Varianz ist abhängig von der Zeit:

$$\text{Var}(Y_t) = \sum_{i=0}^t \text{Var}(\varepsilon_i) = (t+1)\sigma_\varepsilon^2 .$$

Random-Walks mit Drift sind zudem mittelwertinstationär:

$$\mathbb{E}(Y_t) = \mu + t\mu_\varepsilon .$$

Ob die Stationarität eines zugrunde liegenden Prozesses unterstellt werden kann, lässt sich oft schon am Beginn der Analyse einer Zeitreihe am Plot der Reihe erkennen. Die theoretischen Eigenschaften sollten sich in der empirischen Reihe widerspiegeln. Es sollte eine angenäherte Konstanz von Niveau, Streuung und Abhängigkeitsstruktur zu beobachten sein. Andersherum liegt keine Stationarität vor, wenn die Reihe eines der genannten Charakteristika Trend, Saison oder gar eine mit dem Niveau ansteigende Varianz aufweist.

Ein einfacher Check der Stationarität besteht daher darin, die Zeitreihe einige in nicht zu kurze Segmente zu zerlegen und für die Segmente die entsprechenden Maßzahlen zu berechnen. Dass das Niveau einer Zeitreihe sich nicht wesentlich ändert, dass sie insbesondere keinen Trend aufweist, zeigt sich, wenn die zugehörigen arithmetischen Mittel $\bar{y}_i = \sum_{t=i+1}^{i+r} y_t/r$ nicht sehr verschieden sind. Eine Konstanz der Streuung zeigt sich anhand entsprechend geringer Unterschiede bei den Varianzen von Reihensegmenten, $s_i^2 = \sum_{t=i+1}^{i+r} (y_t - \bar{y}_i)^2/r$. Natürlich ist hier die Beurteilung, was „nicht sehr unterschiedlich“ heißt, subjektiv und erfahrungsabhängig. Dies ist auch nur ein heuristischer Check.

Beispiel 2.10 (Regen in LA – Fortsetzung von Seite 13)

Die Zeitreihe REGEN macht einen offensichtlich mittelwert- und varianzstationären Eindruck. Es ergeben sich die folgenden empirischen Kenngrößen:

$$\bar{y} = 15.01 , \quad s_Y^2 = 47.06 .$$

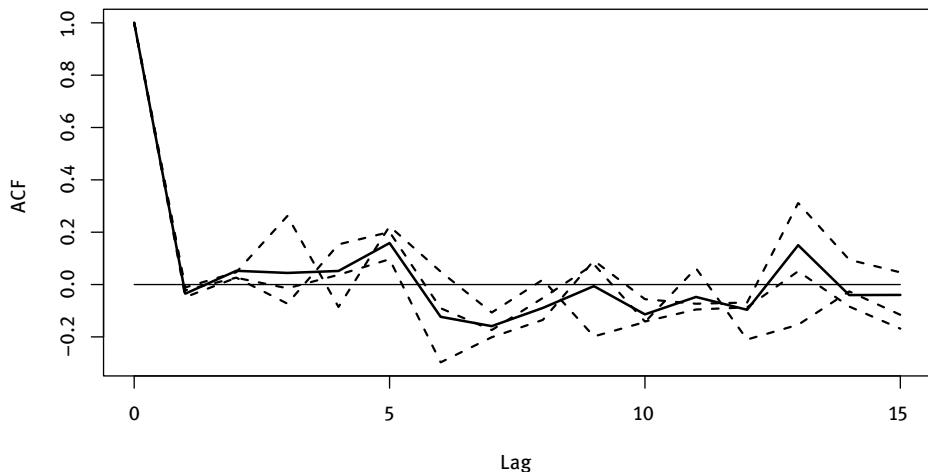


Abb. 2.5: ACF der Reihe REGEN; ganze Reihe (—) und drei Segmente (- - -)

Die ACF aus Segmenten wie auch aus der gesamten Zeitreihe zeigt, dass die Autokorrelationen nur wenig von null abweichen, siehe Abbildung 2.5. Folglich ist die schwache Stationarität als erfüllt zu betrachten. Allerdings lassen die Werte der ACF Zweifel darüber aufkommen, ob in dem der Reihe zugrunde liegenden Prozess wesentliche Abhängigkeiten enthalten sind.

R-Code:

```

1 data(RAINFALL)
2 y <- RAINFALL
3 mean(y)
4 var(y)
5 out <- statcheck(y, 3)
6 a <- acf(y, lag.max=15, lwd=3, plot=F)
7 lines(c(0:15),a$acf,type="l",lwd=2)

```

Die Berechnung der deskriptiven Maßzahlen ist elementar. Um die Kovarianzstationarität zu überprüfen, wird die Funktion `statcheck` eingesetzt; sie ist in `tsapp` enthalten. Diese Funktion erstellt auch die Grafik der ACFs für die Segmente. Es braucht nur noch die grafische Darstellung der ACF der Gesamtreihe. Dies erfolgt in den Zeilen 6 und 7.

Beispiel 2.11 (Anzahlen von Luchspelzen)

Die jährlichen Anzahlen der Pelze von Luchsen, die in den Jahren 1821 bis 1934 in Kanada im McKenzie River District gefangen wurden, werden üblicherweise logarithmisch transformiert (gemäß \log_{10}). Die grafische Darstellung dieser transfor-

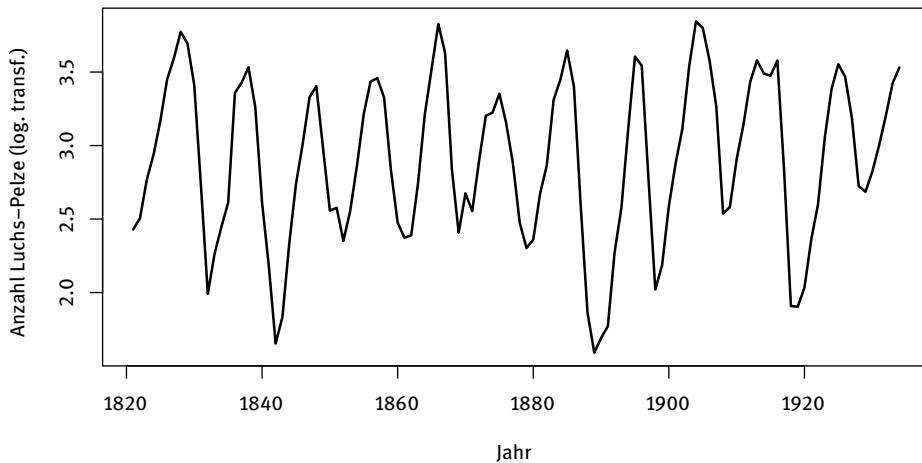


Abb. 2.6: Die Zeitreihe LUCHS

mierten Reihe, die mit LUCHS bezeichnet wird, zeigt, dass keine Veränderung des Niveaus stattfindet. Auch das Streuungsverhalten ist stabil.

Der Stationaritätscheck ergibt bei drei Segmenten:

segment-mean	segment-sd
2.888403	0.5478154
2.823095	0.5815392
2.999493	0.5456087

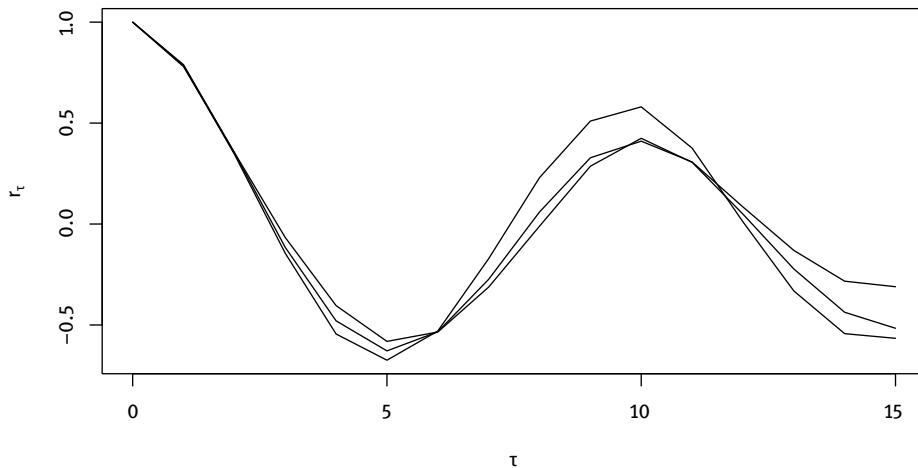


Abb. 2.7: ACFs dreier Segmente der Reihe LUCHS

Hier hat man eine recht gute Übereinstimmung der Maßzahlen der drei Segmente. Eine andere Wahl der Segmente kann jedoch zu größeren Unterschieden führen, insbesondere wenn die Segmente so kurz sind, dass die Anzahlen der Gipfel bzw. Täler in den Segmenten unterschiedlich ist.

Die Reihe weist ein relativ zyklisches Verhalten auf. Die Gipfelpunkte sind alle etwa 9.5 Jahre voneinander entfernt. Das drückt sich auch in der Autokorrelationsfunktion aus. Sie zeigt für Lags um $\tau \approx 10, 20, \dots$ erhöhte Werte. Bei $\tau \approx 5, 15, \dots$ sind die Autokorrelationen entsprechend stark negativ. In der Mitte zwischen zwei Hochpunkten der Reihe liegt ja jeweils eine Talsohle.

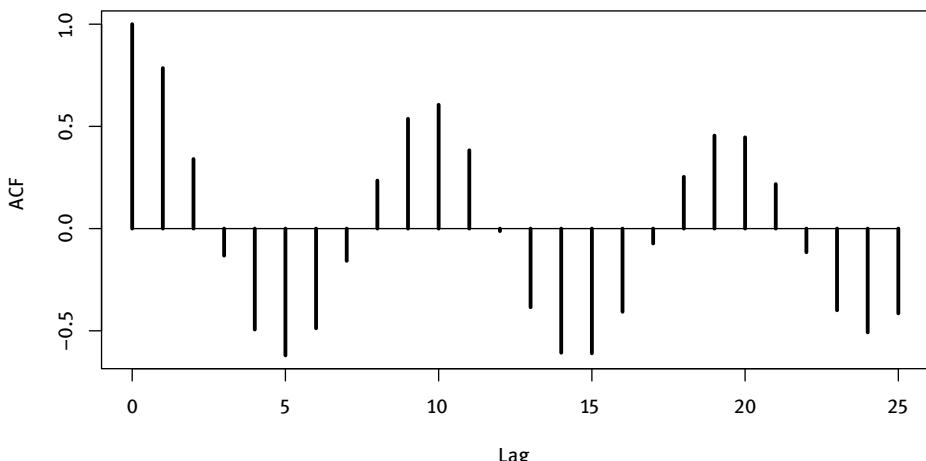


Abb. 2.8: ACF der Reihe LUCHS

R-Code:

```

1 data(LYNX)
2 y <- LYNX
3 y <- log10(y)
4 plot(y)
5 statcheck(y,3)
6 acf(y,lag.max=25,ci=0)

```

Der Code zeigt im Verhältnis zu dem des letzten Beispiels nichts wesentlich Neues. Da in Zeile 6 `y` Zeitreihenattribute besitzt, wird in R die Abszisse auf die zugehörige Zeiteinheit skaliert und nicht auf die Anzahl der Zeitpunkte bezogen. Dies wirkt sich hier allerdings wegen des Zeitreihenattributs `frequency=1` nicht aus.

Schätzen der Kennfunktionen

Modelle für stationäre Reihen bilden die Basis der Methoden zur Analyse von Zeitreihen. Verschiedene Modelle erlauben unterschiedliche Parametrisierungen der Abhängigkeitsstruktur. Diese versucht man dann mit Hilfe der empirischen ACF zu erkennen und daraufhin die Parameter des entsprechenden Modells zu schätzen. In diesem Sinne ist der folgende Satz fundamental. Er sagt, dass die empirischen Mittelwerte, Varianzen und Autokorrelationen unter milden Bedingungen sinnvolle Schätzer für die theoretischen Größen des zugrunde liegenden Prozesses sind. Die Basis ist die Stationarität; ohne sie ist es unmöglich, ohne weitere Annahmen an die theoretischen Größen überhaupt heranzukommen. Zudem muss die Abhängigkeit zwischen den Zeitreihenvariablen mit wachsendem zeitlichem Abstand genügend schnell abnehmen, um geeignete Verallgemeinerungen der Aussagen für unabhängige Zufallsvariablen zuzulassen. Dies ist mit dem Begriff der *Ergodizität* verknüpft. Hier wird die Ergodizität durch die Forderung der absoluten Summierbarkeit der Autokovarianzfunktion, $\sum |\gamma_\tau| < \infty$, sicher gestellt. Diese Forderung ist etwa für $\gamma_\tau = 1/\tau$ nicht erfüllt. Die Werte $1/\tau$ gehen nicht schnell genug gegen null, es ist $\sum_{\tau=1}^{\infty} 1/\tau = \infty$; dagegen gilt $\sum_{\tau=1}^{\infty} a^\tau = 1/(1-a)$ für $0 < a < 1$, so dass diese ACF absolut summierbar ist.

Die Aussagen des folgenden Satzes sind asymptotischer Natur, geben also Eigenschaften von Funktionen der Zeitreihenvariablen für $N \rightarrow \infty$ an. Diese Beschränkung auf asymptotische Aussagen ist typisch für das gesamte Gebiet der Zeitreihenanalyse. Für ‚endliche‘ Reihen, d. h. solche eher geringerer Länge, sind statistische Aussagen über Funktionen von Zeitreihenvariablen in den seltensten Fällen verfügbar. Weiterhin werden der Einfachheit halber nur Gauß-Prozesse betrachtet; diese Forderung kann um den Preis größeren technischen Aufwandes abgeschwächt werden.

Gekennzeichnet werden asymptotische Eigenschaften wie üblich durch einen Punkt über dem Gleichheitszeichen oder dem jeweils geeigneten Symbol. $\text{Var}(\sqrt{N}(\bar{Y}_N - \mu)) \doteq \sigma^2$ bedeutet also, dass die Varianz für $N \rightarrow \infty$ gegen σ^2 strebt. In praktischen Umsetzungen wird schon von einer hinreichenden Gleichheit bei genügend langen Reihen ausgegangen.

Satz 2.12 (Asymptotik von Mittelwert, Kovarianzen und Korrelationen)

Sei (Y_t) ein stationärer Gauß-Prozess mit Erwartungswert $\mu = E(Y_t)$ und ACF $\gamma_\tau = E((Y_t - \mu)(Y_{t+\tau} - \mu))$. (γ_τ) sei absolut summierbar, d. h. $\sum |\gamma_\tau| < \infty$. Dann gilt:

- $\text{Var}(\sqrt{N}(\bar{Y}_N - \mu)) \doteq \gamma_0 + 2 \sum_{\tau=1}^{\infty} \gamma_\tau$.
- $\sqrt{N}(\bar{Y}_N - \mu)$ ist asymptotisch normalverteilt mit Erwartungswert null und oben angegebener Varianz.
- Für die empirischen Kovarianzen $\hat{\gamma}_N(\tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} (Y_t - \bar{Y}_N)(Y_{t+\tau} - \bar{Y}_N)$ gilt asymptotisch:

$$E(\hat{\gamma}_N(\tau)) \doteq \gamma_\tau, \quad \text{Var}(\hat{\gamma}_N(\tau)) \doteq 0.$$

- Für die empirischen Korrelationen $\hat{\rho}_N(\tau) = \hat{y}_N(\tau)/\hat{y}_N(0)$ gilt asymptotisch:

$$E(\hat{\rho}_N(\tau)) \doteq \rho_\tau, \quad \text{Var}(\hat{\rho}_N(\tau)) \doteq 0,$$

und $\sqrt{N}(\hat{\rho}_N(1) - \rho_1), \dots, \sqrt{N}(\hat{\rho}_N(k) - \rho_k)$ sind für jeweils festes k asymptotisch gemeinsam normalverteilt.

Asymptotisch sind die empirischen Autokovarianzen und Autokorrelationen unverfälscht. In kürzeren Zeitreihen kann insbesondere bei den Autokorrelationen der Bias, d. h. die Differenz zwischen dem Erwartungswert des Schätzers und dem zu schätzenden Parameter, also von $\hat{\rho}_N(\tau)$ und ρ_τ , aber erheblich sein.

Beispiel 2.13 (Asymptotik der Korrelationen bei White-Noise)

Für einen White-Noise-Prozess erhält man

$$E(\hat{\rho}_N(\tau)) \doteq -\frac{1}{N} \tag{2.9}$$

$$\text{Cov}(\hat{\rho}_N(\tau), \hat{\rho}_N(\kappa)) \doteq \begin{cases} 1/N & \text{für } \tau = \kappa \\ 0 & \text{für } \tau \neq \kappa. \end{cases} \tag{2.10}$$

Bei den Autokovarianzen ist der Bias auf die Verwendung des Faktors $1/N$ anstelle von $1/(N-\tau)$ zurückzuführen. Dennoch wird dieser Faktor verwendet, weil er zwei wesentliche Vorteile bietet. Zum einen werden die Autokovarianzschätzungen mit größerem Lag immer unsicherer, da ja die Anzahl der eingehenden Summanden entsprechend geringer wird. „Unsicher“ bedeutet dabei eine große Varianz; für $\tau = N-1$ wird ja überhaupt nicht mehr gemittelt, es gibt nur noch einen Summanden. Somit wird durch die Verwendung von $1/N$ zwar ein gewisser Bias induziert, aber dies wird durch die starke Reduzierung der Varianz bei weitem kompensiert. (Sofern die Voraussetzung $y(\tau) \rightarrow 0$ erfüllt ist.) Zudem ist die Folge $\hat{y}(\tau)$ nicht-negativ definit. Dies stellt u. a. sicher, dass die Varianzschätzungen aller Linearkombinationen von Zeitreihenvariablen größer oder gleich null sind, wenn sie auf $\hat{y}(\tau)$ basieren. Als Konsequenz für die Praxis ergibt sich daraus, dass die empirische ACF nur für kleinere Lags bestimmt werden sollte. Auf Box & Jenkins (1976) geht die Empfehlung zurück, $N/4$ als höchstes Lag zu wählen.

Eine Konsequenz aus der mit dem Lag wachsenden Unsicherheit der Autokovarianzen betrifft die Bestimmung der Standardabweichung des arithmetischen Mittels. Für die Varianz von \bar{X}_N aus einer Reihe der Länge N gilt:

$$\text{Var}(\bar{Y}_N) = \frac{1}{N^2} \sum_{i,j=1}^N \text{Cov}(Y_i, Y_j) = \frac{1}{N} \sum_{|h|<N} \left(1 - \frac{|h|}{N} \right) \gamma_h. \tag{2.11}$$

Will man diese Varianz schätzen, indem einfach γ_h durch $\hat{\gamma}_h$ ersetzt wird (Plug-In-Schätzung), so ist der resultierende Schätzer nicht konsistent. Damit ist auch die na-

heligende Bestimmung von approximativen Konfidenzintervallen für μ gemäß

$$\bar{Y}_N - z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\bar{Y}_N)} \leq \mu \leq \bar{Y}_N + z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\bar{Y}_N)}, \quad (2.12)$$

obsolet, wenn $\widehat{\text{Var}}(\bar{Y}_N)$ gemäß (2.11) so bestimmt wird, dass \hat{y}_h für y_h eingesetzt wird. Gleiches gilt für den entsprechenden Test auf der Basis der Teststatistik $T = (\bar{Y}_N - \mu_0) / \sqrt{\widehat{\text{Var}}(\bar{Y}_N)}$.

Bootstrap

Einen Ausweg bietet die *Bootstrap-Methode*. Sie wurde 1979 als computergestütztes Simulationsverfahren zur statistischen Inferenz vorgeschlagen. Inzwischen ist es ein Standardverfahren. Die Basisidee des Bootstrap für unabhängige Beobachtungen besteht darin, zur Schätzung der Varianz eines Schätzers $\hat{\theta}$ aus einer vorliegenden Stichprobe wiederholt Stichproben gleichen Umfangs n mit Zurücklegen zu ziehen. Für jede dieser B Bootstrap-Stichproben wird der Wert $\hat{\theta}_b^*$ der entsprechenden Statistik berechnet. Die Stichprobenvarianz dieser B Werte bildet dann die Schätzung der Varianz von $\hat{\theta}$. Bei Zeitreihendaten geht dieser einfache Weg fehl; die Abhängigkeit der Beobachtungen führt zu inkonsistenten Schätzungen. Das Bootstrap wird daher dahingehend modifiziert, dass nicht einzelne Beobachtungen zufällig aus der vorliegenden Zeitreihe ausgewählt werden, sondern jeweils zusammenhängende Blöcke von Zeitreihenwerten. Es werden also aus einer vorliegenden Zeitreihe $\mathbf{y} = (y_1, \dots, y_N)$ Bootstrap-Reihen (y_1^*, \dots, y_l^*) erzeugt mit (mit $l = k \cdot b \approx N$), indem k Blöcke $B_i = (y_i, \dots, y_{i+b-1})$ der Länge b zufällig und mit Zurücklegen aus der Ausgangsreihe gezogen werden. Die interessierende Statistik wird dann jeweils aus dieser neuen zusammengesetzten Bootstrap-Reihe y_1^*, \dots, y_l^* bestimmt. Ein darauf basierendes Konfidenzintervall ist dann etwa von der Form (2.12), wobei nun für $\widehat{\text{Var}}(\bar{Y}_N)$ die so erhaltene Schätzung der Varianz von \bar{Y}_N eingesetzt wird. Für einen Test kann einfach das Konfidenzintervall verwendet werden oder es ist diese Schätzung entsprechend in die Teststatistik einzusetzen.

Weitere Ausführungen zu Bootstrap-Verfahren in der Zeitreihenanalyse finden sich etwa bei Lahiri (2003) und bei Politis (2003). Zur schwierigen Wahl der Blocklänge sei auf die Arbeiten Politis & White (2004) sowie Politis, White & Patton (2009) hingewiesen.

Politis & White (2004) geben den folgenden Algorithmus zur Generierung einer Block-Bootstrap-Reihe mittels des *zirkulären Bootstrap* an:

1. Starte mit einer zirkulären Verlängerung der Reihe \mathbf{x} , also der Bildung von (\mathbf{x}, \mathbf{x}) .
2. Bestimme die Blocklänge b .
3. Ziehe zufällig mit Zurücklegen aus den Indizes $\{1, 2, \dots, N\}$ die Startpunkte i_0, i_1, \dots der Blöcke.
4. Konstruiere eine Bootstrap-Reihe $(y_1^*, y_2^*, \dots, y_N^*)$, indem für $m = 0, 1, \dots, N-1$ $y_{mb_m+j}^* = y_{i_m+j-1}$ für $j = 1, 2, \dots, b$ gebildet wird.

Beispiel 2.14 (Vergleich zweier Finanzstrategien)

Eine Portfolio-Insurance-Strategie, die Reihe SYNTHETICPUT soll mit einer Geldmarkt-Strategie als Benchmark, der Reihe CASH, verglichen werden, siehe Dichtl & Drobetz (2011). Dazu werden die ersten 250 Tage mit historischen Marktinformationen genommen und die jeweilige Strategie und deren Nutzenwert berechnet. Dann wird das Zeitfenster immer um einen Tag weiter geschoben. Daraus ergeben sich dann insgesamt rund 7 500 Nutzenwerte. Durch die Überlappung der Zeitfenster ist die Autokorrelation sehr hoch. Um die Strategie mit dem Cash-Benchmark zu vergleichen, wird nun die Differenz $y_t = \text{SYNTHETICPUT} - \text{CASH}$ der beiden Reihen betrachtet. Ist deren Niveau signifikant größer als null, so ist SYNTHETICPUT vorteilhafter.

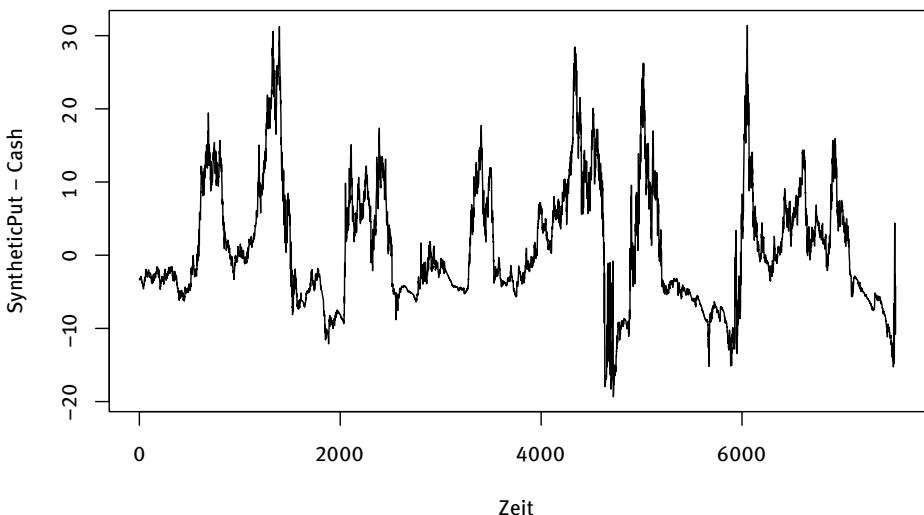


Abb. 2.9: Differenz der Entwicklung zweier Finanzstrategien

Der Test wird zum Niveau $\alpha = 0.05$ durchgeführt, indem ein zweiseitiges 0.9-Konfidenzintervall bestimmt wird. Der Mittelwert der Differenz beträgt $5.1114 - 4.1793 = 0.9321$, die mittels einem blockweisen Bootstrap mit 1000 Bootstrap-Stichproben geschätzte Standardabweichung von \bar{x} ist 1.104. Als Blocklänge wurde 230 gewählt. Das zugehörige asymptotische Konfidenzintervall ist mit $z_{0.95} = 1.6449$ folglich $[-0.884; 2.748]$. Der einseitige Test der Nullhypothese $H_0: \mu \leq 0$ gegen die Alternative $H_1: \mu > 0$ führt nicht zur Ablehnung von H_0 .

R-Code:

```

1 data(FINANCE)
2 y <- FINANCE[,4] - FINANCE[,5]
3 library(np)
4 b <- b.star(y)
5 b <- ceiling(b[2])
6 Boot <- 1000
7 alpha <- 0.1
8 n <- length(y)
9 theta <- mean(y)
10 k <- n%/%b
11 yb <- rep(0,b*k)
12 Index <- c(1:(n+1-b))
13 tstar <- rep(0,Boot)
14 bo <- 1
15 while (bo <= Boot){
16   a <- sample(Index,k,replace=TRUE)
17   for (i in c(1:k)){ yb[(i-1)*b+c(1:b)] <- y[c(a[i]):(a[i]+b-1)] }
18   tstar[bo] <- mean(yb)
19   bo<- bo+1 }
20 v <- var(tstar)
21 z <- qnorm(1-alpha/2)
22 KI <- c(theta-z*sqrt(v),theta+z*sqrt(v))
23 print(KI)

```

Mit der Funktion `b.star` des Paketes `np` wird die für diese Reihe optimale Blocklänge bestimmt (Zeilen 4 bis 6). In Zeile 7 wird die Anzahl der Bootstrap-Reihen festgelegt, in Zeile 8 das Niveau des zweiseitigen Konfidenzintervalls. `b` ist die Länge der Blöcke. In Zeile 16 beginnt die Bootstrap-Schleife. Zuerst werden die Startwerte für die zusammenzufügenden Blöcke per Zufallsauswahl bestimmt. Dann werden die Blöcke ausgewählt und zusammengesetzt. Der zugehörige Mittelwert wird berechnet und gespeichert. Aus den Bootstrap-Mittelwerten wird dann die Standardabweichung und damit das Konfidenzintervall berechnet.

Wegen der starken Asymmetrie der Reihe wäre ein nichtparametrisches Konfidenzintervall angeraten. Das wäre dann signifikant. Dazu siehe Davidson & Hinkley (1997).

2.2 Das Komponentenmodell

Viele der in der Praxis vorkommenden Reihen sind von Vornherein nicht als Realisationen stationärer Prozesse anzusehen. Dafür kann jede der weiter oben genannten Eigenschaften, wie Vorliegen eines Trends, einer Saison oder einer nichtkonstanten Varianz als Grund in Frage kommen. Ein häufig eingeschlagener Weg bei der Analyse solcher nichtstationärer Zeitreihen basiert auf der Vorstellung, dass sich die Zeitreihe aus unterschiedlichen Komponenten zusammensetzt.

Definition 2.15 (Komponentenmodelle)

Komponentenmodelle für (ökonomische) Zeitreihen gehen aus von Zerlegungen der Form

$$Y_t = T_t + S_t + \varepsilon_t \quad (\text{additives Modell}) ; \quad (2.13)$$

$$Y_t = T_t \cdot S_t + \varepsilon_t \quad (\text{quasimultiplikatives Modell}) ; \quad (2.14)$$

$$Y_t = T_t \cdot S_t \cdot \varepsilon_t \quad (\text{multiplikatives Modell}) . \quad (2.15)$$

Dabei bezeichnet (Y_t) den beobachteten Prozess, T_t erfassst die langfristige Veränderung des Mittels, die *glatte Komponente* oder den *Trend*, und S_t ist die *Saisonkomponente*. ε_t ist dann der *Rest*, der die Abweichungen von den beiden Komponenten beinhaltet.

In weiteren Varianten des Komponentenmodells werden noch zusätzliche Komponenten hinzugefügt. So etwa eine *Kalenderkomponente*, die insbesondere die unterschiedlichen Anzahlen von Tagen pro Monat erfassen soll. In seinen verschiedenen Varianten stellt das in der Definition formulierte Modell die Basis für die Ausführungen in den beiden folgenden Abschnitten dar.

Die additive Variante wird in der Regel gewählt, wenn die Saisonausschläge über den gesamten Beobachtungsbereich in etwa gleich stark sind, siehe die Teilgrafik I in der Abbildung 2.10. Falls andererseits die Reihe einen Trend und mit dem Niveau zunehmend größere Ausschläge der Saison aufweist, ist ein multiplikatives oder quasimultiplikatives Komponentenmodell angebracht, vergleiche Teilgrafiken II und III. Beim rein multiplikativen Modell werden auch die Streuungsausschläge mit dem Niveau stärker, siehe III. Beim quasimultiplikativen Modell bleiben die Abweichungen von den deterministischen Komponenten moderat, bei größerem Niveau sieht man die auf die Restkomponente zurückgehenden Abweichungen immer weniger deutlich, vgl. II.

Es gibt keine eindeutige Spezifikation der Komponenten; so werden letztlich die konkreten Modelle über die eingesetzten Verfahren definiert. Dies muss nicht unbedingt als Nachteil angesehen werden, erlaubt es doch eine größere Flexibilität.

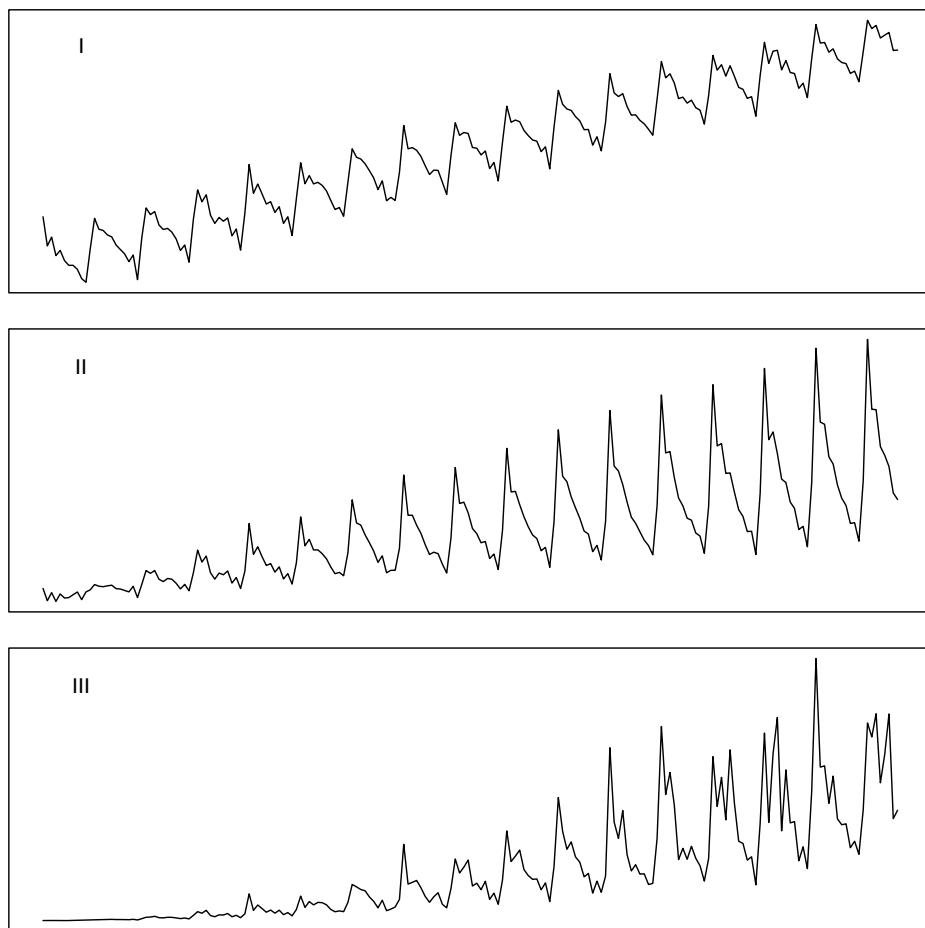


Abb. 2.10: Verschiedenartige Verbindung von Trend, Saison und Rest (siehe Text)

2.3 Deterministische Trends

In diesem Abschnitt werden verschiedene Methoden besprochen, wie der Trend oder die glatte Komponente einer Zeitreihe bestimmt werden kann. Anlass dafür bildet häufig das Bestreben, die Reihe entlang des Trends in die Zukunft zu extrapoliieren. Die andere Zielsetzung ist die Bestimmung einer trendbereinigten Reihe zur weiteren Analyse.

Trendbestimmung mittels linearer Regression

Wird ein einfaches Komponentenmodell ohne Saison unterstellt,

$$Y_t = T_t + \varepsilon_t , \quad (2.16)$$

so bietet sich oft die Modellierung der Trendkomponente durch Polynome an. Die Koeffizienten solcher *polynomialen Trends* können dann mit Regressionstechniken bestimmt werden.

Das lineare Regressionsmodell beschreibt die Abhängigkeit der Zielvariablen Y_t von den fest vorgegebenen Werten der unabhängigen Variablen oder *Regressoren* X_i , $i = 1, \dots, m$ mittels

$$Y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_m x_{mt} + \varepsilon_t \quad (t = 1, \dots, N) . \quad (2.17)$$

Die Störungen ε_t werden klassischer Weise als unabhängig und identisch verteilt mit $E(\varepsilon_t) = 0$ und $\text{Var}(\varepsilon_t) = \sigma^2$ vorausgesetzt. Da dann

$$E(Y_t) = \beta_0 + \beta_1 x_{1t} + \dots + \beta_m x_{mt} \quad (t = 1, \dots, N)$$

gilt, wird also der Erwartungswert von Y_t als lineare Funktion der Werte der Regressoren X_i modelliert. Die Regressoren können dabei von beliebiger Gestalt sein, es können etwa polynomiale Terme oder gar Kosinus- oder Sinusschwingungen sein. Schwingungen als Regressoren werden in Abschnitt 6.2 weitergehend betrachtet.

Erstes Ziel einer Regressionsanalyse ist es, die Koeffizienten β_i zu bestimmen. Damit können im zweiten Schritt Modellüberprüfungen bzw. andere Analysen durchgeführt oder auch Prognosen erstellt werden. Schätzungen der Parameter können nach der *Methode der kleinsten Quadrate* (KQ-Methode) bestimmt werden. Sie sind dann Lösung von

$$\sum_{t=1}^N (y_t - \beta_0 - \beta_1 x_{1t} - \dots - \beta_m x_{mt})^2 \stackrel{!}{=} \min_{\beta_0, \dots, \beta_m} . \quad (2.18)$$

Die Durchführung dieser Minimierungsaufgabe geschieht heutzutage unter Einsatz statistischer Programmpakete. Dabei werden i. d. R. neben den Parameterschätzungen $\hat{\beta}_i$ auch die prognostizierten Werte $\hat{y}_t = \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - \dots - \hat{\beta}_m x_{mt}$ sowie die *Residuen* $\hat{\varepsilon}_t = \hat{y}_t - y_t$ ausgegeben.

Beispiel 2.16 (Herzfrequenz)

Bei mehreren Probanden wurde wiederholt die Herzfrequenz (Schläge/Minute) bestimmt. Ziel der Untersuchung war es, die Auswirkung verschiedener Stimuli zu analysieren. Dazu war es notwendig, die schon bei Stimulus-freien Messungen stark hervortretende Atemkurve zu modellieren. (Bei der Messung der Herzfrequenz wird übrigens i. d. R. so vorgegangen, dass man die Zeit zwischen zwei Herzschlägen bestimmt und dann den Kehrwert bildet. Die Hochrechnung auf eine Minute ergibt dann die Herzfrequenz.)

Gesucht ist also ein Modell der Form „Herzfrequenz = Atemkurve + Rest“, formal:

$$Y_t = T_t + \varepsilon_t .$$

Die durch die Atemkurve bewirkten systematischen Effekte T_t werden in der Medizin üblicherweise durch ganzrationale Funktionen vierten Grades modelliert:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \varepsilon_t \quad (t = 1, \dots, N) .$$

Der KQ-Ansatz

$$\sum_{t=1}^N (y_t - \beta_0 - \beta_1 t - \beta_2 t^2 - \beta_3 t^3 - \beta_4 t^4)^2 \stackrel{!}{\min} \quad (2.19)$$

führt bei $N = 30$ Beobachtungen der vorliegenden Zeitreihe der Herzfrequenzen, der Reihe HERZSCHLAG, zu den folgenden Schätzungen:

$$\hat{\beta}_0 = 64.548, \hat{\beta}_1 = -1.544, \hat{\beta}_2 = 0.178, \hat{\beta}_3 = -0.0095, \hat{\beta}_4 = 0.00015 .$$

Die Ausgangsreihe und das angepasste Polynom sind in Abbildung 2.11 dargestellt.

R-Code:

```

1 data(HEARTBEAT)
2 y <- HEARTBEAT
3 n <- length(y)
4 t <- c(1:n)
5 t2 <- t^2; t3 <- t^3; t4 <- t^4
6 out <- lm(y ~ t + t2 + t3 + t4)
7 plot(y,xlab="Zeit",ylab="Herzschatz")
8 lines(t,out$fitted.values)

```

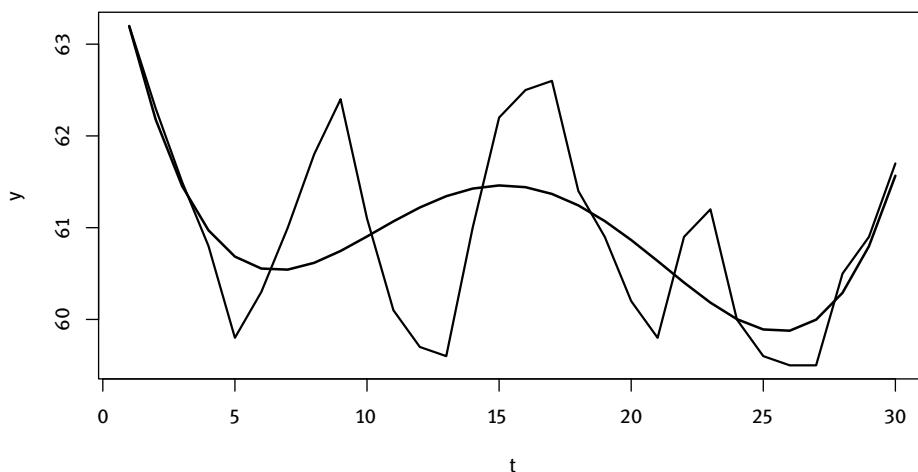


Abb. 2.11: Die Reihe HERZSCHLAG mit Trendpolynom

Lineare Regressionen werden in R mit der Funktion `lm` durchgeführt; zur Syntax siehe die Seite 323. Das Ergebnis ist eine Liste, die umfangreiche Informationen zur Regression enthält. Hier wird in Zeile 7 auf die angepassten Werte zugegriffen.

Die Ergebnisse einer KQ-Schätzung sind allein i. d. R. wenig zufriedenstellend. Zumindest möchte man sie mit Fehlermargen versehen. Auch wird man zu einer Einschätzung kommen wollen, ob das unterstellte Regressionsmodell überhaupt einen sinnvollen Ansatz darstellt.

Sei der letztgenannte Punkt zuerst betrachtet. Die relevante Größe hierfür basiert auf der *Streuungszerlegung* für die Zielgröße Y

$$SS_{\text{Total}} = SS_{\text{Modell}} + SS_{\text{Fehler}} . \quad (2.20)$$

Dabei sind die Quadratsummen (Sums of Squares):

$$\begin{aligned} SS_{\text{Total}} &= \sum_{t=1}^N (y_t - \bar{y})^2 , \quad SS_{\text{Modell}} = \sum_{t=1}^N (\hat{y}_t - \bar{y})^2 \quad \text{und} \\ SS_{\text{Fehler}} &= \sum_{t=1}^N (\hat{y}_t - y_t)^2 = \sum_{t=1}^N \hat{\varepsilon}_t^2 . \end{aligned}$$

Der Streungsanteil SS_{Modell} geht darauf zurück, dass die Regressionsfunktion nicht konstant ist. SS_{Fehler} gibt an, wie stark die Werte der abhängigen Variablen um die Regressionsfunktion schwanken. Die Zerlegung (2.20) zeigt, dass die Fehlerquadratsumme SS_{Fehler} umso kleiner ausfällt, je größer SS_{Modell} ist.

Die *Varianzanalyse-Tafel* einer Regressionsanalyse gibt die globalen Ergebnisse der Kleinstes-Quadrat-Schätzung wieder. Sie zeigt diese Zerlegung der Varianz, genauer der Summe der Abstandsquadrate vom Mittelwert, für die abhängige Variable Y .

Varianzanalyse			
	df	SS	MS
Regr. auf X_1, \dots, X_p	p	SS_{Modell}	SS_{Modell}/p
Fehler	$N - p - 1$	SS_{Fehler}	$SS_{\text{Fehler}}/(N - p - 1)$
Gesamt	$N - 1$	SS_{Total}	

Die mit df bezeichneten *Freiheitsgrade* (df = degrees of freedom) ergeben mit σ^2 multipliziert die jeweiligen Erwartungswerte der Summen der quadrierten Abweichungen.

Die MS , die mittleren Quadratsummen, erhält man dann einfach mittels Division der SS durch die zugehörigen df . MS_{Fehler} gibt eine Schätzung der Varianz der Störungen ε_t an. Ebenso gibt MS_{Modell} eine Schätzung von σ^2 , wenn das Modell wahr ist.

Andernfalls ist MS_{Modell} aufgrund der Fehlspezifikation größer. Dementsprechend ist $F = MS_{\text{Modell}}/MS_{\text{Fehler}}$ eine geeignete Prüfgröße zum Testen, ob das Modell sinnvoll ist. Ihre Verwendung erfordert allerdings die zusätzliche Voraussetzung der Normalverteilung der Fehler ε_t . Dann hat F unter der Nullhypothese eine F -Verteilung mit p Freiheitsgraden des Zählers und $N - p - 1$ Freiheitsgraden des Nenners, i.Z. $F \sim \mathcal{F}_{p, N-p-1}$. Der Test wird als F -Test bezeichnet.

Die Eignung des Modells wird gerne anhand des P -Wertes $P(F > F^*)$ ausgedrückt. Der P -Wert gibt an, wie groß die Wahrscheinlichkeit ist, dass die Statistik F einen noch extremeren Wert annimmt als den aktuell erhaltenen F^* . Ist die Wahrscheinlichkeit sehr klein, etwa < 0.05 oder < 0.01 , so gilt dies als Indikator dafür, dass das Modell wesentliches zur Erklärung der Variation der Y -Werte beiträgt. Neben den Werten der Teststatistiken werden P -Werte bei statistischen Auswertungsprogrammen standardmäßig mit ausgegeben.

Ein weiteres wichtiges Maß zur Beurteilung der Regression ist das *Bestimmtheitsmaß* R^2 . Es ist definiert als der Anteil der erklärten Varianz

$$R^2 = \frac{SS_{\text{Modell}}}{SS_{\text{Total}}} = 1 - \frac{SS_{\text{Fehler}}}{SS_{\text{Total}}}. \quad (2.21)$$

Für das *adjustierte Bestimmtheitsmaß* R_{adj}^2 werden die erwartungstreuen Varianzschätzungen ins Verhältnis gesetzt. Somit resultiert es zwar nicht genau aus der schönen Zerlegungsformel (2.20), ist aber inhaltlich adäquater:

$$R_{\text{adj}}^2 = 1 - \frac{MS_{\text{Fehler}}}{MS_{\text{Total}}} = 1 - \frac{\hat{\sigma}_\varepsilon^2}{s_Y^2}. \quad (2.22)$$

Die Schätzer $\hat{\beta}_i$ der einzelnen Regressionskoeffizienten sind erwartungstreue, sie ergeben also im Mittel den wahren Wert. Für die Beurteilung der $\hat{\beta}_i$ sind ihre Standardfehler zentral. Wird die Normalverteilung der Störungen vorausgesetzt, so ist für $i = 0, 1, \dots, p$ der Quotient $T_i = \hat{\beta}_i/\hat{\sigma}_{\hat{\beta}_i}$ unter der Hypothese $\beta_i = 0$ t -verteilt mit $N - p$ Freiheitsgraden. Eine signifikante Teststatistik deutet auf einen von null verschiedenen Parameter hin. Signifikanz bedeutet also, dass der Parameter – und somit der zugehörige Regressor – für den Modellfit bedeutsam ist. Die Signifikanz der t -Tests kann wieder an den üblicherweise mit ausgegebenen P -Werten $P(|T_i| > |T_i^*|)$ abgelesen werden. Je kleiner die Wahrscheinlichkeit ist, umso wichtiger erscheint der zugehörige Regressor für das Modell. Bei der Interpretation ist aber zu beachten, dass die P -Werte sich jeweils auf nur einen Regressor beziehen und mögliche Zusammenhänge zwischen ihnen unberücksichtigt bleiben. F -Tests, die für einzelne Koeffizienten angegeben werden, sind anders zu interpretieren. Sie messen nämlich, wie gut der Anteil der Varianz erklärt wird, der von den bereits berücksichtigten Regressoren übrig gelassen wurde. Wenn die Regressoren miteinander korreliert sind, sind diese Anteile abhängig von der Reihenfolge, in der die Regressoren aufgeführt sind.

Beispiel 2.17 (Herzfrequenz – Fortsetzung von Seite 28)

Für die Reihe HERZSCHLAG erhält man folgende Werte.

```
Analysis of Variance Table
Response: y
            Df  Sum Sq Mean Sq F value    Pr(>F)
t            1  4.1091  4.1091  5.7205  0.024610
t2           1  0.3551  0.3551  0.4943  0.488502
t3           1  0.2393  0.2393  0.3332  0.568937
t4           1  9.7853  9.7853 13.6226  0.001091
Residuals  25 17.9579  0.7183
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 64.5479075  0.9581350 67.368 < 2e-16
t            -1.5435582  0.4132042 -3.736 0.000974
t2           0.1979815  0.0529720  3.737 0.000969
t3           -0.0095119  0.0025478 -3.733 0.000979
t4            0.0001506  0.0000408  3.691 0.001091
Residual standard error: 0.8475 on 25 degrees of freedom
Multiple R-squared: 0.4465,      Adjusted R-squared: 0.358
F-statistic: 5.043 on 4 and 25 DF, p-value: 0.004045
```

In der Varianzanalysetafel ist der als erstes aufgeführte F -Test des Koeffizienten des linearen Terms signifikant, die für die quadratischen und kubischen Terme sind es nicht. Hier spielt wie erwähnt die Reihenfolge eine Rolle, in der die Regressoren aufgeführt sind. Dass dann der Koeffizient von t^4 wieder einen signifikanten Varianzanteil erklärt zeigt umso mehr seine Bedeutung.

Bei der Betrachtung der t -Tests bedeutet der hohe Wert der t -Statistik für den Achsenabschnitt nur, dass das Mittel der Y -Werte sich deutlich von null unterscheidet. Aber auch alle anderen Koeffizienten haben P -Werte, die so klein sind, dass die zugehörigen Regressoren als für das Modell bedeutsam anzusehen sind.

R-Code:

```
1 anova(out)
2 summary(out)
```

Es wird auf das Ergebnis `out` der Funktion `lm` zurückgegriffen, siehe Beispiel 2.16. Die beiden Funktionen geben die Varianzanalysetafel und die Ergebnisse für die einzelnen Koeffizienten aus. R ermittelt die erklärte Varianz für die Regressoren getrennt.

Aus statistischer Sicht ist es günstiger, ein möglichst *sparsames Modell*, d. h. eines mit möglichst wenigen Parametern zu haben. Das stellt sicher, dass das Modell nicht überangepasst ist. Überanpassung liegt vor, wenn das Modell zu den Daten sehr gut passt, aber auch Besonderheiten des einzelnen Datensatzes in das Modell einbezieht

und somit eine Generalisierung erschwert. Es gibt verschiedene Strategien, bei einem vorgegebenen Satz von Regressoren den günstigsten Teil auszuwählen. Man verzichtet auf Regressoren, wenn dies ohne wesentliche Einbuße an erklärter Varianz möglich ist. Eine einfache Reduzierung der Varianz des Fehlers kann allerdings kein adäquates Optimalitätskriterium darstellen. Die Berücksichtigung eines zusätzlichen Regressors führt nämlich zu einer Reduzierung dieser Varianz.

Eine adäquatere Möglichkeit bietet die C_p -Statistik von Mallows. Sie basiert auf der folgenden Überlegung. Ist ein reduziertes Modell ausreichend, um die abhängige Variable zu erklären, so sollte die standardisierte Summe der quadrierten Residuen des reduzierten Modells mit $q < p$ erklärenden Variablen (einschließlich Konstante) die Residualvarianz σ^2 erwartungstreu schätzen. Ist das reduzierte Modell nicht adäquat, so sollte auf Grund des Bias die zugehörige Quadratsumme eher größer ausfallen. Die C_p -Statistik vergleicht die beiden Schätzungen:

$$C_p(q) = \frac{SS(\hat{\varepsilon}|q)}{MS(\hat{\varepsilon}|p)} - (n - 2q). \quad (2.23)$$

Dabei ist $SS(\hat{\varepsilon}|q)$ die Summe der Fehlerquadrate für das Modell mit q Regressoren und $MS(\hat{\varepsilon}|p)$ die Schätzung für σ^2 im vollen Modell. Ist der Bias des reduzierten Modells null, so sollte also der Quotient auf der rechten Seite von (2.23) im Mittel gleich $n - q$ sein und die Differenz gleich q . Folglich ist das Modell am geeignetsten, für das q möglichst klein ist und $C_p(q)$ am besten mit q übereinstimmt.

Beispiel 2.18 (Herzfrequenz – Fortsetzung von Seite 32)

Für die Reihe HERZSCHLAG ergibt sich bei der Betrachtung aller TeilmODELLE die folgende Tabelle. Dabei steht in der ersten Spalte die Anzahl der Regressoren, in der zweiten bis vierten jeweils eine 1, wenn die zugehörige Potenz von t berücksichtigt ist und eine 0 sonst.

	1	2	3	4	Cp	R2
1	1	0	0	0	13.450	0.127
1	0	1	0	0	14.545	0.102
1	0	0	1	0	15.309	0.085
1	0	0	0	1	16.047	0.069
2	1	1	0	0	14.956	0.138
2	1	0	0	1	15.042	0.136
2	1	0	1	0	15.089	0.135
2	0	0	1	1	15.147	0.133
2	0	1	0	1	15.461	0.126
2	0	1	1	0	15.652	0.122
3	1	1	1	0	16.623	0.145
3	1	1	0	1	16.939	0.138
3	0	1	1	1	16.955	0.138
3	1	0	1	1	16.969	0.137
4	1	1	1	1	5.000	0.447

$C_p(q)$ ist mit Abstand am dichtesten bei q (Die Konstante ist zu beachten!), wenn alle vier Potenzen von t berücksichtigt werden. Hier nimmt auch R^2 einen im Vergleich recht großen Wert an; diese Übereinstimmung der beiden Kriterien muss nicht stets vorkommen.

R-Code:

```
1 library(leaps)
2 l1 <- leaps(x= as.matrix(cbind(t,t2,t3,t4)),y=y,method="Cp")
3 l2 <- leaps(x= as.matrix(cbind(t,t2,t3,t4)),y=y,method="r2")
4 cbind(l1$which,l1$Cp,l2$r2)
```

Der R-Code setzt den vom Beispiel 2.16 fort. Die Funktion `leaps`, die alle Regressions durchzuführen erlaubt, befindet sich in dem Paket `leaps`. Das Ergebnis ist eine Liste, die in der angegebenen Weise das Resultat enthält.

Die Regressionsanalyse hat als Hintergrund die Idee eines wahren Modells. Bei Gültigkeit des Modells sind die $\hat{\beta}_i$ sowie die anderen Größen, die als Ergebnisse anfallen, Schätzungen der Modell-Größen. Die Schätzfunktionen haben unter geeigneten Voraussetzungen Eigenschaften, auf denen die Interpretationen beruhen, die in diesem Abschnitt kurz ausgeführt wurden. Es ist daher einsichtig, dass bei Verletzungen der Modellannahmen die Interpretationen ihre Gültigkeit einbüßen können.

Die Interpretation der F - und der t -Werte basiert auf der Normalverteilung der Störungen. Diese Voraussetzung sollte daher stets überprüft werden. Das kann mit Hilfe eines *QQ-Diagrammes* geschehen. In einem solchen Diagramm werden die der Größe nach geordneten Residuen $\hat{\varepsilon}_t = y_t - \hat{y}_t$ gegen die $(t - 0.5)/N$ -Quantile der Standardnormalverteilung¹ eingetragen. Die geordneten empirischen Werte $\hat{\varepsilon}_{(t)}$ entsprechen gerade den empirischen t/N -Quantilen. Bei den theoretischen bringt man die Stetigkeitskorrektur $-0.5/N$ an, damit die empirische und theoretische Verteilungsfunktion besser übereinanderliegen. Zudem geht dann der größte Wert nicht verloren. Denn für den Wert 1 ist die Inverse der Standardnormalverteilung nicht definiert ($\Phi^{-1}(1) = \infty$). Da die empirischen Quantile vernünftige Schätzungen der theoretischen sind, sollten die resultierenden Punkte bei Gültigkeit der Normalverteilungsannahme unsystematisch um eine Gerade mit dem Achsenabschnitt $\bar{\varepsilon}$ und der Steigung $s_{\hat{\varepsilon}}$ streuen. Eine systematisch von einer Geraden abweichende Struktur der Punktwolke deutet auf eine andere Verteilung hin.

Beispiel 2.19 (Herzfrequenz und Regen in LA – Fortsetzungen von Seite 33 bzw. 17)

Die Residuen der Anpassung eines Polynoms vierter Ordnung an die Reihe HERZ-SCHLAG können als normalverteilt angesehen werden. Das QQ-Diagramm 2.12 zeigt, dass die Punkte unsystematisch und nur geringfügig um die eingezeichnete Gerade streuen.

¹ Die Verteilungsfunktion der Standardnormalverteilung, d. h. der Normalverteilung mit $\mu = 0$ und $\sigma^2 = 1$, wird mit $\Phi(\cdot)$ bezeichnet.

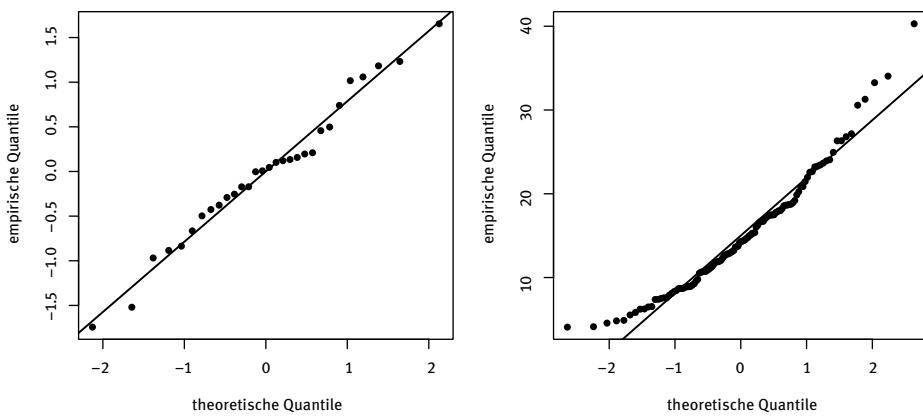


Abb. 2.12: QQ-Diagramm für die Residuen der Regression (2.19) und QQ-Diagramm für die Reihe REGEN

Die Reihe REGEN erscheint nach der bisherigen Analyse als stationär; auch sind die Autokorrelationen verschwindend gering. Es gilt noch, die Normalverteiltheit zu überprüfen. Die Punkte weichen hier recht systematisch von einer Geraden ab: Sowohl kleine als auch große empirische Quantile sind größer als unter der Normalverteilung zu erwarten wäre. Dies deutet auf eine rechtsschiefe Verteilung hin; diese resultiert aus der Beschränkung, dass die Regenmenge nicht kleiner sein kann als null.

R-Code:

```

1 qnorm(out$residuals)
2 abline(mean(out$residuals),sd(out$residuals))
3 data(RAINFALL)
4 z <- RAINFALL
5 qnorm(z)
6 abline(mean(z),sd(z))

```

Der R-Code setzt teilweise den aus Beispiel 2.16 fort. Das Ergebnis der linearen Regression der Herzschlag-Daten ist noch in `out` vorhanden. Die Hilfslinien dienen zur besseren Einschätzung der Übereinstimmung mit einem linearen Verlauf der Punkte.

Eine weitere Voraussetzung ist die Unabhängigkeit der Störungen. Sie ist entscheidend für die Größe der Standardfehler der Schätzungen und damit für die *P*-Werte. Ein Aspekt der Unabhängigkeit kann mit dem *Durbin-Watson-Test* überprüft werden. (Andere Möglichkeiten werden später besprochen.) Die Prüfgröße des Tests ist gegeben durch

$$DW = \frac{\sum_{t=2}^N (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^N \hat{\varepsilon}_t^2} = \frac{\sum_{t=2}^N \hat{\varepsilon}_t^2 + \sum_{t=1}^{N-1} \hat{\varepsilon}_t^2 - 2 \sum_{t=2}^N \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^N \hat{\varepsilon}_t^2}. \quad (2.24)$$

Der Zähler der *DW*-Statistik ist in etwa das Doppelte des Nenners, vermindert um $2 \sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}$. Ist die Korrelation aufeinander folgender Residuen in etwa null, so fällt dieser Teil weg, und der Quotient wird etwa den Wert zwei annehmen. Ist die Korrelation nahe eins, so ist $\hat{\varepsilon}_{t-1} \approx \hat{\varepsilon}_t$ und der Zähler wird in etwa zu null. Im Fall starker negativer Korrelation wird der Quotient bei dem Wert vier liegen. Insgesamt gilt also für diese Statistik

$$DW \approx \begin{cases} 0 & \text{falls die Residuen stark positiv korreliert sind} \\ 2 & \text{falls die Residuen unkorreliert sind} \\ 4 & \text{falls die Residuen stark negativ korreliert sind.} \end{cases}$$

Beispiel 2.20 (Herzfrequenz – Fortsetzung von Seite 34)

Bei der Reihe HERZSCHLAG erhält man für die Residuen des Modells $Y_t = a + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + \varepsilon_t$:

```
lag Autocorrelation D-W Statistic p-value
 1          0.5948255    0.8093489      0
Alternative hypothesis: rho != 0
```

Damit wird die Nullhypothese „die Störungen sind unkorreliert“ zu jedem üblichen Niveau abgelehnt.

R-Code:

```
1 library(car)
2 durbinWatsonTest(out)
```

Das Paket car enthält eine Funktion, mit der der Durbin-Watson-Test durchgeführt werden kann.

Wie im Beispiel werden in aller Regel die Störungen $\hat{\varepsilon}_t$ bei Regressionen mit Zeitreihendaten nicht unabhängig sein, sondern Autokorrelationen aufweisen. Unter der Annahme, dass für die Abhängigkeiten der ε_t eine weiter unten noch zu diskutierende Struktur gilt, kann dies bei der Regressionsanalyse berücksichtigt werden. Darauf wird im Abschnitt zur Regression mit autokorrelierten Störungen eingegangen.

Trendbestimmung mittels nichtlinearer Regression

Bei weitem nicht alle Trendmodelle sind von der Form, dass sie sich mittels einer linearen Regression bestimmen lassen. Bei gewissen Abweichungen ist es aber möglich, durch Transformationen der beobachteten Zeitreihe wieder auf lineare Regressionsmodelle zurückzukommen. Typische Beispiele dafür sind *Exponentialmodelle*

$$y_t = e^{\beta_0 + \beta_1 m_1(t) + \beta_2 m_2(t) + \dots + \beta_p m_p(t) + \varepsilon_t},$$

und *Potenzmodelle*

$$y_t = m_1(t)^{\beta_1} \cdot m_2(t)^{\beta_2} \cdots m_p(t)^{\beta_p} \cdot \varepsilon_t,$$

bei denen der Übergang zu logarithmierten Werten $\ln(y_t)$ ein lineares Modell ergibt.

In anderen Fällen sind die Trendfunktionen nicht auf lineare zurückzuführen. Ein Beispiel ist bereits ein Modell mit exponentiellem Wachstum, bei dem der Störterm nicht multiplikativ sondern additiv auftritt, konkret etwa:

$$y_t = e^{\beta_0 + \beta_1 t} + \varepsilon_t.$$

Generell gehören zu den eigentlich nichtlinearen Trendmodellen solche, bei denen ein nichtlinearer Anstieg mit einer Sättigungsgrenze verbunden ist. Inhaltliche Beispiele sind Wachstumsprozesse wie die Zahl der PKW oder der Mobiltelefone pro Einwohner. Formal gehören dazu die *logistische Trendfunktion*

$$m_t = \frac{\theta_1}{\theta_2 + e^{-\theta_3 t}}, \quad (2.25)$$

welche die Sättigungsgrenze θ_1/θ_2 hat und die *Gomperts-Funktion*

$$m_t = \exp(\theta_1 - \theta_2 e^{-\theta_3 t}) \quad (2.26)$$

mit der Sättigungsgrenze e^{θ_1} .

Derartige Trends führen zu dem Problem, einer Zeitreihe y_1, y_2, \dots, y_N eine nichtlineare Funktion $m(t; \boldsymbol{\beta})$ anzupassen, die von einem (mehrdimensionalen) Parameter $\boldsymbol{\beta}$ abhängt. Das Kleinstquadrat-Kriterium

$$\sum_{t=1}^N (y_t - m(t; \boldsymbol{\beta}))^2 \stackrel{!}{=} \min$$

resultiert hier in der Regel in Normalgleichungen, die nicht mehr explizit lösbar sind. Über eine approximative Linearisierung kann aber eine iterative Lösung bestimmt werden. Diese Methode wird als *Gauß-Newton-Verfahren* bezeichnet. Damit erhält man bei Unterstellung geeigneter Voraussetzungen auch approximative Konfidenzintervalle und Tests für die zu schätzenden Parameter.

Eine wesentliche Schwierigkeit bei der Anpassung einer nichtlinearen Trendfunktion ist die Bestimmung von Startwerten. Sind diese schlecht gewählt, landet man leicht bei einem lokalen Neben-Optimum oder das Verfahren bricht sogar ergebnislos ab.

Beispiel 2.21 (Bevölkerung der USA)

Der Zeitreihe USAPOP der Bevölkerungsumfänge (in Mio Einwohner) der USA von 1630 bis 2010, jeweils im zehn-Jahres-Abstand, wird eine logistische Trendfunktion angepasst.

Zuerst wird eine Startschätzung benötigt. Wie in Schlittgen & Streitberg (2001, S. 24 ff.) ausgeführt, werden die diskretisierten Wachstumsraten $r_t = (y_{t+1} - y_t)/y_{t+1} \approx \theta_3 - \frac{\theta_3}{\eta} m_t$ betrachtet, wobei $\eta = \theta_1/\theta_2$ gesetzt wurde. In dieser Bezie-

hung werden die m_t durch die aktuellen Zeitreihenwerte ersetzt und die Parameter θ_3 und θ_3/η mittels einer linearen Regression bestimmt. Mit den Regressionskoeffizienten der Beziehung $r_t = \beta_0 + \beta_1 y_t + \varepsilon_t$ hat man Ausgangsschätzungen $\hat{\theta}_3^{(0)} = \hat{\beta}_0$ und $\hat{\eta}^{(0)} = -\hat{\beta}_0/\hat{\beta}_1$. Zur Gewinnung einer Startschätzung für θ_2 wird die Ausgangsbeziehung verwendet. Diese lässt sich schreiben als

$$m_t = \frac{\eta \theta_2}{\theta_2 + e^{-\theta_3 t}}.$$

Einsetzen von y_t für m_t und Umformen der Beziehung führt auf

$$\theta_2 \approx \frac{y_t e^{-\theta_3 t}}{\eta - y_t}.$$

Einen Startwert $\hat{\theta}_2^{(0)}$ bekommt man daher durch Bilden des arithmetischen Mittels der rechten Ausdrücke unter Verwendung der Startschätzungen für die beiden Parameter. Schließlich ergibt sich $\hat{\theta}_1^{(0)}$ aus der Beziehung $\theta_1 = \eta \theta_2$. Numerisch erhält man:

$$\hat{\theta}_1^{(0)} = 0.01926, \quad \hat{\theta}_2^{(0)} = 0.000063, \quad \hat{\theta}_3^{(0)} = 0.28589.$$

Die Optimierung mit dem ableitungsfreien Nelder-Mead-Optimierungsverfahren ergibt dann

$$\hat{\theta}_1 = 0.01924, \quad \hat{\theta}_2 = 0.0000567, \quad \hat{\theta}_3 = 0.2950.$$

Die Abbildung lässt erkennen, dass sich das tatsächliche Wachstum nicht der Struktur der vorgegebenen Trendfunktion fügt, die auf eine Begrenzung zustrebt. Dieses Bild widerspricht früheren Anpassungen, bei denen kürzere Zeitreihen verwendet wurden.

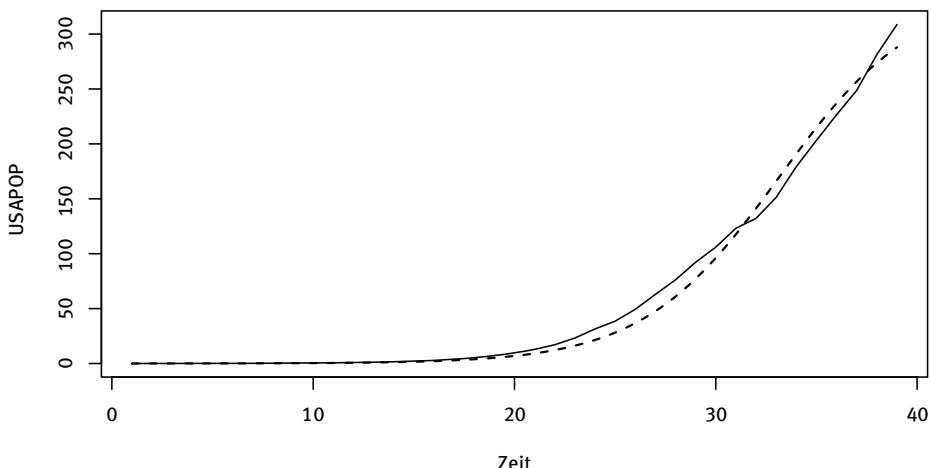


Abb. 2.13: USAPOP mit angepasster logistischer Trendfunktion

R-Code:

```

1 data(USAPOP)
2 y <- as.vector(USAPOP/1000)
3 n <- length(y)
4 t <- c(1:n)
5 r <- (y[-1]-y[-n])/y[-1]
6 out1 <- lm(r~y[-n])
7 theta3 <- out1$coefficients[1]
8 eta <- (-1)*out1$coefficients[1]/out1$coefficients[2]
9 theta2 <- mean(y*exp(-theta3*t)/(eta-y))
10 theta1 <- eta*theta2
11 start <- c(theta1,theta2,theta3)
12 f <- function(par) sum((y-par[1]/(par[2]+exp(-par[3]*c(1:n))))^2)
13 out3 <- optim(start,f)
14 theta <- out3$par
15 fu <- function(x) theta[1]/(theta[2] + exp(-theta[3]*x))
16 ts.plot(y,xlab="Zeit",ylab="USAPOP")
17 curve(fu,add=T,lty=2)

```

Zunächst wird das Zeitreihenattribut mit den zugehörigen 10 Jahresschritten (`frequency=0.1`) gelöscht, um es mit der nichtlinearen Funktion einfacher zu haben. Der Code setzt dann die im Text beschriebene Vorgehensweise zur Bestimmung von Startschätzern um. Die Optimierungsroutine `optim` mit der Nelder-Mead-Methode wird hier gewählt, da die auf Ableitungen basierenden Alternativen nicht zum Ziel führen.

Bestimmung der glatten Komponente

In verschiedenen Situationen ist die Trendbereinigung über die Anpassung eines Polynoms von vornherein suspekt. Das kann etwa der Fall sein, wenn das Verhalten der Reihe durch Zyklen verschiedener Dauer bestimmt wird, oder aus dem Plot deutlich wird, dass der Verlauf nicht über den gesamten Zeitraum durch ein Polynom zu erfassen ist. Die Trendfunktion ist dann eher als glatte Komponente anzusehen. Dann sind *andere Ansätze für Trendfunktionen* mit geeigneten Bestimmungsmethoden heranzuziehen.

Als eine Möglichkeit kann man die als glatt unterstellte deterministische Trendkomponente mittels lokaler Mittelwerte schätzen. Das führt zu *gleitenden Durchschnitten*. Im Wesentlichen wird dabei für alle gleich langen Reihensegmente das zugehörige einfache arithmetische Mittel berechnet und den jeweiligen zeitlichen Mittelpunkten der Segmente zugeordnet. Das Vorgehen ist in der Abbildung 2.14 illustriert.

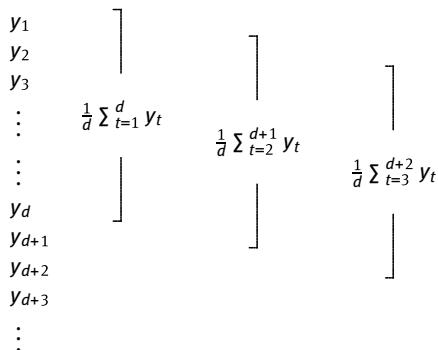


Abb. 2.14: Berechnungsschema für einen einfachen gleitenden Durchschnitt

In der Regel wird d ungerade gewählt; das Ergebnis $\sum_{t=s+1}^{s+d} y_t/d$ wird dann dem mittleren Zeitpunkt $s + (d + 1)/2$ zugeordnet. Ist d gerade, $d = 2q$, so ist diese Zuordnung zum mittleren Zeitpunkt nicht sinnvoll. Man behilft sich so, dass die Länge um eins erhöht wird, dafür aber die Randpunkte nur mit dem halben Gewicht eingehen. Man bildet also $(y_{t-q}/2 + y_{t-q+1} + \dots + y_{t+q-1} + y_{t+q}/2)/d$. Zusammen bezeichnet man diese als *einfache gleitende Durchschnitte*. Bei diesen sind die Gewichte, mit denen die y_{t-u} versehen werden, bis eventuell auf die der beiden Randpunkte, gleich groß.

Beispiel 2.22 (Verkehrsunfälle)

In der Abbildung 2.15 ist die Reihe UNFALL, die monatlichen Anzahlen der Verkehrsunfälle mit Personenschäden in der BRD von Januar 1974 bis Dezember 2017, zusammen mit den gleitenden 24er Durchschnitten dargestellt.

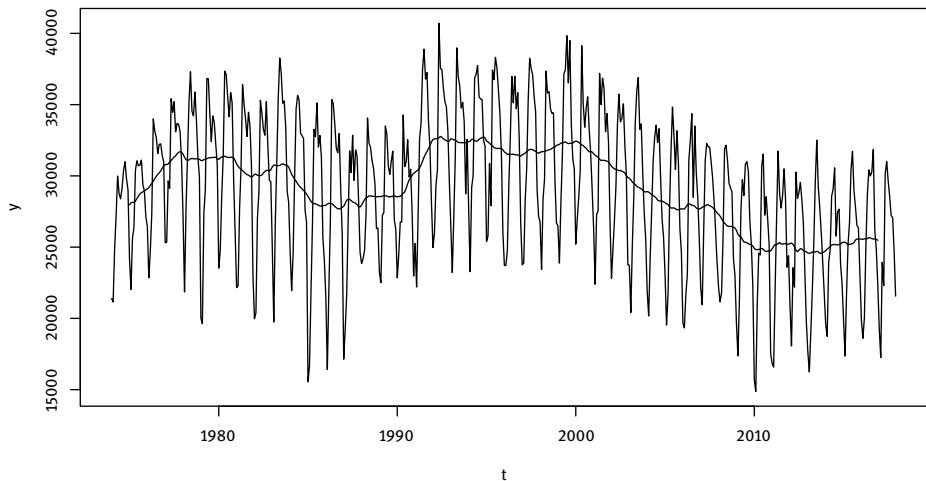


Abb. 2.15: UNFALL mit gleitendem 24er Durchschnitt

R-Code:

```

1 data(ACCIDENT)
2 Unfall <- ACCIDENT
3 g <- moveav(Unfall, 24)
4 plot(Unfall)
5 lines(g)

```

Gleitende Durchschnitte werden mit der Funktion `moveav` aus dem Paket `tsapp` bestimmt. Die Länge darf dabei auch gerade sein.

Anstelle der einfachen können über die jeweiligen Stützbereiche auch gewichtete Mittelwerte $\sum_{t=s+1}^{s+d} a_t y_t$, bei denen nicht alle Koeffizienten a_t gleich groß sind, bestimmt werden. Ein Beispiel dafür ist der *symmetrische gleitende 15er Durchschnitt von Spencer*, der speziell bei Sterbetafeln zur Glättung eingesetzt wird.

$$\frac{1}{320}[-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3].$$

Es lässt sich zeigen, dass die lokale Approximation mittels eines Polynoms einem gleitenden Durchschnitt entspricht. Lokale Approximation meint dabei, dass an einem Abschnitt der Reihe, etwa an die ersten d Werte, ein Trendpolynom angepasst wird. Jedoch wird nur der Wert des Polynoms an dem mittleren Abschnittszeitpunkt als geglätteter Wert genommen. Der Abschnitt wird sodann um eine Zeiteinheit verschoben. Es wird erneut ein Polynom angepasst und wieder der Wert am Abschnittsmittelpunkt genommen. Dieses Verfahren wird dann fortgesetzt, bis man am rechten Rand der Reihe angelangt ist. Siehe dazu Schlittgen & Streitberg (1999). Gleitende Durchschnitte sind in einiger Ausführlichkeit in Kendall (1973) dargestellt.

Ein Beispiel dafür sind die *gleitenden Durchschnitte nach Henderson*. Hier werden kubische Trendpolynome verwendet. Der Stützbereich richtet sich nach der Stärke der Irregularität der Reihe. Für den symmetrischen 9er Durchschnitt gilt etwa:

$$[-0.041, -0.010, 0.119, 0.267, 0.330, 0.267, 0.119, -0.010, -0.041].$$

An den Rändern geht bei gleitenden Durchschnitten eine Anzahl von Punkten verloren. Der so genannte Output ist kürzer als der Input, d. h. die ursprüngliche Reihe. Die bereinigte Reihe als Differenz von Input und Output ist dann also mittelwertstationär aber kürzer.

Einen anderen Ansatz, eine glatte Komponente zu schätzen, bieten *Splines*. Sie wurden als mathematische Gebilde unter dem Vorzeichen entwickelt, dass über kurze Zeitabschnitte Polynome durch die Datenpunkte gelegt werden und dabei die aneinanderstoßenden Polynome einen glatten Übergang aufweisen. *Glättende Splines* sind nunmehr losgelöst von diesem Hintergrund dadurch definiert, dass man eine Funktion g sucht, die einerseits glatt ist und andererseits im Wesentlichen durch die Punkte des Plots verläuft. Das Ziel der Glattheit wird in der Weise operationalisiert, dass die

zweite Ableitung möglichst null ist. Da in dem hier vorliegenden Kontext die Zeitabstände jeweils eine Einheit betragen, werden die zweiten Differenzen betrachtet:

$$(g_t - g_{t-1}) - (g_{t-1} - g_{t-2}) = g_t - 2g_{t-1} + g_{t-2} .$$

Dass diese möglichst verschwinden, führt zu der Forderung

$$\sum (g_t - 2g_{t-1} + g_{t-2})^2 \stackrel{!}{=} \min . \quad (2.27)$$

Zusammen lautet die Approximationsaufgabe dann:

$$\lambda \sum_{t=3}^N (g_t - 2g_{t-1} + g_{t-2})^2 + \sum_{t=1}^N (y_t - g_t)^2 \stackrel{!}{=} \min . \quad (2.28)$$

Der Parameter λ steuert dabei die Glattheit. Ist λ klein, so hat die erste Summe ein kleines Gewicht und die Minimierung führt dazu, dass vor allem die zweite Summe klein wird. Dies ergibt eine gute Übereinstimmung mit den Daten. Ist λ groß, so überwiegt die Forderung nach der Glattheit und man erhält einen eher linearen Verlauf. Die Bestimmung der Lösung von (2.28) ist kein Problem, da sich diese Minimierungsaufgabe auf ein geeignetes Regressionsmodell zurückführen lässt, siehe Schlittgen (1991a). Dort wird auch eine Variante vorgestellt, die gegen Ausreißer unempfindlich ist. Darauf wird im Abschnitt 12.3 eingegangen.

Für eine adäquate datenbasierte Bestimmung von λ liegt es nahe, die Zeitreihen-Kreuzvalidierung einzusetzen. Jedoch ist dies nicht trivial, siehe Opsomer, Wang & Yang (2001). Einen grandiosen Überblick über Splines erhält man aus verschiedenen Artikeln des von Schimek (2000) herausgegebenen Buches. Hier sind auch weiterführende Hinweise zur Zeitreihen-Kreuzvalidierung zu finden.

Die Transformation von y_t in g_t wird auch als *Hodrick-Prescott-Filter* bezeichnet. Jedoch hat diese Glättung bereits Whittacker (1923) vorgestellt. Sie wird auch in Whittacker & Robinson (1924, S. 304 ff.) ausführlich beschrieben. Der einzige Unterschied besteht darin, dass in der Originalquelle dritte Differenzen betrachtet werden. In weiten Bereichen wird sie unter dem Namen *Whittaker-Graduation* geführt. Reinsch (1967) zeigte, dass die Lösung von (2.28) ein kubischer Spline mit Knoten an den Zeitpunkten t ist.

Beispiel 2.23 (Auftragseingänge bei einem Unternehmen)

Für die gemäß $\sqrt[4]{y_t}$ transformierte Reihe AUFTRAGSEINGANG, die monatlichen Auftragseingänge ausgewählter Fertigungsstätten eines Unternehmens (Januar 1968 bis September 1987) ergibt der Splines-Ansatz die in der Abbildung 2.16 dargestellte glatte Komponente. Zu der Transformation siehe den Abschnitt 2.5. Als Glättungsparameter wurde $\lambda = 1000$ gewählt. Das Ansteigen ist offensichtlich nicht mit einer polynomialen Trendfunktion geringerer Ordnung zu erfassen.

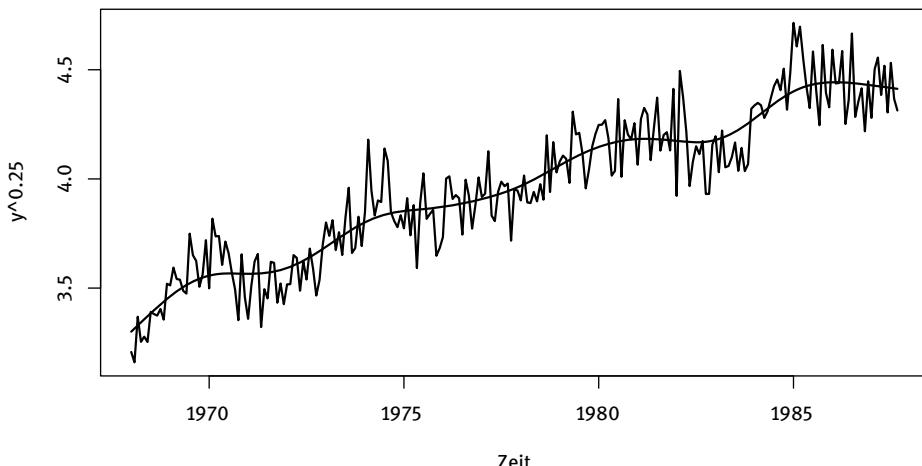


Abb. 2.16: AUFTRAGSEINGANG (transformiert) mit glatter Komponente

R-Code:

```

1 data(INORDER)
2 y <- INORDER
3 g <- smoothls(y^0.25, 100)
4 plot(y^0.25, )
5 lines(seq(1968, 1968+(length(y)-1)/12, 1/12), g)

```

Die Splines-Glättung geschieht mit der Funktion `smoothls` aus dem Paket `tsapp`. Der Glättungsparameter muss bei dieser Funktion per Versuch und Irrtum bestimmt werden, da eine automatische Wahl nicht implementiert ist.

2.4 Saisonbereinigung

Die Saison ist nach herkömmlicher Auffassung ein durch jahreszeitliche Einflüsse bewirktes, sich relativ regelmäßig ausprägendes Phänomen. Die retrospektiv orientierte *Komponentenzerlegung* ökonomischer Zeitreihen hat i. d. R. das Ziel, den Verlauf einer Reihe zu ermitteln, so wie er sich ohne saisonale Einflüsse darstellt. Hauptaugenmerk gilt häufig dem Konjunkturverlauf, der sich nicht in Form eines Trends, sondern vielmehr als glatte Komponente bemerkbar macht. Kurz gefasst reflektiert der verbreitete Einsatz von Verfahren zur *Saisonbereinigung*, d. h. zum Herausrechnen der Saisonkomponente, das Bedürfnis nach einer besseren Einschätzung der momentanen Situation sowie die Notwendigkeit der Adjustierung von Prognosen.

Einfache Ansätze

Ein Weg zur Saisonbereinigung besteht darin, zunächst die Saison zu bestimmen und sie dann von den Zeitreihenwerten abzuziehen. Den Ausgangspunkt für die Diskussion der *einfachen Bestimmungsmethoden* bildet das additive Komponentenmodell (2.13):

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

Das multiplikative Modell kann durch Logarithmieren in das additive überführt werden. Somit ist eine getrennte Betrachtung weitgehend unnötig.

Bei zentrierten Zeitreihen kann die Saison mit Hilfe von so genannten *Dummyvariablen* modelliert werden. Dazu bekommt jeder Monat oder jedes Quartal eine Variable S_i zugeordnet, die genau dann eins ist, wenn der Zeitpunkt zu dem Monat (Quartal) gehört. Sonst ist sie null.

Sei die Saisonperiode s , also $s = 4$ für Quartalsdaten und $s = 12$ für monatliche Beobachtungen. Dann setzt man für Zeitreihen mit Saison, jedoch ohne Trend die lineare Regressionsbeziehung an:

$$Y_t = \beta_0 + \beta_1 S_{1t} + \cdots + \beta_{s-1} S_{s-1,t} + \varepsilon_t .$$

Ohne Konstante muss die Reihe für die Schätzung vorab zentriert werden, damit die Summe der Saisonausschläge über jede Jahresperiode null ist.

Wird für den Trend ein Polynom unterstellt, so können beide Komponenten gemeinsam aus dem Regressionsansatz

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_p t^p + \beta_1 S_{1t} + \cdots + \beta_{s-1} S_{s-1,t} + \varepsilon_t , \quad (2.29)$$

geschätzt werden. Dazu ist eine Effekt-Kodierung der Saison vorzunehmen:

$$S_{i,t} = \begin{cases} 1 & \text{für } t \equiv i \pmod{s} \\ 0 & \text{für } t \not\equiv i, s \pmod{s} \\ -1 & \text{für } t \equiv s \pmod{s} \end{cases} .$$

Die hier nicht angegebene Saisonperiode $S_{s,t}$ hat den Koeffizienten $\beta_s = -(\beta_1 + \cdots + \beta_{s-1})$.

Bei einem linearen Trend und Quartalsdaten lautet das lineare Modell konkret:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 1 \\ 1 & 4 & -1 & -1 & -1 \\ 1 & 5 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

¹ $t \equiv i \pmod{s}$ bedeutet, dass t und i dividiert durch s den gleichen Rest haben.

Beispiel 2.24 (Steuern in Deutschland)

Die Zeitreihe STEUERN beinhaltet die in Deutschland insgesamt monatlich gezahlten Gemeinschaftsteuern in Milliarden Euro von Januar 1999 bis Juni 2019 (Quelle: Statistisches Bundesamt). Unter Gemeinschaftsteuern werden Lohnsteuer, veranlagte Einkommensteuer, nicht veranlagte Steuern vom Ertrag, Zinsabschlag, Körperschaftssteuer, Umsatzsteuer und Einfuhrumsatzsteuer zusammengefasst.

Die Zeitreihe zeigt für den betrachteten Zeitraum einen konstanten positiven Trend und eine recht ausgeprägte und sehr stabile Saisonfigur. Daher eignet sie sich für eine Zerlegung gemäß dem eben skizzierten Regressionsansatz.

Beachtet man die Skalierungen der Ordinaten, so ist aus der Abbildung 2.17 zu erkennen, dass der Ansatz ein zufriedenstellendes Ergebnis liefert.

R-Code:

```
1 data(TAXES)
2 y <-TAXES
3 out <- simpledecomp(y,trend=2,season=12,Plot=TRUE)
```

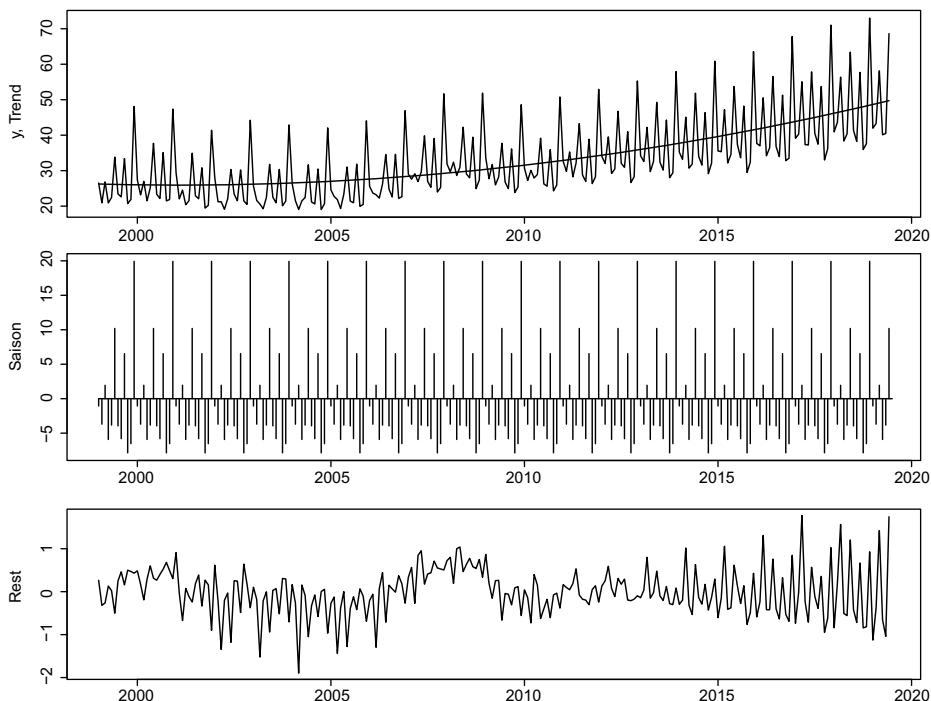


Abb. 2.17: Zerlegung der Zeitreihe STEUERN mit quadratischem Trend und konstanter Saison

Die Funktion `splinedecom` leistet die gewünschte Zerlegung. `out` ist eine dreispaltige Matrix mit Trend, Saison und Rest. Sie wird auch als Grafik ausgegeben, wenn das optionale Argument `Plot` auf wahr gesetzt wird.

In Erweiterung der Trendbereinigung mit Splines gibt es auch die Möglichkeit, Zerlehnungsverfahren für Zeitreihen unter Verwendung der glättenden Splines zu konstruieren. Man erhält dann eine glatte Komponente, eine sich langsam ändernde Saison und einen Rest. Der im Folgenden dargestellte Ansatz ist eine Modifikation eines auf Schlicht (1981) zurückgehenden Vorschlags.

Für die glatte Komponente wird wie bei der Splinesglättung die Forderung $\sum(g_t - 2g_{t-1} + g_{t-2})^2 \stackrel{!}{=} \min$ erhoben. Für die Saison mit der Saisonperiode s wird wie üblich angenommen, dass die Summe von s aufeinander folgenden Saisonkomponenten (in etwa) null ist: $\sum_{\tau=0}^{s-1} S_{t-\tau} \approx 0$ ($t = s, \dots, N$). Damit wird zugleich erreicht, dass auch $S_{t-\tau} \approx S_{t-\tau-s}$ gilt, vgl. Schlittgen (1991b). Diese beiden Forderungen lassen sich nun zu folgendem Minimierungsproblem zusammenfassen:

$$\lambda \sum_{t=3}^N (g_t - 2g_{t-1} + g_{t-2})^2 + \kappa \sum_{t=s}^N \left(\sum_{\tau=0}^{s-1} S_{t-\tau} \right)^2 + \sum_{t=1}^N (y_t - g_t - S_t)^2 \stackrel{!}{=} \min. \quad (2.30)$$

Die Parameter λ und κ steuern dabei die Glattheit der jeweiligen Komponente. Die Berechnung ist kein wesentliches Problem, da sich diese Minimierungsaufgabe auf ein geeignetes Regressionsmodell zurückführen lässt, siehe Schlicht (1981) und Schlittgen (1991b). In der letztgenannten Arbeit wird auch eine Variante vorgestellt, die gegen Ausreißer unempfindlich ist. Darauf wird im Abschnitt 12.3 eingegangen.

Beispiel 2.25 (Produktionsindex des produzierenden Gewerbes)

Einer der wichtigen volkswirtschaftlichen Indikatoren ist der Produktionsindex des produzierenden Gewerbes. Die Beobachtung seiner Entwicklung gehört zum Standard der Konjunkturbeobachtung. Dabei sieht man vielfach die saisonbereinigte Reihe für die Einschätzung der konjunkturellen Entwicklung besser geeignet an als die glatte Komponente. Hier soll zur Illustration die Splineszerlegung der von Januar 2000 bis November 2009 reichenden Reihe des Index auf der Basis 2005 betrachtet werden. Die Reihe wird mit dem Kürzel `PRODINDEX` bezeichnet. Die in der Abbildung 2.18 präsentierte Zerlegung dieser Reihe mittels des Splinesansatzes hat den Vorteil, dass die sich zwar langsam aber doch deutlich ändernde Form der Saison gut ermitteln lässt.

R-Code:

```
1 data(PRODINDEX)
2 y <- PRODINDEX
3 gs <- splinedecom(y, 12, 20, 5, Plot=TRUE)
```

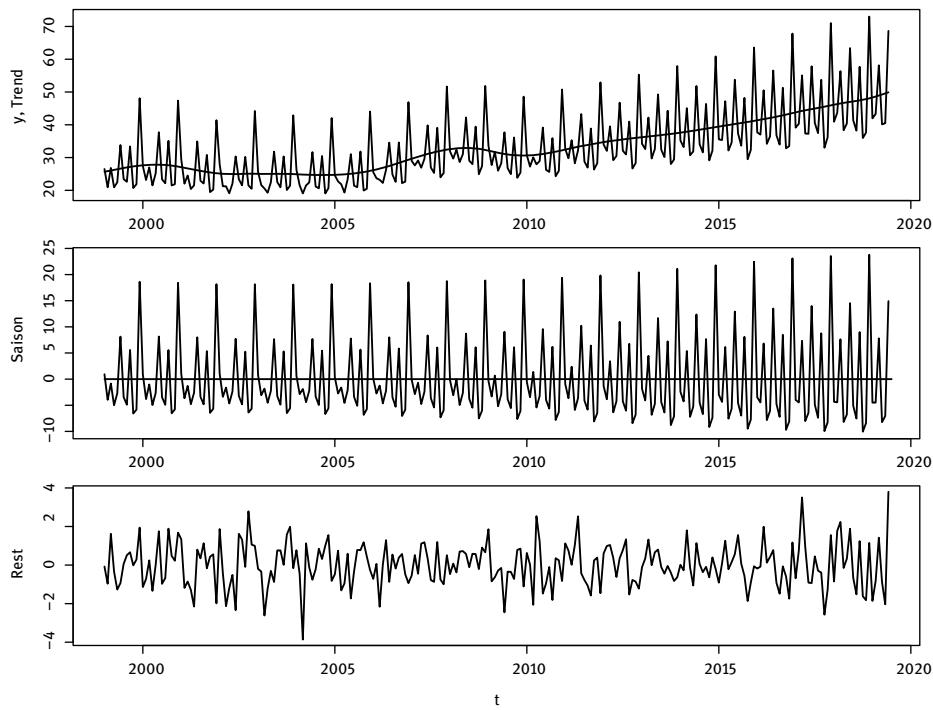


Abb. 2.18: Zerlegung der Zeitreihe PRODINDEX

Bei der Funktion `splinedecomp` ist die Angabe der Saisonperiode notwendig. Die Glättungsparameter sind per Versuch und Irrtum zu bestimmen. Um die Grafik direkt zu bekommen, ist das optionale Argument `Plot` auf wahr zu setzen.

Neben der Darstellung von Trend-, Saison- und ggfs. Restkomponente ist auch die Darstellung der saisonalen Teilreihen aufschlussreich. In einem *saisonalen Teilreihen-Plot* werden für die Saisonkomponente die Mittelwerte der jeweils gleichnamigen Saisonperiode bestimmt und die Abweichungen von diesem Mittelwert eingezzeichnet. Dies erlaubt zu erkennen, wie sich die einzelnen Saisonkomponenten entwickeln.

Beispiel 2.26 (Produktionsindex des produzierenden Gewerbes – Fortsetzung von Seite 46)

An der Darstellung der Saison dieser Zeitreihe sieht man, dass sich die Saisonkomponente im Januar zurückentwickelt, während die Saisonausschläge im Juni immer größer werden. Im Februar, April, Mai, Juli, August, Oktober und November sind die mittleren Werte der Saison am niedrigsten. In diesen Monaten nehmen sie zudem über die gesamte Zeit ab. Ganz anders im Juli und Dezember, wo das

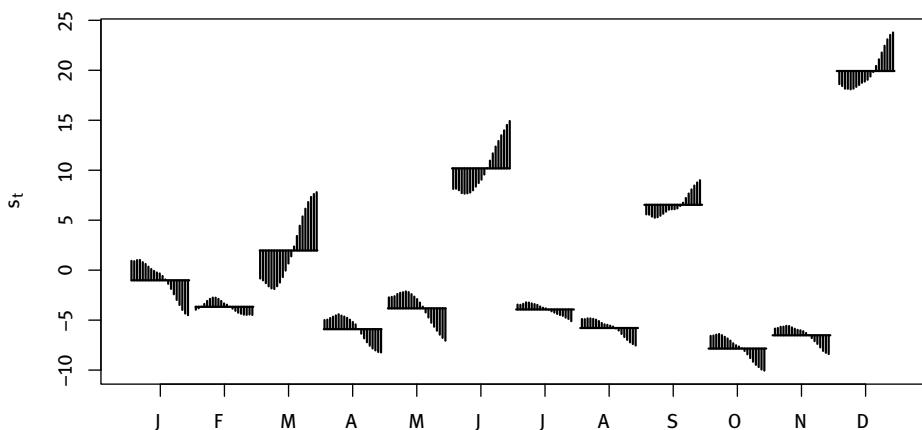


Abb. 2.19: Entwicklung der saisonalen Teilreihen von PRODINDEX

Niveau hoch ist – die Saisonaußschläge liegen über der glatten Komponente – und steigen über die betrachtete Zeit. Über das Jahr gesehen sollte die Summe der Saisonaußschläge ja null sein.

R-Code:

```
1 monthplot(gs[,2], type="h")
```

Der R-Code des Beispiels 2.25 wird fortgesetzt. Da die Saisonkomponente in der zweiten Spalte von gs gespeichert ist und die ursprünglichen Zeitreihenattribute geerbt hat, reicht der hier angegebene Aufruf, um die Grafik zu erstellen.

Etablierte Saisonbereinigungsverfahren

Es gibt etliche *etablierte Saisonbereinigungsverfahren*, die von den statistischen Ämtern und Institutionen regelmäßig eingesetzt werden. Das bekannteste ist das *CENSUS X-11 Verfahren*; es wurde von US Bureau of the Census entwickelt und war das wohl verbreitetste Saisonbereinigungsverfahren. Inzwischen ist es von der Weiterentwicklung *X-12 ARIMA* abgelöst worden. X-12-ARIMA¹ besteht inhaltlich aus drei Verfahrensteilen. Der erste Teil bietet dem Anwender gegenüber X-11 völlig neue Möglichkeiten zur Extremwert- und zur Kalenderbereinigung auf der Basis sogenannter Regressionsmodelle mit ARIMA-Resten, vgl. die Kapitel 3 und 9. Darüber hinaus können Prognosewerte erstellt werden, die dazu dienen sollen, die saisonbereinigten Werte an den

¹ Dieser Skizze des X-12-ARIMA-Verfahrens wurden die „Hinweise(n) zur Zeitreihenanalyse mit X-12-ARIMA“ des Statistischen Bundesamtes zugrunde gelegt.

Rändern des Analysezeitraums zu stabilisieren. Der zentrale Teil des Verfahrens, mit dem die Saisonbereinigung durchgeführt wird, ist dagegen bis auf einige Ergänzungen mit dem alten X-11 von 1965 identisch. Das heißt insbesondere, dass die Saisonbereinigung auf der Basis einer iterativen Anwendung verschiedener gleitender Durchschnitte beruht. Dem Anwender stehen dabei hinsichtlich der Länge der verwendeten Durchschnitte zahlreiche Wahlmöglichkeiten zur Verfügung. Außerdem muss sich der Anwender zwischen einer additiven und einer multiplikativen Verknüpfung der Zeitreihenkomponenten entscheiden. Zur Erstellung qualitativ hochwertiger Ergebnisse ist bei X-12-ARIMA die gut durchdachte Setzung einer Vielzahl von Verfahrensparametern durch den Anwender erforderlich. Hierzu stellt ein dritter Teil einige diagnostische Instrumente bereit, die dem Anwender bei der Überprüfung der Zweckmäßigkeit der gewählten Optionen helfen sollen.

Im Statistischen Amt der Europäischen Gemeinschaften (Eurostat) wird das Verfahren *TRAMO/SEATS* eingesetzt. Eine weitere Neuentwicklung im Bereich der Saisonbereinigungsverfahren ist *STAMP*. Auch diese Verfahren basieren wesentlich auf Ideen aus dem Bereich der ARIMA-Modellierung.

Das *Berliner Verfahren* hat seinen Ursprung in den Bemühungen einer Berliner Arbeitsgruppe und wurde dann im Statistischen Bundesamt weiterentwickelt. Derzeit liegt es in der Version BV4.1 vor. Das Verfahren ist recht konsequent auf der Anwendung von linearen Filtern und auf Algorithmen aufgebaut, die von der Wirkung her durchschaubar sind. Daher soll es etwas ausführlicher dargestellt werden. Dies geschieht entlang der Beschreibung in Speth (2004).

Ausgangspunkt bildet das *Zeitreihenmodell*, das die verschiedenen Einflüsse, die makroökonomische Zeitreihen vielfach gleichartig beeinflussen, durch eine additive Überlagerung entsprechender Komponenten erfasst:

$$y_t = m_t + s_t + c_t + a_t + e_t + u_t \quad (t = 1, \dots, N). \quad (2.31)$$

Dabei sind:

m_t = Trend-Konjunkturkomponente,

s_t = Saisonkomponente,

c_t = Kalenderkomponente,

a_t = Komponente anderer systematischer Einflussgrößen,

e_t = Ausreißerkomponente,

u_t = Komponente von Zufallsschwankungen

Dieses Modell wird konkretisiert im *Basismodell* des BV4.1, indem für die Komponenten Annahmen getroffen werden. Es wird unterstellt, dass sich die Trend-Konjunktur-Komponente lokal durch Polynome niedriger Ordnung p approximieren lässt. Die Saisonkomponente wird als sich nur langsam verändernde periodische Struktur angenommen und entsprechend lokal durch ein Fourierpolynom approximiert, siehe das Kapitel 6. Für den Zufallsprozess (u_t) wird weißes Rauschen unterstellt. Zusammen-

genommen lautet das Basismodell also:

$$y_t = \sum_{j=0}^p \alpha_j t^j + \sum_{j=1}^l \left(\beta_j \cos\left(2\pi \frac{j}{P} t\right) + \gamma_j \sin\left(2\pi \frac{j}{P} t\right) \right) + u_t . \quad (2.32)$$

Dass die gewählten Funktionen jeweils nur lokal die Komponenten erfassen, führt zu einer gleitenden Schätzung der Komponenten anhand des Basismodells. Die Schätzung der Koeffizienten α_j , β_j und γ_j des Trend-Konjunktur-Polynoms und der Fourierreihe des Basismodells geschieht dementsprechend zunächst auf der Grundlage der Beobachtungswerte des Zeitintervalls $U_1 = [t_1, \dots, t_k]$, dann auf $U_2 = [t_2, \dots, t_{k+1}]$ usw. Die Länge k der Schätzbereiche wird dabei ungerade gewählt. Die Schätzung wird jeweils anhand eines Kleinst-Quadrat-Kriteriums durchgeführt. Genauer wird ein gewichtetes Kleinst-Quadrat-Kriterium herangezogen. Die Gewichte w_t für die Zeitreihenwerte sind umso höher, je näher sie in der Mitte liegen:

$$\sum_{t_i \in U_i} w_{t_i} [y_{t_i} - \hat{m}_{t_i} - \hat{s}_{t_i}]^2 \stackrel{!}{=} \min .$$

Trend-Konjunktur- und die Saisonkomponentenschätzung werden nur bezüglich des zentralen Zeitpunktes verwendet. Denn dann weisen die Schätzwerte \hat{m}_{t_i} und \hat{s}_{t_i} keine zeitlichen Vor- oder Nachläufe gegenüber den entsprechenden Komponenten in den Zeitreihen auf.

Die Schätzung der Trend-Konjunktur-Komponente und der Saisonkomponente benötigt unterschiedliche Längen der Schätzbereiche. Für die Saisonkomponente ist es erforderlich, die Zeitreihenwerte mehrerer Jahre heranzuziehen. Bei der Trend-Konjunktur-Entwicklung sind Polynome niedriger Ordnung nur für kürzere Zeitbereiche adäquat. Deshalb wird bei BV4.1 so vorgegangen, dass gemäß dem obigen Schätzansatz zunächst die Trend-Konjunktur-Komponente der zu analysierenden Zeitreihe ermittelt wird, und zwar unter Rahmenbedingungen, die speziell auf das Ziel der Schätzung der Trend-Konjunktur-Komponente abgestellt sind. Die hierbei ebenfalls anfallende Saisonsschätzung wird nicht weiter verwendet, weshalb man das Verfahren als Trend-Konjunktur-Teilschätzung bezeichnet. Es folgt die Schätzung der Saisonkomponente. Dazu wird die Zeitreihe um die geschätzte Trend-Konjunktur-Komponente bereinigt (d. h. Bildung von $y_t - \hat{m}_t$ für $t = 1, \dots, N$) und die bereinigte Reihe wieder gemäß dem Grundprinzip analysiert, jetzt aber unter bezüglich der Schätzung der Saisonkomponente optimierten Bedingungen. Verwendet wird dann nur die Schätzung der Saisonkomponente (Saison-Teilschätzung).

Für die Schätzung der Komponenten für Schätzzeitpunkte an den beiden Enden der Zeitreihe wird das Grundprinzip abgewandelt. Es werden asymmetrische Schätzbereiche verwendet und Modifikationen bezüglich der verwendeten Basisfunktionen und der relativen Lage bezüglich der Schätzstelle vorgenommen.

Das Ergebnis der Schätzung kann durch Ausreißer, einzelne extreme Werte, stark verfälscht werden. Daher wird vor der eigentlichen Zerlegung eine Ausreißererken-

nung durchgeführt. Diese werden dann mittels Dummy-Variablen als Ausreißerkomponente in das Basismodell integriert.

Die Grundlage für die Identifizierung bildet die Annahme, dass die Residualreihe u_t , $t = 1, \dots, n$, bis auf einige wenige Beobachtungswerte die Realisation eines stationären Normalprozesses (U_t) ist. Für gleitende Zeitintervalle mit einer festen Länge M wird dazu jeweils der bedingte Erwartungswert $E(U_t|u_{t-1}, \dots, u_{t-M})$ für den Reihenwert der nächsten Beobachtungsperiode ermittelt. Ein Zeitreihenwert wird als Ausreißer angesehen, wenn die Wahrscheinlichkeit dafür klein ist, dass der Wert von demselben Prozess erzeugt wurde wie die vorausgehenden M Werte. Das Maß hierfür ist die Abweichung des Beobachtungswertes von dem zugehörigen bedingten Erwartungswert der Prozessvariablen U_t um mehr als $\tau \cdot \hat{\sigma}_t$. In der Umsetzung wird dies auf vorläufig trend- und saisonbereinigte Reihe angewendet. Ausreißer werden vorläufig durch die obere oder die untere Grenze des jeweiligen Toleranzbereiches ersetzt.

Die *Kalenderkomponente* erfasst die Auswirkungen von Unterschieden der Beobachtungsperioden hinsichtlich der Anzahl der Tage, der Anzahl der Arbeitstage oder der Anzahlen der einzelnen Wochentage und Feiertage auf die beobachteten wirtschaftlichen Aktivitäten. Der Ansatz zur Modellierung der Kalendereinflüsse im BV4.1 lautet:

$$c_t^T = \sum_{i=1}^8 v_{T,i} d_{i,T} \quad \text{für } t = 1, \dots, n \quad (2.33)$$

Dabei sind $d_{i,T}$ für $i = 1, \dots, 6$ die Anzahlen der Wochentage Montag bis Samstag in der Beobachtungsperiode t , auf die keine Feiertage fallen; $d_{7,T}$ ist die Anzahl der Sonntage und $d_{8,T}$ die Anzahl der Feiertage, die nicht auf einen Sonntag fallen. $v_{T,i}$ sind zugehörige Koeffizienten.

Bezeichnet $\overline{d}_{i,t}$ für $i = 1, \dots, 8$ jeweils die durchschnittliche Anzahl des zu i gehörenden ‚Wochentages‘, so lässt sich die Kalenderkomponente schreiben gemäß:

$$c_t^T = \underbrace{\sum_{i=1}^8 v_{T,i} \overline{d}_i}_{C_1} + \underbrace{\sum_{i=1}^8 v_{T,i} (\overline{d}_{i,t} - \overline{d}_i)}_{C_2} + \underbrace{\sum_{i=1}^8 v_{T,i} (d_{i,t} - \overline{d}_{i,t})}_{C_3}. \quad (2.34)$$

C_1 ist eine Niveaukonstante. Sie wird als Bestandteil der Trend-Konjunktur-Komponente angesehen und für die Ermittlung der Kalenderkomponente nicht berücksichtigt. C_2 stellt einen Periodenlängeneffekt dar. Er hat saisonalen Charakter und wird der Saisonkomponente zugeordnet. C_3 ist schließlich die reine Kalenderkomponente. Erfasst Auswirkungen der Abweichung der tatsächlichen Wochentage- und Feiertagestruktur von der mittleren periodenspezifischen Struktur.

Der *Ablauf des Verfahrens* besteht zusammengefasst in folgenden Schritten:

- Identifikation von (potentiellen) Ausreißern;
- Integrierte Schätzung von Ausreißern, Kalendereinflüssen und Einflüssen von weiteren reihenspezifischen, vom Anwender vorzugebenden Variablen;
- Schätzung der Trend-Konjunktur- und der Saisonkomponente der Zeitreihe.

Schließlich sei darauf hingewiesen, dass das Problem der Saison in der von Helleberg (1992) herausgegebenen Aufsatzsammlung ausführlich diskutiert wird. Hier sind zahlreiche Literaturstellen zu finden. Edel et al. (1997) ist eine weitere Aufsatzsammlung, in der verschiedene Probleme und Ansätze zur Beurteilung der Saisonbereinigungsverfahren angeführt werden. Die Arbeit von Meyer behandelt einige Modifikationen des Census X-11 Verfahrens, die von der Deutschen Bundesbank vorgenommen wurden. Der Artikel von Hebbel gibt eine ausführliche Darstellung des Verallgemeinerten Berliner Verfahrens. Förster stellt die Geschichte der Saisonbereinigung dar. Neuere Entwicklungen werden in dem entsprechenden Kapitel in Peña, Tiao & Tsay (2001) referiert. Die Neuentwicklung TRAMO/SEATS ist in Artikeln in Helleberg (1992) sowie in Peña, Tiao and Tsai (2001) dargestellt. Bzgl. STAMP sei auf das Buch Koopman et al. (1996) hingewiesen. Die Vielzahl der Saisonbereinigungsverfahren wird begleitet von sporadischen Veröffentlichungen mit Vergleichen der verschiedenen Verfahren. Hier sei nur auf Höpfner (1998) sowie auf den Sammelband von Edel et al. (1997) hingewiesen. In dem Artikel von Höpfner und verschiedenen Beiträgen des Sammelbandes kann man sich auch über verschiedene Aspekte informieren, die für solche Verfahrensvergleiche bedeutsam sind.

2.5 Transformationen

In diesem Abschnitt werden zwei ganz unterschiedliche Arten von Transformationen behandelt. Bei der ersten Art von Transformationen werden zu jedem Zeitpunkt die Reihenwerte einzeln (aber gleichartig) modifiziert, um die Voraussetzungen für die Anwendung geeigneter Verfahren sicherzustellen. Zum zweiten geht es um die Kombination von Zeitreihenwerten zu unterschiedlichen Zeitpunkten, die dann zu einer neuen Zeitreihe führt.

Instantane Transformationen

Zu der ersten Art von Transformation, den *instantanen Transformationen*, wird man etwa veranlasst, wenn in einer Reihe ohne Saison eine Varianzinstationarität mit einer Mittelwertinstationarität in der Form einhergeht, dass die Streuung mit dem Niveau der Reihe steigt. Dann wird man vor allen anderen Schritten eine nichtlineare Transformation der Zeitreihenwerte gemäß $z_t = f(y_t)$ vornehmen. Kandidaten sind vor allem die *Box-Cox-Transformationen*

$$Z_t = \begin{cases} \frac{(Y_t + c)^\lambda - 1}{\lambda} & \text{für } \lambda \neq 0 \\ \ln(Y_t + c) & \text{für } \lambda = 0 . \end{cases} \quad (2.35)$$

In der Praxis werden meist einfach die Potenzen Y_t^λ bestimmt. Die Konstante c soll lediglich sicherstellen, dass alle Werte größer als null sind. Die angegebene Form der Transformationen hat den Vorzug zu zeigen, wie sich die logarithmische Transformation einfügt.

Eine Grundlage für die Ermittlung der richtigen Potenz der Transformation basiert darauf, dass die funktionale Abhängigkeit der Standardabweichung von dem Niveau unterstellt wird:

$$\sqrt{\text{Var}(Y_t)} = c \cdot E(Y_t)^b \quad (2.36)$$

Die Potenzierung mit dem Exponenten $1 - b$ führt dann zu einer konstanten Varianz. Linearisierung der Transformation $z = y^{1-b}$ führt nämlich auf:

$$\begin{aligned} \text{Var}(Z_t) &= \text{Var}(Y_t^{1-b}) \approx \text{Var}\left(\mu_t^{1-b} + \frac{\partial Y_t^{1-b}}{\partial Y_t}\Big|_{\mu_t} (Y_t - \mu_t)\right) \\ &= ((1 - b) \cdot (\mu_t)^{-b})^2 \text{Var}(Y_t - \mu_t) \\ &\approx c \cdot \mu_t^{-2b} E(Y_t)^{2b} = \text{konstant}. \end{aligned}$$

Um die Transformation zu bestimmen, wird (2.36) auf beiden Seiten logarithmiert:

$$\ln(\sqrt{\text{Var}(Y_t)}) = \ln(c) + b \cdot \ln(E(Y_t)). \quad (2.37)$$

Diese Gleichung legt folgendes Vorgehen nahe, um eine Schätzung der Potenz zu bestimmen. Es werden für gleichlange Reihenabschnitte $i = 1, \dots, I$ die Mittelwerte und Standardabweichungen, \bar{y}_i und s_i , berechnet und die logarithmierten Werte $(\ln(\bar{y}_i), \ln(s_i))$ in einem Diagramm eingetragen. Daraus, oder aber auch mittels einer einfachen linearen Regression, wird dann \hat{b} bestimmt. Dann ist der geeignete Exponent für die Transformation gegeben durch $\hat{\lambda} = 1 - \hat{b}$. Für $\hat{\lambda}$ reicht es i. d. R., glatte Werte, wie etwa $-2, -1.5, -1, -0.5, 0, 0.5, 1.5, 2$ zu wählen.

Beispiel 2.27 (Auftragseingänge bei einem Unternehmen – Fortsetzung von Seite 42)

Die Reihe AUFTRAGSEINGANG weist eine mit dem Niveau wachsende Streuung auf. Das in der Abbildung 2.21 wiedergegebene Box-Cox-Diagramm beruht auf zehn in etwa gleich langen Teilstücken der Reihe. Die Regressionsgerade ist $\ln(s_i) = -0.045 + 0.638 \ln(\bar{y}_i)$. Somit ergibt sich die geeignete Transformation mit $\hat{\lambda} = 1 - 0.638 = 0.362$. Anstelle dieser Transformation wird man wie oben erwähnt eher die einfachere Transformation $z_t = y_t^{0.3}$ oder gar $z_t = y_t^{0.5} = \sqrt{y_t}$ bevorzugen.

R-Code:

```
1 data(INORDER)
2 y <- INORDER
3 l <- BoxCox(y,10,Plot=T)
```

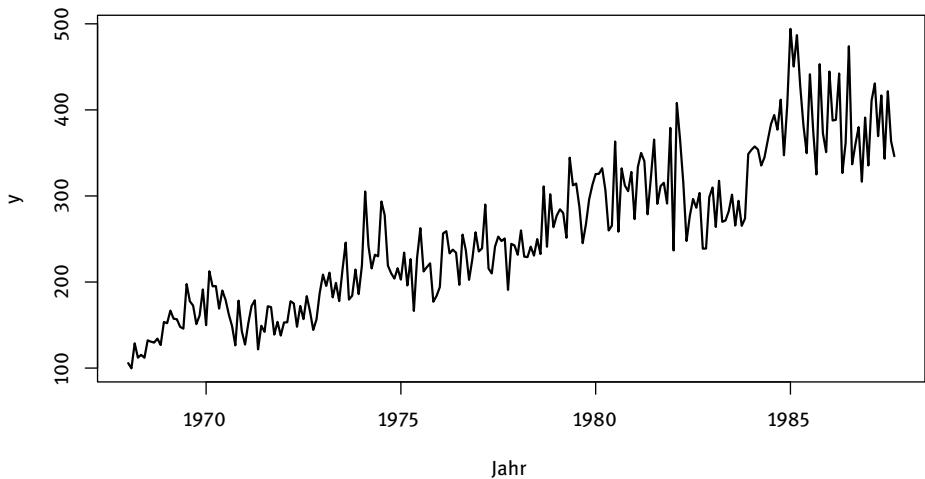


Abb. 2.20: AUFTRAGSEINGANG

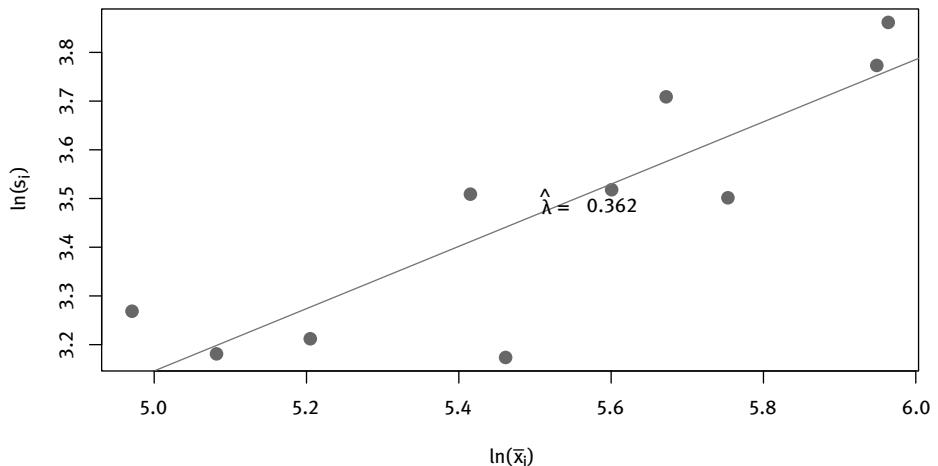


Abb. 2.21: Box-Cox-Diagramm für AUFTRAGSEINGANG

Die Funktion BoxCox braucht neben der Zeitreihe noch die Angabe der Anzahl der Segmente, die der Bestimmung der Potenz zugrunde gelegt werden sollen.

Ein weiterer Anwendungsbereich für die Box-Cox-Transformationen stellt die *Symmetrisierung* der Daten dar. Je symmetrischer die Verteilung der Daten ist, desto besser lassen sie sich insgesamt durch eine Normalverteilung beschreiben. Das ist von Vorteil; wie im ersten Kapitel erwähnt wurde, basieren viele Analysemethoden auf der

Normalverteilung. Weicht die Verteilung der Zeitreihenwerte sehr stark von der Normalverteilung ab, so kann dies Probleme bei der Modellierung verursachen.

Eine Möglichkeit, die Symmetrie zu überprüfen, besteht in der Darstellung der Daten in einem *Symmetriediagramm*. Hier werden die Werte der Größe nach geordnet. Die geordneten Werte seien $y_{(1)}, y_{(2)}, \dots, y_{(N)}$. Bei Symmetrie sollten die Werte $y_{(i)}$ und $y_{(N+1-i)}$ jeweils gleichweit vom Zentrum, dem Median \tilde{y} entfernt sein. Daher ergibt das Eintragen von $y_{(N+1-i)} - \tilde{y}$ in Abhängigkeit von $\tilde{y} - y_{(i)}$ in einem Diagramm Punkte, die bei Symmetrie um die 45° -Achse streuen.

Beispiel 2.28 (Anzahlen von Luchspelzen – Fortsetzung von Seite 18)

In der nicht-transformierten Reihe der Anzahlen von Luchspelzen sind die Peaks hoch und sehr spitz, während die Täler nicht so tief sind und eher einen flachen Eindruck machen. Dementsprechend sind die Daten auch sehr unsymmetrisch verteilt. Die Verteilung der transformierten Daten ist sehr viel symmetrischer.

Mit der Transformation $z_t = y_t^{0.1}$ würde man ein noch schöneres Symmetriediagramm erhalten. Wie aber oben erwähnt, begnügt man sich meist mit ‚glatten‘ Werten für die Transformation; bei biologischen Phänomenen hat sich zudem die logarithmische Transformation eingebürgert.

R-Code:

```
1 data(LYNX)
2 y <- LYNX
3 symplot(y)
4 symplot(log10(y))
```

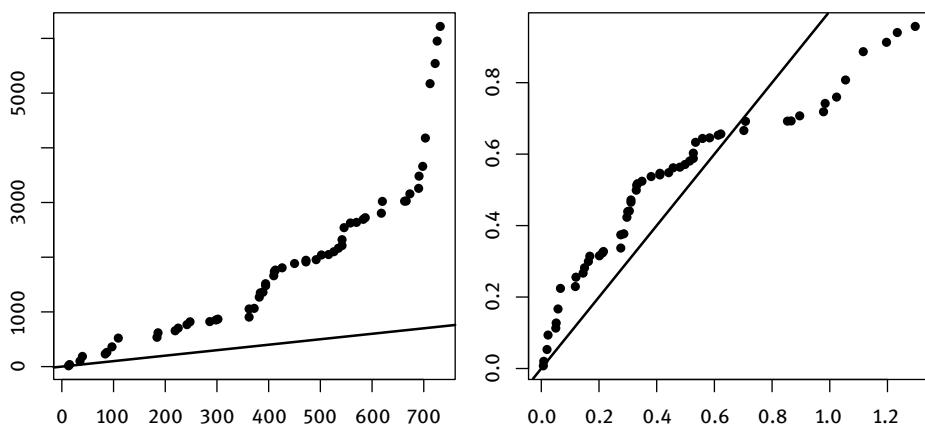


Abb. 2.22: Symmetrieplot der Zeitreihe der Anzahlen von Luchspelzen; links: original; rechts: logarithmisch transformiert

Lineare Filterung von Zeitreihen

Gleitende Durchschnitte gehören zu der zweiten Kategorie der Transformationen, bei denen Zeitreihenwerte unterschiedlicher Zeitpunkte kombiniert werden. Der gleitende 3er-Durchschnitt $z_t = \frac{1}{3}(y_{t+1} + y_t + y_{t-1})$ zeigt beispielsweise, wie die Werte der Reihe y_t verbunden werden, um zu der neuen Reihe z_t zu gelangen. Gleitende Durchschnitte sind Spezialfälle linearer Filter, bei denen dies eine Linearkombination ist.

Definition 2.29 (Linearer Filter)

Ein *linearer Filter* $(a_u)_{u=-r, \dots, s}$ ist eine Folge von Gewichten, so dass die Zeitreihe (y_t) in der folgenden Form transformiert wird:

$$z_t = a_{-r}y_{t+r} + \cdots + a_sy_{t-s} = \sum_{u=-r}^s a_u y_{t-u} \quad (2.38)$$

(z_t) heißt *Output* des Filters, (y_t) wird *Input* des Filters genannt. Die Transformation wird als *Filterung* bezeichnet.

Ein Hilfsmittel zur Beschreibung von linearen Filtern ist der *Backshift-Operator*. Er verschiebt die ganze Reihe um eine Zeiteinheit in die Vergangenheit: $BY_t = Y_{t-1}$

Auch wenn B ein Operator ist, der als Argument eine ganze Zeitreihe braucht, lässt sich damit rechnen, als sei B eine Zahl oder eine Variable. Potenzen des Backshift-Operators sind etwa durch mehrmaliges Anwenden definiert:

$$B^d Y_t = B^{d-1}(BY_t) = B^{d-1}Y_{t-1} = \cdots = Y_{t-d} .$$

Weiter ist $B^{-1}Y_t = Y_{t+1}$ und entsprechend gilt $B^{-d}Y_t = Y_{t+d}$.

Die linearen Filter lassen sich auch in der Form

$$\sum_{u=-r}^s a_u B^u \quad (2.39)$$

angeben. Angewendet auf eine Zeitreihe ergibt dies gerade (2.38):

$$z_t = \sum_{u=-r}^s a_u B^u y_t = \sum_{u=-r}^s a_u y_{t-u} .$$

Eine spezielle lineare Filterung ist die *Differenzenbildung*. Sie spielt eine große Rolle im Rahmen der Bereinigung stochastischer Komponenten. Eine nichtstationäre Zeitreihe kann nämlich durch einen nichtstationären stochastischen Mechanismus erzeugt sein. Ein Beispiel dafür ist etwa ein Random-Walk. Random-Walk-Prozesse bringen Realisationen hervor, die ein langfristiges Ansteigen bzw. Abfallen der Werte aufweisen. Solche Reihen erscheinen als offensichtlich mittelwertinstationär. Aufgrund der Definition des Random-Walk erhält man den stationären Ausgangsprozess durch Bildung sukzessiver Differenzen zurück.

Die einfachen oder *ersten Differenzen* $\Delta Y_t = Y_t - Y_{t-1}$ lassen sich dabei mit dem Backshift-Operator folgendermaßen schreiben:

$$\Delta Y_t = Y_t - Y_{t-1} = (1 - B)Y_t . \quad (2.40)$$

Höhere Differenzen ergeben sich durch mehrmaliges Anwenden der einfachen:

$$\Delta^d Y_t = (1 - B)^d Y_t = (1 - B)^{d-1}((1 - B)Y_t) = \Delta^{d-1}(\Delta Y_t) . \quad (2.41)$$

Es gibt nun keinen Grund, zu verlangen, dass die Differenzenbildung stets zu White-Noise-Prozessen führt. Man nennt einen Prozess $Y(t)$ vom Typ I(1), wenn $Z_t = \Delta Y_t$ stationär ist. Das „J“ steht dabei für „integriert“. Diese Bezeichnungsweise soll kenntlich machen, dass der Ausgangsprozess aus dem stationären durch Summation (Integration) zurückgewonnen werden kann: $Y_t = Z_0 + Z_1 + \dots + Z_t$.

Definition 2.30 (Integrationsgrad)

Ein Prozess (Y_t) heißt *integriert vom Grade k*, kurz vom Typ I(k), wenn $\Delta^k Y_t$ stationär ist. Ein stationärer Prozess heißt auch vom Typ I(0).

Liegt eine Zeitreihe vor, die als Realisation eines integrierten Prozesses angesehen werden kann, so wird man geeignet Differenzen bilden, um zu einer stationären Zeitreihe zu gelangen. Bei einem Prozess vom Typ I(k) sind k -mal Differenzen zu bilden. Auch bei deterministischen Trends hilft die Differenzenbildung, zu stationären Reihen zu gelangen. Dies ergibt sich aus der Aussage, dass ein durch ein Polynom beschreibbarer Anstieg eliminiert wird, wenn die Anzahl der Differenzenbildungen dem Grad des Polynoms entspricht.

Beispielsweise wird bei einem linearen Anstieg mit der Transformation $Z_t = \Delta Y_t$ Stationarität erreicht:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t \quad \Rightarrow \quad Z_t = \Delta Y_t = Y_t - Y_{t-1} = \beta_1 + \varepsilon_t - \varepsilon_{t-1}$$

(Z_t) ist offensichtlich stationär. Allerdings ist (Y_t) auch nach Differenzenbildung nicht bei null zentriert.

Bei quadratischem Anstieg ist die Bildung von Differenzen zweimal durchzuführen, um Stationarität zu erreichen; entsprechend kann man dies für Polynome höherer Ordnung fortsetzen. Bei Anwendungen ist es aber selten nötig, mehr als zweimal Differenzen zu bilden, um eine Zeitreihe stationär zu machen.

Analog zu den stochastischen Trends kann man sich auch die Saisonkomponente durch stochastische Ursachen bewirkt vorstellen. Die Eliminierung der Saison erfolgt dann mittels *saisonaler Differenzen* Δ_s , wenn s die Periodizität der vorliegenden Saisonkomponente ist. Diese erhält man, indem weiter entfernt liegende Zeitreihenwerte einbezogen werden:

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

Mit dieser Differenzenbildung können saisonale Effekte ausgeblendet werden. Wie bei den einfachen Differenzen stellt sich natürlich die Frage, wie oft die saisonalen Differenzen gebildet werden müssen. Hinzu kommt das Problem, dass bei Zeitreihen mit Trend und Saison unter Umständen beide Arten von Differenzenbildungen kombiniert werden müssen. Eine heuristische Lösung bietet sich mit der Methode der *variaten Differenzen* an. Hier bildet man versuchsweise verschiedene Kombinationen von einfachen und saisonalen Differenzen d , D und entscheidet sich für diejenige, welche die kleinste Varianz der transformierten Reihe $(1-B)^d(1-B^s)^D y_t$ bewirkt. Die Vorstellung hinter dieser Entscheidungsregel ist die, dass über die Differenzenbildung zunächst die systematischen Anteile eliminiert werden. Weiteres Differenzenbilden führt dann aber wieder zu einer größeren Streuung durch das mehrfache Auftreten der Störungen, also zu einer Erhöhung der Varianz. Man sucht gewissermaßen einen geeigneten Kompromiss. Das lässt sich schon anhand eines einfachen linearen Trends erkennen:

$$y_t = a + b \cdot t + \varepsilon_t$$

$$\Delta y_t = y_t - y_{t-1} = a + b \cdot t + \varepsilon_t - (a + b \cdot (t-1) + \varepsilon_{t-1}) = b + \varepsilon_t - \varepsilon_{t-1}$$

$$\Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2} = \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}.$$

Bei Unabhängigkeit der Störungen verdoppelt sich die Störvarianz bei einfacher Differenzenbildung, bei zweifacher vervierfacht sie sich.

I. d. R. ist es günstiger, das jeweilige Verhältnis der Varianz zur Varianz der Ausgangsreihe in einer Tabelle darzustellen als die Varianzen selbst; dann gibt es keine Probleme mit den Größenordnungen.

Tab. 2.1: Tafel von Varianzverhältnissen

	$D = 0$	$D = 1$	$D = 2$...
$d = 0$	1	$s_{\Delta_s Y}^2 / s_Y^2$	$s_{\Delta_s^2 Y}^2 / s_Y^2$...
$d = 1$	$s_{\Delta Y}^2 / s_Y^2$	$s_{\Delta \Delta_s Y}^2 / s_Y^2$	$s_{\Delta \Delta_s^2 Y}^2 / s_Y^2$...
$d = 2$	$s_{\Delta^2 Y}^2 / s_Y^2$	$s_{\Delta^2 \Delta_s Y}^2 / s_Y^2$	$s_{\Delta^2 \Delta_s^2 Y}^2 / s_Y^2$...
:	:	:	:	

Beispiel 2.31 (CO₂-Konzentration über Mauna Loa – Fortsetzung von Seite 11)

Für den in der Abbildung 2.3 dargestellten Ausschnitt der Reihe MAUNALOA weist die zugehörige Tafel der variaten Differenzen darauf hin, dass einmal einfache und einmal saisonale Differenzen zu bilden sind, um Stationarität zu erreichen.

Einfachdiff	Saisdiff0	Saisdiff1	Saisdiff2	Saisdiff3
0	1.0000	0.0012	0.0021	0.0065
1	0.0038	0.0005	0.0014	0.0047
2	0.0024	0.0012	0.0037	0.0126
3	0.0035	0.0040	0.0120	0.0406

R-Code:

```

1 data(MAUNALOA)
2 y <- window(MAUNALOA,start=c(1980,1),end= c(2019,5))
3 vartable(y,12)

```

2.6 Einfache Extrapolationsverfahren

Eine der Standardaufgaben der Zeitreihenanalyse ist die Erstellung von Prognosen. Ist eine Zeitreihe y_1, \dots, y_N gegeben, so sind *Prognosen* Schätzungen zukünftiger Realisationen von Y_{N+h} , $h > 0$, auf der Basis der Beobachtungen bis zum Zeitpunkt N . Die Prognose von Y_{N+h} wird mit $\hat{Y}_{N,h}$ bezeichnet; $\hat{y}_{N,h}$ steht dann für den realisierten Wert. h ist der *Prognosehorizont* und $\hat{Y}_{N,h}$ heißt *h -Schritt Prognose*.

Es ist einsichtig, dass Prognosen von dem unterstellten Modell abhängen. In diesem Abschnitt werden nur zwei einfache Prognoseansätze vorgestellt. Weitere Verfahren basieren auf linearen Modellen, siehe Kapitel 3. Auch durch die Berücksichtigung externer, d. h. nicht in der Reihe enthaltener Informationen versucht man, bessere Prognosen zu erzielen. Dies führt dann zu multivariaten Reihen und in das Gebiet der Ökonometrie.

Oft möchte man den Trend bestimmen, um die langfristige Komponente zur Fortschreibung der Entwicklung in die Zukunft zu verwenden. Werden zur Modellierung des Trends Polynom verwendet, so ist eine einfache *Trendextrapolation* jedoch problematisch. Eine gute Anpassung im beobachteten Abschnitt der Zeitreihe, dem so genannten *Stützbereich*, ist noch lange keine Gewähr für eine erfolgreiche Prognose.

Ein wenig mag die Bestimmung von *Prognoseintervallen* helfen, um die Unsicherheiten bei der Trendextrapolation zu veranschaulichen. Prognoseintervalle geben mit einer vorgegebenen Wahrscheinlichkeit an, in welchem zentralen Bereich die zukünftige Zeitreihenvariable Y_{N+h} ihren Wert annehmen wird.

Bei der einfachen linearen Regression lauten die $1 - \alpha$ Prognoseintervalle unter der Voraussetzung normalverteilter Störungen ε_t :

$$\hat{\beta}_0 + \hat{\beta}_1 t \pm t_{N-2,1-\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{N} + \frac{(t - \bar{t})^2}{\sum_{t=1}^N (t - \bar{t})^2}} \quad (2.42)$$

Die Varianzschätzung ist dabei $\hat{\sigma}_\varepsilon^2 = \sum_{t=1}^N (y_t - \bar{y})^2 / (N - 2)$.

Für andere Polynomordnungen gelten geeignet verallgemeinerte Prognoseintervalle.

Beispiel 2.32 (Stromspitzenbedarf – Fortsetzung von Seite 1)

Wie dargestellt, war die Prognose des jährlichen Stromspitzenbedarfs von Berlin (West) Ende der 70er Jahre von Bedeutung. Nunmehr ist eine nachträgliche Beurteilung der Prognose der Reihe JLAST möglich.

Die Anpassung eines Polynoms erster Ordnung an die Werte für die Zeiten von 1954/55 bis 1978/79 ergab die geschätzte Trendgerade

$$\hat{T} = 301.060 + 51.531 \cdot t .$$

(22.699) (1.527)

Dabei wurden die Zeiten einfach mit 1, 2, ... an durchnummertiert. Die Abbildung der bis 1991/92 fortgeschriebene Reihe mit der Trendextrapolation und den 95 % Prognoseintervallen zeigt, dass die Prognose eindeutig zu hoch lag. Ein Grund dafür ist sicherlich die zweite Ölkrise 1979 und die damit einhergehende größere Sparsamkeit beim Energieverbrauch.

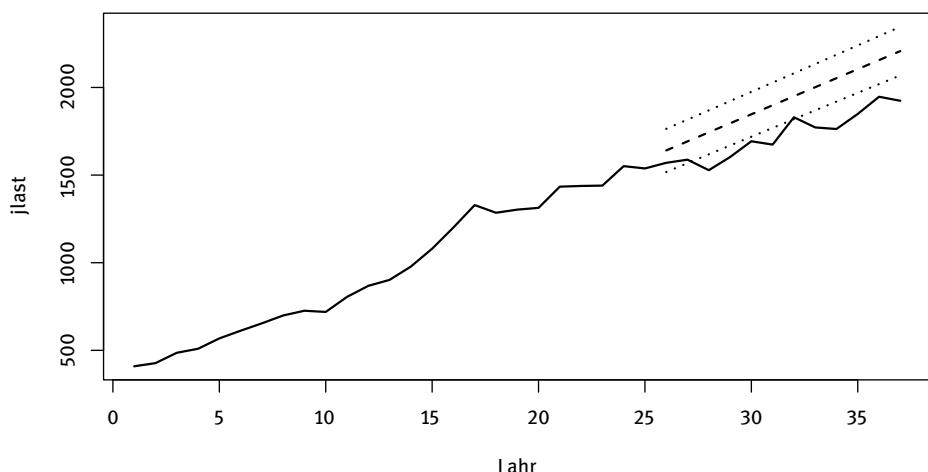


Abb. 2.23: JLAST mit Trend

R-Code:

```

1 data(PPDEMAND)
2 y <- PPDEMAND
3 n <- length(y)
4 t1 <- c(1:25)
5 y1 <- y[1:25]
6 out <- lm(y1~t1)
7 new <- data.frame(t1=c(26:37))
8 out1 <- predict(out,new,interval="prediction")
9 plot(c(1:n),y,type="l",ylim=c(min(y),max(out1[,3])))
10 lines(new$t1,out1[,1],lty=2)
11 lines(new$t1,out1[,2],lty=3)
12 lines(new$t1,out1[,3],lty=3)

```

Das Ergebnis einer linearen Regression mittels `lm` kann als Argument der generischen Funktion `predict` verwendet werden. Dazu sind die neuen Regressorwerte in der angegebenen Form zur Verfügung zu stellen. Das Ergebnis von `predict` ist eine Matrix mit drei Spalten, den prognostizierten Werten und den Prognoseunter- sowie -obergrenzen. Für die grafische Darstellung sind diese geeignet einzubauen.

Einen zweiten Ansatz zur Prognose bilden rekursive Fortschreibungen. Zentral sind Verfahren des *exponentiellen Glättens*. Sie gehen auf Brown (1962) zurück. Sein Ziel war es, eine Prognoseformel zu entwickeln, die mit möglichst wenig Speicherplatz auskommt. Den Hintergrund für diese Aufgabe bildete das Problem, dass die Prognosen häufig aufdatiert werden sollten, dass also stets neue Beobachtungen hinzukamen und die Prognosen unter Verwendung aller vergangenen Werte zu bestimmen waren.

Ausgangspunkt ist das einfache Modell

$$Y_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

Auf der Basis von N Beobachtungen ist das arithmetische Mittel \bar{y}_N die beste Schätzung von μ und auch die beste Prognose zukünftiger Werte.

Unterstellt man nun, dass eine neue Beobachtung hinzukommt, so braucht man zur Bestimmung von \bar{y}_{N+1} nicht auf alle vergangenen Werte zurückzugreifen. Es reichen vielmehr der letzte Mittelwert und die neue Beobachtung:

$$\bar{y}_{N+1} = \frac{N}{N+1} \bar{y}_N + \left(1 - \frac{N}{N+1}\right) y_{N+1}.$$

Es ist plausibel, für Prognosen der neuesten Beobachtung nicht stets kleinere Gewichte zu geben, sondern ein festes. Damit erhält man schon die Prognoseformel des *einfachen exponentiellen Glättens*:

$$\hat{y}_{N,1} = \beta \hat{y}_{N-1,1} + (1 - \beta) y_N = (1 - \alpha) \hat{y}_{N-1,1} + \alpha y_N. \quad (2.43)$$

Das exponentielle Glätten wird auch als *rekursiver Filter* bezeichnet, da das Ergebnis des letzten Schrittes, die Bestimmung von $\hat{y}_{N-1,1}$, im neuen Schritt, die Bestimmung von $\hat{y}_{N,1}$, verwendet wird.

Für die praktische Bestimmung von Prognosen ist die Rekursion irgendwie zu beginnen. Das kann einmal geschehen durch Verwendung der expliziten Form (5.2), wobei die y_t mit $t \leq 0$ durch das allgemeine Mittel ersetzt werden. Meist wird aber $\hat{y}_{1,1} = y_1$ gesetzt; die weiteren $\hat{y}_{t,1}$ werden dann gemäß (2.43) rekursiv bestimmt.

Die Rekursionsformel (2.43) macht deutlich, wie die Wahl von α den Verlauf der Einschritt-Prognosen beeinflusst. Bei einem kleinen Wert von α , $\alpha \approx 0$, wird der Anteil von $\hat{y}_{N-1,1}$ sehr stark gewichtet, der Summand αy_N wiegt sehr gering. Daher reagiert die Prognose nur wenig auf den jeweils letzten Wert. In der Anwendung wird die Reihe der Ein-Schritt-Prognosen sehr stabil sein. Umgekehrt sieht die Situation bei einem großen Wert von α aus, $\alpha \approx 1$. Die Prognosen reagieren sehr empfindlich auf

den jeweils aktuellsten Wert, die Reihe der Ein-Schritt-Prognosen ist verhältnismäßig instabil. In Anwendungen wird α oft so bestimmt, dass die Summe der quadrierten Einschritt-Prognosefehler minimiert wird:

$$\sum_{t=m}^{N-1} (y_{t+1} - \hat{y}_{t-1,1}(\alpha))^2 \stackrel{!}{=} \min_{\alpha}. \quad (2.44)$$

Der Zeitpunkt $m > 1$ wird dabei so groß gewählt, dass der Effekt der Anfangswerte vernachlässigt werden kann.

h -Schritt Prognosen mit $h > 0$ erhält man durch Fortschreiben der Einschritt-Prognose. Da das zugrunde liegende Modell eines mit einem konstanten Niveau ist, wird für $h > 1$ einfach $\hat{y}_{N,h} = \hat{y}_{N,1}$ gesetzt.

Beispiel 2.33 (Tabellenplatz HSV)

Der Hamburger Sportverein erlebte seit dem Beginn der Fußball-Bundesliga einiges an Auf und Ab. Eine Möglichkeit für einen Nichtfachmann, eine Prognose des Tabellenplatzes der nächsten Saison abzugeben, ist der Einsatz des exponentiellen Glättens. In der Abbildung ist der jeweilige Tabellenplatz angegeben sowie die Ein-Schritt-Prognose mit dem exponentiellen Glätten bei Verwendung des Parameters $\alpha = 0.31$. Spitzen und Minima werden nicht nur abgeflacht sondern erscheinen auch in Form eines ‚Nachhinkens‘ in der Bewegung.

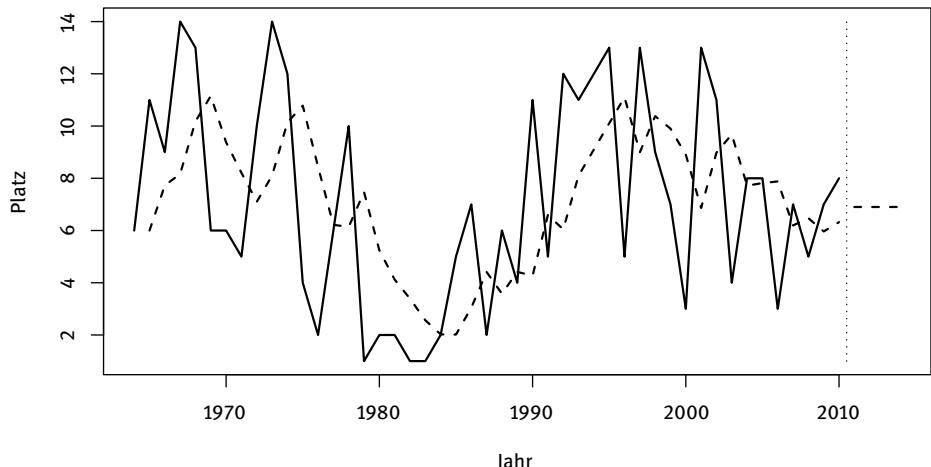


Abb. 2.24: Tabellenplatz HSV mit Prognosen

R-Code:

```

1 data(HSV)
2 y <-HSV
3 out <- HoltWinters(y,alpha=NULL,beta=FALSE,gamma=FALSE)
4 p <- predict(out, n.ahead=4)
5 plot(out,p)

```

Die Funktion HoltWinters ist allgemeiner als das einfache exponentielle Glätten. Letzteres wird aber durchgeführt, wenn die optionalen Parameter beta und gamma auf FALSE gesetzt werden. Mit alpha=NULL wird dieser Parameter automatisch bestimmt. Die Prognose geschieht durch Anwenden der Funktion predict auf das Ergebnis der Glättung.

Um das exponentielle Glätten anwenden zu können, braucht das Niveau der Reihe nicht streng konstant zu sein. Es reicht schon, dass es sich nur langsam ändert, so dass es lokal als konstant angesehen werden kann. Weist eine Zeitreihe jedoch einen deutlichen Trend auf, so sollte das geeignet berücksichtigt werden. Bei einem linearen Trend ist auch die Steigung mit einer dem einfachen exponentiellen Glätten entsprechenden Rekursion fortzuschreiben. Das geschieht bei dem Holt-Winters-Verfahren; dieses erlaubt zusätzlich eine Saison zu berücksichtigen.

Das *Holt-Winters-Verfahren* gibt es in verschiedenen Varianten. An dieser Stelle wird nur auf die additive Version mit einfacherem lokalem linearem Trend eingegangen. Zu Varianten sei auf das Kapitel 5 verwiesen.

Das zugrunde liegende Modell ist

$$Y_t = T_t + S_t + \varepsilon_t \quad (2.45)$$

mit

$$T_t = \alpha + bt, \quad S_t = S_{t-s}. \quad (2.46)$$

Dabei ist s die Saisonperiode. Es wird wieder davon ausgegangen, dass dieser Ansatz zumindest lokal gilt. Die Idee ist dann, Schätzungen für die Parameter in der gleichen Rekursion fortzuschreiben, wie es beim einfachen exponentiellen Glätten geschieht. Dabei ist zu beachten, dass für die Schätzung der Parameter auf entsprechende Differenzen zurückgegriffen werden muss. So ist das Niveau etwa anhand der saisonbereinigten Reihe $Y_t - S_t$ fortzuschreiben. Man muss also dafür sorgen, dass die Definition des Verfahrens nicht zirkulär wird, d. h. man darf nicht auf noch nicht berechnete Größen zugreifen. Das führt zu folgenden Rekursionen:

$$\hat{b}_t = (1 - \beta)\hat{b}_{t-1} + \beta(\hat{T}_{t-1} - \hat{T}_{t-2}) \quad (2.47a)$$

$$\hat{T}_t = (1 - \alpha)(\hat{T}_{t-1} + \hat{b}_t) + \alpha(y_t - \hat{S}_{t-s}) \quad (2.47b)$$

$$\hat{S}_t = (1 - \gamma)\hat{S}_{t-s} + \gamma(y_t - \hat{T}_t) \quad (2.47c)$$

$$\hat{y}_{t,h} = \hat{T}_t + h\hat{b}_t + \hat{S}_{t+h-s} \quad (h = 1, \dots, s). \quad (2.47d)$$

Die Rekursion kann erst bei $N = s + 1$ beginnen. Für die ersten s Werte wählt man plausible Startwerte, etwa:

$$\hat{T}_{s-1} = \hat{T}_s = \bar{y}_s = \frac{1}{s} \sum_{t=1}^s y_t, \quad \hat{b}_s = 0, \quad \hat{S}_t = y_t - \bar{y}_s \quad (t = 1, \dots, s).$$

Beispiel 2.34 (Auftragseingänge für Maschinen)

Ein Unternehmen der Investitionsgüterindustrie verfolgt die Auftragseingänge für Werkzeugmaschinen in einem Fertigungsbetrieb. Sie ist daran interessiert, herauszufinden, wie die Reihen gegebenenfalls zu prognostizieren seien. Betrachtet wird hier die Reihe MASCHINEN der ‚Auftragseingänge für spezielle Maschinen in einem Unternehmen‘ für den Zeitraum 01.1972–08.1987.

Der erste Schritt zur Beantwortung der Fragen besteht in der grafischen Darstellung der Zeitreihe. Sie ist in der Abbildung 2.25 mit wiedergegeben. Da die Reihe eine Saison aufweist soll eine Prognose mit dem Holt-Winters-Verfahren durchgeführt werden. Das Holt-Winters-Verfahren kommt auch mit den sich langsam verändernden Trend- und Saisonstrukturen klar. Weiter hat es Verfahren den Vorteil, dass die einzelnen extremen Werte nicht so stark nachwirken. Die Prognose am aktuellen Rand sollte daher relativ unbeeinflusst von den extremen Werten sein.

Die sich bei der Auswahl der Optimierung ergebenen Koeffizienten sind:

$$\hat{\alpha} = 0.289, \hat{\beta} = 0.033, \hat{\gamma} = 0.245 .$$

Bei $\beta = 0$ wird die Steigung null und es wird ein lokal konstantes Modell verwendet. Es ist einsichtig, dass hier mit 0.033 ein Wert am Rand ausgegeben wird. Die Reihe zeigt ja eine längerfristige Niveauveränderung, aber kein überwiegenderes Ansteigen bzw. Abfallen. Dies führt speziell zu Prognosen, deren Niveau sich nicht ändert.

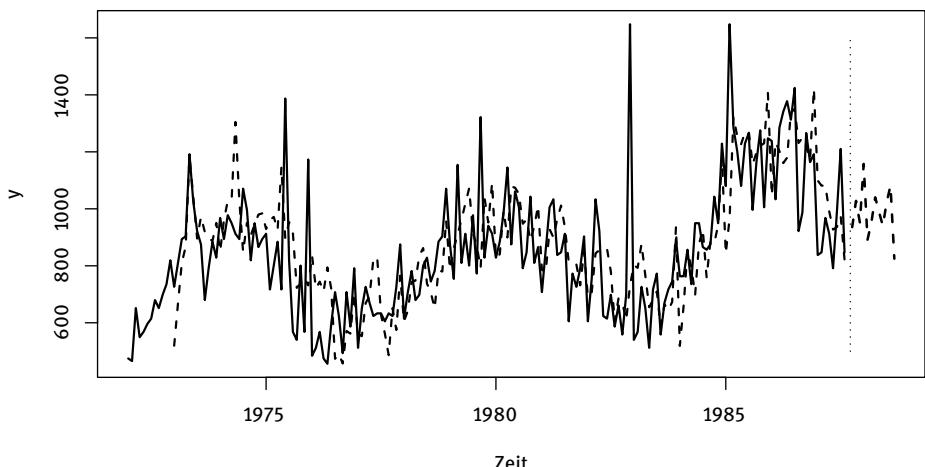


Abb. 2.25: Auftragseingänge für Werkzeugmaschinen mit Prognosen

R-Code:

```

1 data(ENGINES)
2 y <- ENGINES
3 out <- HoltWinters(y)
4 p <- predict(out, n.ahead=12)
5 plot(out,p)

```

Wird bei der Funktion `HoltWinters` nur eine Zeitreihe mit entsprechendem Attribut angegeben, geschieht der Rest automatisch. Die Glättungsparameter werden geeignet optimiert.

2.7 R-Funktionen

Eine grafische Darstellung einer Zeitreihe `y`, die schon ein Zeitreihenobjekt ist, bekommt man mit `plot(y)`. Ist `y` einfach ein Datenvektor, so ruft man `plot(y, type="l")` oder `ts.plot(y)` auf.

Die Funktionen für die Berechnung von Mittelwert, Varianz und Standardabweichung einer Zeitreihe `y` sind `mean(y)`, `var(y)` und `sd(y)`. Die Schiefe und die Wölbung sind weniger für originale Zeitreihen interessant als für Residuen nach dem Anpassen eines Modells. Das Paket `moments` stellt hierfür die beiden Funktionen `skewness` und `kurtosis` bereit.

Bei dem Einsatz der Bootstrap-Methode lässt sich die Blocklänge mit der Funktion `b.star` des Paketes `np` bestimmen. Das Paket `boot` stellt zahlreiche Möglichkeiten für Bootstrap-Methoden bereit. `tsboot` ist speziell für das Bootstrap von Zeitreihen gedacht.

`acf` berechnet die Autokorrelations- oder die Autokovarianzfunktion und stellt sie auch grafisch dar, je nach Wahl der Optionen. Ist `y` ein Zeitreihenobjekt mit einem Zeitreihenattribut `frequency > 1`, also etwa eine Zeitreihe von Quartals- oder Monatsdaten, so werden die Lags in Einheiten der Basiszeit angegeben, und nicht in Anzahlen von Beobachtungen.

Die Funktion `statcheck` aus `tsapp` bestimmt eine Matrix mit den arithmetischen Mitteln und Varianzen von Reihensegmenten und erstellt eine Grafik, in der die ACF's der Reihensegmente übereinander geplottet sind.

Für die Bestimmung einer Kleinsten-Quadrate-Schätzung bietet R vor allem die Funktion `lm`, „linear models“, an. Eine Konstante wird per Voreinstellung automatisch berücksichtigt, mit `-1` wird sie ausgeschlossen. Die Regressoren sind in der Regel vorab zu bestimmen. Es kann aber auch über `I(...)` ein Regressor im Argument als Funktion berechnet werden etwa `I(t^2)`.

Normalverteilungs-QQ-Diagramme werden mit `qqnorm` erstellt. `abline`, mit den beiden Argumenten `Achsenabschnitt` und `Steigung`, fügt eine Ausgleichsgerade hinzu.

Drei Tests auf Normalverteilung sind der Shapiro-Wilk-Test, der Kolmogorov-Smirnov-Test und der bei Ökonometrikern beliebte Jarque-Bera-Test. Der letztergenannte Test kombiniert die Schiefe und die Kurtosis zu einer Teststatistik. Die R-Funktionen dafür sind `shapiro.test`, `ks.test` und `jarque.bera.test` aus dem Paket `tseries`. Das Paket `nor.test` bietet weitere Optionen, wie zum Beispiel die Funktion `ad.test` für den Anderson-Darling-Test.

Für die Anpassung nichtlinearer Trendmodelle ist die Funktion `nls`, nichtlineare Kleinstquadrat-Methode, vorgesehen. Zwei andere Funktionen, mit denen die zugehörige Minimierungsaufgabe angegangen werden kann, sind `nlm` und `optim`. Letztere hat als Standard eine ableitungsfreie Methode zur Bestimmung eines Minimums.

Mit der Funktion `filter` können Filtrationen von Zeitreihen bestimmt werden. Bei dieser Funktion müssen die Gewichte explizit als Vektor angegeben werden. Zur vereinfachten Anwendung gleitender Durchschnitte dient die Funktion `moveav` aus `tsapp`. Differenzen werden mit der Funktion `diff` gebildet. Das Paket `mFilter` offeriert verschiedene Filter.

Glättende Splines können mit der Funktion `smooth.spline` bestimmt werden. Eine andere Implementierung ist in `tsapp` mit der Funktion `smoothls` gegeben. Diese nutzt eine spezielle Struktur des Problems aus, um auch bei langen Reihen eine simple lineare Regression durchführen zu können. Eine weitere Möglichkeit besteht in dem Einsatz der Funktion `hpfilter` des Paketes `mFilter`, welche den Hodrick-Prescott-FILTER implementiert.

R bietet mit `decompose` ein einfaches Saisonbereinigungsverfahren auf der Basis gleitender Durchschnitte. Die Funktion kann additive und multiplikative saisonale Komponenten bestimmen. Allerdings werden für letztere auch multiplikative Fehlerverknüpfungen unterstellt; so kann die multiplikative Variante durch Logarithmieren auf die additive zurückgeführt werden. `y` muss ein Zeitreihenobjekt sein; andernfalls ist die Saisonperiode auch nicht bekannt. Zu einem robusten Saisonbereinigungsverfahren mit der Funktion `stl` siehe das Kapitel 12.

Die Funktion `simpledecomp` in dem Paket `tsapp` basiert auf einer multiplen Regression, bei der die Saisonkomponenten effektkodiert sind. Als Eingabe kann auch ein Vektor von Werten dienen. Die Funktion `splinedecom` ergibt eine Splines-Zerlegung. Sie basiert auf einer sparsamen Form der Speicherung von Matrizen mit Bandstruktur.

Eine saisonal adjustierte Reihe, eine Reihe also, aus der nur die Saison entfernt ist, bekommt man mit `seasadj`. Diese Funktion befindet sich im Paket `forecast`. Die Funktion `seasonal.smooth` aus dem Paket `mar1s` extrahiert die saisonale Komponente einer Zeitreihe indem ihr eine Linearkombination glatter periodischer Funktionen angepasst wird. `seasonal.ave` aus demselben Paket bildet die Durchschnitte der Zeitreihenwerte über die gleiche Saisonperiode und ergibt so eine Saisonschätzung mit konstanter Saisonfigur.

Mit der `tsapp`-Funktion `monthplot` werden die saisonalen Teilreihen grafisch dargestellt, so dass die Entwicklung der saisonalen Komponente deutlich wird. Es wird

jeweils der Durchschnitt der zugehörigen Teilreihe genommen; darauf werden die zugehörigen Saisonaußschläge platziert. Die Funktion `plot.seas.sum` des Paketes `seas` stellt normalisierte saisonale Summen mittels Boxplots dar.

Die Funktion `vartable` in `tsapp` erstellt eine Tafel der variaten Differenzen.

`HoltWinters` ist eine übergreifende Funktion für die Methoden des exponentiellen Glättens. Für das volle Holt-Winters-Verfahren muss `y` ein Zeitreihenobjekt sein.

Eine Funktion zur Zerlegung einer Zeitreihe auf der Basis eines automatisch bestimmten geeigneten Exponentiellen-Glättung-Modells stellt das Paket `forecast` mit der Funktion `ets` bereit. Das einzige notwendige Argument ist die Zeitreihe. Die Automatik hat ihren Preis; es dauert etwas, bis das Ergebnis erreicht ist.

Die Prognose einer Zeitreihe auf der Basis eines mit `ets` bestimmten Modells kann mit der Funktion `predict` und dem anschließenden `plot` erreicht werden.

Komfortabler ist es, die Funktion `forecast` aus `forecast` zu verwenden. Intern wird dann erst einmal das beste Exponentielle-Glättungsmodell bestimmt und dann die Prognose.

Die Pakete `season` und `seas` bieten Verfahren zur Komponentenzerlegung in R speziell für Anwendungen aus der Gesundheitsökonomie und bzw. Klimaforschung. Die Funktion `decomp` im Paket `timsac` ist empfehlenswert. Hier gibt es auch die Möglichkeit, die Unterschiedlichkeit der Arbeitstage pro Monat zu berücksichtigen.

3 Lineare Zeitreihenmodelle

3.1 Autoregressive Modelle

Definition und grundlegende Eigenschaften

Ausgangspunkt der eigentlichen (univariaten) Zeitreihenanalyse bildet das Bestreben, die Reihen aus sich selbst zu erklären. Um diesen Ansatz zu verfolgen, wird von nun an unterstellt, dass die Reihen stationär sind; dementsprechend werden im Folgenden (sofern nicht extra anders angegeben) auch nur stationäre Modelle betrachtet. Schematisch kann die Zielvorstellung folgendermaßen formuliert werden:

$$\text{Wert im Zeitpunkt } t = f \left(\begin{array}{c} \text{Werte in} \\ \text{vorhergehenden} \\ \text{Zeitpunkten} \end{array} \right) + \text{Rest ,}$$

bzw. formaler:

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots) + \varepsilon_t . \quad (3.1)$$

Die einfachste Wahl für die Funktion f ist die Bildung einer Linearkombination der vorangegangenen Werte. Dabei werden nur Werte bis zu einem geeigneten Vergangenheitshorizont berücksichtigt, vom dem man annehmen kann, dass die Auswirkung direkt bis zur jeweiligen Gegenwart reicht. Das Modell wird dann durch die folgende Gleichung festgelegt:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t . \quad (3.2)$$

(ε_t) wird als White-Noise-Prozess mit $E(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$ unterstellt.

Mit ε_t hat auch Y_t den Erwartungswert null, falls $\alpha_0 = 0$ gilt. Der Erwartungswert μ von Y_t ist dabei im Fall $\alpha_0 \neq 0$ nicht gleich α_0 . Er ergibt sich vielmehr durch Erwartungswertbildung auf beiden Seiten der Gleichung (3.2):

$$\mu = \alpha_0 + \alpha_1 \mu + \dots + \alpha_p \mu \quad \Rightarrow \quad \mu = \frac{\alpha_0}{1 - \alpha_1 - \dots - \alpha_p} .$$

Alternativ kann daher die Gleichung (3.2) in die Form

$$(Y_t - \mu) = \alpha_1 (Y_{t-1} - \mu) + \dots + \alpha_p (Y_{t-p} - \mu) + \varepsilon_t \quad (3.3)$$

bringen. Im Folgenden wird in der Regel von den zentrierten Reihen ausgegangen. Für diese kann mit der Modellgleichung (3.2) mit $\alpha_0 = 0$ gearbeitet werden.

Definition 3.1 (Autoregressiver Prozess)

Wenn ein Prozess (Y_t) die Beziehung (3.2) bzw. (3.3) erfüllt, heißt er ein *autoregressiver Prozess p-ter Ordnung*, kurz *AR[p]*-Prozess. Die Innovationen ε_t werden als bei null zentriert mit endlicher Varianz, $\text{Var}(\varepsilon_t) < \infty$, vorausgesetzt.

Die Bezeichnung ist nahe liegend: Die Definitionsgleichung entspricht einem Regressionsansatz, nur sind die Regressoren eben die zeitlich verschobenen Y -Werte selbst. Die Analogie zur Regression legt die Bezeichnung ‚Störungen‘ für die ε_t nahe. Sie werden aber besser als *Innovationen* oder als ‚den AR-Prozess treibende Größen‘ bezeichnet. Wie nämlich später noch deutlich wird, würden die Realisationen oder Pfade des Prozesses bei $\varepsilon_{t+1} = \varepsilon_{t+2} = \dots = 0$ absterben, d. h. für $t \rightarrow \infty$ gegen null gehen. Für einen AR[1]-Prozess sieht man das unmittelbar. Sei $Y_t = \alpha Y_{t-1} + \varepsilon_t$ und seien $\varepsilon_{t+1} = \varepsilon_{t+2} = \dots = 0$. Dann ist $Y_{t+h} = \alpha^h Y_t$ und für $|\alpha| < 1$ geht das gegen null.

Die entscheidende Größe für die Charakterisierung der Eigenschaften autoregressiver Prozesse ist die ACF. Schon anhand eines AR[1]-Prozesses können die wichtigsten Aspekte erkannt werden. Erfüllte also (Y_t) die Gleichung

$$Y_t = \alpha Y_{t-1} + \varepsilon_t \quad (3.4)$$

mit $E(\varepsilon_t) = 0$ und $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$. Für die Varianz erhält man zunächst:

$$\begin{aligned} \text{Var}(Y_t) &= E(Y_t^2) = E((\alpha Y_{t-1} + \varepsilon_t)(\alpha Y_{t-1} + \varepsilon_t)) = E(\alpha^2 Y_{t-1}^2 + 2\alpha Y_{t-1} \varepsilon_t + \varepsilon_t^2) \\ &= \alpha^2 \text{Var}(Y_{t-1}) + \text{Var}(\varepsilon_t). \end{aligned}$$

Damit gilt dann unter der Voraussetzung der Stationarität, speziell mit $\text{Var}(Y_t) = \text{Var}(Y_{t-1})$:

$$\text{Var}(Y_t) = \frac{\sigma_\varepsilon^2}{1 - \alpha^2}.$$

Da dieser Quotient nur für $1 - \alpha^2 > 0$ eine sinnvolle endliche Varianz darstellt, ist schon hier zu erkennen, dass $|\alpha| < 1$ für die Stationarität nötig ist.

Für $|\alpha| < 1$ ergibt sich die Autokovarianz zum Lag 1 mit einer ähnlichen Rechnung wie bei der Varianz:

$$\text{Cov}(Y_{t+1}, Y_t) = E(Y_{t+1} Y_t) = \alpha \frac{\sigma_\varepsilon^2}{1 - \alpha^2}.$$

Das führt zu $\rho_1 = \alpha$. Durch weiteres Einsetzen von $Y_{t+\tau}$ in die Definitionsgleichung und anschließende Erwartungswertbildung erhält man allgemein für einen AR[1]-Prozess:

$$\gamma_\tau = \alpha^{|\tau|} \frac{\sigma_\varepsilon^2}{1 - \alpha^2}, \quad \rho_\tau = \alpha^{|\tau|} \quad (\tau = \pm 1, \pm 2, \dots). \quad (3.5)$$

Die ACF (ρ_τ) klingt für $|\alpha| < 1$ exponentiell ab. Für den Grenzfall $|\alpha| \rightarrow 1$ wird sie konstant 1 oder alternierend ± 1 . Wie man oben sieht, ist die Umformung zur Gewinnung der Varianz dann nicht mehr möglich, da man durch null teilen würde. Startet man aber in der Definitionsgleichung (3.4) beim Zeitpunkt null mit $Y_0 = \varepsilon_0$ und schreibt dann den Prozess fort, so werden die Varianzen mit dem Zeitindex immer größer. Das ist klar, da es sich dann gerade um einen Random-Walk-Prozess handelt. Die Autokorrelationen hängen bei einem Random-Walk-Prozess nicht nur vom zeitlichen Abstand der Prozessvariablen ab. Vielmehr gilt:

$$\text{Corr}(Y_t, Y_{t+\tau}) = \sqrt{\frac{t}{t+\tau}}.$$

Die empirische Autokorrelationsfunktion, berechnet aus einer Realisation eines solchen Prozesses, geht nur langsam gegen null. Sie ergibt jedoch insgesamt noch ein ‚vernünftiges‘ Bild. Daher spricht man von einem grenzstationären Prozess.

Für $|\alpha| > 1$ sind Prozesse, die der Rekursion (3.4) genügen, explodierend, d. h. die Realisationen werden sehr schnell (betragsmäßig) immer größer.

Die Diskussion für AR[1]-Prozesse macht deutlich, dass auch für allgemeine AR[p]-Prozesse geeignete Stationaritätsbedingungen zu fordern sind. Zunächst erhält man durch Multiplikation von (3.3) mit $Y_{t-\tau}$ und anschließender Erwartungswertbildung für die Autokovarianzfunktion eines AR[p]-Prozesses:

$$\gamma_\tau = \begin{cases} \alpha_1 \gamma_1 + \cdots + \alpha_p \gamma_p + \sigma_\varepsilon^2 & \text{für } \tau = 0 \\ \alpha_1 \gamma_{\tau-1} + \cdots + \alpha_p \gamma_{\tau-p} & \text{für } \tau > 0 \end{cases} \quad (3.6)$$

Diese Gleichungen werden als *Yule-Walker-Gleichungen* bezeichnet. Sie können für $\tau > q$ offensichtlich auch in den Korrelationen formuliert werden, da auf beiden Seiten durch γ_0 geteilt werden kann.

Sofern die *charakteristische Gleichung* des AR[p]-Prozesses $Y_t - \alpha_1 Y_{t-1} - \alpha_2 Y_{t-2} - \cdots - \alpha_p Y_{t-p} = \varepsilon_t$,

$$1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_p z^p = 0, \quad (3.7)$$

nur Lösungen z_1, \dots, z_p hat, die betragsmäßig größer als eins sind, mit $|z_1| > 1, \dots, |z_p| > 1$, gilt für die Autokorrelationsfunktion:

$$\rho_\tau = A_1 \left(\frac{1}{z_1} \right)^\tau + A_2 \left(\frac{1}{z_2} \right)^\tau + \cdots + A_p \left(\frac{1}{z_p} \right)^\tau; \quad (3.8)$$

dabei sind A_1, \dots, A_p Konstanten und $\tau \geq 0$. Der AR[p]-Prozess ist dann stationär. Dann folgt aus der Gleichung (3.8) auch $\rho_\tau \rightarrow 0$ für $\tau \rightarrow \infty$.

Da die Lösungen der charakteristischen Gleichung auch komplexe Zahlen sein können, kennzeichnet man diese Eigenschaft in der Weise, dass alle Nullstellen der charakteristischen Gleichung außerhalb des Einheitskreises liegen.

Beispiel 3.2 (Stationarität eines AR[2]-Prozesses)

Für AR[2]-Prozesse lauten die Lösungen der charakteristischen Gleichung $1 - \alpha_1 z - \alpha_2 z^2 = 0$:

$$z_1 = \frac{-\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2}}{2\alpha_2}, \quad z_2 = \frac{-\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2}}{2\alpha_2}.$$

Zusammen mit den Beziehungen zwischen Parametern und Autokorrelationen

$$\alpha_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2} \quad \text{und} \quad \alpha_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},$$

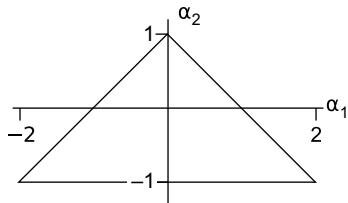


Abb. 3.1: Stationaritätsdreieck für AR[2]-Prozesse

ergeben sich für die Parameterwerte die folgenden Bedingungen für die Stationarität, da beide Autokorrelationen betragsmäßig kleiner als 1 sein müssen:

$$\alpha_1 + \alpha_2 < 1, \quad \alpha_2 - \alpha_1 < 1, \quad \alpha_2 > -1. \quad (3.9)$$

Das führt zu dem in der Abbildung 3.1 dargestellten Stationaritätsdreieck.

AR-Prozesse sind theoretisch am leichtesten zu handhaben, wenn eine unendliche Vergangenheit zugelassen wird. Dieser theoretische Zugang hat allerdings keine empirische Entsprechung; für empirische Zeitreihen wird es irgendwo einen Zeitpunkt „0“ geben. Allerdings hat die endliche Vergangenheit keinen signifikanten Einfluss auf die Momente erster und zweiter Ordnung, wenn erst einmal eine relativ kurze Einschwingphase vergangen ist. Dieses Einschwing-Verhalten wird bei der Simulation von autoregressiven Prozessen ausgenutzt. Man erzeugt unabhängige Zufallszahlen aus der Verteilung der ε_t und wendet die Definitionsgleichung $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t$ mit $Y_t = 0$ für $t \leq 0$ rekursiv darauf an. Der erste Teil der so erhaltenen Reihe wird verworfen und die letzten N Werte werden als Realisation des AR-Prozesses angesehen. Für Gaußsche AR-Prozesse gibt es allerdings effizientere Methoden.

Beispiel 3.3 (Simulation eines AR[2]-Prozesses)

Für den AR[2]-Prozess $Y_t = 0.5 Y_{t-1} - 0.9 Y_{t-2} + \varepsilon_t$ wird eine Realisation der Länge $N = 50$ erzeugt. Es werden 100 Beobachtungen generiert; die ersten 50 werden verworfen.

R-Code:

```

1 n <- 50
2 a <- c(0.5, -0.9)
3 e <- rnorm(n+50)
4 y <- e
5 t <- 3
6 while(t < (n+50)){ t <- t+1; y[t] <- sum(a*y[t-c(1,2)]) + e[t] }
7 ts.plot(y)
8 lines(c(50,50),c(-4,6),lty=2)

```

Die Realisation des AR[2]-Prozesses wird über eine Rekursion entsprechend der Definitionsgleichung erzeugt.

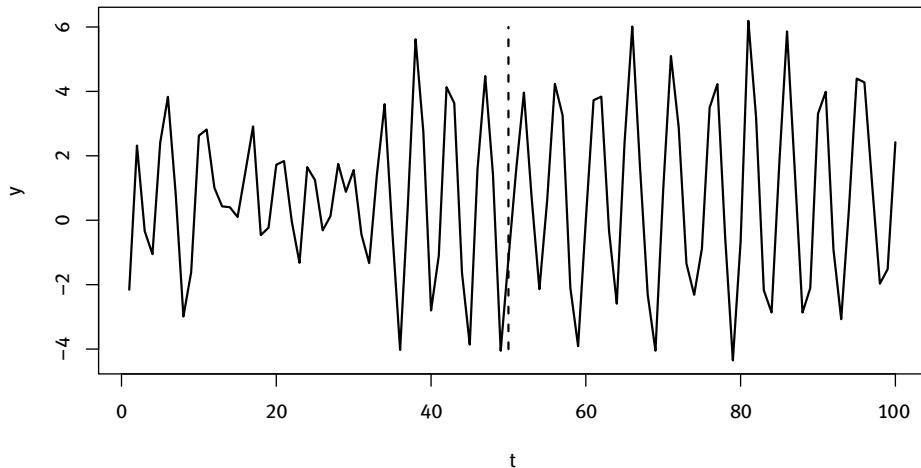


Abb. 3.2: Simulierter AR[2]-Prozess

Schätzen von AR-Parametern

Die formale Ähnlichkeit der Definitionsgleichung des AR-Prozesses mit einer Regressionsbeziehung drängt förmlich den Einsatz der Methoden der Regressionsrechnung für autoregressive Modelle auf. Jedoch sind die Voraussetzungen des Regressionsmodells in AR-Modellen nicht gültig. Erst bei hinreichend langen Zeitreihen können alle Resultate eines Regressionsprogrammes wieder als (approximativ) gültig angesehen werden. Daher sind spezifische *Methoden zum Schätzen der Parameter* in AR-Modellen von Interesse. Die wichtigsten werden im Folgenden aufgeführt. Um die Schätzverfahren und ihre Unterschiede zu verdeutlichen, wird jeweils die Situation betrachtet, dass anhand einer zentrierten Zeitreihe y_1, \dots, y_N die Parameter eines AR[2]-Prozesses $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t$ geschätzt werden sollen.

CLS (Conditional Least Sums of Squares)

Die CLS-Methode ist im Wesentlichen die Schätzung nach der Methode der kleinsten Quadrate, wie man sie aus der klassischen Regressionsanalyse kennt. Es wird, wie oben vereinbart, von einer mittelwertbereinigten Reihe, $\bar{y} = 0$, ausgegangen. $\hat{\alpha}_1, \hat{\alpha}_2$ werden dann über

$$\sum_{t=3}^N [y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2}]^2 \stackrel{!}{=} \min . \quad (3.10)$$

bestimmt. Dies führt auf das Gleichungssystem

$$\hat{\alpha}_1 \frac{1}{N} \sum_{t=3}^N y_{t-1}^2 + \hat{\alpha}_2 \frac{1}{N} \sum_{t=3}^N y_{t-1} y_{t-2} = \frac{1}{N} \sum_{t=3}^N y_t y_{t-1} \quad (3.11a)$$

$$\hat{\alpha}_1 \frac{1}{N} \sum_{t=3}^N y_{t-1} y_{t-2} + \hat{\alpha}_2 \frac{1}{N} \sum_{t=3}^N y_{t-2}^2 = \frac{1}{N} \sum_{t=3}^N y_t y_{t-2} . \quad (3.11b)$$

Die Summation lässt sich auch bei $t = 1$ beginnend schreiben, wenn die beiden Werte y_{-1}, y_0 identisch null gesetzt werden. Gerechtfertigt wird diese Setzung dadurch, dass die Werte damit dem Erwartungswert der Zeitreihenvariablen gleichgesetzt werden. Die Bedingung, dass sie als ‚fest‘ eingehen, ist die Begründung für die Bezeichnung CLS.

ULS (Unconditional Least Sums of Squares)

In einer Verfeinerung der CLS-Methode werden nach einer Vorabschätzung der Parameter die Werte $y_0, y_{-1}, y_{-2}, \dots$ mit Hilfe von Prognosen der Reihe in die Vergangenheit geschätzt. Auf die Prognosen wird später eingegangen. Die ULS-Methode verwendet dann die Lösung der Gleichungen (3.11a), (3.11b), wobei die Summation bei $t = 1$ beginnt. Für y_0, y_{-1} werden die ‚prognostizierten‘ Werte verwendet.

ML (Maximum Likelihood)

Hier wird ein Gauß-Prozess unterstellt. Für die ML-Schätzung geht man dann von der gemeinsamen Normalverteilung der Zufallsvariablen Y_1, \dots, Y_N aus und betrachtet die logarithmierte Likelihoodfunktion. Diese hat bei AR[2]-Prozessen die Gestalt:

$$\begin{aligned} \mathcal{L}(\alpha_1, \alpha_2, \sigma^2) = & -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) + \frac{1}{2} \ln \left((1 + \alpha_2)^2 ((1 - \alpha_2)^2 - \alpha_1^2) \right) \\ & - \frac{1 + \alpha_2}{2\sigma^2} \left[(1 - \alpha_2)y_1^2 - 2\alpha_1 y_1 y_2 + (1 - \alpha_2)y_2^2 \right] \\ & - \frac{1}{2\sigma^2} \sum_{t=3}^N (y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2})^2. \end{aligned} \quad (3.12)$$

Diese Loglikelihood wird bzgl. der Parameter maximiert. Die Parameterwerte, bei denen das Maximum angenommen wird, sind dann die Schätzwerte. Da die letzte Summe mit negativem Vorzeichen eingeht, wäre die Maximierung äquivalent zur Minimierung von $\sum_{t=3}^N (y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2})^2$, wenn nur dieser Term relevant wäre. Die Maximierung der Loglikelihood würde dann also auf die CLS Schätzung führen. In den anderen Teilen der Loglikelihoodfunktion kommen nur die Parameter und die ersten beiden Reihenwerte vor.

YW (Yule-Walker)

Die den autoregressiven Prozessen eigene Rekursionsbeziehung (3.6) der Autokorrelationsfunktion führt zu einem besonderen Zusammenhang von Koeffizienten und Autokorrelationen. Dieser kann mittels der *Levinson-Durbin-Rekursion* zu einer numerisch sehr effizienten Schätzung ausgenutzt werden. Die Rekursion hat zugleich den Vorteil, dass in einem Durchgang auch die Schätzungen für alle AR-Modelle mit einer kleineren Ordnung bestimmt werden. Dies ist für die weiter unten zu besprechende Identifikation von Bedeutung.

Für die Beispielsituation werden die empirischen Autokovarianzen $c_{|i-j|}$ anstelle der standardisierten Summen $\frac{1}{N} \sum_{t=3}^N y_{t-i}y_{t-j}$ in den Gleichungen (3.11a), (3.11b) verwendet. Die Schätzungen ergeben sich damit aus

$$\hat{\alpha}_1 c_0 + \hat{\alpha}_2 c_1 = c_1 \quad (3.13a)$$

$$\hat{\alpha}_1 c_1 + \hat{\alpha}_2 c_0 = c_2 . \quad (3.13b)$$

Burg

Um die Parameter eines AR-Modells ohne den Umweg über die ACF zu ermitteln, kann man neben den Residuen der üblichen Vorwärtsrekursion $\hat{\epsilon}_t = y_t - \alpha_1 y_{t-1} - \dots - \alpha_p y_{t-p}$ auch die aus der rückwärts laufenden Rekursion betrachten: $\hat{\eta}_t = y_t - \alpha_1 y_{t+1} - \dots - \alpha_p y_{t+p}$. Die Minimierung der Summe der quadrierten Residuen aus beiden Beziehungen führt dann zu sehr effizient berechenbaren Schätzungen.

Hier wird also die Summe der quadrierten Vorwärts- und Rückwärtsresiduen minimiert:

$$\sum_{t=3}^N (y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2})^2 + \sum_{t=1}^{N-2} (y_t - \alpha_1 y_{t+1} - \alpha_2 y_{t+2})^2 \stackrel{!}{=} \min . \quad (3.14)$$

Den Hintergrund für diesen Ansatz liefert die Tatsache, dass ein vorwärts laufender stationärer AR[p]-Prozess die gleiche ACF hat wie der zugehörige rückwärts laufende $Y_t - \alpha_1 Y_{t+1} - \dots - \alpha_p Y_{t+p} = \eta_t$. Die Schätzung von AR-Modellen mit dem Burg-Algorithmus wurde zuerst in Burg (1975) beschrieben. Eine weitere Textstelle dazu ist Ulrich & Ooe (1979).

Die knappe Vorstellung der einzelnen Schätzverfahren zeigt, dass sie sich bis auf die Burg-Methode vor allem durch die Behandlung der Werte an den Rändern des Beobachtungssintervalls unterscheiden. Bei längeren Reihen werden die Unterschiede immer weniger relevant. Bei kürzeren Zeitreihen liefert i. d. R. die ML-Schätzung die zufriedenstellenderen Ergebnisse. Nur die Yule-Walker-Methode liefert stets Schätzwerte, die zu stationären Prozessen gehören. Bei ihrer Anwendung sollte die Reihe vorher einer Taper-Modifikation unterzogen werden. Diese wird im Abschnitt über Spektralschätzung erklärt.

Aus der folgenden asymptotischen Verteilungsaussage für die oben angegebenen Schätzer kann man erkennen, dass bei genügend langen Reihen die Schätzer hinreichend genau mit den tatsächlichen Parameterwerten übereinstimmen.

Satz 3.4 (Asymptotische Verteilung der Koeffizientenschätzer)

Für einen stationären AR[p]-Prozess $(Y_t - \mu) = \alpha_1(Y_{t-1} - \mu) + \dots + \alpha_p(Y_{t-p} - \mu) + \varepsilon_t$ mit $E(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$ gilt für die nach einer der oben angegebenen Methoden bestimmten Schätzer:

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

sind gemeinsam asymptotisch normalverteilt mit Erwartungswerten null und Kovarianzmatrix $\sigma^2 \cdot \Sigma_p^{-1}$, wobei $\Sigma_p = [\gamma_{|i-j|}]$ die Kovarianzmatrix von p aufeinander folgenden Variablen des Prozesses (Y_t) ist.

Traut man der Asymptotik bei einer kürzeren Zeitreihe nicht und möchte man trotzdem eine Schätzung der Standardfehler der Parameterschätzer haben, so kann man das Bootstrap-Verfahren einsetzen.

Die *Bootstrap-Methode* für AR-Modelle geht davon aus, dass die beobachtete Zeitreihe y_1, \dots, y_N ein Ausschnitt aus einer Realisation eines AR-Prozesses $Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t$ ist. Unter Verwendung des geschätzten Modells und den sich daraus ergebenden Residuen $\hat{\varepsilon}_t$ wird eine größere Anzahl von Reihen gleicher Länge N simuliert; für alle simulierten Reihen werden die Parameter jeweils neu geschätzt. Für jeden Parameterwert lässt sich die Standardabweichung aus der Menge der erhaltenen Schätzwerte bestimmen. Da die simulierten Reihen angenähert als Realisationen des gleichen Modells wie die Ausgangsreihe angesehen werden können, gibt dies unter milden Annahmen eine bessere Schätzung des tatsächlichen Standardfehlers als die auf der asymptotischen Normalverteilung beruhende. Auch Konfidenzintervalle können darüber ermittelt und Tests durchgeführt werden.

Levinson-Durbin-Rekursion und partielle Autokorrelation

Der Vorteil der Yule-Walker-Schätzer liegt darin, dass sich mit geringem Aufwand AR-Modelle verschiedener Ordnungen schätzen lassen. Diese Aufgabe fällt im Rahmen der Spezifikation an und wird im folgenden Abschnitt ausgeführt.

Die Grundlage für diese Berechnung stellt die *Levinson-Durbin-Rekursion* dar. Damit wird die Möglichkeit gegeben, aus den Schätzungen der Parameter eines AR[p]-Prozesses die Parameterschätzungen des AR-Prozesses der Ordnung $p + 1$ zu bestimmen.

Bevor die Rekursion in einem Satz allgemein formuliert wird, soll sie für den Übergang eines AR[2]- auf ein AR[3]-Modell skizziert werden. Zur Vereinfachung der Darstellung werden die geschätzten Koeffizienten des AR[2]-Modells mit $\hat{\alpha}_1, \hat{\alpha}_2$ bezeichnet und die des AR[3]-Modells mit $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$.

Als Yule-Walker-Schätzer erfüllen die Koeffizienten die Beziehungen

$$\begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \quad (3.15)$$

$$\begin{pmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \quad (3.16)$$

Die Gleichung (3.15) wird „auf den Kopf gestellt“ und die Gleichung (3.16) „zerlegt“:

$$\begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_2 \\ \hat{\alpha}_1 \end{pmatrix} = \begin{pmatrix} r_2 \\ r_1 \end{pmatrix} \quad (3.17)$$

$$\begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} + \begin{pmatrix} r_2 \\ r_1 \end{pmatrix} \hat{\beta}_3 = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \quad (3.18)$$

$$r_2 \hat{\beta}_1 + r_1 \hat{\beta}_2 + \hat{\beta}_3 = r_3 \quad (3.19)$$

Nun werden die Beziehungen (3.15) und (3.17) in die Gleichung (3.18) eingesetzt:

$$\begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} + \begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_2 \\ \hat{\alpha}_1 \end{pmatrix} \hat{\beta}_3 = \begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} \quad (3.20)$$

Multiplizieren von (3.20) mit $\begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix}^{-1}$ und Umstellen führt zu:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} - \begin{pmatrix} \hat{\alpha}_2 \\ \hat{\alpha}_1 \end{pmatrix} \hat{\beta}_3 \quad (3.21)$$

Nun wird die Gleichung (3.19) nach $\hat{\beta}_3$ aufgelöst; dann wird (3.21) eingesetzt:

$$\hat{\beta}_3 = r_3 - (r_2, r_1) \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \Rightarrow \hat{\beta}_3 = r_3 - (r_2, r_1) \left[\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} - \begin{pmatrix} \hat{\alpha}_2 \\ \hat{\alpha}_1 \end{pmatrix} \hat{\beta}_3 \right]. \quad (3.22)$$

Dies ergibt

$$\hat{\beta}_3 = \frac{r_3 - r_2 \cdot \hat{\alpha}_1 - r_1 \cdot \hat{\alpha}_2}{1 - r_2 \cdot \hat{\alpha}_2 - r_1 \cdot \hat{\alpha}_1}. \quad (3.23)$$

Die beiden Gleichungen (3.21) und (3.23) geben die Rekursion von einem AR[2] zu einem AR[3]-Modell.

Satz 3.5 (Levinson-Durbin-Rekursion)

Mit den Startwerten $\hat{\alpha}_1 = r_1$, $Q(1) = 1 - r_1^2$ erhält man aus den Lösungen der Yule-Walker-Gleichungen der Ordnung $p - 1$ unter Verwendung von $\Delta_p = r_p - (\hat{\alpha}_1(p - 1)r_{p-1} - \dots - \hat{\alpha}_{p-1}(p - 1)r_1)$ und $Q(p - 1) = 1 - \hat{\alpha}_1(p - 1)r_1 - \dots - \hat{\alpha}_{p-1}(p - 1)r_{p-1}$:

$$\hat{\alpha}_p(p) = \frac{\Delta_p}{Q(p - 1)} \quad (3.24a)$$

$$\hat{\alpha}_i(p) = \hat{\alpha}_i(p - 1) - \hat{\alpha}_p(p - 1)\hat{\alpha}_{p-i}(p - 1) \quad (i = 1, \dots, p - 1) \quad (3.24b)$$

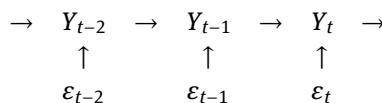
$$Q(p) = Q(p - 1) \cdot (1 - \hat{\alpha}_p^2(p)). \quad (3.24c)$$

Die Levinson-Durbin-Rekursion hat einen weiteren Vorteil. Die Koeffizienten $\hat{\alpha}_p(p)$, die sich bei der rekursiven Bestimmung der Yule-Walker-Schätzungen ergeben, sind gerade die empirischen partiellen Autokorrelationen. Die *partielle Autokorrelationsfunktion*, kurz *PACF* ist eine weitere Kennfunktion stationärer Prozesse, die große Bedeutung bei der Identifikation von linearen Zeitreihenmodellen hat.

Die PACF wird zunächst beispielhaft für einen AR[1]-Prozess eingeführt.

Beispiel 3.6 (Partielle Autokorrelationen bei einem AR[1]-Prozess)

Sei $Y_t = \alpha Y_{t-1} + \varepsilon_t$. Der Wert der unmittelbaren Vergangenheit, also Y_{t-1} , wird fest gehalten und nach dem Zusammenhang, d. h. der Korrelation, von Y_t und Y_{t-2} gefragt. Das folgende Schema legt nahe, dass es dann keinen Zusammenhang mehr gibt.



Formal wird das Festhalten oder Ausschalten von Y_{t-1} so erfasst, dass Y_{t-1} in einer Form von Y_t und von Y_{t-2} abgezogen wird, die den jeweils stärksten Zusammenhang von Y_{t-1} mit den beiden Variablen widerspiegelt. Für den AR[1]-Prozess erhält man:

$$\begin{aligned} E(Y_t - \gamma Y_{t-1})^2 &= E(\alpha Y_{t-1} + \varepsilon_t - \gamma Y_{t-1})^2 \stackrel{!}{=} \min \quad \Rightarrow \quad \gamma = \alpha \\ E(Y_{t-2} - \delta Y_{t-1})^2 &= E(Y_{t-2} - \delta(\alpha Y_{t-1} + \varepsilon_{t-1}))^2 \stackrel{!}{=} \min \quad \Rightarrow \quad \delta = \frac{1}{\alpha}. \end{aligned}$$

Dann gilt für die beiden vom Einfluss von Y_{t-1} bereinigten Variablen $\tilde{Y}_{t-2} = Y_{t-2} - \alpha^{-1} Y_{t-1}$ und $\tilde{Y}_t = Y_t - \alpha Y_{t-1}$: $\text{Cov}(\tilde{Y}_{t-2}, \tilde{Y}_t) = 0$.

Das bleibt auch für $\tau > 2$ richtig: Beim AR[1]-Prozess ist die Korrelation von Y_t und $Y_{t-\tau}$ unter Festhalten von $Y_{t-1}, \dots, Y_{t-\tau+1}$ null.

Definition 3.7 (Partielle Autokorrelationsfunktion)

Die *partielle Autokorrelationsfunktion* (π_τ), kurz *PACF*, eines stationären Prozesses (Y_t) ist die Folge der partiellen Korrelationen von Y_t und $Y_{t-\tau}$ unter Festhalten der Werte von $Y_{t-1}, \dots, Y_{t-\tau+1}$. Für $\tau = 0$ wird $\pi_0 = 1$ und für $\tau = 1$ wird $\pi_1 = \rho_1$ gesetzt.

Wie oben angegeben wird die PACF über den Einsatz der Levinson-Durbin-Rekursion geschätzt. Da die Levinson-Durbin-Rekursion auf der ACF allein aufbaut, ergibt sich speziell, dass die PACF ebenfalls nur von den Autokorrelationen abhängt.

Satz 3.8 (Eigenschaften der PACF)

Für die PACF (π_τ) eines AR[p]-Prozesses und für ihr empirisches Gegenstück ($\hat{\pi}_\tau$) gilt

- $\pi_\tau = 0 \quad \text{für } |\tau| > p.$
- $\hat{\pi}_\tau \sim \mathcal{N}(0, \frac{1}{N}) \quad \text{für } |\tau| > p.$

Spezifikation von AR-Modellen

Die Auswahl eines geeigneten AR-Modells bedeutet, dass die Ordnung p bestimmt werden muss. Dies wird als *Spezifikation* oder *Identifikation eines AR-Modells* bezeichnet. Dazu zieht man im ersten Schritt die ACF heran.

Wenn ein AR[p]-Modell ‚wahr‘ ist, so sind bei der Zeitreihe und der empirischen ACF gewisse Verhaltensmuster zu erwarten. Für AR[1]-Prozesse $Y_t = \alpha Y_{t-1} + \varepsilon_t$ ist etwa $\rho_\tau = \alpha^{|\tau|}$. Die ACF ist exponentiell abklingend. Da für genügend lange Reihen $\hat{\rho}_\tau \approx \rho_\tau$ gilt, sollte die empirische ACF in der gleichen Weise gegen null gehen wie die theoretische.

Für einen stationären AR[p]-Prozess ist das folgende Erscheinungsbild der empirischen ACF charakteristisch:

- „exponentielles Abklingen“ nach dem Lag p .
- näherungsweise „Erfüllung“ der Rekursion $\hat{\rho}_\tau = \hat{\alpha}_1 \hat{\rho}_{\tau-1} + \dots + \hat{\alpha}_p \hat{\rho}_{\tau-p}$, ($\tau > 0$).

Diese Muster in empirischen Reihen wieder zu erkennen ist sicherlich nicht ganz einfach. Daher wird in aller Regel noch die PACF als zweite Kennfunktion betrachtet. Mit der Aussage des Satzes 3.8 ist nämlich die Möglichkeit für einen Modell-Check gegeben: Man zeichnet für die ersten Lags τ die empirischen Größen $\hat{\rho}_\tau$ mit den Grenzen $\pm 2/\sqrt{N}$. Die doppelte Standardabweichung entspricht wegen der Normalverteilung grob einem 95 % Schwankungsintervall. Sobald die Werte von $\hat{\rho}_\tau$ innerhalb der Grenzen verbleiben, kann davon ausgegangen werden, dass das AR-Modell keine von null verschiedenen Koeffizienten mehr besitzt. Somit lässt sich die Ordnung anhand des Verhaltens der empirischen PACF einschätzen.

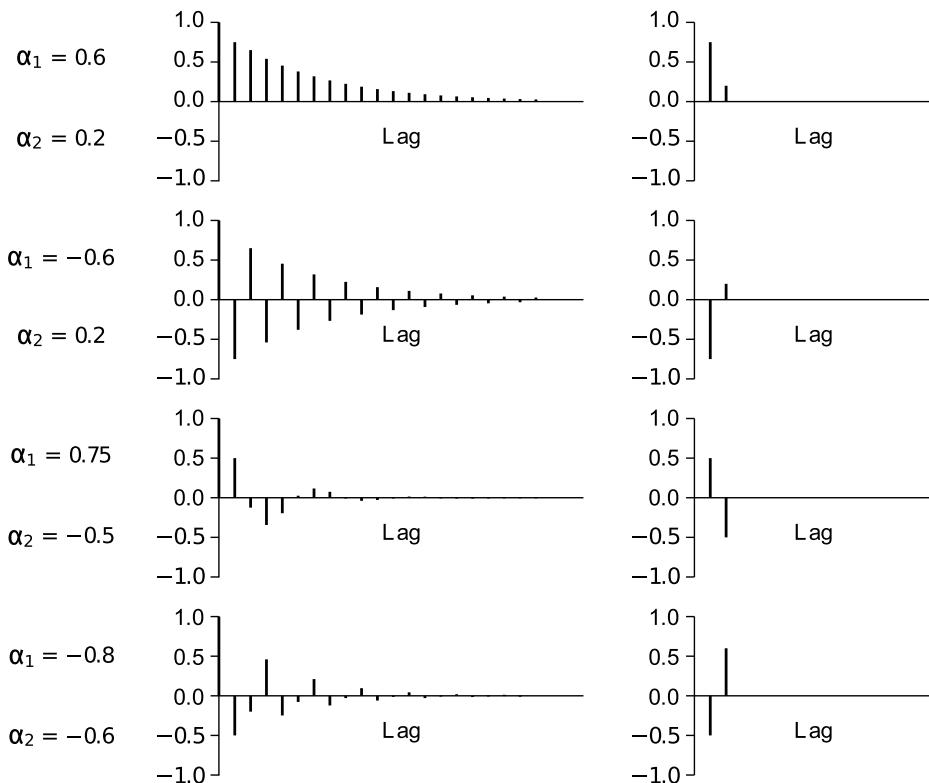


Abb. 3.3: ACF und PACF für AR[2]-Prozesse

Als zweite Möglichkeit der Spezifikation von AR[p]-Modellen bietet sich das sukzessive Anpassen von AR[1]-, AR[2]-, ... Modellen an. Die Auswahl der Ordnung geschieht dann mittels eines Informationskriteriums. Gewählt wird die Ordnung, für die das eingesetzte Kriterium am kleinsten ist.

Das *AIC* von Akaike war eines der zuerst eingeführten Kriterien zur Bestimmung der Ordnung eines linearen Prozessmodells. Es ist informationstheoretisch begründet. Für die Ableitung des AIC hat Akaike vorausgesetzt, dass normalverteilte Daten vorliegen und es mit der Loglikelihoodfunktion ausgedrückt. Im allgemeinen Fall ist es, wenn p die Anzahl der geschätzten Parameter und L der maximierte Wert der Likelihoodfunktion für das geschätzte Modell sind:

$$\text{AIC} = -2 \cdot \ln(L) + 2 \cdot p . \quad (3.25)$$

In der Praxis wird das Kriterium aber auch bei anderen Schätzern angewendet. Dabei wird meist ausgenutzt, dass bei der Normalverteilung die Likelihood im Wesentlichen nur noch von der Maximum-Likelihood-Schätzung $\hat{\sigma}_p^2$ der Varianz der Residuen abhängt. Dann lautet es in vereinfachter, leicht modifizierter Form, wobei RSS die Summe der quadrierten zentrierten Residuen ist:

$$\text{AIC}(p) = \ln\left(\frac{\text{RSS}_p}{N}\right) + \frac{2p}{N} . \quad (3.26)$$

In dieser Form hat das AIC eine einleuchtende Interpretation. Einmal geht die logarithmierte Schätzung der Varianz des Innovationsprozesses ein. Diese wird bei wachsender Ordnung immer kleiner. Das würde dazu führen, ein möglichst großes Modell zu wählen. Daher wird als Zweites die Anzahl der vorkommenden Parameter berücksichtigt. Dies wirkt als Strafterm, wächst er doch mit der Ordnung. Typischer Weise wirkt sich die Verbindung so aus, dass zunächst das AIC mit größer werdendem p sinkt. Die Reduzierung der Residualvarianz ist anfangs stark, dann wird sie geringer und der Strafterm wirkt sich stärker aus. Die Minimalstelle ist ein „optimaler“ Kompromiss. Übrigens ist im Regressionskontext das C_p -Kriterium äquivalent zu dem AIC.

Da sich bei einer genauen Analyse der theoretischen Eigenschaften herausstellte, dass das Kriterium eher zu Modellen mit zu vielen Parametern führt, wurden verschiedene Varianten vorgeschlagen. Sie unterscheiden sich durch den Strafterm:

Die Bias-korrigierte Version des AIC ist das AICC von Hurvich und Tsay:

$$\text{AICC}(p) = \ln\left(\frac{\text{RSS}_p}{N}\right) + \frac{2p}{N-p-2} ; \quad (3.27)$$

das BIC, das Bayessche Informationskriterium von Schwarz (auch als SBC bezeichnet):

$$\text{BIC}(p) = \ln\left(\frac{\text{RSS}_p}{N}\right) + \frac{p \ln N}{N} ; \quad (3.28)$$

sowie das Kriterium von Hannan-Quinn, bei dem $c > 1$ eine beliebig gewählte Konstante ist:

$$\text{HQ}(p) = \ln\left(\frac{\text{RSS}_p}{N}\right) + \frac{2p \cdot c \cdot \ln(\ln N)}{N} \quad (c > 1). \quad (3.29)$$

Beispiel 3.9 (Herzfrequenz – Fortsetzung von Seite 36)

Werden der Reihe HERZSCHLAG AR-Modelle verschiedener Ordnungen unter Verwendung der CLS-Schätzmethode angepasst, so erhält man:

p	1	2	3	4
s2	0.465	0.284	0.263	0.257
AIC	-0.698	-1.179	-1.172	-1.129
BIC	-0.652	-1.085	-1.032	-0.943

In der ersten Zeile sind die jeweiligen Ordnungen angegeben; in der zweiten stehen die Schätzungen der Varianz des Innovationsprozesses. Hier wird also von beiden Kriterien ein AR[2]-Modell ausgewählt. Diese Übereinstimmung ist nicht zwangsläufig. Das ist andeutungsweise auch daran zu erkennen, dass die Werte des AIC für $p = 2$ und $p = 3$ fast identisch sind.

R-Code:

```

1 data(HEARTBEAT)
2 y <- HEARTBEAT
3 n <- length(y)
4 RSS <- rep(0,4)
5 AIC <- rep(0,4)
6 BIC <- rep(0,4)
7 for (p in c(1:4)){
8   out <- ar(y,order.max=p,aic=FALSE,method="ols")
9   res <- out$resid[-c(1:p)]
10  RSS[p] <- sum((res-mean(res))^2)
11  AIC[p] <- log(RSS[p]/n) + 2*p/n
12  BIC[p] <- log(RSS[p]/n) + log(n)*p/n
13 }
14 s2 <- RSS/n
15 round(rbind(c(1:4),s2,AIC,BIC),3)

```

Mit der Funktion `ar` kann ein AR-Modell angepasst werden. Bei `aic=TRUE` wird das Modell mit dem kleinsten Wert des AIC ausgegeben. Allerdings werden nicht die AIC-Werte selbst gespeichert, sondern die Differenzen zu dem Minimum. Die ersten p Residuen sind bei der `ols`-Methode nicht vorhanden. Die fehlenden Werte (NA's) müssen daher aus dem Vektor der Residuen eliminiert werden.

Die automatische Auswahl der Ordnung mittels eines Informationskriteriums ist durchaus problematisch. Bei diesen Kriterien wird nämlich zunächst nicht berücksichtigt, dass einzelne Koeffizienten vernachlässigbar klein sein können. Als Konsequenz wird meist eine zu geringe Ordnung gewählt, da ja die irrelevanten Parameter mit in den Strafterm eingehen. Dann macht es Sinn, *Subsetmodelle* anzupassen. Manchmal kann man schon aus sachlichen Gegebenheiten unterstellen, dass gewisse Lags irrelevant sind. Bisweilen lässt sich dies auch aus der Darstellung von ACF und PACF ersehen. Ein formales Mittel ist wie bei der linearen Regression die Betrachtung der t -Statistiken

$$T_u = \frac{\hat{\alpha}_u}{\hat{\sigma}_{\hat{\alpha}_u}}, \quad (3.30)$$

die unter der Nullhypothese $H_0: \alpha_u = 0$ approximativ $\mathcal{N}(0, 1)$ verteilt sind. Hier werden die P -Werte der einzelnen Koeffizienten separat beurteilt. Die abschließende Überprüfung der Modelleignung erlaubt es zu vernachlässigen, dass es sich eigentlich um ein multiples Testproblem handelt.

Wird man beim Schätzen von AR-Modellen auf Subsetmodelle geführt, so kann man diese mittels der Informationskriterien untereinander und mit den Modellen mit voller Parameterzahl vergleichen. Dazu wird bei den Kriterien die Ordnung p durch die Anzahl der geschätzten Parameter ersetzt. Theoretisch kann man auch von vornherein alle Subsetmodelle an eine Reihe anpassen und das Modell auswählen, bei dem das BIC-Kriterium (oder ein anderes) minimal wird. Jedoch ist der Rechenaufwand bei größeren Ordnungen sehr hoch.

Der – vorläufig – letzte Schritt bei der Anpassung eines AR-Modells besteht in der Überprüfung der Anpassungsgüte. Dazu untersucht man vor allem die Residuen. Wenn das Modell geeignet ist, sollten sich die Residuen verhalten wie eine Realisation eines White-Noise-Prozesses. Dies stimmt zwar wegen der Verwendung der geschätzten Parameter auch dann nicht exakt, wenn der Prozess richtig spezifiziert wurde, jedoch sollten bei nicht zu kurzen Reihen die Unterschiede vernachlässigbar sein.

Es stehen folgende Möglichkeiten zur Überprüfung der Residuen zur Verfügung:

1. Als grafischen Check der Residuen auf White-Noise wird man sich die ACF und die PACF der Residuen ansehen. Sind noch deutlich von null verschiedene Korrelationen bzw. partielle Korrelationen in den Residuen zu sehen, so sind weitere Koeffizienten in das Modell aufzunehmen.
2. Ein Test zur Überprüfung der Unkorreliertheit der Residuen, ist der *Ljung-Box-Pierce-Test*. Er kann als Test auf die Güte der Modellanpassung angesehen werden und wird dementsprechend auch als *Portmanteau-Test* bezeichnet. Er basiert auf der Prüfgröße

$$Q = N(N + 2) \sum_{\tau=1}^k \frac{\tilde{r}_\tau^2}{N - \tau}. \quad (3.31)$$

Dabei ist (\tilde{r}_τ) die ACF der Residualreihe. Falls das Modell korrekt ist, hat Q asymptotisch eine χ^2 -Verteilung mit $k-p$ Freiheitsgraden (df), $Q \sim \chi^2_{k-p}$.

Die Wahl der Anzahl der Summanden k ist nicht offensichtlich. k sollte jedenfalls größer als die höchste berücksichtigte Modellordnung sein. Als praktisch hat sich erwiesen, die Teststatistik für verschiedene k zu bestimmen, etwa für $k = 6, 12, 18, 24, \dots$. Dies erlaubt zu erkennen, ob tatsächlich alle Residualkorrelationen genügend klein sind.

3. Check der AR-Residuen auf Normalverteilung. Dies sollte insbesondere bei Verwendung der ML-Methode durchgeführt werden, da diese ja auf der Normalverteilungsannahme basiert. Abweichungen von der Normalverteilung deuten entweder auf ein schlecht passendes Modell hin oder auf die Notwendigkeit, die Ausgangsreihe nichtlinear zu transformieren.

Beispiel 3.10 (Diabetes Mellitus)

Bei der Inzidenz von Insulin abhängiger Diabetes Mellitus bei Kindern wird von verschiedenen Autoren eine Saisonschwankung gesehen, vgl. Waldhoer, Schober & Tuomilehto (1997). Die Abbildung 3.4 zeigt die Zeitreihe DIABETES der monatlichen Inzidenzen (Anzahl Neuerkrankte pro 100 000) für Kinder und Jugendliche bis 17 Jahren in Österreich. Da auch die Realisationen von AR-Prozessen recht periodisches Verhalten aufweisen können, wird der Reihe ein AR-Modell angepasst.

Dass dieser Ansatz sinnvoll ist, zeigen die in Abbildung 3.5 dargestellten ACF und PACF. Letztere weist noch darauf hin, dass ein AR[3]-Prozess ausreichend sein sollte (Modell 1). Da der P -Wert der t -Statistik für α_2 weit über 0.05 beträgt, wird im zweiten Schritt ein Subsetmodell angepasst (Modell 2). Dann wird der P -Wert des Koeffizienten α_3 relativ groß, so dass versuchsweise noch ein AR[1]-Prozess angepasst wird (Modell 3).

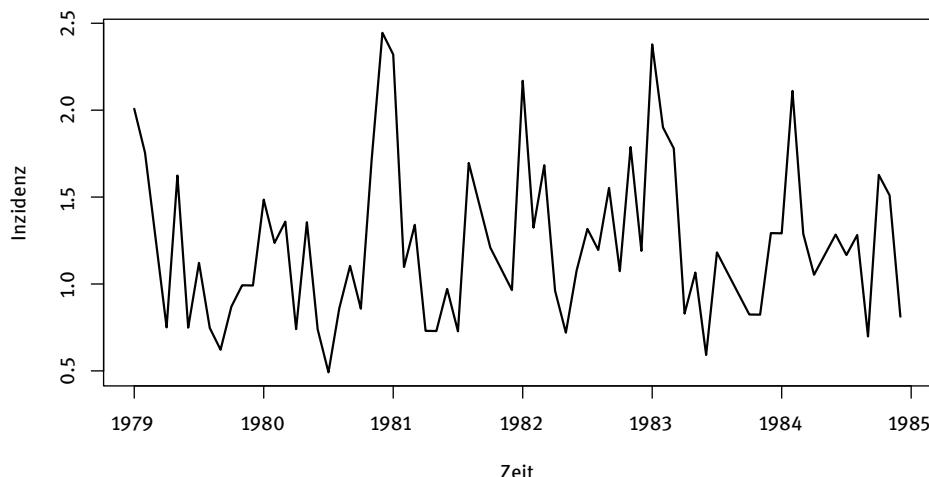


Abb. 3.4: Monatliche Inzidenzen von Diabetes Mellitus

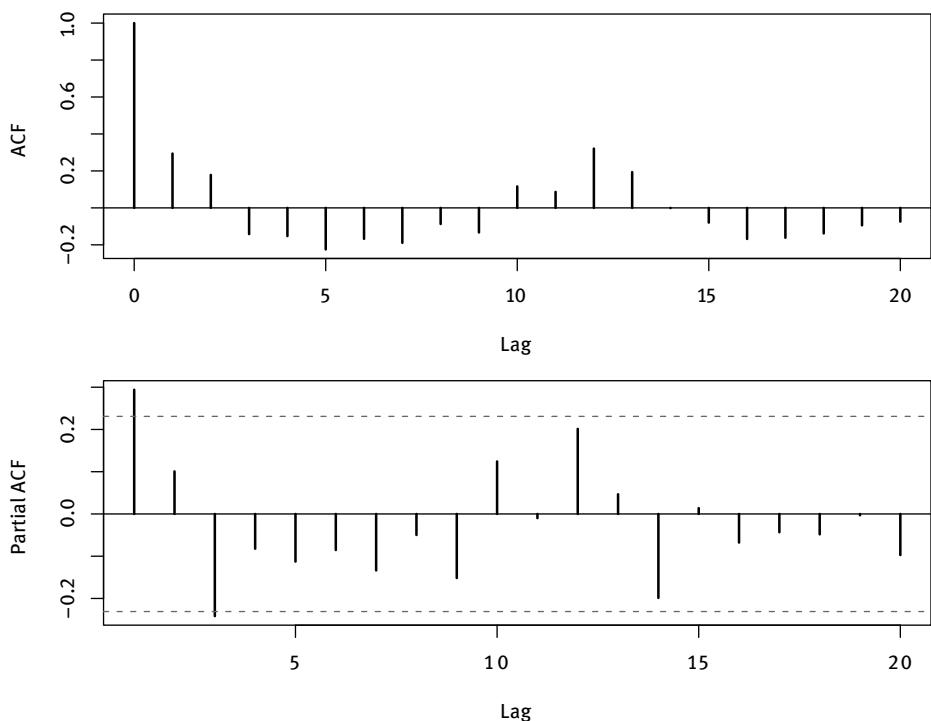


Abb. 3.5: ACF und PACF der Reihe DIABETES

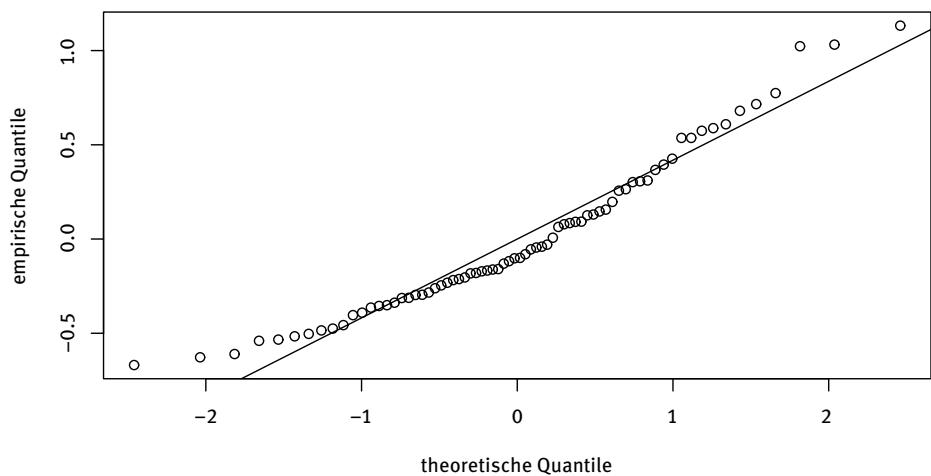


Abb. 3.6: QQ-Diagramm für die Residuen von DIABETES (Modell 2)

Die Ergebnisse für die vorab zentrierte Reihe sind:

Parameter	Modell 1		Modell 2		Modell 3	
	Schätz-wert	Standard-Fehler	Schätz-wert	Standard-Fehler	Schätz-wert	Standard-Fehler
α_1	0.3086	0.1155	0.3496	0.1139	0.3064	0.1145
α_2	0.1742	0.1192				
α_3	-0.2612	0.1162	-0.2184	0.1139		
AIC/BIC:	84.055/88.885		84.15398/86.707		85.708/85.985	

Die Signifikanz der Koeffizienten lässt sich einfach durch Bilden des Quotienten „Schätzwert dividiert durch Standardfehler“ abschätzen. Ist er größer als zwei, so ist er bedeutsam. Dies ergibt die oben angegebene Modellsequenz.

Das AIC wählt das volle AR[3]-Modell (Modell 1), das BIC das AR[1]-Modell (Modell 3) aus. Der Ljung-Box-Pierce-Test für die Residuen ist für die Modelle:

m	Modell 1			Modell 2			Modell 3		
	Test	df	P-Wert	Test	df	P-Wert	Test	df	P-Wert
6	1.68	3	0.6414	3.92	4	0.4165	7.37	5	0.1948
12	10.83	9	0.2876	15.95	10	0.1012	21.24	11	0.0310
18	15.91	15	0.3883	20.67	16	0.1914	25.60	17	0.0821

Beim AR[3]-Modell (Modell 1) sind die P-Werte am größten. Aber auch beim Subsetmodell (Modell 2) sind sie zufriedenstellend. Nur beim AR[1]-Modell verbleiben noch Korrelationen in der Residualreihe. Folgt man dem Prinzip der Sparsamkeit, erscheint das Subset-AR[3]-Modell das geeignetste der drei.

Schließlich wird das im Abschnitt 2.3 vorgestellte QQ-Diagramm für die Residuen gezeichnet, um zu sehen, ob sie hinreichend normalverteilt sind. Das Diagramm weist auf eine leichte Rechtsschiefe ihrer Verteilung hin. Diese ist aber nicht so stark, dass eine Transformation der Ausgangsdaten notwendig wäre.

R-Code:

```

1 data(DIABETES)
2 DiMe <- DIABETES
3 plot(DiMe)
4 y <- ts(DiMe,frequency=1)
5 y <- y-mean(y)
6 n <- length(y)

```

```

7  par(mfrow=c(2,1))
8  acf(y,lag.max=20,ci=FALSE)
9  acf(y,lag.max=20,type="partial",ci=FALSE)
10 par(mfrow=c(1,1))
11 m1 <- arima(y,order=c(3,0,0),include.mean=FALSE)
12 m2 <- arima(y,order=c(3,0,0),include.mean=FALSE,fixed=c(NA,0,NA))
13 m3 <- arima(y,order=c(1,0,0),include.mean=FALSE)
14 -2*m1$loglik+log(n)*3
15 -2*m2$loglik+log(n)*2
16 -2*m3$loglik+log(n)*1
17 LjungBoxPierceTest(m1$residuals,n.par=3,maxlag=18)
18 LjungBoxPierceTest(m2$residuals,n.par=2,maxlag=18)
19 LjungBoxPierceTest(m3$residuals,n.par=1,maxlag=18)
20 qqnorm(m2$resid)
21 abline(0,sd(m2$resid))

```

Für die ACF und die PACF wird die Zeitreihe ihres Zeitreihenattributes beraubt, da die Skalierung der Abszisse sonst in 1/12 Schritten vorgenommen wird. Um beide Funktionen zugleich zu sehen, wird das Grafikfenster geteilt. Dies wird anschließend wieder zurückgesetzt. Mit der Funktion `ar` kann kein Subset-AR-Modell angepasst werden. Daher wird die mächtigere Funktion `arima` verwendet. Der erste der drei in `order` einzutragenden Werte gibt die Ordnung des AR-Modells an. Mit dem optionalen Argument `fixed` werden nur die Parameter geschätzt, die einen NA-Eintrag haben. Gemäß der Voreinstellung wird ein Niveau geschätzt; dies ist ggf. bei der Angabe der festgelegten Parameter zu berücksichtigen. Hier wird dies ausgeschaltet; die Zeitreihe wird ja vorab zentriert. Die BIC-Werte sind nicht direkt vorhanden. Da die ML-Schätzung den voreingestellten Standard bildet, stehen die Werte der Loglikelihood zur Verfügung. Somit können die BIC-Werte unter Rückgriff auf die allgemeine Definition des BIC-Kriteriums entsprechend den Zeilen 14 bis 16 berechnet werden.

Das im Beispiel genannte Prinzip der *Sparsamkeit* verlangt, dass man Modellen mit möglichst wenig Parametern den Vorzug geben sollte. Verwendet man ein Modell mit zu vielen Parametern, so kann es leicht geschehen, dass dieses Modell nicht mehr adäquat ist, wenn die Datenbasis sich etwas ändert, z. B. wenn neue Beobachtungen hinzukommen. Es ist dann überangepasst an die zur Anpassung verwendeten Daten.

Wenn die abschließende Überprüfung der Anpassung eines AR-Modells an eine stationäre Zeitreihe dazu führt, dass ein solches Modell nicht adäquat ist, so ist ein Modell aus einer anderen Modellklasse anzusetzen. Eine der Möglichkeiten bilden die MA-Modelle, die als nächstes vorgestellt werden.

3.2 MA-Modelle

Definition und grundlegende Eigenschaften

Zu einer andern Modellklasse als die der AR-Modelle gelangt man, wenn in der Prozessgleichung für Y_t nicht die beobachteten Werte, wohl aber vergangene Störungen explizit auftauchen dürfen. Das ergibt die *Moving Average-Prozesse*. Ein $MA[q]$ -Prozess (Y_t) ist definiert durch die Beziehung

$$Y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q}. \quad (3.32)$$

Dabei ist (ε_t) ein White-Noise-Prozess. Das Bildungsgesetz eines $MA[q]$ -Prozesses ist einfacher als das eines AR-Prozesses. Es wird hier ja nur eine gewichtete Summe von unabhängigen Zufallsvariablen gebildet. Das hat zur Konsequenz, dass endliche $MA[q]$ -Prozesse stets stationär sind. $MA[\infty]$ -Prozesse sind stationär, wenn die Folge der Gewichte absolut summierbar ist, $\sum |\beta_u| < \infty$.

Die Autokovarianzfunktion eines $MA[q]$ -Prozesses erhält man mit der Unkorreliertheit der ε_t und mit $\beta_0 = -1$ zu

$$\gamma_\tau = \begin{cases} \sigma_\varepsilon^2 \cdot [\beta_0 \beta_\tau + \beta_1 \beta_{\tau+1} + \cdots + \beta_q \beta_{q-\tau}] & \text{für } 0 \leq \tau \leq q \\ 0 & \text{für } \tau > q, \\ \gamma_{-\tau} & \text{für } \tau < 0. \end{cases} \quad (3.33)$$

Speziell ist also $\gamma_0 = \sigma_\varepsilon^2 \cdot [1 + \beta_1^2 + \cdots + \beta_q^2]$. Zudem ist wesentlich, dass $\gamma_\tau = 0$ für $\tau > q$ gilt.

Für die Theorie ist die Einfachheit der Überprüfung der Stationarität von $MA[\infty]$ -Prozessen unter dem Gesichtspunkt wesentlich, dass man nur die Äquivalenz eines Prozesses mit einem entsprechenden $MA[\infty]$ -Prozess zu zeigen braucht, um seine Stationarität nachzuweisen. Bei $AR[p]$ -Prozessen sichert die Stationaritätsbedingung, dass die Wurzeln der charakteristischen Gleichung außerhalb des Einheitskreises liegen, gerade die Möglichkeit der Darstellung als $MA[\infty]$ -Prozess.

Beispiel 3.11 ($MA[\infty]$ -Darstellung eines $AR[1]$ -Prozesses)

Sei $Y_t = \alpha Y_{t-1} + \varepsilon_t$ ein $AR[1]$ -Prozess mit $|\alpha| < 1$. Dann führt das rekursive Einsetzen der Definitionsgleichung zu

$$\begin{aligned} Y_t &= \alpha Y_{t-1} + \varepsilon_t = \alpha^2 Y_{t-2} + \alpha \varepsilon_{t-1} + \varepsilon_t = \alpha^3 Y_{t-3} + \alpha^2 \varepsilon_{t-2} + \alpha \varepsilon_{t-1} + \varepsilon_t \\ &= \cdots = \sum_{u=1}^{\infty} \alpha^u \varepsilon_{t-u} + \varepsilon_t. \end{aligned} \quad (3.34)$$

Wegen $|\alpha| < 1$ ist die Folge (α^u) absolut summierbar.

Umgekehrt kann man sich natürlich auch fragen, wann sich MA- in AR-Prozesse überführen lassen. Diese Frage ist unter Schätz- und Prognosegesichtspunkten ganz wesentlich. Denn für beide Aufgaben stehen ja die beobachteten Zeitreihenwerte zur

Verfügung; mit diesen müssen die nicht-beobachtbaren Residuen geschätzt werden. Anhand des folgenden Beispiels wird deutlich, dass hierfür entsprechend geeignete Bedingungen an die Koeffizienten zu stellen sind.

Beispiel 3.12 (AR[∞]-Darstellung eines MA[1]-Prozesses)

Für einen MA[1]-Prozess erhält man auf die gleiche Weise wie im vorangegangenen Beispiel, dass sich aus $Y_t = \varepsilon_t - \beta\varepsilon_{t-1}$ ergibt:

$$\varepsilon_t = Y_t + \beta\varepsilon_{t-1} = Y_t + \beta(Y_{t-1} + \beta\varepsilon_{t-2}) = \dots = Y_t + \sum_{u=1}^{\infty} \beta^u Y_{t-u}. \quad (3.35)$$

Damit die letzte Summe tatsächlich einen stationären Prozess darstellen kann, muss offensichtlich $|\beta| < 1$ gelten.

Die Bedingung $|\beta| < 1$ für MA[1]-Prozesse kann wieder so formuliert werden, dass die Wurzel der Gleichung $1 - \beta z = 0$, d. h. $1/\beta$, außerhalb des Einheitskreises liegt. In dieser Form lässt sich die Aussage verallgemeinern.

Definition 3.13 (Invertierbarkeit von MA-Prozessen)

Ein MA[q]-Prozess (Y_t) heißt *invertierbar*, wenn die zugehörige charakteristische Gleichung

$$1 - \beta_1 z^1 - \dots - \beta_q z^q = 0 \quad (3.36)$$

nur Nullstellen außerhalb des Einheitskreises besitzt.

Satz 3.14 (Überführungsmöglichkeiten von AR und MA-Prozessen)

Jeder stationäre AR[p]-Prozess lässt sich als MA[∞]-Prozess darstellen. Jeder invertierbare MA[q]-Prozess lässt sich als AR[∞]-Prozess darstellen.

Die Bedeutung der MA-Darstellung liegt, wie bereits erwähnt, nicht zuletzt darin, dass über die gewichteten Summen von unabhängigen Zufallsvariablen ein vergleichsweise einfacher Zugang zur theoretischen Analyse ermöglicht wird.

Alle Prozesse, die sich als MA[∞]-Prozesse schreiben lassen, werden auch als *lineare Prozesse* bezeichnet. Die besondere Relevanz der linearen Prozesse röhrt dabei von dem grundlegenden *Woldschen Zerlegungssatz* her. Dieser sagt, dass jeder stationäre Prozess (Y_t) eindeutig als Summe zweier Prozesse geschrieben werden kann, $Y_t = D_t + R_t$. Dabei ist (R_t) ein MA[∞]-Prozess; (D_t) ist deterministisch. Deterministisch heißt ein Prozess, wenn er sich exakt vorhersagen lässt. Grob gesprochen ist D_t dann eine exakte Funktion der vergangenen Prozessvariablen D_{t-u} . Ein Beispiel dafür ist etwa der Prozess $Y_t = \cos(2\pi 0.25t)$.

Dass sich die Schätzung der Koeffizienten β_1, \dots, β_q eines MA-Prozesses schwieriger darstellt als die Schätzung der Koeffizienten von AR-Prozessen, liegt darin begründet, dass die β_u zwar linear mit den (nicht beobachtbaren) ε_t verknüpft sind, zur Schätzung aber die tatsächliche Zeitreihe (y_1, \dots, y_N) verwendet werden muss. Mit

diesen Werten sind die Koeffizienten nichtlinear verknüpft, wie die Formel (3.35) exemplarisch für einen MA[1]-Prozess zeigt.

Die Invertierbarkeit eines MA-Prozesses sichert nun die Möglichkeit, die Innovationen ε_t sowie die Parameter aufgrund einer Beobachtung, d. h. einer empirischen Reihe zu schätzen. Da zudem verschiedene MA-Prozesse gleicher Ordnung dieselbe Autokorrelationsfunktion hervorbringen können, ist die Invertierbarkeit ein wichtiges zusätzliches Kriterium bei der Bestimmung eines MA-Modells.

Beispiel 3.15 (MA-Prozesse mit gleicher ACF)

Bei dem MA[1]-Prozess $Y_t = \varepsilon_t - \beta\varepsilon_{t-1}$ erhält man die ACF

$$\rho_\tau = \begin{cases} 1 & \text{für } \tau = 0, \\ -\beta/(1 + \beta^2) & \text{für } \tau = \pm 1, \\ 0 & \text{sonst.} \end{cases}$$

Wegen $\beta/(1 + \beta^2) = (1/\beta)/(1 + (1/\beta)^2)$ sind die Autokorrelationsfunktionen der MA[1]-Prozesse mit den Parametern β und $1/\beta$ identisch.

Schätzen und Anpassen von MA-Modellen

Für invertierbare MA[q]-Modelle führt der CLS-Ansatz zu nichtlinearen Schätzungen, die mit numerischen Optimierungsroutinen bestimmt werden müssen. Beim MA[1]-Modell ist etwa

$$\sum_t \varepsilon_t^2 = \sum_t \left(Y_t + \sum_{u \geq 1} \beta^u Y_{t-u} \right)^2$$

zu minimieren. Das Prinzip des ULS-Ansatzes ist dann wieder, mittels der mit der CLS-Methode gewonnene Schätzung die Reihe in die Vergangenheit zu „prognostizieren“, um so zu besseren Schätzungen zu gelangen. Die ML-Methode basiert auch bei diesen Modellen auf der Maximierung der Likelihoodfunktion.

Eine spezielle Schätzmethode für MA-Prozesse ist der *Innovationsalgorithmus* von Brockwell und Davis (1996). Ausgehend von dem MA[q]-Prozess $Y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}$ werden die Ein-Schritt-Prognosen $\hat{Y}_{t,1}$ der Variablen Y_{t+1} sowie die Residuen $\hat{\varepsilon}_t = Y_t - \hat{Y}_{t-1,1}$ auf der Basis der Beobachtungen zu den Zeitpunkten $1, \dots, t$ rekursiv bestimmt. Dieser Algorithmus basiert auf der Autokovarianzfunktion und ergibt die Koeffizienten $\theta_{t,u}$ in der Approximation

$$\hat{Y}_{t-1,1} = \sum_{u=1}^t \theta_{t,u} \hat{\varepsilon}_{t-u} \quad (3.37)$$

Die Koeffizienten $\theta_{t,1}, \dots, \theta_{t,q}$ können dann für genügend großes t als Schätzwerte verwendet werden. Die Schätzer sind asymptotisch gemeinsam normalverteilt. Für

die einzelnen Koeffizienten gilt, wenn (Y_t) ein MA[q]-Prozess ist und m mit N gegen unendlich strebt:

$$\sqrt{N}(\hat{\theta}_{m,i} - \theta_i) \sim \mathcal{N}\left(0, \sum_{u=0}^{i-1} \theta_u^2\right) \quad (i = 1, \dots, q). \quad (3.38)$$

Für die Identifizierung eines MA-Modells sind wieder die Kennfunktionen wesentlich. Im Fall eines MA[q]-Prozesses gilt:

- Die ACF verschwindet für Lags, die größer als die Ordnung des Prozesses sind.
- Die PACF (π_τ) ist für $\tau > q$ exponentiell abklingend.

Die Eigenschaft der PACF ist zumindest für einen MA[1]-Prozess aufgrund der Darstellung als AR[∞]-Prozess einsichtig. Denn die Gleichung (3.35) macht deutlich, dass auch die von $Y_{t-1}, \dots, Y_{t-\tau}$, bereinigten Variablen \tilde{Y}_t und $\tilde{Y}_{t-\tau-1}$ noch eine gewisse Korrelation aufweisen.

Das Verschwinden der ACF für Lags, die größer als die Modellordnung sind, hat zur Konsequenz, dass für die empirischen Korrelationen für Lags $\tau > q$ gilt:

$$E(\hat{\rho}_\tau) \doteq 0 \quad (3.39)$$

$$\text{Var}(\hat{\rho}_\tau) \doteq \frac{1}{N} (1 + 2\rho_1^2 + \dots + 2\rho_q^2). \quad (3.40)$$

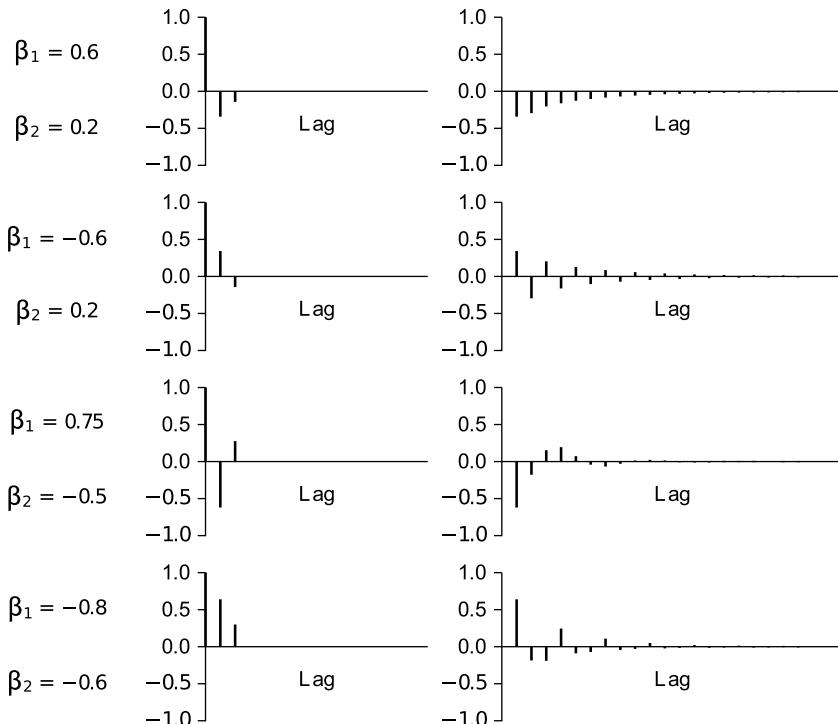


Abb. 3.7: ACF und PACF für MA[2]-Prozesse

Die als *Formel von Bartlett* bezeichnete asymptotische Varianz (3.40) für MA[q]-Prozesse ermöglicht die Spezifikation eines MA-Modells mit Hilfe der empirischen ACF. Dazu werden in der Formel für die Varianzen der $\hat{\rho}_\tau$ die theoretischen Korrelationen durch die empirischen ersetzt. Dann wird für jedes Lag τ die empirische Korrelation sowie die Bartlett-Grenzen, das Doppelte der geschätzten Standardabweichung für einen MA[$\tau - 1$]-Prozess, also von

$$\pm 2 \sqrt{(1 + 2\hat{\rho}_1^2 + \dots + 2\hat{\rho}_{\tau-1}^2)/N},$$

eingezeichnet. Wegen der asymptotischen Normalverteilung der empirischen Autokorrelationen entspricht dies in etwa einem zentralen 95 % Schwankungsintervall, falls die theoretischen Korrelationen für $\tau' > \tau$ null sind. Die grafische Darstellung der Grenzen zeigt eine sich öffnende Trichterform. Um die Grenzen richtig zu interpretieren, sind für jedes Lag τ in der Höhe $\pm 2 \sqrt{(1 + 2\hat{\rho}_1^2 + \dots + 2\hat{\rho}_{\tau-1}^2)/N}$ Parallelen zur Abszisse zu ziehen. Das kleinste Lag, von dem an die Parallelen alle Korrelationen einschließen, ergibt (vorläufig) die Ordnung des MA-Modells zu $q = \tau - 1$.

Beispiel 3.16 (Simulierter MA[7]-Prozess)

Für den MA[7]-Prozess $Y_t = \varepsilon_t + 0.80\varepsilon_{t-1} - 1.80\varepsilon_{t-4} - 1.64\varepsilon_{t-5} - 2.50\varepsilon_{t-7}$ mit $\varepsilon_t \sim \mathcal{N}(0, 1)$ wird eine Reihe der Länge $N = 50$ simuliert.

Die Abbildung zeigt die empirische Autokorrelationsfunktion mit den Bartlett-Grenzen (gestrichelte Hilfslinien). Die durchgezogenen Hilfslinien geben die Höhe der Grenzen, wenn ein MA[2]-Modell angebracht wäre. Lägen also alle Autokorrelationen ab dem Lag 3 innerhalb dieser Grenzen, würde man auf ein MA[2]-Modell tippen. Offensichtlich ist nach dem Lag 7 keine Autokorrelation mehr bedeutsam. So führt das Vorgehen hier auf einen MA[7]-Prozess als Modell.

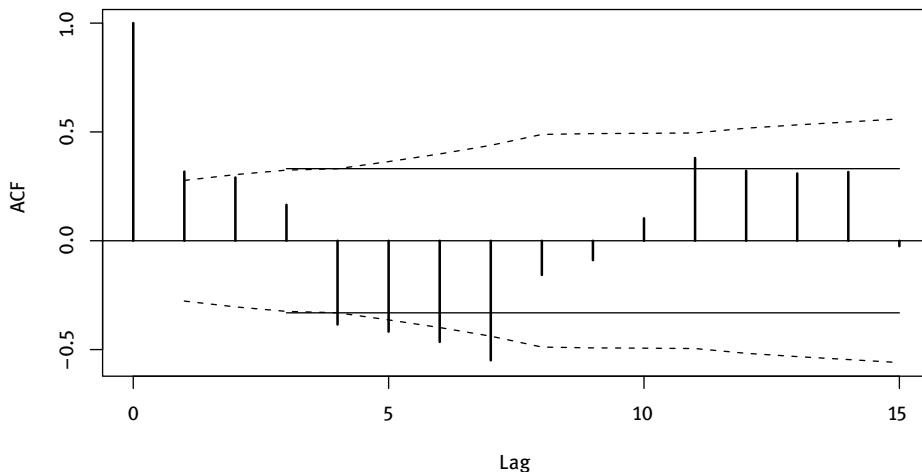


Abb. 3.8: ACF mit Bartlett-Grenzen

R-Code:

```

1 n <- 50
2 e <- rnorm(n+7)
3 y <- e[-c(1:7)] + 0.8*e[-c(1:6,n+7)] - 1.8*e[-c(1:3,n+(4:7))]
4 -1.64*e[-c(1:2,n+c(3:7))]-2.50*e[-(n+c(1:7))]
5 a <- acf(y,lag.max=15,ci.type = "ma")
6 lines(c(3,15),rep(2*sqrt((1+2*a$acf[2]^2+2*a$acf[3]^2)/n),2))
7 lines(c(3,15),rep(-2*sqrt((1+2*a$acf[2]^2+2*a$acf[3]^2)/n),2))

```

Die Simulation von MA-Prozessen kann direkt die Definition umsetzen, ohne eine Schleife zu benötigen. Bei der Funktion `acf` werden standardmäßig die Grenzen eingezeichnet, die für einen White-Noise Prozess gelten. Die Bartlett-Grenzen sind über das optionale Argument `ci.type = "ma"` anzufordern.

3.3 ARMA-Modelle

Definition und grundlegende Eigenschaften

Die nahe liegende Verknüpfung der AR- und der MA-Prozesse führt zu den *Autoregressiven Moving-Average-*, kurz *ARMA-Prozessen*:

$$Y_t = \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q}. \quad (3.41)$$

Diese werden üblicherweise in der Form

$$Y_t - \alpha_1 Y_{t-1} - \cdots - \alpha_p Y_{t-p} = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q},$$

bzw.

$$(1 - \alpha_1 B - \cdots - \alpha_p B^p) Y_t = (1 - \beta_1 B - \cdots - \beta_q B^q) \varepsilon_t$$

oder ganz kurz gemäß $\alpha(B)Y_t = \beta(B)\varepsilon_t$ angegeben. Die Berücksichtigung eines von null verschiedenen Niveaus geschieht am einfachsten mittels Zentrierung von Y_t in der letzten Gleichung:

$$\alpha(B)(Y_t - \mu) = \beta(B)\varepsilon_t. \quad (3.42)$$

Ein ARMA-Prozess ist stationär, wenn die Lösungen der charakteristischen Gleichung des AR-Teils außerhalb des Einheitskreises liegen. Er ist invertierbar, wenn die charakteristische Gleichung des MA-Teils nur Lösungen besitzt, die außerhalb des Einheitskreises liegen. Ein stationärer und invertierbarer ARMA-Prozess kann sowohl als AR[∞]-Prozess als auch als MA[∞]-Prozess geschrieben werden. Der Vorteil

der ARMA-Modelle liegt in ihrer Sparsamkeit bez. der benötigten Parameter. Auch komplexes Verhalten einer Zeitreihe lässt sich durch ARMA-Modelle mit einer eher geringen Zahl von Parametern beschreiben.

Eine Motivation für ARMA-Prozesse ergibt sich aus der Eigenschaft, dass Summen von unabhängigen ARMA-Prozessen wieder ARMA-Prozesse sind. Damit werden z. B. ARMA-Modelle für aggregierte Zeitreihen, wie sie in der Ökonomie häufig anzutreffen sind, nahe gelegt. Genauer gilt:

$$\text{ARMA}[p, q] + \text{ARMA}[p', q'] = \text{ARMA}[p + q', \max\{q + p', p + q'\}] . \quad (3.43)$$

Speziell ist

$$\text{AR}[p] + \text{White Noise} = \text{ARMA}[p, p] . \quad (3.44)$$

Wird also die Beobachtung eines $\text{AR}[p]$ -Prozesses zufällig gestört, so resultiert ein $\text{ARMA}[p, p]$ -Modell. Das Beispiel 3.17 illustriert diesen Sachverhalt.

Choi (1992) gibt einen Überblick über die verschiedenen Ansätze zur Identifikation von ARMA-Modellen. Dort wird auch auf die von Tsay & Tiao (1984) eingeführte ESACF eingegangen.

Schätzen der Parameter

Für die *Schätzung der Parameter* von ARMA-Prozessen stehen bis auf die Yule-Walker- und die Burg-Schätzer die gleichen Methoden zur Verfügung, die schon bei den AR-Prozessen erwähnt wurden. Das sind also die CLS-, ULS- und ML-Schätzer. Diese sind wieder asymptotisch gemeinsam normalverteilt mit den wahren Parameterwerten als (asymptotischen) Erwartungswerten. Daher kann man die Signifikanz der einzelnen Parameterschätzungen mittels t -Tests abschätzen, oder äquivalent dazu, mit dem Check, ob das jeweils zugehörige Konfidenzintervall den Wert null einschließt.

Speziell die Anwendung der ML-Methode setzt voraus, dass die Störungen normalverteilt sind. Dies sollte wie bei den AR-Modellen zumindest nach Anpassung des Modells anhand der Residuen überprüft werden.

Als spezifisches Schätzverfahren kommt der *Ansatz von Durbin-Hannan-Rissanen* hinzu. Dabei wird zuerst ein AR-Modell hoher Ordnung geschätzt. Die sich daraus ergebenden Residuen $\hat{\varepsilon}_t$ werden benutzt, als ob sie die korrekten Beobachtungen des White-Noise-Prozesses wären. Dann liefert der Regressionsansatz

$$Y_t = \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} - \beta_1 \hat{\varepsilon}_{t-1} - \cdots - \beta_q \hat{\varepsilon}_{t-q} + \eta_t .$$

Schätzungen der Parameter des ARMA-Modells. Mit den Residuen dieses Modells kann das Verfahren dann iteriert werden. Die Schätzer sind asymptotisch normalverteilt aber nicht effizient, sie haben eine größere Varianz als die drei oben angegebenen.

Spezifikation von ARMA-Modellen

Die **Spezifikation von ARMA-Modellen** ist keine sehr leichte Aufgabe mehr. Im Rahmen des *Box-Jenkins-Ansatzes* betrachtet man zuerst die ACF und die PACF der Ausgangsreihe und versucht, aus typischen Mustern die Modellordnungen zu erraten. Wie die Abbildung 3.9 für ARMA[1, 1]-Prozesse erkennen lässt, ist das Verhalten dieser Kennfunktionen schon in gemischten Modellen niedrigerer Ordnung durchaus komplex. Die Tabelle 3.1 fasst deren Verhalten für allgemeine ARMA[p, q]-Prozesse zusammen.

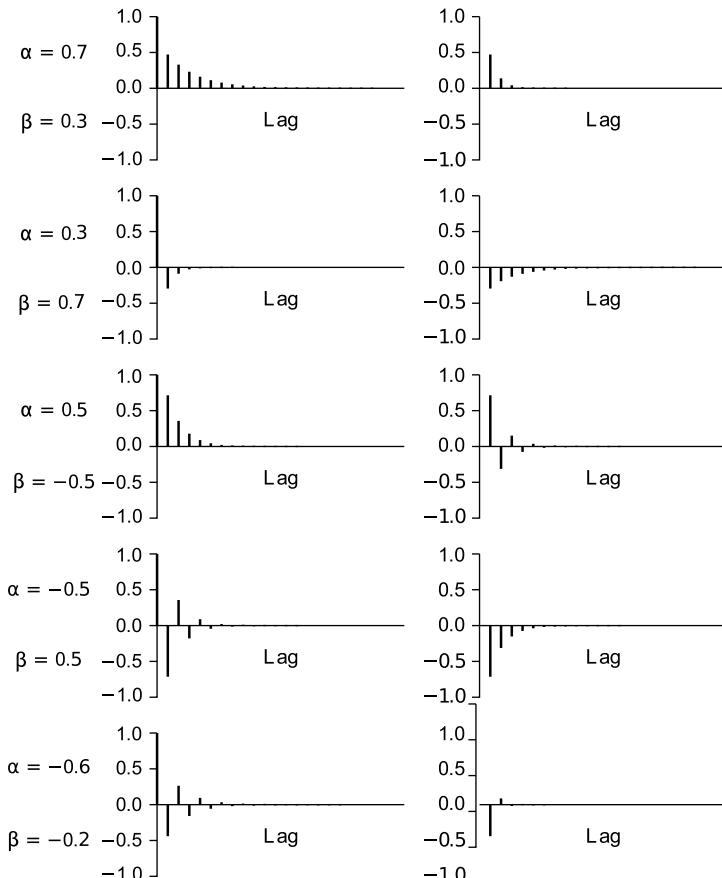


Abb. 3.9: ACF und PACF für ARMA[1, 1]-Prozesse

Tab. 3.1: Typisches Verhalten von ACF und PACF in stationären und invertiblen ARMA-Modellen

	AR[p]	MA[q]	ARMA[p, q]
ACF	klingt ab	bricht nach dem Lag q ab	klingt ab
PACF	bricht nach dem Lag p ab	klingt ab	klingt ab

Somit fällt das Erkennen der zugrunde liegenden Struktur nicht leicht. Hat man sich dennoch für ein Modell entschieden, so werden die Parameter geschätzt. Die Analyse der Residuen, insbesondere der ACF und der PACF der Residualreihe, gibt Hinweise, ob alle Autokorrelationen erfasst sind. Formal lässt sich dies mit dem Ljung-Box-Pierce-Test überprüfen, vgl. (3.31). Falls noch Autokorrelationen in der Residualreihe verblieben sind, ist die Schätzung für ein neues, die noch vorhandenen Abhängigkeiten berücksichtigendes Modell erneut durchzuführen. Dies gibt eine neue Residualreihe. So tastet man sich schließlich auf iterative Weise an ein geeignetes Modell heran. Auch sind die Korrelationen der Parameterschätzungen zu untersuchen. Starke Korrelationen zwischen AR- und MA-Parametern kommt vor allem dann vor, wenn die eine oder andere Ordnung zu hoch gewählt wurde. Andererseits sind die Korrelationen der AR-Parameter untereinander und ebenso der MA-Parameter untereinander oft recht hoch, ohne dass dies ein Hinweis auf Spezifikationsprobleme wäre.

Beispiel 3.17 (Tremor)

Als Tremor bezeichnet man das Zittern, das durch unwillkürliche, weitgehend rhythmisch aufeinander folgende Kontraktionen der Muskeln bewirkt wird. Dabei ist der physiologische Tremor normalerweise nicht sichtbar. Er kann jedoch mit einer besonderen Vorrichtung gemessen werden. Die Reihe TREMOR ist einen Ausschnitt aus einer solchen Messung.

Die ACF und die PACF zeigen beide ein ‚exponentielles Abklingen‘; sie ‚brechen nicht abrupt‘ ab. Das deutet auf ein gemischtes Modell hin. Das Schwingungsverhalten der ACF legt ein AR-Polynom mindestens der Ordnung 2 nahe. Ebenso weist die PACF auf eine MA-Ordnung von mindestens zwei hin. In der Tat ergibt

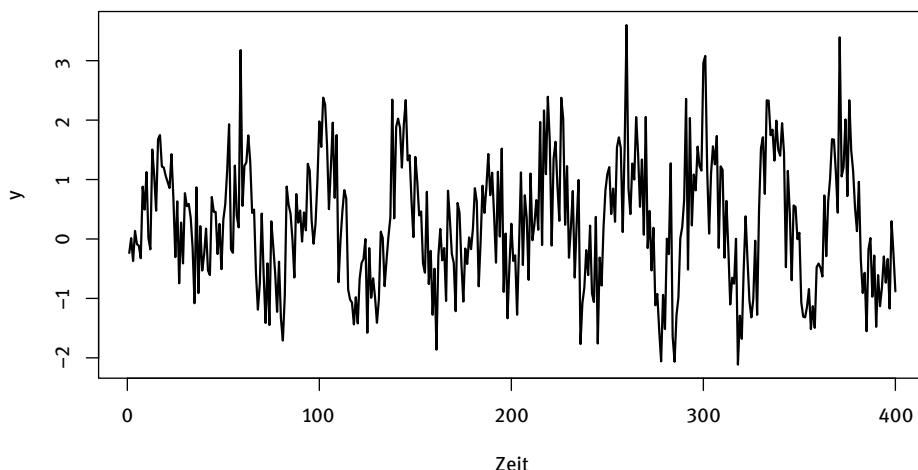


Abb. 3.10: Physiologischer TREMOR (Ausschnitt)

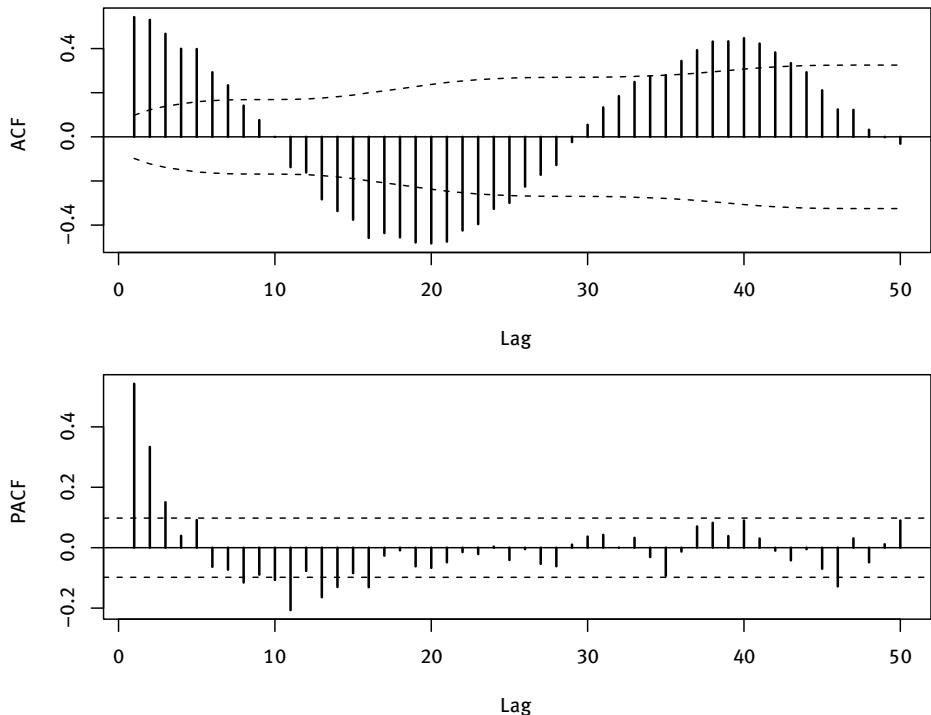


Abb. 3.11: ACF und PACF der Zeitreihe TREMOR

das ARMA[2, 2]-Modell

$$(1 - 1.9630B + 0.9885B^2)(Y_t - 0.2815) = (1 - 1.8215B + 0.8470B^2)\varepsilon_t$$

eine zufrieden stellende Anpassung. Für alle Koeffizienten zeigen die Quotienten aus Schätzwert und zugehörigem Standardfehler, dass die zugehörigen *P*-Werte praktisch null sind. Die Koeffizienten werden daher alle im Modell behalten:

Coefficients:

	ar1	ar2	ma1	ma2	intercept
1.	1.9630	-0.9885	-1.8215	0.8470	0.2815
s.e.	0.0076	0.0077	0.0309	0.0326	0.0387
sigma^2 estimated as 0.581: log likelihood = -460.35,					
aic = 932.7					

Die Korrelationen zwischen den Schätzern ist teilweise recht hoch, insbesondere die zwischen $\hat{\alpha}_1$ und $\hat{\alpha}_2$ sowie die zwischen $\hat{\beta}_1$ und $\hat{\beta}_2$, siehe die folgende Abbildung. Das erklärt, warum bei dieser Reihe die Informationskriterien eher ein ARMA[1, 1] auswählten. Jedoch verbleiben dann relevante Korrelationen in

	ar1	ar2	ma1	ma2	intercept		
ar1	1.000	-0.981	-0.542	0.522	0.017		
lags	df	statistic	p-value	lags	df	statistic	p-value
ar2	-0.981	1.000	0.563	-0.563	-0.020		
ma1	-0.542	0.563	1.000	-0.982	-0.011		
ma2	0.522	-0.563	-0.982	1.000	0.012		
intercept	0.017	-0.020	-0.011	0.012	1.000		
12	8	9.136	0.331	36	32	31.035	0.515
18	14	15.404	0.351	42	38	39.116	0.420
24	20	16.930	0.658	48	44	45.192	0.422

Abb. 3.12: Korrelationen der Schätzungen (oben) und Ergebnisse der Ljung-Box-Pierce-Tests (unten)

den Residuen. Dass die Korrelation zwischen den beiden Parametermengen des AR- und MA-Teils moderat ist, weist auf die Gültigkeit des Modells hin. Beim ARMA[2, 2]-Modell sind keine signifikanten Abhängigkeiten mit dem Ljung-Box-Pierce-Test mehr festzustellen. Im Gegenteil, die *P*-Werte sind für die verschiedenen Anzahlen von Summanden alle ausreichend hoch, um das Modell zu akzeptieren.

Nach Timmer (1998), dem die Daten zu verdanken sind, stellt ein AR[2]-Prozess für derartige Zeitreihen von den physiologischen Gegebenheiten her ein adäquates Modell dar. Allerdings ist der Messvorgang selbst nicht störungsfrei. Der AR[2]-Prozess wird durch einen White-Noise-Prozess überlagert. Daher ist es wegen (3.44) nicht verwunderlich, dass für die empirische Reihe ein ARMA[2, 2]-Modell resultiert.

R-Code:

```

1 data(TREMOR)
2 y <- TREMOR
3 plot(y)
4 a <- acfpacf(y,50)
5 out <- arima(y,order=c(2,0,2))
6 out
7 S <- vcov(out)
8 D <- diag(1/sqrt(diag(S)))
9 R <- round(D%*%S%*%D,3)
10 colnames(R)<-rownames(R)<-c("ar1","ar2","ma1","ma2","intercept")
11 print(R)
12 LjungBoxPierceTest(out$resid,4,48)

```

Die Schätzung von ARMA-Modellen erfolgt mit der Funktion `arima`. Die gewählten Ordnungen p und q sind der erste bzw. dritte Wert des Arguments `order`. In dem Ergebnis `out` sind u. a. die Kovarianzen der Parameterschätzungen enthalten; sie werden in der Zeile 7 extrahiert. Die folgenden vier Zeilen machen daraus die Korrelationsmatrix und geben diese aus. Der hier aufgerufene Ljung-Box-Pierce-Test gehört zum Paket `tsapp`.

Als zusätzliches Hilfsmittel wurde die *inverse ACF*, oder kurz *IACF*, eingeführt. Die IACF von $\alpha(B)Y_t = \beta(B)\varepsilon_t$ ist die ACF des Prozesses $\beta(B)Y_t = \alpha(B)\varepsilon_t$. Dementsprechend vertauschen sich die Verhaltensmuster, und die IACF sollte bei AR-Prozessen für Lags, die größer sind als die Modellordnung, praktisch null sein. Diese Kennfunktion ist aber nicht hilfreicher als die PACF, da sie das grundlegende Identifikationsproblem nicht beseitigt.

Eine wesentliche Verbesserung für die Bestimmung der Ordnung versprechen die *erweiterten Stichprobenautokorrelationen*, *ESACF*, von Tsay & Tiao (1984).

Zu einer Tabelle mit einem durch die Ordnung des ARMA-Modells bedingten Bereich von Nullen führt die ESACF. Der Ausgangspunkt ist die folgende Überlegung.

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

$$(1 - \alpha_1 B - \dots - \alpha_p B^p)Y_t = (1 - \beta_1 B - \dots - \beta_q B^q)\varepsilon_t .$$

Dann ist $Z_t = (1 - \alpha_1 B - \dots - \alpha_p B^p)Y_t$ ein MA[q]-Prozess:

$$Z_t = (1 - \beta_1 B - \dots - \beta_q B^q)\varepsilon_t .$$

Dies legt nahe, einer Zeitreihe sukzessive AR[p]-Modelle mit wachsenden p anzupassen und die Residuen daraufhin zu untersuchen, ob sie als Realisationen eines MA[q]-Prozesses aufgefasst werden können. Der Pferdefuß liegt darin, dass die KQ-Schätzungen der AR-Parameter bei einem gemischten Modell nicht konsistent sind. Die direkte Umsetzung der Idee würde folglich leicht zu Fehlspezifikationen führen. Tsay und Tiao haben aber eine iterative KQ-Schätzung entwickelt, die dieses Problem behebt. Die erweiterte Stichprobenautokorrelation $r_{u(k)}$ ist dann die empirische Autokorrelation zum Lag u von (Z_t) , wenn ein AR[k]-Modell gemäß Tsay und Tiaos Vorgehen angepasst wurde. Die ESACF ist nützlich, da für einen ARMA[p, q]-Prozess gilt:

$$r_{u(k)} \doteq 0 \quad \text{für} \quad u - q > k - p \geq 0 . \tag{3.45}$$

Die anderen $r_{u(k)}$ sind von null verschiedene Konstanten oder stetige Zufallsvariablen mit Werten zwischen -1 und $+1$.

Als Resultat ergibt sich in der Theorie bei einer Tafel der ESACF, bei der die Zeilen durch die AR-Ordnungen und die Spalten durch die MA-Ordnungen bestimmt sind, ein Keil von Nullen, bei dem die Spitze die Ordnung des ARMA-Modells anzeigt. Für

einen ARMA(1,2)-Prozess sieht das z. B. folgendermaßen aus; die Eintragungen \times sind dabei nicht spezifizierbar:

AR	MA								...
	0	1	2	3	4	5	6	...	
0	\times	...							
1	\times	\times	\times	0	0	0	0	0	...
2	\times	\times	\times	\times	0	0	0	0	...
3	\times	\times	\times	\times	\times	0	0	0	...
4	\times	\times	\times	\times	\times	\times	0	0	...
5	\times	0	...						
:	:	:	:	:	:	:

In der Anwendung wird eine entsprechende Tafel erstellt, in der die Zellen mit einem Kreuz o. ä. markiert sind, die sich signifikant von null unterscheiden. Es ist die Hoffnung, dass sich ein der Theorie entsprechendes Muster ergibt, das die zu wählenden Modellordnungen anzeigen. Leider hat sich die ESACF erfahrungsgemäß als wenig zuverlässig erwiesen.

Beispiel 3.18 (Tremor – Fortsetzung von Seite 95)

Für die Zeitreihe TREMOR erhält man die folgende Tafel der ESACF. Das sich im Beispiel 3.17 als gutes Modell ergebende ARMA[2, 2]-Modell ist daraus nicht so recht zu erkennen.

AR / MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	\times	o	o	x	x	x	x	x						
1	x	o	o	o	x	o	o	o	o	o	x	o	o	o
2	x	x	o	o	x	o	o	o	o	o	x	o	o	o
3	x	x	x	o	x	o	o	o	o	o	o	o	x	o
4	x	x	o	x	x	o	o	o	o	o	o	o	o	o
5	x	x	x	x	o	o	o	o	o	o	o	o	o	o
6	x	o	x	x	o	o	o	o	o	o	o	o	o	o
7	x	x	x	o	x	x	o	o	o	o	o	o	o	o

R-Code:

```
1 library(TSA)
2 eacf(y, ar.max = 7, ma.max = 13)
```

Die benötigte Funktion ist im Paket TSA; der Aufruf ist elementar.

Neben den Kennfunktionen können wieder die *Informationskriterien* herangezogen werden, um zwischen verschiedenen, eine Reihe in etwa gleich gut erfassenden Modellen auszuwählen. Sie nehmen hier die Form an:

$$\text{AIC}(p, q) = \ln\left(\frac{\text{RSS}_{p,q}}{N}\right) + \frac{2(p+q)}{N} \quad (3.46)$$

$$\text{AICC}(p, q) = \ln\left(\frac{\text{RSS}_{p,q}}{N}\right) + \frac{2(p+q)}{N-p-q-2} \quad (3.47)$$

$$\text{BIC}(p, q) = \ln\left(\frac{\text{RSS}_{p,q}}{N}\right) + \frac{(p+q)\ln(N)}{N} \quad (3.48)$$

$$\text{HQ}(p, q) = \ln\left(\frac{\text{RSS}_{p,q}}{N}\right) + \frac{2(p+q) \cdot c \cdot \ln(\ln(N))}{N} \quad (c > 1). \quad (3.49)$$

Sie sind auch im Falle von Subsetmodellen von Bedeutung, bei Modellen also, bei denen Parameter null gesetzt werden können. Dann ist in den Formeln für $p+q$ die Anzahl der tatsächlich geschätzten Parameter einzusetzen. Die Auswahl eines Subsetmodells wird im Beispiel 3.10 für ein AR-Modell betrieben.

Die *Spezifikation* von ARMA-Modellen verläuft in folgenden Schritten:

1. Analyse von ACF, PACF (IACF) und ESACF:
Identifizieren von p und q über „matching“ von empirischen und theoretischen Kennfunktionen bzw. Erkennen des geeigneten Nullblocks der durch die ESACF gegebenen Tafel.
2. Schätzen des Modells mit einer der im Rahmen der AR-Modelle angesprochenen Methoden (CLS, ULS, ML) bzw. mit dem Verfahren von Durbin, Hannan und Rissanen.
3. Überprüfen, welche Parameterwerte signifikant von null verschieden sind. Restriktive Neuschätzung unter Nullsetzen der nichtsignifikanten Parameter. Check der Korrelationen der Parameterschätzungen.
4. Check der Residuen auf White-Noise. Dazu stehen wieder die im Rahmen der AR-Modelle besprochenen Methoden zur Verfügung. Falls diese noch Abhängigkeiten aufweisen, gehe man zurück nach 1.

3.4 ARIMA-Modelle

Definition und Spezifikation von ARIMA-Modellen

Viele empirische Zeitreihen sind nicht stationär, sondern weisen eine Trendkomponente auf. Box und Jenkins haben daher ARMA-Modelle in der Weise verallgemeinert, dass nach geeignetem Differenzenbilden ein stationärer ARMA-Prozess vorliegt. Damit bekommt man zunächst Modelle für Zeitreihen mit einer Trendkomponente. Das Modell für die Ausgangsreihe erhält man also durch „Integrieren“ aus einem ARMA-Prozess. Diese Modelle heißen *Autoregressive Integrierte Moving Average Prozesse*,

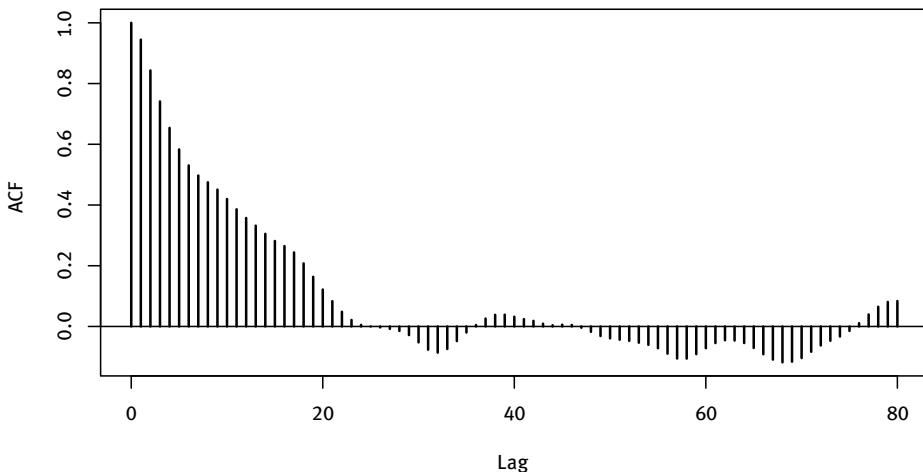


Abb. 3.13: ACF der Realisation ($N = 200$) eines integrierten AR-Prozesses

kurz *ARIMA*[p, d, q]-Prozesse. Sie lassen sich in der folgenden Form schreiben:

$$\alpha(B) \underbrace{(1 - B)^d Y_t}_{\text{stationär}} = \beta(B) \varepsilon_t .$$

Für die Identifikation von ARIMA-Modellen ist zusätzlich zu den bei ARMA-Modellen vorzunehmenden Schritten noch der Grad der Differenzenbildung d zu bestimmen. Zunächst einmal kann auch aus der empirischen ACF erkannt werden, ob Differenzenbilden notwendig ist. Falls das AR-Polynom eine Einheitswurzel hat, sich also in der Form $\alpha(B)(1 - B)$ schreiben lässt, gilt nämlich, dass die empirische ACF nur langsam von +1 herkommend abklingt, vgl. Abbildung 3.13. Anschließend (oder alternativ) bietet es sich an, die Methode der *variaten Differenzen*, vgl. Kapitel 2, einzusetzen.

Will man nun einer empirischen Reihe nach Differenzenbildung ein ARMA-Modell anpassen, so empfiehlt es sich, für den MA-Teil eine Ordnung zu unterstellen, die mindestens so groß ist wie der Grad der Differenzenbildung. Wie sich aus dem folgenden Satz ergibt, wird durch jede Differenzenbildung eines ARMA-Prozesses eine Einheitswurzel bei dem MA-Teil produziert. War also die Reihe schon vor der Differenzenbildung stationär, so verliert die den Schätzungen zugrunde liegende Theorie wegen der Einheitswurzel ihre Gültigkeit. Die Standardfehler und die darauf basierenden t -Statistiken und P -Werte für die Koeffizienten stimmen also nicht mehr.

Satz 3.19 (Verhalten der Differenzen stationärer ARMA-Prozesse)

Sei (Y_t) ein stationärer ARMA[p, q]-Prozess, $\alpha(B)Y_t = \beta(B)\varepsilon_t$. Die Wurzeln des charakteristischen Polynoms des autoregressiven Teils seien betragsmäßig alle größer als eins, die des charakteristischen Polynoms des MA-Teils, $1 - \beta_1 z - \dots - \beta_q z^q = 0$, seien z_1, \dots, z_q . Dann bilden die ersten Differenzen $Y_t - Y_{t-1}$ einen ARMA[$p, q+1$]-Prozess, bei dem der autoregressive Teil sich nicht ändert, der MA-Teil die Wurzeln $z_1, \dots, z_q, 1$ hat.

Beispiel 3.20 (Kaffee-Konsum)

Die Abbildung zeigt die Reihe des logarithmierten Kaffee-Konsums von 1910 bis 1970 in den USA sowie die ACF und PACF der Originalreihe (links) und der einmal differenzierten Reihe (rechts). Dieser Zeitreihe KAFFEE wurden verschiedene Modelle angepasst. Geben Sie zur Übung – bevor Sie weiterlesen – mit Begründung an, welches der Modelle Sie für das geeignetste halten.

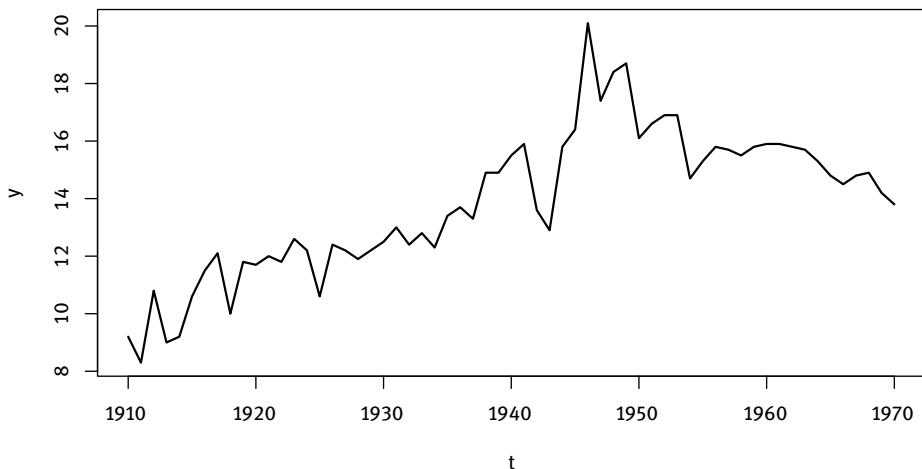


Abb. 3.14: Logarithmierte Werte des jährlichen Kaffee-Konsums in den USA

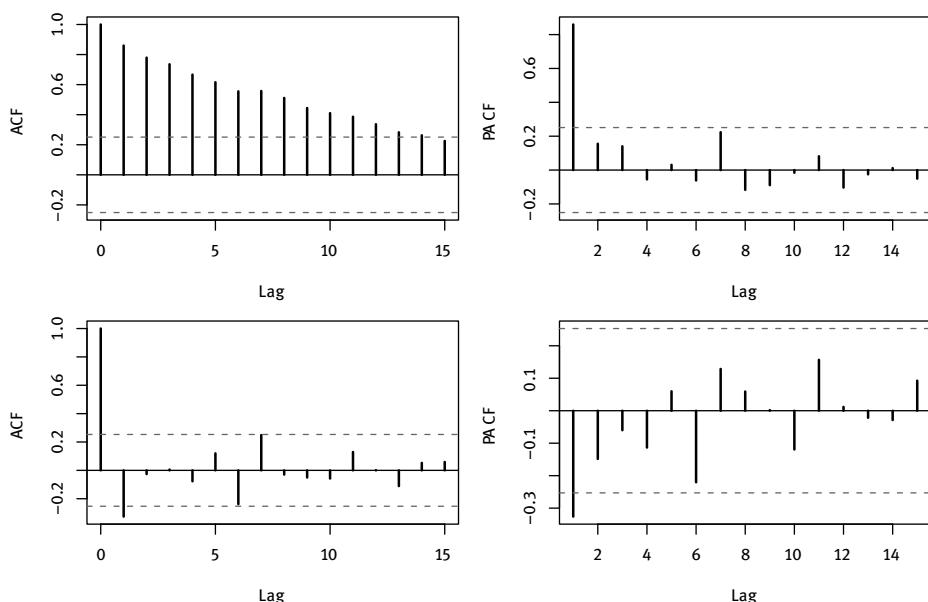


Abb. 3.15: ACF/PACF der nicht-differenzierten (oben) und der differenzengebildeten (unten) Reihe KAFFEE

Modell 1: ARIMA[1, 0, 0]

Coefficients:

	ar1	intercept
s.e.	0.8950	13.3379
sigma^2	0.0563	1.2687
estimated	as 1.335:	log likelihood=-96.17 , aic=198.35

LjungBoxPierceTest(out1\$residuals, 1, 20)

lags	df	statistic	p-value
6	5	9.016	0.108
12	11	14.963	0.184
18	17	17.759	0.404

Modell 2: ARIMA[0, 0, 1]

Coefficients:

	ma1	intercept
s.e.	0.7776	13.8339
sigma^2	0.0888	0.3841
estimated	as 2.889:	log likelihood=-119.38 , aic=244.76

LjungBoxPierceTest(out2\$residuals, 1, 20)

lags	df	statistic	p-value
6	5	82.317	0
12	11	126.044	0
18	17	135.331	0

Modell 3: ARIMA[1, 0, 1]

Coefficients:

	ar1	ma1	intercept
s.e.	0.9639	-0.3578	12.8027
sigma^2	0.0343	0.1351	1.9172
estimated	as 1.210:	log likelihood=-93.35 , aic=194.69	

LjungBoxPierceTest(out3\$residuals, 2, 20)

lags	df	statistic	p-value
6	4	3.592	0.464
12	10	7.681	0.660
18	16	9.955	0.869

Modell 4: ARIMA[1, 1, 0]

```
Coefficients:
      ar1
      -0.3176
  s.e.   0.1218
sigma^2 estimated as 1.260: log likelihood=-92.13,
                           aic=188.27
LjungBoxPierceTest(out4$residuals,1,20)
  lags df statistic p-value
      6   5     3.879   0.567
     12  11     9.398   0.585
     18  17    11.593   0.824
```

Modell 5: ARIMA[0, 1, 1]

```
Coefficients:
      ma1
      -0.3854
  s.e.   0.1252
sigma^2 estimated as 1.231: log likelihood=-91.45,
                           aic=186.9
LjungBoxPierceTest(out5$residuals,1,20)
  lags df statistic p-value
      6   5     2.536   0.771
     12  11     6.870   0.810
     18  17     8.329   0.959
```

Modell 6: ARIMA[1, 1, 1]

```
Coefficients:
      ar1      ma1
      0.1016  -0.4735
  s.e.  0.3337  0.2981
sigma^2 estimated as 1.229: log likelihood=-91.4,
                           aic=188.81
LjungBoxPierceTest(out6$residuals,2,20)
  lags df statistic p-value
      6   4     2.595   0.628
     12  10     6.835   0.741
     18  16     8.206   0.942
```

Die ACF- und PACF-Darstellung in der Abbildung legen für die nicht-differenzengebildete Reihe eher ein AR[1]-Modell nahe mit einem großen AR-Koeffizienten. Denn bei der PACF bricht nach dem ersten Lag praktisch ab, während die ACF abklingt, allerdings recht langsam. Für die differenzengebildete Reihe legen die Kennfunktionen dagegen ein MA[1]-Modell nahe. Denn die Autokorrelation zum Lag eins ist noch groß, während die weiteren schon praktisch null sind; die ACF bricht also ab. Die PACF zeigt dagegen kein so abruptes Abbrechen.

Das Modell 2 scheidet sofort aus, weil der Ljung-Box-Pierce-Test zeigt, dass bei den Residuen eine hohe Autokorrelation verblieben ist. Alle anderen Modelle sind von diesem Gesichtspunkt her akzeptabel.

Beim Modell 6 ist der AR[1]-Koeffizient nicht signifikant. Das weist auf eine Überparametrisierung hin. Folglich ist auch dieses Modell eher nicht zu wählen.

Bei allen anderen Modellen sind die Koeffizienten in Ordnung. Zur Auswahl des besten Modells kann nun auf das AIC-Kriterium zurückgegriffen werden. Das ist für das Modell 5 am kleinsten. Dies wird folglich als Modell ausgewählt.

Es zeigt sich also, dass die Vorab-Einschätzung bestätigt wird.

Saisonale ARIMA-Modelle

Reihen, die ein saisonales Verhalten aufweisen, können im Prinzip ARMA-Modelle der Form

$$\alpha(B) \underbrace{(1 - B)^d (1 - B^s)^D Y_t}_{\text{stationär}} = \beta(B) \varepsilon_t$$

angepasst werden. Hier wird die Saison (mit der Periode s) durch eine geeignete Anzahl D von Differenzenbildungen eliminiert; weiter wird die Reihe durch d -maliges einfaches Differenzenbilden trendbereinigt. Solche Reihen weisen jedoch häufig signifikante Werte der ACF bzw. PACF für die ersten Lags und dann für Lags um die Saisonperioden $s, 2s, \dots$ auf, während die Werte für dazwischenliegende Lags vernachlässigbar sind. Das ist Ausdruck der Tatsache, dass etwa bei Monatsdaten die Werte eines Monats nicht nur von den Werten der unmittelbar davor liegenden Monate abhängen, sondern auch von denen des gleichnamigen Monats des Vorjahres sowie vielleicht noch von dessen Vormonaten. Statt nun ein Modell der Art

$$(1 - \alpha_1 B - \alpha_s B^s - \alpha_{s+1} B^{s+1})(1 - B)^d (1 - B^s)^D Y_t = (1 - \beta_1 B - \beta_s B^s - \beta_{s+1} B^{s+1}) \varepsilon_t$$

anzupassen, kann es sparsamer sein, die beiden Polynome $\alpha(B)$ und $\beta(B)$ als Produkte anzusetzen:

$$(1 - \phi B)(1 - \Phi B^s)(1 - B)^d (1 - B^s)^D Y_t = (1 - \psi B)(1 - \Psi B^s) \varepsilon_t .$$

Den Koeffizienten zum Lag $s + 1$ auf diese Weise zu berücksichtigen, ist stimmig. Sei dazu $Z_t = (1 - B)^d(1 - B^s)^D Y_t$. Dann lässt sich die letzte Gleichung schreiben als

$$(1 - \phi B - \Phi B^s + \phi \Phi B^{s+1})Z_t = (1 - \psi B - \Psi B^s + \psi \Psi B^{s+1})\varepsilon_t . \quad (3.50)$$

Stelle man sich nun Monatsdaten vor; der Einfachheit der Erläuterung wegen sei der November aktuell. Der aktuelle Wert hängt mit dem Faktor ϕ vom Oktoberwert und mit dem Faktor Φ von dem des November des letzten Jahres ab. Beide Werte hängen aber auch vom Wert des Oktober des letzten Jahres ab. Somit wird dessen Einfluss quasi doppelt berücksichtigt. Bei der Produktbildung wird er mittels $-\phi\Phi$ einmal ‚herausgenommen‘.

Die von Box und Jenkins eingeführten *saisonalen ARIMA-Prozesse*, SARIMA $[p, d, q] \times [P, D, Q]_s$ -Prozesse, lassen sich allgemein in der folgenden Weise schreiben:

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D Y_t = \psi(B)\Psi(B^s)\varepsilon_t . \quad (3.51)$$

Für die additive Variante des Census X-11 Saisonbereinigungsverfahrens für Monatsdaten wurde ein spezielles saisonales ARIMA-Modell gefunden, dessen Anpassung praktisch mit der Saisonbereinigung übereinstimmt. (Jedenfalls wenn bei der Reihe keine Ausreißer oder sonstigen Kalenderunregelmäßigkeiten vorkommen.) Den Ausgangspunkt bildet das Modell

$$Y_t = T_t + S_t + \varepsilon_t ,$$

bei dem Trend T_t und Saison S_t einfachen ARIMA-Prozessen folgen. Das dann von Van Winkel (1982) gefundene Modell lautet:

$$(1 - B)^2(1 - B^{12})Y_t = (1 - 0.34B + 0.486B^2)(1 - 0.485B^{12} + 0.042B^{24})\varepsilon_t . \quad (3.52)$$

Das ist ein reines SARIMA-Modell vom Typ $[0, 2, 2] \times [0, 1, 2]_{12}$.

SARIMA-Modelle bilden eine elegante Möglichkeit, saisonale Abhängigkeiten in die Klasse der ARIMA-Modelle aufzunehmen. Wie aber die Gleichung (3.50) zeigt, hängen die Vorzeichen der Koeffizienten für das Lag $s + 1$ von den beiden Koeffizienten ϕ und Φ bzw. ψ und Ψ ab. Stimmt das aber im Anwendungsfall nicht mit den tatsächlichen Verhältnissen überein, und das kommt durchaus häufig vor, so ist das Subset-ARIMA-Modell das geeignete.

Beispiel 3.21 (Verkehrsunfälle – Fortsetzung von Seite 40)

Es wird die transformierte Reihe $y_t = \ln(\text{Unfall}_t)$ für den Zeitraum bis Ende 1990 untersucht. Ab 1991 liegen die Werte ja nur für die Gesamtzahlen vor, bei denen die neuen Bundesländer eingeschlossen sind. Die in der Abb. 2.15 wiedergegebene Reihe UNFALL weist eine starke Saisonalität, jedoch kaum einen Trend auf. Die Tafel der variaten Differenzen für y_t legt dementsprechend nahe, nur einmal saisonale Differenzen zu bilden, um Stationarität zu erreichen. Die in der Abbil-

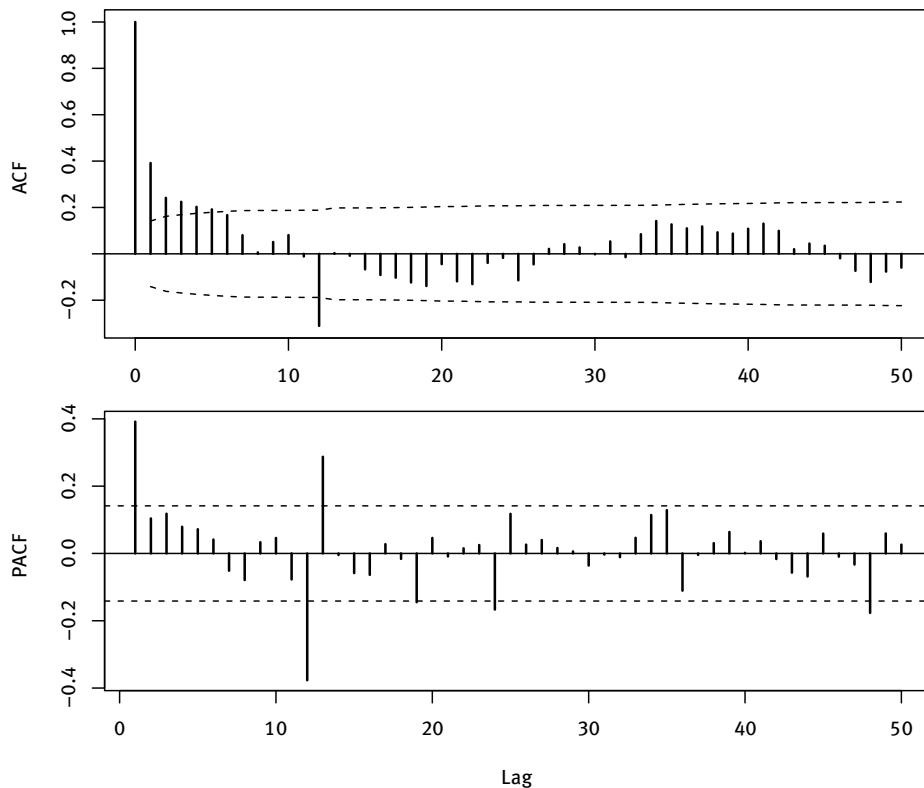


Abb. 3.16: ACF und PACF der saisonalen Differenzen von $\log(\text{UNFALL})$ (bis 1990)

dung 3.16 wiedergegebenen ACF und PACF von $\Delta_{12}y_t$ zeigen, dass die Abhangigkeit im kurzfristigen Bereich wohl durch einen AR[1]-Term erfasst werden kann. Die saisonale weist auf einen MA-Term oder AR-Term hin.

In der weiteren Analyse zeigen sich neue Abhangigkeiten bei hoheren Lags. Insbesondere scheint das Lag 11 eine besondere Rolle zu spielen. Das zeigen auch die beiden Kennfunktionen. Das abschlieende Modell ist dann

$$(1 - \phi_1 B - \phi_{11} B^{11})(1 - B^{12})Y_t = (1 - \Psi B^{12})\varepsilon_t .$$

Die ML-Schatzungen ergeben die folgenden Resultate:

Coefficients:

ar1	ar11	sma1
0.5667	0.2559	-0.9370
s.e.	0.0570	0.0641
sigma^2	estimated as 0.004315:	log likelihood=239.69 ,
		aic=-471.38

Die Koeffizienten sind alle relevant. Die P -Werte des Ljung-Box-Pierce-Tests Q sind in Ordnung (Die Anzahl der in Q berücksichtigten Lags muss die Ordnung übersteigen!):

lags	df	statistic	p-value
18	15	13.120	0.593
24	21	17.643	0.671
30	27	20.272	0.819
36	33	25.855	0.807

R-Code:

```

1 data(ACCIDENT)
2 y <- window(ACCIDENT,start=c(1974,1),end=c(1990,12))
3 y <- log(y)
4 y1 <- as.vector(diff(y,12))
5 acfpacf(y1,50)
6 fix <- c(NA,rep(0,9),NA,NA)
7 out <- arima(y,order=c(11,0,0),seasonal=list(order=c(0,1,1),
8     period=12),fixed=fix)
9 LjungBoxPierceTest(out$resid,3,36)

```

SARIMA-Modelle werden mittels der Funktion `arima` über optionale Argumente wie in den Zeilen 7 und 8 geschätzt.

Konstante Terme in ARIMA-Modellen

Die Standardannahme bei der Anpassung eines ARIMA[p, d, q]-Modells ist, dass $(1 - B)^d Y_t$ ein stationärer, bei null zentrierter Prozess ist. Sofern die Reihe nicht differenziert werden muss, wird sie daher oft mittelwertbereinigt. Wie aber im Abschnitt 3.1 ausgeführt wurde, hängt der Erwartungswert eines linearen Prozesses mit den Parametern zusammen. Für ARMA[p, q]-Prozesse gilt bei $E(Y_t) \neq 0$:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q};$$

das mündet in folgende Relation:

$$\alpha_0 = \mu(1 - \alpha_1 - \cdots - \alpha_p).$$

Hieran zeigt sich, dass es sinnvoll sein kann, das Niveau simultan mit den Modellparametern zu schätzen. Umgesetzt wird dies nur im Rahmen der ML-Schätzung; bei den Kleinste-Quadrat-Methoden führt dagegen das Zielkriterium automatisch zu dem Reihenmittel als Niveauschätzer.

Hat eine Zeitreihe nach einfacher Differenzenbildung noch ein von null verschiedenes Niveau, so liegt in der ursprünglichen Reihe ein linearer deterministischer Trend vor, siehe die Diskussion im Anschluss an Definition 2.30. Wird dieser bei der Spezifikation eines ARIMA-Modells außer Acht gelassen, so gestaltet sich die Bestimmung der Modellordnung meist schwierig, da Korrelationen zu inhaltlich nicht erklärbaren Lags signifikant werden. Ein Beispiel dazu wird in Abschnitt 9.1 gegeben.

3.5 R-Funktionen

Simulieren lassen sich ARIMA-Modelle mit `arima.sim` aus dem Paket `stats`. Dies ist ein Standardpaket, d. h. `stats` braucht nicht gesondert installiert zu werden, es ist in der R-Basiskonfiguration automatisch enthalten. Mit der Angabe einer Differenzenordnung in `arima.sim` wird daraus ein entsprechend integrierter Prozess geschätzt. Die Funktion `SimulateGaussianAR` des Paketes `FitAR` simuliert ebenfalls null zentrierte Gaußsche AR-Prozesse. Will man stationäre AR-Modelle höherer Ordnung simulieren, so ist es i. d. R. nicht selbstverständlich, wie die Koeffizienten zu wählen sind. In dem Paket `tsapp` gibt es daher die Funktion `polymake`, die aus vorgegebenen Nullstellen die zugehörigen AR-Koeffizienten berechnet. Werden die Nullstellen betragsmäßig größer als Eins gewählt, ist das Modell stationär.

`ARMAacf` berechnet die theoretische ACF bzw. PACF eines ARMA-Modells. Die Koeffizienten der MA-Darstellung eines ARMA-Modells bekommt man mit `ARMAtoMA`.

Für Schätzung von AR-Modellen mit den verschiedenen Schätzmethoden gibt es die Funktion `ar`. Per Voreinstellung wird ein Modell angepasst, bei dem die beste Ordnung unter allen möglichen bis zur maximalen, angegebenen Obergrenze `order.max` gewählt wird. Die Auswahl geschieht mittels des AIC-Kriteriums. Wird dieses auf `FALSE` gesetzt, so erhält man ein volles `order.max`-Modell. Subset-AR-Modelle können mit der Funktion `arima` aus `stats` oder mit `FitAR` des Paketes `FitAR` angepasst werden. `arima` ist die Standardfunktion für das Anpassen von MA-, ARMA-, ARIMA- und SARIMA-Modellen. `FitARMA` aus dem gleichnamigen Paket schätzt ARIMA-Modelle mit einem schnellen ML-Algorithmus.

In R beruhen ARMA-Modelle auf eine etwas unterschiedliche Definition: die Koeffizienten β_1, \dots, β_q des MA-Teils werden hier mit Plusvorzeichen genommen:

$$Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}, \quad (3.53)$$

(vgl. Abschnitt 3.3). Sie müssen entsprechend interpretiert werden!

Die PACF erhält man mit der Funktion `acf`, indem das optionale Argument `type="partial"` angegeben wird. Sie kann auch mit `pacf` direkt aufgerufen werden. Die ESACF ist mit der Funktion `eacf` im Paket `TSA` realisierbar.

Zur Durchführung des Ljung-Box-Pierce-Tests gibt es die Funktion `Box.test`. Mit der Option `type="Ljung-Box"` erhält man den Ljung-Box-Pierce-Test. Die in dePaket `tsapp` enthaltene Funktion `LjungBoxPierceTest` berechnet die relevanten Werte des Tests für eine Suite von einbezogenen Autokorrelationen.

Das Paket `TSTutorial` ist ein Statistiklabor für die interaktive Modellanpassung nach dem Box-Jenkins-Ansatz und Prognose.

4 Differenzen- und Trendinstationarität

4.1 Instationaritätstypen und ihre Implikationen¹

In vielen Fällen ist die Vorstellung angebracht, dass eine makroökonomische Zeitreihe eine Summe von Komponenten mit unterschiedlichen Eigenschaften ist. Ein einfacher Ansatz wäre hier etwa, wenn saisonale Einflüsse beiseitegelassen werden:

$$Y_t = DT_t + Z_t ,$$

wobei (DT_t) ein deterministischer Trend ist, etwa $DT_t = \beta_0 + \beta_1 t$, und (Z_t) ein ARIMA-Prozess, $\alpha(B)Z_t = \theta(B)\varepsilon_t$. $\alpha(B)$ ist dabei das ausmultiplizierte Polynom in B , das sonst als $\alpha(B)(1 - B)^d$ geschrieben wird. Es ergeben sich unterschiedliche Situationen, je nachdem, ob das AR-Polynom Wurzeln auf dem Einheitskreis ($d > 0$) hat oder nur solche außerhalb des Einheitskreises ($d = 0$).

Liegen die Wurzeln von $\alpha(B)$ außerhalb des Einheitskreises, so ist (Z_t) stationär. Man spricht von einem *trendstationären Modell*. Bei einem solchen Modell schlängeln sich die Pfade von (Y_t) um die durch das Trendpolynom gegebene Niveaulinie.

Wenn mindestens eine Wurzel von $\alpha(B)$ auf dem Einheitskreis liegt, so ist (Z_t) nicht stationär. Stationarität der Zeitreihe (Y_t) lässt sich durch geeignetes Differenzenbildnen erreichen. Es liegt ein *differenzenstationäres Modell* vor. Sei für das Folgende die Situation dahingehend eingeschränkt, dass nur eine Wurzel auf dem Einheitskreis liegt. Dann ist $(1 - B)Z_t$ stationär und $(1 - B)Y_t$ ist stationär mit einem von Null verschiedenen Niveau β_1 . In diesem stochastischen Trendmodell pendelt kein Pfad von (Y_t) um irgendeine Niveaulinie. Die unterschiedlichen Verhaltensmuster der beiden Typen sind in der Abbildung 4.1 illustriert. Die dargestellten Reihen weisen zudem jeweils einen Schock auf. Dabei wird der Schock in der Weise eingebracht, dass der Innovationsprozess (ε_t) an einer Stelle einen extrem großen oder kleinen Wert hat. Die Auswirkungen von Schocks sind bei den beiden Modellen ganz unterschiedlich. Das aus der Abbildung 4.1 erkennbare Muster gilt allgemein: Die Wirkung ist bei einer trendstationären Reihe vorübergehend, bei einer differenzenstationären persistent.

Im Rahmen der Makroökonomie, speziell im Rahmen der Konjunkturforschung, führt dies zu inhaltlichen Interpretationen. Dazu wird etwa eine Rezession als ein solcher Schock aufgefasst. Ist die zugrunde liegende Reihe, beispielsweise das reale Bruttoinlandsprodukt zu konstanten Preisen, trendstationär, so kann argumentiert werden, dass die Rezession einen zeitlich begrenzten Rückgang darstellt, der beim nächsten Aufschwung wieder kompensiert wird. Bei Differenzenstationarität lässt sich eine derartig optimistische Haltung daraus nicht ableiten.

¹ Die Ausführungen in diesem Abschnitt basieren weitgehend auf Rinne & Specht (2002, Abschnitt 9.3).

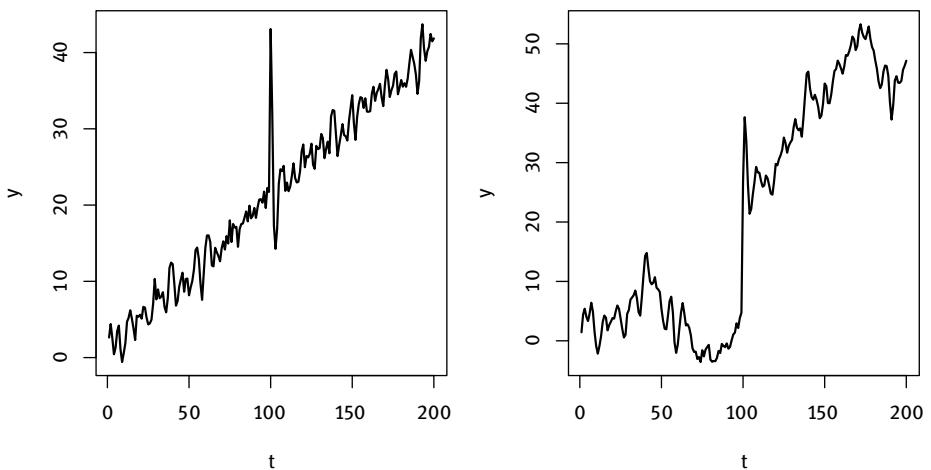


Abb. 4.1: Trendstationäre (links) und differenzenstationäre (rechts) Reihen jeweils mit einem Schock bei $t_0 = 100$

Für eine weitere Betrachtung im Rahmen der makroökonomischen Gegebenheiten wird die Komponente (Z_t) nochmals zerlegt in eine zyklische Komponente C_t und stochastische Trendkomponente (ST_t). (C_t) wird als mittelwertstationärer stochastischer Prozess unterstellt; (ST_t) enthält alle Zufallseinflüsse, die einen permanenten Effekt auf das Niveau von (Y_t) haben. Der gesamte Trend ist damit die Summe von deterministischem und stochastischem Trend, $DT_t + ST_t$.

Ein wichtiges Ziel der empirischen Makroökonomie besteht dann in der Isolierung der zyklischen Komponente durch Bilden der trendbereinigten Reihe $Y_t - (DT_t + ST_t)$. Bei einem trendstationären Modell ist die Zerlegung trivial. Wegen der Stationarität von (Z_t) ist $ST_t = 0$ und somit $Z_t = C_t$. Ist das Modell differenzenstationär, so gibt es die Zerlegung

$$Z_t = ST_t + C_t = k \cdot \sum_{u=1}^t \varepsilon_t + \psi(B)\varepsilon_t ,$$

bei der die Wachstumskomponente (ST_t) ein bei null zentrierter Random Walk ist. Die säkulare oder langfristige Entwicklung wird hier nicht nur durch den deterministischen Trend beschrieben, sondern zusätzlich durch einen instationären Prozess. Die zyklische Komponente $\psi(B)\varepsilon_t$ ist stationärer MA[∞]-Prozess; sie erfasst die Variation um den deterministischen Trend. Sie kann aber nicht nur der zyklischen Komponente (= Konjunkturkomponente) allein zugeschrieben werden.

Bei dem Stationarisieren solcher Prozesse ist bei einem deterministischen Trendmodell die Regressionsschätzung der deterministischen Komponente angesagt. Die Differenz zwischen der beobachteten Zeitreihe und der Regressionsfunktion liefert stationäre Residuen. Beim stochastischen Trendmodell gelangt man über Differenzen-

bildung zu einer stationären Zeitreihe. Es soll nun betrachtet werden, welche Folgen eine Fehlspezifikation des Trendtyps und die nicht angemessene Wahl der Stationarisierung hat.

Sei die Zeitreihe (Y_t) zunächst differenzenstationär. Die Schätzung der Konjunkturkomponente als Abweichung zwischen den beobachteten Zeitreihenwerten und denen eines (falschlich unterstellten) deterministischen Trends führt in der Regel zur Überschätzung der Höhe und Dauer der zyklischen Komponente und Unterschätzung der Wachstumskomponente.

Die Folgen einer Fehlspezifikation dieses Trendtyps ergeben sich aus dem Verhalten der Kleinst-Quadrate-Schätzer. Im Modell

$$Y_t = y_0 + \delta t + \sum_{u=1}^t \varepsilon_u$$

ergab sich durch Simulationen für mittels der KQ-Methode geschätzte lineare Trendfunktionen, siehe Chan et al. (1977) und Nelson & Kang (1981, 1984):

1. R^2 liegt bei 0.44 wenn $\delta = 0$, egal wie lang die Reihe ist.
2. R^2 geht mit $t \rightarrow \infty$ gegen 1, wenn $\delta \neq 0$.
3. Residuen haben eine Varianz von $0.14 \cdot \sigma_\varepsilon^2$, egal wie groß t und δ sind.
4. Autokorrelationskoeffizient der Residuen liegt bei $1 - 1/t$.
5. t -Test auf Signifikanz von $\hat{\delta}$: $H_0: \delta = 0$ wird bei einem Signifikanzniveau von $\alpha = 0.05$ in 87 % der Fälle irrtümlich verworfen.

Sei die Zeitreihe (Y_t) trendstationär. Zur Vereinfachung wird das Modell

$$Y_t = \delta_0 + \delta_1 t + \varepsilon_t$$

unterstellt. (Y_t) werde fälschlich als differenzenstationär spezifiziert; dementsprechend werde also die Reihe der ersten Differenzen betrachtet:

$$(1 - B)Y_t = \delta_1 + (\varepsilon_t - \varepsilon_{t-1})$$

Hier ist $(1 - B)Y_t$ ein nicht invertierbarer MA-Prozess. Allerdings bleibt der Kleinst-Quadrate-Schätzer für δ_1 unverzerrt und approximativ t -verteilt. Die folgende Analyse wird also nicht gänzlich falsch.

Insgesamt ist festzuhalten, dass die zweite betrachtete Fehlspezifikation, Trendstationarität wird als Differenzenstationarität angesehen, keine heftigen Konsequenzen mit sich bringt. Bei der ersten sind die Auswirkungen wesentlich gravierender; die Schlussfolgerungen können grob falsch sein.

4.2 Einheitswurzeltests

Die Überlegungen des voranstehenden Abschnittes rechtfertigen die Bemühungen, das Vorliegen einer Einheitswurzel mit einem statistischen Test zu überprüfen. Den Ausgangspunkt bildet ein einfaches Modell, das eine deterministische und stochastische Trendkomponente als Möglichkeiten enthält:

$$(1 - \alpha B)(Y_t - \theta_0 - \theta_1 t) = \varepsilon_t . \quad (4.1)$$

Dabei ist wie üblich (ε_t) ein White-Noise-Prozess mit $E(\varepsilon_t) = 0$ und $E(\varepsilon_t^2) = \sigma_\varepsilon^2$.

Sofern $|\alpha| < 1$, wird durch (4.1) ein trendstationäres Modell formuliert. Mit

$$\mu = \theta_0(1 - \alpha) + \theta_1\alpha \quad \text{und} \quad \beta = \theta_1(1 - \alpha) \quad (4.2)$$

lässt sich (4.1) auch schreiben als

$$Y_t = \mu + \beta t + \alpha Y_{t-1} + \varepsilon_t . \quad (4.3)$$

In dieser Formulierung kommen die Relationen der Parameter untereinander nicht mehr zur Geltung. Könnten in (4.3) die Parameter frei variieren, so wäre dies ein Modell mit einem deterministischen quadratischen Trend.

Sei nun $\alpha = 1$ betrachtet. Das Modell ist differenzenstationär. Aus den Beziehungen (4.2) folgt $\mu = \theta_1$ und $\beta = 0$. Aus (4.3) wird damit

$$Y_t = \theta_1 + Y_{t-1} + \varepsilon_t . \quad (4.4)$$

(Y_t) ist daher ein Random-Walk und zwar einer mit Drift, sofern $\theta_1 \neq 0$.

Um nun die Differenzenstationarität mit einem Test zu überprüfen, wird als Testproblem formuliert:

$$H_0: \alpha = 1 \quad H_1: |\alpha| < 1$$

Dass H_0 als einfache Hypothese formuliert wird, resultiert daraus, dass $\alpha > 1$ a priori ausgeschlossen ist. Ein Wert größer als eins würde ja zu einem explodierenden Prozess führen. Weiterhin werden die Hypothesen nicht andersherum formuliert, da diese Hypothesenwahl die Bereitschaft ausdrückt, solange von einer Differenzenstationarität als Arbeitshypothese auszugehen, wie sich kein starker Hinweis für das Gegenteil finden lässt. Die Wahrscheinlichkeit für einen Fehler 2. Art, eine Beibehaltung der Nullhypothese, obwohl sie falsch ist, kann allerdings sehr groß sein. Von der statistischen Sicht her macht dies Tests auf Einheitswurzeln praktisch wertlos. Vom ökonometrischen Standpunkt lässt sich aber ins Feld führen, dass die fälschliche Beibehaltung der Nullhypothese nach den Überlegungen des voranstehenden Abschnittes die weniger gravierende Fehlentscheidung ist.

Der *Unit-Root-Test von Dickey und Fuller*, kurz *DF-Test*, verwendet eine Teststatistik aus einem Regressionsansatz, der sich aus (4.3) durch Subtraktion von Y_{t-1} auf beiden Seiten ergibt:

$$(1 - B)Y_t = \mu + \beta t + \pi Y_{t-1} + \varepsilon_t ; \quad (4.5)$$

dabei ist $\pi = \alpha - 1$.

Die Nullhypothese ist $H_0: \beta = \pi = 0$; aufgrund der Interdependenzen hängen ja diese beiden Parameter zusammen. Sie kann mit der üblichen F -Teststatistik geprüft werden. Da die Verteilung der Kleinstes-Quadrat-Schätzer unter der Nullhypothese keine Standardverteilungen sind, sind als kritische Werte jedoch die bei Dickey & Fuller (1981, table VI) angegebenen zu verwenden. Dickey und Fuller haben sie mittels aufwendiger Monte-Carlo-Simulationen bestimmt. Wird die Nullhypothese abgelehnt, so ist die Reihe als nicht differenzenstationär nachgewiesen. Es kann natürlich noch sein, dass $\theta_1 \neq 0$ gilt, die Reihe also einen deterministischen Trend hat. Dies lässt sich dann mit einem Standardvorgehen überprüfen, da nun von dem Modell (4.1) mit $\alpha < 1$ bzw. $\pi < 0$ ausgegangen werden kann. Wird bei diesem Testproblem $H'_0: \theta_1 = 0$ $H'_1: \theta_1 \neq 0$ die Nullhypothese beibehalten, ist ein stationäres Modell zu verwenden. Andernfalls ist die Reihe als trendstationär nachgewiesen. Von der Testlogik her ist nur die Entscheidung „trend- aber nicht differenzenstationär“ verlässlich. Die beiden Situationen „Differenzenstationarität“ und „Stationarität“ können nur als nicht abzulehnende Arbeitshypothesen angesehen werden.

Auch die t -Teststatistik für $H_0: \pi = 0$ ist mit den alternativen kritischen Werten zum Testen der Differenzenstationarität geeignet. Der F -Test ist aber vorzuziehen, da er die Interdependenzen der Parameter unter der Nullhypothese berücksichtigt.

Problematisch an dem einfachen DF-Test ist, dass weitere zeitliche Abhängigkeiten nicht berücksichtigt werden. Eine Erweiterung ist dann das Modell

$$(1 - \alpha B)(1 - \alpha_1 B - \cdots - \alpha_{p-1} B^{p-1})(Y_t - \theta_0 - \theta_1 t) = \varepsilon_t, \quad (4.6)$$

mit den gleichen Annahmen bzgl. (ε_t) wie bisher. Für das AR-Polynom der Ordnung $p - 1$ wird gefordert, dass alle Lösungen der charakteristischen Gleichung außerhalb des Einheitskreises liegen. Über eine Reparametrisierung konnte Fuller zeigen, dass die t -Teststatistik für das Testen von $H_0: \alpha = 1$ unter der Nullhypothese nicht von den anderen Parametern abhängt.

Die Reparametrisierung erfolgt konkret für $\theta_0 = \theta_1 = 0$ und $p = 3$ folgendermaßen:

$$\begin{aligned} (1 - \alpha B)(1 - \alpha_1 B - \cdots - \alpha_{p-1} B^{p-1})Y_t &= \varepsilon_t \\ \Rightarrow Y_t - \alpha_1 Y_{t-1} - \alpha_2 Y_{t-2} - \alpha_3 Y_{t-3} - \alpha Y_{t-1} + \alpha \alpha_1 Y_{t-2} + \alpha \alpha_2 Y_{t-3} + \alpha \alpha_3 Y_{t-4} &= \varepsilon_t \\ \Rightarrow Y_t = (\alpha_1 + \alpha)Y_{t-1} + (\alpha_2 - \alpha \alpha_1)Y_{t-2} + (\alpha_3 - \alpha \alpha_2)Y_{t-3} - \alpha \alpha_3 Y_{t-4} + \varepsilon_t & \\ \Rightarrow Y_t = (\alpha_1 + \alpha + \alpha_2 - \alpha \alpha_1 + \alpha_3 - \alpha \alpha_2 - \alpha \alpha_3)Y_{t-1} & \\ &\quad - (\alpha_2 - \alpha \alpha_1 + \alpha_3 - \alpha \alpha_2 - \alpha \alpha_3)(Y_{t-1} - Y_{t-2}) \\ &\quad - (\alpha_3 - \alpha \alpha_2 - \alpha \alpha_3)(Y_{t-2} - Y_{t-3}) + \alpha \alpha_3(Y_{t-3} - Y_{t-4}) + \varepsilon_t \\ \Rightarrow Y_t - Y_{t-1} = (-1 + \alpha_1 + \alpha_2 + \alpha_3 - \alpha(-1 + \alpha_1 + \alpha_2 + \alpha_3))Y_{t-1} & \\ &\quad - (\alpha_2 + \alpha_3 - \alpha(\alpha_1 + \alpha_2 + \alpha_3))(Y_{t-1} - Y_{t-2}) \\ &\quad - (\alpha_3 - \alpha(\alpha_2 + \alpha_3))(Y_{t-2} - Y_{t-3}) + \alpha \alpha_3(Y_{t-3} - Y_{t-4}) + \varepsilon_t \end{aligned}$$

Der Koeffizient von Y_{t-1} auf der rechten Seite der letzten Gleichung ist Null, wenn $\alpha = 1$ gilt.

Allgemein führt die Reparametrisierung zu

$$(1 - B)Y_t = DT_t \boldsymbol{\beta}_t + \pi Y_{t-1} + \sum_{u=1}^k a_u (1 - B) Y_{t-u} + \epsilon_t . \quad (4.7)$$

Dabei kann $DT_t = 0$, $DT_t = 1$ oder $DT_t = (1, t)$ sein; der Koeffizientenvektor ist geeignet zu interpretieren.

Für die Durchführung des *erweiterten Dickey-Fuller-Tests*, kurz *ADF-Tests*, wird zuerst die Ordnung der Autoregression bestimmt. Dazu wird meist eine maximale Ordnung k_{\max} gewählt und das Modell angepasst. Falls der Koeffizient des Regressor mit dem höchsten Lag nicht signifikant ist, wird er entfernt. Die üblichen *t*-Tests für die Koeffizienten sind dabei anwendbar. Dies wird iteriert, bis man bei einem signifikanten maximalen Lag gelandet ist. Dann kann der *F*- bzw. der *t*-Test in der gleichen Weise wie der einfache DF-Test durchgeführt werden.

Beispiel 4.1 (Bruttoinlandsprodukt)

Es wird die saison- und kalenderbereinigte Zeitreihe des preisbereinigten Bruttoinlandsproduktes (2000=100) im Zeitraum vom ersten Quartal 1970 bis zum vierten Quartal 2007 betrachtet. Die Zeitreihe wird mit BIP bezeichnet.

Die Darstellung der ACF und PACF der differenzengebildeten Reihe legt eine Autokorrelation der Ordnung vier nahe. (Die erhöhten Werte bei den höheren Lags werden ignoriert. Auch führen sie nicht zu einer Signifikanz beim Ljung-Box-Pierce-Test, wenn ein AR[4]-Modell an die differenzengebildete Reihe angepasst wird.)

Da die Differenzen noch ein von null verschiedenes Niveau aufweisen, wird der Dickey Fuller-Test für die Hypothese, dass das Modell gegeben ist mit $\beta_1 = 0$ und $\pi = 0$, ausgewählt:

$$\Delta Y_t = \beta_1 + \pi Y_{t-1} + \sum_{u=1}^p \gamma_u \Delta Y_{t-u}$$

Die Anwendung des Tests in der Variante des *F*-Tests führt dann zu einem Wert der Teststatistik von 5.69. Die zugehörigen kritischen Werte von Dickey und Fuller sind: $c_{0.01} = 6.52$, $c_{0.05} = 4.63$, $c_{0.10} = 3.81$. Damit wird die Nullhypothese zum Niveau $\alpha = 0.05$ abgelehnt. Die Zeitreihe ist als differenzenstationär anzusehen. Jetzt könnte die Signifikanz des linearen Trends mit einem Standardverfahren getestet werden. Dazu siehe den Abschnitt 9.1.

R-Code:

```

1 data(GDP)
2 bip <- GDP[, 3]
3 y <- window(bip, start=c(1970, 1), end=c(2007, 4))
4 plot(y)
5 acfpacf(diff(y), 40)
6 mean(diff(y))

```

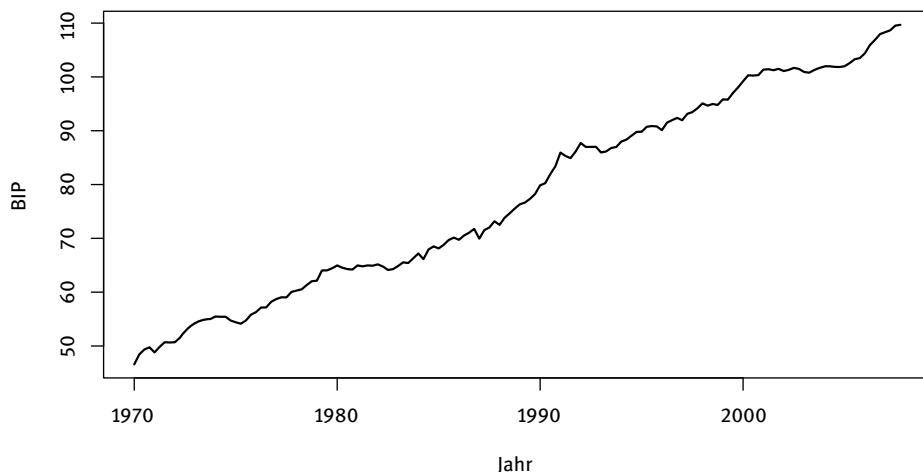


Abb. 4.2: Bruttoinlandsprodukt

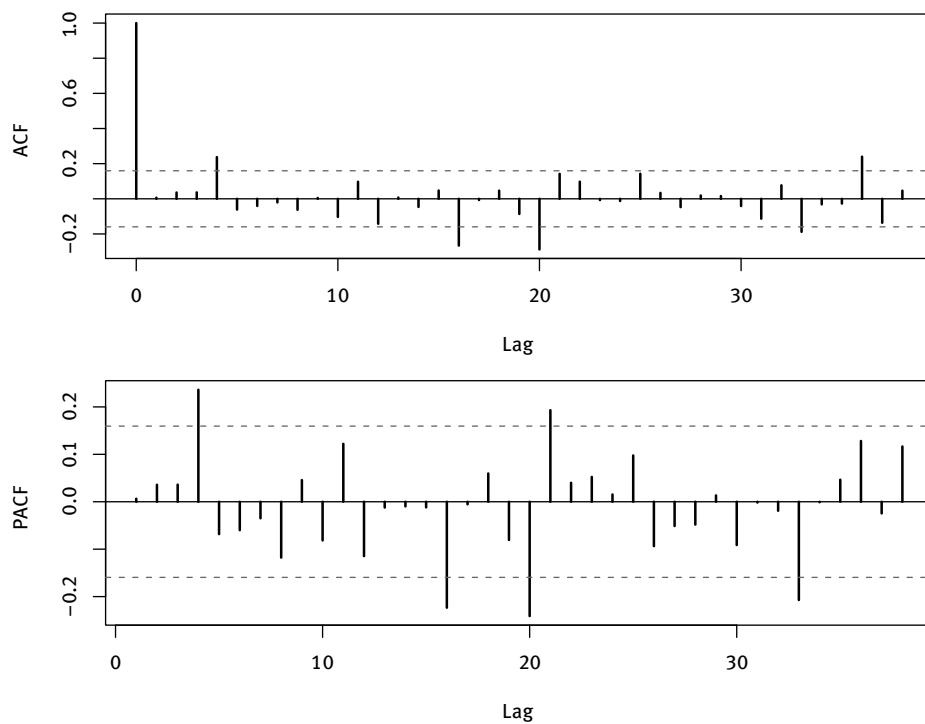


Abb. 4.3: ACF und PACF der differenzengebildeten Reihe BIP

```

7 library(urca)
8 out<-ur.df(y, type="drift", lags=4, selectlags="Fixed")
9 summary(out)

```

Es wird das Paket `urca` geladen, da es mit der Funktion `ur.df` auch den *F*-Test zur Verfügung stellt. Die Funktion gibt bei der gewählten Variante zwei Tests aus. Der relevante *F*-Test ist hier der unter `phi1` angegebene.

Es gibt etliche alternative Ansätze, um das Vorliegen einer Einheitswurzel zu testen, speziell den Phillips-Perron-Test (1988), den HEGY-Test (1990), das ADF-GLS-Verfahren (1996) und den KPSS-Test (1992), der die Stationarität als Nullhypothese formuliert. Dhrymes (1998) stellt den formalen Rahmen der Unit-Root-Tests dar. Ein weiterer Text dazu ist Hatanaka (1996).

4.3 R-Funktionen

Das Paket `tseries` stellt einige Funktionen zum Überprüfen der Nullhypothese bereit, dass die Zeitreihe eine Einheitswurzel aufweist. Hierzu gehören der erweiterte Dickey-Fuller-Test `adf.test`, der die *t*-Teststatistik verwendet, der Phillips-Perron-Test `pp.test`, und der KPSS-Test `kpss.test`. Alternative Implementierungen befinden sich im Paket `urca`. Es enthält noch weitere Tests auf Einheitswurzeln, z. B. das ADF-GLS-Verfahren `ur.ers`. Das Paket `fUnitRoots` ist ebenfalls einschlägig. Die Funktion `ndiffs` im Paket `forecast` ermittelt mit Hilfe von Einheitswurzeltests die Anzahl der benötigten einfachen Differenzen, um eine Zeitreihe stationär zu machen.

Spezielle Verfahren für das Vorliegen einer Einheitswurzel im Zusammenhang mit einer saisonalen Periodizität sind im Paket `uroot` enthalten. Explodierendes Verhalten wie z. B. eine Preisblase lässt sich mit einem erweiterten ADF-Test aus dem Paket `MultipleBubbles` ausschließen.

5 Prognosen mit univariaten Zeitreihenmodellen

Punktprognosen für nicht zu weit in der Zukunft liegende Zeitpunkte sind das zentrale Anliegen von Zeitreihenprognosen. Daneben ist aber auch höchst bedeutsam, wie zuverlässig solche Punktprognosen sind. Die Zuverlässigkeit hängt dabei von verschiedenen Einflüssen ab, vgl. Chatfield (1993):

- Die Unsicherheit bzgl. der Modellwahl. Möglicherweise ist ein anderes Modell korrekt, oder es ist keines der in Betracht gezogenen.
- Auch bei einem korrekten Modell führen die zukünftigen stochastischen Innovationen zu nicht vermeidbaren Unsicherheiten.
- Parameter sind bei den Modellen zu schätzen. Diese Schätzungen sind selbst mit Zufallsfehlern behaftet.

Diese Einflüsse müssten eigentlich alle berücksichtigt werden; allerdings ist dies kaum möglich. Nur der als zweites genannte Einfluss wird für die Angabe der Zuverlässigkeit verwendet. Zusätzlich wird i. d. R. noch die Normalverteilung der stochastischen Innovationen vorausgesetzt. Dann kann die Zuverlässigkeit mit *Prognoseintervallen* bestimmt werden. Das sind Intervalle, so dass die zukünftige Zeitreihenvariablen Y_{N+h} mit vorgegebener Wahrscheinlichkeit $1 - \alpha$ einen Wert aus dem zugehörigen Intervall annimmt. Meist werden die Grenzen der Intervalle für einzelne Prognosezeitpunkte angegeben, obwohl sie dann als Bänder in grafischen Darstellungen wiedergegeben werden. Auch wenn nicht gesagt werden kann, dass der zukünftige Pfad mit der Wahrscheinlichkeit $1 - \alpha$ in den Grenzen verbleibt, geben solche Bänder zumindest einen Eindruck der Zuverlässigkeit.

5.1 Verfahren der exponentiellen Glättung

Das exponentielle Glätten wurde bereits in Abschnitt 2.6 eingeführt. Damit wird im einfachsten Fall die Rekursion

$$\hat{y}_{N,1} = (1 - \alpha)\hat{y}_{N-1,1} + \alpha y_N \quad (5.1)$$

bezeichnet. Es gibt einige Aspekte, die auch beim einfachen exponentiellen Glätten erwähnenswert sind.

Der Name wird erklärlich, wenn (2.43) durch iteratives Einsetzen auf der rechten Seite bei Zulassen einer unendlichen Vergangenheit umgeschrieben wird:

$$\begin{aligned} \hat{y}_{N,1} &= (1 - \alpha)\hat{y}_{N-1,1} + \alpha y_N \\ &= (1 - \alpha)((1 - \alpha)\hat{y}_{N-2,1} + \alpha y_{N-1}) + \alpha y_N = (1 - \alpha)^2\hat{y}_{N-2,1} + (1 - \alpha)\alpha y_{N-1} + \alpha y_N \\ &= (1 - \alpha)^3\hat{y}_{N-3,1} + (1 - \alpha)^2\alpha y_{N-2} + (1 - \alpha)\alpha y_{N-1} + \alpha y_N = \dots \\ &= \alpha \sum_{u=-\infty}^N (1 - \alpha)^u y_{N-u}. \end{aligned} \quad (5.2)$$

Die Prognose lässt sich also als exponentiell gewichtete Summe vergangener Zeitreihenwerte darstellen.

Neben der rekursiven (2.43) und der expliziten Form (5.2) kann das exponentielle Glätten noch in der *Fehlerkorrektur-Darstellung* angegeben werden:

$$\hat{y}_{N,1} = \hat{y}_{N-1,1} + \alpha(y_N - \hat{y}_{N-1,1}). \quad (5.3)$$

Zudem gibt es einen Zusammenhang mit ARIMA-Modellen. Sei nämlich ein ARIMA-[0, 1, 1]-Prozess gegeben, $(1 - B)Y_t = (1 - \beta B)\varepsilon_t$. Dann ist $(1 - B)Y_t$ ein MA[1]-Prozess mit einer AR[∞]-Darstellung:

$$\begin{aligned}\varepsilon_t &= (1 - B)Y_t + \sum_{u=1}^{\infty} \beta^u (1 - B)Y_{t-u} = Y_t - Y_{t-1} + \sum_{u=1}^{\infty} (\beta^u Y_{t-u} - \beta^u Y_{t-u-1}) \\ &= Y_t - Y_{t-1} + \sum_{u=1}^{\infty} \beta^u Y_{t-u} - \sum_{u=1}^{\infty} \beta^u Y_{t-u-1} = Y_t + \beta \sum_{u=1}^{\infty} \beta^{u-1} Y_{t-u} - \sum_{u=0}^{\infty} \beta^u Y_{t-u-1} \\ &= Y_t + (\beta - 1) \sum_{u=1}^{\infty} \beta^{u-1} Y_{t-u}\end{aligned}$$

Umstellen dieser Gleichung ergibt:

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

Mit $1 - \beta = \alpha$ ist dies die mit einem Zufallsfehler ausgestattete Beziehung (5.2).

Wie schon die Angabe des Holt-Winter-Verfahrens in Abschnitt 2.6 deutlich macht, ist das exponentielle Glätten in verschiedener Weise verallgemeinert worden, um die Idee für Zeitreihen unterschiedlicher Struktur nutzbar zu machen. Hyndman et al. (2008) geben verschiedene Konstellationen für die Trend- und Saisonkomponente an. 15 sind in der folgenden Tabelle schematisch zusammengefasst. Durch die zusätzliche Unterscheidung, ob der Fehler additiv bzw. multiplikativ einwirkt, ergeben sich weitere Konstellationen.

Trend-Komponente		Saisonale Komponente		
		N (keine)	A (additiv)	M (multiplikativ)
N	(keine)	N,N	N,A	N,M
A	(additiv)	A,N	A,A	A,M
A _d	(additiv gedämpft)	A _d ,N	A _d ,A	A _d ,M
M	(multiplikativ)	M,N	M,A	M,M
M _d	(multiplikativ gedämpft)	M _d ,N	M _d ,A	M _d ,M

In diesen Kombinationen sind auch die bekannten Verfahren wiederzufinden. Das Modell N,N ist das einfache exponentielle Glätten, A,A das additive und A,M das multiplikative Holt-Winters-Verfahren.

Die Form des Einwirkens des Fehlers kann bei der Punktprognose vernachlässigt werden. Bei der Bestimmung von Prognoseintervallen ist die Unterscheidung jedoch wichtig, wie die Fehler in das Modell eingehen. Werden die Fehler als Gaußsche White-Noise-Prozesse unterstellt, so lassen sich für einige Modelle damit Prognoseintervalle aufstellen. Vor allem bei multiplikativer Verknüpfung von Trend und/oder Saison lassen sich die Verteilungen der Prognosefehler allerdings nicht berechnen. Dann sind die Prognoseintervalle mittels Simulationen zu bestimmen.

Beispiel 5.1 (Bierproduktion in Australien)

Für die Zeitreihe BIER der monatlichen Bierproduktion in Australien von Januar 1990 bis August 1995, veröffentlicht vom Australischen Bureau of Statistics, wird ein exponentielles Glättungsmodell angepasst und eine zugehörige Prognose für zwei Jahre erstellt. Das Modell ist vom Typ (M, A_d, M) . Der erste Buchstabe der Modellkomponente gibt dabei an, dass der Fehlerterm multiplikativ berücksichtigt wird. Die anderen beiden Buchstaben entsprechen den Eintragungen der oben angegebenen Tabelle. In der Abbildung sind zudem die 80 %- und 95 %-Prognoseintervalle dargestellt.

R-Code:

```
1 data(BEER)
2 y <- window(BEER, start=c(1990, 1))
3 library("forecast")
4 out <- ets(y, damped=T)
5 p <- predict(out, n.ahead=24)
6 plot(p)
```

Die Anpassung erfolgt mit der Funktion `ets` des Paketes `forecast`. Die Modellwahl wird dabei automatisch vollzogen.

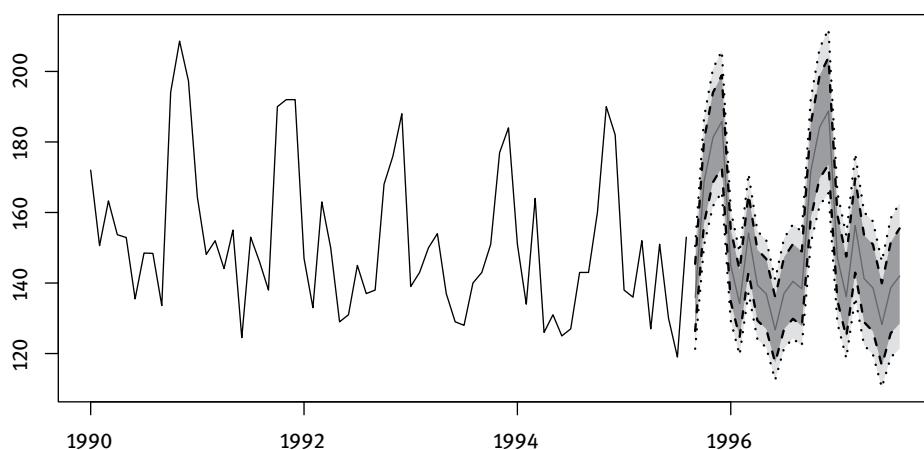


Abb. 5.1: Bierproduktion in Australien mit Prognosen auf Basis des Modells (M, A_d, M)

Die klassische Standardreferenz zum exponentiellen Glätten ist das Buch von Brown (1962). Gardner (1985) gibt einen Überblick über verschiedene Varianten des exponentiellen Glättens. Hyndman et al. (2008) ist eine neuere Darstellung, die auch Grundlage des oben verwendeten R-Paketes `forecast` ist.

5.2 Prognosen mit ARIMA-Modellen

Die Aufgabe der Zeitreihenprognose besteht darin, zukünftige Werte, d. h. Realisationen von Y_{N+h} , $h = 1, 2, \dots$, auf der Basis einer beobachteten Zeitreihe vorherzusagen. Das kann natürlich nur mit Hilfe geeigneter Modelle getan werden. Hat man sich für eine Klasse von Modellen, z. B. die ARIMA Modelle, entschieden, so wird man versuchen, die im Rahmen dieser Modelle optimale Prognose zu bestimmen. Optimal ist dabei eine Prognose $\hat{Y}_{t,1}$, bei der ein möglichst kleiner Prognosefehler zu erwarten ist:

$$E(\hat{Y}_{t,1} - Y_{t+1})^2 \stackrel{!}{=} \min. \quad (5.4)$$

Ein White-Noise-Prozess (ε_t) mit $E(\varepsilon_t) = 0$ wird unter diesem Gesichtspunkt am besten durch den Wert Null prognostiziert. Das reflektiert die Tatsache, dass in den vorhandenen Beobachtungen keine Information steckt, die die Abweichung der zukünftigen Variablen vom Erwartungswert beeinflussen kann. Keine Information bedeutet statistisch gesehen, dass die Abweichung der Prognosen von den zu prognostizierenden Variablen unkorreliert ist mit den Variablen, mit denen die Prognose bestimmt wird. (Natürlich darf auch keine andere Form von Abhängigkeit vorhanden sein.) Diese Eigenschaft ist für die Optimalität einer linearen Prognose entscheidend.

Satz 5.2 (Charakterisierung der optimalen linearen Prognose)

(Y_t) sei ein stationärer Prozess. Eine lineare Prognose $\hat{Y}_{t,h} = \sum_{u=0}^n c_u Y_{t-u}$ hat einen minimalen mittleren quadratischen Fehler, wenn der Prognosefehler $\hat{Y}_{t,h} - Y_{t+h}$ und die in die Prognose eingehenden Zufallsvariablen unkorreliert sind.

Die Definitionsgleichung der AR-Prozesse legt nahe, dass bei bekannten Parametern zukünftige Werte einfach prognostiziert werden, indem die Definitionsgleichung fortgeschrieben wird. Zukünftige Störungen sind dabei durch ihren Erwartungswert, also durch Null, zu ersetzen. Durch dieses Vorgehen wird für $h = 1$ die beste lineare Prognose bestimmt, sind doch die zukünftigen Innovationen $\varepsilon_{N+1}, \varepsilon_{N+2}, \dots$ unkorreliert mit Y_N, Y_{N-1}, \dots . Die Fortsetzung der Rekursion für $h = 2, 3, \dots$ unter Verwendung der bis dahin bestimmten Prognosen ergibt insgesamt die optimalen linearen Prognosen.

Die gleiche Vorgehensweise führt auch bei MA- und ARMA-Modellen zu optimalen Prognosen; dabei sind die Modellresiduen $\hat{\varepsilon}_t$ für die nicht beobachtbaren Innovationen ε_t zu verwenden. Die Ein-Schritt-Prognose eines ARMA[p, q]-Modells ist

$$\hat{Y}_{N,1} = \hat{\alpha}_1 Y_N + \dots + \hat{\alpha}_p Y_{N+1-p} - \hat{\beta}_1 \hat{\varepsilon}_N - \dots - \hat{\beta}_q \hat{\varepsilon}_{N+1-q}.$$

Für die h -Schritt Prognose ergibt sich, wenn $\hat{Y}_{N,v} = Y_{N+v}$ für $v \leq 0$ sowie $\hat{\varepsilon}_{N,v} = 0$ für $v > 0$ und $= \hat{\varepsilon}_{N+v}$ für $v \leq 0$ gesetzt werden:

$$\hat{Y}_{N,h} = \hat{\alpha}_1 \hat{Y}_{N,h-1} + \hat{\alpha}_2 \hat{Y}_{N,h-2} + \cdots + \hat{\alpha}_p \hat{Y}_{N,h-p} - \hat{\beta}_1 \hat{\varepsilon}_{N,h-1} - \cdots - \hat{\beta}_q \hat{\varepsilon}_{N,h-q}. \quad (5.5)$$

Dieses rekursive Verfahren ist als *Box-Jenkins-Prognose* bekannt.

Die Optimalität der Prognosen bei AR-, MA- und ARMA-Modellen gilt streng genommen nur bei Verwendung der tatsächlichen Parameterwerte und Innovationen. Die Eigenschaften der Schätzer gibt aber zumindest asymptotisch die Berechtigung, auch bei Verwendung der Schätzwerte und empirischen Residuen von optimalen Prognosen zu sprechen.

Mehrschrittprognosen gehen bei stationären ARMA-Prozessen mit wachsendem Prognosehorizont gegen den Erwartungswert. Dies ist leicht zu sehen, wenn man für einen zentrierten Prozess die MA[∞]-Darstellung $Y_t = -\sum_{u=0}^{\infty} c_u \varepsilon_{t-u}$ mit $c_0 = -1$, verwendet. Hier gilt:

$$\hat{Y}_{N,h} = - \sum_{u=h}^{\infty} c_u \hat{\varepsilon}_{N+h-u}. \quad (5.6)$$

Da die Folge der Koeffizienten gegen Null geht, werden auch die Prognosen betragsmäßig immer kleiner.

Bei stationären Prozessen wird auch die Varianz des Prognosefehlers,

$$\text{Var}(\hat{Y}_{N,h} - Y_{N+h}) = \sigma_{\varepsilon}^2 \sum_{u=0}^{h-1} c_u^2, \quad (5.7)$$

bei wachsendem h nach einer Anfangsphase stabil. Sie geht für $h \rightarrow \infty$ gegen die Prozessvarianz.

Bei Unterstellung eines Gauß-Prozesses kann man auch Prognoseintervalle bestimmen, Intervalle, in denen mit einer vorgegebenen Wahrscheinlichkeit $1 - \alpha$ der zukünftige Wert der Zeitreihenvariablen zu liegen kommen wird. Diese haben die Form

$$P(\hat{Y}_{N,h} - z_{1-\alpha/2} \sqrt{V} \leq Y_{N+h} \leq \hat{Y}_{N,h} + z_{1-\alpha/2} \sqrt{V}) = 1 - \alpha. \quad (5.8)$$

Dabei ist V die Varianz des Prognosefehlers, und $z_{1-\alpha/2}$ ist das entsprechende Quantil der Standardnormalverteilung. Die Gültigkeit dieser Prognoseintervalle hängt dementsprechend von der Voraussetzung normalverteilter Innovationen ab.

I. d. R. werden Prognosen für mehrere Zeitpunkte berechnet und dann als Verlängerung des beobachteten Abschnittes der Zeitreihe gezeichnet. Die Prognoseintervalle – meist mit $100\alpha = 75\%$ oder $= 95\%$ – ergeben dann *Prognosebänder*. Diese sind aber mit Vorsicht zu betrachten, da die Prognoseintervalle ja nur für einzelne Zeitpunkte gelten.

Werden Prognosen für Zeitreihen erstellt, für die ein ARIMA-Modell mit $d \geq 1$ angepasst wurde, so ergibt die Rücktransformation (zum Differenzenbildern) Prognosen für die ursprüngliche Zeitreihe. Die Prognoseintervalle öffnen sich dann mit zunehmendem Prognosehorizont. Es können aber auch einfach geeignete Verallgemeinerungen von (5.5) und (5.7) ausgenutzt werden, um die Prognosen und Prognoseintervalle direkt zu bestimmen.

Beispiel 5.3 (Australischer Rotwein)

Es soll der monatliche Verkauf von australischem Rotwein (in 1000 l) prognostiziert werden. Als Basis liegt die entsprechende Zeitreihe ROTWEIN von Januar 1980 bis Juli 1995 vor.

Zunächst weist die Reihe ROTWEIN (vgl. Abbildung 5.2) einen Trend und eine Saison auf. Die Ausschläge werden dabei mit ansteigendem Niveau größer. Um die Varianz zu stabilisieren, wird im ersten Schritt eine Box-Cox-Transformation vorgenommen. Das Box-Cox-Diagramm führt eigentlich auf die Transformation mit der Potenz $\lambda = 0.1$. Das rechtfertigt die einfachere logarithmische Transformation. Die Methode der variaten Differenzen, angewendet auf die Reihe $\ln(y_t)$ legt die einmalige saisonale Differenzenbildung nahe.

Die Werte von ACF und PACF der saisonal differenzierten Reihe deuten für den nicht-saisonalen Teil auf einen AR[1]-Teil hin, für den saisonalen ist ein MA-Term angebracht, da die ACF eher als abbrechend zu bezeichnen ist, vergleiche die Abbildung 5.3. Bei der weiteren Analyse erweist sich, dass für den ersten Teil sogar ein gemischtes Modell adäquat ist. Als Modell ergibt sich schließlich:

$$(1 - 0.9975B)Z_t = (1 + 0.8230B)(1 + 0.7436B^{12})\varepsilon_t .$$

(0.0037)	(0.0458)	(0.0651)
----------	----------	----------

In den Klammern unterhalb der Koeffizienten sind die Standardfehler angegeben. Sie sind alle so klein, dass die Relevanz der Koeffizienten schon daraus deutlich wird. Der Ljung-Box-Pierce-Test zeigt keine Signifikanz; die P -Werte sind alle größer als 0.3.

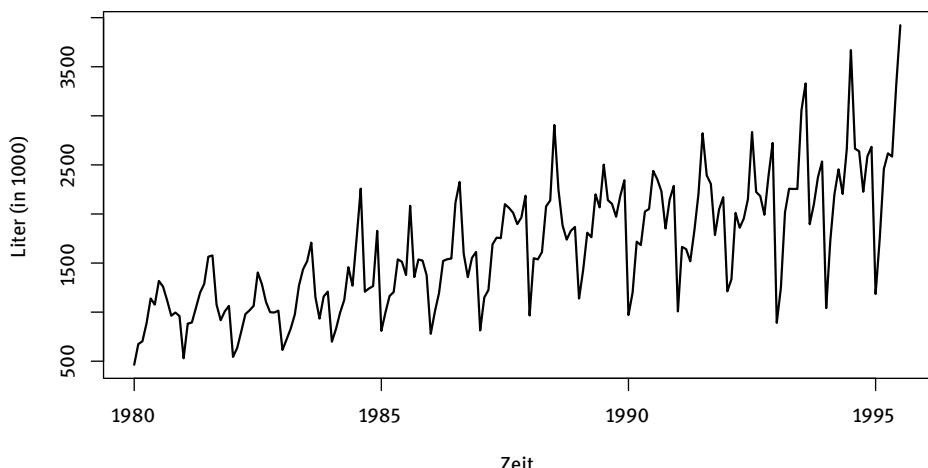


Abb. 5.2: Verkäufe von australischem ROTWEIN

Für die Prognose bis zu 24 Monate in die Zukunft wird nun das geschätzte Modell eingesetzt. Um die Prognosen auf der ursprünglichen Skala zu haben, werden die Prognosen wie auch die unteren und oberen Grenzen der Prognoseintervalle antilogarithmisch transformiert.

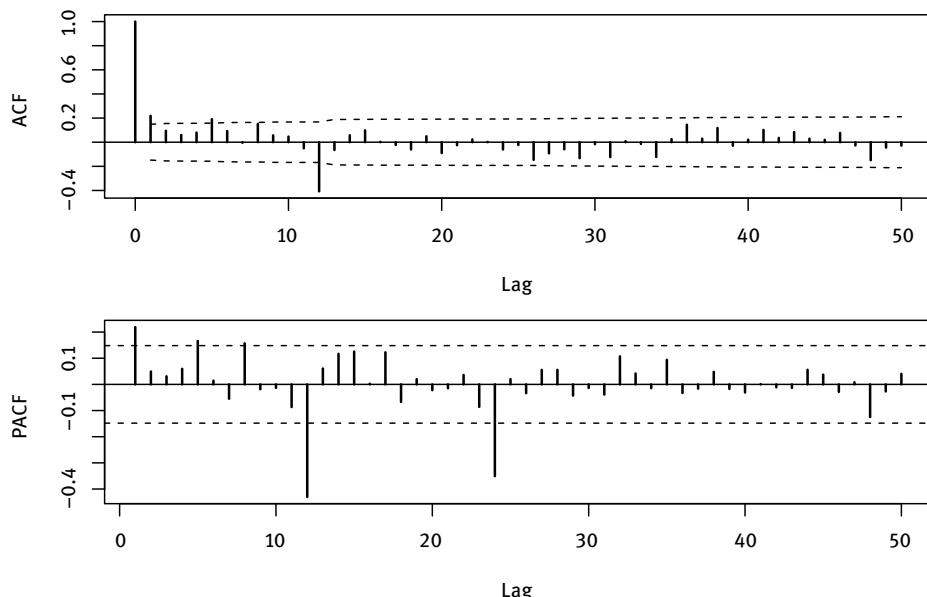


Abb. 5.3: ACF und PACF für $z_t = (1 - B^{12}) \ln(y_t)$

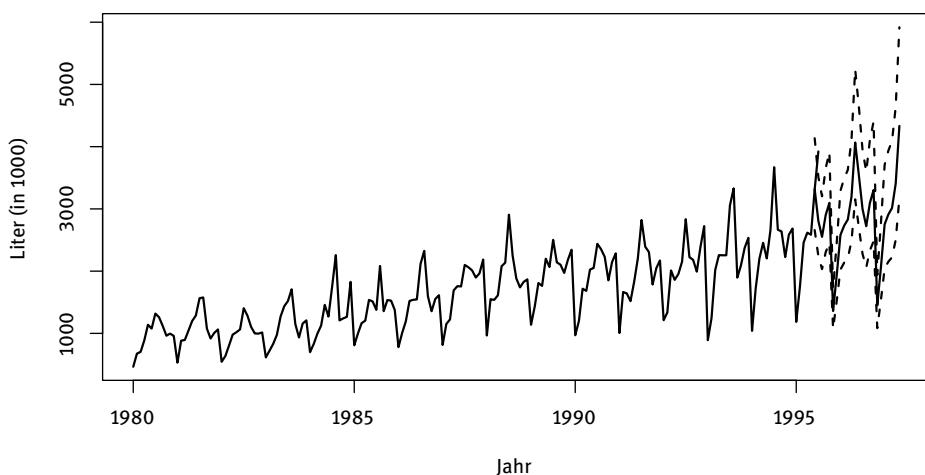


Abb. 5.4: Verkäufe von Rotwein mit Prognosen und 95 %-Prognosegrenzen

R-Code:

```

1 data(REDWINE)
2 plot(REDWINE)
3 BoxCox(REDWINE,10)
4 y <- log(REDWINE)
5 vartable(y,12)
6 acfpacf(diff(y,12),50)
7 out <- arima(y,order=c(1,0,1),seasonal=list(order=c(0,1,1),
8     period=12))
9 LjungBoxPierceTest(out$residuals,4,50)
10 p <- predict(out,n.ahead=24)
11 new.dat.1 <- ts(exp(p$pred),start=c(1995,6),frequency=12)
12 new.dat.2 <- ts(exp(p$pred-1.96*p$se),start=c(1995,6),frequency=12)
13 new.dat.3 <- ts(exp(p$pred+1.96*p$se),start=c(1995,6),frequency=12)
14 ts.plot(exp(y),new.dat.1,new.dat.2,new.dat.3, lty=c(1,1,2,2))

```

Für die grafische Darstellung werden die prognostizierte Werte und die Intervallgrenzen mit dem Zeitreihenattribut versehen (Zeilen 13–15). Mit der Exponentiation werden dabei die Daten zurücktransformiert. Der Wert 1.96 ist das 0.975-Quantil der Standardnormalverteilung.

Die optimale Prognose gemäß (5.4) kann auch direkt bestimmt werden. Ersetzen der Prognose durch die im Satz 5.2 angegebene Form und Umstellen der beiden Terme in der Klammer ergibt für eine Zeitreihe (y_t) die Optimierungsaufgabe

$$\sum_{t \geq n+1} \left[y_{t+1} - \sum_{u=0}^n c_u y_{t-u} \right]^2 \stackrel{!}{=} \min. \quad (5.9)$$

Im Abschnitt 3.1 haben wir die Parameter α_1, α_2 eines AR[2]-Prozesses mit der gleichen Überlegung geschätzt. Gleichung (3.10) ist nichts anderes als ein Spezialfall von (5.9) für den Wert $n = 1$. Das Yule-Walker-Verfahren legt nahe, die empirischen Autokovarianzen zu verwenden. Analog zu (3.13a), (3.13b) können wir die optimale lineare Prognose durch Lösen eines Systems mit $n + 1$ Gleichungen berechnen. Die Levinson-Durbin-Rekursion kommt hier wieder zum Einsatz. Mit diesem Verfahren können Prognosen für verschiedene Prozesse berechnet werden. Im Gegensatz dazu ist die Box-Jenkins-Prognose speziell auf die ARIMA-Modelle zugeschnitten. Die Funktion TrenchForecast aus dem Paket 1tsa verwendet eine nach Trench verbesserte Variante der Levinson-Durbin-Rekursion.

5.3 Trendextrapolation mit ARMA-Störungen

Die Prognosen mit ARIMA-Modellen unterstellen Stationarität oder stochastische Trends. Nicht selten liegen aber auch deterministische Trendfunktionen zu Grunde. Kann man von unkorrelierten Störungen ausgehen, so führt die Prognoseaufgabe einfach zur Regression und Extrapolation der Regressionsfunktion. Bei autokorrelierten Störungen ist eine Modellierung für die Fehler wesentlich. Dies gilt nicht nur für die Punktprognosen bei kürzeren Prognosehorizonten, sondern auch für etwaige Prognoseintervalle. Hier wird nur auf diesen Aspekt eingegangen; die Anpassung wird im Abschnitt 9.1 dargestellt.

Sei zur Illustration der Bedeutung der Fehlermodellierung die Trendfunktion ein einfacher linearer Trend, aber die Residuen sollen einem ARMA-Modell folgen:

$$Y_t = \theta_0 + \theta_1 \cdot t + U_t, \quad \text{mit} \quad \alpha(B)U_t = \beta(B)\eta_t$$

Unterstellt, dass das zugehörige Modell identifiziert und die Parameter geschätzt wurden, führt die Bestimmung der Prognosen $\hat{u}_{N,h}$ für die Fehler

$$\hat{u}_t = y_t - (\hat{\beta}_0 + \hat{\beta}_1 \cdot t)$$

zu:

$$\hat{y}_{N,1} = \hat{\beta}_0 + \hat{\beta}_1 \cdot (N + 1) + \hat{u}_{N,1}$$

$$\hat{y}_{N,2} = \hat{\beta}_0 + \hat{\beta}_1 \cdot (N + 2) + \hat{u}_{N,2}.$$

Diese beiden Zeilen lassen die weiteren Schritte erkennen.

Beispiel 5.4 (Stromspitzenbedarf – Fortsetzung von Seite 59)

Die Prognose des Stromspitzenbedarfs mittels einer Trendgeraden ist an sich problematisch. Wenn sie betrieben wird, sollte wenigstens die zeitliche Korrelation der Residuen berücksichtigt werden. Um ein geeignetes Modell zu identifizieren, werden die Residuen der einfachen KQ-Anpassung auf Autokorrelationen gecheckt. Die ACF und die PACF zeigen, dass ein AR[1]-Modell ausreichen sollte, um die Autokorrelation zu erfassen. Die Schätzung des Modells

$$Y_t = \beta_0 + \beta_2 t + e_t \quad \text{mit} \quad e_t - \alpha e_{t-1} = \varepsilon_t$$

ergibt:

	ar1	intercept	t1
	0.6301	-97547.897	50.1115
s.e.	0.1553	5220.056	2.6551

Die Residuen zeigen keine Autokorrelationen mehr. Daher wird dieses Modell nun verwendet, um den Stromspitzenbedarf in die Zukunft fortzuschreiben.

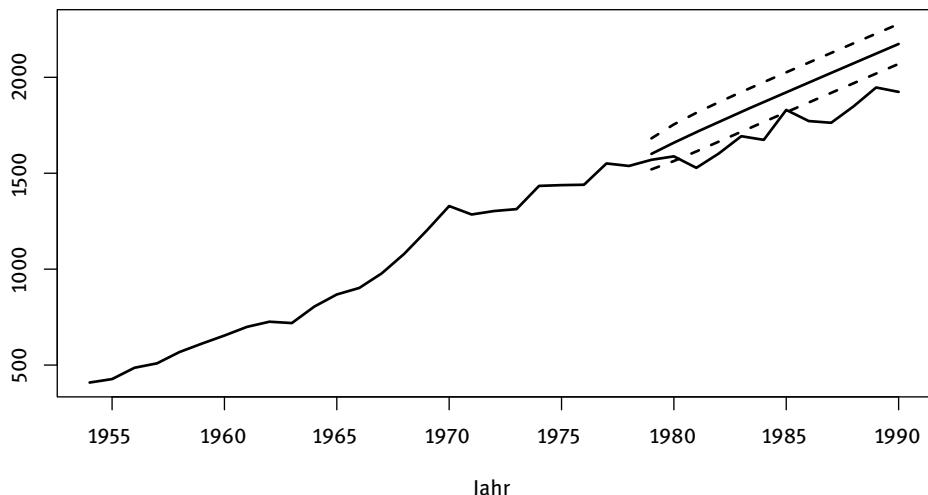


Abb. 5.5: Stromspitzenbedarf mit Prognosen und 95 %-Prognosegrenzen

Das Bild hat sich gegenüber Abbildung 2.23 qualitativ natürlich nicht geändert. Die Prognoseintervalle sind allerdings ein wenig breiter geworden.

R-Code:

```

1 data(PPDEMAND)
2 y <- PPDEMAND
3 y1 <- window(y,start=1954,end=1978)
4 t1<- 1954 + c(1:length(y1))
5 out <- lm(y1~t1[1:25])
6 acfpacf(out$residuals,10)
7 out <- arima(y1,order=c(1,0,0),xreg=t1)
8 out
9 p <- predict(out,n.ahead=12,newxreg = c(1979:1990),se.fit=TRUE)
10 new.dat.1 <- ts(p$pred,start=1979,frequency=1)
11 new.dat.2 <- ts(p$pred-1.96*p$se,start=1979,frequency=1)
12 new.dat.3 <- ts(p$pred+1.96*p$se,start=1979,frequency=1)
13 par(mar=c(4,4,1,1))
14 ts.plot(y,new.dat.1,new.dat.2,new.dat.3,lty=c(1,1,2,2))

```

Im Vergleich zu dem R-Code des Beispiels 5.3 kommt nichts wesentlich Neues hinzu. Für die Prognose muss der Trend über das Argument `newxreg` in die Zukunft fortgesetzt werden. Über die Option `se.fit=TRUE` werden die Standardabweichungen der Punktprognosen angefordert.

5.4 Zur Auswahl eines Prognosemodells

Anhand einiger Beispiele soll diskutiert werden, welches Prognoseverfahren ggf. bei einer konkreten Zeitreihe eingesetzt werden sollte. Die Leserin bzw. der Leser ist aber aufgefordert, erst einmal selbst diese Frage jeweils anhand der Abbildung zu beantworten.

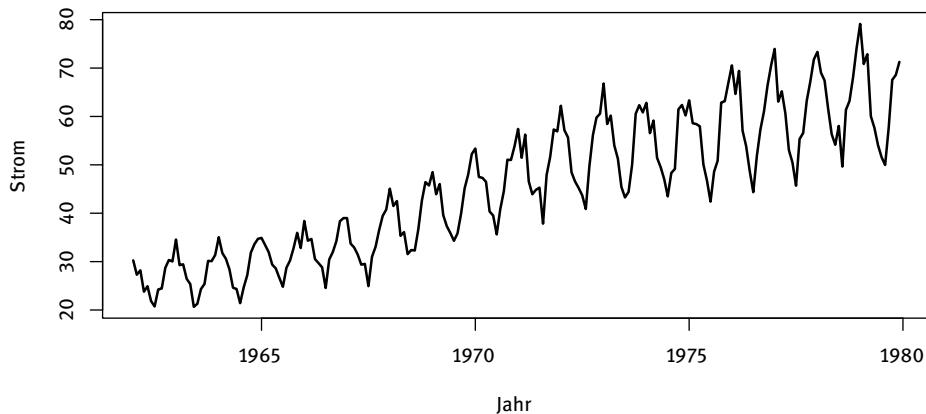


Abb. 5.6: Monatlicher Stromverbrauch der Stadt Braunschweig, Jan. 1962–Dez. 1979

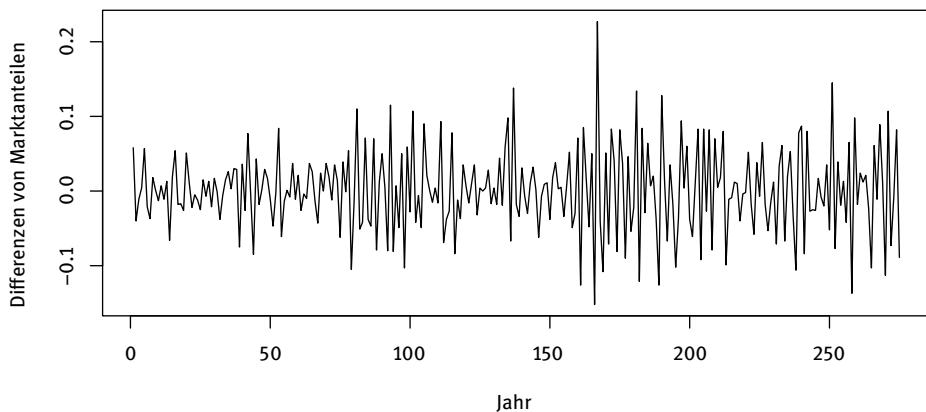


Abb. 5.7: Änderungen des wöchentlichen Marktanteils der Zahnpasta Crest, Januar 1958 bis April 1963

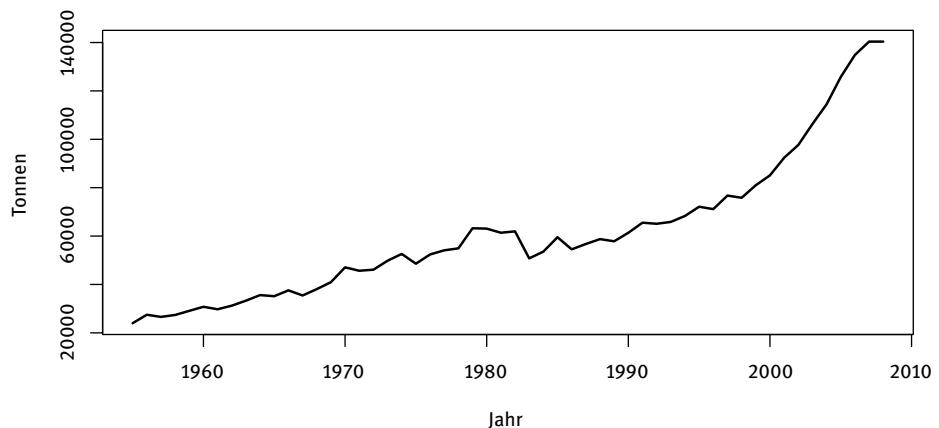


Abb. 5.8: Seegüterumschlag im Hamburger Hafen seit 1955 (gesamt, in 1000 t)

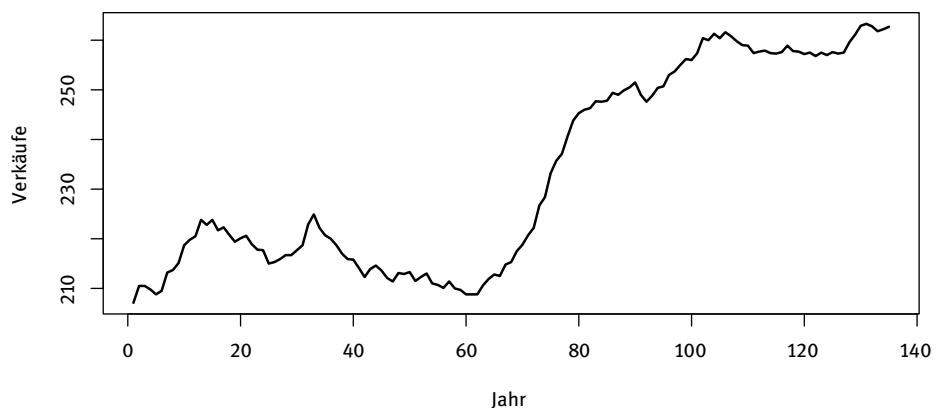


Abb. 5.9: Sales-Reihe aus Box-Jenkins, Serie M



Abb. 5.10: Autoverkäufe in den USA

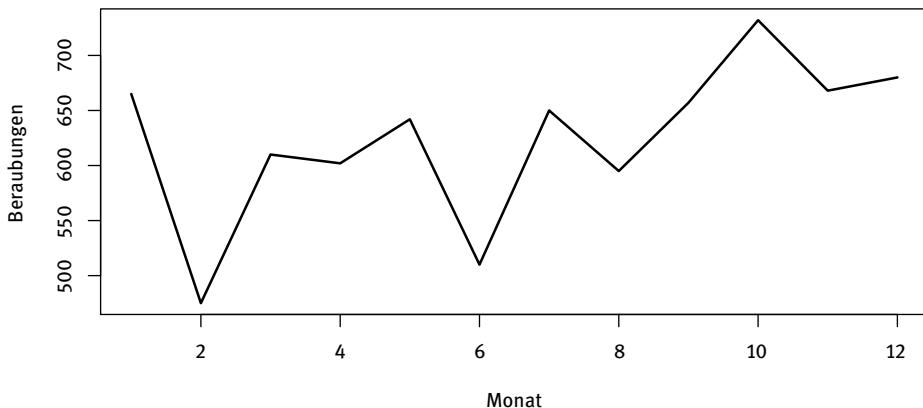


Abb. 5.11: Beraubungen 2008 pro Monat im Staat Virginia

Der in der Abbildung 5.6 dargestellte monatliche Stromverbrauch der Stadt Braunschweig weist eine starke Saisonalität und ein Anwachsen der Ausschläge mit steigendem Niveau auf. Das deutet auf ein multiplikatives Komponentenmodell hin. Die multiplikative Variante des Holt-Winters-Verfahrens wäre dementsprechend eine geeignete Wahl. Alternativ könnte man die Reihe logarithmisch transformieren und dann ein (saisonales) ARIMA-Modell anpassen. Anschließend ist natürlich die Prognose mittels Exponentiation zurück zu transformieren. Hier ist aber fraglich, ob sich dieser größere Aufwand lohnt. Der mögliche Gewinn der Box-Jenkins-Prognose geht bei größerem Prognosehorizont ja verloren. Dann liegt der Hauptteil der Struktur in dem Trend und der Saison; diese werden mit dem Holt-Winters-Verfahren adäquat erfasst.

Die Änderungen des wöchentlichen Marktanteils der Zahnpasta Crest, siehe Abbildung 5.7, weist keinen Trend und keine saisonale Struktur auf. Will man eine zuverlässige kurzfristige Prognose haben, so wird man ein ARMA-Modell anpassen. Die Prognose mittels exponentiellem Glätten würde nur das Niveau forschreiben.

Beim jährlichen Seegüterumschlag im Hamburger Hafen, Abbildung 5.8 ist vor allem ein Anwachsen des Niveaus zu erkennen. Das Anwachsen ist eher von quadratischer oder gar exponentieller Gestalt. Hier wird man die Daten (logarithmisch) transformieren und dann ein ARIMA-Modell anpassen. Alternativ könnte man auch den Trend mittels eines quadratischen Polynoms modellieren. Die Extrapolation einer solchen Trendfunktion ist aber immer gewagt.

Die in der Abbildung 5.9 dargestellten Verkäufe eines Produktes weist einen Trend auf, der kaum mit einer einfachen Funktion zu erfassen ist. Andere deterministische Trendfunktionen, wie etwa gleitende Durchschnitte oder Splines, eignen sich aber für die Prognose nicht. Da der Trend zudem im mittleren Bereich kurzfristig stark ansteigt, ist ein exponentielles Glättungsverfahren über den ganzen Zeitbereich kaum sinnvoll. Man wird den Ansatz eines ARIMA-Modells wählen und mit Hilfe eines solchen Modells eine Prognose erstellen.

Die Autoverkäufe in den USA, Abbildung 5.10, zeigen einen nur leicht steigenden Trend. Die Saisonalität ist ausgeprägt, aber ganz offensichtlich nicht sehr starr. Mit nur drei Saisonperioden ist die Reihe eher zu kurz, um von einem ausreichenden Einschwingen des Holt-Winters-Verfahrens auszugehen. Günstiger scheint ein ARIMA-Ansatz zu sein. Da die Fluktuation auf eine eher konstante Varianz hinweist, wird man auf eine logarithmische Transformation verzichten.

Zu der in der Abbildung 5.11 dargestellten Reihe: Es liegen nur 12 Beobachtungen vor. Daher ist ein möglichst einfaches Modell angebracht. Da die Reihe einen Trend aufweist, wäre wohl ein einfaches lineares Trendmodell mit autokorrelierten Störungen das angemessenste Modell. Man könnte auch noch ein ARIMA $[p, 1, q]$ -Modell versuchen. Hier würde man aber möglichst $q = 0$ und $p \leq 2$ ansetzen, da sonst die Schätzung auf zu wenigen Beobachtungen beruht.

5.5 R-Funktionen

Mit dem Output verschiedener Funktionen, u. a. `arima`, als Argument kann die Funktion `predict` direkt Prognosen erstellen. Bei `arima` ist es dabei sinnvoll, die Differenzbildung im Rahmen der Option `order` der Funktion `arima` durchzuführen. Dann erhält man gleich die Prognose der Ausgangsreihe.

Die Funktion `TrenchForecast` aus dem Paket `ltsa` berechnet die optimale lineare Prognose. Sie benötigt als Argument die Autokovarianzen des Modells. Speziell für ARMA $[p, q]$ -Modelle kann die Autokovarianzfunktion mit der Prozedur `tacvfARMA` ermittelt werden.

Die Funktion `ets` des Paketes `forecast` bestimmt das geeignete verschiedener Modelle des exponentiellen Glättens auf der Basis eines Informationskriteriums. Prognosen können dann mit `predict` oder `forecast` aus dem gleichen Paket erstellt werden. Die Funktion `auto.arima` dieses Paketes schätzt das beste ARIMA-Modell, gemessen anhand des AIC-, AICC- oder BIC-Wertes.

Das Paket `BootPR` enthält Funktionen zur Bias-korrigierten Prognose von AR-Modellen. Für Prognosen mit Methoden des Maschinellen Lernens gibt es das Paket `forecastML`.

Durch Gegenüberstellen von Prognosen und wahren Werten können Prognosen Ex-Post ausgewertet werden. Die Funktion `accuracy` aus dem Paket `forecast` berechnet verschiedene empirische Gütemaße, z. B. den mittleren absoluten/quadratischen Prognosefehler. Der Diebold-Mariano-Test testet die Nullhypothese, dass zwei Prognoseverfahren dieselbe Genauigkeit besitzen. Er ist in der Funktion `dm.test` implementiert. Die Testvariante mit mehr als zwei Prognoseverfahren wird im Paket `multDM` behandelt.

6 Periodizitäten in Zeitreihen

6.1 Periodizitäten

Viele empirische Zeitreihen weisen in offensichtlicher Weise ein wiederkehrendes Muster auf. Bei anderen Reihen sind solche Muster auf den ersten Blick nicht zu erkennen. Dennoch stellt sich bisweilen die Frage, ob und gegebenenfalls welche Zyklen auch in solch einer Reihe vorhanden sind. Beispiele sind etwa der Schweinezyklus und der Konjunkturzyklus. In diesem Kapitel werden Ansätze zur Entdeckung und Beschreibung von Zyklen in Zeitreihen sowie der zugehörige modellmäßige Rahmen vorgestellt.

Beispiel 6.1 (Anzahlen von Luchspelzen – Fortsetzung von Seite 55)

Die in Abbildung 2.6 dargestellte Zeitreihe LUCHS der logarithmierten Anzahlen von jährlichen Verkäufen von Luchspelzen ist durch ein relativ zyklisches Verhalten charakterisiert. Die Gipfpunkte sind alle in etwa gleich weit voneinander entfernt. Der Abstand beträgt ca. 9.5 Jahre.

Beispiel 6.2 (Schweinezyklus – Fortsetzung von Seite 2)

Die Originalzeitreihe der Schweinepreise, die Hanau (1928) auf den Begriff des Schweinezyklus brachte, ist in der Abbildung 6.1 wiedergegeben. Hier stellt sich die Frage, ob neben der saisonalen Komponente eine weitere zyklische Komponente vorhanden ist.

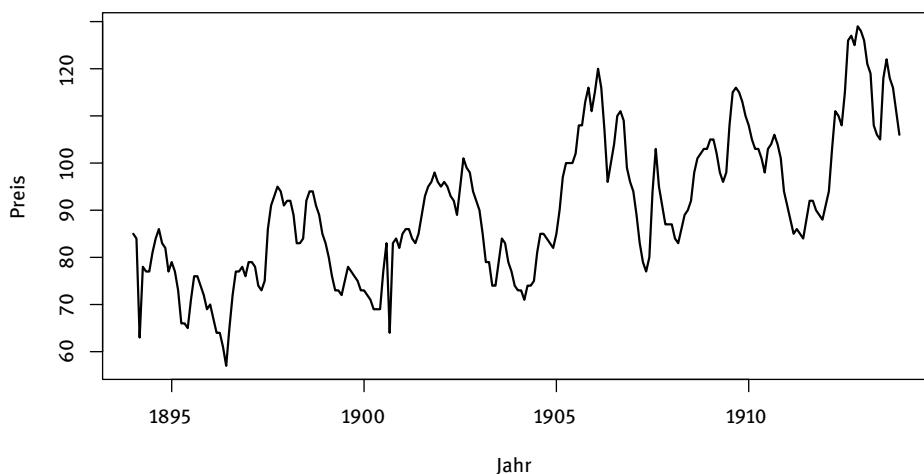


Abb. 6.1: Monatliche Preise für Schweine in Mark je 1900 kg, (80 bis 100 kg Lebendgewicht)

Die Dauer eines Zyklus in einer Zeitreihe wird als *Periode* bezeichnet. Formal heißt eine Funktionen f periodisch mit der Periode P , wenn sie die Beziehung

$$f(t + P) = f(t) \quad (-\infty < t < \infty)$$

erfüllt. Alternativ kann das zyklische Verhalten auch durch die *Frequenz*, den Kehrwert der Periode, beschrieben werden. Die Frequenz gibt an, welcher Teil des Zyklus pro Zeiteinheit vergangen ist.

Wichtige periodische Funktionen sind die Sinus- und Kosinusfunktion. Sie sind zunächst einmal periodisch mit der Periode 2π . Nach Umrunden des Einheitskreises, der gerade den Umfang 2π hat, starten sie wieder am Ausgangspunkt:

$$\sin(t + 2\pi) = \sin(t), \quad \cos(t + 2\pi) = \cos(t).$$

Durch den Übergang von t zu $2\pi t$ erhält man dann Funktionen, die periodisch sind mit der Periode eins. Andere Perioden P lassen sich durch Einfügen der zugehörigen Frequenz $\lambda = 1/P$ erreichen; das führt von dem Argument t zu dem Argument $2\pi\lambda t$.

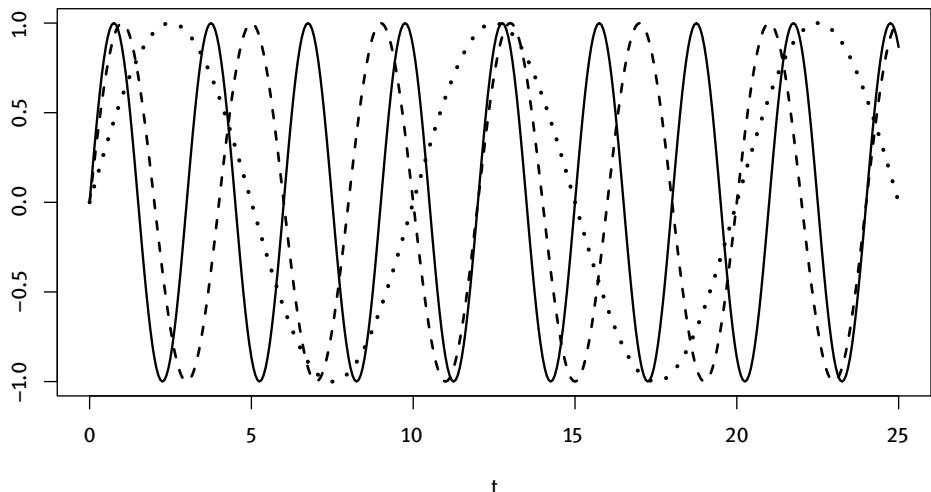


Abb. 6.2: Sinuskurven; — : $\lambda = 1/3$; - - - : $\lambda = 1/4$; ··· : $\lambda = 1/10$

Eine Sinusfunktion der Form $a \cdot \sin(2\pi\lambda t + \phi)$ heißt eine *harmonische Schwingung* oder *harmonische Welle*. Die Größe ϕ ist die *Phase*; sie verschiebt die Schwingung, so dass sie nicht mehr bei null anfängt. Die Verschiebung kann auch mittels einer Kombination von Sinus- und Kosinusfunktion dargestellt werden. Aufgrund der Beziehungen zwischen den beiden Funktionen gilt:

$$a \sin(2\pi\lambda t + \phi) = a \cos(\phi) \sin(2\pi\lambda t) + a \sin(\phi) \cos(2\pi\lambda t). \quad (6.1)$$

Jede Funktion $f(t)$, die periodisch ist mit der Periode P/k , $k = 1, 2, \dots$, ist auch periodisch mit der Periode P :

$$f\left(t + \frac{P}{k}\right) = f(t) \quad \Rightarrow \quad f(t + P) = f(t) \quad \text{mit } k = 1, 2, \dots .$$

Die Frequenzen $2\lambda, 3\lambda, 4\lambda, \dots$, die zu den Perioden P/k gehören, werden als *Oberschwingungen* von λ bezeichnet. λ heißt auch *Fundamentalsfrequenz* der Folge $\lambda, 2\lambda, 3\lambda, \dots$.

Linearkombinationen oder, wie man auch sagt, Überlagerungen von harmonischen Wellen mit gleichen Perioden sind wieder periodisch. Periodische Funktionen lassen sich andersherum unter geeigneten Bedingungen als Überlagerung harmonischer Wellen mit dieser Periode und ihren Oberschwingungen darstellen. Die Bedingungen sind so allgemein, dass sie in der Praxis als erfüllt angesehen werden können.

Beispiel 6.3 (Sägezahnfunktion)

Die Sägezahnfunktion

$$f(t) = \begin{cases} t - 4n & \text{für } 4n \leq t < 4n + 3 \\ -\frac{1}{3}(t - 4n) + 12 & \text{für } 4n + 3 \leq t < 4n + 4 \end{cases} \quad (n = 0, 1, \dots)$$

weist eine Periode von 4 auf. Um diese Funktion als Überlagerung harmonischer Wellen zu beschreiben, sind Sinus- und Kosinusfunktionen $\sin(2\pi\lambda t), \cos(2\pi\lambda t)$ mit $\lambda = 1/4, 1/2, 3/4, 1, 5/4, \dots$ zu verwenden.

Über die Approximation rein periodischer Funktionen hinaus können auch beliebige endliche Folgen mittels Überlagerungen harmonischer Wellen dargestellt werden. Dabei werden allerdings Wellen unterschiedlicher Frequenzen benötigt. (Die Periodizität würde sich durch zyklische Wiederholung der endlichen Folge ergeben.) Damit können auch Zeitreihen entsprechend dargestellt werden. Zudem ist zu berücksichtigen, dass die Wellen geeignet verschoben werden müssen, um eventuelle Gipfel- und Tiefpunkte treffen zu können. Aufgrund der Beziehung (6.1) kann dies geschehen, indem Summen von Sinus- und Kosinuswellen gleicher Frequenz berücksichtigt werden.

Beispiel 6.4 (Überlagerung harmonischer Wellen)

Zur Illustration wurden drei harmonische Wellen für $0 \leq t \leq 90$ erzeugt.

$$\begin{aligned} y_{1t} &= 2 \cos(2 \cdot \pi \cdot 0.025t) + 1.5 \cdot \sin(2 \cdot \pi \cdot 0.025t), \\ y_{2t} &= 2 \cos(2 \cdot \pi \cdot 0.08t) + 3 \cdot \sin(2 \cdot \pi \cdot 0.08t), \\ y_{3t} &= 5 \cos(2 \cdot \pi \cdot 0.23t) + 3 \cdot \sin(2 \cdot \pi \cdot 0.23t). \end{aligned}$$

Diese drei Funktionen und ihre Summe sind in der Abbildung 6.3 dargestellt.

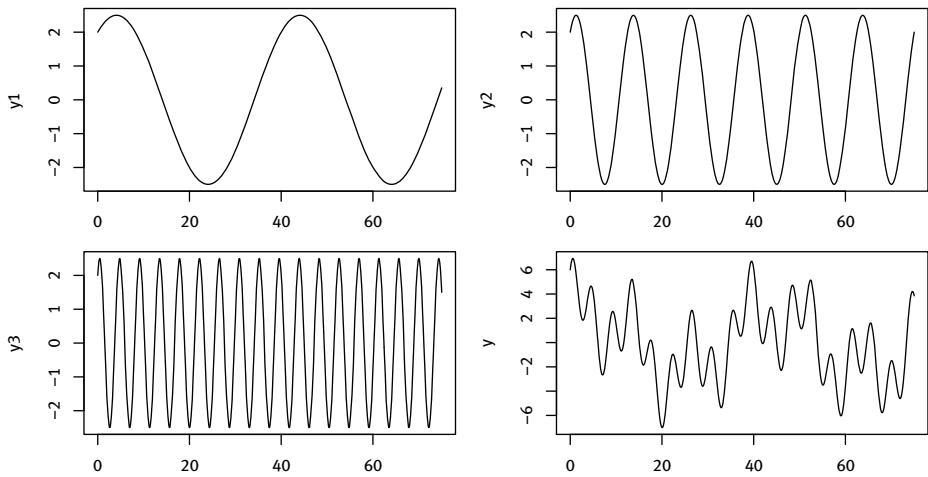


Abb. 6.3: Harmonische Wellen und ihre Überlagerung

Die theoretische Absicherung liefert das *Spektral-Repräsentationstheorem*, welches sagt, dass jede stationäre Zeitreihe als Überlagerung von Sinus- und Kosinuswellen verschiedener Frequenzen dargestellt werden kann. Das ist die sogenannte *Cramér-Darstellung*. Um eine Idee davon zu liefern sei ein Kosinoid ($A \cos(2\pi\lambda t) + B \sin(2\pi\lambda t)$) betrachtet, bei dem A und B Zufallsvariablen sind. Mit $E(A) = E(B) = 0$, $\text{Var}(A) = \text{Var}(B)$ und $\text{Cov}(A, B) = 0$ ist dieses ein stationärer stochastischer Prozess. Dann ist auch die Überlagerung von k Kosinoiden verschiedener Frequenzen $\lambda_1, \dots, \lambda_k$,

$$Y_t = \sum_{j=1}^k (A_j \cos(2\pi\lambda_j t) + B_j \sin(2\pi\lambda_j t)) \quad (6.2)$$

ein stationärer Prozess, wenn zusätzlich die (A_i, B_i) und (A_j, B_j) für $i \neq j$ unabhängig sind. (Y_t) heißt dann ein *harmonischer Prozess*. Für einen harmonischen Prozess gilt $\text{Var}(Y_t) = \sum_{j=1}^k \text{Var}(A_j)$. Die Varianz $\text{Var}(A_j)$ der Amplitude A_j gibt an, welchen Beitrag die Schwingungskomponente mit der Frequenz λ_j zur Gesamtvarianz leistet. Die Berücksichtigung aller Frequenzen $0 \leq \lambda \leq 0.5$ führt dann zu der angegebenen Cramér-Darstellung.

6.2 Periodische Trends

Das einfachste Modell, in dem eine Periodizität bei einer Zeitreihe berücksichtigt wird, ist das *eines periodischen Trends*, bei dem die Frequenz λ bekannt ist, und bei dem die Störungen ε_t einen White-Noise-Prozess mit $E(\varepsilon_t) = 0$ bilden:

$$Y_t = c + a \cos(2\pi\lambda t) + b \sin(2\pi\lambda t) + \varepsilon_t . \quad (6.3)$$

Das Modell ist linear in den Koeffizienten a, b , sofern die Frequenz bekannt ist. Die Schätzung oder Inferenz bezüglich dieser unbekannten Koeffizienten erfolgt daher im Rahmen der linearen Regression. Geht man von der zentrierten Zeitreihe aus, so erhält man Schätzwerte für a, b mit dem Ansatz der kleinsten Quadrate gemäß

$$\sum_{t=1}^N [y_t - \bar{y} - \{a \cos(2\pi\lambda t) + b \sin(2\pi\lambda t)\}]^2 \stackrel{!}{=} \min_{a,b} .$$

Neben den Punktschätzungen stehen mit der zusätzlichen Normalverteilungsannahme der Störungen die bekannten Konfidenzintervalle und Tests zur Verfügung.

Beispiel 6.5 (Anzahlen von Luchspelzen – Fortsetzung von Seite 133)

Bei der logarithmisch transformierten Reihe LUCHS, den Anzahlen von versteigerten Luchsfallen, sind die Gipfpunkte im Mittel 9.5 Jahre voneinander entfernt. Daher wird λ mit $1/9.5$ festgesetzt. Es ergeben sich mit $\bar{y} = 2.90$ als Koeffizienten der harmonischen Welle die Werte: $\hat{a} = 0.363$, $\hat{b} = -0.491$.

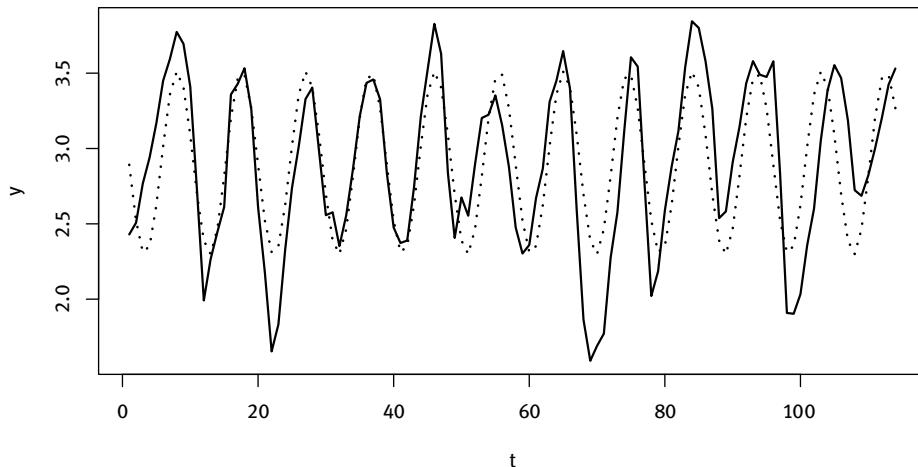


Abb. 6.4: Transformierte Reihe LUCHS mit periodischem Trend

R-Code:

```

1 data(LYNX)
2 y <- log10(LYNX)
3 plot(y)
4 lambda <- 1/9.5
5 t <- c(1:length(y))
6 co <- cos(2*pi*lambda*t)
7 si <- sin(2*pi*lambda*t)
8 out <- lm(as.vector (y-mean(y))~co+si-1)
9 plot(t,y,type="l")
10 lines(t,mean(y)+out$fitted.values,lty=3)

```

Die Sinus- und die Kosinuswelle werden elementar erzeugt. Mit der Funktion $1m$ wird eine lineare Regression durchgeführt. Um der Praxis gerecht zu werden, dass Reihen i. d. R. mit dem arithmetischen Mittel zentriert werden, wird dies hier entsprechend umgesetzt. Die „-1“ im Argument der Funktion $1m$ bewirkt, dass kein Achsenabschnitt bei der Regression berücksichtigt wird.

6.3 Das Periodogramm

Definition des Periodogramms

Für die Fourierfrequenzen $\lambda_j = j/N, j = 0, 1, \dots, [N/2]^1$ führt der Ansatz der kleinsten Quadrate beim periodischen Trend, d. h. die Minimierung von

$$SS(a_j, b_j) = \sum_{t=1}^N [(y_t - \bar{y}) - \{a_j \cdot \cos(2\pi\lambda_j t) + b_j \cdot \sin(2\pi\lambda_j t)\}]^2$$

bzgl. a_j, b_j auf die Schätzwerte

$$\hat{a}_j = 2 \cdot \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) \cos(2\pi\lambda_j t) \quad (6.4a)$$

$$\hat{b}_j = 2 \cdot \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) \sin(2\pi\lambda_j t). \quad (6.4b)$$

Setzt man diese Werte in die Summe der quadrierten Abweichungen ein, so erhält man

$$SS(\hat{a}_j, \hat{b}_j) = \sum_{t=1}^N (y_t - \bar{y})^2 - \frac{N}{2} \cdot [\hat{a}_j^2 + \hat{b}_j^2].$$

Die Größe $I(\lambda_j)/N = [\hat{a}_j^2 + \hat{b}_j^2]/4$ hat eine eigene statistische Interpretation. Dazu wird die Kreuzkovarianz zweier Reihen (x_t) und (y_t) benötigt. Diese ist definiert als

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

$I(\lambda_j)/N$ ist nun gerade die maximale quadrierte Kreuzkovarianz von (y_t) mit ($a \cdot \cos(2\pi\lambda_j t) + b \cdot \sin(2\pi\lambda_j t)$), wenn a, b variieren dürfen. Allerdings müssen diese Parameter die Nebenbedingung $a^2 + b^2 = 1$ erfüllen; sie sichert, dass die Kovarianz nicht einfach aufgrund größerer Werte der Koeffizienten ansteigt. Für die Interpretation von $I(\lambda_j)$ ist hauptsächlich die relative Größe relevant. Je größer $I(\lambda_j)$, desto größer (absolut gesehen) ist auch die Kreuzkovarianz der Reihe mit dieser Welle, desto stärker ist die Frequenz in der Reihe ausgeprägt.

¹ $\lfloor \cdot \rfloor$ ist die Gaußsche Klammer; $\lfloor x \rfloor$ bezeichnet die größte ganze Zahl, die kleiner oder gleich x ist.

$I(\lambda_j)$ ist zunächst nur für die Fourierfrequenzen erklärt. Es liegt jedoch nahe, $I(\lambda)$ auch für andere Frequenzen in gleicher Weise zu definieren. Dann ist die Interpretation als maximale quadrierte Kreuzkovarianz zwar nicht mehr ganz korrekt, man kann sie aber noch angenähert verwenden.

Definition 6.6 (Periodogramm)

Das *Periodogramm* der Zeitreihe (y_t) ist die Funktion

$$I(\lambda) = N \left[\left(\frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) \cos(2\pi\lambda t) \right)^2 + \left(\frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}) \sin(2\pi\lambda t) \right)^2 \right]. \quad (6.5)$$

Man braucht sich das Periodogramm nur für $0 \leq \lambda \leq 0.5$ anzusehen. Denn die kleinste entdeckbare Periodizität beträgt zwei, während die größte abhängig von der Reihenlänge ist. Die Beziehung $2 \leq P < \infty$ führt aber für die Frequenz $1/P$ zu dem angegebenen Bereich. Auch formal ist dieses Intervall ausreichend. Wegen der Periodizität der trigonometrischen Funktionen ist auch das Periodogramm periodisch mit der Periode eins:

$$\begin{aligned} I(\lambda + 1) &= \frac{1}{N} \left[\left(\sum (y_t - \bar{y}) \cos(2\pi(\lambda + 1)t) \right)^2 + \left(\sum (y_t - \bar{y}) \sin(2\pi(\lambda + 1)t) \right)^2 \right] \\ &= \frac{1}{N} \left[\left(\sum (y_t - \bar{y}) \cos(2\pi\lambda t) \right)^2 + \left(\sum (y_t - \bar{y}) \sin(2\pi\lambda t) \right)^2 \right]. \end{aligned}$$

Daher reicht z. B. die Betrachtung von $I(\lambda)$ über dem Intervall $[-0.5, 0.5]$ aus. Da zudem $I(\lambda) = I(-\lambda)$ gilt, braucht $I(\lambda)$ nur über dem Intervall $[0, 0.5]$ betrachtet zu werden.

Die Bestimmung des Periodogramms kann auf verschiedene Weisen erfolgen. Einmal kann die Definitionsgleichung direkt umgesetzt werden. Das ist aber vor allem bei längeren Reihen recht aufwändig. Eine zweite Möglichkeit basiert auf der folgenden Beziehung:

$$I(\lambda) = c_0 + 2 \sum_{\tau=1}^{N-1} c_\tau \cos(2\pi\lambda\tau). \quad (6.6)$$

Dabei ist (c_τ) die empirische Autokovarianzfunktion der Zeitreihe. Als dritte Möglichkeit sei die schnelle Fouriertransformation, kurz FFT, erwähnt, die in der Tat eine sehr schnelle Berechnung erlaubt. Die FFT geht auf Gauß zurück. Sie wurde dann von Cooley und Tukey (1965) wieder entdeckt und gilt als bedeutendster Beitrag der Numerik zur Zeitreihenanalyse allgemein. (Zur Fouriertransformation siehe S. 152.) Sie ist für Reihen geeignet, deren Länge eine Potenz von zwei ist. Daher wird eine Zeitreihe ggf. so oft um den Mittelwert verlängert, bis sie eine Länge hat, die sich als Potenz von zwei darstellen lässt, $N^* = 2^K$. Das Periodogramm wird dann an den Frequenzen i/N^* bestimmt. Bei der Anwendung bereits implementierter Programme ist darauf zu achten, ob die Reihe vorab zentriert werden muss, da zur Verlängerung häufig automatisch Nullen angehängt werden.

Probleme bei der Interpretation des Periodogramms

Aus Gründen der Übersichtlichkeit wird das Periodogramm i. d. R. an 50–300 gleichabständigen Punkten des Intervalls $[0, 0.5]$ gezeichnet. Oft werden die Fourierfrequenzen $\lambda_j = j/N$ gewählt. Aus der grafischen Darstellung versucht man dann, Aufschlüsse über die in einer Reihe vorhandenen Periodizitäten zu erhalten. Dabei sind die folgenden *Interpretationshürden* zu beachten.

1. *Aliasing* (Maskierung). Aufgrund der Wahl eines zu großen Zeitintervalls für die Beobachtung eines ‚kontinuierlich‘ ablaufenden Phänomens kann es vorkommen, dass man die tatsächliche Periodizität nicht sieht, sondern eine längerwellige. Formal drückt sich das in der Beziehung $I(\lambda) = I(k + \lambda) = I(k - \lambda)$ aus.
2. *Leakage* (Durchsickern). Wegen der Kürze von empirischen Zeitreihen ist mit $I(\lambda)$ auch $I(\lambda')$ erhöht, wenn λ' dicht bei λ liegt.
3. *Oberschwingungen*. Bei nicht-sinusförmigen Periodizitäten sind mit $I(\lambda)$ auch $I(2\lambda), I(3\lambda), \dots$ erhöht.

Das Aliasing-Problem hängt wie angedeutet mit dem Beobachtungsintervall und dem zugrundeliegenden Prozess zusammen. Hat man eine Reihe durch diskrete Beobachtung eines kontinuierlich ablaufenden Prozesses $Y(t)$ erhalten, so enthält die Zeitreihe natürlich weniger Information als eine feiner abgetastete, d. h. eine mit dichter beieinander liegenden Beobachtungszeitpunkten. Sei Δ der Zeitabstand, in dem der Prozess abgetastet wird. Dann ist

$$Y_t = Y(t \cdot \Delta) \quad (t = 0, 1, 2, \dots).$$

Anhand von Realisationen des Prozesses (Y_t) sind Perioden nur bis hinunter zu 2Δ beobachtbar. Die maximale Frequenz von $1/(2\Delta)$ pro Zeiteinheit heißt *Nyquist Frequenz*.

Beispiel 6.7 (Wasserstand)

Wird der Wasserstand eines Meeres an einer Stelle der Küste einmal die Woche oder alle 7 Tage gemessen, so gibt das die Nyquist Frequenz von $\frac{1}{2 \times 7} = \frac{1}{14}$. Ein Zyklus, der einmal in zwei Wochen bzw. in 14 Tagen vollendet ist, kann also beobachtet werden.

Gäbe es nun einen Gezeiteneffekt, der nach sechs Tagen abgeschlossen ist, so wäre dieser nicht mehr zu erkennen. Er erschien vielmehr als langwelliger Siebenwocheneffekt. Das ergibt sich aus folgender Beziehung:

$$\cos\left(2\pi \frac{6t}{7}\right) = \cos\left(2\pi t - 2\pi \frac{1}{7}t\right) = \cos\left(2\pi \frac{1}{7}t\right).$$

Man muss also schon bei der Erhebung kontinuierlicher oder möglicherweise schneller ablaufender Vorgänge berücksichtigen, welche Perioden bzw. Frequenzen man noch betrachten will. Das gilt sinngemäß auch für die Diskretisierung bereits vorliegender kontinuierlicher Signale.

Zur Illustration der Oberschwingungen sei auf das Beispiel 6.3 mit der Sägezahnreihe verwiesen. Um diese durch eine Überlagerung von harmonischen Wellen zu approximieren, wird nicht nur die Welle derselben Periode benötigt, sondern auch schneller schwingende Wellen, die jedoch auch noch die gleiche Periode haben. Das müssen also Oberschwingungen sein.

Eine der Konsequenzen des Auftretens von Oberschwingungen ist die Erhöhung von Periodogrammordinaten bei saisonbehafteten Reihen nicht nur an der Saisonfrequenz $1/s$, sondern auch an den Stellen $2/s, 3/s, \dots$. Hat die Reihe einen Trend, so erscheint dieser als eine Welle mit langer Schwingung, einer Periode die der Reihenlänge entspricht. Dies führt zu einer Erhöhung der Periodogrammordinaten für die Frequenzen $1/N, 2/N$ usw. Dass diese beiden Komponenten die bestimmenden Faktoren des Periodogramms ökonomischer Reihen sind, hat Granger (1966) in einer berühmten Arbeit ausgeführt.

Beispiel 6.8 (Steuern in Deutschland – Fortsetzung von Seite 45)

Die Zeitreihe STEUERN der in Deutschland insgesamt monatlich gezahlten Gemeinschaftsteuern, in Milliarden Euro von Januar 1999 bis Juni 2019, ist in der Abbildung 2.17 wiedergegeben. Hier ist das zugehörige Periodogramm grafisch dargestellt.

Die Struktur des Periodogramms lässt sich aufgrund der Eigenheiten der Zeitreihe erklären. Zunächst gibt es in der Zeitreihe eine Saisonperiode von einem Jahr. Diese wird durch das jahresweise gleiche Muster, insbesondere durch die gleich weit voneinander weg liegenden hohen Gipfpunkte, die Dezemberwerte, charakterisiert. Die Fundamentalfrequenz beträgt $1/12 = 0.0833$. Da die Saisonkomponente nicht sinusförmig ist, kommt es zu Oberschwingungen an den Frequenzen

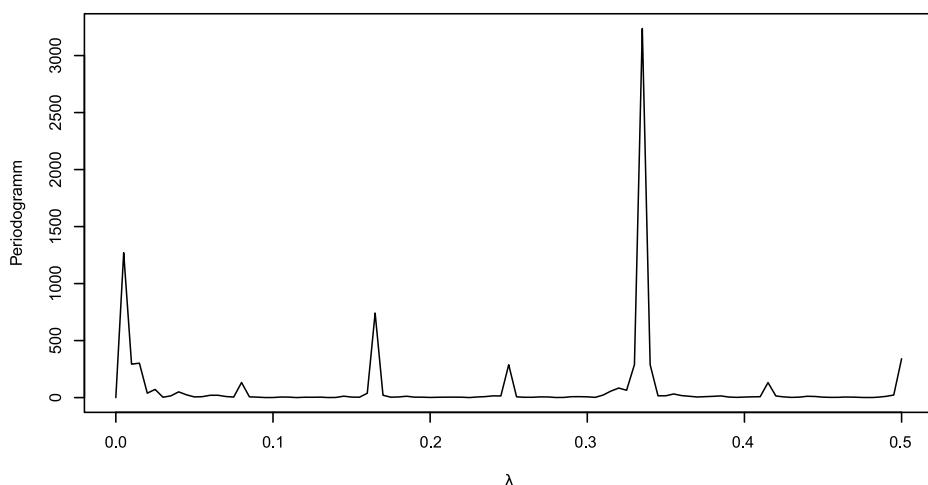


Abb. 6.5: Periodogramm der Reihe STEUERN

$2/12, 3/12, 4/12$ und $5/12$. Weiter gibt es in der Reihe eine Periodizität der Länge 3. Diese resultiert aus den Quartalswerten; am Ende eines jeden Quartals sind vielfach Steuervorauszahlungen fällig. Als Konsequenz ist das Periodogramm bei $1/3 = 4/12 = 0.33$ erhöht. Bei der Frequenz 0.33 gibt es also zwei Einflüsse, die eigenständige Periodizität und die aus der Jahresperiodik. Der Trend in der Zeitreihe lässt sich an dem hohen Wert nahe bei der Frequenz Null erkennen.

R-Code:

```

1 data(TAXES)
2 y <- TAXES
3 fp <- periodogram(y,100)
4 plot(fp[,1],fp[,2],type="l")

```

Um das Periodogramm der Reihe darzustellen wird die Funktion `periodogram` aus `tsapp` benutzt. Sie berechnet die Periodogrammordinaten an der gewünschten Anzahl von gleichabständigen Stellen im Intervall $[0,1/2]$ über die Beziehung (6.6).

Das Leakage-Problem wird anhand der Abbildung 6.6 deutlich. Hier ist einmal eine Kosinuswelle mit der Frequenz $\lambda = 0.1$ an zwanzig Zeitpunkten beobachtet worden (durchgezogene Kurve). Die optimal angepasste harmonische Welle mit der Frequenz $\lambda = 0.125$ (gestrichelte Kurve) ist offensichtlich mit dieser Ausgangsreihe stark korreliert. Also ist die zugehörige Periodogrammordinate auch an der Stelle $\lambda = 0.125$ erhöht, obwohl die Ausgangsreihe eine reine Schwingung mit der Frequenz $\lambda = 0.1$

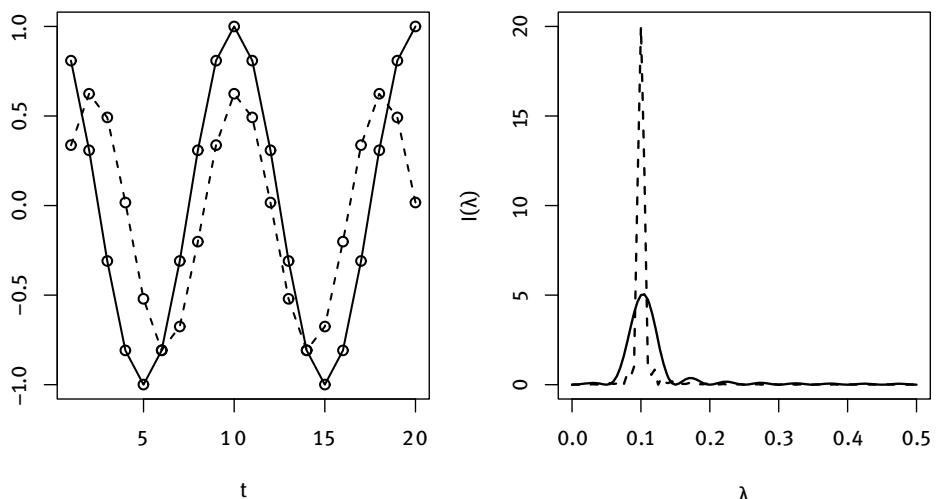


Abb. 6.6: Links: Kosinuswelle mit $\lambda = 0.1$ (—) und beste Approximation einer harmonischen Welle mit $\lambda = 0.125$ (- - -); rechts: Periodogramme einer Kosinuswelle mit $\lambda = 0.1$ für $N = 20$ (—) u. $N = 80$ (- - -)

ist. Diese Korrelation resultiert aus der Endlichkeit des Beobachtungsabschnitts. Sie verringert sich bei längeren Reihen, wie aus dem rechten Teil der Abbildung 6.6 deutlich wird.

Test auf eine Periodizität

Das Periodogramm kann u. a. ausgenutzt werden, um einen *Test auf eine Periodizität* durchzuführen, wenn man die Periode nicht von vornherein kennt. Das ist eine relevante Verallgemeinerung des einfachen Tests auf das Vorhandensein einer harmonischen Welle mit bekannter Frequenz. Die Grundlage bildet das *Modell einer verborgenen Periodizität*, bei dem nunmehr λ unbekannt ist:

$$Y_t = a \cdot \cos(2\pi\lambda t) + b \cdot \sin(2\pi\lambda t) + \varepsilon_t \quad (6.7)$$

mit $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, $\lambda \in (0, 0.5)$.

Die Verteilungseigenschaften des Periodogramms unter der Nullhypothese $H_0: a = 0, b = 0$ sind im folgenden Satz zusammengestellt. Dass die Periodogrammordinaten danach von λ unabhängige Erwartungswerte und Varianzen haben, ist dabei leicht einsichtig. Denn für White-Noise-Prozesse gilt ja $E(c_\tau) \approx 0$, $\tau \neq 0$. Damit ist

$$E(I(\lambda)) = E(c_0) + 2 \sum_{\tau=1}^{N-1} E(c_\tau) \cos(2\pi\lambda\tau) \approx E(c_0) = \sigma^2. \quad (6.8)$$

Satz 6.9 (Verteilung der modifizierten Periodogrammordinaten)

Sei (Y_t) ein Gaußscher White-Noise-Prozess, $Y_t \sim \mathcal{N}(0, \sigma^2)$. Das modifizierte Periodogramm $J(\lambda_j)$ sei aus der nicht-zentrierten Reihe berechnet:

$$J(\lambda) = N \left[\left(\frac{1}{N} \sum_{t=1}^N Y_t \cos(2\pi\lambda t) \right)^2 + \left(\frac{1}{N} \sum_{t=1}^N Y_t \sin(2\pi\lambda t) \right)^2 \right] \quad (6.9)$$

Dann haben die normalisierten Periodogrammordinaten $J(\lambda_j)/(2\sigma^2)$ an den Fourierfrequenzen $\lambda_j = j/N$, $j = 1, \dots, M = \lfloor (N-1)/2 \rfloor$ eine χ^2 -Verteilung mit 2 Freiheitsgraden.¹ Zudem sind $J(\lambda_j), J(\lambda_{j'})$, $j \neq j'$, unabhängig.

Unter H_0 sollten die Periodogrammordinaten also unregelmäßig um eine Konstante schwanken. Ist andererseits die Nullhypothese $H_0: a = 0, b = 0$ falsch, so gibt es ei-

¹ Bei Verteilungsaussagen über Periodogrammordinaten muss man den Sonderfall $\lambda = 0.5$ extra behandeln, da der Sinus ja Null ist. Um nicht zu technisch zu werden, wird diese Situation hier ausgebündet.

ne harmonische Welle, die der beobachteten Reihe inhärent ist. Daher sollte an einer Stelle das Periodogramm einen großen Wert zeigen, während an den anderen Stellen des Intervalls $[0, 0.5]$ das Periodogramm vergleichsweise klein ist. Als Prüfgröße bietet sich folglich die größte Periodogrammordinate an. Die theoretische Verteilungseigenschaft wird im folgenden Satz formuliert. Im Anschluss wird auf die Umsetzung eingegangen.

Satz 6.10 (Verteilung der maximalen Periodogrammordinate)

(Y_t) sei ein Gaußscher White-Noise-Prozess mit $\text{Var}(Y_t) = \sigma^2$. Es seien $\lambda_j = j/N$, $j = 1, \dots, M$, $M = \lfloor (N - 1)/2 \rfloor$; dann gilt für die aus dem modifizierten Periodogramm $J(\lambda)$, siehe (6.9), gebildete Prüfgröße $T = \max_{1 \leq j \leq M} J(\lambda_j)/\sigma^2$:

$$P(T \geq t) = 1 - (1 - e^{-t})^M. \quad (6.10)$$

Der zu einem vorgegebenen Niveau α gehörige kritische Wert c der Prüfgröße wird also bestimmt, indem die rechte Seite gleich α gesetzt wird. Dies ergibt:

$$c = -\ln(1 - (1 - \alpha)^{1/M}). \quad (6.11)$$

Alternativ kann zu dem beobachteten Wert t^* von T der P -Wert, d. h. die Wahrscheinlichkeit $P(T \geq t^*)$, bestimmt werden. Ist er kleiner als α , so ist die Nullhypothese zu verwerfen.

In der Regel hat man es mit einer nicht bei null zentrierten Reihe zu tun; zudem ist die Varianz σ^2 unbekannt. Dann wird man die Reihe zunächst zentrieren; folglich wird das übliche Periodogramm $I(\lambda_j)$ anstelle von $J(\lambda_j)$ benutzt. Weiter wird die unbekannte Varianz durch die Schätzung $\hat{\sigma}^2 = \frac{1}{M} \sum_{j=1}^M I(\lambda_j)$ ersetzt. Dass hier zur Schätzung der Varianz ein arithmetisches Mittel verwendet wird, hängt damit zusammen, dass unter H_0 der Erwartungswert der Periodogrammordinaten gerade gleich σ^2 ist. Bei Verwendung dieser Schätzung kann ab $N \approx 100$ der kritische Wert wie oben bestimmt werden. Für kleinere N muss man auf die exakte Verteilung zurückgreifen.

Grenander & Rosenblatt (1957) haben die auch exakte Verteilung der zweit- und drittgrößten Periodogrammordinaten bestimmt. Daher kann man die größten Periodogrammordinaten sukzessive testen. Dabei sollte aber eine Adjustierung des Testniveaus im Sinne der multiplen Tests vorgenommen werden.

Beispiel 6.11 (Schweinezyklus – Fortsetzung von Seite 133)

Um einen nicht offensichtlichen Zyklus in dieser Reihe zu entdecken, wird sie zunächst mit der Funktion `stl` trend- und saisonbereinigt. Zu der Funktion `stl` siehe das Kapitel 12. Die saison- und trendbereinigte Reihe hat das in der Abbildung 6.7 dargestellte Periodogramm.

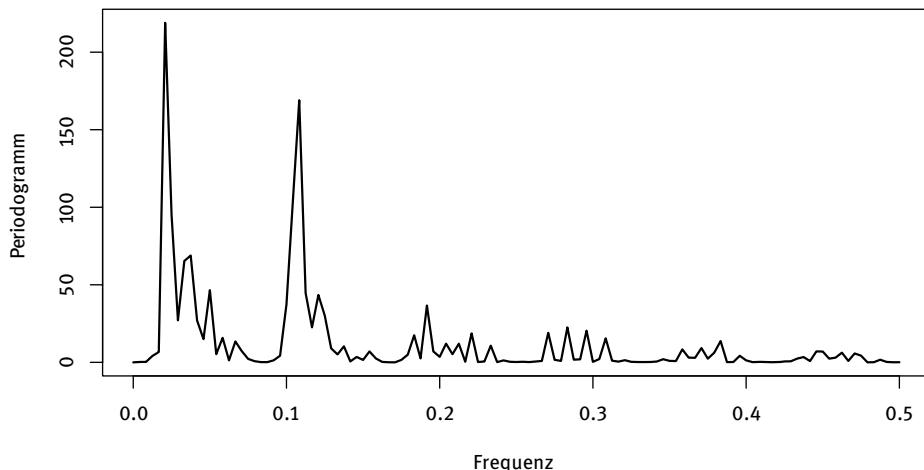


Abb. 6.7: Periodogramm der saison- und trendbereinigten Preise für Schweine

Die Frequenz mit der größten Ordinate ist 0.02083. Bei einer Reihenlänge von $n = 240$ entspricht dies gerade $n/5$ oder einer Periode von 4 Jahren. Dies ist also der berühmte „Schweinezyklus“. Er ist auch signifikant. Mit dem arithmetischen Mittel von 11.97591 der Periodogrammordinaten ergibt sich:

$$t = \frac{218.914}{11.97591} = 18.2795 \quad \Rightarrow \quad P(T \geq t) = 1 - (1 - e^{-18.2795})^{119} = 1.3 \cdot 10^{-6}.$$

Die nächst höhere Periodogrammordinate befindet sich bei der Frequenz $0.108333 = 26/n$. Das ist in etwa eine Zwei-Jahres-Periode. Eine Interpretation fällt schwer; es ist eher eine nicht ausreichende Saisonbereinigung oder eine Oberschwingung zu vermuten.

R-Code:

```

1 data(PIGPRICE)
2 y <- PIGPRICE
3 out <- stl(y,s.window=6)
4 e <- out$time.series[,3]
5 n <- length(e)
6 Ffreq <- length(seq(1:(n/2))/n)
7 p <- periodogram(e,Ffreq)
8 plot(p[,1],p[,2],type="l")
9 m <- mean(p[,2])
10 1-(1-exp(-max(p[,2])/m))^trunc((n-1)/2)
11 out <- periodotest(e)

```

Das Periodogramm wird mit der Funktion `periodogram` aus dem Paket `tsapp` berechnet. Diese Funktion setzt die Beziehung zur ACF direkt um, wenn das optionale Argument `ACF` auf wahr gesetzt wird. So kann das Periodogramm an einer frei gewählten Anzahl gleichabständiger Stellen $0 \leq \lambda \leq 0.5$ bestimmt werden. In Zeile 6 wird sichergestellt, dass das Periodogramm an den Fourierfrequenzen berechnet wird. Ergänzt wird der R-Code durch die Zeilen 9 und 10 zur Berechnung des P -Wertes des Tests auf das Vorliegen einer verborgenen Periodizität. Dies ist hier zur Illustration ausgeführt. Für eine einfache Anwendung ist das Vorgehen mit der Funktion `periodotest` in `tsapp` implementiert. Die letzte Zeile fasst also die Zeilen 5 bis 9 zusammen.

Test auf White-Noise

Auf dem Periodogramm basiert ebenfalls ein *Test auf White-Noise*. Falls die Nullhypothese H_0 : „ (Y_t) ist ein Gaußscher White-Noise-Prozess“ gilt, hat das (aus dem modifizierten Periodogramm bestimmte) *kumulierte Periodogramm*

$$S_r = \frac{\sum_{k=1}^r J(\lambda_k)}{\sum_{k=1}^M J(\lambda_k)}, \quad r = 1, \dots, M-1, \quad M = \lfloor (N-1)/2 \rfloor, \quad (6.12)$$

die gleiche gemeinsame Verteilung wie eine geordnete Stichprobe aus einer Rechteckverteilung über dem Intervall $(0, 1)$. Daher sollten die Werte von S_r bei Gültigkeit von H_0 um die Winkelhalbierende streuen. Dies führt zu *Bartletts Kolmogorov-Test* mit der Prüfgröße $C = \max_{0 < r < M} |S_r - r/M|$, bei der wieder das in der Praxis übliche Periodogramm zugrunde gelegt wird. Kritische Werte ergeben sich gemäß der Approximation:

$$C = \frac{\sqrt{-0.5 \ln(\alpha/2)}}{\sqrt{M-1} + 0.2 + 0.68/\sqrt{M-1}} - \frac{0.4}{M-1}. \quad (6.13)$$

Bei längeren Zeitreihen, die Realisationen von White-Noise-Prozessen mit Nicht-Gaußschen Verteilungen sind, gilt die Verteilungsaussage und damit diese Bestimmungsformel für die kritischen Werte zumindest noch approximativ.

Beispiel 6.12 (Regen in LA – Fortsetzung von Seite 34)

Bei der Reihe `REGEN` stellte sich im Kapitel 2 die Frage, ob überhaupt Abhängigkeiten vorhanden sind. Zur Beantwortung soll nun Bartlett's Kolmogorov-Test auf White-Noise durchgeführt werden.

Für die Teststatistik C erhält man den Wert 0.086, während der approximative P -Wert gleich 0.6867 ist. In der Grafik, siehe Abbildung 6.8, verbleibt dementsprechend der Streckenzug des kumulierten Periodogramms in den gestrichelt eingezeichneten Grenzen. Die Nullhypothese, dass die Reihe eine Realisation eines White-Noise-Prozesses ist, kann daher nicht abgelehnt werden.

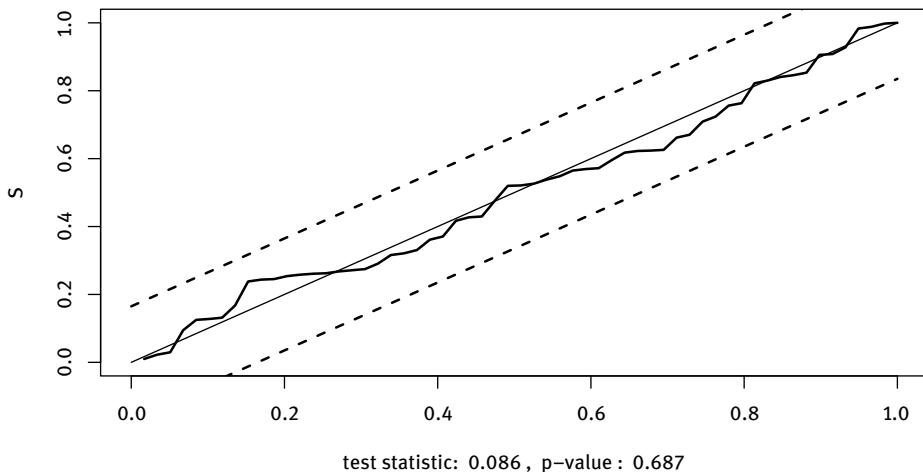


Abb. 6.8: White-Noise-Test für REGEN

Die Frage, ob die vorliegende Reihe als Realisation eines White-Noise-Prozesses angesehen werden kann, taucht auch im Zusammenhang mit der Untersuchung der Residuen nach Anpassung eines linearen Prozesses auf. Für diese Situation ist der Test nicht geeignet. Denn Residuen einer Regression oder einer Autoregression sind generell korreliert. Eine Erweiterung des Tests auf White-Noise erlaubt es aber, zu überprüfen, ob die Residuen einer linearen Regression als Realisation einer Folge unabhängiger Zufallsvariablen angesehen werden können. Bei dieser Erweiterung wird ein Paar von kritischen Werten ermittelt. Fällt die Prüfgröße in den Bereich außerhalb der äußeren Grenze, so wird die Nullhypothese der Unabhängigkeit abgelehnt. Ist sie kleiner als die innere Grenze, so wird sie beibehalten; ansonsten kann keine Entscheidung getroffen werden.

Beispiel 6.13 (Schweinezyklus – Fortsetzung von Seite 144)

Im Rahmen des Modells der verborgenen Periodizität wurde ein dominierender Einfluss der harmonischen Schwingung mit der Frequenz $0.02083 = 1/48$ bei den Residuen e_t eines Komponentenmodells festgestellt; dies deutet auf eine Vierjahresperiode hin. Zur Überprüfung des Modells wird daher eine Regression

$$E_t = \beta_0 + \beta_1 \cos\left(2\pi \frac{1}{48} t\right) + \beta_2 \sin\left(2\pi \frac{1}{48} t\right) + \varepsilon_t$$

durchgeführt, und die Residuen werden auf White-Noise getestet. Als Regressionsbeziehung ergibt sich $\hat{e}_t = 0.02373 + 1.89716 \cos(2\pi \frac{1}{48} t) - 0.22214 \cdot \sin(2\pi \frac{1}{48} t)$.

Das kumulierte Periodogramm verlässt die äußeren Grenzen, die sich bei einem Testniveau von $\alpha = 0.05$ ergeben, siehe Abbildung 6.9. Somit weist der White-Noise-Test die Abhängigkeit der Residuen nach. Das Modell der verborgenen Peri-

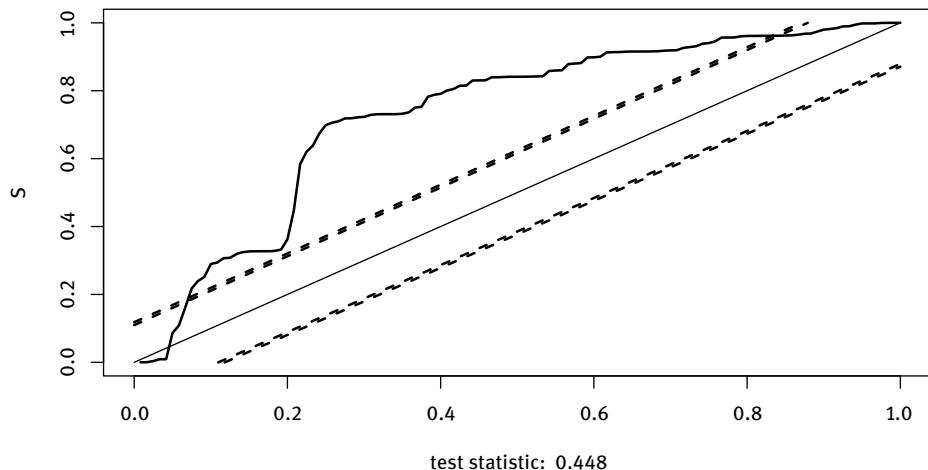


Abb. 6.9: White-Noise-Test der Residuen der Reihe SCHWEINEPREIS

odizitäten ist für diese Reihe also nicht passend. Das ist kein Wunder, wurde doch schon im Beispiel 6.11 das Vorhandensein einer weiteren Periodizität spekuliert.

R-Code:

```

1 t <- c(1:n)
2 co <- cos(2*pi*(1/48)*t)
3 si <- sin(2*pi*(1/48)*t)
4 out <- lm(e~co+si)
5 wntest(out$resid, 0.05, k=3)

```

In Fortsetzung des R-Codes von Beispiel 6.11 wird die Regression des Rest-Terms der Zerlegung der Reihe SCHWEINEPREIS durchgeführt. Dazu werden zuerst die beiden harmonischen Wellen als Regressoren definiert. Für die Residuen der Regression wird dann der Test `wntest` aus `tsapp` angewendet. Ein *P*-Wert kann in dieser Situation nicht berechnet werden.

6.4 Spektren

Definition und Eigenschaften

Es gilt gemäß (6.6) die folgende Beziehung zwischen der empirischen Autokovarianzfunktion und dem Periodogramm einer Zeitreihe:

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

Nach Satz 2.12 gilt approximativ $E(c_\tau) = \gamma_\tau$, wenn (γ_τ) die ACF des zugrundeliegenden Prozesses ist. Unter geeigneten, recht schwachen Bedingungen erhält man daraus für $N \rightarrow \infty$ die Approximation:

$$E(I(\lambda)) \approx \gamma_0 + 2 \sum_{\tau=1}^{N-1} \gamma_\tau \cos(2\pi\lambda\tau) \rightarrow \gamma_0 + 2 \sum_{\tau=1}^{\infty} \gamma_\tau \cos(2\pi\lambda\tau).$$

Definition 6.14 (Spektrum)

Der Prozess (Y_t) habe die ACF (γ_τ) , für die $\sum |\gamma_\tau| < \infty$ gelte. Dann heißt die Funktion

$$f(\lambda) = \gamma_0 + 2 \sum_{\tau=1}^{\infty} \gamma_\tau \cos(2\pi\lambda\tau) \quad (6.14)$$

das *Spektrum* oder die Spektraldichte von (Y_t) .

Das Spektrum eines Prozesses ist nicht negativ und stetig. Weiter erfüllt es die im folgenden Satz angegebene Beziehung.

Satz 6.15 (Parsevalsches Theorem)

Sei $f(\lambda)$ das Spektrum eines Prozesses mit der Autokovarianzfunktion (γ_τ) . Dann gilt die Umkehrformel

$$\gamma_\tau = \int_{-0.5}^{0.5} f(\lambda) \cos(2\pi\lambda\tau) d\lambda, \quad (6.15)$$

und speziell für die Varianz:

$$\gamma_0 = \int_{-0.5}^{0.5} f(\lambda) d\lambda. \quad (6.16)$$

Da das Spektrum symmetrisch um null ist, führt die Relation (6.16)t zu der Interpretation, dass der Anteil der Fläche unter $f(\lambda)$ zwischen zwei Frequenzpunkten an der Gesamtfläche von 0 bis 0.5 bis auf den Faktor 1/2 dem Anteil der Varianz entspricht, der auf harmonische Wellen mit Frequenzen aus diesem Bereich zurückgeht. Das ist konsistent damit, dass das Periodogramm $I(\lambda)$ die Stärke der Korrelation mit einer harmonischen Welle der entsprechenden Frequenz angibt. Je größer diese Korrelation ist, desto stärker ist die Welle in der Reihe präsent, desto größer ist auch die Spektraldichte an dieser Stelle. Dies ist vor dem Hintergrund der Cramér-Darstellung zu sehen, die ja sagt, dass jede stationäre Zeitreihe als Überlagerung von Sinus- und Kosinuswellen verschiedener Frequenzen dargestellt werden kann.

Auch gibt diese Relation Anlass zu dem Begriff *Varianzspektrum*. Wird in (6.14) von der Autokorrelationsfunktion ausgegangen, so ist die Fläche unter $f(\lambda)$ gleich eins. Somit besitzt $f(\lambda)$ dann die formalen Eigenschaften einer Dichtefunktion; $f(\lambda)$ heißt daher auch *Spektraldichte*. Zudem zeigt der Satz, dass das Spektrum die gleiche Information enthält wie die Autokovarianzfunktion, nur nach Frequenzen und nicht nach Lags geordnet.

Beispiel 6.16 (Weißes Licht)

Sei (Y_t) ein White-Noise-Prozess. Dieser hat dann das Spektrum $f(\lambda) = \sigma^2$. Das Spektrum ist also konstant. Dies ist auch der Grund für die Bezeichnung. Wie das „weiße“ Licht eine gleichmäßige Mischung aller Frequenzen ist, setzt sich der Prozess gleichmäßig aus Schwingungen aller Frequenzen zusammen.

Beispiel 6.17 (MA[1]- und AR[1]-Prozess)

Sei (Y_t) ein MA[1]-Prozess, $Y_t = \varepsilon_t - \beta\varepsilon_{t-1}$, mit $E(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$. Dann ergibt (6.14) mit (3.33):

$$f(\lambda) = \sigma_\varepsilon^2(1 + \beta^2) - 2\sigma_\varepsilon^2\beta \cos(2\pi\lambda) .$$

Für einen stationären AR[1]-Prozess $Y_t = \alpha Y_{t-1} + \varepsilon_t$ mit $E(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ führt (6.14) mit (3.5) zu:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{1 - \alpha^2} \left(1 + 2 \sum_{\tau=1}^{\infty} \alpha^\tau \cos(2\pi\lambda\tau) \right) = \frac{\sigma_\varepsilon^2}{1 + \alpha^2 - 2\alpha \cos(2\pi\lambda)} .$$

Die beiden Spektraldichten sind in Abbildung 6.10 illustriert.

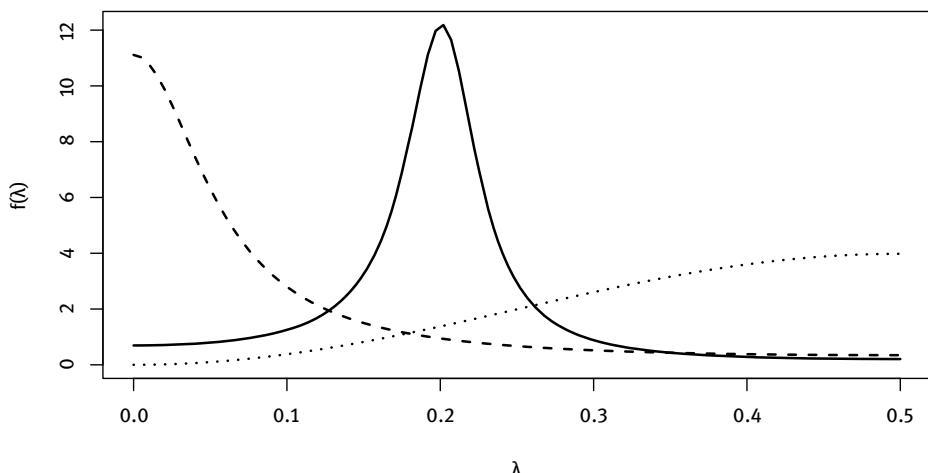


Abb. 6.10: Spektren von AR- und MA-Prozessen; MA[1] mit $\beta = 0.85$: ···, AR[1] mit $\alpha = 0.75$: ---, AR[2]: —

Beispiel 6.18 (Zeitreihen und ihre Spektralschätzungen)

Die sich aus dem Parsevalschen Theorem ergebende Interpretation in Bezug auf die Varianzanteile, die auf harmonische Wellen mit Frequenzen aus entsprechenden Frequenzbändern, zurückzuführen sind, lassen bei empirischen Zeitreihen die Gestalt der zugehörigen geschätzten Spektren schon vor der Durchführung der

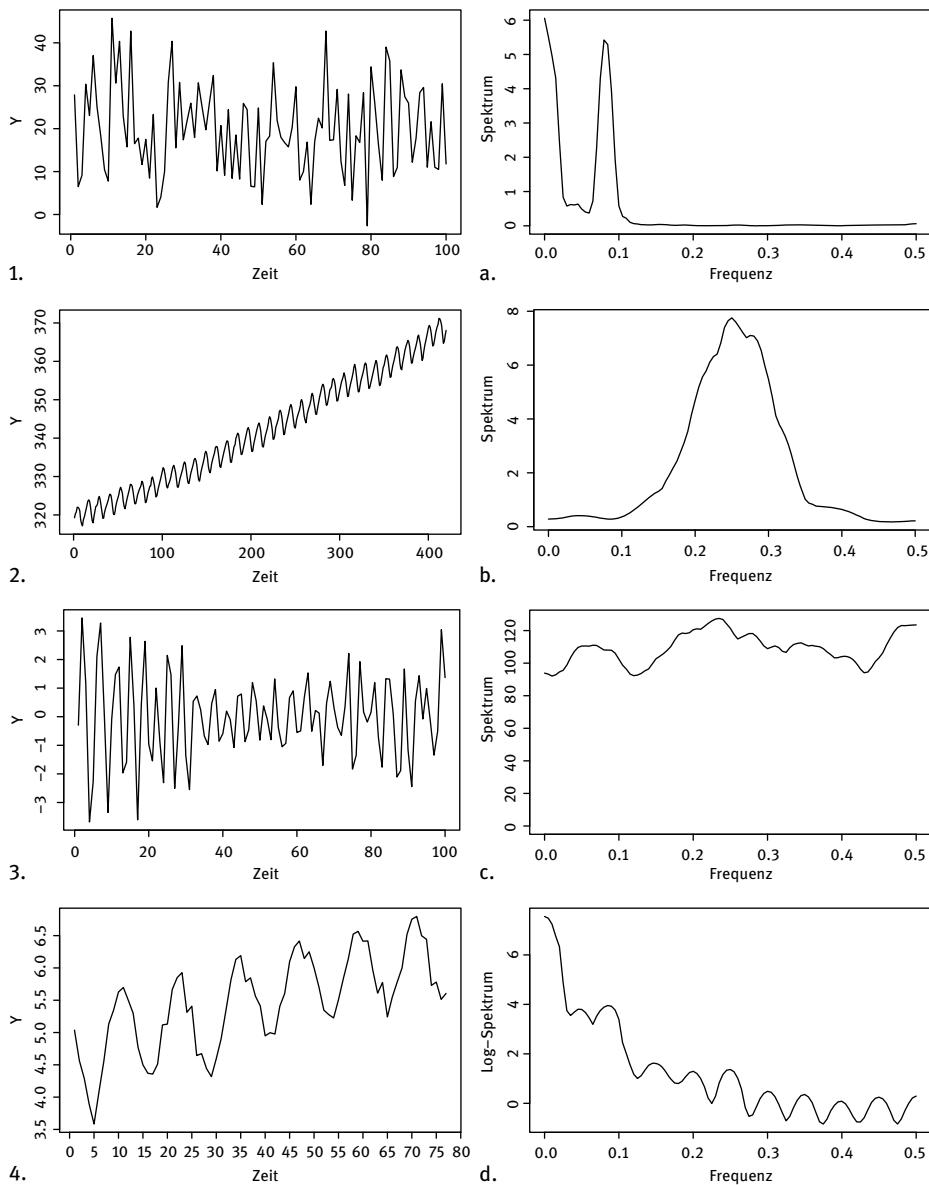


Abb. 6.11: Zeitreihen und Spektralschätzungen

Schätzung vorhersagen. Um die Leserin/ den Leser zu animieren, sich dies klar zu machen, sind in der Abbildung 6.11 die Zuordnungen der Zeitreihen und der zu gehörigen Spektralschätzungen durcheinander gebracht. (Auch wenn diese noch zu behandeln sind.) Versuchen Sie, die richtige Zuordnung zu bestimmen, bevor Sie weiterlesen.

Bei der Zeitreihe 1. lässt sich keine Struktur erkennen. Weder hat sie einen Trend noch eine Saison, auch ein zyklisches Verhalten lässt sich dem Plot nicht entnehmen. Sie macht also den Eindruck einer Realisation eines White-Noise-Prozesses. Die zugehörige Spektralschätzung sollte im Wesentlichen flach verlaufen. Die richtige Zuordnung ist daher c.

Die Zeitreihe 2. weist einen Trend und eine Saison auf. Der Trend führt zu erhöhten Werten der Spektralschätzung bei null. Die Saison bringt Erhöhungen bei der Saisonfrequenz mit sich und ggf. bei den Oberschwingungen. Im Verhältnis zum Trend werden diese Erhöhungen jedoch geringer sein, da der Trend stark dominiert. Dieses Muster zeigt gerade die Spektralschätzung d.

Einen Trend hat die Reihe 3. nicht, aber bei genauerer Betrachtung erkennt man, dass sie recht zyklisch verläuft. Bei der zugehörigen Frequenz wird die zugehörige Spektralschätzung einen Peak haben. Dies ist gerade bei b. zu sehen.

Reihe 4. ist der Reihe 2. von der Struktur her ähnlich. Jedoch ist der Trend weniger stark, dafür sind die Saisonausschläge stärker. Zudem hat die Saison eine recht sinusförmige Gestalt. Es ist also ein Spektrum mit einer Erhöhung bei null und einem ähnlich hohen Wert bei der Saisonfrequenz zu erwarten. Bei den Oberschwingungen wird eher wenig zu sehen sein. Die Spektralschätzung a. zeigt dieses Muster.

Lineare Filter im Frequenzbereich

Ein Bereich, in dem die Betrachtung von Spektren stochastischer Prozesse eine große Rolle spielt, ist die Untersuchung der Auswirkungen, die die Anwendung linearer Filter auf Zeitreihen haben, vgl. Definition 2.29. Die Anwendung eines Filters auf eine Zeitreihe führt zunächst zu einer eher komplizierten Beziehung zwischen der Autokovarianzfunktion des Inputs (Y_t) und des Outputs $Z_t = \sum a_u Y_{t-u}$:

$$\gamma_Z(\tau) = \sum_u \sum_v a_u a_v \gamma_Y(\tau + u - v).$$

Der Zusammenhang für die Spektren der beiden Prozesse (Y_t) und (Z_t) stellt sich wesentlich einfacher dar.

Um dies anzugeben, benötigt man den Begriff der *Fouriertransformation*. Dies wiederum setzt komplexe Zahlen voraus. In diesem Text wird so wenig wie möglich von ihnen Gebrauch gemacht; dennoch ist Folgendes wesentlich. Eine komplexe Zahl ist durch einen Realteil und einen Imaginärteil bestimmt, $z = a + bi$, wobei bi mit $i = \sqrt{-1}$ der Imaginärteil ist. Weiter gilt mit $|z| = \sqrt{a^2 + b^2}$ und $\tan(\phi) = b/a$ die Relation:

$$z = a + bi = |z| \cos(\phi) + |z| \sin(\phi)i = |z|e^{i\phi}.$$

Damit lässt sich die Fouriertransformation definieren.

Definition 6.19 (Fouriertransformation)

Die *Fouriertransformation* einer Folge (a_u) , ist definiert durch

$$F_a(\lambda) = \sum_u a_u e^{i2\pi\lambda u}. \quad (6.17)$$

Ist (a_u) ein linearer Filter, so wird die Fouriertransformation $F_a(\lambda)$ als *Transferfunktion* des Filters bezeichnet. Der Betrag der Transferfunktion heißt die *Gainfunktion* des Filters, $G_a(\lambda) = |F_a(\lambda)|$.

Das Spektrum erhält man mit diesen Vorbereitungen auch als Fouriertransformation der Folge (y_τ) :

$$f(\lambda) = \sum_{\tau=-\infty}^{\infty} y_\tau e^{i2\pi\lambda\tau}. \quad (6.18)$$

Satz 6.20 (Beziehung zwischen den Spektren von In- und Output eines linearen Filters)

Sei (Y_t) ein stationärer Prozess und sei $Z_t = \sum_{u=-\infty}^{\infty} a_u Y_{t-u}$ mit $\sum |a_u| < \infty$. Dann gilt:

$$f_Z(\lambda) = G_a(\lambda)^2 f_Y(\lambda). \quad (6.19)$$

Dabei ist $G_a(\lambda)$ die Gainfunktion des Filters.

Aus dem Satz erhält man für das Spektrum eines MA[q]-Prozesses $Y_t = \varepsilon_t - \sum_{u=1}^q \beta_u \varepsilon_{t-u}$, da für einen White-Noise-Prozess das Spektrum konstant σ_ε^2 ist:

$$f_Y(\lambda) = \left| 1 - \sum_{u=1}^q \beta_u e^{i2\pi\lambda u} \right|^2 \sigma_\varepsilon^2. \quad (6.20)$$

Für einen stationären AR[p]-Prozess $Y_t - \sum_{u=1}^p \alpha_u Y_{t-u} = \varepsilon_t$ gilt entsprechend:

$$\left| 1 - \sum_{u=1}^p \alpha_u e^{i2\pi\lambda u} \right|^2 f_Y(\lambda) = \sigma_\varepsilon^2. \quad (6.21)$$

Zusammengekommen erhält man für die Spektren stationärer und invertierbarer ARMA[p, q]-Prozesse $\alpha(B)Y_t = \beta(B)\varepsilon_t$:

$$f(\lambda) = \frac{\left| 1 - \sum_{u=1}^q \beta_u e^{i2\pi\lambda u} \right|^2}{\left| 1 - \sum_{u=1}^p \alpha_u e^{i2\pi\lambda u} \right|^2} \sigma_\varepsilon^2. \quad (6.22)$$

Beispiel 6.21 (AR[2]-Prozess)

Für einen AR[2]-Prozess ergibt (6.21)

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{1 + \alpha_1^2 + \alpha_2^2 + 2(\alpha_1 \alpha_2 - \alpha_1) \cos(2\pi\lambda) - 2\alpha_2 \cos(4\pi\lambda)}. \quad$$

Für die Parameter $\alpha_1 = 0.5$, $\alpha_2 = -0.75$ ist das zugehörige Spektrum in Abbildung 6.10 mit dargestellt. Es lässt erahnen, dass AR[2]-Prozesse sehr periodische

Reihen hervorbringen können. Dies lässt sich auch formal belegen. Die Differenzengleichung

$$y_t = \alpha y_{t-1} - y_{t-2}$$

hat nämlich als Lösung die Funktion $y_t = a \cdot \sin(2\pi\lambda t + b)$, wobei a, b und λ in einer geeigneten Beziehung zu α stehen. Ein AR[2]-Prozess resultiert nun aus der gestörten Beobachtung der Differenzengleichung. Yule, auf den die AR-Modelle zurückgehen, hat dafür in seiner Arbeit (1927) den folgenden Rahmen gewählt: Die Bewegung eines Pendels wird in gleichen Zeitabständen gemessen; dadurch wird eine reine trigonometrische Schwingung beschrieben. Dann wird die Messvorrichtung unbeaufsichtigt gelassen. Unglücklicherweise kommt ein Junge in den Raum und beginnt, mit einem Blasrohr Erbsen auf das Pendel zu schießen, mal von der einen, mal von der anderen Seite. Dadurch ändern sich jeweils Amplitude und Phase der Schwingung zufällig.

Über die Gainfunktionen werden Filter klassifiziert. Speziell lassen *Low-pass Filter* nur langwellige Schwingungskomponenten passieren. Das sind ja gerade die mit kleinen Frequenzen. *High-pass Filter* eliminieren dagegen langwellige Schwingungen und lassen nur die Komponenten mit hohen Frequenzen in der Reihe.

Beispiel 6.22 (Gleitender Durchschnitt und einfache Differenzen)

Der gleitende 3er Durchschnitt

$$a_u = \begin{cases} 1/3 & \text{für } u = -1, 0, 1 \\ 0 & \text{sonst} \end{cases}$$

hat die Gainfunktion

$$G_a(\lambda) = \frac{1}{3} \cdot |1 + 2 \cos(2\pi\lambda)| .$$

Wie die Darstellung der Gainfunktion in der Abbildung 6.12 zeigt, wird bei der Multiplikation des Ausgangsspektrums mit $G_a(\lambda)^2$ die Frequenz $1/3$ ganz ausgeblendet. Das ist einsichtig, da Dreierzyklen bei der Zusammenfassung von je 3 aufeinander folgenden Werten nicht mehr sichtbar sind. Zudem ist $G_a(\lambda)$ für höhere Frequenzen recht klein; die zugehörigen Schwingungen werden also gedämpft. Im Niederfrequenzbereich ($\lambda \approx 0$) bleibt das Spektrum fast unverändert, der Filter lässt also niedrige Frequenzen passieren. Somit ist er ein Low-pass Filter.

Den ersten Differenzen entsprechen die Filtergewichte $b_u = 1$ für $u = 0$, $b_u = -1$ für $u = 1$ und $b_u = 0$ sonst. Sie bilden einen High-pass Filter. Seine Transferfunktion ist $F_b(\lambda) = e^{i2\pi\lambda 0} - e^{i2\pi\lambda 1} = 1 - e^{i2\pi\lambda}$. Damit erhält man für die Gainfunktion

$$|F_b(\lambda)| = 2 \sqrt{(1 - \cos(2\pi\lambda))/2} = 2 \sin(\pi\lambda) .$$

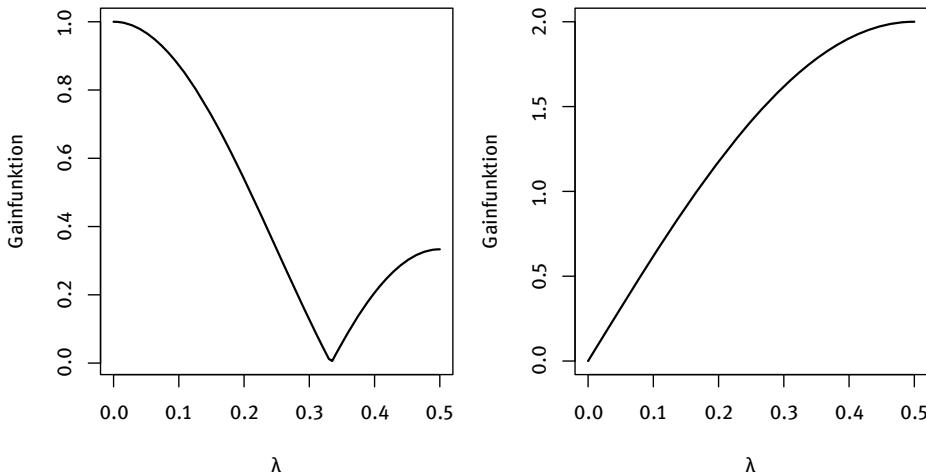


Abb. 6.12: Gainfunktionen, Links: gleitender 3er-Durchschnitt, rechts: einfache Differenzen

Ein idealer Low-pass-Filter (f_u), der nur Frequenzen $\lambda \leq \lambda_c$ passieren lässt, in anderer Sprechweise die Cutoff-Frequenz λ_c hat, ist offensichtlich durch folgende Gainfunktion charakterisiert:

$$H_f(\lambda) = \begin{cases} 1 & \text{für } \lambda \leq \lambda_c \\ 0 & \text{für } \lambda > \lambda_c . \end{cases} \quad (6.23)$$

Solch ein idealer Filter existiert nicht; man kann ihn aber beliebig gut approximieren, wenn ein genügend komplexer Filter verwendet wird. Ein üblicher Zugang zur Konstruktion eines Filters mit vorgegebener Anzahl von Gewichten ist der Ansatz der Kleinsten Quadrate. Dieser führt auf Filter der Form

$$h_u = \begin{cases} c \cdot \frac{\sin(2\pi\lambda_c u)}{\pi u} \frac{\sin(2\pi u/(2q+1))}{2\pi u/(2q+1)} & \text{für } |u| = 1, \dots, q \\ c \cdot 2\lambda_c & \text{für } |u| = 0 . \end{cases} \quad (6.24)$$

Dabei ist λ_c die Cutoff-Frequenz, bis zu der die Gainfunktion praktisch eins sein soll. Wird ein solcher Filter modifiziert gemäß

$$g_u = h_u \cdot \cos(2\pi\lambda_0 u) , \quad (6.25)$$

so ist (g_u) ein bei der Frequenz λ_0 zentrierter *Band-pass-Filter*. Diese Modifikation wird als *komplexe De- und Remodulation* bezeichnet.

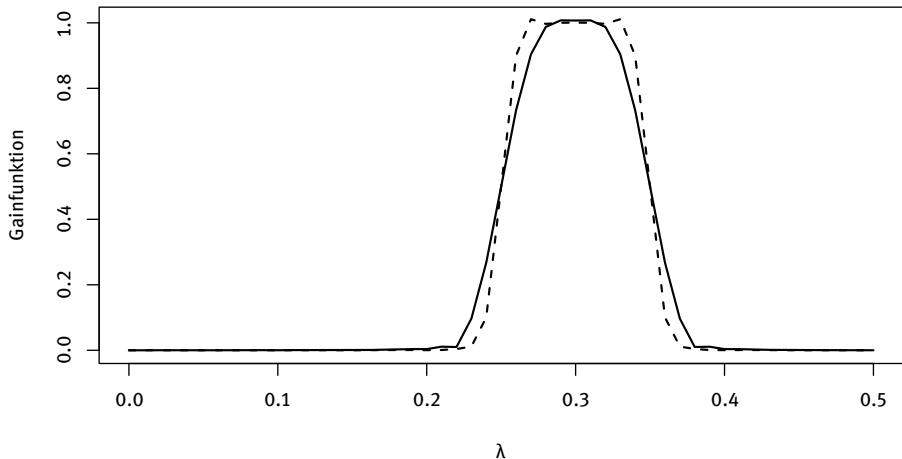


Abb. 6.13: Gainfunktion eines KQ-Band-pass-Filters: $\lambda_0 = 0.3$, Cutoff-Frequenz $\lambda_c = 0.05$, $q = 20$ (—) und $q = 40$ (- - -)

Beispiel 6.23 (Bruttoinlandsprodukt)

In der Volkswirtschaftlichen Gesamtrechnung wird das preisbereinigte Bruttoinlandsprodukt aus den Daten des Statistischen Bundesamtes ermittelt. Die vorliegende Zeitreihe BIPORIG besteht aus Laspeyres-Indexwerten zur Basis 2000. Die Reihe wurde Band-Pass-gefiltert, so dass nur Periodizitäten zwischen drei und fünf den Filter passieren sollten. Bei Quartalsdaten entspricht das einer Saisonbestimmung. In der Abbildung 6.14 ist die Zeitreihe dargestellt. Die Band-pass-

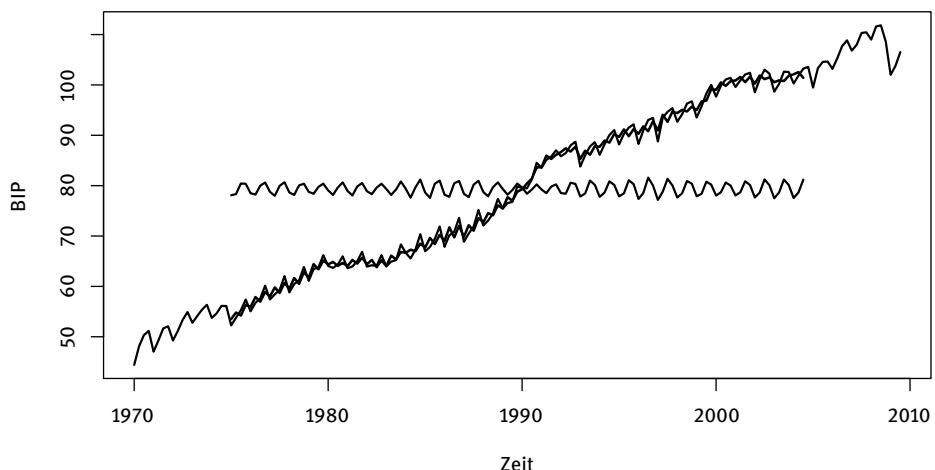


Abb. 6.14: Komplexe De- und Remodulation: Bruttoinlandsprodukt (preisbereinigt) mit Schwingungskomponenten für $1/5 < \lambda < 1/3$ und Differenz

gefilterte Version wurde aus Darstellungsgründen um das arithmetische Mittel der Ausgangsreihe verschoben dargestellt. Die entsprechende Differenz ist verstärkt eingezzeichnet.

R-Code:

```

1 data(GDPORIG)
2 y <- GDPORIG
3 yf <- bandfilt(y-mean(y), 20, 3, 5)
4 plot(y,xlab="Zeit",ylab="BIP")
5 lines(yf+mean(y))
6 lines(y-yf)

```

In der Zeile 4 wird die komplexe De- und Remodulation durchgeführt. Die nachfolgenden Befehle dienen zur grafischen Darstellung.

Beispiel 6.24 (Berliner Verfahren)

Die zentralen Filter des Berliner Verfahrens BV4.1 zur Saisonbereinigung von monatlichen Zeitreihen haben die in der Abbildung 6.15 dargestellten Gainfunktionen.

Man sieht, dass die Filter die Saisonfrequenz und die Frequenzen der Oberschwingungen vollständig ausblenden.

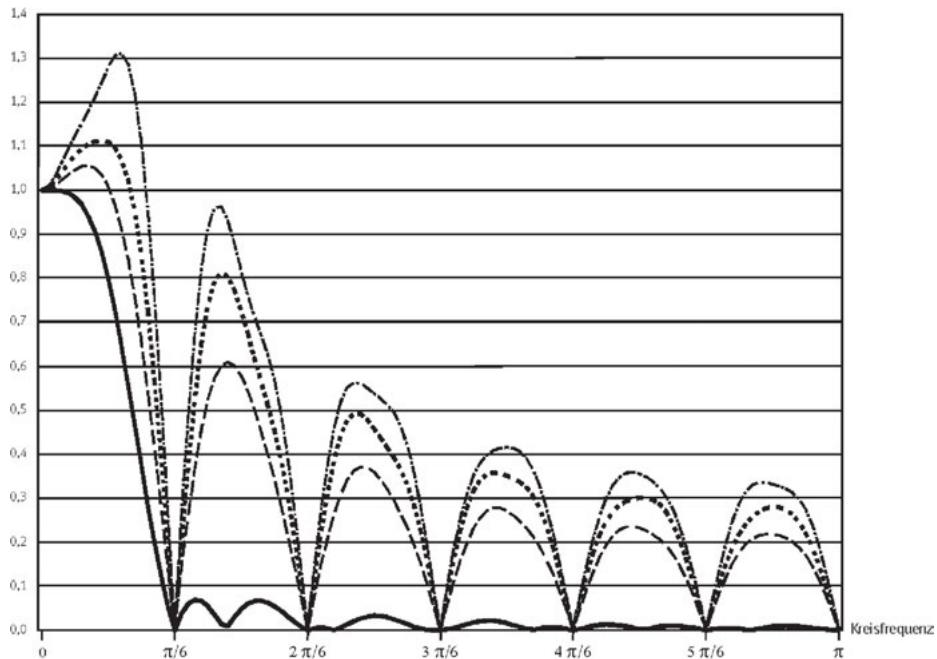


Abb. 6.15: Gainfunktionen der Filter des Berliner Verfahrens; aus Speth (2004)

Der *Baxter-King-Filter* ist ebenfalls ein Band-pass-Filter. Er ist speziell auf ökonomische Gegebenheiten zugeschnitten. Das Interesse besteht in der Bestimmung von Konjunkturzyklen. Dazu wird festgelegt, in welchem Periodenbereich ein solcher angesiedelt ist und die Zeitreihe entsprechend gefiltert. Zu diesem Filter sei auf Baxter & King (1999) verwiesen.

Beispiel 6.25 (Bruttoinlandsprodukt – Fortsetzung von Seite 156)

Die in der Abbildung 6.16 dargestellte Saisonbereinigung der Zeitreihe BIPORIG wurde mit der Funktion `bkfilter` aus dem Paket `mFilter` erzeugt.

R-Code:

```
1 library(mFilter)
2 ybk <- bkfilter(y, pl=3, pu=5)
3 plot(ybk)
```

Die Funktion `bkfilter` erlaubt zusätzlich die Bereinigung eines deterministischen Trends. Dies erfolgt mit der Option `drift=TRUE`.

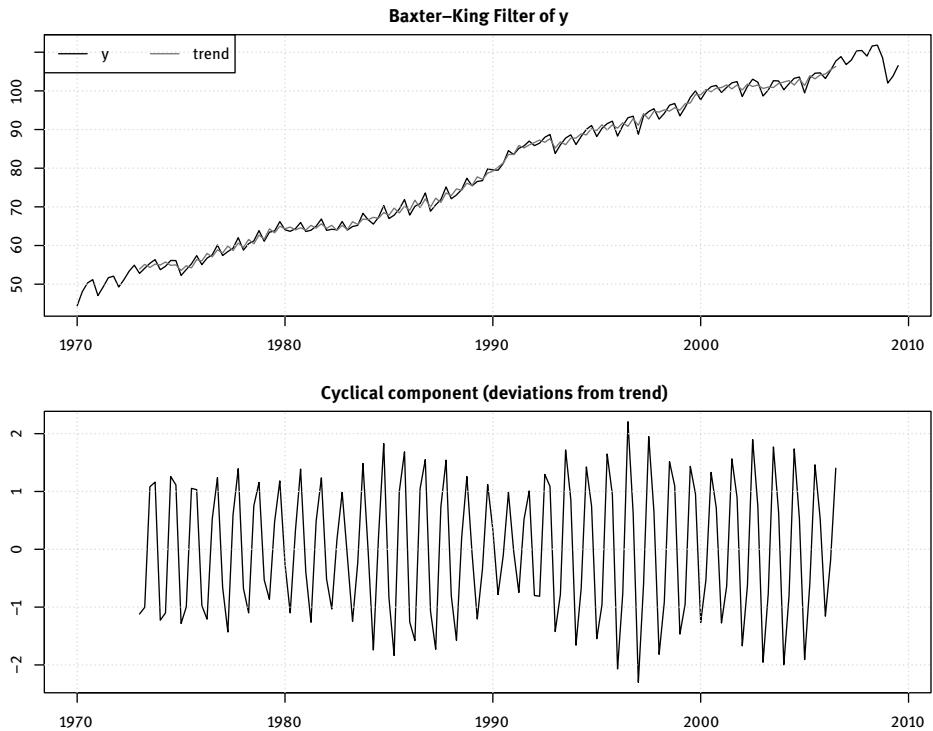


Abb. 6.16: Baxter-King-Filter: Bruttoinlandsprodukt (preisbereinigt) mit Schwingungskomponenten für $1/5 < \lambda < 1/3$ und Differenz

Häufig wird nicht nur ein linearer Filter auf eine Reihe angewendet, sondern mehrere. Das Studium des Zusammenwirks ist im Frequenzbereich besonders einfach, da die Hintereinanderausführung von linearen Filtern einen neuen linearen Filter ergibt; dessen Gainfunktion ist einfach das Produkt der einzelnen Gainfunktionen.

So erhält man auch eine einfache Begründung des so genannten *Slutzky-Effektes*. Damit wird die Tatsache bezeichnet, dass durch mehrfaches Anwenden von gleitenden Mitteln und Differenzenbilden aus einem White-Noise-Prozess eine Reihe mit deutlichem periodischem Schwingungsverhalten erzeugt werden kann. Bei dem Versuch, Trend und Saison zu eliminieren, kann es also vorkommen, dass Zyklen in die Reihe hineingerechnet werden. So ist die in der Abbildung 6.17 oben dargestellte White-Noise-Reihe mittels zehnmaliger Anwendung von einfachen Differenzen und gleitenden Durchschnitten in die unten abgebildete transformiert worden.

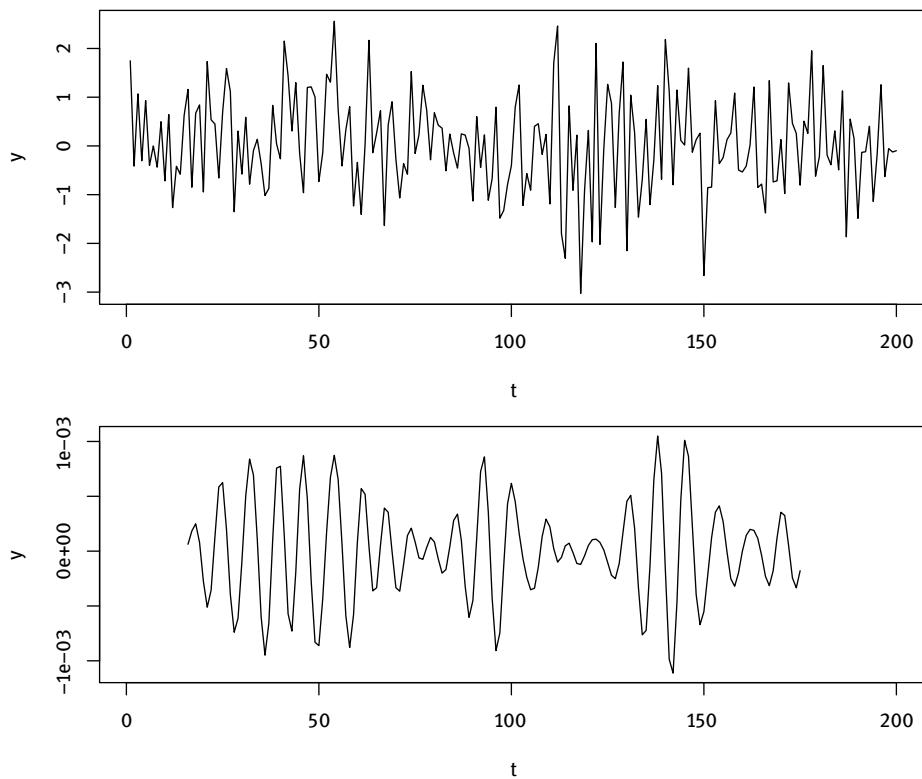


Abb. 6.17: Zum Slutsky-Effekt

6.5 Spektralschätzung

Um die Probleme beim *Schätzen des Spektrums* eines linearen Prozesses $Y_t = \sum_{u=0}^{\infty} \beta_u \varepsilon_{t-u}$ mit $\sum |\beta_u| < \infty$ zu verstehen, ist es notwendig, sich die beiden wichtigsten Eigenschaften des Periodogramms, das aus einer Realisation eines solchen Prozesses bestimmt wird, klar zu machen. Ist $G_\beta(\lambda)$ die Gainfunktion des Filters (β_u), so gilt wegen $f_Y(\lambda) = G_\beta(\lambda)^2 \sigma_\varepsilon^2$ für das Periodogramm einer Realisation von (Y_t):

$$I_Y(\lambda) \approx G_\beta(\lambda)^2 \cdot I_\varepsilon(\lambda).$$

Das hat zur Folge, dass an Fourierfrequenzen $\lambda_j, \lambda_{j'}, j \neq j'$ gilt:

$$I_Y(\lambda_j) \approx \mathcal{E}(f(\lambda_j)) \quad (6.26)$$

$$I_Y(\lambda_j), I_Y(\lambda_{j'}) \text{ sind approximativ unabhängig.} \quad (6.27)$$

Bei der Exponentialverteilung $\mathcal{E}(\theta)$ mit der Dichte $p(u) = e^{-u/\theta}/\theta$, $u \geq 0$, ist der Erwartungswert gleich dem Kehrwert des Parameters und die Varianz gleich dem Kehrwert des Quadrates des Parameters. Daher stimmt das Periodogramm zwar im Mittel mit dem Spektrum überein, die Varianzen der Periodogrammordinaten bleiben aber auch bei wachsender Reihenlänge von der Größenordnung $f(\lambda)^2$. Das steht im Gegensatz zu den üblichen Eigenschaften von Schätzern, deren Verteilungen sich bei wachsendem Stichprobenumfang immer enger um den zugehörigen theoretischen Parameter zusammenziehen. Zudem erhält man, auch wenn $f_Y(\lambda)$ stetig und mithin glatt ist, wegen der asymptotischen Unabhängigkeit sogar bei beliebig großem N ein unruhiges Bild, wenn das Periodogramm gezeichnet wird. Dieses andersartige Verhalten ist darin begründet, dass mit wachsender Reihenlänge die Anzahl der Fourierfrequenzen steigt. Somit liefert die Verlängerung der Reihe an immer mehr Frequenzpunkten Information über das Spektrum. An einer speziellen Frequenz wird sie allerdings nicht präziser.

Direkte Spektralschätzung

Das Periodogramm eignet sich nach dem eben Ausgeführten offensichtlich nicht als Schätzer für das Spektrum. Die approximative Unabhängigkeit benachbarter Periodogrammordinaten und ihre zunehmende Dichte erlaubt aber, konsistente Schätzer über die Bildung von lokalen Mittelwerten zu konstruieren.

Definition 6.26 (Direkter Spektralschätzer)

Ein *direkter Spektralschätzer* ist eine Schätzung der Spektraldichte von der Form

$$\hat{f}(\lambda) = \sum_{u=-q}^q w_u I\left(\lambda + \frac{u}{N}\right). \quad (6.28)$$

Derartige Folgen von Gewichten (w_u) werden als *diskrete Periodogrammfenster* bezeichnet.

Mit den Beziehungen (6.26), (6.27) gelten die Näherungen

$$\mathrm{E}(\hat{f}(\lambda)) \approx \sum_{u=-q}^q w_u f\left(\lambda + \frac{u}{N}\right) \quad (6.29)$$

$$\mathrm{Var}(\hat{f}(\lambda)) \approx \sum_{u=-q}^q w_u^2 f\left(\lambda + \frac{u}{N}\right)^2. \quad (6.30)$$

Damit ein direkter Spektralschätzer erwartungstreue wird, also $\mathrm{E}(\hat{f}(\lambda)) = f(\lambda)$ gilt, muss die Summe der Gewichte eins ergeben. Denn in einer (kleinen) Umgebung der Frequenz λ stimmt das Spektrum mit $f(\lambda)$ wegen der Stetigkeit näherungsweise überein, $f(\lambda + u/N) \approx f(\lambda)$.

Zudem soll die Varianz eines Spektralschätzers mit wachsender Reihenlänge gegen null gehen, $\mathrm{Var}(\hat{f}(\lambda)) \rightarrow 0$. Mit $\sum w_u = 1$ und $w_u \geq 0$ geht die Summe der quadrierten Koeffizienten gegen null, wenn die Anzahl der Koeffizienten größer wird. Nun darf die Anzahl der Gewichte nicht zu schnell wachsen, da sonst die asymptotische Erwartungstreue nicht zu erreichen wäre. Daher verlangt man $q \rightarrow \infty$ und $q/N \rightarrow 0$ für $N \rightarrow \infty$. Eine mögliche Wahl ist beispielsweise $q = \sqrt{N}$.

Beispiel 6.27 (Daniell- und Dreieckfenster)

Zwei Beispiele sind das Daniell-Fenster

$$w_u = \frac{1}{2q+1}, \quad (u = -q, \dots, 0, \dots, q)$$

und das Dreieckfenster

$$w_u = \frac{q+1-|u|}{(q+1)^2}, \quad (u = -q, \dots, 0, \dots, q).$$

Für diese beiden Fenster ist $\sum_{u=-q}^q w_u = 1$ erfüllt.

Beim Daniell-Fenster ist zudem

$$\sum_{u=-q}^q w_u^2 = (2q+1) \frac{1}{(2q+1)^2} = \frac{1}{2q+1}.$$

Somit geht die Varianz bei der Wahl $q = \sqrt{N}$ gegen null.

Das Fenster von Daniell ist recht nahe liegend. Es ist sogar asymptotisch varianzminimal. Dennoch ist es für die Praxis nicht zu empfehlen. Diese Nicht-Empfehlung beruht auf Eigenschaften, die mit dem Bias zusammenhängen.

Der *Bias*, d. h. die systematische Verfälschung eines Spektralschätzers wird durch die Differenz $\mathrm{E}(\hat{f}(\lambda)) - f(\lambda)$ erfasst. Dieser Bias wird aus zwei Quellen gespeist. Einmal kommt er dadurch zustande, dass über einen nichtverschwindenden Bereich von Frequenzen gemittelt wird. Insbesondere in Bereichen, wo das Spektrum sich stark ändert, wird die Mittelung nicht unbedingt den korrekten Wert treffen. Speziell werden Gipfel abgeflacht und Täler angehoben. Zum zweiten resultiert der Bias auch aus dem

Leakage, der Eigenschaft des Periodogramms, nicht nur an den Stellen eine Erhöhung der Ordinatenwerte zu zeigen, an denen tatsächlich eine harmonische Komponente vorliegt, sondern auch an benachbarten Frequenzen.

Die erstgenannte Quelle der Verfälschung wird durch die *Bandbreite* eines Spektralschätzers beschrieben. Anschaulich ist zunächst einmal die Bandbreite des Daniell-Fensters; sie ist nämlich $1/(N \sum w_u^2) = (2q + 1)/N$. Das ist (bis auf den Faktor 2) der Anteil der Fourierfrequenzen, über die bei der Schätzung gemittelt wird. Diese allgemeine Definition geht auf Parzen zurück.

Definition 6.28 (Äquivalente Bandbreite)

Sei $\hat{f}(\lambda) = \sum w_u I(\lambda + u/N)$ ein direkter Spektralschätzer mit $\sum w_u = 1$. Seine *äquivalente Bandbreite EBW* ist

$$EBW(\hat{f}(\lambda)) = \frac{1}{N \cdot \sum w_u^2}. \quad (6.31)$$

Das Maß *EBW* kann als derjenige Abstand zweier Peaks im Spektrum interpretiert werden, die vom Schätzer noch getrennt werden können. Damit zeigt sich das Dilemma der Spektralschätzung. Wie die beiden Gleichungen (6.30) und (6.31) zeigen, kann nicht gleichzeitig die Varianz und die Bandbreite eines Spektralschätzers klein gehalten werden. Vielmehr gilt die *Unschärferelation der Spektralschätzung*:

$$EBW(\hat{f}(\lambda)) \cdot \text{Var}(\hat{f}(\lambda)) = \frac{f^2(\lambda)}{N}. \quad (6.32)$$

Kurz gesagt ist also für jedes $0 \leq \lambda \leq 0.5$

$$\text{Bandbreite} \cdot \text{Varianz} = \text{konstant}.$$

Diese Beziehung legt die Schwäche des Daniell-Fensters offen. Es hat zwar eine sehr kleine Varianz, dafür aber eine große Bandbreite. Spektrale Peaks lassen sich bei Verwendung dieses Fensters nicht gut trennen. Dementsprechend wählt man andere Fenster, bevorzugt solche, die im Zentrum erhöht sind und symmetrisch abfallen.

Beispiel 6.29 (Faltungen des Daniell-Fensters und direktes Bartlett-Priestley-Fenster)

Eine Möglichkeit das Daniell-Fenster zu verbessern, besteht in der Faltung mehrerer Daniell-Fenster. Tatsächlich lässt sich das Dreieck-Fenster durch die Hintereinanderausführung zweier Daniell-Fenster gewinnen. In der Abbildung 6.18 wurden Daniell-Fenster der Längen 9, 9, 11 und 13 gefaltet. Dabei sind die einzelnen Fenster in der Weise modifiziert, dass die Gewichte an den Rändern halbiert sind. Dies ist in Abbildung 6.18 dargestellt.

Das direkte Bartlett-Priestley-Fenster ist gegeben durch:

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

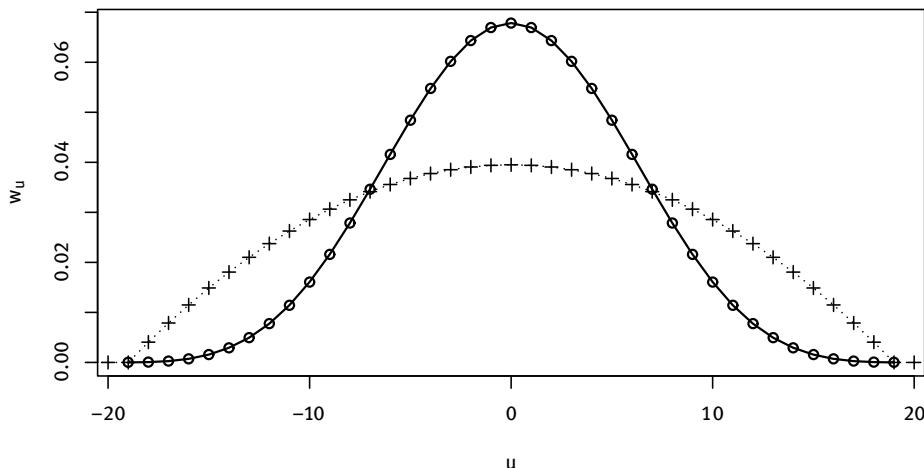


Abb. 6.18: Faltung von vier modifizierten Daniell-Fenstern (+) und direktes Bartlett-Priestley-Fenster (•)

In der folgenden Tabelle sind für dieses und die beiden anderen angesprochenen Fenster die relevanten Summen der quadrierten Gewichte für verschiedene q angegeben.

q	$\sum w_u^2$				
	Daniell-F.	Dreieck-F.	Bartlett-Priestley-F.	gefaltete Daniell-F.	Länge der einzelnen F.
15	0.03226	0.04454	0.04009	0.06369	5,7,9,11
20	0.02439	0.03338	0.03004	0.04908	9,9,11,13
25	0.01961	0.02669	0.02402	0.04379	9,9,11,11,13

Man sieht, dass das Bartlett-Priestley-Fenster eine Mittelstellung zwischen dem Daniell-Fenster und dem Dreieck-Fenster einnimmt. Damit hat es zwar eine etwas größere Varianz als das Daniell-Fenster, kann aber auch spektrale Peaks besser trennen.

Für die Anwendung resultiert aus dieser Diskussion die Vorgehensweise, dass i. d. R. Spektralschätzungen für verschiedene Bandbreiten durchgeführt werden. In Abhängigkeit von der gewünschten Auflösung spektraler Peaks und der Glattheit der Spektralschätzung wählt man die geeignete EBW . Sofern eines der elaborierteren Fenster verwendet wird, ist der Einfluss des Fensters erfahrungsgemäß weniger bedeutsam.

Die andere Quelle der Verfälschung, das Leakage, resultiert wie auf Seite 142 ausgeführt daher, dass nur ein endlicher Ausschnitt aus der unendlichen Realisation des Prozesses beobachtet wird. Um den Effekt des endlichen Ausschnittes zu mildern, geht man von der Vorstellung aus, dass der nicht beobachtete Teil identisch mit dem

Mittelwert des beobachteten Teils ist. Dann kann der Leakage-Effekt gemildert werden, indem ein glatter Übergang des beobachteten in den nicht beobachteten Teil der Zeitreihe hergestellt wird. Dazu dienen *Taper* oder *Datenfenster*. Die Anwendung eines Tapers wird als *Tapermodifikation* bezeichnet. Mit Datenfenstern werden die Ränder einer zentrierten Zeitreihe heruntergewichtet.

Ein Beispiel ist das Kosinus-Taper; es besteht etwa am Beginn aus einem aufsteigenden Teil einer Kosinusfunktion, ist in der Mitte konstant eins und am rechten Rand gleich dem absteigenden Teil einer Kosinusfunktion:

$$h_t = \begin{cases} \frac{1}{2} \left(1 - \cos \left(\frac{t\pi}{m} \right) \right) & t = 1, 2, \dots, m \\ 1 & t = m + 1, m + 2, \dots, N - m \\ \frac{1}{2} \left(1 - \cos \left(\frac{(N+1-t)\pi}{m} \right) \right) & t = N - m + 1, \dots, N. \end{cases} \quad (6.33)$$

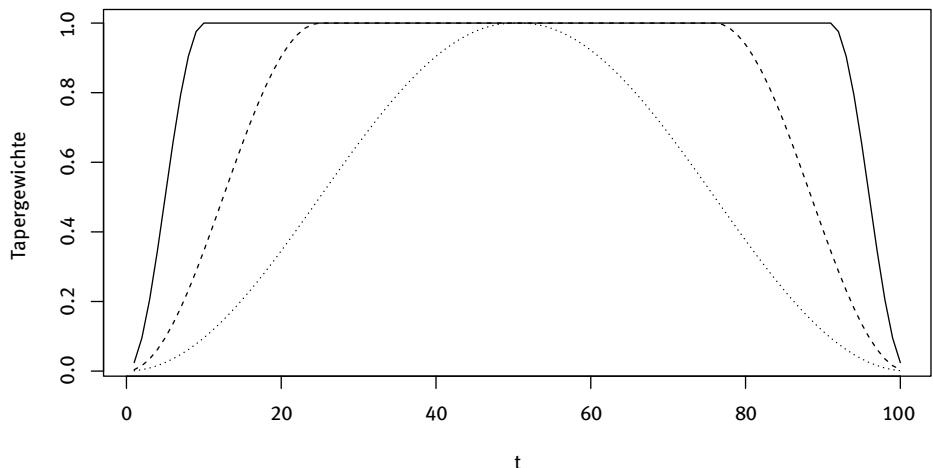


Abb. 6.19: Tapergewichte, Taperanteil: – 10 %, – – 25 % und - - - 50 %

Beispiel 6.30 (Tabellenplatz HSV – Fortsetzung von Seite 62)

Für die in der Abbildung 2.24 dargestellte Zeitreihe wird eine Spektralschätzung durchgeführt. Dazu wird die Originalreihe, die zu 25 % und die zu 50 % getaperierten Versionen verwendet. Die Dichteschätzungen werden mit einer Bandbreite von 0.025 an den Fourierfrequenzen berechnet. Man sieht, dass höherer Anteil von tapermodifizierten Werten zu mehr spektraler Masse am peak der Spektralschätzung führt, während der Rest erst bei 50 % Tapermodifikation deutlicher gedämpft wird. Dabei wurde die normierte Version der Spektralschätzung verwendet. Sonst würde die Spektralschätzung wegen der geringeren Varianz der Reihe insgesamt flacher verlaufen.

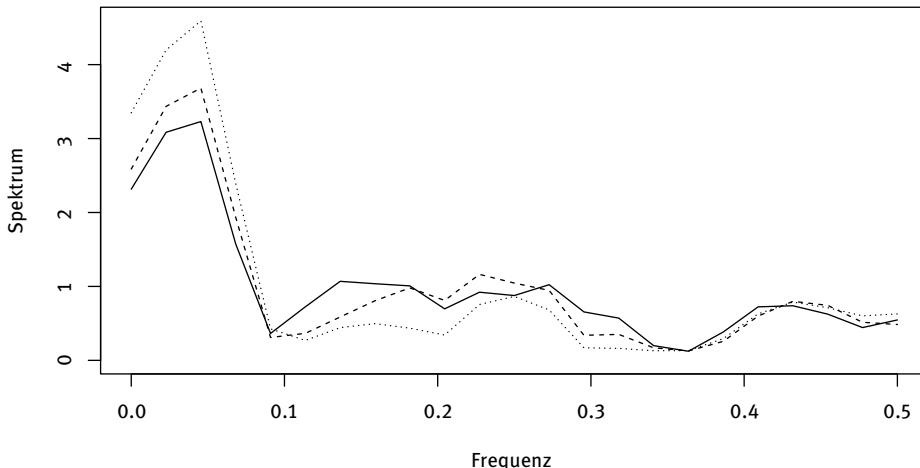


Abb. 6.20: Spektralschätzung der Reihe HSV, — original, – – 25 % und - · - 50 % getapert

Der Peak ist bei der Frequenz $\lambda = 0.04545$ lokalisiert. Das entspricht einer Periode von $1/\lambda = 22$, genau der halben Reihenlänge.

Die asymptotische Verteilung von direkten Spektralschätzern lässt sich aus der asymptotischen Verteilung des Periodogramms herleiten. Sie lässt sich zur Gewinnung von Konfidenzintervallen verwenden. Hier wird nur das für die Umsetzung relevante Ergebnis angegeben.

Satz 6.31 (Konfidenzgrenzen für das Spektrum)

Sei (Y_t) ein linearer Prozess mit einer genügend schnell gegen Null gehenden Koeffizientenfolge und Spektrum $f(\lambda) > 0$. $\hat{f}(\lambda) = \sum_{u=-q_N}^{q_N} w_u I(\lambda - u/N)$ sei ein direkter Spektralschätzer. Dann gilt für festes $\lambda_j = j/N$:

$$P\left(\frac{\hat{f}(\lambda_j)v}{\chi^2_{v;1-\alpha/2}} \leq f(\lambda_j) \leq \frac{\hat{f}(\lambda_j)v}{\chi^2_{v;\alpha/2}}\right) = 1 - \alpha. \quad (6.34)$$

Die Anzahl v der Freiheitsgrade der χ^2 -Verteilung ist dabei $v = 2/(\sum_{u=-q_N}^{q_N} w_u^2)$.

Wie der Satz lehrt, hat man die Spektralschätzung mit den durch den Satz gegebenen Grenzen zu zeichnen, um ein Konfidenzband zu erhalten. Dabei ist zu beachten, dass nur für eine einzelne Frequenz die Wahrscheinlichkeit approximativ $1 - \alpha$ beträgt, dass das Konfidenzintervall den unbekannten Wert des Spektrums einschließt. Somit gibt das Konfidenzband als Ganzes einen eher heuristischen Anhaltspunkt über die Zuverlässigkeit der Spektralschätzung.

Beispiel 6.32 (Sekretions-Rhythmus von Hormonen – Fortsetzung von Seite 2)

Bei einem Bullen wurde viertelstündlich von 0.00 Uhr bis 24.00 Uhr die Konzentration des Wachstumshormons (in ng/ml) im Blutplasma ermittelt. Die Zeitreihe besteht also aus 97 Werten. Um das zyklische Verhalten zu untersuchen, wird für die um den ersten Wert reduzierte Reihe WHORMON eine Spektralschätzung ermittelt. Bei einer Länge von $n = 96$ lassen sich die Frequenzen direkt in interpretierbare Perioden umrechnen. Die Zeitreihe und die Spektralschätzung mit zugehörigen 0.95-Konfidenzgrenzen ist in der Abbildung 6.21 dargestellt.

Bei $\lambda = 0.04166667$ ist der erste Peak im Spektrum zu erkennen. Diese Frequenz entspricht einer Periode von $1/0.04166667 = 24$, bei 1/4-Stundenwerten also von sechs Stunden. Der zweite, kleinere ist bei $\lambda = 0.16666667$; die zugehörige Periode beträgt eineinhalb Stunden. Biologen bzw. Mediziner müssen entscheiden, ob dies eine eigenständige Periodizität ist oder ob es sich um eine Oberschwingung handelt.

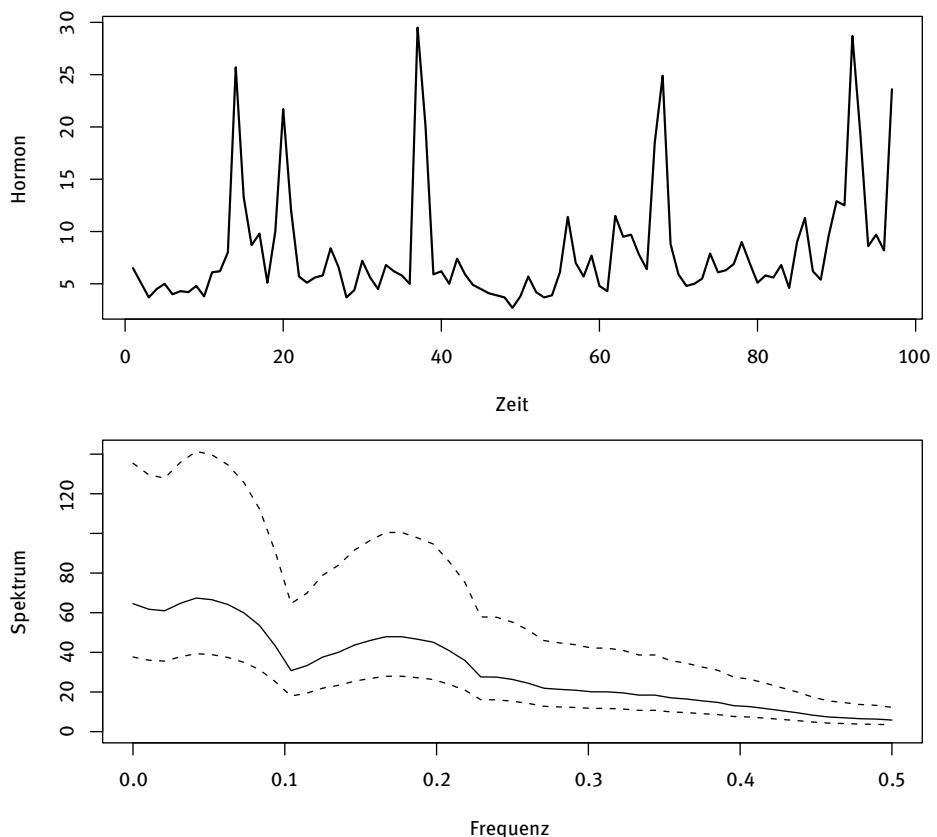


Abb. 6.21: Reihe WHORMON (Wachstumshormon eines Bullen) und zugehörige Spektralschätzung

R-Code:

```

1 data(WHORMONE)
2 y <- WHORMONE
3 out <- specest(y[-1], 48, 0.1, win="perwinba", conf=0.95)
4 par(mfrow=c(2,1))
5 plot(y)
6 specplot(out)
7 par(mfrow=c(1,1))

```

Für die Spektralschätzung in Zeile 3 wird die Funktion `specest` aus `tsapp` verwendet. Sie ermittelt zugleich die Konfidenzgrenzen. Mit der Option `win="perwinba"` wird das direkte Bartlett-Priestley-Fenster angefordert. Im Paket `tsapp` befindet sich auch die Funktion `specplot` zur grafischen Darstellung dieser Spektralschätzung.

Weitere Ansätze zur Spektralschätzung

Indirekte Spektralschätzer erhält man durch Gewichtungen der ACF. Dazu werden so genannte *Lagfenster* (κ_τ) mit $\kappa_{-\tau} = \kappa_\tau$ verwendet. Die Lagfenster werden so gewählt, dass die Kovarianzen für große Lags kleine Gewichte erhalten oder sogar mit Null gewichtet werden. Dass der Zugang über die Gewichtung der ACF stimmig ist, zeigt folgende Überlegung. Die Schätzung der Spektraldichte mit dem Periodogramm scheitert grob gesprochen daran, dass beim Periodogramm zu viele – nämlich alle – Autokorrelationen berücksichtigt werden. Für große Lags sind diese aber sehr schlechte Schätzungen, da ja die Anzahl der einbezogenen Summanden mit wachsendem Lag immer geringer wird.

Die Motivation für die Betrachtung dieser Schätzer röhrt von dem Bedürfnis her, den Berechnungsaufwand für die Bestimmung des Periodogramms zu reduzieren. Dazu benötigt man ja die ACF bis zum Lag $N - 1$. Mit den heutigen Computern ist die Bedeutung der indirekten Spektralschätzer in den Hintergrund gerückt.

Die Schätzung ist dann

$$\hat{f}(\lambda) = \sum_{\tau=-(N-1)}^{N-1} \kappa_\tau c_\tau \cos(2\pi\lambda\tau) = \int_{-0.5}^{0.5} W(\omega) I(\lambda - \omega) d\omega . \quad (6.35)$$

$W(\omega)$ heißt das zu dem Lagfenster κ_τ gehörige *stetige Periodogrammfenster*. Diese Beziehung zeigt, dass die gewichteten Summen der direkten Spektralschätzer bei den indirekten Spektralschätzern durch geeignete Integrale ersetzt werden.

Beispiel 6.33 (Indirektes Bartlett-Priestley-Fenster)

Das Ziel, ein unter geeigneten Gesichtspunkten optimales kontinuierliches Periodogrammfenster zu finden, führt auf das indirekte *Bartlett-Priestley-Fenster*. Das Lagfenster ist

$$k_\tau = \begin{cases} 1 & \text{für } \tau = 0 \\ \frac{3T^2}{(\pi\tau)^2} \left(\frac{\sin(\pi\tau/T)}{\pi\tau/T} - \cos(\pi\tau/T) \right) & \text{für } |\tau| \leq T, \tau \neq 0 \\ 0 & \text{für } |\tau| > T. \end{cases}$$

Das zugehörige Periodogrammfenster lautet:

$$W(\omega) = \begin{cases} \frac{3}{2} T(1 - (2T\omega)^2) & \text{für } |\omega| < \frac{1}{2T} \\ 0 & \text{für } |\omega| \geq \frac{1}{2T}. \end{cases}$$

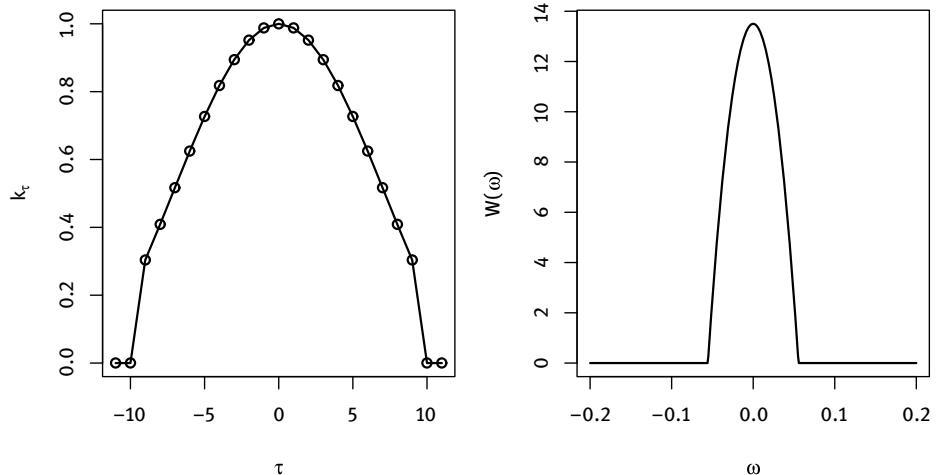


Abb. 6.22: Bartlett-Priestley-Lag- und Periodogrammfenster ($T = 9$)

Hieran kann man gut sehen, dass eine größere Anzahl von einbezogenen Lags beim Lagfenster mit einem schmäleren Bereich von Frequenzen korrespondiert, über den mit dem zugehörigen stetigen Periodogrammfenster gemittelt wird.

Das kontinuierliche Periodogrammfenster kann diskretisiert werden, um einen direkten Spektralschätzer zu erhalten. Die Approximation ergibt die in Beispiel 6.29 angegebene Folge von Gewichten.

Parametrische Spektralschätzer sind neben den nichtparametrischen direkten und indirekten Spektralschätzern sind von Bedeutung. Sie gehen von der Möglichkeit aus, einen stationären Prozess (Y_t), dessen Spektraldichte

$$0 < c_1 \leq f_Y(\lambda) \leq c_2 < \infty$$

erfüllt, als ARMA[p, q]-Prozess darzustellen. Dabei ist zwar theoretisch $p, q = \infty$ zugelassen, für die Spektralschätzung wird aber einfach ein ARMA[p, q]-Prozess angepasst und das zugehörige theoretische Spektrum unter Verwendung der geschätzten Parameter berechnet. Meist begnügt man sich mit AR-Modellen, da diese leicht zu schätzen sind. Bei autoregressiven Spektralschätzern stehen übrigens Konfidenzbänder für die gesamte Spektraldichte zur Verfügung, siehe Newton & Pagano (1984).

Die Wahl der Ordnung entspricht dabei der Problematik der Bandbreitenwahl bei den nichtparametrischen Spektralschätzern. Hier ist es keine gute Idee, die Ordnung des anzupassenden Modells mit einem der Informationskriterien zu bestimmen. In der Regel wird dann nämlich eine für die Spektralschätzung zu geringe Ordnung bestimmt; einzelne spektrale Peaks werden somit nicht sichtbar. Dies ist am Beispiel der Reihe WHORMON verdeutlicht. Dort wird über das AIC-Kriterium ein AR[2]-Prozess ausgewählt. Aber erst ein AR-Modell mit relativ hoher Ordnung erlaubt, die beiden spektralen Peaks zu trennen, siehe die Abbildung 6.23.

Werden zur Schätzung die Yule-Walker-Schätzer verwendet, so ist es vorteilhaft, die Zeitreihe einer Taper-Modifikation zu unterwerfen.

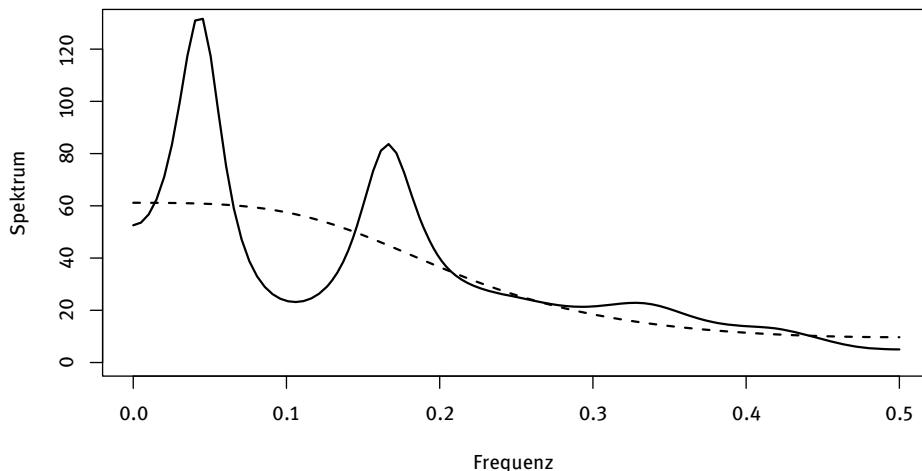


Abb. 6.23: Autoregressive Spektralschätzungen für WHORMON, AR[2]-Modell (---) und AR[10]-Modell (—)

Die *dynamische Spektralschätzung* verfolgt eine andere Idee. Für die Betrachtung einer Zeitreihe im Frequenzbereich wird sie als stationär vorausgesetzt. Das macht Sinn, da die Sinus- und Kosinuswellen, als deren Überlagerung die Reihe dabei betrachtet wird, ja periodisch sind und sich somit selbst im Laufe der Zeit nicht ändern. Nun sind aber viele der empirisch vorkommenden Zeitreihen nicht stationär. Dann ist es bisweilen von größerem Interesse, die zeitliche Entwicklung der Reihe zu analysieren als sie mittels einer geeigneten Transformation in eine stationäre zu verwandeln um sie einer Analyse im Frequenzbereich zugänglich zu machen. Sofern die Instationarität sich nicht abrupt zeigt sondern langsam in der Zeit vollzieht, kann eine Spektralschätzung in der Weise erfolgen, dass kürzere Segmente der Reihe noch als (in etwa) stationär angesehen werden und für diese jeweils eine Schätzung des Spektrums bestimmt wird. Man erhält damit eine dynamische Spektralschätzung.

Beispiel 6.34 (IBM-Aktienkursreihe)

Die Reihe der börsentäglichen Aktienkurse der IBM-Aktie ab dem 17. Mai 1961 erscheint selbst instationär; daher wird zu den ersten Differenzen der logarithmierten Reihe $\Delta \ln(y_t)$ übergegangen, siehe Abbildung 6.24. Diese Zeitreihe wird mit IBM bezeichnet.

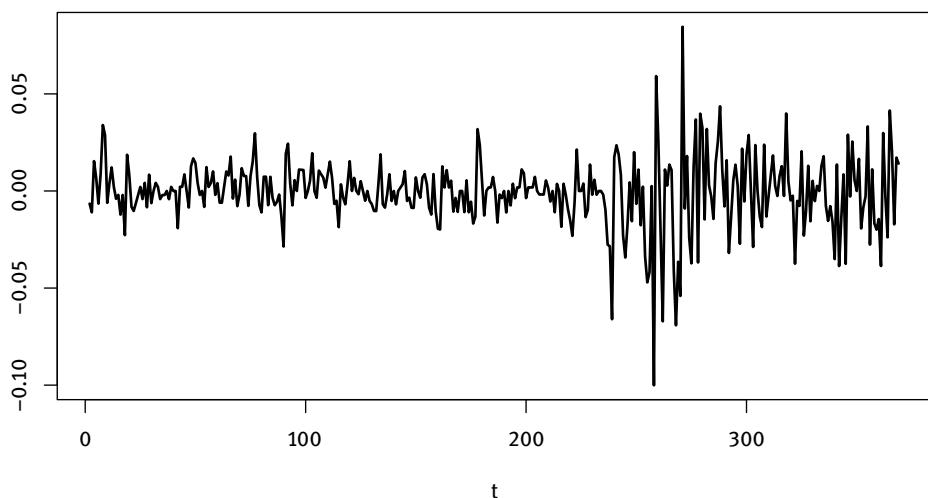


Abb. 6.24: IBM-Aktie (Differenzen)

Der zweite Teil der Reihe der Differenzen weist eine erheblich höhere Variabilität auf als der erste. Eine dynamische Spektralschätzung quantifiziert dieses Bild. Während der ersten Hälfte der Zeitpunkte ist die Spektralschätzung niedrig und vergleichsweise flach. Ab dem dritten Drittel ändert sich die Dynamik drastisch. Man erkennt die starke kurzfristige Bewegung im Hochfrequenzbereich. Zudem gibt es einen Peak der Spektralschätzung bei $0.13 \leq \lambda \leq 0.16$ und einen bei $0.31 \leq \lambda \leq 0.32$.

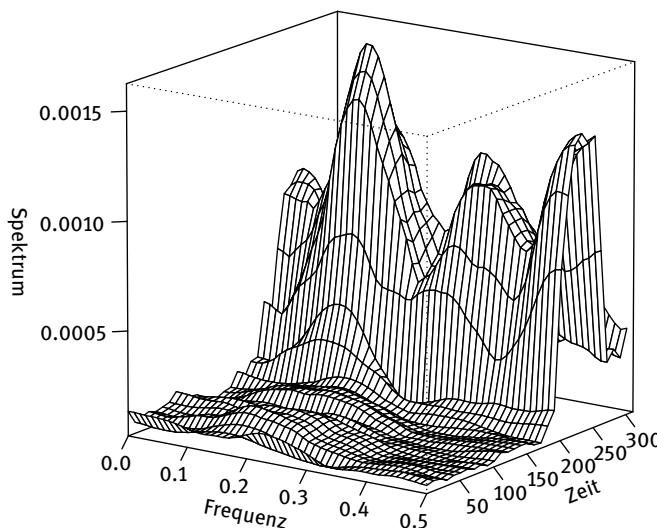


Abb. 6.25: Dynamische Spektralschätzung für die ersten Differenzen der Reihe IBM

R-Code:

```

1 data(IBM)
2 y <- diff(log(IBM))
3 plot(y)
4 out <- dynspecest(y,60,50,0.2,theta=0,phi=15,d=1,Plot=TRUE)
5 persp(out$f,out$t,out$spec,ticktype="detailed",xlab="Frequenz",
6       ylab="Zeit",zlab="Spektrum",theta=35,d=35)

```

Die Funktion `dynspecest` aus dem Paket `tsapp` führt eine dynamische Spektralschätzung durch und stellt das Ergebnis dreidimensional dar. Da ohne Kenntnis der Gestalt des dynamischen Spektrums die Einstellung der Betrachtungskoordinaten nicht möglich ist, erfolgt im Anschluss die Modifikation der Grafik unter Verwendung der ermittelten Werte für mit der Funktion `persp`.

6.6 R-Funktionen

In R werden Periodogramme und Spektralschätzungen mit der Funktion `spectrum` ermittelt. Wird die Funktion auf eine Zeitreihe `y` angewendet, so wird das Spektrum für einen Bereich angegeben, der mit dem Zeitreihenattribut `frequency` zusammenhängt. Die Spektraldichte wird zu einer Dichte über dem Intervall `(-frequency(y)/2, +frequency(y)/2)`. Um die übliche und der Interpretation zugänglichere Skalierung auf das Intervall `(-0.5, 0.5]` zur Verfügung zu haben, braucht man nur die `frequency`-Information der Zeitreihe zu löschen.

Per Voreinstellung werden die Reihen Taper-modifiziert und mit Nullen ergänzt, so dass über eine Fouriertransformation eine sehr schnelle Berechnung der Periodogrammordinaten möglich ist. Dies kann über die Setzung der optionalen Argumente ausgeschaltet werden. Dann werden sie an den Fourierfrequenzen bestimmt. Über das optionale Argument `method = "ar"` kann auch eine autoregressive Spektralschätzung durchgeführt werden.

Die Funktionen `periodogram` und `specest` in `tsapp` erlauben, selbst die Abszissenwerte festzulegen und auch mit unterschiedlichen Spektralfenstern zu operieren. Die zweitgenannte berechnet auch Konfidenzgrenzen für die Spektraldichte. Mit `spec.plot` wird das Ergebnis grafisch dargestellt. Die Funktion `wntest` aus `tsapp` dient zur grafischen Durchführung des Tests auf White Noise. Die Anzahl der ggf. verwendeten Regressoren ist als Parameter anzugeben.

Die komplexe De- und Remodulation lässt sich mit der Funktion `bandfilt` aus dem Paket `tsapp` durchführen. Der Baxter-King-Filter ist in der Funktion `bkfilter` aus dem Paket `mFilter` implementiert. Die Komponentenzerlegung mit der Funktion `ARIMADec` aus dem Paket `tsdecomp` erfolgt mittels einer parametrischen Spektralschätzung. Dabei werden ARIMA-Modelle für jede Komponente (Trend, Saison und Rest) bestimmt. `ArDec` geht von der autoregressiven Spektralschätzung aus.

Um theoretische Spektren von ARMA-Modellen mit empirischen Spektralschätzungen zu vergleichen, kann die Funktion `armathspec` aus `tsapp` verwendet werden.

7 Prozesse mit langem Gedächtnis

7.1 Einführung der Prozesse

Im Abschnitt ‚Schätzen der Kennfunktionen‘, siehe Seite 21, wurde darauf hingewiesen, dass für Aussagen über das Schätzen die absolute Summierbarkeit der Autokovarianzfunktion (γ_τ) hinreichend ist. Die Eigenschaft $\sum |\gamma_\tau| < \infty$ stellt sicher, dass die Abhängigkeiten der Zeitreihenvariablen mit wachsendem Lag rasch gegen null gehen (exponentiell). Man spricht dann auch von einem *kurzen Gedächtnis* des zugehörigen Prozesses. Ein kurzes Gedächtnis geht damit einher, dass die Spektraldichte bei null beschränkt ist, die langwelligen Bewegungen in der Reihe also nicht zu stark sind. Alle stationären ARMA-Prozesse sind von dieser Sorte. In Anwendungen kommen aber auch Zeitreihen vor, die diesem Bild nicht entsprechen. Granger & Joyeux (1980) sowie Granger (1980) bemerken, dass bisweilen der langwellige Anteil von Zeitreihen durch das übliche Differenzenbilden zu stark ausgeblendet wird. Dann scheint das Log-Spektrum der Originalreihe bei null unbeschränkt zu sein, während das Spektrum der ein- bzw. zweimal differenzierten Reihe bei null praktisch verschwindet. Für stationäre und invertible ARMA-Prozesse verschwindet die Spektraldichte aber bei null nicht.

Um dies modellmäßig zu erfassen, sei (X_t) ein Prozess, so daß $Y_t = (1 - B)X_t$ ein bei null positives und beschränktes Spektrum $f_Y(\lambda)$ besitzt. Da der Differenzenfilter die Gainfunktion $G(\lambda) = 2 \sin(\pi\lambda)$ hat, siehe Beispiel 6.22, könnte man nach Satz 6.20 für ein Spektrum des Ausgangsprozesses $f_X(\lambda) = (2 \sin(\pi\lambda))^{-2} f_Y(\lambda)$ ansetzen. Hier führt die Differenzenbildung zur Multiplikation mit $G(\lambda)^2$, also zu $f_Y(\lambda)$. Ein Modell, bei dem das einfache Differenzenbilden ‚zu viel‘ ist, erhält man daher, indem $f_X(\lambda) = (2 \sin(\pi\lambda))^{-2d} f_Y(\lambda)$ mit $0 < d < 1$ angesetzt wird.

Prozesse mit einem solchen Spektrum erfüllen die Voraussetzungen für die erwähnte Beobachtung, dass die zugehörigen Zeitreihen eine Differenzenbildung nahelegen, dies aber ‚zu viel des Guten‘ ist: Da $f_Y(\lambda)$ bei null positiv und beschränkt ist, gilt $f_X(\lambda) \rightarrow \infty$ für $\lambda \rightarrow 0$. Andererseits führt das einfache Differenzenbilden zur Elimination der spektralen Masse bei null:

$$(2 \sin(\pi\lambda))^2 \cdot (2 \sin(\pi\lambda))^{-2d} f_Y(\lambda) = (2 \sin(\pi\lambda))^{2(1-d)} f_Y(\lambda) \rightarrow 0 \quad \text{für } \lambda \rightarrow 0.$$

Prozesse mit einem solchen Spektrum sind stationär, falls $d < 0.5$. Sie haben Autokovarianzfunktionen, die nicht mehr absolut summierbar sind, $\sum |\gamma_\tau| = \infty$. Daher spricht man von Prozessen mit *langem Gedächtnis*.

Beispiel 7.1 (Aggregation von AR[1]-Prozessen)

Granger (1980) hat darauf hingewiesen, dass die Aggregation von AR[1]-Prozessen zu Modellen mit langem Gedächtnis führen kann. Möge sich etwa die Entwicklung des Konsums von Familien einer Population jeweils mittels eines AR[1]-Prozesses

erfassen lassen, $Y_{jt} = \alpha_j Y_{jt-1} + \varepsilon_{jt}$, $j = 1, \dots, J$. Dabei seien die White-Noise-Prozesse (ε_{jt}) unabhängig, bei null zentriert und mögen die Varianzen $(1 - \alpha_j^2)\sigma_j^2$ haben.

Für den Gesamtkonsum $Z_t = \sum_{j=1}^J Y_{jt}$ gilt dann:

$$\text{Var}(Z_t) = \sum_{j=1}^J \sigma_j^2 \quad \text{und} \quad \gamma_Z(\tau) = \alpha_j^\tau \sigma_j^2 \quad \tau > 0.$$

Um bei großem J die Varianz von Z_t in Grenzen zu halten, wird $\sigma_j^2 = c/J$ unterstellt. Weiter wird unterstellt, dass sich die Koeffizienten gleichmäßig über das Intervall $(0;1)$ verteilen. Dann sind die Kovarianzen approximativ

$$\gamma_Z(\tau) = \sum_{j=1}^J \left(\frac{j}{J}\right)^\tau \approx \frac{c}{\tau + 1}.$$

Wegen $\sum \tau^{-1} = \infty$ ist (Z_t) ein Prozess mit langem Gedächtnis.

Die Verbindung von langem Gedächtnis und ARMA-Modellen geschieht nun in der folgenden Definition. Sie wurde von Granger & Joyeux (1980) und Hosking (1981) propagiert.

Definition 7.2 (Fraktionell integrierter ARMA-Prozess)

(Y_t) heißt ein fraktionell integrierter ARMA-Prozess, kurz ARFIMA[p, d, q]-Prozess, falls $(l - B)^d Y_t$ ein ARMA[p, q]-Prozess ist. Dabei ist für nicht ganzzahliges $d > 0$ die Differenz über die Binomialentwicklung definiert durch:

$$(l - B)^d = \sum_{u=0}^{\infty} \pi_u B^u, \quad (7.1)$$

mit $\pi_0 = 1$ und

$$\pi_u = \frac{\Gamma(u - d)}{\Gamma(u + 1)\Gamma(-d)} = \prod_{k=1}^d \frac{k - u - d}{k} \quad u = 1, 2, \dots$$

wobei $\Gamma(u)$ die Gamma-Funktion ist: $\Gamma(u) = \int_0^\infty s^{u-1} e^{-s} ds$ für $u > 0$, $\lim_{u \rightarrow 0+} \Gamma(u) = \infty$ und der Fortsetzung für negative u unter Ausnutzung der Eigenschaft $u\Gamma(u) = \Gamma(u + 1)$.

Ein ARFIMA[$0, d, 0$]-Prozess mit $-0.5 < d < 0.5$ heißt auch fraktionell differenzierter White-Noise-Prozess oder einfach FDWN-Prozess.

Satz 7.3 (Eigenschaften eines FDWN-Prozesses)

Sei (Y_t) ein FDWN-Prozess mit dem Parameter d . Dann ist (Y_t) für $-1/2 < d < 1/2$ stationär und hat die AR[∞]-Darstellung

$$\sum_{u=0}^{\infty} \frac{\Gamma(u - d)}{\Gamma(u + 1)\Gamma(-d)} Y_{t-u} = \varepsilon_t,$$

sowie die $MA[\infty]$ -Darstellung

$$Y_t = \sum_{u=0}^{\infty} \frac{\Gamma(u+d)}{\Gamma(u+1)\Gamma(d)} \varepsilon_{t-u}. \quad (7.2)$$

Die Autokovarianzfunktion und die Autokorrelationsfunktion sind gegeben durch:

$$\gamma_\tau = \frac{(-1)^\tau \Gamma(1-2d)}{\Gamma(\tau+1-d)\Gamma(-\tau+1-d)} \quad \tau = 0, 1, 2, 3, \dots \quad (7.3a)$$

$$\rho_\tau = \frac{\Gamma(\tau+d)\Gamma(1-d)}{\Gamma(\tau+1-d)\Gamma(d)} \quad \tau = 0, 1, 2, 3, \dots \quad (7.3b)$$

Dass ein FDWN-Prozess (Y_t) im Fall $d > 0$ ein Prozess mit langem Gedächtnis ist, ersieht man wegen der Divergenz der harmonischen Reihe $\sum n^\alpha$ für $\alpha \geq -1$ aus der folgenden Relation, die sich mit der Stirlingschen Approximation für große τ ergibt:

$$\rho_\tau \approx \frac{\Gamma(1-d)}{\Gamma(d)} \tau^{2d-1}.$$

Auch ARFIMA[p, d, q]-Prozesse $\alpha(B)(1-B)^d Y_t = \beta(B)\varepsilon_t$ sind für $-0.5 < d < 0.5$ stationär, sofern die Wurzeln der charakteristischen Gleichung des AR-Polynoms außerhalb des Einheitskreises liegen. Weiter ist ein solcher Prozess invertibel, wenn zudem die Wurzeln des charakteristischen Polynoms von $\beta(B)$ außerhalb des Einheitskreises liegen. Daher braucht die Situation $d \geq 0.5$ nicht gesondert betrachtet zu werden; einfaches Differenzenbilden führt auf sie. Dabei ist $d < 0$ für die Anwendung uninteressant. Man spricht auch von *antipersistenten* Prozessen. Deren Realisationen tendieren dazu, häufiger als ein White-Noise-Prozess von einer Seite das Niveaus zur anderen zu springen.

Die Eigenschaften der FDWN-Prozesse sind vor allem durch die Korrelationsstruktur bei sehr großen Lags gekennzeichnet. ARMA-Prozesse dagegen weisen für solche großen Lags keine nennenswerten Korrelationen mehr auf. Daher ist es nicht verwunderlich, dass die ARFIMA[p, d, q]-Prozesse sich bzgl. der langfristigen Korrelation ganz ähnlich verhalten wie die FDWN-Prozesse.

Beispiel 7.4 (Hurst-Effekt)

Der Hydrologe Hurst (1951) stellte die Charakteristik der langen Abhängigkeit fest, als er untersuchte, wie der Nil am besten reguliert werden könnte. Er betrachtete dazu ein Reservoir und den kumulierten Zufluss zwischen den Zeitpunkten t und $t+k$. Ist X_s der Zufluss zum Zeitpunkt s , so ist die ideale Kapazität des Reservoirs mit $\bar{X}_{t,k} = \sum_{i=1}^k X_i/k$ gegeben durch

$$R(t, k) = \max_{0 \leq i \leq k} \left(\sum_{j=0}^i X_{t+j} - i \cdot \bar{X}_{t,k} \right) - \min_{0 \leq i \leq k} \left(\sum_{j=0}^i X_{t+j} - i \cdot \bar{X}_{t,k} \right) \quad (7.4)$$

$R(t, k)$ wird als *adjustierte Spannweite (Range)* bezeichnet. Sie wird standardisiert mittels

$$S(t, k) = \sqrt{\frac{1}{k} \sum_{i=t+1}^{t+k} (X_i - \bar{X}_{t,k})^2}. \quad (7.5)$$

Die R/S -Statistik, die *reskalierte adjustierte Spannweite*, ist dann

$$R/S = \frac{R(t, k)}{S(t, k)}. \quad (7.6)$$

Für diese Statistik fand Hurst die Gesetzmäßigkeit

$$\mathbb{E}(R/S) \rightarrow c \cdot k^H \quad \text{für } k \rightarrow \infty. \quad (7.7)$$

Stellt man also die logarithmierten Werte von R/S in Abhängigkeit von verschiedenen Werten von $\ln(k)$ dar, so streuen die Punkte für große k um eine Gerade mit einer Steigung H . Hurst fand auf diese Weise bei etlichen Reihen einen Wert $H > 1/2$. Dieser empirisch ermittelte *Hurst-Effekt* ist nicht vereinbar mit dem Verhalten von Prozessen, die ein kurzes Gedächtnis haben. Denn für solche Prozesse ist die Variabilität der Mittelwerte $\bar{X}_{t,k}$ von der Ordnung \sqrt{k} . Dem entspricht eine Steigung von $H = 1/2$. Mandelbrot zeigte, dass der Effekt mittels FDWN-Prozessen modelliert werden kann, siehe Mandelbrot (1977) und die dort angegebenen Quellen.

Der im letzten Beispiel eingeführte *Hurst-Koeffizient* H steht bei ARFIMA[p, d, q]-Prozessen in direktem Zusammenhang zum Parameter d :

$$d = H - 0.5. \quad (7.8)$$

Die Betrachtung der empirischen Autokorrelationen mit den individuellen Standardgrenzen ist bei Prozessen mit langem Gedächtnis nicht sinnvoll. Einzeln können die Autokorrelationen innerhalb der Grenzen bleiben, aber in der Gesamtheit dennoch einen Eindruck einer Erhöhung vermitteln. Sie sind in ihrer Gesamtheit zu beurteilen, da sich das lange Gedächtnis auf die Struktur der ACF auswirkt.

Beispiel 7.5 (Wasserstände des Nil)

Die jährlichen minimalen Wasserstände des Nil liegen als sehr lange Zeitreihe vor. Im R-Paket `longmemo` gibt es die Zeitreihe `NIL` (als `NileMin`) mit den jährlichen minimalen Wasserständen von 622 uZ bis 1284 uZ. Das lange Gedächtnis dieser Reihe lässt sich schon aus dem langsamen Abklingen der ACF erkennen. Dieses Bild steht wie gesagt im Einklang mit dem Hurst-Effekt.

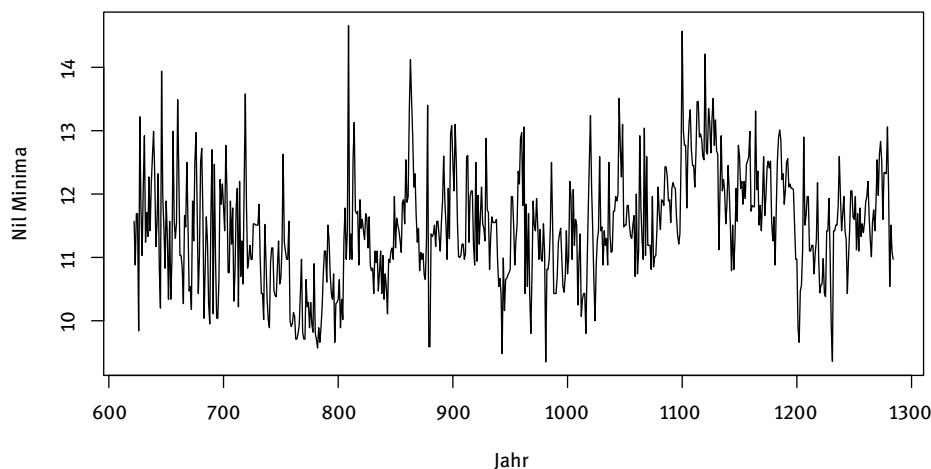


Abb. 7.1: Jährliche minimale Wasserstände des Nil

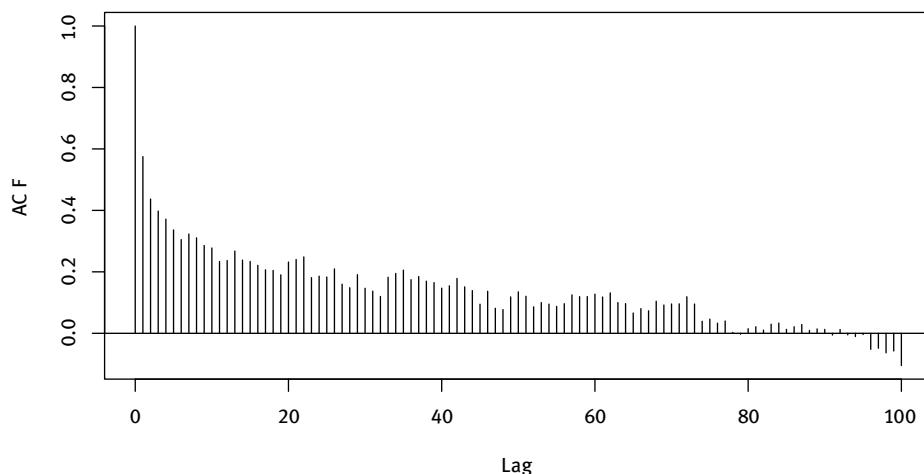


Abb. 7.2: ACF der Zeitreihe NIL

7.2 Bestimmung des fraktionellen Exponenten

Wie das Beispiel zum Hurst-Effekt zeigt, schlug man sich mit dem Phänomen ‚langes Gedächtnis‘ herum, lange bevor ein adäquater theoretischer Rahmen existierte. Dementsprechend wurden verschiedene heuristische Methoden entwickelt, die heute eher als rein explorative Werkzeuge verwendet werden. Auch bei der Schätzung standen am Anfang heuristische Methoden.

Die R/S-Analyse gehört zu den explorativen Methoden. Sie setzt den im Beispiel 7.4 skizzierten Ansatz um und schätzt den Hurst-Parameter mittels gewöhnlicher linearer Regression. Dieser Ansatz bietet einige Wahlmöglichkeiten. Da das Resultat nur für große k gilt, macht es Sinn, die Regression auch nur auf relativ große k zu stützen. Zu jedem k gibt es weiter $N - k$ Werte der R/S-Statistik, entsprechend der Wahl von $t = 1, 2, \dots, N - k$. Für nahe beieinander liegende Zeitpunkte t und t' werden sich die Werte der Statistik nur wenig unterscheiden. Daher lässt es sich rechtfertigen, nur überlappungsfreie Zeitreihensegmente zu betrachten.

Beispiel 7.6 (Nil – Fortsetzung von Seite 176)

Für die Reihe NIL werden für $20 \leq k \leq 200$ aus jeweils allen nicht-überlappenden Segmenten die Werte der R/S-Statistik berechnet und in doppeltlogarithmischer Achseninteilung dargestellt. Die Regression ergibt die Steigung $H = 0.876$. Die Regressionsgerade ist ebenfalls eingezeichnet.

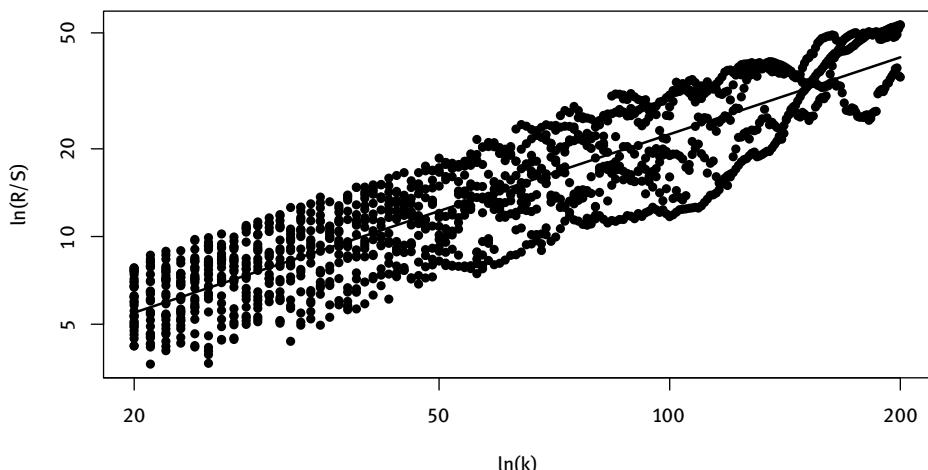


Abb. 7.3: Reskalierte adjustierte Spannweite für Reihe NIL

R-Code:

```

1 library(longmemo)
2 data(NileMin)
3 k <- c(20:200)
4 RNil <- RS(NileMin,k)
5 plot(RNil[,1],RNil[,3], type = "p", log ="xy",pch=16,xlab="ln(k)",
6      ylab="ln(R/S)")
7 lk <- log(RNil[,1])
8 lR <-log(RNil[,3])

```

```

9  coef <- coefficients(lm(lR~lk))
10 print(coef)
11 a <- coef[1]
12 b <- coef[2]
13 curve(exp(a)*x^b, add = T, lwd=2)

```

Die Nil-Daten werden bereitgestellt. In dem Paket `tsapp` befindet sich eine Implementierung der *R/S*-Statistik. Diese wird auf die Zeitreihe für verschiedene k angewendet. Die grafische Darstellung nutzt die Option für die logarithmische Skalierung beider Achsen. Allerdings ist dann die Regressionsgerade in die ursprüngliche Skala zu transformieren.

Ein erster Ansatz, den fraktionellen Exponenten im Rahmen von fraktionell integrierten Prozessen zu schätzen, wurde von Geweke & Porter-Hudak (1983) vorgestellt. Er basiert auf dem Frequenzbereich. Sei (Y_t) ein fraktionell integrierter Prozess, $(1 - B)^d Y_t = Z_t$, wobei (Z_t) ein stationärer Prozess mit beschränkter Spektraldichte sei, die zudem eine positive untere Schranke besitze. Für d wird $-0.5 < d < 0.5$ unterstellt. Dann ist (Y_t) stationär und hat das Spektrum

$$f_Y(\lambda) = (2 \cdot \sin(\pi\lambda))^{-2d} \cdot f_Z(\lambda).$$

Das aus der (nicht-zentrierten) Reihe y_1, \dots, y_N berechnete Periodogramm erfüllt dann die Identität

$$\ln I_Y(\lambda) = \ln(f_Z(0)) - d \cdot \ln(4 \cdot \sin^2(\pi\lambda)) + \ln\left(\frac{I_Y(\lambda)}{f_Y(\lambda)}\right) + \ln\left(\frac{f_Y(\lambda)}{f_Y(0)}\right) \quad (7.9)$$

Der letzte Term kann im Vergleich zu den anderen vernachlässigt werden, wenn nur kleine λ betrachtet werden. Die Standard-Theorie legt dann nahe, dass an Fourierfrequenzen $\lambda_j = 2\pi j/N$ die normierten Periodogrammordinaten $I(\lambda_j)/f_Y(\lambda_j)$ approximativ identisch verteilt und unabhängig sind. Somit kann (7.9) als Regressionsgleichung aufgefasst und d nach der Methode der kleinsten Quadrate bestimmt werden.

Es hat sich aber herausgestellt, dass die Standardvoraussetzungen nicht erfüllt sind. Für sehr kleine λ_j sind die Periodogrammordinaten asymptotisch weder unabhängig noch identisch verteilt, siehe Robinson (1995). Allerdings hat Robinson in dieser Arbeit auch gezeigt, dass der Ansatz unter gewissen Bedingungen zu konsistenten, jedoch nicht effizienten Schätzern führt. Dabei sind auf der einen Seite die kleinsten Frequenzen auszublenden; auf der anderen muss sich der Bereich der in die Regression eingehenden Frequenzen immer stärker bei null zusammenziehen.

Satz 7.7 (Asymptotik des Regressionsschätzers für d)

Sei (Y_t) ein stochastischer Prozess, so dass für ein $-0.5 < d < 0.5$ $(1 - B)^d Y_t = Z_t$ ein stationärer Gaußscher Prozess ist mit der Darstellung $Z_t = \sum_{u=-\infty}^{\infty} c_u \varepsilon_{t-u}$.

Seien weiter $\lambda_j = 2\pi j/N$ die Fourierfrequenzen, $J \geq 1$ eine feste Zahl und seien $h = h_N$ eine (vom Anwender gewählte) Trimm-Zahl und $m = m_N$ eine (ebenfalls vom Anwender gewählte) Bandbreite, h und m mögen mit N größer werden, aber

geeignet langsamer. Sei

$$W_k = \ln \left(\sum_{u=1}^J I_Z(\lambda_{k+j-J}) \right), \quad k = h+J, h+2J, \dots, m.$$

Dann ist der Kleinsten-Quadrat-Schätzer \hat{d} der Regressionsbeziehung

$$W_k = c - d \cdot \ln(4 \cdot \sin^2(\pi \lambda_k)) + U_k, \quad k = h+J, h+2J, \dots, m$$

asymptotisch normalverteilt:

$$\frac{\hat{d} - d}{\hat{\sigma}_{\hat{d}}} \sim \mathcal{N}(0, 1).$$

Dabei ist

$$\begin{aligned} \hat{\sigma}_{\hat{d}}^2 = & \left\{ \frac{j}{m-h} \sum_k \hat{U}_k^2 \right\} \Bigg/ \left\{ \sum_k \left[\ln \left(4 \cdot \sin^2 \left(\frac{\pi k}{N} \right) \right) \right]^2 \right. \\ & \left. - \frac{j}{m-h} \left[\sum_k \ln \left(4 \cdot \sin^2 \left(\frac{\pi k}{N} \right) \right) \right]^2 \right\}. \end{aligned}$$

Hier gehen die Summen über $k = h+J, h+2J, \dots, m$.

Die Bedingungen an h und m sind etwa für $h = m^{3/4} \ln(m)$ und $m = N^{1/2}$ erfüllt. Bei wachsendem J kann man mit einem Effizienzgewinn rechnen. Samarov, Smith & Taqqu (1992) zeigen, dass der KQ-Schätzer für $h_n/\ln(N) \rightarrow \infty$ asymptotisch unverfälscht ist.

Die angegebenen Bedingungen gelten asymptotisch. Für die Anwendung bei endlichen Reihen stellt sich die Frage nach der Wahl der Unter- und der Obergrenze h und m der in die Regression einzubeziehenden Frequenzen. Geweke & Porter-Hudak empfehlen aufgrund von Erfahrungen aus Simulationen $m = \sqrt{N}$. Die Wahl von h ist in gewisser Hinsicht noch kritischer. Denn es sind ja gerade die Frequenzen nahe null, die für das long-memory-Verhalten entscheidend sind. Da man weiß, dass die Periodogrammordinaten erratisch sehr groß sein können, empfiehlt es sich, anstelle der Kleinsten-Quadrat-Regression eine robuste Regression durchzuführen. Eine solche ist etwa eine M-Regression. Hier wird der quadratische Abstand zwischen Modell und Daten durch eine andere Abstandsfunktion ersetzt, die für große Abstände weniger schnell wächst oder sogar ab einem Punkt konstant bleibt. An Stelle der Abstandsfunktion wird i. d. R. ihre Ableitung, die sogenannte Psi-Funktion, betrachtet und das resultierende Gleichungssystem numerisch gelöst. Auf das Verfahren der M-Schätzung wird im Abschnitt 12.3 näher eingegangen.

Beispiel 7.8 (Nil – Fortsetzung von Seite 178)

Die Reihe NIL diente von Anfang an als eines der zentralen Studienobjekte bei der Entwicklung der long-memory-Modelle.

Die Abbildung 7.4 zeigt das erratische Verhalten der Periodogrammordinaten. Für $m = 26 \approx \sqrt{N}$ und $J = 1$ wurde für verschiedene h der Parameter d mittels

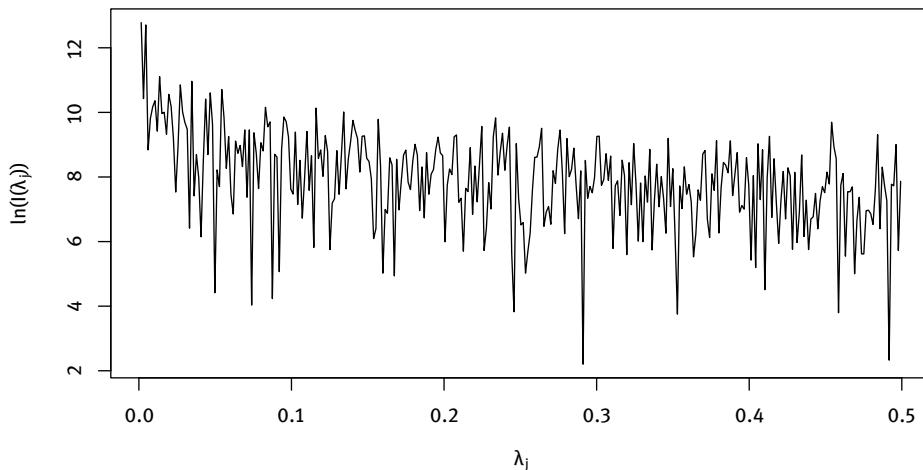


Abb. 7.4: Logarithmiertes Periodogramm der Zeitreihe NIL an den Fourierfrequenzen

Kleinstes-Quadrat- und M-Regression geschätzt. Bei letzterer wurde die Biweight-Psi-Funktion verwendet. Man sieht, dass die M-Schätzung durch die Wahl von h wesentlich weniger stark beeinflusst wird.

h	KQ-Reg.	M-Reg.
2	0.504	0.464
3	0.433	0.454
4	0.516	0.446
5	0.316	0.434

Zur Erinnerung sei auf das Ergebnis aus Beispiel 7.6 hingewiesen. Dort ergab sich der Hurst-Koeffizient $H = 0.876$. Wegen $d = H - 0.5$ erhält man hier Schätzungen, die höher sind.

R-Code:

```

1 out <- periodogram(NileMin, 663/2)
2 x <- log(4*sin(pi*out[,1])^2)
3 y <- log(out[,2])
4 h <- 4
5 out1 <- lm(y[h:26]~x[h:26])
6 library(MASS)
7 out2 <- rlm(y[h:26],x[h:26],method = "M",psi = psi.bisquare)

```

Um das Periodogramm an den Fourierfrequenzen zu erhalten, wird die Funktion `periodogram` aus dem Paket `tsapp` verwendet. Dort ist die erste Frequenz null,

so dass sie aus dem Ergebnis eliminiert werden muss. Weiterhin werden bei dieser Funktion die Frequenzen über das zweite Argument nf gemäß $j/(2*nf)$ bestimmt. Um die Fourierfrequenzen zu erhalten, ist die hier vorgenommene Angabe nötig. Für die robuste Regression wird die Funktion `r1m` aus dem Paket MASS eingesetzt.

Bei Unterstellung eines Normalprozesses kann für einen ARFIMA[p, d, q]-Prozess die Likelihoodfunktion aufgeschrieben werden. Die Bestimmung der ML-Schätzer kann dann im Prinzip über eine Maximierung der Likelihoodfunktion erfolgen. Deren Effizienz und asymptotische Normalität wurden von Fox & Taqqu (1986) und Dahlhaus (1989) nachgewiesen.

Satz 7.9 (Asymptotik der ML-Schätzer bei ARFIMA-Prozessen)

Sei (Y_t) ein stationärer Gaußscher ARFIMA[p, d, q]-Prozess mit $-0.5 < d < 0.5$, $\alpha(B)(1 - B)^d Y_t = \beta(B)\varepsilon_t$. Sei $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q, d, \sigma_\varepsilon^2)'$ und $\boldsymbol{\theta}_0$ der wahre Parameter des Prozesses. Sei $\hat{\boldsymbol{\theta}}_N$ der ML-Schätzer, der die exakte Loglikelihoodfunktion bei N Beobachtungen maximiert. Sind die in Dahlhaus (1989) angegebenen Regularitätsbedingungen erfüllt, so gilt:

- $\hat{\boldsymbol{\theta}}_N \rightarrow \boldsymbol{\theta}_0$ in Wahrscheinlichkeit. (Konsistenz)
- $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}_0))$ für $N \rightarrow \infty$, wobei $\boldsymbol{\Gamma}(\boldsymbol{\theta}_0) = (\Gamma_{ij}(\boldsymbol{\theta}_0))$ mit

$$\Gamma_{ij}(\boldsymbol{\theta}_0) = \int_{-0.5}^{0.5} \frac{\partial \ln(k(2\pi\lambda, \boldsymbol{\theta}))}{\partial \theta_i} \cdot \frac{\partial \ln(k(2\pi\lambda, \boldsymbol{\theta}))}{\partial \theta_j} d\lambda$$

$$\text{und } k(2\pi\lambda, \boldsymbol{\theta}) = |\sum_{u=0}^{\infty} \psi_u(\boldsymbol{\theta}) e^{i2\pi\lambda u}|^2. \quad (\text{asymptotische Normalität})$$

- $\hat{\boldsymbol{\theta}}_N$ ist ein effizienter Schätzer für $\boldsymbol{\theta}_0$. (Effizienz)

Der parametrische Ansatz verlangt, dass die Modellordnung vorab bekannt sein oder bestimmt werden muss, da die Schätzer im Fall von fehlspezifizierten Modellen inkonsistent sind. So ist das oben dargestellte Schätzverfahren zumindest für die Gewinnung von Startwerten geeignet. Die Maximierung der Likelihoodfunktion ist jedoch numerisch sehr aufwändig. Auch ist mit numerischen Schwierigkeiten zu rechnen. Insgesamt ist das kein sehr gangbarer Weg. Daher wurden verschiedene Approximationen vorgeschlagen. Eine schnelle und akkurate Approximation der Likelihoodfunktion haben Haslett & Raftery (1989) vorgeschlagen. Hier wird auch das Startwert-Problem für die Schätzung von d umgangen. Zudem erhält man bei ihrem Ansatz die asymptotischen Standardfehler.

Beispiel 7.10 (Nil – Fortsetzung von Seite 180)

Die Parameter des ARFIMA-Modells werden mit der approximativen Methode von Haslett and Raftery (1989) geschätzt. Zuerst wird eine Anzahl von autoregressiven Parametern zugelassen. Dann ist nur der erste Koeffizient bedeutsam. Dass auch diese kurzfristige Abhängigkeit nicht signifikant ist, zeigt dann die Schätzung eines ARFIMA[1, d , 0]-Modells. Es bleibt ein FGWN-Modell für die Reihe NIL.

ARFIMA[10, d , 0]-Modell

	Estimate	Std. Error	z value	Pr(> z)
d	0.309585	0.029517	10.488	< 2e-16
ar1	0.129493	0.049105	2.637	0.00836
ar2	-0.005100	0.041673	-0.122	0.90260
:				
ar10	0.035344	0.039100	0.904	0.36603

ARFIMA[1, d , 0]-Modell

	Estimate	Std. Error	z value	Pr(> z)
d	0.35530	0.02861	12.419	<2e-16
ar	0.06590	0.04875	1.352	0.176

ARFIMA[0, d , 0]-Modell

	Estimate	Std. Error	z value	Pr(> z)
d	3.933e-01	3.959e-05	9934	<2e-16

R-Code:

```
1 library(fracdiff)
2 out <- fracdiff(NileMin, nar=10)
3 summary(out)
```

Die Approximation von Haslett and Raftery ist im Paket `fracdiff` implementiert. Hier ist nur der Aufruf für das Modell mit den zehn AR-Parametern angegeben. Die anderen beiden Ergebnisse ergeben sich durch Setzung der Option `nar=1` bzw. durch deren Weglassen.

In der Ergebnistabelle des Beispiels lässt sich erkennen, dass bei der Berücksichtigung von AR- (und ebenso von MA-) Parametern die Höhe von \hat{d} tendenziell reduziert wird. Es ist offensichtlich schwierig, den auf die kurzfristigeren Abhängigkeiten zurückzuführenden Schwingungsanteile von den langfristigen zu trennen. In anderen Anwendungen kann das sogar zu antipersistenten Schätzwerten \hat{d} führen, auch wenn die Anpassung eines FDWN-Prozesses einen Wert $0 < \hat{d} < 0.5$ ergibt.

7.3 Prognosen mit ARFIMA-Modellen

ARFIMA-Modelle können in ganz derselben Weise prognostiziert werden wie die ARIMA-Modelle. Dazu wird der fraktionelle Differenzenfilter in eine Reihe entwickelt und mit dem AR-Filter multipliziert. Diese muss natürlich bei einer geeigneten Stelle abgebrochen werden. Da aber die Gewichte sehr viel langsamer gegen null gehen als bei den ARMA-Modellen wird bei den fraktionellen Exponenten die gesamte zur Verfügung stehende Vergangenheit zur Prognose genutzt.

Beispiel 7.11 (Mittlere jährliche Temperatur)

Zur Illustration wird die Zeitreihe TMPYR der mittleren jährlichen Temperatur in Zentralengland von 1659 bis 1976 genommen. Die Werte bis 1900 werden verwendet, um ein Modell anzupassen. Die restlichen 76 Werte werden prognostiziert. Die Anpassung eines ARFIMA[1, d , 1]-Prozesses ergibt (phi ist der AR-Parameter und theta der MA-Parameter):

```

      Estimate   Std. Error
phi(1)     -0.7563356  0.1424079
theta(1)    -0.6533207  0.1666733
d.f        0.2623565  0.0597327
Fitted mean 9.0550289  0.1589969
---
sigma^2 estimated as 0.366609; Log-likelihood = 122.633;
AIC = -235.265; BIC = -217.821

```

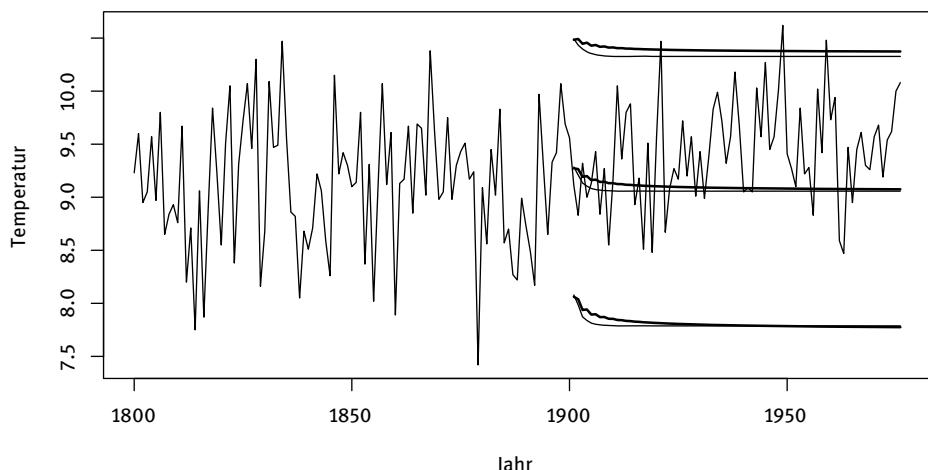


Abb. 7.5: Jährliche mittlere Temperatur in Zentralengland mit Prognosen; starke Linien: ARFIMA-Modell, feine Linien: AR[2]-Modell

Die Darstellung der ACF und PACF der Residuen für jeweils 100 Lags zeigen bei der PACF zwei irreguläre und geringfügige Verletzungen der 95 %-Grenzen. Diese scheinen aber ignorierbar. In der Literatur wird darauf hingewiesen, dass ein ARFIMA[1, d, 0]-Prozess eine fast ebenso gute Anpassung ergibt. Allerdings ist dann der AR-Koeffizient nicht mehr signifikant.

Mit diesem Modell wird die Prognose erstellt. Zum Vergleich wird ein ARMA-Modell angepasst und ebenfalls prognostiziert. Als Modell ergibt sich ein AR[2]-Prozess. Die Prognosen mit den 95 %-Prognosegrenzen sind in der Abbildung 7.5 dargestellt. Aus Darstellungsgründen wurde der erste Teil der Zeitreihe abgeschnitten. Die Prognosegrenzen des ARFIMA-Modells wurden mit einem Bootstrap bestimmt. Der Vergleich der beiden Prognosen macht deutlich, dass die ARFIMA-Prognose langsamer gegen den Mittelwert der Reihe geht als die des AR-Modells. Das spiegelt die allgemeine oben angegebene Tendenz wider.

R-Code:

```

1 library(arfima)
2 data(tmpyr)
3 tmpyrs <- window(tmpyr,start=1659,end=1900)
4 out <- arfima(tmpyrs,order=c(1,0,1))
5 res <- residuals(out)
6 r <- res[[1]]
7 acfpacf(r,100)
8 summary(out)
9 p <- predict(out,n.ahead=76)
10 out1 <- arima(tmpyrs,order=c(2,0,0))
11 p1 <- predict(out1,n.ahead=76)
12 plot(window(tmpyr,start=1800), xlab="Jahr",ylab="Temperatur",
13     type="l")
14 lines(c(1901:1976),p[[1]]$Forecast,lwd=2)
15 lines(c(1901:1976),p[[1]]$Forecast + 1.96*p[[1]]$exactSD,lwd=2)
16 lines(c(1901:1976),p[[1]]$Forecast - 1.96*p[[1]]$exactSD,lwd=2)
17 lines(c(1901:1976),p1$pred)
18 lines(c(1901:1976),p1$pred + 1.96* p1$se)
19 lines(c(1901:1976),p1$pred - 1.96* p1$se)
```

Die Zeitreihe TMPYR befindet sich im Paket `arfima`. Geschätzt wird das Modell mit der Funktion `arfima`. Bei der anzugebenden Ordnung ist bei den Differenzen nur die ganzzahlige Ordnung anzugeben. Die fraktionelle wird immer geschätzt. Das Ergebnis der Funktion enthält standardmäßig nur die Angaben für das erste lokale Maximum. Um vorab die Ordnung zu bestimmen, wurde das Modell auch mit der Funktion `fracdiff` des gleichnamigen Paketes geschätzt. Um die Größen für die Prognose zu extrahieren, ist zu beachten, dass das Ergebnis der Funktion `arfima` eine Liste von Listen ist.

7.4 R-Funktionen

Zur Analyse von Zeitreihen mit langer Abhängigkeit stehen mehrere Pakete zur Auswahl. FGN ist zur Analyse fraktionellem weißen Rauschen gedacht. Die Funktion RS zur Hurst-Statistik ist in tsapp programmiert. longmemo setzt die in der Monographie zu Long-Memory-Prozessen von Beran (1994) beschriebenen Funktionen und Methoden um. arfima ist ein recht umfassendes Paket zum Schätzen und zur Prognose von ARFIMA-Modellen. fracdiff ist das neweste Paket. Es scheint auch das am einfachsten zu bedienende, ist allerdings von den Möglichkeiten her begrenzt. Das Paket LongMemoryTS enthält Funktionen für das Testen vom echten vs. unechten langen Gedächtnis. Nicht stationäre ARFIMA-Modelle können mit dem Paket nsarfima simuliert und geschätzt werden.

8 Mehrdimensionale Zeitreihen

In den vorangegangenen Kapiteln werden Methoden und Modelle für univariate Zeitreihen behandelt. Liegen Folgen von Beobachtungen verschiedener Variablen vor, kann man die entsprechenden Zeitreihen mit dem bisher dargestellten Instrumentarium jeweils getrennt analysieren. Bei einer solchen Datensituation ist aber in der Regel die gemeinsame Analyse von größerem Interesse, insbesondere in Hinsicht auf Abhängigkeiten und spezielle Formen von Interaktionen zwischen den Reihen. Mittels komplexerer, alle Reihen umfassender Modelle können gegebenenfalls auch genauere Prognosen erstellt werden.

Die Ausführungen beschränken sich hier auf zweidimensionale Zeitreihen. Der Übergang zu höheren Dimensionen sollte nach dem Studium dieses Kapitels leicht fallen. Mehr Dimensionen in der Darstellung zu berücksichtigen, würde jedoch zu einem unangemessenen Mehraufwand führen.

8.1 Kenngrößen mehrdimensionaler Zeitreihen

Kenngrößen im Zeitbereich

(X_t) , (Y_t) seien zwei stationäre Prozesse. Anstatt diese als getrennte Größen zu betrachten, können sie auch als Komponenten einer zweidimensionalen Zeitreihe angesehen werden. Die Betrachtung von (X_t) , (Y_t) und von $(X_t, Y_t)' = \left(\begin{smallmatrix} X_t \\ Y_t \end{smallmatrix}\right)$ wird im Folgenden als gleichwertig angesehen. Das hat etwa keinen Einfluss auf die Darstellung. Sie wird praktischer Weise sowieso in der Form geschehen, dass die Reihen als Funktionen der Zeit in einem oder in zwei Diagrammen gezeichnet werden.

Beispiel 8.1 (Jahresringe von Bäumen – Fortsetzung von Seite 3)

Die Fragestellung der Jahresringe von Bäumen wurde bereits im Beispiel 1.5 angegangen. Um für eine Region die Möglichkeit der Altersbestimmung mittels Jahresringen zu schaffen, wurden am Tree Ring Laboratory der Universität von Arizona in Tucson für die Jahre 1907 bis 1972 die jährlichen Niederschlagsmengen (X) sowie für eine Baumart die durchschnittliche Dicke der Jahresringe (Y) ermittelt. Zu den Daten und der Studie siehe Fritts et al. (1971). Die beiden Reihen sind in der Abbildung 8.1 dargestellt. Aus Darstellungsgründen wurde der Niederschlag um das Niveau der anderen Reihe angehoben.

Neben den zugehörigen, auf die einzelnen Prozesse bezogenen Größen μ_X , μ_Y und ACF's $\gamma_X(\tau)$, $\gamma_Y(\tau)$ ist auch die Erfassung des Zusammenhangs der beiden Prozesse relevant. Dies geschieht mit den Kreuzkovarianzen bzw. den Kreuzkorrelationen. Damit dies sinnvoll wird, muss die gemeinsame Stationarität gefordert werden, d. h. auch die Kreuzkovarianzen dürfen nur von der Differenz der Zeitpunkte abhängen.

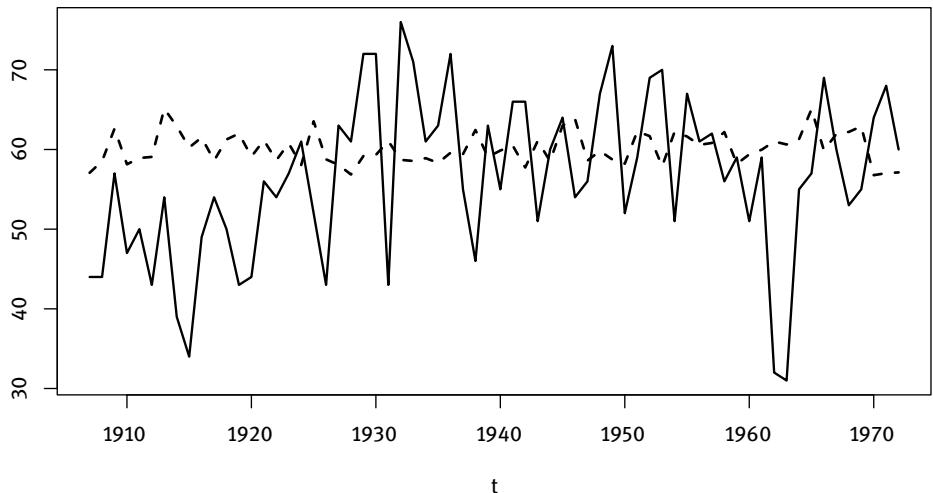


Abb. 8.1: BAUMRING-Reihen — : Ringstärke, - - - : Niederschlag

Definition 8.2 (Stationarität bivariater Prozesse)

Ein bivariater Prozess (X_t, Y_t) heißt *stationär* oder die beiden Prozesse (X_t) und (Y_t) heißen *gemeinsam stationär*, wenn sie einzeln stationär sind, und wenn die Kreuzkovarianzen $E((X_{t+\tau} - \mu_X)(Y_t - \mu_Y))$ nur vom Lag τ und nicht vom Zeitpunkt t abhängen.

Für gemeinsam stationäre Prozesse sind die *Kreuzkovarianz- und -korrelationsfunktionen*, kurz CCF, definiert durch:

$$\gamma_{XY}(\tau) = E((X_{t+\tau} - \mu_X)(Y_t - \mu_Y)), \quad (8.1)$$

$$\rho_{XY}(\tau) = \frac{\gamma_{XY}(\tau)}{\sqrt{\gamma_X(0)\gamma_Y(0)}}. \quad (8.2)$$

Dass neben der Kreuzkorrelationsfunktion gelegentlich auch die Kreuzkovarianzfunktion mit CCF abgekürzt wird, ist wie bei der ACF unproblematisch, da sie bis auf einen Faktor übereinstimmen. Für die Beschreibung der Abhängigkeitsstruktur wählt man vorzugsweise die Kreuzkorrelationsfunktion; die Varianzen der beiden Komponentenprozesse können sehr unterschiedlich sein.

In der Regel gilt $\gamma_{XY}(\tau) \neq \gamma_{XY}(-\tau)$, so dass die CCF für positive wie auch für negative Lags zu bestimmen ist. Andererseits ist

$$\gamma_{XY}(-\tau) = E((X_{t+(-\tau)} - \mu_X)(Y_t - \mu_Y)) = E((X_t - \mu_X)(Y_{t+\tau} - \mu_Y)) = \gamma_{YX}(\tau).$$

Die Kreuzkorrelation zum Lag 0 misst den (linearen) Zusammenhang zwischen den beiden zeitlich nicht gegeneinander verschobenen Prozessen (X_t) und (Y_t) . Diese können unkorreliert oder nur mäßig stark korreliert sein; so muss $\rho_{XY}(0)$ natürlich nicht gleich eins sein.

Beispiel 8.3 (Ein AR[1]-Prozess und seine Innovationen)

$Y_t = \alpha Y_{t-1} + \varepsilon_t$ sei ein stationärer AR[1]-Prozess mit $E(Y_t) = 0$. Es soll $\gamma_{Y\varepsilon}(\tau)$ bestimmt werden. Dazu wird (Y_t) als unendlicher MA-Prozess geschrieben, $Y_t = \sum_{u=0}^{\infty} \alpha^u \varepsilon_{t-u}$. Daraus ergibt sich (ggfs. mit der Setzung $0^0 = 1$):

$$\gamma_{Y\varepsilon}(\tau) = E\left(\left[\sum_{u=0}^{\infty} \alpha^u \varepsilon_{t+\tau-u}\right] \varepsilon_t\right) = \begin{cases} 0 & \tau < 0 \\ \alpha^\tau \sigma_\varepsilon^2 & \tau \geq 0 \end{cases}.$$

Weiter gilt wegen $\sigma_Y^2 = \sigma_\varepsilon^2 / (1 - \alpha^2)$:

$$\rho_{Y\varepsilon}(\tau) = \frac{\gamma_{Y\varepsilon}(\tau)}{\sigma_Y \sigma_\varepsilon} = \begin{cases} 0 & \tau < 0 \\ \alpha^\tau \sqrt{1 - \alpha^2} & \tau \geq 0 \end{cases}.$$

Speziell ist also $\gamma_{\varepsilon Y}(0) = \sqrt{1 - \alpha^2}$ für $\alpha \neq 0$ von 1 verschieden.

Die *empirische Kreuzkvarianzfunktion* zweier Zeitreihen $(x_t), (y_t)$ wird analog zur empirischen ACF gebildet. Allerdings muss wie bei der theoretischen CCF auf den Unterschied bei positiven und negativen Lags geachtet werden:

$$c_{XY}(\tau) = \begin{cases} \frac{1}{N} \sum_{t=1}^{N-\tau} (x_{t+\tau} - \bar{x})(y_t - \bar{y}) & \text{für } \tau \geq 0 \\ \frac{1}{N} \sum_{t=1-\tau}^N (x_{t+\tau} - \bar{x})(y_t - \bar{y}) & \text{für } \tau < 0 \end{cases}. \quad (8.3)$$

Die zugehörige *empirische Kreuzkorrelationsfunktion* lautet:

$$r_{XY}(\tau) = \frac{c_{XY}(\tau)}{\sqrt{c_X(0)} \sqrt{c_Y(0)}}. \quad (8.4)$$

Beispiel 8.4 (Jahresringe von Bäumen – Fortsetzung von Seite 187)

Für die beiden Reihen Baumringstärke und Niederschlag erhält man die in Abb. 8.2 dargestellte Kreuzkorrelationsfunktion. Die Kreuzkorrelation beim Lag null ist negativ. Das lässt vermuten, dass ein erhöhter Niederschlag sich zunächst hinderlich auf das Wachstum auswirkt. Beim Lag -9 ist die Korrelation erhöht. Dies würde auf einen positiven Einfluss erhöhten Niederschlags nach neun Jahren hinweisen. Allerdings sind die Korrelationen insgesamt so schwach, dass diese Interpretationen wohl als nicht angemessen einzustufen sind.

Werden die Kennfunktionen für jedes Lag in einer gemeinsamen *Kovarianzmatrix* zusammen gefasst,

$$\boldsymbol{\Gamma}_{XY}(\tau) = \begin{pmatrix} \gamma_X(\tau) & \gamma_{XY}(\tau) \\ \gamma_{YX}(\tau) & \gamma_Y(\tau) \end{pmatrix}, \quad (8.5)$$

so ergibt sich aus den oben angegebenen Beziehungen:

$$\boldsymbol{\Gamma}(-\tau) = \boldsymbol{\Gamma}(\tau)'.$$

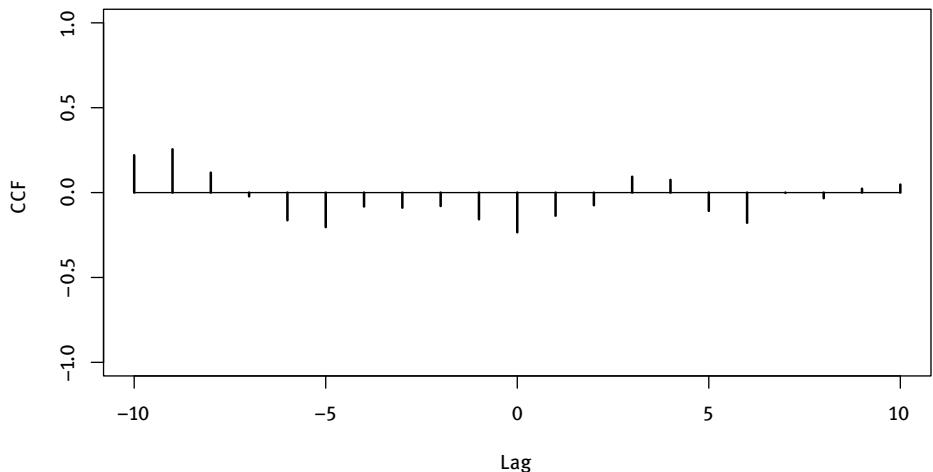


Abb. 8.2: Kreuzkorrelation von Baumringdicke mit Niederschlag

Die gleiche Eigenschaft hat die *Korrelationsmatrix*:

$$\mathbf{R}_{XY}(\tau) = \begin{pmatrix} \rho_X(\tau) & \rho_{XY}(\tau) \\ \rho_{YX}(\tau) & \rho_Y(\tau) \end{pmatrix}.$$

Der Einfachheit halber nennt man $\Gamma_{XY}(\tau)$ die *Kovarianzfunktion* des bivariaten Prozesses $(X_t, Y_t)'$ und $\mathbf{R}_{XY}(\tau)$ seine *Korrelationsfunktion*.

Beispiel 8.5 (Ein spezieller bivariater Prozess)

Sei $(X_t, Y_t)'$ ein bivariater Prozess, dessen Komponenten die Beziehung

$$Y_t = \beta X_t + \varepsilon_t$$

erfüllen, wobei (ε_t) ein mit (X_t) unkorrelierter, bei null zentrierter White-Noise-Prozess ist. Dann gilt:

$$\gamma_Y(\tau) = \beta^2 \gamma_X(\tau) + \gamma_\varepsilon(\tau),$$

$$\gamma_{XY}(\tau) = \beta \gamma_X(\tau),$$

$$\gamma_{YX}(\tau) = \gamma_{XY}(-\tau) = \beta \gamma_X(\tau).$$

Zusammen ergibt dies:

$$\Gamma_{XY}(\tau) = \begin{cases} \gamma_X(0) \begin{pmatrix} 1 & \beta \\ \beta & \beta^2 + \sigma_\varepsilon^2 / \gamma_X(0) \end{pmatrix} & \text{für } \tau = 0 \\ \gamma_X(\tau) \begin{pmatrix} 1 & \beta \\ \beta & \beta^2 \end{pmatrix} & \text{für } \tau \neq 0. \end{cases}$$

Definition 8.6 (Bivariater White-Noise-Prozess)

Ein bivariater Prozess $(U_t, V_t)'$ heißt *bivariater White-Noise-Prozess*, wenn für beliebige Zeitpunkte t_1, \dots, t_k die zugehörigen Zufallsvektoren $(U_{t_1}, V_{t_1})'$, $(U_{t_2}, V_{t_2})', \dots, (U_{t_k}, V_{t_k})'$ vollständig unabhängig sind.

Die Unabhängigkeitsforderung in der Definition bedeutet nicht, dass für jedes t die Komponenten ebenfalls unabhängig oder unkorreliert sein müssen. Vielmehr ist die folgende Kovarianzfunktion zugelassen:

$$\boldsymbol{\Gamma}_{UV}(\tau) = \begin{cases} \boldsymbol{\Sigma} & \text{für } \tau = 0 \\ \mathbf{0} & \text{für } \tau \neq 0. \end{cases}$$

Beispiel 8.7 (Bivariater White-Noise-Prozess mit linear abhängigen Komponenten)

Sei U_t ein univariater White-Noise-Prozess und sei $V_t = cU_t$. Zwischen U_t und V_t liegt also eine exakte lineare Beziehung vor. Für die CCF gilt aber:

$$\boldsymbol{\Gamma}_{UV}(0) = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & c\sigma^2 \\ c\sigma^2 & c^2\sigma^2 \end{pmatrix}, \quad \boldsymbol{\Gamma}_{UV}(\tau) = \mathbf{0} \quad \text{falls } \tau \neq 0.$$

Trotz der exakten linearen Beziehung ist $(U_t, V_t)'$ ein zweidimensionaler White-Noise-Prozess.

Die empirischen Gegenstücke zu den theoretischen Kreuzkovarianzen sind analog zu den univariaten Größen für kovarianzstationäre Prozesse unter geeigneten Voraussetzungen konsistente Schätzer. Die Varianz der Schätzer hängt nicht nur von den Kreuzkorrelationen, sondern auch von den Autokovarianzen ab. Es wird hier nur der Spezialfall unabhängiger Prozesse betrachtet.

Satz 8.8 (Asymptotik der Kreuzkovarianzen bei unabhängigen Prozessen)

Seien (X_t) und (Y_t) zwei unabhängige, gemeinsam stationäre Prozesse. Dann sind $\hat{\rho}_{XY}(\tau)$, $\hat{\rho}_{XY}(\kappa)$ asymptotisch gemeinsam normalverteilt mit Erwartungswerten $\rho_{XY}(\tau) = \rho_{XY}(\kappa) = 0$ und Kovarianzen

$$\text{Cov}(\hat{\rho}_{XY}(\tau), \hat{\rho}_{XY}(\kappa)) \doteq \frac{1}{N} \sum_{v=-\infty}^{\infty} \rho_X(v + \tau - \kappa) \rho_Y(v). \quad (8.6)$$

Für die Varianzen der empirischen Kreuzkorrelationen gilt also unter den Voraussetzungen des Satzes

$$\text{Var}(\hat{\rho}_{XY}(\tau)) \doteq \frac{1}{N} \sum_{v=-\infty}^{\infty} \rho_X(v) \rho_Y(v). \quad (8.7)$$

Ist speziell einer der beiden Prozesse White-Noise, so reduziert sich die Varianz zu:

$$\text{Var}(\hat{\rho}_{XY}(\tau)) \doteq \frac{1}{N}. \quad (8.8)$$

Beispiel 8.9 (Zwei AR[1]-Prozesse)

Seien $X_t = \alpha X_{t-1} + \varepsilon_t$ und $Y_t = \beta Y_{t-1} + \eta_t$ zwei unabhängige AR[1]-Prozesse. Dann sind $\rho_X(\tau) = \alpha^\tau$ und $\rho_Y(\tau) = \beta^\tau$. Aus (8.7) folgt

$$\text{Var}(\hat{\rho}_{XY}(\tau)) \doteq \frac{1}{N} \frac{1 + \alpha\beta}{1 - \alpha\beta}.$$

Sind also die Koeffizienten α, β groß, so ist die Varianz der Kreuzkorrelationen allein aufgrund der inneren Struktur der beiden Prozesse hoch. Für $N = 100$, $\alpha = 0.9$ und $\beta = 0.8$ kann man 50 % der Korrelationen außerhalb des Bereiches ± 0.167 und noch 10 % außerhalb von ± 0.40 erwarten, auch wenn es keine Abhängigkeit zwischen den beiden Reihen gibt.

In manchen Anwendungen wünscht man eine formale Überprüfung der Hypothese, dass zwei Zeitreihen Realisationen unkorrelierter Prozesse sind. Nun können wie im Beispiel 8.9 die Kreuzkovarianzen allein deswegen sehr groß sein, weil die Reihen eine hohe Autokorrelation aufweisen. Andererseits ist es nahe liegend, die Unabhängigkeit zweier univariater Zeitreihen zu testen, indem man die Hypothese überprüft, dass die Kreuzkovarianzen alle null sind.

Die Lösung des Problems wird durch (8.8) nahe gelegt. Man betrachtet die CCF von gefilterten Reihen anstelle der der Reihen selbst. Dabei wird (mindestens) eine Reihe so transformiert, dass der Output White-Noise ist. Diese Filterung wird als *Pre-whitening* bezeichnet. Mit (8.8) sollten dann die Output-Reihen eine Kreuzkorrelation von null haben. Die Rechtfertigung für den Rückschluss auf die Ausgangsreihen gibt der folgende Satz.

Satz 8.10 (Unkorreliertheit von Zeitreihen)

Zwei gemeinsam kovarianzstationäre Zeitreihen (X_{1t}) und (X_{2t}) sind genau dann unkorreliert (und folglich unabhängig, wenn sie zusätzlich Gauß-Prozesse sind), wenn die gefilterten Reihen $Y_{1t} = g_1(B)X_{1t}$ und $Y_{2t} = g_2(B)X_{2t}$ unkorreliert sind. Dabei sind $g_1(B)$ und $g_2(B)$ zwei invertierbare Filter.

Aufgrund des Satzes kann zum Testen der Unabhängigkeit zweier gemeinsam normalverteilter, kovarianzstationärer Zeitreihen folgendermaßen vorgegangen werden:

1. Den Reihen werden univariate ARMA-Modelle angepasst.
2. Für die Residualreihen wird die Kreuzkorrelationsfunktion berechnet.
3. Es wird ein Test durchgeführt, ob die Kreuzkorrelationen sich signifikant von null unterscheiden.

Beim formalen Testen ist zu beachten, dass unter der Nullhypothese die Kreuzkorrelationen der Residualreihen approximativ unabhängig und identisch verteilt sind mit Erwartungswert 0 und Varianz $1/N$. Wenn also $2M + 1$ Kreuzkorrelationen (üblicherweise für die Lags $-M$ bis M) getestet werden, sollte man das Signifikanzniveau $\alpha^{1/(2M+1)}$ für die einzelnen Lags verwenden, um insgesamt das Signifikanzniveau α sicherzustellen. Meist begnügt man sich allerdings mit der Darstellung der CCF, wo bei die Grenzen $\pm 2/\sqrt{N}$ mit eingezeichnet sind.

Bei der Anpassung von Modellen verzichtet man in aller Regel auf solche α -Adjustierungen. Dann sucht man „nur“ ein geeignetes Modell, zum Beispiel eines, mit dem im Anschluss Prognosen erstellt werden sollen. Es geht hier um eine Approximation an die Daten und nicht um eine formale Aussage im Sinne, dass eine Theorie oder Behauptung widerlegt werden soll.

Beispiel 8.11 (Temperatur und Speiseeisverbrauch)

Beeinflusst die Temperatur signifikant den Speiseeisverbrauch? Für den jeweils 30 tägigen Verbrauch Y_t von März 1951 bis Juli 1953 (in pints pro Kopf) und die Durchschnittstemperatur X_t (in Grad Fahrenheit) erhielt man die folgenden Reihen. Diese bivariate Zeitreihe SPEISEEIS ist Koteswara (1970) entnommen.

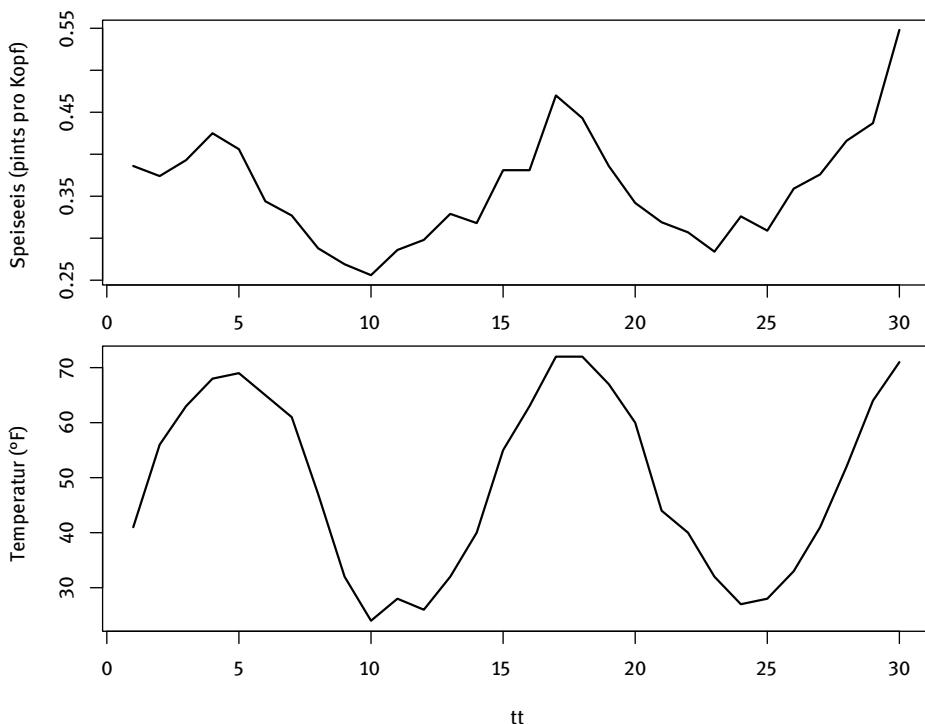


Abb. 8.3: Speiseeisverbrauch und Temperatur

Die Kreuzkorrelationsfunktion der Originalreihen, siehe Abbildung 8.4, weist zunächst auf einen starken Zusammenhang hin. Werden jedoch den Reihen jeweils AR-Prozesse angepasst, so wird in beiden Fällen ein AR[3]-Modell ausgewählt:

$$X_t = 1.302X_{t-1} - 0.310X_{t-2} - 0.308X_{t-3} + \varepsilon_t$$

$$Y_t = 0.857Y_{t-1} + 0.227Y_{t-2} - 0.525Y_{t-3} + \eta_t .$$

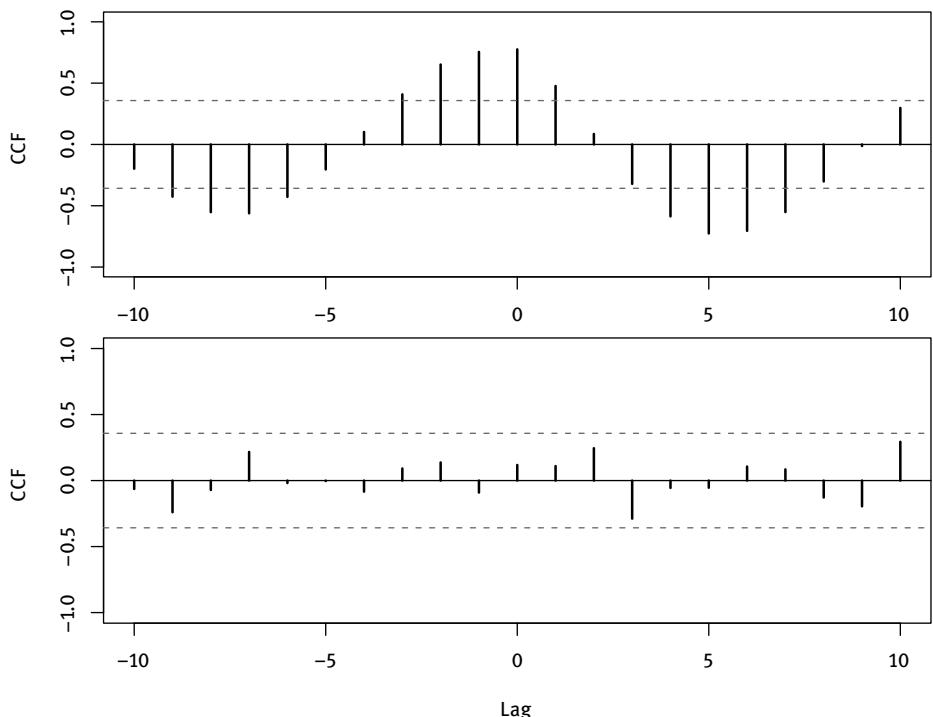


Abb. 8.4: r_{XY} für die originalen (oben) und gefilterten (unten) Reihen $X = \text{Temperatur}$ und $Y = \text{Speiseisverbrauch}$

Die Residualreihen zeigen keine signifikante Kreuzkorrelation mehr, vgl. Abbildung 8.4. Die inneren Strukturen erklären die gesamte Dynamik. Diese ist dabei natürlich für beide Zeitreihen durch die Jahreszeit bedingt. Somit liegt hier ein Fall von Scheinkorrelation vor.

R-Code:

```

1 data(ICECREAM)
2 y <- ICECREAM
3 plot(y)
4 a1 <- arima(y[,1],order=c(3,0,0))
5 a2 <- arima(y[,2],order=c(3,0,0))
6 par(mfrow=c(2,1))
7 ccf(y[,1],y[,2],lag.max=10)
8 ccf(a1$resid,a2$resid,lag.max=10)
9 par(mfrow=c(1,1))

```

Die Zeitreihen liegen als bivariate Matrix (mit Zeitreihenattribut) vor. Mit der Funktion `plot` werden sie entsprechend grafisch dargestellt. Für beide Reihen werden AR[3]-Modelle geschätzt. Dann werden die Kreuzkorrelationsfunktionen der Originalreihen und der Residualreihen berechnet und zugleich geplottet.

Kreuzspektren

$(X_t, Y_t)'$ sei ein stationärer bivariater Prozess mit den zugehörigen univariaten Autokovarianzfunktionen $\gamma_X(\tau)$, $\gamma_Y(\tau)$, den Spektren

$$f_X(\lambda) = \gamma_X(0) + 2 \sum_{\tau=1}^{\infty} \gamma_X(\tau) \cos 2\pi\lambda\tau ,$$

$$f_Y(\lambda) = \gamma_Y(0) + 2 \sum_{\tau=1}^{\infty} \gamma_Y(\tau) \cos 2\pi\lambda\tau ,$$

und der Kreuzkovarianzfunktion $\gamma_{XY}(\tau)$.

Um den Zusammenhang der beiden Komponenten des Prozesses im Frequenzbereich zu beschreiben, wird analog zum univariaten Fall die Fouriertransformierte von $\gamma_{XY}(\tau)$ betrachtet. Da die CCF jedoch nicht symmetrisch ist, $\gamma_{XY}(\tau) \neq \gamma_{XY}(-\tau)$, ist die Fouriertransformierte

$$f_{XY}(\lambda) = \sum_{\tau=-\infty}^{\infty} \gamma_{XY}(\tau) e^{i2\pi\lambda\tau} \quad (8.9)$$

eine komplexwertige Funktion. Die drei Darstellungsformen komplexer Zahlen

$$z = a + i \cdot b = |z|(\cos(\phi) + i \sin(\phi)) = |z|e^{i\phi} ,$$

mit $|z| = \sqrt{a^2 + b^2}$ und $\tan(\phi) = b/a$, führen zu verschiedenen reellwertigen Funktionen, die jeweils wichtige Aspekte von $f_{XY}(\lambda)$ erfassen.

Definition 8.12 (Kreuzspektrum)

$(X_t, Y_t)'$ sei ein stationärer bivariater Prozess mit der Kreuzkovarianzfunktion $\gamma_{XY}(\tau)$ und den Spektraldichten $f_X(\lambda)$, $f_Y(\lambda)$. Dann heißt die in (8.9) angegebene Funktion $f_{XY}(\lambda)$ das *Kreuzspektrum*. Weiter heißen

$$C_{XY}(\lambda) = \sum_{\tau=-\infty}^{+\infty} \gamma_{XY}(\tau) \cos 2\pi\lambda\tau \quad \text{das Kospektrum,} \quad (8.10a)$$

$$Q_{XY}(\lambda) = \sum_{\tau=-\infty}^{+\infty} \gamma_{XY}(\tau) \sin 2\pi\lambda\tau \quad \text{das Quadraturspektrum,} \quad (8.10b)$$

$$\phi_{XY}(\lambda) = \tan^{-1} \left(\frac{-Q_{XY}(\lambda)}{C_{XY}(\lambda)} \right) \quad \text{die Phase und} \quad (8.10c)$$

$$K_{XY}(\lambda) = \frac{\sqrt{C_{XY}(\lambda)^2 + Q_{XY}(\lambda)^2}}{\sqrt{f_X(\lambda)} \sqrt{f_Y(\lambda)}} \quad \text{die Kohärenz.} \quad (8.10d)$$

Kospektrum und Quadraturspektrum sind die beiden Bausteine, aus denen sich die beiden anderen reellwertigen Funktionen ergeben. Diese sind direkt interpretierbar. Die Phase beschreibt für jedes λ die Verschiebung der in den beiden Reihen enthaltenen Wellen dieser Frequenz gegeneinander. Mit der Phase können also Lead- und Lag-Relationen zwischen zwei Zeitreihen studiert werden. Es ist dabei möglich, dass in der einen Reihe die kurzen Wellen einen Vorlauf gegenüber denen in der zweiten haben, während sich das Bild bei langen Wellen umkehrt. Mit der Kohärenz wird für jede Frequenz λ die Stärke des Zusammenhangs der in den beiden Reihen enthaltenen harmonischen Wellen erfasst. Für die Kohärenz gilt

$$0 \leq K_{XY}(\lambda) \leq 1 .$$

Die beiden Komponenten hängen bei der Frequenz λ umso stärker linear zusammen, je näher $K_{XY}(\lambda)$ bei eins liegt.

Eine Kreuzspektralanalyse ist auch unter dem Gesichtspunkt von Interesse, den Effekt der Anwendung eines linearen Filters auf eine univariate Zeitreihe zu beschreiben. Dann kann man eine Kreuzspektralanalyse auch einsetzen, um zu sehen, ob von einer bivariaten Zeitreihe die eine als gefilterte Version der anderen betrachtet werden kann. Dies ist das bivariate Zeitreihen-Analogon zur einfachen Regression.

Bei der Schätzung des Ko- und des Quadraturspektrums stößt man auf die gleichen Probleme wie beim Schätzen der univariaten Spektren. Daher sind speziell die empirischen Gegenstücke zum Kospektrum und zum Quadraturspektrum geeignet zu glätten. Das kann ganz analog zum univariaten Fall geschehen, indem über benachbarte Frequenzen gemittelt oder ein Lagfenster auf die CCF angewendet wird. Hat man diese Schätzungen, so können sie in die Formeln für Kohärenz und Phase eingesetzt werden, um zu Schätzungen dieser Größen zu gelangen. Werden alle Funktionen in der gleichen Weise geglättet, so gilt:

$$0 \leq \frac{\hat{C}_{XY}(\lambda)^2 + \hat{Q}_{XY}(\lambda)^2}{\hat{f}_X(\lambda)\hat{f}_Y(\lambda)} = \hat{K}_{XY}^2(\lambda) \leq 1 .$$

Beispiel 8.13 (Grundwasserspiegel)

Die Stadt Werther versorgt ihre Bürger durch eigene Tiefenbrunnen. Als 1979 ein zusätzlicher Brunnen angelegt wurde, kam es im Rat der Stadt und unter den Bürgern zu kontroversen Auffassungen in einer Reihe von Fragen. Dazu gehörte auch die Frage, in welcher Weise der Regen die Förderung an den Brunnen den Grundwasserspiegels wieder ausgleicht. Um diese Fragestellung im Ansatz zu untersuchen, wird die bivariate Zeitreihe LITH betrachtet, deren Komponenten tägliche Messwerte sind. Die eine ist die Reihe der Niederschlagsmengen X_t , die andere der Grundwasserstand des Peilbrunnens Lith; Genauer ist Y_t die Peiltiefe; damit

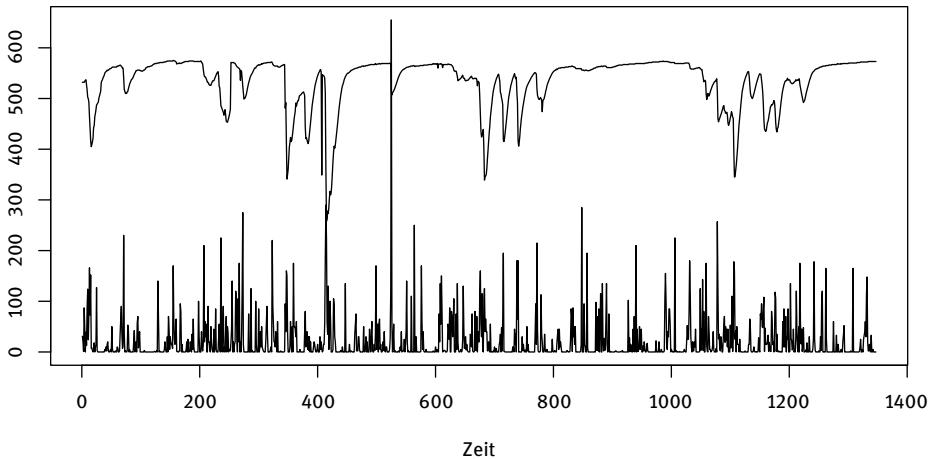


Abb. 8.5: Niederschlag und Grundwasserspiegel des Peilbrunnens Lith

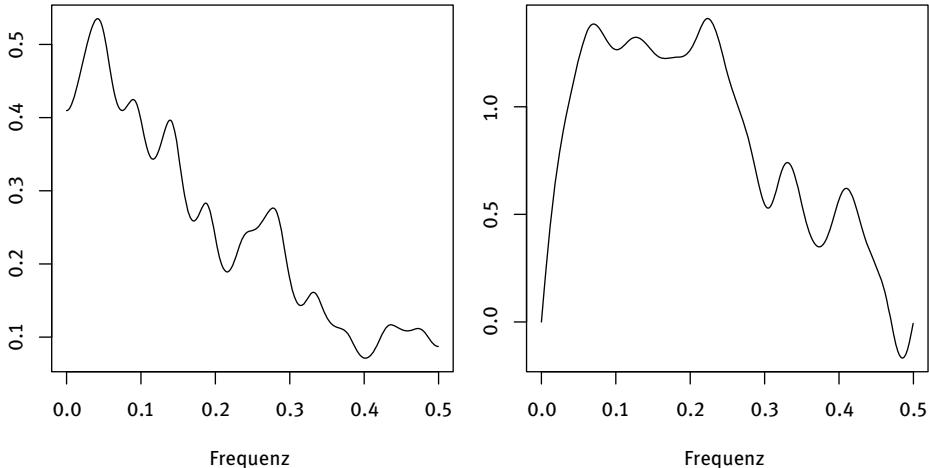


Abb. 8.6: Kohärenz- (links) und Phasendiagramm (rechts) von Niederschlag und Grundwasserspiegel

sind die Werte von Y_t umso größer, je geringer der Grundwasserstand ist. Der Zeitraum reicht vom 23.1.1980 bis zum 30.9.1983. Ausführlich wird dieser Fall in Naeve und Steinecker (1983) dargestellt.

Kohärenz und Phasendiagramm sind in der Abbildung 8.6 dargestellt. Das Kohärenzdiagramm zeigt, dass für kleinere Frequenzen der Zusammenhang stärker ist, mit höheren Frequenzen zunächst abnimmt und dann noch einmal geringfügig ansteigt. Der Zusammenhang für kleine Frequenzen spiegelt wohl die Jahreszeit-

lichen Schwankungen wider. Kurzfristig wird jedoch der Grundwasserpegel von der Niederschlagsmenge nur wenig beeinflusst.

R-Code:

```

1 data(LITH)
2 N <- LITH[,1]
3 G <- LITH[,2]
4 bi <- bispecies(N,G,20,win="bartlett")
5 par(mfrow=c(1,2))
6 plot(bi[,1],bi[,2],type="l")
7 plot(bi[,1],bi[,3],type="l")
8 par(mfrow=c(1,1))

```

Die Funktion `bispecies` zum Schätzen der Kohärenz und Phase befindet sich in dem Paket `tsapp`. In der ersten Spalte des Outputs sind die Frequenzen, in den beiden folgenden die zugehörigen Werte von Kohärenz und Phase.

Der auf der Kreuzkorrelationsfunktion basierende Test auf Unkorreliertheit zweier Zeitreihen hat in der Form kein Gegenstück im Frequenzbereich. Dies ist eine Konsequenz aus dem zweiten Punkt des folgenden Satzes.

Satz 8.14 (Kohärenz und Kreuzkorrelation)

Sei $(X_t, Y_t)'$ ein kovarianzstationärer bivariater Prozess. Dann gilt:

1. Sind (\tilde{X}_t) und (\tilde{Y}_t) gefilterte Versionen von (X_t) und (Y_t) , dann hat $(\tilde{X}_t, \tilde{Y}_t)'$ die gleiche Kohärenz wie $(X_t, Y_t)'$.
2. Die univariaten Prozesse (X_t) und (Y_t) sind genau dann nicht kreuzkorreliert, wenn das Kohärenzspektrum für alle Frequenzen null ist.
3. Unter geeigneten formalen Voraussetzungen gilt für λ mit $K_{XY}(\lambda) = 0$, wenn die Spektralschätzungen auf den ungewichteten Auto- und Kreuzkorrelationen der Lags $-L/2, \dots, L/2$ basieren:

$$(L-1) \frac{\hat{K}_{XY}^2(\lambda)}{1 - \hat{K}_{XY}^2(\lambda)} \sim F_{2,2L-2}. \quad (8.11)$$

Auch wenn wie gesagt aufgrund des Punktes 1 kein Test auf Unabhängigkeit möglich ist, erlaubt doch der Punkt 2 zusammen mit der asymptotischen Aussage 3 einen heuristischen Check auf der Basis der Kohärenz. Dazu ist einfach in der Höhe des $1-\alpha$ -Prozentpunktes der $F_{2,2L-2}$ -Verteilung eine Hilfslinie in die grafische Darstellung der entsprechend (8.11) transformierten Kohärenzen einzutragen. An allen Frequenzen, an denen diese Hilfslinie übertroffen wird, liegt ein Zusammenhang vor.

Das *integrierte Kospektrum* (der transformierten Reihen) kann als Grundlage für einen visuellen Test dienen. Dieser visuelle Test wurde von Jenkins & Watts (1968,

S. 366) vorgeschlagen.

$$\frac{2}{N\sqrt{c_X(0)c_Y(0)}} \sum_{j=0}^k \hat{C}_{XY}\left(\frac{j}{N}\right), \quad k = 1, \dots, M = \lfloor \frac{N-1}{2} \rfloor.$$

Sind die beiden zugrunde liegenden Prozesse unkorreliert, so ist das theoretische integrierte Kospektrum identisch null. Folglich sollte die empirische Funktion in diesem Fall ebenfalls um null schwanken. In anderen Fällen wird man eine systematische Abweichung von null erleben.

Auf dem *kumulierten Phasenspektrum* basiert ein anderer Test. Für die einzelnen Fourierfrequenzen sind die Ordinaten des Phasenspektrums bei nicht-korrelierten White-Noise-Prozessen Rechteck-verteilt über dem Intervall $[-0.5, 0.5]$:

$$\hat{\phi}(\lambda_j) = \sim \mathcal{R}(-0.5, 0.5).$$

Das kumulierte Phasenspektrum $\hat{\Phi}_{XY}(\lambda_j) = \sum_{k=1}^j \hat{\phi}(\lambda_j)$ sollte sich daher um die 45° -Achse schlängeln. Als kritische Grenzen zum Niveau $\alpha = 0.05$ ergeben sich dabei entsprechend dem Kolmogorov-Anpassungstest $\pm 1.36\sqrt{N/2}$.

8.2 Mehrdimensionale lineare Zeitreihenmodelle

VARMA-Prozesse

Die in Kapitel 3 vorgestellten linearen Zeitreihenmodelle lassen sich zumindest konzeptionell leicht auf den mehrdimensionalen Fall erweitern. Oft ist es so, dass gerade mehrdimensionale reale Vorgänge – bisweilen über geeignete Approximationen – zu linearen Modellen führen. Das soll anhand eines Beispiels ausgeführt werden.

Beispiel 8.15 (Luchse und Schneehasen)

Bezeichne X_t die Anzahl der Luchse im Zeitpunkt t und Y_t die Anzahl der Schneehasen. Dann kann man versuchen, die ökologische Beziehung zwischen beiden Populationen durch ein dynamisches Modell zu modellieren. Idealisiert kann man als erste Näherung eine lineare Beziehung unterstellen:

$$\begin{aligned} X_t &= aX_{t-1} + bY_{t-1} + \varepsilon_{1t} \\ Y_t &= -cX_{t-1} + dY_{t-1} + \varepsilon_{2t}. \end{aligned}$$

Dabei sind ε_{1t} und ε_{2t} Zufallsschocks, die Komponenten eines bivariaten White-Noise-Prozesses ($\boldsymbol{\varepsilon}_t$) mit $\boldsymbol{\mu} = \mathbf{0}$ und $\boldsymbol{\Gamma}_{\boldsymbol{\varepsilon}}(0) = \boldsymbol{\Sigma}$ seien. Die Parameter $a, d > 0$ drücken die „natürlichen“ Geburtenraten aus. Man kann $d > a$ annehmen, da sich Schneehasen fleißiger vermehren als Luchse (dank ihrer Verwandtschaft mit Kaninchen). Die Parameter $b, c > 0$ erfassen die Interrelationen zwischen beiden

Populationen. Dabei geht b mit positivem Vorzeichen ein, da Luchse eher überleben, wenn sie Schneehasen fressen können, während c negativ eingeht, da es die Wirkung der Jäger auf die Beute beschreibt. Mittels der Matrzenschreibweise können die beiden Gleichungen zusammengefasst werden:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} a & b \\ -c & d \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

oder kurz, mit offensichtlichen Setzungen:

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t.$$

Kennt man die Parameter des Prozesses, also \mathbf{A} , $\boldsymbol{\mu}$ und die Kovarianzmatrix $\boldsymbol{\Sigma}$ des bivariaten White-Noise-Prozesses, so kann man den Prozess simulieren und sehen, ob er zu ähnlichen Realisationen führt wie die empirisch beobachteten Reihen.

Die in der Abbildung 8.7 dargestellten Reihen wurden mit den Parametern

$$\mathbf{A} = \begin{pmatrix} 0.5 & 0.3 \\ -1.5 & 0.8 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 2.0 & 1.8 \\ 1.8 & 3.0 \end{pmatrix},$$

erzeugt. Die hohe Korrelation zwischen den beiden Komponenten des White-Noise-Prozesses soll dabei suggerieren, dass Zufallsveränderungen in den äußereren Bedingungen für beide Populationen ähnlich positive bzw. negative Auswirkungen haben. Die Reihen zeigen dabei die Schwankungen um das jeweilige Niveau.

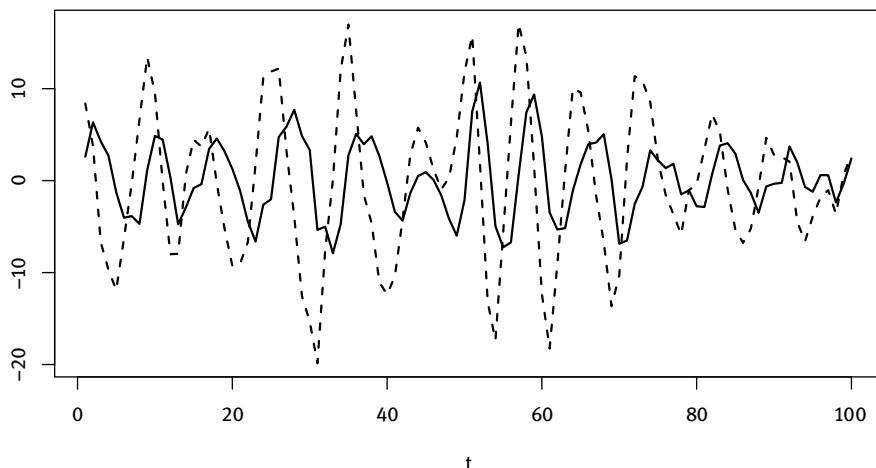


Abb. 8.7: Die Reihe LUCHSHASE, eine Realisation eines VAR[1]-Prozesses

Sieht man von dem multivariaten Rahmen ab, so entspricht die im Beispiel erhaltene Modellgleichung formal einem AR[1]-Prozess. Die bivariate Version stellt dementsprechend den einfachsten Fall eines multivariaten AR-Prozesses dar. Da jede Komponente von der jeweils anderen abhängt, ist eine gemeinsame Betrachtung beider nötig.

Allgemein spricht man bei mehrdimensionalen AR-Prozessen von *Vektor-Autoregressiven Prozessen*, kurz von *VAR-Prozessen*.

Als Bedingung dafür, dass ein VAR[p]-Prozess

$$\mathbf{y}_t - \mathbf{A}_1 \mathbf{y}_{t-1} - \cdots - \mathbf{A}_p \mathbf{y}_{t-p} = \boldsymbol{\varepsilon}_t$$

eine stationäre Lösung besitzt, erhält man in Analogie zum univariaten Fall, dass alle Nullstellen des durch $\det(\mathbf{I} - \mathbf{A}_1 z - \cdots - \mathbf{A}_p z^p)$ gegebenen Polynoms außerhalb des Einheitskreises liegen müssen; es muss also gelten:

$$\det(\mathbf{I} - \mathbf{A}_1 z - \cdots - \mathbf{A}_p z) \neq 0 \quad \text{für } |z| \leq 1 .$$

So wie univariate AR-Modelle zu mehrdimensionalen VAR-Modellen führen, lassen sich auch MA-Prozesse auf den multivariaten Fall übertragen. Es resultieren *Vektor-Moving-Average-Prozesse*, kurz *VMA-Prozesse*. Wird der Einfachheit halber ein Erwartungswertvektor von $\boldsymbol{\mu} = \mathbf{0}$ unterstellt, so lautet die Modellgleichung:

$$\mathbf{y}_t = \boldsymbol{\varepsilon}_t - \mathbf{B}_1 \boldsymbol{\varepsilon}_{t-1} - \cdots - \mathbf{B}_q \boldsymbol{\varepsilon}_{t-q} .$$

VMA-Prozesse lassen sich unter geeigneten Bedingungen als VAR-Prozesse darstellen. Diese Eigenschaft ist aus dem Eindimensionalen bekannt. Die Begründung läuft ganz ähnlich. Sei exemplarisch ein VMA-Prozess der Ordnung 1, ein VMA[1]-Prozess $\mathbf{y}_t = \boldsymbol{\varepsilon}_t - \mathbf{B}_1 \boldsymbol{\varepsilon}_{t-1}$ betrachtet. Man erhält rekursiv, wenn $(\mathbf{B}_1)^u \rightarrow 0$ für $u \rightarrow \infty$:

$$\boldsymbol{\varepsilon}_t = \mathbf{y}_t + \mathbf{B}_1 \boldsymbol{\varepsilon}_{t-1} = \mathbf{y}_t + \mathbf{B}_1 \mathbf{y}_{t-1} + \mathbf{B}_1 \mathbf{B}_1 \boldsymbol{\varepsilon}_{t-2} = \cdots = \mathbf{y}_t + \sum_{u=1}^{\infty} (\mathbf{B}_1)^u \mathbf{y}_{t-u}$$

Damit $(\mathbf{B}_1)^u$ gegen null geht, müssen die Eigenwerte von \mathbf{B}_1 betragsmäßig alle kleiner als eins sein. Äquivalent dazu ist, dass gilt: $\det(\mathbf{I} - \mathbf{B}_1 z) \neq 0$ für $|z| \leq 1$. Das ist die *Invertibilitätsbedingung* für einen VMA[1]-Prozess. Allgemein ist ein VMA[q]-Prozess *invertibel*, wenn die Wurzeln des Polynoms $\det(\mathbf{I} - \mathbf{B}_1 z - \cdots - \mathbf{B}_q z^q)$ außerhalb des Einheitskreises liegen:

$$\det(\mathbf{I} - \mathbf{B}_1 z - \mathbf{B}_2 z^2 - \cdots - \mathbf{B}_q z^q) \neq 0 \quad \text{für } |z| \leq 1 . \quad (8.12)$$

Die erstmals von Quenouille (1957) ausführlich betrachteten *Vektor-Autoregressive-Moving-Average-Prozesse*, *VARMA-Prozesse*, sind definiert durch

$$\mathbf{y}_t - \mathbf{A}_1 \mathbf{y}_{t-1} - \cdots - \mathbf{A}_p \mathbf{y}_{t-p} = \boldsymbol{\varepsilon}_t - \mathbf{B}_1 \boldsymbol{\varepsilon}_{t-1} - \cdots - \mathbf{B}_q \boldsymbol{\varepsilon}_{t-q} .$$

Dafür wird wieder kurz geschrieben:

$$\mathbf{A}(B) \mathbf{y}_t = \mathbf{B}(B) \boldsymbol{\varepsilon}_t . \quad (8.13)$$

(\mathbf{y}_t) kann auch in der Form eines VARMA-Prozesses geschrieben werden, bei dem die links stehende Matrizen $\mathbf{C}_1, \dots, \mathbf{C}_p$ des Matrixpolynoms $\mathbf{C}(B) = \mathbf{I} - \mathbf{C}_1 B - \cdots - \mathbf{C}_p B^p$

Diagonalgestalt haben:

$$\mathbf{C}(B)\mathbf{y}_t = \mathbf{D}(B)\boldsymbol{\varepsilon}_t . \quad (8.14)$$

Damit ist jede Komponente von \mathbf{y}_t durch eine endliche Anzahl eigener vergangener Elemente und zusammengesetzter Störterme der Form

$$\sum_{j=1}^q d_{ij}\varepsilon_{jt}$$

darstellbar. Jede Komponente besitzt also für sich eine univariate ARMA-Darstellung. Diese ist nicht eindeutig. Speziell kann man die Koeffizienten so wählen, dass die Kovarianzmatrix $\boldsymbol{\Sigma}$ des Noise-Prozesses $\boldsymbol{\varepsilon}_t$ gleich der Einheitsmatrix ist.

Neben die Stationaritätsbedingung für den VAR-Teil tritt die Invertibilitätsbedingung für den MA-Teil. Unter der Voraussetzung der Stationarität erhält man eine VMA[∞]-Darstellung des VARMA-Prozesses (\mathbf{y}_t) zu

$$\mathbf{y}_t = \mathbf{A}^{-1}(B)\mathbf{B}(B)\boldsymbol{\varepsilon}_t , \quad (8.15)$$

bzw. ausführlicher mit $\Delta(B) = \mathbf{A}^{-1}(B)\mathbf{B}(B)$:

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \sum_{u=0}^{\infty} \begin{pmatrix} \delta_{(11)u} & \delta_{(12)u} \\ \delta_{(21)u} & \delta_{(22)u} \end{pmatrix} \begin{pmatrix} \varepsilon_{(1)t-u} \\ \varepsilon_{(2)t-u} \end{pmatrix} . \quad (8.16)$$

Es gibt eine multivariate Version des Woldschen Zerlegungssatzes, vgl. Seite 88. Diese weist auf die Bedeutung von VMA-Modellen für stationäre Zeitreihen hin. Damit werden dann auch VARMA-Modelle als Approximationen gerechtfertigt.

Ein besonderer Nutzen der VMA[∞]-Darstellung (8.16) des VAR-Prozesses (\mathbf{y}_t) besteht darin, dass die Koeffizientenfolgen ($\delta_{(ij)u}$) als *Impakt-Multiplikatoren* interpretiert werden können. Man kann hieran studieren, wie sich einzelne Impulse oder Zufallsschocks in einer Komponente auf die andere Komponente auswirken. So erfasst $\delta_{(21)0}$ die momentane Auswirkung eines Impulses in der ersten Komponente, konkret den Effekt einer Änderung von ε_{1t} um 1 Einheit, auf Y_{2t} . Diese Art von Untersuchungen rangiert auch unter dem Namen *Impulsresponseanalyse*. Bisweilen wird der Response der Komponente Y_{it} auf einen Einheits-Schock in der Komponente Y_{jt} grafisch in Abhängigkeit von den Lags dargestellt, um eine visuelle Einschätzung der dynamischen Wechselwirkungen zu erlauben.

Beispiel 8.16 (Investition, Einkommen und Konsum)

Lütkepohl (2005) betrachtet den dreidimensionalen Vektor-Prozess $(Y_{1t}, Y_{2t}, Y_{3t})'$ aus den Komponenten Investition, verfügbares Einkommen und Konsumausgaben. Von Benkowitz (2000) wurde mit einem in der Programmiersprache GAUSS geschriebenen Programm die in Abbildung 8.8 wiedergegebene Schätzung der Impuls-Response-Funktionen für die einmal differenzierten Reihen auf der Basis von Quartalswerten von 1960 bis 1978 bestimmt. Überlagert sind Konfidenzgrenzen, die mit dem Bootstrap-Verfahren ermittelt wurden.

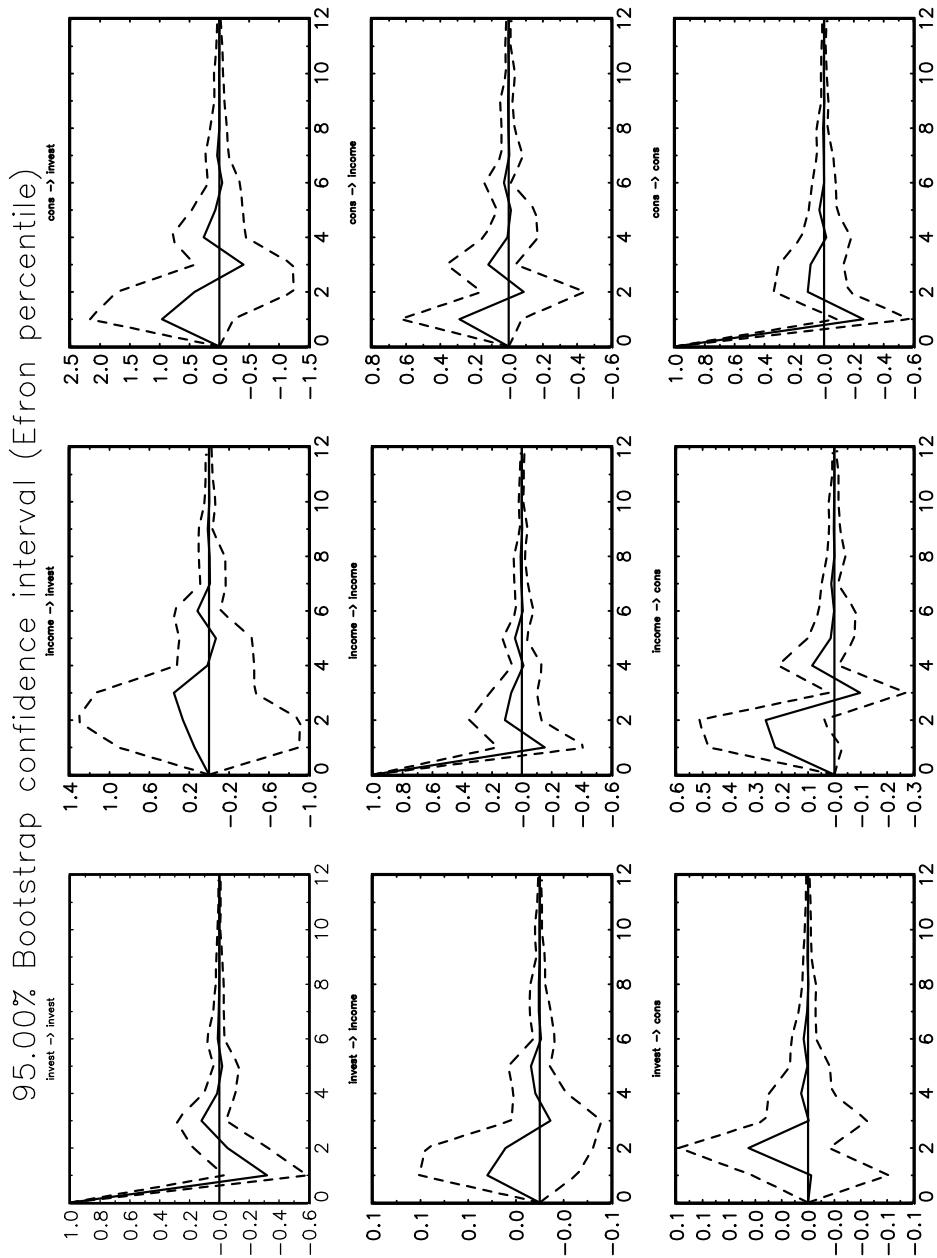


Abb. 8.8: Schätzung der Impulsantwortfunktionen mit Konfidenzgrenzen

In praktischen Anwendungen wird fast ausschließlich die Modellklasse der VAR-Modelle für multivariate Zeitreihen herangezogen. Einmal ist die Anpassung von VARMA-Modellen natürlich sehr aufwändig; zum anderen kann man sich höchstens bei VAR-Prozessen noch eine inhaltliche Interpretation des letztlich gefundenen Modells erhoffen. Allerdings wurde mittels Simulationen gezeigt, dass durch die Verwendung von VARMA- anstelle von VAR-Modellen bessere Prognosen erzielt werden können. Ein weiteres Argument für die Berücksichtigung der vollen Modelle resultiert daraus, dass in praktischen Anwendungen die interessierenden Variablen häufig vor dem Modellierungsprozess transformiert werden. Das kann beispielsweise eine Saisonbereinigung oder eine zeitliche Aggregation sein. Sind Quartalsdaten von Interesse, können die entsprechenden Monatswerte addiert und Quartals-Durchschnitte gebildet werden. Auch werden häufig Aggregationen über Haushalte oder Regionen vorgenommen. Für Strukturanalysen und Prognosen sind daher die Auswirkungen linearer Transformationen von Interesse. Dies ist besonders wichtig vor dem Hintergrund, dass viele der benannten Aggregationen lineare Transformationen darstellen. Bei VARMA-Prozessen bleibt man dann in dieser Modellklasse.

Sei zur Illustration ein K -dimensionaler invertierbarer VMA[q]-Prozess $\mathbf{y}_t = \mathbf{u}_t + \mathbf{M}_1 \mathbf{u}_{(t-1)} + \cdots + \mathbf{M}_q \mathbf{u}_{(t-q)}$ gegeben. Eine lineare Transformation von \mathbf{y}_t entspricht der Multiplikation mit einer Matrix. Sei also \mathbf{F} eine $(M \times K)$ -Matrix vom Rang M . Multiplikation von \mathbf{y}_t mit \mathbf{F} ergibt den M -dimensionalen Prozess $\mathbf{z}_t = \mathbf{F}\mathbf{y}_t$, der eine invertierbare VMA[\tilde{q}]-Darstellung mit $\tilde{q} \leq q$ besitzt. Dieses Resultat kann auf einen stationären und invertierbaren VARMA[p, q]-Prozess $\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t + \mathbf{M}_1 \mathbf{u}_{t-1} + \cdots + \mathbf{M}_q \mathbf{u}_{t-q}$ übertragen werden. Der linear transformierte Prozess $\mathbf{z}_t = \mathbf{F}\mathbf{y}_t$ besitzt eine stationäre und invertierbare VARMA[\tilde{p}, \tilde{q}]-Darstellung mit $\tilde{p} \leq (K - M + 1)p$ und $\tilde{q} \leq (K - M)p + q$. VARMA-Modelle sind also abgeschlossen gegenüber linearen Transformationen; ein linear transformierter VARMA-Prozess endlicher Ordnung ergibt erneut einen VARMA-Prozess endlicher Ordnung. Die Darstellung als VARMA[\tilde{p}, \tilde{q}]-Modell gibt obere Grenzen für die Ordnungen der VAR- und VMA-Operatoren des linear transformierten VARMA[p, q]-Prozesses an. Wird ein VARMA[$p, 0$], d. h. ein VAR[p]-Prozess $\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t$ linear transformiert zu $\mathbf{z}_t = \mathbf{F}\mathbf{y}_t$, weist \mathbf{z}_t eine VARMA[\tilde{p}, \tilde{q}]-Darstellung mit $p \leq Kp$ und $q \leq (K - 1)p + q$ auf (da $M = 1$ gilt). Es lässt sich zeigen, dass im Allgemeinen eine Darstellung mit geringeren Ordnungen nicht existiert, wenn für Transformationen obere Grenzen der Ordnungen bestehen. Lineare Transformationen von VAR[p]-Prozessen werden daher i. d. R. einer gemischten VARMA-Darstellung folgen. Da wie gesagt Transformationen von Variablen in praktischen Anwendungen üblich sind, stellt dies ein Argument zugunsten der allgemeineren VARMA-Klasse dar.

Spezifikation und Schätzung von VARMA-Modellen

Die Parameter eines VARMA-Prozesses sollten i. d. R. mit der ML-Methode geschätzt werden. Diese ergibt konsistente, asymptotisch effiziente und normalverteilte Schätzer. Im Allgemeinen führt die Maximierung der Likelihoodfunktion aber auf einen komplizierten nichtlinearen Optimierungsprozess. Falls jedoch ein reiner VAR-Prozess ohne MA-Teil vorliegt, ergeben sich wesentliche Vereinfachungen. So gibt es multivariate Verallgemeinerung der Yule-Walker-Schätzer. Auch Kleinste-Quadrat-Schätzer sind verfügbar. Asymptotisch weisen die Komponenten der Kleinste-Quadrat-Schätzer unter geeigneten Voraussetzungen die gleichen Eigenschaften auf, wie die im üblichen multivariaten Regressionsmodell. Die Voraussetzungen sind bei Normalprozessen erfüllt. Es gibt auch allgemeinere Bedingungen, unter denen dies gilt. Dazu sei auf Lütkepohl (2005) verwiesen.

Satz 8.17 (Asymptotik der KQ-Schätzer bei VAR-Prozessen)

Sei (y_t) ein stationärer k -dimensionaler $VAR[p]$ -Prozess. Der Residualprozess sei ein Gaußscher White-Noise-Prozess. Dann ist der KQ-Schätzer konsistent und asymptotisch normalverteilt.

Bei der Kleinste-Quadrat-Methode kann jede Gleichung getrennt mittels einer einfachen linearen Regression geschätzt werden. Mit dem asymptotischen Verteilungsresultat können dann die Parameter bez. ihrer Relevanz gecheckt werden. Ggf. können sie dann gleich null gesetzt und ein restringiertes Modell angepasst werden.

Parameterrestriktionen sind bei multivariaten Prozessen generell von Interesse. Ist nämlich die Zahl der Komponenten bzw. die Ordnung des VARMA-Prozesses groß, so ergibt sich das Problem, dass die Anzahl der zu schätzenden Parameter enorm hoch ist. Große Standardfehler oder, anders gesagt, unzuverlässige Schätzungen sind die Konsequenz. Kann man z. B. unterstellen, dass die Parametermatrizen symmetrisch sind, so reduziert das die Anzahl der Parameter erheblich.

Eine Strategie zur Identifikation bewegt sich entlang des klassischen Box/Jenkins-Ansatzes zur Spezifikation von univariaten ARMA-Modellen. Für eine beobachtete multivariate Zeitreihe y_1, y_2, \dots, y_N wird das zugrunde liegende Modell anhand der Struktur der Korrelationsmatrizen und der partiellen Korrelationsmatrizen identifiziert. Nichtstationäre Reihen sind dabei vorab stationär zu machen. In der Regel geschieht das mittels geeigneter Differenzierung. Dabei ist es nicht notwendig, dass die Differenzierung bei allen Komponenten in der gleichen Weise vorgenommen wird. Dementsprechend ist eine formale Angabe eines VARIMA-Modells wesentlich komplexer als die eines univariaten ARIMA-Modells. Hier sei darauf verzichtet und die weitere Darstellung auf den stationären Fall beschränkt.

Ein Wort ist zur Differenzenbildung bei multivariaten Modellen noch angebracht. Diese ist kompliziert und sollte mit Vorsicht betrieben werden. Überdifferenzierung kann nämlich zu erhöhten Schwierigkeiten bei der Modellbildung führen. Speziell kann auch bei instationären Ausgangskomponenten die Nichtidentifizierbarkeit einzelner Komponenten resultieren, vgl. Satz 3.19.

Beispiel 8.18 (Melanom-Inzidenzen in Connecticut)

Das Krebs-Register von Connecticut, USA, ist das älteste der Welt. Aus den jährlichen Inzidenzen (Anzahl aller neuen registrierten Fälle) in den Jahren 1935–1979, siehe Cusano & Young (1986), wurden unter Heranziehung der Angaben zur Bevölkerung die Raten der Neuerkrankungen an Melanomen berechnet. Von Interesse ist nun, ob es einen Zusammenhang mit der Sonnenfleckaktivität gibt. Die beiden Komponenten der Zeitreihe MELANOM, (R_t) und (S_t), sind in der Abbildung 8.9 dargestellt. Während die Raten eine wachsende Tendenz zeigen, weisen die Sonnenfleckaktivitäten ein zyklisches, aber kein trendmäßiges Verhalten auf. Man wird hier also nur bei der ersten Reihe Differenzen bilden.

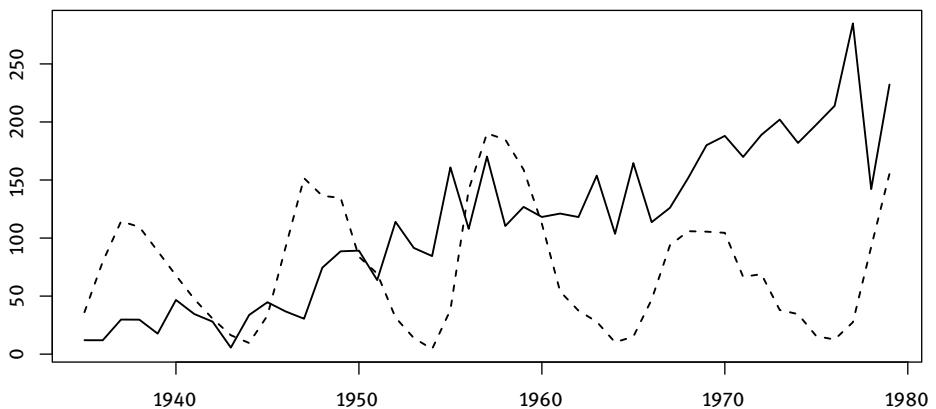


Abb. 8.9: Melanom-Inzidenzraten in Connecticut (—) und Sonnenfleckenaktivität (- - -)

Für die Elemente $\hat{\rho}_{ij}(\tau)$ der empirische Korrelationsmatrizen lassen sich analog zu dem univariaten Fall asymptotische Aussagen ableiten, speziell die Konsistenz und asymptotische Normalverteiltheit. Wenn der zugrunde liegende Prozess White-Noise ist, ergibt sich

$$\text{Cov}(\hat{\rho}_{ij}(\tau), \hat{\rho}_{ij}(\tau + \kappa)) \approx \frac{1}{N - \tau} \quad (8.17)$$

und

$$\text{Var}(\hat{\rho}_{ij}(\tau)) \approx \frac{1}{N - \tau} \quad (8.18)$$

Für genügend lange Reihen wird $1/(N - \tau)$ durch $1/N$ ersetzt.

Die empirischen Kreuzkorrelationsmatrizen können zur Identifizierung von VMA[q]-Modellen herangezogen werden, da sie für Lags $\tau > q$ theoretisch gleich $\mathbf{0}$ sind. Da die Bestimmung der Ordnung mehrere Komponenten betrifft, ist das Ausnutzen der Theorie allerdings nicht einfach. Eine der empfohlenen Methoden, dies zu vereinfachen, basiert auf der Ersetzung der aktuellen Werte in den Matrizen durch die Symbole +, - und ·. Dabei steht

- + für einen Wert oberhalb des doppelten Standardfehlers $2/\sqrt{N}$,
- für einen Wert unterhalb des doppelten Standardfehlers $-2/\sqrt{N}$,
- für einen Wert zwischen diesen beiden Grenzen.

Bei der Verwendung dieser Größen ist zu beachten, dass diese Grenzen nicht mehr korrekt sind, wenn die Zeitreihen stark autokorriert sind. In der Regel wird dann eine Signifikanz zu schnell angezeigt. Folglich kann die Verwendung der Indikatormatrizen leicht zu einer Überparametrisierung führen. Dementsprechend sollten sie mit Vorsicht betrachtet werden.

Beispiel 8.19 (Simulierter VMA[1]-Prozess)

Für den VMA[1]-Prozess

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 0.5 & 0.3 \\ -1.5 & 0.8 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad \text{mit } \boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 2.0 & 1.8 \\ 1.8 & 3.0 \end{pmatrix}$$

wurde eine Realisation der Länge $n = 100$ erzeugt. Die zugehörige Folge der Autokorrelationsmatrizen zeigt das zu erwartende Muster.

Lag τ	Kreuzkorrelationen	Indikator-Symbole
1	$\begin{pmatrix} 0.394 & 0.297 \\ -0.746 & -0.156 \end{pmatrix}$	$\begin{pmatrix} + & + \\ - & . \end{pmatrix}$
2	$\begin{pmatrix} -0.183 & 0.091 \\ 0.005 & -0.133 \end{pmatrix}$	$\begin{pmatrix} . & . \\ . & . \end{pmatrix}$
3	$\begin{pmatrix} -0.172 & -0.038 \\ 0.136 & -0.070 \end{pmatrix}$	$\begin{pmatrix} . & . \\ . & . \end{pmatrix}$
4	$\begin{pmatrix} -0.056 & -0.046 \\ 0.074 & 0.013 \end{pmatrix}$	$\begin{pmatrix} . & . \\ . & . \end{pmatrix}$
5	$\begin{pmatrix} -0.033 & 0.103 \\ 0.012 & 0.015 \end{pmatrix}$	$\begin{pmatrix} . & . \\ . & . \end{pmatrix}$

R-Code:

```

1 n<-101
2 komp <- 2
3 S <- matrix(c(2.0 , 1.8 , 1.8 , 3.0),2,2, byrow=T)
4 C <- chol(S)

```

```

5 e <- matrix(rnorm(2*(n+komp)),2,n+komp)
6 e <- C%*%e
7 A <- matrix(c(0.5,0.3,-1.5,0.8),2,2,byrow=T)
8 y <- e
9 for(t in c(2:(n+komp))){ 
10   y[,t] <- A%*%e[,t-1] + e[,t]
11 }
12 y <- t(y[, (komp+1):(n+komp)])
13 a <- acfmat(y,5)

```

Im Code wird hauptsächlich ein bivariater VMA[1]-Prozess simuliert. In den Zeilen 4 bis 6 wird erstmal ein bivariater White-Noise-Prozess mit Kovarianzmatrix S generiert. Die Cholesky-Zerlegung der Matrix S wird mit der Funktion `chol` durchgeführt. Die Autokorrelationsmatrizen ergeben sich dann über Aufruf der Funktion `acfmat` aus dem Paket `tsapp`.

Die partielle Autokorrelationen lassen sich bei univariaten Reihen mit der Levinson-Durbin-Rekursion bestimmen. Die Rekursion liefert die PACF als jeweils letzte Koeffizienten der sukzessiv angepassten AR-Modelle. Darauf wird im multivariaten Fall zurückgegriffen und die Folge der *partiellen Autokorrelationsmatrizen* eines Vektorprozesses als Folge der Koeffizientenmatrizen definiert, die sich bei der sukzessiven Anpassung von VAR-Modelle immer höherer Ordnung ergeben; dabei wird aber nicht unbedingt auf die multivariate Variante der Levinson-Durbin-Rekursion zurückgegriffen. Vielmehr werden die $\boldsymbol{\Pi}(\tau)$ meist geschätzt, indem sukzessive VAR-Modelle mit der KQ-Methode geschätzt werden. meist wird von wiederholten KQ-Schätzungen ausgegangen. Für die zugrunde liegende theoretische Matrizenfolge $\boldsymbol{\Pi}(\tau)$ gilt bei VAR[p]-Prozessen:

$$\boldsymbol{\Pi}(\tau) = \begin{cases} \mathbf{A}_p, & \tau = p, \\ \mathbf{0}, & \tau > p. \end{cases}$$

Die Schätzungen erlauben auch die Bestimmung der (approximativen) Standardfehler. Da die Schätzungen $\widehat{\mathbf{A}}_1, \dots, \widehat{\mathbf{A}}_p$ gemeinsam asymptotisch normalverteilt sind, können die partiellen Autokorrelationsmatrizen somit zur Ordnungsbestimmung von VAR-Modellen herangezogen werden. Es gelten die üblichen Grenzen von ± 2 , ob die entsprechenden standardisierten Koeffizienten als signifikant von null verschieden angesehen werden können. Dafür wird die gleiche Muster-Hervorhebung wie bei den Autokorrelationen mittels $+$, $-$ und \cdot empfohlen.

Zudem gibt es summarische Likelihood-Quotienten-Tests, die die jeweilige Nullhypothese $\boldsymbol{\Pi}(\tau) = \mathbf{0}$ zu überprüfen erlauben. Um etwa $H_0: \boldsymbol{\Pi}_\tau = \mathbf{0}$ bzw. $H_0: \mathbf{A}_\tau = \mathbf{0}$ zu testen, wird die Matrix der Residuenquadratsummen und Kreuzprodukte nach Anpassung eines VAR[τ]-Modells bestimmt:

$$\mathbf{S}(\tau) = \sum_{t=\tau+1}^N (\mathbf{y}_t - \widehat{\mathbf{A}}_1 \mathbf{y}_{t-1} - \cdots - \widehat{\mathbf{A}}_\tau \mathbf{y}_{t-\tau}) (\mathbf{y}_t - \widehat{\mathbf{A}}_1 \mathbf{y}_{t-1} - \cdots - \widehat{\mathbf{A}}_\tau \mathbf{y}_{t-\tau})' . \quad (8.19)$$

Die Teststatistik ist bei einem Vektorprozess mit k Komponenten:

$$M(\tau) = -(N - p - 1 - 0.5 - \tau k) \ln \left(\frac{\det(\mathbf{S}(\tau))}{\det(\mathbf{S}(\tau - 1))} \right). \quad (8.20)$$

Sie ist unter der Nullhypothese χ^2 -verteilt mit k^2 Freiheitsgraden. $N - p - 1$ ist die effektive Anzahl der Beobachtungen unter der Voraussetzung, dass eine Konstante im Modell enthalten ist.

Schließlich wird man bei der Bestimmung der Ordnung eines Vektorprozesses auch die Diagonalelemente der geschätzten Kovarianzmatrix des Residuenprozesses heranziehen. Solange die geschätzten Residualvarianzen bei der Erhöhung der Ordnung stark abnehmen, ist die Erhöhung der Modellordnung angeraten. Werden sie nur wenig kleiner, spricht dies gegen eine Erhöhung.

Die Bestimmung der Modellordnung eines VAR-Prozesses mit Hilfe eines der Informationskriterien hat die mögliche Korrelation der Residualprozesse zu berücksichtigen. An die Stelle der bei univariaten Reihen verwendeten Varianz tritt hier die Determinante der geschätzten Kovarianzmatrix $\hat{\Sigma}_\epsilon$ von ϵ_t . Das BIC-Kriterium lautet damit:

$$\text{BIC} = \ln(\det(\hat{\Sigma}_\epsilon)) + \frac{\ln(N) \cdot \text{Anzahl geschätzter Parameter}}{N}. \quad (8.21)$$

Die Verwendung des BIC-Kriteriums führt aber bei Subsetmodellen leicht dazu, die höheren Lags ganz zu vernachlässigen. Daher ist oft ist das AIC-Kriterium mit seiner Tendenz, höhere Ordnungen zu wählen, vorzuziehen. Dann können Koeffizienten niedrigerer Lags null gesetzt werden. Das zu wählende Modell hängt jedoch letztlich von der Zielsetzung ab. Im Box-Jenkins-Ansatz sucht man sparsame Modelle. So weit wie möglich werden nicht-signifikante Parameter null gesetzt. Das Ziel ist hier eine hohe Prognosequalität des Modells. Von einigen Ökonometrikern wird dagegen die die Auffassung vertreten, dass das Ziel darin besteht, relevante ökonometrische Relationen zwischen den Variablen aufzufinden. Daher werden überparametrisierte Modelle für besser gehalten als solche, bei denen nicht-signifikante Parameter auf null gesetzt werden.

Schließlich wendet die abschließende Überprüfung der Modellanpassung die gleichen Kenngrößen wie zur Modellbestimmung auf die Residualreihe an. Weder die Autokorrelationsmatrizen noch die partiellen Autokorrelationsmatrizen sollten noch signifikant Einträge aufweisen noch sollten die summarischen Likelihood-Quotienten-Tests Signifikanz signalisieren.

Beispiel 8.20 (Luchse und Schneehasen – Fortsetzung von Seite 199)

Die in der Abbildung 8.7 dargestellte Reihe ergab sich durch Simulation eines VAR[1]-Prozesses. Es soll untersucht werden, was die angegebenen Möglichkeiten zu Bestimmung eines VAR-Modells ergeben. Man sieht anhand der Tabelle, dass genau das herauskommt, was hineingesteckt wurde. Die Vorzeichen der Teststatistiken für die Koeffizienten des VAR[1]-Modells stimmen mit denen der verwendeten Koeffizienten überein. Der Test auf die Nullhypothese $H_0: \mathbf{A}_\tau = \mathbf{0}$ ist nur für

das erste Lag signifikant. Das BIC-Kriterium (wie das AIC-Kriterium auch) wählt ebenfalls die verwendete Ordnung. Die Varianzen der beiden Residualreihen bleiben recht stabil.

Lag τ	Indikator-Symbole	$M(\tau)$	P-Wert	BIC(n)	$\text{diag}(\widehat{\Sigma}_{\varepsilon\varepsilon})$
1	$\begin{pmatrix} + & + \\ - & + \end{pmatrix}$	580.67	0	1.01	3.646 1.284
2	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	2.96	0.565	1.08	3.587 1.290
3	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	3.03	0.553	1.14	3.594 1.287
4	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	7.41	0.116	1.21	3.491 1.206
5	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	1.23	0.873	1.24	3.500 1.211

R-Code:

```

1 n <- 300
2 S <- matrix(c(2.0 , 1.8 , 1.8 , 3.0),2,2, byrow=T)
3 C <- chol(S)
4 A <- matrix(c(0.5 , 0.3 , -1.5 , 0.8),2,2, byrow=T)
5 e <- matrix(rnorm(2*(n+200)),2,n+200)
6 y <- e
7 for (t in c(2:(n+200))){ y[,t] <- A%*%y[,t-1] + C%*%e[,t] }
8 y <- t(y[,201:(n+200)])
9 out <- pacfmat(y,5)
10 VARselect(y,lag.max = 5, type = "none")

```

Wie bei dem MA-Beispiel beinhaltet der Code hauptsächlich die Simulation eines VAR-Prozesses. Mit der Funktion `pacfmat`, die sich in dem Paket `tsapp` befindet, bekommt man die relevanten Größen zur partiellen Autokorrelation. Diese sind Bestandteile einer Liste und entsprechend zu extrahieren. Die Ergebnisse sind in der obenstehenden Tabelle angegeben. Die Werte der Informationskriterien können mit der Funktion `VARselect` aus dem Paket `vars` berechnet werden. Es wird hier nicht mit dem `library`-Befehl geladen, da dies schon durch das Laden des Paketes `tsapp` erfolgte. In der Tabelle wurde exemplarisch das Bayessche Informationskriterium (BIC) von Schwarz ausgegeben (im Output von `VARselect` als SC bezeichnet).

Beispiel 8.21 (Papierproduktion)

Bei einer papierproduzierenden Maschine wurden während laufender Produktion zwei Größen in gleichen Zeitabständen gemessen. Die erste war die Höhe einer

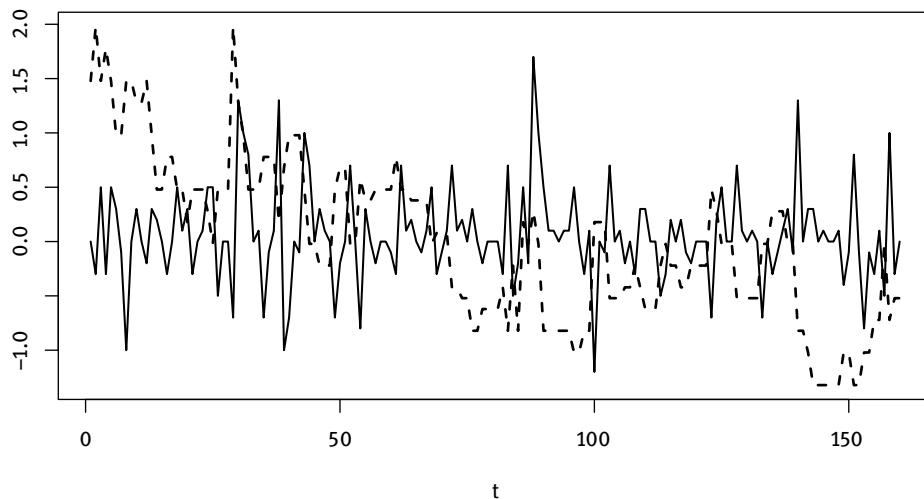


Abb. 8.10: Höhe (—) und zentrierte Gewichtsabweichung (---) bei einer Papierproduktion

Öffnung an der Maschine und die zweite die Abweichung des Papierge wichtes von einem Zielwert. Die Höhe kann als bei null zentriert betrachtet werden. Die Zeitreihe PAPIER ist Janacek & Swift (1993) entnommen.

Dieser bivariate Zeitreihe soll ein VAR-Modell angepasst werden. Die relevanten Kenngrößen für die ersten fünf Lags sind in der folgenden Tabelle angegeben. Die partiellen Korrelationsmatrizen legen einen VAR-Prozess mit der maximalen Ordnung drei nahe. Die Tests wie auch das AIC-Kriterium weisen eindeutiger auf ein VAR[3]-Modell hin. Auch die Varianzen der Residualreihen kann man dahingehend interpretieren, dass sie bis zur Ordnung drei deutlicher zurückgehen; danach nur noch geringfügig. Die Anpassung eines VAR[3]-Modells ergibt keine Abhängigkeiten bei den Residuen mehr.

Lag τ	Indikator-Symbole	$M(\tau)$	P-Wert	AIC(n)	$\text{diag}(\hat{\Sigma}_\varepsilon)$
1	$\begin{pmatrix} \cdot & + \\ \cdot & + \end{pmatrix}$	400.868	0.000	-4.719	0.183 0.103
2	$\begin{pmatrix} \cdot & \cdot \\ \cdot & + \end{pmatrix}$	20.400	0.000	-4.754	0.179 0.093
3	$\begin{pmatrix} \cdot & - \\ \cdot & \cdot \end{pmatrix}$	18.900	0.001	-4.827	0.165 0.092
4	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	6.874	0.143	-4.797	0.161 0.091
5	$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$	4.530	0.339	-4.768	0.160 0.091

Die summarischen Schätzergebnisse sind:

```
Estimation results for equation height:
=====
height = height.l1 + weight.l1 + height.l2 + weight.l2 +
         height.l3 + weight.l3
      Estimate Std. Error t value Pr(>|t|)
height.l1  0.25453   0.11555   2.203  0.029122
weight.l1  0.53573   0.16164   3.314  0.001150
height.l2  0.34562   0.11768   2.937  0.003835
weight.l2  0.16616   0.19760   0.841  0.401737
height.l3 -0.04734   0.07726  -0.613  0.540991
weight.l3 -0.59515   0.16142  -3.687  0.000316

Estimation results for equation weight:
=====
weight = height.l1 + weight.l1 + height.l2 + weight.l2 +
          height.l3 + weight.l3
      Estimate Std. Error t value Pr(>|t|)
height.l1 -0.24060   0.08571  -2.807  0.00566
weight.l1  0.50909   0.11990   4.246  3.79e-05
height.l2 -0.14468   0.08730  -1.657  0.09952
weight.l2  0.26544   0.14658   1.811  0.07216
height.l3  0.05247   0.05732   0.915  0.36144
weight.l3  0.12847   0.11974   1.073  0.28502
```

Das Weglassen der nicht signifikanten Koeffizienten ergibt das geschätzte Modell für die zentrierten Reihen:

$$\begin{pmatrix} H_t \\ W_t \end{pmatrix} = \begin{pmatrix} 0.312 & 0.625 \\ -0.242 & 0.523 \end{pmatrix} \begin{pmatrix} H_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} 0.298 & 0 \\ -0.071 & 0.372 \end{pmatrix} \begin{pmatrix} H_{t-2} \\ W_{t-2} \end{pmatrix} + \begin{pmatrix} 0 & -0.517 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} H_{t-3} \\ W_{t-3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad \text{mit } \hat{\Sigma}_{\varepsilon\varepsilon} = \begin{pmatrix} 0.172 & -0.097 \\ -0.097 & 0.096 \end{pmatrix}.$$

Bei diesem Modell sind die Koeffizienten alle signifikant. Die Residuen weisen keine Auto- oder Kreuzkorrelationen mehr auf, die summarischen Statistiken zeigen ebenfalls, dass das Modell alle Abhängigkeiten erfasst. Die starke negative Korrelation der beiden Residualprozesse von -0.75 weist darauf hin, dass die Höhe der Öffnung an der Maschine und die Abweichung des Papiergegewichtes wesentlich zusammenhängen.

R-Code:

```

1 data(PAPER)
2 y <- scale(PAPER,center=c(0,mean(PAPER[,2])),scale=FALSE)
3 ts.plot(y,lty=c(1,2),lwd=c(1.5,2),xlab="",ylab="")
4 out <- pacfmat(y,10)
5 out$pval
6 print(out$M1,quote=F)
7 VARselect(y,lag.max = 10, type = "none")
8 out <- VAR(y,p=3,type = "none")
9 summary(out)
10 res <- resid(out)
11 out1 <- pacfmat(res,10)
12 restriction <- matrix(c(1, 1, 1, 0, 0, 1,
13                         1, 1, 1, 1, 0, 0),
14                         nrow=2, ncol=6, byrow=TRUE)
15 out1 <- restrict(out,method="man",resmat=restriction)
16 out2 <- pacfmat(resid(out1),10)
17 out2$pval
18 print(out2$M1,quote=F)

```

Die Daten werden eingelesen und geplottet (Zeilen 1–3). `plot` würde die Zeitreihen getrennt darstellen, wie bei einer Matrix üblich. In der Zeile 4 werden die mit den PACF-Matrizen zusammenhängenden Größen berechnet. Die *P*-Werte der summarischen Tests (Zeile 5) und die Indikatormatrizen (Zeile 6) erlauben die Wahl der Ordnung des VAR-Modells. Die zur VAR-Modellierung benötigten Funktionen sind Bestandteil des Paketes `vars`, das mit `tsapp` geladen wurde. Die Schätzung des restringierten Modells in Zeile 15 geschieht mit der Erklärung einer Restriktionsmatrix, aus `0en` (wird nicht geschätzt) und `1en` (wird geschätzt).

VARMA-Modelle sind in ihrer Standard-Darstellung meist nicht identifizierbar. Das heißt, es gibt unterschiedliche Modelle mit denselben Momenten erster und zweiter Ordnung. Damit lassen sich die Parameter nicht mehr eindeutig aus den Beobachtungen schätzen. Voraussetzung für eine konsistente Schätzung der Parameter ist also eine identifizierbare Darstellung. Es gibt im Wesentlichen drei Ansätze zur Gewinnung von solchen Darstellungen. Einmal wurde die finale Gleichungs-Form als identifizierbare VARMA-Darstellung vorgeschlagen. Dann werden die Modelle in der Echelon-Form über die Bestimmung von Kronecker-Indizes spezifiziert. Als dritter Ansatz ist die Skalarkomponenten-Methode zu nennen. Schon die Beschreibung dieser Ansätze erfordert einen recht hohen formalen Aufwand. Daher wird hier darauf verzichtet und auf Lütkepohl (2005) und Reinsel (1993) verwiesen.

Granger-Kausalität

Bei der Betrachtung bivariater Zeitreihen möchte man oft wissen, ob eine Komponente, etwa (Y_t) , von der anderen, etwa (X_t) , abhängt. Diese Frage hat zu der Formulierung der *Granger-Kausalität* geführt. Grangers Überlegung war, dass ein Ereignis, das kausal für ein anderes ist, diesem vorausgeht. Damit kann es u. U. zur Verbesserung der Prognose beitragen. Im Zeitreihenkontext bedeutet das, dass X_t für die Prognose von Y_t bedeutsam ist. Der Begriff „Kausalität“ hat bei dieser Form der Operationalisierung kaum mehr etwas mit Ursache und Wirkung zu tun. Er ist aber fest eingeführt, so dass er auch hier verwendet wird.

Im Rahmen bivariater VAR[p]-Prozesse

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} + \begin{pmatrix} a_{(11)1} & a_{(12)1} \\ a_{(21)1} & a_{(22)1} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} a_{(11)2} & a_{(12)2} \\ a_{(21)2} & a_{(22)2} \end{pmatrix} \begin{pmatrix} X_{t-2} \\ Y_{t-2} \end{pmatrix} + \dots + \begin{pmatrix} a_{(11)p} & a_{(12)p} \\ a_{(21)p} & a_{(22)p} \end{pmatrix} \begin{pmatrix} X_{t-p} \\ Y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \quad (8.22)$$

ist (X_t) nicht Granger-kausal für (Y_t) , wenn alle Koeffizienten $a_{(21)1}, \dots, a_{(21)p}$ null sind. Andernfalls liegt Granger-Kausalität vor. Analog ist (Y_t) Granger-kausal für (X_t) , wenn nicht alle $a_{(12)u}$ identisch null sind. Granger-Kausalität kann also in beiden Richtungen vorliegen. Zwischen den beiden Variablen gibt es dann eine *Feedback-Beziehung*. Eine umfassende Diskussion des von Granger (1969) eingeführten Kausalitätsbegriffes ist bei Kirchgässner (1981) zu finden.

Ein ökonometrischer Ansatz, um zu testen, ob (X_t) Granger-kausal für (Y_t) ist, basiert auf der Kleinsten-Quadrat-Schätzung von

$$Y_t = \mu_1 + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \beta_1 X_{t-1} + \dots + \beta_p X_{t-p} + \varepsilon_t \quad (8.23)$$

mit dem anschließenden Testen der Hypothese

$$H_0: \beta_1 = \dots = \beta_p = 0.$$

Die zugehörige Prüfgröße des F -Tests ist

$$F = \frac{(SS_{\text{Fehler}0} - SS_{\text{Fehler}1})/p}{SS_{\text{Fehler}1}/(N - 2p - 1)}.$$

Dabei ergibt sich die Residuen-Quadratsumme $SS_{\text{Fehler}1}$ aus (8.23) und die Residuen-Quadratsumme $SS_{\text{Fehler}0}$ aus der entsprechenden Beziehung, bei der dann die X_{t-u} nicht auftauchen. Unter H_0 ist die Teststatistik asymptotisch $\mathcal{F}_{p, N-2p-1}$ -verteilt.

Die Ergebnisse können sehr sensitiv bzgl. der unterstellten Ordnung p sein.

Bei dem folgenden Verfahren, das von Hsiao (1979, 1981) vorgeschlagen wurde, wird die Verbesserung der Prognosegüte durch die Berücksichtigung einer zusätzlichen Variablen anhand von Modellselektionskriterien gemessen. Der Ansatz ist zugleich ein praktikabler Weg zur Identifizierung von VAR-Modellen. Das Verfahren läuft in folgenden Schritten ab:

1. Das VAR-Modell mit der Ordnung p ,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} + \sum_{u=1}^p \begin{pmatrix} a_{(11)u} & a_{(12)u} \\ a_{(21)u} & a_{(22)u} \end{pmatrix} \begin{pmatrix} X_{t-u} \\ Y_{t-u} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix},$$

wird in die beiden Teilbeziehungen zerlegt:

$$\begin{aligned} X_t &= \mu_X + \sum_{u=1}^{p_{11}} a_{(11)u} X_{t-u} + \sum_{u=1}^{p_{12}} a_{(12)u} Y_{t-u} + \varepsilon_{1t} \\ Y_t &= \mu_Y + \sum_{u=1}^{p_{21}} a_{(21)u} X_{t-u} + \sum_{u=1}^{p_{22}} a_{(22)u} Y_{t-u} + \varepsilon_{2t}. \end{aligned}$$

Für alle Ordnungen $0 \leq p_{11}, p_{12}, p_{21}, p_{22} \leq p$ werden die beiden Modelle mittels der CLS-Methode geschätzt und der Wert eines der Informationskriterien bestimmt. (Dabei ist die Kreuzkovarianz der Residualreihen zu berücksichtigen!)

2. Aus der Gesamtheit der Informationskriterien werden drei Mengen gebildet und das zugehörige Minimum bestimmt:

Als erstes sei m_1 der minimale Wert des Kriteriums für alle betrachteten Lagstrukturen.

m_2 sei der minimale Wert des Kriteriums für die Lagstrukturen, bei denen die Restriktion $a_{(12)u} = 0$, $u = 1, \dots, p$ erfüllt ist.

Der minimale Wert der Menge, bei der die Lagstrukturen $a_{(21)u} = 0$, $u = 1, \dots, p$, erfüllen, sei m_3 .

3. Die ermittelten Prognosegüten der drei Modellklassen werden verglichen:

- Falls $m_2 > m_1$, ist Y Granger-kausal zu X .
- Falls $m_3 > m_1$, ist X Granger-kausal zu Y .

Beispiel 8.22 (Anbindung der Währung Österreichs an die DM)

Seit Anfang der achtziger Jahre betrieb Österreich eine enge Anbindung seiner Währung an die DM, um bei der Inflationsbekämpfung von der Reputation der Deutschen Bundesbank zu profitieren. Als Ergebnis dieser Hartwährungspolitik wird in der ökonomischen Literatur angenommen, dass die österreichische Wirtschaft voll in die erweiterte DM-Zone integriert sei. Jaenicke & Neck (1996) konfrontierten diese Annahme mit konkreten Daten. Insbesondere untersuchten sie in ihrer Arbeit drei Teihypothesen, um die „Deutsche Dominanz“-Hypothese zu überprüfen. Hier sei nur ihre erste Hypothese betrachtet:

Der deutsche Zinssatz ist Granger-kausal für den österreichischen Zinssatz (Abhängigkeit von Deutschland).

Die zur Überprüfung verwendeten Reihen sind die Umlaufrenditen öffentlicher Anleihen in Österreich (Y_t) und in Deutschland (X_t) für die Zeit von Januar 1981 bis November 1994; sie bilden die bivariate Reihe DOMINANZ. Da die Reihen vom Typ I(1) sind, gehen die Autoren von bivariaten VAR-Modellen für die ersten Differenzen X_t^* und Y_t^* aus.

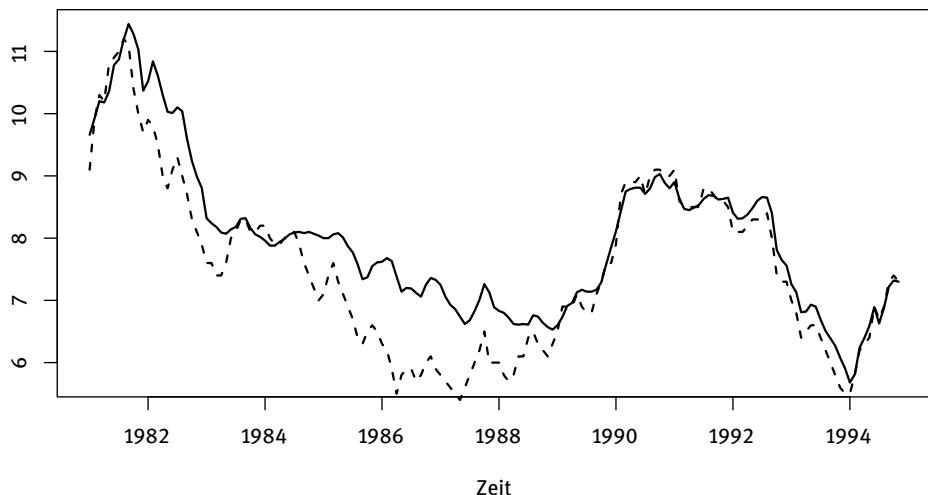


Abb. 8.11: Langfristige Zinssätze, $Y_t = -$, $X_t = \dots$

Die Nullhypothese, dass X_t^* nicht Granger-kausal für Y_t^* ist, wird bei Zulassen von vier Lags abgelehnt; der P -Wert des F -Tests ist kleiner als 0.001. (Andersherum wird die Nullhypothese, dass Y_t^* keinen kausalen Einfluss auf die Reihe X_t^* hat, nicht verworfen. Der P -Wert beträgt 0.262.) Dieses Ergebnis spricht also für die Dominanzhypothese.

Das Verfahren nach Hsiao ergibt andererseits bei maximal zugelassener Ordnung $p = 4$ unter Verwendung des BIC das unrestringierte Modell

$$\begin{aligned} X_t^* &= -0.0179 + 0.5236X_{t-1}^* - 0.2997Y_{t-1}^* + \varepsilon_{1t} \\ Y_t^* &= -0.011 + 0.474X_{t-1}^* + \varepsilon_{2t}. \end{aligned}$$

Zudem liegt zwischen den ersten Differenzen von Y_t und X_t eine Feedback-Beziehung vor: $m_1 = -7.55$, $m_2 = -7.50$, $m_3 = -7.35$. Wegen der Feedback-Beziehung kann mit diesem Ansatz die Dominanzhypothese nicht bestätigt werden.

R-Code:

```

1 data(DOMINANCE)
2 dzins <- diff(DOMINANCE)
3 est <- VAR(dzins,p=4)
4 causality(est,cause="X")
5 causality(est,cause="Y")
6 out <- Grangercaus(dzins[,1],dzins[,2],4)

```

In dem Paket `vars` gibt es die Funktion `causality`, die zwei F -Tests durchführt, einen, ob vergangene Werte der wirkenden Variablen signifikante Einflüsse haben und dann, ob es eine instantane kausale Relation gibt. Die Funktion `Grangercaus` aus `tsapp` setzt das Verfahren von Hsiao um.

Kointegration

In ökonomischen Theorien wird oft nahe gelegt, dass gewisse Paare von ökonomischen Variablen auf lange Sicht ein Gleichgewicht in dem Sinne aufweisen, dass ihr Verhältnis zueinander stabil bleibt. Auch wenn die Variablen bisweilen von diesem Gleichgewicht abweichen, sollten sie aufgrund der ökonomischen Kräfte doch zu dem Gleichgewicht zurückfinden. Beispiele solcher Variablen sind etwa das verfügbare Einkommen und der Konsum oder die Geldmenge und das Preisniveau. Natürlich gibt es keinen Grund, sich auf zwei Variablen zu beschränken; dies geschieht hier wie eingangs gesagt nur der Vereinfachung wegen.

Die von den ökonomischen Variablen erzeugten Zeitreihen sind oft vom Typ I(1), oder scheinen es zumindest zu sein, wenn Einheitswurzeltests angewendet werden. Solche Prozesse divergieren für $N \rightarrow \infty$, da ihre Varianz proportional zu N ist. Daher mag es unplausibel erscheinen, dass solche Variablen eine Art langfristige gleichgewichtige Relation aufweisen. Es ist jedoch möglich, dass beide Variablen I(1) sind, eine geeignete Linearkombination dagegen vom Typ I(0). In diesem Fall heißen die beiden Variablen *kointegriert*. Sind zwei oder mehr Variablen kointegriert, so gibt es ein langfristiges Gleichgewicht, auch wenn sie bei kürzerem Zeithorizont davon wesentlich abweichen können.

Definition 8.23 (Kointegrierte Prozesse)

Seien $(Y_{1t}), (Y_{2t})$ zwei I(1)-Prozesse. Sie sind kointegriert, falls es einen Vektor $\eta = (1, -\eta_2)'$ gibt, so dass

$$Z_t = (1, -\eta_2) \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix}$$

stationär ist. Der Vektor η wird als *Kointegrationsvektor* bezeichnet.

Der Kointegrationsvektor ist zunächst nicht eindeutig, da er z. B. mit einer Konstanten multipliziert werden kann. Daher ist die allgemeine, auch in der Definition wiederzufindende Vereinbarung, dass der erste Koeffizient eins gesetzt wird.

Beispiel 8.24 (Ein kointegrierter zweidimensionaler Prozess)

Sei (u_t) ein zweidimensionaler White-Noise-Prozess mit

$$\mathbf{u}_t = \begin{pmatrix} U_t \\ V_t \end{pmatrix} = \begin{pmatrix} b\varepsilon_t \\ c\varepsilon_t \end{pmatrix},$$

wobei (ε_t) mit $E(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$ ein univariater White-Noise-Prozess ist. Dann hat (u_t) den Erwartungswertvektor $\mathbf{0}$ und es ist

$$\boldsymbol{\Gamma}(0) = \boldsymbol{\Sigma}_{UV} = \sigma^2 \begin{pmatrix} b^2 & bc \\ bc & c^2 \end{pmatrix}.$$

Weiter sei

$$\mathbf{y}_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} U_1 \\ V_1 \end{pmatrix} + \begin{pmatrix} U_2 \\ V_2 \end{pmatrix} + \dots + \begin{pmatrix} U_t \\ V_t \end{pmatrix}.$$

Zwischen den Variablen des Prozesses existieren exakte lineare Beziehungen, so dass $a_1 Y_{1t} + a_2 Y_{2t}$ wieder stationär ist, genauer ist $Y_{1t} - \frac{b}{c} Y_{2t} = 0$.

Beispiel 8.25 (Zwei nicht kointegrierte Prozesse)

Seien (Y_t) und (Z_t) zwei von unabhängigen White-Noise-Prozessen überlagerte Random-Walks:

$$Y_t = \sum_{u=0}^t \varepsilon_{(1)u} + \eta_t, \quad Z_t = \sum_{u=0}^t \varepsilon_{(2)u} + \zeta_t,$$

wobei $(\varepsilon_{(1)t})$, $(\varepsilon_{(2)t})$, (η_t) und (ζ_t) unabhängige Gaußsche White-Noise-Prozesse mit Erwartungswerten null und Varianzen $\sigma_{\varepsilon_{(1)t}}^2 = 1$, $\sigma_{\eta_t}^2 = 0.5$ und $\sigma_{\zeta_t}^2 = 0.0625$ sind.

In der obersten Grafik der Abbildung 8.12 sind die beiden Zeitreihen (y_t) und (z_t) dargestellt. Z_t wird dann gegen Y_t regressiert; dies ist die mittlere Grafik, in der natürlich die zeitliche Anordnung verloren gegangen ist. In der untersten werden dann die Residuen wieder in der zeitlichen Anordnung präsentiert. Wie sich aus dem Verhalten der Residuen der Kleinste-Quadrat-Regression von (Z_t) auf (Y_t) erkennen lässt, sind (Y_t) und (Z_t) nicht kointegriert. Sie wandern genauso instationär wie die Ausgangsreihen.

R-Code:

```

1 y <- cumsum(rnorm(100)) + rnorm(100,sd=sqrt(0.5))
2 z <- cumsum(rnorm(100)) + rnorm(100,sd=sqrt(0.0625))
3 out <- lm(z~y)
4 plot(y,ylim=c(min(y,z),max(y,z)),type="l")
5 lines(z,lwd=1.75,lty=2)
6 plot(y,z ,xlab="x",ylab="y")
7 abline(out)
8 plot(out$residuals,type="l")

```

In Zeile 1 und 2 des R-Codes werden zwei unabhängige Random-Walks simuliert, die von White-Noise-Prozessen überlagert sind. In Zeile 3 wird die Regression durchgeführt. Es folgen Befehle zur Erstellung der Teilgrafiken.

Das im Beispiel 8.25 praktizierte Vorgehen stellt ein heuristisches und praktisch leicht umsetzbares Vorgehen dar, um abzuklären, ob zwei Zeitreihen als Realisationen kointegrierter Prozesse anzusehen sind. Genauer wird zuerst untersucht, ob beide Reihen

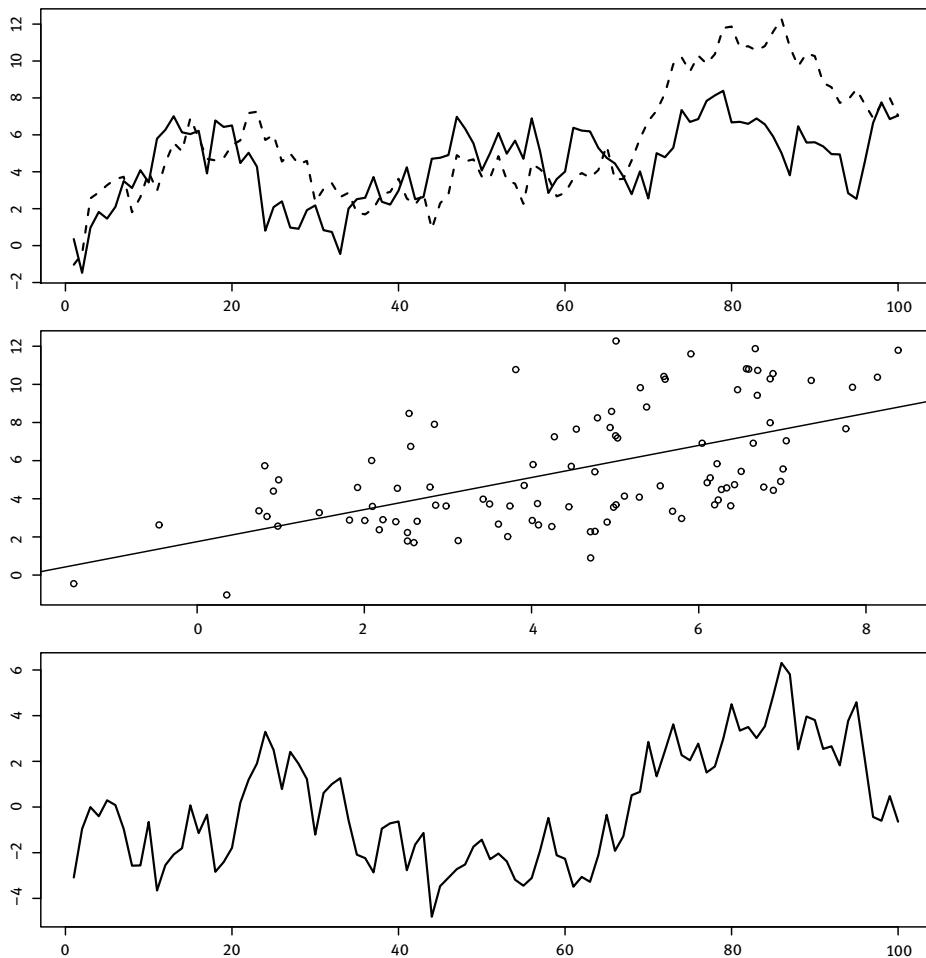


Abb. 8.12: Nicht-kointegrierte Prozesse; siehe Text

vom Typ I(1) sind. Ist dies der Fall, so wird eine Kleinsten-Quadrat-Regression der Form

$$\sum_{t=1}^N (z_t - \beta_0 - \beta_1 y_t)^2 \stackrel{!}{=} \min$$

angesetzt. Das Verhalten der Residuen gibt Aufschluss über die Kointegration.

In der Ökonometrie werden häufig formale Test auf das Vorliegen von Kointegrationsbeziehungen durchgeführt. Um die Diskussion einfach zu halten, wird zunächst unterstellt, dass der Kointegrationsvektor bekannt sei. Dann besteht die formale Variante des eben angesprochenen heuristischen Vorgehens zur Überprüfung der Koin-

tegration aus den beiden folgenden Schritten:

1. Es wird die Integrationsordnung der zu untersuchenden Zeitreihen bestimmt. Dies geschieht unter Verwendung eines Einheitswurzeltests wie z. B. des DF-Tests oder ADF-Tests. Haben sie unterschiedliche Integrationsordnungen oder ist gar einer vom Typ I(0), so können sie nicht kointegriert sein.
2. Falls beide Zeitreihen (Y_t) und (Z_t) dieselbe Integrationsordnung haben, lässt sich aus der Integrationsordnung der Linearkombination $U_t = \eta_1 Y_t + \eta_2 Z_t$ auf bestehende oder nicht bestehende Kointegration schließen. Daher wird dies mit einem Einheitswurzeltest überprüft. Wird die Nullhypothese, dass (U_t) eine Einheitswurzel hat, abgelehnt, so sind die beiden Zeitreihen kointegriert.

Der Kointegrationsvektor wird in der Regel nicht bekannt sein. Die *Engle-Granger-Methode zum Testen auf Kointegration* schiebt daher zwischen den beiden Schritten einen weiteren ein, in dem eine KQ-Schätzung der potentiellen Langfristbeziehung vorgenommen wird. Dann werden nachfolgend die Residuen der Regression auf eine Einheitswurzel überprüft. Falls die Nullhypothese, die Residuen sind vom Typ I(k) mit $k \geq 1$ sind, abgelehnt wird, besteht zwischen (Y_t) und (Z_t) eine Kointegrationsbeziehung. In der ökonometrischen Interpretation besteht ein langfristiges ökonomisches Gleichgewicht.

Da der Integrationsvektor im neuen zweiten Schritt geschätzt wird und statt der einfachen Linearkombination der Zeitreihenvariablen die Residuen der Regression im (neuen) dritten Schritt verwendet, gelten die üblichen kritischen Werte des (erweiterten) Dickey-Fuller-Tests nicht mehr. Kritische Werte sind aber von MacKinnon (1991) ermittelt worden.

Beispiel 8.26 (Anbindung der Währung Österreichs an die DM – Fortsetzung von Seite 215)

Es soll beispielhaft untersucht werden, ob die beiden Zinsreihen sich im langjährigen Gleichgewicht befinden. Zunächst werden nun beide auf eine Einheitswurzel getestet. Die Mittelwerte der differenzengebildeten Reihen sind nahe null (-0.011 und -0.014). Zudem zeigen beide ein stationäres Verhalten. Daher wird der Dickey-Fuller-Test beide Male in der Version „ohne Trend und ohne Drift“ durchgeführt. Die Darstellungen der ACF und PACF der differenzengebildeten Reihen weisen auf eine autoregressive Abhängigkeit erster Ordnung hin. Entsprechend wird der augmentierte DF-Test eingesetzt. Die Nullhypothese wird beide Male nicht abgelehnt:

```
Test von x:  Value of test-statistic is: -1.1085
Test von y:  Value of test-statistic is: -1.0493
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Die Regression der y_t gegen die x_t ergibt die Residuen e_t , die weiter zu betrachten sind. Wie die Kennfunktionen zeigen, ist von einer Autokorrelation erster Ordnung auszugehen. Auf die Residualreihe wird also der erweiterte Dickey-Fuller-Test angewendet. Der Wert der Teststatistik ist -2.2345 . Der kritische Wert zum Niveau 0.05 ergibt sich nach MacKinnon zu -1.942749 . Damit wird die Nullhypothese einer Einheitswurzel abgelehnt; die beiden Zeitreihen sind kointegriert.

R-Code:

```

1 data(DOMINANCE)
2 X <- DOMINANCE[,1]
3 Y <- DOMINANCE[,2]
4 library(urca)
5 out <- ur.df(X,type="none",lags=1,selectlags="Fixed")
6 summary(out)
7 out <- ur.df(Y,type="none",lags=1,selectlags="Fixed")
8 summary(out)
9 out <- lm(Y~X)
10 out <- ur.df(out$residuals,type="none",lags=1,selectlags="Fixed")
11 summary(out)
12 qunitroot(0.05,N=length(Y),trend="nc")

```

Die Anwendung des DF-Tests wird bereits im Beispiel 4.1 illustriert. Die AR-Ordnungen $lags=1$ für den DF-Test ergeben sich wie im Text angegeben. In der Zeile 12 wird das 0.05-Quantil der Teststatistik nach MacKinnon bestimmt.

Das Testen von Kointegrationsbeziehungen ist vor dem Hintergrund der ökonometrischen Fragestellung zu sehen. Von der statistischen Seite her stellt sich das Problem der Güte der Tests genauso wie bei den Unit Root Tests; es sind ja auch derartige Tests, die hier zum Einsatz kommen.

Kointegration hängt zusammen mit *Fehlerkorrekturmodellen*. Mit solchen Modellen wird beschrieben, wie die kurzfristige Dynamik eines Systems durch Abweichungen vom asymptotischen Gleichgewicht beeinflusst wird. Das sei anhand eines Beispiels dargestellt.

Beispiel 8.27 (Dynamische Konsumfunktion)

Das einfache Modell einer dynamischen Konsumfunktion mit dem gesamtwirtschaftlichen Konsum C_t und dem verfügbaren Einkommen E_t sei (nach Schröer, 1997):

$$C_t = \alpha C_{t-1} + \beta E_{t-1} + \varepsilon_{1t},$$

$$E_t = c + \zeta E_{t-1} + \varepsilon_{2t}.$$

Für $|\alpha|, |\zeta| < 1$ ist das Modell stationär. Fallen alle zukünftigen Innovationen weg, $\varepsilon_{1,t+\tau} = \varepsilon_{2,t+\tau} = 0$, $\tau > 0$, so konvergieren die Zeitpfade gegen die Gleichgewichtswerte

$$\mu_C = \frac{\beta c}{(1 - \alpha)(1 - \zeta)}, \quad \mu_E = \frac{c}{1 - \zeta}.$$

Damit hat man auch die Relation $\mu_C - \phi \cdot \mu_E = 0$, wobei der Faktor gegeben ist durch $\phi = \beta/(1 - \alpha)$.

Umformen der ersten Zeile in der Weise, dass auf der linken Seite die ersten Differenzen von C_t stehen, ergibt:

$$(1 - B)C_t = (\alpha - 1)[C_{t-1} - \phi E_{t-1}] + \varepsilon_{1t}.$$

$[C_{t-1} - \phi E_{t-1}]$ ist gerade die Abweichung vom Gleichgewicht $\mu_C - \phi \cdot \mu_E = 0$ zwischen C und E in der Periode $t - 1$. Eine positive Differenz bedeutet, dass in dieser Periode zu viel konsumiert wurde, was wegen $\alpha - 1 < 0$ eine negative Veränderung auf $(1 - B)C_t$ hervorruft.

Sei nun \mathbf{y}_t ein zweidimensionaler VAR[p]-Prozess. Er lässt sich schreiben als

$$(1 - B)\mathbf{y}_t = \sum_{u=1}^{p-1} \mathbf{C}_u (1 - B)\mathbf{y}_{t-u} - \boldsymbol{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \quad (8.24)$$

Die Matrix $\boldsymbol{\Pi}$ spielt dabei die herausragende Rolle in der Kointegrationstheorie. Sie ist bei einem Prozess mit kointegrierten Komponenten singulär und lässt sich als Produkt einer Matrix mit einem Kointegrationsvektor darstellen.

Speziell zeigt die als *Vektor-Fehlerkorrekturmodell-Darstellung* bezeichnete Gleichung (8.24), dass die Anpassung eines VAR-Modells an die Differenzen fehlspezifiziert ist.

8.3 R-Funktionen

Die Kreuzkorrelationsfunktion zweier Zeitreihen wird mit `ccf` bestimmt. Für die Kohärenz und Phase ist die Funktion `bispecies` aus `tsapp` geschrieben.

Multivariate Gaußsche White-Noise-Prozesse können mit der Funktion `rmvnorm` des Paketes `mvtnorm` simuliert werden. Argumente sind die Reihenlänge, der unterstellte Erwartungswertvektor und die Kovarianzmatrix.

VAR-Modelle lassen sich mit der Funktion `ar` schätzen. Das Paket `vars` enthält weitergehende Funktionen zum Arbeiten mit VAR-Modellen. Das Paket `mAr` bietet mit der Funktion `mAr.est` eine schrittweise Kleinstes-Quadrate-Schätzung eines VAR[p]-Modells. Die Methode basiert auf einem Algorithmus von Neumaier & Schneider (2001). Eine optimierte Version dieses Algorithmus steht im Paket `BigVAR`, Funktion `VARXFit` zur Verfügung.

Ein eigenes, umfangreiches Paket zur Analyse von multivariaten Zeitreihenmodellen ist MTS. Es bietet Funktionalitäten für die Identifikation, Schätzung und Prognose einer Vielzahl von Modellen, z. B. VAR-, VMA-, VARMA- und saisonale VARMA-Modelle. VAR-Modelle mit exogenen Regressoren, so genannte VARX-Modelle, werden ebenfalls berücksichtigt. Die Funktionen VARpred, VARMApred, SVARMApred und VARXpred ermitteln Prognosen von angepassten Objekten und bestimmen den mittleren absoluten Prognosefehler. VARMAirf und VARXirf berechnen und plotten die Impulse-Response-Funktion (IRF) für den entsprechenden Prozess. Die Funktion GrangerTest führt einen Granger-Kausalitätstest durch.

Die Funktion Grangercaus dem Paket tsapp setzt den Vorschlag von Hsiao (1979, 1981) um. Es werden die drei BIC-Werte für die unterschiedlichen Lag-Strukturen berechnet und die beiden Gleichungen für das Modell mit dem kleinsten BIC-Wert ausgegeben. Fraktionell integrierte VARMA[1, 1]-Modelle werden im Paket LongMemoryTS behandelt.

Die Schätzung hochdimensionaler VARMA- und VARX-Modelle wird mit den Paketen BigVAR, bigtime und onlineVAR bewerkstelligt. Hochdimensionale Modelle sind mit dem so genannten *Fluch der hohen Dimension* und mit der damit einhergehenden Überparametrisierung konfrontiert. Hier wird die Dimension durch den Einsatz einer Straffunktion erstmal reduziert.

urca ist ein Paket speziell zur Integrations- und Kointegrationsanalyse. Mit qunitroot wird eine Funktion zur Bestimmung der Quantile nach MacKinnon bereitgestellt. In dem Paket tseries findet man die Funktion po.test, mit der der Phillips-Ouliaris-Kointegrationstest durchgeführt werden kann. Vektor-Fehlerkorrektur-Modelle können mit der Prozedur ECMvar aus dem Paket MTS geschätzt werden.

9 Regressionsmodelle für Zeitreihen

In diesem Kapitel werden multivariate zeitlich erhobene Phänomene in unsymmetrischer Weise betrachtet. Genauer geht es um Einflüsse von deterministischen oder stochastischen Größen auf eine interessierende Variable Y bzw. auf einen Prozess (Y_t).

9.1 Regression mit autokorrelierten Störungen

Die Vorstellung, dass eine Zeitreihe (y_t) von anderen Reihen $(x_{1t}), (x_{2t}), \dots, (x_{mt})$ linear abhängig ist, kann formal durch einen linearen Regressionsansatz ausgedrückt werden:

$$Y_t = \sum_{j=1}^m \beta_j x_{jt} + U_t . \quad (9.1)$$

In diesem Abschnitt werden die (x_{jt}) als feste, vorliegende Reihen aufgefasst. Die Analyse erfolgt somit als durch diese Werte bedingt. Soweit entspricht der Ansatz dem der linearen Regression aus Kapitel 2. Jedoch ist im Zeitreihenkontext die Unabhängigkeit der Störungen U_t selten gegeben. Damit sind die im Rahmen der einfachen linearen Regression bestimmten Standardfehler der Koeffizientenschätzungen $\hat{\beta}_j$ nicht mehr korrekt.

Bei positiven Autokorrelationen streuen die Schätzungen stärker als bei unkorrelierten Störungen. Dies resultiert daraus, dass die Beobachtungen dann länger auf einer Seite der Regressionsgeraden verharren als bei Unabhängigkeit, vgl. Abbildung 9.1. Daher kommen in experimentellen Situationen bei Wiederholungen öfter relativ große bzw. relativ kleine Steigungen vor und Achsenabschnitte, die stärker von den zugrundeliegenden abweichen. Die Standardfehler werden beim Einsatz der KQ-Methode, die die positiven Autokorrelationen nicht berücksichtigt, unterschätzt. Negative Autokorrelationen bewirken dagegen, dass die Residuen eine Tendenz haben, von einer Seite der Geraden auf die andere zu springen; das erlaubt eine präzisere Schätzung. Mit anderen Worten werden dann die Standardfehler bei der Verwendung der KQ-Methode überschätzt.

Beispiel 9.1 (Lineares Trendmodell)

Sei zur Illustration das einfache Trendmodell

$$Y_t = \beta_0 + \beta_1 t + U_t$$

betrachtet, bei dem (U_t) ein AR[1]-Prozess ist, $U_t = \alpha U_{t-1} + \varepsilon_t$. Die Abbildung 9.1 zeigt die Regressionsgerade mit drei Realisationen für die Parameter $\beta_0 = 0$, $\beta_1 = 1$, $\alpha = 0.9$ und $N = 20$. Zwei der Reihen verlaufen gänzlich auf einer der beiden Seiten der zugrundeliegenden Geraden, die dritte weist eine deutlich geringere Steigung auf.

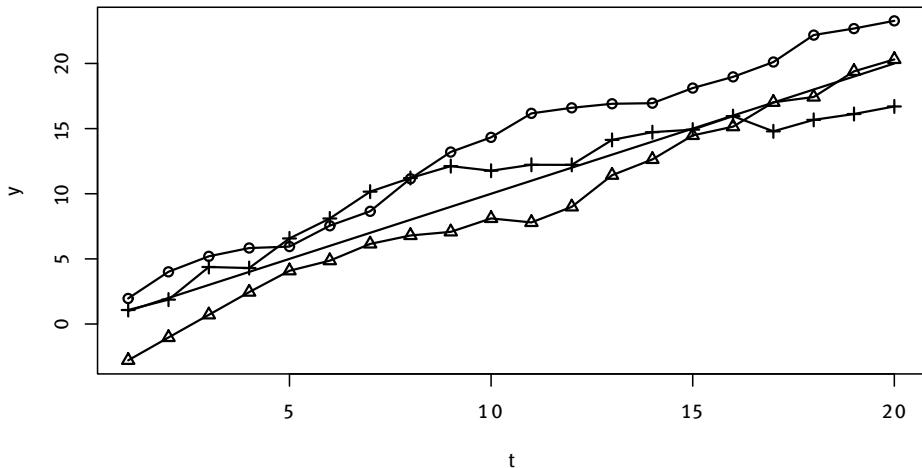


Abb. 9.1: Regressionsgerade mit autokorrelierten Störungen

Eine kleine Simulation mit den der Abbildung zugrundeliegenden Parametern er gab bei 100 Wiederholungen die folgenden Durchschnittswerte und Standardab weichungen für $\hat{\beta}_1$:

α	$\bar{\hat{\beta}}_1 = \frac{1}{100} \sum_{b=1}^{100} \hat{\beta}_{1,b}$	$\hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{1}{100} \sum_{b=1}^{100} (\hat{\beta}_{1,b} - \bar{\hat{\beta}}_1)^2}$
0.7	1.0032	0.0995
0.0	1.0037	0.0380
-0.7	1.0029	0.0303

Die Autokorrelation führt im Mittel nicht zu Fehlschätzungen; die Standardfehler sind in der oben angedeuteten Weise von der Autokorrelation der Störungen abhängig.

Der Regressionsansatz (9.1) ist bei Zeitreihen so zu modifizieren, dass für die Fehler eine geeignete Abhängigkeitsstruktur zugelassen wird. Mit einem ARMA-Fehlerprozess (Z_t) lautet das Modell dann:

$$Y_t = \sum_{j=1}^m \beta_j x_{jt} + Z_t \quad \text{mit} \quad \alpha(B)Z_t = \theta(B)\varepsilon_t ; \quad (9.2)$$

dabei ist (ε_t) ein White-Noise-Prozess.

Die ML-Schätzung kann aufgrund eines allgemeinen Ergebnisses ermittelt werden, indem die Parameter aufgeteilt und abwechselnd verschiedene Parametermengen bestimmt werden. Genauer wird bei gegebenen Startwerten von α, θ eine *generalisierte*

Kleinste-Quadrat-Schätzung für den Parametervektor β durchgeführt. Generalisiert heißt dabei, dass die Kovarianzen der Störungen entsprechend dem ARMA-Modell mit diesen Parameterwerten berücksichtigt werden. Dann werden mit den Residuen dieser Regression die Parameter α, θ des ARMA-Modells mit der ML-Methode geschätzt. Die abwechselnde Schätzung wird iteriert, bis sich die Lösungen stabilisieren. Für die Startwerte werden zuerst die Koeffizienten β_j geschätzt, als ob die Störungen unkorreliert wären. Dann wird auf der Basis der Residuen die Abhängigkeitsstruktur bestimmt. Diese wird im zweiten Durchgang berücksichtigt. Oft wird schon nach einem Iterationszyklus abgebrochen. Über die Fortsetzung der Iteration erhält man sogar die ML-Schätzer mit asymptotisch korrekten Standardfehlern. Die Schätzer $\hat{\alpha}$ und $\hat{\beta}$ sind asymptotisch unabhängig. Das allgemeine Ergebnis, nach dem die Bestimmung von ML-Schätzern in der Weise geschehen kann, dass iterativ über Teilmengen des Parameters maximiert wird, wurde von Oberhofer und Kmenta (1974) bewiesen.

Beispiel 9.2 (Bisam-Felle)

Zur Analyse der Zeitreihe der jährlichen Verkäufe von Bisam-Fellen durch die Hudson's Bay Company in den Jahren 1848–1911 werden wie üblich die logarithmisch transformierten Werte betrachtet. In ihrer Analyse Chan & Wallis (1978) bei der Reihe BISAM einen linearen Trend. Wie die Autoren ausführen, befürworteten andere, die die Reihe modelliert haben, eine quadratische Trendfunktion. In der Abbildung 9.2 sind beide Trendverläufe eingezeichnet.

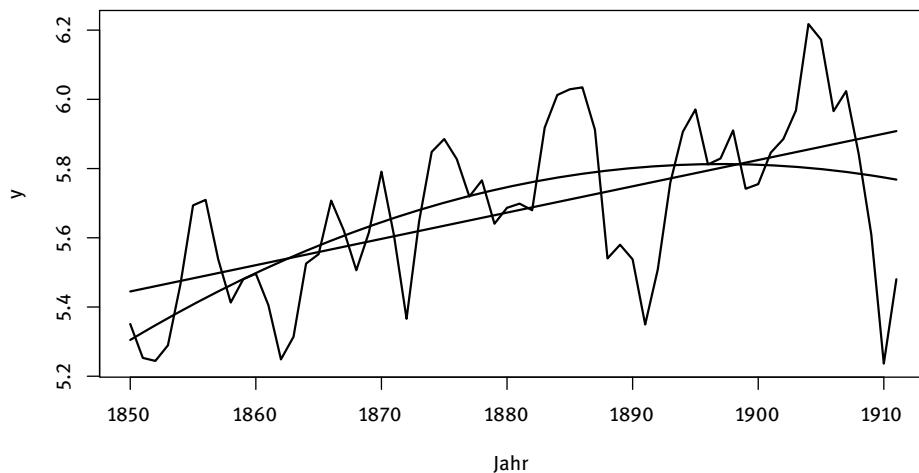


Abb. 9.2: Die Zeitreihe BISAM mit linearem und quadratischem Trend

Die Residuen der Schätzung eines quadratischen Trends führen zu einem AR[2]-Modell für den Fehlerprozess. Die Ergebnisse der Startschätzung und der vollen ML-Schätzung lauten:

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.283e+00	7.564e-02	69.846	< 2e-16
t1	2.206e-02	5.540e-03	3.983	0.000190
t2	-2.296e-04	8.523e-05	-2.694	0.009167

```
Coefficients:
```

	ar1	ar2	intercept	t1	t2
s.e.	0.9527	-0.4483	5.3121	0.0195	-2e-04
	0.1258	0.1594	0.0996	0.0074	2e-04

Bei den Schätzergebnissen der ersten Stufe sind beide Koeffizienten signifikant. Der Koeffizient des quadratischen Terms ist bei der vollen ML-Schätzung dann nicht mehr signifikant; der *P*-Wert des linearen Terms ist um einen Faktor von ca. 40 größer geworden.

R-Code:

```
1 data(MUSKRAT)
2 y <- log10(as.vector(MUSKRAT))
3 n <- length(y)
4 t1 <- c(1:n)
5 t2 <- t1^2
6 out1 <- lm(y~t1)
7 out2 <- lm(y~t1+t2)
8 plot(y)
9 lines(t1,out1$fitted.values)
10 lines(t1,out2$fitted.values)
11 acfpacf(out2$residuals,15)
12 out3 <- arima(y,order=c(2,0,0),xreg=cbind(t1,t2))
13 out3
```

In der Zeile 11 werden die Residuen aus der quadratischen Regression zur Identifikation des Fehlerprozesses verwendet. Die Funktion arima erlaubt es auch, Regressoren einzubeziehen. Die Regressoren t1 und t2 werden über die Option xreg=cbind(t1,t2) ins Modell eingespeist.

Im Abschnitt 3.4 wurde darauf hingewiesen, dass sich die Modellierung von Zeitreihen schwierig gestalten kann, wenn Differenzen gebildet werden und dann noch ein von null verschiedenes Niveau vorliegt. Der Hintergrund ist dann, dass zusätzlich eine (lineare) Trendfunktion berücksichtigt werden muss.

Beispiel 9.3 (Verkäufe eines Unternehmens)

Die monatlichen Verkäufe eines Unternehmens, siehe Newton (1988), sollen prognostiziert werden; dazu wird der Reihe ein ARIMA-Modell angepasst.

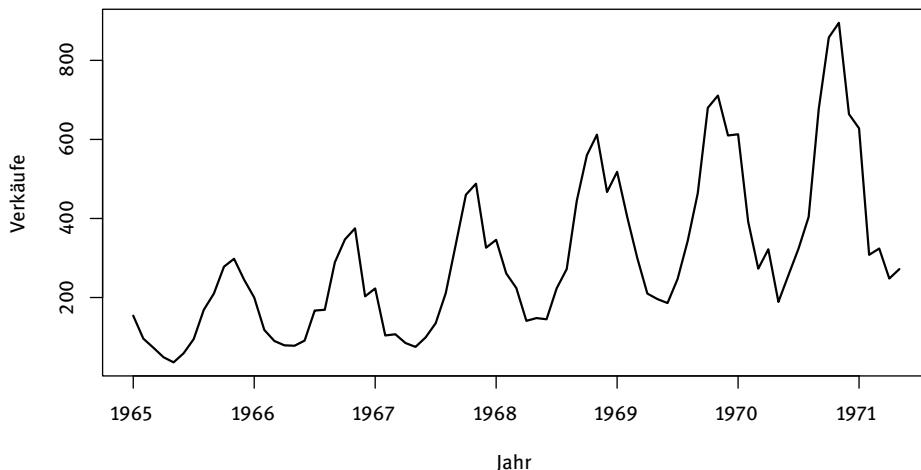


Abb. 9.3: VERKÄUFE eines Unternehmens

Zunächst weist die Reihe VERKÄUFE einen Trend und eine Saison auf, siehe Abbildung 9.3. Die Ausschläge werden dabei mit ansteigendem Niveau größer. Um die Varianz zu stabilisieren, werden die Daten logarithmiert. Die Methode der variaten Differenzen, angewendet auf die Reihe $\ln(y_t)$, legt einmalige saisonale Differenzen nahe. Die Werte von ACF und PACF der saisonal differenzierten logarithmierten Reihe deuten auf ein gemischtes Modell hin, siehe Abbildung 9.4. Bei der Schätzung und der anschließenden Residuenanalyse erweist sich die zunächst nur geringfügig erscheinende Erhöhung der PACF bei dem Lag sieben als relevant. Versuche, dies durch AR- und/oder MA-Terme zu erfassen, führen zu einer weiteren Vergrößerung der Kennfunktionen beim Lag elf. Diese lässt sich nicht modellieren. Der Ljung-Box-Pierce-Test zeigt durchgängig kleine P -Werte.

Die Berücksichtigung eines linearen Trends behebt das Problem. Es ergibt sich als Modell $(1 - \Phi B^{12})(1 - B^{12})(Y_t - \theta t) = (1 - \beta_1 B - \beta_2 B^2)\varepsilon_t$ mit den Schätzungen

Coefficients:

	ma1	ma2	sar1	t
	0.3400	0.5757	-0.3199	0.0203
s.e.	0.1061	0.1005	0.1371	0.0027

Nun ist der Ljung-Box-Pierce-Test nicht mehr signifikant:

lags	df	statistic	p-value
6	3	3.941	0.268
12	9	13.273	0.151
18	15	17.054	0.316
24	21	26.968	0.172

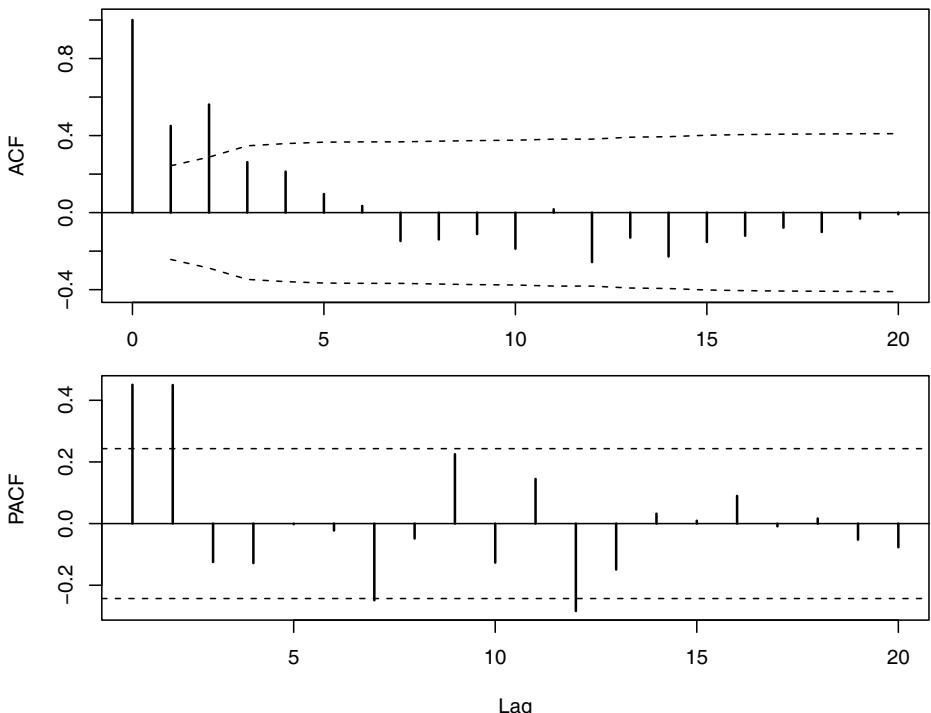


Abb. 9.4: ACF und PACF für $z_t = (1 - B^{12}) \ln(y_t)$

R-Code:

```

1 data(SALES)
2 plot(SALES)
3 y <- log(SALES)
4 vartable(y,12)
5 acfpacf(diff(y,12),20)
6 t <- c(1:length(y))
7 out <- arima(y,order=c(0,0,2),seasonal=list(order=c(1,1,0),
8     period=12), xreg=t)
9 LjungBoxPierceTest(out$residuals,3,24)

```

Der R-Code bringt im Vergleich zu dem von Beispiel 9.2 nichts wesentlich Neues.

Eine Alternative bietet die konsistente Schätzung der Standardfehler für die KQ-Koeffizientenschätzer im Modell (9.1) unter Berücksichtigung der Autokorrelationsstruktur der Störungen, die von Newey & West (1987) eingeführt wurde, kurz *HAC-Standardfehler*, aus dem Englischen „heteroskedasticity and autocorrelation consistent“. Die aus den Termen von (9.1) gebildeten Reihen $(x_{1t} \hat{U}_t), (x_{2t} \hat{U}_t), \dots, (x_{mt} \hat{U}_t)$ bilden eine

m -dimensionale Zeitreihe. In Anlehnung an (8.5) hat diese Zeitreihe eine $m \times m$ Kovarianzmatrix $\Gamma(\tau)$ für jedes Lag τ , d. h. es gibt m^2 Kreuzkovarianzen für jedes Lag. Die geschätzten Kreuzkovarianzen führen dann zu einer Korrektur der Standardfehler.

Beispiel 9.4 (Bisam-Felle – Fortsetzung von Seite 227)

Im Vergleich zu der KQ-Startschätzung fallen die HAC-Standardfehler für den quadratischen Term deutlich höher aus. Wie auch für die volle ML-Schätzung kann dem quadratischen Term auch mit dem HAC-Verfahren keinen signifikanten Einfluß auf die Verkäufe von Bisam-Fellen beigemessen werden.

```
t test of coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.28301289 0.08401791 62.8796 < 2.2e-16
t            0.02206344 0.00765832  2.8810  0.005519
t2           -0.00022965 0.00013417 -1.7117  0.092212
```

R-Code:

```
1 vcovNW <- NeweyWest(out2,lag=4,prewhite=FALSE)
2 coeftest(out2,vcov=vcovNW,df=out2$df.residual)
```

Die Funktion `NeweyWest` gehört zum Paket `sandwich`. Dieses bietet mehrere Funktionen für die Berechnung von HAC-Schätzern, die sich in der Gewichtung der Kreuzkovarianzmatrizen $\hat{\Gamma}_\tau$, dem sogenannten *Lagfenster* unterscheiden. Das Lagfenster wird so gewählt, dass die Kovarianzmatrizen für große Lags kleine Gewichte erhalten oder sogar mit Null gewichtet werden. Die Funktion `NeweyWest` implementiert den HAC-Schätzer in Newey & West (1987). Mit der Option `lag=4` gehen nur die Lags 1 bis 4 in die Berechnung ein. Für die Signifikanztests in Zeile 2 wurde die Funktion `coeftest` aus dem Paket `lmtest` verwendet.

Die gleichen Ergebnisse können mit Hilfe der Funktion `HAC` aus dem R-Paket `tsapp` generiert werden. Da sich diese Funktion die schnelle Fouriertransformation zunutze macht, ist sie besonders für hochdimensionale Modelle relevant, siehe Heberle & Sattarhoff (2017).

R-Code:

```
1 matregr<- matrix(c(rep(1,62),t,t2),nrow=62,ncol=3)
2 matq <- t(matregr)%*%mat_regr/62
3 matxu <- matrix(c(out2$residuals,t*out2$residuals,
4                  t2*out2$residuals),nrow=62,ncol=3)
5 hac <- HAC(matxu, method="Bartlett", 4)
6 vcovHAC <- solve(matq)%*%hac%*%solve(matq)/62
7 coeftest(out2,vcov=vcovHAC,df=out2$df.residual)
```

Im Code wird die Kovarianzmatrix `vcovHAC` der KQ-Koeffizientenschätzer explizit berechnet. Dafür ist die Regressormatrix `matregr` erstmals nötig. Die so genannte Matrix Q (`matq`) wird später mit der Funktion `solve` invertiert. Die Reihen $(x_{1t} \hat{U}_t), (x_{2t} \hat{U}_t), \dots, (x_{mt} \hat{U}_t)$ werden in `matxu` gespeichert.

Nicht umsonst wird die geschätzte Kovarianzmatrix auch *Sandwich-Schätzer* genannt, siehe Greene (2018). In der Zeile 6 kann man die folgende Struktur erkennen: der „Brotbelag“ `hac` liegt zwischen zwei „Brotscheiben“.

9.2 Interventionsanalysen

Nicht selten gibt es im Zeitablauf ein Ereignis, das auf das Niveau der beobachteten Zeitreihe eine kurzfristige oder auch eine lang andauernde Auswirkung auf das Niveau der Zeitreihe hat. Dann stellt sich die Frage, wie dieser Effekt in ein Modell für die Zeitreihe eingebaut werden kann. Ein Beispiel für einen lang andauernden Effekt sind etwa Anzahlen von Verkehrsunfällen nach der deutschen Wiedervereinigung. Hier macht sich schon die größere Zahl der Autofahrer bemerkbar. Das folgende Beispiel illustriert einen kürzer währenden Effekt.

Beispiel 9.5 (Black-Out in New York – Fortsetzung von Seite 4)

Um die Möglichkeit der Auswirkung des Black-Out in New York vom 9. November 1965 zu überprüfen, wird der Black-Out dann als einmalige Intervention einbezogen. Sofern er Auswirkungen gehabt hat, müssten sie in der Abbildung 9.5 dargestellten Reihe BLACKOUT der Anzahlen der wöchentlichen Lebendgeburten in New York City um die zweite Woche im Juli 1966 herum, $t \approx 293$, in der Weise zu sehen sein, dass eine vorübergehende Erhöhung der Anzahlen zu verzeichnen ist.

Um Niveau-Veränderungen in einem ARIMA-Modell zu berücksichtigen, wurden von Box & Tiao (1975) die so genannten *Interventionsmodelle* eingeführt. Sie können als Sonderfall der Regression mit autokorrelierten Störungen angesehen werden; jedoch haben sie eine eigenständige Struktur. Dabei geschieht hier die Beschränkung auf eine einzelne Intervention, um die Darstellung einfach zu halten. Im Ansatz wird also eine unabhängige Variable mit entsprechenden Lags als Regressoren berücksichtigt:

$$Y_t = \sum_{u=0}^{\infty} \nu_u X_{t-u} + U_t . \quad (9.3)$$

Hierbei ist (ν_u) eine absolut summierbare Folge. $\sum_{u=0}^{\infty} \nu_u X_{t-u}$ ist das Niveau der Reihe, sofern (U_t) ein bei null zentrierter Prozess ist. (X_t) sowie die Folge (ν_u) der Gewichte bestimmen daher, wie der Einfluss auf das Niveau der beobachteten Reihe modifiziert wird. Dabei wird die Form der *Interventionsvariable* (X_t) möglichst einfach gehalten;

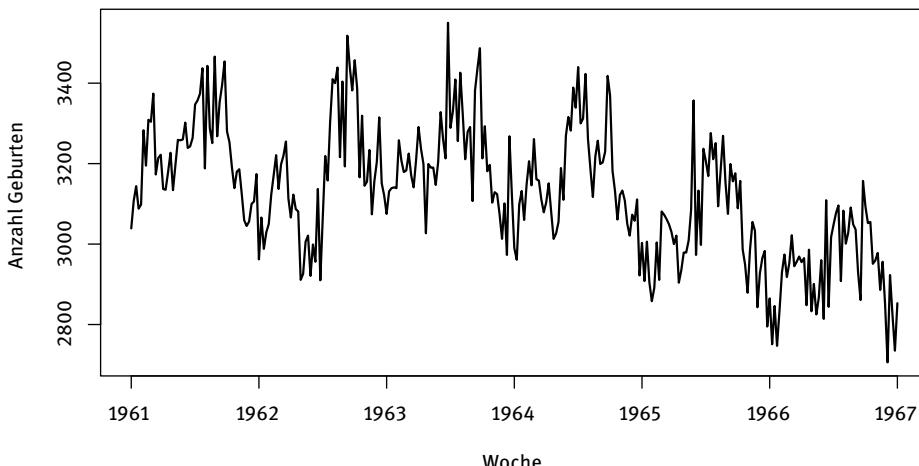


Abb. 9.5: Geburten in New York (Reihe BLACKOUT)

über die Struktur der Koeffizienten werden die Einflüsse der Intervention erfasst. Für

$$I_t = \begin{cases} 1 & \text{für } t = t_0 \\ 0 & \text{sonst} \end{cases} \quad \text{und} \quad v_u = \begin{cases} v & \text{für } u = 0 \\ 0 & \text{sonst} \end{cases} \quad (9.4)$$

ist die Interventionsvariable ein Impuls, der nicht nachklingt. Falls jedoch

$$I_t = \begin{cases} 1 & \text{für } t = t_0 \\ 0 & \text{sonst} \end{cases} \quad \text{und} \quad v_u = \begin{cases} \omega \cdot v^u & \text{für } u \geq 0 \\ 0 & \text{sonst,} \end{cases} \quad (9.5)$$

so klingt der sich zum Zeitpunkt t_0 ereignende Impuls erst (allmählich) ab. Falls sich das Niveau nach der Intervention auf ein anderes Niveau einpegelt, mit anderen Worten einen Niveaushift bei t_0 bewirkt, so ist eine Sprungfunktion zur Modellierung geeignet:

$$H_t = \begin{cases} 0 & \text{für } t < t_0 \\ 1 & \text{für } t \geq t_0. \end{cases} \quad (9.6)$$

Im Ansatz (9.3) sind erst einmal unendlich viele Lags und damit Koeffizienten zugelassen. Von der praktischen Seite her dürfen schon aus schätztechnischen Gründen nur endlich viele Parameter von Null verschieden sein. Dies wird erreicht, indem die Folge der Gewichte durch eine rationale Approximation angenähert wird:

$$v(B) = \frac{(\omega_0 - \omega_1 B - \dots - \omega_s B^s)}{(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r)} B^b \quad (9.7)$$

Obwohl hier nur wenige freie Parameter zur Verfügung stehen, können eine Vielzahl von (unendlichen) Folgen (v_u) zumindest approximativ erzeugt werden. In der Tabelle 9.1 sind einige typische Formen von Impulsresponsefunktionen (v_u) zusammengestellt.

Tab. 9.1: Typische Impulsresponsefunktionen

$$v(B) = \frac{(\omega_0 - \omega_1 B - \dots - \omega_s B^s)}{(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r)} B^b$$

(b,r,s)	IR-Funktion	Typische Impulsgewichte
Impulsresponsefunktionen für $r = 0$		
(2,0,0)	$v(B)x_t = \omega_0 x_{t-2}$	
(2,0,1)	$v(B)x_t = (\omega_0 - \omega_1 B)x_{t-2}$	
(2,0,2)	$v(B)x_t = (\omega_0 - \omega_1 B - \omega_2 B^2)x_{t-2}$	
Impulsresponsefunktionen für $r = 1$		
(2,1,0)	$v(B)x_t = \frac{\omega_0}{(1 - \delta_1 B)} x_{t-2}$	
(2,1,1)	$v(B)x_t = \frac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B)} x_{t-2}$	
(2,1,2)	$v(B)x_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B)} x_{t-2}$	
Impulsresponsefunktionen für $r = 2$		
(2,2,0)	$v(B)x_t = \frac{\omega_0}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-2}$	
(2,2,1)	$v(B)x_t = \frac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-2}$	
(2,2,2)	$v(B)x_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B - \delta_2 B^2)} x_{t-2}$	

Das zugehörige *allgemeine Interventionsmodell* mit nur einer Impuls-Intervention, bei dem Stationarität erst durch geeignete Differenzenbildung hergestellt wird, lautet dementsprechend:

$$Y_t = \frac{(\omega_0 - \omega_1 B - \dots - \omega_s B^s)}{(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r)} B^b X_t + U_t . \quad (9.8)$$

Hier ist die Möglichkeit einer zeitlichen Verzögerung zwischen dem Ereignis und der Auswirkung durch B^b mit aufgenommen. Bei den Geburten in New York ist beispielsweise die Intervention auf die Woche zu legen, in der der Black-Out stattfand; mittels der Setzung $b = 36$ ist die Zeitdifferenz einzubeziehen, bis die möglichen Auswirkungen sichtbar werden konnten.

Eine Interventionsanalyse besteht erstens in der Bestimmung des Interventionseffektes, genauer der Bestimmung der Intervention gemäß (9.5) bzw. (9.6) sowie der Polynome $\omega(B)$ und $\delta(B)$. Damit wird die Auswirkung der Intervention quantifiziert. Zweitens ist die *Noise-Reihe* (U_t) als ARMA-Prozess zu modellieren. In (9.8) ist also das Modell für die Noise-Reihe $\alpha(B)U_t = \beta(B)\eta_t$. Dieses kann aus dem Abschnitt vor der Intervention mittels der für den ersten Abschnitt gültigen Beziehung $U_t = Y_t$ aus den Beobachtungen direkt ermittelt werden. Alternativ kann man das ARMA-Modell unter Verwendung einer Grobschätzung der Interventionsparameter identifizieren. Im zweiten Schritt ist dann eine Gesamtschätzung vorzunehmen. Das resultierende Modell kann natürlich auch zur Prognose verwendet werden.

Falls der Reihenabschnitt vor der Intervention lang genug ist und falls die Intervention die Form eines Impulses aufweist, kann das *Vorgehen bei einer Interventionsanalyse* folgendermaßen zusammengefasst werden:

1. Spezifizierte das ARIMA-Modell für die Zeitreihe U_t aus dem Abschnitt der Reihe vor Intervention.
2. Passe das Modell an und prognostizierte U_t für $t > t_0$; die Prognosen seien \hat{U}_t .
3. Untersuche $Y_t - \hat{U}_t$ für $t > t_0$, um $\omega(B)$ und $\delta(B)$ zu spezifizieren.
4. Führe eine gemeinsame Schätzung unter Verwendung aller Daten durch.
5. Überprüfe das geschätzte Modell bzgl. von Modelldefekten.

Beispiel 9.6 (Black-Out – Fortsetzung von Seite 232)

Für die Zeitreihe (Y_t) der wöchentlichen Anzahlen von Geburten in New York wird im nicht-gestörten Fall, d. h. ohne Black-Out, ein ARIMA-Modell unterstellt. Dabei ist aus dem Plot der Reihe ersichtlich, dass sie instationär ist und u. U. eine Saisonalität vorliegt. Weil es sich um Wochenwerte handelt, ist die Saisonperiode $s = 52$.

Sofern der Black-Out Auswirkungen gehabt hat, müssten sie in der vorliegenden Reihe um den Zeitpunkt $t = 293$ herum in der Weise zu sehen sein, dass eine Erhöhung der Anzahlen zu verzeichnen ist. Die Zeitpunkte für die Intervention ergeben sich dabei aus der entsprechend lang zurück reichenden Erhebung. Izenman & Zabell (1981) schlagen unter Berücksichtigung der Tatsache, dass nicht jedes Kind genau nach 36 Wochen geboren wird, eine Intervention der folgenden Form vor:

$$I_t = \begin{cases} 1 & \text{für } t = 292, 293, 294, 295 \\ 0 & \text{sonst.} \end{cases}$$

Das Gewicht wird für die Zeitpunkte $t = 292, 293, 294, 295$ als gleich angesetzt. Daher lautet der allgemeine Modellansatz für die Reihe mit Intervention:

$$\alpha(B)(1 - B)^d(1 - B^s)^D(Y_t - \omega I_t) = \beta(B)\eta_t .$$

Nun wird zunächst der erste, nicht beeinflusste Teil der Reihe analysiert.

Die Anwendung der Methode der variaten Differenzen auf die ersten 290 Anzahlen der wöchentlichen Geburten führt zu

Einfachdiff	Saisdiff0	Saisdiff1	Saisdiff2	Saisdiff3
0	1.000	0.822	2.349	7.010
1	0.494	0.841	2.749	9.903
2	1.456	2.494	8.206	30.127
3	4.961	8.249	27.075	100.041

Somit sind nur einmal einfache Differenzen zu bilden. Die ACF zeigt nur für das Lag 1 einen großen Wert, danach bricht sie ab. Die PACF ist bei diesem Lag ebenfalls erhöht; sie weist aber eher ein abklingendes Verhalten auf. Dies motiviert die Berücksichtigung eines MA[1]-Terms. Ein solcher ist, wie im Rahmen der ARIMA-Modellierung ausgeführt, bei einfachen Differenzen stets angezeigt. Zudem könnte man spekulieren, dass dieser Modellteil das Verhalten von Ärzten widerspiegelt, die an Wochenenden anstehenden Geburten möglichst hinauszögern. Dann weist die PACF einen größeren Wert bei den Lags 49, 52 und 54 auf. Davon hat nur das Lag 52 eine inhaltlich interpretierbare Bedeutung. Zunächst wird daher nur dieses berücksichtigt. Es wird als MA-Term modelliert, da auch die kurzfristigen Abhängigkeiten mittels MA-Termen modelliert werden. Mit einem saisonalen Ansatz kann man hoffen, auch das Lag 54 zu erfassen. Jedoch zeigt sich, dass ein MA[52]-Modell, bei dem nur die beiden Koeffizienten zu den Lags 1 und 52 berücksichtigt werden, eine adäquatere Anpassung ergibt.

Mit dem für den ersten Reihenabschnitt gefundenen Modell wird das allgemeine Modell für die ganze Zeitreihe spezifiziert:

$$(1 - B)(Y_t - \omega I_t) = (1 - \beta_1 B - \beta_{52} B^{52})\varepsilon_t .$$

Die Koeffizienten des ARMA-Modells sind signifikant, der des Regressors jedoch nicht:

ma1	ma52	It
-0.5941	0.2511	9.2516
s.e.	0.0414	55.3797

Der Ljung-Box-Pierce-Test zeigt, dass in den Residuen keine signifikanten Autokorrelationen verbleiben:

lags	df	statistic	p-value
54	52	60.532	0.195
60	58	64.772	0.252

Eine Intervention ist folglich nicht nachweisbar. Die entsprechenden (immer wieder auftauchenden!) Zeitungsmeldungen sind als nicht belegt einzustufen.

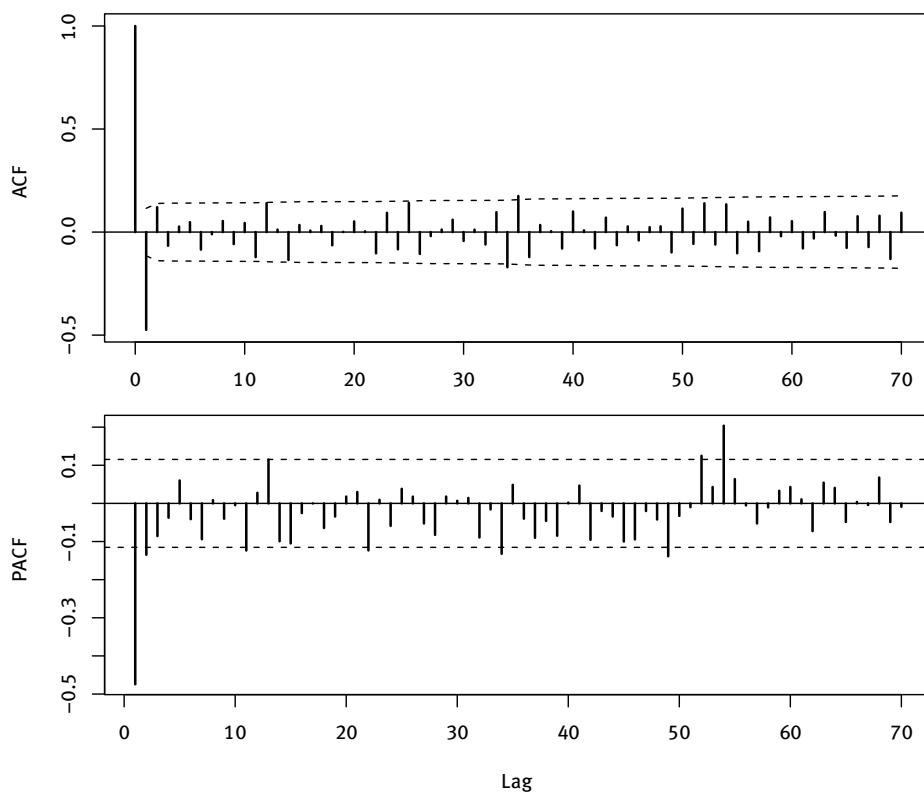


Abb. 9.6: Kennfunktionen für den ersten Abschnitt der Reihe BLACKOUT nach Differenzenbildnen

R-Code:

```

1 data(BLACKOUT)
2 y <- BLACKOUT
3 plot(y,xlab="Woche",ylab="Anzahl Geburten")
4 n <- length(y)
5 It <- rep(0,n)
6 It[292:295] <- 1
7 y1 <- window(y,start=c(1961,1),end=c(1966,30))
8 vartable(y1,52)
9 acfpacf(diff(y1),70)
10 out <- arima(y1,order=c(0,1,52),fixed=c(NA,rep(0,50),NA))
11 acfpacf(out$residuals,70)
12 LjungBoxPierceTest(out$residuals,2,60)
13 out <- arima(y,order=c(0,1,52),xreg=It,fixed=c(NA,rep(0,50),NA,NA))
14 out
15 LjungBoxPierceTest(out$residuals,2,60)

```

In den Zeilen 5 und 6 wird die Interventionsvariable definiert. Sonst bringt der R-Code im Vergleich zu dem von Beispiel 9.2 nichts wesentlich Neues.

Beispiel 9.7 (Verkehrsunfälle – Fortsetzung von Seite 106)

Die Anzahlen der Verkehrsunfälle mit Personenschäden liegt von Januar 1974 bis Ende 1990 für das Gebiet der alten Bundesrepublik vor, ab Januar 1991 für die ganze BRD. Daher liegt es nahe, eine Intervention der Form einer Sprungfunktion anzusetzen:

$$I_t = \begin{cases} 0 & \text{für } t \leq 204 \\ 1 & \text{für } t > 204. \end{cases}$$

Im Beispiel 3.21 wurde den logarithmierten Werten des ersten Abschnitts der Zeitreihe ein SARIMA[11, 0, 0] \times [0, 1, 1]₁₂-Modell angepasst. Es zeigt sich, dass nun gemischte Modelle sowohl für die kurzfristigen und als auch für die saisonalen Abhängigkeiten für eine zufriedenstellende Anpassung nötig sind. Dies hat weitgehende Auswirkungen. Das vollständige Modell für die logarithmierten Werte ist schließlich:

$$(1 - \phi_1 B - \phi_{11} B^{11})(1 - \Phi B^{12})(1 - B^{12})(Y_t - \omega I_t) = (1 - \psi B)(1 - \Psi B^{12})\varepsilon_t.$$

Das Ergebnis der Anpassung des Modells an die ganze Zeitreihe ist:

Coefficients:

ar1	ar11	ma1	sar1	sma1	I
0.7186	0.2545	-0.3179	-0.1465	-0.9205	0.1362
s.e.	0.0399	0.0371	0.0621	0.0498	0.0303

sigma^2 estimated as 0.003982: log likelihood = 682.71,
aic = -1351.42

Die Kennfunktionen der Residuen zeigen nur marginale Verletzungen der üblichen Grenzen, der Ljung-Box-Pierce-Test weist keine signifikante P-Werte auf.

lags	df	statistic	p-value
30	24	26.665	0.320
36	30	29.533	0.490
42	36	30.865	0.711
48	42	33.068	0.836

Die Erhöhung der Unfallzahlen ist also deutlich.

R-Code:

```

1 data(ACCIDENT)
2 y <- log(ACCIDENT)
3 y1 <- window(y,start=c(1974,1),end=c(1990,12))
4 vartable(y1,12)
5 acfpacf(diff(y1,12),50)
6 out <- arima(y1,order=c(11,0,0),
7           seasonal=list(order=c(0,1,1),period=12))
8 acfpacf(out$residuals,50)
9 LjungBoxPierceTest(out$residuals,3,36)
10 I <- rep(1,length(y))
11 I[1:length(y1)] <- 0
12 fix <- c(NA,rep(0,9),NA,NA,NA,NA,NA)
13 out <- arima(y,order=c(11,0,1),seasonal=list(order=
14           c(1,1,1),period=12),xreg=I,fixed=fix)
15 acfpacf(out$residuals,50)
16 LjungBoxPierceTest(out$residuals,6,48)

```

9.3 Transferfunktionsmodelle

Bei den Interventionsanalysen wird der Einfluss einer deterministischen Größe auf eine Zeitreihe betrachtet. Zeitliche Verzögerungen werden in einer spezifischen Weise berücksichtigt. Transferfunktionsmodelle bilden nun einen Ansatz, um den Einfluss von zeitreihenmäßig erhobenen, zufälligen Variablen auf die interessierende Zeitreihe zu erfassen. Dabei werden vergangene Werte des Systeminputs in ähnlicher Weise wie bei den Interventionsanalysen berücksichtigt.

Der Einfachheit halber wird nur eine Input-Zeitreihe berücksichtigt. Um nun den Einfluss des Prozesses (X_t) auf den Prozess (Y_t) zu modellieren, wird analog zu (9.3) angesetzt:

$$Y_t = \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \cdots + U_t . \quad (9.9)$$

Dabei ist (U_t) ein Stör-Prozess, der nicht ‚White‘ zu sein braucht. Er wird jedoch als unabhängig vom Input-Prozess (X_t) vorausgesetzt. Die Folge (ν_u) wird als *Transferfunktion* bezeichnet. Üblich ist wieder die abgekürzte Schreibweise:

$$Y_t = \nu(B)X_t + U_t . \quad (9.10)$$

Das Ziel einer Transferfunktionsmodellierung ist in der Regel, die Transferfunktion $\nu(B)$ zu identifizieren und zu schätzen. Oft möchte man damit eine Verbesserung von

Prognosen erreichen. Um auf eine akzeptable Anzahl von Parametern zu kommen, wird für $v(B)$ wieder eine rationale Approximation verwendet. Wie bei den Interventionsmodellen kann man mittels der Aufnahme einer Zeitverschiebung um b Lags zusätzlich eine *Totzeit* berücksichtigen, falls die Wirkung von X auf Y erst nach b Zeiteinheiten beginnt:

$$v(B) = \frac{\omega_0 - \omega_1 B - \dots - \omega_s B^s}{1 - \delta_1 B - \dots - \delta_r B^r} \cdot B^b . \quad (9.11)$$

Die Transferfunktion v_0, v_1, v_2, \dots lässt sich aus ω_τ und δ_τ gewinnen, indem in (9.11) der Nenner auf die andere Seite gebracht und ein Koeffizientenvergleich vorgenommen wird. Für $b = 0, s = 2, r = 2$ ergibt das z. B.:

$$\begin{aligned} v_0 &= \omega_0 \\ v_1 - \delta_1 v_0 &= -\omega_1 \\ v_2 - \delta_1 v_1 - \delta_2 v_0 &= -\omega_2 \\ v_3 - \delta_1 v_2 - \delta_2 v_1 &= 0 \\ v_4 - \delta_1 v_3 - \delta_2 v_2 &= 0 . \end{aligned}$$

Daraus können nun die Koeffizienten der Transferfunktion rekursiv bestimmt werden. Die Tabelle 9.1 zeigt einige typische Folgen v_u .

Erst wenn man die Parameter b, r und s kennt, kann man die Koeffizienten schätzen. Zur Identifikation der Transferfunktion vergleicht man im Rahmen der Box-Jenkins-Philosophie empirische Größen mit typischen Mustern theoretischer Transferfunktionen.

Die geeignete empirische Größe, die zur Identifikation einer Transferfunktion herangezogen werden kann, ist die Kreuzkovarianz- bzw. Kreuzkorrelationsfunktion. Die Beziehung zwischen der CCF des Inputs und des Outputs eines Transferfunktionsmodells und der Transferfunktion ist i. d. R. jedoch kompliziert. Nur für den Fall, dass (X_t) ein White-Noise-Prozess ist, ist sie einfach und für die Praxis verwendbar.

Satz 9.8 (Transferfunktion bei White-Noise-Input)

Ist in der Transferfunktionsbeziehung $Y_t = v(B)X_t + U_t$ der Input-Prozess (X_t) White-Noise, so gilt:

$$v_\tau = \frac{\sigma_Y}{\sigma_X} \rho_{XY}(-\tau) . \quad (9.12)$$

Beispiel 9.9 (AR-Prozess mit überlagerndem White-Noise)

Sei $X_t = \alpha X_{t-1} + \varepsilon_t$ ein AR[1]-Prozess, der durch einen White-Noise überlagert wird: $Y_t = X_t + \eta_t$. Dann ist

$$Y_t = \sum_{u=0}^{\infty} \alpha^u \varepsilon_{t-u} + \eta_t .$$

(Y_t) kann also als Output und (ε_t) als Input eines Transferfunktionsmodells interpretiert werden. Es folgt:

$$\text{Cov}(Y_t, \varepsilon_{t-u}) = \alpha^u \sigma_\varepsilon^2$$

$$\text{Corr}(Y_t, \varepsilon_{t-u}) = \frac{\alpha^u \sigma_\varepsilon^2}{\sigma_Y \sigma_\varepsilon} = \frac{\sigma_\varepsilon}{\sigma_Y} \alpha^u$$

$$\nu_u = \frac{\sigma_Y}{\sigma_\varepsilon} \text{Corr}(Y_t, \varepsilon_{t-u}) = \alpha^u .$$

Will man also die Kreuzkorrelationsfunktion als Muster für die Identifizierung einer Transferfunktion verwenden, so ist die Ausgangsbeziehung (9.10) in der Weise umzuformen, dass (X_t) durch einen White-Noise-Prozess ersetzt wird. Dies ist möglich, wenn (X_t) ein ARMA-Prozess ist. Dadurch wird ein breiter, für praktische Belange weitgehend ausreichender Anwendungsbereich dieser Modelle eröffnet.

Sei $Y_t = \nu(B)X_t + U_t$ mit $\alpha(B)X_t = \beta(B)\varepsilon_t$. Die Definitionsgleichung wird nun mit $\alpha(B)/\beta(B)$ multipliziert; man erhält:

$$\frac{\alpha(B)}{\beta(B)} Y_t = \nu(B) \frac{\alpha(B)}{\beta(B)} X_t + \frac{\alpha(B)}{\beta(B)} U_t = \nu(B)\varepsilon_t + \frac{\alpha(B)}{\beta(B)} U_t , \quad (9.13)$$

oder kurz, mit offensichtlicher Notation:

$$\tilde{Y}_t = \nu(B)\varepsilon_t + \tilde{U}_t . \quad (9.14)$$

Aufgrund von (9.12) gilt nun:

$$\nu_\tau = \frac{\sigma_{\tilde{Y}}}{\sigma_\varepsilon} \cdot \rho_{\varepsilon \tilde{Y}}(-\tau) . \quad (9.15)$$

Kennt man also die Parameter des ARMA-Prozesses $\alpha(B)X_t = \beta(B)\varepsilon_t$, so lässt sich der Prozess (Y_t) transformieren und die CCF von (\tilde{Y}_t) und (ε_t) bestimmen. Die letzte Gleichung ergibt dann (bis auf einen konstanten Faktor) die Transferfunktion ν_τ .

Für die praktische Umsetzung ist noch bedeutsam, dass für die empirischen Kreuzkorrelationen zweier unkorrelierter Prozesse (Y_t) und (ε_t) , wobei der zweite White-Noise ist, nach Satz 8.8 gilt:

$$\hat{\rho}_{\varepsilon Y}(\tau) \stackrel{d}{\sim} \mathcal{N}(0, N^{-1}) . \quad (9.16)$$

Daher kann man die empirischen Kreuzkorrelationen mit den Grenzen $\pm 2/\sqrt{N}$ vergleichen, um zu sehen, für welche Lags sie signifikant sind.

Insgesamt läuft die *Identifikation eines Transferfunktionsmodells* in folgenden Schritten ab:

1. Aufstellen des Ansatzes $Y_t = \nu(B)X_t + U_t$.
2. Anpassen eines ARMA-Modells für die zentrierte (!) Reihe (X_t) : $\hat{\alpha}(B)X_t = \hat{\beta}(B)\varepsilon_t$.

3. Durchführen der als *Prewitthing* bezeichneten Transformation der Input-Reihe: $\hat{\alpha}(B)x_t = \hat{\varepsilon}_t$. (Hier kann man einfach die Residuen $\hat{\varepsilon}_t$ der Modellanpassung nehmen.)
4. Filtern der Output-Reihe: $\frac{\hat{\alpha}(B)}{\hat{\beta}(B)}y_t = \tilde{y}_t$.
5. Bestimmen der CCF von $(\hat{\varepsilon}_t)$ und (\tilde{y}_t) sowie von $\hat{\sigma}_{\tilde{Y}}$, $\hat{\sigma}_{\hat{\varepsilon}}$, um \hat{v}_τ zu ermitteln:

$$\hat{v}_\tau = \frac{\hat{\sigma}_{\tilde{Y}}}{\hat{\sigma}_{\hat{\varepsilon}}} r_{\hat{\varepsilon}\tilde{Y}}(-\tau) .$$

6. Identifizieren von b , r und s über Vergleich von $\hat{v}(B)$ mit „typischen Mustern“ aus der in der Tabelle 9.1 angegebenen Übersicht der Polynomquotienten

$$\hat{v}(B) = \frac{\omega_0 - \omega_1 B - \dots - \omega_s B^s}{1 - \delta_1 B - \dots - \delta_r B^r} \cdot B^b .$$

b ist speziell das erste Lag, für das die Kreuzkorrelation signifikant von Null verschieden ist.

7. Bilden von $\hat{U}_t = Y_t - \hat{v}(B)X_t$ und Identifizieren des ARMA-Modells $\phi(B)\hat{U}_t = \theta(B)\eta_t$.
8. Das gesamte Modell als ARMA-Modell unter Verwendung der identifizierten Transferfunktion (d. h. von r , s und b) und des im letzten Schritt identifizierten Modells für U_t schätzen.

Für die Identifikation einer Transferfunktion schlagen Muller & Wei (1997) ein formelles Vorgehen auf der Basis iterativer Regressionen vor, das auch automatisierbar ist.

Beispiel 9.10 (Alkoholkonsum und Einkommen)

Der Alkoholkonsum hängt auch vom Einkommen ab. Für die Jahre 1870 bis 1938 geben Durbin & Watson (1951) die logarithmierten Werte des Alkoholkonsums pro Kopf, Y_t , und des Realeinkommens pro Kopf, X_t , an, die bivariate Reihe ALK-EINK. Während die Ausgangsreihen einfach eine gegenläufige Tendenz zeigen, weisen die ersten Differenzen auf einen Gleichklang bzw. ein tendenzielles Nachhinken des Alkoholkonsums bei den Zuwächsen hin, vgl. die Abbildung 9.7. Folglich macht es Sinn, den Ansatz

$$(1 - B)Y_t = v(B)(1 - B)X_t + U_t$$

aufzustellen. Die Input-Reihe $(1 - B)x_t$ wird zunächst zentriert; dies ergibt x_t^* . Dann wird für x_t^* ein ARMA-Modell ermittelt. Die in der Abbildung 9.8 dargestellten Kennfunktionen von x_t^* weisen auf die Eignung eines AR[2]-Modells hin.

Das geschätzte Modell lautet

$$X_t^* - 0.3036X_{t-1}^* + 0.4035X_{t-2}^* = \varepsilon_t$$

(0.1106)	(0.1102)
----------	----------

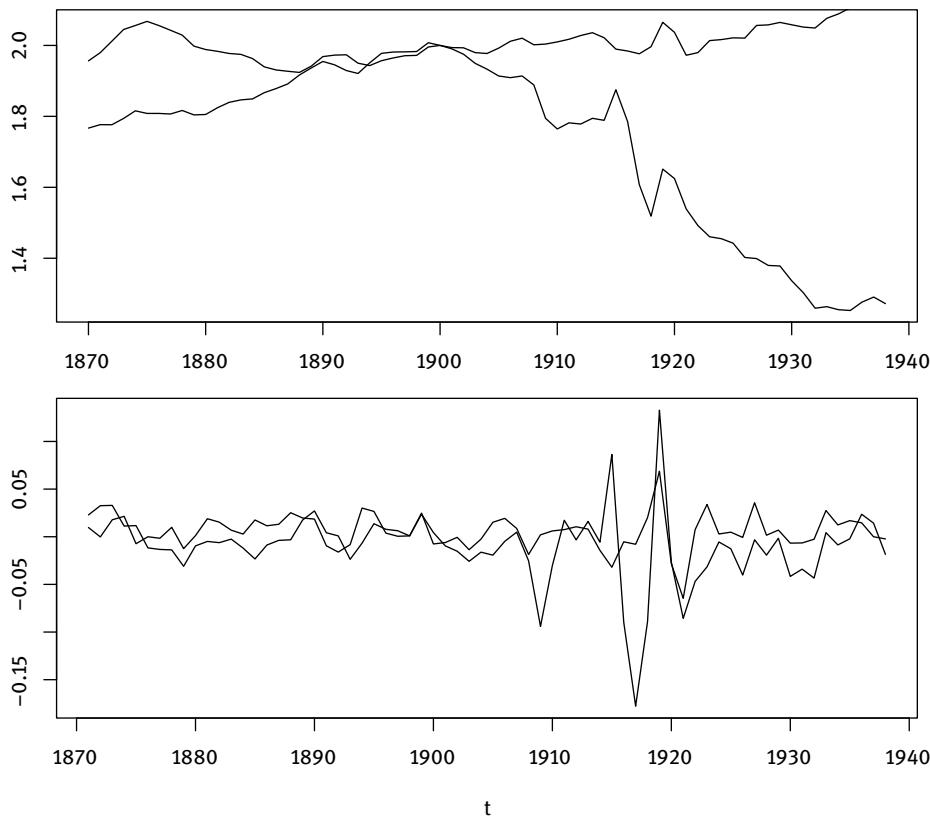


Abb. 9.7: Einkommen (—) und Alkoholkonsum (- - -) in UK; oben: Originalreihen, unten: erste Differenzen

Das Modell gibt einen guten Fit; die Residuen dieses Modells weisen keinerlei Autokorrelationen mehr auf. Mit dem Filter $(1 - 0.3036B + 0.4035B^2)$ werden beide Reihen $y_t^* = (1 - B)y_t$ und x_t^* gefiltert. Die resultierenden Reihen seien mit y_t^w , x_t^w bezeichnet. Ihre CCF zeigt beim Lag $\tau = 0$ einen deutlich von Null verschiedenen Wert, siehe Abbildung 9.9. Zudem weisen die Kreuzkorrelationen $c_{x^w y^w}(-2)$ und $c_{x^w y^w}(-4)$ auf verzögerte Einflüsse hin. Der erhöhte Wert $c_{x^w y^w}(-4)$ zeigt für sich betrachtet einen positiven Zusammenhang zwischen X_{t-4} und Y_t . Das bedeutet, dass die Erhöhung (Verringerung) des Einkommens verzögert zu einer Erhöhung (Verringerung) der Ausgaben für Alkohol führt. Andererseits ist die negative Kreuzkorrelation $c_{x^w y^w}(-2)$ stärker als der positive Wert der CCF bei Lag -4. Somit hat der Gesamteinfluss eher die Form einer gegenläufigen Tendenz.

Daher wird für die Transferfunktion entsprechend dem Punkt 6 der Liste der Identifikation eines Transferfunktionsmodells das sich aus $\hat{v} = r_{\hat{\epsilon}\tilde{Y}}(-\tau)\hat{\sigma}_{\tilde{Y}}/\hat{\sigma}_\epsilon$ ergibt.

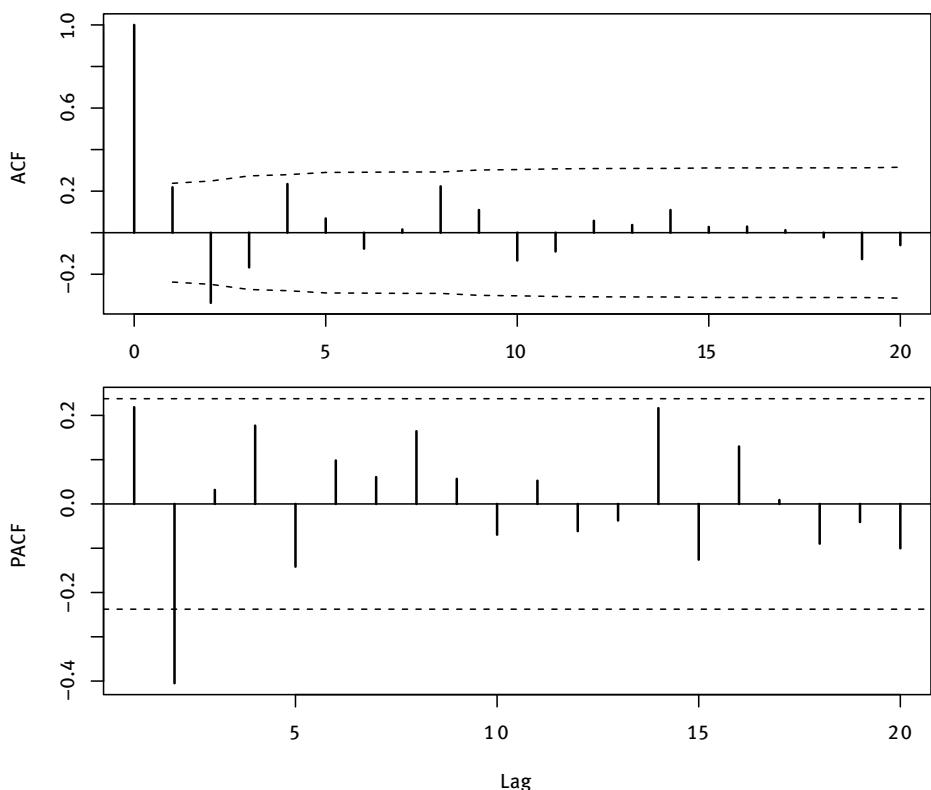


Abb. 9.8: ACF und PACF für $(1 - B)x_t$

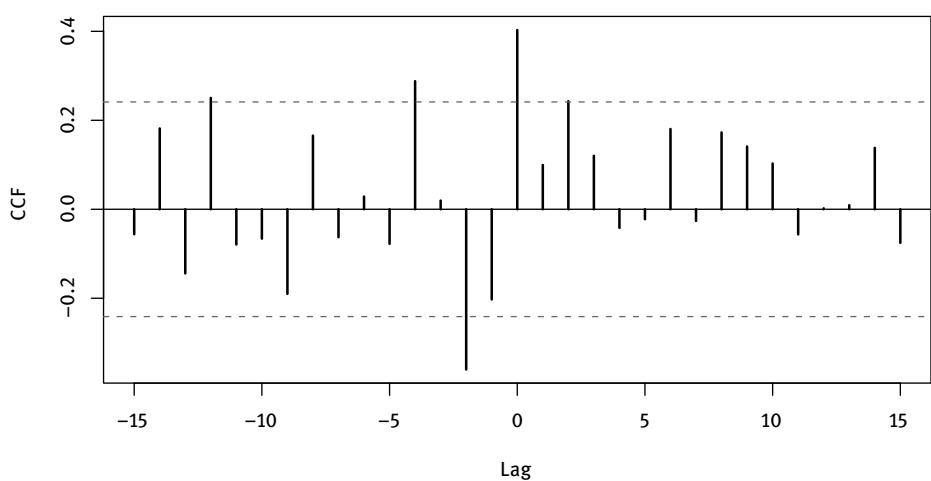


Abb. 9.9: CCF von x_t^w und y_t^w

bende Polynom in B angesetzt:

$$\hat{v}(B) = 0.182 - 0.163B^2 + 0.130B^4.$$

Die Identifizierung der Reihe $\hat{U}_t = Y_t^* - 0.182X_t^* + 0.163X_{t-2}^* - 0.130X_{t-4}^*$ führt auf ein AR[2]-Modell. Damit wird eine Gesamtschätzung durchgeführt. Als abschließendes Modell ergibt sich:

$$Y_t^* = 0.0050 + 0.1517X_t^* - 0.1820X_{t-2}^* + 0.1038X_{t-4}^* + U_t$$

mit $U_t = 0.2243U_{t-1} - 0.3623U_{t-2} + \eta_t$.

R-Code:

```

1 data(ALCINCOME)
2 dEink <- diff(ALCINCOME[,2])
3 dEink <- dEink-mean(dEink)
4 dAlk <- diff(ALCINCOME[,1])
5 acfpacf(dEink,20)
6 out <- arima(dEink,order=c(2,0,0),include.mean=FALSE)
7 out
8 acfpacf(out$resid,20)
9 xw <- out$resid
10 yw <- filter(dAlk,c(1,-out$coef),method="convolution",sides=1)
11 yw <- yw[-c(1,2)]
12 xw <- xw[-c(1,2)]
13 cc <- ccf(xw,yw,lwd=2)
14 nu <- cc[c(0,-2,-4)]
15 nu <- c(nu$acf[1],0,nu$acf[2],0,nu$acf[3])
16 se <- sd(xw)
17 sy <- sd(yw)
18 nu <- nu*sy/se
19 xw <- filter(dEink,nu,method="convolution",sides=1)
20 U <- dAlk-xw
21 U <- U[-c(1:4)]
22 acfpacf(U,20)
23 out <- arima(U,order=c(2,0,0))
24 out
25 acfpacf(out$resid,20)
26 n <- length(dEink)
27 trans <- cbind(dEink[-c(1:4)],dEink[-c(1:2,(n-1):n)],
28                 dEink[-c((n-3):n)])
29 out <- arima(dAlk[-c(1:4)],order=c(2,0,0),xreg=trans)
30 out

```

Der Code orientiert sich an dem im Text beschriebenen Vorgehen. Zu beachten ist, dass für die Abschlusschätzung die relevanten zeitlich verschobenen Inputvariablen „händisch“ produziert werden.

Bei der Bestimmung eines Transferfunktions-Modells sind die allgemeinen Voraussetzungen für diese Modelle zu überprüfen. Gegebenenfalls sind geeignete ‚Korrekturen‘ vorzunehmen; schlimmstenfalls ist auf einen derartigen Ansatz ganz zu verzichten. Ein Teil der Voraussetzungen ist natürlich vor der Identifikation zu checken, ein anderer benötigt die im Rahmen der Identifikation anfallenden Größen. Eine *Checkliste der Voraussetzungen* umfasst folgende Punkte:

1. Stationarität von (Y_t) und (X_t) . Falls dies nicht gegeben ist, ist das Modell unsinnig. Hier können aber vorab mittels der im zweiten Kapitel besprochenen Techniken die Reihen in stationäre überführt werden.
2. Unkorreliertheit von (ε_t) und der Realisation des Prozesses (\tilde{Y}_t) , der sich gemäß (9.14) ergibt. Um dies zu überprüfen, ist die empirische CCF von $\hat{\varepsilon}_t$ und dem Residualprozess $\hat{\eta}_t$ aus Punkt 7 des Identifikationsschemas zu untersuchen. Die CCF sollte
 - kein systematisches Verhalten aufweisen
 - im Bereich $\pm 2/\sqrt{N}$ bleiben (Die Grenzen sind das Doppelte der Standardabweichung von $\hat{\rho}_{\varepsilon, \eta_t}(\tau)$ falls (ε_t) White-Noise ist und (ε_t) und (η_t) unkorreliert sind.)

Zusätzlich kann der Portmanteau-Test angewendet werden. Er basiert hier auf der Prüfgröße

$$Q = m(m+2) \sum_{\tau=0}^k \frac{1}{m-\tau} \hat{\rho}_{\hat{\varepsilon}, \hat{\eta}}(\tau),$$

die bei Unkorreliertheit approximativ χ^2_{k+1-m} -verteilt ist. Dabei ist m die Anzahl der Zeitpunkte t , für die \tilde{Y}_t bestimmt wurde. k sollte nicht zu klein sein, jedenfalls größer als die höchste berücksichtigte Modellordnung.

9.4 R-Funktionen

Die Prozedur `arima` ist so angelegt, dass feste Regressoren über das optionale Argument `xreg` berücksichtigt werden können. Damit können folglich auch Interventionsanalysen mit einfachen Impulsinterventionen oder mit Niveaushifts angepasst werden. Einfachere Transferfunktionsmodelle können ebenso geschätzt werden. Das gleiche gilt für `arfima` aus dem gleichnamigen Paket. Komplexe Interventionen und Transferfunktionen (siehe Tabelle 9.1) einzubeziehen erlaubt die Funktion `arimax` des Paketes `TSA`. Einige einschlägige Beispiele können in Cryer & Chan (2010), das Buch zum Paket `TSA` konsultiert werden.

HAC-Schätzer mit verschiedenen Lagfenstern können mit der Funktion `lrvar` im Paket `sandwich` implementiert werden.

10 Zustandsraummodelle und Kalman-Filter

10.1 Zustandsraummodelle

Den Ausgangspunkt der Zustandsraummodelle bilden sich zeitlich entwickelnde Systeme, die mit Störungen beobachtet werden. Dazu gehört etwa der in Beispiel 3.17 betrachtete physiologische Tremor. Hier handelt es sich um einen AR[2]-Prozess, der nicht rein beobachtet werden kann, sondern nur überlagert von einer durch die Messvorrichtung verursachten Störung. Um dies formal zu beschreiben, ist einmal die zugrunde liegende Prozessgleichung anzugeben, welche die Entwicklung des *Systems* beschreibt:

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \eta_t ; \quad (t = 1, 2, 3, \dots).$$

X_{-1}, X_0 sind dabei geeignete Startvariablen. Zusätzlich ist zu formalisieren, dass die *Beobachtung* gestört erfolgt:

$$Y_t = X_t + \varepsilon_t .$$

Dabei wird (η_t) wie (ε_t) als ein bei null zentrierter White-Noise-Prozess vorausgesetzt; seine Varianz ist σ_η^2 .

Die Entwicklung des Systems wird bei Zustandsraummodellen stets in der Weise angegeben, dass der aktuelle Zustand nur von dem letzten davor liegenden Zustand abhängt. Das ist möglich, wenn man (trotz univariater Größen) zu Vektoren und Matrizen übergeht. Dann erhält man:

$$\begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \eta_t . \quad (10.1)$$

Hier tauchen zwar die drei Zeitpunkte $t, t - 1, t - 2$ noch auf; werden X_t, X_{t-1} aber als Komponenten eines Zufallsvektors x_t geschrieben, so verschwindet diese weiter zurückliegende Abhängigkeit zumindest formal.

Die Gleichung, welche die Beobachtung mit dem Zustand verknüpft, lässt sich entsprechend modifizieren:

$$Y_t = (1 \ 0) \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} + \varepsilon_t . \quad (10.2)$$

Wie in dieser Situation benötigt man zur Formulierung eines Zustandsraummodells generell zwei Gleichungen. Die erste Gleichung wird als Zustands- oder Systemgleichung und die zweite als Mess- oder Beobachtungsgleichung bezeichnet. Die Generalisierung führt dann zu der folgenden Definition. Die Bezeichnungen werden dabei in Anlehnung an Harvey (1989) gewählt. Seine Notation wird in zahlreichen Publikationen adaptiert. Hierbei wird speziell der Zustand mit dem Symbol α bezeichnet um zu verdeutlichen, dass er einen Parameter(vektor) darstellt.

Definition 10.1 (Zustandsraummodell)

Ein *Zustandsraummodell* für eine multivariate Zeitreihe verknüpft die Beobachtungen $\mathbf{y}_t = (Y_{t,1}, \dots, Y_{t,r})'$ mit einem k -dimensionalen *Zustandsvektor* $\boldsymbol{\alpha}_t$ gemäß der *Beobachtungsgleichung*

$$\begin{pmatrix} Y_{1,t} \\ \vdots \\ Y_{r,t} \end{pmatrix} = \begin{pmatrix} z_{11} & \dots & z_{1k} \\ \vdots & & \vdots \\ z_{r1} & \dots & z_{rk} \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \vdots \\ \alpha_{k,t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{r,t} \end{pmatrix},$$

kurz

$$\mathbf{y}_t = \mathbf{Z}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t. \quad (10.3)$$

Die zeitliche Dynamik des im Allgemeinen nicht beobachtbaren Zustandes entspricht einer Autoregression erster Ordnung. Sie ist gegeben durch die *Systemgleichung*,

$$\begin{pmatrix} \alpha_{1,t} \\ \vdots \\ \alpha_{k,t} \end{pmatrix} = \begin{pmatrix} t_{11} & \dots & t_{1k} \\ \vdots & & \vdots \\ t_{k1} & \dots & t_{kk} \end{pmatrix} \begin{pmatrix} \alpha_{1,t-1} \\ \vdots \\ \alpha_{k,t-1} \end{pmatrix} + \begin{pmatrix} r_{11} & \dots & r_{1l} \\ \vdots & & \vdots \\ r_{k1} & \dots & r_{kl} \end{pmatrix} \begin{pmatrix} \eta_{1,t} \\ \vdots \\ \eta_{l,t} \end{pmatrix},$$

kurz

$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \mathbf{R}\boldsymbol{\eta}_t. \quad (10.4)$$

Hierbei sind die Matrizen \mathbf{Z} , \mathbf{T} und \mathbf{R} bekannt. $\boldsymbol{\varepsilon}_t$ und $\boldsymbol{\eta}_t$ sind r - bzw. l -dimensionale White-Noise-Prozesse, die bei null zentriert sind und die Kovarianzmatrizen \mathbf{H} bez. \mathbf{Q} besitzen. $\boldsymbol{\alpha}_0$ ist ein geeigneter Startvektor.

Es werden zusätzlich die folgenden vereinfachenden Annahmen über die Zufallsvektoren der beiden Gleichungen (10.4) und (10.3) getroffen:

1. Die beiden Störungen $\boldsymbol{\varepsilon}_t$ und $\boldsymbol{\eta}_t$ sind für alle Kombinationen von Zeitpunkten s und t unkorreliert, d. h. es gilt: $E(\boldsymbol{\varepsilon}_t \boldsymbol{\eta}_s') = \mathbf{0}$, für alle s und t .
2. Sowohl $\boldsymbol{\varepsilon}_t$ als auch $\boldsymbol{\eta}_t$ sind multivariat normal verteilt.
3. $\boldsymbol{\alpha}_1$ ist sowohl mit $\boldsymbol{\varepsilon}_t$ als auch mit $\boldsymbol{\eta}_t$, $t = 1, 2, \dots$, unkorreliert.

Die Kovarianzmatrizen \mathbf{H} und \mathbf{Q} heißen auch *Hyperparameter*; die Parametermatrix \mathbf{Z} bzw. \mathbf{T} wird als *Übergangsmatrix* bezeichnet. In verschiedenen Anwendungen lässt man zeitabhängige Modellmatrizen zu. Dies wird hier nicht betrachtet.

Neben den AR-Prozessen gibt es eine Vielzahl von Zeitreihenmodellen, die sich in die Form eines Zustandsraummodells bringen lassen. Speziell in ökonomischen Anwendungen wählt man häufig Modelle, die das stochastische Verhalten einer Zeitreihe durch die Überlagerung von verschiedenen, inhaltlich interpretierbaren Komponenten erklären. Sind die Komponenten deterministisch, so liegen sehr einfache Zu-

standsraummodelle vor. Werden die Komponenten selbst als stochastisch angesetzt, so spricht man von *strukturellen Modellen*. Strukturelle Modelle gehören ebenfalls zu der Klasse der Zustandsraummodelle. Deren Ursprung liegt allerdings im ingenieurwissenschaftlichen Bereich. Harvey (1989) behandelt strukturelle Zeitreihenmodelle unter dem Blickwinkel eines Anwenders aus der Ökonomie. Die anwendungsorientierten Bücher Pole, West & Harrison (1994) sowie West & Harrison (1997) sind gänzlich dem Zustandsraummodell gewidmet. Die Autoren nehmen dabei eine Bayessche Perspektive ein.

Beispiel 10.2 (Komponentenmodell mit deterministischem Trend)

Komponentenmodelle mit deterministischer Trendfunktion können als Zustandsraummodelle geschrieben werden. Sei etwa $Y_t = \mu_t + \varepsilon_t$ mit einem White-Noise-Prozess (ε_t) und einer linearen Trendfunktion $\mu_t = a + bt$. Mit $\mu_t = \mu_{t-1} + b$ ist hier der Zustand $\alpha_t = (\mu_t \ 1)'$. Daher lautet die Systemgleichung

$$\begin{pmatrix} \mu_t \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ 1 \end{pmatrix} .$$

Als Beobachtungsgleichung ergibt sich

$$Y_t = (1 \ 0) \begin{pmatrix} \mu_t \\ 1 \end{pmatrix} + \varepsilon_t .$$

Beispiel 10.3 (Random-Walk mit überlagertem White-Noise-Prozess)

Gegeben sei das Komponentenmodell $Y_t = \mu_t + \varepsilon_t$, wobei (ε_t) ein White-Noise-Prozess ist und der Trend einem Random-Walk folgt: $\mu_t = \mu_{t-1} + \eta_t$. Als Beobachtungs- und Systemgleichung interpretiert, ist das offensichtlich ein einfaches Zustandsraummodell.

Beispiel 10.4 (Splinesglättung)

Der Zerlegungsansatz auf der Basis der Splinesglättung lässt die Interpretation zu, dass die Forderung (2.27) für die glatte Komponente implizit einen stochastischen Ansatz beinhaltet:

$$g_t = 2g_{t-1} - g_{t-2} + \eta_t .$$

Zufallsfehler sind auch als Abweichungen von der glatten Komponente zu berücksichtigen. Diese stellen die Beobachtungsfehler dar. Zusammen lauten die der Splinesglättung zugrunde liegenden Gleichungen dann:

$$\begin{aligned} \begin{pmatrix} g_t \\ g_{t-1} \end{pmatrix} &= \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} g_{t-1} \\ g_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \eta_t \\ Y_t &= (1 \ 0) \begin{pmatrix} g_t \\ g_{t-1} \end{pmatrix} + \varepsilon_t . \end{aligned}$$

Beispiel 10.5 (Additives Holt-Winters-Verfahren)

Die additive Variante des Holt-Winters-Verfahrens für Quartalsdaten gibt Anlass zu folgendem Zustandsraummodell. Beobachtet wird eine Reihe Y_t , die sich aus Trend, Saison und Rest zusammensetzt:

$$Y_t = \mu_t + s_t + \varepsilon_t .$$

Der Trend ist lokal linear; dabei wird die Steigung ebenfalls rekursiv aufdatiert. Das führt zu den beiden Gleichungen:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad \text{und} \quad \beta_t = \beta_{t-1} + \zeta_t .$$

Analog gilt für die Saison unter Berücksichtigung der Forderung $\sum_{u=0}^3 s_{t-u} \approx 0$:

$$s_t = -s_{t-1} - s_{t-2} - s_{t-3} + \nu_t .$$

Die letzten drei Gleichungen werden zur Systemgleichung verbunden:

$$\begin{pmatrix} \mu_t \\ \beta_t \\ s_t \\ s_{t-1} \\ s_{t-2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ s_{t-1} \\ s_{t-2} \\ s_{t-3} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_t \\ \zeta_t \\ \nu_t \end{pmatrix} .$$

Die Beobachtungsgleichung pickt aus dem Zustandsvektor die erste und dritte Komponente heraus und verknüpft sie additiv noch mit dem Rest ε_t :

$$Y_t = (1 \ 0 \ 1 \ 0 \ 0) \begin{pmatrix} \mu_{t-1} \\ \beta_t \\ s_t \\ s_{t-1} \\ s_{t-2} \end{pmatrix} + \varepsilon_t .$$

Hyndman et al. (2008) haben die verschiedenen Ansätze des exponentiellen Glättens in einem etwas allgemeiner gefassten Zustandsraummodell subsumiert. Es hat für die additiven Fehler die Form

$$\alpha_t = f(\alpha_{t-1}) + g(\alpha_{t-1})\eta_t , \tag{10.5}$$

$$Y_t = w(\alpha_{t-1}) + \varepsilon_t . \tag{10.6}$$

(ε_t) wird dabei als Gaußscher White-Noise-Prozess mit $E(\varepsilon_t) = 0$ und $\text{Var}(\varepsilon_t) = \sigma^2$ unterstellt.

Bei Zustandsraummodellen hängt der Zustand zum Zeitpunkt t gemäß der Systemgleichung stets ausschließlich vom letzten Zeitpunkt ab. Weiter zurückreichende Abhängigkeiten können einbezogen werden, indem die Dimension des Zustandes erhöht wird. Dies gilt nicht nur wie in (10.1) für AR- sondern allgemein auch für ARMA-Prozesse. Im folgenden Beispiel wird dies anhand einer konkreten Situation vorgeführt.

Beispiel 10.6 (ARMA-Prozess)

Sei (Y_t) ein ARMA[4, 3]-Prozess, $Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \theta_3 Y_{t-3} + \theta_4 Y_{t-4} + \eta_t - \beta_1 \eta_{t-1} - \beta_2 \eta_{t-2} - \beta_3 \eta_{t-3}$. Um die Systemgleichung zu formulieren, wird eine Rekursion so aufgebaut, dass durch sukzessives Einsetzen die in die weitere Vergangenheit reichende Abhängigkeit berücksichtigt wird. Das Modell ist

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \\ Y_{4,t} \end{pmatrix} = \begin{pmatrix} \theta_1 & 1 & 0 & 0 \\ \theta_2 & 0 & 1 & 0 \\ \theta_3 & 0 & 0 & 1 \\ \theta_4 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \\ Y_{4,t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ -\beta_1 \\ -\beta_2 \\ -\beta_3 \end{pmatrix} \eta_t . \quad (10.7)$$

Die ersten beiden Zeilen sind, getrennt aufgeschrieben:

$$\begin{aligned} Y_{1,t} &= \theta_1 Y_{1,t-1} + Y_{2,t-1} + \eta_t \\ Y_{2,t} &= \theta_2 Y_{1,t-1} + Y_{3,t-1} - \beta_1 \eta_t . \end{aligned}$$

Die zweite Zeile wird in die erste eingesetzt; dabei ist der Zeitshift zu berücksichtigen:

$$Y_{1,t} = \theta_1 Y_{1,t-1} + \theta_2 Y_{1,t-2} + Y_{3,t-2} + \eta_t - \beta_1 \eta_{t-1} .$$

Weiteres Einsetzen von $Y_{3,t-2}$ und dann von $Y_{4,t-3}$ ergibt schließlich:

$$Y_{1,t} = \theta_1 Y_{1,t-1} + \theta_2 Y_{1,t-2} + \theta_3 Y_{1,t-3} + \eta_t - \beta_1 \eta_{t-1} - \beta_2 \eta_{t-2} - \beta_3 \eta_{t-3} .$$

Damit erfüllt die erste Komponente des Zustandes gerade die gewünschte ARMA-Gleichung. Die Beobachtungsgleichung lautet folglich:

$$Y_t = (1 \ 0 \ 0 \ 0) \begin{pmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \\ Y_{4,t} \end{pmatrix} .$$

Das im Beispiel praktizierte Vorgehen ist auf offensichtliche Weise verallgemeinerbar. Auch nichtstationäre ARIMA-Modelle können in diese Form gebracht werden. Da aber die Zustandsraumdarstellung für nichtstationäres Verhalten der Beobachtungen meistens mittels struktureller Modellierung von Trend und Saison umgesetzt wird, ist die Zustandsraumdarstellung eines ARIMA-Prozesses weniger üblich als die eines stationären ARMA-Prozesses.

Die Möglichkeit, ARMA-Modelle als Zustandsraummodelle zu schreiben, führt auf die Frage, wann ein Zustandsraummodell stationär ist. Hier spielt die in (10.4) eingeführte Übergangsmatrix \mathbf{T} die entscheidende Rolle. Das lässt sich unmittelbar einsehen. Rekursives Einsetzen des Systemvektors in die Beobachtungsgleichung (10.3) führt auf

$$\mathbf{y}_t = \dots + \mathbf{ZT}^{t+1} \mathbf{R} \boldsymbol{\varepsilon}_{-1} + \mathbf{ZT}^t \mathbf{R} \boldsymbol{\varepsilon}_0 + \mathbf{ZT}^{t-1} \mathbf{R} \boldsymbol{\varepsilon}_1 + \dots + \mathbf{ZTR} \boldsymbol{\varepsilon}_{t-1} + \mathbf{ZR} \boldsymbol{\varepsilon}_t + \boldsymbol{\eta}_t .$$

Damit dies ein sinnvoller Ausdruck wird, müssen Bedingungen erfüllt sein, die denen für die charakteristischen Gleichungen von AR-Prozessen verwandt sind.

Als großer Vorteil von Zustandsraummodellen gegenüber dem Box-Jenkins Ansatz zur Modellierung von Zeitreihen auf der Basis von ARIMA-Modellen wird von manchen Autoren angesehen, dass verschiedene Komponenten einer Zeitreihe, wie Trend und Saison, einzeln modelliert werden, um sie dann in einem Zustandsraummodell zusammenzufassen. Somit lassen sich spezielle Eigenarten der Zeitreihe identifizieren, und ihnen mit entsprechender Modellierung Rechnung tragen. Beim Box-Jenkins Ansatz werden diese Komponenten dagegen wegen der generellen Differenzenbildung zur Erfüllung der Stationaritätsforderung recht grob behandelt. Zudem sind Zeitreihen aus dem ökonomischen und sozialen Anwendungsbereich oft auch nach der Differenzenbildung nicht wirklich stationär. Dann stellt sich die schwierige Frage, ob man nah genug an der Stationarität herangekommen ist. Bei Zustandsraummodellen wird die Stationarität von vornherein nicht vorausgesetzt.

10.2 Kalman-Filter

Das vorwiegende Interesse bei der Analyse einer Zeitreihe mittels eines Zustandsraummodells besteht in der Bestimmung des nicht beobachtbaren Signals oder Zustandes α_t bei gegebenen Daten y_1, \dots, y_s bis zum Zeitpunkt s . Bei $s = t$ liegt ein eigentliches Schätzproblem vor. Man nennt es auch ein *Filterungsproblem*. Der Fall $s > t$ wird als *Glättung* geführt. Im Fall $s < t$ spricht man von *Prognose* oder *Prädiktion*.

Filterung und *Glättung* basieren darauf, dass die Normalverteilung nicht nur der Störterme sondern auch des Startvektors α_0 gilt. Dann sind die Zustandsvektoren $\alpha_0, \alpha_1, \dots, \alpha_t$ gemeinsam normalverteilt und auch die Beobachtungen y_1, \dots, y_s . Diese $t+1+s$ Zufallsvektoren können zu einem gemeinsam multivariat normalverteilten Zufallsvektor zusammengefügt werden. Sei dieser $(z'_1, z'_2)'$, wobei der erste Teil aus den Zuständen und der zweite aus den Beobachtungen besteht. Dann erlaubt die Theorie, den bedingten Erwartungswert $E(z_1|z_2)$ in Abhängigkeit von den Werten von z_2 und den Erwartungswerten so wie den Kovarianzmatrizen von z_1, z_2 anzugeben. Für $t = s$ ist dieser bedingte Erwartungswert der Vektor der gefilterten Werte, für $s > t$ der Vektor der geglätteten Werte. Die formale Aussage gibt das folgende Lemma an.

Lemma 10.7 (Bedingter Erwartungswert und Kovarianzmatrix)

$(z'_1, z'_2)'$ sei ein partitionierter multivariat normalverteilter Zufallsvektor. Dann gilt

$$E(z_1|z_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(z_2 - \mu_2), \quad (10.8)$$

$$\text{Var}(z_1|z_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'. \quad (10.9)$$

Prognosen ergeben sich aus der rekursiven Natur des Modells in naheliegender Weise. Einmal möchte man den Zustand prognostizieren. Dann wird die Systemgleichung

(10.4) unter Verwendung der gefilterten Zustandsvektoren $\hat{\alpha}_t$ fortgeschrieben; zukünftige Störungen werden dabei durch ihre Erwartungswerte, also durch Null, ersetzt. Die Einschritt-Prognose zum Zeitpunkt t ist daher:

$$\hat{\alpha}_{t,1} = \mathbf{T}\hat{\alpha}_t . \quad (10.10)$$

Bisweilen möchte man den beobachteten Prozess prognostizieren. Da (ε_t) ein bei null zentrierter White-Noise-Prozess ist, lautet die optimale lineare Prognose für y_{t+1} :

$$\hat{y}_{t,1} = \mathbf{Z}\hat{\alpha}_{t,1} = \mathbf{ZT}\hat{\alpha}_t . \quad (10.11)$$

Kalman-Filterung

Die für die Filterung und Glättung zu Beginn dieses Abschnittes skizzierte Lösung hat den Nachteil, dass sie numerisch schwer umsetzbar ist. (Es müssen die Inversen sehr großer Matrizen berechnet werden.) Dies vermeidet der *Kalman-Filter*. Er ist ein generelles Rekursionsschema, das die Schätzung auf der Basis der letzten Ein-Schritt-Prognose und der neuesten Beobachtung erlaubt. Dieser rekursive Ansatz für die Schätzung soll anhand eines einfachen Modells verdeutlicht werden.

Beispiel 10.8 (Einfaches Mittelwertmodell)

Gegeben sei das Modell mit konstantem Erwartungswert; die Systemgleichung ist also $\mu = \mu + 0 \cdot \eta$ und die Beobachtungsgleichung lautet:

$$Y_t = \mu + \varepsilon_t , \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) .$$

Es mögen $N-1$ Beobachtungen vorliegen. Auf der Basis dieser Beobachtungen ist das arithmetische Mittel \bar{y}_{N-1} die beste Schätzung von μ , $\hat{\mu}_{N-1} = \bar{y}_{N-1}$. Zugleich ist $\hat{\mu}_{N-1}$ die Prognose für den Zustand zum Zeitpunkt N .

Kommt nun eine neue Beobachtung y_N hinzu, so kann die Schätzung verbessert werden. Die alte Prognose wird im Lichte der neuen Beobachtung korrigiert. Um dies deutlich herauszustellen, wird die neue Schätzung mit Hilfe der alten und dem letzten Schätzfehler ausgedrückt:

$$\begin{aligned} \hat{\mu}_N &= \bar{y}_N = \frac{N-1}{N}\bar{y}_{N-1} + \frac{1}{N}y_N = \bar{y}_{N-1} + \frac{1}{N}(y_N - \bar{y}_{N-1}) \\ &= \hat{\mu}_{N-1} + \frac{1}{N}(y_N - \hat{\mu}_{N-1}) . \end{aligned}$$

Diese Formel ist sehr einleuchtend: Als neue Schätzung dient die alte Prognose, die unter Ausnutzung der Kenntnis des Prognosefehlers geeignet korrigiert wird. Die Korrektur geschieht dabei so, dass der Fehler mit einem Faktor multipliziert wird, welcher sicherstellt, dass die neue Schätzung die kleinstmögliche Fehlervarianz hat.

Das Modell entspricht auch der Ausgangssituation des einfachen exponentiellen Glättens; die eben erhaltene Relation korrespondiert mit der in Gleichung (5.3) angegebenen Fehlerkorrektur-Darstellung.

Ist $\hat{\alpha}_{t-1}$ die optimale Schätzung des Zustandes auf der Basis der Beobachtungen y_1, \dots, y_{t-1} , so ist die nahe liegende Prognose von α_t gegeben durch $\hat{\alpha}_{t-1,1} = T\hat{\alpha}_{t-1}$, vgl. (10.10). Die Beobachtung y_t wird gemäß (10.11) prognostiziert durch $\hat{y}_{t-1,1} = Z\hat{\alpha}_{t-1,1}$. Die Korrektur unter Berücksichtigung der neuen Beobachtung ergibt die neue Schätzung des Zustandes, $\hat{\alpha}_t$.

Zusätzlich zu den Schätzungen des Zustandes werden auch die relevanten geschätzten Kovarianzmatrizen des Schätzfehlers und des Prognosefehlers iterativ aufdatiert. Sei die Kovarianzmatrix des Schätzfehlers $\hat{\alpha}_t - \alpha_t$ gegeben durch

$$\mathbf{P}_t = E((\hat{\alpha}_t - \alpha_t)(\hat{\alpha}_t - \alpha_t)') .$$

Für die Kovarianzmatrix des Prognosefehlers $\alpha_t - \hat{\alpha}_{t-1,1}$ ergibt sich:

$$\hat{\mathbf{P}}_{t-1,1} = T\hat{\mathbf{P}}_{t-1}T' + RQR' . \quad (10.12)$$

Für die Varianz des Prognosefehlers der Beobachtungen gilt entsprechend:

$$\text{Var}(y_t - \hat{y}_{t-1,1}) = Z\hat{\mathbf{P}}_{t-1,1}Z' + \sigma_\varepsilon^2 . \quad (10.13)$$

Der wesentliche Punkt im Korrekturschritt besteht wie im Beispiel 10.8 in der Bildung einer Linearkombination der alten Prognose und dem Prognosefehler, bei dem nun eine Gewichtungsmatrix K_t , die *Kalman-Verstärkung*, eingeht. Sie wird dabei so bestimmt, dass die Varianz des Schätzfehlers minimal wird.

Definition 10.9 (Kalman-Filter)

Der *Kalman-Filter* im Zustandsraummodell mit der Systemgleichung $\alpha_t = T\alpha_{t-1} + R\varepsilon_t$ und der Beobachtungsgleichung $y_t = Z\alpha_t + \eta_t$ ist durch folgende Rekursion bestimmt. Diese besteht aus zwei Schritten, dem *Prädiktionsschritt*:

$$\hat{\alpha}_{t-1,1} = T\hat{\alpha}_{t-1} \quad (10.14)$$

$$\hat{y}_t = Z\hat{\alpha}_{t-1,1} \quad (10.15)$$

$$\hat{\mathbf{P}}_{t-1,1} = T\hat{\mathbf{P}}_{t-1}T' + RQR' \quad (10.16)$$

und dem *Korrekturschritt*. In diesem werden aus den Prognosen $\hat{\alpha}_{t-1,1}$ und $\hat{\mathbf{P}}_{t-1,1}$ Schätzungen ermittelt. Das geschieht unter Verwendung des mittels der neuen Beobachtung bestimmten Prognosefehlers $v_t = y_t - Z\hat{\alpha}_{t-1,1}$:

$$\hat{K}_t = T\hat{\mathbf{P}}_{t-1,1}Z'(Z\hat{\mathbf{P}}_{t-1,1}Z' + H)^{-1} \quad (10.17)$$

$$\hat{\alpha}_t = \hat{\alpha}_{t-1,1} + \hat{K}_t v_t \quad (10.18)$$

$$\hat{\mathbf{P}}_t = \hat{\mathbf{P}}_{t-1,1} - \hat{K}_t Z \hat{\mathbf{P}}_{t-1,1} . \quad (10.19)$$

Die Berücksichtigung der Messgenauigkeit bei der Korrektur von $\hat{\alpha}_t$ ist einfach zu erkennen. Sei dazu die Beobachtung univariat. Hat ε_t nur eine kleine Varianz, so fällt $H_t = \text{Var}(\varepsilon_t)$ bei der Aufdatierung der Kalman-Verstärkung nicht ins Gewicht und

die Kalman-Verstärkung wird hauptsächlich durch die Systemgegebenheiten determiniert. Ist sie groß, so wird K_{t+1} klein und die alte Prognose wird nur wenig geändert; der neuen Beobachtung wird dann nicht viel Vertrauen entgegengebracht.

Um die Kalman-Rekursion durchführen zu können, muss man den Algorithmus mit Startwerten $\hat{\alpha}_0$ und \hat{P}_0 initialisieren. Sind die Systemmatrizen zeitinvariant, so kann der Startvektor $\hat{\alpha}_0$ mit $\mathbf{0}$ initialisiert werden. Startwerte für \hat{P}_0 erhält man aus der Lösung des Gleichungssystems

$$\hat{P}_0 = \mathbf{T}\hat{P}_0\mathbf{T}' + \mathbf{R}\mathbf{Q}\mathbf{R}' .$$

Bei nicht-zeitinvarianten Systemmatrizen ist es üblich, für den Ausgangszustand einfach den Nullvektor und $\hat{P}_0 = k \cdot \mathbf{I}$ mit einer großen Zahl k zu wählen. Dieser Zugang, Unterstellen sehr großer Streuungen für die Fehler und beliebige Wahl des Startvektors, wird im Bayesianischen Sprachgebrauch als *diffuse Prior* notiert.

Ist der zugrunde liegende Prozess stationär, so erreicht der Kalman-Filter schon nach wenigen Rekursionen einen Gleichgewichtszustand. Das heißt, dass dann die Kovarianzmatrizen \hat{P}_t , \tilde{P}_t sowie $\text{Var}(Y_{t+1} - \hat{Y}_{t,1})$ und die Kalman-Verstärkung praktisch konstant sind. Daher ist die Wahl der Startwerte nicht kritisch.

Besonders einfach gestaltet sich die Behandlung *fehlender Werte* bei Zustandsraummodellen. Hier wird der Kalman-Filter wie üblich für den Prognoseschritt eingesetzt. Eine Korrektur wird nur vorgenommen, wenn die entsprechende Beobachtung vorhanden ist. Fehlt sie, so wird der Korrekturschritt übersprungen. Fehlende Werte führen also zu Prognosen der Zustandsvektoren auf der Basis der vorangehenden Beobachtungen.

Glättung

Bisweilen möchte man retrospektiv das Verhalten eines Systems rekonstruieren, um ein den Beobachtungen zugrunde liegendes Phänomen zu verstehen. Dazu wird man die geglätteten Zustandsvektoren heranziehen. Glättung bedeutet ja in dem Kontext des Kalman-Filters, dass die Zustandsvektoren unter Verwendung aller Beobachtungen bestimmt werden.

Die Glättung des Zustandsvektors $\hat{\alpha}_t$ erfolgt mit dem *Kalman-Smoother*. Dies ist ein rückwärts-rekursives Verfahren, das auf Größen zurückgreift, die bei der Kalman-Filterung des Zustandsvektors anfallen. Der Kalman-Smoother kann also erst nach dem Kalman-Filter angewandt werden.

Die zentralen Rekursionsgleichungen des Kalman-Smoothers lauten:

$$\hat{\alpha}_{t|N} = \hat{\alpha}_t + \hat{P}_t^*(\hat{\alpha}_{t+1|N} - \mathbf{T}\hat{\alpha}_t) \quad (10.20)$$

$$\hat{P}_{t|T} = \hat{P}_t + \hat{P}_t^*(\hat{P}_{t+1|T} - \hat{P}_{t+1|t})\hat{P}_t^{*'} \quad (10.21)$$

Dabei ist

$$\hat{P}_t^* = \hat{P}_t \mathbf{T} \hat{P}_{t+1|t}^{-1} . \quad (10.22)$$

Die Rekursionen werden rückwärts in der Zeit durchgeführt, $t = N - 1, \dots, 1$ mit den Anfangsbedingungen $\hat{\alpha}_{N|N} = \hat{\alpha}_N$ und $\hat{P}_{N|N} = \hat{P}_N$. (Es ist nicht garantiert, dass die Matrizen $\hat{P}_{t+1|t}$ nicht-singulär ist. Dann wird anstelle der Inversen eine generalisierte Inverse verwendet.)

Auch geglättete Versionen der Störungen sind von Interesse. Denn nach Anpassung eines Modells sind wie üblich die Basisaspekte der Normalverteilung der Fehler, die Homoskedastizität und die serielle Unkorreliertheit zu überprüfen. Dies sollte unter Verwendung der Residuen der Schätzungen, die jeweils auf alle Beobachtungen basieren, geschehen. Zur Standardisierung der Residuen sollten ebenfalls die aus Glättung resultierenden Größen verwendet werden. Bei multivariaten Modellen sind die verschiedenen Komponenten jedoch korreliert. Daher sollten sie transformiert werden gemäß

$$\boldsymbol{\varepsilon}_t^* = \mathbf{G}_t \boldsymbol{\varepsilon}_t, \quad \hat{\mathbf{P}}_t^{-1} = \mathbf{G}_t' \mathbf{G}_t;$$

dann können die einzelnen Komponenten getrennt univariat diagnostiziert werden. Rekursionen für diese Versionen der Störungen sind bei Durbin & Koopman (2001) angegeben.

Bei der Glättung von Zeitreihen, bei denen Werte fehlen, wird analog zur Filterung verfahren. Hier handelt es sich dann um eine Interpolation der entsprechenden Lücke(n).

Parameterschätzung

In den Anwendungen müssen natürlich die Hyperparameter \mathbf{H} und \mathbf{Q} geschätzt werden. Zudem wurde bisher unterstellt, dass die Übergangsmatrix \mathbf{T} sowie die in der Systemgleichung an den Störvektor $\boldsymbol{\varepsilon}$ heranmultiplizierte Matrix \mathbf{B} bekannt sind. Oft hängen diese aber von unbekannten Parametern ab. Die rekursive Struktur der Zustandsraummodelle erlaubt nun eine recht einfache Angabe der Likelihoodfunktion. Mit der gemeinsamen Dichte $p(\mathbf{y})$ der Beobachtungen y_1, \dots, y_N ist sie, wenn die unbekannten Parameter im Vektor $\boldsymbol{\theta}$ gesammelt sind:

$$L(\boldsymbol{\theta}) = p(y_1, \dots, y_N; \boldsymbol{\theta}) = \prod_{t=1}^N p(y_t | y_{t-1}; \boldsymbol{\theta}), \quad (10.23)$$

wobei $p(y_1 | y_0; \boldsymbol{\theta}) = p(y_1; \boldsymbol{\theta})$.

Mit der Normalverteilungsannahme lässt sich dies nach Logarithmierung unter Verwendung von $\hat{\eta}_t = \mathbf{y}_t - \mathbf{T}_t \boldsymbol{\alpha}_t$ umformen zu

$$\ln(L(\boldsymbol{\theta})) = -\frac{N \cdot r}{2} \ln(\pi) - \frac{1}{2} \sum_{t=1}^N (\ln(\det(\mathbf{P}_t)) + \hat{\eta}_t' \mathbf{P}^{-1} \hat{\eta}_t) \quad (10.24)$$

Hier kommen nur Größen vor, die bei der Kalman-Filterung routinemäßig bestimmt werden. Somit kann $\ln(L(\boldsymbol{\theta}))$ leicht berechnet werden. Die leichte Berechenbarkeit der

Loglikelihood-Funktion ist aber nur eine Seite. Zur Bestimmung der ML-Schätzer ist noch eine numerische Optimierung durchzuführen. Dabei ist zu beachten, dass die Optimierung oft nicht so glatt geht, wie das in manchen Lehrbüchern suggeriert wird. Die Likelihoodfunktion hat in der Regel viele lokale Maxima, so dass die Wahl der Startwerte ganz wesentlich ist. Auch führt ein zu komplex angesetztes Modell schnell dazu, dass die Optimierungsroutine nicht konvergiert.

Beispiel 10.10 (CO₂-Konzentration über Schauinsland)

Von der Messstation Schauinsland des Umweltbundesamtes wird seit 1972 die CO₂-Konzentration im Bereich des Schwarzwaldes aufgenommen. Die vorliegende Zeitreihe umfasst die Monatsmittel der atmosphärischen CO₂-Konzentrationen in parts per million (ppm) von 2005 bis 2010. Zwei von den 72 Werten fehlen. Die Zeitreihe SCHAUINSLAND stammt aus Krämer (2011).

An die Zeitreihe soll ein Zustandsraummodell angepasst werden. Die Auswahl eines passenden Modells liegt in der Verantwortung des Analysten. Die in der Abbildung 10.1 dargestellte Reihe legt einen additiven Holt-Winters-Ansatz mit lokalem linearem Trend nahe, vgl. das Beispiel 10.5. Das hat folgende Systemgleichung zur Folge:

$$\begin{pmatrix} \mu_t \\ \beta_t \\ s_t \\ s_{t-1} \\ \vdots \\ s_{t-10} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & -1 & \dots & -1 & -1 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ s_{t-1} \\ s_{t-2} \\ \vdots \\ s_{t-11} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_t \\ \zeta_t \\ \omega_t \end{pmatrix}.$$

Die Beobachtungsgleichung lautet:

$$Y_t = (1 \ 0 \ 1 \ 0 \ \dots \ 0) \alpha_t + \varepsilon_t .$$

Die Wahl der Startwerte für den Ausgangsvektor α_0 und die zugehörige Kovarianzmatrix P_0 sind durchaus kritisch. Hier wurde der Anfangslevel in der Nähe der ersten Zeitreihenwerte gewählt; die anderen Startparameter sind null gesetzt. Die Varianzen für die Niveauparameter werden klein gewählt. Werden sie zu groß vorgegeben, so wird die Trendschätzung zu instabil, sie inkorporiert dann auch

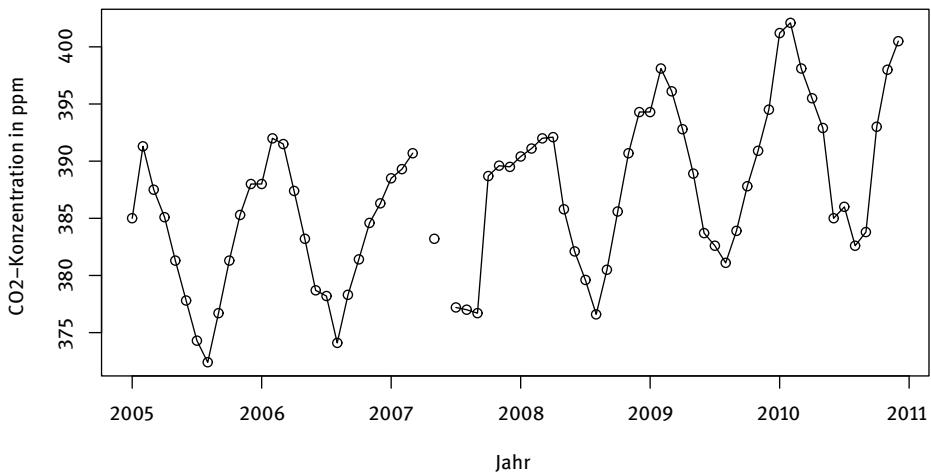


Abb. 10.1: CO₂-Konzentration über Schauinsland

die Saison.

$$\alpha_0 = (382, 0.1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)' ,$$

$$\mathbf{P}_0 = \text{diag}(0.1, 0.1, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100).$$

Für die mit der ML-Methode geschätzte Kovarianzmatrix \mathbf{Q}_t und die Varianz σ_ϵ^2 ergeben sich damit

$$\widehat{\mathbf{Q}}_t = 10^{-8} \cdot \text{diag}(121.88, 39523.52, 0.0000017), \quad \hat{\sigma}_{\varepsilon}^2 = 2.68.$$

Das Ergebnis der Filterung und Glättung der Reihe unter Verwendung der ermittelten Parameter ist in der Abbildung 10.2 dargestellt.

Schließlich wird der geglättete Verlauf ein Jahr in die Zukunft prognostiziert. In der Abbildung 10.3 ist der prognostizierte Verlauf mit den 95 % Prognosegrenzen dargestellt.

R-Code:

```

1 library(dlm)
2 data(SCHAUINSLAND)
3 y <- SCHAUINSLAND
4 dlmco <- dlmModPoly(2) + dlmModSeas(12)
5 m1 <- c(382,0.1,rep(0,11))
6 c1 <- diag(c( 0.1, 0.1, rep(100,11)))
7 buildFun <- function(x) {
8   W(dlmco)[1:3,1:3] <- diag(exp(x[1:3]))
9   V(dlmco) <- exp(x[4])

```

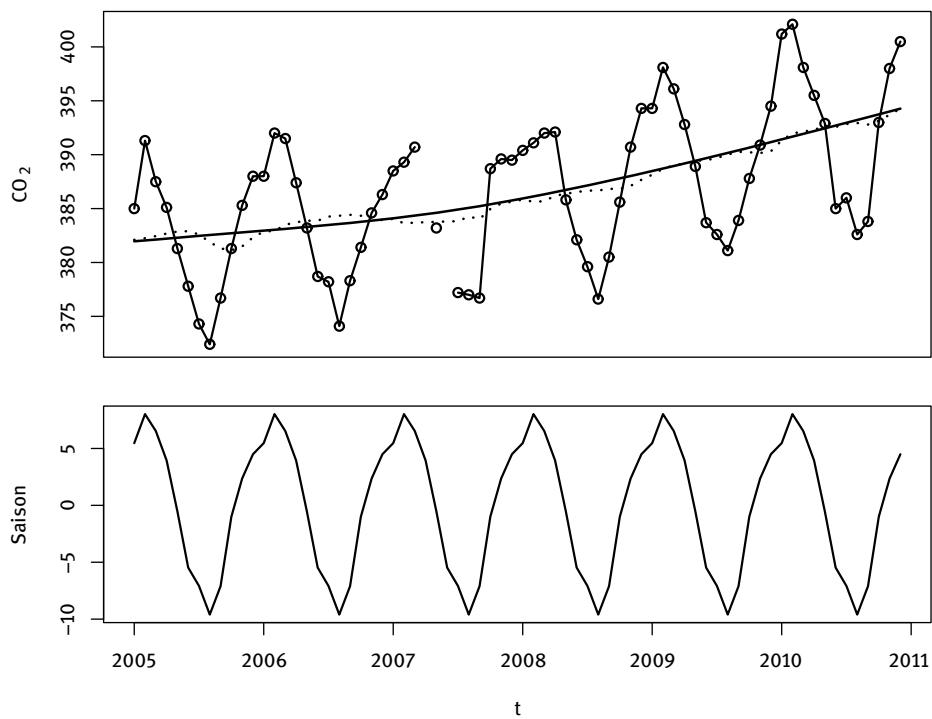


Abb. 10.2: CO₂-Konzentration über Schauinsland mit gefilterter (- - -) und geglätteter Trendlinie (—) sowie geglättete Saisonkomponente

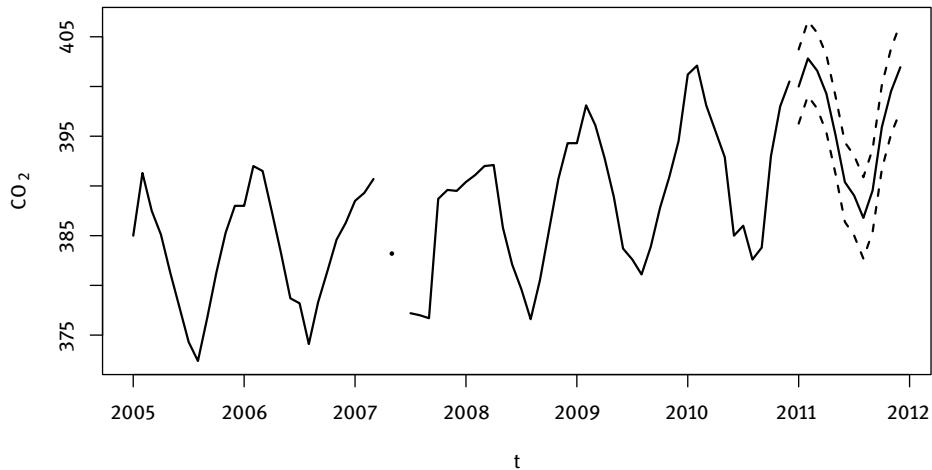


Abb. 10.3: CO₂-Konzentration über Schauinsland mit Prognosen

```

10   C0(dlmco) <- c1
11   m0(dlmco) <- m1
12   return(dlmco) }
13 fit <- dlmMLE(y, parm=rep(1,4), build=buildFun)
14 fit$conv
15 dlm.y <- buildFun(fit$par)
16 coFilter <- dlmFilter(y, mod=dlm.y)
17 coSmooth <- dlmSmooth(coFilter)
18 plot(y,type="o")
19 lines(dropFirst(coFilter$m[,1]))
20 lines(dropFirst(coSmooth$s[,1]))
21 plot(dropFirst(coSmooth$s[,3]))
22 fut1 <- dlmForecast(coFilter, n=12)
23 x <- cbind(y,fut1$f,fut1$f-1.96*sqrt(unlist(fut1$Q)),
24           fut1$f+1.96*sqrt(unlist(fut1$Q)))
25 plot.ts(x,plot.type="s",lty=c(1,1,2,2))
26 points(2005+28/12,y[29])

```

Für die Analyse wird das Paket `dlm` verwendet. Dies wird zuerst geladen. Dann wird die Zeitreihe aktiviert. Mit der Zeile 4 werden die Matrizen und Vektoren des Zustandsraummodells erzeugt. Dies ist für Standardmodelle extrem einfach. Man kann aber auch andere detailliert konstruieren. Auf 5 und 6 werden die Anfangswerte von α_0 und P_0 gesetzt. Die dann in den Zeilen 7–12 definierte Funktion wird für `dlmMLE` benötigt. Mit dem Aufruf wird die Schätzung durchgeführt. Es ist immer sinnvoll nachzusehen, ob die iterative Schätzung konvergiert ist. (Zeile 14 muss eine 0 ergeben.) Die ermittelten Schätzwerte werden dann in das Modell eingepflegt. So kann die Reihe gefiltert und geglättet werden. Die für die Darstellung benötigten Komponenten sind aus dem Output herauszupicken. Die Prognose ist ebenfalls einfach, die Aufbereitung für die grafische Darstellung aufwendiger.

Die Diskussion des Kalman-Filters in diesem Kapitel berührt praktisch nur die Oberfläche. Ausführliche Darstellungen zum Kalman-Filter sind etwa Anderson and Moore (1979) sowie Catlin (1989). Das Buch von Aoki (1990) ist ganz dem Zustandsraummodell gewidmet. Hier wird der Rahmen für einen Algorithmus aufgezeigt, der gewissen Besonderheiten ökonomischer Zeitreihen Rechnung trägt. Der Sammelband Aoki und Havenner (1997) stellt Anwendungen dieses Ansatzes vor. Die anwendungsorientierten Bücher Pole, West & Harrison (1994) sowie West & Harrison (1997) sind gänzlich dem Zustandsraummodell gewidmet. Die Autoren nehmen dabei eine Bayessche Perspektive ein. Ein neueres empfehlenswertes Buch zu den Zustandsraummodellen ist Durbin & Koopman (2001).

10.3 R-Funktionen

Das Paket `stats` bietet verschiedene Funktionen für Zustandsraummodelle. Die Funktion `StructTS` passt ein strukturelles Modell an eine Zeitreihe an. Dabei wird eine Maximum-Likelihood-Schätzung durchgeführt. Eine feste-Intervall-Glättung einer univariaten Zeitreihe führt die Funktion `tsSmooth` über die Anpassung eines Zustandsraummodells durch.

Die genannten Funktionen basieren auf Umsetzungen der Kalman-Filterung. Damit können die Gaußsche Loglikelihood bestimmt werden, Prognosen erstellt oder Zeitreihen geglättet werden.

Mit den Funktionen des Paketes `dynlm` stehen weitere Möglichkeiten bereit, dynamische lineare Modelle in R zu spezifizieren und anzupassen. Die Funktion `dynlm` ist weitgehend gleichwertig zu der Funktion `lm`, jedoch erhält sie die Zeitreihenattribute und erlaubt eine Instrumentalvariablen-Regression mittels der zweistufigen KQ-Methode. Fortgeschrittene, auch multivariate dynamische lineare Modelle können mit dem Paket `dse` umgesetzt werden.

Petris, Petrone & Campagnoli (2009) geben eine ausführliche Beschreibung der Möglichkeiten des Paketes `dlm`. Hier sei nur darauf hingewiesen, dass das Paket verschiedene Funktionen enthält, um dynamische Lineare Modelle zu erzeugen. Die generelle Funktion hierfür ist `dlm`. Eine Kalman-Filterung wird durch den Aufruf der Funktion `dlmFilter` durchgeführt. Dabei werden die Einschritt-Prognosen automatisch mit erstellt. Weiterreichende Prognosen erstellt die Funktion `dlmForecast`.

Ein weiteres Paket für Zustandsraummodellierung ist `MARSS`. Es ist für multivariate Zeitreihen entwickelt worden; siehe dazu Holmes, Ward & Scheuerell (2018).

11 Nichtlineare Modelle

11.1 Nichtlinearität in Zeitreihen

Nichtlineares bedingtes Niveau

Trotz des großen Anwendungsspektrums der ARMA-Modelle werden in der Praxis vielfach Zeitreihen erhoben, für die diese Modellklasse augenscheinlich nicht passt. Einmal weisen empirische Beobachtungen häufig Strukturbrüche auf. Ein geeignetes Modell sollte diese erklären, um realistische Vorhersagen zu treffen. Zudem sollte eine mittels eines Gaußschen ARMA-Prozesses modellierbare Reihe eine Symmetrie bei den An- und Abstiegen aufweisen. Um dies erkennbar zu machen, sollte bei der Darstellung das Aspekt-Verhältnis a von Höhe zu Breite des Plots so gewählt werden, dass die Steigungen der Verbindungslien von je zwei aufeinander folgenden Zeitreihenwerten sich an der 45°- bzw. der 135°-Achse orientieren, vgl. den Abschnitt 2.1.

Den Hintergrund für diese Aussage bildet die *Zeitreversibilität* von Gaußschen ARMA-Prozessen $\alpha(B)Y_t = \beta(B)\varepsilon_t$. Zeitreversibel bedeutet dabei, dass der vorwärtslaufende Prozess

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

nicht nur die gleiche Kovarianzstruktur hat wie der rückwärts laufende Prozess

$$Y_t - \alpha_1 Y_{t+1} - \cdots - \alpha_p Y_{t+p} = \eta_t - \beta_1 \eta_{t+1} - \cdots - \beta_q \eta_{t+q},$$

sondern dass die Störungen η_t auch verteilt sind wie die ε_t .

Wenn also die Plots der Ausgangsreihe und der zeitlich umgedrehten Reihe nicht das gleiche Bild liefern, so kann vermutet werden, dass ein lineares Modell nicht geeignet ist, die Reihe zu erfassen.

Beispiel 11.1 (Anzahlen von Luchspelzen – Fortsetzung von Seite 137)

Die Reihe LUCHS weist langsamere Anstiege als Abstiege auf. Das macht z. B. die Abbildung 11.1 deutlich, bei der die vorwärts laufende Reihe sowie die zeitlich umgekehrte Reihe dargestellt sind. Das stellt die Eignung eines AR-Modells in Frage.

Einer der Gründe, nichtlineare Zeitreihenmodelle zu betrachten, liegt in der Hoffnung verbesserter Prognosen. Da Prognosen mit ARIMA-Modellen gut analysiert und auch erfolgreich sind, liegen Erweiterungen der linearen ARMA-Prozesse nahe. Eine Klasse solcher Erweiterungen lässt zu, dass in verschiedenen Zeitabschnitten unterschied-

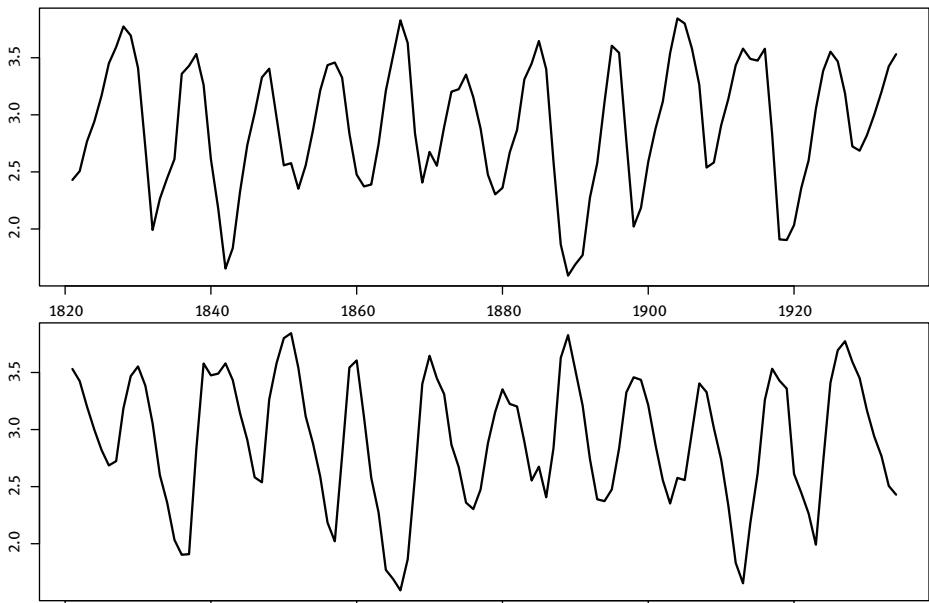


Abb. 11.1: Reihe LUCHS und die zeitlich umgekehrte Reihe

liche Regimes gelten. Von diesen werden hier nur die Markov-switching-Prozesse genauer betrachtet, da sie als einzige eine stärkere Anwendungsbreite erreicht haben.

Nichtlineare bedingte Streuung

Ein anderes Phänomen, das zu nichtlinearen Modellen führt, ist das Vorhandensein von Abschnitten unterschiedlicher Streuung. Solche Reihen werden oft bei spekulativen Märkten, etwa bei Aktien- und Wechselkursen, beobachtet. In ein Modell lässt sich eine zeitlich variierende Streuung dadurch einbauen, dass die Varianz von vergangenen Werten abhängig gemacht wird. Damit wird man zur bedingten Varianz geführt. Da bei linearen Prozessen die bedingte Varianz stets konstant ist, kann eine sich zeitlich ändernde Streuung mit ihnen nicht erfasst werden. Allgemein interessiert man sich bei Finanzmarktreihen für die Entwicklung der bedingten Varianz und nicht für den bedingten Erwartungswert.

Beispiel 11.2 (Kurse der IBM-Aktie – Fortsetzung von Seite 170)

Wie im Beispiel 6.34 angemerkt wurde, weist der zweite Teil der Reihe der Differenzen eine erheblich höhere Variabilität auf als der erste, siehe die Abbildung 6.24. Verschiedene Autoren, die diese Zeitreihe als Beispiel verwenden, teilen deshalb

den Datensatz und modellieren beide Teile getrennt. Will man ein geschlossenes Modell für die gesamte Reihe angeben, muss dieses die Unterschiedlichkeit der Streuung erfassen.

Spezifikation nichtlinearer Zeitreihenmodelle

Bei der Spezifikation nichtlinearer Zeitreihen wird von Granger, einem der Pioniere der Beschäftigung mit nichtlinearen Zeitreihenmodellen, ein mehrschrittiges Vorgehen empfohlen, siehe Granger & Teräsvirta (1993):

1. Spezifizierte ein lineares AR-Modell für die zu modellierende Zeitreihe.
2. Teste die Nullhypothese der Linearität gegen den Typ des nichtlinearen Modells.
3. Schätze die Parameter des nichtlinearen Modells.
4. Evaluiere das Modell mittels diagnostischer Tests.
5. Modifizierte das Modell, wenn nötig.
6. Verwende das Modell zur Beschreibung oder zu Prognosen.

Mit der Spezifizierung eines linearen AR-Modells zu beginnen ist ratsam, da die Anpassung nichtlinearer Modelle um einiges aufwendiger ist als die eines AR-Prozesses. Nur wenn ein Test es dann angeraten sein lässt, ein solches Modell zu verwenden, sollte man das tun. Allerdings setzen manche Tests die Anpassung des nichtlinearen Modells voraus.

11.2 Markov-switching Modelle

Markov-switching-Modelle oder auch Markov-regime-switching Modelle sind nichtlineare Zeitreihenmodelle. Diese Modelle werden seit langem in der Elektrotechnik und in den Wirtschaftswissenschaften eingesetzt. Wirtschaftswissenschaftliche Anwendungen finden sich etwa bei Finanz- und makroökonomischen Zeitreihen, die abrupte Verhaltensänderungen aufweisen. Gründe dafür können bspw. Änderungen der gesetzlichen und politischen Rahmenbedingungen sein, aber auch Finanzkrisen und Bubbles. Bei Konjunkturzyklen stellen die Übergänge von Hause- und Baisse-Phasen solche Anlässe zu Regime-Wechseln dar.

Ein einfaches Beispiel stand am Anfang des Interesses im ökonomischen Anwendungsbereich. Das von mexikanischen Banken gehaltene Dollarvermögen im Verhältnis zu ihrem Pesovermögen weist in 1984 einen abrupten Einbruch in 1982 auf, um dann auf geringerem Niveau genau so weiterzugehen wie zuvor, vgl. die Abbildung in Hamilton (1994, S. 678). Das zugehörige Modell ist ein AR-Prozess, bei dem das Niveau von einem nicht beobachtbaren Prozess abhängt, dessen Variablen nur zwei unterschiedliche Werte annehmen können. Dieser Prozess wird als *Markov-Kette* unterstellt.

Markov-Ketten

Definition 11.3 (Markov-Kette)

Eine *Markov-Kette* (S_t) ist ein stochastischer Prozess, bei denen die Zufallsvariablen S_t nur endlich viele Werte oder *Zustände* annehmen können. Diese Zustände werden einfach mit $1, \dots, k$ bezeichnet. Eine Markov-Kette ist dadurch charakterisiert, dass die bedingten Wahrscheinlichkeiten

$$P(S_t = j \mid S_{t-1} = i) = p_{ij}, \quad (11.1)$$

die als *Übergangswahrscheinlichkeiten* bezeichnet werden, unabhängig von der weiter zurückliegenden Vergangenheit sind:

$$P(S_t = j \mid S_{t-1} = i_1) = P(S_t = j \mid S_{t-1} = i_1, S_{t-2} = i_2, S_{t-3} = i_3, \dots). \quad (11.2)$$

(Genauer ist dies eine *homogene* Markov-Kette erster Ordnung. Bei nicht-homogenen hängen die Übergangswahrscheinlichkeiten von der Zeit ab.)

Die Eigenschaft, dass sich die gesamte Information der Vergangenheit im letzten Zeitpunkt kondensiert, ist auch bei AR[1]-Prozessen gegeben. Sie haben also auch die Markov-Eigenschaft. Das Besondere bei Markov-Ketten ist aber, dass die Prozessvariablen diskret sind.

Für die Matrix der Übergangswahrscheinlichkeiten

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \vdots & \vdots & & \vdots \\ p_{k1} & p_{12} & \dots & p_{kk} \end{pmatrix} \quad (11.3)$$

gilt, dass alle Zeilensummen eins sind, $p_{i1} + p_{i2} + \dots + p_{ik} = 1$. Das bewirkt speziell, dass ein Eigenwert von \mathbf{P} gleich eins ist:

$$(\mathbf{I} - 1 \cdot \mathbf{P})\mathbf{1} = \mathbf{0}, \quad (11.4)$$

wobei $\mathbf{1}$ ein Spaltenvektor von Einsen und $\mathbf{0}$ einer von Nullen ist.

Eine Markov-Kette heißt *nicht-reduzierbar*, wenn aus keinem Zustand mit Wahrscheinlichkeit eins der Zustand des nächsten Zeitpunktes folgt, wenn $p_{ij} < 1$ für alle $i, j = 1, \dots, k$.

Eine nicht-reduzierbare Markov-Kette ist *ergodisch*, falls ein Eigenwert von \mathbf{P} eins ist und alle anderen Eigenwerte innerhalb des Einheitskreises liegen. Bei ergodischen Markov-Ketten existieren unbedingte Wahrscheinlichkeiten $P(S_t = 1), \dots, P(S_t = k)$.

Durch die Konstruktionsweise ist eine ergodische Markov-Kette trotz der Einheitswurzel ein stationärer Prozess. Allerdings sind die üblichen Kennfunktionen in diesem Zusammenhang weniger von Interesse.

Markov-switching autoregressive Prozesse

Es wird zunächst nur das Niveau in Abhängigkeit von dem Zustand der Markov-Kette modelliert. Schon hier gibt es unterschiedliche Möglichkeiten. Bei dem Modell

$$Y_t - \mu_{S_t} = \alpha_1(Y_{t_1} - \mu_{S_{t-1}}) + \cdots + \alpha_p(Y_{t_p} - \mu_{S_{t-p}}) + \varepsilon_t \quad (11.5)$$

verhalten sich die Zeitpfade anders als bei

$$Y_t = \nu_{S_t} + \alpha_1 Y_{t_1} + \cdots + \alpha_p Y_{t_p} + \varepsilon_t. \quad (11.6)$$

Während im ersten Modell ein Level-Shift sich unmittelbar auf das Niveau der Reihe auswirkt, wird das neue Niveau bei dem zweiten erst allmählich angenommen.

Beispiel 11.4 (Zwei simulierte Markov-switching-Prozesse)

Die beiden Möglichkeiten, einen Level-Shift einzubeziehen, wurden in einem Simulationsexperiment umgesetzt. Es liegt ein AR[1]-Modell mit $\alpha = 0.8$ und standardnormalverteilten Innovationen zugrunde. Die beiden Niveaus sind $\mu_1 = 0$ und $\mu_2 = 10$, die Übergangsmatrix ist $P = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$. Die Abbildung 11.2 oben zeigt die gemäß (11.5) simulierte Reihe und unten die mittels (11.6) simulierte.

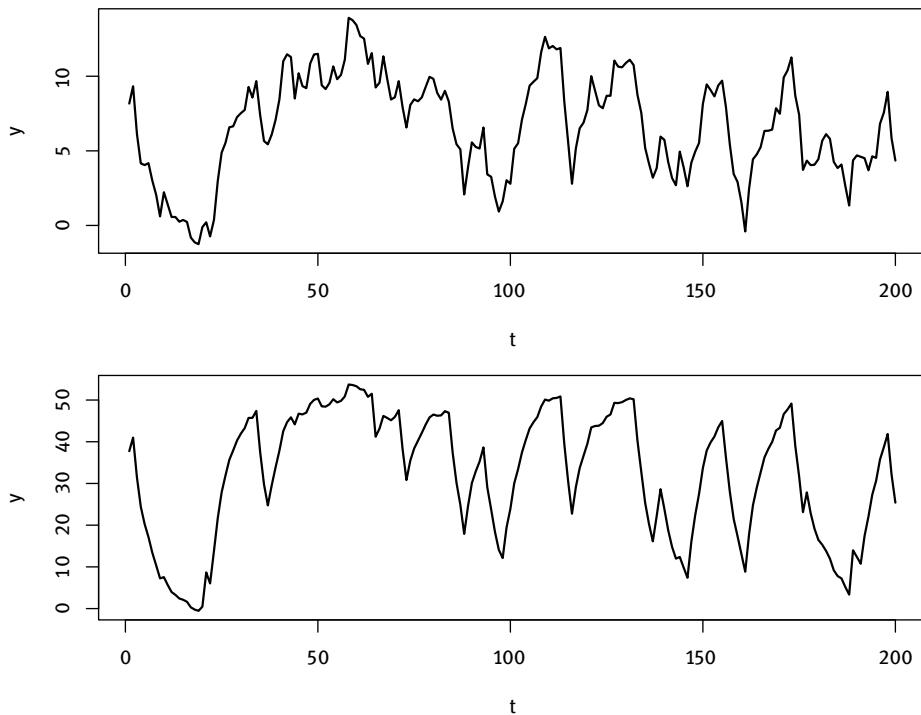


Abb. 11.2: Zwei Markov-switching-Prozesse

R-Code:

```

1 P <- matrix(c(0.8,0.2,0.1,0.9),2,2,byrow=TRUE)
2 e <- rnorm(250)
3 S <- runif(250)
4 y1 <- rep(1,250)
5 y2 <- rep(1,250)
6 m1 <- 0
7 m2 <- 10
8 S[1] <- 1*(S[1]<=0.5) + 2*(S[1]>0.5)
9 for (t in c(2:250)){
10   if (S[t-1] == 1){ S[t] <- 1*(S[t]<=P[1,1])+2*(S[t]>P[1,1]) }
11   if (S[t-1] == 2){ S[t] <- 1*(S[t]<=P[2,1])+2*(S[t]>P[2,1]) }
12   if (S[t] == 1){y1[t] <- m1+0.8*(y1[t-1]-m1)+e[t]
13     y2[t] <- m1+0.8*y2[t-1]+ e[t] }
14   if (S[t] == 2){y1[t] <- m2+0.8*(y1[t-1]-m2)+e[t]
15     y2[t] <- m2+0.8*y2[t-1]+e[t] }
16 }
17 ts.plot(y1[51:250])
18 ts.plot(y2[51:250])

```

Der Code setzt die Definitionen direkt um. Je nachdem welchen Zustand S zum Zeitpunkt $t - 1$ hatte, wird der neue Zustand gemäß der zugehörigen Wahrscheinlichkeitsverteilung aus der Matrix P erzeugt. Dies bestimmt dann das Regime für den AR[1]-Prozess. Verwendet werden nur die letzten 200 Werte, die ersten werden verworfen.

In ihrer allgemeinen Form werden die Modelle wie folgt definiert.

Definition 11.5 (Markov-switching autoregressive Prozesse)

Ein Prozess, der der Gleichung

$$Y_t = \nu_{S_t} + \alpha_{S_t,1} Y_{t_1} + \cdots + \alpha_{S_t,p} Y_{t_p} + \varepsilon_t . \quad (11.7)$$

genügt, bei der S_t eine Markov-Kette erster Ordnung ist, heißt *Markov-switching autoregressiver Prozess der Ordnung p, MSAR[p]-Prozess*.

Die Stationarität von (Y_t) hängt sowohl von den Übergangswahrscheinlichkeiten p_{ij} als auch von den Koeffizienten des AR-Prozesses ab, sowie von der Existenz der ergodischen Wahrscheinlichkeiten $P(S_t = 1), \dots, P(S_t = k)$. Es gibt also einen gewissen Trade-off zwischen Koeffizienten und Übergangswahrscheinlichkeiten. Das hat den Vorteil, dass ein asymmetrisches Verhalten innerhalb eines Regimes nicht notwendigerweise die Stationarität des ganzen Prozesses zerstört. Der Nachteil liegt darin, dass die notwendige Stationaritätsbedingung für komplexere Systeme schwierig zu berechnen ist. Meist beschränkt man sich daher auf hinreichende Bedingungen.

Das Hauptinteresse bei diesen Modellen besteht in der Bestimmung der Zeitpunkte, an denen ein Regimewechsel stattfindet. Dabei wird i. d. R. von nur wenigen Regimes oder Zuständen ausgegangen, meist sind es nur zwei. Somit können etwa Wendepunkte in der Konjunktur eines Landes bestimmt werden.

Inferenz

Die zu schätzenden Modellparameter $\boldsymbol{\theta}$ hängen nicht-linear von nicht beobachtbaren Regimen ab. Daher ist ein numerisches Verfahren zu Bestimmung der Maximum-Likelihood-Schätzer unabdingbar. Zudem sind auch die nicht beobachtbaren Regimes zu berücksichtigen.

Einen Lösungsansatz bietet der *Expectation-Maximization-Algorithmus*, kurz *EM-Algorithmus*. Sei zur Illustration das MSAR[p]-Modell mit zwei Zuständen gegeben:

$$Y_t = \begin{cases} \alpha_{1,0} + \alpha_{1,1}y_{t-1} + \dots + \alpha_{1,p}y_{t-p} + \varepsilon_t & \text{falls } S_t = 1, \\ \alpha_{2,0} + \alpha_{2,1}y_{t-1} + \dots + \alpha_{2,p}y_{t-p} + \varepsilon_t & \text{falls } S_t = 2, \end{cases}$$

oder, kurz

$$Y_t = \alpha_{S_t,0} + \alpha_{S_t,1}y_{t-1} + \dots + \alpha_{S_t,p}y_{t-p} + \varepsilon_t.$$

Bei Normalverteilung der Störungen gilt für die bedingte Dichte von Y_t gegeben die vorangegangenen Beobachtungen Ω_{t-1} und Zustände:

$$p(y_t | S_t = j, \Omega_{t-1}, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(\frac{-(y_t - \boldsymbol{\alpha}_j' \mathbf{x}_t)^2}{2\sigma^2}\right), \quad (11.8)$$

wobei $\mathbf{x}_t = (1, y_{t-1}, \dots, y_{t-p})'$ und der Vektor $\boldsymbol{\theta}$ alle Parameter des Modells enthält, also die AR-Koeffizienten, die Varianzen und die Übergangswahrscheinlichkeiten. Daraus ergibt sich die bez. S_t unbedingte Dichte von Y_t zu

$$p(y_t | \Omega_{t-1}, \boldsymbol{\theta}) = \sum_{j=1}^2 p(y_t | \Omega_{t-1}, \boldsymbol{\theta}) P(S_t = j | \Omega_{t-1}, \boldsymbol{\theta}). \quad (11.9)$$

Im *Erwartungswert*-Schritt werden die nicht beobachtbaren Zustände s_t durch ihre Wahrscheinlichkeiten geschätzt. Die Schätzungen der Wahrscheinlichkeiten müssen allerdings noch geglättet werden, da die einfachen Schätzungen zu erratisch über die Zeit sind. Im *Maximierungs*-Schritt werden die Modellparameter mit ML-Methode geschätzt. Dabei werden die für die Zustände die Schätzungen aus dem E-Schritt verwendet. Beide Schritte werden jeweils iteriert, bis ein geeignetes Zielkriterium erfüllt ist.

Die Überprüfung der Parameter kann unter geeigneten Voraussetzungen wie üblich erfolgen. Man wird also die t -Statistiken zu Rate ziehen, ob AR-Koeffizienten bedeutsam sind.

Von Interesse ist auch ein Test auf die Anzahl der Regimes. Als naheliegendes Testproblem ergibt sich die Hypothese H_0 : „Es liegen K Regimes zugrunde.“, die gegen

die Alternative H_1 : „Es liegen $K + 1$ Regimes zugrunde.“ zu testen ist. Hier sind aber unter H_0 einige der Parameter, die unter H_1 auftauchen, nicht definiert. Das hat zur Folge, dass der Likelihood-Ratio Test nicht χ^2 -verteilt ist. Als Lösung bietet sich an, einen modifizierten Likelihood-Ratio Test einzusetzen. Die Bestimmung der kritischen Werte muss dann allerdings mittels Simulation erfolgen. Dieser Test ist für $K = 1$ auch ein diagnostischer Test auf die Notwendigkeit überhaupt ein Markov-switching Modell anstelle eines reinen AR-Modells zu verwenden.

Alternativ kann auf Tests zurückgegriffen werden, die nicht auf der Likelihood-Ratio-Statistik beruhen. Auch können die Modelle über ihre Prognosekraft verglichen werden. Hierzu sei auf Hamilton (1994) und Krolzig (1997) verwiesen.

Die üblichen Werkzeuge der Modellüberprüfung bei ARMA-Modellen, wie ACF und PACF der Residuen und der Portmanteau-Test zur Überprüfung der Anpassung, stehen hier nicht zur Verfügung. Da die Verteilungstheorie nicht geklärt ist, kann ihre Verwendung höchstens deskriptiven Charakter haben.

Beispiel 11.6 (Bruttoinlandsprodukt)

Aus dem realen Bruttoinlandsprodukt wurde mit einem Baxter-King-Filter die zyklische Komponente extrahiert, siehe die Abbildung 11.3. Der resultierenden Zeitreihe wurde ein Markov-switching-AR[4]-Modell mit zwei Zuständen angepasst.

```
Coefficients: Regime 1
-----
            Estimate Std. Error t value Pr(>|t|)    
L(S)        0.0171    0.0298   0.5738  0.5661    
y2_1(S)     2.6074    0.1111  23.4689 < 2.2e-16 *** 
y2_2(S)    -3.1174    0.2634 -11.8352 < 2.2e-16 *** 
y2_3(S)     2.0606    0.2754   7.4822 7.305e-14 *** 
y2_4(S)    -0.6733    0.1205  -5.5876 2.302e-08 *** 
---
Residual standard error: 0.1293528

Coefficients: Regime 2
-----
            Estimate Std. Error t value Pr(>|t|)    
L(S)        -0.0286    0.0157  -1.8217  0.0685 .      
y2_1(S)     2.3845    0.1365  17.4689 <2e-16 *** 
y2_2(S)    -2.8514    0.2347 -12.1491 <2e-16 *** 
y2_3(S)     2.1627    0.1762  12.2741 <2e-16 *** 
y2_4(S)    -0.7719    0.0776  -9.9472 <2e-16 *** 
---
Residual standard error: 0.08979698
```

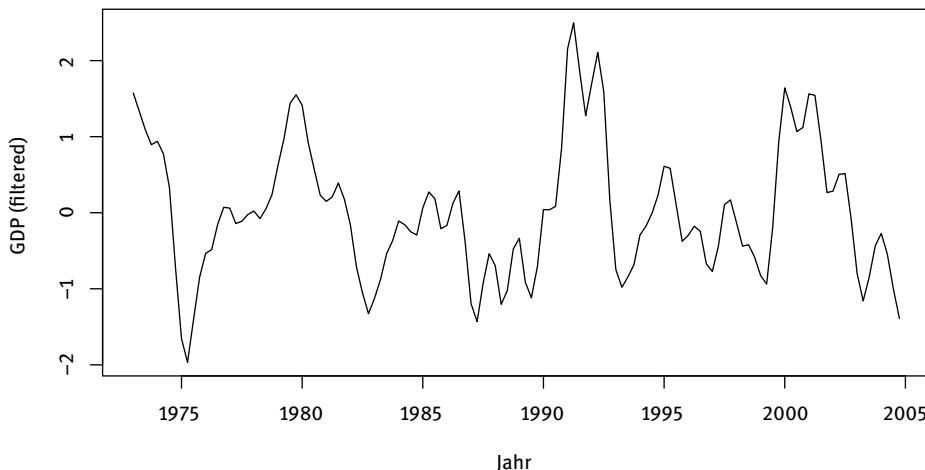


Abb. 11.3: Zyklische Komponente des BIP

```
Transition probabilities:
    Regime 1   Regime 2
Regime 1  0.8806047  0.118772
Regime 2  0.1193953  0.881228
```

Die in der Abbildung 11.4 dargestellten geschätzten Zustandswahrscheinlichkeiten zeigen, wie wahrscheinlich sich die Konjunktur in einer Hochphase bzw. in keiner Hochphase befand.

Die Ordnung 4 des AR-Modells ist offensichtlich notwendig, wie die Signifikanz der Koeffizientenschätzungen beweist. Die abschließenden diagnostischen Plots der Residuen zeigen, dass das Regime 1 mit einem AR[4]-Modell nicht recht zufriedenstellend erfasst wird. Eher ist ein ARMA-Modell von Nöten. Die Normalverteilung ist ausreichend gut gegeben.

R-Code:

```
1 data(GDP)
2 y <- GDP
3 y1 <- window(y,start=c(1970,1),end=c(2007,4))
4 library(mFilter)
5 out <- bkfilter(y1,pl=4.4,pu=35,nfix=12,drift=FALSE)
6 plot(out)
7 y2 <- na.omit(out$cycle)
8 library(MSwM)
9 L <- rep(1,length(y2))
10 mod <- lm(y2 ~ L - 1)
```

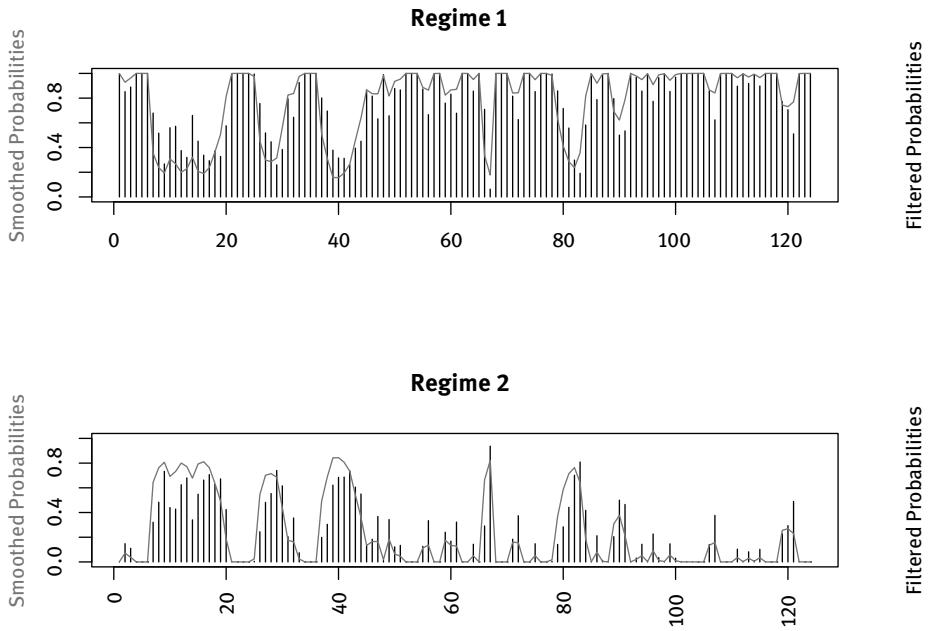


Abb. 11.4: Wahrscheinlichkeiten für $S_t = 1$ bzw. $S_t = 2$

```

11 out <- msmFit(mod, k=2, sw=c(T,T,T,T,T,T), p=4,
                  control = list(maxiter=5000))
12 summary(out)
13 plotProb(out,1)
14 plotDiag(out,1,3)
15 plotDiag(out,2,3)

```

Zuerst wird die Reihe aktiviert, der gewünschte Zeitabschnitt wird selektiert. Für die Filterung mit dem Baxter-King-Filter muss das Paket `mFilter` geladen werden. Die Enden der gefilterten Reihe sind mit NAs aufgefüllt; diese werden entfernt. Die Anpassung erfolgt mit der Funktion `msmFit` des Paketes `MSwM`. Diese benötigt als Argument das Ergebnis der Anpassung eines linearen Modells. Daher wird der Mittelwert in der angegebenen Form mittels `lm` geschätzt. Um stabile Ergebnisse zu erhalten sind wesentlich mehr Iterationen des EM-Algorithmus nötig, als die 100 der Voreinstellung. Die abschließenden diagnostischen Grafiken sind hier nur mittels der Zeilen 14 und 15 angedeutet.

Bei Frühwirt-Schnatter (2006, S. 366 f.) findet sich eine Diskussion der Anpassungen der GDP-(Bruttoinlandsprodukt-)Reihe der USA bei verschiedenen Autoren. Dort wird deutlich, wie sensitiv die verschiedenen Beurteilungsmaße gegen die gewählte AR-Ordnung und die Anzahl der Zustände der Markov-Kette sind. Das Spezifikationsproblem ist bei weitem nicht gelöst.

Während Markov-switching-Modelle den linearen Zeitreihenmodellen darin überlegen sind, gewisse Phänomene in Zeitreihen zu erfassen, hat sich eine Verbesserung bez. der Prognosen mit solchen Modellen in empirischen Untersuchungen nicht erkennen lassen. AR-Modelle sind doch recht robust gegen Abweichungen von den Modellannahmen, wenn es um die Vorhersageproblematik geht, siehe etwa Clements & Krolzig (1998).

11.3 Threshold-Modelle

Beim so genannten *Threshold autoregressiven Modell der Ordnung p*, kurz: *TAR[p]-Modell*, handelt es sich um eine andere Möglichkeit, unterschiedliche AR-Modelle zu kombinieren. Regimewechsel werden hier durch Vergangenheitswerte der Prozessvariablen induziert:

$$Y_t = \begin{cases} \alpha_0 + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t, & \text{falls } Y_{t-d} < r \\ \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + \varepsilon_t, & \text{falls } Y_{t-d} \geq r. \end{cases} \quad (11.10)$$

Dabei wird d als Verzögerungsparameter und r als Schwellenwert (engl. threshold) bezeichnet. Die TAR-Prozesse werden auch als *SETAR-Prozesse* bezeichnet, von ‚Self Exciting Threshold Autoregressiv‘.

Der Modellkonstruktion liegt die Beobachtung zugrunde, dass dynamische Systeme bisweilen ihr Verhalten ändern, sobald ein bestimmter Schwellenwert über- oder unterschritten wird. Als Resultat eines solchen Verhaltens können sich z. B. asymmetrische Zyklen ergeben. Im physikalischen Kontext handelt es sich bei Grenzzyklen um einen stationären Zustand von andauernden Oszillationen. Dabei sind die Zyklen oft asymmetrisch, d. h. sie zeigen ein anderes Auf- als Abschwungverhalten. TAR-Modelle erfassen nun asymmetrische Zyklen bei Zeitreihen mit periodischem Verhalten, indem sie beim Erreichen des Gipfels der Aufschwungphase, z. B. eines Konjunkturzyklus, auf ein anderes autoregressives Modell umschalten.

Beispiel 11.7 (IBM-Aktienkursreihe – Fortsetzung von Seite 264)

Tong (1990) deutet an, dass die ersten 219 Werte der Differenzen der IBM-Aktienkursreihe durch ein TAR-Modell beschrieben werden könnten. Eine Neuschätzung ergibt das Modell:

$$(1 - B)Y_t = \begin{cases} -0.565 - 0.416 Y_{t-2} - 0.187 Y_{t-3} + 0.183 Y_{t-4} \\ \quad + 0.355 Y_{t-5} - 0.667 Y_{t-6} + 0.322 Y_{t-7} + \varepsilon_t & \text{für } Y_{t-1} \leq -4 \\ -0.700 + 0.467 Y_{t-1} + 0.166 Y_{t-4} + 0.072 Y_{t-6} \\ \quad - 0.132 Y_{t-9} + \varepsilon_t & \text{für } Y_{t-1} > -4 \end{cases} \quad (11.11)$$

An diese Aktienkursreihe wurden auch andere nichtlineare Modelle angepasst.

Für TAR-Modelle lassen sich Prognosen wie für AR-Prozesse durch Fortschreiben der Rekursion bei Nullsetzen zukünftiger Innovationen erstellen. Prognoseintervalle sind nicht geschlossen angebar, da die Prognosefehler $\hat{Y}_{t,h} - Y_{t+h}$ für $h > 1$ auch bei normalverteilten Innovationen nicht mehr normalverteilt sind.

Der TAR-Ansatz lässt sich auf verschiedene, nahe liegende Weisen verallgemeinern. Zunächst kann die Zahl der Schwellenparameter erhöht werden. So sind in der Abbildung 11.5 zwei Schwellen symmetrisch zu Null gegeben (± 2); im zentralen Bereich liegt der nicht-stationäre AR[1]-Prozess $Y_t = 2Y_{t-1} + \varepsilon_t$ vor, in den Randbereichen jeweils ein White-Noise-Prozess (ε_t). Es ist jeweils $\varepsilon_t \sim N(0, 0.1^2)$. In der Praxis werden selten allgemeinere TAR-Modelle als mit einem oder zwei Schwellenwerten verwendet. Ein Grund ist schon darin zu sehen, dass in empirischen Zeitreihen meistens nicht genügend Beobachtungen vorliegen, um komplexere Modelle schätzen zu können. Aber auch der Rechenaufwand steigt mit der Zahl der zugelassenen Schwellen beträchtlich.

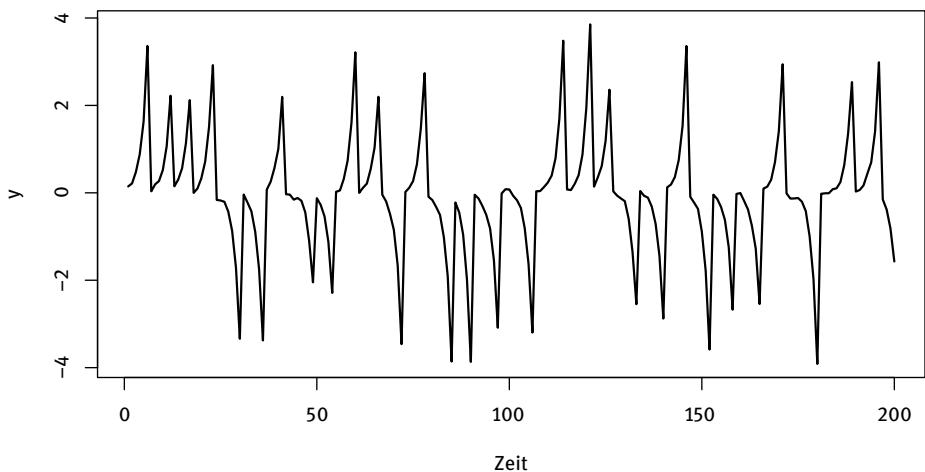


Abb. 11.5: Realisierung eines TAR-Prozesses (siehe Text)

Eine weitere Verallgemeinerung der TAR-Modelle wird nahe gelegt, wenn die Ausgangsgleichung der TAR-Modelle unter Verwendung einer Indikatorfunktion $I_r(y) = 0$ für $y \leq r$ und $= 1$ für $y > r$ in folgender Form geschrieben wird:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + (\beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p}) \cdot I_r(Y_{t-d}) + \varepsilon_t .$$

Nun ist es plausibel, I_r durch eine glatte Funktion F mit Werten im Intervall $[0, 1]$ zu ersetzen. Für diese Funktionen wird meist die Verteilungsfunktion der Standardnormalverteilung oder die der logistischen Verteilung verwendet. Die resultierenden

Modelle

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + (\beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p}) \cdot F(\gamma(Y_{t-d} - r)) + \varepsilon_t .$$

werden als *Smooth transition Autoregressive Zeitreihenmodelle* der Ordnung p , kurz *STAR[p]-Modelle*, bezeichnet. Die Modelle wurden auf verschiedene ökonomische Reihen angewendet. Eine andere Variante benutzt die Exponentialfunktion $F(z) = 1 - 2 \exp(-z^2)$; diese entspricht einem TAR-Modell mit zwei zu Null symmetrischen Schwellenwerten. Mit den Parametern $d = 1$ und $\zeta = 0$ erhält man ein *EAR-Modell*, ein exponentiell autoregressives Modell.

Die TAR- und STAR-Modelle wurden verschiedentlich zur Modellierung eingesetzt. Bei der Interpretation der Modelle ist dann aber sehr große Vorsicht angebracht. Noack (2003) hat die SETAR-Modellierung detailliert untersucht und kommt zu dem Fazit: „Nur bei ungewöhnlich großen Zeitreihenlängen kann von einer sinnvollen Anpassung eines SETAR-Modells ausgegangen werden, da nur dann eine befriedigende Identifizierbarkeit gewährleistet ist.“ Bei den STAR-Modellen scheint die Situation noch schwieriger. Schlittgen (1997) erhielt bei seinen Untersuchungen, dass die Likelihoodfunktion vielfach sehr flach verläuft und die Schätzungen dementsprechend eine sehr große Varianz aufweisen.

11.4 Bedingt heteroskedastische Modelle

Das ARCH-Modell

Bei Anlageentscheidungen sind die wichtigsten Kriterien für einen potenziellen Investor die erwartete Rendite und das Risiko, das er beim Investment eingeht. Ist K_t der Kurs einer Anlage zum Zeitpunkt t , so kann die Rendite näherungsweise durch

$$R_t = \ln(K_t) - \ln(K_{t-1})$$

gemessen werden. Das Risiko äußert sich dabei in den Renditeschwankungen. Die Stärke der Renditeschwankungen wird als *Volatilität* bezeichnet.

Vor allem für Optionen und Optionsscheine ist die Analyse der Volatilität der Aktienkursrenditen ein entscheidender Faktor. Je höher die Volatilität, desto größer wird der Wert einer Option bzw. eines Optionsscheines; denn mit der Volatilität erhöht sich die Wahrscheinlichkeit, dass sich ein Kurs einstellt, bei dem die Ausübung der Option lohnt. Das größere Ausübungsrisiko ist dem Optionsverkäufer in einer höheren Prämie zu entgelten.

Für Aktienkursreihen wurde recht früh festgestellt, dass deren Varianzen nicht konstant über die Zeit sind. Große (kleine) Kursbewegungen neigen dazu, von großen (kleinen) Kursbewegungen gefolgt zu werden. Die Volatilität ist gewissermaßen

in Clustern stark ausgeprägt oder schwach, vgl. die Abbildung 6.24. Der Übergang ist dabei jeweils nicht abrupt, sondern erfolgt gewissermaßen glatt. Um dies zu modellieren, wird man die Streuung von den vergangenen Werten abhängen lassen. Dann sind die *bedingten Varianzen*, gegeben die jeweils vergangenen Werte von Y_t , von der Zeit abhängig, nicht die Varianz der Zeitreihenvariablen. Das Modell ist auch geeignet für Residuen von Regressionen, bei denen es um zeitliche angeordnete Beobachtungen geht.

Definition 11.8 (ARCH-Prozess)

Ein *autoregressiver bedingt heteroskedastischer Prozess*, kurz *ARCH[q]-Prozess*, ist definiert durch

$$Y_t = \varepsilon_t \cdot \sqrt{\beta_0 + \sum_{j=1}^q \beta_j Y_{t-j}^2}. \quad (11.12)$$

Dabei sind die ε_t unabhängig und identisch verteilt mit Erwartungswert 0 und Varianz 1. Die Koeffizienten erfüllen $\beta_0 > 0$, $\beta_1, \dots, \beta_q \geq 0$.

Die Nichtnegativitätsbedingungen für die Modellparameter sind notwendig, um eine positive bedingte Varianz der Beobachtung Y_t sicherzustellen. Diese ist hierbei eine lineare Funktion der quadrierten vergangenen Beobachtungen:

$$\text{Var}(Y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \dots) = \sigma^2 (\beta_0 + \beta_1 y_{t-1}^2 + \dots + \beta_q y_{t-q}^2)$$

Notwendig und hinreichend für die Existenz von ARCH-Prozessen der Form (11.12) ist, dass die Wurzeln der charakteristischen Gleichung $1 - \beta_1 z - \dots - \beta_q z^q = 0$ außerhalb des Einheitskreises liegen. Bei Berücksichtigung der Nichtnegativitätsbedingungen $\beta_u \geq 0$, $u = 1, \dots, q$ vereinfacht sich das Kriterium zu $\sum_{i=1}^q \beta_i < 1$.

Oft wird vorausgesetzt, dass die ε_t normalverteilt sind. Trotz dieser Verteilungsannahme weisen die Randverteilungen der Variablen bei ARCH-Modellen stärkere Flanken auf, als sie die Normalverteilung hat. Dies lässt sich etwa anhand des Exzesses der Randverteilungen zeigen. Bei einem ARCH[1]-Modell mit normalverteilten ε_t ist sie etwa, wenn die Stationaritätsbedingung $\beta_1 \geq 0$ erfüllt ist:

$$3 \cdot \frac{1 - \beta_1^2}{1 - 3\beta_1^2} - 3 > 0,$$

Damit ist der Exzess größer als bei der Normalverteilung. Somit wird dem oft beobachtbaren großen Anteil sehr extremer Werte in Rendite-Reihen zum mindestens ansatzweise Rechnung getragen. Zusätzlich wurden aber auch weitere Verteilungen, die schon selbst viele extreme Werte hervorbringen, ins Spiel gebracht.

Die Struktur des Modells impliziert, dass große Werte von $y_{t-1}^2, \dots, y_{t-q}^2$ eine große bedingte Varianz für Y_t implizieren. Y_t nimmt folglich selbst einen (absolut) großen Wert an. Dies zeigt, dass ARCH-Modelle die erwähnten Volatilitäts-Cluster erzeugen können.

Beispiel 11.9 (DAX)

Der DAX ist der wichtigste deutsche Aktienindex. Er spiegelt die Entwicklung der 30 größten und umsatzstärksten, an der Frankfurter Wertpapierbörsen gelisteten Unternehmen wider. Die hier betrachtete Reihe DAX umfasst 522 werktägliche Werte des Performance-Index für die Jahre 2009 und 2010. Die Werte werden in Renditen transformiert gemäß $y_t = \ln(DAX_t) - \ln(DAX_{t-1})$.

Die beiden Grafiken der Abbildung 11.6 illustrieren die vorstehend angegebenen ARCH-Effekte. Einmal ist deutlich, dass es zahlreiche extreme Werte in der Zeitreihe der Renditen gibt. Zudem gibt es zu Beginn eine Phase größerer Volatilität, diese nimmt im zweiten Teil der Zeitreihe ab. Damit geht einher, dass die quadrierten Werte zunächst tendenziell größer sind.

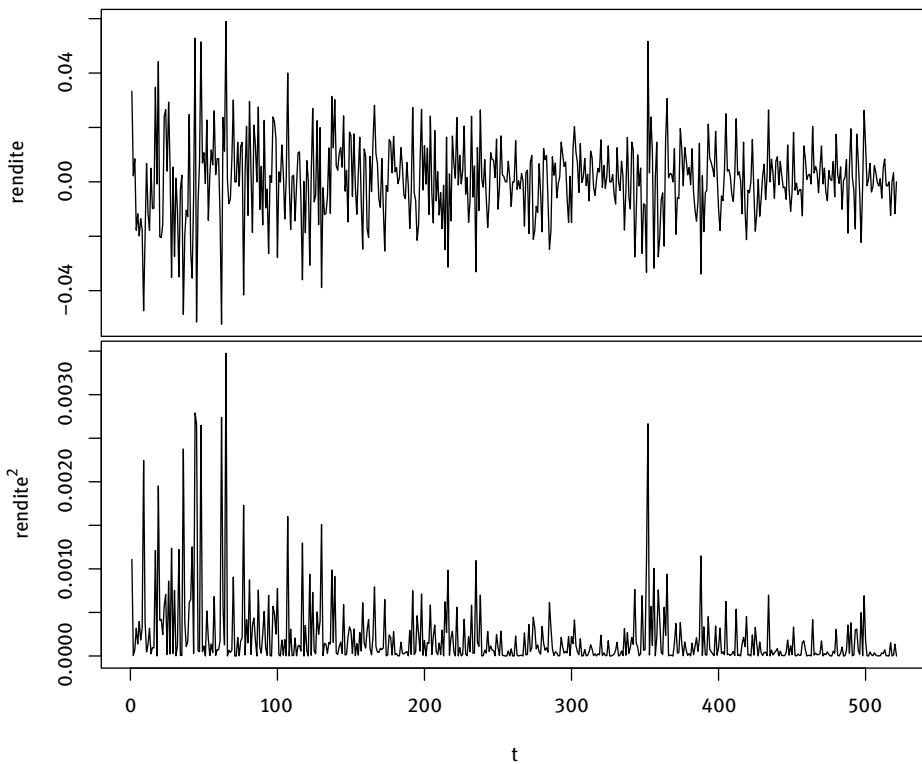


Abb. 11.6: DAX-Renditen (oben) und quadrierte Zeitreihe (unten)

Wie die ACF in der Abbildung 11.7 zeigt, sind die Beobachtungen unkorreliert. Bei der PACF der Reihe der quadrierten Werte gibt es aber etliche deutliche Verletzungen der Grenzen. Dies deutet darauf hin, dass die Variablen der Originalreihe nicht unabhängig sind und die Reihe insgesamt einige ARCH-Effekte aufweist.

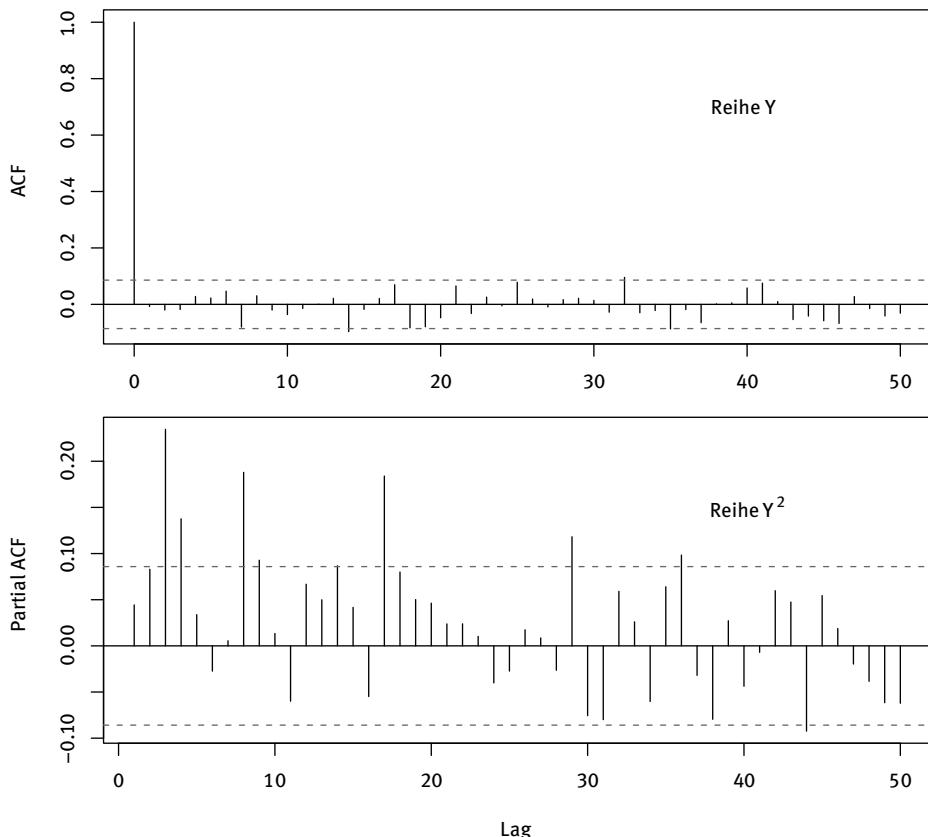


Abb. 11.7: ACF der DAX-Renditen (oben) und PACF der quadrierten DAX-Renditen (unten)

R-Code:

```

1 data(DAX)
2 dax<-DAX[(DAX[,3]>2008)&(DAX[,3]<2011),4]
3 rendite <- ts(diff(log(dax)))
4 par(mfrow=c(2,1))
5 plot(rendite,xlab="Zeitindex",ylab="rendite",xaxt="n")
6 plot(rendite^2,xlab="Zeitindex",ylab=expression(rendite^2))
7 acfpacf(rendite,50)

```

Modellanpassung und Parameterschätzung

Die Identifikation eines ARCH-Prozesses nutzt die Tatsache aus, dass bei einem ARCH[q]-Prozess (Y_t), der Prozess der quadrierten Beobachtungen, (Y_t^2), unter milden Bedingungen ein AR[q]-Prozess ist. Dementsprechend kann der Ljung-Box-Pierce-Test

eingesetzt werden, um ARCH-Strukturen in der Ausgangsreihe zu überprüfen. Auch kann die PACF inspiriert werden, um die Ordnung des ARCH-Modells zu bestimmen. Einen Lagrange-Multiplikatortest zum Testen auf die Lag-Länge der ARCH-Struktur wurde von Engle (1982) vorgeschlagen. Die von ihm betrachtete Situation war die einer Regression mit unter Umständen autokorrelierten Fehlern. Dann wäre die im Folgenden die Zeitreihe die der Residuen, welche sich bei der Anpassung eines adäquaten AR-Modells ergeben:

1. Bestimme die Quadrate y_t^2 der Zeitreihenwerte und regressiere diese gegen eine Konstante und q gelagte Werte:

$$y_t^2 = \hat{\alpha}_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \eta_t$$

2. Gibt es keine ARCH-Komponenten, so gilt $\alpha_i = 0$ für alle $i = 1, \dots, q$. Diese Nullhypothese wird gegen die Alternative getestet, dass, bei vorhandenen ARCH-Komponenten, mindestens einer der geschätzten Koeffizienten signifikant von Null verschieden ist.

Die Teststatistik ist $N \cdot R^2$, wobei N die Anzahl der Beobachtungen ist und R^2 das Bestimmtheitsmaß. Unter der Nullhypothese folgt die Teststatistik einer χ^2 -Verteilung mit q Freiheitsgraden. Ist $N \cdot R^2$ größer als das $1 - \alpha$ -Quantil dieser Chi-quadrat-Verteilung, so wird die Nullhypothese verworfen; dann liegt ein ARCH-Effekt in dem ARMA-Modell vor. Andernfalls behandelt man die Reihe so, als wäre ein entsprechender Effekt nicht vorhanden.

Natürlich kann zur Bestimmung der Ordnung eines ARCH-Modells auch eines der Informationskriterien herangezogen werden. Hier ist auf die ursprüngliche Definition des AIC zurückzugehen und das Kriterium geeignet zu adaptieren, vgl. Brooks & Burke (1998).

Geschätzt werden ARCH-Modelle mittels der Maximum-Likelihood Methode. Die Loglikelihoodfunktion ist einfach anzugeben; sie ist bis auf Konstante gleich, siehe Engle (1982):

$$L = -\frac{1}{N} \sum_{t=1}^N \left(\ln(h_t) + \frac{1}{2} \frac{y_t^2}{h_t} \right). \quad (11.13)$$

Um die Anpassung zu überprüfen ist auf Residuen zurückzugreifen. Bei ARCH-Modellen ist nicht ohne weiteres klar, was als Residuen anzusehen ist. Oft werden die Beobachtungen selbst genommen. Da sie aber unterschiedliche bedingte Varianzen haben, werden die standardisierten Werte als Residuen verwendet:

$$\hat{y}_t = y_t \sqrt{\hat{\beta}_0 + \sum_{u=1'} \hat{\beta}_u y_{t-u}^2}. \quad (11.14)$$

Auf diesen standardisierten Residuen basiert ein diagnostischer Test von Li & Mak (1994). Die Teststatistik lautet

$$Q(r, M) = N \cdot \sum_{i=r+1}^M \hat{r}_i^2 ; \quad (11.15)$$

dabei ist $\hat{r}_k = (\sum_{t=k+1}^N (\hat{y}_t^2 - \bar{\hat{y}})(\hat{y}_{t-k}^2 - \bar{\hat{y}})) / (\sum_{t=1}^N (\hat{y}_t^2 - \bar{\hat{y}})^2)$

Die Teststatistik ist asymptotisch χ^2 -verteilt mit $M-r$ Freiheitsgraden, wenn $r > q$, wenn also nur über quadrierte Autokorrelationen summiert wird, deren Index die Ordnung des ARCH-Modells überschreitet.

Beispiel 11.10 (DAX – Fortsetzung von Seite 277)

Die PACF der quadrierten Werte, siehe Abbildung 11.7, weist darauf hin, dass eine höhere Lag-Ordnung für die ARCH-Beziehung nötig ist. Der Ljung-Box-Pierce-Test ist für Lags ≥ 6 deutlich signifikant.

Eine Regression der y_t^2 gegen zeitlich verschobenen Reihen y_{t-1}^2 bis y_{t-8}^2 ergibt einen Wert von $R^2 = 0.1173$. Multipliziert mit der Reihenlänge $N = 521$ ergibt eine klare Signifikanz, hat doch das 0.95-Quantil der χ_8^2 -Verteilung den Wert 15.507.

Da mit der Erfassung der Abhängigkeiten niedrigerer Ordnung unter Umständen auch die höheren mit eingefangen werden, wird ein ARCH[8]-Modell angepasst. Die geschätzten Koeffizienten sind:

Lag u	$\hat{\beta}_u$	Std.Fehler	t-Wert	P-Wert	Lag u	$\hat{\beta}_u$	Std.Fehler	t-Wert	P-Wert
0	0.00005	0.00001	3.637	0.0003	5	0.1287	0.03702	3.476	0.0005
1	0.0063	0.03089	0.203	0.8394	6	0.0390	0.04680	0.834	0.4045
2	0.1697	0.06023	2.818	0.0048	7	0.1077	0.04154	2.593	0.0095
3	0.1670	0.05803	2.878	0.0040	8	0.1260	0.05443	2.315	0.0206
4	0.0953	0.05451	1.748	0.0805					

Nicht alle Koeffizienten sind signifikant; das Lag 8 ist aber offensichtlich notwendig.

Das Modell ist nicht vollständig zufriedenstellend, da die PACF der quadrierten Residuen noch einige starke Verletzungen der Grenzen aufweist, vgl. Abbildung 11.8. Jedoch ist der diagnostische Test von Li und Mak für $M = 100$ und alle r von 9 bis 60 fern von der Signifikanz. Auch sind die Residuen recht gut normalverteilt.

R-Code:

```
1 library(tseries)
2 out <- garch(rendite,order=c(0,8))
```

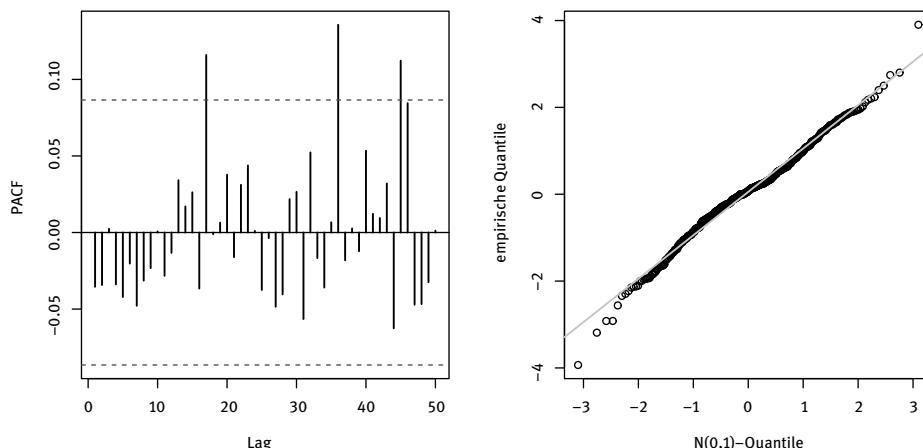


Abb. 11.8: PACF der Reihe der quadrierten standardisierten Residuen des ARCH[8]-Modells und Normalverteilungs-QQ-Diagramm der standardisierten Residuen

```

3 summary(out)
4 res <- out$residuals[-c(1:8)]
5 par(mfrow=c(1,2))
6 acf(res^2,lag.max=50,type="partial")
7 qqnorm(res,main="")
8 abline(mean(res),sd(res),col="grey",lwd=2)
9 M <- 100
10 out1 <- acf(res^2,M)
11 a <- out1$a[-1]
12 n <- length(res)
13 r2 <- n*rev(cumsum(rev(a^2)))
14 out1 <- NULL
15 for (i in c(9:61)){out1<-rbind(out1, round(c(i,r2[i+1],
16 1-pchisq(r2[i],M-i)),2)) }
17 print(out1)

```

Für die im Beispiel 11.9 bestimmten Renditen wird ein GARCH-Modell angepasst. Das Modell wird mit der Funktion `garch` geschätzt. Auf den hier 0 gesetzten Parameter wird weiter unten eingegangen. Ab Zeile 9 werden die Residuen untersucht. Die führenden NA's werden dazu entfernt. Ab Zeile 13 wird dann der Test von Li und Mak berechnet. Das geschieht hier nur zur Illustration. Der Test kann auch mit der Funktion `Weighted.LM.test` aus dem Paket `WeightedPortTest` durchgeführt werden.

Modellerweiterungen

In der Form (11.12) kann das ARCH-Modell nur solche Prozesse beschreiben, deren Erwartungswert und bedingter Erwartungswert bei null zentriert sind. Um flexiblere Modelle für die Zeitreihenanalyse zu erhalten, wird das ARCH-Modell in der Form (11.12) nicht für die Ausgangsreihe selbst, sondern zur Beschreibung des Residualprozesses eines autoregressiven Modells eingesetzt. AR[p]-Modelle mit ARCH[q]-Störungen, kurz *ARARCH[p, q]-Modelle*, sind einfach in geschlossener Form anzugeben. Sie sind definiert durch:

$$Y_t = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + \varepsilon_t \cdot \sqrt{\beta_0 + \sum_{j=1}^q \beta_j e_{t-j}^2}; \quad (11.16)$$

dabei ist $e_t = Y_t - \alpha_0 - \sum_{i=1}^p \alpha_i Y_{t-i}$. Es wird natürlich wieder $\beta_0 > 0$, $\beta_1, \dots, \beta_q \geq 0$ gefordert.

Beispiel 11.11 (Simulierter ARARCH-Prozess)

Es wird ein ARARCH[1, 2]-Prozess mit den Parametern $\alpha_0 = 0$, $\alpha_1 = -0.9$, $\beta_0 = 0.01$, $\beta_1 = 0.4$, $\beta_2 = 0.4$ simuliert. Die Abbildung 11.9 zeigt die resultierende Zeitreihe.

R-Code:

```

1 n <- 300
2 a <- -0.9
3 b <- c(0.01, 0.4, 0.4)
4 e <- rnorm(n+100, 0.2)
5 y <- e
6 t <- 4
7 while(t < (n+100)){
8   t <- t+1
9   e[t] <- e[t]*sqrt(b[1] + sum(b[3:2]*(e[(t-2):(t-1)]^2)))
10  y[t] <- a*y[t-1] + e[t]
11 }
12 y <- y[101:(n+100)]
13 ts.plot(y)

```

In den Zeilen 9 und 10 wird die Definitionsgleichung direkt umgesetzt. Die ersten Werte der über die Iteration erzeugte Reihe werden in Zeile 12 abgeschnitten, da der Prozess sich erst einschwingen muss, vgl. den Abschnitt vor dem Beispiel 3.3.

Bei ARARCH-Modellen sind die Schätzer der AR-Parameter und der ARCH-Parameter asymptotisch unabhängig. So kann analog zu den Regressionsmodellen mit AR-Störungen iterativ vorgegangen werden. Man schätzt im ersten Schritt die AR-Parameter;

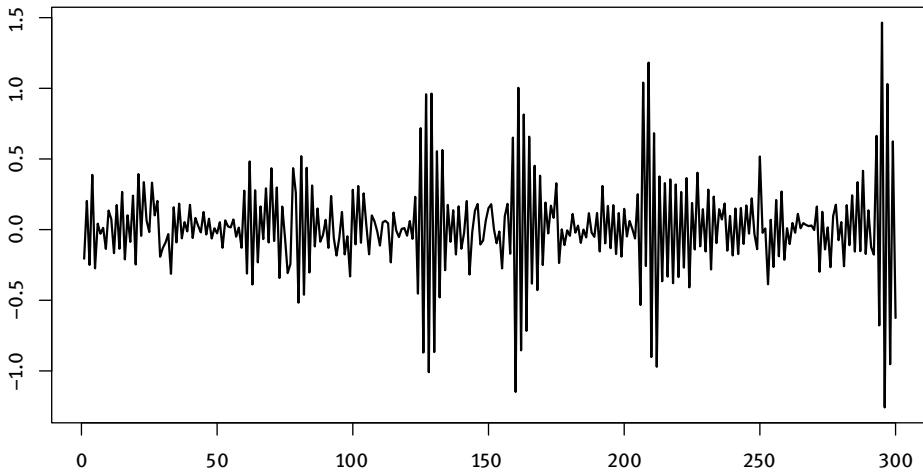


Abb. 11.9: Realisierung eines ARARCH[1, 2]-Prozesses

sodann werden die Residuen verwendet um die ARCH-Parameter mit dem ML-Ansatz zu schätzen. Die Berücksichtigung der bedingten Varianzen führt zu einer verbesserten Schätzung des AR-Teils, usw.

Eine häufig anzutreffende Form für die Angabe von solchen Modellen ist:

$$Y_t = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + e_t \quad (11.17a)$$

$$e_t | (Y_{t-1}, Y_{t-2}, \dots) \sim \mathcal{N}(0, \sigma_t^2), \quad \sigma_t^2 = \beta_0 + \sum_{j=1}^q \beta_j e_{t-j}^2. \quad (11.17b)$$

Dabei ist der erste Teil der zweiten Gleichung so zu lesen, dass die bedingte Verteilung von e_t bei gegebenen Y_{t-1}, Y_{t-2}, \dots eine Normalverteilung ist. In dieser Notation spiegelt sich die Tatsache wider, dass eine getrennte Modellierung von bedingter Erwartung und bedingter Varianz erfolgt.

Die Schreibweise (11.17) lässt eine ganze Reihe von möglichen Verallgemeinerungen des ARCH[q]-Modells deutlich werden. Eine Erweiterung des ARCH-Modells, die sich auch für praktische Anwendungen durchgesetzt hat, ist das *generalisierte ARCH-Modell*, kurz *GARCH-Modell*. Die Erweiterung besteht darin, dass vergangene bedingte Varianzen mit in die Modellgleichung einbezogen werden. Die bedingte Varianz eines GARCH[p, q]-Prozesses ist gegeben durch

$$\sigma_t^2 = \beta_0 + \sum_{j=1}^q \beta_j e_{t-j}^2 + \sum_{k=1}^p \delta_k \sigma_{t-k}^2, \quad (11.18)$$

wobei $\beta_0 > 0, \beta_1, \dots, \beta_q, \delta_1, \dots, \delta_s \geq 0$.

Den Hintergrund für die GARCH-Modelle bildet die oftmals benötigte hohe Ordnung für ARCH[q]-Modelle. Das bereitet schnell Probleme bei der Schätzung wegen

der Stationaritäts- und Nichtnegativitätsbedingungen. Dass auf diese Weise die hohen Ordnungen umgangen werden können, ist aufgrund der Diskussion der Zusammenhänge von AR- und ARMA-Prozessen naheliegend.

Eine Bedingung für die Existenz eines stationären Prozesses (e_t) mit bedingter Varianz gemäß (11.18) ist $\sum_{j=1}^q \beta_j + \sum_{k=1}^s \delta_k < 1$. Jedoch können streng stationäre GARCH-Prozesse existieren, deren Parameter dieses Kriterium nicht erfüllen. In solchen Fällen existieren die Momente $E(e_t^2)$ nicht. Dies gilt speziell für die SGARCH-Modelle, bei denen für die Innovationen eine stabile Verteilung unterstellt wird. Diese Verteilungen haben i. d. R. eine unendliche Varianz. Damit versucht man, die z. T. extremen Ausschläge bei empirischen Reihen adäquat zu modellieren. Für das gleiche Ziel wurden die TGARCH-Modelle vorgeschlagen, bei denen die Innovationen eine t -Verteilung haben. Ihr Vorteil ist es, dass weiterhin ML-Methoden existieren. Dies sind aber nur zwei spezielle Vertreter der vielen verschiedenen Mitglieder der GARCH-Modell-Familie. Auch multivariate Varianten stehen zur Verfügung.

In Analogie zur Situation bei der Anpassung von ARMA-Modellen wird man diese allgemeinere Modellklasse den reinen ARCH-Modellen vorziehen, da man in der Regel, dem Prinzip der Sparsamkeit gehorchend, mit weniger Modellparametern auskommen kann. Dies ist besonders wichtig, da bei der Schätzung von ARCH-Modellen mit hoher Ordnung oft die Nichtnegativitätsbedingungen verletzt werden.

Die Identifikation eines GARCH-Modells ist keine einfache Angelegenheit. Ein Weg zur Identifizierung besteht darin, mit einem Modell größeren Werten für p und q zu starten und unter den Einsatz geeigneter Tests zu einem Modell mit kleineren Ordnungen zu gelangen. Im Prinzip kann zur Identifikation eines GARCH-Modells ausgenutzt werden, dass die Autokorrelationsstruktur des Prozesses (e_t^2) der eines ARMA-Modells entspricht. Auch lassen sich die Informationskriterien einsetzen, um eine adäquate Modellordnung zu bestimmen. Auch die Schätzung bereitet aufgrund der nichtlinearen Optimierung schnell Schwierigkeiten. In den meisten publizierten Anwendungen werden nur Modelle mit kleinen Ordnungen verwendet, oft sind es nur GARCH[1, 1]-Modelle.

Beispiel 11.12 (MDAX)

Es wird die Zeitreihe des MDAX für die beiden Jahre 2009 und 2010 betrachtet. Dieser Aktienindex umfasst 50 Werte – vorwiegend aus klassischen Branchen –, die in der Rangliste nach Marktkapitalisierung des Streubesitzes und des Börsenumsatz auf die Werte des DAX folgen. Der MDAX spiegelt damit die Kurssentwicklung von Aktien mittelgroßer deutscher oder überwiegend in Deutschland tätiger Unternehmen wider.¹ Aus den Werten der Reihe MDAX werden zuerst wieder die Renditen gewonnen, $y_t = \ln(\text{MDAX}_t) - \ln(\text{MDAX}_{t-1})$. Diese MDAX-Renditen sind in der Abbildung 11.10 wiedergegeben. Die ACF und PACF der quadrierten Renditen weist auf Abhängigkeiten höherer Ordnung hin.

¹ Aus Wikipedia; <http://de.wikipedia.org/wiki/MDAX>.

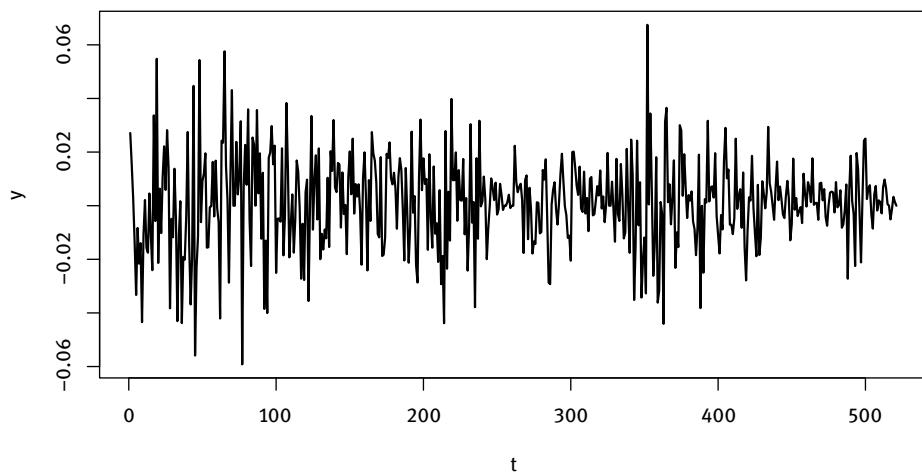


Abb. 11.10: Renditen des MDAX

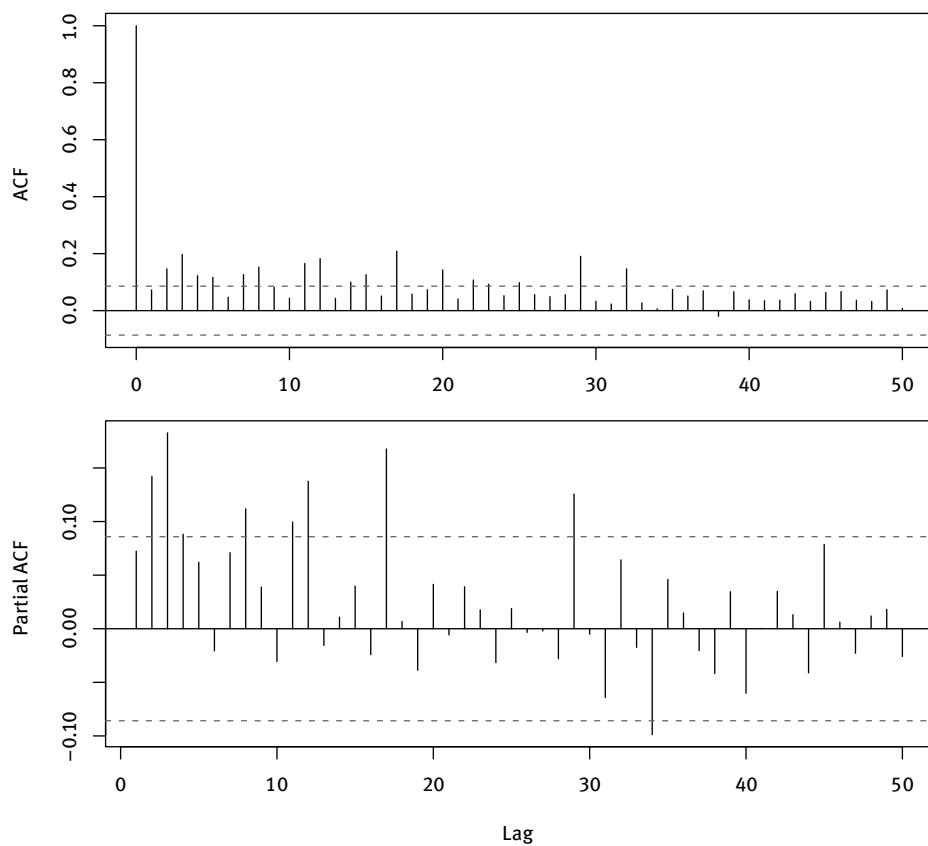


Abb. 11.11: ACF und PACF der quadrierten Renditen des MDAX

Die Anpassung eines GARCH[1, 2]-Modells an die Renditen ergibt:

Coefficient(s):

	Estimate	Std. Error	t value	Pr(> t)
a0	6.156e-06	2.439e-06	2.524	0.0116
a1	1.033e-02	4.456e-02	0.232	0.8166
a2	1.457e-01	5.812e-02	2.507	0.0122
b1	8.259e-01	3.264e-02	25.303	<2e-16

Diagnostic Tests:

Jarque Bera Test

data: Residuals

X-squared = 7.3024, df = 2, p-value = 0.02596

Die Ordnungen sind offensichtlich nötig, da die jeweilig höchsten Koeffizienten signifikant sind. Bei den ARCH-Parametern könnte aber der erste Koeffizient weggelassen werden. Die Signifikanz des Jarque-Bera-Tests ist nicht bedeutsam; mit $N = 521$ Werten ist die Reihe eher lang. Auch zeigt das Normalverteilungs-QQ-Diagramm eine hinreichende Gruppierung um eine Gerade. Schließlich ist das Modell ausreichend, da der Ljung-Box-Pierce-Test keine Korrelationen bei den quadrierten Residuen mehr feststellt.

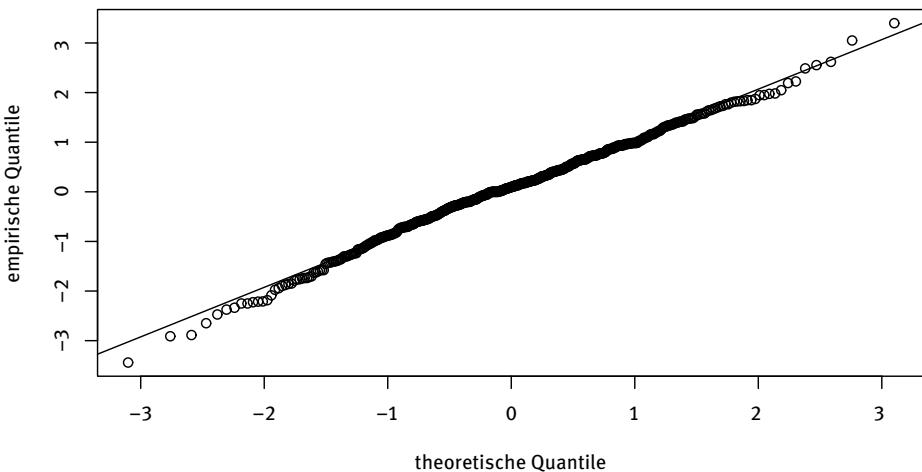


Abb. 11.12: Normalverteilungs-QQ-Diagramm für die Residuen des GARCH-Modells

lags	df	statistic	p-value	lags	df	statistic	p-value
6	3	2.771	0.428	36	33	27.520	0.736
12	9	7.507	0.584	42	39	29.677	0.859
18	15	12.644	0.630	48	45	32.697	0.914
24	21	16.345	0.750	54	51	34.775	0.960
30	27	20.991	0.787	60	57	39.229	0.965

R-Code:

```

1 data(MDAX)
2 y <- MDAX[(MDAX[,3]>2008)&(MDAX[,3]<2011),4]
3 rendite <- ts(diff(log(y)))
4 library(tseries)
5 out <- garch(rendite,order=c(1,2))
6 summary(out)
7 res <- out$residuals[-c(1:2)]
8 qqnorm(res)
9 abline(mean(res),sd(res))
10 LjungBoxPierceTest(res^2,3,60)

```

Der Code gibt im Vergleich zu dem von Beispiel 11.10 nichts wesentlich Neues.

11.5 R-Funktionen

Funktionen zum Arbeiten mit Markov-switching-Modellen stellt das Paket MSwM bereit. Die Schätzung verschiedener nichtlinearer autoregressiver Modelle, speziell auch TAR- und STAR-Modelle, erlauben die Pakete tsDyn und NTS. Tsay & Chen (2018) geben eine ausführliche Beschreibung der Möglichkeiten des Paketes NTS.

Die Funktion garch passt ein GARCH $[p, q]$ -Modell an eine Zeitreihe an. Sie befindet sich im Paket tseries. Das Paket TSA enthält die Funktion garch.sim, mit der solche Reihen simuliert werden können. fGarch und gogarch sind zwei weitere Pakete, die für spezielle Verallgemeinerungen des ARCH-Modells geschrieben wurden.

Ein eigenes, umfangreiches Paket zur Analyse von GARCH-Modellen ist rugarch. Hier kann die Modellierung von Finanzmarktrenditen mit Modellen der Klasse ARFIMA umgesetzt werden, die der extrahierten Volatilität selbst – mit GARCH. Das Paket rmgarch wurde speziell für multivariate bedingt heteroskedastische Modelle entwickelt. Wie im R-Code zum Beispiel 11.10 erwähnt, kann der Test von Li & Mak mit der Funktion Weighted.LM.test aus dem Paket WeightedPortTest durchgeführt werden.

12 Spezielle Probleme

12.1 Fehlende Werte

Es ist eher häufig, dass bei Zeitreihen Werte fehlen; das können vereinzelte sein oder Blöcke von Werten. Außer dem Wunsch, auch in dieser Situation eine ‚normale Analyse‘ durchführen zu wollen, also die Kennfunktionen zu bestimmen und Prozessmodelle anzupassen, stellt sich als neue Aufgabe die Schließung der Lücken, d. h. die Ergänzung bzw. Schätzung der fehlenden Werte.

Auf die Behandlung fehlender Werte bei Zustandsraummodellen wurde schon im Kapitel 10 eingegangen, siehe insbesondere Seite 255.

Außerhalb der Zustandsraummodelle ist für die Modellschätzung bei den meisten Methoden die ACF die entscheidende Größe. Diese ist daher vorab (unter der Voraussetzung der Stationarität) zu schätzen. Dazu werden die fehlenden Werte durch den Mittelwert ersetzt, da die Nichtberücksichtigung der Lücken zu fundamentalen Fehleinschätzungen der Abhängigkeitsstruktur führen würde.

Das für die Zentrierung benötigte arithmetische Mittel $\bar{y} = \frac{1}{N^*} \sum y_t$ ist natürlich auch aus der Reihe mit Lücken bestimmbar. N^* ist dabei die Anzahl der vorhandenen Beobachtungen. Im Folgenden wird von der mit diesem Mittel zentrierten Reihe ausgegangen.

Dann erhält man

$$c_\tau = \frac{1}{N^*} \sum_{t=1}^{N-\tau} y_t y_{t+\tau}. \quad (12.1)$$

Die Auffüllung der Lücken der zentrierten Reihe mit Nullen bewirkt, dass die Kreuzprodukte an den Stellen verschwinden, an denen keine Beobachtungen vorliegen. Um einer Verzerrung aufgrund des Nenners entgegenzuwirken, wird mit der Anzahl der tatsächlich vorliegenden Beobachtungen standardisiert.

Für multivariate Reihen ist ein analoges Vorgehen nahe liegend, bei dem jede Reihe einzeln durch den Mittelwert der Reihe aufgefüllt wird.

Will man einer Zeitreihe ein spezielles Modell, etwa einen AR-Prozess, anpassen, so kann die einfache Ersetzung von Lücken durch Null bzw. den Mittelwert verbessert werden, indem mit dem zunächst angepassten Modell die fehlenden Werte ‚prognostiziert‘ werden. Mit den so geschätzten Werten erhält man eine Verbesserung der Schätzung der ACF, die ihrerseits eine verbesserte Schätzung des Modells bewirkt. Da sich die Likelihoodfunktion bei unterstellter Normalverteilung auch für Reihen mit Lücken aufschreiben lässt, gehen aufwendigere Ansätze von einem ARMA-Modell aus und setzen die Maximum-Likelihood-Methode zur Modellschätzung ein.

Interessiert man sich für die möglichst gute Schätzung der fehlenden Werte selbst, so steht die Methode der kleinsten Quadrate zur Verfügung. Diese geht von einem stationären AR[∞]-Prozess aus, der mit $\pi_0 = -1$ in der Form $e_t = \sum_{u=0} \pi_u Y_{t-u}$ geschrieben

wird. Ziel ist es, die Summe der quadrierten Residuen zu minimieren. Bei einem einzelnen fehlenden Wert zum Zeitpunkt s lautet das Problem also:

$$\sum_t e_t^2 \stackrel{!}{=} \min_{Y_s} .$$

Für die praktische Umsetzung geht man von einem angepassten AR-Modell $Y_t = \hat{\alpha}_1 Y_{t-1} + \dots + \hat{\alpha}_p Y_{t-p} + \varepsilon_t$ aus und minimiert

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

in Abhängigkeit von y_{t_1}, \dots, y_{t_k} , wenn $t_1 < \dots < t_k$ die Zeitpunkte kennzeichnen, an denen die Beobachtungen fehlen. Die Schätzwerte erhält man durch Nullsetzen der partiellen Ableitungen $\partial SS / \partial y_{t_i}$ als Lösung eines linearen Gleichungssystems.

Grundprinzip der KQ-Interpolation ist also die Schätzung der fehlenden Werte mittels Linearkombinationen der vorhandenen. Die dabei auftretenden Gewichte ergeben sich als die Koeffizienten der IACF des zugrundeliegenden Prozesses.

Beispiel 12.1 (Grundwasserspiegel – Fortsetzung von Seite 196)

In der bereits im Beispiel 8.13 angesprochenen Studie zum Grundwasser in der Region Werther weisen die erhobenen Reihen zum Teil erhebliche Lücken auf, insbesondere bei dem Grundwasserpegel des Peilbrunnens L921 für die Zeit vom 1.2. bis 31.12.1980. Diese wurden nach Angabe der Autoren von einem erfahrenen Statistiker am Bildschirm aufgefüllt.

In der Abbildung sind diese Ersetzungen („Hand“-Methode) den Ersetzungen mittels der zwei Methoden, AR- und KQ-Methode, gegenübergestellt. Es ergibt sich eine recht große Übereinstimmung der AR- und der KQ-Methode. Die Unterschiede sind erst bei einer Vergrößerung der Grafik richtig zu erkennen. Auch der Statistiker hat nicht schlecht geschätzt. Die anderen Methoden führen jedoch zu mehr Struktur, wenn eine größere Anzahl von Beobachtungen zusammenhängend fehlt. Dies wird um die Tage mit den Nummern 200 bis 250 besonders deutlich. Zudem handelt es sich bei dem manuell eingefügten Wert an der Stelle 151 offensichtlich um einen Schreibfehler.

R-Code:

```

1 data(L921)
2 fehl <- L921[is.na(L921[,2]),1]
3 hand <- L921[fehl,3]
4 y <- L921[,2]
5 outar <- missar(y,10)
6 outls <- missls(y,10)
7 ts.plot(outls, col='gray')
8 lines(L921[,1],outar$y)
9 points(fehl,hand,pch=8)

```

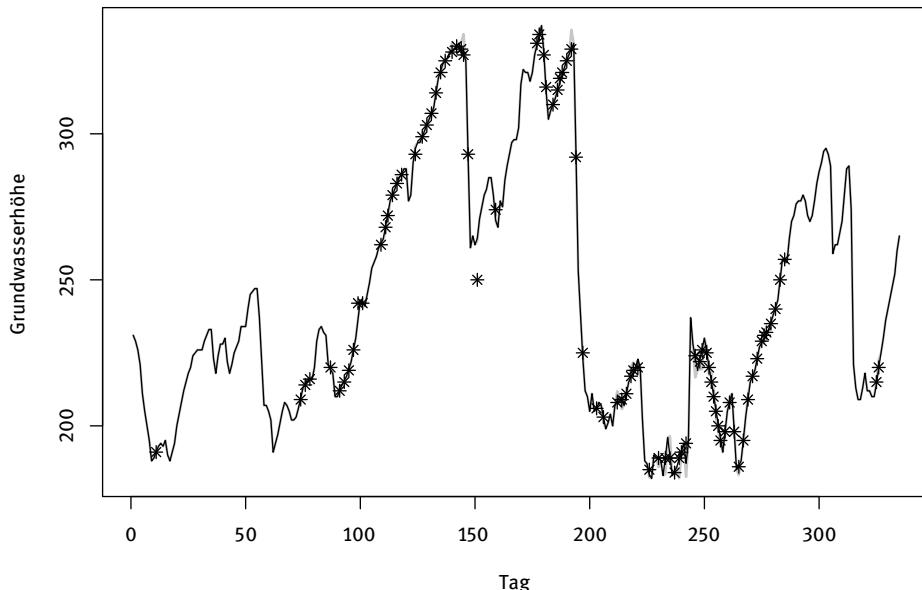


Abb. 12.1: Aufgefüllte Reihe L921; schwarz: AR, grau: KQ, Stern: per Hand

Für die Darstellung der per Hand eingesetzten Punkte wird zu Beginn ermittelt, an welchen Stellen Beobachtungen fehlen. Die beiden Funktionen zur Schätzung der fehlenden Werte befinden sich in dem Paket tsapp. Das Ergebnis ist dann nur noch grafisch aufzubereiten.

Es gibt weitere Vorschläge zur Schätzung fehlender Werte. Bei nur wenigen fehlenden Werten ist auch der Einsatz von Indikatorvariablen angemessen. Dies kann etwa umgesetzt werden, indem eine Interventionsanalyse durchgeführt wird. Dazu sind die Zeitpunkte mit den fehlenden Werten zu ermitteln, die Lücken mit einem beliebigen Wert zu schließen und dann diese Zeitpunkte als Interventionszeitpunkte zu definieren. Zum Einsatz von Indikatorvariablen bei fehlenden Werten sei auch auf Fuller (1996, S. 458 f.) hingewiesen.

Andere, nicht durch Zeitreihenmodelle geleitete Auffüllungsverfahren können katastrophale Auswirkungen haben. Dies gilt z. B. für Splines-Interpolationen wie auch für die lineare Interpolation der die Lücken begrenzenden Werte. Diese Verfahren sind gänzlich untauglich bei Prozessen mit viel spektraler Masse im Hochfrequenzbereich, bei Prozessen also, die Reihen mit schnellen Schwingungen hervorbringen. Bei Prozessen, die andererseits eher glatte Reihen als Realisierungen haben, sind die Verfahren alle mehr oder weniger gleich gut.

12.2 Nicht gleichabständige Beobachtungen

Bei Zeitreihen mit vielen Lücken ist man in einer Grenzsituation dazu, dass Beobachtungen von vornherein nicht gleichabständig erhoben sind. Erhebungen zu nicht gleichabständigen Zeitpunkten fallen z. B. bei Finanzmärkten an. Dort werden innerhalb eines Tages einzelne Geschäfte mit ihrem Zeitpunkt, ihrem Handelsumfang und dem zugrundeliegenden Preis erhoben. Aber auch in geologischen Anwendungen sind solche Zeitreihen häufig. Hier handelt es sich bei dem Index jedoch i. d. R. um eine räumliche Dimension.

Beispiel 12.2 (Gehalt eines Oxygen Isotops)

Eine in der Form der Datenerhebung unübliche Zeitreihe wurde von Shackleton erhoben, siehe Belcher, Hampton & Tunnicliffe Wilson (1994). Sie besteht aus 164 Werten des relativen Gehaltes eines Oxygen Isotops ($\delta^{18}\text{O}$) an einer Stelle des Meeresbodens. Von Interesse ist die Zeitreihe OXYGEN, da die Werte die klimatischen Bedingungen zu den entsprechenden Zeiten widerspiegeln. Die Messungen wurden in regelmäßigen Abständen von 10 cm durchgeführt. Aufgrund der variierenden Ablagerungsrate korrespondieren die Messpunkte mit unterschiedlichen zeitlichen Abständen. In der Abbildung 12.2 sind die Zeitpunkte, zu denen Beobachtungen vorliegen, markiert, um den unterschiedlichen zeitlichen Abstand zu verdeutlichen. Die Reihe ist hier mittelwertbereinigt dargestellt.

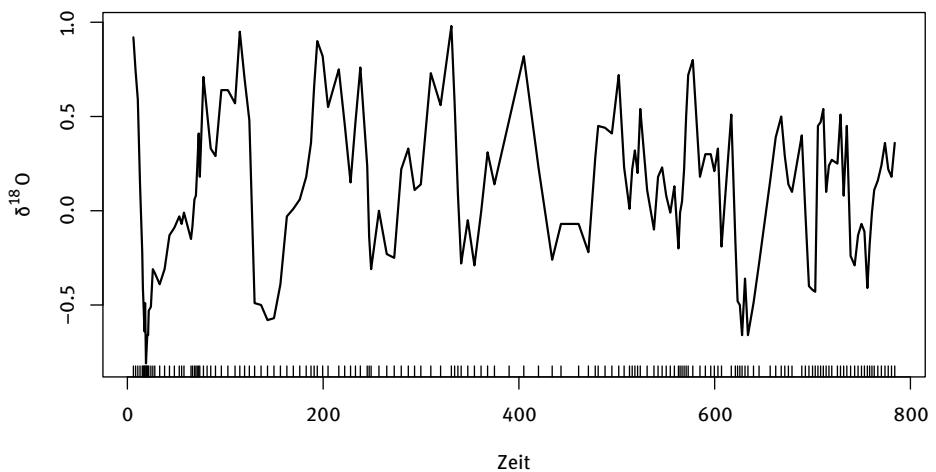


Abb. 12.2: OXYGEN, Oxygen-Isotop-Messungen

R-Code:

```

1 data(OXYGEN)
2 t <- OXYGEN[,1]
3 y <-OXYGEN[,2]
4 plot(t,y,type="l",xlab="Zeit",
5      ylab=expression(paste(delta^18,"0")))
6 rug(t)

```

Für das Markieren der Zeitpunkte wird die Funktion `rug` verwendet.

Die Datensituation nicht gleichabständiger Zeitpunkte stellt an die Zeitreihenanalyse erhöhte Anforderungen. Um zu einem – allerdings eingeschränkten – Auswertungsansatz zu kommen, wird nun von einem Prozess ($Y(t)$) in stetiger Zeit ausgegangen. Modelle für Zeitreihen mit stetigem Zeitparameter sind um etliches schwieriger als für diskrete, gleichabständige Zeitpunkte. Bei der Definition von White-Noise-Prozessen mit stetiger Zeit muss man unendliche Varianzen zulassen. Für die Formulierung von autoregressiven Prozessen sind anstelle der Differenzen Ableitungen zu betrachten; man befindet sich dann im Gebiet der stochastischen Differentialgleichungen. Hierzu sei das Kapitel 3.7 von Priestley (1981) empfohlen. Bergstrom (1990) entwickelte einen eigenen Ansatz zur ökonometrischen Modellierung in stetiger Zeit.

Die Datensituation nicht gleichabständiger Zeitpunkte stellt an die Zeitreihenanalyse erhöhte Anforderungen. So kann schon die Autokorrelationsfunktion nicht mehr auf die übliche Weise geschätzt werden. Um zu einem Auswertungsansatz zu kommen, wird von einem Prozess ($Y(t)$) in stetiger Zeit ausgegangen. Wenn der Prozess dann stationär ist, hängen die Verteilungen der Differenzen $Y(t) - Y(t - \tau)$ nicht von t ab. Zudem ist, wenn $\gamma(\tau)$ die Autokovarianz- und $\rho(\tau)$ die Autokorrelationsfunktion von ($Y(t)$) bezeichnet:

$$E(Y(t) - Y(t - \tau))^2 = 2\gamma(0) - 2\gamma(\tau) = 2\gamma(0)(1 - \rho(\tau)) .$$

Das auf diesem Resultat aufbauende *Variogramm*

$$V(\tau) = \frac{1}{2}E(Y(t) - Y(t - \tau))^2 = \gamma(0)(1 - \rho(\tau)) \quad (12.2)$$

bildet den Schlüssel zur Schätzung der ACF.

Die Schätzung aus einer Zeitreihe $y(t_i), t_1 < \dots < t_N$ geschieht so, dass zunächst die $\frac{1}{2}N(N - 1)$ empirischen Größen

$$v(t_i - t_j) = \frac{1}{2}(y_{t_i} - y_{t_j})^2 \quad (t_i > t_j, i, j = 1, \dots, N)$$

gebildet werden. Alle $v(t_i - t_j)$ werden in Abhängigkeit vom Beobachtungsabstand $t_i - t_j$ grafisch dargestellt. Dies ergibt das *empirische* oder *Stichproben-Variogramm*. Falls

gleiche Zeitabstände $t_i - t_j$ mehrfach vorkommen, werden die zugehörigen $v(t_i - t_j)$ gemittelt. Gibt es nur wenige Zeitabstände, bei denen die Mittelung vorgenommen werden kann, so weist das empirische Variogramm erratiche Ausschläge auf. Am besten ist es dann, eine Glättung vorzunehmen. Dies kann etwa über eine lokale Mittelung oder mit dem Splines-Ansatz geschehen. Die resultierenden Größen seien mit $\tilde{v}(\tau_k)$ bezeichnet.

Da für stationäre Prozesse die ACF mit zunehmendem Lag gegen Null geht, tendieren die $\tilde{v}(\tau_k)$ dazu, sich mit wachsendem τ_k bei einer Konstanten zu stabilisieren. Liegt dieses Erscheinungsbild vor, so kann die Konstante aufgrund von (12.2) einfach durch

$$\hat{y}(0) = \frac{1}{N} \sum_{i=1}^N (y(t_i) - \bar{y})^2 \quad (12.3)$$

geschätzt werden. Das führt dann zu den Schätzungen

$$\hat{\rho}(\tau_k) = 1 - \frac{\tilde{v}(\tau_k)}{\hat{y}(0)}. \quad (12.4)$$

Zeigen die $\tilde{v}(\tau_k)$ andererseits keine Stabilisierungstendenz, so weist dies entweder auf eine weit reichende Abhängigkeit oder auf Instationarität hin.

Beispiel 12.3 (Gehalt eines Oxygen Isotops – Fortsetzung von Seite 292)

Für die Oxygen-Isotop-Daten wird das Variogramm für die Lags bis 100 bestimmt. Da die Reihe sehr irregulär ist, gibt es nicht viele gleiche Abstände $t_i - t_k$. Daher ist eine Glättung notwendig. Dies ergibt die Abbildung 12.3. Aus dem Variogramm könnte nun eine Darstellung der Autokorrelationsfunktion gewonnen werden. Erst diese gibt dann in gewohnter Weise Aufschluss über die Abhängigkeiten.

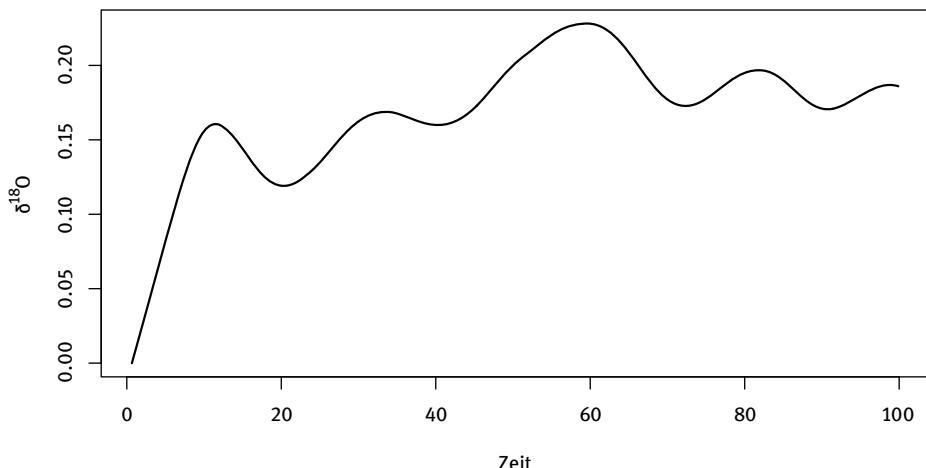


Abb. 12.3: Variogramm ($\tilde{v}(k)$) für die Reihe OXYGEN

R-Code:

```

1 library(nlme)
2 d <- Variogram(y,dist(t))
3 attach(d)
4 s <- smooth.spline(dist,variog)
5 x <- s$x[s$x<100]
6 y <- s$y[s$x<100]
7 plot(x,y,type="l")

```

Der Code setzt den vom Beispiel 12.2 fort. Um das Variogramm der Reihe OXYGEN zu bestimmen, wird die Funktion `Variogram` aus dem Paket `nlme` verwendet. Das Ergebnis der Funktion ist eine Liste, die unter `dist` die Abstände und unter `variog` die zugehörigen Werte des Semi-Variogramms enthält. Diese werden noch mit einem glättenden Spline geglättet.

Um in dieser Datensituation lineare Zeitreihenmodelle an Zeitreihen anzupassen, geht man von einem stetigen Zeitparameter aus. Dann ist einiger zusätzlicher theoretischer Aufwand nötig. Schon die Definition eines White-Noise-Prozesses in stetiger Zeit ist nicht einfach. Für die Schätzung von ARMA-Modellen mit stetigem Zeitparameter kann dann die Maximum-Likelihood-Methode eingesetzt werden.

12.3 Ausreißer und robuste Verfahren

Einzelne extreme Werte in Zeitreihen werden als Ausreißer bezeichnet. Weist eine Zeitreihe Ausreißer auf, so wird die Bestimmung der deterministischen glatten Komponente oder der Saisonkomponente mit der Methode der kleinsten Quadrate u. U. stark beeinflusst. Das hängt damit zusammen, dass dabei die einzelnen quadratischen Abweichungen alle gleich gewichtet eingehen. Folglich ergeben beispielsweise 100 Abstände der Größe 1 das gleiche Gewicht wie ein Abstand von 10. Wegen der Auswirkungen von Ausreißern sind z. B. bei den Saisonbereinigungsverfahren Schritte zur Erkennung und Bereinigung von Ausreißern eingebaut; siehe z. B. die Beschreibung des Census X-11 Verfahrens. Natürlich kann auch die Identifizierung und Schätzung von ARIMA-Modellen stark beeinflusst werden. Generell muss man sich also vor Ausreißern schützen. Die R-Funktion `stl` basiert ganz auf robusten Verfahren, auf solchen die von vornherein unempfindlich gegen Ausreißer sind.

Ausreißer

In Zeitreihen können zwei Typen von Ausreißern vorkommen. Das eine sind die so genannten *additiven Ausreißer*, oder AO's (=additive Outlier's). Dies sind Werte, die

an einzelnen Stellen vom sonstigen Gesamtverlauf abweichen; im Prinzip können sie schon aus dem Plot der Reihe erkannt werden. *Innovative Ausreißer*, IO's, sind dagegen solche, die bei linearen Zeitreihenmodellen extreme Innovationen hervorbringen. IO's haben daher einen Nachwirkungseffekt. Es ist das Verdienst von Fox (1972), darauf hingewiesen zu haben, dass bei Zeitreihen zwei Typen von Ausreißern zu unterscheiden sind.

Beispiel 12.4 (AR-Prozess)

Sei (Y_t) ein AR[p]-Prozess. Um die Situation eines einzelnen additiven Ausreißers zum Zeitpunkt t_0 zu erfassen, wird die Modellgleichung

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

als Systemgleichung im Sinne eines Zustandsraummodells aufgefasst, vergleiche (10.1), und zusätzlich eine Beobachtungsgleichung eingeführt:

$$Z_t = \begin{cases} Y_t & \text{für } t \neq t_0 \\ Y_t + \omega & \text{für } t = t_0 . \end{cases} \quad (12.5)$$

Ein Ausreißer vom IO-Typ bei t_0 wird dagegen folgendermaßen im Modell berücksichtigt:

$$Y_t = \begin{cases} \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t & \text{für } t \neq t_0 \\ \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \omega + \varepsilon_t & \text{für } t = t_0 . \end{cases} \quad (12.6)$$

Die Auswirkungen einzelner AO's auf die ACF sind i. d. R. so, dass die Autokorrelationen gedämpft werden. Dies wird dadurch bewirkt, dass im Nenner die Quadrate der Differenzen zum Mittelwert vorkommen, während im Zähler die einfachen Differenzen mit anderen, nicht so extremen multipliziert werden. Bei der Anpassung von ARMA-Modellen auf der Basis der ACF werden folglich die Modellordnungen unterschätzt.

Innovative Ausreißer haben nicht diese Auswirkungen, da ihr Nachwirkungseffekt für eine Art Selbstheilung sorgt. AR-Prozesse lassen sich dann sogar besser schätzen!

Von Chang, Chen & Tiao (1988) wurden zweit Teststatistiken λ_{1,t_0} und λ_{2,t_0} vorgeschlagen, mit denen sich das Vorliegen eines additiven bzw. innovativen Ausreißers zum Zeitpunkt t_0 überprüfen lässt. Sei an die Zeitreihe ein ARMA-Modell angepasst und seien π_1, π_2, \dots die AR-Koeffizienten der AR[∞]-Darstellung des Modells, so dass:

$$\varepsilon_t = Y_t - \pi_1 Y_{t-1} - \pi_2 Y_{t-2} - \cdots .$$

Werden die geschätzten Größen wieder mit einem ‚Dach‘ versehen, so führt bei additiven Ausreißern (12.5) zu der Beziehung

$$\begin{pmatrix} \hat{\varepsilon}_1 \\ \vdots \\ \hat{\varepsilon}_{T-1} \\ \hat{\varepsilon}_T \\ \hat{\varepsilon}_{T+1} \\ \hat{\varepsilon}_{T+2} \\ \vdots \\ \hat{\varepsilon}_N \end{pmatrix} = \omega \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ -\hat{\pi}_1 \\ -\hat{\pi}_2 \\ \vdots \\ \hat{\pi}_{N-T} \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_{T-1} \\ a_T \\ a_{T+1} \\ a_{T+2} \\ \vdots \\ a_N \end{pmatrix}. \quad (12.7)$$

ω kann mittels einer linearen Regression bestimmt werden. Bei einem innovativen Ausreißer ist der Designvektor auf der rechten Seite der Gleichung einfach der Vektor $I_t^{(T)}$, wobei $I_t^{(T)} = 1$ für $t = T$ und = 0 sonst. Der Kleinsten-Quadrate-Schätzer von ω ist damit einfach $\hat{\omega} = \hat{\varepsilon}_T$. Das führt zu den Teststatistiken

$$\lambda_{1,T} = \frac{\hat{\varepsilon}_T - \sum_{j=1}^{N-T} \hat{\pi}_j \hat{\varepsilon}_{T+j}}{\sqrt{\hat{\sigma}_{\varepsilon}^2 \sum_{j=0}^{n-T} \hat{\pi}_j^2}} \quad \text{und} \quad \lambda_{2,T} = \frac{\hat{\varepsilon}_T}{\sqrt{\hat{\sigma}_{\varepsilon}^2}}. \quad (12.8)$$

Beide Teststatistiken sind im Fall eines Normalprozesses unter der Nullhypothese, dass kein Ausreißer zum Zeitpunkt T vorliegt, $\mathcal{N}(0, 1)$ -verteilt.

Wei (1990) hat eine *iterative Prozedur zur Erkennung von Ausreißern* unter Verwendung dieser Statistiken vorgeschlagen. Eine Variante davon geht folgendermaßen:

1. Der (stationären) Zeitreihe ein ARMA-Modell angepasst und die AR-Koeffizienten der AR[∞]-Darstellung des Modells bestimmt. Die Residualvarianz der Innovativen $\hat{\varepsilon}_t$ sei $\hat{\sigma}_{\varepsilon}^2$.
2. Für $t = 1, 2, \dots, N$ werden $\hat{\lambda}_{1,t}$ und $\hat{\lambda}_{2,t}$ unter Verwendung des Residuen und der Parameter des geschätzten Modells bestimmt. Das Maximum sei

$$\hat{\lambda}_T = \max_t \max_i \{|\hat{\lambda}_{i,t}|\},$$

wobei T den Zeitindex bezeichne, bei dem das gesamte Maximum auftritt. Ist $\hat{\lambda}_T > c$, wobei c eine fest gewählte Konstante ist, etwa $c = z_{1-\alpha/2}$, so werden die Residuen bereinigt. Bei einem additiven Ausreißer wird $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \hat{\omega} \hat{\pi}(B) I_t^{(T)}$ gebildet, bei einem innovativen $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \hat{\omega} I_t^{(T)}$.

Mit den modifizierten Residuen wird ein neuer Schätzwert $\hat{\sigma}_{\varepsilon}^2$ bestimmt.

3. Es wird zu Schritt zwei zurückgegangen und das Vorgehen für die t , bei denen bisher kein Ausreißer festgestellt wurde, unter Verwendung der modifizierten Residuen iteriert. Die Iteration hört nach dem Schritt auf, in dem kein neuer Ausreißer identifiziert wurde.

4. Es wird zum Abschluss das ARMA-Modell neu geschätzt. Die Ausreißer werden gemäß als feste Regressoren gemäß (12.7) bzw. gemäß der entsprechenden Beziehung für innovative Ausreißer berücksichtigt.

Der Vorteil des Vorgehens liegt darin, dass am Ende der Iteration sowohl die einzelnen Ausreißer ermittelt als auch bei der Modellanpassung berücksichtigt sind.

Beispiel 12.5 (Hormonausschüttung)

Die Identifizierung von Ausreißern unterschiedlichen Typs wurde zur Bestimmung von Zeitpunkten der Hormonausschüttung in hormonalen Zeitreihen ausgenutzt. Die typischen steilen Anstiege in solchen Reihen resultieren von der Ausschüttung des Hormons in das Blutplasma; die folgenden langsameren Abfälle kommen durch biochemische Abbau-Prozesse zustande. Solche Hormonausschüttungen wirken in den Zeitreihen wie IO's. Zusätzlich ist aber zu berücksichtigen, dass auch einfache Messfehler vorkommen. Diese können als AO's betrachtet werden.

Thomas, Plu & Thalabard (1992) betrachten eine Reihe LUHORMON des im 15 Minuten-Abstand bei einer Kuh erhobenen Luteinzing-Hormons. Sie schlagen ein spezielles Identifizierungsverfahren vor, wobei sie als Basismodell von einem AR[1]-Prozess ausgehen. Hier lässt sich aber auch die iterative Ausreißer-Erkennungsprozedur einsetzen. (Die hier betrachtete Reihe wurde aus der genannten Veröffentlichung (Abbildung 1) rekonstruiert.)

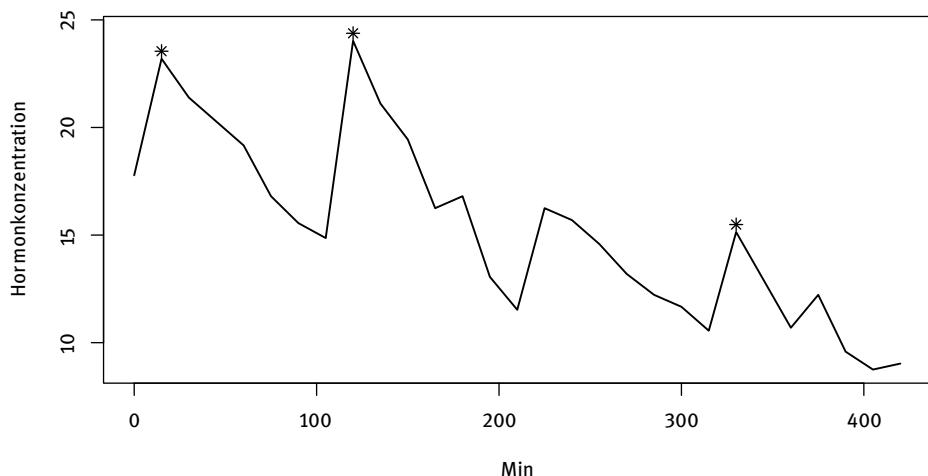


Abb. 12.4: Ausreißererkennung in der Reihe LUHORMON

Die Anwendung des iterativen Verfahrens auf die Reihe LUHORMON führt zu dem Ergebnis:

ind	type	lambda
9	2	3.756
2	2	2.984
23	1	2.329

Es werden also drei Ausreißer identifiziert, zwei vom IO-Typ und ein additiver Ausreißer. Diese sind in der Abbildung markiert. Die mit einem Stern markierten Stellen sind die entdeckten Ausreißer.

Die abschließende Gesamtschätzung ergibt das Modell

Coefficients:

ar1	intercept	2	9	23
0.8429	14.5343	3.6833	6.0918	1.4477
s.e.	0.0994	2.0707	1.5616	1.5600
sigma^2 estimated as 4.154: loglikelihood = -62.42 ,				
aic = 136.83				

Damit ist der additive Ausreißer nicht mehr signifikant. Es verbleiben die beiden IO's als relevante Einflüsse. Zu diesen Zeitpunkten (und an zwei weiteren) wurden mit einer anderen Methode Hormonausschüttungen festgestellt. Die Autoren erhielten mit ihrer Methode zudem einen weiteren IO-Ausreißer bei 300 Minuten. Dieser konnte als Zeitpunkt einer Hormonausschüttung nicht verifiziert werden. Zusätzlich entdeckten sie zwei (scheinbare) AO's.

Der AIC-Wert beim ursprünglichen AR[1]-Modell beträgt übrigens 147.23, die Schätzung der Residualvarianz 7.403. Beide Werte haben sich bei Berücksichtigung der Ausreißer deutlich verbessert.

R-Code:

```

1 data(LUHORMONE)
2 t <- LUHORMONE[,1]
3 y <- LUHORMONE[,2]
4 out <- arima(y,order=c(1,0,0))
5 out1 <- outidentify(y,out,alpha=0.05)
6 plot(y,type="l")
7 points(out1$outlier[,1],y[out1$outlier[,1]]+0.35,pch=8)
8 out1$arima.out

```

Die Daten werden eingelesen; es wird gleich ein AR[1]-Modell angepasst, da dieses Modell in der Quelle gewählt wurde. Der Aufruf von `outidentify` ergibt in der Matrix `out1$outlier` die Angaben zu den identifizierten Ausreißern. Das wird zur grafischen Darstellung ausgenutzt. Die letzte Zeile ergibt dann die Ausgabe der abschließenden ARMA-Schätzung.

Ist schon aus anderen Gründen oder aus dem Plot der Zeitreihe klar, an welcher Stelle ein Wert u. U. als Ausreißer anzusehen ist, kann eine einfache Interventionsanalyse helfen zu entscheiden, ob der Wert tatsächlich so extrem ist, dass er als Ausreißer zu betrachten ist.

Beispiel 12.6 (Fichteneinschlag)

Fichtenholz wird in der Bundesrepublik vor allem zu Zwecken der Papierherstellung und als Bauholz geschlagen. Die Entwicklung des Einschlags wird u. a. durch die Nachfrage, den Import und das Altpapieraufkommen beeinflusst. Die allgemeine Entwicklungstendenz wird punktuell von Schadensereignissen überlagert. Dazu gehören Käferkalamitäten sowie Stürme.

Von besonderem Ausmaß war der Jahrhundertsturm von 1990, der nach dem regulär im Winter vorgenommenen Einschlag auftrat und daher nicht durch einen geringeren Einschlag ausgeglichen werden konnte. Dieser Sturm soll als einmalige Intervention in ein Modell für diese Zeitreihe einbezogen werden. Die Intervention hat also die Form

$$I_t = \begin{cases} 0 & \text{für } t \neq 1990 \\ 1 & \text{für } t = 1990 \end{cases}$$

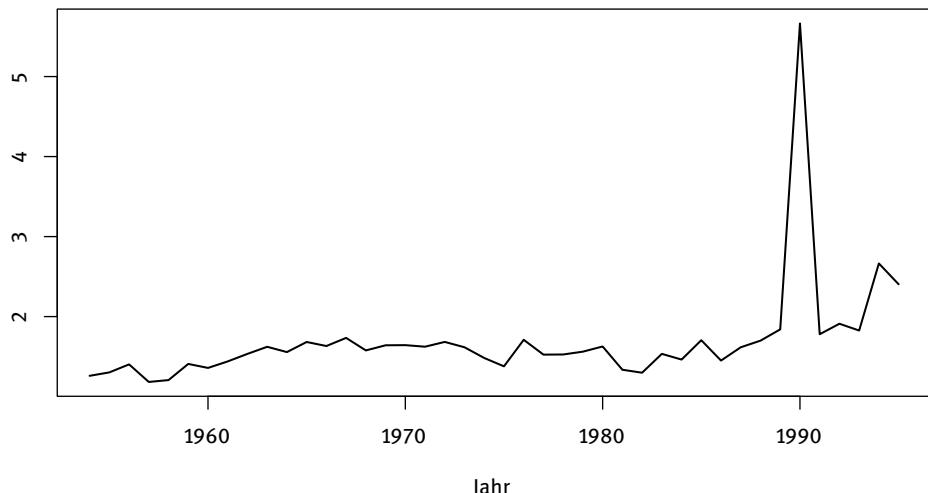


Abb. 12.5: Einschlagmengen von Fichtenstammholz in Bayern (in Mio.)

Da der Plot der Zeitreihe FICHTE einen leichten Trend zeigt, wird zu den ersten Differenzen übergegangen.

Die Kennfunktionen der bis 1989 betrachteten Differenzen $(1 - B)Y_t$ legen ein MA[1]-Modell nahe. Die Schätzung ergibt dann für das Modell $(1 - B)(Y_t - \omega I_t) = \mu + (1 - \beta)\epsilon_t$ aus der gesamten Reihe:

Coefficients:

ma1	Imp
-0.3739	3.8516
s.e.	0.1341 0.1480

```
sigma^2 estimated as 0.0318: log likelihood = 12.13,
                                aic = -18.27
```

Die Residuen zeigen dann keine Korrelation mehr. Der für verschiedene m berechnete Ljung-Box-Pierce-Test hat durchweg P -Werte, die größer als 0.5 sind. Somit ist das Modell angemessen. Der Koeffizient von I erlaubt die Quantifizierung der Auswirkung des Sturmes; er gibt an, wie viel *mehr* Fichtenholz in diesem Jahr eingeschlagen wurde.

R-Code:

```
1 data(SPRUCE)
2 y <- SPRUCE
3 t <- c(tsp(y)[1]:tsp(y)[2])
4 plot(y)
5 acfpacf(diff(y[1:35]),10)
6 Imp <- rep(0,length(y))
7 Imp[t == 1990] <- 1
8 out <- arima(y,order=c(0,1,1),xreg=Imp)
9 out
```

Nach dem Einlesen wird mit Zeile 5 das Modell identifiziert. Bei der Erzeugung der Interventionsvariablen wird ausgenutzt, dass es sich um Jahreszahlen handelt, das Zeitreihenattribut frequency gleich eins ist. Dann ist die Schätzung einfach.

Robuste Verfahren

Anstatt sich um Ausreißer einzeln zu kümmern, kann man auch den Ansatz wählen, Zeitreihen robust zu analysieren. Man spricht von einem *robusten* oder genauer von einem *Ausreißer-resistenten Verfahren*, wenn es durch Ausreißer relativ wenig beeinflusst wird.

Eine einfache Variante der robusten Analyse von Zeitreihen, besteht darin, die extremen Werte an die Reihe heranzuschieben, die Reihe zu winsorisieren. Dafür benötigt man zuerst eine robuste Niveauschätzung sowie eine robuste Skalenschätzung.

Als Niveauschätzer kann man den Median nehmen. Ein robustes Streuungsmaß ist der *MAD*, der *Median der absoluten Abweichungen vom Median*,

$$\hat{\sigma} = 1.483 \cdot \text{median}\{|y_t - \text{median}\{y_s\}|\}. \quad (12.9)$$

Der Faktor 1.483 macht den Schätzer im Fall Gaußscher White-Noise-Prozesse erwartungstreu für σ_Y .

Bei der Winsorisierung werden die Zeitreihenwerte, die außerhalb des dreifachen Streubereiches um den Median liegen, auf diese Grenze gesetzt:

$$y_t^* = \begin{cases} \tilde{y} - 3 * \hat{\sigma} & \text{für } y_t < \tilde{y} - 3 * \hat{\sigma} \\ y_t & \text{für } \tilde{y} - 3 * \hat{\sigma} \leq y_t \leq \tilde{y} + 3 * \hat{\sigma} \\ \tilde{y} + 3 * \hat{\sigma} & \text{für } y_t > \tilde{y} + 3 * \hat{\sigma} \end{cases}.$$

Dieses Vorgehen ist ähnlich dem im CENSUS-Verfahren implementierten, nur dass hier robuste Lage- und Streuungsparameter verwendet werden.

Ein anderer Zugang ist die Übertragung von *robusten Schätzverfahren* auf Zeitreihen. Dies wird anhand der M-Schätzer kurz skizziert.

Bei einer univariaten Stichprobe y_1, \dots, y_N erhält man das arithmetische Mittel \bar{y} als Lösung der Minimierungsprobleme

$$\sum_{t=1}^N (y_t - a)^2 \stackrel{!}{=} \min.$$

Das arithmetische Mittel reagiert deswegen sehr stark auf Ausreißer, weil extreme Werte quadratisch in das Zielkriterium eingehen. Dementsprechend liegt es nahe, die quadratische Abstandsfunktion durch eine langsam mit u wachsende $\rho(u)$ zu ersetzen. Man erhält dann robuste Lageparameter über den Minimierungsansatz

$$\sum_{t=1}^N \rho((y_t - a)/\sigma) \stackrel{!}{=} \min. \quad (12.10)$$

Für die Lösung wird die Minimierungsaufgabe unter Setzung von $\psi(e; \beta_j) = \rho'(e; \beta_j)$ transformiert in die in der Regel iterativ zu lösenden Gleichungen

$$\sum_{t=1}^N \psi((y_t - a)/\sigma) = 0. \quad (12.11)$$

Die resultierenden Schätzer werden als *M-Schätzer* bezeichnet. Die Aufnahme eines Skalenparameters σ ist hier nötig, da bei den meisten der üblichen ρ -Funktionen die Entfernung von y_t zu a sonst nicht adäquat erfasst werden kann. σ wird robust durch den MAD geschätzt.

Über eine geeignete Wahl der Funktion ρ bzw. ψ versucht man nun, zwei Ziele möglichst gleichzeitig zu erreichen. Einmal eine große Unempfindlichkeit gegen Ausreißer. Dazu müssen extreme Werte u herunter gewichtet werden. Zum anderen möchte man Schätzer, die im Fall der Normalverteilung keine wesentlich größere Varianz als \bar{y} aufweisen, die also möglichst effizient sind. Zudem wird für die Lösung des Problems (12.10) eine Startschätzung benötigt. Dazu nimmt man meist den Median, da er sehr unempfindlich gegen Ausreißer ist. Einige ρ - und ψ -Funktionen sind in der folgenden Tabelle aufgeführt. In der Abbildung 12.6 sind sie grafisch dargestellt.

Tab. 12.1: ρ - und ψ -Funktionen verschiedener M-Schätzer

Schätzer	$\rho(u)$	$\psi(u)$	Grenzen
KQ	$u^2/2$	u	$ u < \infty$
LAD	$ u $	$\text{sign}(u)$	$ u < \infty$
Huber	$u^2/2$ $k u - \frac{1}{2}k^2$	u $k \cdot \text{sign}(u)$	$ u \leq k$ $ u > k$
Biweight	$B^2/6\{1 - (1 - (u/B)^2)^3\}$ $B^2/6$	$u(1 - (u/B)^2)^2$ 0	$ u \leq B$ $ u > B$

Empfehlungen für die Konstanten: $k = 1.345$ und $B = 4.685$

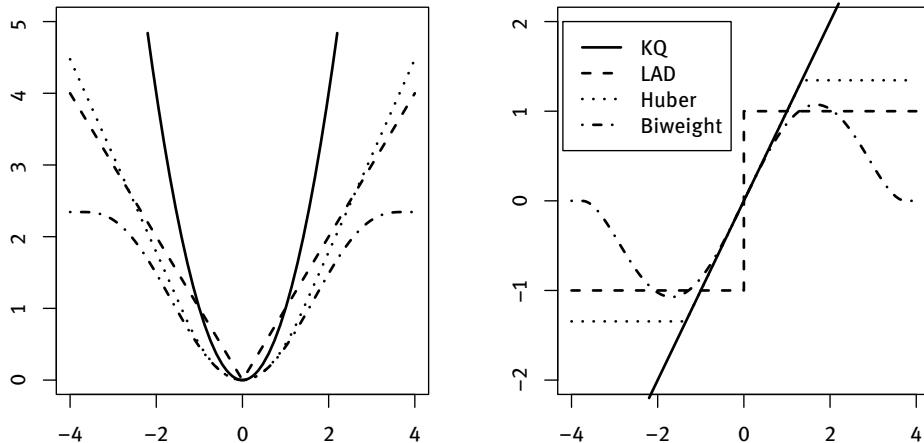


Abb. 12.6: Einige ρ -Funktionen (links) und zugehörige ψ -Funktionen (rechts)

Der Ansatz der M-Schätzer lässt sich nun auf verschiedene Probleme der Zeitreihenanalyse übertragen. Zunächst wird das Regressionsmodell

$$Y_t = \beta_1 m_1(t) + \cdots + \beta_k m_k(t) + U_t ,$$

mit bekannten Regressoren $m_1(t), \dots, m_k(t)$ betrachtet. Hier lautet das Zielkriterium für die Schätzung des Parametervektors β , das auf der Abstandsfunktion ρ basiert:

$$\sum_{t=1}^N \rho\left(\frac{y_t - \beta_1 m_1(t) - \dots - \beta_k m_k(t)}{\hat{\sigma}}\right) \stackrel{!}{=} \min_{\beta}.$$

Mit diesem Ansatz können beispielsweise deterministische Trends robust bestimmt werden.

Der Zugang eröffnet auch eine Möglichkeit, die Splines-Glättung zu robustifizieren. In dem Zielkriterium, siehe Gleichung (2.28), werden einfach entsprechend der M-Schätzung die Abstandsquadrate $(y_t - g_t)^2$ durch andere Abstände $\rho(y_t - g_t)$ ersetzt.

Beispiel 12.7 (Fichteneinschlag – Fortsetzung von Seite 300)

Die Zeitreihe FICHTE weist einen gewissen Anstieg auf, siehe die Abbildung 12.5. Die glatte Komponente wird mit der robusten Splines-Glättung bestimmt. Zum Vergleich ist die nicht-robust ermittelte glatte Komponente ebenfalls eingezeichnet. Im Beispiel 12.6 wurde der Zeitreihe ein MA[1]-Modell angepasst, wobei ein Impuls als Regressor berücksichtigt wurde. Es bleibt also zu untersuchen, ob die Abweichungen von der glatten Komponente noch eine Autokorrelation aufweisen. Um den Impuls ebenfalls zu berücksichtigen, wird eine lineare Regression der Differenz der Zeitreihe und glatter Komponente gegen den Impuls gerechnet. Die ACF und PACF der Reihe der Residuen zeigen keine serielle Korrelationen.

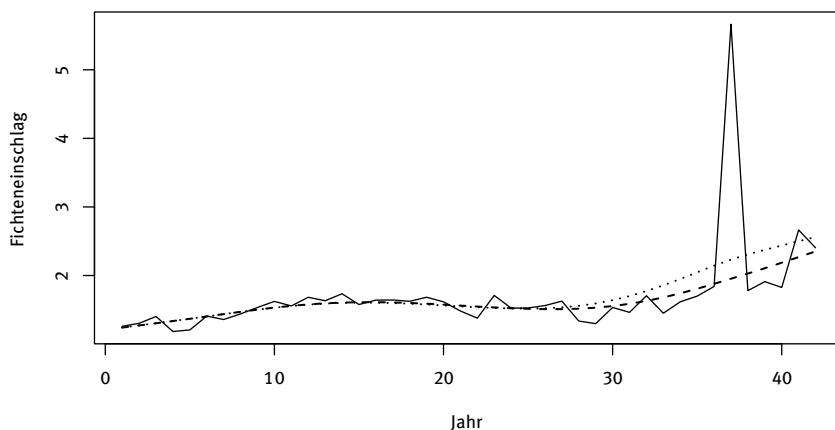


Abb. 12.7: Fichteneinschlag mit glatter Komponente: - - - robust, ⋯⋯⋯ kleinste Quadrate

R-Code:

```

1 g <- smoothrb(y, 15, 5)
2 out <- lm((y-g)~Imp)
3 acfpacf(out$resid, 15)

```

Aus dem Beispiel 12.6 stehen y , Imp noch zur Verfügung. So kann das im Text beschriebene Vorgehen hier mit der Funktion `smoothrb` aus dem Paket `tsapp` entsprechend umgesetzt werden.

Saisonbereinigungsverfahren sollten in der einen oder anderen Form mögliche Ausreißer berücksichtigen. In R steht das Verfahren *Seasonal Decomposition of Time Series by Loess*, kurz *STL*, zur Verfügung, das von vornherein auf robusten Methoden basiert. Die Funktion `stl` führt die Zeitreihenzerlegung kurz gefasst folgendermaßen durch. Der saisonbedingte Bestandteil wird durch den Einsatz eines robusten Glätters bestimmt. Damit werden die saisonbedingten Teilfolgen glättet; dies sind bei Monatsdaten die Folgen aller Januar-Werte, die aller Februar-Werte usw. Die Saisonwerte werden aus der Reihe entfernt, der Rest wird geglättet, um den Trend zu bestimmen. Das Saisonniveau wird von den Saisonkomponenten abgezogen und dem Trend zugeschlagen. Dieser Prozess wird einige Male iteriert. Den Rest bilden dann die Residuen vom ‚Saison plus Trend Fit‘.

Beispiel 12.8 (Auftragseingänge für Maschinen – Fortsetzung von Seite 64)

Da es bei der Zeitreihe **MASCHINEN**, den Auftragseingängen für Werkzeugmaschinen, einige Monate mit extremen Auftragseingängen gibt, ist bei der Bestimmung von glatter und Saisonkomponente darauf zu achten, dass Ausreißer adäquat behandelt werden. Dies legt den Einsatz eines robusten Verfahrens wie dem hier verwendeten *STL* nahe. Man sieht im untersten Teil der Abbildung 12.8, dass die extremen Werte alle den Resten zugeschlagen werden. Damit sollten die glatte Komponente und die Saisonkomponente recht zuverlässig bestimmt sein.

R-Code:

```
1 data(ENGINES)
2 y <- ENGINES
3 out <- stl(y,s.window=9)
4 plot(out)
```

Bei der im standardmäßig zur Verfügung stehenden Paket `stats` implementierten Funktion `stl` ist neben einer Zeitreihe, die das Attribut einer solchen haben muss, noch anzugeben, über wie viele Zeitpunkte die Saison geglättet werden soll.

Auch die in Abschnitt 2.4 angesprochene Zerlegung mittels glättender Splines lässt sich robustifizieren, siehe Schlittgen (1991b). Dazu wird die Minimierungsaufgabe (2.30) modifiziert zu

$$\lambda \sum_{t=3}^N (g_t - 2g_{t-1} + g_{t-2})^2 + \kappa \sum_{t=s}^N \left(\sum_{\tau=0}^{s-1} S_{t-\tau} \right)^2 + \sum_{t=1}^N \rho((y_t - g_t - S_t)/\hat{\sigma}) \stackrel{!}{=} \min. \quad (12.12)$$

Hierbei ist $\rho(u)$ eine der obigen Abstandsfunktionen.

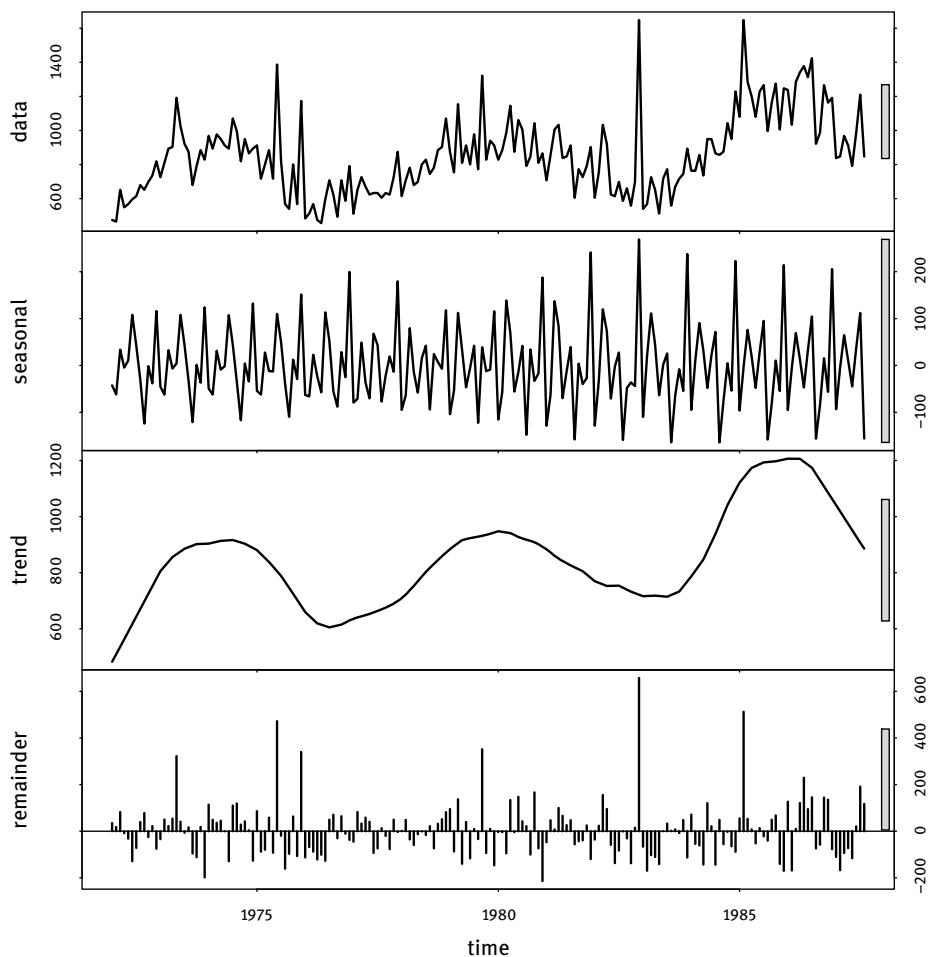


Abb. 12.8: STL-Zerlegung der Zeitreihe MASCHINEN

Beispiel 12.9 (Auftragseingänge für Maschinen – Fortsetzung von Seite 305)

Die Zerlegung der Zeitreihe MASCHINEN mittels des robustifizierten Splines-Ansatzes bringt ein ähnliches Ergebnis wie STL. Allerdings erscheint die glatte Komponente als glatter, die Saisonkomponente macht einen regelmäßigeren Eindruck. Der Unterschied röhrt daher, dass STL lokal arbeitet, während die Splines-Zerlegung ein Gesamtansatz ist.

R-Code:

```
1 gs <- robsplinedecom(y,12,20,5,Plot=T)
```

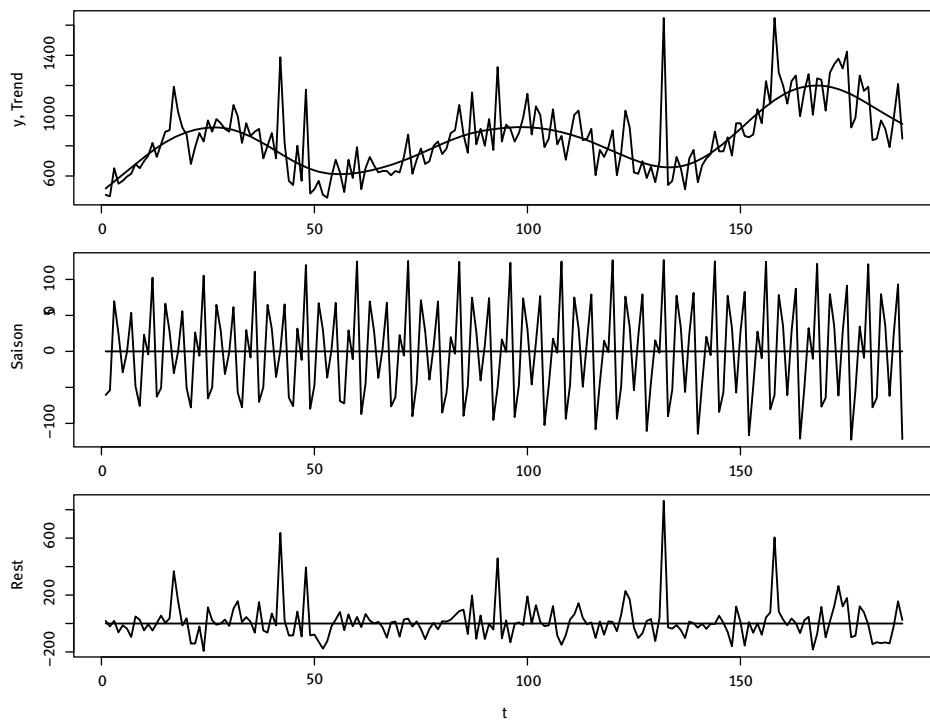


Abb. 12.9: Robuste Splines-Zerlegung der Zeitreihe Auftragseingänge für Werkzeugmaschinen

Die Reihe steht bereits zur Verfügung. Die Zeitreihenattribute gehen aber bei der in tsapp implementierten Funktion verloren. Der Hintergrund dafür ist das Paket Matrix. Mit den dort implementierten Funktionen lässt sich die Berechnung platzsparend durchführen. Allerdings erlauben sie keine Zeitreihenattribute.

Autoregressive Modelle können als direkte Übertragungen der klassischen Regressionsmodelle auf Zeitreihen angesehen werden. Daher liegt der Zugang zur robusten Schätzung von AR-Modellen mittels des Ansatzes der M-Schätzer auf der Hand. Für das AR[p]-Modell

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

führt dies zu dem Minimierungsproblem

$$\sum_{t=p+1}^N \rho\left(\frac{1}{\hat{\sigma}}(y_t - \alpha_1 y_{t-1} - \cdots - \alpha_p y_{t-p})\right) \stackrel{!}{=} \min .$$

Da hier die Ausreißer nicht nur in der abhängigen Variablen (Y_t) sondern auch in den „unabhängigen“ (Y_{t-1}, \dots, Y_{t-p}) vorkommen, sind M-Schätzer für AR-Modelle im Fall von AO's nicht mehr robust.

Um robuste Schätzer von AR-Modellen aus Reihen mit AO-Ausreißern zu erhalten, kann man das aus der partiellen Ableitung der Zielfunktion resultierende Gleichungssystem für die M-Schätzer so modifizieren, dass der Effekt einzelner extremer y -Werte neutralisiert wird. Die resultierenden *generalisierten M-Schätzer*, kurz GM-Schätzer, besitzen bei AO's gute Eigenschaften. Bei IO's sind sie besser als die Kleinstes-Quadrat-Schätzer, haben jedoch größere Streuungen als die einfachen M-Schätzer.

Bei ARMA-Modellen wird die Situation wirklich schwierig, da auch GM-Schätzer nicht mehr robust sind. Schon ein einzelner Ausreißer kann in einer Weise nachwirken, dass die Schätzung zusammenbricht. Anders als bei AR-Prozessen hängen nämlich die zu modifizierenden Faktoren in den Normalgleichungen von allen vergangenen y -Werten und zusätzlich von den zu bestimmenden Parametern ab.

Neben der erwähnten iterativen Prozedur stellt die rekursive generalisierte M-Schätzung, kurz RGM-Schätzung, einen praktikablen Weg der robusten Schätzung dar. Dies ist eine Robustifizierung des Ansatzes von Durbin-Hannan-Rissanen zur Schätzung von ARMA-Modellen; siehe dazu Allende & Heiler (1992).

12.4 R-Funktionen

Die Schätzung der fehlenden Werte in Zeitreihen ist mit den beiden Funktionen `missar` und `missls` möglich. Sie gehören zu den Funktionen der Datei `tsapp`. Das Paket `imputeTS` bietet ausführliche Funktionalitäten für das Ersetzen (Imputation) fehlender Werte.

Irregulär beobachtete Zeitreihen können als Objekte vom Typ `zoo` mit dem Paket `zoo` analysiert werden. `zoo`-Objekte haben neben den Zeitreihenwerten selbst einen vollständigen Zeitindex, nicht nur den Beginn, das Ende und die Frequenz wie die `ts`-Objekte. Das Variogramm lässt sich mit der Funktion `Variogram` des Paketes `n1me` bestimmen. Umfangreiche Möglichkeiten zur Analyse geologischer Zeitreihen bietet das Paket `gstat`.

Die Funktion `outidentify` zur iterativen Ausreißererkennung befindet sich in dem Paket `tsapp`. Innovative Ausreißer können auch mit der Funktion `detectIO` und additive mit `detectAO` aus dem Paket `TSA` aufgespürt werden. Allerdings ist bei diesen beiden Funktionen eine α -Adjustierung implementiert, die das Niveau der einzelnen Tests auf α/N senkt. Dies führt dazu, dass nur sehr extreme Ausreißer erkannt werden. Die Funktion `tsoutliers` aus dem Paket `forecast` ist für das Identifizieren und Ersetzen von Ausreißern geschrieben, `tsclean` füllt gleichzeitig auch die fehlenden Werte auf. Ein auf die Ausreißerproblematik spezialisierte Paket ist `tsoutliers`.

`smoothrb` aus dem Paket `tsapp` ist eine robustifizierte Variante eines glättenden Splines. Eine robuste Zerlegungsmethode wird mit `stl` im Paket `stats` bereit gestellt. Ebenfalls in `tsapp` ist `robsplinedecom` vorhanden.

13 Einführung in R¹

Hier kann nur eine knappe Einführung in R gegeben werden. Es gibt aber einiges an Literatur, das zusätzlich konsultiert werden kann. Da die Syntax von R der von S entspricht, lassen sich auch dafür konzipierte Texte verwenden. Ein paar Texte sind Chambers & Hastie (1992), Dalgaard (2002), Süsselbeck (1993), Venables (2000) und Venables & Ripley (1994). Zudem sind im Internet verschiedene, auch deutschsprachige Skripten zu R und S zu finden.

13.1 Erste Schritte

R wird betreut von dem R-Core Team, siehe R Development Core Team (2011). Die Nutzergemeinschaft von R ist sehr groß. Die R-Gemeinschaft ist sehr aktiv; viele haben zusätzliche Pakete, Bündel von Funktionen, geschrieben, die größtenteils von der Projekt-Seite bezogen werden können. Die Aktivität bringt recht häufige Aktualisierungen mit sich, die bisweilen recht umfangreich sein können. Zur Zeit der Fertigstellung ist die Version 3.6.2 aktuell.

Die R-Software ist eine freie Software unter der GNU-Lizenz und ist unter diesen Bedingungen kostenfrei aus dem Internet zu beziehen. Die Adresse lautet cran.r-project.org/. R läuft auf einer Vielzahl von Plattformen, insbesondere Linux, Mac OS und Windows. Um R zu beziehen, gehen Sie auf die angegebene Web-Seite, laden Sie sich die geeignete Version herunter und installieren Sie sie. Die Beschreibung hier ist auf das Windows Betriebssystem beschränkt. Dafür gibt es eine exe-Datei, zu der man durch Anklicken von „Download R for Windows“ gelangt. Die heruntergeladene Datei wird ausgeführt und dabei den Angaben gefolgt, die auf dem Bildschirm erscheinen. Bevor Sie die Software zum ersten Mal starten, sollten Sie ein eigenes Verzeichnis anlegen, in dem Sie Ihre Daten-Dateien und die R-Kommando-Dateien halten.

Die R-Konsole und Skripte

Nach dem Starten befindet man sich in der R-Konsole. Ist der Cursor nach dem Promptzeichen zu sehen, so ist R bereit, Befehle entgegenzunehmen und zu verarbeiten. Man kann Befehle direkt auf der Konsole eingeben und mittels Betätigen der Enter-Taste ausführen lassen. Ein Beispiel ist etwa der Befehl `1+1`. Nach dem Abschicken erhält man das Ergebnis auf der Konsole angezeigt. Auf dem Bildschirm sieht die vollständige Einheit wie folgt aus:

¹ Das Kapitel basiert auf dem Kapitel „Einführung in R“ im Lehrbuch „Statistische Auswertungen mit R“, Schlittgen (2004).

```
> 1+1
[1] 2
```

Wird versehentlich ein Befehl abgeschickt, bevor er vollständig ist, so fragt R mittels eines +-Zeichens nach der Vervollständigung:

```
> (1+1)*
+ 3
[1] 6
```

Mehrere Befehle dürfen auf einer Zeile platziert werden; sie sind dann durch Semikolons zu trennen:

```
> (1+1); 2^3
[1] 2
[1] 8
```

Die Konsole bietet einige Steuerungsmöglichkeiten. Vom Arbeiten auf der Konsole wird aber dringend abgeraten! Die Navigationsmöglichkeiten sind zu eingeschränkt, als dass dabei Freude aufkommen könnte. Besser ist es, ein Skript anzulegen und in diesem die Kommandos einzugeben. Später können Sie dann ein bereits vorhandenes Skript zu weiterer Bearbeitung Ihres Projektes öffnen. Skripte erreicht man über die Menüleiste unter dem Menüpunkt „Datei“. Haben Sie ein Skript geöffnet, so wird die Befehlszeile, in der sich der Cursor befindet durch die Tastenkombination Ctrl+R in die Konsole übertragen und dort ausgeführt. Ist im Skript ein Text markiert, so wird mit dem gesamten markierten Text in dieser Weise verfahren. Damit dürfen auch mehrere Operationen angefordert werden. Diese müssen durch ein Semikolon oder durch einen Zeilenumbruch voneinander getrennt sein. Das Zeichen # beendet die weitere Abarbeitung der Zeile; es wird zur nächsten Zeile übergegangen. Alles, was nach # auf der Zeile steht, wird als Kommentar angesehen.

Über die rechte Maustaste bzw. über den Menüpunkt Datei lässt sich das Skript abspeichern und kann bei einer nachfolgenden Sitzung wieder geladen werden. Im Folgenden wird davon ausgegangen, dass Sie Befehle über ein Skript absetzen.

R kann als Taschenrechner verwendet werden; es lassen sich die gängigen Rechenoperationen durchführen. Die anzuwendenden Befehlskombinationen bedürfen in vielen Fällen keiner Erklärung.

Beispiel 13.1

- 1+2 Addiert 1 und 2,
- 2-1 Subtrahiert 1 von 2,
- 2*3 Multipliziert 2 und 3,
- 2/3 Dividiert 2 durch 3,
- 2^3 Berechnet 2 „hoch“ 3.

Weiterhin gibt es natürlich Funktionen, welche entsprechend der üblichen mathematischen Schreibweise aufgerufen werden:

Beispiel 13.2

- `sqrt(2)` Quadratwurzel von 2,
- `exp(2)` Exponentialfunktion an der Stelle 2,
- `log(2)` Natürlicher Logarithmus von 2.

Will man das errechnete Ergebnis eines Befehls zur späteren Weiterverarbeitung im Direktzugriff zur Verfügung haben, so kann die auszuführende Operation auf ein zu erstellendes Objekt zuwiesen werden. In einfacheren Fällen sind es Variablen; eine Variable in R ist ein über den Variablenamen ansprechbarer Wert oder eine Anzahl von Werten.

Variablen werden in R durch ihre erste Wertzuweisung initialisiert. Die Zuweisung eines Wertes an eine Variable geschieht durch den Operator `<-`. (Das Kleiner-Zeichen und das Minus-Zeichen hintereinander gesetzt.) Dabei kann die Variable auch einen noch zu berechnenden Ausdruck zugewiesen bekommen. Nur im einfachsten Fall handelt es sich bei dem Ausdruck in der Zuweisung Variable `<- Ausdruck` um eine Konstante.

Beispiel 13.3

- `A <- 4.5` A hat den Wert 4.5
- `B <- 1+1` B hat den Wert 2.

Was unter einem Variablenamen zwischengespeichert ist, kann während der Laufzeit einer R-Sitzung ändern. Variablennamen sind „case sensitive“, d. h. R unterscheidet zwischen großen und kleinen Buchstaben. Die Variablennamen können aus beliebig vielen Buchstaben und dem Punkt bestehen. Reservierte Namen wie NA, NULL, TRUE, T, FALSE und F dürfen nicht als Variablennamen verwendet werden.

Hilfen

Eine globale Möglichkeit Hilfefunktion aufzurufen, bieten die menügesteuerten Hilfen. Sie können über den Menüpunkt Help (obere Bildschirmleiste) angefordert werden. In den ersten Punkten gibt es einige Informationen zum Arbeiten mit der R-Konsole und zu R. Die eigentliche Hilfe befindet sich unter „R Funktionen (Text)...“ bzw. „HTML Hilfe“. In der erstgenannten Standardhilfe wird nach Aufruf in einem Dialogfenster nach dem Befehl gefragt, über den Informationen gewünscht sind. Bei der Hilfe im HTML-Format wird ein Browser geöffnet und eine lokale Datei angezeigt; man muss sich bei Verwendung dieser Hilfe also nicht im Internet anmelden. Hier erhält man durch Anklicken von „Packages“ eine Liste aller installierten Pakete. Auswählen

etwa von „base“ zeigt dann eine alphabetisch geordnete Beschreibung aller verfügbaren Funktionen des Basispaketes. Relevant sind vor allem die Pakete „base“, „graphics“, „stats“, „tseries“ und „utils“. Ein Vorteil der HTML-Hilfe ist, dass man bei Hinweisen auf andere, in diesem Zusammenhang wichtige Funktionen direkt durch Anklicken zugreifen kann, während die Standardhilfe für den „Folgebefehl“ wieder neu aufgerufen werden muss. Die jeweilige Hilfeseite kann durch Klicken auf das Kreuz rechts oben in der Ecke geschlossen werden. Die alphabetische Liste der Funktionen steht auch als PDF-Datei unter „Handbücher (PDF)“, und dort als „R Reference Manual“ zur Verfügung, wenn bei der Installation die entsprechende Option ausgewählt wurde. Um sie zu nutzen, muss der Acrobat Reader installiert sein. Er ist kostenfrei aus dem Internet herunterzuladen.

Es gibt auch eine direkte Hilfefunktion in R. Diese kann mit `help(befehl)` aufgerufen werden. Hierbei muss der in Klammern gesetzte Ausdruck `befehl` durch den Befehl ersetzt werden, über den eine Auskunft gewünscht wird. Alternativ zu `help(befehl)` kann auch das Kommando `?befehl` verwendet werden.

R Studio

Die Software RStudio ist eine graphische Benutzeroberfläche für R, welche das Arbeiten mit R erleichtert. Eine freie Softwareversion steht unter der AGPL-Lizenz zur Verfügung. Die Adresse lautet rstudio.com/products/rstudio/. Durch das Starten des Programms wird automatisch auch R geöffnet.

Das Arbeitsumfeld von RStudio ist in vier Fenstern unterteilt. Das Fenster *Console* ist die R-Konsole. Hier wird R ausgeführt. Im Fenster *Source* werden Skripte bearbeitet und verwaltet. Eine Befehlszeile, in der sich der Cursor gerade befindet, kann durch Drücken auf die Schaltfläche „Run“ oder durch die Tastenkombination Strg+Enter in die Konsole übertragen und dort ausgeführt werden. Die Schaltfläche „Source“ lässt das gesamte Skript ausführen. Die erzeugten Variablen werden im Fenster *Environment* angezeigt. Schließlich erlaubt das Fenster *Files* die Navigation in den Dateien und Verzeichnissen. Auch werden hier die mit R erstellten Abbildungen angezeigt.

13.2 Datentypen und Objekte

Datentypen

Die einfachen Datentypen in R sind

- Zahlen,
- Wahrheitswerte,
- Zeichenketten.

Zahlkonstanten bestehen aus einem optionalen Vorzeichen und beliebig vielen Ziffern, die durch einen Dezimalpunkt in Vor- und Nachkommastellen unterteilt werden. Der angelsächsischen Konvention gemäß wird ausschließlich der Dezimalpunkt anstelle des im deutschen Sprachraum gebräuchlichen Kommas verwendet! Zusätzlich wird von R auch die Exponentialschreibweise unterstützt. Beispiele für Zahlkonstanten sind 1, 13.5674, .04, -45, 34e-12, pi. Die Konstante pi ist dabei die bekannte Kreiszahl 3.1415926535897931 (usw.).

Zusätzlich kennt R die Werte +Inf und -Inf, also $\pm\infty$. Dies ergibt sich z. B. immer, wenn eine von null verschiedene Zahl durch null dividiert wird.

Zeichenketten sind beliebige Folgen von Ziffern, Buchstaben und Sonderzeichen. Konstante Zeichenketten erkennt man an den umschließenden doppelten Ausführungszeichen. Zeichenketten bilden neben den Zahlen die andere Form von Konstanten. Beispiele sind "Hallo" und "Dies ist ein Test."

Wahrheitswerte sind die Zustände wahr und falsch. Sie werden in R durch die Booleschen Konstanten TRUE und FALSE (kurz T bzw. F) repräsentiert. Auch numerische Konstanten tragen Wahrheitswerte, und zwar den Wert TRUE, wenn die Konstante von null verschieden ist und FALSE, wenn sie gleich null ist.

NA, „Not Available“, ist eine keinem dieser Datentypen zugehörige Konstante, die R kennt. NA und wird immer dann verwendet, wenn in einem Datensatz ein Wert nicht verfügbar ist. Wenn eine mathematische Operation oder Funktion keinen sinnvollen Wert berechnen kann, wird NaN, „Not a Number“ (keine Zahl), ausgegeben.

R-Vektoren

Die einfachste Art, mehrere Daten zusammenzustellen um damit insgesamt operieren zu können, ist die Bildung von *R-Vektoren*. Ein R-Vektor ist eine Folge von gleichartigen Objekten, im häufigsten Fall von Zahlkonstanten. Vektoren können mit einem Buchstaben bzw. Namen angesprochen werden. Wesentlich für die Generierung von Vektoren ist die Funktion c, mit der eine Werteliste zu einem Vektor zusammengefügt werden kann.

Beispiel 13.4

A <- c(1, 2, 3, 4) A hat den Datentyp numerischer Vektor und den Inhalt 1,2,3,4.
Gibt man die Zuweisung in der R-Konsole ein und anschließend A, so hat die Ausgabe die Gestalt

```
[1] 1 2 4 6
```

Auch wenn die Ausgabe als eine Zeile erscheint, ist A ein R-Vektor, der sich in vielen Umständen wie ein Spaltenvektor verhält.

Mit dem binären Sequenzoperator, der durch einen Doppelpunkt angegeben wird, kann ein Vektor von aufeinanderfolgenden Zahlen erzeugt werden. Die Zahlen haben den Abstand 1. Der erste Operand gibt den Startwert, der zweite die Obergrenze der Sequenz an. Falls der zweite Operand kleiner ist als der erste, wird eine absteigende Sequenz erzeugt. Wenn die Obergrenze sich nicht um eine ganze Zahl von der Untergrenze unterscheidet, ist die letzte Zahl der Sequenz kleiner als die Obergrenze.

Beispiel 13.5

`c(1:6)`, `seq(6)` oder einfach `1:6` erzeugt den Vektor `(1, 2, 3, 4, 5, 6)`

`6:-1` erzeugt den Vektor `(6, 5, 4, 3, 2, 1, 0, -1)`

`1.2:3` erzeugt den Vektor `(1.2, 2.2)`

Ein Vektor kann wie eine Zahl oder Zeichenkette einer Variablen zugewiesen werden.

Beispiel 13.6

`A <- c(1:6); A` hat den Datentyp Vektor und den Wert `(1, 2, 3, 4, 5, 6)`.

Weitere Objekte

R ist eine objektorientierte Programmierumgebung. Das bedeutet, dass Methoden bei der Anwendung auf unterschiedliche Objekttypen unterschiedlich reagieren. Dazu verfügen die Objekte über entsprechende Attribute.

Eine *Matrix* unterscheidet sich von einem R-Vektor darin, dass sie zwei Dimensionen hat. Um das zu realisieren, wird ein Vektor mit dem Attribut `dim` versehen. `dim` ist dabei ein Vektor der Länge zwei, der die Anzahl der Zeilen und Spalten enthält. Als optionales Attribut kann einer Matrix auch das Attribut `dimnames` angeheftet sein. Standardmäßig wird eine Matrix aus den Vektorelementen gebildet, indem diese spaltenweise aufgefüllt wird. Aus R-Vektoren können mit den folgenden Funktionen Matrizen erzeugt werden:

<code>matrix(x, z, s)</code>	Ordnet einem Vektor die Dimension <code>(z,s)</code> zu; m.a.Worten: <code>matrix</code> transformiert einen Vektor <code>x</code> in eine Matrix vom Typ <code>(z,s)</code> .
<code>dim(x)<-c(z, s)</code>	Transformiert einen Vektor <code>x</code> in eine Matrix vom Typ <code>(z,s)</code> .
<code>diag(x)</code>	Erzeugt eine Diagonalmatrix, bei der <code>x</code> auf der Diagonale steht.
<code>cbind(x, y)</code>	Setzt Vektoren (und Matrizen) nebeneinander zu einer Matrix zusammen.
<code>rbind(x, y)</code>	Setzt Vektoren (und Matrizen) untereinander zu einer Matrix zusammen.
<code>structure(x, y)</code>	Gibt das eingegebene Objekt <code>x</code> mit den in <code>y</code> gesetzten Attributen zurück.

Mit `is.matrix` stellt man fest, ob ein Objekt eine Matrix ist. Nicht nur die Konvertierung eines R-Vektors in eine Matrix, sondern auch die Änderung der Dimension einer Matrix und die Abfrage der Dimension einer Matrix ist mit `dim` möglich.

Beispiel 13.7

```

1. A <- c(1,2,3,4,5,6); B <- matrix(A,2,3); B
   [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6

2. C <- structure(1:6, dim = 2:3); B==C
   [,1] [,2] [,3]
[1,] TRUE TRUE TRUE
[2,] TRUE TRUE TRUE

3. x <- 1:3; y <- 4:6; z <- cbind(x,y) ; z
   x  y
[1,] 1  4
[2,] 2  5
[3,] 3  6
is.matrix(z)
[1] TRUE

4. x <- 1:12 ; dim(x) <- c(3,4); x
   [,1] [,2] [,3] [,4]
[1,]    1    4    7   10
[2,]    2    5    8   11
[3,]    3    6    9   12
dim(x)
[1] 3    4

```

Datensätze bilden den Ausgangspunkt der meisten statistischen Aktivitäten. Sie bestehen in der Statistik aus einem rechteckigen Schema von Werten, so dass eine Zeile jeweils eine Beobachtungseinheit repräsentiert. In den Spalten sind jeweils die zu einer Variablen gehörigen Beobachtungen angeordnet. Die Variablen können fantasievolle Namen tragen wie etwa Ozon, Geschlecht oder Zigarette; sie können aber auch einfach mit V01, V02 usw. bezeichnet sein. Bei Datensätzen sind die Spalten mit den Variablenbezeichnungen verknüpft. Die Funktionen in R arbeiten mit Datensätzen weitgehend so wie mit Datenmatrizen. Im Hintergrund sind Datensätze aber gänzlich anders organisiert, nämlich als *Listen*. Das sind übergeordnete Objekte, die unterschiedliche Elemente in einem Objekt verbinden. Auf die einzelnen Objekte kann mittels der Angabe eines enthaltenen Objektes, das durch ein \$-Zeichen mit dem Na-

men des Datensatzes verbunden ist, zugegriffen werden. Bei einem Datensatz etwa mittels `dat$var`. Alternativ dazu sind die Variablen des Datensatzes mittels `attach` als einfache Datenvektoren zur Verfügung zu stellen.

Beispiel 13.8

Im November 2001 veröffentlichte DMEuro Ergebnisse aus einer europäischen Studie, u. a. mit den Angaben Land, dem Hauptsitz der Firma und Inter, der Internationalität (gemessen als außereuropäischer Umsatzanteil in Prozent). Dieser Datensatz sei unter dem Namen `firmen` in der Konsole vorhanden. Beide folgende Befehle geben den Inhalt der Variablen Land aus:

```
firmen$Land
attach(firmen); Land
```

Reguläre *Zeitreihen* haben das Attribut `tsp`. Regulär sind sie, wenn die zeitlichen Abstände zweier aufeinander folgender Beobachtungen jeweils gleich sind. Sie sind Vektoren (eindimensionale Zeitreihen) oder Matrizen (mehrdimensionale Zeitreihen) mit zusätzlichen Parametern für die Zeitkomponente: `start`, `end`, und `frequency` (= Anzahl von Beobachtungen pro Zeitintervall). Diese Konstruktion wird hauptsächlich verwendet, um Reihen mit periodischen Unterstrukturen wie Monats- oder Quartalsdaten zu bearbeiten. Mit `ts(y)` wird `y` zur Zeitreihe, mit `tsp(y)` werden die Attribute eines Zeitreihenobjekts abgefragt.

Beispiel 13.9

Mit den folgenden Anweisungen wird eine Zeitreihe von Monatswerten erzeugt und ihr Attribut abgefragt:

```
y <- ts(rnorm(50), start=c(1974,1), frequency=12)
tsp(y)
[1] 1974.000 1978.083    12.000
```

Indizierung

Man kann auf Teile oder einzelne Elemente eines Vektors, einer Matrix, einer Zeitreihe oder eines Datensatzes zugreifen, um sie auszuwählen bzw. zu entfernen. Dies geschieht durch die *Indizierung*, die Angabe der Indizes innerhalb eckiger Klammern. Da in der mathematischen und statistischen Literatur die Indizes vornehmlich als Subskripte notiert werden, spricht man hier auch von *Subscription*. Zeitreihen entsprechen hierbei den Vektoren und Datensätzen den Matrizen.

Bei Vektoren erfolgt die Indizierung gemäß `vektor[a]`. Bei `a` kann es sich um einen Vektor mit positiven oder negativen Komponenten handeln, oder um einen Vektor von booleschen Werten.

Ist a eine positive Zahl, so wird bei einem Vektor das an der entsprechenden Stelle stehende Element ausgewählt. Das erste Element eines Vektors hat dabei den Index 1. Bei nicht ganzzahligen Indizes wird abgerundet. Ist a ein Vektor von positiven Werten, so werden diese als Liste von Indizes interpretiert und die entsprechenden Elemente werden ausgewählt.

Da das Komma für die Trennung der Dimensionen bei Matrizen dient, darf es nicht zur Trennung von Indizes eines Vektors verwendet werden. Ist der angegebene Index größer als die Länge des Vektors, so ist das Ergebnis der Operation NA.

Beispiel 13.10

`A <- 3:8`

`A[2]` liefert den Wert 4.

`A[9]` liefert NA.

`A[2:4]` liefert den Vektor (4, 5, 6).

`A[1,3]` liefert die Fehlermeldung „incorrect number of dimensions“.

`A[c(2:4,9)]` liefert den Vektor (2 3 4 NA).

Wird ein negativer Index angegeben, so ist das Resultat ein Vektor, aus dem das an dieser Stelle befindliche Element entfernt wurde. Ist der Betrag des angegebenen negativen Index größer als die Länge des Vektors, so gibt R eine Fehlermeldung aus. Es gilt die Einschränkung, dass alle Indizes entweder positiv oder negativ sein müssen.

Beispiel 13.11

`A <- 3:8`

`A[-2]` liefert den Vektor (3, 5, 6, 7, 8).

`A[-9]` produziert die Fehlermeldung "Error: subscript out of bounds".

`A[-2:-4]` liefert den Vektor (3, 7, 8).

`A[c(-1,-3)]` liefert den Vektor (4, 6, 7, 8).

`A[c(-1,3)]` führt zu der Meldung:

‘Error: only 0’s may mix with negative subscripts’.

Anstelle von numerischen Indizes können auch Vektoren von booleschen Werten verwendet werden. Hierbei werden aus dem indizierten Vektor nur die Elemente ausgewählt, deren Pendant im Indexvektor den Wert TRUE hat. Ist der Indexvektor kürzer als der indizierte Vektor, so findet wie schon bei den arithmetischen Operatoren ein „wrap around“ statt.

Beispiel 13.12

`A <- 1:6; A[c(T,F,F,T,F,T)]` liefert den Vektor (1, 4, 6);

`A <- 1:6; A[c(T,F)]` liefert den Vektor (1, 3, 5).

Anstelle eines konstanten Vektors von booleschen Werten kann natürlich als Index auch ein beliebiger Ausdruck verwendet werden, der in einem Vektor von booleschen Werten resultiert.

Beispiel 13.13

`A[A<4]` liefert genau wie

`A<-1:6; B<-A<4; A[B]` den Vektor (1, 2, 3)

Bei Matrizen gibt es eine Zeilen- und eine Spaltendimension. Es sind also zwei Angaben zu machen; diese werden durch ein Komma getrennt. Dann wird auf die jeweiligen Zeilen bzw. Spalten zugegriffen. Das Komma muss angeführt werden. Werden nur Indizes vor dem Komma spezifiziert, so betrifft die Auswahl die ganze(n) vor dem Komma spezifizierten Zeile(n). Entsprechendes gilt für die Angabe von Indizes nach dem Komma bzgl. der Auswahl der ganze(n) nach dem Komma spezifizierten Spalte(n).

`Matrix[a], Matrix[,b], Matrix[a,b]`

Bei a und b kann es sich um einen Vektor mit positiven oder negativen Komponenten handeln, oder um einen Vektor von booleschen Werten. Auch hier dürfen innerhalb eines Vektors von Indizes positive und negative Werte nicht gemischt werden. Ansonsten gelten die für Vektoren angegebenen Regeln.

13.3 Operatoren und Funktionen

Mathematische Operatoren

R unterstützt die folgenden binären mathematischen Operationen:

`a + b` Addition,

`a - b` Subtraktion,

`a * b` Multiplikation,

`a / b` Division,

`a ^ b` Potenzierung,

`a %% b` Divisionsrest,

`a %/% b` Ganzzahlige Division (Division ohne Rest).

Über die genannten Operatoren können sowohl zwei Konstanten oder Variablen vom Datentyp Zahl, wie auch eine Zahl und ein Vektor sowie zwei Vektoren miteinander verknüpft werden. Sind beide Operanden Zahlen, so ist das Ergebnis wiederum eine Zahl, ist zumindest einer der beiden Operanden ein Vektor, so ist auch das Ergebnis ein Vektor.

Bei der Verknüpfung eines Vektors mit einer Zahl wird jedes Element des Vektors mit der Zahl verknüpft. Ist die Verknüpfung nicht möglich, so wird in den Resultatevektor an dieser Stelle NaN eingefügt; das Ergebnis ist keine Zahl. Bei einer Division durch Null wird der Wert +Inf ausgegeben.

Beispiel 13.14

```
1:4 + 2    liefert den Vektor (3, 4, 5, 6),
(2:-1)^0.5 liefert (1.414214, 1.000000, 0.000000, NaN),
12 / 4:0    liefert den Vektor (3, 4, 6, 12, Inf).
```

Bei Durchführung einer mathematischen Operation auf zwei Vektoren L1 und L2 besteht der Ergebnisvektor aus (L11 op L21, L12 op L22, ... L1n op L2n). Der Ergebnisvektor ist so lang wie der längere der beiden Eingangsvektoren. Ist einer der Operanden kürzer als der andere, so wird er mehrfach hintereinander gesetzt, es findet ein „wrap around“ statt.

Beispiel 13.15

```
1:4 + 1:4    liefert den Vektor (2, 4, 6, 8),
1:4 * 1:4    liefert den Vektor (1, 4, 9, 16),
1:4 + 1:8    liefert den Vektor (2, 4, 6, 8, 6, 8, 10, 12),
1:4 + 1:5    liefert den Vektor (2, 4, 6, 8, 6).
```

Im letzten Fall wird eine Warnung ausgegeben, dass die Längen der Vektoren sich nicht um ein ganzzahliges Vielfaches unterscheiden:

```
Warning message:
longer object length
is not a multiple of shorter object length in: 1:4 + 1:5
```

Vergleichsoperatoren

Neben den oben aufgeführten mathematischen Operationen kennt R eine Reihe von Vergleichsoperatoren, die jeweils einen booleschen Wert, TRUE oder FALSE, kurz T oder F, bzw. einen Vektor von booleschen Werten zurückliefern. Für skalare Werte (Zahlen, boolesche Werte und Zeichenketten) sind diese Operatoren wie folgt definiert:

- a < b T, wenn a kleiner als b ist.
- a <= b T, wenn a kleiner oder gleich b ist.
- a = = b T, wenn a gleich b ist.
- a != b T, wenn a ungleich b ist
- a >= b T, wenn a größer oder gleich b ist.
- a > b T, wenn a größer als b ist.

Bei der Durchführung einer Vergleichsoperation auf einen Vektor und einen skalaren Wert, wird jedes Element des Vektors mit diesem Wert verglichen und ein entsprechender Resultatevektor zurückgeliefert.

Beispiel 13.16

`1:8 < 4` liefert den Vektor (T, T, T, F, F, F, F, F).

Der Vergleich zweier Vektoren führt wie bei den arithmetischen Operationen zu einem paarweisen Vergleich der Werte beider Vektoren.

Boolesche Operatoren

<code>!a</code>	Negation
<code>a & b</code>	Und
<code>a b</code>	Oder

Der *Negationsoperator* `!` negiert einen booleschen Wert, bzw. alle Werte eines Vektors von booleschen Werten.

Beispiel 13.17

`!(1:8 < 4)` liefert den Vektor (F, F, F, T, T, T, T, T).

Die booleschen Operatoren `&` (und) und `|` (oder) führen eine boolesche Verknüpfung auf zwei Skalaren, bzw. einem Skalar und einem Vektor, bzw. zwei Vektoren durch. Hierbei werden die nicht-booleschen Datentypen wie folgt interpretiert:

0 hat den Wert FALSE, alle anderen Zahlen haben den Wert TRUE. Die leere Zeichenkette `" "` hat den Wert FALSE, alle anderen Zeichenketten haben den Wert TRUE. NA hat immer den Wert FALSE. Die Verknüpfung von Vektor und Skalar, bzw. von zwei Vektoren geschieht wie bei den arithmetischen Operatoren durch paarweise Verknüpfung.

Beispiel 13.18

```
A <- 1:8
B <- (A < 6) & (A > 2)
B enthält den Vektor (F, F, T, T, T, F, F, F)
```

Matrix-Operationen

Matrizen können wie üblich mittels `*` mit Skalaren multipliziert werden. Das eigentliche Matrizenprodukt ist `%%`. Die Transponierung einer Matrix A wird durch `t(A)` geleistet. Die Inverse einer regulären Matrix erhält man mit `solve`.

Beispiel 13.19

```
A <- matrix(c(1:6), 2, 3)
B <- t(A)
B
 [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    5    6
B %*% A
 [,1] [,2] [,3]
[1,]    5   11   17
[2,]   11   25   39
[3,]   17   39   61
solve(diag(c(1,1,1))+ (B %*% A))
 [,1]          [,2]          [,3]
[1,]  0.7844828 -0.1637931 -0.1120690
[2,] -0.1637931  0.7155172 -0.4051724
[3,] -0.1120690 -0.4051724  0.3017241
```

Um eine Funktion, die eigentlich für Vektoren gedacht ist, zeilen- oder spaltenweise auszuführen, steht die Funktion `apply` zur Verfügung.

Beispiel 13.20

`A <- matrix(c(1:6), 2, 3).apply(A, 1, sum)` gibt die Summe über die Zeilen:

```
[1] 9 12
```

`apply(A, 2, sum)` gibt die Summe über die Spalten:

```
[1] 3 7 11
```

Dass Matrizenoperationen zur Verfügung stehen, führt dazu, dass man möglichst viele der bei einer Berechnung auszuführenden Operationen als Matrix-Operationen schreibt.

Beispiel 13.21

Die folgenden Befehle erzeugen eine Matrix `x`, bestimmen die Randsummen und den Wert der χ^2 -Teststatistik.

```
x <- 1:24 ; dim(x) <- c(4,6)
x
 [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    5    9   13   17   21
[2,]    2    6   10   14   18   22
[3,]    3    7   11   15   19   23
[4,]    4    8   12   16   20   24
```

```
r <- apply(x, 1, sum)
u <- apply(x, 2, sum)
e <- r %*% t(u) / sum(x)
xe <- (x - e)^2
xe <- scale(xe, center = F, scale = u)
xe <- scale(t(xe), center = F, scale = r)
sum(xe) * sum(x)
```

Speziell ist die Verknüpfung von R-Vektoren mit Matrizen. R-Vektoren sind ja keine Zeilen- oder Spaltenvektoren im herkömmlichen Sinne. Solche sind einfach Matrizen der Dimension (1,n) bzw. (n,1). Wird ein R-Vektor x mittels $*$ mit einer Matrix a multipliziert, so geschieht das elementweise, und zwar so, dass die erste Zeile der Matrix mit dem ersten Element des R-Vektors durchmultipliziert wird, die zweite Zeile mit dem zweiten Element usw.

Beispiel 13.22

`A<-matrix(c(1:6), 2, 3).` Dann erzeugt `c(4, 5)*A` die Ausgabe

```
[,1] [,2] [,3]
[1,]    4   12   20
[2,]   10   20   30
```

Will man diese Operation mit einer einspaltigen oder einzeiligen Matrix durchführen, so ist sie zuerst in einen R-Vektor zu verwandeln. Das geschieht mit `as.vector`.

Beispiel 13.23

```
A<-matrix(c(1:6), 2, 3); x <- as.vector(A[,1]); print(x*A)
[,1] [,2] [,3]
[1,]    1    3    5
[2,]    4    8   12
```

Funktionen

Wie bei jeder anderen Programmiersprache wird auch der R-Sprachumfang durch eine Reihe von Funktionen ergänzt. Einige wurden schon in der bisherigen Beschreibung angegeben. Allgemein geschieht der Aufruf einer Funktion in R wie folgt:

`funktionsname(argument1, argument2, ...)`

Funktionen können dabei auch keine oder optionale Argumente besitzen. Wenn die optionale Argumente nicht mit angeführt sind, werden die Voreinstellungen verwendet.

Beispiel 13.24

`round(2.331)` und `round(2.331, 2)` sind legitime Aufrufe der Funktion `round`. 2 ist dabei ein optionales Argument, das die Anzahl der Stellen angibt, bei denen gerundet werden soll. Fehlt es, so wird per Voreinstellung beim Dezimalpunkt gerundet.

Um einen Eindruck von der großen Anzahl von Funktionen zu geben, sind in der Tabelle 13.1 einige wichtige mathematische Funktionen zusammengestellt.

Tab. 13.1: Einige mathematische Funktionen

Funktionsname	Funktionsbeschreibung
<code>abs(x)</code>	Werte der Elemente von x.
<code>ceiling(x)</code>	Aufrunden der Werte in x (auf ganze Zahlen).
<code>chol(x)</code>	Bestimmen des Cholesky-Faktors der Matrix x.
<code>choose(n, x)</code>	Binomialkoeffizient n über x.
<code>cos(x)</code>	Kosinus von x (x in Bogenmaß).
<code>diag(x)</code>	Extrahieren der Hauptdiagonalen einer Matrix oder Erzeugen einer Diagonalmatrix.
<code>exp(x)</code>	Exponentialfunktion für die in x enthaltenen Werte.
<code>floor(x)</code>	Abrunden der Werte in x (auf ganze Zahlen).
<code>log(x)</code>	Logarithmus zur Basis e (=2.718282.....).
<code>log10(x)</code>	Logarithmus zur Basis 10.
<code>round(x)</code>	Rundet die in x enthaltenen Werte. Über das optionale Argument <code>digits = d</code> kann die Nachkommastelle bestimmt werden, bei der gerundet werden soll.
<code>sign(x)</code>	Bestimmt die Vorzeichen der Elemente von x.
<code>sin(x)</code>	Sinus von x (x in Bogenmaß).
<code>sqrt(x)</code>	Quadratwurzel der Elemente von x.
<code>tan(x)</code>	Tangens von x (x in Bogenmaß).
<code>trunc(x)</code>	Abrunden der Werte von x (auf ganze Zahlen).

Wesentlich zu Verbreitung von R hat der große Umfang an Funktionen beigetragen, die zur Bestimmung statistischer Maßzahlen und zur Durchführung statistischer Analysen zur Verfügung stehen. Tatsächlich sind die in den Beispielen der voranstehenden Kapitel aufgeführten R-Codes voll von verschiedenen für die Zeitreihenanalyse relevanten Funktionen dieser Art. Auch hier wird mittels einer Auflistung in der Tabelle 13.2 ein Eindruck von der großen Anzahl der vorhandenen Funktionen gegeben.

Die Funktion `lm()` zur Anpassung von Regressionsmodellen ist grundlegend; eine von allen Optionen befreite Version des Aufrufs der Regression von y gegen x hat die einfache Struktur:

```
fitted.model <- lm(y ~ x)
```

Tab. 13.2: Einige statistische Funktionen

Funktionsname	Funktionsbeschreibung
attach(dat)	Zugriff zu Variablen des Datensatzes dat.
c(x,y,z,...)	Erzeugen eines Vektors aus den Objekten x, y z, ...
cbind(x,y)	Zusammenfügen zweier Vektoren zu einem Datensatz.
cor(x,y)	Korrelationskoeffizient der Vektoren x und y. Wenn x eine Matrix ist und y weggelassen wird, werden die Korrelationen der Spalten von x berechnet. Wird zusätzlich y angegeben, so werden die Korrelationen der Spalten von x mit denen von y bestimmt.
cumsum(x)	Kumulierte Summe aller in dem Vektor x enthaltenen Elemente.
data.frame(x)	Umwandeln des Objektes x in einen Datensatz.
is.na(x)	Überprüfen, ob ein Objekt einen fehlenden Wert enthält.
length(x)	Länge des Vektors x bzw. Anzahl der Elemente der Matrix x.
lm(y~ x)	Führt eine Regression mit der abhängigen Variablen y und der unabhängigen Variablen x durch. Dabei wird ein konstantes Glied berücksichtigt.
mad(x)	Median der absoluten Abweichungen der Elemente von x vom Median von x.
matrix(x,z,s)	Erzeugen einer Matrix aus den Elementen in x mit z Zeilen und s Spalten.
max(x)	Maximum der Elemente von x.
mean(x)	Arithmetisches Mittel der Elemente von x.
median(x)	Median der Elemente von x.
min(x)	Minimum der Elemente von x.
qnorm(p)	Quantil der Normalverteilung
qqnorm(x)	QQ-Diagramm für den Vektor x.
quantile(x,p)	Bestimmt die empirische Quantile von x zu den in p angegebenen Anteilen. Dabei wird linear interpoliert.
range(x)	Gibt den Vektor zurück, der aus dem kleinsten und dem größten der in x enthaltenen Werte besteht.
rank(x)	Gibt die Rangwerte eines numerischen Vektors x wieder. Bei Bindungen oder Ties werden mittlere Ränge berechnet.
read.csv("dat")	Einlesen eines Datensatzes aus einer csv-Datei.
read.table("dat")	Einlesen eines Datensatzes aus einer ASCII-Datei.
rep(x,n)	Erzeugen eines Vektors durch n-fache Wiederholung von x.
rev(x)	Vektor x in umgekehrter Reihenfolge.
rnorm(n)	Vektor mit n normalverteilten Zufallszahlen.
runif(n)	Vektor mit n gleichverteilten Zufallszahlen zwischen 0 und 1.
sample(x,n)	Zufällige Auswahl von n Elementen aus dem Objekt x.
scale(x,...)	Zentriert und/oder skaliert die numerischen Matrix x spaltenweise. Die Auswahl nur einer der Operationen geschieht, indem bei center=TRUE, scale=TRUE einer der beiden TRUE auf FALSE gesetzt wird. Es können auch Vektoren angegeben werden, die zur Zentrierung bzw. Skalierung verwendet werden.

Tab. 13.2: (Fortsetzung)

Funktionsname	Funktionsbeschreibung
scan("dat")	Einlesen eines Datenvektors aus einer ASCII-Datei.
sd(x)	Standardabweichung der in x enthaltenen Daten.
seq(n)	Erzeugen eines Vektors mit Einträgen 1, 2, ..., n.
sum(x)	Summe aller in dem Vektor oder der Matrix x enthaltenen Elemente.
summary(x)	Je nach Art der Eingabe gibt diese Funktion eine Übersicht über die wesentlichen statistischen Charakteristika von x aus: univariater Datensatz, Vektor: 5-Zahlen-Zusammenfassung, ohne Umfang der Daten aber mit arithmetischem Mittel. Ergebnis der Funktion <code>lm</code> : Übersicht über das Regressionsergebnis
var(x,...)	Varianz oder Kovarianz der Variablen x bzw. der Variablen x und y.

Bei mehr als einem Regressor werden diese mit einem + verbunden. Normalerweise wird ein Achsenabschnitt berücksichtigt. Dies kann aber durch Einfügen von -1 auf der rechten Seite ausgeschaltet werden.

Mit `summary(fitted.model)` werden die Informationen über das angepasste Modell extrahiert.

Da Zeitreihenobjekte selbst Vektoren oder Matrizen mit einem zusätzlichen Attribut sind, können sie mit Funktionen für Vektoren bzw. Matrizen bearbeitet werden. Die Komponenten einer multivariaten Zeitreihe können eigenständige Namen tragen. Jedoch muss die Indizierung verwendet werden, um auf einzelne Komponenten zuzugreifen.

Beispiel 13.25

```
> Z <- ts(matrix(rnorm(100), 50, 2), start=1, frequency=1)
> colnames(Z) <- c("X", "Y")
> print(Z[1,])
      X Y
1.43886459 0.08714056
> print(X)
Fehler in print(X) : Objekt 'X' nicht gefunden
```

Tab. 13.3: Einige Funktionen zur Bearbeitung von Zeitreihen

Funktionsname	Funktionsbeschreibung
acf(y)	Die empirische ACF. Abbildung enthält Schwankungsintervall $-1/N \pm 1.96 \cdot 1/\sqrt{N}$. Mit der Option ci.type = "ma" werden die Bartlett-Grenzen angezeigt.
acfmat(y)	Die empirischen Autokorrelationsmatrizen einer multivariaten Zeitreihe y. (Funktion im R-Paket tsapp enthalten.)
acfpacf(y)	Die empirischen ACF und PACF; Abbildung enthält die für die Identifikation eines ARMA-Modells notwendigen Schwankungsintervalle. (Funktion im R-Paket tsapp enthalten.)
ar(y)	Schätzung eines AR[p]-Modells.
arima(y)	Schätzung eines ARIMA[p, d, q]-Modells. Speziell können auch SARIMA-Modelle sowie der Einfluss exogener Regressoren geschätzt werden.
as.vector(y)	Zeitreihe wird in einen Vektoren umgewandelt.
ccf(x,y)	Die Kreuzkorrelationsfunktion zweier Zeitreihen.
diff(y)	Erzeugen der Zeitreihe der einfachen Differenzen der Reihe y. Die Funktion berechnet die saisonalen Differenzen, wenn für das optionale Argument lag ein Wert angegeben ist.
end(y)	Der Zeitpunkt des letzten Zeitreihenwertes in Vektordarstellung. Z. B. für eine Reihe von Monatsdaten der Zeitpunkt Februar 2020 wird in Vektordarstellung "2020 2" ausgegeben.
ets(y)	Schätzung eines exponentiellen Glättungsmodells. (Funktion im R-Paket forecast enthalten.)
filter(y)	Lineare Filterung einer Zeitreihe.
frequency(y)	Anzahl Beobachtungen pro Zeitintervall.
is.na(y)	Überprüfen, ob y einen fehlenden Wert (NA) enthält.
is.ts(y)	Überprüfen, ob y eine Zeitreihe ist.
na.omit(y)	Entfernt die fehlenden Werte (NA's) aus der Zeitreihe y.
pacf(y)	Die empirische PACF; Abbildung enthält Schwankungsintervall $-1/N \pm 1.96 \cdot 1/\sqrt{N}$.
pacfmat(y)	Die empirischen partiellen Autokorrelationsmatrizen einer multivariaten Zeitreihe y. (Funktion im R-Paket tsapp enthalten.)
periodogram(y)	Berechnung und grafische Darstellung des Periodogramms der Reihe y an einer vorgegebenen Anzahl von gleichabständigen Frequenzen im Intervall [0,1/2]. (Funktion im R-Paket tsapp enthalten.)
predict(out)	Prognose einer Zeitreihe auf der Basis eines bestimmten Modells. out ist der Output der Modellschätzung, welcher z. B. mit der Funktion lm, ets oder arima generiert wurde.
start(y)	Der Zeitpunkt des ersten Zeitreihenwertes in Vektordarstellung.
ts(y)	Erzeugen einer Zeitreihe aus dem Vektor/der Matrix y.
ts.plot(y)	Zeitreihenplot. Die Zeitkomponente wird automatisch auf der x-Achse aufgetragen.
tsp(y)	Abfragen der Zeitreihenattribute.
window(y)	Erzeugen einer Zeitreihe aus einem Abschnitt der Zeitreihe y. Mit den Argumenten start und end wird der Abschnitt spezifiziert.
y[-c(1,2)]	Die 1. und die 2. Beobachtung werden aus Zeitreihe y gelöscht.

13.4 Bibliotheken und Programmierung

Bibliotheken

Es gibt neben den Funktionen in den Basispaketen viele zusätzliche Funktionen in weiteren Paketen. Mit dem Aufruf `library()` erhält man einen Überblick über die aktuell implementierten Pakete.

Der einfachste Weg, um zusätzliche Pakete zu installieren, besteht in dem Anklicken von „Installiere Paket(e)…“ bzw. „Installiere Paket(e) aus lokalen Zip-Dateien…“. Für das erste muss eine Internetverbindung bestehen. Hier geschieht bei den entsprechend vorbereiteten Paketen alles automatisch. Sogar die HTML-Hilfe wird aktualisiert.

Mit dem Befehl `library(Paket)` bekommt man Zugriff auf die Funktionen eines Paketes; für Paket ist natürlich das spezielle gewünschte Paket anzugeben, also etwa `tseries`.

Jedes Paket verfügt über eine Anleitung im Format pdf, in der die Funktionen beschrieben und mit Anwendungsbeispielen illustriert werden. Mit dem Befehl `data()` werden die verwendeten Datensätze aufgelistet. Diese Datensätze stellen auch ein guter Startpunkt für die ersten eigenständigen Analysen dar.

Eine weitere Option bietet das Paket „Time Series Data Library“ von Rob Hyndman, welche mehr als 600 Zeitreihen, klassifiziert nach Anwendungsbereich, enthält. Die Installationshinweise können unter <https://pkg.yangzhuoranyang.com/tsdl/> aufgerufen werden.

Kontroll-Strukturen

Das Schreiben von eigenen Funktionen macht vor allem Sinn, wenn man häufiger die gleiche Anweisungsfolge bei eventuell unterschiedlichen Parameterkonstellationen ausführen will. Dann benötigt man Kontroll-Strukturen, die den Fluss eines Programms regeln. R kennt zwei Arten von Kontroll-Strukturen, nämlich bedingte Anweisungen und Schleifen.

Bedingte Anweisungen werden über die `if`-Anweisung realisiert:

```
if (Bedingung) { Anweisungen }
```

Die in den geschweiften Klammern enthaltenen Anweisungen werden nur ausgeführt, falls die angegebene Bedingung wahr ist, d. h. den Wert TRUE besitzt. Optional kann die if-Anweisung auch einen `else`-Zweig enthalten, der ausgeführt wird, wenn die Bedingung falsch ist:

```
if (Bedingung) { Anweisungen } else { Anweisungen }
```

Beispiel 13.26

```
if (x==10){ k <- x+1 } else { k <- x-1 }
```

Die einfachste Form der *Schleife* in R ist die `while`-Anweisung. Auch hier folgt in einer in runden Klammern gesetzten Bedingung die auszuführende(n) Operation(en) in geschweiften Klammern. Diese in den geschweiften Klammern angegebene Anweisung (oder Menge von Anweisungen) wird solange ausgeführt, wie die Bedingung wahr ist.

```
while (Bedingung) { Anweisungen }
```

Beispiel 13.27

```
i <- 1; while (i < 4) { i <- i + 1; ;print(i) }
[1] 2
[1] 3
[1] 4
```

Diese Schleife wird also genau dreimal durchlaufen. Im letzten Durchlauf bekommt `i` den Wert 4; daher wird die weitere Bearbeitung gestoppt.

Eine weitere Form der *Schleife* in R ist die `for`-Schleife. Die Syntax der `for`-Anweisung ist sehr elegant:

```
for (Variable in Vektor) { Anweisungen }
```

Die Laufvariable `Variable` nimmt nacheinander die Werte von `Vektor` an, d. h. beim ersten Schleifendurchlauf hat sie den Wert des ersten Vektorelements, beim zweiten den Wert des zweiten und so weiter.

Beispiel 13.28

```
for(i in 1:5) { print(1:i) }
[1] 1
[1] 1 2
[1] 1 2 3
[1] 1 2 3 4
[1] 1 2 3 4 5
```

Mit Hilfe der Anweisung `break` kann eine Schleife unmittelbar verlassen werden. Die Anweisung `next` führt zum Beenden des aktuellen Schleifendurchlaufs.

Eigene Funktionen

Um eine Funktion zu schreiben, sind die Befehle in einem Skript als Text (ohne jede Formatierung) zu schreiben. Man kann auch einen anderen Texteditor verwenden.

Hierfür gibt es die Möglichkeit, von R aus einen Editor zu starten. Dies geschieht mit der Funktion `edit` oder mit `fix`:

```
fix(Funktionsname)
```

Über eine Option kann ein beliebiger Text-Editor ausgewählt werden. In dem Editor wird dann die Funktion geschrieben und als ‚Source-Code‘ in einer externen Datei mit der Dateierweiterung `r` abgespeichert. Die externe Datei kann auch mehrere selbstdefinierte Funktionen enthalten.

Eine Funktion wird nun mittels

```
MeineFunktion <- function(Argumente) { Anweisungen }
```

definiert. Diese Funktion steht nicht automatisch zur Verfügung. Sie muss vielmehr erst mit dem Befehl

```
source("myfunc.r")
```

initialisiert werden. Hierbei ist angenommen, dass `MeineFunktion` in der Datei `myfunc.r` abgespeichert ist.

Beispiel 13.29

```
# Funktion zur Bestimmung des Standardfehlers des Medians
# mittels Bootstrap
sdmedian <- function(x,B) {
  n <- length(x)
  m <- rep(0,B)
  for(b in 1:B){
    x1 <- sample(x,n,replace=TRUE)
    m[b] <- median(x1)}
  sd(m) }
```

13.5 Einlesen und Exportieren von Daten

Aus einer externen ASCII-Datei können Daten, die dort einfach hintereinander und/oder untereinander gelistet sind, mittels `scan` in einen Vektor eingelesen werden.

Beispiel 13.30

```
y <- scan("jlast.dat")
Read 37 items
```

Auch ein Datensatz lässt sich aus einer ASCII-Datei einlesen. In der ersten Zeile dieser Datei seien die Variablennamen eingetragen (in " " eingeschlossen und durch Leerzeichen separiert). Zudem muss die Datei schon die Struktur haben, die der Datensatz

dann bekommen soll. Die zu einer Variablen gehörigen Werte sind spaltenweise angeordnet; die Werte einer Zeile gehören zu einer Beobachtungseinheit. Die Angaben sind durch Leerzeichen voneinander getrennt. Dann erfolgt das Einlesen mittels der Funktion `read.table`.

Beispiel 13.31

```
firmen <- read.table("c:\\d\\mydata\\firmen2001.dat", header=TRUE)
```

Sind in der ersten Zeile der Datei die Namen der Variablen nicht angegeben, so lautet die Option `header=FALSE`.

`read.table` hat zahlreiche optionale Argumente; dazu gehören die Möglichkeiten, Dezimalkommata anstelle der Dezimalpunkte korrekt zu identifizieren und andere Trennungssymbole als Leerzeichen anzugeben.

Excel-csv-Dateien können in ähnlicher Weise mittels `read.csv` eingelesen werden.

13.6 Grafik

Es gibt etliche Typen von Grafiken, die auf einfache Weise Datensätze bzw. Vektoren oder Matrizen darstellen. Die Grafikfunktionen produzieren dabei Grafik-Typen, die von dem jeweiligen Objekt-Typ der Eingabe abhängen. Diese Grafikfunktionen gehören zu den sogenannten High-Level-Grafikfunktionen. Der Aufruf einer dieser Funktionen führt dazu, dass eine bereits vorhandene Grafik durch eine neue ersetzt wird. Um in einem Grafik-Fenster mehr als eine Grafik unterzubringen, kann der Befehl `par` eingesetzt werden. So wird durch `par(mfrow=c(1, 2))` die Grafik senkrecht geteilt. Andere Funktionen, die für diese Problemstellung angezeigt sind, sind `split.screen` und `layout`.

Zwei Grafik-Typen zur Darstellung univariater Daten bzw. Vektoren sind `hist` und `qqnorm`. `hist` bietet ein einfaches *Histogramm*. Hier wird neben dem Datenvektor die Angabe der Anzahl der Klassen erwartet. Erst durch Setzen von `include.lowest=TRUE` werden Werte berücksichtigt, die mit der untersten Klassengrenze übereinstimmen. Es kann auch ein Vektor von Klassengrenzen angegeben werden. Dann ist allerdings der gesamte Wertebereich abzudecken; die unterste Klassengrenze muss kleiner oder gleich dem kleinsten Wert in `x` sein und die oberste größer oder gleich dem größten `x`-Wert.

Beispiel 13.32

```
hist(runif(500), breaks=10)
hist(runif(500), breaks=10, include.lowest=TRUE)
hist(runif(500), breaks=c(0, 0.1, 0.25, 0.5, 0.75, 1))
```

`qqnorm` bietet ein *Normalverteilungs-Quantil-Quantil-Diagramm*. Auch diese Funktion ist einfach aufzurufen. Als Argumente werden zwei Vektoren angegeben. Sie müssen nicht gleich lang sein. Es werden die durch den kürzeren Vektor bestimmten Anteile als p -Werte für die Quantile genommen. Zu beachten ist hier, dass in R Quantile stets über lineare Interpolation ermittelt werden.

Beispiel 13.33

```
x <- rnorm(50)
qqnorm(x)
```

`plot` ist die Universalfunktion zur Darstellung bivariater Daten im *Streudiagramm* und für *Linienzüge*. Der Aufruf ist denkbar einfach: Sind `x` und `y` gleichlange Vektoren, so wird mit `plot(x, y, type="p")` ein Streudiagramm der (x,y) -Werte erstellt. Das optionale Argument `type="p"` verlangt dabei die Darstellung als Punkte. Möchte man einen Linienzug haben, so ist `type="l"` anzugeben. `type="o"` zeichnet beides, Linien und Punkte.

Die Darstellung einer Zeitreihe `y` mittels `plot` hängt davon ab, ob `y` Zeitreihenattribute besitzt. Dann leistet `plot(y)` schon alles. Andernfalls kann man `ts.plot(y)` aufrufen, um eine adäquate Darstellung zu erhalten.

Ein wichtiger Aspekt ist die Achsenbezeichnung. Automatisch werden die Variablennamen gewählt, die für die Platzhalter `x` und `y` eingesetzt werden. Möchte man das ändern, so ist `xlab` als Parameter für die x-Achse zu definieren; entsprechend steht `ylab` für die y-Achse.

Die vollen Möglichkeiten von `plot` sind ausufernd. Einige Optionen, die analog zu `type` angegeben werden können, sind:

Parameter	Effekt	Beispiel
<code>col</code>	Farbe	<code>col="red"</code>
<code>lwd</code>	Strichstärke	<code>lwd=2</code>
<code>main</code>	Gesamttitel	<code>main="Gesamttitel"</code>
<code>xlab</code>	Titel für die x-Achse	<code>xlab="X"</code>
<code>ylab</code>	Titel für die y-Achse	<code>ylab="Y"</code>
<code>pch</code>	Symbol für Punkte	<code>pch="o"</code>
<code>cex</code>	Symbolgröße	<code>cex=2</code>

Alternativ zur direkten Angabe der Optionen können diese auch mittels eines vorangestellten `par`-Befehls gesetzt werden. `par` wurde bereits zum Beginn des Abschnittes zur Unterteilung des Grafik-Objektes erwähnt.

Ist eine der oben angegebenen Grafiken erstellt, so können Linien, Punkte und Bezeichnungen hinzugefügt werden. Dies geschieht mit den sogenannten *Low-Level-Grafikfunktionen*. Ihr Aufruf erzeugt keine neue Grafik, sondern verändert eine bereits vorhandene. Die wichtigsten sind

Low-Level-Grafikfunktionen

abline	Fügt Linien zu dem aktuellen Plot hinzu. Anzugeben sind Achsenabschnitt und Steigung, welche die Gerade bestimmen
lines	Hinzufügen von Linien zu dem aktuellen Plot.
points	Hinzufügen von Punkten zu dem aktuellen Plot.
rug	Hinzufügen eines rug zum aktuellen Plot.
text	Hinzufügen von Textsymbolen zum aktiven Plot.
title	Hinzufügen eines Titels zum aktiven Plot.

Die Funktion `abline` ist vor allem günstig, um Regressionslinien zu Streudiagrammen oder QQ-Diagrammen hinzuzufügen. Verlangt werden als Argumente der Achsenabschnitt `a` und die Steigung `b`. Mit weiteren Argumenten können die Farbe der ab-Geraden sowie die Strichstärke eingestellt werden. Für eine Umsetzung sei auf das Beispiel 2.19 verwiesen.

Linienzüge mittels der Angabe von Punkten werden durch die Funktion `lines` erzeugt. Im Aufruf `lines(x, y)` sind `x` und `y` zwei gleichlange Vektoren. Sollen mehr als ein Linienzug einem Plot hinzugefügt werden, so ist `lines` mehrmals aufzurufen.

Beispiel 13.34

Zum Einfügen einer Ausgleichsgeraden brauchen jeweils nur Vektoren mit den Anfangs- und Endkoordinaten angegeben zu werden:

```
y <- 3 + 2*c(1:30) + rnorm(30)
plot(c(1:30),y)
lines(c(1,30),c(3+2*1,3+2*30))
```

Das Hinzufügen von Punkten mittels `points` erfolgt gemäß der gleichen Logik. Unter einem `rug` wird ein am Rande angebrachtes Stabdiagramm verstanden. Das ist z. B. sinnvoll, wenn man eine übersichtliche Darstellung mittels eines Histogramms haben und dennoch nicht auf eine detailliertere Angabe verzichten möchte, wo die Werte tatsächlich liegen; eine Anwendung ist im Beispiel 12.2 zu finden.

Als Titel kann der Grafik nun beispielsweise „Hübsches Bild“ gegeben werden. Der Text ist dabei in Anführungszeichen zu setzen: `title("Hübsches Bild")`.

Weiterer Text wird mit `text` in die Grafik eingefügt: `text(1,5,"Beginn")`. Der angegebene Punkt ist dann das Zentrum des Textfeldes. Nicht mehr ganz in das Ausgabefeld passen würde dementsprechend `text(1,5,"Beginn der Regressionsgeraden")`.

Mit der Anweisung `pdf(file="myfile.pdf")` wird die anschließend erzeugte Grafik als PDF-Datei in der Datei `myfile.pdf` gespeichert. Mit `dev.off()` wird das Abspeichern beendet.

Die aufgerufenen R-Funktionen

Die folgende Tabelle gibt einen Überblick über die bei den Beispielen aufgerufenen Funktionen, sofern sie nicht schon in einer der beiden Tabellen 13.1 oder 13.2 angegeben sind. Sie sind in alphabetischer Reihenfolge gelistet. Sofern sie zu einem der Pakete gehören, die jeweils extra geladen werden müssen, wird das zugehörige Paket mit angegeben. Die Pakete sind alle vom CRAN zu beziehen.

Funktionsname	Paket	Funktionsbeschreibung
acf		ACF und PACF inklusive Darstellung
acfmat	tsapp	Autokorrelationsmatrizen und damit zusammenhängende Größen
acfpacf	tsapp	Darstellung der ACF und PACF
ar		Anpassen eines AR-Modells
arima		Schätzen eines ARIMA-Modells
as.vector		Kovertireung in einen Vektor
aspectratio	tsapp	Bestimmung des Aspektverhältnisses
b.star	np	Bestimmen der Blocklänge beim Zeitreihenbootstrap
bandfilt	tsapp	Band-pass-Filterung
bkfilter	mFilter	Baxter-King-Filterung
bispecies	tsapp	Schätzen der des bivariaten Spektrums
BoxCox	tsapp	Bestimmen des Exponenten einer Box-Cox-Transformation
causality	vars	Kausalitätstest
ccf		CCF
coeftest	lmtest	Signifikanztests im linearen Regressionsmodell
diff		Bilden von Differenzen
dlmFilter	dlm	Kalman-Filterung
dlmForecast	dlm	Prognose eines mit dlmMLE geschätzten Objektes.
dlmMLE	dlm	Schätzen eines Zustandsraummodells.
dlmModPoly	dlm	Hilfsfunktion zu dlmMLE.
dlmModSeas	dlm	Hilfsfunktion zu dlmMLE.
dlmSmooth	dlm	Kalman-Glättung
durbinWatsonTest	car	Durchführung des Durbin-Watson-Tests.
eacf	TSA	ESACF
ets	forecast	Anpassen eines Exponentiellen-Glättungs-Modells.
filter		Filterung einer Zeitreihe.
fracdiff	fracdiff	Schätzen eines ARFIMA-Modells.
garch	tseries	Schätzen eines GARCH-Modells.
Grangercaus	tsapp	Überprüfen der Granger-Kausalität
HoltWinters		Holt-Winters-Glättung
leaps	leaps	Alle Teilregressionen
LjungBoxPierceTest	tsapp	Suite von Ljung-Box-Pierce-Tests.

Funktionsname	Paket	Funktionsbeschreibung
missar	tsapp	Auffüllen der Lücken mittels der AR-Methode.
missls	tsapp	Auffüllen der Lücken mittels der LS-Methode.
monthplot		Erzeugen eines saisonalen Teilreihenplots.
moveav	tsapp	Gleitende Durchschnitte
msmFit	MSwM	Anpassen eines Markov-Switching-Modells
NeweyWest	sandwich	HAC-Standardfehler
outidentify	tsapp	Identifizieren von Ausreißern in Zeitreihen.
pacfmat	tsapp	Bestimmung der mit PACF-Matrizen zusammenhängenden Größen.
periodogram	tsapp	Bestimmung des Periodogramms.
plotDiag	MSwM	diagnostische Grafik für ein Markov-Switching-Modell
plotProb	MSwM	Grafik für ein Markov-Switching-Modell
predict		Erstellen von Prognosen.
qunitroot	urca	Bestimmen der Quantile des DF-Unit-Root-Tests.
residuals		Bestimmen der Residuen eines angepassten Modells.
rlm	MASS	robuste Regression
RS	tsapp	R/S-Statistik
robsplinedecom	tsapp	Zerlegung einer Zeitreihe mittels robustifizierter Splines.
simpledecomp	tsapp	Einfache Zeitreihenzerlegung mit konstanter Saison und polynomialen Trend.
smoothls	tsapp	Splinesglättung.
smoothrb	tsapp	Robuste Splinesglättung.
specest	tsapp	Spektralschätzung.
specplot	tsapp	Plot einer Spektralschätzung
splinedecom	tsapp	Zerlegung einer Zeitreihe mittels Splines.
statchec	tsapp	Stationaritätscheck
stl		Robuste Zeitreihenzerlegung
symplot	tsapp	Darstellung eines Symmetriediagramms.
tacfARMA	Itsa	Berechnung der theoretischen Autokovarianzen eines ARMA-Modells.
TrenchForecast	Itsa	Optimale lineare Prognose.
ts		Macht einen Vektor zu einer Zeitreihe
ts.plot		Erstellt einen Zeitreihenplot aus dem Vektor x.
ur.df	urca	DF-Unit-Root-Test
VAR	vars	Schätzen eines VAR-Modells.
VARselect	vars	Bestimmung der Ordnung eines VAR-Modells.
Variogram	nlme	Bestimmung eines Variogramms.
vartable	tsapp	Bestimmung der Tafel der variaten Differenzen.
vcov		Bestimmung der Varianz-Kovarianzmatrix aus dem Ergebnis einer anderen Funktion.
window		Herausschneiden eines Zeitreihensegments.
wntest	tsapp	Test auf White-Noise.

Die Zeitreihen

Bezeichnung	Beschreibung	Seiten
ALKEINK (ALCINCOME)	Alkoholkonsum und Einkommen	242
AUFTRAGSEINGANG (INORDER)	Auftragseingänge bei einem Unternehmen	42, 53
BAUMRING (TREERING)	Jahresringe von Bäumen	3, 187, 189
BIER (BEER)	Bierproduktion in Australien	121
BIP (GDP)	preisbereinigtes Bruttoinlandsprodukt	116, 270
BIPORIG (GDPORIG)	Bruttoinlandsprodukt, Laspeyres-Indexwerte zur Basis 2000	156, 158
BISAM (MUSKRAT)	jährliche Verkäufe von Bisam-Fellen	227, 231
BLACKOUT	wöchentliche Anzahlen der Geburten in New York	4, 232, 235
DAX	Kurse des DAX	277, 280
DIABETES	Inzidenzen von Insulin abhängiger Diabetes Mellitus	83
DOMINANZ (DOMINANCE)	Umlaufrenditen öffentlicher Anleihen in Österreich und Deutschland	215, 220
FICHTE (SPRUCE)	Entwicklung des Einschlags von Fichtenholz	300, 304
FINANZ (FINANCE)	zwei Portfolio-Insurance-Strategien	24
HERZSCHLAG (HEARTBEAT)	Herzfrequenz von Patienten	28, 32–34, 36, 81
HSV	Tabellenplatz HSV	62, 164
IBM	Kurse der IBM-Aktie	170, 264, 273
JLAST (PPDEMAND)	Stromspitzenbedarf in Berlin	1, 59, 127
KAFFEE (COFFEE)	Kaffee-Konsums in den USA	102
L921	Grundwasserspiegel und Niederschlag am Brunnen L921	290
LITH	Niederschlagsmengen und Grundwasserstand am Brunnen LITH	196
LUCHS (LYNX)	jährliche Anzahlen von versteigerten Pelzen von Luchsen	18, 55, 133, 137, 263
LUCHSHASE (LYNXHARE)	Luchse und Schneehasen, simulierter VAR[1]-Prozess	199, 209
LUHORMON (LUHORMONE)	Level des Luteinzing-Hormons bei einer Kuh	298
MASCHINEN (ENGINES)	Auftragseingänge für Maschinen	64, 305, 306
MAUNALOA	CO ₂ -Konzentration über Mauna Loa	11, 58
MDAX	Kurse des MDAX	284
MELANOM	Melanom-Inzidenzen in Connecticut	206
NIL (NILEMIN)	jährliche minimale Wasserstände des Nil	176, 178, 180, 183
OXYGEN	Gehalt eines Oxygen Isotops	292, 294

Bezeichnung	Beschreibung	Seiten
PAPIER (PAPER)	zwei Größen bei einer papierproduzierenden Maschine	210
PRODINDEX	Produktionsindex des produzierenden Gewerbes	46, 47
REGEN (RAINFALL)	jährliche Niederschlagsmengen in LA	9, 13, 17, 34, 146
ROTWEIN (REDWINE)	monatlicher Verkauf von australischem Rotwein	124
SCHAUINSLAND	CO ₂ -Konzentration über Schauinsland	257
SCHWEINEPREIS (PIGPRICE)	Monatliche Preise für Schweine	2, 133, 144, 147
SPEISEEIS (ICECREAM)	Temperatur und Speiseeisverbrauch	193
STEUERN (TAXES)	monatlich in Deutschland gezahlte Gemeinschaftsteuern	45, 141
TMPYR	jährliche Durchschnittstemperatur in Zentralengland	184
TREMOR	Messungen des physiologischen Tremors	95, 99
UNFALL (ACCIDENT)	monatliche Anzahlen der Verkehrsunfälle mit Personenschäden in der BRD	40, 106, 238
USAPOP	Bevölkerung der USA	37
VERKÄUFE (SALES)	monatliche Verkäufe eines Unternehmens	228
WHORMON (WHORMONE)	Konzentration des Wachstumshormons eines Bullen	166, 170

Literatur

- Allende, H. und Heiler, S. (1992). Recursive Generalized M-Estimates for Autoregressive Moving-average Models. *Journal of Time Series Analysis*, **13**, 1–18.
- Anderson, B. O. und Moore, J. B. (1979). *Optimal Filtering*. Englewood Cliffs: Prentice Hall.
- Aoki, M. (1990). *State Space Modeling of Time Series*, 2. Aufl. Berlin: Springer.
- Aoki, M. und Havenner, A. M. (1997). *Applications of Computer Aided Time Series Modeling*. Berlin: Springer.
- Baxter, M. und King, R. G. (1999). Measuring business cycles: approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, **81**, 575–593.
- Belcher, J., Hampton, J. S. und Tunnicliffe Wilson, G. (1994). Parameterization of continuous time autoregressive models for irregularly sampled time series data. *Journal of the Royal Statistical Society, Series B*, **56**, 141–155.
- Benkwitz, A. (1999). Computing Bootstrap Confidence Intervals for Impulse responses. Unveröffentlichtes Manuskript.
- Beran, J. (1994). *Statistics for Long-Memory Processes*. London: Chapman & Hall.
- Bergstrom, A. R. (1990). *Continuous Time Econometric Modelling*. Oxford: Oxford University Press.
- Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*. New York: Wiley.
- Bollerslev, T. (1988). On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process. *Journal of Time Series Analysis*, **9**, 121–131.
- Bollerslev, T., Chou, R. Y. und Kroner, K. F. (1992). ARCH Modelling in Finance: a Review of the Theory and Empirical Evidence. *Journal of Econometrics*, **52**, 55–59.
- Box, G. E. P. und Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*, 2. Aufl. San Francisco: Holden-Day.
- Box, G. E. P. und Tiao, G. (1975). Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of the American Statistical Association*, **70**, 70–79.
- Brillinger, D. R. (1975). *Time Series, Data Analysis and Theory*. New York: Holt, Rinehart and Winston.
- Brockwell, P. J. und Davis, R. A. (1996). *Introduction to Time Series and Forecasting*. Berlin: Springer.
- Brooks, C. und Burke, S. P. (1998). Forecasting exchange rate volatility using conditional variance models selected by information criteria. *Economics Letters*, **61**, 273–278.
- Brown, R. G. (1962). *Smoothing, Forecasting, and Prediction*. Englewood Cliffs, N. J.: Prentice-Hall.
- Burg, J. P. (1975). *Maximum entropy spectral analysis*. Ph. D. Thesis, Stanford University, Palo Alto, CA.
- Catlin, D. E. (1989). *Estimation, Control, and the Discrete Kalman Filter*. Berlin: Springer.
- Chambers, J. M. und Hastie, T. J. (1992). *Statistical Models in S*. Pacific Grove: Wadsworth & Brooks/Cole.
- Chan, K. H., Hayya, J. L. und Ord, J. K. (1977). A Note on Trend Removal Methods – The Case of Polynomial Versus Variate Differencing. *Econometrica*, **45**, 737–744.
- Chan, W.-Y. T. und Wallis, K. F. (1978). Multiple Time Series Modelling: Another Look at the Mink-Muskrat Interaction. *Applied Statistics*, **27**, 168–175.
- Chang, I. und Tiao, G. C. (1983). Estimation of Time Series Parameters in the Presence of Outliers. Technical Report 8, Statistics Research Center, Univ. Chicago.
- Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business & Economic Statistics*, **11**, 121–135.
- Choi, B. (1992). *ARMA Model Identification*. Berlin: Springer.
- Clements, M. P. und Krolzig, H.-M. (1998). A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. *Econometric Journal*, **1**, 47–75.

- Cleveland, W. S. (1993). A Model for Studying Display Methods of Statistical Graphics. *J. Comp. Graph. Statistics*, **2**, 323–343.
- Cleveland, W. S. (1994). *The Elements of Graphing Data, Revised Edition*. Lafayette: Hobart Press.
- Cleveland, W. P. und Tiao, W. P. (1976). Decomposition of seasonal time series: a model for the Census X-11 Program. *Journal of the American Statistical Association*, **71**, 581–587.
- Cooley, J. W. und Tukey, J. W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Math. Computation*, **19**, 297–301.
- Cryer, J. D. und Chan, K.-S. (2010). *Time Series Analysis With Applications in R*. Berlin: Springer.
- Cusano, M. M. und Young, J. L. (1986). *Forty-five years of cancer incidence in Connecticut, 1935–79*. National Cancer Institute in Bethesda, Maryland.
- Dahlhaus, R. (1989). Efficient Parameter Estimation for Self-Similar Processes. *Annals of Statistics*, **17**, 1749–1766.
- Dalgaard, P. (2002). *Introductory Statistics with R*. Berlin: Springer Verlag.
- Davison, A. C. und Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.
- Dhrymes, P. (1998). *Time Series, Unit Roots, and Cointegration*. San Diego: Academic Press.
- Dichtl, H. und Drobetz, W. (2011). Portfolio insurance and prospect theory investors: Popularity and optimal design of capital protected financial products. *Journal of Banking & Finance*, **35**, 1683–1697.
- Draper, N. R. und Smith, H. (1981). *Applied Regression Analysis*, 2. Aufl. New York: Wiley.
- Durbin, J. und Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- Durbin, J. und Watson, G. S. (1951). Testing for serial correlation in least squares regression II. *Biometrika*, **38**, 159–178.
- Edel, K., Schäffer, K.-A. und Stier, W. (1997). *Analyse saisonaler Zeitreihen*. Heidelberg: Physica-Verlag.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7**, 1–26.
- Enders, W. (1995). *Applied Econometric Time Series*. New York: Wiley.
- Engle, R. F. und Granger, C. W. J. (1987). Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, **55**, 251–276.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **5**, 987–1007.
- Engle, R. F. (1995). *ARCH: Selected Reading*. Oxford: Oxford University Press.
- Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society, Series B*, **34**, 350–363.
- Fox, R. und Taqqu, M. S. (1986). Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series. *Annals of Statistics*, **14**, 517–532.
- Fritts, H. C. (1971). Multivariate techniques for specifying tree-growth and climatic relationships and for reconstructing anomalies in paleoclimate. *Journal of Applied Meteorology*, **10**, 845–864.
- Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2. Aufl. New York: Wiley.
- Förster, W. (1997). Zur Historie der Saisonbereinigung und Komponentenzerlegung bei ökonomischen Zeitreihen. In: Edel u. a. (1997), S. 3–21.
- Frühwirt-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Berlin: Springer.
- Gardner, Jr., E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, **4**, 1–28.
- Geweke, J. und Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221–238.
- Granger, C. W. J. (1966). The Typical Spectral Shape of an Economic Variable. *Econometrica*, **34**, 150–161.

- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, **37**, 424–438.
- Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, **14**, 227–238.
- Granger, C. W. J. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics*, **16**, 121–130.
- Granger, C. W. J. und Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Göttingen: Vandenhoeck & Ruprecht.
- Granger, C. W. J. und Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15–29.
- Granger, C. W. J. und Newbold, P. (1986). *Forecasting Economic Time Series*. San Diego: Academic Press.
- Granger, C. W. J. und Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.
- Greene, W. H. (2018). *Econometric Analysis*, 8th Edition. Harlow, Essex: Pearson Education.
- Grenander, U. und Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. New York: Wiley.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton N.J.: Princeton University Press.
- Hamilton, J. D. (1989). A new Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, **57**, 357–384.
- Hanau, A. (1928). *Die Prognose der Schweinepreise*. Sonderheft 7 der Reihe Vierteljahreshefte zur Konjunkturforschung. 2., ergänzte Auflage des Sonderheft 2 von 1927. Berlin: Verlag Reimar Hobbing.
- Hart, J. D. (1994). Automated kernel smoothing of dependent data using time series cross-validation. *Journal of the Royal Statistical Society, Series B*, **56**, 529–542.
- Harvey, A. (1994). *Ökonometrische Analyse von Zeitreihen*. München: R. Oldenbourg Verlag.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Haslett, J. und Raftery, A. E. (1989). Space-time modeling with long-memory dependence: assessing Ireland's wind power resource. *Applied Statistics*, **38**, 1–50.
- Hatanake, M. (1996). *Time-Series-Based-Econometrics*. Oxford: Oxford University Press.
- Hebbel, H. (1997). Verallgemeinertes Berliner Verfahren VBV. In: Edel u. a. (1997), S. 83–93.
- Heberle, J. und Sattarhoff, C. (2017). A Fast Algorithm for the Computation of HAC Covariance Matrix Estimators. *Econometrics*, **5**, 9.
- Holmes, E. E., Ward, E. J. und Scheuerell, M. D. (2018). *Analysis of multivariate timeseries using the MARSS package, version 3.10.10*. Northwest Fisheries Science Center, Seattle, WA.
- Höpfner, B. (1998). Ein empirischer Vergleich neuerer Verfahren zur Saisonbereinigung und Komponentenzerlegung. *Wirtschaft und Statistik*, **12/1998**, 949–959.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, **68**, 165–176.
- Hsiao, C. (1979). Autoregressive Modeling of Canadian Money and Income Data. *Journal of the American Statistical Association*, **74**, 553–560.
- Hsiao, C. (1981). Autoregressive Modelling and Money-Income Causality Detection. *Journal of Monetary Economics*, **7**, 85–106.
- Hurst, H. E. (1951). Long Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770–799.
- Hylleberg, S. (1992). *Modelling Seasonality*. Oxford: Oxford University Press.
- Hyndman, R. J. (2010). *Why every statistician should know about cross-validation*. URL: <http://robjhyndman.com/researchtips/crossvalidation/>.

- Hyndman, R. J. (2011). *Time series cross-validation: an R example*. URL: <http://robjhyndman.com/researchtips/tscvexample/>.
- Hyndman, R. J., Koehler, A. B., Ord, J. K. und Snyder, R. D. (2008). *Forecasting with Exponential Smoothing*. Berlin: Springer.
- Izenman, A. J. und Zabell, S. L. (1981). Babies and the Blackout: The Genesis of a Misconception. *Social Science research*, **10**, 282–299.
- Jaenicke, J. und Neck, R. (1996). Ist Österreich voll in die DM-Zone integriert? Eine ökonometrische Analyse. *Österreichische Zeitschrift für Statistik*, **25**, 33–64.
- Janacek, G. (2001). *Practical Time Series*. London: Arnold.
- Jenkins, G. M. und Watts, D. G. (1968). *Spectral Analysis and its applications*. San Francisco: Holden-Day.
- Jones, R. H. (1981). Fitting a continuous time autoregression to discrete data. In: Findley, D. F. (Hrsg.), *Applied Time Series Analysis II*. New York: Academic Press.
- Jones, R. H. (1980). Maximum likelihood fitting of ARIMA models to time series with missing data. *Technometrics*, **22**, 389–395.
- Keeling, C. D. und Whorf, T. P. (2000). *Atmospheric CO₂ concentrations (ppmv) derived from in situ air samples collected at Mauna Loa Observatory, Hawaii*. Scripps Institution of Oceanography (SIO), University of California, La Jolla.
- Keeling, C. D., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M. und Meijer, H. A. (2001). *Exchanges of atmospheric CO₂ and 13CO₂ with the terrestrial biosphere and oceans from 1978 to 2000*. I. Global aspects, SIO Reference Series, No. 01–06. Scripps Institution of Oceanography, San Diego.
- Kendall, M. G. (1973). *Time Series*. London: Griffin.
- Kirchgässner, G. (1981a). *Einige neuere statistische Verfahren zur Erfassung kausaler Beziehungen zwischen Zeitreihen*. Göttingen: Vandenhoeck & Ruprecht.
- Kirchgässner, G. (1981b). *Einige neuere statistische Verfahren zur Erfassung kausaler Beziehungen zwischen Zeitreihen*. Göttingen: Vandenhoeck & Ruprecht.
- Koopman, S. J., Harvey, A. C., Doornik, J. A. und Shepard, N. (1996). *Stamp: Structural Time Series Analyser, Modeller and Predictor*. London: Chapman & Hall.
- Koopmans, L. H. (1995). *The Spectral Analysis of Time Series*. San Diego: Academic Press.
- Koteswara, R. K. (1970). Testing for the independence of regression disturbances. *Econometrica*, **38**, 97–117.
- Krämer, J. (2011). Lineare Gaußsche Zustandsraummodelle und der Kalman Filter. Studienarbeit im Hochschulübergreifenden Studiengang Wirtschaftingenieurwesen, Hamburg.
- Krolzig, H.-M. (1997). *Markov-Switching Vektor Autoregressions*. Berlin: Springer.
- Li, W. K. und Mak, T. K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, **15**, 627–636.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.
- MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, **11**, 601–618.
- Mandelbrot, B. B. (1977). *Fractals: Form, Chance and Dimension*. San Francisco: Freeman.
- Meyer, N. (1997a). Die Bedeutung der glatten Komponente für die aktuelle Konjunkturdiagnose. In: Edel u. a. (1997), S. 101–108.
- Meyer, N. (1997b). Modifikationen des Census-Verfahrens durch die Deutsche Bundesbank. In: Edel u. a. (1997), S. 95–96.
- Muller, D. und Wei, W. W. S. (1997). Iterative least squares estimation and identification of the transfer function model. *Journal of Time Series Analysis*, **18**, 579–592.
- Naeve, P. und Steinecker, J. (1983). Ökologie und Statistik, Eine Zeitreihenanalytische Fallstudie. Diskussionsarbeit No. 133/1983, Universität Bielefeld, Fakultät für Wirtschaftswissenschaften.

- Nelson, C. R. und Kang, H. (1981). Spurious Periodicity in Inappropriately Detrended Time Series. *Econometrica*, **49**, 741–751.
- Nelson, C. R. und Kang, H. (1984). Pitfalls in the Use of Time as an Explanatory Variable in Regression. *Journal of Business & Economic Statistics*, **2**, 73–82.
- Nerlove, M., Grether, D. M. und Carvalho, J. L. (1979). *Analysis of Economic Time Series*. New York: Academic Press.
- Neumaier, A. und Schneider, T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, **27**, 1, 27–57.
- Newey, W. K. und West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.
- Newton, H. J. (1988). *TIMESLAB: A time series analysis laboratory*. Pacific Grove, CA: Wadsworth & Brooks/Cole.
- Newton, H. J. und Pagano, M. (1984). Simultaneous confidence bands for autoregressive spectra. *Biometrika*, **71**, 197–202.
- Noack, T. (2003). *Probleme bei der SETAR-Modellierung in der Zeitreihenanalyse*. Berlin: Logos Verlag.
- Noack, T. und Schlittgen, R. (2000). Nonparametric estimation of missing values in time series. *Allgemeines Statistisches Archiv*, **84**, 23–32.
- Oberhofer, W. und Kmenta, J. (1974). A general procedure for obtaining maximum likelihood estimates in generalizing regression models. *Econometrica*, **42**, 579–590.
- Opsomer, J., Wang, Y. und Yang, Y. (2001). Nonparametric Regression with Correlated Errors. *Statistical Science*, **16**, 134–153.
- Parzen, E. (1984). *Time Series Analysis of Irregularly Observed Data*. New York: Springer.
- Peña, D., Tiao, G. C. und Tsay, R. S. (2001). *A Course in Time Series Analysis*. New York: Wiley.
- Petris, G., Petrone, S. und Campagnoli, P. (2009). *Dynamic Linear Models with R*. Berlin: Springer.
- Pole, A., West, M. und Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. New York: Chapman & Hall.
- Politis, D. N., Romano, J. P. und Wolf, M. (1999). *Subsampling*. Berlin: Springer.
- Politis, D. N. und White, H. (2004). Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews*, **23**, 53–70.
- Politis, D. N., White, H. und Patton, A. J. (2009). Correction: Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews*, **28**, 372–375.
- Pollock, D. S. G. (1999). *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. San Diego: Academic Press.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series, Vol 1: Univariate Series, Vol 2: Multivariate Series, Prediction and Control*. London: Academic Press.
- Quenouille, M. H. (1957). *The Analysis of Multiple Time Series*. London: Charles Griffin.
- R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org>.
- Reinsch, C. (1967). Smoothing by spline functions. *Numerische Mathematik*, **10**, 177–183.
- Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*. Berlin: Springer.
- Rinne, H. und Specht, K. (2002). *Zeitreihen, Statistische Modellierung, Schätzung und Prognose*. München: Vahlen.
- Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, **23**, 1048–1072.
- Schimek, M. G. (2000). *Smoothing and Regression*. New York: Wiley.
- Schlittgen, R. (1990). Robuste Glättung ökonomischer Zeitreihen. *Allgemeines Statistisches Archiv*, **74**, 223–250.

- Schlittgen, R. (1991a). Ein moderner Zugang zur Analyse von Zeitreihen mit Ausreißern. *Allgemeines Statistisches Archiv*, **75**, 75–102.
- Schlittgen, R. (1991b). Resistant decomposition of economic time series. *Empirica*, **18**, 47–63.
- Schlittgen, R. (1997). Fitting of threshold models for time series. Discussion paper 97-13, UCSD.
- Schlittgen, R. (2004). *Statistische Auswertungen mit R*. München: R. Oldenbourg Verlag.
- Schlittgen, R. und Streitberg, B. H. J. (1999). *Zeitreihenanalyse*, 9. Aufl. München: R. Oldenbourg Verlag.
- Schneider, T. und Neumaier, A. (2001). A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, **27**, 1, 58–65.
- Schröer, G. (1998). *Schätz- und Testverfahren im strukturellen Kointegrationsmodell*. Frankfurt am Main: Peter Lang.
- Shumway, R. H. und Stoffer, D. S. (2000). *Time Series Analysis and its Applications*. Berlin: Springer.
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, **48**, 1–48.
- Samarov, A., Smith, R. L. und Taqqu, M. (1992). *Estimating the intensity of long-range dependence through regression*. Arbeitspapier.
- Speth, H.-T. (2004). *Komponentenzerlegung und Saisonbereinigung ökonomischer Zeitreihen mit dem Verfahren BV4.1*. Methodenberichte, Heft 3. Wiesbaden: Statistisches Bundesamt.
- Süsselbeck, B. (1993). *S und S-Plus, Eine Einführung in Programmierung und Anwendung*. Stuttgart: Gustav Fischer Verlag.
- Teräsvirta, T. (1998). Modelling economic relationships with smooth transition regression. In: Ullah, A. und Giles, D. E. A. (Hrsg.), *Handbook of Applied Economic Statistics*, S. 507–552.
- Thomas, G., Plu, G. und Thalabard, J.-C. (1992). Identification of pulses in hormone time series using outlier detection methods. *Statistics in Medicine*, **11**, 2133–2145.
- Timmer, J. (1998). Modeling noisy time series: Physiological tremor. *International Journal of Bifurcation and Chaos*, **8**, 1505–1516.
- Tong, H. (1990). *Non-linear Time Series; a Dynamical System Approach*. Oxford: Clarendon Press.
- Tsay, R. S. und Chen, R. (2018). *Nonlinear Time Series Analysis*. New York: Wiley.
- Tsay, R. S., Peña, D. und Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrika*, **87**, 789–804.
- Tsay, R. S. und Tiao, G. C. (1984). Consistent estimates of autoregressive parameters an extended sample autocorrelation function for stationary and non-stationary ARIMA models. *Journal of the American Statistical Association*, **79**, 84–96.
- Ulrich, T. J. und Ooe, M. (1979). Autoregressive and Mixed Autoregressive-Moving Average Models and Spectra. In: Haykin, S. (Hrsg.), *Nonlinear Methods of Spectral Analysis*, S. 73–125. Berlin: Springer.
- Van Winkel, E. G. F. (1982). Interpretation of Time Series Models: A Simplified Model for the Census X-11 Program. In: Anderson, O. D. (Hrsg.), *Time Series Analysis, Theory and Practice 1*. Amsterdam: North Holland.
- Venables, W. N. (2000). *S-Programming*. Berlin: Springer Verlag.
- Venables, W. N. und Ripley, B. D. (1994). *Modern Applied Statistics with S-Plus*. Berlin: Springer Verlag.
- Waldhoer, T., Schober, E. und Tuomilehto, J. (1997). Long-term patterns in seasonality of insulin-dependent diabetes mellitus in Austrian children. *Journal of Clinical Epidemiology*, **50**, 159–165.
- Wei, W. S. (1990). *Time Series Analysis, Univariate and Multivariate Methods*. New York: Addison Wesley.
- West, M. und Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2. Aufl. Berlin: Springer.

- Whittaker, E. T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41, 63–75.
- Whittaker, E. T. und Robinson, G. (1924). *The Calculus of Observations*. London: Old Bailey.
- Yaffee, R. (2000). *Introduction to Time Series Analysis and Forecasting*. San Diego: Academic Press.
- Yule, G. U. (1927). On a method of investigating prieridicities in disturbed seies, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Roal Society of London A*, 226, 267–298.

Abkürzungen und Symbole

Abkürzungen

ACF	14	SGARCH	284
AIC	80, 100	STAR	275
AICC	80, 100	STL	305
AO	295	TAR	273
AR	69	ULS	74
ARARCH	282	VAR	201
ARCH	276	VARMA	201
ARFIMA	174	VMA	201
ARIMA	101		
ARMA	92	YW	74
BIC	80, 100, 209		

CCF 188
CLS 73

Symbole

df	30	\doteq, \approx	21
DW	36	\equiv	44
		$\lfloor \cdot \rfloor$	138
EAR	275	$A(B)$	201
EBW	162	$\alpha(B)$	92
ESACF	98	B, B^d	56
		$\beta(B)$	92
FDWN	174	$B(B)$	201
FFT	139	(c_τ)	14
GARCH	283	$C_p(q)$	33
HAC	230	$c_{XY}(\tau)$	189
HQ	81, 100	$C_{XY}(\lambda)$	195
IACF	98	Δ, Δ^d	57
IO	296	Δ_s	57
		ε	160
MA	87	$f(\lambda)$	149
MAD	302	$f_{XY}(\lambda)$	195
ML	74	$F_a(\lambda)$	153
mod	44	\mathcal{F}_{m_1, m_2}	31
MSAR	268	$G_a(\lambda)$	153
PACF	78	$\gamma(\tau)$	293
SARIMA	106	$\gamma_Y(\tau)$	17
SBC	80	γ_τ	16
SETAR	273	$\gamma_{XY}(\tau)$	188
		$\Gamma(u)$	174
		$\Gamma(\tau)$	231
		$\Gamma_{XY}(\tau)$	190
		$\hat{\gamma}_N(\tau)$	21
		H	176

$I(k)$	57	$\rho_Y(\tau)$	17
$I(\lambda)$	139	$\rho_{XY}(\tau)$	188
$J(\lambda)$	143	$\hat{\rho}_{ij}(\tau)$	206
$K_{XY}(\lambda)$	195	R/S	176
\mathcal{L}	74	SS_{Fehler}	30
$\lambda_{1,t_0}, \lambda_{2,t_0}$	296	SS_{Modell}	30
$M(\tau)$	209	SS_{Total}	30
N	5	$S(\tau)$	208
$v(B)$	233	Σ_p	75
(π_t)	78	$V(\tau)$	293
$\Pi(\tau)$	208	$Y(t)$	293
Q	82	$y_{(i)}$	55
$Q_{XY}(\lambda)$	195	$\hat{y}_{N,1}$	61
(r_τ)	14	$\hat{Y}_{N,h}$	59
$r_{XY}(\tau)$	189	$\hat{Y}_{N,v}$	123
R^2	31	$\hat{Y}_{t,1}, \hat{Y}_{t,h}$	122
R_{adj}^2	31	ϕ	199
$R_{XY}(\tau)$	190	$\phi(\cdot)$	34
\mathcal{R}	199	$\phi_{XY}(\lambda)$	195
ρ_τ	16	χ_n^2	82
$\rho(\tau)$	293	$\psi(u)$	303

Sachindex

- absolut summierbar 21, 87
Aliasing 140
AR-Prozess *siehe* Prozess, autoregressiver
ARARCH-Prozess *siehe* Prozess, ARARCH-
ARCH[q]-Prozess *siehe* Prozess, bedingt
heteroskedastischer
ARFIMA-Prozess *siehe* Prozess, ARFIMA-
ARMA-Prozess *siehe* Prozess, Autoregressiver
Moving-Average-
Aspektverhältnis 11
Ausreißer 4, 50, 295
additiver 295
innovativer 296
iterative Erkennung 297
Autokorrelationsfunktion 14
AR-Prozess 71
ARMA-Prozess 94
asymptotische Eigenschaften 22
empirische 14
erweiterte Stichproben- 98
FDWN-Prozess 175
inverse 98
MA-Prozess 90
partielle 77, 78, 94
MA-Prozess 90
Random-Walk 70
theoretische 16
Autokorrelationsmatrix, partielle 208
Autokovarianzfunktion 14
AR-Prozess 71
asymptotische Eigenschaften 21
empirische 14
FDWN-Prozess 175
gefilterte Reihe 152
MA-Prozess 87
theoretische 16

Backshift-Operator 56
Bandbreite
äquivalente 162
eines Spektralschätzers 162, 169
Bartlett, Formel von 91
Bartlett-Priestley-Fenster 168
Beobachtungsgleichung 248
Bestimmtheitsmaß 31
adjustiertes 31

Bias 22
eines Spektralschätzers 161
Bootstrap 23, 76
zirkulärer 23
Box-Cox-Transformation 52
Box-Jenkins-Ansatz 94
Box-Jenkins-Prognose 123
Burg *siehe* Schätzung, Burg

 C_p -Statistik von Mallows 33
charakteristische Gleichung 92
AR-Prozess 71
MA-Prozess 88
circadian *siehe* Zykluslänge
Conditional Least Sums of Squares *siehe*
Schätzung, Conditional Least Sums of
Squares
Cramér-Darstellung 136

Datenfenster *siehe* Taper
Differenzen 57, 101
saisonale 57, 105
variate 58, 101
Drift 5
Dummyvariable 44
Durchschnitt, gleitender 39, 56
15er von Spencer 41
von Henderson 41

Einheitskreis 71
EM-Algorithmus 269
Ergodizität 7, 21
Exponentialmodell 36
exponentielles Glätten 61

Fehlerkorrektur-Darstellung 120, 253
Fehlerkorrekturmodell 221
Fenster
Lag- 167, 231, 246
Periodogramm-
diskretes 160
stetiges 167
Filter
Band-pass 155
Baxter-King- 158
High-pass 154
Hodrick-Prescott- 42

- Kalman- 252, 254
 - linearer 56, 152
 - Low-Pass 154
 - rekursiver 61
 - Filterung 56, 252
 - Formel von Bartlett 91
 - Fourierfrequenz *siehe* Frequenz, Fourier-Frequenz 152, 153
 - schnelle 139, 231
 - Freiheitsgrad 30
 - Frequenz 134
 - Cutoff- 155
 - Fourier- 138
 - Nyquist 140
 - Gainfunktion 153, 154
 - Gamma-Funktion 174
 - GARCH-Prozess *siehe* Prozess, GARCH-Prozess 37
 - Gauß-Newton-Verfahren 37
 - Gauß-Prozess *siehe* Prozess, Normal-Gaußsche Klammer 138
 - Gedächtnis
 - kurzes 173
 - langes 173
 - generalisierte Inverse 256
 - Glätten, exponentielles 61
 - Glättung 252
 - Splines 46
 - Glattheit 42
 - Gleichgewichtszustand 255
 - GM-Schätzer *siehe* Schätzung, generalisierte M-
 - Granger-Kausalität 214
 - Grenzyklus 273
 - HAC-Standardfehler 230
 - Hodrick-Prescott-Filter 42
 - Holt-Winters-Verfahren 63, 250
 - Hurst-Effekt 176
 - Hurst-Koeffizient 176
 - Hyperparameter 248
 - Identifikation *siehe* Spezifikation
 - Identifizierbarkeit 213
 - Impakt-Multiplikator 202
 - Impulsresponseanalyse 202
 - Impulsresponsefunktion 233
 - Informationskriterium 100
 - von Akaike 80, 100
 - von Hannan-Quinn 81, 100
 - von Hurvich-Tsay 80, 100
 - von Schwarz 80, 100, 209
 - infradian *siehe* Zykluslänge
 - Innovationen 70
 - Innovationsalgorithmus 89
 - Input 41, 56
 - Integrationsgrad *siehe* Prozess, integrierter
 - Interpolation, KQ- 290
 - Intervention
 - sanalyse 232, 291
 - svariable 232
 - Invertibilitätsbedingung 201
 - Invertierbarkeit 88
 - Irrfahrtsprozess *siehe* Prozess, Random-Walk-
 - Kalman-Filter *siehe* Filter, Kalman-Smooth 255
 - Kalman-Verstärkung 254
 - Kohärenz 195
 - Kointegration 217
 - Kointegrationsvektor 217
 - komplexe De- und Remodulation 155
 - Komponente
 - glatte 26, 27
 - Saison- 26
 - Komponentenmodell *siehe* Modell, Komponentenzerlegung 43
 - Korrekturschritt *siehe* Filter, Kalman-Korrektur 254
 - Korrelationsfunktion, bivariate 190
 - Kovarianzfunktion, bivariate 190
 - KQ-Ansatz 29
 - KQ-Methode *siehe* Methode, der kleinsten Quadrate 188, 241
 - empirische 189, 241
 - Kreuzkorrelationsfunktion 188, 241
 - Kreuzkovarianz 138
 - Kreuzkovarianzfunktion 188
 - empirische 189
 - Kreuzspektrum 195
 - Kreuzvalidierung 42
 - Lag 13
 - Lagfenster 167, 231, 246
 - Leakage 140, 142, 163
 - Levinson-Durbin-Rekursion 74, 77, 126
 - M-Schätzer 180, 302
 - MA-Prozess *siehe* Prozess, Moving-Average-Prozess 290

- Markov-Kette 266
 Maximum Likelihood *siehe* Schätzung,
 Maximum Likelihood-
 Median der absoluten Abweichungen vom
 Median 302
 Methode
 der kleinsten Quadrate 28
 generalisierte Kleinst-Quadrate- 227
 ML-Schätzung *siehe* Schätzung, Maximum
 Likelihood-
 Modell
 differenzenstationäres 111
 Interventions- 234
 Komponenten- 26, 27, 44
 lineares Regressions- 28
 lineares Zeitreihen- 69
 sparsames 32
 strukturelles 249
 Subset- 82
 Transferfunktions- 239
 trendstationäres 111
 Zustandsraum- 247, 248
 Modellvorstellung 7

 Noise-Reihe 235

 Oberschwingung 135, 140, 141
 Output 41, 56

 P-Wert 31
 Parsevalsches Theorem 149
 partielle Autokorrelationsfunktion *siehe*
 Autokorrelationsfunktion, partielle
 Periode 134
 Periodizität 133
 Modell einer verborgenen 143
 Test auf 143
 Periodogramm 138, 139
 kumulierte 146
 modifiziertes 143
 Periodogrammfenster *siehe* Fenster,
 Periodogramm-
 Phase 134, 195
 Potenzmodell 37
 Prädiktion *siehe* Prognose
 Prädiktionschritt *siehe* Filter, Kalman-
 Prewhitening 192, 242
 Prior, diffuse 255

 Prognose 59, 252
 -band 123
 -fehler 122, 123
 -horizont 59
 -intervall 59, 119, 123
 -h-Schritt 59, 123
 ARIMA-Modell 122
 optimale 122, 126
 prognostizierter Wert 28
 Prozess
 antipersistenter 175
 AR- 69
 ARARCH- 282
 ARFIMA- 174
 ARIMA- 101
 saisonaler 106
 ARMA- 92, 250
 autoregressiver 69
 Autoregressiver Integrierter Moving Average-
 101
 Autoregressiver Moving-Average- 92
 bedingt heteroskedastischer 276
 deterministischer 88
 GARCH- 283
 harmonischer 136
 in stetiger Zeit 293
 integrierter 57
 invertierbarer 88
 kointegrierte -e 217
 linearer 88, 165
 Markov-switching autoregressiver 268
 Moving-Average- 87
 Normal- 5
 Random-Walk- 5, 56
 SETAR- 273
 Smooth Transition Autoregressiver 275
 stationärer 16
 stochastischer 4
 Threshold Autoregressiver 273
 unkorrelierte -e 192
 Vektor-Autoregressiver 201
 Vektor-Autoregressiver-Moving-Average- 201
 Vektor-Moving-Average- 201
 VMA 201
 White-Noise- 5
 bivariater 191

 QQ-Diagramm 34

- Random-Walk *siehe* Prozess, Random-Walk-Regression
 lineare 28
 mit autokorrelierten Störungen 225
 Regressionskoeffizienten 31
 Regressoren 28
 Residuen 28
 RGM-Schätzung *siehe* Schätzung, rekursiv
 generalisierte M-
 Robustheit 301
- Saison 4, 63, 105
 -aler Teilreihenplot 47
 -bereinigung 43
 Berliner Verfahren 49
 CENSUS X-11 48
 CENSUS X-12 ARIMA 48
 -komponente 26
 Schätzung
 Ansatz von Durbin-Hannan-Rissanen 93
 Burg 75
 Conditional Least Sums of Squares 73, 89, 93
 generalisierte M- 308
 M- 180, 302
 Maximum Likelihood- 74, 89, 93
 rekursiv generalisierte M- 308
 Spektral-
 direkte 160
 dynamische 170
 indirekte 167
 parametrische 169
 Unconditional Least Sums of Squares 74, 89,
 93
 Yule-Walker- 74, 169
 multivariate 205
 Schwingung, harmonische 134
 SETAR-Prozess *siehe* Prozess, SETAR-
 Slutsky-Effekt 159
 Spannweite, reskalierte adjustierte 176
 Sparsamkeit 86
 Spektral-Repräsentationstheorem 136
 Spektralschätzer *siehe* Schätzung, Spektral-
 Spektrum 149
 AR-Prozess 153
 ARMA-Prozess 153
 Ko- 195
 integriertes 198
 Kreuz- 195
 kumulierte Phasen- 199
- MA-Prozess 153
 Quadratur- 195
 Schätzen des -s 160
 White-Noise- 150
 Spezifikation
 AR-Modell 78
 ARIMA-Modell 101
 ARMA-Modell 94, 100
 Splines 41, 46, 291
 glättende 41
 Stützbereich 59
 STAR-Prozess *siehe* Prozess, Smooth Transition
 Autoregressiver
 Stationarität 7, 16
 gemeinsame 188
 Mittelwert- 16
 schwache 16
 strenge 16
 Varianz- 16
 Stationaritätsbedingungen 71
 Stationaritätsdreieck 72
 stochastischer Prozess *siehe* Prozess,
 stochastischer
 Streuungszerlegung 30
 Symmetriediagramm 55
 Symmetrisierung 54
 Systemgleichung 248
- Taper 164
 Tapermodifikation 164
 TAR-Prozess *siehe* Prozess, Threshold
 Autoregressiver
 Test
 ADF- 116
 auf Ausreißer 296
 auf Kointegration
 Engle-Granger-Methode 220
 auf Periodizität 143
 auf White-Noise 146, 147
 Bartletts Kolmogorov- 146
 DF- 114
 Dickey-Fuller-Unit-Root- 114
 Durbin-Watson- 35
 Einheitswurzel- 114
 erweiterter Dickey-Fuller- 116
 F- 31, 214
 Ljung-Box-Pierce- 82
 Portmanteau- 82, 246

- t*- 31
- Unit-Root- 114
- Theorem
 - Parsevalsches 149
 - Spektral-Repräsentations- 136
- Totzeit 240
- Trajektorie 4
- Transferfunktion 153, 239
- Transformation
 - Box-Cox- 52
 - instantane 52
- Trend 4, 26, 27
 - bereinigung 27
 - bestimmung 27
 - extrapolation 59
 - nicht-polynomialer 39
 - periodischer 136
 - polynomialer 28
- Trendfunktion
 - Gomberts- 37
 - logistische 37
- Trendkomponente 28, 39
- Überanpassung 32, 86
- Übergangsmatrix 248
- Übergangswahrscheinlichkeit 266
- ultradian *siehe* Zykluslänge
- Unconditional Least Sums of Squares *siehe*
 - Schätzung, Unconditional Least Sums of Squares
- Unschärferelation 162
- VAR-Prozess *siehe* Prozess,
 - Vektor-Autoregressiver
- Variable
 - Indikator- 291
 - kointegrierte -n 217
- Varianzanalyse-Tafel 30
- Variogramm 293
- VARMA-Prozess *siehe* Prozess,
 - Vektor-Autoregressiver-Moving-Average-
- Verfahren von Hsiao 214
- Volatilität 275
- Welle, harmonische 134
- Wert, fehlender 255, 289
- White-Noise-Prozess *siehe* Prozess,
 - White-Noise-
- Whittaker-Graduation 42
- Winsorisierung 302
- Wold 202
- Woldscher Zerlegungssatz 88
- Yule-Walker *siehe* Schätzung, Yule-Walker-Gleichungen 71
- Zeitpfad 4
- Zeitreihe 4
 - zweidimensionale 187
- Zeitreversibilität 263
- Zentrierung 14
- Zustand 266
- Zustandsraummodell 248
- Zyklus 4, 134, 273
- Zykluslänge 2

