

Thomas Mazzoni

Zeitreihenanalyse

Einstieg und Aufgaben

**wirtschafts
wissenschaft**



FernUniversität in Hagen

Vorwort

Die Analyse ökonomischer Zeitreihen ist ein komplexes Gebiet der quantitativen Wirtschaftsforschung. Das notwendige mathematische Instrumentarium ist anspruchsvoll und in der Regel nicht oder nur teilweise Bestandteil der wirtschaftswissenschaftlichen Ausbildung. Dieses Skript verfolgt daher das Ziel, den Einstieg in die Welt der quantitativen Analyse ökonomischer Zeitreihen zu erleichtern. Es findet sich hier ein buntes Sammelsurium aus veranschaulichenden Erklärungen, Aufgaben, Kochrezepten und mathematischen Basics, die Studierenden der Wirtschaftswissenschaften den Einstieg in die quantitativen Konzepte erleichtern sollen.

Dieses Skript kann sowohl als begleitendes Material zum eigentlichen Kurs parallel bearbeitet sowie auch dem Kurs vorgeschaltet werden. Letztgenannte Strategie empfiehlt sich für Studierende ohne ausgesprochene mathematische Vorbildung. Keinesfalls ist dieser Einstieg geeignet, das Studium des Kurses 00889 zu ersetzen, da dort noch viele weiterführende Konzepte behandelt werden. Er soll vielmehr als Generator für das Verständnis der hinter den formalen Verfahren liegenden Ideen fungieren. Mit dem Verständnis dieser Ideen degeneriert der zugehörige mathematische Ballast zu einer alternativen Darstellung, die sich nun einfacher verstehen lässt, da eine intuitive Referenz vorhanden ist.

Mein Dank gilt Herrn Prof. Dr. Hermann Singer und meinen Kolleginnen Marina Lorenz, Daniela Doliwa und Anja Bittner für gewissenhafte Lektüre und vielfältige Unterstützung.

Inhaltsverzeichnis

1	Zeitreihen im Allgemeinen	1
2	AR-Prozesse und Elementares	5
2.1	Der AR(1)-Prozess	5
2.1.1	Zentrierter AR(1)-Prozess	6
2.1.2	Stationarität des AR(1)-Prozesses	6
2.1.3	Wissenswertes über den AR(1)-Prozess	11
2.2	Autokovarianz und Autokorrelation	12
2.2.1	Autokovarianz	14
2.2.2	Autokorrelation	15
2.2.3	Partielle Autokorrelation	16
2.3	Parameterschätzung	20
2.3.1	Parameterschätzung und Stationarität	22
2.3.2	Kleinste Quadrate (KQ)	24
2.3.3	Maximum Likelihood (ML)	28
2.3.4	Konfidenzintervalle	32
2.4	Der AR(2)-Prozess	36
2.4.1	Operatoren	37
2.4.2	Stationarität des AR(2)-Prozesses	39
2.4.3	Darstellung in Polar-Koordinaten	43
2.4.4	Autokovarianz und Autokorrelation	46
2.4.5	Parameterschätzung	50
2.5	AR(p)-Prozesse	50
2.5.1	Stationarität des AR(p)-Prozesses	51
2.5.2	Autokovarianz und Autokorrelation	51
2.5.3	Parameterschätzung	52
3	MA-Prozesse	55
3.1	AR(1)- und MA(unendlich)-Darstellung	56
3.2	Der MA(1)-Prozess	56
3.2.1	Parameterschätzung	59
3.3	MA(q)-Prozesse	61
3.3.1	Autokovarianz und Autokorrelation	61
3.3.2	Invertierbarkeit	64
3.3.3	Parameterschätzung	66

INHALTSVERZEICHNIS

4	ARMA-Prozesse und Erweiterungen	69
4.1	Der ARMA(1,1)-Prozess	70
4.1.1	Parameterschätzung	73
4.2	ARIMA und SARIMA	74
4.2.1	ARIMA(p, d, q)-Modelle	74
4.2.2	Seasonal ARIMA-Modelle	77
4.3	ARMAX-Modelle	78
4.3.1	Das ARMAX(1,0,1)-Modell	79
4.3.2	ARMAX(p, s, q)-Modelle	81
4.4	Modellidentifikation	82
4.4.1	<i>Akaike</i> -Informationskriterium	83
4.4.2	<i>Bayes</i> -Informationskriterium	85
4.4.3	Modellselektion	86
5	Bedingt heteroskedastische Prozesse	89
5.1	ARCH-Modelle	91
5.1.1	Das ARCH(1)-Modell	91
5.1.2	Stationarität des ARCH(1)-Prozesses	93
5.1.3	Kurtosis im ARCH(1)-Modell	95
5.1.4	Parameterschätzung im ARCH(1)-Modell	98
5.1.5	Das ARCH(p)-Modell	101
5.2	GARCH-Modelle	103
5.2.1	Das GARCH(1,1)-Modell	104
5.2.2	Das GARCH(p, q)-Modell	105
5.3	ARMA-GARCH und asymmetrische Erweiterungen	106
5.3.1	ARMA-GARCH	107
5.3.2	Asymmetrische GARCH-Modelle allgemein	108
5.3.3	Threshold ARCH	109
5.3.4	Quadratic GARCH	110
5.3.5	Exponential GARCH	110
5.3.6	Symmetrische vs. asymmetrische Modelle	113
5.3.7	ARCH in Mean	114
5.4	Modellidentifikation und Selektion	116
5.4.1	ARMA vs. GARCH	116
5.4.2	Modellordnung und Selektion	118
6	Zustandsraummodelle und <i>Kalman</i>-Filter	119
6.1	Das lineare Filterproblem	119
6.2	Zustandsraummodelle	122
6.2.1	Architektur des Zustandsraummodells	123
6.2.2	AR(p)-Modelle und VAR(1)-Darstellung	126

6.2.3	ARMA-Modelle in Zustandsraumform	127
6.3	<i>Kalman</i> -Filter	129
6.3.1	Die <i>Bayes</i> -Formel	131
6.3.2	Mess-Update	132
6.3.3	Innovation und effizientes <i>Kalman</i> -Update	135
6.3.4	Time-Update und <i>Kalman</i> -Filter-Algorithmus	136
6.3.5	Zustandsschätzung und Prognose	138
6.3.6	Parameterschätzung	140
6.4	ARV-Modelle	141
6.4.1	Zustandsraumform des ARV-Modells	141
7	Tipps und Tricks	145
7.1	Zentrale Momente einer Normalverteilung	145
7.2	Einfache Berechnung stationärer Momente	145
7.3	Vereinfachte Kovarianzformel	147
7.4	<i>Hodrick-Prescott</i> -Filter	147
8	Lösungen	149

Zeitreihen im Allgemeinen

Eine Zeitreihe ist zunächst, wie ihr Name schon nahelegt, eine Reihe oder Sequenz zeitlich geordneter Daten. Traditionell wird die Zeitreihe durch einen kleinen Buchstaben mit Index, bspw. $(y_t)_{t=1,\dots,T}$, gekennzeichnet. Dabei handelt es sich um die Menge

$$(y_t)_{t=1,\dots,T} = \{y_1, y_2, \dots, y_{T-1}, y_T\} \quad (1.1)$$

zeitlich geordneter Werte. Der Index t muss dabei nicht zwangsläufig bei eins beginnen oder bei T enden, häufig beginnen Zeitreihen auch mit einem (bekannten) Anfangswert y_0 woraus folgt, dass der Index der Zeitreihe (1.1) auch bei $t = 0$ beginnen könnte.

Diese Definition einer Zeitreihe hat zunächst noch nichts mit Zufallsvariablen oder Ähnlichem zu tun, sie besagt nur, dass die erfassten Werte in einem zeitlich geordneten Verhältnis zueinander stehen. Ziel der Zeitreihenanalyse ist es eine Systematik in dieser Zeitordnung zu identifizieren und mathematisch zu erfassen, um die Zeitreihe damit weiteren mathematischen Diagnoseverfahren zuführen zu können oder um mit Hilfe des gefundenen Zusammenhangs Vorhersagen über den weiteren Verlauf der Zeitreihe machen zu können.

Beispiel 1.1:

Die Zeitreihe

$$(y_t)_{t=0,\dots,5} = \left\{ 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32} \right\}$$

hat ganz offensichtlich eine Struktur. Der Wert zum Zeitpunkt t ist immer die Hälfte des Wertes des vorangegangenen Zeitpunktes $t - 1$. Daher kann die Zeitreihe durch die Vorschrift

$$y_t = \frac{1}{2} y_{t-1}$$

repräsentiert werden.

Mit Hilfe der erkannten Struktur lassen sich nun auch Vorhersagen für zukünftige Werte der Zeitreihe machen. Dies kann entweder rekursiv geschehen, indem man ausgehend von einem bekannten Wert solange den folgenden Wert berechnet, bis die Prognose für den gewünschten Zeitpunkt verfügbar ist, oder durch Ermitteln einer expliziten Funktion. Man verifiziert leicht den folgenden Zusammenhang für die Zeitreihe in Beispiel 1.1

$$\begin{aligned} y_1 &= \frac{1}{2} y_0 = \frac{1}{2} \\ y_2 &= \frac{1}{2} y_1 = \left(\frac{1}{2}\right)^2 y_0 = \frac{1}{4} \\ y_3 &= \frac{1}{2} y_2 = \left(\frac{1}{2}\right)^3 y_0 = \frac{1}{8}, \end{aligned} \tag{1.2}$$

woraus die explizite Form

$$y_t = \left(\frac{1}{2}\right)^t y_0 = \left(\frac{1}{2}\right)^t \tag{1.3}$$

hervorgeht. Das zweite Gleichheitszeichen in (1.3) gilt weil die Anfangsbedingung $y_0 = 1$ ist. Mit Hilfe der expliziten Form (1.3) lässt sich auch ein Grenzwert für $t \rightarrow \infty$ berechnen

$$\lim_{t \rightarrow \infty} y_t = \lim_{t \rightarrow \infty} \left(\frac{1}{2}\right)^t = 0. \tag{1.4}$$

Die Existenz eines solchen Grenzwertes ist eng mit dem Begriff der Stationarität von Zeitreihen verbunden. Vereinfacht ausgedrückt bedeutet

Stationarität, dass die Eigenschaften einer Zeitreihe auf lange Sicht (also für große t) nicht mehr spürbar vom Index t abhängen.

In der Realität sind die Zusammenhänge natürlich nicht so einfach wie in Beispiel 1.1 dargestellt. Verschiedene Komplikationen können beim Betrachten von Zeitreihen auftreten. Einige sollen an dieser Stelle explizit aufgeführt werden:

- Die Werte einer Zeitreihe lassen sich für gewöhnlich keiner so klaren Systematik zuführen wie es im Beispiel 1.1 der Fall war. Meistens sind sie „verrauscht“. Daher wird es in der Regel notwendig sein, eine weitere Variable in die Funktionsvorschrift aufzunehmen, die diese Abweichungen kompensiert, bspw.

$$y_t = \phi y_{t-1} + \epsilon_t.$$

Der Parameter ϕ war in Beispiel 1.1 mit $\phi = \frac{1}{2}$ gewählt worden, er kann jedoch auch andere Werte annehmen. Die zweite Variable ϵ_t wird in der Regel als zufallsabhängig angenommen, mit einer an dieser Stelle nicht näher spezifizierten Verteilung. Daher auch der Begriff „verrauscht“, weil man annimmt, dass ϵ_t den irregulären und nicht vorhersagbaren Anteil des Zusammenhangs darstellt.

- Die Präsenz eines Zufallseinflusses in der Funktionsgleichung hat weitreichende Folgen. Die Werte y_t der Zeitreihe sind damit selbst Zufallsvariablen, die mit Hilfe angemessener statistischer Instrumente wie Verteilungs- bzw. Dichtefunktionen, Erwartungswerte etc. analysiert werden müssen.
- Es ist weiterhin denkbar, dass der Wert y_t nicht ausschließlich systematisch vom vorangegangenen Wert y_{t-1} abhängt, sondern vielleicht ebenfalls noch von y_{t-2} . Solche Prozesse werden allgemein als autoregressiv bezeichnet (AR-Prozesse), wobei auch häufig die Form AR(p)-Prozess zur Charakterisierung verwendet wird. Der Parameter p gibt dabei die (größte) zeitliche Verzögerung an, mit der vergangene Werte der Zeitreihe noch auf den gegenwärtigen Wert wirken. Beispiel 1.1 stellt also einen AR(1)-Prozess dar, da der Gegenwartswert y_t lediglich von y_{t-1} abhängt. Wäre zusätzlich noch eine Abhängigkeit von y_{t-2} zu berücksichtigen gewesen, würde es sich um einen AR(2)-Prozess handeln.
- Die Zeitreihe könnte für $t \rightarrow \infty$ divergieren. Eine solche Nicht-Stationarität hat schwerwiegendere Folgen als man zunächst vermuten

könnte. Da es sich im Allgemeinen bei den Werten y_t der Zeitreihe um Zufallsvariablen handelt würde hier der Erwartungswert für $t \rightarrow \infty$ divergieren. In der Statistik beruhen aber außerordentlich viele wichtige Schätzfunktionen und Teststatistiken auf der Bedingung $E|X| < \infty$, also darauf, dass der Erwartungswert einer Zufallsvariable X eben nicht divergiert.

Diese Probleme zu meistern oder Bedingungen anzugeben, unter denen solche Komplikationen behandelbar sind, ist die größte Herausforderung in der Zeitreihenanalyse.

2

AR-Prozesse und Elementares

Wir haben bereits in Beispiel 1.1 eine einfache Manifestation des AR(1)-Prozesses kennengelernt. Wir werden nun einige Eigenschaften dieser Prozessklasse zusammentragen und dabei auf allgemeinere Eigenschaften schließen. Das notwendige mathematische Instrumentarium wird dabei an den entsprechenden Stellen in Form eines kurzen Repetitoriums (Kochkurs) eingeschoben.

2.1 Der AR(1)-Prozess

In seiner allgemeinen Form wird der AR(1)-Prozess als

$$y_t = \phi y_{t-1} + \epsilon_t \quad (2.1)$$

notiert. Der Parameter ϕ stellt dabei den AR-Parameter dar, der angibt, wie stark der vergangene Wert y_{t-1} bei der Berechnung des gegenwärtigen Wertes y_t berücksichtigt wird. ϕ besitzt deshalb den Charakter eines Gewichts. Die Zufallsvariable ϵ_t sei normalverteilt mit Erwartungswert 0 und Varianz σ^2 .

Einschub: Normalverteilung

Eine stetige Zufallsvariable X ist normalverteilt mit $E[X] = \mu$ und $\text{Var}[X] = \sigma^2$, wenn ihre Wahrscheinlichkeitsdichte die Form

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

besitzt. Die Normalverteilung ist in der Statistik so wichtig, weil eine große Summe von unabhängigen Zufallsvariablen durch eine entsprechende Normalverteilung approximiert werden kann (zentraler Grenzwertsatz).

Der Fehlerterm ϵ_t in (2.1) wird mit Erwartungswert 0 spezifiziert, da jeder

andere Wert ja einen systematischen Einfluss auf y_t hätte, der Zufallsfehler aber gerade die nichtsystematischen Einflüsse repräsentiert. Die Normalverteilungsannahme ist deswegen plausibel, weil man davon ausgeht, dass unter ϵ_t eine nahezu unbegrenzte Anzahl von voneinander unabhängigen Einflüssen subsummiert wird, die in Konsequenz des zentralen Grenzwertsatzes durch eine Normalverteilung approximiert werden können. Nichtsdestotrotz können auch andere Verteilungen spezifiziert werden!

2.1.1 Zentrierter AR(1)-Prozess

Es ist in wirtschaftswissenschaftlichen Anwendungen gelegentlich auch erforderlich die Abweichungen von bestimmten Werten über die Zeit zu untersuchen. Beispielsweise die Abweichung der Arbeitslosigkeit von der natürlichen Arbeitslosenquote oder ähnliches. In diesem Fall kann der AR(1)-Prozess (2.1) mit der Substitution $y_t = x_t - \mu$ betrachtet werden

$$x_t - \mu = \phi(x_{t-1} - \mu) + \epsilon_t. \quad (2.2)$$

μ repräsentiert dabei den interessierenden (langfristigen) Gleichgewichtszustand. Nach Ausmultiplizieren von (2.2) und Umstellen erhält man

$$x_t = (1 - \phi)\mu + \phi x_{t-1} + \epsilon_t, \quad (2.3)$$

was insofern erstaunlich ist, als dass x_t offenbar als gewichtetes Mittel aus dem Gleichgewichtswert und dem vergangenen Wert gebildet wird, falls $0 < \phi < 1$ gilt und die Zufallseinflüsse einmal vernachlässigt werden.

Die Verwendung des Buchstabens μ für den langfristigen Gleichgewichtswert an dieser Stelle ist natürlich suggestiv, da μ für gewöhnlich für den Erwartungswert verwendet wird. Wir werden im nächsten Abschnitt sehen, was es damit auf sich hat.

2.1.2 Stationarität des AR(1)-Prozesses

Wir leiten nun die Bedingung ab, unter der der AR(1)-Prozess stationär ist. Wir haben in (1.3) eine explizite Formel für y_t hergeleitet, die jedoch nicht ohne weiteres auf den AR(1)-Prozess (2.1) zu übertragen ist, da hier ein Zufallseinfluss involviert ist. Es könnten aber ganz ähnliche Formeln zum Tragen kommen, die sich nicht auf y_t selbst beziehen, sondern auf die Momente der Verteilung von y_t .

Einschub: Momente

Mit Momenten werden wichtige Kenngrößen einer Wahrscheinlichkeitsverteilung bezeichnet. Das erste Moment ist der Erwartungswert, der sich für eine stetige Zufallsvariable X aus

$$\mu = \int xp(x)dx$$

berechnet, wobei $p(x)$ die Wahrscheinlichkeitsdichtefunktion darstellt. Höhere Momente werden in Form von Abweichungen vom Erwartungswert angegeben (zentrale Momente)

$$m_k = \int (x - \mu)^k p(x) dx.$$

Das zweite zentrale Moment entspricht damit gerade der Varianz, $m_2 = \sigma^2$. Wir werden später sehen, dass auch höhere Momente bei der Analyse von Zeitreihen eine Rolle spielen können.

Wir betrachten den AR(1)-Prozess $y_t = \phi y_{t-1} + \epsilon_t$ bzw. den zentrierten Prozess $y_t = x_t - \mu$. Der Zufallsfehler sei normalverteilt mit $\epsilon_t \sim N(0, \sigma^2)$ und der Anfangswert y_0 sei bekannt. Für den Erwartungswert des (zentrierten) AR(1)-Prozesses erhalten wir

$$E[y_t] = \phi E[y_{t-1}] \quad \text{bzw.} \quad E[x_t] - \mu = \phi(E[x_{t-1}] - \mu). \quad (2.4)$$

Der Erwartungswert des Fehlerterms ϵ_t ist gerade null, daher taucht er nicht mehr in der Erwartungswertgleichung (2.4) auf. Der Gleichgewichtswert μ ist eine deterministische Größe, daher gilt $E[\mu] = \mu$. Es kann nun dasselbe Rekursionsprinzip wie in (1.2) auf Seite 2 angewendet werden, um eine explizite Darstellung für den Erwartungswert von y_t abzuleiten. Man erhält so

$$E[y_t] = \phi^t E[y_0] = \phi^t y_0 \quad (2.5)$$

bzw. einen analogen Ausdruck für den zentrierten Prozess $y_t = x_t - \mu$. Berechnet man weiterhin den Grenzwert für großes t ,

$$\lim_{t \rightarrow \infty} E[y_t] = \lim_{t \rightarrow \infty} \phi^t y_0, \quad (2.6)$$

erkennt man sofort, dass der Erwartungswert nur dann gegen null strebt, wenn die Bedingung $|\phi| < 1$ erfüllt ist. Für $|\phi| > 1$ divergiert der Prozess, woraus hervorgeht, dass kein stationärer Erwartungswert existiert. Der Fall

$|\phi| = 1$ ist ein Spezialfall, da der stationäre Erwartungswert hier der Anfangsbedingung entspricht. Diesen Fall bezeichnet man als Einheitswurzel. Dieser Begriff wird später klarer, wenn wir das AR(2)-Modell betrachten. Im Fall einer Einheitswurzel liegt dennoch kein stationärer Prozess vor, da die Varianz divergiert, wie wir im Anschluss zeigen werden. Bis hierher haben wir den stationären Erwartungswert ermittelt, der unter der Bedingung $|\phi| < 1$ existiert

$$E[y_{st.}] = 0 \quad \text{bzw.} \quad E[x_{st.}] = \mu \quad \text{für } |\phi| < 1. \quad (2.7)$$

Für die Varianz des AR(1)-Prozesses gilt die rekursive Formel

$$\text{Var}[y_t] = \text{Var}[\phi y_{t-1} + \epsilon_t] = \phi^2 \text{Var}[y_{t-1}] + \sigma^2. \quad (2.8)$$

Dieses Ergebnis mag nicht unmittelbar einleuchten, daher hier ein kurzer Einschub zu den Rechenregeln von Varianzen.

Einschub: Varianzen

Seien X und Y zwei unabhängige Zufallsvariablen und c eine beliebige Konstante, dann gilt

1. $\text{Var}[c] = 0$
2. $\text{Var}[cX] = c^2 \text{Var}[X]$
3. $\text{Var}[X + Y] = \text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y]$.

Voraussetzung für die Gültigkeit der letzten Eigenschaft ist die Unabhängigkeit der Zufallsvariablen.

ϵ_t ist von ϕy_{t-1} unabhängig, da der zum Zeitpunkt $t - 1$ manifestierte Wert der Zeitreihe keinen Einfluss auf die Realisierung der Zufallsvariable ϵ_t zum Zeitpunkt t hat. Daher zerfällt die gemeinsame Varianz in (2.8) in die Summe der Einzelvarianzen. ϕ ist eine Konstante und rückt deshalb quadratisch vor die Varianz von y_{t-1} . Die Varianz von ϵ_t ist bereits bekannt und für jeden Zeitpunkt gleich, $\text{Var}[\epsilon_t] = \sigma^2$.

Der nächste Schritt ist nun wieder eine explizite Vorschrift für $\text{Var}[y_t]$ abzuleiten und dann den Grenzwert für $t \rightarrow \infty$ zu betrachten. Diese Iteration gestaltet sich geringfügig sperriger als die des Erwartungswertes, daher kommt an dieser Stelle zunächst eine kurze Auffrischung zum Summenoperator und der geometrischen Reihe.

Einschub: Summenzeichen und geometrische Reihe

Das Summenzeichen ist eine bequeme Art lange Additionsketten zu schreiben

$$\sum_{k=1}^n x_k = x_1 + x_2 + x_3 + \dots + x_{n-1} + x_n.$$

Der Index unter und über dem Operator gibt dabei die „Laufgrenzen“ der Summation an. Im Weiteren werden wir ein Polynom betrachten, das durch folgende Summe gebildet wird

$$\sum_{k=0}^n x^k = 1 + x + x^2 + \dots + x^{n-1} + x^n.$$

Diese Summe wird als geometrische Reihe bezeichnet und ist ein wichtiges Instrument in der Zeitreihenanalyse. Die geometrische Reihe besitzt folgende nützliche Eigenschaft

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad \text{für } |x| < 1,$$

die wir uns im Folgenden zu Nutze machen. Man beachte, dass die Summe unendlich viele Glieder besitzt.

Wir beginnen nun (2.8) vorwärts zu iterieren

$$\begin{aligned} \text{Var}[y_1] &= \phi^2 \text{Var}[y_0] + \sigma^2 = \sigma^2 \\ \text{Var}[y_2] &= \phi^2 \text{Var}[y_1] + \sigma^2 = \phi^2 \sigma^2 + \sigma^2 = \sigma^2(1 + \phi^2) \\ \text{Var}[y_3] &= \phi^2 \text{Var}[y_2] + \sigma^2 = \phi^4 \sigma^2 + \phi^2 \sigma^2 + \sigma^2 = \sigma^2(1 + \phi^2 + \phi^4). \end{aligned} \tag{2.9}$$

Man erkennt schnell, dass sich bei entsprechendem Umstellen der Terme in (2.9) eine geometrische Reihe bezüglich ϕ^2 bildet. Daher lässt sich die Varianz für ein beliebiges Glied der Zeitreihe in der kompakten Form

$$\text{Var}[y_t] = \sigma^2 \sum_{k=0}^{t-1} (\phi^2)^k \tag{2.10}$$

angeben. Hier zeigt sich wieder die Effizienz des Summenzeichens bei der Notation langer Additionsketten. Wer mit dem Handling noch nicht vertraut ist kann im Rahmen der Aufgaben am Ende dieses Abschnittes ein wenig Erfahrung sammeln.

Die Frage nach der stationären Varianz steht noch im Raum. Dazu bilden wir wieder den Grenzwert für $t \rightarrow \infty$ und verwenden die entsprechende Eigenschaft der geometrischen Reihe

$$\text{Var}[y_{st.}] = \lim_{t \rightarrow \infty} \sigma^2 \sum_{k=0}^{t-1} (\phi^2)^k = \frac{\sigma^2}{1 - \phi^2} \quad \text{für } \phi^2 < 1. \quad (2.11)$$

Die Bedingung $\phi^2 < 1$ für die Konvergenz der geometrischen Reihe entspricht genau der Bedingung $|\phi| < 1$, die wir bereits im Rahmen der stationären Erwartungswertgleichung (2.7) bestimmen konnten. Es ist jetzt ebenfalls klar, dass im Fall der Einheitswurzel mit $\phi = 1$ die Varianz divergiert. Wir erhalten

$$\lim_{t \rightarrow \infty} \text{Var}[y_t] = \lim_{t \rightarrow \infty} \sigma^2 \sum_{k=0}^{t-1} 1^k = \lim_{t \rightarrow \infty} \sigma^2 \sum_{k=1}^t 1 = \lim_{t \rightarrow \infty} \sigma^2 t = \infty. \quad (2.12)$$

Wir halten zusammenfassend fest, dass die beiden stationären Momente, (2.7) und (2.11) nicht mehr vom Index t abhängen und konstant sind. Damit gelangen wir unmittelbar zur Definition der schwachen Stationarität:

Unter schwacher Stationarität versteht man die Zeitinvarianz der ersten beiden Momente eines Prozesses.

Im Fall der Normalverteilung ist diese Definition besonders vorteilhaft, da eine beliebige Normalverteilung durch ihre ersten beiden Momente vollständig spezifiziert ist. Für andere Verteilungen wie beispielsweise die *Student-t*-Verteilung oder die χ^2 -Verteilung gilt das nicht. Im Fall einer solchen Verteilung käme eine Definition der Stationarität im strengen Sinn zum Tragen:

Unter Stationarität (im strengen Sinn) versteht man die Zeitinvarianz der gesamten Verteilung eines Prozesses.

Würde man beispielsweise den zentrierten AR(1)-Prozess (2.3) mit $|\phi| < 1$ und der zufälligen Anfangsbedingung $x_0 \sim N(\mu, \frac{\sigma^2}{1-\phi^2})$ starten, wäre der resultierende Prozess augenblicklich stationär und zwar im strengen Sinn, da die Fehlerterme normalverteilt sind und die ersten beiden Momente für jedes weitere Glied x_t mit $t = 1, \dots, \infty$ unverändert bleiben, wie durch Einsetzen in (2.4) und (2.8) leicht zu zeigen ist.

Aufgabe 2.1

Berechnen Sie die folgenden Summen:

$$\begin{array}{llll} \text{a)} \sum_{k=1}^5 k, & \text{b)} \sum_{k=1}^3 \frac{1}{2k}, & \text{c)} \sum_{k=0}^4 2^k, & \text{d)} \sum_{i=1}^2 \sum_{j=1}^3 (i+j)^2 \\ \text{e)} \sum_{i=1}^3 \sum_{j=0}^i i \cdot j, & \text{f)} \sum_{k=0}^{\infty} \left(\frac{3}{4}\right)^k, & \text{g)} \sum_{k=1}^{\infty} \frac{1}{2^k} \end{array}$$

Aufgabe 2.2

Schreiben Sie die folgenden Terme mit dem Summenoperator:

$$\begin{array}{l} \text{a)} 2 + 4 + 6 + \dots + 16 \\ \text{b)} -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \frac{1}{6} \\ \text{c)} 0 + \frac{1}{2} + \frac{2}{3} + \frac{3}{4} + \dots + \frac{45}{46} \\ \text{d)} 1 + 4 + 27 + 256 \\ \text{e)} \frac{2}{1 - \frac{1}{2}} \end{array}$$

Anmerkung: Es gibt unter Umständen mehrere Möglichkeiten die Summen zu notieren. In der Aufgabenlösung wird lediglich der kompakteste Ansatz wiedergegeben.

2.1.3 Wissenswertes über den AR(1)-Prozess

Der AR(1)-Prozess (2.1) nimmt eine gewisse Sonderstellung bei der Betrachtung dynamischer Systeme ein. Er ist ein kausaler Prozess, was bedeutet, dass die Wirkung (linke Seite der Gleichung) nach der Ursache (rechte Seite der Gleichung) eintritt. Es gibt auch simultane, respektive akausale Modelle, wo beispielsweise auf beiden Seiten der Gleichung y_t steht oder der Gegenwartswert von zukünftigen Werten abhängt. Wir werden mit der *Wold-Zerlegung* eine noch allgemeinere Bedingung für die Kausalität eines Prozesses kennenlernen.

Im Hinblick auf die Prognose im AR(1)-Modell lässt sich sagen, dass der nächste vorherzusagende Wert y_{t+1} lediglich vom Gegenwartswert y_t abhängt. Die Kenntnis der Vergangenheit y_{t-1}, \dots, y_0 ist zur Erstellung

der Prognose nicht erforderlich, sie erklärt nur den Weg, den der Prozess bis zur Gegenwart genommen hat. Da wir den Gegenwartswert y_t aber kennen, liefert uns die Vergangenheit keine zusätzlichen Erkenntnisse mehr bezüglich der Zukunft. Prozesse, die ihre gesamte Vergangenheit sozusagen im Gegenwartswert speichern, bezeichnet man als *Markov*-Prozesse. *Markov*-Prozesse sind in vielen Wissenschaftsgebieten von Interesse, da mit ihnen eine außerordentlich große Zahl natürlich auftretender Phänomene abgebildet werden kann.

Weiterhin können AR-Prozesse höherer Ordnung immer in ein System von gekoppelten AR(1)-Prozessen überführt werden. Diese gekoppelten Prozesse werden meistens in Vektorform zusammengefasst und man spricht dann von einem VAR-Prozess (Vector Auto Regressive). Wir werden später sehen, wie sich das AR(2)-Modell in ein VAR-Modell der Dimension $d = 2$ umschreiben lässt.

Der Spezialfall der Einheitswurzel, $|\phi| = 1$ wird als Random Walk oder als einfache Irrfahrt bezeichnet. Gelegentlich begegnet man auch der etwas unterhaltsameren Metapher des Weges eines Betrunkenen. Die Struktur $y_t = y_{t-1} + \epsilon_t$ zeigt, dass die nächste „Position“ sich aus der letzten, zuzüglich eines zufälligen „Schrittes“ ergibt. Wobei wie bei einem Betrunkenen nicht vorherzusagen ist, in welche Richtung der nächste Schritt wohl erfolgen wird. Ferner wird hier natürlich vernachlässigt, dass der Betrunkenen stolpern und hinfallen könnte, was möglicherweise zur Stationarität des Betrunkenen führt aber nicht des Prozesses. Auch der Random Walk ist ein wichtiges dynamisches Modell in vielen Wissenschaftsdisziplinen.

2.2 Autokovarianz und Autokorrelation

Wir sind nun soweit genauer zu verstehen, um was für ein Objekt es sich bei einer Zeitreihe eigentlich handelt:

Eine Zeitreihe ist eine Manifestation eines darunterliegenden (stochastischen) Prozesses.

Das bedeutet, ein in der Regel unbekannter Prozess erzeugt eine Reihe von Realisierungen in der Zeit, die wir beobachten können. Diese Beobachtungsreihe bezeichnen wir als Zeitreihe. Da wir nicht wissen, welcher Prozess diese Zeitreihe erzeugt hat, bspw. ein AR(1)-Prozess oder ein AR(2)-Prozess oder eine völlig andere Prozessklasse, sind wir darauf angewiesen die Hinweise, die in der Zeitreihe verborgen sind, sehr genau auszuwerten.

2.2. AUTOKOVARIANZ UND AUTOKORRELATION

Die Zeitreihe ähnelt in vielerlei Hinsicht einer beliebigen Zufallsstichprobe, mit einem entscheidenden Unterschied: Die Realisierungen sind nicht unabhängig voneinander. Die Abhängigkeit wird sofort klar, wenn wir für den Moment einmal annehmen, der Prozess, der die Zeitreihe erzeugt hat, sei ein AR(1)-Prozess

$$y_t = \phi y_{t-1} + \epsilon_t. \quad (2.13)$$

Der Wert y_t besteht aus der Realisation des Zufallsfehlers zuzüglich des mit dem AR-Parameter ϕ gewichteten vorangegangenen Wert y_{t-1} . Für die Wahl $\phi = 0$ würde die Zeitreihe in der Tat einer unabhängigen Zufallsstichprobe entsprechen, in der Regel gilt aber $\phi \neq 0$.

Die Abhängigkeit der Zeitreihenwerte wird mit Hilfe ihrer Kovarianz bzw. ihrer Korrelation charakterisiert. Da es sich gewissermaßen um eine Korrelation des Merkmals mit sich selbst in der Zeit handelt, verwendet man die Begriffe Autokovarianz bzw. Autokorrelation.

Einschub: Kovarianz und Korrelation

Seien X und Y zwei Zufallsvariablen, die auf nicht näher spezifizierte Weise voneinander abhängen, dann ergibt sich die Kovarianz zwischen ihnen aus

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

Für unabhängige Zufallsvariablen ist die Kovarianz null. Die Varianz einer Zufallsvariablen ist ein Spezialfall der Kovarianz

$$\text{Cov}[X, X] = E[(X - E[X])^2] = \text{Var}[X].$$

Bei der Korrelation ρ handelt es sich um eine Normierung der Kovarianz in den Bereich $[-1, 1]$. Diese Normierung wird durch folgende Operation bewirkt

$$\rho[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}}.$$

Die quasi Korrelation mit sich selbst bringt einen angenehmen Effekt mit sich, nämlich dass der Grad der Korrelation im stationären Zustand lediglich vom zeitlichen Abstand der Werte abhängt, jedoch nicht von der Zeit selbst. Das heißt nichts anderes, als dass beispielsweise die Autokorrelation $\rho[y_t, y_{t-1}]$ gleich der Autokorrelation $\rho[y_{t-1}, y_{t-2}]$ ist usw. Infolge dessen werden Auto-

korrelationen und Autokovarianzen lediglich mit einem Verschiebungsindex geschrieben, der die Lücke (Lag) zwischen den beiden Werten angibt

$$\gamma_k = \text{Cov}[y_t, y_{t-k}]. \quad (2.14)$$

Wir können nun die einmal getroffene Konvention (2.14) ausnutzen um die stationäre Varianz des AR(1)-Prozesses als Autokovarianz mit Lag 0 zu schreiben

$$\gamma_0 = \text{Var}[y_t] = \frac{\sigma^2}{1 - \phi^2}. \quad (2.15)$$

Auch die stationäre Varianz bleibt für alle t gleich, denn das war ja gerade eines der Kriterien für Stationarität. Da die Varianz aber ein Spezialfall der Kovarianz, nämlich mit Verschiebung (Lag) 0 ist, lässt sich intuitiv verstehen, warum die Autokovarianz im stationären Zustand nur vom Lag abhängen darf. In der Tat wird die schwache Stationarität auch als Kovarianzstationarität bezeichnet und es wird im Allgemeinen definiert:

Ein Prozess ist schwach stationär (kovarianzstationär), wenn seine ersten beiden Momente endlich und konstant (invariant gegenüber Zeitverschiebungen) sind und die Autokovarianz lediglich von der Verschiebung aber nicht von der Zeit abhängt.

Dies ist eine Erweiterung gegenüber unserer vorangegangenen Definition schwacher Stationarität. Wir kannten jedoch zu diesem Zeitpunkt noch nicht die Rolle der Autokovarianz. Ferner ist es nicht nötig bezüglich der Autokovarianz den zentrierten Prozess $y_t = x_t - \mu$ gesondert zu betrachten, da

$$\text{Cov}[x_t, x_{t-k}] = E[(x_t - \mu)(x_{t-k} - \mu)] = E[y_t y_{t-k}] = \text{Cov}[y_t, y_{t-k}] \quad (2.16)$$

gilt.

2.2.1 Autokovarianz

Wir werden nun die Autokovarianz für den AR(1)-Prozess ausrechnen. Wir beginnen naheliegenderweise mit dem Lag 1. Man sollte dabei im Hinterkopf behalten, dass der AR(1)-Prozess (2.13) den stationären Erwartungswert $E[y_t] = 0$ und die stationäre Varianz $\text{Var}[y_t] = \gamma_0$ für alle t besitzt.

$$\begin{aligned} \gamma_1 &= \text{Cov}[y_t, y_{t-1}] = E[y_t y_{t-1}] \\ &= E[(\phi y_{t-1} + \epsilon_t) y_{t-1}] = E[\phi y_{t-1}^2] + E[\epsilon_t y_{t-1}] \\ &= \phi \text{Var}[y_{t-1}] + \text{Cov}[\epsilon_t, y_{t-1}] \\ &= \phi \gamma_0 \end{aligned} \quad (2.17)$$

2.2. AUTOKOVARIANZ UND AUTOKORRELATION

Es wurde in der Herleitung (2.17) wieder das frühere Ergebnis benutzt, dass der Zufallsfehler ϵ_t nicht vom Vergangenheitswert y_{t-1} abhängt. Folglich ist deren Kovarianz null.

Als nächstes soll nun die Autokovarianz mit Lag 2 berechnet werden. Wir modifizieren dafür den Ansatz (2.17) zu

$$\begin{aligned}\gamma_2 &= \text{Cov}[y_t, y_{t-2}] = E[y_t y_{t-2}] \\ &= E[(\phi y_{t-1} + \epsilon_t) y_{t-2}] = E[\phi y_{t-1} y_{t-2}] + E[\epsilon_t y_{t-2}] \\ &= \phi \text{Cov}[y_{t-1}, y_{t-2}] + \text{Cov}[\epsilon_t, y_{t-2}] \\ &= \phi \gamma_1 = \phi^2 \gamma_0.\end{aligned}\tag{2.18}$$

Wir haben hier wieder ausgenutzt, dass der Fehlerterm ϵ_t von vergangenen Zeitreihenwerten unabhängig ist.

Aufgabe 2.3

Berechnen Sie die Autokovarianz mit Lag 3.

Wir erhalten offensichtlich eine AR-Repräsentation der Autokovarianz mit beliebigem Lag k

$$\gamma_k = \phi \gamma_{k-1}.\tag{2.19}$$

Diese Repräsentation (2.19) wird *Yule-Walker-Gleichung* genannt. Sie stellt die Verbindung zwischen Autokovarianz bzw. Autokorrelation (wie wir noch sehen werden) und dem AR-Prozess her. Die *Yule-Walker-Gleichung* lässt sich in zwei Schritten aus der Funktionsvorschrift des AR-Prozesses gewinnen. Im ersten Schritt werden alle Terme mit verzögerten Werten mit dem entsprechenden Lag multipliziert und im zweiten Schritt wird der Erwartungswert gebildet. (2.19) lässt sich somit einfach aus (2.13) herleiten

$$\begin{aligned}y_t &= \phi y_{t-1} + \epsilon_t \quad | \cdot y_{t-k} \\ \Rightarrow y_t y_{t-k} &= \phi y_{t-1} y_{t-k} + \epsilon_t y_{t-k} \quad | E[\dots] \\ \Rightarrow E[y_t y_{t-k}] &= \phi E[y_{t-1} y_{t-k}] + E[\epsilon_t y_{t-k}] \\ \Rightarrow \gamma_k &= \phi \gamma_{k-1}.\end{aligned}\tag{2.20}$$

2.2.2 Autokorrelation

Die Autokorrelation ist das eigentlich wichtigere Maß der Abhängigkeit der Zeitreihenwerte, da sie auf das Intervall $[-1, 1]$ normiert ist. Die Autokorrelation berechnet sich recht einfach für beliebiges Lag k

$$\rho_k = \frac{\text{Cov}[y_t, y_{t-k}]}{\sqrt{\text{Var}[y_t]} \sqrt{\text{Var}[y_{t-k}]}} = \frac{\gamma_k}{\sqrt{\gamma_0} \sqrt{\gamma_0}} = \frac{\gamma_k}{\gamma_0}.\tag{2.21}$$

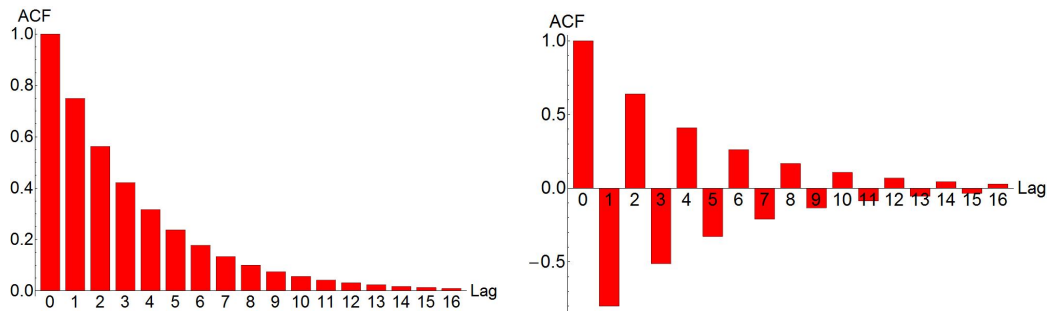


Abbildung 2.1: Autokorrelationsfunktion mit $\phi = 0.75$ (links) und $\phi = -0.8$ (rechts)

Formel (2.21) birgt mehr Information, als man auf Anhieb vermuten könnte. Setzen wir beispielsweise $k = 0$, so ergibt sich die einfache Korrelation $\rho_0 = 1$. Nutzen wir die *Yule-Walker*-Gleichung (2.19) resp. (2.20) für die Autokovarianz aus, erhalten wir

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi \gamma_{k-1}}{\gamma_0} = \phi \rho_{k-1}. \quad (2.22)$$

Offensichtlich gilt die *Yule-Walker*-Gleichung auch für die Autokorrelation. Vielmehrnoch kann die Autokorrelation mit Lag k sehr bequem explizit geschrieben werden

$$\rho_k = \phi^k \rho_0 = \phi^k, \quad (2.23)$$

da ja die Korrelation mit Lag 0 genau eins ist. Man sieht in (2.23) sehr schön, dass die Autokorrelation exponentiell abfällt, wenn die Verschiebung k größer wird. Alle bis hierher gemachten Aussagen gelten natürlich nur für stationäre Prozesse, was im Fall des AR(1)-Prozesses bedeutet $|\phi| < 1$.

Abbildung 2.1 zeigt die Autokorrelationsfunktion (ACF) für verschiedene Werte des AR-Parameters. Für $\phi = 0.75$ ergibt sich ein schöner exponentieller Abfall der Autokorrelation. Bei negativem AR-Parameter, $\phi = -0.8$, klingt die Autokorrelation ebenfalls exponentiell ab, jedoch mit alternierenden Werten. Dennoch kann ebenfalls beobachtet werden, dass die ACF für große Lags k gegen null strebt.

2.2.3 Partielle Autokorrelation

Die Herleitung der partiellen Autokorrelation ist formal kompliziert¹, interessierte Leser seien deshalb an dieser Stelle auf Spezialliteratur verwiesen

¹Insbesondere wird dabei die Symmetrieeigenschaft der Autokorrelation $\rho_k = \rho_{-k}$ ausgenutzt.

2.2. AUTOKOVARIANZ UND AUTOKORRELATION

(siehe Hauptskript). Stattdessen soll zunächst die zugrundeliegende Idee erläutert werden.

Abbildung 2.1 zeigt das Autokorrelogramm für einen AR(1)-Prozess. Wir wissen inzwischen, dass y_t im AR(1)-Prozess eigentlich nur systematisch von y_{t-1} abhängt, die Autokorrelation klingt jedoch nur langsam ab, weil der Einfluss früherer Werte noch wie ein „Echo“ präsent ist. In AR-Modellen höherer Ordnung treten zusätzlich noch „Interferenzen“ mit Zwischenwerten hinzu. Wir wissen beispielsweise, dass ein AR(2)-Prozess die Form $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$ hat. Wie ermittelt man nun den isolierten (partiellen) Einfluss der Variable mit Lag 2, ohne die Überlagerung die von y_{t-1} erzeugt wird?

Dieses Problem wird vereinfacht gesprochen durch das Verschieben der Autokorrelationen gegeneinander gelöst. Dabei werden quasi Überlagerungen und Echos herausgerechnet. Diese Verschiebung wird durch das Lösen des linearen Gleichungssystems

$$\begin{pmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & \rho_0 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_0 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} \quad (2.24)$$

bewirkt. Zunächst gilt natürlich $\rho_0 = 1$, daher besteht die Hauptdiagonale der Matrix in (2.24) aus Einsen. ϕ_{k1} , ϕ_{k2} usw. sind die partiellen Autokorrelationen. Der erste Index k gibt hier die Gesamtzahl der berücksichtigten Zwischenkorrelationen an. Wir sehen beispielsweise in Abbildung 2.1, dass die Autokorrelation irgendwann für große Lags gegen null geht, daher ist anzunehmen, dass k in übersichtlichen Größenordnungen bleibt. Der zweite Index gibt dann die Partielle Autokorrelation mit der entsprechenden Verschiebung, bereinigt um die Einflüsse von insgesamt k Lags. Zunächst aber eine kleine Auffrischung zur Matrixnotation!

Einschub: Matrizen und Vektoren

Matrizen und Vektoren sind eine bequeme Art lineare Gleichungssysteme zu schreiben. Beispielsweise das System

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2\end{aligned}$$

kann in Matrixform als

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

geschrieben werden. Matrizen und Vektoren werden oft mit fettgedruckten Buchstaben oder Symbolen abgekürzt, wodurch die obige Matrixgleichung in der kompakten Form $\mathbf{Ax} = \mathbf{b}$ notiert werden kann. Die Rechenregeln für das Produkt der Matrix \mathbf{A} mit dem Vektor \mathbf{x} gehen unmittelbar aus der obigen Darstellung hervor.

Der Vorteil der Matrixdarstellung ist, dass ein lineares Gleichungssystem der Form $\mathbf{Ax} = \mathbf{b}$ simultan durch eine Operation gelöst werden kann, die als Matrixinversion bezeichnet wird. Diese Operation wird zu einem späteren Zeitpunkt noch genauer beleuchtet. Man erhält dann eine Matrixgleichung der Form $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ wodurch die Lösungen des linearen Gleichungssystems sofort im Vektor \mathbf{x} abgelesen werden können.

Wir wollen nun die partiellen Autokorrelationen für den AR(1)-Prozess bis zum Lag $k = 2$ berechnen. Mit Hilfe des Gleichungssystems (2.24) erhalten wir

$$\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}. \quad (2.25)$$

Um an die partiellen Autokorrelation zu kommen muss nun die (2×2) -Matrix in (2.25) invertiert werden.

Einschub: Determinante und Inverse einer (2×2) -Matrix

Beim Teilen durch einen Skalar a muss sichergestellt werden, dass $a \neq 0$ gilt, da durch null nicht geteilt werden darf. Bei Matrizen gilt eine ähnliche Bedingung, hier darf die Determinante nicht null werden. Die Determinante einer (2×2) -Matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

wird durch die Differenz der Produkte der Diagonalelemente gegeben

$$\det \mathbf{A} = a_{11}a_{22} - a_{12}a_{21}.$$

Die Determinante ist wieder ein Skalar und für $\det \mathbf{A} \neq 0$ ergibt sich als Inverse Matrix zu \mathbf{A}

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Skalare wie $1/\det \mathbf{A}$ werden einfach mit jedem Element einer Matrix oder eines Vektors multipliziert oder können wie oben „vor die Matrix“ gezogen werden.

Als Determinante der Matrix in (2.25) erhalten wir $1 - \rho_1^2$ und damit die Bedingung $|\rho_1| < 1$ für die Existenz einer eindeutigen Lösung für das Gleichungssystem der partiellen Autokorrelationen. Für den AR(1)-Prozess entspricht das gerade der Stationaritätsbedingung, da ja $\rho_1 = \phi$ gilt. Wir erhalten also im stationären Fall

$$\begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \frac{1}{1 - \rho_1^2} \begin{pmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \frac{1}{1 - \rho_1^2} \begin{pmatrix} \rho_1 - \rho_1\rho_2 \\ -\rho_1^2 + \rho_2 \end{pmatrix}, \quad (2.26)$$

bzw. einzeln notiert

$$\phi_{21} = \frac{\rho_1 - \rho_1\rho_2}{1 - \rho_1^2} \quad \text{und} \quad \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}. \quad (2.27)$$

Setzt man nun die Autokorrelationsvorschrift des AR(1)-Prozesses $\rho_k = \phi^k$ (Gleichung (2.23) auf Seite 16) in (2.27) ein, so erhält man

$$\begin{aligned} \phi_{21} &= \frac{\phi - \phi^3}{1 - \phi^2} = \frac{\phi(1 - \phi^2)}{1 - \phi^2} = \phi = \rho_1 \\ \phi_{22} &= \frac{\phi^2 - \phi^2}{1 - \phi^2} = 0. \end{aligned} \quad (2.28)$$

Mit anderen Worten, die Partielle Autokorrelation (PACF) mit Lag 1 entspricht genau der herkömmlichen Autokorrelation (ACF) und die Partielle Autokorrelation mit Lag 2 verschwindet. Bei genauerem Hinsehen ist das genau das Ergebnis, dass sich mit der Idee der partiellen Autokorrelation deckt. Da beim AR(1)-Prozess keine Interferenzen zu erwarten sind (es wird ja lediglich ein verzögerter Wert berücksichtigt), entspricht die PACF gerade der ACF. Ferner werden die Echos für Lags $k > 1$ beseitigt. Wir haben das hier zwar nur für $k = 2$ gezeigt, es ist jedoch eine wichtige Eigenschaft der AR(p)-Prozesse, dass die Partielle Autokorrelation für Lags $k > p$ augenblicklich auf null abfällt.

Aufgabe 2.4

Lösen Sie, falls möglich, das lineare Gleichungssystem $\mathbf{Ax} = \mathbf{b}$ mit

$$\begin{array}{ll} \text{a) } \mathbf{A} = \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 3 \\ 6 \end{pmatrix} & \text{b) } \mathbf{A} = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -2 \\ 8 \end{pmatrix} \\ \text{c) } \mathbf{A} = \begin{pmatrix} 3 & 2 \\ 6 & 4 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} & \text{d) } \mathbf{A} = \begin{pmatrix} 2 & 3 \\ 6 & 4 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \\ \text{e) } \mathbf{A} = \begin{pmatrix} -1 & 12 \\ 0 & -4 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -3 \\ -8 \end{pmatrix} & \end{array}$$

2.3 Parameterschätzung

Wir haben bereits gelernt, dass die Zeitreihe lediglich eine Manifestation von Werten ist, die von einem (meist unbekannten) stochastischen Prozess generiert werden. Das wirft unmittelbar zwei wichtige Probleme auf: Erstens, woher weiß man, welcher Prozess die Datenreihe generiert hat und zweitens, wie sehen die Parameter dieses Prozesses aus. Die erste Fragestellung betrifft die Modellidentifikation, auf die wir später zurückkommen werden, die zweite betrifft die Parameterschätzung.

Eine der angenehmsten Eigenschaften der Stationarität ist, dass die Parameterschätzer mit denen aus einfachen Stichprobensituationen übereinstimmen. Wir bezeichnen im Folgenden die Schätzer mit einem Dach über dem Symbol, bspw. $\hat{\mu}$ ist ein Schätzer für den Erwartungswert. $(y_t)_{t=1,\dots,T}$ sei eine Zeitreihe, die von einem uns bekannten stationären Prozess erzeugt wurde.

2.3. PARAMETERSCHÄTZUNG

Dann ergibt sich der Mittelwert- und der Varianzschätzer wie gewohnt durch

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t \quad (2.29)$$

$$\hat{\gamma}_0 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2. \quad (2.30)$$

Die Form von (2.29) und (2.30) ist vertraut, dennoch gibt es kleine Unterschiede zu gewöhnlichen Querschnittsstichproben. Zunächst ist die Stichprobengröße nicht N sondern T . Weiterhin steht im Nenner von (2.30) nicht $T - 1$, wie wir es von erwartungstreuen Varianzschätzern kennen. Der Grund dafür ist, dass Zeitreihen meist sehr lang sind (im Gegensatz zu einfachen Zufallsstichproben) und daher kaum ein Unterschied zwischen T^{-1} und $(T - 1)^{-1}$ besteht. Weiterhin wird die Varianz des stochastischen Prozesses auch nicht mit σ^2 bezeichnet, sondern mit γ_0 . Diese Unterscheidung ist vital, da wir ja die Varianz des Zufallsfehlers ϵ_t bereits mit σ^2 bezeichnet haben.

Die Schätzer für Autokovarianz und Autokorrelation ergeben sich auf ähnlich natürliche Weise aus

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^T (y_t - \hat{\mu})(y_{t-k} - \hat{\mu}) \quad (2.31)$$

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}. \quad (2.32)$$

Es gilt zu beachten, dass in (2.31) weniger Werte zur Verfügung stehen, da der Summationsindex bei $k + 1$ beginnt. Ist der zugrundeliegende Prozess ein AR(1)-Prozess, können wir die Tatsache ausnutzen, dass $\rho_1 = \phi$ gilt und mit Hilfe der stationären Varianzformel (2.15) auf Seite 14 einen Schätzer für die Varianz des Fehlerterms konstruieren

$$\hat{\sigma}^2 = \hat{\gamma}_0(1 - \hat{\phi}^2) = \hat{\gamma}_0(1 - \hat{\rho}_1^2). \quad (2.33)$$

Wir werden im Anschluss ein Beispiel betrachten, das verdeutlicht, warum Stationarität im Zusammenhang mit der Parameterschätzung so wichtig ist.

2.3.1 Parameterschätzung und Stationarität

Beispiel 2.1: Simulierter AR(1)-Prozess

Zur Veranschaulichung wurde ein AR(1)-Prozess $y_t = \phi y_{t-1} + \epsilon_t$ simuliert, wobei $\epsilon_t \sim N(0, 1)$ angenommen wurde. Der AR-Parameter wurde auf $\phi = 0.5$ gesetzt. Abbildung 2.2 zeigt die Zeitreihe $(y_t)_{t=0, \dots, 1000}$, die durch den simulierten Prozess erzeugt wurde. Der selbe Prozess wurde zusätzlich mit einem deterministischen Trend, $\nu_t = \frac{t}{100}$, überlagert, was natürlich zur Folge hat, dass die Stationarität nicht mehr gegeben ist, da der Erwartungswert nun von der Zeit t abhängt und nicht mehr konstant ist. Abbildung 2.3 auf der nächsten Seite zeigt die resultierende Zeitreihe. Die entsprechenden Parameterschätzer sind in Tabelle 2.1 auf der nächsten Seite zusammengefasst.

Die im stationären Modell ermittelten empirischen Parameterschätzungen stimmen sehr gut mit den wahren Werten überein, während die Werte im trendüberlagerten Modell völlig unbrauchbar sind. Die Ursache für diese Fehlschätzungen ist in Abbildung 2.3 auf der nächsten Seite angedeutet. Die horizontalen Linien geben den geschätzten Mittelwert (durchgezogene Linie) plus/minus zwei Standardabweichungen (gestrichelte Linien) wider. Man sieht deutlich, dass der „Trendkanal“ hier vollkommen unberücksichtigt bleibt. Es wird einfach über das volle Spektrum der Werte gemittelt. Daher sind die resultierenden Schätzer natürlich nicht brauchbar. Das Einrechnen

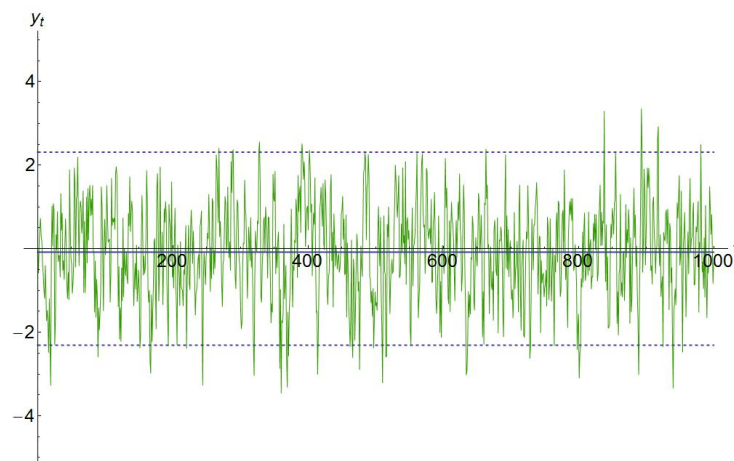


Abbildung 2.2: Stationäre Zeitreihe

2.3. PARAMETERSCHÄTZUNG

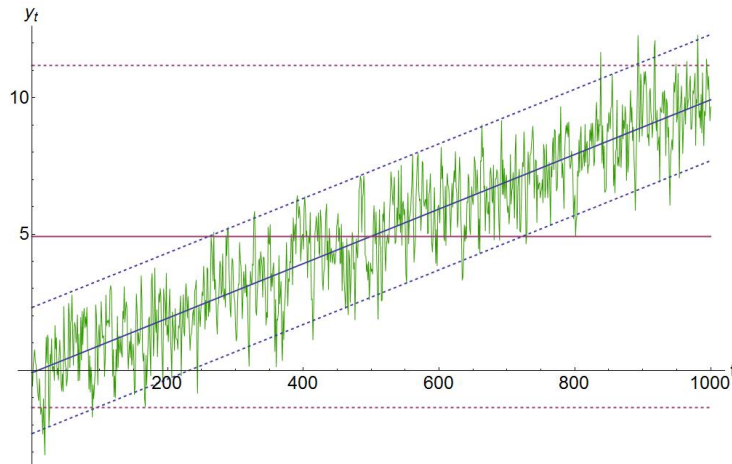


Abbildung 2.3: Zeitreihe mit deterministischem Trend

dieses Trends durch $y_t = x_t - \nu_t$ würde die richtigen Mittelwerte bzw. Standardabweichungen ergeben. Diese sind in Abbildung 2.3 ebenfalls in Form des besagten Trendkanals eingezeichnet.

Wir können nun besser verstehen, warum Stationarität einen so hohen Stellenwert in der Zeitreihenanalyse hat. Wir haben ebenfalls bereits eine Möglichkeit angedeutet Stationarität herzustellen, falls der ursprüngliche Prozess nicht stationär ist, nämlich durch Differenzenbildung (wir werden später auf diese Methode zurückkommen). Weiterhin haben wir verstanden, dass die Zeitreihe nur die sichtbare Manifestation eines Zufallsprozesses darstellt und aus diesem Grund aus den Zeitreihenwerten Schätzungen für die wahren Parameter des darunterliegenden Prozesses berechnet werden können.

Damit wird unmittelbar klar, dass selbst wenn ein und derselbe Prozess zweimal hintereinander eine Zeitreihe mit identischer Länge erzeugt, die

Parameter	Wahrer Wert	Stationär	Nicht-Stationär
μ	0	-0.08	4.91
γ_0	$1.\bar{3}$	1.33	9.83
γ_1	$0.\bar{6}$	0.69	9.16
ϕ	0.5	0.52	0.93
σ^2	1	0.98	1.29

Tabelle 2.1: Parameterschätzungen im AR(1)-Modell

Parameterschätzer nicht die selben sind, da die Zeitreihen ja Manifestationen eines Zufallsprozesses sind. Daher sind die Parameterschätzer ebenfalls vom Zufall abhängig und besitzen sogar eine eigene Wahrscheinlichkeitsverteilung. Auf solchen Verteilungen bauen verschiedene statistische Tests, wie beispielsweise der *Jarque-Bera*-Test auf. Wir werden Verteilungen von Schätzern nur am Rande betrachten, es sollte jedoch nicht außer Acht gelassen werden, dass verschiedene Schätzer, bspw. KQ-Schätzer (Kleinste Quadrate) oder ML-Schätzer (Maximum Likelihood) unter Umständen verschiedene asymptotische Eigenschaften besitzen. Damit ist gemeint, dass sich die Verteilungen der Schätzer für $T \rightarrow \infty$ unterschiedlich verhalten können. Man bevorzugt natürlich solche Schätzer, deren Eigenschaften möglichst angenehm oder möglichst gut an die Analysesituation angepasst sind. Wir werden im Folgenden zwei zentrale Schätzverfahren kennenlernen, die in der Statistik eine fundamentale Rolle spielen.

Aufgabe 2.5

Gegeben sei die (stationäre) Zeitreihe

$$(y_t)_{t=1,\dots,10} = \{2, 3, 4, 5, 4, 1, 0, 1, 3, 2\}.$$

Gehen Sie davon aus, dass diese Zeitreihe von einem AR(1)-Prozess erzeugt wurde, der einen normalverteilten Fehlerterm besitzt. Bestimmen Sie die empirischen Schätzer für μ , γ_0 , γ_1 , γ_2 , ρ_1 , ρ_2 , ϕ_{21} , ϕ_{22} und σ^2 .

2.3.2 Kleinste Quadrate (KQ)

Die Kleinste-Quadrate-Methode (KQ-Methode) ist eine intuitiv eingängige Schätzmethode. Sie ist sehr alt und geht auf Carl Friedrich Gauß zurück, der mit ihr astronomische Berechnungen vorgenommen hat. Die Idee ist recht einfach, wir werden sie zunächst am Beispiel des Erwartungswertes erörtern. Die Beobachtungswerte können als Abweichungen vom Erwartungswert $y_t = \mu + \epsilon_t$ geschrieben werden. Der Fehler ϵ_t fängt dabei gerade die Abweichungen auf, wobei hier keine Verteilung für ϵ_t spezifiziert werden muss. Man kann nun einfach die Summe der quadratischen Fehler minimieren

$$\min \sum_{t=1}^T \epsilon_t^2 = \min \sum_{t=1}^T (y_t - \mu)^2. \quad (2.34)$$

2.3. PARAMETERSCHÄTZUNG

Das Quadrieren des Fehlerterms ist dabei aus zwei Gründen sinnvoll, zum einen spielt die Richtung der Abweichung (positiv oder negativ) keine Rolle mehr, zum anderen fallen große Fehler stärker ins Gewicht. Wir finden das Minimum durch Nullsetzen der Ableitung der Summe in (2.34) nach μ

$$\frac{\partial}{\partial \mu} \sum_{t=1}^T (y_t - \mu)^2 = -2 \sum_{t=1}^T y_t + 2T\mu \stackrel{!}{=} 0. \quad (2.35)$$

Der partielle Differentialoperator ∂ wurde hier verwendet, weil die Summe von y_t und von μ abhängt. Einfaches Umstellen und Auflösen ergibt den bekannten KQ-Schätzer für den Erwartungswert, vgl. (2.29) auf Seite 21

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t. \quad (2.36)$$

Das gesamte AR(1)-Modell bietet ebenfalls die Möglichkeit der Parameterschätzung mittels Kleinste-Quadrate-Methode. Wir notieren den Prozess einmal in seiner allgemeinen Form

$$y_t = \theta_0 + \phi y_{t-1} + \epsilon_t. \quad (2.37)$$

Der zusätzliche Parameter θ_0 erscheint hier, weil der Erwartungswert in der Regel ungleich null ist, wie der zentrierte Prozess (2.3) auf Seite 6 zeigt. Das konstante Glied wurde einfach im Parameter $\theta_0 = (1 - \phi)\mu$ zusammengefasst. Diese Schreibweise wird sich noch als vorteilhaft erweisen, da sich die Konstante für Modelle höherer Ordnung anders zusammensetzt.

(2.37) hat die Form eines linearen Regressionsmodells, bei dem die Regressoren aus verzögerten Werten der Zeitreihe bestehen. Bei einer linearen Regression stellt man sich vor, dass eine Reihe von Datenpunkten durch eine Gerade möglichst gut repräsentiert werden soll. Zu diesem Zweck wird die Gerade genau so platziert, dass die Summe der quadrierten Abstände der Geraden zu allen Datenpunkten möglichst klein wird. Abbildung 2.4 zeigt die lineare Regression der Zeitreihe aus Aufgabe 2.5 auf der vorherigen Seite. Die Abstandskquadrate sind ebenfalls eingezeichnet.

Die Parameterschätzung mit der Kleinste-Quadrate-Methode erfolgt wieder durch Minimierung der Fehlerquadratsumme

$$\min \sum_{t=2}^T \epsilon_t^2 = \min \sum_{t=2}^T (y_t - \theta_0 - \phi y_{t-1})^2 = \min Q(\theta_0, \phi). \quad (2.38)$$

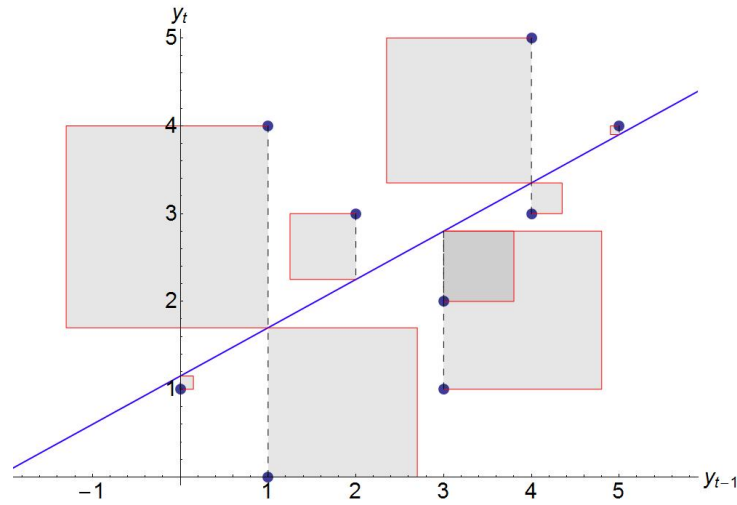


Abbildung 2.4: Kleinste-Quadrate Regression

Die Quadratsumme wurde hier mit Q bezeichnet, was uns an späterer Stelle viel Schreibarbeit ersparen wird. Beachten Sie, dass der Summationsindex bei $t = 2$ beginnt, da innerhalb der Quadratsumme $Q(\theta_0, \phi)$ in (2.38) ja y_{t-1} steht. Die KQ-Schätzer für θ_0 und ϕ ergeben sich nun aus den Bedingungen

$$\frac{\partial}{\partial \theta_0} Q(\theta_0, \phi) = 0 \quad \text{und} \quad \frac{\partial}{\partial \phi} Q(\theta_0, \phi) = 0. \quad (2.39)$$

Anstatt die Bedingungen (2.39) für jeden Parameter einzeln zu berechnen, was insbesondere bei Modellen höherer Ordnung sehr mühsam wäre, geht man in der Praxis einen effizienteren Weg. Zunächst aber eine kleine Auffrischung zur Vektor- und Matrixtransposition.

Einschub: Transponierte Matrix

Die zu einer Matrix \mathbf{A} mit

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

transponierte Matrix \mathbf{A}' entsteht einfach durch Vertauschung der Zeilen- und Spaltenindizes

$$\mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}.$$

Für eine symmetrische Matrix gilt demzufolge $\mathbf{A} = \mathbf{A}'$. Weiterhin gilt allgemein

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'.$$

Da ein Zeilenvektor als $(1 \times n)$ -Matrix und ein Spaltenvektor als $(n \times 1)$ -Matrix aufgefasst werden kann, wird ein transponierter Zeilenvektor zum Spaltenvektor und umgekehrt.

Wir können nun beispielsweise die Parameter θ_0 und ϕ in einem Vektor $\boldsymbol{\theta} = (\theta_0, \phi)'$ zusammenfassen und die Minimierungsbedingung äquivalent zu (2.39) als

$$\frac{\partial}{\partial \boldsymbol{\theta}'} Q(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}'} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) = 0 \quad (2.40)$$

formulieren, wobei

$$\mathbf{y} = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_2 \end{pmatrix} \quad \text{und} \quad \mathbf{X} = \begin{pmatrix} 1 & y_{t-1} \\ 1 & y_{t-2} \\ \vdots & \vdots \\ 1 & y_1 \end{pmatrix} \quad (2.41)$$

gilt. Auflösen von Bedingung (2.40) führt dann zum KQ-Schätzer, wobei wir den Rechenweg hier nicht erörtern wollen. Wichtiger ist das Ergebnis

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}. \quad (2.42)$$

Dieser Schätzer ist in der Literatur sehr bekannt und wird gegebenenfalls in Abhängigkeit von der Problemstellung modifiziert (beispielsweise bei nicht konstanter Varianz des Fehlerterms).

Aufgabe 2.6

Gegeben sei die (stationäre) Zeitreihe

$$(y_t)_{t=0,\dots,4} = \{1, 2, 1, 0, 1\}.$$

Gehen Sie zunächst davon aus, dass diese Zeitreihe von einem AR(1)-Prozess erzeugt wurde, der einen normalverteilten Fehlerterm besitzt. Bestimmen Sie den KQ-Schätzer für $\boldsymbol{\theta} = (\theta_0, \phi)'$.

2.3.3 Maximum Likelihood (ML)

Das Maximum-Likelihood-Prinzip wurde 1922 von Ronald Aylmer Fisher vorgeschlagen und ist heute eins der meist verwendeten Schätzprinzipien in der Statistik. Der Erfolg von ML leitet sich aus verschiedenen Komponenten ab, wie zum Beispiel angenehme asymptotische Eigenschaften, Anwendbarkeit für jede spezifizierbare Verteilung, Nähe zu Teststatistiken usw. Wir werden den Grundgedanken zunächst an einem einfachen Beispiel demonstrieren.

Beispiel 2.2: Maximum-Likelihood-Prinzip

Seien $\mathbf{x} = (x_1, x_2, x_3)$ drei unabhängige Realisierungen einer normalverteilten Zufallsvariable $X \sim N(\mu, \sigma^2)$. Dann zerfällt aufgrund der Unabhängigkeit die gemeinsame Dichtefunktion von \mathbf{x} in das Produkt der Dichten von x_1 , x_2 und x_3

$$p(x_1, x_2, x_3) = \prod_{k=1}^3 p(x_k) = p(x_1)p(x_2)p(x_3),$$

das Produktzeichen funktioniert hier genauso wie das Summenzeichen. Nach dem Maximum-Likelihood-Prinzip wird nun dieses Produkt von Dichten als Funktion der Parameter aufgefasst. Die Likelihood-Funktion ist dann einfach

$$L(\mu, \sigma^2) = \prod_{k=1}^3 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_k - \mu}{\sigma}\right)^2}.$$

Hier wurde die Dichtefunktion der Normalverteilung verwendet, da die Zufallsvariable X ja normalverteilt ist. Maximierung der Likelihoodfunktion ergibt dann den gesuchten Schätzer.

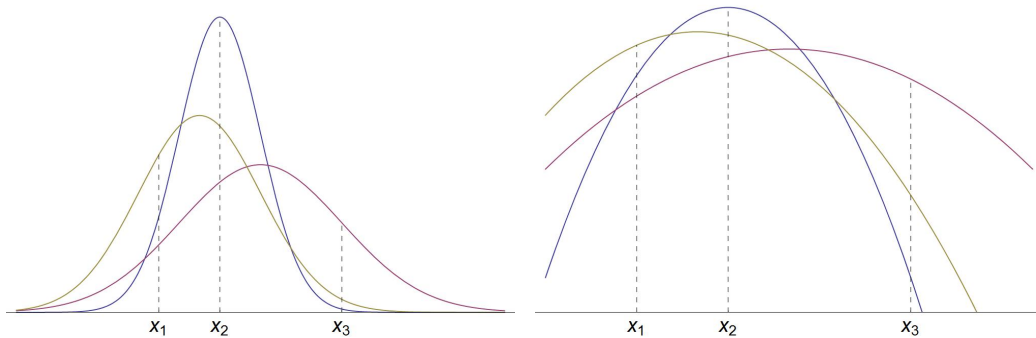


Abbildung 2.5: Likelihood- (links) und LogLikelihood-Funktion (rechts)

Die Parameter selbst sind keine Zufallsvariablen, daher die Bezeichnung „Likelihood“ in Abgrenzung zu „Probability“. Abbildung 2.5 (links) zeigt die Likelihood-Funktion für drei beliebige Datenpunkte aus Beispiel 2.2. Man sieht deutlich, dass die Funktionswerte der *Gauß*-Glocken an den Datenpunkten stark von der Lokalisation und der Breite der Glocken abhängen. Das Produkt dieser Funktionswerte wird in Abhängigkeit von den Parametern maximiert.

Aus technischen Gründen wird in der Praxis häufig nicht das Produkt der Dichten als Funktion der Parameter maximiert, sondern die Summe der logarithmierten Dichten.

Einschub: Logarithmus

Der (natürliche) Logarithmus ist die Umkehrfunktion der Exponentialfunktion, das bedeutet

$$x = \log[e^x].$$

Als Faustformel ist es nützlich sich einzuprägen, dass der Logarithmus alle Rechenoperationen um eine Ebene „herunterdrückt“, d.h. aus einer Potenz wird ein Produkt und aus einem Produkt wird eine Summe

$$\log[ax^b] = \log[a] + b \log[x].$$

Im Fall einer *Gauß*-Dichte ist das Logarithmieren besonders praktisch, da ja eine Exponentialfunktion involviert ist. Man schreibt daher die

LogLikelihood-Funktion für die unabhängigen Realisierungen (x_1, \dots, x_n)

$$l(\mu, \sigma^2) = \log L(\mu, \sigma^2) = -\frac{n}{2} \log[2\pi\sigma^2] - \frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \mu)^2. \quad (2.43)$$

In (2.43) steht nun aufgrund der Logarithmierung kein Produkt mehr, sondern lediglich eine Summe. Abbildung 2.5 (rechts) zeigt die Funktionswerte der LogLikelihood-Funktion an den Datenpunkten x_1 bis x_3 aus Beispiel 2.2 auf Seite 28 für verschiedene Parameterkonfigurationen.

Wir wollen nun ML-Schätzer für den AR(1)-Prozess ausrechnen. Hier stoßen wir jedoch auf die Schwierigkeit, dass die Realisierungen y_t nicht unabhängig sind. Dadurch zerfällt die gemeinsame Dichte nicht in ein Produkt der Einzeldichten. Wir erinnern uns jedoch an die *Markov*-Eigenschaft des AR(1)-Prozesses, speziell daran, dass die gesamte Vergangenheit des Prozesses in der letzten Realisierung gespeichert ist. Das bedeutet, bei Kenntnis des letzten Wertes y_{t-1} ist die Information über weiter zurückliegende Realisationen des Prozesses im Hinblick auf y_t irrelevant. Daher zerfällt die gemeinsame Dichte $p(y_T, \dots, y_1)$ in ein Produkt von Einzeldichten unter der Bedingung, dass die jeweils vorangegangene Realisierung, insbesondere y_0 , bekannt ist

$$p(y_T, y_{T-1}, \dots, y_1) = \prod_{t=1}^T p(y_t | y_{t-1}). \quad (2.44)$$

Die Dichte $p(y_t | y_{t-1})$ wird bedingte Dichte genannt, da sie an die Bedingung der Kenntnis von y_{t-1} geknüpft ist. Für unseren allgemeinen AR(1)-Prozess $y_t = \theta_0 + \phi y_{t-1} + \epsilon_t$ mit $\epsilon_t \sim N(0, \sigma^2)$ und y_0 bekannt, gilt dann

$$p(y_t | y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(y_t - \theta_0 - \phi y_{t-1})^2}{\sigma^2}}. \quad (2.45)$$

Damit kann ebenfalls die LogLikelihood-Funktion für den gesamten Prozess, vgl. (2.43), angegeben werden

$$l(\theta_0, \phi, \sigma^2) = \sum_{t=1}^T \log p(y_t | y_{t-1}) = -\frac{T}{2} \log[2\pi\sigma^2] - \frac{1}{2\sigma^2} Q(\theta_0, \phi). \quad (2.46)$$

Wir haben hier wieder die Abkürzung Q für die Quadratsumme verwendet, vgl. (2.38) auf Seite 25. Wir stellen fest, dass der Parameter σ^2 offenbar nicht in der Quadratsumme steht und leiten deswegen zunächst nach ihm ab

$$\frac{\partial}{\partial \sigma^2} l(\theta_0, \phi, \sigma^2) = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} Q(\theta_0, \phi) \stackrel{!}{=} 0. \quad (2.47)$$

2.3. PARAMETERSCHÄTZUNG

Auflösen von Bedingung (2.47) liefert den ML-Schätzer für σ^2

$$\hat{\sigma}^2 = \frac{1}{T} Q(\theta_0, \phi) = \frac{1}{T} \sum_{t=1}^T (y_t - \theta_0 - \phi y_{t-1})^2. \quad (2.48)$$

Dieser Schätzer hat eine uns bekannte Struktur, nämlich

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - E[y_t])^2, \quad (2.49)$$

wodurch wir den Schätzer für die stationäre Varianz (2.30) auf Seite 21 nun besser verstehen können. Wir hatten an dieser Stelle argumentiert, dass Zeitreihen in der Regel sehr lang sind und deswegen im Nenner nicht $T - 1$ sondern T steht. Wir sehen nun, dass es sich bei diesem Schätzer um den Maximum-Likelihood-Schätzer handelt, was natürlich ein fundierteres Argument ist.

Wir haben auch gesehen, dass der Parameter σ^2 in der LogLikelihood-Funktion (2.46) eine gewisse Sonderstellung einnimmt, da er nicht innerhalb der Quadratsumme Q steht. Wird der ML-Schätzer $\hat{\sigma}^2 = \frac{1}{T} Q(\theta_0, \phi)$ in die LogLikelihood-Funktion eingesetzt, erhält man die sog. konzentrierte LogLikelihood

$$l(\theta_0, \phi) = -\frac{T}{2} \log \left[\frac{2\pi}{T} \right] - \frac{T}{2} \log Q(\theta_0, \phi) - \frac{T}{2}. \quad (2.50)$$

Wir sehen, dass lediglich der mittlere Term auf der rechten Seite von (2.50) von den Parametern abhängt. Fassen wir θ_0 und ϕ wieder in einem Parametervektor $\boldsymbol{\theta} = (\theta_0, \phi)'$ zusammen und leiten die konzentrierte LogLikelihood-Funktion nach $\boldsymbol{\theta}$ ab, finden wir die Bedingung

$$\frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}) = -\frac{T}{2} \frac{\partial \log Q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{T}{2Q(\boldsymbol{\theta})} \cdot \frac{\partial Q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \stackrel{!}{=} 0. \quad (2.51)$$

Das entspricht aber nach dem Wegmultiplizieren der entsprechenden Terme gerade der Bedingung für den KQ-Schätzer

$$\frac{\partial}{\partial \boldsymbol{\theta}} Q(\boldsymbol{\theta}) = 0, \quad (2.52)$$

vgl. (2.40) auf Seite 27. Wir sehen also, dass für normalverteilte Fehlerterme KQ- und ML-Schätzer der Regressionsparameter übereinstimmen.

Aufgabe 2.7

Expandieren Sie die folgenden Terme:

- a) $\log[ab]$, b) $\log\left[\frac{a}{b}\right]$, c) $\log[2a^b]$
 d) $\log[\sqrt{a}]$, e) $\log[a^3\sqrt[4]{b}]$

2.3.4 Konfidenzintervalle

Konfidenzintervalle sind deshalb interessant, weil mit ihrer Hilfe eingeschätzt werden kann, wie verlässlich die Informationen, die wir aus der Zeitreihe gewinnen, im Hinblick auf die darunterliegende Prozessstruktur tatsächlich sind. Die Frage nach dem Konfidenzintervall ist auch eng mit der Frage nach der Verteilung des involvierten Schätzers verknüpft.

Um das Prinzip zu verstehen nehmen wir zunächst an, dass eine (stationäre) Zeitreihe $(y_t)_{t=1,\dots,T}$ vorliegt, die von einem reinen Zufallsprozess $y_t = \epsilon_t$ mit Fehlerterm $\epsilon_t \sim N(0, \sigma^2)$ erzeugt wurde. Wir können den Erwartungswert μ mit der bekannten Formel

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t \quad (2.53)$$

schätzen, vgl. (2.29) auf Seite 21. Dieser Schätzer hängt jedoch von den Werten der Zeitreihe ab, die ihrerseits vom Zufall abhängen. Damit hängt der Schätzer $\hat{\mu}$ selbst auch vom Zufall ab, nicht jedoch der wahre, aber unbekannte Parameter μ . In diesem Fall ist der Schätzer $\hat{\mu}$ ebenfalls normalverteilt, was bedeutet, dass wir nur seine ersten beiden Momente ermitteln müssen. Wir erhalten

$$E[\hat{\mu}] = E\left[\frac{1}{T} \sum_{t=1}^T y_t\right] = \frac{1}{T} \sum_{t=1}^T E[y_t] = \mu \quad (2.54)$$

$$\text{Var}[\hat{\mu}] = \text{Var}\left[\frac{1}{T} \sum_{t=1}^T y_t\right] = \frac{1}{T^2} \sum_{t=1}^T \text{Var}[y_t] = \frac{\sigma^2}{T}. \quad (2.55)$$

Die erste Eigenschaft, (2.54), bezeichnet man als Erwartungstreue. Erwartungstreue Schätzer, deren Abweichung vom wahren Parameter asymptotisch gegen null geht bezeichnet man als konsistent. Diese Eigenschaft wird durch (2.55) sichergestellt. Wir haben also bereits einige wichtige Eigenschaften des Schätzers $\hat{\mu}$ aufgedeckt.

2.3. PARAMETERSCHÄTZUNG

Das Konfidenzintervall selbst ist ein symmetrischer, um den Schätzer herum aufgebauter Bereich, der so konstruiert ist, dass er den wahren Parameter mit einer vorgegebenen Wahrscheinlichkeit überdeckt. Die Grenzen des Konfidenzintervalls lassen sich insbesondere für (asymptotisch) normalverteilte Schätzer sehr einfach durch Rückgriff auf die tabellierte Verteilungsfunktion der Standard-Normalverteilung berechnen. Für $\hat{\mu}$ muss gelten

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\hat{\mu} - \mu}{\frac{\sigma}{\sqrt{T}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha, \quad (2.56)$$

mit $z_{\kappa} = \Phi^{-1}(\kappa)$. Umformen liefert die gewünschten Grenzen. Das 95%-Konfidenzintervall für $\hat{\mu}$ beispielsweise hat die obere und untere Grenze

$$c_{o/u} = \hat{\mu} \pm 1.96 \frac{\sigma}{\sqrt{T}}, \quad (2.57)$$

wobei 1.96 der Wert der inversen Verteilungsfunktion der Standard-Normalverteilung an der Stelle $z = 0.975$ ist.

Beispiel 2.3: Konfidenzintervalle

Es werden per Simulation 100 Zeitreihen $(y_t)_{t=1,\dots,T}$ erzeugt. Der stochastische Prozess, der die Zeitreihen generiert ist $y_t = \epsilon_t$ mit $\epsilon_t \sim N(0, 1)$. Anschließend wird der (stationäre) Erwartungswert durch $\hat{\mu}$ geschätzt und die entsprechenden Konfidenzintervalle zum Konfidenzniveau $1 - \alpha = 0.95$ berechnet.

Abbildung 2.6 auf der nächsten Seite zeigt die Konfidenzintervalle aus Beispiel 2.3. Der wahre Mittelwert ist durch die horizontale Achse gekennzeichnet. Es ist deutlich zu erkennen, dass einige Intervalle den wahren Parameter nicht überdecken. Das war zu erwarten, da wir ja die Wahrscheinlichkeit der Überdeckung mit dem Konfidenzniveau auf 95% festgelegt hatten. Wir würden genaugenommen erwarten, dass fünf aus hundert Intervallen nicht den wahren Parameter überdecken. In der Tat sind es aber sechs. Diese Diskrepanz ist auf die kleine Stichprobe von 100 simulierten Zeitreihen zurückzuführen, für unendlich viele Zeitreihen würde die Anzahl der nichtüberdeckenden Konfidenzintervalle genau gegen 5% gehen. Da wir praktisch nie den Luxus genießen mit asymptotisch exakten Ergebnissen zu arbeiten, sollten wir immer im Hinterkopf behalten, dass auch der Zufall einen schlechten Tag haben kann.

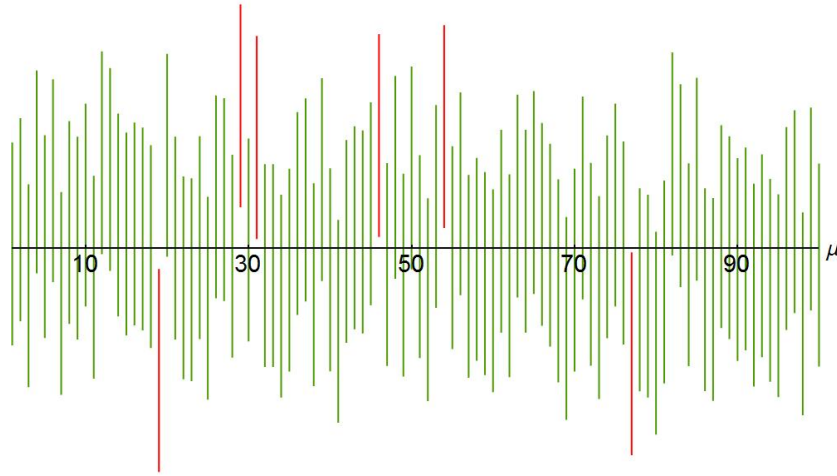


Abbildung 2.6: Konfidenzintervalle für μ

Wir haben bisher mit dem einfachen Zufallsprozess eine sehr simple Konstruktion betrachtet um das Prinzip des Konfidenzintervalls zu verstehen. In der Praxis sind die involvierten Prozesse natürlich komplizierter und es gibt andere Parameter, deren Schätzung und Analyse von vitaler Bedeutung für die Identifikation des stochastischen Prozesses sind. Wir haben beispielsweise gelernt, dass die partielle Autokorrelationsfunktion (PACF) eines $AR(p)$ -Prozesses für Lags $m > p$ auf null abfällt. Finden wir also eine Zeitreihe vor, deren empirische Schätzungen der PACF für Lag 2 und größer nahe null liegen, können wir vermuten, dass es sich um einen $AR(1)$ -Prozess handelt. In dieser Situation können Konfidenzintervalle für die Schätzer der Autokorrelationen und der partiellen Autokorrelationen unschätzbare Dienste bei der Beurteilung solcher Befunde leisten.

Wir kehren zurück zum $AR(1)$ -Prozess. Der zentrale Parameter bezüglich der Autokorrelation im $AR(1)$ -Modell ist ϕ . Wird nun aus den Daten per Maximum-Likelihood-Methode ein Schätzer $\hat{\phi}$ berechnet, so ist dieser Schätzer asymptotisch ebenfalls erwartungstreu und normalverteilt. Er besitzt die inverse *Fisher*-Information als Varianz, $\hat{\phi} \stackrel{a}{\sim} N(\phi, F^{-1})$. Wir wollen an dieser Stelle nicht näher beleuchten wie die *Fisher*-Information berechnet wird, sondern begnügen uns damit sie anzugeben

$$F(\phi) = \frac{T}{1 - \phi^2}. \quad (2.58)$$

Die *Fisher*-Information (2.58) hängt selbst vom wahren Parameter ϕ ab, den wir in der Regel natürlich nicht kennen und muss daher selbst mit Hilfe von $\hat{\phi}$

2.3. PARAMETERSCHÄTZUNG

geschätzt werden. Daher erhalten wir als asymptotisches Konfidenzintervall für ϕ zum Signifikanzniveau $1 - \alpha$

$$c_{o/u} = \hat{\phi} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{1 - \hat{\phi}^2}{T}}. \quad (2.59)$$

Im Fall der ACF und PACF werden die Konfidenzintervalle häufig auch um null herum gelegt, was lediglich eine Umkehrung der Betrachtungsweise beinhaltet. Man interessiert sich für gewöhnlich dafür, ob das Konfidenzintervall um den Schätzer den Wert null einschließt. In diesem Fall würde man argumentieren, dass sich keine signifikante (partielle) Autokorrelation nachweisen lässt. Gruppiert man das Konfidenzintervall um null herum, so ist genau dann eine signifikante (partielle) Autokorrelation nachweisbar, wenn der Schätzer aus diesem Bereich herausragt.

Beispiel 2.4: AR(1)-Prozess und Korrelogramm

Es wurde der AR(1)-Prozess

$$y_t = 0.8y_{t-1} + \epsilon_t$$

mit $\epsilon_t \sim N(0, 1)$ simuliert. Gleichzeitig wurden die Autokorrelationen und partiellen Autokorrelationen mit der Analysesoftware „EViews“ geschätzt.

Abbildung 2.7 zeigt die vom simulierten AR(1)-Prozess erzeugte Zeitreihe sowie das zugehörige Autokorrelogramm. Die vertikalen Markierungen im Korrelogramm kennzeichnen die Grenzen des 95%-Konfidenzbereichs. Die

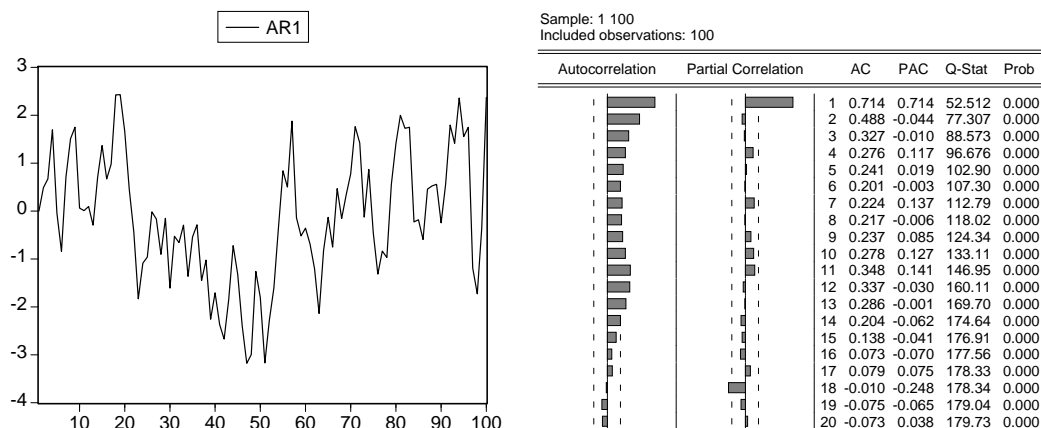


Abbildung 2.7: AR(1)-Prozess und Korrelogramm

Autokorrelation klingt zunächst auch ab, wie wir es erwarten würden, erzeugt dann aber vorübergehend noch eine Schwingung, die einen zusätzlichen Einfluss suggeriert, der gar nicht vorhanden ist. Die partielle Autokorrelation fällt wie erwartet nach Lag 1 sofort unter die Konfidenzgrenze ab, es wird jedoch zusätzlich eine signifikante partielle Autokorrelation mit Lag 18 geschätzt. Diese ist natürlich nicht vorhanden, da wir ja einen AR(1)-Prozess simuliert haben. Solche zufälligen Überschreitungen passieren aufgrund der Fehlerwahrscheinlichkeit von $\alpha = 5\%$. Hier ist bei realen Problemen ökonomischer Sachverstand gefragt um einschätzen zu können, ob solche „exotischen“ Korrelationen lediglich Schätzartefakte darstellen oder ob sie auf eine inhärente Fehlspezifikation des Modells hinweisen.

Schließlich wird der AR-Parameter ϕ mit $\hat{\phi} = 0.714$ geschätzt. Wir konstruieren das entsprechende Konfidenzintervall zum Konfidenzniveau $1 - \alpha = 95\%$

$$c_{o/u} = 0.714 \pm 1.96 \sqrt{\frac{1 - 0.714^2}{100}} \Rightarrow \phi \in [0.577, 0.851]. \quad (2.60)$$

Der wahre Wert des AR-Parameters war $\phi = 0.8$. Er wird somit voll vom Konfidenzintervall überdeckt.

Aufgabe 2.8

Gegeben sei die stationäre Zeitreihe

$$(y_t)_{t=0,\dots,8} = \{1, 2, 3, 3, 4, 2, 0, 2, 1\}.$$

Gehen Sie davon aus, dass ein AR(1)-Prozess zugrunde liegt. Schätzen Sie den AR-Parameter ϕ und bilden Sie das 95%-Konfidenzintervall mit Hilfe der geschätzten *Fisher-Information*.

2.4 Der AR(2)-Prozess

Der AR(2)-Prozess ist ebenfalls ein ausgesprochen wichtiger Repräsentant seiner Zunft. Er ist der sparsamste Prozess, der in der Lage ist schwingende Systeme zu beschreiben. Wir werden deshalb einige Eigenschaften des AR(2)-Prozesses näher beleuchten und gleichzeitig die Gelegenheit nutzen einiges, was wir über den AR(1)-Prozess gelernt haben, allgemeiner zu formulieren. Zunächst werden wir aber eine andere, platzsparende Schreibweise für AR(p)-Prozesse kennenlernen.

2.4.1 Operatoren

Operatoren sind Rechenvorschriften, die nach allgemeiner Konvention auf den Ausdruck angewendet werden, der unmittelbar nach ihnen steht. Wir haben bereits im vorangegangenen Abschnitt mit Operatoren zu tun gehabt, nämlich als wir die Ableitung der Quadratsumme nach dem Parametervektor gebildet haben, vgl. beispielsweise (2.52) auf Seite 31. Die Ableitung einer beliebigen Funktion $f(x)$ kann nämlich in Operatorschreibweise dargestellt werden

$$f'(x) = \frac{df(x)}{dx} = \frac{d}{dx}f(x). \quad (2.61)$$

Der Operator im Fall (2.61) ist $\frac{d}{dx}$, der Differentialoperator. Er wirkt auf den Ausdruck nach ihm, nämlich $f(x)$ und verursacht eine Manipulation dieses Ausdrucks nach einer festen Operationsvorschrift. Für Terme der Form x^a beispielsweise lautet diese Vorschrift $x^a \rightarrow ax^{a-1}$. Die Möglichkeit, den Operator wie in (2.61) einfach vor die Funktion zu ziehen, suggeriert, dass er denselben Rechenvorschriften unterliegt wie alle anderen Ausdrücke. In der Tat kann mit Operatoren in vielen Fällen einfach „gerechnet“ werden, was sie zu einem ausgesprochen leistungsfähigen Instrument macht.

Wir werden nun als ersten Operator den sog. Backshift-Operator B kennenlernen. Er manipuliert den ihm folgenden Ausdruck gemäß der Rechenvorschrift

$$B : \#_t \rightarrow \#_{t-1}. \quad (2.62)$$

Die $\#$ -Symbole in (2.62) sind lediglich Platzhalter. Wir erhalten beispielsweise $By_t = y_{t-1}$, $Bx_{t-k} = x_{t-k-1}$ usw. Die Darstellung (2.62) erlaubt es uns auch genauer zu verstehen, was für ein Objekt der Operator eigentlich ist. (2.62) hat die Gestalt einer Abbildungsvorschrift, nur das hier nicht eine Zahl auf eine andere Zahl abgebildet wird (Funktion, Relation) sondern eine Funktion auf eine andere Funktion (Operator). Wir können nun den Backshift-Operator selbst durch Rechenvorschriften manipulieren. Für normale Ausdrücke der Form $a^2x = a(ax)$ verursacht der Exponent, dass die ursprüngliche Rechenoperation (Multiplikation mit a) zweimal hintereinander ausgeführt wird. Für den Backshift-Operator gilt dasselbe

$$B^2y_t = B(By_t) = By_{t-1} = y_{t-2}. \quad (2.63)$$

Wir können auch den zu B inversen Operator B^{-1} erzeugen. Für normale Ausdrücke gilt $a^{-1}ax = x$. Fordert man dasselbe für den Backshift-Operator, so folgt

$$B^{-1}By_t = y_t \quad \Rightarrow \quad B^{-1}y_t = y_{t+1}. \quad (2.64)$$

KAPITEL 2. AR-PROZESSE UND ELEMENTARES

Der inverse Operator B^{-1} wird auch Forward-Operator genannt, weil er eine Zeitverschiebung in Vorwärtsrichtung verursacht.

Operatoren können auch zusammengesetzte Ausdrücke enthalten. Ein Beispiel für solch einen Operator ist der Rückwärts-Differenzen- oder ∇ -Operator („Nabla“-Operator). Er wird definiert durch $\nabla = 1 - B$, mit dem Backshift-Operator B . Er erzeugt somit die erste Rückwärtsdifferenz

$$\nabla y_t = (1 - B)y_t = y_t - y_{t-1}. \quad (2.65)$$

Das volle Potential dieser Operatorschreibweise lässt sich erkennen, wenn wir die zweite Rückwärtsdifferenz bilden, also die Differenz zwischen der aktuellen Rückwärtsdifferenz $y_t - y_{t-1}$ und der vergangenen Rückwärtsdifferenz $y_{t-1} - y_{t-2}$. Wir können dies sehr bequem tun, indem wir den ∇ -Operator quadrieren und mittels binomischer Formel berechnen

$$\nabla^2 y_t = (1 - B)^2 y_t = (1 - 2B + B^2)y_t = y_t - 2y_{t-1} + y_{t-2}. \quad (2.66)$$

Eine kurze Probe in Form von Separation der Rückwärtsdifferenzen offenbart, dass das Ergebnis absolut korrekt ist.

Um nun die Schreibweise von $\text{AR}(p)$ -Prozessen tatsächlich zu vereinfachen (oder zumindest abzukürzen) definiert man einen neuen, zusammengesetzten Operator, den AR-Operator

$$\phi(B) = 1 - \sum_{k=1}^p \phi_k B^k. \quad (2.67)$$

Diese Schreibweise ist in der Tat äußerst kompakt, was sich bereits erahnen lässt, wenn man den bekannten $\text{AR}(1)$ -Prozess mit ihrer Hilfe notiert

$$\phi(B)y_t = y_t - \phi_1 y_{t-1} = \epsilon_t. \quad (2.68)$$

Der $\text{AR}(2)$ -Prozess, den wir im Folgenden ausgiebig untersuchen werden, notiert genauso kompakt

$$\phi(B)y_t = y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = \epsilon_t. \quad (2.69)$$

Aufgabe 2.9

Schreiben Sie die folgenden Operatorausdrücke aus:

- a) $\nabla^3 y_t$, b) $B\nabla y_t$, c) $B^{-1}\nabla y_t$, d) $B^{-2}y_t$
e) $\nabla^2 B^{-1}y_t$, f) $B^{-2}\nabla^2 y_t$

2.4.2 Stationarität des AR(2)-Prozesses

Wir haben beim AR(1)-Prozess gesehen, dass die Voraussetzung für Stationarität durch den AR-Parameter geschaffen wurde. Wir können die dort abgeleitete Bedingung in Verbindung mit dem Backshift-Operator formulieren, um sie auch für Modelle höherer Ordnung nutzen zu können. Das AR(1)-Modell notiert in Operatorschreibweise

$$\phi(B)y_t = (1 - \phi_1 B)y_t = \epsilon_t, \quad (2.70)$$

mit dem AR-Parameter ϕ_1 . Der AR(1)-Prozess kann nur dann stationär sein, wenn der Einfluss der vergangenen Realisation betragsmäßig kleiner als eins ist oder anders formuliert, wenn für einen beliebigen Koeffizient λ vor dem Backshift-Operator $|\lambda| < 1$ gilt. Wie sieht die Sache nun für den AR(2)-Prozess

$$\phi(B)y_t = y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = \epsilon_t \quad (2.71)$$

aus? Wir wissen, dass die benötigte Information ausschließlich im Operator steckt und zerlegen deshalb den Operator $\phi(B)$ wie ein gewöhnliches Polynom

$$1 - \phi_1 B - \phi_2 B^2 = (1 - \lambda_1 B)(1 - \lambda_2 B) = 1 - (\lambda_1 + \lambda_2)B + \lambda_1 \lambda_2 B^2. \quad (2.72)$$

Gleichung (2.72) zeigt, dass der AR(2)-Operator offensichtlich in ein Produkt aus zwei AR(1)-Operatoren faktorisiert werden kann. Hier können wir nun für jeden AR-Operator separat die Stationaritätsbedingung fordern. Weiterhin sehen wir, dass die Koeffizienten λ_1 und λ_2 gewissermaßen elementare Bausteine der AR-Parameter ϕ_1 und ϕ_2 sind, denn wir erhalten aus (2.72) die Bedingung

$$\phi_1 = \lambda_1 + \lambda_2 \quad \text{und} \quad \phi_2 = -\lambda_1 \lambda_2. \quad (2.73)$$

Die λ -Koeffizienten werden auch als Wurzeln oder „roots“ bezeichnet, oder noch genauer als „inverted roots“. Was es damit auf sich hat erfahren wir im Anschluss noch genauer; in der Tat hat es nicht ausschließlich damit zu tun, dass sie für die AR-Parameter metaphorisch als Wurzeln fungieren.

Wie findet man nun die Wurzeln λ_1 und λ_2 im Allgemeinen? Gleichung (2.72) hat die Form einer quadratischen Gleichung, jedoch macht eine Lösung für die Backshift-Operatoren keinen Sinn. Deshalb setzt man in den AR-Operator anstelle von B eine Variable z ein

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z). \quad (2.74)$$

Die rechte Seite von (2.74) verrät uns, dass die Nullstellen bei $z = 1/\lambda_1$ und $z = 1/\lambda_2$ liegen, während die linke Seite eine quadratische Gleichung ist, deren Nullstellen aus dem Ansatz

$$1 - \phi_1 z - \phi_2 z^2 = 0 \quad (2.75)$$

folgen. Die Lösungen z_1 und z_2 sind die eigentlichen Wurzeln oder roots. Für $z = \lambda^{-1}$ gilt aber $z_{1/2} = \lambda_{1/2}^{-1}$. Das erklärt auch die Begriffe „roots“ und „inverted roots“. Für die eigentlichen Wurzeln $z_{1/2}$ gilt natürlich auch die entsprechend umgekehrte Stationaritätsbedingung $|z_k| > 1$ für $k = 1, 2$. Um keine größere Verwirrung zu stiften werden wir im Folgenden die „inverted roots“ $\lambda_{1/2}$ betrachten. Wir sollten jedoch im Hinterkopf behalten, dass beide Formulierungen äquivalent sind. Wir multiplizieren (2.74) mit $z^{-2} = \lambda^2$ und erhalten

$$\lambda^2 - \phi_1 \lambda - \phi_2 = (\lambda - \lambda_1)(\lambda - \lambda_2). \quad (2.76)$$

Gleichung (2.76) hat eine noch klarere Struktur als (2.74). Die rechte Seite wird trivialerweise null für $\lambda = \lambda_{1/2}$, während sich λ_1 und λ_2 selbst als Nullstellen der quadratischen Gleichung auf der linken Seite von (2.76) ergeben.

Einschub: Quadratische Gleichungen

Eine quadratische Gleichung der Form

$$x^2 + px + q = 0$$

kann durch die sog. pq -Formel gelöst werden. Die Lösungen ergeben sich zu

$$x_{1/2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}.$$

Die pq -Formel liefert zwei Lösungen, da ein Polynom n -ten Grades immer genau n nicht notwendigerweise reelle Nullstellen besitzt.

Wir erhalten also aus der quadratischen Gleichung (2.76) mit Hilfe der pq -Formel die Wurzeln (inverted roots)

$$\lambda_{1/2} = \frac{\phi_1}{2} \pm \sqrt{\frac{\phi_1^2}{4} + \phi_2}. \quad (2.77)$$

Wir wollen Gleichung (2.76) für einen kurzen Moment aus einem anderen Blickwinkel betrachten. Das AR(2)-Modell kann alternativ als VAR(1)-Modell (Vector Auto Regressive) geschrieben werden. Ohne auf Details ein-

zugehen verifiziert man leicht, dass die Schreibweise

$$\underbrace{\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}}_{\mathbf{\Phi}} \underbrace{\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}}_{\mathbf{y}_{t-1}} + \underbrace{\begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}}_{\boldsymbol{\epsilon}_t} \quad (2.78)$$

das AR(2)-Modell exakt reproduziert. Die Berechnung der Eigenwerte von $\mathbf{\Phi}$ führt nun zu

$$\det[\mathbf{\Phi} - \lambda \mathbf{I}] = -\lambda(\phi_1 - \lambda) - \phi_2 = \lambda^2 - \phi_1 \lambda - \phi_2, \quad (2.79)$$

dem charakteristischen Polynom aus (2.74), respektive (2.76).

Einschub: Eigenwerte

Jede quadratische $(n \times n)$ -Matrix \mathbf{A} besitzt n Eigenwerte $\lambda_1, \dots, \lambda_n$ die der Eigenwertgleichung

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

genügen. Ein passender Vektor $\mathbf{v} \neq 0$ wird Eigenvektor genannt. Die Eigenwerte können dann mit Hilfe der charakteristischen Gleichung

$$\det[\mathbf{A} - \lambda \mathbf{I}] = 0,$$

mit der passenden Einheitsmatrix \mathbf{I} gewonnen werden. Die Eigenwerte legen einige elementare Eigenschaften der Matrix fest. Ist beispielsweise ein Eigenwert null so ist die Matrix singulär.

Gleichung (2.79) liefert damit eine alternative Herleitung des charakteristischen Polynoms des AR(2)-Prozesses. Eine Frage haben wir bislang jedoch nicht erörtert. Wir haben in (2.77) die Nullstellen $\lambda_{1/2}$ berechnet, was passiert jedoch, wenn der resultierende Term unter der Wurzel negativ ist?

Bekannterweise hat ein Polynom n -ten Grades auch genau n Nullstellen. Diese Nullstellen müssen aber nicht zwangsläufig in den Bereich der reellen Zahlen fallen sondern können auch teilweise oder vollständig komplex sein.

Einschub: Komplexe Zahlen

Eine Zahl z heißt Element der komplexen Zahlen $z \in \mathbb{C}$, wenn sie die Struktur

$$z = a + ib \quad \text{mit} \quad a, b \in \mathbb{R}$$

besitzt. i bezeichnet hier die imaginäre Einheit, die die Eigenschaft $i^2 = -1$ besitzt. Sie ermöglicht die Berechnung negativer Wurzeln, da für beliebiges $c \in \mathbb{R}^+$

$$\sqrt{-c} = i\sqrt{c}$$

gilt. Die beiden Bestandteile einer komplexen Zahl werden Realteil, $\text{Re}[z] = a$, und Imaginärteil, $\text{Im}[z] = b$, genannt.

Im Zusammenhang mit den Wurzeln des AR(2)-Prozesses ist für uns von vitalem Interesse, wie der Betrag einer komplexen Zahl gebildet wird. Dazu ist es hilfreich sich die Struktur der komplexen Zahlenebene anzuschauen. Abbildung 2.8 auf der nächsten Seite zeigt den reellen und imaginären Zahlenstrahl, die senkrecht aufeinander stehen und so die komplexe Ebene bilden. Eine komplexe Zahl ist nun durch ihre Koordinaten auf der reellen und imaginären Achse festgelegt. Für gewöhnliche reelle Zahlen ist der Betrag als Abstand vom Ursprung, also als Abstand von null definiert. Dieselbe Definition gilt im Rahmen der komplexen Zahlen. Der Unterschied ist hier, dass die reellen Zahlen nur aus einem Zahlenstrahl bestehen, die komplexen Zahlen aber aus einer Ebene. Um hier den Abstand vom Ursprung zu ermitteln müssen wir den Satz des Pythagoras bemühen. Demnach ergibt sich der Betrag von z als

$$|z| = r = \sqrt{a^2 + b^2}. \quad (2.80)$$

Die magische Grenze $|z| = 1$ wird also in der komplexen Zahlenebene zu einem Kreis, der in Abbildung 2.8 angedeutet ist. Damit lässt sich die notwendige Stationaritätsbedingung für das AR(2)-Modell allgemeingültig formulieren:

Ein AR(2)-Prozess kann nur dann stationär sein, wenn seine Wurzeln außerhalb des komplexen Einheitskreises liegen.

Für die „inverted roots“ $\lambda_{1/2}$ bedeutet das natürlich genau das Gegenteil, sie müssen beide innerhalb des komplexen Einheitskreises liegen, damit der Prozess stationär ist.

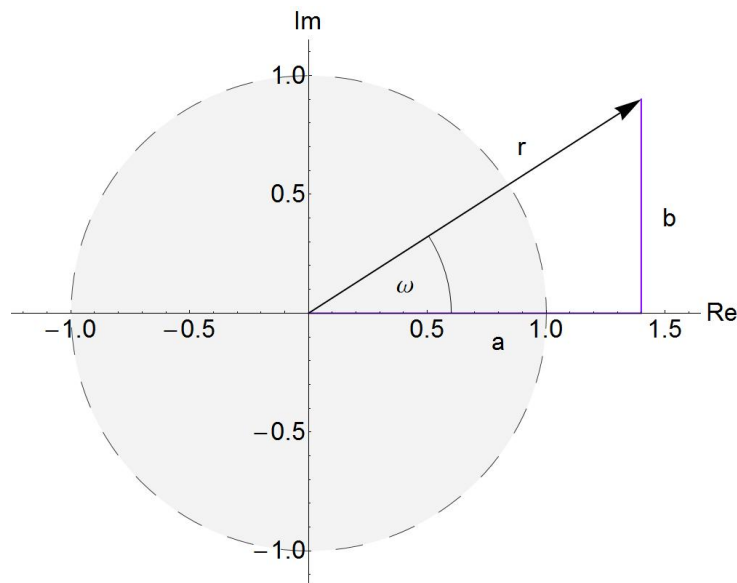


Abbildung 2.8: Komplexe Zahlenebene

2.4.3 Darstellung in Polar-Koordinaten

Es ist oft bequemer eine komplexe Zahl in Polar-Koordinaten darzustellen, da diese besser an die Problemstellung angepasst sind als kartesische Koordinaten. In Polar-Koordinaten wird z nicht durch die Real- und Imaginärteile a und b festgelegt, sondern durch den Radius r und den Winkel ω . Einen Zusammenhang zwischen dem Radius und den kartesischen Koordinaten haben wir schon mit Hilfe des Satzes von Pythagoras gefunden, siehe (2.80). Um einen Ausdruck für die Winkelkoordinate ω abzuleiten erinnern wir uns an zwei elementare Zusammenhänge im rechtwinkligen Dreieck in Abbildung 2.8, nämlich

$$\cos \omega = \frac{a}{r} \quad \text{und} \quad \sin \omega = \frac{b}{r}. \quad (2.81)$$

Daraus können wir durch geschicktes Arrangieren folgenden Zusammenhang gewinnen

$$\frac{b}{a} = \frac{r \sin \omega}{r \cos \omega} = \tan \omega. \quad (2.82)$$

Nun müssen wir nur noch die Umkehrfunktion des Tangens bemühen und erhalten den gesuchten Ausdruck

$$\omega = \arctan \left[\frac{b}{a} \right]. \quad (2.83)$$

Es gilt hier zu beachten, dass die Definitionsbereiche beim Übergang von kartesischen zu Polar-Koordinaten ebenfalls verändert werden. Während a und b jeden beliebigen reellen Wert annehmen konnten, $a, b \in \mathbb{R}$, gilt für die Radius- und Winkelkoordinaten $r \in [0, \infty)$ und $\omega \in [-\pi, \pi]$. Wir fassen die beiden alternativen Darstellungen, vgl. (2.82), zusammen

$$z = a + ib = r(\cos \omega + i \sin \omega). \quad (2.84)$$

Wir haben am Anfang des Abschnitts unterstellt, dass im Fall komplexer Wurzeln Polar-Koordinaten besser an die Problemstellung angepasst sind. Wir sehen in (2.84), dass die Polar-Darstellung trigonometrische Funktionen enthält. Daraus können wir im Umkehrschluss folgern, dass es im Falle komplexer Wurzeln offenbar zu Schwingungen des Systems kommt. Genau das ist auch der Fall.

Die Polar-Darstellung der komplexen Zahl z ist aber noch viel leistungsfähiger als es bislang scheint. Das wird klar, wenn wir versuchen die Potenz einer komplexen Zahl zu berechnen. In kartesischen Koordinaten bekämen wir als Lösung

$$z^t = (a + ib)^t = \sum_{k=0}^t \binom{t}{k} a^{t-k} (ib)^k. \quad (2.85)$$

In (2.85) wurde der binomische Lehrsatz verwendet, den wir ohne weiteren Kommentar stehen lassen, da wir nicht die Absicht haben den Ausdruck mühsam auszurechnen. Im Exponenten steht etwas suggestiv bereits t , was sich im Anschluss aufklären wird. Wir verwenden stattdessen Polar-Koordinaten in Verbindung mit der *Eulerschen* Formel.

Einschub: Eulersche Formel

Zwischen den trigonometrischen Funktionen und der Exponentialfunktion besteht folgender Zusammenhang

$$e^{i\omega} = \cos \omega + i \sin \omega.$$

Der Zusammenhang entsteht aus der Reihenentwicklung der involvierten Funktion

$$\begin{aligned} e^{i\omega} &= \sum_{k=0}^{\infty} \frac{(i\omega)^k}{k!} = 1 + i\omega - \frac{\omega^2}{2!} - i\frac{\omega^3}{3!} + \frac{\omega^4}{4!} + \dots \\ &= \left(1 - \frac{\omega^2}{2!} + \frac{\omega^4}{4!} - \dots\right) + i \left(\omega - \frac{\omega^3}{3!} + \frac{\omega^5}{5!} - \dots\right) \\ &= \cos \omega + i \sin \omega. \end{aligned}$$

Der Ausdruck $k!$ (Fakultät) kann dabei rekursiv definiert werden

$$k! = k \cdot (k-1)! \quad \text{und} \quad 0! = 1.$$

Das heißt, wir erhalten für die Potenz einer komplexen Zahl in Polar-Darstellung

$$z^t = r^t (\cos \omega + i \sin \omega)^t = r^t e^{i\omega t} = r^t (\cos[\omega t] + i \sin[\omega t]). \quad (2.86)$$

Dieser Ausdruck ist wesentlich einfacher zu handhaben und er verrät uns einiges über die Struktur der Schwingungen. Zunächst wird klar, dass die Schwingungen für $r < 1$ in der Zeit abnehmen (gedämpfte Schwingung) denn es gilt ja

$$\lim_{t \rightarrow \infty} r^t = 0 \quad \text{für} \quad r < 1. \quad (2.87)$$

Weiterhin können wir die Schwingungszeit (Periode) einer kompletten Schwingung berechnen. Wir wissen ja, dass ein kompletter Zyklus der Sinus- bzw. Cosinusfunktion genau dem Kreisumfang 2π entspricht. Bezeichnet hier T die Periode erhalten wir durch Gleichsetzen

$$\omega T = 2\pi \quad \Rightarrow \quad T = \frac{2\pi}{\omega}. \quad (2.88)$$

Wir verstehen nun auch besser was mit dem Terminus „an die Problemstellung angepasst“ gemeint war. Bei der Berechnung der Schwingungszeit in (2.88) wurde implizit $\omega > 0$ angenommen. Diese Annahme ist unproblematisch, da für Sinus und Cosinus mit $\omega \in [-\pi, \pi]$

$$\cos[-\omega] = \cos \omega \quad \text{und} \quad \sin[-\omega] = -\sin \omega \quad (2.89)$$

gilt. Das heißt nichts anderes, als dass für konjugiert-komplexe Wurzeln

$$\lambda_{1/2} = a \pm ib = r(\cos \omega \pm i \sin \omega) \quad \text{mit} \quad \omega = \arctan \left| \frac{b}{a} \right| \quad (2.90)$$

gilt. Damit ist ω nicht negativ und die Periode $T \geq 0$.

Aufgabe 2.10

Gegeben sei der folgende AR(2)-Prozess

$$y_t = y_{t-1} - \frac{1}{2}y_{t-2} + \epsilon_t$$

mit $\epsilon_t \sim N(0, \sigma^2)$.

- a) Bestimmen Sie die Wurzeln λ_1 und λ_2 ,
- b) Berechnen Sie die Schwingungsdauer T ,
- c) Nach welcher Zeit τ sind die Schwingungen bis auf 10% ihrer Anfangsamplitude abgeklungen?

2.4.4 Autokovarianz und Autokorrelation

Wie im Fall des AR(1)-Prozesses hängt die Autokovarianz bzw. die Autokorrelation im stationären Fall ausschließlich vom Lag k ab. Wir gehen hier ganz analog zu (2.20) auf Seite 15 vor um die *Yule-Walker*-Gleichung abzuleiten

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \quad | \cdot y_{t-k} \\ \Rightarrow y_t y_{t-k} &= \phi_1 y_{t-1} y_{t-k} + \phi_2 y_{t-2} y_{t-k} + \epsilon_t y_{t-k} \quad | E[\dots] \\ \Rightarrow E[y_t y_{t-k}] &= \phi_1 E[y_{t-1} y_{t-k}] + \phi_2 E[y_{t-2} y_{t-k}] + E[\epsilon_t y_{t-k}] \\ \Rightarrow \gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}. \end{aligned} \quad (2.91)$$

Für die Autokorrelation ergibt sich eine identische Vorschrift, da lediglich alle Terme durch γ_0 zu teilen sind. Wir können daher die *Yule-Walker*-Gleichung ebenfalls in Kurzform als

$$\phi(B)\gamma_k = 0 \quad \text{bzw.} \quad \phi(B)\rho_k = 0 \quad (2.92)$$

schreiben. Mit (2.92) gewinnen wir eine rekursive Gleichung für die Autokovarianz bzw. die Autokorrelation. Wie gelangen wir aber an eine explizite Formel und müssen wir zwischen reellen und imaginären Lösungen unterscheiden? In einem solchen Fall machen Mathematiker einen „Ansatz“,

was nichts anderes heißt, als dass sie raten. Bei einem solchen Ansatz wird aber nicht blind irgendetwas geraten, sondern man hat eine Vorstellung von der Form und Struktur der Lösung und rät dann einen allgemeinen Lösungstyp, dessen Koeffizienten man durch geschicktes Ausnutzen der Randbedingungen zu bestimmen versucht. Das hört sich sehr kompliziert an, ist es aber letztendlich nicht. Wir wollen einmal Schritt für Schritt die Autokorrelationsfunktion des AR(2)-Prozesses „erraten“.

Zunächst rufen wir uns noch einmal ins Gedächtnis, wie die Autokorrelation des AR(1)-Prozesses aufgebaut war. Im AR(1)-Modell haben wir zunächst nur eine einzige Wurzel, nämlich $\lambda = \phi$, was im stationären Fall zur expliziten Autokorrelation

$$\rho_k = \lambda^k, \quad (2.93)$$

vgl. (2.23) auf Seite 16, führt. Wir „raten“ nun, dass die Autokorrelation im AR(2)-Modell aus einer Linearkombination der Wurzeln besteht

$$\rho_k = A_1 \lambda_1^k + A_2 \lambda_2^k, \quad (2.94)$$

mit den Koeffizienten $A_1, A_2 \in \mathbb{R}$. Wir nehmen zunächst an, dass $\lambda_1 \neq \lambda_2$ gilt. Für $k = 0, 1$ erhält man das lineare Gleichungssystem

$$\begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix}. \quad (2.95)$$

Für die Autokorrelation mit Lag 0 gilt definitionsgemäß $\rho_0 = 1$. Die Autokorrelation mit Lag 1 kann mit einem kleinen Trick berechnet werden, indem man den Zusammenhang $\rho_1 = \rho_{-1}$ ausnutzt und die *Yule-Walker*-Gleichung (2.92) beginnend mit $k = 1$ anschreibt

$$\rho_1 = \phi_1 + \phi_2 \rho_1 \quad \Rightarrow \quad \rho_1 = \frac{\phi_1}{1 - \phi_2}. \quad (2.96)$$

Damit erhalten wir die Koeffizienten des Ansatzes (2.94) durch Lösen des linearen Gleichungssystems

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \frac{\phi_1}{1 - \phi_2} \end{pmatrix}. \quad (2.97)$$

Die Inverse Matrix in (2.97) existiert in jedem Fall, wenn $\lambda_1 \neq \lambda_2$ gilt, und man erhält explizit

$$A_{1/2} = \pm \frac{\lambda_{2/1} - \frac{\phi_1}{1 - \phi_2}}{\lambda_2 - \lambda_1}. \quad (2.98)$$

KAPITEL 2. AR-PROZESSE UND ELEMENTARES

Allerdings war für die Koeffizienten gefordert, dass Sie in der Menge der reellen Zahlen liegen. Daher ist diese Lösung nur für $\lambda_1, \lambda_2 \in \mathbb{R}$ brauchbar.

Als nächstes machen wir uns klar, dass eine Lösung mit komplexen Wurzeln auch den Fall $\lambda_1 = \lambda_2$ umfasst, da jede reelle Zahl auch Element der komplexen Zahlen ist, wobei lediglich der Imaginärteil null ist. Aus dieser Überlegung folgt, dass wir mit einer Lösung für komplexe Wurzeln automatisch auch die Lösung für identische reelle Wurzeln gefunden haben. Wir gehen zweckmäßigerweise wieder zu Polar-Koordinaten über

$$\begin{aligned}\rho_k &= A_1 r^k (\cos \omega + i \sin \omega)^k + A_2 r^k (\cos \omega - i \sin \omega)^k \\ &= A_1 r^k e^{i\omega k} + A_2 r^k e^{-i\omega k} \\ &= (A_1 + A_2) r^k \cos[\omega k] + i(A_1 - A_2) r^k \sin[\omega k]\end{aligned}\tag{2.99}$$

Wir wissen, dass die Autokorrelation selbst eine reelle Größe ist, oder allgemeiner formuliert, eine komplexe Größe mit Imaginärteil null. Dieses Wissen schöpfen wir nun aus, um etwas über die Koeffizienten zu erfahren

$$\text{Im}[\rho_k] = (A_1 - A_2) r^k \sin[\omega k] = 0.\tag{2.100}$$

Da A_1 und A_2 reell sind folgt aus (2.100) unmittelbar $A_1 = A_2 = A$. Damit vereinfacht sich die Korrelationsgleichung (2.99) bereits drastisch und wir erhalten

$$\rho_k = 2A r^k \cos[\omega k].\tag{2.101}$$

Um den verbleibenden Koeffizienten A zu determinieren, nutzen wir wieder die Autokorrelation mit Lag null, woraus in Verbindung mit (2.101)

$$\rho_0 = 2A = 1 \quad \Rightarrow \quad A = \frac{1}{2}\tag{2.102}$$

folgt. Damit können wir explizit die Autokorrelation mit Lag k in kartesischen sowie in Polar-Koordinaten angeben

$$\rho_k = \frac{1}{2}(\lambda_1^k + \lambda_2^k) = r^k \cos[\omega k].\tag{2.103}$$

Für $\lambda_1 = \lambda_2 = \lambda$ führt Gleichung (2.103) wieder auf die Autokorrelationsfunktion des AR(1)-Prozesses

$$\rho_k = \lambda^k,\tag{2.104}$$

vgl. (2.23) auf Seite 16. In der Tat ist ein solcher Prozess nicht anhand seiner Korrelationsstruktur von einem AR(1)-Prozess mit $\phi = \lambda$ unterscheidbar.

2.4. DER AR(2)-PROZESS

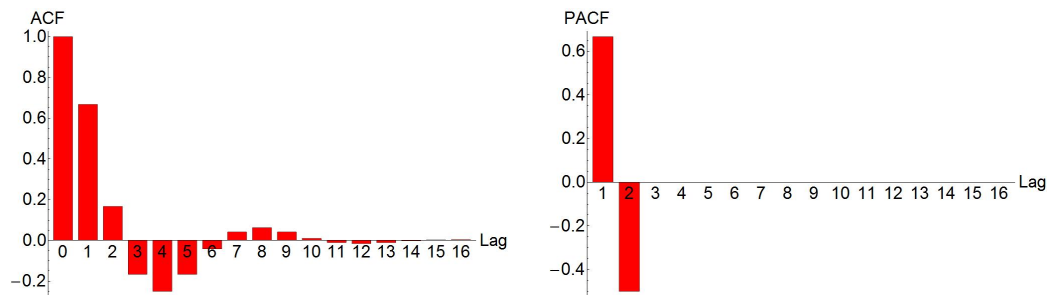


Abbildung 2.9: ACF und PACF des AR(2)-Prozesses mit $\phi_1 = 1$ und $\phi_2 = -0.5$

Die partielle Autokorrelation ergibt sich wieder aus der Lösung des bekannten Gleichungssystems (2.24) auf Seite 17. Hier ergibt sich nichts Neues, die PACF beseitigt die „Interferenzen“ zwischen den ersten beiden Lags und die „Echos“ für $k > 2$. Abbildung 2.9 zeigt die Autokorrelationsfunktion (links) und die partielle Autokorrelationsfunktion (rechts) des AR(2)-Prozesses aus Aufgabe 2.10 auf Seite 46. Die PACF offenbart die isolierten Einflüsse der AR-Quellen.

Aufgabe 2.11

Gegeben sei der AR(2)-Prozess

$$y_t = 0.2y_{t-1} + 0.35y_{t-2} + \epsilon_t.$$

Überprüfen Sie die Stationarität und geben Sie die Korrelationsfunktion für beliebiges Lag k an.

Aufgabe 2.12

Für den AR(2)-Prozess aus Aufgabe 2.10 auf Seite 46 haben wir die Wurzeln (inverted roots)

$$\lambda_{1/2} = \frac{1}{2} \pm i\frac{1}{2}$$

errechnet. Ermitteln Sie die Autokorrelationen ρ_2 und ρ_4 für diesen Prozess.

2.4.5 Parameterschätzung

Die Parameterschätzung im AR(2)-Modell hält keine großen Überraschungen bereit, es gelten weiterhin die Schätzer (2.29) bis (2.32) auf Seite 21 für die Stationären Momente.

Der Kleinste-Quadrate Schätzer für die „Regressionskoeffizienten“ des AR(2)-Modells entsteht nun aus der Bedingung

$$\min \sum_{t=3}^T \epsilon_t^2 = \min \sum_{t=3}^T (y_t - \theta_0 - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2. \quad (2.105)$$

Daraus leitet sich für den Parametervektor $\boldsymbol{\theta} = (\theta_0, \phi_1, \phi_2)'$ abermals der Schätzer (2.42) von Seite 27

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (2.106)$$

ab, wobei nun für den Vektor \mathbf{y} und die Matrix \mathbf{X}

$$\mathbf{y} = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_3 \end{pmatrix} \quad \text{und} \quad \mathbf{X} = \begin{pmatrix} 1 & y_{t-1} & y_{t-2} \\ 1 & y_{t-2} & y_{t-3} \\ \vdots & \vdots & \vdots \\ 1 & y_2 & y_1 \end{pmatrix} \quad (2.107)$$

gilt.

2.5 AR(p)-Prozesse

Der AR(p)-Prozess ist die natürlichen Erweiterung der bisher behandelten AR(1)- und AR(2)-Prozesse. Er wird im Allgemeinen effizient in Operator-schreibweise notiert

$$\phi(B)y_t = y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \epsilon_t. \quad (2.108)$$

Die Analyse von AR(p)-Prozessen wird mit wachsender Ordnung p schnell sehr komplex und ist in der Regel nur durch den Einsatz spezieller Software möglich. Der Grund hierfür wird schnell klar, wenn die Stationaritätseigenschaften des Prozesses betrachtet werden.

2.5.1 Stationarität des AR(p)-Prozesses

Wie in den vorangegangenen Beispielen stecken auch hier alle Informationen zum Auffinden der Wurzeln des Prozesses im Operator $\phi(B)$. Nach Einsetzen der Variable z erhalten wir wieder das charakteristische Polynom

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p. \quad (2.109)$$

Verwenden wir die Darstellung mit „inverted roots“ $\lambda = z^{-1}$, erhalten wir die Wurzeln des AR(p)-Prozesses aus der Gleichung

$$\lambda^p - \phi_1 \lambda^{p-1} - \dots - \phi_{p-1} \lambda - \phi_p = 0. \quad (2.110)$$

(2.110) ist eine polynomiale Gleichung n -ter Ordnung. Sie kann für $n > 4$ nur numerisch gelöst werden. Es können sowohl komplexe wie reelle Wurzeln auftreten, komplexe Wurzeln kommen jedoch nur als konjugiert komplexe Paare $z_{k/k+1} = a \pm ib$ mit $1 \leq k < p$ vor.

Ein AR(p)-Prozess ist stationär, wenn alle Wurzeln z_1, \dots, z_p außerhalb des komplexen Einheitskreises liegen.

Äquivalent kann wieder gefordert werden, dass alle inversen Wurzeln $\lambda_1, \dots, \lambda_p$ innerhalb des komplexen Einheitskreises liegen müssen. In der Praxis ist es oft ein Problem zu entscheiden, ob ein vorgelegter Prozess stationär ist oder nicht. Wir haben bereits beim AR(2)-Prozess gesehen, dass die Wurzeln in komplizierter Weise von den AR-Parametern abhängen, vgl. Gleichung (2.77) auf Seite 40. In praktischen Anwendungen sind die wahren Parameter jedoch in der Regel unbekannt und müssen geschätzt werden. Die Konsequenz ist, dass die Wurzeln des Prozesses ebenfalls nur geschätzt werden können. Es werden dann häufig sog. „unit root“ Tests verwendet, die auf statistischem Wege klären sollen, ob eine oder mehrere Wurzeln des Prozesses auf dem komplexen Einheitskreis liegen könnten.

2.5.2 Autokovarianz und Autokorrelation

Auch hier stellen AR(p)-Prozesse nur eine Verallgemeinerung der bisher vorgestellten AR(1)- und AR(2)-Prozesse dar. Autokovarianz und Autokorrelation hängen weiterhin ausschließlich vom Lag k ab. Die *Yule-Walker*-Gleichungen können ebenfalls analog durch Multiplikation mit y_{t-k} und anschließende Erwartungswertbildung hergeleitet werden, vgl. (2.91) auf Seite 46. Wir erhalten wiederum in Kurzform

$$\phi(B)\gamma_k = 0 \quad \text{bzw.} \quad \phi(B)\rho_k = 0. \quad (2.111)$$

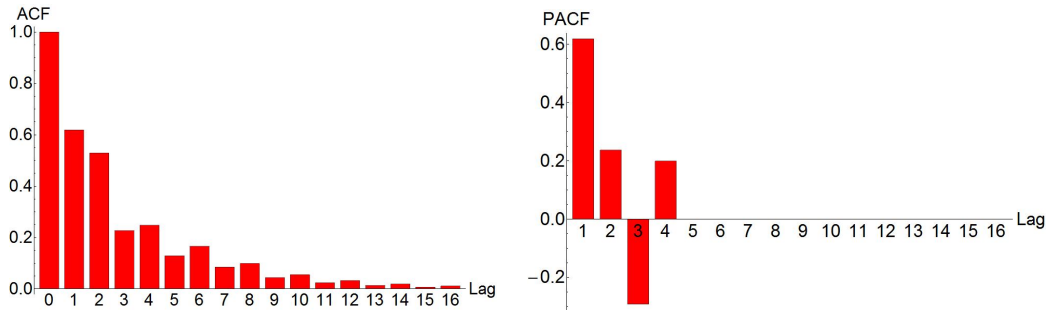


Abbildung 2.10: ACF und PACF eines AR(4)-Prozesses mit $\phi_1 = 0.6$, $\phi_2 = 0.3$, $\phi_3 = -0.4$ und $\phi_4 = 0.2$

Ein Ansatz für die explizite Berechnung der Autokorrelation kann wieder über eine Linearkombination der Wurzeln gebildet werden. Man erhält

$$\rho_k = \sum_{j=1}^p A_j \lambda_j^k, \quad (2.112)$$

mit den Koeffizienten $A_1, \dots, A_p \in \mathbb{R}$. Hier liegt erneut eine polynomiale Gleichung p -ten Grades mit unbekannten Koeffizienten vor. Die Lösung lässt sich daher nicht so einfach bestimmen wie im Fall des AR(2)-Prozesses.

Die partiellen Autokorrelationen werden wieder aus dem Gleichungssystem (2.24) auf Seite 17 gewonnen, wenn die Autokorrelationen vorliegen. Abbildung 2.10 zeigt exemplarisch die Autokorrelation (links) und die partielle Autokorrelation (rechts) für einen AR(4)-Prozess. Solche Prozesse werden häufig verwendet um saisonale Schwankungen in Quartalsdaten zu modellieren. Die grundlegenden Charakteristiken eines AR-Prozesses bleiben offensichtlich auch für Prozesse höherer Ordnung erhalten. Die partielle Autokorrelation fällt für Lags $k > p$ augenblicklich auf null ab. Die Autokorrelation selbst klingt exponentiell ab. Offenbar wird in diesem Fall das Abfallen der Autokorrelation mit einer Schwingung überlagert, was auf komplexe Wurzeln des Prozesses hindeutet.

2.5.3 Parameterschätzung

Die Parameter des AR(p)-Prozesses können wieder mit der Methode der kleinsten Quadrate geschätzt werden. Das Minimierungsproblem lautet nun

$$\min \sum_{t=p+1}^T \epsilon_t^2 = \min \sum_{t=p+1}^T (y_t - \theta_0 - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p})^2. \quad (2.113)$$

Für den Parametervektor $\boldsymbol{\theta} = (\theta_0, \phi_1, \dots, \phi_p)'$ gilt wieder der KQ-Schätzer (2.42) von Seite 27

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}, \quad (2.114)$$

wobei der Vektor \mathbf{y} und die Matrix \mathbf{X} durch

$$\mathbf{y} = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{p+1} \end{pmatrix} \quad \text{und} \quad \mathbf{X} = \begin{pmatrix} 1 & y_{t-1} & \dots & y_{t-p} \\ 1 & y_{t-2} & \dots & y_{t-p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & y_p & \dots & y_1 \end{pmatrix} \quad (2.115)$$

gegeben sind. Für AR-Modelle höherer Ordnung müssen also umfangreiche Matrizen berechnet und invertiert werden, was in der Praxis ebenfalls mit speziellen Computerprogrammen durchgeführt wird.

Die Schätzfunktionen für die stationären Momente, (2.29) bis (2.32) auf Seite 21, gelten auch weiterhin uneingeschränkt für AR(p)-Prozesse.

3

MA-Prozesse

Im vorangegangenen Kapitel wurden die Eigenschaften von AR-Prozessen ausführlich diskutiert. Bei der AR-Formulierung steht ein quasi lokaler Beschreibungsansatz im Vordergrund. Der aktuelle Systemzustand wird durch eine beschränkte Anzahl vergangener Systemzustände (beim $AR(p)$ -Prozess exakt durch p vergangene Zustände) erklärt, wobei noch ein aktueller Zufallseinfluss hinzu addiert wird. Es existiert jedoch noch ein alternativer Formulierungsansatz, der als global oder separiert charakterisiert werden könnte. Wir steigen in diese Thematik mit einem zentralen Theorem ein, das die Konstruktionsidee dieses Ansatzes sehr schön verdeutlicht.

Theorem 3.1: Wold-Zerlegung

Jeder kovarianzstationäre Prozess kann in eine lineare deterministische Komponente und eine rein stochastische Komponente zerlegt werden

$$x_t = \kappa_t + \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k},$$

mit

$$\psi_0 = 1 \quad \text{und} \quad \sum_{k=0}^{\infty} \psi_k^2 < \infty.$$

Der Term κ_t ist der rein deterministische Anteil, der sowohl aus einem konstanten Mittelwert als auch aus einem linearen, polynomialen, exponentiellen oder zyklischen Trend bestehen kann. Er ist für alle $k = 0, \dots, \infty$ von der Rauschkomponente ϵ_{t-k} unabhängig.

Die Separation zwischen systematischen (deterministischen) Einflüssen und reinem Zufallsrauschen wird hier global über die gesamte Historie des Prozesses vollzogen. Die Quadratsummenrestriktion für die Koeffizienten ψ_k ist

eine technische Bedingung, die gewährleisten soll, dass die Varianz des Prozesses endlich bleibt, vorausgesetzt die Varianz des Zufallsfehlers ϵ_t ist selbst endlich. Wir werden nun zeigen, dass die AR-Prozesse aus dem letzten Kapitel eine alternative MA-Darstellung besitzen, die durch die *Wold*-Zerlegung vermittelt wird.

3.1 AR(1)- und MA(∞)-Darstellung

Wir nehmen nun an, dass der lineare systematische Anteil κ_t der *Wold*-Zerlegung lediglich aus einem konstanten Mittelwert μ besteht. Wir erhalten dann gemäß Theorem 3.1 auf der vorherigen Seite

$$x_t - \mu = y_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}. \quad (3.1)$$

Weiterhin sei ein stationärer AR(1)-Prozess gegeben, der mit Hilfe seines Lag-Polynoms in der Form

$$(1 - \phi_1 B)y_t = \epsilon_t \quad (3.2)$$

geschrieben werden kann, vgl. (2.70) auf Seite 39. Da der AR(1)-Prozess stationär ist und daher $|\phi_1| < 1$ gilt, kann der Faktor $1 - \phi_1 B$ auf die rechte Seite gebracht und in eine geometrische Reihe entwickelt werden

$$y_t = \frac{1}{1 - \phi_1 B} \epsilon_t = (1 + \phi_1 B + \phi_1^2 B^2 + \dots) \epsilon_t = \sum_{k=0}^{\infty} \phi_1^k \epsilon_{t-k}. \quad (3.3)$$

Ein Vergleich der Koeffizienten in (3.1) und (3.3) zeigt, dass offenbar $\psi_k = \phi_1^k$ gilt. Der AR(1)-Prozess lässt sich also in eine *Wold*-Darstellung umschreiben. Allgemeiner vereinbart man die Moving Average (MA) Darstellung

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}. \quad (3.4)$$

Der AR(1)-Prozess besitzt also eine äquivalente MA(∞)-Darstellung mit $\theta_k = \phi_1^k$.

3.2 Der MA(1)-Prozess

MA-Prozesse stellen eine selbständige Prozessklasse dar, die auch losgelöst von der *Wold*-Zerlegung formuliert und analysiert werden kann. Das einfachste Mitglied dieser Klasse ist der MA(1)-Prozess, der ähnlich wie der AR(1)-Prozess in Operatorschreibweise notiert werden kann

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} = (1 + \theta_1 B) \epsilon_t = \theta(B) \epsilon_t. \quad (3.5)$$

Der Fehlerterm wird hier unabhängig und identisch verteilt angenommen, mit endlicher Varianz σ^2 . Der MA-Operator wird in allgemeiner Form analog zum AR-Operator (2.67) auf Seite 38 definiert als

$$\theta(B) = 1 + \sum_{k=1}^q \theta_k B^k, \quad (3.6)$$

was im Fall des MA(1)-Modells gerade zur Darstellung (3.5) führt.

Für einen beliebigen vorgelegten MA(1)-Prozess ergeben sich nun zwei interessante Fragen. Erstens, ist der MA(1)-Prozess stationär? Zweitens, besitzt er eine äquivalente AR-Darstellung? Wir werden die Frage nach der Stationarität hier auf die Kovarianzstationarität beschränken; zum einen, da wir den Fehlerterm in der Regel als normalverteilt annehmen, zum anderen wurde in der *Wold*-Zerlegung (Theorem 3.1 auf Seite 55) lediglich dieses Kriterium gefordert. Kovarianzstationarität ist gegeben, wenn die ersten beiden Momente sowie die Autokovarianzfunktion nicht von der Zeit, respektive lediglich vom Lag abhängen. Für den Erwartungswert gilt

$$E[y_t] = E[\epsilon_t + \theta_1 \epsilon_{t-1}] = E[\epsilon_t] + \theta_1 E[\epsilon_{t-1}] = 0. \quad (3.7)$$

Da die Fehler per Definition voneinander unabhängig sind, ergibt sich die Varianz als

$$\text{Var}[y_t] = \text{Var}[\epsilon_t] + \theta_1^2 \text{Var}[\epsilon_{t-1}] = (1 + \theta_1^2) \sigma^2. \quad (3.8)$$

Beide Momente (3.7) und (3.8) hängen damit nicht von der Zeit oder der Zeitverschiebung ab. Für die Kovarianzfunktion mit Lag 1 erhalten wir

$$\begin{aligned} \text{Cov}[y_t, y_{t-1}] &= E[y_t y_{t-1}] = E[\epsilon_t \epsilon_{t-1} + \theta_1 \epsilon_t \epsilon_{t-2} + \theta_1 \epsilon_{t-1}^2 + \theta_1^2 \epsilon_{t-1} \epsilon_{t-2}] \\ &= \theta_1 \text{Var}[\epsilon_{t-1}] = \theta_1 \sigma^2. \end{aligned} \quad (3.9)$$

Für Lags $k > 1$ ist unmittelbar aus (3.9) ersichtlich, dass die Kovarianz null ist, da keine zeitliche Überschneidung der Fehlerterme mehr zustande kommt. Die allgemeine Kovarianzfunktion ist deshalb

$$\gamma_k = \begin{cases} (1 + \theta_1^2) \sigma^2 & \text{für } k = 0 \\ \theta_1 \sigma^2 & \text{für } k = 1 \\ 0 & \text{für } k > 1. \end{cases} \quad (3.10)$$

Wir können zusammenfassen, dass alle Momente endlich sind und nicht von der Zeit t abhängen. Die Kovarianzfunktion γ_k hängt lediglich vom Lag k ab, daher ist der MA(1)-Prozess in jedem Fall stationär, und zwar unabhängig von der Wahl des Koeffizienten θ_1 .

Die zweite Frage war nun, ob der MA(1)-Prozess eine äquivalente AR-Darstellung besitzt. Es spricht einiges dafür, beispielsweise, dass ein stationärer AR(1)-Prozess eine äquivalente MA-Darstellung besitzt. Im umgekehrten Fall hatten wir den AR-Operator auf die andere Seite der Gleichung gebracht und in eine geometrische Reihe entwickelt. Der Gedanke liegt nahe dasselbe mit Gleichung (3.5) zu tun. Dazu muss aber gefordert werden, dass $|\theta_1| < 1$ gilt, da sonst die benötigte Reihenentwicklung nicht möglich ist. Im Fall des AR(1)-Prozesses war die Bedingung $|\phi_1| < 1$ automatisch erfüllt, da Stationarität gefordert wurde. Der MA(1)-Prozess ist hingegen immer stationär. Hier wird die notwendige Restriktion des MA-Koeffizienten als Invertierbarkeit bezeichnet. Ein MA(1)-Prozess besitzt also eine alternative AR-Darstellung, wenn er invertierbar ist. Die Berechnung dieser Form steht als Aufgabe am Ende dieses Abschnitts zur Verfügung. Wir wollen zunächst zeigen, dass diese Bedingung in der Praxis eine gewissermaßen schwächere Einschränkung darstellt als die Stationaritätsforderung an den AR(1)-Prozess.

Wir berechnen zunächst die Autokorrelation des MA(1)-Prozesses

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1}{1 + \theta_1^2}. \quad (3.11)$$

Gleichung (3.11) gilt für alle möglichen Werte des Parameters θ_1 . Weiterhin bleibt die Autokorrelationsfunktion unverändert, wenn an Stelle von θ_1 der Kehrwert $1/\theta_1$ eingesetzt wird. Diese Behauptung erscheint zunächst völlig kontraintuitiv. Nachrechnen zeigt jedoch

$$\rho_1 = \frac{\frac{1}{\theta_1}}{1 + \frac{1}{\theta_1^2}} = \frac{\theta_1^2 \cdot \frac{1}{\theta_1}}{\theta_1^2 \cdot \left(1 + \frac{1}{\theta_1^2}\right)} = \frac{\theta_1}{\theta_1^2 + 1}. \quad (3.12)$$

Das bedeutet, jeder nicht-invertierbare MA(1)-Prozess mit $|\theta_1| > 1$ besitzt ein Gegenstück oder einen Zwillingsprozess, der über dieselbe Autokorrelationsstruktur verfügt und invertierbar ist.

Aufgabe 3.1

Gegeben sei der MA(1)-Prozess

$$y_t = \theta(B)\epsilon_t = (1 + \theta_1 B)\epsilon_t,$$

mit

$$|\theta_1| < 1 \quad \text{und} \quad \epsilon_t \sim N(0, \sigma^2).$$

Leiten Sie die AR-Representation des MA(1)-Prozesses her.

3.2.1 Parameterschätzung

Die Parameterschätzung wird für MA-Modelle in der Regel mit der Maximum-Likelihood-Methode durchgeführt. Genauer wird meist die logarithmierte bedingte Likelihood Funktion maximiert. Wir werden hier das Prinzip exemplarisch Schritt für Schritt am MA(1)-Modell

$$y_t = \theta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} \tag{3.13}$$

nachvollziehen. Die Konstante θ_0 stellt hier den Mittelwert dar. Sie wurde eingefügt, um die Notation hinsichtlich der Parameterschätzung in AR-Modellen zu vereinheitlichen.

Ausgangspunkt für die ML-Schätzung ist die bedingte Dichte von y_t , vgl. (2.45) auf Seite 30. Hier muss jedoch auf ϵ_{t-1} bedingt werden, da eine MA-Spezifikation vorliegt. Bei normalverteilten Fehlertermen ergibt sich die bedingte Dichte

$$p(y_t | \epsilon_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(y_t - \theta_0 - \theta_1 \epsilon_{t-1})^2}{\sigma^2}}. \tag{3.14}$$

Die Frage lautet nun, wie kann ϵ_{t-1} ermittelt werden, wenn für gewöhnlich nur die Zeitreihe $(y_t)_{t=1, \dots, T}$ bekannt ist? Wir nehmen einmal an, der nullte Fehlerterm sei bekannt und setzen $\epsilon_0 = 0$. Wir erkennen sofort, dass daraus

$$\epsilon_1 = y_1 - \theta_0 \tag{3.15}$$

folgt. Damit ist ϵ_1 ebenfalls bekannt, und das selbe Argument kann rekursiv eingesetzt werden. Man erhält

$$\epsilon_t = y_t - \theta_0 - \theta_1 \epsilon_{t-1}. \tag{3.16}$$

KAPITEL 3. MA-PROZESSE

Iterieren von (3.16) und beachten der Anfangsbedingung $\epsilon_0 = 0$, führt zu einer expliziten Formel für ϵ_t ,

$$\epsilon_t = \sum_{k=0}^{t-1} (-1)^k \theta_1^k (y_{t-k} - \theta_0). \quad (3.17)$$

Theoretisch lassen sich die Fehlerterme ϵ_t bequem rekursiv berechnen, was können wir also aus der expliziten Formel lernen? Rufen wir uns zunächst noch einmal ins Gedächtnis, dass der MA(1)-Prozess in jedem Fall stationär ist, also auch für $|\theta_1| > 1$. Ist das der Fall, wirkt sich die Wahl der Anfangsbedingung mit zunehmender Zeit immer stärker auf ϵ_t aus. Um das zu erkennen machen wir uns klar, dass der letzte Term der Summe in (3.17) den Fehler ϵ_1 enthält. Einsetzen von (3.15) liefert für diesen finalen Term $(-1)^{t-1} \theta_1^{t-1} \epsilon_1$. Der Fehler ϵ_1 berechnet sich aber genaugenommen aus $\epsilon_1 = y_1 - \theta_0 - \theta_1 \epsilon_0$. Zusammenfassend schließt man also, dass der Fehler, der aufgrund der willkürlichen Festsetzung von ϵ_0 in die Berechnung von ϵ_t einfließt, mit θ_1^t potenziert wird. Ist $|\theta_1| < 1$, verliert dieser Fehler immer mehr an Gewicht und verschwindet für große t . Für $|\theta_1| > 1$ sind die Folgen jedoch fatal. Die bedingten Wahrscheinlichkeitsdichten werden zunehmend falsch und die vorgeschlagene ML-Schätzung ist nicht anwendbar.

Wir gehen daher im Folgenden davon aus, dass der vorgelegte MA-Prozess invertierbar ist, wodurch eben genau $|\theta_1| < 1$ sichergestellt wird. Die logarithmierte Likelihoodfunktion ergibt sich dann zu

$$l(\theta_0, \theta_1, \sigma^2) = \sum_{t=1}^T \log p(y_t | \epsilon_{t-1}) = -\frac{T}{2} \log[2\pi\sigma^2] - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2, \quad (3.18)$$

vgl. (2.46) auf Seite 30. Diese Funktion muss numerisch maximiert werden. Die dazu notwendigen Verfahren, wie *Newton-Raphson*-Algorithmus oder Scoring-Verfahren sind nicht Gegenstand dieser Vorlesung, sondern ein eigener Bereich der numerischen Mathematik.

Aufgabe 3.2

Gegeben sei die stationäre Zeitreihe

$$(y_t)_{t=1,\dots,6} = \{2, 3, 0, -2, 1, -2\}.$$

Nehmen Sie an, dass die Zeitreihe von dem MA(1)-Prozess

$$y_t = \epsilon_t + \theta \epsilon_{t-1},$$

mit $\epsilon_0 = 0$ erzeugt wurde. Geben Sie die logarithmierte Likelihoodfunktion an für

a) $\theta = \sigma^2 = \frac{1}{2},$

b) $\theta = \frac{1}{4}$ und $\sigma^2 = 1.$

3.3 MA(q)-Prozesse

MA(q)-Prozesse sind die natürliche Erweiterung des MA(1)-Prozesses, genau wie der AR(1)-Prozess im vorigen Kapitel auf AR(p)-Prozesse erweitert wurde. Sie notieren allgemein in Operatorschreibweise

$$y_t = \theta(B)\epsilon_t, \quad (3.19)$$

mit dem MA-Operator

$$\theta(B) = 1 + \sum_{k=1}^q \theta_k B^k. \quad (3.20)$$

Dies ist völlig analog zum AR-Operator (2.67) auf Seite 38. Wir werden in diesem Abschnitt einige allgemeine Eigenschaften von MA-Prozessen zusammentragen, die lediglich eine Generalisierung der MA(1)-Eigenschaften darstellen. Da MA-Prozesse ihre Hauptwirkung als Bestandteil der zusammengesetzten ARMA-Klasse entfalten, wird an dieser Stelle auf detaillierte Beweise verzichtet. Der interessierte Leser sei auf das Hauptskript und die darin zitierte Literatur verwiesen.

3.3.1 Autokovarianz und Autokorrelation

Analog zur Argumentation im MA(1)-Fall lässt sich die Varianz, respektive die Autokovarianz, durch Auswertung der Gleichung

$$\gamma_k = \text{Cov}[y_t, y_{t-k}] = E[y_t y_{t-k}] \quad (3.21)$$

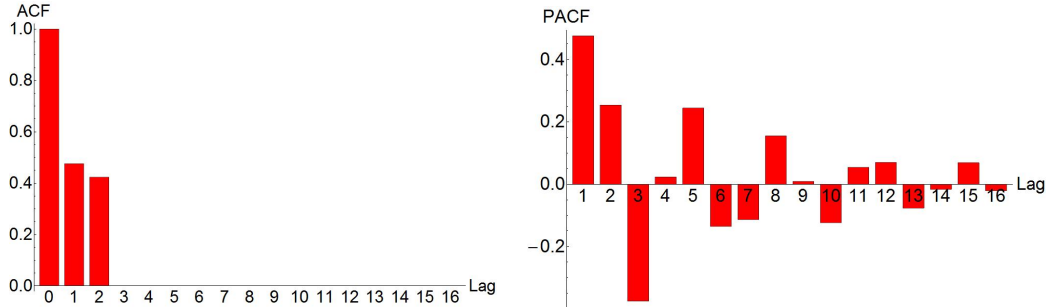


Abbildung 3.1: ACF und PACF des MA(2)-Prozesses mit $\theta_1 = 0.5$ und $\theta_2 = 0.8$

bestimmen, vgl. (3.8) und (3.9) auf Seite 57. Beachten Sie, dass $\text{Cov}[y_t, y_t] = \text{Var}[y_t]$ gilt. Man erhält explizit

$$\gamma_k = \begin{cases} \left(1 + \sum_{j=1}^q \theta_j^2\right) \sigma^2 & \text{für } k = 0 \\ \left(\theta_k + \sum_{j=1}^{q-k} \theta_{k+j} \theta_j\right) \sigma^2 & \text{für } 1 \leq k \leq q \\ 0 & \text{für } k > q. \end{cases} \quad (3.22)$$

Man überprüft leicht, dass sich mit Hilfe von (3.22) für $q = 1$ wieder die Formeln aus (3.10) für das MA(1)-Modell ergeben. Die Autokorrelation ergibt sich einfach durch Normierung mit γ_0

$$\rho_k = \begin{cases} 1 & \text{für } k = 0 \\ \frac{\theta_k + \theta_{k+1}\theta_1 + \dots + \theta_q\theta_{q-k}}{1 + \theta_1^2 + \dots + \theta_q^2} & \text{für } 1 \leq k \leq q \\ 0 & \text{für } k > q. \end{cases} \quad (3.23)$$

In (3.23) wurden lediglich die Summenzeichen aus (3.22) expandiert. Um die partielle Autokorrelation zu berechnen, müssen die Autokorrelationen nun wieder in das lineare Gleichungssystem (2.25) auf Seite 18 eingesetzt werden. Wir ersparen uns an dieser Stelle eine formale Demonstration und diskutieren stattdessen, welche Struktur die partielle Autokorrelationsfunktion eines MA-Prozesses aufweist.

Wir erinnern uns, dass der MA(1)-Prozess eine äquivalente, exponentiell abklingende AR-Representation besitzt, siehe Aufgabe 3.1 auf Seite 59. Ein solcher Prozess hat eine ebenfalls gegen unendlich abklingende partielle Autokorrelationsfunktion. Dies ist die schematische Struktur, die wir bei einem reinen MA-Modell erwarten dürfen. Abbildung 3.1 zeigt exemplarisch die ACF und PACF für einen MA(2)-Prozess. Es ist deutlich erkennbar, dass die Autokorrelation nach Lag 2 augenblicklich auf null abfällt, während

	AR(p)	MA(q)
ACF	Exponentiell abklingend	0 für Lags größer q
PACF	0 für Lags größer p	Exponentiell abklingend
Überführbar	Wenn stationär	Wenn invertierbar

Tabelle 3.1: Eigenschaften von AR(p)- und MA(q)-Prozessen

die partielle Autokorrelation exponentiell schwingend abklingt. AR- und MA-Prozesse besitzen hinsichtlich ihrer (partiellen) Autokorrelationsstruktur also genau entgegengesetzte Charakteristiken, die in Tabelle 3.1 zusammengefasst sind.

Aufgabe 3.3

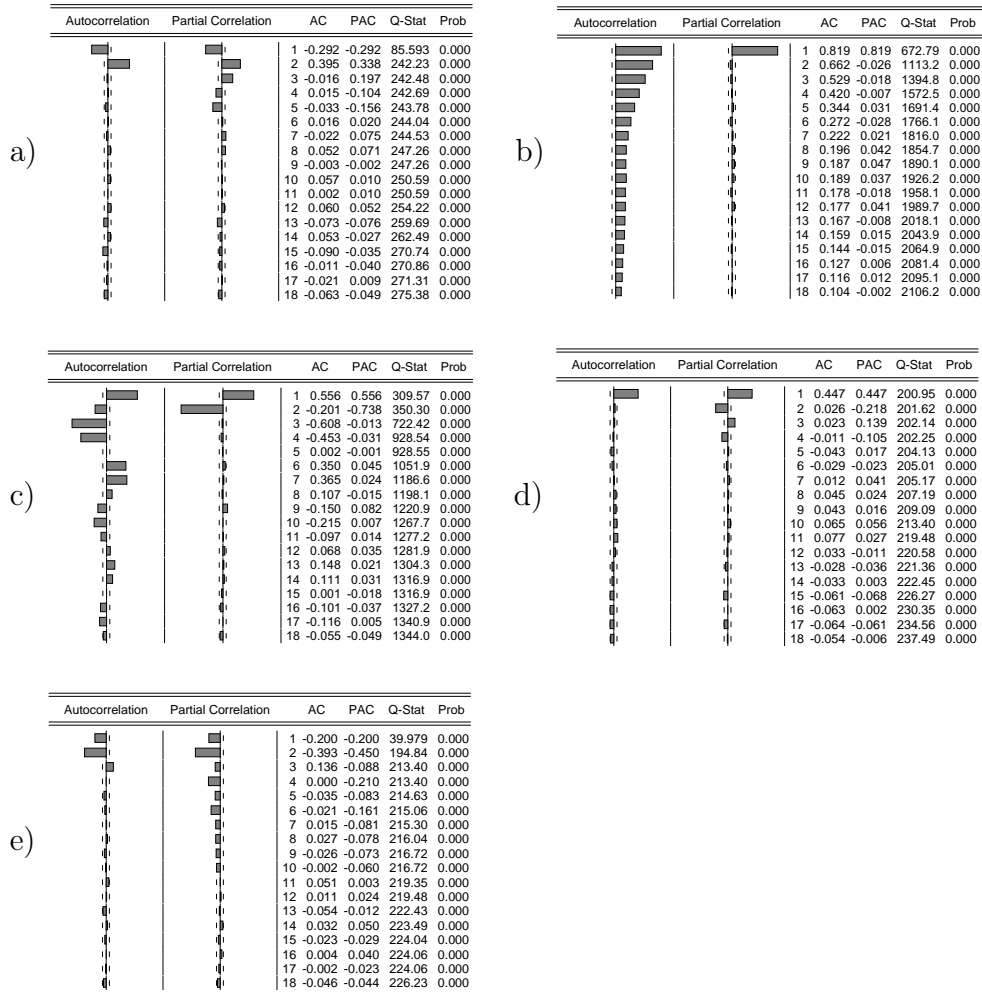
Bestimmen Sie die Autokovarianzfunktionen des MA(2)-Modells

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}.$$

Nehmen Sie an, dass ϵ_t unabhängig $N(0, \sigma^2)$ -verteilt ist.

Aufgabe 3.4

Die folgenden Ausdrücke enthalten mit der Software EViews geschätzte ACF und PACF verschiedener Prozesse. Ordnen Sie jeweils die Prozessklasse zu.



3.3.2 Invertierbarkeit

Um zu ermitteln, ob ein vorgelegter $MA(q)$ -Prozess invertierbar ist, wird genau wie für $AR(p)$ -Prozesse der zusammengesetzte Operator analysiert. Die Faktorisierung des charakteristischen Polynoms ergibt

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q = (1 - \lambda_1 z) \cdot \dots \cdot (1 - \lambda_q z), \quad (3.24)$$

mit den Nullstellen („Inverted Roots“) $\lambda_1, \dots, \lambda_q$. Für $z = \lambda^{-1}$ erhält man nach Multiplikation mit λ^q die gewohnte Form

$$\lambda^q + \theta_1 \lambda^{q-1} + \dots + \theta_q = (\lambda - \lambda_1) \cdot \dots \cdot (\lambda - \lambda_q). \quad (3.25)$$

In (3.25) wird deutlicher, dass die inversen Wurzeln $\lambda_1, \dots, \lambda_q$ tatsächlich die Nullstellen der Gleichung sind. Völlig analog zur Stationarität bei AR-Prozessen folgert man nun, dass der betreffende MA-Prozess invertierbar ist, wenn gilt $|\lambda_k| < 1$ für $k = 1, \dots, q$. Äquivalent lässt sich wieder für die Wurzeln z_1, \dots, z_q postulieren:

Ein MA(q)-Prozess ist invertierbar, wenn alle Wurzeln z_1, \dots, z_q außerhalb des komplexen Einheitskreises liegen.

Zur Äquivalenz der Darstellung mit Wurzeln und inversen Wurzeln vergleiche Abschnitt 2.5.1 auf Seite 51.

Wir haben am Beispiel eines nicht invertierbaren MA(1)-Prozesses gesehen, dass es einen zweiten („Zwillings-“) Prozess gibt, der dieselbe Autokorrelationsstruktur besitzt wie der ursprüngliche Prozess, aber invertierbar ist. Dieses Ergebnis lässt sich auf MA(q)-Prozesse verallgemeinern. Wir müssen dazu allerdings den MA-Operator unter Rückgriff auf die faktorierte Darstellung (3.24) schreiben

$$\theta(B) = \prod_{k=1}^q (1 - \lambda_k B). \quad (3.26)$$

Nehmen wir nun an, die inversen Wurzeln $\lambda_1, \dots, \lambda_{q'-1}$ liegen innerhalb des komplexen Einheitskreises und $\lambda_{q'}, \dots, \lambda_q$ liegen außerhalb. Dann lässt sich ein Prozess mit identischer Autokorrelationsstruktur definieren, der aber invertierbar ist. Diesen Prozess findet man, indem man den Kehrwert der Wurzeln einsetzt, die die Invertierbarkeitsbedingung verletzen. Man erhält also

$$y_t = \prod_{k=1}^{q'-1} (1 - \lambda_k B) \cdot \prod_{k=q'}^q \left(1 - \frac{1}{\lambda_k} B\right) \epsilon_t. \quad (3.27)$$

Der so definierte Prozess ist in jedem Fall invertierbar, weil alle seine Wurzeln innerhalb des komplexen Einheitskreises liegen.

Aufgabe 3.5

Schreiben Sie Gleichung (3.24) in kompakter Form unter Rückgriff auf das Summen- und Produktzeichen.

Aufgabe 3.6

Gegeben sei der MA(2)-Prozess

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}.$$

- a) Berechnen Sie in allgemeiner Form die inversen Wurzeln λ_1 und λ_2 .

Nehmen Sie die Parameterwerte $\theta_1 = -2$ und $\theta_2 = 0.75$ an.

- b) Bestimmen Sie die Wurzeln. Ist der Prozess invertierbar?
c) Ermitteln Sie die Parameter des Zwillingsprozesses.

3.3.3 Parameterschätzung

Die Parameterschätzung im MA(q)-Modell funktioniert völlig analog zum MA(1)-Fall. Wir gehen wieder davon aus, dass die ϵ_t unabhängig und identisch $N(0, \sigma^2)$ -verteilt sind. Es ergibt sich daher genau wie in (3.18) die bedingte logarithmierte Likelihoodfunktion

$$\begin{aligned} l(\theta_0, \dots, \theta_q, \sigma^2) &= \sum_{t=1}^T \log p(y_t | \epsilon_{t-1}, \dots, \epsilon_{t-q}) \\ &= -\frac{T}{2} \log[2\pi\sigma^2] - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2. \end{aligned} \quad (3.28)$$

Zur Erinnerung, der Parameter θ_0 repräsentiert den zeitunabhängigen Mittelwert. Anders formuliert, die LogLikelihood-Funktion bezieht sich auf das allgemeine Modell

$$y_t = \theta_0 + \theta(B)\epsilon_t. \quad (3.29)$$

Wir müssen wiederum eine Annahme über die Fehlerterme vor $t = 1$ machen, um rekursiv die Zeitreihe $(\epsilon_t)_{t=1, \dots, T}$ berechnen zu können. Die simpelste Annahme ist wieder

$$\epsilon_0 = \epsilon_{-1} = \dots = \epsilon_{-q+1} = 0. \quad (3.30)$$

Damit ergibt sich die Rekursion

$$\epsilon_t = y_t - \theta_0 - \sum_{k=1}^q \theta_k \epsilon_{t-k} \quad (3.31)$$

für alle $t = 1, \dots, T$. Beachten Sie, dass analog zur Problematik im MA(1)-Prozess die Berechnung der bedingten Likelihood lediglich Sinn

macht, wenn der vorgelegte Prozess invertierbar ist. Ansonsten potenzieren sich die Fehler aus der willkürlichen Festsetzung der ϵ_t für $t \leq 0$ über die Zeit.

Aufgabe 3.7

Gegeben sei die stationäre Zeitreihe

$$(y_t)_{t=1,\dots,6} = \{2, 3, 0, -2, 1, -2\}.$$

Nehmen Sie an, dass die Zeitreihe vom MA(2)-Prozess

$$y_t = 1 + \epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2},$$

mit $\epsilon_t \sim N(0, 1)$ erzeugt wurde. Berechnen Sie die bedingte logarithmierte Likelihoodfunktion mit $\epsilon_t = 0$ für $t \leq 0$.

Aufgabe 3.8

- a) Ist der MA(2)-Prozess aus Aufgabe 3.7 invertierbar?
- b) Ist die bedingte ML-Parameterschätzung in diesem Fall sinnvoll?

4

ARMA-Prozesse und Erweiterungen

Die Prozessklasse der ARMA-Modelle stellt eine natürliche Erweiterung der AR- bzw. MA-Modelle dar. Sie fasst die Komponenten der beiden Klassen in einem Modell zusammen. Das im Weiteren etwas genauer betrachtete ARMA(1,1)-Modell besitzt beispielsweise folgende Darstellung

$$y_t = \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}. \quad (4.1)$$

Mit Hilfe der in Abschnitt 2.4.1 eingeführten Operatorschreibweise und den entsprechenden AR- und MA-Operatoren (2.67) und (3.20) kann das ARMA(p, q)-Modell beliebiger Ordnung in ausgesprochen effizienter Weise notiert werden

$$\phi(B)y_t = \theta(B)\epsilon_t. \quad (4.2)$$

Die Schreibweise (4.2) ist zunächst rein formal und wird erst durch die Angabe der Modellordnung (p und q) konkret. Sie erweist sich jedoch bereits vorher als nützlich, da sie formal manipuliert werden kann. Um das zu sehen, überführen wir das ARMA(p, q)-Modell (4.2) formal in ein MA(∞)-Modell. Dies geschieht durch Invertieren des AR-Operators

$$y_t = \phi(B)^{-1} \theta(B) \epsilon_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}. \quad (4.3)$$

Das ist wieder die Form der *Wold-Zerlegung* (Theorem 3.1 auf Seite 55). Die Koeffizienten ψ_k können aus dem zusammengesetzten Operator $\phi(B)^{-1} \theta(B)$ berechnet werden. Dazu muss natürlich wiederum die Modellordnung bekannt sein. Durch Invertieren des MA-Operators hätte der ARMA-Prozess natürlich auch in einen AR(∞)-Prozess überführt werden können, was den Nutzen solcher formaler Manipulationen verdeutlicht. Das bis hierher gesagte gilt natürlich nur unter der Prämisse, dass der betreffende Prozess stationär, respektive invertierbar ist.

4.1 Der ARMA(1,1)-Prozess

Der ARMA(1,1)-Prozess ist eine der nützlichsten Modellvarianten in der Zeitreihenanalyse. Die Ursache hierfür ist die Sparsamkeit der Parametrisierung. Wir haben schon gesehen, dass ein ARMA-Modell in ein AR(∞)- oder auch in ein MA(∞)-Modell überführt werden kann, was die besondere Flexibilität dieser Mischprozesse unterstreicht. Der ARMA(1,1)-Prozess ist aber der sparsamste von allen, da er lediglich einen AR- und einen MA-Parameter benötigt. Wir werden im Folgenden einige Eigenschaften des ARMA(1,1)-Prozesses herausarbeiten. Die Verallgemeinerung auf den ARMA(p, q)-Fall ist in der Regel aufwendig und wird mit Computerunterstützung durchgeführt. Daher werden wir uns darauf beschränken, einige Parallelen zu dieser allgemeinen Prozessklasse aufzuzeigen.

Unser Ausgangspunkt ist das ARMA(1,1)-Modell (4.2) auf der vorherigen Seite, mit den entsprechenden Operatoren

$$\phi(B) = 1 - \phi B \quad \text{und} \quad \theta(B) = 1 + \theta B. \quad (4.4)$$

Wir überführen dieses Modell zunächst in die MA(∞)-Form, indem wir den AR-Operator invertieren

$$\begin{aligned} y_t &= \frac{1 + \theta B}{1 - \phi B} \epsilon_t \\ &= (1 + \phi B + \phi^2 B^2 + \dots)(1 + \theta B) \epsilon_t \\ &= (1 + (\phi + \theta)(B + \phi B^2 + \phi^2 B^3 + \dots)) \epsilon_t. \end{aligned} \quad (4.5)$$

In der zweiten Zeile wurde wieder die Eigenschaft der geometrischen Reihe ausgenutzt, vgl. Seite 9. Wir gewinnen mit (4.5) also eine *Wold*-Darstellung

$$y_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}, \quad (4.6)$$

bei der die Koeffizienten durch

$$\psi_0 = 1 \quad \text{und} \quad \psi_k = (\phi + \theta)\phi^{k-1} \quad (4.7)$$

für $k \geq 1$ gegeben sind.

4.1. DER ARMA(1,1)-PROZESS

Wir berechnen nun die Autokovarianzen für den ARMA(1,1)-Prozess. Für Lag $k = 0$ erhalten wir aus (3.22) auf Seite 62 in Verbindung mit den Wold-Koeffizienten (4.7)

$$\begin{aligned}\gamma_0 &= \sigma^2 \left(1 + \sum_{j=1}^{\infty} \psi_j^2 \right) = \sigma^2 \left(1 + (\phi + \theta)^2 \sum_{j=0}^{\infty} \phi^{2j} \right) \\ &= \sigma^2 \left(1 + \frac{(\phi + \theta)^2}{1 - \phi^2} \right) = \sigma^2 \frac{1 - \phi^2 + (\phi + \theta)^2}{1 - \phi^2} \\ &= \sigma^2 \frac{1 + 2\phi\theta + \theta^2}{1 - \phi^2}.\end{aligned}\tag{4.8}$$

Offensichtlich ist die Varianz des Prozesses bereits eine nicht triviale Funktion der Parameter. Die Komplexität der Ausdrücke steigt mit höherer Modellordnung beträchtlich. Die Autokovarianz mit Lag $k = 1$ ist gegeben durch

$$\gamma_1 = \phi\gamma_0 + \theta\sigma^2.\tag{4.9}$$

Die Herleitung dieser Beziehung wird dem Leser als Übungsaufgabe am Ende dieses Abschnitts überlassen. Eine interessante Eigenschaft der Kovarianzfunktion (4.9) ist, dass sie strukturell die Summe aus den Autokovarianzen des AR(1)-Prozesses und des MA(1)-Prozesses darstellt. Beachte: Diese Struktureigenschaft gilt nicht für γ_0 ! Für größere Lags $k \geq 2$ überschneiden sich die Fehlerterme des MA-Teils nicht mehr, wodurch unmittelbar klar ist, dass der AR(1)-Teil den einzigen Beitrag zur Autokovarianz liefern muss. Daher gilt insgesamt

$$\gamma_k = \begin{cases} \sigma^2 \frac{1+2\phi\theta+\theta^2}{1-\phi^2} & \text{für } k = 0 \\ \phi\gamma_0 + \theta\sigma^2 & \text{für } k = 1 \\ \phi\gamma_{k-1} & \text{für } k \geq 2. \end{cases}\tag{4.10}$$

Ganz allgemein gilt für ARMA(p, q)-Prozesse, dass die Autokovarianz für Lags $k > q$ nur noch vom AR-Teil des Prozesses gespeist wird. Die Autokorrelation ergibt sich wieder aus γ_k/γ_0 . ARMA-Modelle besitzen aber eine zusätzliche Komplikation. Es stellt sich heraus, dass die partielle Autokorrelation ebenfalls (betragsmäßig) langsam abfällt, wodurch eine Modellidentifikation anhand der ACF und PACF in der Regel nicht möglich ist. Abbildung 4.1 zeigt die ACF und PACF für den ARMA(1,1)-Prozess mit $\phi = 0.7$ und $\theta = 0.8$. Wir kommen zum Problem der Modellidentifikation zu einem späteren Zeitpunkt zurück. Zunächst wollen wir die wichtige Frage nach Stationarität bzw. Invertierbarkeit diskutieren.

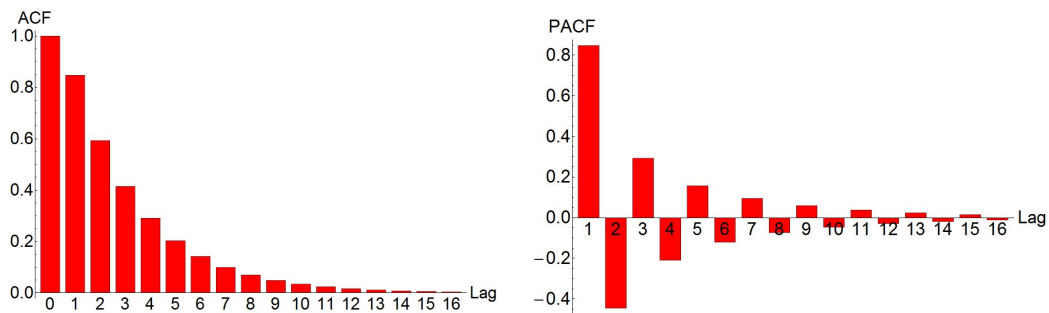


Abbildung 4.1: ACF und PACF des ARMA(1,1)-Prozesses mit $\phi = 0.7$ und $\theta = 0.8$

Wir haben im Fall der MA-Prozesse bereits gesehen, dass jeder nicht-invertierbare Prozess ein Gegenstück besitzt, das über dieselbe Autokorrelationsstruktur verfügt wie der ursprüngliche Prozess, aber die geforderten Bedingungen für Invertierbarkeit erfüllt. Diese Eigenschaft erstreckt sich auch auf den MA-Teil des ARMA-Prozesses. Invertierbarkeit scheint also eine eher unproblematische Eigenschaft zu sein. Bezüglich der Stationarität erwartet uns eine angenehme Überraschung. Wir wissen, dass ein MA-Prozess immer stationär ist. Wir haben weiterhin das ARMA(1,1)-Modell in (4.6) als MA(∞)-Modell dargestellt. Die Bedingung für Stationarität ist also dieselbe, unter der der ARMA(1,1)-Prozess in MA-Form überführt werden kann. Eine kurze Inspektion von (4.5) zeigt, dass die kritische Bedingung $|\phi| < 1$ ist, da sonst die geometrische Reihenentwicklung unzulässig wäre. Das entspricht aber genau der Stationaritätsbedingung des AR(1)-Modells. Es stellt sich heraus, dass diese Eigenschaft auch für ARMA(p, q)-Modelle gilt:

Ein ARMA(p, q)-Prozess ist genau dann stationär, wenn sein AR-Teil die Stationaritätsbedingung erfüllt.

Das bedeutet konkret, die Wurzeln des AR-Teils müssen außerhalb des komplexen Einheitskreises liegen, oder anders herum, die inversen Wurzeln (inverted roots) müssen innerhalb des komplexen Einheitskreises liegen.

Aufgabe 4.1

Zeigen Sie, dass die Autokovarianzfunktion mit Lag 1 im ARMA(1,1)-Modell durch

$$\gamma_1 = \phi\gamma_0 + \theta\sigma^2$$

gegeben ist.

Aufgabe 4.2

Zeigen Sie, dass ein ARMA(1,1)-Prozess mit $\theta = -\phi$ ein einfacher Rauschprozess ist. Hinweis: Verwenden Sie die Operatorschreibweise und setzen Sie die korrekten Operatoren ein.

4.1.1 Parameterschätzung

Ähnlich wie im Fall von MA-Prozessen wird die Parameterschätzung in ARMA-Modellen im einfachsten Fall mit Hilfe der bedingten Likelihood-Funktion abgewickelt¹. Auch hier gilt, AR- und MA-Teile sollten stationär und invertierbar sein, da sich sonst Fehler, die bei der Festsetzung der Anfangsbedingungen gemacht werden, potenzieren. Angenommen, der Fehlerterm ϵ_t sei $N(0, \sigma^2)$ -verteilt, dann kann die bedingte logarithmierte Likelihoodfunktion als

$$l(\phi_1, \theta_0, \theta_1, \sigma^2) = \sum_{t=1}^T \log p(y_t | y_{t-1}, \epsilon_{t-1}) = -\frac{T}{2} \log[2\pi\sigma^2] - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2 \quad (4.11)$$

geschrieben werden, wobei der Fehlerterm durch

$$\epsilon_t = y_t - \phi_1 y_{t-1} - \theta_0 - \theta_1 \epsilon_{t-1} \quad (4.12)$$

gegeben ist. (4.11) bzw. (4.12) ist wieder eine allgemeine Form, d.h. sie berücksichtigt einen gegebenenfalls vorhandenen Intercept θ_0 . Die Anfangsbedingungen y_0 und ϵ_0 sind in der Regel nicht verfügbar und können durch ihre Erwartungswerte ersetzt werden. Zu Referenzzwecken soll auch die bedingte LogLikelihood-Funktion für ARMA(p, q)-Modelle angegeben werden:

$$\begin{aligned} l(\phi_1, \dots, \phi_p, \theta_0, \dots, \theta_q, \sigma^2) &= \sum_{t=1}^T \log p(y_t | y_{t-1}, \dots, y_{t-p}, \epsilon_{t-1}, \dots, \epsilon_{t-q}) \\ &= -\frac{T}{2} \log[2\pi\sigma^2] - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2, \end{aligned} \quad (4.13)$$

¹Moderne Programmpakete wie EViews wählen zunächst eine AR(∞)-Darstellung. Diese Darstellung wird dann durch ein AR-Modell hoher Ordnung approximiert, dessen Parameter mit der Kleinste-Quadrate-Methode geschätzt werden (Long-AR-Estimator). Anschließend werden die Residuen berechnet, die ihrerseits Schätzer für die Zufallsfehler sind und so die Kleinste-Quadrate-Schätzung der ursprünglichen ARMA-Parameter erlauben. Die resultierenden Schätzer werden dann als Startlösung für die iterative Maximum-Likelihood-Schätzung verwendet.

mit

$$\epsilon_t = y_t - \sum_{j=1}^p \phi_j y_{t-j} - \theta_0 - \sum_{k=1}^q \theta_k \epsilon_{t-k}. \quad (4.14)$$

Aufgabe 4.3

Gegeben sei die stationäre Zeitreihe

$$(y_t)_{t=1,\dots,6} = \{2, 3, 1, -1, -2, 0.5\}.$$

Nehmen Sie an, dass die Zeitreihe von dem ARMA(1,1)-Prozess

$$y_t = \frac{1}{2}y_{t-1} + \epsilon_t + \frac{1}{2}\epsilon_{t-1},$$

mit $\epsilon_t \sim N(0, 1)$ erzeugt wurde. Verwenden Sie die Anfangsbedingung $y_0 = \epsilon_0 = 0$ und berechnen Sie die logarithmierte Likelihoodfunktion.

4.2 ARIMA und SARIMA

Wir werden in diesem Abschnitt einige Erweiterungen des ARMA-Modells kennenlernen, die eine Anpassung an spezielle Eigenschaften von Zeitreihen erlauben. Das ARIMA-Modell ist so konzipiert, dass auch nicht stationäre Zeitreihen einer Modellanalyse im ARMA-Kontext unterworfen werden können. SARIMA-Modelle adressieren zusätzlich noch zyklische Eigenschaften, wie sie im Bereich volkswirtschaftlicher Modelle häufig zu finden sind.

4.2.1 ARIMA(p, d, q)-Modelle

Die Idee, die hinter ARIMA-Modellen steckt, ist sehr einfach. Sollte eine Zeitreihe vorliegen, die offensichtlich nicht stationär ist, bspw. das Brutto-Inlandsprodukt (BIP) seit dem 01.01.2000 (Abbildung 4.3 rechts auf Seite 78), so lässt sich diese Zeitreihe möglicherweise trotzdem mit den bekannten Mitteln der Zeitreihenanalyse behandeln. Der Schlüssel lautet hier: „Transformation des Prozesses“. Genauer gesagt bildet man die Differenz zwischen jeweils zwei aufeinander folgenden Elementen der Zeitreihe. In unserem Beispiel hieße das, wir berechnen die BIP-Zuwächse. Ein solcher Zuwachs-Prozess wird aller Wahrscheinlichkeit nach stationär sein, anderenfalls wären wir bereits unermesslich reich. Formal betrachtet heißt das nichts

anderes, als dass wir ein Zeitreihenmodell für $x_t = y_t - y_{t-1} = \nabla y_t$ bilden. Zur Definition des Nabla-Operators vgl. Abschnitt 2.4.1, insbesondere (2.65) auf Seite 38.

Ganz allgemein ist diese Strategie natürlich nicht auf Differenzen erster Ordnung beschränkt, vielmehr können d -fache Differenzen gebildet werden, soviele wie nötig sind, um den Prozess stationär zu machen. Im besten Fall resultiert ein einfacher Rauschprozess und das entsprechende Modell kann als

$$x_t = \nabla^d y_t = \epsilon_t \quad (4.15)$$

geschrieben werden. Der ursprüngliche Prozess kann durch Aufsummieren der gebildeten Differenzen zurückgewonnen werden. Der Summation im diskreten Fall entspricht die Integration im stetigen Fall, daher werden Prozesse wie (4.15) d -fach integrierte Prozesse genannt. Der nächste logische Schritt ist, einen Differenzenprozess, der ja stationär ist, als ARMA-Prozess zu modellieren. Man gewinnt so die ARIMA(p, d, q)-Form

$$\phi(B)\nabla^d y_t = \theta(B)\epsilon_t. \quad (4.16)$$

Zwei Sätze zur Vertiefung sind hier angebracht. Zunächst steht die Abkürzung ARIMA, wie zu vermuten war, für AutoRegressive Integrated Moving Average. Darüber hinaus ist die Anordnung dieser Abkürzungen, ebenso wie die Reihenfolge der Modellordnungsparameter (p, d, q) nicht zufällig. (4.16) zeigt, dass zuerst der AR-Operator, dann der Differenzen-Operator und am Ende der MA-Operator steht. Die Modellbezeichnung legt also eindeutig die Reihenfolge der Modellelemente fest.

Abbildung 4.2 links zeigt die Zeitreihe der logarithmierten DAX-Kurse vom 01.01.1991 bis zum 31.10.2009. Diese Zeitreihe ist augenscheinlich

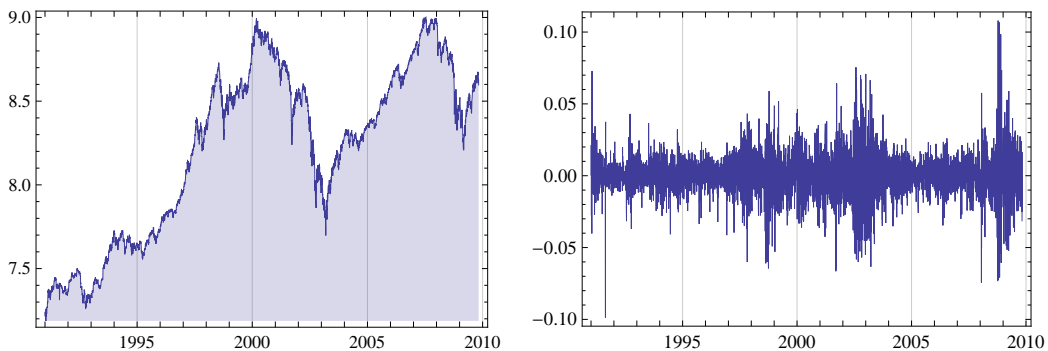


Abbildung 4.2: Logarithmierte DAX-Zeitreihe und Differenzen erster Ordnung

nicht stationär, wohingegen die Differenzen erster Ordnung (rechts) einen Rauschprozess bilden, dessen Streuung sich zwar offenbar über die Zeit ändert, der aber auf jeden Fall stationär ist. Wir werden zu einem späteren Zeitpunkt im Rahmen von ARCH- und GARCH-Modellen lernen, wie solche zeitvariablen Streuungen zu behandeln sind. Im Augenblick geht es darum, dass der ursprünglich instationäre Prozess durch Differenzbildung einer Zeitreihenanalyse zugänglich gemacht wurde.

Wir wollen einen Augenblick beim Beispiel von Aktienkursen bleiben, um das bisher gesagte noch etwas genauer theoretisch zu untermauern. Entgegen gewöhnlichen AR-Prozessen hängt die Dynamik von Aktienkursen, bzw. deren Logarithmus, offenbar nicht vom Niveau des Kurses selbst ab. Nehmen wir im einfachsten Fall an, dass ein AR(1)-Prozess vorliegt, mit $\phi(B) = 1 - \phi B$, dann lässt sich diese Eigenschaft formal durch

$$\phi(B)(y_t + c) = \phi(B)y_t \quad (4.17)$$

ausdrücken, mit einer beliebigen Konstante $c \neq 0$. Daraus folgt sofort für den invariablen Teil der Gleichung

$$\phi(B)c = \phi(1)c = (1 - \phi)c = 0. \quad (4.18)$$

Da $c \neq 0$ muss $\phi = 1$ gelten, damit (4.18) erfüllt ist. Folglich hat die charakteristische Gleichung $\phi(z) = 0$ die Lösung $z = 1$, was zeigt, dass der Prozess eine Einheitswurzel besitzt. Der Differenzenprozess $x_t = \nabla y_t$ muss daher ein stationärer Rauschprozess sein.

Wir zeigen nun noch, dass die formale Rücktransformation den integrierten (aufsummierten) Prozess liefert

$$\begin{aligned} y_t &= \nabla^{-1}x_t = (1 - B)^{-1}x_t \\ &= \sum_{k=0}^{\infty} B^k x_t = \sum_{k=0}^{\infty} x_{t-k}, \end{aligned} \quad (4.19)$$

wie gefordert.

Aufgabe 4.4

Schreiben Sie folgende Modelle aus:

- a) ARIMA(1,1,0)
- b) ARIMA(0,1,2)
- c) ARIMA(1,2,1).

Aufgabe 4.5

Gegeben sei die Zeitreihe

$$(y_t)_{t=1,\dots,6} = \{1, 3, 2, 4, 2, 0\}.$$

Gehen Sie davon aus, dass die Zeitreihe durch das ARIMA(0,1,1)-Modell

$$\nabla y_t = \epsilon_t + \epsilon_{t-1},$$

mit $\theta_0 = 0$ und $\epsilon_t \sim N(0, 1)$ beschrieben werden kann. Verwenden Sie die Anfangsbedingung $y_0 = \epsilon_0 = 0$ und berechnen Sie die logarithmierte Likelihoodfunktion.

4.2.2 Seasonal ARIMA-Modelle

Seasonal ARIMA- oder SARIMA-Modelle sind eine natürliche Erweiterung der ARIMA-Modelle, um möglichst effizient saisonale Einflüsse behandeln zu können. Sie spielen in der Makroökonomie und der volkswirtschaftlichen Gesamtrechnung eine große Rolle und sollen aus diesem Grund hier besprochen werden. Da SARIMA-Modelle im Modul Zeitreihenanalyse nicht behandelt werden, enthält dieser Abschnitt auch keine prüfungsrelevante Information und kann übersprungen werden.

Da ARIMA-Modelle ihrerseits eine Erweiterung des ARMA-Konzepts darstellen, soll die Idee der saisonalen Modellierung zuerst anhand von ARMA-Modellen erklärt, und anschließend auf ARIMA erweitert werden. Man betrachte einen Prozess, der nur von verzögerten Werten abhängt, deren Lag ein ganzzahliges Vielfaches s einer Basisperiode ist. Abbildung 4.3 zeigt zwei Zeitreihen mit saisonalen Effekten. Links abgebildet sind monatliche Passagierdaten einer Luftfahrtgesellschaft von Januar 1949 bis Dezember 1960. Hier würde man von einer monatlichen Basisperiode ausgehen und daher $s = 12$ wählen. Auf der rechten Seite ist das Brutto-Inlandsprodukt (BIP) der Bundesrepublik Deutschland von 2000 bis 2008 gegeben. Da die BIP-Daten im Rahmen der volkswirtschaftlichen Gesamtrechnung quartalsmäßig erfasst werden, wäre hier $s = 4$ zu wählen. Ein ARMA-Modell über diese saisonalen Effekte würde dann folgendermaßen spezifiziert werden

$$\Phi(B^s)y_t = \Theta(B^s)\epsilon_t, \quad (4.20)$$

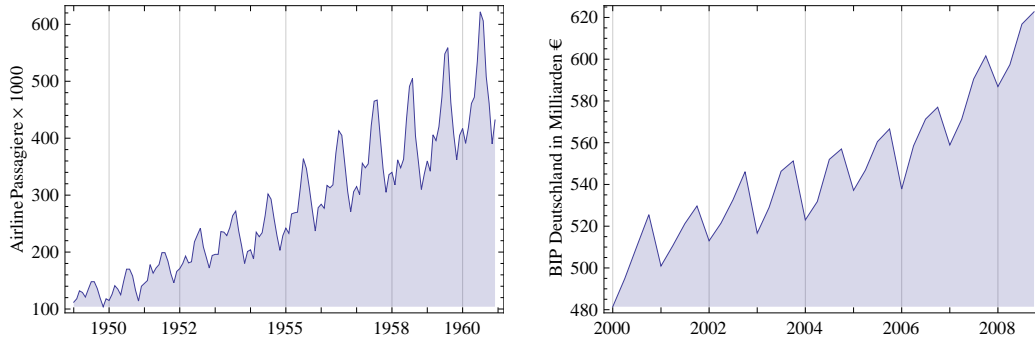


Abbildung 4.3: Airline Passagiere und Brutto-Inlandsprodukt

mit den saisonalen AR- und MA-Operatoren

$$\Phi(B^s) = 1 - \sum_{j=1}^P \Phi_j B^{js} \quad \text{und} \quad \Theta(B^s) = 1 + \sum_{k=1}^Q \Theta_k B^{ks}. \quad (4.21)$$

Ein solches Modell ist natürlich nicht besonders realistisch, da hier genaue genommen s verschiedene, entkoppelte Zeitreihen modelliert werden, die keine gemeinsame Korrelation aufweisen. Man würde aber erwarten, dass auch zwischen den einzelnen Basisperioden zumindest eine schwache Abhängigkeit besteht. Diese Annahme lässt sich durch Kombination von saisonalen und regulären Operatoren in das Modell integrieren und man erhält

$$\Phi(B^s)\phi(B)y_t = \Theta(B^s)\theta(B)\epsilon_t. \quad (4.22)$$

Die Zeitreihen in Abbildung 4.3 weisen weiterhin einen eindeutigen Wachstumstrend auf. Es wird also notwendig, integrierte Prozesse $x_t = \nabla^d y_t$ zu betrachten. Da wir es aber zusätzlich mit saisonalen Operatoren zu tun haben, definieren wir zunächst den entsprechenden saisonalen Differenzenoperator

$$\nabla_s^D = (1 - B^s)^D. \quad (4.23)$$

Damit kann das Konzept der Saisonalität einfach auf seine allgemeinste Form erweitert werden und man erhält das SARIMA(p, d, q)(P, D, Q) $_s$ -Modell

$$\Phi(B^s)\nabla_s^D\phi(B)\nabla^d y_t = \Theta(B^s)\theta(B)\epsilon_t. \quad (4.24)$$

4.3 ARMAX-Modelle

Wie der Zusatz „X“ im Modelltyp signalisiert, gestatten ARMAX-Modelle die Berücksichtigung exogener Einflüsse oder Kontrollvariablen. Solche Einflüsse sind in der Volkswirtschaftslehre, insbesondere in der Makroökonomie,

von besonderem Interesse. Beispielsweise stellen Einflussgrößen wie nominale Geldmenge oder Staatsausgaben in traditionellen Makromodellen oft exogene Größen dar, die vorgegeben werden um die Gleichgewichtseigenschaften von Geld-, respektive Gütermarkt unter diesen Vorgaben zu untersuchen. ARMAX-Modelle übersetzen diesen Gedanken in eine dynamische Umgebung. Wir werden sehen, dass diese Aufgabe sehr elegant und einfach durch die ARMAX-Spezifikation gelöst wird, wodurch Mehrperiodenmodelle effizient formuliert werden können. Darüber hinaus besitzt die ARMAX-Klasse eine enge Anbindung zu Regressionsmodellen mit autokorrelierten Fehlern und sogenannten Distributed-Lag-Modellen. An diesem Punkt soll jedoch die eigentliche Aufgabe von ARMAX-Modellen herausgearbeitet werden, nämlich die Behandlung von exogenen Einflüssen.

4.3.1 Das ARMAX(1,0,1)-Modell

Das ARMAX(1,0,1)-Modell ist das einfachste ARMA-Modell, dass exogene Einflüsse abbilden kann. Es kann in ausgeschriebener Form wie folgt notiert werden

$$y_t = \phi_1 y_{t-1} + \chi_0 x_t + \epsilon_t + \theta_1 \epsilon_{t-1}. \quad (4.25)$$

Im Gegensatz zum normalen ARMA(1,1)-Modell kommt hier noch der exogene Einfluss $\chi_0 x_t$ hinzu, der ebenfalls einen Zeitindex trägt. Die Koeffizienten sind hier etwas suggestiv mit Indizes versehen worden, um einen wichtigen Punkt aufzuzeigen. Wir haben im vorangegangenen Abschnitt gesehen, dass ein ARIMA(1,0,1)-Modell einem ARMA(1,1)-Modell entspricht. Das gilt für die ARMAX-Klasse nicht! Ganz allgemein gilt:

Die ARMAX-Ordnung $(p,0,q)$ zeigt an, dass die exogene Einflussgröße unverzögert, das heißt mit Lag 0, in die Modellgleichung einfließt.

Um die Wichtigkeit von Modellen der ARMAX-Klasse zu verdeutlichen, sind in Abbildung 4.4 verschiedene Anwendungsbeispiele gegeben, die in den Wirtschaftswissenschaften besonders relevant sind. Es wurde hier ein ARMAX(1,0,0)- oder verkürzt ARX(1,0)-Modell verwendet, in dem der MA-Teil entfällt. Im ersten Fall, Abbildung 4.4 links oben, wurde mit Hilfe des exogenen Modellteils ein Schock zum Zeitpunkt $t = 100$ herbeigeführt. Modelltechnisch bedeutet das für die exogene Variable

$$x_t = \begin{cases} 1 & \text{für } t = 100 \\ 0 & \text{sonst.} \end{cases} \quad (4.26)$$

KAPITEL 4. ARMA-PROZESSE UND ERWEITERUNGEN

Damit wird zum Zeitpunkt $t = 100$ ein Schock in Höhe des Koeffizienten χ_0 induziert, dessen Resorption im Rahmen des Modells untersucht werden kann. Eine weitere häufige Problemstellung wird durch Niveauverschiebungen gegeben. Solche Verschiebungen treten zumeist ein, wenn sich gewisse Rahmenbedingungen geändert haben. In Abbildung 4.4 rechts oben ist eine solche Niveauverschiebung dargestellt. Sie wurde durch die Vorschrift

$$x_t = \begin{cases} 0 & \text{für } t < 100 \\ 1 & \text{für } t \geq 100 \end{cases} \quad (4.27)$$

herbeigeführt. Als letztes Beispiel wurde ein deterministischer Wachstumstrend in den Prozess eingeschleust. Dies geschieht einfach durch die Beziehung

$$x_t = t. \quad (4.28)$$

Die Steigung des Wachstumstrends wird dann durch den ARMAX-Koeffizienten χ_0 kontrolliert. Das Ergebnis ist in Abbildung 4.4 links unten zu sehen.

Wir können bereits an dieser Stelle erkennen, wie extrem nützlich ARMAX-Modelle in Bereichen wie Makroökonomie, Fiskal- oder Stabilitäts-

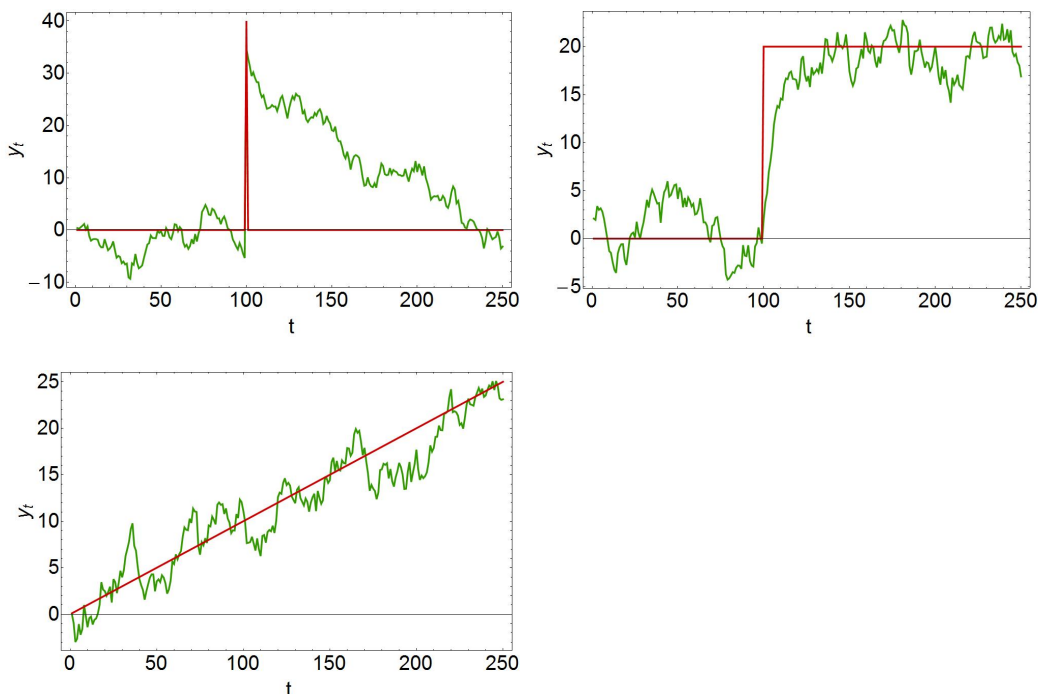


Abbildung 4.4: Verschiedene Applikationen des ARX(1,0)-Modells

politik sein können, kurzum überall, wo dynamische Reaktionen auf äußere Einflüsse abgebildet werden sollen.

Aufgabe 4.6

Nehmen Sie an, die Gleichungsfehler seien $N(0, \sigma^2)$ -verteilt. Schreiben Sie die bedingte LogLikelihood-Funktion für das ARMAX(1,0,1)-Modell in allgemeiner Form (inklusive Intercept θ_0).

4.3.2 ARMAX(p, s, q)-Modelle

Natürlich lassen sich auch ARMAX-Modelle in allgemeiner und kompakter Form mit Hilfe von Operatoren schreiben. Man notiert

$$\phi(B)y_t = \chi(B)x_t + \theta(B)\epsilon_t, \quad (4.29)$$

mit dem neuen Operator

$$\chi(B) = \sum_{k=0}^s \chi_k B^k \quad (4.30)$$

für die exogenen Kontrollgrößen. Beachten Sie, dass der Summationsindex in (4.30) bei null beginnt, d.h. der erste äußere Einfluss wirkt unverzögert, also mit Lag 0. Die Generalisierung (4.29) ist nun keine Überraschung mehr, wir haben das Prinzip bereits bei vorangegangenen Modellklassen gesehen. Wir werden den Rest des Abschnitts darauf verwenden, die Zusammenhänge der ARMAX-Klasse mit anderen etablierten und eventuell bereits bekannten Modelltypen aufzuzeigen, um einen Eindruck von der Vielseitigkeit der ARMAX-Modelle zu bekommen.

Wir leiten zunächst den Zusammenhang mit Regressionsmodellen anhand eines einfachen Beispiels her. Nehmen Sie ein lineares Regressionsmodell

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad (4.31)$$

an, dessen Fehlerterm u_t jedoch nicht wie üblich unabhängig und identisch verteilt ist, sondern mit sich selbst rückwärts in der Zeit korreliert ist. Präziser gesagt vereinbaren wir eine AR(1)-Struktur für den Fehlerterm

$$u_t = \phi u_{t-1} + \epsilon_t. \quad (4.32)$$

Wir hätten (4.32) natürlich auch in Operatorschreibweise $\phi(B)u_t = \epsilon_t$ formulieren können. Der Punkt ist folgender: Erweitern wir die ursprüngliche Regressionsgleichung mit dem entsprechenden AR-Operator, erhalten wir ein ARX-Modell. Wir zeigen diese Behauptung Schritt für Schritt

$$\begin{aligned}\phi(B)y_t &= \phi(B)\beta_0 + \phi(B)\beta_1 x_t + \phi(B)u_t \\ &= \phi(1)\beta_0 + \phi(B)\beta_1 x_t + \epsilon_t \\ &= (1 - \phi)\beta_0 + \beta_1 x_t - \phi\beta_1 x_{t-1} + \epsilon_t \\ &= \theta_0 + \chi_0 x_t + \chi_1 x_{t-1} + \epsilon_t.\end{aligned}\tag{4.33}$$

Wir erhalten also präzise ein ARX(1,1)-Modell, mit den Parameterrestriktionen

$$\theta_0 = (1 - \phi)\beta_0, \quad \chi_0 = \beta_1 \quad \text{und} \quad \chi_1 = -\phi\beta_1.\tag{4.34}$$

Betrachtet man ein ARMAX(0,∞,0)-Modell, erhält man definitionsgemäß die allgemeine Gleichung eines Distributed-Lag-Modells

$$y_t = \theta_0 + \sum_{k=0}^{\infty} \chi_k x_{t-k} + \epsilon_t.\tag{4.35}$$

Unter bestimmten Restriktionen für die Koeffizienten χ_k ergeben sich dann etablierte Modelltypen. Beispielsweise führt die Vorschrift $\chi_k = \lambda\chi_{k-1}$ für $0 < \lambda < 1$ und $k \geq 1$ auf das bekannte *Koyck*-Modell.

Aufgabe 4.7

Leiten Sie das ARMAX-Modell und die zugehörigen Parameterrestriktionen her, die einem linearen Regressionsmodell mit ARMA(2,1)-verteilten Fehlern entsprechen.

4.4 Modellidentifikation

Nachdem wir nun verschiedene Erweiterungen von ARMA-Modellen kennengelernt haben, wollen wir zur Ausgangsklasse zurückkehren und versuchen, eine andere Frage zu beantworten, nämlich die der Modellidentifikation. In Kapitel 2 und 3 haben wir gesehen, dass sich AR- und MA-Prozesse entlang ihrer Autokorrelationsstruktur, respektive partiellen Autokorrelationsstruktur, unterscheiden lassen. Liegen solche Prozesse vor, lässt sich in der Regel sogar die Modellordnung relativ zuverlässig dem Autokorrelogramm entnehmen. Für die Mischprozesse der ARMA-Klasse gelingt das im Allgemeinen

nicht, da auch die Struktureigenschaften der (partiellen) Autokorrelation vermischt werden, vgl. Abbildung 4.1 auf Seite 72. Wir benötigen also höher entwickelte, formale Instrumente um eine geeignete Modellklasse festzulegen. Diese Instrumente sind Informationskriterien.

4.4.1 *Akaike*-Informationskriterium

Das historisch gesehen erste, und wohl auch am weitesten verbreitete Informationskriterium stammt von Hirotugu Akaike. Der Begriff Informationskriterium bezieht sich dabei nicht in oberflächlicher Form auf den Bestimmungszweck, sondern verweist auf den Ursprung des Kriteriums in der Informationstheorie. Die große Leistung von Akaike war es, das führende Paradigma der Informationstheorie, die *Kullback-Leibler*-Entropie, und das führende Paradigma der Statistik, das Maximum-Likelihood-Prinzip, zu vereinigen. Ohne auf die genaue Herleitung des *Akaike*-Informationskriteriums (AIC) einzugehen sei angemerkt, dass mit Hilfe der *Kullback-Leibler*-Information die Abweichung zwischen dem wahren, unbekannten Modell und dem mittels Maximum-Likelihood geschätzten Modell berechnet werden kann. Für das Informationskriterium ergibt sich dadurch die verblüffend einfache Form

$$\text{AIC} = -\frac{2}{T}(l - u), \quad (4.36)$$

wobei l der Wert der logarithmierten Likelihoodfunktion ist und u die Anzahl der geschätzten Parameter (inklusive $\hat{\sigma}^2$) angibt. Es ist hilfreich Formel (4.36) ein wenig zu analysieren.

Der hintere Term $l - u$ wird oft als „kompensierte“ Likelihood beschrieben. Das ist folgendermaßen zu verstehen. Ein Modell, das viele Parameter enthält, lässt sich genauer an eine gegebene Zeitreihe anpassen. Daher wird der Wert der logarithmierten Likelihoodfunktion bei einem solchen Modell auch größer sein. Zu stark angepasste Modelle besitzen jedoch zwei entscheidende Nachteile: Es muss eine große Zahl von Parametern geschätzt werden, und sie versagen häufig bei Prognosen, da sie zu stark auf die vorhandenen Daten zugeschnitten sind. Daher müsste ein im informationstheoretischen Sinne optimales Modell einen Kompromiss zwischen Anpassung an vorhandene Daten und Prognosefähigkeit darstellen. Im Fall des AIC wird dieser Kompromiss erreicht, indem die Anzahl der geschätzten Parameter vom Wert der logarithmierten Likelihoodfunktion abgezogen wird. Daher auch der Begriff „Likelihood-Kompensator“. Die Modellanpassung wird also im Nachhinein durch die Zahl der dafür notwendigen Parameter korrigiert, respektive bestraft. Der Term $2/T$ kann als eine Art asymptotische

Korrektur verstanden werden. Da der Betrag von l mit der Länge der Zeitreihe unweigerlich steigt, vgl. Formel (4.13) auf Seite 73, muss dieser Effekt neutralisiert werden, um für jede beliebige Länge der Zeitreihe den richtigen Kompensator zu erhalten. Das Minuszeichen schließlich überführt das Likelihood-Maximierungsproblem in ein Minimierungsproblem. Das ist notwendig, da der informationstheoretische Ansatz von der Minimierung von Modellabweichungen ausgeht. Das bedeutet, ein möglichst kleiner Wert des AIC ist erstrebenswert.

Im Falle normalverteilter Fehler lässt sich noch eine Vereinfachung vornehmen. In der Regel ist die Varianz σ^2 unbekannt und muss ebenfalls geschätzt werden. In diesem Fall wird das Informationskriterium oft in der Form

$$\text{AIC} = \log[\hat{\sigma}^2] + \frac{2u}{T} \quad (4.37)$$

geschrieben². Dieser Ausdruck ist proportional zu (4.36). Wir wollen einmal Schritt für Schritt überlegen, wieso das so ist. Unser Ausgangspunkt ist das AIC (4.36) und die bedingte LogLikelihood-Funktion (4.13) auf Seite 73. Wir setzen überall Schätzer ein und erhalten

$$\begin{aligned} \text{AIC} &= -\frac{2l}{T} + \frac{2u}{T} \\ &= \log[2\pi\hat{\sigma}^2] + \underbrace{\frac{1}{\hat{\sigma}^2} \cdot \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2}_{\hat{\sigma}^2} + \frac{2u}{T} \\ &= \log[2\pi] + \log[\hat{\sigma}^2] + 1 + \frac{2u}{T} \\ &\propto \log[\hat{\sigma}^2] + \frac{2u}{T}. \end{aligned} \quad (4.38)$$

Die Verknüpfung \propto in der letzten Zeile bedeutet „ist proportional zu“. Da die vernachlässigten Summanden $\log[2\pi] + 1$ weder von Parametern noch von T abhängen, liefern sie für das Minimierungsproblem keinen relevanten Beitrag und können weggelassen werden.

²Oft findet sich in der Literatur auch die Definition

$$\text{AIC} = \log[\hat{\sigma}^2] + \frac{2(p+q)}{T}.$$

Hier wird der Parameter σ^2 vernachlässigt, da er als Funktion der Daten und der übrigen Parametern ausgedrückt werden kann. Dies ist eine spezielle Eigenschaft der *Gaußschen* Likelihoodfunktion (siehe die Diskussion zur konzentrierten LogLikelihood-Funktion auf Seite 31).

4.4.2 *Bayes*-Informationskriterium

Das *Bayes*-Kriterium nach Schwarz (BIC oder SIC) ist kein Informationskriterium im eigentlichen Sinne, da es sich nicht aus der Informationstheorie ableitet. Es resultiert vielmehr aus der maximum a posteriori Selektion aus verschiedenen Alternativmodellen, unter der Bedingung, dass alle Varianten a priori dieselbe Eintrittswahrscheinlichkeit besitzen. Diese eher abstrakte Herleitungsskizze wird uns nicht weiter beschäftigen, wichtig sind nur die Konsequenzen, die sich daraus für das BIC (SIC) ergeben:

- BIC (SIC) erfordert nicht, dass das wahre Modell aus der Klasse der betrachteten Modelle stammt oder dass überhaupt ein wahres Modell existiert. Im Fall einer fehlspezifizierten Modellklasse liefert AIC kein verlässliches Ergebnis.
- BIC (SIC) ist asymptotisch konsistent, während AIC dazu tendiert in großen Stichproben die Modellordnung zu überschätzen.
- AIC besitzt die stärkere theoretische Fundierung, während BIC (SIC) lediglich das wahrscheinlichste Modell auswählt, wenn vorher keine Information vorliegt. Jedoch sind an diese Auswahl auch weniger Voraussetzungen geknüpft.

Als Essenz der Herleitung aus der *Bayes*-Formel ergibt sich für das BIC (SIC)

$$\text{BIC} = -\frac{2}{T} \left(l - \frac{u}{2} \log[T] \right). \quad (4.39)$$

Wie im Fall des *Akaike*-Informationskriteriums lässt sich auch hier für normalverteilte Fehlerterme die Vereinfachung

$$\text{BIC} = \log[\hat{\sigma}^2] + \frac{u \log[T]}{T} \quad (4.40)$$

vornehmen, die proportional zum originalen Kriterium (4.39) ist. Der Beweis soll dem Studierenden als Aufgabe am Ende dieses Abschnitts überlassen werden.

Der Vollständigkeit halber soll noch darauf hingewiesen werden, dass eine Zahl weiterer Informationskriterien existiert, die sich aus Endlichkeitskorrekturen der Stichprobe oder einer alternativen Gewichtung des Abstands der wahren Verteilung zur geschätzten Verteilung, im Rahmen der *Kullback-Leibler*-Information, ergeben. Ein weiteres im Hauptskript aufgeführtes Kriterium ist das *Hannan-Quinn*-Informationskriterium

$$\text{HQ} = -\frac{2}{T} (l - u \log \log[T]). \quad (4.41)$$

Aufgabe 4.8

Zeigen Sie, dass für normalverteilte Fehlerterme die Form

$$\text{BIC} = \log[\hat{\sigma}^2] + \frac{u \log[T]}{T}$$

proportional zum Info-Kriterium (4.39) ist. Hinweis: Setzen Sie die *Gaußsche* LogLikelihood-Funktion (4.13) auf Seite 73 ein.

4.4.3 Modellselektion

Die Modellselektion mit Hilfe von Informationskriterien ist denkbar einfach. Es werden zunächst alle in Frage kommenden Modellvarianten geschätzt. Anschließend werden die gewünschten Informationskriterien berechnet. Man entscheidet sich dann für die Modellvariante, bei der das Kriterium den geringsten Wert annimmt.

Beispiel 4.1: Modellselektion

Es liegt eine Zeitreihe mit $T = 100$ Beobachtungen vor und es stehen zwei Modellvarianten zur Disposition, ein reiner Rauschprozess und ein ARMA(0,0)-Modell. Die Werte der LogLikelihood-Funktion an den ML-Schätzern seien

$$l(\hat{\sigma}^2) = -12.453 \quad \text{und} \quad l(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) = -10.887.$$

Für die einzelnen Informationskriterien berechnet man

Info-Kriterium	Rauschen	ARMA(1,1)
AIC	0.2691	0.2778
BIC	0.2951	0.3559
HQ	0.2796	0.3094

Offensichtlich wird in Beispiel 4.1 unter jedem Informationskriterium der Rauschprozess als Modell bevorzugt, obwohl die LogLikelihood-Funktion für das angepasste ARMA(1,1)-Modell größer ist. Achtung: Verschiedene Informationskriterien müssen nicht immer dasselbe Modell präferieren! Es kann vorkommen, dass bei Verwendung unterschiedlicher Kriterien andere Modelle bevorzugt werden. Das liegt an der abweichenden Definition bzw.

Herleitung der verschiedenen Informationskriterien.

Aufgabe 4.9

Aus einer Zeitreihe mit $T = 40$ Beobachtungen wurden verschiedene ARMA(p, q)-Modelle geschätzt. Zusätzlich sei bekannt, dass der Fehlerterm normalverteilt ist. Es ergeben sich folgende Schätzer $\hat{\sigma}^2$ für die Varianz des Fehlerterms:

(p, q)	0	1	2
0	2.82	2.61	2.23
1	2.12	1.99	1.91
2	2.09	1.90	1.85
3	1.88	1.82	1.77

Berechnen Sie die zugehörigen Werte für AIC und BIC (SIC). Welche Modelle sind zu bevorzugen?

5

Bedingt heteroskedastische Prozesse

Unter dem Begriff „bedingt heteroskedastische“ Modelle werden in der Regel alle Variationen und Spezifikationen von ARCH- bzw. GARCH-Modellen zusammengefasst. Diese Modelle haben in den vergangenen zwanzig Jahren in den quantitativen Wirtschaftswissenschaften einen immensen Stellenwert erlangt. In der Finanzwirtschaft basieren beispielsweise Verfahren zur Bewertung von Derivaten auf GARCH-Modellen. Obwohl bedingt heteroskedastische Prozesse eine zusätzliche Komplikation mit sich bringen (sie sind in der Regel zwar unkorreliert, aber nicht unabhängig; wir werden später klären was das zu bedeuten hat), verfügen sie dennoch über einige sehr nützliche Eigenschaften. Wir werden aber zunächst die beiden offensichtlichen Fragen beantworten: Was bedeutet bedingt heteroskedastisch und wozu braucht man eine solche Modellklasse?

Zur ersten Frage: der Begriff „Heteroskedastizität“ stammt aus dem Griechischen (skedastikós: zum Zerstreuen gehörend) und bezeichnet eine Variabilität in der Streuung eines Prozesses. Wir haben vorher ARMA-Modelle in der Form

$$\phi(B)y_t = \theta(B)\epsilon_t \quad (5.1)$$

formuliert, wobei wir angenommen haben, dass für den Fehlerterm $\epsilon_t \sim N(0, \sigma^2)$ gilt. Das wäre ein Beispiel für einen homoskedastischen Prozess, die Streuung bleibt immer gleich, egal welchen Wert t annimmt. Das heißt aber im Umkehrschluss, wir müssen uns auf einen Ausdruck der Form σ_t^2 einstellen, der sich im Zeitablauf ändern kann, das ist die Bedeutung von Heteroskedastizität. Das Adjektiv „bedingt“ bedeutet hier nicht, dass der Prozess nur unter bestimmten Umständen heteroskedastisch ist, sondern dass die Verteilung im *Bayesschen* Sinne auf eine gegebene Information bedingt ist. Das heißt in diesem Zusammenhang, wenn auf den Fehler zum Zeitpunkt $t - 1$ bedingt wird, dann ist die Kenntnis von ϵ_{t-1} erforderlich, um die Verteilung (speziell die Streuung) von ϵ_t angeben zu können. Man würde dann schreiben: die Verteilung von $\epsilon_{t|t-1}$ ($t|t-1$ bedeutet zur Zeit t ,

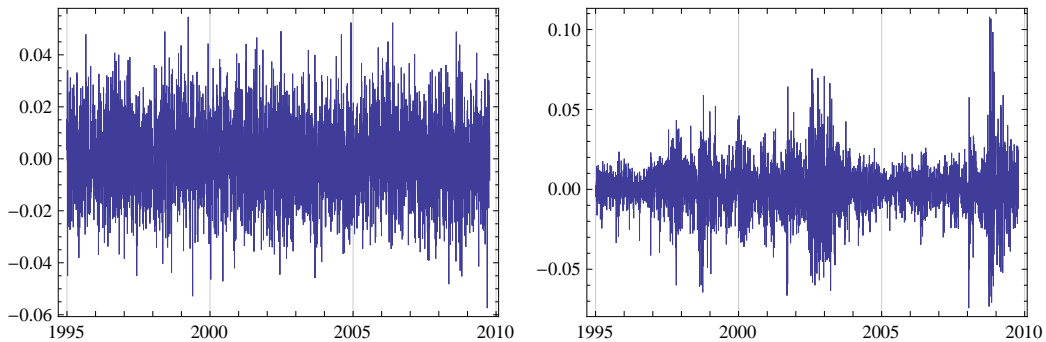


Abbildung 5.1: Homoskedastischer Rauschprozess und DAX-Rendite

gegeben die Information von Zeitpunkt $t - 1$) ist $N(0, \sigma_t^2)$. Machen Sie sich Folgendes klar: Die Verteilung von $\epsilon_{t+1}|t-1$ ist nicht bekannt, weil σ_{t+1}^2 von der Realisation ϵ_t abhängt, diese ist aber nicht in der Informationsmenge zum Zeitpunkt $t - 1$ enthalten. Daher hängt σ_{t+1}^2 selbst vom Zufall ab, ist also kein bekannter Verteilungsparameter, zumindest nicht bedingt auf die Information, die zum Zeitpunkt $t - 1$ vorliegt. Diese Diskussion hat an dieser Stelle natürlich noch keinen rigorosen Charakter, da wir die entsprechenden Modelle noch nicht in mathematischer Weise formuliert haben. Wenden wir uns deshalb zunächst der zweiten Frage zu.

Betrachten Sie Abbildung 5.1. Links ist die Simulation eines homoskedastischen Rauschprozesses abgebildet, wie er häufig als Modell für Renditen verwendet wird. Auf der rechten Seite befindet sich die Renditezeitreihe des Deutschen Aktien Index (DAX). Es wird sofort klar, dass der einfache homoskedastische Rauschprozess höchstwahrscheinlich kein gutes Modell für Wertpapierrenditen ist. Es fallen zwei Charakteristika der DAX-Rendite Zeitreihe unmittelbar ins Auge. Zum einen ist die Streuung offenbar tatsächlich nicht konstant, sondern variiert mit der Zeit. Zum anderen scheint sie sich nicht willkürlich zu verändern, sondern in zusammenhängenden Gebieten (Zeiträumen). Dieses Phänomen wird als „Volatility Clustering“ bezeichnet und ist ein starker Hinweis auf ARCH-Effekte in einer Zeitreihe. Es wird nun Zeit zu klären, wie bedingt heteroskedastische Modelle formuliert werden und was die Bezeichnungen ARCH und GARCH bedeuten.

5.1 ARCH-Modelle

Der Begriff ARCH steht für AutoRegressive Conditional Heteroskedasticity, also autoregressive bedingte Heteroskedastizität. Wir haben autoregressive Modelle bereits in Kapitel 2 behandelt und kennen einige ihrer Eigenschaften. Ein ARCH-Modell ist ganz ähnlich, nur dass hier gewissermaßen die Varianz als AR-Modell formuliert wird. Bitte behalten Sie stets im Hinterkopf:

Ein ARCH-Modell ist ein Modell für die Streuung eines Prozesses.

5.1.1 Das ARCH(1)-Modell

Das ARCH(1)-Modell ist der einfachste Vertreter der bedingt heteroskedastischen Modellklasse. Es hat darüber hinaus den Vorzug, dass seine Parameter allein durch Momentenschätzer bestimmt werden können, aber dazu später mehr. Wir haben oben gesagt, dass ein ARCH-Modell ein Modell für die Streuung eines Prozesses ist. Die Streuung ist ihrerseits ein Attribut des Fehlerterms, wir beginnen unsere Diskussion deshalb mit der folgenden, im ARCH-Kontext üblichen Notationsvariante

$$\epsilon_t = \sigma_t z_t. \quad (5.2)$$

Vereinbarungsgemäß ist z_t ein Zufallsfehler mit Erwartungswert null (Innovation) und Varianz eins, also ein standardisierter, nicht notwendigerweise normalverteilter Fehlerterm. Wir werden jedoch im Weiteren annehmen, dass $z_t \sim N(0, 1)$ gilt, wenn nicht explizit etwas anderes gilt.

Einschub: Lineare Transformation

Sei X eine Zufallsvariable und $Y = a + bX$, mit beliebigen Konstanten a und b , dann wird Y eine lineare Transformation von X genannt. Y ist ebenfalls eine Zufallsvariable und es gilt:

1. $E[Y] = a + bE[X]$
2. $\text{Var}[Y] = b^2 \text{Var}[X]$

Weiterhin gilt insbesondere: Ist X normalverteilt, dann ist Y ebenfalls normalverteilt.

Wir sehen nun, dass ϵ_t in (5.2) eine lineare Transformation von z_t ist, mit $a = 0$ und $b = \sigma_t$. Wenn z_t standard-normalverteilt ist, wie ist dann ϵ_t verteilt? Die Antwort ist einfach, $\epsilon_t \sim N(0, \sigma_t^2)$. Gleichung (5.2) ist gewissermaßen eine Mikroskopierung des Fehlerterms, den wir bereits aus

KAPITEL 5. BEDINGT HETEROSKEDASTISCHE PROZESSE

den vorangegangenen Kapiteln kennen. Das bedeutet aber auch, dass die Modelle, die wir noch untersuchen werden, nahtlos in die bereits bekannten Modelle eingepasst werden können. Die folgende Modellierung von σ_t stellt also eine weitere, filigranere Modellebene dar.

Nun genug der Vorarbeit. Das ARCH(1)-Modell wird allgemein üblich in folgender Form notiert

$$y_t = \epsilon_t = \sigma_t z_t \quad (5.3)$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2. \quad (5.4)$$

Der erste Teil von Gleichung (5.3) zeigt, dass ein reines Rauschmodell angenommen wird, $y_t = \epsilon_t$. Der zweite Teil von (5.3) ist die uns schon bekannte Transformation (Mikroskopierung) des Fehlerterms. Gleichung (5.4) stellt die eigentliche ARCH-Dynamik her. Demnach wird die Varianz σ_t^2 zum Zeitpunkt t erklärt durch eine Art Basisstreuung ω (kleines Omega) und eine mit α gewichtete quadrierte Realisation des Fehlers zum vorherigen Zeitpunkt $t-1$. ϵ_{t-1} muss also bekannt sein, damit σ_t^2 berechnet werden, und somit die Verteilung von y_t angegeben werden kann. Man sagt, y_t ist bedingt $N(0, \sigma_t^2)$ -verteilt und die Bedingung ist, dass die Information von Zeitpunkt $t-1$ vorliegt (also ϵ_{t-1} bekannt ist). Mit Hilfe der bedingten Notation aus dem vorherigen Abschnitt schreibt man formal¹ $y_{t|t-1} \sim N(0, \sigma_t^2)$.

Wieso ausgerechnet ein Rauschprozess mit Erwartungswert null? ARCH-Modelle wurden entwickelt, um den Besonderheiten von Renditezeitreihen von Finanzanlagen Rechnung zu tragen, vgl. Abbildung 5.1 auf Seite 90. Angenommen eine Wertpapieranlage hat eine erwartete jährliche Rendite von 10%, dann ist bei ca. 250 Handelstagen im Jahr die erwartete Tagesrendite etwa von der Größenordnung 10^{-4} . Demgegenüber liegt im Schnitt die tägliche Volatilität σ_t der meisten Finanzanlagen zwischen 2% und 5%. Der Erwartungswert wird also um zwei Größenordnungen von der Standardabweichung dominiert und kann daher in erster Näherung vernachlässigt werden. Sind genauere Ergebnisse gewünscht, kann analog zum Vorgehen in Kapitel 2 die Substitution $y_t = x_t - \mu$ vorgenommen werden. Man erhält dann anstelle von (5.3)

$$x_t = \mu + \sigma_t z_t. \quad (5.5)$$

Damit ist x_t unter den üblichen Annahmen bedingt $N(\mu, \sigma_t^2)$ -verteilt. Bedingt wird wiederum auf die Information zum Zeitpunkt $t-1$ (also ϵ_{t-1}),

¹Beachten Sie, dass es viele unterschiedliche Variationen der bedingten Schreibweise gibt. Wichtig ist, dass immer auf eine bestimmte Information bedingt werden muss, sobald Sie ein ‘|’ Zeichen sehen.

da σ_t^2 nur berechnet werden kann, wenn diese Information vorliegt.

Gibt es Restriktionen für die Parameter? Allerdings. Die Varianz ist das Quadrat der Standardabweichung. Es wäre sehr peinlich, wenn die Varianz negative Werte annehmen würde. Deshalb wird gefordert, dass $\omega > 0$ und $\alpha > 0$ gilt. Durch diese Einschränkung bleibt σ_t^2 immer positiv. Gibt es weitere Einschränkungen? Möglicherweise, es hängt davon ab, ob wir wünschen, dass der Prozess y_t stationär sein soll. Rufen Sie sich noch einmal die Definitionen der starken und schwachen Stationarität ins Gedächtnis (Seite 10 und 14). Wir werden uns darauf beschränken, Bedingungen für schwache Stationarität anzugeben.

5.1.2 Stationarität des ARCH(1)-Prozesses

Um herauszufinden, unter welchen Bedingungen der ARCH(1)-Prozess stationär ist, müssen wir die ersten beiden Momente und die Autokovarianz ausrechnen. Diese Berechnung birgt eine gewisse Subtilität. Wir kennen nur die bedingte Verteilung von y_t , nämlich $y_{t|t-1} \sim N(0, \sigma_t^2)$. Wir brauchen aber die Momente von y_t , also die Momente der unbedingten Verteilung. Lassen Sie uns sehen, ob wir diese berechnen können. Beginnen wir mit dem Erwartungswert

$$E[y_t] = E[\sigma_t z_t] = E[\sigma_t]E[z_t] = 0. \quad (5.6)$$

Einschub: Produkterwartungswert

Seien X und Y unabhängige Zufallsvariablen, dann gilt für das Produkt $X \cdot Y$:

$$E[X \cdot Y] = E[X] \cdot E[Y].$$

Beachten Sie, dass diese Beziehung nur für unabhängige oder zumindest unkorrelierte X und Y gilt!

Der unbedingte Erwartungswert ist also konstant, soweit so gut. Als nächstes berechnen wir die Autokovarianz mit Lag $k = 1$

$$\begin{aligned} \gamma_1 &= \text{Cov}[y_{t-1}, y_t] = E[y_{t-1}y_t] \\ &= E[\sigma_{t-1}z_{t-1}\sigma_t z_t] \\ &= E[\sigma_{t-1}z_{t-1}\sigma_t]E[z_t] = 0. \end{aligned} \quad (5.7)$$

KAPITEL 5. BEDINGT HETEROSKEDASTISCHE PROZESSE

Wieder wurde ausgenutzt, dass z_t ein reiner Zufallsfehler ist, der auf keine Information bedingt werden muss. Da z_t $N(0, 1)$ -verteilt ist, folgt natürlich wie in (5.6) sofort $E[z_t] = 0$. Die Lags $k \geq 2$ müssen wir nicht abprüfen, da die Varianzgleichung (5.4) nur verzögerte Einflüsse bis $k = 1$ enthält. Es folgt also

$$\gamma_k = 0 \quad \text{für } k \geq 1. \quad (5.8)$$

Damit haben wir gleichzeitig gezeigt, dass der ARCH(1)-Prozess ein reiner Rauschprozess ist. Wir haben gezeigt, dass y_t nicht autokorreliert ist, gleichzeitig wissen wir aber, dass die Verteilung von y_t von der Information zur Zeit $t - 1$ abhängt. Dieser etwas seltsame Befund wird oft so formuliert: „ y_t ist zwar unkorreliert, aber nicht unabhängig.“ Aus dieser Eigenschaft lässt sich (zumindest heuristisch) wiederum folgender, noch seltsamerer Schluss ziehen: Da lediglich unter Normalverteilung gilt, dass unkorrelierte Zufallsvariablen auch unabhängig sind, kann die stationäre Verteilung von y_t keine Normalverteilung sein. Wir werden später zeigen, dass dieser Schluss in der Tat korrekt ist.

Bisher verlief unsere Analyse völlig unproblematisch. Der Haken muss also in der letzten Bedingung stecken. Berechnen wir nun die Varianz von y_t

$$\begin{aligned} \gamma_0 &= \text{Var}[y_t] = E[\sigma_t^2 z_t^2] = E[\sigma_t^2] E[z_t^2] = E[\sigma_t^2] \\ &= E[\omega + \alpha \sigma_{t-1}^2 z_{t-1}^2] = \omega + \alpha E[\sigma_{t-1}^2] E[z_{t-1}^2] \\ &= \omega + \alpha E[\sigma_{t-1}^2]. \end{aligned} \quad (5.9)$$

In der ersten Zeile wurde wieder ausgenutzt, dass z_t unabhängig von verzögerter Information ist. In der Tat ist z_t^2 eine unabhängig χ_1^2 -verteilte (Chi-Quadrat mit einem Freiheitsgrad) Zufallsvariable². Da $z_t \sim N(0, 1)$ gilt, folgt unmittelbar $E[z_t^2] = 1$. In Zeile zwei wurde wieder verwendet, dass $\epsilon_{t-1} = \sigma_{t-1} z_{t-1}$ gilt. Der wichtigste Ausdruck, der sich aus (5.9) ergibt, ist

$$\text{Var}[y_t] = \omega + \alpha \text{Var}[y_{t-1}]. \quad (5.10)$$

Das ist aber exakt dieselbe Representation wie im AR(1)-Modell, vgl. Gleichung (2.8) auf Seite 8, mit den Ersetzungen $\sigma^2 \rightarrow \omega$ und $\phi^2 \rightarrow \alpha$. Nun wird auch klar, dass ω und α beide größer null sein müssen. Um die stationäre Varianz zu berechnen, bzw. um die notwendige Bedingung abzuleiten, geht

²Für die Summe von quadrierten standard-normalverteilten Zufallsvariablen Z_1, \dots, Z_N gilt

$$\sum_{n=1}^N Z_n^2 \sim \chi_N^2.$$

man nun völlig analog zu Abschnitt 2.1.2 vor. Am Ende dieses Abschnitts können Sie die Ableitungsschritte erneut im Rahmen einer Übungsaufgabe nachvollziehen. Man erhält durch Vorwärtsiteration

$$\text{Var}[y_t] = \omega \sum_{k=0}^{t-1} \alpha^k. \quad (5.11)$$

Die stationäre Varianz ergibt sich im Grenzwert für $t \rightarrow \infty$ unter Verwendung der geometrischen Reihe

$$\text{Var}[y_{st.}] = \lim_{t \rightarrow \infty} \omega \sum_{k=0}^{t-1} \alpha^k = \frac{\omega}{1 - \alpha} \quad \text{für } \alpha < 1. \quad (5.12)$$

Die gesuchte Bedingung für schwache Stationarität lautet also $\alpha < 1$. Der Parameter α wird im ökonomischen Kontext auch als Volatilitätspersistenz bezeichnet. Er bestimmt, wie lange Volatilitätsschocks im System nachwirken, bspw. in makroökonomischen Modellen. Zusammengefasst lauten die Parameterrestriktionen für ein schwach stationäres ARCH(1)-Modell $\omega > 0$ und $0 < \alpha < 1$.

Aufgabe 5.1

Zeigen Sie, dass im ARCH(1)-Modell die unbedingte Varianz von y_t (gegeben y_0)

$$\text{Var}[y_t] = \omega \sum_{k=0}^{t-1} \alpha^k$$

ist. Iterieren Sie die ersten drei Zeitpunkte vorwärts.

5.1.3 Kurtosis im ARCH(1)-Modell

In diesem Abschnitt werden wir bestimmte Aspekte der unbedingten Verteilung eines ARCH(1)-Prozesses untersuchen. Genauer gesagt, werden wir die Kurtosis (Wölbung) herleiten. In Abbildung 5.2 ist die Verteilung der DAX-Log-Renditen aus Abbildung 4.2 auf Seite 75 (rechts) dargestellt.

Einschub: Rendite und Log-Rendite

Sie werden sich wundern, warum einmal die Rendite und ein anderes mal die Log-Rendite (logarithmierte Rendite) betrachtet wird? Die erste Hälfte der Antwort lautet: Es besteht kein großer Unterschied zwischen beiden. Angenommen die relative Änderung eines Wertpapierkurses (Rendite) von einem Tag zum nächsten ist nicht allzu groß, dann gilt folgende Näherung

$$\frac{S_{t+1} - S_t}{S_t} = \frac{S_{t+1}}{S_t} - 1 \approx \log \left[\frac{S_{t+1}}{S_t} \right] = \log S_{t+1} - \log S_t.$$

Links steht die Rendite, rechts die Log-Rendite. Die zweite Hälfte der Antwort hängt damit zusammen, dass die Log-Rendite als Differenz geschrieben werden kann, was oft sehr praktisch ist.

Die zugehörige Normalverteilung ist ebenfalls als Referenz eingezeichnet. Es fällt sofort auf, dass die empirische Verteilung (Histogramm) deutlich mehr Wahrscheinlichkeitsmasse im Hauptbereich besitzt. Es kommt ebenfalls häufiger vor, dass extreme Werte auftreten. Das bedeutet, es befindet sich im Vergleich zur Normalverteilung auch mehr Wahrscheinlichkeitsmasse in den Ausläufern der Verteilung. Dieser Befund wird in der Praxis oft mit dem Begriff „Fat Tails“ beschrieben. Korrekterweise müsste man sagen, die unbedingte Verteilung der DAX-Log-Renditen ist stärker gewölbt als eine Normalverteilung. Die Wölbung einer Verteilung wird als Kurtosis bezeichnet und eine Überschusswölbung gegenüber der Normalverteilung heißt Leptokurtosis. Merke:

Bei Leptokurtosis befindet sich mehr Wahrscheinlichkeitsmasse im

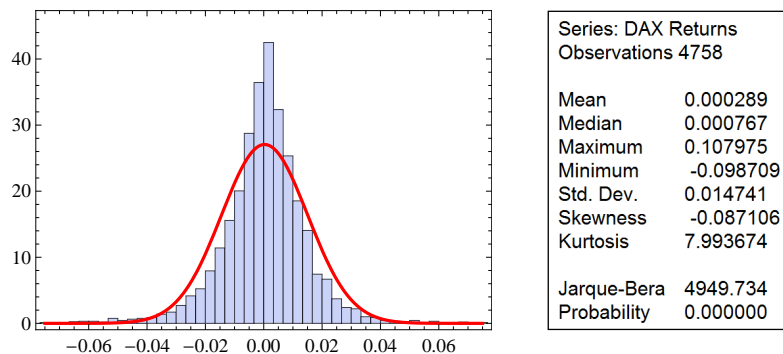


Abbildung 5.2: Histogramm (links) und Statistiken der unbedingten DAX-Log-Rendite Verteilung (rechts)

Zentralbereich und in den Ausläufern als bei einer Normalverteilung.

Die Normalverteilung hat immer eine Kurtosis von 3. Die exakte Definition der Kurtosis lautet

$$K = E \left[\frac{x - \mu}{\sigma} \right]^4 = \frac{m_4}{m_2^2}, \quad (5.13)$$

mit m_k als k -tem zentralen Moment (Siehe Einschub in Abschnitt 2.1.2). Werte über 3 werden als Überschusskurtosis oder Kurtosis-Exzess bezeichnet.

Wenn das ARCH(1)-Modell tatsächlich eine bessere Beschreibung von Wertpapierrenditen (respektive Log-Renditen) liefern soll, sollte es besser eine unbedingte Verteilung mit Kurtosis-Exzess generieren können. Wir werden das nachprüfen. Die stationäre Varianz haben wir bereits berechnet. Damit steht uns der Term im Nenner von (5.13) bereits zur Verfügung (erinnern Sie sich, dass $m_2 = \sigma^2$ gilt). Bezüglich dem vierten Moment gehen wir analog vor

$$\begin{aligned} E[y_t^4] &= E[\epsilon_t^4] = E[\sigma_t^4 z_t^4] = E[\sigma_t^4] E[z_t^4] \\ &= 3E[(\omega + \alpha \epsilon_{t-1}^2)^2] \\ &= 3E[\omega^2 + 2\omega\alpha \epsilon_{t-1}^2 + \alpha^2 \epsilon_{t-1}^4] \\ &= 3\omega^2 + 6\frac{\omega^2\alpha}{1-\alpha} + 3\alpha^2 E[\epsilon_{t-1}^4] \\ &= 3\frac{\omega^2(1-\alpha)}{1-\alpha} + 6\frac{\omega^2\alpha}{1-\alpha} + 3\alpha^2 E[\epsilon_{t-1}^4] \\ &= \frac{3\omega^2(1+\alpha)}{1-\alpha} + 3\alpha^2 E[y_{t-1}^4]. \end{aligned} \quad (5.14)$$

In der ersten Zeile wurde wieder ausgenutzt, dass z_t unabhängig von σ_t ist (was auch für beliebige Potenzen gilt). In Zeile zwei wurde dann die Kurtosis der Normalverteilung eingesetzt, $E[z_t^4] = 3$. In der vierten Zeile wurde schließlich die stationäre Varianz (5.12) für $E[\epsilon_{t-1}^2]$ eingesetzt.

Erkennen Sie die endgültige Form von Gleichung (5.14) wieder? Es handelt sich um eine rekursive Gleichung, die strukturell völlig identisch mit Gleichung (5.10) auf Seite 94 ist. Nur die Koeffizienten sind hier etwas komplizierter. Wenn wir aber zusätzlich fordern, dass $3\alpha^2 < 1$ sein soll, so dass die entsprechende geometrische Reihe konvergiert, können wir die stationäre Lösung in vollkommen analoger Weise ablesen

$$E[y_{st.}^4] = \frac{3\omega^2(1+\alpha)}{(1-\alpha)(1-3\alpha^2)} = \frac{\omega^2}{(1-\alpha)^2} \cdot 3 \frac{(1+\alpha)(1-\alpha)}{1-3\alpha^2}. \quad (5.15)$$

Sind wir damit am Ende? Noch nicht ganz. Wir müssen noch durch das Quadrat der stationären Varianz teilen. Die Terme sind in (5.15) bereits sehr vorteilhaft gruppiert und die dritte binomische Formel führt uns ans Ziel

$$K = \frac{E[y_{st.}^4]}{\text{Var}[y_{st.}]^2} = 3 \frac{1 - \alpha^2}{1 - 3\alpha^2} > 3. \quad (5.16)$$

Das ARCH(1)-Modell generiert also tatsächlich eine stationäre Überschusskurtosis. Es ist damit ein guter Kandidat zur Modellierung von Renditezeitreihen von Wertpapieren. Beachten Sie, dass die stationäre Kurtosis im Grenzwert für $\alpha \rightarrow 0$ gegen 3 geht. Das ARCH-Modell würde also in diesem Fall auf ein ganz gewöhnliches Rauschmodell mit konstanter Varianz ω reduziert.

Aufgabe 5.2

In Abschnitt 7.1 auf Seite 145 ist ein Schema für die Berechnung höherer Momente der Normalverteilung angegeben. Leiten Sie die stationäre Schiefe (Skewness)

$$S = E \left[\frac{x - \mu}{\sigma} \right]^3 = \frac{m_3}{m_2^{3/2}}$$

für das ARCH(1)-Modell her.

Aufgabe 5.3

Wie müssen die Parameter des ARCH(1)-Modells gewählt werden, wenn die unbedingte Verteilung die Eigenschaften

a) $\sigma^2 = \frac{3}{2}$ und $K = 4$

b) $\sigma^2 = 4$ und $K = 9$

besitzt? Hinweis: Berechnen Sie zuerst α , dann ω .

5.1.4 Parameterschätzung im ARCH(1)-Modell

Das ARCH(1)-Modell ist nicht nur hinsichtlich seiner Architektur das einfachste seiner Klasse, es erlaubt darüberhinaus eine sehr einfache und elegante Schätzung seiner Parameter mit der Momentenmethode.

Das Standardverfahren zur Parameterschätzung in ARCH- respektive GARCH-Modellen ist Maximum-Likelihood. Wir werden dieses Verfahren im Anschluss motivieren. Zunächst soll aber die Momentenmethode erörtert werden.

Eine Bestimmung der Parameter mit Hilfe von empirisch geschätzten Momenten wird dadurch ermöglicht, dass die Kurtosis (5.16) eine direkte Beziehung zu einem der beiden zu schätzenden Parameter herstellt, nämlich zu α . Die Kurtosis kann aber wiederum aus geschätzten Momenten berechnet werden

$$\hat{K} = \frac{\hat{m}_4}{\hat{m}_2^2}. \quad (5.17)$$

Wird Formel (5.16) nun entsprechend umgestellt, vgl. Aufgabe 5.3, erhält man einen Schätzer für den Parameter α

$$\hat{\alpha} = \sqrt{\frac{\hat{K} - 3}{3(\hat{K} - 1)}}. \quad (5.18)$$

Genaugenommen erhält man sogar zwei Schätzer für α , da die negative Wurzel ebenfalls eine Lösung der quadratischen Gleichung für α^2 darstellt. Eine Konstruktionsbedingung des ARCH(1)-Modells war jedoch, dass $\alpha > 0$ gilt. Somit können wir die negative Lösung außer Acht lassen. Aus der geschätzten Varianz lässt sich mit Hilfe von $\hat{\alpha}$ und Gleichung (5.12) sofort ein Schätzer für den zweiten Parameter ω angeben

$$\hat{\omega} = \hat{m}_2(1 - \hat{\alpha}). \quad (5.19)$$

Die Momentenschätzer selbst werden analog zu den Schätzern für Erwartungswert und Varianz (2.29) und (2.30) auf Seite 21 definiert

$$\hat{m}_k = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^k. \quad (5.20)$$

Damit steht ein sehr effizientes zweistufiges Verfahren zur Schätzung der Parameter zur Verfügung: Berechne im ersten Schritt \hat{m}_2 und \hat{m}_4 aus der beobachteten Zeitreihe und bestimme im zweiten Schritt $\hat{\alpha}$ und $\hat{\omega}$ aus den Momentenschätzern. Eine numerische Maximierung, wie sie unter ML benötigt wird, entfällt hier. Leider ist eine vergleichbar einfache Schätzung mit der Momentenmethode in Modellen höherer Ordnung nicht möglich. Man greift daher auf das bewährte Arbeitspferd der modernen Statistik zurück: Maximum-Likelihood.

Wir haben bereits in Kapitel 2 gesehen, dass auf verzögerte Realisierungen des Zufallsprozesses bedingt wurde, damit die Likelihoodfunktion eine faktorisierte Form annimmt, also als Produkt von (bedingten) Einzeldichten geschrieben werden kann. Im Falle der bedingt heteroskedastischen Prozesse funktioniert derselbe Trick, nur dass hier die Verteilung selbst, bzw. ihr Parameter σ_t , auf die vergangene Information bedingt werden muss. Wir gehen ganz analog vor. Zunächst wird angenommen, dass z_t unabhängig und identisch standard-normalverteilt ist, und dass weiterhin $\epsilon_0 = 0$ gilt. Dann lässt sich die gemeinsame Dichte von y_1, \dots, y_T wie folgt schreiben

$$\begin{aligned} p(y_T, y_{T-1}, \dots, y_1) &= \prod_{t=1}^T p(y_t | \epsilon_{t-1}) \\ &= \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{1}{2} \frac{y_t^2}{\sigma_t^2}} \\ &= \prod_{t=1}^T \frac{1}{\sqrt{2\pi(\omega + \alpha\epsilon_{t-1}^2)}} e^{-\frac{1}{2} \frac{y_t^2}{\omega + \alpha\epsilon_{t-1}^2}}. \end{aligned} \tag{5.21}$$

Von Schritt zwei nach drei wurde einfach die ARCH-Gleichung (5.4) auf Seite 92 eingesetzt. Die logarithmierte Likelihoodfunktion ist einfach wieder der Logarithmus der gemeinsamen Dichte (5.21), aufgefasst als Funktion der Parameter ω und α

$$\begin{aligned} l(\omega, \alpha) &= \sum_{t=1}^T \log p(y_t | \epsilon_{t-1}) \\ &= -\frac{1}{2} \sum_{t=1}^T \left(\log[2\pi(\omega + \alpha\epsilon_{t-1}^2)] + \frac{y_t^2}{\omega + \alpha\epsilon_{t-1}^2} \right) \\ &= -\frac{T}{2} \log[2\pi] - \frac{1}{2} \sum_{t=1}^T \left(\log[\omega + \alpha\epsilon_{t-1}^2] + \frac{y_t^2}{\omega + \alpha\epsilon_{t-1}^2} \right). \end{aligned} \tag{5.22}$$

Gleichung (5.22) muss wieder numerisch maximiert werden, um einen ML-Schätzer für die beiden Parameter zu erhalten. Das überlassen wir getrost dem Computer, unsere Arbeit hier ist getan.

Aufgabe 5.4

Gegeben sei die Zeitreihe

$$(y_t)_{t=1,\dots,8} = \{0, 0, 2, 0, 0, -2, 0, 0\}.$$

Gehen Sie davon aus, dass ein ARCH(1)-Prozess zugrunde liegt und schätzen Sie die Parameter ω und α mit der Momentenmethode.

Aufgabe 5.5

Nehmen Sie an, dass der ARCH(1)-Prozess aus Aufgabe 5.4 eine standard-normalverteilte Innovation z_t besitzt. Geben Sie den Wert der logarithmierten Likelihoodfunktion für $\epsilon_0 = 0$ und die Parameterwerte $\omega = \frac{2}{3}$ und $\alpha = \frac{1}{3}$ an.

5.1.5 Das ARCH(p)-Modell

Die Erweiterung der ARCH-Klasse auf ARCH(p)-Modelle³ ist kaum eine unerwartete Neuigkeit. Das bis hierher Gesagte besitzt mit den offensichtlichen Modifikationen weiterhin Gültigkeit. Wir werden diese Modifikationen dennoch Schritt für Schritt durchgehen. Zunächst soll das ARCH(p)-Modell formal notiert werden

$$y_t = \sigma_t z_t \tag{5.23}$$

$$\sigma_t^2 = \omega + \sum_{k=1}^p \alpha_k \epsilon_{t-k}^2. \tag{5.24}$$

Die redundante Definition $\epsilon_t = \sigma_t z_t$ wird im Folgenden weggelassen, es ist völlig ausreichend im Hinterkopf zu behalten, dass der Term auf der rechten Seite von (5.23) der Fehlerterm ist. Neu ist hier, dass jeder mit Lag k verzögerte Fehler einen eigenen Parameter α_k erhält. Dadurch wird das Modell natürlich anpassungsfähiger, da nun mehr „Stellschrauben“ zur Verfügung stehen. Andererseits wird die Parameterschätzung auch aufwendiger. Sie werden bereits ahnen, dass man in der Praxis nun auch wieder auf das alte Problem trifft, wie das geeignete Modell ausgesucht werden soll, das am besten zu einer vorliegenden Zeitreihe passt. All diese Probleme haben wir prinzipiell schon analysiert und gelöst und wir werden sehen, dass das

³Gelegentlich wird auch die Bezeichnung ARCH(q) verwendet.

Vorgehen völlig analog zum Vorgehen bei AR-, MA- bzw. ARMA-Modellen ist. Zunächst kümmern wir uns aber um die unmittelbaren Fragen. Die erste Frage lautet: Welche Einschränkungen müssen die Parameter erfüllen, damit die Varianz für alle Zeitpunkte positiv bleibt? Die Antwort ist offensichtlich und völlig analog zum ARCH(1)-Fall, nämlich $\omega, \alpha_1, \dots, \alpha_k > 0$. Die zweite Frage lautet: Unter welchen Bedingungen erhalten wir einen stationären Prozess? Dieser Frage könnten wir nun durch aufwendige Reihenentwicklung der verzögerten Terme nachgehen. Wir werden uns stattdessen eines kleinen Tricks bedienen, der in Abschnitt 7.2 auf Seite 145 erläutert wird. Zunächst wird die Varianzgleichung (5.24) für die unendlich weit entfernte Zukunft geschrieben

$$\sigma_\infty^2 = \omega + \sum_{k=1}^p \alpha_k \epsilon_\infty^2 = \omega + \sigma_\infty^2 z_\infty^2 \sum_{k=1}^p \alpha_k. \quad (5.25)$$

Da $\text{Var}[y_t] = E[\sigma_t^2]$ gilt, wird als nächstes der Erwartungswert auf beiden Seiten von (5.25) berechnet

$$E[\sigma_\infty^2] = \omega + E[\sigma_\infty^2] \sum_{k=1}^p \alpha_k. \quad (5.26)$$

Weil die stationäre Varianz genau dem Erwartungswert von σ_t^2 für $t \rightarrow \infty$ entspricht, erhält man den strukturell vertrauten Ausdruck

$$\text{Var}[y_{st.}] = \frac{\omega}{1 - \sum_{k=1}^p \alpha_k} \quad \text{für} \quad \sum_{k=1}^p \alpha_k < 1, \quad (5.27)$$

vgl. mit der stationären ARCH(1)-Varianz auf Seite 95. Die Bedingung ist also prinzipiell dieselbe geblieben, die Parameter $\alpha_1, \dots, \alpha_p$ müssen in der Summe (nicht einzeln!) kleiner als eins sein.

Last but not least berechnen wir die LogLikelihood-Funktion, da im ARCH(p)-Modell eine Schätzung mit der Momentenmethode nicht mehr ohne weiteres möglich ist. Wir führen hier eine generische Schreibweise ein, die auch für die GARCH-Modelle in Abschnitt 5.2 Gültigkeit behält. Bezeichne ϵ^{t-1} die gesamte Information, die zum Zeitpunkt t vorliegt, dann gilt im ARCH(p)-Fall $p(y_t | \epsilon^{t-1}) = p(y_t | \epsilon_{t-1}, \dots, \epsilon_{t-p})$, und die logarithmierte Likelihoodfunktion besitzt die Form

$$\begin{aligned} l(\omega, \alpha_1, \dots, \alpha_p) &= \sum_{t=1}^T \log p(y_t | \epsilon^{t-1}) \\ &= -\frac{T}{2} \log[2\pi] - \frac{1}{2} \sum_{t=1}^T \left(\log \sigma_t^2 + \frac{y_t^2}{\sigma_t^2} \right), \end{aligned} \quad (5.28)$$

mit dem zugehörigen Modell für die bedingte Streuung

$$\sigma_t^2 = \omega + \sum_{k=1}^p \alpha_k \epsilon_{t-k}^2. \quad (5.29)$$

Um die LogLikelihood-Funktion für beliebige Modelle anzugeben, muss abgesehen von den Funktionsargumenten in (5.28) nur die Varianzgleichung (5.29) modifiziert werden.

Aufgabe 5.6

Zeigen Sie, dass der ARCH(∞)-Prozess

$$y_t = \sigma_t z_t$$

$$\sigma_t^2 = \omega + \sum_{k=1}^{\infty} \alpha_k \epsilon_{t-k}^2,$$

mit $\omega > 0$, $\alpha_k = \alpha^k$ und $0 < \alpha < \frac{1}{2}$ eine stationäre Varianz besitzt.

5.2 GARCH-Modelle

Wir wollen unsere Diskussion der GARCH-Modelle (Generalized AutoRegressive Conditional Heteroscedasticity) über einen kleinen Umweg beginnen. Betrachten Sie den Ausdruck⁴ $\nu_t = \epsilon_t^2 - \sigma_t^2$ („Ny“) und lassen Sie uns den bedingten Erwartungswert ausrechnen

$$E[\nu_t | \epsilon^{t-1}] = E[\epsilon_t^2 | \epsilon^{t-1}] - \sigma_t^2 = \sigma_t^2 E[z_t^2] - \sigma_t^2 = 0. \quad (5.30)$$

Die Differenz ν_t besitzt eine gewisse Ähnlichkeit mit einem Fehlerterm. Sie ist darüber hinaus unkorreliert, aber nicht unabhängig! Lassen Sie uns nun folgendes ARMA(1,1)-Modell in $y_t^2 = \epsilon_t^2 = \sigma_t^2 z_t^2$ anschreiben⁵

$$y_t^2 = \omega + (\alpha + \beta) y_{t-1}^2 + \nu_t - \beta \nu_{t-1}. \quad (5.31)$$

Dieses Modell ist im Grunde identisch mit dem typischen ARMA(1,1)-Modell (4.1) aus Kapitel 4 auf Seite 69. Es ist lediglich in y_t^2 formuliert und enthält

⁴Die Differenz ν_t wird auch als Martingaldifferenz bezeichnet. Der Begriff Martingal stammt ursprünglich aus dem Glücksspiel. In der Wahrscheinlichkeitstheorie gilt er als Synonym für ein faires Spiel.

⁵Eine äquivalente Argumentationskette für das ARMA(p, q)-Modell kann im Hauptskript nachgelesen werden.

speziell gewählte Koeffizienten. Lassen Sie uns sehen, wohin uns ein paar algebraische Manipulationen führen. Aufgrund der Definition $\nu_t = \epsilon_t^2 - \sigma_t^2$ gilt $y_t^2 = \epsilon_t^2 = \nu_t + \sigma_t^2$. Setzen wir diesen Ausdruck für y_t^2 , respektive y_{t-1}^2 in (5.31) ein, erhalten wir nach trivialen Umformungen

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \quad (5.32)$$

Gleichung (5.32) ist unsere erste generalisierte ARCH-Gleichung. Das Prinzip lässt sich leicht auf höhere Ordnungen verallgemeinern.

Aufgabe 5.7

Zeigen Sie, dass die Innovation $\nu_t = \epsilon_t^2 - \sigma_t^2$ bedingt unkorreliert ist, also dass

$$\text{Cov}[\nu_t \nu_{t-1} | \epsilon^{t-1}] = 0$$

gilt. Tipp: Verwenden Sie die Vereinfachung aus Abschnitt 7.3 auf Seite 147.

Aufgabe 5.8

Zeigen Sie, dass die Innovation $\nu_t = \epsilon_t^2 - \sigma_t^2$ sogar unbedingt unkorreliert ist, also dass

$$\text{Cov}[\nu_t \nu_{t-1}] = 0$$

gilt. Tipp: Schreiben Sie die Innovation als $\nu_t = (z_t^2 - 1)\sigma_t^2$.

5.2.1 Das GARCH(1,1)-Modell

Im Grunde haben wir das GARCH(1,1)-Modell schon hergeleitet. Der Vollständigkeit halber wird es hier noch einmal notiert

$$y_t = \sigma_t z_t \quad (5.33)$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (5.34)$$

mit $\omega, \alpha, \beta > 0$. Das GARCH(1,1)-Modell ist das in der Praxis wahrscheinlich am häufigsten verwendete bedingt heteroskedastische Modell. Ähnlich wie im ARMA-Fall benötigt man in der Regel relativ viele verzögerte Fehlerterme, um eine gute Modellanpassung mit einem reinen ARCH-Modell zu erzielen. Die GARCH(1,1)-Variante schafft meistens eine ausreichend gute Anpassung bei minimaler Parametrisierung.

Die Stationaritätsbedingung resultiert wiederum lediglich aus der Anforderung an die stationäre Varianz. Am einfachsten erkennt man die notwendigen Bedingungen durch Umschreiben von (5.34) und anschließende Erwartungswertbildung

$$\begin{aligned} E[\sigma_t^2] &= E[\omega + \alpha\sigma_{t-1}^2 z_{t-1}^2 + \beta\sigma_{t-1}^2] \\ &= \omega + (\alpha + \beta)E[\sigma_{t-1}^2]. \end{aligned} \quad (5.35)$$

Vergleicht man diesen Ausdruck mit (5.10) auf Seite 94, wird sofort klar, dass die stationäre Varianz durch

$$\text{Var}[y_{st.}] = \frac{\omega}{1 - \alpha - \beta} \quad (5.36)$$

gegeben ist, und dass die eingesetzte geometrische Reihe nur konvergiert, wenn $\alpha + \beta < 1$ gilt. Letzteres ist die gesuchte Bedingung für Stationarität eines GARCH(1,1)-Prozesses.

Bei der Formulierung der bedingten Likelihoodfunktion können wir auf die generische Form (5.28) zurückgreifen und erhalten unmittelbar

$$l(\omega, \alpha, \beta) = -\frac{T}{2} \log[2\pi] - \frac{1}{2} \sum_{t=1}^T \left(\log \sigma_t^2 + \frac{y_t^2}{\sigma_t^2} \right), \quad (5.37)$$

mit

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \quad (5.38)$$

Auch beim GARCH(1,1)-Modell kann der Maximum-Likelihood-Schätzer für ω , α und β nur durch numerische Maximierung der logarithmierten Likelihoodfunktion (5.37) gefunden werden.

Aufgabe 5.9

Berechnen Sie die Kurtosis eines GARCH(1,1)-Prozesses. Nehmen Sie wie gewöhnlich an, dass z_t unabhängig und identisch standard-normalverteilt ist.

5.2.2 Das GARCH(p, q)-Modell

Die offensichtliche Erweiterung des GARCH(1,1)-Modells entsteht durch Lags höherer Ordnung, sowohl in den verzögerten Innovationen, als auch

in den verzögerten Varianzen. Das GARCH(p, q)-Modell⁶ notiert

$$y_t = \sigma_t z_t \quad (5.39)$$

$$\sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2. \quad (5.40)$$

Hier gilt ebenfalls die Bedingung $\omega, \alpha_j, \beta_k > 0$, für $j = 1, \dots, p$ und $k = 1, \dots, q$. Zur Erinnerung, j und k sind lediglich Laufvariablen, vgl. (5.40).

Die Analyse der Stationaritätseigenschaften erfolgt nach dem altbekannten Schema und birgt keine Überraschungen. Der GARCH(p, q)-Prozess ist kovarianzstationär, wenn

$$\sum_{j=1}^p \alpha_j + \sum_{k=1}^q \beta_k < 1 \quad (5.41)$$

gilt.

Die Struktur der LogLikelihood-Funktion wird Sie nun sicher nicht mehr vor ein unlösbares Rätsel stellen. Daher bleibt dieses Highlight der anschließenden Aufgabe überlassen.

Aufgabe 5.10

Geben Sie die logarithmierte Likelihood-funktion für das GARCH(p, q)-Modell an. Nehmen Sie wieder an, dass für die Innovation $z_t \sim N(0, 1)$ gilt.

5.3 ARMA-GARCH und asymmetrische Erweiterungen

Wie eingangs erwähnt, stellt die GARCH-Klasse elaborierte Modelle für den Fehlerterm bereit. Diese Fehlermodelle lassen sich auf natürliche Weise in die bereits bekannten, konventionellen Modellklassen einpassen. Man erhält auf diesem Wege die erste wichtige Erweiterung der GARCH-Modelle.

⁶Beachten Sie auch hier, dass eine gebräuchliche alternative Schreibweise die Modellordnungen p und q vertauscht.

5.3. ARMA-GARCH UND ASYMMETRISCHE ERWEITERUNGEN

5.3.1 ARMA-GARCH

Fassen wir die Ergebnisse aus Kapitel 4 und diesem zusammen und pressen sie in ein einziges, komprimiertes Modell, erhalten wir das ARMA-GARCH-Modell

$$\phi(B)y_t = \theta_0 + \theta(B)\epsilon_t \quad (5.42)$$

$$\epsilon_t = \sigma_t z_t \quad (5.43)$$

$$\sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2. \quad (5.44)$$

Dieses Modell besitzt nun die Charakteristika sowohl von ARMA-, wie auch von GARCH-Modellen. Ein Prozess, der (5.42) bis (5.44) genügt, ist beispielsweise nur stationär, wenn die Wurzeln seines AR-Teils außerhalb des komplexen Einheitskreises liegen und wenn die Summe aller α 's und β 's kleiner als eins ist.

Die Likelihoodfunktion muss nun auf verzögerte Beobachtungen y_{t-1}, \dots und auf verzögerte Fehlerterme ϵ_{t-1}, \dots bedingt werden. Wieviele Beobachtungen, respektive Fehler aus der Vergangenheit benötigt werden, hängt von der Modellordnung ab. Um eine einheitliche Notation zu schaffen vereinbaren wir, dass für beliebige Ordnungen p und q gilt $p(y_t|Y^{t-1}) = p(y_t|y_{t-1}, \dots, y_{t-p}, \epsilon_{t-1}, \dots, \epsilon_{t-q})$. Y^{t-1} stellt also die Menge der gesamten zum Zeitpunkt t benötigten Information dar. Nun kann die bedingte LogLikelihood-Funktion eines ARMA(m, n)-GARCH(p, q)-Modells als

$$\begin{aligned} l(\boldsymbol{\theta}) &= p(y_t|Y^t) \\ &= -\frac{T}{2} \log[2\pi] - \frac{1}{2} \sum_{t=1}^T \left(\log \sigma_t^2 + \frac{\epsilon_t^2}{\sigma_t^2} \right), \end{aligned} \quad (5.45)$$

mit

$$\epsilon_t = y_t - \sum_{j=1}^m \phi_j y_{t-j} - \theta_0 - \sum_{k=1}^n \theta_k \epsilon_{t-k} \quad (5.46)$$

$$\sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2 \quad (5.47)$$

geschrieben werden. Alle Parameter des gemischten Modells wurden in (5.45) in einem Parametervektor $\boldsymbol{\theta} = (\phi_1, \phi_2, \dots, \beta_q)'$ zusammengefasst.

ARMA-GARCH-Modelle bilden die allgemeinste konventionelle Modellklasse, die wir bisher kennengelernt haben. Konventionell bedeutet hier, dass einige „exotische“ Erweiterungen wie SARIMA oder ARMAX nicht explizit enthalten sind. Die erforderlichen Modifikationen können aber problemlos vorgenommen werden. Wir werden in den nächsten Abschnitten noch weitere Exoten kennenlernen, jedoch handelt es sich dann um Erweiterungen des GARCH-Teils.

Aufgabe 5.11

Geben Sie die logarithmierte Likelihoodfunktion für das ARMA(1,2)-GARCH(2, 1)-Modell explizit an. Nehmen Sie wieder an, dass für die Innovation $z_t \sim N(0, 1)$ gilt.

Aufgabe 5.12

Unter welchen Bedingungen ist der zugehörige Prozess aus Aufgabe 5.11 stationär?

5.3.2 Asymmetrische GARCH-Modelle allgemein

Die Notwendigkeit asymmetrische Effekte zu implementieren steht wieder in engem Zusammenhang mit empirisch beobachteten Phänomenen in Finanzzeitreihen. Bei Aktienrenditen stellt man beispielsweise fest, dass durch eine negative Realisation am Vortag (Kurssturz) eine höhere Volatilität für den Folgetag induziert wird, als bei einer betragsmäßig gleichgroßen positiven Rendite. Dieses Phänomen wird „Leverage-Effekt“ genannt. Man geht davon aus, dass dieser Zusammenhang auf ein risikoaverses Verhalten von Anlegern zurückzuführen ist. Vereinfacht ausgedrückt bedeutet das, dass ein Wirtschaftssubjekt einen deutlichen Verlust überproportional schlimmer einstuft als einen Gewinn derselben Größenordnung. Solche Annahmen sind in den Wirtschaftswissenschaften seit langem bekannt⁷ und akzeptiert. Die Entwicklung asymmetrischer GARCH-Modelle trägt diesem Sachverhalt theoretisch wie empirisch Rechnung.

Wir werden im Folgenden verschiedene Ansätze kennenlernen, die solche Asymmetrien abbilden können. Wir werden uns dabei auf einige Spezialfälle konzentrieren, die vom Standpunkt der Parameterschätzung einfach zu

⁷Vgl. die klassische Nutzentheorie, speziell die Annahme, dass Nutzenfunktionen konkav sind. Die Risikoaversion wird dann durch den *Pratt-Arrow*-Koeffizienten ausgedrückt.

5.3. ARMA-GARCH UND ASYMMETRISCHE ERWEITERUNGEN

handhaben sind, aber dennoch die Idee der asymmetrischen Modellierung gut transportieren. Der Einfachheit halber soll die Darstellung auf den Fall der GARCH(1,1)-Erweiterung beschränkt werden, da zum einen höhere Modellordnungen in der Praxis selten sind, und zum anderen keine qualitativ neuen Erkenntnisse aus der Verallgemeinerung gewonnen werden. Eine formale Darstellung für GARCH(p, q)-Ordnungen befindet sich im Hauptskript.

5.3.3 Threshold ARCH

Beim Threshold-ARCH oder TARCH Ansatz wird das wohl einfachste Prinzip zur Realisierung von Asymmetrien verfolgt. Zunächst wird eine Indikatorvariable definiert, die sozusagen anzeigt, ob die Rendite negativ oder positiv ist

$$d_t = \begin{cases} 0 & \text{für } \epsilon_t \geq 0 \\ 1 & \text{für } \epsilon_t < 0. \end{cases} \quad (5.48)$$

Diese Indikatorvariable wird dann in die Varianzgleichung des GARCH-Modells eingeschleust, um die heutige Varianz um einen zusätzlichen Faktor γ zu erhöhen, wenn die gestrige Innovation negativ war. Man erhält dann das folgende TARCH-Modell

$$y_t = \sigma_t z_t \quad (5.49)$$

$$\sigma_t^2 = \omega + (\alpha + \gamma d_{t-1}) \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (5.50)$$

mit $\omega, \alpha, \beta, \gamma > 0$.

Eine wichtige Eigenschaft, die wir im Folgenden zwar begründen, aber nicht mathematisch herleiten wollen, sorgt dafür, dass das TARCH-Modell einfach zu handhaben und theoretisch gut beherrschbar bleibt. Die Indikatorvariable d_t ist ihrerseits selbst eine Zufallsvariable. Sie ist zwar vollständig von ϵ_t abhängig, aber unter den üblichen Annahmen ebenfalls vollkommen unkorreliert mit ϵ_t^2

$$\text{Cov}[d_t, \epsilon_t^2] = E[d_t \epsilon_t^2] - E[d_t] E[\epsilon_t^2] = 0. \quad (5.51)$$

Wird z_t wieder standard-normalverteilt angenommen, ist ϵ_t normalverteilt mit Erwartungswert null. Das bedeutet, die Wahrscheinlichkeit, dass ϵ_t größer oder kleiner als null ist, ist jeweils $\frac{1}{2}$. d_t ist also eine *Bernoulli*-verteilte Zufallsvariable mit $\pi = \frac{1}{2}$. Damit ergibt sich aber für die Varianz des TARCH-Prozesses

$$\begin{aligned} E[\sigma_t^2] &= \omega + \alpha E[\epsilon_{t-1}^2] + \gamma E[d_{t-1}] E[\epsilon_{t-1}^2] + \beta E[\sigma_{t-1}^2] \\ &= \omega + \left(\alpha + \frac{\gamma}{2} + \beta \right) E[\sigma_{t-1}^2]. \end{aligned} \quad (5.52)$$

Es wird sofort klar, auf welche Form die stationäre Varianz hinauslaufen wird und unter welchen Bedingungen der TAR-CH-Prozess überhaupt stationär ist. Nach Abspulen der inzwischen etablierten Routine erhält man

$$\text{Var}[y_{st.}] = \frac{\omega}{1 - \alpha - \beta - \frac{\gamma}{2}}, \quad (5.53)$$

unter der Bedingung, dass $\alpha + \beta + \frac{\gamma}{2} < 1$ gilt.

Die LogLikelihood-Funktion lässt sich ebenfalls sofort anschreiben. Sie ist durch die generische Form (5.28), respektive (5.37), in Verbindung mit der Varianzgleichung (5.50) gegeben.

5.3.4 Quadratic GARCH

Quadratic GARCH oder QGARCH bietet möglicherweise einen noch intuitiveren Zugang zur Abbildung asymmetrisch bedingter Volatilitäten. Dieser Ansatz hat im Zusammenhang mit der Bewertung von Finanzderivaten einige Prominenz erlangt. Zu Anschauungszwecken soll hier nicht der allgemeinste Fall diskutiert werden, sondern die spezielle Form

$$y_t = \sigma_t z_t \quad (5.54)$$

$$\sigma_t^2 = \omega + \alpha(\epsilon_{t-1} - \gamma)^2 + \beta\sigma_{t-1}^2. \quad (5.55)$$

Hier wird die Asymmetrie in Form des Modellparameters γ in unmittelbarer Nähe des Fehlerterms platziert, noch bevor das Quadrat gebildet wird. Dieses Vorgehen ist sowohl direkt wie auch effizient. Die formale Behandlung des Modells bleibt strukturell genauso einfach wie die Idee selbst. Die Formulierung der Likelihoodfunktion folgt dem trivialen Schema der vorangegangenen Abschnitte.

Aufgabe 5.13

Berechnen Sie die stationäre Varianz im QGARCH-Modell und geben Sie die Stationaritätsbedingung für den QGARCH-Prozess an.

5.3.5 Exponential GARCH

Exponential GARCH oder kurz EGARCH ist eine Modellvariante mit wesentlich höherem Komplexitätsgrad als die bisher vorgestellten Modelle.

5.3. ARMA-GARCH UND ASYMMETRISCHE ERWEITERUNGEN

Seine Attraktivität gewinnt das EGARCH-Modell vorwiegend durch seine theoretischen Eigenschaften⁸. Wir werden Schritt für Schritt die wichtigsten Punkte herausarbeiten.

Zunächst wird das EGARCH-Modell nicht in den Varianzen σ_t^2 formuliert, sondern in den logarithmierten Varianzen $h_t = \log \sigma_t^2$. Dadurch entfallen die Positivitätsrestriktionen der Parameter, da eine beliebige Größe, positiv wie negativ, durch Rücktransformation positiv wird, $e^{h_t} > 0$ für alle $h_t \in \mathbb{R}$. Die zweite Besonderheit besteht darin, dass das EGARCH Modell als verzögerte Variable z_t und nicht ϵ_t verwendet. Eine Umrechnung ist natürlich immer möglich durch $z_t = \frac{\epsilon_t}{\sigma_t}$. Die dritte Modifikation betrifft den funktionalen Zusammenhang zwischen h_t und z_t . Wir werden diesen Zusammenhang diskutieren, nachdem wir das EGARCH-Modell formuliert haben

$$y_t = e^{\frac{h_t}{2}} z_t \quad (5.56)$$

$$h_t = \omega + g(z_{t-1}) + \beta h_{t-1} \quad (5.57)$$

$$g(z_t) = \alpha(|z_t| - E[|z_t|]) + \gamma z_t. \quad (5.58)$$

Die Spezifikation der verzögerten Innovation $g(z_{t-1})$ erlaubt eine separate Modellierung der Wirkung der Größenordnung und des Vorzeichens von z_{t-1} . Beachten Sie, dass $E[|z_t|] \neq 0$ gilt. Für die übliche Annahme der standardnormalverteilten Innovation gilt $E[|z_t|] = \sqrt{\frac{2}{\pi}}$.

⁸Insbesondere die asymptotischen Eigenschaften sind theoretisch bedeutsam, da EGARCH-Prozesse für immer kleinere Zeitschritte $\Delta t \rightarrow 0$ gegen bestimmte kontinuierliche Prozesse konvergieren, die in der quantitativen Finanzwirtschaft von außerordentlich großer Bedeutung sind.

Einschub: $E[|Z|]$

Die Betragsfunktion $|Z|$ kann formal geschrieben werden als

$$|Z| = \begin{cases} Z & \text{für } Z \geq 0 \\ -Z & \text{für } Z < 0. \end{cases}$$

Für standard-normalverteiltes Z ergibt sich damit der Erwartungswert von $|Z|$ als

$$\begin{aligned} E[|Z|] &= \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= - \int_{-\infty}^0 z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \int_0^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \int_0^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \int_0^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z e^{-\frac{z^2}{2}} dz. \end{aligned}$$

Von Schritt zwei auf drei wurden die Integrationsgrenzen unter Vorzeichenwechsel vertauscht und eine Substitution $y = -z$ mit anschließender Rückersetzung $y \rightarrow z$ der Integrationsvariable durchgeführt. Das verbleibende Integral in Schritt vier lässt sich leicht lösen und ergibt eins. Damit ist der gesuchte Erwartungswert $E[|Z|] = \sqrt{\frac{2}{\pi}}$.

Um die Likelihoodfunktion angeben zu können, wollen wir nun das EGARCH-Modell in eine vertrautere Form bringen. Die Definition $h_t = \log \sigma_t^2$ lässt sich durch algebraische Manipulation auf die Form $\sigma_t = e^{\frac{h_t}{2}}$ bringen. Wenn Ihnen diese Umstellung nicht klar ist, konsultieren Sie kurz die Rechenregeln für Logarithmen auf Seite 29. Damit reduziert sich Gleichung (5.56) wieder auf die bekannte Form $y_t = \sigma_t z_t$. Setzen wir weiterhin konsequent $z_t = \frac{\epsilon_t}{\sigma_t}$ ein, erhalten wir für die Varianzgleichung

$$\begin{aligned} \sigma_t^2 &= \exp \left[\omega + g \left(\frac{\epsilon_{t-1}}{\sigma_{t-1}} \right) + \beta \log \sigma_{t-1}^2 \right] \\ &= \exp \left[\omega + \alpha \left(\left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right] \sigma_{t-1}^{2\beta}. \end{aligned} \tag{5.59}$$

Damit sind wir zumindest in der Lage, die LogLikelihood-Funktion im

5.3. ARMA-GARCH UND ASYMMETRISCHE ERWEITERUNGEN

Rahmen des bewährten Schemas anzugeben. Gleichung (5.59) zeigt aber deutlich, dass bereits eine Diskussion der Stationaritätseigenschaften beträchtliche mathematische Probleme verursacht.

Wir werden an dieser Stelle auf tiefere Analysen verzichten und stattdessen noch eine Beobachtung bezüglich der Modellierung des Leverage-Effekts im EGARCH-Modell machen. Gleichung (5.58) hat bereits gezeigt, dass die Einflüsse der Größenordnung und des Vorzeichens der Innovation voneinander separiert werden. Die Varianzgleichung (5.59) offenbart aber noch eine weitere Quelle von Asymmetrie, die durch den nichtlinearen Verlauf der Exponentialfunktion eingebracht wird. Das EGARCH-Modell stellt also einen hoch komplexen Leverage-Mechanismus bereit. Der Preis für diesen Mechanismus ist die analytisch schwierige Handhabbarkeit des Modells.

5.3.6 Symmetrische vs. asymmetrische Modelle

In Abbildung 5.3 sind die bedingten Varianzen verschiedener GARCH-Modelle als Funktionen des verzögerten Fehlers abgebildet. Für die

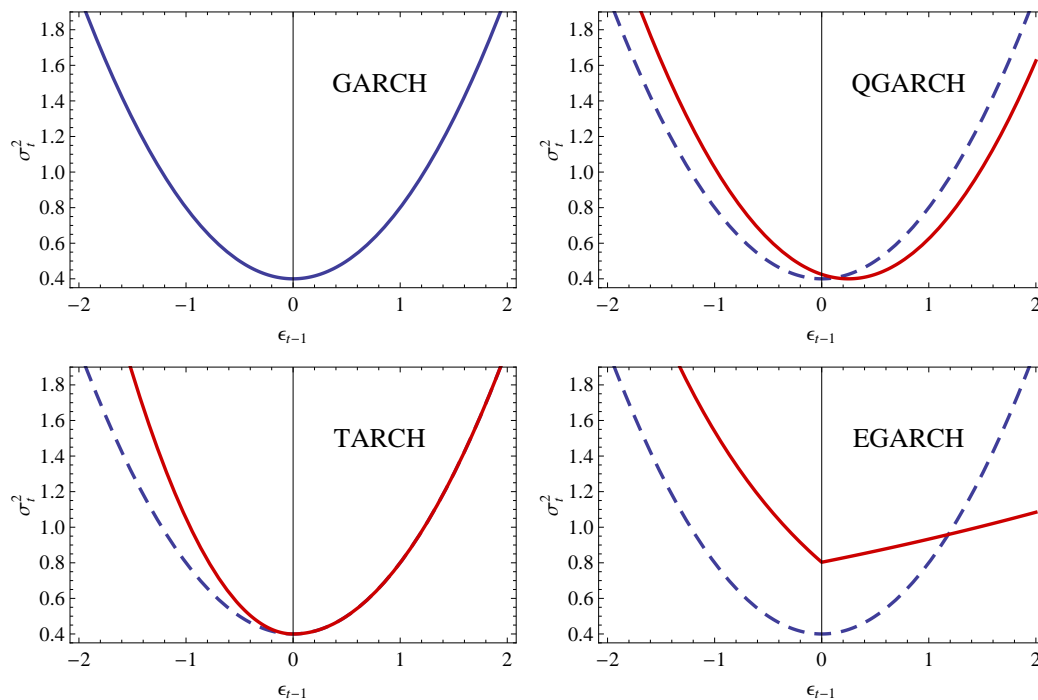


Abbildung 5.3: Bedingte Varianz symmetrischer und asymmetrischer GARCH-Modelle

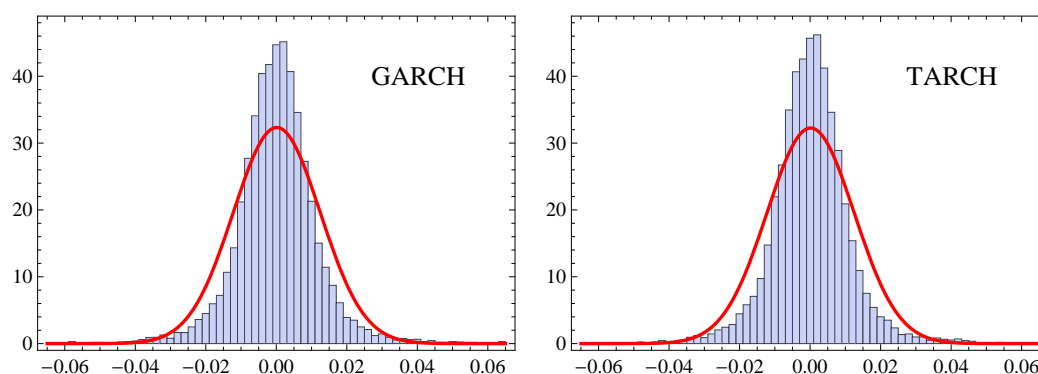


Abbildung 5.4: Unbedingte GARCH- und TARCH-Verteilung mit identischer stationärer Varianz

verzögerte Varianz wurde vereinfachend $\sigma_{t-1}^2 = 1$ angenommen. Die übrigen Parameter sind für alle Modelle gleich gewählt⁹. Das einzige symmetrische Modell ist das GARCH-Modell. Der Verlauf der bedingten GARCH-Varianz ist in allen anderen Abbildungen als Referenz gestrichelt eingezeichnet. Es ist gut erkennbar, dass die unterschiedlichen Modelle sehr individuelle Asymmetrien erzeugen. Während QGARCH lediglich die Varianzparabel verschiebt, wird im TARCH-Modell ein Ast aufgebogen. EGARCH erzeugt nicht einmal eine Parabel, sondern zwei unterschiedliche, exponentiell gekrümmte Äste.

Ein möglicher Leverage-Effekt drückt sich im Allgemeinen nicht in der unbedingten Verteilung durch eine Schiefe aus. Abbildung 5.4 zeigt eine simulierte unbedingte Verteilung für ein GARCH- und ein TARCH-Modell, die über eine identische stationäre Varianz verfügen. Für die Simulation wurden in beiden Fällen dieselben Zufallszahlen verwendet. Obgleich beide Histogramme bei genauer Inspektion nicht vollkommen identisch sind, gibt es in keiner Verteilung ein Anzeichen für eine nennenswerte Schiefe.

5.3.7 ARCH in Mean

Wir kommen nun zu einer besonderen Modellklasse, den ARCH in Mean oder kurz ARCH-M-Modellen. Diese Klasse ist aufgrund ihrer Eigenschaften überaus interessant für theoretische ökonomische Fragestellungen. Ein über die quantitative Finanzwirtschaft hinaus verbreitetes Konzept ist das Prämien-Prinzip. Vereinfacht besagt dieses Prinzip, dass ein Entscheider für

⁹Im EGARCH-Modell muss das Vorzeichen des Leverage-Parameters definitionsbedingt umgekehrt werden.

5.3. ARMA-GARCH UND ASYMMETRISCHE ERWEITERUNGEN

die Übernahme von Risiko durch eine höhere erwartete Rendite entschädigt werden will. Die ARCH-M-Formulierung erfasst genau diese Überlegung

$$y_t = \mu + \lambda\sigma_t + \sigma_t z_t \quad (5.60)$$

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \quad (5.61)$$

Das Modell wurde hier wieder mit GARCH(1,1)-Varianz formuliert, also als GARCH-M. Der Term $\lambda\sigma_t$ wird häufig als Risikoprämie bezeichnet. Die Risikoprämie hängt also von der Ungewissheit (Standardabweichung¹⁰) zum Zeitpunkt t ab. Der fixe Anteil des Erwartungswertes wird in finanzwirtschaftlichen Anwendungen häufig mit dem risikolosen Zinssatz $\mu = r$ identifiziert.

Das (G)ARCH-M-Modell ist einfach zu analysieren, hält aber trotzdem einige Überraschungen bereit. Zunächst einige Fakten. Die stationäre Varianz und damit auch die Bedingung für Stationarität des zugehörigen Prozesses hängt nur von der Varianzspezifikation (5.61) ab¹¹. Der stationäre Erwartungswert ist demzufolge

$$E[y_{st.}] = \mu + \lambda\sqrt{\text{Var}[y_{st.}]} \quad (5.62)$$

Ist z_t wieder $N(0,1)$ -verteilt, können wir die bedingte LogLikelihood-Funktion mit einer kleinen Modifikation angeben

$$l(\theta) = -\frac{T}{2} \log[2\pi] - \frac{1}{2} \sum_{t=1}^T \left(\log \sigma_t^2 + \frac{(y_t - \mu - \lambda\sigma_t)^2}{\sigma_t^2} \right) \quad (5.63)$$

Für σ_t^2 wird einfach wieder die entsprechende Varianzspezifikation, in unserem Fall (5.61), eingesetzt.

Abschließend soll noch eine Besonderheit analysiert werden. Wir haben in Abschnitt 5.1.2 gezeigt, dass ein ARCH(1)-Prozess ein reiner Rauschprozess ist. Diese Eigenschaft hat sich auf alle bisher diskutierten Prozesse erstreckt. Im ARCH-M Kontext hängt aber der Erwartungswert selbst von der realisierten Streuung ab. Das bedeutet, wenn unter Leverage-Wirkung eine höhere Varianz durch eine negative Innovation heute induziert wird, sollte morgen eine höhere Risikoprämie, sprich ein höherer Erwartungswert realisiert wer-

¹⁰Häufig wird auch die bedingte Varianz σ_t^2 verwendet.

¹¹Das lässt sich einfach überprüfen, indem man die Gleichung $\text{Var}[y_t] = E[(\sigma_t(\lambda + z_t) - \lambda E[\sigma_t])^2] = E[\sigma_t^2]$ beweist.

den. Um diesen Sachverhalt zu belegen, analysieren wir das quadratische QGARCH-M-Modell

$$y_t = \mu + \lambda\sigma_t^2 + \sigma_t z_t \quad (5.64)$$

$$\sigma_t^2 = \omega + \alpha(\epsilon_{t-1} - \gamma)^2 + \beta\sigma_{t-1}^2. \quad (5.65)$$

Um die gesuchte Eigenschaft nachzuweisen zeigen wir, dass y_{t+1} und ϵ_t korreliert sind. Wir erhalten

$$\begin{aligned} \text{Cov}[y_{t+1}, \epsilon_t] &= E[y_{t+1}\epsilon_t] \\ &= E[(\mu + \lambda\sigma_{t+1}^2 + \epsilon_{t+1})\epsilon_t] \\ &= \lambda E[\omega\epsilon_t + \alpha(\epsilon_t - \gamma)^2\epsilon_t + \beta\sigma_t^2\epsilon_t] \\ &= -2\lambda\alpha\gamma E[\sigma_t^2]. \end{aligned} \quad (5.66)$$

Da $E[\sigma_t^2]$ immer positiv ist, folgt die Behauptung für $\lambda > 0$. Die Struktur von (5.66) zeigt deutlich, dass ohne asymmetrische Varianzspezifikation keine solche Korrelation zustande kommen würde, da ϵ_t bedingt normalverteilt ist und ungerade Momente verschwinden.

5.4 Modellidentifikation und Selektion

Für die Identifikation und Selektion sind zwei Fragestellungen von großer Bedeutung. Erstens, wie kann ein GARCH-Prozess von einem gewöhnlichen ARMA-Prozess unterschieden werden und zweitens, wie kann eine adäquate Modellordnung festgelegt werden?

5.4.1 ARMA vs. GARCH

Bezüglich der ersten Frage ist es hilfreich Abbildung 5.5 zu betrachten. Dort ist ein typischer AR(2)-Prozess mit dem dazugehörigen Autokorrelogramm gegeben. Dieser Prozess kann einfach anhand der Eigenschaften seiner ACF und PACF identifiziert werden. In der unteren Zeile in Abbildung 5.5 ist ein GARCH(1,1)-Prozess illustriert. Aus seiner ACF und PACF lässt sich keine Information gewinnen. Beachten Sie, dass scheinbar signifikante Autokorrelationen mit Lag 7 und 8 ausgewiesen werden. Diese Information ist nicht verlässlich, da wir bereits analytisch nachgewiesen haben, dass ein GARCH-Prozess ein unkorrelierter Rauschprozess ist. Der Sachverhalt bezüglich der (P)ACF-Überschreitungen ist im Hinblick auf zwei wichtige Punkte zu würdigen. Zum einen sind die Überschreitungen betragsmäßig klein, zum anderen werden die Konfidenzgrenzen unter der Annahme eines weißen Rauschens mit fixer Varianz berechnet. Diese Voraussetzung ist aber gerade bei GARCH-Prozessen verletzt.

5.4. MODELLIDENTIFIKATION UND SELEKTION

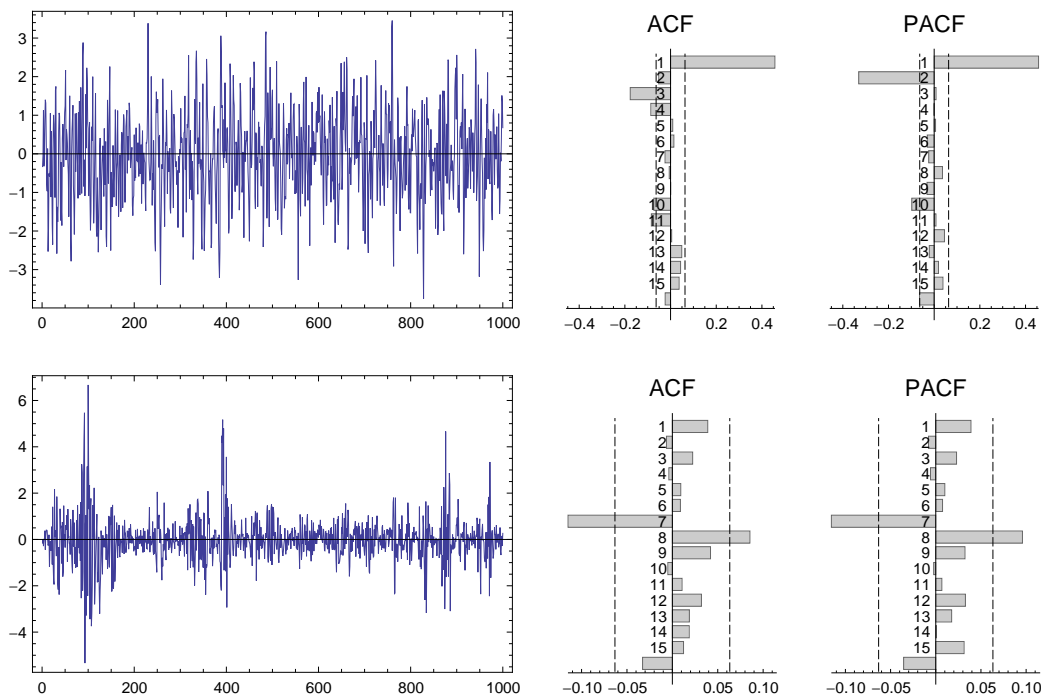


Abbildung 5.5: Simulierter AR(2)-Prozess (oben) und GARCH(1,1)-Prozess (unten) – ACF und PACF rechts

Stattdessen lässt sich der GARCH-Prozess durch optische Inspektion klar von einem gewöhnlichen Rauschprozess oder einer reinen ARMA-Spezifikation unterscheiden. Das Rauschspektrum des AR(2)-Prozesses ist gut erkennbar konstant. Diese Eigenschaft ist typisch für alle ARMA-Prozesse. Der GARCH-Prozess hingegen zeigt deutlich andere Charakteristika. Seine Varianz schwankt und bildet Gebiete mit größerer Rauschbreite aus (Volatility-Clustering). Diese beiden Eigenschaften, zusammen mit der deutlichen Leptokurtosis ($Kurtosis > 3$) der unbedingten Verteilung, sind die primären Identifikationsmerkmale eines GARCH-Prozesses.

Bei der Herleitung des GARCH(1,1)-Modells in Abschnitt 5.2 haben wir gesehen, dass ein GARCH-Modell als ARMA-Form in y_t^2 repräsentiert werden kann. Infolgedessen sind Autokorrelationen in den quadrierten Residuen ein Hinweis auf GARCH-Effekte in den Fehlern. Solche quadratische Restkorrelationen können auch formal abgetestet werden (siehe Hauptskript). Darüber hinaus kann detaillierte Information über die Modellordnung genau wie bei gemischten ARMA-Modellen nur mit Hilfe von Informationskriterien gewonnen werden. Auch ein reiner ARCH-Prozess kann nicht anhand seiner Autokorrelationsstruktur identifiziert werden!

5.4.2 Modellordnung und Selektion

Die Identifikation und Selektion einer geeigneten Modellvariante verläuft im Fall von GARCH-Modellen oder Mischformen völlig analog zur Selektion bei ARMA-Modellen, vgl. Abschnitt 4.4. Das heißt, man stützt sich wieder auf Informationskriterien bzw. auf die Beurteilung durch eine kompensierte Likelihoodfunktion. Der Übersicht halber sind die wichtigsten Informationskriterien nachfolgend noch einmal zusammengefasst:

$$\text{AIC} = -\frac{2}{T}(l - u) \quad (\text{Akaike}) \quad (5.67)$$

$$\text{BIC} = -\frac{2}{T}\left(l - \frac{u}{2} \log T\right) \quad (\text{Schwarz}) \quad (5.68)$$

$$\text{HQ} = -\frac{2}{T}(l - u \log[\log T]) \quad (\text{Hannan-Quinn}) \quad (5.69)$$

Dabei bezeichnet l den Wert der logarithmierten Likelihoodfunktion am Maximum $\hat{\boldsymbol{\theta}}_{ML}$, und u die Anzahl der geschätzten Parameter, also die Anzahl der Elemente in $\boldsymbol{\theta}$. Vergleicht man mehrere alternative Modelle, wird wieder jenes bevorzugt, dessen Informationskriterium am kleinsten ist.

Aufgabe 5.14

An eine Zeitreihe mit $T = 60$ Beobachtungen wurden drei verschiedene Modelle angepasst, ein ARMA(2,1)-GARCH(1,1)- (inklusive Intercept), ein ARCH(7)- und ein TAR-M-Modell. Nach Maximieren der LogLikelihood-Funktion ergaben sich jeweils die Werte $l = -10.4$, $l = -9.1$ und $l = -12.3$. Beurteilen Sie die drei Varianten unter AIC, BIC und HQ. Welche Modelle sollten jeweils bevorzugt werden?

6

Zustandsraummodelle und *Kalman*-Filter

In diesem Kapitel werden wir zwei neue Konzepte kennenlernen, Zustandsraummodelle und Filter. Insbesondere werden wir uns mit einem der leistungsfähigsten Instrumente der Zeitreihenanalyse auseinandersetzen, dem *Kalman*-Filter. Obwohl die bloße Erwähnung des *Kalman*-Filters so manchen gestandenen Ökonomen in Angst und Schrecken versetzt, soll hier eine Herangehensweise gewählt werden, die es Ihnen erlaubt, Schritt für Schritt alle wichtigen und notwendigen Konzepte zu erlernen und zu verstehen. Wir werden unterwegs eine ganze Reihe offener Enden verknüpfen, sodass Sie nicht nur Theorie und Umgang mit Zustandsraummodellen und Filtern erlernen, sondern auch ein tieferes Verständnis für viele bereits besprochene Sachverhalte entwickeln können.

Alles was wir an Handwerkszeug benötigen ist ein wenig grundlegende Matrix/Vektor-Rechnung. Wir werden daher einen kleinen Umweg über das Konzept des im Hauptkript kurz erwähnten linearen Filters machen, bevor wir zum Zustandsraummodell kommen. Dieser kleine Umweg erlaubt es uns bereits, ein intuitives und formales Verständnis des Filterprinzips zu entwickeln, während wir die notwendigen Rechentechniken wiederholen und üben.

6.1 Das lineare Filterproblem

Lassen Sie uns unsere Diskussion des optimalen linearen Filters mit einer Definition beginnen:

Ein Filter ist eine Maschine, die ein (möglicherweise verrauschtes) Eingangssignal verarbeitet und ein Ausgangssignal ausgibt. Die Signalverarbeitung erfolgt in optimaler Weise bezüglich eines bestimmten Kriteriums.

KAPITEL 6. ZUSTANDSRAUMMODELLE UND KALMAN-FILTER

Ein Filter wird als linear bezeichnet, wenn sein Ausgangssignal eine Linearkombination seiner Eingangssignal- und Rauschkomponenten ist. Sie sehen an dieser Stelle bereits, warum wir ein wenig lineare Algebra benötigen werden.

Wir beschränken uns im Weiteren auf kausale Filter, also solche, die keine Signalinformation aus der Zukunft benötigen. Bezeichnen wir das Eingangssignal mit x und das Ausgangssignal mit y , können wir das lineare Filterproblem in der Form

$$y_t = \sum_{k=0}^p w_k x_{t-k} + \epsilon_t, \quad (6.1)$$

mit $p \in \mathbb{N}_0$ schreiben. Die Koeffizienten w_k sind die Filtergewichte. Sie müssen in einer optimalen Weise hinsichtlich eines noch zu bestimmenden Kriteriums gewählt werden. Es ist an dieser Stelle auch nicht notwendig, irgendeine Annahme über die Verteilung des Fehlers ϵ_t zu machen, außer natürlich, dass der Erwartungswert null ist. Was könnten wir als Eingangssignal verwenden? Im Rahmen der Zeitreihenanalyse ist es sicher naheliegend, verzögerte Ausgangssignalwerte $x_t = y_{t-1}$ zu verwenden. Wir erhalten damit den *Wiener-Levinson*-Filter

$$y_t = \sum_{k=1}^p w_k y_{t-k} + \epsilon_t. \quad (6.2)$$

Im Folgenden wird wieder angenommen, dass $E[y_t] = 0$ gilt. Eine transformation ist auch hier immer durch die Substitution $y_t = x_t - \mu$ möglich. Wir werden uns jetzt um ein Optimalitätskriterium bemühen, um die Filtergewichte w_k berechnen zu können. Lassen Sie uns dazu Gleichung (6.2) umstellen, quadrieren und den Erwartungswert bilden

$$\begin{aligned} E[\epsilon_t^2] &= E\left[\left(y_t - \sum_{k=1}^p w_k y_{t-k}\right)^2\right] \\ &= \gamma_0 - 2 \sum_{k=1}^p w_k \gamma_k + \sum_{j=1}^p \sum_{k=1}^p w_j w_k \gamma_{j-k}. \end{aligned} \quad (6.3)$$

Zur Erinnerung: γ_k ist die Autokovarianz $E[y_t y_{t-k}]$, und es gilt weiterhin $\gamma_k = \gamma_{-k}$. Damit kann das Problem wesentlich kompakter in Matrix/Vektor-Form geschrieben werden

$$E[\epsilon_t^2] = \gamma_0 - 2\mathbf{w}'\boldsymbol{\gamma} + \mathbf{w}'\mathbf{T}\mathbf{w} \quad (6.4)$$

mit

$$\mathbf{w} = (w_1, \dots, w_p)', \quad \boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)', \quad \boldsymbol{\Gamma} = \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{p-2} \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_{p-1} & \dots & \gamma_1 & \gamma_0 \end{pmatrix}. \quad (6.5)$$

Wenn Ihnen die Schreibweise (6.4) und (6.5) nicht klar ist, wiederholen Sie bitte an dieser Stelle die Auffrischungen zur Matrixrechnung auf Seite 18 f. und gegebenenfalls die Aufgaben am Ende des Abschnitts. Als Optimalitätskriterium soll jetzt gefordert werden, dass die Fehlervarianz $E[\epsilon_t^2]$ möglichst klein wird. Sie erkennen hier deutlich das Kleinste-Quadrate-Prinzip (KQ). Wir müssen also den gesamten Vektor \mathbf{w} so bestimmen, dass das Optimalitätskriterium erfüllt ist. Die klassische Vorgehensweise ist Bilden der ersten Ableitung und Nullsetzen.

Einschub: Ableitung nach einem Vektor

Für die passend dimensionierten Vektoren \mathbf{a}, \mathbf{x} und die quadratische Matrix \mathbf{B} gelten folgende Ableitungsregeln

$$\begin{aligned} \frac{d}{d\mathbf{x}'} \mathbf{x}' \mathbf{a} &= \mathbf{a} \\ \frac{d}{d\mathbf{x}'} \mathbf{x}' \mathbf{B} \mathbf{x} &= (\mathbf{B} + \mathbf{B}') \mathbf{x} \\ \frac{d}{d\mathbf{x}'} (\mathbf{a} + \mathbf{B} \mathbf{x})' (\mathbf{a} + \mathbf{B} \mathbf{x}) &= 2\mathbf{B}' (\mathbf{a} + \mathbf{B} \mathbf{x}). \end{aligned}$$

Für symmetrisches \mathbf{B} ergibt sich die Vereinfachung

$$\frac{d}{d\mathbf{x}'} \mathbf{x}' \mathbf{B} \mathbf{x} = 2\mathbf{B} \mathbf{x}.$$

In diesem Fall sind die Ableitungsregeln völlig analog zum skalaren Kalkül.

Wir erhalten also aus (6.4) eine Gleichung für die optimale Wahl der Filtergewichte

$$-2\boldsymbol{\gamma} + 2\boldsymbol{\Gamma} \mathbf{w} = 0. \quad (6.6)$$

Lassen Sie uns die Filtergewichte umbenennen $w_k \rightarrow \phi_k$, um den Zusammenhang mit der Zeitreihenanalyse zu verdeutlichen. Umstellen von (6.6) führt dann zu folgender Lösung

$$\boldsymbol{\gamma} = \boldsymbol{\Gamma} \boldsymbol{\phi} \quad \Leftrightarrow \quad \boldsymbol{\phi} = \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}. \quad (6.7)$$

KAPITEL 6. ZUSTANDSRAUMMODELLE UND KALMAN-FILTER

Erkennen Sie die erste Gleichung wieder? Lassen Sie sich nicht von der Matrix-Schreibweise in die Irre führen. Komponentenweise erhält man

$$\gamma_k = \sum_{j=1}^p \phi_j \gamma_{k-j} \quad \Leftrightarrow \quad \phi(B) \gamma_k = 0. \quad (6.8)$$

Das sind die *Yule-Walker*-Gleichungen für einen $\text{AR}(p)$ -Prozess, vgl. (2.111) auf Seite 51, nur in Matrix/Vektor-Form notiert. Die Koeffizienten ϕ_1, \dots, ϕ_p des $\text{AR}(p)$ -Prozesses gehen also aus der optimalen Lösung des linearen Filterproblems hervor, wenn das KQ-Prinzip als Optimalitätskriterium zugrundegelegt wird. Genaugenommen müssten wir noch zeigen, dass die Lösung (6.7) tatsächlich ein Minimum ist. Aufgrund der Problemstellung kommt aber nur ein Minimum als Extremum in Frage, weshalb der formale Nachweis entfallen kann. Stattdessen halten wir fest, dass wir einen ersten Zusammenhang zwischen AR-Prozessen und linearen Filtern hergestellt haben, unser erstes verknüpftes Ende.

Ein weiterer, vor allem für explorative Anwendungen in der Ökonomie sehr wichtiger Filter ist der *Hodrick-Prescott*-Filter, der ebenfalls auf dem KQ-Prinzip basiert. Dieser Filter wird im Hauptskript nicht besprochen, weil er keinen unmittelbaren Zusammenhang mit der Theorie der Zeitreihenanalyse aufweist. Er ist jedoch in Standard-Softwarepaketen wie EViews implementiert. Der neugierige Leser kann ihn in Abschnitt 7.4 unter die Lupe nehmen.

Aufgabe 6.1

Leiten Sie den KQ-Schätzer $\hat{\theta}$ im linearen Modell

$$\mathbf{y} = \mathbf{X}\theta + \epsilon$$

aus Abschnitt 2.3 her. Hinweis: Verwenden Sie die Optimalitätsbedingung (2.40) auf Seite 27.

6.2 Zustandsraummodelle

Das Zustandsraummodell ist eine Erweiterung des bisher diskutierten Konzepts der Zeitreihen. Es ist gleichsam eine generellere und sehr viel flexiblere Modellklasse.

6.2.1 Architektur des Zustandsraummodells

Lassen Sie uns nocheinmal zurückblicken an den Anfang unserer Reise, zum AR(1)-Prozess

$$y_t = \theta_0 + \phi y_{t-1} + \epsilon_t. \quad (6.9)$$

Wir sind bisher immer davon ausgegangen, dass die Zeitreihe, die von Prozessen wie (6.9) erzeugt wird, direkt beobachtet werden kann. Das muss nicht zwangsläufig der Fall sein, denken Sie beispielsweise an das Brutto-Inlandsprodukt, das quartalsmäßig erfasst wird und später nachkorrigiert werden muss, um Rückstellungen und Rechnungsabgrenzungsposten zu korrigieren. Hier treten zwei Probleme im Zusammenhang mit der Beobachtung der Zeitreihe auf. Zum Einen kann die zeitliche Auflösung des eigentlichen Prozesses feiner sein als die Beobachtung, und zum Anderen kann die Messung mit einer gewissen Beobachtungsunschärfe belastet sein. Das Zustandsraummodell trägt diesem Sachverhalt Rechnung, indem es ein zweites Modell, das Messmodell, integriert

$$z_t = Hy_t + d + \delta_t. \quad (6.10)$$

Der beobachtbare Prozess ist nun z_t , der Messprozess. Der Systemprozess y_t ist latent. z_t ist eine lineare Funktion (im einfachsten Fall) von y_t , wird aber von einer weiteren Zufallsfluktuation, dem Messfehler δ_t , überlagert. Man nimmt üblicherweise an, dass System- und Messrauschen nicht miteinander korrelieren und reine Zufallsfluktuationen sind.

Es ist üblich, im Rahmen des Zustandsraummodells eine leicht von der Zeitreihenkonvention abweichende Notation zu verwenden. Man schreibt allgemein¹

$$y_t = Ay_{t-1} + b + \epsilon_t \quad (6.11)$$

$$z_t = Hy_t + d + \delta_t. \quad (6.12)$$

Für die Zufallsfehler wird weiterhin $\epsilon_t \sim N(0, \Omega)$, $\delta_t \sim N(0, R)$ und $\text{Cov}[\epsilon_t, \delta_t] = 0$ angenommen. Sie werden sich wundern, warum manche Parameter groß geschrieben wurden und andere nicht. Die Antwort mag Ihre schlimmsten Befürchtungen bestätigen; die Formulierung des Zustandsraummodells erstreckt sich auch auf vektorielle System- und Messmodelle. In diesem Fall wären A , H , Ω und R Matrizen. Aber keine Sorge, Sie werden sehen,

¹Eine noch allgemeinere Formulierung lässt auch eine Zeitabhängigkeit der Parameter zu: $A \rightarrow A_t$, $b \rightarrow b_t$ usw. (siehe Hauptskript). Wir werden diese Komplikation im Folgenden ignorieren, da sie qualitativ nicht zu neuen Erkenntnissen führt und später durch die offensichtlichen Modifikationen berücksichtigt werden kann.

KAPITEL 6. ZUSTANDSRAUMMODELLE UND KALMAN-FILTER

dass sich diese Komplikation fast von selbst behandelt, wenn es soweit ist. Wir wollen uns zunächst etwas genauer anschauen, wie das Zustandsraummodell arbeitet. Abbildung 6.1 zeigt einen Workflow des gesamten Zustandsraummodells. Das Systemmodell (6.11) ist in der oberen Hälfte repräsentiert, das Messmodell (6.12) in der unteren. Die beobachtbaren Größen sind grau unterlegt. Behalten Sie stets im Hinterkopf, dass der eigentlich interessierende Systemprozess nun nicht mehr direkt beobachtet werden kann.

Der Systemprozesszustand oder kurz Systemzustand y_t entwickelt sich jeweils in Abhängigkeit von der Realisation des vorangegangenen Zeitpunktes y_{t-1} , überlagert von einem aktuellen Zufallsfehler ϵ_t . Diese Dynamik hat klar eine AR(1)-Struktur und besitzt daher die *Markov*-Eigenschaft. Zur Erinnerung: Die *Markov*-Eigenschaft besagt, dass die gesamte Information der Vergangenheit in der letzten Realisation des Prozesses gespeichert ist. Daher muss die fernere Vergangenheit nicht mehr betrachtet werden, weil keine zusätzliche Information aus ihr gewonnen werden kann. Da hier der Systemzustand nicht mehr direkt beobachtet werden kann, spricht man auch von „Hidden *Markov*-Models“.

Der Systemprozess liefert durch seine Realisationen (nicht zwangsläufig durch jede Realisation) ein Eingangssignal für den Messprozess z_t . Dieses

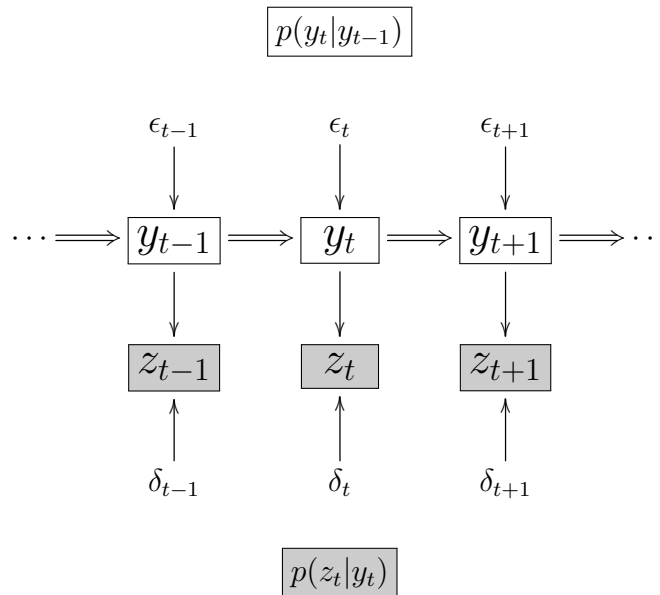


Abbildung 6.1: Zustandsraummodell – Systemmodell und Messmodell (grau)

Signal wird im Rahmen der Messung verarbeitet und einem Messfehler unterworfen, der die Beobachtungsunschärfe modelliert. Der Signalausgang des Messprozesses generiert schließlich die Zeitreihe, die wir beobachten können. Machen Sie sich anhand von Abbildung 6.1 klar, dass das Messmodell selbst keine *Markov*-Struktur aufweist. Dennoch sind die einzelnen Messungen nicht voneinander unabhängig, weil sie über die verborgene *Markov*-Kette des Systemmodells induziert werden.

Wir haben Prozess- und Messfehler als normalverteilte Zufallsvariablen definiert. Damit sind die beiden bedingten Wahrscheinlichkeitsdichten in Abbildung 6.1 ebenfalls *Gauß*-Dichten. Man bezeichnet

$$p(y_t|y_{t-1}) = \frac{1}{\sqrt{2\pi\Omega}} e^{-\frac{1}{2} \frac{(y_t - Ay_{t-1} - b)^2}{\Omega}} \quad (6.13)$$

als Übergangsdichte oder *Markov*-Kern. Die bedingte Messdichte ist durch

$$p(z_t|y_t) = \frac{1}{\sqrt{2\pi R}} e^{-\frac{1}{2} \frac{(z_t - Hy_t - d)^2}{R}} \quad (6.14)$$

gegeben. Ω und R sind bereits Varianzen! Handelt es sich bei System- und/oder Messmodell um vektorielle Modelle, ergibt sich jeweils eine bedingte multivariate Normalverteilung.

Einschub: Multivariate Normalverteilung

Eine vektorielle Zufallsvariable \mathbf{X} ist multivariat normalverteilt, mit Erwartungswertvektor $\boldsymbol{\mu}$ und Kovarianzmatrix $\boldsymbol{\Sigma}$, wenn die zugehörige Wahrscheinlichkeitsdichte durch

$$p(\mathbf{x}) = \frac{1}{\sqrt{\det[2\pi\boldsymbol{\Sigma}]}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

gegeben ist. Im bivariaten Fall $\mathbf{X} = (X_1, X_2)'$ kann die Kovarianz $\sigma_{12} = \rho\sigma_1\sigma_2$ mit Hilfe der Korrelation dargestellt werden und man erhält

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

6.2.2 AR(p)-Modelle und VAR(1)-Darstellung

Wir haben im vorangegangenen Abschnitt festgestellt, dass das Systemmodell eine *Markov*-Struktur besitzt. Bedeutet das nun, dass nur AR(1)-Modelle im Rahmen der Zustandsraumbetrachtung analysiert werden können? Nein. Hier kommt die Matrix/Vektor-Erweiterung ins Spiel. Betrachten Sie folgendes Zustandsraummodell:

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \epsilon_t \quad (6.15)$$

$$z_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}. \quad (6.16)$$

Lassen Sie uns zunächst die Komponenten analysieren. Vergleichen wir (6.15) und (6.16) mit (6.11) und (6.12) auf Seite 123, erhalten wir die Komponenten

$$\begin{aligned} \mathbf{y}_t &= \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} \Omega & 0 \\ 0 & 0 \end{pmatrix}, \\ \mathbf{H} &= \begin{pmatrix} 1 & 0 \end{pmatrix}, d = 0 \text{ und } R = 0. \end{aligned} \quad (6.17)$$

(6.15) und (6.16) bilden also ein gültiges Zustandsraummodell. Lösen wir die einzelnen Gleichungen auf

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \quad (6.18)$$

$$y_{t-1} = y_{t-1} \quad (6.19)$$

$$z_t = y_t. \quad (6.20)$$

Gleichung (6.19) ist lediglich eine Identität und (6.20) sagt uns, dass der latente Systemzustand y_t direkt beobachtet werden kann, die Messfehlervarianz R war schließlich null. Wir erhalten also kurzgesagt das AR(2)-Modell in z_t

$$\phi(B)z_t = \epsilon_t \quad \text{mit} \quad \phi(B) = 1 - \phi_1 B - \phi_2 B^2. \quad (6.21)$$

Man bezeichnet die Darstellung in (6.15) als VAR(1)-Darstellung (Vector Auto Regressive). Wir sind dieser Darstellung bereits in Abschnitt 2.4.2 begegnet, wo es um die Stationarität des AR(2)-Prozesses ging. Wir haben dort gesehen, dass alle benötigten Informationen in der Systemmatrix \mathbf{A} stecken. Berechnet man das charakteristische Polynom aus der Eigenwertgleichung

$$\det[\mathbf{A} - \lambda \mathbf{I}] = -\lambda(\phi_1 - \lambda) - \phi_2 = \lambda^2 - \phi_1 \lambda - \phi_2, \quad (6.22)$$

erhält man als Lösung von (6.22) die inversen Wurzeln (inverted roots). Wenn Ihnen das Eigenwertproblem oder das Lösen quadratischer Gleichungen nicht mehr im Gedächtnis ist, wiederholen Sie kurz die Einschübe auf Seite 40 und 41 und den Text dazwischen. Mit der Herkunft der inverted roots haben wir nun ein weiteres offenes Ende Verknüpft.

Jeder beliebige $AR(p)$ -Prozess kann als $VAR(1)$ -Prozess dargestellt werden. Die allgemeine Form des Zustandsraummodells für $AR(p)$ -Prozesse ist

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-(p-1)} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \epsilon_t \quad (6.23)$$

$$z_t = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-(p-1)} \end{pmatrix}. \quad (6.24)$$

Die Darstellung in Zustandsraumform ist jedoch nicht eindeutig, beispielsweise hätte die $AR(2)$ -Spezifikation (6.15) und (6.16) auch folgendermaßen geschrieben werden können

$$\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \phi_2 & \phi_1 \end{pmatrix} \begin{pmatrix} y_{t-2} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \epsilon_t \quad (6.25)$$

$$z_t = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix}, \quad (6.26)$$

vgl. Hauptskript.

Aufgabe 6.2

Schreiben Sie beide Varianten des Zustandsraummodells für den $AR(3)$ -Prozess.

Aufgabe 6.3

Formulieren Sie das zu (6.23) und (6.24) alternative Zustandsraummodell für den $AR(p)$ -Prozess.

6.2.3 ARMA-Modelle in Zustandsraumform

$ARMA(p, q)$ -Modelle können ebenfalls vollständig im Rahmen der Zustandsraumformulierung dargestellt werden. Wir beginnen wieder mit einem kon-

KAPITEL 6. ZUSTANDSRAUMMODELLE UND KALMAN-FILTER

kreten Beispiel. Betrachten Sie folgendes Zustandsraummodell

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \epsilon_t \quad (6.27)$$

$$z_t = (1 \quad \theta_1) \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}. \quad (6.28)$$

Lassen Sie uns die Terme einzeln analysieren. In (6.27) liefert lediglich die erste Gleichung das AR(2)-Modell, die zweite Gleichung ist nur eine Identität, vgl. (6.15) und (6.16). Lassen Sie uns diese Gleichung mit Backshift-Operatoren schreiben und nach y_t auflösen

$$\begin{aligned} y_t &= (\phi_1 B + \phi_2 B^2) y_t + \epsilon_t \\ &= \frac{1}{1 - \phi_1 B - \phi_2 B^2} \epsilon_t. \end{aligned} \quad (6.29)$$

Gleichung (6.28) kann ebenfalls mit Hilfe des Backshift-Operators geschrieben werden. Man erhält

$$\begin{aligned} z_t &= (1 + \theta_1 B) y_t \\ &= \frac{1 + \theta_1 B}{1 - \phi_1 B - \phi_2 B^2} \epsilon_t, \end{aligned} \quad (6.30)$$

nach Einsetzen von (6.29). Multiplikation mit dem AR-Polynom $1 - \phi_1 B - \phi_2 B^2$ führt zu einem überraschenden aber bekannten Ergebnis

$$z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2} = \epsilon_t + \theta_1 \epsilon_{t-1}, \quad (6.31)$$

dem ARMA(2,1)-Modell in z_t . Beachten Sie: Damit alle Restriktionen richtig verknüpft werden können, muss der Systemzustand die Dimension $d = \max[p, q + 1]$ aufweisen! Die eventuell nicht vorhandenen AR- oder MA-Parameter müssen null gesetzt werden.

Beispiel 6.1: ARMA(2,2)-Zustandsraumform

Die benötigte Dimension des Zustandsraums ergibt sich aus $d = \max[2, 2 + 1] = 3$. Man formuliert

$$\begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \epsilon_t$$

$$z_t = (1 \quad \theta_1 \quad \theta_2) \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix}.$$

Der nicht vorhandene AR-Parameter ϕ_3 wurde null gesetzt.

Diese äußerst elegante Form wird als „phase canonical form“ bezeichnet². Sie können am Ende des Abschnitts im Rahmen einer Übungsaufgabe überprüfen, dass die Formulierung des ARMA(2,2)-Modells in Beispiel 6.1 in der Tat korrekt ist. Die allgemeine „phase canonical form“ für das ARMA(p, q)-Modell lautet

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-(d-1)} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_d \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-d} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \epsilon_t \quad (6.32)$$

$$z_t = \begin{pmatrix} 1 & \theta_1 & \theta_2 & \dots & \theta_{d-1} \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-(d-1)} \end{pmatrix}. \quad (6.33)$$

Aufgabe 6.4

Zeigen Sie durch Explizieren der Gleichungen, dass das Zustandsraummodell in Beispiel 6.1 auf der vorherigen Seite einem ARMA(2,2)-Modell in z_t entspricht.

Aufgabe 6.5

Formulieren Sie folgende Modelle in Zustandsraumform:

- a) MA(2)
- b) ARMA(3,1).

6.3 Kalman-Filter

Bisher haben wir uns darauf beschränkt, bekannte Modellklassen in Zustandsraumform umzuschreiben. Wir haben gesehen, dass eine solche äquivalente Formulierung immer dann möglich ist, wenn der Messfehler verschwindet. Ist jedoch ein Messfehler involviert, gilt nicht mehr $z_t = y_t$. Der Systemzustand ist nun latent und kann nicht mehr direkt beobachtet werden. An dieser Stelle wird ein Mechanismus, oder besser ein Algorithmus benötigt, der aus der Beobachtung z_t die maximale Information über den Zustand y_t herausfiltert. Dieser Algorithmus ist der *Kalman*-Filter.

²Es gibt auch alternative Formulierungen des ARMA-Modells, siehe Hauptskript.

KAPITEL 6. ZUSTANDSRAUMMODELLE UND KALMAN-FILTER

Bevor wir beginnen, lassen Sie uns noch ein paar nützliche Konventionen zur Notation besprechen. Das Zustandsraummodell (6.11) und (6.12) auf Seite 123 wurde in skalarer Notation geschrieben. Die Parameter sind aber bereits durch Groß- und Kleinbuchstaben so repräsentiert, dass eine Erweiterung in Matrix/Vektor-Form intuitiv klar ist. Wir wollen diese hybride Notation von nun an nutzen, um alle folgenden Aussagen in einer Weise zu formulieren, die gleichsam für skalare wie vektorielle Problemstellungen Gültigkeit besitzt. Folgende Übersicht fasst die wichtigsten Unterscheidungen zusammen.

	Hybrid	Skalar	Matrix/Vektor
Inneres Produkt	$a'b$	$ab = c$	$\mathbf{a}'\mathbf{b} = c$
Äußeres Produkt	ab'	$ab = c$	$\mathbf{a}\mathbf{b}' = \mathbf{C}$
Matrix-Produkt	AB	$ab = ba$	$\mathbf{AB} \neq \mathbf{BA}$
Inverse Matrix	$A^{-1}A$	$\frac{1}{a}a = 1$	$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
Transposition	$(AB)'$	$ab = ba$	$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

Tabelle 6.1: Übersicht zur Hybridnotation

Beachten Sie stets, dass Matrizen bzw. Vektoren die passenden Dimensionen aufweisen müssen, damit ein Produkt definiert ist. Weiterhin kann nur eine quadratische, nicht singuläre Matrix invertiert werden. Diese Bedingung ist das Pendant zum Verbot des Teilens durch Null.

Als erste Anwendung der Hybridnotation werden wir das Konzept der linearen Transformation von Seite 91 präzisieren.

Theorem 6.1: Affine Transformation

Sei X eine (vektorielle) Zufallsvariable und $Y = a + BX$, mit beliebigen Konstanten a und B , dann wird Y eine affine Transformation von X genannt. Y ist ebenfalls eine Zufallsvariable und es gilt:

1. $E[Y] = a + BE[X]$
2. $\text{Var}[Y] = B\text{Var}[X]B'$

Weiterhin gilt insbesondere: Ist X (multivariat) normalverteilt, dann ist Y ebenfalls (multivariat) normalverteilt.

Für skalare Zufallsvariablen reduziert die affine Transformation sich wieder auf die Definition der linearen Transformation von Seite 91. Im Fall von vektorwertigen Zufallsvariablen sind a und B passende Vektoren respektive Matrizen.

6.3.1 Die *Bayes*-Formel

Wir haben uns in den vergangenen Kapiteln schon oft mit bedingten Größen auseinandergesetzt, dabei aber eine rigorose Definition umgangen, indem wir die entsprechende Bedingung formuliert haben (meistens wurde auf Information aus vergangenen Perioden bedingt). Die *Bayes*-Formel wird uns einen Einblick in bedingte Wahrscheinlichkeiten und Verteilungen auf wesentlich elementarerer Ebene gewähren. Wir werden als erstes die *Bayes*-Formel anschreiben und dann anhand von Beispielen und Erweiterungen aufdecken, wie weitreichend ihre Konsequenzen sind.

Angenommen A und B bezeichnen zwei Ereignisse, und $A|B$ (A gegeben B) sei der Eintritt von Ereignis A , unter der Bedingung, dass auch Ereignis B eintritt oder bereits eingetreten ist. Die *Bayes*-Formel macht nun folgende Aussage über die Eintrittswahrscheinlichkeit des bedingten Ereignisses

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (6.34)$$

Lassen Sie uns an einem Beispiel beleuchten, was diese Formel leisten kann.

Beispiel 6.2: *Bayes*-Formel

Jeder Student der FernUni besucht zumindest einmal im Rahmen eines Seminars die schöne Stadt Hagen. Die Winter im Sauerland können sich sehen lassen. Insbesondere kommt es im Schnitt an 60 Tagen zu Verkehrsbehinderungen durch Schneefall. Insgesamt entfallen 95% der Schneefälle auf die Wintermonate November bis Februar. Wie hoch ist die Wahrscheinlichkeit bei einem Seminar im Januar von Schneeglätte überrascht zu werden?

$$\begin{aligned} P(\text{Schnee}|\text{Winter}) &= \frac{P(\text{Winter}|\text{Schnee})P(\text{Schnee})}{P(\text{Winter})} \\ &= \frac{\frac{95}{100} \cdot \frac{60}{365}}{\frac{120}{365}} = \frac{19}{40} < \frac{1}{2}. \end{aligned}$$

Der risikoneutrale Student würde also auf den Kauf von Winterreifen verzichten.



KAPITEL 6. ZUSTANDSRAUMMODELLE UND KALMAN-FILTER

Die *Bayes*-Formel erstreckt sich in analoger Form auf Wahrscheinlichkeitsdichten. Seien Y und Z Zufallsvariablen, dann gilt für die bedingte Wahrscheinlichkeitsdichte

$$p(y|z) = \frac{p(z|y)p(y)}{p(z)}. \quad (6.35)$$

Die Variablennamen wurden hier natürlich schon in suggestiver Weise gewählt, um den Zusammenhang mit dem Zustandsraummodell herzustellen. Normalerweise nimmt der Ausdruck auf der rechten Seite von (6.35) eine komplizierte Form an, da hier verschiedene Dichten miteinander multipliziert bzw. dividiert werden. Sind jedoch alle involvierten Wahrscheinlichkeitsdichten *Gaußsch*, ist auch die resultierende bedingte Dichte eine *Gauß*-Dichte, die sich durch Erwartungswert(vektor) und Kovarianz(matrix) repräsentieren lässt. Das folgende, äußerst wichtige Theorem gibt die Momente der bedingten Verteilung an:

Theorem 6.2: Normal-Korrelation

Für multivariat normalverteilte Zufallsvariablen Y und Z ,

$$\begin{pmatrix} Y \\ Z \end{pmatrix} \sim N \left(\begin{pmatrix} E[Y] \\ E[Z] \end{pmatrix}, \begin{pmatrix} \text{Var}[Y] & \text{Cov}[Y, Z] \\ \text{Cov}[Z, Y] & \text{Var}[Z] \end{pmatrix} \right),$$

ist auch die bedingte Verteilung von $Y|Z$ eine Normalverteilung mit den bedingten Momenten

$$E[Y|Z] = E[Y] + \text{Cov}[Y, Z]\text{Var}[Z]^{-1}(Z - E[Z])$$

$$\text{Var}[Y|Z] = \text{Var}[Y] - \text{Cov}[Y, Z]\text{Var}[Z]^{-1}\text{Cov}[Z, Y].$$

Die Normal-Korrelation liefert also die Momente der bedingten Dichte auf der linken Seite von (6.35), für den Fall, dass alle involvierten Verteilungen normal sind. Beachten Sie, dass allgemein $\text{Cov}[Y, Z] = \text{Cov}[Z, Y]'$ gilt, da die Kovarianz möglicherweise eine (rechteckige) Matrix ist!

6.3.2 Mess-Update

Wir sind nun schon fast in der Lage das *Kalman*- oder Mess-Update zu formulieren. Rufen wir uns noch einmal das Zustandsraummodell ins Gedächtnis

$$y_t = Ay_{t-1} + b + \epsilon_t \quad (6.36)$$

$$z_t = Hy_t + d + \delta_t. \quad (6.37)$$

Im Augenblick nehmen wir an, dass y_t $N(\mu_t, \Sigma_t)$ -verteilt ist. Wir werden später die notwendige Bedingung dafür angeben. Dann müssen wir laut Theorem 6.2 auf der vorherigen Seite lediglich $E[z_t]$, $\text{Cov}[y_t, z_t]$ und $\text{Var}[z_t]$ berechnen. Die Erwartungswert- und Varianzterme sind trivial, da $Hy_t + d$ eine affine Transformation ist (Einschub auf Seite 130). Man erhält

$$E[z_t] = HE[y_t] + d + E[\delta_t] = H\mu_t + d. \quad (6.38)$$

Beim Varianzterm wird ausgenutzt, dass der Fehler δ_t unkorreliert ist. Daher ergibt sich

$$\text{Var}[z_t] = H\text{Var}[y_t]H' + \text{Var}[\delta_t] = H\Sigma_tH' + R. \quad (6.39)$$

Der Kovarianzterm ist ein bisschen widerspenstiger. Wir können aber erneut ausnutzen, dass der Messfehler mit keiner anderen Modellkomponente korreliert und erhalten

$$\begin{aligned} \text{Cov}[y_t, z_t] &= \text{Cov}[y_t, Hy_t + d + \delta_t] \\ &= \text{Cov}[y_t, Hy_t + d] + \underbrace{\text{Cov}[y_t, \delta_t]}_{=0} \\ &= E[(y_t - E[y_t])(Hy_t + d - HE[y_t] - d)'] \\ &= E[(y_t - E[y_t])(y_t - E[y_t])'H'] \\ &= \text{Var}[y_t]H' \\ &= \Sigma_tH'. \end{aligned} \quad (6.40)$$

Um das gesamte *Kalman*-Update algorithmisch in einfacherer Form schreiben zu können wird der Term $\text{Cov}[y_t, z_t]\text{Var}[z_t]^{-1}$ zu

$$K_t = \Sigma_tH'(H\Sigma_tH' + R)^{-1} \quad (6.41)$$

zusammengefasst. K_t bezeichnet man als *Kalman*-Gain. Vereinfacht gesprochen wird die Information der Messung im *Kalman*-Gain kumuliert. Es fungiert dann gewissermaßen als Filtergewicht, mit dem die Momente des latenten Zustands y_t korrigiert werden, bedingt auf die Messung. Lassen Sie uns das Mess-Update in seiner endgültigen Form resümieren

$$\mu_{t|t} = \mu_t + K_t(z_t - H\mu_t - d) \quad (6.42)$$

$$\Sigma_{t|t} = \Sigma_t - K_tH\Sigma_t \quad (6.43)$$

Die Schreibweise $\mu_{t|t}$ bzw. $\Sigma_{t|t}$ ist eine Kurzform für $E[y_t|z_t]$ und $\text{Var}[y_t|z_t]$. Bemerkung: Erinnern Sie sich, dass für symmetrische Matrizen $A = A'$ gilt. Varianzen sind generell symmetrisch, weil sie aus äußeren Produkten bestehen. Daher folgt für den Kovarianzterm in (6.43) $\text{Cov}[z_t, y_t] = \text{Cov}[y_t, z_t]' = H\Sigma_t' = H\Sigma_t$.

KAPITEL 6. ZUSTANDSRAUMMODELLE UND KALMAN-FILTER

Lassen Sie uns hier kurz verweilen und zusammenfassen, was wir bisher erreicht haben:

- Wir haben gesehen, dass das *Kalman*-Update die a priori Momente (Momente vor der Messung) μ_t und Σ_t als Eingangssignal verwendet und sie einem linearen Messprozess unterwirft. Das Ausgangssignal sind die a posteriori Momente (Momente nach der Messung) $\mu_{t|t}$ und $\Sigma_{t|t}$, die entlang der *Bayes*-Formel generiert werden.
- Als Filtergewicht fungiert der *Kalman*-Gain (6.41). Dieses Gewicht liefert die optimale lineare Schätzung für y_t im Hinblick auf einen minimalen Fehler im Quadratmittel (siehe Hauptskript). Der *Kalman*-Gain enthält die Kovarianz zwischen y_t und z_t . Sind Systemzustand und Messung unkorreliert, folgt $K_t = 0$ und die a priori Momente werden nicht korrigiert, da keine Information aus der Messung gewonnen werden kann.
- Der Filter wirkt bei der Korrektur des Erwartungswertes in (6.42) nur auf den nicht antizipierten Teil $z_t - E[z_t]$. Dieser Teil bildet die sog. Innovation. Wir werden an späterer Stelle noch einmal auf diesen Punkt zu sprechen kommen.
- Der Filterprozess wirkt varianzreduzierend, vgl. (6.43). Durch die Information, die aus der Messung gewonnen wird, wird nicht nur die Schätzung des latenten Systemzustands verbessert, sondern auch die Ungewissheit reduziert. Ist der Messfehler $R = 0$, kollabiert $\Sigma_{t|t}$ sogar vollständig; der Systemzustand y_t wird beobachtbar.

Aufgabe 6.6

Gegeben seien folgende Parameter des Messmodells

$$H = \begin{pmatrix} 1 & 0 \end{pmatrix}, d = 1 \text{ und } R = 1.$$

Weiterhin seien a priori Momente und Messung durch

$$\mu_t = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \Sigma_t = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \text{ und } z_t = 2$$

gegeben. Berechnen Sie den *Kalman*-Gain K_t , sowie die a posteriori Momente $\mu_{t|t}$ und $\Sigma_{t|t}$.

Wenn Sie ein korrektes Ergebnis für Aufgabe 6.6 berechnet haben werden Sie feststellen, dass zwar nur die erste Komponente des Systemzustands gemessen wurde (zweidimensionaler latenter Prozess), aber trotzdem beide Komponenten durch die Information der Messung korrigiert wurden. Mehr noch, die a posteriori Varianz der gemessenen Komponente ist sogar kleiner als die Messfehlervarianz. Verblüffend, nicht wahr?

6.3.3 Innovation und effizientes *Kalman*-Update

In diesem kurzen Abschnitt soll die Berechnung des Mess-Updates optimiert werden. Die resultierende Schreibweise wird sich zu einem späteren Zeitpunkt als überaus nützlich erweisen. Zusätzlich kann Rechenaufwand eingespart werden. Wir haben gesehen, dass der *Kalman*-Gain in Gleichung (6.42) lediglich auf die Innovation $z_t - E[z_t]$ wirkt. Deshalb definieren wir formal

$$\nu_t = z_t - H\mu_t - d. \quad (6.44)$$

Die Innovation hat Erwartungswert null, $E[z_t - E[z_t]] = 0$, und Varianz $\text{Var}[z_t - E[z_t]] = \text{Var}[z_t]$. Beachten Sie, dass $E[z_t]$ eine fixe Größe ist, keine Zufallsvariable, und damit für die Varianz keine Rolle spielt. Man schreibt für die Varianz von ν_t formal

$$\Gamma_t = H\Sigma_t H' + R. \quad (6.45)$$

Lassen Sie uns nun sehen, welche Update-Gleichungen wir mit diesen Definitionen erhalten

$$K_t = \Sigma_t H' \Gamma_t^{-1} \quad (6.46)$$

$$\mu_{t|t} = \mu_t + K_t \nu_t \quad (6.47)$$

$$\Sigma_{t|t} = \Sigma_t - K_t \Gamma_t K_t'. \quad (6.48)$$

Die letzte Gleichung wurde durch Erweitern mit $\Gamma_t \Gamma_t^{-1}$ gewonnen. Warum ist die Form (6.44) bis (6.48) effizienter? In der Regel besitzt das Messmodell weniger Dimensionen als das Systemmodell. Wird beispielsweise nur eine Systemkomponente beobachtet, sind ν_t und Γ_t sogar Skalare. Dadurch wird die Berechnung des Filter-Gains und der a posteriori Momente natürlich deutlich vereinfacht.

Aufgabe 6.7

Betrachten Sie noch einmal das Modell aus Aufgabe 6.6 auf der vorherigen Seite. Berechnen Sie ν_t , Γ_t und K_t , sowie die a posteriori Momente $\mu_{t|t}$ und $\Sigma_{t|t}$ entlang (6.44) bis (6.48).

6.3.4 Time-Update und *Kalman*-Filter-Algorithmus

Wir haben beim *Kalman*-Update angenommen, dass die a priori Verteilung von y_t eine Normalverteilung ist. Diese Annahme steckt implizit in der Verwendung der Normal-Korrelation. Was ist aber notwendig, damit diese Annahme auch richtig ist? Nehmen wir einmal an, dass der Initialzustand y_0 normalverteilt ist. Die Zustandsgleichung

$$y_1 = Ay_0 + b + \epsilon_1 \quad (6.49)$$

besteht aus einer affinen Transformation von y_0 und einer weiteren unkorreliert normalverteilten Zufallsvariable ϵ_1 . Damit ist y_1 ebenfalls normalverteilt, mit den a priori Momenten

$$\mu_1 = A\mu_0 + b \quad (6.50)$$

$$\Sigma_1 = A\Sigma_0A' + \Omega. \quad (6.51)$$

Damit habe wir den Kreis schon geschlossen. Wird zum Zeitpunkt $t = 1$ ein Mess-Update durchgeführt, ist $y_{1|1}$ ebenfalls wieder normalverteilt und nimmt den Platz von y_0 in der Zustandsgleichung (6.49) ein. Es ist also lediglich notwendig zu fordern, dass der Anfangszustand y_0 normalverteilt ist. Algorithmus 6.1 fasst den gesamten *Kalman*-Filter zusammen.

Algorithmus 6.1: *Kalman*-Filter

Anfangsbedingung μ_0, Σ_0 ▷ Initialisierung

for $t = 1, \dots, T$ **do**

Time-Update

$$\mu_{t|t-1} = A\mu_{t-1|t-1} + b$$

$$\Sigma_{t|t-1} = A\Sigma_{t-1|t-1}A' + \Omega$$

Mess-Update

$$\nu_t = z_t - H\mu_{t|t-1} - d \quad \text{▷ Innovation}$$

$$\Gamma_t = H\Sigma_{t|t-1}H' + R$$

$$K_t = \Sigma_{t|t-1}H'\Gamma_t^{-1} \quad \text{▷ Kalman-Gain}$$

$$\mu_{t|t} = \mu_{t|t-1} + K_t\nu_t$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t\Gamma_tK_t'$$

end for

Lassen Sie uns etwas genauer beleuchten, welche Möglichkeiten wir beim Festlegen der Anfangsbedingungen haben. Ist y_0 tatsächlich normalverteilt wird der Filter einfach mit μ_0 und Σ_0 initialisiert. Was passiert aber, wenn $y_0 = c$ bekannt ist, also keine Zufallsvariable mehr ist? In diesem Fall liegt eine **deterministische Anfangsbedingung** vor und man setzt

$$\mu_0 = c \quad \text{und} \quad \Sigma_0 = 0. \quad (6.52)$$

Vereinfacht gesprochen verwendet man eine degenerierte Normalverteilung, deren Varianz null ist. Der umgekehrte Fall ist ebenfalls denkbar. Angenommen wir haben keinerlei Information über die Anfangsverteilung, dann ist es möglich, eine Normalverteilung mit sehr großer Streuung (bspw. 10^{16}) zu wählen, die de facto so flach ist, dass sie völlig nicht-informativ ist. Eine solche Bedingung wird als **diffuse Anfangsbedingung** bezeichnet,

$$\mu_0 = 0 \quad \text{und} \quad \Sigma_0 = 10^{16} I. \quad (6.53)$$

Ist der Systemprozess stationär, kann der *Kalman*-Filter beispielsweise auch mit den stationären Momenten initialisiert werden³. Es wird nun Zeit ein wenig Praxis im Handling des *Kalman*-Filters zu bekommen.

Aufgabe 6.8

Gegeben Sei der latente ARMA(2,1)-Prozess

$$y_t = y_{t-1} - 2y_{t-2} + \epsilon_t + \frac{1}{2}\epsilon_{t-1}.$$

Nehmen Sie an, dass ϵ_t normalverteilt ist, mit Varianz $\Omega = 2$. Weiterhin ist y_t nicht direkt beobachtbar, sondern wird von einem standard-normalverteilten Messfehler überlagert. Zwei Messwerte liegen vor:

$$z_1 = \frac{3}{2} \quad \text{und} \quad z_2 = \frac{1}{2}.$$

Formulieren Sie das zugehörige Zustandsraummodell in „phase canonical form“ und berechnen Sie die bedingten Momente $\mu_{2|2}$ und $\Sigma_{2|2}$. Initialisieren Sie den *Kalman*-Filter mit der deterministischen Anfangsbedingung $\mu_0 = (1, 0)'$.

³Zur Berechnung der stationären Momente werden Vektorisierungsoperatoren und die Spektraldarstellung benötigt. Die explizite Vorgehensweise ist im Hauptskript erläutert.

6.3.5 Zustandsschätzung und Prognose

Wenn Sie bis hierher durchgehalten haben, ist der Rest ein Klacks. Wir werden nun lediglich noch einige Aspekte des *Kalman*-Filters genauer beleuchten, um ein geordnetes Verständnis seiner Eigenschaften zu erzeugen.

Was war der gravierende Unterschied zur Diskussion von Zeitreihen in den vorangegangenen Kapiteln? Wir haben die Annahme aufgegeben, dass der Systemprozess, der die relevante Zeitreihe erzeugt, direkt beobachtbar ist. Infolge dessen konnten wir keine präzisen Aussagen mehr über den Zustand des Prozesses zu einem bestimmten Zeitpunkt machen, sondern nur noch über seine Verteilung (genauer seine Momente). Der *Kalman*-Filter hat uns eine optimale Schätzung des Systemzustands ermöglicht. Diese Zustandsschätzung muss von der Parameterschätzung, mit der wir vorher schon zu tun hatten, scharf abgegrenzt werden. Eine Parameterschätzung kann auch mit dem *Kalman*-Filter durchgeführt werden, wir werden aber auf diesen Punkt erst später zu sprechen kommen.

Abbildung 6.2 zeigt eine solche Zustandsschätzung per *Kalman*-Filter für ein ARMA(2,1)-Modell, ähnlich dem aus Aufgabe 6.8. Die einzelnen Aspekte sollen im Folgenden erläutert werden:

- Der bedingte Erwartungswert $\mu_{t|t}$ bzw. $\mu_{t|t-1}$ ist grün eingezeichnet. An den Stellen, an denen neue Messinformation verfügbar ist, wird die Zustandsschätzung in unstetiger Weise korrigiert. Sie erkennen kleinere

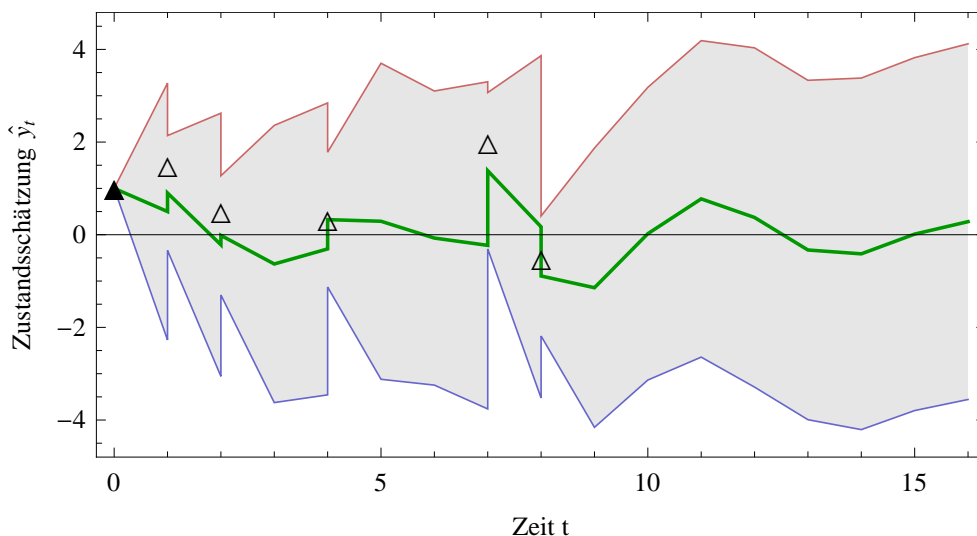


Abbildung 6.2: *Kalman*-Filter Zustandsschätzung und 95% HPD-Intervall

oder auch größere Sprünge in der Kurve, je nach dem, ob die Messung nahe am prognostizierten Messwert liegt oder weit davon entfernt.

- Die Messungen selbst sind durch Dreiecke repräsentiert. Der Prozess wurde mit einer deterministischen Anfangsbedingung initialisiert, daher ist der Anfangswert gewissermaßen eine Messung ohne Messfehler. Die Messungen liegen unter Umständen weit ab von der Zustandsschätzung, da sie von einem Messfehler überlagert sind. Genaugenommen hängt die Zustandsschätzung von allen vorangegangenen Messungen $Z^t = z_1, \dots, z_t$ ab. Der *Kalman*-Filter generiert also für $s \geq t$ die bedingten Momente $E[y_s|Z^t]$ und $\text{Var}[y_s|Z^t]$. Die Notation $\mu_{s|t}$ und $\Sigma_{s|t}$ ist also in diesem Sinne zu lesen.
- Für $s > t$ generiert der *Kalman*-Filter eine Prognose. Innerhalb des „normalen“ Durchlaufs der Filteriterationen handelt es sich dabei immer um eine Einschnitt-Prognose, die anschließend durch ein Mess-Update korrigiert wird. Für $s > T$ sind aber keine Messungen mehr verfügbar. In diesem Fall liefern die nicht vorhandenen Messungen auch keine Information und der *Kalman*-Gain ist $K_s = 0$. Ein kurzer Blick in Algorithmus 6.1 auf Seite 136 zeigt, dass die bedingten Momente in diesem Fall im Mess-Update nicht korrigiert werden. Dasselbe gilt für fehlende Messungen (Missing Data). In Abbildung 6.2 fehlt bspw. die Messung zum Zeitpunkt $t = 3$. Das bedeutet $K_3 = 0$ und die a priori Momente sind gleichzeitig die neuen a posteriori Momente.
- Die grau unterlegten Bereiche sind 95% HPD-Bereiche (Highest Posterior Density). Das HPD-Intervall ist dem Konfidenzintervall ähnlich, es besteht jedoch ein subtiler Unterschied zwischen beiden. Das Konfidenzintervall wird gebildet, wenn ein fixer aber unbekannter Parameter geschätzt wird. Hier wurde eine Zufallsvariable geschätzt, die ihrerseits eine Verteilung besitzt. Daher muss das HPD-Intervall verwendet werden. Die Berechnung ist in diesem Fall analog zum Konfidenzintervall bei Normalverteilung, nur dass für die Varianz des Schätzers die entsprechende Komponente von $\Sigma_{s|t}$ verwendet wird⁴. Beachten Sie, dass die Breite der HPD-Bereiche in Abbildung 6.2 beim Eintreffen neuer Messinformation deutlich reduziert wird.

⁴In einem zweidimensionalen Zustandsraum wird y_s durch $\hat{y}_{s|t} \sim N(\mu_{s|t}, \Sigma_{s|t})$, mit

$$\mu_{s|t} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{und} \quad \Sigma_{s|t} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

geschätzt. Damit gilt für die erste Komponente von y_s das 95% HPD-Intervall $\mu_1 \pm 1.96\sqrt{\sigma_{11}}$. Das HPD-Intervall für die zweite Komponente ergibt sich analog.

6.3.6 Parameterschätzung

Die Parameterschätzung mit Maximum-Likelihood kann im Rahmen von Zustandsraummodellen über die sog. Prognose-Fehler-Zerlegung abgewickelt werden. Im Grunde wird dort ausgenutzt, dass die Innovationen unkorreliert und (multivariat) normalverteilt sind⁵. Die Likelihoodfunktion ist wieder eine Funktion der Parameter, und zwar aller Parameter des System- und Messmodells

$$L(A, b, \Omega, H, d, R) = \prod_{t=1}^T \frac{1}{\sqrt{\det[2\pi\Gamma_t]}} e^{-\frac{1}{2}\nu_t'\Gamma_t^{-1}\nu_t}. \quad (6.54)$$

In der Praxis werden so gut wie nie die kompletten, unrestringierten Matrizen bzw. Vektoren geschätzt. Beispielsweise gilt für ein AR(2)-Systemmodell

$$A = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}. \quad (6.55)$$

Die unrestringierte Matrix A hat vier zu bestimmende Komponenten, während die restringierte Form (6.55) lediglich zwei unbekannte Parameter enthält, ϕ_1 und ϕ_2 .

Wird die Likelihoodfunktion (6.54) wieder in ihrer logarithmierten Form maximiert, ergibt sich eine besonders vorteilhafte Situation, da über die einzelnen Likelihoodbeiträge summiert wird

$$\begin{aligned} l(A, b, \Omega, H, d, R) &= \sum_{t=1}^T l_t \\ &= \sum_{t=1}^T -\frac{1}{2}(\log \det[2\pi\Gamma_t] + \nu_t'\Gamma_t^{-1}\nu_t). \end{aligned} \quad (6.56)$$

Die Inkredenzien ν_t und Γ_t des LogLikelihood-Beitrags l_t werden aber vom *Kalman*-Filter bei jedem Mess-Update quasi als Nebenprodukt berechnet, vgl. Algorithmus 6.1 auf Seite 136. Jetzt ist auch klar, warum diese algorithmische Form so effizient ist. Man erhält die zur Parameterschätzung benötigten Terme sozusagen kostenlos.

⁵Wir werden in Abschnitt 6.4 einen Fall kennenlernen, in dem die Innovationen nicht normalverteilt sind, da der Messfehler zwar normalverteilt angenommen wird, in Wirklichkeit aber eine andere Verteilung besitzt. Das resultierende Schätzverfahren nennt man Quasi-Maximum-Likelihood. Es liefert trotz falscher Verteilungsannahmen asymptotisch erwartungstreue Schätzer.

Aufgabe 6.9

Betrachten Sie noch einmal Aufgabe 6.8 auf Seite 137. Berechnen Sie den Wert der logarithmierten Likelihood-funktion $l = l_1 + l_2$. Hinweis: Für Skalare gilt $\det A = a$.

6.4 ARV-Modelle

Wir kommen nun zu einer Modellklasse, die für gewöhnlich extreme Probleme hinsichtlich der Parameterschätzung verursacht. Diese Probleme lassen sich aber erheblich durch den Einsatz des *Kalman*-Filters vereinfachen. Es handelt sich um die Klasse der ARV-Modelle⁶ (Autoregressive Random Volatility). Lassen Sie uns zunächst das Modell formulieren und dann seine Komponenten analysieren

$$y_t = \mu + e^{\frac{h_t}{2}} z_t \quad (6.57)$$

$$h_t - \bar{h} = \lambda(h_{t-1} - \bar{h}) + \gamma \nu_t. \quad (6.58)$$

Die erste Gleichung besitzt enorme Ähnlichkeit mit dem EGARCH-Modell (5.56) aus Abschnitt 5.3.5, lediglich ergänzt um einen Erwartungswert μ . z_t ist hier wieder eine standard-normalverteilte Innovation⁷. Es liegt mit (6.57) also ein bedingt heteroskedastischer Prozess vor. Die Varianzgleichung (6.58) ist eine Überraschung. Es handelt sich strukturell um einen AR(1)-Prozess in h_t , mit Mittelwert \bar{h} , AR-Parameter λ und Zufallsfehler $\nu_t \sim N(0, 1)$. Es wird im Allgemeinen angenommen, dass z_t und ν_t unkorreliert sind.

Wir haben vorher keine Zufallsfehler in der Varianzgleichung gesehen und die Komplikationen, die dieser Fehler mit sich bringt, sind enorm (siehe Hauptskript). Die Verteilung von y_t muss nun nicht nur auf verzögerte Varianzen bedingt werden, sondern auch auf die verzögerten Realisationen des zweiten Zufallsfehlers. Per Definition ist der Varianzprozess aber gar nicht beobachtbar, wir stehen also vor einem Dilemma.

6.4.1 Zustandsraumform des ARV-Modells

Wir haben bereits gesehen, dass unbeobachtbare Prozesse in Zustandsraumform mit Hilfe des *Kalman*-Filters geschätzt werden konnten. Daher lautet die relevante Frage: Können wir das ARV-Modell in Zustandsraumform

⁶Gelegentlich findet man auch die Bezeichnung SVola-Modell (Stochastic Volatility).

⁷Die Notation ist etwas unglücklich, da z_t im Zustandsraummodell für die Messung steht. Diese Notation ist jedoch etabliert und wird deshalb beibehalten.

KAPITEL 6. ZUSTANDSRAUMMODELLE UND KALMAN-FILTER

transformieren? Lassen Sie es uns versuchen. Beginnen wir mit dem latenten Prozess (6.58), der unseren Systemzustand beschreibt. Hier erhalten wir bereits nach trivialen algebraischen Manipulationen

$$h_t = \lambda h_{t-1} + (1 - \lambda)\bar{h} + \epsilon_t, \quad (6.59)$$

mit $\epsilon_t \sim N(0, \gamma^2)$. Das war leicht, wir können sofort die Zustandsparameter identifizieren: $A = \lambda$, $b = (1 - \lambda)\bar{h}$ und $\Omega = \gamma^2$. Jetzt kommt der schwere Teil. Bringen wir zunächst den Erwartungswert μ auf die linke Seite von Gleichung (6.57). Anschließend quadrieren wir beide Seiten, um einen Term in e^{h_t} zu bekommen, und bilden den Logarithmus. Wir erhalten

$$\log[(y_t - \mu)^2] = h_t + \log z_t^2. \quad (6.60)$$

Das sieht schon relativ brauchbar aus. Wir können eine modifizierte Messung $z_t^* = \log[(y_t - \mu)^2]$ definieren, dazu muss lediglich die ursprünglich beobachtete Zeitreihe mittelwertbereinigt, quadriert und anschließend logarithmiert werden. Allerdings haben wir einen äußerst seltsamen Zufallsfehler erhalten, $\log z_t^2$. z_t war ursprünglich standard-normalverteilt, daher gilt $z_t^2 \sim \chi_1^2$. Wir können schon erahnen, dass der Logarithmus von z_t^2 eine sehr unerfreuliche Verteilung haben wird. Diese Verteilung kann berechnet werden, jedoch wollen wir uns nicht damit auseinandersetzen und beschränken uns darauf, ihren Erwartungswert⁸, $E[\log z_t^2] \approx -1.27$, und ihre Varianz, $\text{Var}[\log z_t^2] = \frac{1}{2}\pi^2$, anzugeben. Wir ignorieren nun einfach die Tatsache, dass die Verteilung von $\log z_t^2$ keine Normalverteilung ist und arbeiten einfach mit diesen beiden Momenten, sprich wir definieren $d = -1.27$ und $\delta_t = \log z_t^2 - d$. Die verbleibenden Identifikationen sind $H = 1$ und $R = \frac{1}{2}\pi^2$. Das komplette approximierte Zustandsraummodell ist dann durch

$$h_t = \lambda h_{t-1} + (1 - \lambda)\bar{h} + \epsilon_t \quad (6.61)$$

$$z_t^* = h_t - 1.27 + \delta_t \quad (6.62)$$

gegeben, mit $\epsilon_t \sim N(0, \gamma^2)$ und $\delta_t \sim N(0, \frac{1}{2}\pi^2)$.

Der *Kalman*-Filter berechnet nun mit jedem Mess-Update den jeweiligen LogLikelihood-Beitrag. Ein kompletter Filterdurchlauf generiert also die

⁸Der exakte Erwartungswert lautet $E[\log z_t^2] = -\gamma - \log 2$. Dabei bezeichnet γ die *Euler-Mascheroni*-Konstante, die über die Reihenentwicklung

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \log n \right) \approx 0.5772$$

dargestellt werden kann.

vollständige LogLikelihood-Funktion, die dann maximiert werden kann. Es gibt einen kleinen Haken: die berechnete LogLikelihood-Funktion ist falsch. Erinnern Sie sich, dass die Verteilung von δ_t gar nicht wirklich eine Normalverteilung ist, wir haben das nur angenommen, um das ARV-Modell in die Zustandsraumform zu pressen. Nichtsdestotrotz liefert die ML-Schätzung erwartungstreue Parameterschätzer. Diese Art der eigentlich falschen ML-Schätzung wird Quasi-Maximum-Likelihood Methode genannt. Die resultierenden Schätzer sind zwar noch asymptotisch erwartungstreu, besitzen aber nicht mehr die kleinstmögliche Streuung. Dieser Defekt kann jedoch korrigiert werden, siehe Hauptskript.

Aufgabe 6.10

Schreiben Sie den *Kalman*-Filter Algorithmus speziell mit den Parametern und Werten des ARV-Modells (6.61) und (6.62).

7

Tipps und Tricks

7.1 Zentrale Momente einer Normalverteilung

Die höheren Momente einer Normalverteilung können sehr einfach durch folgende Formel berechnet werden

$$m_k = \begin{cases} m_2^{k/2} (k-1)!! & \text{für } k \text{ gerade} \\ 0 & \text{für } k \text{ ungerade.} \end{cases}$$

Der Ausdruck für gerade k enthält ein doppeltes Fakultätszeichen ‘!!’. Die einfache Fakultät wird in der Regel rekursiv durch $k! = k \cdot (k-1)!$ und die Festsetzung $0! = 1$ definiert. Man erhält also $k! = k \cdot (k-1) \cdot \dots \cdot 2 \cdot 1$. Die formale Schreibweise der Doppelfakultät ist so zu interpretieren, dass die rekursive Definition mit Lag 2 erfolgt, $k!! = k \cdot (k-2) \cdot \dots \cdot 3 \cdot 1$. Der Ausdruck $7!!$ wird also durch $7!! = 7 \cdot 5 \cdot 3 \cdot 1 = 105$ berechnet.

Beispiel 7.1:

Angenommen z_t sei standard-normalverteilt. Welchen Wert erhält man für das sechste zentrale Moment? Zunächst gilt $m_2 = \sigma^2 = 1$, und damit

$$m_6 = 1^3 \cdot 5!! = 5 \cdot 3 \cdot 1 = 15.$$

Alle ungeraden Momente verschwinden definitionsgemäß.

7.2 Einfache Berechnung stationärer Momente

Bei der Berechnung stationärer Momente wird häufig folgende Strategie angewendet. Zunächst wird die betreffende Momentengleichung so umgeschrieben, dass eine rekursive Form entsteht. Diese wird dann rückwärts

vom Zeitpunkt t an bis zu einem bekannten Startwert oder einer bekannten Anfangsbedingung iteriert. Als letzter Schritt wird dann der Grenzwert für $t \rightarrow \infty$ gebildet und eine Reihenentwicklung, bspw. die geometrische Reihe, eingesetzt, um einen expliziten Ausdruck zu erhalten.

Liegen Modelle vor, die nur endliche Lags $k < \infty$ berücksichtigen, kommt man mit einem kleinen „dirty trick“ schnell und bequem zum Ziel. Zuerst werden alle Zeitargumente formal durch ihren Grenzwert ersetzt, das heißt $y_t \rightarrow y_\infty$ usw. Beachten Sie, dass bei verzögerten Werten in unendlich ferner Zukunft das Lag keine Rolle mehr spielt. Das bedeutet formal $y_{t-k} \rightarrow y_\infty$. Im zweiten Schritt wird nun das entsprechende Moment, Erwartungswert oder Varianz etc., gebildet. Auflösen der Gleichung liefert dann unmittelbar das gesuchte stationäre Moment. Dieses Vorgehen lässt sich am besten am Beispiel verdeutlichen.

Beispiel 7.2:

Für den nicht zentrierten AR(1)-Prozesses (2.3) auf Seite 6 mit $|\phi| < 1$ erhält man in unendlicher Zukunft

$$x_\infty = (1 - \phi)\mu + \phi x_\infty + \epsilon_\infty.$$

Für Erwartungswert und Varianz erhält man folgende Gleichungen

$$\begin{aligned} E[x_\infty] &= (1 - \phi)\mu + \phi E[x_\infty] \\ \text{Var}[x_\infty] &= \phi^2 \text{Var}[x_\infty] + \sigma^2. \end{aligned}$$

Auflösen dieser Gleichungen führt auf die bereits bekannten Ergebnisse $E[x_{st.}] = \mu$ und $\text{Var}[x_{st.}] = \frac{\sigma^2}{1 - \phi^2}$, mit $x_{st.} = x_\infty$.

Beispiel 7.3:

Für die stationäre Varianz des GARCH(1,1)-Prozesses, mit $\text{Var}[y_t] = E[\sigma_t^2]$, erhält man in unendlicher Zukunft

$$E[\sigma_\infty^2] = \omega + \alpha E[\sigma_\infty^2] + \beta E[\sigma_\infty^2].$$

Auflösen führt auf $\text{Var}[y_{st.}] = \frac{\omega}{1 - \alpha - \beta}$, mit $y_{st.} = \sigma_\infty z_\infty$.

7.3 Vereinfachte Kovarianzformel

Bei der Berechnung von Kovarianzen können oft Vereinfachungen erzielt werden, wenn der Erwartungswert einer Zufallsvariable null ist. Eine erste Vereinfachung ergibt sich bereits durch umstellen der Kovarianzformel. Für zwei beliebige Zufallsvariablen X und Y gilt

$$\begin{aligned}\text{Cov}[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - 2E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y].\end{aligned}$$

Insbesondere ergibt sich damit für Die Varianz einer Zufallsvariable

$$\text{Var}[X] = \text{Cov}[X, X] = E[X^2] - E[X]^2.$$

Ist nun weiterhin bekannt, dass der Erwartungswert mindestens einer der beiden Zufallsvariablen null ist, also $E[X] = 0$ oder $E[Y] = 0$, erhält man unmittelbar

$$\text{Cov}[X, Y] = E[XY].$$

7.4 Hodrick-Prescott-Filter

Die Idee des *Hodrick-Prescott*-Filters basiert auf der Zerlegung einer Zeitreihe y_t in eine Summe aus zwei Zeitreihen

$$y_t = g_t + c_t, \quad t = 1, \dots, T,$$

wobei g_t nur den Wachstums- oder Trendanteil enthält (growth) und c_t den periodischen (cyclic) Anteil. Die Separation zwischen Trend und Zyklus wird durch die Lösung des Minimierungsproblems

$$\min_{\{g_t\}_{t=1}^T} \left[\sum_{t=1}^T (y_t - g_t)^2 + \lambda \sum_{t=2}^{T-1} (\nabla^2 g_{t+1})^2 \right]$$

ermöglicht. Der Parameter λ gewichtet dabei einen „Strafterm“, ∇ ist wieder der Rückwärts-Differenzenoperator, vgl. (2.65) auf Seite 38. Der zweimal angewendete ∇ -Operator ist ein Maß für die Abweichung von der Linearität. Im Falle einer stetig differenzierbaren Funktion würde der Differenzenoperator dem Differentialoperator entsprechen, wobei die zweite Ableitung bekanntlich ein Maß für die Krümmung einer Funktion ist.

KAPITEL 7. TIPPS UND TRICKS

Aufgrund der Struktur des Differenzenoperators kann das Minimierungsproblem in Matrixschreibweise notiert werden

$$z(\mathbf{g}) = (\mathbf{y} - \mathbf{g})'(\mathbf{y} - \mathbf{g}) + \lambda(\mathbf{F}\mathbf{g})'(\mathbf{F}\mathbf{g}) \rightarrow \min!$$

mit der Tridiagonalmatrix

$$\mathbf{F} = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}.$$

Die übliche notwendige Bedingung für die Existenz eines Minimums ist das Verschwinden der ersten Ableitung der Zielfunktion

$$\frac{\partial}{\partial \mathbf{g}'} z(\mathbf{g}) = -2(\mathbf{y} - \mathbf{g}) + 2\lambda \mathbf{F}'\mathbf{F}\mathbf{g} \stackrel{!}{=} \mathbf{0},$$

woraus nach elementaren Umformungen die Zerlegung des Vektors \mathbf{y} folgt

$$\begin{aligned} \mathbf{g} &= (\lambda \mathbf{F}'\mathbf{F} + \mathbf{I})^{-1} \mathbf{y} \\ \mathbf{c} &= \mathbf{y} - \mathbf{g}. \end{aligned}$$

Aufgrund der Struktur des Problems kommt als Extremum in der Tat nur ein Minimum in Frage. Hodrick und Prescott empfehlen für Quartalsdaten den Glätterwert $\lambda = 1600$.

8

Lösungen

Lösungen zu Kapitel 2

Lösung 2.1

$$\text{a) } \sum_{k=1}^5 k = 1 + 2 + 3 + 4 + 5 = 15$$

$$\text{b) } \sum_{k=1}^3 \frac{1}{2k} = \frac{1}{2} + \frac{1}{4} + \frac{1}{6} = \frac{6}{12} + \frac{3}{12} + \frac{2}{12} = \frac{11}{12}$$

$$\text{c) } \sum_{k=0}^4 2^k = 1 + 2 + 4 + 8 + 16 = 31$$

$$\begin{aligned} \text{d) } \sum_{i=1}^2 \sum_{j=1}^3 (i+j)^2 &= (1+1)^2 + (1+2)^2 + (1+3)^2 + (2+1)^2 + (2+2)^2 \\ &\quad + (2+3)^2 = 4 + 9 + 16 + 9 + 16 + 25 = 79 \end{aligned}$$

$$\begin{aligned} \text{e) } \sum_{i=1}^3 \sum_{j=0}^i i \cdot j &= 1 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 + 2 \cdot 1 + 2 \cdot 2 + 3 \cdot 0 + 3 \cdot 1 + 3 \cdot 2 \\ &\quad + 3 \cdot 3 = 1 + 2 + 4 + 3 + 6 + 9 = 25 \end{aligned}$$

$$\text{f) } \sum_{k=0}^{\infty} \left(\frac{3}{4}\right)^k = \frac{1}{1 - \frac{3}{4}} = 4$$

$$\text{g) } \sum_{k=1}^{\infty} \frac{1}{2^k} = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k - 1 = \frac{1}{1 - \frac{1}{2}} - 1 = 2 - 1 = 1$$

Index in Aufgabenteil g) beachten!

Lösung 2.2

- a) $2 + 4 + 6 + \dots + 16 = \sum_{k=1}^8 2k$
- b) $-1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \frac{1}{6} = \sum_{k=1}^6 \frac{(-1)^k}{k}$
- c) $0 + \frac{1}{2} + \frac{2}{3} + \frac{3}{4} + \dots + \frac{45}{46} = \sum_{k=0}^{45} \frac{k}{k+1}$
- d) $1 + 4 + 27 + 256 = \sum_{k=1}^4 k^k$
- e) $\frac{2}{1 - \frac{1}{2}} = 2 \cdot \frac{1}{1 - \frac{1}{2}} = 2 \sum_{k=0}^{\infty} \frac{1}{2^k}$
-

Lösung 2.3

$$\begin{aligned}\gamma_3 &= \text{Cov}[y_t, y_{t-3}] = E[y_t y_{t-3}] \\ &= E[(\phi y_{t-1} + \epsilon_t) y_{t-3}] = E[\phi y_{t-1} y_{t-3}] + E[\epsilon_t y_{t-3}] \\ &= \phi \text{Cov}[y_{t-1}, y_{t-3}] + \text{Cov}[\epsilon_t, y_{t-3}] \\ &= \phi \gamma_2 = \phi^2 \gamma_1 = \phi^3 \gamma_0\end{aligned}$$

Lösung 2.4

- a) $\det \mathbf{A} = 2, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2 & -2 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 6 - 12 \\ -3 + 12 \end{pmatrix} = \begin{pmatrix} -3 \\ 4.5 \end{pmatrix}$
- b) $\det \mathbf{A} = 10, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 3 & -1 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} -2 \\ 8 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} -6 - 8 \\ 4 + 32 \end{pmatrix} = \begin{pmatrix} -1.4 \\ 3.6 \end{pmatrix}$
- c) $\det \mathbf{A} = 0$, \mathbf{A} ist nicht invertierbar!
- d) $\det \mathbf{A} = -10, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -\frac{1}{10} \begin{pmatrix} 4 & -3 \\ -6 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = -\frac{1}{10} \begin{pmatrix} -6 \\ 4 \end{pmatrix} = \begin{pmatrix} 0.6 \\ -0.4 \end{pmatrix}$
- e) $\det \mathbf{A} = 4, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} -4 & -12 \\ -0 & -1 \end{pmatrix} \begin{pmatrix} -3 \\ -8 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 108 \\ 8 \end{pmatrix} = \begin{pmatrix} 27 \\ 2 \end{pmatrix}$
-

Lösung 2.5

$$\hat{\mu} = \frac{1}{10} \sum_{t=1}^{10} y_t = 2.5$$

$$\hat{\gamma}_k = \frac{1}{10} \sum_{t=k+1}^{10} (y_t - \hat{\mu})(y_{t-k} - \hat{\mu}) \Rightarrow \hat{\gamma}_0 = 2.25, \hat{\gamma}_1 = 1.225, \hat{\gamma}_2 = -0.3$$

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \Rightarrow \hat{\rho}_1 = 0.5\bar{4}, \hat{\rho}_2 = -0.1\bar{3}$$

$$\hat{\phi}_{21} = \frac{\hat{\rho}_1 - \hat{\rho}_1 \hat{\rho}_2}{1 - \hat{\rho}_1^2} = 0.877$$

$$\hat{\phi}_{22} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} = -0.611$$

$$\hat{\sigma}^2 = \hat{\gamma}_0(1 - \hat{\rho}_1^2) = 1.583$$

Lösung 2.6

$$\mathbf{y} = (1, 0, 1, 2)', \quad \mathbf{X}' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} 4 & 4 \\ 4 & 6 \end{pmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{8} \begin{pmatrix} 6 & -4 \\ -4 & 4 \end{pmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \quad \hat{\boldsymbol{\theta}} = \frac{1}{8} \begin{pmatrix} 8 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Lösung 2.7

a) $\log[ab] = \log[a] + \log[b]$

b) $\log\left[\frac{a}{b}\right] = \log[ab^{-1}] = \log[a] - \log[b]$

c) $\log[2a^b] = \log[2] + b \log[a]$

d) $\log[\sqrt{a}] = \log[a^{\frac{1}{2}}] = \frac{1}{2} \log[a]$

e) $\log[a^3 \sqrt[4]{b}] = 3 \log[a] + \frac{1}{4} \log[b]$

Lösung 2.8

$$\hat{\mu} = \frac{1}{9} \sum_{t=0}^8 y_t = 2$$

$$\hat{\gamma}_0 = \frac{1}{9} \sum_{t=0}^8 (y_t - 2)^2 = \frac{4}{3}$$

$$\hat{\gamma}_1 = \frac{1}{9} \sum_{t=1}^8 (y_t - 2)(y_{t-1} - 2) = \frac{1}{3} \quad \Rightarrow \quad \hat{\phi} = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{1}{4}$$

$$F(\hat{\phi}) = \frac{9}{1 - \frac{1}{16}} = \frac{144}{15}$$

$$c_{o/u} = \frac{1}{4} \pm 1.96 \sqrt{\frac{15}{144}} \quad \Rightarrow \quad \phi \in [-0.383, 0.883]$$

Lösung 2.9

$$\text{a) } \nabla^3 y_t = (1 - B)^3 y_t = (1 - 3B + 3B^2 - B^3) y_t = y_t - 3y_{t-1} + 3y_{t-2} - y_{t-3}$$

$$\text{b) } B \nabla y_t = B(1 - B) y_t = (B - B^2) y_t = y_{t-1} - y_{t-2}$$

$$\text{c) } B^{-1} \nabla y_t = (B^{-1} - B^{-1} B) y_t = (B^{-1} - 1) y_t = y_{t+1} - y_t$$

$$\text{d) } B^{-2} y_t = B^{-1} B^{-1} y_t = B^{-1} y_{t+1} = y_{t+2}$$

$$\text{e) } \nabla^2 B^{-1} y_t = (1 - 2B + B^2) B^{-1} y_t = (B^{-1} - 2 + B) y_t = y_{t+1} - 2y_t + y_{t-1}$$

$$\begin{aligned} \text{f) } B^{-2} \nabla^2 y_t &= B^{-2} (1 - 2B + B^2) y_t = (B^{-2} - 2B^{-1} + 1) y_t \\ &= y_{t+2} - 2y_{t+1} + y_t \end{aligned}$$

Lösung 2.10

$$\text{a) } \lambda_{1/2} = \frac{\phi_1}{2} \pm \sqrt{\frac{\phi_1^2}{4} + \phi_2} = \frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{1}{2}} = \frac{1}{2} \pm i \sqrt{\frac{1}{4}} = \frac{1}{2} \pm i \frac{1}{2}$$

$$\text{b) } \omega = \arctan \left| \frac{b}{a} \right| = \arctan 1 = \frac{\pi}{4} \quad \Rightarrow \quad T = \frac{2\pi}{\omega} = \frac{8\pi}{\pi} = 8$$

Lösung 2.10 (Fortsetzung)

$$\begin{aligned} \text{c) } r^\tau &= \frac{r}{10} \quad \Leftrightarrow \quad \tau \log r = \log r - \log 10 \quad \Leftrightarrow \quad \tau = 1 - \frac{\log 10}{\log r} \\ r &= \sqrt{\frac{1}{4} + \frac{1}{4}} = \frac{1}{\sqrt{2}} \quad \Rightarrow \quad \tau = 1 - \frac{\log 10}{-\frac{1}{2} \log 2} = 1 + 2 \frac{\log 10}{\log 2} = 7.644 \end{aligned}$$

Lösung 2.11

$$\begin{aligned} \lambda_{1/2} &= \frac{\phi_1}{2} \pm \sqrt{\frac{\phi_1^2}{4} + \phi_2} = 0.1 \pm \sqrt{0.36} \quad \Rightarrow \quad \lambda_1 = 0.7, \lambda_2 = -0.5 \\ &\Rightarrow \quad \text{Prozess ist stationär!} \\ \rho_1 &= \phi_1 + \phi_2 \rho_1 \quad \Rightarrow \quad \rho_1 = \frac{\phi_1}{1 - \phi_2} = \frac{0.2}{0.65} = 0.3077 \\ \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}^{-1} \begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix} = -\frac{5}{6} \begin{pmatrix} -0.5 & -1 \\ -0.7 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0.3077 \end{pmatrix} = \begin{pmatrix} 0.6731 \\ 0.3269 \end{pmatrix} \\ &\Rightarrow \quad \rho_k = 0.6731 \cdot 0.7^k + 0.3269 \cdot (-0.5)^k \end{aligned}$$

Lösung 2.12

$$\begin{aligned} r &= \sqrt{a^2 + b^2} = \sqrt{\frac{1}{4} + \frac{1}{4}} = \frac{1}{\sqrt{2}}, \quad \omega = \arctan \left| \frac{b}{a} \right| = \arctan 1 = \frac{\pi}{4} \\ \rho_k &= r^k \cos[\omega k] = 2^{-\frac{k}{2}} \cos \left[\frac{\pi k}{4} \right] \quad \Rightarrow \quad \rho_2 = 0, \quad \rho_4 = -\frac{1}{4} \end{aligned}$$

Lösungen zu Kapitel 3

Lösung 3.1

$$\frac{1}{1 + \theta_1 B} y_t = (1 - \theta_1 B + \theta_1^2 B^2 - \dots) y_t = \left(1 - \sum_{k=1}^{\infty} (-1)^{k-1} \theta_1^k B^k \right) y_t = \epsilon_t$$

$$\Rightarrow \phi(B) y_t = \left(1 - \sum_{k=1}^{\infty} \phi_k B^k \right) y_t = \epsilon_t \quad \text{mit} \quad \phi_k = (-1)^{k-1} \theta_1^k$$

\Rightarrow MA(1)-Prozess besitzt AR(∞)-Darstellung

Lösung 3.2

a) $(\epsilon_t)_{t=0,\dots,6} = \{0, 2, 2, -1, -1.5, 1.75, -2.875\}$

$$\Rightarrow l(0.5, 0.5) = -3.434 - 22.578 = -26.012$$

b) $(\epsilon_t)_{t=0,\dots,6} = \{0, 2, 2.5, -0.625, -1.844, 1.461, -2.365\}$

$$\Rightarrow l(0.25, 1) = -5.514 - 10.884 = -16.398$$

Lösung 3.3

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2) \sigma^2$$

$$\gamma_1 = (\theta_1 + \theta_2 \theta_1) \sigma^2$$

$$\gamma_2 = \theta_2 \sigma^2$$

$$\gamma_3 = \gamma_4 = \dots = 0$$

Lösung 3.4

a) MA(2)-Prozess b) AR(1)-Prozess

c) AR(2)-Prozess d) MA(1)-Prozess

e) MA(3)-Prozess

Lösung 3.5

$$\theta(z) = 1 + \sum_{k=1}^q \theta_k z^k = \prod_{k=1}^q (1 - \lambda_k z)$$

Lösung 3.6

a) $\lambda_{1/2} = -\frac{\theta_1}{2} \pm \sqrt{\frac{\theta_1^2}{4} - \theta_2}$ (pq-Formel)

b) $\lambda_{1/2} = 1 \pm \frac{1}{2} \Rightarrow \lambda_1 = \frac{3}{2}, \lambda_2 = \frac{1}{2}$
 \Rightarrow Prozess ist nicht invertierbar!

c) $y_t = \left(1 - \frac{2}{3}B\right)\left(1 - \frac{1}{2}B\right)\epsilon_t = \epsilon_t - \frac{2}{3}\epsilon_{t-1} - \frac{1}{2}\epsilon_{t-1} + \frac{1}{3}\epsilon_{t-2}$
 $= \epsilon_t - \frac{7}{6}\epsilon_{t-1} + \frac{1}{3}\epsilon_{t-2} \Rightarrow \theta_1 = -1.17, \theta_2 = 0.33$

Lösung 3.7

$$(\epsilon_t)_{t=-1,\dots,6} = \{0, 0, 1, 4, 6, 5, 4, 0\}$$

$$\Rightarrow l(1, -2, 1, 1) = -5.514 - 47 = -52.514$$

Lösung 3.8

a) $\lambda_{1/2} = -\frac{\theta_1}{2} \pm \sqrt{\frac{\theta_1^2}{4} - \theta_2} = 1 \pm \sqrt{1 - 1} = 1$
 \Rightarrow Prozess ist nicht invertierbar!

b) Bedingte ML-Schätzung ist hier nicht sinnvoll, da sich Fehler in den Anfangsbedingungen potenzieren.

Lösungen zu Kapitel 4

Lösung 4.1

$$\begin{aligned}\gamma_1 &= \sigma^2 \left(\psi_1 + \sum_{j=1}^{\infty} \psi_{j+1} \psi_j \right) = \sigma^2 \left(\phi + \theta + (\phi + \theta)^2 \sum_{j=1}^{\infty} \phi^{2j-1} \right) \\ &= \sigma^2 \left(\phi + \theta + \phi(\phi + \theta)^2 \sum_{k=0}^{\infty} \phi^{2k} \right) = \sigma^2 \left(\phi + \theta + \frac{\phi(\phi + \theta)^2}{1 - \phi^2} \right) \\ &= \phi \sigma^2 \left(1 + \frac{(\phi + \theta)^2}{1 - \phi^2} \right) + \theta \sigma^2 = \phi \gamma_0 + \theta \sigma^2\end{aligned}$$

Lösung 4.2

$$\begin{aligned}\phi(B)y_t &= \theta(B)\epsilon_t \\ \Rightarrow (1 - \phi B)y_t &= (1 + \theta B)\epsilon_t = (1 - \phi B)\epsilon_t \\ \Rightarrow y_t &= \epsilon_t\end{aligned}$$

Lösung 4.3

$$\begin{aligned}(\epsilon_t)_{t=0,\dots,6} &= \{0, 2, 1, -1, -1, -1, 2\} \\ \Rightarrow l(0.5, 0, 0.5, 1) &= -5.514 - 6 = -11.514\end{aligned}$$

Lösung 4.4

$$\begin{aligned}\text{a) } (1 - \phi B)\nabla y_t &= \epsilon_t \\ \Leftrightarrow (1 - \phi B)(1 - B)y_t &= \epsilon_t \\ \Leftrightarrow (1 - B - \phi B + \phi B^2)y_t &= \epsilon_t \\ \Rightarrow y_t &= (1 + \phi)y_{t-1} - \phi y_{t-2} + \epsilon_t\end{aligned}$$

Lösung 4.4 (Fortsetzung)

$$\text{b) } \nabla y_t = (1 + \theta_1 B + \theta_2 B^2) \epsilon_t$$

$$\Leftrightarrow (1 - B)y_t = (1 + \theta_1 B + \theta_2 B^2) \epsilon_t$$

$$\Rightarrow y_t = y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

$$\text{c) } (1 - \phi B) \nabla^2 y_t = (1 + \theta B) \epsilon_t$$

$$\Leftrightarrow (1 - \phi B)(1 - 2B + B^2)y_t = (1 + \theta B) \epsilon_t$$

$$\Leftrightarrow (1 - 2B + B^2 - \phi B + 2\phi B^2 - \phi B^3)y_t = (1 + \theta B) \epsilon_t$$

$$\Rightarrow y_t = (2 + \phi)y_{t-1} - (1 + 2\phi)y_{t-2} + \phi y_{t-3} + \epsilon_t + \theta \epsilon_{t-1}$$

Lösung 4.5

$$(\nabla y_t)_{t=1,\dots,6} = \{1, 2, -1, 2, -2, -2\}$$

$$(\epsilon_t)_{t=0,\dots,6} = \{0, 1, 1, -2, 4, -6, 4\}$$

Bedingte LogLikelihood-Funktion:

$$l(\theta_0, \theta_1, \sigma^2) = -\frac{T}{2} \log[2\pi\sigma^2] - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2 \quad \text{mit} \quad \epsilon_t = \nabla y_t - \theta_0 - \theta_1 \epsilon_{t-1}$$

$$\Rightarrow l(0, 1, 1) = -5.514 - 37 = -42.514$$

Lösung 4.6

$$l(\phi_1, \chi_0, \theta_0, \theta_1, \sigma^2) = -\frac{T}{2} \log[2\pi\sigma^2] - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2 \quad \text{mit}$$

$$\epsilon_t = y_t - \phi_1 y_{t-1} - \chi_0 x_t - \theta_0 - \theta_1 \epsilon_{t-1}$$

Lösung 4.7

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

$$\phi(B)u_t = \theta(B)\epsilon_t$$

Lösung 4.7 (Fortsetzung)

$$\begin{aligned}
 \Rightarrow \quad \phi(B)y_t &= \phi(1)\beta_0 + \phi(B)\beta_1x_t + \theta(B)\epsilon_t \\
 &= (1 - \phi_1 - \phi_2)\beta_0 + \beta_1x_t - \phi_1\beta_1x_{t-1} - \phi_2\beta_1x_{t-2} + \epsilon_t + \theta_1\epsilon_{t-1} \\
 &= \theta_0 + \chi_0x_t + \chi_1x_{t-1} + \chi_2x_{t-2} + \epsilon_t + \theta_1\epsilon_{t-1} \\
 \Rightarrow \quad \theta_0 &= (1 - \phi_1 - \phi_2)\beta_0, \quad \chi_0 = \beta_1, \quad \chi_1 = -\phi_1\beta_1, \quad \chi_2 = -\phi_2\beta_1
 \end{aligned}$$

Lösung 4.8

$$\begin{aligned}
 \text{BIC} &= -\frac{2l}{T} + \frac{u \log[T]}{T} \\
 &= \log[2\pi\hat{\sigma}^2] + \frac{1}{\hat{\sigma}^2 T} \sum_{t=1}^T \hat{\epsilon}_t^2 + \frac{u \log[T]}{T} \\
 &= \log[2\pi] + \log[\hat{\sigma}^2] + 1 + \frac{u \log[T]}{T} \\
 &\propto \log[\hat{\sigma}^2] + \frac{u \log[T]}{T}.
 \end{aligned}$$

Lösung 4.9

AIC	0	1	2
0	1.087	1.059	0.952
1	0.851	0.838	0.847
2	0.887	0.842	0.865
3	0.831	0.849	0.871

BIC	0	1	2
0	1.129	1.144	1.079
1	0.936	0.965	1.016
2	1.014	1.011	1.076
3	1.000	1.060	1.124

Das Modell mit jeweils kleinstem Info-Kriterium sollte gewählt werden.

\Rightarrow Unter AIC: AR(3)

\Rightarrow Unter BIC: AR(1)

Lösungen zu Kapitel 5

Lösung 5.1

$$\text{Var}[y_1] = \omega + \alpha \text{Var}[y_0] = \omega$$

$$\text{Var}[y_2] = \omega + \alpha \text{Var}[y_1] = \omega + \alpha\omega = \omega(1 + \alpha)$$

$$\text{Var}[y_3] = \omega + \alpha \text{Var}[y_2] = \omega + \alpha\omega + \alpha^2\omega = \omega(1 + \alpha + \alpha^2)$$

$$\Rightarrow \quad \text{Var}[y_t] = \omega \sum_{k=0}^{t-1} \alpha^k$$

Lösung 5.2

$$E[y_t^3] = E[\sigma_t^3 z_t^3] = E[\sigma_t^3] E[z_t^3] = 0$$

$$\Rightarrow \quad S = \frac{E[y_t^3]}{\text{Var}[y_t]^{3/2}} = 0$$

Lösung 5.3

$$K = \frac{3 - 3\alpha^2}{1 - 3\alpha^2}$$

$$\Leftrightarrow K(1 - 3\alpha^2) = 3 - 3\alpha^2$$

$$\Leftrightarrow K = 3 + 3(K - 1)\alpha^2$$

$$\Leftrightarrow \alpha^2 = \frac{K - 3}{3(K - 1)} \quad \Leftarrow \quad \alpha = \sqrt{\frac{K - 3}{3(K - 1)}}$$

$$\sigma^2 = \frac{\omega}{1 - \alpha}$$

$$\Leftrightarrow \omega = \sigma^2(1 - \alpha)$$

$$\text{a) } \alpha = \frac{1}{3}, \omega = 1$$

$$\text{b) } \alpha = \frac{1}{2}, \omega = 2$$

Lösung 5.4

$$\hat{\mu} = \frac{1}{8} \sum_{t=1}^8 y_t = \frac{2-2}{8} = 0$$

$$\hat{m}_2 = \frac{1}{8} \sum_{t=1}^8 (y_t - \hat{\mu})^2 = \frac{4+4}{8} = 1$$

$$\hat{m}_4 = \frac{1}{8} \sum_{t=1}^8 (y_t - \hat{\mu})^4 = \frac{16+16}{8} = 4$$

$$\Rightarrow \hat{K} = \frac{\hat{m}_4}{\hat{m}_2^2} = 4$$

$$\Rightarrow \hat{\alpha} = \sqrt{\frac{\hat{K} - 3}{3(\hat{K} - 1)}} = \sqrt{\frac{1}{9}} = \frac{1}{3}, \quad \hat{\omega} = \hat{m}_2(1 - \hat{\alpha}) = \frac{2}{3}$$

Lösung 5.5

$$\begin{aligned} l(\omega, \alpha) &= -\frac{T}{2} \log[2\pi] - \frac{1}{2} \sum_{t=1}^T \left(\log[\omega + \alpha \epsilon_{t-1}^2] + \frac{y_t^2}{\omega + \alpha \epsilon_{t-1}^2} \right) \\ &= -4 \log[2\pi] - \frac{1}{2} \left(6 \log \frac{2}{3} + 2 \log 2 + 2 \cdot 6 \right) \\ &= -4 \log 2 - 4 \log \pi - \frac{1}{2} (8 \log 2 - 6 \log 3 + 12) \\ &= -8 \log 2 + 3 \log 3 - 4 \log \pi - 6 \approx -12.828 \end{aligned}$$

Lösung 5.6

$$\text{Bedingung: } \sum_{k=1}^{\infty} \alpha_k < 1$$

$$\Rightarrow \sum_{k=1}^{\infty} \alpha^k = \sum_{k=0}^{\infty} \alpha^k - 1 = \frac{1}{1-\alpha} - 1 = \frac{\alpha}{1-\alpha}$$

$$\Rightarrow \frac{\alpha}{1-\alpha} < 1 \quad \Leftrightarrow \quad \alpha < 1-\alpha \quad \Leftrightarrow \quad 2\alpha < 1 \quad \Leftrightarrow \quad \alpha < \frac{1}{2} \quad \checkmark$$

Lösung 5.7

$$E[\nu_t | \epsilon^{t-1}] = 0$$

$$\Rightarrow \text{Cov}[\nu_t, \nu_{t-1} | \epsilon^{t-1}] = E[\nu_t \nu_{t-1} | \epsilon^{t-1}] = E[\nu_t | \epsilon^{t-1}] \nu_{t-1} = 0$$

Lösung 5.8

$$E[\nu_t] = E[(z_t^2 - 1)\sigma_t^2] = E[z_t^2 - 1]E[\sigma_t^2] = 0$$

$$\begin{aligned} \Rightarrow \text{Cov}[\nu_t, \nu_{t-1}] &= E[(z_t^2 - 1)(z_{t-1}^2 - 1)\sigma_t^2 \sigma_{t-1}^2] \\ &= E[z_t^2 - 1]E[(z_{t-1}^2 - 1)\sigma_t^2 \sigma_{t-1}^2] = 0 \end{aligned}$$

Lösung 5.9

$$\begin{aligned} E[y_t^4] &= 3E[\sigma_t^4] = 3E[(\omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2)^2] \\ &= 3E[(\omega + (\alpha z_{t-1}^2 + \beta)\sigma_{t-1}^2)^2] \\ &= 3\omega^2 + 6\omega E[\alpha z_{t-1}^2 + \beta]E[\sigma_{t-1}^2] + 3E[(\alpha z_{t-1}^2 + \beta)^2]E[\sigma_{t-1}^4] \\ &= 3\omega^2 + 6\omega(\alpha + \beta)E[\sigma_{t-1}^2] + (3\alpha^2 + 2\alpha\beta + \beta^2)E[\epsilon_{t-1}^4] \\ &= 3\omega^2 + 6\omega^2 \frac{\alpha + \beta}{1 - \alpha - \beta} + (3\alpha^2 + 2\alpha\beta + \beta^2)E[y_{t-1}^4] \end{aligned}$$

$$\begin{aligned} \Rightarrow E[y_{st.}^4] &= \frac{3\omega^2 \left(1 + \frac{2(\alpha+\beta)}{1-\alpha-\beta}\right)}{1 - 3\alpha^2 - 2\alpha\beta - \beta^2} \\ &= \frac{\omega^2}{(1 - \alpha - \beta)^2} \cdot 3 \frac{(1 - \alpha - \beta)(1 + \alpha + \beta)}{1 - 2\alpha^2 - (\alpha + \beta)^2} \end{aligned}$$

$$\Rightarrow K = \frac{E[y_{st.}^4]}{\text{Var}[y_{st.}]^2} = 3 \frac{1 - (\alpha + \beta)^2}{1 - 2\alpha^2 - (\alpha + \beta)^2}$$

Lösung 5.10

$$l(\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q) = -\frac{T}{2} \log[2\pi] - \frac{1}{2} \sum_{t=1}^T \left(\log \sigma_t^2 + \frac{y_t^2}{\sigma_t^2} \right)$$

$$\text{mit } \sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2$$

Lösung 5.11

$$l(\phi, \theta_0, \theta_1, \theta_2, \omega, \alpha_1, \alpha_2, \beta) = -\frac{T}{2} \log[2\pi] - \frac{1}{2} \sum_{t=1}^T \left(\log \sigma_t^2 + \frac{\epsilon_t^2}{\sigma_t^2} \right)$$

$$\text{mit } \epsilon_t = y_t - \phi y_{t-1} - \theta_0 - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \beta \sigma_{t-1}^2$$

Lösung 5.12

$$\text{ARMA-Teil: } |\phi| < 1 \quad \text{und} \quad \text{GARCH-Teil: } \alpha_1 + \alpha_2 + \beta < 1$$

Lösung 5.13

$$\begin{aligned} E[\sigma_t^2] &= \omega + \alpha E[\epsilon_{t-1}^2] - 2\alpha\gamma E[\epsilon_{t-1}] + \alpha\gamma^2 + \beta E[\sigma_{t-1}^2] \\ &= \omega + \alpha\gamma^2 + (\alpha + \beta)E[\sigma_{t-1}^2] \end{aligned}$$

$$\Rightarrow \quad \text{Var}[y_{st.}] = \frac{\omega + \alpha\gamma^2}{1 - \alpha - \beta}$$

$$\Rightarrow \quad \text{Stationaritätsbedingung: } \alpha + \beta < 1$$

Lösung 5.14

ARMA(2,1)-GARCH(1,1) : $\boldsymbol{\theta} = (\phi_1, \phi_2, \theta_0, \theta_1, \omega, \alpha, \beta)'$ $\Rightarrow u = 7$

ARCH(7) : $\boldsymbol{\theta} = (\omega, \alpha_1, \dots, \alpha_7)'$ $\Rightarrow u = 8$

TARCH-M : $\boldsymbol{\theta} = (\mu, \lambda, \omega, \alpha, \beta, \gamma)'$ $\Rightarrow u = 6$

	AIC	BIC	HQ
ARMA-GARCH	0.58	0.824	0.676
ARCH	0.57	0.849	0.679
TARCH-M	0.61	0.819	0.692

Das Modell mit jeweils kleinstem Info-Kriterium sollte gewählt werden.

\Rightarrow Unter AIC: ARCH(7)

\Rightarrow Unter BIC: TARCH-M

\Rightarrow Unter HQ: ARMA(2,1)-GARCH(1,1)

Lösungen zu Kapitel 6

Lösung 6.1

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\theta}'} Q(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}'} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \\ &= -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\theta} \stackrel{!}{=} 0\end{aligned}$$

$$\Rightarrow \hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Lösung 6.2

$$\text{a) } \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \epsilon_t$$

$$z_t = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix}$$

$$\text{b) } \begin{pmatrix} y_{t-2} \\ y_{t-1} \\ y_t \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \phi_3 & \phi_2 & \phi_1 \end{pmatrix} \begin{pmatrix} y_{t-3} \\ y_{t-2} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \epsilon_t$$

$$z_t = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-2} \\ y_{t-1} \\ y_t \end{pmatrix}$$

Lösung 6.3

$$\begin{pmatrix} y_{t-(p-1)} \\ \vdots \\ y_{t-1} \\ y_t \end{pmatrix} = \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 \\ \phi_p & \dots & \phi_2 & \phi_1 \end{pmatrix} \begin{pmatrix} y_{t-p} \\ \vdots \\ y_{t-2} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \epsilon_t$$

$$z_t = \begin{pmatrix} 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-(p-1)} \\ \vdots \\ y_{t-1} \\ y_t \end{pmatrix}$$

Lösung 6.4

$$y_t = (\phi_1 B + \phi_2 B^2)y_t + \epsilon_t = \frac{1}{1 - \phi_1 B - \phi_2 B^2} \epsilon_t$$

$$z_t = (1 + \theta_1 B + \theta_2 B^2)y_t = \frac{1 + \theta_1 B + \theta_2 B^2}{1 - \phi_1 B - \phi_2 B^2} \epsilon_t$$

$$\Rightarrow z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2} = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \quad \checkmark$$

Lösung 6.5

$$\text{a) } \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \epsilon_t$$

$$z_t = \begin{pmatrix} 1 & \theta_1 & \theta_2 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix}$$

$$\text{b) } \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \epsilon_t$$

$$z_t = \begin{pmatrix} 1 & \theta_1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix}$$

Lösung 6.6

$$\begin{aligned} K_t &= \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left((1 \ 0) \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 1 \right)^{-1} \\ &= \begin{pmatrix} 1 \\ 2 \end{pmatrix} (1 + 1)^{-1} = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \mu_{t|t} &= \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} \left(2 - (1 \ 0) \begin{pmatrix} 2 \\ 1 \end{pmatrix} - 1 \right) \\ &= \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ 0 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \Sigma_{t|t} &= \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} (1 \ 0) \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 1 \\ 1 & 2 \end{pmatrix} \end{aligned}$$

Lösung 6.7

$$\nu_t = 2 - (1 \ 0) \begin{pmatrix} 2 \\ 1 \end{pmatrix} - 1 = -1$$

$$\Gamma_t = (1 \ 0) \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 1 = 1 + 1 = 2$$

$$K_t = \frac{1}{2} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}$$

$$\mu_{t|t} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ 0 \end{pmatrix}$$

$$\begin{aligned} \Sigma_{t|t} &= \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} - 2 \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} - 2 \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 1 \\ 1 & 2 \end{pmatrix} \end{aligned}$$

Lösung 6.8

$$\begin{aligned}\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} &= \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \epsilon_t &\Rightarrow \Omega = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \\ z_t &= \begin{pmatrix} 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} + \delta_t &\Rightarrow R = 1\end{aligned}$$

Initialisierung:

$$\mu_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \Sigma_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Time-Update 1:

$$\begin{aligned}\mu_{1|0} &= A\mu_0 = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \Sigma_{1|0} &= A\Sigma_0A' + \Omega = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -2 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\end{aligned}$$

Mess-Update 1:

$$\begin{aligned}\nu_1 &= z_1 - H\mu_{1|0} = \frac{3}{2} - \begin{pmatrix} 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0 \\ \Gamma_1 &= H\Sigma_{1|0}H' + R = \begin{pmatrix} 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 1 = 3 \\ K_1 &= \Sigma_{1|0}H'\Gamma_1^{-1} = \frac{1}{3} \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} \\ \mu_{1|1} &= \mu_{1|0} + K_1\nu_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \Sigma_{1|1} &= \Sigma_{1|0} - K_1\Gamma_1K_1' = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} - 3 \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} \begin{pmatrix} \frac{2}{3} & 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & 0 \end{pmatrix}\end{aligned}$$

Lösung 6.8 (Fortsetzung)

Time-Update 2:

$$\begin{aligned}\mu_{2|1} &= A\mu_{1|1} = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ \Sigma_{2|1} &= A\Sigma_{1|1}A' + \Omega = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -2 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{8}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{pmatrix}\end{aligned}$$

Mess-Update 2:

$$\begin{aligned}\nu_2 &= z_2 - H\mu_{2|1} = \frac{1}{2} - \begin{pmatrix} 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 1 \\ \Gamma_2 &= H\Sigma_{2|1}H' + R = \begin{pmatrix} 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{8}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} + 1 = \frac{9}{2} \\ K_2 &= \Sigma_{2|1}H'\Gamma_2^{-1} = \frac{2}{9} \begin{pmatrix} \frac{8}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix} \\ \mu_{2|2} &= \mu_{2|1} + K_2\nu_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} \\ \frac{11}{9} \end{pmatrix} \\ \Sigma_{2|2} &= \Sigma_{2|1} - K_2\Gamma_2K_2' = \begin{pmatrix} \frac{8}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{pmatrix} - \frac{9}{2} \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix} \begin{pmatrix} \frac{2}{3} & \frac{2}{9} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & \frac{4}{9} \end{pmatrix}\end{aligned}$$

Lösung 6.9

$$l = l_1 + l_2 = -\frac{\log[6\pi]}{2} - \frac{\log[9\pi] + \frac{2}{9}}{2} \approx -3.25$$

Lösung 6.10

ARV-*Kalman*-Filter

Anfangsbedingung μ_0, Σ_0

▷ Initialisierung

for $t = 1, \dots, T$ **do**

Time-Update

$$\mu_{t|t-1} = \lambda \mu_{t-1|t-1} + (1 - \lambda) \bar{h}$$

$$\Sigma_{t|t-1} = \lambda^2 \Sigma_{t-1|t-1} + \gamma^2$$

Mess-Update

$$\nu_t = z_t^* - \mu_{t|t-1} + 1.27$$

▷ Innovation

$$\Gamma_t = \Sigma_{t|t-1} + \frac{\pi^2}{2}$$

$$K_t = \frac{\Sigma_{t|t-1}}{\Gamma_t}$$

▷ *Kalman*-Gain

$$\mu_{t|t} = \mu_{t|t-1} + K_t \nu_t$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t^2 \Gamma_t$$

end for
