

Zeitreihenanalyse mit R

Matti Schneider, Sebastian Mentemeier

SS 2010

Inhaltsverzeichnis

1	Klassische Zeitreihenanalyse	4
1.1	Einführung	4
1.1.1	Das klassische Komponentenmodell	4
1.1.2	Additives und multiplikatives Modell	5
1.1.3	multiplikatives Modell	5
1.2	Darstellung von Zeitreihen	7
1.2.1	Gleitender Durchschnitt	7
1.2.2	Trendbestimmung mit saisonalen Einflüssen	8
1.2.3	Bestimmung der Saisonkomponente	8
1.2.4	Vergleich mit Vorjahresdaten	9
1.3	Parametrische Modelle für Trend und Saisonkomponente	9
1.3.1	Ohne Saisonkomponente	10
1.3.2	Regressionsmodelle für die Saisonkomponente	10
1.3.3	Prognose	11
1.4	Exponentielle Glättung und adaptive Prognose	11
1.4.1	Exponentielle Glättung	11
1.4.2	Bestimmung des optimalen α	12
1.4.3	Exponentielles Glätten mit Wachstum	12
1.4.4	Holt-Winters-Verfahren	13
1.5	R-Befehle	13
1.5.1	Saisonkomponenten aus linearen Modellen berechnen	14
1.5.2	Vorhersagen aus linearen Modellen	15
2	AR-Modelle	16
2.1	schwach Stationäre Prozesse	16
2.1.1	Definition und Beispiele	16
2.1.2	empirische Kenngrößen	18
2.2	AR(p) Modelle	20
2.2.1	Schätzung von ϕ_1, \dots, ϕ_p	23
2.2.2	Schätzung von σ^2	24
2.2.3	Vorhersagen in AR(p) Modellen	25
2.3	Tests für die Zeitreihenanalyse	25
2.3.1	Box-Pierce Test	26
2.3.2	Ljung-Box Test	26
2.3.3	Turning Point Test	27
2.4	R-Befehle	27

3	ARMA-Modelle	28
3.1	Partielle Autokorrelationsfunktion	28
3.1.1	Optimale lineare Prognose in AR-Modellen	28
3.1.2	Die partielle Autokorrelationsfunktion	29
3.1.3	Schätzung der PACF	30
3.2	MA(q)-Modelle	30
3.2.1	Invertierbare MA-Prozesse	32
3.2.2	Schätzen der Parameter $\theta_1, \dots, \theta_n$	33
3.2.3	Optimale Prognose in MA(q)-Modellen	34
3.3	Modellwahl: Akaike Information Criterion	34
3.4	ARMA-Modelle	35
3.4.1	ACVF eines kausalen ARMA(1,1)-Prozesses	36
3.4.2	Box-Jenkins-Methode der Modellwahl	37
3.5	R-Befehle	37
4	ARIMA-Modelle	39
4.1	Integrierte Prozesse	39
4.1.1	ARIMA-Modelle	41
4.2	Tests auf (In-)Stationarität	41
4.2.1	Test auf Differenzen-Stationarität: Dickey-Fuller-Test	41
4.2.2	Test auf (Trend-)Stationarität: KPSS-Test	42
4.2.3	Abschließende Bemerkungen	42
4.3	R-Befehle	43
5	GARCH-Modell	44
5.1	Heteroskedastische Modelle	44
5.2	ARCH(q) Modell	46
5.2.1	Der Prozess (X_t^2)	48
5.3	GARCH(p, q) Modelle	48
5.4	Parameterschätzung	49
5.5	R-Befehle	50

1 Klassische Zeitreihenanalyse

1.1 Einführung

Eine *Zeitreihe* ist eine zeitlich geordnete Folge $(X_t)_{t \in T}$ von Beobachtungen einer Größe, z.B. der Temperatur am Flughafen Münster-Osnabrück, der Studierendenzahl an der WWU oder eines Konjunkturindicators. Wir beschränken uns hier auf den Fall $T \subset \mathbb{Z}$, wobei reale Zeitreihen stets endliche Indexmengen (also Beobachtungszeiträume) besitzen werden, für eine elegante theoretische Darstellung wird jedoch meist $T = \mathbb{Z}$ oder $T = \mathbb{N}$ gewählt.

1.1.1 Das klassische Komponentenmodell

Betrachten wir zwei exemplarische Zeitreihen: Jährliche Werte für das deutsche Bruttoinlandsprodukt im Vergleich zum Wert des Jahres 2005 (welches dem Wert 100 entspricht), und die monatlichen Durchschnitte der Tagesmitteltemperaturen am Flughafen Münster-Osnabrück:

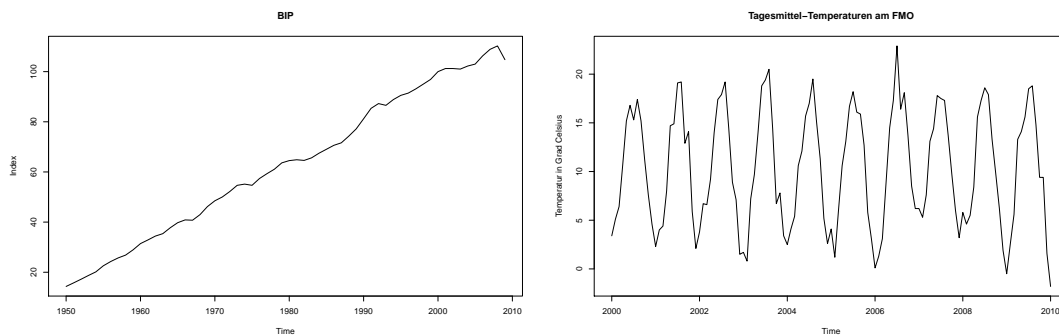


Abbildung 1.1: Bruttoinlandsprodukt Jahresdaten 1950-2009 sowie monatliche Durchschnitte der Tagesmitteltemperaturen am Flughafen Münster-Osnabrück

Wir beobachten gewisse Muster, so ist das BIP im Großen und Ganzen gesehen wachsend, die Temperaturkurve hat einen wiederkehrenden jährlichen Verlauf, der alle weiteren Informationen deutlich überlagert. Doch welche Informationen sind überhaupt gemeint?

Im *klassischen Komponentenmodell* (vgl. [4, S. 9]) geht man für Zeitreihen von folgenden vier Komponenten aus:

- (a) dem *Trend*, das ist eine langfristige systematische Veränderung des mittleren Niveaus der Zeitreihe;
- (b) einer *Konjunkturkomponente*, die eine mehrjährige, nicht notwendig regelmäßige Schwankung darstellt;

- (c) der *Saison*, das ist eine jahreszeitlich bedingte Schwankungskomponente, die sich relativ unverändert jedes Jahr wiederholt
- (d) der *Restkomponente*, die die nicht zu erklärenden Einflüsse oder Störungen zusammenfasst.

Die ersten beiden Komponenten werden bisweilen zu einer einzigen, der sogenannten *glatten Komponente* zusammengefasst. Alternativ fasst man manchmal die Komponenten (ii) und (iii) zur sogenannten *zyklischen Komponente* zusammen. Die Summe der Saisonkomponenten sollte stets Null ergeben.

Da die Unterscheidung zwischen Trend und Konjunkturkomponente rein wirtschaftswissenschaftlicher Natur ist, werden wir uns im Folgenden auf die definitorisch besser voneinander abgrenzbaren Komponenten Trend, Saison und Residuen (Restkomponente) beschränken.

Die Zerlegung in die genannten Komponenten ist dennoch nicht eindeutig, und wir werden im Laufe des Praktikums verschiedene Ansätze kennenlernen, deren Plausibilität stets auch von Hintergrundinformationen über die Zeitreihe abhängen wird, wie z.B. zugrundeliegenden physikalischen Gesetzen oder ökonomischen Theorien.

Der Trend μ_t ist je nach Ansatz eine Gerade, oder auch eine langsam variierende Funktion der Zeit, die Saisonkomponente s_t ist typischerweise periodisch (12-periodisch bei Monatsdaten, 4-periodisch bei Quartalsdaten usw.), die Restkomponente r_t ist in der stochastisch fundierten Zeitreihenanalyse die interessante Komponente, da sie ja gerade die zufällige / nicht beschreibbare Komponente ist. Mit der Auswahl geeigneter stochastischer Modelle für die Residuen werden wir uns hauptsächlich in diesem Praktikum beschäftigen; zunächst lernen wir jedoch einige Ansätze zur Zerlegung gegebener Zeitreihen in die beschriebenen Komponenten kennen.

1.1.2 Additives und multiplikatives Modell

Für das Zusammenspiel der Komponenten gibt es im wesentlichen zwei Ansätze, zwischen denen zumeist bereits durch grafische Darstellung der Zeitreihe unterschieden werden kann. Ist die Streubreite um einen augenscheinlichen Trend zeitlich konstant, so kann von einem *additiven Modell*

$$x_t = \mu_t + s_t + r_t \tag{1.1.1}$$

ausgegangen werden, welche wir auch im Folgenden untersuchen wollen.

Ein Beispiel für eine Zeitreihe, die mit einem additiven Modell behandelt werden kann, sind die monatlichen Daten über die US-amerikanische Arbeitslosenquote:

1.1.3 multiplikatives Modell

Wächst hingegen die Streubreite mit wachsenden Werten der Zeitreihe, so sollte von einem *multiplikativen Modell*

$$x_t = \mu_t \cdot s_t \cdot r_t \tag{1.1.2}$$

ausgegangen werden.

Ein Beispiel für eine Zeitreihe, die mit einem multiplikativen Modell behandelt werden kann, sind die monatlichen Passagierzahlen einer US-amerikanischen Fluglinie.

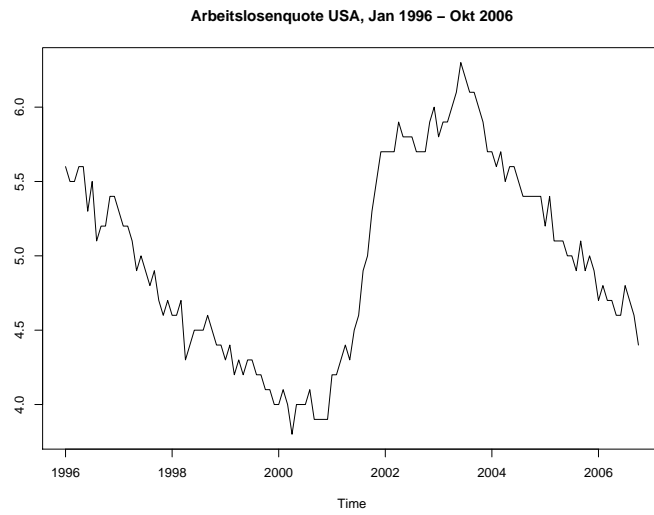


Abbildung 1.2: Arbeitslosenquote der Vereinigten Staaten, monatliche Daten von Januar 1996 bis Oktober 2006

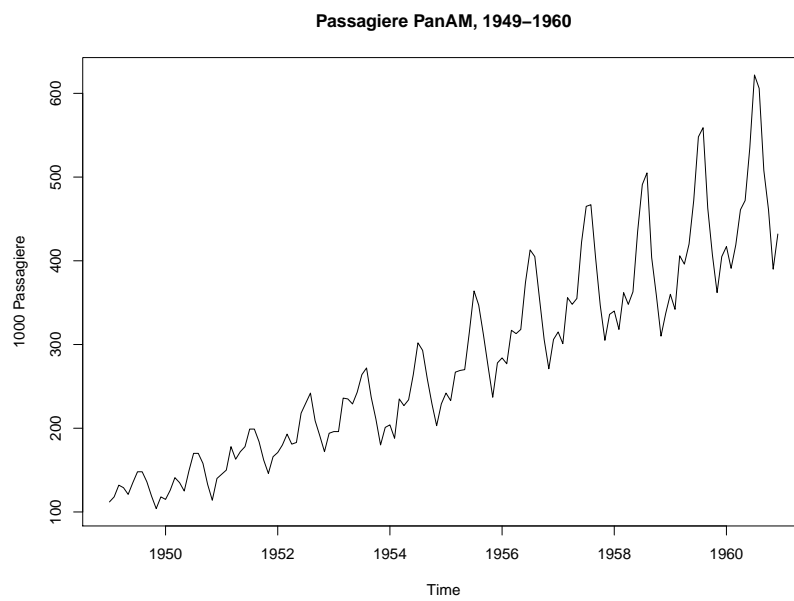


Abbildung 1.3: Passagierzahlen (in Tausend) der PanAm, 1949-1960, Monatsdaten

Multiplikative Modelle können durch Logarithmieren der Beobachtungswerte in ein additives Modell überführt werden. Auch andere Transformationen, um die Streubreite zu harmonisieren, sind möglich (\rightarrow Box-Cox-Transformationen). Will man jedoch die im transformierten Modell gewonnenen Erkenntnisse, z.B. Prognosen, auf das ursprüngliche Modell übertragen, so muss man Vorsicht walten lassen, bereits beim Übergang vom multiplikativen zum additiven Modell. Diese Problematik kann jedoch erst im Rahmen stochastischer Modelle für Zeitreihen behandelt werden.

1.2 Darstellung von Zeitreihen

Wir lernen zunächst ein (auch vom Rechenaufwand her) einfaches Verfahren kennen, um Zeitreihen in die oben beschriebenen Komponenten zu zerlegen, und damit zugleich die in der Zeitreihe enthaltenen Informationen aufzubereiten.

1.2.1 Gleitender Durchschnitt

Denken wir an die Darstellung der Arbeitslosenquote in den USA, so haben wir vermutlich zuerst auf den groben Verlauf geachtet, und uns eine geeignete Kurve vorgestellt, um welche die Werte zu schwanken schienen. Möchten wir diese Kurve nun berechnen, so müssen die Schwankungen eliminiert werden. Diese werden typischerweise durch Mittelbildung ausgeglichen.

Möchte man also den Trend als eine langsam variierende Funktion aus den beobachteten Daten berechnen, so bietet es sich an, zu jedem Zeitpunkt nicht den beobachteten Wert selbst, sondern eine Mittelung aus benachbarten Werten zu betrachten. Gewiss nicht das Mittel über die gesamte Reihe, dies würde eine zeitlich konstante neue Reihe liefern, sondern etwa über a linke wie rechte Nachbarwerte. Dies ist die Methode des gleitenden Durchschnitts: Für $a \in \mathbb{N}$ definiert man

$$\mu_t := \frac{1}{2a+1} \sum_{k=-a}^a x_{t+k}. \quad (1.2.1)$$

An der Stelle t wird also das arithmetische Mittel aus x_t , und den a links und rechts benachbarten Werten gebildet. Große Werte für a führen zu glatten Trends, kleine Werte für a zu weniger glatten Trends.

Beachte, dass so am Anfang und Ende des Beobachtungszeitraum a Werte für μ_t undefiniert bleiben. Dies muss insbesondere bei der Berechnung von Saisonkomponenten beachtet werden.

Betrachtet man statt der konstanten Gewichtung mit $\frac{1}{2a+1}$ allgemein absolut summierbare Folgen $(\lambda_i)_{i \in \mathbb{Z}}$ (sogenannte *lineare Filter*), so ist die *gefilterte Zeitreihe* gegeben durch

$$\mu_t := \sum_{k=-\infty}^{\infty} \lambda_k x_{t+k}. \quad (1.2.2)$$

Der gleitende Durchschnitt ist also der *zentrierte* Filter $\frac{1}{2a+1}(1, 1, \dots, 1)$. Dabei bedeutet zentriert, dass der $a+1$ te (also der mittlere) Eintrag des Filters gerade dem Gewicht für x_t (bei der Berechnung von μ_t) entspricht.

In R benutzt man den Befehl `filter` mit `method="c"`, sowie `sides=2` für zentrierte Filter.

1.2.2 Trendbestimmung mit saisonalen Einflüssen

Liegen keine saisonalen Einflüsse vor, so kann der Trend mittels des gleitenden Durchschnitts und beliebig gewähltem a bestimmt werden, wobei ein großes a das Augenmerk auf langfristige Änderungen legt, kleines a hingegen kurzfristige Änderungen im mittleren Niveau widerspiegelt.

Liegen hingegen (über verschiedene Jahre hinweg konstante) saisonale Einflüsse vor, so muss die Breite so gewählt werden, dass stets über ganze Perioden gemittelt wird, sonst würden die saisonalen Einflüsse im Trend noch erhalten bleiben. Ein intuitiver Ansatz für monatliche Daten wäre, den gleitenden Durchschnitt über 12 Monate zu bilden. Aber welchem Monat sollen wir beispielsweise das Mittel

$$\frac{1}{12}x_1 + \dots + \frac{1}{12}x_{12}$$

zuordnen? Juni oder Juli?

Deshalb wird stattdessen der durch

$$\left(\frac{1}{24}, \underbrace{\frac{1}{12}, \dots, \frac{1}{12}}_{11 \text{ mal}}, \frac{1}{24}\right)$$

definierte zentrierte Filter angewendet, d.h.

$$\mu_t = \frac{1}{24}x_{t-6} + \frac{1}{12} \sum_{k=-5}^5 x_{t+k} + \frac{1}{24}x_{t+6}. \quad (1.2.3)$$

Hier ist jeder Monat mit dem gleichen Gewicht vertreten, und die Summe ist um x_t zentriert. Bei Quartalsdaten verfährt man analog.

Den so berechneten Trend für die US-Arbeitslosenquote sehen wir hier:

1.2.3 Bestimmung der Saisonkomponente

Ist der Trend nach obiger Methode saisonbereinigt berechnet, so gilt im additiven Modell

$$s_t = (x_t - \mu_t) - r_t.$$

Da die Saisonkomponente gerade als periodisch definiert wurde, also bspw. bei Jahresdaten $s_t = s_{t+12}$ gelten soll, würde ohne Restkomponente gelten

$$(x_1 - \mu_1) = s_1 = s_{13} = (x_{13} - \mu_{13}) = \dots$$

Da die tatsächlichen Differenzen $(x_t - \mu_t)$ aber nicht periodisch sind, sondern gerade Schwankungen aufweisen, berechnen wir die Saisonkomponenten als arithmetisches Mittel aus allen zum jeweiligen Monat (bzw. Quartal) zugehörigen Differenzen $(x_t - \mu_t)$, d.h. für $n = k \cdot 12$ Beobachtungen

$$s_i = \frac{1}{k} \sum_{t=0}^{k-1} (x_{i+12t} - \mu_{i+12t}) \quad (1.2.4)$$

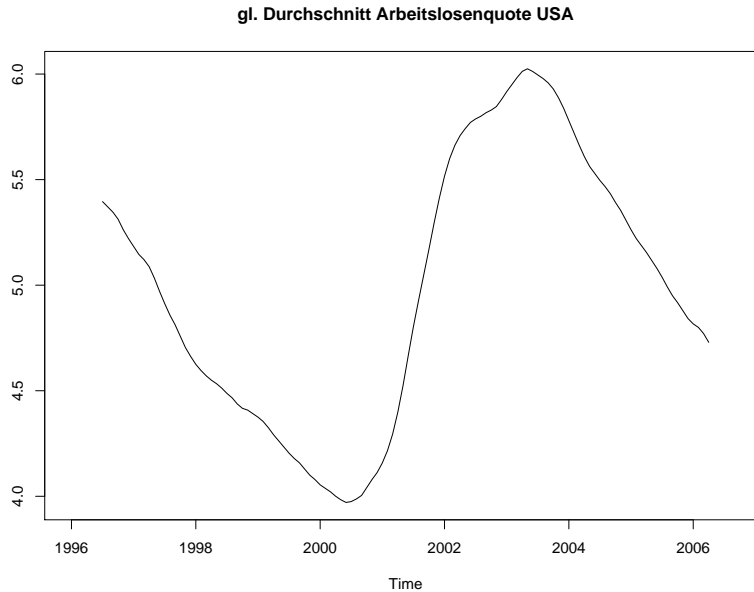


Abbildung 1.4: Gleitender Durchschnitt (Jahresfilter) der Arbeitslosenquote der Vereinigten Staaten

für $1 \leq i \leq 12$. Beachte, dass bei Bestimmung von μ_t als gleitendem Durchschnitt Randwerte für μ_t fehlen, die entsprechenden Summanden werden dann bei der Berechnung der Saisonkomponente ausgelassen.

Eventuell ist es nötig, die Saisonkomponenten anschließend zu *zentrieren*, dazu wird ihr arithmetische Mittel berechnet, und (falls $\neq 0$), von allen Saisonkomponenten abgezogen. Diese additive Konstante wird dann dem Trend zugerechnet.

Die hier beschriebene Zerlegung ist in R in der Funktion `decompose` implementiert.

1.2.4 Vergleich mit Vorjahresdaten

Bei Zeitreihen mit Saisoneinflüssen wird oft auch ein Vergleich mit dem Vorjahreszeitraum gezogen, da sich in einem additiven Modell die Saisoneinflüsse so aufheben. Man wendet den einseitigen Filter $\lambda_{-12} = -1$, $\lambda_0 = 1$, alle übrigen Einträge 0, an, welcher zur gefilterten Zeitreihe $y_t = x_t - x_{t-12}$ führt. Für einseitige Filter muss in R die Option `sides=1` gewählt werden, die Gewichte werden dann als $(\lambda_0, \lambda_{-1}, \dots)$ interpretiert.

1.3 Parametrische Modelle für Trend und Saisonkomponente

Nimmt man über die Annahme eines additiven Modells hinaus weiterhin an, dass der Trend eine gewisse (meist lineare) Funktion der Zeit ist, so können die zugehörigen Parameter (Achsenabschnitt und Steigung) mit Methoden der linearen Regression bestimmt werden. Da in

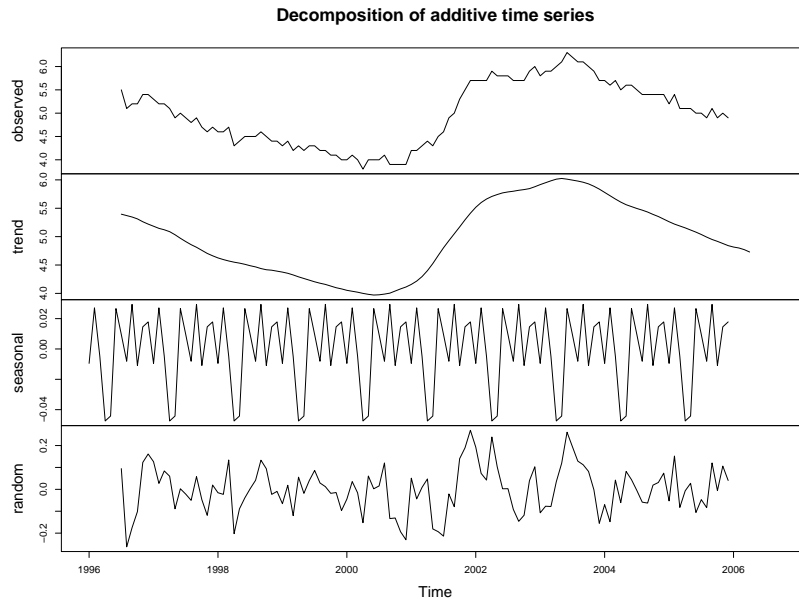


Abbildung 1.5: Anwendung von `decompose` auf die Arbeitslosenraten

diese Berechnung alle beobachteten Werte einfließen, spricht man von einem globalen Ansatz - der gleitende Durchschnitt ist ein lokaler Ansatz, da nur benachbarte Werte in die Berechnung des Trends zum Zeitpunkt t einfließen.

1.3.1 Ohne Saisonkomponente

Ein typischer Ansatz in einem Modell ohne Saisonkomponente ist

$$x_t = \alpha + \beta \cdot t + r_t$$

mit unbekanntem Parametern α und β , die mit der Methode der Kleinsten Quadrate geschätzt werden können. In R ist dies im Befehl `lm` (für lineares Modell) implementiert. Für Stochastiker: Beachte, dass im Allgemeinen die Annahmen eines linearen Modells nicht erfüllt sind, da wir bei Zeitreihen gerade gewisse Abhängigkeiten der Fehler- oder Restterme r_t unterstellen. Es gibt Korrekturmöglichkeiten, fürs erste genügt uns aber, dass die korrigierten Ergebnisse nicht stark von den mit KQS-Schätzern erhaltenen abweichen.

1.3.2 Regressionsmodelle für die Saisonkomponente

Es gibt zwei Ansätze: Den zuvor verfolgten Ansatz einer Folge von 12 monatlichen Komponenten, welche (bei linearem Trend) dem Modell

$$x_t = \alpha + \beta \cdot t + s_i + r_t; \quad i \equiv t(\bmod 12)$$

entspricht; und die Modellierung mittels trigonometrischer Funktionen

$$x_t = \alpha + \beta \cdot t + \left(\sum_{k=1}^6 \gamma_k \cos\left(\frac{2\pi}{k}t\right) + \eta_k \sin\left(\frac{2\pi}{k}t\right) \right) + r_t$$

mit zu bestimmenden Parametern $\alpha, \beta, \gamma_k, \eta_k, 1 \leq k \leq 6$. Dabei kann auf η_1 verzichtet werden - überlege, welche Werte $\sin(2\pi t)$ für $t \in \mathbb{Z}$ annimmt.

1.3.3 Prognose

Hat man eine Zeitreihe bis zum Zeitpunkt T beobachtet, und einen linearen Trend und eventuelle Saisonkomponenten berechnet, so kann man *unter der Annahme*, dass das Modell auch für zukünftige Werte zutreffend ist (insbesondere sich also der Trend fortsetzt), Prognosewerte berechnen:

$$\hat{x}_{T+k|T} = \alpha + \beta \cdot (T + k) + S_{[T+k]}.$$

Dabei bezeichnet $S_{[T+k]}$ die zum Zeitpunkt $T + k$ gehörige Saisonkomponente.

Dieses Verfahren hat jedoch zwei Nachteile: Zum einen setzt es explizit einen linearen (oder zumindest funktional festgelegten) Trend voraus, zum anderen ist ein Update aufwendig: Liegen neue Beobachtungen vor, so muss die gesamte Regression erneut durchgeführt werden.

1.4 Exponentielle Glättung und adaptive Prognose

1.4.1 Exponentielle Glättung

Ein weiteres Verfahren, einen Trend zu berechnen, ist das sogenannte exponentielle Glätten; welches insbesondere für die Prognose den Vorteil hat, dass es sich ohne großen Rechenaufwand updaten lässt.

Ihm liegt die Idee zugrunde, dass sich künftige Beobachtungen eher auf dem Niveau von Werten aus der jüngeren Vergangenheit befinden werden, als durch weit zurückliegende Werte beeinflusst zu werden - man denke an die Zeitreihe der Bundesligaplatzierungen eines beliebigen Vereins. Die Erfolge der 60er Jahre mögen allenfalls ein finanzielles Polster angelegt haben, wesentlich relevanter ist jedoch die Qualität der aktuellen Mannschaft, die sich am besten durch die Vorjahresplatzierung einschätzen lässt.

Formal führen wir zunächst rekursiv das Niveau ein (der Trend in der Sprechweise der exponentiellen Glättung). Unsere Beobachtungen mögen zum Zeitpunkt 0 beginnen, dann setzen wir $n_0 = x_0$, und

$$n_t := \alpha x_t + (1 - \alpha)n_{t-1}$$

für $t \geq 1$. Der auftauchende Parameter α heißt Glättungsparameter, und entstammt sinnvollerweise dem Intervall $(0, 1)$. Großes α bedeutet starke "Gegenwartsorientierung" des Niveaus, ein kleines α räumt vergangenen Werten mehr Gewicht ein, wie auch die explizite Formel

$$n_t := \sum_{k=0}^{t-1} (1 - \alpha)^k \alpha x_{t-k} + (1 - \alpha)^t x_0$$

zeigt.

Als Prognose nehmen wir nun einfach an, dass zukünftige Werte auf dem berechneten Niveau liegen, d.h.

$$\hat{x}_{T+k|T} = n_T.$$

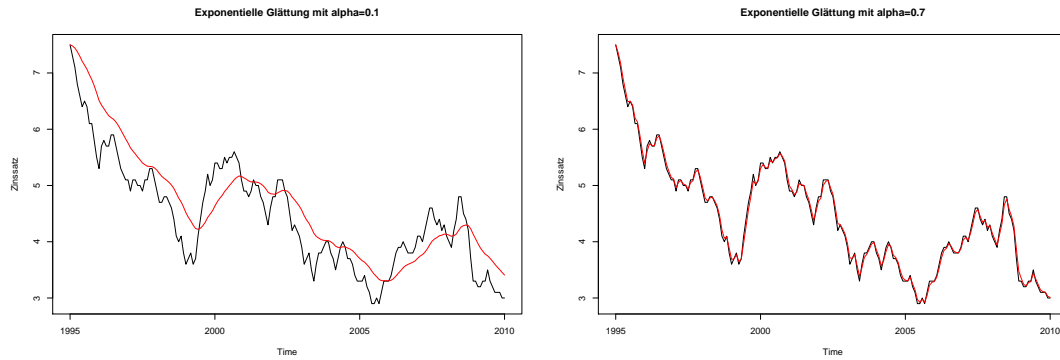


Abbildung 1.6: Umlaufrendite inländischer Inhaberschuldverschreibungen, exponentiell geglättet mit Parametern $\alpha = 0.1$ bzw $\alpha = 0.9$.

Aus der rekursiven Definition wird deutlich, dass die Aktualisierung des Niveaus sehr einfach ist, neben α müssen nur der neue beobachtete Wert x_{T+1} und das aktuelle Niveau n_T bekannt sein. Bei Regressionsverfahren hätten wir alle bisher beobachteten Werte speichern müssen.

1.4.2 Bestimmung des optimalen α

Wie aber ist α zu wählen? Es lässt sich ein Gütekriterium angeben, und zwar die sogenannte mittlere Summe der quadrierten 1-Schritt-Prognose-Fehler: Für jedes $t \geq 1$ wird n_{t-1} (mittels α) berechnet, und der prognostizierte Wert n_{t-1} mit dem tatsächlich beobachteten Wert x_t verglichen. Als optimal sehen wir nun ein α an, welches die mittlere Summe der quadrierten Fehler minimiert.

$$\alpha \stackrel{!}{=} \arg \min \frac{1}{T} \sum_{t=1}^T (x_t - \hat{x}_{t|t-1}(\alpha))^2 = \arg \min \frac{1}{T} \sum_{t=1}^T (x_t - n_{t-1}(\alpha))^2 \quad (1.4.1)$$

1.4.3 Exponentielles Glätten mit Wachstum

Betrachte nochmals die mit $\alpha = 0.1$ geglättete Zeitreihe der Umlaufrendite. Bei fallenden Werten liegt das Niveau meist über dem tatsächlichen Verlauf, umgekehrt bei steigenden Werten. Das dies eine generelle Schwäche ist, bestätigt ein erneuter Blick auf die Berechnungsformel. Bei fallenden Werten wird ein gewichtetes Mittel aus dem aktuellen Wert, und zurückliegenden, höheren Werten gebildet, so dass das Ergebnis höher ist als der aktuelle Wert.

Deshalb wird das Verfahren erweitert, indem neben dem Niveau auch eine Steigung m_t mit der Methode des exponentiellen Glättens berechnet wird:

$$\begin{aligned} n_0 &= x_0, & n_1 &= x_1; \\ m_0 &= x_1 - x_0, & m_1 &= m_0. \end{aligned}$$

$$\begin{aligned} n_t &= \alpha x_t + (1 - \alpha)(n_{t-1} + m_{t-1}) \\ m_t &= \beta(n_t - n_{t-1}) + (1 - \beta)m_{t-1} \end{aligned}$$

So wird das vergangene Niveau entsprechend angehoben / abgesenkt, bevor es in die Berechnung des aktuellen Niveaus einfließt.

1.4.4 Holt-Winters-Verfahren

Schließlich kann das Niveau durch saisonale Einflüsse verfälscht werden. Führt man deshalb noch Saisonkomponenten s_t ein, die ebenfalls permanent aktualisiert werden, so gelangt man zum Holt-Winters-Verfahren, welches exponentielle Glättung mit Saison und Wachstum darstellt. Die Aktualisierungsformeln lauten dann für Zeitreihen mit Periode p und additiver Saisonkomponente

$$\begin{aligned}n_t &= \alpha(x_t - s_{t-p}) + (1 - \alpha)(n_{t-1} + m_{t-1}) \\m_t &= \beta(n_t - n_{t-1}) + (1 - \beta)m_{t-1} \\s_t &= \gamma(x_t - a_t) + (1 - \gamma)s_{t-p}.\end{aligned}$$

Etwas subtiler ist die Berechnung der Startwerte; diese müssen offensichtlich für eine gesamte Periode angegeben werden. Am einfachsten ist die Steigung: $m_t = \frac{1}{p}(x_{p-1} - x_0)$ für $0 \leq t < p$. Eine Gerade mit dieser gegebenen Steigung wird nun mit der Methode der kleinsten Quadrate an die Daten x_0, \dots, x_{p-1} angepasst (im Endeffekt also nur der Achsenabschnitt berechnet), und s_t wird dann als Abweichung von x_t von der berechneten Gerade festgesetzt. Dies führt zur Minimierung von $\sum_{t=0}^{p-1} s_t^2$ (bei gegebener Steigung der Geraden). Schließlich wird $n_t = x_t - s_t$ gesetzt für $0 \leq t < p$.

Das Holt-Winters-Verfahren ist in R als `HoltWinters` implementiert, und umfasst alle beschriebenen Funktionen. Die Parameter α, β, γ können angegeben werden, ansonsten werden sie als (möglicherweise nur lokale!) Minima der Summe der quadrierten Ein-Schritt-Prognose-Fehler bestimmt. Vorhersagen macht die Funktion `predict`.

1.5 R-Befehle

`ts` In R erstellt man Zeitreihen mit dem Befehl `ts(data, start, frequency)`. Dabei kann `data` ein Vektor, eine Matrix oder eine Datentabelle sein, je nachdem, wieviele Zeitreihen man in einer Variablen zusammenfassen will; `start` muss ein Vektor mit 2 Komponenten sein: Die erste gibt das Startjahr an, die zweite den Startmonat. Der Parameter `frequency` gibt die Häufigkeit der Beobachtungen pro Zeiteinheit an, etwa 12, wenn man 12 Beobachtungen pro Jahr hat

`window` Mit dem Befehl `window(x, start, end, frequency)` kann man einen Ausschnitt aus einer Zeitreihe herausnehmen

`frequency(x)` Gibt die Frequenz der Zeitreihe `x` aus

`start(x)` gibt das Startjahr und den Startmonat von `x` aus, analog `end(x)`

`plot(x)` plottet die Zeitreihe `x`

`ts.plot` Will man mehrere Zeitreihen in einem Plot anzeigen lassen, so ist diese Funktion nützlich: In ihr lassen sich beliebig viele Zeitreihen gleicher Frequenz übergeben.

- decompose** Zerlegt eine Zeitreihe in Trend-, Saison- und Restkomponente. Der Trend wird als gleitender Durchschnitt mit dem zentrierten Filter $\frac{1}{p}(\frac{1}{2}, 1, \dots, \frac{1}{2})$ berechnet, die Saisonkomponenten als mittlere monatliche Abweichungen. Auf die gewonnenen Zerlegungen kann (als Zeitreihen) mittels `decompose(Zeitreihe)$seasonal`, `$trend` bzw. `$random` zugegriffen werden.
- filter** Wendet einen angegebenen Filter auf eine Zeitreihe an, und liefert die gefilterte Zeitreihe zurück. Für die oben beschriebenen Filter muss als Argument `method="convolution"` übergeben werden, die Filter selbst werden als zentrierte Filter interpretiert, und als Vektor übergeben. Das exponentielle Glätten kann als rekursiver Filter angesehen werden (`method="recursive"`), hier müssen zusätzlich Startwerte übergeben werden.
- predict** Nimmt als Argument ein Modell (z.B. die Ausgabe von `HoltWinters`, oder `lm`), und liefert eine Zeitreihe mit anhand des Modells vorhergesagten Werten. Die Anzahl der Vorhersagen wird als Argument `n.ahead=Anzahl` übergeben.
- HoltWinters** Implementierung des exponentiellen Glättens, nimmt als Argument eine Zeitreihe, und Werte für α , β und γ . Soll keine Steigung und/oder Saisonkomponente berechnet werden, so müssen β bzw. γ Null gesetzt werden. Ohne Angabe von Werten werden diese anhand der 1-Schritt-Prognose-Fehler bestimmt.
- lm** Methode zur Behandlung von linearen Modellen, insbesondere für lineare Regression. Der Ansatz über die funktionale Abhängigkeit wird in der Form `x ~ time(x)` beispielsweise für einen linearen Trend übergeben. Um Saisonkomponenten mit zubestimmen, muss der Befehl
- ```
lm(Zeitreihe ~ time(Zeitreihe) +factor(cycle(Zeitreihe)))
```
- verwendet werden. Die Funktionen `cycle` und `time`, angewendet auf eine Zeitreihe, liefern das vermutete. Der Befehl `factor` sorgt insbesondere dafür, dass die Saisonkomponente nicht als lineare Funktion des Monatswertes (z.B. Juni=6), sondern für jeden Monat als einzelner Wert bestimmt wird.
- Sollen die Residuen weiter untersucht werden, so müssen diese wieder in eine Zeitreihe umgewandelt werden, dies geschieht mit dem Befehl `ts`.

### 1.5.1 Saisonkomponenten aus linearen Modellen berechnen

Ein lineares Modell in der oben beschriebenen Form

```
lm(Zeitreihe ~ time(Zeitreihe) +factor(cycle(Zeitreihe)))
```

liefert nur 11 Saisonkomponenten  $\tilde{s}_2, \dots, \tilde{s}_{12}$ , sowie die additive Konstante (Intercept)  $\tilde{a}$ . In diesem ist die erste Saisonkomponente bereits eingerechnet, für die zentrierten Saisonkomponenten  $s_i$  und die zugehörige additive Konstante  $a$  gilt:

$$\tilde{a} = (a + s_1), \quad \tilde{s}_i = (s_i - s_1)$$

Um zu den zentrierten Saisonkomponenten  $s_i$  zu gelangen, bilden wir zunächst den Vektor

$$(\tilde{a}, \tilde{s}_2 + \tilde{a}, \dots, \tilde{s}_{12} + \tilde{a}) = (a + s_1, \dots, a + s_{12}),$$

berechnen dessen arithmetisches Mittel, welches gerade  $a$  entspricht, und erhalten schließlich die zentrierten Saisonkomponenten, indem wir von jedem Eintrag  $a$  subtrahieren.

Diese Umrechnungen sind jedoch nur notwendig, wenn Sie die zentrierten Saisonkomponenten angeben oder plotten wollen, für Vorhersagen ist die R-interne Darstellung äquivalent.

### 1.5.2 Vorhersagen aus linearen Modellen

Diese geschehen mit der Funktion `predict`, allerdings funktioniert der Parameter `n.ahead` bei linearen Modellen (z.B. Regressionsmodelle zur Bestimmung eines linearen Trends) nicht. Stattdessen muss `newdata` angegeben werden, und zwar in subtiler Form. Diese lässt sich anhand eines Beispiels am besten nachvollziehen. Sei `x` eine Zeitreihe. Die Variablenzuweisungen zu Beginn sind hier nötig.

```
ZEIT<-time(x); SAISON<-factor(cycle(x))
modell<-lm(x ~ ZEIT + SAISON)
y<-ts(start=c(...),end=c(...),f=12)
Bereich<-data.frame(ZEIT=time(y), SAISON=factor(cycle(y)))
y.dat<-predict(modell,newdata=Bereich)
y<-ts(y.dat,start=c(...),end=c(...),f=12)
```

Die Zeitreihe `y` enthält die vorhergesagten Werte, bei Start, Ende und Frequenz sind die entsprechenden Daten einzutragen, typischerweise wird der Beginn von `y` nach dem Ende von `x` liegen, und die Frequenz muss gleich bleiben.

## 2 AR-Modelle

Bislang wurden nur deskriptive und modellfreie Methoden vorgestellt und kein stochastisches Modell, mit dem wir „rechnen“ können. Dies werden wir in diesem Kapitel ändern: Wir werden den Begriff der schwachen Stationarität, AR(p) Modelle und einige – für die Zeitreihenanalyse interessante – Tests kennenlernen.

### 2.1 schwach Stationäre Prozesse

Für theoretische Überlegungen nehmen wir weiterhin an, dass eine Zeitreihen-Beobachtung  $x_1, \dots, x_n$  von einem stochastischen Prozess (im Folgenden abkürzend ZR-Prozess genannt)

$$X = (X_t)_{t \in \mathbb{Z}} \quad (2.1.1)$$

stammt, der weit weit in der Vergangenheit gestartet ist und noch bis in alle Ewigkeit läuft.

Als potentielle Modelle wollen wir nicht *alle* stochastischen Prozesse zulassen, sondern uns auf diejenigen beschränken, die „brauchbare“ Eigenschaften besitzen. Im Besonderen sind wir natürlich daran interessiert, Vorhersagen für konkrete Datensätze treffen zu können.

#### 2.1.1 Definition und Beispiele

Das einfachste Beispiel eines stochastischen Prozesses kennen wir bereits aus elementaren Stochastik-Vorlesungen, auch wenn es dort üblicher ist von einer „Folge von Zufallsgrößen“ zu sprechen.

**Beispiel 2.1.1** (IID White Noise) Der Prozess  $(W_t)_{t \in \mathbb{Z}}$  bestehe aus stochastisch unabhängig und identisch verteilten Zufallsgrößen mit endlichem Erwartungswert  $\mathbb{E}W_0 = \mu$  und endlicher Varianz  $\text{Var}(W_0) = \sigma^2$ . Im Folgenden schreiben wir abkürzend  $(W_t) \sim \text{IID}(\mu, \sigma^2)$  für diesen Prozess.

Für einen  $\text{IID}(\mu, \sigma^2)$  ist keine Vorhersage möglich (Unabhängigkeit!), er ist aber ein wichtiger Baustein für weitere Prozesse, etwa in dem wir die Partialsummen  $S_t = W_1 + \dots + W_t$  bilden. Dabei müssen wir einen „Startpunkt“  $S_0$  festlegen, d. h. die Zeitparametermenge ist in diesem Fall nicht  $\mathbb{Z}$ , sondern  $\mathbb{N}_0$ . Für unsere Zwecke ist es ausreichend wenn wir  $S_0 = 0$  setzen, genauer:

**Beispiel 2.1.2** (Random Walk) Der Prozess  $(S_t)_{t \in \mathbb{N}_0}$  sei definiert durch

$$S_t = \sum_{i=1}^t W_i = S_{t-1} + W_t \quad (2.1.2)$$

für  $t = 1, 2, \dots$ , wobei  $(W_t) \sim \text{IID}(\mu, \sigma^2)$  und  $S_0 = 0$ .  $(S_n)_{n \in \mathbb{N}_0}$  heißt *Random-Walk* mit Startpunkt  $S_0$  und Zuwachsverteilung  $F =$  Verteilung von  $W_0$ .



Der Pfad eines Random-Walkes ist wesentlich „glatter“ als der eines IID( $\mu, \sigma^2$ ) Prozesses, was Abbildung 2.1 verdeutlicht.

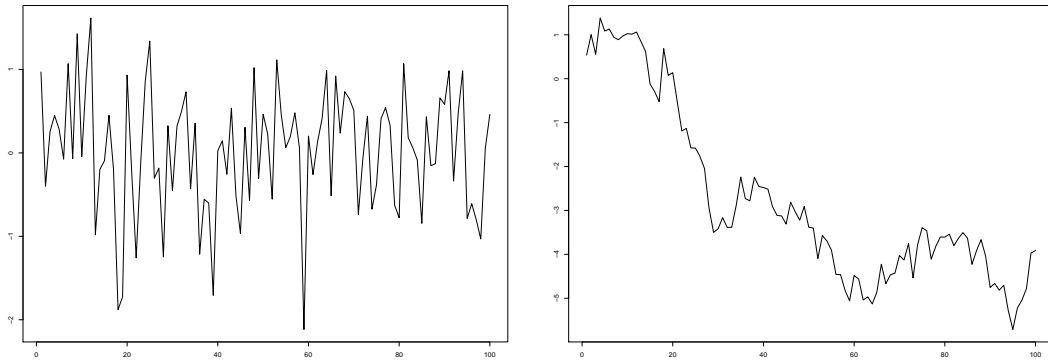


Abbildung 2.1: Sample eines IID(0,0.5) Prozesses (Normalverteilung) und des zugehörigen Random-Walkes (hier ist  $S_0 = W_0$ ).

Wie schon erwähnt wollen wir nicht die (viel zu große) „Menge aller Prozesse“ untersuchen, sondern uns auf eine Teilmenge einschränken, die gewisse Eigenschaften besitzt. Aus diesem Grund führen wir nun die Klasse der schwach stationären Prozesse ein: Sie ist „groß genug“ für die Modellierung vieler Zeitreihen und besitzt zudem gute Vorhersage-Eigenschaften.

**Definition 2.1.3** (schwache Stationarität) Ein ZR-Prozess  $X = (X_t)_{t \in \mathbb{Z}}$  heißt *schwach stationär*, falls jedes  $X_i$  endliche Varianz besitzt und es gilt:

- (a)  $\mathbb{E}X_t = \mu$  für alle  $t \in \mathbb{Z}$
- (b)  $\text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_0, X_h)$  für alle  $t, h \in \mathbb{Z}$

(b) besagt, dass die Kovarianz von  $X_t$  und  $X_s$  nur vom Abstand zwischen  $t$  und  $s$  abhängt und nicht von  $t$  und  $s$  selber. Für diese Kenngröße ist es also egal, ob wir  $X$  zwischen 12 Uhr und 13 Uhr, oder zwischen 18 Uhr und 19 Uhr beobachten.

Die Funktion  $(t, s) \mapsto \text{Cov}(X_t, X_s)$  ist so wichtig, dass sie einen eigenen Namen erhält.

**Definition 2.1.4** Es sei  $X = (X_t)_{t \in \mathbb{Z}}$  ein ZR-Prozess.

- (a) Die Funktion  $\gamma(t, s) = \text{Cov}(X_t, X_s)$  heißt die zu  $X$  gehörige *Kovarianzfunktion*.
- (b) Ist  $X$  schwach stationär, so gilt

$$\gamma(t, t + h) = \gamma(0, 0 + h)$$

für alle  $t, h \in \mathbb{Z}$ . In diesem Fall können wir

$$\gamma(h) = \gamma(t, t + h) \tag{2.1.3}$$

definieren.  $\gamma$  heißt *Auto-Kovarianzfunktion von  $X$* ,  $h$  heißt auch *Lag*. Es bezeichne weiter

$$\varrho(h) = \frac{\gamma(h)}{\gamma(0)}, \tag{2.1.4}$$

die *Autokorrelationsfunktion (ACF) von  $X$* .

Wir beachten, dass die ACF betragsmäßig stets  $\leq 1$  ist. Beispiele schwach stationärer Prozesse werden wir in Kürze kennenlernen, siehe etwa Beispiele 2.1.5 und 2.2.1.

## 2.1.2 empirische Kenngrößen

Dieser Abschnitt dient der Einführung einiger empirischer Kenngrößen für die Beobachtung  $x = (x_1, \dots, x_n)$ .

Das *Stichprobenmittel* bezeichnen wir mit

$$\bar{x}_n = \frac{1}{n} \sum_{t=1}^n x_t. \quad (2.1.5)$$

In R erhalten wir es mit dem Befehl `mean(x)`. Den erwartungstreuen Schätzer für die Varianz

$$\frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x}_n)^2 \quad (2.1.6)$$

bekommen wir mit `var(x)`.

Ein Ansatz, die Autokovarianzfunktion  $\gamma$  zu schätzen ist es,

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x}_n) \cdot (x_t - \bar{x}_n) \quad (2.1.7)$$

für  $h = 0, 1, \dots, n-1$  zu setzen. Wir nennen  $\hat{\gamma}$  die *empirische Auto-Kovarianzfunktion* und beachten, dass  $\hat{\gamma}(0)$ , bis auf den Faktor  $\frac{n-1}{n}$ , dem Ausdruck (2.1.6) entspricht.

Die *empirische Autokorrelationsfunktion* (*eACF*) definieren wir als

$$\hat{\varrho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \quad (2.1.8)$$

Um (2.1.8) definieren zu können, müssen wir  $\hat{\gamma}(0) > 0$  voraussetzen. Der Fall  $\hat{\gamma}(0) = 0$  ist für uns aber auch nicht sonderlich interessant, denn es würde  $x_1 = \dots = x_n$  gelten. Daher schließen wir dies immer aus.

Wir definieren weiter die *empirische Auto-Kovarianzmatrix* (*bis zum Lag p*) als die  $p \times p$  Matrix

$$\hat{\Gamma}_p = (\hat{\gamma}(|i-j|))_{i,j=1,\dots,p} = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \dots & \hat{\gamma}(p-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \dots & \hat{\gamma}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(p-1) & \hat{\gamma}(p-2) & \dots & \hat{\gamma}(0) \end{pmatrix} \quad (2.1.9)$$

und die *empirische Auto-Korrelationsmatrix* (*bis zum Lag p*) als

$$\hat{R}_p = (\hat{\varrho}(|i-j|))_{i,j=1,\dots,p} = \frac{1}{\hat{\gamma}(0)} \cdot \hat{\Gamma}_p \quad (2.1.10)$$

Wegen  $\hat{\varrho}(0) = 1$  stehen auf der Diagonalen von  $\hat{R}_p$  nur Einsen.  $\hat{\Gamma}_p$  und  $\hat{R}_p$  sind positiv semidefinit, symmetrisch und invertierbar, denn es gilt

$$\hat{\Gamma}_p = n^{-1} \cdot T \cdot T^t$$

mit der  $p \times 2p$  Matrix

$$T = \begin{pmatrix} 0 & \cdots & 0 & 0 & y_1 & \cdots & y_{p-1} & y_p \\ 0 & \cdots & 0 & y_1 & y_2 & \cdots & y_p & 0 \\ \vdots & & & & & & & \vdots \\ 0 & y_1 & \cdots & y_{p-1} & y_p & 0 & \cdots & 0 \end{pmatrix}$$

und  $y_i = x_i - \bar{x}_n$

Die eACF ist eine wichtige Kenngröße für die Zeitreihen-Analyse, daher sollten wir von wichtigen Zeitreihen-Modellen die (theoretische) ACF kennen, um später anhand einer Beobachtung ein „vernünftiges“ Modell wählen zu können.

In R erhalten wir die eACF mit dem Befehl `acf(x)`, wobei `x` sowohl ein `ts`-Objekt, als auch ein gewöhnlicher Vektor sein kann. Führen wir dies für die Beobachtungen aus Abbildung 2.1 aus, so erkennen wir, dass die empirische ACF der i. i. d. Folge nahezu 0 ist (ab dem Lag 1) und die des Random Walkes linear abfällt, vgl. Abbildung 2.2.

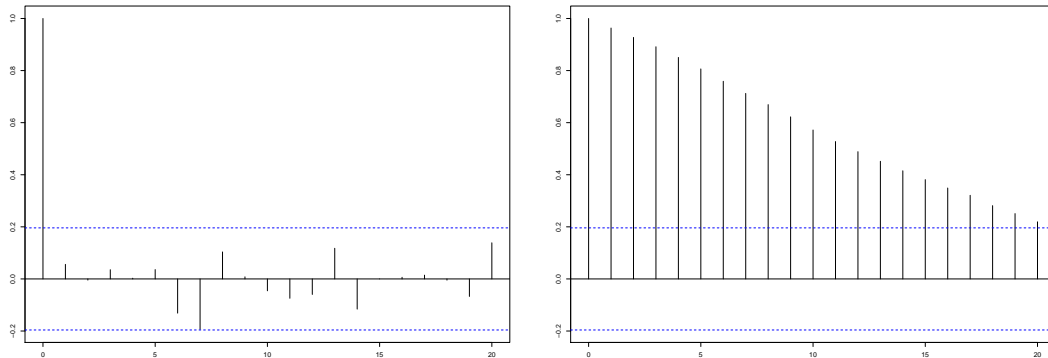


Abbildung 2.2: Empirische Autokorrelationsfunktionen der Zeitreihen aus Abbildung 2.1.

In den nächsten beiden Beispielen zeigt sich, dass der  $\text{IID}(\mu, \sigma^2)$  schwach stationär ist, ein Random-Walk jedoch nicht.

**Beispiel 2.1.5** (White Noise Prozess) Sei  $(W_t) \sim \text{IID}(\mu, \sigma^2)$ , ein IID White Noise Prozess. Für diesen gilt

$$\gamma(t, t+h) = \text{Cov}(W_t, W_{t+h}) = \begin{cases} \text{Var}(W_t) = \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

denn im Fall  $h \neq 0$  sind  $W_t$  und  $W_{t+h}$  stoch. unabh. und somit unkorreliert. Also ist  $W$  schwach stationär mit

$$\varrho(h) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases} \quad (2.1.11)$$

Allgemeiner heie ein Prozess  $(W_t)_{t \in \mathbb{Z}}$  *White Noise Prozess* (kurz  $\text{WN}(\mu, \sigma^2)$ ), falls die  $(W_t)_{t \in \mathbb{Z}}$  paarweise unkorreliert sind,  $\mathbb{E}W_t = \mu$  und  $\text{Var}(W_t) = \sigma^2$  für alle  $t \in \mathbb{Z}$  gilt. Mit der gleichen Rechnung sehen wir, dass auch ein  $\text{WN}(\mu, \sigma^2)$  Prozess schwach stationär mit ACF  $\varrho$  ist.

**Beispiel 2.1.6** (Random Walk) Sei  $(W_t) \sim \text{IID}(0, \sigma^2)$  und  $(S_t)_{t \in \mathbb{N}_0}$  der zugehörige Random-Walk mit Start in 0, d.h.  $S_t = S_{t-1} + W_t$  und  $S_0 = 0$ . Für diesen Prozess gilt  $\mathbb{E}S_t = 0$ ,  $\text{Var}(S_t) = t \cdot \sigma^2$  für alle  $t \geq 1$  und

$$\begin{aligned} \gamma(t, t+h) &= \text{Cov}(S_t, S_{t+h}) \\ &= \text{Cov}(S_t, S_t + W_{t+1} + W_{t+2} + \dots + W_{t+h}) \\ &= \underbrace{\text{Cov}(S_t, S_t)}_{=\text{Var}(S_t)} + \underbrace{\text{Cov}(S_t, W_{t+1})}_{=0} + \dots + \underbrace{\text{Cov}(S_t, W_{t+h})}_{=0} = t \cdot \sigma^2. \end{aligned}$$

Da dies von  $t$  abhängt, ist  $(S_t)_{t \in \mathbb{N}_0}$  nicht schwach stationär.

**Definition 2.1.7** Für einen ZR-Prozess  $(X_t)_{t \in \mathbb{Z}}$  sei der *Differenzen-Prozess*  $(\Delta X_t)_{t \in \mathbb{Z}}$  definiert als

$$\Delta X_t = X_t - X_{t-1} \quad t \in \mathbb{Z}. \quad (2.1.12)$$

Ist  $(S_t)$  ein Random-Walk, so bildet der Differenzen-Prozess  $(\Delta S_t)_{t=1,2,\dots}$  eine Folge unabhängig identisch verteilter Zufallsgrößen. Durch Anwenden des Operators  $\Delta$  wird in diesem Fall aus einem nicht schwach stationären Prozess ein schwach stationärer. Häufig ist es so, dass mehrmaliges Anwenden von  $\Delta$  einen schwach stationären Prozess liefert.

In R berechnen wir den Differenzen-Prozess mit `diff(x)`. Die Funktion bildet aus dem  $n$ -elementigen Vektor  $\mathbf{x} = (\mathbf{x}[1], \dots, \mathbf{x}[n])$  den Differenzenvektor  $(\mathbf{x}[2] - \mathbf{x}[1], \dots, \mathbf{x}[n] - \mathbf{x}[n-1])$  der Länge  $n-1$ .

## 2.2 AR(p) Modelle

In diesem Abschnitt führen wir eine spezielle Klasse stochastischer Prozesse ein, sogenannte autoregressive Prozesse:  $X_t$  soll sich in Abhängigkeit seiner vergangenen Wert  $X_{t-1}, X_{t-2}, \dots$  schreiben lassen, etwa

$$X_t = f(X_{t-1}, X_{t-2}, \dots)$$

für eine Funktion  $f$ . Zusätzlich nehmen wir an, dass zu jedem Zeitpunkt eine Störung (White Noise) in die Beobachtung mit eingeht, d.h.  $X_t$  ist von der Bauart

$$X_t = f(X_{t-1}, X_{t-2}, \dots) + W_t$$

für einen  $\text{WN}(0, \sigma^2)$  Prozess  $(W_t)$ . Bei AR Modellen beschränken wir uns auf *lineare* Funktionen  $f$  und sollten eigentlich von *linearen autoregressiven Modellen* sprechen. Bevor wir diese Modelle in voller Allgemeinheit einführen untersuchen wir das einfache Beispiel  $f(x) = \phi \cdot x$ . Für  $\phi = 1$  erhalten wir einen Random-Walk, d.h. einen nicht schwach stationären Prozess. Für  $|\phi| < 1$  wird der Prozess schwach stationär.

**Beispiel 2.2.1** (AR(1) Modelle) Der Prozess  $(X_t)_{t \in \mathbb{Z}}$  sei definiert durch

$$X_t = \phi \cdot X_{t-1} + W_t \quad (2.2.1)$$

mit  $(W_t) \sim \text{WN}(0, \sigma^2)$  und  $\phi \in \mathbb{R}$  mit  $|\phi| < 1$ . Jedes  $X_t$  lässt sich schreiben als

$$\begin{aligned} X_t &= W_t + \phi \cdot W_{t-1} + \phi^2 \cdot W_{t-2} + \dots \\ &= \sum_{i=0}^{\infty} \phi^i \cdot W_{t-i}, \end{aligned} \tag{2.2.2}$$

denn die Reihe in (2.2.2) konvergiert fast sicher wegen  $|\phi| < 1$ .

Es gilt somit  $\mathbb{E}X_t = 0$  für alle  $t \in \mathbb{Z}$  und

$$\begin{aligned} \gamma(t, t+h) &= \text{Cov}(X_t, X_{t+h}) \\ &= \text{Cov} \left( \sum_{i=0}^{\infty} \phi^i \cdot W_{t-i}, \sum_{j=0}^{\infty} \phi^j \cdot W_{t+h-j} \right) \\ &= \sum_{i=0}^{\infty} \phi^i \cdot \phi^{h+i} \cdot \underbrace{\text{Cov}(W_{t-i}, W_{t-i})}_{=\sigma^2}, \end{aligned}$$

denn für  $j \neq h+i$  sind  $W_{t-i}$  und  $W_{t+h-j}$  unkorreliert (vgl. Definition von WN). Wir erhalten unter erneuter Beachtung von  $|\phi| < 1$

$$\begin{aligned} \gamma(t, t+h) &= \phi^h \cdot \sigma^2 \cdot \sum_{i=0}^{\infty} \phi^{2i} \\ &= \frac{\phi^h \cdot \sigma^2}{1 - \phi^2} \end{aligned}$$

und damit die schwache Stationarität von  $(X_t)$  mit ACF

$$\varrho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h. \tag{2.2.3}$$

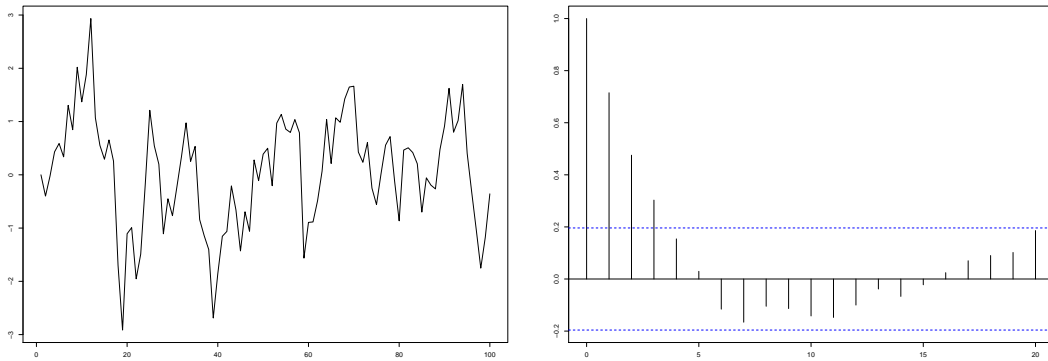


Abbildung 2.3: Sample eines AR(1) Prozesses mit  $\phi = 0.7$  und zugehöriger eACF.

**Definition 2.2.2** (AR(p) Modelle) Ein stochastischer Prozess  $(X_t)$  heißt *autoregressiver Prozess der Ordnung p* (kurz: *AR(p)*) Prozess, falls jedes  $X_t$  die Darstellung

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + W_t \tag{2.2.4}$$

für einen  $\text{WN}(0, \sigma^2)$  Prozess  $(W_t)$  und Konstanten  $\phi_1, \dots, \phi_p \in \mathbb{R}$  mit  $\phi_p \neq 0$  besitzt.  
Im Fall  $p = \infty$ , d. h.

$$X_t = W_t + \sum_{i=1}^{\infty} \phi_i X_{t-i} \quad (2.2.5)$$

müssen wir

$$\sum_{i=1}^{\infty} |\phi_i| < \infty \quad (2.2.6)$$

fordern, um die fast sichere Konvergenz in (2.2.5) zu gewährleisten.

Mit Hilfe des Backshift-Operators  $\mathbf{B}$ , der durch  $\mathbf{B}X_t = X_{t-1}$  (und damit induktiv  $\mathbf{B}^n X_t = X_{t-n}$ ) definiert ist, lässt sich (2.2.4) äquivalent als

$$(1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2 - \dots - \phi_p \mathbf{B}^p) X_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = W_t$$

schreiben. Definieren wir  $\Phi(\mathbf{B}) = 1 - \phi_1 \mathbf{B} - \dots - \phi_p \mathbf{B}^p$ , so lautet die Forderung (2.2.4) in Kurzform:

$$\Phi(\mathbf{B})X_t = W_t \quad \text{für alle } t \in \mathbb{Z}. \quad (2.2.7)$$

Obwohl  $\Phi$  ein Polynom in  $\mathbf{B}$  ist können wir damit rechnen, als wäre es eine Polynomabbildung  $\Phi : \mathbb{C} \rightarrow \mathbb{C}$ . Insbesondere können wir von den Nullstellen von  $\Phi$  sprechen.

Für  $p = 1$  und  $\phi_1 = 1$  erhalten wir einen Random-Walk. Es stellt sich also die Frage, unter welchen Voraussetzungen an die Konstanten  $\phi_1, \dots, \phi_p$  ein  $\text{AR}(p)$  Prozess schwach stationär ist. In der Tat ist dafür wichtig, wo die Nullstellen von  $\Phi$  liegen.

**Satz 2.2.3** (a) Für die Gleichung (2.2.4) gibt es genau dann eine schwach stationäre Lösung  $(X_t)$ , wenn alle Nullstellen von  $\Phi$  außerhalb des Einheitskreises  $\{z \in \mathbb{C} : |z| = 1\}$  liegen. Diese schwach stationäre Lösung ist zudem eindeutig.

(b) Gilt sogar  $|z| > 1$  für alle Nullstellen  $z \in \mathbb{C}$  von  $\Phi$ , so können wir jedes  $X_t$  als

$$X_t = \sum_{i=0}^{\infty} \psi_i \cdot W_{t-i} \quad (2.2.8)$$

schreiben, wobei die Konstanten  $\psi_0, \psi_1, \dots$  sich aus

$$\Psi(z) = \sum_{i=0}^{\infty} \psi_i \cdot z^i = \frac{1}{\Phi(z)} \quad \text{für alle } z \in \mathbb{C} \quad (2.2.9)$$

ergeben und  $\sum_{i=0}^{\infty} |\psi_i| < \infty$  erfüllen.

$(X_t)$  heißt in dem Fall *kausal* bzgl.  $(W_t)$  und die  $\psi_0, \psi_1, \dots$  lassen sich rekursiv durch

$$\psi_j = \sum_{i=1}^p \phi_i \cdot \psi_{j-i} \quad j = 1, 2, \dots \quad (2.2.10)$$

bestimmen. Dabei sei  $\psi_0 = 1$  und  $\psi_{-j} = 0$  für  $j \geq 1$  gesetzt.

Bevor wir der Fragestellung nachgehen, wie wir für einen Datensatz  $x_1, \dots, x_n$  „gute“ Schätzwerte für  $\phi_1, \dots, \phi_p$  erhalten, betrachten wir einige Beispiele.

- Beispiel 2.2.4** (a) Für ein AR(1) Modell mit  $\phi = \phi_1$  ergibt sich  $\Phi(z) = 1 - \phi \cdot z$  mit Nullstelle  $1/\phi$ . Demnach ist für  $\phi = \pm 1$  (Random Walk) der Prozess nicht schwach stationär, für  $|\phi| \neq 1$  ist er schwach stationär und für  $|\phi| < 1$  sogar kausal mit  $\psi_j = \phi^j$ .
- (b) Das AR(2) Modell  $X_t = X_{t-1} - \frac{1}{4}X_{t-2} + W_t$  ist schwach stationär, denn hier ist  $\Phi(z) = 1 - z + \frac{1}{4}z^2 = \frac{1}{4} \cdot (z - 2)^2$ , also 2 die einzige Nullstelle von  $\Phi$ .
- (c) Das AR(2) Modell  $X_t = \frac{1}{2}X_{t-1} + \frac{1}{2}X_{t-2} + W_t$  ist *nicht* schwach stationär, denn wir erhalten  $\Phi(z) = -\frac{1}{2} \cdot (z - 1) \cdot (2 + z)$  mit den Nullstellen 1 und  $-2$ .

Ein numerischer Algorithmus zur Nullstellensuchen von Polynomen ist in R in die Funktion `polyroot(t)` implementiert, wobei  $t = (t[1], \dots, t[n])$  die Koeffizienten des Polynoms  $z \mapsto t[1] + t[2] \cdot z + \dots + t[n] \cdot z^{n-1}$  angeben. Im Beispiel (b) müssten wir demnach `polyroot(c(1, -1, 1/4))` aufrufen, für (c) wäre es `polyroot(c(1, -1/2, -1/2))`.

## 2.2.1 Schätzung von $\phi_1, \dots, \phi_p$

Haben wir die Ordnung  $p$  für eine Beobachtung  $x_1, \dots, x_n$  fixiert und möchten die Koeffizienten  $\phi_1, \dots, \phi_p$  schätzen, so gibt es verschiedene Ansätze, z. B. durch die Minimierung der Summe der Fehlerquadrate (klassische Regression), durch einen Maximum-Likelihood-Ansatz (unter der zusätzlichen Annahme, dass jedes  $W_t$  normalverteilt ist) oder durch die sogenannten *Yule-Walker Gleichungen*, auf die wir im Folgenden genauer eingehen werden.

Wir nehmen an,  $(X_t)$  ist ein AR(p)-Prozess,  $\Phi(\mathbf{B})X_t = W_t$ , der kausal bzgl.  $(W_t) \sim \text{WN}(0, \sigma^2)$  ist. Er besitzt also die Darstellung

$$X_t = \sum_{i=0}^{\infty} \psi_i \cdot W_{t-i} \quad (2.2.11)$$

mit  $(\psi_i)_{i \in \mathbb{N}_0}$  gemäß (2.2.10). Multiplizieren wir für  $j = 1, \dots, p$  beide Seiten von  $\Phi(\mathbf{B})X_t = W_t$  mit  $X_{t-j}$  und bilden den Erwartungswert, so gilt

$$\mathbb{E}(X_{t-j} \cdot \Phi(\mathbf{B})X_t) = \mathbb{E}(X_{t-j} \cdot W_t) = 0, \quad (2.2.12)$$

denn aufgrund von (2.2.11) sind  $W_t$  und  $X_{t-j}$  unkorreliert für  $j \geq 1$ . Andererseits gilt wegen  $\mathbb{E}(X_t) = 0$

$$\begin{aligned} \mathbb{E}(X_{t-j} \cdot \Phi(\mathbf{B})X_t) &= \mathbb{E}(X_{t-j} \cdot X_t) - \phi_1 \mathbb{E}(X_{t-j} \cdot X_{t-1}) - \dots - \phi_p \mathbb{E}(X_{t-j} \cdot X_{t-p}) \\ &= \text{Cov}(X_{t-j}, X_t) - \phi_1 \text{Cov}(X_{t-j}, X_{t-1}) - \dots - \phi_p \text{Cov}(X_{t-j}, X_{t-p}) \\ &= \gamma(j) - \phi_1 \gamma(1-j) - \dots - \phi_p \gamma(p-j). \end{aligned} \quad (2.2.13)$$

(2.2.12) und (2.2.13) zusammen ergeben

$$\gamma(j) = \phi_1 \gamma(1-j) + \dots + \phi_p \gamma(p-j) \quad \text{für alle } j = 1, \dots, p,$$

oder kürzer

$$\Gamma_p \cdot \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix} \quad (2.2.14)$$

wobei

$$\Gamma_p = (\gamma(i-j))_{i,j=1,\dots,p} = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{pmatrix} \quad (2.2.15)$$

die Auto-Kovarianzmatrix (bis zum Lag  $p$ ) sei. Die Gleichungen (2.2.14) nennen wir auch *Yule-Walker Gleichungen*.

Ersetzen wir  $\gamma(1), \dots, \gamma(p)$  und  $\Gamma_p$  in (2.2.14) durch ihre empirischen Gegenstücke (wir beachten die Analogie zum Momentenmethode Verfahren: dort werden die theoretischen Momente den empirischen Momenten gleichgesetzt), so erhalten wir

$$\hat{\Gamma}_p \cdot \begin{pmatrix} \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = \begin{pmatrix} \hat{\gamma}(1) \\ \vdots \\ \hat{\gamma}(p) \end{pmatrix} \quad (2.2.16)$$

Da  $\hat{\Gamma}_p$  invertierbar ist (sofern nicht gerade  $x_1 = \dots = x_n$  gilt), erhalten wir auf diese Art Schätzer für  $\phi_1, \dots, \phi_p$  durch

$$\begin{pmatrix} \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = (\hat{\Gamma}_p)^{-1} \cdot \begin{pmatrix} \hat{\gamma}(1) \\ \vdots \\ \hat{\gamma}(p) \end{pmatrix} = (\hat{R}_p)^{-1} \cdot \begin{pmatrix} \hat{\varrho}(1) \\ \vdots \\ \hat{\varrho}(p) \end{pmatrix} \quad (2.2.17)$$

Wollen wir in R den Datensatz  $x_1, \dots, x_n$  an ein AR(p)-Modell fitten, so benutzen wir den Befehl `ar(x, aic = FALSE, order.max = p, method = "yule-walker")`.

Andere mögliche Methoden sind "ols" (Minimierung der Summe der Fehlerquadrate) und "mle" (Maximum-Likelihood Verfahren).

Geben wir  $p$  nicht an und setzen `aic = TRUE`, so wählt R automatisch ein  $p$  mit gutem `aic` Wert (siehe dazu Abschnitt 3.3).

## 2.2.2 Schätzung von $\sigma^2$

Neben den Parameter  $\phi_1, \dots, \phi_p$  wollen wir noch die Varianz des White-Noise Prozesses  $(W_t)_{t \in \mathbb{Z}}$  schätzen. In der Situation eines kausalen AR(p)-Prozesses  $(X_t)$  multiplizieren wir  $\Phi(\mathbf{B})X_t = W_t$  dazu auf beiden Seiten mit  $X_t = \sum_{i=0}^{\infty} \psi_i \cdot W_{t-i}$  und bilden den Erwartungswert. Dies liefert zum einen

$$\mathbb{E}(X_t \cdot W_t) = \mathbb{E}\left(\sum_{i=0}^{\infty} \psi_i \cdot W_{t-i} \cdot W_t\right) = \mathbb{E}\left(\underbrace{\psi_0}_{=1} \cdot W_t \cdot W_t\right) = \text{Var}(W_t) = \sigma^2 \quad (2.2.18)$$



und zum anderen

$$\begin{aligned}
 \mathbb{E}(X_t \cdot W_t) &= \mathbb{E}(\Phi(\mathbf{B})X_t \cdot X_t) = \mathbb{E}(X_t^2 - \phi_1 \cdot X_{t-1} \cdot X_t - \dots - \phi_p \cdot X_{t-p} \cdot X_t) \\
 &= \text{Var}(X_t) - \phi_1 \cdot \gamma(1) - \dots - \phi_p \cdot \gamma(p) \\
 &= \gamma(0) - (\phi_1 \ \dots \ \phi_p) \cdot \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix}
 \end{aligned} \tag{2.2.19}$$

Zusammen erhalten wir

$$\sigma^2 = \gamma(0) - (\phi_1 \ \dots \ \phi_p) \cdot \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix}$$

Der Yule-Walker Schätzer  $\hat{\sigma}^2$  für  $\sigma^2$  ist demnach durch

$$\hat{\sigma}^2 = \hat{\gamma}(0) - (\hat{\phi}_1 \ \dots \ \hat{\phi}_p) \cdot \begin{pmatrix} \hat{\gamma}(1) \\ \vdots \\ \hat{\gamma}(p) \end{pmatrix} = \hat{\gamma}(0) \cdot \left( 1 - (\hat{\phi}_1 \ \dots \ \hat{\phi}_p) \cdot \begin{pmatrix} \hat{\varrho}(1) \\ \vdots \\ \hat{\varrho}(p) \end{pmatrix} \right) \tag{2.2.20}$$

gegeben und wird in R beim Aufruf der `ar` Funktion (`method = "yule-walker"`) mit berechnet.

### 2.2.3 Vorhersagen in AR(p) Modellen

In einem AR(p) Modell  $\Phi(\mathbf{B})X_t = W_t$  ist eine Vorhersage für den Zeitpunkt  $n + 1$  durch

$$\hat{x}_{n+1} = \phi_1 x_n + \phi_2 x_{n-1} + \dots + \phi_p x_{n-p+1} \tag{2.2.21}$$

gegeben.

In R erhalten wir diese wieder mit der Funktion `predict(ar(x))`.

Achtung: Die Funktion `ar` arbeitet immer mit den zentrierten Daten  $x_1 - \bar{x}, \dots, x_n - \bar{x}$ , d. h. die Vorhersage ist durch

$$\hat{x}_{n+1} = \bar{x} + \hat{\phi}_1 \cdot (x_n - \bar{x}) + \dots + \hat{\phi}_p \cdot (x_{n-p+1} - \bar{x}) \tag{2.2.22}$$

gegeben.

## 2.3 Tests für die Zeitreihenanalyse

Im folgenden werden drei Tests vorgestellt, mit denen sich signifikant nachweisen lässt, dass eine Beobachtung  $x_1, \dots, x_n$  nicht von einer i. i. d. Folge stammt.

Ausgangspunkt der beiden ersten Tests ist der folgende Satz, der sich aus dem schwachen Gesetz der großen Zahlen folgern lässt.

**Satz 2.3.1** Ist  $(W_t) \sim \text{IID}(\mu, \sigma^2)$ , und  $x_1, \dots, x_n$  eine Beobachtung von  $(W_t)$  mit eACF  $\hat{\varrho}_n$ , so gilt für alle  $h \geq 1$

$$\sqrt{n} \cdot (\hat{\varrho}_n(1), \dots, \hat{\varrho}_n(h)) \xrightarrow{d} \mathcal{N}(0, 1)^h \quad \text{für } n \rightarrow \infty. \quad (2.3.1)$$

Der Satz besagt, dass für große  $n$  die eACF  $(\hat{\varrho}_n(h))_{h \geq 1}$  approximativ eine unabhängige Folge von  $\mathcal{N}(0, 1/n)$  Beobachtungen ist. Es fallen demnach approximativ 95% der Werte von  $\hat{\varrho}_n(h)$  in das Intervall  $\pm 1.96/\sqrt{n}$  (denn `qnorm(.975)`, das  $1 - 0.975 = 2.5\%$  Quantil der Standardnormalverteilung, ist in etwa gleich 1.96). Dieses Intervall wird in R im Plot der `acf` Funktion blau gestrichelt dargestellt.

### 2.3.1 Box-Pierce Test

Statt sich die ganze Folge  $(\hat{\varrho}(h))_{h \geq 1}$  anzuschauen, können wir auch die Statistik

$$Q = n \cdot \sum_{h=1}^L \hat{\varrho}(h)^2 \quad (2.3.2)$$

betrachten, die für große  $n$  verteilt ist wie  $\sum_{h=1}^L Y_h^2$ , mit  $(Y_i)$  i. i. d. und standardnormalverteilt, d. h. für große  $n$  ist  $Q$  approximativ  $\chi_L^2$  verteilt.

Die Hypothese

$$H : \hat{\varrho}(1) = \dots = \hat{\varrho}(L) = 0 \quad (2.3.3)$$

wird zum Niveau  $\alpha$  verworfen, wenn  $Q > \chi_{L, 1-\alpha}^2$  (`qchisq(1 - alpha, L)`) ist.

Für  $n < 100$  arbeitet dieser Test nicht zufriedenstellend, ebenso ist die Wahl von  $L$  schwierig. Als Faustregel wählen wir  $L$  etwa als  $2 \cdot \sqrt{n}$ , es schadet aber nicht, wenn wir alle  $p$ -Werte bis zu diesem  $L$  berechnen lassen und anhand aller  $p$ -Werte eine Entscheidung fällen.

In R erhalten wir diesen Test durch `Box.test(x, lag)`, wobei `x` wie gewohnt die Zeitreihe ist, und `lag` dem  $L$  von oben entspricht.

### 2.3.2 Ljung-Box Test

Für kleine Beobachtungslängen  $n$  greifen wir statt des Box-Pierce Tests auf den Ljung-Box Test zurück, bei dem statt  $Q$  die Statistik

$$Q' = n \cdot (n + 2) \cdot \sum_{h=1}^L \frac{1}{n - h} \cdot \hat{\varrho}(h)^2 \quad (2.3.4)$$

berechnet wird, die ebenfalls für große  $n$  approximativ  $\chi_L^2$  verteilt ist.

In R erhalten wir diesen Test, indem wir beim `Box.test` das optionale Argument `type = "Ljung-Box"` wählen.

### 2.3.3 Turning Point Test

Wir sagen, die Beobachtung  $x_1, \dots, x_n$  besitzt einen *Wendepunkt* in  $i$  ( $1 < i < n$ ), wenn  $x_{i-1} < x_i$  und  $x_i > x_{i+1}$  gilt, oder wenn  $x_{i-1} > x_i$  und  $x_i < x_{i+1}$  gilt. Es sei  $T_n$  die Anzahl der Wendepunkte von  $x_1, \dots, x_n$ .

**Satz 2.3.2** Ist  $X_1, \dots, X_n$  eine i. i. d. Folge und die Verteilung von  $X_1$  stetig, so gilt

$$\begin{aligned} \mathbb{E}T_n &= (n-2) \cdot \frac{2}{3} \\ \text{Var}(T_n) &= \frac{16n-29}{90} \\ T_n^* = \frac{T_n - \mathbb{E}T_n}{\sqrt{\text{Var}(T_n)}} &\xrightarrow{d} \mathcal{N}(0, 1) \end{aligned}$$

d. h. für große  $n$  ist  $T_n$  approximativ  $\mathcal{N}((n-2) \cdot \frac{2}{3}, \frac{16n-29}{90})$  verteilt.

Ein großer Wert von  $T_n - \mathbb{E}T_n$  deutet darauf hin, dass die Zeitreihe zu viel hin und her springt, um von einer i. i. d. Folge zu kommen. Ein kleiner Wert von  $T_n - \mathbb{E}T_n$  deutet auf eine Korrelation zwischen benachbarten Beobachtungen hin. Satz 2.3.2 liefert einen plausiblen Test zum Niveau  $\alpha$  für die Hypothese

$$H : x_1, \dots, x_n \text{ stammt von einer i. i. d. Beobachtung.} \quad (2.3.5)$$

Wir verwerfen  $H$ , wenn  $|T_n^*| > \Phi_{1-\alpha/2}$  ist.  $\Phi_{1-\alpha/2}$  ist dabei das  $1 - \alpha/2$ -Quantil der Standardnormalverteilung ( $= \text{qnorm}(1 - \alpha/2)$ ).

## 2.4 R-Befehle

`acf(x)` Plottet die empirische Autokorrelationsfunktion (ACF) der Zeitreihe (oder des Vektors)  $x$ . Auf die Zahlenwerte können wir mit `acf(x)$acf` zugreifen.

`diff(x)` Bildet aus  $x = (x[1], \dots, x[n])$  den Differenzenvektor  $(x[2] - x[1], \dots, x[n] - x[n-1])$

`polyroot(t)` Berechnet numerisch die Nullstellen des Polynoms  $z \mapsto t[1] + t[2] \cdot z + \dots + t[n] \cdot z^{n-1}$

`ar(x)` Passt an  $x$  ein  $\text{AR}(p)$  Modell an. Optionale Argumente:

Setzen wir `aic = FALSE`, so wird `p = order.max` als Ordnung gewählt, im Fall `aic = TRUE` wählt R das  $p \leq \text{order.max}$ , so dass das gefittete Modell minimalen `aic` Wert besitzt.

Mit `method` geben wir an, wie die Koeffizienten  $\phi_1, \dots, \phi_p$  geschätzt werden sollen. Standard ist `"yule-walker"`, möglich sind noch `"mle"` (Maximum-Likelihood), `"ols"` (Minimierung der Fehlerquadrate) und `"burg"` (Burg's Algorithmus).

`Box.test(x, lag)` Führt einen Box-Pierce Test für  $x$  durch. Die Wahl von `lag` ist schwierig: Am Besten die  $p$ -Werte für `lag = 1, \dots, 2 \cdot \sqrt{\text{length}(x)}` berechnen.

Für einen Ljung-Box Test muss `type = "Ljung-Box"` gewählt werden.

`qnorm(1-a)` Das  $1-a$ -Quantil einer Standardnormalverteilung.

`qchisq(1-a,L)` Das  $1-a$ -Quantil einer  $\chi^2$ -Verteilung mit  $L$  Freiheitsgraden.

# 3 ARMA-Modelle

In diesem Kapitel lernen wir Methoden kennen, ein geeignetes Modell für eine vorliegende Zeitreihe zu bestimmen; sowie zwei weitere Klassen von Modellen.

## 3.1 Partielle Autokorrelationsfunktion

### 3.1.1 Optimale lineare Prognose in AR-Modellen

Gegeben ein kausales AR(p)-Modell

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + W_t$$

mit  $(W_t) \sim WN(0, \sigma^2)$ , können wir leicht einen gleichmäßig besten linearen erwartungstreuen Schätzer für  $X_t$ , gegeben  $(X_{t-k})_{k \geq 1}$ , bestimmen. Dabei nennen wir den linearen Schätzer  $\hat{X}_t$  einen gleichmäßig besten, wenn er

$$\mathbb{E} \left( \hat{X}_t - X_t \right)^2$$

unter allen linearen Schätzern für alle denkbaren  $\sigma^2$  minimiert.

Es ist einfach

$$\hat{X}_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p},$$

denn aufgrund der Kausalität sind  $W_t$  und  $(X_{t-k})_{k \geq 1}$  unkorreliert, also gilt für jeden weiteren linearen erwartungstreuen Schätzer  $g = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + \dots$

$$\begin{aligned} \mathbb{E} (g - X_t)^2 &= \mathbb{E} \left( g - \hat{X}_t + \hat{X}_t - X_t \right)^2 \\ &= \text{Var} \left( g - \hat{X}_t \right) + 2\text{Cov} \left( g - \hat{X}_t, \hat{X}_t - X_t \right) + \text{Var} \left( \hat{X}_t - X_t \right) \\ &= \text{Var} \left( g - \hat{X}_t \right) + 2\text{Cov} \left( \underbrace{g - \hat{X}_t}_{\text{unkorreliert mit } W_t}, W_t \right) + \text{Var} (W_t) \\ &= \text{Var} \left( g - \hat{X}_t \right) + 0 + \sigma^2 \\ &\geq \mathbb{E} \left( \hat{X}_t - X_t \right)^2 \quad (= \sigma^2) \end{aligned}$$

Die naheliegende Prognose

$$\hat{x}_{t+1|t} = \phi_1 x_t + \dots + \phi_p x_{t-p}$$

für beobachtete Zeitreihen, hinter denen wir ein AR(p)-Modell vermuten, ist also auch die theoretisch beste (zumindest unter den linearen).

Im Folgenden benötigen wir eine einfache Verallgemeinerung, nämlich einen gleichmäßig besten linearen erwartungstreuen Schätzer für  $X_t$ , gegeben nicht die gesamte Prozessvergangenheit,

sondern nur eine Auswahl von Zufallsgrößen, bspw.  $(X_{t-1}, X_{t-2})$ . Wir suchen also eine Linearkombination  $a_0 + a_1 X_{t-1} + a_2 X_{t-2}$  mit der Eigenschaft, dass

$$\mathbb{E} \left[ (X_t - a_0 + a_1 X_{t-1} + a_2 X_{t-2})^2 \right]$$

minimal wird. Wir bezeichnen diese mit

$$\Pr[X_t \mid X_{t-1}, X_{t-2}],$$

wobei die Abkürzung  $\Pr$  dadurch gerechtfertigt wird, dass wir hier gerade die Projektion von  $X_t$  auf den von  $1, X_{t-1}, X_{t-2}$  erzeugten Unterraum des Hilbertraums  $\mathcal{L}^2(\Omega, \mathfrak{A}, \mathbb{P})$  berechnen. Insbesondere hat sie die Eigenschaft, dass

$$\text{Cov}(X_t - \Pr[X_t \mid X_{t-1}, X_{t-2}], X_{t-i}) = 0$$

für  $i \in \{1, 2\}$ , sie enthält also alle Korrelationen zwischen  $X_t$  und  $X_{t-1}, X_{t-2}$ . Dies beispielhaft zur Einführung, die allgemeinen Schreibweisen und Beziehungen gelten natürlich analog.

Beachte abschließend den Unterschied zum bedingten Erwartungswert! Dieser ist die Projektion auf den Unterraum  $\mathcal{L}^2(\Omega, \sigma(X_{t-1}, \dots, X_{t-h+1}), \mathbb{P})$ , welcher wesentlich größer ist.

### 3.1.2 Die partielle Autokorrelationsfunktion

Betrachten wir die ACF eines AR(1)-Prozesses: Diese fällt exponentiell gegen Null ab, ist aber insbesondere auch für Lags größer 1 positiv, obwohl wir aus der Struktur des Prozesses wissen, dass  $X_t$  und  $X_{t-2}$  nicht direkt voneinander abhängen. Die Korrelation ergibt sich vielmehr, weil  $X_t$  mit  $X_{t-1}$ , und  $X_{t-1}$  wiederum mit  $X_{t-2}$  korreliert ist. Die Autokorrelationsfunktion berücksichtigt also auch indirekte Korrelationen zwischen  $X_t$  und  $X_{t-h}$ .

Zudem lässt sich anhand der ACF nicht einschätzen, wie der Parameter  $p$  zu wählen ist;  $p \in \{1, 2\}$  liefert eine qualitativ gleiche ACF.

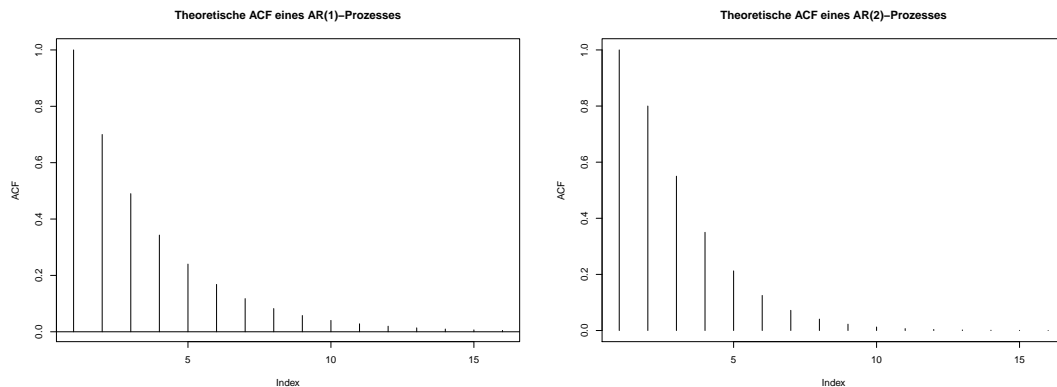


Abbildung 3.1: Theoretische ACF: AR(1)-Prozesses,  $\phi = 0.7$ , sowie AR(2)-Prozess,  $\phi_1 = 1$  und  $\phi_2 = -0.25$

Deshalb suchen wir nun nach einer Funktion, welche nur direkte Korrelationen berücksichtigt, und mehr Aussagekraft für die Wahl eines geeigneten  $p$  besitzt; dies wird die *partielle Autokorrelationsfunktion* leisten.

**Definition 3.1.1** (Partielle Autokorrelationsfunktion)

Sei  $(X_t)_{t \in T}$  ein schwach stationärer Prozess. Die partielle Autokorrelationsfunktion  $\varphi(h)$  ist für  $h \geq 1$  definiert durch

$$\begin{aligned}\varphi(1) &= \frac{\text{Cov}(X_2, X_1)}{(\text{Var}(X_2)\text{Var}(X_1))^{\frac{1}{2}}} = \varrho(1) \\ \varphi(h) &= \frac{\text{Cov}(X_{h+1} - \Pr[X_{h+1} | X_2, \dots, X_h], X_1)}{(\text{Var}(X_{h+1} - \Pr[X_{h+1} | X_2, \dots, X_h])\text{Var}(X_1))^{\frac{1}{2}}}, \quad h > 1.\end{aligned}$$

Man betrachtet also nur die Korrelationen zwischen  $X_1$  und  $X_{h+1}$ , die verbleiben, wenn alle Korrelationen mit den dazwischenliegenden Variablen entfernt wurden, diese sind ja gerade in  $\Pr[X_{h+1} | X_2, \dots, X_h]$  enthalten.

**Beispiel 3.1.2** (PACF eines AR(1)-Prozesses)

Sei  $(X_t)_{t \in \mathbb{Z}}$  ein kausaler AR(1)-Prozess mit  $X_t = \phi X_{t-1} + W_t$ . Dann ist also für alle  $h \geq 2$   $\Pr(X_{h+1} | X_2, \dots, X_h) = \phi X_h$ , und somit

$$\begin{aligned}\varphi(1) &= \text{corr}(X_2, X_1) = \varrho(1) = \phi \\ \varphi(h) &= \text{corr}(X_{h+1} - \Pr(X_h | X_2, \dots, X_{h-1}), X_1) = \text{corr}(W_t, X_1) = 0\end{aligned}$$

für alle  $h > 1$ .

**3.1.3 Schätzung der PACF**

Die Vermutung, dass für beliebige AR(p)-Prozesse die Koeffizienten gerade den partiellen Autokorrelationen entsprechen, trifft nicht zu. Die partiellen Autokorrelationen sind betraglich durch 1 beschränkt, die Koeffizienten nicht. Es gilt aber folgender Satz:

**Satz 3.1.3** Für einen AR(p)-Prozess gilt  $\varphi(p) = \phi(p)$ , und  $\varphi(h) = 0$  für  $h > p$ .

Dies zeigt eine analoge Rechnung zu oben.

Dieser Satz liefert auch die Idee für eine alternative Definition der PACF (die obiger in der Tat äquivalent ist, siehe [4], Satz 4.1.2.6):

**Definition 3.1.4** Die partielle Autokorrelation  $\varphi(h)$  ist der Koeffizient  $\phi(h)$  von  $X_{t-h}$  in  $\Pr[X_{h+1} | X_2, \dots, X_h]$ .

Empirisch bedeutet dies, dass wir die ePACF berechnen können, indem wir sukzessive AR(h)-Modelle,  $h \geq 11$  an die Daten fitten. Mittels der Yule-Walker-Gleichungen berechnet sich die empirische partielle Autokorrelationsfunktion dann als

$$\hat{\varphi}(h) = \left[ \hat{\Gamma}_h^{-1} \hat{\gamma}_h \right]_h.$$

**3.2 MA(q)-Modelle**

Wir lernen nun eine zweite wichtige Prozessklasse kennen, die sich in gewisser Weise invers zur Klasse der AR-Prozesse verhält. Es sind die sogenannten Moving-Average-Prozesse.

**Definition 3.2.1** (Moving Average-Prozess)

$(X_t)_{t \in \mathbb{Z}}$  heißt *Moving Average-Prozess der Ordnung  $q$* , kurz *MA( $q$ )-Prozess*, falls jedes  $X_t$  die Darstellung

$$X_t = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$$

für  $(W_t)$   $WN(0, \sigma^2)$  und Konstanten  $\theta_1, \dots, \theta_q$  mit  $\theta_q \neq 0$  besitzt.

Der Name erklärt sich dadurch, dass der Prozess als gleitender Durchschnitt der zurückliegenden White Noise gebildet wird.

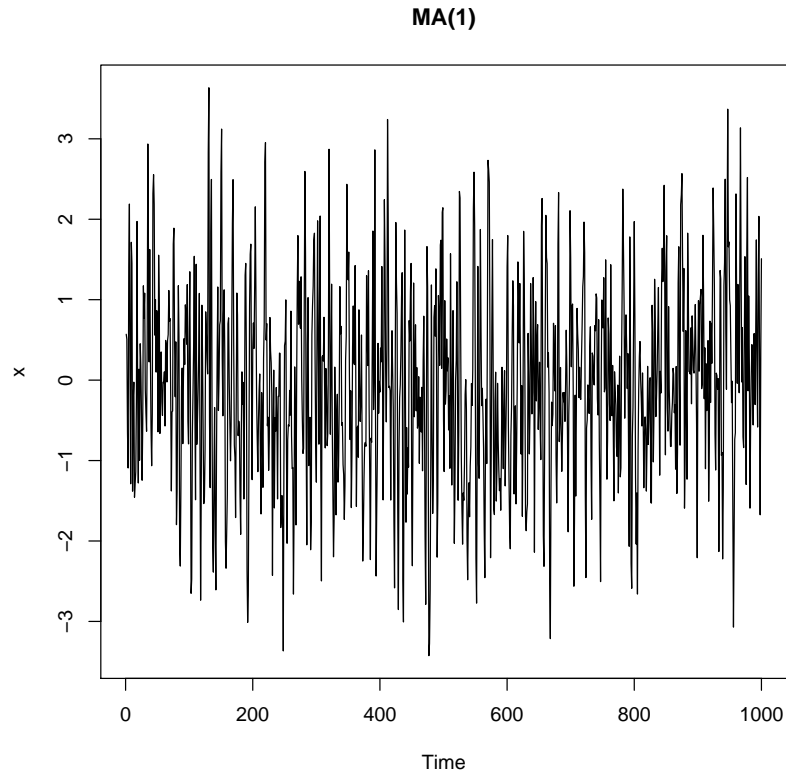


Abbildung 3.2: Reihe mit 1000 Realisierungen eines MA(1)-Prozesses

Ein MA( $q$ )-Prozess ist stets schwach stationär (ohne Voraussetzungen an  $(\theta_i)_{1 \leq i \leq q!}$ ), da er eine Linearkombination unkorrelierter schwach stationärer Prozesse ist.

Die ACF eines MA(1)-Prozess lässt sich leicht berechnen, es gilt

$$\mathbb{E}(X_t) = 0$$

$$\gamma(0) = \text{Var}(W_1 + \theta W_0) = \text{Var}(W_1) + \theta^2 \text{Var}(W_0) + 2\text{Cov}(W_1, \theta W_0) = (1 + \theta^2)\sigma^2$$

$$\gamma(1) = \text{Cov}(W_2 + \theta W_1, W_1 + \theta W_0) = \theta \text{Cov}(W_1, W_1) = \theta\sigma^2$$

$$\gamma(h) = \text{Cov}(W_{h+1} + \theta W_h, W_1 + \theta W_0) = 0$$

für alle  $h \geq 2$ .

Damit folgt für die ACF

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta}{1+\theta^2} & h = 1 \\ 0 & h > 1 \end{cases}$$

Allgemein gilt:

**Satz 3.2.2** Die ACF eines MA(q)-Prozesses ist gegeben durch

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\sum_{i=0}^q \theta_i^2} & h = 1, \dots, q \\ 0 & h > q \end{cases} \quad (3.2.1)$$

wobei  $\theta_0 = 1$ .

Der Beweis bleibt dem Leser zur Übung überlassen.

Werfen wir nun einen Blick auf die eACF und ePACF der oben gezeigten Realisierung eines MA(1)-Prozesses. Daran mag deutlich werden, inwiefern der MA(1)-Prozess sich invers zum AR(1)-Prozess verhält.

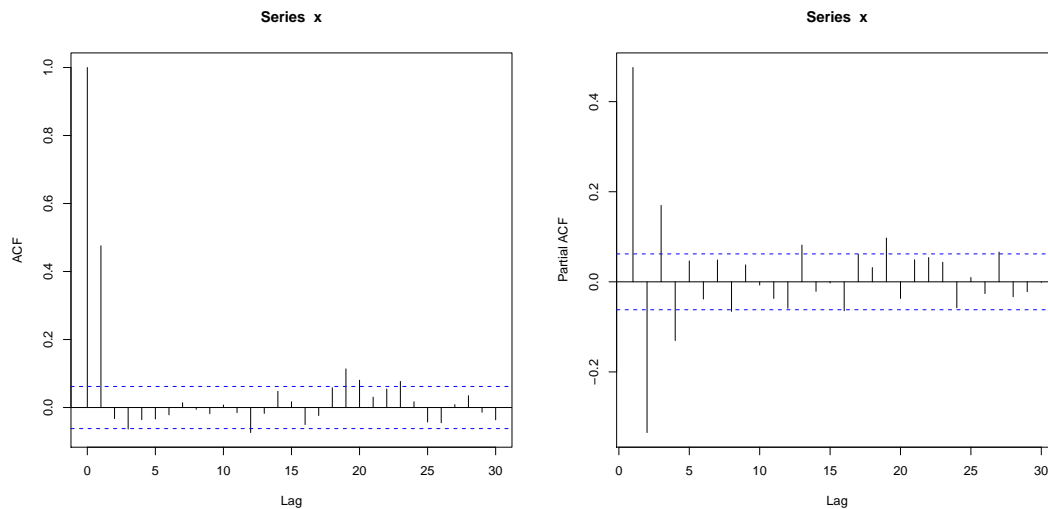


Abbildung 3.3: Empirische ACF und PACF der obigen Reihe

Das Verhalten von ACF und PACF ist also bei diesen Prozessen gerade vertauscht. Steigen wir nun etwas tiefer in die wechselseitigen Beziehungen ein.

### 3.2.1 Invertierbare MA-Prozesse

Obige Definition lässt sich auf  $q = \infty$  erweitern, falls wir zusätzlich fordern, dass  $\sum_{i=1}^{\infty} |\theta_i| < \infty$ .

Erinnern wir uns an kausale AR-Prozesse, so stellen wir fest, dass die Kausalitätseigenschaft gerade besagt, dass der Prozess als MA( $\infty$ )-Prozess dargestellt werden kann (insbesondere



konvergierte die Summe der Koeffizienten). Dies erklärt das Nichtverschwinden der ACF eines kausalen AR-Prozesses.

Eine analoge Aussage gilt auch für MA-Prozesse, die entsprechende Eigenschaft heißt hier Invertierbarkeit. Dazu schreiben wir mit Hilfe des Back-Shift-Operators die Differenzengleichung eines MA(q)-Prozesses als

$$X_t = \Theta(\mathbf{B})W_t.$$

**Definition 3.2.3** (Invertierbare MA(q)-Prozesse)

Ein MA(q)-Prozess  $(X_t)$  heißt *invertierbar*, falls  $(X_t)$  eine Darstellung als AR( $\infty$ )-Prozess besitzt, d.h.  $\sum_{i=1}^{\infty} |\phi_i| < \infty$  und

$$X_t = -\sum_{i=1}^{\infty} \phi_i X_{t-i} + W_t.$$

**Satz 3.2.4** Ein MA(q)-Prozess ist invertierbar genau dann, wenn  $\Theta(z) \neq 0$  für alle  $z \in \mathbb{C}$  mit  $|z| \leq 1$ . Dann berechnen sich die Koeffizienten der AR( $\infty$ )-Darstellung aus

$$\frac{1}{\Theta(z)} = 1 + \sum_{i=1}^{\infty} \phi_i z^i.$$

Daraus folgt, dass die PACF eines MA(q)-Prozesses überall von Null verschiedene Einträge besitzt, da sie der PACF eines AR( $\infty$ )-Prozesses entspricht.

Auf eine exakte Berechnung der PACF, die insbesondere exponentielles Abfallen nachweisen würde, verzichten wir hier. Die Berechnung erfolgt mittels der Yule-Walker-Gleichungen.

**Satz 3.2.5** Die PACF des MA(1)-Prozesses  $X_t = W_t + \theta W_{t-1}$  ist gegeben durch

$$\varphi(h) = -\frac{(-\theta)^h}{1 + \theta^2 + \dots + \theta^{2h}} = -\frac{(-\theta)^h(1 - \theta^2)}{1 - \theta^{2(h+1)}}$$

für alle  $h \geq 1$ , erfüllt also  $\varphi(h) = c(-\theta)^h$  für eine geeignete Konstante  $c$ .

### 3.2.2 Schätzen der Parameter $\theta_1, \dots, \theta_n$

Gab es bei einem AR-Modell zahlreiche Ansätze, die Parameter zu schätzen, so ist dies bei einem MA(q)-Modell weitaus schwieriger, weshalb zumeist auf AR-Modelle zurückgegriffen wird, zumal invertierbare MA-Modelle eine AR( $\infty$ )-Darstellung besitzen. MA-Modelle werden hauptsächlich aus Gründen der Parametereffizienz betrachtet, dazu kommen wir jedoch später.

Wie aber werden nun die Parameter geschätzt? Eine Möglichkeit sind Maximum-Likelihood-Schätzungen, wobei hierfür angenommen wird, dass die White Noise unabhängig identisch normal verteilt ist. Diese Einschränkung sollte keine Bauchschmerzen bereiten, da wir ja nur ein Modell suchen, das die Daten beschreiben kann, nicht aber die wirklichen Gesetzmäßigkeiten hinter den Daten ergründen können.

Ein anderer Ansatz ist die Minimierung der Summe der quadrierten 1-Schritt-Prognosefehler, konkret ist also die Funktion

$$S(\theta_1, \dots, \theta_q) = \sum_{t=1}^n (x_t - (\theta_1 \hat{w}_{t-1} + \dots + \theta_q \hat{w}_{t-q}))^2$$

zu minimieren. Das Problem dabei ist, dass die  $\hat{w}_t$  nicht beobachtet wurden, sondern gerade die 1-Schritt-Prognose-Fehler sind, und rekursiv berechnet werden müssen. Dafür wiederum müssen Startwerte für  $\theta_1, \dots, \theta_q$  vorgegeben werden; sowie Startwerte  $w_{-q} = \dots = w_0 = 0$ . Damit werden zunächst alle  $\hat{w}_t$  berechnet, und die so entstehende Summe nun bzgl.  $\theta_1, \dots, \theta_q$  minimiert. Daraus können dann neue Schätzwerte für  $\hat{w}_t$  bestimmt werden, und die so entstehende Summe minimiert werden. Das Verfahren wird iteriert.

Ein alternativer, von Durbin vorgeschlagener Ansatz ist, zunächst ein AR(p) Modell hoher Ordnung anzupassen; darin die empirischen Fehler  $\hat{w}_t$  (die Residuen nach Modellanpassung) zu bestimmen, und dann einmalig  $S$  zu minimieren.

In R berechnet `arima(Zeitreihe, order=(0,0,q), include.mean=T)$coef` die Koeffizienten eines MA(q)-Modells mittels Maximum-Likelihood.

### 3.2.3 Optimale Prognose in MA(q)-Modellen

Eine gleichmäßig bester linearer erwartungstreuer Schätzer für  $X_t$ , gegeben  $(W_{t-k})_{k \geq 1}$ , ist

$$\hat{X}_t = \theta_1 W_{t-1} + \dots + \theta_q W_{t-1}.$$

Weitergehende Prognosen erhält man im Gegensatz zu AR(p)-Modellen nicht durch Iterieren, sondern es ist

$$\begin{aligned} \hat{X}_{t+j} &= \mathbb{E}(W_{t+j} + \theta_1 W_{t+j-1} + \dots + \theta_j W_t) + \theta_{j+1} W_{t-1} + \dots + \theta_q W_{t+j-q} \\ &= 0 + \theta_{j+1} W_{t-1} + \dots + \theta_q W_{t+j-q}, \end{aligned}$$

insbesondere also  $\hat{X}_{t+j} = 0$  für  $j \geq q$ .

Dies beweist man ähnlich wie für AR(p)-Modelle.

Alternativ kann für invertierbare Prozesse die AR( $\infty$ )-Darstellung gewählt werden, und dort wie üblich eine Vorhersage berechnet werden.

## 3.3 Modellwahl: Akaike Information Criterion

Ein allgemeines Prinzip in der Modellierung ist die Sparsamkeit: Je mehr freie Parameter ein Modell besitzt, umso genauer kann es zwar an die Daten angepasst werden, verliert damit aber zugleich an Vorhersagekraft (jede weitere Beobachtung würde eine große Zahl von Parametern ändern), und ist in diesem Sinne nicht robust.

Man denke an die Temperaturkurve im Verlauf eines Jahres. Sie wird bis zum Sommer ansteigen, danach abfallen. Dieser Sachverhalt wird durch eine periodische, stückweise lineare Funktion (Dreiecksfunktion), oder eine Sinuskurve ausreichend beschrieben. Nimmt man hingegen ein Polynom vom Grad 364, so kann dieses perfekt an die Daten angepasst werden, es stellt aber keinen *typischen* Verlauf dar, sondern gibt stattdessen bspw. Höchsttemperaturen Mitte April, sowie einen kalten September an, was in einem beliebigen Jahr nur mit sehr geringer Wahrscheinlichkeit zutrifft; der grobe Verlauf Sommer-Winter wird jedoch in jedem Jahr wieder auftauchen. Sprich, die Dreiecksfunktion spiegelt den Temperaturverlauf in *jedem* Jahr wieder.

Dies ist die Grundidee des *Akaike Information Criterion (AIC)*: Ein Modell für Datensatz  $A$  ist umso besser, je besser es auch als Modell für Datensatz  $B$  taugt, wobei  $B$  vom gleichen datengenerierenden Prozess (z.B. einem Thermometer am FMO) stammt wie  $A$ . Formal lässt sich dies mittels der (log-) Likelihood-Funktion ausdrücken:  $l_M(B|\hat{\theta}(A))$  bezeichne die (logarithmierte) Wahrscheinlichkeitsdichte des Datensatzes  $B$  in einem festen Modell  $M$ , dessen Parameter  $\hat{\theta}$  anhand von Datensatz  $A$  gewählt wurden.

Unter der Annahme, dass tatsächlich das Modell  $M$  den Daten generierenden Prozess beschreibt (mit tatsächlichem Parameter  $\theta_0$ ), können wir der Modellierung mittels  $M$  und  $\hat{\theta}(\cdot)$  die Zahl

$$\mathbb{E}_{M,\theta_0} \left[ l_M(B|\hat{\theta}(A)) \right]$$

zuordnen, wobei  $B$  und  $A$  zwei unabhängig identische, anhand  $M$  und  $\theta_0$  generierte Datensätze seien.

Diese Zahlen sind also ein Gütekriterium für Modelle: jedem Modell kann eine Zahl zugeordnet werden, und wir entscheiden uns für das Modell mit der höchsten Zahl.

Allerdings ist obige Formel keine durchführbare Rechenvorschrift; mit wachsendem Stichprobenumfang gilt jedoch, dass

$$-\frac{1}{2}AIC \approx \mathbb{E}_{M,\theta_0} \left[ l_M(B|\hat{\theta}(A)) \right],$$

das AIC muss also *minimiert* werden. Zur Herleitung dieser Beziehung siehe [4, Bemerkung 6.3.3.9]. Das AIC selbst schließlich ist, bei gegebenem Datensatz  $A$ , definiert als

$$AIC = -2 \cdot l_M(A|\hat{\theta}(A)) + 2 \cdot \text{Anzahl Modellparameter.}$$

Beachte dabei, dass  $\sigma^2$  ebenfalls ein Modellparameter ist. Die Vorzeichen und -faktoren sind historisch bedingt; zudem berechnet R das AIC wie zuletzt angegeben. Das AIC eines gefitteten Modells kann in R meist durch Anhängen von `$AIC` an den Modellaufruf, oder durch `AIC(Model)` ausgegeben werden.

In der Methode `ar` geschieht die Modellwahl standardmäßig anhand des AIC, dies erklärt, warum wir zunächst `aic=F` gesetzt hatten.

### 3.4 ARMA-Modelle

Wie wir gesehen haben, können wir sowohl AR(p)-Prozesse als auch (invertierbare) MA-Prozesse einfach als AR-Modelle mit hoher Ordnung behandeln, jedoch kann dies schnell zu einer hohen Parameterzahl führen, welche im AIC bestraft wird. Einen Ausweg weist eine Kombination aus beiden Prozessen, die sogenannten *ARMA(p,q)-Prozesse*.

**Definition 3.4.1** (ARMA(p,q)-Prozess)

$(X_t)$  heißt ARMA(p,q)-Prozess, falls  $(X_t)$  stationär ist, und es Polynome  $\Phi$  und  $\Theta$  vom Grad  $p$  bzw.  $q$  gibt, sodass alle  $X_t$  der Darstellung

$$\Phi(\mathbf{B})X_t = \Theta(\mathbf{B})W_t$$

mit  $(W_t) \sim WN(0, \sigma^2)$  genügen. Der Prozess heißt *kausal* bzw. *invertierbar* (bzgl.  $W_t$ ), wenn die Polynomfunktion  $\Phi(z)$  bzw.  $\Theta(z)$  für alle  $z \in \mathbb{C}$  mit  $|z| \leq 1$  invertierbar ist.

Dabei ist  $(X_t)$  stationär genau dann, wenn der AR-Teil stationär ist, wenn also  $\Phi(z) \neq 0$  für alle  $z \in \mathbb{C}$  mit  $|z| = 1$ . Ist  $(X_t)$  invertierbar, so existiert eine Darstellung als AR( $\infty$ )-Prozess, mit AR-Polynom  $\tilde{\Phi}$ , gegeben durch

$$\tilde{\Phi}(z) = \frac{\Phi(z)}{\theta(z)}.$$

Ist  $(X_t)$  kausal, so existiert eine Darstellung als MA( $\infty$ )-Prozess, mit MA-Polynom  $\Psi$ , gegeben durch

$$\Psi(z) = \frac{\Phi(z)}{\theta(z)}.$$

Ein ARMA(p,q)-Modell benötigt bei gleichem AIC zumeist weniger Parameter als ein reines AR- oder MA-Modell; bedenke, dass es zugleich einem AR( $\infty$ ) wie einem MA( $\infty$ )-Modell entsprechen kann (s.u.). Zudem kann eine gemeinsame Nullstelle von  $\Phi$  und  $\Theta$  gekürzt werden; die Definitionsgleichung ist in einem solchen Fall redundant.

### 3.4.1 ACVF eines kausalen ARMA(1,1)-Prozesses

Sei  $X_t$  gegeben durch

$$(1 - \phi\mathbf{B})X_t = (1 + \theta\mathbf{B})W_t.$$

Da der Prozess kausal ist, können wir schreiben

$$\begin{aligned} X_t &= (1 - \phi\mathbf{B})^{-1}(1 + \theta\mathbf{B})W_t \\ &= \left( \sum_{i=0}^{\infty} \phi^i \mathbf{B}^i \right) (1 + \theta\mathbf{B})W_t \\ &= \left( 1 + \sum_{i=0}^{\infty} \phi^{i+1} \mathbf{B}^{i+1} + \sum_{i=0}^{\infty} \phi^i \theta \mathbf{B}^{i+1} \right) W_t \\ &= W_t + (\phi + \theta) \sum_{i=1}^{\infty} \phi^{i-1} W_{t-i}, \end{aligned}$$

und erhalten somit

$$\gamma(0) = \text{Var}(X_0) = \sigma^2 + \sigma^2(\phi + \theta)^2(1 - \phi^2)^{-1}$$

sowie

$$\text{Cov}(X_0, X_h) = (\phi + \theta)\phi^{h-1}\sigma^2 + (\phi + \theta)^2\sigma^2\phi^h(1 - \phi^2)^{-1}.$$

Die ACF berechnet sich damit zu

$$\varrho(h) = \frac{\phi^{h-1}(\phi + \theta)(1 + \phi\theta)}{1 + \phi\theta + \theta^2},$$

insbesondere gilt  $\varrho(h) = \phi \cdot \varrho(h - 1)$ . Dies entspricht gerade dem von einem AR(1)-Prozess bekannten Verhalten der ACF für  $h > 1$ .

### 3.4.2 Box-Jenkins-Methode der Modellwahl

Allerdings wird die PACF eines ARMA(1,1)-Prozesses auch für Lags größer 1 nicht verschwinden, im Gegensatz zum AR(1)-Prozess. Eine Zusammenstellung unseres Wissens über ACF und PACF führt uns zur sogenannten *Box-Jenkins-Methode* der Modellwahl.

|           | ACF $\rho(h)$                             | PACF $\varphi(h)$                            |
|-----------|-------------------------------------------|----------------------------------------------|
| AR(p)     | exponentielle Hüllkurve                   | <b>endlich.</b> $\varphi(h) = 0$ für $h > p$ |
| MA(q)     | <b>endlich.</b> $\rho(h) = 0$ für $h > q$ | exponentielle Hüllkurve                      |
| ARMA(p,q) | wie AR(p) ab $h > q$                      | wie MA(q) ab $h > p$                         |

Durch Betrachtung von eACF und ePACF können wir also (als Alternative zum AIC) ein geeignetes Modell auswählen, insbesondere wenn eine ab einem gewissen Lag unter die Signifikanzgrenzen verschwindet.

Diese Vorgehensweise ist jedoch nur empfehlenswert für  $p + q \leq 2$ , da ansonsten die Überlagerungen das Bild zu sehr verwischen.

### 3.5 R-Befehle

**pacf** Plottet die empirische partielle Autokorrelationsfunktion einer gegebenen Zeitreihe. Auf die Zahlenwerte kann interessanterweise mit `pacf$acf` zugegriffen werden. Kein Schreibfehler!

**arima** Gegeben eine Zeitreihe, berechnet die Funktion Schätzwerte für AR- und MA-Parameter. Die Ordnung muss als `order=c(p,0,q)` angegeben werden. Der mittlere Parameter entspricht dem  $I$  in ARIMA, und steht für die Integrationsordnung (dazu später mehr). Das Argument `include.mean=FALSE` muss verwendet werden, wenn zentrierte Zeitreihen vorliegen, ansonsten wird auch ein Mittelwert geschätzt, und als Intercept ausgegeben. Typische Syntax:

```
arima(Zeitreihe, order=c(p,0,q), include.mean=F)$aic
```

Das angehängte `$aic` bewirkt die Ausgabe des AIC für das gefittete Modell; die Koeffizienten erhält man mit `$coef`, die Varianz der WN mit `$sigma2`. Der einfache Aufruf liefert all diese Daten auf einen Blick. Die Schätzung der Koeffizienten geschieht standardmäßig mit der Maximum-Likelihood-Methode.

Durch Angabe von `xreg=time(Zeitreihe)` kann zugleich ein linearer Trend geschätzt werden. Bei Vorhersagen mittels `predict` muss dann neben `n.ahead` zusätzlich `newxreg=neueZeit` angegeben werden; wobei `neueZeit` ein Vektor mit den Daten (=Monaten bzw. Jahren) sein muss, für welche die Vorhersage berechnet werden soll. Diesen erhält man am einfachsten als `time(neueZeitreihe)`.

**arima.sim** Simuliert AR(I)MA-Prozesse. Typische Syntax:

```
arima.sim(n=100, list(order=c(1,0,2), ar=0.3, ma=c(0.2,0.5)),sd=1)
```

`n=100` gibt die Anzahl der Realisierungen an, die Modellspezifikation geschieht durch die

Angaben in der Liste, wobei bei ARMA-Modellen auf die Angabe von `order` verzichtet werden kann, es genügt, die Koeffizienten ungleich Null anzugeben. Durch `sd` wird die Standardabweichung  $\sigma$  der WN festgelegt.

## 4 ARIMA-Modelle

Dieses Kapitel widmet sich der Untersuchung von nicht schwach-stationären Zeitreihenprozessen; insbesondere von Zeitreihen, die einen stochastischen Trend aufweisen, der sich nicht als einfache Funktion der Zeit beschreiben lässt.

### 4.1 Integrierte Prozesse

Erinnern wir uns an die Zeitreihe der Umlaufrenditen inländischer Inhaberschuldverschreibungen. Augenscheinlich besitzt diese Zeitreihe einen Trend, das mittlere Niveau ändert sich langfristig. Einen linearen Trend anzunehmen, würde jedoch den ökonomischen Gesetzmäßigkeiten widersprechen, da die Zinssätze keine negativen Werte annehmen können, und auch gewisse Grenzen in der Vergangenheit nie überschritten haben. Vielmehr können wir hier von einem stochastischen Trend sprechen, das heißt, der Trend kann durch einen Random Walk beschrieben werden.

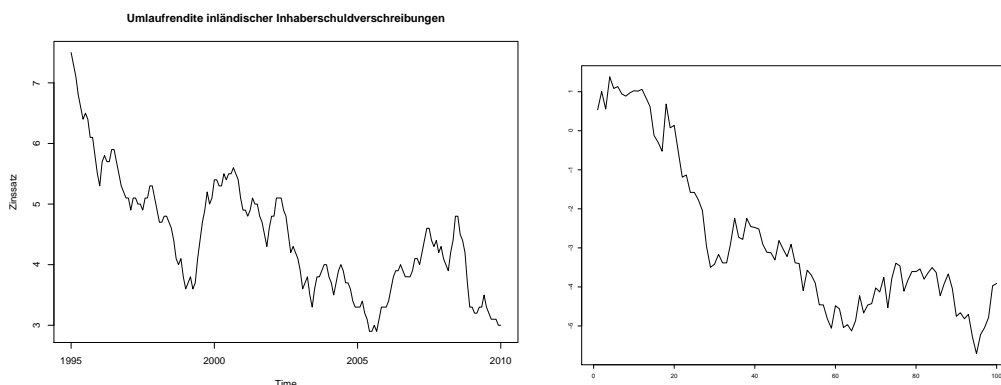


Abbildung 4.1: Umlaufrendite inländischer Inhaberschuldverschreibungen, daneben 100 Realisierungen eines RW

Zur Unterscheidung nicht schwach-stationärer Zeitreihen führen wir zwei Begriffe ein; wir beschränken uns dabei auf den Fall, dass die Zeitreihe durch einmaliges Differenzieren (d.h. Bilden der Zeitreihe  $\Delta X_t = X_t - X_{t-1}$ ) in eine schwach stationäre Zeitreihe überführt werden kann. Im Folgenden ist die Zeitparametermenge  $T$  üblicherweise die Menge der natürlichen Zahlen.

**Definition 4.1.1** (Trend-stationär)

$(X_t)$  heißt *Trend-stationär*, falls für jedes  $t \in T$  die Darstellung

$$X_t = a + b \cdot t + \Psi(\mathbf{B})W_t$$

mit  $\sum_{j=0}^{\infty} |\Psi_j| < \infty$  und  $(W_t) \sim WN(0, \sigma^2)$  gilt.

Da  $b = 0$  nicht ausgeschlossen wurde, ist auch jeder schwach stationäre Prozess, der eine  $MA(\infty)$ -Darstellung besitzt, Trend-stationär.

**Definition 4.1.2** (Differenzen-stationär)

$(X_t)$  heißt *integriert der Ordnung 1* oder *Differenzen-stationär*, falls  $\Delta X_t = X_t - X_{t-1}$  die Differenzgleichung

$$\Delta X_t = (1 - \mathbf{B})X_t = b + \Psi(\mathbf{B})W_t$$

mit  $\Psi(1) \neq 0$ ,  $(W_t) \sim WN(0, \sigma^2)$  und  $\sum_{i=0}^{\infty} i |\Psi_i| < \infty$  erfüllt.

Die Bedingung  $\Psi(1) \neq 0$  heißt, dass 1 keine Nullstelle von  $\Psi$  ist; dies schließt also gerade aus, dass  $(X_t)$  Trend bzw. schwach stationär ist, da sonst der Linearfaktor  $(1 - \mathbf{B})$  auftaucht, denn für  $X_t = a + b \cdot t + \tilde{\Psi}(\mathbf{B})W_t$  gilt

$$\begin{aligned} \Delta X_t &= (b) + \tilde{\Psi}(\mathbf{B})W_t - \tilde{\Psi}(\mathbf{B})W_{t-1} = (b) + \tilde{\Psi}(\mathbf{B})W_t - \tilde{\Psi}(\mathbf{B})\mathbf{B}W_t \\ &= (b) + (1 - \mathbf{B})\tilde{\Psi}(\mathbf{B})W_t, \end{aligned}$$

hierbei entspricht das Polynom (die Potenzreihe)  $(1 - \mathbf{B})\tilde{\Psi}(\mathbf{B})$  dem Polynom  $\Psi(\mathbf{B})$  in der Definition.

Die erweiterte Summierbarkeitsbedingung sorgt dafür, dass  $X_t = X_0 + \sum_{i=1}^t \Delta X_i$  existiert (s.u.) Die allgemeine Struktur Differenzen-stationären Prozessen liefert die *Beveridge-Nelson-Zerlegung*:

**Satz 4.1.3** (Beveridge-Nelson-Zerlegung) Jeder Differenzen-stationäre Prozess  $(X_t)_{t \geq 0}$  besitzt die folgende Zerlegung:

$$X_t = \underbrace{X_0 + b \cdot t}_{\text{linearer Trend}} + \underbrace{\Psi(1) \sum_{j=1}^t W_j}_{\text{Random Walk}} + \underbrace{\tilde{\Psi}(\mathbf{B})W_0 - \tilde{\Psi}(\mathbf{B})W_t}_{\text{stationäre Komponente}}.$$

Die Reihe  $\tilde{\Psi}$  erhält man dabei aus der Beziehung  $\Psi(z) = \Psi(1) + (z - 1)\tilde{\Psi}(z)$ ,  $z \in \mathbb{C}$ .

Da bei Differenzen-stationären Prozessen gerade  $\Psi(1) \neq 0$ , besitzen diese also tatsächlich stets einen Random-Walk-Anteil.

*Beweis.* Die Existenz und Eindeutigkeit von  $\tilde{\Psi}$  erhält man durch Taylorentwicklung von  $\Psi$  um den Entwicklungspunkt 1. Man sieht, dass  $\tilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i$ , und damit auch

$$\sum_{j=0}^{\infty} |\tilde{\Psi}_j| = \sum_{j=0}^{\infty} \left| \sum_{i=j+1}^{\infty} \Psi_i \right| \leq \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} |\Psi_i| = \sum_{j=0}^{\infty} j |\Psi_j| < \infty. \quad (4.1.1)$$



Aus der Definition erhält man nun durch Einsetzen und Erkennen einer Teleskopsumme:

$$\begin{aligned}
X_t &= X_0 + \sum_{j=1}^t \Delta X_j \\
&= X_0 + \sum_{j=1}^t \left( b + [\Psi(1) + (\mathbf{B} - 1)\tilde{\Psi}(\mathbf{B})]W_j \right) \\
&= X_0 + b \cdot t + \Psi(1) \sum_{j=1}^t W_j + \sum_{j=1}^t (\mathbf{B} - 1)\tilde{\Psi}(\mathbf{B})W_j
\end{aligned}$$

Wegen (4.1.1) ist der letzte Term tatsächlich stationär, da die Koeffizienten von  $\tilde{\Psi}$  absolut summierbar sind. □

### 4.1.1 ARIMA-Modelle

Bei Differenzen-stationären Prozessen kann im Allgemeinen angenommen werden, dass der stationäre Prozess ein ARMA-Prozess ist. Solche Prozesse heißen *integrierte ARMA-Prozesse*, kurz *ARIMA(p,d,q)-Prozesse*, wobei der Parameter d angibt, wie oft der Prozess differenziert werden muss, bis er stationär wird.

**Definition 4.1.4** ( $X_t$ ) heißt ARIMA(p,d,q)-Prozess, wenn Polynome  $\Phi$  und  $\Theta$  vom Grad p bzw. q existieren, sodass für jedes  $t$  die Darstellung

$$\Phi(\mathbf{B})(1 - \mathbf{B})^d X_t = \Theta(\mathbf{B})W_t$$

mit  $(W_t)$  WN gilt.

Die R-Funktion `arima` passt ein ARIMA-Modell an gegebene Zeitreihen an. Liegt augenscheinlich zusätzlich ein linearer Trend vor, so kann dieser durch Angabe von `xreg=time(Zeitreihe)` mit bestimmt werden. Diese lineare Regression wird durchgeführt, *bevor* der Prozess differenziert wird.

## 4.2 Tests auf (In-)Stationarität

### 4.2.1 Test auf Differenzen-Stationarität: Dickey-Fuller-Test

Ein Ansatz, einen Test zu entwickeln, ist folgender: Wir nehmen als Hypothese, eine vorgelegte Zeitreihe sei Differenzen-stationär. Wenn der stationäre Teil des Differenzenprozesses eine AR(p)-Darstellung besitzt, so kann der Prozess selbst als AR(p+1) Prozess (mit lin. Trend) der Form

$$(1 - \mathbf{B})\Phi(\mathbf{B})X_t = b + (1 - \mathbf{B})(1 - \phi_1\mathbf{B} - \dots - \phi_p\mathbf{B}^p)X_t = W_t$$

angesehen werden. Dies formen wir um zu

$$\begin{aligned}
X_t &= b + X_{t-1} + \phi_1(1 - \mathbf{B})X_{t-1} + \dots + \phi_p(1 - \mathbf{B})X_{t-p} + W_t \\
&= b + X_{t-1} + \phi_1\Delta X_{t-1} + \dots + \phi_p\Delta X_{t-p} + W_t.
\end{aligned}$$

Mittels des Regressionsansatzes

$$x_t = b + a_1 x_{t-1} + a_2 (x_{t-1} - x_{t-2}) + \dots + a_{p+1} (x_{t-p} - x_{t-p-1})$$

können nun KQS-Schätzer  $\hat{a}_1, \dots, \hat{a}_{p+1}$  bestimmt werden; und wir können unsere Hypothese formulieren als

$$H: a_1 = 1,$$

und den Test anhand von  $\hat{a}_1$  durchführen. Leider besitzt  $\hat{a}_1$  unter der Hypothese keine Standardverteilung, die Statistik liegt nur vertafelt vor. Als Alternative wird üblicherweise  $K: |a_1| < 1$  gewählt; also (Trend-)Stationarität.

Dieser Test wurde von Dickey und Fuller vorgeschlagen, und ist als `adf.test` (Augmented Dickey-Fuller-Test) in R im Paket `tseries` enthalten.

#### 4.2.2 Test auf (Trend-)Stationarität: KPSS-Test

Als Gegenstück zum ADF-Test lernen wir nun einen Test kennen, dessen Hypothese ist, eine vorgelegte Zeitreihe sei (Trend-)stationär. Beide Tests zusammen können also signifikant angeben, ob Trend- oder Differenzen-Stationarität vorliegt.

Wir beschreiben zunächst nur den Test auf Stationarität; einen Test auf Trend-Stationarität erhält man durch Regression eines linearen Trends, und Betrachtung der Residuen. Dem *Kwiatkowski-Phillips-Schmidt-Shin-Test* liegt die Idee zugrunde, dass nach der Beveridge-Nelson-Zerlegung jeder Differenzen-stationäre Prozess eine Random-Walk-Komponente besitzt. Da die übrigen Komponenten stationär bzw. deterministisch sind, besitzen diese eine konstante Varianz; die Varianz der RW-Komponente wächst jedoch linear in  $t$ . Betrachte die Teststatistik

$$S_T = \frac{\sum_{t=1}^T X_t^2}{T^2 \hat{J}_T},$$

wobei  $\hat{J}_T$  ein Schätzer für die sogenannte langfristige Varianz ist; die hier an die Stelle von  $\sigma^2$  treten muss, aber von gleicher Größenordnung ist. Besitzt  $X_t$  eine RW-Komponente, so wachsen Zähler und Nenner mit der gleichen Ordnung; im stationären Fall hingegen konvergiert die Varianz der Teststatistik gegen Null, und somit auch die Teststatistik selbst in Wahrscheinlichkeit.

Auch hier liegen die Werte der Teststatistik nur vertafelt vor, und auch dort wurden nur einige kritische Werte berechnet.

Der in R implementierte `kpss.test` aus dem Paket `tseries` liefert deshalb nur p-Werte zwischen 0.01 und 0.1; bei potentiell größeren oder kleineren Werten zusammen mit einer Warnung, dass der tatsächliche p-Wert davon abweicht. Standardmäßig testet R auf reine Stationarität, wird ein linearer Trend vermutet, so muss der Test mit dem zusätzlichen Argument `null="Trend"` aufgerufen werden.

#### 4.2.3 Abschließende Bemerkungen

Wir haben nun zahlreiche Verfahren kennengelernt, wie Zeitreihen modelliert werden können. Welche Methode angewendet wird, sollte stets durch vorhandenes Hintergrundwissen über

mögliche Trends oder Saisoneinflüsse; sowie durch Betrachtung des Graphen entschieden werden. Schließlich mag dies auch immer von der Zielsetzung abhängen, ob man an einer eleganten Darstellung oder vielleicht einer weitreichenden Prognose interessiert ist.

Eine klassische Zerlegung, wie auch Differenzierung können beide geeignet sein, Saison und Trend auszublenden, um zu einem stationären Prozess zu gelangen. Eine Richtlinie kann hier wohl nur die Erfahrung geben.

### 4.3 R-Befehle

`adf.test` Augmented Dickey-Fuller Test. Enthalten im Paket `tseries`.

Hypothese: Prozess ist Differenzen-stationär, d.h. integriert von der Ordnung 1,  
Alternative: Prozess ist Trend-stationär.

`kpss.test` KPSS-Test. Enthalten im Paket `tseries`.

Hypothese: Prozess ist stationär (Standard) bzw. Trend-stationär (Option `null="Trend"`)  
Alternative: Prozess ist Differenzen-stationär.

# 5 GARCH-Modell

## 5.1 Heteroskedastische Modelle

Kurse von Finanzmarkttiteln sind meist starken und in ihrer Amplitude nicht konstanten Schwankungen unterworfen: Perioden mit hektischen Kursausschlägen werden von ruhigen Perioden abgelöst, d. h. Perioden in denen die Schwankungen geringer sind. Innerhalb einer Periode sind die Amplituden der Ausschläge positiv autokorreliert: Hohe Ausschläge werden mit großer Wahrscheinlichkeit von hohen Ausschlägen gefolgt und niedrige Ausschläge mit hoher Wahrscheinlichkeit von niedrigen.

Mit anderen Worten: Die bedingte Varianz des Einschrittprognosefehlers ist nicht konstant (das wäre der *homoskedastische* Fall) sondern variabel (das nennen wir *heteroskedastisches* Verhalten). Bei Hochfrequentierten Daten (z. B. Tagesdaten) tritt der heteroskedastische Fall häufig ein.

**Beispiel 5.1.1** Die Renditen des *Standard & Poor's 500* Aktienindex aus den 90er Jahren weisen drei Perioden auf: Im zweiten Drittel des Jahrzehnts sind die Ausschläge geringer und im letzten Drittel sehr hoch, vgl. Abbildung 5.1.

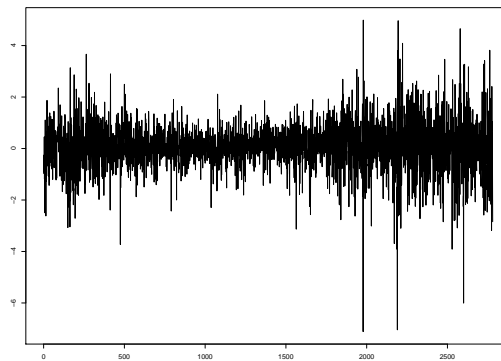


Abbildung 5.1: Renditen des SP500 Aktienindex (1990 bis 1999).

In ARMA Modellen existiert ein solches Verhalten nicht, denn sie sind homoskedastisch. Wir betrachten dazu den einfachsten Fall des AR(1) Modelles.

**Beispiel 5.1.2** (Einschrittprognosefehler im AR(1) Modell) Sei  $X$  ein AR(1) Prozess, etwa

$$X_t = \phi \cdot X_{t-1} + W_t$$

mit  $(W_t) \sim \text{WN}(0, \sigma^2)$ . Haben wir eine Beobachtung  $x_1, \dots, x_n$  gegeben, so ist der beste lineare

Schätzer für den Zeitpunkt  $n + 1$  von  $X$  durch

$$\hat{X}_{n+1} = \phi \cdot X_n \quad (5.1.1)$$

gegeben (wobei  $\phi$  im Anwendungsfall durch einen Schätzer  $\hat{\phi}$  ersetzt werden muss).

Die bedingte Varianz des Einschrittprognosefehlers ist in diesem Fall

$$\mathbb{E}((X_{n+1} - \hat{X}_{n+1})^2 | X_1, \dots, X_n) = \mathbb{E}(W_{n+1}^2 | X_1, \dots, X_n) = \mathbb{E}(W_{n+1}^2) = \sigma^2, \quad (5.1.2)$$

d. h. konstant gleich  $\sigma^2$ . Die gleiche Rechnung könnten wir für ein AR(p) Modell durchführen: Auch hier ist die bedingte Varianz des Einschrittprognosefehlers konstant  $\sigma^2$ .

Beachten wir, dass hier

$$\hat{X}_{n+1} = \mathbb{E}(X_{n+1} | X_n, X_{n-1}, \dots) \quad (5.1.3)$$

gilt, so können wir (5.1.2) auch als

$$\text{Var}(X_{n+1} | X_n, X_{n-1}, \dots) = \sigma^2 \quad (5.1.4)$$

schreiben.

Diese theoretische Überlegung lässt sich auch empirisch überprüfen, etwa indem man eine AR(2) Simulation schrittweise an ein AR(2) Modell fittet. Zum Vergleich zeigt Abbildung 5.2 den Einschrittprognosefehler, wenn man die SP500 Zeitreihe aus Abbildung 5.1 schrittweise an ein AR(2) Modell fittet.

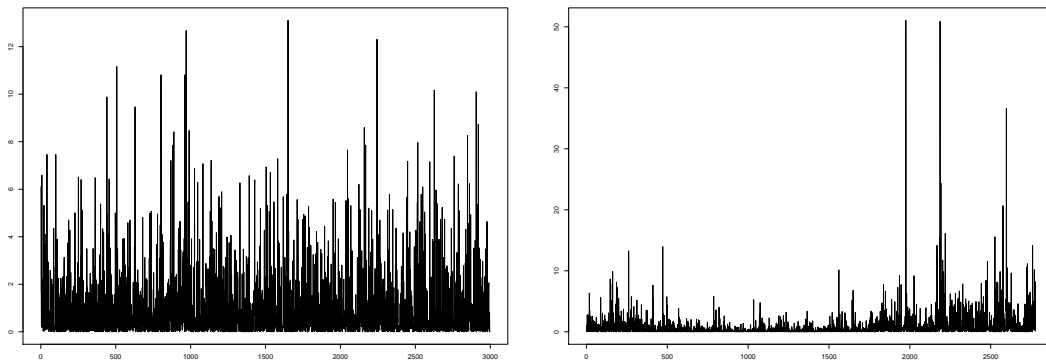


Abbildung 5.2: Verlauf des Einschrittprognosefehlers einer AR(2) Simulation und von SP500 (schrittweise an ein AR(2) Modell gefittet).

Für ARMA Modell ist die bedingte Varianz von  $X_t$ , gegeben die Vergangenheit, konstant. Das Ziel der folgenden Modellbildung ist es, die bedingte Varianz (5.1.4) variabel zu gestalten, d. h. wir hätten gerne einen Prozess  $(X_t)$ , der

$$\text{Var}(X_t | X_{t-1}, X_{t-2}, \dots) = \sigma_t^2 \quad (5.1.5)$$

erfüllt.  $(\sigma_t^2)$  soll dabei eine (nichtkonstante) Folge von Zufallsgrößen sein, den wir als *Volatilitätsprozess* (*Varianzprozess*) bezeichnen.

Ein naheliegender Ansatz für  $(X_t)_{t \in \mathbb{Z}}$  ist

$$X_t = N_t \cdot \sigma_t \quad (5.1.6)$$

mit einer i. i. d. Folge von standardnormalverteilten Zufallsgrößen  $(N_t)$ , so dass  $N_t$  auch unabhängig von  $\sigma_t$  ist für jedes  $t \in \mathbb{Z}$ . Den Prozess  $(N_t)$  nennen wir auch *Innovationsprozess*.

Eine einfache Methode ist es, den Varianzprozess als Funktion der Vergangenheit modellieren, d. h. jedes  $\sigma_t^2$  hat die Gestalt

$$\sigma_t^2 = f(X_{t-1}, X_{t-2}, \dots) \quad (5.1.7)$$

für eine Funktion  $f$ .

Setzen wir die Momentenannahme  $\mathbb{E}(\sigma_t^2) < \infty$  voraus, so gilt zum einen

$$\begin{aligned} \mathbb{E}(X_t | X_{t-1}, X_{t-2}, \dots) &= \mathbb{E}(\sigma_t \cdot N_t | X_{t-1}, X_{t-2}, \dots) \\ &= \sigma_t \cdot \mathbb{E}(N_t | X_{t-1}, X_{t-2}, \dots) \\ &= \sigma_t \cdot \mathbb{E}(N_t) \\ &= 0 \end{aligned}$$

denn  $\sigma_t$  ist messbar bzgl.  $X_{t-1}, X_{t-2}, \dots$  und  $N_t$  ist  $\mathcal{N}(0, 1)$ -verteilt und unabhängig von  $X_{t-1}, X_{t-2}, \dots$ . Wir erhalten

$$\begin{aligned} \text{Var}(X_t | X_{t-1}, X_{t-2}, \dots) &= \mathbb{E}(X_t^2 | X_{t-1}, X_{t-2}, \dots) \\ &= \mathbb{E}(\sigma_t^2 \cdot N_t^2 | X_{t-1}, X_{t-2}, \dots) \\ &= \sigma_t^2 \cdot \mathbb{E}(N_t^2 | X_{t-1}, X_{t-2}, \dots) \\ &= \sigma_t^2 \cdot \mathbb{E}(N_t^2) \\ &= \sigma_t^2, \end{aligned}$$

d. h. (5.1.5).

Vor der Einführung von ARCH Modellen mittelte man die Quadrate  $x_{t-1}^2, x_{t-2}^2, \dots$  in einem Zeitfenster der Breite  $B$ , d. h. man setzte

$$f(x_{t-1}, x_{t-2}, \dots) = \frac{1}{B} \cdot \sum_{i=1}^B x_{t-i}^2. \quad (5.1.8)$$

## 5.2 ARCH(q) Modell

ARCH Modelle machen den Ansatz, die Summe in (5.1.8) noch zu gewichten. Durch die „Erfindung“ dieses Modells im Jahr 1982 gelangte Robert F. Engle in 2003 an den Nobelpreis für Wirtschaftswissenschaften.

**Definition 5.2.1** Ein Prozess  $X = (X_t)_{t \in \mathbb{Z}}$  heißt *autoregressiver, bedingt heteroskedastischer Prozess der Ordnung  $q$*  (kurz: *ARCH( $q$ )*), falls

$$X_t = N_t \cdot \sigma_t \quad \text{mit} \quad \sigma_t^2 = \alpha_0 + \alpha_1 \cdot X_{t-1}^2 + \alpha_2 \cdot X_{t-2}^2 + \dots + \alpha_q \cdot X_{t-q}^2 \quad (5.2.1)$$

für alle  $t \in \mathbb{Z}$  gilt. Dabei sei der Innovationsprozess  $(N_t)$  i. i. d. und standardnormalverteilt, und zudem sei  $N_t$  unabhängig von  $\sigma_t$  für jedes  $t \in \mathbb{Z}$ . Es wird außerdem

$$\alpha_0 > 0 \quad \text{und} \quad \alpha_i \geq 0 \quad \text{für alle } i = 1, \dots, q \quad (5.2.2)$$

angenommen, damit der Volatilitätsprozess  $(\sigma_t^2)$  positiv ist.

Gemäß (5.1.5) gilt

$$\text{Var}(X_t | X_{t-1}, X_{t-2}, \dots) = \sigma_t^2 = \alpha_0 + \alpha_1 \cdot X_{t-1}^2 + \dots + \alpha_q \cdot X_{t-q}^2,$$

was auch die Namensgebung erklärt: Die bedingte Varianz von  $X_t$ , gegeben die Vergangenheit, wird autoregressiv (d. h. abhängig von der Vergangenheit) modelliert. Wir beachten aber, dass im Fall  $\alpha_1 = \dots = \alpha_q = 0$  der Homoskedastische Fall vorliegt.

Die Standardnormalverteilungsannahme an den Innovationsprozess  $(N_t)$  sichert sogar, dass

$$X_t | X_{t-1}, X_{t-2}, \dots, X_{t-q} \sim \mathcal{N}(0, \alpha_0 + \alpha_1 \cdot X_{t-1}^2 + \dots + \alpha_q \cdot X_{t-q}^2) \quad (5.2.3)$$

gilt. Das heißt, die bedingte Verteilung von  $X_t$ , gegeben die Vergangenheit, ist eine Normalverteilung mit Mittelwert 0 und Varianz  $\sigma_t^2$ . Wir beachten aber, dass  $X_t$  selber *nicht* normalverteilt ist.

Die Normalverteilungsannahme an die  $(N_t)$  ist nicht wesentlich für ARCH-Effekte und wir könnten auch einen beliebigen  $\text{WN}(0, 1)$  Prozess für  $(N_t)$  zulassen. Der Einfachheit halber bleiben wir aber bei einer i. i. d.  $\mathcal{N}(0, 1)$  Folge.

Für den Fall  $q = \infty$  müssen wir wie gewohnt

$$\sum_{i=1}^{\infty} |\alpha_i| < \infty \quad (5.2.4)$$

fordern, um die Konvergenz sicher zu stellen.

Wie schon bei ARMA Prozesses stellt sich die natürliche Frage, unter welchen Voraussetzungen an die Parameter eine schwach stationäre Lösung  $(X_t)$  der Gleichung (5.2.1) existiert. Die befriedigende Antwort liefert der folgende Satz.

**Satz 5.2.2** Es existiert genau dann eine schwach stationäre Lösung  $(X_t)$  von (5.2.1), wenn gilt

$$\sum_{i=1}^q \alpha_i < 1. \quad (5.2.5)$$

*Proof.* Die Notwendigkeit von (5.2.5) lässt sich einfach beweisen, indem wir  $\text{Var}(X_t)$  berechnen. Aufgrund von  $\mathbb{E}(X_t) = \mathbb{E}(N_t \cdot \sigma_t) = \mathbb{E}(N_t) \cdot \mathbb{E}(\sigma_t) = 0$  ist

$$\begin{aligned} \text{Var}(X_t) &= \mathbb{E}(X_t^2) \\ &= \underbrace{\mathbb{E}(N_t^2)}_{=1} \cdot \mathbb{E}(\sigma_t^2) \\ &= \alpha_0 + \alpha_1 \cdot \mathbb{E}(X_{t-1}^2) + \dots + \alpha_q \cdot \mathbb{E}(X_{t-q}^2) \\ &= \alpha_0 + \alpha_1 \cdot \text{Var}(X_{t-1}^2) + \dots + \alpha_q \cdot \text{Var}(X_{t-q}^2). \end{aligned}$$

Ist  $(X_t)$  schwach stationär, so gilt  $\text{Var}(X_t) = \dots = \text{Var}(X_{t-q}^2)$  und wir erhalten

$$\text{Var}(X_t) = \frac{\alpha_0}{1 - \alpha_1 - \dots - \alpha_q} \quad (5.2.6)$$

für alle  $t \in \mathbb{Z}$ . Da eine Varianz stets nichtnegativ ist, impliziert dies

$$1 - \alpha_1 - \dots - \alpha_q \geq 0.$$

Im Fall  $1 - \alpha_1 - \dots - \alpha_q = 0$  wäre  $\text{Var}(X_t) = \infty$ , was für einen schwach stationären Prozess  $(X_t)$  per Definition ausgeschlossen wird. Dies zeigt die Notwendigkeit von (5.2.5).

Dass (5.2.5) auch hinreichend ist wollen wir hier nicht beweisen. Eine weitere äquivalente Charakterisierung von (5.2.5), die man im Beweis benötigt, lautet: Alle Nullstellen  $z \in \mathbb{C}$  von  $\Phi(z) = 1 - \alpha_1 \cdot z - \dots - \alpha_q \cdot z^q$  sind betragsmäßig größer als 1.  $\square$

### 5.2.1 Der Prozess $(X_t^2)$

Die starke Korrelation der Quadrate  $(X_t^2)$ , die wir im SP500 Beispiel erkannt haben, ist in einem ARCH(q) Prozess  $(X_t)$  vorhanden:

Definieren wir  $W_t = X_t^2 - \sigma_t^2$ , so gilt  $\mathbb{E}(W_t) = 0$  und mit (5.2.1)

$$X_t^2 = \alpha_0 + \alpha_1 \cdot X_{t-1}^2 + \dots + \alpha_q \cdot X_{t-q}^2 + W_t, \quad (5.2.7)$$

d. h.  $(X_t^2)$  besitzt eine autoregressive Struktur.

## 5.3 GARCH(p, q) Modelle

Aus den Bedürfnissen der Praxis entstanden zahlreiche Erweiterungen des ARCH Modells. Es zeigt sich etwa bei Finanzreihen, dass die Korrelation der Quadrate weit in die Vergangenheit zurückreicht, d. h. für eine gute Schätzung ist ein großes  $q$  von Nöten.

Eine sparsamere Parametrisierung erlaubt dagegen das *verallgemeinerte ARCH Modell der Ordnung (p, q)* (kurz: *GARCH(p, q)*). Wir gehen wieder von  $X_t = \sigma_t \cdot N_t$  aus und definieren den Volatilitätsprozess als

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \cdot X_{t-i}^2 + \sum_{i=1}^p \beta_i \cdot \sigma_{t-i}^2 \quad (5.3.1)$$

für alle  $t \in \mathbb{Z}$ . Wir fordern dabei

$$\alpha_0 > 0, \alpha_i \geq 0 \text{ für alle } i = 1, \dots, q \text{ und } \beta_j \geq 0 \text{ für alle } j = 1, \dots, p \quad (5.3.2)$$

um  $\sigma_t^2 > 0$  zu gewährleisten.

Der Fall  $p = 0$  entspricht offenbar dem ARCH(q) Modell (also  $\text{ARCH}(q) \equiv \text{GARCH}(0, q)$ ).

**Satz 5.3.1** Für das GARCH(p,q) existiert genau dann eine schwach stationäre Lösung  $(X_t)$ , wenn

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1 \quad (5.3.3)$$



erfüllt ist. In diesem Fall gilt

$$\text{Var}(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i - \sum_{i=1}^p \beta_i} \quad (5.3.4)$$

für alle  $t \in \mathbb{Z}$ .  $(X_t)$  lässt sich zudem stets als ARCH( $\infty$ ) Prozess auffassen, d. h. es existieren Konstanten  $(\gamma_i)_{i \in \mathbb{N}_0}$  mit  $\gamma_0 > 0$ ,  $\gamma_i \geq 0$  für alle  $i \geq 1$  und

$$\sigma_t^2 = \gamma_0 + \sum_{i=1}^{\infty} \gamma_i \cdot X_{t-i}^2 \quad \text{sowie} \quad \sum_{i=1}^{\infty} |\gamma_i| < \infty. \quad (5.3.5)$$

Wir merken an, dass für den quadrierten Prozess  $(X_t^2)$  und  $W_t = X_t^2 - \sigma_t^2$  (auch hier ist  $\mathbb{E}W_t = 0$ )

$$X_t^2 = \alpha_0 + \sum_{i=1}^{q \vee p} (\alpha_i + \beta_i) \cdot X_{t-i}^2 - \sum_{i=1}^p \beta_i \cdot W_{t-i} + W_t$$

gilt. Dabei sei  $\alpha_i = 0$  für alle  $i > q$  sowie  $\beta_i = 0$  für alle  $i > p$  gesetzt.  $(X_t^2)$  besitzt also eine ARMA( $q \vee p, p$ ) Struktur.

## 5.4 Parameterschätzung

Ausgangspunkt für die Maximum-Likelihood Schätzung der  $\alpha$ 's und  $\beta$ 's ist das folgende Lemma.

**Lemma 5.4.1** Es sei  $f_1$  die Dichte einer Zufallsgröße  $X_1$  und  $f_t(\cdot | x_{t-1}, \dots, x_1)$  die (bedingte) Dichte von  $X_t$  gegeben  $X_{t-1} = x_{t-1}, \dots, X_1 = x_1$ . Dann besitzt  $(X_1, \dots, X_n)$  die Dichte  $f$ , gegeben durch

$$f(x_1, \dots, x_n) = f_1(x_1) \cdot \prod_{t=2}^n f_t(x_t | x_{t-1}, \dots, x_1). \quad (5.4.1)$$

Ist  $(X_t)$  eine ARCH( $q$ ) Zeitreihe mit normalverteilten Innovationen, so ist die bedingte Verteilung von  $X_t$  gegeben  $X_{t-1}, \dots, X_{t-q}$  eine Normalverteilung mit Mittelwert 0 und Varianz  $\sigma_t^2$ . Die Marginaldichte  $f_1$  und die Dichten  $f_t(\cdot | x_{t-1}, \dots, x_1)$  für  $2 \leq t \leq q$  sind dagegen nur schwer bestimmbar. Daher maximiert man in (5.4.1) nur die Faktoren  $t > q$ , d. h. die Quasi-Likelihood-Funktion

$$L_x(\alpha_0, \dots, \alpha_q) = \prod_{t=q+1}^n f_t(x_t | x_{t-1}, \dots, x_1) = \prod_{t=q+1}^n \sigma_t^{-1} \cdot \varphi(x_t \cdot \sigma_t^{-1}) \quad (5.4.2)$$

wobei  $\varphi$  die Dichte der Standardnormalverteilung sei.

Für ein GARCH( $p, q$ ) Fitting betrachten wir auch die Quasi-Likelihood-Funktion  $L_x$ , um die Parameter zu schätzen. Wir schreiben dann  $L_x(\alpha_0, \dots, \alpha_q, \beta_1, \dots, \beta_p)$  und beachten, dass  $\sigma_t$  in diesem Fall auch wirklich von  $\beta_1, \dots, \beta_p$  abhängt.

Die Wahl von  $q$  und  $p$  können wir mit einem modifizierten Akaike Kriterium durchführen:

$$\text{AICC} = -2 \cdot \frac{n}{n-q} \cdot \log(L_x(\hat{\alpha}_0, \dots, \hat{\alpha}_q, \hat{\beta}_1, \dots, \hat{\beta}_p)) + 2 \cdot (q+p+2) \cdot \frac{n}{n-p-q-3} \quad (5.4.3)$$

Das Paket `tseries` in R stellt eine Funktion `garch(x)` bereit, mit der man ein GARCH(p, q) Modell an einen Datensatz `x` fitten kann. Das AICC Kriterium ist leider in diese Funktion nicht implementiert (Übung).

## 5.5 R-Befehle

`garch(x)` Fittet `x` an ein GARCH(p, q) Modell. `p` und `q` müssen dabei im Vektor `order` übergeben werden (Standardwert ist `p=q=1`)

Setzt man `trace = FALSE`, so verschwindet der Output, den das Optimierungsverfahren für die Maximumssuche ausgibt.

In `$n.likeli` ist  $-\log(L_x(\hat{\alpha}_0, \dots, \hat{\alpha}_q, \hat{\beta}_1, \dots, \hat{\beta}_p))$  gespeichert.

Die Funktion ist Teil des Packetes `tseries`, welches man mit `library(tseries)` laden muss.

`summary(obj)` Gibt eine Übersicht zu einem Objekt aus und berechnet einige Diagnosetests im Fall eines `garch` Objektes. Man kann daran ablesen, wie signifikant die geschätzten Parameter  $\hat{\alpha}_0, \dots, \hat{\alpha}_q, \hat{\beta}_1, \dots, \hat{\beta}_p$  in das Modell eingehen: (\*\*\*) bedeutet starke Signifikanz, kein Sternchen sehr schwache. Im letzteren Fall sollte man `x` an ein GARCH Modell mit kleinerer Ordnung fitten.

# Literaturverzeichnis

- [1] P. Cowpertwait, A. Metcalfe: Introductory Time Series with R, Springer 2009
- [2] P. Brockwell, R. Davis: Introduction to Time Series and Forecasting, Springer 1996
- [3] K. Neusser: Zeitreihenanalyse in den Wirtschaftswissenschaften, Teubner 2006
- [4] R. Schlittgen, B. Streitberg: Zeitreihenanalyse, 5. Auflage, R. Oldenbourg Verlag München Wien 1994
- [5] W. Zucchini: Vorlesungsskript Zeitreihenanalyse. Göttingen, 2004. online unter: <http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer04/>