

Zeitreihenanalyse

Prof. Dr. Hajo Holzmann
Fachbereich Mathematik und Informatik, Universität Marburg

Wintersemester 2008/09
(Stand: 26. Januar 2009)

Inhaltsverzeichnis

1	Beispiele und Grundlagen univariater Zeitreihen	1
1.1	Beispiele für Zeitreihen	1
1.2	Entfernung von Trend und Saisonkomponente	4
1.3	Stochastische Prozesse, Stationarität und Autokorrelation	7
1.4	Schätzung von Erwartungswert und Autokorrelation	13
1.5	Periodogramm und Spektralschätzung	18
1.6	Literaturverweise und Bemerkungen	20
1.7	Anwendung auf Beispiele	21
2	Autoregressive moving average Modelle	26
2.1	Definition und grundlegende Eigenschaften von ARMA Modellen	26
2.2	Autokovarianz und Autokorrelation	31
2.3	Die partielle Autokorrelation	35
2.4	Vorhersage	39
2.5	Parameterschätzung	42
2.6	Ordnungswahl und Modellierung	47
2.7	Literaturhinweise und Bemerkungen	49
2.8	Anwendung auf Datenbeispiele	50
3	Erweiterungen der ARMA Modellklasse	51
3.1	Integrierte Modelle für nicht-stationäre Zeitreihen	51
3.2	Seasonal ARIMA Modelle	54
3.3	Long range dependence und fractional ARMA Modelle	56
3.4	Unit Root Tests	63
3.5	Lineare Zeitreihenmodelle mit heavy tails	72

3.6	Literaturhinweise und Bemerkungen	73
4	Spektraltheorie	74
4.1	Zyklische Zeitreihen	74
4.2	Das Spektralmaß	75
4.3	Nichtparametrische Spektraldichteschätzung	79
4.4	Der Spektralsatz	80
4.5	Filterung	84
4.6	Anwendungen auf ARMA Modelle	87
4.7	Anwendungen auf fraktionale ARMA Modelle	91

1 Beispiele und Grundlagen univariater Zeitreihen

1.1 Beispiele für Zeitreihen

Wir verstehen zunächst unter einer Zeitreihe n reellwertige Beobachtungen x_1, \dots, x_n , die zeitlich in (im wesentlichen) gleichen Abständen aufeinander folgen. Diese werden wie üblich als Realisierungen von Zufallsvariablen X_1, \dots, X_n interpretiert.

Beispiel 1.1

1. Datensatz `milk` aus package `TSA`. Enthält “Average monthly milk production per cow in the US” über den Zeitraum 1994-2005. Die Zeitreihe ist in Abb. 1 geplottet. Man erkennt deutlich eine jährliche Saisonkomponente sowie einen (eventuell linearen) ansteigenden Trend.
2. Datensatz `airmiles` aus package `TSA`. Enthält “Monthly U.S. airline passenger-miles” über den Zeitraum 1996-2005. Die Zeitreihe ist in Abb. 1 geplottet. Man erkennt neben einer jährlichen Saisonkomponente und einem (eventuell linearen) ansteigenden Trend eine Strukturbruch nach dem 11. September 2001.
3. Datensatz `airpass` aus package `TSA`. Enthält “Monthly total international airline passengers” über den Zeitraum 1960-1971. Die Zeitreihe ist in Abb. 1 geplottet. Man erkennt wiederum eine jährliche Saisonkomponente und einen ansteigenden Trend. Darüber hinaus wächst jedoch auch die Varibilität der Reihe mit steigendem Level.
4. Datensatz `beersales` aus package `TSA`. Enthält “Monthly beer sales in millions of barrels” über den Zeitraum 1975-1990. Die Zeitreihe ist in Abb. 2 geplottet. Man erkennt wiederum eine jährliche Saisonkomponente. Zunächst liegt ein ansteigender Trend bis ca. 1981 vor, danach verharrt die Reihe auf einem mittleren Niveau.
5. Datensatz `tempdub` aus package `TSA`. Enthält “Monthly average temperature (in degrees Fahrenheit) recorded in Debuque” über den Zeitraum 1964-1975. Die Zeitreihe ist in Abb. 2 geplottet. Die Zeitreihe scheint durch eine nahezu deterministische jährliche Saisonkomponente beschrieben zu werden.

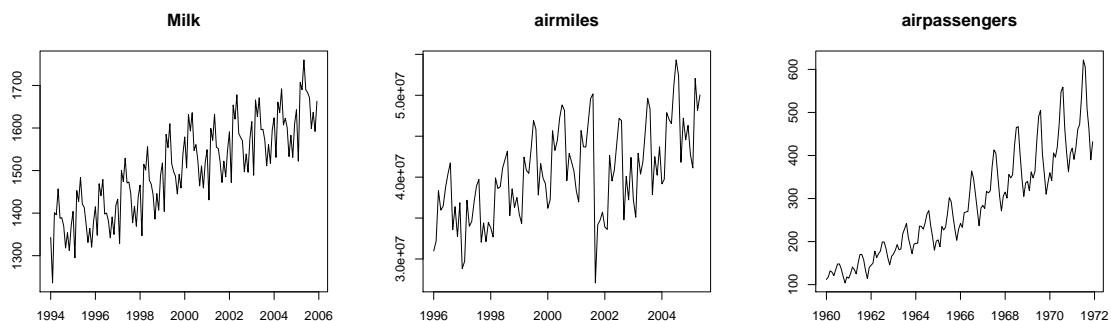


Abbildung 1: Milk, Airmiles und Airpassengers

6. Datensatz `prod.dat` von Shumway und Stoffer (2006). Enthält den monatlichen US “Federal Reserve Production Index” Zeitraum im Zeitraum 1948-1978. Die Zeitreihe ist in Abb. 2 geplottet. Man erkennt in erster Linie einen steigenden Trend, der jedoch nicht regelmäßig (sondern stochastisch) verläuft.
7. Datensatz `lynx` (package `datasets`, welches stets verfügbar ist). Enthält “Annual numbers of lynx trappings in Canada” im Zeitraum 1821 – 1934 im Mackenzie River district in nordwest Kanada. Die Gesamtpopulation an Luchsen in diesem Distrikt wird durch die Reihe wiedergespiegelt. Die Zeitreihe ist in Abb. 11 geplottet. Es gibt keinen Trend und auch keine direkt erkennbare Saisonkomponente, allerdings verhält sich die Reihe “periodisch”.
8. Datensatz `sunspot.year` (package `datasets`, welches stets verfügbar ist). Enthält “Yearly numbers of sunspots” im Zeitraum 1700 – 1988. Die Zeitreihe ist in Abb. 11 geplottet, sie verhält sich “periodisch”.
9. Datensatz `gnp96.dat` von Shumway und Stoffer (2006). Enthält das US BIP (engl: gross national product) im Zeitraum 1947 - 2002 (als Quartalsdaten, bis zum 3. Quartal 2002). Genauer handelt es sich um inflationsbereinigte Daten (Basisjahr 1996), die auch saisonbereinigt wurden. Die Zeitreihe ist in Abb. 12 geplottet. Man erkennt in erster Linie einen wachsenden Trend.

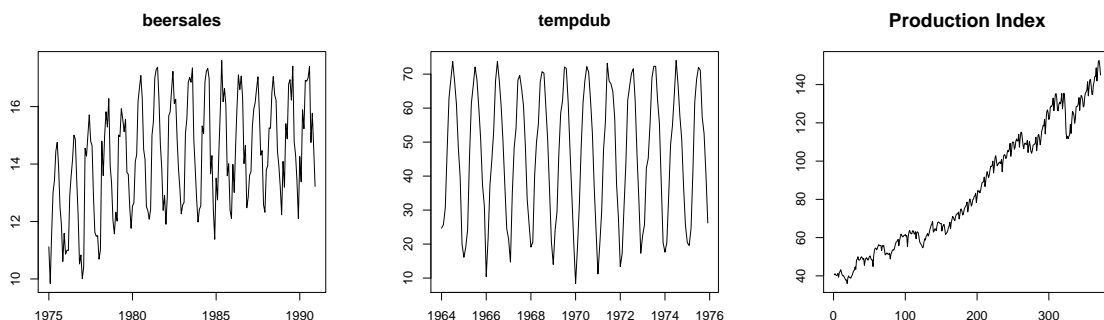


Abbildung 2: Beersales, Tempdub und Production Index

10. Datensatz `oil.price` aus package `TSA`. Enthält “Monthly spot price for crude oil, Cushing, OK (in U.S. dollars per barrel)” über den Zeitraum 1986–2006 (nur erster Monat für 2006). Die Zeitreihe ist in Abb. 12 geplottet. Man erkennt zunächst abgesehen vom starken Preisanstieg ab 2002 relativ wenig Struktur, eventuell noch eine jährliche Saisonalität.
11. Datensatz `d.sp9003lev` aus package `FinTS`. Enthält den S&P500 Index (tägliche Spotpreise) im Zeitraum 1990-2003. Die Zeitreihe ist in Abb. 12 geplottet. Man erkennt wiederum ein tendenziell steigendes Niveau sowie eine steigende Schwankung bei steigendem Niveau.
12. Datensatz `d.fxjp00` aus package `FinTS`. Enthält den täglichen \$/Yen Wechselkurs (tägliche Spotpreise) im Zeitraum 2000-2004. Die Zeitreihe ist in Abb. 12 geplottet.

Man erkennt wenig systematische Struktur.

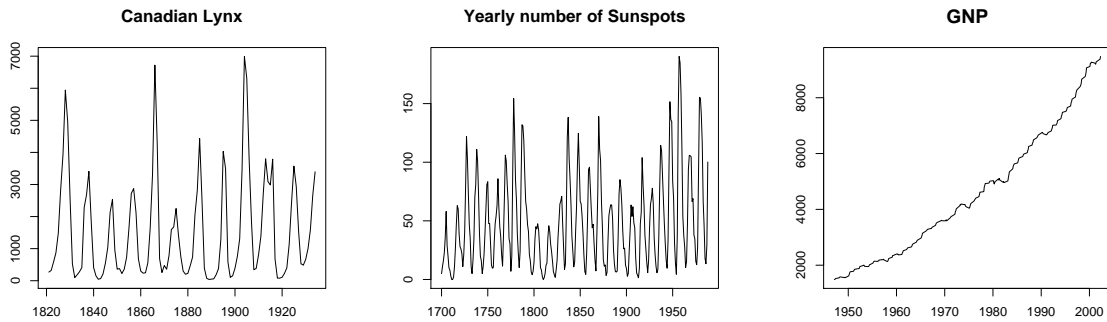


Abbildung 3: Lynx, Sunspots und GNP

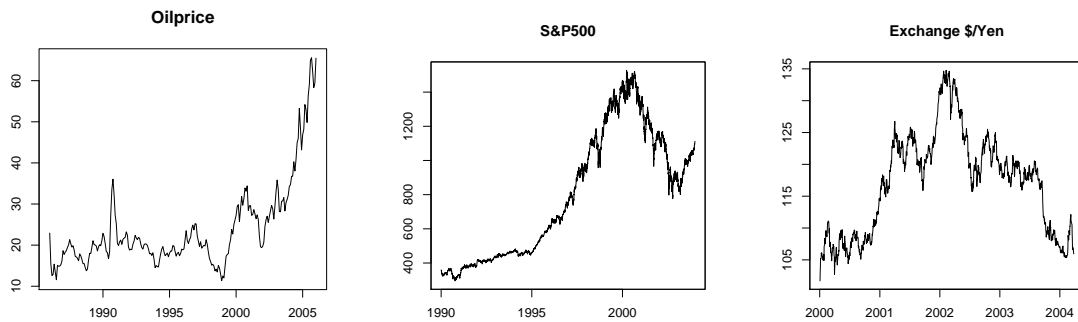


Abbildung 4: Oilprice, Sp500 und Yen/\$ Wechselkurs

1.2 Entfernung von Trend und Saisonkomponente

Klassischer Modellierungsansatz

Im klassischen Modellierungsansatz zerlegt man die Zeitreihe in deterministische Trend- und Saisonkomponenten sowie einen zufälligen, stationären Anteil. Man setzt also an

$$x_t = m_t + s_t + e_t, \quad (1)$$

wobei

- m_t : deterministische Trendfunktion ($m_t = m(t)$),
- s_t : deterministische Saisonkomponente ($s_t = s(t)$).
Falls etwa a Saisons vorliegen, gilt $s(t+a) = s(t)$. Weiter nimmt zur Identifizierbarkeit an, dass diese um 0 herum variiert: $s(1) + \dots + s(a) = 0$.
- e_t : zufällige, stationäre Fehler. Dabei bedeutet stationär, dass sich die gemeinsame Verteilung endlich vieler Komponenten im Zeitverlauf nicht verändert. Zur Identifizierung nimmt man noch $Ee_t = 0$ an.

Vorgehen in der klassischen Zeitreihenanalyse

- a. Schätzen von m_t und s_t , etwa \hat{m}_t oder \hat{s}_t .
- b. Residuenbildung $\hat{e}_t = x_t - \hat{m}_t - \hat{s}_t$, Modellierung von \hat{e}_t über ein Modell für stationäre Zeitreihen.

Bemerkungen zum klassischen Modellierungsansatz

Vorteile:

- a. einfaches, transparentes Modell
- b. explizite Schätzung von m_t und s_t ermöglicht gute Visualisierung.

Nachteil:

Trend und Saisonkomponente sind rein deterministisch, dadurch werden Vorhersagen (insbesondere bei Extrapolation des Trends) häufig unplausibel.

1. Trendschätzung und -entfernung ohne Saisonalität

a. parametrische Regressionsmodelle

Gängige Modelle sind

polynomialer Trend: $m(t) = a_k t^k + \dots + a_1 t + a_0$, $a_i \in \mathbb{R}$, $k \geq 1$,

exponentieller Trend: $m(t) = a_0 + a_1 \exp(a_2 t)$, $a_i \in \mathbb{R}$, $a_2 > 0$.

b. Moving average

Für $q \in \mathbb{N}$ setze

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^q x_{t+j}, \quad q+1 \leq t \leq n-q. \quad (2)$$

Statt der Gewichte $w_j = 1/(2q+1)$ kann man auch allgemeine Gewichte zulassen (die sich zu 1 summieren). Auf diese Weise erhält man etwa Kern-Schätzer (z. B. den lokal-linearen

Schätzer). Diese erlauben insbesondere eine Randkorrektur, d.h. die Werte des Trends können für alle $1 \leq t \leq n$ geschätzt werden.

c. Exponentielles Glätten

Für $a \in (0, 1)$ setze

$$\hat{m}_t = x_1, \quad \hat{m}_t = ax_t + (1-a)\hat{m}_{t-1}, \quad t = 2, \dots, n.$$

Man erhält also

$$\hat{m}_t = \sum_{j=0}^{t-2} a(1-a)^j x_{t-j} + (1-a)^{t-1} x_1.$$

d. Differenzenbildung

Bei der Differenzenbildung kann der Trend entfernt werden, ohne ihn explizit zu schätzen. Man setzt $Dx_t = x_t - x_{t-1}$, $t \geq 2$, sowie induktiv

$$D^j x_t = D(D^{j-1} x_t) = \sum_{k=0}^j \binom{j}{k} (-1)^k x_{t-k}.$$

Für eine polynomiellen Trend $m(t) = \sum_{k=0}^j a_k t^k$ ist dann $D^j m_t$ konstant (dazu zeige, dass einfache Differenzenbildung den Grad des polynomiellen Trends um eins verringert). Man passt dann häufig an die Differenzenzeitreihe ein stationäres Zeitreihenmodell an. Im Modell (1) wäre dies (bei erfolgreicher Entfernung des Trends) ein Modell für $D^j e_t$ und nicht für e_t selbst. Mit Differenzenbildung kann man auch zufällige Trends bzw. Abweichungen von der Stationarität in Form von random walk Komponenten entfernen.

2. Schätzung und Entfernung von Trend und Saisonkomponente

a. parametrische Regressionsmodelle

Ein geeignetes Modell bei einer Saisonkomponente mit a Saisons ist ein trigonometrisches Polynom eines bestimmten Grades

$$s(t) = \sum_{j=1}^p \left(b_j \cos(jt \frac{2\pi}{a}) + c_j \sin(jt \frac{2\pi}{a}) \right), \quad b_j, c_j \in \mathbb{R}. \quad (3)$$

Für $p \leq [(a-1)/2]$, wobei $[x]$ den ganzzahligen Anteil von $x \in \mathbb{R}$ bezeichnet, gilt für Modelle der Form (3) auch $s(1) + \dots + s(a) = 0$, vgl. Lemma 1.33 (es kann für gerades a noch der Kosinus Term $\cos(\pi t)$ dazugenommen werden).

Insbesondere kann man translatierte Kosinus (oder äquivalent Sinus) Terme der Form $s(t) = b \cos(t2\pi/a + \omega)$ in der Form (3) schreiben (Additionstheorem).

Man schätzt dann $m(t)$ und $s(t)$ in einem gemeinsamen Schritt mit der Methode der kleinsten Quadrate.

b. Seasonal moving average

Ist die Saisonkomponente $a = 2q$ gerade, so setzt man

$$\hat{m}_t = \frac{1}{a} (x_{t-q/2} + x_{t-q/2+1} + \dots + x_{t+q/2-1} + x_{t+q/2}), \quad q+1 \leq t \leq n-q,$$

Für $a = 2q+1$ nutze direkt (2) für diese Wahl von q . Beachte, dass sich die saisonale Komponente wegen $s(1) + \dots + s(a) = 0$ sowie der Periodizität gerade heraushebt. Man kann

auch stärker glätten, indem man Vielfache von q in beide Richtungen zurückgeht.

Für $k = 1, \dots, a$ setze

$$\hat{w}_k = \text{mean} \{ (x_{k+ia} - \hat{m}_{k+ia}, \quad i = 0, 1, \dots \text{ mit } k + ia \leq n) \}.$$

Diese Menge enthält entweder $[n/a]$ oder $[n/a] + 1$ Elemente, wobei $[x]$ für $x \geq 0$ die größte ganze Zahl $\leq x$ bezeichnet. Dann setze

$$\hat{s}_k = w_k - \frac{1}{a} \sum_{j=1}^a w_j.$$

c. *Differenzen zum Lag a*

Setze $D_a = x_t - x_{t-a}$. Dann erhält man im Modell (1)

$$D_a x_t = m_t - m_{t-a} + e_t - e_{t-a}.$$

Durch eine weitere Differenzenbildung $D^j(D_a x_t)$ kann dann der Trend $m_t - m_{t-a}$ von $D_a x_t$ entfernt werden, so dass $D^j(D_a e_t)$ verbleibt.

3. Transformation

Manchmal muss zunächst x_t geeignet transformiert werden, also zu $y_t = f(x_t)$ für eine Funktion f übergegangen werden, um auf ein Modell der Form (1) zu kommen. Ist etwa

$$x_t = m_t s_t e_t, \quad m_t, s_t, e_t > 0,$$

so bietet sich logarithmieren an. Allgemeiner werden häufig für $x_t > 0$ Transformationen aus der Box-Cox-Familie angewendet, dabei ist die Transformation f_λ gegeben durch

$$f_\lambda(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(x), & \lambda = 0 \end{cases}$$

1.3 Stochastische Prozesse, Stationarität und Autokorrelation

Definition 1.2

Ein reellwertiger (komplexwertiger) stochastischer Prozess $\mathcal{X} = (X_t)_{t \in T}$, $T \subset \mathbb{R}$, auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) heißt reelle (komplexe) *Zeitreihe*, falls für die Indexmenge $T = \mathbb{N}_0$ oder $T = \mathbb{Z}$ gilt.

Wir lassen komplexwertige Zeitreihen aus technischen Gründen mit zu.

Reelle Zeitreihen

Beispiel 1.3

a. *weißes Rauschen*. Eine reelle Zeitreihe $(e_t)_{t \in \mathbb{Z}}$ mit $Ee_t = 0$, $Ee_t^2 = \sigma_e^2 < \infty$, $Ee_{t_1}e_{t_2} = 0$, $t_1 \neq t_2$ heißt weißes Rauschen.

b. *random walk*. Für ein weißes Rauschen $(e_t)_{t \in \mathbb{Z}}$ heißt \mathcal{X} mit $X_0 = 0$, $X_t = e_t + X_{t-1}$, $t \geq 1$, $X_t = X_{t+1} - e_{t+1}$, $t \leq -1$, ein (zweiseitiger) random walk, es gilt also stets $X_{t+1} = X_t + e_{t+1}$.

c. *AR(1)* Gilt für $\mathcal{X} = (X_t)_{t \in T}$ für ein weißes Rauschen $(e_t)_{t \in \mathbb{Z}}$ und ein $\phi \in \mathbb{R}$, $|\phi| \neq 1$, $X_t = \phi X_{t-1} + e_t$, $t \in \mathbb{Z}$, so heißt $(X_t)_{t \in T}$ eine autoregressive Zeitreihe der Ordnung 1 (AR(1)).

d. *MA(1)* Gilt für $\mathcal{X} = (X_t)_{t \in T}$ für ein weißes Rauschen $(e_t)_{t \in \mathbb{Z}}$ und ein $\theta \in \mathbb{R}$ $X_t = e_t + \theta e_{t-1}$, $t \in \mathbb{Z}$, so heißt $(X_t)_{t \in T}$ eine moving average Zeitreihe der Ordnung 1 (MA(1)).

Definition 1.4 (Randverteilungen)

Sei $\mathcal{X} = (X_t)_{t \in T}$ ein reellwertiger stochastischer Prozess. Die Verteilungen der Form

$$F_{t_1, \dots, t_r}^{\mathcal{X}}(x_1, \dots, x_r) = P(X_{t_1} \leq x_1, \dots, X_{t_r} \leq x_r), \quad x_i \in \mathbb{R},$$

wobei $t_1 < \dots < t_r$, $t_i \in T$, $r \geq 1$, heißen die endlichdimensionalen Randverteilungen von \mathcal{X} .

Die endlichdimensionalen Verteilungen bestimmen die Verteilung des Prozesses eindeutig.

Definition 1.5 (Konsistente Verteilungen)

Eine Familie von endlichdimensionalen Verteilungen der Form

$$\mathfrak{F} = \left\{ F_{t_1, \dots, t_r}(x_1, \dots, x_r), \quad x_i \in \mathbb{R}, t_1 < \dots < t_r, t_i \in T, r \geq 1 \right\}, \quad (4)$$

heißt *konsistent*, falls für alle $t_1 < \dots < t_r$, $t_i \in T$, $r \geq 1$, $1 \leq j \leq r$

$$F_{t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_r}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_r) = \lim_{x \rightarrow \infty} F_{t_1, \dots, t_r}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_r). \quad (5)$$

Bei Bildung von Randverteilungen bleibt man also in der Familie \mathfrak{F} .

Satz 1.6 (Existenzsatz von Kolmogorov)

Eine Familie \mathfrak{F} von endlichdimensionalen Verteilungen der Form (4) ist genau dann die Familie der endlichdimensionalen Randverteilungen eines geeigneten stochastischen Prozesses $\mathcal{X} = (X_t)_{t \in T}$, falls \mathfrak{F} konsistent ist.

Definition 1.7 (Starke Stationarität)

Ein reellwertiger stochastischer Prozess $\mathcal{X} = (X_t)_{t \in T}$ heißt *stark stationär*, falls für alle $t_1 < \dots < t_r, t_i \in T, r \geq 1$, sowie $t \in T$ gilt

$$F_{t_1, \dots, t_r}^{\mathcal{X}}(x_1, \dots, x_r) = F_{t_1+t, \dots, t_r+t}^{\mathcal{X}}(x_1, \dots, x_r), \quad x_i \in \mathbb{R}.$$

Wir nehmen dabei an, dass die Indexmenge T unter Addition abgeschlossen ist.

Die endlichdimensionalen Verteilungen sind also invariant unter Translation der Zeit.

Bemerkung 1.8

Für eine Zeitreihe \mathcal{X} genügt es für die starke Stationarität zu fordern, dass die Prozesse \mathcal{X} und $\mathcal{Y} = (Y_t)_{t \in T}$ mit $Y_t = X_{t+1}$ die gleichen endlichdimensionalen Verteilungen haben (d.h. man kann sich auf Zeittranslationen um eins beschränken).

Satz 1.9 (zweiseitige Fortsetzung)

Sei $(X_t)_{t \in \mathbb{N}_0}$ eine reelle, stark-stationäre Zeitreihe. Dann existiert eine reelle, stark-stationäre Zeitreihe $(\tilde{X}_t)_{t \in \mathbb{Z}}$, so dass $(X_t)_{t \in \mathbb{N}_0}$ und $(\tilde{X}_t)_{t \in \mathbb{N}_0}$ die gleiche Verteilung (auf $\mathbb{R}^{\mathbb{N}_0}$) haben.

Beweis

Wir definieren eine konsistente Familie von Verteilungen auf $\mathbb{R}^{\mathbb{Z}}$ durch

$$F_{t_1, \dots, t_r}(x_1, \dots, x_r) = P(X_0 \leq x_1, \dots, X_{t_r-t_1} \leq x_r), \quad t_1 < \dots < t_r, t_i \in \mathbb{Z}. \quad \blacksquare$$

Nun zeigt man, dass die derart definierte Familie konsistent ist. Dann folgt die Existenz mit Satz 1.6 (Kolmogorov), und die Stationarität ist unmittelbar klar nach Definition der endlichdimensionalen Verteilungen.

Zur Konsistenz: Wir prüfen die Bedingung (5) nach. Diese ist klar, falls $j \geq 2$. Für $j = 1$ erhält man wegen der Stationarität von $(X_t)_{t \in \mathbb{N}_0}$

$$\begin{aligned} F_{t_2, \dots, t_r}(x_2, \dots, x_r) &= P(X_0 \leq x_2, \dots, X_{t_r-t_2} \leq x_r) \\ &= P(X_{t_2-t_1} \leq x_2, \dots, X_{t_r-t_1} \leq x_r) \\ &= \lim_{x_1 \rightarrow \infty} P(X_{t_1} \leq x_1, X_{t_2-t_1} \leq x_2, \dots, X_{t_r-t_1} \leq x_r). \end{aligned}$$

Definition 1.10 (Schwache Stationarität)

Eine reelle Zeitreihe $(X_t)_{t \in \mathbb{Z}}$ heißt L_2 -Zeitreihe, falls $E|X_t|^2 < \infty$ für alle $t \in \mathbb{Z}$. Eine L_2 Zeitreihe heißt *schwach stationär* oder einfach *stationär*, falls für ein $\mu \in \mathbb{R}$

$$EX_t = \mu \quad \text{für alle } t \in \mathbb{Z},$$

und falls darüber hinaus

$$\text{Cov}(X_{t+h}, X_t) = E(X_{t+h} - \mu)(X_t - \mu), \quad \text{ist unabhängig von } t \in \mathbb{Z} \text{ für alle } h \in \mathbb{Z} \quad (6)$$

Die Funktion $\gamma(h) = E(X_{t+h} - \mu)(X_t - \mu)$, $h \in \mathbb{Z}$, heißt dann *Autokovarianzfunktion*, und falls $\gamma(0) \geq 0$, so heißt $\rho(h) = \gamma(h)/\gamma(0)$ die *Autokorrelationsfunktion* der Zeitreihe.

Die schwache Stationarität ist eine weniger restriktive Forderung an eine L_2 Zeitreihe als die starke Stationarität. Für Gaußsche Zeitreihen stimmen diese Bedingungen jedoch überein.

Definition 1.11 (Gaußsche Zeitreihe)

Eine reelle Zeitreihe $(X_t)_{t \in T}$ heißt *Gaußsche Zeitreihe*, falls die zugehörigen endlichdimensionalen Randverteilungen (multivariat) normalverteilt sind.

Satz 1.12

Eine Gaußsche Zeitreihe $(X_t)_{t \in T}$ ist genau dann schwach stationär, wenn sie stark stationär ist.

Satz 1.13

Seien $\mathcal{X}_k = (X_{t,k})_{t \in T}$ für $k \geq 1$ Gaußsche Zeitreihen auf einem gemeinsamen Wahrscheinlichkeitsraum. Gilt dann für eine Zeitreihe $\mathcal{X} = (X_t)_{t \in T}$

$$E(X_{t,k} - X_t)^2 \rightarrow 0, \quad k \rightarrow \infty, \quad \text{für alle } t \in T, \quad (7)$$

so ist auch \mathcal{X} eine Gaußsche Zeitreihe.

Beweis

Zunächst folgt aus (7) die Konvergenz $EX_{t,k} \rightarrow EX_t$, $k \rightarrow \infty$, sowie $\text{Cov}(X_{t_1,k}, X_{t_2,k}) \rightarrow \text{Cov}(X_{t_1}, X_{t_2})$, $k \rightarrow \infty$, $t_1, t_2 \in T$. Wir setzen nun für $t_1 < \dots < t_r$

$$\mathbf{X}_k = (X_{t_1,k}, \dots, X_{t_r,k})^T, \quad \mathbf{X} = (X_{t_1}, \dots, X_{t_r})^T.$$

Dann folgt aus der L_2 -Konvergenz auch die Konvergenz $\mathbf{X}_k \rightarrow \mathbf{X}$ in Wahrscheinlichkeit, und somit auch in Verteilung. Dies impliziert punktweise Konvergenz der charakteristischen Funktionen

$$E \exp(i\mathbf{z}^T \mathbf{X}_k) \rightarrow E \exp(i\mathbf{z}^T \mathbf{X}) \quad k \rightarrow \infty, \quad \mathbf{z} \in \mathbb{R}^r.$$

Nun ist mit $E\mathbf{X}_k = \boldsymbol{\mu}_k$, $\text{Cov} \mathbf{X}_k = \Sigma_k$,

$$E \exp(i\mathbf{z}^T \mathbf{X}_k) = e^{i\mathbf{z}^T \boldsymbol{\mu}_k} \exp(-\mathbf{z}^T \Sigma_k \mathbf{z} / 2), \quad \mathbf{z} \in \mathbb{R}^r,$$

und wegen $\boldsymbol{\mu}_k \rightarrow \boldsymbol{\mu} = E\mathbf{X}$, $\Sigma_k \rightarrow \Sigma = \text{Cov} \mathbf{X}_k$ dann weiter

$$E \exp(i\mathbf{z}^T \mathbf{X}) = e^{i\mathbf{z}^T \boldsymbol{\mu}} \exp(-\mathbf{z}^T \Sigma \mathbf{z} / 2),$$

daher ist \mathbf{X} multinomial normalverteilt. ■

Definition 1.14 (Lineare Zeitreihe)

a. Ist (e_t) ein weißes Rauschen und sind $\psi_j \in \mathbb{R}$ mit $\sum_{j \in \mathbb{Z}} \psi_j^2 < \infty$ und $\mu \in \mathbb{R}$, so heißt

$$X_t = X_t = \mu + \sum_{j \in \mathbb{Z}} \psi_j e_{t-j} \quad (8)$$

ein (zweiseitiger) $MA(\infty)$ Prozess. Falls $\psi_j = 0$ für $j < 0$, so spricht man von einem einseitigen $MA(\infty)$ Prozess.

b. Ist das weiße Rauschen in (8) sogar u.i.v. und gilt darüber hinaus für die Koeffizienten

$$\sum_{j \in \mathbb{Z}} |\psi_j| < \infty, \quad (9)$$

so heißt $(X_t)_{t \in \mathbb{Z}}$ eine *lineare Zeitreihe*. Diese heißt *kausal*, falls $\psi_j = 0$ für $j < 0$.

Beispiel 1.15

Ein $MA(\infty)$ Prozess ist stets (schwach) stationär mit Erwartungswert μ und Kovarianzfunktion

$$\gamma(h) = \sigma_e^2 \sum_{j \in \mathbb{Z}} \psi_j \psi_{j+h}. \quad (10)$$

Eine lineare Zeitreihe ist sogar stets streng stationär. Ist das weiße Rauschen in (8) Gaußsch (und somit u.i.v.), so ist nach Satz 1.13 auch die $MA(\infty)$ Reihe Gaußsch und daher streng stationär.

Definition 1.16 (Reell positiv semidefinit)

Eine Funktion $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ heißt reell positiv semidefinit, falls für alle $\lambda_1, \dots, \lambda_r \in \mathbb{R}, t_1, \dots, t_r \in \mathbb{Z}, r \geq 1$, gilt

$$\sum_{j,k=1}^r \lambda_j \lambda_k \gamma(t_j - t_k) \geq 0. \quad (11)$$

Satz 1.17

Genau dann ist eine Funktion $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ reell positiv semidefinit und symmetrisch, falls γ Autokovarianzfunktion einer (geeignet gewählten) reellen schwach-stationären Zeitreihe ist. Diese kann als Gaußsche Zeitreihe gewählt werden.

Beweis

a. Angenommen, γ ist reell positiv semidefinit und symmetrisch. Für $t_1 < \dots < t_r$ sind die Matrizen

$$\Gamma_{t_1, \dots, t_r} = ((\gamma(t_j - t_k))_{j,k=1, \dots, r},$$

dann ebenfalls symmetrisch und positiv semidefinit. Weiter entsteht $\Gamma_{t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_r}$ aus Γ_{t_1, \dots, t_r} durch Streichung der j -ten Zeile und Spalte. Somit folgt, dass die Familie von Verteilungen

$$F_{t_1, \dots, t_r} \sim N(\mathbf{0}, \Gamma_{t_1, \dots, t_r})$$

konsistent ist. Somit folgt die Existenz einer geeigneten Gaußschen Zeitreihe mit dem Satz von Kolmogoroff.

b. Ist $(X_t)_{t \in \mathbb{Z}}$ stationär mit Autokovarianzfunktion γ , so ist die Symmetrie klar. Da $EX_{t_j} X_{t_k} = \gamma(t_j - t_k)$, gilt

$$\sum_{j,k=1}^r \lambda_j \lambda_k \gamma(t_j - t_k) = E \left(\sum_{j=1}^r \lambda_j X_{t_j} \right)^2 \geq 0. \quad \blacksquare$$

Komplexe Zeitreihen

Für eine komplexe Zufallsvariable X zerlege $X = U + iV$, wobei U, V reellwertige Z.V. sind. Die Verteilung von X ist dann die gemeinsame Verteilung von (U, V) . Sind U und V integrierbar, so setze $EX = EU + iEV$.

Allgemeiner ist die Verteilung des Zufallsvektors $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{C}^n$ definiert als die Verteilung von $(U_1, V_1, \dots, U_n, V_n) \in \mathbb{R}^{2n}$. Sind X_1, X_2 komplexe L_2 Zufallsvariablen, also $E|X_i|^2 < \infty, i = 1, 2$, so setze

$$\text{Cov}(X_1, X_2) = E(X_1 - EX_1) \overline{(X_2 - EX_2)},$$

wobei \bar{z} das komplex-konjugierte von $z \in \mathbb{C}$ bezeichnet. Die Definitionen von komplexer Zeitreihe, starke und schwache Stationarität sowie von komplexen Gaußschen Zeitreihen übertragen sich nun ganz analog vom reellen auf den komplexen Fall. Die Autokovarianz definieren wir dabei mit Konjugation in der zweiten Komponente:

$$\gamma(h) = \text{Cov}(X_{t+h}, X_t) = E(X_{t+h} - \mu) \overline{(X_t - \mu)}, \quad t, h \in \mathbb{Z}.$$

Beachte, dass für die komplexe Autokovarianzfunktion eines stationären Prozesses gilt $\gamma(h) = \overline{\gamma(-h)}$, $h \in \mathbb{Z}$.

Die Sätze 1.6, 1.9 und 1.13 übertragen sich nun ganz analog auf den komplexen Fall. Man beachte jedoch, dass Satz 1.12 für komplexe Gaußsche Zeitreihen *nicht* gilt, da die komplexe Kovarianz die zweidimensionale reelle Kovarianzmatrix nicht eindeutig bestimmt.

Wir kommen nun noch zur Übertragung von Definition 1.16 und Satz 1.17.

Definition 1.18

Eine Funktion $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ heißt *komplex positiv semidefinit* oder einfach positiv semidefinit, falls für alle $\lambda_1, \dots, \lambda_r \in \mathbb{C}$, $t_1, \dots, t_r \in \mathbb{Z}$, $r \geq 1$, gilt

$$\sum_{j,k=1}^r \lambda_j \overline{\lambda_k} \gamma(t_j - t_k) \geq 0. \quad (12)$$

Bemerkung 1.19

Eine einfache Rechnung zeigt, dass eine Funktion $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ genau dann komplex positiv semidefinit ist, falls sie reell positiv semidefinit und symmetrisch ist.

Satz 1.20

Genau dann ist eine Funktion $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ positiv semidefinit, falls γ Autokovarianzfunktion einer (geeignet gewählten) komplexen schwach-stationären Zeitreihe ist. Diese kann als Gaußsche Zeitreihe gewählt werden.

Beweis (Skizze)

Dass eine Autokovarianzfunktion einer komplexen, stationären Zeitreihe positiv semidefinit ist, zeigt man analog wie im reellen Fall.

Angenommen, γ ist positiv semidefinit. Wir zerlegen $\gamma(h) = \gamma_{re}(h) + i\gamma_{im}(h)$ für $\gamma_{re}, \gamma_{im} : \mathbb{Z} \rightarrow \mathbb{R}$, und setzen für $t_1 < \dots < t_r$

$$\Gamma_{re} = (\gamma_{re}(t_j - t_k))_{j,k=1,\dots,r}, \quad \Gamma_{im} = (\gamma_{im}(t_j - t_k))_{j,k=1,\dots,r},$$

wobei wir die Abhängigkeit von t_1, \dots, t_r unterdrückt haben. Weiter

$$\Gamma_{t_1,\dots,t_r} = \frac{1}{2} \begin{pmatrix} \Gamma_{re} & \Gamma_{im} \\ -\Gamma_{im} & \Gamma_{re} \end{pmatrix} \quad (13)$$

Aus (12) erhält man leicht $\gamma_{re}(h) = \gamma_{re}(-h)$ sowie $\gamma_{im}(h) = -\gamma_{im}(-h)$. Damit hat man für $\mathbf{u}, \mathbf{v} \in \mathbb{R}^r$

$$[\mathbf{u}^T, \mathbf{v}^T] \Gamma_{t_1,\dots,t_r} [\mathbf{u}^T, \mathbf{v}^T]^T = \frac{1}{2} \mathbf{u}^T \Gamma_{re} \mathbf{u} + \frac{1}{2} \mathbf{v}^T \Gamma_{re} \mathbf{v} + \mathbf{u}^T \Gamma_{im} \mathbf{v},$$

und für $\lambda_j = u_j + iv_j$ in (12) erhält man damit leicht, dass Γ_{t_1,\dots,t_r} positiv semidefinit ist. Sei nun $(Y_{t_1}, \dots, Y_{t_r}, Z_{t_1}, \dots, Z_{t_r})^T \sim N(0, \Gamma_{t_1,\dots,t_r})$, dann ist die Familie der Verteilungen

$$F_{t_1,\dots,t_r}(x_1, \dots, x_r) = P(Y_{t_1} \leq y_1, Z_{t_1} \leq z_1, \dots, Y_{t_r} \leq y_r, Z_{t_r} \leq z_r), \quad x_j = (y_j, z_j)$$

konsistent (wiederum entsprechende Zeilen - und Spaltenstreichungen), und definiert einen bivariaten (Gauß) Prozess $((Y_j, Z_j))_{j \in \mathbb{Z}}$. Weiter ist

$$EY_{t+h}Y_t = EZ_{t+h}Z_t = \frac{1}{2}\gamma_{re}(h), \quad EY_{t+h}Z_t = -EZ_{t+h}Y_t = \frac{1}{2}\gamma_{im}(h). \quad \blacksquare$$

Dann ist $X_t = Y_t - iZ_t$ ein komplexer Gauß Prozess mit Kovarianzfunktion γ .

Man beachte, dass die Wahl von Γ_{t_1, \dots, t_r} in (13) und somit die der Verteilung des Gauß-Prozesse nicht eindeutig ist, man hätte in (13) auch andere Gewichte für die Γ_{re} Matrizen wählen können, die positiv sind und sich zu 1 addieren.

1.4 Schätzung von Erwartungswert und Autokorrelation

Schätzung des Erwartungswerts

Wir schätzen den Erwartungswert einer stationären Zeitreihe über den Mittelwert der n Beobachtungen.

Satz 1.21

Sei $(X_n)_{n \in \mathbb{Z}}$ ein (schwach) stationäre, reelle Zeitreihe mit Erwartungswert μ und Kovarianzfunktion $\gamma(h)$. Setze $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Dann ist $E\bar{X}_n = \mu$, und für $n \rightarrow \infty$ gilt

$$\begin{aligned} \text{Var} \bar{X}_n = E(\bar{X}_n - \mu)^2 &\rightarrow 0 && \text{falls } \gamma(h) \rightarrow 0, \\ n \text{Var} \bar{X}_n &\rightarrow \sum_{h \in \mathbb{Z}} \gamma(h) && \text{falls } \sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty. \end{aligned} \quad (14)$$

Beweis

Die Erwartungstreue von \bar{X}_n ist leicht einzusehen.

Wir verwenden die Formel

$$\sum_{k,s=1}^n g(r-s) = \sum_{h=-(n-1)}^{n-1} (n-|h|)g(h), \quad (15)$$

wobei g komplexwertig sein darf. Dazu bemerke, dass $h = r - s \in \{- (n - 1), \dots, n - 1\}$ tritt gerade $n - |h|$ -mal auf, für $h \geq 0$ gerade falls $r = h + 1, \dots, n$, und für $h < 0$ für $r = 1, \dots, n + h$.

Wegen $\text{Cov}(X_j, X_k) = \gamma(j - k)$, erhalten wir mit (15)

$$\begin{aligned} n \text{Var} \bar{X}_n &= \frac{1}{n} \sum_{j,k=1}^n \text{Cov}(X_j, X_k) \\ &= \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h) \\ &= \sum_{|h| < n} |\gamma(h)|. \end{aligned} \quad (16)$$

Falls $\gamma(h) \rightarrow 0$, so (Konvergenz des Mittelwertes) auch $\frac{1}{n} \sum_{|h| < n} |\gamma(h)| \rightarrow 0$, somit folgt die erste Gleichung in (14). Zur zweiten bemerken wir, dass die Folgen

$$\left(\left(1 - \frac{|h|}{n}\right) \gamma(h) \mathbf{1}_{|h| \leq n-1} \right)_{h \in \mathbb{Z}}$$

■

punktweise gegen $\gamma(h)$ konvergieren für jeden festen Index h und für $n \rightarrow \infty$, und durch $(|\gamma(h)|)_{h \in \mathbb{Z}}$ majorisiert werden. Da diese Folge summierbar ist, folgt mit dem Satz von der majorisierten Konvergenz für das Zählmaß auf \mathbb{Z} , dass man in (16) zum Grenzwert übergehen darf und den zweiten Teil von (15) erhält.

Beispiel 1.22

Für lineares $X_t = \mu + \sum_{j \in \mathbb{Z}} \psi_j e_{t-j}$ mit $\sum_{j \in \mathbb{Z}} |\psi_j| < \infty$ ist mit (10)

$$\sum_{h \in \mathbb{Z}} |\gamma(h)| \leq \sigma^2 \sum_{h \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} |\psi_j \psi_{j+h}| = \sigma^2 \sum_{h \in \mathbb{Z}} |\psi_j| \sum_{j \in \mathbb{Z}} |\psi_{j+h}| = \sigma^2 \left(\sum_{j \in \mathbb{Z}} |\psi_j| \right)^2$$

und somit auch

$$\sum_{h \in \mathbb{Z}} \gamma(h) = \sigma^2 \left(\sum_{j \in \mathbb{Z}} \psi_j \right)^2.$$

Satz 1.23

Es sei $X_t = \mu + \sum_{j \in \mathbb{Z}} \psi_j e_{t-j}$ eine lineare Zeitreihe mit $\sum_{j \in \mathbb{Z}} |\psi_j| < \infty$ und $\sum_{j \in \mathbb{Z}} \psi_j \neq 0$. Dann gilt mit $\sigma^2 = Ee_t^2$

$$\sqrt{n}(\bar{X}_n - \mu) \rightarrow N\left(0, \sigma^2 \left(\sum_{j \in \mathbb{Z}} \psi_j \right)^2\right).$$

Wir wollen auf den Beweis verzichten, s. etwa Brockwell and Davis (1991, Kap. ??).

Bemerkung 1.24

Falls die $(e_t)_{t \in \mathbb{Z}}$ in Satz 1.23 sogar u.i.v. $N(0, \sigma^2)$ verteilt sind, so ist auch (X_t) ein Gauß-Prozess und

$$\sqrt{n}(\bar{X}_n - \mu) \sim N\left(0, \sigma^2 \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h)\right)$$

mit $\gamma(h)$ wie in (10).

Schätzung der Autokovarianzfunktion

Es sei $(X_t)_{t \in \mathbb{Z}}$ eine reelle stationäre Zeitreihe mit Autokovarianzfunktion $\gamma(h)$. Als Schätzer für $\gamma(h)$ verwendet man am häufigsten

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n), \quad 0 \leq h \leq n-1, \quad (17)$$

wobei noch $\hat{\gamma}(h) = 0$ für $h \geq n$ sowie $\hat{\gamma}(h) = \hat{\gamma}(-h)$ für $h < 0$ gesetzt wird. Wir setzen weiter

$$\Gamma_h = (\gamma(j-k))_{k,j=1,\dots,h} = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(h-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(h-2) \\ \vdots & \vdots & & \vdots \\ \gamma(h-1) & \gamma(h-2) & \dots & \gamma(0) \end{pmatrix} \quad (18)$$

und analog $\hat{\Gamma}_h = (\hat{\gamma}(j-k))_{k,j=1,\dots,h}$.

Satz 1.25

Sei $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ eine reelle Kovarianzfunktion mit $\gamma(h) \rightarrow 0$ für $h \rightarrow \infty$, sowie $\gamma(0) > 0$. Dann ist Γ_h in (18) positiv definit für jedes $h \geq 0$.

Beweis

Als Kovarianzmatrix ist Γ_h positiv semidefinit für jedes h . Sei (X_t) stationärer, zentrierter Prozess mit Kovarianzfunktion γ . Angenommen, die Aussage gilt nicht, und sei r derart, dass Γ_h nicht ausgeartet für $h \leq r$, und Γ_{r+1} sei ausgeartet. Dann existiert $\mathbf{c} \in \mathbb{R}^{r+1}$ mit $c_{r+1} = 1$ (da ja Γ_r nicht ausgeartet) und

$$E \left(\sum_{j=0}^{r+1} c_j X_{t+j} \right)^2 = \mathbf{c}^T \Gamma_{r+1} \mathbf{c} = 0 \quad \forall t \in \mathbb{Z}.$$

Somit ist $X_t \in \text{span}\{X_{t-1}, \dots, X_{t-r}\}$ für alle t , und durch Induktion $X_t \in \text{span}\{X_1, \dots, X_r\}$ für alle $t \geq 1$, etwa $X_t = \mathbf{a}_t^T \mathbf{X}$, $\mathbf{a}_t \in \mathbb{R}^r$, $\mathbf{X} = (X_1, \dots, X_r)^T$. Multipliziere $X_t = \mathbf{a}_t^T \mathbf{X}$ mit \mathbf{X}^T , bilde Erwartungswert, dann erhalte

$$\gamma_t^T = \mathbf{a}_t^T \Gamma_r, \quad \gamma_t = (\gamma(t-1), \dots, \gamma(t-r))^T.$$

Somit gilt $X_t = \gamma_t^T \Gamma_r^{-1} \mathbf{X}$ für alle $t \geq 1$, somit

$$\begin{aligned} \gamma(0) &= \|X_t\|^2 \\ &= \gamma_t^T \Gamma_r^{-1} \Gamma_r \Gamma_r^{-1} \gamma_t \\ &= \gamma_t^T \Gamma_r^{-1} \gamma_t \rightarrow 0, \quad t \rightarrow \infty, \end{aligned}$$

da ja $\gamma(h) \rightarrow 0$. Dies liefert aber wegen $\gamma(0) > 0$ einen Widerspruch. ■

Satz 1.26

Die Funktion $\hat{\gamma} : \mathbb{Z} \rightarrow \mathbb{R}$ ist reell positiv semidefinit und symmetrisch, äquivalent ist jede Matrix $\hat{\Gamma}_h$ positiv semidefinit. Ist $\hat{\gamma}(0) > 0$, so ist $\hat{\Gamma}_h$ sogar stets positiv definit (also nicht singulär).

Beweis

Setze $Y_i = X_i - \bar{X}_n$,

$$A_h = \begin{pmatrix} Y_1 & \dots & Y_n & 0 & \dots & 0 \\ 0 & Y_1 & \dots & Y_n & 0 & \dots \\ \vdots & \dots & & & & \vdots \\ 0 & \dots & 0 & Y_1 & \dots & Y_n \end{pmatrix} \in \mathbb{R}^{h \times (n+h-1)}$$

Dann ist $\frac{1}{n} A_h A_h^T = \hat{\Gamma}_h$, und für $\mathbf{v} \in \mathbb{R}^h$ ist

$$\mathbf{v}^T \hat{\Gamma}_h \mathbf{v} = \frac{1}{n} (A_h^T \mathbf{v})^T (A_h^T \mathbf{v}).$$

Daher ist $\hat{\Gamma}_h$ positiv semidefinit für jedes $h \geq 1$, und $\hat{\gamma}$ positiv semidefinit. Nach Satz 1.17 ist $\hat{\gamma}(h)$ die Autokovarianzfunktion eines stationären Prozesses, und falls $\hat{\gamma}(0) > 0$, so ist dann nach Satz 1.25 $\hat{\Gamma}_h$ sogar nicht singulär, also positiv definit. ■

Wir setzen

$$\gamma_h = (\gamma(0), \dots, \gamma(h))^T \in \mathbb{R}^{h+1}, \quad \hat{\gamma}_h = (\hat{\gamma}(0), \dots, \hat{\gamma}(h))^T.$$

Satz 1.27

Es sei $X_t = \mu + \sum_{j \in \mathbb{Z}} \psi_j e_{t-j}$ eine lineare Zeitreihe mit $Ee_t^4 = \eta\sigma_e^4 < \infty$ und Kovarianzfunktion $\gamma(h)$. Dann gilt für jedes $h \geq 0$

$$\sqrt{n}(\hat{\gamma}_h - \gamma_h) \rightarrow N(\mathbf{0}, V),$$

wobei $V = (v_{p,q})_{p,q=0,\dots,h}$ mit

$$v_{p,q} = (\eta - 3)\gamma(p)\gamma(q) + \sum_{k \in \mathbb{Z}} \left(\gamma(k)\gamma(k-p+q) + \gamma(k+q)\gamma(k-p) \right).$$

Wir verzichten wiederum auf den Beweis, s. Brockwell and Davis (1991, Kap. ??).

Schätzung der Autokorrelationsfunktion

Als Schätzer der Autokorrelation wählt man $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$ (für $\hat{\gamma}(0) > 0$). Nach Satz 1.26 ist auch $\hat{\rho}(h)$ eine positiv semidefinite Funktion.

$$\boldsymbol{\rho}_h = (\rho(1), \dots, \rho(h))^T \in \mathbb{R}^h, \quad \hat{\boldsymbol{\rho}}_h = (\hat{\rho}(1), \dots, \hat{\rho}(h))^T.$$

Satz 1.28

Es sei $X_t = \mu + \sum_{j \in \mathbb{Z}} \psi_j e_{t-j}$ eine lineare Zeitreihe mit $Ee_t^4 = \eta\sigma_e^4 < \infty$ und Autokorrelationsfunktion $\rho(h)$. Dann gilt für jedes $h \geq 0$

$$\sqrt{n}(\hat{\boldsymbol{\rho}}_h - \boldsymbol{\rho}_h) \rightarrow N(\mathbf{0}, W),$$

wobei $W = (w_{p,q})_{p,q=1,\dots,h}$ mit

$$w_{p,q} = \sum_{k \in \mathbb{Z}} \left(\rho(k+p)\rho(k+q) + \rho(k-p)\rho(k+q) + 2\rho(p)\rho(q)\rho^2(k) - 2\rho(p)\rho(k)\rho(k+q) - 2\rho(q)\rho(k)\rho(k+p) \right). \quad (19)$$

Bemerkung 1.29

Die Formel (19) heißt Bartlett's Formel, man kann diese auch umschreiben zu

$$w_{p,q} = \sum_{k=1}^{\infty} (\rho(k+p) + \rho(k-p) - 2\rho(p)\rho(k)) (\rho(k-q) + \rho(k+q) - 2\rho(q)\rho(k)). \quad (20)$$

Beweis

Sei $g : \mathbb{R}^{h+1} \setminus (\{0\} \times \mathbb{R}^h) \rightarrow \mathbb{R}^h$, $g(x_0, \dots, x_h) = (x_1/x_0, \dots, x_h/x_0)^T$. Dann ist

$$Dg(x_0, \dots, x_h) = \frac{1}{x_0} \begin{pmatrix} -\frac{x_1}{x_0} & 1 & 0 & \dots & 0 \\ -\frac{x_2}{x_0} & 0 & 1 & \dots & 0 \\ \vdots & & & \dots & \\ -\frac{x_h}{x_0} & 0 & \dots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{h \times (h+1)}$$

Nach der δ -Methode ist dann

$$\sqrt{n}(\hat{\boldsymbol{\rho}}_h - \boldsymbol{\rho}_h) \rightarrow N(\mathbf{0}, Dg(\boldsymbol{\gamma}_{h+1})V_{h+1}Dg(\boldsymbol{\gamma}_{h+1}^t)).$$

Eine Rechnung mit Blockmatrizen liefert dann

$$w_{p,q} = \frac{1}{\gamma(0)^2} (v_{p,q} - \rho(p)v_{0,q} - \rho(q)v_{p,0} + \rho(p)\rho(q)v_{0,0}). \quad \blacksquare$$

Beispiel 1.30

Für ein weißes Rauschen (e_t) erhalten wir in Bartletts Formel (20) $w_{p,q} = \delta_{p,q}$, wobei $\delta_{p,q}$ das Dirac-delta bezeichnet. Insbesondere ist $\hat{\rho}(h) \rightarrow N(0, 1)$ für alle $h \geq 1$.

1.5 Periodogramm und Spektralschätzung

Seien X_1, \dots, X_n komplexwertige Zufallsvariable.

Definition 1.31

Das *Periodogramm* von (X_1, \dots, X_n) ist definiert durch

$$I_n(\omega) = \frac{1}{n} \left| \sum_{t=1}^n X_t e^{-it\omega} \right|^2. \quad (21)$$

Periodogramm und finite Fourier Transformation

Das Periodogramm wird insbesondere an den *Fourier-Frequenzen* $\omega_j = 2\pi j/n$ für

$$j \in \mathcal{F}_n = \left\{ [-(n-1)/2], \dots, [n/2] \right\}, \quad n \geq 2,$$

ausgewertet, wobei $[x]$ den ganzzahligen Anteil von x bezeichnet. Explizit ist

$$\begin{aligned} \mathcal{F}_n &= \left\{ -(n-1)/2, \dots, (n-1)/2, \quad n \text{ ungerade} \right\}, \\ \mathcal{F}_n &= \left\{ -n/2 + 1, \dots, n/2, \quad n \text{ gerade} \right\}. \end{aligned}$$

Lemma 1.32

Die n Vektoren $\mathbf{b}_j = n^{-1/2} (e^{i\omega_j}, \dots, e^{in\omega_j})^T \in \mathbb{C}^n$, $j \in \mathcal{F}_n$, bilden eine Orthonormalbasis des \mathbb{C}^n .

Beweis

Offensichtlich ist $\|\mathbf{b}_j\|^2 = 1$. Für $\omega \notin 2\pi\mathbb{Z}$ ist

$$\sum_{t=1}^n e^{-it\omega} = e^{-i\omega} \frac{1 - e^{-in\omega}}{1 - e^{-i\omega}}, \quad \omega \notin 2\pi\mathbb{Z}, \quad (22)$$

Da

$$\omega_j - \omega_k = 2(j-k)\pi/n \notin 2\pi\mathbb{Z}, \quad j \neq k, \quad j, k \in \mathcal{F}_n,$$

erhält man

$$\mathbf{b}_j^T \bar{\mathbf{b}}_k = \frac{1}{n} \sum_{t=1}^n e^{it(\omega_j - \omega_k)} = 0,$$

da $n(\omega_j - \omega_k) \in 2\pi\mathbb{Z}$ ($n \geq 2$), und somit der Zähler in (22) verschwindet. ■

Nach Lemma 1.32 kann man schreiben

$$\mathbf{X} = (X_1, \dots, X_n)^T = \sum_{j \in \mathcal{F}_n} a_j \mathbf{b}_j, \quad a_j = \frac{1}{\sqrt{n}} \sum_{t=1}^n X_t e^{-it\omega_j}, \quad (23)$$

so dass

$$I_n(\omega_j) = |a_j|^2, \quad j \in \mathcal{F}_n, \quad \|\mathbf{X}\|^2 = \sum_{j \in \mathcal{F}_n} I_n(\omega_j).$$

Das Periodogramm gibt also Aufschluss darüber, welcher Teil der Variation in X_1, \dots, X_n auf die Fourier-Frequenz ω_j , $j \in \mathcal{F}_n$ entfällt.

Reelle Beobachtungen

Für reelle Beobachtungen X_j gilt offenbar in (23) $a_{-j} = \overline{a_j}$ für $j, -j \in \mathcal{F}_n$, und somit auch $I(\omega_j) = I(\omega_{-j})$. Man plottet daher das Periodogramm nur für positive Fourier Frequenzen, also $\omega_j \geq 0$, $j \in \mathcal{F}_n$. Die Darstellung (23) kann man nun umschreiben zu

$$\mathbf{X} = (X_1, \dots, X_n)^T = a_0 \mathbf{b}_0 + \sum_{j=1}^{[(n-1)/2]} \sqrt{2} (\Re(a_j) \mathbf{c}_j - \Im(a_j) \mathbf{s}_j) + a_{n/2} \mathbf{b}_{n/2}, \quad (24)$$

wobei der letzte Term nur für n gerade auftritt und

$$\begin{aligned} \mathbf{c}_j &= \sqrt{2/n} (\cos(\omega_j), \dots, \cos(n\omega_j))^T, \\ \mathbf{s}_j &= \sqrt{2/n} (\sin(\omega_j), \dots, \sin(n\omega_j))^T. \end{aligned}$$

Auch in (24) gilt

Lemma 1.33

Die Vektoren $\mathbf{b}_0, \mathbf{c}_j, \mathbf{s}_j$, $j = 1, \dots, [(n-1)/2]$ und $\mathbf{b}_{n/2}$ (falls n gerade) bilden eine orthonormale Basis des \mathbb{R}^n .

Testen auf deterministische periodische Komponenten

Man kann das Periodogramm auch dazu verwenden, um zu testen, ob bestimmte deterministische periodische Komponenten in der reellen Reihe (X_t) vorhanden sind. Wir betrachten als Beispiel das Regressionsmodell

$$X_t = \mu + \beta_1 \cos(\omega_j t) + \beta_2 \sin(\omega_j t) + \epsilon_t,$$

für $\beta_1, \beta_2 \in \mathbb{R}$ konstant und ϵ_t i.i.d. $N(0, \sigma^2)$, sowie ein $j \in \mathcal{F}_n$ mit $j > 0$. Darin untersuchen wir die Hypothese

$$H : \beta_1 = \beta_2 = 0.$$

Die i.i.d. Fehlerstruktur ist natürlich in der Praxis häufig zweifelhaft, man wird auch korrelierte Fehler erwarten. Man schätzt μ, β_1 und β_2 über die Methode der kleinsten Quadrate und erhält mit Lemma 1.33 und (24)

$$\hat{\mu} = a_0, \quad \hat{\beta}_1 = \sqrt{2} \Re(a_j), \quad \hat{\beta}_2 = -\sqrt{2} \Im(a_j).$$

Beim Bilden der F-Statistik zur Hypothese H beachtet man im Zähler

$$\hat{\beta}_1^2 + \hat{\beta}_2^2 = 2|a_j|^2 = 2I(\omega_j) \sim \sigma^2 \chi_2^2,$$

sowie im Nenner

$$\|\mathbf{X}\|^2 - \hat{\mu}^2 - \hat{\beta}_1^2 - \hat{\beta}_2^2 = \|\mathbf{X}\|^2 - I_n(0) - 2I_n(\omega_j) \sim \sigma^2 \chi^2(n-3).$$

Somit ist unter H

$$F = \frac{(n-3)I_n(\omega_j)}{\|\mathbf{X}\|^2 - I_n(0) - 2I_n(\omega_j)} \sim F(2, n-3).$$

Man verwirft H für große Werte von F .

1.6 Literaturverweise und Bemerkungen

Zur linearen Regression bei abhängigen Fehlern.

Zur nichtparametrischen Regression in der Zeitaxe: s. etwa Fan and Yao (2003), Kapitel ?.

Beweise des Satzes 1.6 (Komogorov): Billingsley (1979, Kap. 7, Sec. 36).

Zu Satz 1.21: Kreiß und Neumann (2006, Satz ??) geben eine allgemeinere Version mit Hilfe des Spektralsatzes. Relation zum Ergodensatz von von Neumann.

Zu Satz 1.23: Brockwell und Davis (1991) geben einen Beweis über ein Truncation Argument und den CLT für m -abhängige Prozesse. Ein anderer Zugang basiert auf der Approximation der partialsummen durch ein Martingal mit stationären Inkrementen. Für einen Beweis von Satz 1.23 mit diesem Zugang s. Durrett (200?). Heyde (1974) nutzt diesen Zugang, um den CLT unter der schwächeren Bedingung der Stetigkeit der Spektraldichte

$$f(\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{k \in \mathbb{Z}} \psi_k e^{ik\omega} \right|^2$$

bei 0 zu zeigen. Es gibt auch Versionen, in denen ein abhängiges weißes Rauschen zugelassen wird, s. etwa Peligrad und Utev (2006).

Zu Satz 1.27: Brockwell und Davis (1991) geben wiederum einen Beweis über ein Truncation Argument und den CLT für m -abhängige Prozesse. MEHR.

Zu Satz 1.28: Brockwell und Davis (1991) geben auch eine (überraschende) Version, die ohne die Annahme endlicher vieter Momente der ϵ_k auskommt, sie benötigen allerdings die zusätzliche Annahme $\sum_{j \in \mathbb{Z}} \psi_j^2 |j| < \infty$. MEHR

Zu Section 1.5: Brockwell und Davis (1991) diskutieren weitere Tests auf periodische Komponenten, insbesondere falls nicht eine Fourier-Frequenz vorliegt. Anderson (19??, Kapitel ??) diskutiert derartige Tests ausführlich, und untersucht deren Optimalität (uniformly most powerful).

Referenzen

- Anderson, T. W. (1971) *The statistical analysis of time series*. John Wiley & Sons
- Billingsley, P. (1979). *Probability and Measure*, John Wiley and Sons.
- Brockwell P., und Davis, R. A. (1991) *Time Series Analysis: Theory and Methods*
- Durrett, R. (1996) *Probability: theory and examples*. Second edition. Duxbury Press, Belmont.
- Fan, J. and Yao, Q. (2003) *Nonlinear time series. Nonparametric and parametric methods*. Springer, New York.
- Kreiss, J.-P. und Neuhaus, G. (2006) *Einführung in die Zeitreihenanalyse*. Springer, Berlin.
- Heyde, C. C. (1974) On the central limit theorem for stationary processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **30**, 315-320.
- Peligrad, Magda; Utev, Sergey (2006) Central limit theorem for stationary linear processes. *Ann. Probab.* **34**, 1608–1622.

1.7 Anwendung auf Beispiele

1. Datensatz Milk

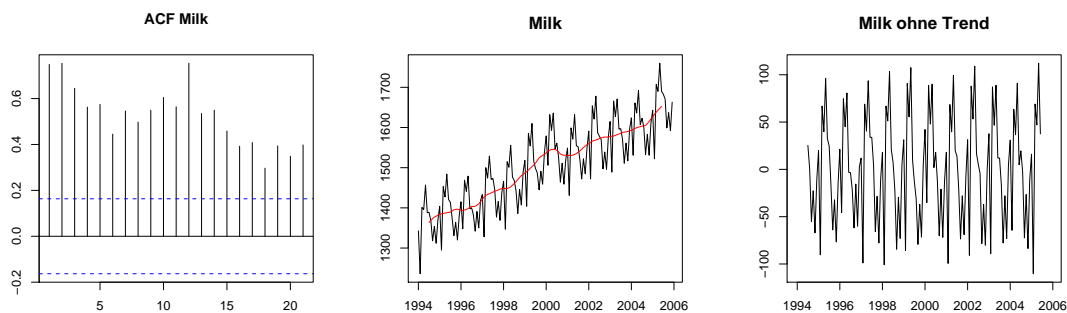
Wir wollen in dem Datensatz `milk` Trend und Saisonkomponente entfernen. Zunächst plotten wir die Autokorrelationsfunktion (mit `acf`) des ursprünglichen Datensatzes (vgl. Abb. 5), diese verharrt auf hohem Niveau, was typisch bei einem Trend ist. Man beachte die blauen gestrichelten Linien, welche bei $\pm q_{0.975}/\sqrt{n}$ verlaufen (q_α das α -Quantil der Normalverteilung), und also punktweise Konfidenzbänder für weißes Rauschen darstellen (vgl. Bemerkung 1.30). Zunächst schätzen wir mit saisonalen moving average den Trend und bilden Residuen

```
filcoef<-c(1/2,rep(1,11),1/2)/12
milk.fil<-filter(milk, filcoef)
milk.det<-milk-milk.fil
```

In Abb. 5 sind Zeitreihe mit geschätztem Trend und die Residuen nach Trendentfernung abgebildet. In Abb. 6 sind deren Autokorrelation sowie ein Periodogramm der Residuen (`spec.pgram`) geplottet. Bei der Autokorrelation erkennt man den Ausschlag bei Lag 12, was auf jährliche Saisonalität hindeutet. Beim Periodogramm ist zu beachten, dass die Frequenzen auf einer 2π -Skala abgebildet sind, und dass die y -Achse eine logarithmische Skala hat. Man erkennt deutlich einen Peak bei $1/12 = 0.083$ (vertikale Linie, `abline(v=1/12)`), welches zur jährlichen Saisonkomponente gehört.

Wir entfernen nun die jährliche Saisonalität durch moving average. Zunächst berechnen wir

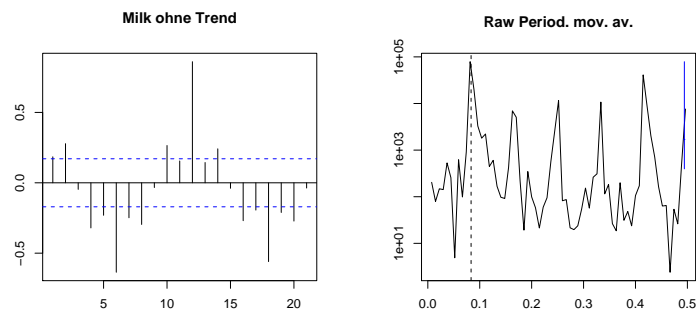
Abbildung 5: Milk mit seasonal moving average Trend



die \hat{w}_k auf zwei Methoden, einmal direkt, und einmal durch lineare Regression auf monatliche Dummyvariablen

```
month<-season(milk)
# Direkte Berechnung
wk<-tapply(na.omit(milk.det), month[is.na(milk.det)==F], mean)
#
g<-lm(milk.det~month-1)
wk2<-g$coef
```

Abbildung 6: Diagnose der Residuen nach Trendentfernung



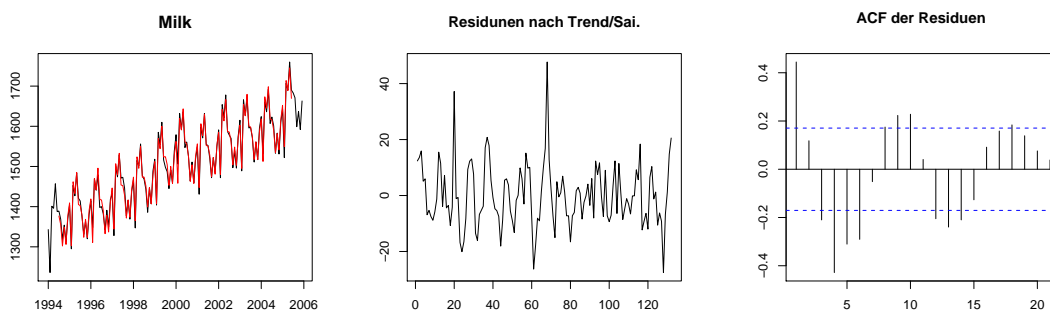
Dabei enthalten w_k und w_{k2} die geschätzten \hat{w}_k . Im nächsten Schritt bilden wir die Residuen, die um den Mittelwert der \hat{w}_k verschoben werden.

```
mcoef<-mean(g$coef)
milk.res<-residuals(g)+ mcoef
milk.res<-residuals(g)
sai<-c(rep(NA,6),fitted.values(g),rep(NA,6))
milk.fit<-milk.fil+sai
```

Dabei wurden sowohl fitted values mit Trend und Saisonschwankung (in `milk.fit`) sowie die entsprechenden Residuen (in `milk.res`) gebildet. Abb. 7 enthält die Zeitreihe mit den gefitteten Werten, sowie einen Plot der Residuen und deren Autokorrelation. Es ist noch relativ viel zyklische Struktur in den Residuen enthalten.

In einem zweiten Ansatz entfernen wir einen linearen Trend sowie die Saisonschwankungen

Abbildung 7: Fit Milk mit moving average Trend und Saisonkomponente sowie Residuenanalyse



durch moving average wie folgt

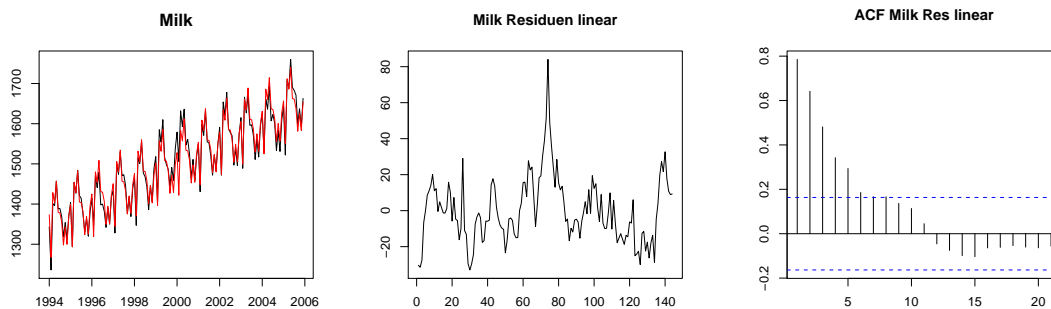
```
g1<-lm(milk~month-1+time(milk))
```

```
milk.res<-residuals(g1)
milk.fit<-fitted.values(g1)
```

Dabei wurden wieder die fitted values mit Trend und Saisonschwankung (in `milk.fit`) sowie die entsprechenden Residuen (in `milk.res`) gebildet. Abb. 8 enthält die Zeitreihe mit den gefitteten Werten, sowie einen Plot der Residuen und deren Autokorrelation. Die Autokorrelationsstruktur ist nun weniger zyklisch und fällt relativ schnell ab.

Man kann nun in einem der beiden Ansätze die Residuen durch ein stationäres Zeitreihenmodell modellieren, wobei man hier den zweiten Ansatz über lineare Trendentfernung bevorzugen würde.

Abbildung 8: Milk und Airmiles



Im Datensatz `airpassengers` geht man zu den Logarithmen über, um die unterschiedliche Variabilität zu entfernen, und führt dann eine Analyse wie im obigen Beispiel durch. Im Datensatz `airmiles` muss zunächst der Strukturbruch am 11. September 2001 explizit modelliert bzw. entfernt werden.

2. Datensatz Beersales

Wir wollen im Datensatz `beersales` Differenzenbildung zur Entfernung der Saisonalität und des Trends benutzen. Wir bilden Differenzen zum lag 12 und danach bilden wir noch eine weitere einfache Differenz wie folgt

```
beerds<-diff(beersales,12)
beerds2<-diff(beerds)
```

In Abb. 9 sind die ACF der ursprünglichen Daten, die Residuen nach Differenzenbildung zum Lag 12 (in `beerds`) und deren Autokorrelation abgebildet. Die ACF zeigt noch für höhere Lags größere Ausschläge. Man kann daher nochmal einfache Differenzen bilden (in `beerds2`), diese und deren Autokorrelation sind in Abb. 10 geplottet. Diese haben nur zum Lag 1 noch Autokorrelation. Ob die zweite Differenzenbildung tatsächlich notwendig ist, kann an dieser

Stelle nicht eindeutig beantwortet werden.

Abbildung 9: Beersales: Desaisonalisierung durch Differenzen zum Lag 12

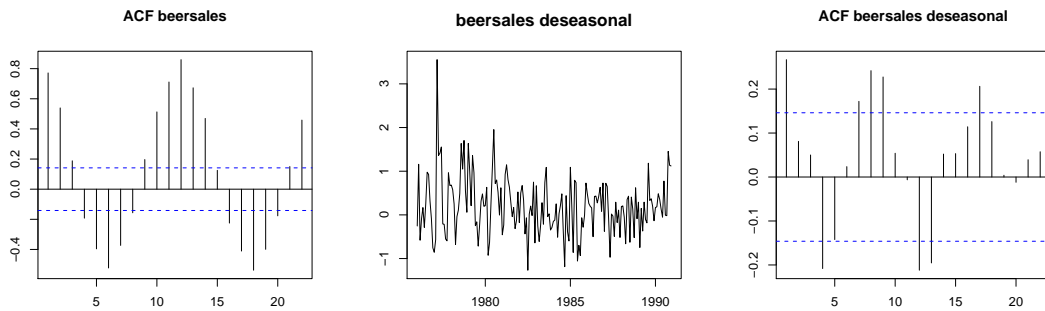
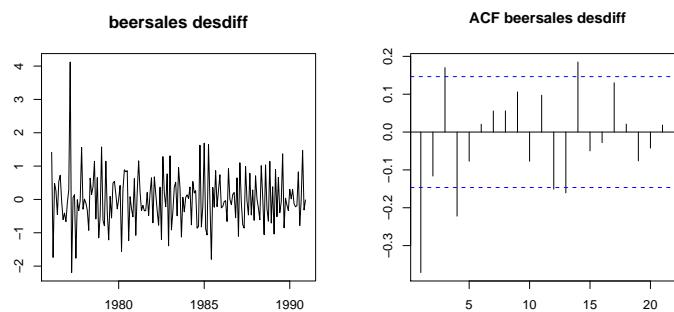


Abbildung 10: Beersales: Weitere Differenzbildung



3. Datensatz Tempdub

Im Datensatz `tempdub` passen wir zur Entfernung der Saisonalität eine sinusoidale Komponente an wie folgt

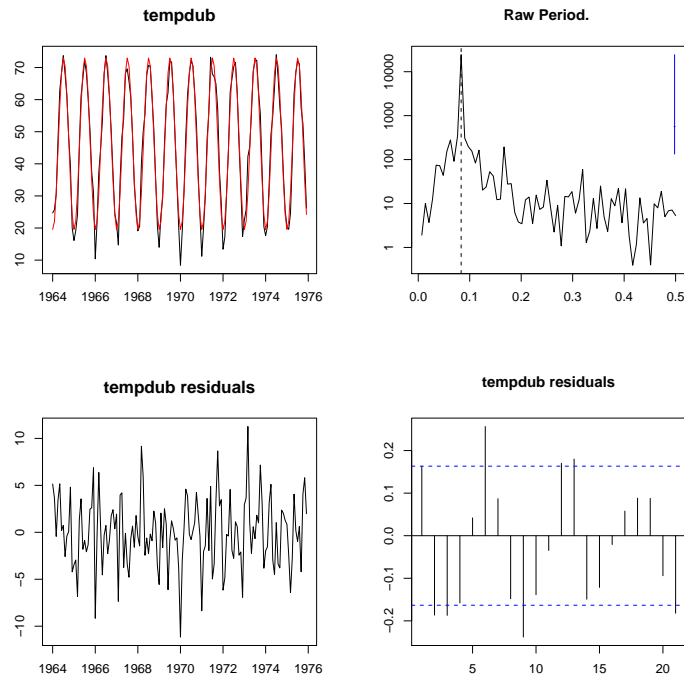
```
t<-time(tempdub)
cs<-cos(t*2*pi)
ss<-sin(t*2*pi)
g<-lm(tempdub~cs+ss)
temp_fit<-as.vector(fitted.values(g))
```

Dabei enthält `tempfit` die angepasste Reihe. Abb. 11 enthält die Zeitreihe mit der angepassten Reihe, einen Plot des Periodogramms (starker Peak bei $1/12$), die Residuen sowie deren Autokorrelation. Es ist kaum Autokorrelation zu erkennen, die Reihe ist also im wesentlichen eine sinusoidale Schwingung mit weißem Rauschen.

3. Datensatz GNP

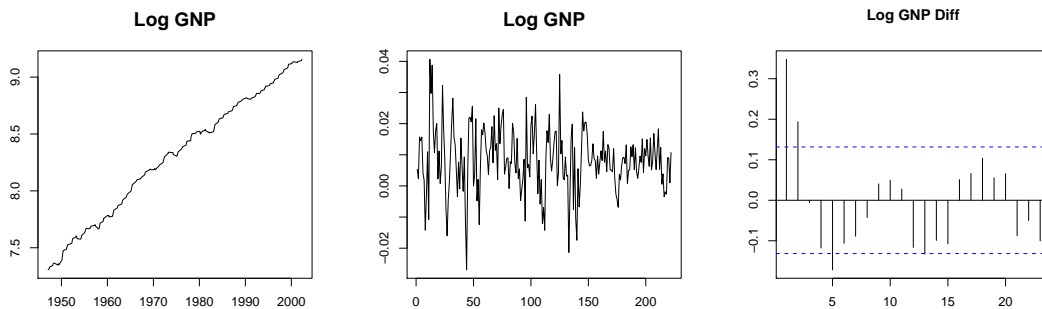
Im Datensatz `gnp` ist im wesentlichen der Trend zu entfernen. Hier bietet sich insbesondere

Abbildung 11: Tempdub: Entfernung der Saisonkomponente über sinusoiden Term



Differenzenbildung an. Man kann dies direkt auf die Beobachtungen anwenden, wir wollen jedoch zunächst zu den Logarithmen übergehen, da der Trend im GNP auch weniger linear als vielmehr exponentiell zu sein scheint. Die logarithmierten Werte, deren erste Differenzen mit Autokorrelation sind in Abb. 12 geplottet. Wir modellieren also die log-returns oder exponentielle Wachstumsrate.

Abbildung 12: GNP: Modellierung der exponentiellen Wachstumsrate



2 Autoregressive moving average Modelle

2.1 Definition und grundlegende Eigenschaften von ARMA Modellen

Definition 2.1

Eine zentrierte, stationäre, reelle Zeitreihe $(X_t)_{t \in \mathbb{Z}}$ folgt einem *autoregressive moving average Modell* mit *autoregressiver Ordnung* p und *moving average Ordnung* q , ARMA(p,q), falls Konstanten $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q \in \mathbb{R}$, $\phi_p \neq 0$, $\theta_q \neq 0$ sowie ein weißes Rauschen $(e_t)_{t \in \mathbb{Z}}$ existieren, so dass

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q}, \quad \forall t \in \mathbb{Z}. \quad (25)$$

Bemerkung 2.2

- Man nimmt gewöhnlich weiter an, dass $(e_t)_{t \in \mathbb{Z}}$ u.i.v. nach $N(0, \sigma_e^2)$ verteilt sind.
- Für nicht zentrierte X_t , etwa $EX_t = \mu$, ersetzt man in (25) X_s durch $X_s - \mu$, $s = t-p, \dots, p$.

Definition 2.3

- Ein ARMA(p,q) Modell (25) mit $q = 0$ heißt *autoregressives Modell* der Ordnung p .
- Ein ARMA(p,q) Modell (25) mit $p = 0$ heißt *moving average Modell* der Ordnung q .
- TERMINOLOGIE ARMA GLEICHUNG; NUR BEZIEHUNG ZWISCHEN WEIßEN RAUSCHEN UND ZEITREIHE X

Definition 2.4

Das AR-Polynom $\phi(z)$ sowie das MA Polynom $\theta(z)$ eines ARMA(p,q) Modells (25) sind definiert durch

$$\begin{aligned} \phi(z) &= \phi_0 - \phi_1 z - \dots - \phi_p z^p, \\ \theta(z) &= \theta_0 + \theta_1 z + \dots + \theta_q z^q \end{aligned}$$

wobei $\phi_0 = \theta_0 = 1$.

Definition 2.5

Der Lag Operator B operiert auf einer stationären Zeitreihe (X_t) durch $B^k X_t = X_{t-k}$. Für ein Polynom $a(z) = a_0 + a_1 z + \dots + a_r z^r$ setze $a(B) = a_0 + a_1 B + \dots + a_r B^r$, so dass $a(B)X_t = a_0 + a_1 X_t + \dots + a_r X_{t-r}$.

Der autoregressive Operator eines ARMA(p,q) Modells (25) ist gegeben durch $\phi(B)$, der moving average Operator durch $\theta(B)$, wobei $\phi(z)$ und $\theta(z)$ die autoregressiven und moving average Polynome sind. Es gilt also $\phi(B)X_t = \theta(B)e_t$, $t \in \mathbb{Z}$.

Beispiel 2.6 (AR(1))

ÜBERARBEITEN!!!

Wir betrachten $X_t = \phi X_{t-1} + e_t$, wobei wir $\phi \in \mathbb{R}$ annehmen.

1. Fall: $|\phi| < 1$.

Man erhält durch Induktion für $k \geq 1$

$$\begin{aligned} X_t &= \phi(\phi X_{t-2} + e_{t-1}) + e_t \\ &= \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j e_{t-j}. \end{aligned}$$

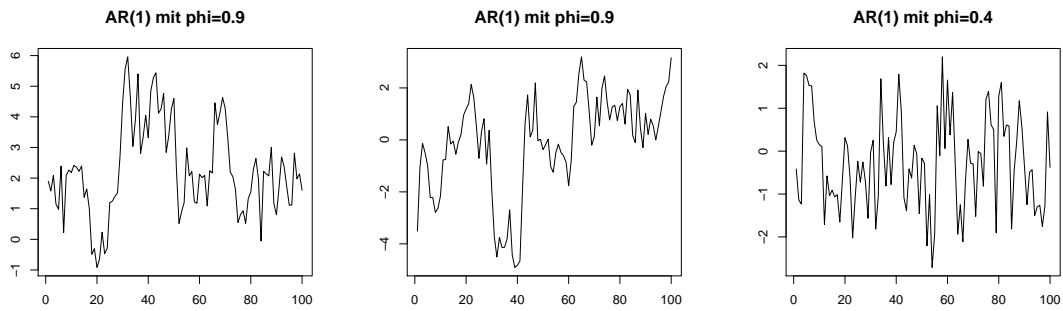


Abbildung 13: Simulations AR(1) Prozess für positive Werte von ϕ

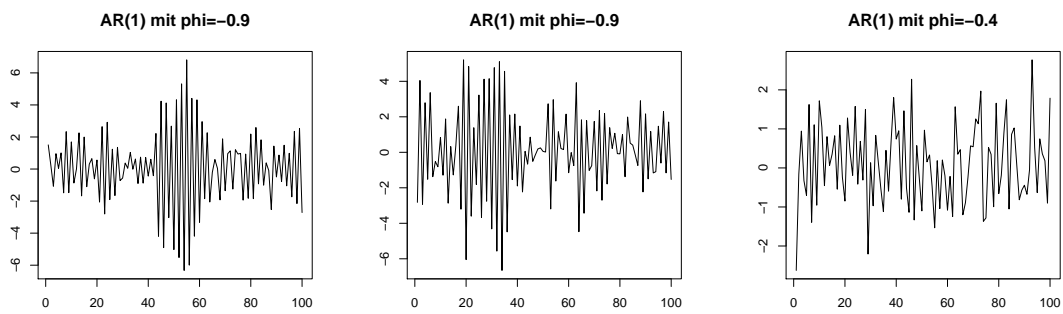


Abbildung 14: Simulations AR(1) Prozess für negative Werte von ϕ

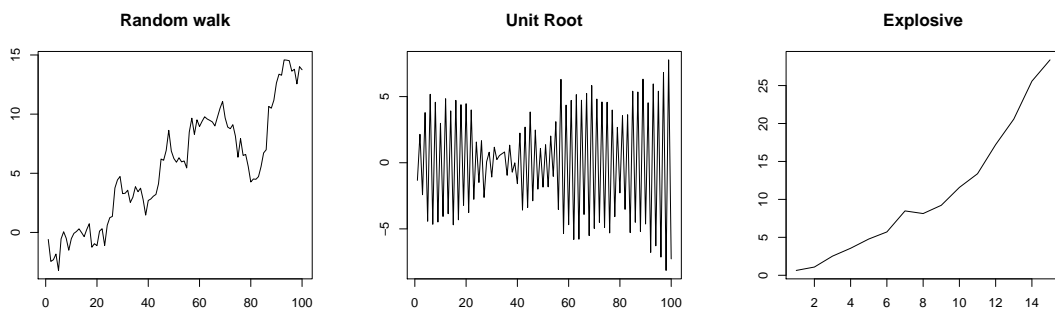


Abbildung 15: Nicht-stationäre Fälle und ein explosives AR(1) Modell

Ist X_t stationär, so folgt wegen $|\phi| < 1$

$$X_t = \sum_{j=0}^{\infty} \phi^j e_{t-j} \quad (26)$$

Umgekehrt definiert (26) eine stationäre Lösung der AR(1) Gleichung, die also für gegebenes Rauschen eindeutig bestimmt ist.

2. Fall: $\phi = 1$

Dann ist $X_t = X_{t-1} + e_t$ ein random walk, also nicht stationär und somit kein AR(1) Modell. Ebenso für $\phi = -1$. FORMAL ZEIGEN!!!!

3. Fall: $|\phi| > 1$

Die Darstellung (26) gilt offenbar nicht. Man kann jedoch vorwärts entwickeln und erhält

$$\begin{aligned} X_t &= \phi^{-1} X_{t+1} - \phi^{-1} e_{t+1} \\ &= \phi^{-k} X_{t+k} + \sum_{j=1}^k \phi^{-j} e_{t+j}. \end{aligned}$$

Wegen $|\phi| > 1$ erhält man

$$X_t = \sum_{j=1}^{\infty} \phi^{-j} e_{t+j} \quad (27)$$

als eindeutige stationäre Lösung der AR(1) Gleichung. Hier hängt aber X_t von den zukünftigen Fehlern e_{t+j} ab, was unerwünscht ist. Solche AR(1) Prozesse heißen auch *explosiv* (explosive). Falls eine Darstellung (26) existiert, spricht man dagegen von einem *kausalen* AR(1) Modell. UMSCHREIBEN DES EXPLOSIVEN IN EIN KAUSALES AR(1) MODELL; KAUSAL ALSO NUR IM ZUSAMMENHANG DER BEIDEN PROZESSE ZU SEHEN:

In den Abbildungen 15 und 15 sind einige Pfade von kausalen AR(1) Prozessen mit verschiedenen Werten von ϕ simuliert (mit standard-normalverteilten weißen Rauschen). Zum Vergleich enthält Abb. 15 zwei nicht-stationäre Fälle, den random walk $X_t = X_{t-1} + e_t$ und den unit root Fall $X_t = -X_{t-1} + e_t$ sowie ein explosives AR(1) Modell $X_t = 1.2X_{t-1} + e_t$.

Die Pfade der kausalen AR(1) Modelle wurden mit `arma.sim` generiert.

Beispiel 2.7 (MA(1))

Hier sei $X_t = e_t + \theta e_{t-1}$, $\theta \in \mathbb{R}$. Dann ist $EX_t = 0$,

$$\gamma(h) = \text{Cov}(X_{t+h}, X_t) = \begin{cases} \sigma_e^2(1 + \theta^2), & h = 0, \\ \sigma_e^2\theta, & h = 1, \\ 0, & h \geq 2. \end{cases}$$

und

$$\rho(h) = \text{Corr}(X_{t+h}, X_t) = \begin{cases} \frac{\theta}{(1+\theta^2)}, & h = 1, \\ 0, & h \geq 2. \end{cases}$$

Daher ist die Korrelation für θ und $1/\theta$ gleich. Weiter liefern die Parameterpaare (θ, σ_e^2) und $(\tilde{\theta}, \tilde{\sigma}_e^2) = (1/\theta, \theta^2\sigma_e^2)$ die gleiche Kovarianzfunktion. Somit sind die Parameter nur unter

zusätzlichen Restriktionen eindeutig bestimmt. Man wählt die $MA(1)$ Darstellung, für die man auflösen kann

$$e_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j} \quad (28)$$

welches für $|\theta| < 1$ möglich ist (in Analogie zur Kausalität bei $AR(1)$ Modellen). Solche $MA(1)$ Modelle heißen *invertibel*.

Beispiel 2.8 (Parameter Redundanz)

Falls $X_t = e_t$ ein weißes Rauschen ist, erfüllt es auch z.B. die $ARMA(1,1)$ Gleichung $X_t = 0,5X_{t-1} + e_t - 0,5e_{t-1}$. Hierbei liegt der Fall $\phi(z) = \theta(z)$ vor, insbesondere haben die beiden Polynome eine gemeinsame Nullstelle.

Definition 2.9

Die $ARMA(p,q)$ Gleichung $\phi(B)X_t = \theta(B)e_t$ heißt *reduziert*, falls $\phi(z)$ und $\theta(z)$ keine (möglicher Weise komplexe) gemeinsame Nullstelle haben.

Satz 2.10 (Existenzsatz für ARMA Modelle)

Zu einem gegebenen weißen Rauschen $(e_t)_{t \in \mathbb{Z}}$ existiert genau dann eine, dann eindeutig bestimmte, stationäre, zentrierte, reelle Lösung $(X_t)_{t \in \mathbb{Z}}$ einer reduzierten $ARMA(p,q)$ Gleichung $\phi(B)X_t = \theta(B)e_t$, falls $\phi(z) \neq 0$ für alle $|z| = 1$.

Definition 2.11 (Kausalität)

Ein reduziertes $ARMA(p,q)$ (25) Modell heißt *kausal*, falls eine stationäre, zentrierte Lösung $(X_t)_{t \in \mathbb{Z}}$ der Form

$$X_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}, \quad \sum_{j=0}^{\infty} |\psi_j| < \infty, \quad (29)$$

mit $\psi_j \in \mathbb{R}$ existiert, wobei $\psi_0 = 1$.

Satz 2.12 (Kriterium für Kausalität)

Ein reduziertes $ARMA(p,q)$ Modell $\phi(B)X_t = \theta(B)e_t$ ist genau dann kausal, falls $\phi(z) \neq 0$ für alle $|z| \leq 1$. Die Koeffizienten in (29) können aus der Taylor Reihe

$$\frac{\theta(z)}{\phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j \quad (30)$$

bestimmt werden, die für $|z| \leq \rho$ für ein $\rho > 1$ konvergiert.

Bemerkung 2.13

a. Die Reihe (29) ist eine einseitige $MA(\infty)$ Darstellung für die kausale $ARMA(p,q)$ Reihe. Da (30) für $z = \rho > 1$ konvergiert, gilt $|\psi_j| \rho^j \rightarrow 0, j \rightarrow \infty$, also für ein $C > 0$

$$|\psi_j| \leq C \rho^{-j}, j \geq 0, \quad (31)$$

insbesondere gilt also die Summierbarkeitsbedingung (9). Für u.i.v weißes Rauschen ist eine kausale $ARMA(p,q)$ Reihe also eine kausale lineare Zeitreihe, für Gaußsches weißes Rauschen insbesondere dann auch Gaußsch.

b. Wegen $\theta(z) = \phi(z) \sum_{j=0}^{\infty} \psi_j z^j$ kann man die Koeffizienten ψ_j rekursiv berechnen über

$$\psi_j - \sum_{k=1}^{\min(j,p)} \psi_{j-k} \phi_k = \begin{cases} \theta_j, & k \leq q, \\ 0, & k \geq q+1, \end{cases} \quad (32)$$

Über die Theorie der Differenzgleichungen kann man auch die Struktur der ψ_j aus (35) bestimmen, wir werden dies für die Autokorrelationen, die eine ähnliche Differenzgleichung erfüllen, im nächsten Abschnitt durchführen.

Definition 2.14 (Invertibilität)

Ein reduziertes ARMA(p,q) (25) Modell mit $\phi(z) \neq 0$ für $|z| = 1$ heißt *invertibel*, falls die stationäre, zentrierte Lösung $(X_t)_{t \in \mathbb{Z}}$ die Darstellung

$$e_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad \sum_{j=0}^{\infty} |\pi_j| < \infty, \quad (33)$$

mit $\pi_j \in \mathbb{R}$ erlaubt, wobei $\psi_0 = 1$.

Satz 2.15 (Kriterium für Invertibilität)

Ein reduziertes ARMA(p,q) Modell $\phi(B)X_t = \theta(B)e_t$ ist genau dann invertibel, falls $\phi(z) \neq 0$ für $|z| = 1$ sowie $\theta(z) \neq 0$ für alle $|z| \leq 1$. Die Koeffizienten in (33) können aus der Taylor Entwicklung

$$\frac{\phi(z)}{\theta(z)} = \sum_{j=0}^{\infty} \pi_j z^j$$

bestimmt werden, die für $|z| \leq \rho$ für ein $\rho > 1$ konvergiert.

Satz 2.16 (Identifizierbarkeit der Parameter)

Innerhalb der Klasse der reduzierten, kausalen, invertiblen ARMA Modelle sind die Ordnungen p und q , die Koeffizienten ϕ_1, \dots, ϕ_p und $\theta_1, \dots, \theta_q$ sowie (wegen der Invertibilität) auch die Varianz des weißen Rauschens σ_e^2 eindeutig bestimmt.

Dies bedeutet konkret: Sei (X_t) eine stationäre, zentrierte Zeitreihe, die einer reduzierten, kausalen, invertiblen ARMA(p,q) Gleichung genügt mit einem weißen Rauschen (e_t) . Dann erfüllt (X_t) keine andere reduzierte, kausale, invertible ARMA(p*,q*) Gleichung mit einem, möglicher Weise anderen weißen Rauschen (e_t^*) .

Die Beweise werden wir innerhalb der Spektraltheorie führen.

2.2 Autokovarianz und Autokorrelation

Im Folgenden sei $(X_t)_{t \in \mathbb{Z}}$ ein reeller, kausaler ARMA(p,q) Prozess. Nach der Darstellung (29) ist (X_t) ein einseitiger MA(∞) Prozess, also

$$\gamma(h) = EX_{t+h}X_t = \sum_{j,k=0}^{\infty} \psi_j \psi_k E(e_{t+h-j}e_{t-k}) = \sigma_e^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}. \quad (34)$$

Beispiel 2.17 (Autokorrelation eines MA(q) Modells)

Für ein MA(q) Modell $X_t = e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q}$ ist mit $\theta_0 = 1$

$$\gamma(h) = \begin{cases} \sigma_e^2 \sum_{k=0}^{q-h} \theta_k \theta_{h+k} & 0 \leq h \leq q, \\ 0, & h \geq q+1, \end{cases} \quad (35)$$

Beispiel 2.18 (ARMA(1,1))

Seien $\phi, \theta \in \mathbb{R}$, $\phi \neq -\theta$, $|\phi| < 1$, und $X_t = \phi X_{t-1} + e_t + \theta e_{t-1}$. Wir bestimmen zunächst die Koeffizienten in (29) mit Hilfe von (30). Es ist für $|z| < 1/|\phi|$

$$\begin{aligned} \frac{\theta(z)}{\phi(z)} &= (1 + \theta z) \sum_{j=0}^{\infty} (\phi z)^j \\ &= 1 + \sum_{j=0}^{\infty} (\phi + \theta) \phi^{j-1} z^j, \end{aligned}$$

somit ($\psi_0 = 1$) $\psi_j = (\phi + \theta) \phi^{j-1}$, $j \geq 1$. Mit (34) erhält man

$$\begin{aligned} \gamma(0) &= \sigma_e^2 \sum_{j=0}^{\infty} \psi_j^2 \\ &= \sigma_e^2 (1 + (\phi + \theta)^2 \sum_{j=0}^{\infty} (\phi^2)^j) \\ &= \sigma_e^2 \left(1 + \frac{(\phi + \theta)^2}{1 - \phi^2}\right) \end{aligned}$$

sowie

$$\begin{aligned} \gamma(h) &= \sigma_e^2 \left((\phi + \theta) \phi^{h-1} + (\phi + \theta)^2 \sum_{j=0}^{\infty} \phi^{j-1} \phi^{j+h-1} \right) \\ &= \sigma_e^2 \left((\phi + \theta) \phi^{h-1} + \frac{\phi^h (\phi + \theta)^2}{1 - \phi^2} \right) \\ &= \sigma_e^2 \phi^{h-1} \frac{(\phi + \theta)(1 + \phi\theta)}{1 - \phi^2}. \end{aligned}$$

Satz 2.19

Für die Autokovarianz eines kausalen ARMA(p,q) Prozesses gilt $|\gamma(h)| \leq C \rho^{-|h|}$, $h \in \mathbb{Z}$, für ein $\rho > 1$.

Beweis

Aus (34) folgt mit (31) (also $|\psi_j| \leq c\rho^{-j}$, $j \geq 0$), dass

$$|\gamma(h)| \leq c\rho^{-h} \frac{\sigma^2}{1 - \rho^{-2}}, \quad h \geq 0. \quad \blacksquare$$

Differenzgleichung für die Kovarianzfunktion

Es ist mit $\psi_0 = \theta_0 = 1$

$$\begin{aligned} \gamma(h) = E(X_{t+h}X_t) &= E\left(\sum_{j=1}^p \phi_j X_{t+h-j} X_t + \sum_{k=0}^q \theta_k e_{t+h-k} X_t\right) \\ &= \sum_{j=1}^p \phi_j \gamma(h-j) + \begin{cases} \sigma_e^2 \sum_{k=h}^q \theta_k \psi_{k-h}, & h \leq q, \\ 0, & h \geq q+1, \end{cases} \end{aligned} \quad (36)$$

wobei

$$E(e_{t+h-k} X_t) = E\left(\sum_{j=0}^{\infty} e_{t-j} \psi_j e_{t+h-k}\right) = \begin{cases} \psi_{k-h} \sigma_e^2, & k \geq h, \\ 0, & k < h \end{cases}.$$

Man kann nun (36) sukzessive lösen, nachdem man die Koeffizienten ψ_1, \dots, ψ_q etwa mit (18) bestimmt hat. Wir wollen die Struktur mit Hilfe der Theorie der Differenzgleichungen untersuchen.

Lemma 2.20

Seien $d_1, \dots, d_p \in \mathbb{C}$. Der p -dimensionale Raum der komplexen Lösungen $(a_k)_{k \geq 0}$ der homogenen Differenzgleichung

$$a_k + \sum_{j=1}^p d_j a_{k-j} = 0, \quad k = p, p+1, \dots \quad (37)$$

ist gegeben durch Lösungen der Form

$$a_k = \sum_{i=1}^l \sum_{n=0}^{r_i} c_{i,n} k^n z_i^{-k}, \quad c_{i,n} \in \mathbb{C}, \quad (38)$$

wobei z_1, \dots, z_l die (komplexen) Nullstellen des Polynoms $1 + d_1 z + \dots + d_p z^p$ sind mit den Vielfachheiten r_1, \dots, r_l .

Sind beliebige Anfangsglieder a_0, \dots, a_{p-1} gegeben, so existiert offenbar genau eine Lösungsfolge von (37) (Rekursion), und daher sind dann auch die Koeffizienten $c_{i,n}$ in (38) eindeutig bestimmt.

Für reelle Koeffizienten d_1, \dots, d_p ist man besonders an reellen Lösungen interessiert. Ist in diesem Fall z_i eine reelle Nullstelle, so erhält man in (38) Summanden der Form $c_{i,n} k^n z_i^{-k}$, wobei $c_{i,n} \in \mathbb{R}$. Für zwei komplex konjugierte Nullstellen z und \bar{z} wählt man die Koeffizienten c_n ebenfalls komplex konjugiert, und kommt mit Polarkoordinaten auf Summanden der Form

$$a k^n |z|^{-k} \cos(k\phi + \omega), \quad a > 0, \quad 0 \leq \omega < 2\pi,$$

wobei $z = |z|e^{i\phi}$ und $c_n = ae^{-i\omega}/2$.

Wir wollen dies nun auf die Differenzgleichung (36) der Autokovarianzen anwenden. Hier haben wir die homogene Gleichung

$$\gamma(h) - \sum_{j=1}^p \phi_j \gamma(h-j) = 0, \quad h \geq \max(p, q+1).$$

(wir benötigen $h \geq \max(p, q+1)$, damit stets $h-p \geq 0$). Für $s = \max(p-1, q)$ sind also durch $\gamma(s-p+1), \dots, \gamma(s)$ alle weiteren Kovarianzen eindeutig bestimmt, und können in der aus (38) abgeleiteten reellen Form dargestellt werden. Insbesondere erhält man sinusoidale Terme, falls das AR-Polynom Paare von komplex-konjugierten Nullstellen hat. Die Werte $\gamma(0), \dots, \gamma(s)$ erhält man aus den ersten s Gleichung in (36).

Yule-Walker Gleichungen

Für ein AR(p) Modell mit Parameter Vektor $\phi = (\phi_1, \dots, \phi_p)^T$ setze für $p_0 \geq p$ $\phi(p_0) = (\phi_1, \dots, \phi_p, 0, \dots, 0)^T \in \mathbb{R}^{p_0}$. Die Gleichung (36) liefert dann mit der Matrixnotation aus (18) die sogenannten Yule-Walker Gleichungen

$$\begin{aligned} \Gamma_{p_0} \phi(p_0) &= \gamma_{p_0}^1, & \gamma_{p_0}^1 &= (\gamma(1), \dots, \gamma(p_0))^T \\ \sigma_e^2 &= \gamma(0) - \sum_{j=1}^p \phi_j \gamma(j). \end{aligned} \quad (39)$$

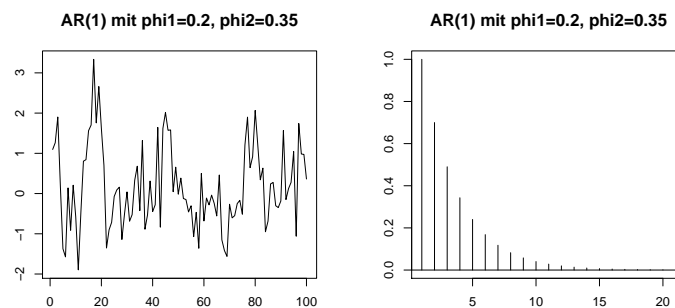


Abbildung 16: Simulation und ACF AR(2) Prozess bei zwei reellen Nullstellen

Beispiel 2.21

Wir betrachten AR(2) Prozesse.

1. Fall: zwei reelle Nullstellen. Wir betrachten konkret die Nullstellen $a = -0.5$, $b = 0.7$, also $\phi_1 = 0.2$, $\phi_2 = 0.35$. Abb. 16 enthält die theoretische ACF (generiert mit `ARMAacf`) sowie einen simulierten Pfad. Man erkennt wenig Unterschied zu einem AR(1) Prozess.

2. Fall: zwei komplex-konjugierte Nullstellen. Wir betrachten $(0.7 + / - i0.6)$. Abb. 17 enthält die theoretische ACF (generiert mit `ARMAacf`) sowie einen simulierten Pfad. Man erkennt ein zyklisches Verhalten des Prozesses.

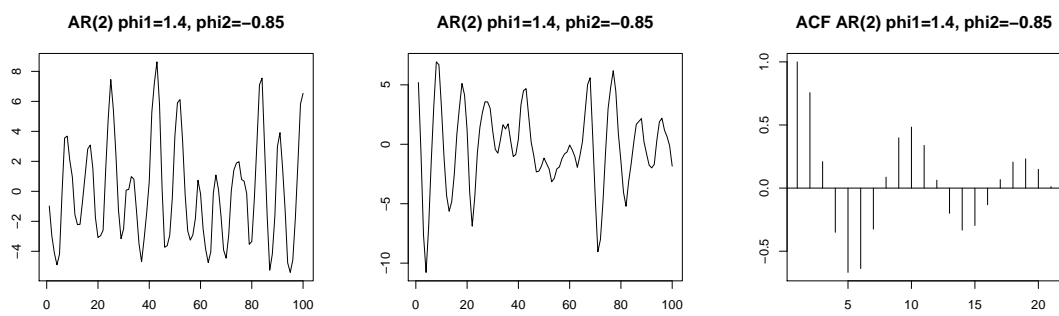


Abbildung 17: Simulation und ACF AR(2) Prozess bei zwei komplex-konjugierten Nullstellen

2.3 Die partielle Autokorrelation

Es sei $(X_t)_{t \in \mathbb{Z}}$ ein zentrierter, stationärer, reellwertiger Prozess. Setze für $h \geq 1$

$$H_h = \text{span} \{X_1, \dots, X_h\} \subset L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P),$$

sowie $H_0 = \{0\}$. Weiter sei P_{H_h} die orthogonale Projektion auf H_h , diese existiert, da H_h endlichdimensional und somit insbesondere abgeschlossen ist. Wir bezeichnen die Norm in $L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P)$ mit $\|\cdot\|$.

Definition 2.22

Die *partielle Autokorrelation* $\pi(h)$ zum Lag $h \geq 1$ ist definiert durch

$$\pi(h) = \text{Corr} \left((X_h - P_{H_{h-1}} X_h), (X_0 - P_{H_{h-1}} X_0) \right).$$

Bemerkung 2.23

- Für $h = 1$ ist $\pi(1) = \text{Corr}(X_1, X_0) = \rho(1)$.
- Man kann in den Definitionen von H_h und $\pi(h)$ auch alle Indices um t verschieben, wegen der Stationarität ist dann das resultierende $\pi(h)$ unabhängig von t .
- Die partielle Korrelation wird insbesondere zur Ordnungswahl in AR-Modellen benutzt, s. Bemerkung 2.26.

Wir gehen nun näher auf die orthogonale Projektion P_{H_h} ein. Diese ist bestimmt durch die Gleichungen

$$E \left((P_{H_h} X_{h+k} - X_{h+k}) X_{h+1-s} \right) = 0, \quad s = 1, \dots, h. \quad (40)$$

Für $h \geq 1$, $k \in \mathbb{Z}$, ist für Koeffizienten $a_1^{h,k}, \dots, a_h^{h,k}$

$$P_{H_h} X_{h+k} = a_1^{h,k} X_h + \dots + a_h^{h,k} X_1 \quad (41)$$

Einsetzen in (40) liefert

$$\sum_{j=1}^h a_j^{h,k} E X_{h+1-j} X_{h+1-s} = E X_{h+k} X_{h+1-s}, \quad s = 1, \dots, h,$$

bzw.

$$\sum_{j=1}^h a_j^{h,k} \gamma(s-j) = \gamma(k+s-1), \quad s = 1, \dots, h. \quad (42)$$

Schreibe in Matrixschreibweise mit

$$\begin{aligned} \boldsymbol{\gamma}_h^k &= (\gamma(k), \dots, \gamma(k+h-1))^T, \\ \mathbf{a}^{h,k} &= (a_1^{h,k}, \dots, a_h^{h,k})^T \end{aligned} \quad (43)$$

(42) um zu

$$\Gamma_h \mathbf{a}^{h,k} = \boldsymbol{\gamma}_h^k. \quad (44)$$

Unter den Voraussetzungen von Satz 1.25, insbesondere also für kausale ARMA(p,q) Modelle, ist die Lösung von (44) stets eindeutig bestimmt.

Satz 2.24

Für einen zentrierten, stationären, reellwertigen Prozess $(X_t)_{t \in \mathbb{Z}}$ sei Γ_h stets nicht ausgeartet für alle $h \geq 0$. Dann ist die partielle Autokorrelation bestimmt durch die Koeffizienten in der Darstellung (44), genauer gilt

$$\pi(h) = a_h^{h,1}.$$

Beweis

Zunächst zeigen wir den Zusammenhang

$$a_j^{h,-h} = a_{h+1-j}^{h,1}, \quad j = 1, \dots, h. \quad (45)$$

Die Gleichung (42) liefert für die Koeffizienten $a_j^{h,-h}$

$$\sum_{j=1}^h a_j^{h,-h} \gamma(s-j) = \gamma(s-h-1), \quad s = 1, \dots, h.$$

Mit $\gamma(h) = \gamma(-h)$ und Übergang zu $t = h+1-s$ sowie $i = h+1-j$ schreibt man dies äquivalent um zu

$$\sum_{i=1}^h a_{h+1-i}^{h,-h} \gamma(t-i) = \gamma(t), \quad t = 1, \dots, h.$$

Ein Vergleich mit (42) liefert dann sofort (45). Wegen $(\Gamma_{h-1})_{j,k} = (\Gamma_{h-1})_{h-j, h-k}$ erhält man mit (45)

$$\|P_{H_{h-1}} X_h\|^2 = (\mathbf{a}^{h-1,1})^T \Gamma_{h-1} \mathbf{a}^{h-1,1} = (\mathbf{a}^{h-1,-h+1})^T \Gamma_{h-1} \mathbf{a}^{h-1,-h+1} = \|P_{H_{h-1}} X_0\|^2$$

und somit auch

$$\|X_h - P_{H_{h-1}} X_h\|^2 = \|X_h\|^2 - \|P_{H_{h-1}} X_h\|^2 = \|X_0\|^2 - \|P_{H_{h-1}} X_0\|^2 = \|X_0 - P_{H_{h-1}} X_0\|^2. \quad (46)$$

Wegen $E(P_{H_{h-1}} X_h (X_0 - P_{H_{h-1}} X_0)) = 0$ ist somit

$$\pi(h) = \frac{E(X_h (X_0 - P_{H_{h-1}} X_0))}{\|X_0 - P_{H_{h-1}} X_0\|^2}. \quad (47)$$

Setze nun $H = \text{span}\{X_0, \dots, X_{h-1}\}$, dann gilt für die orthogonale Projektion P_H auf H wegen Stationarität (s. Bemerkung 2.23 b., man kann dies auch durch Zurückführen auf die Gleichungen (42) nachrechnen) auch

$$P_H X_h = a_1^{h,1} X_{h-1} + \dots + a_h^{h,1} X_0. \quad (48)$$

Weiter ist $H = H_{h-1} \oplus \text{span}(X_0 - P_{H_{h-1}} X_0)$ (orthogonale Summe), und mit (47) erhalten wir

$$\begin{aligned} P_H X_h &= P_{H_{h-1}} X_h + \pi(h) (X_0 - P_{H_{h-1}} X_0) \\ &= P_{H_{h-1}} X_h - \pi(h) P_{H_{h-1}} X_0 + \pi(h) X_0. \end{aligned} \quad (49)$$

Da $P_{H_{h-1}} X_h - \pi(h) P_{H_{h-1}} X_0 \in H_{h-1}$, liefert ein Koeffizientenvergleich von (49) und (48) (mit der Eindeutigkeit der Koeffizienten) die Behauptung. ■

Korollar 2.25 (Levinson Rekursion)

Setze $\Delta_h(1) = \|X_{h+1} - P_{H_h} X_{h+1}\|^2$, der quadratische Einschnitt Vorhersagefehler. Dann gilt unter den Voraussetzungen von Satz 2.24 mit $a_1^{h,1} = \gamma(1)/\gamma(0)$, $\Delta_0(1) = \gamma(0)$, die Levinson Rekursion

$$\begin{aligned} a_h^{h,1} &= \frac{\gamma(h) - \sum_{j=1}^{h-1} a_j^{h-1,1} \gamma(h-j)}{\Delta_{h-1}(1)} \\ a_j^{h,1} &= a_j^{h-1,1} - a_h^{h,1} a_{h-j}^{h-1,1}, \\ \Delta_h(1) &= \Delta_{h-1}(1)(1 - (a_h^{h,1})^2). \end{aligned}$$

Beweis

Für die erste Gleichung benutzen wir $\pi(h) = a_h^{h,1}$ und (47). Der Nenner in (47) ist wegen (46) gerade $\Delta_{h-1}(1)^2$. Für den Zähler haben wir mit (45)

$$\begin{aligned} E(X_h(X_0 - P_{H_{h-1}} X_0)) &= EX_h X_0 - \sum_{i=1}^{h-1} a_i^{h-1, -h+1} EX_{h-i} X_h \\ &= \gamma(h) - \sum_{j=1}^{h-1} a_{h-j}^{h-1,1} \gamma(h-j), \end{aligned}$$

also die erste Gleichung.

Die zweite Gleichung ergibt sich durch Koeffizientenvergleich von (48) und (49), wobei in (49) die Darstellung

$$P_{H_{h-1}} X_0 = a_{h-1}^{h-1,1} X_{h-1} + \dots + a_1^{h-1,1} X_1 \quad \blacksquare$$

verwendet wird.

Die dritte Gleichung ergibt sich aus (46) mit der Notation des Beweises des Satzes:

$$\begin{aligned} \Delta_h(1) &= \|X_h - P_H X_h\|^2 \\ &= \|X_h - P_{H_{h-1}} X_h - \pi(h)(X_0 - P_{H_{h-1}} X_0)\|^2 \\ &= \|X_h - P_{H_{h-1}} X_h\|^2 + \pi(h)^2 \|X_0 - P_{H_{h-1}} X_0\|^2 - 2\pi(h)E(X_h(X_0 - P_{H_{h-1}} X_0)) \\ &= \Delta_{h-1}(1) + (a_h^{h,1})^2 \Delta_{h-1}(1) - 2(a_h^{h,1})^2 \Delta_{h-1}(1), \end{aligned}$$

wobei (46) und (47) benutzt wurden.

Bemerkung 2.26 (PACF und Yule Walker Gleichungen für AR(p))

Ein Vergleich der Yule Walker Gleichung (39) mit (44) liefert für einen AR(p) Prozess für den Koeffizientenvektor $\mathbf{a}^{h,1}$ für $h \geq p$

$$\mathbf{a}^{h,1} = \phi(h).$$

Insbesondere ist mit Satz 2.24 die partielle Autokorrelation $\pi(h) = 0$ für $h \geq p+1$, und $\pi(p) = \phi_p$.

Empirische partielle Autokorrelation

Ersetzt man in (44) die theoretische Autokorrelation durch die Schätzwerte $\hat{\gamma}(h)$, so liefert

dies die *empirische partielle Autokorrelation* $\hat{\pi}(h)$. Die asymptotische Verteilung der einzelnen empirischen partiellen Autokorrelationen $\hat{\pi}(h)$ ergibt sich dann im Prinzip aus Satz 1.28 und der Konsistenz des Schätzers $\hat{\Gamma}_h$. Diese ist aber im allgemeinen zu kompliziert, um nützlich zu sein. Man erhält aber für einen AR(p)-Prozess für $h > p$ für $\sqrt{n}\hat{\pi}(h)$ asymptotisch eine Standardnormalverteilung, s. Satz 2.37.

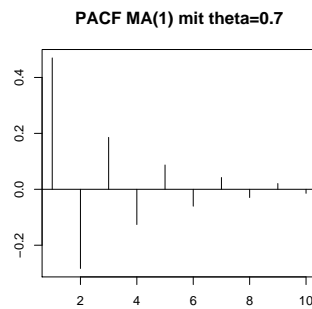


Abbildung 18: PACF eines MA(1) Prozesse mit $\theta = 0.7$

Beispiel 2.27

Wir betrachten ein MA(1) Modell mit $\theta = 0.7$. Die partielle Autokorrelation ist in Abb. 18 geplottet. Diese fällt exponentiell schnell ab, für die explizite Formel s. etwa Brockwell und Davis (1991).

2.4 Vorhersage

Vorhersage bei endlicher Vergangenheit

Sei $(X_t)_{t \in \mathbb{Z}}$ ein stationärer, reellwertiger, zentrierter Prozess. Wir setzen

$$\mathcal{M}_t = \{f(X_1, \dots, X_t) \in L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P), \quad f: \mathbb{R}^t \rightarrow \mathbb{R}\}.$$

Lemma 2.28

\mathcal{M}_t ist ein linearer, abgeschlossener Unterraum von $L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P)$.

Beweis

Die Messbarkeit überträgt sich auf L_2 -Limiten, da jede L_2 -konvergente Folge eine fast-sicher konvergente Teilfolge (mit demselben Limes) besitzt, und Messbarkeit sich bei fast-sicheren Limiten überträgt. ■

Definition 2.29

Die *beste Vorhersage* von X_{t+h} aus X_1, \dots, X_t , $t, h \geq 1$, ist die orthogonale Projektion $P_{\mathcal{M}_t} X_{t+h}$ von X_{t+h} auf \mathcal{M}_t . Der Vorhersagefehler ist gegeben durch $E(X_{t+h} - P_{\mathcal{M}_t} X_{t+h})$.

Bemerkung 2.30

Es gilt $P_{\mathcal{M}_t} X_{t+h} = E(X_{t+h} | X_1, \dots, X_t)$.

Definition 2.31

Die *beste lineare Vorhersage* \hat{X}_{t+h} von X_{t+h} aus X_1, \dots, X_t , $t, h \geq 1$, ist die orthogonale Projektion P_{H_t} von X_{t+h} auf $H_t = \text{span}\{X_1, \dots, X_t\}$. Der lineare Vorhersagefehler ist gegeben durch

$$\Delta_t(h) = E(X_{t+h} - \hat{X}_{t+h})^2.$$

Bemerkung 2.32

Da $H_t \subset \mathcal{M}_t$, gilt stets

$$E(X_{t+h} - P_{\mathcal{M}_t} X_{t+h})^2 \leq E(X_{t+h} - \hat{X}_{t+h})^2.$$

Man benutzt dennoch in der Regel die beste lineare Vorhersage, da

- i. diese leichter zu berechnen ist,
- ii. diese nur von der Kovarianzfunktion des Prozesses abhängt und nicht von der vollständigen Verteilung,
- iii. falls $(X_t)_{t \in \mathbb{Z}}$ ein Gaußprozess ist, so ist $\hat{X}_{t+h} = P_{\mathcal{M}_t} X_{t+h}$.

Beweis von iii. $P_{\mathcal{M}_t} X_{t+h}$ ist eindeutig bestimmt durch

$$E((X_{t+h} - P_{\mathcal{M}_t} X_{t+h})W) = 0, \quad W \in \mathcal{M}_t.$$

Zeige, dass \hat{X}_{t+h} dies erfüllt: Zunächst ist

$$E((X_{t+h} - \hat{X}_{t+h})X_j) = 0, \quad j = 1, \dots, t.$$

Also sind $(X_{t+h} - \hat{X}_{t+h})$ und X_j unkorreliert, $j = 1, \dots, t$, und da $((X_{t+h} - \hat{X}_{t+h}), X_1, \dots, X_t)$ gemeinsam normalverteilt sind, sind $(X_{t+h} - \hat{X}_{t+h})$ und X_1, \dots, X_t auch unabhängig. Daher ergibt sich für $W \in \mathcal{M}_t$

$$E((X_{t+h} - \hat{X}_{t+h})W) = E(X_{t+h} - \hat{X}_{t+h})EW = 0,$$

da (X_t) zentriert ist und somit $E(X_{t+h} - \hat{X}_{t+h}) = 0$.

Bemerkung 2.33

Falls $\mu \neq 0$.

Berechnung von \hat{X}_{t+h} und $\Delta_t(h)$

1. In der Notation von (41) ist

$$P_{H_t} X_{t+h} = a_1^{t,h} X_t + \dots + a_t^{t,h} X_1.$$

Der Koeffizientenvektor $\mathbf{a}^{t,h}$ und somit \hat{X}_{t+h} können über das Gleichungssystem (44) berechnet werden. Dies ist für große t direkt jedoch nicht sehr praktikabel.

2. Für $h = 1$ kann man die Levinson Rekursion (Korollar 2.25) verwenden.

3. Allgemein über (iteratives) Gram-Schmidt Verfahren, dies wird in diesem Kontext auch Innovation Algorithmus genannt.

4. Speziell für ARMA-Modelle: Über Zustandsraummodell Darstellung sowie Kalman Vorhersage.

Vorhersageintervall

Ist $(X_t)_{t \in \mathbb{Z}}$ ein stationärer, reellwertiger, zentrierter Gauß-Prozess, so ist $X_{t+h} - \hat{X}_{t+h} \sim N(0, \Delta_t(h))$. Daher kann ein $1 - \alpha$ Vorhersageintervall für X_{t+h} bestimmt werden durch

$$[\hat{X}_{t+h} - u_{1-\alpha/2} \Delta_t(h)^{1/2}, \hat{X}_{t+h} + u_{1-\alpha/2} \Delta_t(h)^{1/2}].$$

Beispiel 2.34

Speziell für $AR(p)$, $MA(q)$.

Vorhersage bei unendlicher Vergangenheit

Nun sei $(X_t)_{t \in \mathbb{Z}}$ ein kausaler, invertibler (zentrierter) reellwertiger ARMA(p,q)-Prozess mit $\phi(B)X_t = \theta(t)e_t$, $t \in \mathbb{Z}$. Setzte

$$H_{t,-\infty} = \overline{\text{span}}\{X_j, -\infty < j \leq t\}.$$

Definition 2.35

Die Zufallsvariable $\tilde{X}_{t+h} = P_{H_{t,-\infty}} X_{t+h}$ heißt beste lineare Vorhersage bei unendlicher Vergangenheit $(X_j)_{j \leq t}$. $E(X_{t+h} - \tilde{X}_{t+h})^2$ heißt linearer Vorhersagefehler bei unendlicher Vergangenheit.

\tilde{X}_{t+h} wird in erster Linie als Approximation an \hat{X}_{t+h} verwendet, und $E(X_{t+h} - \tilde{X}_{t+h})^2$ an $\Delta_t(h)$.

Satz 2.36

Sei $(X_t)_{t \in \mathbb{Z}}$ ein kausaler, invertibler (zentrierter) reellwertiger ARMA(p,q)-Prozess mit $\phi(B)X_t = \theta(t)e_t$, $t \in \mathbb{Z}$. Dann gilt

$$\tilde{X}_{t+h} = - \sum_{j=1}^{\infty} \pi_j \tilde{X}_{t+h-j}, \quad \frac{\phi(z)}{\theta(z)} = \sum_{j=1}^{\infty} \pi_j z^j, \quad (50)$$

$$\tilde{X}_{t+h} = \sum_{j=h}^{\infty} \psi_j e_{t+h-j}, \quad \frac{\theta(z)}{\phi(z)} = \sum_{j=1}^{\infty} \psi_j z^j. \quad (51)$$

Weiter ist $E(X_{t+h} - \tilde{X}_{t+h})^2 = \sigma^2 \sum_{j=1}^{h-1} \psi_j^2$.

Beweis

Zu (50): Es ist $e_{t+h} = X_{t+h} + \sum_{j=1}^{\infty} \pi_j X_{t+h-j}$. Anwendung von $P_{H_{t,-\infty}}$ liefert dann (50), da wegen $Ee_{t+h}X_j = 0$ für $j < t+h$ die Variable e_{t+h} orthogonal auf $H_{t,-\infty}$ ist. Mit $X_{t+h} = \sum_{j=1}^{\infty} \psi_j e_{t+h-j}$ ergibt sich analog (51), da wegen der Invertibilität $e_{t+h-j} \in H_{t,-\infty}$, $j \geq h$. Damit folgt $\tilde{X}_{t+h} - X_{t+h} = \sum_{j=1}^{h-1} \psi_j e_{t+h-j}$, und somit die letzte Behauptung des Satzes.

2.5 Parameterschätzung

ÜBERALL HÖHERE ORDNUNGEN ZULASSEN

1. AR(p) Modelle

Sei (X_t) ein stationärer, reeller, zentrierter AR(p) Prozess mit $\phi(B)X_t = e_t$, $t \in \mathbb{Z}$, wobei $(e_t)_{t \in \mathbb{Z}}$ ein weißes Rauschen ist.

a. *Yule-Walker Schätzer:* Für $p_0 \geq p$ gilt nach Bemerkung 2.26 die Yule-Walker Gleichungen (39). Ersetzt man darin die Kovarianzen durch die geschätzten Kovarianzen und löst nach den Parametern auf, so erhält man

$$\begin{aligned}\hat{\phi}(p_0) &= \hat{\Gamma}_{p_0}^{-1} \hat{\gamma}_{p_0}^1, \\ \hat{\sigma}_e^2 &= \hat{\gamma}(0) - \sum_{j=1}^p \hat{\phi}_j \hat{\gamma}(j).\end{aligned}$$

Dabei ist die Matrix $\hat{\Gamma}_{p_0}$ invertierbar nach Satz 1.25. Den sich ergebenden Schätzer, der von der Bauart ein *Momentenschätzer* ist, nennt man den *Yule-Walker Schätzer*.

Satz 2.37

Sei $(X_t)_{t \in \mathbb{Z}}$ ein stationärer, reeller, zentrierter, kausaler AR(p) Prozess mit $\phi(B)X_t = e_t$, $t \in \mathbb{Z}$, wobei $(e_t)_{t \in \mathbb{Z}}$ ein u.i.v. weißes Rauschen ist mit $Ee_t^2 = \sigma_e^2$, $Ee_t^4 < \infty$. Dann gilt für $p_0 \geq p$ für den Yule Walker Schätzer $\hat{\phi}(p_0)$ zur Ordnung p_0

$$\sqrt{n} (\hat{\phi}(p_0) - \phi(p_0)) \xrightarrow{\mathcal{L}} N(0, \sigma_e^2 \Gamma_{p_0}^{-1}), \quad (52)$$

wobei $\phi(p_0) = (\phi_1, \dots, \phi_p, 0, \dots, 0)^T \in \mathbb{R}^{p_0}$, und $\hat{\sigma}_e^2$ ist konsistent für σ_e^2 . Für $p_0 > p$ ist $\sigma_e^2 (\Gamma_{p_0}^{-1})_{p_0, p_0} = 1$, so dass für die partielle Autokorrelation

$$\sqrt{n} \hat{\phi}(p_0)_{p_0} = \sqrt{n} \hat{\pi}(p_0) \xrightarrow{\mathcal{L}} N(0, 1).$$

NOCH: KOVARIANZEN DES RESULTIERENDEN PROZESSES DIE EMPIRISCHEN KOVARIANZEN BIS ZUR ORDNUNG P; KAUSALITÄT STETS GEWÄHRLEISTET? STICHPUNKTARTIG SCHREIBEN!!!

b. *Maximum Likelihood:* Seien X_{1-p}, \dots, X_n , $p \geq 1$, Zufallsvariablen mit Dichte $f_{X_{1-p}, \dots, X_n}(x_{1-p}, \dots, x_n)$, und entsprechenden Marginaldichten. Sei

$$f_{X_t | X_{t-1}, \dots, X_{1-p}}(x_t | x_{t-1}, \dots, x_{1-p}) = \frac{f_{X_{1-p}, \dots, X_t}(x_{1-p}, \dots, x_t)}{f_{X_{1-p}, \dots, X_{t-1}}(x_{1-p}, \dots, x_{t-1})}, \quad t \leq p,$$

die bedingte Dichte von X_t gegeben X_{t-1}, \dots, X_{1-p} , also die Dichte der entsprechenden bedingten Verteilung. Dann gilt

$$f_{X_{1-p}, \dots, X_n}(x_{1-p}, \dots, x_n) = \prod_{t=1-p}^n f_{X_t | X_{t-1}, \dots, X_{1-p}}(x_t | x_{t-1}, \dots, x_{1-p}).$$

Falls nun $(X_t)_{t \in \mathbb{Z}}$ ein AR(p) Prozess mit u.i.v. $N(0, \sigma_e^2)$ -verteilterm Rauschen $(e_t)_{t \in \mathbb{Z}}$, dann ist $(X_t)_{t \in \mathbb{Z}}$ ein Gauß-Prozess, und die bedingten Dichten sind auch Gaußsch und durch die Kovarianzstruktur $\gamma(h)$ bestimmt. Diese kann aus den Parametern berechnet werden, und somit lässt sich die Likelihood-Funktion in Abhängigkeit von ϕ_1, \dots, ϕ_p und σ_e^2 berechnen. Speziell erhält man für $t \geq 1$ wegen $X_t = \sum_{j=1}^p \phi_j X_{t-j} + e_t$ und der Unabhängigkeit von $X_{t-j}, j \geq 1$ und e_t

$$f_{X_t|X_{t-1}, \dots, X_{1-p}}(x_t|x_{t-1}, \dots, x_{1-p}) = \frac{1}{(2\pi\sigma_e^2)^{1/2}} \exp\left(-\frac{1}{2\sigma_e^2}\left(x_t - \sum_{j=1}^p \phi_j x_{t-j}\right)^2\right), \quad t \geq 1.$$

Somit erhält man für die Log-Likelihood mit $\phi = (\phi_1, \dots, \phi_p)$

$$\begin{aligned} \mathcal{L}(\phi, \sigma_e^2) &= -\frac{1}{2\sigma_e^2} \sum_{t=1}^n \left(X_t - \sum_{j=1}^p \phi_j X_{t-j}\right)^2 - \frac{n}{2} \log \sigma_e^2 - \frac{n}{2} \log(2\pi) \\ &\quad + \sum_{t=1-p}^0 f_{X_t|X_{t-1}, \dots, X_{1-p}}(x_t|x_{t-1}, \dots, x_{1-p}). \end{aligned} \quad (53)$$

Explizit für AR(1) : Hier ist $X_t = \phi X_{t-1} + e_t$, und es ist nur $f_{X_0}(x_0)$ zu berechnen. Mit $\gamma(0) = \sigma_e^2/(1 - \phi^2)$ ist $f_{X_0}(x_0)$ die Dichte von $N(0, \gamma(0))$. Somit ist hier

$$\begin{aligned} \mathcal{L}(\phi, \sigma_e^2) &= -\frac{1}{2\sigma_e^2} \sum_{t=1}^n (X_t - \phi X_{t-1})^2 - \frac{n+1}{2} \log \sigma_e^2 \\ &\quad - \frac{n+1}{2} \log(2\pi) - \frac{X_0^2(1 - \phi^2)}{\sigma_e^2} + \frac{1}{2} \log(1 - \phi^2). \end{aligned} \quad (54)$$

Eine Maximalstelle $(\hat{\phi}, \hat{\sigma}_e^2)$ von (58) mit kausalem Parametervektor ϕ heißt ein Maximum Likelihood (ML) Schätzer, Bezeichnung $(\hat{\phi}_{ML}, \hat{\sigma}_{ML}^2)$ ist somit nichtlineare Optimierung erforderlich. Für die asymptotische Verteilung gilt der folgende Satz

mit GENAU FORMULIERUNG bzgl EXISTENZ!!!!

Satz 2.38

Sei $(X_t)_{t \in \mathbb{Z}}$ ein stationärer, reeller, zentrierter, kausaler AR(p) Prozess mit $\phi(B)X_t = e_t$, $t \in \mathbb{Z}$, wobei $(e_t)_{t \in \mathbb{Z}}$ ein u.i.v. weißes Rauschen ist mit $Ee_t^2 = \sigma_e^2$, $Ee_t^4 < \infty$. Dann gilt für den Maximum Likelihood Schätzer $\hat{\phi}^{MLE}$, also eine kausale Maximalstelle von (58), die mit gegen 1 konvergierender Wahrscheinlichkeit existiert,

$$\sqrt{n+p}(\hat{\phi}^{MLE} - \phi) \xrightarrow{\mathcal{L}} N(0, \sigma_e^2 \Gamma_p^{-1}), \quad (55)$$

und $\hat{\sigma}_{ML}^2 \rightarrow \sigma_e^2$.

Bemerkung 2.39

a. Man beachte, dass der Schätzer zwar unter der Annahme eines normalverteilten weißen Rauschens hergeleitet wurde, aber die Asymptotik in Satz 2.38 allgemein gilt. Es handelt sich also um einen quasi maximum Likelihood Schätzer

b. Statt mit X_1, \dots, X_{n+p} haben wir die Beobachtungen zur angenehmeren Notation mit X_{-p+1}, \dots, X_n indiziert. Es wurden aber insgesamt $n+p$ Beobachtungen gemacht (so wird n gerade definiert).

c. *Bedingter ML bzw. kleinste Quadrate Schätzer*

Für einen stationären, reellen, zentrierten AR(p) Prozess betrachten wir nun die bedingte Log-Likelihood Funktion, gegeben X_0, \dots, X_{1-p} . Da

$$f_{X_n, \dots, X_1 | X_0, \dots, X_{1-p}}(x_n, \dots, x_1 | x_0, \dots, x_{1-p}) = \prod_{t=1}^n f_{X_t | X_{t-1}, \dots, X_{1-p}}(x_t | x_{t-1}, \dots, x_{1-p}),$$

erhalten wir als bedingte Log-Likelihood

$$\mathcal{L}_{\text{cond}}(\boldsymbol{\phi}, \sigma_e^2) = -\frac{1}{2\sigma_e^2} \sum_{t=1}^n \left(X_t - \sum_{j=1}^p \phi_j X_{t-j} \right)^2 - \frac{n}{2} \log \sigma_e^2 - \frac{n}{2} \log(2\pi) \quad (56)$$

Somit erhalten wir als bedingten ML-Schätzer den (bedingten) kleinste Quadrate Schätzer

$$\hat{\boldsymbol{\phi}}^{CLS} = \operatorname{argmin}_{(\phi_1, \dots, \phi_p)} \sum_{t=1}^n \left(X_t - \sum_{j=1}^p \phi_j X_{t-j} \right)^2, \quad (57)$$

und als Schätzer für σ_e^2

$$\hat{\sigma}_e^2 = \frac{1}{n} \sum_{t=1}^n \left(X_t - \sum_{j=1}^p \hat{\phi}_j X_{t-j} \right)^2.$$

Man kann die Lösung von (57) leicht über ein Gleichungssystem ausdrücken, welches finit jedoch nicht eindeutig lösbar sein muss (asymptotisch ist dies aber der Fall). Man erhält wiederum einen Konvergenzsatz wie Satz 2.37 und 2.38. Alle Schätzer haben somit asymptotisch die gleiche Kovarianzmatrix. Insbesondere ist der Yule-Walker Schätzer, ein Momentenschätzer, in diesem speziellen Fall ebenso effizient wie der MLE.

Bemerkung 2.40 (Schätzen des Erwartungswertes μ)

Bei Yule-Walker Mittelwert, dann zentrieren, sonst in die Likelihood-Funktion aufnehmen, dann mitschätzen. WAS PAAIERT MIT DER ASYMPTOTIK??

2. Parameterschätzung von ARMA(p,q)-Prozessen

a. *Maximum-Likelihood Schätzung*

Sei $(X_t)_{t \in \mathbb{Z}}$ ein kausaler, invertibler, zentrierter ARMA(p,q)-Prozess, $\phi(B)X_t = \theta(B)e_t$.

Die Einschnitt Vorhersage und der entsprechende Vorhersagefehler hängen von der Kovarianzstruktur $\gamma(h)$, und somit von den Parametern $(\boldsymbol{\phi}^T, \boldsymbol{\theta}^T, \sigma_e^2)$ ab. Wir schreiben $\boldsymbol{\beta}^T = (\boldsymbol{\phi}^T, \boldsymbol{\theta}^T)$ und

$$P_{H_t} X_{t+1} = \hat{X}_{t+1} = \hat{X}_{t+1}(\boldsymbol{\beta}, \sigma_e^2), \quad \Delta_t(1) = E(X_{t+1} - \hat{X}_{t+1})^2 = \Delta_t(1, \boldsymbol{\beta}, \sigma_e^2).$$

Lemma 2.41

Man kann schreiben

$$\Delta_t(1, \boldsymbol{\beta}, \sigma_e^2) = \sigma_e^2 r_t(\boldsymbol{\beta}), \quad \hat{X}_{t+1}(\boldsymbol{\beta}, \sigma_e^2) = \hat{X}_{t+1}(\boldsymbol{\beta}).$$

Beweis

Dies folgt aus der Levinson Rekursion, da sich zunächst schreiben lässt $\gamma(t) = \sigma_e^2 b(\boldsymbol{\beta}, t)$, und dann $a_j^{t,1} = a_j^{t,1}(\boldsymbol{\beta})$ (σ_e^2 kürzt sich zunächst bei $a_t^{t,1}$ heraus, dann Induktion), und daher $\hat{X}_{t+1}(\boldsymbol{\beta}, \sigma_e^2) = \hat{X}_{t+1}(\boldsymbol{\beta})$ (Linearkombination in den Koeffizienten $a_j^{t,1}$). Weiter ergibt sich direkt die zweite Behauptung aus der letzten Gleichung der Levinson Rekursion. ■

Zur Herleitung nehmen wir nun wieder Normalverteilung der (e_t) an. Dann ist $(X_1, \dots, X_t) \sim N(0, \Gamma_t)$, und über die Formeln für bedingte Verteilungen bei gemeinsamer Normalverteilung zeigt man $X_t | (X_{t_1}, \dots, X_1) \sim N(\mu_t, \sigma_t^2)$, mit

$$\begin{aligned}\mu_t &= (X_1, \dots, X_{t-1}) \Gamma_{t-1}^{-1} \boldsymbol{\gamma}_{t-1}^1 = \hat{X}_t(\boldsymbol{\beta}), \\ \sigma_t^2 &= \gamma(0) - (\boldsymbol{\gamma}_{t-1}^1)^T \Gamma_{t-1}^{-1} \boldsymbol{\gamma}_{t-1}^1 = E(X_t - \hat{X}_t(\boldsymbol{\beta}))^2 = \sigma_e^2 r_{t-1}(\boldsymbol{\beta}),\end{aligned}$$

also

$$X_t | (X_{t_1}, \dots, X_1) \sim N(\hat{X}_t(\boldsymbol{\beta}), \sigma_e^2 r_t(\boldsymbol{\beta})).$$

und daher

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, \sigma_e^2) &= \sum_{t=1}^n \log f(X_t | X_{t-1}, \dots, X_1) \\ &= -\frac{1}{2\sigma_e^2} \sum_{t=1}^n \frac{(X_t - \hat{X}_t(\boldsymbol{\beta}))^2}{r_t(\boldsymbol{\beta})} - \frac{n}{2} \log \sigma_e^2 - \frac{n}{2} \log(2\pi) \\ &\quad - \frac{1}{2} \log(r_1(\boldsymbol{\beta}) \cdot \dots \cdot r_n(\boldsymbol{\beta})).\end{aligned}\tag{58}$$

Ein ML Schätzer $(\hat{\boldsymbol{\beta}}^{ML}, \hat{\sigma}_{ML}^2)$ von $\mathcal{L}(\boldsymbol{\beta}, \sigma_e^2)$ ist eine kausale, invertible Maximalstelle von (58). Durch partielles Ableiten nach σ^2 und Nullsetzen erhält man die Relation

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{t=1}^n \frac{(X_t - \hat{X}_t(\hat{\boldsymbol{\beta}}^{ML}))^2}{r_t(\hat{\boldsymbol{\beta}}^{ML})}.$$

Setzt man dies in (58) ein, so ergeben sich $(\hat{\boldsymbol{\beta}}^{ML})$ als Minimalstellen von

$$l(\boldsymbol{\beta}) = \log \left(\frac{1}{n} \sum_{t=1}^n \frac{(X_t - \hat{X}_t(\boldsymbol{\beta}))^2}{r_t(\boldsymbol{\beta})} \right) + \frac{1}{n} \sum_{t=1}^n \log r_t(\boldsymbol{\beta}).\tag{59}$$

Für feste Werte von $\boldsymbol{\beta}$ kann man $\hat{X}_t(\boldsymbol{\beta})$ und $r_t(\boldsymbol{\beta})$ aus der Levinson Rekursion berechnen (mit Korrelationen statt den Kovarianzen). Für das asymptotische Verhalten gilt nun der folgende

Satz 2.42

Sei $(X_t)_{t \in \mathbb{Z}}$ ein zentrierter, kausaler und invertibler ARMA(p,q) Prozess mit $\phi(B)X_t = \theta(B)e_t$, $t \in \mathbb{Z}$, wobei $(e_t)_{t \in \mathbb{Z}}$ ein u.i.v. weißes Rauschen ist mit $Ee_t^2 = \sigma_e^2 < \infty$. Dann gilt für den Maximum Likelihood Schätzer $(\hat{\boldsymbol{\beta}}^{ML})$ von $\boldsymbol{\beta}^T = (\boldsymbol{\phi}^T, \boldsymbol{\theta}^T)$

$$\sqrt{n} ((\hat{\boldsymbol{\beta}}^{ML} - \boldsymbol{\beta})) \xrightarrow{\mathcal{L}} N(0, \sigma_e^2 V((\boldsymbol{\beta}))),\tag{60}$$

wobei

$$V(\boldsymbol{\beta}) = \begin{pmatrix} E\mathbf{U}_t\mathbf{U}_t^T & E\mathbf{U}_t\mathbf{V}_t^T \\ E\mathbf{V}_t\mathbf{U}_t^T & E\mathbf{V}_t\mathbf{V}_t^T \end{pmatrix}^{-1},$$

wobei $\mathbf{U}_t = (U_t, \dots, U_{t-p+1})^T$, $\mathbf{V}_t = (V_t, \dots, V_{t-q+1})^T$, und (U_t) , (V_t) sind die autoregressiven Prozesse

$$\phi(B)U_t = e_t, \quad \theta(B)V_t = e_t, \quad t \in \mathbb{Z}.$$

Für den Beweis verweisen wir auf Brockwell und Davis (1991, Kapitel 10.8).

Beispiel 2.43

1. AR(p). Hier ist wie in Satz 2.38

$$V(\phi) = (E\mathbf{U}_t\mathbf{U}_t^T)^{-1} = \Gamma_p^{-1}.$$

2. MA(q). Hier ist

$$V(\theta) = (E\mathbf{V}_t\mathbf{V}_t^T)^{-1} = (\Gamma_q^*)^{-1},$$

wobei Γ_q^* die Kovarianzmatrix des autoregressiven Prozesses $V_t + \theta_1 V_{t-1} + \dots + \theta_q V_{t-q} = e_t$ ist. Insbesondere ist für $q = 1$

$$\sigma^2 V(\theta) = (1 - \theta^2).$$

Für AR(p) Prozesse hatten wir gesehen, dass der Momentenschätzer (basierend auf den empirischen Kovarianzen) ebenso effizient ist wie der MLE. Dies ist bei MA(q) Prozessen nicht der Fall. Es ist $\rho(1) = \theta/(1 + \theta^2)$, und somit stets $|\rho(1)| < 1/2$. Falls $|\hat{\rho}(1)| < 1/2$, können wir auflösen und erhalten als invertiblen Schätzer

$$\hat{\theta}^{MO} = \frac{1 - \sqrt{1 - 4\hat{\rho}(1)^2}}{2\hat{\rho}(1)}.$$

MEHR!!!!

3. ARMA(1,1). Hier zeigt man mit $U_t = \sum_{j=0}^{\infty} \phi^j e_{t-j}$, $V_t = \sum_{j=0}^{\infty} (-\theta)^j e_{t-j}$, dass

$$V(\phi, \theta) = \begin{pmatrix} (1 - \phi^2)^{-1} & (1 + \phi\theta)^{-1} \\ (1 + \phi\theta)^{-1} & (1 - \theta^2)^{-1} \end{pmatrix}^{-1}$$

2.6 Ordnungswahl und Modellierung

1. Zunächst Zeitreihe auf stationäre Form reduzieren.

Ordnungswahl

reine MA(q) Modelle: Hier ist die Autokorrelation = 0 für lags $h > q$. Man betrachtet dann die empirische Autokorrelation mit den zugehörigen Konfidenzschranken, die zu einem weißen Rauschen gehören, die also auf der asymptotischen Standardnormalverteilung von $\sqrt{n}\hat{\rho}(h)$ basieren. Diese sind allerdings schmaler als für ein MA(q) Modell, da hier für $h > q$ gilt

$$\sqrt{n}\hat{\rho}(h) \rightarrow N(0, w_{hh}), \quad w_{hh} = (1 + 2\rho^2(1) + \dots + 2\rho^2(q)).$$

Man wählt also tendenziell zu große Modelle.

reine AR(p) Modelle: Hier ist die partielle Autokorrelation = 0 für lags $h > p$. Man betrachtet dann die empirische partielle Autokorrelation, für die für lags $h > p$ gilt $\sqrt{n}\hat{\pi}(h) \rightarrow N(0, 1)$.

In einem ARMA(p,q) Modell hat man keine einfachen Kriterien mehr für die Ordnungswahl. Man orientiert sich dennoch an ACF und PACF als oberen Schranken für die jeweiligen Ordnungen.

Modellwahlkriterien Insbesondere bei gemischten ARMA(p,q) Modellen betrachtet man zur Ordnungswahl formale Modellwahlkriterien, insbesondere das AIC und das BIC. Diese sind definiert durch

$$\begin{aligned} \text{AIC} &= -2\mathcal{L}(\hat{\beta}^{ML}, \hat{\sigma}_{ML}^2) + 2(p + q + 1), \\ \text{AICC} &= -2\mathcal{L}(\hat{\beta}^{ML}, \hat{\sigma}_{ML}^2) + 2\frac{(p + q + 1)n}{n - p - q - 2}, \\ \text{BIC} &= -2\mathcal{L}(\hat{\beta}^{ML}, \hat{\sigma}_{ML}^2) + (p + q + 1) \log n, \end{aligned}$$

wobei $\beta^T = (\phi^T, \theta^T)$ und n die Länge der Zeitreihe bezeichnen. Man wählt nun das Modell mit minimalen AIC (bzw. AICC oder BIC).

Modelldiagnostik

Residuenanalyse

Die Residuen sind definiert durch

$$\hat{e}_t = \frac{X_t - \hat{X}_t(\hat{\beta})}{\hat{\sigma}(r_{t-1}(\hat{\beta}))^{1/2}},$$

wobei wir die Notation aus Lemma 2.41 verwenden. Falls (X_t) tatsächlich einem ARMA(p,q) Modell mit Parametern β folgen würde, so wäre (\hat{e}_t) ein weißes Rauschen (da dann $X_t - \hat{X}_t(\beta)$ orthogonal auf H_{t-1}). Da aber $\hat{\beta}$ nur ein Schätzer ist, sind die (\hat{e}_t) selbst bei korrekt spezifiziertem Modell nur approximativ ein weißes Rauschen.

Deskriptive Residuenanalyse: QQPlot (gegen Normalverteilung), ACF und PACF, Plot

der Reihe der Residuen.

Autokorrelation der Residuen: Sei $\hat{\rho}_e(h)$, $h \geq 1$, die empirische Autokorrelation der Residuen (\hat{e}_t). Dann sind für festes $h \geq 1$ $\sqrt{n}(\hat{\rho}_e(1), \dots, \hat{\rho}_e(h))$ asymptotisch korreliert, s. Brockwell und Davis (1991) (im Gegensatz zu weißen Rauschen, vgl. Bemerkung ??).

Portmanteau Tests: Um das Modell zu überprüfen und zu untersuchen, ob die Residuen (\hat{e}_t) die Eigenschaften haben, die man bei Gültigkeit des Modells erwarten kann, betrachtet man auch formale Hypothesentests.

$$Q_e = n \sum_{j=1}^h \hat{\rho}_e^2(j) \quad \text{Box-Pierce}$$

$$\tilde{Q}_e = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}_e^2(j)}{n-j} \quad \text{Ljung-Box}$$

Falls $h = h(n) \rightarrow \infty$ derart, dass

a. $h = O(n^{1/2})$,

b. $\psi_j = O(n^{-1/2})$, $j \geq h$ (wobei $X_t = \sum_{j \geq 0} \psi_j e_{t-j}$),

dann sind

$$Q_e, \tilde{Q}_e \sim \chi^2(h - p - q).$$

Ein verwandte Statistik ergibt sich über die Autokorrelation $\hat{\rho}_{ee}(h)$ der Quadrate (\hat{e}_t). Hier ist

$$\tilde{Q}_{ee} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}_{ee}^2(j)}{n-j} \quad \text{McLeod-Li}$$

asymptotisch $\chi^2(h)$ verteilt.

2.7 Literaturhinweise und Bemerkungen

2.8 Anwendung auf Datenbeispiele

3 Erweiterungen der ARMA Modellklasse

3.1 Integrierte Modelle für nicht-stationäre Zeitreihen

Definition 3.1

Für $d \in \mathbb{N}$ heißt $(X_t)_{t \in \mathbb{Z}}$ eine ARIMA(p,d,q) Zeitreihe, falls $Y_t = (1 - B)^d X_t$ eine kausale ARMA(p,q) Reihe ist.

Beispiel 3.2 (Fall $d = 1$)

Wir betrachten den Fall $d = 1$. Hier gilt dann für $t \geq 1$

$$X_t = X_0 + \sum_{k=1}^t Y_k, \quad (61)$$

wobei (Y_t) einem ARMA(p,q) Modell folgt.

Beispiel 3.3 (ARIMA(0,1,1))

Wir betrachten noch spezieller ein ARIMA(0,1,1) Modell, welches wir mit $\lambda = \theta$ in der Form

$$X_t = X_{t-1} + e_t - \lambda e_{t-1} \quad (62)$$

schreiben können, wobei (e_t) ein weißes Rauschen ist und $\lambda \in (-1, 1)$. Für $Y_t = X_t - X_{t-1}$ ist dann

$$e_t = \sum_{k \geq 0} \lambda^k Y_{t-k}.$$

Einsetzen in (62) führt nach kurzer Rechnung zu

$$X_t = (1 - \lambda) \sum_{j=1}^{\infty} \lambda^{j-1} X_{t-j} + e_t. \quad (63)$$

Nimmt man nun an, dass e_t senkrecht auf $H_{t-1, -\infty} = \overline{\text{span}}\{X_{t-1}, X_{t-2}, \dots\}$ ist, dann ist die Projektion \hat{X}_t von X_t auf $H_{t-1, -\infty}$ mit kurzer Rechnung

$$\hat{X}_t = (1 - \lambda) \sum_{j=1}^{\infty} \lambda^{j-1} X_{t-j} = (1 - \lambda) X_{t-1} + \lambda \hat{X}_{t-1}.$$

Dies führt also auf das exponentielle Glätten.

Parameterschätzen und Modellwahl

Zunächst bestimmt man den Parameter d als minimalen Wert, für den die Reihe $Y_t = (1 - B)^d X_t$ stationär wird. Hier kann die Autokorrelationsfunktion betrachtet werden sowie auch formale Tests auf Stationarität (s. Sektion ??).

Danach passt man an (Y_t) ein ARMA(p,q) Modell an.

Vorhersage

Die L_2 Eigenschaften eines ARIMA Modells sind nicht eindeutig bestimmt. Dies erkennt man etwa an der Gleichung (61) für ein ARIMA(p,1,q) Modell, bei der X_0 beliebig gewählt werden kann. Somit ist auch die beste lineare Vorhersage ohne weitere Annahmen nicht eindeutig bestimmt.

Vorhersage im ARIMA(p,1,q) Modell. Wir suchen die orthogonale Projektion \hat{X}_{t+h} von X_{t+h} auf

$$\tilde{H}_t = \text{span} \{X_0, \dots, X_t\} = \text{span} \{X_0, Y_1, \dots, Y_t\}, \quad t \geq 1,$$

wobei $Y_t = X_t - X_{t-1}$. Es ist

$$\hat{X}_{t+h} = P_{\tilde{H}_t} X_{t+h} = P_{\tilde{H}_t} (X_0 + Y_1 + \dots + Y_{t+h}) = X_t + \sum_{j=1}^h P_{\tilde{H}_t} Y_{t+j}.$$

Zur Berechnung von $P_{\tilde{H}_t} Y_{t+j}$ benötigt man $E(X_0 Y_t)$, $t \geq 1$ (sowie EX_0^2).

Annahme. Es ist $EX_0 Y_k = 0$ für $k = 1, \dots, t+h$.

In diesem Fall ist $\tilde{H}_t = H_t \oplus \text{span} \{X_0\}$ wobei $H_t = \text{span} \{Y_1, \dots, Y_t\}$ (orthogonale Summe), und $P_{\tilde{H}_t} = P_{H_t} + P_{X_0}$. Da $P_{X_0} Y_{t+j} = 0$, erhält man $P_{\tilde{H}_t} Y_{t+j} = P_{H_t} Y_{t+j}$, die j -Schritt Vorhersage bei einem stationären ARMA(p,q) Prozess. Insgesamt hat man unter der Annahme der Unkorreliertheit also

$$\hat{X}_{t+h} = X_t + \sum_{j=1}^h P_{H_t} Y_{t+j}.$$

Als Vorhersagefehler ergibt sich

$$E(X_{t+h} - \hat{X}_{t+h})^2 = E\left(\sum_{j=1}^h (Y_{t+j} - P_{H_t} Y_{t+j})\right)^2.$$

Explizite Berechnung von Vorhersage und Vorhersagefehler:

- Innovation Algorithmus (s. Brockwell und Davis (1991)).
- Zustandsraum Darstellung eines ARIMA Modells und Kalman Filterung.

Vorhersage im ARIMA(p,d,q) Modell.

Enwickelt man die rechte Seite von $Y_t = (1 - B)^d X_t$ und löst nach X_t auf, erhält man

$$X_t = Y_t - \sum_{j=1}^d \binom{d}{j} (-1)^j X_{t-j}, \quad t = 1, 2, \dots \quad (64)$$

Indiziere die Beobachtungen X_{1-d}, \dots, X_t und setze

$$\tilde{H}_t = \text{span} \{X_{1-d}, \dots, X_t\} = \text{span} \{X_{1-d}, \dots, X_0, Y_1, \dots, Y_t\}, \quad t \geq 1,$$

wobei die zweite Gleichung aus (64) folgt. Anwendung von $P_{\tilde{H}_t}$ auf X_{t+h} liefert mit (64)

$$\hat{X}_{t+h} = P_{\tilde{H}_t} X_{t+h} = P_{\tilde{H}_t} Y_{t+h} - \sum_{j=1}^d \binom{d}{j} (-1)^j P_{\tilde{H}_t} X_{t+h-j}, \quad t = 1, 2, \dots \quad (65)$$

Falls also $P_{\tilde{H}_t} Y_{t+h}$ berechnet werden kann, kann man mit (65) die Vorhersagen \hat{X}_{t+h} rekursiv berechnen (beachte $P_{\tilde{H}_t} X_{t+h-j} = X_{t+h-j}$ für $j \geq h$.) Mit $H_t = \text{span} \{Y_1, \dots, Y_t\}$ und der Annahme, dass H_t und $\text{span} \{X_{1-d}, \dots, X_0\}$ orthogonal aufeinander stehen, erhält man dann wie oben $P_{\tilde{H}_t} Y_{t+h} = P_{H_t} Y_{t+h}$.

3.2 Seasonal ARIMA Modelle

MOTIVATION VERBESSERN; AUCH NICHT MULTIPLIKATIVE MODELLE ERWÄHNEN!!!

Häufig ist es nicht sinnvoll, die Saisonkomponente einer Zeitreihe rein deterministisch zu modellieren, vielmehr gibt es auch im saisonalen Verlauf zufällige Schwankungen. Hierfür wollen wir eine Klasse von Zeitreihenmodellen einführen, die auf den ARIMA Modellen basieren.

TABLE MIT DEN MONATLICH GEORDNETEN DATEN

Jede Spalte kann als eine Zeitreihe gesehen werden, die den Verlauf in der jeweiligen Saison (hier dem Monat) beschreibt.

Wir nehmen nun an, diese folgen dem gleichen ARMA(P,Q) Modell, und bezeichnen die Anzahl der Saisons allgemein mit s . Dann kann man schreiben für $j = 1, \dots, s$ und $t \in \mathbb{Z}$

$$X_{j+st} = \Phi_1 X_{j+s(t-1)} + \dots + \Phi_P X_{j+s(t-P)} + U_{j+st} + \Theta_1 U_{j+s(t-1)} + \dots + \Theta_Q U_{j+s(t-Q)},$$

wobei Φ_1, \dots, Φ_P und $\Theta_1, \dots, \Theta_Q$ die Koeffizienten des ARMA Modells sind, und $EU_t^2 = \sigma_U^2$. Es ist dann also $(U_{j+st})_{t \in \mathbb{Z}}$ ein weißes Rauschen für $j = 1, \dots, s$. Man kann dann auch einheitlich schreiben

$$X_t = \Phi_1 X_{t-s} + \dots + \Phi_P X_{t-Ps} + U_t + \Theta_1 U_{t-s} + \dots + \Theta_Q U_{t-Qs}$$

oder

$$\Phi(B^s)X_t = \Theta(B^s)U_t, \quad t \in \mathbb{Z}, \quad (66)$$

wobei

$$\begin{aligned} \Phi(z) &= 1 - \Phi_1 z - \dots - \Phi_P z^P \\ \Theta(z) &= 1 + \Theta_1 z + \dots + \Theta_Q z^Q. \end{aligned}$$

Es ist nun natürlich unrealistisch anzunehmen, dass die s saisonalen Zeitreihen völlig unkorreliert voneinander verlaufen. Vielmehr modelliert man die (U_t) als stationäre Zeitreihe über ein ARMA(p,q) Modell

$$\phi(B)U_t = \theta(B)e_t, \quad t \in \mathbb{Z}, \quad (67)$$

wobei e_t ein weißes Rauschen mit Varianz σ_e^2 ist. In diesem Modell sind die Reihen $(U_{j+st})_{t \in \mathbb{Z}}$ kein weißes Rauschen mehr, falls jedoch die Saisonkomponente s hinreichend groß ist und die Korrelation im ARMA(p,q) Modell (67) hinreichend schnell abklingt, wird die Korrelation in diesen Reihen klein sein.

Aus (66) und (67) ergibt sich zusammen

$$\phi(B)\Phi(B^s)X_t = \phi(B)\Theta(B^s)U_t = \Theta(B^s)\phi(B)U_t = \Theta(B^s)\theta(B)e_t. \quad (68)$$

Man nennt (68) ein saisonales ARMA Modell, Bezeichnung SARMA(p,q) \times (P,Q)_s. Dies ist offenbar ein spezielles ARMA(p + sP, q + sQ) Modell.

Um eventuelle Nichtstationerität zu entfernen, werden in einem ersten Schritt noch Differenzen gebildet

$$Y_t = (1 - B)^d(1 - B^s)^D X_t = (1 - B^s)^D(1 - B)^d X_t. \quad (69)$$

Definition 3.4

Eine Zeitreihe folgt einem saisonalen ARIMA Modell $SARIMA(p, d, q) \times (P, D, Q)_s$, falls die Differenzen (69) die Gleichung (68) für ein weißes Rauschen $(e_t)_{t \in \mathbb{Z}}$ erfüllen.

Wir bemerken, dass das $SARMA(p, q) \times (P, Q)_s$ Modell (3.4) genau dann kausal ist, falls $\phi(z) \neq 0$ und $\Phi(z) \neq 0$ für $|z| \leq 1$ (da das AR Polynom gegeben ist durch $\phi(z)\Phi(z^s)$).

Parameterschätzung und Modellwahl

Zunächst wählt man die Ordnungen d und D derart, dass man eine stationäre Zeitreihe erhält.

Für die Reihe der Differenzen betrachtet man dann die partielle Autokorrelation zunächst zu Lags, die Vielfache von s sind, also $\hat{\rho}(ks)$, $k \geq 0$. Diese sollten einem $ARMA(P, Q)$ Prozess folgen. Danach bestimmt man die Ordnungen p, q , entweder direkt oder aus den Residuen des ersten Schrittes.

Die Parameter können etwa mit Maximum Likelihood geschätzt werden, da das SARMA Modell ein Teilmodell eines $ARMA(p + sP, q + sQ)$ Modells ist, wobei dessen Parameter eine differenzierbare Funktion des Parametervektors $\beta^T = (\phi^T, \Phi^T, \theta^T, \Theta^T)$ ist.

Man kann dann zur Ordnungbestimmung auch formale Modellwahlkriterien wie das AIC oder das BIC anwenden.

Vorhersage

Ähnlich wie bei ARIMA Modellen muss man Annahmen bzgl. der Unkorreliertheit der ersten Differenzen von den weiteren Beobachtungen machen (s. Brockwell und Davis (1991)).

3.3 Long range dependence und fractional ARMA Modelle

1. Long range dependence

Definition 3.5

Ein stationärer Prozess $(X_t)_{t \in \mathbb{Z}}$ heißt *long memory* Prozess oder Prozess mit *long-range Korrelationen*, falls für die Korrelationsfunktion $\rho(k)$ für ein $\alpha \in (0, 1)$ und ein $c_\rho > 0$ gilt

$$\lim_{k \rightarrow \infty} \frac{\rho(k)}{c_\rho k^{-\alpha}} = 1. \quad (70)$$

Bemerkung 3.6

- Hat (X_t) long memory, dann ist $\sum_{k \geq 1} \rho(k) = \infty$, da $\rho(k) \geq c_\rho k^{-\alpha}/2$ für große k .
- Gilt $\rho(k) \sim c_\rho k^{-\alpha}$ für ein $\alpha > 1$, so ist $\sum_{k \geq 1} \rho(k) < \infty$. Man spricht dann manchmal von intermediate memory (im Gegensatz zu short memory, wo die Korrelationen exponentiell schnell abklingen).

Für einen long memory Prozess $(X_t)_{t \in \mathbb{Z}}$ heißt $H = 1 - \alpha/2$ der *Hurst Parameter*, in dieser Notation ist long memory äquivalent zu

$$\rho(k) \sim c_\rho k^{2H-2}. \quad (71)$$

Satz 3.7

Für einen long memory Prozess $(X_t)_{t \in \mathbb{Z}}$ gilt mit $\sigma^2 = \text{Var} X_t$

$$\lim_{n \rightarrow \infty} \frac{\text{Var}(\sum_{i=1}^n X_i)}{c_\rho \sigma^2 n^{2H}} = \frac{1}{H(2H-1)}. \quad (72)$$

Bemerkung 3.8

Zu (72) äquivalent ist

$$\lim_{n \rightarrow \infty} n^{2(1-H)} \frac{\text{Var}(\frac{1}{n} \sum_{i=1}^n X_i)}{c_\rho \sigma^2} = \frac{1}{H(2H-1)}.$$

Die Varianz des Mittelwertes fällt somit langsamer ab als mit einer $n^{-1/2}$ Rate.

REFERENZ UND BEWEIS FÜR DEN OBIGEN SATZ!!!

Fraktionale Brownsche Bewegung und fraktionales weißes Rauschen.

Definition 3.9

Ein (zeitstetiger) Prozess $(Y_t)_{t \in \mathbb{R}}$ heißt *self-similar* mit self-similarity Parameter $H \in (0, 1)$, falls gilt

$$(Y_t)_{t \geq 0} \stackrel{d}{=} (c^{-H} Y_{ct})_{t \geq 0} \quad \forall c > 0,$$

wobei $\stackrel{d}{=}$ Gleichheit in Verteilung bezeichnet.

BEMERKUNG ZU $H \in (0, 1)$.

Aus der Gleichung

$$Y_t \stackrel{d}{=} t^H Y_1$$

leitet man unter der Annahme $P(Y_1 = 0) = 0$ die folgende Beziehungen her:

$$|Y_t| \stackrel{d}{\rightarrow} \infty \quad \text{für } t \rightarrow \infty, \quad |Y_t| \stackrel{d}{\rightarrow} 0 \quad \text{für } t \rightarrow 0.$$

Wir erinnern daran, dass ein Prozess $(Y_t)_{t \geq 0}$ stationäre Inkremente hat, falls für alle $0 \leq t_1 < \dots < t_{k+1}$ die Verteilung von $(Y_{t_2+c} - Y_{t_1+c}, \dots, Y_{t_{k+1}+c} - Y_{t_k+c})$ nicht von $c \geq 0$ abhängt.

Im folgenden nehmen wir an, dass $(Y_t)_{t \geq 0}$ self-similar ist mit self-similarity Parameter $H > 0$ sowie stationäre Inkremente hat. Außerdem nehmen wir noch $EY_t = 0$ für alle t sowie $Y_0 = 0$ an. Dann ist für $s < t$

$$Y_t - Y_s \stackrel{d}{=} Y_{t-s} \stackrel{d}{=} (t-s)^H Y_1.$$

Mit $\sigma^2 = EY_1^2$ ergibt sich

$$E(Y_t - Y_s)^2 = \sigma^2 (t-s)^{2H}.$$

Wegen $E(Y_t - Y_s)^2 = \sigma^2(t^{2H} + s^{2H}) - 2\text{Cov}(Y_t, Y_s)$ erhält man somit für die Kovarianz eines self-similar Prozesses mit stationären Inkrementen

$$\text{Cov}(Y_t, Y_s) = \frac{\sigma^2}{2} (t^{2H} + s^{2H} - (t-s)^{2H}), \quad t > s. \quad (73)$$

Definition 3.10

Ein Gaußscher Prozess mit $(Y_t)_{t \geq 0}$ mit $Y_0 = 0$, Erwartungswertfunktion 0 und Kovarianzfunktion (73) für $H > 0$ heißt eine *fraktionale Brownsche Bewegung*.

Indem man die obigen Rechnungen zurückverfolgt, erkennt man, dass eine fraktionale Brownsche Bewegung gerade ein Gaußscher self-similar Prozess mit stationären Inkrementen ist. Für $H = 1/2$ erhält man $\text{Cov}(Y_t, Y_s) = \sigma^2 \min(t, s)$, also die Kovarianzfunktion der Brownschen Bewegung.

Bisher haben wir noch nicht gezeigt, dass es self-similar Prozesse mit stationären Inkrementen überhaupt gibt, oder mit anderen Worten, dass (73) eine gültige Kovarianzfunktion definiert.

Satz 3.11

Sei $B(t)$ eine zweiseitige standard Brownsche Bewegung. Dann definiert für $H \in (0, 1)$ das stochastische Integral

$$B_H(t) = \int_{-\infty}^0 ((t-u)^{H-1/2} - (-u)^{H-1/2}) dB(u) + \int_0^t (t-u)^{H-1/2} dB(u), \quad t \geq 0, \quad (74)$$

eine fraktionale Brownsche Bewegung.

Beweis

Wir zeigen, dass (74) self-similar mit Parameter H ist und stationäre Inkremente hat. Dazu setzen wir

$$w(t, u) = \begin{cases} 0, & u > t, \\ (t-u)^{H-1/2} - (-u)^{H-1/2}, & u < 0 \\ (t-u)^{H-1/2}, & 0 \leq u < t \end{cases}$$

so dass $B_H(t) = \int_{\mathbb{R}} w(t, u) dB(u)$.

a. *self-similar*. Da

$$w(ct, u) = c^{H-1/2} w(t, u/c), \quad c > 0$$

folgt

$$\begin{aligned} B_H(ct) &= \int_{\mathbb{R}} w(ct, u) dB(u) = c^{H-1/2} \int_{\mathbb{R}} w(t, u/c) dB(u) \\ &= c^{H-1/2} \int_{\mathbb{R}} w(t, v) dB(cv) \stackrel{d}{=} c^{H-1/2} c^{1/2} \int_{\mathbb{R}} w(t, v) dB(v). \end{aligned}$$

b. *stationäre Inkremente*. Wir betrachten zunächst nur ein Inkrement $B_H(t_2) - B_H(t_1)$ für $0 \leq t_1 < t_2$ und $c \geq 0$. Dann ist

$$\begin{aligned} &B_H(t_2 + c) - B_H(t_1 + c) \\ &= \int_{-\infty}^{t_1+c} ((t_2 + c - u)^{H-1/2} - (t_1 + c - u)^{H-1/2}) dB(u) + \int_{t_1+c}^{t_2+c} (t_2 + c - u)^{H-1/2} dB(u) \\ &= \int_{-\infty}^{t_1} ((t_2 - u)^{H-1/2} - (t_1 - u)^{H-1/2}) dB(u - c) + \int_{t_1}^{t_2} (t_2 - u)^{H-1/2} dB(u - c) \\ &\stackrel{d}{=} \int_{-\infty}^{t_1} ((t_2 - v)^{H-1/2} - (t_1 - v)^{H-1/2}) dB(v) + \int_{t_1}^{t_2} (t_2 - v)^{H-1/2} dB(v) \\ &= B_H(t_2) - B_H(t_1), \end{aligned}$$

da die Brownsche Bewegung stationäre Inkremente hat. Mit derselben Translation lässt sich das Argument auf mehrere Inkremente übertragen. ■

Definition 3.12

Für einen self-similar Prozess $(Y_t)_{t \geq 0}$ mit self-similarity Parameter $H > 0$, stationären Inkrementen, Erwartungswertfunktion 0 und $Y_0 = 0$ heißt $X_k = Y_k - Y_{k-1}$, $k \geq 1$, ein *fraktionales Rauschen*. Falls $(Y_t)_{t \geq 0}$ eine fraktionale Brownsche Bewegung ist, so heißt $(X_k)_{k \geq 1}$ *fraktionales Gaußsches Rauschen*.

Da $(X_k)_{k \geq 1}$ stark stationär ist, können wir direkt die zweiseitige Erweiterung betrachten, $(X_k)_{k \in \mathbb{Z}}$.

Satz 3.13

Die Korrelationsfunktion $\rho(k)$ eines fraktionales Rauschens ist

$$\rho(k) = \frac{1}{2} ((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}), \quad k \geq 1. \quad (75)$$

Diese erfüllt für $H \neq 1/2$,

$$\frac{\rho(k)}{H(2H-1)k^{2H-2}} \rightarrow 1, \quad k \rightarrow \infty.$$

insbesondere ist das fraktionale Rauschen für $H \in (1/2, 1)$ long-range dependent mit Hurst Parameter H .

Beweis

Zunächst ist (X_k) offenbar (stark) stationär. Dann berechnen wir

$$\begin{aligned}\gamma(k) &= \text{Cov}(X_i, X_{i+k}) = \text{Cov}(X_1, X_{k+1}) \\ &= \text{Cov}(Y_1, Y_{1+k}) + \text{Cov}(Y_0, Y_k) - \text{Cov}(Y_1, Y_k) - \text{Cov}(Y_0, Y_{k+1}) \\ &= \frac{\sigma^2}{2} ((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}),\end{aligned}$$

wobei wir $Y_0 = 0$ benutzt haben. Wegen $EX_1^2 = EY_1^2 = \sigma^2$ ergibt sich (75). Sei nun

$$g(x) = (1+x)^{2H} - 2 + (1-x)^{2H}, \quad x \in (0, 1),$$

so dass

$$\rho(k) = k^{2H} g(k^{-1})/2. \quad (76)$$

■

Dann ist $g(0) = 0$,

$$\begin{aligned}g'(x) &= 2H((1+x)^{2H-1} - (1-x)^{2H-1}), & g'(0) &= 0, \\ g''(x) &= 2H(H-1)((1+x)^{2H-2} + (1-x)^{2H-2}), & g''(0) &= 4H(H-1).\end{aligned}$$

Daher ist $g(x) = 2H(H-1)x^2 + o(x^2)$ für $x \rightarrow 0$. Damit und mit (76) folgt (75).

2. Rauschen aus fraktionalen Differenzen

Das fraktionale Rauschen wurde in Definition 3.12 als diskrete Inkremente eines zeitstetigen Prozesses erhalten, welcher wiederum als fraktionales Integral bzgl. der Brownschen Bewegung in (74) definiert wurde.

Wir wollen nun direkt einen zeitdiskreten stationären Prozess definieren, der ähnliche Eigenschaften hat wie das fraktionale Rauschen. Dazu bemerken wir, dass das zweidiskrete Analogon zur Brownschen Bewegung der random walk ist, und somit das zeitdiskrete Analogon zum fraktionalen Integral bzgl. der Brownschen Bewegung in (74) eine fraktionale Summe bzgl. eines zeitdiskreten weißen Rauschens.

Binomische Reihenentwicklung. Für $|z| \leq 1$ gilt $(1-z)^d = \sum_{k=0}^{\infty} \alpha_k z^k$, wobei

$$\begin{aligned}\alpha_k &= (-1)^k \binom{d}{k} = (-1)^k \frac{d(d-1)\dots(d-k+1)}{k!} \\ &= \frac{(k-d-1)\dots(1-d)(-d)}{k!} = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)}.\end{aligned}$$

Fraktionale Differenzen-Operator. Für $d > -1$ setze

$$\Delta^d = (1-B)^d = \sum_{k=0}^{\infty} \alpha_k B^k.$$

Definition 3.14

Ein stationärer Prozess (X_t) heißt ein *Rauschen aus fraktionalen Differenzen*, bez. FARMA(0,d,0) mit einem $d \in (-0.5, 0.5)$, falls $\Delta^d X_t$ in L_2 konvergiert und

$$\Delta^d X_t = e_t, \quad (e_t) \quad \text{weißes Rauschen.} \quad (77)$$

Für ein Rauschen aus fraktionalen Differenzen heißt $H = d + 1/2$ der Hurst Parameter.

Wir sagen weiter, die Lösung (X_t) von (77) sei kausal, falls $X_t = \sum_{j \geq 0} \psi_j e_{t-j}$ mit Koeffizienten $\sum_{j \geq 0} \psi_j^2 < \infty$.

Satz 3.15

Für $d \in (-0.5, 0.5)$ und gegebenes weißes Rauschen (e_t) existiert genau eine stationäre, kausale Lösung (X_t) von (77). Diese hat die Darstellung

$$X_t = \sum_{j \geq 0} \psi_j e_{t-j} = \Delta^{-d} e_t, \quad \psi_j = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}.$$

Diese hat Varianz $\gamma(0) = \sigma^2 \Gamma(1-2d)/\Gamma^2(1-d)$ sowie Autokorrelationsfunktion und partielle Autokorrelationsfunktion

$$\rho(h) = \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)} = \prod_{k=1}^h \frac{k-1+d}{k-d}, \quad \pi(h) = \frac{d}{h-d}, \quad h = 0, 1, 2, \dots \quad (78)$$

Wir werden den Beweis von Satz 3.15 in der Spektraltheorie führen.

DEFINITION DER GAMMA FUNKTION FÜR NEGATIVES X .

Bemerkung 3.16

Da nach Stirlings Formel (REFERENZ) $\Gamma(x) \sim \sqrt{2\pi} e^{-x+1} (x-1)^{x-1/2}$, erhält man

$$\alpha_k \sim \frac{k^{-d-1}}{\Gamma(-d)}, \quad \psi_k \sim \frac{k^{d-1}}{\Gamma(d)}, \quad k \rightarrow \infty, \quad (79)$$

sowie

$$\rho(h) \sim \frac{h^{2d-1}\Gamma(1-d)}{\Gamma(d)}, \quad h \rightarrow \infty. \quad (80)$$

Für $d \in (0, 0.5)$ hat man in der Tat einen long range dependent Prozess im Sinne von Definition 3.5, es gilt (71) mit $H = d + 1/2$ und $c_\rho = \Gamma(1-d)/\Gamma(d)$.

Beweis von (79): Es ist $\psi_k = \Gamma(k+d)/(\Gamma(k+1)\Gamma(d))$, und

$$\begin{aligned} \frac{\Gamma(k+d)}{\Gamma(k+1)} &\sim \frac{e^{-k-d+1} (k+d-1)^{k+d-1/2}}{e^{-k} k^{k+1/2}} \\ &\sim k^{d-1} e^{-d+1} \left(1 + \frac{d-1}{k}\right)^{k+d-1/2} \sim k^{d-1}. \end{aligned}$$

Der andere Teil von (79) und (80) ergeben sich ähnlich.

3. Fraktionale ARMA Prozesse.

Rauschen aus fraktionalen Differenzen für Modellierung zu eng. Daher Erweiterung.

Definition 3.17

Eine stationäre Zeitreihe (X_t) heißt ein fraktionaler ARMA Prozess FARMA(p,d,q) mit $d \in (-0.5, 0.5)$, falls X_t die Gleichung

$$\phi(B)\Delta^d X_t = \theta(B)e_t \quad (81)$$

mit einem weißen Rauschen (e_t) erfüllt, wobei $\phi(z)$ und $\theta(z)$ Polynome vom Grad p und q sind.

Bemerkung 3.18

Mit $Y_t = \Delta^{-d}e_t$ ist (Anwendung der Lag-Operatoren Polynome kommutiert)

$$\phi(B)X_t = \theta(B)Y_t.$$

Somit kann man einen FARMA(p,d,q) Prozess auch interpretieren als einen ARMA Prozess, der durch Rauschen aus fraktionalen Differenzen gesteuert wird.

Man spricht in (81) wiederum von einer reduzierten Gleichung, falls $\phi(z)$ und $\theta(z)$ keine gemeinsamen Nullstellen haben. Kausalität und Invertibilität sind wie in den Definitionen erklärt, wobei jedoch nur die Quadratsummierbarkeit der Koeffizienten (und nicht die absolute Summierbarkeit) gefordert wird.

Satz 3.19

Wir betrachten eine reduzierte FARMA(p,d,q) Gleichung mit $d \in (-0.5, 0.5)$.

- Es existiert genau dann eine (dann eindeutig bestimmte) stationäre Lösung (X_t) der FARMA(p,d,q) Gleichung, falls $\phi(z) \neq 0$ für alle $|z| = 1$.
- Die Lösung ist genau dann kausal, falls $\phi(z) \neq 0$ für alle $|z| \leq 1$.
- Die Lösung ist genau dann invertibel, falls $\theta(z) \neq 0$ für alle $|z| \leq 1$.
- Für eine kausale und invertible Lösung gilt für die Autokorrelationsfunktion

$$\rho(h) \sim Ch^{2d-1} \quad h \rightarrow \infty,$$

für ein $C \neq 0$.

Für die Modellierung sind insbesondere long range FARMA(p,d,q) Prozesse, also mit $d \in (0, 1/2)$ relevant.

Parameterschätzung*1. Schätzung des Erwartungswertes*

In Satz 3.7 hatten wir gesehen, dass für einen long-range dependent Prozess der Erwartungswert mit einer langsameren Rate als $n^{-1/2}$ geschätzt wird. Die asymptotische Verteilung ist auch i.A. nicht mehr normal, s. etwa Taqqu (1975).

2. Schätzung der Autokorrelation

Autokorrelationen werden über die empirische Autokorrelation geschätzt. Es gibt zur asymptotischen Verteilung nur partielle Resultate. Für einen FARMA(0,d,0) Prozess erhält man:

- für $d \in (0, 1/4)$ ist $\hat{\rho}(k)$ asymptotisch normalverteilt mit \sqrt{n} Rate und der Kovarianz aus Bartlett's Formel.
- für $d = 1/4$ ist $\hat{\rho}(k)$ asymptotisch normalverteilt mit $(n/\log(n))^{1/2}$ Rate,
- für $d \in (1/4, 1/2)$ hat $\hat{\rho}(k)$ die Rate $n^{1/2-d}$ und keine asymptotische Normalverteilung.

3. Schätzung in einem FARMA (p,d,q) Modell

Die Parameter $d, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ in einem FARMA(p,d,q) Modell können über Maximum Likelihood geschätzt werden. Der ML Schätzer ist asymptotisch normalverteilt mit \sqrt{n} -Rate und Kovarianzmatrix gegeben durch die Inverse der Fisher Informationsmatrix.

Da die Log-Likelihood schwer zu evaluieren ist (wegen der komplizierten Gestalt der Kovarianzfunktion eines FARMA(p,d,q) Modells, wurden auch Approximationen an die log-likelihood konstruiert und somit auch approximative ML Schätzer, die die gleiche Verteilung

haben. REFERENZEN!!!

4. *Schätzung des long-range Parameters*

Schließlich gibt es direkte semiparametrische Schätzer Parameters d . REFERENZ

3.4 Unit Root Tests

Eine grundsätzliche Frage bei der Untersuchung einer Zeitreihe ist die der Stationarität. Von Stationarität kann u.a. über deterministische Trends als auch über random walk Komponenten abgewichen werden.

In diesem Abschnitt wollen wir Tests auf random walk Komponenten betrachten.

Es gibt eine Reihe von statistischen Tests, die ausgehend von parametrischen ARIMA Modellen die Stationarität einer Zeitreihe untersuchen. Hier gibt es zwei Gruppen: Sogenannte Unit-root Tests, bei denen als Hypothese Nichtstationarität angenommen wird, sowie Stationaritätstests, bei denen die Hypothese Stationarität ist.

Zur Gruppe der Unit root Tests gehören der klassische Dickey Fuller Test, der augmented Dickey Fuller Tests, der Phillips-Perron Test und sowie weitere aktuellere Tests, auf die weiter unten noch eingegangen wird.

Zur Gruppe der Stationaritätstest gehört insbesondere der Kwiatkowski - Phillips Test.

Der klassische Dickey-Fuller Test

Dieser basiert auf einem AR(1) Modell.

a. *kein Drift oder Trend.* Das Modell ist

$$X_t = \phi X_{t-1} + \epsilon_t, \quad \epsilon_t \text{ u.i.v..} \quad (82)$$

Das Testproblem ist dann gegeben durch

$$H : \phi = 1 \quad \text{gegen} \quad K : |\phi| < 1. \quad (83)$$

Der KQ-Schätzer für ϕ sowie die zugehörige t-Statistik basierend auf X_0, \dots, X_n sind gegeben durch

$$\hat{\phi} = \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}, \quad t_\phi = \frac{\hat{\phi} - 1}{\hat{\sigma}} \left(\sum_{t=1}^n X_{t-1}^2 \right)^{1/2}, \quad (84)$$

wobei $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{t=1}^n (X_t - \hat{\phi} X_{t-1})^2$. Unter H haben sowohl $\hat{\phi}$ als auch t_ϕ ein nicht-standard Verhalten, wie der folgende Satz zeigt.

Satz 3.20

Für u.i.v. Fehler ϵ_t mit $E\epsilon_t = 0$, $E\epsilon_t^2 = \sigma^2 < \infty$ gilt unter der Hypothese H

$$n(\hat{\phi} - 1) \xrightarrow{d} \frac{1}{2} \frac{W(1)^2 - 1}{\int_0^1 W(r)^2 dr}, \quad (85)$$

sowie

$$t_\phi \xrightarrow{d} \frac{1}{2} \frac{W(1)^2 - 1}{\left(\int_0^1 W(r)^2 dr \right)^{1/2}}, \quad (86)$$

wobei $(W(r), 0 \leq r \leq 1)$ eine standard Brownsche Bewegung ist.

Bemerkung 3.21

- a. Der Schätzer $\hat{\phi}$ hat nach (85) unter H eine n -Rate, ist also super-efficient (schnellere Rate als das parametrische \sqrt{n}), und hat keine asymptotische Normalverteilung.
- b. Die t-Statistik t_ϕ konvergiert nach (86) in Verteilung, jedoch ist sie nicht asymptotisch normalverteilt sondern hat eine nicht-standard Grenzverteilung.
- c. Man kann den Zähler in (85) und (86) auch umschreiben mit $(W(1)^2 - 1)/2 = \int_0^1 W(r) dW(r)$.
- d. In der Originalarbeit von Dickey und Fuller (198?) haben die Autoren eine andere (äquivalente) Darstellung der Grenzverteilung erhalten. Die Darstellung in Satz 3.20 geht auf Phillips (1987) FRÜHERE REFERENZEN!!!! zurück.
- e. Unter der Alternative $K : |\phi| < 1$ ist $\sqrt{n}(\hat{\phi} - \phi)$ sowie t_ϕ asymptotisch normalverteilt GENAUER!!!.

Dickey Fuller Test

Man erhält nun den Dickey-Fuller Test in seiner einfachsten Form. Dieser ist offenbar ein einseitiger Test, da eine einseitige Alternative vorliegt. Sei q_α das α -Quantil der Grenzverteilung in (86), dann wird H zum asymptotischen Niveau $\alpha > 0$ verworfen, falls $t_\phi < q_\alpha$.

Wir kommen zum Beweis von Satz 3.20 und zeigen zunächst folgendes Lemma.

Lemma 3.22

Unter den Bedingungen von Satz 3.20 gilt

$$\left(n^{-2} \sum_{t=1}^n X_{t-1}^2, n^{-1} \sum_{t=1}^n X_{t-1} \epsilon_t \right)^T \xrightarrow{d} \sigma^2 \left(\int_0^1 W(r)^2 dr, (W(1)^2 - 1)/2 \right)^T. \quad (87)$$

Beweis (des Lemmas)

Sei $S_t = \sum_{j=1}^t \epsilon_j$, $S_0 = 0$,

$$Z_n(r) = \frac{1}{\sqrt{n\sigma}} S_{[nr]} = \frac{1}{\sqrt{n\sigma}} S_{j-1}, \quad \frac{j-1}{n} \leq r < \frac{j}{n}. \quad (88)$$

Dann besagt das Invarianzprinzip von Donsker

$$Z_n(r) \xrightarrow{d} W(r), \quad 0 \leq r \leq 1, \quad (89)$$

wobei die schwache Konvergenz sich auf $D[0, 1]$ bezieht. Mit $X_t = X_0 + S_t$ gilt

$$\begin{aligned} \frac{1}{n^2} \sum_{t=1}^T X_{t-1}^2 &= \frac{1}{n^2} \sum_{t=1}^T \left(\sum_{j=1}^{t-1} \epsilon_j + X_0 \right)^2 \\ &= \frac{1}{n^2} \sum_{t=1}^T \left(S_{t-1}^2 + 2X_0 S_{t-1} + X_0^2 \right) \\ &= \sigma^2 \sum_{t=1}^T \int_{(t-1)/n}^{t/n} \frac{1}{n\sigma^2} S_{[nr]}^2 dr + 2\sigma X_0 n^{-1/2} \sum_{t=1}^T \int_{(t-1)/n}^{t/n} \frac{1}{\sqrt{n\sigma}} S_{[nr]} dr + \frac{X_0^2}{n} \\ &= \sigma^2 \sum_{t=1}^T \int_{(t-1)/n}^{t/n} Z_n(r)^2 dr + 2\sigma X_0 n^{-1/2} \sum_{t=1}^T \int_{(t-1)/n}^{t/n} Z_n(r) dr + \frac{X_0^2}{n} \\ &= \sigma^2 \int_0^1 Z_n(r)^2 dr + O_P(n^{-1/2}). \end{aligned} \quad (90)$$

Analog ist mir $\bar{\epsilon}_n = \sum_{j=1}^n \epsilon_j/n$

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^T X_{t-1} \epsilon_t &= \frac{1}{n} \sum_{t=1}^T (S_{t-1} + X_0) \epsilon_t \\
&= \frac{1}{n} \sum_{t=1}^T S_{t-1} \epsilon_t + X_0 \bar{\epsilon}_n \\
&= \frac{1}{2n} \sum_{t=1}^T (S_t^2 - S_{t-1}^2 - \epsilon_t^2) + X_0 \bar{\epsilon}_n \\
&= \frac{\sigma^2}{2} \frac{S_n^2}{\sigma^2 n} - \frac{1}{2n} \sum_{t=1}^n \epsilon_t^2 + X_0 \bar{\epsilon}_n \\
&= \frac{\sigma^2}{2} (Z_n(1)^2 - 1) + o_P(1),
\end{aligned} \tag{91}$$

da $\bar{\epsilon}_n \rightarrow 0$ (P) und $n^{-1} \sum_{t=1}^n \epsilon_t^2 \rightarrow \sigma^2$ (P). Wir betrachten nun die stetigen Funktionen

$$G : D[0, 1] \rightarrow \mathbb{R}^2, \quad G(\phi) = \left(\int_0^1 \phi^2(t) dt, (\phi^2(1) - 1)/2 \right)^T. \quad \blacksquare$$

Dann gilt nach dem oben Bewiesenen und dem Stetigkeitssatz

$$\begin{aligned}
\left(n^{-2} \sum_{t=1}^n X_{t-1}^2, n^{-1} \sum_{t=1}^n X_{t-1} \epsilon_t \right)^T &= \sigma^2 G(Z_n(r)) + o_p(1) \\
&\xrightarrow{d} \sigma^2 \left(\int_0^1 W(r)^2 dr, (W(1)^2 - 1)/2 \right)^T
\end{aligned}$$

Beweis (von Satz 3.20)

Zu (85): Es ist

$$n(\hat{\phi} - 1) = \frac{n^{-1} \sum_{t=1}^n \epsilon_t X_{t-1}}{n^{-2} \sum_{t=1}^n X_{t-1}^2},$$

und da $\int_0^1 W(r)^2 dr > 0$ f.s., ist $H(x, y) = y/x$ f.s. stetig bzgl. des Grenzzaßes in (87), so dass (85) wiederum mit dem Stetigkeitssatz folgt.

Zu (86): Es ist

$$t_\phi = \frac{n^{-1} \sum_{t=1}^n \epsilon_t X_{t-1}}{(\sigma + o_P(1)) (n^{-2} \sum_{t=1}^n X_{t-1}^2)^{1/2}},$$

und die Behauptung folgt wiederum mit dem Stetigkeitssatz. \blacksquare

b. deterministischer Drift. a. Hier ist das Modell

$$X_t = \mu + \phi X_{t-1} + \epsilon_t, \quad \epsilon_t \text{ u.i.v.}, \tag{92}$$

wobei $\mu \in \mathbb{R}$ der deterministische Drift ist. Das Testproblem ist wiederum gegeben durch

$$H : \phi = 1 \quad \text{gegen} \quad K : |\phi| < 1.$$

Sei $\hat{\phi}_\mu$ der KQ-Schätzer von ϕ in (92) mit $t_{\phi,\mu}$ die entsprechende t-Statistik. Dann gilt der folgende Satz. Hier sei $(W(r), 0 \leq r \leq 1)$ eine standard Brownsche Bewegung und $W^*(r) = W(r) - \int_0^1 W(s) ds$.

Satz 3.23

Für u.i.v. Fehler ϵ_t mit $E\epsilon_t = 0$, $E\epsilon_t^2 = \sigma^2 < \infty$ gilt unter der Hypothese H

a. für $\mu = 0$

$$n(\hat{\phi}_\mu - 1) \xrightarrow{d} \frac{\int_0^1 W^*(r) dW(r)}{\int_0^1 W^*(r)^2 dr}, \quad (93)$$

sowie

$$t_{\phi,\mu} \xrightarrow{d} \frac{\int_0^1 W^*(r) dW(r)}{(\int_0^1 W^*(r)^2 dr)^{1/2}}, \quad (94)$$

b. für $\mu \neq 0$

$$t_{\phi,\mu} \xrightarrow{d} N(0, 1). \quad (95)$$

Bemerkung 3.24

a. Zunächst bemerken wir, dass sich Grenzverteilungen auch für $\mu = 0$ in (93) und (94) gegenüber denen in (85) und (86) unterscheiden WIE GENAU!!. Der Dickey Fuller Test hängt also stark von der Spezifizierung des Modells ab.

b. Auch unter der Hypothese $H : \phi = 1$ unterscheidet sich die Grenzverteilungen, je nachdem, ob $\mu = 0$ oder $\mu \neq 0$. Man kann zeigen, dass $N(0, 1)$ auf $(-\infty, 0)$ stochastisch größer ist als die Grenzverteilung in (94). Basiert man somit die Testentscheidung auf dem kritischen Wert q_α der Grenzverteilung in (94) (verwerfe also H , falls $t_{\phi,\mu} < q_\alpha$), so erhält man für $\mu \neq 0$ einen konservativen Test.

Dickey Fuller Test

Der Dickey Fuller Test erfolgt nun analog zum Fall ohne Drift. Es ist aber essentiell, die Testentscheidung auf der t-Statistik und nicht auf der Verteilung des Schätzers selbst zu basieren, denn dieser konvergiert im Fall $\mu = 0$ schneller gegen 1 als im Fall $\mu \neq 0$. Sei \tilde{q}_α das α -Quantil der Grenzverteilung in (94), dann wird H zum asymptotischen Niveau $\alpha > 0$ verworfen, falls $t_{\phi,\mu} < \tilde{q}_\alpha$.

c. *Drift und linearer Trend.* Hier ist das Modell

$$X_t = \mu + \beta t + \phi X_{t-1} + \epsilon_t, \quad \epsilon_t \text{ u.i.v.}, \quad (96)$$

wobei $\mu \in \mathbb{R}$ der deterministische Drift ist. Das Testproblem ist wiederum gegeben durch

$$H : \phi = 1 \quad \text{gegen} \quad K : |\phi| < 1.$$

Sei $\hat{\phi}_\beta$ der KQ-Schätzer von ϕ in (96) mit $t_{\phi,\beta}$ die entsprechende t-Statistik.

Satz 3.25

Für u.i.v. Fehler ϵ_t mit $E\epsilon_t = 0$, $E\epsilon_t^2 = \sigma^2 < \infty$ gilt unter der Hypothese H

a. für $\beta = 0$ (und μ beliebig)

$$n(\hat{\phi}_\beta - 1) \xrightarrow{d} A, \quad (97)$$

sowie

$$t_{\phi, \beta} \xrightarrow{d} B, \quad (98)$$

wobei die Grenzzufallsvariablen A und B in Phillips und Perron (1988) angegeben sind.

b. für $\beta \neq 0$ (und μ beliebig) ist

$$t_{\phi, \beta} \xrightarrow{d} N(0, 1). \quad (99)$$

Lemma 3.26

Der KQ Schätzer für ϕ in (96) ist unabhängig vom Wert von μ .

GENAUE FORMULIERUNG UND BEWEIS DES LEMMAS????

Das AR(1)-Modell im klassischen DF-Test ist in der Regel zu einfach, so dass das Modell, in dem auf unit root getestet wird, misspezifiziert ist. Somit sind auch die Testergebnisse nur approximativ gültig. Es gibt verschiedene Erweiterungen, die diese Problem überwinden. Wir besprechen den erweiterten Dickey-Fuller Test (ADF Test) sowie den Phillips-Perron Test (PP-Test).

Der erweiterte Dickey-Fuller Test (Augmented Dickey Fuller Test, ADF Test)

Man beobachtet X_0, \dots, X_n nach dem Modell $X_t = d_t + Z_t$, wobei d_t ein deterministischer Drift (also $d_t = 0$ oder $d_t = \mu$ oder $d_t = \mu + \beta t$), und Z_t als ARIMA Prozess modelliert wird. Das Testproblem im ADF-Test ist

$$H : (Z_t) \text{ folgt } ARIMA(r, 1, q) \text{ für ein (unbekanntes) Paar } (r, q),$$

gegen

$$K : (Z_t) \text{ folgt } ARMA(r + 1, q) \text{ für ein (unbekanntes) Paar } (r, q).$$

Man nimmt dabei nicht unbedingt an, dass man stationär startet, die Anfangsverteilung spielt für die Asymptotik jedoch keine Rolle.

Man approximiert nun die Zeitreihenstruktur für Z_t durch ein AR(p) Modell, und lässt die Ordnung p wachsen, um gültige Resultate auch für Reihen mit einem MA-Anteil zu erhalten. Wir setzen also an

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \epsilon_t.$$

Schreibe

$$Z_t = \rho Z_{t-1} + \sum_{i=1}^{p-1} a_i \Delta Z_{t-i} + \epsilon_t, \quad \Delta Z_{t-i} = Z_{t-i} - Z_{t-i-1},$$

so dass $\rho + a_1 = \phi_1$, $a_j - a_{j-1} = \phi_j$, $j = 2, \dots, p-1$, $\phi_p = -a_{p-1}$, oder

$$\rho = \sum_{j=1}^p \phi_j, \quad a_j = - \sum_{i=j+1}^p \phi_i.$$

Eine Einheitswurzel $\phi(1) = 0$ in dem AR(p) Polynom $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ ist äquivalent

zu $\rho = 1$. Wir schreiben nun

$$\begin{aligned} X_t &= d_t + \rho Z_{t-1} + \sum_{i=1}^{p-1} a_i \Delta Z_{t-i} + \epsilon_t \\ &= d_t - \rho d_{t-1} - \sum_{i=1}^{p-1} a_i \Delta d_{t-i} + \rho X_{t-1} + \sum_{i=1}^{p-1} a_i \Delta X_{t-i} + \epsilon_t. \end{aligned}$$

Subtraktion von X_{t-1} liefert

$$\Delta X_t = d_t - \rho d_{t-1} - \sum_{i=1}^{p-1} a_i \Delta d_{t-i} + (\rho - 1) X_{t-1} + \sum_{i=1}^{p-1} a_i \Delta X_{t-i} + \epsilon_t.$$

Man erhält nun die folgenden ADF-Regressionen:

Für $d_t = 0$ verschwinden alle Trend Terme.

Für $d_t = \mu$ erhält man mit $a = (1 - \rho)\mu$, $\alpha = \rho - 1$

$$\Delta X_t = a + \alpha X_{t-1} + \sum_{i=1}^{p-1} a_i \Delta X_{t-i} + \epsilon_t.$$

Für $d_t = \mu + \beta t$ erhält man mit $b = (1 - \rho)\mu + \rho\beta + \beta \sum_{i=1}^{p-1} a_i$, $c = (1 - \rho)\beta$, $\alpha = \rho - 1$

$$\Delta X_t = b + ct + \alpha X_{t-1} + \sum_{i=1}^{p-1} a_i \Delta X_{t-i} + \epsilon_t.$$

Man schätzt nun für festes p die Parameter $\alpha, a_1, \dots, a_{p-1}$ sowie gegebenenfalls noch a oder (b, c) durch KQ-Schätzer, und t_α sei die t-Statistik in jeweiligen Modell zur Hypothese $\alpha = 0$.

Wir kommen nun zur Asymptotik des ADF Tests.

Annahme A: Sei $p = p(n) \rightarrow \infty$ derart, dass $n^{-1/2}p \rightarrow 0$ und $cp > n^{1/r}$ für ein $c, r > 0$.

Satz 3.27

Angenommen, $d_t = 0$ und (X_t) liegt ein ARIMA(r,1,q) Modell mit u.i.v. Fehlern und Parametern ϕ_1, \dots, ϕ_r sowie $\theta_1, \dots, \theta_q$ zugrunde. Dann gilt unter der Annahme A für den KQ-Schätzer $\hat{\alpha}$ von α

$$\frac{1 + \theta_1 + \dots + \theta_q}{1 - \phi_1 - \dots - \phi_r} n \hat{\alpha} \xrightarrow{d} \frac{1}{2} \frac{W(1)^2 - 1}{\int_0^1 W(r)^2 dr}, \quad (100)$$

sowie

$$t_\alpha \xrightarrow{d} \frac{1}{2} \frac{W(1)^2 - 1}{\left(\int_0^1 W(r)^2 dr\right)^{1/2}}, \quad (101)$$

wobei $(W(r), 0 \leq r \leq 1)$ eine standard Brownsche Bewegung ist.

Bemerkung 3.28

a. Verteilung von $\hat{\alpha}$ hängt von unbekanntem Parametern ab.

b. Verteilung von t_α nicht, kann zum testen verwendet werden.

Weiter gelten analoge Versionen der Sätze 3.23 und 3.25. Insbesondere bleiben die Verteilungen der t-Statistiken gültig (also nuisance Parameter frei). REFERENZEN; MEHR TEXT; WIE IST ES GENAU????

Phillips-Perron Test

Während der ADF-Test innerhalb der ARIMA Klasse getestet, kann der Phillips-Perron Test auf allgemeinere Situationen angewendet werden.

a. *Kein deterministischer Trend.* Man setzt an wie im klassischen Dickey-Fuller Test in (82), nur dass man für die (ϵ_t) nur (starke) Stationariät fordert. Die Hypothese ist wie in (83) und Schätzer und Teststatistik sind wie in (84) gegeben.

Wir nehmen an, dass $E\epsilon_t = 0$, $E\epsilon_t^2 = \sigma_e^2 < \infty$ und dass (ϵ_t) ergodisch ist, so dass

$$\frac{1}{n} \sum_{t=1}^n \epsilon_t \rightarrow 0, \quad \frac{1}{n} \sum_{t=1}^n \epsilon_t^2 \rightarrow \sigma_e^2,$$

wobei beide Aussagen fast sicher, also insbesondere in Wahrscheinlichkeit gelten. Darüber hinaus nehmen wir an, dass der Grenzwert

$$\lim_{n \rightarrow \infty} \frac{1}{n} E \left(\sum_{t=1}^n \epsilon_t \right)^2 = \sigma^2 < \infty \quad (102)$$

existiert, und dass die Aussage des Donskerschen Invarianzprinzips gilt, also für den Polygonzug (88) (wobei hier das σ^2 aus (102) verwendet wird), die Grenzwertaussage (89) gilt. Dies ist unter allgemeinen Bedingungen für stationäre Prozesse der Fall, s. etwa REFERENZEN!!!

Satz 3.29

Unter den obigen Annahmen gilt

$$n(\hat{\phi} - 1) \xrightarrow{d} \frac{1}{2} \frac{W(1)^2 - \sigma_e^2/\sigma^2}{\int_0^1 W(r)^2 dr}, \quad (103)$$

sowie

$$t_\phi \xrightarrow{d} \frac{\sigma}{2\sigma_e} \frac{W(1)^2 - \sigma_e^2/\sigma^2}{\left(\int_0^1 W(r)^2 dr\right)^{1/2}}. \quad (104)$$

Beweis

Wir skizzieren den Beweis. Die Rechnung (90) bleibt unverändert gültig, während wir in (91) erhalten

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^T X_{t-1} \epsilon_t &= \frac{\sigma^2}{2} \frac{S_n^2}{\sigma^2 n} - \frac{1}{2n} \sum_{t=1}^n \epsilon_t^2 + X_0 \bar{\epsilon}_n \\ &= \frac{\sigma^2}{2} Z_n(1)^2 - \frac{\sigma_e^2}{2} + o_P(1). \end{aligned}$$

Mit dem Stetigkeitssatz erhalten wir (103) und (104). Für (104) benötigen wir noch, dass $\hat{\sigma}^2 \rightarrow \sigma_e^2$. Hierfür rechnen wir mit $\hat{\phi} \rightarrow 1$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-1} \sum_{t=1}^n (X_t^2 - 2\hat{\phi}X_tX_{t-1} + X_{t-1}^2)^2 \\ &= \frac{1}{n-1} \sum_{t=1}^n (X_t - X_{t-1})^2 + o_p(1) \\ &= \frac{1}{n-1} \sum_{t=1}^n \epsilon_t^2 + o_p(1) \rightarrow \sigma_e^2.\end{aligned}$$

Die Grenzverteilungen in (103) und (104) hängen von nuisance Parametern ab. Diese können zwar geschätzt werden, aber die Grenzverteilungen müssten für jeden möglichen Wert des Quotienten tabularisiert werden. Daher geht man zu Statistiken über, deren Grenzverteilung asymptotisch nicht von nuisance Parametern abhängt.

Korollar 3.30

Seien $\hat{\sigma}^2$ und $\hat{\sigma}_e^2$ konsistente Schätzer für σ^2 und σ_e^2 . Dann gilt unter $H : \phi = 1$ für die modifizierte t Statistik

$$Z_t = \frac{\hat{\phi} - 1}{\hat{\sigma}(\sum_{t=1}^n X_{t-1}^2)^{1/2}} - \frac{1}{2} \frac{\hat{\sigma}^2 - \hat{\sigma}_e^2}{\hat{\sigma}(n^{-2} \sum_{t=1}^n X_{t-1}^2)^{1/2}}.$$

Beweis

Zunächst konvergiert

$$\frac{\hat{\phi} - 1}{\hat{\sigma}(\sum_{t=1}^n X_{t-1}^2)^{1/2}} \xrightarrow{d} \frac{1}{2} \frac{W(1)^2 - \sigma_e^2/\sigma^2}{(\int_0^1 W(r)^2 dr)^{1/2}}.$$

Weiter ist

$$\hat{\sigma}(n^{-2} \sum_{t=1}^n X_{t-1}^2)^{1/2} \xrightarrow{d} \sigma^2 \left(\int_0^1 W(r)^2 dr \right)^{1/2}.$$

Dies ergibt nun leicht die Behauptung. ■

b. Trend Fälle. Man kann analoge Resultate auch in den Fällen eines deterministischen Drifts oder Trends erhalten, vgl. Phillips und Perron (198?). Man muss wiederum die Statistiken geeignet modifizieren, um Grenzverteilungen zu erhalten, die nicht von den Nuisance Parametern σ^2 und σ_e^2 anhängen. Allerdings ist deren Verhalten nur für $\mu = 0$ (Drift Fall) bzw. $\beta = 0$ (Trend Fall) explizit bekannt. Was unter $\mu \neq 0$ (Drift Fall) bzw. $\beta \neq 0$ (Trend Fall) passiert, scheint nicht bekannt zu sein.

Tests auf saisonale Unit roots

Bisher haben wir die Hypothese getestet, ob eine nicht stationäre Zeitreihe vorliegt, deren Differenzen (erster Ordnung) stationär sind. Es gibt auch Tests auch saisonale Unit Roots,

also darauf, ob das saisonale Muster einer Zeitreihe stationär ist oder eine random walk Komponente hat. Hier seien die Arbeiten von Hylleberg, Engle, Granger und Yoo (1990) sowie F. Canova and B.E. Hansen (1995), in denen derartige Tests vorgeschlagen werden, erwähnt.

Tests auf Stationarität

Falls man mit einem Unit root Test die Hypothese der Unit root nicht verwerfen kann, so ist deren Existenz damit jedoch noch nicht bestätigt. Daher sind auch Tests von Interesse, deren Hypothese die Stationarität ist. Falls dies zugunsten einer Unit Root verworfen werden kann, hat man mit einem bestimmten Fehlerniveau eine Unit root nachgewiesen.

Ein derartiger Test wurde von Kiatkowski, Phillips, Schmidt und Shin (1992) (KPSS Test) vorgeschlagen. Das Modell lautet

$$X_t = \mu + \beta t + r_t + \epsilon_t, \quad t = 1, \dots, n,$$

wobei $\mu, \beta \in \mathbb{R}$, $r_0 = 0$ und $r_t = r_{t-1} + u_t$ mit u_t u.i.v. mit $Eu_t = 0$, $Eu_t^2 = \sigma_u^2 < \infty$ und (ϵ_t) eine stationäre Zeitreihe mit $E\epsilon_t = 0$, $E\epsilon_t^2 = \sigma_\epsilon^2 < \infty$. Dabei ist also r_t die random walk Komponente, diese verschwindet, falls $\sigma_u^2 = 0$. Somit lautet der Test:

$$H : \sigma_u^2 = 0 \quad \text{gegen} \quad K : \sigma_u^2 > 0.$$

MEHR; ASYMPTOTISCHE VETEILUNG; GENAUE ZUSAMMENHÄNGE!!!!!!!

3.5 Lineare Zeitreihenmodelle mit heavy tails

3.6 Literaturhinweise und Bemerkungen

4 Spektraltheorie

4.1 Zyklische Zeitreihen

Seien für $-\pi < \omega_1 < \dots < \omega_n \leq \pi$ $Z(\omega_1), \dots, Z(\omega_n)$ komplexwertige Zufallsvariablen mit $EZ(\omega_j) = 0$, $E|Z(\omega_j)|^2 = \sigma_j^2 < \infty$, $j = 1, \dots, n$, $E(Z(\omega_j)\overline{Z(\omega_k)}) = 0$, $j \neq k$. Dann heißt

$$X_t = \sum_{j=1}^n Z(\omega_j) e^{it\omega_j}, \quad t \in \mathbb{Z} \quad (105)$$

eine zyklische Zeitreihe. Offenbar gilt $EX_t = 0$,

$$\text{Cov}(X_{t+h}, X_t) = EX_{t+h}\overline{X_t} = \sum_{j=1}^n \sigma_j^2 e^{ih\omega_j} = \gamma(h),$$

insbesondere ist X_t schwach-stationär. Man kann also schreiben

$$\gamma(h) = \int_{(-\pi, \pi]} e^{i\omega h} d\mu(\omega), \quad \mu = \sum_{j=1}^n \sigma_j^2 \delta_{\omega_j}, \quad (106)$$

wobei δ_ω das Dirac-Maß bei ω bezeichnet. Die Formeln (105) und (106) sind Spezialfälle der Spektraldarstellung eines stationären Prozesses sowie seiner Autokorrelationsfunktion.

Reelle zyklische Zeitreihen

Die Zeitreihe (105) ist reellwertig, falls

- $\omega_j = \overline{\omega_{n-j}}$ für $\omega_j, \omega_{n-j} \neq \pi$, sowie $Z(\omega_j) = \overline{Z(\omega_{n-j})}$,
- für $\omega_n = \pi$ ist $\Im(Z(\pi)) = 0$.

Zu a.: Schreibe $Z(\omega_j) = C(\omega_j) + iD(\omega_j)$, dann

$$Z(\omega_j) e^{i\omega_j t} + Z(\omega_{n-j}) e^{i\omega_{n-j} t} = 2(C(\omega_j) \cos(\omega_j t) - D(\omega_j) \sin(\omega_j t)),$$

so dass (die 2 geht in der Summe sein)

$$X_t = \sum_{j=1}^n (C(\omega_j) \cos(\omega_j t) - D(\omega_j) \sin(\omega_j t)) \quad (107)$$

Wegen

$$\sigma_j^2 = E|Z(\omega_j)|^2 = EC(\omega_j)^2 + ED(\omega_j)^2$$

sowie

$$0 = EZ(\omega_j)\overline{Z(\omega_{n-j})} = EZ(\omega_j)^2 = EC(\omega_j)^2 - ED(\omega_j)^2 + 2iEC(\omega_j)D(\omega_j),$$

gilt für reelle zyklische Zeitreihen in der Darstellung (107)

$$EC(\omega_j)^2 = ED(\omega_j)^2 = \sigma_j^2, \quad EC(\omega_j)D(\omega_j) = 0.$$

Weiter kann man aus der Unkorreliertheit der $Z(\omega_j)$ und a. noch die Unkorreliertheit aller $C(\omega_j), D(\omega_k)$ (auch für verschiedene j, k) folgern.

4.2 Das Spektralmaß

Lemma 4.1

Sei μ ein endliches Borel-Maß auf $[-\pi, \pi]$. Dann liegt der Raum $V = \text{span}\{e^{ih\omega}, h \in \mathbb{Z}\}$ (trigonometrische Polynome) dicht in $L_2^{\mathbb{C}}(\mu)$, und μ ist durch die Werte

$$\int_{-\pi}^{\pi} e^{ih\omega} d\mu(\omega) \quad (108)$$

eindeutig bestimmt.

Beweis

Die trigonometrischen Polynome V liegen dicht in den stetigen, 2π -periodischen Funktionen $C_{per}[-\pi, \pi]$ bzgl. der max-Norm $\|\cdot\|_{[-\pi, \pi]}$ (Weierstrassscher Approximationssatz). Weiter ist $C_{per}[-\pi, \pi]$ dicht in $L_2^{\mathbb{C}}(\mu)$ bzgl. der $L_2 - \mu$ Norm. Dann ist V auch dicht in $L_2^{\mathbb{C}}(\mu)$, da μ endlich ist. Somit wird ein Maß auch eindeutig durch seine Werte (108) festgelegt, da Integration stetig bzgl. der L_2 Norm. ■

Bemerkungen: positiv semidefinit = Autokovarianzfunktion.

Satz 4.2 (Herglotz, Existenz des Spektralmaßes)

Sei $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ positiv semidefinit. Dann existiert genau ein endliches Maß μ auf $(-\pi, \pi]$, so dass

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\omega} d\mu(\omega), \quad h \in \mathbb{Z} \quad (109)$$

Umgekehrt ist jede Funktion der Form (109) positiv semidefinit. Das Maß μ heißt Spektralmaß von γ . Hat μ eine stetige Dichte f bzgl. $\lambda_{(-\pi, \pi]}$, (Lebesgue Maß), so heißt f Spektraldichte von γ .

Beweis

a. Funktionen der Form (109) sind positiv semidefinit:

$$\begin{aligned} \sum_{j,k=1}^r \lambda_j \bar{\lambda}_k \gamma(t_j - t_k) &= \sum_{j,k=1}^r \lambda_j \bar{\lambda}_k \int_{(-\pi, \pi]} e^{i\omega(t_j - t_k)} d\mu(\omega) \\ &= \int_{(-\pi, \pi]} \sum_{j,k=1}^r \lambda_j \bar{\lambda}_k e^{i\omega(t_j - t_k)} d\mu(\omega) \\ &= \int_{(-\pi, \pi]} \left| \sum_{j=1}^r \lambda_j e^{i\omega t_j} \right|^2 d\mu(\omega) \geq 0. \end{aligned}$$

b. Existenz des Spektralmaßes μ : Wir können $\gamma(0) > 0$ annehmen, da sonst $\gamma(h) = 0$ für alle h (da Kovarianzfunktion, somit Cauchy Schwarz), und dann kann $\mu = 0$ gesetzt werden. Setzte $\rho(k) = \gamma(k)/\gamma(0)$, $f_n : [-\pi, \pi] \rightarrow \mathbb{C}$ mit

$$\begin{aligned} f_n(\omega) &= \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} (1 - |h|/n) \gamma(h) e^{-ih\omega} \\ &= \frac{1}{2\pi} \sum_{j,k=1}^n \gamma(j-k) e^{-ij\omega} e^{ik\omega} \geq 0, \end{aligned} \quad (110)$$

da γ positiv semidefinit, wobei (15) verwendet wurde. Man berechnet

$$\begin{aligned} \int_{-\pi}^{\pi} f_n(\omega) e^{ih\omega} d\omega &= \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} (1 - |k|/n) \gamma(k) \int_{-\pi}^{\pi} e^{i(h-k)\omega} d\omega \\ &= \begin{cases} (1 - |h|/n) \gamma(h), & |h| \leq n-1, \\ 0, & \text{sonst.} \end{cases} \end{aligned} \quad (111)$$

Insbesondere ist $\int_{-\pi}^{\pi} f_n(\omega) d\omega = \gamma(0)$, daher definiert

$$\mu_n(B) = \frac{1}{\gamma(0)} \int_B f_n(\omega) d\omega, \quad B \in [-\pi, \pi],$$

eine Folge von W-Maßen auf der kompakten Menge $[-\pi, \pi]$, und daher existiert nach Prohorov eine schwach-konvergente Teilfolge $\mu_{n_k} \rightarrow \mu_0$, $k \rightarrow \infty$. Dann gilt nach Definition der schwachen Konvergenz und (111)

$$\begin{aligned} \int_{[-\pi, \pi]} e^{ih\omega} d\mu_0(\omega) &= \lim_k \int_{[-\pi, \pi]} e^{ih\omega} d\mu_{n_k}(\omega) \\ &= \lim_{k \rightarrow \infty, n_k \geq |h|} \left(1 - \frac{|h|}{n_k}\right) \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(h)}{\gamma(0)}. \end{aligned}$$

Setzt man nun $\mu(B) = (\mu(B) + \mu_0(-\pi)1_{\pi}(B))\gamma(0)$, $B \subset (-\pi, \pi]$, so erfüllt μ (109).

c. Eindeutigkeit: Dies folgt direkt aus der Bedingung (109) und Lemma 4.1. ■

Satz 4.3 (Inversionsformel)

Sei $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ positiv semidefinit mit

$$\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty. \quad (112)$$

Dann ist

$$f(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) e^{-i\omega h} \quad (113)$$

die stetige Spektraldichte von γ .

Beweis

a. Stetigkeit von f : Die Partialsummen in (113) sind stetig, und wegen (112) und $|e^{-i\omega h}| = 1$ konvergiert die Reihe in (113) gleichmäßig, also ist auch der Limes stetig.

b. $f \geq 0$: Für die f_n in (110) gilt $f_n(\omega) \geq 0$, und

$$f(\omega) - f_n(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \frac{|k|}{n} \gamma(k) e^{-ik\omega} + \frac{1}{2\pi} \sum_{|h| \geq n} \gamma(h) e^{-i\omega h} = I + II.$$

Der Term $II \rightarrow 0$, $n \rightarrow \infty$, da die Reihe in (113) konvergiert. Zu I : Die Folgen

$$\left(\frac{|k|}{n} \gamma(k) e^{-ik\omega} 1_{|k| \leq n-1} \right)_{k \in \mathbb{Z}}$$
■

konvergieren gegen 0 für jeden festen Index k und für $n \rightarrow \infty$, und werden majorisiert durch $(|\gamma(k)|)_{k \in \mathbb{Z}}$. Da diese Folge nach (112) summierbar ist, folgt mit dem Satz von der majorisierten Konvergenz für das Zählmaß auf \mathbb{Z} die Konvergenz $I \rightarrow 0$, $n \rightarrow \infty$. Also konvergiert punktweise $f_n(\omega) \rightarrow f(\omega)$, und da $f_n(\omega) \geq 0$ für alle n , folgt auch $f(\omega) \geq 0$.

c. f Spektraldichte: Wir rechnen die Eigenschaft (109) nach. Mit dem Satz von der majorisierten Konvergenz für das Lebesgue Maß auf $[-\pi, \pi]$, sowie (112), erhalten wir

$$\begin{aligned} \int_{(-\pi, \pi]} f(\omega) e^{ih\omega} &= \frac{1}{2\pi} \int_{(-\pi, \pi]} \sum_{k \in \mathbb{Z}} \gamma(k) e^{i\omega(h-k)} d\omega \\ &= \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma(k) \int_{(-\pi, \pi]} e^{i\omega(h-k)} d\omega = \gamma(h). \end{aligned}$$

Bemerkung 4.4

Die Sätze 4.2 und 4.3 sind klassische Resultate der harmonischen Analysis auf kompakten Abelschen Gruppen (hier speziell für S^1 formuliert). Das Haarsche Maß ersetzt dann allgemein das Lebesgue-Maß, und die Charaktere der Gruppe die Funktionen $e^{it\omega}$.

Beispiel 4.5

In den Beispielen berechnen wir jeweils die Spektraldichte f der Autokorrelationsfunktion.

a. Weißes Rauschen. In (113) haben wir $\rho(h) = 0$ für $h \neq 0$, also $f(\omega) = 1/(2\pi)$.

b. AR(1). Es ist $X_t = \phi X_{t-1} + e_t$, und $\rho(h) = \phi^{|h|}$. Somit

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \phi^{|h|} e^{-ih\omega} = \frac{1}{2\pi} \left(1 + \sum_{h \geq 1} (\phi e^{-i\omega})^h + \sum_{h \geq 1} (\phi e^{i\omega})^h \right) \\ &= \frac{1}{2\pi} \left(\frac{1}{1 + \phi e^{-i\omega}} + \frac{1}{1 + \phi e^{i\omega}} - 1 \right) = \frac{1}{2\pi} \left(\frac{1 + \phi e^{-i\omega} + 1 + \phi e^{i\omega} - |1 + \phi e^{i\omega}|^2}{|1 + \phi e^{i\omega}|^2} \right) \\ &= \frac{1}{2\pi} \frac{1 - \phi^2}{1 + 2\phi \cos(\omega) + \phi^2}. \end{aligned}$$

Die Dichte f hat das Maximum bei $\omega = \pi$ falls $-1 < \phi < 0$, und bei $\omega = 0$ falls $1 > \phi > 0$.

c. Spezielles AR(2): Betrachte $X_t = \phi X_{t-2} + e_t$, dann ist $\rho(h) = \phi^{|h|/2}$ für h gerade, und $\rho(h) = 0$ für h ungerade. Man berechnet

$$f(\omega) = \frac{1}{2\pi} \frac{1 - \phi^2}{1 + 2\phi \cos(2\omega) + \phi^2}.$$

Für $1 > \phi > 0$ hat f ein Maximum bei $\pi/2$.

d. MA-Reihen: Sei $X_t = \sum_{j=0}^q \theta_j e_{t-j}$, wobei $\theta_0 = 1$. Dann ist

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{\sum_{j=0}^q \theta_j^2}, & h = 0, \dots, q \\ 0, & h > q. \end{cases}$$

Somit ist

$$f(\omega) = \frac{1}{2\pi} \sum_{|h| \leq q} \rho(h) e^{-i\omega h} = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{h=1}^q \rho(h) \cos(\omega h)$$

Für MA(1) ist etwa

$$f(\omega) = \frac{1}{2\pi} + \frac{1}{\pi} \frac{\theta}{1 + \theta^2} \cos(\omega)$$

4.3 Nichtparametrische Spektraldichteschätzung

4.4 Der Spektralsatz

Stochastische Integration

Definition 4.6 (Maß mit orthogonalen Werten)

Sei (X, \mathcal{B}) ein Messraum und (H, \langle, \rangle) ein komplexer Hilbertraum. Dann heißt eine Abbildung $Z : \mathcal{B} \rightarrow H$ ein *Maß mit orthogonalen Werten* (MOW), falls

$$\mu_Z(B) = \|Z(B)\|^2, \quad B \in \mathcal{B}, \quad (114)$$

ein (dann endliches) Maß ist, für das

$$\langle Z(B_1), Z(B_2) \rangle = \mu_Z(B_1 \cap B_2), \quad B_1, B_2 \in \mathcal{B}. \quad (115)$$

Bemerkung 4.7

- a. Aus (115) folgt: Ist $B_1 \cap B_2 = \emptyset$, so ist $\langle Z(B_1), Z(B_2) \rangle = 0$.
 b. Sind $(B_j)_{j \geq 1}$ paarweise disjunkt, so gilt

$$Z\left(\bigcup_{j \geq 1} B_j\right) = \sum_{j=1}^{\infty} Z(B_j). \quad (116)$$

Beweis zu (116) Zunächst zeige für $B_1 \cap B_2 = \emptyset$

$$Z(B_1 \cup B_2) = Z(B_1) + Z(B_2).$$

Dazu betrachte $Y = Z(B_1 \cup B_2) - Z(B_1) - Z(B_2)$. Dann ist mit (115)

$$\begin{aligned} \|Y\|^2 &= \|Z(B_1 \cup B_2)\|^2 + \|Z(B_1)\|^2 + \|Z(B_2)\|^2 \\ &\quad - 2(\langle Z(B_1 \cup B_2), Z(B_1) \rangle + \langle Z(B_1 \cup B_2), Z(B_2) \rangle - \langle Z(B_2), Z(B_1) \rangle) \\ &= \mu_Z(B_1 \cup B_2) + \mu_Z(B_1) + \mu_Z(B_2) - 2\mu_Z(B_1) - 2\mu_Z(B_2) + 0 = 0, \end{aligned}$$

da μ_Z endlich additiv ist. Somit ist für alle $n \geq 1$

$$Z\left(\bigcup_{j \geq 1} B_j\right) = \sum_{j=1}^n Z(B_j) + Z\left(\bigcup_{j \geq n+1} B_j\right).$$

Da

$$\left\|Z\left(\bigcup_{j \geq n+1} B_j\right)\right\|^2 = \mu_Z\left(\bigcup_{j \geq n+1} B_j\right) = \sum_{j \geq n+1} \mu_Z(B_j) \rightarrow 0.$$

Also konvergieren die Partialsummen in (116).

Satz 4.8 (Fortsetzungssatz für isometrische Abbildungen)

Es seien $(H_1, \langle, \rangle_1)$, $(H_2, \langle, \rangle_2)$ komplexe Hilberträume, $A \subset H_1$, $\psi : A \rightarrow H_2$ mit

$$\langle \psi(a), \psi(b) \rangle_2 = \langle a, b \rangle_1, \quad a, b \in A. \quad (117)$$

Sei $V = \overline{\text{lin}}(A)$. Dann existiert genau eine lineare Abbildung $\tilde{\psi} : V \rightarrow H_2$, mit

$$\langle \tilde{\psi}(a), \tilde{\psi}(b) \rangle_2 = \langle a, b \rangle_1, \quad a, b \in V,$$

und $\tilde{\psi}|_A = \psi$. Man nennt dann $\tilde{\psi}$ die isometrische Fortsetzung von ψ , und bezeichnet diese auch wieder mit ψ .

Beweisskizze. Definiere zunächst $\tilde{\psi}$ auf $\text{lin}(A)$. Für $v = \sum_{j=1}^n \lambda_j a_j$, $\lambda_j \in \mathbb{C}$, $a_j \in A$, setze

$$\tilde{\psi}(v) = \sum_{j=1}^n \lambda_j \psi(a_j). \quad (118)$$

Zeige nun der Reihe nach mit Hilfe von (117):

- $\tilde{\psi}$ ist in dieser Darstellung auf $\text{lin}(A)$ wohldefiniert, also unabhängig von der speziellen Darstellung von v ,
- linear und isometrisch auf $\text{lin}(A)$,
- setze durch Stetigkeit auf $\overline{\text{lin}(A)}$ fort, dann wieder wohldefiniert.

Andererseits muss $\tilde{\psi}$ (118) erfüllen, und ist dann durch Stetigkeit eindeutig bestimmt.

Lemma 4.9

- Für ein MOW $Z : \mathcal{B} \rightarrow H$ wird durch

$$\psi_Z(1_B) := Z(B), \quad B \in \mathcal{B}, \quad (119)$$

eindeutig eine Isometrie $\psi_Z : L_2^{\mathbb{C}}(X, \mathcal{B}, \mu_Z) \rightarrow H$ definiert.

- Ist $\psi_Z : L_2^{\mathbb{C}}(X, \mathcal{B}, \mu_Z) \rightarrow H$ eine Isometrie, so definiert $Z(B) := \psi(1_B)$, $B \in \mathcal{B}$, ein MOW mit $\mu_Z = \mu$.

Beweis

- Es ist

$$\begin{aligned} \langle 1_{B_1}, 1_{B_2} \rangle_{\mu_Z} &= \int 1_{B_1} 1_{B_2} d\mu_Z \\ &= \mu_Z(B_1 \cap B_2) = \langle Z(B_1), Z(B_2) \rangle. \end{aligned}$$

Da die Indikatorfunktionen dicht in $L_2^{\mathbb{C}}(X, \mathcal{B}, \mu_Z)$ sind, folgt die Behauptung mit dem Fortsetzungssatz 4.8.

- Es ist $\|Z(B)\|^2 = \|\psi(1_B)\|^2 = \int 1_B^2 d\mu = \mu(B)$. Eine analoge Rechnung liefert die gewünschte Gleichung für das Skalarprodukt. ■

Definition 4.10

Ist $Z : \mathcal{B} \rightarrow H$ ein MOW und ψ_Z die in Lemma 4.9 a. definierte Isometrie, so heißt für $f \in L_2^{\mathbb{C}}(X, \mathcal{B}, \mu_Z)$

$$\psi_Z(f) = \int_X f dZ$$

stochastisches Integral von f bzgl. Z .

Lemma 4.11 (Rechenregeln für das stochastische Integral)

Für $f, g, f_n \in L_2^{\mathbb{C}}(X, \mathcal{B}, \mu_Z)$, $a, b \in \mathbb{C}$ gelten

- $\int (af + bg) dZ = a \int f dZ + b \int g dZ$.
- $f_n \rightarrow f$ in $L_2^{\mathbb{C}}(X, \mathcal{B}, \mu_Z) \Leftrightarrow \int f_n dZ \rightarrow \int f dZ$ in H .
- $\langle \int f dZ, \int g dZ \rangle_H = \langle f, g \rangle_{\mu_Z}$.

$$4. \quad \|\int f dZ\|_H^2 = \|f\|_{\mu_Z}^2.$$

Das Lemma folgt unmittelbar aus der Definition, da ψ_Z eine Isometrie ist.

Bemerkung 4.12 (Zusammenhang zur stochastischen Integration)

Sei $Z(\omega)_{\omega \in [-\pi, \pi]}$ ein zentrierter, quadratintegrierbarer, komplexwertiger Prozess mit orthogonalen Inkrementen

$$E(Z(\omega_2) - Z(\omega_1)) \overline{(Z(\omega_4) - Z(\omega_3))} = 0, \quad -\pi \leq \omega_1 < \omega_2 < \omega_3 < \omega_4 \leq -\pi$$

und rechtsstetigen Pfaden in L_2 :

$$E|Z(\omega + \delta) - Z(\omega)|^2 \rightarrow 0, \quad \delta \searrow 0, \quad \omega \in [-\pi, \pi].$$

Ein Beispiel wäre ein quadratintegrierbares Martingal mit rechtsstetigen Pfaden. Dann kann

$$\mu_Z((\omega_1, \omega_2]) = E|Z(\omega_2) - Z(\omega_1)|^2$$

zu einem Maß μ_Z auf $(-\pi, \pi]$ fortgesetzt werden. Die Abbildung

$$\psi(1_{(-\omega_1, \omega_2]}) = Z(\omega_2) - Z(\omega_1)$$

kann zu einer Isometrie $\psi : L_2((-\pi, \pi], \mu_Z) \rightarrow L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ fortgesetzt werden. Mit Lemma 4.9 b. erhält man dann ein MOW, dessen stochastisches Intergal gleich dem über ψ definierten stochastischen Intergal bzgl. des Prozesses $Z(\omega)$ ist.

Spektralsatz

Satz 4.13 (Spektralsatz für stationäre Prozesse)

Sei $(X_t)_{t \in \mathbb{Z}}$ ein stationärer, zentrierter, komplexwertiger Prozess mit Kovarianzfunktion γ und Spektralmaß μ . Dann definiert

$$\psi : e^{it \cdot} \rightarrow X_t, \quad t \in \mathbb{Z}, \quad (120)$$

eindeutig eine Isometrie $\psi : L_2^{\mathbb{C}}((-\pi, \pi], \mu) \rightarrow L_2(\Omega, \mathcal{A}, P)$. Weiter definiert $Z(B) = \psi(1_B)$ ein MOW mit $\mu_Z = \mu$ und

$$X_t = \int_{(-\pi, \pi]} e^{it\omega} dZ(\omega). \quad (121)$$

Das MOW $Z = Z_X$ ist durch $\mu_Z = \mu$ und (121) eindeutig bestimmt.

Beweis

$$\begin{aligned} \langle e^{it_1 \cdot}, e^{it_2 \cdot} \rangle_{L_2^{\mathbb{C}}(\mu)} &= \int_{(-\pi, \pi]} e^{i\omega(t_1 - t_2)} \mu(\omega) \\ &= \gamma(t_1 - t_2) = EX_{t_1} \overline{X_{t_2}}. \end{aligned}$$

Da $(e^{it \cdot})_{t \in \mathbb{Z}}$ dicht in $L_2^{\mathbb{C}}((-\pi, \pi], \mu)$ nach Lemma 4.1, kann ψ auf ganz $L_2^{\mathbb{C}}((-\pi, \pi], \mu)$ fortgesetzt werden. Dann ist Z ein MOW nach Lemma 4.9 b. Die Gleichung (121) ergibt sich aus der Eindeutigkeit der Fortsetzung von ψ , ausgehend von den Indikatorfunktionen. Zur Eindeutigkeit von $Z = Z_X$: da die Funktionen $(e^{it \cdot})_{t \in \mathbb{Z}}$ dicht in $L_2^{\mathbb{C}}((-\pi, \pi], \mu)$ sind, ist die Isometrie $f \rightarrow \int f dZ$ für alle $f \in L_2^{\mathbb{C}}((-\pi, \pi], \mu)$ eindeutig bestimmt, und dann auch auf den Indikatorfunktionen, was Z selbst eindeutig festlegt. ■

Bemerkung 4.14 (Stochastische Integral ist zentriert)

Sei $(X_t)_{t \in \mathbb{Z}}$ ein stationärer, zentrierter, komplexwertiger Prozess mit Kovarianzfunktion γ und Spektralmaß μ . Dann ist für $f \in L_2^{\mathbb{C}}((-\pi, \pi], \mu)$ die Zufallsvariable $\int_{(-\pi, \pi]} f dZ_X$ stets zentriert.

Beweis

Für endliche Linearkombinationen von Funktionen $(e^{it \cdot})_{t \in \mathbb{Z}}$ gilt dies wegen (121), da die X_t zentriert sind. Da $V = \text{span}\{(e^{it \cdot})_{t \in \mathbb{Z}}\}$ dicht in $L_2^{\mathbb{C}}((-\pi, \pi], \mu)$, existiert zu $f \in L_2^{\mathbb{C}}((-\pi, \pi], \mu)$ eine Folge $f_n \in V$ mit $f_n \rightarrow f$ und $E \int f_n dZ = 0$. Wegen

$$\int f_n dZ \rightarrow \int f dZ \quad \text{in } L_2^{\mathbb{C}}(\Omega, \mathring{A}, P).$$

folgt auch $0 = E \int f_n dZ \rightarrow E \int f dZ$. ■

4.5 Filterung

Satz 4.15 (Substitutionsregel für stochastische Integrale)

Sei $Z : \mathcal{B} \rightarrow H$ ein MOW, $g \in L_2(X, \mathcal{B}, \mu_Z)$. Dann definiert

$$Z_1(B) = \int_X g(\omega) 1_B(\omega) dZ(\omega), \quad B \in \mathcal{B},$$

ein MOW mit

$$\mu_{Z_1}(B) = \int_B |g|^2 d\mu_Z, \quad (122)$$

und für $f \in L_2(X, \mathcal{B}, \mu_{Z_1})$ gilt $fg \in L_2(X, \mathcal{B}, \mu_Z)$ und

$$\int f dZ_1 = \int fg dZ. \quad (123)$$

Beweis

a. Zeige: Z_1 ist ein MOW mit μ_{Z_1} definiert nach (122). Dann ist

$$\begin{aligned} \langle Z_1(B_1), Z_1(B_2) \rangle &= \left\langle \int g 1_{B_1} dZ, \int g 1_{B_2} dZ \right\rangle \\ &= \left\langle \int_X g 1_{B_1} \bar{g} 1_{B_2} d\mu_Z, \int_{B_1 \cap B_2} |g|^2 d\mu_Z \right\rangle. \end{aligned}$$

Da (122) ein Maß auf (X, \mathcal{B}) definiert, folgt hiermit, dass Z_1 ein MOW mit μ_{Z_1} der Form (123) ist.

b. Die Gleichung (123) gilt wegen der Linearität des stochastischen Integrals bzgl. Z_1 und Z für Linearkombinationen von Indikatorfunktionen. Seien f_n Elementarfunktionen, $f_n \rightarrow f$ in $L_2(X, \mathcal{B}, \mu_{Z_1})$, also $\int f_n dZ_1 \rightarrow \int f dZ_1$. Da mit (122) folgt

$$\int_X |f_n - f|^2 d\mu_{Z_1} = \int_X |f_n - f|^2 |g|^2 d\mu_Z,$$

also $f_n g \rightarrow fg$ in $L_2(X, \mathcal{B}, \mu_Z)$, und somit auch $\int f_n dZ_1 = \int f_n g dZ \rightarrow \int fg dZ$. ■

Bemerkung 4.16

Ist $|g| > 0$ μ_Z -f.s., so gilt auch

$$Z(B) = \int 1_B \frac{1}{g} dZ_1.$$

Beweis

Setze $g^{-1} = 0$ für $g = 0$, dann ist $gg^{-1} = 1$ μ_Z -f.s., und

$$\int |g^{-1} 1_B|^2 d\mu_{Z_1} = \int |g^{-1}|^2 |g|^2 1_B d\mu_Z = \mu_Z(B) < \infty, \quad B \in \mathcal{B},$$

also $g^{-1} \in L_2(X, \mathcal{B}, \mu_{Z_1})$, und somit folgt die Behauptung mit (123). ■

Definition 4.17 (Filter)

Sei $(Y_t)_{t \in \mathbb{Z}}$ eine zentrierte, stationäre, komplexe Zeitreihe mit Spektralmaß μ_Y und MOW Z_Y . Für eine *Transferfunktion* (oder *Filter*) $\varphi \in L_2(\mu_Y)$ setze

$$X_t = \int_{(-\pi, \pi]} e^{it\omega} \varphi(\omega) dZ_Y, \quad t \in \mathbb{Z}, \quad (124)$$

die *gefilterte Zeitreihe*. Die Funktion $|\varphi|^2$ heißt dann *Filterfunktion*.

Satz 4.18

Die gefilterte Zeitreihe (124) ist stationär, zentriert mit Autokorrelation

$$\gamma_X(h) = \int_{(-\pi, \pi]} e^{ih\omega} |\varphi(\omega)|^2 d\mu(\omega), \quad h \in \mathbb{Z}, \quad (125)$$

MOW

$$Z_X(B) = \int 1_B \varphi dZ_Y, \quad B \subset (-\pi, \pi], \quad (126)$$

Spektralmaß

$$\mu_X(B) = \int_B |\varphi|^2 d\mu_Y, \quad B \subset (-\pi, \pi], \quad (127)$$

sowie Spektraldichte (falls Spektraldichte von f_Y von μ_Y existiert)

$$f_X = |\varphi|^2 f_Y. \quad (128)$$

Beweis

Die gefilterte Zeitreihe (124) ist zentriert nach Bemerkung 4.14.

$$\begin{aligned} \langle X_{t+h}, X_t \rangle &= \int \varphi(\omega) e^{i(t+h)\omega} \bar{\varphi}(\omega) e^{-it\omega} d\mu_Y(\omega) \\ &= \int e^{ih\omega} |\varphi|^2 d\mu_X(\omega), \end{aligned}$$

somit gelten (125) und (127).

Zu (126): Dies definiert nach Lemma 4.9 b. ein MOW, und nach (123)

$$\int e^{it\omega} dZ_X(\omega) = \int e^{it\omega} \varphi(\omega) dZ_Y(\omega) = X_t,$$

daher folgt (126), da das MOW einer stationären Zeitreihe diese Eigenschaft sowie sein zugeordnetes Maß eindeutig bestimmt ist, wobei (126) als zugeordnetes Maß offenbar (127) hat. ■

Bemerkung 4.19 (Hintereinanderschaltung von Filtern)

Hintereinanderschaltung von Filtern φ und χ ergibt Filter $\varphi\chi$.

Satz 4.20 (Filter in Reihendarstellung)

Für $\varphi \in L_2^{\mathbb{C}}(\mu_Y)$ der Form

$$\varphi(\omega) = \sum_{j \in \mathbb{Z}} c_j e^{-ij\omega}, \quad c_j \in \mathbb{C}, \quad (129)$$

gilt für (124)

$$X_t = \sum_{j \in \mathbb{Z}} c_j Y_{t-j}, \quad t \in \mathbb{Z}. \quad (130)$$

Beweis

$$\begin{aligned} X_t &= \int_{(-\pi, \pi]} \varphi(\omega) e^{it\omega} dZ_Y \\ &= \sum_{j \in \mathbb{Z}} c_j \int_{(-\pi, \pi]} e^{i\omega(t-j)} dZ_Y = \sum_{j \in \mathbb{Z}} c_j Y_{t-j} \end{aligned}$$

Bemerkung 4.21

a. Da $\text{lin}\{e^{it}, t \in \mathbb{Z}\}$ dicht in $L_2^{\mathbb{C}}(\mu_Y)$, existiert stets eine Darstellung (129). Für $Y_t = e_t$ weißes Rauschen ist $\mu_Y = \sigma_e^2 / (2\pi) \lambda_{(-\pi, \pi]}$ (Lebesgues Maß), und (129) die Fourier Entwicklung von φ .

b. Ist in der Praxis ein Filter der Form (130) gegeben, so bestimmt die Transferfunktion φ durch (129) und Koeffizientenvergleich.

c. Sind die Koeffizienten c_j absolut summierbar, so definiert (129) eine beschränkte stetige Funktion, die also für jedes endliche Maß μ in $L_2(\mu)$ ist.

Beispiel 4.22

1. Differenzenoperatoren. Für $X_t = DY_t = Y_t - Y_{t-1}$ sind in (129) $c_0 = 1, c_1 = -1, c_j = 0$ sonst. Somit

$$\varphi(\omega) = 1 - e^{i\omega} = e^{i\omega/2} (e^{-i\omega/2} - e^{i\omega/2}) = -2ie^{i\omega/2} \sin(\omega/2).$$

Als Filterfunktion erhalte

$$|\varphi(\omega)|^2 = 4 \sin^2(\omega/2) = 2(1 - \cos \omega).$$

Für Differenzenoperatoren höherer Ordnung, die durch

$$D^n Y_t = D^{n-1} Y_t - D^{n-1} Y_{t-1} = \sum_{r=0}^n (-1)^r \binom{n}{r} Y_{t-r}, \quad t \in \mathbb{Z}, n \geq 1,$$

(Induktion!) definiert sind. Aus der Rechenregel zur Hintereinschaltung von Filtern gilt für den Filter von D^n : $\varphi(\omega) = (-2i)^n e^{in\omega/2} \sin^n(\omega/2)$, und für die Filterfunktion $\varphi(\omega) = 2^n (1 - \cos \omega)^n$.

2. Moving average. Falls

$$X_t = \frac{1}{2q+1} \sum_{j=-q}^q Y_{t-j}, \quad t \in \mathbb{Z},$$

somit $c_j = 1/(2q+1), j = t-q, \dots, t+q$, und $c_j = 0$, sonst. Man erhält $\varphi(0) = 1$ sowie

$$\varphi(\omega) = \frac{1}{2q+1} \frac{\sin(\omega(2q+1)/2)}{\sin(\omega/2)}, \quad \omega \neq 0.$$

4.6 Anwendungen auf ARMA Modelle

Angenommen, die stationäre, zentrierte Zeitreihe $(X_t)_{t \in \mathbb{Z}}$ erfüllt eine $ARMA(p, q)$ -Gleichung (25) mit einem weißen Rauschen $(e_t)_{t \in \mathbb{Z}}$. Seien Z_X und Z_e die zugehörigen MOWs (nach Satz 4.13). Mit (121) kann man dann die $ARMA(p, q)$ Gleichung schreiben als

$$\int_{(-\pi, \pi]} (1 - \phi_1 e^{-i\omega} - \dots - \phi_p e^{-ip\omega}) e^{it\omega} dZ_X(\omega) = \int_{(-\pi, \pi]} (1 + \theta_1 e^{-i\omega} + \dots + \theta_q e^{-iq\omega}) e^{it\omega} dZ_e(\omega). \quad (131)$$

Mit den AR- und MA-Polynomen $\phi(z)$ und $\theta(z)$ schreibe

$$\varphi(\omega) = \phi(e^{-i\omega}), \quad \vartheta(\omega) = \theta(e^{-i\omega}). \quad (132)$$

Damit kann man (131) schreiben als

$$\int_{(-\pi, \pi]} \varphi(\omega) e^{it\omega} dZ_X(\omega) = \int_{(-\pi, \pi]} \vartheta(\omega) e^{it\omega} dZ_e(\omega). \quad (133)$$

Dann sind auch die Spektralmaße der linken und rechten Seite von (133) gleich, und mit (127) und $d\mu_e(\omega) = \sigma_e^2 / (2\pi) d\omega$ kann man schreiben

$$\int_B |\varphi(\omega)|^2 d\mu_X(\omega) = \int_B |\vartheta(\omega)|^2 \frac{\sigma_e^2}{2\pi} d(\omega), \quad B \subset (-\pi, \pi]. \quad (134)$$

Beweis (von Satz 2.10)

a. Angenommen, $\phi(z) \neq 0$ für $|z| = 1$, also $\varphi(\omega) \neq 0$ für alle $\omega \in (-\pi, \pi]$. Da φ stetig und 2π periodisch ist (ist ein trigonometrisches Polynom), ist φ^{-1} beschränkt, also in $L_2(\nu)$ für alle endlichen Maße ν . Somit erhält man durch Anwendung des Filters φ^{-1} auf (133) mit Bemerkung 4.19 und (121)

$$X_t = \int_{(-\pi, \pi]} e^{it\omega} \frac{\vartheta(\omega)}{\varphi(\omega)} dZ_e(\omega). \quad (135)$$

Dies zeigt die Eindeutigkeit einer Lösung. Andererseits ist ein über (135) definierter Prozess (X_t) zentriert und stationär (Satz 4.18), und mit Bemerkung 4.19 gilt

$$\int_{(-\pi, \pi]} \varphi(\omega) e^{it\omega} dZ_X(\omega) = \int_{(-\pi, \pi]} e^{it\omega} \varphi(\omega) \frac{\vartheta(\omega)}{\varphi(\omega)} dZ_e(\omega) = \int_{(-\pi, \pi]} \vartheta(\omega) e^{it\omega} dZ_e(\omega).$$

b. Angenommen, (X_t) sei eine stationäre, zentrierte Lösung der ARMA Gleichung (133). Wir zeigen dann $\varphi(\omega) \neq 0$ für alle $\omega \in (-\pi, \pi]$, und führen dazu einen Widerspruchsbeweis. Angenommen, $\varphi(\omega_0) = 0$, setze $z_0 = e^{i\omega_0}$. Dann kann man für das AR Polynom für ein Polynom $\tilde{\phi}(z)$ schreiben

$$\phi(z) = (1 - z/z_0)\tilde{\phi}(z).$$

Es ist

$$\left| 1 - \frac{e^{i\omega}}{e^{i\omega_0}} \right| = |e^{i\omega_0} - e^{i\omega}| = 2|\sin(\omega - \omega_0)|.$$

Setze nun

$$U_t = \int_{(-\pi, \pi]} \tilde{\phi}(e^{-i\omega}) e^{it\omega} dZ_X(\omega), \quad t \in \mathbb{Z}.$$

Dann gilt für das Spektralmaß μ_U von (U_t) nach Bemerkung 4.19

$$\int_B 4 \sin^2(\omega - \omega_0) d\mu_U(\omega) = \int_B |\varphi(\omega)|^2 d\mu_X, \quad B \subset (-\pi, \pi].$$

Damit und mit (134) erhält man

$$\int_B 4 \sin^2(\omega - \omega_0) d\mu_U(\omega) = \int_B |\vartheta(\omega)|^2 \frac{\sigma_e^2}{2\pi} d(\omega), \quad B \subset (-\pi, \pi]. \quad (136)$$

Da $\phi(z)$ und $\theta(z)$ keine gemeinsame Nullstelle besitzen (Gleichung liegt in reduzierter Form vor), existieren $\epsilon_0 > 0$, $c > 0$, so dass $|\vartheta(\omega)| \geq c$ für $\omega \in (\omega_0 - \epsilon_0, \omega_0 + \epsilon_0) \cap (-\pi, \pi] =: I(\epsilon_0)$. Aus (136) mit $B = I(\epsilon)$, $\epsilon \leq \epsilon_0$, und $|\sin x| \leq |x|$ folgt

$$4\mu_U(I(\epsilon))\epsilon^2 \geq c^2 \epsilon \frac{\sigma_e^2}{\pi}.$$

Division durch ϵ und $\epsilon \rightarrow 0$ liefert einen Widerspruch. ■

Bemerkung 4.23

Falls Gleichung nicht in reduzierter Form vorliegt.

Gleichung (135) liefert direkt mit Satz 4.18, Gleichung (128) die Spektraldichte eines $ARMA(p, q)$ Prozesses.

Satz 4.24

Es sei $\mathcal{X} = (X_t)_{t \in \mathbb{Z}}$ ein $ARMA(p, q)$ - Prozess zur Gleichung (133). Dann hat \mathcal{X} die Spektraldichte

$$f_{\mathcal{X}}(\omega) = \frac{\sigma_e^2}{2\pi} \frac{|\vartheta(\omega)|^2}{|\varphi(\omega)|^2}. \quad (137)$$

Bemerkung 4.25

Falls $\mathcal{X} = (X_t)_{t \in \mathbb{Z}}$ reell, dann $f_{\mathcal{X}}$ symmetrisch.

Bemerkung 4.26

parametrische Spektraldichteschätzung

Beweis (von Satz 2.12)

a. Ist (X_t) kausal, so kann man nach (29) mit Satz 4.20 schreiben

$$X_t = \int P(e^{-it\cdot}) e^{it\cdot} dZ_e, \quad P(z) = \sum_{j=0}^{\infty} \psi_j z^j. \quad (138)$$

Anwendung des Filters φ auf (138) liefert

$$\int \varphi e^{it\cdot} dZ_X = \int \varphi P(e^{-it\cdot}) e^{it\cdot} dZ_e.$$

Mit (133) und Eindeutigkeit des Filters in L_2 folgt dann

$$\vartheta(\omega) = \varphi(\omega)P(e^{-i\omega}) \quad \text{in } L_2(\mu_e). \quad (139)$$

Wegen Stetigkeit gilt die Gleichheit auch überall auf $(-\pi, \pi]$. Nun ist $\phi(z)P(z)$ eine Potenzreihe, die für $|z| \leq \rho$ konvergiert, und auf $|z| = 1$ mit $\theta(z)$ übereinstimmt. Somit folgt nach dem Identitätssatz für holomorphe Funktionen

$$\theta(z) = \phi(z)P(z), \quad |z| \leq 1.$$

Somit wäre eine Nullstelle von ϕ für $|z| \leq 1$ auch eine Nullstelle von $\theta(z)$, welches aber wegen der Reduziertheit ausgeschlossen ist. Also $\phi(z) \neq 0$ für $|z| \leq 1$.

b. Ist $\phi(z) \neq 0$ für $|z| \leq 1$, so konvergiert die Potenzreihe

$$\frac{\theta(z)}{\phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j$$

für $|z| \leq \rho$ mit $\rho > 1$, und die Koeffizienten ψ_j sind absolut summierbar (vgl. Remmert ?). Daher ergibt sich (29) mit der Darstellung (135) und Satz 4.20, wobei noch $\phi(0) = \theta(0) = 1$ zu beachten ist. ■

Der Beweis von Satz 2.15 (Invertibilität) verläuft ganz analog, nur die Rollen von φ und ϑ werden vertauscht.

Bemerkung 4.27

Falls nicht kausal, existiert eine 2-seitige Entwicklung.

Beweis (von Satz 2.16)

a. Wir betrachten zunächst die Klasse der kausalen AR-Modelle. Ist $(X_t)_{t \in \mathbb{Z}}$ ein kausaler AR(p)-Prozess, so ist wegen Satz 2.19 und Satz 1.25 die Kovarianzmatrix Γ_h in (43) für jedes $h \geq 1$ invertierbar. Ist nun neben der (kausalen) AR(p)-Darstellung mit Koeffizienten ϕ_1, \dots, ϕ_p eine weitere Darstellung mit Koeffizienten $\phi_1^*, \dots, \phi_{p^*}^*$, $p^* \geq p$ gegeben, so setzen wir

$$\phi(p^*) = (\phi_1, \dots, \phi_p, 0, \dots, 0)^T \in \mathbb{C}^{p^*}, \quad \phi^*(p^*) = (\phi_1^*, \dots, \phi_{p^*}^*)^T.$$

Beide Vektoren erfüllen die Yule-Walker Gleichung (??, uns somit folgt mit der Invertierbarkeit von Γ_{p^*}

$$\phi(p^*) = \Gamma_{p^*}^{-1} \gamma_{p^*}^1 = \phi^*(p^*).$$

und somit $p = p^*$ und $\phi_j = \phi_j^*$.

b. Nun betrachten wir eine stationäre Zeitreihe $\mathcal{X} = (X_t)_{t \in \mathbb{Z}}$, die der reduzierten, invertiblen, kausalen ARMA(p, q) Gleichung (133) mit weißem Rauschen (e_t) genügt, so dass die Spektraldichte von \mathcal{X} durch (137) gegeben ist. Gilt darüber hinaus eine weitere reduzierte, invertible, kausale ARMA(p*, q*) Gleichung $\phi^*(B)X_t = \theta^*(B)e_t^*$ mit einem weißen Rauschen (e_t^*) , so dass für die Spektraldichte $f_{\mathcal{X}}$ auch gilt

$$f_{\mathcal{X}}(\omega) = \frac{\sigma_{e^*}^2}{2\pi} \frac{|\vartheta^*(\omega)|^2}{|\varphi^*(\omega)|^2}, \quad (140)$$

wobei ϑ^* und φ^* wie in (132) definiert sind.

Wir betrachten nun die kausale AR-Gleichung

$$\phi(B)\theta^*(B)Y_t = \eta_t, \quad (141)$$

wobei (η_t) ein weißes Rausche mit $E\eta_t^2 = 1$ sei. Dann hat die eindeutig bestimmt Lösung (Satz 2.10) $(Y_t)_{t \in \mathbb{Z}}$ von (141) die Spektraldichte

$$f_Y(\omega) = \frac{1}{2\pi} |\varphi(\omega)\vartheta^*(\omega)|^{-2}.$$

Setze nun $\eta_t^* = \phi^*(B)\theta(B)Y_t$, dann gilt für die Spektraldichte f_{η^*} von (η_t^*) nach Bemerkung 4.19

$$f_{\eta^*}(\omega) = \frac{1}{2\pi} \frac{|\varphi^*(\omega)\vartheta(\omega)|^2}{|\varphi(\omega)\vartheta^*(\omega)|^2} = \frac{1}{2\pi}$$

nach (137) und (140). Somit ist auch (η_t^*) ein weißes Rauschen, und nach Teil a. folgt somit die Gleichheit der beiden kausalen AR-Darstellungen von (Y_t) ,

$$\phi(z)\theta^*(z) = \phi^*(z)\theta(z), \quad z \in \mathbb{C}. \quad (142)$$

■

Da die Polynome $\phi(z)$ und $\theta(z)$ sowie $\phi^*(z)$ und $\theta^*(z)$ jeweils keine gemeinsamen Nullstellen haben, müssen somit wegen (142) die Nullstellen von $\phi(z)$ und $\phi^*(z)$ sowie von $\theta(z)$ und $\theta^*(z)$ übereinstimmen, so dass sich die Polynome jeweils nur um einen Faktor unterscheiden. Da aber $\phi(0) = \phi^*(0) = \theta(0) = \theta^*(0) = 1$, folgt die Gleichheit $\phi(z) = \phi^*(z)$ und $\theta(z) = \theta^*(z)$. Dann folgt aus der Invertibilität auch die Gleichheit $e_t = e_t^*$ und insbesondere $\sigma_e^2 = \sigma_{e^*}^2$.

Bemerkung 4.28

Falls reduzierte ARMA Gleichung nicht kausal oder invertibel, so kann man (bei geändertem weißen Rauschen) zu einer kausalen, invertiblen Gleichung übergehen.

4.7 Anwendungen auf fraktionale ARMA Modelle

Beweis (von Satz 3.15)

Wir führen den Beweis nur für $d \in (0, 0.5)$. Mit (79) gilt in diesem Fall

$$\sum_{j \geq 0} \psi_j^2 < \infty \quad \text{sowie} \quad \sum_{j \geq 0} |\alpha_j| < \infty. \quad (143)$$

Eindeutigkeit. Angenommen, (X_t) ist eine stationäre Lösung von (77) mit Spektralmaß Z_X . Da die α_j absolut summierbar sind (s. (143)), folgt mit Bemerkung 4.21 c.

$$e_t = (1 - B)^d X_t = \int_{(-\pi, \pi]} e^{it\omega} (1 - e^{-i\omega})^d dZ_X(\omega).$$

Da e_t nach Voraussetzung ein weißes Rauschen ist, gilt für das Spektralmaß $d\mu_e = \sigma^2/2\pi dx$ und somit wegen (143) $\sum_{j \geq 0} \psi_j e^{-ij\omega} \in L_2(\mu_e)$. Daher ist $(1 - B)^{-d} e_t$ definiert (der Filter hat gerade die Koeffizienten ψ_j) und mit Bemerkung 4.19

$$\begin{aligned} (1 - B)^{-d} e_t &= \int_{(-\pi, \pi]} e^{it\omega} (1 - e^{-i\omega})^{-d} dZ_e(\omega) \\ &= \int_{(-\pi, \pi]} e^{it\omega} (1 - e^{-i\omega})^{-d} (1 - e^{-i\omega})^d dZ_X(\omega) \\ &= \int_{(-\pi, \pi]} e^{it\omega} dZ_X(\omega) = X_t. \end{aligned}$$

Existenz. Wir definieren nun (wohldefiniert mit (143) wie oben)

$$X_t = (1 - B)^{-d} e_t. \quad (144)$$

Anwendung des Filters $(1 - B)^d$ (erlaubt wegen (143) wie oben) liefert dann (77), also die Existenz.

Spektraldichte. Diese ergibt sich direkt aus (144) mit (128) zu

$$f(\omega) = |1 - e^{-i\omega}|^2 \frac{\sigma^2}{2\pi} = |2 \sin(\omega/2)|^{-2d} \frac{\sigma^2}{2\pi}. \quad (145)$$

Autokovarianzen. Diese berechnen sich zu

$$\begin{aligned} \gamma(h) &= \int_{(-\pi, \pi]} e^{ih\omega} f(\omega) d\omega = \frac{\sigma^2}{2\pi} \int_0^{2\pi} \cos(h\omega) f(\omega) d\omega \\ &= \frac{\sigma^2}{2^{2d}\pi} \int_0^\pi \cos(2hx) |\sin(x)|^{-2d} dx = \frac{(-1)^h \Gamma(1 - 2d)}{\Gamma(h - d + 1) \Gamma(1 - h - d)} \sigma^2, \end{aligned}$$

da (vgl. REFERENZ)

$$\int_0^\pi \cos(kx) |\sin(x)|^{\nu-1} dx = \frac{\pi \cos(k\pi/2) \Gamma(\nu + 1) 2^{1-\nu}}{\nu \Gamma((\nu + k + 1)/2) \Gamma((\nu - k + 1)/2)}.$$

Hieraus ergeben sich direkt die Autokorrelationen. Die partiellen Autokorrelationen können mit Induktion aus der Levinson Rekursion bestimmt werden (vgl. Brockwell und Davis, 1991, p. 324). ■