

Thema:ARMA-Prozesse

Johannes Blank

8. Juni 2010

Abstract

Zeitreihen, also stoch. Prozesse mit Indexmenge \mathbb{Z} , spielen in der Ökonometrie (der mathematisch formalisierte Teil der Wirtschaftswissenschaften) und in der Finanzmathematik eine wichtige Rolle. Beispiele wären die Entwicklung eines Aktienkurses oder die statistische Datenerhebung über einen längeren Zeitraum. Von Interesse sind dabei insbesondere die Eigenschaften einer gegebenen Zeitreihe, die eine Prognose ermöglichen, d.h. die von einem gegebenem Satz an Daten Rückschlüsse auf zukünftige Entwicklungen erlauben. Eine solche Eigenschaft, die wir im folgenden untersuchen werden, ist die Stationarität. ARMA-Prozesse sind besondere Zeitreihen, und viele stationäre Prozesse lassen sich als ARMA-Prozesse modellieren. Daher bietet sich eine Analyse dieser Zeitreihen an.

0 Motivation

In der Finanzmathematik begegnen uns Zeitreihen in natürlicher Weise in Form von Kursen, z.B. von Aktien, Indices, allgemein in Form von Preisen von Finanzgütern im Lauf der Zeit. Betrachten wir z.B. den Verlauf des DAX-Kurses (vgl. Seite 2). Während die absoluten DAX-Werte keinen festen Erwartungswert vermuten lassen (die Werte schwanken zwischen 3500 und 8000 Punkten), sieht es bei den Logreturn-Daten bereits besser aus; insbesondere zwischen 2004 und 2007 schwanken die Werte nur geringfügig um die Nullmarke, hier entspricht die Grafik eher unserer Vorstellung eines "stationären" Prozesses. Ziel ist es, dertartige Prozesse verstehen und modellieren zu können, um damit Voraussagen für die zukünftige Entwicklung treffen zu können.

1 Grundlegende Definitionen

Definition 1.1. *Eine Zeitreihe ist im folgenden ein reellwertiger stochastischer Prozess $(X_t)_{t \in \mathbb{Z}}$ auf einem W -Raum (Ω, \mathcal{F}, P) .*

Anmerkung 1.2. *Zeitreihen lassen sich oftmals in der Form*

$$X_t = m_t + s_t + Y_t$$

aufschlüsseln; dabei ist s_t die sogenannte Saison-Komponente mit Periodenlänge d , d.h. $s_{t+d} = s_t$ f.a. t , m_t wird als Trendkomponente bezeichnet und stellt einen (längerfristigen) Trend der Zeitreihenentwicklung dar. Y_t schließlich ist die für

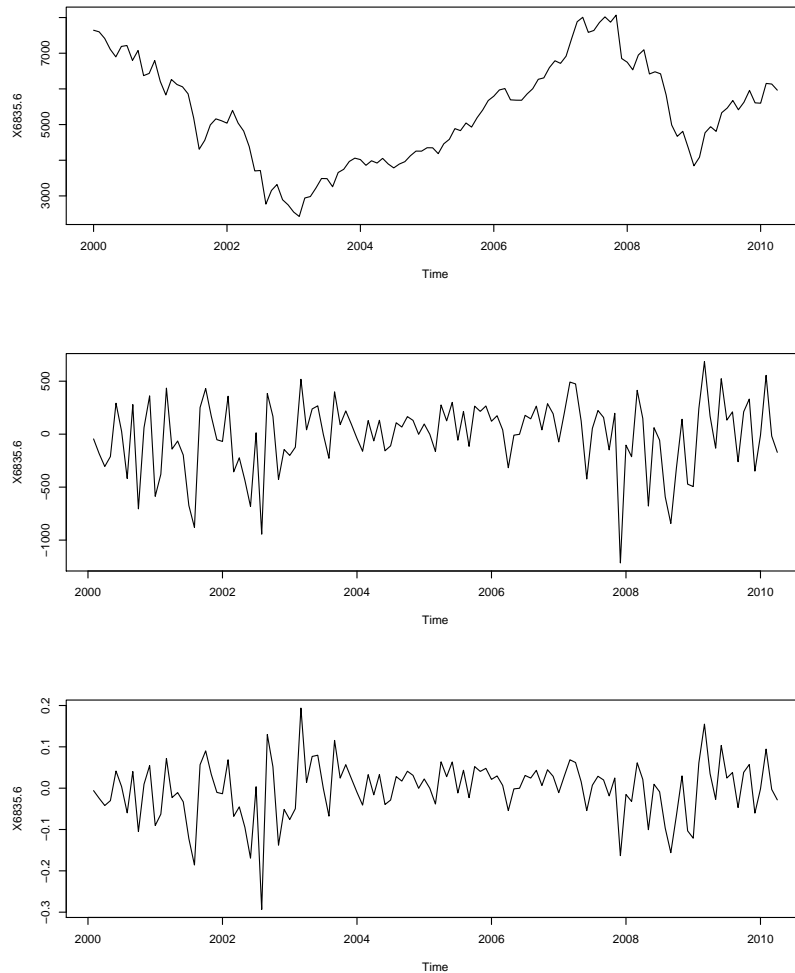


Figure 1: DAX Absolutwerte, Differenzen und Logreturns. Monatsdaten der Bundesbank

uns entscheidende Größe, die Restkomponente, welche zufällige Störungen modelliert. Unsere Analyse konzentriert sich v.a. auf diese nicht-deterministische Komponente. Am Ende von Kapitel 2 werden wir einige Methoden einführen, Saison- und Trendkomponente aus einer Zeitreihe herauszurechnen.

Definition 1.3 (Mittel und Autokovarianz-Funktionen). $\mu(t) = \mathbb{E}(X_t)$ wird als das Mittel, $\gamma(t, s) = \mathbb{E}((X_t - \mu(t))(X_s - \mu(s))) = \text{cov}(X_t, X_s)$ als Autokovarianz-Funktion von X bezeichnet.

Für uns von besonderer Bedeutung ist der Begriff der *Stationarität*. Hiermit wird eine Eigenschaft von speziellen Zeitreihen eingeführt, die für sämtliche Prozesse entscheidend ist, die wir betrachten werden.

Definition 1.4 (Stationär).

(X_t) heißt **strikt stationär**, falls f.a. $t_1, \dots, t_n \in \mathbb{Z}, n \in \mathbb{N}$ gilt

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+k}, \dots, X_{t_n+k})$$

(X_t) heißt **schwach stationär** oder einfach **stationär**, wenn $\text{var}(X_t) < \infty$ f.a. $t \in \mathbb{Z}$ und

$$\mu(t) = \mu, \gamma(t, s) = \gamma(t+k, s+k)$$

gilt f.a. $t, s, k \in \mathbb{Z}$

Eigenschaften stationärer Prozesse:

- Strikt stationäre Prozesse sind schwach stationär, falls $\text{var} X_t < \infty$
- Ist X_t normalverteilt für alle t und (schwach) stationär, so ist X strikt stationär
- $\gamma(t-s, 0) = \gamma(t, s) = \gamma(s, t) = \gamma(s-t, 0)$ f.a. s, t , falls X schwach stationär ist; definiere dann $\gamma(h) := \gamma(h, 0)$, $h \in \mathbb{Z}$ heißt dann *lag*

Definition 1.5 (ACF). $\rho_X(h) := \rho(h) := \rho(X_h, X_0) := \gamma(h)/\gamma(0)$ für $h \in \mathbb{Z}$ heißt **Autokorrelationsfunktion** eines schwach stationären Prozesses X . Nach Cauchy-Schwarz ist $\rho(h) \in (-1, 1)$ f.a. $h \in \mathbb{Z}$

Definition 1.6 (Weißes Rauschen). Ein **Weißes Rauschen** ist ein schwach stationärer stoch. Prozess (X_t) mit $\rho(h) = \begin{cases} 1, & h = 0 \\ 0, & h \neq 0 \end{cases}$, d.h. die X_t sind paarweise unkorreliert.

$WN(0, \sigma^2)$ bezeichne im Folgenden ein Weißes Rauschen mit $\text{var}(X_t) = \sigma^2$, $\mu(t) = 0$.

(X_t) heißt **striktes Weißes Rauschen**, falls (X_t) iid ist mit $\text{var}(X_t) < \infty$. $SWN(0, \sigma^2)$ bezeichne im Folgenden ein striktes Weißes Rauschen mit $\text{var}(X_t) = \sigma^2$, $\mu(t) = 0$.

Definition 1.7 (Martingaldifferenz). (X_t) wird **Martingaldifferenzfolge** bzgl. einer Filtration $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ genannt, wenn $\mathbb{E}|X_t| < \infty$, X_t ist \mathcal{F}_t -messbar und $\mathbb{E}(X_t | \mathcal{F}_{t-1}) = 0 \forall t \in \mathbb{Z}$

Für eine MDF gilt $\mathbb{E}(X_t) = \mathbb{E}(\mathbb{E}(X_t | \mathcal{F}_{t-1})) = 0 \forall t$ sowie

$$\gamma(t, s) = \text{cov}(X_t, X_s) = \mathbb{E}(X_t X_s) = 0 \text{ für } t \neq s \quad (1.1)$$

Ist $\text{var}(X_t)$ konstant für alle t , so ist X_t also Weißes Rauschen.

2 ARMA-Prozesse (Auto-Regressive-Moving-Average)

Definition 2.1 (ARMA-Prozess). Sei $(\varepsilon_t)_{t \in \mathbb{Z}} WN(0, \sigma_\varepsilon^2)$. Der Prozess $(X_t)_{t \in \mathbb{Z}}$ ist **0-Mittel ARMA(p,q)-Prozess**, wenn er schwach stationär ist und er Gleichungen der Form

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad \forall t \in \mathbb{Z}$$

genügt für gewisse $\phi_i, \theta_i \in \mathbb{R}$. (X_t) ist ein ARMA-Prozess mit Mittel μ , falls $(X_t - \mu)$ ein 0-Mittel ARMA(p, q)-Prozess ist. Der ARMA-Prozess (X_t) heißt *kausal* (bzgl. (ε_t)), falls

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \quad (2.2)$$

für gewisse $\psi_i \in \mathbb{R}$ mit

$$\sum_{i=0}^{\infty} |\psi_i| < \infty \quad (2.3)$$

Anmerkung 2.2. Sei L der Lag-Operator, d.h. es gelte $L(X_t) = (X_{t-1})$. Sei $\Phi(z) = 1 - \phi_1 z - \dots - \phi_q z^q$, $\Theta(z) = 1 + \theta_1 z + \dots + \theta_p z^p$. Dann lässt sich die Differenzengleichung in Def. 2.1 ausdrücken durch diese beiden Polynome, und zwar ist

$$\Phi(L)X_t = \Theta(L)\varepsilon_t \quad \forall t \in \mathbb{Z}$$

Diese Polynome heißen *charakteristische Polynome* des ARMA-Prozesses.

Die Kausalität in obiger Definition drückt dabei aus, dass der Zustand X_t in t Ergebnis vergangener Impulse ε_t ist. In Verallgemeinerung des Konzepts des char. Polynoms führen wir sogenannte *Filter* ein

Definition 2.3 (Allgemeiner linearer Prozess). Sei (ε_t) Weißes Rauschen und $(\psi_i)_{i \in \mathbb{Z}} \subseteq \mathbb{R}$ absolut summierbar, also $\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$. Dann wird (ψ_i) absolut summierbarer Filter genannt und

$$Y_t := \sum_{i=-\infty}^{\infty} \psi_i \varepsilon_{t-i} \quad (2.4)$$

bezeichnet man als *allgemeinen linearen Prozess*.

Satz 2.4. Sei (X_t) stoch. Prozess mit $\sup_t \mathbb{E}(|X_t|) < \infty$ und sei (ψ_i) absolut summierbarer Filter. Dann ist $\sum_{i=-\infty}^{\infty} |\psi_i X_{t-i}| < \infty$ P-f.s., d.h. $Y_t := \sum_{i=-\infty}^{\infty} \psi_i X_{t-i}$ existiert f.s. Ausserdem gilt dann $E(|Y_t|) < \infty$ sowie

$$E(Y_t) = \lim_{n \rightarrow \infty} \sum_{i=-n}^n \psi_i \mathbb{E}(X_{t-i}), \quad t \in \mathbb{Z} \quad (2.5)$$

Y_t wird dann von (ψ_i) gefilterter Prozess genannt.

Ist (ψ_i) absolut summierbarer Filter, so wird (analog zu Anmerkung 2.2) die *Laurent-Reihe*

$$\chi(z) := \sum_{i \in \mathbb{Z}} \psi_i z^i$$

als *charakteristisches Polynom* von (ψ_i) bezeichnet. Komplexe Analysis liefert uns: $\chi(z)$ existiert entweder auf einem Kreisring $KR := \{z \in \mathbb{C} : r < |z| < R\}$

oder fast nirgends, und im ersten Fall sind die Koeffizienten ψ_i eindeutig durch χ bestimmt.

Mit Hilfe des Cauchyproduktes kann ein Produkt $(\vartheta_i) = (\psi_i) * (\phi_i)$ auf der Menge der absolut summierbaren Filter definiert werden. Dafür seien $(\phi_i), (\psi_i)$ absolut summierbare Filter. Definiere dann

$$\vartheta_m := \sum_{i+j=m} \psi_i \phi_j \quad (2.6)$$

Wir nennen (ϑ_i) den *Produktfilter* von (ϕ_i) und (ψ_i) . Haben $(\phi_i), (\psi_i)$ die char. Polynome χ_1 bzw. χ_2 , so hat der Produktfilter das char. Polynom $\chi := \chi_1 \chi_2$. Als nächstes betrachten wir *inverse* Filter. Dazu seien $(\phi_i), (\psi_i)$ abs. summierbare Filter mit char. Polynomen χ_1 bzw. χ_2 , welche auf einem Kreisring existieren und $\chi_1 \chi_2 = 1$ erfüllen. Damit folgt für den Produktfilter ϑ : $\vartheta_0 = 1$, $\vartheta_i = 0$ sonst. Ist (Z_t) stationärer Prozess, so gilt also fast sicher

$$Y_t = \sum_{i \in \mathbb{Z}} \psi_i Z_{t-i} \quad \text{und} \quad \sum_{j \in \mathbb{Z}} \phi_j Y_{t-j} = Z_t, \quad t \in \mathbb{Z} \quad (2.7)$$

Der Filter (ϕ_i) wird *inverser Filter* von (ψ_i) genannt. Die Menge der absolut summierbaren Filter mit den Verknüpfungen $*$ und $+$ (koeffizientenweise Addition der Reihen) bildet insgesamt einen Ring; 0 ist das neutrale Element der Addition, 1 das der Multiplikation.

Als *kausalen* Filter bezeichnet man analog zu (2.2)-Filter (ψ_i) mit $\psi_i = 0$ f.a. $i < 0$.

Das folgende Lemma garantiert Stationarität von Prozessen, welche die Kausalitätsbedingungen (2.2) und (2.3) erfüllen und erleichtert uns die Berechnung der Autokorrelationsfunktion.

Lemma 2.5. *Erfüllt ein Prozess (X_t) (2.2) und (2.3), so ist er schwach stationär mit*

$$\rho(h) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}}{\sum_{i=0}^{\infty} \psi_i^2} \quad (2.8)$$

Beweis. $\mathbb{E}(X_t) = 0$ wegen (2.2), $\text{var}(X_t) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty$ wegen 2.3. Weiterhin gilt

$$\text{cov}(X_t, X_{t+h}) = \mathbb{E}(X_t X_{t+h}) = \mathbb{E}\left(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t+h-j}\right) \quad (2.9)$$

Da (ε_t) Weißes Rauschen ist, gilt also $\mathbb{E}(\varepsilon_{t-i} \varepsilon_{t+h-j}) \neq 0$ gdw $j = i + h$ und somit

$$\gamma(h) = \text{cov}(X_t, X_{t+h}) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+|h|} \quad (2.10)$$

für $h \in \mathbb{Z}$ □

Anmerkung 2.6. *Haben Φ und Θ keine gemeinsamen Nullstellen, so gilt: Der ARMA-Prozess X_t ist kausal gdw Φ keine Nullstellen im Einheitskreis hat. Die*

ψ_i in (2.2) ergeben sich dann durch

$$\sum_{i=0}^{\infty} \psi_i z^i = \frac{\Theta(z)}{\Phi(z)}, \quad |z| < 1 \quad (2.11)$$

Im Folgenden betrachten wir die Spezialfälle $p = 0$ (*Moving Average*) und $q = 0$ (*Autoregressive*).

Beispiel 2.7 (Moving Average-Prozess). *Ein Moving Average Prozess $MA(q)$ der Form*

$$X_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (2.12)$$

genügt (2.2) für geeignete θ_i . Es gilt dann $\rho(h) = 0$ f.a. h mit $|h| > q$, vgl. (2.8). Man spricht deshalb von einem "cut-off" bei lag q . Also sind MA-Prozesse insbesondere kausale ARMA-Prozesse.

Interpretation: Der $MA(q)$ -Prozess modelliert eine Abhängigkeit des Prozesses X von gewissen Schocks (auch Innovationen oder Impulse genannt) ε_i . Diese werden gewichtet durch die θ_i , daher der Name Moving Average.

Beispiel 2.8 (Autoregressive-Prozess (1)). *Der $AR(1)$ -Prozess erfüllt*

$$X_t = \phi X_{t-1} + \varepsilon_t \quad \forall t \quad (2.13)$$

Daraus folgt

$$X_t = \phi^{k+1} X_{t-k-1} + \sum_{i=0}^k \phi^i \varepsilon_{t-i} \quad (2.14)$$

durch k -malige Anwendung und somit: X_t ist kausal gdw $|\phi| < 1$. Insbesondere ist X_t für $|\phi| < 1$ stationär mit

$$\gamma(h) = \frac{\phi^{|h|} \sigma_\varepsilon^2}{1 - \phi^2}, \quad \rho(h) = \phi^{|h|}, \quad h \in \mathbb{Z} \quad (2.15)$$

Betrachte für $|\phi| < 1$ den Filter $\phi_j := \begin{cases} 0, & j < 0 \\ \phi^j, & j \geq 0 \end{cases}$ und den dadurch gefilterten

Prozess $(\sum_{i=0}^{\infty} \phi_i \varepsilon_{t-i})_t$ (Existenz folgt aus 2.4). Aus obigen Gleichungen folgt, dass dieser Prozess eine Lösung von (2.13) ist; wegen dieser Darstellung spricht man auch von einem $MA(\infty)$ -Prozess, vgl. Beispiel 2.7.

Beispiel 2.9 (Autoregressive-Prozess (p)). *Ein allgemeiner $AR(p)$ -Prozess hat die Form*

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t, \quad t \in \mathbb{Z} \quad \text{wobei } \varepsilon \text{ WN sei} \quad (2.16)$$

Interpretation: Der $AR(p)$ -Prozess modelliert solche Prozesse, die zum Zeitpunkt t von ihrer Vergangenheit bis in die $t - p$ -te Periode abhängen, sowie von der (zufälligen) Innovation ε_t .

*$AR(p)$ -Prozesse müssen aber i.A. **nicht** stationär sein! Allerdings existiert eine stationäre Lösung von (2.16), falls die Nullstellen des charakterischen*

Polynoms Φ ausserhalb des Einheitskreises liegen (wir sprechen dann von AR-Stationarität). Diese ist dann gegeben durch

$$X_t := \sum_{i \geq 0} \eta_i \varepsilon_{t-i} \quad (2.17)$$

wobei (η_i) der absolut summierbare inverse kausale Filter von $c_0 = 1, c_i = \phi_i$ für $i=1, \dots, p, c_i = 0$ sonst ist.

Im obigen AR(1)-Beispiel ist $\Phi(z) = 1 - \phi z$, also existiert insbesondere eine stationäre Lösung für $|\phi| < 1$, vgl. obiges Beispiel

Mit Hilfe des folgenden Lemmas ergibt sich eine "cut off"-Eigenschaft für AR(p)-Prozesse, analog zu der für MA(q)-Prozesse

Lemma 2.10 (Yule-Walker-Gleichungen). Sei $X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t$ AR-stationärer AR(p)-Prozess. Dann gilt

$$\rho(k) = \sum_{i=1}^p \phi_i \rho(k-i) \quad (2.18)$$

Beweis. Es gilt für $\mu := EX_0$

$$X_t - \mu = \sum_{i=1}^p \phi_i (X_{t-i} - \mu) + \varepsilon_t - \mu \left(1 - \sum_{i=1}^p \phi_i\right) \quad (2.19)$$

Bildet man den Erwartungswert beider Seiten, folgt $\mu \left(1 - \sum_{i=1}^p \phi_i\right) = \mathbb{E}\varepsilon_0$ da X insbesondere stationär ist. Multipliziert man erst mit $X_{t-k} - \mu$ und nimmt wieder den Erwartungswert, so folgt

$$\begin{aligned} \gamma(k) &= \mathbb{E}((X_t - \mu)(X_{t-k} - \mu)) \\ &= \sum_{i=1}^p \phi_i \mathbb{E}((X_{t-i} - \mu)(X_{t-k} - \mu)) + \mathbb{E}((\varepsilon_t - \mathbb{E}\varepsilon_0)(X_{t-k} - \mu)) \\ &= \sum_{i=1}^p \phi_i \gamma(k-i) \end{aligned}$$

Zu *: X_{t-k} und ε_t sind unkorreliert für $k > 0$. Teilen durch $\gamma(0)$ liefert die Behauptung. □

Auch für AR(p)-Prozesse gibt es eine "cut-off"-Eigenschaft, allerdings nicht für die ACF, sondern für die sogenannte Partielle ACF. Dafür sei

$$R_k := \frac{1}{\gamma(0)} (\text{cov}(X_i, X_j))_{1 \leq i, j \leq k} \quad (2.20)$$

R_k ist pos. semidefinit, da cov ein Skalarprodukt ist. Definiere, falls R_k invertierbar ist,

$$a_k := (a_{k1}, \dots, a_{kk}) := R_k^{-1}(\rho(1), \dots, \rho(k)) \quad (2.21)$$

Dann bezeichnet man $\alpha(k) := a_{kk}$ als PACF-Koeffizient bei einem lag von k .

Sei jetzt (X_t) wie in Lemma 2.10. Dann gilt mit Yule-Walker

$$\rho(h) = \sum_{i=1}^p \phi_i \rho(h-i) \stackrel{(2.21)}{=} \sum_{i=1}^h \rho(h-i) a_{hi}$$

Ist also $h > p$, so ist $a_{hh} = 0$ wegen der Eindeutigkeit der a_k 's. Dies liefert die gewünschte "cut off"-Eigenschaft der PACF.

Beispiel 2.11 (ARMA(1,1)-Prozess). Für X_t gegeben durch

$$X_t - \phi X_{t-1} = \varepsilon_t + \theta \varepsilon_{t-1} \quad \forall t \in \mathbb{Z} \quad (2.22)$$

gilt $\Phi(z) = 1 - \phi z$ sowie $\Theta(z) = 1 + \theta z$; für $\phi + \theta \neq 0$ liegen keine gemeinsamen Nullstellen vor. Aus

$$\sum_{i=0}^{\infty} \psi_i z^i = \frac{1 + \theta z}{1 - \phi z} \stackrel{\text{geom. Reihe}}{=} (1 + \theta z)(1 + \phi z + \phi^2 z^2 + \dots), \quad |z| < 1 \quad (2.23)$$

erhält man eine Darstellung von X_t durch

$$X_t = \varepsilon_t + (\phi + \theta) \sum_{i=1}^{\infty} \phi^{i-1} \varepsilon_{t-i} \quad (2.24)$$

Für $h \neq 0$ ist

$$\rho(h) = \frac{\phi^{|h|-1} (\phi + \theta)(1 + \phi\theta)}{1 + \theta^2 + 2\phi\theta} \quad (2.25)$$

Liegen Beobachtungen für einen ARMA(p,q)-Prozess X vor, so ist zwar der Zustand X_t zum Zeitpunkt t sichtbar, der Impuls ε_t jedoch nicht. Prozesse, die es erlauben, aus der Beobachtung der X_t auf den Impuls ε_t zu schließen, heißen *invertierbar*, genauer:

Ein ARMA(p,q)-Prozess heißt invertierbar (bzgl. (ε_t)), falls es einen absolut summierbaren Filter (π_i) gibt mit

$$\varepsilon_t = \sum_{i=0}^{\infty} \pi_i X_{t-i} \quad (2.26)$$

Ist (X_t) ARMA(p,q)-Prozess, so dass die char. Polynome keine gemeinsamen Nullstellen haben, so kann man zeigen: (X_t) ist invertierbar gdw. Θ keine Nullstellen im Einheitskreis hat.

Liegt ein invertierbares, nicht notwendig zentriertes ARMA-Modell vor, so können wir

$$X_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu + \sum_{i=1}^p (\phi_i X_{t-i} - \mu) + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (2.27)$$

für das Modell schreiben. Dabei ist μ_t messbar bzgl. $\mathcal{F}_{t-1} = \sigma(\{X_s : s \leq t-1\})$

Um nicht-saisonale Trends aus einer Zeitreihe herauszurechnen, kann ein ARIMA-Modell betrachtet werden.

Definition 2.12 (Arima-Modelle (Autoregressive Integrated Moving-Average)). Die Zeitreihe (Y_t) heißt ARIMA(p,d,q)-Prozess, falls $X_t := \Delta^d Y_t$ ein ARMA(p,q)-Prozess ist. Dabei bezeichne Δ den Differenzenoperator, d.h. $\Delta = 1 - L$

Anmerkung 2.13 (Polynomialer Trend). Sei $X_t = m_t + Y_t$ wie in Anmerkung 1.2 mit $\mathbb{E}Y_t = 0$ f.a.t und einem polynomialen Trend m_t vom Grad k , d.h.

$m_t = \sum_{j=0}^k a_j t^j$. Dann ist

$$\Delta^k X_t = k!a_k + \Delta^k Y_t \quad (2.28)$$

stationär mit $\mathbb{E}(\Delta^k X_t) = k!a_k$

Beweis. Wegen der Linearität von Δ reicht es zu zeigen, dass $\Delta^k t^k = k!$ gilt. Zunächst ist

$$\Delta t^k = t^k - (t-1)^k \stackrel{\text{Bin. Lehrsatz}}{=} P_{k-1}(t), \quad \deg P_{k-1} \leq k-1 \quad (2.29)$$

für beliebiges k mit

$$P_{k-1}(t) = - \sum_{i=0}^{k-1} \binom{k}{i} (-1)^{k-i} t^i \quad (2.30)$$

Für $k=1$ ist $\Delta t = t - (t-1) = 1 = k!$. Nehme an, die Behauptung gelte für k . Dann

$$\Delta^{k+1} t^{k+1} = \Delta^k (\Delta t^{k+1}) = \Delta^k (P_k(t)) \stackrel{(2.29), (2.30)}{=} \Delta^k ((k+1)t^k) \stackrel{\text{I.V.}}{=} (k+1)!$$

Der Rest der Behauptung ist klar. \square

Anmerkung 2.14 (Saisonale Einflüsse). Gegeben sei eine Zeitreihe $X_t = S_t + R_t$, wobei S ein saisonaler Einfluss sei mit Periodenlänge s , also gelte $S_{t+s} = S_t$ f.a. $t \in \mathbb{Z}$. (R_t) sei stationär. Mit $X_t^* := \Delta_s X_t := (1 - L^s)X_t = X_t - X_{t-s} = R_t - R_{t-s}$ erhält man einen stationären Prozess, eine **saisonbereinigte Zeitreihe**. Falls (X_t^*) ARMA(p, q) ist, nennt man X SARMA $_s(p, q)$ -Prozess. Ist $(\Delta^d(X_t - X_{t-s}))_{t \in \mathbb{Z}}$ ARMA(p, q)-Prozess, so spricht man von einem SARIMA $_s(p, q)$ -Prozess X .

Mit Hilfe dieser beiden Methoden kann eine Zeitreihe der Form $X_t = m_t + s_t + Y_t$ in eine stationäre Zeitreihe transformiert werden.

3 Analyse

Sei $(X_t)_{t \in \mathbb{Z}}$ schwach stationäre Zeitreihe und X_1, \dots, X_n eine zufällige Stichprobe. Für diese Stichprobe definiert

$$\hat{\gamma}(h) := \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n), \quad 0 \leq h < n \quad (3.31)$$

die empirische Autokovarianz der Probe, mit $\bar{X}_n := \frac{1}{n} \sum_{t=1}^n X_t$ als Näherung für den Erwartungswert μ . Aus diesen Werten kann die Autokorrelationsfunktion der Probe errechnet werden:

$$\hat{\rho} = \hat{\gamma}(h)/\hat{\gamma}(0), \quad 0 \leq h < n \quad (3.32)$$

Als *Korrelogramm* wird der Graph $\{(h, \hat{\rho}(h)) : h = 0, 1, 2, \dots\}$ bezeichnet.

Der folgende wichtige Satz hilft uns, ein passendes Modell zu einer Stichprobe von Daten einer Zeitreihe zu finden. Zunächst brauchen wir den Begriff der mehrdimensionalen Normalverteilung

Definition 3.1 (d-dimensionale Normalverteilung). Sei $W \in \mathbb{R}^{d \times d}$ SPD-Matrix, $\mu \in \mathbb{R}^d$ und $X = (X_1, \dots, X_d)$ Zufallsvektor. X nennt man *d-dimensionale normalverteilt* mit Erwartungswert μ , Kovarianzmatrix W , falls

$$\varphi(x) = \frac{1}{\sqrt{(2\pi)^d \det(W)}} \exp\left(-\frac{1}{2} \langle x - \mu, W^{-1}(x - \mu) \rangle\right) \quad (3.33)$$

Dichte von X ist. Schreibe dann $X \sim N_d(\mu, W)$.

Satz 3.2. Sei $(X_t)_{t \in \mathbb{Z}}$ gegeben durch

$$X_t - \mu = \sum_{i=0}^{\infty} \psi_i Z_{t-i}, \quad \text{dabei sei } \sum_{i=0}^{\infty} |\psi_i| < \infty \text{ und } (Z_t)_{t \in \mathbb{Z}} \sim \text{SWN}(0, \sigma_Z^2)$$

Unter der Voraussetzung, dass $\mathbb{E}(Z_t^4) < \infty$ oder $\sum_{i=0}^{\infty} i\psi_i^2 < \infty$, gilt

$$\sqrt{n}(\hat{\rho}(h) - \rho(h)) \xrightarrow{d} N_h(\mathbf{0}, W) \quad \forall h \in \{1, 2, \dots\}$$

wobei

$$\hat{\rho}(h) = (\hat{\rho}(1), \dots, \hat{\rho}(h))^T, \quad \rho(h) = (\rho(1), \dots, \rho(h))^T$$

mit $W_{ij} := \sum_{k=1}^{\infty} (\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k))(\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k))$

Für kausale ARMA-Prozesse ist $\sum_{i=0}^{\infty} i\psi_i^2 < \infty$, also lässt sich der Satz anwenden. Für SWN gilt mit obigem Satz

$$\sqrt{n}\hat{\rho} \xrightarrow{d} N_h(\mathbf{0}, I_h)$$

Ist also n groß genug, sollten 95% der geschätzten Autokorrelationen in $(-1.96/\sqrt{n}, 1.96/\sqrt{n})$ liegen; weichen mehr als 5% der Daten davon ab, deutet das auf die Falschheit der Nullhypothese hin, es handele sich um ein striktes Weißes Rauschen.

4 Statistische Auswertung

Will man die Daten einer gegebenen Zeitreihe X_1, \dots, X_n statistisch auswerten, bietet sich folgendes Programm an. Dabei ist zunächst abzuwägen, wie weit die erhobenen Daten zurückreichen sollen; für genaue stat. Schätzungen sind große Datenmengen erforderlich, andererseits möchte man keine "veralteten" Daten einbeziehen. Letzteres kann passieren, wenn grundlegende Änderungen der Rahmenbedingungen stattgefunden haben, z.B. entscheidende Änderungen in der Wirtschaftspolitik.

- **Vorabanalyse**

Dabei werden die Daten geplottet und die Hypothese der Stationarität geprüft.

- **Analyse**

Mit grafischen (Korrelogramm) und numerischen (z.B. Box-Ljung, vgl.4.1.3) Methoden, angewandt auf die vorliegenden Daten sowie deren Absolutwerte, wird die SWN-Hypothese überprüft.

- **Modellanpassung**

Mithilfe eines Korrelograms (oder anderer Methoden, s.u.) die Ordnung eines passenden ARMA-Prozesses bestimmen und fitten.

- **Residuen analysieren und Modelle vergleichen**

Benutze

$$\hat{\varepsilon}_t = X_t - \hat{\mu}_t, \quad \hat{\mu}_t = \hat{\mu} + \sum_{i=1}^p \hat{\phi}_i (X_{t-i} - \hat{\mu}) + \sum_{j=1}^q \hat{\theta}_j \hat{\varepsilon}_{t-j}$$

Dabei müssen zunächst Startwerte gesetzt werden (da nur endlich viele Daten vorliegen), z.B. $\hat{\varepsilon}_{-q+1} = \hat{\varepsilon}_{-q+2} = \dots = \hat{\varepsilon}_0 = 0, X_{-p+1} = X_{-p+2} = \dots = X_0 = \bar{X}$. Die Residuen $\hat{\varepsilon}_t$ sollten sich verhalten wie ein Weißes Rauschen, da dass ja gerade unsere Annahme für die Innovationen ist. Ein Plot des Korrelograms der (empirischen) ACF der Residuen kann dabei Aufschluss geben.

Das *Box-Jenkins-Programm*, das im Folgenden vorgestellt wird, ist eine konkrete Möglichkeit, einen gegebenen Datensatz an ein ARMA(p,q)-Modell anzupassen und die Güte dieser Anpassung zu überprüfen.

4.1 Box-Jenkins-Programm

Von einer Zeitreihe X wird vermutet, dass es sich um einen ARMA(p,q)-Prozess handelt; gegeben ist der Datensatz x_1, \dots, x_n , so dass etwaige Trends und saisonale Effekte bereits herausgerechnet wurden. Gesucht ist jetzt eine Schätzung der Parameter, d.h. von p, q sowie der Koeffizienten der char. Polynome, $\phi = (\phi_1, \dots, \phi_p), \theta = (\theta_1, \dots, \theta_q)$. Liegen geschätzte Werte vor, sollen diese geprüft werden und mit ihrer Hilfe soll eine Voraussage getroffen werden.

4.1.1 Schätzung von p,q

Wird ein MA(q)-Prozess vermutet, kann ein Plot der empirischen ACF Aufschluss über q geben, denn für $h > q$ sollte aufgrund des cut-offs $\rho(h) = 0$ im MA(q)-Modell bedeuten. Für die empirische ACF bedeutet dies, dass $\hat{\rho}(h)$ nicht "signifikant" von 0 abweichen sollte für $h \geq q$. Entsprechend wird q so gewählt, dass $\hat{\rho}(q)$ signifikant größer als Null ist und $\hat{\rho}(h)$ nahe Null ist für $h > q$.

Ähnlich im AR(p) Modell für die empirische PACF mit einem cut-off bei p . Im Falle eines ARMA(p,q)-Modell wird versucht, (p, q) so zu wählen, dass eine bestimmte Funktion minimiert wird, zum Beispiel

$$AIC(p, q) := \log(\hat{\sigma}_{p,q}^2) + 2 \frac{p+q+1}{n+1} \quad (\text{Akaike Information Criterion}) \quad (4.34)$$

mit $\hat{\sigma}_{p,q}^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_i)^2$.

4.1.2 Schätzung der Koeffizienten

Nach Festlegung von p, q wird eine Schätzung von ϕ und θ gesucht. Zunächst gehen wir von einem normalverteilten Modell aus, d.h. (X_1, \dots, X_n) habe eine Dichte φ wie in Def. 3.1. Mit Hilfe des *Maximum-Likelihood-Prinzips* werden

Schätzungen für den Erwartungswert μ und die Kovarianzmatrix W berechnet. Dabei nehmen wir an, dass $X_t = \sum_{i \geq 0} \alpha_i \varepsilon_{t-i}$, mit $\alpha_i = \alpha_i(\phi, \theta)$. Damit gilt für $h \geq 0$

$$\gamma(h) = \text{cov}(X_0, X_h) = \sum_{i \geq 0} \sum_{j \geq 0} \alpha_i \alpha_j \text{cov}(\varepsilon_{-i}, \varepsilon_{h-j}) = \sigma^2 \sum_{i \geq 0} \alpha_i \alpha_{h+i} \quad (4.35)$$

also für $W' := \sigma^{-2}W$:

$$W' = W'(\phi, \theta)$$

Definiere weiter

$$p(y_1, \dots, y_n) := \varphi(y_1, \dots, y_n) = (2\pi\sigma^2)^{-n/2} (\det(W'))^{-(1/2)} \exp\left(-\frac{1}{2\sigma^2} Q(\nu|y_1, \dots, y_n)\right)$$

mit $\nu := (\sigma^2, \mu, \phi, \theta)$ und $Q(\nu|y_1, \dots, y_n) := ((y_1, \dots, y_n) - \mu)W'^{-1}((y_1, \dots, y_n) - \mu)^T$. Sei $\hat{\nu}$ Maximierer der Likelihood-Funktion $L(\nu|y_1, \dots, y_n) := p(y_1, \dots, y_n|\nu)$, d.h. $L(\hat{\nu}|y_1, \dots, y_n) = \sup_{\nu} L(\nu|y_1, \dots, y_n)$. Also gilt

$$l(\hat{\nu}|x_1, \dots, x_n) := \log L(\hat{\nu}|x_1, \dots, x_n) = \sup_{\nu} \left(-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log(\det W') - \frac{1}{2\sigma^2} Q(\nu|x_1, \dots, x_n)\right)$$

Damit kann ν und somit Schätzungen für ϕ, θ bestimmt werden.

4.1.3 Prüfen der Diagnose

Als nächstes wollen wir die Güte der geschätzten Werte überprüfen. Dazu bietet sich eine Analyse der Residuen $\hat{\varepsilon}_t, t = 1, \dots, n$ an, welche sich wie Realisationen eines Weißen Rauschens verhalten sollten. Dazu betrachtet man die empirische ACF $\hat{\rho}_\varepsilon$ von $\hat{\varepsilon}$ wie in (3.32) definiert. Für $h = 1, \dots, n-1$ sollte $\hat{\rho}_\varepsilon$ "nahe" Null sein. Um diese Bedingung zu präzisieren, betrachten wir die Box-Pierce-Statistik

$$Q(K) := n \sum_{i=1}^K \hat{\rho}_\varepsilon^2(i)$$

welche asymptotisch χ^2 -verteilt ist für $n \rightarrow \infty$ mit $K - p - q$ Freiheitsgraden, falls X_t tatsächlich ARMA(p,q) ist. Ist der Wert $1 - \chi_{K-p-q}^2(Q(K))$ zu klein, wird die Annahme eines ARMA(p,q)-Modells verworfen. Die Methode kann für kleinen Stichprobenumfang unbefriedigende Resultate liefern; besser geeignet ist dann die Box-Ljung-Statistik

$$Q^*(K) := n \sum_{i=1}^K \left(\left(\frac{n+2}{n-i} \right)^{1/2} \hat{\rho}_\varepsilon(i) \right)^2 = n(n+2) \sum_{i=1}^K \frac{1}{n-i} \hat{\rho}_\varepsilon^2(i)$$

4.1.4 Vorhersage

Wir wollen eine Vorhersage für die zukünftige Entwicklung der Zeitreihe X treffen. Dazu bestimmen wir zunächst Gewichte $c_0^*, \dots, c_{n-1}^* \in \mathbb{R}$ so dass

$$\mathbb{E} \left(\left(X_{n+h} - \sum_{i=0}^{n-1} c_i^* X_{n-i} \right)^2 \right) = \min_{c_0, \dots, c_{n-1} \in \mathbb{R}} \mathbb{E} \left(\left(X_{n+h} - \sum_{i=0}^{n-1} c_i X_{n-i} \right)^2 \right)$$

Dann minimiert $\hat{X}_{n+h} := \sum_{i=0}^{n-1} c_i^* X_{n-i}$ den mittleren quadratischen Fehler; \hat{X}_{n+h} wird als beste h -Schritt-Vorhersage bezeichnet gegeben X_1, \dots, X_n . Das folgende Lemma sagt uns, wann die c_i^* optimal sind.

Lemma 4.1. *Sei X_t stoch. Prozess mit endlichen zweiten Momenten. Gilt für $c_0^*, \dots, c_{n-1}^* \in \mathbb{R}$, dass*

$$\mathbb{E} \left(X_i \left(X_{n+h} - \sum_{j=0}^{n-1} c_j^* X_{n-j} \right) \right) = 0, \quad i = 1, \dots, n \quad (4.36)$$

so ist $\hat{X}_{n+h} := \sum_{i=0}^{n-1} c_i^* X_{n-i}$ eine beste h -Schritt-Vorhersage.

Beweis. Ist $\tilde{X}_{n+h} := \sum_{i=0}^{n-1} c_i X_{n-i}$ Vorhersage für X_{n+h} gegeben X_1, \dots, X_n , so gilt

$$\begin{aligned} \mathbb{E}((X_{n+h} - \tilde{X}_{n+h})^2) &= \mathbb{E}((X_{n+h} - \hat{X}_{n+h} + \hat{X}_{n+h} - \tilde{X}_{n+h})^2) \\ &= \mathbb{E}((X_{n+h} - \hat{X}_{n+h})^2) + 2 \sum_{i=0}^{n-1} (c_i^* - c_i) \mathbb{E}(X_{n-i}(X_{n+h} - \hat{X}_{n+h})) \\ &\quad + \mathbb{E}((\hat{X}_{n+h} - \tilde{X}_{n+h})^2) \\ &= \mathbb{E}((X_{n+h} - \hat{X}_{n+h})^2) + \mathbb{E}((\hat{X}_{n+h} - \tilde{X}_{n+h})^2) \\ &\geq \mathbb{E}((X_{n+h} - \hat{X}_{n+h})^2) \end{aligned}$$

□

Ist X stationär mit Erwartungswert 0, so schreibt sich (4.36)

$$\rho(h+s) = \sum_{i=0}^{n-1} c_i^* \rho(s-i), \quad s := n-i = 0, 1, \dots, n-1$$

Ist also R_n (vgl.(2.20)) invertierbar, so ist

$$\begin{pmatrix} c_0^* \\ \cdot \\ \cdot \\ \cdot \\ c_{n-1}^* \end{pmatrix} := R_n^{-1} \begin{pmatrix} \rho(h) \\ \cdot \\ \cdot \\ \cdot \\ \rho(h+n-1) \end{pmatrix} \quad (4.37)$$

die gesuchte Lösung.

Die folgenden beiden Sätze erweitern dieses Resultat.

Satz 4.2. *Sei X_t AR-stationärer $AR(p)$ -Prozess mit $\mathbb{E}(X_0) = 0$ und $n \geq p$. Dann ist die beste Ein-Schritt-Vorhersage gegeben durch*

$$\hat{X}_{n+1} = \phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_p X_{n+1-p}$$

und rekursiv ergibt sich die beste h -Schritt-Vorhersage

$$\hat{X}_{n+h} = \phi_1 \hat{X}_{n+h-1} + \dots + \phi_{h-1} \hat{X}_{n+1} + \phi_h X_n + \dots + \phi_p X_{n+h-p}$$

Satz 4.3. Sei X_t AR-stationärer ARMA(p, q) Prozess mit $\mathbb{E}(\varepsilon_0) = 0$ und sei $n + q - p \geq 0$. Dann ist die beste h -Schritt-Vorhersage für $h > q$ gegeben durch die Rekursion

$$\hat{X}_{n+h} = \sum_{i=1}^p \phi_i \hat{X}_{n+h-i}$$

5 Prognosen

Gegeben sei eine Zeitreihe, die bis zum Zeitpunkt Z beobachtet wurde, bspw. der Verlauf eines Aktienkurses. Dann ist natürlich von Interesse, wie sich diese Zeitreihe in Zukunft entwickelt. Aus dem bekannten vergangenen Verlauf soll eine Prognose erstellt werden, mit der man die künftige Entwicklung voraussagen kann. Eine Möglichkeit haben wir bereits im Box-Jenkins-Programm kennengelernt.

Box-Jenkins-Ansatz

Sei ein invertierbares ARMA-Modell gegeben und (ε_t) habe die Martingaldifferenzeigenschaft bzgl. \mathcal{F}_t ; insbesondere gilt dann (2.27).

Bezeichne den Datensatz mit X_{t-n+1}, \dots, X_t und unsere Prognose mit $P_t X_{t+h}$. Für unsere Prognose verwenden wir $\mathbb{E}(X_{t+h} | \mathcal{F}_t)$. Dies minimiert den Fehler $\mathbb{E}((X_{t+h} - P_t X_{t+h})^2)$ unter allen Prognosen von X_{t+h} gegeben der Verlauf von X bis t . Für $h \geq 1$ wird $\mathbb{E}(X_{t+h} | \mathcal{F}_t)$ rekursiv berechnet; dabei benutzen wir $\mathbb{E}(\varepsilon_{t+h} | \mathcal{F}_t) = 0$ (MDF von (ε_t)) und die Invertierbarkeit des Modells. Am Beispiel des ARMA(1,1) Modells wird dies illustriert:

Beispiel 5.1 (ARMA(1,1)-Modell). Sei ein ARMA(1,1)-Modell mit Parametern μ, ϕ, θ der Form (2.27) gegeben. Mit $\mathbb{E}(\varepsilon_{t+1} | \mathcal{F}_t) = 0$ folgt

$$\mathbb{E}(X_{t+1} | \mathcal{F}_t) = \mu_{t+1} = \mu + \phi(X_t - \mu) + \theta \varepsilon_t \quad (5.38)$$

und

$$\mathbb{E}(X_{t+h} | \mathcal{F}_t) = \mu + \phi^h (X_t - \mu) + \phi \theta \varepsilon_t \quad (5.39)$$

Exponentielle Glättung

Die Daten seien jetzt Realisationen von Zufallsvariablen Y_{t-n}, \dots, Y_t (es wird keine Stationarität gefordert). Um Y_{t+1} vorherzusagen benutzen wir die Rekursionsgleichung

$$P_t Y_{t+1} = \alpha Y_t + (1 - \alpha) P_{t-1} Y_t, \quad \text{mit } 0 < \alpha < 1 \quad (5.40)$$

Es gilt dann

$$P_t Y_{t+1} = \sum_{i=0}^{n-1} \alpha (1 - \alpha)^i Y_{t-i} \quad (5.41)$$

α wird *Glättungsparameter* genannt; dieser gewichtet den Einfluss, den vergangene Werte haben. Kleines α entspricht dabei großer Gewichtung vergangener Werte im Vergleich zum gegenwärtigen Wert.