

ARMA Modelle

Lernziele

- Stationäre und nicht-stationäre Prozesse:
White noise und random walk
- ARMA: Autoregressive moving average Modelle
- Modellbildung
- Schätzung von ARMA Modellen
- Modellwahl und Modellüberprüfung
- Prognose
- Integrierte ARMA Modelle: ARIMA

Schwach stationäre Prozesse

Kovarianz oder schwach stationäre Prozesse

$$\{y_t\}, \quad t = \dots, -2, -1, 0, 1, 2, \dots$$

haben die Eigenschaft:

- **mittelwertstationär:**

$$E(y_t) = E(y_{t-s}) = \mu$$

- **kovarianzstationär:**

$$V(y_t) = E(y_t - \mu)^2 = V(y_{t-s}) = \sigma_y^2$$

$$\text{Cov}(y_t, y_{t-s}) = E(y_t - \mu)(y_{t-s} - \mu) = \text{Cov}(y_{t-j}, y_{t-s-j}) = \gamma_s$$

μ , σ_y^2 und γ_s sind **konstant** und **unabhängig** von t .

Autokorrelationsfunktion (ACF)

Die **Autokorrelation** zwischen y_t und y_{t-s} ist definiert als

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \text{Corr}(y_t, y_{t-s})$$

$$\gamma_s = \text{Cov}(y_t, y_{t-s}), \quad \gamma_0 = \sigma_y^2.$$

Im Speziellen gilt:

$$\rho_0 = 1 \quad \text{und} \quad -1 \leq \rho_s \leq 1.$$

Fasst man die $\rho_s, s \geq 0$, zusammen, erhält man die **Autokorrelationsfunktion, ACF**:

$$1, \rho_1, \rho_2, \rho_3, \dots$$

Beispiel

White noise, WN

Ein Prozess $\{y_t\}, y_t = \epsilon_t$, mit

$$E(\epsilon_t) = 0, \quad V(\epsilon_t) = \sigma_\epsilon^2, \quad \rho_0 = 1, \quad \rho_1 = \rho_2 = \rho_3 = \dots = 0$$

heißt *white noise*, WN, oder **Weißes Rauschen**.

Ein *white noise* Prozess ist (kovarianz-) stationär.

Wold Darstellung

Jeder schwach stationäre Prozess $\{y_t\}$ lässt sich als unendliche gewichtete Summe eines vergangenen white noise Prozesses darstellen:

$$y_t - \mu = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

Die ψ_j heißen **Impulse-response-Koeffizienten**; die Funktion $\{\psi_j, j \geq 0\}$, **Transferfunktion**, Impuls-Antwort-Funktion.

Sie erfüllt die Bedingung

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty$$

D.h., die ψ_j sind **quadratisch summierbar**.

Wold Darstellung / (2)

- Aus $E(\epsilon_t) = 0$ folgt

$$E(y_t) = \mu$$

- Aus $V(\epsilon_t) = \sigma_\epsilon^2$ und der quadratischen Summierbarkeit der ψ_j folgt

$$V(y_t) = \sum \psi_j^2 \sigma_\epsilon^2 = \sigma_\epsilon^2 \sum \psi_j^2$$

da die ϵ_t unkorreliert sind.

Lagoperator

Der **Lagoperator** L ist definiert als

$$Ly_t = y_{t-1}$$

Durch Mehrfaches anwenden des Lagoperators erhält man

$$L^2 y_t = Ly_{t-1} = y_{t-2}$$

$$L^3 y_t = L^2 y_{t-1} = Ly_{t-2} = y_{t-3},$$

...

$$L^s y_t = y_{t-s}$$

Beispiel:

$$(1 - L)y_t = y_t - y_{t-1},$$

$$(1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3)y_t = y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2} - \alpha_3 y_{t-3}.$$

$(1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3)$ heißt auch Lagpolynom der Ordnung 3.

Wold Darstellung mittels Lagoperator

Die Wold Darstellung kann auch mit Hilfe eines unendlichen Lagpolynoms $\Psi(L)$ angegeben werden:

$$y_t - \mu = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \sum_{j=0}^{\infty} \psi_j (L^j \epsilon_t) = \sum_{j=0}^{\infty} (\psi_j L^j) \epsilon_t = \Psi(L) \epsilon_t$$

wobei $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ ist.

Random Walk (RW)

Ein Prozess $\{y_t\}$ mit $y_t = y_{t-1} + \epsilon_t$
heißt *random walk*, RW, (ohne Drift).

Ein Prozess $\{y_t\}$ mit

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

heißt *random walk mit Drift*. c ist der Driftparameter.

Der Prozess ist in rekursiver Darstellung gegeben.

Explizite Darstellung des RWs:

$$y_t = y_0 + \sum_{j=1}^t \epsilon_j \quad \text{bzw.} \quad y_t = y_0 + ct + \sum_{j=1}^t \epsilon_j$$

Ein *random walk* ist **nicht stationär**.

Bedingter und unbedingter Erwartungswert

Die Informationsmenge I_t ist

$$I_t = \{y_t, \epsilon_t, y_{t-1}, \epsilon_{t-1}, \dots, y_1, \epsilon_1, y_0\}$$

Der **bedingte Erwartungswert** eines *random walks* y_t bezüglich der Informationsmengen I_{t-1} , I_{t-s} und I_0 ist

$$\begin{aligned} E(y_t | I_{t-1}) &= c + y_{t-1} \\ E(y_t | I_{t-s}) &= sc + y_{t-s} \\ E(y_t | I_0) &= tc + y_0 \end{aligned}$$

Die Abhängigkeit des bedingten Erwartungswertes vom Anfangswert verschwindet nicht mit $s \rightarrow \infty$.

Der unbedingte Erwartungswert $E(y_t)$ existiert nicht.

Bedingte und unbedingte Varianz

Die **bedingte Varianz** eines *random walks* y_t ist

$$\begin{aligned} V(y_t | I_{t-1}) &= \sigma_\epsilon^2 \\ V(y_t | I_{t-s}) &= s \sigma_\epsilon^2 \\ V(y_t | I_0) &= t \sigma_\epsilon^2 \end{aligned}$$

Die bedingte Varianz ist nicht konstant und nimmt ausgehend von $t = 0$ mit t zu.

Die unbedingte Varianz existiert nicht.

Die Kovarianz $\text{Cov}(y_t, y_{t+s})$ ist $t \sigma_\epsilon^2$.

Random Walk und dynamischer Multiplikator

Der *random walk* Prozess hat die Eigenschaft, dass vergangene Schocks nicht vergessen werden. Jeder (vergangene) Schock, ϵ_{t-s} , geht zur Gänze in das aktuelle Niveau, y_t , ein. Kein Schock wird vergessen.

$$y_t = y_0 + ct + \sum_{j=1}^t \epsilon_j \quad \Rightarrow \quad \frac{\partial y_t}{\partial \epsilon_{t-s}} = 1$$

Man sagt auch die **Persistenz** eines Schocks ist Eins.

Mit diesem Modell können *irreversible ökonomische Entscheidungen* beschrieben werden.

ARMA

Ein *autoregressiver moving average Prozess* der Ordnung (p, q) , **ARMA** (p, q) , ist ein schwach stationärer Prozess mit dem Bildungsgesetz

$$\alpha_p(L)(y_t - \mu) = \beta_q(L)\epsilon_t$$

wobei

$$\begin{aligned}\alpha_p(L) &= 1 - \alpha_1 L - \dots - \alpha_p L^p \\ \beta_q(L) &= 1 - \beta_1 L - \dots - \beta_q L^q\end{aligned}$$

$\alpha_p(L)$... **AR-Polynom** der Ordnung p

$\beta_q(L)$... **MA-Polynom** der Ordnung q

ϵ_t ... *white noise*.

ARMA / (2)

Beispiele:

ARMA(0,0), $\mu = 0$: $y_t = \epsilon_t$ white noise

ARMA(0,0), $\mu \neq 0$: $y_t = \mu + \epsilon_t$

AR(1): $(1 - \alpha_1 L)(y_t - \mu) = \epsilon_t$

MA(1): $(y_t - \mu) = (1 - \beta_1 L)\epsilon_t$

ARMA(1,1): $(1 - \alpha_1 L)(y_t - \mu) = (1 - \beta_1 L)\epsilon_t$

ARMA(1,2): $(1 - \alpha_1 L)(y_t - \mu) = (1 - \beta_1 L - \beta_2 L^2)\epsilon_t$

AR(12): $(1 - \alpha_1 L - \dots - \alpha_{12} L^{12})(y_t - \mu) = \epsilon_t$

Vergleich ARMA und Wold Darstellung

ARMA Modelle liefern eine Approximation der $\Psi(L)$ -Polynoms aus der Wold Darstellung mittels einer rationalen Funktion.

Durch Division (sofern zulässig) erhält man aus

$$\alpha_p(L)(y_t - \mu) = \beta_q(L)\epsilon_t$$

$$y_t - \mu = \frac{\beta(L)}{\alpha(L)}\epsilon_t = \Psi(L)\epsilon_t$$

Beispiel:

$$\text{ARMA}(1,0): y_t - \mu = \frac{1}{1 - \alpha_1 L}\epsilon_t = \sum_{j=0}^{\infty} \alpha_1^j \epsilon_{t-j} = \Psi(L)\epsilon_t$$

$$\text{ARMA}(1,1): y_t - \mu = \frac{1 - \beta_1 L}{1 - \alpha_1 L}\epsilon_t = \Psi(L)\epsilon_t$$

$$\text{MA}(\infty): y_t - \mu = (1 - \beta_1 L - \beta_2 L^2 - \dots)\epsilon_t = \beta_{\infty}(L)\epsilon_t = \Psi(L)\epsilon_t$$

Principle of Parsimony

Mittels ARMA-Modellen können alle (schwach-) stationären Prozesse dargestellt werden, sofern die Ordnung der Polynome groß genug gewählt wird: $p \rightarrow \infty$ oder $q \rightarrow \infty$.

In der Regel muss man annehmen, dass der zugrundeliegende Prozess sehr kompliziert ist, das eigentlich ein $\text{MA}(\infty)$ zur Modellierung notwendig wäre.

Bei der **Modellbildung** wird der zugrundeliegenden Prozess durch ein **sparsam parametrisiertes** ARMA-Modell (ARMA-Modell mit niedriger Ordnung) approximiert: *principle of parsimony*.

Das Problem besteht darin ein „gutes“ und zugleich sparsam parametrisiertes Modell zu finden.

AR(1) Prozess

Das Modell für einen AR(1) Prozess lautet:

$$(1 - \alpha L)(y_t - \mu) = \epsilon_t \quad \text{mit } |\alpha| < 1$$

oder

$$y_t - \alpha y_{t-1} = c + \epsilon_t \quad \text{mit } c = (1 - \alpha)\mu$$

Für $\mu = 0$ erhalten wir $y_t - \alpha y_{t-1} = \epsilon_t$.

Explizite Darstellung:

$$y_t = \alpha^t y_0 + \sum_{j=1}^t \alpha^{t-j} \epsilon_j \quad \text{bzw.} \quad y_t = \alpha^t y_0 + c \sum_{j=1}^{\tau-1} \alpha^j + \sum_{j=1}^t \alpha^{t-j} \epsilon_j$$

Erwartungswert und Varianz eines AR(1)

Aus der expliziten Darstellung erhalten wir direkt den **bedingten Erwartungswert** und die **bedingte Varianz** für $\mu = 0$:

$$E(y_t|y_0) = \alpha^t y_0 \quad \text{und} \quad V(y_t|y_0) = \sigma_\epsilon^2 \sum_{j=1}^t \alpha^{2(t-j)}$$

Die Abhängigkeit vom Anfangswert verschwindet mit $|\alpha| < 1$, wenn wir den Prozess im Zeitpunkt $-\infty$ starten lassen:

$$E(y_t|y_0) = \alpha^t y_0 \quad \rightarrow \quad E(y_t|y_{-\infty}) = E(y_t) = 0$$

$$V(y_t|y_0) = \sigma_\epsilon^2 \sum_{j=1}^t \alpha^{2(t-j)} \quad \rightarrow \quad V(y_t|y_{-\infty}) = V(y_t) = \frac{\sigma_\epsilon^2}{1 - \alpha^2}$$

Der **unbedingte Erwartungswert** ist konstant und gleich Null:

$$E(y_t) = 0.$$

Stationarität eines AR(1)

- Der AR(1) Prozess ist für

$$|\alpha| < 1$$

ein stationärer Prozess.

- Ein Prozess, gegeben durch die Differenzgleichung

$$y_t - \alpha y_{t-1} = y_t - \alpha L y_t = 0$$

ist stationär, wenn er einen **stabilen Fixpunkt** hat.

Das ist genau dann der Fall, wenn die Wurzeln des charakteristische Polynoms

$$1 - \alpha z = 0$$

außerhalb des Einheitskreises liegen. (Hier: $|z| = |1/\alpha| > 1$)

Autokorrelationsfunktion eines AR(1)

Die ACF fällt geometrisch (exponentiell):

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \frac{\text{Cov}(y_t, y_{t-s})}{V(y_t)} = \alpha^s$$

Ein AR(1) Prozess beschreibt ein **Vergessen vergangener Schocks**. Ein Schock, der s Perioden zurückliegt, wird mit $\psi_s = \alpha^s$ gewichtet, die ACF fällt daher mit α^s .

Allgemein gilt für AR(p)-Prozesse, dass die ACF (betragsmäßig) geometrisch fällt. Sie muss aber nicht monoton fallen wie beim AR(1).

Varianz und ACF sind von μ unabhängig.

Prognose eines AR(1)

Die τ -Schritt Prognose ist der bedingte Erwartungswert, $E(y_{t+\tau}|y_t)$.

Die Varianz des Prognosefehlers der τ -Schritt Prognose ist

$$V(y_{t+\tau}|y_t) = E\left([y_{t+\tau} - E(y_{t+\tau}|y_t)]^2|y_t\right)$$

Aus der Definition einer AR(1) erhalten wir $y_{t+1} = c + \alpha y_t + \epsilon_{t+1}$ mit $c = \mu(1 - \alpha)$.

$\tau = 1$ (1-Schritt Prognose):

$$E(y_{t+1}|y_t) = c + \alpha y_t, \quad V(y_{t+1}|y_t) = \sigma_\epsilon^2$$

$\tau = 2$ (2-Schritt Prognose):

$$E(y_{t+2}|y_t) = c(1 + \alpha) + \alpha^2 y_t, \quad V(y_{t+2}|y_t) = \sigma_\epsilon^2(1 + \alpha^2)$$

Prognose eines AR(1)

Allgemein (τ -Schritt Prognose):

$$E(y_{t+\tau}|y_t) = c \sum_{j=1}^{\tau-1} \alpha^j + \alpha^\tau y_t$$

$$V(y_{t+\tau}|y_t) = \sigma_\epsilon^2 \sum_{j=1}^{\tau} \alpha^{2(\tau-j)}$$

- Die Prognose konvergiert mit $\tau \rightarrow \infty$ gegen das arithmetische Mittel.
- Die Prognosevarianz konvergiert mit $\tau \rightarrow \infty$ gegen die unbedingte Varianz.
- Je größer $|\alpha|$ ist, desto langsamer erfolgt die Konvergenz.

MA(1) Prozess

Das **Modell** für einen MA(1) lautet

$$y_t - \mu = \epsilon_t - \beta \epsilon_{t-1} = (1 - \beta L)\epsilon_t$$

bzw.

$$y_t = \mu + \epsilon_t - \beta \epsilon_{t-1} = \mu + (1 - \beta L)\epsilon_t$$

Erwartungswert eines MA(1)

Der **bedingte Erwartungswert** $E(y_t|y_{t-1})$ ergibt sich als:

$$\begin{aligned} E(y_t|y_{t-1}) &= E(\mu + \epsilon_t - \beta \epsilon_{t-1} | \mu + \epsilon_{t-1} - \beta \epsilon_{t-2}) \\ &= E(\epsilon_t - \beta \epsilon_{t-1} | \epsilon_{t-1}, \epsilon_{t-2}) \\ &= \mu + \beta \epsilon_{t-1} \end{aligned}$$

$$\begin{aligned} E(y_t|y_{t-2}) &= E(\mu + \epsilon_t - \beta \epsilon_{t-1} | \mu + \epsilon_{t-2} - \beta \epsilon_{t-3}) \\ &= E(\mu + \epsilon_t - \beta \epsilon_{t-1} | \epsilon_{t-2}, \epsilon_{t-3}) \\ &= \mu \end{aligned}$$

$$E(y_t|y_{t-s}) = \mu \quad \text{für } s > 1$$

Die bedingten Erwartungswerts $E(y_t|y_{t-s})$ für $s > 1$ sind gleich dem **unbedingten Erwartungswert**:

$$E(y_t|y_{t-s}) = \mu = E(y_t) \quad (\text{für } s > 1)$$

Der Prozess hat ein Gedächtnis von genau einer Periode.

Varianz eines MA(1)

Für die **bedingten Varianzen** erhalten wir

$$V(y_t|y_{t-1}) = \sigma_\epsilon^2$$

$$V(y_t|y_{t-2}) = \sigma_\epsilon^2(1 + \beta^2)$$

$$V(y_t|y_{t-s}) = \sigma_\epsilon^2(1 + \beta^2) \quad \text{für } s > 1$$

da die Kovarianzen der ϵ_t Null sind.

Die bedingten Varianzen sind für $s > 1$ gleich der **unbedingten Varianz**:

$$V(y_t|y_{t-s}) = \sigma_\epsilon^2(1 + \beta^2) = V(y_t) \quad (\text{für } s > 1)$$

Die Varianz von y_t existiert immer, unabhängig davon, welchen Wert β annimmt.

Autokovarianz eines MA(1)

Die **Autokovarianzen** haben ebenfalls eine einfache Struktur.

$$\begin{aligned} \text{Cov}(y_t, y_{t-1}) &= \text{Cov}(\epsilon_t - \beta \epsilon_{t-1}, \epsilon_{t-1} - \beta \epsilon_{t-2}) \\ &= -\beta \text{Cov}(\epsilon_{t-1}, \epsilon_{t-1}) \\ &= -\beta V(\epsilon_{t-1}) \\ &= -\beta \sigma_\epsilon^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}(y_t, y_{t-2}) &= \text{Cov}(\epsilon_t - \beta \epsilon_{t-1}, \epsilon_{t-2} - \beta \epsilon_{t-3}) \\ &= 0 \end{aligned}$$

$$\text{Cov}(y_t, y_{t-s}) = 0 \quad \text{für } s > 1$$

Autokorrelationsfunktion eines MA(1)

Für die **Autokorrelationsfunktion** erhalten wir

$$\text{Corr}(y_t, y_{t-1}) = \frac{\text{Cov}(y_t, y_{t-1})}{V(y_t)} = \frac{-\beta \sigma_\epsilon^2}{(1+\beta^2)\sigma_\epsilon^2} = -\frac{\beta}{1+\beta^2}$$

$$\text{Corr}(y_t, y_{t-s}) = 0 \quad \text{für } s > 1$$

- Die ACF bricht nach dem Lag 1 ab.
- Für $s \geq 2$ zeigt die ACF das Muster der ACF eines white noise.
- Allgemein bricht die ACF eines MA(q) Prozesses nach dem Lag q ab.

Invertierbarkeitsbedingung

MA(1) Prozesse mit Parameter β und Parameter $(1/\beta)$ besitzen die selbe ACF, und haben somit die selben stochastischen Eigenschaften. Daher beschränkt man sich der Eindeutigkeit wegen auf den Bereich $|\beta| < 1$.

Diese Bedingung heißt **Invertierbarkeitsbedingung**.

Beispiel:

Der MA(1) Prozess $\{y_t\}$ mit $E(y_t) = 3$, $V(y_t) = 2.50$ und $\rho_1 = 0.40$ lässt sich auf 2 Arten darstellen:

$$y_t - 3 = u_t - 0.5 u_{t-1} \quad \text{mit } u_t \sim N(0, 2) \quad (\text{iid})$$

und

$$y_t - 3 = v_t - \frac{1}{0.5} v_{t-1} \quad \text{mit } v_t \sim N(0, 0.5) \quad (\text{iid})$$

Prognose eines MA(1)

Die τ -**Schritt Prognose** ist der bedingte Erwartungswert $E(y_{t+\tau}|y_t)$.

$\tau = 1$ (1-Schritt Prognose) für MA(1):

$$E(y_{t+1}|y_t) = \mu - \beta \epsilon_t, \quad V(y_{t+1}|y_t) = \sigma_\epsilon^2$$

τ -Schritt Prognose ($\tau > 1$):

$$E(y_{t+\tau}|y_t) = \mu, \quad V(y_{t+\tau}|y_t) = (1 + \beta^2) \sigma_\epsilon^2$$

- Ein MA(1) liefert nur für die 1-Schritt-Prognose eine kleinere Prognosevarianz als das arithmetische Mittel.
- Ein MA(q) liefert nur bis zur q -Schritt-Prognose eine kleinere Prognosevarianz als das arithmetische Mittel.

ARMA(p, q) Prozess

Zur Modellierung sollen ARMA Prozesse folgende Bedingungen erfüllen:

- **Stationarität:** die Wurzeln (Nullstellen) des AR-Polynoms liegen außerhalb des Einheitskreises.
- **Invertierbarkeit:** die Wurzeln des MA-Polynoms liegen ausserhalb des Einheitskreises.

Die Parameter des Prozesses müssen geschätzt, die Modellvoraussetzungen überprüft werden.

Schätzer für Mittel und Varianz

Der **Mittelwert** (arithmetisches Mittel) einer Reihe y_1, \dots, y_T der Länge T ist

$$\bar{y} = \hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$$

μ wird in der Regel durch \bar{y} ersetzt.

Die **Varianz** ist gegeben als

$$\hat{\sigma}_y^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2$$

Sie ist nur dann brauchbar, wenn nur wenig Autokorrelation in y_t vorliegt oder die Reihe sehr lang ist.

t -Test für das Mittel

Das Mittel ist für große T ungefähr normalverteilt

$$\bar{y} \sim N(\mu, \sigma_y^2/T)$$

sofern die zugrundeliegende Verteilung eine Varianz besitzt und die Autokorrelationen in der Reihe nicht zu stark sind.

Die Verteilung des Stichprobenmittels kann zum Testen auf das Mittel verwendet werden.

$$H_0 : \mu = 0 \quad H_1 : \mu > 0 \quad Z = \frac{\bar{y}}{\sqrt{\hat{\sigma}_y^2/T}} \sim N(0,1)$$

Die Prüfgröße Z ist für große T (näherungsweise) standardnormalverteilt.

Test auf die Differenz zweier Mittel

Test auf Gleichheit der Mittelwerte in A und B :

$$H_0 : \mu_A = \mu_B \quad H_1 : \mu_A \neq \mu_B$$

Die Prüfgröße

$$Z = \frac{(\bar{y}_A - \bar{y}_B)}{\sqrt{\hat{\sigma}_A^2/T_A + \hat{\sigma}_B^2/T_B}} \sim N(0,1)$$

ist (näherungsweise) standardnormalverteilt.

Voraussetzung:

- Stichproben sind voneinander unabhängig.
- Beobachtungen sind nicht oder nur schwach autokorreliert.
- Die Stichprobenumfänge T_A und T_B sind hinreichend groß.

Schiefe einer Verteilung

Eine Maßzahl für die **Schiefe** ist

$$\hat{S} = \frac{1}{T\hat{\sigma}_y^3} \sum_{t=1}^T (y_t - \bar{y})^3$$

Die Verteilung ist

$$\left. \begin{array}{l} \text{linksschief} \\ \text{symmetrisch} \\ \text{rechtsschief} \end{array} \right\} \dots S \left\{ \begin{array}{l} < \\ = \\ > \end{array} \right\} 0 \dots \mu \left\{ \begin{array}{l} < \\ = \\ > \end{array} \right\} \tilde{\mu}$$

$\tilde{\mu}$ bezeichnet den Median.

Die Normalverteilung ist symmetrisch. Ihre Schiefe ist Null.

Kurtosis einer Verteilung

Eine Maßzahl für die **Kurtosis** (Wölbung)

$$\hat{K} = \frac{1}{T\hat{\sigma}_y^4} \sum_{t=1}^T (y_t - \bar{y})^4$$

$(\hat{K} - 3)$ heißt **Exzess-Kurtosis**.

- Die Normalverteilung besitzt eine Kurtosis von 3.
- Ist die Kurtosis größer als 3, so besitzt die Verteilung höhere Schwänze als die Normalverteilung (**leptokurtische** Verteilung).

Beispiel:

Die t -Verteilung mit ν Freiheitsgraden ist symmetrisch.

Ihre Kurtosis beträgt $3(\nu - 2)/(\nu - 4)$.

Test auf Normalverteilung

Die Schätzer \hat{S} und \hat{K} sind bei Zugrundeliegen einer Normalverteilung näherungsweise normalverteilt:

$$\hat{S} \sim N(0, 6/T) \quad \hat{K} \sim N(3, 24/T)$$

Der **Jarque-Bera Test** testet simultan auf $S = 0$ und $K = 3$ einer Verteilung.

Unter der Nullhypothese einer normalverteilten Zufallsvariable ist die Teststatistik χ^2 verteilt mit 2 Freiheitsgraden:

$$\frac{T-k}{6} \left[\hat{S}^2 + \frac{1}{4} (\hat{K} - 3)^2 \right] \sim \chi^2_2$$

- Beim Test der Verteilung einer Variablen ist $k = 0$.
- Testet man Residuen einer Regression auf Normalverteilung, ist k die Anzahl der Regressoren.

Autokorrelation eines ARMA(p, q)

Die Autokorrelationsfunktion eines ARMA(p, q)-Prozesses ergibt sich durch Überlagerung der ACF des AR- und MA-Teils.

Die ersten q Werte werden durch die AR- und MA-Komponenten gemeinsam, das Ausklingen durch das AR-Polynom bestimmt.

Der Schätzer für die ACF ist die

Stichprobenautokorrelationsfunktion (SACF):

$$\hat{\rho}_k = r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Die Autokorrelationen der Ordnungen $0, 1, 2, 3, \dots$ werden zum Korrelogramm zusammengefasst.

Verteilung von $\hat{\rho}_k (= r_k)$

Die SACF eines (normalverteilten) **stationären Prozesses**, $\{y_t\}$, mit $\rho_k \neq 0$ für $k \leq m$ und $\rho_s = 0$ für $s > m$ hat für $s > m$

$$E(\hat{\rho}_s) = -\frac{1}{T} \quad \text{und} \quad V(\hat{\rho}_s) = \frac{1}{T} \left(1 + 2 \sum_{k=1}^{s-1} \rho_k^2 \right)$$

Für einen *white noise* $y_t = \epsilon_t$ gilt: $m = 0$, $\rho_s = 0$ für $s > m = 0$.

$$E(\hat{\rho}_s) = -\frac{1}{T} (\approx 0) \quad \text{und} \quad V(\hat{\rho}_s) = \frac{1}{T}$$

Die $\hat{\rho}_k$ sind normalverteilt: $\hat{\rho}_k \sim N(-\frac{1}{T}, \frac{1}{T})$. (\Rightarrow Test auf *white noise*)

Test auf white noise: Omnibustest

Die **Box-Pierce-Statistik**, Q , testet, ob bestimmte Autokorrelationskoeffizienten eines stationären Prozesses gleich Null sind.

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_s = 0 \quad H_1 : \rho_k \neq 0 \quad \text{für ein } k, 1 \leq k \leq s$$

Die Box-Pierce-Statistik ist χ^2 -verteilt mit s Freiheitsgraden.

$$Q = T \sum_{k=1}^s \hat{\rho}_k^2 \sim \chi_s^2$$

Die **Ljung-Box-Statistik**, \tilde{Q} , ist für kleine T besser geeignet.

$$\tilde{Q} = T(T+2) \sum_{k=1}^s \frac{\hat{\rho}_k^2}{T-k} \sim \chi_s^2$$

Schätzen eines ARMA Modells

Schätzen eines Modells bedeutet, die Parameter des Modells so zu wählen, dass die Eigenschaften des geschätzten Modells möglichst gut mit den Eigenschaften der beobachteten Reihe übereinstimmen.

Man sagt auch: **Das Modell wird an die Daten angepasst.**

Es gibt verschiedene Methoden ARMA Modelle zu schätzen, u.a.

- eine Kleinst-Quadrat-Methode, die *Unconditional Least Squares* Methode, ULS.
- die *Maximum Likelihood* Methode, ML, basierend auf der Normalverteilung

Unconditional Least Squares Methode, ULS

Wir schreiben das ARMA(p, q) Modell wie folgt an

$$y_t = c + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \epsilon_t - \beta_1 \epsilon_{t-1} - \dots - \beta_q \epsilon_{t-q}$$

Die Parameter α_i und β_j werden analog zur linearen Regression so gewählt, dass

$$\sum_{t=1}^T \epsilon_t^2 \rightarrow \min$$

Für die ersten Beobachtungen fehlen aber die notwendigen Werte fuer y_{t-i} und ϵ_{t-i} wenn $t-i \leq 0$ ist.

Diese werden durch die Mittel \bar{y} bzw. 0 geschätzt:

$$y_{-j} = \bar{y} \quad \text{und} \quad \epsilon_{-k} = 0 \quad \text{für } j, k \geq 0$$

Eigenschaften der ULS

Die Schätzmethode heißt **unconditional least squares** methode (unbedingte Methode der Kleinsten Quadrate).

Die Schätzung verwendet das unbedingte Mittel der Zufallsvariablen für die nicht beobachteten Werte.

Die geschätzten Parameter können (in großen Stichproben) analog zum bekannten Regressionsmodell getestet werden.

Diese sind

- asymptotisch unverzerrt,
- konsistent und
- asymptotisch normal verteilt.

Maximum Likelihood Schätzer

Beim *Maximum Likelihood* Schätzer (ML) werden die Parameter so gewählt, dass sie am besten zu den Beobachtungen passen; also am wahrscheinlichsten (*likely*) sind.

Dazu wird die sogenannte *Likelihood*-Funktion maximiert.

Beispiel:

Wir suchen den ML Schätzer für den Mittelwert μ einer normalverteilten ZV mit Varianz $\sigma^2 = 1$:

$$L(\mu; x_1, \dots, x_n) = 1/(\sqrt{2\pi})^n \exp\left(-\sum_{i=1}^n (x_i - \mu)^2 / 2\right)$$

Differenzieren nach μ und Null setzen ergibt

$$\sum_{i=1}^n (x_i - \mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

Maximum Likelihood Methode für ARMA

Wir gehen einem multinormalverteilten Zufallsvektor $\mathbf{y} = (y_1, \dots, y_T)'$ aus mit

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$\boldsymbol{\mu}$ ist der Vektor der (konstanten) Mittel für die einzelnen Zeitpunkte.

$\boldsymbol{\Sigma}$ ist die zugehörige Kovarianzmatrix.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & & \gamma_{T-3} \\ \vdots & & & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \dots & \gamma_0 \end{pmatrix}$$

Likelihood für AR(1)

Die Kovarianzen eines AR(1) Prozesses sind

$$\text{Cov}(y_t, y_{t+s}) = \gamma_s = \alpha^s \sigma_y^2 \quad \text{wobei} \quad \sigma_y^2 = \gamma_0 = \sigma_\epsilon^2 / (1 - \alpha^2)$$

Daher ist Σ eine Funktion von α und σ_ϵ^2 .

$$\Sigma = \Sigma(\alpha, \sigma_\epsilon^2) = \frac{\sigma_\epsilon^2}{1 - \alpha^2} \begin{pmatrix} 1 & \alpha & \dots & \alpha^{T-1} \\ \alpha & 1 & & \alpha^{T-2} \\ \dots & & \ddots & \dots \\ \alpha^{T-1} & \alpha^{T-2} & \dots & 1 \end{pmatrix}$$

Die *Likelihood*-Funktion ist bei gegebenen Daten $\mathbf{y} = (y_1, \dots, y_T)'$

$$L(\mu, \alpha, \sigma_\epsilon^2 | \mathbf{y}) = (2\pi)^{-T/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})'\right)$$

Likelihood für MA(1)

Die Kovarianzen eines MA(1) Prozesses sind

$$\text{Cov}(y_t, y_t) = \gamma_0 = (1 + \beta^2) \sigma_\epsilon^2,$$

$$\text{Cov}(y_t, y_{t+1}) = \gamma_1 = -\beta \sigma_\epsilon^2, \text{ und}$$

$$\text{Cov}(y_t, y_{t+s}) = \gamma_s = 0, \text{ für } s > 1.$$

Daher ist Σ eine Funktion von β und σ_ϵ^2 .

$$\Sigma = \Sigma(\beta, \sigma_\epsilon^2) = \sigma_\epsilon^2 \begin{pmatrix} 1 + \beta^2 & -\beta & 0 & \dots & 0 \\ -\beta & 1 + \beta^2 & -\beta & & 0 \\ \dots & & & \ddots & 0 \\ 0 & 0 & 0 & \dots & 1 + \beta^2 \end{pmatrix}$$

Die *Likelihood*-Funktion ist bei gegebenen Daten $\mathbf{y} = (y_1, \dots, y_T)'$

$$L(\mu, \beta, \sigma_\epsilon^2 | \mathbf{y}) = (2\pi)^{-T/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})'\right)$$

Maximierung der Log-Likelihood

Allgemein ist für ein beliebiges ARMA(p, q) Modell Σ eine Funktion von allen Parametern und der Fehlervarianz

$$\Sigma = \Sigma(\alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q; \sigma_\epsilon^2)$$

μ wird durch \bar{y} ersetzt, und die Log-Likelihood Funktion numerisch maximiert (bzw. minus Log-Likelihood minimiert).

Numerische Minimierungsverfahren sind u.a.

- Grid-Search
- Simplex-Verfahren
- Methode des steilsten Abstiegs
(*method of steepest decent*)

Numerische Verfahren

- **Grid-Search:** Dieses einfachste Verfahren dient nur zur Illustration. Man gibt ein grobes Raster (Gitter) z.B. im Suchraum $(\alpha, \sigma_\epsilon^2)$ vor. Dann probiert man alle gewählten Paare durch. In der Umgebung des gefunden Minimums verfeinert man das Gitter, usw.
- **Methode des steilsten Abstiegs:** Der Gradient nach den zu minimierenden Parametern gibt die Richtung des steilsten Abstiegs. In diese Richtung wird ausgehend von einem Startwert der nächste Wert gesucht, usw., bis sich der Funktionswert kaum mehr ändert.
- **Simplex-Verfahren:** Dieses ist nicht mit der Linearen Optimierung zu verwechseln. Es umgeht die Berechnung der Ableitungen in der Methode des steilsten Abstiegs.

Eigenschaften der ML

Die geschätzten Parameter können (in großen Stichproben) analog zum bekannten Regressionsmodell getestet werden.

Diese sind

- asymptotisch unverzerrt,
- konsistent,
- asymptotisch normal verteilt, und
- asymptotisch effizient.

Signifikanz der Parameter

Wir wollen nun ein geschätztes ARMA Modell auf seine Eignung überprüfen.

Angenommen wir haben eine Ordnung (p, q) gewählt und die Parameter des ARMA(p, q) geschätzt.

Kriterien für ein brauchbares Modell:

- 1 **Parameter α_p und β_q signifikant (von Null verschieden).**
Man führt dazu den t -Test für die Parameter durch, die zum größten Lag im AR- bzw MA-Polynom gehören, $\hat{\alpha}_p$ und $\hat{\beta}_q$. Ist z.B. $\hat{\alpha}_p$ nicht signifikant von Null verschieden, so sollte die AR-Ordnung um 1 reduziert oder erhöht werden. „Mittlere“ Parameter „dürfen“ Null sein.

Residuen

④ Residuen des geschätzten Modells sind *white noise*.

Ist die gesamte Autokorrelationsstruktur der Reihe durch unser Modell erfasst?

Dazu plottet man die Residuen und kontrolliert visuell die SACF der Residuen. Dann führt man den Ljung-Box Test auf Autokorrelation in den Residuen durch. Sind die Residuen nicht *white noise*, so hat das Modell nicht die gesamte Autokorrelationsstruktur der Reihe erfasst: p, q sind zu klein.

④ Modellierte Reihe erfasst weitgehend die wichtigen Charakteristika.

Dies sollte stets die erste Frage sein. Dazu vergleicht man visuell den Pfad der Reihe, den das Modell erzeugt, mit dem beobachteten. Interessant dabei ist z.B. ob Wendepunkte (Richtungswechsel), und wieviele davon, erfasst werden.

Prognose

④ Modell liefert vernünftige Prognosen.

Dazu führt man eine Prognose, sowohl *ex post*, als auch *ex ante* durch, und misst die Prognosegüte mit z.B. dem RMSE.

In der Praxis stellt sich oft eine anderes als das „beste“ geschätzte Modell als bestes Prognosemodell heraus.

Nun stellt sich die Frage, ob die Modellwahl, die Wahl der Ordnungen p und q , richtig war, selbst wenn das geschätzte Modell alle obigen Kriterien zu erfüllen scheint.

Modellwahl

Ziel der Anpassung eines Modells ist es,

- die Fehlerquadratsumme zu minimieren, oder
- die Likelihood zu maximieren.

Nimmt man eine weitere Variable in die Gleichung auf, so wird sie i.A. die Fehlerquadratsumme zumindest ein wenig verringern.

Die (vermeintliche?) Anpassung eines Modells an die Daten wird immer besser, je mehr Parameter in das Modell aufgenommen werden.

Frage: Wie viele Variable (hier Lags) sollen aufgenommen werden?

Die oben besprochenen Kriterien sind nicht immer einfach zu handhaben.

Informationskriterium und *Parsimony*

Die am leichtesten handhabbaren Instrumente zur Modellwahl sind **Informationskriterien**. Wir geben

- **Akaikes Informationskriterium**, AIC, und das
- **Schwarz' Kriterium**, auch Schwarz'-Bayes'sche Informationskriterium, SIC (SBC) an.

Sie helfen auch (non nested) Modelle wie ARMA(1,2) und ARMA(2,1) zu vergleichen.

Informationskriterien können durch das **Prinzip der Sparsamkeit** (*parsimony*) motiviert werden.

Um die vermeintliche Anpassung des Modells an die Daten bei Hinzunahme weiterer Parameter in das Modell zu vermeiden, wird eine eine Bestrafung, ein **Pönale**, für jede eingeführte Variable verwendet.

Informationskriterien: AIC und SIC

- **Akaike's Informationskriterium, AIC**, ist für ARMA(p, q) Prozesse gegeben durch

$$\text{AIC}(p, q) = \log(\hat{\sigma}_{\epsilon, p+q}^2) + \frac{2}{T}(p + q)$$

- **Schwarz'-Informationskriterium, SIC**, bzw. **Schwarz-Bayes'sche Informationskriterium, SBC**, ist gegeben durch

$$\text{SIC}(p, q) = \log(\hat{\sigma}_{\epsilon, p+q}^2) + \frac{\log(T)}{T}(p + q)$$

$\hat{\sigma}_{\epsilon, p+q}^2$ steht für die minimierte Fehlervarianz.
 $\log(\cdot)$ steht stets für den natürlichen Logarithmus.

Informationskriterien: Anzahl an Lags

Frage: Wie viele Variable (hier Lags) sollen aufgenommen werden?

- Laut AIC wird nur dann eine Variable (Lag) aufgenommen, wenn sie zumindest die (logarithmierte) Fehlerquadratsumme um $(2/T)$ reduziert.
- Ähnlich, aber mit einem größeren Pönale bewertet das SIC: $\log(T)/T$.

Es werden somit solange Variable (Lags) aufgenommen, bis die letzte Variable das geforderte Mindestmaß an Reduktion der Fehlerquadratsumme unterschreitet.

Modellsuche

Praktisch geht man so vor, dass man *a priori* eine maximale Ordnung für p und q annimmt, und alle Modelle bis zu diesen Ordnungen schätzt.

Die maximale Ordnung hängt davon ab, ob man saisonale Effekte berücksichtigen muss oder nicht, und, welche Struktur die ACF erahnen lässt.

Beispiel: $p_{max} = 4$ und $q_{max} = 4$.

Es sind $(4 + 1) \times (4 + 1) = 25$ Modelle zu schätzen, aus denen wir eines auswählen werden.

- Für jedes in Betracht kommende Modell werden AIC oder SIC berechnet.
- Es wird nun das Modell gewählt, das das kleinste AIC bzw. kleinste SIC aufweist.

AIC, SIC in Log-Likelihood-Funktion

Eine andere Darstellung für die Informationskriterien erhält man, wenn folgende Näherung verwendet wird:

$$\log(\hat{\sigma}_{\varepsilon, p+q}^2) \approx -\frac{2}{T} l(\hat{\theta}_{p+q})$$

- $\hat{\theta}_{p+q}$ ist der Vektor der geschätzten Parameter des Modells,

$$\hat{\theta}_{p+q} = (\hat{\alpha}_1, \dots, \hat{\alpha}_p, \hat{\beta}_1, \dots, \hat{\beta}_q, \hat{\sigma}_\varepsilon^2)$$

- $l(\hat{\theta}_{p+q})$ steht für den Logarithmus der maximierten Likelihood-Funktion des Modells, $l(\theta_{p+q}) = \log(L(\theta_{p+q}))$.

Die Likelihood-Funktion wird **maximiert**, während die Fehlerquadratsumme **minimiert** wird.

Trendstationarität

Ein Prozess heißt **trendstationär** wenn er durch Berücksichtigung eines linearen Trends in einen stationären Prozess transformiert werden kann.

Man modelliert üblicherweise Trend und stationären Teil simultan.

Beispiel: x_t trendstationär

$$x_t = a + b t + u_t \quad \text{bzw.} \quad x_t - (a + b t) = u_t,$$

wobei u_t ein stationärer Prozess ist.

Man regrestriert x_t auf die Zeit t , und modelliert den stationären Teil u_t (z.B. als AR(3)).

Differenzenstationarität

Ein Prozess heißt **differenzstationär** wenn er durch Differenzenbildung in einen stationären Prozess transformiert werden kann.

Hier wird die Veränderung zwischen den Beobachtungen x_t und x_{t-1} berechnet und als neue Variable modelliert:

$$y_t = (1 - L)x_t$$

Beispiel: x_t differenzstationär

$$x_t = c + x_{t-1} + v_t \quad \text{bzw.} \quad x_t - x_{t-1} = c + v_t$$

wobei v_t ein stationärer Prozess ist.

Beispiel: *random walk*, $x_t = x_{t-1} + \epsilon_t$.

Differenzenstationär: ARIMA(p, d, q)

Man bezeichnet differenzenstationäre Prozesse als **ARIMA(p, d, q)**, **integrierte ARMA Modelle** mit Integrationsordnung d .

$d = 1$, ARIMA($p, 1, q$):

Hier wird eine einmalige Differenzenbildung durchgeführt.

Durch **Integration** (Summierung) der y_t , $y_t = (1 - L)x_t$, kann wieder der ursprüngliche Prozess x_t erzeugt werden: $x_t = x_{t-1} + y_t$

$$x_t = x_0 + \sum_{j=1}^t y_j$$

Integrierte Prozesse zeigen in der SACF einen **sehr langsamen Abfall**, vergleichbar mit dem eines random walks.

ARIMA(p, d, q) – Beispiele

- Ein ARIMA($p, 0, q$) Prozess ist ein ARMA Prozess. $d = 0$.
- Ein Random Walk ist ein ARIMA($0, 1, 0$), $d = 1$.
- Ist x_t ein ARIMA($p, 2, q$), dann ist die 2-malige Differenz

$$y_t = (1 - L)^2 x_t$$

stationär.

Saisonaler ARMA (stationär)

Saison ist ein mit fixer Periodizität, S , wiederkehrendes (stochastisches) Verhalten.

Beispiele:

- Wochentagsmuster in Tagesdaten von 5-Tages-Wochen, $S = 5$,
- Monatsmuster in Monatsdaten, $S = 12$,
- Quartalsmuster in Quartalsdaten, $S = 4$.

Die ACF zeigt hier Ausschläge bei Lag S und Vielfachen davon.

Im ARMA Modell führt man dann einen Lag von S (evt. Vielfache davon) ein.

Additive und multiplikative SARMA (stationär)

Man unterscheidet

- **additive saisonale ARMA**, z.B.,

$$(1 + \alpha_S L^S + \alpha_{(2S)} L^{2S})(y_t - \mu) = \epsilon_t$$

mit dem saisonalen AR-Polynom $\alpha_P(L^S)$, und

- **multiplikative saisonale ARMA** Modelle

$$\alpha_p(L) A_P(L^S)(y_t - \mu) = \beta_q(L) B_Q(L^S) \epsilon_t$$

mit den saisonalen AR-Polynom $A_P(L^S)$ und dem saisonalen MA-Polynom $B_Q(L^S)$. Man schreibt auch

$$\text{ARMA}(p, q)(P, Q)_S$$