

Dynamische Systeme und Zeitreihenanalyse

Autoregressive moving average Modelle *Kapitel 12*

Statistik und Mathematik – WU Wien

Michael Hauser

michael.hauser@wu-wien.ac.at – (2003)

Dynamische Systeme und Zeitreihenanalyse // Autoregressive moving average Modelle – 12 – p.0??

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

michael.hauser@wu-wien.ac.at – (2003)

Dynamische Systeme und Zeitreihenanalyse // Autoregressive moving average Modelle – 12 – p.1??

Stationäre Prozesse

michael.hauser@wu-wien.ac.at – (2003)

Dynamische Systeme und Zeitreihenanalyse // Autoregressive moving average Modelle – 12 – p.2??

Schwach stationäre Prozesse

Wir betrachten nun Zufallsvariable y_t und nennen die Folge

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

einen **stochastischen Prozess in diskreter Zeit**.

Prozesse mit den folgenden Eigenschaften heißen **kovarianz** oder **schwach stationär**:

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

● **kovarianzstationär**

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

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

$\mu, \sigma_y^2, \gamma_s$ sind *konstant* und unabhängig von t .

ACF

Die Kovarianz ist symmetrisch:

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

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

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

$\gamma_0 = \sigma_y^2$. γ_s sind die Kovarianzen.

Im Speziellen gilt: $\rho_0 = 1$ und $-1 < \rho_s < 1$.

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

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

Beispiel: White noise, WN

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

$$E(\epsilon_t) = 0, \quad V(\epsilon_t) = \sigma_\epsilon^2,$$

$$\rho_0 = 1, \quad \rho_1 = 0, \quad \rho_2 = 0, \quad \rho_3 = 0, \quad \dots$$

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

Der white noise Prozess ist (kovarianz)stationär.

Bemerkung: Für ϵ_t gilt:

$$V(\epsilon_t) = E[(\epsilon_t - E[\epsilon_t])^2] = E[\epsilon_t^2]$$

Für $s \neq 0$:

$$\begin{aligned} \gamma(s) &= \text{Cov}(\epsilon_t, \epsilon_{t+s}) = E[(\epsilon_t - E[\epsilon_t])(\epsilon_{t+s} - E[\epsilon_{t+s}])] \\ &= E[\epsilon_t \epsilon_{t+s}] = 0 \end{aligned}$$

Wold Darstellung

Wold Darstellung eines stationären Prozesses

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** oder als Ganzes $\{\psi_j, j \geq 0\}$ **Transferfunktion**, Impuls-Antwort-Funktion. Sie erfüllen die Bedingung

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

Dh, die ψ_j sind *quadratisch summierbar*.

Wold Darstellung (Fs)

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

$$E(y_t) = \mu$$

- Aus 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, Differenzenoperator

Der **Lagoperator** L ist definiert als (vgl Δ bei LDGen)

$$Ly_t = y_{t-1}, \quad L^2 y_t = y_{t-2}, \quad L^3 y_t = y_{t-3}, \quad \dots$$

Zum Beispiel ist

$$(1 - L)y_t = y_t - y_{t-1}, \quad \text{oder}$$

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

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

Der **Differenzenoperator** ist definiert als

$$\Delta = 1 - L$$

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

$$\Delta^2 y_t = (1 - L)^2 y_t = (1 - 2L + L^2)y_t = y_t - 2y_{t-1} + y_{t-2}$$

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

Ein nicht-stationärer Prozess: Der random walk, RW

Rekursion für den random walk

Sei ϵ_t white noise. 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.

Ein random walk ist **nicht stationär**.

Explizite Darstellung des RWs

Gehen wir von einem RW aus, der in $t = 0$ mit y_0 (fix) startet:

$$y_0$$

$$y_1 = y_0 + \epsilon_1$$

$$y_2 = y_1 + \epsilon_2 = (y_0 + \epsilon_1) + \epsilon_2 = y_0 + (\epsilon_1 + \epsilon_2)$$

$$y_3 = y_2 + \epsilon_3 = (y_1 + \epsilon_2) + \epsilon_3 = y_0 + (\epsilon_1 + \epsilon_2 + \epsilon_3)$$

...

$$y_t = y_{t-1} + \epsilon_t = y_0 + \sum_{j=1}^t \epsilon_j$$

Explizite Darstellung des RWs:

$$y_t = y_0 + \sum_{j=1}^t \epsilon_j$$

Bedingte und unbedingte Erwartung

Sei die Informationsmenge I_t

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

Die **bedingte Erwartung** von y_t bezüglich der Informationsmengen I_{t-1} , I_{t-s} und I_0 ist

$$E(y_t | I_{t-1}) = y_{t-1}$$

$$E(y_t | I_{t-s}) = y_{t-s}$$

$$E(y_t | I_0) = y_0$$

Die Abhängigkeit der bedingten Erwartung vom Anfangswert verschwindet nicht mit $s \rightarrow \infty$.

Die *unbedingte Erwartung* $E(y_t)$ existiert nicht.

Berechnungen

Es gilt für ϵ_t , y_t und 2 beliebige ZVen X und Y :

$$E(\epsilon_t | I_{t-s}) = E(\epsilon_t) = 0, \quad s > 1$$

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

$$E(a + bX + cY) = a + bE(X) + cE(Y)$$

$$V(a + bX + cY) = b^2 V(X) + 2bc \text{Cov}(X, Y) + c^2 V(Y)$$

Berechnungen (Fs)

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

$$E(y_t | y_{t-1}) = y_{t-1} + E(\epsilon_t) = y_{t-1}$$

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

$$I_0: y_t = y_0 + \sum_{j=1}^t \epsilon_j$$

$$E(y_t | y_0) = y_0 + E(\sum_{j=1}^t \epsilon_j) = y_0$$

$$V(y_t | y_0) = V(\sum_{j=1}^t \epsilon_j) = E(\sum_{j=1}^t \epsilon_j)^2 = E(\sum_{j=1}^t \epsilon_j^2) = t \sigma_\epsilon^2$$

$$\gamma(t, t+s) = \text{Cov}(y_t, y_{t+s}): y_{t+s} = y_t + \sum_{j=1}^s \epsilon_{t+j}$$

$$\text{Cov}(y_t, y_{t+s}) = \text{Cov}(y_t, y_t + \sum_{j=1}^s \epsilon_{t+j}) = \text{Cov}(y_t, y_t) = V(y_t)$$

Bedingte und unbedingte Varianz

Die bedingte Varianz ist

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

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

$$V(y_t | I_0) = t \sigma_\epsilon^2$$

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

Die unbedingte Varianz existiert nicht.

Die Kovarianz $\text{Cov}(y_t, y_{t+s} | I_0)$ ist $t \sigma_\epsilon^2$.

Die Korrelation $\text{Corr}(y_t, y_{t+s} | I_0) = \sqrt{t} / \sqrt{t+s}$.

Explizite Darstellung des RWs mit Drift

Gehen wir von einem RW mit Drift aus, der in $t = 0$ mit y_0 fix startet:

$$y_0$$

$$y_1 = c + y_0 + \epsilon_1$$

$$y_2 = c + y_1 + \epsilon_2 = c + (c + y_0 + \epsilon_1) + \epsilon_2 = 2c + y_0 + (\epsilon_1 + \epsilon_2)$$

$$y_3 = c + y_2 + \epsilon_3 = c + (c + y_1 + \epsilon_2) + \epsilon_3 = 3c + y_0 + (\epsilon_1 + \epsilon_2 + \epsilon_3)$$

...

$$y_t = c + y_{t-1} + \epsilon_t = y_0 + ct + \sum_{j=1}^t \epsilon_j$$

Explizite Darstellung des RWs mit Drift:

$$y_t = y_0 + ct + \sum_{j=1}^t \epsilon_j$$

Bedingte/unbed Erwartung, RW mit Drift

Sei die Informationsmenge I_t

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

Die **bedingte Erwartung** von y_t bezüglich der Informationsmengen I_{t-1} , I_{t-s} und I_0 ist

$$E(y_t | I_{t-1}) = c + y_{t-1}$$

$$E(y_t | I_{t-s}) = s c + y_{t-s}$$

$$E(y_t | I_0) = t c + y_0$$

Die Abhängigkeit der bedingten Erwartung vom Anfangswert verschwindet nicht mit $s \rightarrow \infty$. (vgl LDG $y_t - y_{t-1} = c$)

Die **unbedingte Erwartung** $E(y_t)$ existiert nicht.

Bedingte/unbed Varianz, RW mit Drift

Die **bedingte Varianz** ist

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

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

$$V(y_t | I_0) = t \sigma_\epsilon^2$$

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

Die unbedingte Varianz existiert nicht.

Die Kovarianz zwischen $t, t + s$ ist $t \sigma_\epsilon^2$

RW und dynamische 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. Der dynamische Multiplikator ist Eins.

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

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

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

Der ARMA Prozess

ARMA

Gehorcht ein schwach stationärer Prozess dem Bildungsgesetz

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

so heißt er **ARMA(p, q)**, **autoregressiver moving average Prozess** der Ordnung (p, q) .

Wobei

$$\alpha_p(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p$$

$$\beta_q(L) = 1 - \beta_1 L - \dots - \beta_q L^q$$

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

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

ϵ_t ist white noise.

Beispiele

Sei ϵ_t WN.

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

zB für Monatsdaten mit Saison

Vergleich ARMA und Wold Darstellung

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

Dividiert man die Definitionsgleichung eines ARMA Modells durch das AR-Polynom (sofern das zulässig ist), so erhält man mit $\alpha_p(L)(y_t - \mu) = \beta_q(L)\epsilon_t$

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

Eigentlich werden wir das wahre $\Psi(L)$, das wir iA nicht kennen, nur durch die rationale Funktion approximieren.

Vergleich ARMA und Wold Darstellung (Fs)

Beispiele:

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

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

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

$$y_t - \mu = \beta_{\infty}(L)\epsilon_t = \Psi(L)\epsilon_t$$

Bemerkung: $\sum_{i=0}^{\infty} q^i = 1/(1 - q)$

Vergleich ARMA und Wold Darstellung (Fs)

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, dass eigentlich ein $\text{MA}(\infty)$ zur Modellierung notwendig ist.

Kürzende Wurzeln

Zu beachten ist, dass es zu **keiner Kürzung** der Faktoren im Zähler- und im Nennerpolynom kommt.

Zerlegung von $\alpha_p(L)$ und $\beta_q(L)$ in Linearfaktoren:

$$\alpha_p(L) = (1 - z_1^\alpha L) \cdot \dots \cdot (1 - z_p^\alpha L)$$

$$\beta_q(L) = (1 - z_1^\beta L) \cdot \dots \cdot (1 - z_q^\beta L)$$

$$y_t - \mu = \frac{\beta_q(L)}{\alpha_p(L)} \epsilon_t = \frac{(1 - z_1^\beta L) \cdot \dots \cdot (1 - z_q^\beta L)}{(1 - z_1^\alpha L) \cdot \dots \cdot (1 - z_p^\alpha L)} \epsilon_t$$

Kürzende Wurzeln liegen vor, wenn für ein i bzw j gilt:

$$z_i^\alpha = z_j^\beta$$

Kürzende Wurzeln - Beispiel

Beispiel: ARMA(1,1) mit $\alpha_1 = \beta_1$ ist ein ARMA(0,0).

$$y_t - \mu = \frac{1 - \beta_1 L}{1 - \alpha_1 L} \epsilon_t = \epsilon_t$$

Die Parameter α_1 und β_1 sind nicht identifiziert.

Modellbildung, principle of parsimony

Modellbildung

Das zentrale Konzept der Modellbildung ist, den zugrundeliegenden Prozess durch ein **sparsam parametrisiertes** ARMA-Modell (ARMA-Modell mit niedriger Ordnung) zu approximieren: **principle of parsimony**.

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

(Parsimony ... Sparsamkeit)

Der AR(1) Prozess, $|\alpha| < 1$

Das Modell mit $\mu \neq 0$, die Konstante c

Das Modell für einen AR(1) lautet:

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

oder

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

Berechnung:

$$\begin{aligned} (1 - \alpha L)(y_t - \mu) &= (1 - \alpha L)y_t - (1 - \alpha L)\mu \\ &= (1 - \alpha L)y_t - (1 - \alpha)\mu = (1 - \alpha L)y_t - c \end{aligned}$$

da

$$L\mu = \mu \text{ ist.}$$

Das Modell für $\mu = 0$

Das Modell für einen AR(1) lautet unter der Annahme
 $\mu = E(y_t) = 0$:

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

bzw

$$y_t - \alpha y_{t-1} = \epsilon_t \quad \text{mit } |\alpha| < 1$$

Erwartung und Varianz

Bedingte und unbedingte Erwartung und Varianz

Lassen wir den Prozess $\{y_t\}$ wiederum im Zeitpunkt 0 starten:

$$y_0$$

$$y_1 = \alpha y_0 + \epsilon_1 = \alpha y_0 + \epsilon_1$$

$$y_2 = \alpha y_1 + \epsilon_2 = \alpha(\alpha y_0 + \epsilon_1) + \epsilon_2 = \alpha^2 y_0 + \alpha \epsilon_1 + \epsilon_2$$

...

$$y_t = \alpha y_{t-1} + \epsilon_t = \alpha^t y_0 + \epsilon_t + \alpha \epsilon_{t-1} + \alpha^2 \epsilon_{t-2} + \dots + \alpha^{t-1} \epsilon_1$$

$$y_t = \alpha^t y_0 + \sum_{j=1}^t \alpha^{t-j} \epsilon_j$$

Explizite Darstellung

Erwartung und Varianz

Somit ist die **bedingte Erwartung** und die **bedingte Varianz**

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

Die Beziehung für den Erwartungswert sieht man sofort, da $E(\epsilon_j) = 0$ ist.

Die Varianz folgt aus:

$$\begin{aligned} V(y_t|y_0) &= V(\sum_{j=1}^t \alpha^{t-j} \epsilon_j) = E(\sum_{j=1}^t \alpha^{t-j} \epsilon_j)^2 = \\ &= \sum_{j=1}^t \alpha^{2(t-j)} V(\epsilon_j) \\ &= \sigma_\epsilon^2 \sum_{j=1}^t \alpha^{2(t-j)} \end{aligned}$$

Die Fehler haben konstante Varianz und sind unkorreliert.

Erwartung und Varianz

Die bedingte Erwartung verliert die Abhängigkeit vom Anfangswert, wenn wir den Prozess im Zeitpunkt $-\infty$ starten lassen:

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

Die **unbedingte Erwartung** ist konstant und gleich Null laut obiger Annahme.

$$E(y_t) = 0$$

Die bedingte Varianz verliert mit $|\alpha| < 1$ ebenfalls die Abhängigkeit von t , wenn der Prozess in $-\infty$ startet:

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

Stationarität und Lsg der Differenzengl

- Der AR(1) Prozess ist für

$$|\alpha| < 1$$

eine stationärer Prozess.

- Der Prozess ist stationär, wenn die Differenzgleichung

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

eine stabile Lösung besitzt.

Dh, die Wurzeln $b = \alpha$ des charakteristischen Polynoms

$$b - \alpha = 0$$

liegen innerhalb des Einheitskreises, ($b = \alpha$), $|\alpha| < 1$.

Stationarität und Lsg der Differenzengl

- In der Zeitreihenanalyse lautet das charakteristische Polynom

$$1 - \alpha z = 0$$

mit der Lösung $z = 1/\alpha$.

Wir sagen daher, die Wurzeln müssen **ausserhalb des Einheitskreises** liegen. Die **inversen Wurzeln**, $1/z$, liegen

daher innerhalb des Einheitskreises. (Vgl den EViews Oputput.)

Die ACF, ρ_1

Die **Autokorrelationsfunktion, ACF**, ist die Folge $\{\rho_0, \rho_1, \rho_2, \dots\}$.

ρ_s ist definiert als

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

$\gamma_0 = V(y_t)$ kennen wir schon: $\gamma_0 = \frac{\sigma_\epsilon^2}{1 - \alpha^2}$

Nehmen wir wiederum $E(y_t) = \mu = 0$ an. Dann gilt:

ρ_1 :

$$\begin{aligned} Cov(y_t, y_{t-1}) &= E[(\alpha y_{t-1} + \epsilon_t) y_{t-1}] = \alpha E(y_{t-1}^2) + E(\epsilon_t y_{t-1}) \\ &= \alpha V(y_t) \end{aligned}$$

$$\rho_1 = \alpha$$

Die ACF, ρ_2, \dots

ρ_2 :

$$\begin{aligned} \text{Cov}(y_t, y_{t-2}) &= E[(\alpha y_{t-1} + \epsilon_t) y_{t-2}] \\ &= E[(\alpha (\alpha y_{t-2} + \epsilon_{t-1}) + \epsilon_t) y_{t-2}] \\ &= E[(\alpha^2 y_{t-2} + \alpha \epsilon_{t-1} + \epsilon_t) y_{t-2}] \\ &= \alpha^2 E(y_{t-2}^2) + \alpha E(\epsilon_{t-1} y_{t-2}) + E(\epsilon_t y_{t-2}) \\ &= \alpha^2 V(y_t) \end{aligned}$$

$$\rho_2 = \alpha^2$$

...

ρ_s :

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

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

Alternative Herleitung: $\gamma(s)$

$$y_t - \alpha y_{t-1} = \epsilon_t : \text{oBdA } E(y_t) = \mu = 0$$

$\gamma(1)$: Wir multiplizieren mit y_{t-1} , und nehmen die Erwartung.

$$\begin{aligned} E(y_{t-1} y_t) - \alpha E(y_{t-1} y_{t-1}) &= E(y_{t-1} \epsilon_t) \rightarrow \\ \gamma(1) - \alpha \gamma(0) &= 0 \end{aligned}$$

$\gamma(2)$: Wir multiplizieren mit y_{t-2} , und nehmen die Erwartung.

$$\begin{aligned} E(y_{t-2} y_t) - \alpha E(y_{t-2} y_{t-1}) &= E(y_{t-2} \epsilon_t) \rightarrow \\ \gamma(2) - \alpha \gamma(1) &= 0 \end{aligned}$$

...

Allgemein, mit $\gamma(0)$ als Anfangswert,

$$\gamma(s) - \alpha \gamma(s-1) = 0$$

Die LDGen in y_t und $\gamma(s)$ sind gleich.

Die ACF (Fs)

Die Autokorrelationsfunktion fällt geometrisch (exponentiell).

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

Ist das Mittel nicht wie oben angenommen Null, so lautet das Modell entweder

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

oder

$$(1 - \alpha L)y_t = c + \epsilon_t \quad \text{mit } |\alpha| < 1 \quad \text{wobei } E(y_t) = \mu = \frac{c}{1 - \alpha}$$

Die Varianz und die ACF sind von μ unabhängig.

Die ACF (Fs)

Ein AR(1) Prozess beschreibt ein **Vergessen vergangener Schocks**. Ein Schock, der s Perioden zurückliegt, wird mit $\psi_s = \alpha^s$ gewichtet. Das wird auch in der ACF sichtbar. Die Autokorrelation fällt mit α^s .

Mit $\alpha \rightarrow 1$ wird der Abfall der ACF immer geringer. Die theoretische ACF eines RW mit Startwert y_0 ist konstant 1.

$$\gamma_{RW}(t, t+s) = V(y_t|y_0) = t \sigma_\epsilon^2$$

Die **Stichproben-ACF eines Random walks fällt - lt Konstruktion - ungefähr linear ab**.

Prognose eines AR(1), $\mu \neq 0$

Die τ -Schritt Prognose ist die bedingte Erwartung

$$E(y_{t+\tau}|y_t)$$

Die Varianz der τ -Schritt Prognose, des Prognosefehlers, ist

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

AR(1): $y_{t+1} = c + \alpha y_t + \epsilon_{t+1}$ mit $c = \mu(1 - \alpha)$ bzw

$$y_{t+\tau} = c \sum_{j=0}^{\tau-1} \alpha^j + \alpha^\tau y_t + \sum_{j=1}^{\tau} \alpha^{\tau-j} \epsilon_{t+j}$$

Prognose eines AR(1), $\mu \neq 0$

$\tau = 1$: Die **1-Schritt Prognose** und zugehörige **Prognosevarianz**

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

$\tau = 2$: Die **2-Schritt Prognose** und zugehörige **Prognosevarianz**

$$\begin{aligned} E(y_{t+2}|y_t) &= E[E(y_{t+2}|y_{t+1})|y_t] = E[c + \alpha y_{t+1}|y_t] \\ &= c + \alpha E[y_{t+1}|y_t] = c + \alpha [c + \alpha y_t] \\ &= c(1 + \alpha) + \alpha^2 y_t \end{aligned}$$

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

Allgemein

$$E(y_{t+\tau}|y_t) = c \sum_{j=0}^{\tau-1} \alpha^j + \alpha^\tau y_t \quad V(y_{t+\tau}|y_t) = \sigma_\epsilon^2 \sum_{j=1}^{\tau} \alpha^{2(\tau-j)}$$

Prognose eines AR(1), $\mu \neq 0$

- 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.

Der MA(1) Prozess

Das Modell

Das Modell für einen MA(1) lautet

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

bzw

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

Erwartung und Varianz

Bedingte und unbedingte Erwartung und Varianz

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

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

$$E(y_t|y_{t-1}) = \mu - \beta \epsilon_{t-1}$$

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

Die bedingten Erwartungen $E(y_t|y_{t-s})$ sind gleich der **unbedingten Erwartung** für $s > 1$:

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

Der Prozess hat ein Gedächtnis von genau einer Periode.

Erwartung und Varianz

Die **bedingte und unbedingte Varianz** sind

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

$V(y_t|y_{t-2}) = V(y_t) = \sigma_\epsilon^2(1 + \beta^2)$, da die Kovarianzen der ϵ_t Null sind.

$$V(y_t|y_{t-3}) = V(y_t) = \sigma_\epsilon^2(1 + \beta^2), \text{ etc.}$$

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

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

Die ACF

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

ρ_1 :

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

ρ_2 :

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

...

ρ_s :

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

Die ACF

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

$$\text{Corr}(y_t, y_{t-s}) = 0 \quad \text{für } s \geq 2$$

- 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

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

und

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

Prognose eines MA(1)

Die τ -Schritt Prognose ist die bedingte Erwartung $E(y_{t+\tau}|I_t)$.

I_t ist die gesamte Information bis t . $I_t = \{y_t, \epsilon_t, y_{t-1}, \epsilon_{t-1}, \dots\}$.

Die Varianz der τ -Schritt Prognose, des Prognosefehlers, ist

$$V(y_{t+\tau}|I_t) = E\{[y_{t+\tau} - E(y_{t+\tau}|I_t)]^2|I_t\}.$$

$$\text{MA}(1): \quad y_{t+1} = \mu + \epsilon_{t+1} - \beta \epsilon_t \quad y_{t+2} = \mu + \epsilon_{t+2} - \beta \epsilon_{t+1}$$

$\tau = 1$: **1-Schritt Prognose** und zugeh **Prognosevarianz** sind

$$E(y_{t+1}|I_t) = \mu - \beta \epsilon_t$$

$$V(y_{t+1}|I_t) = E\{[y_{t+1} - E(y_{t+1}|I_t)]^2|I_t\} = \sigma_\epsilon^2$$

τ : **τ -Schritt Prognose**, $\tau > 1$ und zugeh **Prognosevarianz** sind

$$E(y_{t+\tau}|I_t) = \mu$$

$$V(y_{t+\tau}|I_t) = E\{[y_{t+\tau} - E(y_{t+\tau}|I_t)]^2|I_t\} = (1 + \beta^2) \sigma_\epsilon^2$$

Prognose eines MA(q)

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

Die ACF eines ARMA Prozesses

Stationarität und Invertierbarkeit

- Für **Stationarität** verlangen wir für ARMA(p, q) Prozesse, dass die Wurzeln (Nullstellen) des AR-Polynoms ausserhalb des Einheitskreises liegen.
- Für **Invertierbarkeit** verlangen wir für ARMA(p, q) Prozesse, dass die Wurzeln (Nullstellen) des MA-Polynoms ausserhalb des Einheitskreises liegen.

Der ARMA Prozess ist dann **stationär** und **invertierbar**.

Wir zerlegen zB ein AR-Lag-Polynom in seine Linearfaktoren

$$\alpha_p(L) = (1 - z_1 L) \cdot \dots \cdot (1 - z_q L),$$

lösen $1 - z_i z = 0$, erhalten $z = 1/z_i$ und verlangen $1/|z_i| > 1$, was gleichbedeutend mit $|z_i| < 1$ ist.

Die ACF eines ARMA Prozesses

ACF eines ARMA(p, q)

Die Autokorrelationsfunktion eines ARMA(p, q)-Prozesses

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.

Schätzen

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, da der ML-Schätzer nur numerisch aufwendig ist, und die Berücksichtigung der Autokorrelationsstruktur an Effizienz kaum einen Gewinn bringt.

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.

Test für das Mittel, t -Test

Verteilung des Stichprobenmittels:

Das Mittel ist für große T ungefähr normalverteilt mit einem Erwartungswert μ und einer Varianz (σ_y^2/T), sofern die zugrundeliegende Verteilung eine Varianz besitzt und die Autokorrelationen in der Reihe nicht zu stark sind.

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

t-Test: Die Verteilung des Stichprobenmittels kann zum Testen auf das Mittel verwendet werden. ZB:

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

Die 95%-, 97.5%- und 99%-Quantile sind 1.645, 1.960, 2.325.

Test auf die Differenz zweier Mittel

Beide Stichproben sind voneinander unabhängig.

Die Beobachtungen sind nicht oder nur schwach autokorreliert.

Die Stichprobenumfänge T_A und T_B sind so groß, dass der zentrale Grenzwertsatz wirksam wird.

Die Varianzen der beiden Grundgesamtheiten können verschieden sein.

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

Die Zufallsvariable

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

ist (ungefähr) standardnormalverteilt.

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** ist:

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

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

- Die Normalverteilung besitzt eine Kurtosis von 3.
- Ist die Kurtosis einer Verteilung größer als 3, so besitzt sie breitere Ränder als die N-Vtlg: **Leptokurtische Vtlg**

Beispiel:

Die t -Verteilung ist symmetrisch. Die Kurtosis hängt von ihren Freiheitsgraden, n , ab.

Tabelle: t_n -Vtlg, $E(t_n)$, $V(t_n)$, Kurtosis

Freiheitsgrade	μ	σ^2	S	K
1	∞	∞	∞	∞
2	0	∞	∞	∞
3	0	3	∞	∞
4	0	2	0	∞
5	0	5/3	0	9
...
n	0	$n/(n-2)$	0	$3(n-2)/(n-4)$
...
∞	0	1	0	3

Die t -Verteilung mit einem Freiheitsgrad ist die Cauchy-Verteilung.

Mit n gegen unendlich konvergiert t_n gegen die $N(0, 1)$.

Test auf Normalverteilung

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

- mit Erwartung 0 und Varianz $6/T$,

$$\hat{S} \sim N(0, 6/T)$$

- mit Erwartung 3 und Varianz $24/T$.

$$\hat{K} \sim N(3, 24/T)$$

Der folgende **Jarque-Bera Test** testet simultan auf

$$\text{Schiefe} = 0 \quad \text{und} \quad \text{Kurtosis} = 3$$

einer Verteilung.

Test auf Normalvtlg: Jarque-Bera Test

Unter der Nullhypothese, dass die Zufallsvariable Y normal verteilt ist, ist die Statistik

$$\frac{n-k}{6} \left[\hat{S}^2 + \frac{1}{4} (\hat{K} - 3)^2 \right] \sim \chi_2^2$$

χ^2 verteilt mit 2 Freiheitsgraden.

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

Schätzer für die ACF, SACF

Der Schätzer für die ACF ist die **Stichprobenautokorrelationsfunktion**, auch als Korrelogramm bezeichnet.

Der **Stichprobenautokorrelationskoeffizient k -ter Ordnung** ist definiert als

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

Fasst man die Autokorrelationen der Ordnungen $0, 1, 2, 3, \dots$ zusammen, erhält man das Korrelogramm mit $r_0 = 1$:

$$1, r_1, r_2, r_3, \dots$$

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 $s > m$

Erwartungswert

$$E(\hat{\rho}_s) = -\frac{1}{T}$$

und Varianz

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

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

Für einen white noise $y_t = \varepsilon_t$ gilt:

$$m = 0, \quad \rho_s = 0 \quad \text{für} \quad s > m = 0$$

und damit

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

für $s > 0$.

Die $\hat{\rho}_k$ sind normalverteilt

$$\hat{\rho}_k \sim N\left(-\frac{1}{T}, \frac{1}{T}\right)$$

Test auf white noise: Test für $\hat{\rho}_k$

Test auf white noise:

Unter den Hypothesen

$$H_0 : \rho_k = 0 \quad H_1 : \rho_k \neq 0 \quad k > 0$$

ist $\hat{\rho}_k$ normalverteilt

$$\hat{\rho}_k \sim N\left(0, \frac{1}{T}\right) \quad \text{bzw} \quad \hat{\rho}_k \sim N\left(-\frac{1}{T}, \frac{1}{T}\right)$$

Test auf white noise: Omnibustest

Test auf eine Gruppe von Autokorrelationskoeffizienten:

Die **Box-Pierce-Statistik**, Q , testet, ob in einer Gruppe von Autokorrelationskoeffizienten eines stationären Prozesses alle Null sind.

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_s = 0 \quad H_1 : \rho_k \neq 0 \quad 1 \leq k \leq s$$

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

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

Besser geeignet für kleine T ist die **Ljung-Box-Statistik** \tilde{Q} .

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

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, ua

- 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

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

Unsere Beobachtungen beginnen bei Nummer 1 und enden bei Nummer T . Das Modell für die 1. Beobachtung lautet

$$y_1 = c + \alpha_1 y_0 + \dots + \alpha_p y_{-p+1} + \epsilon_1 - \beta_1 \epsilon_0 - \dots - \beta_q \epsilon_{-q+1}$$

y_0, \dots, y_{-p+1} und $\epsilon_0, \dots, \epsilon_{-q+1}$ müssten aber bekannt sein, um c über die Minimierung des Fehlers ϵ_1 berechnen zu können.

Diese Werte stehen aber nicht zur Verfügung.

ULS (Fs)

Setzt man die nicht beobachteten y_{-j} gleich dem Mittel der Reihe, \bar{y} , und die nicht beobachteten ϵ_{-k} gleich ihrem Erwartungswert, Null,

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

wird die Anwendung der LS-Methode möglich. Zu den T Beobachtungen gibt es nun T Störungen.

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

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

ULS: Eigenschaften, Tests

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

Sie sind

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

Maximum Likelihood Methode

Wir gehen von dem Vektor der ZVen $\mathbf{y} = (y_1, \dots, y_T)'$ aus, und nehmen an, dass sie gemeinsam normalverteilt sind mit

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

$\boldsymbol{\mu}$ ist der Vektor der (konstanten) Mittel für die einzelnen Zeitpunkte, $t = 1, \dots, T$, $\boldsymbol{\mu} = (\mu, \dots, \mu)'$ und

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

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

Likelihood für AR(1)

Die Kovarianzen eines AR(1) Prozesses sind

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

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

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

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

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

Likelihood für MA(1)

Die Kovarianzen eines MA(1) Prozesses sind

$$\text{Cov}(y_t, y_t) = \gamma_0 = (1 + \beta^2) \sigma_\epsilon^2, \quad \text{Cov}(y_t, y_{t+1}) = \gamma_1 = -\beta \sigma_\epsilon^2,$$

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

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

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

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

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

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 (nun reduced) Log-Likelihood numerisch maximiert, **minus Log-Likelihood minimiert**.

Numerische Minimierungsverfahren sind ua

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

Min $\{-\log(L)\}$: Numerische Verfahren

- **Grid-Search:** Dieses einfachste Verfahren dient nur zur Illustration. Man gibt ein grobes Raster (Gitter) zB 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.
- **Simplex-Verfahren:** Dieses ist nicht mit der Linearen Optimierung zu verwechseln. Es umgeht die Berechnung der Ableitungen in der Methode des steilsten Abstiegs.
- **Methode des steilsten Abstiegs:** Der Vektor der partiellen Ableitungen (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.

ML: Eigenschaften, Tests

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

Sie sind

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

Modellüberprüfung

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 zB $\hat{\alpha}_p$ nicht signifikant von Null verschieden, so sollte die AR-Ordnung um 1 reduziert oder erhöht werden. "Mittlere" Parameter "dürfen" Null sein.

Residuen, in-sample Prognose

- (2) Residuen des geschätzten Modells 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 zu klein.

- (3) Modellerte Reihe erfasst weitgehend die wichtigen Charakteristika.

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

Prognose

- (4) Modell liefert vernünftige Prognosen.

Dazu führt man eine Prognose, sowohl in-sample, als auch out-of-sample durch, und misst die Prognosegüte mit zB dem RMSE. In der Praxis stellt sich oft ein 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, Informationskriterien

Problem

Die obigen Kriterien sind nicht immer einfach zu handhaben. ZB wie unterscheidet man zwischen einem ARMA(1,2) und einem ARMA(2,1), an Hand der t -Tests, wenn für beide Modelle die Parameter vor den größten Lags signifikant von Null verschieden sind?

Informationskrit und Prinzip d Sparsamkeit

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

- Akaike's Informationskriterium, AIC, und das
- Schwarz' Kriterium, auch Schwarz'-Bayes'sche Informationskriterium, SIC (SBC, SC) 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.

Da die vermeintliche Anpassung des Modells an die Daten immer besser wird, je mehr Parameter im Modell aufgenommen werden, schlagen die Informationskriterien eine Bestrafung, ein **Pönale**, für jede eingeführte Variable vor.

Informationskriterien: AIC, SIC

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

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

Schwarz'-Informationskriterium, bzw Schwarz'-Bayes'sche Informationskriterium, SIC (SBC), ist gegeben durch

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

$\hat{\sigma}_{\epsilon, p+q}^2$ steht für die minimierte Fehlervarianz.

$\log(\cdot)$ steht stets für den natürlichen Logarithmus.

Informationskriterien (Fs)

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 iA die Fehlerquadratsumme zumindest ein wenig verringern.

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

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

Informationskriterien (Fs)

Es werden somit solange Variable (Lags) aufgenommen, bis die letzte Variable (Lag) 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 (und die PACF) 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 bzw SIC berechnet.
- Es wird nun *jenes Modell gewählt*, welches das **kleinste AIC** bzw **kleinste SIC** aufweist.

Modellsuche (Fs)

Bemerkung:

Manchmal wird das AIC (SIC) mit einem negativen Vorzeichen definiert. Dann ist umgekehrt das Modell mit dem größten AIC (SIC) zu suchen. (Das trifft auf den EViews Output aber nicht zu!)

AIC, SIC in Log-Likelihood

Eine andere Darstellung für die Informationskriterien erhält man, wenn die Näherung

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

verwendet wird.

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

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

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

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

AIC, SIC im EViews

Das AIC und SIC in EViews werden wie folgt berechnet und sind zu minimieren:

$$\text{AIC}(p, q) = -2l(\hat{\theta}_{p+q})/T + 2(p + q)/T$$

$$\text{SIC}(p, q) = -2l(\hat{\theta}_{p+q})/T + \log(T)(p + q)/T$$

Eine Version, die in der Literatur oft verwendet wird:

Multipliziert man mit T , ergibt sich

$$\text{AIC}(p, q) = -2l(\hat{\theta}_{p+q}) + 2(p + q)$$

$$\text{SIC}(p, q) = -2l(\hat{\theta}_{p+q}) + \log(T)(p + q)$$

Modellierung von nicht-stationären Prozessen

Trendstationarität

Wir diskutieren kurz sogenannte trendstationäre und differenzenstationäre Prozesse.

- Kann ein Prozess durch Berücksichtigung eines linearen Trends in einen stationären transformiert werden, so heißt er **trendstationär**.

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 **regressiert** x_t auf die Zeit t , und modelliert den stationären Teil u_t zB als AR(3).

Differenzenstationarität

- Kann eine Prozess durch Differenzenbildung in einen stationären transformiert werden, so heißt er **differenzenstationär**.

Hier wird die Veränderung zwischen den Beobachtungen, $y_t = (1 - L)x_t$, berechnet und als neue Variable, y_t , modelliert.

x_t differenzenstationär

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

wobei v_t ein stationärer Prozess ist, wird stationär modelliert als

$$y_t = c + v_t$$

mit $y_t = (1 - L)x_t = x_t - x_{t-1}$ bzw $x_t = x_{t-1} + y_t$.

Differenzenstationarität - Beispiele

Beispiele:

- Ein random walk mit $\{\epsilon_t\}$ WN ist differenzstationär.

$$x_t = x_{t-1} + \epsilon_t$$

- Sei v_t ein ARMA(p, q).

$$y_t = c + v_t$$

Dann ist x_t differenzenstationär:

$$x_t = x_{t-1} + y_t$$

Differenzenstationär: ARIMA(p, d, q)

Man bezeichnet differenzenstationäre Proz 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 Stichprobenautokorrelationsfunktion 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. $d = 0$.
Er ist stationär.
- Ein Random Walk, y_t , ist ein ARIMA(0,1,0), $d = 1$.
Er ist differenzenstationär. $y_t - y_{t-1} = (1 - L)y_t = \Delta y_t$ ist stationär und white noise.
- Ist x_t ein ARIMA($p, 2, q$), dann ist die 2-malige Differenz

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

stationär.

Saison

Saisonaler ARMA (stationär)

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

Beispiele sind:

- 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 (stat)

Man unterscheidet

- **additive saisonale ARMA**, zB,

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

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

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

Saisonaler ARIMA (differenzen-stationär)

Sei hier $S = 4$. Statt der ersten Differenz, $(1 - L)$, wird die saisonale 4-te Differenz, $(1 - L^4)$, gebildet um eine stationäre Reihe zu erhalten.

Die Idee ist, dass die Veränderung vom 1.Quartal des Vorjahres zum heutigen 1.Quartal die Jahresveränderung gut wiedergibt. Analog für 2.Quartal des Vorjahres zum jetzigen 2-ten Quartal.

Die Veränderung $(1 - L)$ beschreibt diese Struktur nicht geeignet, da sie auf die Veränderung von einem Quartal zum Folgenden zielt: Veränderung vom 3.Quartal zum 4-ten, vom 4-ten zum ersten des Folgejahres, etc.

Saisonaler ARIMA (differenzen-stationär)

Das multiplikative saisonale ARIMA schreibt man als

$$\text{ARIMA}(p, d, q)(P, D, Q)_S$$

Beispiel: $\text{ARIMA}(0, 1, 1)(0, 1, 1)_{12}$

$$(1 - L)(1 - L^{12}) y_t = (1 - \beta_1 L)(1 - B_1 L^{12}) \epsilon_t$$