Zeitreihenmodellierung der ARMA-Klasse

Ziel: endogene Modellierung einer Zeitreihe (***Synthese***)

Motivation: künstliche Verlängerung gegebener Zeitreihen unter Wahrung wichtiger statistischer Eigenschaften zum Zwecke der

- operativen Vorhersage (*one-step-ahead-Prediction*)
- Anwendung statistischer Verfahren, die lange Datensätze voraussetzen (*Bsp. Extremwert-Statistik, wird mittlerweile aber kritisch gesehen*)

Ansatz: Nutzung der Autokorrelationsfunktion

Extrapolation einer Zeitreihe

Gegeben: Zeitreihe mit n (äquidistanten) Beobachtungen

Gesucht: bestmögliche Prognose der nächsten m Werte („*m steps-ahead-prediction*")

Ansatz:

- Liegen keine weiteren Informationen über statistische Eigenschaften der Zeitreihe vor, ist die beste Schätzung der Mittelwert
- Zeitreihen weisen aber meist signifikante Autokorrelationen auf => lässt sich für eine kurzfristige Vorhersage nutzen

AR(p)-Modell

= ***A****uto-****R****egressive-Modell der Ordnung* ***p***

- **Prinzip**: Ausnutzung der Autokorrelation über p Zeitschritte zur Verlängerung einer gegebenen Zeitreihe:

$$x(t_i) = \sum_{j=1}^{p} \alpha_j \cdot x(t_{i-j}) + \eta$$

- der Rauschterm (= Fehler, Residuen) $\eta_i = x(t_i)_{prog} - x(t_i)_{beob}$ sollte dabei minimiert werden

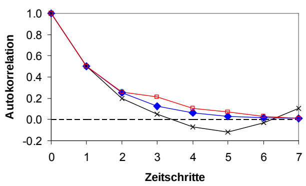
- der Rauschterm sollte möglichst wenig Rest-Informationen enthalten
  - $\overline{\eta} = 0$
  - Keine Autokorrelation
  - z.B. Gaußsches Rauschen, Weißes Rauschen

Partielle Autokorrelation

Autokorrelation für ***1*** Zeitschritt = $\rho_{1\cdot\Delta t}$

=> Autokorrelation für ***2*** Zeitschritte = $\rho_{1\cdot\Delta t} \cdot \rho_{1\cdot\Delta t} = \rho_{1\cdot\Delta t}^2$

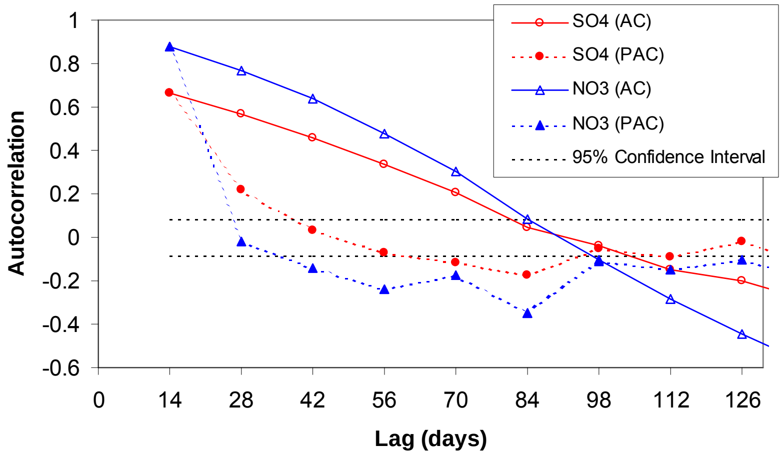
=> Autokorrelation für ***n*** Zeitschritte = $\rho_{1\cdot\Delta t}^n$



- Die beobachtete Autokorrelation setzt sich aus den Beiträgen der einzelnen Zeitdifferenzen Δ***t*** zusammen (***Partielle Autokorrelationen***)
- Die partielle Autokorrelation ist eine bedingte Wahrscheinlichkeit:

$$Corr\left(x(t_i), x(t_{i-j}) \mid x(t_{i-1}), x(t_{i-2}), \ldots, x(t_{i-j+1})\right)$$

Bsp.: Autokorrelation



Bestimmung des *p*-Wertes

Ziel: Minimierung der Abweichungsquadrate (*least square*): $\sum \eta_i^2 \overset{!}{=} \min$

Ansatz: Yule-Walker-Gleichung $C(k) = \sum_{j=1}^{p} \alpha_j \, C(k-j)$

mit den partiellen Autokorrelationen ρ: $\rho_i = \frac{C(i)}{C(0)}$, $\rho_0 = 1$

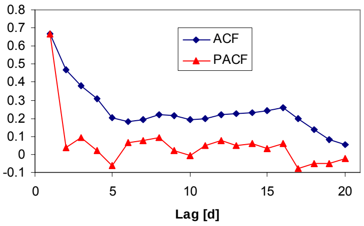
=> Gleichungssystem:

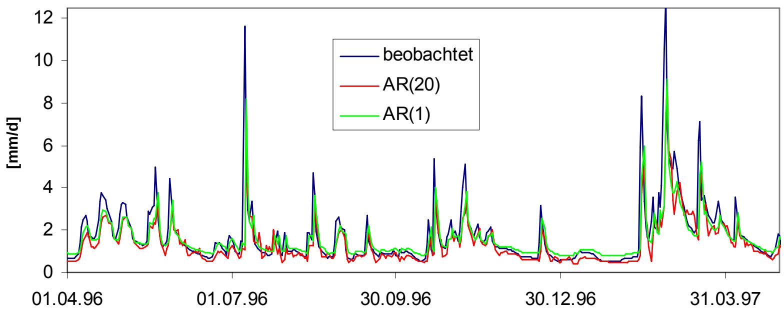
$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \dots & \dots & \dots & & \dots \\ \dots & \dots & \dots & & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & \dots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ \rho_p \end{pmatrix}$$

AR(1)-Modell: Thomas-Fiering-Modell

- $x(t_i) = \alpha_1 \cdot x(t_{i-j}) + \eta$
- erhält 1. und 2. Momente
- geeignet für kurzfristige (operationelle) Vorhersage
- Generierung synthetischer Datensätze (Abflussganglinien)
- weit verbreitet

Beispiel: Lehstenbach-Abfluss





Der Rückwärtsschiebe-Operator

Operator = Rechenvorschrift

Der Rückwärtsschieboperator ***B*** : $Bx(t_k) = x(t_{k-1}) \qquad B^j x(t_k) = x(t_{k-j})$

Bsp. AR-Modell: $x(t_i) = \sum_{j=1}^{p} \alpha_j \cdot x(t_{i-j}) + \eta$ => $x(t_i) - \sum_{j=1}^{p} \alpha_j \cdot x(t_{i-j}) = \eta$

alternativ: $x(t_i) - \sum_{j=1}^{p} \alpha_j \cdot B^j x(t_i) = \eta$

allgemein: $\alpha(B) = 1 - \sum_{j=1}^{p} \alpha_j B^j$

=> alternative Schreibweise für das AR-Modell: $x(t) = \alpha(B)x(t) + \eta(t)$

MA(q)-Modell

= ***M**oving **A**verage*-Modell der Ordnung ***q***

Motivation: Die Residuen eines AR(p)-Modells sind oft ebenfalls autokorreliert

=> transiente Trends: der Mittelwert der ***x(t)*** für kurze Fenster ist nicht konstant

=> Verbesserung des Modells durch gleitende Mittelwertsbildung (*Moving Average*)

Ansatz: Nutzung der Autokorrelation der Residuen:

$$x(t_i) = \eta_0(t) + \sum_{k=1}^{q} \beta_k \, \eta(t_{i-k})$$

MA(q)-Modell: Vorgehensweise

- Bestimmung der Koeffizienten des MA(q)-Modells anhand der Autokorrelationsfunktion der Residuen:

$$\rho_k = \begin{cases} \dfrac{-\beta_k + \beta_1\beta_{k+1} + \ldots + \beta_{q-k}\beta_k}{1+\beta_1^2+\ldots+\beta_q^2} & \text{für} \quad k = 1,2,\ldots,q \\ 0 & \text{für} \quad k > q \end{cases}$$

- i.d.R. numerisches Lösungsverfahren

ARIMA(p,d,q)-Modell

= ***A***uto-***R***egressive ***I***ntegrated ***M***oving ***A***verage-Modell der Ordnung (*p*,*d*,*q*)

Prinzip: Erweiterung des ARMA-Modells: anstelle der $x(t)$-Werte werden die Differenzen ($x(t+i) - x(t)$) (für $d = 1$), bzw. die Differenzen der Differenzen (für $d = 2$), etc., verwendet

Motivation: Differenzen sind oft eher stationär als die ursprünglichen Daten

Der Differenzenoperator

Beobachtung: Differenzen aufeinanderfolgender Werte einer Zeitreihe sind oft wesentlich besser stationär (und stärker verrauscht) als das Original

Beispiel: $x(t_i) = x_0 + at_i + y(t_i)$

mit stationärer Zeitreihe $y(t)$

Die erste Differenz ist stationär:

$$x'(t_i) \equiv x(t_i) - x(t_{i-1}) = (1-B)x(t_i) = a\Delta t + y(t_i) - y(t_{i-1})$$

Zweite Differenz (= *Differenz der Differenzen*):

$$x''(t_i) = x(t_i) - 2x(t_{i-1}) + x(t_{i-2}) = (1-B)^2\, x(t_i)$$

usw.

Beispiel: ARIMA(0,1,0)

$$x(t) = x(t_{i-1}) + \eta(t)$$

= Random Walk Modell

Anwendungen:

- Brownsche Molekularbewegung
- Börsenkurse
- Kristallwachstum
- Diffusion
- ...

Spezialfälle von ARIMA-Modellen

Modell	Bezeichnung	Kommentar
ARIMA(0,1,0)	Random Walk-Model	Beispiele: Brownsche Molekularbewegung, Diffusion, Kristallwachstum
ARIMA(1,0,0)	Thomas-Fiering-Modell	verbreitet im Bereich der Hydrologie
ARIMA(0,1,1)	Exponentielle Glättung	

Der Rückwärtsschiebe-Operator

Der Rückwärtsschieboperator $\boldsymbol{B}$: $Bx(t_k) = x(t_{k-1})$ $\quad B^j x(t_k) = x(t_{k-j})$

- AR-Modell: $\alpha(B) = 1 - \sum_{j=1}^{p} \alpha_j B^j$
- MA-Modell: $\beta(B) = 1 + \sum_{k=1}^{q} \beta_k B^k$
- ARMA-Modell: $\alpha(B)x(t) = \beta(B)\eta(t)$

$$\left(1 - \sum_{j=1}^{p} \alpha_j B^j\right) x(t) = \left(1 + \sum_{k=1}^{q} \beta_k B^k\right) \eta(t)$$

$$x(t) - \sum_{j01}^{p} \alpha_j B^j \, x(t) = \eta(t) + \sum_{k=1}^{q} \beta_k B^k \, \eta(t)$$

- ARIMA-Modell: $\alpha(B)\,(1-B)^d\, x(t) = \beta(B)\ \eta(t)$

Strategie der Extrapolation einer Zeitreihe

Wenn ...	Beste Vorhersage
Weißes Rauschen der $x(t)$	Mittelwert
signifikante Autokorrelationen der $x(t)$	**AR**-Modell
signifikante Autokorrelationen der Residuen	AR**MA**-Modell
signifikante Autokorrelationen der $x(t)$ für viele k => Trend	AR**I**MA-Modell
ausgeprägtes Langzeit-Gedächtnis der $x(t)$	**F**ARIMA-Modell

FARIMA-Modelle

AR: **A**uto-**R**egression → auf der Basis der Autokorrelationsfunktion (lags of the time series)

MA: **M**oving **A**verage → autokorrelierte Residuen (lags of the residues)

I: **I**ntegrated → Anwendung auf Differenzen zwischen aufeinanderfolgenden Werten, da eher stationär

F: **F**ractal → Ansatz zur Brücksichtigung langreichweitiger Korrelationen

Vorgehensweise

1. Normierung der Zeitreihe auf den Mittelwert = 0
2. Enttrenden der Zeitreihe
3. Desaisonalisierung
4. Bestimmung der p, d, q-Werte

Aufgabe

1. Bestimmen Sie die partiellen Autokorrelationen für den Niederschlags-Temperatur-Abfluss-Datensatz.
2. Erstellen Sie das bestmögliche ARIMA(p,d,q)-Modell für den Niederschlags-Temperatur-Abfluss-Datensatz (mit $0 \leq p \leq 5$ und $0 \leq q \leq 5$, ev. mit Desaisonalisierung). Achten Sie auf die Autokorrelation der Residuen, sowie auf den Standardfehler und das Konfidenzintervall der Prognose. Stellen Sie die Prognosewerte für ausgewählte Zeitabschnitte grafisch dar.
3. Erstellen Sie das bestmögliche ARIMA(p,d,q)-Modell, nachdem Sie die Zeitreihen auf Mittelwert = 0 normalisiert und um den jeweiligen Trend bereinigt haben.

Anmerkung: Ausführliche Tipps für eine Optimierung von ARIMA-Modellen finden sich z.B. unter http://www.duke.edu/~rnau/411arim2.htm