

Universität Potsdam

Institut für Informatik
Lehrstuhl Maschinelles Lernen



Sequenzanalyse

Michael Brückner (Vertretung)

Christoph Sawade/Niels Landwehr

Paul Prasse

Tobias Scheffer

Literatur

- Klaus Neusser: Zeitreihenanalyse in den Wirtschaftswissenschaften.
- Mike Hüftle: Stochastische Prozesse in der Zeitreihenanalyse.
<http://www.ivh.uni-hannover.de/optiv/Methoden/StochMet/StochMet.pdf>

Überblick

- Lernen aus einer Sequenz.
 - ◆ Stochastischer Prozess.
 - ◆ Datentransformation.
 - ◆ Parameter eines stationären stochastischen Prozesses schätzen.
 - ◆ Prognose.

Sequenzanalyse

Problemstellung

- Gegeben: Sequenz $\{x_t \in \mathbb{R} : t = 1 \dots n\}$ von n geordneten Datenpunkten (Messpunkte).
- Gesucht: Stochastisches Modell welches Sequenz gut erklärt.
- Ziele:
 - ◆ Beschreibung: Untersuchen der Charakteristika der Sequenz über Kennzahlen und Diagramme.
 - ◆ Modellierung: Verstehen des zugrunde liegenden, Daten-generierenden Modells.
 - ◆ Prognose: Vorhersage zukünftiger Werte aufgrund des gelernten Modells.

Sequenzanalyse

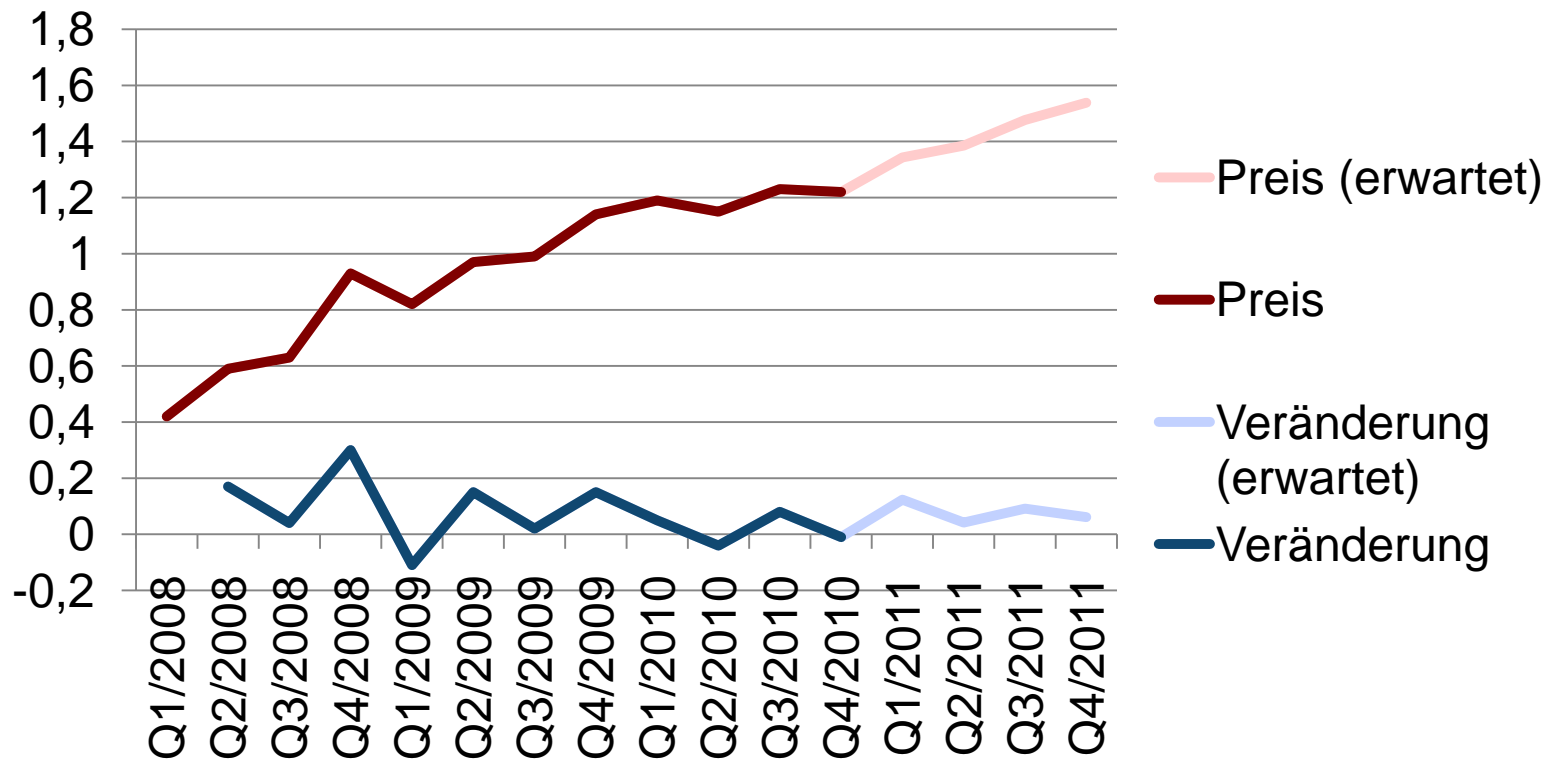
Gegeben Daten

- Sequenzen sind durch Paare (x_i, t_i) gegeben:
 - ◆ $x_i \in \mathbb{R}^m$ gibt den Wert der Sequenz an der Position t_i an.
 - ◆ Äquidistante Folge (gleiche Abstände $t_{i+1} - t_i$) $\{x_t : t = 1 \dots n\}$
 \Rightarrow *diskreter stochastischer Prozess*
 - ◆ Nicht-äquidistante Folge $\{x(t_i) : i = 1 \dots n\}$
 \Rightarrow *stetiger stochastischer Prozess*
- Betrachten im Weiteren nur univariate, äquidistante Zeitreihen $\{x_t \in \mathbb{R} : t = 1 \dots n\}$ wie bspw.
 - ◆ Tägliche Messung der Temperatur an einem Ort,
 - ◆ Entwicklung der Einwohnerzahl eines Landes,
 - ◆ Aktienkurse an aufeinanderfolgenden Börsentagen.

Lernen aus einer Sequenz

Beispiel

- Vorhersage der Preisentwicklung eines Produktes:



Stochastischer Prozess

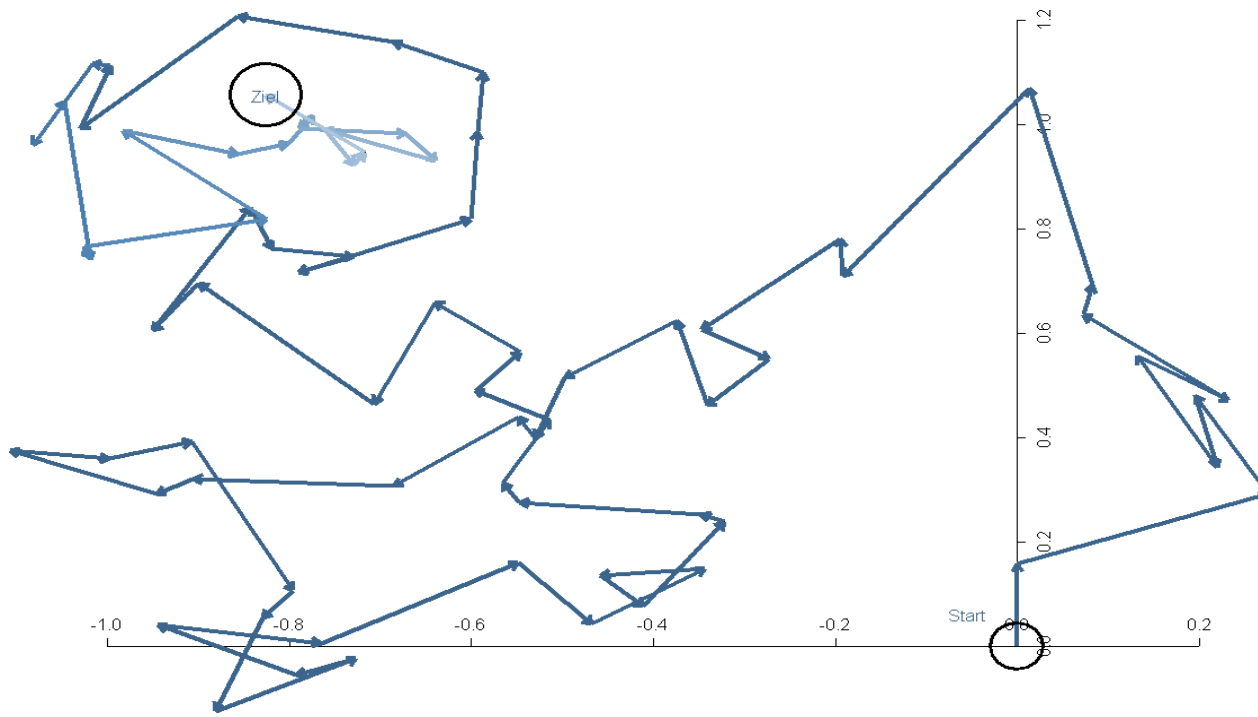
Definition

- Stochastischer Prozess:
 - ◆ Abbildung $X(\omega, t)$ aus $\Omega \times T$ auf Menge der reellen Zahlen die für jedes fixierte $t \in T$ eine Zufallsgröße X_t und für jedes fixierte $\omega \in \Omega$ eine gewöhnliche Funktion $x(t)$ darstellt.
 - ◆ Jede Zufallsgröße X_t nimmt für einen Versuchsausgang einen Wert (Realisierung) x_t an.
- Annahme: Gegebene Sequenz $\{x_t \in \mathbb{R} : t = 1 \dots n\}$ ist Folge von Realisierungen eines stochastischen Prozesses.

Stochastischer Prozess

Beispiel

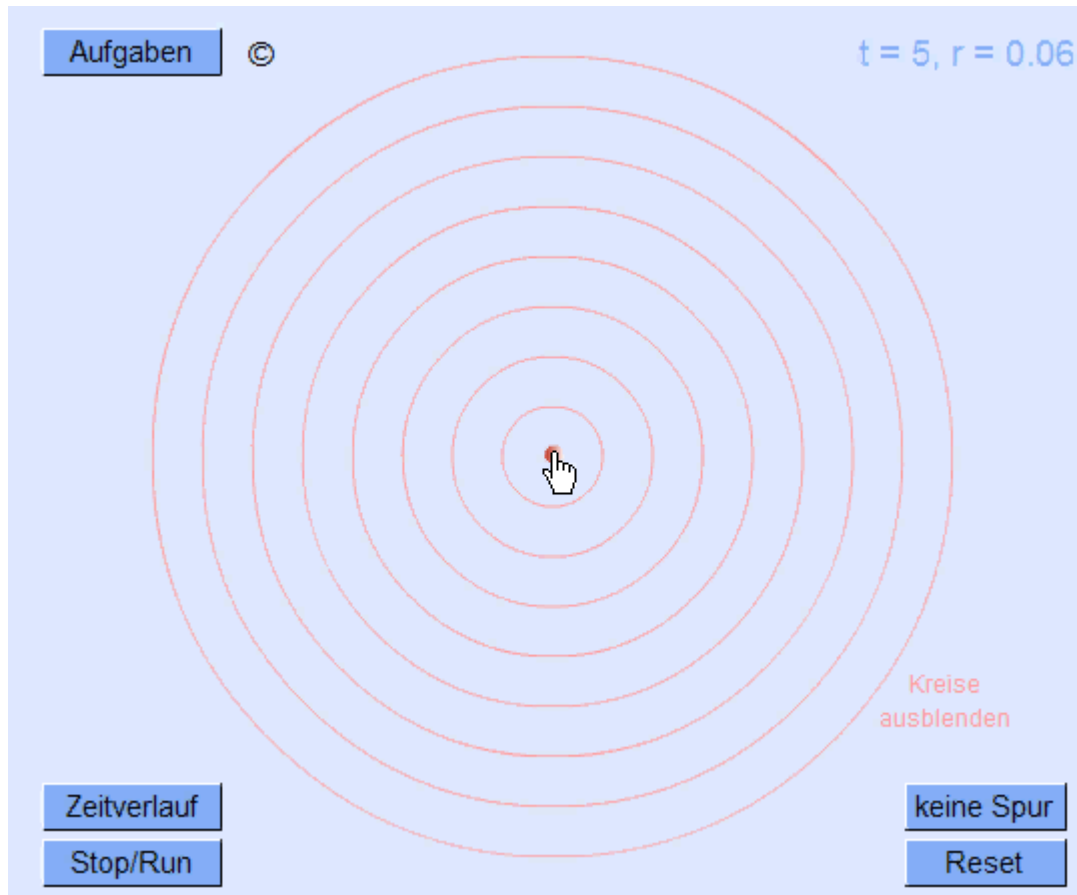
- Brown'sche Bewegung:
 - ◆ Beschreibt zufällige Bewegung von Teilchen.



Stochastischer Prozess

Beispiel

■ Random Walk:



Stochastischer Prozess

Kenngrößen

■ Mittelwertfunktion:

- ◆ Erwartungswert $\mu_X(t) = E[X_t]$ der Zufallsgröße X_t zum Zeitpunkt t .
- ◆ Durchschnittliche Sequenz, um welche die tatsächlichen Beobachtungen des stochastischen Prozesses schwanken.

■ Varianzfunktion:

- ◆ Varianz $\sigma_X^2(t) = E[(X_t - \mu_X(t))^2]$ der Zufallsgröße X_t zum Zeitpunkt t .
- ◆ Mittlere quadratische Abweichung vom erwarteten Verlauf des stochastischen Prozesses.

Stochastischer Prozess

Kenngrößen

■ Autokovarianzfunktion:

- ◆ Autokovarianz $\gamma_X(s, t) = E[(X_s - \mu_X(s))(X_t - \mu_X(t))]$ der Zufallsgrößen X_s und X_t zu den Zeitpunkten $s, t \in T$.

■ Autokorrelationsfunktion:

- ◆ Autokorrelation $\rho_X(s, t)$ der Zufallsgrößen X_s und X_t zu den Zeitpunkten $s, t \in T$ mit

$$\rho_X(s, t) = \frac{\gamma_X(s, t)}{\sigma_X(s)\sigma_X(t)}.$$

- ◆ Gibt an ob die Zufallsgrößen eines stochastischer Prozesses linear korreliert sind, d.h. ob der Prozess zyklische Muster beschreibt.

Stochastischer Prozess

Eigenschaften

Ein stochastischer Prozess heißt

- *stationär* falls für $\forall r, s, t \in T$
 - ◆ Erwartungswert $\mu_X(t) = \mu_X$ konstant,
 - ◆ Varianz $\sigma_X^2(t) < \infty$,
 - ◆ Autokovarianz $\gamma_X(s, t) = \gamma_X(s + r, t + r)$, d.h. die Autokovarianzfunktion hängt nur vom zeitlichen Abstand $h = |s - t|$ ab; somit gilt

$$\gamma_X(s, t) = \gamma_X(t, s) = \gamma_X(h, 0) = \gamma_X(-h, 0).$$

- *diskret* falls die Zeitpunkte $t \in T$ abzählbar sind.
- *stetig* falls T ein Intervall der reellen Zahlen ist.

Stochastischer Prozess

Schätzer der Kenngrößen für stationäre Prozesse

- Schätzer für Mittelwertfunktion:

$$\hat{\mu}_X(t) = \frac{1}{t} \sum_{i=1}^t x_i$$

- Schätzer für Varianzfunktion:

$$\hat{\sigma}_X^2(t) = \frac{1}{t} \sum_{i=1}^t (x_i - \hat{\mu}_X(t))^2$$

- Schätzer für Autokovarianzfunktion:

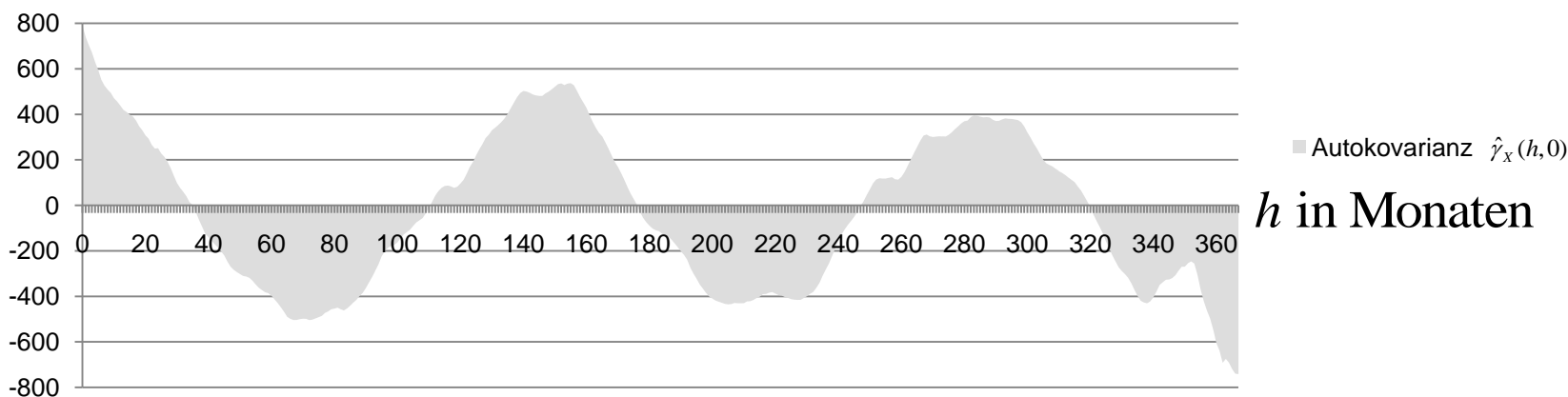
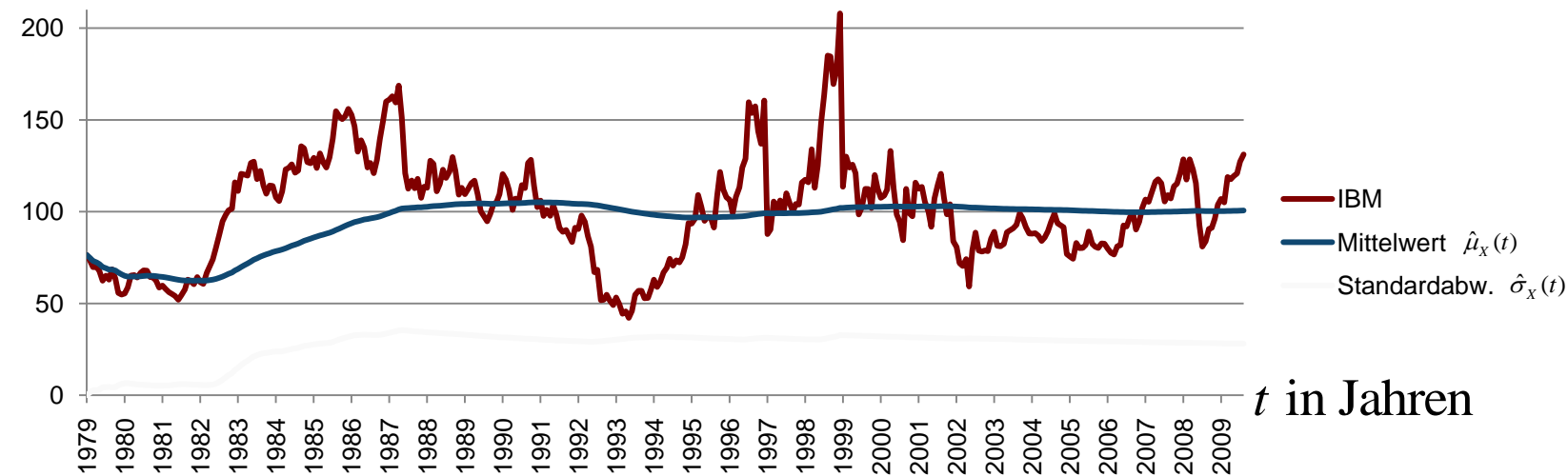
$$\hat{\gamma}_X(h, 0) = \frac{1}{n} \sum_{i=1}^{n-h} (x_i - \hat{\mu}_X(n))(x_{i+h} - \hat{\mu}_X(n))$$

- Schätzer für Autokorrelationsfunktion

$$\hat{\rho}_X(h, 0) = \frac{\hat{\gamma}_X(h, 0)}{\hat{\gamma}_X(0, 0)}$$

Stochastischer Prozess

Beispiel: Schätzer für Kenngrößen



Stochastischer Prozess

Lernen des Prozesses

- Stochastischer Prozess durch Mittelwert-, Varianz- und Autokovarianzfunktion hinreichend genau gegeben.
- Schätzung dieser Funktionen für nicht-stationäre Prozesse benötigt sehr viele Daten!
- Ansatz:
 1. Daten transformieren um Stationarität sicherzustellen.
 2. Stationären Prozess wählen und dessen Parameter schätzen.
 3. Ursprüngliche Datentransformation umkehren und gelerntes Modell anwenden (Prognose).

Datentransformation

Trendbereinigung

- In der Praxis Mittelwert oft nicht konstant über die Zeit.
- Idee: Modellierung des Mittelwerts als Funktion der Zeit, z.B.

◆ Linearer Trend: $\hat{\mu}_X(t) = \beta_0 + \beta_1 t$

◆ Polynomieller Trend: $\hat{\mu}_X(t) = \sum_{i=0}^p \beta_i t^i$

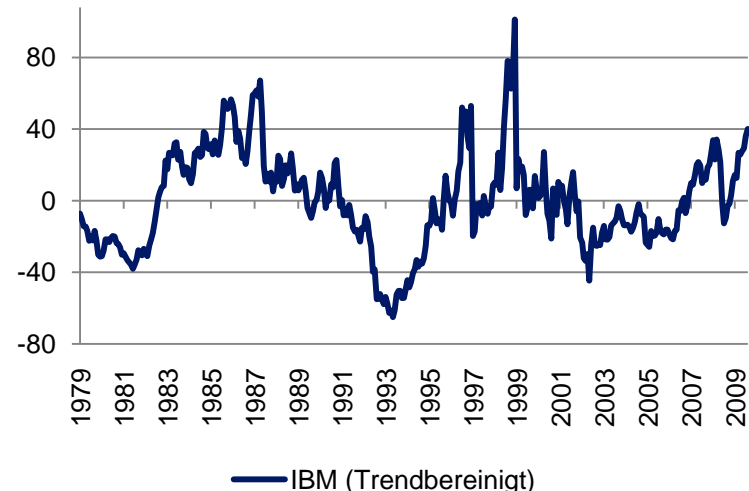
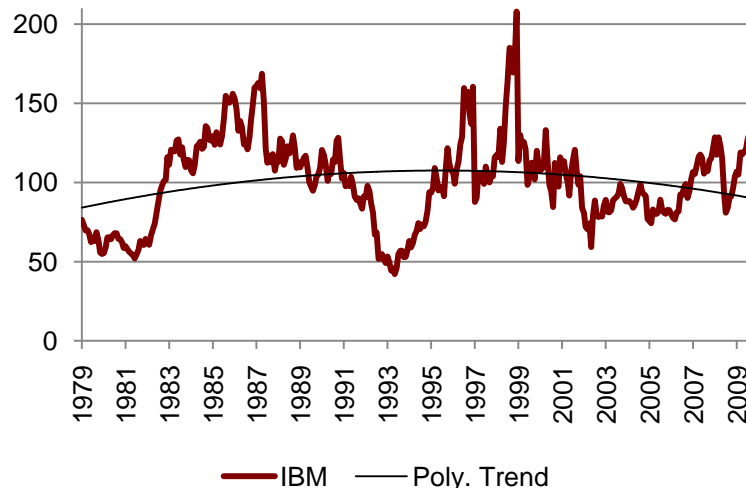
◆ Exponentieller Trend: $\hat{\mu}_X(t) = \sum_{i=0}^p e^{\beta_i t}$

◆ Potenzieller Trend: $\hat{\mu}_X(t) = \sum_{i=0}^p t^{\beta_i}$

Datentransformation

Trendbereinigung

- Schätzung der Parameter β_i durch Minimierung des quadratischen Fehlers $\min_{\beta_i} \sum_{t=1}^n (\hat{\mu}_X(t) - x_t)^2$.
- Bereinigte Sequenz $x'_t = x_t - \hat{\mu}_X(t)$, $t = 1 \dots n$ ist (approximativ) stationär bzgl. des Mittelwertes mit $\hat{\mu}'_X = 0$.



Datentransformation

Filter

■ Entfernung nicht-stationärer Komponenten durch

- ◆ Anwenden des Differenzenfilters:

$$x'_t = \Delta x_t = x_{t+1} - x_t$$

Entfernt lineare Trends

- ◆ d -maliges Anwenden des Differenzenfilters:

$$x'_t = \Delta^d x_t = \Delta^{d-1} x_{t+1} - \Delta^{d-1} x_t$$

Entfernt poly. Trends

- ◆ Anwenden des Differenzenfilters mit Fenstergröße Δt :

$$x'_t = \Delta_{\Delta t} x_t = x_{t+\Delta t} - x_t$$

Entfernt saisonale Trends

- ◆ Glättung (gleitender Durchschnitt mit Fenstergröße q):

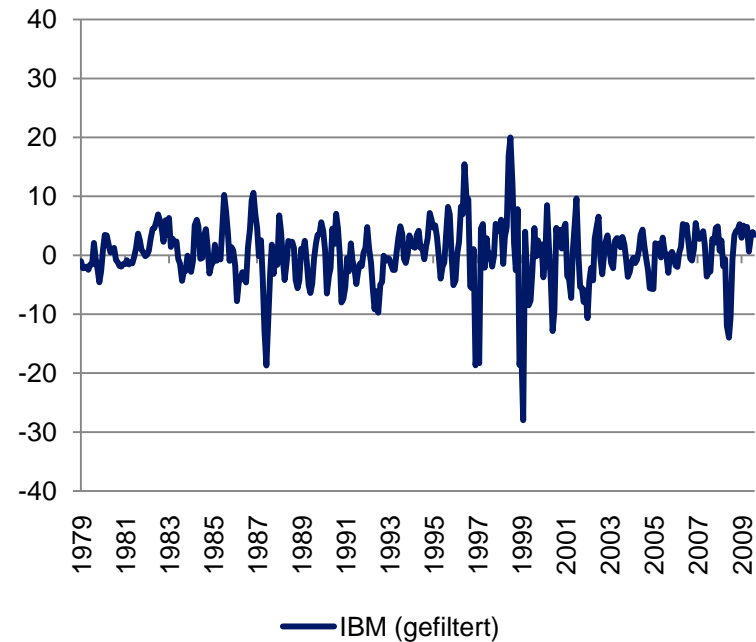
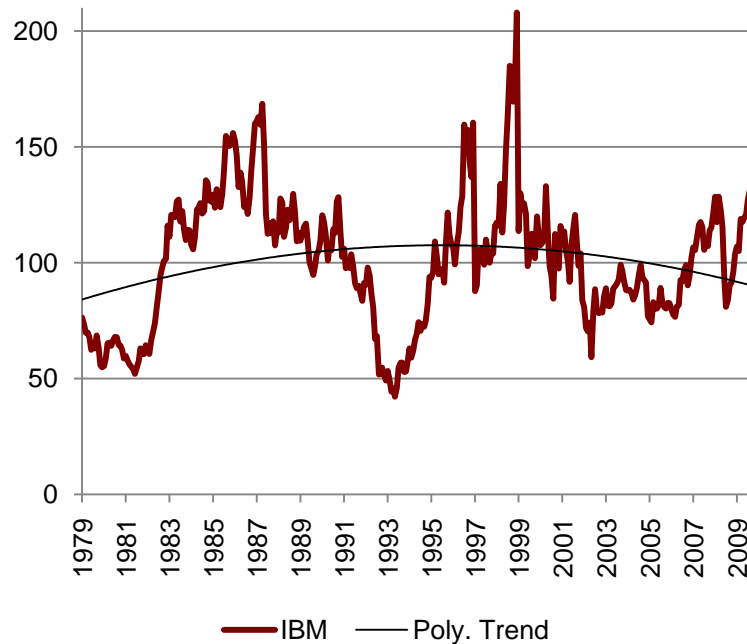
$$x'_t = \frac{1}{2q+1} \sum_{i=-q}^q x_{t+i}$$

Entfernt lokale Trends

Datentransformation

Filter

- Beispiel: Einmaliges Anwenden des Differenzenfilters und Glättung mit $q=1$.



Stationärer stochastischer Prozess

Beispiel für stationäre Prozesse

- Starke stationäre Prozesse:
 - ◆ White-Noise-Prozess (Rauschen).
- Lineare stationäre Prozesse:
 - ◆ Moving-Average-Prozess q -ter Ordnung (MA_q).
- Autoregressive Prozesse:
 - ◆ Autoregressiver Prozesse p -ter Ordnung (AR_p).
 - ◆ Autoregressiver Moving-Average-Prozess ($ARMA_{p,q}$).
- Zählprozesse.
- Poisson-Prozesse.
- (Hidden) Markov-Prozesse & Markov-Ketten.
- Wiener-Prozesse.

Starke stationäre Prozesse

White-Noise-Prozess

■ Annahmen:

- ◆ Aufenthaltsort $x_t = \varepsilon_t$ zum Zeitpunkt t ist unabhängig vom Aufenthaltsort zu einem vorherigen Zeitpunkt.

■ Eigenschaften:

- ◆ Mittelwertfunktion: $\mu_X(t) = 0$
- ◆ Varianzfunktion: $\sigma_X^2(t) = \sigma_X^2$
- ◆ Autokovarianzfunktion: $\gamma_X(h, 0) = 0$

■ Beispiel:

- ◆ Unabhängig normalverteilte Beobachtungen
 $X_t \sim N(0, \sigma_X^2)$.

Lineare stationäre Prozesse

Moving-Average-Prozess

- Beispiel: Ein Eisverkäufer möchte Verkaufszahlen von Eiscreme vorhersagen.
 - ◆ Nur bisherige Verkaufszahlen x_t gegeben.
- Annahme: Kunden essen nur Eis wenn in den letzten Tagen die Sonne oft zu sehen war.
- Modell:
 - ◆ (Unbeobachtete) Rauschwerte ε_t sind die Sonnenstunden pro Tag normiert auf Mittelwert 0.
 - ◆ Beispiel für Modell mit Modellparametern:

$$x_t = \varepsilon_t + 0,8 \cdot \varepsilon_{t-1} + 0,5 \cdot \varepsilon_{t-2} + 0,2 \cdot \varepsilon_{t-3}$$

Lineare stationäre Prozesse

Moving-Average-Prozess

- Annahmen:

- ◆ Aufenthaltsort x zum Zeitpunkt t ist das gewichtete Mittel aus $q+1$ Rauschwerten zu den Zeitpunkten $t-q, \dots, t$:

$$x_t = \varepsilon_t + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}$$

mit Rauschwerten ε_j wobei $E[\varepsilon_i] = 0$ und $\text{Var}[\varepsilon_i] = \sigma_\varepsilon^2$.

- Ziel: Berechnung der Gewichte α_j .

Lineare stationäre Prozesse

Moving-Average-Prozess

- Berechnung der Gewichte α_j am Beispiel eines MA_1 -Prozesses: $x_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1}$

- Umformung des Prozesses ergibt:

$$\begin{array}{l} \varepsilon_j = x_j - \alpha_1 \varepsilon_{j-1} \\ \varepsilon_{j-1} = x_{j-1} - \alpha_1 \varepsilon_{j-2} \\ \vdots \\ \varepsilon_1 = x_1 \end{array} \Rightarrow \begin{array}{l} \varepsilon_j = x_j - \alpha_1 \varepsilon_{j-1} \\ = x_j - \alpha_1 (x_{j-1} - \alpha_1 \varepsilon_{j-2}) \\ = x_j + (-\alpha_1)^1 x_{j-1} + (-\alpha_1)^2 \varepsilon_{j-2} \\ \vdots \end{array} \Rightarrow \varepsilon_j(\boldsymbol{\alpha}) = \sum_{i=0}^{j-1} (-\alpha_1)^i x_{j-i}$$

- Minimieren des mittleren quadratischen Fehlers zwischen Modell und Sequenz mit $E[\varepsilon_t] = \mu_\varepsilon = 0$:

$$\boldsymbol{\alpha}^* = \arg \min_{\boldsymbol{\alpha}} \sum_{j=1}^n (\varepsilon_j(\boldsymbol{\alpha}) - \mu_\varepsilon)^2 = \arg \min_{\boldsymbol{\alpha}} \sum_{j=1}^n \varepsilon_j(\boldsymbol{\alpha})^2$$

Autoregressive Prozesse

Einfacher Autoregressiver Prozess

- Beispiel: Ein Eisverkäufer möchte Verkaufszahlen von Eiscreme vorhersagen.
 - ◆ Nur bisherige Verkaufszahlen x_t gegeben.
- Annahme: Kunden essen nur Eis wenn sie in den letzten Tagen wenig Eis gegessen haben.
- Modell:
 - ◆ Verkaufszahl x_t ist nur abhängig von den vorherigen Verkaufszahlen und einem Rauschwert ε_t (Laune der Kunden).
 - ◆ Beispiel für Modell mit Modellparametern:

$$x_t = \varepsilon_t - 0,7 \cdot x_{t-1} - 0,6 \cdot x_{t-2} + 0,1 \cdot x_{t-3}$$

Autoregressive Prozesse

Einfacher Autoregressiver Prozess

- Annahmen:

- ◆ Aufenthaltsort x zum Zeitpunkt t ist die Summe aus dem Rauschwert zum Zeitpunkt t und dem gewichteten Mittel der p Aufenthaltsorte zu den Zeitpunkten $t-p, \dots, t-1$:

$$x_t = \varepsilon_t + \sum_{j=1}^p \beta_j x_{t-j}$$

mit Rauschwerten ε_t wobei $E[\varepsilon_t] = 0$, $\text{Var}[\varepsilon_t] = \sigma_\varepsilon^2$
und $|\beta_j| \neq 1$.

- Ziel: Berechnung der Gewichte β_j .

Autoregressive Prozesse

Einfacher Autoregressiver Prozess

- Berechnung der Gewichte β_j eines AR_p -

Prozesses:
$$x_t = \varepsilon_t + \sum_{j=1}^p \beta_j x_{t-j}$$

- Idee: Minimieren des mittleren quadratischen Fehlers.

$$\beta^* = \arg \min_{\beta} \frac{1}{n-p-1} \sum_{i=p+1}^n \left(x_i - \varepsilon_i - \sum_{j=1}^p \beta_j x_{i-j} \right)^2 = \arg \min_{\beta} \sum_{i=p+1}^n \left(x_i - \sum_{j=1}^p \beta_j x_{i-j} \right)^2$$

Erwartungswert Null & konstante Varianz

$$\beta^* = \arg \min_{\beta} \left(\begin{bmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} x_p & x_{p-1} & \cdots & x_1 \\ x_{p+1} & x_p & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_{n-p} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \right)^2 \Rightarrow \beta^* = \mathbf{A}^+ \mathbf{y}$$

y

A

Pseudo-Inverse von **A**

Autoregressive Prozesse

Autoregressiver Moving-Average-Prozess

- Motivation:
 - ◆ Fast jeder endliche Datensatz gut an MA- oder AR-Modell mit hoher Ordnung anpassbar.
 - ◆ Je größer die Ordnung desto mehr Parameter.
 - ◆ Ziel: Modell mit möglichst wenigen Parametern welches zudem reale Bedeutung/Interpretation hat.
- Idee: Kombination von MA- und AR-Modellen zu $ARMA_{p,q}$ -Prozess.

$$x_t = \varepsilon_t + \sum_{j=1}^q \alpha_j \varepsilon_{t-j} + \sum_{j=1}^p \beta_j x_{t-j}$$

Autoregressive Prozesse

Autoregressiver Moving-Average-Prozess

- Beispiel: Vorhersage der Werbeausgaben zweier konkurrierender Unternehmen.
 - ◆ Nur bisherige Werbeausgaben x_t und y_t gegeben.
- Modell:
 - ◆ Werbeausgabe x_t ist abhängig von vorherigen Werbeausgabe des Konkurrenten y_{t-1} und Rauschwert.
 - ◆ Werbeausgabe y_t ist abhängig von der vorherigen Werbeausgabe des Konkurrenten x_{t-1} und Rauschwert.

$$\begin{aligned} x_t &= \varepsilon_t + \beta_1 y_{t-1} \\ y_t &= \varepsilon'_t + \beta'_1 x_{t-1} \end{aligned} \quad \Rightarrow \quad x_t = \varepsilon_t + \beta_1 (\varepsilon'_{t-1} + \beta'_1 x_{t-2}) = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \beta_2 x_{t-2}$$

ARMA_{2,1}-Prozess

Autoregressive Prozesse

Autoregressiver Moving-Average-Prozess

- Überlagerungssatz von ARMA-Prozessen:
 - ◆ x_t und y_t seien zwei unabhängige ARMA-Prozesse der Ordnung (p_1, q_1) und (p_2, q_2) .
 - ◆ Summe $z_t = x_t + y_t$ ist wieder ein ARMA-Prozess der Ordnung (p, q) .
 - ◆ Für AR-Ordnung gilt: $p \leq p_1 + p_2$
 - ◆ Für MA-Ordnung gilt: $q \leq \max(p_1 + q_2, p_2 + q_1)$
- Folgerung:
 - ◆ Summe zweier MA-Prozesse ergibt wieder MA-Prozess.
 - ◆ Summe zweier AR-Prozesse ergibt ARMA-Prozess.

Autoregressive Prozesse

Autoregressiver Moving-Average-Prozess

- Berechnung der Gewichte α_j und β_j eines ARMA-Prozesses analog zu MA-Prozessen.
- Abhängig von der Datentransformation unterscheidet man verschiedene ARMA-Prozesse, z.B.:
 - ◆ ARIMA $_{p,d,q}$: x_t ist nach d -maligem Anwenden des Differenzenfilters ein stationärer ARMA $_{p,q}$ -Prozess.
 - ◆ ARMAX: Nach vorheriger Trendbereinigung sind die Residuen (bereinigte Werte x'_t) Realisierungen eines ARMA-Prozesses.

Prognose

Autoregressiver Moving-Average-Prozess

- Vorhersage durch gelernten stochastischen Prozess (Prognose für x_{t+h}):
 - ◆ x_t, x_{t-1}, \dots, x_1 entsprechen den tatsächlichen Beobachtungen.
 - ◆ x_{t+h}, \dots, x_{t+1} werden durch ihre Prognosen ersetzt.
 - ◆ Störterme $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_1$ entsprechen den Prognosefehlern der 1-Schrittprognosen in der Vergangenheit.
 - ◆ Störungen $\varepsilon_{t+h}, \dots, \varepsilon_{t+1}$ werden durch ihren Erwartungswert Null ersetzt.
- Datentransformation (Trendbereinigung, Filter etc.) für prognostizierte Werte x_{t+h} umkehren.

Zusammenfassung

- Idee:
 - ◆ Sequenzen sind Realisierungen eines stochastischen Prozesses (z.B. MA-, AR-, ARMA-Prozess).
- Vorgehen:
 - ◆ Bereinigung der Daten von nicht-stationären Komponenten (Trendbereinigung, Filtern).
 - ◆ Wahl des Prozesses und Schätzen seiner Parameter.
 - ◆ Umkehrung der Datentransformation und Prognose neuer Werte.