Zeitreihenanalyse

Steffen Bickel

Zeitreihen

- Zeitreihen entstehen bei Messung eines Merkmals zu verschiedenen Zeitpunkten.
- z.B.:
  - tägliche Messung der Temperatur an einem Ort,
  - monatliche Verkaufszahlen für ein Produkt,
  - EKG, EEG eines Patienten,
  - Zahl der Einwohner eines Landes,
  - Aktienkurse an aufeinanderfolgenden Börsentagen.

Zeitreihen

- Eine Zeitreihe bezeichnen wir durch $\{x_t, t = 1,..., n\}$.
- Index t indiziert Zeitpunkt, an dem (metrische) Beobachtung x_t gemacht wurde.
- Beobachtungen sind unter Umständen nicht in gleichen Abständen gemacht worden,
  - in einem solchen Fall verwenden wir die Notation $\{x(t_i), i = 1,..., n\}$, wobei t_i Zeitpunkt der i-ten Beobachtung ist.

Aufgaben der Zeitreihenanalyse

- Beschreibung
  - Untersuchen der Charakteristika der Zeitreihe über Kennzahlen und Diagramme.
- Modellierung und Prognose
  - Formulierung eines stochastischen Modells und Schätzung der Modellparameter.
  - Vorhersage zukünftiger Werte aufgrund des angepassten Modells.
- Kontrolle und Regelung
  - Optimale Steuerung des datenerzeugenden Prozesses.

Empirische Kennzahlen von Zeitreihen

- Arithmetisches Mittel

$$\bar{x} = \frac{1}{N} \sum_{t=1}^{N} x_t$$

- Varianz

$$s^2 = \frac{1}{N} \sum_{t=1}^{N} (x_t - \bar{x})^2$$

Empirische Kennzahlen von Zeitreihen

- Zentrale Stellung bei der Untersuchung von Zeitreihen: Abhängigkeiten zwischen verschiedenen Zeitpunkten
- Empirische Kovarianz zweier Zufallsvariablen:

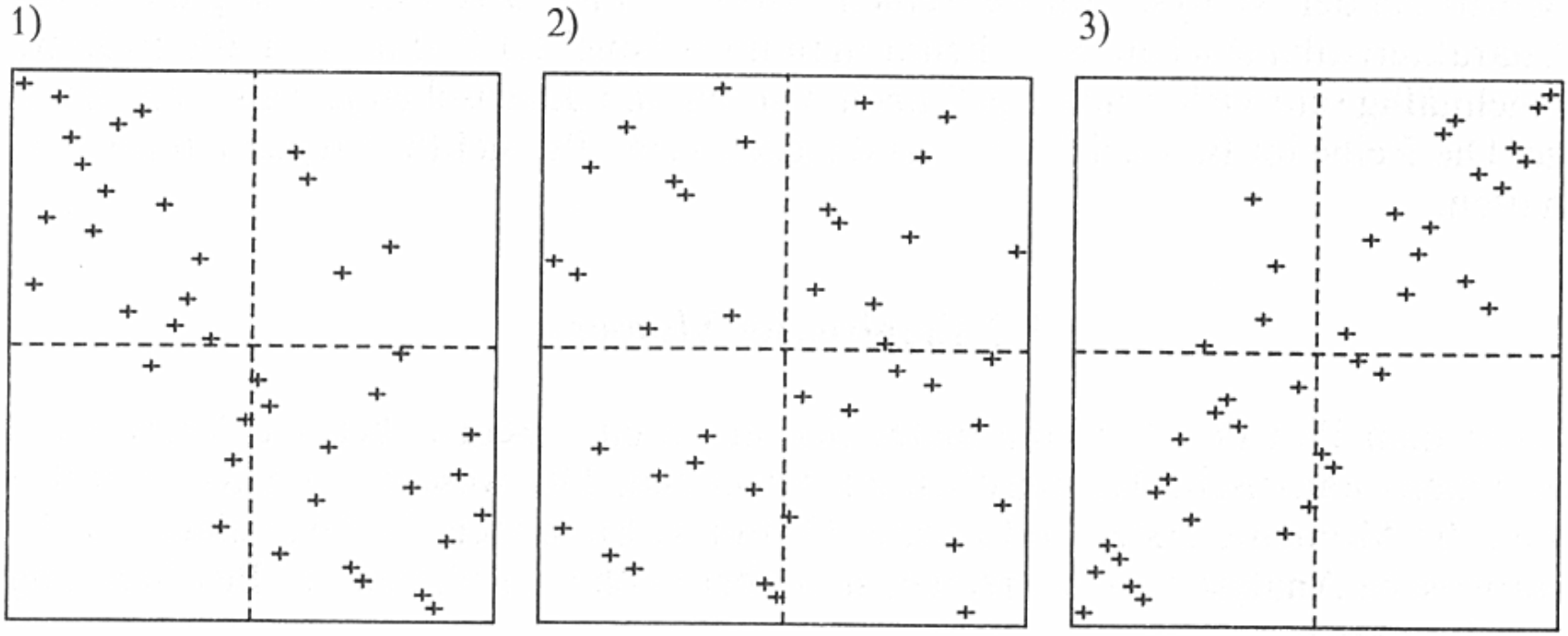
$$c = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

- Empirischer Korrelationskoeffizient r (normierte Kovarianz, r $\in$ [-1;+1]):

$$r = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

- Korrelation = lineare Abhängigkeit

Beispiel: Korrelationskoeffizient



1) $r = -0.72,\ c < 0$ 2) $r = -0.01,\ c \approx 0$ 3) $r = 0.87,\ c > 0$

Empirische Kennzahlen von Zeitreihen

- Autokovarianzfunktion (Kovarianz der Zeitreihe mit sich selbst verschoben um k):

$$c(k) = \frac{1}{N} \sum_t (x_t - \bar{x})(x_{t+k} - \bar{x})$$

- Autokorrelationsfunktion:

$$r(k) = \frac{c(k)}{\sqrt{\sigma^2}\sqrt{\sigma^2}}$$

- Plot der Autokorrelationsfunktion heißt Korrelogramm.
- Enthält wesentliche Informationen über zeitliche Abhängigkeiten.

Beispiel: Autokorrelationsfunktion

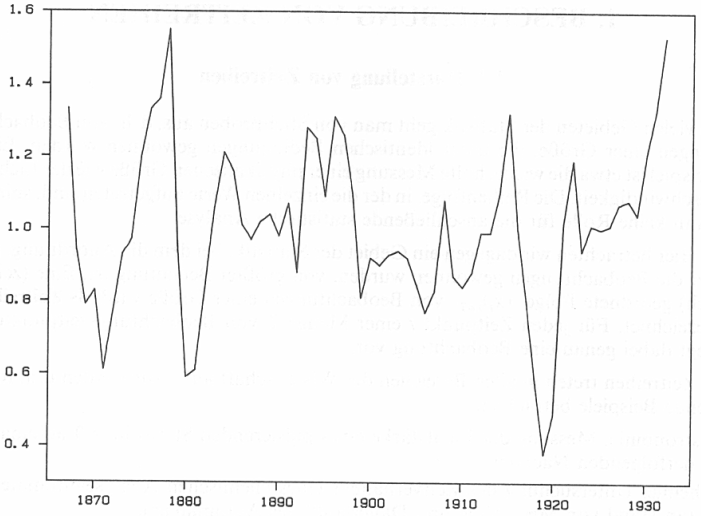


Abb. 1.1.1 *KONKURSE in den USA, (Quelle: B.Greenstein [1935])*

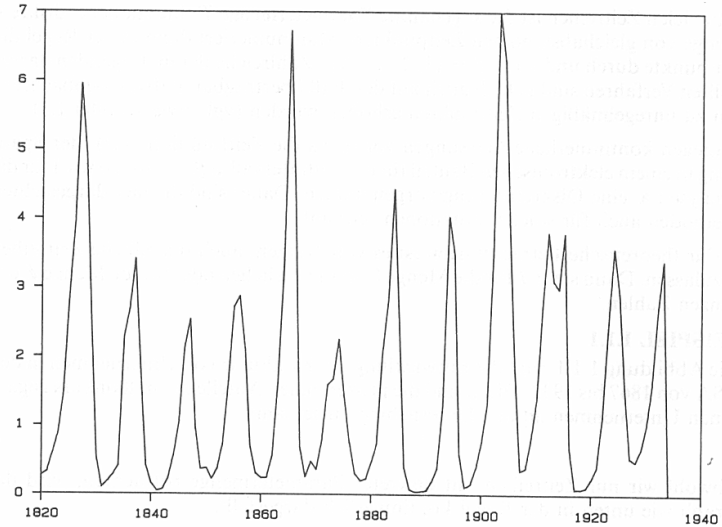


Abb. 1.1.2 *Anzahl gefangener Luchse (Quelle: M.J. Campbell and A.M. Walker [1977])*

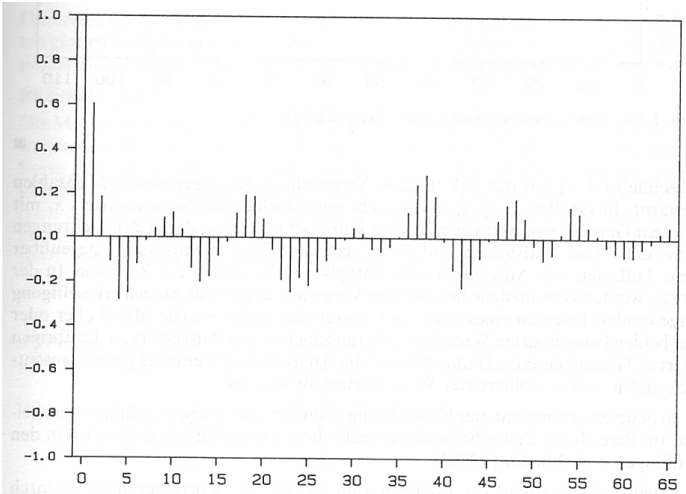


Abb. 1.2.3 *Autokorrelationsfunktion der Reihe KONKURSE*

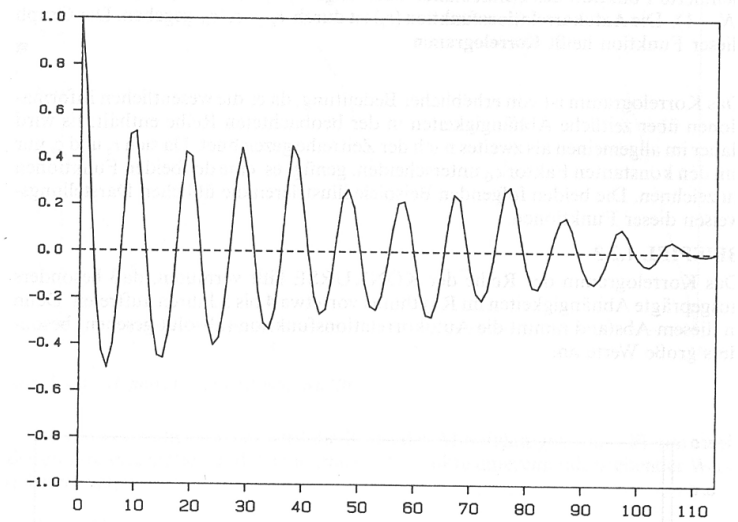


Abb. 1.2.4 *Autokorrelationsfunktion der Reihe LUCHS*

Komponentenmodell

Ursachen für Variation von Zeitreihen:

- Trendkomponente m(t)
  - Langfristige Änderung des Mittelwerts der Zeitreihe
- Saisonkomponente s(t)
  - Kalenderbedingte Schwankungen, die sich relativ regelmäßig wiederholen (Tag, Woche, Jahr)
- Andere zyklische Schwankungen k(t)
  - Sonstige längerfristige wiederkehrende Schwankungen (z.B. Konjunkturzyklus)
- Restkomponente u(t)
  - Zusammenfassung aller sonstigen Einflüsse und Störungen

Komponentenmodell

- Additives Modell:
  - $x_t = m(t) + s(t) + k(t) + u(t)$
- Multiplikatives Modell:
  - $x_t = m(t) \bullet s(t) \bullet k(t) \bullet u(t)$
  - ist angebracht, wenn die Größe der Schwankungen mit dem Niveau der Zeitreihe zunimmt.
  - kann durch Logarithmieren in ein additives Modell transformiert werden:
    - $\log(x_t) = \log(m(t)) + \log(s(t)) + \log(k(t)) + \log(u(t))$

Modellierung der Trendkomponente

- Einfachste Trendfunktion ist Gerade:
  - $m(t) = \beta_1 + \beta_2 t$
- Parameterschätzung über lineare Regression - die Summe des quadrierten Fehlers soll minimal werden:

$$\text{minimiere} \sum_{t=1}^{N} (x_t - m(t))^2 = \sum_{t=1}^{N} (x_t - \beta_1 - \beta_2 t)^2$$

- Partielle Ableitungen null setzen und auflösen, siehe Entscheidungsbaumfolien

$$\hat{\beta}_1 = \bar{x} - \hat{\beta}_2 \bar{t} \qquad \hat{\beta}_2 = \frac{\sum_{t=1}^{N} (\bar{x} - x_t)(\bar{t} - t)}{\sum_{t=1}^{N} (\bar{t} - t)^2}$$

Beispiel: Anpassung Trendgerade

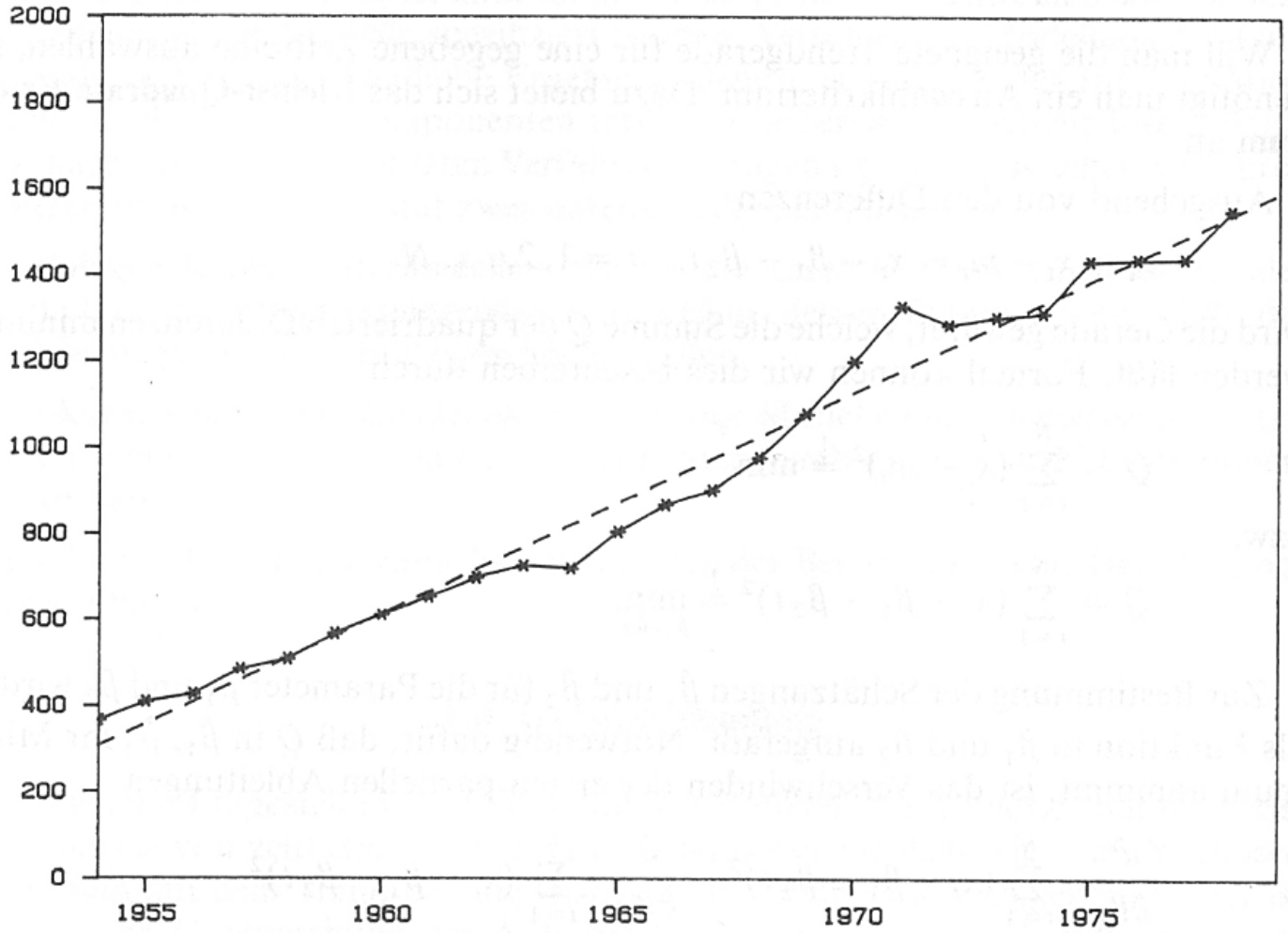


Abb. 1.4.1.1 *Jahreshöchstlasten und angepaßte Trendgerade*

Modellierung der Trendkomponente

- Allgemeineres lineares Trendmodell - Linearkombination von k bekannten Funktionen:

$$m(t) = \beta_1 m_1(t) + \beta_2 m_2(t) + \ldots + \beta_k m_k(t)$$

- Beispielsweise polynomialer Trend (k-1)´ter Ordnung:

$$m(t) = \beta_1 + \beta_2 t + \beta_3 t^2 + \ldots + \beta_k t^{k-1}$$

- entsteht aus dem allg. linearen Trendmodell durch die Wahl:

$$m_1(t) = 1, m_2(t) = t, m_k(t) = t^{k-1}$$

- Parameterschätzung von β_1, β_2 ,..., β_k wieder über Kleinstquadratekriterium
  - Partielle Ableitungen der Summe der Fehlerquadrate null setzen und Gleichungssysteme auflösen.

Modellierung der Trendkomponente

- Transformation nichtlinearer in lineare Trendmodelle oft möglich durch Logarithmieren
  - dann wieder Parameterschätzung mit linearer Regression.

- Beispiele:
  - Exponentionalmodelle

$$x_t = e^{\beta_1 m_1(t) + \beta_2 m_2(t) + \ldots + \beta_k m_k(t)}$$

  - Potenzierte Modelle

$$x_t = m_1(t)^{\beta_1} \cdot m_2(t)^{\beta_2} \cdot m_k(t)^{\beta_k}$$

Komponentenmodell

- Trendkomponente m(t)
- Saisonkomponente s(t)
- Andere zyklische Schwankungen k(t)
- Restkomponente u(t)

- Nachdem Trendkomponente bestimmt und von Zeitreihe subtrahiert, Saisonkomponente schätzen.

Modellierung der Saisonkomponente

- Starre Saisonfigur: s(t) = s(t+p)
- p ist Saisonperiode, p=12 für Jahressaison
- Modellierung mittels p Dummy-Variablen $s_r(t)$, r = 1,...,p

$$s_r(t) = \begin{cases} 1 & \text{wenn der Zeitpunkt t zum Monat r gehört} \\ 0 & \text{sonst} \end{cases}$$

- Modell: $$s(t) = \sum_{r=1}^{12} \beta_r s_r(t)$$

- Man nimmt sich die Reihe für ein Jahr heraus und schätzt Parameter β_r mit Kleinstquadrateansatz:

$$\text{minimiere} \sum_t (x_t - s(t))^2$$

- Ergibt welchen Schätzer für β_r?

Modellierung der Saisonkomponente

- Starre Saisonfigur: s(t) = s(t+p)
- p ist Saisonperiode, p=12 für Jahressaison
- Modellierung mittels p Dummy-Variablen $s_r(t)$, r = 1,...,p

$$s_r(t) = \begin{cases} 1 & \text{wenn der Zeitpunkt t zum Monat r gehört} \\ 0 & \text{sonst} \end{cases}$$

- Modell: $$s(t) = \sum_{r=1}^{12} \beta_r s_r(t)$$

- Man nimmt sich die Reihe für ein Jahr heraus und schätzt Parameter β_r mit Kleinstquadrateansatz:

$$\text{minimiere} \sum_t (x_t - s(t))^2$$

- Ergibt welchen Schätzer für β_r?

$$\beta_{\mathrm{r}} = \frac{1}{|\, s_r(t) = 1 \,|} \sum_{t: s_r(t)=1} x_t$$

Komponentenmodell

- Trendkomponente m(t)
- Saisonkomponente s(t)
- Andere zyklische Schwankungen k(t)
- Restkomponente u(t)

- Nachdem Trend- und Saisonkomponenten bestimmt und von Zeitreihe subtrahiert, andere zyklische Schwankungen schätzen.

Analyse von zyklischen Schwankungen

- Analyse über Fouriertransformation oder
- über fitten von Kosinus-Kurven mit verschiedenen Frequenzen.
  - Ergebnis $I(\lambda)$ darstellbar als Periodogramm.
  - x-Achse: Frequenz; y-Achse: $I(\lambda)$ „Stärke“ in der Ausgangsreihe.

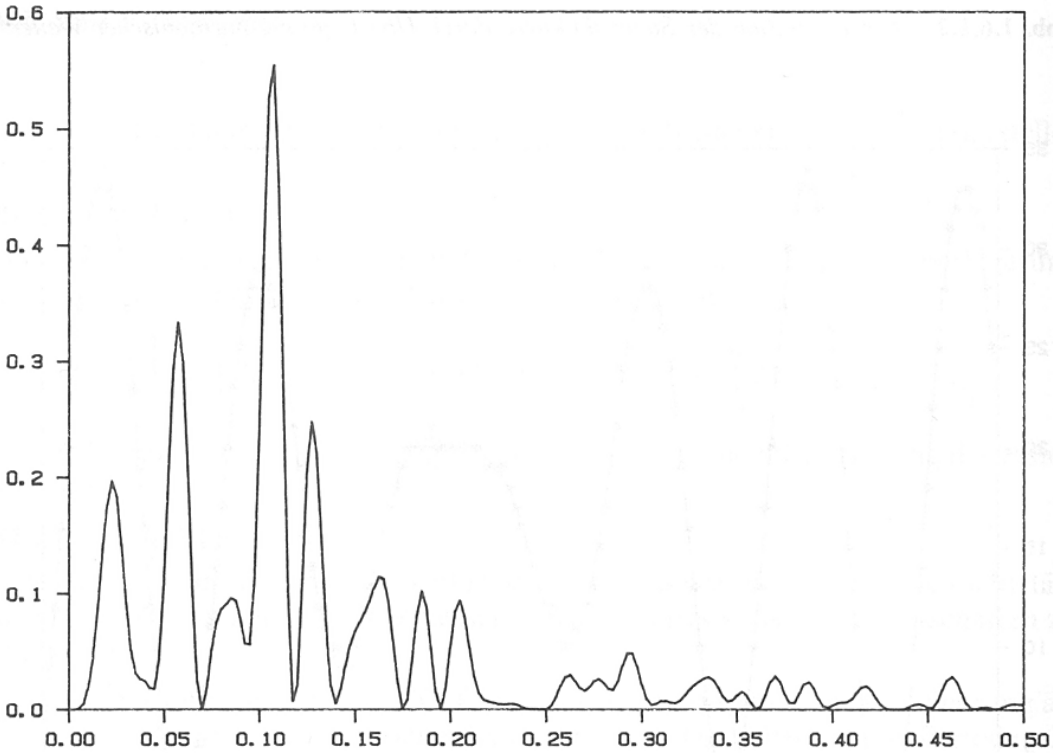


Abb. 1.6.2.1 *Periodogramm der Reihe KONKURSE*

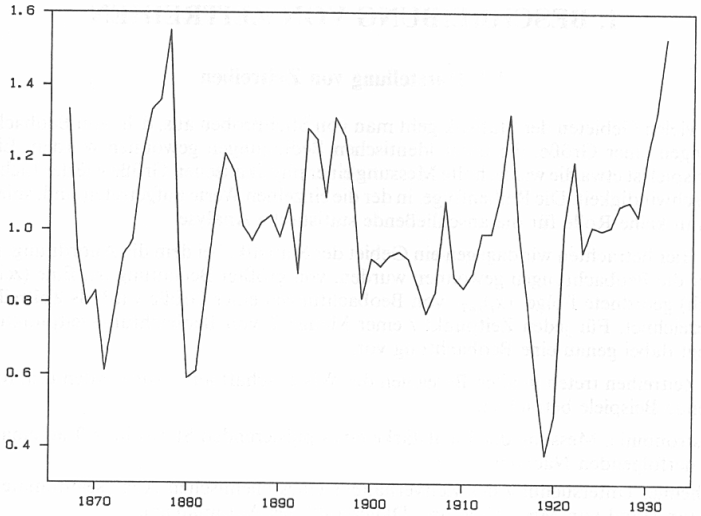


Abb. 1.1.1 *KONKURSE in den USA, (Quelle: B.Greenstein [1935])*

Bestimmung von I(λ) („Stärke“ der Frequenz λ)

- Idee: Fitten einer Kosinus-Kurve mit Amplitude A, Frequenz λ und Phasenverschiebung v:

$$k_\lambda(t) = A\cos(2\pi\lambda t - v)$$

- Phasenverschiebung schwierig zu fitten, deshalb Linearisierung durch Umformung:

$$k_\lambda(t) = A\cos(2\pi\lambda t - v) \equiv \beta_1 \cos(2\pi\lambda t) + \beta_2 \sin(2\pi\lambda t)$$

- Parameter β_1 und β_2 einfach mit Kleinste-Quadrate-Ansatz schätzen:

$$\text{minimiere} \sum_t (x_t - k_\lambda(t))^2$$

- I(γ) soll groß sein, wenn Residuen klein:

$$I(\lambda) \approx \frac{1}{\sum_t (x_t - \hat{k}_\lambda(t))^2}$$

Analyse von zyklischen Schwankungen

- Periodogramm gibt Informationen über regelmässige Zyklen.
- Starke Zyklen können als Komponenten im Komponentenmodell verwendet werden.
  - Beispielsweise die 3 stärksten Zyklen:

$$k(t) = k_{\lambda_1}(t) + k_{\lambda_2}(t) + k_{\lambda_3}(t)$$

  - Einzelne Zyklen ergeben sich aus den vorher gefitteten Kosinus/Sinus-Funktionen:

$$k_\lambda(t) = \beta_{1\lambda}\cos(2\pi\lambda t) + \beta_{2\lambda}\sin(2\pi\lambda t)$$

Komponentenmodell

- Trendkomponente m(t)
- Saisonkomponente s(t)
- Andere zyklische Schwankungen k(t)
- Restkomponente u(t)

- Zur Analyse der Restkomponente müssen alle anderen Komponenten entfernt werden.
- 2 Möglichkeiten:
  - Modelle für Trend, Saison, Zyklen schätzen und empirische Reihe von geschätzter Reihe subtrahieren,
  - Anwendung von Filtern zum Entfernen von Trend und Saison.

Transformation durch Filter

- Filter glätten die Zeitreihe.
- Gleitende Durchschnitte:

$$y_t = \frac{1}{2q+1} \sum_{u=-q}^{+q} x_{t-u} \qquad t = q+1, \ldots, N-q$$

- Anwendung zur Trendentfernung:

$$z_t = x_t - y_t$$

- Oder direkte Weiterverwendung von y_t,
  - Vorteil: Trendmodellschätzung kommt mir Polynomen niedrigerem Grads aus.

Differenzenfilter

- Theorem: f(x) sei ein Polynom vom Grad p > 0:

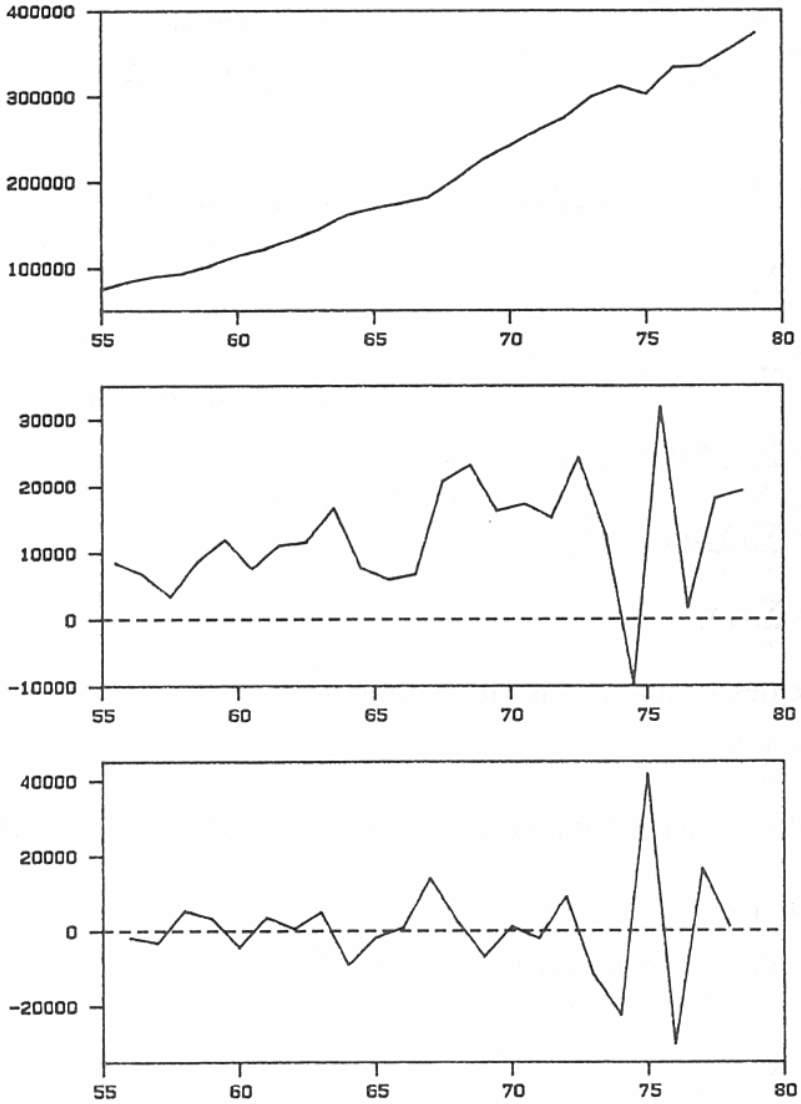
$$f(t) = a_0 + a_1 t + \ldots + a_p t^p$$

Dann ist $g(t) = f(t) - f(t-1)$ ein Polynom vom Grad höchsten p-1.

- Beweis? Ausmultiplizieren, Term mit höchstem Polynomgrad kürzt sich weg.

- Differenzenbildung reduziert Polynomgrad um 1.
- (p-1)-maliger Anwendung auf Polynom p-ten Grades ergibt konstanten Wert.
- „Numerische Differenzierung“.

Beispiel: Differenzenfilter



Jährliche Stromproduktion, erste und zweite Differenzen

Differenzenfilter

- Einmaliges Anwenden des Differenzenfilters entfernt linearen Trend
- Zweimaliges Anwenden des Differenzenfilters entfernt quadratischen Trend, usw.

- Mit Differenzenfiltern können auch saisonale Schwankungen gefiltert werden
  - für Jahressaisonentfernung beispielsweise:

$$g(t) = f(t) - f(t-12)$$

Komponentenmodell

- Trendkomponente $m(t)$
- Saisonkomponente $s(t)$
- Andere zyklische Schwankungen $k(t)$
- Restkomponente $u(t)$

- Wir haben $m(t), s(t), k(t)$ entfernt und wollen nun Modell für Restkomponente anpassen.
- Modellschätzung der Restkomponente mit stochastischen Prozessen.

Stochastische Prozesse

- Ein stochastischer Prozess ist ein Folge von Zufallsvariablen X_t , $t \in T$.
- Annahme: Zeitreihe ist Realisierung eines zeitdiskreten stochastischen Prozesses

Wichtige Prozessmodelle für Zeitreihen

- White Noise-Prozess
- Random Walk-Prozess
- Moving Average-Prozess
- Autoregressiver-Prozess

White Noise-Prozess

- X_t , $t \in T$ ist eine Folge von gleich und unabhängig verteilten Zufallsvariablen ε_t mit Erwartungswert 0.
- Reiner White Noise-Prozess uninteressant, aber als Grundbaustein für komplexere Prozesse wichtig.
- Eigenschaften:
  - $\mu(t) = 0$
  - $\sigma^2(t) = \sigma_\varepsilon^2$
  - $\gamma(k) = 0$ (Autokovar.), für $k <> 0$ (k ist Verschiebung)

Random Walk-Prozess

$$X_t = X_{t-1} + \varepsilon_t = X_0 + \sum_{t=1}^{t} \varepsilon_t$$

- In jedem Zeitschritt wird unabhängig von Vergangenheit per Zufall entschieden in welche Richtung der Prozess fortschreitet.
- ε_t ist ein White Noise-Prozess.
- Anwendung bei der Modellierung von Aktienkursen.
  - Annahme: Jeden Tag macht ein Aktienkurs einen Zufallssprung unabhängig von der Vergangenheit.
- Eigenschaften:
  - $\mu(t) = X_0$
  - $\sigma^2(t) = t\sigma_\varepsilon^2$ (nicht stationär)
- Prozess der Zuwächse ergibt White Noise-Prozess:
  - $Y_t = X_t - X_{t-1}$
- Beste Prognose
  - $E(X_{t+1}|X_t) = X_t$
  - $E(X_{t+k}|X_t) = X_t$
  - $Var(X_{t+k}|X_t) = k\sigma_\varepsilon^2$

Moving Average-Prozess

- Ist ε_t ein White Noise-Prozess, dann ist

$$X_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \ldots - \beta_q \varepsilon_{t-q}$$

  ein Moving Average-Prozess der Ordnung q (MA(q)-Prozess).
- Beobachtungen sind gewichtete Summen aus aktuellen und vergangenen Zufallsschocks
- Subtraktion anstelle Addition hat historische Gründe.
- Beispiel: MA(1)-Prozess
  - $X_t = \varepsilon_t - \beta\varepsilon_{t-1}$

MA(3)-Prozess: Konstruiertes Beispiel

- Langnese möchte Verkaufszahlen von Eiscreme vorhersagen.
  - Nur bisherige Verkaufszahlen x_t gegeben, sonst keine Daten.
- Annahme/konstruierter Zusammenhang:
  - Kunden essen nur Eis, wenn in den letzten 3 Tagen die Sonne oft zu sehen war.
- MA(3)-Prozess:
  - Unbeobachtete Zufallsschocks sind die Sonnenstunden pro Tag.
  - ε_t = [Sonnenstunden an Tag t] (normiert auf Mittelwert 0)
  - Modell mit Modellparametern z.B.:

$$X_t = \varepsilon_t + 0.8\varepsilon_{t-1} + 0.5\varepsilon_{t-2} + 0.2\varepsilon_{t-3}$$

Autoregressiver Prozess

- Ist ε_t ein White Noise-Prozess, dann ist

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \varepsilon_t$$

  ein Autoregressiver Prozess der Ordnung p (AR(p)-Prozess).

- Entspricht formal einem linearen Regressionsmodell für X_t mit vergangenen Werten als erklärende Variablen.
- Beispiel: AR(1)-Prozess
  - $X_t = \alpha X_{t-1} + \varepsilon_t$

AR(3)-Prozess: Konstruiertes Beispiel

- Langnese möchte Verkaufszahlen von Eiscreme vorhersagen.
  - Nur bisherige Verkaufszahlen x_t gegeben, sonst keine Daten.
- Annahme/konstruierter Zusammenhang:
  - Kunden essen nur Eis, wenn sie in den letzten 3 Tagen wenig Eis gegessen haben.
- AR(3)-Prozess:
  - Verkaufszahlen sind nur abhängig von den vorherigen Verkaufszahlen und einem Zufallsschock ε_t (Laune der Kunden).
  - Modell mit Modellparametern z.B.:

$$X_t = -0.8X_{t-1} - 0.5X_{t-2} - 0.2X_{t-3} + \varepsilon_t$$

ARMA-Prozess

- MA- und AR-Prozesse sind Grundbausteine der Zeitreihenanalyse.
- Je größer die Ordnung, desto mehr Parameter stehen zur Verfügung.
- Man kann fast jeden endlichen Datensatz gut an MA- oder AR-Modell mit hoher Ordnung anpassen.
- Man möchte aber Modell
  - mit möglichst wenigen Parametern
  - möglichst gut anpassen,
  - welche zudem reale Bedeutung (Interpretation) haben.
- Erfahrungsgemäß kann man Zeitreihe durch Kombination von MA- und AR-Prozessen zu ARMA-Prozessen mit weniger Parametern modellieren.

ARMA-Prozess

- Ein ARMA[*p, q*]-Prozess besteht aus *p* AR-Termen und aus q MA-Termen.

$$X_t =$$

$$\alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} \qquad \text{AR-Terme}$$

$$+ \varepsilon_t$$

$$- \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \ldots - \beta_q \varepsilon_{t-q} \qquad \text{MA-Terme}$$

Wie entstehen ARMA-Prozesse?

- Überlagerungssatz von ARMA Prozessen:
  - X_t und Y_t seien zwei unabhängie ARMA-Prozesse der Ordnung $[p_1,q_1]$ und $[p_2,q_2]$. Dann ist die Summe $Z_t = X_t + Y_t$ wieder eine ARMA-Prozess der Ordnung $[p,q]$.
  - Für $[p,q]$ gilt:
    - $p \leq p_1 + p_2$ (p: AR-Ordung)
    - $q \leq \max(p_1+q_2,p_2+q_1)$ (q: MA-Ordung)
  - Beweis: Siehe Granger/Morris (1976).
- Folgerung:
  - Summe zweier MA-Prozesse ergibt wieder MA-Prozess.
  - Summe zweier AR-Prozesse ergibt ARMA-Prozess.
  - Auch wenn MA in der Praxis einzeln selten vorkommt, kann es durch AR-Überlagerung entstehen.

Beispiel für ARMA-Prozess

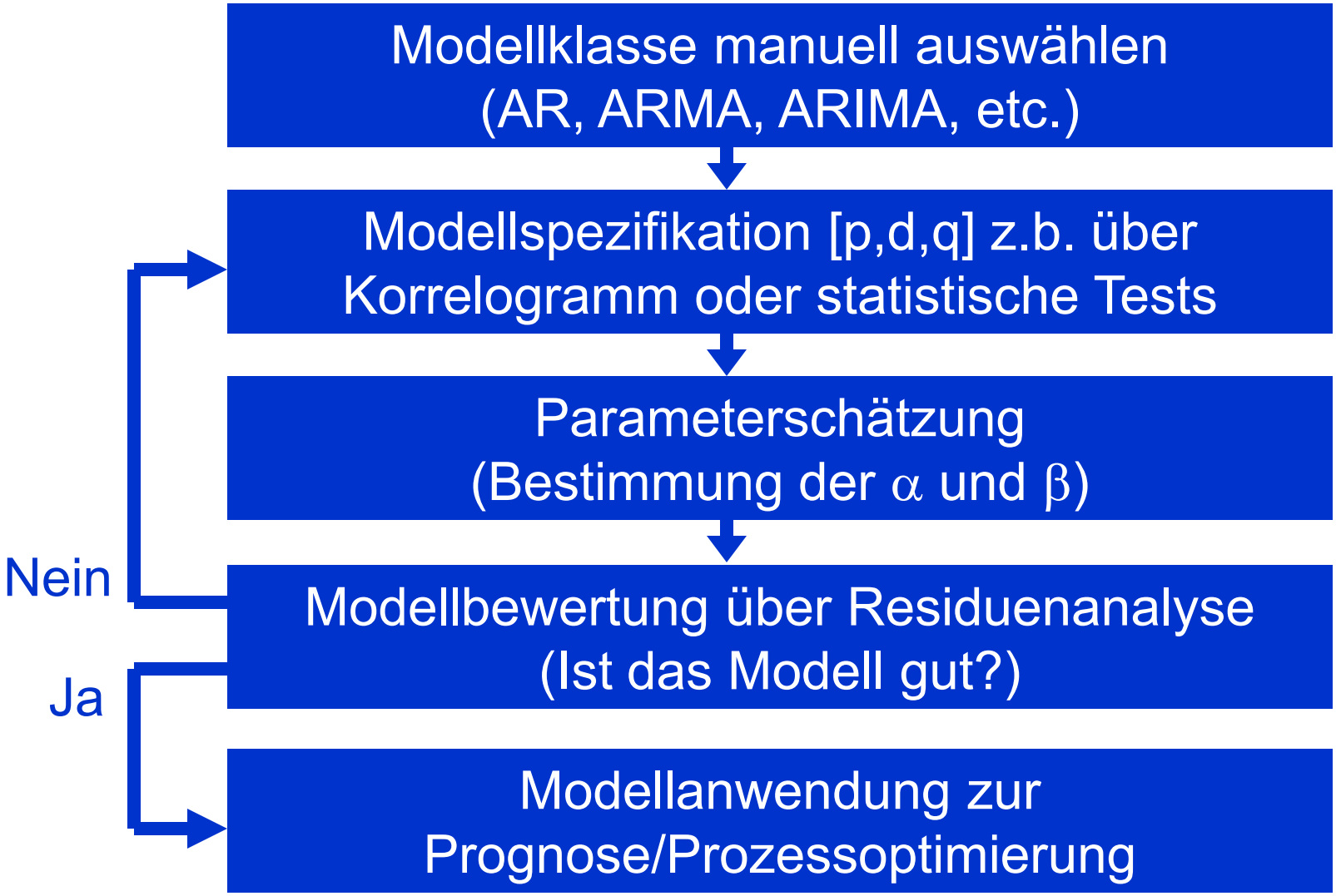
- In der Ökonometrie häufig studierte Feedback-Systeme:
  - $X_t = \alpha Y_{t-1} + \varepsilon_t$
  - $Y_t = \beta X_{t-1} + \varepsilon'_t$
- Beispielsweise Werbeausgaben X_t und Y_t zweier konkurrierender Unternehmen
- Einsetzen von Y_t in X_t ergibt:
  - $X_t = \alpha(\beta X_{t-2} + \varepsilon'_{t-1}) + \varepsilon_t = \alpha\beta X_{t-2} + \alpha\varepsilon'_{t-1} + \varepsilon_t$
- Feedback von Werbeausgaben t-2.
- Entspricht ARMA(2,1)-Prozess, Kombination aus AR(2) und MA(1).

ARIMA-Prozess

- ARIMA: Autoregressiver-Integrierter-Moving-Average-Prozess.
- Einbindung eines Differenzenfilters in ARMA-Modell.

- Ist X_t nach d-maligem Anwenden des Differenzenfilters ein stationärer ARMA(p,q)-Prozess, so bezeichnet man den Prozess X_t als ARIMA(p,d,q)-Prozess
- Der Name ”I=Integriert” kommt daher, dass man die Elemente des ARMA-Prozesses aufsummieren muss, um wieder X_t zu erhalten.

Box-Jenkins-Verfahren zur Modellierung von Zeitreihen



Parameterschätzung von AR-Modellen

- Es gibt verschiedene Möglichkeiten Parameter für MA, AR und ARMA-Modelle zu schätzen.
- Einfache Methode für AR-Modelle: Lineare Regression.

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \varepsilon_t$$

- Die Ausprägungen x_{t-k} für X_{t-k} , k = 0, …, p sind bekannt.
- Es kann eine lineare Regression mit Kleinste-Quadrate-Ansatz geschätzt werden:

$$\text{minimiere} \sum_{t=p+1}^{N} \left(x_t - \left(\sum_{k=1}^{p} \alpha_k x_{t-k} \right) - \varepsilon_t \right)^2$$

- Zufallsschocks kennen wir nicht, deshalb mit Erwartungswert $E(\varepsilon_t) = 0$ ersetzen:

$$\text{minimiere} \sum_{t=p+1}^{N} \left(x_t - \sum_{k=1}^{p} \alpha_k x_{t-k} \right)^2$$

Prognosen mit ARIMA-Modellen

- Wir haben angepasstes ARIMA(p,d,q)-Modell:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \ldots - \beta_q \varepsilon_{t-q}$$

- Prognose für x_{t+h}:

$$\hat{x}_{t+h} = \alpha_1 x_{t+h-1} + \ldots + \alpha_p x_{t+h-p} + \varepsilon_{t+h} - \beta_1 \varepsilon_{t+h-1} - \ldots - \beta_q \varepsilon_{t+h-q}$$

  - x_t, x_{t-1},… entsprechen den tatsächlichen Beobachtungen
  - x_{t+h}, x_{t+h-1},…,x_{t+1} werden durch ihre Prognosen ersetzt
  - Die Störterme ε_t, ε_{t-1},… entsprechen den Prognosefehlern der 1-Schrittprognosen in der Vergangenheit
    $\varepsilon_t = x_t - \hat{x}_t, \varepsilon_{t-1} = x_{t-1} - \hat{x}_{t-1}, \ldots$
  - Die Störungen ε_{t+h}, ε_{t+h-1},…, ε_{t+1} werden durch ihren Erwartungswert null ersetzt.

Prognosen mit ARIMA-Modellen

- Am Ende d-mal über die ganze Reihe summieren, um den Differenzenfilter rückgängig zu machen.

- Dieses Prognoseverfahren minimiert den erwarteten quadratischen Fehler der Prognose:

$$E[(X_{N+h} - \hat{x}_{N+h})^2]$$

Modelldiagnose über Residuenanalyse

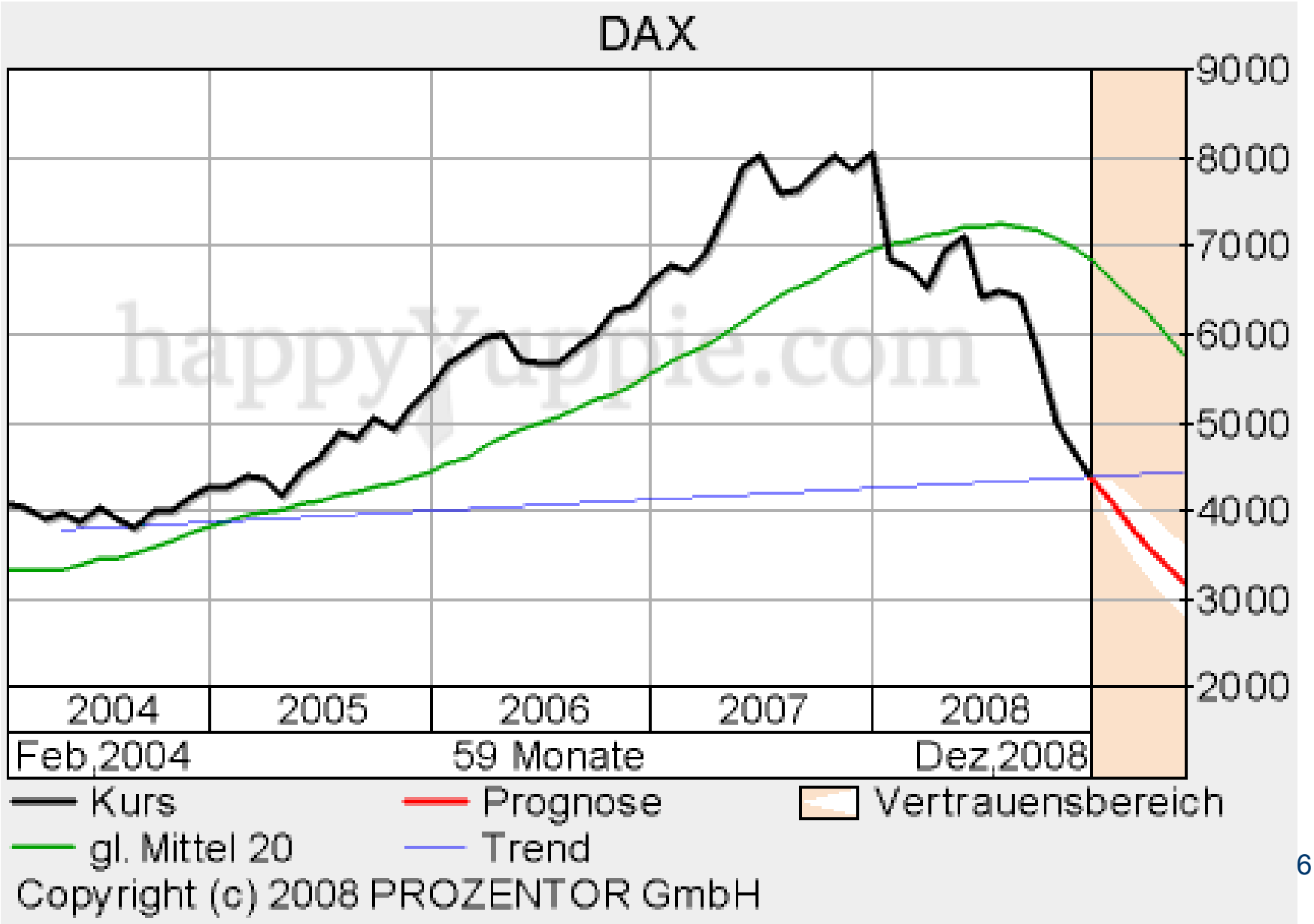
- Residuen = Daten minus ”gefittete” Werte

- Gut angepasstes Modell:
  - Residuen haben zufälligen Charakter (White Noise-Prozess),
  - man hat alle Regelmässigkeiten des wirklichen Prozesses modelliert.

- Plotten der Residuen:
  - Zyklen oder andere Regelmässigkeiten deuten auf schlechtes Modell hin.
- Plotten des Korrelogramm der Residuen:
  - Werte deutlich verschieden von null deuten auf schlechtes Modell hin.

- Es gibt eine Vielzahl von statistischen Tests für die Modelldiagnose.

Praxisbeispiel: ARIMA-Prognosen mit Konfidenzintervallen



Weitere Zeitreihenanalyseprobleme

- Klassifikation
  - Bsp.: Klassifikation von EKG-Zeitreihen zur Vorhersage von Herzproblemen
  - Kernel-Klassifikator benötigt Kernel (Ähnlichkeitsfunktion)
- Anomalieerkennung (unähnlich zu bisherigem Verlauf)
  - Bsp.: Erdbeben-/Tsunami-Vorhersage
- Clustering, Gruppieren von ähnlichen Zeitreihen
- Retrieval, Zeitreihensuche
  - Bsp.: Query-by-Humming

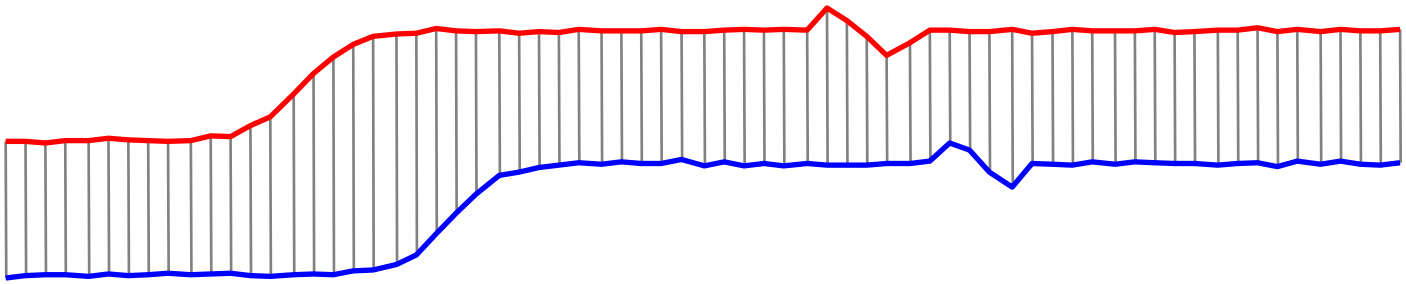
- $\rightarrow$ Ähnlichkeitsmaß für Zeitreihen benötigt

Euklidische Distanz

- Zeitreihen als Vektoren auffassen
- Euklidische Distanz zwischen Vektoren:

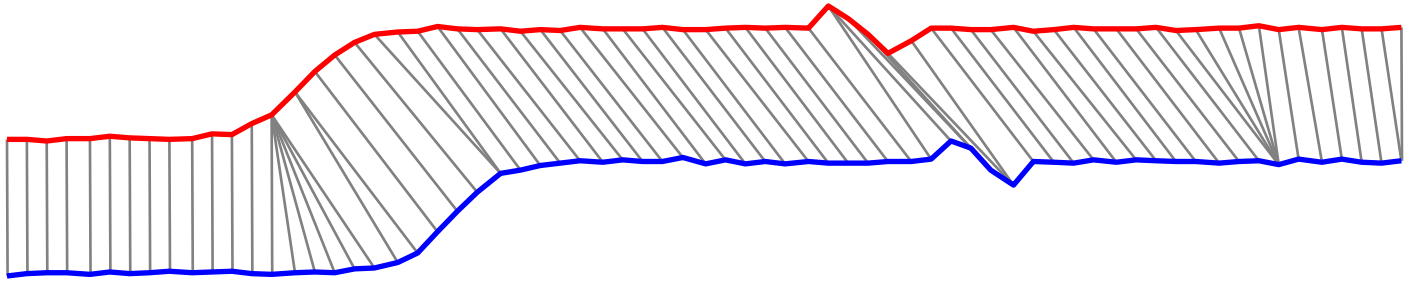
$$D_{eucl}(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

- Verschobene/gedehnte Motive werden nicht berücksichtigt.

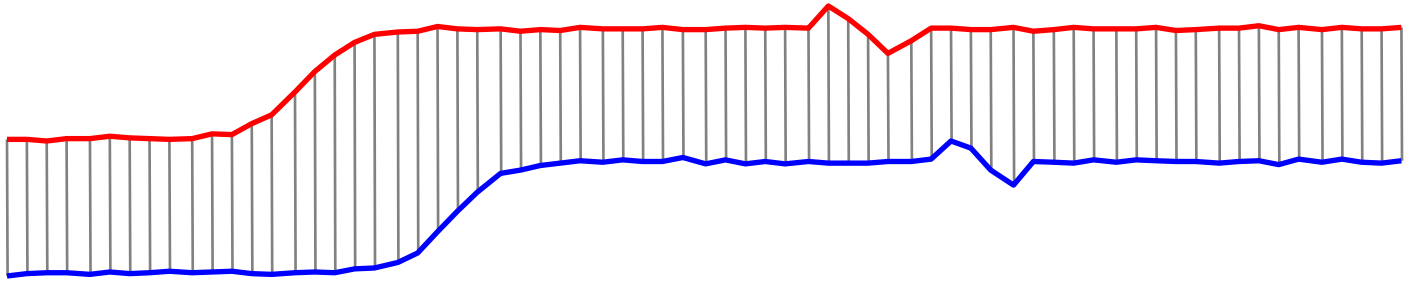


Dynamic Time Warping (DTW)

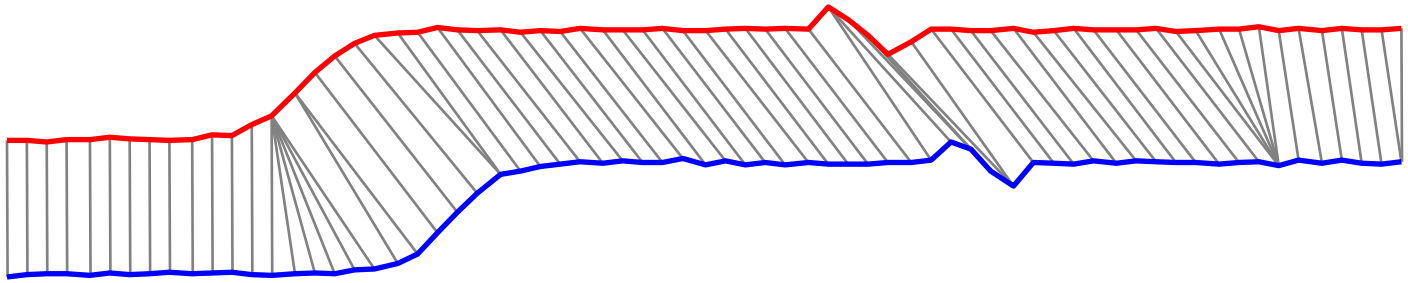
- Mit Dynamic Time Warping



- Ohne Dynamic Time Warping



Dynamic Time Warping (DTW)



- Zuordnungsfunktionen $\pi_1(i)$ und $\pi_2(i)$
- DTW-Distanz ist minimale verschobene euklidische Distanz (quadriert):

$$D_{DTW}(x, y) = \underset{\pi}{\operatorname{argmin}} \sum_i (x_{\pi_1(i)} - y_{\pi_2(i)})^2$$

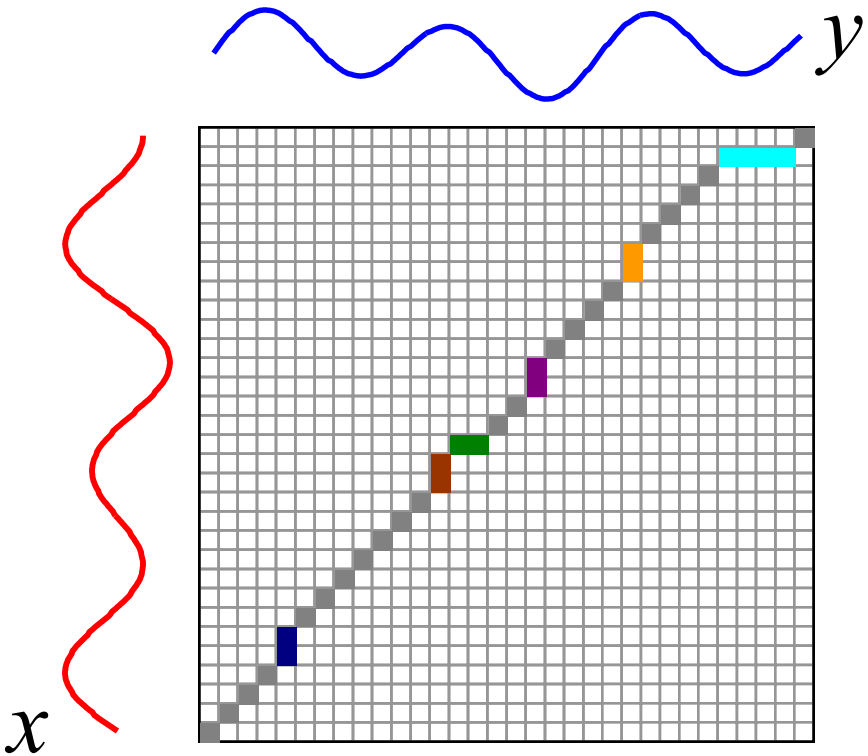
- Minimierung mit dynamischer Programmierung

DTW-Berechnung mit Dynamischer Programmierung

- Rekursive Definition von $\gamma(i,j)$, minimale quadrierte verschobene eukl. Distanz bis zu Zeitpunkten i und j

$$\gamma(i,j) = (x_i - y_j)^2 + \min(\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1))$$

- $\gamma(i,j)$ entspricht kürzestem gewichteten Pfad durch Gitter bis zu Koordinaten (i,j):



Algorithmus für DTW-Distanz

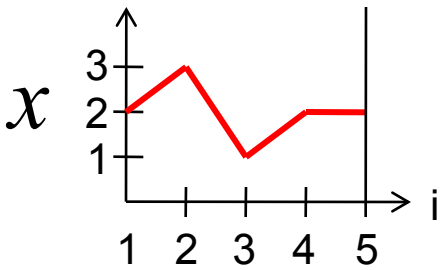
- Initialisierung:
  - n = [Länge von Zeitreihe x], m = [Länge von Zeitreihe y]
  - $\gamma(0,0)=0$, $\gamma(i,0)=\text{inf}$, $\gamma(0,j)=\text{inf}$, für alle i,j

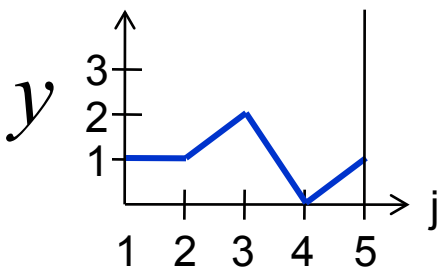
- for i=1 to n do
  - for j=1 to m do

$$\gamma(i,j) = (x_i - y_j)^2 + \min(\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1))$$

  - end
- end
- return: $\gamma(n,m)$

DTW-Distanz: Beispiel





Gamma-Tabelle

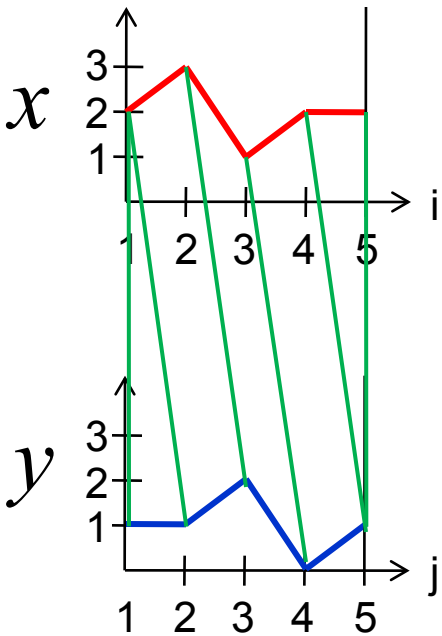
i=5						
i=4						
i=3						
i=2						
i=1						
i=0						
	j=0	j=1	j=2	j=3	j=4	j=5

x

y

$$\gamma(i,j) = (x_i - y_j)^2 + \min(\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1))$$

DTW-Distanz: Beispiel



Gamma-Tabelle

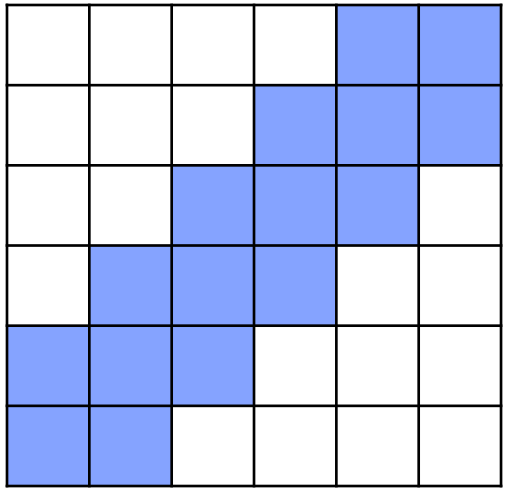
x	j=0	j=1	j=2	j=3	j=4	j=5
i=5	inf	7	7	4	8	6
i=4	inf	6	6	4	8	5
i=3	inf	5	5	4	4	4
i=2	inf	5	5	3	11	10
i=1	inf	1	2	2	6	7
i=0	0	inf	inf	inf	inf	inf

y

$$\gamma(i,j) = (x_i - y_j)^2 + \min(\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1))$$

DTW-Beschleunigung

- Original-Algorithmus hat quadratische Laufzeit O(n*m)
- Beschleunigung durch Beschränkung auf kurze Verschiebung, hier z.B. maximal ein Zeitschritt:
- Nur Pfade im blauen Bereich zugelassen.



- Mit speziellen Indexierungsverfahren und unteren Schranken von DTW lässt sich Laufzeit stark verkürzen.

DTW-Kernel

- DTW-Distanz ist nicht positiv definit.
- Kernel-Methoden brauchen pos. def. Kernel.
- Exponentieren ergibt pos. def. Kernel, λ ist Parameter:

$$k_{DTW}(x, y) = \exp(-\lambda D_{DTW}(x, y))$$

Verarbeitungsschritte für Ereignisvorhersage mit Zeitreihendaten

1. Fenstergröße festlegen.
2. Fenster über Zeitreihe schieben und fenstergroße Unterzeitreihen erzeugen.
3. Unterzeitreihen normieren auf Mittelwert=0 und Varianz=1.
4. Jede Unterzeitreihe markieren mit Label (nach Ende von Unterzeitreihe trat Ereignis auf/nicht auf).
5. DTW-Kernel berechnen für alle paarweisen Kombinationen von Unterzeitreihen.
6. Kernel-Klassifikator lernen (kommt später noch).
7. Neue Zeitreihenstücke in Kernel-Klassifikator stecken $\rightarrow$ Vorhersage.