

Grundlegende statistische Analyse von Zeitreihen

Eine Zeitreihe ist eine Folge von Beobachtungen (y_t) zu verschiedenen Zeitpunkten (t):

$$t : 1, 2, 3, 4, \dots, T$$

$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$

$$y_t : y_1, y_2, y_3, y_4, \dots, y_T$$

Typisch: die Zeitpunkte sind äquidistant.

Komponenten einer Zeitreihe:

1. NIVEAU : durchschnittliches Niveau der Zeitreihe.
2. TREND : langfristige Auf- oder Abwärtsbewegung
3. SAISONELLE KOMPONENTE : regelmäßig wiederkehrende Muster
4. ZUFALLSKOMPONENTE : (Rest/Noise) : unregelmäßige Schwankungen

A. Additives Modell. $y_t = L_t + T_t + S_t + \varepsilon_t$

L_t : Level S_t : Saisonell Additiv ist sinnvoll, wenn die Schwankungen
 T_t : Trend ε_t : Zufallsteil in ähnlicher Größe bleiben, egal wie hoch das Niveau ist.

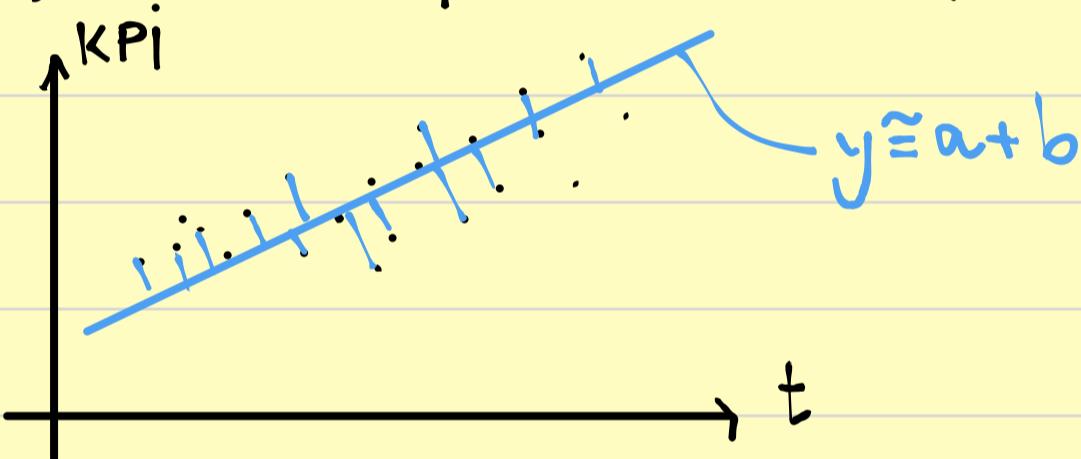
8. Multiplikatives Modell. $y_t = L_t \cdot T_t \cdot S_t \cdot E_t$

Multiplikativ ist sinnvoll, wenn die Schwankungen proportional zum Niveau sind (z.B. höhere Umsätze → große absolute Schwankungen).

Methoden

I. TRENDANALYSE. (Lineare Regression)

Ein einfaches Modell ist ein linearer Trend $y_t \approx a + b t$



II. GLEITENDER DURCHSCHNITT. (Moving Average)

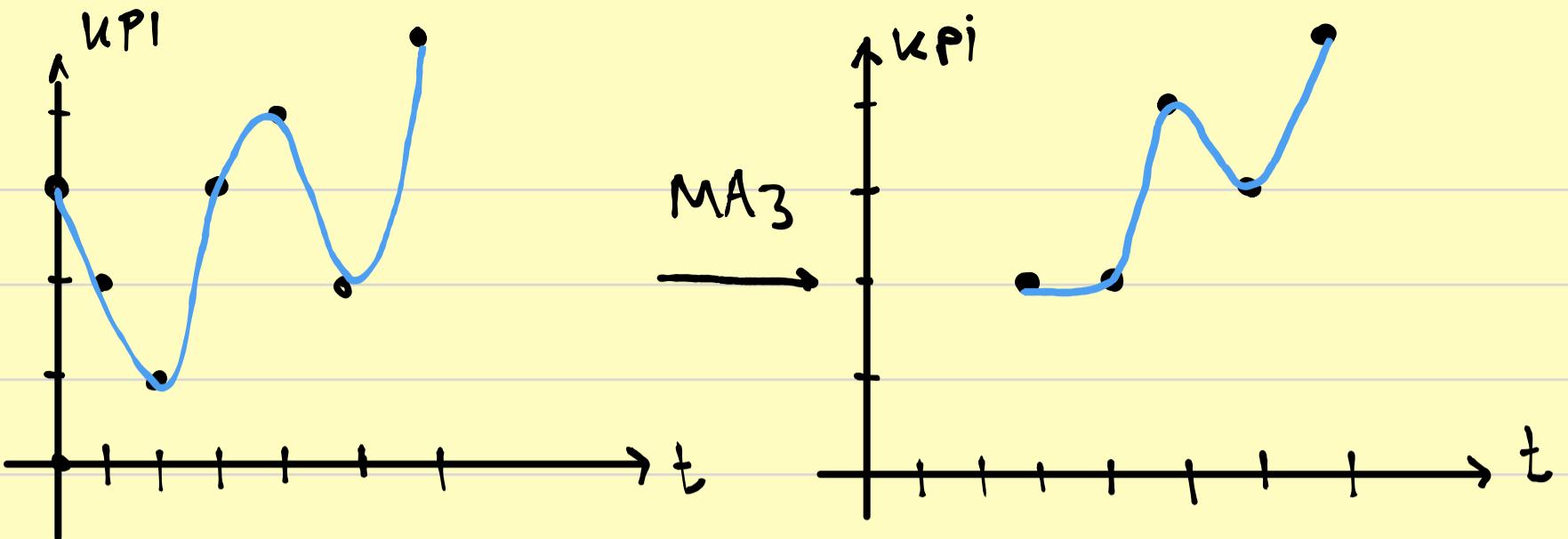
Glättung der Zeitreihe, um kurzzeitige Schwankungen zu reduzieren und den Trend besser zu sehen.

Definition (einfacher gleitender Durchschnitt der Länge k)

$$MA_K = \frac{1}{K} \sum_{j=0}^{K-1} y_{t-j}$$

z.B. $K=3$:

$$MA_3 = \frac{1}{3} (y_t + y_{t-1} + y_{t-2})$$



III. SAISONALE MUSTER (Wochentage, Monate, ...)
Zeitreihen haben oft periodische Muster.

$$y_t = \frac{\text{Level}}{\text{Trend}_t} + \text{saisonale Effekte} + \text{Zufall}$$

1. Level/Trend t schätzen (z.B. Trendlinie)

2. ... Residuen berechnen:

$$\text{Residuum}_t = y_t - \text{Trend}_t$$

IV. AUTOKORRELATION (Grundlage für LSTM)

Begriff: Lag (Verzögerung)

Lag 1: Zusammenhang zw. y_t & y_{t-1}

Lag 2: " " y_t & y_{t-2}

Einfache (Stichproben-) Autokorrelation bei Lag 1:

$$r_1 = \frac{\sum_{t=2}^T (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

- $r_1 > 0$: hohe Werte folgen tendenziell auf hohe Werte (Trägheit)
- $r_1 < 0$: hohe Werte werden tendenziell von niedrigen gefolgt (Wechsel)
- $r_1 \approx 0$: weniger linearer Zusammenhang

Prognoseideen.

1. Naiver Prognoseansatz : $\hat{y}_{T+1} = \hat{y}_T$

2. Durchschnittsmodell : $\hat{y}_{T+1} = \bar{y}$

3. Trendmodell :
 $\hat{y}_{T+h} = \hat{a} + \hat{b}(T+h)$
 (die Parameter werden mit Regression ermittelt)

4. Trend + Saison

$$\hat{y}_{T+h} = \text{Trend}_{T+h} + \text{Index}$$

LSTM . Long Short Term Memory .

