Zeitreihen

Wir lernen, zeitliche struktur in den Daten zu erkennen und zu modellieren: Trend & Saisonalität sicher unterscheiden, Stationarität als Bedingung verstehen, Autokorrelation bzw. partielle Autokorrelation, durch Differänzen stationären machen, Zeit sauber in Lern- & Prüfphase trennen, Vergleichsmodelle bauen, ...

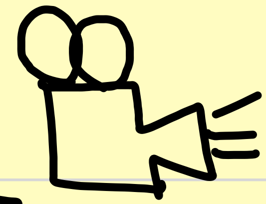
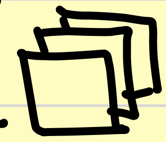
Theorie.

A1. Zeitreihe & Zeitindex - die Reihenfolge ist heilig.

· Eine Zeitreihe ist eine geordnete Liste von Messwerten $\{y_1, y_2, \ldots, y_T\}$, die in gleichen Abständen erhoben werden (i.e. täglich, wöchentlich, ...)

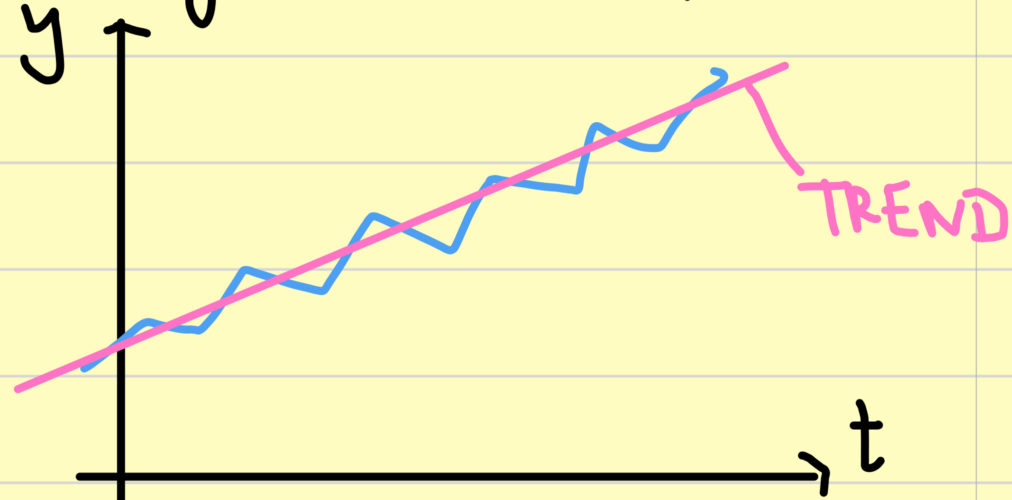
· Der Zeitindex (Datum/Uhrzeit) gehört zur Information. Die Reihenfolge darf nicht gemischt werden, denn Abhängigkeiten in der Zeit wurde zerstört.

· Bei unregelmäßigen Messzeitpunkten wird zunächst auf eine sinnvolle, gleichmäßige, Freqvenz umgerechnet. Fehlende Zeitpunkte werden explizit sichtbar gemacht.

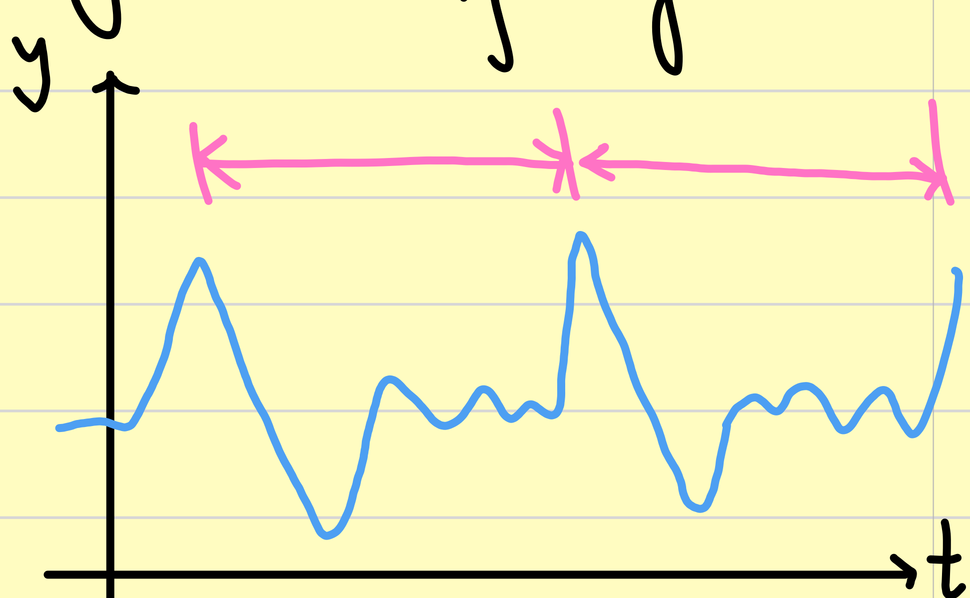
PRAXIS: Eine Zeitreihe ist wie ein Film , nicht eine Ansammlung von Bildern Σ .

A2. Zerlegung in Trend, Saisonalität & Rest.

- Trend beschreibt die langsame Veränderung des mittleren Niveaus über längere Zeiträume, zB ein stetiges Ansteigen.



- Saisonalität ist ein wiederkehrendes Muster mit fester Periodenlänge: i.e. ein Wochenrhythmus mit Periodenlänge 7 Tage. Auch mehrer Saisonalitäten sind gleichzeitig möglich.



- **Rest** sind kurzfristige Schwankungen, die weder Trend noch Saisonalität erklären.

- **ADDITIVES MODELL** für Zeitreihenanalyse:

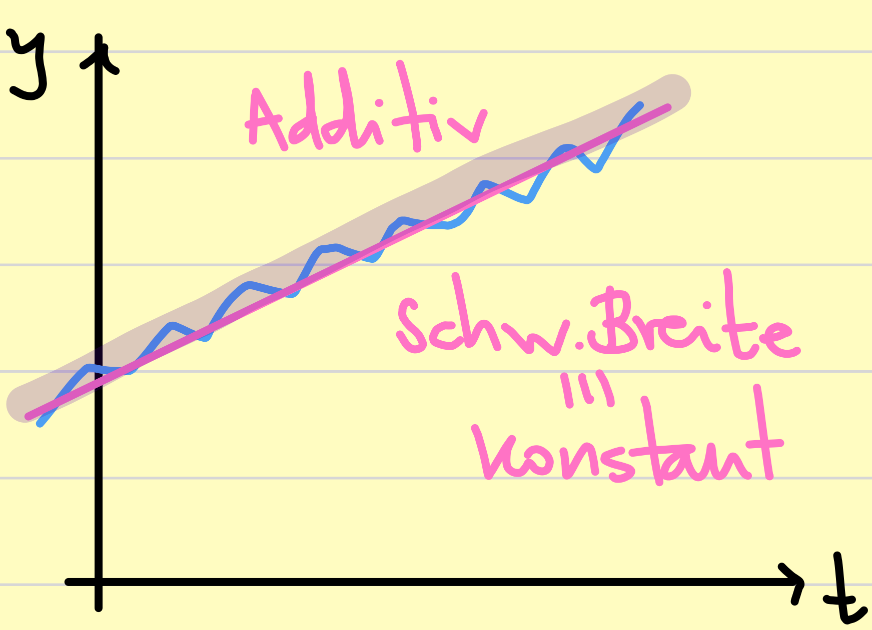
$$\boxed{y_t = \text{Trend}_t + \text{Saison}_t + \text{Rest}_t}$$

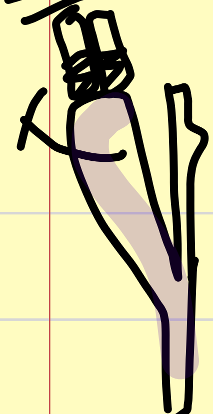
Diese Darstellung passt, wenn die Schwankungsbreite ungefähr konstant bleibt.

- **MULTIPLIKATIVES MODELL**:

$$\boxed{y_t = \text{Trend}_t \cdot \text{Saison}_t \cdot \text{Rest}_t}$$

Diese Darstellung passt, wenn die Schwankungsbreite mit dem Niveau mitwächst.





Bei positiven Zeitreihen, deren Schwankungen mit dem Niveau wachsen, hilft oft die Logarithmus-transformation. Der Logarythmus verwandelt die multiplikative Struktur in eine additive, was die Analyse vereinfacht.

$\log y$ vs $\log t$

$$\log[y_M] = \log[\text{Trend} \cdot \text{Sais.} \cdot \text{Rest}] =$$
$$= \log[\text{Trend}] + \log[\text{Sais}] + \log[\text{Rest}]$$

?David

A3. Stationarität als Arbeitsbedingung

- Eine schwach stationäre Zeitreihe hat ungefähr einen konstanten Mittelwert, eine ungefähr konstante Varianz und eine Abhängigkeit, die nur vom zeitlichen Abstand zweier Messungen abhängt.

- Viele klassische Modelle für Zeitreihen arbeiten zuverlässig, wenn diese Bedingung erfüllt ist.

- Anzeichen für fehlende Stationarität sind ein klarer Trend, eine Varianz, die im Verlauf zu oder abnimmt, und eine Autokorrelation, die langsam abfällt.

· WICHTIGES WERKZEUG ist die „ERSTE DIFFERENZ"

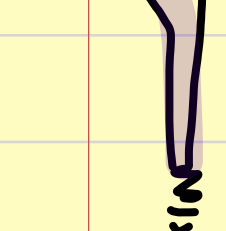
$\boxed{\nabla y_t = y_t - y_{t-1}}$. Sie entfernt lineare Trends und macht die Serie häufig stationär.

· Ein weiteres Werkzeug die Saisonale Differenz mit Periodenlänge „s"

$\boxed{\nabla_s y_t = y_t - y_{t-s}}$. Sie entfernt periodische Muster in „s".

· Beide können kombiniert werden.

· Zur Stabilisierung der Varianz, helfen Transformationen wie der Logarythmus.



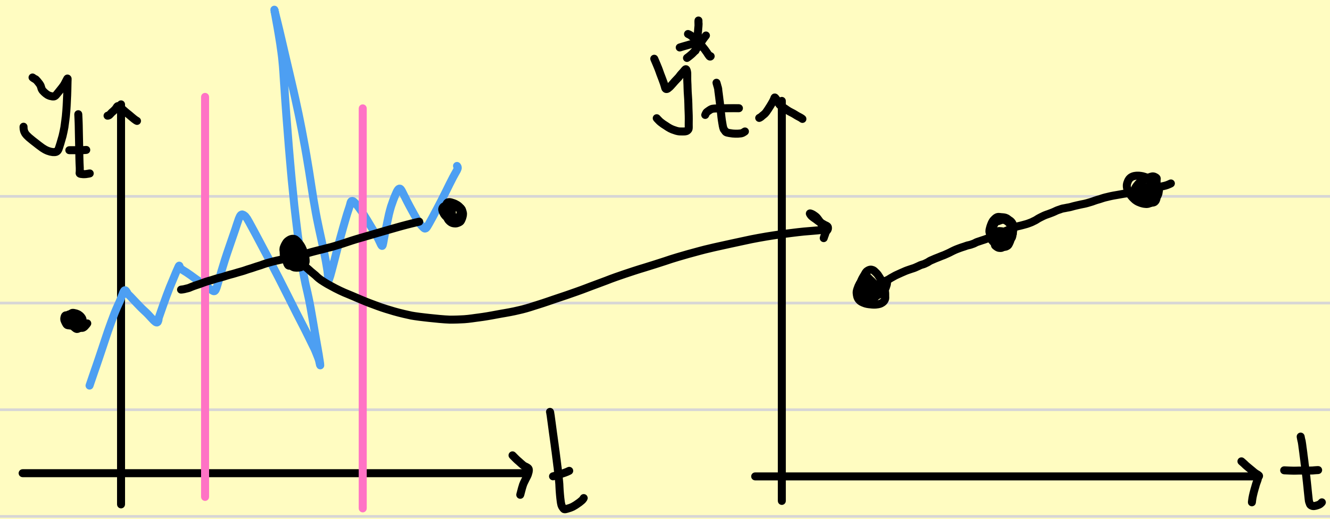
Wir verwenden so wenig Differenzen wie möglich.

A4. Verzögerungen Autokorrelationsfunktion und partielle Autokorrelationsfunktion

- Ein Verzögerung der Ordnung „k" vergleich den Wert y_t mit dem Wert y_{t-k}.
- Die Autokorrelationsfunktion misst für jede Verzögerung den linearen Zusammenhang zw. y_t & y_{t-k}. Ein sich wiederholendes Muster führt zu ausgeprägten Spitzen bei Vielfachen der Periodenlänge.)
- Die partielle Autokorrelationsfunktion misst den direkten Zusammenhang zw y_t & y_{t-k} nachdem der Einfluss aller kleineren Verzögerungen herausgerechnet wurde.

A5. Glätten und Umrechnung der Frequenz

- Der gleitende Mittelwert mit Fensterbreite „w" ersetzt den aktuellen Wert durch den Durchschnitt der letzten „w" Werte. Dadurch werden kurzfristige „Zacken" geglättet.



· Ein zentrierter gleitender Mittelwert ist für die Darstellung sinnvoll, darf aber nicht als Eingabe für ein Vorhersagemodell verwendet werden.

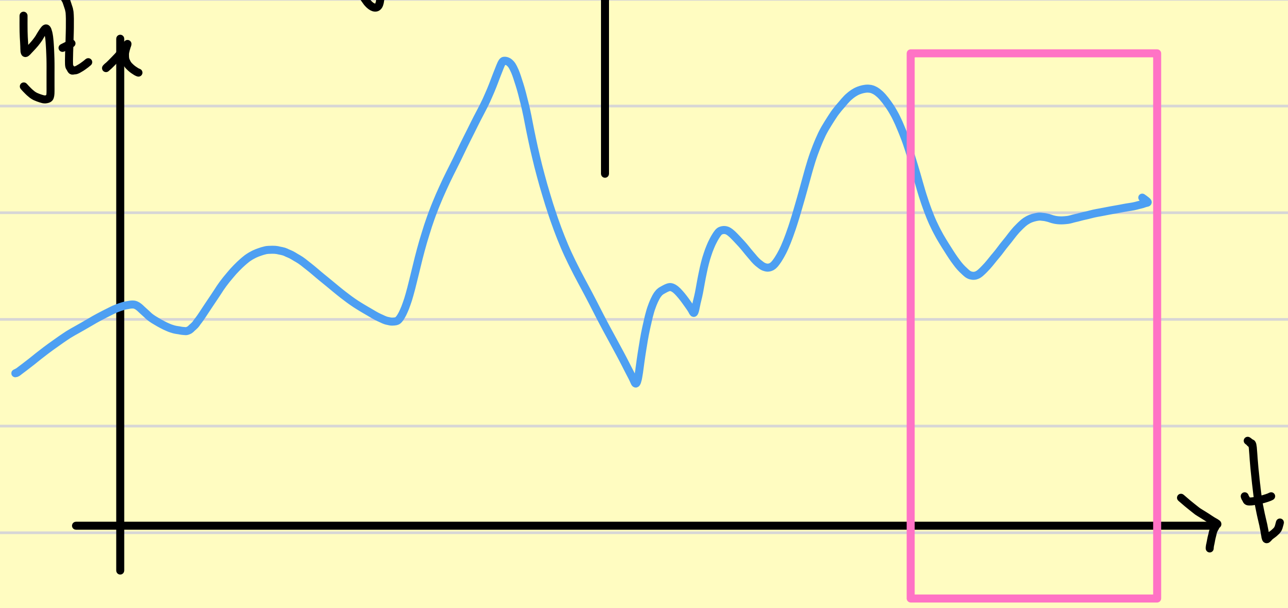
· Beim Umrechnen der Frequenz werden Messwerte in größere Zeitblöcke zusammengefasst, z.B von Tagen auf Wochen.

A6. Saubere Aufteilung in Lern- & Prüfphasen.

· Die zeitliche Reihenfolge bleibt unverändert.

· Die Lernphase besteht aus einem frühen, zusammenhängenden Zeitraum, die Prüfphase aus einem späteren zusammenhängenden Zeitraum.

· Eine robuste Variante ist die s.g. rollierende Ursprungsvalidierung: Dabei wird der Schnittpunkt mehrfach vorwärts geschoben und jedes mal neu gelernt.



A7. Modellfamilien.

· AUTOREGRESSIVES MODELL der Ordnung „p".
Der neue Wert wird als lineare Kombination der letzten „p" Werte und eines Zufallsteils beschrieben.

$$AR(p) \equiv y_t = c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

· MODELL mit GLEITENDEM MITTELWERT der Ordnung „q"
Der neue Wert entsteht als lineare Kombination vergangener Zufallsteile und eines neuen Zufallsteils.

$$MA(q) \equiv y_t = c + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

· AUTOREGRESSIVES INTEGRIERTES MODELL MIT GLEITENDEM MITTELWERT.

Zuerst wird die Zeitreihe eine bestimmte Anzahl von malen differenziert, um Stationarität zu erreichen. Anschließend wird ein autoregessives Modell mit einem Modell des gleitenden Mittelwerts kombiniert.

ARIMA

A8. Messung der Vorhersagegüte.

- MAE : $MAE = \frac{1}{m} \sum |y - \hat{y}|$
- RSME
- MAPE : $\frac{100}{m} \sum \left| \frac{y - \hat{y}}{y} \right|$