Zeitreihen vorhersagen (Statistik II): ARIMA-Modelle

Session 4 (Montag 15:15 – 17:00)

ENABLE AI
MASTER YOUR DATA
Hmetrica

Zeitreihen und ihre Merkmale visualisieren

- Autoregressive- (AR) und Moving Average (MA)
- Mit ARMA und ARIMA-Modellen vorhersagen
- Mit SARIMA-Modellen Saisonalität berücksichtigen

Das AR(p)-Modell “autoregressive”

Beispiel

Was ist der Preis von 1 MWh Strom (Intraday Preis) an der Strombörse in einer Stunde?

Was wäre Ihre Schätzung, wenn ich Ihnen sage, dass er gerade bei 42.6 Euro/MWh liegt?

Naïve Vorhersage

$$\hat{x}_{t+h} = x_t$$

Wie können wir die verbessern?

Ein **autoregressives Modell der Ordnung p** (AR(p)) kann geschrieben werden als

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \epsilon_t$$

wobei ϵ_t weißes Rauschen ist.

Das MA(q)-Modell
“moving average”

Naive Vorhersage II (Durchschnittsmethode)

$$\hat{x}_{t+h} = \frac{1}{t}\sum_{i=1}^{t} x_i$$

Wie können wir die verbessern?

Beispiel

Wie viel Strom wird heute, am 25.2.2021, in Bayern aus Windkraft erzeugt?

Wenn ich Ihnen jetzt sage, dass es im Durchschnitt täglich 12,6 Gwh sind?

Wenn ich zusätzlich sage, dass meine Schätzung gestern +0,6 Gwh daneben lag?

Ein **gleitendes Durchschnittsmodell der Ordnung q** (MA(q)) kann geschrieben werden als

$$x_t = c + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q} + \epsilon_t$$

wobei ϵ_t weißes Rauschen ist.

Das ARMA (p,q)-Modell

Ein **ARMA(p,q)-Modell** (AutoRegressive Moving Average) wird geschrieben als

$$x_t \; = c + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

wobei ϵ_t weißes Rauschen ist.

Das ARIMA (p,d,q)-Modell

Differenzen- und Backshiftoperator

$$\Delta_h x_t = x_t - x_{t-h},$$
$$B x_t = x_{t-1}$$
also: $\Delta_1 x_t = x_t - x_{t-1} = (1-B)x_t$

Ein **ARIMA(p,d,q)-Modell** (AutoRegressive Integrated Moving Average) wird geschrieben als

$$x'_t = c + \phi_1 x'_{t-1} + \cdots + \phi_p x'_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

wobei ϵ_t weißes Rauschen und $x'_t = (1-B)^d x_t$ ist.

Das SARIMA(p,d,q)(P,D,Q)$_m$ Modell

Erweiterung: Saisonales ARIMA-Modell

SARIMA(p,d,q)(P,D,Q)$_m$

- Beispiel SARIMA(1,1,1)(1,1,1)$_4$

$$\underbrace{(1-\phi_1 B)}_{\text{Non-seasonal AR(1)}}\ \underbrace{(1-\Phi_1 B^4)}_{\text{Seasonal AR(1)}}\ \underbrace{(1-B)}_{\text{Non-seasonal difference}}\ \underbrace{(1-B^4)}_{\text{Seasonal difference}}y_t = \underbrace{(1+\theta_1 B)}_{\text{Non-seasonal MA(1)}}\ \underbrace{(1+\Theta_1 B^4)}_{\text{Seasonal MA(1)}}e_t.$$

Beispiel

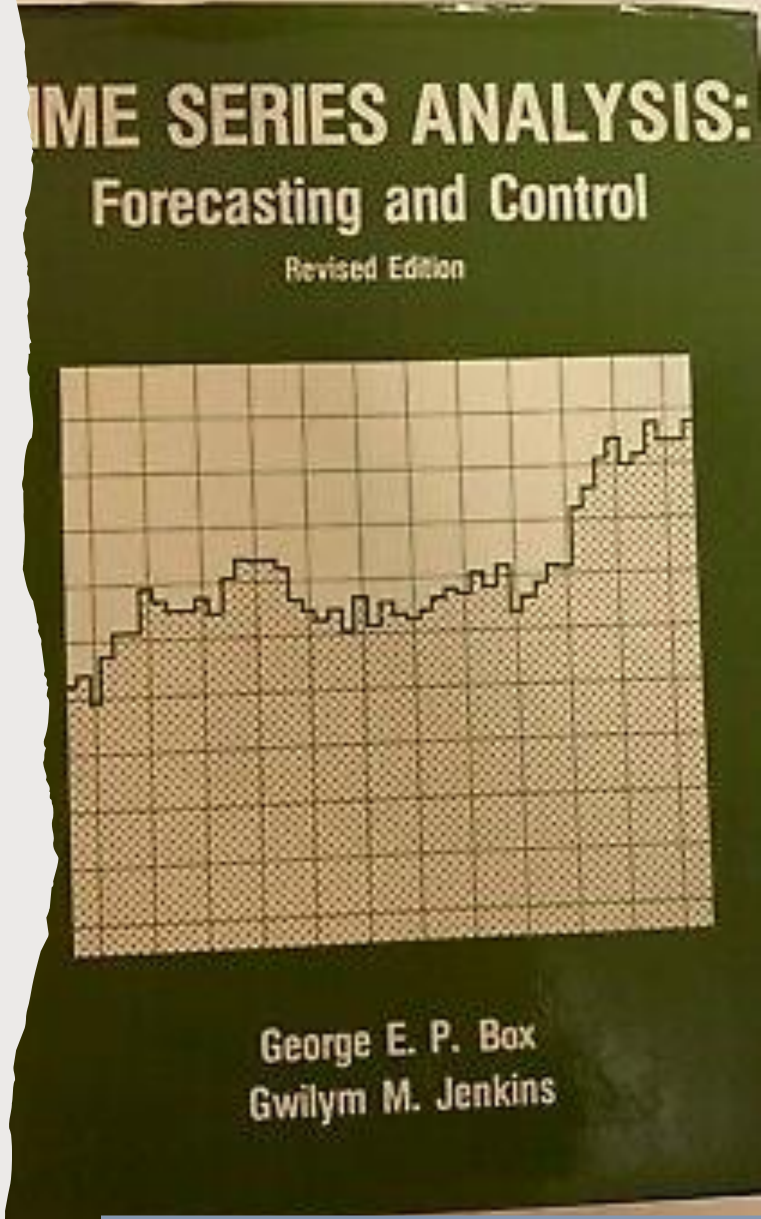
Ich messe die Raumtemperatur in meinem Schlafzimmer jede Stunde. Es ist neun Uhr und ich will die Temperatur für zehn Uhr vorhersagen. Welche Informationen hätten Sie gern von mir?

Image source:Wikipedia, XYZ

Box-Jenkins-Methode: Mit SARIMA Vorhersagen

"Alle Modelle sind falsch, aber einige sind nützlich."

George Edward Pelham Box
(1919 - 2013)
britischer Statistiker



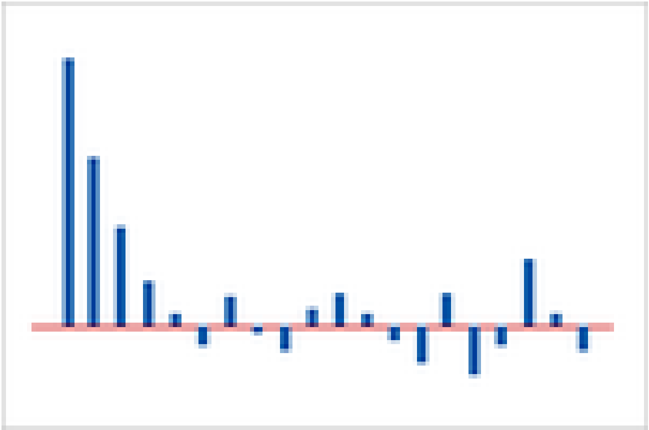
ENABLE AI MASTER YOUR DATA
Hmetrica

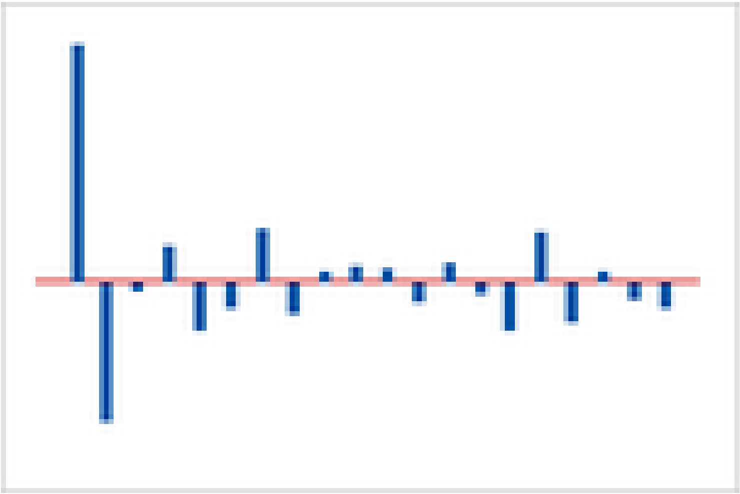
Box-Jenkins-Methode: Mit SARIMA Vorhersagen

1. **Modellidentifikation**
   - Stationär? Sonst d Differenzen bilden
   - Passende p und q wählen (ACF, PACF; AIC**, AICc, BIC**)
2. **Schätzung**
   - Software schätzt $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_{q,} \Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_q$
3. **Validierung**
   - z.B. prüfen, ob die Residuen, also die geschätzten ϵ_t unkorreliert sind und sich wie weißes Rauschen verhalten
4. **Anwendung: Vorhersage**
   - Einschritt-Prognose: Differenzengleichung des geschätzten ARMA-Modells eine Periode in die Zukunft schieben und den bedingten Erwartungswert berechnen.
   - Mehrschritt-Prognosen: dies rekursiv wiederholen

Image source:Wikipedia, XYZ

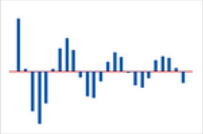
ACF und PACF nutzen





ACF

- Die Autokorrelationsfunktion ist ein Maß für die Korrelation zwischen Beobachtungen einer Zeitreihe, die durch k Zeiteinheiten (yt und yt-k) getrennt sind.

Pattern	What the pattern indicates	Example
Large spike at lag 1 that decreases after a few lags.	An autoregressive term in the data. Use the partial autocorrelation function to determine the order of the autoregressive term.	
Large spike at lag 1 followed by a decreasing wave that alternates between positive and negative correlations.	A higher order autoregressive term in the data. Use the partial autocorrelation function to determine the order of the autoregressive term.	
Significant correlations at the first or second lag, followed by correlations that are not significant.	A moving average term in the data. The number of significant correlations indicates the order of the moving average term.	

https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/autocorrelation/interpret-the-results/autocorrelation-function-acf/

PACF

- Die partielle Autokorrelationsfunktion ist ein Maß für die Korrelation zwischen Beobachtungen einer Zeitreihe, die durch k Zeiteinheiten (yt und yt-k) voneinander getrennt sind, nach Bereinigung um alle anderen Terme mit kürzerer Verzögerung (yt-1, yt-2, ..., yt-k-1).

Pattern	Indicates	Example
Large spike at lag 1 that decreases after a few lags.	A moving average term in the data. Use the autocorrelation function to determine the order of the moving average term.	
Large spike at lag 1 followed by a damped wave that alternates between positive and negative correlations.	A higher order moving average term in the data. Use the autocorrelation function to determine the order of the moving average term.	
Significant correlations at the first or second lag, followed by correlations that are not significant.	An autoregressive term in the data. The number of significant correlations indicate the order of the autoregressive term.	

https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/autocorrelation/interpret-the-results/autocorrelation-function-acf/

CODING

Vorhersage Stromlast im Gebäude

- Vorhersagen generell
- AR, MA, ARMA, ARIMA,
- SARIMA
- Vergleich (Errormaße)

Ab ins Jupyter Notebook

Analyse und Visualisierung von Zeitreihen-Daten in Python

Tag 2

ENABLE AI
MASTER YOUR DATA
Hmetrica

Inhalte – Was haben wir vor?

Tag 1

1. Einführung in Zeitreihendaten in Python
2. Zeitreihen und ihre Merkmale visualisieren
3. Zeitreihen vorhersagen (Statistik I): Exponentielle Glättung und Holt-Winters
4. Zeitreihen vorhersagen (Statistik II): ARIMA-Modelle

Tag 2

5. Einblick in andere Zeitreihenmodelle
6. Machine Learning für Zeitreihen: Überblick, Vorbereitung und Klassifikation
7. Machine Learning für Zeitreihen: Clustering
8. Deep Learning für Zeitreihen (Einblick)

Feedback

Vielen Dank

ENABLE AI
MASTER YOUR DATA

Hmetrica