INTERNATIONALE HOCHSCHULE

Online Conference Algorithms

AVOIDING TEMPORAL CONFOUNDING IN TIMESERIES FORECASTING USING MACHINE LEARNING

Ulrich Kerzel, Felix Wick

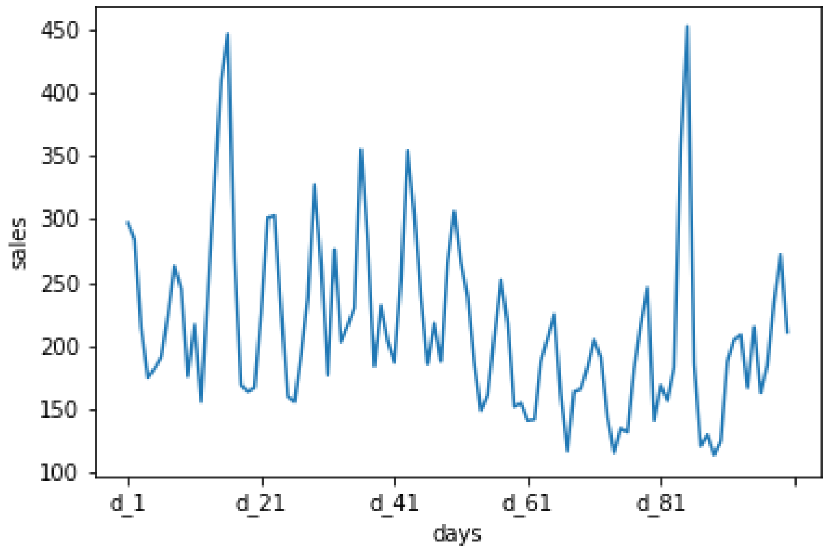
Timeseries

Timeseries play an important role in many operational tasks

Examples:
- Sales of products
- Number of customers per day
- Price of shares
- …

Timeseries Forecasting: Determine future value of a timeseries
- Order new products
- Buy/Sell shares (nor not)
- …

Timeseries for some group of products

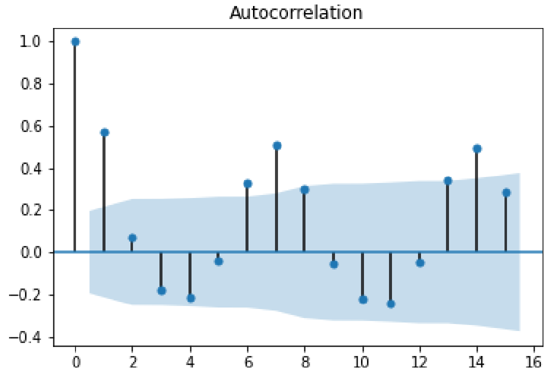


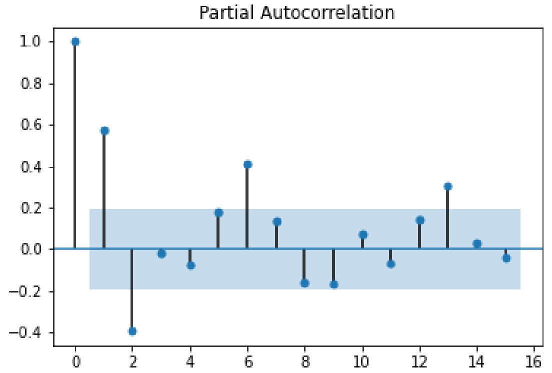
Data: Walmart M5 forecasting challenge (https://www.kaggle.com/c/m5-forecasting-accuracy)

Timeseries Forecasting

„Traditional approach“ --- ARIMA

- Analyze (partial) auto-correlation
- Determine order of AR and MA contribution
- If applicable:
  - Differencing for stationary time-series (ARIMA)
  - Plus: Seasonal components (SARIMA)
  - Plus: External influences (SARIMAX)

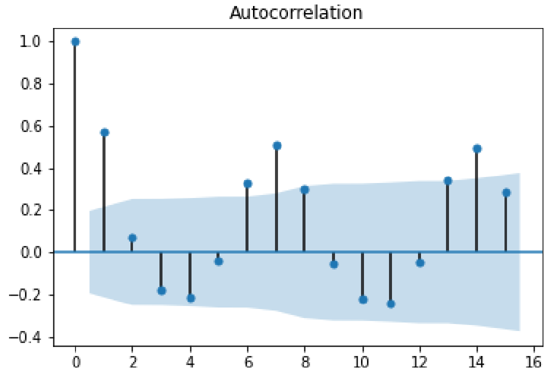




Timeseries Forecasting

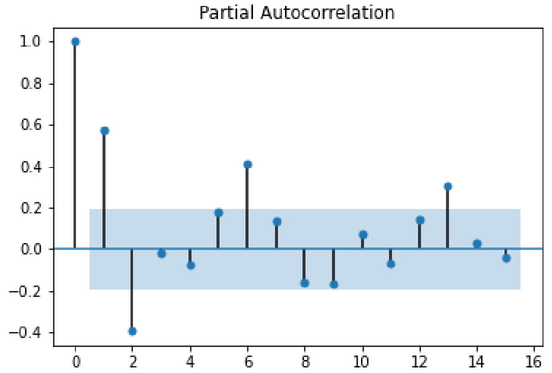
„Traditional approach“ --- ARIMA

- Analyze (partial) auto-correlation
- Determine order of AR and MA contribution
- If applicable: determine additional seasonal lags,…



Key point:

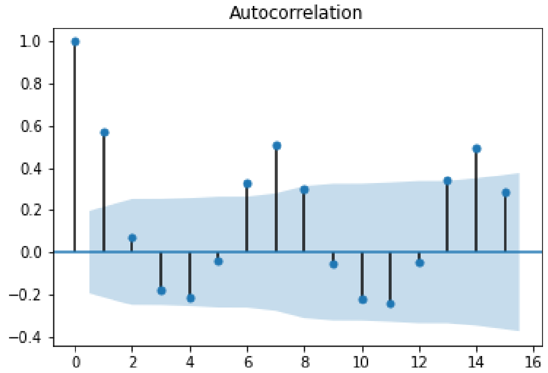
- Exploit auto-correlation between subsequent data-points
- Idea:
  Future behaviour of the timeseries can be learnt from its past behaviour.



Timeseries Forecasting

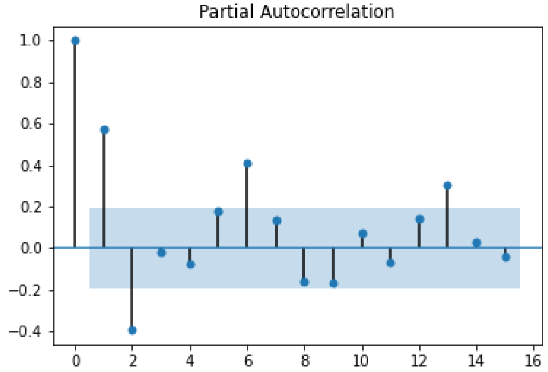
„Traditional approach" --- ARIMA

- Analyze (partial) auto-correlation
- Determine order of AR and MA contribution
- If applicable: determine additional seasonal lags,…



Key point:

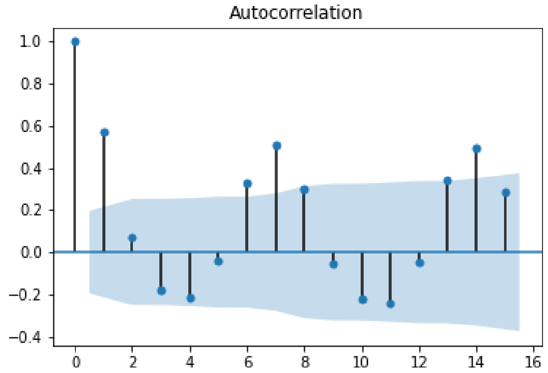
- Exploit auto-correlation between subsequent data-points
- Idea:
  Future behaviour of the timeseries can be learnt from its past behaviour.
- Same underlying concept for other methods:
  - Holt-Winters, exponential smoothing,
  - LSTM
  - Transformer
  - …



Timeseries Forecasting - Challenge

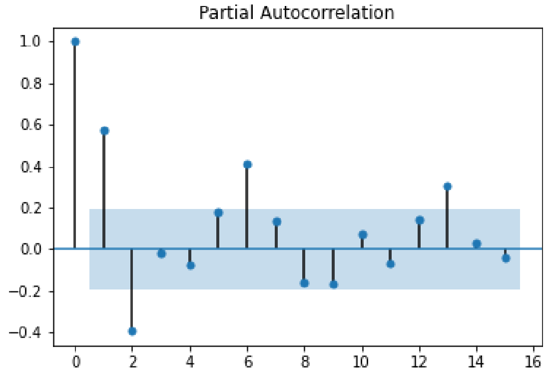
„Traditional approach“ --- ARIMA

- Analyze (partial) auto-correlation
- Determine order of AR and MA contribution
- If applicable: determine additional seasonal lags and model contributions



Key point:

- Exploit auto-correlation between subsequent data-points



Key challenge:

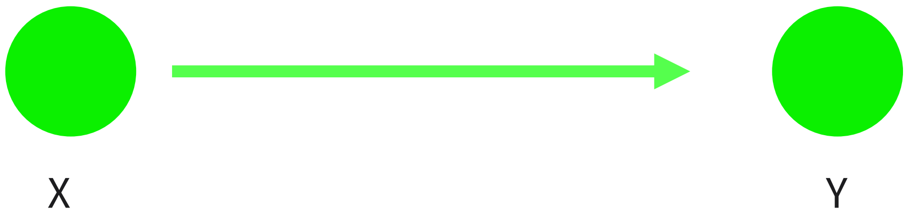
Assumes that the auto-correlation is genuine (not spurious)

Confounding

INTERNATIONALE
HOCHSCHULE

Causal analysis: Variable X has a causal influence on variable Y

- X and Y are correlated (as are Y and X)
- A change in X causes a change in Y --- but not vice versa

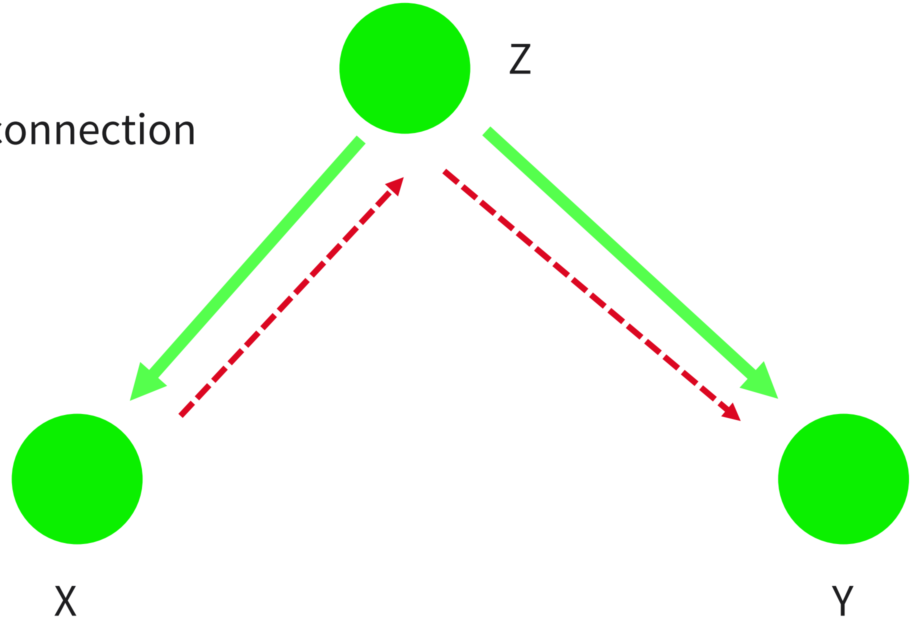


Confounding

Common cause Z

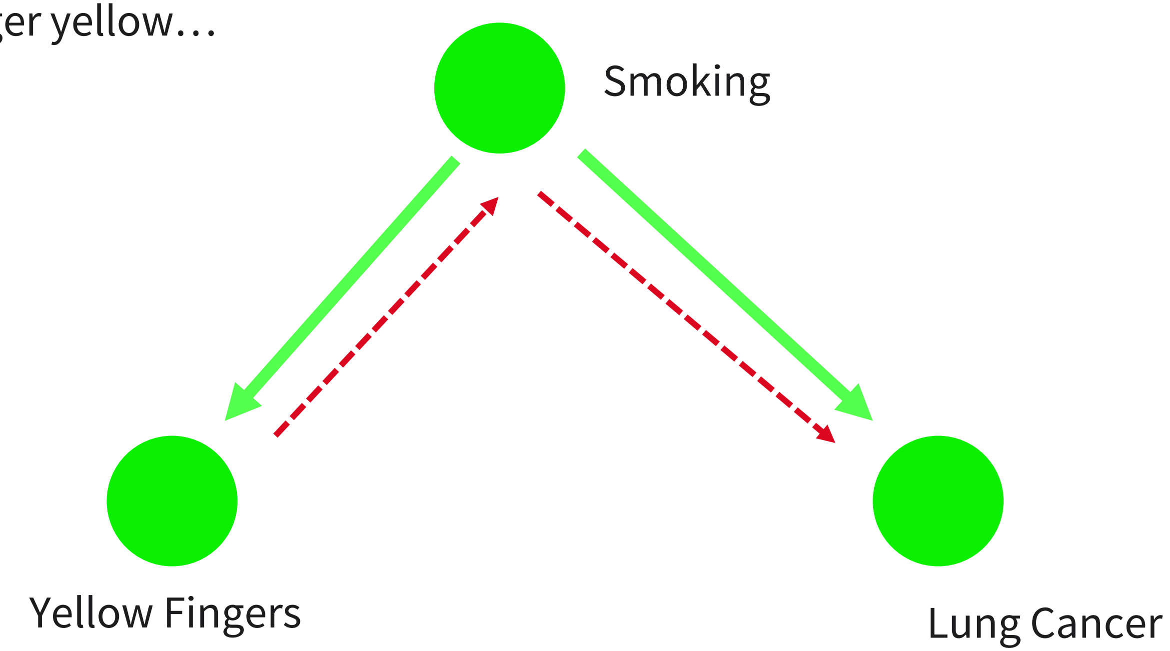
- Influences both X and Y
- Opens a „backdoor path“ X – Z – Y
- X and Y become correlated (spurious correlation)
- **even if** no direct influence / connection between X and Y

- Approach: „control for“ confounder



Confounding - Example

- Smooking causes yellow stains and lung cancer.
- People who smoke often have yellow fingers and a higher risk of lung cancer
  - Lung cancer and yellow stains are correlated
- Yellow fingers do not cause lung cancer (or vice versa)
  - Safe to paint your finger yellow…



Confounding in Timeseries

INTERNATIONALE HOCHSCHULE

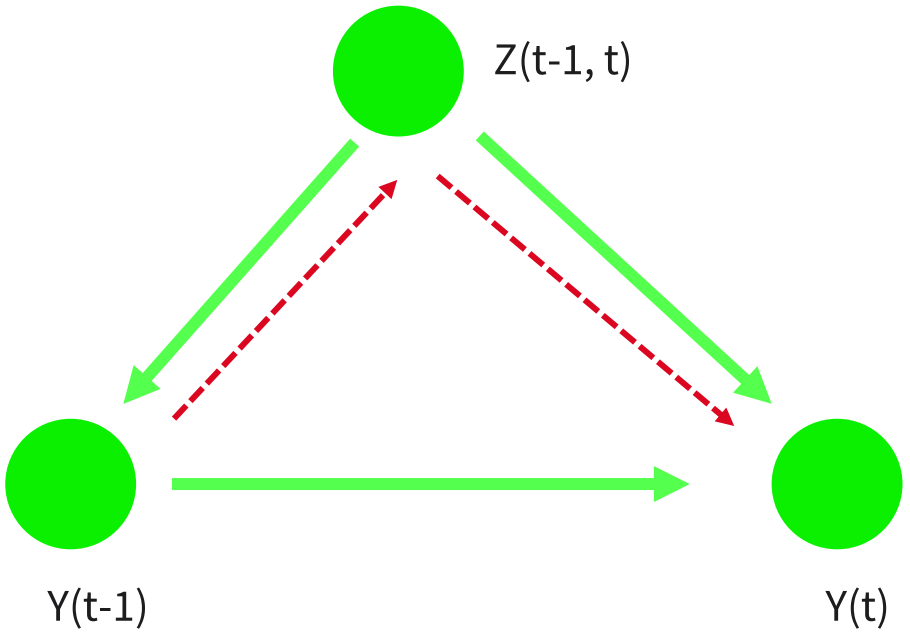
Assumption:

- (Partial) auto-correlation is genuine
- Learning temporal structure of timeseries, lags, … allows to forecast future values

Confounding in Timeseries

Confounder affecting timeseries at two different timeperiods, e.g. at time t-1 and t

- Opens backdoor path
- Even if a genuine correlation exists: spurious correlation now present

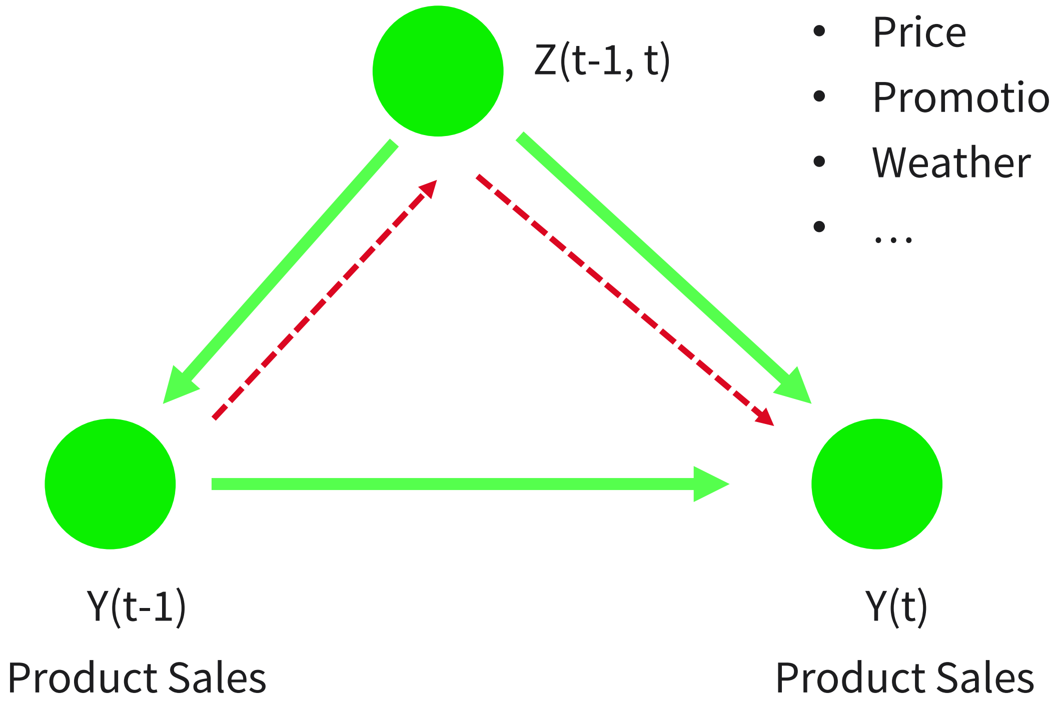


Confounding in Timeseries - Example

Confounder affecting timeseries at two different timeperiods, e.g. at time t-1 and t

— Opens backdoor path

— Even if a genuine correlation exists: spurious correlation now present



Examples for temporal confounders:

- Price
- Promotion / Advertisement
- Weather
- …

Confounding in Timeseries

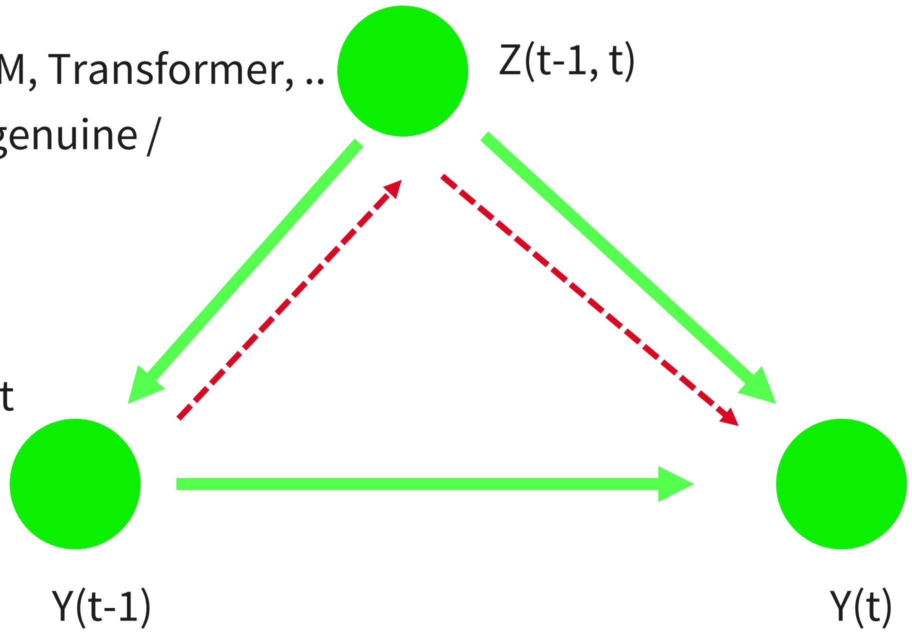
Confounder affecting timeseries at two different timeperiods, e.g. at time t-1 and t

- Opens backdoor path
- Even if a genuine correlation exists: spurious correlation now present

<u>Challenge:</u>

- Methods such as ARIMA, LSTM, Transformer, .. cannot distinguish between genuine / spurious correlation

- Decrease in generalizability: Spurious auto-correlation not an indicator of future behaviour based on past values

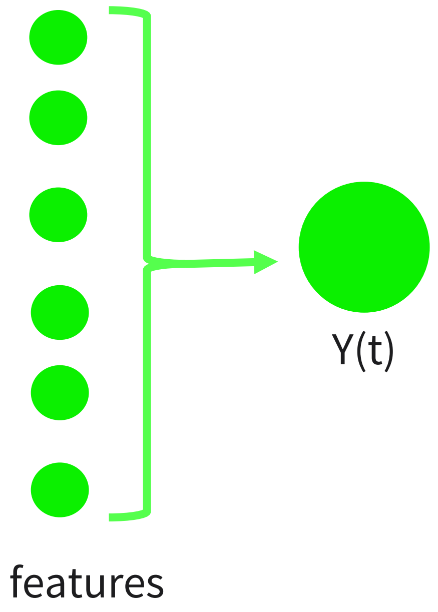


Avoiding Confounding in Timeseries

General idea:

- Treat timeseries forecasting as supervised learning approach
  - Do not rely on auto-correlation

- Learn relationship between „meaningful“ variables
  - „control for“ confounders



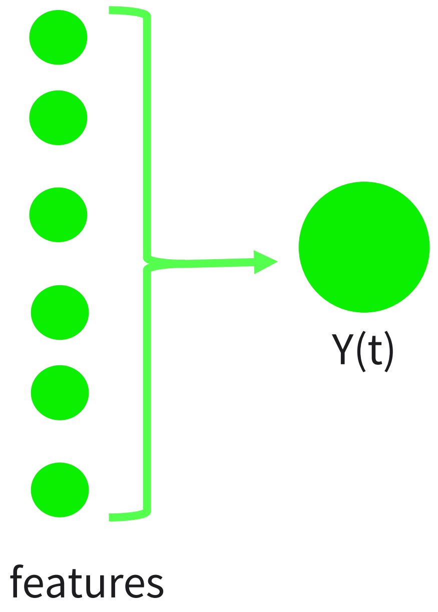
Avoiding Confounding in Timeseries

Choose variables with causal influence:

- Include variables with „meaningful“ causal influence
  - Deep expert domain knowledge needed

- Avoid variables with temporal information
  - Damped / weighted moving average,
  - Reference periods
  - …

- If required:
  include residual correction based on e.g., ratio of weighted moving averages of prediction and past timeseries
  - Can capture any residual “real” auto-correlations
  - But: may open up backdoor path again – need to investigate thoroughly if all variables with causal meaning have been captured.

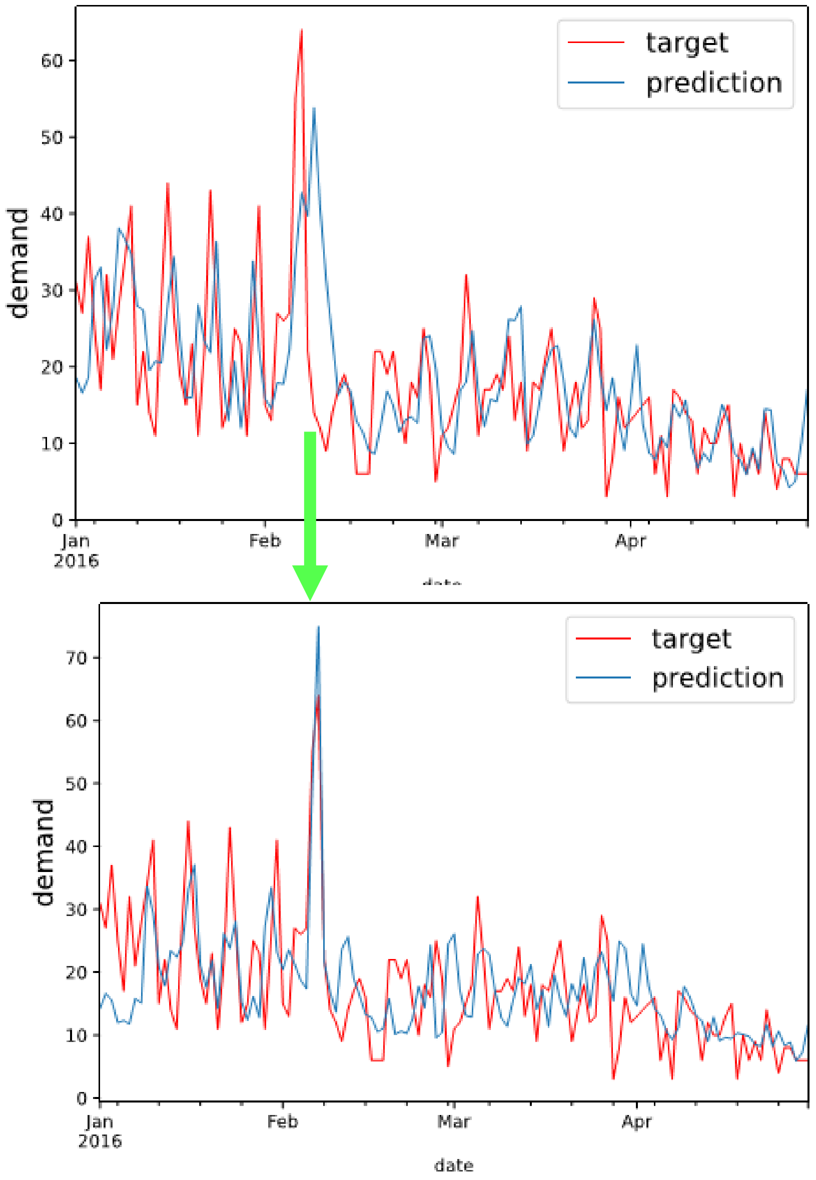


Example:
Demand Forecasting

Features used:

Product ID, store ID, special events, sales price and derived variables

- Top: „traditional“ approach:
  - include time-series features, e.g. eponentially weighed moving average as features
- Bottom: only use features with causal meaning
  - Much better prediction of target behaviour
  - Particularly noticeable for „difficult“ events



Data: Walmart M5 forecasting challenge (https://www.kaggle.com/c/m5-forecasting-accuracy),

Summary

INTERNATIONALE HOCHSCHULE

- Timeseries (TS) forecasting remains crucial for many applications

- „Traditional“ approaches exploit auto-correlation of events:
  sequence of events carries forecasting power --- „defining characteristic“
  - ARIMA, Holt-Winters, exponential smoothing, …
  - Recurrent neural networks, LSTM, Transformer, …

- Assumption: auto-correlation is genuine
  - Temporal confounders: external factor influencing TS at multiple times
  - Opens backdoor path --- spurious auto-correlation
  - TS forecasts lose generalizeability as non-causal influences are picked up.

- Approach:
  Use supervised machine learning and avoid features exploiting temporal sequences.