STL decomposition for outlier detection

Outliers

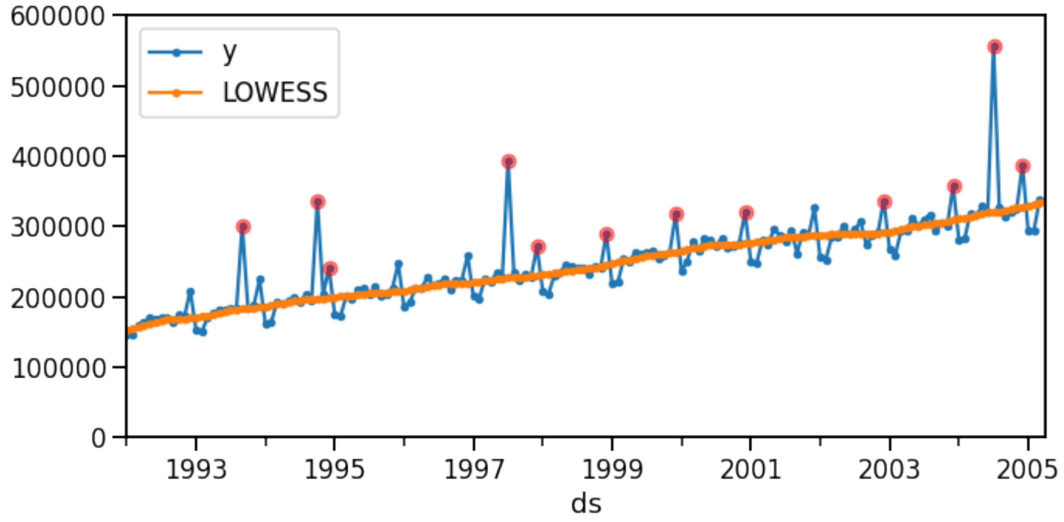
Contents

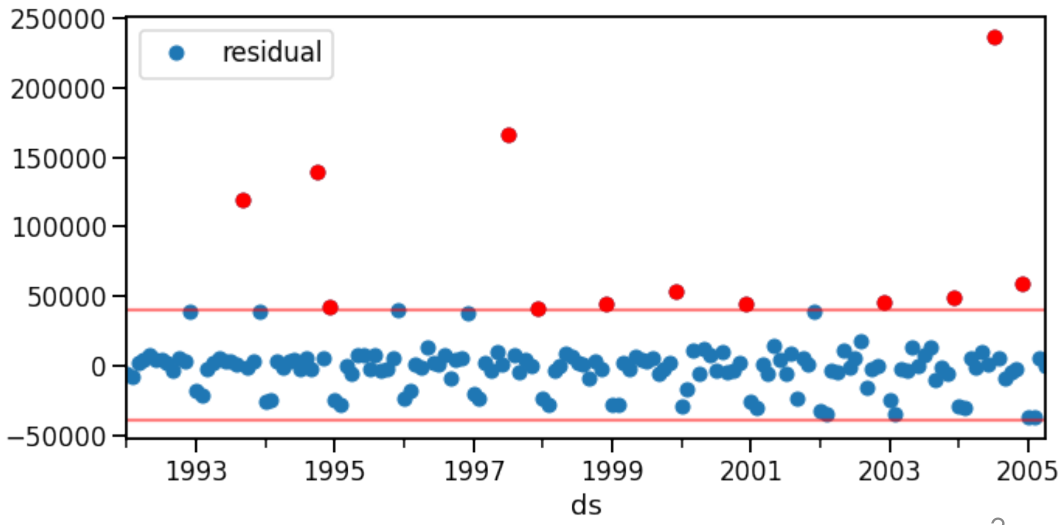
STL DECOMPOSITION

USE RESIDUALS TO IDENTIFY OUTLIERS

Seasonality has been problematic so far

- Seasonal spikes can be incorrectly detected as outliers
- So far recommendation is to de-seasonalise the data first
- Is there a method which can handle seasonality directly?
- Yes! Using STL decomposition!

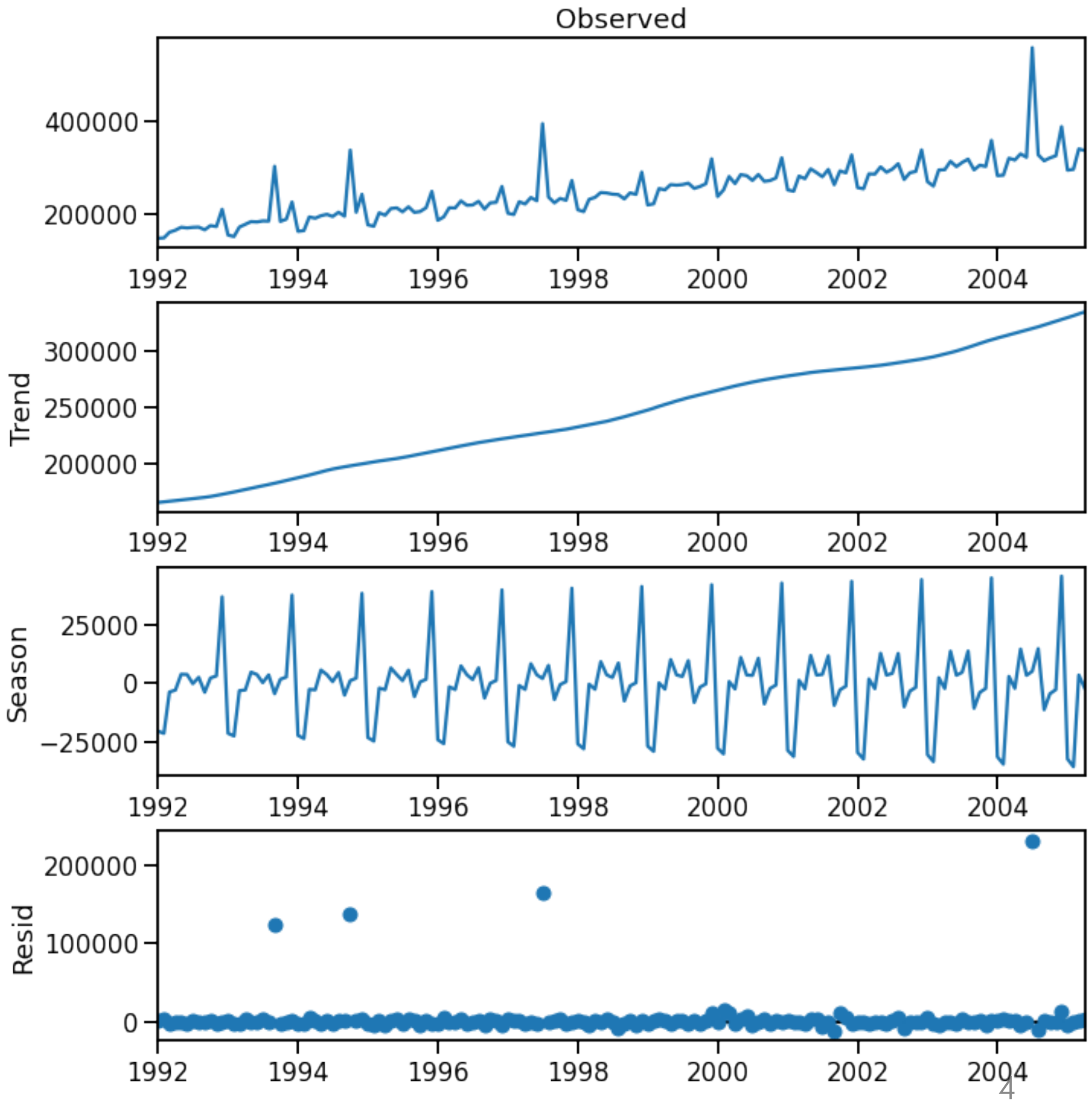




STL recap

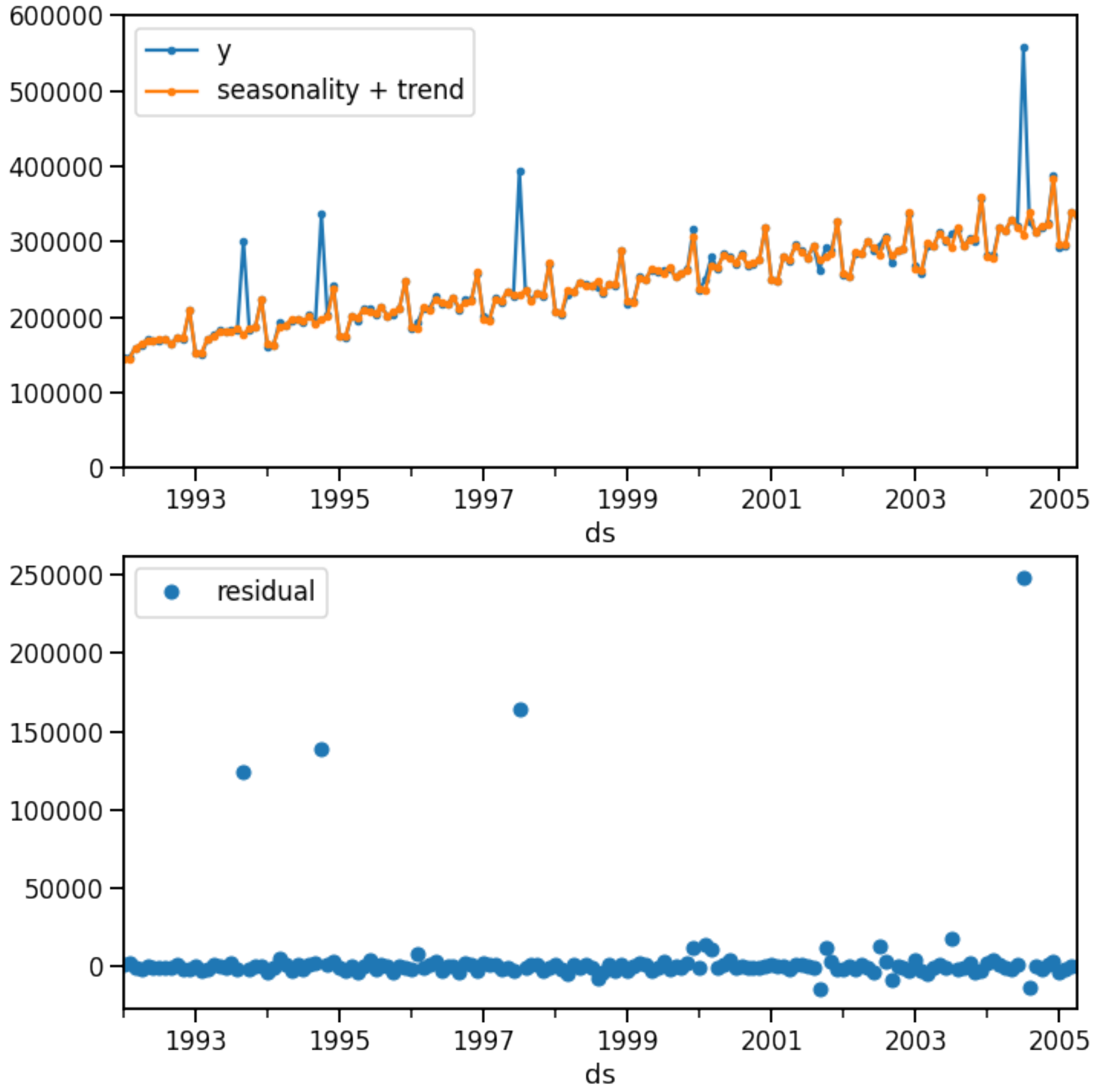
- **S**easonal and **T**rend decomposition using **L**owess
- y = trend + seasonality + residual
- STL extracts trend, seasonal, and residual component
- Can use the residual component from STL which is the same as:

$$e_t = y_t - \ trend_t - seasonality_t$$



Consider residuals from STL decomposition

$$e_t = y_t - \hat{y}_t$$
$$\hat{y}_t = seasonality_t - trend_t$$

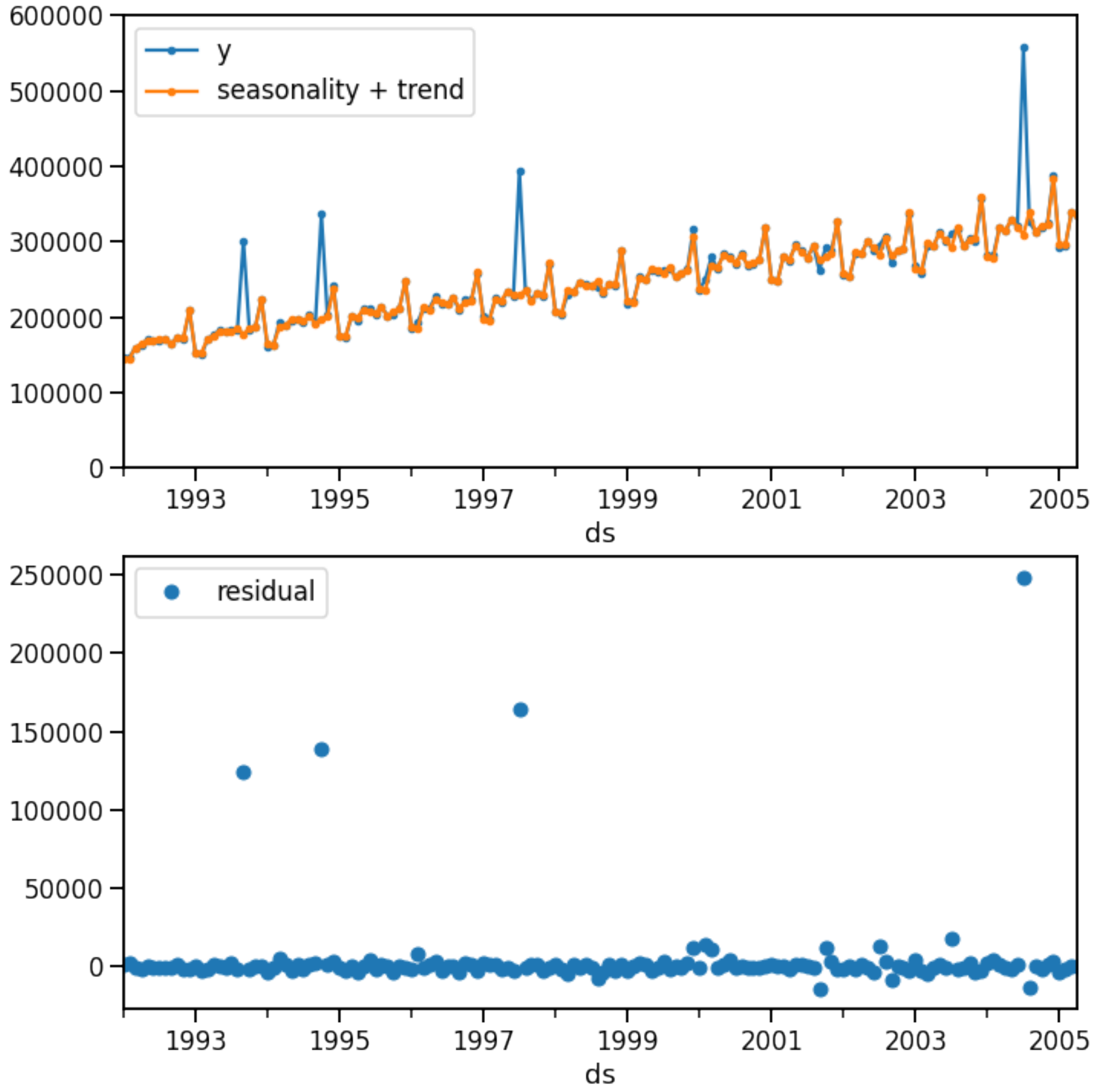


- The residuals look stationary

- Determine outliers using IQR:

$$e_t > \delta_{upper} = Q3 + \alpha \times IQR$$
$$e_t < \delta_{lower} = Q1 - \alpha \times IQR$$

- We set $\alpha = 3$ so that only more extreme outliers are detected



- The residuals look stationary

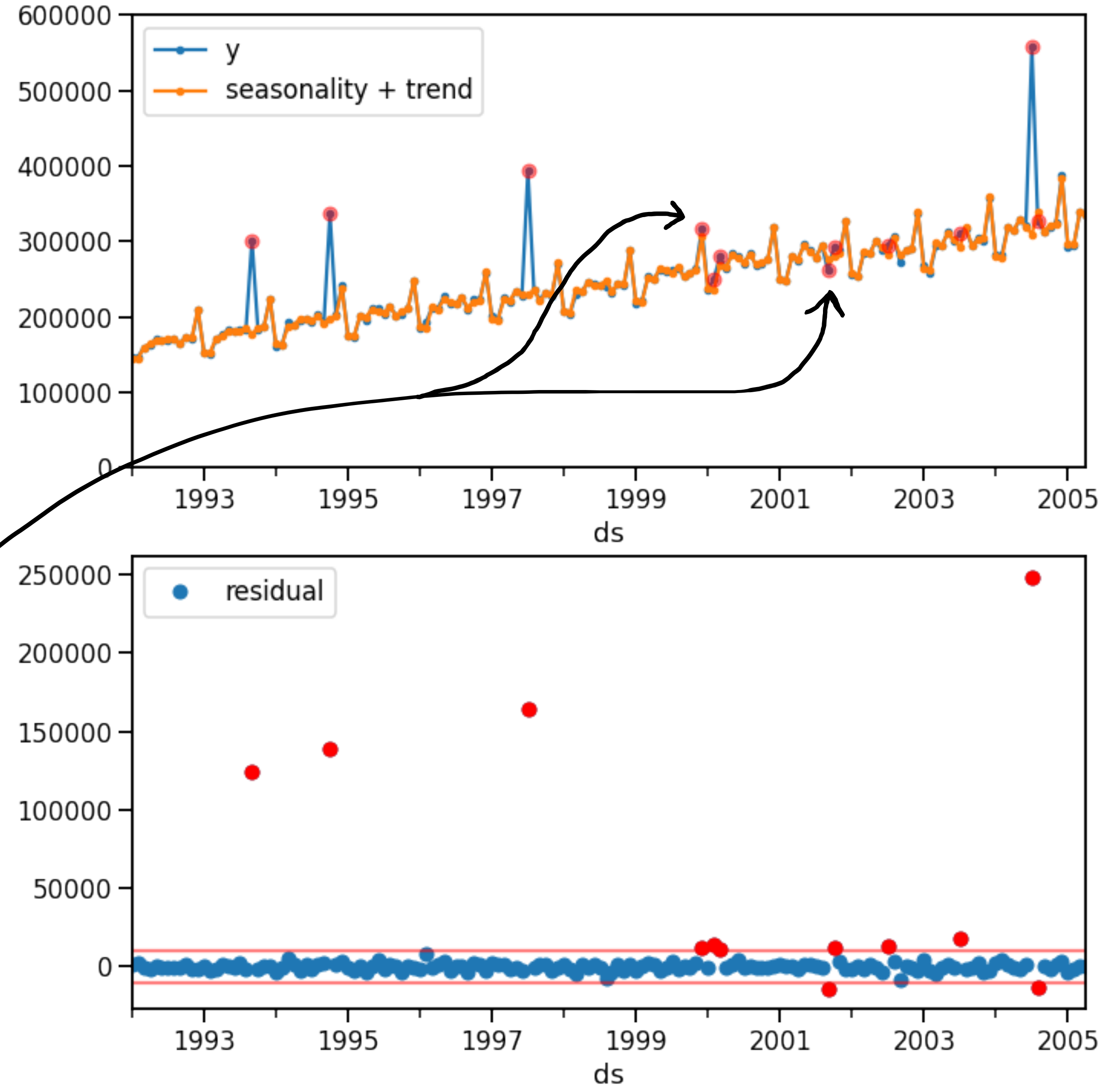
- Determine outliers using IQR:

$$e_t > \delta_{upper} = Q3 + \alpha \times IQR$$
$$e_t < \delta_{lower} = Q1 - \alpha \times IQR$$

- We set $\alpha = 3$ so that only more extreme outliers are detected

- Other points which deviate from expected value, albeit by less

- A larger threshold is a potential solution



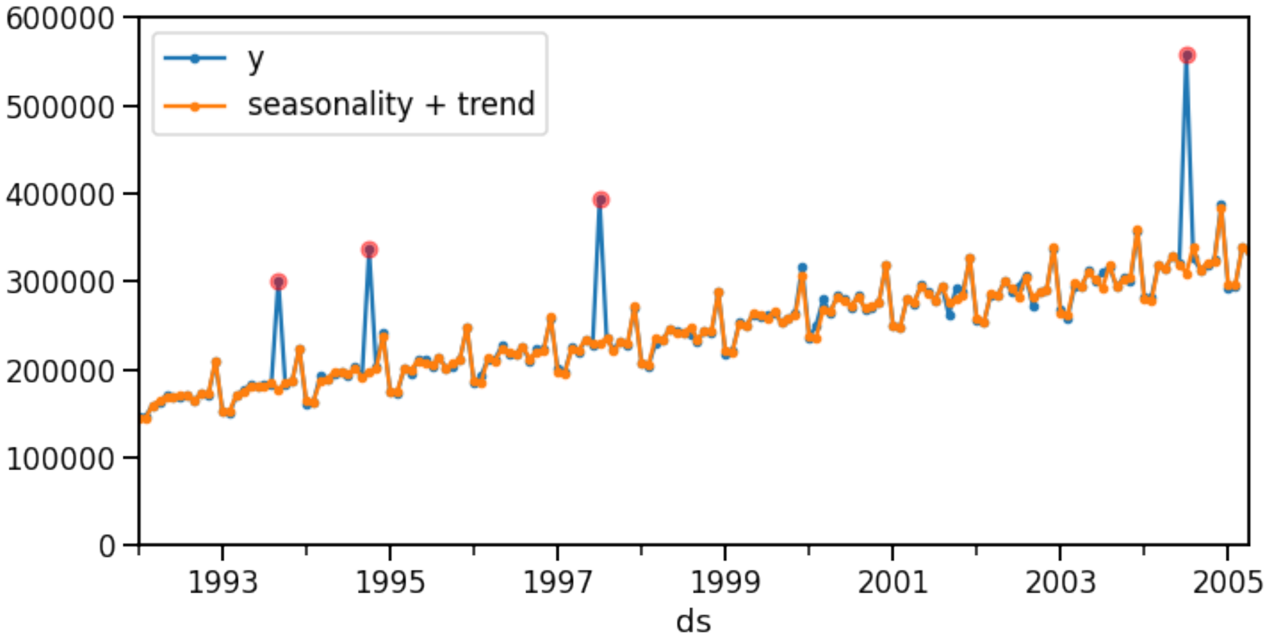
- The residuals look stationary

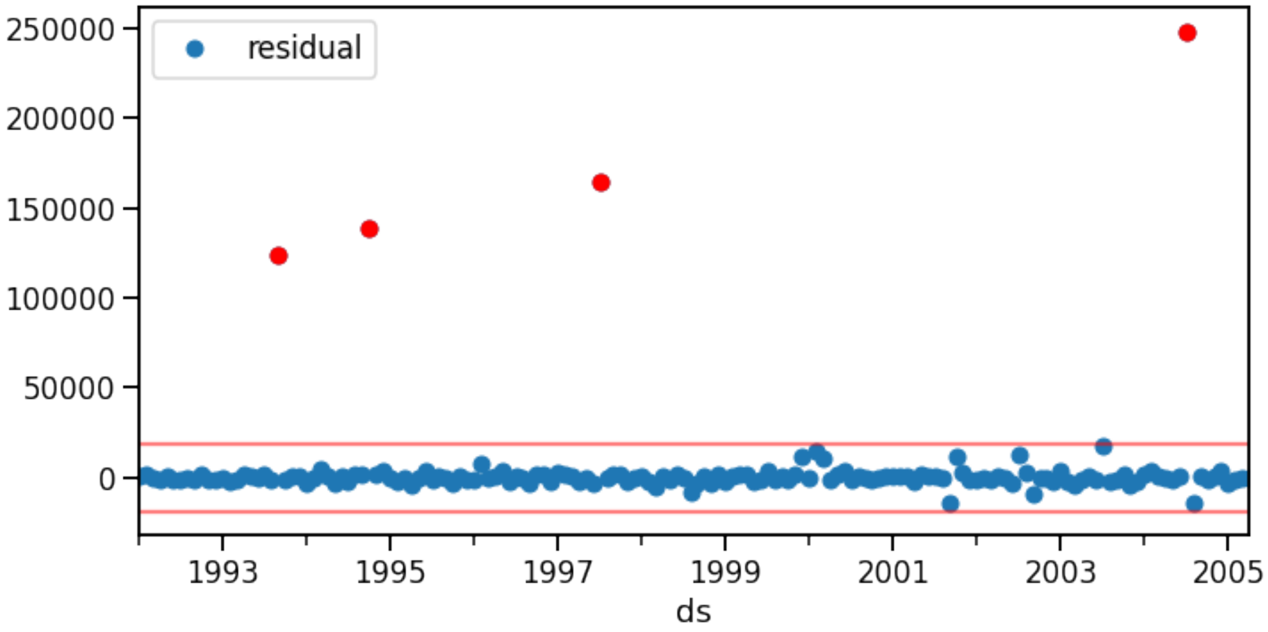
- Determine outliers using IQR:
$$e_t > \delta_{upper} = Q3 + \alpha \times IQR$$
$$e_t < \delta_{lower} = Q1 - \alpha \times IQR$$

- We set $\alpha > 3$ so that only more extreme outliers are detected

- Other points which deviate from expected value, albeit by less

- A larger threshold is a potential solution





STL - summary

- Parameters:
  - STL parameters
    - Seasonal
    - Period
  - Threshold parameter α

- Pros:
  - Robust to outliers
  - No missing data at edges
  - Captures rapid changes in the trend
  - Handles seasonality

- Cons:
  - Computationally more intensive

Summary

STL can extract seasonality and trend. This can be used to compute an expected value for a time series

The residuals can be used to identify outliers

Thresholds still need to be assessed and set depending on the data and extremity of the outliers