

# Residuals

---

Outliers

# Contents



INTRODUCE IDEA OF  
RESIDUALS



HOW TO IDENTIFY  
OUTLIERS FROM RESIDUALS

# Residuals

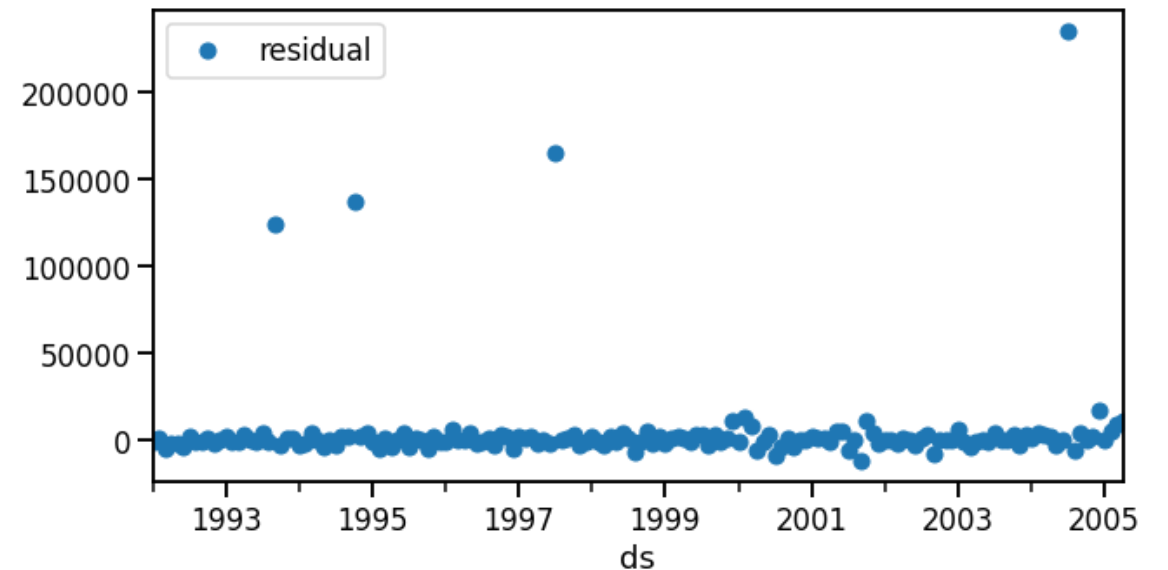
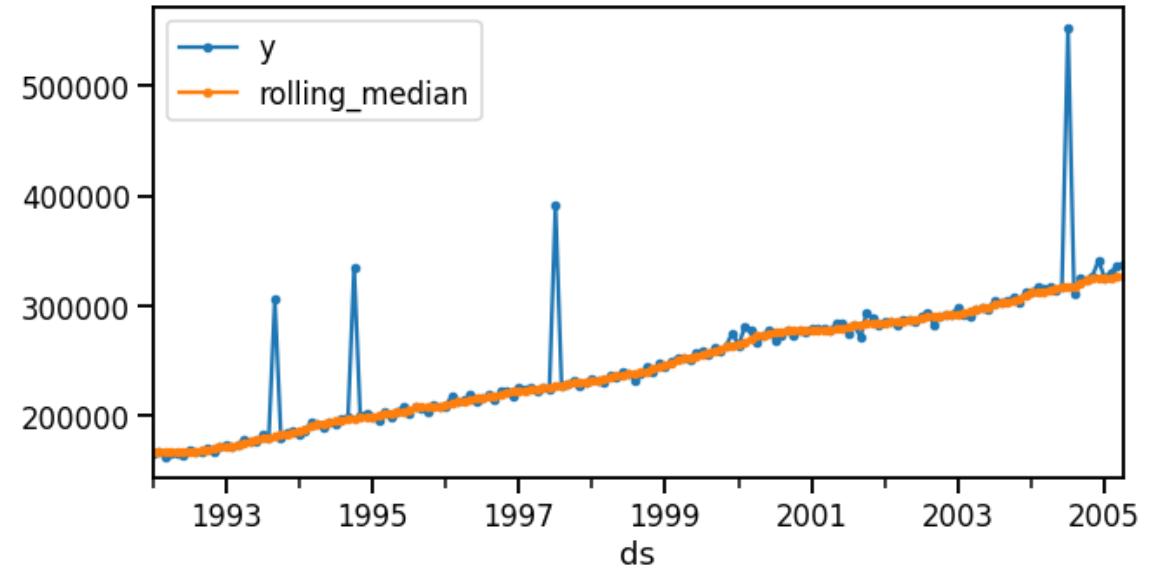
- Residuals are difference between observation  $y_t$  and the expected or forecasted value  $\hat{y}_t$ :

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

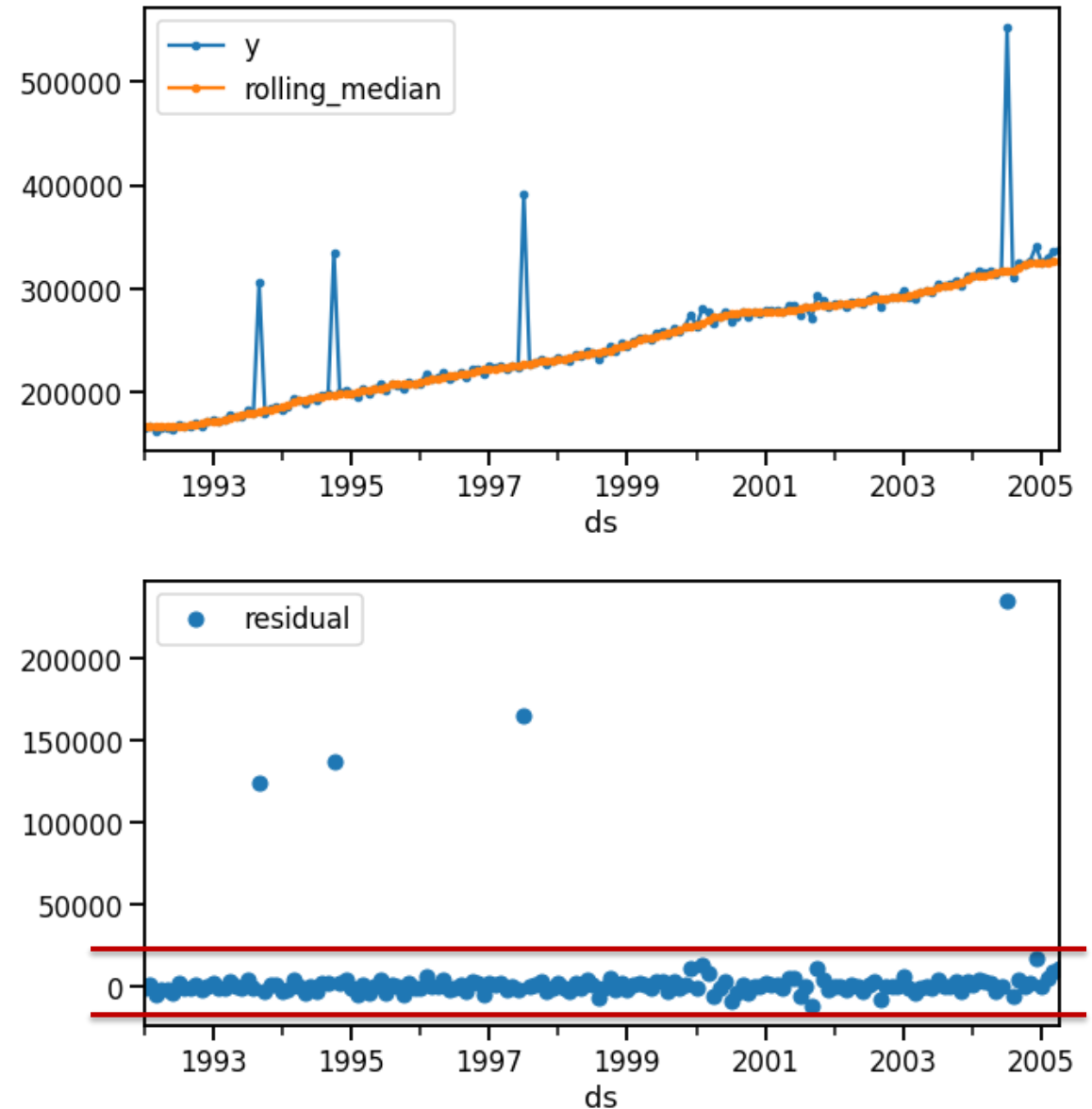
- Residuals can be used to diagnose the performance of a forecast (out of scope for this lecture)
- We shall focus on how to use them to identify outliers

# Consider residuals from the rolling median

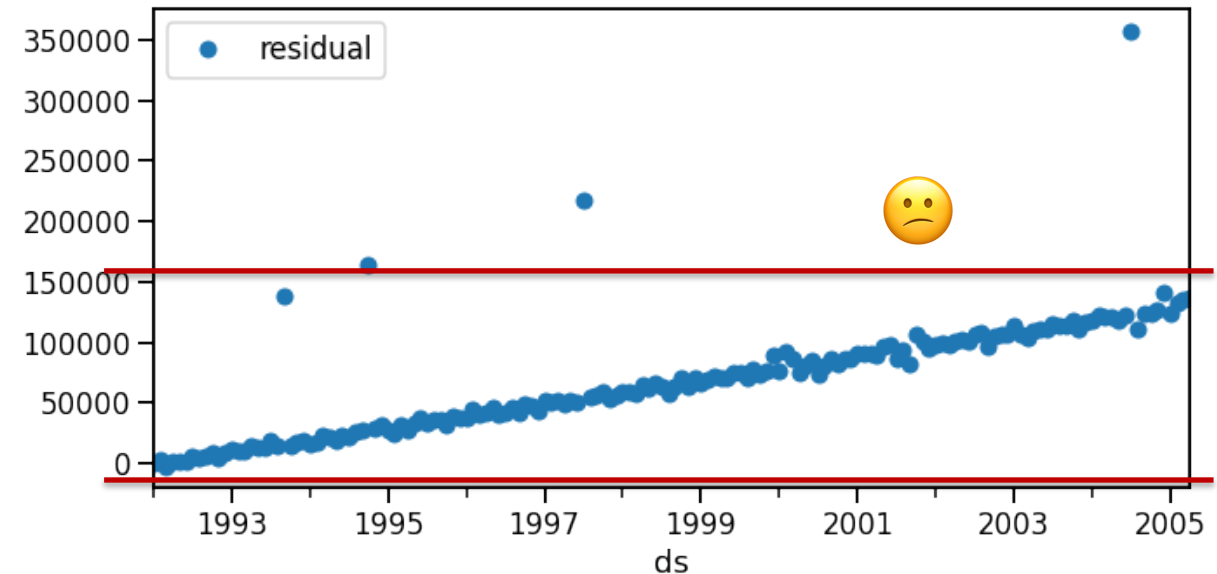
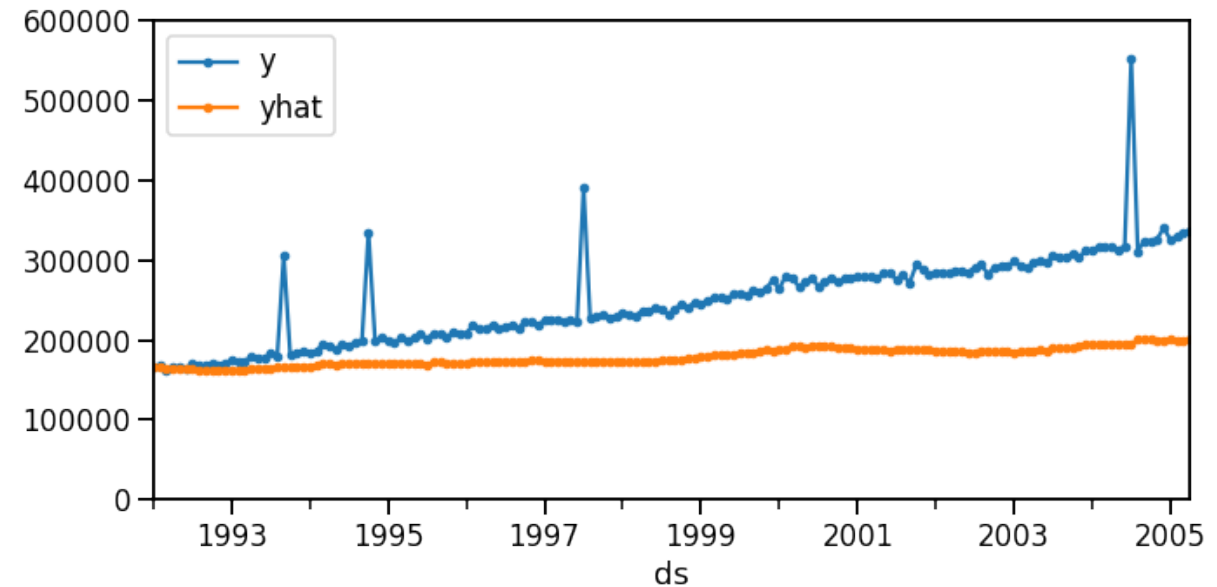
$$e_t = y_t - \hat{y}_t$$
$$\hat{y}_t = \text{median}(y_{t-T}, \dots, y_{t-1}, y_t, y_{t+1}, \dots, y_{t+T})$$



- The outliers are visually discernible in the residuals
- **Intuition:** Can we just take a threshold?  $e_t = |y_t - \hat{y}_t| > \delta$
- Yes! If the residuals are stationary (i.e., mean and variance do not change with time)
- If residuals not stationary, a simple threshold would not be effective

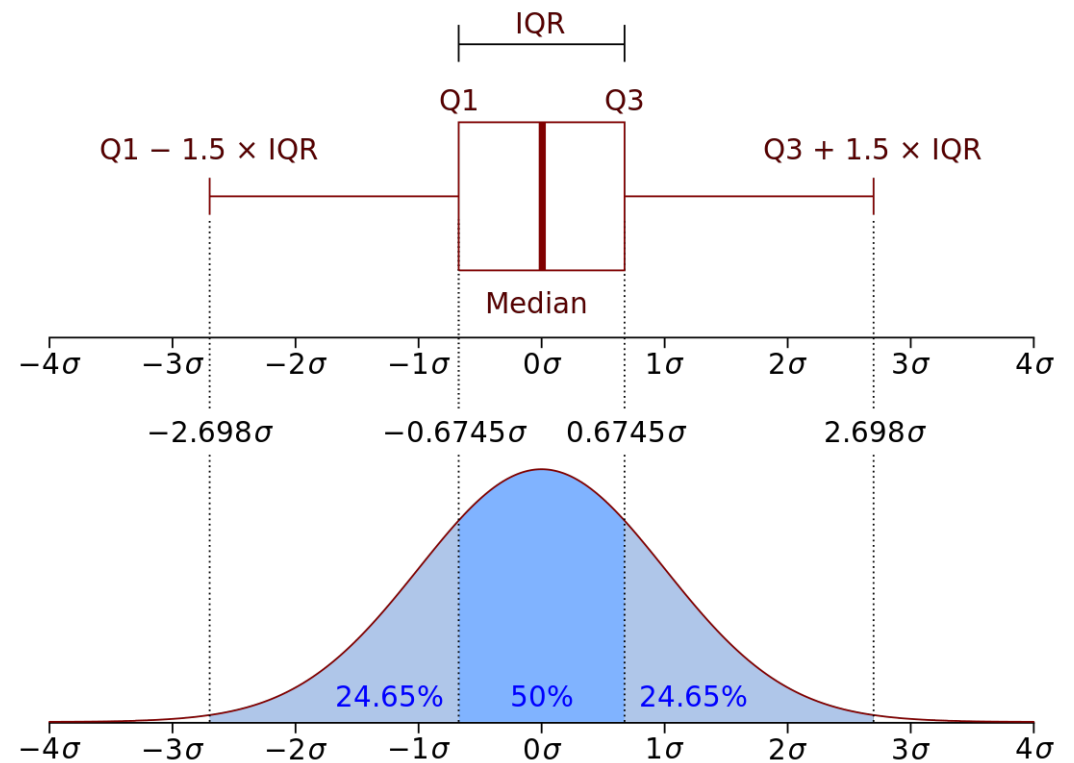
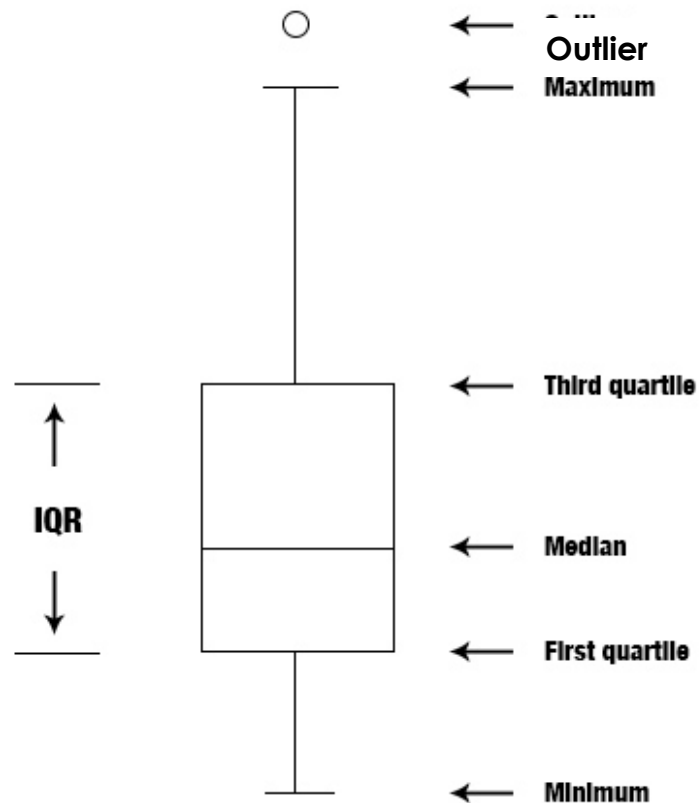


- The outliers are visually discernible in the residuals
- **Intuition:** Can we just take a threshold?  $e_t = |y_t - \hat{y}_t| > \delta$
- Yes! If the residuals are stationary (i.e., mean and variance do not change with time)
- If residuals not stationary, a simple threshold would not be effective
- So, how to pick a threshold?



# Recap: 1.5 x IQR rule

- A value is an outlier if it lies outside  $1.5 \times \text{IQR}$  from the 1<sup>st</sup> or 3<sup>rd</sup> quartile



# Using residuals to identify outliers

1. Compute  $\hat{y}_t$  (e.g., rolling median)
2. Compute residuals  $e_t = y_t - \hat{y}_t$
3. Compute Q1, Q3, and IQR of residuals
4. Flag as outlier any points which satisfy:

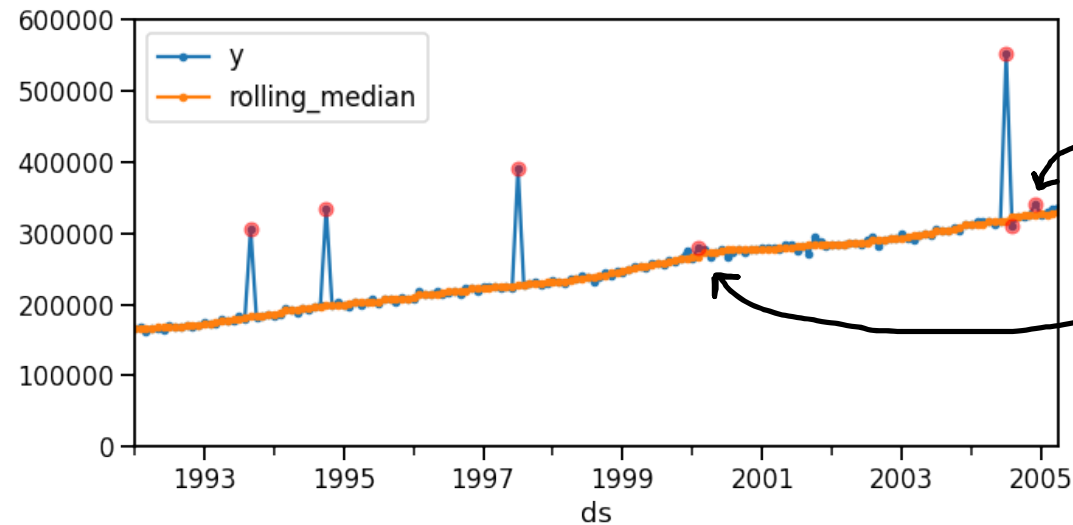
$$e_t > \delta_{upper} = Q3 + \alpha \times IQR$$

$$e_t < \delta_{lower} = Q1 - \alpha \times IQR$$

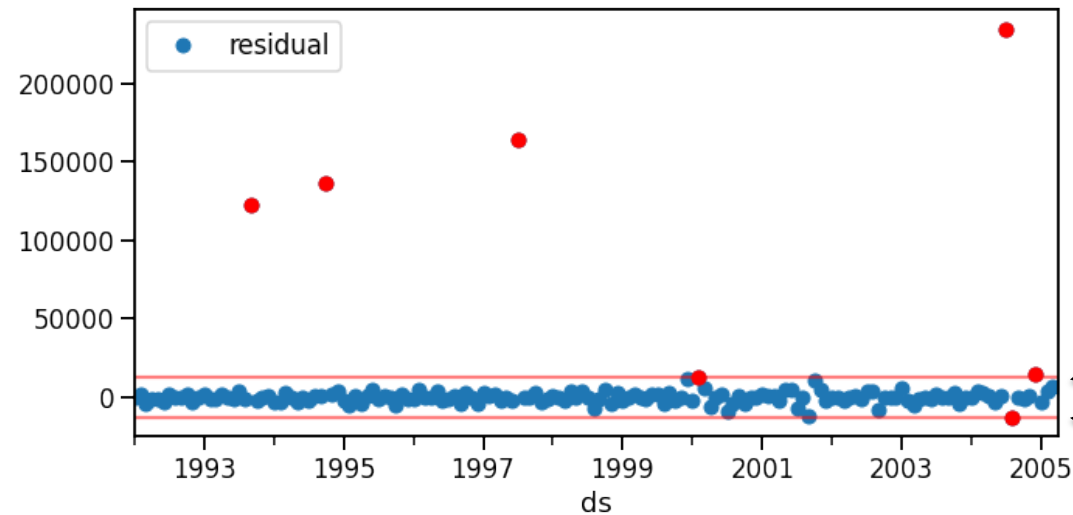
where  $\alpha = 1.5$ , however can be adjusted (e.g.,  $\alpha = 3$  to pick extreme outliers)



# Using residuals to identify outliers



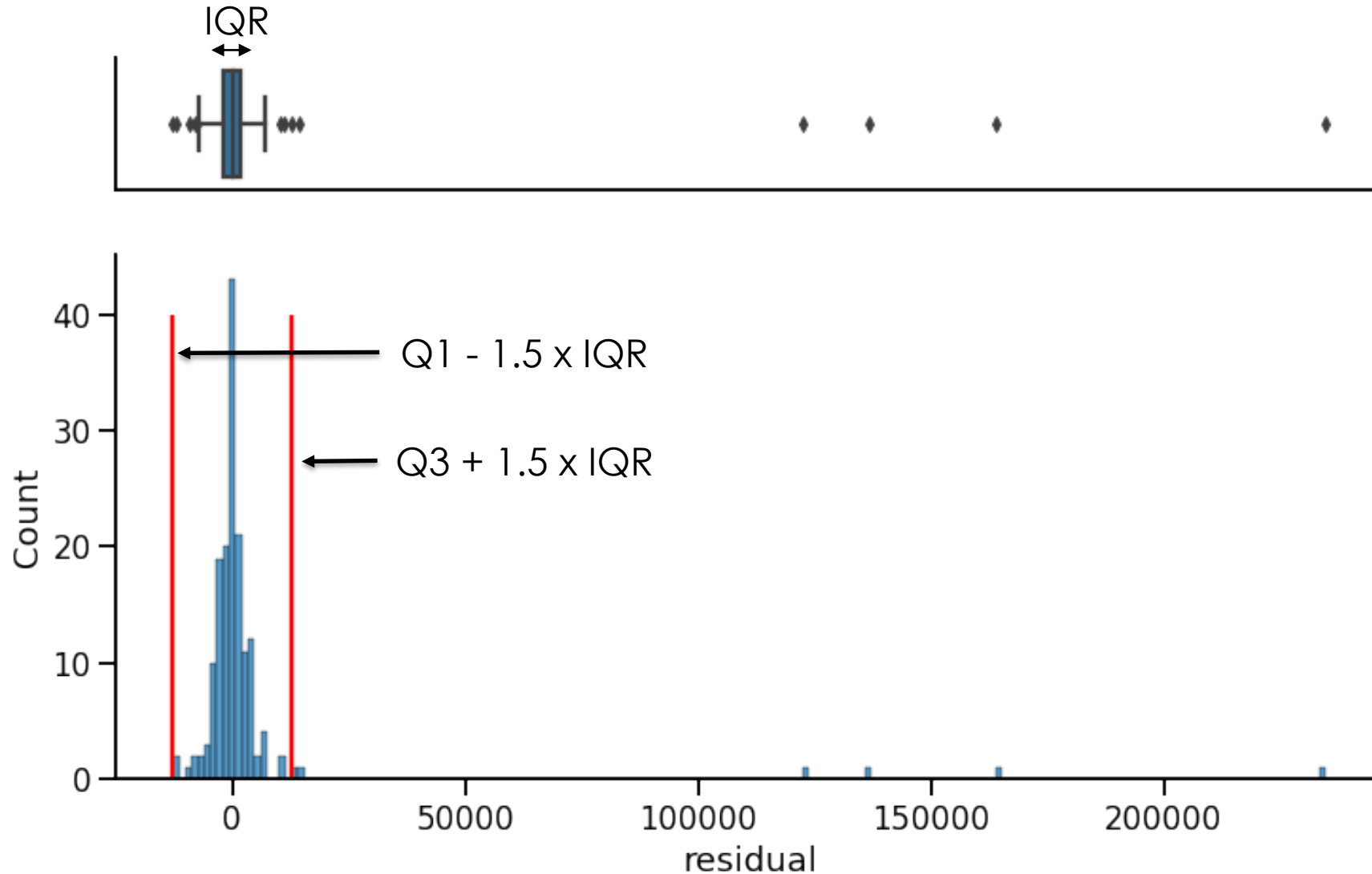
These data points did deviate more from the median. Consider increasing threshold not to capture them.



$Q3 + 1.5 \times IQR$

$Q1 - 1.5 \times IQR$

# Using residuals to identify outliers



# Comments

- Pros
  - Applicable to any estimation method
  - Threshold based on entire dataset, not just a window
- Cons
  - Require stationary residuals
  - Need to examine residuals for stationarity
  - Still need to adjust threshold for desired sensitivity

# Summary

Stationary residuals can be used to identify outliers

Can use non time-series methods such as the IQR rule

Applicable to any estimation method