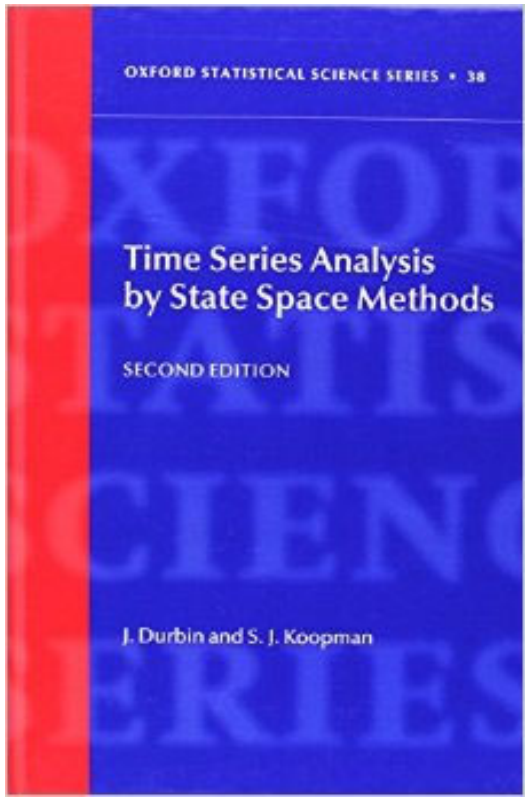
# Detection of outliers and structural breaks using standardized residuals

See the chapter on outlier and structural break detection in the HWS (MARSS User Guide)

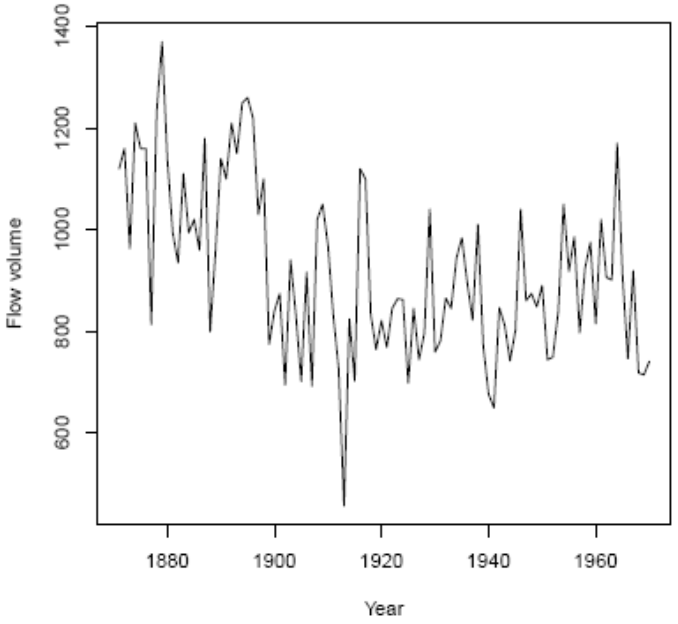
de Jong, P. and Penzer, J. 1998. Diagnosing shocks in time series. Journal of the American Statistical Association 93:796-806.

Durbin and Koopman. 2012. Time series analysis by state-space methods. Chapter 2, Section 12

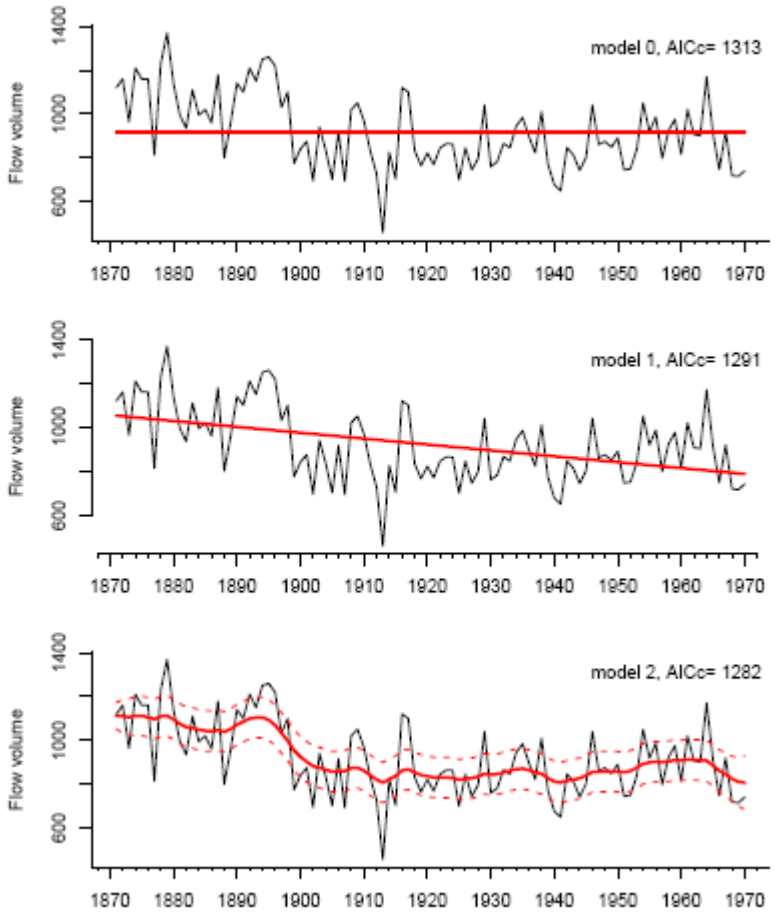


# Back to the Nile River data

River flow by year




Three different models



# Observation outlier detection

Observation outlier: observation (data) at time t is different than what you would expect given the model.

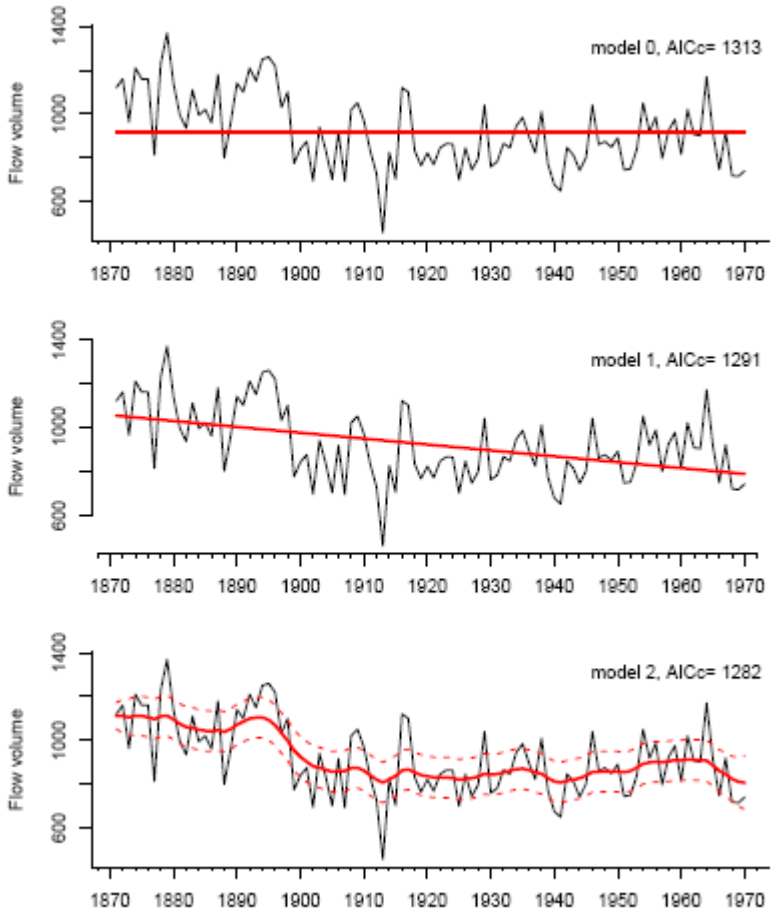
obs. residual = data – fitted value

$$\hat{v}_t = y_t - \hat{y}_{t|T}$$

$$e_t = \frac{1}{\sqrt{\mathrm{var}(\hat{v}_t)}}\hat{v}_t$$

we standardize by the estimated variance and get a t-distributed standardized residual

Three different models




*This idea hinges on v(t) being normal so that means it hinges on the model being able to fit the data (= put a line through the data)*

# Observation residual in the context of state-space models

$$\hat{v}_t = y_t - \hat{y}_{t|T}$$

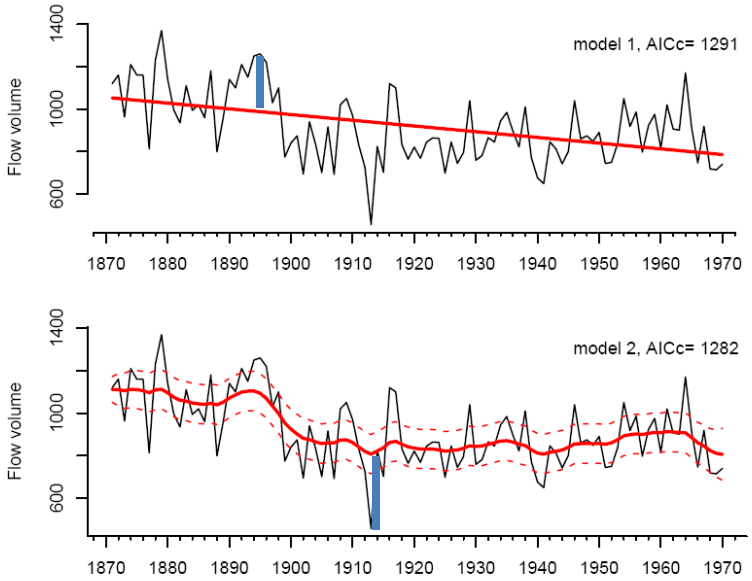
$$e_t = \frac{1}{\sqrt{\mathrm{var}(\hat{v}_t)}}\hat{v}_t$$

obs. residual = data – fitted value

$$\hat{y}_{t|T} = \hat{Z}\tilde{x}_{t|T} + \hat{a}$$

you need to standardize by the variance of that, which is a bit hairy but algorithms for computing it are worked out.

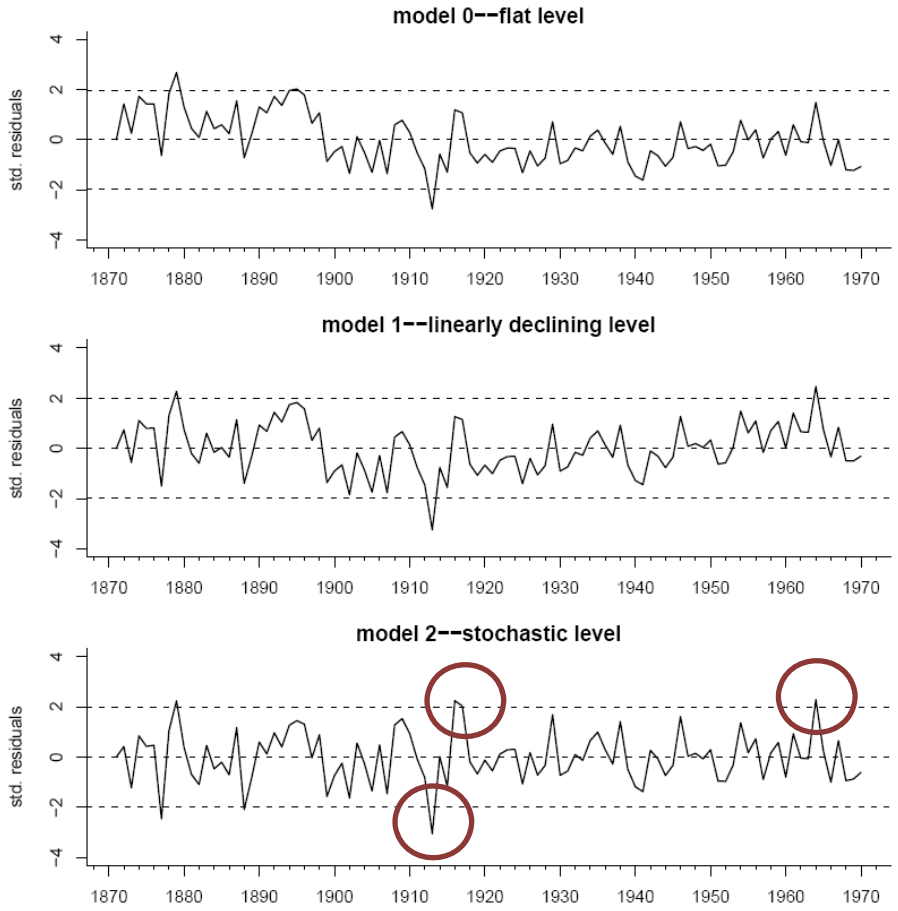
for a linear regression, 'fitted y' is easy.




for a state-space model, there isn't one 'fitted y'. 'fitted y' has a distribution.

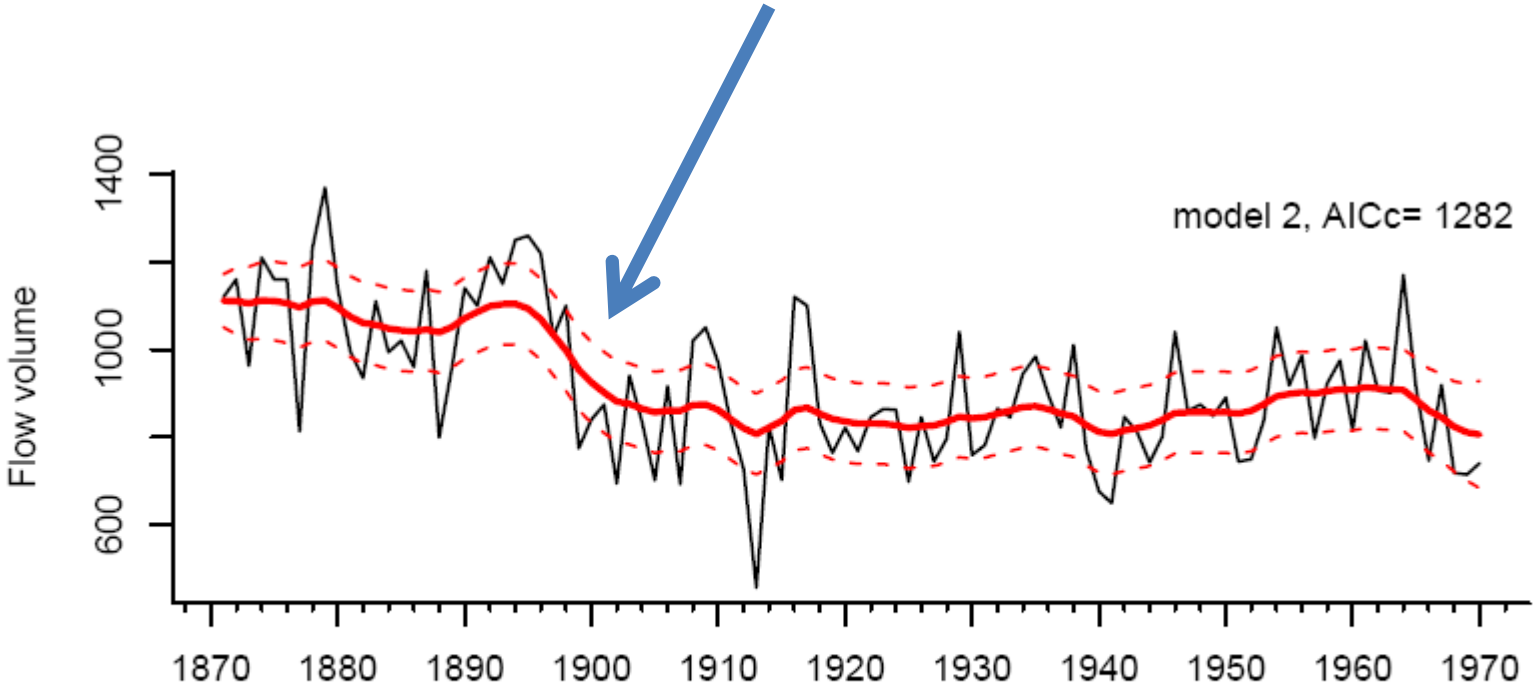
```
resids.0=residuals(kem.0)$std.residuals
resids.1=residuals(kem.1)$std.residuals
resids.2=residuals(kem.2)$std.residuals
```

*Note, the standard concerns regarding setting test levels for multiple tests exist*



# “Structural break detection” aka testing state outliers

*Idea is to test whether observed changes in the stochastic state (in this example level) were more unusual than you would expect given the estimated MAR model for the state.*



# “Structural break detection” aka testing state outliers

State outlier: estimated state at time t+1 is different than what you would expect given the model.

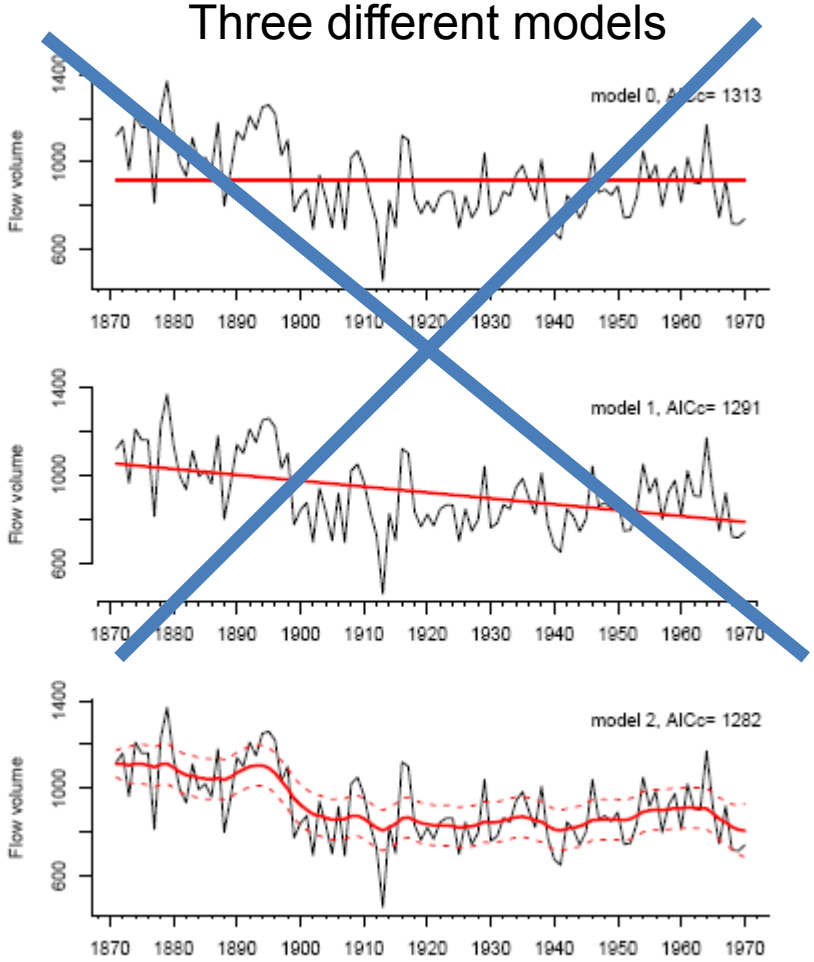
state. residual =

$$\hat{w}_t = \tilde{x}_{t|T} - \tilde{x}_{t-1|T}$$

$$f_t = \frac{1}{\sqrt{\mathrm{var}(\hat{w}_t)}}\hat{w}_t$$

we standardize by the estimated variance and get a t-distributed standardized residual

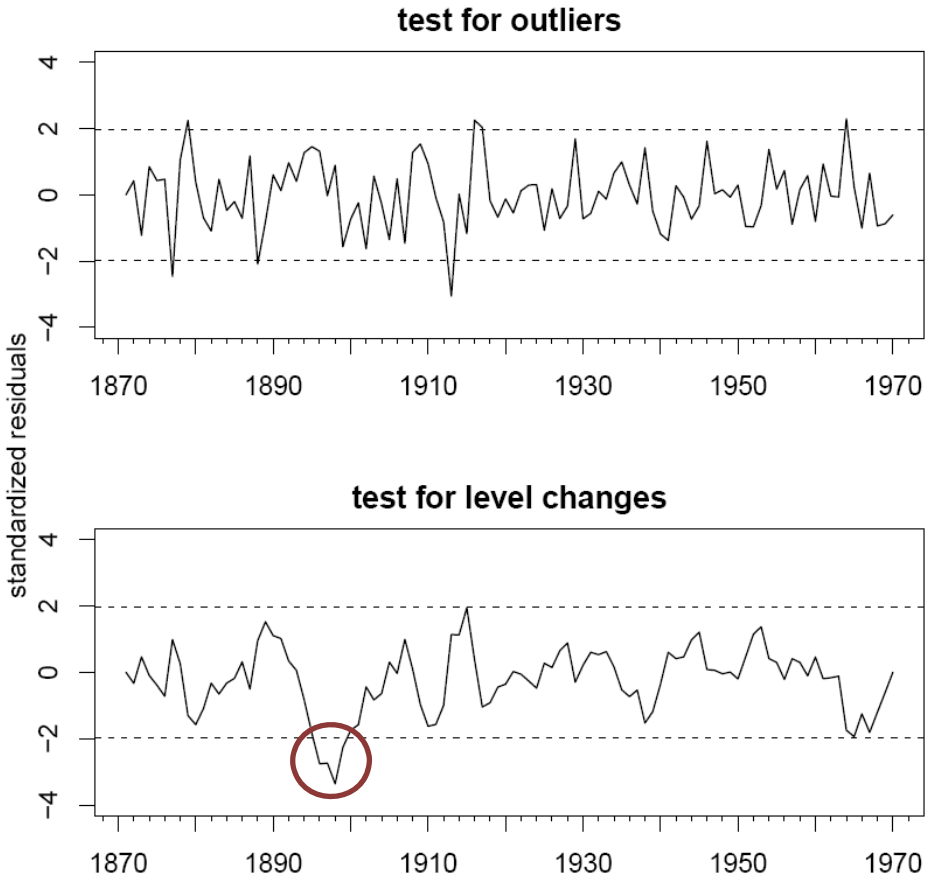
Three different models




*Again this idea hinges on w(t) being normal so that means it hinges on the model being able to fit the data (= put a line through the data)*

```
resids.0=residuals(kem.0)$std.residuals
resids.1=residuals(kem.1)$std.residuals
resids.2=residuals(kem.2)$std.residuals
```

*Note, the standard concerns regarding setting test levels for multiple tests exist*



# Summary

- Residual analysis is a diagnostic tool to look for observation or state outliers and evidence of times when the underlying model is violated, but there is no cause involved.
- Intervention analysis is more suited to a mechanistic analysis of changes/breaks that may or may not have occurred.