

Model Order Selection: formal methods

Outline of topics:

- 1) Likelihood Ratio Tests
- 2) Form of Likelihood Ratio Tests for ARMA(p, q) models.
- 3) Final Prediction Error
- 4) Akaike's Information Criterion
- 5) Example of use in model selection

General set up: data X ; joint density $f(x; \phi, \psi)$.
Dimension of ϕ is p .

Goal: test $H_o : \phi = \phi_0$.

Method: likelihood ratio test.

Maximize log-likelihood $\ell(\phi, \psi)$ twice.

First: find unrestricted MLEs $\hat{\phi}, \hat{\psi}$ by maximizing ℓ over all possibilities.

Second: find restricted MLEs $\phi_0, \hat{\psi}_0$ by maximizing $\ell(\phi_0, \psi)$ over ψ .

Likelihood ratio statistic is

$$\frac{f(X, \hat{\phi}, \hat{\psi})}{f(X; \phi_0, \hat{\psi}_0)}$$

Usual test statistic is 2 times log likelihood ratio:

$$\Lambda = 2 \left\{ \ell(\hat{\phi}, \hat{\psi}) - \ell(\phi_0, \hat{\psi}_0) \right\}$$

Large sample theory: if H_o is true then

$$\Lambda \sim \chi_p^2$$

Example: Compare $AR(p_0)$ to $AR(p_0 + p)$.
Take $\mu = 0$.

Model is

$$X_t = a_1 X_{t-1} + \cdots + a_{p_0+p} X_{t-p_0-p} + \epsilon_t$$

Take

$$\begin{aligned}\psi &= (a_1, \dots, a_{p_0}, \sigma) \\ \phi &= (a_{p_0+1}, \dots, a_{p_0+p}) \\ \phi_0 &= (0, \dots, 0)\end{aligned}$$

Write out likelihood:

$$\begin{aligned}f_{X_0, \dots, X_{T-1}} &= f_{X_0, \dots, X_{p_0+p-1}} \\ &\quad \times f_{X_{p_0+p}, \dots, X_{T-1} | X_0, \dots, X_{p_0+p-1}}\end{aligned}$$

Take logs to get

$$\ell(\phi, \psi) = \ell_M(\phi, \psi) + \ell_C(\phi, \psi)$$

Subscript C for conditional, M for marginal.

Two approaches common in software: maximize only ℓ_C or maximize ℓ .

Call

$$\Lambda_C = 2 \left\{ \ell_C(\hat{\phi}_C, \hat{\psi}_C) - \ell_C(\phi_0, \hat{\psi}_{0,C}) \right\}$$

Subscript C on ests means maximize ℓ_C .

Large sample theory still valid, that is,

$$\Lambda_C \sim \chi_p^2$$

asymptotically if H_o true.

One way to use this.

Choose between $AR(p_0)$ and $AR(p_0 + p)$:

Simplify ℓ_C :

$$\begin{aligned}\ell_C &= -\frac{1}{2\sigma^2} \sum_{p_0+p}^{T-1} (X_t - \sum a_j X_{t-j})^2 \\ &\quad - (T - p - p_0) \{\log(\sigma) + \log(2\pi)/2\} \\ &= -\frac{1}{2\sigma^2} \sum_{p_0+p}^{T-1} \epsilon_t^2 \\ &\quad - (T - p - p_0) \{\log(\sigma) + \log(2\pi)/2\}\end{aligned}$$

Suffices to minimize

$$\frac{1}{2\sigma^2} \sum_{p_0+p}^{T-1} \epsilon_t^2 + (T - p - p_0) \log(\sigma)$$

Steps:

- 1) Minimize $\sum \epsilon_t^2$. (Notice notational tactic – think of ϵ_t as depending on data and a values.) Get \hat{a}_j by ordinary least squares.
- 2) Get two sets of residuals $\hat{\epsilon}_t$ and $\hat{\epsilon}_{t,0}$.

3) Compare

$$\frac{\sum \hat{\epsilon}_t^2}{2\sigma^2} + (T - p - p_0) \log(\sigma)$$

and

$$\frac{\sum \hat{\epsilon}_{t,0}^2}{2\sigma^2} + (T - p - p_0) \log(\sigma)$$

Minimize over σ to find

$$\hat{\sigma}^2 = \frac{\sum \hat{\epsilon}_t^2}{T - p - p_0}$$

$$\hat{\sigma}_0^2 = \frac{\sum \hat{\epsilon}_{t,0}^2}{T - p - p_0}$$

Get

$$\begin{aligned}\Lambda_C &= (T - p - p_0) \log(\hat{\sigma}_0^2 / \hat{\sigma}^2) \\ &= (T - p - p_0) \log \hat{\sigma}_0^2 - (T - p - p_0) \log \hat{\sigma}^2.\end{aligned}$$

Difficulties arising:

- 1) Which two models to compare? What p and p_0 ?
- 2) What about $\text{ARMA}(p, q)$?
- 3) Can compare only *nested* models. Can't, e.g., compare $\text{MA}(2)$ and $\text{AR}(2)$.
- 4) Must be careful to condition on same number of data points in fitting both models.
- 5) Really want to compare many models. Try variety of p, q pairs. Not a hypothesis testing problem.

Akaike's suggestions:

1) Choose model order to minimize Final Prediction Error:

$$\left(\frac{T+K}{T-K}\right) \frac{\sum \epsilon_{K,t}^2}{T}.$$

K is number of parameters; subscript K on ϵ means residuals from that model.

2) Minimize Akaike's Information Criterion:

$$AIC = \log(\sum \epsilon_{K,t}^2/T) + 2K/T$$

Note:

$$\begin{aligned} \log(FPE) &= \log(\sum \epsilon_{K,t}^2/T) + \log(1 + 2K/T) \\ &\quad + O(T^{-2}) \\ &= AIC + 2K/T + O(T^{-2}). \end{aligned}$$

Idea: compare many models with different numbers K of parameters by computing

$$AIC_K \equiv \log(\hat{\sigma}_K^2) + 2K/T$$

or equivalently

$$AIC_K \equiv T \log(\hat{\sigma}_K^2) + 2K$$

Latter is equivalent to -2 times log likelihood +2K.

In time series: must make sure to use same data points to compute

$$\hat{\sigma}_{p+q}^2 = \frac{\sum \epsilon_t^2}{\# \text{ data pts}}$$

Problems:

- 1) Plot AIC_p against p for $p = 0, 1, \dots, P$. How to select P the largest order tried?
- 2) The method is not consistent; overfitting is likely.

The good news:

$$\text{Prob}_{p_0}(\hat{p} < p_0) \rightarrow 0$$

but:

$$\text{Prob}_{p_0}(\hat{p} > p_0) \not\rightarrow 0$$

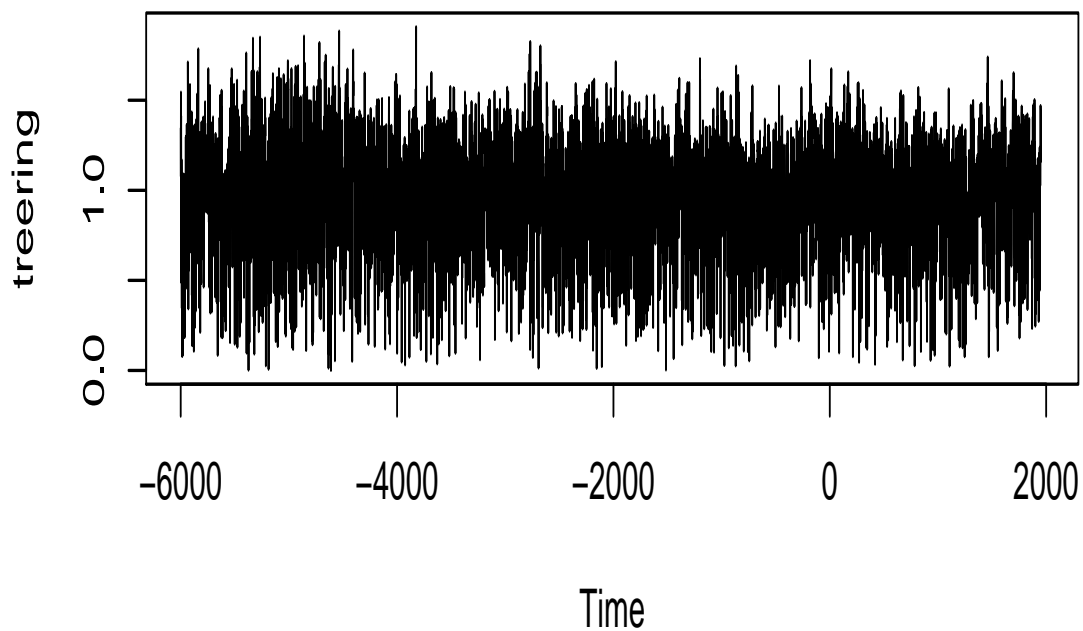
Alternative suggestions. Maximize

$$\ell(\hat{\phi}_p, \hat{\psi}_p) - \text{ftn}(T) \times p$$

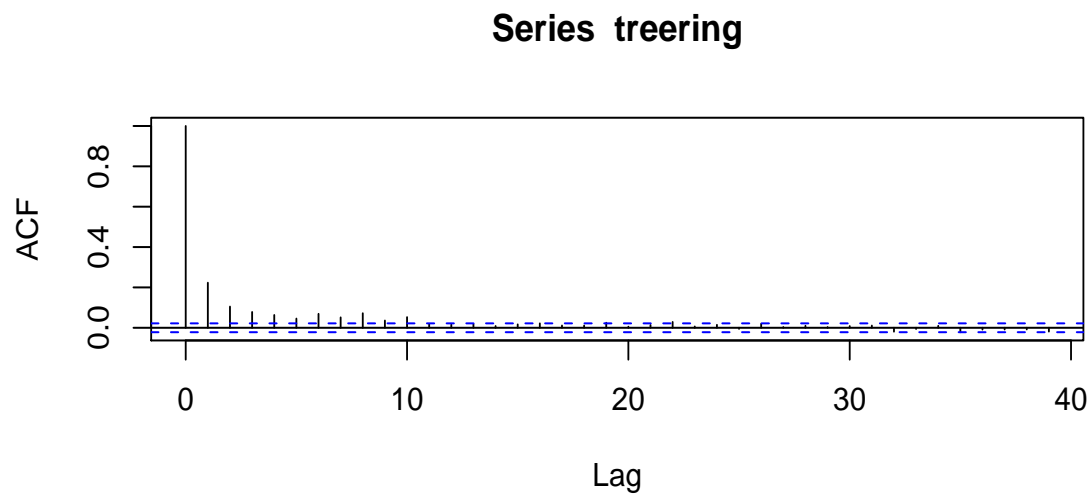
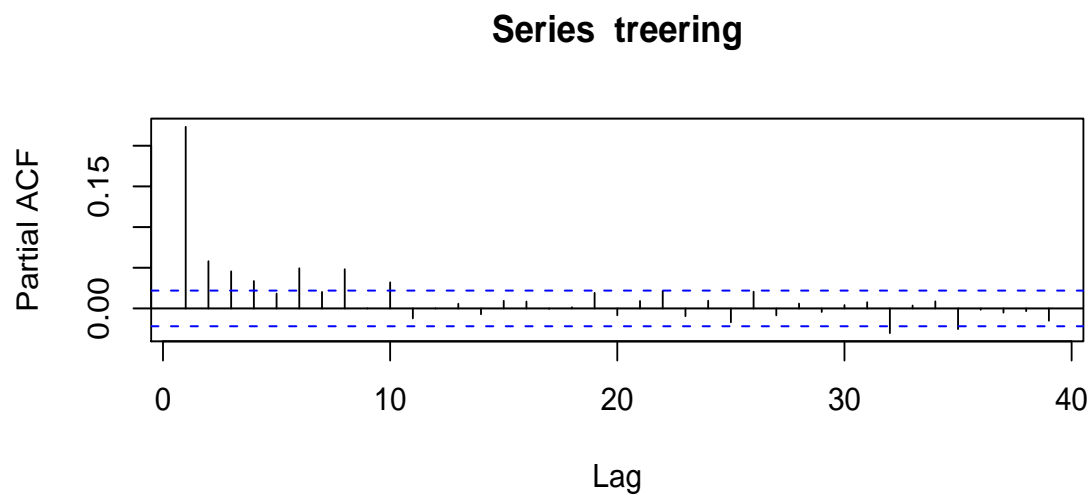
Leads, eg, to BIC.

Tree Ring Data from R

Dataset `treering` in **R** is “normalized tree-ring widths in dimensionless units.” Time plot:



Here are the ACF and PACF:



Note neither plot goes quickly to 0 suggesting no low order AR or MA will provide good fit.

So: try a variety of low order ARMA(p, q).

For purposes of example I tried all p and q from 1 to 10.

Code:

```
modaic <- matrix(0,10,10)
for( p in 1:10){
  for( q in 1:10){
    modaic[p,q] <- arima0(treering,c(p,0,q))$aic
  }
}
```

Top left corner of modaic:

```
> modaic[1:3,1:3]
      [,1]      [,2]      [,3]
[1,] 3003.639 2968.884 2968.967
[2,] 2966.934 2969.322 2963.461
[3,] 2968.921 2964.367 2965.404
```

To examine values: subtract smallest value, round to 2 digits.

```
> round(modaic-min(modaic),2)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,] 43.37 8.62  8.70 10.61 12.60  8.02  9.92  6.63 17.70  2.90
[2,]  6.67 9.06  3.20  5.23  4.58  2.42  4.41  2.12  3.46  4.45
[3,]  8.66 4.10  5.14 16.31 15.55  4.41  6.12  4.00  5.16  5.00
[4,] 10.63 5.29 14.56 16.34  7.86  7.73 11.12  7.95  7.83  7.22
[5,] 12.18 6.30 16.57 10.13  5.93  6.36  5.47  5.76 10.26  9.13
[6,]  7.86 2.37  4.43  6.76  2.13  7.53 12.95  6.63 13.26 10.17
[7,]  9.77 4.46  6.07  7.66  6.31  8.25  0.84 13.25 13.69  6.23
[8,]  6.30 2.31  2.38  4.27  5.94  7.07  2.43 13.52  6.59  8.32
[9,]  8.28 4.25  5.37  6.10  3.68 10.77  4.63  6.04  0.00  0.47
[10,]  1.54 3.47  5.41  7.26  9.23 11.22 13.27 15.34 17.21  3.05
```

Notice: least AIC at ARMA(9,9).

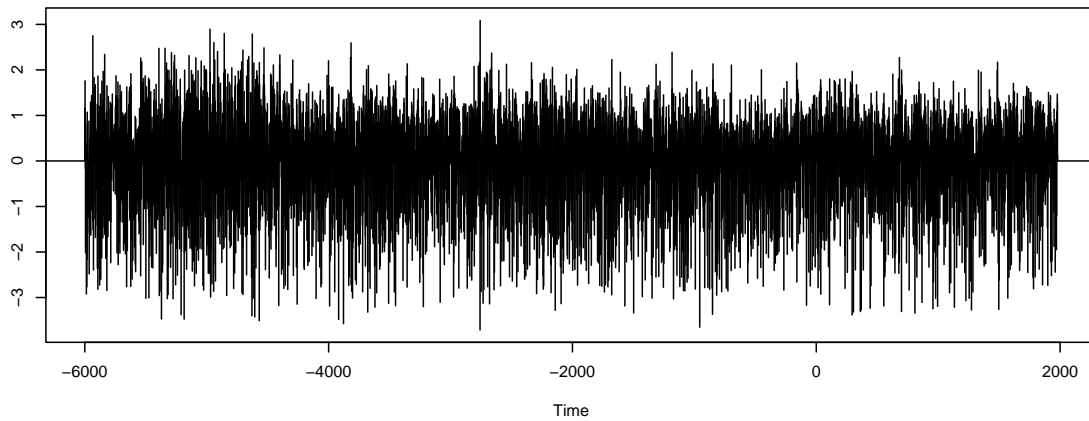
But also try some low order models – for parsimony.

Try say ARMA(2,3).

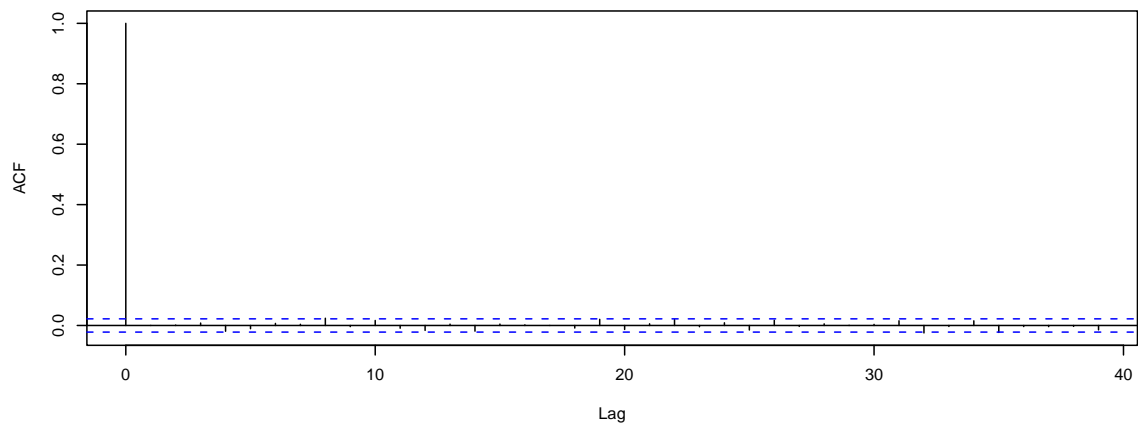
Summary of situation: both models have good diagnostics but ARMA(9,9) looks like overfitting.

ARMA(2,3)

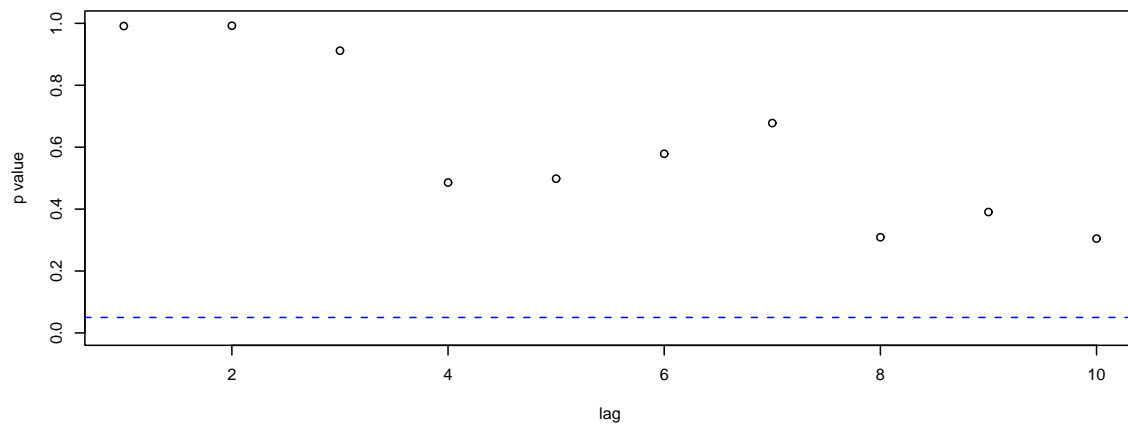
Standardized Residuals



ACF of Residuals

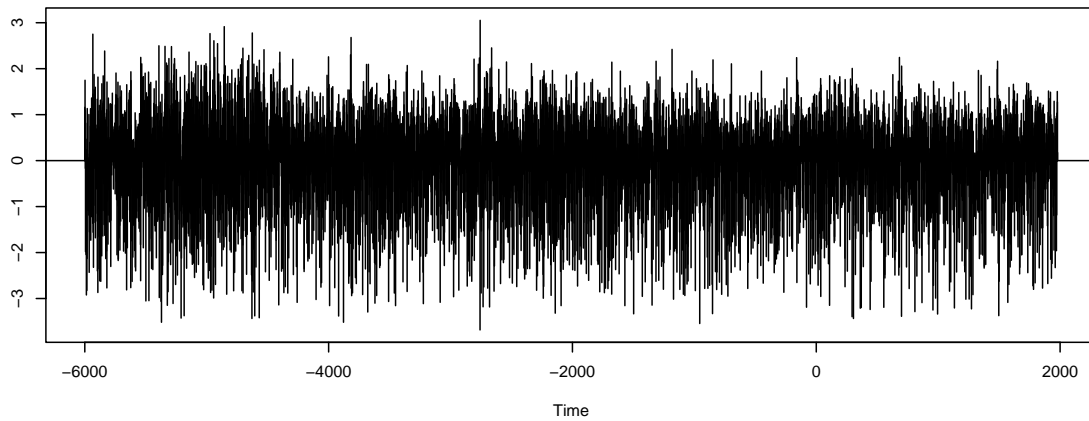


p values for Ljung-Box statistic

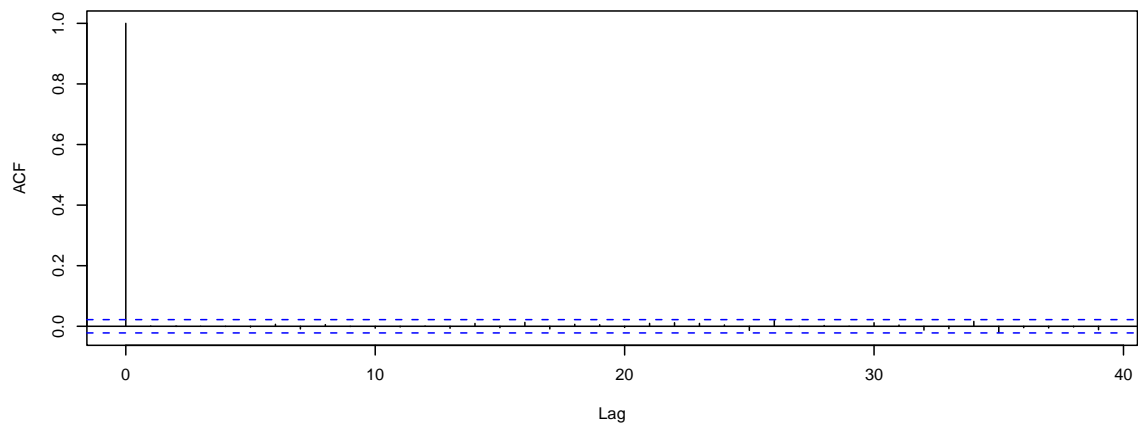


ARMA(9,9)

Standardized Residuals



ACF of Residuals



p values for Ljung-Box statistic

