ARMA Modeling and Forecasting (Chap 5)

5.1 Preliminary Estimation

Useful for

- order identification (requires the fitting of a number of competing models).
- initial parameter estimates for likelihood optimization.

ARMA(p,q) Model: Based on observations $x_1,\ldots, x_n$,

from the model

$$\phi(B)\, X_t = \theta(B)\, Z_t\,, \qquad \{Z_t\} \sim \mathrm{WN}(0,\sigma^2),$$

want to estimate $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$,

where the orders p and q are assumed known (for the moment).

AR(p) Processes:

Yule-Walker Estimation (moment estimates)

Burg Estimation

ARMA(p,q) Processes:

Innovations Algorithm

Hannan-Rissanen Estimates

Yule-Walker Estimates (for AR(p) processes):

Recall from Chapter 3, that

$$\gamma(0)-\phi_1\gamma(1) - \cdots - \phi_p \gamma(p) = \sigma^2$$

$$\gamma(k)-\phi_1\gamma(k-1) - \cdots - \phi_p \gamma(k-p) = 0, \quad k=1,\ldots,p$$

or in matrix form,

$$\Gamma_p \phi = \gamma_p$$

where Γ_p is the covariance matrix of $X_1, \ldots, X_p$, and $\phi = (\phi_1, \ldots, \phi_p)'$, $\gamma_p = (\gamma(1), \ldots, \gamma(p))'$. The Y-W estimates are then found by replacing the ACVF $\gamma(\ )$, by its estimated value $\hat{\gamma}(\ )$.

3

Sample Yule-Walker Equations:

$$\hat{\phi} = \hat{\Gamma}_p^{-1}\hat{\gamma}_p,$$

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\gamma}_p'\hat{\Gamma}_p^{-1}\hat{\gamma}_p.$$

(These estimates are computed in ITSM .)

Remarks:

- The fitted model is causal.
- The sample ACVF and fitted model ACVF agree at lags h=0,1,. . . ,p.
- Estimates are asymptotically efficient (i.e. have the same limit behavior as MLE).

$$\hat{\phi} \approx N(\phi, n^{-1}\sigma^2\Gamma_p^{-1}).$$

4

Order Selection for AR Models

Usually there is no true AR model. Goal is to find an AR model which represents the data in some sense.

Two Techniques.

- If $\{X_t\}$ is an AR(p) process with $\{Z_t\} \sim \text{IID}(0,\sigma^2)$, then $\hat{\phi}_{mm} = \hat{\alpha}(m) \sim N(0, n^{-1})$ for all $m > p$. ($\hat{\alpha}(m)$ is the sample PACF). Estimate p as the smallest value m such that

$$|\hat{\alpha}(k)| < 1.96\, n^{-.5} \text{ for } k > m.$$

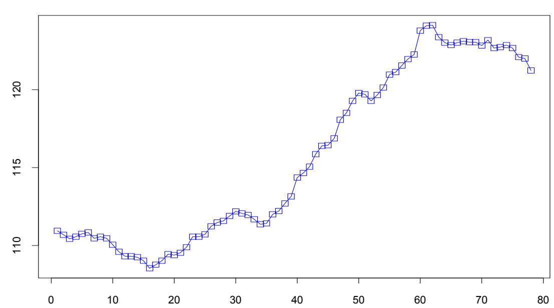
- Estimate p by minimizing the AICC statistic

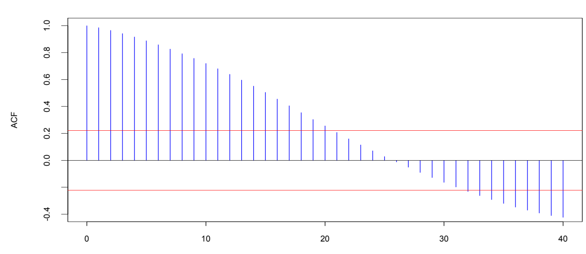
$$\text{AICC} = -2\ln L(\phi_p) + 2(p+1)n/(n-p-2)$$

where L( ) denotes the Gaussian likelihood.

5

Ex 5.1.1 (Dow-Jones Utilities Index; DOWJ.DAT).
Plot of $D_1, \ldots, D_{78}$





6

Fig 5.1 ACF of differenced series $Y_t = D_t - D_{t-1}$

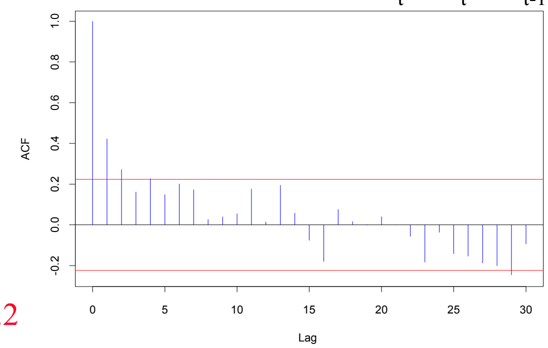
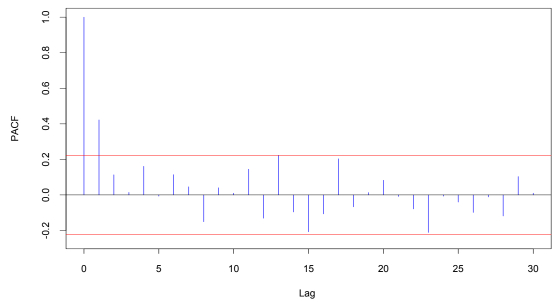


Fig 5.2



Preliminary Model for Dow-Jones:

Using the Preliminary Estimation option of ITSM, the fitted model is

$$X_t = \underset{(.094)}{.4219}\, X_{t-1} + Z_t\,, \quad \{Z_t\} \sim \mathrm{WN}(0, .1479)$$

where

$$X_t = Y_t - .1336, \quad Y_t = D_t - D_{t-1}\,,$$

Remark: We can also arrive at this model using the automatic AICC minimization option in Preliminary Estimation.

5.1.2 Burg's Algorithm

The Yule-Walker estimates $(\hat{\phi}_1, \ldots, \hat{\phi}_p)$ satisfy

$$\tilde{P}_p X_{p+1} = \hat{\phi}_1 X_p + \cdots + \hat{\phi}_p X_1$$

where $\tilde{P}_p$ is the prediction operator relative to the fitted Y-W model. As can be seen from the Durbin-Levinson algorithm, these coefficients are determined from the sample PACF ,

$$\hat{\alpha}(h) = \hat{\phi}_{hh}\ , \ h = 0, \ldots, p.$$

Burg's algorithm uses a different estimate of the PACF based on minimizing the forward and backward one-step prediction errors.

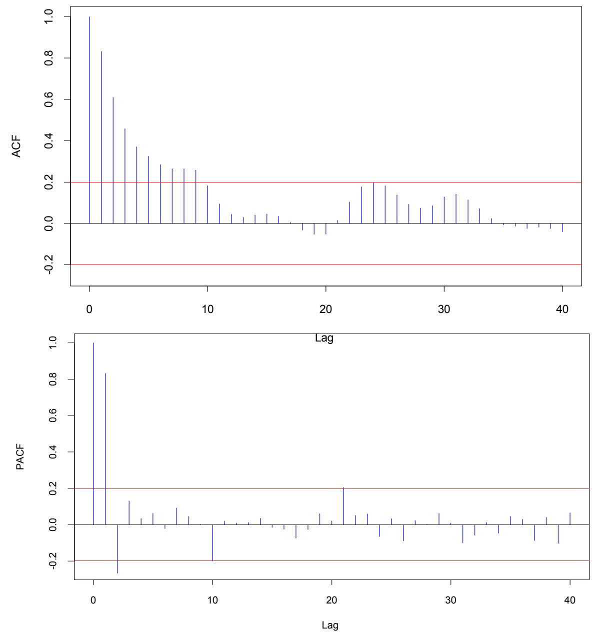
9

Properties of Burg's estimates:

- The fitted model is causal.
- Estimates are asymptotically efficient (i.e. have the same limit behavior as MLE).
- Estimates tend to perform better than the Y-W estimates especially if a zero of the AR polynomial is close to the unit circle.

10

Ex 5.1.4 (Lake Data).



11

Model for the Lake data.

The sample ACF and PACF's suggest an AR(2) model for the mean corrected data

$$X_t = Y_t - 9.0041.$$

Burg's model:

$$X_t = 1.0449\, X_{t-1} - .2456\, X_{t-2} + Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, .4705)$$

Y-W fitted model:

$$X_t = 1.0583\, X_{t-1} - .2668\, X_{t-2} + Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, .4920)$$

Minimum AICC model using Burg or Y-W is p=2.

Burg's gives smaller WN variance and larger likelihood

12

5.1.3 The Innovations Algorithm

The Yule-Walker estimates were found by applying the Durbin-Levinson Algorithm with the ACF replaced by the sample ACF. The Innovations algorithm estimates are obtained in the same fashion--the ACF is replaced by the sample ACF.

The fitted innovations MA(m) model:

$$X_t = Z_t + \hat{\theta}_{m1} Z_{t-1} + \cdots + \hat{\theta}_{mm} Z_{t-m}, \ \{Z_t\} \sim \mathrm{WN}(0, \hat{v}_m)$$

where $\hat{\theta} = (\hat{\theta}_{m1}, \ldots, \hat{\theta}_{mm})'$ and $\hat{v}_m$ are found from the innovations algorithm with the ACVF replaced by the sample ACVF.

Innovations estimates for MA(q) model.

$$X_t = Z_t + \hat{\theta}_{m1} Z_{t-1} + \cdots + \hat{\theta}_{mq} Z_{t-q}, \ \{Z_t\} \sim \mathrm{WN}(0, \hat{v}_m),$$

where m_n is a sequence of integers,

$$m_n \longrightarrow \infty, \quad m_n = o(n^{1/3}).$$

(These estimates are computed in ITSM.)

Order Selection for MA Models

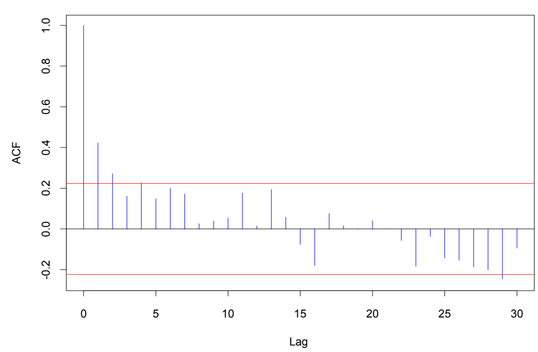
Three Techniques.

- If $\{X_t\}$ is an MA(q) process with $\{Z_t\} \sim \text{IID}(0,\sigma^2)$, then $\hat{\rho}(m) \sim N(0, n^{-1}(1+2\rho^2(1) + \cdots + 2\rho^2(q))$ for all $m > q$. Estimate q as the smallest value m such that $\hat{\rho}(k)$ is not significantly different from 0, for all $k > m$.
- Examine the coefficient vector in the fitted innovations MA(m) model to see which are not significantly different from 0.
- Estimate q by minimizing the AICC statistic

$$\text{AICC} = -2 \ln L(\theta_q) + 2(q+1)n/(n-q-2)$$

15

Ex 5.1.5 (Dow-Jones Data). $X_t = D_t - D_{t-1} - .1336$



ACF suggests MA(2).

MA(2) model using innovations estimates: (m=17)

$$X_t = Z_t + \underset{(.114)}{.4269}\, Z_{t-1} + \underset{(.124)}{.2704}\, Z_{t-2}\,,\ \{Z_t\} \sim \text{WN}(0,.1470)$$

16

MA(17) model:

MA Coefficient							
.427	.270	.118	.159	.135	.157	.128	-.006
.015	-.003	.197	-.046	.202	.129	-.021	-.258
.076							

COEFFICIENT/(1.96*STANDARD ERROR)							
1.911	1.113	.473	.631	.533	.613	.497	-.023
.057	-.006	.759	-.176	.767	.480	-.079	-.956
.276							

17

Innovations Algorithm Estimates When $p > 0$

Step 1. Use Innovations algorithm to fit MA(m) model

$$X_t = Z_t + \hat{\theta}_{m1} Z_{t-1} + \cdots + \hat{\theta}_{mm} Z_{t-m}, \ \{Z_t\} \sim \mathrm{WN}(0, \hat{v}_m),$$

where $m > p + q$.

Step 2. Replace ψ_j by $\hat{\theta}_{mj}$ in the following equations:

$$\psi_j = \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \psi_{j-i}, \ j = 1, \ldots, p+q.$$

Step 3. Solve the equations in Step 2 for ϕ and θ.

18

5.1.4 Hannan-Rissanen Algorithm

Step 1. Fit a high order AR(m) (m > max(p,q)) using Yule -Walker estimates and compute estimated residuals

$$\hat{Z}_t = X_t - \hat{\phi}_{m1} X_{t-1} - \cdots - \hat{\phi}_{mm} X_{t-m}, \quad t = m+1, \ldots, n.$$

Step 2. Estimate ϕ and θ, by regressing X_t onto $(X_{t-1}, \ldots, X_{t-p}, \hat{Z}_{t-1}, \ldots, \hat{Z}_{t-q})$, i.e. by minimizing the sum of squares

$$S(\phi, \theta) = \sum_{t=m+1}^{n} (X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \theta_1 \hat{Z}_{t-1} - \cdots - \theta_q \hat{Z}_{t-q})^2$$

19

Ex 5.1.7 (Lake Data). We fit an ARMA(1,1) model using the Innovations and Hannan-Rissanen estimates to the mean corrected data $X_t = Y_t - 9.0041$.

Innovations fitted model :

$$X_t - \underset{(.115)}{.7234} X_{t-1} = Z_t + \underset{(.099)}{.3596} Z_{t-1}, \quad \{Z_t\} \sim \mathrm{WN}(0,.4757)$$

Hannan-Rissanen fitted model :

$$X_t - \underset{(.078)}{.6961} X_{t-1} = Z_t + \underset{(.147)}{.3788} Z_{t-1}, \quad \{Z_t\} \sim \mathrm{WN}(0,.4757)$$

20

5.2 Maximum Likelihood Estimation

Suppose $\{X_t\}$ is a causal ARMA(p,q) process.

The Gaussian Likelihood :

$$L(\phi, \theta, \sigma^2) = (2\pi\sigma^2)^{-n/2}(r_0 \cdots r_{n-1})^{-.5} \exp\{-\frac{1}{2\sigma^2} \sum_{t=1}^{n} \frac{(X_t - \hat{X}_t)^2}{r_{t-1}}\}$$

where $\hat{X}_t = P_{t-1} X_t$, $\sigma^2 r_{t-1} = E(X_t - \hat{X}_t)^2$.

21

Maximum Likelihood Estimators:

$$\hat{\sigma}^2 = n^{-1} S(\hat{\phi}, \hat{\theta}),$$

where

$$S(\hat{\phi}, \hat{\theta}) = \sum_{t=1}^{n} (X_t - \hat{X}_t)^2 / r_{t-1},$$

and $\hat{\phi}, \hat{\theta}$ are the values of ϕ, θ which minimize the **reduced likelihood**

$$l(\phi, \theta) = \ln(n^{-1} S(\phi, \theta)) + n^{-1} \sum_{t=1}^{n} \ln(r_{t-1}).$$

Remark: Minimization of $l(\phi, \theta)$ must be done numerically which is carried out in ITSM.

22

Order Selection

AICC Criterion:

Choose p, q, ϕ_p and θ_q to minimize

$$\text{AICC} = -2 \ln L(\phi_p , \theta_q) + 2(p+q+1)n/(n-p-q-2).$$

For fixed p, q, AICC is minimized when ϕ_p , θ_q are the maximum likelihood estimates.

Large Sample Behavior of MLE :

For a large sample,

$(\hat{\phi} , \hat{\theta})'$ is approx $N((\phi , \theta)' , n^{-1}V(\phi , \theta))$ (see p.162)

Ex. 5.2.5 (The Lake Data).

Preliminary ARMA(1,1) model:

$$X_t - .7234\, X_{t-1} = Z_t + .3596\, Z_{t-1},\ \{Z_t\} \sim \text{WN}(0,.4757)$$

(Model fitted to the mean-corrected data, $X_t = Y_t - 9.004$ using the innovations algorithm found in the Preliminary Estimation option of ITSM.)

Maximum likelihood ARMA(1,1) model:

$$X_t - \underset{(.0771)}{.7453}\, X_{t-1} = Z_t + \underset{(.1116)}{.3208}\, Z_{t-1},\ \{Z_t\} \sim \text{WN}(0,.4750)$$

Estimates are found using option ARMA Parameter Estimation in ITSM.

Maximum Likelihood AR(2) Model:

$X_t - 1.0415\, X_{t-1} + .2494\, X_{t-2} = Z_t$, $\{Z_t\} \sim \mathrm{WN}(0,.4790)$

AICC = 213.54 (AR(2) Model)

AICC = 212.77 (ARMA(1,1) model)

Based on AICC, ARMA(1,1) model is slightly better.

5.3 Diagnostic Checking

Residuals:

$$\hat{W}_t = (X_t - \hat{X}_t)/\, r_{t-1}^{1/2}, \quad t = 1, \ldots, n.$$

Properties of $\{\hat{W}_t\}$ should reflect those of the WN sequence $\{Z_t\}$, namely $\{\hat{W}_t\}$ should be approximately

- uncorrelated if $\{Z_t\} \sim \mathrm{WN}(0,\sigma^2)$
- independent if $\{Z_t\} \sim \mathrm{IID}(0,\sigma^2)$
- normally distributed if $Z_t \sim \mathrm{N}(0,\sigma^2)$

Rescaled Residuals :

$$\hat{R}_t = \hat{W}_t \;/\hat{\sigma}$$

Diagnostic Checks :

Graph of $\{\hat{R}_t\}$.

- Inspect for departures from WN (eg. trend, cyclical component, non-constant variance, ...)

Sample ACF of $\{\hat{R}_t\}$.

- Compare with bounds $\pm$ 1.96 $n^{-.5}$.
- Ljung-Box Portmanteau Test
- McLeod-Li Portmanteau Test

27

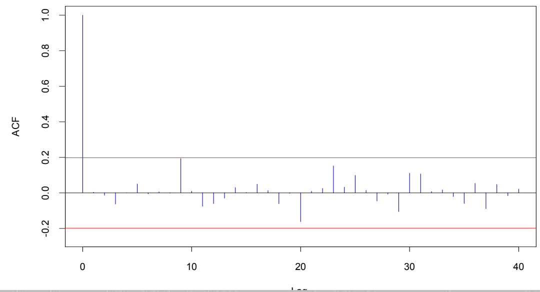
Tests for Randomness

- Turning points
- Difference Sign Test
- Rank Test
- Minimum AICC YW Model for Residuals

All these tests can be carried out in ITSM after model fitting by selecting Statistics>Residual Analysis>Tests of Randomness.

28

Applied to residuals of Lake Data from ARMA(1,1) fit:




```
Ljung - Box statistic = 10.234 Chi-Square ( 20 ), p-value = .96372

McLeod - Li statistic = 16.548 Chi-Square ( 22 ), p-value = .78777

# Turning points = 69.000~AN(64.000,sd = 4.1352), p-value = .22661

# Diff sign points = 50.000~AN(48.500,sd = 2.8723), p-value = .60151

Rank test statistic = .20830E+04~AN(.23765E+04,sd = .16290E+03), p-value = .07160

Jarque-Bera test statistic (for normality) = .28493 Chi-Square (2), p-value = .86722

Order of Min AICC YW Model for Residuals = 0
```

5.4 Forecasting

Ex 5.4.1. (Overshorts).

Maximum likelihood MA(1) Model :

$$X_t + 4.035 = Z_t - .818\, Z_{t-1},\ \{Z_t\} \sim \text{WN}(0,2040.75)$$

To predict the next 5 days of overshorts, we have

$$P_{57} X_{57+h} = \begin{cases} -4.035 + \theta_{57,1}(X_{57} - \hat{X}_{57}), & \text{if h=1,} \\ -4.035, & \text{if h>1.} \end{cases}$$

#	XHAT	SQRT(MSE)	XHAT+MEAN
58	1.0097	45.1753	-3.0254
59	0.0000	58.3602	-4.0351
59	0.0000	58.3602	-4.0351
60	0.0000	58.3602	-4.0351
61	0.0000	58.3602	-4.0351

95% Prediction Interval for X_{61} :

$$-4.0351 \mp 1.96(58.3602) = (-118.42, 110.35)$$

This interval assumes that the noise is Gaussian.

31

Effect of parameter estimation on forecast error :

Data : $X_1, \ldots, X_n$

Model : $X_t = \phi X_{t-1} + Z_t$, $\{Z_t\} \sim \mathrm{IID}(0,\sigma^2)$

MLE estimate : $\hat{\phi}$

One-step forecast : $\hat{X}_{n+1} = \hat{\phi} X_n$

Mean-square error of forecast :

$$E(X_{n+1} - \hat{\phi} X_n)^2 = E((\phi - \hat{\phi})X_n + Z_{n+1})^2,$$

$$= E((\phi - \hat{\phi})X_n)^2 + \sigma^2.$$

Now, $E((\phi - \hat{\phi})^2|X_n) \sim E(\phi - \hat{\phi})^2 X_n^2 \sim ((1-\phi^2)/n)X_n^2$,

so, $E(X_{n+1} - \hat{\phi} X_n)^2 \sim n^{-1}(1-\phi^2)(1-\phi^2)^{-1}\sigma^2 + \sigma^2$

$$= \sigma^2(1 + 1/n)$$

Error in parameter estimation contributes σ^2/n to MSE.

32

5.5 Order Selection

AICC Statistic :

$$-2 \ln L_X(\phi,\theta, S_X(\phi,\theta)/n) + 2(p+q+1)n/(n-p-q-2)$$

AIC Statistic :

$$-2 \ln L_X(\phi,\theta, S_X(\phi,\theta)/n) + 2(p+q+1)$$

BIC Statistic :

$$(n-p-q) \ln (n\,\sigma^2/(n-p-q)) + n(1 + .5 \ln (2\pi)) + (p+q) \ln \Big(\Big(\sum_{t=1}^{n} X_t^2 - n\,\sigma^2\Big)/(p+q)\Big).$$

33

Properties of AICC, AIC, BIC :

- BIC is consistent; estimated orders converge to true orders.
- AICC and AIC are not consistent. Designed to be 'unbiased' estimates of Kullback Leibler index.
- AICC and AIC are efficient; fitted model achieves optimal rate of convergence for the mean square (one-step) prediction error (see Section 9.3 of TSTM).

BIC AICC AIC

fewer parameters ⟵⟶ more parameters

34