Ch5. Estimation of ARMA(p, q) processes

estimation of parameters

method of moment estimator

need an initial estimator

1. Estimation using MME/MLE/LSE
2. Diagnostic Checking
3. Order Selection
4. Forecasting
5. Data analysis example with R

Estimation of ARMA(p, q) parameters

AR(p)

- Want to estimate parameters $\phi = (\phi_1, \dots, \phi_p)'$,
MA(q)
$\theta = (\theta_1, \dots, \theta_q)'$ and σ^2 from the observations $x_1, \dots, x_n$ of the causal ARMA(p, q) processes

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

- Since MLE is known to have very good properties, we will ultimately use MLE. In practice, however, it requires **numerical optimization** so we need some good initial values.
- Strategy: MME or LSE for initial estimation
  - Do some preliminary estimation. (Yule-Walker and variants, LSE)
  - Follow-up with (Gaussian) maximum likelihood estimation (MLE).

MME: Yule-Walker equation for AR(p)

- Consider causal AR(p) model

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t, \quad Z_t \sim WN(0, \sigma^2).$$

- Multiply X_{t-k} on both sides and taking expectation gives

$$E(X_{t-k}X_t) = E((\phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t)X_{t-k})$$

$$\gamma(k) = \phi_1\gamma(k-1) + \ldots + \phi_p\gamma(k-p) + \underline{\mathrm{Cov}(Z_t, X_{t-k})}.$$

Note from the causality,

$$\mathrm{Cov}(Z_t, X_{t-k}) = \mathrm{Cov}\left(Z_t, \sum_{j=0}^{\infty} \psi_j Z_{t-k-j}\right) = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \geq 1. \end{cases}$$

$Z_{t-k} + \psi_1 Z_{t-k-1} + \psi_2 Z_{t-k-2}$

- Yule-Walker equation

$$\gamma(k) = \phi_1\gamma(k-1) + \ldots + \phi_p\gamma(k-p), \quad k \geq 1$$

$$\gamma(0) = \phi_1\gamma(-1) + \ldots + \phi_p\gamma(-p) + \sigma^2$$, $k=0$

Yule-Walker equation

if you solve the system of linear equations for given $\phi_1, \dots, \phi_p \Rightarrow \gamma(k)$.
Now, we are solving for $\phi_1, \dots, \phi_p$ with given ACVF's.

- In a matrix form

$$\underbrace{\begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix}}_{\gamma_p} = \underbrace{\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(p-1) & \dots & \gamma(1) & \gamma(0) \end{pmatrix}}_{\Gamma_p} \underbrace{\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}}_{\phi}$$

- Thus, Yule-Walker equation is given by

$$\phi = \Gamma_p^{-1} \gamma_p$$
$$\sigma^2 = \gamma(0) - \phi_1 \gamma(1) - \dots - \phi_p \gamma(p)$$

- Note that (when mean is zero)

$$\gamma(k) = E X_0 X_k \approx \overbrace{\frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k}}^{\text{MME of } \gamma(k)} = \widehat{\gamma}(k)$$

Yule-Walker equation

- Thus, YW estimator is given by replacing theoretical ACF by SACF

$$\begin{aligned} \widehat{\phi} &= \widehat{\Gamma}_p^{-1}\widehat{\gamma}_p \\ \widehat{\sigma}^2 &= \widehat{\gamma}(0) - \phi_1\widehat{\gamma}(1) - \ldots - \phi_p\widehat{\gamma}(p), \end{aligned}$$

where

$$\widehat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(X_t - \overline{X})(X_{t+k} - \overline{X})$$

MLE for ARMA(p, q) assume gaussianity $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim MVN(0, \Sigma)$, $\frac{1}{|2\pi \Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2} X' \Sigma^{-1} X)$

- MLE is defined as

$$\widehat{\boldsymbol{\theta}}^{MLE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ L_n(X_1, X_2, \ldots, X_n).$$

However, numerically not easy to compute $n \times n$ matrix inverse. = we have to get Σ^{-1}

- Let $X_0 = 0$ and consider one-step ahead prediction $P_n X_{n+1}$ which is the projection (orthogonalization) of X_{n+1} onto $\{X_0, X_1, \ldots, X_n\}$. Denote $\widehat{X}_n = P_{n-1} X_n$. Then, Cholesky Decomposition

↳ Orthogonalize by projection

$$\widehat{\boldsymbol{\theta}}^{MLE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ L_n(X_1, X_2 - \widehat{X}_2, \ldots, X_n - \widehat{X}_n)$$

They are all orthogonal so Σ only have diagonal terms

- $(X_1 - \widehat{X}_1)$, $(X_2 - \widehat{X}_2), \ldots, (X_n - \widehat{X}_n)$ are uncorrelated and

$$E(X_t - \widehat{X}_t) = 0$$

$$\mathrm{Var}(X_t - \widehat{X}_t) = E(X_t - \widehat{X}_t)^2 =: \sigma^2 r_{t-1}\ \text{(MSPE)}$$

diagonal entries

MLE for ARMA(p, q)

- MLE is given by (numerically) minimizing

$$\boxed{-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\frac{1}{n}\sum_{t=1}^{n}\frac{(X_t - \widehat{X}_t)^2}{r_{t-1}}\right) - \frac{1}{2}\sum_{t=1}^{n}\log r_{t-1} - \frac{n}{2}}$$

☆ $Var(X_t - \hat{X}_t) = E(X_t - \hat{X}_t)^2 \approx \sigma^2 \cdot r_{t-1}$

with

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\frac{(X_t - \widehat{X}_t)^2}{r_{t-1}}$$

- YW is a good initial estimator. There are fast algorithm (Innovations algorithm Ch. 2. 5.2) to calculate $(X_t - \widehat{X}_t)$ and r_t.

$$\widehat{X}_t = \begin{cases} \sum_{j=1}^{n}\theta_{nj}(X_{n+1-j} - \widehat{X}_{n+1-j}), & 1 \le n \le \max(p, q+1) \\ \phi_1 X_n + \ldots + \phi_p X_{n+1-p} + \sum_{j=1}^{q}\theta_{nj}(X_{n+1-j} - \widehat{X}_{n+1-j}), & n \ge \max(p, q+1) \end{cases}$$

- First few initial values are hard to calculate. If you just ignore them, it is called **conditional likelihood**.

Inference for MLE

- Since MLE is BAN (Best Asymptotic Normal)

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(0, I(\boldsymbol{\theta})^{-1}), \quad I(\boldsymbol{\theta}) = -E\left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\log L(\boldsymbol{\theta})\right)$$

no need to remember

- Some examples for ARMA models: $I(\boldsymbol{\theta})^{-1}$ is
  - AR(1): $1-\phi_1^2$
  - AR(2): $\begin{pmatrix} 1-\phi_1^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}$
  - MA(1): $1-\theta_1^2$
  - MA(2): $\begin{pmatrix} 1-\theta_1^2 & \theta_1(1+\theta_2) \\ \theta_1(1+\theta_2) & 1-\theta_2^2 \end{pmatrix}$
  - ARMA(1,1):

$$\frac{1+\phi_1\theta_1}{(\phi_1+\theta_1)^2}\begin{pmatrix} (1-\phi_1^2)(1+\phi_1\theta_1) & -(1-\phi_1^2)(1-\theta_1^2) \\ -(1-\phi_1^2)(1-\theta_1^2) & (1-\theta_1^2)(1+\phi_1\theta_1) \end{pmatrix}$$

LSE for ARMA(p, q)

- Recall from the prediction from the infinite past;

$$\widetilde{X}_t = \widetilde{P}_{t-1} X_t = -\sum_{j=1}^{\infty} \pi_j X_{t-j}$$

$$\text{MSPE} = E(X_t - \widetilde{X}_t)^2 = \sigma^2 = v_{t-1}, \quad r_{t-1} = 1$$

- This suggests that even for the prediction for the finite past, as $t \to \infty, r_t \to 1$. Thus, MLE is approximately equal to minimize

$$\sum_{r=1}^{n} (X_t - \widetilde{X}_t)^2$$

$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$

$Z_t = \sum \pi_j X_{t-j}$

- Equivalently, observe that

$$X_t - \widetilde{X}_t = Z_t - \sum_{j=1}^{\infty} \pi_j X_{t-j} - \widetilde{X}_t = Z_t$$

LSE for ARMA(p, q)

- LSE is given by

$$\boxed{\widehat{\boldsymbol{\theta}}^{LSE} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{t=1}^{n} Z_t^2 = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{t=1}^{n} (X_t - \widetilde{X}_t)^2}$$

- Example: ARMA(1,1)

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$$

$$\iff Z_t = X_t - \phi X_{t-1} - \theta Z_{t-1}$$

Data version: minimize $\sum_{t=1}^{n} z_t^2$

$$\begin{aligned}
z_1 &= x_1 - \phi \overset{0}{x_0} - \theta \overset{0}{z_0} \\
z_2 &= x_2 - \phi x_1 - \theta z_1 \\
z_3 &= x_3 - \phi x_2 - \theta z_2 \\
&\vdots
\end{aligned}$$

LSE for ARMA(p, q)

- We do not know x_0 and z_0, how to estimate them? Estimate by expected value gives $x_0 = 0$ and $z_0 = 0$. This is called the conditional sum of squares (CSS).
- Some variations:
  - $\widetilde{X}_t$ can be replace by finite sample version $\widehat{X}_t$.
  - Weighted LSE

$$\widehat{\boldsymbol{\theta}}^{WLS} = \operatorname*{argmin}_{\boldsymbol{\theta}} \sum_{t=1}^{n} \frac{Z_t^2}{r_{t-1}} = \operatorname*{argmin}_{\boldsymbol{\theta}} \sum_{t=1}^{n} \frac{(X_t - \widetilde{X}_t)^2}{r_{t-1}}$$

$\frac{1}{r_{t-1}}$: weights

Diagnostics

After fitting ARMA(p, q) model, we need to assess the goodness of fit. Basic idea is similar to regression. "residuals"

- Look at the estimates of parameters. Is it valid, i.e, $\theta \neq 0$? Asymptotic normality of MLE
  Also, check whether coefficients are satisfying stationary/causal/invertible/identifiability conditions.
- Next, check residuals. Residuals are given by

$$\boxed{\widehat{R}_t = \frac{X_t - \widehat{X}_t(\hat{\boldsymbol{\theta}})}{\widehat{\sigma}\sqrt{\hat{r}_{t-1}(\hat{\boldsymbol{\theta}})}} \approx WN(0,1)}$$

AR(p) : $\hat{X}_t = \hat{\phi}_1 X_{t-1} + \hat{\phi}_2 X_{t-2} + \cdots \hat{\phi}_p X_{t-p}$

Diagnostics

General guidelines to check residuals:

- Plot residuals and look for trend / outliers / seasonality / heteroscedasticity.
- Plot sample ACF/PACF for $\widehat{R}_t$. $WN(0,\sigma^2)$
- Check QQ plot for normality. ⇒ b/c MLE assumes Gaussianity
- Perform other formal tests such as
  - IID/WN: Ljung-Box/McLeod-Li / Different sign test
  - No remaining trend: Turning point test / Rank test
  - Gaussianity: Jarque-Bera test

However, $\widehat{R}_t \approx WN(0,1)$ are indeed from that the true model is ARMA(p, q) model. But, in practice, we do not know the orders of p and q.

Order selection by information criteria

- Similar to regression analysis, we can do order selection by looking at residuals, parameter estimates together with ACF/PACF.

  For pure AR(p), PACF cuts off at lag p
  For pure MA(q), ACF cuts off at lag q
- Alternatively, consider that
  - Too many parameters (overfitting) $\Rightarrow$ accurate fit, but model is hard to interpret and estimation can be unstable.
  - Too smaller parameters (under fitting) $\Rightarrow$ lack of fit, estimation increases mean squared error.
  - Best model will have smaller error and few parameters (parsimonious).

  Need some balance between

penalization =

measure of fit	+	# of parameters
(Goodness of fit)	+	(model complexicy)
log-likelihood / SSE		

$\Rightarrow$ It leads to (automatic) model selection by information criteria.

Order selection by information criteria

$$(\text{Approx. SSE}) + (\text{penalty on \# of parameters})$$

Several examples of information criteria:

most basic

$$\text{AIC} : -2\log L_n(\widehat{\boldsymbol{\theta}}) + 2m$$

~ under iid

$$\text{AICC} : -2\log L_n(\widehat{\boldsymbol{\theta}}) + \frac{2mn}{n-m+1}$$

corrects bias caused by time dependence

$$\text{BIC} : -2\log L_n(\widehat{\boldsymbol{\theta}}) + m\log n$$

$$m : \ p+q+1 \ \ (\text{total \# of parameters})$$

Best model: minimize information criteria

- $-2\log L_n(\widehat{\boldsymbol{\theta}})$ approximates SSE for MVN
- AIC (Akaike Information Criteria), AICC (AIC bias corrected) for TS, BIC (Bayesian Information Criteria)

Forecasting ARMA(p, q) processes

1. Select order of p, q
2. estimate parameters using MLE
3. Forecasting

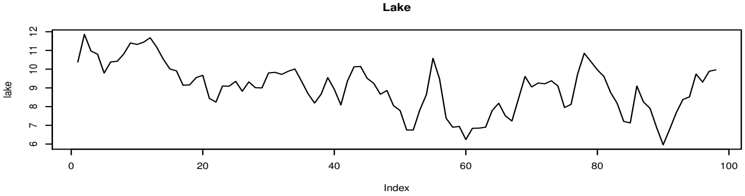
- Once, you find the best model and having parameter estimates, then we can do forecasting.
- Exactly apply $P_n X_{n+h}$ based on the finite past prediction formula and obtain MSPE.
- Approximately 95% CI is given by

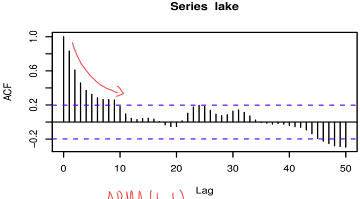
$$P_n X_{n+h} \pm 1.96\sqrt{\text{MSPE}}$$

Computation will be done by R using

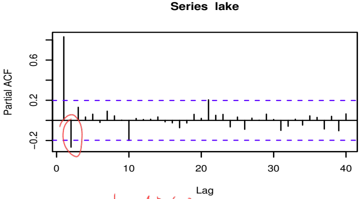
- `arima()` MLE estimation
- `forecast()` in `forecast` package

Lake data analysis





ARMA(1,1)



maybe AR(2)

Which ARMA model would you prefer here?

Lake data analysis

AR(2) parameter estimates from YW

gives Yule-Walker Equation results

```
> ar.yw(lake, aic=FALSE, order.max=2, demean=FALSE)
Coefficients:
      1        2
 1.0747  -0.0923

Order selected 2  sigma^2 estimated as  2.7
```

include mean

Fitting MLE gives

AR(2) 이니까

```
> ar2.out = arima(lake, order=c(2,0,0), include.mean=TRUE)
> ar2.out
Coefficients:
         ar1      ar2  intercept
      1.0436  -0.2495     9.0473
s.e.  0.0983   0.1008     0.3319

sigma^2 estimated as 0.4788:  log likelihood = -103.63,  aic = 215.27
```

Lake data analysis

Order selection by information criteria. `forecast` library is useful

```
>library(forecast)
> fit=auto.arima(lake, d=0)
```
automatically select the best order form

```
Series: lake
ARIMA(1,0,1) with non-zero mean

Coefficients:
         ar1     ma1  intercept
      0.7449  0.3206     9.0555
s.e.  0.0777  0.1135     0.3501

sigma^2 estimated as 0.4749:  log likelihood=-103.25
AIC=214.49   AICc=214.92   BIC=224.83
```

Test whether coefficients are zero:

```
> 2*(1-pnorm(fit$coef/(sqrt(diag(fit$var.coef)))))
        ar1         ma1   intercept
0.000000000 0.004745202 0.000000000
```

p-values are $<.05$, hence coefficients are not equal to zero.

Lake data analysis

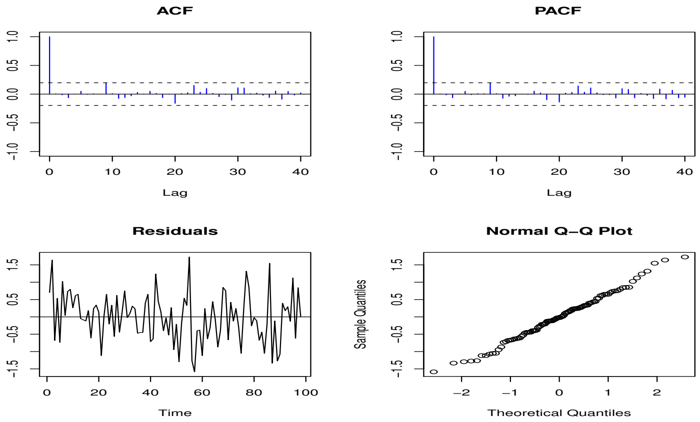
Formal testing for residuals:

```
> library(itsmr)
> test(residuals(fit))
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                  Q ~ chisq(20)     10.14    0.9656
McLeod-Li Q                  Q ~ chisq(20)     16.43    0.6899
Turning points T      (T-64)/4.1 ~ N(0,1)        69    0.2266
Diff signs S        (S-48.5)/2.9 ~ N(0,1)        50    0.6015
Rank P          (P-2376.5)/162.9 ~ N(0,1)      2083    0.0716
```

What about IID assumption? Any remaining trend?

Lake data analysis

Diagnostic plots are "residuals"



Lake data analysis

Forecasting next 30 observations will give you ARMA(1,1)

```
> detach("package:itsmr")
> library(forecast)
> forecast(fit, 30)
    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
 99       9.733373 8.850180 10.61657 8.382646 11.08410
100       9.560436 8.269866 10.85100 7.586680 11.53419
101       9.431615 7.962965 10.90027 7.185508 11.67772
102       9.335656 7.776946 10.89437 6.951814 11.71950
103       9.264177 7.657671 10.87068 6.807237 11.72112
104       9.210932 7.578508 10.84336 6.714356 11.70751
105       9.171270 7.524641 10.81790 6.652969 11.68957
106       9.141726 7.487268 10.79618 6.611451 11.67200
107       9.119718 7.460932 10.77850 6.582824 11.65661
108       9.103325 7.442142 10.76451 6.562765 11.64388
109       9.091113 7.428602 10.75362 6.548522 11.63370
110       9.082017 7.418769 10.74526 6.538299 11.62574
```

Lake data analysis