# Week 7 Lecture Note

Hyoeun Lee

## Estimation strategy

- We need to have specific values of p,d and q for the ARIMA model
- We assume the *dth* difference of the original time series is a stationary series
- Series $X_1, X_2, \ldots, X_n$ is an appropriate difference of the original series and will be our *observed* stationary process
- We will discuss: Method of moments, Least squares and Maximum Likelihood estimators

# ARMA Series Parameter Estimation

# ARMA Series Parameter Estimation

We assume

- we have $n$ observations, $X_1, \ldots, X_n$,
- from an ARMA$(p, q)$ process, in which $p$, $q$ are known
- Goal: estimate the parameters,
  - $\mu$,
  - $\phi_1, \ldots, \phi_p$,
  - $\theta_1, \ldots, \theta_q$, $\sigma_W^2$.

# Method of Moment Estimator

# Method of Moment Estimator

- Easiest but not very efficient

population moments.

- Equate sample moments to theoretical moments and solve resulting equations for unknown parameters

$$\frac{1}{n}\sum_{t=1}^{n} x_t^k$$

$$E[x_t^k].$$

sample $k^{th}$ moment.

theoretical moment

# AR(p) case

$\gamma(h) = Cov(X_t, X_{t-h}) = Cov(\phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t, X_{t-h})$ $h = 1, \ldots, p,$

$= Cov(\phi_1 X_{t-1}, X_{t-h}) + Cov(\phi_2 X_{t-2}, X_{t-h}) \cdots + Cov(\phi_p X_{t-p}, X_{t-h})$

$= \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2) + \cdots + \phi_p \gamma(h-p).$

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t$$

**Yule-walker equation:**

$$\boxed{\rho(h) = \phi_1 \rho(h-1) + \ldots + \phi_p \rho(h-p), \quad h = 1, 2, \ldots, p}$$

$$\boxed{\sigma_W^2 = \gamma(0)[1 - \phi_1 \rho(1) - \cdots - \phi_p \rho(p)]}$$

# Yule-Walker Estimation

The estimator obtained by replacing

- $\gamma(0)$ with its estimate, $\hat{\gamma}(0)$,
- and $\rho(h)$ with its estimate, $\hat{\rho}(h)$,

are called the Yule-Walker Estimators.

# Recall: Autocovariance and Autocorrelation Estimation of (Stationary) series

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X})$$

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

# Yule-walker estimation for AR(1)

$$\boxed{X_t - \mu = \phi(X_{t-1} - \mu) + W_t}$$

, mean estimate is $\hat{\mu} = \bar{X}$,

and $\rho(1) = \phi \underbrace{\rho(0)}_{=1} = \phi$,

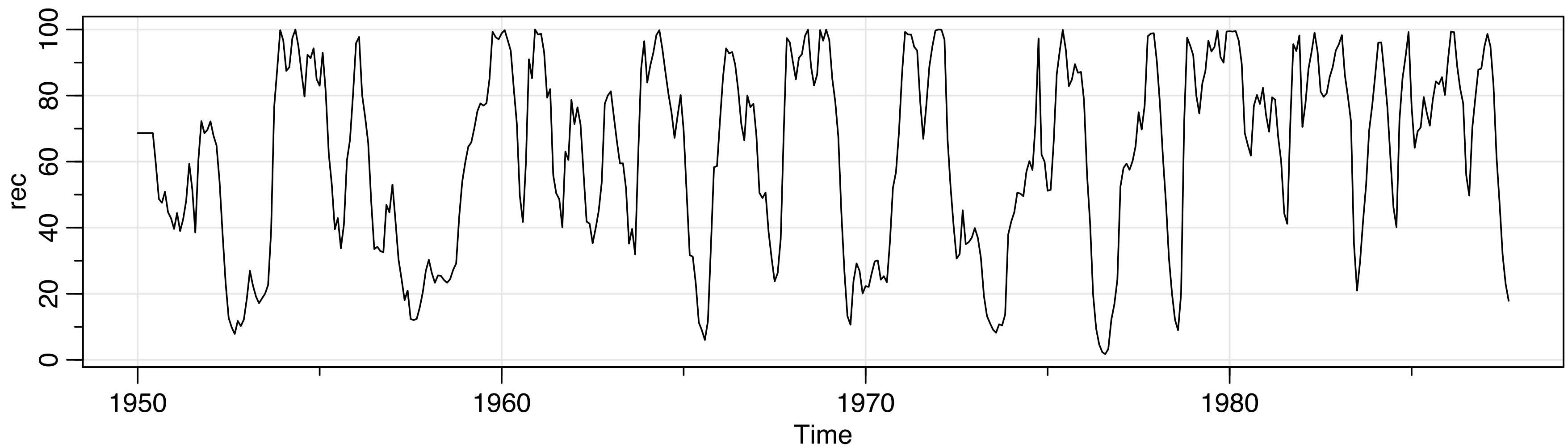
so

$$\hat{\phi} = \hat{\rho}(1) = \frac{\sum_{t=1}^{n-1}(X_{t+1} - \bar{X})(X_t - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

$$\hat{\sigma}_W^2 = \hat{\gamma}(0)[1 - \hat{\phi}^2]$$

# Example: Recruitment Series

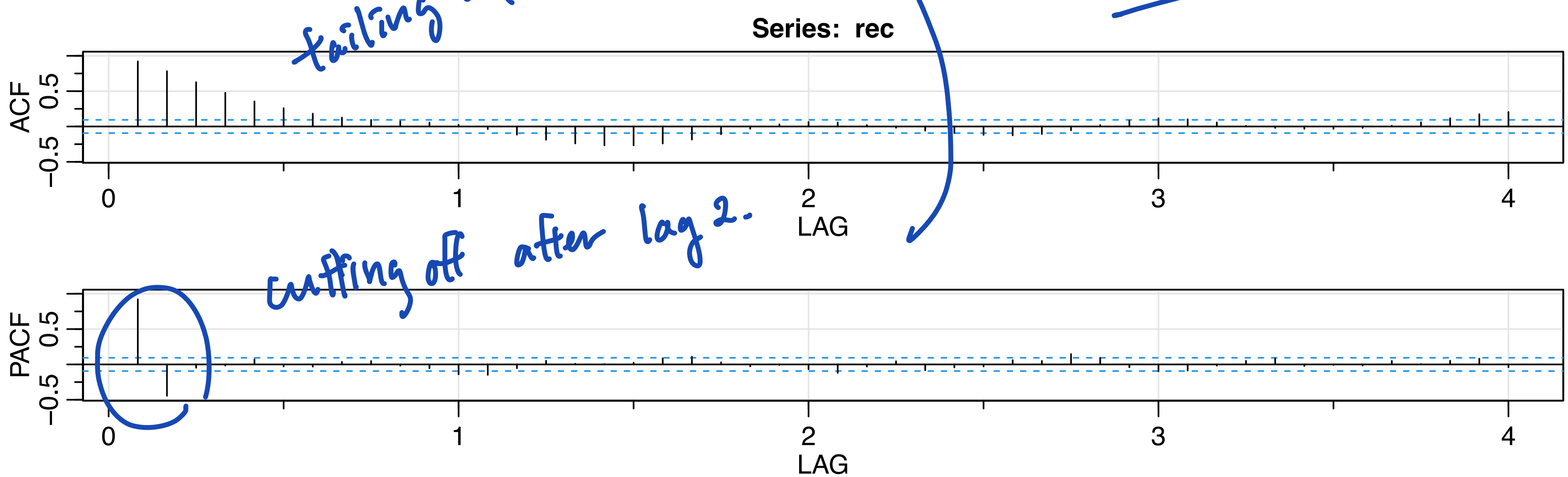
```
tsplot(rec)
```




```
acf2(rec)
```

tailing off.

→ AR(2).




cutting off after lag 2.

$$\hat{\rho}(1) = \hat{\phi}_1 \hat{\rho}(0) + \hat{\phi}_2 \hat{\rho}(1)$$
$$\hat{\rho}(2) = \hat{\phi}_1 \hat{\rho}(1) + \hat{\phi}_2 \hat{\rho}(0)$$

$$\Rightarrow \begin{bmatrix} \hat{\rho}(0) & \hat{\rho}(1) \\ \hat{\rho}(1) & \hat{\rho}(0) \end{bmatrix} \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = \begin{bmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \end{bmatrix}.$$

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = \begin{bmatrix} \hat{\rho}(0) & \hat{\rho}(1) \\ \hat{\rho}(1) & \hat{\rho}(0) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \end{bmatrix}$$

$\hat{\rho}(1) = 0.92$, $\hat{\rho}(2) = 0.78$. Use this information to find $\hat{\phi}_1, \hat{\phi}_2$.

1, 0.92, 0.92, 1

```r
mat.A = matrix(c(0.92, 1, 1, 0.92), nrow = 2)
mat.A
```

```
##      [,1] [,2]
## [1,] 0.92 1.00
## [2,] 1.00 0.92
```

$$\begin{bmatrix} 1 & 0.92 \\ 0.92 & 1 \end{bmatrix}$$

```r
phi.hat = solve(mat.A, c(0.78, 0.92))
phi.hat
```

0.92 0.78.

```
## [1] 1.3177083 -0.4322917
```

$$\begin{bmatrix} 1 & 0.92 \\ 0.92 & 1 \end{bmatrix} \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = \begin{bmatrix} 0.92 \\ 0.78 \end{bmatrix}.$$

```r
rec.yw = ar.yw(rec, order = 2)
rec.yw$x.mean  # mean estimate
```

$\hat{\mu} = \bar{X}$.

```
## [1] 62.26278
```

```r
rec.yw$ar  # phi parameter estimates
```

```
## [1]  1.3315874 -0.4445447
```

```r
sqrt(diag(rec.yw$asy.var.coef))  # their standard errors
```

```
## [1] 0.04222637 0.04222637
```

```r
rec.yw$var.pred  # error variance estimate
```

```
## [1] 94.79912
```

# Yule-Walker AR(p) (Matrix Notation)

$R_p$ : $p \times p$ Matrix $\quad R_p(i,j) = \rho(i-j)$. , $\quad \rho_p : \begin{bmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{bmatrix}$.

$$\begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(p) \\ \rho(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho(1) \\ \rho(p) & \cdots & \rho(1) & \rho(0) \end{bmatrix}$$

$$\phi = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix}$$

$$\hat{\phi} = \underset{\sim}{\hat{R}_p^{-1}} \, \underset{\sim}{\hat{\rho}_p} .$$

$$\hat{\sigma}_w^2 = [1 - \hat{\rho}_p^T \hat{\phi}] \hat{\gamma}(0).$$

# MoM Estimation for MA(1)

$$X_t = W_t + \theta W_{t-1},$$

with $|\theta| < 1$, can be written as

$$X_t = -\sum_{j=1}^{\infty} (-\theta)^j X_{t-j} + W_t$$

The first two population autocovariances are:

$$\gamma(0) = \sigma_W^2 (1 + \theta^2)$$

$$\gamma(1) = \sigma_W^2 \theta$$

$$\hat{\rho}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{\hat{\theta}}{1 + \hat{\theta}^2}.$$

For invertible MA(1),

$|\rho(1)| < \frac{1}{2}$.

Two solutions exits, pick the invertible one.

- If $|\hat{\rho}(1)| \leq \frac{1}{2}$, solutions exist.
- If $|\hat{\rho}(1)| > \frac{1}{2}$, real solutions do not exist.

$\rho(1) = \frac{0.9}{1+0.9^2} = 0.497$.

```r
set.seed(2)
ma1 = arima.sim(list(order = c(0, 0, 1), ma = 0.9), n = 50)
acf1(ma1, plot = FALSE)[1]
```

```
## [1] 0.5066599
```

$> 0.5$

Maximum Likelihood Estimator.

'most likely'

Likelihood: Joint density of $x_1, \dots, x_n$

MLE

AR(1).

$$\left.\begin{aligned} &X_t = \phi X_{t-1} + W_t, \quad |\phi| < 1, \\ &W_t \sim N(0, \sigma_w^2). \end{aligned}\right\}$$

$$L(\phi, \sigma_w^2) = f(x_1, \dots, x_n \mid \phi, \sigma_w^2).$$

$$f(x_1, \cdots, x_n)$$

$$= f(x_1) f(x_2 | x_1) f(x_3 | x_2, x_1) \cdots\cdots f(x_n | x_{n-1}, \cdots, x_1).$$

For AR(1),

$$= f(x_1) f(x_2|x_1) f(x_3|x_2) \cdots f(x_n|x_{n-1}).$$

for $t = 2, \cdots, n$. $\quad x_t = \underline{\phi x_{t-1}} + \underline{w_t}$.

$$x_t | x_{t-1} \sim N(\phi x_{t-1}, \sigma_w^2).$$

$t=1$, $f(x_1)$? Causal Representation, $x_1 = \sum_{j=0}^{\infty} \phi^j w_{1-j}$,

$$Var(x_1) = \frac{\sigma_w^2}{1-\phi^2} . \left( = \sigma_w^2 (1+\phi^2+\phi^4+\phi^6 \cdots ) \right).$$

$$x_1 \sim N\left(0, \frac{\sigma_w^2}{1-\phi^2}\right).$$

$$L(\phi, \sigma_w^2) = \frac{\sqrt{1-\phi^2}}{\sqrt{2\pi}\sqrt{\sigma_w^2}} \exp\left(-\frac{1-\phi^2}{2\sigma_w^2} x_1^2\right) \prod_{t=2}^{n} \left\{ \frac{1}{\sqrt{2\pi}\sqrt{\sigma_w^2}} \exp\left(-\frac{(x_t - \phi x_{t-1})^2}{2\sigma_w^2}\right) \right\}$$

$$L(\phi, \sigma_w^2) = (2\pi\sigma_w^2)^{-n/2} (1-\phi^2)^{1/2} \exp\left\{ -\frac{S(\phi)}{2\sigma_w^2} \right\}$$

$$S(\phi) := \boxed{(1-\phi^2) x_1^2} + \sum_{t=2}^{n} (x_t - \phi x_{t-1})^2.$$

↳ unconditional Sum of Squares.

Conditional Sum of Squares:

$$S_c(\phi) = \sum_{t=2}^{n} (x_t - \phi x_{t-1})^2.$$

Conditional Likelihood:

$$L(\phi, \sigma_w^2) = (2\pi\sigma_w^2)^{-(n-1)/2} \exp\left\{ -\frac{S_c(\phi)}{2\sigma_w^2} \right\}$$

$$\hat{\phi} = \frac{\sum_{t=2}^{n} x_t x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2}. \quad \hat{\sigma}_w^2 = \frac{S_c(\hat{\phi})}{(n-1)}.$$

large sample size(n) ⇒ Unconditional & Conditional equiv.

small n → unconditional MLE.

for AR → conditional LS method = OLS method.

ar.ols

ARMA(p,q) & MA(q): OLS cannot be used.

Gauss-Newton Method.

$$S(\beta) = \sum_{t=1}^{n} w_t^2(\beta).$$

ARMA(p,q). $\beta := [\phi_1, \cdots, \phi_p, \theta_1, \cdots \theta_q]$.

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + w_t + \sum_{k=1}^{q} \theta_k w_{t-k} .$$

$$w_t(\beta) = x_t - \sum_{j=1}^{p} \phi_j x_{t-j} - \sum_{k=1}^{q} \theta_k w_{t-k}(\beta)$$

problem : we don't observe $x_t$, $w_t$ for $t \le 0$.

$\boxed{\text{Conditional LS}}$ → condition on $x_1, \cdots, x_p$ $(p>0)$.

and set $w_p = w_{p-1} = \cdots = w_{p+1-q} = 0$ (for $q>0$)

ARMA(1,1)

$$x_t = \phi x_{t-1} + \theta w_{t-1} + w_t .$$

condition on $x_1$.

set $w_1 = 0$

$$\left\{\begin{aligned} w_2 &= x_2 - \phi x_1 - \theta w_1 = x_2 - \phi x_1 \\ w_3 &= x_3 - \phi x_2 - \theta w_2 \\ &\vdots \\ w_n &= x_n - \phi x_{n-1} - \theta w_{n-1} . \end{aligned}\right\} \rightarrow S_c = \sum_{t=2}^{n} w_t^2 .$$

$$S(\beta) = \sum_{t=1}^{n} w_t^2(\beta) \quad , \quad S_c(\beta) = \sum_{t=p+1}^{n} w_t^2(\beta).$$

---

ARMA, MA $\to$ Gauss-Newton Method

Iterative method to solve minimizing

conditional error SS $\quad S_c(\beta) = \sum_{t=p+1}^{n} w_t^2(\beta)$

Gauss-Newton Method for MA(1)

$$x_t = w_t + \theta w_{t-1}.$$

$$w_t(\theta) = x_t - \theta w_{t-1}(\theta), \quad t = 1, \dots, n.$$

condition: $\quad w_0(\theta) = 0.$

Goal: find $\theta$ minimizes $S_c(\theta) = \sum_{t=1}^{n} w_t^2(\theta)$.

$$S_c(\theta) = \sum_{t=1}^{n} w_t^2(\theta) \approx \sum_{t=1}^{n} \left( w_t(\theta_0) - (\theta - \theta_0) z_t(\theta_0) \right)^2$$

$$w_t(\theta) \approx w_t(\theta_0) + \left. \frac{\partial w_t(\theta)}{\partial \theta} \right|_{\theta=\theta_0} (\theta - \theta_0)$$

$$z_t(\theta_0) = - \left. \frac{\partial w_t(\theta)}{\partial \theta} \right|_{\theta=\theta_0}.$$

$$\frac{\partial}{\partial\theta} w_t(\theta) = -w_{t-1}(\theta) - \theta\frac{\partial w_{t-1}(\theta)}{\partial\theta}, \quad t=1,\cdots n.$$

With $\frac{\partial w_0(\theta)}{\partial\theta} = 0.$

$$z_t(\theta) = w_{t-1}(\theta) - \theta z_{t-1}(\theta), \quad t=1,\cdots n.$$

$$z_0(\theta) = 0.$$

$$S_c(\theta) = \sum_{t=1}^{n} w_t^2(\theta) \approx \boxed{\sum_{t=1}^{n}\Big(\underbrace{w_t(\theta_0)}_{y_t} - \underbrace{(\theta-\theta_0)}_{\beta}\underbrace{z_t(\theta_0)}_{z_t.}\Big)^2} := Q(\theta).$$

$$y_t = \beta z_t + \varepsilon_t.$$

$$\hat{\beta} = \widehat{(\theta-\theta_0)} = \frac{\sum_{t=1}^{n} z_t(\theta_0)\, w_t(\theta_0)}{\sum_{t=1}^{n} z_t^2(\theta_0)}.$$

$$\boxed{\hat{\theta} = \theta_0 + \frac{\sum_{t=1}^{n} z_t(\theta_0)\, w_t(\theta_0)}{\sum_{t=1}^{n} z_t^2(\theta_0)}}.$$

$$\theta_{(j+1)} = \theta_{(j)} + \frac{\sum_{t=1}^{n} z_t(\theta_{(j)})\, w_t(\theta_{(j)})}{\sum_{t=1}^{n} z_t^2(\theta_{(j)})}.,$$

Stop when $|\theta_{(j+1)} - \theta_{(j)}|$

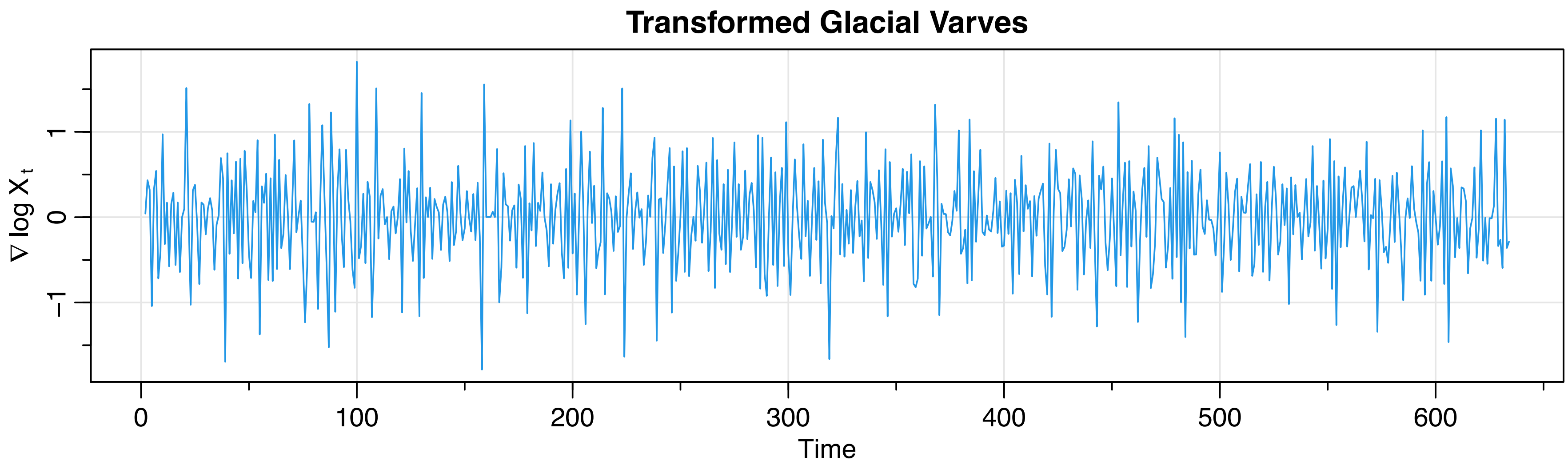
$$|Q(\theta_{(j+1)}) - Q(\theta_{(j)})|$$

are smaller than some preset amount.

# Conditional Least Square

# Conditional Least Square

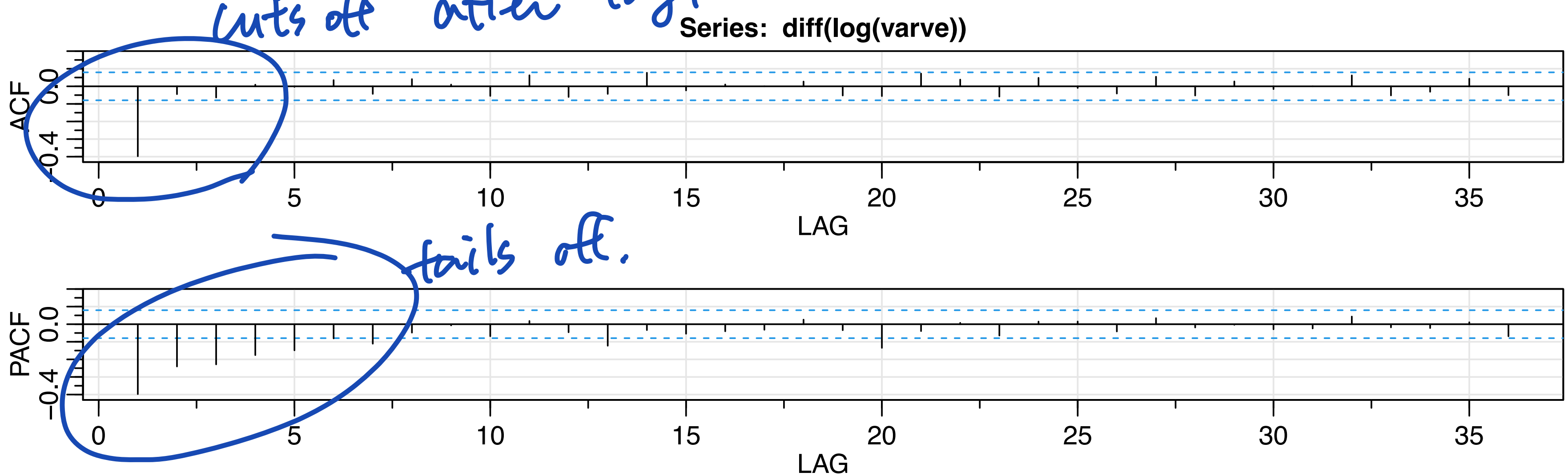
```
tsplot(diff(log(varve)), col = 4, ylab = expression(nabla
```




```
acf2(diff(log(varve)))
```

MA(1).

cuts off after lag1




tails off.

$W_t$. $Z_t$

```
x = diff(log(varve))  # data
r = acf1(x, 1, plot = FALSE)  # acf(1)
w = c(0)
z = c(0)
Sc = c(0)
Sz = c(0)
Szw = c(0)
para = c(0)
# c(0) -> w -> z -> Sc -> Sz -> Szw -> para # initialize
num = length(x)  # = 633
## Estimation
para[1] = (1 - sqrt(1 - 4 * (r^2)))/(2 * r)  # MoM
```

$S_c = \sum W_t^2$

$S_z = \sum Z_t^2$,

$S_{zw} = \sum Z_t W_t$.

$\dfrac{\theta}{1+\theta^2} = r$.

$r\theta^2 - \theta + r = 0$

$\dfrac{1 - \sqrt{1-4r^2}}{2r}$ $\theta_0$.

$W_t = x_t - \theta W_{t-1}$.

$$\theta_{(j+1)} = \theta_{(j)} + \frac{\sum_{t=1}^{n} z_t(\theta_{(j)})\, w_t(\theta_{(j)})}{\sum_{t=1}^{n} z_t^2(\theta_{(j)})}.$$

```
niter = 12
for (j in 1:niter) {
    for (i in 2:num) {
        w[i] = x[i] - para[j] * w[i - 1]
        z[i] = w[i - 1] - para[j] * z[i - 1]
    }
    Sc[j] = sum(w^2)
    Sz[j] = sum(z^2)
    Szw[j] = sum(z * w)
    para[j + 1] = para[j] + Szw[j]/Sz[j]
}
```

$$z_t(\theta) = w_{t-1}(\theta) - \theta z_{t-1}(\theta).$$

$$\theta_{j+1} = \theta_j + \frac{\sum z_t(\theta_j)\, w_t(\theta_j)}{\sum z_t^2(\theta_j)}.$$

```r
# Results
cbind(iteration = 1:niter - 1, thetahat = para[1:niter], Sc
```

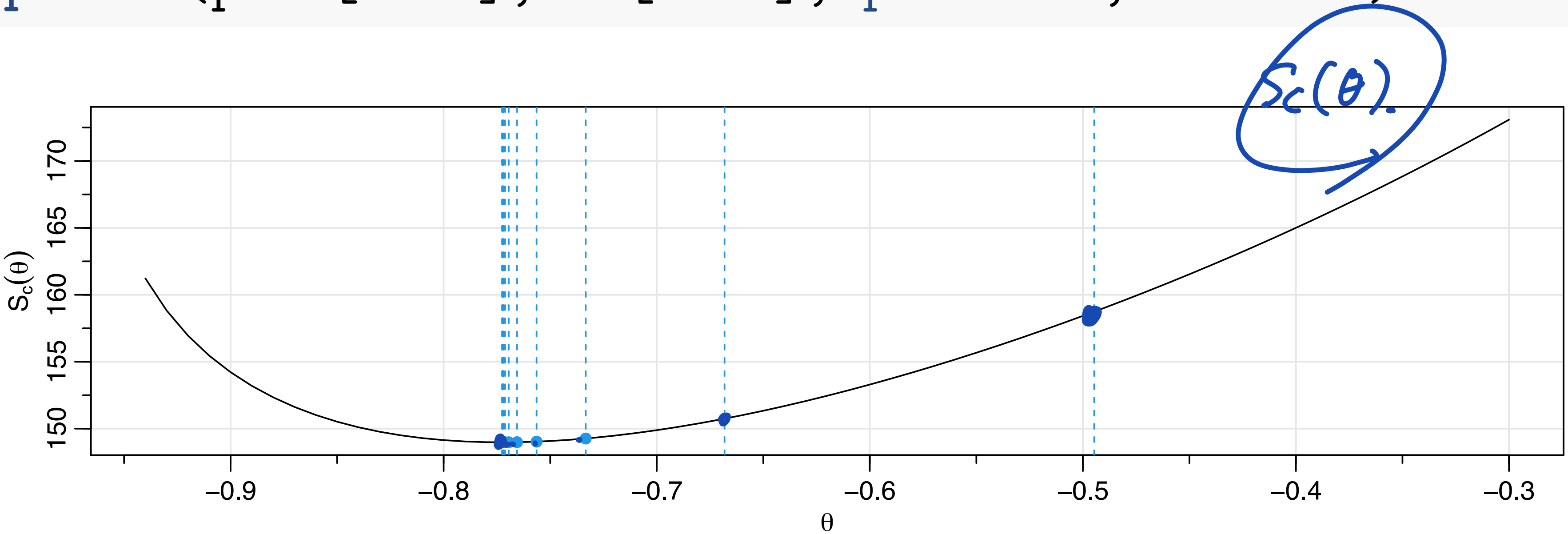
```
##         iteration   thetahat     Sc       Sz
##  [1,]           0 -0.4946886 158.7393 171.2397
##  [2,]           1 -0.6681759 150.7468 235.2660
##  [3,]           2 -0.7333163 149.2644 300.5618
##  [4,]           3 -0.7563893 149.0309 336.8234
##  [5,]           4 -0.7655639 148.9897 354.1729
##  [6,]           5 -0.7694700 148.9818 362.1665
##  [7,]           6 -0.7711856 148.9803 365.8014
##  [8,]           7 -0.7719497 148.9800 367.4456
##  [9,]           8 -0.7722921 148.9799 368.1875
## [10,]           9 -0.7724460 148.9799 368.5220
## [11,]          10 -0.7725153 148.9799 368.6727
## [12,]          11 -0.7725465 148.9799 368.7406
```

$\Sigma w_t^2$. $\Sigma z_t^2$ $j=$ $j=$ $S_c(\hat{\theta})$.

```r
## Plot conditional SS
cSS <- w <- c(0)
th = -seq(0.3, 0.94, 0.01)
for (p in 1:length(th)) {
```

$\hat{\theta} = -0.773.$ $S_c(\hat{\theta}) = 148.98.$ $\hat{\sigma}_w^2 = \frac{S_c(\hat{\theta})}{n-1} = \frac{148.98}{632} = 0.236\ldots$

```r
tsplot(th, cSS, ylab = expression(S[c](theta)), xlab = expr
abline(v = para[1:12], lty = 2, col = 4)  # add previous re
points(para[1:12], Sc[1:12], pch = 16, col = 4)
```



# Unconditional Lest Squares, MLE

sarima( ~~~~~~ , p= __ , d= ___ , q= ___ ,

no.constant=TRUE ).

# Unconditional Lest Squares, MLE

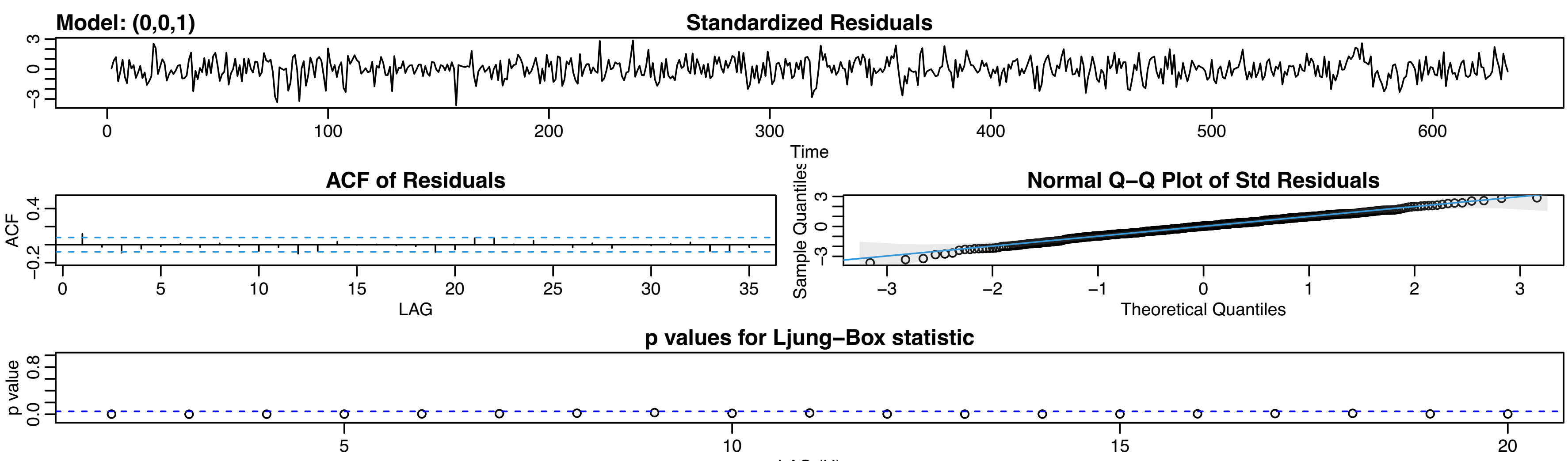
```
sarima(diff(log(varve)), p = 0, d = 0, q = 1, no.constant = TRUE)
```

```
## initial  value -0.551778
## iter   2 value -0.671626
## iter   3 value -0.705973
## iter   4 value -0.707314
## iter   5 value -0.722372
## iter   6 value -0.722738
## iter   7 value -0.723187
## iter   8 value -0.723194
## iter   9 value -0.723195
## iter   9 value -0.723195
## iter   9 value -0.723195
## final  value -0.723195
## converged
```

Conditional SS.

```
## initial  value -0.722700
## iter   2 value -0.722702
## iter   3 value -0.722702
## iter   3 value -0.722702
## iter   3 value -0.722702
## final  value -0.722702
## converged
```

unconditional SS (MLE).



# Large Sample Distributions

large n.

AN : approximately normal for large n.

AR(1) : $\hat{\phi}_1 \sim AN\left[\phi_1, \underbrace{n^{-1}(1-\phi_1^2)}_{\text{variance}}\right].$

$100(1-\alpha)\%$ CI : $\hat{\phi}_1 \pm z_{\alpha/2} \sqrt{\dfrac{1-\hat{\phi}_1^2}{n}}$.

AR(2):

$$\hat{\phi}_1 \sim AN\left[\phi_1, \underbrace{n^{-1}(1-\phi_2^2)}\right], \quad \hat{\phi}_2 \sim AN\left[\phi_2, n^{-1}(1-\phi_2^2)\right].$$

MA(1)

$$\hat{\theta}_1 \sim AN\left[\theta_1, n^{-1}(1-\theta^2)\right].$$

MA(2)

$$\hat{\theta}_1 \sim AN\left[\theta_1, n^{-1}(1-\theta_2^2)\right], \quad \hat{\theta}_2 \sim AN\left[\theta_2, n^{-1}(1-\theta_2^2)\right]$$

AR(1) $\to$ we fit AR(2)

$\to$ what happens?

overfit

$Var(\hat{\phi}_1) = n^{-1}(1-\phi_1^2)$ If fit AR(1).

# Overfitting

$Var(\hat{\phi}_1) \approx n^{-1}(1-\phi_2^2)$. If fit AR(2).

$= n^{-1}$