

3. ARMA Models

3.1 Basic Properties

Consider the ARMA(p, q) process $\{X_t\}$,

$$\Phi(B)X_t = \Theta(B)Z_t$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, where

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = \prod_{j=1}^p (1 - \xi_j z)$$

$$\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q = \prod_{j=1}^q (1 - \omega_j z)$$

We assume that there are no common factors in $\Phi(z)$ in $\Theta(z)$.

- $\{X_t\}$ is causal and stationary if and only if $|\xi_j| < 1, j = 1, \dots, p$
- $\{X_t\}$ is invertible if and only if $|\omega_j| < 1, j = 1, \dots, q$.

3.1 Basic Properties

Note that by using the power series expansion of

$$\Psi(z) = \frac{\Theta(z)}{\Phi(z)}$$

we can construct the linear process (MA(∞)) representation

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

We must have

$$\begin{aligned}(1 - \phi_1 z - \cdots - \phi_p)(\psi_0 + \psi_1 z + \cdots +) \\ = (1 + \theta_1 z + \cdots + \theta_q z^q)\end{aligned}$$

so equating coefficients, we have $\psi_0 = 1$, and for $1, 2, \dots$, we have the recursion

$$\psi_j = \theta_j + \sum_{k=1}^p \phi_k \psi_{j-k}$$

3.1 Basic Properties

By a similar argument, under invertibility, we can construct an $\text{AR}(\infty)$ representation

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

where $\pi_0 = 1$, and $j = 0, 1, 2, \dots$

$$\pi_j = -\phi_j - \sum_{k=1}^q \theta_k \pi_{j-k}$$

3.1 Basic Properties

ACF: The ACF $\gamma_X(h)$ is given by previous results for linear processes; for integer h

$$\gamma_X(h) = \mathbb{E}[X_t X_{t+h}] = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}$$

Example (MA(q))

By the linear process representation, $\psi_0 = 1$, $\psi_j = \theta_j$ for $j = 1, \dots, q$, and $\psi_j = 0$ for $j > q$, so therefore

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} \quad 0 \leq h \leq q$$

and zero otherwise.

3.1 Basic Properties

Example (ARMA(1, 1))

Suppose we have the ARMA(1, 1) process based on a $WN(0, \sigma^2)$ process, with $|\phi| < 1$. Then from the result on page 106, recall that

$$\begin{aligned}\psi_0 &= 1 \\ \psi_j &= (\theta + \phi)\phi^{j-1}, j \geq 1.\end{aligned}$$

give the $MA(\infty)$ representation.

3.1 Basic Properties

Example (ARMA(1, 1) stationary and causal)

$$\begin{aligned}\gamma_X(0) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 \\ &= \sigma^2 \left[\psi_0^2 + \sum_{j=1}^{\infty} \psi_j^2 \right] \\ &= \sigma^2 \left[1 + (\theta + \phi)^2 \sum_{j=0}^{\infty} \phi^{2j} \right] \\ &= \sigma^2 \left[1 + \frac{(\theta + \phi)^2}{1 - \phi^2} \right]\end{aligned}$$

3.1 Basic Properties

Example (ARMA(1, 1) stationary and causal)

$$\begin{aligned}\gamma_X(1) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} \\ &= \sigma^2 \left[(\theta + \phi) + \sum_{j=1}^{\infty} \phi^{2j-1} \right] \\ &= \sigma^2 \left[(\theta + \phi) + (\theta + \phi)^2 \phi \sum_{j=0}^{\infty} \phi^{2j} \right] \\ &= \sigma^2 \left[(\theta + \phi) + \frac{\phi(\theta + \phi)^2}{1 - \phi^2} \right]\end{aligned}$$

3.1 Basic Properties

Example (ARMA(1, 1) stationary and causal)

For $h \geq 2$

$$\begin{aligned}\gamma_X(h) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \\ &= \sigma^2 \phi \sum_{j=0}^{\infty} \psi_j \psi_{j+h-1} \\ &= \phi \gamma_X(h-1)\end{aligned}$$

3.1 Basic Properties

For a direct computation for the general ARMA case, note that

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

yields, multiplying through by X_{t-k} , $k = 0, 1, 2, \dots$,

$$\begin{aligned} X_t X_{t-k} - \phi_1 X_{t-1} X_{t-k} - \cdots - \phi_p X_{t-p} X_{t-k} \\ = Z_t X_{t-k} + \theta_1 Z_{t-1} X_{t-k} + \cdots + \theta_q Z_{t-q} X_{t-k}. \end{aligned}$$

Taking expectations we have for $j = 0, 1, \dots, p$

$$\mathbb{E}[X_{t-j} X_{t-k}] = \mathbb{E}[X_{t+k-j} X_t] = \gamma_X(k-j)$$

so for the left-hand side we have

$$\gamma_X(k) - \phi_1 \gamma_X(k-1) - \cdots - \phi_p \gamma_X(k-p).$$

3.1 Basic Properties

For the right hand side, using the $\text{MA}(\infty)$ form

$$X_{t-k} = \sum_{l=0}^{\infty} \psi_l Z_{t-k-l}.$$

Thus for $j = 0, 1, \dots, q$,

$$\mathbb{E}[X_{t-k}Z_{t-j}] = \mathbb{E}\left[\sum_{l=0}^{\infty} \psi_l Z_{t-k-l}Z_{t-j}\right] = \sum_{l=0}^{\infty} \psi_l \mathbb{E}[Z_{t-k-l}Z_{t-j}]$$

Now

$$\mathbb{E}[Z_{t-k-l}Z_{t-j}] = \begin{cases} \sigma^2 & l = j - k \\ 0 & \text{otherwise} \end{cases}$$

3.1 Basic Properties

Therefore, for fixed k , as $\psi_j = 0$ if $j < 0$, we have

$$\mathbb{E}[X_{t-k}Z_{t-j}] = \sigma^2\psi_{j-k}$$

for $j = k, \dots, q$, and zero otherwise. Hence, writing $\theta_0 = 1$, we have

$$\mathbb{E}\left[\sum_{j=0}^q \theta_j Z_{t-j} X_{t-k}\right] = \sigma^2 \sum_{j=k}^q \theta_j \psi_{j-k} = \sigma^2 \sum_{j=0}^{q-k} \theta_{j+k} \psi_j$$

Note that if $k > q$, $\theta_{j+k} = 0$ for all $j \geq 0$, so in this case the right hand side is zero.

3.1 Basic Properties

If $m = \max\{p, q + 1\}$, we therefore have the infinite system of equations

$$\gamma_X(k) - \sum_{j=1}^p \phi_j \gamma_X(k-j) = \begin{cases} \sigma^2 \sum_{j=0}^m \theta_{j+k} \psi_j & 0 \leq k < m \\ 0 & k \geq m \end{cases}$$

This system can be solved using two approaches.

3.1 Basic Properties

- ▶ **Direct approach:** this is a system of linear difference equations for which the analytic solution exists, and is expressed via the roots $\zeta_1, \dots, \zeta_p$ of the AR polynomial equation $\Phi(z) = 0$

$$\gamma_X(h) = \sum_{j=1}^p \alpha_j \zeta_j^{-h} \quad h \geq m - p$$

where the initial conditions determine $\alpha_1, \dots, \alpha_p$ and $\gamma_X(0), \dots, \gamma_X(h - p - 1)$ (see Brockwell and Davis, p 90).

- ▶ **Recursive approach:** The first $p + 1$ equations form a linear system in the autocovariances $\gamma_X(0), \dots, \gamma_X(p)$, which can be solved using matrix inversion, Levinson-Durbin, or innovations algorithm. The remaining autocovariances $\gamma_X(p + 1), \gamma_X(p + 1), \dots$ can be computed using the recursion.

3.1 Basic Properties

In R, the function `ARMAacf` can be used to compute the acf for given parameter settings

```
ARMAacf(ar = numeric(0), ma = numeric(0),  
lag.max = r, pacf = FALSE)
```

Note also the function `ARMAtoMA` to obtain the $MA(\infty)$ coefficients.

3.1 Basic Properties

Autocovariance Generating Function

The *autocovariance generating function* (acvgf) can also be used for computing the ACV sequences. For a real sequence $\{a_n\}$, the generating function

$$G(z) = \sum_n a_n z^n \quad z \in \mathbb{C}$$

is defined for $|z|$ in a neighbourhood of 1, that is

$$\frac{1}{r} < |z| < r$$

for $r > 1$.

3.1 Basic Properties

Suppose that $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, and

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

Suppose that there exists an $r > 1$ such that

$$\sum_{j=-\infty}^{\infty} |\psi_j| z^j \quad \frac{1}{r} < |z| < r$$

is convergent. Then define the acvgf for $\{X_t\}$ by

$$G_X(z) = \sum_{h=-\infty}^{\infty} \gamma_X(h) z^h \quad \frac{1}{r} < |z| < r$$

3.1 Basic Properties

Recall that

$$\gamma_X(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|h|}$$

Therefore

$$\begin{aligned} G_X(z) &= \sigma^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|h|} z^h \\ &= \sigma^2 \left[\sum_{j=-\infty}^{\infty} \psi_j^2 + \sum_{h=1}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} (z^h + z^{-h}) \right] \\ &= \sigma^2 \left[\sum_{j=-\infty}^{\infty} \psi_j z^j \right] \left[\sum_{h=-\infty}^{\infty} \psi_h z^{-h} \right] \\ &= \sigma^2 \Psi(z) \Psi(z^{-1}), \text{ say.} \end{aligned}$$

3.1 Basic Properties

For the stationary ARMA(p, q) model

$$\Phi(B)X_t = \Theta(B)Z_t$$

where $\Phi(z) \neq 0$ for $|z| = 1$, we have

$$X_t = \{\Phi(B)\}^{-1} \Theta(B)Z_t = \Psi(B)Z_t$$

say, so

$$G_X(z) = \sigma^2 \Psi(z) \Psi(z^{-1}) = \sigma^2 \frac{\Theta(z) \Theta(z^{-1})}{\Phi(z) \Phi(z^{-1})}$$

which is well defined and finite as the denominator is non zero.

3.1 Basic Properties

Example (WN($0, \sigma^2$))

Note that if $\{Z_t\} \sim \text{WN}(0, \sigma^2)$,

$$G_Z(z) = \sigma^2$$

and that this is a unique characteristic of white noise processes.

In general, the ACV and acvgf are not unique.

3.1 Basic Properties

Note: We have seen that the $\text{ARMA}(p, q)$ process

$$\Phi(B)X_t = \Theta(B)Z_t$$

is

- **causal** if the AR reciprocal roots satisfy

$$|\xi_j| < 1 \quad j = 1, 2, \dots, p$$

- **invertible** if the MA reciprocal roots satisfy

$$|\omega_j| < 1 \quad j = 1, 2, \dots, q$$

If the process is non-causal and/or non-invertible, we can transform to a causal and invertible process with the same autocovariance sequence.

3.1 Basic Properties

Example (MA(1))

If $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, and

$$X_t = Z_t - 2Z_{t-1}$$

so that

$$\Theta(z) = 1 - 2z,$$

as $|2| > 1$, the process is non-invertible. But let

$$Z_t^* = (1 - B/2)^{-1}(1 - 2B)Z_t.$$

Then

$$(1 - B/2)Z_t^* = (1 - 2B)Z_t = X_t$$

3.1 Basic Properties

Example (MA(1))

That is

$$\Phi(B)Z_t^* = \Theta(B)Z_t$$

where

$$\Phi(z) = 1 - \frac{1}{2}z \quad \Theta(z) = 1 - 2z.$$

Thus

$$Z_t^* = \Psi(B)Z_t.$$

where

$$\Psi(z) = \frac{\Theta(z)}{\Phi(z)} = \frac{(1 - 2z)}{(1 - z/2)}$$

3.1 Basic Properties

Example (MA(1))

We have

$$Z_t^* = (1 - B/2)^{-1} X_t = \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^j X_{t-j}$$

and the acvgf for $\{Z_t^*\}$ is

$$G_{Z^*}(z) = \sigma^2 \Psi(z) \Psi(z^{-1}) = \sigma^2 \frac{(1 - 2z)(1 - 2z^{-1})}{(1 - z/2)(1 - 1/2z)} = 4\sigma^2$$

so therefore $\{Z_t^*\} \sim \text{WN}(0, 4\sigma^2)$. Thus

$$X_t = \Phi(B)Z_t^* = Z_t^* - \frac{1}{2}Z_{t-1}^*$$

has an **invertible** representation in terms of $\{Z_t^*\}$.

3.1 Basic Properties

For a general transformation approach for ARMA(p, q), if

$$\Phi(B)X_t = \Theta(B)Z_t$$

where $\Phi(z) \neq 0$, $\Theta \neq 0$ when $|z| = 1$, then there exist polynomials

$$\Phi^*(z) \quad \Theta^*(z)$$

and a white noise process $\{Z_t^*\}$ such that

$$\Phi(B)^* X_t = \Theta(B)^* Z_t^*$$

is a causal and invertible representation.

3.1 Basic Properties

Suppose that the reciprocal roots are ordered such that

$$\xi_{r+1}, \dots, \xi_p \quad \omega_{s+1}, \dots, \omega_q$$

lie outside of the unit circle. Then let

$$\begin{aligned}\Phi^*(z) &= \Phi(z) \prod_{j=r+1}^p \left(\frac{1 - \xi_j z}{1 - \xi_j^{-1}} \right) \\ \Theta^*(z) &= \Theta(z) \prod_{j=s+1}^q \left(\frac{1 - \omega_j z}{1 - \omega_j^{-1} z} \right)\end{aligned}$$

and consider

$$Z_t^* = \{\Theta^*(B)\}^{-1} \Phi^*(B) X_t$$

3.1 Basic Properties

After some algebra, it follows that

$$G_{Z^*}(z) = \sigma^2 \left\{ \prod_{j=r+1}^p |\xi_j|^2 \right\} \left\{ \prod_{j=s+1}^q |\omega_j|^2 \right\}$$

and hence that $\{Z_t^*\}$ is a white noise process and

$$\Phi(B)^* X_t = \Theta(B)^* Z_t^*$$

By construction, this is a causal and invertible representation of $\{X_t\}$.

3.1 Basic Properties

Partial Autocorrelation: The PACF, $\alpha_X(h)$, defined by

$$\alpha_X(0) = 1$$

$$\alpha_X(h) = \phi_{h,h}, \quad h \geq 1.$$

$\phi_{h,h}$ is the last component of

$$\phi_h = \Gamma_h^{-1} \gamma_{1:h}$$

Can carry out testing for the estimated PACF sequence as for ACF sequence.

3.1 Basic Properties

Forecasting: The innovations algorithm can be used for ARMA(p, q). Let $m = \max(p, q)$, and define $\{W_t\}$

$$W_t = \begin{cases} \frac{1}{\sigma} X_t & t = 1, 2, \dots, m \\ \frac{1}{\sigma} \Phi(B) X_t & t > m \end{cases}$$

Let

$$\kappa(i, j) = \mathbb{E}[W_i W_j] = \begin{cases} \frac{1}{\sigma^2} \gamma_X(i - j) & 1 \leq i, j \leq m \\ \frac{1}{\sigma^2} \gamma_X(i - j) & \min(i, j) \leq m \leq \max(i, j) \leq 2m \\ \sum_{l=0}^q \theta_l \theta_{l+|i-j|} & \min(i, j) \geq m \\ 0 & \text{otherwise} \end{cases}$$

3.1 Basic Properties

The innovations algorithm yields

$$\hat{W}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{n,j} (W_{n+1-j} - \hat{W}_{n+1-j}) & 1 \leq n \leq m \\ \sum_{j=1}^q \theta_{n,j} (W_{n+1-j} - \hat{W}_{n+1-j}) & n > m \end{cases}$$

Note that $\theta_{n,j} = 0$ if $n \geq m$ and $j > q$.

3.1 Basic Properties

Denoting by

$$\hat{W}_{n+1} = \hat{P}_n W_{n+1} \qquad \hat{X}_{n+1} = \hat{P}_n X_{n+1}$$

it follows that

$$\hat{W}_t = \begin{cases} \frac{1}{\sigma} X_t & t = 1, 2, \dots, m \\ \frac{1}{\sigma} (X_t - \phi_1 \hat{X}_{t-1} - \dots - \phi_p \hat{X}_{t-p}) & t > m \end{cases}$$

so for all t , $X_t - \hat{X}_t = (W_t - \hat{W}_t)/\sigma$, and hence

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}) & 1 \leq n \leq m \\ \sum_{j=1}^p \phi_j X_{n+1-j} + \sum_{j=1}^q \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}) & n > m \end{cases}$$

3.2 Estimation for ARMA models

Estimation of the parameters of ARMA models proceeds using one of a number of different approaches

- Yule-Walker approach
- Burg's algorithm
- Maximum likelihood

Previous approaches use moment-based approaches in a non-parametric fashion.

3.2 Estimation for ARMA models

Yule-Walker approach for AR(p) models: If

$$\Phi(B)X_t = Z_t$$

then using the previous technique, we premultiply by X_{t-k} for $k = 0, \dots, p$ and take expectations, to yield

$$\gamma_X(k) = \phi_1 \gamma_X(k-1) + \dots + \gamma_X(k-p) \quad k = 0, \dots, (p-1)$$

or equivalently, if $\phi_{1:p} = (\phi_1, \dots, \phi_p)^\top$,

$$\Gamma_p \phi_{1:p} = \gamma_{1:p}$$

and

$$\sigma^2 = \gamma_X(0) - \phi_{1:p}^\top \gamma_{1:p}$$

3.2 Estimation for ARMA models

Replacing the theoretical quantities by sample-based analogues, these equations yield estimates of $\phi_{1:p}$ and σ^2 ; using the sample correlations, we have

$$\hat{\phi}_{1:p} = \hat{R}_p^{-1} \hat{\gamma}_{1:p}$$

As $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\phi}_{1:p} - \phi_{1:p}) \sim \mathcal{AN}(0, \sigma^2 \Gamma_p)$$

The estimated parameters yield a stationary and causal procedure.

Note that to construct a stationary and causal AR(p) process we may use the relationship to define $\phi_{1:p}$ from $\gamma_X(h)$.

3.2 Estimation for ARMA models

Note: In the $\text{AR}(p)$ case, if the series of models $\text{AR}(1), \text{AR}(2), \dots, \text{AR}(p), \dots$ are fitted, then the sequence of estimates

- $\hat{\phi}_{1,1}$ for $\text{AR}(1)$,
- $\hat{\phi}_{2,1}, \hat{\phi}_{2,2}$ for $\text{AR}(2)$,
- $\dots$
- $\hat{\phi}_{p,1}, \dots, \hat{\phi}_{p,p}$ for $\text{AR}(p)$

coincide with the Levinson-Durbin coefficients.

The constants $\hat{\phi}_{h,h}$, $h = 0, 1, 2, \dots$ yield the PACF $\alpha_X(h)$.

3.2 Estimation for ARMA models

Yule-Walker for ARMA(p, q) Models

If $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, and $\{X_t\}$ is an ARMA(p, q) model

$$\Phi(B)X_t = \Theta(B)Z_t$$

Assuming this model admits a causal linear process representation

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

the equivalent calculation yields, for $0 \leq k \leq p + q$,

$$\gamma_X(k) - \phi_1 \gamma_X(k-1) - \cdots - \phi_p \gamma_X(k-p) = \sigma^2 \sum_{j=k}^q \theta_j \psi_{j-k}.$$

These $p + q + 1$ equations are non-linear in the parameters, and hence more difficult to solve when sample moments are used.

3.2 Estimation for ARMA models

Burg's Algorithm

Burg's algorithm is used to compute the estimates for an $AR(p)$ model by direct computation of the PACF series. For $0 \leq i < n$, define

$$U_i(t) = X_{n+1+i-t} - \hat{X}_{n+1+i-t}^F$$

$$V_i(t) = X_{n+1+i-t} - \hat{X}_{n+1+i-t}^B$$

where $\hat{X}_{n+1+i-t}^F$ and $\hat{X}_{n+1+i-t}^B$ are the forward and backward forecasts of $x_{n+1+i-t}$ based on the preceding/following i observations.

3.2 Estimation for ARMA models

Then $U_0(t) = V_0(t) = X_{n+1-t}$, and

$$U_i(t) = U_{i-1}(t-1) - \phi_{i,i} V_{i-1}(t)$$

$$V_i(t) = V_{i-1}(t) - \phi_{i,i} U_{i-1}(t-1)$$

This yields an estimation procedure: we choose $\phi_{i,i}$ to minimize

$$\sigma_i^2 = \frac{1}{2(n-i)} \sum_{t=i+1}^n (u_i(t)^2 + v_i(t)^2)$$

that is, the sums of forward and backward prediction errors.
This minimization can be achieved analytically.

3.2 Estimation for ARMA models

Initialize: Set

$$d(1) = \sum_{t=2}^n (u_0(t-1)^2 + v_0(t)^2) = \sum_{t=2}^n (x_{n+2-t}^2 + x_{n+1-t}^2)$$

Recursion: For $i = 1, 2, \dots, p$, set

$$\hat{\phi}_{i,i} = \frac{2}{d(i)} \sum_{t=i+1}^n u_i(t-1)v_i(t)$$

$$d(i+1) = (1 - \hat{\phi}_{i,i}^2)d(i) - v_i(i+1)^2 - u_i(n)^2$$

$$\sigma_i^2 = \frac{(1 - \hat{\phi}_{i,i}^2)d(i)}{2(n-i)}$$

These moment-based estimators have the same asymptotic properties as the previous ones.

3.2 Estimation for ARMA models

Maximum Likelihood Estimation

Maximum likelihood (ML) provides the most efficient approach to estimation, and usually estimators have good small sample properties. However, outside of the Gaussian case, the likelihood function is not often available.

We consider first Gaussian time series $\{X_t\}$, presumed to be zero mean, with autocovariance

$$\mathbb{E}[X_i X_j] = \kappa(i, j)$$

Let

$$\hat{X}_{t+1} = \mathbb{E}[X_{t+1} | X_1, \dots, X_t] = \hat{P}_t X_{t+1} \quad t = 1, 2, \dots$$

which is always available for the Gaussian case.

3.2 Estimation for ARMA models

If $X_{1:t} = (X_1, \dots, X_t)^\top$, then

$$\hat{X}_{t+1} = \gamma_{t+1,t:1} \Gamma_t^{-1} X_{1:t}$$

where

$$\gamma_{t+1,t:1} = \mathbb{E}[X_{t+1} X_{1:t}] \qquad \Gamma_t = \mathbb{E}[X_{1:t} X_{1:t}^\top]$$

are derived from the properties of the general multivariate normal distribution (see Appendix, Brockwell and Davis).

3.2 Estimation for ARMA models

Let

$$\Gamma_n = \mathbb{E}[X_{1:n} X_{1:n}^T]$$

then the likelihood for the parameters of the covariance matrix is

$$L_n(\Gamma_n) = \frac{1}{(2\pi)^{n/2}} |\Gamma_n|^{-1/2} \exp \left\{ -\frac{1}{2} x_{1:n}^T \Gamma_n^{-1} x_{1:n} \right\}.$$

If n is large, computation of

$$|\Gamma_n| = \det \Gamma_n \qquad \Gamma_n^{-1}$$

may not be straightforward numerically. However, the innovations algorithm can be used to compute the determinant and inverse.

3.2 Estimation for ARMA models

The innovations algorithm returns the quantities

$$\theta_{i,j}, i = 1, 2, \dots, n-1, j = 1, \dots, i.$$

If

$$C_n = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 1 & 0 & 0 & \cdots & 0 \\ \theta_{2,2} & \theta_{2,1} & 1 & 0 & \cdots & 0 \\ \theta_{3,3} & \theta_{3,2} & \theta_{3,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \theta_{n-1,n-4} & \cdots & 1 \end{bmatrix}$$

Then, as before $X_{1:n} = C_n(X_{1:n} - \hat{X}_{1:n})$, with the elements of the $n \times 1$ vector $(X_{1:n} - \hat{X}_{1:n})$ uncorrelated, with covariance matrix $D_n = \text{diag}(v_0, v_1, \dots, v_{n-2})$.

3.2 Estimation for ARMA models

Therefore

$$\Gamma_n = C_n D_n C_n^\top$$

and

$$X_{1:n}^\top \Gamma_n^{-1} X_{1:n} = (X_{1:n} - \hat{X}_{1:n})^\top D_n^{-1} (X_{1:n} - \hat{X}_{1:n}) = \sum_{j=1}^n (X_j - \hat{X}_j)^2 / v_{j-1}$$

and

$$|\Gamma_n| = \prod_{j=0}^{n-1} v_j.$$

and thus the innovations algorithm facilitates the Gaussian likelihood computation for any n .

3.2 Estimation for ARMA models

For ARMA(p, q), we saw that

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j}) & 1 \leq n \leq m \\ \sum_{j=1}^p \phi_j X_{n+1-j} + \sum_{j=1}^q \theta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j}) & n > m \end{cases}$$

where $m = \max(p, q)$, and

$$v_n = \mathbb{E}[(X_{n+1} - \hat{X}_{n+1})^2] = \sigma^2 r_n$$

say.

3.2 Estimation for ARMA models

ML estimation requires that the likelihood is maximized, which in turn requires that the innovations algorithm is run for each new parameter setting given by the iterative algorithm selected (Newton-Raphson, Quasi-Newton etc.).

Recall that the likelihood $L(x_{1:n})$ depends on the parameters via the ACF $\gamma_X(h)$, which appears in the calculation via Γ_n , or the parameters output by the innovations algorithm.

The properties of the ML estimates are well-established.

3.2 Estimation for ARMA models

We have that

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

where

$$R(\phi, \theta) = \prod_{j=1}^n r_{j-1} \qquad S(\phi, \theta) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j(\phi, \theta))^2}{r_{j-1}}$$

Note that estimation of σ^2 and (ϕ, θ) can be separated.

3.2 Estimation for ARMA models

Notes:

1. There is no guarantee that the resulting estimates will lead to a causal or invertible process. These constraints can be imposed during the optimization, or a reparameterization to the reciprocal roots can be performed.
2. The identical optimization can be carried out for **non-Gaussian** processes as well, leading to maximum pseudo-likelihood estimates.

3.2 Estimation for ARMA models

Conditional Likelihood for AR(p)

A *conditional likelihood* approach uses the first p data points to condition upon to facilitate ease of likelihood construction. We have that

$$f(y_{1:n}|\phi, \sigma^2) = f(y_{1:p}|\phi, \sigma^2)f(y_{(p+1):n}|y_{1:p}, \phi, \sigma^2) = L_1(\phi, \sigma^2)L_2(\phi, \sigma^2)$$

say, where

$$f(y_{(p+1):n}|\phi, \sigma^2) = \prod_{j=p+1}^n f(y_j|y_{(j-p-1):(j-1)}, \phi, \sigma^2).$$

The conditional likelihood approach uses only $L_2(\phi, \sigma^2)$ as the objective function.

3.2 Estimation for ARMA models

Least-squares

A least-squares procedure based on $S(\phi, \theta)$, that is

$$(\hat{\phi}, \hat{\theta}) = \arg \min S(\phi, \theta)$$

with

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

can be used.

A *conditional* least-squares approach can also be used for $\text{AR}(p)$ models.

3.2 Estimation for ARMA models

In R the main function for ARMA estimation is `arima`:

```
1 arima(x, order = c(0, 0, 0),  
2       seasonal = list(order = c(0, 0, 0), period = NA),  
3       xreg = NULL, include.mean = TRUE,  
4       transform.pars = TRUE,  
5       fixed = NULL, init = NULL,  
6       method = c("CSS-ML", "ML", "CSS"),  
7       n.cond, optim.method = "BFGS",  
8       optim.control = list(), kappa = 1e6)
```

For example

```
arima(x, order=c(1, 0, 1))
```

fits an ARMA(1,1) model to the data in `x`.

3.2 Estimation for ARMA models

The library `fArma` contains the `armaFit` function, which provides a number of alternative fitting approaches.

Other functions for AR models include:

```
ar  
ar.burg  
ar.yw  
ar.mle
```

3.2 Estimation for ARMA models

Example (Lake Huron data (de-trended))

Four potential models:

- AR(1) `arima(Lake.Yt, order=c(1, 0, 0))`
- AR(2) `arima(Lake.Yt, order=c(2, 0, 0))`
- MA(1) `arima(Lake.Yt, order=c(0, 0, 1))`
- ARMA(1,1) `arima(Lake.Yt, order=c(1, 0, 1))`

3.2 Estimation for ARMA models

Example (Lake Huron data (de-trended))

```
1 > arima(Lake.Yt,order=c(1,0,0))
2 Call:arima(x = Lake.Yt, order = c(1, 0, 0))
3 Coefficients:
4           ar1  intercept
5       0.7829    0.0797
6 s.e.  0.0634    0.3178
7 sigma^2 estimated as 0.4972:  log likelihood = -105.29,
  aic = 216.58
8
9 > arima(Lake.Yt,order=c(2,0,0))
10 Call: arima(x = Lake.Yt, order = c(2, 0, 0))
11 Coefficients:
12           ar1          ar2  intercept
13       1.0047   -0.2919    0.0196
14 s.e.  0.0977    0.1004    0.2351
15 sigma^2 estimated as 0.4571:  log likelihood = -101.25,
  aic = 210.5
```

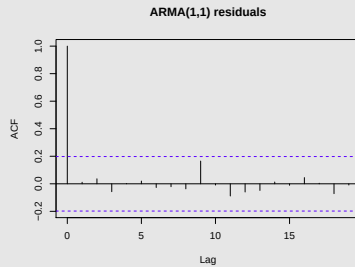
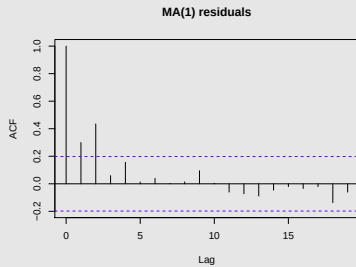
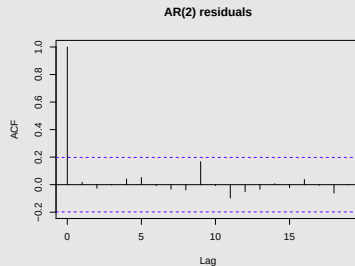
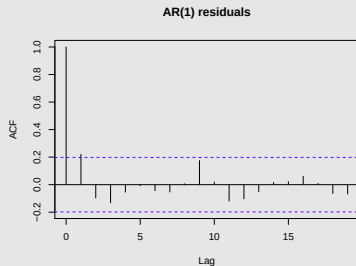
3.2 Estimation for ARMA models

Example (Lake Huron data (de-trended))

```
1 > arima(Lake.Yt,order=c(0,0,1))
2 Call:arima(x = Lake.Yt, order = c(0, 0, 1))
3 Coefficients:
4           ma1    intercept
5      0.7821    -0.0037
6 s.e.  0.0651     0.1390
7 sigma^2 estimated as 0.6013:  log likelihood = -114.6,
  aic = 235.2
8
9 arima(Lake.Yt,order=c(1,0,1))
10 Call:arima(x = Lake.Yt, order = c(1, 0, 1))
11 Coefficients:
12           ar1           ma1    intercept
13      0.6517     0.3571         0.0332
14 s.e.  0.0945     0.1150         0.2610
15 sigma^2 estimated as 0.4572:  log likelihood = -101.26,
  aic = 210.52
```

3.2 Estimation for ARMA models

Example (Lake Huron data (de-trended))



3.2 Estimation for ARMA models

Model selection

A series of ARMA models might be fitted, and then compared with reference to different aspects:

- fidelity to the data
- model complexity
- residuals
- forecasting performance

3.2 Estimation for ARMA models

Akaike Information Criterion (AIC)

The AIC model selection method is widely used; it trades off *fidelity to the observed data* (goodness of fit) with *model complexity* (number of parameters fitted).

For model M with parameters β yielding likelihood $L_M(\beta)$, the AIC is defined as

$$\text{AIC}(M, \hat{\beta}) = -2 \log L_M(\hat{\beta}) + 2d$$

where d is the dimension of β .

For the ARMA(p, q) model, there are $p + q + 1$ parameters $(\phi_{1:p}, \theta_{1:q}, \sigma^2)$.

3.2 Estimation for ARMA models

For the ARMA(p, q) model,

$$\text{AIC}(M_{p,q}, \hat{\beta}) = -2\ell_M(\hat{\phi}_{1:p}, \hat{\theta}_{1:q}, S(\hat{\phi}_{1:p}, \hat{\theta}_{1:q})/n) + 2(p + q + 1)$$

where ℓ_M is the log-likelihood, and

$$\hat{\sigma}^2 = \frac{1}{n} S(\hat{\phi}_{1:p}, \hat{\theta}_{1:q})$$

The model with the **smallest** (ie most negative) AIC is selected.

3.2 Estimation for ARMA models

Notes

- ▶ The AIC is a “within sample” criterion, and fits to the observed data. This might not always reflect the objective of the analysis we might be more interested in forecasting performance.
- ▶ The AIC is sometimes regarded to give an insufficient penalty for complexity, and chooses models that are too complex.
- ▶ As $n \rightarrow \infty$, the AIC is not guaranteed to choose the correct model.

3.2 Estimation for ARMA models

Corrected AIC

To fix the small sample issue, a correction was suggested:

$$AICC(M_{p,q}, \hat{\beta}) = -2\ell_M(\hat{\phi}_{1:p}, \hat{\theta}_{1:q}, S(\hat{\phi}_{1:p}, \hat{\theta}_{1:q})/n) + \frac{2(p+q+1)n}{n-p-q-2}$$

that is the penalty $2d$ is replaced by $2dn/(n-p-q-2)$. This uses a bias-corrected estimator of σ^2 .

For large n , the criterion reverts to the AIC.

3.2 Estimation for ARMA models

Bayesian Information Criterion (BIC)

The BIC (or Schwarz criterion) uses a different penalty

$$BIC(M_{p,q}, \hat{\beta}) = -2\ell_M(\hat{\phi}_{1:p}, \hat{\theta}_{1:q}, S(\hat{\phi}_{1:p}, \hat{\theta}_{1:q})/n) + (p + q + 1) \log n$$

that is the penalty $2d$ is replaced by $d \log n$.

For large n , the criterion correctly selects the true model (provided the true model is one that has been fitted).

3.2 Estimation for ARMA models

Residual checking

The residual sequence $\{\hat{W}_t\}$ estimated by

$$\hat{w}_t = \frac{x_t - \hat{x}_t}{\sqrt{r_{t-1}}}$$

(that is, the one-step ahead prediction errors standardized by their standard deviation) should have the property that $\{\hat{W}_t\}$ is an approximate white noise sequence. In fact, if the model fitted is the correct model

$$W_t \xrightarrow{\text{m.s.}} Z_t \sim \text{WN}(0, \sigma^2)$$

so the observed residual sequence should be approximately uncorrelated, zero mean and variance σ^2 . We can also compute $\hat{R}_t = \hat{W}_t / \hat{\sigma}$, which should be a unit variance white noise process.