Forecasting based on ARMA (3.4 - 3.5)

Suppose we know the set of past values $X_{1:n} = \{X_1, X_2, \dots, X_n\}$ and we want to estimate the future value X_{n+m}. Intuitively, the value we want to know is $E(X_{n+m} \mid X_{1:n})$. Or, we may consider

$$\min_g E\left[(X_{n+m} - g(X_{1:n}))^2 \mid X_{1:n}\right] \qquad (g \text{ is a function of } X_{1:n}$$

$$= \min_g E\left[X_{n+m}^2 - 2X_{n+m}g + g^2 \mid X_{1:n}\right] \qquad \text{given } X_{1:n},\ g \text{ is fixed})$$

$$= \min_g \left(E\left[X_{n+m}^2 \mid X_{1:n}\right] - 2gE\left[X_{n+m} \mid X_{1:n}\right] + g^2\right)$$

$$\Rightarrow \quad g_{min} = E(X_{n+m} \mid X_{1:n}) \overset{let}{=} X_{n+m}^n$$

Consider an AR(p) model $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + W_t$, with W_t iid

Assume $n > p$, then $X_{n+1}^n = E(X_{n+1} \mid X_{1:n}) = \phi_1 X_n + \dots + \phi_p X_{n-p+1}$

$$X_{n+2}^n = E(X_{n+2} \mid X_{1:n}) = E\left[\phi_1 X_{n+1} + \phi_2 X_n + \dots + \phi_p X_{n+2-p} \mid X_{1:n}\right]$$

$$= \phi_1 E(X_{n+1} \mid X_{1:n}) + \phi_2 X_n + \dots + \phi_p X_{n+2-p}$$

$$= (\phi_1^2 + \phi_2) X_n + (\phi_1\phi_2 + \phi_3) X_{n-1} \cdots + (\phi_1\phi_{p-1} + \phi_p) X_{n-(p-2)} + \phi_1\phi_p X_{n-(p-1)}$$

So, we expect X_{n+m}^n is a linear combination of $X_1, \dots, X_n$.

Let $X_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k X_k$, we determine α_k's by the fact that X_{n+m}^n is the solution of $\min_g E\left[(X_{n+m} - g(X_{1:n}))^2 \mid X_{1:n}\right]$ for any given $X_{1:n}$, which implies X_{n+m}^n is also the solution of $\min_g E\left[(X_{n+m} - g(X_{1:n}))^2\right]$

As we know X_{n+m}^n is of the form $\alpha_0 + \sum_{k=1}^n \alpha_k X_k$, we consider

$$\min_{\alpha_0, \alpha_1, \dots, \alpha_n} E\left[\left(X_{n+m} - \left(\alpha_0 + \sum_{k=1}^n \alpha_k X_k\right)\right)^2\right] \qquad (\text{Set } X_0 = 1)$$

Consider $E\left[(X_{n+m} - X_{n+m}^n)^2\right] = E(X_{n+m}^2) - 2\sum_{k=0}^n \alpha_k E(X_{n+m}X_k) + \sum_{k=0}^n \alpha_k^2 E(X_k^2) + 2\sum_{i<j} \alpha_i\alpha_j E(X_iX_j)$

Set $\frac{d}{d\alpha_k} = 0 \Rightarrow -2E(X_{n+m}X_k) + 2\alpha_k E(X_k^2) + 2\sum_{i=0}^{k-1} \alpha_i E(X_iX_k) + 2\sum_{j=k+1}^n \alpha_j E(X_kX_j) = 0$

$$\Rightarrow \quad E\left[\left(X_{n+m} - \sum_{i=0}^n \alpha_i X_i\right) X_k\right] = 0$$

Property 3.3 Best Linear Prediction (BLP) for Stationary Processes

Given data $X_1, \dots, X_n$, the best linear predictor, $X^n_{n+m} = \alpha_0 + \sum_{k=1}^{n} \alpha_k X_k$, of X_{n+m}, for $m \geq 1$, is found by solving

$$E[(X_{n+m} - X^n_{n+m}) X_k] = 0, \quad k = 0, 1, \dots, n,$$

where $X_0 = 1$, for $\alpha_0, \alpha_1, \dots, \alpha_n$.

If $E(X_t) = \mu$, then $E[(X_{n+m} - X^n_{n+m}) X_0] = 0$

$$\Rightarrow E(X_{n+m}) = E(X^n_{n+m}) = \mu$$

$$\therefore \mu = E(X^n_{n+m}) = E\left(\sum_{k=0}^{n} \alpha_k X_k\right) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \mu \Rightarrow \alpha_0 = \mu\left(1 - \sum_{k=1}^{n} \alpha_k\right)$$

$\therefore$ the BLP $\quad X^n_{n+m} = \left(\mu - \sum_{k=1}^{n} \alpha_k \mu\right) + \sum_{k=1}^{n} \alpha_k X_k$

$$\Rightarrow X^n_{n+m} - \mu = \sum_{k=1}^{n} \alpha_k (X_k - \mu)$$

$\therefore$ By considering $X_t - \mu$ as before, we can assume $\mu = 0$, in which case $\alpha_0 = 0$

In particular, we are interested in $X^n_{n+1} = E(X_{n+1} \mid X_1, \dots, X_n)$

let $X^n_{n+1} = \alpha_{n1} X_n + \alpha_{n2} X_{n-1} + \dots + \alpha_{nn} X_1$, we have

$$E\left[\left(X_{n+1} - \sum_{j=1}^{n} \alpha_{nj} X_{n+1-j}\right) X_{n+1-k}\right] = 0, \quad k = 1, \dots, n$$

$$\Rightarrow \gamma(k) = E(X_{n+1} X_{n+1-k}) = \sum_{j=1}^{n} \alpha_{nj} E(X_{n+1-j} X_{n+1-k}) = \sum_{j=1}^{n} \alpha_{nj} \gamma(k-j)$$

$$\Rightarrow \vec{\gamma}_n = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(-1) & \cdots & \gamma(1-n) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(2-n) \\ \vdots & & \ddots & \vdots \\ \gamma(n-1) & \cdots & \cdots & \gamma(0) \end{pmatrix} \begin{pmatrix} \alpha_{n1} \\ \vdots \\ \alpha_{nn} \end{pmatrix} = \vec{\Gamma}_n \vec{\alpha}_n$$

For ARMA models, the fact that $\sigma_w^2 > 0$ and $\gamma(h) \to 0$ as $h \to \infty$ is enough to ensure $\vec{\Gamma}_n$ is positive definite (Note that $Var(a_1 X_1 + \dots + a_n X_n) = \vec{a}^T \vec{\Gamma}_n \vec{a}$)

$\therefore \vec{\alpha}_n = \vec{\Gamma}_n^{-1} \vec{\gamma}_n \quad$ and $\quad X^n_{n+1} = \vec{\alpha}_n^T \vec{X}, \quad \vec{X} = (X_n, \dots, X_1)^T$

let
$$P^n_{n+1} = E(X_{n+1} - X^n_{n+1})^2 = E(X_{n+1}^2) - 2E(X_{n+1} \vec{\alpha}_n^T \vec{X}) + E(\vec{\alpha}_n^T \vec{X} \vec{X}^T \vec{\alpha}_n)$$
$$= \gamma(0) - 2\vec{\gamma}_n^T \vec{\Gamma}_n^{-1} E\begin{pmatrix} X_{n+1} X_n \\ \vdots \\ X_{n+1} X_1 \end{pmatrix} + \vec{\gamma}_n^T \vec{\Gamma}_n^{-1} E(\vec{X}\vec{X}^T) \vec{\Gamma}_n^{-1} \vec{\gamma}_n$$
$$= \gamma(0) - 2\vec{\gamma}_n^T \vec{\Gamma}_n^{-1} \vec{\gamma}_n + \vec{\gamma}_n^T \vec{\Gamma}^{-1} \vec{\gamma}_n$$
$$= \gamma(0) - \vec{\gamma}_n^T \vec{\Gamma}_n^{-1} \vec{\gamma}_n$$

Example 3.19 $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + W_t$ with X_1 observed

$$X_2^1 = E(X_2 | X_1) = \alpha_{11} X_1 \quad \text{where } \alpha_{11} = \vec{\Gamma}_1^{-1} \vec{\gamma}_1 = \frac{\gamma(1)}{\gamma(0)} = \rho(1) \ (= \phi_{11})$$

Recall that for ARMA(p,q) model

$$\gamma(h) - \phi_1 \gamma(h-1) - \dots - \phi_p \gamma(h-p) = 0 \quad \text{for } h \geq \max(p, q+1)$$
$$\gamma(h) - \phi_1 \gamma(h-1) - \dots - \phi_p \gamma(h-p) = \sigma_w^2 \sum_{j=h}^{q} \theta_j \psi_{j-h} \quad \text{for } h < \max(p, q+1)$$

For AR(2) model, $\max(p, q+1) = 2$, $\gamma(h) - \phi_1 \gamma(h-1) - \phi_2 (h-2) = 0$

Put $h=1$, $\gamma(1) - \phi_1 \gamma(0) - \phi_2 \gamma(-1) = (1-\phi_2)\gamma(1) - \phi_1 \gamma(0) = 0$

$$\Rightarrow \rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\phi_1}{1-\phi_2}$$

Now, suppose $X_3^2 = E(X_3 | X_1, X_2) = \alpha_{21} X_2 + \alpha_{22} X_1$

$$\vec{\alpha}_2 = \vec{\Gamma}_2^{-1} \vec{\gamma}_n \Rightarrow \begin{pmatrix} \alpha_{21} \\ \alpha_{22} \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} \gamma(1) \\ \gamma(2) \end{pmatrix}$$

We can check that $\begin{pmatrix} \alpha_{21} \\ \alpha_{22} \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$.

$$E[(X_3 - (\phi_1 X_2 + \phi_2 X_1)) X_k] = E[W_3 X_k] = 0 \quad \text{for } k = 1, 2$$

In general, for causal AR(p) process and $n \geq p$,

$$X_{n+1}^n = \phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_p X_{n-p+1}$$

Property 3.4 The Durbin-Levinson Algorithm

$$\vec{\alpha}_n = \begin{pmatrix} \alpha_{n1} \\ \vdots \\ \alpha_{nn} \end{pmatrix} = \vec{\Gamma}_n^{-1} \vec{\gamma}_n \quad \text{and} \quad P_{n+1}^n = \gamma(0) - \vec{\gamma}_n^T \vec{\Gamma}_n^{-1} \vec{\gamma}_n$$

Can be solved iteratively as follows:

$$\alpha_{00} = 0 \qquad P_1^0 = \gamma(0)$$

$$\alpha_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \alpha_{n-1,k} \rho(n-k)}{1 - \sum_{k=1}^{n-1} \alpha_{n-1,k} \rho(k)}, \qquad P_{n+1}^n = P_n^{n-1} (1 - \alpha_{nn}^2), \quad \text{for } n \geq 1$$

where, for $n \geq 2$,

$$\alpha_{nk} = \alpha_{n-1,k} - \alpha_{nn} \alpha_{n-1,n-k}, \quad k = 1, 2, \dots, n-1$$

Example 3.20 For $n=1$, $\alpha_{11} = \frac{\rho(1)}{1} = \rho(1)$ $\quad P_2^1 = \gamma(0)\cdot(1-\alpha_{11}^2)$

For $n=2$, $\alpha_{22} = \dfrac{\rho(2) - \alpha_{11}\rho(1)}{1 - \alpha_{11}\rho(1)}$, $\alpha_{21} = \alpha_{11} - \alpha_{22}\alpha_{11}$

$$P_3^2 = P_2^1(1-\alpha_{22}^2) = \gamma(0)(1-\alpha_{11}^2)(1-\alpha_{22}^2)$$

In general $P_{n+1}^n = \gamma(0)\prod_{j=1}^{n}(1-\alpha_{jj}^2)$

Property 3.5 The PACF of a stationary process X_t, $\phi_{nn} = \alpha_{nn}$

For AR(p) model and $n=p$, we have

$$\begin{aligned} X_{p+1}^p &= \alpha_{p1}X_p + \alpha_{p2}X_{p-1} + \dots + \alpha_{pp}X_1 \\ &= \phi_1 X_p + \phi_2 X_{p-1} + \dots + \phi_p X_1 \quad \Rightarrow \quad \phi_{pp} = \alpha_{pp} = \phi_p \end{aligned}$$

Consider AR(2), $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + W_t$, we have

$$\gamma(h) - \phi_1\gamma(h-1) - \phi_2\gamma(h-2) = 0 \quad \text{or} \quad \rho(h) - \phi_1\rho(h-1) - \phi_2\rho(h-2) = 0$$

$\rho(0) = 1$, $\rho(1) = \frac{\phi_1}{1-\phi_2}$ $\Rightarrow$ $\rho(2) = \phi_1\rho(1) + \phi_2$

$\rho(3) = \phi_1\rho(2) + \phi_2\rho(1)$

$$\therefore \quad \phi_{11} = \alpha_{11} = \rho(1) = \frac{\phi_1}{1-\phi_2}$$

$$\begin{aligned} \phi_{22} = \alpha_{22} = \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2} &= \frac{\phi_1\left(\frac{\phi_1}{1-\phi_2}\right) + \phi_2 - \left(\frac{\phi_1}{1-\phi_2}\right)^2}{1-\left(\frac{\phi_1}{1-\phi_2}\right)^2} \\ &= \frac{(1-\phi_2)\left(\frac{\phi_1}{1-\phi_2}\right)^2 + \phi_2 - \left(\frac{\phi_1}{1-\phi_2}\right)^2}{1-\left(\frac{\phi_1}{1-\phi_2}\right)^2} = \phi_2 \end{aligned}$$

$$\phi_{21} \overset{\text{def}}{=} \alpha_{21} = \phi_{11} - \phi_{22}\phi_{11} = \frac{\phi_1}{1-\phi_2}(1-\phi_2) = \phi_1$$

$$\phi_{33} = \frac{\rho(3) - \phi_{21}\rho(2) - \phi_{22}\rho(1)}{1 - \phi_{21}\rho(1) - \phi_{22}\rho(2)} = \frac{\rho(3) - \phi_1\rho(2) - \phi_2\rho(1)}{1 - \phi_1\rho(1) - \phi_2\rho(2)} = 0$$

Now, we consider $X_{n+m}^n = E(X_{n+m} \mid X_1, \dots, X_n) = \phi_{n1}^{(m)} X_n + \dots + \phi_{nn}^{(m)} X_1$

By Property 3.3, $\sum_{j=1}^{n} \phi_{nj}^{(m)} E(X_{n+1-j}X_{n+1-k}) = E(X_{n+m}X_{n+1-k})$, $k=1,\dots,n$

or $\sum_{j=1}^{n} \phi_{nj}^{(m)}\gamma(k-j) = \gamma(m+k-1)$ $\quad \boxed{\begin{aligned}\vec{\Phi}_n^{(m)} &= (\phi_{n1}^{(m)}, \dots, \phi_{nn}^{(m)})^T \\ \vec{\gamma}_n^{(m)} &= (\gamma(m), \dots, \gamma(m+n-1))^T\end{aligned}}$

$\Rightarrow \vec{\Gamma}_n \vec{\Phi}_n^{(m)} = \vec{\gamma}_n^{(m)}$ and $\boxed{P_{n+m}^n = E(X_{n+m} - X_{n+m}^n)^2 = \gamma(0) - \vec{\gamma}_n^{(m)T}\vec{\Gamma}_n^{-1}\vec{\gamma}_n^{m}}$

Again, we can compute X_{n+m}^{n} by solving $\vec{\phi}_n^{(m)} = \vec{\Gamma}_n^{-1} \vec{\gamma}_n^{(m)}$. However, it could be very hard to compute $\vec{\Gamma}_n^{-1}$ when n is large.

Another way to compute X_{n+m}^{n} is to apply the Innovations Algorithm in Property 3.6.

We have been assuming $X_{n+m}^{n} = E(X_{n+m} | X_1, \dots, X_n)$ is equal to the BLP of X_{n+m}, i.e. $X_{n+m}^{n} = \phi_{n1}^{(m)} X_n + \phi_{n2}^{(m)} X_{n-1} + \dots + \phi_{nn}^{(m)} X_1$

We have seen that it is true for AR(p) models with $W_t \sim iid(0, \sigma_w^2)$

Actually, it is true for general Gaussian process, e.g. ARMA(p,q) models with $W_t \sim iid\ N(0, \sigma_w^2)$

$\boxed{\text{We will assume } X_t \text{ is a causal and invertible ARMA}(p,q) \text{ process, } \phi(B) X_t = \theta(B) W_t, \text{ where } W_t \sim iid\ N(0,\sigma_w^2) \text{ for forecasting.}}$

Because $X_{n+m} = \phi^{-1}(B)\theta(B) W_{n+m} = \sum_{j=0}^{\infty} \psi_j W_{n+m-j}$, $\psi_0 = 1$, ——— (1)

$W_{n+m} = \theta^{-1}(B)\phi(B) X_{n+m} = \sum_{j=0}^{\infty} \pi_j X_{n+m-j} = X_{n+m} + \sum_{j=1}^{\infty} \pi_j X_{n+m-j}$, $\pi_0 = 1$ ——— (2)

It is easier to compute $\tilde{X}_{n+m} = E(X_{n+m} | X_n, X_{n-1}, \dots, X_1, X_0, X_{-1}, \dots)$ that $X_{n+m}^{n} = E(X_{n+m}|X_1,\dots,X_n)$. The idea is that $\tilde{X}_{n+m}$ should be close to X_{n+m}^{n} when n is large.

Consider $\tilde{X}_{n+m} = \sum_{j=0}^{\infty} \psi_j E(W_{n+m-j} | X_n, X_{n-1}, \dots) = \sum_{j=m}^{\infty} \psi_j W_{n+m-j}$ ——— (3)

$\because E(W_t | X_n, X_{n-1}, \dots) = \begin{cases} 0 & t > n \\ W_t & t \leq n \end{cases}$

Also

$$E(W_{n+m} | X_n, X_{n-1}, \dots) = \sum_{j=0}^{\infty} \pi_j \tilde{X}_{n+m-j} = \tilde{X}_{n+m} + \sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} + \sum_{j=m}^{\infty} \pi_j X_{n+m-j}$$

$$\Rightarrow \tilde{X}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j X_{n+m-j} \quad \text{——— (4)}$$

(1) − (3), $\Rightarrow$ $X_{n+m} - \tilde{X}_{n+m} = \sum_{j=0}^{m-1} \psi_j W_{n+m-j}$

and $P_{n+m}^{n} = E(X_{n+m} - \tilde{X}_{n+m})^2 = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$

Since we only have $X_1, \dots, X_n$, we consider the truncated predictor by setting $\sum_{j=n+m}^{\infty} \pi_j X_{n+m-j} = 0$, i.e. $\underbrace{W_{n+m} = X_{n+m} + \sum_{j=1}^{n+m-1} \pi_j X_{n+m-j}}_{AR(n+m-1)}$

$$\tilde{X}_{n+m}^{n} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j}^{n} - \sum_{j=m}^{n+m-1} \pi_j X_{n+m-j},$$

So that $\tilde{X}_{n+1}^{n} = -\sum_{j=1}^{n} \pi_j X_{n+1-j} = -\pi_1 X_n - \pi_2 X_{n-1} - \dots - \pi_n X_1$

$$\tilde{X}_{n+2}^{n} = -\pi_1 \tilde{X}_{n+1}^{n} - \sum_{j=2}^{n+1} \pi_j X_{n+2-j} = -\pi_1 \tilde{X}_{n+1}^{n} - \pi_2 X_n - \pi_3 X_{n-1} - \dots - \pi_{n+1} X_1$$

For mean square prediction error, we still use $P_{n+m}^{n} = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$ as an estimate.

In terms of original coefficients in $\phi(B) X_t = \theta(B) w_t$, we have

$\boxed{\text{Property 3.7}}$ $\tilde{X}_{n+m}^{n} = \phi_1 \tilde{X}_{n+m-1}^{n} + \dots + \phi_p \tilde{X}_{n+m-p}^{n} + \theta_1 \tilde{w}_{n+m-1}^{n} + \dots + \theta_q \tilde{w}_{n+m-q}^{n}$

where $\tilde{X}_t^n = X_t$ for $1 \le t \le n$ and $\tilde{X}_t^n = 0$ for $t \le 0$

$\tilde{w}_t^n = 0$ for $t \le 0$ or $t > n$ and

$\tilde{w}_t^n = \phi(B) \tilde{X}_t^n - \theta_1 \tilde{w}_{t-1}^n - \dots - \theta_q \tilde{w}_{t-q}^n$ for $1 \le t \le n$

$\boxed{\text{Example 3.24}}$ $X_{n+1} = \phi X_n + W_{n+1} + \theta W_n$, given $X_1, \dots, X_n$

From Property 3.7, $\tilde{X}_{n+m}^{n} = \phi \tilde{X}_{n+m-1}^{n} + \theta \tilde{w}_{n+m-1}^{n}$

Note that $\tilde{w}_t^n = 0$ for $t \le 0$ or $t > n$, and $\tilde{X}_t^n = X_t$ for $1 \le t \le n$

for $m=1$, $\tilde{X}_{n+1}^{n} = \phi X_n + \tilde{w}_n^n$

for $m \ge 2$, $\tilde{X}_{n+m}^{n} = \phi \tilde{X}_{n+m-1}^{n}$,

where $\tilde{w}_t^n = (1-\phi B) \tilde{X}_t^n - \theta \tilde{w}_{t-1}^n = X_t - \phi X_{t-1} - \theta \tilde{w}_{t-1}^n$ for $1 \le t \le n$

$\therefore$ Given $\tilde{w}_0^n = 0$ and $X_0 = 0$, we can compute $\tilde{w}_1^n, \tilde{w}_2^n, \dots, \tilde{w}_n^n$ and hence $\tilde{X}_{n+m}^n$

Note that $X_{n+m} - \tilde{X}_{n+m} = \sum_{j=0}^{m-1} \psi_j W_{n+m-j} \overset{\text{let}}{=} e_{n+m}$

$\therefore E(e_{n+m}) = 0$ $\quad Var(e_{n+m}) = E(e_{n+m}^2) = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2 = P_{n+m}^n$

If $W_t \sim iid\ N(0, \sigma_w^2)$, then $e_{n+m} \sim N(0, Var(e_{n+m}))$ or $X_{n+m} \sim N(\tilde{X}_{n+m}, Var(e_{n+m}))$

and hence a $(1-\alpha)$ prediction interval for X_{n+m} is $\tilde{X}_{n+m} \pm z_{1-\frac{\alpha}{2}} \sqrt{P_{n+m}^n}$,

where $z_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$th quantile of $N(0,1)$.

An approximate 95% prediction interval is $\tilde{X}_{n+m}^n \pm 2\sqrt{P_{n+m}^n}$

From Example 3.12, $\psi_j = (\phi+\theta)\phi^{j-1}$, $\psi_0 = 1$ $\Rightarrow P^n_{n+m} = \sigma_w^2\left(1+(\theta+\phi)^2\sum_{j=1}^{m-1}\phi^{2(j-1)}\right)$

$$= \sigma_w^2\left(1+(\theta+\phi)^2\frac{1-\phi^{2(m-1)}}{1-\phi^2}\right)$$

Now, given a ARMA(p, q) model $\phi(B)X_t = \theta(B)w_t$, we will talk about how to estimate $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ and σ_w^2.

We keep assuming the model is causal and invertible Gaussian ARMA(p,q) i.e. $X_t = \phi^{-1}(B)\theta(B)w_t = \psi(B)w_t$, $w_t = \theta^{-1}(B)\phi(B)X_t = \pi(B)X_t$ and $w_t \sim \text{iid } N(0,\sigma_w^2)$.

We first consider AR(p) model $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + w_t$

$$\Rightarrow E(X_{t-h}X_t) = \phi_1\gamma(h-1) + \dots + \phi_p\gamma(h-p) + E(X_{t-h}w_t)$$

$$\Rightarrow \gamma(h) = \phi_1\gamma(h-1) + \dots + \phi_p\gamma(h-p) \quad \text{for } 1\le h\le p \qquad (3.98)$$

Also $E(X_tX_t) = \phi_1\gamma(1) + \dots + \phi_p\gamma(p) + E(X_tw_t)$

$$\Rightarrow \sigma_w^2 = \gamma(0) - \phi_1\gamma(1) - \dots - \phi_p\gamma(p) \qquad (3.99)$$

$$(3.98) \Rightarrow \vec{\gamma}_p = \begin{pmatrix}\gamma(1)\\ \gamma(2)\\ \vdots\\ \gamma(p)\end{pmatrix} = \begin{pmatrix}\gamma(0) & \gamma(1) & \cdots & \gamma(p-1)\\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2)\\ \vdots & & & \\ \gamma(p-1) & & \cdots & \gamma(0)\end{pmatrix}\begin{pmatrix}\phi_1\\ \vdots\\ \phi_p\end{pmatrix} = \vec{\Gamma}_p\vec{\phi}$$

By replacing $\gamma(h)$ by the sample covariance $\hat\gamma(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h}-\bar X)(X_t-\bar X)$, we get the Yule-Walker estimators

$$\hat\phi_{YW} = \hat\Gamma_p^{-1}\hat\gamma_p \qquad \hat\sigma^2_{w,YW} = \hat\gamma(0) - \hat\phi_{YW}^T\hat\gamma_p = \hat\gamma(0) - \hat\gamma_p^T\hat\Gamma_p^{-1}\hat\gamma_p$$

By factoring $\hat\gamma(0)$, we also have

$$\hat\phi_{YW} = \hat R_p^{-1}\hat\rho_p \qquad \hat\sigma^2_{w,YW} = \hat\gamma(0)\left[1-\hat\rho_p^T\hat R_p^{-1}\hat\rho_p\right],$$

where $\hat R_p = (\hat\rho(k-j))_{1\le j,k\le p}$ and $\hat\rho_p = (\hat\rho(1),\dots,\hat\rho(p))^T$.

$\boxed{\text{Property 3.8}}$ For AR(p), causal Gaussian process,

$$\sqrt{n}(\hat\phi_{YW} - \phi) \xrightarrow{d} N(0, \sigma_w^2\vec{\Gamma}_p^{-1})$$

$$\hat\sigma^2_{w,YW} \xrightarrow{p} \sigma_w^2 \qquad \text{as } n\to\infty$$

If $\phi_p = 0$, we can show that the (p,p)-entry of Γ_p^{-1} is $(\sigma_w^2)^{-1}$ and hence we have $\sqrt{n}(\hat{\phi}_p - 0) \xrightarrow{d} N(0,1)$

Recall that $\phi_{pp} = \phi_p$ for AR(p) model. Therefore we have

Property 3.9 For a causal AR(p) process,

$$\sqrt{n}\,\hat{\phi}_{hh} = \sqrt{n}\,\hat{\phi}_h \xrightarrow{d} N(0,1) \qquad \text{for } h>p$$

Example 3.27 $X_t = 1.5X_{t-1} - 0.75X_{t-2} + W_t$, $n = 144$, $w_t \sim \text{iid } N(0,1)$

For these data, $\hat{\gamma}(0) = 8.903$, $\hat{\rho}(1) = 0.849$ and $\hat{\rho}(2) = 0.519$

$$\therefore \hat{\phi}_{YW} = \begin{pmatrix}\hat{\phi}_1\\ \hat{\phi}_2\end{pmatrix} = \begin{pmatrix}1 & 0.849\\ 0.849 & 1\end{pmatrix}^{-1}\begin{pmatrix}0.849\\ 0.519\end{pmatrix} = \begin{pmatrix}1.463\\ -0.723\end{pmatrix}$$

and $\hat{\sigma}^2_{w,YW} = 8.903\left[1 - (1.463, -0.723)\begin{pmatrix}0.849\\ 0.519\end{pmatrix}\right] = 1.187$

The asymptotic covariance matrix of $\hat{\phi}_{YW}$ is

$$n^{-1}\hat{\sigma}^2_{w,YW}\hat{\Gamma}_2^{-1} = (1.187)(144)^{-1}\hat{\gamma}(0)^{-1}\begin{pmatrix}1 & \hat{\rho}(1)\\ \hat{\rho}(1) & 1\end{pmatrix}^{-1} = \begin{pmatrix}0.058^2 & -0.003\\ -0.003 & 0.058^2\end{pmatrix}$$

Example 3.29 Method of Moments (MM) Estimation

$$X_t = W_t + \theta W_{t-1} = \sum_{j=1}^{\infty}(-\theta)^j X_{t-j} + W_t \qquad \text{for } |\theta|<1$$

Since $\gamma(0) = \sigma_w^2(1+\theta^2)$ and $\gamma(1) = \sigma_w^2\theta$, so an MM estimate of θ is

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

$$\Rightarrow \hat{\theta} = \frac{\hat{\rho}(1)^{-1} \pm \sqrt{\hat{\rho}(1)^{-2} - 4}}{2}$$

So, if $|\hat{\rho}(1)| > \frac{1}{2}$, $\hat{\theta}$ is not real. In such case, the MA(1) model may not hold or $|\rho(1)| \approx \frac{1}{2}$. Suppose $|\hat{\rho}(1)| \le \frac{1}{2}$, then we pick $\hat{\theta} = \frac{1-\sqrt{1-4\hat{\rho}(1)^2}}{2\hat{\rho}(1)}$ so that the MA(1) process is invertible.

From Theorem A.7, $\hat{\rho}(1) \xrightarrow{d} N(\rho(1), n^{-1}W_{11})$, where

$$W_{11} = \sum_{u=1}^{\infty}(\rho(u+1) + \rho(u-1) - 2\rho(1)\rho(u))^2$$

$$= (\rho(0) - 2\rho(1)^2)^2 + \rho(1)^2$$

$$= (1 - 4\rho(1)^2 + 4\rho(1)^4) + \rho(1)^2 = 1 - 3\frac{\theta^2}{(1+\theta^2)^2} + 4\frac{\theta^4}{(1+\theta^2)^4} = \frac{1+\theta^2+4\theta^4+\theta^6+\theta^8}{(1+\theta^2)^4}$$

Now, for any particular value of $\hat{\rho}(1)$ and a differentiable function $g(\cdot)$, by Taylor expansion,

$$g(\hat{\rho}(1)) = g(\rho(1)) + g'(\xi)(\hat{\rho}(1) - \rho(1)) \qquad \xi \in (\hat{\rho}(1), \rho(1))$$

Now, as $n \to \infty$, we have $\hat{\rho}(1) \xrightarrow{P} \rho(1)$ $\quad (\Rightarrow \xi \xrightarrow{P} \rho(1) \Rightarrow g'(\xi) \xrightarrow{P} g(\rho$

and $\quad \hat{\rho}(1) - \rho(1) \xrightarrow{d} N(0, n^{-1} W_{11})$

Together with $g'(\xi) \xrightarrow{d} g(\rho(1))$, we have

$$g(\hat{\rho}(1)) \xrightarrow{d} g(\rho(1)) + g'(\rho(1))\, N(0, n^{-1} W_{11})$$
$$= N\left(g(\rho(1)),\ g'(\rho(1))^2 n^{-1} W_{11}\right)$$

Now consider $\theta = g(\rho(1))$ with $g(z) = \dfrac{1-\sqrt{1-4z^2}}{2z} = \frac{1}{2} z^{-1}(1-(1-4z^2)^{\frac{1}{2}})$

$$g'(z) = -\frac{1}{2} z^{-2}\left(1-(1-4z^2)^{\frac{1}{2}}\right) + \frac{1}{2} z^{-1}\left[-\frac{1}{2}(1-4z^2)^{-\frac{1}{2}}(-8z)\right]$$
$$= -\frac{1-\sqrt{1-4z^2}}{2z^2} + 2\frac{1}{\sqrt{1-4z^2}}$$

Note that $\rho(1) = \dfrac{\theta}{1+\theta^2} \Rightarrow 1 - 4\rho(1)^2 = 1 - \dfrac{4\theta^2}{(1+\theta^2)^2} = \dfrac{1+2\theta^2+\theta^4-4\theta^2}{(1+\theta^2)^2} = \dfrac{(1-\theta^2)^2}{(1+\theta^2)^2}$

$$\Rightarrow \sqrt{1-4\rho(1)^2} = \frac{1-\theta^2}{1+\theta^2} \qquad (\because |\theta|<1)$$

$$\Rightarrow g'(\rho(1)) = -\frac{1-\frac{1-\theta^2}{1+\theta^2}}{2\frac{\theta^2}{(1+\theta^2)^2}} + 2\frac{1+\theta^2}{1-\theta^2}$$

$$= -\frac{(1+2\theta^2+\theta^4)-(1-\theta^4)}{2\theta^2} + \frac{2(1+\theta^2)}{1-\theta^2}$$

$$= -(1+\theta^2)\frac{1-\theta^2}{1-\theta^2} + \frac{2+2\theta^2}{1-\theta^2} = \frac{-(1-\theta^4)+2+2\theta^2}{1-\theta^2} = \frac{(1+\theta^2)^2}{1-\theta^2}$$

$$\therefore \hat{\theta} = g(\hat{\rho}(1)) \xrightarrow{d} N\left(\theta, \frac{1+\theta^2+4\theta^4+\theta^6+\theta^8}{n(1-\theta^2)^2}\right)$$

Maximum Likelihood Estimation

We first focus on the causal AR(1) model $X_t = \mu + \phi(X_{t-1} - \mu) + W_t$ where $|\phi| < 1$ and $W_t \sim iid\ N(0, \sigma_w^2)$. Given $X_1, X_2, \dots, X_n$, we consider the likelihood

$$L(\mu, \phi, \sigma_w^2) = f(X_1, \dots, X_n \mid \mu, \phi, \sigma_w^2)$$
$$= f(X_n \mid X_{n-1}, \dots, X_1, \mu, \phi, \sigma_w^2)\, f(X_{n-1}, \dots, X_1 \mid \mu, \phi, \sigma_w^2)$$

Since X_t does not depend on $X_{t-2}, X_{t-3}, \dots$ given X_{t-1}, so

$$L(\mu, \phi, \sigma_w^2) = f(X_n \mid X_{n-1}) f(X_{n-1} \mid X_{n-2}) \cdots f(X_2 \mid X_1) f(X_1)$$

$W_t \sim iid\ N(0, \sigma_w^2) \Rightarrow$
$$X_t = \mu + \phi(X_{t-1} - \mu) + W_t$$
$$\sim N(\mu + \phi(X_{t-1} - \mu), \sigma_w^2) \quad \text{given } X_{t-1}$$

For $X_1 = \mu + \sum_{j=0}^{\infty} \phi^j W_{1-j}$, $E(X_1) = \mu$ and $Var(X_1) = \sum_{j=0}^{\infty} (\phi^j)^2 = \frac{1}{1-\phi^2}$

$$\therefore L(\mu, \phi, \sigma_w^2) = f(X_1) \prod_{t=2}^{n} f(X_t \mid X_{t-1})$$
$$= \frac{1}{\sqrt{2\pi \sigma_w^2/(1-\phi^2)}} e^{-\frac{(X_1-\mu)^2}{2\sigma_w^2/(1-\phi^2)}} \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi \sigma_w^2}} e^{-\frac{(X_t - \mu - \phi(X_{t-1}-\mu))^2}{2\sigma_w^2}}$$
$$= (2\pi \sigma_w^2)^{-\frac{n}{2}} (1-\phi^2)^{\frac{1}{2}} e^{-\frac{1}{2\sigma_w^2} S(\mu, \phi)},$$

where $S(\mu, \phi) = (1-\phi^2)(X_1 - \mu)^2 + \sum_{t=2}^{n} [(X_t - \mu) - \phi(X_{t-1} - \mu)]^2$ is called the unconditional sum of squares.

Consider
$$\log L(\mu, \phi, \sigma_w^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma_w^2 + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma_w^2} S(\mu, \phi)$$

$\frac{\partial}{\partial \sigma_w^2} = 0 \Rightarrow$
$$-\frac{n}{2}\frac{1}{\sigma_w^2} + \frac{1}{2}\frac{1}{(\sigma_w^2)^2} S(\mu, \phi) = 0 \Rightarrow \hat{\sigma}^2_{w,MLE} = \frac{1}{n} S(\hat{\mu}_{MLE}, \hat{\phi}_{MLE})$$

Putting $\sigma_w^2 = \frac{1}{n} S(\mu, \phi)$ into $\log L(\mu, \phi, \sigma_w^2)$, $\hat{\mu}_{MLE}$ and $\hat{\phi}_{MLE}$ are the values that minimize
$$\ell(\mu, \phi) = \log\left(\tfrac{1}{n} S(\mu, \phi)\right) - \tfrac{1}{n}\log(1-\phi^2)$$

If we estimate μ and ϕ by minimizing $S(\mu, \phi)$, they are called the unconditional least squares estimators.

Since $S(\mu, \phi)$ and $\ell(\mu, \phi)$ are complicated functions of μ and ϕ, people may consider the likelihood with X_1 assumed to be nonrandom.

Since X_1 is non random, we don't have $f(X_1)$ any more and so

$$L(\mu, \phi, \sigma_w^2 \mid X_1) = \prod_{t=2}^{n} f(X_t \mid X_{t-1}) = (2\pi\sigma_w^2)^{-\frac{n-1}{2}} e^{-\frac{1}{2\sigma_w^2} S_c(\mu,\phi)},$$

where $S_c(\mu,\phi) = \sum_{t=2}^{n} [(X_t-\mu) - \phi(X_{t-1}-\mu)]^2$ is the conditional sum of squares

Such likelihood is call the conditional likelihood.

The conditional MLE of σ_w^2 is $\hat{\sigma}_{w,c}^2 = \frac{1}{n-1} S_c(\hat{\mu}_c, \hat{\phi}_c)$ where $\hat{\mu}_c$ and $\hat{\phi}_c$ are the MLE estimators. In particular, plugging $\sigma_w^2 = \frac{1}{n-1} S_c(\mu,\phi)$ into

$$\log L(\mu, \phi, \sigma_w^2 \mid X_1) = -\frac{n-1}{2}\log 2\pi - \frac{n-1}{2}\log\left(\frac{1}{n-1}S_c(\mu,\phi)\right) - \frac{n-1}{2S_c(\mu,\phi)} S_c(\mu,\phi)$$

$\hat{\mu}_c$ and $\hat{\phi}_c$ minimize $\log L(\mu, \phi, \sigma_w^2 \mid X_1)$ and also $S_c(\mu,\phi)$

Let $\alpha = \mu(1-\phi)$, then $S_c(\mu,\phi) = \sum_{t=2}^{n} [X_t - (\alpha + \phi X_{t-1})]^2$

Using the results of linear regression, we have

$$\hat{\mu}_c = \frac{1}{1-\hat{\phi}_c}\left(\frac{1}{n-1}\sum_{t=2}^{n} X_t - \frac{1}{n-1}\sum_{t=1}^{n-1} X_t\right) = \frac{\bar{X}_{(2)} - \hat{\phi}_c \bar{X}_{(1)}}{1-\hat{\phi}_c} \approx \frac{(1-\hat{\phi}_c)\mu}{1-\hat{\phi}_c} = \mu$$

$$\hat{\phi}_c = \frac{\sum_{t=2}^{n}(X_t - \bar{X}_{(2)})(X_{t-1} - \bar{X}_{(1)})}{\sum_{t=2}^{n}(X_{t-1} - \bar{X}_{(1)})^2} \approx \frac{\frac{1}{n}\sum_{t=2}^{n}(X_t-\mu)(X_{t-1}-\mu)}{\frac{1}{n}\sum_{t=2}^{n}(X_{t-1}-\mu)^2} \approx \frac{\gamma(1)}{\gamma(0)} = \rho(1)$$

Note that $\hat{\phi}_c \approx \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \hat{\phi}_{YW}$. Actually we have Yule-Walker estimators and the conditional least squares estimators are approximately the same for AR(p) models.

MLE for ARMA(p,q) $\quad X_t = \mu + \sum_{i=1}^{p} \phi_i (X_{t-i} - \mu) + W_t + \sum_{j=1}^{q} \theta_j W_{t-j}$

let $\beta = (\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ be the $(p+q+1)$-dimensional vector

The likelihood is $L(\beta, \sigma_w^2) = f(X_n \mid X_{n-1}, \dots, X_1) f(X_{n-1} \mid X_{n-2}, \dots, X_1) \cdots f(X_2 \mid X_1) f(X_1)$

with $X_1 - \mu = \sum_{j=0}^{\infty} \psi_j W_{t-j}$, $W_t \sim \text{iid } N(0, \sigma_w^2) \Rightarrow X_1 \sim N(\mu, \sigma_w^2 \sum_{j=0}^{\infty} \psi_j^2)$

for $X_t \mid X_{t-1}, \dots, X_1$, note that since X_t is a Guassian process, so $X_t \mid X_{t-1}, \dots, X_1$ is normal with mean $E(X_t \mid X_{t-1}, \dots, X_1) = X_t^{t-1}$ and $Var(X_t \mid X_{t-1}, \dots, X_1) = E[(X_t - X_t^{t-1})^2 \mid X_{t-1}, \dots, X_1)$ which is hard to compute in general.

For AR(1), it is equal to $Var(W_t) = \sigma_w^2$ as we have seen before.

Note that the textbook states that $Var(X_t | X_{t-1}, \dots, X_1) = P_t^{t-1} = E[(X_t - X_t^{t-1})^2]$.
To understand why the condition $X_t, \dots, X_1$ can be dropped out, note that X_t^{t-1} is the projection of X_t on the space $span\{X_t, \dots, X_1\}$, hence $X_t - X_t^{t-1}$ is orthogonal (uncorrelated) to $span\{X_t, \dots, X_1\}$. For normal distribution, it means $X_t - X_t^{t-1}$ is independent of $X_t, \dots, X_1$.
Therefore, we have

$$Var(X_t | X_{t-1}, \dots, X_1) = P_t^{t-1} = \gamma(0) \prod_{j=1}^{t-1} (1 - \phi_{jj}^2)$$

$$= \left(\sigma_w^2 \sum_{j=0}^{\infty} \psi_j^2\right)\left(\prod_{j=1}^{t-1} (1 - \phi_{jj}^2)\right)$$

$$= \sigma_w^2 r_t$$

By defining $x_1^0(\beta) = \mu$, $r_1 = \sum_{j=0}^{\infty} \psi_j^2$, we have $X_t | X_{t-1}, \dots, X_1 \sim N(X_t^{t-1}(\beta), \sigma_w^2 r_t)$,
$X_1 \sim N(X_1^0(\beta), \sigma_w^2 r_1)$

$$\therefore L(\beta, \sigma_w^2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi \sigma_w^2 r_t(\beta)}} e^{-\frac{(X_t - X_t^{t-1}(\beta))^2}{2\sigma_w^2 r_t(\beta)}} = (2\pi\sigma_w^2)^{-\frac{n}{2}} [r_1(\beta) \cdots r_n(\beta)]^{-\frac{1}{2}} e^{-\frac{S(\beta)}{2\sigma_w^2}}$$

where $S(\beta) = \sum_{t=1}^{n} \frac{(X_t - X_t^{t-1}(\beta))^2}{r_t(\beta)}$

Let $\hat{\beta} = (\hat{\mu}, \hat{\phi}_1, \dots, \hat{\phi}_p; \hat{\theta}_1, \dots, \hat{\theta}_q)^T$ be the MLE, the maximizer of $L(\beta, \sigma_w^2)$, then we have $\hat{\sigma}_w^2 = \frac{1}{n} S(\hat{\beta})$ and $\hat{\beta}$ is also the maximizer of

$$l(\beta) = \log\left(\frac{1}{n} S(\beta)\right) + \frac{1}{n} \sum_{t=1}^{n} \log r_t(\beta)$$

The unconditional least squares estimator is the minimizer of $S(\beta)$
For conditional least squares estimator, we assume $x_1, \dots, x_p$ (if $p>0$) are non-random and $w_p = w_{p-1} = \dots = w_{1-q} = 0$. Consider

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

then $w_{p+1}(\beta), w_{p+2}(\beta), \dots, w_n(\beta)$ can be evaluted given β (and the observed $x_{p+1}, \dots, x_n$)
We search β such that $S_c(\beta) = \sum_{t=p+1}^{n} w_t^2(\beta)$ is minimum

Property 3.10 Large sample distribution of the estimators

Under appropriate conditions, for causal and invertible ARMA processes, the MLE, unconditional, conditional least squares estimators, each initialized by the method of moments estimator, all provide optimal estimators of σ_w^2 and β, i.e. $\hat{\sigma}_w^2 \xrightarrow{P} \sigma_w^2$ and the asymptotic distribution of $\hat{\beta}$ is the best asymptotic normal distribution

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \sigma_w^2 \Gamma_{p,q}^{-1}), \qquad \beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$$

where $\Gamma_{p,q} = \begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta} \end{pmatrix}$, $\Gamma_{\phi\phi} = (\gamma_x(i-j))_{1\le i,j\le p}$ with $\phi(B)x_t = w_t$

and $\Gamma_{\theta\theta} = (\gamma_y(i-j))_{1\le i,j\le q}$ with $\theta(B)y_t = w_t$ and

$$\Gamma_{\phi\theta} = (\gamma_{xy}(i-j))_{1\le i\le p, 1\le j\le q} = (Cov(x_{t+i}, y_{t+j}))_{1\le i\le p, 1\le j\le q} \in \mathbb{R}^{p\times q}$$

and $\Gamma_{\theta\phi} = \Gamma_{\phi\theta}^T$

Example 3.34 For AR(1), $x_t = \phi x_{t-1} + w_t$, $p=1, q=0$

$$\phi(B) = 1-\phi B \qquad \Gamma_{1,0} = \Gamma_{\phi\phi} = \gamma_x(1-1) = \gamma_x(0) = \frac{\sigma_w^2}{1-\phi^2} \Rightarrow \sigma_w^2 \Gamma_{1,0}^{-1} = 1-\phi^2$$

$$\therefore \hat{\phi} \xrightarrow{d} N(\phi, n^{-1}(1-\phi^2))$$

For AR(2), $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$, $p=2, q=0$

we can check that $\gamma_x(0) = \left(\frac{1-\phi_2}{1+\phi_2}\right)\frac{\sigma_w^2}{(1-\phi_2)^2-\phi_1^2}$ and $\gamma_x(1) = \phi_1\gamma_x(0) + \phi_2\gamma_x(1)$

$$\Rightarrow \gamma_x(1) = \frac{\phi_1\gamma_x(0)}{1-\phi_2}$$

$$\Gamma_{2,0} = \Gamma_{\phi\phi} = \begin{pmatrix} \gamma_x(0) & \gamma_x(1) \\ \gamma_x(1) & \gamma_x(0) \end{pmatrix}$$

$$\Rightarrow \Gamma_{2,0}^{-1} = (\sigma_w^2)^{-1}\begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, n^{-1}\begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}\right)$$

For MA(1), $x_t = w_t + \theta w_{t-1}$ $\qquad \theta(B) = 1+\theta B$

Consider y_t satisfies $\theta(B)y_t = w_t$ i.e. $y_t = -\theta y_{t+1} + w_t$

$$\Gamma_{0,1} = \Gamma_{\theta\theta} = \gamma_y(1-1) = \gamma_y(0) = \frac{\sigma_w^2}{1-\theta^2} \Rightarrow \sigma_w^2 \Gamma_{0,1}^{-1} = 1-\theta^2$$

$$\therefore \hat{\theta} \xrightarrow{d} N(\theta, n^{-1}(1-\theta^2))$$

For MA(2), $X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2}$ $\quad \theta(B) = 1 + \theta_1 B + \theta_2 B^2$

Consider $\theta(B) y_t = w_t \Rightarrow y_t = -\theta_1 y_{t-1} - \theta_2 y_{t-2} + w_t$

$$E(y_t^2) = -\theta_1 E(y_t y_{t-1}) - \theta_2 E(y_t y_{t-2}) + E(w_t y_t)$$

$$\Rightarrow \gamma_y(0) = -\theta_1 \gamma_y(1) - \theta_2 \gamma_y(2) + \sigma_w^2$$

$$E(y_t y_{t-1}) = \gamma_y(1) = -\theta_1 \gamma_y(0) - \theta_2 \gamma_y(1) \Rightarrow \gamma_y(1) = \frac{-\theta_1}{1+\theta_2} \gamma_y(0)$$

$$E(y_t y_{t-2}) = \gamma_y(2) = -\theta_1 \gamma_y(1) - \theta_2 \gamma_y(0)$$

$$= \left(\frac{\theta_1^2}{1+\theta_2} - \theta_2\right) \gamma_y(0)$$

$$\therefore \gamma_y(0) + \theta_1 \left(\frac{-\theta_1}{1+\theta_2}\right) \gamma_y(0) + \theta_2 \left(\frac{\theta_1^2}{1+\theta_2} - \theta_2\right) \gamma_y(0) = \sigma_w^2$$

$$\gamma_y(0) \left[1 - \theta_2^2 - \theta_1^2 \frac{1}{1+\theta_2} (1 - \theta_2) \right] = \sigma_w^2$$

$$\Rightarrow \gamma_y(0) = \left(\frac{1+\theta_2}{1-\theta_2}\right) \frac{\sigma_w^2}{(1+\theta_2)^2 - \theta_1^2}$$

$$\gamma_y(1) = -\frac{\theta_1}{1-\theta_2} \frac{\sigma_w^2}{(1+\theta_2)^2 - \theta_1^2}$$

$$\therefore \Gamma_{0,2} = \Gamma_{\theta\theta} = \begin{pmatrix} \gamma_y(0) & \gamma_y(1) \\ \gamma_y(1) & \gamma_y(0) \end{pmatrix} \Rightarrow \Gamma_{0,2}^{-1} = \sigma_w^{-2} \begin{pmatrix} 1-\theta_2^2 & \theta_1(1-\theta_2) \\ \theta_1(1-\theta_2) & 1-\theta_2^2 \end{pmatrix}$$

(Note that there is a typo in (3.137))

$$\therefore \begin{pmatrix} \hat\theta_1 \\ \hat\theta_2 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, n^{-1} \begin{pmatrix} 1-\theta_2^2 & \theta_1(1-\theta_2) \\ \theta_1(1-\theta_2) & 1-\theta_2^2 \end{pmatrix} \right)$$

For ARMA(1,1) $\quad X_t = \phi X_{t-1} + W_t + \theta W_{t-1}$ $\quad p = q = 1, \ \phi(B) = 1 - \phi B$, $\ \theta(B) = 1 + \theta B$

Consider $\phi(B) X_t = W_t \Rightarrow X_t = \phi X_{t-1} + W_t$

$\theta(B) y_t = W_t \Rightarrow y_t = -\theta y_{t-1} + W_t$

$$\Gamma_{1,1} = \begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta} \end{pmatrix} = \begin{pmatrix} \gamma_x(0) & \gamma_{xy}(0) \\ \gamma_{xy}(0) & \gamma_y(0) \end{pmatrix}$$

$$\gamma_x(0) = \frac{\sigma_w^2}{1-\phi^2} \quad \gamma_y(0) = \frac{\sigma_w^2}{1-\theta^2} \quad \gamma_{xy}(0) = Cov(X_t, y_t) = Cov(\phi X_{t-1} + W_t, -\theta y_{t-1} + W_t)$$

$$= -\phi\theta \gamma_{xy}(0) + \sigma_w^2$$

$$\therefore \gamma_{xy}(0) = \frac{\sigma_w^2}{1+\phi\theta}$$

$$\therefore \begin{pmatrix} \hat\phi \\ \hat\theta \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \phi \\ \theta \end{pmatrix}, n^{-1} \begin{pmatrix} (1-\phi^2)^{-1} & (1+\phi\theta)^{-1} \\ (1+\phi\theta)^{-1} & (1-\theta^2)^{-1} \end{pmatrix}^{-1} \right)$$

Example 3.36 Consider $X_t = \mu + \phi(X_{t-1} - \mu) + W_t$

where $\mu = 50$, $\phi = 0.95$, $W_t \sim iid(0, 8)$ but not normal

We can still use the Yule-Walker estimators $\hat{\mu} = \bar{X}$, $\hat{\phi} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p$

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

but the asymptotic normal approximation in Property 3.10 can be poor when n is small ($n=100$ in Ex. 3.36) or the parameters are close to the boundaries. In such case, bootstrap can be helpful.

Suppose we have $X_1, X_2, \dots, X_n$, the we can approximate W_t by

$$\hat{W}_t = X_t - \hat{\mu} - \hat{\phi}(X_{t-1} - \hat{\mu}) \qquad \text{for } t = 2, \dots, n$$

For general ARMA(p,q),

$$\hat{W}_t = (X_t - \hat{\mu}) - \sum_{j=1}^{p} \hat{\phi}_j (X_{t-j} - \hat{\mu}) - \sum_{k=1}^{q} \hat{\theta}_k \hat{W}_{t-k}(\hat{\beta}) \qquad t = p+1, \dots, n$$

with $\hat{W}_{p-1} = \hat{W}_{p-2} = \dots = 0$

Since W_t's are iid, we resample $\{W_2^*, \dots, W_n^*\}$ from $\{\hat{W}_2, \dots, \hat{W}_n\}$ to form

$$X_t^* = \hat{\mu} + \hat{\phi}(X_{t-1} - \hat{\mu}) + W_t^* \qquad \text{for } t = 2, \dots, n$$

$$X_1^* = X_1$$

Then, for each set of $\{X_1^*, \dots, X_n^*\}$, we can compute the corresponding $\{\hat{\mu}^*, \hat{\phi}^*, (\hat{\sigma}_w^2)^*\}$. Repeat the process B times to get $\{\hat{\mu}(b), \hat{\phi}(b), \hat{\sigma}_w^2(b)\ ;\ b = 1, \dots, B\}$. the we can approximate the distribution of $\hat{\phi} - \phi$ by the empirical distribution $\{\hat{\phi}(b) - \hat{\phi}\ ;\ b = 1, \dots, B\}$