$$\gamma(1) = \mathrm{E}(y_t y_{t-1}) = \mathrm{E}((\epsilon_t + \theta_1 \epsilon_{t-1})(\epsilon_{t-1} + \theta_1 \epsilon_{t-2})) = \theta_1 \sigma_\epsilon^2$$

$$\gamma(2) = \mathrm{E}(y_t y_{t-2}) = \mathrm{E}((\epsilon_t + \theta_1 \epsilon_{t-1})(\epsilon_{t-2} + \theta_1 \epsilon_{t-3})) = 0$$

3. **Autocorrelation Function of MA(1) Process:**

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \begin{cases} \dfrac{\theta_1}{1 + \theta_1^2}, & \text{for } \tau = 1, \\ 0, & \text{for } \tau = 2, 3, \cdots. \end{cases}$$

Let x be $\rho(1)$.

$$\frac{\theta_1}{1 + \theta_1^2} = x, \qquad \text{i.e.,} \qquad x\theta_1^2 - \theta + x = 0.$$

θ_1 should be a real number.

$$1 - 4x^2 > 0, \qquad \text{i.e.,} \qquad -\frac{1}{2} \le \rho(1) \le \frac{1}{2}.$$

4. **Invertibility Condition of MA(1) Process:**

$$
\begin{aligned}
\epsilon_t &= -\theta_1 \epsilon_{t-1} + y_t \\
&= (-\theta_1)^2 \epsilon_{t-2} + y_t + (-\theta_1) y_{t-1} \\
&= (-\theta_1)^3 \epsilon_{t-3} + y_t + (-\theta_1) y_{t-1} + (-\theta_1)^2 y_{t-2} \\
&\quad \vdots \\
&= (-\theta_1)^s \epsilon_{t-s} + y_t + (-\theta_1) y_{t-1} + (-\theta_1)^2 y_{t-2} + \cdots + (-\theta_1)^{t-s+1} y_{t-s+1}
\end{aligned}
$$

When $(-\theta_1)^s \epsilon_{t-s} \longrightarrow 0$, the MA(1) model is written as the AR(∞) model, i.e.,

$$
y_t = -(-\theta_1) y_{t-1} - (-\theta_1)^2 y_{t-2} - \cdots - (-\theta_1)^{t-s+1} y_{t-s+1} - \cdots + \epsilon_t
$$

5. **Likelihood Function of MA(1) Process:**

The autocovariance functions are: $\gamma(0) = (1 + \theta_1^2)\sigma_\epsilon^2$, $\gamma(1) = \theta_1 \sigma_\epsilon^2$, and $\gamma(\tau) = 0$ for $\tau = 2, 3, \cdots$.

The joint distribution of $y_1, y_2, \cdots, y_T$ is:

$$f(y_1, y_2, \cdots, y_T) = \frac{1}{(2\pi)^{T/2}} |V|^{-1/2} \exp\left(-\frac{1}{2} Y' V^{-1} Y\right)$$

where

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, \qquad V = \sigma_\epsilon^2 \begin{pmatrix} 1+\theta_1^2 & \theta_1 & 0 & \cdots & 0 \\ \theta_1 & 1+\theta_1^2 & \theta_1 & \ddots & \vdots \\ 0 & \theta_1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1+\theta_1^2 & \theta_1 \\ 0 & \cdots & 0 & \theta_1 & 1+\theta_1^2 \end{pmatrix}.$$

6. **MA(1) +drift:** $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$

Mean of MA(1) Process:

$$y_t = \mu + \theta(L)\epsilon_t,$$

where $\theta(L) = 1 + \theta_1 L$.

Taking the expectation,

$$\mathrm{E}(y_t) = \mu + \theta(L)\mathrm{E}(\epsilon_t) = \mu.$$

Example: MA(2) Model: $y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$

1. Autocovariance Function of MA(2) Process:

$$\gamma(\tau) = \begin{cases} (1 + \theta_1^2 + \theta_2^2)\sigma_\epsilon^2, & \text{for } \tau = 0, \\ (\theta_1 + \theta_1\theta_2)\sigma_\epsilon^2, & \text{for } \tau = 1, \\ \theta_2\sigma_\epsilon^2, & \text{for } \tau = 2, \\ 0, & \text{otherwise.} \end{cases}$$

2. let $-1/\beta_1$ and $-1/\beta_2$ be two solutions of x from $\theta(x) = 0$.

   For invertibility condition, both β_1 and β_2 should be less than one in absolute value.

   Then, the MA(2) model is represented as:

$$\begin{aligned} y_t &= \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \\ &= (1 + \theta_1 L + \theta_2 L^2)\epsilon_t \\ &= (1 + \beta_1 L)(1 + \beta_2 L)\epsilon_t \end{aligned}$$

AR(∞) representation of the MA(2) model is given by:

$$\begin{aligned}\epsilon_t &= \frac{1}{(1+\beta_1 L)(1+\beta_2 L)} y_t \\ &= \left(\frac{\beta_1/(\beta_1 - \beta_2)}{1+\beta_1 L} + \frac{-\beta_2/(\beta_1 - \beta_2)}{1+\beta_2 L} \right) y_t\end{aligned}$$

3. **Likelihood Function:**

$$f(y_1, y_2, \cdots, y_T) = \frac{1}{(2\pi)^{T/2}} |V|^{-1/2} \exp\left(-\frac{1}{2} Y' V^{-1} Y\right)$$

where

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, \quad V = \sigma_\epsilon^2 \begin{pmatrix} 1+\theta_1^2+\theta_2^2 & \theta_1+\theta_1\theta_2 & \theta_2 & & 0 \\ \theta_1+\theta_1\theta_2 & 1+\theta_1^2+\theta_2^2 & \theta_1+\theta_1\theta_2 & \ddots & \\ \theta_2 & \theta_1+\theta_1\theta_2 & \ddots & \ddots & \theta_2 \\ & \ddots & \ddots & 1+\theta_1^2+\theta_2^2 & \theta_1+\theta_1\theta_2 \\ 0 & & \theta_2 & \theta_1+\theta_1\theta_2 & 1+\theta_1^2+\theta_2^2 \end{pmatrix}$$

4. **MA(2) +drift:** $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$

Mean:

$$y_t = \mu + \theta(L)\epsilon_t,$$

where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2$.

Therefore,

$$\mathrm{E}(y_t) = \mu + \theta(L)\mathrm{E}(\epsilon_t) = \mu$$

Example: MA(q) Model: $y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}$

1. **Mean of MA(q) Process:**

$$\mathrm{E}(y_t) = \mathrm{E}(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}) = 0$$

2. **Autocovariance Function of MA(q) Process:**

$$\gamma(\tau) = \begin{cases} \sigma_\epsilon^2(\theta_0\theta_\tau + \theta_1\theta_{\tau+1} + \cdots + \theta_{q-\tau}\theta_q) = \sigma_\epsilon^2 \sum_{i=0}^{q-\tau} \theta_i\theta_{\tau+i}, & \tau = 1, 2, \cdots, q, \\ 0, & \tau = q+1, q+2, \cdots, \end{cases}$$

   where $\theta_0 = 1$.

3. MA(q) process is stationary.

4. **MA(q) +drift:** $y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}$

Mean:

$$y_t = \mu + \theta(L)\epsilon_t,$$

where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$.

Therefore, we have:

$$\mathrm{E}(y_t) = \mu + \theta(L)\mathrm{E}(\epsilon_t) = \mu.$$

6.5 ARMA Model

ARMA (Autoregressive Moving Average, 自己回帰移動平均) Process

1. ARMA(p, q)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q},$$

   which is rewritten as:

$$\phi(L) y_t = \theta(L) \epsilon_t,$$

   where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$.

2. Likelihood Function:

   The variance-covariance matrix of Y, denoted by V, has to be computed.

Example: ARMA(1,1) Process: $y_t = \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$

Obtain the autocorrelation coefficient.

The mean of y_t is to take the expectation on both sides.

$$\mathrm{E}(y_t) = \phi_1 \mathrm{E}(y_{t-1}) + \mathrm{E}(\epsilon_t) + \theta_1 \mathrm{E}(\epsilon_{t-1}),$$

where the second and third terms are zeros.

Therefore, we obtain:

$$\mathrm{E}(y_t) = 0.$$

The autocovariance of y_t is to take the expectation, multiplying $y_{t-\tau}$ on both sides.

$$\mathrm{E}(y_t y_{t-\tau}) = \phi_1 \mathrm{E}(y_{t-1} y_{t-\tau}) + \mathrm{E}(\epsilon_t y_{t-\tau}) + \theta_1 \mathrm{E}(\epsilon_{t-1} y_{t-\tau}).$$

Each term is given by:

$$\mathrm{E}(y_t y_{t-\tau}) = \gamma(\tau), \qquad \mathrm{E}(y_{t-1} y_{t-\tau}) = \gamma(\tau - 1),$$

$$\mathrm{E}(\epsilon_t y_{t-\tau}) = \begin{cases} \sigma_\epsilon^2, \ \tau = 0, \\ 0, \quad \tau = 1, 2, \cdots, \end{cases} \qquad \mathrm{E}(\epsilon_{t-1} y_{t-\tau}) = \begin{cases} (\phi_1 + \theta_1)\sigma_\epsilon^2, \ \tau = 0, \\ \sigma_\epsilon^2, \qquad \tau = 1, \\ 0, \qquad \tau = 2, 3, \cdots. \end{cases}$$

Therefore, we obtain;

$$\gamma(0) = \phi_1 \gamma(1) + (1 + \phi_1\theta_1 + \theta_1^2)\sigma_\epsilon^2,$$

$$\gamma(1) = \phi_1 \gamma(0) + \theta_1 \sigma_\epsilon^2,$$

$$\gamma(\tau) = \phi_1 \gamma(\tau - 1), \qquad \tau = 2, 3, \cdots.$$

From the first two equations, $\gamma(0)$ and $\gamma(1)$ are computed by:

$$\begin{pmatrix} 1 & -\phi_1 \\ -\phi_1 & 1 \end{pmatrix} \begin{pmatrix} \gamma(0) \\ \gamma(1) \end{pmatrix} = \sigma_\epsilon^2 \begin{pmatrix} 1 + \phi_1\theta_1 + \theta_1^2 \\ \theta_1 \end{pmatrix}$$

$$\begin{pmatrix} \gamma(0) \\ \gamma(1) \end{pmatrix} = \sigma_\epsilon^2 \begin{pmatrix} 1 & -\phi_1 \\ -\phi_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 + \phi_1\theta_1 + \theta_1^2 \\ \theta_1 \end{pmatrix}$$

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

Thus, the initial value of the autocorrelation coefficient is given by:

$$\rho(1) = \frac{(1+\phi_1\theta_1)(\phi_1+\theta_1)}{1 + 2\phi_1\theta_1 + \theta_1^2}.$$

We have:

$$\rho(\tau) = \phi_1\rho(\tau - 1).$$

ARMA(p, q) +drift:

$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$.

Mean of ARMA(p, q) Process: $\phi(L) y_t = \mu + \theta(L)\epsilon_t$,

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$.

$$y_t = \phi(L)^{-1}\mu + \phi(L)^{-1}\theta(L)\epsilon_t.$$

Therefore,

$$\mathrm{E}(y_t) = \phi(L)^{-1}\mu + \phi(L)^{-1}\theta(L)\mathrm{E}(\epsilon_t) = \phi(1)^{-1}\mu = \frac{\mu}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}.$$

6.6 ARIMA Model

Autoregressive Integrated Moving Average (ARIMA，自己回帰和分移動平均) Model

ARIMA(p, d, q) Process

$$\phi(L)\Delta^d y_t = \theta(L)\epsilon_t,$$

where $\Delta^d y_t = \Delta^{d-1}(1-L)y_t = \Delta^{d-1}y_t - \Delta^{d-1}y_{t-1} = (1-L)^d y_t$ for $d = 1, 2, \cdots$, and $\Delta^0 y_t = y_t$.

例：ARIMA(0,1,0) Model

Consider the model: $\Delta y_t = y_t - y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \quad y_0 = 0,$

which is rewritten as: $y_t = \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1.$

$$\mathrm{E}(y_t) = 0, \quad \gamma(0) = \mathrm{V}(y_t) = \sigma^2 t, \quad \gamma(\tau) = \mathrm{Cov}(y_t, y_{t-\tau}) = \mathrm{E}(y_t y_{t-\tau}) = \sigma^2(t-\tau),$$

which implies that $\gamma(\tau)$ is time-dependent. $\Longrightarrow$ y_t is not stationary.

$$\rho(\tau) = \frac{\mathrm{Cov}(y_t, y_{t-\tau})}{\sqrt{\mathrm{V}(y_t)}\sqrt{\mathrm{V}(y_{t-\tau})}} = \frac{t-\tau}{\sqrt{t}\sqrt{t-\tau}} = \sqrt{\frac{t-\tau}{t}}.$$

That is, $\rho(\tau)$ gradually decreases with slow speed.

6.7 SARIMA Model

Seasonal ARIMA (SARIMA) Process:

1. SARIMA(p, d, q)

$$\phi(L)\Delta^d \Delta_s y_t = \theta(L)\epsilon_t,$$

   where

$$\Delta_s y_t = (1 - L^s) y_t = y_t - y_{t-s}.$$

   $s = 4$ when y_t denotes quarterly date and $s = 12$ when y_t represents monthly data.

6.8 Optimal Prediction

1. AR(p) Process: $y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t$

(a) Define:

$$E(y_{t+k}|Y_t) = y_{t+k|t},$$

where Y_t denotes all the information available at time t.

Taking the conditional expectation of $y_{t+k} = \phi_1 y_{t+k-1} + \cdots + \phi_p y_{t+k-p} + \epsilon_{t+k}$ on both sides,

$$y_{t+k|t} = \phi_1 y_{t+k-1|t} + \cdots + \phi_p y_{t+k-p|t},$$

where $y_{s|t} = y_s$ for $s \le t$.

(b) Optimal prediction is given by solving the above differential equation.

2. MA(q) Process: $y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$

(a) Let $\hat{\epsilon}_T, \hat{\epsilon}_{T-1}, \cdots, \hat{\epsilon}_1$ be the estimated errors.

(b) $y_{t+k} = \epsilon_{t+k} + \theta_1\epsilon_{t+k-1} + \cdots + \theta_q\epsilon_{t+k-q}$

(c) Therefore,

$$y_{t+k|t} = \epsilon_{t+k|t} + \theta_1 \epsilon_{t+k-1|t} + \cdots + \theta_q \epsilon_{t+k-q|t},$$

where $\epsilon_{s|t} = 0$ for $s > t$ and $\epsilon_{s|t} = \hat{\epsilon}_s$ for $s \leq t$.

3. ARMA(p, q) Process: $y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$

   (a) $y_{t+k} = \phi_1 y_{t+k-1} + \cdots + \phi_p y_{t+k-p} + \epsilon_{t+k} + \theta_1 \epsilon_{t+k-1} + \cdots + \theta_q \epsilon_{t+k-q}$

   (b) Optimal prediction is:

$$y_{t+k|t} = \phi_1 y_{t+k-1|t} + \cdots + \phi_p y_{t+k-p|t} + \epsilon_{t+k|t} + \theta_1 \epsilon_{t+k-1|t} + \cdots + \theta_q \epsilon_{t+k-q|t},$$

where $y_{s|t} = y_s$ and $\epsilon_{s|t} = \hat{\epsilon}_s$ for $s \leq t$, and $\epsilon_{s|t} = 0$ for $s > t$.

6.9 Identification (識別，または，同定)

We have the following two approaches for model specification.

1. Based on AIC or SBIC given d, s, we obtain p, q.

   (a) AIC (Akaike's Information Criterion，赤池の情報量基準)

   $$\text{AIC} = -2\log(\text{likelihood}) + 2k,$$

   where $k = p + q$, which is the number of parameters estimated.

   (b) SBIC (Shwarz's Bayesian Information Criterion)

   $$\text{SBIC} = -2\log(\text{likelihood}) + k\log T,$$

   where T denotes the number of observations.

2. From the sample autocorrelation coefficient function $\hat{\rho}(k)$ and the sample partial autocorrelation coefficient function $\hat{\phi}_{k,k}$ for $k = 1, 2, \cdots$, we obtain p, d, q, s.

	AR(p) Process	MA(q) Process
Autocorrelation Function	Gradually decreasing	$\rho(k) = 0$, $k = q+1, q+2, \cdots$
Partial Autocorrelation Function	$\phi(k,k) = 0$, $k = p+1, p+2, \cdots$	Gradually decreasing

(a) Compute $\Delta_s y_t$ to remove seasonality.

Compute the autocovariance functions of $\Delta_s y_t$.

If the autocovariance functions have period s, we take $(1 - L^s)$, again.

(b) Determine the order of difference.

Compute the partial autocovariance functions every time.

If the autocovariance functions decrease as τ is large, go to the next step.

(c) Determine the order of AR terms (i.e., p).

Compute the partial autocovariance functions every time.

The partial autocovariance functions are close to zero after some τ, go to the next step.

(d) Determine the order of MA terms (i.e., q).

Compute the autocovariance functions every time.

If the autocovariance functions are randomly around zero, end of the procedure.