5. **Partial Autocorrelation Function of MA(1) Process:**

$$\phi_{1,1} = \rho(1) = \frac{\theta_1}{1+\theta_1^2} \neq 0$$

$$\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{2,1} \\ \phi_{2,2} \end{pmatrix} = \begin{pmatrix} \phi(1) \\ \phi(2) \end{pmatrix}$$

$$\Longrightarrow \quad \phi_{2,2} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{-\rho(1)^2}{1-\rho(1)^2} = \frac{-\theta_1^2}{1+\theta_1^2+\theta_1^4} \neq 0$$

$$\begin{pmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{3,1} \\ \phi_{3,2} \\ \phi_{3,3} \end{pmatrix} = \begin{pmatrix} \phi(1) \\ \phi(2) \\ \phi(3) \end{pmatrix}$$

$$\Longrightarrow \quad \phi_{3,3} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & 0 \\ 0 & \rho(1) & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & 0 \\ \rho(1) & 1 & \rho(1) \\ 0 & \rho(1) & 1 \end{vmatrix}} = \frac{\rho(1)^3}{1 - 2\rho(1)^2} \neq 0$$

$$\begin{pmatrix} 1 & \rho(1) & \rho(2) & \rho(3) \\ \rho(1) & 1 & \rho(1) & \rho(2) \\ \rho(2) & \rho(1) & 1 & \rho(1) \\ \rho(3) & \rho(2) & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{4,1} \\ \phi_{4,2} \\ \phi_{4,3} \\ \phi_{4,4} \end{pmatrix} = \begin{pmatrix} \phi(1) \\ \phi(2) \\ \phi(3) \\ \phi(4) \end{pmatrix}$$

$$\Longrightarrow \quad \phi_{4,4} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \rho(1) \\ \rho(1) & 1 & \rho(1) & \rho(2) \\ \rho(2) & \rho(1) & 1 & \rho(3) \\ \rho(3) & \rho(2) & \rho(1) & \rho(4) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \rho(3) \\ \rho(1) & 1 & \rho(1) & \rho(2) \\ \rho(2) & \rho(1) & 1 & \rho(1) \\ \rho(3) & \rho(2) & \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho(1) & 0 & \rho(1) \\ \rho(1) & 1 & \rho(1) & 0 \\ 0 & \rho(1) & 1 & 0 \\ 0 & 0 & \rho(1) & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & 0 & 0 \\ \rho(1) & 1 & \rho(1) & 0 \\ 0 & \rho(1) & 1 & \rho(1) \\ 0 & 0 & \rho(1) & 1 \end{vmatrix}} \neq 0$$

As a result, $\phi_{k,k} \neq 0$ for all $k = 1, 2, \cdots$

6. **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

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

7. **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:

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

$$= (1 + \theta_1 L + \theta_2 L^2)\epsilon_t$$

$$= (1 + \beta_1 L)(1 + \beta_2 L)\epsilon_t$$

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:

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

5.4 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:

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

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

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,

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

5.5 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 $\Delta^0 y_t = y_t$.

5.6 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.

5.7 Optimal Prediction

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

(a) Define:

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

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 $s \leq t$.

(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 $s > t$.

5.8 Identification

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

   We choose p and q, where AIC or SBIC is minimized.

   (a) AIC (Akaike's Information Criterion)

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

   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 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.

5.9 Example of SARIMA using Consumption Data

Construct SARIMA model using monthly and seasonally unadjusted consumption expenditure data and STATA12.

Estimation Period: Jan., 1970 — Dec., 2012 ($T = 516$)

```
. gen time=_n

. tsset time
        time variable:  time, 1 to 516
                delta:  1 unit

. corrgram expend

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.8488   0.8499   373.88  0.0000          |------            |------
2        0.8231   0.3858   726.18  0.0000          |------            |---
3        0.8716   0.5266     1122  0.0000          |------            |----
4        0.8706   0.4025   1517.6  0.0000          |------            |---
5        0.8498   0.3447   1895.3  0.0000          |------            |--
6        0.8085   0.0074   2237.9  0.0000          |------            |
7        0.8378   0.1528   2606.5  0.0000          |------            |-
```

```
8        0.8460   0.1467     2983  0.0000          |------            |-
9        0.8342   0.3006   3349.9  0.0000          |------            |--
10       0.7735  -0.1518     3666  0.0000          |------           -|
11       0.7852  -0.1185   3992.3  0.0000          |------            |
12       0.9234   0.9442   4444.5  0.0000          |-------           |-------
13       0.7754  -0.5486   4764.1  0.0000          |------        ----|
14       0.7482  -0.3248   5062.1  0.0000          |-----           --|
15       0.7963  -0.2392   5400.5  0.0000          |------           -|


. gen dexp=expend-l.expend
(1 missing value generated)

. corrgram dexp

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1       -0.4316  -0.4329   96.485  0.0000       ---|               ---|
2       -0.2546  -0.5441   130.13  0.0000        --|              ----|
3        0.1721  -0.4091   145.53  0.0000          |-              ---|
4        0.0667  -0.3459   147.85  0.0000          |                --|
5        0.0715  -0.0036   150.52  0.0000          |                  |
6       -0.2428  -0.1489   181.36  0.0000         -|                 -|
7        0.0711  -0.1400   184.01  0.0000          |                 -|
8        0.0668  -0.2900   186.36  0.0000          |                --|
9        0.1704   0.1681   201.64  0.0000          |-                 |-
10      -0.2485   0.1306   234.21  0.0000         -|                  |-
11      -0.4293  -0.9305   331.56  0.0000       ---|           -------|
12       0.9773   0.6768   837.12  0.0000          |-------           |-----
13      -0.4152   0.3778   928.56  0.0000       ---|                  |---
14      -0.2583   0.2688   964.03  0.0000        --|                  |--
15       0.1712   0.0406   979.63  0.0000          |-                 |
```

```
. gen sdex=dexp-l12.dexp
(13 missing values generated)

. corrgram sdex

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1       -0.4752  -0.4753   114.28  0.0000       ---|               ---|
2       -0.0244  -0.3235   114.58  0.0000          |                --|
3        0.1163  -0.0759   121.46  0.0000          |                  |
4       -0.1246  -0.1365   129.37  0.0000          |                 -|
5        0.0341  -0.1016   129.96  0.0000          |                  |
6       -0.0151  -0.1136   130.08  0.0000          |                  |
7       -0.0395  -0.1413   130.88  0.0000          |                 -|
8        0.1123   0.0092   137.35  0.0000          |                  |
9       -0.0664  -0.0100   139.62  0.0000          |                  |
10       0.0168   0.0069   139.76  0.0000          |                  |
11       0.1642   0.2422   153.68  0.0000          |-                 |-
12      -0.3888  -0.2469    231.9  0.0000       ---|                 -|
13       0.2242  -0.1205   257.96  0.0000          |-                 |
14      -0.0147  -0.0941   258.07  0.0000          |                  |
15      -0.0708  -0.0591   260.68  0.0000          |                  |



. arima sdex, ar(1,2) ma(1)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -5107.4608
Iteration 1:   log likelihood =  -5102.391
```

```
Iteration 2:   log likelihood = -5099.9071
Iteration 3:   log likelihood = -5099.4216
Iteration 4:   log likelihood = -5099.2463
(switching optimization to BFGS)
Iteration 5:   log likelihood = -5099.2361
Iteration 6:   log likelihood = -5099.2346
Iteration 7:   log likelihood = -5099.2346
Iteration 8:   log likelihood = -5099.2346

ARIMA regression

Sample:  14 - 516                               Number of obs      =       503
                                                Wald chi2(3)       =    973.93
Log likelihood = -5099.235                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
        sdex |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
sdex         |
       _cons |  -15.64573   59.17574    -0.26   0.791     -131.628    100.3366
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .1271774   .0581883     2.19   0.029     .0131304    .2412244
         L2. |   .1009983    .053626     1.88   0.060    -.0041068    .2061034
             |
          ma |
         L1. |  -.8343264   .0419364   -19.90   0.000    -.9165202   -.7521326
-------------+----------------------------------------------------------------
      /sigma |   6111.128   139.0105    43.96   0.000     5838.673    6383.584
------------------------------------------------------------------------------
```

```
Note: The test of the variance against zero is one sided, and the two-sided
    confidence interval is truncated at zero.



. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |    503           .   -5099.235      5     10208.47    10229.57
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

6 Unit Root (単位根) and Cointegration (共和分)

6.1 Unit Root (単位根) Test (Dickey-Fuller (DF) Test)

1. Why is a unit root problem important?

   (a) Economic variables increase over time in general.

   One of the assumptions of OLS is stationarity on y_t and x_t.

   This assumption implies that $\dfrac{1}{T}X'X$ converges to a fixed matrix as T is large.

   That is, asymptotic normality of OLS estimator does not hold.

   (b) In nonstationary time series, the unit root is the most important.

   In the case of unit root, OLSE of the first-order autoregressive coefficient is consistent.

OLSE is $\sqrt{T}$-consistent in the case of stationary AR(1) process, but OLSE is T-consistent in the case of nonstationay AR(1) process.

(c) A lot of economic variables increase over time.

It is important to check an economic variable is trend stationary (i.e., $y_t = a_0 + a_1 t + \epsilon_t$) or difference stationary (i.e., $y_t = b_0 + y_{t-1} + \epsilon_t$).

Consider k-step ahead prediction for both cases.

$$\text{(Trend Stationarity)} \qquad y_{t+k|t} = a_0 + a_1(t+k)$$

$$\text{(Difference Stationarity)} \qquad y_{t+k|t} = b_0 k + y_t$$

2. **The Case of $|\phi_1| < 1$:**

$$y_t = \phi_1 y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{i.i.d. } N(0, \sigma_\epsilon^2), \qquad y_0 = 0, \qquad t = 1, \cdots, T$$

Then, OLSE of ϕ_1 is:

$$\hat{\phi}_1 = \frac{\displaystyle\sum_{t=1}^{T} y_{t-1} y_t}{\displaystyle\sum_{t=1}^{T} y_{t-1}^2}.$$

In the case of $|\phi_1| < 1$,

$$\hat{\phi}_1 = \phi_1 + \frac{\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow \phi_1 + \frac{\mathrm{E}(y_{t-1}\epsilon_t)}{\mathrm{E}(y_{t-1}^2)} = \phi_1.$$

Note as follows:

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow \mathrm{E}(y_{t-1}\epsilon_t) = 0.$$

By the central limit theorem,

$$\frac{\overline{y\epsilon} - \mathrm{E}(\overline{y\epsilon})}{\sqrt{\mathrm{V}(\overline{y\epsilon})}} \longrightarrow N(0, 1)$$

where

$$\overline{y\epsilon} = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t.$$

$$
\begin{aligned}
\mathrm{E}(\overline{y\epsilon}) &= 0, \\
\mathrm{V}(\overline{y\epsilon}) &= \mathrm{V}(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t) = \mathrm{E}\Big((\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t)^2\Big) \\
&= \frac{1}{T^2}\mathrm{E}\Big(\sum_{t=1}^{T}\sum_{s=1}^{T} y_{t-1}y_{s-1}\epsilon_t\epsilon_s\Big) = \frac{1}{T^2}\mathrm{E}\Big(\sum_{t=1}^{T} y_{t-1}^2\epsilon_t^2\Big) = \frac{1}{T}\sigma_\epsilon^2\gamma(0).
\end{aligned}
$$

Therefore,

$$
\frac{\overline{y\epsilon}}{\sqrt{\sigma_\epsilon^2\gamma(0)/T}} = \frac{1}{\sigma_\epsilon\sqrt{\gamma(0)}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow N(0,1),
$$

which is rewritten as:

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow N(0,\sigma_\epsilon^2\gamma(0)).
$$

Using $\dfrac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \mathrm{E}(y_{t-1}^2) = \gamma(0)$, we have the following asymptotic distribution:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) = \frac{\dfrac{1}{\sqrt{T}}\displaystyle\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow N\left(0, \frac{\sigma_\epsilon^2}{\gamma(0)}\right) = N\left(0, 1 - \phi_1^2\right).$$

Note that $\gamma(0) = \dfrac{\sigma_\epsilon^2}{1 - \phi_1^2}$.

3. In the case of $\phi_1 = 1$, as expected, we have:

$$\sqrt{T}(\hat{\phi}_1 - 1) \longrightarrow 0.$$

That is, $\hat{\phi}_1$ has the distribution which converges in probability to $\phi_1 = 1$ (i.e., degenerated distribution).

Is this true?

4. **The Case of $\boldsymbol{\phi_1 = 1}$:** $\Longrightarrow$ Random Walk Process

$y_t = y_{t-1} + \epsilon_t$ with $y_0 = 0$ is written as:

$$y_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \cdots + \epsilon_1.$$

Therefore, we can obtain:

$$y_t \sim N(0, \sigma_\epsilon^2 t).$$

The variance of y_t depends on time t. $\Longrightarrow$ y_t is nonstationary.

5. Remember that $\hat{\phi}_1 = \phi_1 + \dfrac{\sum y_{t-1}\epsilon_t}{\sum y_{t-1}^2}$.

   (a) First, consider the numerator $\sum y_{t-1}\epsilon_t$.

   We have $y_t^2 = (y_{t-1} + \epsilon_t)^2 = y_{t-1}^2 + 2y_{t-1}\epsilon_t + \epsilon_t^2$.

Therefore, we obtain:

$$y_{t-1}\epsilon_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - \epsilon_t^2).$$

Taking into account $y_0 = 0$, we have:

$$\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}y_T^2 - \frac{1}{2}\sum_{t=1}^{T}\epsilon_t^2.$$

Divided by $\sigma_\epsilon^2 T$ on both sides, we have the following:

$$\frac{1}{\sigma_\epsilon^2 T}\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\left(\frac{y_T}{\sigma_\epsilon\sqrt{T}}\right)^2 - \frac{1}{2\sigma_\epsilon^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2.$$

From $y_t \sim N(0, \sigma_\epsilon^2 t)$, we obtain the following result:

$$\left(\frac{y_T}{\sigma_\epsilon\sqrt{T}}\right)^2 \sim \chi^2(1).$$

Moreover, the second term is derived from:

$$\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \longrightarrow \mathrm{E}(\epsilon_t^2) = \sigma_\epsilon^2.$$

Therefore,

$$\frac{1}{\sigma_\epsilon^2 T}\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\left(\frac{y_T}{\sigma_\epsilon\sqrt{T}}\right)^2 - \frac{1}{2\sigma_\epsilon^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \longrightarrow \frac{1}{2}(\chi^2(1)-1).$$

(b) Next, consider $\sum y_{t-1}^2$.

$$\mathrm{E}\left(\sum_{t=1}^{T} y_{t-1}^2\right) = \sum_{t=1}^{T}\mathrm{E}(y_{t-1}^2) = \sum_{t=1}^{T}\sigma_\epsilon^2(t-1) = \sigma_\epsilon^2\frac{T(T-1)}{2}.$$

Thus, we obtain the following result:

$$\frac{1}{T^2}\mathrm{E}\left(\sum_{t=1}^{T} y_{t-1}^2\right) \longrightarrow \text{ a fixed value, i.e., } \frac{\sigma_\epsilon^2}{2}.$$

Therefore,

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \text{ a distribution.}$$

6. Summarizing the results up to now, $T(\hat{\phi}_1 - \phi_1)$, not $\sqrt{T}(\hat{\phi}_1 - \phi_1)$, has limiting distribution in the case of $\phi_1 = 1$.

$$T(\hat{\phi}_1 - \phi_1) = \frac{(1/T)\sum y_{t-1}\epsilon_t}{(1/T^2)\sum y_{t-1}^2} \longrightarrow \text{ a distribution.}$$

7. Basic Concepts of Random Walk Process:

   (a) Model: $y_t = y_{t-1} + \epsilon_t, \qquad y_0 = 0, \qquad \epsilon_t \sim N(0,1).$

   Then,

   $$y_t = \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1.$$

   Therefore,

   $$y_t \sim N(0,t).$$

$\Longrightarrow$ Nonstationary Process (i.e., variance depends on time t.)

Difference between y_s and y_t ($s > t$) is:

$$y_s - y_t = \epsilon_s + \epsilon_{s-1} + \cdots + \epsilon_{t+2} + \epsilon_{t+1}.$$

The distribution of $y_s - y_t$ is:

$$y_s - y_t \sim N(0, s-t).$$

(b) Rewrite as follows:

$$\begin{aligned} y_t &= y_{t-1} + \epsilon_t \\ &= y_{t-1} + e_{1,t} + e_{2,t} + \cdots + e_{N,t}, \end{aligned}$$

where $\epsilon_t = e_{1,t} + e_{2,t} + \cdots + e_{N,t}$.

$e_{1,t}, e_{2,t}, \cdots, e_{N,t}$ are iid with $e_{i,t} \sim N(0, 1/N)$.

That is, suppose that there are N subperiods between time t and time $t+1$.