平成**26**年度

計量経済学特別講義

時系列分析入門

2 月 **6** 日（金）

13:00 – 14:30　入門時系列モデル

14:45 – 16:15　**VAR (vector Auto Regression)** モデル分析

2 月 **13** 日（金）

14:45 – 16:15　単位根・共和分析

16:30 – 18:00　**Volatility Models (**セミナー**)**

場所： 経済研究所 第一共同研究室 （経済研究所 本館 ４ **F**）

代表的テキスト：

・J.D. Hamilton (1994) *Time Series Analysis*

沖本・井上訳 (2006)『時系列解析 (上・下)』

・A.C. Harvey (1981) *Time Series Models*

国友・山本訳 (1985)『時系列モデル入門』

・沖本竜義 (2010)『経済・ファイナンスデータの計量時系列分析』

1 Time Series Analysis (時系列分析)

1.1 Introduction

1. **Stationarity (定常性) :**

   Let $y_1, y_2, \cdots, y_T$ be time series data.

   (a) **Weak Stationarity (弱定常性) :**

   $$\mathrm{E}(y_t) = \mu,$$

   $$\mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \gamma(\tau), \qquad \tau = 0, 1, 2, \cdots$$

   The first moment does not depend on time.

   The second moment depends only on time difference.

(b) **Strong Stationarity (強定常性) :**

Let $f(y_{t_1}, y_{t_2}, \cdots, y_{t_r})$ be the joint distribution of $y_{t_1}, y_{t_2}, \cdots, y_{t_r}$.

$$f(y_{t_1}, y_{t_2}, \cdots, y_{t_r}) = f(y_{t_1+\tau}, y_{t_2+\tau}, \cdots, y_{t_r+\tau})$$

All the moments are same for all τ.

2. **Auto-covariance Function (自己共分散関数) :**

$$\mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \gamma(\tau), \qquad \tau = 0, 1, 2, \cdots$$

$$\gamma(\tau) = \gamma(-\tau)$$

3. **Auto-correlation Function (自己相関関数) :**

$$\rho(\tau) = \frac{\mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu))}{\sqrt{\mathrm{Var}(y_t)}\sqrt{\mathrm{Var}(y_{t-\tau})}} = \frac{\gamma(\tau)}{\gamma(0)}$$

Note that $\mathrm{Var}(y_t) = \mathrm{Var}(y_{t-\tau}) = \gamma(0)$ in the case of stationary process.

4. **Partial Autocorrelation Coefficient (偏自己相関係数), $\boldsymbol{\phi_{k,k}}$:**

The partial autocorrelation coefficient between y_t and y_{t-k}, denoted by $\phi_{k,k}$, is a measure of strength of the relationship between y_t and y_{t-k}, after removing influence of $y_{t-1}, \cdots, y_{t-k+1}$.

$$\phi_{1,1} = \rho(1)$$

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

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

$$\vdots$$

$$\begin{pmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & & \rho(k-3) & \rho(k-2) \\ \vdots & \vdots & & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \vdots \\ \phi_{k,k} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{pmatrix}$$

Use Cramer's rule (クラメールの公式) to obtain $\phi_{k,k}$.

$$\phi_{k,k} = \frac{\begin{vmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(1) \\ \rho(1) & 1 & & \rho(k-3) & \rho(2) \\ \vdots & \vdots & & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & \rho(k) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & & \rho(k-3) & \rho(k-2) \\ \vdots & \vdots & & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & 1 \end{vmatrix}}$$

5. **Sample Mean (標本平均) :**

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

6. **Sample Auto-covariance (標本自己共分散) :**

$$\hat{\gamma}(\tau) = \frac{1}{T}\sum_{t=\tau+1}^{T} (y_t - \hat{\mu})(y_{t-\tau} - \hat{\mu})$$

7. **Correlogram (コレログラム, or 標本自己相関関数) :**

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}$$

8. **Lag Operator (ラグ作要素) :**

$$L^{\tau} y_t = y_{t-\tau}, \qquad \tau = 1, 2, \cdots$$

9. **Likelihood Function (尤度関数)** — Innovation Form :

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

$$\begin{aligned} f(y_1, , y_2, \cdots, y_T) &= f(y_T | y_{T-1}, \cdots, y_1) f(y_{T-1}, \cdots, y_1) \\ &= f(y_T | y_{T-1}, \cdots, y_1) f(y_{T-1} | y_{T-2}, \cdots, y_1) f(y_{T-2}, \cdots, y_1) \\ &\vdots \\ &= f(y_T | y_{T-1}, \cdots, y_1) f(y_{T-1} | y_{T-2}, \cdots, y_1) \; \cdots \; f(y_2 | y_1) f(y_1) \\ &= f(y_1) \prod_{t=2}^{T} f(y_t | y_{t-1}, \cdots, y_1). \end{aligned}$$

Therefore, the log-likelihood function is given by:

$$\log f(y_1, y_2, \cdots, y_T) = \log f(y_1) + \sum_{t=2}^{T} \log f(y_t | y_{t-1}, \cdots, y_1).$$

Under linear model with normality assumption, $f(y_t | y_{t-1}, \cdots, y_1)$ is given by the normal distribution with mean $\mathrm{E}(y_t | y_{t-1}, \cdots, y_1)$ and variance $\mathrm{Var}(y_t | y_{t-1}, \cdots, y_1)$.

1.2 Time Series Models (時系列モデル)

Autoregressive Model (自己回帰モデル or AR モデル): AR(p)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

Moving Average Model (移動平均モデル or MA モデル): MA(q)

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

ARMA Model: ARMA(p, q)

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

ARIMA Model: $\text{ARIMA}(p, d, q)$

$$\Delta y_t = y_t - y_{t-1} = (1 - L)y_t,$$

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = (1 - L)^2 y_t,$$

$$\vdots$$

$$\Delta^d y_t = (1 - L)^d y_t.$$

$$\Delta^d y_t \sim \text{ARMA}(p, q) \iff y_t \sim \text{ARIMA}(p, d, q)$$

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

SARIMA Model: $\text{SARIMA}(p, d, q)$

$\Delta_s y_t = y_t - y_{t-s}$, $\quad s = 4$ for quarterly data, and $s = 12$ for monthly data

$$\Delta_s \Delta^d y_t \sim \text{ARMA}(p, q) \iff \Delta_s y_t \sim \text{ARIMA}(p, d, q)$$

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

1.3 Autoregressive Model (自己回帰モデル or AR モデル)

1. **AR(p) Model :**

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

   which is rewritten as:

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

   where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p.$$

2. **Stationarity (定常性) :**

   Suppose that all the p solutions of x from $\phi(x) = 0$ are real numbers

   When the p solutions are greater than one in absolute value, y_t is stationary.

Suppose that the p solutions include imaginary numbers.

When the p solutions are outside unit circle, y_t is stationary.

Example: AR(1) Model: $y_t = \phi_1 y_{t-1} + \epsilon_t$ for $\epsilon_t \sim$ iid $N(0, \sigma^2)$

1. The stationarity condition is: the solution of $\phi(x) = 1 - \phi_1 x = 0$, i.e., $x = 1/\phi_1$, is greater than one in absolute value, or equivalently, $|\phi_1| < 1$.

2. Rewriting the AR(1) model,

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \epsilon_t \\ &= \phi_1^2 y_{t-2} + \epsilon_t + \phi_1 \epsilon_{t-1} \\ &= \phi_1^3 y_{t-3} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} \\ &\;\vdots \\ &= \phi_1^\tau y_{t-\tau} + \epsilon_t + \phi_1 \epsilon_{t-1} + \cdots + \phi_1^{\tau-1} \epsilon_{t-\tau+1}. \end{aligned}$$

As τ goes to infinity, ϕ_1^τ approaches zero. $\Longrightarrow$ Stationarity condition

3. For stationarity, $y_t = \phi_1 y_{t-1} + \epsilon_t$ is rewritten as:

$$y_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots$$

   MA representation of AR model, i.e., AR(1)=MA(∞)

4. Mean of AR(1) process, μ

$$\begin{aligned}\mu = \mathrm{E}(y_t) &= \mathrm{E}(\epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots) \\ &= \mathrm{E}(\epsilon_t) + \phi_1 \mathrm{E}(\epsilon_{t-1}) + \phi_1^2 \mathrm{E}(\epsilon_{t-2}) + \cdots = 0\end{aligned}$$

5. Variance of AR(1) process, $\gamma(0)$

$$\begin{aligned}\gamma(0) = \mathrm{V}(y_t) &= \mathrm{V}(\epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots) \\ &= \mathrm{V}(\epsilon_t) + \mathrm{V}(\phi_1 \epsilon_{t-1}) + \mathrm{V}(\phi_1^2 \epsilon_{t-2}) + \cdots\end{aligned}$$

$$= \mathrm{V}(\epsilon_t) + \phi_1^2 \mathrm{V}(\epsilon_{t-1}) + \phi_1^4 \mathrm{V}(\epsilon_{t-2}) + \cdots$$

$$= \sigma^2 (1 + \phi_1^2 + \phi_1^4 + \cdots) = \frac{\sigma^2}{1 - \phi_1^2}$$

Note that $\mathrm{V}(aX + b) = a^2 \mathrm{V}(X)$ and $\mathrm{V}(X + Y) = \mathrm{V}(X) + \mathrm{V}(Y)$ when two random variables X and Y are independent.

6. Autocovariance and autocorrelation functions of the AR(1) process:

   Rewriting the AR(1) process, we have:

$$y_t = \phi_1^\tau y_{t-\tau} + \epsilon_t + \phi_1 \epsilon_{t-1} + \cdots + \phi_1^{\tau-1} \epsilon_{t-\tau+1}.$$

   Therefore, for $\tau = 1, 2, \cdots$, the autocovariance function of AR(1) process is:

$$\gamma(\tau) = \mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \mathrm{E}(y_t y_{t-\tau})$$

$$= \mathrm{E}\Big((\phi_1^\tau y_{t-\tau} + \epsilon_t + \phi_1 \epsilon_{t-1} + \cdots + \phi_1^{\tau-1} \epsilon_{t-\tau+1}) y_{t-\tau} \Big)$$

$$= \phi_1^\tau \mathrm{E}(y_{t-\tau} y_{t-\tau}) + \mathrm{E}(\epsilon_t y_{t-\tau}) + \phi_1 \mathrm{E}(\epsilon_{t-1} y_{t-\tau}) + \cdots + \phi_1^{\tau-1} \mathrm{E}(\epsilon_{t-\tau+1} y_{t-\tau})$$

$$= \phi_1^\tau \gamma(0).$$

The autocorrelation function of AR(1) process is:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \phi_1^\tau.$$

Or, multiply $y_{t-\tau}$ on both sides of the AR(1) process and take the expectation:

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

$$\gamma(\tau) = \begin{cases} \phi_1 \gamma(\tau - 1), & \text{for } \tau \neq 0, \\ \phi_1 \gamma(\tau - 1) + \sigma^2, & \text{for } \tau = 0. \end{cases}$$

Using $\gamma(\tau) = \gamma(-\tau)$, $\gamma(\tau)$ for $\tau = 0$ is given by:

$$\gamma(0) = \phi_1 \gamma(1) + \sigma^2 = \phi_1^2 \gamma(0) + \sigma^2.$$

Note that $\gamma(1) = \phi_1\gamma(0)$. Therefore, $\gamma(0)$ is given by: $\gamma(0) = \dfrac{\sigma^2}{1 - \phi_1^2}$

7. Partial autocorrelation function of AR(1) process:

$$\phi_{1,1} = \rho(1) = \phi_1$$

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

8. Estimation of AR(1) model:

   (a) Likelihood function

$$\log f(y_T, \cdots, y_1) = \log f(y_1) + \sum_{t=1}^{T} \log f(y_t | y_{t-1}, \cdots, y_1)$$

$$
\begin{aligned}
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi_1^2}\right) - \frac{1}{\sigma^2/(1-\phi_1^2)}y_1^2 \\
&\quad -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})^2
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}\log\left(\frac{1}{1-\phi_1^2}\right) \\
&\quad -\frac{1}{2\sigma^2/(1-\phi_1^2)}y_1^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})^2
\end{aligned}
$$

Note as follows:

$$
f(y_1) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi_1^2)}}\exp\left(-\frac{1}{2\sigma^2/(1-\phi_1^2)}y_1^2\right)
$$

$$
f(y_t|y_{t-1},\cdots,y_1) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(y_t - \phi_1 y_{t-1})^2\right)
$$

$$\frac{\partial \log f(y_T, \cdots, y_1)}{\partial \sigma^2} = -\frac{T}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4/(1-\phi_1^2)}y_1^2 + \frac{1}{2\sigma^4}\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})^2 = 0$$

$$\frac{\partial \log f(y_T, \cdots, y_1)}{\partial \phi_1} = -\frac{\phi_1}{1-\phi_1^2} + \frac{\phi_1}{\sigma^2}y_1^2 + \frac{1}{\sigma^2}\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})y_{t-1} = 0$$

The MLE of ϕ_1 and σ^2 satisfies the above two equation.

$$\tilde{\sigma}^2 = \frac{1}{T}\left((1-\tilde{\phi}_1^2)y_1^2 + \sum_{t=2}^{T}(y_t - \tilde{\phi}_1 y_{t-1})^2\right)$$

$$\tilde{\phi}_1 = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} + \left(\tilde{\phi}_1 y_1^2 - \frac{\tilde{\sigma}^2\tilde{\phi}_1}{1-\tilde{\phi}_1^2}\right)\Big/\sum_{t=2}^{T} y_{t-1}^2$$

(b) Ordinary Least Squares (OLS) Method

$$S(\phi_1) = \sum_{t=2}^{T} (y_t - \phi_1 y_{t-1})^2$$

is minimized with respect to ϕ_1.

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

OLSE of ϕ_1 is a consistent estimator.

The following equations are utilized.

$$\mathrm{E}(y_{t-1}\epsilon_t) = 0$$

$$\mathrm{E}(y_{t-1}^2) = \mathrm{Var}(y_{t-1}) = \gamma(0)$$

9. Some formulas:

(a) Central Limit Theorem

Random variables $x_1, x_2, \cdots, x_T$ are mutually independently distributed with mean μ and variance σ^2. Define $\overline{x} = (1/T)\sum_{t=1}^{T} x_t$.

Then,

$$\frac{\overline{x} - \mathrm{E}(\overline{x})}{\sqrt{\mathrm{V}(\overline{x})}} = \frac{\overline{x} - \mu}{\sigma/\sqrt{T}} \longrightarrow N(0, 1)$$

(b) Central Limit Theorem II

Random variables $x_1, x_2, \cdots, x_T$ are distributed with mean μ and variance σ^2. Define $\overline{x} = (1/T)\sum_{t=1}^{T} x_t$.

Then,

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

(c) Let x and y be random variables.

y converges in distribution to a distribution, and x converges in probability to a fixed value. Then, xy converges in distribution.

For example, consider: $y \longrightarrow N(\mu, \sigma^2)$ and $x \longrightarrow c$.

Then, we obtain: $xy \longrightarrow N(c\mu, c^2\sigma^2)$.

10. Asymptotic distribution of OLSE $\hat{\phi}_1 = \dfrac{\sum_{t=2}^T y_{t-1}y_t}{\sum_{t=2}^T y_{t-1}^2} = \phi_1 + \dfrac{(1/T)\sum_{t=2}^T y_{t-1}\epsilon_t}{(1/T)\sum_{t=2}^T y_{t-1}^2}$:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) \longrightarrow N(0, 1 - \phi_1^2)$$

Proof:

$y_{t-1}\epsilon_t$ is distributed with mean $\mathrm{E}(y_{t-1}\epsilon_t) = 0$ and variance $\mathrm{V}(y_{t-1}\epsilon_t) = \mathrm{V}(y_{t-1})\mathrm{V}(\epsilon_t) = \sigma^2\gamma(0) = \dfrac{\sigma^4}{1 - \phi_1^2}$.

$y_{t-1}\epsilon_t$, $t = 1, 2, \cdots, T$, are iid, because $\mathrm{Cov}(y_{t-1}\epsilon_t,\ y_{s-1}\epsilon_s) = \mathrm{E}(y_{t-1}y_{s-1}\epsilon_t\epsilon_s) = 0$ for $t > s$.

From the central limit theorem,

$$\frac{(1/T)\sum_{t=1}^{T} y_{t-1}\epsilon_t - \mathrm{E}((1/T)\sum_{t=1}^{T} y_{t-1}\epsilon_t)}{\sqrt{\mathrm{V}((1/T)\sum_{t=1}^{T} y_{t-1}\epsilon_t)}} = \frac{(1/T)\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\sqrt{\sigma^4/(1-\phi_1^2)}/\sqrt{T}} \longrightarrow N(0,1).$$

Rewriting,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow N(0, \frac{\sigma^4}{1-\phi_1^2}).$$

Next,

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \mathrm{E}(y_{t-1}^2) = \gamma(0) = \frac{\sigma^2}{1-\phi_1^2}$$

yields:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) = \frac{(1/\sqrt{T})\sum_{t=1}^{T} y_{t-1}\epsilon_t}{(1/T)\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow N(0, 1-\phi_1^2).$$

11. **AR(1) +drift:** $y_t = \mu + \phi_1 y_{t-1} + \epsilon_t$

Mean:

Using the lag operator,

$$\phi(L) y_t = \mu + \epsilon_t$$

where $\phi(L) = 1 - \phi_1 L$.

Multiply $\phi(L)^{-1}$ on both sides. Then, when $|\phi_1| < 1$, we have:

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

Taking the expectation on both sides,

$$\begin{aligned} \mathrm{E}(y_t) &= \phi(L)^{-1}\mu + \phi(L)^{-1}\mathrm{E}(\epsilon_t) \\ &= \phi(1)^{-1}\mu = \frac{\mu}{1 - \phi_1} \end{aligned}$$

Example: AR(2) Model: Consider $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$.

1. The stationarity condition is: two solutions of x from $\phi(x) = 1 - \phi_1 x - \phi_2 x^2 = 0$ are outside the unit circle.

2. Rewriting the AR(2) model,

$$(1 - \phi_1 L - \phi_2 L^2) y_t = \epsilon_t.$$

   Let $1/\alpha_1$ and $1/\alpha_2$ be the solutions of $\phi(x) = 0$.

   Then, the AR(2) model is written as:

$$(1 - \alpha_1 L)(1 - \alpha_2 L) y_t = \epsilon_t,$$

   which is rewritten as:

$$y_t = \frac{1}{(1 - \alpha_1 L)(1 - \alpha_2 L)} \epsilon_t$$

$$= \left(\frac{\alpha_1/(\alpha_1 - \alpha_2)}{1 - \alpha_1 L} + \frac{-\alpha_2/(\alpha_1 - \alpha_2)}{1 - \alpha_2 L} \right) \epsilon_t$$

3. **Mean of AR(2) Model:**

   When y_t is stationary, i.e., α_1 and α_2 are outside the unit circle,

$$\mu = \mathrm{E}(y_t) = \mathrm{E}(\phi(L)\epsilon_t) = 0$$

4. **Autocovariance Function of AR(2) Model:**

$$\begin{aligned}
\gamma(\tau) &= \mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \mathrm{E}(y_t y_{t-\tau}) \\
&= \mathrm{E}\Big((\phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t) y_{t-\tau}\Big) \\
&= \phi_1 \mathrm{E}(y_{t-1} y_{t-\tau}) + \phi_2 \mathrm{E}(y_{t-2} y_{t-\tau}) + \mathrm{E}(\epsilon_t y_{t-\tau}) \\
&= \begin{cases} \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2), & \text{for } \tau \neq 0, \\ \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2) + \sigma^2, & \text{for } \tau = 0. \end{cases}
\end{aligned}$$

The initial condition is obtained by solving the following three equations:

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2,$$

$$\gamma(1) = \phi_1\gamma(0) + \phi_2\gamma(1),$$

$$\gamma(2) = \phi_1\gamma(1) + \phi_2\gamma(0).$$

Therefore, the initial conditions are given by:

$$\gamma(0) = \left(\frac{1-\phi_2}{1+\phi_2}\right)\frac{\sigma^2}{(1-\phi_2)^2-\phi_1^2},$$

$$\gamma(1) = \frac{\phi_1}{1-\phi_2}\gamma(0) = \left(\frac{\phi_1}{1-\phi_2}\right)\left(\frac{1-\phi_2}{1+\phi_2}\right)\frac{\sigma^2}{(1-\phi_2)^2-\phi_1^2}.$$

Given $\gamma(0)$ and $\gamma(1)$, we obtain $\gamma(\tau)$ as follows:

$$\gamma(\tau) = \phi_1\gamma(\tau-1) + \phi_2\gamma(\tau-2), \qquad \text{for } \tau = 2, 3, \cdots.$$

5. **Another solution for $\gamma(0)$:**

From $\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2$,

$$\gamma(0) = \frac{\sigma^2}{1 - \phi_1\rho(1) - \phi_2\rho(2)}$$

where

$$\rho(1) = \frac{\phi_1}{1-\phi_2}, \qquad \rho(2) = \phi_1\rho(1) + \phi_2 = \frac{\phi_1^2 + (1-\phi_2)\phi_2}{1-\phi_2}.$$

6. **Autocorrelation Function of AR(2) Model:**

Given $\rho(1)$ and $\rho(2)$,

$$\rho(\tau) = \phi_1\rho(\tau-1) + \phi_2\rho(\tau-2), \qquad \text{for } \tau = 3, 4, \cdots,$$

7. $\boldsymbol{\phi_{k,k}}$ **= Partial Autocorrelation Coefficient of AR(2) Process:**

$$\begin{pmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & & \rho(k-3) & \rho(k-2) \\ \vdots & \vdots & & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \vdots \\ \phi_{k,k-1} \\ \phi_{k,k} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{pmatrix},$$

for $k = 1, 2, \cdots$.

$$
\phi_{k,k} = \frac{\begin{vmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(1) \\ \rho(1) & 1 & & \rho(k-3) & \rho(2) \\ \vdots & \vdots & & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & \rho(k) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & & \rho(k-3) & \rho(k-2) \\ \vdots & \vdots & & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & 1 \end{vmatrix}}
$$

Autocovariance Functions:

$$\gamma(1) = \phi_1\gamma(0) + \phi_2\gamma(1),$$

$$\gamma(2) = \phi_1\gamma(1) + \phi_2\gamma(0),$$

$$\gamma(\tau) = \phi_1\gamma(\tau-1) + \phi_2\gamma(\tau-2), \qquad \text{for } \tau = 3, 4, \cdots.$$

Autocorrelation Functions:

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

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

$$\rho(\tau) = \phi_1\rho(\tau-1) + \phi_2\rho(\tau-2), \qquad \text{for } \tau = 3, 4, \cdots.$$

$$\phi_{1,1} = \rho(1) = \frac{\phi_1}{1 - \phi_2}$$

$$\phi_{2,2} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} = \phi_2$$

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

$$= \frac{(\rho(3)-\rho(1)\rho(2)) - \rho(1)^2(\rho(3)-\rho(1)) + \rho(2)\rho(1)(\rho(2)-1)}{(1-\rho(1)^2) - \rho(1)^2(1-\rho(2)) + \rho(2)(\rho(1)^2-\rho(2))} = 0.$$

8. **Log-Likelihood Function — Innovation Form:**

$$\log f(y_T, \cdots, y_1) = \log f(y_2, y_1) + \sum_{t=3}^{T} \log f(y_t|y_{t-1}, \cdots, y_1)$$

where

$$f(y_2, y_1) = \frac{1}{2\pi} \begin{vmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{vmatrix}^{-1/2} \exp\left(-\frac{1}{2}(y_1\ y_2) \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right),$$

$$f(y_t|y_{t-1}, \cdots, y_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2 \right).$$

Note as follows:

$$\begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix} = \gamma(0) \begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} = \gamma(0) \begin{pmatrix} 1 & \phi_1/(1-\phi_2) \\ \phi_1/(1-\phi_2) & 1 \end{pmatrix}.$$

9. **AR(2) +drift:** $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$

Mean:

Rewriting the AR(2)+drift model,

$$\phi(L)y_t = \mu + \epsilon_t$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$.

Under the stationarity assumption, we can rewrite the AR(2)+drift model as follows:

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

Therefore,

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

Example: AR(p) model: Consider $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$.

1. **Variance of AR(p) Process:**

   Under the stationarity condition (i.e., the p solutions of x from $\phi(x) = 0$ are outside the unit circle),

   $$\gamma(0) = \frac{\sigma^2}{1 - \phi_1\rho(1) - \cdots - \phi_p\rho(p)}.$$

   Note that $\gamma(\tau) = \rho(\tau)\gamma(0)$.

   Solve the following simultaneous equations for $\tau = 0, 1, \cdots, p$:

   $$\begin{aligned}\gamma(\tau) &= \mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)) = \mathrm{E}(y_t y_{t-\tau}) \\ &= \begin{cases} \phi_1\gamma(\tau-1) + \phi_2\gamma(\tau-2) + \cdots + \phi_p\gamma(\tau-p), & \text{for } \tau \neq 0, \\ \phi_1\gamma(\tau-1) + \phi_2\gamma(\tau-2) + \cdots + \phi_p\gamma(\tau-p) + \sigma^2, & \text{for } \tau = 0. \end{cases}\end{aligned}$$

2. **Estimation of AR(p) Model:**

1. OLS:

$$\min_{\phi_1,\cdots,\phi_p} \sum_{t=p+1}^{T} (y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p})^2$$

2. MLE:

$$\max_{\phi_1,\cdots,\phi_p} \log f(y_T,\cdots,y_1)$$

where

$$\log f(y_T,\cdots,y_1) = \log f(y_p,\cdots,y_2,y_1) + \sum_{t=p+1}^{T} \log f(y_t|y_{t-1},\cdots,y_1),$$

$$f(y_p,\cdots,y_2,y_1) = (2\pi)^{-p/2}|V|^{-1/2} \exp\left(-\frac{1}{2}(y_1\ y_2\ \cdots\ y_p)V^{-1}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}\right)$$

$$V = \gamma(0) \begin{pmatrix} 1 & \rho(1) & \cdots & \rho(p-2) & \rho(p-1) \\ \rho(1) & 1 & & \rho(p-3) & \rho(p-2) \\ \vdots & \vdots & & \vdots & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & \rho(1) & 1 \end{pmatrix}$$

$$f(y_t | y_{t-1}, \cdots, y_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p})^2 \right)$$

3. Yule=Walker (ユール・ウォーカー) Equation:

Multiply y_{t-1}, y_{t-2}, $\cdots$, y_{t-p} on both sides of $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t = y_t$, take expectations for each case, and divide by variance $\gamma(0)$. Moreover, replace the autocorrelation function $\rho(\tau)$ by the correlogram $\hat{\rho}(\tau)$.

$$\begin{pmatrix} 1 & \hat{\rho}(1) & \cdots & \hat{\rho}(p-2) & \hat{\rho}(p-1) \\ \hat{\rho}(1) & 1 & & \hat{\rho}(p-3) & \hat{\rho}(p-2) \\ \vdots & \vdots & & \vdots & \vdots \\ \hat{\rho}(p-1) & \hat{\rho}(p-2) & \cdots & \hat{\rho}(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{pmatrix} = \begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(p) \end{pmatrix}$$

where

$$\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=\tau+1}^{T} (y_t - \hat{\mu})(y_{t-\tau} - \hat{\mu}), \qquad \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} y_t, \qquad \hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}.$$

3. **AR(p) +drift:** $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots \phi_p y_{t-p} + \epsilon_t$

Mean:

$$\phi(L) y_t = \mu + \epsilon_t$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$.

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

Taking the expectation on both sides,

$$\begin{aligned} \mathrm{E}(y_t) &= \phi(L)^{-1}\mu + \phi(L)^{-1}\mathrm{E}(\epsilon_t) = \phi(1)^{-1}\mu \\ &= \frac{\mu}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \end{aligned}$$

4. **Partial Autocorrelation of AR(p) Process:**

$\phi_{k,k} = 0$ for $k = p + 1, p + 2, \cdots$.

1.4 MA Model

MA (Moving Average，移動平均) Model:

1. MA(q)

$$y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q},$$

which is rewritten as:

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

where

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

2. **Invertibility (反転可能性):**

   The q solutions of x from $\theta(x) = 1 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_q x^q = 0$ are outside the unit circle.

   $\Longrightarrow$ MA(q) model is rewritten as AR(∞) model.

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

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

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

2. **Autocovariance Function of MA(1) Process:**

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

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

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

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.

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

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

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

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

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

$$\vdots$$

$$\begin{pmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & & \rho(k-3) & \rho(k-2) \\ \vdots & \vdots & & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \vdots \\ \phi_{k,k} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{pmatrix}$$

$\phi_{k,k} \neq 0$ for $k = 1, 2, 3, \cdots$

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

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

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

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

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

The autocovariance functions are: $\gamma(0) = (1 + \theta_1^2)\sigma^2$, $\gamma(1) = \theta_1 \sigma^2$, 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

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

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:

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

2. Autocorrelation Function of MA(2) Process:

$$\rho(\tau) = \begin{cases} \dfrac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, & \text{for } \tau = 1, \\ \dfrac{\theta_2}{1 + \theta_1^2 + \theta_2^2} & \text{for } \tau = 2, \\ 0, & \text{for } \tau = 3, 4, 5, \cdots. \end{cases}$$

3. Partial Autocorrelation Function of MA(2) Process: $\phi_{k,k} \neq 0$ for $k = 1, 2, \cdots$.

4. 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:

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

5. **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

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

6. **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:**

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

   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:

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

1.5 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:

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

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

Therefore, we obtain;

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

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

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

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

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

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

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

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,

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

1.6 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.

1.7 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.

1.8 Identification (識別，または，同定)

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 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.

1.9 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


. predict resid, r
(13 missing values generated)

. corrgram resid

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1       -0.0132  -0.0132  .08814  0.7666          |                  |
2       -0.0095  -0.0097   .1341  0.9351          |                  |
3        0.1248   0.1246  8.0433  0.0451          |                  |
4       -0.0644  -0.0624  10.154  0.0379          |                  |
5       -0.0001   0.0011  10.154  0.0710          |                  |
6       -0.0138  -0.0309  10.252  0.1144          |                  |
7       -0.0032   0.0126  10.257  0.1745          |                  |
8        0.0958   0.0938   14.97  0.0597          |                  |
9       -0.0317  -0.0255  15.487  0.0784          |                  |
```

10	0.0126	0.0112	15.569	0.1127
11	-0.0053	-0.0305	15.583	0.1573
12	-0.3773	-0.3837	89.235	0.0000
13	0.0408	0.0258	90.098	0.0000
14	-0.0233	-0.0307	90.381	0.0000
15	-0.0911	-0.0059	94.703	0.0000

1.10 ARCH and GARCH Models

Autoregressive Conditional Heteroskedasticity (ARCH)

Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

1. ARCH (p) Model

$$\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \cdots, \epsilon_1 \sim N(0, \sigma_t^2),$$

   where,

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2.$$

   The unconditional variance of ϵ_t is:

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p}$$

2. GARCH (p, q) Model

$$\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \cdots, \epsilon_1 \sim N(0, \sigma_t^2),$$

where

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_p\epsilon_{t-p}^2 + \beta_1\sigma_{t-1}^2 + \cdots + \beta_q\sigma_{t-q}^2.$$

3. Application to OLS (Case of ARCH(1) Model):

$$y_t = x_t\beta + \epsilon_t, \qquad \epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, \cdots, \epsilon_1 \sim N(0, \alpha_0 + \alpha_1\epsilon_{t-1}^2).$$

The joint density of $\epsilon_1, \epsilon_2, \cdots, \epsilon_T$ is:

$$\begin{aligned} f(\epsilon_1, \cdots, \epsilon_T) &= f(\epsilon_1)\prod_{t=2}^{T} f(\epsilon_t|\epsilon_{t-1}, \cdots, \epsilon_1) \\ &= (2\pi)^{-1/2}\left(\frac{\alpha_0}{1-\alpha_1}\right)^{-1/2} \exp\left(-\frac{1}{2\alpha_0/(1-\alpha_1)}\epsilon_1^2\right) \\ &\quad \times (2\pi)^{-(T-1)/2}\prod_{t=2}^{T}(\alpha_0 + \alpha_1\epsilon_{t-1}^2)^{-1/2} \exp\left(-\frac{1}{2}\sum_{t=2}^{T}\frac{\epsilon_t^2}{\alpha_0 + \alpha_1\epsilon_{t-1}^2}\right). \end{aligned}$$

The log-likelihood function is:

$$
\begin{aligned}
&\log L(\beta, \alpha_0, \alpha_1; y_1, \cdots, y_T) \\
&\quad = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\alpha_0}{1-\alpha_1}\right) - \frac{1}{2\alpha_0/(1-\alpha_1)}(y_1 - x_1\beta)^2 \\
&\qquad -\frac{T-1}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\log\left(\alpha_0 + \alpha_1(y_{t-1} - x_{t-1}\beta)^2\right) \\
&\qquad -\frac{1}{2}\sum_{t=2}^{T}\frac{(y_t - x_t\beta)^2}{\alpha_0 + \alpha_1(y_{t-1} - x_{t-1}\beta)^2}.
\end{aligned}
$$

Obtain α_0, α_1 and β such that the log-likelihood function is maximized.

$\alpha_0 > 0$ and $\alpha_1 > 0$ have to be satisfied.

These two conditions are explicitly included, when the model is modified to:

$\mathrm{E}(\epsilon_t^2 | \epsilon_{t-1}, \epsilon_{t-2}, \cdots, \epsilon_1) = \alpha_0^2 + \alpha_1^2 \epsilon_{t-1}^2$.

Testing the ARCH(1) Effect:

(a) Estimate $y_t = x_t\beta + u_t$ by OLS, and compute $\hat{\beta}$ and $\hat{u}_t = y_t - x_t\hat{\beta}$.

(b) Estimate $\hat{u}_t^2 = \alpha_0 + \alpha_1\hat{u}_{t-1}^2$ by OLS. If $\hat{\alpha}_1$ is significant, there is the ARCH(1) effect in the error term.

This test corresponds to LM test.

Example: GARCH(1,1) Model

```
. arch sdex l.sdex l2.sdex, arch(1) garch(1)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -5089.3558
Iteration 1:   log likelihood = -5086.7468
.........
.........

Iteration 22:  log likelihood = -5064.9328  (backed up)
Iteration 23:  log likelihood = -5064.9328

ARCH family regression
```

```
Sample: 16 - 516                                   Number of obs    =       501
Distribution: Gaussian                             Wald chi2(2)     =    225.19
Log likelihood = -5064.933                         Prob > chi2      =    0.0000

------------------------------------------------------------------------------
             |                 OPG
        sdex |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
sdex         |
        sdex |
         L1. |  -.6357273   .0426939   -14.89   0.000    -.7194059   -.5520488
         L2. |   -.370862   .0466222    -7.95   0.000    -.4622398   -.2794842
             |
       _cons |  -55.28043   261.2057    -0.21   0.832    -567.2341    456.6733
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |    .041632   .0123474     3.37   0.001     .0174317    .0658324
             |
       garch |
         L1. |   .9526041   .0148639    64.09   0.000     .9234715    .9817367
             |
       _cons |   312143.8   227564.3     1.37   0.170    -133873.9    758161.6
------------------------------------------------------------------------------
```

2 Vector Autoregressive (VAR) Model – Causality, Impulse Response Function and etc

Vector Autoregressive Process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

where

$$y_t : k \times 1, \qquad \mu : k \times 1, \qquad \epsilon_t : k \times 1, \qquad \phi_i : k \times k.$$

Rewriting the above equation,

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

where $\phi(L) = I_k - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$.

VAR(1) Model:

$$y_t = \phi_1 y_{t-1} + \epsilon_t, \qquad \text{i.e.,} \qquad (I_k - \phi_1 L) y_t = \epsilon_t.$$

When y_t is stationary, we obtain:

$$\begin{aligned} y_t &= (I_k - \phi_1 L)^{-1} \epsilon_t \\ &= (I_k + \phi_1 L + \phi_1^2 L^2 + \phi_1^3 L^3 + \cdots) \epsilon_t \\ &= \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \phi_1^3 \epsilon_{t-3} + \cdots \end{aligned}$$

VAR(1)=VMA(∞)

VAR(2) Model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \qquad \text{i.e.,} \qquad (I_k - \phi_1 L - \phi_2 L^2) y_{t-1} = \epsilon_t.$$

When y_t is stationary, we obtain:

$$y_{t-1} = (I_k - \phi_1 L - \phi_2 L^2)^{-1}\epsilon_t$$

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

VAR(2)=VMA(∞)

VAR(p) Model:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

i.e.,

$$(I_k - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)y_{t-1} = \epsilon_t.$$

When y_t is stationary, we obtain:

$$y_t = (I_k - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)^{-1}\epsilon_t$$

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

VAR(p)=VMA(∞)

2.1 Autocovariance Matrix and Autocorrelation Matrix

Let y_t be a $k \times 1$ vector.

Autocovariance Function Matrix:

$$\Gamma(\tau) = \mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)'), \qquad \tau = 0, 1, 2, \cdots,$$

where $\mathrm{E}(y_t) = \mu$. $\Gamma(\tau)$ is a $k \times k$ matrix.

$$\Gamma(\tau) = \Gamma(-\tau)'$$

Note that $\Gamma(-\tau) = \mathrm{E}((y_{t-\tau} - \mu)(y_t - \mu)')$ and $\Gamma(-\tau)' = \mathrm{E}((y_t - \mu)(y_{t-\tau} - \mu)') = \Gamma(\tau)$.

Autocorrelation Function Matrix:

$$\rho(\tau) = D^{-1/2}\Gamma(\tau)D^{-1/2},$$

where the (i, j)th element of D is given by $\gamma_{ii}(\tau) = \mathrm{V}(y_{it})$ for $i = j$ and zero otherwise.

$$\rho(\tau) = \rho(-\tau)'$$

2.2 Granger Cuasality Test (グレンジャー因果性テスト)

Consider a bivariate case.

Unrestricted Model (Sum of Squared Residuals, denoted by SSR_1):

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \phi_{11,p} & \phi_{12,p} \\ \phi_{21,p} & \phi_{22,p} \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

$H_0:\ \phi_{12,1} = \phi_{12,2} = \cdots = \phi_{12,p} = 0$

When H_0 is correct, we say there is no causality from y_2 to y_1.

$\Longrightarrow$ Granger Causality Test.

Restricted Model (Sum of Squared Residuals, denoted by SSR_0):

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \phi_{11,1} & 0 \\ \phi_{21,1} & \phi_{22,1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \phi_{11,p} & 0 \\ \phi_{21,p} & \phi_{22,p} \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

Asymptotically, we have the following distribution:

$$F = \frac{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/p}{\mathrm{SSR}_1/(T-2p-1)} \sim F(p, T-2p-1),$$

or

$$pF \sim \chi^2(p).$$

In general, we consider testing the Granger causality from y_j to y_i.

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

$y_t : k \times 1, \qquad \mu : k \times 1, \qquad \phi_p : k \times k, \qquad \epsilon_t : k \times 1.$

The null hypothesis is: $H_0 : \phi_{ij,1} = \phi_{ij,2} = \cdots = \phi_{ij,p} = 0.$

The alternative hypothesis is: H_1 : not H_0.

SSR_0 = Sum of Squared Residuals under H_0

SSR_1 = Sum of Squared Residuals under H_1

Under H_0, the asymptotic distribution is given by:

$$F = \frac{(\text{SSR}_0 - \text{SSR}_1)/p}{\text{SSR}_1/(T - kp - 1)} \sim F(p, T - kp - 1),$$

or

$$pF \sim \chi^2(p).$$

Example:

Data:　1994 年第一四半期～2014 年第一四半期

`gdp` = GDP (実質，10 億円，季調済，内閣府 HP から取得)

`def` = GDP デフレータ (季調済，内閣府 HP から取得)

`r` = 貸出約定平均金利 (%，新規，総合・国内銀行，日銀 HP から取得)

`m` = 通貨流通高（平均発行高，億円，季調済，日銀 HP から取得)

```
. gen time=_n

. tsset time
        time variable:  time, 1 to 81
                delta:  1 unit

. gen lgdp=log(gdp)

. gen lm=log(m/(def/10))

. varsoc d.lgdp d.r d.lm

   Selection-order criteria
   Sample:  6 - 81                                    Number of obs      =        76
  +---------------------------------------------------------------------------+
  |lag |    LL      LR      df    p      FPE       AIC      HQIC      SBIC    |
```

```
|----+---------------------------------------------------------------------------|
|  0 |   541.22                              1.4e-10  -14.1637  -14.1269  -14.0717 |
|  1 |  571.181  59.923*    9  0.000  8.2e-11* -14.7153* -14.5682* -14.3473* |
|  2 |  575.715  9.0675     9  0.431  9.2e-11  -14.5978  -14.3404  -13.9537 |
|  3 |   579.55  7.6704     9  0.568  1.1e-10  -14.4619  -14.0942  -13.5418 |
|  4 |  583.767  8.4328     9  0.491  1.2e-10   -14.336   -13.858  -13.1399 |
+---------------------------------------------------------------------------------+
  Endogenous:  D.lgdp D.r D.lm
   Exogenous:  _cons

. var d.lgdp d.r d.lm, lags(1)

Vector autoregression

Sample:  3 - 81                                    No. of obs      =        79
Log likelihood =  592.2334                         AIC             = -14.68945
FPE            =  8.38e-11                         HQIC            = -14.54526
Det(Sigma_ml)  =  6.18e-11                         SBIC            = -14.32954

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_lgdp                4     .010717   0.0422   3.480972   0.3232
D_r                   4     .087186   0.2553    27.0782   0.0000
D_lm                  4     .009434   0.2903   32.30929   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
D_lgdp       |
        lgdp |
         LD. |   .2031129   .1119361     1.81   0.070    -.0162778    .4225037
             |
```

r						
LD.	.0045431	.0120151	0.38	0.705	-.0190061	.0280922
lm						
LD.	.0152162	.1086739	0.14	0.889	-.1977807	.228213
_cons	.0019504	.0019124	1.02	0.308	-.0017978	.0056986
D_r						
lgdp						
LD.	.4341641	.9106374	0.48	0.634	-1.350652	2.218981
r						
LD.	.5085677	.0977469	5.20	0.000	.3169874	.7001481
lm						
LD.	.1845222	.8840978	0.21	0.835	-1.548278	1.917322
_cons	-.0202984	.0155578	-1.30	0.192	-.0507912	.0101943
D_lm						
lgdp						
LD.	-.1972406	.098541	-2.00	0.045	-.3903774	-.0041037
r						
LD.	-.029395	.0105773	-2.78	0.005	-.0501261	-.0086639
lm						
LD.	.4472679	.0956691	4.68	0.000	.2597599	.634776
_cons	.0071036	.0016835	4.22	0.000	.0038039	.0104033

```
. vargranger
```

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
D_lgdp	D.r	.14297	1	0.705
D_lgdp	D.lm	.0196	1	0.889
D_lgdp	ALL	.15705	2	0.924
D_r	D.lgdp	.22731	1	0.634
D_r	D.lm	.04356	1	0.835
D_r	ALL	.3039	2	0.859
D_lm	D.lgdp	4.0064	1	0.045
D_lm	D.r	7.7232	1	0.005
D_lm	ALL	10.798	2	0.005

2.3 Impulse Response Function (インパルス応答関数):

$$\frac{\partial y_{i,t+m}}{\partial \epsilon_{j,t}}, \qquad m = 1, 2, \cdots,$$

where i, $j = 1, 2, \cdots, k$.

Example: AR(p) Process:

When y_t is stationary, we obtain:

$$\begin{aligned} y_t &= (I_k - \phi_1 L - \phi_2 L^2 - \ \cdots \ - \phi_p L^p)^{-1} \epsilon_t \\ &= \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ \cdots \end{aligned}$$

The impulse response function is:

$$\frac{\partial y_{i,t+k}}{\partial \epsilon_{j,t}} = \theta_{ij,k}, \qquad k = 1, 2, \cdots,$$

where $\theta_{ij,k}$ denotes the (i, j)th element of θ_k.

$$
\begin{aligned}
y_t &= \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots \\
&= PP^{-1}\epsilon_t + \theta_1 PP^{-1}\epsilon_{t-1} + \theta_2 PP^{-1}\epsilon_{t-2} + \cdots \\
&= \Omega_0 \eta_t + \Omega_1 \eta_{t-1} + \Omega_2 \eta_{t-2} + \cdots,
\end{aligned}
$$

where $\mathrm{V}(\eta_t) = I_k$, and $\Omega_i = \theta_i P$ for $i = 0, 1, 2, \cdots$ and $\Omega_0 = P$.

$$
\frac{\partial y_{i,t+m}}{\partial \eta_{j,t}}, \qquad m = 1, 2, \cdots,
$$

where $i, j = 1, 2, \cdots, k$.

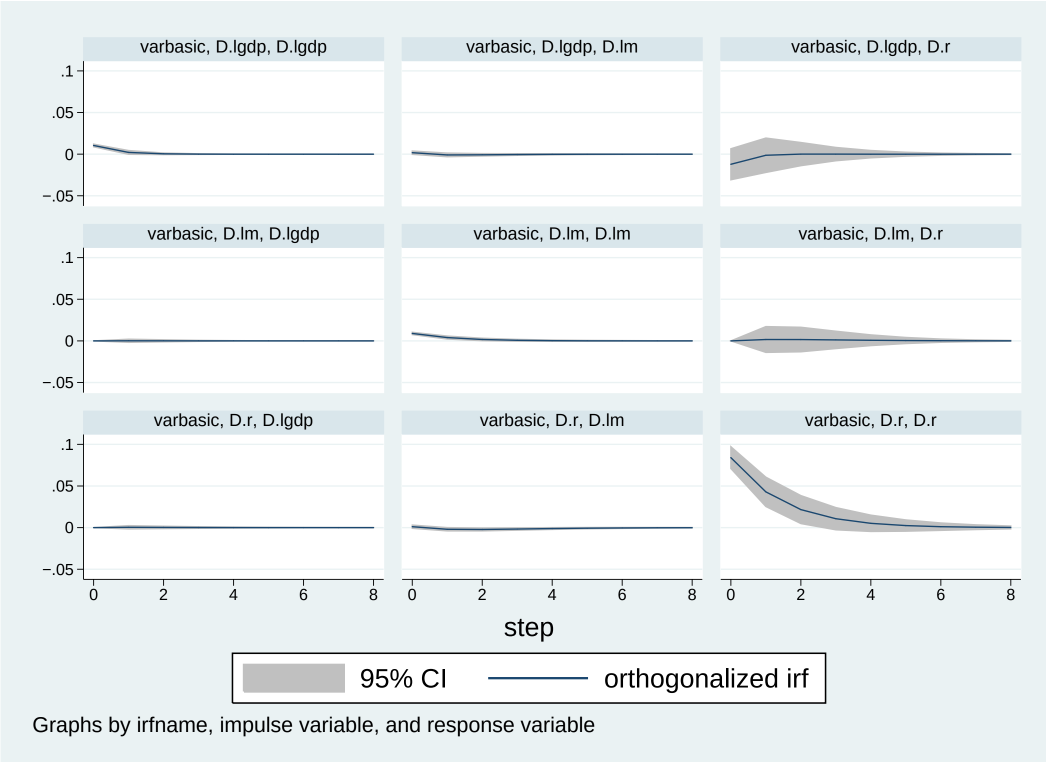
$\Longrightarrow$ **Orthogonalized Impulse Response Function (直交化インパルス応答関数)**

Example:

```
. varbasic d.lgdp d.r d.lm, lags(1)

Vector autoregression
```

```
Sample:  3 - 81                                    No. of obs      =        79
Log likelihood =  592.2334                         AIC             = -14.68945
FPE            =  8.38e-11                         HQIC            = -14.54526
Det(Sigma_ml)  =  6.18e-11                         SBIC            = -14.32954

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_lgdp               4     .010717   0.0422   3.480972   0.3232
D_r                  4     .087186   0.2553    27.0782   0.0000
D_lm                 4     .009434   0.2903   32.30929   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
D_lgdp       |
        lgdp |
         LD. |   .2031129   .1119361     1.81   0.070    -.0162778    .4225037
             |
           r |
         LD. |   .0045431   .0120151     0.38   0.705    -.0190061    .0280922
             |
          lm |
         LD. |   .0152162   .1086739     0.14   0.889    -.1977807     .228213
             |
       _cons |   .0019504   .0019124     1.02   0.308    -.0017978    .0056986
-------------+----------------------------------------------------------------
D_r          |
        lgdp |
         LD. |   .4341641   .9106374     0.48   0.634    -1.350652    2.218981
             |
           r |
```

```
         LD. |   .5085677   .0977469     5.20   0.000     .3169874    .7001481
             |
          lm |
         LD. |   .1845222   .8840978     0.21   0.835    -1.548278    1.917322
             |
       _cons |  -.0202984   .0155578    -1.30   0.192    -.0507912    .0101943
-------------+----------------------------------------------------------------
D_lm         |
        lgdp |
         LD. |  -.1972406    .098541    -2.00   0.045    -.3903774   -.0041037
             |
           r |
         LD. |   -.029395   .0105773    -2.78   0.005    -.0501261   -.0086639
             |
          lm |
         LD. |   .4472679   .0956691     4.68   0.000     .2597599     .634776
             |
       _cons |   .0071036   .0016835     4.22   0.000     .0038039    .0104033
------------------------------------------------------------------------------
```



varbasic, D.lgdp, D.lgdp
varbasic, D.lgdp, D.lm
varbasic, D.lgdp, D.r
varbasic, D.lm, D.lgdp
varbasic, D.lm, D.lm
varbasic, D.lm, D.r
varbasic, D.r, D.lgdp
varbasic, D.r, D.lm
varbasic, D.r, D.r
.1
.05
0
−.05
0
2
4
6
8
step
95% CI
orthogonalized irf
Graphs by irfname, impulse variable, and response variable

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

3.1 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 goes 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 $|\phi_1| < 1$:**

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

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 $|\phi_1| < 1$,

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

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

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

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

Therefore,

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

which is rewritten as:

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

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

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

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

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:

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

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 $\sigma^2 T$ on both sides, we have the following:

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

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

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

Note taht $Y \sim \chi^2(1)$ for $X \sim N(0,1)$ and $Y = X^2$. Moreover, the second term is derived from:

$$\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \longrightarrow \sigma^2.$$

Therefore,

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

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

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

Thus, we obtain the following result:

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

Therefore,

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

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 $\phi_1 = 1$.

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

More formally, we introduce **standard Brownian motion**.

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) **Definition:**

The **standard Brownian motion** or **Wiener process** $W(t)$ denotes a continuous-time variable at time t and a stochastic function.

$W(t)$ for $t \in [0, 1]$ satisfies the following:

i. $W(0) = 0$

ii. For any time periods $0 \leq r_1 < r_2 < \cdots < r_k \leq 1$, $W(r_2) - W(r_1)$, $W(r_3) - W(r_2)$, $\cdots$, $W(r_k) - W(r_{k-1})$ are independently multivariate normal with $W(s) - W(t) \sim N(0, s - t)$ for $s > t$.

iii. $W(t)$ is continuous in t with probability 1.

An example:

$$\sigma W(t) \sim N(0, \sigma^2 t),$$

which denotes the Brownian motion with variance σ^2.

Another example;

$$W(t)^2 \sim t \times \chi^2(1).$$

(c) Assume $\epsilon_t \sim$ iid $(0, \sigma^2)$. Define $X_T(r)$ for $r \in [0, 1]$ as follows:

$$X_T(r) = \begin{cases} 0, & 0 \le r < \dfrac{1}{T} \\ \dfrac{\epsilon_1}{T}, & \dfrac{1}{T} \le r < \dfrac{2}{T} \\ \dfrac{\epsilon_1 + \epsilon_2}{T}, & \dfrac{2}{T} \le r < \dfrac{3}{T} \\ \quad\vdots & \quad\vdots \\ \dfrac{\epsilon_1 + \epsilon_2 + \cdots + \epsilon_T}{T}, & r = 1 \end{cases}$$

Let $[Tr]$ be the largest integer which is less than or equal to $T \times r$.

$$X_T(r) \equiv \frac{1}{T} \sum_{t=1}^{[Tr]} \epsilon_t, \qquad \sqrt{T} X_T(r) \longrightarrow N(0, r\sigma^2).$$

Note that

$$X_T(r) = \frac{1}{T}\sum_{t=1}^{[Tr]} \epsilon_t = \frac{[Tr]}{T}\frac{1}{[Tr]}\sum_{t=1}^{[Tr]} \epsilon_t,$$

$$\frac{[Tr]}{T} \longrightarrow r, \qquad \frac{1}{\sqrt{[Tr]}}\sum_{t=1}^{[Tr]} \epsilon_t \longrightarrow N(0,\sigma^2),$$

$$\sqrt{T}X_T(r) = \frac{[Tr]}{T}\sqrt{\frac{T}{[Tr]}}\frac{1}{\sqrt{[Tr]}}\sum_{t=1}^{[Tr]} \epsilon_t, \qquad \sqrt{\frac{T}{[Tr]}} \longrightarrow \frac{1}{\sqrt{r}}.$$

Therefore, we obtain:

$$\sqrt{T}X_T(r) \longrightarrow N(0, r\sigma^2).$$

Moreover, we have the following results:

$$\frac{\sqrt{T}\,(X_T(r_2) - X_T(r_1))}{\sigma} \longrightarrow N(0, r_2 - r_1) = W(r_2) - W(r_1),$$

$$\frac{\sqrt{T}X_T(r)}{\sigma} \longrightarrow N(0, r) = W(r)$$

For example, consider the case of $r = 1$:

$$X_T(1) = \frac{1}{T} \sum_{t=1}^{T} \epsilon_t.$$

Then,

$$\frac{\sqrt{T} X_T(1)}{\sigma} = \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{T} \epsilon_t \longrightarrow N(0,1) = W(1).$$

(d) Consider $y_t = y_{t-1} + \epsilon_t$, $y_0 = 0$ and $\epsilon_t \sim N(0, \sigma^2)$.

$X_T(r)$ is defined as follows:

$$X_T(r) = \begin{cases} 0, & 0 \le r < \dfrac{1}{T}, \\ \dfrac{y_1}{T}, & \dfrac{1}{T} \le r < \dfrac{2}{T}, \\ \dfrac{y_2}{T}, & \dfrac{2}{T} \le r < \dfrac{3}{T}, \\ \vdots & \vdots \\ \dfrac{y_T}{T}, & r = 1. \end{cases}$$

To obtain $\int_0^1 X_T(r)\mathrm{d}r$, we compute a sum of rectangulars as follows:

$$\int_0^1 X_T(r)\mathrm{d}r = \frac{y_1}{T}\left(\frac{2}{T} - \frac{1}{T}\right) + \frac{y_2}{T}\left(\frac{3}{T} - \frac{2}{T}\right) + \cdots + \frac{y_{T-1}}{T}\left(1 - \frac{T-1}{T}\right)$$

$$= \frac{y_1}{T^2} + \frac{y_2}{T^2} + \cdots + \frac{y_{T-1}}{T^2} = \frac{1}{T^2}\sum_{t=1}^{T} y_t,$$

We have already known that $\dfrac{\sqrt{T}X_T(r)}{\sigma} \longrightarrow W(r)$.

Therefore,

$$\int_0^1 \frac{\sqrt{T}X_T(r)}{\sigma}\mathrm{d}r = \frac{1}{T^{3/2}\sigma}\sum_{t=1}^{T} y_t \longrightarrow \int_0^1 W(r)\mathrm{d}r.$$

(e) Define $S_T(r)$ as follows:

$$S_T(r) = \left(\sqrt{T}X_T(r)\right)^2 = \begin{cases} 0, & 0 \le r < \dfrac{1}{T}, \\ \dfrac{y_1^2}{T}, & \dfrac{1}{T} \le r < \dfrac{2}{T}, \\ \dfrac{y_2^2}{T}, & \dfrac{2}{T} \le r < \dfrac{3}{T}, \\ \quad\vdots & \quad\vdots \\ \dfrac{y_T^2}{T}, & r = 1. \end{cases}$$

To obtain $\int_0^1 S_T(r)\mathrm{d}r$, we compute a sum of rectangulars as follows:

$$\begin{aligned} \int_0^1 S_T(r)\mathrm{d}r &= \frac{y_1^2}{T}\left(\frac{2}{T} - \frac{1}{T}\right) + \frac{y_2^2}{T}\left(\frac{3}{T} - \frac{2}{T}\right) + \cdots + \frac{y_{T-1}^2}{T}\left(1 - \frac{T-1}{T}\right) \\ &= \frac{y_1^2}{T^2} + \frac{y_2^2}{T^2} + \cdots + \frac{y_{T-1}^2}{T^2} = \frac{1}{T^2}\sum_{t=1}^{T} y_t^2. \end{aligned}$$

We have already known that $\sqrt{T} X_T(r) \longrightarrow \sigma W(r)$.

From $S_T(r) \equiv \left(\sqrt{T} X_T(r) \right)^2$,

$$\frac{S_T(r)}{\sigma^2} = \left(\frac{\sqrt{T} X_T(r)}{\sigma} \right)^2 \longrightarrow W(r)^2,$$

which is called the continuous mapping theorem.

(*) **Continuous Mapping Theorem (連続写像定理):**

if $x_T \longrightarrow x$ (convergence in distribution) and $g(\cdot)$ is a continuous function,

then $g(x_T) \longrightarrow g(x)$ (convergence in distribution).

Threfore, we have the follwoing result:

$$\int_0^1 \frac{S_T(r)}{\sigma^2} \mathrm{d}r = \int_0^1 \left(\frac{\sqrt{T} X_T(r)}{\sigma} \right)^2 \mathrm{d}r = \frac{1}{T^2 \sigma^2} \sum_{t=1}^{T} y_t^2 \longrightarrow \int_0^1 W(r)^2 \mathrm{d}r.$$

Asymptotic Distribution of AR(1) Model:

$\boldsymbol{H_0 : y_t = y_{t-1} + \epsilon_t}$ **and** $\boldsymbol{H_1 : y_t = \phi_1 y_{t-1} + \epsilon_t}$ **for** $\boldsymbol{|\phi_1| < 1}$

OLSE of ϕ_1, denoted by $\hat{\phi}_1$, is given by:

$$\hat{\phi}_1 = \frac{\sum_{t=1}^T y_{t-1}y_t}{\sum_{t=1}^T y_{t-1}^2} = \phi_1 + \frac{\sum_{t=1}^T y_{t-1}\epsilon_t}{\sum_{t=1}^T y_{t-1}^2}$$

Using $\phi_1 = 1$ and some formulas shown above, we obtain:

$$T(\hat{\phi}_1 - 1) = \frac{T^{-1}\sum_{t=1}^T y_{t-1}\epsilon_t}{T^{-2}\sum_{t=1}^T y_{t-1}^2} \longrightarrow \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2 \mathrm{d}r}$$

Note that

$$T^{-1}\sum_{t=1}^T y_{t-1}\epsilon_t \longrightarrow \frac{1}{2}\sigma^2(W(1)^2 - 1) \text{ and } T^{-2}\sum_{t=1}^T y_{t-1}^2 \longrightarrow \sigma^2 \int_0^1 W(r)^2 \mathrm{d}r,$$

where $W(1)^2 = \chi^2(1)$.

We say that $\hat{\phi}_1$ is **super-consistent (超一致性)** or ***T*-consistent**.

Remember that when $|\phi_1| < 1$ we have $\sqrt{T}(\hat{\phi}_1 - \phi_1) \longrightarrow N(0, 1 - \phi_1^2)$, and in this case we say that $\hat{\phi}_1$ is **$\sqrt{T}$-consistent**.

Next, consider the conventional t test statistic:

$$t_T = \frac{\hat{\phi}_1 - 1}{s_\phi},$$

where

$$s_\phi = \left(\frac{s_T^2}{\sum_{t=1}^T y_{t-1}^2} \right)^{1/2} \qquad \text{and} \qquad s_T^2 = \frac{1}{T-1} \sum_{t=1}^T (y_t - \hat{\phi}_1 y_{t-1})^2.$$

The t test statistic, denoted by t_T, is represented as follows:

$$t_T = \frac{\hat{\phi}_1 - 1}{s_\phi} = \frac{T(\hat{\phi}_1 - 1)}{T s_\phi}$$

The denominator is:

$$T s_{\phi} = \left(\frac{s_T^2}{T^{-2} \sum_{t=1}^{T} y_{t-1}^2} \right)^{1/2} \longrightarrow \left(\frac{\sigma^2}{\sigma^2 \int_0^1 W(r)^2 \mathrm{d}r} \right)^{1/2} = \left(\int_0^1 W(r)^2 \mathrm{d}r \right)^{-1/2},$$

where $s_T^2 \longrightarrow \sigma^2$ is utilized.

Therefore, we have the following asymptotic distribution:

$$t_T = \frac{\hat{\phi}_1 - 1}{s_{\phi}} \longrightarrow \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2 \mathrm{d}r} \Bigg/ \Big(\int_0^1 W(r)^2 \mathrm{d}r \Big)^{-1/2} = \frac{\frac{1}{2}(W(1)^2 - 1)}{(\int_0^1 W(r)^2 \mathrm{d}r)^{1/2}}.$$

Therefore, the distribution of the t_T statistic shown above is different from the *t* distribution.

$\Longrightarrow$ Compare *t* distribution with (a) – (c).

$\Longrightarrow$ **Unit Root Test (単位根検定, or Dickey-Fuller (DF) Test)**

t Distribution

T	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
25	−2.49	−2.06	−1.71	−1.32	1.32	1.71	2.06	2.49
50	−2.40	−2.01	−1.68	−1.30	1.30	1.68	2.01	2.40
100	−2.36	−1.98	−1.66	−1.29	1.29	1.66	1.98	2.36
250	−2.34	−1.97	−1.65	−1.28	1.28	1.65	1.97	2.34
500	−2.33	−1.96	−1.65	−1.28	1.28	1.65	1.96	2.33
∞	−2.33	−1.96	−1.64	−1.28	1.28	1.64	1.96	2.33

(a) $\boldsymbol{H_0 : y_t = y_{t-1} + \epsilon_t}$

$\boldsymbol{H_1 : y_t = \phi_1 y_{t-1} + \epsilon_t}$ **for** $\boldsymbol{\phi_1 < 1}$

T	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
25	−2.66	−2.26	−1.95	−1.60	0.92	1.33	1.70	2.16
50	−2.62	−2.25	−1.95	−1.61	0.91	1.31	1.66	2.08
100	−2.60	−2.24	−1.95	−1.61	0.90	1.29	1.64	2.03
250	−2.58	−2.23	−1.95	−1.62	0.89	1.29	1.63	2.01
500	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00
∞	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00

Note that testing $H_0: \phi_1 = 1$ and $H_1: \phi_1 < 1$ in the model:

$$y_t = \phi_1 y_{t-1} + \epsilon_t,$$

is equivalent to testing $H_0: \rho = 0$ and $H_1: \rho < 0$ in the model:

$$\Delta y_t = \rho y_{t-1} + \epsilon_t,$$

where $\rho = \phi_1 - 1$.

(b) $\boldsymbol{H_0: y_t = y_{t-1} + \epsilon_t}$

$\boldsymbol{H_1: y_t = \alpha_0 + \phi_1 y_{t-1} + \epsilon_t}$ **for** $\boldsymbol{\phi_1 < 1}$

T	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
25	−3.75	−3.33	−3.00	−2.63	−0.37	0.00	0.34	0.72
50	−3.58	−3.22	−2.93	−2.60	−0.40	−0.03	0.29	0.66
100	−3.51	−3.17	−2.89	−2.58	−0.42	−0.05	0.26	0.63
250	−3.46	−3.14	−2.88	−2.57	−0.42	−0.06	0.24	0.62
500	−3.44	−3.13	−2.87	−2.57	−0.43	−0.07	0.24	0.61
∞	−3.43	−3.12	−2.86	−2.57	−0.44	−0.07	0.23	0.60

(c) $\boldsymbol{H_0: \ y_t = \alpha_0 + y_{t-1} + \epsilon_t}$

$\boldsymbol{H_1: \ y_t = \alpha_0 + \alpha_1 t + \phi_1 y_{t-1} + \epsilon_t}$ **for** $\boldsymbol{\phi_1 < 1}$

T	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
25	−4.38	−3.95	−3.60	−3.24	−1.14	−0.80	−0.50	−0.15
50	−4.15	−3.80	−3.50	−3.18	−1.19	−0.87	−0.58	−0.24
100	−4.04	−3.73	−3.45	−3.15	−1.22	−0.90	−0.62	−0.28
250	−3.99	−3.69	−3.43	−3.13	−1.23	−0.92	−0.64	−0.31
500	−3.98	−3.68	−3.42	−3.13	−1.24	−0.93	−0.65	−0.32
∞	−3.96	−3.66	−3.41	−3.12	−1.25	−0.94	−0.66	−0.33

3.2 Serially Correlated Errors

Consider the case where the error term is serially correlated.

3.2.1 Augmented Dickey-Fuller (ADF) Test

Consider the following AR(p) model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t, \qquad \epsilon_t \sim \text{iid}(0, \sigma^2),$$

which is rewritten as: $\qquad \phi(L) y_t = \epsilon_t.$

When the above model has a unit root, we have $\phi(1) = 0$, i.e., $\phi_1 + \phi_2 + \cdots + \phi_p = 1$.

The above AR(p) model is written as:

$$y_t = \rho y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where $\rho = \phi_1 + \phi_2 + \cdots + \phi_p$ and $\delta_j = -(\phi_{j+1} + \phi_{j+2} + \cdots + \phi_p)$.

The null and alternative hypotheses are:

$$H_0:\ \rho = 1 \text{ (Unit root)},$$

$$H_1:\ \rho < 1 \text{ (Stationary)}.$$

That is, the above test is equivalent to the following:

$$\Delta y_t = \alpha y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where $\alpha = \rho - 1$. The null and alternative hypotheses are:

$$H_0:\ \alpha = 0 \text{ (Unit root)},$$

$$H_1:\ \alpha < 0 \text{ (Stationary)}.$$

Use the t test, where we have the same asymptotic distributions.

Therefore, we can utilize the same tables as before.

Use AIC or SBIC to choose p.

Use $N(0,1)$ to test $H_0:\ \delta_j = 0$ against $H_1:\ \delta_j \neq 0$ for $j = 1, 2, \cdots, p-1$.

Reference

Kurozumi (2008) "Economic Time Series Analysis and Unit Root Tests: Development and Perspective," *Japan Statistical Society*, Vol.38, Series J, No.1, pp.39 – 57.

Download the above paper from:

`http://ci.nii.ac.jp/vol_issue/nels/AA11989749/ISS0000426576_ja.html`

Example of ADF Test

```
. gen time=_n

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

. gen sexpend=expend-l12.expend
(12 missing values generated)

. corrgram sexpend

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1         0.7177   0.7184   261.14  0.0000          |-----             |-----
2         0.7036   0.3895    512.6  0.0000          |-----             |---
3         0.7031   0.2817   764.23  0.0000          |-----             |--
4         0.6366   0.0456   970.94  0.0000          |-----             |
5         0.6413   0.1116   1181.1  0.0000          |-----             |
6         0.6267   0.0815   1382.2  0.0000          |-----             |
7         0.6208   0.0972     1580  0.0000          |----              |
8         0.6384   0.1286   1789.5  0.0000          |-----             |-
9         0.5926  -0.0205   1970.5  0.0000          |----              |
10        0.5847  -0.0014   2146.9  0.0000          |----              |
11        0.5658  -0.0185   2312.6  0.0000          |----              |
12        0.4529  -0.2570   2418.9  0.0000          |---             --|
13        0.5601   0.2318   2581.8  0.0000          |----              |-
14        0.5393   0.1095   2733.2  0.0000          |----              |
15        0.5277   0.0850   2878.4  0.0000          |----              |
```

```
. varsoc d.sexpend, exo(l.sexpend) maxlag(25)

   Selection-order criteria
   Sample:  39 - 516                                  Number of obs      =       478
  +---------------------------------------------------------------------------+
  |lag |    LL      LR      df    p      FPE       AIC      HQIC      SBIC    |
  |----+----------------------------------------------------------------------|
  |  0 |  -4917.7                      5.1e+07   20.5845   20.5914   20.6019  |
  |  1 | -4878.69  78.013    1  0.000  4.3e+07   20.4255   20.4358   20.4516  |
  |  2 | -4858.95  39.481    1  0.000  4.0e+07   20.3471   20.3608    20.382  |
  |  3 | -4858.46  .97673    1  0.323  4.0e+07   20.3492   20.3664   20.3928  |
  |  4 | -4855.44  6.0461    1  0.014  4.0e+07   20.3407   20.3613   20.3931  |
  |  5 | -4853.84  3.1904    1  0.074  4.0e+07   20.3383   20.3623   20.3993  |
  |  6 | -4851.58  4.5304    1  0.033  4.0e+07    20.333   20.3604   20.4027  |
  |  7 | -4847.61   7.942    1  0.005  3.9e+07   20.3205   20.3514    20.399  |
  |  8 | -4847.51  .20154    1  0.653  3.9e+07   20.3243   20.3586   20.4115  |
  |  9 | -4847.51  .00096    1  0.975  3.9e+07   20.3285   20.3662   20.4244  |
  | 10 | -4847.43  .16024    1  0.689  4.0e+07   20.3323   20.3735    20.437  |
  | 11 | -4831.38  32.094    1  0.000  3.7e+07   20.2694   20.3139   20.3828  |
  | 12 | -4818.46  25.834    1  0.000  3.5e+07   20.2195   20.2675   20.3416* |
  | 13 | -4815.64  5.6341    1  0.018  3.5e+07   20.2119   20.2633   20.3427  |
  | 14 | -4813.98   3.321    1  0.068  3.5e+07   20.2091    20.264   20.3487  |
  | 15 | -4813.38  1.2007    1  0.273  3.5e+07   20.2108   20.2691   20.3591  |
  | 16 | -4810.57  5.6184    1  0.018  3.5e+07   20.2032    20.265   20.3603  |
  | 17 |  -4808.7  3.7539    1  0.053  3.5e+07   20.1996   20.2647   20.3653  |
  | 18 | -4806.12  5.1557    1  0.023  3.4e+07    20.193   20.2616   20.3674  |
  | 19 |  -4804.6  3.0319    1  0.082  3.4e+07   20.1908   20.2628    20.374  |
  | 20 |  -4804.6  2.7e-05   1  0.996  3.5e+07    20.195   20.2704   20.3869  |
  | 21 | -4797.33  14.542    1  0.000  3.4e+07   20.1688   20.2476   20.3694  |
  | 22 |  -4794.2  6.2571*   1  0.012  3.3e+07*  20.1598*  20.2422*  20.3692  |
  | 23 | -4793.42  1.5626    1  0.211  3.3e+07   20.1608   20.2465   20.3788  |
  | 24 | -4792.85  1.1533    1  0.283  3.3e+07   20.1625   20.2517   20.3893  |
```

```
  | 25 | -4792.78  .13518    1  0.713  3.4e+07   20.1664    20.259    20.402  |
  +---------------------------------------------------------------------------+
   Endogenous:  D.sexpend
    Exogenous:  L.sexpend  _cons


. dfuller sexpend, lags(23)

Augmented Dickey-Fuller test for unit root         Number of obs   =       480

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.754            -3.442            -2.871            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.4033

. dfuller sexpend, lags(13)

Augmented Dickey-Fuller test for unit root         Number of obs   =       490

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.129            -3.441            -2.870            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.2329

.
```

3.3 Cointegration (共和分)

1. For a scalar y_t, when $\Delta y_t = y_t - y_{t-1}$ is a white noise (i.e., iid), we write $\Delta y_t \sim I(1)$.

2. **Definition of Cointegration:**

   Suppose that each series in a $g \times 1$ vector y_t is $I(1)$, i.e., each series has unit root, and that a linear combination of each series (i.e, $a'y_t$ for a nonzero vector a) is $I(0)$, i.e., stationary.

   Then, we say that y_t has a cointegration.

3. **Example:**

   Suppose that $y_t = (y_{1,t},\ y_{2,t})'$ is the following vector autoregressive process:

$$y_{1,t} = \phi_1 y_{2,t} + \epsilon_{1,t},$$

$$y_{2,t} = y_{2,t-1} + \epsilon_{2,t}.$$

   Then,

$$\Delta y_{1,t} = \phi_1 \epsilon_{2,t} + \epsilon_{1,t} - \epsilon_{1,t-1}, \quad \text{(MA(1) process)},$$

$$\Delta y_{2,t} = \epsilon_{2,t},$$

where both $y_{1,t}$ and $y_{2,t}$ are $I(1)$ processes.

The linear combination $y_{1,t} - \phi_1 y_{2,t}$ is $I(0)$.

In this case, we say that $y_t = (y_{1,t},\ y_{2,t})'$ is cointegrated with $a = (1,\ -\phi_1)$.

$a = (1,\ -\phi_1)$ is called the cointegrating vector, which is not unique.

Therefore, the first element of a is set to be one.

4. Suppose that $y_t \sim I(1)$ and $x_t \sim I(1)$.

   For the regression model $y_t = x_t\beta + u_t$, OLS does not work well if we do not have the β which satisfies $u_t \sim I(0)$.

   $\Longrightarrow$ **Spurious regression (見せかけの回帰)**

5. Suppose that $y_t \sim I(1)$, y_t is a $g \times 1$ vector and $y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}$.

   $y_{2,t}$ is a $k \times 1$ vector, where $k = g - 1$.

   Consider the following regression model:

$$y_{1,t} = \alpha + \gamma' y_{2,t} + u_t, \qquad t = 1, 2, \cdots, T.$$

OLSE is given by:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \sum y'_{2,t} \\ \sum y_{2,t} & \sum y_{2,t} y'_{2,t} \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{1,t} \\ \sum y_{1,t} y_{2,t} \end{pmatrix}.$$

Next, consider testing the null hypothesis $H_0: \ R\gamma = r$, where R is a $m \times k$ matrix ($m \leq k$) and r is a $m \times 1$ vector.

The F statistic, denoted by F_T, is given by:

$$F_T = \frac{1}{m}(R\hat{\gamma} - r)' \left(s_T^2 \begin{pmatrix} 0 & R \end{pmatrix} \begin{pmatrix} T & \sum y'_{2,t} \\ \sum y_{2,t} & \sum y_{2,t} y'_{2,t} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ R' \end{pmatrix} \right)^{-1} (R\hat{\gamma} - r),$$

where

$$s_T^2 = \frac{1}{T-g} \sum_{t=1}^{T} (y_{1,t} - \hat{\alpha} - \hat{\gamma}' y_{2,t})^2.$$

When we have the γ such that $y_{1,t} - \gamma y_{2,t}$ is stationary, OLSE of γ, i.e., $\hat{\gamma}$, is not statistically equal to zero.

When the sample size T is large enough, H_0 is rejected by the F test.

6. Phillips, P.C.B. (1986) "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, Vol.33, pp.95 – 131.

Consider a $g \times 1$ vector y_t whose first difference is described by:

$$\Delta y_t = \Psi(L)\epsilon_t = \sum_{s=0}^{\infty} \Psi_s \epsilon_{t-s},$$

for ϵ_t an i.i.d. $g \times 1$ vector with mean zero , variance $\mathrm{E}(\epsilon_t \epsilon_t') = PP'$, and finite fourth moments and where $\{s\Psi_s\}_{s=0}^{\infty}$ is absolutely summable.

Let $k = g - 1$ and $\Lambda = \Psi(1)P$.

Partition y_t as $y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}$ and $\Lambda\Lambda'$ as $\Lambda\Lambda' = \begin{pmatrix} \Sigma_{11} & \Sigma_{21}' \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, where $y_{1,t}$ and Σ_{11} are scalars, $y_{2,t}$ and Σ_{21} are $k \times 1$ vectors, and Σ_{22} is a $k \times k$ matrix.

Suppose that $\Lambda\Lambda'$ is nonsingular,and define $\sigma_1^{*2} = \Sigma_{11} - \Sigma_{21}'\Sigma_{22}^{-1}\Sigma_{21}$.

Let L_{22} denote the Cholesky factor of Σ_{22}^{-1}, i.e., L_{22} is the lower triangular matrix satisfying $\Sigma_{22}^{-1} = L_{22}L_{22}'$.

Then, (a) – (c) hold.

(a) OLSEs of α and γ in the regression model $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$, denoted by $\hat{\alpha}_T$ and $\hat{\gamma}_T$, are characterized by:

$$\begin{pmatrix} T^{-1/2}\hat{\alpha}_T \\ \hat{\gamma}_T - \Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix} \longrightarrow \begin{pmatrix} \sigma_1^* h_1 \\ \sigma_1^* L_{22} h_2 \end{pmatrix},$$

where $\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 1 & \int_0^1 W_2^*(r)'\mathrm{d}r \\ \int_0^1 W_2^*(r)\mathrm{d}r & \int_0^1 W_2^*(r)W_2^*(r)'\mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 W_1^*(r)\mathrm{d}r \\ \int_0^1 W_2^*(r)W_1^*(r)\mathrm{d}r \end{pmatrix}.$

$W_1^*(r)$ and $W_2^*(r)$ denote scalar and g-dimensional standard Brownian motions, and $W_1^*(r)$ is independent of $W_2^*(r)$.

(b) The sum of squared residuals, denoted by $\mathrm{RSS}_T = \sum_{t=1}^T \hat{u}_t^2$, satisfies

$$T^{-2}\mathrm{RSS}_T \longrightarrow \sigma_1^{*2}H,$$

where $H = \int_0^1 (W_1^*(r))^2\mathrm{d}r - \left(\begin{pmatrix} \int_0^1 W_1^*(r)\mathrm{d}r \\ \int_0^1 W_2^*(r)W_1^*(r)\mathrm{d}r \end{pmatrix}' \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}\right)^{-1}.$

(c) The F_T test satisfies:

$$\begin{aligned} T^{-1}F_T \longrightarrow\ & \frac{1}{m}(\sigma_1^* R^* h_2 - r^*)' \\ & \times\left(\sigma_1^{*2}H \begin{pmatrix} 0 & R^* \end{pmatrix} \begin{pmatrix} 1 & \int_0^1 W_2^*(r)'\mathrm{d}r \\ \int_0^1 W_2^*(r)\mathrm{d}r & \int_0^1 W_2^*(r)W_2^*(r)'\mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} 0 & R^* \end{pmatrix}'\right)^{-1} \\ & \times(\sigma_1^* R^* h_2 - r^*), \end{aligned}$$

where $R^* = RL_{22}$ and $r^* = r - R\Sigma_{22}^{-1}\Sigma_{21}$.

Summary:

(a) indicates that OLSE $\hat{\gamma}_T$ is not consistent.

(b) indicates that $s_T^2 = \dfrac{1}{T-g}\sum_{t=1}^{T}\hat{u}_t^2$ diverges.

(c) indicates that F_T diverges.

$\Longrightarrow$ **Spurious regression (見せかけの回帰)**

7. **Resolution for Spurious Regression:**

Suppose that $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ is a spurious regression.

(1) Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + \phi y_{1,t-1} + \delta y_{2,t-1} + u_t$.

Then, $\hat{\gamma}_T$ is $\sqrt{T}$-consistent, and the t test statistic goes to the standard normal distribution under $H_0:\ \gamma = 0$.

(2) Estimate $\Delta y_{1,t} = \alpha + \gamma' \Delta y_{2,t} + u_t$. Then, $\hat{\alpha}_T$ and $\hat{\beta}_T$ are $\sqrt{T}$-consistent, and the t test and F test make sense.

(3) Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ by the Cochrane-Orcutt method, assuming that u_t is the first-order serially correlated error.

Usually, choose (2).

However, there are two exceptions.

(i) The true value of ϕ is not one, i.e., less than one.

(ii) $y_{1,t}$ and $y_{2,t}$ are the cointegrated processes.

In these two cases, taking the first difference leads to the misspecified regression.

8. **Cointegrating Vector:**

Suppose that each element of y_t is $I(1)$ and that $a' y_t$ is $I(0)$.

a is called a **cointegrating vector (共和分ベクトル)**, which is not unique.

Set $z_t = a'y_t$, where z_t is scalar, and a and y_t are $g \times 1$ vectors.

For $z_t \sim I(0)$ (i.e., stationary),

$$T^{-1}\sum_{t=1}^{T} z_t^2 = T^{-1}\sum_{t=1}^{T}(a'y_t)^2 \longrightarrow \mathrm{E}(z_t^2).$$

For $z_t \sim I(1)$ (i.e., nonstationary, i.e., a is not a cointegrating vector),

$$T^{-2}\sum_{t=1}^{T}(a'y_t)^2 \longrightarrow \lambda^2 \int_0^1 (W(r))^2\,\mathrm{d}r,$$

where $W(r)$ denotes a standard Brownian motion and λ^2 indicates variance of $(1-L)z_t$.

If a is not a cointegrating vector, $T^{-1}\sum_{t=1}^{T} z_t^2$ diverges.

$\Longrightarrow$ We can obtain a consistent estimate of a cointegrating vector by minimizing $\sum_{t=1}^{T} z_t^2$ with respect to a, where a normalization condition on a has to be imposed.

The estimator of the a including the normalization condition is super-consistent (T-consistent).

● Stock, J.H. (1987) "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, Vol.55, pp.1035 – 1056.

Proposition:

Let $y_{1,t}$ be a scalar, $y_{2,t}$ be a $k \times 1$ vector, and $(y_{1,t}, y'_{2,t})'$ be a $g \times 1$ vector, where $g = k + 1$.

Consider the following model:

$$y_{1,t} = \alpha + \gamma' y_{2,t} + z_t^*, \qquad \begin{pmatrix} z_t^* \\ u_{2,t} \end{pmatrix} = \Psi^*(L)\epsilon_t,$$
$$\Delta y_{2,t} = u_{2,t},$$

ϵ_t is a $g \times 1$ i.i.d. vector with $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{E}(\epsilon_t \epsilon_t') = PP'$.

OLSE is given by: $\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \sum y'_{2,t} \\ \sum y_{2,t} & \sum y_{2,t} y'_{2,t} \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{1,t} \\ \sum y_{1,t} y_{2,t} \end{pmatrix}.$

Define λ_1^*, which is a $g \times 1$ vector, and Λ_2^*, which is a $k \times g$ matrix, as follows:

$$\Psi^*(1)\, P = \begin{pmatrix} \lambda_1^{*\prime} \\ \Lambda_2^* \end{pmatrix}.$$

Then, we have the following results:

$$\begin{pmatrix} T^{1/2}(\hat{\alpha}-\alpha) \\ T(\hat{\gamma}-\gamma) \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \left(\Lambda_2^* \int W(r)\mathrm{d}r\right)' \\ \Lambda_2^* \int W(r)\mathrm{d}r & \Lambda_2^* \left(\int (W(r))\,(W(r))'\,\mathrm{d}r \right) \Lambda_2^{*\prime} \end{pmatrix}^{-1} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix},$$

where $\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \lambda_1^{*\prime} W(1) \\ \Lambda_2^* \left(\int W(r)\,(\mathrm{d}W(r))' \right) \lambda_1^* + \sum_{\tau=0}^{\infty} \mathrm{E}(u_{2,t} z_{t+\tau}^*) \end{pmatrix}.$

$W(r)$ denotes a g-dimensional standard Brownian motion.

1) OLSE of the cointegrating vector is consistent even though u_t is serially correlated.

2) The consistency of OLSE implies that $T^{-1}\sum \hat{u}_t^2 \longrightarrow \sigma^2$.

3) Because $T^{-1}\sum(y_{1,t}-\bar{y}_1)^2$ goes to infinity, a coefficient of determination, R^2, goes to one.

3.4 Testing Cointegration

3.4.1 Engle-Granger Test

$y_t \sim I(1)$

$y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$

- $u_t \sim I(0) \implies$ Cointegration
- $u_t \sim I(1) \implies$ Spurious Regression

Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ by OLS, and obtain $\hat{u}_t$.

Estimate $\hat{u}_t = \rho \hat{u}_{t-1} + \delta_1 \Delta \hat{u}_{t-1} + \delta_2 \Delta \hat{u}_{t-2} + \cdots + \delta_{p-1} \Delta \hat{u}_{t-p+1} + e_t$ by OLS.

ADF Test:

- $H_0: \ \rho = 1$ (Sprious Regression)
- $H_1: \ \rho < 1$ (Cointegration)

$\implies$ **Engle-Granger Test**

For example, see Engle and Granger (1987), Phillips and Ouliaris (1990) and Hansen (1992).

Asymmptotic Distribution of Residual-Based ADF Test for Cointegration

# of Refressors,	(a) Regressors have no drift				(b) Some regressors have drift			
excluding constant	1%	2.5%	5%	10%	1%	2.5%	5%	10%
1	−3.96	−3.64	−3.37	−3.07	−3.96	−3.67	−3.41	−3.13
2	−4.31	−4.02	−3.77	−3.45	−4.36	−4.07	−3.80	−3.52
3	−4.73	−4.37	−4.11	−3.83	−4.65	−4.39	−4.16	−3.84
4	−5.07	−4.71	−4.45	−4.16	−5.04	−4.77	−4.49	−4.20
5	−5.28	−4.98	−4.71	−4.43	−5.36	−5.02	−4.74	−4.46

J.D. Hamilton (1994), *Time Series Analysis*, p.766.

3.4.2 Error Correction Representation

VAR(p) model:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

where y_t, α and ϵ_t indicate $g \times 1$ vectors for $t = 1, 2, \cdots, T$, and ϕ_s is a $g \times g$ matrix for $s = 1, 2, \cdots, p$.

Rewrite:

$$y_t = \alpha + \rho y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where

$$\rho = \phi_1 + \phi_2 + \cdots + \phi_p,$$

$$\delta_s = -(\phi_{s+1} + \delta_{s+2} + \cdots + \phi_p), \qquad \text{for } s = 1, 2, \cdots, p-1.$$

Again, rewrite:

$$\Delta y_t = \alpha + \delta_0 y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where

$$\delta_0 = \rho - I_g = -\phi(1),$$

for $\phi(L) = I_g - \delta_1 L - \delta_2 L^2 - \cdots - \delta_p L^p$.

If y_t has h cointegrating relations, we have the following error correction representation:

$$\Delta y_t = \alpha - BA' y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where $A' y_{t-1}$ is a stationary $h \times 1$ vector (i.e., h I(0) processes), and B and A are $g \times h$ matrices.

Note that $\phi(1) = BA'$ for $\phi(L) = I_g - \delta_1 L - \delta_2 L^2 - \cdots - \delta_p L^p$.

Each row of A' denotes the cointegrating vector, i.e., A' consists of h cointegrating vectors.

Suppose that $\epsilon_t \sim N(0, \Sigma)$. The log-likelihood function is:

$$\begin{aligned}
&\log l(\alpha, \delta_1, \cdots, \delta_{p-1}, B|A) \\
&\quad = -\frac{Tg}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma| \\
&\qquad -\frac{1}{2}\sum_{t=1}^{T}(\Delta y_t - \alpha + BA' y_{t-1} - \delta_1 \Delta y_{t-1} - \cdots - \delta_{p-1}\Delta y_{t-p+1})'\Sigma^{-1} \\
&\qquad\qquad \times(\Delta y_t - \alpha + BA' y_{t-1} - \delta_1 \Delta y_{t-1} - \cdots - \delta_{p-1}\Delta y_{t-p+1})
\end{aligned}$$

Given A and h, maximize $\log l$ with respect to $\alpha, \delta_1, \cdots, \delta_{p-1}, B$.

Then, given h, how do we estimate A? $\Longrightarrow$ Johansen (1988, 1991)

(*) Canonical Correlatoion (正準相関)

$x' = (x_1, x_2, \cdots, x_n)$ and $y' = (y_1, y_2, \cdots, y_m)$, where $n \leq m$.

$$u = a'x = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n, \qquad v = b'y = b_1 y_1 + b_2 y_2 + \cdots + b_m y_m,$$

where $\mathrm{V}(u) = \mathrm{V}(v) = 1$ and $\mathrm{E}(x) = \mathrm{E}(y) = 0$ for simplicity.

Define:

$$\mathrm{V}(x) = \Sigma_{xx}, \qquad \mathrm{E}(xy') = \Sigma_{xy}, \qquad \mathrm{V}(y) = \Sigma_{yy}, \qquad \mathrm{E}(yx') = \Sigma_{yx} = \Sigma'_{xy}.$$

The correlation coefficient between u and v, denoted by ρ, is:

$$\rho = \frac{\mathrm{Cov}(u, v)}{\sqrt{\mathrm{V}(u)}\sqrt{\mathrm{V}(v)}} = a'\Sigma_{xy}b,$$

where $\mathrm{V}(u) = a'\Sigma_{xx}a = 1$ and $\mathrm{V}(v) = b'\Sigma_{yy}b = 1$.

Maximize $\rho = a'\Sigma_{xy}b$ subject to $a'\Sigma_{xx}a = 1$ and $b'\Sigma_{yy}b = 1$.

The Lagrangian is:

$$L = a'\Sigma_{xy}b - \frac{1}{2}\lambda(a'\Sigma_{xx}a - 1) - \frac{1}{2}\mu(b'\Sigma_{yy}b - 1).$$

Take a derivative with respect to a and b.

$$\frac{\partial L}{\partial a} = \Sigma_{xy}b - \lambda\Sigma_{xx}a = 0, \qquad \frac{\partial L}{\partial b} = \Sigma'_{xy}a - \mu\Sigma_{yy}b = 0.$$

Using $a'\Sigma_{xx}a = 1$ and $b'\Sigma_{yy}b = 1$, we obtain:

$$\lambda = \mu = a'\Sigma_{xy}b.$$

From the first equation, we obtain:

$$a = \frac{1}{\lambda}\Sigma_{xx}^{-1}\Sigma_{xy}b,$$

which is substituted into the second equation as follows:

$$\frac{1}{\lambda}\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}b - \lambda\Sigma_{yy}b = 0, \implies (\Sigma_{yy}^{-1}\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy} - \lambda^2 I_m)b = 0, \implies |\Sigma_{yy}^{-1}\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy} - \lambda^2 I_m| = 0.$$

The solution of λ^2 is given by the maximum eigen value of $\Sigma_{yy}^{-1}\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}$, and b is the corresponding eigen vector.

Back to the Cointegration:

Estimate the following two regressions:

$$\Delta y_t = b_{1,0} + b_{1,1}\Delta y_{t-1} + b_{1,2}\Delta y_{t-2} + \cdots + b_{1,p-1}\Delta y_{t-p+1} + u_{1,t}$$

$$y_{t-1} = b_{2,0} + b_{2,1}\Delta y_{t-1} + b_{2,2}\Delta y_{t-2} + \cdots + b_{2,p-1}\Delta y_{t-p+1} + u_{2,t}$$

Obtain $\hat{u}_{i,t}$ for $i = 1, 2$ and $t = 1, 2, \cdots, T$, and compute as follow:

$$\hat{\Sigma}_{11} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{1,t}\hat{u}'_{1,t}, \qquad \hat{\Sigma}_{22} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{2,t}\hat{u}'_{2,t}, \qquad \hat{\Sigma}_{12} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{1,t}\hat{u}'_{2,t}, \qquad \hat{\Sigma}_{21} = \hat{\Sigma}'_{12}.$$

From $\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$, compute h biggest eigenvalues, denoted by $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\cdots$, $\hat{\lambda}_h$, and the corresponding eigen vectors, denoted by $\hat{a}_1$, $\hat{a}_2$, $\cdots$, $\hat{a}_h$, where $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_h$,

The estimate of A, $\hat{A}$, is given by $\hat{A} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_h)$.

How do we obtain h?

3.5 Testing the Number of Cointegrating Vectors

Trace Test (トレース検定):

$H_0: \lambda_{h+1} = 0$ and $H_1: \lambda_h > 0.$

$$2(\log l_1 - \log l_0) = -T \sum_{i=h+1}^{g} \log(1 - \hat{\lambda}_i) \longrightarrow \mathrm{tr}(Q),$$

where $Q = \left(\int_0^1 W(r)\mathrm{d}W(r)'\right)' \left(\int_0^1 W(r)W(r)'\mathrm{d}r\right)^{-1} \left(\int_0^1 W(r)\mathrm{d}W(r)'\right).$

Trace Test for # of Cointegrating Relations

# of Random	(a) Regressors have no drift				(b) Some regressors have drift			
Walks $(g-h)$	1%	2.5%	5%	10%	1%	2.5%	5%	10%
1	11.576	9.658	8.083	6.691	6.936	5.332	3.962	2.816
2	21.962	19.611	17.844	15.583	19.310	17.299	15.197	13.338
3	37.291	34.062	31.256	28.436	35.397	32.313	29.509	26.791
4	55.551	51.801	48.419	45.248	53.792	50.424	47.181	43.964
5	77.911	73.031	69.977	65.956	76.955	72.140	68.905	65.063

J.D. Hamilton (1994), *Time Series Analysis*, p.767.

Largest Eigenvalue Test (最大固有値検定):

$H_0: \lambda_{h+1} = 0$ and $H_1: \lambda_h > 0.$

$$2(\log l_1 - \log l_0) = -T \log(1 - \hat{\lambda}_{h+1}) \longrightarrow \text{maxmum eigen value of } Q,$$

Maximum Eigenvalue Test for # of Cointegrating Relations

# of Random Walks ($g - h$)	(a) Regressors have no drift 1%	2.5%	5%	10%	(b) Some regressors have drift 1%	2.5%	5%	10%
1	11.576	9.658	8.083	6.691	6.936	5.332	3.962	2.816
2	18.782	16.403	14.595	12.783	17.936	15.810	14.036	12.099
3	26.154	23.362	21.279	18.959	25.521	23.002	20.778	18.697
4	32.616	29.599	27.341	24.917	31.943	29.335	27.169	24.712
5	38.858	35.700	33.262	30.818	38.341	35.546	33.178	30.774

J.D. Hamilton (1994), *Time Series Analysis*, p.768.