計量モデル分析I

グローバル計量モデル分析

Thr., 8:50-10:20

Room # 4 (法経講義棟)

- The prerequisite of this class is **Basic Statistics (統計基礎)** (by Prof. Oya, Tue., 16:20-17:50, this semester) and **Econometrics (エコノメトリックス)** (undergraduate level, next semester, 『計量経済学』山本 拓 著，新世社).
- The class of **Introductory Econometrics (計量経済学基礎)** (by Prof. Takeuchi, Mon., 16:20-17:50, this semester) should be registered.

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』

Statistics Test (統計検定) on June 22 (Sun.)

- **Exams：** Level 2 (2 級) – Level 4 (4 級)

  Note that Level 4 is Junior high school level,

  Level 3 is High school level, and

  Level 2 is the 1st or 2nd year statistics in undergraduate school.

  See **http://www.toukei-kentei.jp/index.html** in more detail.

- **Qualification for Exam (受験資格)：**

  Undergraduate and Graduate Students in Osaka University

- **Application Period (受験申込期間)：** April 14 (Mon.) — May 14 (Wed.)

- **Application Fee (受験料)：** Free

受験料は，平成 24 年度に採択された文部科学省の大学間連携共同推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」から支払われる。

連携校：　東京大学，**大阪大学**，総合研究大学院大学，青山学院大学（代表校），多摩大学，　立教大学，早稲田大学，同志社大学

ちなみに、連携大学以外の人の受験料は,

統計検定 2 級	10:30～12:00	5,000 円
統計検定 3 級	13:30～14:30	4,000 円
統計検定 4 級	10:30～11:30	3,000 円

となる。

- **Exam Date (試験日)**：　June 22 (Sun.)
- **Exam Place (場所)**：　法経講義棟 # 1, 2, 4

1　最小二乗法について

経済理論に基づいた線型モデルの係数の値をデータから求める時に用いられる手法 $\Longrightarrow$ 最小二乗法

1.1　最小二乗法と回帰直線

$(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ のように n 組のデータがあり，X_i と Y_i との間に以下の線型関係を想定する。

$$Y_i = \alpha + \beta X_i,$$

X_i は説明変数，Y_i は被説明変数，α, β はパラメータとそれぞれ呼ばれる。

上の式は回帰モデル (または，回帰式) と呼ばれる。目的は，切片 α と傾き β をデータ $\{(X_i, Y_i), i = 1, 2, \cdots, n\}$ から推定すること，

データについて：

1. タイム・シリーズ (時系列)・データ： i が時間を表す (第 i 期)。

2. クロス・セクション (横断面)・データ： i が個人や企業を表す (第 i 番目の家計，第 i 番目の企業)。

1.2 切片 α と傾き β の推定

次のような関数 $S(\alpha,\beta)$ を定義する。

$$S(\alpha,\beta) = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

このとき,

$$\min_{\alpha,\beta} S(\alpha,\beta)$$

となるような α,β を求める (最小自乗法)。このときの解を $\widehat{\alpha},\widehat{\beta}$ とする。

最小化のためには,

$$\frac{\partial S(\alpha,\beta)}{\partial \alpha} = 0$$

$$\frac{\partial S(\alpha,\beta)}{\partial \beta} = 0$$

を満たす α,β が $\widehat{\alpha},\widehat{\beta}$ となる。　　すなわち，$\widehat{\alpha},\widehat{\beta}$ は,

$$\sum_{i=1}^{n}(Y_i - \widehat{\alpha} - \widehat{\beta}X_i) = 0, \quad (1)$$

$$\sum_{i=1}^{n} X_i(Y_i - \widehat{\alpha} - \widehat{\beta}X_i) = 0, \quad (2)$$

を満たす。　　さらに,

$$\sum_{i=1}^{n} Y_i = n\widehat{\alpha} + \widehat{\beta}\sum_{i=1}^{n} X_i, \quad (3)$$

$$\sum_{i=1}^{n} X_iY_i = \widehat{\alpha}\sum_{i=1}^{n} X_i + \widehat{\beta}\sum_{i=1}^{n} X_i^2,$$

行列表示によって，

$$\begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix},$$

逆行列の公式：

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$\widehat{\alpha}, \widehat{\beta}$ について，まとめて，

$$\begin{aligned} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} &= \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} \\ &= \frac{1}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \begin{pmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} \end{aligned}$$

さらに，$\widehat{\beta}$ について解くと，

$$\widehat{\beta} = \frac{n\sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

$$= \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

連立方程式の (3) 式から，

$$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{X}$$

となる。ただし，

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

とする。

数値例：　以下の数値例を使って，回帰式 $Y_i = \alpha + \beta X_i$ の α，β の推定値 $\widehat{\alpha}$，$\widehat{\beta}$ を求める。

i	Y_i	X_i
1	6	10
2	9	12
3	10	14
4	10	16

$\widehat{\alpha}$，$\widehat{\beta}$ を求めるための公式は

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

$$\widehat{\alpha} = \overline{Y} - \widehat{\beta X}$$

なので，必要なものは $\overline{X}$，$\overline{Y}$，$\sum_{i=1}^{n} X_i^2$，$\sum_{i=1}^{n} X_i Y_i$ である。

i	Y_i	X_i	X_iY_i	X_i^2
1	6	10	60	100
2	9	12	108	144
3	10	14	140	196
4	10	16	160	256
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$
	35	52	468	696
平均	$\overline{Y}$	$\overline{X}$		
	8.75	13		

よって,

$$\widehat{\beta} = \frac{468 - 4 \times 13 \times 8.75}{696 - 4 \times 13^2} = \frac{13}{20} = 0.65$$

$$\widehat{\alpha} = 8.75 - 0.65 \times 13 = 0.3$$

となる。

注意事項：

1. α, β は真の値で未知

2. $\widehat{\alpha}, \widehat{\beta}$ は α, β の推定値でデータから計算される

回帰直線は

$$\widehat{Y_i} = \widehat{\alpha} + \widehat{\beta} X_i,$$

として与えられる。

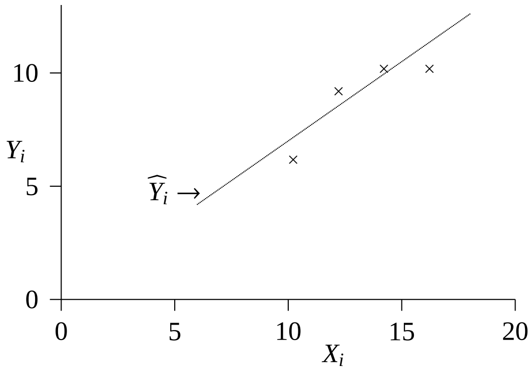
上の数値例では,

$$\widehat{Y_i} = 0.3 + 0.65 X_i$$

となる。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$
1	6	10	60	100	6.8
2	9	12	108	144	8.1
3	10	14	140	196	9.4
4	10	16	160	256	10.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$
	35	52	468	696	35.0
平均	$\overline{Y}$	$\overline{X}$			
	8.75	13			

図 2： Y_i， X_i， $\widehat{Y_i}$



$\widehat{Y_i}$ を実績値 Y_i の予測値または理論値と呼ぶ。

$$\widehat{u_i} = Y_i - \widehat{Y_i},$$

$\widehat{u}_i$ を残差と呼ぶ。

$$Y_i = \widehat{Y}_i + \widehat{u}_i = \widehat{\alpha} + \widehat{\beta} X_i + \widehat{u}_i,$$

さらに，$\overline{Y}$ を両辺から引いて，

$$(Y_i - \overline{Y}) = (\widehat{Y}_i - \overline{Y}) + \widehat{u}_i,$$

1.3　残差 $\widehat{u}_i$ の性質について

$\widehat{u}_i = Y_i - \widehat{\alpha} - \widehat{\beta} X_i$ に注意して，(1) 式から，

$$\sum_{i=1}^{n} \widehat{u}_i = 0,$$

を得る。　　(2) 式から，

$$\sum_{i=1}^{n} X_i \widehat{u}_i = 0,$$

を得る。　$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} X_i$ から，

$$\sum_{i=1}^{n} \widehat{Y}_i \widehat{u}_i = 0,$$

を得る。なぜなら，

$$\begin{aligned}\sum_{i=1}^{n} \widehat{Y}_i \widehat{u}_i &= \sum_{i=1}^{n} (\widehat{\alpha} + \widehat{\beta} X_i) \widehat{u}_i \\ &= \widehat{\alpha} \sum_{i=1}^{n} \widehat{u}_i + \widehat{\beta} \sum_{i=1}^{n} X_i \widehat{u}_i \\ &= 0\end{aligned}$$

である。

i	Y_i	X_i	$\widehat{Y}_i$	$\widehat{u}_i$	$X_i\widehat{u}_i$	$\widehat{Y}_i\widehat{u}_i$
1	6	10	6.8	−0.8	−8.0	−5.44
2	9	12	8.1	0.9	10.8	7.29
3	10	14	9.4	0.6	8.4	5.64
4	10	16	10.7	−0.7	−11.2	−7.49
合計	$\sum Y_i$	$\sum X_i$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$	$\sum X_i\widehat{u}_i$	$\sum \widehat{Y}_i\widehat{u}_i$
	35	52	35.0	0.0	0.0	0.00

1.4 決定係数 R^2 について

次の式

$$(Y_i - \overline{Y}) = (\widehat{Y}_i - \overline{Y}) + \widehat{u}_i,$$

の両辺を二乗して，総和すると，

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 &= \sum_{i=1}^{n}\left((\widehat{Y_i} - \overline{Y}) + \widehat{u_i}\right)^2 \\ &= \sum_{i=1}^{n}(\widehat{Y_i} - \overline{Y})^2 + 2\sum_{i=1}^{n}(\widehat{Y_i} - \overline{Y})\widehat{u_i} + \sum_{i=1}^{n}\widehat{u_i^2} \\ &= \sum_{i=1}^{n}(\widehat{Y_i} - \overline{Y})^2 + \sum_{i=1}^{n}\widehat{u_i^2}\end{aligned}$$

となる。まとめると，

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(\widehat{Y_i} - \overline{Y})^2 + \sum_{i=1}^{n}\widehat{u_i^2}$$

を得る。さらに，

$$1 = \frac{\sum_{i=1}^{n}(\widehat{Y_i} - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} + \frac{\sum_{i=1}^{n}\widehat{u_i^2}}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

それぞれの項は，

1. $\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \Longrightarrow y$ の全変動

2. $\sum_{i=1}^{n}(\widehat{Y_i} - \overline{Y})^2 \Longrightarrow \widehat{Y_i}$ (回帰直線) で説明される部分

3. $\sum_{i=1}^{n}\widehat{u}_i^2 \Longrightarrow \widehat{Y_i}$ (回帰直線) で説明されない部分

となる。

回帰式の当てはまりの良さを示す指標として，決定係数 R^2 を以下の通りに定義する。

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{Y_i} - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

または,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2},$$

として書き換えられる。

または，$Y_i = \widehat{Y}_i + \widehat{u}_i$ と

$$\begin{aligned}\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 &= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y} - \widehat{u}_i) \\ &= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y}) - \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})\widehat{u}_i \\ &= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y})\end{aligned}$$

を用いて，

$$\begin{aligned}R^2 &= \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \\ &= \frac{\left(\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2\right)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2} \\ &= \left(\frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}}\right)^2\end{aligned}$$

と書き換えられる。　　すなわち，R^2 は Y_i と $\widehat{Y}_i$ の相関係数の二乗と解釈される。

$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n}\widehat{u}_i^2$ から，明らかに，

$$0 \leq R^2 \leq 1,$$

となる。R^2 が 1 に近づけば回帰式の当てはまりは良いと言える。しかし，t 分布のような数表は存在しない。したがって,「どの値よりも大きくなるべき」というような基準はない。

慣習的には，メドとして 0.9 以上を判断基準にする。

数値例：　決定係数の計算には以下の公式を用いる。

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}Y_i^2 - n\overline{Y}^2}$$

計算に必要なものは，$\widehat{u}_i = Y_i - (\widehat{\alpha} + \widehat{\beta} X_i)$，$\overline{Y}$，$\sum_{i=1}^{n} Y_i^2$ である。

i	Y_i	X_i	$\widehat{Y}_i$	$\widehat{u}_i$	$\widehat{u}_i$	Y_i^2
1	6	10	6.8	−0.8	0.64	36
2	9	12	8.1	0.9	0.81	81
3	10	14	9.4	0.6	0.36	100
4	10	16	10.7	−0.7	0.49	100
合計	$\sum Y_i$	$\sum X_i$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$	$\sum \widehat{u}_i^2$	$\sum Y_i^2$
	35	52	35.0	0.0	2.30	317

$\sum \widehat{u}_i^2 = 2.30$，$\overline{X} = 13$，$\overline{Y} = 8.75$，$\sum_{i=1}^{n} Y_i^2 = 317$ なので，

$$R^2 = 1 - \frac{2.30}{317 - 4 \times 8.75^2} = 1 - \frac{2.30}{10.75} = 0.786$$

1.5　まとめ

$\widehat{\alpha}$，$\widehat{\beta}$ を求めるための公式は

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

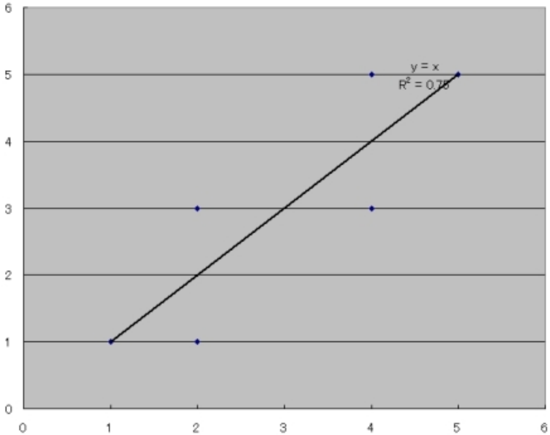
$$\widehat{\alpha} = \overline{Y} - \widehat{\beta \overline{X}}$$

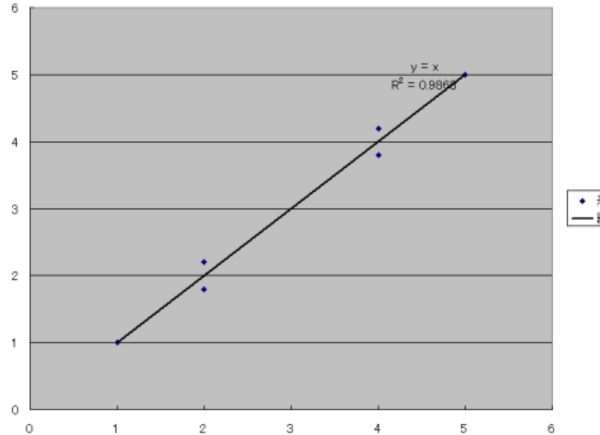
なので，必要なものは $\overline{X}$，$\overline{Y}$，$\displaystyle\sum_{i=1}^{n} X_i^2$，$\displaystyle\sum_{i=1}^{n} X_i Y_i$ である。

決定係数の計算には以下の公式を用いる。

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \widehat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^{n} \widehat{u}_i^2}{\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2}$$

計算に必要なものは，$\sum \widehat{u}_i^2$，$\overline{Y}$，$\displaystyle\sum_{i=1}^{n} Y_i^2$ である。


y = x
R^2 = 0.75
0 1 2 3 4 5 6
0 1 2 3 4 5 6


y = x
R^2 = 0.9868
0 1 2 3 4 5 6
0 1 2 3 4 5 6

2 Regression Analysis (回帰分析)

2.1 Setup of the Model

When (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) are available, suppose that there is a linear relationship between y and x, i.e.,

$$y_i = \beta_1 + \beta_2 x_i + u_i, \tag{4}$$

for $i = 1, 2, \cdots, n$. x_i and y_i denote the ith observations.

$\longrightarrow$ **Single (or simple) regression model (単回帰モデル)**

y_i is called the **dependent variable (従属変数)** or the **explained variable (被説明変数)**, while x_i is known as the **independent variable (独立変数)** or the **explanatory (or explaining) variable (説明変数)**.

β_1 = **Intercept (切片),** β_2 = **Slope (傾き)**

β_1 and β_2 are unknown **parameters (パラメータ，母数)** to be estimated.

β_1 and β_2 are called the **regression coefficients (回帰係数)**.

u_i is the unobserved **error term (誤差項)** assumed to be a random variable with mean zero and variance σ^2.

σ^2 is also a parameter to be estimated.

x_i is assumed to be **nonstochastic (非確率的)**, but y_i is **stochastic (確率的)** because y_i depends on the error u_i.

The error terms $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed, which is called ***iid***.

It is assumed that u_i has a distribution with mean zero, i.e., $\mathrm{E}(u_i) = 0$ is assumed.

Taking the expectation on both sides of (4), the expectation of y_i is represented as:

$$\begin{aligned} \mathrm{E}(y_i) &= \mathrm{E}(\beta_1 + \beta_2 x_i + u_i) = \beta_1 + \beta_2 x_i + \mathrm{E}(u_i) \\ &= \beta_1 + \beta_2 x_i, \end{aligned} \tag{5}$$

for $i = 1, 2, \cdots, n$.

Using $\mathrm{E}(y_i)$ we can rewrite (4) as $y_i = \mathrm{E}(y_i) + u_i$.

(5) represents the true regression line.

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be estimates of β_1 and β_2.

Replacing β_1 and β_2 by $\hat{\beta}_1$ and $\hat{\beta}_2$, (4) turns out to be:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + e_i, \tag{6}$$

for $i = 1, 2, \cdots, n$, where e_i is called the **residual (残差)**.

The residual e_i is taken as the experimental value (or realization) of u_i.

We define $\hat{y}_i$ as follows:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i, \tag{7}$$

for $i = 1, 2, \cdots, n$, which is interpreted as the **predicted value (予測値)** of y_i.

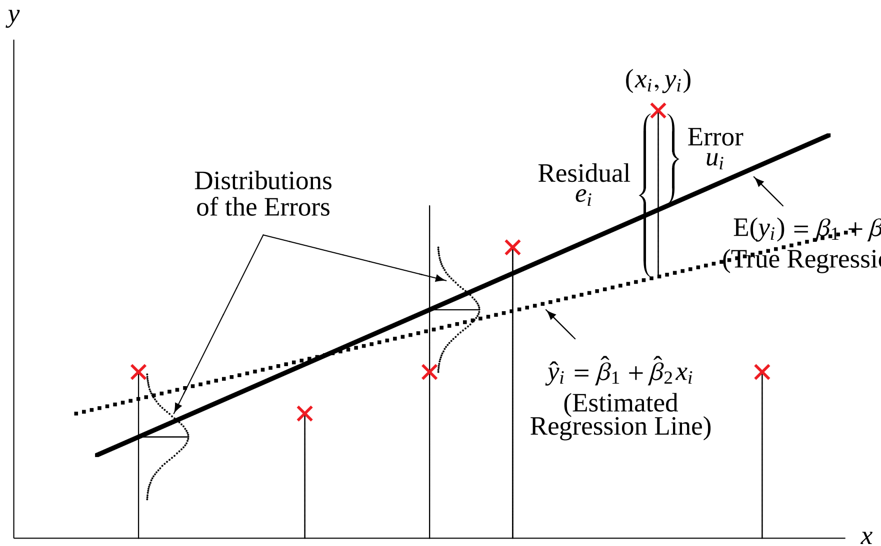
(7) indicates the estimated regression line, which is different from (5).

Moreover, using $\hat{y}_i$ we can rewrite (6) as $y_i = \hat{y}_i + e_i$.

(5) and (7) are displayed in Figure 1.

Consider the case of $n = 6$ for simplicity. $\times$ indicates the observed data series.

Figure 1. True and Estimated Regression Lines (回帰直線)



The true regression line (5) is represented by the solid line, while the estimated regression line (7) is drawn with the dotted line.

Based on the observed data, β_1 and β_2 are estimated as: $\hat{\beta}_1$ and $\hat{\beta}_2$.

In the next section, we consider how to obtain the estimates of β_1 and β_2, i.e., $\hat{\beta}_1$ and $\hat{\beta}_2$.

2.2 Ordinary Least Squares Estimation

Suppose that (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) are available.

For the regression model (4), we consider estimating β_1 and β_2.

Replacing β_1 and β_2 by their estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, remember that the residual e_i is given by:

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i.$$

The sum of squared residuals is defined as follows:

$$S(\hat{\beta}_1, \hat{\beta}_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2.$$

It might be plausible to choose the $\hat{\beta}_1$ and $\hat{\beta}_2$ which minimize the sum of squared residuals, i.e., $S(\hat{\beta}_1, \hat{\beta}_2)$.

This method is called the **ordinary least squares estimation (最小二乘法，OLS)**.

To minimize $S(\hat{\beta}_1, \hat{\beta}_2)$ with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$, we set the partial derivatives equal to zero:

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0,$$

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0.$$

The second order condition for minimization is:

$$\begin{pmatrix} \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_1^2} & \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_1 \partial \hat{\beta}_2} \\ \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_2 \partial \hat{\beta}_1} & \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_2^2} \end{pmatrix} = \begin{pmatrix} 2n & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2\sum_{i=1}^n x_i^2 \end{pmatrix}$$

should be a positive definite matrix.

The diagonal elements $2n$ and $2\sum_{i=1}^n x_i^2$ are positive.

The determinant:

$$\begin{vmatrix} 2n & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2\sum_{i=1}^n x_i^2 \end{vmatrix} = 4n\sum_{i=1}^n x_i^2 - 4(\sum_{i=1}^n x_i)^2 = 4n\sum_{i=1}^n (x_i - \overline{x})^2$$

is positive. $\Longrightarrow$ The second-order condition is satisfied.

The first two equations yield the following two equations:

$$\overline{y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{x}, \tag{8}$$

$$\sum_{i=1}^n x_i y_i = n\overline{x}\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^n x_i^2, \tag{9}$$

where $\overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$.

Multiplying (8) by $n\overline{x}$ and subtracting (9), we can derive $\hat{\beta}_2$ as follows:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n x_i y_i - n\overline{x}\overline{y}}{\sum_{i=1}^n x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2}. \tag{10}$$

From (8), $\hat{\beta}_1$ is directly obtained as follows:

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}. \tag{11}$$

When the observed values are taken for y_i and x_i for $i = 1, 2, \cdots, n$, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimate**s (or simply the **least squares estimate**s, 最小二乗推定値) of β_1 and β_2.

When y_i for $i = 1, 2, \cdots, n$ are regarded as the random sample, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimator**s (or the **least squares estimator**s, 最小二乗推定量) of β_1 and β_2.

2.3 Properties of Least Squares Estimator

Equation (10) is rewritten as:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sum_{i=1}^n (x_i - \overline{x})y_i}{\sum_{i=1}^n (x_i - \overline{x})^2} - \frac{\overline{y}\sum_{i=1}^n (x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}$$
$$= \sum_{i=1}^n \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} y_i = \sum_{i=1}^n \omega_i y_i. \qquad (12)$$

In the third equality, $\sum_{i=1}^n (x_i - \overline{x}) = 0$ is utilized because of $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

In the fourth equality, ω_i is defined as: $\omega_i = \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2}$.

ω_i is nonstochastic because x_i is assumed to be nonstochastic.

ω_i has the following properties:

$$\sum_{i=1}^n \omega_i = \sum_{i=1}^n \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sum_{i=1}^n (x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0, \qquad (13)$$

$$\sum_{i=1}^{n} \omega_i x_i = \sum_{i=1}^{n} \omega_i (x_i - \overline{x}) = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} = 1, \tag{14}$$

$$\sum_{i=1}^{n} \omega_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \right)^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{\left(\sum_{i=1}^{n} (x_i - \overline{x})^2\right)^2} = \frac{1}{\sum_{i=1}^{n} (x_i - \overline{x})^2}. \tag{15}$$

The first equality of (14) comes from (13).

From now on, we focus only on $\hat{\beta}_2$, because usually β_2 is more important than β_1 in the regression model (4).

In order to obtain the properties of the least squares estimator $\hat{\beta}_2$, we rewrite (12) as:

$$\begin{aligned} \hat{\beta}_2 &= \sum_{i=1}^{n} \omega_i y_i = \sum_{i=1}^{n} \omega_i (\beta_1 + \beta_2 x_i + u_i) \\ &= \beta_1 \sum_{i=1}^{n} \omega_i + \beta_2 \sum_{i=1}^{n} \omega_i x_i + \sum_{i=1}^{n} \omega_i u_i = \beta_2 + \sum_{i=1}^{n} \omega_i u_i. \end{aligned} \tag{16}$$

In the fourth equality of (16), (13) and (14) are utilized.

[Review] Random Variables:

Let X_1, X_2, $\cdots$, X_n be n random variavles, which are mutually independently and identically distributed.

mutually independent $\Longrightarrow$ $f(x_i, x_j) = f_i(x_i) f_j(x_j)$ for $i \neq j$.

$f(x_i, x_j)$ denotes a joint distribution of X_i and X_j.

$f_i(x)$ indicates a marginal distribution of X_i.

identical $\Longrightarrow$ $f_i(x) = f_j(x)$ for $i \neq j$.

[End of Review]

[Review] Mean and Variance:

Let X and Y be random variables (continuous type), which are independently distributed.

Definition and Formulas:

- $\mathrm{E}(g(X)) = \int g(x)f(x)\mathrm{d}x$ for a function $g(\cdot)$ and a density function $f(\cdot)$.
- $\mathrm{V}(X) = \mathrm{E}((X-\mu)^2) = \int (x-\mu)^2 f(x)\mathrm{d}x$ for $\mu = \mathrm{E}(X)$.
- $\mathrm{E}(aX+b) = a\mathrm{E}(X)+b$ and $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$.
- $\mathrm{E}(X \pm Y) = \mathrm{E}(X) \pm \mathrm{E}(Y)$ and $\mathrm{V}(X \pm Y) = \mathrm{V}(X) + \mathrm{V}(Y)$.

[End of Review]

Mean and Variance of $\hat{\beta}_2$: $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2, but they are not necessarily normal.

Remember that we do not need normality assumption to obtain mean and variance but the normality assumption is required to test a hypothesis.

From (16), the expectation of $\hat{\beta}_2$ is derived as follows:

$$\mathrm{E}(\hat{\beta}_2) = \mathrm{E}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \beta_2 + \mathrm{E}(\sum_{i=1}^{n} \omega_i u_i) = \beta_2 + \sum_{i=1}^{n} \omega_i \mathrm{E}(u_i) = \beta_2. \qquad (17)$$

It is shown from (17) that the ordinary least squares estimator $\hat{\beta}_2$ is an **unbiased estimator (不偏推定量)** of β_2.

From (16), the variance of $\hat{\beta}_2$ is computed as:

$$\begin{aligned} \mathrm{V}(\hat{\beta}_2) &= \mathrm{V}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \mathrm{V}(\sum_{i=1}^{n} \omega_i u_i) = \sum_{i=1}^{n} \mathrm{V}(\omega_i u_i) = \sum_{i=1}^{n} \omega_i^2 \mathrm{V}(u_i) \\ &= \sigma^2 \sum_{i=1}^{n} \omega_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}. \end{aligned} \tag{18}$$

The third equality holds because $u_1, u_2, \cdots, u_n$ are mutually independent.

The last equality comes from (15).

Thus, $\mathrm{E}(\hat{\beta}_2)$ and $\mathrm{V}(\hat{\beta}_2)$ are given by (17) and (18).

Gauss-Markov Theorem (ガウス・マルコフ定理): $\hat{\beta}_2$ has minimum variance within a class of the linear unbiased estimators.

$\longrightarrow$ **best linear unbiased estimator (BLUE, 最良線型不偏推定量)**

(Proof is omitted.)

Distribution of $\hat{\beta}_2$: We discuss the small sample properties of $\hat{\beta}_2$.

In order to obtain the distribution of $\hat{\beta}_2$ in small sample, the distribution of the error term has to be assumed.

Therefore, the extra assumption is that $u_i \sim N(0, \sigma^2)$.

Writing (16), again, $\hat{\beta}_2$ is represented as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i.$$

First, we obtain the distribution of the second term in the above equation.

It is well known that sum of normal random variables results in a normal distribution.

Therefore, $\sum_{i=1}^{n} \omega_i u_i$ is distributed as:

$$\sum_{i=1}^{n} \omega_i u_i \sim N(0,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2).$$

Therefore, $\hat{\beta}_2$ is distributed as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i \sim N(\beta_2,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2),$$

or equivalently,

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma \sqrt{\sum_{i=1}^{n} \omega_i^2}} = \frac{\hat{\beta}_2 - \beta_2}{\sigma / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \sim N(0, 1),$$

for any n.

Moreover, replacing σ^2 by its estimator $s^2 = \dfrac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$, it is known that we have:

$$\frac{\hat{\beta}_2 - \beta_2}{s / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \sim t(n-2),$$

where $t(n-2)$ denotes t distribution with $n-2$ degrees of freedom.

Thus, under normality assumption on the error term u_i, the $t(n-2)$ distribution is used for the confidence interval and the testing hypothesis in small sample.

Or, taking the square on both sides,

$$\Big(\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\Big)^2 \sim F(1, n-2).$$

[Review] Confidence Interval (信頼区間，区間推定)):

Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Then, we can obtain: $\dfrac{\overline{X} - \mu}{S / \sqrt{n}} \sim t(n-1)$, where $S^2 = \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n} (X_i - \overline{X})^2$.

That is,

$$P\Big(-t_{\alpha/2}(n-1) < \frac{\overline{X} - \mu}{S / \sqrt{n}} < t_{\alpha/2}(n-1)\Big) = 1 - \alpha$$

i.e.,

$$P\Big(\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\Big) = 1 - \alpha.$$

Note that $t_{\alpha/2}(n-1)$ is obtained from the t distribution table, given α and $n-1$.

Then, replacing $\overline{X}$ by $\overline{x}$, we obtain the $100(1-\alpha)\%$ confidence interval of μ as follows:

$$(\overline{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ \overline{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}).$$

[End of Review]

In the case of OLS,

$$P\left(-t_{\alpha/2}(n-2) < \frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} < t_{\alpha/2}(n-2)\right) = 1 - \alpha,$$

where $t_{\alpha/2}(n-2)$ denotes $100 \times \alpha/2\%$ point from the $t(n-2)$ distribution.

Rewriting,

$$P\left(\hat{\beta}_2 - t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} < \beta_2 < \hat{\beta}_2 + t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\right) = 1 - \alpha.$$

Replacing $\hat{\beta}_2$ and s^2 by observed data, the $100(1-\alpha)\%$ confidence interval of β_2 is given by:

$$\left(\hat{\beta}_2 - t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}},\ \hat{\beta}_2 + t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\right).$$

[Review] Testing the Hypothesis (仮説検定):

Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Then, we obtain: $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$, where $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, which is known as the unbiased estimator of σ^2.

- The null hypothesis $H_0: \mu = \mu_0$, where μ_0 is a fixed number.
- The alternative hypothesis $H_1: \mu \neq \mu_0$

Under the null hypothesis, we have the disribution: $\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$.

Replacing $\overline{X}$ and S^2 by $\overline{x}$ and s^2, compare $\dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$ and $t(n-1)$.

H_0 is rejected when $\left|\dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}\right| > t_{\alpha/2}(n-1)$.

$t_{\alpha/2}(n-1)$ is obtained from the significance level α and the degrees of freedom $n-1$.

[End of Review]

In the case of OLS, the hypotheses are as follows:

- The null hypothesis $H_0: \beta_2 = \beta_2^*$
- The alternative hypothesis $H_1: \beta_2 \neq \beta_2^*$

Under H_0,

$$\frac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} \sim t(n-2).$$

Replacing $\hat{\beta}_2$ and s^2 by the observed data, compare $\dfrac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}$ and $t(n-2)$.

H_0 is rejected at significance level α when $\left|\dfrac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\right| > t_{\alpha/2}(n-1)$.

(*) $\hat{\beta}_2$ = Coefficient, $\dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}$ = Standard Error,

s = Standard Error of Regression

3　多重回帰

n 組のデータ $(Y_i, X_{1i}, X_{2i}, \cdots, X_{ki})$, $i = 1, 2, \cdots, n$ を用いて，k 変数の多重回帰モデルを考える。

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,$$

ただし，X_{ji} は j 番目の説明変数の第 i 番目の観測値を表す。u_i は誤差項 (または，攪乱項) で，同じ仮定を用いる (すなわち，$u_1, u_2, \cdots, u_n$ は互いに独立に，平均ゼロ，分散 σ^2 の正規分布に従う)。

$\beta_1, \beta_2, \cdots, \beta_k$ は推定されるべきパラメータである。

すべての i について，$X_{1i} = 1$ とすれば，β_1 は定数項として表される。

次のような関数 $S(\beta_1, \beta_2, \cdots, \beta_k)$ を定義する。

$$S(\beta_1, \beta_2, \cdots, \beta_k) = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (Y_i - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2$$

このとき,

$$\min_{\beta_1,\beta_2,\cdots,\beta_k} S\,(\beta_1,\beta_2,\cdots,\beta_k)$$

となるような $\beta_1, \beta_2, \cdots, \beta_k$ を求める。$\Longrightarrow$ 最小自乗法

このときの解を $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ とする。

最小化のためには,

$$\frac{\partial S\,(\beta_1,\beta_2,\cdots,\beta_k)}{\partial\beta_1} = 0, \qquad \frac{\partial S\,(\beta_1,\beta_2,\cdots,\beta_k)}{\partial\beta_2} = 0, \qquad \cdots, \qquad \frac{\partial S\,(\beta_1,\beta_2,\cdots,\beta_k)}{\partial\beta_k} = 0$$

を満たす $\beta_1, \beta_2, \cdots, \beta_k$ が $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ となる。

すなわち, $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ は,

$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})X_{1i} = 0,$$

$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})X_{2i} = 0,$$

$$\vdots$$

$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})X_{ki} = 0,$$

を満たす。

さらに,

$$\sum_{i=1}^{n} X_{1i}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}^2 + \widehat{\beta}_2 \sum_{i=1}^{n} X_{1i}X_{2i} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{1i}X_{ki},$$

$$\sum_{i=1}^{n} X_{2i}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{2i} + \widehat{\beta}_2 \sum_{i=1}^{n} X_{2i}^2 + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{2i}X_{ki},$$

$$\vdots$$

$$\sum_{i=1}^{n} X_{ki}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{ki} + \widehat{\beta}_2 \sum_{i=1}^{n} X_{2i}X_{ki} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{ki}^2,$$

行列表示によって，

$$\begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix} \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix},$$

が得られ，$\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ についてまとめると，

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix},$$

を解くことになる。$\Longrightarrow$ コンピュータによって計算

3.1　推定量の性質

$\beta_1, \beta_2, \cdots, \beta_k$ の最小二乗推定量は $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ とする。

誤差項 (または，攪乱項) u_i の分散 σ^2 の推定量 s^2 は，

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n}\widehat{u}_i^2 = \frac{1}{n-k}\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})^2$$

として表される。

このとき，

$$\mathrm{E}(\widehat{\beta}_j) = \beta_j, \qquad \mathrm{E}(s^2) = \sigma^2,$$

を証明することが出来る。(証明略)

分布について： $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ の分散は以下のように表される。

$$
\mathrm{V}\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix} = \begin{pmatrix} \mathrm{V}(\widehat{\beta}_1) & \mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) & \cdots & \mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_k) \\ \mathrm{Cov}(\widehat{\beta}_2, \widehat{\beta}_1) & \mathrm{V}(\widehat{\beta}_2) & \cdots & \mathrm{Cov}(\widehat{\beta}_2, \widehat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\widehat{\beta}_k, \widehat{\beta}_1) & \mathrm{Cov}(\widehat{\beta}_k, \widehat{\beta}_2) & \cdots & \mathrm{V}(\widehat{\beta}_k) \end{pmatrix}
$$

$$
= \sigma^2 \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix}^{-1}
$$

$\widehat{\beta}_j$ の分散 (すなわち，上の逆行列の j 番目の対角要素) を，

$$
\mathrm{V}(\widehat{\beta}_j) = \sigma^2_{\widehat{\beta}_j},
$$

として，その推定量を $s^2_{\widehat{\beta}_j}$ とする。

このとき，

$$\widehat{\beta}_j \sim N(\beta_j, \sigma^2_{\widehat{\beta}_j}),$$

となり，標準化すると，

$$\frac{\widehat{\beta}_j - \beta_j}{\sigma_{\widehat{\beta}_j}} \sim N(0, 1),$$

が得られる。さらに，

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k),$$

となり (証明略)，しかも，$\widehat{\beta}_j$ と s^2 の独立性から (証明略)，

$$\frac{\widehat{\beta}_j - \beta_j}{s_{\widehat{\beta}_j}} \sim t(n-k)$$

となる。

よって，通常の区間推定や仮説検定を行うことが出来る。

決定係数について：　また，決定係数 R^2 についても同様に表される。

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

ただし，$\widehat{Y}_i = \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_k X_{ki}$，$Y_i = \widehat{Y}_i + \widehat{u}_i$ である。

R^2 は，説明変数を増やすことによって，必ず大きくなる。なぜなら，説明変数が増えることによって，$\sum_{i=1}^{n}\widehat{u}_i^2$ が必ず減少するからである。

R^2 を基準にすると，被説明変数にとって意味のない変数でも，説明変数が多いほど，よりよいモデルということになる。この点を改善するために，自由度修正済み決定係数 $\overline{R}^2$ を用いる。

$$\overline{R}^2 = 1 - \frac{\sum_{i=1}^{n}\widehat{u}_i^2/(n-k)}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2/(n-1)},$$

$\sum_{i=1}^{n}\widehat{u}_i^2/(n-k)$ は u_i の分散 σ^2 の不偏推定量であり，$\sum_{i=1}^{n}(Y_i-\overline{Y})^2/(n-1)$ は Y_i の分散の不偏推定量である。

R^2 と $\overline{R}^2$ との関係は，

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k},$$

となる。さらに，

$$\frac{1-\overline{R}^2}{1-R^2} = \frac{n-1}{n-k} \geq 1,$$

という関係から，$\overline{R}^2 \leq R^2$ という結果を得る。($k = 1$ のときのみに，等号が成り立つ。)

数値例：　今までと同じ数値例で，$\overline{R}^2$ を計算する。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$	$\widehat{u}_i$
1	6	10	60	100	6.8	−0.8
2	9	12	108	144	8.1	0.9
3	10	14	140	196	9.4	0.6
4	10	16	160	256	10.7	−0.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$
	35	52	468	696	35	0
平均	$\overline{Y}$	$\overline{X}$				
	8.75	13				

まず R^2 は,

$$R^2 = 1 - \frac{\sum \widehat{u}_i^2}{\sum Y_i^2 - n\overline{Y}^2} = 1 - \frac{(-0.8)^2 + 0.9^2 + 0.6^2 + (-0.7)^2}{35 - 4 \times 8.75^2} = 1 - \frac{2.30}{10.75} = 0.786$$

となり，$\overline{R}^2$ は，

$$\overline{R}^2 = 1 - \frac{\sum \widehat{u}_i^2/(n-k)}{(\sum Y_i^2 - n\overline{Y}^2)/(n-1)} = 1 - \frac{2.30/(4-2)}{10.75/(4-1)} = 0.679$$

となる。

注意： R^2 や $\overline{R}^2$ を比較する場合，被説明変数が同じことが必要である。被説明変数が異なる場合 (例えば，被説明変数を上昇率とするかそのままの値を用いるかによって，被説明変数が異なる)，誤差項 u_i の標準誤差で比較すべきである (標準誤差の小さいモデルを採用する)。$\Longrightarrow$ 関数型の選択

4　系列相関：*DW* について

4.1　*DW* について

最小自乗法の仮定の一つに,「攪乱項 $u_1, u_2, \cdots, u_n$ はそれぞれ独立に分布する」というものがあった。ダービン・ワトソン比 (DW) とは，誤差項の系列相関，すなわち，u_i と u_{i-1} との間の相関の有無を検定するために考案された。

$\Longrightarrow$ 時系列データのときのみ有効

$u_1, u_2, \cdots, u_n$ の系列について，それぞれの符号が，+ + + - - - - + + - - - + + のように，プラスが連続で続いた後で，マイナスが連続で続くというような場合，$u_1, u_2, \cdots, u_n$ は正の系列相関があると言う。また，+ - + - + - + - + のように交互にプラス，マイナスになる場合，$u_1, u_2, \cdots, u_n$ 負の系列相関があると言う。

特徴：$u_1, u_2, \cdots, u_i$ から u_{i+1} の符号が予想できる。$\Longrightarrow$「$u_1, u_2, \cdots, u_n$ はそれぞれ独立に分布する」という仮定に反する。

すなわち，ダービン・ワトソン比とは，回帰式が

$$Y_i = \alpha + \beta X_i + u_i,$$

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときに，H_0：$\rho = 0$, H_1：$\rho \neq 0$ の検定である。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

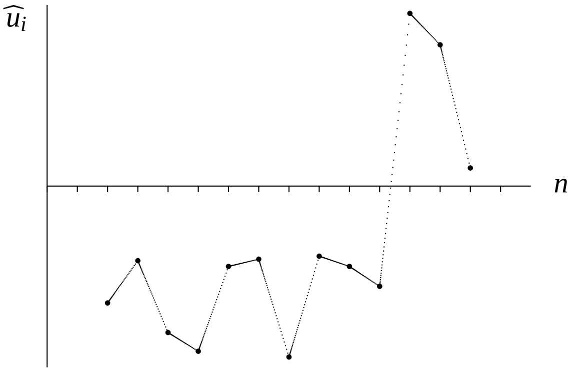


図 4： 正の系列相関

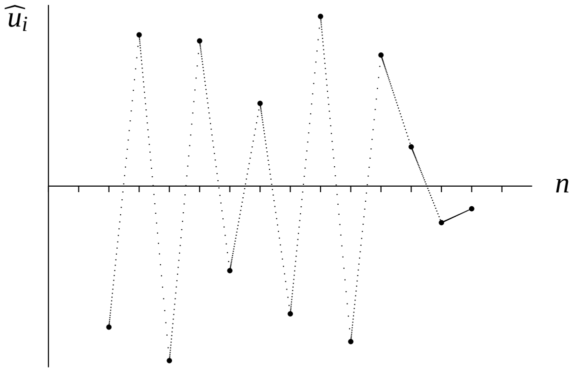


図 5： 負の系列相関

ダービン・ワトソン比の定義は次の通りである。

$$DW = \frac{\sum_{i=2}^{n}(\widehat{u}_i - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2}$$

DW は近似的に，次のように表される。

$$\begin{aligned}DW &= \frac{\sum_{i=2}^{n}(\widehat{u}_i - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2} = \frac{\sum_{i=2}^{n}\widehat{u}_i^2 - 2\sum_{i=2}^{n}\widehat{u}_i\widehat{u}_{i-1} + \sum_{i=2}^{n}\widehat{u}_{i-1}^2}{\sum_{i=1}^{n}\widehat{u}_i^2} \\ &= \frac{2\sum_{i=1}^{n}\widehat{u}_i^2 - (\widehat{u}_1^2 + \widehat{u}_n^2)}{\sum_{i=1}^{n}\widehat{u}_i^2} - 2\frac{\sum_{i=2}^{n}\widehat{u}_i\widehat{u}_{i-1}}{\sum_{i=1}^{n}\widehat{u}_i^2} \approx 2(1-\widehat{\rho}),\end{aligned}$$

以下の 2 つの近似が用いられる。

$$\begin{aligned}&\frac{\widehat{u}_1^2 + \widehat{u}_n^2}{\sum_{i=1}^{n}\widehat{u}_i^2} \approx 0, \\ &\frac{\sum_{i=2}^{n}\widehat{u}_i\widehat{u}_{i-1}}{\sum_{i=1}^{n}\widehat{u}_i^2} = \frac{\sum_{i=2}^{n}\widehat{u}_i\widehat{u}_{i-1}}{\sum_{i=2}^{n}\widehat{u}_{i-1}^2 + \widehat{u}_n^2} \approx \frac{\sum_{i=2}^{n}\widehat{u}_i\widehat{u}_{i-1}}{\sum_{i=2}^{n}\widehat{u}_{i-1}^2} = \widehat{\rho},\end{aligned}$$

すなわち，$\widehat{\rho}$ は $\widehat{u}_i$ と $\widehat{u}_{i-1}$ の回帰係数である。$u_i = \rho u_{i-1} + \epsilon_i$ において，u_i, u_{i-1} の代わりに $\widehat{u}_i, \widehat{u}_{i-1}$ に置き換えて，ρ の推定値 $\widehat{\rho}$ を求める。

1. DW の値が 2 前後のとき，系列相関なし ($\widehat{\rho} = 0$ のとき，$DW \approx 2$)。

2. DW が 2 より十分に小さいとき，正の系列相関と判定される。

3. DW が 2 より十分に大きいとき，負の系列相関と判定される。

正確な判定には，データ数 n とパラメータ数 k に依存する。表 1 を参照せよ。
k' は定数項を除くパラメータ数を表すものとする。

See `http://www.stanford.edu/~clint/bench/dwcrit.htm` for the DW table.

Table 1: ダービン・ワトソン統計量の 5 % 点の上限と下限

	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$		$k' = 6$		$k' = 7$		$k' = 8$		$k' = 9$		$k' = 10$		k'
n	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl
6	0.610	1.400	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
7	0.700	1.356	0.467	1.896	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8	0.763	1.332	0.559	1.777	0.367	2.287	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	—	—	—	—	—	—	—	—	—	—	—	—	—
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	—	—	—	—	—	—	—	—	—	—	—
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	—	—	—	—	—	—	—	—	—
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	—	—	—	—	—	—	—
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	—	—	—	—	—
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	—	—	—
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438	—
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304	0.09
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184	0.13
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073	0.17
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974	0.22
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885	0.26
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806	0.30
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735	0.34
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670	0.39
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613	0.43
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560	0.47
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513	0.50
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470	0.54
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.959	0.874	2.071	0.798	2.188	0.723	2.309	0.649	2.431	0.57

$$DW = \frac{\sum_{i=2}^{n}(\widehat{u}_i - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2} \approx 2(1-\widehat{\rho}) \longrightarrow 2(1-\rho)$$

$-1 < \rho < 1$ なので (証明略)，近似的に $0 \leq DW \leq 4$ となる。

- $0 \leq DW \leq dl \quad \longrightarrow \quad u_i$ に正の系列相関
- $dl \leq DW \leq du \quad \longrightarrow \quad u_i$ に正の系列相関と判定できない
- $du \leq DW \leq 4 - du \quad \longrightarrow \quad u_i$ に系列相関なし
- $4 - du \leq DW \leq 4 - dl \quad \longrightarrow \quad u_i$ に負の系列相関と判定できない
- $4 - dl \leq DW \leq 4 \quad \longrightarrow \quad u_i$ に負の系列相関

数値例：　今までと同じ数値例で，DW を計算する。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$	$\widehat{u}_i$
1	6	10	60	100	6.8	−0.8
2	9	12	108	144	8.1	0.9
3	10	14	140	196	9.4	0.6
4	10	16	160	256	10.7	−0.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$
	35	52	468	696	35	0
平均	$\overline{Y}$	$\overline{X}$				
	8.75	13				

$$
\begin{aligned}
DW &= \frac{\sum_{i=2}^{n}(\widehat{u}_i - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2} \\
&= \frac{(-0.8-0.9)^2 + (0.9-0.6)^2 + (0.6-(-0.7))^2}{(-0.8)^2 + 0.9^2 + 0.6^2 + (-0.7)^2} = \frac{4.67}{2.30} = 2.03
\end{aligned}
$$

推定結果の表記方法：　回帰モデル：

$$Y_i = \alpha + \beta X_i + u_i,$$

の推定の結果，$\widehat{\alpha} = 0.3, \widehat{\beta} = 0.65$, $s_{\widehat{\alpha}} = \sqrt{10.0005} = 3.163$, $s_{\widehat{\beta}} = \sqrt{0.0575} = 0.240$, $\dfrac{\widehat{\alpha}}{s_{\widehat{\alpha}}} = 0.095$, $\dfrac{\widehat{\beta}}{s_{\widehat{\beta}}} = 2.708$, $s^2 = 1.15$ (すなわち，$s = 1.07$), $R^2 = 0.786$, $\overline{R}^2 = 0.679$, $DW = 2.03$ を得た。

これらをまとめて，

$$Y_i = \underset{(0.095)}{0.3} + \underset{(2.708)}{0.65}\, X_i,$$

$$R^2 = 0.786, \quad \overline{R}^2 = 0.679, \quad s = 1.07, \quad DW = 2.03,$$

ただし，係数の推定値の下の括弧内は t 値を表すものとする。

または,

$$Y_i = \underset{(3.163)}{0.3} + \underset{(0.240)}{0.65} X_i,$$

$$R^2 = 0.786, \quad \overline{R}^2 = 0.679, \quad s = 1.07, \quad DW = 2.03,$$

ただし，係数の推定値の下の括弧内は標準誤差を表すものとする。

のように書く。$s = \sqrt{1.15} = 1.07$ に注意。

4.2 系列相関のもとで回帰式の推定

回帰式が

$$Y_i = \alpha + \beta X_i + u_i,$$

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときの推定を考える。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

u_i を消去すると，

$$(Y_i - \rho Y_{i-1}) = \alpha(1-\rho) + \beta(X_i - \rho X_{i-1}) + \epsilon_i,$$

となり，

$Y_i^* = (Y_i - \rho Y_{i-1}),\ X_i^* = (X_i - \rho X_{i-1})$
を新たな変数として，

$$Y_i^* = \alpha' + \beta X_i^* + \epsilon_i,$$

に最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とするなので，最小二乗法を適用が可能となる。ただし，$\alpha' = \alpha(1-\rho)$ の関係が成り立つことに注意。

より一般的に，回帰式が

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,$$

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときの推定を考える。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

u_i を消去すると，

$$(Y_i - \rho Y_{i-1}) = \beta_1(X_{1i} - \rho X_{1,i-1}) + \beta_2(X_{1i} - \rho X_{2,i-1}) + \cdots + \beta_k(X_{1i} - \rho X_{k,i-1}) + \epsilon_i,$$

となり，

$Y_i^* = (Y_i - \rho Y_{i-1}),\ X_{1i}^* = (X_{1i} - \rho X_{1,i-1}),\ X_{2i}^* = (X_{2i} - \rho X_{2,i-1}),\ \cdots,\ X_{ki}^* = (X_{ki} - \rho X_{k,i-1})$ を新たな変数として，

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i$$

最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とするなので，最小二乗法を適用が可能となる。

ρ の求め方について (その 1):　DW は近似的に $DW \approx 2(1 - \widehat{\rho})$ と表されるので，DW から ρ の推定値 $\widehat{\rho}$ を逆算して，

$Y_i^* = (Y_i - \widehat{\rho} Y_{i-1})$, $X_{1i}^* = (X_{1i} - \widehat{\rho} X_{1,i-1})$, $X_{2i}^* = (X_{2i} - \widehat{\rho} X_{2,i-1})$, $\cdots$, $X_{ki}^* = (X_{ki} - \widehat{\rho} X_{k,i-1})$ を新たな変数として,

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i,$$

に最小二乗法を適用する。

ρ の求め方について (その 2): 収束計算によって求める。 $\longrightarrow$ コクラン・オーカット法

1. $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, 2, \cdots, n$

   を最小二乗法で推定する。 $\longrightarrow$ $\widehat{\beta}_1, \cdots, \widehat{\beta}_k, \widehat{u}_i$ を得る。

2. $\widehat{u}_i = \rho \widehat{u}_{i-1} + \epsilon_i, \quad i = 2, 3, \cdots, n$

   を最小二乗法で推定する。 $\longrightarrow$ $\widehat{\rho}$ を得る。

3. $\rho^{(m-1)} = \widehat{\rho}$ とおく。

4. $Y_i^* = (Y_i - \rho^{(m-1)} Y_{i-1}),\ X_{1i}^* = (X_{1i} - \rho^{(m-1)} X_{1,i-1}),\ X_{2i}^* = (X_{2i} - \rho^{(m-1)} X_{2,i-1}),\ \cdots,$
$X_{ki}^* = (X_{ki} - \rho^{(m-1)} X_{k,i-1})$ を計算する。

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i, \quad i = 2, 3, \cdots, n$$

を最小二乗法で推定する。 $\longrightarrow \widehat{\beta}_1, \cdots, \widehat{\beta}_k$ を得る。

5. $\widehat{u}_i = Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki}, \quad i = 1, 2, \cdots, n$

を計算する。

6. ステップ 2 に戻り，$m = 1, 2, \cdots$ について繰り返す。

収束先を $\beta_1, \beta_2, \cdots, \beta_k, \rho$ の推定値とする。

5　不均一分散 (不等分散)

回帰式が

$$Y_i = \alpha + \beta X_i + u_i$$

の場合を考える。X_i が外生変数，Y_i は内生変数，u_i は互いに独立な同一の分布を持つ攪乱項 (最小二乗法に必要な仮定) とする。「独立な同一の分布」の意味は「攪乱項 $u_1, u_2, \cdots, u_n$ はそれぞれ独立に平均ゼロ，分散 σ^2 の分布する」である。分散が時点に依存する場合，代表的には，分散が他の変数 (例えば，z_i) に依存する場合，すなわち，u_i の平均はゼロ，分散は $\sigma_*^2 z_i^2$ の場合は，最小二乗法の仮定に反する。そのため，単純には，$Y_i = \alpha + \beta X_i + u_i$ に最小二乗法を適用できない。以下のような修正が必要となる。

$$\frac{Y_i}{z_i} = \alpha \frac{1}{z_i} + \beta \frac{X_i}{z_i} + \frac{u_i}{z_i} = \alpha \frac{1}{z_i} + \beta \frac{X_i}{z_i} + u_i^*$$

このとき，新たな攪乱項 u_i^* は平均ゼロ，分散 σ_*^2 の分布となる (すなわち,「同

一の」分布)。

$$\mathrm{E}(u_i^*) = \mathrm{E}\left(\frac{u_i}{z_i}\right) = \left(\frac{1}{z_i}\right)\mathrm{E}(u_i) = 0$$

u_i の仮定 $\mathrm{E}(u_i) = 0$ が使われている。

$$\mathrm{V}(u_i^*) = \mathrm{V}\left(\frac{u_i}{z_i}\right) = \left(\frac{1}{z_i}\right)^2 \mathrm{V}(u_i) = \sigma_*^2$$

u_i の仮定 $\mathrm{V}(u_i) = \sigma_*^2 z_i^2$ が最後に使われている。
よって，$\dfrac{Y_i}{z_i}, \dfrac{1}{z_i}, \dfrac{X_i}{z_i}$ を新たな変数として，最小二乗法を適用することができる。

不均一分散の検定について

$$\widehat{u}_i^2 = \gamma z_i + \epsilon_i$$

を推定し，γ の推定値 $\widehat{\gamma}$ の有意性の検定を行う (通常の t 検定)。

z_i は回帰式に含まれる変数でもよい。例えば，u_i の平均はゼロ，分散は $\sigma_*^2 X_i^2$ の

場合，各変数を X_i で割って，

$$\frac{Y_i}{X_i} = \alpha\frac{1}{X_i} + \beta + \frac{u_i}{X_i} = \alpha\frac{1}{X_i} + \beta + u_i^*$$

を推定すればよい。β は定数項として推定されるが，意味は限界係数 (すなわち，傾き) と同じなので注意すること。

6 Time Series Analysis (時系列分析)

6.1 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. **Ergodicity (エルゴード性) :**

As time difference between two data is large, the two data become independent.

$y_1, y_2, \cdots, y_T$ is said to be ergodic in mean when $\bar{y}$ converges in probability to $\mathrm{E}(y_t)$.

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

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

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

4. **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 $\mathrm{Var}(y_t) = \mathrm{Var}(y_{t-\tau}) = \gamma(0)$.

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

6.2 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: 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: $\mathrm{SARIMA}(p, d, q)$

${}^{s}\Delta y_t = y_t - y_{t-s}$, $s = 4$ for quarterly data $s = 12$ for monthly data

$$ {}^{s}\Delta\Delta^{d} y_t \;\sim\; \mathrm{ARMA}(p, q) \iff y_t \;\sim\; \mathrm{SARIMA}(p, d, q) $$

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

6.3 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.

3. **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 $y_{t-1}, \cdots, y_{t-k+1}$.

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

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

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

$$\vdots$$

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

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

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

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

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,

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

As s is large, ϕ_1^s 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.

(MA will be discussed later.)

4. Mean of AR(1) process, μ

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

5. 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, the autocovariance function of AR(1) process is:

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

The autocorrelation function of AR(1) process is:

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

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:

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

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

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

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

7. 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.

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

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

8. Asymptotic distribution of OLSE $\hat{\phi}_1$:

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

Proof:

$y_{t-1}\epsilon_t$, $t = 1, 2, \cdots, T$, are distributed with mean zero and variance $\dfrac{\sigma_\epsilon^4}{1 - \phi_1^2}$.

From the central limit theorem,

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

Rewriting,

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

Next,

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

yields:

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

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:

$$y \longrightarrow N(\mu, \sigma^2), \qquad x \longrightarrow c.$$

Then, we obtain:

$$xy \longrightarrow N(c\mu, c^2\sigma^2)$$

10. **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:**

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

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

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

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

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

Therefore, the initial conditions are given by:

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

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

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

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

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

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

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

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),

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

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

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

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

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

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

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

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

$$f(y_t|y_{t-1},\cdots,y_1) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}}\exp\Bigl(-\frac{1}{2\sigma_\epsilon^2}(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \ \cdots\ - \phi_p y_{t-p})^2\Bigr)$$

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 the sample variance $\hat{\gamma}(0)$.

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

where

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

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 $k = p + 1, p + 2, \cdots$.

6.4 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:**

$$\begin{aligned}\gamma(0) &= \mathrm{E}(y_t^2) = \mathrm{E}(\epsilon_t + \theta_1 \epsilon_{t-1})^2 = \mathrm{E}(\epsilon_t^2 + 2\theta_1 \epsilon_t \epsilon_{t-1} + \theta_1^2 \epsilon_{t-1}^2) \\ &= \mathrm{E}(\epsilon_t^2) + 2\theta_1 \mathrm{E}(\epsilon_t \epsilon_{t-1}) + \theta_1^2 \mathrm{E}(\epsilon_{t-1}^2) = (1 + \theta_1^2)\sigma_\epsilon^2\end{aligned}$$

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

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

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

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

Let x be $\rho(1)$.

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

θ_1 should be a real number.

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

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

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

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

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

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

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

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

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

where

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

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

Mean of MA(1) Process:

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

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

Taking the expectation,

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

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

1. Autocovariance Function of MA(2) Process:

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

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

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

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

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

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

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

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

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

3. **Likelihood Function:**

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

where

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

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

Mean:

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

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

Therefore,

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

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

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

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

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

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

   where $\theta_0 = 1$.

3. MA(q) process is stationary.

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

Mean:

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

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

Therefore, we have:

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

6.5 ARMA Model

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

1. ARMA(p, q)

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

   which is rewritten as:

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

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

2. Likelihood Function:

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

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

Obtain the autocorrelation coefficient.

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

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

where the second and third terms are zeros.

Therefore, we obtain:

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

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

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

Each term is given by:

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

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

Therefore, we obtain;

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

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

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

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

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

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

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

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

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

We have:

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

ARMA(p, q) +drift:

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

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

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

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

Therefore,

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

6.6 ARIMA Model

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

ARIMA(p, d, q) Process

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

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

例： **ARIMA(0,1,0) Model**

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

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

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

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

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

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

6.7 SARIMA Model

Seasonal ARIMA (SARIMA) Process:

1. SARIMA(p, d, q)

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

   where

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

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

6.8 Optimal Prediction

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

   (a) Define:

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

   where Y_t denotes all the information available at time t.

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

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

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

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

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

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

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

(c) Therefore,

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

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

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

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

(b) Optimal prediction is:

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

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

6.9 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)

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

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

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

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

   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.

6.10 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           |                    |
```

6.11 ARCH and GARCH Models

Autoregressive Conditional Heteroskedasticity (ARCH)

Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

1. ARCH (p) Model

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

   where,

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

   The unconditional variance of ϵ_t is:

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

2. GARCH (p, q) Model

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

where

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

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

sdex	Coef.	OPG 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

7 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(∞)

7.1 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.

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

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.

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

7.2 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

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

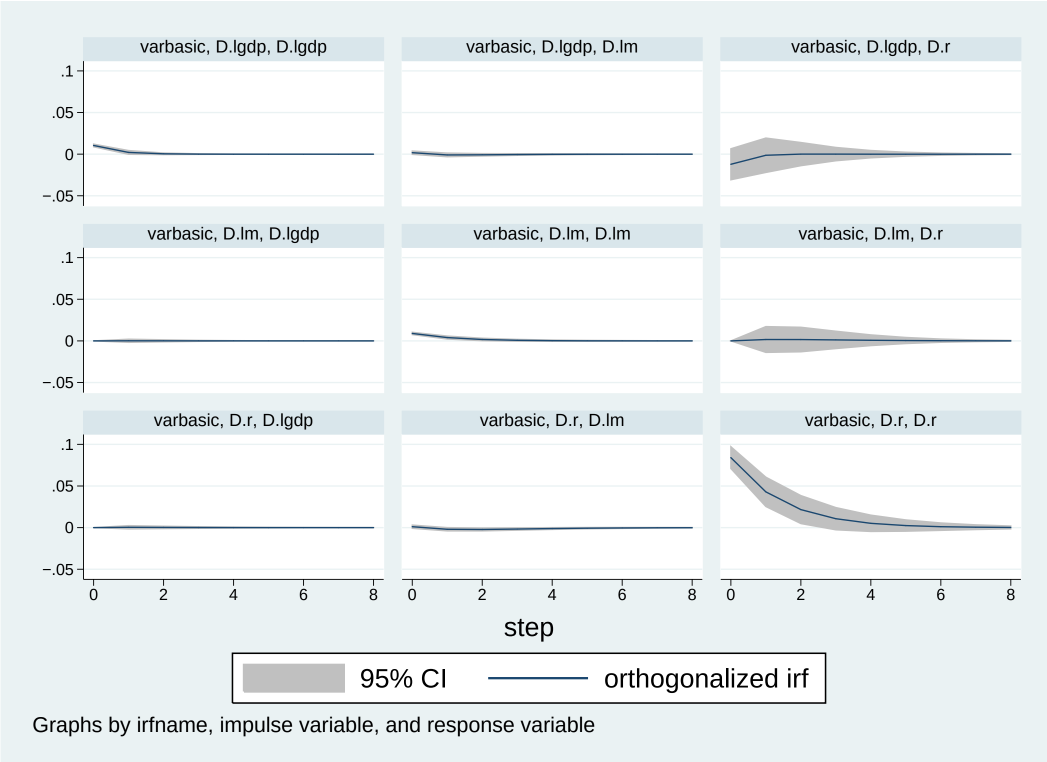
$\Longrightarrow$ **Orthogonalized Impulse Response Function (直交化インパルス応答関数)**

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

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

8.1 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 $\frac{1}{T}X'X$ 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 $|\phi_1| < 1$:**

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

Then, OLSE of ϕ_1 is:

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

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

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

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,

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

which is rewritten as:

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

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

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

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

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

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

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

Is this true?

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

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

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

Therefore, we can obtain:

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

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

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

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

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

Therefore, we obtain:

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

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

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

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

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

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

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

Moreover, the second term is derived from:

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

Therefore,

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

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

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

Thus, we obtain the following result:

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

Therefore,

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

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

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

The distributions of the t statistic: $\dfrac{\hat{\phi}_1 - 1}{s_\phi}$, where s_ϕ denotes the standard error of $\hat{\phi}_1$.

$\Longrightarrow$ 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** $\boldsymbol{\phi_1 < 1}$ **or** $\boldsymbol{-1 < \phi_1}$

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

(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** $\boldsymbol{\phi_1 < 1}$ **or** $\boldsymbol{-1 < \phi_1}$

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** $\boldsymbol{\phi_1 < 1}$ **or** $\boldsymbol{-1 < \phi_1}$

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

8.2 Serially Correlated Errors

Consider the case where the error term is serially correlated.

8.2.1 Augmented Dickey-Fuller (ADF) Test

Consider the following AR(p) model:

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

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:

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

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

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

We can utilize the same tables as before.

Choose p by AIC or SBIC.

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

.
```

8.3 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:

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

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:

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

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.

8.4 Testing Cointegration

8.4.1 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.