ECONOMETRICS NOTE

Junyong Kim[1]

Last Update: December 21st, 2012

Last Update: December 21st, 2012

Last Update: December 21st, 2012

Last Update: December 21st, 2012

Last Update: December 21st, 2012

[1] Kyung Hee University, mindhunter@khu.ac.kr, http://www.junyongkim.com/,

README

2012-12-21 "SAS를이용한계량경제학" 2012년 가을학기 강의에서 할당되었던 8개 과제 첨부를 완료하였습니다.

2012-11-08 "SAS를이용한고급계량경제학" 강의 노트 13장의 추가 작성이 완료되었습니다. 작성자에게 강의 노트 1장, 10장이 없어 현재 비어 있습니다. "금융시계열분석" 강의 노트를 추가하고자 계획하고 있습니다.

2012-09-23 "SAS를이용한계량경제학" 강의 노트 3번 과제 뒷부분에 SAS 프로그래밍을 처음 접하시는 분들에게 유용할 Short Commentary를 첨부하였습니다. File input 및 output, Univariate analysis, Regression analysis 및 Time-series analysis와 간단한 simulation method 예제를 포함하고 있습니다.

2012-09-09 "SAS를이용한계량경제학" 2012년 가을학기 강의에 한하여 강의록마다 할당된 과제(Assignment)를 첨부하고 있습니다.

2012-09-02 경희대 경제학부 이기석 교수님의 "SAS를이용한고급계량경제학" 강의 노트가 추가되고 있습니다. 작성자의 Curriculum Vitae가 추가되었습니다.

2012-08-20 본고는 작성자가 계량경제학 학습의 수월성을 기하여 작성했습니다. 작성 중 경희대 경제학부 이기석 교수님의 "경제통계학", "SAS를이용한계량경제학" 등 수업의 강의 노트 및 경희대 주가예측연구회 자료, 그 외 다수의 통계학, 계량경제학 문서들을 참고했습니다. 결함을 보완하고자 향후 기타 수업의 강의 노트를 추가할 계획입니다.

주가예측연구회 지도교수님이신 이기석 교수님 및 본고 제작에 도움을 주신 남상욱 선배님, 강은경 선배님, 김승기 학우님에게 진심으로 감사의 말씀을 전합니다. 개선 및 정오 등 향후 변동은 Readme에 작성됩니다.

Contents

Statistics

$\boxed{\text{Now Updating, Coming Soon}}$

$\boxed{\text{Now Updating, Coming Soon}}$

$\boxed{\text{Now Updating, Coming Soon}}$

$\boxed{\text{Now Updating, Coming Soon}}$

$\boxed{\text{Now Updating, Coming Soon}}$

Basic Econometrics

Last Update: December 21st, 2012

Last Update: December 21st, 2012

Last Update: December 21st, 2012

Last Update: December 21st, 2012

Last Update: December 21st, 2012

Lecture Note 1

1 Econometrics

Econometrics is *Economic Measurement*, measurement in terms of statistics and mathematics.

Why should we learn this subject?

1 Economic theories may be justified by the data; *Deductive (Theoretical) Approach*

2 A set of data may help find new economic theories; *Inductive (Empirical) Approach*

2 SAS Code

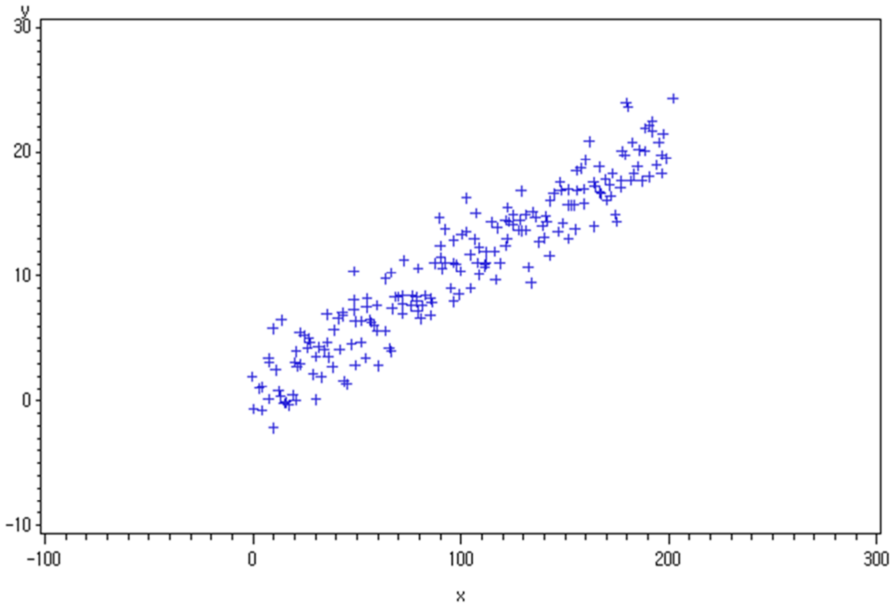
Code 1.2.1 *Generating and plotting the data*

```
data fun;
	seed1=12;
	seed2=14;
	do i=1 to 200;
		j=i;
		x=j+rannor(seed1);
		u=2*rannor(seed2);
		y=1+0.1*x+u;
		output;
	end;
run;

proc gplot data=fun;
	plot y*x;
run;

quit;
```

Figure 1.2.1 *Result of Code 1.2.1*



Code 1.2.2 *Estimating Regression model, Durbin-Watson Test, Plotting*

```
proc reg data=fun;
        model y=x/dw;
run;

proc gplot data=fun;
        plot y*j;
run;

quit;
```

Figure 1.2.2 *Result of Code 1.2.2 (Regression model)*

SAS 시스템

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	200
Number of Observations Used	200

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6899.18515	6899.18515	1731.14	<.0001
Error	198	789.09597	3.98533		
Corrected Total	199	7688.28112			

Root MSE	1.99633	R-Square	0.8974
Dependent Mean	10.87773	Adj R-Sq	0.8968
Coeff Var	18.35244		

Parameter Estimates

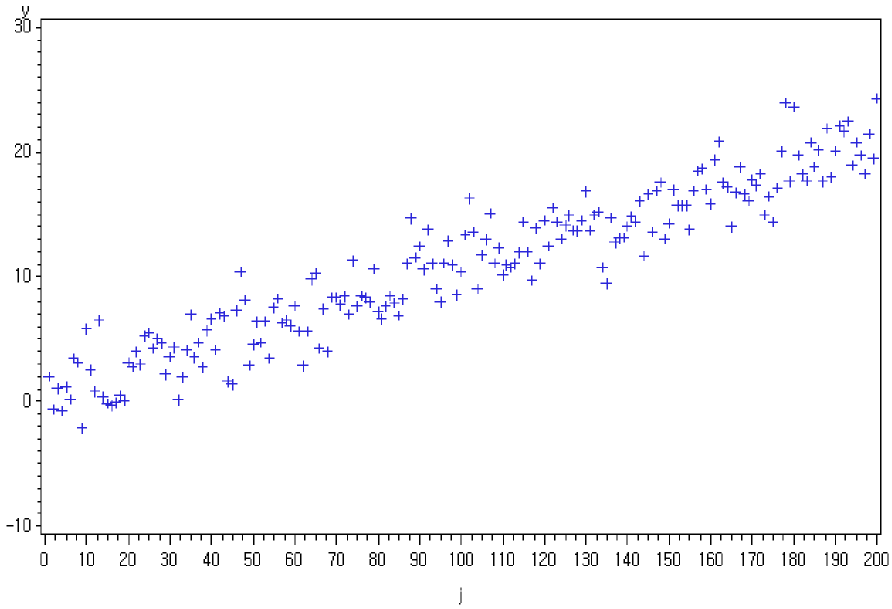
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.64562	0.28356	2.28	0.0239
x	1	0.10178	0.00245	41.61	<.0001

SAS 시스템

The REG Procedure
Model: MODEL1
Dependent Variable: y

Durbin-Watson D	1.731
Number of Observations	200
1st Order Autocorrelation	0.127

Figure 1.2.3 *Result of Code 1.2.2 (Plotting)*



Code 1.2.3 *Generating and plotting the data*

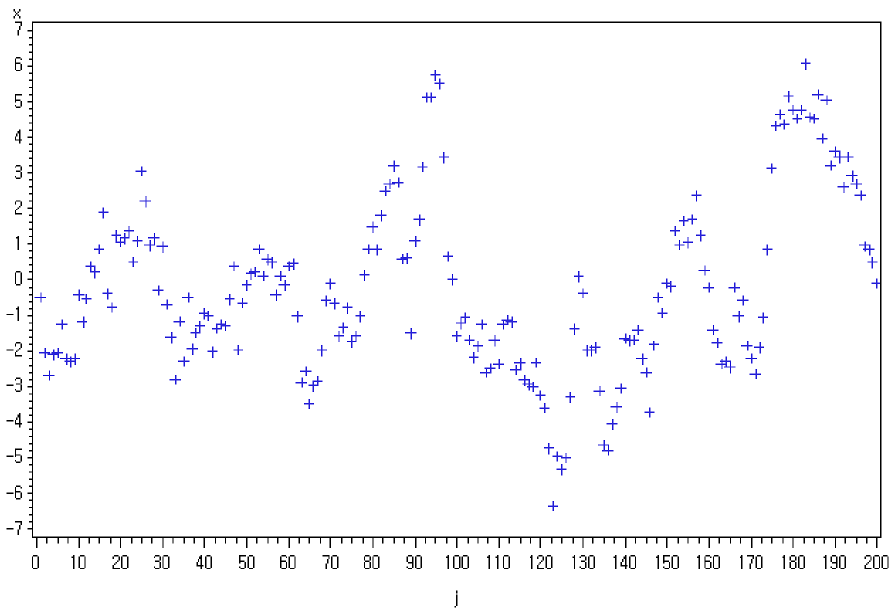
```
data fun;
        seed1=12;
        seed2=1234;
        xlag=rannor(seed1);
        alpha=0.9;
        do i=1 to 200;
                j=i;
                u=rannor(seed2);
                x=alpha*xlag+u;
                output;
                xlag=x;
        end;
run;

proc print data=fun;
        var x xlag;
run;

proc gplot data=fun;
        plot x*j;
run;

quit;
```

Figure 1.2.4 *Result of Code 1.2.3 (Plotting)*



Code 1.2.4 *Regression model estimation, Univariate Analysis*

```
proc reg data=fun corr;
        model x=xlag/dw;
run;

proc univariate data=fun;
        var u x;
run;

quit;
```

Figure 1.2.5 *Result of Code 1.2.4 (Univariate Analysis)*

SAS 시스템

UNIVARIATE 프로시저
변수: u

적률

N	200	가중합	200
평균	-0.0194906	관측치 합	-3.8981133
표준편차	0.98244306	분산	0.96519438
왜도	0.00105061	첨도	-0.2463149
제곱합	192.149657	수정 제곱합	192.073681
변동계수	-5040.6081	평균의 표준오차	0.06946922

기본 통계 측도

위치측도		변이측도	
평균	-0.01949	표준편차	0.98244
중위수	-0.00324	분산	0.96519
최빈값	.	범위	4.86517
		사분위 범위	1.33986

위치모수 검정: Mu0=0

검정	--통계량---		-------p-값-------	
스튜던트의 t	t	-0.28056	Pr > |t|	0.7793
부호	M	-1	Pr >= |M|	0.9437
부호 순위	S	-175	Pr >= |S|	0.8315

ASSIGNMENT 1

Exercise 1.1 (PROC UNIVARIATE) *Consider the following model*

$$Y_t = \mu + u_t$$

Where $\mu = 10.0$ and u_t follows $\mathcal{N}(0, 1^2)$ distribution.

(1) Generate 200 numbers of observations of Y_t and estimate μ using the sample average

$$\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$$

And denote it as m_1; use SAS command *PROC UNIVARIATE* to obtain the sample average.

(2) Generate 500 numbers of observations of Y_t and estimate μ using $\bar{Y}$ and denote it as m_2; compare the standard errors of m_1 and m_2. Is there any considerable difference?

(3) Generate 1000 numbers of observations of Y_t and estimate μ using $\bar{Y}$ and denote it as m_3; compare the standard errors of m_1, m_2 and m_3. Can you find any insight from exercises?

(Hint: Generate the data with SEED=12)

Exercise 1.2 (PROC REG) *Suppose the econometric model of the consumption function can be described as*

$$Y_t = \alpha + \beta X_t + u_t \tag{1.1}$$

Where

Y_t : Monthly log-growth rate of Industrial Production (IP)

X_t : Monthly log-return on S&P 500 stock index (SP500)

u_t : Error term of the model (1.1)

(1) Obtain the estimates of α and β

(2) Obtain the standard errors of estimates α and β

(3) Obtain the residuals e_t and plot them regarding to its time or its index number.

(4) Can you find any systematic pattern from the plot of extracted residuals?

(5) Obtain the predicted values of Y_t and plot them.

(6) Try six regression models as (1.1) by substituting X_t by $X_{t-1}, X_{t-2}, \cdots, X_{t-6}$ and construct a table in your notebook for the estimates, standard errors, and R^2 of a and b.

(7) Find the model that gives the highest R^2 value; i.e. which one out of $X_{t-1}, X_{t-2}, \cdots, X_{t-6}$ renders the highest R^2 value?

(8) Suggest an economic interpretation on the finding in (7) above.

Exercise 1.3 (PROC REG with the residuals) *Suppose below regression model*

$$e_t = \phi e_{t-1} + \varepsilon_t \tag{1.2}$$

Where

e_t : Estimated u_t from the regression model (1.1)

ε_t : Error term of the model (1.2)

(1) Obtain the estimate and its *t*-value and *p*-value of ϕ; is the *p*-value smaller than 0.05?

(2) What should be the utility of above regression?

(3) Can you suggest any variation of above regression?

(Hint: Use PROC REG with NOINT option)

Exercise 1.4 (PROC REG with actual data) *Suppose below regression model*

$$Y_t = \alpha + \beta X_{t-1} + u_t \tag{1.3}$$

Where

Y_t : Nightly log-return on KOSPI index (KPD); i.e. $\log(P_t^O) - \log(P_{t-1}^C)$

P_t^O : Opening value of KOSPI index (KPDO) observed the time t

P_{t-1}^C : Closing value of KOSPI index (KPDC) observed the time $(t-1)$

X_{t-1} : Daily log-return on S&P 500 index (SP500D) on the time $(t-1)$

(1) What is the estimate of β and its *t*-value?

(2) What can be inferred from above regression model?

(3) Do you think you can make money by knowing that the X_{t-1} affects Y_t?

Code Assignment 1

```
/**************************************************
Exercise 1.1
**************************************************/

data exercise11;
	mu=10;
	seed=12;
	do t=1 to 200;
		u=rannor(seed);
		y=mu+u;
		output;
	end;
run;

proc univariate data=exercise11;
	var y;
run;

data exercise11;
	mu=10;
	seed=12;
	do t=1 to 500;
		u=rannor(seed);
		y=mu+u;
		output;
	end;
run;

proc univariate data=exercise11;
	var y;
run;

data exercise11;
	mu=10;
	seed=12;
	do t=1 to 1000;
		u=rannor(seed);
		y=mu+u;
		output;
	end;
run;

proc univariate data=exercise11;
	var y;
run;

/**************************************************
Exercise 1.2
**************************************************/
```

```
data ip;
        infile "c:₩ip.prn";
        input month ip;
        ipg=dif(log(ip))*100;
run;

data sp500;
        infile "c:₩sp500.prn";
        input month sp500;
        sp500g=dif(log(sp500))*100;
run;

data exercise12;
        merge ip sp500;
        by month;
run;

proc reg data=exercise12;
        model ipg=sp500g;
        output out=exercise13 r=resid;
run;

data exercise13;
        set exercise13;
        obs=_n_;
run;

proc gplot data=exercise13;
        symbol i=join;
        plot resid*obs;
run;

data exercise12;
        set exercise12;
        sp500g1=lag1(sp500g);
        sp500g2=lag2(sp500g);
        sp500g3=lag3(sp500g);
        sp500g4=lag4(sp500g);
        sp500g5=lag5(sp500g);
        sp500g6=lag6(sp500g);
run;

proc reg data=exercise12;
        model ipg=sp500g1;
        model ipg=sp500g2;
        model ipg=sp500g3;
        model ipg=sp500g4;
        model ipg=sp500g5;
        model ipg=sp500g6;
run;
```

```
/**************************************************
Exercise 1.3
**************************************************/

data exercise13;
        set exercise13;
        resid1=lag1(resid);
run;

proc reg data=exercise13;
        model resid=resid1/noint;
run;

/**************************************************
Exercise 1.4
**************************************************/

data kpd;
        infile "c:₩kpd.prn";
        input date kpdo kpdc;
        kpdc1=lag1(kpdc);
        nkpdg=log(kpdo/kpdc1)*100;
run;

data sp500d;
        infile "c:₩sp500d.prn";
        input date sp500d;
        sp500dg=dif(log(sp500d))*100;
        sp500dg1=lag1(sp500dg);
run;

data exercise14;
        merge kpd sp500d;
        by date;
run;

proc reg data=exercise14;
        model nkpdg=sp500dg1;
run;

quit;
```

Lecture Note 2

1 Regression Model

1.1 Two Variable Regression Model

Definition *Simple Regression Model; suppose we have a regression model (2.1)*

$$\begin{aligned} Y_i &= f(X_i) \\ &= \mathbb{E}(Y_i|X_i) \\ &= \alpha + \beta X_i + u_i \end{aligned} \tag{2.1}$$

Where

Y_i : Dependent variable

X_i : Independent variable

u_i : Error term ($= \varepsilon_i$, similarly)

α, β : Unknown parameters

We think God has (2.1) population regression model between Y_i and X_i above. For this, we can think about below (2.2) sample regression model.

$$Y_i = a + bX_i + e_i \tag{2.2}$$

Where

e_i : Residual term ($= \hat{\varepsilon}_i$, similarly)

a, b : Sample parameters ($= \hat{\alpha}, \hat{\beta}$, see below)

Model (2.2) is often expressed as below (2.3), which contains Greeks alphabet, but intuitively same as (2.2), totally.

$$Y_i = \hat{\alpha} + \hat{\beta}X_i + \hat{u}_i \tag{2.3}$$

1.2 Criteria of Parameter Estimation

Criteria 1 *The sum of residuals equals zero.*

$$
\begin{aligned}
Y_i &= a + bX_i + e_i \\
e_i &= Y_i - \mathbb{E}(Y_i) \\
&= Y_i - \hat{Y}_i \\
&= Y_i - (a + bX_i) \\
&= Y_i - a - bX_i \\
\sum_{i=1}^{n} e_i &= \sum_{i=1}^{n} (Y_i - a - bX_i)
\end{aligned}
$$

Guessing about α and β amounts to find a straight line since once a line in found we obtain its intercept a and the slope b above.

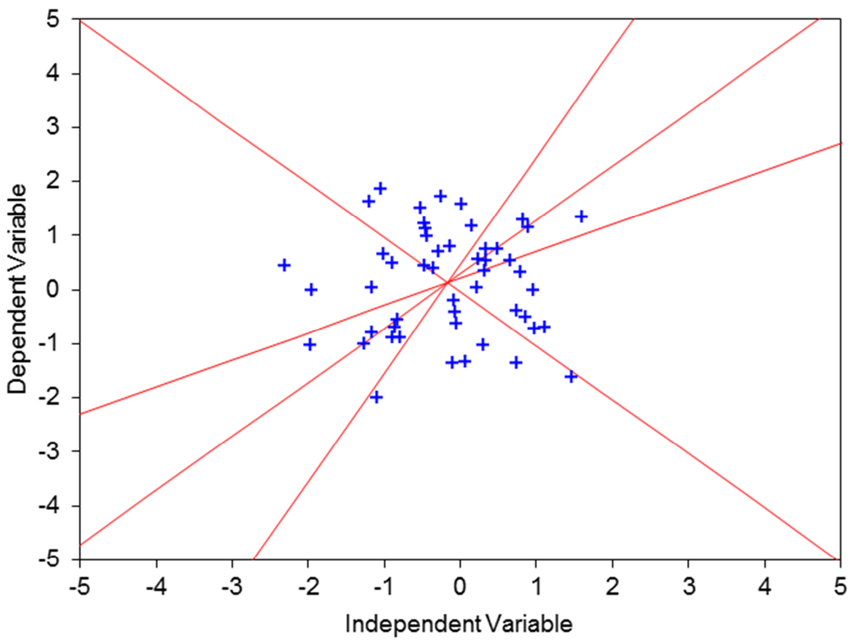
$$
\begin{aligned}
\hat{Y}_i &= \text{Estimated (predicted) } Y_i \\
\sum_{i=1}^{n} e_i &= 0 \\
&= \sum_{i=1}^{n} (Y_i - a - bX_i) \\
&= \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bX_i \\
&= n\bar{Y} - na - nb\bar{X} \\
&= \bar{Y} - a - b\bar{X}
\end{aligned}
$$

Or

$$\bar{Y} = a + b\bar{X} \tag{2.4}$$

Statement (2.4) implies the regression line passes through $(\bar{X}, \bar{Y})$ coordinate.

Figure 2.1.1 *Regression line with the statement (2.4)*



Criteria 2 *Minimize the sum of absolute deviations (residuals), i.e.*

$$\min_{\tilde{a},\tilde{b}} \sum_{i=1}^{n} |e_i|$$

This criterion gives a single line, i.e. single guessed values of $\bar{a}$ and $\bar{b}$ and they are called *Minimum Absolute Deviation Estimators*. However, the disadvantage of this criterion is that the mathematical expression of $\tilde{a}$ (which is called as a-tilde)[2] and $\tilde{b}$ (b-tilde) is too complicated.

Criteria 3 (Ordinary Least Squares; OLS) *Find a line in such a way that the sum of squared residuals is minimized.*

$$\min_{\tilde{a},\tilde{b}} \sum_{i=1}^{n} e_i^2 \tag{2.5}$$

Statement (2.5) implies that there exists only single line and it can be expanded as the statement (2.6) below; with the parameter (a, b) and the variable (X_i, Y_i) form.

[2] Tilde expression is adjusted for the term which is not constantly settled yet.

$$\min_{\tilde{a},\tilde{b}} \sum_{i=1}^{n} e_i^2 \quad = \quad \min_{\tilde{a},\tilde{b}} \sum_{i=1}^{n} \left(Y_i - \tilde{a} - \tilde{b}X_i\right)^2 \tag{2.6}$$

1.3 First Order Condition

Condition (First Order Condition; FOC) *Note that the function* $y = f(x)$ *is maximized where* $f'(x) = 0$ *and* $f''(x) < 0$*, so*

$$\begin{aligned}
\frac{\partial}{\partial \tilde{a}} \sum_{i=1}^{n} e_i^2 &= \frac{\partial}{\partial \tilde{a}} \sum_{i=1}^{n} \left(Y_i - \tilde{a} - \tilde{b}X_i\right)^2 \\
&= (-2) \sum_{i=1}^{n} (Y_i - a - bX_i) \\
&= 0 \qquad (2.7) \\
\frac{\partial^2}{\partial \tilde{a}^2} \sum_{i=1}^{n} e_i^2 &= \frac{\partial^2}{\partial \tilde{a}^2} \sum_{i=1}^{n} \left(Y_i - \tilde{a} - \tilde{b}X_i\right)^2 \\
&= \frac{\partial}{\partial \tilde{a}} (-2) \sum_{i=1}^{n} \left(Y_i - \tilde{a} - \tilde{b}X_i\right) \\
&= -1
\end{aligned}$$

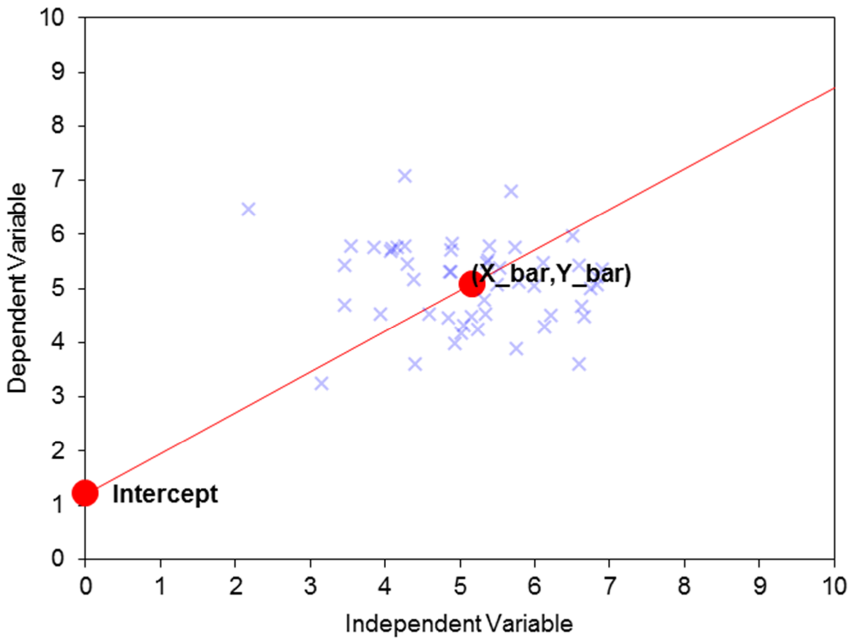
Thus, (2.5) is maximized at (2.7). And

$$\begin{aligned}
(-2) \sum_{i=1}^{n} \left(Y_i - \tilde{a} - \tilde{b}X_i\right) &= 0 \\
&= \sum_{i=1}^{n} (Y_i - a - bX_i) \\
&= \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bX_i \\
&= n\bar{Y} - na - nb\bar{X} \\
&= \bar{Y} - a - b\bar{X} \\
\bar{Y} &= a + b\bar{X} \qquad (2.8)
\end{aligned}$$

Note that $\bar{Y} = \sum Y_i / n$ so $n\bar{Y} = \sum Y_i$

The regression found by applying criterion 3 (OLS criterion) should pass through the point $(\bar{X},\bar{Y})$ coordinate. If the true regression has an intercept α

Figure 2.1.2 *Implication of the statement (2.8)*



From FOC we can derive two simultaneous equations through two partial derivatives.

$$\frac{\partial}{\partial \tilde{a}}\sum_{i=1}^{n} e_i^2 \quad , \quad \frac{\partial}{\partial \tilde{b}}\sum_{i=1}^{n} e_i^2$$

The first normal equation is already expressed above, and the second normal equation can be derived from another partial derivative.

$$\begin{aligned}\frac{\partial}{\partial \tilde{b}}\sum_{i=1}^{n} e_i^2 &= \frac{\partial}{\partial \tilde{b}}\sum_{i=1}^{n}\left(Y_i-\tilde{a}-\tilde{b}X_i\right)^2 \\ &= \sum_{i=1}^{n}(-2X_i)(Y_i-a-bX_i) \\ &= 0 \end{aligned} \tag{2.9}$$

$$\frac{\partial^2}{\partial \tilde{b}^2}\sum_{i=1}^{n} e_i^2 = \frac{\partial^2}{\partial \tilde{b}^2}\sum_{i=1}^{n}\left(Y_i-\tilde{a}-\tilde{b}X_i\right)^2$$

$$
\begin{aligned}
&= \frac{\partial}{\partial \tilde{b}} \sum_{i=1}^{n} (-2X_i)(Y_i - a - bX_i) \\
&= \sum_{i=1}^{n} 2X_i^2 \qquad (2.10)
\end{aligned}
$$

Since the variable X_i is real, (2.10) is always non-negative. Thus (2.5) is maximized at (2.9) also, which is another pillar of simultaneous equation.

Thus, two of simultaneous equations can be arranged as

$$
\begin{aligned}
\sum_{i=1}^{n} (Y_i - a - bX_i) &= 0 \\
\sum_{i=1}^{n} X_i(Y_i - a - bX_i) &= 0
\end{aligned}
$$

Those two equations can be transformed as below through distributing.

$$
\begin{aligned}
\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bX_i &= 0 \\
\sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} aX_i - \sum_{i=1}^{n} bX_i^2 &= 0 \\
\sum_{i=1}^{n} X_iY_i - na\bar{X} - b\sum_{i=1}^{n} X_i^2 &= \qquad (2.11)
\end{aligned}
$$

To estimate those two parameters, especially b, below equation should be arranged by substitution of a. By substituting a in (2.11) by (2.8) we obtain

$$
\begin{aligned}
\sum_{i=1}^{n} X_iY_i - n(\bar{Y} - b\bar{X})\bar{X} - b\sum_{i=1}^{n} X_i^2 &= 0 \\
\sum_{i=1}^{n} X_iY_i - n\bar{X}\bar{Y} + bn\bar{X}^2 - b\sum_{i=1}^{n} X_i^2 &= \\
b\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) &= \sum_{i=1}^{n} X_iY_i - n\bar{X}\bar{Y} \\
b &= \sum_{i=1}^{n} X_iY_i - n\bar{X}\bar{Y} \Bigg/ \sum_{i=1}^{n} X_i^2 - n\bar{X}^2
\end{aligned}
$$

Therefore, two OLS estimators are

$$
a = \hat{\alpha}
$$

$$
\begin{aligned}
&= \bar{Y} - b\bar{X} \\
b &= \hat{\beta} \\
&= \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2}
\end{aligned}
$$

Cf. (Estimator and Estimate) In Econometrics, *estimator* is a formula, which is algebraically abstract. In contrast, *estimate* is a number computed through estimator.

Ex. (Estimator) $a = \bar{Y} - b\bar{X}$, $b = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2}$

Ex. (Estimate) $a = 0.0731$, $b = 1.4523$

Above two estimators can be rewritten as below. To simplify those expressions, we can adopt another notation, by which the deviation of each variable can be expressed easily.

$$
\begin{aligned}
x_i &= X_i - \bar{X} \\
y_i &= Y_i - \bar{Y} \\
b &= \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2 \qquad (2.12)
\end{aligned}
$$

Above b also can be expressed as basic Statistics term as

$$
\begin{aligned}
b &= \frac{1}{n-1} \times \left(\sum_{i=1}^{n} x_i y_i \right) \Big/ \frac{1}{n-1} \times \left(\sum_{i=1}^{n} x_i^2 \right) \\
&= \mathbb{c}ov(X_i, Y_i) / \mathbb{v}ar(X_i) \\
&= s_{XY} / s_X^2
\end{aligned}
$$

2 SAS Code

Use above codes 1.2.3 and 1.2.4 and watch those results.

Figure 2.2.1 *Result of Code 1.2.4 (Regression Analysis)*

SAS 시스템

The REG Procedure
Model: MODEL1
Dependent Variable: x

Number of Observations Read	200
Number of Observations Used	200

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1008.05820	1008.05820	1040.90	<.0001
Error	198	191.75304	0.96845		
Corrected Total	199	1199.81124			

Root MSE	0.98410	R-Square	0.8402
Dependent Mean	-0.24077	Adj R-Sq	0.8394
Coeff Var	-408.72994		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.01547	0.06994	-0.22	0.8251
xlag	1	0.91634	0.02840	32.26	<.0001

Figure 2.2.2 *Result of Code 1.2.4 (Durbin-Watson Test)*

SAS 시스템

The REG Procedure
Model: MODEL1
Dependent Variable: x

Durbin-Watson D	1.918
Number of Observations	200
1st Order Autocorrelation	0.040

ASSIGNMENT 2

Exercise 2.1 (PROC REG) *Suppose the econometric model of the consumption function can be described as*

$$Y_t = \alpha + \beta X_t + u_t \tag{2.1}$$

Where

Y_t : Monthly log-growth rate of Industrial Production (IP)

X_t : Monthly log-return on S&P 500 stock index (SP500)

u_t : Error term of the model (2.1)

(1) Obtain the predicted values of Y_t and plot them.

(2) Try six regression models as (2.1) by substituting X_t by $X_{t-1}, X_{t-2}, \cdots, X_{t-6}$ and construct a table in your notebook for the estimates, standard errors, and R^2 of a and b.

(3) Find the model that gives the highest R^2 value; i.e. which one out of $X_{t-1}, X_{t-2}, \cdots, X_{t-6}$ renders the highest R^2 value?

(4) Suggest an economic interpretation on the finding in (1) above.

Exercise 2.2 (PROC REG with the residuals) *Suppose below regression model*

$$e_t = \phi e_{t-1} + \varepsilon_t \tag{2.2}$$

Where

e_t : Estimated u_t from the regression model (2.1)

ε_t : Error term of the model (2.2)

(1) Obtain the estimate and its *t*-value and *p*-value of ϕ; is the *p*-value smaller than 0.05?

(2) What should be the utility of above regression?

(3) Can you suggest any variation of above regression?

(4) Try same process for the model (2.3) and compare those results from (2.2) and (2.3)

$$e_t = \gamma + \phi e_{t-1} + \delta X_t + \varepsilon_t \tag{2.3}$$

Cf. You should know that all models below means exactly same; the regression model consists of dependent variable Y_t, independent variable X_t, OLS intercept estimator (usually α or β_0), OLS slope estimator (usually β or β_1), and the residual as a result of the model (usually u_t)

For instance, there is no difference between (2.4) and (2.5)

$$Y_t = \alpha + \beta X_t + u_t \tag{2.4}$$

$$Y_t = \beta_0 + \beta_1 X_t + u_t \tag{2.5}$$

However, (2.6) means something different; (2.4) and (2.5) is *Population Regression Function*, but (2.6) is *Sample Regression Function*; If you want to study in detail, refer to 2.2 and 2.6 of Chapter 2, Gujarati, *Basic Econometrics*.

$$Y_t = a + bX_t + e_t \tag{2.6}$$

Below expression also means Sample Regression Function.

$$Y_t = \hat{\alpha} + \hat{\beta} X_t + \hat{u}_t$$

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_t + \hat{u}_t$$

Figure Assignment 2 *Population Parameter and Sample Estimator*

Population Parameter		**Sample Estimator**
μ, σ^2, α (or β_0), β (or β_1)	$\Leftrightarrow$	$\bar{X}$, s^2, a (or $\hat{\alpha}, \hat{\beta}_0$), b (or $\hat{\beta}, \hat{\beta}_1$)
Which is unobservable, but we really want to know because they are helpful information to understand various economic phenomena.		We consider this as a proxy of Population Parameter. Sample Estimator can be obtained from the sample and proper reasoning.

LECTURE NOTE 3

1 Random Property of *b*

OLS estimators of regression model is

$$\begin{aligned} a &= \bar{Y} - b\bar{X} \\ b &= \frac{\sum x_i y_i}{\sum x_i^2} \end{aligned}$$

We can transform the regression model (2.1) through differencing.

$$\begin{aligned} Y_i &= \alpha + \beta X_i + u_i \\ \bar{Y} &= \alpha + \beta \bar{X} + \bar{u} \\ Y_i - \bar{Y} &= \alpha - \alpha + \beta X_i - \beta \bar{X} + u_i - \bar{u} \\ &= \beta(X_i - \bar{X}) + u_i - \bar{u} \\ y_i &= \beta x_i + u_i - \bar{u} \end{aligned}$$

Note that

$$\sum_{i=1}^{n} e_i = 0 \quad \rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} e_i = \bar{e} = 0$$

If we substitute the above into the estimator b equation (2.12), we obtain

$$\begin{aligned} b &= \frac{\sum x_i[\beta x_i + (u_i - \bar{u})]}{\sum x_i^2} \\ &= \frac{\sum(\beta x_i^2 + x_i u_i - \bar{u} x_i)}{\sum x_i^2} \\ &= \frac{\beta \sum x_i^2 + \sum x_i u_i - \bar{u} \sum x_i}{\sum x_i^2} \end{aligned}$$

$$= \quad \beta + \frac{\sum x_i u_i}{\sum x_i^2} - \frac{\bar{u}\sum x_i}{\sum x_i^2}$$

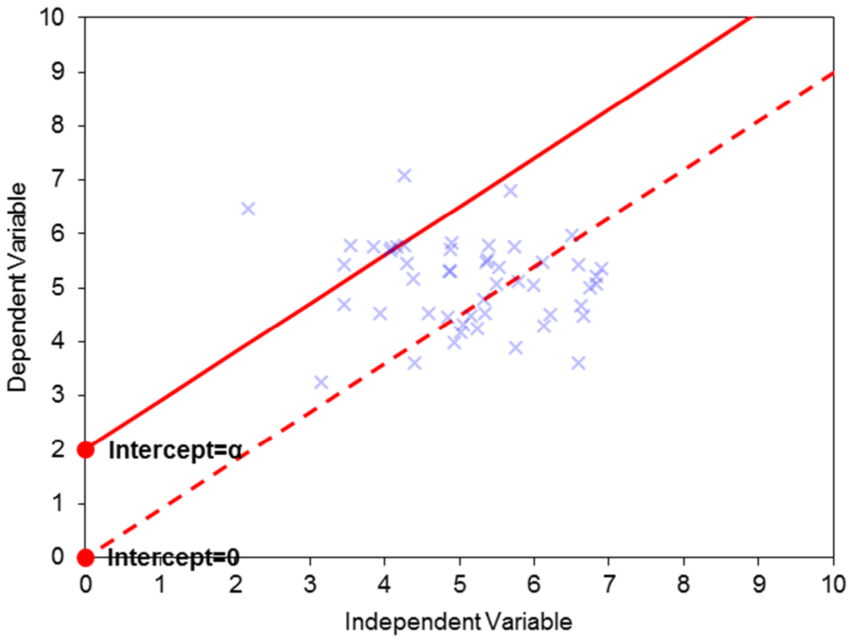
Note that

$$\sum_{i=1}^{n} x_i \quad = \quad \sum_{i=1}^{n}(X_i - \bar{X}) \quad = \quad \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X}$$

$$= \quad n\bar{X} - n\bar{X}$$

$$= \quad 0$$

Figure 3.1.1 *Regression line with zero-intercept (β is fixed number)*



Therefore, b can be rewritten as

$$b \quad = \quad \beta + \frac{\sum x_i u_i}{\sum x_i^2}$$

$$= \quad \beta + \frac{\left(\frac{1}{n-1}\right)\sum x_i u_i}{\left(\frac{1}{n-1}\right)\sum x_i^2}$$

$$= \quad \beta + \frac{\mathbb{c}ov(X_i, u_i)}{\mathbb{v}ar(X_i)}$$

$$= \quad \beta + \frac{s_{Xu}}{s_X^2}$$

Since the error u_i of right hand side of above equation is random variable, the OLS estimator b is also random variable, which can be fluctuated with u_i's movement.

2 Classical Assumption

Assumption (Classical Assumption) *For the regression model* $Y_t = \alpha + \beta X_t + u_t$,

(1) The independent variable is fixed non-random.

(2) $\forall t,\ \mathbb{E}(u_t) = 0$

(3.a) $\forall t,\ \mathbb{E}(u_t^2) = \sigma^2$

(3.b) $\forall t \neq s,\ \mathbb{E}(u_t u_s) = 0$

2.1 Unbiasedness of *b*

$$\begin{aligned}
Y_i &= \alpha + \beta X_i + u_i \\
u_i &: \text{Coin throw} \\
&= \begin{cases} +1, & \text{if top} \\ -1, & \text{if bottom} \end{cases} \\
\mathbb{E}(u_i) &= \underbrace{\left(\frac{1}{2}\right)}_{P_1} \times (+1) + \underbrace{\left(\frac{1}{2}\right)}_{P_2} \times (-1) \\
P_1 &: \text{Weight; probability of obtaining } +1 \\
P_2 &: \text{Weight; probability of obtaining -1} \\
\mathbb{E}(b) &= \mathbb{E}\left(\beta + \frac{\sum x_i u_i}{\sum x_i^2}\right) \\
&= \mathbb{E}\left(\beta + \frac{\sum x_i u_i}{\sum x_i^2}\right) \\
&= \mathbb{E}(\beta) + \mathbb{E}\left(\frac{\sum x_i u_i}{\sum x_i^2}\right) \qquad (3.1)
\end{aligned}$$

Classical Assumption (1) can be applied for above (3.1). Thus,

$$\mathbb{E}(b) \quad = \quad \beta + \frac{\sum x_i \mathbb{E}(u_i)}{\sum x_i^2} \tag{3.2}$$

Similarly, Classical Assumption (2) can be applied for (3.2), by which

$$\mathbb{E}(b) \quad = \quad \beta \tag{3.3}$$

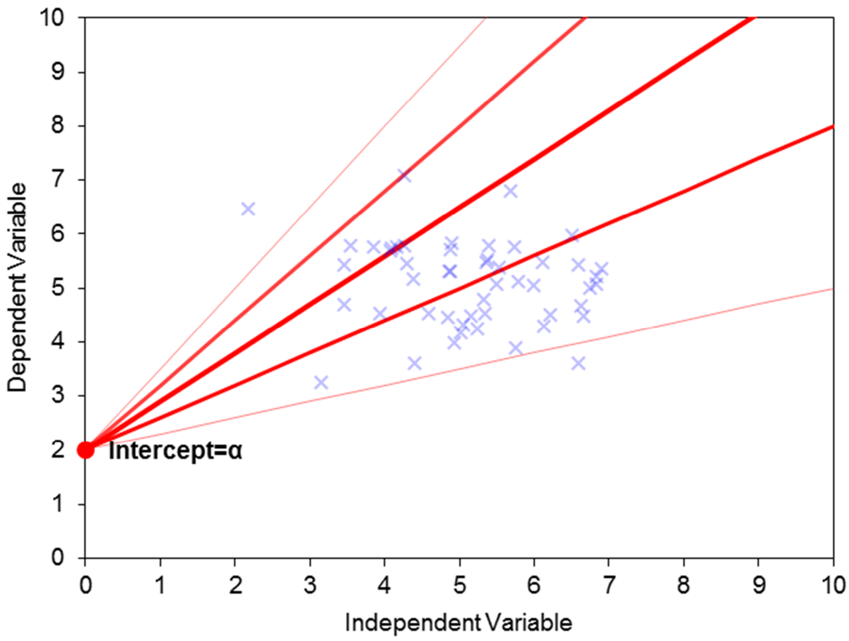
To prove (3.3), Classical Assumption (1) and (2) is used. Thus, b is *unbiased*; expected value of the estimator is exactly same as the parameter.

Cf. (Unbiasedness) The estimator $\hat{\theta}$ is unbiased since $\mathbb{E}(\hat{\theta}) = \theta$

Ex. (Unbiasedness) $b_1 = 0.52$, $b_2 = 0.43$, $b_3 = 0.47$, $\cdots$, $\lim_{n\to\infty} \frac{1}{n} \sum b_i = \beta$

2.2 Variance of b

Figure 3.1.2 *How much b varies?*



$$\mathbb{var}(b) \quad = \quad \mathbb{E}[b - \mathbb{E}(b)]^2$$

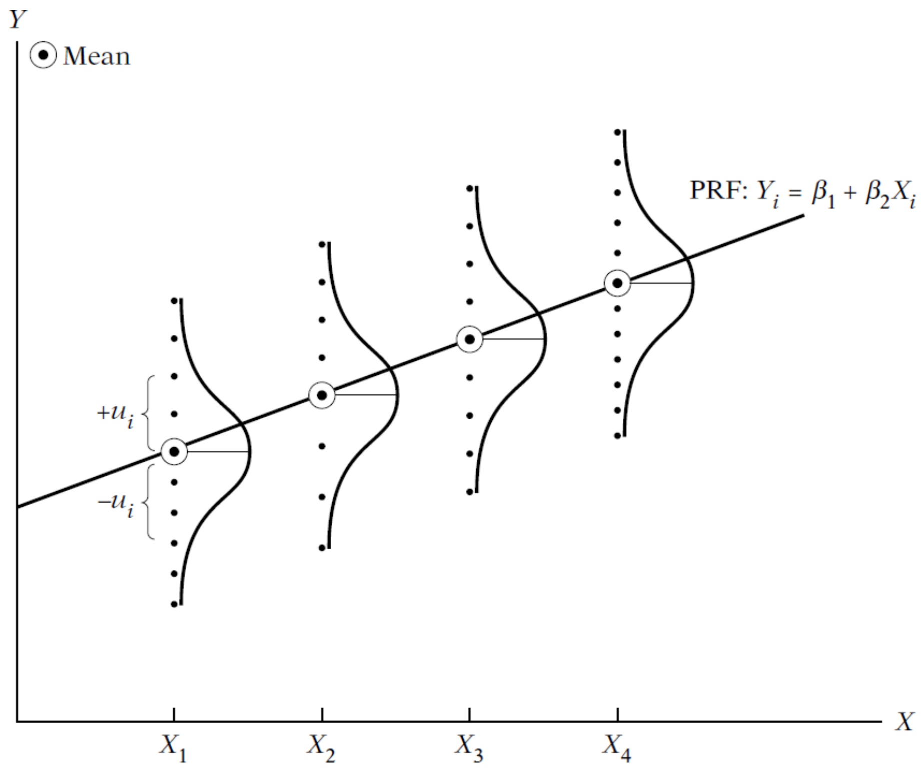
Classical Assumption (1) and (2) can be applied for the unbiasedness of b and hence

$$\mathbb{var}(b) \quad = \quad \mathbb{E}[b - \beta]^2$$

$$
\begin{aligned}
&= \mathbb{E}\left[\left(\beta+\frac{\sum x_i u_i}{\sum x_i^2}\right)-\beta\right]^2 \\
&= \mathbb{E}\left(\frac{\sum x_i u_i}{\sum x_i^2}\right)^2 \\
&= \mathbb{E}\left[\frac{(\sum x_i u_i)^2}{(\sum x_i^2)^2}\right] \leftarrow \text{Classical Assumption (1)} \\
&= \frac{1}{(\sum x_i^2)^2}\mathbb{E}\left[\left(\sum_{i=1}^{n} x_i u_i\right)^2\right] \\
&= \frac{1}{(\sum x_i^2)^2}\mathbb{E}\left[\sum_{i=1}^{n} x_i^2 u_i^2 + 2\sum_{i=1}^{n}\sum_{j>i}^{n} x_i x_j u_i u_j\right] \leftarrow \text{Classical Assumption (1)} \\
&= \frac{1}{(\sum x_i^2)^2}\sum_{i=1}^{n} x_i^2 \underbrace{\mathbb{E}(u_i^2)}_{=\sigma^2} + 2\sum_{i=1}^{n}\sum_{j>i}^{n} x_i x_j \underbrace{\mathbb{E}(u_i u_j)}_{=0} \\
&= \frac{1}{(\sum x_i^2)^2}\sum_{i=1}^{n} x_i^2 \sigma^2 \\
&= \frac{\sigma^2}{\sum x_i^2}
\end{aligned}
$$

Cf. $\bar{e}=0,\ \bar{u}\neq 0$

Figure 3.1.3 *Conditional distribution of u_i; the disturbances*[3]



3 Gauss-Markov Theorem

Theorem (Gauss-Markov Theorem) *Under Classical Assumption, OLS estimator b is BLUE; Best Linear Unbiased Estimator.*

Proof *To the contrary, suppose there is an estimator $\tilde{b}$ whose variance is smaller than variance of OLS estimator, and let*

$$\tilde{b} \quad = \quad b + \sum_{t=1}^{n} C_t Y_t$$

Where C_t $(t = 1,2,3,\cdots,n)$ are some fixed constants. Note that $b = \frac{1}{\sum x_t^2} \sum x_t y_t$. Thus, b is a linear estimator with respect to y_t.

$$\mathbb{E}(\tilde{b}) \quad = \quad \mathbb{E}\left(b + \sum C_t Y_t\right)$$

[3] Damodar N. Gujarati, 2004, *Basic Econometrics*

$$= \mathbb{E}\left[b + \sum C_t(\alpha + \beta X_t + u_t)\right]$$

$$= \mathbb{E}(b) + \mathbb{E}\left[\sum C_t(\alpha + \beta X_t + u_t)\right]$$

$$= \beta + \mathbb{E}\left(\sum C_t\alpha + \sum C_t\beta X_t + \sum C_t u_t\right)$$

$$= \beta + \mathbb{E}\left(\alpha \sum C_t + \beta \sum C_t X_t + \sum C_t u_t\right)$$

According to Classical Assumption (1) and above assumption

$$\mathbb{E}(\tilde{b}) = \beta + \alpha \sum C_t + \beta \sum C_t X_t + \sum C_t \mathbb{E}(u_t)$$

According to Classical Assumption (2)

$$\mathbb{E}(\tilde{b}) = \beta + \alpha \sum C_t + \beta \sum C_t X_t$$

In order for $\tilde{b}$ to be unbiased, we need

$$\sum C_t = 0 \quad \text{and} \quad \sum C_t X_t = 0 \tag{3.4}$$

Then, we gain

$$\mathbb{E}(\tilde{b}) = \beta$$

Therefore, as a consequence

$$\tilde{b} = b + \sum C_t Y_t$$

$$= b + \sum C_t(\alpha + \beta X_t + u_t)$$

$$= b + \sum C_t u_t$$

$$\mathbb{var}(\tilde{b}) = \mathbb{E}[\tilde{b} - \mathbb{E}(\tilde{b})]^2$$

$$= \mathbb{E}\left[b + \sum C_t u_t - \mathbb{E}\left(b + \sum C_t u_t\right)\right]^2$$

$$= \mathbb{E}\left[b + \sum C_t u_t - \beta\right]^2$$

$$= \mathbb{E}\left[(b - \beta) + \sum C_t u_t\right]^2$$

$$= \mathbb{E}\left[(b - \beta)^2 + \left(\sum C_t u_t\right)^2 + 2(b - \beta)\left(\sum C_t u_t\right)\right]$$

Note that $x_t = X_t - \bar{X}$ and similarly $y_t = Y_t - \bar{Y}$. Since $b = \frac{\sum x_t y_t}{\sum x_t^2} = \beta + \frac{\sum x_t u_t}{\sum x_t^2}$

$$\mathbb{v}ar(\tilde{b}) \quad = \quad \mathbb{E}(b-\beta)^2 + \mathbb{E}\left(\sum C_t^2 u_t^2 + 2\sum_{t=1}^{n}\sum_{s>t}^{n} C_t C_s u_t u_s\right) + 2\mathbb{E}\left(\frac{\sum x_t u_t}{\sum x_t^2}\right)\left(\sum C_t u_t\right)$$

According to Classical Assumption (3.b)

$$\begin{aligned}
\mathbb{v}ar(\tilde{b}) \quad &= \quad \mathbb{E}(b-\beta)^2 + \mathbb{E}\left(\sum C_t^2 u_t^2\right) + 2\mathbb{E}\left(\frac{\sum x_t u_t}{\sum x_t^2}\right)\left(\sum C_t u_t\right) \\
&= \quad \mathbb{E}(b-\beta)^2 + \mathbb{E}\left(\sum C_t^2 u_t^2\right) + \frac{2}{\sum x_t^2}\mathbb{E}\left(\sum x_t u_t \sum C_t u_t\right) \\
&= \quad \mathbb{E}(b-\beta)^2 + \mathbb{E}\left(\sum C_t^2 u_t^2\right) + \frac{2}{\sum x_t^2}\mathbb{E}\left(\sum x_t C_t \sigma^2\right) \\
&= \quad \mathbb{E}(b-\beta)^2 + \mathbb{E}\left(\sum C_t^2 u_t^2\right) + \frac{2\sigma^2}{\sum x_t^2}\mathbb{E}\left(\sum x_t C_t\right) \\
&= \quad \mathbb{E}(b-\beta)^2 + \mathbb{E}\left(\sum C_t^2 u_t^2\right) + \frac{2\sigma^2}{\sum x_t^2}\mathbb{E}\left[\sum (X_t - \bar{X}) C_t\right] \\
&= \quad \mathbb{E}(b-\beta)^2 + \mathbb{E}\left(\sum C_t^2 u_t^2\right) + \frac{2\sigma^2}{\sum x_t^2}\mathbb{E}\left(\sum C_t X_t - \bar{X}\sum C_t\right)
\end{aligned}$$

Similarly, according to the statement (3.4)

$$\mathbb{v}ar(\tilde{b}) \quad = \quad \mathbb{E}(b-\beta)^2 + \mathbb{E}\left(\sum C_t^2 u_t^2\right)$$

Lastly, according to Classical Assumption (3.a)

$$\begin{aligned}
\mathbb{v}ar(\tilde{b}) \quad &= \quad \mathbb{E}(b-\beta)^2 + \sum C_t^2 \sigma^2 \\
&= \quad \mathbb{v}ar(b) + \sum C_t^2 \sigma^2
\end{aligned}$$

Since C_t and σ is real and $\sum C_t^2 \sigma^2$ is always non-negative

$$\mathbb{v}ar(\tilde{b}) \quad \geq \quad \mathbb{v}ar(b)$$

This completes the proof. ■

ASSIGNMENT 3

Exercise 3.1 (PROC REG of different periods) *Suppose the model as*

$$Y_t = \alpha + \beta X_t + u_t \tag{3.1}$$

Y_t : Daily log-return on Apple Inc. (AAPL)

X_t : Daily log-return on NASDAQ index (NASDAQ)

Run the regression model (3.1) above

(1) For the period from 2011-01-01 to 2011-10-04

(2) For the period from 2011-10-15 to 2012-08-31

(3) Run the same as in question (1) by replacing X_t by S&P500 index; find and report any difference in the estimates and explain the reason.

(4) Run the same as in question (2) by replacing X_t by S&P500 index; find and report any difference in the estimates and explain the reason.

Figure 3.1 *Basic reporting form of the regression model*

Regression model for (3.1); $R^2 = 0.3927$, $\bar{R}^2 = 0.3836$

Parameter Estimates	Standard Error	t value	p value
1.2544	0.3901	3.2152	0.0008
0.7852	0.1501	5.2325	<.0001

Or simplified form which is often used in various econometric articles is

$$\hat{Y}_t = \underset{(3.2152)}{1.2544} + \underset{(5.2325)}{0.7852} X_t$$

Figure 3.2 *NASDAQ and S&P500 index time-series*

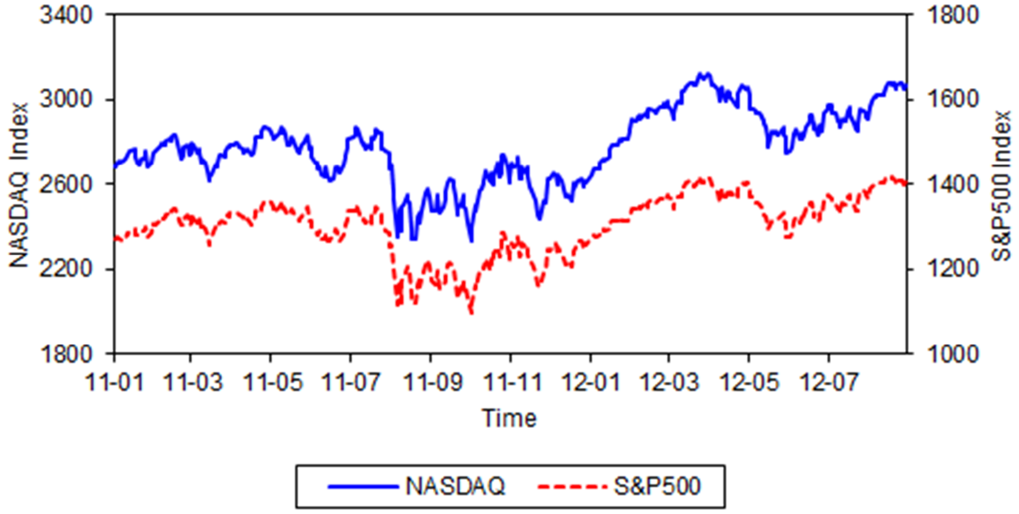
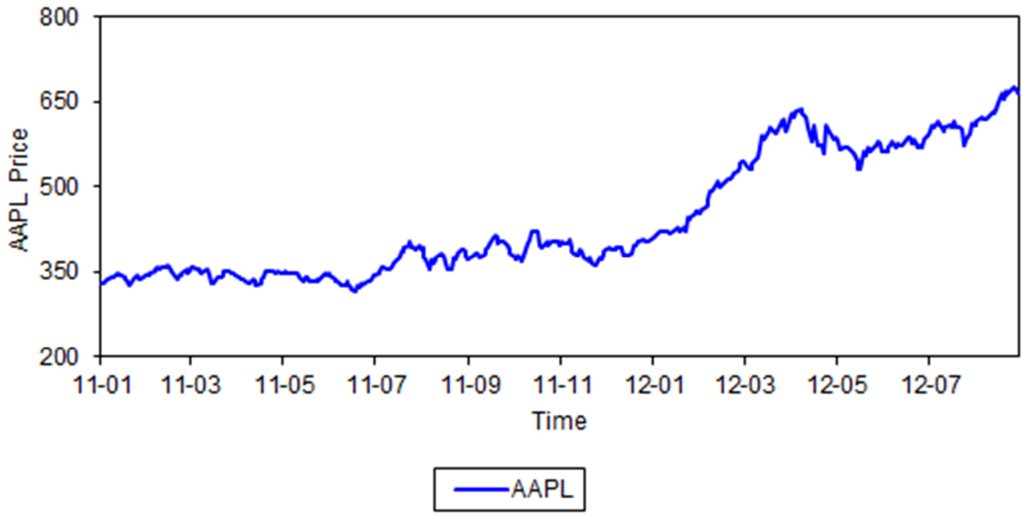


Figure 3.3 *AAPL stock price time-series*



Exercise 3.2 (PROC REG with dummy variables; structural change) *Suppose the regression (3.1) is replaced by*

$$Y_t = \beta_0 + \beta_1 X_t^+ + \beta_2 X_t^- + u_t \tag{3.2}$$

$$X_t^+ := \begin{cases} X_t, & \text{for } X_t \geq 0 \\ 0, & \text{for } X_t < 0 \end{cases}$$

$$X_t^- \quad := \quad \begin{cases} 0, & \text{for } X_t \geq 0 \\ X_t, & \text{for } X_t < 0 \end{cases}$$

(1) Compare the estimates and their t values of β_1 and β_2; what do the results tell us?

(2) Replace NASDAQ index by S&P500 index and do (3.2) above again; are the results analogous to those obtained in (1) of Exercise 3.2?

Exercise 3.3 (PROC REG with lagged variables) *Run the regression model below*

$$Y_t \quad = \quad \alpha + \beta X_t + \gamma W_t + \delta V_t + u_t \tag{3.2}$$

Y_t : Daily Log-return on KOSPI200 index

X_t : Daily Log-return on NASDAQ index

$$W_t \quad := \quad \frac{Y_{t-1} + Y_{t-2} + Y_{t-3} + Y_{t-4} + Y_{t-5}}{5}$$

$$V_t \quad := \quad \frac{Y_{t-1}^2 + Y_{t-2}^2 + Y_{t-3}^2 + Y_{t-4}^2 + Y_{t-5}^2}{5}$$

(1) For the period from 2005-01-01 to 2008-09-14

(2) For the period from 2008-09-15 to 2012-08-31

(3) Run the same as in question (1) of Exercise 3.3 by replacing X_t by S&P500 index; find the difference between the models and explain.

(4) Run the same as in question (2) of Exercise 3.3 by replacing X_t by S&P500 index; find the difference between the models and explain.

(5) Compare of the answers for (1) & (2), and (3) & (4)

(6) Redefine W_t and V_t as

$$W_t \quad := \quad \frac{X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4} + X_{t-5}}{5}$$

$$V_t \quad := \quad \frac{X_{t-1}^2 + X_{t-2}^2 + X_{t-3}^2 + X_{t-4}^2 + X_{t-5}^2}{5}$$

and retry the (1), (2), (3), and (4) with the model (3.2) above; check the significance level (i.e. p value) of γ and δ of (3.2)

(7) What can be inferred from the parameters γ and δ, which is included above models; think about the interpretation. (There is no exact answer.)

Code Assignment 3.1

```
/***************************************************
Exercise 3.1
***************************************************/

/**********Data Input**********/

data aapl;
	infile "c:₩aapld.prn";
	input date aapl;
	dla=dif(log(aapl));
run;

data nasdaq;
	infile "c:₩nasdaqd.prn";
	input date nasdaq;
	dln=dif(log(nasdaq));
run;

data sp500;
	infile "c:₩sp500d.prn";
	input date sp500;
	dls=dif(log(sp500));
run;

/**********Regression model for (1)**********/

data exercise31;
	merge aapl nasdaq;
	by date;
	where 20110101<=date<=20111004;
run;

proc reg data=exercise31;
	model dla=dln;
run;

/**********Regression model for (2)**********/

data exercise31;
	merge aapl nasdaq;
	by date;
	where 20111005<=date<=20120831;
run;

proc reg data=exercise31;
	model dla=dln;
run;

/**********Regression model for (3)**********/
```

```
data exercise31;
        merge aapl sp500;
        where 20110101<=date<=20111004;
run;

proc reg data=exercise31;
        model dla=dls;
run;

/**********Regression model for (4)**********/

data exercise31;
        merge aapl sp500;
        where 20111005<=date<=20120831;
run;

proc reg data=exercise31;
        model dla=dls;
run;

/**************************************************
Exercise 3.2
**************************************************/

/**********Regression model for (1)**********/

data exercise32;
        merge aapl nasdaq;
        by date;
        where 20110101<=date<=20120831;
        if dln>=0 then do;
                dlnp=dln;
                dlnn=0;
        end;
        else do;
                dlnp=0;
                dlnn=dln;
        end;
run;

proc reg data=exercise32;
        model dla=dlnp dlnn;
run;

/**********Regression model for (2)**********/

data exercise32;
        merge aapl sp500;
        by date;
        where 20110101<=date<=20120831;
        if dls>=0 then do;
```

```
                dlsp=dls;
                dlsn=0;
        end;
        else do;
                dlsp=0;
                dlsn=dls;
        end;
run;

proc reg data=exercise32;
        model dla=dlsp dlsn;
run;

/**************************************************
Exercise 3.3
**************************************************/

data kp200;
        infile "c:₩kp200d.prn";
        input date kp200;
        dlk=dif(log(kp200));
run;

data exercise33;
        merge kp200 nasdaq;
        by date;
        y1=lag1(dlk);
        y2=lag2(dlk);
        y3=lag3(dlk);
        y4=lag4(dlk);
        y5=lag5(dlk);
        w=(y1+y2+y3+y4+y5)/5;
        ysq1=y1**2;
        ysq2=y2**2;
        ysq3=y3**2;
        ysq4=y4**2;
        ysq5=y5**2;
        v=(ysq1+ysq2+ysq3+ysq4+ysq5)/5;
run;

/*Regression model for (1)*/

proc reg data=exercise33;
        model dlk=dln w v;
        where 20050101<=date<=20080914;
run;

/*Regression model for (2)*/

proc reg data=exercise33;
        model dlk=dln w v;
```

```
        where 20080915<=date<=20120831;
run;

data exercise33;
        merge exercise33 sp500;
        by date;
run;

/*Regression model for (3)*/

proc reg data=exercise33;
        model dlk=dls w v;
        where 20050101<=date<=20080914;
run;

/*Regression model for (4)*/

proc reg data=exercise33;
        model dlk=dls w v;
        where 20080915<=date<=20120831;
run;

data exercise33;
        merge kp200 nasdaq;
        by date;
        x1=lag1(dln);
        x2=lag2(dln);
        x3=lag3(dln);
        x4=lag4(dln);
        x5=lag5(dln);
        w=(x1+x2+x3+x4+x5)/5;
        xsq1=x1**2;
        xsq2=x2**2;
        xsq3=x3**2;
        xsq4=x4**2;
        xsq5=x5**2;
        v=(xsq1+xsq2+xsq3+xsq4+xsq5)/5;
run;

/*Regression model for (6-1)*/

proc reg data=exercise33;
        model dlk=dln w v;
        where 20050101<=date<=20080914;
run;

/*Regression model for (6-2)*/

proc reg data=exercise33;
        model dlk=dln w v;
        where 20080915<=date<=20120831;
```

```
run;

data exercise33;
	merge kp200 sp500;
	by date;
	x1=lag1(dls);
	x2=lag2(dls);
	x3=lag3(dls);
	x4=lag4(dls);
	x5=lag5(dls);
	w=(x1+x2+x3+x4+x5)/5;
	xsq1=x1**2;
	xsq2=x2**2;
	xsq3=x3**2;
	xsq4=x4**2;
	xsq5=x5**2;
	v=(xsq1+xsq2+xsq3+xsq4+xsq5)/5;
run;

/*Regression model for (6-3)*/

proc reg data=exercise33;
	model dlk=dls w v;
	where 20050101<=date<=20080914;
run;

/*Regression model for (6-4)*/

proc reg data=exercise33;
	model dlk=dls w v;
	where 20080915<=date<=20120831;
run;

quit;
```

Short Commentary

Code Assignment 3.2

```
/***************************************************
1. 외부 데이터를 SAS 데이터셋으로 만들기
***************************************************/

data (만들 데이터셋 이름);
	infile "(파일이 저장된 위치)";
	input (1열 변수 이름) (2열 변수 이름) (3열 변수 이름);
	(로그 차분된 변수 이름)=dif(log((로그 차분할 변수 이름)));
run;
```

```
/*****************************************************
2. 두 개 이상의 데이터를 합쳐 하나로 만들기
*****************************************************/

data (합쳐서 만들어질 데이터셋 이름);
        merge (합칠 데이터셋 1번) (합칠 데이터셋 2번);
        by (기준으로 삼을 변수 이름, 주로 날짜);
        where (시작하는 날짜)<=(날짜 변수)<=(끝나는 날짜);
run;

/*****************************************************
3. 특정 데이터셋을 불러와서 넘버링하기
*****************************************************/

data (가져올 데이터셋 이름);
        set (가져올 데이터셋 이름);
        (넘버링 변수 이름)=_n_;
run;

/*****************************************************
4. 회귀분석 후 예측값, 잔차 새로운 데이터셋으로 내보내기
*****************************************************/

proc reg data=(분석할 데이터셋 이름);
        model (종속변수 이름)=(독립변수 1) (독립변수 2) (독립변수 …);
        output out=(내보낼 데이터셋) p=(예측값 저장할 변수) r=(잔차 저장할 변수);
run;

/*****************************************************
5. 데이터셋 안에 있는 변수로 플로팅하기
*****************************************************/

proc gplot data=(플로팅할 데이터셋 이름);
        symbol i=(직선으로 그리려면 join, 곡선으로 그리려면 spline);
        plot (Y축에 들어갈 변수 이름)*(X축에 들어갈 변수 이름);
run;

/*****************************************************
6. 데이터셋 불러와서 변수 시간 지연시키기
*****************************************************/

data (변수가 저장된 데이터셋 이름);
        set (변수가 저장된 데이터셋 이름);
        (지연된 변수 이름)=lag1(지연시킬 변수 이름);
run;

/*****************************************************
7. 절편(Intercept) 없이 회귀분석 실행하기
*****************************************************/
```

```
proc reg data=(분석할 데이터셋 이름);
        model (종속변수)=(독립변수 1) (독립변수 2) (독립변수 …)/noint;
run;

/**************************************************
8. 특정 시간대만 잘라서 회귀분석 실행하기
**************************************************/

proc reg data=(분석할 데이터셋 이름);
        model (종속변수)=(독립변수 1) (독립변수 2) (독립변수 …);
        where (시작하는 날짜)<=(날짜 변수)<=(끝나는 날짜);
run;

/**************************************************
9. 단변량분석 실행하기
**************************************************/

proc univariate data=(분석할 데이터셋 이름);
        var (분석할 변수 이름);
        histogram; /*분석되는 변수의 도수분포표를 작성해준다.*/
run;

/**************************************************
10. 시계열 시뮬레이션 (표현이 다소 난해하여 예제만 수록)
**************************************************/

/*10-1. AR(1) 모형, Y(t)=0.2+0.7Y(t-1)+e(t), 1000번 시뮬레이션*/

data example;
        mu=0.2;
        phi=0.7;
        seed1=1234;
        seed2=5678;
        ylag=rannor(seed1);
        do t=1 to 1000;
                e=rannor(seed2);
                y=mu+phi*ylag+e;
                ylag=y;
                output;
        end;
run;

proc arima data=example;
        identify var=y;
run;

/*10-2. MA(1) 모형, Y(t)=e(t)+0.7e(t-1)*/

data example;
        theta=0.7;
        seed1=1234;
```

```
          seed2=5678;
          elag=rannor(seed1);
          do t=1 to 1000;
                    e=rannor(seed2);
                    y=e+theta*elag;
                    elag=e;
                    output;
          end;
run;

proc arima data=example;
          identify var=y;
run;

/*10-3. AR(2) 모형, Y(t)=0.2+0.7Y(t-1)+0.2Y(t-2)+e(t)*/

data example;
          mu=0.2;
          phi1=0.7;
          phi2=0.2;
          seed1=1234;
          seed2=5678;
          seed3=9012;
          ylag1=rannor(seed1);
          ylag2=rannor(seed2);
          do t=1 to 1000;
                    e=rannor(seed3);
                    y=mu+phi1*ylag1+phi2*ylag2+e;
                    ylag2=ylag1;
                    ylag1=y;
                    output;
          end;
run;

proc arima data=example;
          identify var=y;
run;

/*10-4. MA(2) 모형, Y(t)=e(t)+0.7e(t-1)+0.2e(t-2)*/

data example;
          theta1=0.7;
          theta2=0.2;
          seed1=1234;
          seed2=5678;
          seed3=9012;
          elag1=rannor(seed1);
          elag2=rannor(seed2);
          do t=1 to 100000;
                    e=rannor(seed3);
                    y=e+theta1*elag1+theta2*elag2;
```

```
                elag2=elag1;
                elag1=e;
                output;
        end;
run;

proc arima data=example;
        identify var=y;
run;

/*10-5. ARMA(1,1) 모형, Y(t)=0.2+0.5Y(t-1)+e(t)+0.3e(t-1)*/

data example;
        mu=0.2;
        phi=0.5;
        theta=0.3;
        seed1=1234;
        seed2=5678;
        seed3=9012;
        ylag=rannor(seed1);
        elag=rannor(seed2);
        do t=1 to 1000;
                e=rannor(seed3);
                y=mu+phi*ylag+e+theta*elag;
                ylag=y;
                elag=e;
                output;
        end;
run;

proc arima data=example;
        identify var=y;
run;

quit;
```

Prof. Kiseok Lee (econklee@khu.ac.kr)
TA Junyong Kim (mindhunter@khu.ac.kr)

LECTURE NOTE 4

1 Standardization of b

Assumption (Homoscedasticity) $\forall i = \{1,2,3,\cdots,n\}$, u_i *are distributed as normal with mean* 0 *and variance* σ^2, *i.e.*

$$u_i \sim \mathcal{N}(0,\sigma^2)$$

To reflect above property, Probability Density Function (PDF) of u_i can be assumed as

$$\begin{aligned}\mathbb{P}(u_i) &= \mathbb{P}(u_i;\sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{u_i}{\sigma^2}\right)^2} \\ &= (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{1}{2}\times\frac{u_i^2}{\sigma^2}\right)\end{aligned}$$

Remind that

$$\begin{aligned}b &= \frac{\sum x_i y_i}{\sum x_i^2} \\ &= \beta + \frac{\sum x_i u_i}{\sum x_i^2}\end{aligned}$$

Since $x_i = X_i - \bar{X}$ are some numbers we don't care about by Classical Assumption (1), we know that

$$\begin{aligned}\mathbb{E}(b) &= \beta \leftarrow \text{Classical Assumption (1) and (2)} \\ \mathbb{v}ar(b) &= \frac{\sigma^2}{\sum x_i^2} \leftarrow \text{Classical Assumption (1), (2), (3a) and (3b)}\end{aligned}$$

Since a linear combination of normally distributed random variables is normally distributed, we can write

$$b \sim \mathcal{N}\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right)$$

Since we know its mean and variance, we can standardize it in following manner.

Standardization of b

$$z = \frac{b-\beta}{\sqrt{\frac{\sigma^2}{\sum x_i^2}}}$$

$$\sim \mathcal{N}(0,1)$$

$$\mathbb{v}ar(u_i) = \sigma^2$$

$$\mathbb{v}ar(2u_i) = 4\sigma^2$$

$$\mathbb{v}ar(3u_i) = 9\sigma^2$$

$$\mathbb{v}ar\left(\frac{u_i}{\sqrt{\sigma^2}}\right) = 1$$

Figure 4.1.1 *Probability Density Function of* b

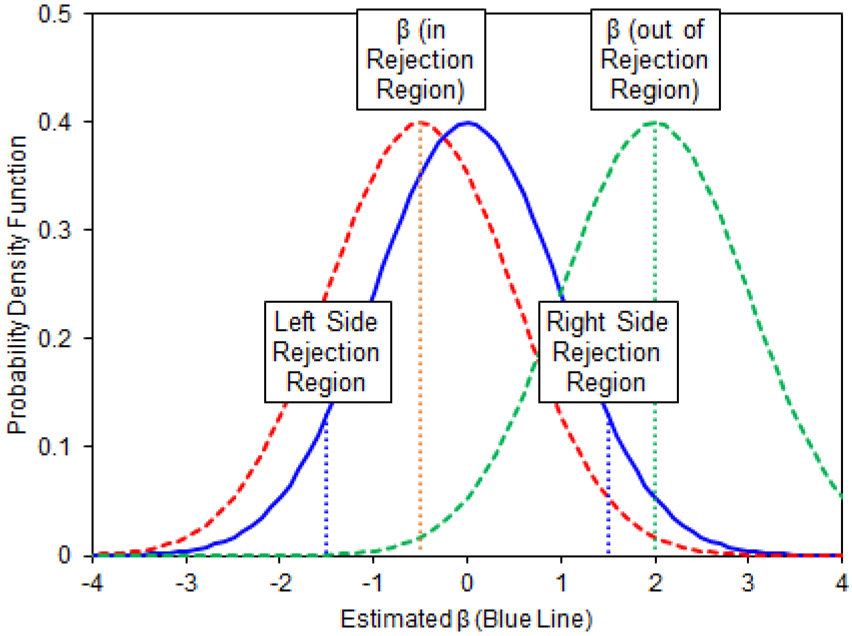


Figure 4.1.2 *Homoscedasticity;* $\forall i = \{1,2,3,\cdots,n\},\ \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots = \sigma_n^2 = \sigma^2$

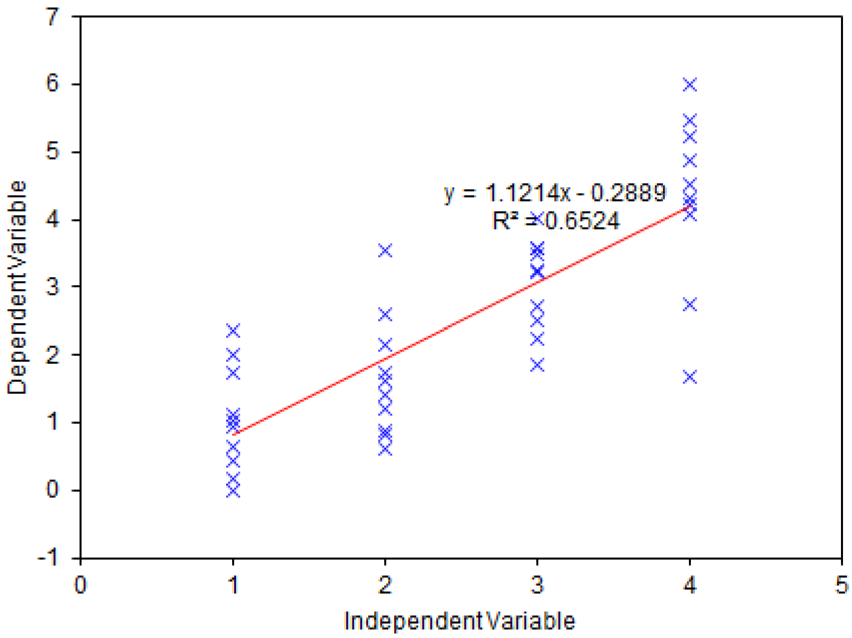
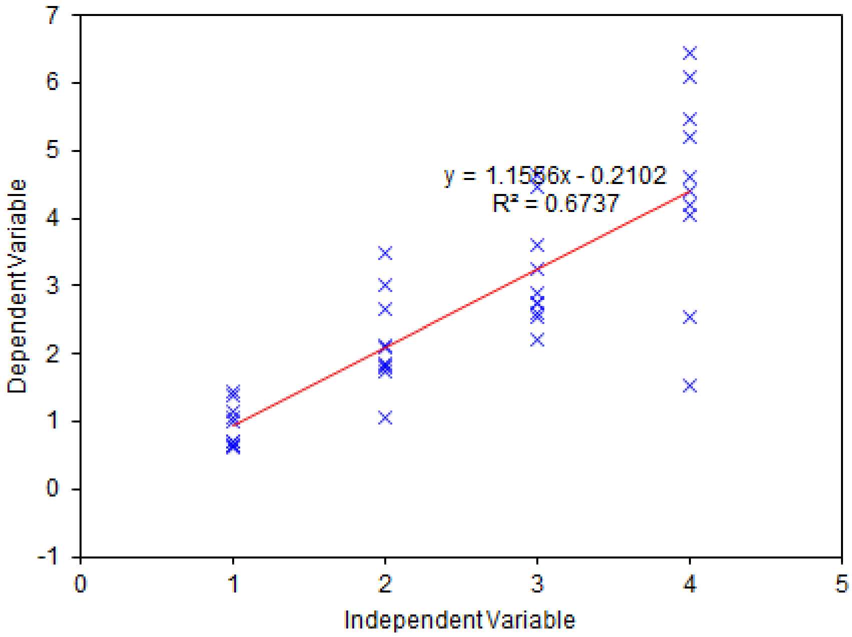


Figure 4.1.3 *Heteroscedasticity;* $\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2 \neq \cdots \sigma_n^2 \neq \sigma^2$

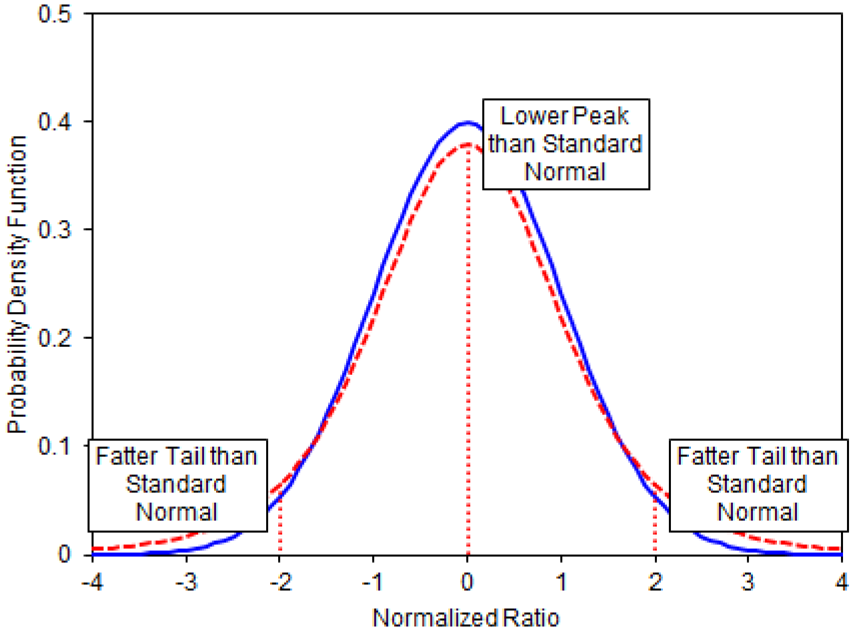


2 t Test Statistic

The ratio t, which is defined as below follows t distribution with k degrees of freedom, by which Standard Normal distribution can be approximated.

$$
\begin{aligned}
t &= \frac{v}{\sqrt{\frac{w}{k}}} \\
&\sim \text{Student } t_k \text{ (Student } t \text{ distribution with } k \text{ degrees of freedom)} \\
v &\sim \mathcal{N}(0,1) \\
w &\sim \chi_k^2 \\
k &= \text{Degrees of freedom}
\end{aligned}
$$

Figure 4.1.4 *Properties of t distribution compared to Standard Normal distribution*



Definition (Probability Density Function of t Distribution) *Student t distribution has the probability density function given by*

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

ν : The number of degrees freedom

Γ : Gamma function

$$\Gamma(k) = (k-1)(k-2)(k-3)\cdots(2)(1)$$

$$\mathbb{E}(t) = 0$$

$$\mathbb{V}ar(t) = \frac{\nu}{\nu-2}, \forall \nu > 2$$

One of the important properties of Student t distribution is that if $k \to \infty$, then $t \to \mathcal{N}(0,1)$. Note that

$$w = \frac{\sum e_i^2}{\sigma^2} \sim \chi^2_{(n-2)}$$

And it is known that $z = \frac{x-\mu}{\sigma}$ and w are independent. The ratio then,

$$
\begin{aligned}
t &= \frac{\dfrac{b-\beta}{\sqrt{\dfrac{\cancel{\sigma^2}}{\sum x_i}}}}{\sqrt{\dfrac{\sum e_i^2}{\cancel{\sigma^2}(n-2)}}} \sim t_{(n-2)} \\
&= \frac{z}{\sqrt{\dfrac{w}{(n-2)}}} \\
&= \frac{(b-\beta)\sqrt{\sum x_i^2}}{\sqrt{\dfrac{\sum e_i^2}{n-2}}} \\
&= \frac{b-\beta}{\sqrt{\dfrac{\sum e_i^2}{n-2} \times \dfrac{1}{\sum x_i^2}}} \\
&= \frac{b-\beta}{\underbrace{\sqrt{\dfrac{\sum e_i^2}{n-2}}}_{s^2} \times \dfrac{1}{\sqrt{\sum x_i^2}}}
\end{aligned}
$$

If σ^2 is unknown then

$$
s^2 = \frac{\sum e_i^2}{n-2} \leftarrow \text{unbiased} \tag{4.1}
$$

$$
\hat{\sigma}^2 = \frac{\sum e_i^2}{n} \leftarrow \text{biased}
$$

Note that

$$
\begin{aligned}
\mathbb{V}ar(u_i) &= \sigma^2 \\
&= \mathbb{E}\left[u_i - \underbrace{\mathbb{E}(u_i)}_{0}\right]^2 \\
&= \mathbb{E}(u_i)^2
\end{aligned}
$$

(4.1) shows that s^2, not $\hat{\sigma}^2$, is the unbiased estimator of σ^2. Thus,

$$
\begin{aligned}
\mathbb{E}(s^2) &= \sigma^2 \\
t &= \frac{b-\beta}{\sqrt{\dfrac{s^2}{\sum x_i^2}}}
\end{aligned}
$$

$$= \frac{\mathrm{b} - \upbeta}{\widehat{\text{Standard Error}} \text{ of } \beta}$$

$$\mathbb{v}ar(b) = \frac{\sigma^2}{\sum x_i^2}$$

$$\widehat{\mathbb{v}ar}(b) = \frac{s^2}{\sum x_i^2}$$

$$\mathrm{S.E.}(b) = \sqrt{\mathbb{v}ar(b)}$$

$$= \sqrt{\frac{\sigma^2}{\sum x_i^2}}$$

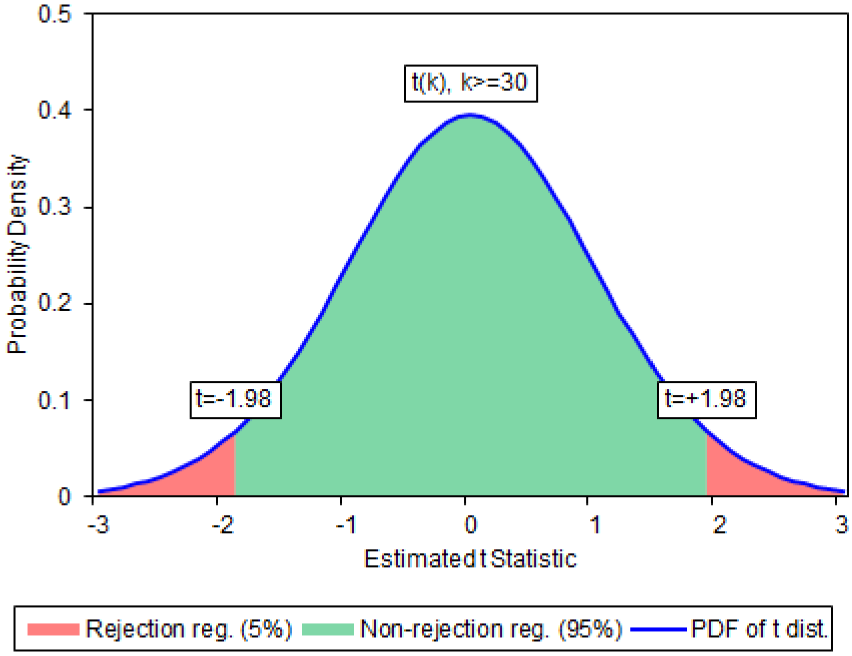
$$\widehat{\mathrm{S.E.}}(b) = \sqrt{\frac{s^2}{\sum x_i^2}}$$

Therefore, the test statistic t for the null hypothesis $\beta = 0$ can be denoted as

$$t = \frac{b}{\widehat{\mathrm{S.E.}}\ (b)}$$

$$\sim \mathcal{N}(0,1^2)$$

Figure 4.1.4 *Critical value for 95% confidence interval*



ASSIGNMENT 4

Exercise 4.1 *Suppose the model* (4.1) *and* (4.2) *for* $19540701 \leq t \leq 20120701$

$$Y_t = \alpha + \beta X_{t-\tau} + u_t \tag{4.1}$$

$$Y_t = \alpha^+ d_t^+ + \alpha^- d_t^- + \beta^+ X_{t-\tau}^+ + \beta^- X_{t-\tau}^- + u_t \tag{4.2}$$

Y_t ≔ Annualized growth rate of Industrial Production

X_t ≔ The τ-th lag of the Federal Funds Rate

$$d_t^+ := \begin{cases} 1, & \forall t < 19830101 \\ 0, & \forall t \geq 19830101 \end{cases}$$

$$d_t^- := 1 - d_t^+$$

$$X_t^+ := X_t d_t^+$$

$$X_t^- := X_t d_t^-$$

(1) Obtain $a(=\hat{\alpha})$and $b\left(=\hat{\beta}\right)$ of (4.1) for $\tau = 0,1,2,\cdots,6$ and select the best τ that produces the largest absolute t-value. Denote Y_t with selected lag as $X_{t-\tau}$

(2) Obtain a^+, a^-, b^+, b^- of (4.2) for $\tau = 0$ and denote Y_t with the estimate and X_t

(3) Test $H_0 : \alpha^+ = \alpha^-$ and $\beta^+ = \beta^-$

(4) Change the definition of d_t^+ as below (Then, the definition of d_t^-, X_t^+, X_t^- will also be changed automatically.) and do above experiment once again.

$$d_t^+ := \begin{cases} 1, & \forall X_t \geq X_{t-1} \\ 0, & \forall X_t < X_{t-1} \end{cases}$$

(5) Consider $e_t(=\hat{u}_t)$ of the (4.1). Try the following regression and test $H_0 : \gamma_1 = 0$

$$e_t = \gamma_0 + \gamma_1 e_{t-1} + \zeta_t$$

(6) Consider the same of two (4.2). Try the following regression and test $H_0 : \delta_1 = 0$

$$e_t = \delta_0 + \delta_1 e_{t-1} + \zeta_t$$

Code Assignment 4

```
data ip;
        infile "c:₩ip.prn";
        input month ip;
        ipg=dif(log(ip))*1200;
run;

data fyff;
        infile "c:₩fyff.prn";
        input month fyff;
        fyff1=lag1(fyff);
        fyff2=lag2(fyff);
        fyff3=lag3(fyff);
        fyff4=lag4(fyff);
        fyff5=lag5(fyff);
        fyff6=lag6(fyff);
run;

data assign4;
        merge ip fyff;
        where 19540701<=month<=20120701;
run;

proc reg data=assign4;
        model ipg=fyff;
        model ipg=fyff1;
        model ipg=fyff2;
        model ipg=fyff3;
        model ipg=fyff4;
        model ipg=fyff5;
        model ipg=fyff6;
run;

proc reg data=assign4;
        model ipg=fyff4;
        output out=assign41 r=resid;
run;

data assign4;
        set assign4;
        if month<19830101 then dpos=1;
        else dpos=0;
        dneg=1-dpos;
        fyffpos=fyff*dpos;
        fyffneg=fyff*dneg;
run;

proc reg data=assign4;
        model ipg=dpos dneg fyffpos fyffneg/noint;
```

```
          test dpos=dneg,fyffpos=fyffneg;
          output out=assign42 r=resid;
run;

data assign4;
          set assign4;
          if fyff>=lag1(fyff) then dpos=1;
          else dpos=0;
          dneg=1-dpos;
          fyffpos=fyff*dpos;
          fyffneg=fyff*dneg;
run;

proc reg data=assign4;
          model ipg=dpos dneg fyffpos fyffneg/noint;
          test dpos=dneg,fyffpos=fyffneg;
          output out=assign43 r=resid;
run;

data assign41;
          set assign41;
          resid1=lag1(resid);
run;

data assign42;
          set assign42;
          resid1=lag1(resid);
run;

data assign43;
          set assign43;
          resid1=lag1(resid);
run;

proc reg data=assign41;
          model resid=resid1;
run;

proc reg data=assign42;
          model resid=resid1;
run;

proc reg data=assign43;
          model resid=resid1;
run;

quit;
```

Lecture Note 5

1 Confidence Interval

The t statistic

$$\begin{aligned} t &= \frac{b-\beta_0}{\sqrt{\frac{s^2}{\sum x_i^2}}} \\ &\sim t_{n-2} \end{aligned}$$

Where β_0 is *the null hypothesis* and $s^2 = \frac{\sum e_i^2}{n-2}$ under the null hypothesis, i.e. if the null is true, that is $\beta_0 = \beta$, the above random variable t is distributed as student-t with $(n-2)$ degrees of freedom. Note that the variable z below coming from Classical Assumption (3.a)

$$\begin{aligned} z &= \frac{\overbrace{b}^{\text{random}} - \overbrace{\beta}^{\text{true}}}{\sqrt{\frac{\sigma^2}{\sum x_i^2}}} \\ &\sim \mathcal{N}(0,1^2) \end{aligned}$$

The estimators of σ^2 is

$$s^2 = \frac{\sum e_i^2}{n-2} \tag{5.1}$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n} \tag{5.2}$$

Since Classical Assumption (3.a)

$$\begin{aligned} \mathbb{E}(s^2) &= \sigma^2 \\ \mathbb{E}(\hat{\sigma}^2) &\neq \sigma^2 \end{aligned}$$

$$\mathbb{V}ar(s^2) \quad > \quad \mathbb{V}ar(\hat{\sigma}^2) \leftarrow \text{which is not needed yet}$$

All above means that if the null hypothesis is false, then

$$\beta_0 \neq \beta$$

$$\mathbb{E}(b - \beta_0) \neq 0$$

$$\mathbb{E}(b) - \beta_0 = \beta - \beta_0 \neq 0$$

Under the significance level of α,

$$\mathbb{P}\left[\left|\frac{b-\beta_0}{\sqrt{\frac{s^2}{\sum x_i^2}}}\right| < t_\alpha\right] = 1-\alpha$$

Where t_α is some constant

$$\mathbb{P}\left[-t_\alpha < \frac{b-\beta_0}{\sqrt{\frac{s^2}{\sum x_i^2}}} < t_\alpha\right] = 1-\alpha$$

$$\mathbb{P}\left[-t_\alpha\sqrt{\frac{s^2}{\sum x_i^2}} < b-\beta_0 < t_\alpha\sqrt{\frac{s^2}{\sum x_i^2}}\right] = 1-\alpha$$

$$\mathbb{P}\left[b-t_\alpha\sqrt{\frac{s^2}{\sum x_i^2}} < \beta_0 < b+t_\alpha\sqrt{\frac{s^2}{\sum x_i^2}}\right] = 1-\alpha$$

Suppose $\alpha = 0.05$, the 95% confidence interval is

$$\left(b - \underbrace{t_{0.05}}_{1.98}\underbrace{\sqrt{\frac{s^2}{\sum x_i^2}}}_{\text{S.E.}(b)}, b + \underbrace{t_{0.05}}_{1.98}\underbrace{\sqrt{\frac{s^2}{\sum x_i^2}}}_{\text{S.E.}(b)}\right)$$

If $b = 2.0$, $\text{S.E.}(b) = 1.0$, then $(2.00 - 1.98 \times 1, 2.00 + 1.98 \times 1) = (0.02, 3.98)$

The 95% confidence interval implies that if one constructs the interval many times, 95% of the intervals constructed will contain the true β. However, once the interval is constructed like $(0.02,3.98)$, the probability that this interval contains the true value. β is either 1 or 0.

2 Hypothesis Testing

If the hypothesis testing is performed using 5% significance level, then if the null falls inside the confidence interval, then the null is not rejected. Otherwise, it is rejected.

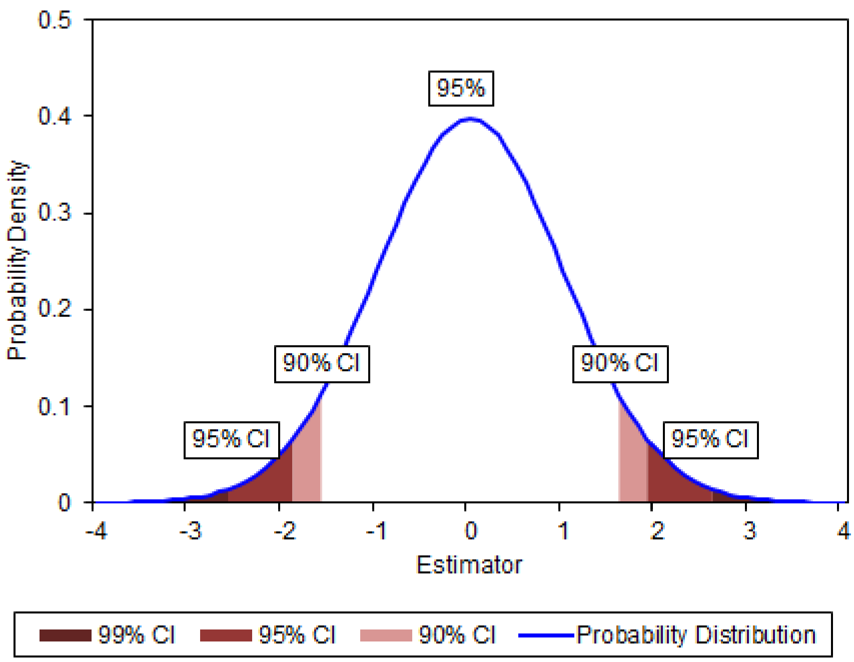
Ex. Confidence Interval = $(0.02, 3.98) \rightarrow H_0: \beta_0 = -1.0$(be rejected) & $H_0: \beta_0 = 2.0$(cannot be)

Below arguments are the errors involved in hypothesis testing.

2.1 Type 1 Error

Definition (Type 1 Error) *The error of rejecting the correct hypothesis. Suppose we have the correct hypothesis.*

Figure 5.2.1 *Probability distribution of estimator*

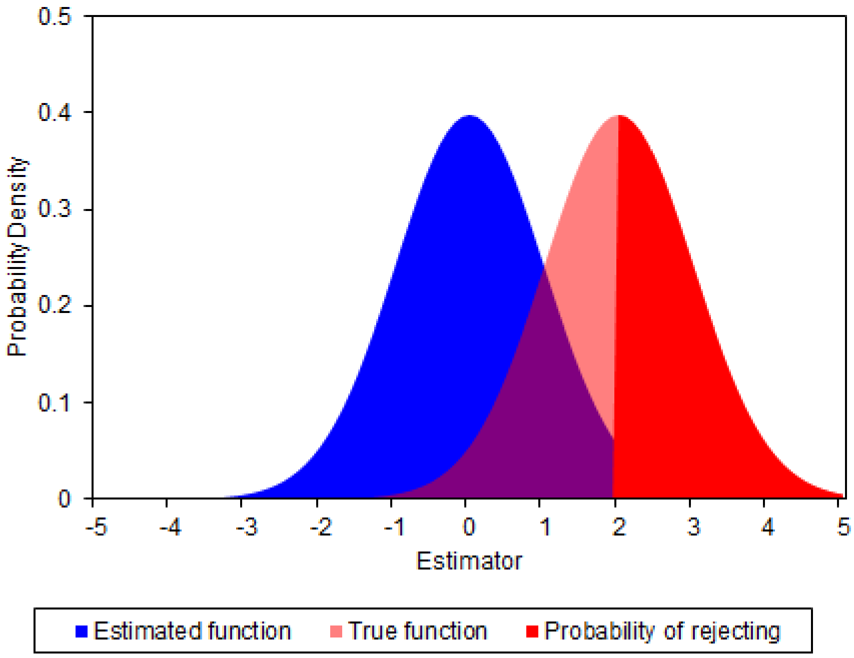


When the hypothesis is tested based upon the significance level of 5%; the confidence level of 95%, the probability of Type 1 Error is 5%.

2.2 Type 2 Error

Definition (Type 2 Error) *The error of accepting the wrong (incorrect) hypothesis.*

Figure 5.2.2 *Probability of rejecting the null; non-faded red region*



Sum of non-faded red region is the probability of rejecting the hypothesis $\beta_0 = 0$ (while $\beta = 2$) area of faded red is the probability of accepting the incorrect hypothesis, i.e. probability of committing Type 2 Error.

Schematically, we have

Figure 5.2.3 *Type 1 and Type 2 Errors*

Decision	State of Nature	
	H_0 is true	H_0 is false
Reject	Type 1 Error	No error
Do not reject	No error	Type 2 Error

2.3 Power of Test

Power of test can be denoted as $1 - \mathbb{P}(\text{Type 2 Error})$

Figure 5.2.4 *Shaded area indicates the probability that $\bar{X}$ will fall into the critical region.*[4]

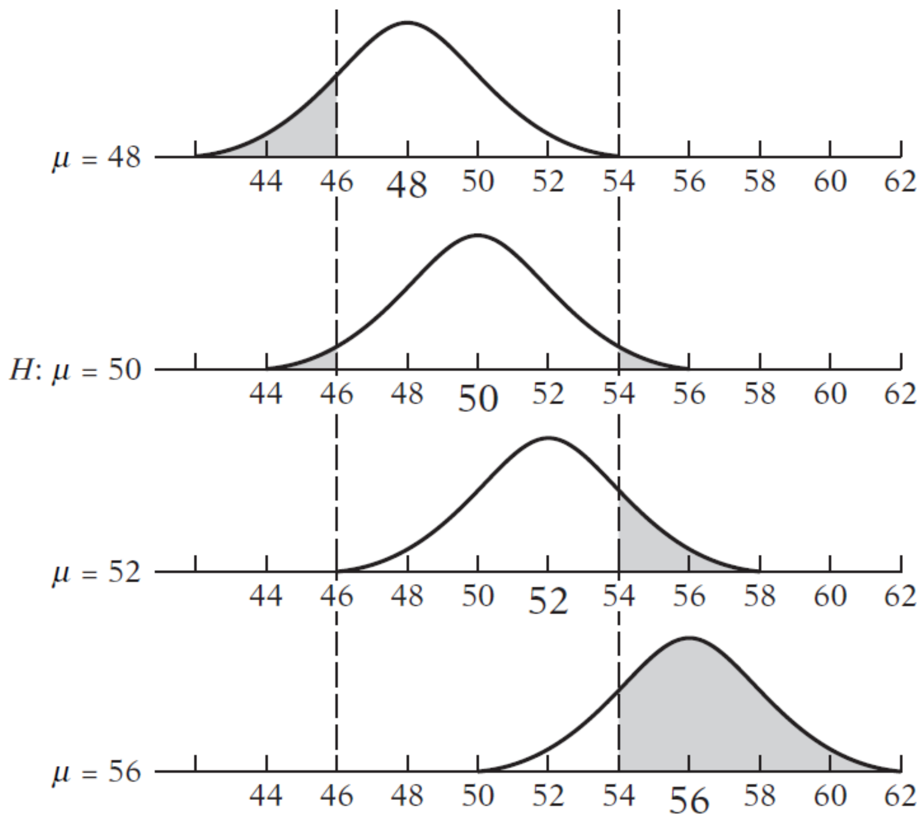


Figure 5.2.4 shows the distribution of X when $N = 25$, $\sigma = 10$, and $\mu = 48$, 50, 52, or 56. Under $H\!: \mu = 50$, the critical region with $\alpha = 0.05$ is $\bar{X} < 46.1$ and $\bar{X} > 53.9$. The shaded area indicates the probability that $\bar{X}$ will fall into the critical region. This probability is

$$\mathbb{P} = \begin{cases} 0.17 \text{ if } \mu = 48 \\ 0.05 \text{ if } \mu = 50 \\ 0.17 \text{ if } \mu = 52 \\ 0.85 \text{ if } \mu = 56 \end{cases}$$

Farther from true value, β, μ, etc., Power increases and Type 2 Error decreases.

Sharp curve is better than blunt curve. (Figure 5.2.5) Type 2 Error can be avoided through expanding the sample size.

There exists a trade-off between the probability of Type 1 Error and the probability of Type 2 Error. (Figure 5.2.6)

[4] Damodar N. Gujarati, 2004, *Basic Econometrics*

Figure 5.2.5 *Power function of test of hypothesis with* $\mu = 50$

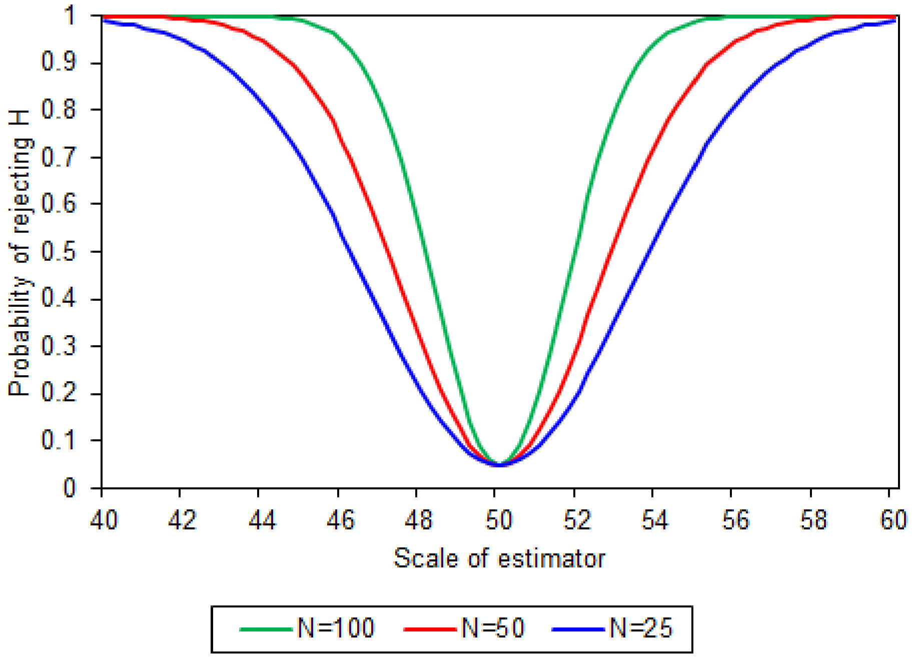
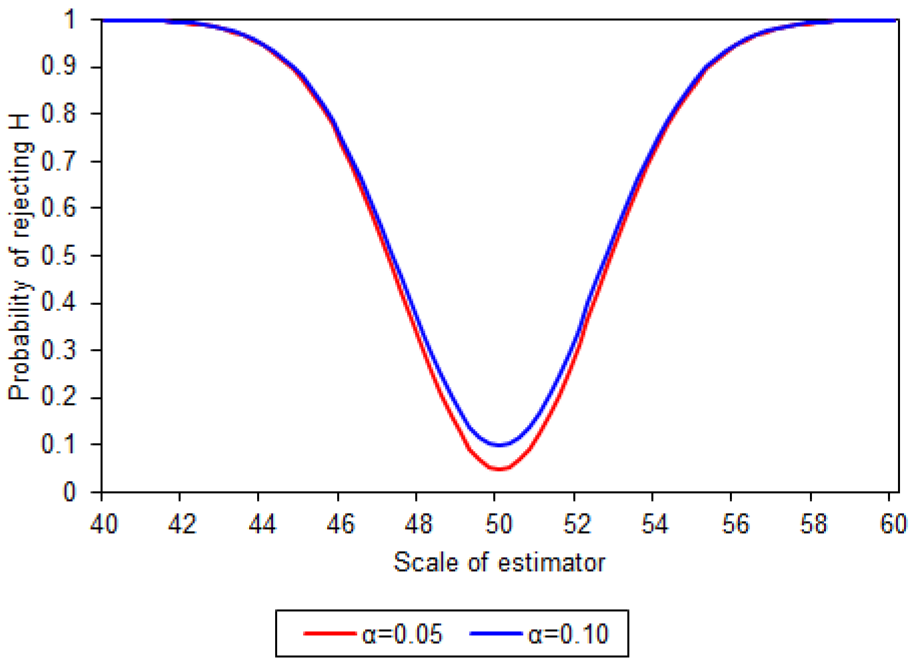


Figure 5.2.5 *Trade-off between the probability of Type 1 Error and Type 2 Error*



3 SAS Code

Code 5.3.1 *Regression model with Industrial Production and Federal Funds Rate*

```
data ip;
        infile "c:₩ip.prn";
        input month ip;
        ipg=dif(log(ip))*1200;
run;

data ffr;
        infile "c:₩ffr.prn";
        input month ffr;
run;

data all1;
        merge ip ffr;
        by month;
run;

proc reg data=all1;
        model ipg=ffr/dw;
        output out=out1 r=resid;
run;

proc gplot data=out1;
        symbol i=join;
        plot resid*month;
run;

quit;
```

Figure 5.3.1 *Results of Regression model of Code 5.3.1*

SAS 시스템

The REG Procedure
Model: MODEL1
Dependent Variable: ipg

Number of Observations Read	697
Number of Observations Used	697

Analysis of Variance

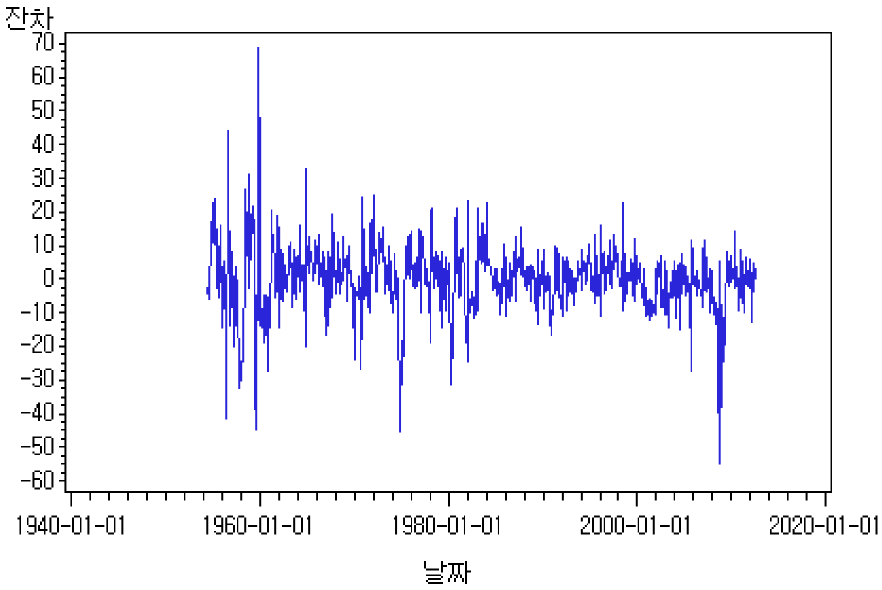
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	912.78654	912.78654	8.02	0.0048
Error	695	79129	113.85406		
Corrected Total	696	80041			

Root MSE	10.67024	R-Square	0.0114
Dependent Mean	2.91386	Adj R-Sq	0.0100
Coeff Var	366.18944		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.64777	0.73372	6.33	<.0001
fyff	1	-0.32820	0.11591	-2.83	0.0048

Figure 5.3.2 *Residual Plot of Code 5.3.1*



ASSIGNMENT 5

Exercise 5.1 *Suppose the regression model (5.1) below; fixed sample period from January 1990 to February 2009*

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + u_t \tag{5.1}$$

$$Y_t \coloneqq \text{Annualized growth rate of Industrial Production}$$

$$X_t \coloneqq \text{Federal Funds Rate}$$

(1) Obtain the coefficients of the model (5.1). Check the Durbin-Watson statistic. Is there serial correlation in u_t?

(2) Run the regression model (5.2) following. Are the estimates of γ_1 and γ_2 statistically significant? What can be inferred from the model (5.2)?

$$e_t = \gamma_0 + \gamma_1 e_{t-1} + \gamma_2 X_t + \varepsilon_t \tag{5.2}$$

(3) Test the hypothesis $H_0 : \gamma_1 = \gamma_2 = 0$

(4) Do above (1), (2), and (3) without the explanatory variable Y_{t-1}; what the different can be captured?

(5) Correct the model (5.1) with below and check the coefficients.

$$Y_t = \beta_0 + \beta_1 X_t + \beta Y_{t-1} + u_t$$

$$u_t = -\phi u_{t-1} + \nu_t$$

$$\nu_t \sim \mathcal{N}(0, \sigma_\nu^2)$$

(6) Consider the following (5.3) with e_t obtained from the model (5.2) and obtain the coefficients α_0 and α_1 below.

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \xi_t$$

Code Assignment 5

```
data ip;
        infile "c:₩ip.prn";
        input month ip;
        ipg=dif(log(ip))*1200;
        ipg1=lag1(ipg);
run;

data fyff;
        infile "c:₩fyff.prn";
        input month fyff;
        fyff1=lag1(fyff);
run;

data hw51;
        merge ip fyff;
        where 19900101<=month<20100101;
run;

proc reg data=hw51;
        model ipg=fyff ipg1/dwprob;
        output out=hw52 r=resid;
run;

data hw52;
        set hw52;
        resid1=lag1(resid);
run;

proc reg data=hw52;
        model resid=resid1 ipg;
run;

proc reg data=hw51;
        model ipg=fyff/dwprob;
        output out=hw53 r=resid;
run;

data hw53;
        set hw53;
        resid1=lag1(resid);
run;

proc reg data=hw53;
        model resid=resid1;
run;

proc autoreg data=hw51;
        model ipg=fyff ipg1/p=1 method=ml;
run;
```

```
data hw52;
        set hw52;
        residsq=resid**2;
        residsq1=resid1**2;
run;

proc reg data=hw52;
        model residsq=residsq1;
run;

quit;
```

LECTURE NOTE 6

1 Coefficient of Determination

1.1 R-square

$$Y_i = a + bX_i + e_i \tag{6.1}$$

$$\bar{Y} = a + b\bar{X} \tag{6.2}$$

Through difference (6.2) from (6.1) below equation can be constructed.

$$(Y_i - \bar{Y}) = b(X_i - \bar{X}) + e_i$$

With $y_i = Y_i - \bar{Y}$ and $x_i = X_i - \bar{X}$ above equation can be transformed as

$$\begin{aligned} y_i &= bx_i + e_i \\ y_i^2 &= b^2x_i^2 + e_i^2 + 2bx_ie_i \\ \sum_{i=1}^{n} y_i^2 &= b^2\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} e_i^2 + 2b\underbrace{\sum_{i=1}^{n} x_ie_i}_{0} \end{aligned}$$

Note that

$$\begin{aligned} \sum_{i=1}^{n} x_ie_i &= \sum_{i=1}^{n} (X_i - \bar{X})e_i \\ &= \sum_{i=1}^{n} X_ie_i - \bar{X}\sum_{i=1}^{n} e_i \\ &= 0 \end{aligned}$$

So

$$\underbrace{\sum_{i=1}^{n} y_i^2}_{\text{Total SS}} = \underbrace{b^2 \sum_{i=1}^{n} x_i^2}_{\text{Explained SS}} + \underbrace{\sum_{i=1}^{n} e_i^2}_{\text{Residual SS}}$$

Definition (R-square) *R-square, the coefficient of determination, can be defined as*

$$R^2 = \frac{ESS}{TSS}$$

$$= 1 - \frac{RSS}{TSS}$$

When R^2 should be large?

(1) The proportion of ESS is large, then R^2 should be large.

(2) When the slope b is large, the R^2 should be large; R^2 is determined by the model specified. In other words, R^2 will be large if the regression model has a large slope b, and vice versa.

(3) R^2 rises whenever an independent variable is added.

Ex. $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$ where Y: rice production in Korea, X_1: precipitation in Korea, and X_2: precipitation in South Korea

Figure 6.1.1 R^2 *of the regression model* $GDP = f(t)$

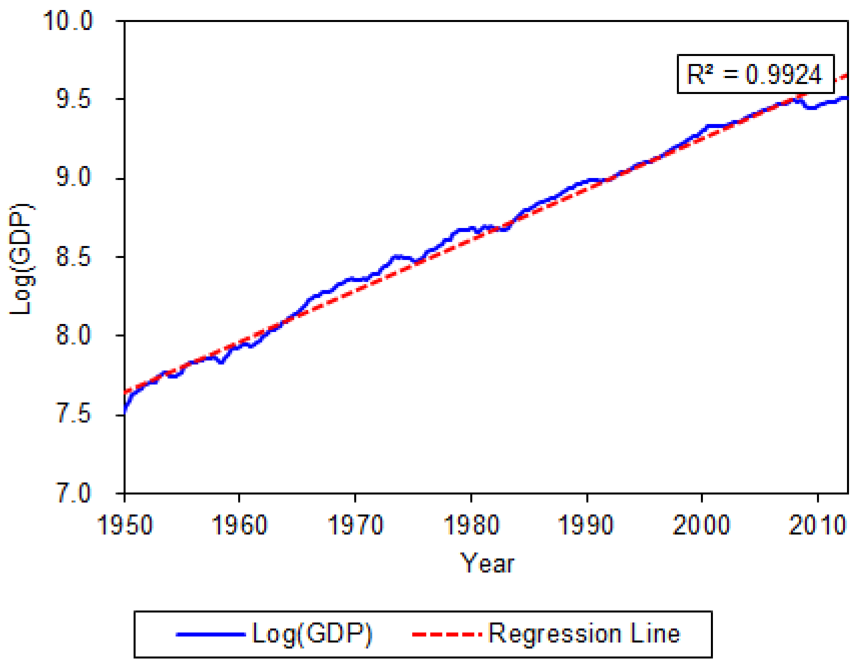
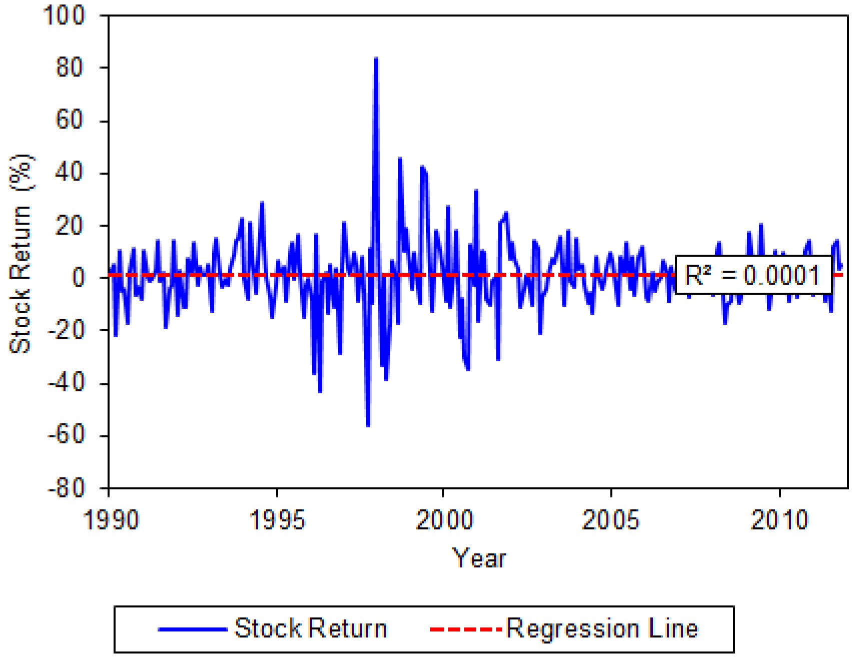


Figure 6.1.2 *R^2 of the regression model $r_t = f(t)$*



1.2 Adjusted R-square

Definition (Adjusted R-square) *Adjusted R-square can be defined as*

$$\bar{R}^2 = 1 - \frac{\frac{RSS}{n-k}}{\frac{TSS}{n-1}}$$

Prediction using estimated regression model; suppose we obtained the following regression

$$Y_i = a + bX_i + e_i$$

Where

Y_i : Consumption

X_i : Income

The predicted value of Y_i is $\hat{Y}_i = a + bX_i$, so if one wants to know the point prediction of consumption for a specific value of X_i, then it is easy to obtain

$$\hat{Y}_1 = a + bX_1$$

$$\hat{Y}_2 = a + bX_2$$

$$\vdots$$

$$\hat{Y}_n = a + bX_n \leftarrow n \text{ is the last sample point}$$

$$\hat{Y}_{n+1} = a + bX_{n+1} \leftarrow \text{unknown (income of the next year)}$$

Confidence interval of the forecast does not exist; therefore, it is impossible to obtain the predicted value of the next year's consumption.

Figure 6.1.3 *Regression Model and Time-series Model*

Pros and Cons	Model	
	Regression	Time-series
Pros	Economic theory can be reflected in the model.	Forecasting is very easy.
Cons	Forecasting is very difficult.	Economic theory cannot be reflected in the model.

2 Multicollinearity Problem

2.1 Symptom

When two or more independent variables are correlated (co-linear) then we have the following difficulties.

(1) High R^2 but few statistically significant parameter estimates.

(2) Small changes in the sample size cause big changes in the regression results.

(3) The signs of estimates are often counter-intuitive; not realistic.

Ex. $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$ where Y: consumption, X_1: income, and X_2: wealth. If the model is estimated, R^2 will be more than 0.9 with $b_1 = 1.2$ and $b_2 = -0.3$; "Less wealthier, more consumption" is counter-intuitive.

2.2 Remedy

(1) Collect more data if it is possible.

(2) Deleting an independent variable to alleviate the multicollinearity problem may induce specification errors.

Cf. (Property of estimators) Still the OLS estimators are, however, BLUE.

Cf. (Multivariate analysis) Factor Analysis? Principal Component Analysis?

Exercise *Estimate below regression models.*

$$IPG_i = \beta_0 + \beta_1 FYFF_i + \beta_2 FYGM3_i + u_i$$

$$IPG_i = \beta_0 + \beta_1 FYFF_i + u_i$$

$$IPG_i = \beta_0 + \beta_1 FYGM3_i + u_i$$

IPG_i : Log return of Industrial Production

$FYFF_i$: Federal Funds Interest Rate

$FYGM3_i$: 3 Month Treasury Bill Interest Rate

3 Structure Change

Suppose we want to test the structural change with the null

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

And the alternative

$$Y_i = \alpha_1 d_1 + \alpha_2 d_2 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

H_0 : $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$

H_1 : $\alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$

Exercise *Estimate below null (6.3) and alternative (6.4) models.*

$$IPG_i = \alpha + \beta FYFF_i + u_i \tag{6.3}$$

$$IPG_i = \alpha_1 d_1 + \alpha_2 d_2 + \beta_1 FYFF_{1i} + \beta_2 FYFF_{2i} + u_i \tag{6.4}$$

Suppose t^*: end of 1983 then

Figure 6.3.1 *Estimated Line w/o Structure Change (Y=0.0092+1.1453X)*

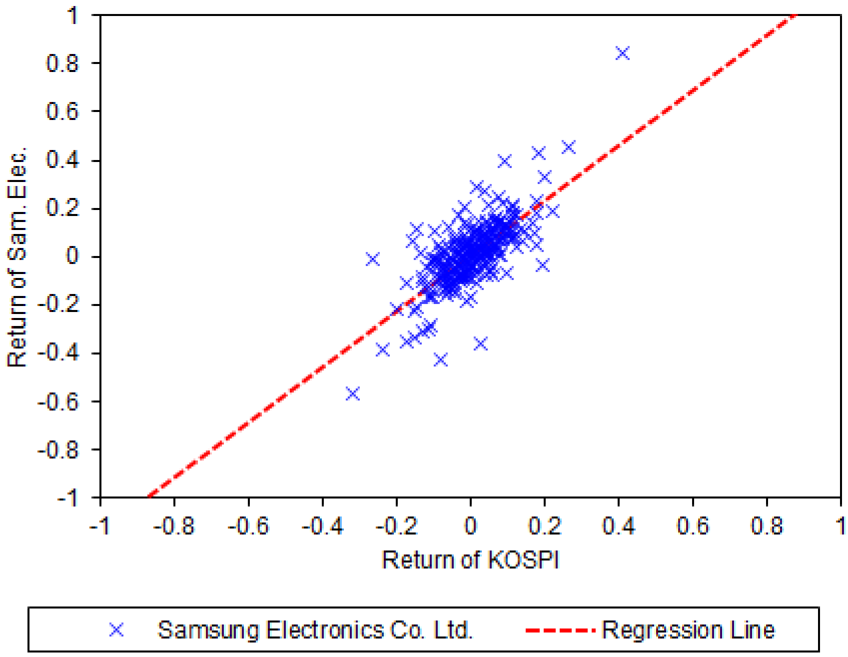


Figure 6.3.2 *Structure Change; slope determined* $(Y=0.0095+1.2677X_1+0.9738X_2)$

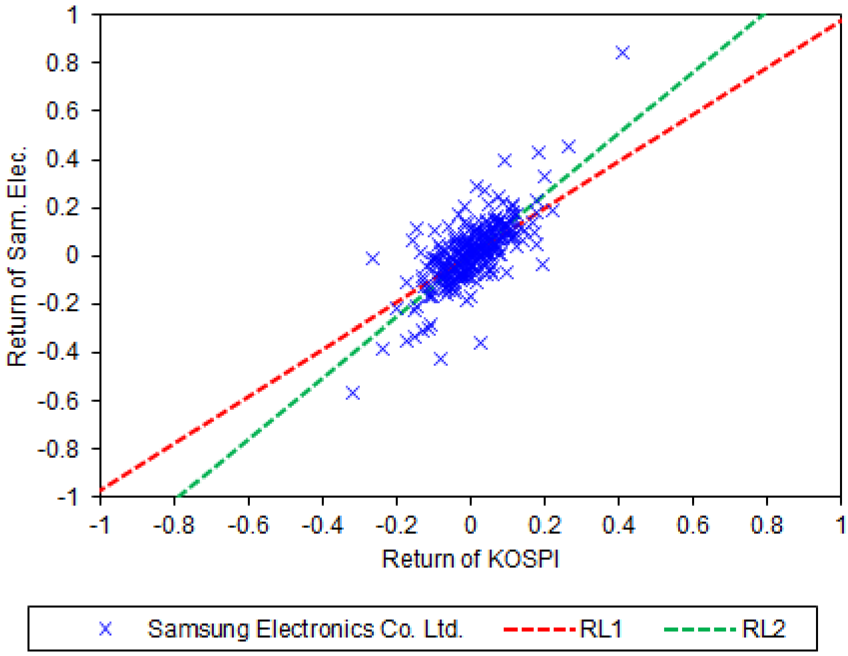


Figure 6.3.3 *St. Ch.; entirely determined* $(Y=0.0139d_1+0.0060d_2+1.2672X_1+0.9764X_2)$

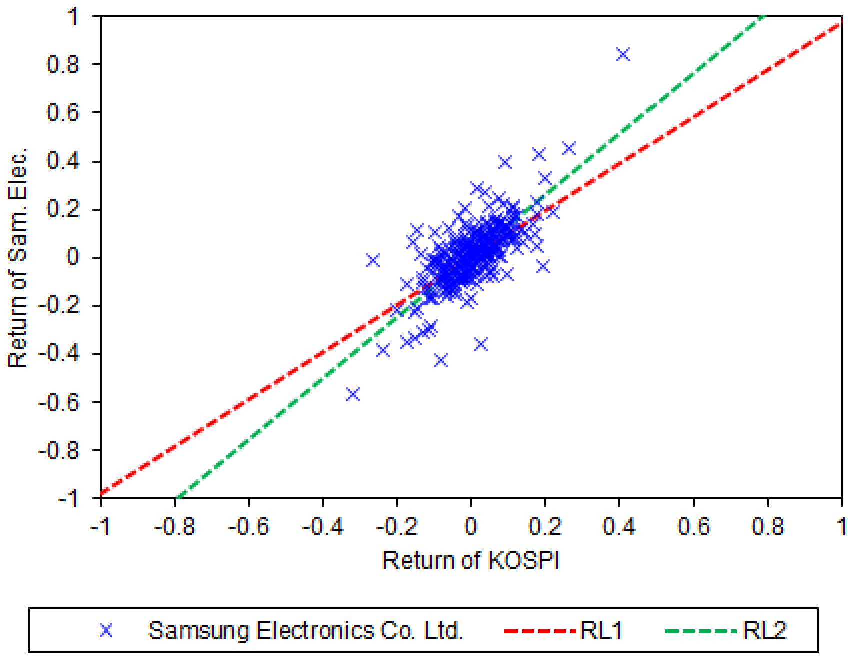


Figure 6.3.4 *Required data for Regression model w/ Structure Change*

date	kospi	samsung	kospig	samsungg	d1	d2	samsungg1	samsungg2
19900131	896.16	42800						
19900228	861.59	43500	-0.0393	0.0162	1	0	-0.0393	0.0000
19900331	840.89	45700	-0.0243	0.0493	1	0	-0.0243	0.0000
19900430	688.66	36700	-0.1997	-0.2193	1	0	-0.1997	0.0000
19900531	797.95	40800	0.1473	0.1059	1	0	0.1473	0.0000
...	...	...	...	...	...	...	...	...
20000131	943.88	279000	-0.0854	0.0477	0	1	0.0000	-0.0854
20000229	828.38	256000	-0.1305	-0.0860	0	1	0.0000	-0.1305
20000331	860.94	335000	0.0386	0.2690	0	1	0.0000	0.0386
20000428	725.39	300000	-0.1713	-0.1103	0	1	0.0000	-0.1713
20000531	731.88	308000	0.0089	0.0263	0	1	0.0000	0.0089
...	...	...	...	...	...	...	...	...

4 SAS Code

Code 6.4.1 *Regression model estimation w/ Structure Change*

```
data ip;
        infile "c:₩ip.prn";
        input month ip;
        ipg=dif(log(ip))*1200;
run;

data fyff;
        infile "c:₩fyff.prn";
        input month fyff;
run;

data fygm3;
        infile "c:₩fygm3.prn";
        input month fygm3;
run;

data all1;
        merge ip fyff fygm3;
        by month;
        if month<=198312 then do;
                d1=1;
                d2=0;
                fyff1=fyff;
                fyff2=0;
        end;
        else do;
                d1=0;
                d2=1;
                fyff1=0;
                fyff2=fyff;
        end;
        obs=_n_;
run;

proc reg data=all1;
        model ipg=fyff fygm3/dw;
run;

proc reg data=all1;
        model ipg=fyff1 fyff2;
        test fyff1=fyff2;
run;

proc reg data=all1;
        model ipg=d1 d2 fyff1 fyff2/noint;
        test d1=d2,fyff1=fyff2;
run;

quit;
```

ASSIGNMENT 6

Exercise 6.1 (Spurious Regression) *Suppose the variables below for* $t \in \{1,2,\cdots,2000\}$ *and the regression model (6.1), which includes those variables*

$$\begin{aligned}
X_t &= X_{t-1} + \varepsilon_{1t} \\
Y_t &= Y_{t-1} + \varepsilon_{2t} \\
\varepsilon_{1t} &\sim^{iid} \mathcal{N}(0,1) \\
\varepsilon_{2t} &\sim^{iid} \mathcal{N}(0,1) \\
X_1 &= \varepsilon_{11} \\
Y_1 &= \varepsilon_{21} \\
Y_t &= \beta_0 + \beta_1 X_t + u_t \qquad (6.1)
\end{aligned}$$

(1) Generate ε_{1t} and ε_{2t} sequence with the seeds 100 (for ε_{1t}) and 200 (for ε_{2t}) and plot each ε_{1t}, ε_{2t}, X_t, and Y_t against t.

(2) Run (6.1) and obtain the table of estimated coefficients; if the variables are generated properly, the model will be estimated as below with $R^2 = 0.1984$; what can be inferred from this result?

$$Y_t = -\underbrace{3.2821}_{(-8.18)} - \underbrace{0.4978}_{(-22.23)} X_t + e_t$$

(3) Run the same model with the "*first-differenced*" variables that can be defined as below and the model should be estimated with $R^2 = 0.0002$

$$\begin{aligned}
\Delta X_t &:= X_t - X_{t-1} \\
&= \varepsilon_{1t} \\
\Delta Y_t &:= Y_t - Y_{t-1} \\
&= \varepsilon_{2t}
\end{aligned}$$

$$\Delta Y_t = \beta_0 + \beta_1 \Delta X_t + u_t$$

$$\Delta Y_t = -\underbrace{0.0170}_{(-0.75)} - \underbrace{0.0135}_{(-0.60)} \Delta X_t + e_t$$

(4) Plot the scatter diagram of the independent (x-axis) and dependent (y-axis) variables above in the "*PROC REG*" procedure.

(5) Obtain autocorrelation function of those four variables above and describe the properties of them.

(6) Briefly write down what you discovered from above analysis.

Exercise 6.2 (Time-series Simulation) *Generate the variables below for* $t \in \{1,2,\cdots,2000\}$ *with the seed 300 (for* ε_t*)*

$$Y_t = 0.2 + 0.5Y_{t-1} + 0.3Y_{t-2} + \varepsilon_t$$

$$\varepsilon_t \sim^{iid} \mathcal{N}(0,1)$$

(1) Generate and plot them with Y_{-1} and Y_0 set to be zero and obtain autocorrelation and partial autocorrelation function of them.

(2) Check whether the variable is "*White-noise*" or not through the table of autocorrelation check from SAS.

(3) Write down the model estimated; AR(1) and AR(2). If the variable is generated appropriately, SAS will display the numbers below.

$$Y_t - \underbrace{0.7307}_{(47.82)} Y_{t-1} = \underbrace{1.1157}_{(12.67)} + \varepsilon_t$$

$$Y_t - \underbrace{0.5013}_{(23.59)} Y_{t-1} - \underbrace{0.3141}_{(14.78)} Y_{t-2} = \underbrace{1.1189}_{(9.22)} + \varepsilon_t$$

(4) Obtain Akaike Information Criteria (AIC) and Schwartz Bayesian Information Criteria (AIC) for the AR(1), ARMA(1,1), and AR(2) models and compare them; describe which one is better.

(5) Generate and plot Y_t with following structure and the seed 400 and do the same thing above as below, with ε_0 set to be zero also; estimate the AR(1), ARMA(1,1), AR(2), and ARMA(2,1) models.

$$Y_t = 0.2 + 0.5Y_{t-1} + 0.3Y_{t-2} + \varepsilon_t - 0.2\varepsilon_{t-1}$$

$$\varepsilon_t \sim^{iid} \mathcal{N}(0,1)$$

Code Assignment 6

```
data assign1;
        do t=1 to 2000;
                dy=rannor(100);
                dx=rannor(200);
                output;
        end;
run;

data assign1;
        set assign1;
        y+dy;
        x+dx;
run;

proc gplot data=assign1;
        symbol i=join;
        plot y*t x*t/overlay;
run;

proc gplot data=assign1;
        plot dy*t dx*t/overlay;
run;

proc reg data=assign1;
        symbol;
        model y=x;
        plot y*x;
run;

proc reg data=assign1;
        model dy=dx;
        plot dy*dx;
run;

proc arima data=assign1;
        identify var=y;
        identify var=x;
        identify var=dy;
        identify var=dx;
run;

data assign21;
        alpha0=0.2;
        alpha1=0.5;
        alpha2=0.3;
        y1=0;
        y2=0;
        do t=1 to 2000;
                e=rannor(300);
                y=alpha0+alpha1*y1+alpha2*y2+e;
```

```
                output;
                y2=y1;
                y1=y;
        end;
run;

proc gplot data=assign21;
        plot y*t;
run;

proc arima data=assign21;
        identify var=y;
        estimate p=1;
run;

proc arima data=assign21;
        identify var=y;
        estimate p=2;
run;

data assign22;
        alpha0=0.2;
        alpha1=0.5;
        alpha2=0.3;
        beta1=-0.2;
        y1=0;
        y2=0;
        e1=0;
        do t=1 to 2000;
                e=rannor(400);
                y=alpha0+alpha1*y1+alpha2*y2+e+beta1*e1;
                output;
                y2=y1;
                y1=y;
                e1=e;
        end;
run;

proc gplot data=assign22;
        plot y*t;
run;

proc arima data=assign22;
        identify var=y;
        estimate p=1;
run;

proc arima data=assign22;
        identify var=y;
        estimate p=2;
run;
```

```
proc arima data=assign22;
        identify var=y;
        estimate p=2 q=1;
run;

quit;
```

LECTURE NOTE 7

1 Heteroskedasticity

1.1 Heteroskedasticity Problem

Suppose the following model

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

Where u_i has mean 0 and variance σ^2 (Figure 3.1.3)

Classical Assumption (3) is

(3.a) $\forall i,\ \mathbb{E}(u_i^2) = \sigma^2$

(3.b) $\forall i \neq j,\ \mathbb{E}(u_i u_j) = 0$

Term 'Heteroskedasticity' and 'Homoskedasticity' can be depicted as

$$\mathbb{E}(u_i^2) = \sigma_i^2 \leftarrow \text{Heteroskedasiticy} \tag{7.1}$$

$$\mathbb{E}(u_i^2) = \sigma^2 \leftarrow \text{Homoskedasticity}$$

By (7.1) the model violates Classical Assumption (3.a)

Suppose that σ_i^2 are all known a priori $\forall i = 1,2,\cdots,n$; then, by taking the following transformation as

$$Y_i/\sigma_i = (\alpha + \beta X_i + u_i)/\sigma_i$$

$$\frac{Y_i}{\sigma_i} = \frac{\alpha}{\sigma_i} + \beta\frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i} \leftarrow \text{new error}$$

$$Y_i^* = \alpha^* + \beta X_i^* + u_i^* \leftarrow \text{new model}$$

Thereby properties of new error will be different from original error as

$$\begin{aligned}
\mathbb{E}(u_i^*) &= \mathbb{E}\left(\frac{u_i}{\sigma_i}\right) \\
&= 0 \\
\mathbb{V}ar(u_i^*) &= \mathbb{E}[u_i^* - \mathbb{E}(u_i^*)]^2 \\
&= \mathbb{E}[(u_i^*)^2] \\
&= \mathbb{E}\left[\left(\frac{u_i}{\sigma_i}\right)^2\right] \\
&= \mathbb{E}\left(\frac{u_i^2}{\sigma_i^2}\right) \\
&= \frac{1}{\sigma_i^2}\mathbb{E}(u_i^2) \\
&= \frac{1}{\sigma_i^2}\sigma_i^2 \\
&= 1 \leftarrow \text{homoskedastic error term}
\end{aligned}$$

Since the new regression model has the homoscedastic error term, the OLS estimators of the parameters in the new model become BLUE.

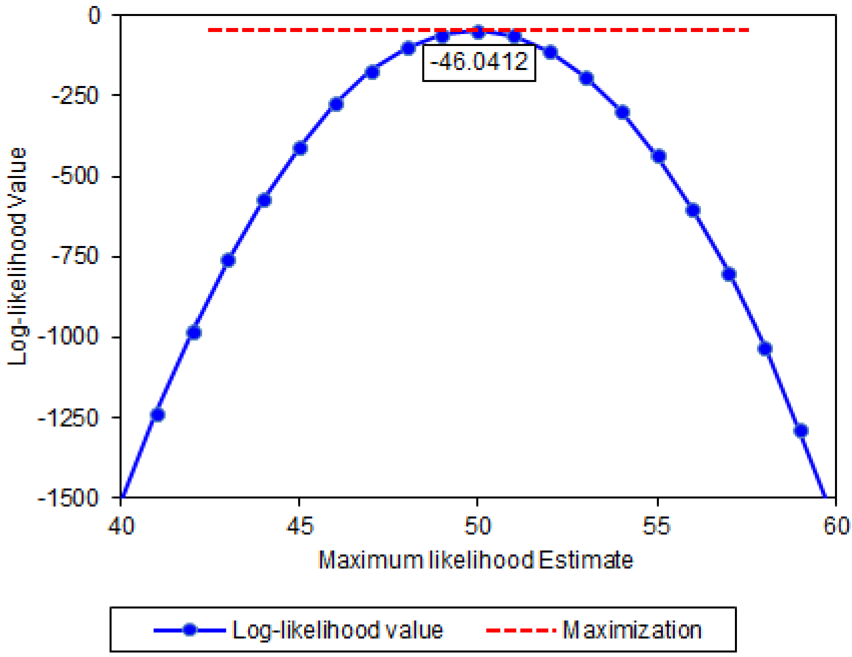
Suppose σ_i^2 are unknown; then, there are $(n+2)$ unknowns while there are only n observations; which means $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2, \alpha, \beta$

Suppose $\sigma_i^2 = \delta X_i$ or $\sigma_i^2 = \delta_0 + \delta_1 X_i$; if this is the case, then one may use Maximum Likelihood Estimation (MLE) method to estimate α, β, δ_0, and δ_1

Below model is non-linear as

$$\begin{aligned}
Y_i &= \alpha + \beta X_i + u_i \\
u_i &\sim \mathcal{N}(0, \sigma_i^2) \\
\sigma_i^2 &= \delta_0 + \delta_1 X_i
\end{aligned}$$

Figure 7.1.1 *Idea of Maximum Likelihood Estimation when* $Y_i = \hat{\mu} + u_i$, $Y \sim \mathcal{N}(50, 1^2)$



1.2 Testing for Presence of Heteroskedasticity

Let the model be

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

To test the presence of Heteroskedasticity, we may use the following model as

$$Y_i = a + bX_i + e_i \tag{7.2}$$

$$e_i^2 = \gamma_0 + \gamma_1 X_i + \gamma_2 X_i^2 + \gamma_3 X_i^3 + \hat{v}_i$$

Where e_i is the residual of the regression equation (7.2). The Homoscedasticity u_i can be described as

$$H_0 \quad : \quad \gamma_1 = \gamma_2 = \gamma_3 = 0$$

$$H_1 \quad : \quad \text{At least one of them is non-zero.}$$

The consequence of using the OLS method to estimate parameters ignoring the presence of Heteroskedasticity since

$$\mathbb{v}ar(b) = \mathbb{E}(b - \beta)^2$$

$$= \left(\frac{\sum x_i u_i}{\sum x_i^2}\right)^2$$

$$= \frac{1}{(\sum x_i^2)^2}\mathbb{E}\left(\sum_{i=1}^{n} x_i^2 u_i^2 + 2\sum_{i=1}^{n}\sum_{j>i}^{n} x_i x_j u_i u_j\right)$$

$$= \frac{1}{(\sum x_i^2)^2}\sum_{i=1}^{n} x_i^2 \underbrace{\mathbb{E}(u_i^2)}_{\sigma_i^2}$$

$$= \frac{1}{(\sum x_i^2)^2}\sum_{i=1}^{n} x_i^2 \sigma_i^2 \leftarrow \text{Heteroskedasticity}$$

Note that

$$\mathbb{v}ar(b) = \frac{\sigma^2}{\sum x_i^2} \leftarrow \text{Homoskedasticity}$$

$$\widehat{\mathbb{v}ar}(b) = \frac{s^2}{\sum x_i^2}$$

Thus, the true variance estimator should be

$$\widehat{\mathbb{v}ar}(b) = \frac{1}{(\sum x_i^2)^2}\sum_{i=1}^{n} x_i^2 s_i^2$$

Since s_i^2 is impossible to calculate, the variance of b, by which t statistic for hypothesis testing is constructed, is not applicable.

2 AR(1) Process

2.1. Autocovariance Function of AR(1)

Suppose the following model

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

Where

$$\mathbb{E}(u_i) = 0$$

$$\mathbb{E}(u_i u_j) = \sigma^2 \leftarrow \forall i = j$$

$$\neq 0 \leftarrow \forall i \neq j$$

$$\mathbb{E}(u_i u_{i-1}) \neq 0$$

$$u_i = \phi u_{i-1} + \varepsilon_i \tag{7.3}$$

(7.3) is first-order Auto-Regressive (AR) model; which can be denoted as AR(1); where ε_i is *independent and identically distributed* (*i.i.d.*) error term with mean zero and variance σ_ε^2

$$\begin{aligned}
\mathbb{E}(\varepsilon_i) &= 0 \\
\mathbb{E}(\varepsilon_i \varepsilon_j) &= \sigma_\varepsilon^2 \leftarrow \forall i = j \\
&= 0 \leftarrow \forall i \neq j
\end{aligned}$$

(7.3) can be expanded as

$$\begin{aligned}
u_{i-1} &= \phi u_{i-2} + \varepsilon_{i-1} \\
u_{i-2} &= \phi u_{i-3} + \varepsilon_{i-2} \\
&\vdots
\end{aligned}$$

Then through this structure u_i can be rewritten as

$$u_i = \varepsilon_i + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots$$

Thus the structure of the variance of u_i is

$$\begin{aligned}
\gamma_0 &= \mathbb{v}\mathrm{ar}(u_i) \\
&= \mathbb{E}(\varepsilon_i + \phi \varepsilon_{i-1} + \phi^2 \varepsilon_{t-2} + \cdots) \\
&= \sigma_\varepsilon^2 + \phi^2 \sigma_\varepsilon^2 + \phi^4 \sigma_\varepsilon^2 + \cdots \\
&= \frac{\sigma_\varepsilon^2}{1 - \phi^2} \leftarrow \forall |\phi| < 1 \\
\gamma_1 &= \mathbb{c}ov(u_i u_{i-1}) \\
&= \mathbb{E}(u_i u_{i-1}) \\
&= \mathbb{E}[(\varepsilon_i + \phi \varepsilon_{i-1} + \phi^2 \varepsilon_{t-2} + \cdots)(\varepsilon_{i-1} + \phi \varepsilon_{i-2} + \phi^2 \varepsilon_{t-3} + \cdots)] \\
&= \phi \sigma_\varepsilon^2 + \phi^3 \sigma_\varepsilon^2 + \phi^5 \sigma_\varepsilon^2 + \cdots \\
&= \frac{\phi \sigma_\varepsilon^2}{1 - \phi^2} \leftarrow \text{first} - \text{order Autocovariance} \\
\gamma_2 &= \mathbb{c}ov(u_i u_{i-2}) \\
&= \mathbb{E}(u_i u_{i-2})
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[(\varepsilon_i + \phi\varepsilon_{i-1} + \phi^2\varepsilon_{t-2} + \cdots)(\varepsilon_{i-2} + \phi\varepsilon_{i-3} + \phi^2\varepsilon_{t-4} + \cdots)] \\
&= \phi^2\sigma_\varepsilon^2 + \phi^4\sigma_\varepsilon^2 + \phi^6\sigma_\varepsilon^2 + \cdots \\
&= \frac{\phi^2\sigma_\varepsilon^2}{1-\phi^2} \leftarrow \text{second} - \text{order Autocovariance} \\
\gamma_3 &= \frac{\phi^3\sigma_\varepsilon^2}{1-\phi^2} \\
\gamma_4 &= \frac{\phi^4\sigma_\varepsilon^2}{1-\phi^2} \\
&\vdots
\end{aligned}
$$

With those $\gamma_1, \gamma_2, \cdots$, Autocovariance function can be constructed.

2.2 Autocorrelation Function of AR(1)

$$
\text{Autocorrelation} = \frac{\text{Autocovariance}}{\text{Variance}}
$$

Therefore

$$
\begin{aligned}
\rho_1 &= \text{first} - \text{order Autocorrelation} \\
&= \frac{\gamma_1}{\gamma_0} \\
&= \frac{\frac{\phi\sigma_\varepsilon^2}{1-\phi^2}}{\frac{\sigma_\varepsilon^2}{1-\phi^2}} \\
&= \phi \\
\rho_2 &= \frac{\gamma_2}{\gamma_0} \\
&= \frac{\frac{\phi^2\sigma_\varepsilon^2}{1-\phi^2}}{\frac{\sigma_\varepsilon^2}{1-\phi^2}} \\
&= \phi^2 \\
\rho_3 &= \phi^3 \\
&\vdots
\end{aligned}
$$

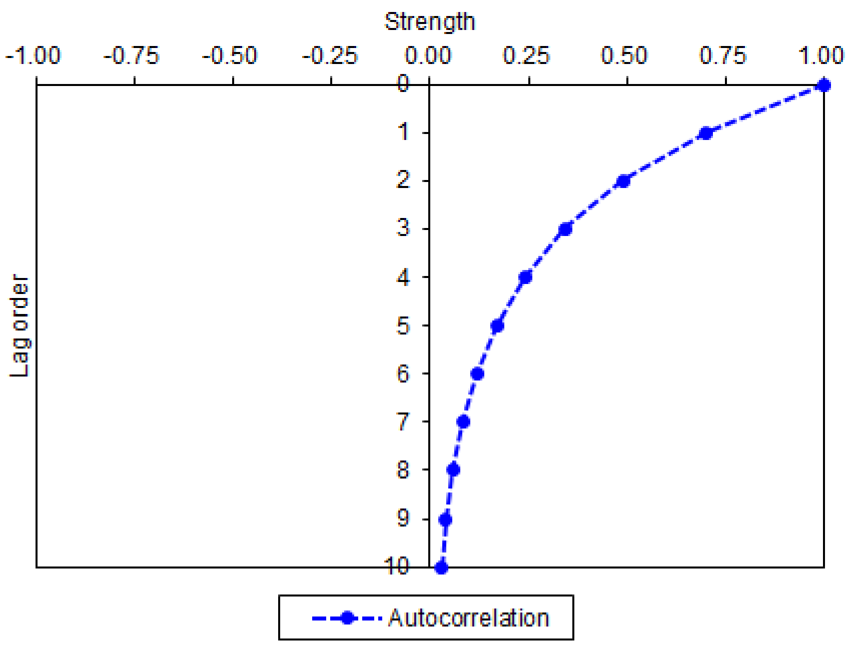
Remind Figure 6.1.3 again.

Cf. (Regression model) $Y_{n+1} = \alpha + \beta X_{n+1} + u_{n+1}$; next year's income? How? There will not be confidence interval for Y_{n+1} at all.

Cf. (AR(2) model) $u_i = \phi_1 u_{i-1} + \phi_2 u_{i-2} + \varepsilon_i$

Cf. (MA(1) model) $u_i = \varepsilon_i + \theta \varepsilon_{i-1}$

Figure 7.2.1 *Autocorrelation function of AR(1) Process ($\phi = 0.7$)*



3 SAS Code

Exercise *Estimate below regression models.*

$$LOGGMDC_i = \beta_0 + \beta_1 LOGIP_i + u_i$$

$$u_i^2 = \gamma_0 + \gamma_1 LOGIP_i + \gamma_2 LOGIP_i^2 + v_i^2$$

$GMDC_i$: Consumer Price Deflator

IP_i : Industrial Production

Code 7.3.1 *Testing for the presence of Heteroskedasticity*

```
data gmdc;
        infile "c:₩gmdc.prn";
        input month gmdc;
        loggmdc=log(gmdc);
run;

data ip;
        infile "c:₩ip.prn";
        input month ip;
        logip=log(ip);
run;

data all1;
        merge gmdc ip;
        by month;
run;

proc reg data=all1;
        model loggmdc=logip;
        output out=out1 r=resid;
run;

data out1;
        set out1;
        resid2=resid**2;
        logip2=logip**2;
        obs=_n_;
run;

proc reg data=out1;
        model resid2=logip logip2;
        test logip,logip2;
run;

proc gplot data=out1;
        symbol i=join;
        plot resid2*obs;
run;

quit;
```

ASSIGNMENT 7

Exercise 7.1 (Dickey-Fuller Stationarity Test) *Suppose the variables below for* $t \in \{1,2,\cdots,2000\}$

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_{1t} \\ Y_t &= Y_{t-1} + \varepsilon_{2t} \\ \varepsilon_{1t} &\sim^{iid} \mathcal{N}(0,1) \\ \varepsilon_{2t} &\sim^{iid} \mathcal{N}(0,1) \\ X_1 &= \varepsilon_{11} \\ Y_1 &= \varepsilon_{21} \end{aligned}$$

(1) Try three models of X (and Y also) below and obtain the estimates of β_1 (i.e. $\hat{\beta}_1$)

$$\begin{aligned} \Delta X_t &:= X_t - X_{t-1} \\ &= \beta_1 X_{t-1} + u_t \\ &= \beta_0 + \beta_1 X_{t-1} + u_t \\ &= \beta_0 + \gamma t + \beta_1 X_{t-1} + u_t \end{aligned}$$

(2) Test the hypothesis $H_0: \beta_1 = 0.0$; is there any $\hat{\beta}_1$ that can reject the null?

(3) Explain the implication of above tests.

(4) Try another three regression models of X, Y below and obtain same coefficients mentioned already.

$$\begin{aligned} \Delta^2 X_t &:= \Delta X_t - \Delta X_{t-1} \\ &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= \beta_1 \Delta X_{t-1} + u_t \\ &= \beta_0 + \beta_1 \Delta X_{t-1} + u_t \end{aligned}$$

$$= \quad \beta_0 + \gamma t + \beta_1 \Delta X_{t-1} + u_t$$

(5) Test the hypothesis already referred at (2)

(6) Compare the t statistic can be observed from "*PROC REG*" procedure and the τ statistic can be obtained from "*PROC ARIMA*" procedure.

Code Assignment 7

```
/**************************************************/
/*Augmented Dickey-Fuller Stationarity Test*/
/**************************************************/

data assign71;
        do t=1 to 2000;
                dx=rannor(100);
                dy=rannor(200);
                output;
        end;
run;

data assign71;
        set assign71;
        x+dx;
        y+dy;
        x1=lag1(x);
        y1=lag1(y);
        ddx=dif(dx);
        ddy=dif(dy);
        dx1=lag1(dx);
        dy1=lag1(dy);
run;

/**************************************************/
/*Test for Variable X - Zero Mean, Single Mean, Trend*/
/**************************************************/

proc reg data=assign71;
        model dx=x1/noint;
        model dx=x1;
        model dx=t x1;
run;

/**************************************************/
/*Test for Variable ΔX - Zero Mean, Single Mean, Trend*/
/**************************************************/

proc reg data=assign71;
        model ddx=dx1/noint;
        model ddx=dx1;
        model ddx=t dx1;
run;

/**************************************************/
/*Test for Variable Y - Zero Mean, Single Mean, Trend*/
/**************************************************/

proc reg data=assign71;
        model dy=y1/noint;
```

```
          model dy=y1;
          model dy=t y1;
run;

/**************************************************/
/*Test for Variable  ΔY - Zero Mean, Single Mean, Trend*/
/**************************************************/

proc reg data=assign71;
          model ddy=dy1/noint;
          model ddy=dy1;
          model ddy=t dy1;
run;

/**************************************************/
/*PROC ARIMA Application for ADF Test*/
/**************************************************/

proc arima data=assign71;
          identify var=x stationarity=(adf);
          identify var=dx stationarity=(adf);
          identify var=y stationarity=(adf);
          identify var=dy stationarity=(adf);
run;

quit;
```

LECTURE NOTE 8

1 Partial Autocorrelation

In general, the AR(p) model can be represented as

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t$$

All AR models have the autocorrelation functions that are hard to identify the order of AR through eye inspection. Therefore, in order to identify the order of AR, the partial autocorrelation function is used.

The partial autocorrelation function is the collection of the parameters, α_{11}, α_{22}, $\cdots$, α_{pp} of AR models.

The partial autocorrelations are

$$\begin{aligned} Y_t &= \boldsymbol{\alpha_{11}} Y_{t-1} + \varepsilon_t \\ Y_t &= \alpha_{21} Y_{t-1} + \boldsymbol{\alpha_{22}} Y_{t-2} + \varepsilon_t \\ &\vdots \\ Y_t &= \alpha_{p1} Y_{t-1} + \alpha_{p2} Y_{t-2} + \cdots + \boldsymbol{\alpha_{pp}} Y_{t-p} + \varepsilon_t \end{aligned}$$

Figure 8.1.1 shows typical structure of partial autocorrelation function of AR(1) structure; compare with Figure 7.2.1, the autocorrelation function.

Figure 8.1.1 *Partial autocorrelation function of AR(1) Process* ($\phi = 0.5$)

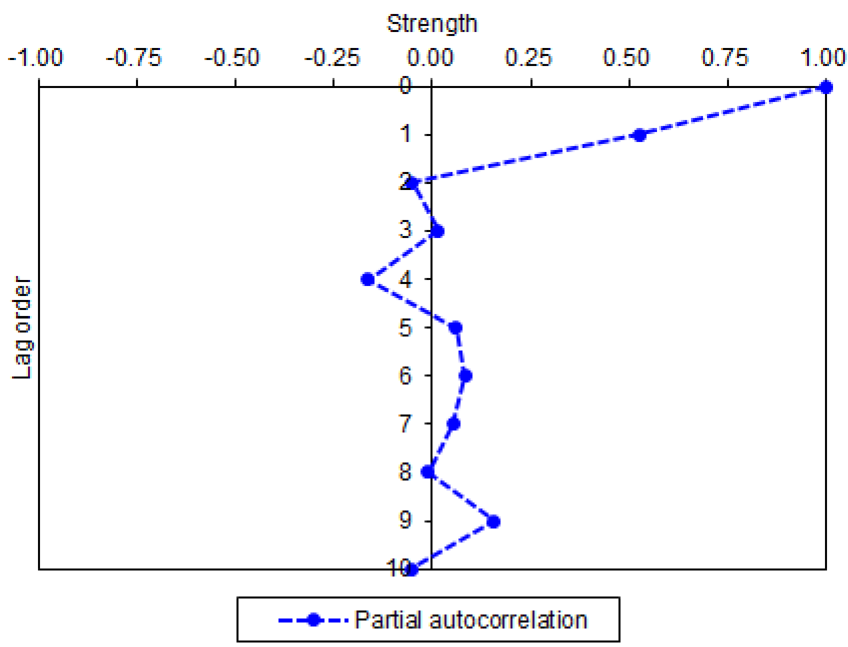


Figure 8.1.2 *AR(1) time-series* ($\phi = 0.5$)

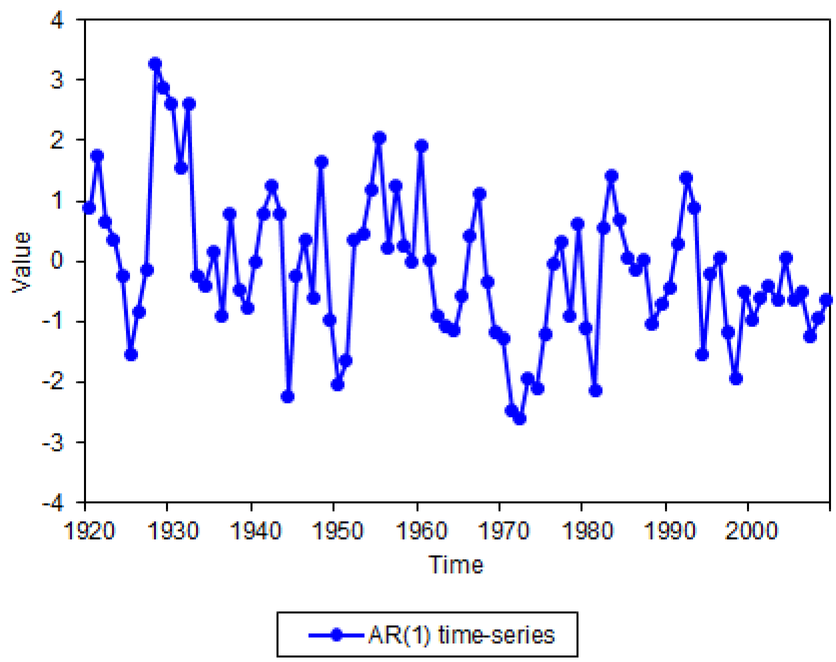
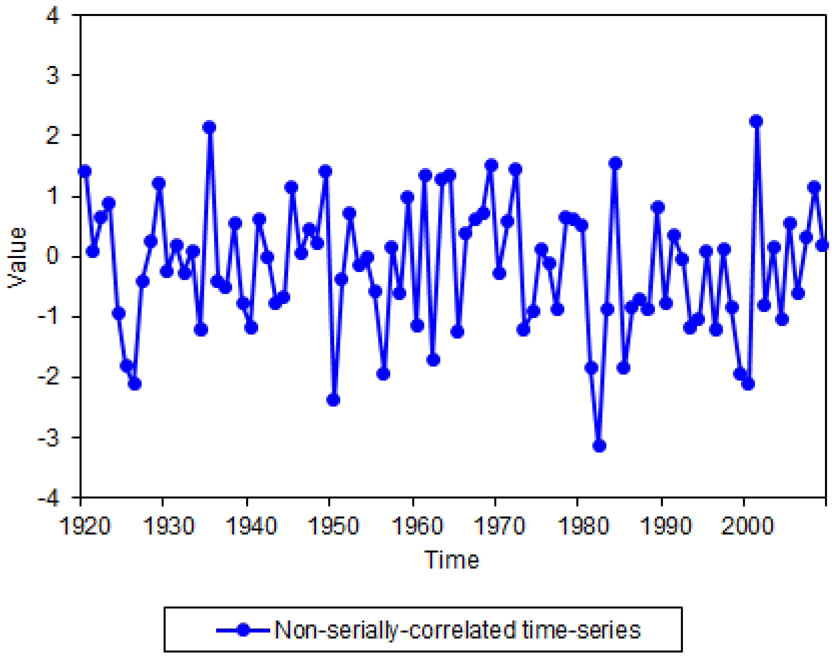


Figure 8.1.3 *Time-series w/o any timely correlated structure*



2 Forecasting of AR(p) Model

2.1 AR(1) Model

$$Y_t = \alpha Y_{t-1} + \varepsilon_t$$

Suppose we want to forecast

$$Y_{t+1} = \alpha Y_t + \varepsilon_{t+1}$$

Where ε_{t+1} is *i.i.d* white noise. Based on the information Y_1, Y_2, $\cdots$, Y_t, then we can estimate α with t number of observations, and denote it as $\hat{\alpha}$ and use the forecasting model.

$$Y_{t+1|t} = \hat{\alpha} Y_t$$

Where $t+1|t$ denotes that the forecast of $t+1$ period based on the information available at time t, and also

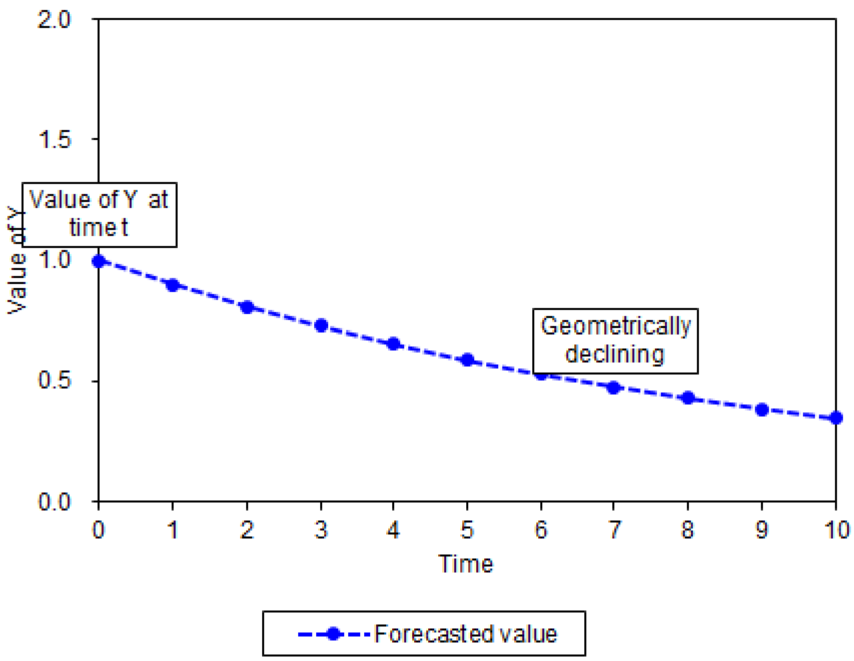
$$Y_{t+2} = \alpha Y_{t+1} + \varepsilon_{t+2}$$

$$Y_{t+2|t} = \hat{\alpha}Y_{t+1|t} + \varepsilon_{t+2|t}$$

Since we don't have any information about the white noise ε_{t+1} at the time t, $\varepsilon_{t+2|t} = 0$, so

$$\begin{aligned} Y_{t+2|t} &= \hat{\alpha}(\hat{\alpha}Y_t) \\ &= \hat{\alpha}^2 Y_t \\ Y_{t+3|t} &= \hat{\alpha}^3 Y_t \\ &\vdots \end{aligned}$$

Figure 8.2.1 *Forecasted value of* $Y_{t+p|t}$ *when* $Y_t = 1$ $(\hat{\alpha} = 0.9)$



2.2 AR(2) Model

$$\begin{aligned} Y_t &= \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t \\ Y_{t+1} &= \alpha_1 Y_t + \alpha_2 Y_{t-1} + \varepsilon_{t+1} \end{aligned}$$

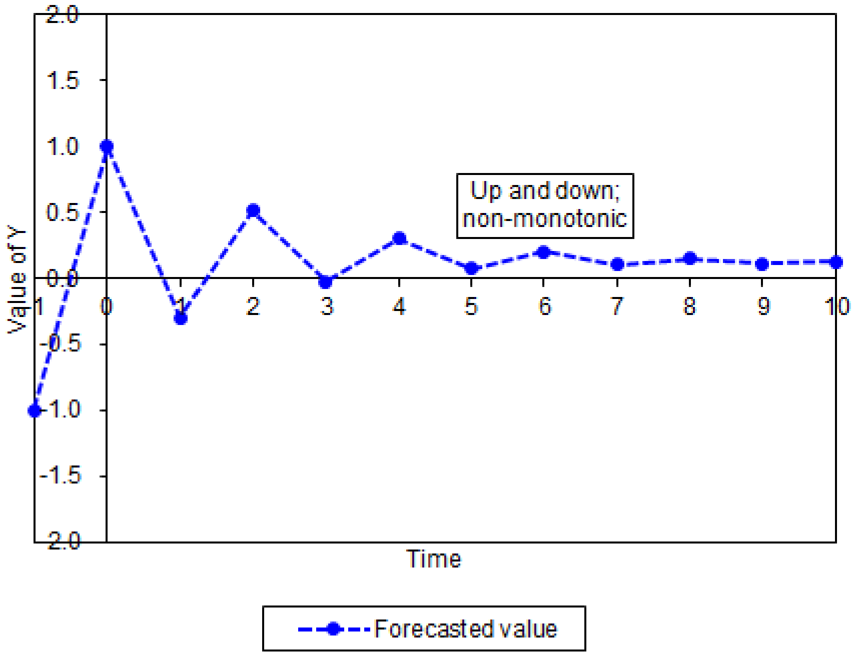
Therefore, since $\varepsilon_{t+1|t} = 0$

$$Y_{t+1|t} = \hat{\alpha}_1 Y_t + \hat{\alpha}_2 Y_{t-1}$$

Similarly, since $\varepsilon_{t+2|t} = 0$

$$
\begin{aligned}
Y_{t+2} &= \alpha_1 Y_{t+1} + \alpha_2 Y_t + \varepsilon_{t+2} \\
Y_{t+2|t} &= \hat{\alpha}_1 Y_{t+1|t} + \hat{\alpha}_2 Y_t \\
&= \hat{\alpha}_1(\hat{\alpha}_1 Y_t + \hat{\alpha}_2 Y_{t-1}) + \hat{\alpha}_2 Y_t \\
&= \hat{\alpha}_1^2 Y_t + \hat{\alpha}_1\hat{\alpha}_2 Y_{t-1} + \hat{\alpha}_2 Y_t \\
&= (\hat{\alpha}_1^2 + \hat{\alpha}_2) Y_t + \hat{\alpha}_1\hat{\alpha}_2 Y_{t-1}
\end{aligned}
$$

Figure 8.2.2 *Forecasted value of* $Y_{t+p|t}$ *when* $Y_{t-1} = -1,\ Y_t = 1\ (\hat{\alpha}_1 = 0.3,\ \hat{\alpha}_2 = 0.6)$



3 MA(1) Process

3.1 Autocovariance Function of MA(1)

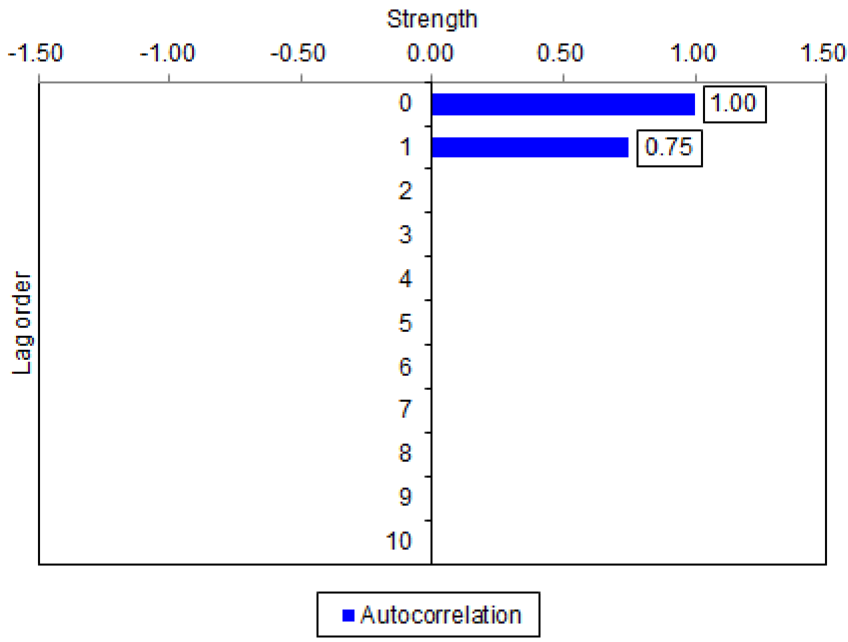
$$
\begin{aligned}
Y_t &= \varepsilon_t + \beta\varepsilon_{t-1} \\
\mathbb{V}ar(Y_t) &= \gamma_0 \\
&= \mathbb{E}(\varepsilon_t + \beta\varepsilon_{t-1})^2
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}(\varepsilon_t^2 + \beta^2\varepsilon_{t-1}^2 + 2\beta\varepsilon_t\varepsilon_{t-1}) \\
&= (1+\beta^2)\sigma_\varepsilon^2 \\
\mathbb{c}ov(Y_t, Y_{t-1}) &= \gamma_1 \\
&= \mathbb{E}[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-1} + \beta\varepsilon_{t-2})] \\
&= \beta\sigma_\varepsilon^2 \\
\mathbb{c}ov(Y_t, Y_{t-2}) &= \gamma_2 \\
&= \mathbb{E}[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-2} + \beta\varepsilon_{t-3})] \\
&= 0 \\
\mathbb{c}ov(Y_t, Y_{t-3}) &= \gamma_3 \\
&= 0 \\
&\vdots
\end{aligned}
$$

3.2 Autocorrelation Function of MA(1)

$$
\begin{aligned}
\mathbb{c}orr(Y_t, Y_{t-1}) &= \rho_1 \\
&= \frac{\beta\sigma_\varepsilon^2}{(1+\beta^2)\sigma_\varepsilon^2} \\
&= \frac{\beta}{1+\beta^2} \\
\mathbb{c}orr(Y_t, Y_{t-2}) &= \rho_2 \\
&= 0 \\
\mathbb{c}orr(Y_t, Y_{t-3}) &= \rho_3 \\
&= 0 \\
&\vdots
\end{aligned}
$$

Figure 8.3.1 *Autocorrelation function of MA(1) process ($\beta = 0.75$)*



4 MA(2) Process

Suppose below MA(2), short memory process, then

$$Y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}$$

4.1 Autocovariance Function of MA(2)

$$\begin{aligned}
\mathbb{v}ar(Y_t) &= \gamma_0 \\
&= \mathbb{E}(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2})^2 \\
&= \mathbb{E}[(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2})(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2})] \\
&= \mathbb{E}[\varepsilon_t^2 + \beta_1^2 \varepsilon_{t-1}^2 + \beta_2^2 \varepsilon_{t-2}^2 + 2(\beta_1 \varepsilon_t \varepsilon_{t-1} + \beta_2 \varepsilon_t \varepsilon_{t-2} + \beta_1 \beta_2 \varepsilon_{t-1} \varepsilon_{t-2})] \\
&= \mathbb{E}(\varepsilon_t^2 + \beta_1^2 \varepsilon_{t-1}^2 + \beta_2^2 \varepsilon_{t-2}^2) \\
&= (1 + \beta_1^2 + \beta_2^2)\sigma_\varepsilon^2
\end{aligned}$$

$$\begin{aligned}
\mathbb{c}ov(Y_t, Y_{t-1}) &= \gamma_1 \\
&= \mathbb{E}[(\varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2})(\varepsilon_{t-1} + \beta_1\varepsilon_{t-2} + \beta_2\varepsilon_{t-3})] \\
&= \mathbb{E}(\beta_1\varepsilon_{t-1}^2 + \beta_1\beta_2\varepsilon_{t-2}^2 + \cdots) \\
&= (\beta_1 + \beta_1\beta_2)\sigma_\varepsilon^2 \\
\mathbb{c}ov(Y_t, Y_{t-2}) &= \gamma_2 \\
&= \mathbb{E}[(\varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2})(\varepsilon_{t-2} + \beta_1\varepsilon_{t-3} + \beta_2\varepsilon_{t-4})] \\
&= \mathbb{E}(\beta_2\varepsilon_{t-2}^2 + \cdots) \\
&= \beta_2\sigma_\varepsilon^2 \\
\mathbb{c}ov(Y_t, Y_{t-3}) &= \gamma_3 \\
&= \mathbb{E}[(\varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2})(\varepsilon_{t-3} + \beta_1\varepsilon_{t-4} + \beta_2\varepsilon_{t-5})] \\
&= 0 \\
&\vdots
\end{aligned}$$

4.2 Autocorrelation Function of MA(2)

$$\begin{aligned}
\mathbb{c}orr(Y_t, Y_{t-1}) &= \rho_1 \\
&= \frac{\gamma_1}{\gamma_0} \\
&= \frac{(\beta_1 + \beta_1\beta_2)\sigma_\varepsilon^2}{(1 + \beta_1^2 + \beta_2^2)\sigma_\varepsilon^2} \\
&= \frac{\beta_1 + \beta_1\beta_2}{1 + \beta_1^2 + \beta_2^2} \\
\mathbb{c}orr(Y_t, Y_{t-2}) &= \rho_2 \\
&= \frac{\gamma_2}{\gamma_0} \\
&= \frac{\beta_2\sigma_\varepsilon^2}{(1 + \beta_1^2 + \beta_2^2)\sigma_\varepsilon^2} \\
&= \frac{\beta_2}{1 + \beta_1^2 + \beta_2^2}
\end{aligned}$$

$$corr(Y_t, Y_{t-2}) = 0$$

$$\vdots$$

5 SAS Code

Code 8.5.1 *Exercise for understanding AR and MA process*

```
data arma;
        seed1=22;
        seed2=24;
        seed3=26;
        seed4=28;
        alpha1=0.9;
        alpha2=0.0;
        ylag1=rannor(seed1);
        ylag2=rannor(seed2);
        do t=1 to 500;
                e=rannor(3);
                y=alpha1*ylag1+alpha2*ylag2+e;
                ylag2=ylag1;
                ylag1=y;
                output;
        end;
run;

proc gplot data=arma;
        symbol i=join;
        plot y*t;
run;

proc arima data=arma;
        identify var=y;
run;

quit;
```

Figure 8.5.1 *Randomly generated Y; AR(2) process with* $\alpha_1 = 0.9$ *and* $\alpha_2 = 0.0$

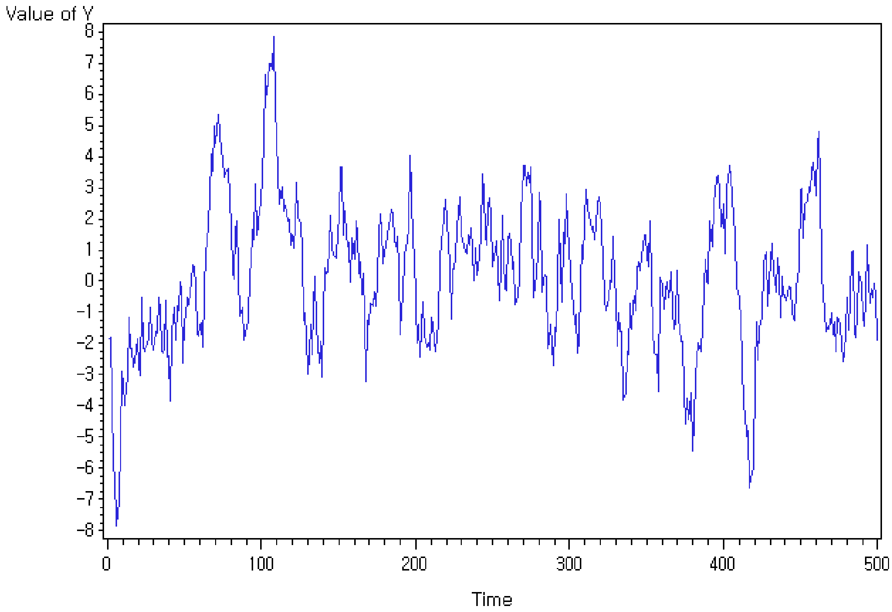


Figure 8.5.2 *Autocorrelation function resulted from Code 8.5.1*

Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	5.253024	1.00000	|******************** |	0
1	4.703434	0.89538	. |****************** |	0.044721
2	4.170775	0.79398	. |**************** |	0.072158
3	3.728724	0.70982	. |************** |	0.087911
4	3.281191	0.62463	. |************ |	0.098711
5	2.854150	0.54333	. |*********** |	0.106322
6	2.355905	0.44849	. |********* |	0.111738
7	1.959481	0.37302	. |******* |	0.115282
8	1.656563	0.31535	. |****** |	0.117671
9	1.358782	0.25867	. |***** |	0.119349
10	1.047449	0.19940	. |****. |	0.120465
11	0.730160	0.13900	. |*** . |	0.121123
12	0.474655	0.09036	. |** . |	0.121442
13	0.162468	0.03093	. |* . |	0.121576
14	-0.064228	-.01223	. | . |	0.121592
15	-0.252485	-.04806	. *| . |	0.121595
16	-0.377678	-.07190	. *| . |	0.121633
17	-0.331732	-.06315	. *| . |	0.121718
18	-0.279174	-.05315	. *| . |	0.121783
19	-0.227702	-.04335	. *| . |	0.121829
20	-0.208490	-.03969	. *| . |	0.121860
21	-0.196938	-.03749	. *| . |	0.121886
22	-0.159424	-.03035	. *| . |	0.121909
23	-0.093444	-.01779	. | . |	0.121924
24	-0.041578	-.00791	. | . |	0.121930

"." marks two standard errors

Figure 8.5.3 *Partial autocorrelation function resulted from Code 8.5.1*

```
                         Partial Autocorrelations

 Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1        0.89538     |                    . |******************  |
   2       -0.03895     |                    .*| .                  |
   3        0.03081     |                    . |*.                  |
   4       -0.05280     |                    .*| .                  |
   5       -0.02638     |                    .*| .                  |
   6       -0.12194     |                    **| .                  |
   7        0.03696     |                    . |*.                  |
   8        0.02572     |                    . |*.                  |
   9       -0.02478     |                    . | .                  |
  10       -0.05348     |                    .*| .                  |
  11       -0.04905     |                    .*| .                  |
  12       -0.00480     |                    . | .                  |
  13       -0.10798     |                    **| .                  |
  14        0.04438     |                    . |*.                  |
  15       -0.01272     |                    . | .                  |
  16        0.03352     |                    . |*.                  |
  17        0.12143     |                    . |**                  |
  18        0.01655     |                    . | .                  |
  19       -0.00766     |                    . | .                  |
  20       -0.04664     |                    .*| .                  |
  21       -0.01340     |                    . | .                  |
  22       -0.00215     |                    . | .                  |
  23        0.05092     |                    . |*.                  |
  24        0.00188     |                    . | .                  |
```

Figure 8.5.4 *Check for white noise resulted from Code 8.5.1*

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1424.78	6	<.0001	0.895	0.794	0.710	0.625	0.543	0.448
12	1615.05	12	<.0001	0.373	0.315	0.259	0.199	0.139	0.090
18	1623.04	18	<.0001	0.031	-0.012	-0.048	-0.072	-0.063	-0.053
24	1626.26	24	<.0001	-0.043	-0.040	-0.037	-0.030	-0.018	-0.008

ASSIGNMENT 8

Exercise 8.1 (Stationarity Test) *Try Dickey-Fuller stationarity test for below variables, which have been used as the samples of exercises of throughout this semester; verify whether those variables contain unit-root or not with full sample uploaded on the Internet homepage of this course.*

$$\Delta Y_t = \beta_0 + \beta_1 Y_t + \beta_2 t + u_t$$

(1) Industrial Production

(2) Federal Funds Rate

(3) S&P 500 index

(4) KOSPI 200 index

(5) CD rate (Korean, 91 days; often used as proxy of risk-free rate)

(6) Producer Price index (Korean)

Exercise 8.2 (ARMA Modeling) *Build ARMA forecasting model of CD rate time-series used at (5) above; follow Box-Jenkins step to estimate the ARMA structure. The basic ARMA form of the results given by SAS is often expressed as below. And report lead 10 forecasted observations as a chart.*

$$\begin{aligned} A(L)Y_t &= \mu + B(L)\varepsilon_t \\ (1 + \alpha_1 L + \alpha_2 L^2 + \cdots)Y_t &= \mu + (1 + \beta_1 L + \beta_2 L^2 + \cdots)\varepsilon_t \\ Y_t + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots &= \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots \\ Y_t &= \mu - \alpha_1 Y_{t-1} - \cdots + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots \end{aligned}$$

Exercise 8.3 (GARCH-type Modeling)

(1) Build ARMA forecasting model of KOSPI 200 time-series used at (4).

(2) Check the Autocorrelation of the residuals obtained from proper model.

(3) Check the Autocorrelation of the squared residuals.

(4) Try the ARCH model below and obtain the μ, α_0, and α_1 below.

$$
\begin{aligned}
Y_t &= \mu + \varepsilon_t \\
\varepsilon_t &\sim \mathcal{N}(0, h_t) \\
h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2
\end{aligned}
$$

Lecture Note 9

1 Unit Root

MA(1) process and MA(2) process are

$$\begin{aligned} Y_t &= \varepsilon_t + \beta_1 \varepsilon_{t-1} \\ Y_t &= \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} \end{aligned}$$

AR(1) process is

$$\begin{aligned} Y_t &= \alpha_1 Y_{t-1} + \varepsilon_t \\ Y_{t-1} &= \alpha_1 Y_{t-2} + \varepsilon_{t-1} \\ Y_{t-2} &= \alpha_1 Y_{t-3} + \varepsilon_{t-2} \\ &\vdots \end{aligned}$$

This AR(1) process can be transformed into MA(∞) as

$$Y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_1^2 \varepsilon_{t-2} + \cdots$$

In the case of MA(1)

$$\begin{aligned} Y_t &= \varepsilon_t + \beta_1 \varepsilon_{t-1} \\ \mathbb{v}ar(Y_t) &= \gamma_0 \\ &= (1 + \beta_1^2)\sigma_\varepsilon^2 \\ \mathbb{c}ov(Y_t, Y_{t-1}) &= \gamma_1 \\ &= \beta_1 \sigma_\varepsilon^2 \\ \mathbb{c}orr(Y_t, Y_{t-1}) &= \rho_1 \end{aligned}$$

$$= \frac{\beta_1}{1+\beta_1^2}$$

Ex. $\beta_1 = 0.5$ then $\rho_1 = \frac{0.5}{1+0.25} = 0.4$, $\beta_1 = 2.0$ and $\rho_1 = \frac{2}{1+4} = 0.4$; two different parameters have same autocorrelation. Which means that parameter cannot be identified from autocorrelation that causes a problem.

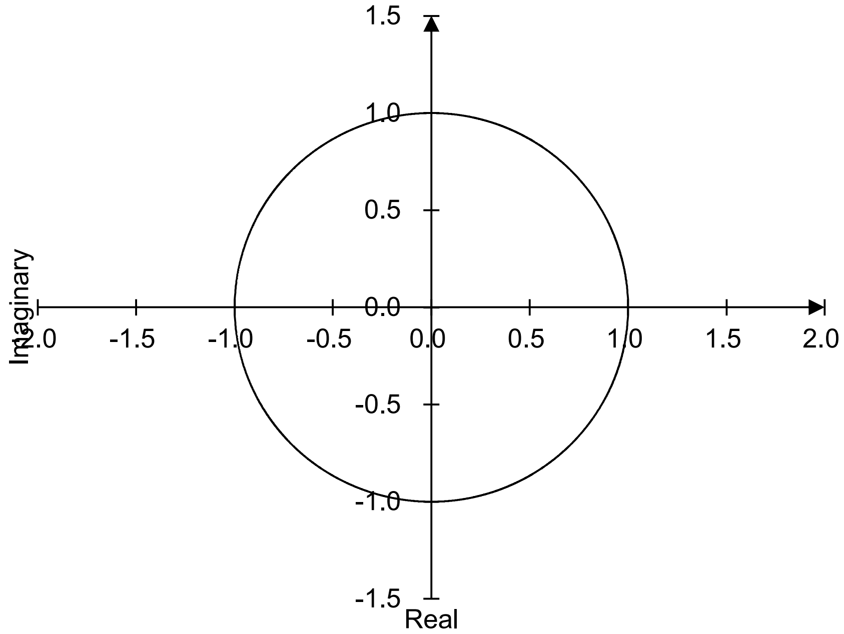
To avoid this kind of difficulty of identifying the parameter values of MA models, we impose the invertibility condition. The invertibility condition (only MA model) is

$$\begin{aligned} Y_t &= \varepsilon_t + \beta_1 \varepsilon_{t-1} \\ &= (1+\beta_1 L)\varepsilon_t \end{aligned}$$

The notation L is the lag operator such that $L\varepsilon_t = \varepsilon_{t-1}$ and $L^2\varepsilon_t = \varepsilon_{t-2}$ etc.

Solve $(1+\beta z) = 0$ in terms of z and hence $z = -\beta^{-1}$, the invertibility condition; the root(s) of the polynomial equation in terms of z should lie outside unit circle.

Figure 9.1.1 *Unit circle and existence of unit root*



The invertibility of condition of MA(2)

$$Y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}$$

$$= \quad (1+\beta_1 L+\beta_2 L^2)\varepsilon_t$$

The roots of the polynomial equation $1+\beta_1 z+\beta_2 z^2=0$ should lie outside the unit circle. Remind that, for the polynomial equation $ax^2+bx+c=0$, $x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$

2 ARMA(p,q) Process

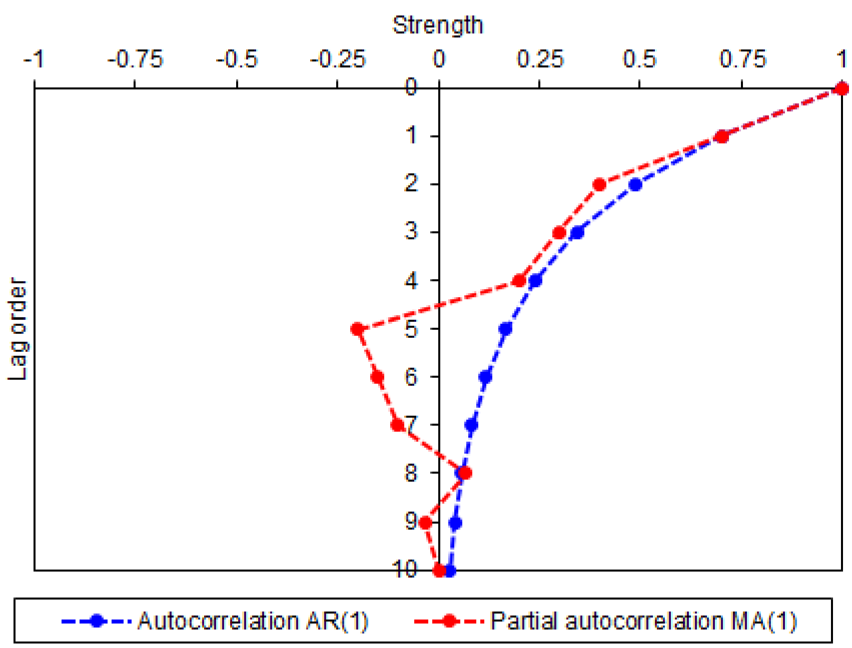
ARMA(1,1) process is

$$Y_t \quad = \quad \alpha Y_{t-1}+\varepsilon_t+\beta\varepsilon_{t-1}$$

Since this is a mixture of AR and MA process, it is impossible to determine neither the order of AR by looking at the partial autocorrelation function nor the order of MA by looking at the autocorrelation function.

The presence of AR process makes the autocorrelations dying down as the lag increases and the presence of MA process makes the partial autocorrelation dying down as the lag increases.

Figure 9.2.1 *Autocorrelation and partial autocorrelation function of AR(1), MA(1)*



MA(1)

$$\begin{aligned} Y_t &= \varepsilon_t + \beta_1 \varepsilon_{t-1} \\ &= (1 + \beta_1 L)\varepsilon_t \\ \frac{Y_t}{1+\beta_1 L} &= \varepsilon_t \end{aligned}$$

AR(∞)

$$\begin{aligned} \varepsilon_t &= Y_t - \beta_1 \varepsilon_{t-1} \\ \varepsilon_{t-1} &= Y_{t-1} - \beta_1 \varepsilon_{t-2} \\ \varepsilon_t &= Y_t - \beta_1 (Y_{t-1} - \beta \varepsilon_{t-2}) \\ &= Y_t - \beta_1 Y_{t-1} + \beta_1^2 (Y_{t-2} - \beta_1 \varepsilon_{t-3}) \\ &= Y_t - \beta_1 Y_{t-1} + \beta_1^2 Y_{t-2} - \beta_1^3 Y_{t-3} + \cdots \\ Y_t &= \beta_1 Y_{t-1} - \beta_1^2 Y_{t-2} + \beta_1^3 Y_{t-3} - \cdots + \varepsilon_t \end{aligned}$$

Thus, we can transform MA(1) into AR(∞) and also AR(1) into MA(∞)

3 Principle of Parsimony

Principle of parsimony, in words, the shorter is the better in terms of the order of Auto-Regressive and Moving Average.

Ex. Suppose ARMA(1,1) and ARMA(2,1) match the date equally well.

Definition (Akaike Information Criterion) *In Econometrics, AIC statistic to determine the fit of the estimated model can be defined as*

$$AIC = -\ell^* + 2k$$

Where

ℓ^* : Maximized log-likelihood value

k : The number of parameters estimated

n : The number of observation

Definition (Schwartz Bayesian Information Criterion) *SBC fit statistic for economic model can be defined as*

$$SBC = -\ell^* + k\log n$$

For AIC and SBC, $2k$ and $k\log n$ can be viewed as penalty for the number of parameters. When $n \geq 8$, $\log n \geq 2$; thus, SBC has more penalty than AIC.

4 SAS Code

Code 9.4.1 *Exercise for heterogeneous AR and MA process*

```
data arma;
	seed1=40;
	seed2=42;
	seed3=44;
	alpha=0.9;
	beta=0.8;
	ylag=rannor(seed1);
	elag=rannor(seed2);
	do t=1 to 500;
		e=rannor(seed3);
		y=alpha*ylag+e+beta*elag;
		ylag=y;
		elag=e;
		output;
	end;
run;

proc arima data=arma;
	identify var=y;
	estimate p=1 q=0 method=ml maxiter=100 noint;
	estimate p=1 q=1 method=ml maxiter=100 noint;
	estimate p=2 q=1 method=ml maxiter=100 noint;
run;

quit;
```

Figure 9.4.1 *Descriptive statistics for estimated ARMA models; AIC and SBC*

Model	Statistic	
	AIC	SBC
ARMA(1,0)	1572.12	1576.33
ARMA(1,1)	**1368.43**	**1376.86**
ARMA(2,1)	1368.67	1381.32

Choose the model with the smallest AIC and SBC.

$$
\begin{aligned}
(1-\alpha L)Y_t &= (1-\beta L)\varepsilon_t \\
Y_t &= \alpha Y_{t-1} + \varepsilon_t + \beta\varepsilon_{t-1} \\
&= \alpha Y_{t-1} + \varepsilon_t - \beta\varepsilon_{t-1} \leftarrow \text{When interpret the output of SAS result}
\end{aligned}
$$

Code 9.4.2 *Exercise for ARMA estimation; Industrial Production case*

```
data ip;
        infile "c:₩ip.prn";
        input month ip;
        ipg=dif(log(ip))*1200;
run;

proc arima data=ip;
        identify var=ipg;
        estimate p=1 q=0 method=ml maxiter=100 noint;
        estimate p=1 q=1 method=ml maxiter=100 noint;
        estimate p=2 q=1 method=ml maxiter=100 noint;
run;

quit;
```

All finite order MA processes are stationary regardless of their parameter values.

Cf. $|\alpha| \geq 1$ for AR model estimation?

Lecture Note 10

1 Durbin-Watson d Test

Suppose the model

$$Y_t = \alpha + \beta X_t + u_t$$

Assumption (Durbin-Watson d Test) *DW d test assumes*

(1) Dependent variables are fixed non-random. Cf. Classical Assumption (1)

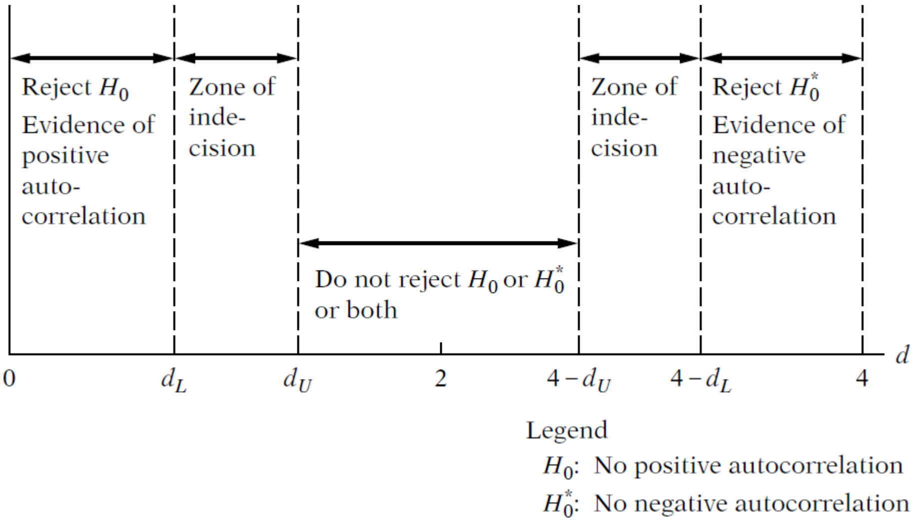
(2) Alternative hypothesis, H_1, is $u_t = \phi u_{t-1} + \varepsilon_t$ with non-zero ϕ, i.e. AR(1)

Definition (Durbin-Watson d Statistic) *d statistic for DW test*[5] *can be defined as*

$$\begin{aligned}
d &= \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \\
&= \frac{\sum e_t^2 + \sum e_{t-1}^2 - 2\sum e_t e_{t-1}}{\sum e_t^2} \\
&\approx 2\left(1 - \frac{\sum e_t e_{t-1}}{\sum e_t^2}\right) \\
&\approx 0 \leftarrow \text{in the case of severe positive autocorrelation} \\
&\approx 2 \leftarrow \text{without autocorrelation} \\
&\approx 4 \leftarrow \text{severe negative autocorrelation}
\end{aligned}$$

[5] James Durbin, and Geoffrey S. Watson, 1951, "Testing for Serial Correlation in Least-Squares Regression", *Biometrika* 38

Figure 8.2.1 *Durbin-Watson d statistic*[6]



If $0 \leq d \leq dL$ then reject the null $\phi = 0$ in favor of $\phi > 0$

If $dL < d < dU$ then the test is inconclusive.

If $dU \leq d \leq 2$ then do not reject the null.

If $2 \leq d \leq 4 - dU$ then do not reject the null.

If $4 - dU < d < 4 - dL$ then the test is inconclusive.

If $4 - dL \leq d \leq 4$ then reject the null in favor of $\phi < 0$

2 Durbin's *h* Test

When one or more of independent variables are random; dynamic model, Durbin-Watson d test will be invalid; instead of the test, Durbin's h test is used.

Durbin's h test involves 3 steps.

(1) Run the OLS for the original regression model

$$Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + u_t$$

(2) Obtain the residuals e_t and run the following regression

$$e_t = \delta_0 + \delta_1 e_{t-1} + \delta_2 X_t + \delta_3 Y_{t-1} + \nu_t$$

(3) Test the hypothesis δ_1 using the usual t test.

[6] Damodar N. Gujarati, 2004, *Basic Econometrics*

Code 10.2.1 *Exercise for Durbin's h test application*

```
data fygm3;
          infile "c:₩fygm3.prn";
          input month fygm3;
          ylag=lag1(fygm3);
run;

data fyff;
          infile "c:₩fyff.prn";
          input month fyff;
run;

data all1;
          merge fygm3 fyff;
          by month;
run;

proc reg data=all1;
          model fygm3=fyff/dw;
          output out=out1 r=resid1;
run;

proc reg data=all1;
          model fygm3=fyff/dw;
          output out=out2 r=resid2;
run;

data out2;
          set out2;
          elag=lag1(resid2);
          obs=_n_;
run;

proc reg data=all1;
          model resid2=elag fyff ylag/dw;
run;

quit;
```

$$Y_t = \alpha + \beta X_t + u_t \tag{10.1}$$

Where

$$u_t = \phi u_{t-1} + \varepsilon_t$$

$$Y_{t-1} = \alpha + \beta X_{t-1} + u_{t-1}$$

$$\phi Y_{t-1} = \phi\alpha + \phi\beta X_{t-1} + \phi u_{t-1} \tag{10.2}$$

Subtract (10.2) from (10.1) then,

$$Y_t - \phi Y_{t-1} = (1-\phi)\alpha + \beta X_t - \phi\beta X_{t-1} + \underbrace{(u_t - \phi u_{t-1})}_{\varepsilon_t}$$

$$Y_t = \underbrace{(1-\phi)\alpha}_{\gamma_1} + \underbrace{\beta}_{\gamma_2} X_t - \underbrace{\phi\beta}_{\gamma_3} X_{t-1} + \underbrace{\phi}_{\gamma_4} Y_{t-1} + \varepsilon_t$$

$$\gamma_3 = -\gamma_2\gamma_4 \tag{10.3}$$

(10.3) means non-linear model; by which AR(1) is also non-linear model.

Code 10.2.2 *Estimate regression model w/ AR(1) process through MLE*

```
proc autoreg data=all1;
        model y=x/nlag=1 method=ml maxiter=100;
run;

proc autoreg data=all1;
        model y=x/nlag=5 method=ml maxiter=100;
run;

proc autoreg data=all1;
        model y=x/nlag=(1 3 5) method=ml maxiter=100;
run;

quit;
```

Note that

$$u_t = \phi u_{t-1} + \varepsilon_t$$

$$u_t - \phi u_{t-1} = \varepsilon_t$$

$$(1-\phi L)u_t = \varepsilon_t$$

Cf. (SAS technical issue) In this case, the result from SAS implies $-\phi$; thereby, the estimate should be interpreted in opposite way.

Lecture Note 11

1 Time-series Analysis

Time-series analysis is consist of 3 steps as

(1) Identification (of the model)

(2) Estimation (of the model parameters)

(3) Diagnostic check (of the chosen model)

Code 11.1.1 *Estimate ARMA(1,0) and ARMA(1,1) model for Industrial Production*

```
data ip;
        infile "c:₩ip.prn";
        input month iop;
        ipg=dif(log(ip))*1200;
run;

proc arima data=ip;
        identify var=ipg;
        estimate p=1 q=0 method=ml maxiter=100;
        estimate p=1 q=1 method=ml maxiter=100;
run;

quit;
```

Remind Figure 8.5.4; the check for white noise.

For “To Lag 6” below hypothesis test will be applied as

$$H_0 \quad : \quad \rho_1 = \rho_2 = \cdots = \rho_6 = 0$$

H_1 : At least one of them is non-zero.

Figure 11.1.1 *White noise check; recall of Figure 8.5.4*

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1424.78	6	<.0001	0.895	0.794	0.710	0.625	0.543	0.448
12	1615.05	12	<.0001	0.373	0.315	0.259	0.199	0.139	0.090
18	1623.04	18	<.0001	0.031	-0.012	-0.048	-0.072	-0.063	-0.053
24	1626.26	24	<.0001	-0.043	-0.040	-0.037	-0.030	-0.018	-0.008

All of the low of Figure 11.1.1 rejects the null hypothesis; thereby it can be recognized that autocorrelation among the lags still remains in the time-series. Simply, the time-series analyzed above is not white noise; the time-series is still including analyzable information.

Code 11.1.2 *Generate and Analyze the white noise X for exercise*

```
data a;
        do t=1 to 100;
                x=rannor(1);
                output;
        end;
run;

proc arima data=a;
        identify var=x;
run;

quit;
```

Figure 11.1.2 *White noise check for the generated variable X*

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.81	6	0.7026	-0.103	0.036	-0.133	0.018	-0.020	-0.077
12	4.97	12	0.9589	-0.013	-0.035	-0.017	0.023	0.052	-0.072
18	14.78	18	0.6773	-0.080	-0.066	0.163	0.003	0.069	-0.196
24	21.27	24	0.6226	0.021	-0.130	-0.026	0.011	0.152	-0.091

Any of above hypothesis cannot reject the null, which means there is no Statistical autocorrelation among the lags; thus, the variable X is white noise.

$$Y_t \quad = \quad \alpha Y_{t-1} + \varepsilon_t$$

Above statement will be well-constructed and desirable model if ε_t is white noise.

2 Maximum Likelihood Estimation

$$\begin{aligned} Y_t &= \alpha + \beta X_t + u_t \\ u_t &= \phi u_{t-1} + \varepsilon_t \end{aligned}$$

The autocorrelation in the error term makes the model non-linear. Proof is given by (10.3) above.

Consider the model

$$Y_t = \alpha + \beta X_t + u_t$$

Where

$$u_t \sim \mathcal{N}(0, \sigma^2) \tag{11.1}$$

Assumption (Classical Assumption) *For the regression model* $Y_t = \alpha + \beta X_t + u_t$*,*

(1) The independent variable is fixed non-random.

(2) $\forall t,\ \mathbb{E}(u_t) = 0$

(3.a) $\forall t,\ \mathbb{E}(u_t^2) = \sigma^2$

(3.b) $\forall t \neq s,\ \mathbb{E}(u_t u_s) = 0$

Now we need another assumption (11.1); since u_t is distributed as normal with mean zero and variance σ^2, we can write

$$\mathbb{P}(u_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{u_t^2}{\sigma^2}}$$

Above is the probability density function (PDF) of normally distributed u_t.

Since u_t and u_s, $\forall t \neq s$, are independent, i.e. the joint density of u_t and u_s is

$$\mathbb{P}(u_t, u_s) = \mathbb{P}(u_t)\mathbb{P}(u_s)$$

In other words, the joint density of u_t, $\forall t = 1,2,\cdots,n$ become

$$\begin{aligned} \mathbb{P}(u_1, u_2, \cdots, u_n) &= \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{u_t^2}{\sigma^2}} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2}\frac{\sum u_t^2}{\sigma^2}} \end{aligned}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\frac{\sum u_t^2}{\sigma^2}\right)$$

The statement "Normally distributed" plus u_t and u_s are uncorrelated implies u_t and u_s are independent. (But, not exactly)

$$\mathbb{P}(u_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{u_1^2}{\sigma^2}\right)$$

$$\mathbb{P}(u_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{u_2^2}{\sigma^2}\right)$$

$$\begin{aligned}\mathbb{P}(u_1, u_2) &= \mathbb{P}(u_1)\mathbb{P}(u_2) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\frac{\sum u_t^2}{\sigma^2}\right)\end{aligned}$$

The likelihood function of Y_1, Y_2, $\cdots$, Y_n with given parameters α and β becomes

$$\begin{aligned}\mathcal{L}(Y_1, Y_2, \cdots, Y_n | \alpha, \beta) &= \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(Y_t - \alpha - \beta X_t)^2}{\sigma^2}} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2}\frac{\sum (Y_t - \alpha - \beta X_t)^2}{\sigma^2}} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\frac{\sum (Y_t - \alpha - \beta X_t)^2}{\sigma^2}\right)\end{aligned}$$

Logarithm of $\mathcal{L}$ is transformed as

$$\begin{aligned}\ell(Y_1, Y_2, \cdots, Y_n | \alpha, \beta) &= \log \mathcal{L} \\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(Y_t - \alpha - \beta X_t)^2\end{aligned}$$

Maximum likelihood estimators are obtained by maximizing the log likelihood function with respect to the unknown parameters α, β, and σ^2

$$\max_{\alpha,\beta,\sigma^2} \ell(Y_1, Y_2, \cdots, Y_n | \alpha, \beta, \sigma^2)$$

$$\Rightarrow \max_{\alpha,\beta,\sigma^2} -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(Y_t - \alpha - \beta X_t)^2 \tag{11.2}$$

(11.2) can be transformed as

$$\Rightarrow \max_{\alpha,\beta} -\sum_{t=1}^{n}(Y_t - \alpha - \beta X_t)^2$$

$$\Rightarrow \min_{\alpha,\beta} \sum_{t=1}^{n} (Y_t - \alpha - \beta X_t)^2$$

Cf. (criteria) Compare with (2.5).

Thus, Maximum Likelihood Estimators become

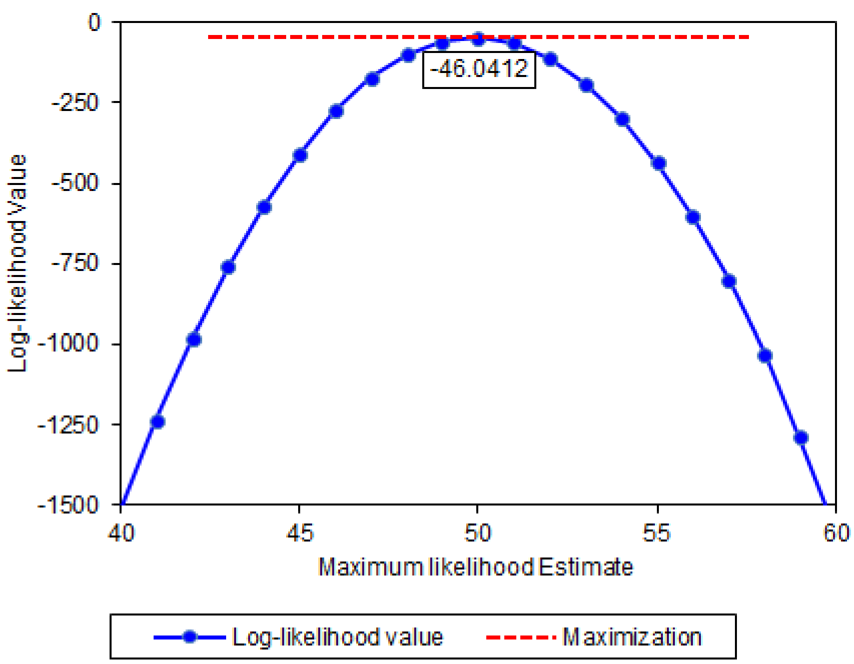
$$\begin{aligned} \hat{\beta} &= \frac{\sum x_i y_i}{\sum x_i^2} \\ &= b \leftarrow \text{OLS estimator} \\ \hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{X} \\ &= a \leftarrow \text{OLS estimator} \end{aligned}$$

Cf. (variance) Note that

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{n} \neq \frac{\sum e_i^2}{n-2} = s^2$$

Remind Figure 7.1.1; primary principle of Maximum Likelihood Estimation.

Figure 11.2.1 *Idea of Maximum Likelihood Estimation; recall of Figure 7.1.1*



Exercise *Estimate below regression models.*

$$INF_t = \alpha + \beta M2G_t + u_t$$

INF : Inflation based on wholesale price index

$M2G$: Growth rate of money supply measured in M2

Code 11.2.1 *Regression model estimation between INF and M2G*

```
data pw;
        infile "c:₩pw.prn";
        input month pw;
        inf=dif(log(pwg));
run;

data fm2;
        infile "c:₩fm2.prn";
        input month fm2;
        m2g=dif(log(fm2));
run;

data all1;
        merge pw fm2;
        by month;
run;

proc autoreg data=all1;
        model inf=m2g/nlag=1 method=ml maxiter=100;
run;

quit;
```

3 Stationarity Test

3.1 Stationary Condition

Stationary condition for AR(1) model is

$$\begin{aligned} Y_t &= \alpha Y_{t-1} + u_t \\ (1+\alpha L)Y_t &= u_t \end{aligned}$$

The root of the polynomial equation should lie outside the unit circle.

$$(1-dz) = 0$$

The root is $z = 1/\alpha$ and this should lie outside the circle; which implies $|\alpha| < 1$.

Stationary condition for AR(2) model is

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + u_t$$

$$(1 - \alpha_1 L - \alpha_2 L^2) Y_t = u_t$$

Polynomial equation for above equation is

$$1 - \alpha_1 z - \alpha_2 z^2 = 0$$

The stationary condition is "*The roots of the above polynomial equation should lie outside the unit circle.*"

3.2 Non-stationary Time-series Models

(1) Random-walk Model

$$Y_t = Y_{t-1} + \varepsilon_t$$

Where ε_t is white noise; that is

$$Y_t = Y_{t-1} + \varepsilon_t$$

$$Y_{t-1} = Y_{t-2} + \varepsilon_{t-1}$$

$$\vdots$$

$$Y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots$$

Statistical properties of non-stationary process; OLS estimators of

$$Y_t = \phi Y_{t-1} + u_t$$

Does not have asymptotic normal distribution, and hence the usual t test cannot be applied.

(2) The asymptotic distribution of OLS estimators of ϕ of the following models are different.

$$Y_t = \phi Y_{t-1} + u_t$$

$$Y_t = \mu + \phi Y_{t-1} + u_t$$

$$Y_t = \mu + \gamma t + \phi Y_{t-1} + u_t$$

When the true $\phi = 1$

(3) The probability of a non-stationary process comes back to its starting value is zero.

Figure 11.3.1 *Absence of mean-reverting in Random-walk time-series*

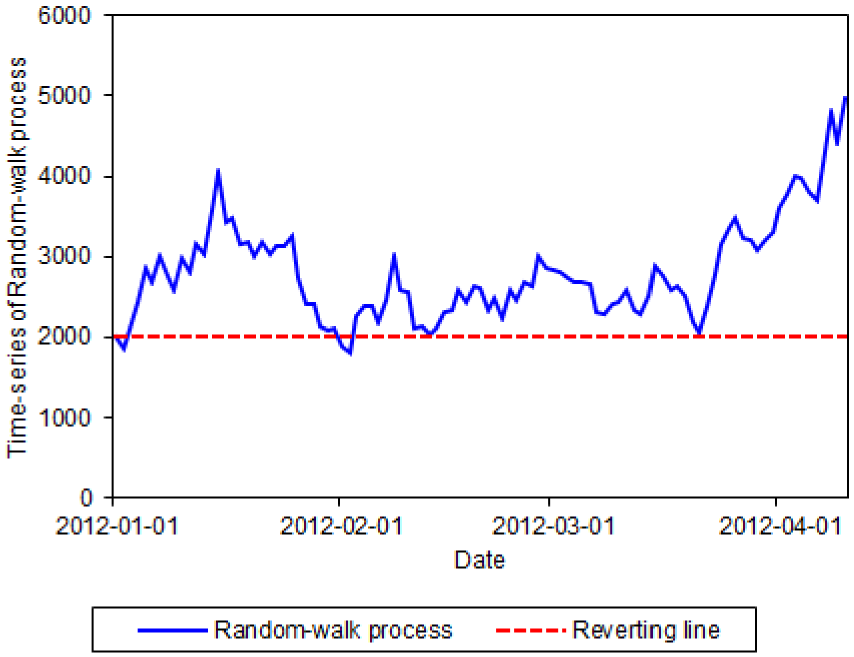
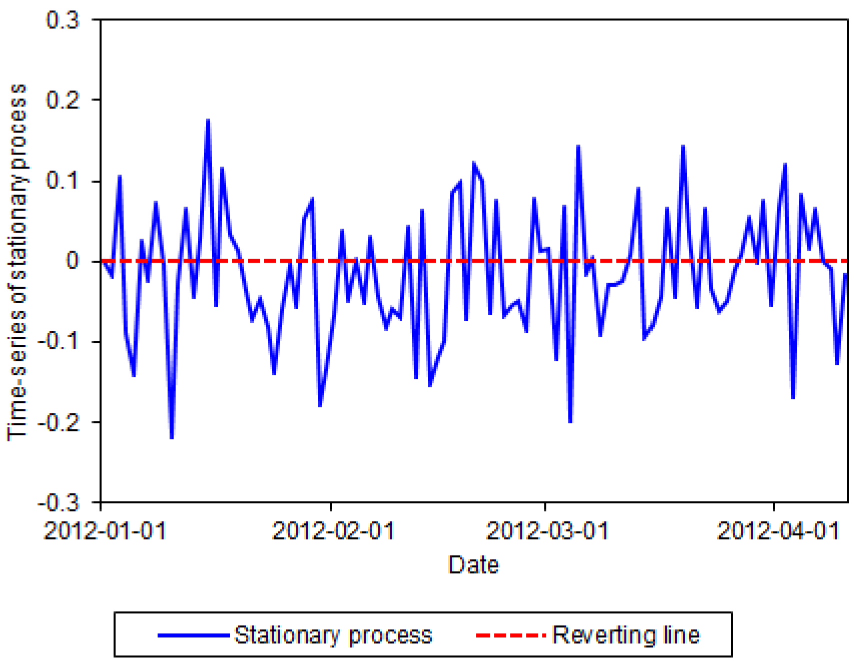


Figure 11.3.2 *Presence of mean-reverting in stationary time-series*



LECTURE NOTE 12

1 Conditional Heteroskedasticity

1.1 ARCH Model

Consider the following model

$$\begin{aligned} Y_t &= \alpha + \beta X_t + u_t \\ u_t &\sim^{i.i.d.} \mathcal{N}(0, \sigma^2) \\ h_t &= \mathbb{E}(u_t^2 | \Omega_{t-1}) \\ &= \gamma_0 + \gamma_1 u_{t-1}^2 \qquad (12.1) \end{aligned}$$

Ω_{t-1} : The information set that is available at time $(t-1)$

The above equation (12.1) implies the existence of the conditional heteroskedasticity; this equation is called as the first-order Auto-Regressive Conditional Heteroskedasticity[7], ARCH(1). The equation indicates that the conditional heteroskedasticity model is non-linear and hence OLS estimation cannot be used to estimate parameters.

What is the advantage of using conditional heteroskedasticity model?

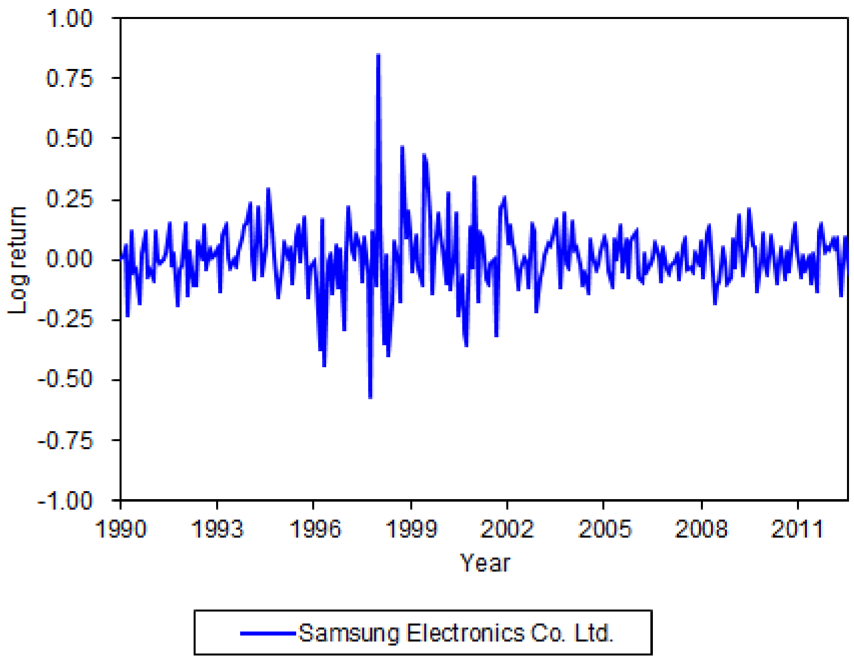
(1) The models provide a scientific measure of uncertainty which plays an important role in economic analysis.

(2) If the conditional heteroskedasticity is incorporated in estimating α and β, the standard errors will become smaller than those of BLUE.

Stock return time-series data usually shows strong evidence of conditional heteroskedasticity.

[7] Robert F. Engle, 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica* 50

Figure 12.1.1 *Is there any evidence of time-varying heteroskedasticity?*



Similarly, ARCH(2) can be denoted as

$$h_t = \mathbb{E}(u_t^2|\Omega_{t-1})$$

$$= \gamma_0 + \gamma_1 u_{t-1}^2 + \gamma_2 u_{t-2}^2$$

Necessary condition; all parameters of a conditional heteroskedasticity model should be non-negative.

$$\gamma_0, \gamma_1, \gamma_2 \geq 0$$

1.2 GARCH Model

GARCH model, which is firstly proposed by Bollerslev (1986)[8], can be depicted as below mathematical expressions.

$$h_t = \mathbb{E}(u_t^2|\Omega_{t-1})$$

[8] Tim Bollerslev, 1986, "Generalized Autoregressive Conditional Heteroskedasticity" , *Journal of Econometrics* 31

$$= \gamma_0 + \gamma_1 u_{t-1}^2 + \gamma_2 h_{t-1}$$

Where

$$\gamma_0, \gamma_1, \gamma_2 \geq 0$$

Code 12.1.1 *Exercise for estimation of ARCH model and GARCH model*

```
proc autoreg data=all1;
        model y=x/garch=(q=1) method=ml maxiter=100;
run;

proc autoreg data=all1;
        model y=x/garch=(q=1,p=1) method=ml maxiter=100;
run;

quit;
```

Similarly, GARCH-in-mean model can be depicted as

$$Y_t = \alpha + \beta X_t + \delta h_t + u_t$$

$$h_t = \mathbb{E}(u_t^2 | \Omega_{t-1})$$

$$= \gamma_0 + \gamma_1 u_{t-1}^2 + \gamma_2 h_{t-1}$$

Code 12.1.2 *Exercise for estimation of GARCH-in-mean model*

```
proc autoreg data=all1;
        model y=x/garch=(q=1,p=1,mean=linear) method=ml maxiter=100;
run;

quit;
```

LECTURE NOTE 13

1 Structure Change

Consider the model $\forall t = 1,2,\cdots,n$

$$Y_t = \alpha + \beta X_t + u_t$$

If the model experiences an abrupt change in parameter values at a certain point in time t^*, then we say that the model undergoes structural change.

Ex. Suppose the following statement

$$Y_t = \alpha_1 + \beta_1 X_t + u_t \leftarrow \text{for } t = 1,2,\cdots,(t^*-1)$$

$$Y_t = \alpha_2 + \beta_2 X_t + u_t \leftarrow \text{for } t = t^*,(t^*+1),\cdots,n$$

The test statistic τ can be defined as

$$\tau = \frac{RSS_R - RSS_U/q}{RSS_U/(n-k)}$$

Where

RSS_R	:	Residual sum of squares under the restriction
RSS_U	:	Residual sum of squares without the restriction
n	:	Total number of observations
k	:	Total number of parameters without the restriction
q	:	The number of restrictions on the parameters
H_0	:	Restriction $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$; no structural change
H_1	:	At least one of them is significantly different.

Code 13.1.1 *Exercise for structure change model*

```
data all1;
	set all1;
	if month<=19971201 then do;
		int1=1;
		int2=0;
		x1=x;
		x2=0;
	end;
	else do;
		int1=0;
		int2=1;
		x1=0;
		x2=x;
	end;
run;

proc reg data=all1;
	model y=x1 x2;
	test x1=x2;
run;

proc reg data=all1;
	model y=int1 int2 x1 x2/noint;
	test int1=int2,x1=x2;
run;

quit;
```

2 Autocorrelation

Consider the following model

$$\begin{aligned} Y_t &= \alpha + \beta X_t + u_t \\ u_t &= \phi u_{t-1} + \varepsilon_t \end{aligned}$$

Exercise

(1) How can we test the presence of auto correlation in u_i? Durbin-Watson d test

(2) If one or more explanatory variables are random, then what test should be used to test the existence of autocorrelations in the error term? Durbin's h test

3 Maximum Likelihood Estimation

Consider the model

$$Y_t = \alpha + \beta X_t + u_t$$

Where

$$u_t \sim \mathcal{N}(0, \sigma^2)$$

The likelihood function of Y_1, Y_2, ⋯, Y_n with given parameters α and β becomes

$$\mathcal{L}(Y_1, Y_2, \cdots, Y_n | \alpha, \beta) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(Y_t - \alpha - \beta X_t)^2}{\sigma^2}}$$

Logarithm of $\mathcal{L}$ is transformed as

$$\ell(Y_1, Y_2, \cdots, Y_n | \alpha, \beta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(Y_t - \alpha - \beta X_t)^2$$

Maximum likelihood estimators are obtained by maximizing the log likelihood function as

$$\max_{\alpha, \beta, \sigma^2} \ell(Y_1, Y_2, \cdots, Y_n | \alpha, \beta, \sigma^2)$$

4 Conditional Heteroskedasticity

Exercise

(1) Motivation? Modeling volatility clustering

(2) Utility? (a) Scientific measure of uncertainty of an economic time-series (b) Reflection of the conditional heteroskedasticity improves the fit of the original regression.

Remind the following mathematical expression

$$\begin{aligned} Y_t &= \alpha + \beta X_t + u_t \\ h_t &= \mathbb{E}(u_t^2 | \Omega_{t-1}) \\ &= \gamma_0 + \gamma_1 u_{t-1}^2 \end{aligned}$$

(3) SAS codes for GARCH-family model estimation?

5 Cointegration

Suppose $Y_t \sim I(1)$ and $X_t \sim I(1)$; where $I(1)$ means "first order integrated." This means that each differenced variable ΔX_t and ΔY_t are stationary. If a linear combination of Y_t and X_t become stationary, X_t and Y_t are cointegrated with each other.

Ex. Let $Y_t - X_t = z_t$. If z_t is stationary, then Y_t and X_t are cointegrated.

Ex. $r_{f,30y} - r_{f,10y} = \text{stationary}$

Ex. $PQ = MV$, where P: price, Q: quantity, M: money stock, and V: velocity; those variables can be modeled as $\ln P + \ln Q - \ln M - \ln V = \text{stationary}$, although every variables are non-stationary.

$$
\begin{aligned}
Y_t - X_t &= u_t \\
u_t &= u_{t-1} + \varepsilon_t \leftarrow I(1) \\
\Delta u_t &= \varepsilon_t \\
u_t &= 0.9u_{t-1} + \varepsilon_t \leftarrow \text{stationary} \\
\Delta u_t &= -0.1u_{t-1} + \varepsilon_t \\
r_{f,30y} &= c + \alpha r_{f,10y} + u_t
\end{aligned}
$$

Prof. Kiseok Lee (econklee@khu.ac.kr)
TA Junyong Kim (mindhunter@khu.ac.kr)

Advanced Econometrics

Last Update: November 8th, 2012

Last Update: November 8th, 2012

Last Update: November 8th, 2012

Last Update: November 8th, 2012

Last Update: November 8th, 2012

Lecture Note 1

Now Updating, Coming Soon

Now Updating, Coming Soon

Now Updating, Coming Soon

Now Updating, Coming Soon

Now Updating, Coming Soon

Lecture Note 2

1 Ordinary Least Squares Estimator

The Ordinary Least-Square (OLS) Estimator a and b is

$$\begin{aligned} a &= \hat{\alpha} \\ &= \bar{Y} - b\bar{X} \\ b &= \hat{\beta} \\ &= \frac{\sum x_t y_t}{\sum x_t^2} \\ &= \frac{\mathbb{c}ov(X_t, Y_t)}{\mathbb{v}ar(X_t)} \end{aligned}$$

Where

$$\begin{aligned} x_t &= X_t - \bar{X} \\ y_t &= Y_t - \bar{Y} \end{aligned}$$

Why should we use the OLS estimator?

2 Gauss-Markov Theorem

Population regression model is

$$Y_t = \alpha + \beta X_t + u_t$$

And corresponding sample regression model is

$$Y_t = a + bX_t + e_t$$

Here, what are the assumptions utilized to yield the equation?

(1) Y_t are random variables, but X_t are fixed numbers.

(2) Sum of errors is equal to zero.

$$\sum_{t=1}^{n} e_t = 0$$

(3) The denominator should not be equal to zero.

$$\begin{aligned} \frac{1}{n-1}\sum_{t=1}^{n} x_t^2 &= \mathbb{V}ar(X_t) \\ &\neq 0 \end{aligned}$$

If the denominator is zero, then the OLS estimator is undefined.

Assumption (Classical Assumption) *For the regression model* $Y_t = \alpha + \beta X_t + u_t$*,*

(1) The independent variable is fixed non-random.

(2) $\forall t,\ \mathbb{E}(u_t) = 0$

(3.a) $\forall t,\ \mathbb{E}(u_t^2) = \sigma^2$

(3.b) $\forall t \neq s,\ \mathbb{E}(u_t u_s) = 0$

Theorem (Gauss-Markov Theorem) *Under Classical Assumption, OLS estimator* b *is BLUE; Best Linear Unbiased Estimator.*

2.1 Best

$$b = \frac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} x_t^2}$$

Since

$$\begin{aligned} y_t &= Y_t - \bar{Y} \\ &= (\alpha + \beta X_t + u_t) - (\alpha + \beta \bar{X} + \bar{u}) \\ &= \beta X_t - \beta \bar{X} + u_t - \bar{u} \\ &= \beta (X_t - \bar{X}) + (u_t - \bar{u}) \end{aligned}$$

$$= \quad \beta x_t + (u_t - \bar{u})$$

Therefore

$$b = \frac{\sum_{t=1}^{n} x_t[\beta x_t + (u_t - \bar{u})]}{\sum_{t=1}^{n} x_t^2}$$

$$= \frac{\beta \sum_{t=1}^{n} x_t^2 + \sum_{t=1}^{n} x_t u_t - \bar{u} \sum_{t=1}^{n} x_t}{\sum_{t=1}^{n} x_t^2}$$

$$= \frac{\beta \sum_{t=1}^{n} x_t^2}{\sum_{t=1}^{n} x_t^2} + \frac{\sum_{t=1}^{n} x_t u_t}{\sum_{t=1}^{n} x_t^2} - \frac{\bar{u} \sum_{t=1}^{n} x_t}{\sum_{t=1}^{n} x_t^2}$$

$$= \beta + \frac{\sum_{t=1}^{n} x_t u_t}{\sum_{t=1}^{n} x_t^2} - \frac{\bar{u} \overbrace{\sum_{t=1}^{n} x_t}^{0}}{\sum_{t=1}^{n} x_t^2}$$

Because $\sum_{t=1}^{n} x_t = \sum_{t=1}^{n}(X_t - \bar{X}) = n\bar{X} - n\bar{X} = 0$, thus

$$b = \beta + \frac{\sum_{t=1}^{n} x_t u_t}{\sum_{t=1}^{n} x_t^2}$$

The estimator b is random variable. Because β and X_t are fixed constants, but u_t is random variable. So the dependent variable Y_t is always random variable.

$$\mathbb{V}ar(b) = \mathbb{E}[b - \mathbb{E}(b)]^2$$

$$= \frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}$$

Assume two estimators b_1 and b_2 that

Figure 2.2.1 *Properties of estimators in terms of variance*

	Estimator b_1	Estimator b_2
Sample 1	3.0	3.0
Sample 2	3.5	4.0
Sample 3	3.3	2.0

Then

$$\mathbb{V}ar(b_1) < \mathbb{V}ar(b_2)$$

Thus, the variance of OLS estimators is the smallest; in terms of the size of variance of estimators.

2.2 Linear

If an estimator is a function of the dependent variable Y_t, then the estimator is a linear estimator.

$$\begin{aligned} b &= \frac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} x_t^2} \\ &= \frac{1}{\sum_{t=1}^{n} x_t^2} \sum_{t=1}^{n} x_t y_t \\ &= \frac{1}{\sum_{t=1}^{n} x_t^2} \sum_{t=1}^{n} x_t (Y_t - \bar{Y}) \end{aligned}$$

Thus, b is linear in Y_t

2.3 Unbiased

For the parameter θ and the estimator $\hat{\theta}$, if $\mathbb{E}(\hat{\theta}) = \theta$, then $\hat{\theta}$ is the unbiased estimator of θ. Therefore

$$\begin{aligned} b &= \beta + \frac{\sum_{t=1}^{n} x_t u_t}{\sum_{t=1}^{n} x_t^2} \\ \mathbb{E}(b) &= \mathbb{E}\left(\beta + \frac{\sum_{t=1}^{n} x_t u_t}{\sum_{t=1}^{n} x_t^2}\right) \\ &= \mathbb{E}(\beta) + \mathbb{E}\left(\frac{\sum_{t=1}^{n} x_t u_t}{\sum_{t=1}^{n} x_t^2}\right) \\ &= \beta + \mathbb{E}\left(\frac{x_t \sum_{t=1}^{n} u_t}{\sum_{t=1}^{n} x_t^2}\right) \\ &= \beta + \frac{x_t}{\sum_{t=1}^{n} x_t^2} \underbrace{\mathbb{E}\left(\sum_{t=1}^{n} u_t\right)}_{0} \\ &= \beta \end{aligned}$$

The second term of above equation is eliminated by Classical Assumption (2).

$$b = \frac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} x_t^2}$$

To prove those properties

(Best) Classical Assumption (1), (2), and (3.a)

(Linear) No assumption

(Unbiased) Assumption (1) and (2)

Therefore, the OLS estimator b is the best among the estimators.

3 Variance Estimator

Taking one's expected value can be applied only when its distribution is given.

$$\mathbb{E}(u_t u_s) = \begin{cases} \sigma^2, & \forall t = s \\ 0, & \forall t \neq s \end{cases}$$

Above condition is given by the Classical Assumption (3.a)

$$\begin{aligned} \mathbb{E}(u_t u_s) &= \begin{cases} \sigma^2, & \forall t = s \\ 0, & \forall t \neq s \end{cases} \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{t=1}^{n} e_t^2 \\ s^2 &= \frac{1}{n-2}\sum_{t=1}^{n} e_t^2 \\ &\neq \hat{\sigma}^2 \end{aligned}$$

Those statistics are not OLS estimators because they are computed without minimize the objective function. The Classical Assumption is only used when certify that the estimators a and b are the Best Linear Unbiased Estimators. Therefore, we need an additional apparatus except the Classical Assumption.

$$u_t \sim \mathcal{N}(0, \sigma^2)$$

There are heterogeneous methods to estimate the unknown parameters and OLS method is a simple way to calculate those parameters. However, still other methods can be used to estimate them.

Note that $\sum_{t=1}^{n} x_t u_t$ is a linear combination of normal random variable and hence it is itself normally distributed because a linear combination of normally distributed random variable is also normal. Therefore, estimator b is normal with the mean

$$\mathbb{E}(b) = \beta$$

And the variance

$$\mathbb{V}ar(b) = \frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}$$

Or

$$b \sim \mathcal{N}\left(\beta, \frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}\right)$$

This is kind of the fourth assumption; the normality assumption.

Then, we have to standardize the normal distribution into the standard normal distribution.

3.1 Standardization

$$\begin{aligned} \mathbb{E}(b-\beta) &= \mathbb{V}ar\left(\frac{b}{\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}}}\right) \\ &= 1 \end{aligned}$$

Suppose $\mathbb{V}ar(u_t) = \sigma^2$ then

$$\begin{aligned} \mathbb{V}ar(u_t|\sigma) &= \mathbb{E}[u_t/\sigma - \mathbb{E}(u_t/\sigma)]^2 \\ &= \mathbb{E}(u_t/\sigma)^2 \\ &= \mathbb{E}(u_t)^2/\sigma^2 \\ &= \sigma^2/\sigma^2 \\ &= 1 \end{aligned}$$

Therefore, the distribution of below z is standard normal.

$$\begin{aligned} z &= \frac{b-\beta}{\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}}} \\ &\sim \mathcal{N}(0,1^2) \end{aligned}$$

From above

$$\begin{aligned}
\mathbb{P}(-1.96 < z < 1.96) &= 0.95 \\
\mathbb{P}\left(-1.96 < \frac{b-\beta}{\sqrt{\sigma^2/\sum_{t=1}^{n} x_t^2}} < 1.96\right) &= \\
\mathbb{P}\left(-1.96\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}} < b-\beta < 1.96\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}}\right) &= \\
\mathbb{P}\left(b-1.96\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}} < \beta < b+1.96\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}}\right) &=
\end{aligned}$$

Hence, the 95% confidence interval of β becomes

$$\left(b-1.96\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}}, b+1.96\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}}\right)$$

So, the hypothesis testing for a particular hypothesis that $\beta_0 = \beta$ can proceed by checking whether the value β_0 falls inside the 95% confidence interval. For example, suppose the 95% confidence interval is (-1.2, 2.2) and

$$H_0 \quad : \quad \beta_0 = 0.0$$

Then, the above hypothesis cannot be rejected. But

$$H_0 \quad : \quad \beta_0 = 3.0$$

It can be rejected.

3.2 Weakness of Hypothesis Test

(1) If the distribution of β is not normal, 1.96 won't have any meaning.

(2) σ^2 is unknown; since (2), we can't use the normal distribution. Fortunately, if we replace σ^2 by s^2, then the random variable follows the t distribution with $(n-2)$ degrees of freedom.

$$\begin{aligned}
t &= \frac{b-\beta}{\sqrt{s^2/\sum_{t=1}^{n} x_t^2}} \\
&\sim t_{(n-2)}
\end{aligned}$$

When the degrees of freedom go to infinity, the t distribution becomes the normal distribution and for reasonably large $(n-2)$

$$
\begin{aligned}
\mathbb{P}(-1.98 < t < 1.98) &= 0.95 \\
\mathbb{P}\left(-1.98 < \frac{b-\beta}{\sqrt{s^2/\sum_{t=1}^{n} x_t^2}} < 1.98\right) &= \\
\mathbb{P}\left(-1.98\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}} < b-\beta < 1.98\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}}\right) &= \\
\mathbb{P}\left(-1.98\sqrt{\frac{s^2}{\sum_{t=1}^{n} x_t^2}} < b-\beta < 1.98\sqrt{\frac{s^2}{\sum_{t=1}^{n} x_t^2}}\right) &= \\
\mathbb{P}\left(b-1.98\sqrt{\frac{s^2}{\sum_{t=1}^{n} x_t^2}} < \beta < b+1.98\sqrt{\frac{s^2}{\sum_{t=1}^{n} x_t^2}}\right) &=
\end{aligned}
$$

Then, the 95% confidence level of β becomes

$$
\left(b-1.98\sqrt{\frac{s^2}{\sum_{t=1}^{n} x_t^2}}, b+1.98\sqrt{\frac{s^2}{\sum_{t=1}^{n} x_t^2}}\right)
$$

Therefore,

$$
\begin{aligned}
\mathbb{v}ar(b) &= \frac{\sigma^2}{\sum_{t=1}^{n} x_t^2} \\
\mathrm{S.E.}(b) &= \sqrt{\frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}}
\end{aligned}
$$

And

$$
\begin{aligned}
\widehat{\mathbb{v}ar}(b) &= \frac{s^2}{\sum_{t=1}^{n} x_t^2} \\
\widehat{\mathrm{S.E.}}(b) &= \sqrt{\frac{s^2}{\sum_{t=1}^{n} x_t^2}}
\end{aligned}
$$

Code 2.3.1 *Exercise for simulated regression model*

```
data ex1;
        seed1=30;
        seed2=32;
        seed3=34;
        alpha=2.0;
        beta=1.0;
        ulag1=rannor(seed1);
        rho=0.0;
        do t=1 to 200;
                x=0.1*t+1.0*rannor(seed2);
                e=rannor(seed3);
                u=rho*ulag1+e;
                y=alpha+beta*x+u;
                ulag1=u;
                output;
        end;
run;

proc gplot data=ex1;
        symbol i=join;
        plot y*t u*t/overlay;
run;

proc reg data=ex1;
        model y=x;
run;

quit;
```

Prof. Kiseok Lee (econklee@khu.ac.kr)
TA Junyong Kim (mindhunter@khu.ac.kr)

Lecture Note 3

In regression model, if there is more than 3 parameters to estimate, too complex to express through the scalar form. Here, the OLS estimator simply can be found with the matrix theory and which is why we study the matrix theory.

1 Multiple Regression Model

Consider the following model

$$Y_t \quad = \quad \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t$$

Above model is a k number of parameters regression model. Suppose there are n number of observations.

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k1} + u_1 \\ Y_2 &= \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_k X_{k2} + u_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \cdots + \beta_k X_{kn} + u_n \end{aligned}$$

Then transform all variables into vector form as

$$\begin{aligned} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} &= \begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_3 \end{pmatrix} \\ \vec{Y} &= \vec{X}\vec{B} + \vec{U} \\ \vec{Y} &: (n \times 1) \text{ vector of dependent variable} \\ \vec{X} &: (n \times k) \text{ vector of independent variable} \end{aligned}$$

$$\vec{B} \quad : \quad (k \times 1) \text{ vector of parameter}$$

$$\vec{U} \quad : \quad (n \times 1) \text{ vector of residual}$$

Thus, multiple-regression model can be written as

$$\underbrace{\vec{Y}}_{(n\times 1)} = \underbrace{\vec{X}}_{(n\times k)} \underbrace{\vec{B}}_{(k\times n)} + \underbrace{\vec{U}}_{(n\times 1)}$$

Then, the OLS estimator of the $(k \times 1)$ vector $\vec{B}$ is just like this.

$$\underbrace{\vec{B}}_{(k\times 1)} = \underbrace{\left(\underbrace{\vec{X}'}_{(k\times n)} \underbrace{\vec{X}}_{(n\times k)}\right)^{-1}}_{(k\times k)} \underbrace{\underbrace{\vec{X}'}_{(k\times n)} \underbrace{\vec{Y}}_{(n\times 1)}}_{(k\times 1)}$$

Note that

$$b = \frac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} x_t^2}$$

$$= \left(\sum_{t=1}^{n} x_t^2\right)^{-1} \sum_{t=1}^{n} x_t y_t$$

From here, arrow sign for vector notation will be omitted.

2 Matrix Algebra

Definition (Vector) *A vector is an ordered sequence of elements arranged in a row or column like*

$$\underbrace{A}_{(1\times 3)} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

$$\underbrace{B}_{(3\times 1)} = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix}$$

The vector A is a column vector and B is a row vector. Wherever a vector is defined, it is defined in terms of a column vector.

Definition (Transposition)

$$A' = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$$

$$
\begin{aligned}
B' &= \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \\
A + B &= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 & 5 & 6 \end{pmatrix} \leftarrow \text{Impossible} \\
A + B' &= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \\
&= \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix}
\end{aligned}
$$

Definition (Multiplication by Scalar)

$$
\begin{aligned}
3A &= 3\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\
&= \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix}
\end{aligned}
$$

Definition (Subtraction)

$$
\begin{aligned}
B' - A &= \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\
&= \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix} \\
&= 3\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}
\end{aligned}
$$

Definition (Linear Combination)

$$
\begin{aligned}
3B' - 2A &= \begin{pmatrix} 12 \\ 15 \\ 18 \end{pmatrix} - \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} \\
&= \begin{pmatrix} 10 \\ 11 \\ 12 \end{pmatrix}
\end{aligned}
$$

Definition (Multiplication) *Multiplication between* (1×3) *and* (3×1) *vector becomes* (1×1) *vector, i.e. scalar*

$$
\begin{aligned}
BA &= \begin{pmatrix} 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\
&= 4 + 10 + 18 \\
&= 32
\end{aligned}
$$

Multiplication between (3×1) and (1×3) vector becomes (3×3) vector

$$
\begin{aligned}
AB &= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 4 & 5 & 6 \end{pmatrix} \\
&= \begin{pmatrix} 4 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{pmatrix}
\end{aligned}
$$

For the vector A

$$
A = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & \cdots & a_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{kn} \end{pmatrix}
$$

Transposition is

$$
\begin{aligned}
A &= \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \\
A' &= \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}
\end{aligned}
$$

Definition (Symmetric Matrix) *Symmetric matrix has mirror image around. For any vector X, $(X'X)$ is always defined as symmetric matrix.*

$$
A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 4 & 1 \end{pmatrix}
$$

Definition (Matrix Multiplication) *The first and second vector terms of the left hand side is rectangular and the vector term of the right hand side is square. Multiplication between (3×2) and (2×3) vector becomes (3×3) vector.*

$$
\begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 3 \\ 3 & 6 & 3 \\ 2 & 4 & 2 \end{pmatrix}
$$

Definition (Transpose of a product)

$$
\begin{aligned}
C &= AB \\
C' &= (AB)' \\
&= B'A'
\end{aligned}
$$

Definition (Identity Matrix)

$$
I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}
$$

Definition (Diagonal Matrix) *If all elements in off-diagonal positions are 0, then the matrix is diagonal.*

$$
\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}
$$

Definition (Idempotent Matrix) *If $A^2 = A$, then A is an idempotent matrix.*

$$
\begin{aligned}
A &= I_{(n\times n)} - \frac{1}{n}\left(i_{(n\times 1)} i'_{(1\times n)}\right) \\
A^2 &= AA \\
&= \left[I - \frac{1}{n}(ii')\right]\left[I - \frac{1}{n}(ii')\right] \\
&= I - \frac{1}{n}(ii') - \frac{1}{n}(ii') + \frac{1}{n^2}(ii')(ii')
\end{aligned}
$$

Since $i'i = n$, then

$$
\begin{aligned}
\frac{1}{n^2}(ii')(ii') &= \frac{1}{n}(ii') \\
A^2 &= I - \frac{1}{n}(ii') - \frac{1}{n}(ii') + \frac{1}{n}(ii') \\
&= I - \frac{1}{n}(ii')
\end{aligned}
$$

Definition (Trace)

$$
\operatorname{tr}(I_n) = n
$$

Definition (Matrix Differentiation) *For* $Y = XB + U$ *and* $Y = Xb + e$

$$
\begin{aligned}
e &= Y - Xb \\
e'e &= (Y - Xb)'(Y - Xb) \\
&= [Y' - (Xb)'](Y - Xb) \\
&= (Y' - b'X')(Y - Xb) \\
&= Y'Y - Y'Xb - b'X'Y + b'X'Xb \\
&= Y'Y - 2b'X'Y + b'X'Xb \\
\frac{\partial e'e}{\partial b} &= 0 - 2X'Y + 2X'Xb
\end{aligned}
$$

Note that

$$
\begin{aligned}
\frac{\partial b'X'Xb}{\partial b} &= 2(b'X'X)' \\
&= 2X'Xb
\end{aligned}
$$

Thus

$$
\begin{aligned}
e'e &= \begin{pmatrix} e_1 & e_2 & \cdots & e_n \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \\
&= \sum_{t=1}^{n} e_t^2
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\min_{\tilde{\alpha},\tilde{\beta}} \sum_{t=1}^{n} e_t^2 &= \min_{\tilde{B}} e'e \\
\frac{\partial e'e}{\partial b} &= -2X'Y + 2X'Xb \\
&= 0 \\
-X'Y + X'Xb &= 0 \\
X'Xb &= X'Y \\
b &= (X'X)^{-1}X'Y
\end{aligned}
$$

Definition (Inverse Matrix) *For the matrix A, if there exists A^{-1} which gives*

$$A^{-1}A = I$$

Then A^{-1} is called the inverse matrix of the matrix A.

Property (Inverse Matrix) *Some properties of inverse matrices are*

(1) $(A')^{-1} = (A^{-1})'$

(2) The inverse of an upper (lower) triangular matrix is also upper (lower) triangular matrix.

(3) Inverse matrix does not exist for singular matrices.

Definition (Determinant)

$$\begin{aligned}
|A| &= \det(A) \\
&= \begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} \\
&= (-1)^{1+1}a_{11}\begin{vmatrix} a_{22} & a_{32} \\ a_{23} & a_{33} \end{vmatrix} + (-1)^{1+2}a_{21}\begin{vmatrix} a_{12} & a_{32} \\ a_{13} & a_{33} \end{vmatrix} + (-1)^{1+3}a_{31}\begin{vmatrix} a_{12} & a_{22} \\ a_{13} & a_{23} \end{vmatrix} \\
&= a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{21}(a_{12}a_{33} - a_{32}a_{13}) + a_{31}(a_{12}a_{23} - a_{22}a_{13})
\end{aligned}$$

Suppose

$$\begin{aligned}
|A| &= \begin{vmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 3 & 0 & 1 \end{vmatrix} \\
&= (-1)^2(1)\begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} + (-1)^3(1)\begin{vmatrix} -1 & 1 \\ 3 & 1 \end{vmatrix} + (-1)^4(2)\begin{vmatrix} -1 & 2 \\ 3 & 0 \end{vmatrix} \\
&= (2-0) - (-1-3) + 2(0-6) \\
&= -6
\end{aligned}$$

Property (Determinant) *Some properties of determinant are*

(1) If a row (or column) of A is linearly dependent $|A| = 0$, other wise $|A| \neq 0$; non-singular matrices have non-zero determinants and singular matrices have zero determinants.

(2) The determinant of a triangular matrix is equal to the product of main diagonal elements.

(3) Multiplying any row (or column) of a matrix by a constant multiplies the determinant by the same constant. Multiplying a matrix of order n by a constant multiplies the determinant by the constant raised to the nth power

Definition (Rank of Matrix) *The rank is defined as the maximum number of linearly independent columns or rows in the matrix. If a matrix has the rank equal to the number of columns, then the matrix has the full column rank.*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 1 \\ 8 & 10 & 2 \end{pmatrix}$$

Let a_1, a_2, and a_3 as each row of the vector A, then a_3 is equal to $2a_2$ and hence

$$\begin{aligned} \text{rank}(A) &= 3 - 1 \\ &= 2 \end{aligned}$$

Rank can be applied to rectangle matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 8 & 12 \end{pmatrix}$$

From the rows

$$\text{rank}(A) = 2 - 1 = 1$$

And from the columns

$$\text{rank}(A) = 3 - 2 = 1$$

Rank of vector is the smaller one among the above numbers.

Property (Rank of Matrix)

$$\begin{aligned} \text{rank}(A) &= \text{rank}(A') \\ &= \text{rank}(A'A) \\ \text{rank}(AB) &\leq \min[\text{rank}(A), \text{rank}(B)] \end{aligned}$$

Definition (Inverse) *If there exists a matrix A^{-1} in such a way that $A^{-1}A = I$ then A^{-1} is called the inverse of the matrix A.*

$$\begin{aligned} A^{-1} &= \frac{1}{|A|}\text{adjoint}(A) \\ \text{adjoint}(A) &= \text{cofactor}(A)' \\ A^{-1} &= \frac{1}{|A|}\text{adjoint}(A) \end{aligned}$$

$$= \frac{1}{|A|}\begin{pmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{pmatrix}$$

Where c_{ij} is a cofactor of the element a_{ij} of A.

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

$$A^{-1} = -\frac{1}{2}\begin{bmatrix} (-1)^{1+1}4 & (-1)^{1+2}3 \\ (-1)^{2+1}2 & (-1)^{2+2}1 \end{bmatrix}'$$

$$= -\frac{1}{2}\begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix}'$$

$$= -\frac{1}{2}\begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}$$

$$A^{-1}A = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

$$= \begin{pmatrix} -2+3 & -4+4 \\ 3/2-3/2 & 6/2-4/2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= I$$

For example, for the vector B

$$B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

$$B^{-1} = \frac{1}{1}\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

And for the vector C

$$C = \begin{pmatrix} 1 & 3 & 4 \\ 1 & 2 & 1 \\ 2 & 4 & 5 \end{pmatrix}$$

$$|C| = (+1)(10-4) + (-3)(5-2) + (+4)(4-4)$$

$$= -3$$

$$\text{cofactor}(C) = \begin{pmatrix} 10-4 & 5-2 & 4-4 \\ 15-16 & 5-8 & 4-6 \\ 3-8 & 1-4 & 2-3 \end{pmatrix}$$

$$= \begin{pmatrix} 6 & 3 & 0 \\ -1 & -3 & -2 \\ -5 & -3 & -1 \end{pmatrix}$$

$$\text{adjoint}(C) = \begin{pmatrix} 6 & -1 & -5 \\ 3 & -3 & -3 \\ 0 & -2 & -1 \end{pmatrix}$$

$$C^{-1} = -\frac{1}{3}\begin{pmatrix} 6 & -1 & -5 \\ 3 & -3 & -3 \\ 0 & -2 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} -2 & 1/3 & 5/3 \\ -1 & 1 & 1 \\ 0 & 2/3 & 1/3 \end{pmatrix}$$

Definition (Eigenvalues) *Suppose a polynomial equation*

$$|A - \lambda I| = 0$$

In λ, the k roots of the above equation are referred to as *eigenvalues* of the matrix A. The above equation is called the characteristic equation and eigenvalues are sometimes called the characteristic roots or latent roots.

Ex. Suppose $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

$$|A - \lambda I| = \left|\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right|$$

$$= \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix}$$

$$= (2-\lambda)^2 - 1$$

$$= \lambda^2 - 4\lambda + 3$$

$$= 0$$

$$\lambda = \begin{cases} 3 \\ 1 \end{cases}$$

Definition (Eigenvectors) *Suppose the equation*

$$(A - \lambda_1 I)C_1 = 0$$

Where λ_1 is an eigenvalue of A; the vector C_1 that satisfies the above equation is called *eigenvector* corresponding to the eigenvalue λ_1.

Ex. $(A - 3I)C_1 = 0$

$$(A - 3I)C_1 = \left[\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - 3\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\begin{pmatrix} C_{11} \\ C_{12} \end{pmatrix}$$

$$= \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} C_{11} \\ C_{12} \end{pmatrix}$$

$$= \begin{pmatrix} -C_{11} + C_{12} \\ C_{11} - C_{12} \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Thus, the eigenvector corresponding to the eigenvalue 3 becomes $\begin{pmatrix} r \\ r \end{pmatrix}$ for any real number r.

$$\begin{aligned} (A - 1I)C_2 &= \left[\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - 1\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\begin{pmatrix} C_{12} \\ C_{22} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} C_{12} \\ C_{22} \end{pmatrix} \\ &= \begin{pmatrix} C_{12} + C_{22} \\ C_{12} + C_{22} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned}$$

Thus, the eigenvector corresponding to the eigenvalue 1 becomes $C_2 = \begin{pmatrix} s \\ -s \end{pmatrix}$ for any real number s.

Often, the norms of C_1 and C_2 vectors are equated to 1; that is

$$C_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

$$C_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

Norm of the vector C_1

$$= \sqrt{C_1' C_1}$$

$$= ||C_1||$$

Ex. $B = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$

$$\begin{aligned} \left|\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right| &= \begin{vmatrix} 2-\lambda & 2 \\ 1 & 3-\lambda \end{vmatrix} \\ (2-\lambda)(3-\lambda) - 2 &= 0 \\ 6 - 2\lambda - 3\lambda + \lambda^2 - 2 &= \\ \lambda^2 - 5\lambda + 4 &= \\ \lambda &= \begin{cases} 1 \\ 4 \end{cases} \end{aligned}$$

Property (Eigenvalue, Eigenvector)

(1) Eigenvalues of a symmetric matrix is real.

(2) When all eigenvalues are distinct, the eigenvector matrix C will have k linearly independent column and so $C^{-1}AC = \Lambda$ or $A = C\Lambda C^{-1}$.

(3) When all eigenvalues of a symmetric matrix are distinctive, eigenvectors are pairwise orthogonal; if H is a orthogonal matrix, then $H^{-1} = H'$.

(4) When the eigenvalues are not all distinctive, there are usually fewer than k linearly independent eigenvectors.

(5) The sum of all eigenvalues equals to the trace $C^{-1}AC = \Lambda$;

$$\text{trace}(C^{-1}AC) = \text{trace}(\Lambda)$$

Note that

$$\begin{aligned}\text{trace}(C^{-1}AC) &= \text{trace}(ACC^{-1}) \\ &= \text{trace}(A) \\ \text{trace}(A) &= \text{trace}(\Lambda)\end{aligned}$$

(6) The product of eigenvalues is equal to the determinant of A.

$$\begin{aligned}|\Lambda| &= |C^{-1}AC| \\ &= |C^{-1}||A||C| \\ &= \frac{1}{|C|}|A||C| \\ &= |A|\end{aligned}$$

(7) The rank of a matrix is equal to the number of non-zero eigenvalues.

(8) The eigenvalues of $\Pi = I - A$ are the complements of the eigenvalues of A, but the eigenvectors of the two matrices are the same.

(9) The eigenvalues of A^2 are the squares of the eigenvalues of A, but the eigenvectors of bot matrices are the same.

(10) The eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A, but the eigenvectors of both matrices are the same.

(11) Each eigenvalues of an idempotent matrix is either zero or one.

Code 3.2.1 *Matrix algebra exercises 1*

```
proc iml;
        reset noprint;

        a={2 1,1 2};

        deta=det(a);
        tracea=trace(a);
        call eigen(evala,eveca,a);
        print ,"a matrix" a;
        print ,"determinant of a" deta;
        print ,"trace of a" tracea;
        print ,"eigenvalues of a" evala;
        print ,"eigenvectors of a" eveca;

        hh=eveca`*eveca;
        eig=eveca`*a*eveca;
        print ,"hh" hh;
        print ,"eig" eig;

        inva=inv(a);
        print ,"inverse of a" inva;
quit;
```

Code 3.2.2 *Matrix algebra exercises 2*

```
proc iml;
	reset noprint;

	x={1 1,1 2,1 1,1 3};

	xtx=x`*x;
	detxtx=det(xtx);
	tracextx=trace(xtx);
	invxtx=inv(xtx);
	call eigen(evalxtx,evecxtx,xtx);
	print ,"x matrix" x;
	print ,"xtx matrix" xtx;
	print ,"determinant of xtx" detxtx;
	print ,"trace of xtx" tracextx;
	print ,"inverse of xtx" invxtx;
	print ,"eigenvalues of xtx" evalxtx;
	print ,"eigenvectors of xtx" evecxtx;

	eye=i(4);

	a=eye-x*invxtx*x`;
	deta=det(a);
	tracea=trace(a);
	call eigen(evala,eveca,a);
	print ,"a matrix" a;
	print ,"determinant of a" deta;
	print ,"trace of a" tracea;
	print ,"eigenvalues of a" evala;
	print ,"eigenvectors of a" eveca;

	hhxtx=evecxtx`*evecxtx;
	hha=eveca`*eveca;
	eigxx=evecxtx`*xtx*evecxtx;
	eiga=eveca`*a*eveca;
	print ,"hhxx" hhxx;
	print ,"hha" hha;

	rownum=nrow(x);
	one=j(rownum,1,1);
	onet=one`;
	invrownum=1/rownum;
	b=i(rownum)-invrownum*one*onet;
	dmx=b*x;
	print ,"demeaned x" dmx;
quit;
```

Lecture Note 4

$$
\begin{aligned}
Y_t &= \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \\
\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} &= \begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_3 \end{pmatrix} \\
\vec{Y} &= \vec{X}\vec{B} + \vec{U} \\
\vec{Y} &: (n \times 1) \ \text{vector of dependent variable} \\
\vec{X} &: (n \times k) \ \text{vector of independent variable} \\
\vec{B} &: (k \times 1) \ \text{vector of parameter} \\
\vec{U} &: (n \times 1) \ \text{vector of residual}
\end{aligned}
$$

1 Least Squares Criterion

$$
\begin{aligned}
Y &= X\tilde{b} + \tilde{e} \\
e'e &= \sum_{t=1}^{n} e_t \\
\min_{\tilde{b}} e'e &= \min_{\tilde{a},\tilde{b}} \sum_{t=1}^{n} e_t \\
\frac{\partial e'e}{\partial \tilde{b}} &= \frac{\partial \left(Y - X\tilde{b}\right)'\left(Y - X\tilde{b}\right)}{\partial \tilde{b}} \\
&= \frac{\partial \left(Y'Y - Y'X\tilde{b} - \tilde{b}'X'Y + \tilde{b}'X'X\tilde{b}\right)}{\partial \tilde{b}}
\end{aligned}
$$

$$= \quad 0-(Y'X)'-X'Y+2X'Xb$$

$$= \quad -2X'Y+2X'Xb$$

To minimize $e'e$, the first order condition of it should be equal to zero.

$$-2X'Y+2X'Xb = 0$$

$$X'Xb = X'Y$$

If $(X'X)^{-1}$ exists, then

$$b = (X'X)^{-1}X'Y$$

This completes the proof. ■

What assumptions have used to obtain the OLS estimator b?

(1) All observation of independent variable should not be same.

$$\mathbb{v}ar(X_t) \neq 0$$

(2) If the column rank of X is smaller than the number of columns k, $(X'X)^{-1}$ is singular.

$$A = \begin{pmatrix} 120 & 240 \\ 240 & 480 \end{pmatrix}$$

$$\text{rank}(A) = 2-1=1$$

$$(A'A)^{-1} = \text{Singular}$$

(3) The number of observation n should be much more than the number of parameter k which should be estimated.

Therefore, in order for $(X'X)^{-1}$ to be non-singular, X must be a full-rank matrix. This means that there should not be linearity between independent variables.

2 Characteristic of Regression Model

Where $(n \times k)$ matrix X has a full column rank with

$$e = Y-Xb$$

$$X'e = X'(Y-Xb)$$

$$= X'[Y-X(X'X)^{-1}X'Y]$$

$$= \quad X'Y - X'X(X'X)^{-1}X'Y$$

$$= \quad X'Y - X'Y$$

$$= \quad 0$$

Therefore, the columns of X are orthogonal to the $(n \times 1)$ vector e.

Suppose a and b both are $(n \times 1)$ matrices. If $a'b = 0$ then a and b are orthogonal.

$$X'e \quad = \quad \sum_{t=1}^{n} X_t e_t$$

$$= \quad 0$$

Remember in scalar expression, the first

$$\frac{\partial}{\partial \tilde{\alpha}} \sum_{t=1}^{n} e_t^2 \quad = \quad \frac{\partial}{\partial \tilde{\alpha}} \sum_{t=1}^{n} \left(Y_t - \tilde{\alpha} - \tilde{\beta} X_t\right)^2$$

$$= \quad (2)(-1) \sum_{t=1}^{n} (Y_t - a - bX_t)$$

$$= \quad \sum_{t=1}^{n} e_t$$

$$= \quad 0$$

And the second

$$\frac{\partial}{\partial \tilde{\beta}} \sum_{t=1}^{n} e_t^2 \quad = \quad \frac{\partial}{\partial \tilde{\beta}} \sum_{t=1}^{n} \left(Y_t - \tilde{\alpha} - \tilde{\beta} X_t\right)^2$$

$$= \quad (2)(-X_t) \sum_{t=1}^{n} (Y_t - a - bX_t)$$

$$= \quad \sum_{t=1}^{n} X_t e_t$$

$$= \quad 0$$

3 Coefficient of Determination

$$Y \quad = \quad Xb + e$$

$$\begin{aligned} Y'Y &= (Xb+e)'(Xb+e) \\ &= b'X'Xb + b'\underbrace{X'e}_{0} + \underbrace{e'X}_{0}b + e'e \\ &= b'X'Xb + e'e \end{aligned}$$

In scalar expression, note that

$$\underbrace{\sum_{t=1}^{n} y_t^2}_{\text{Total SS}} = \underbrace{b^2 \sum_{t=1}^{n} x_t^2}_{\text{Explained SS}} + \underbrace{\sum_{t=1}^{n} e_t^2}_{\text{Residual SS}}$$

Since

$$\begin{aligned} \sum_{t=1}^{n} y_t^2 &= \sum_{t=1}^{n} (Y_t - \bar{Y})^2 \\ &= \sum_{t=1}^{n} (Y_t - \bar{Y})(Y_t - \bar{Y}) \\ &= \sum_{t=1}^{n} (Y_t^2 - 2\bar{Y}Y_t + \bar{Y}^2) \\ &= \sum_{t=1}^{n} Y_t^2 - 2\bar{Y}\sum_{t=1}^{n} Y_t + n\bar{Y}^2 \\ &= \sum_{t=1}^{n} Y_t^2 - 2n\bar{Y}^2 + n\bar{Y}^2 \\ &= \sum_{t=1}^{n} Y_t^2 - n\bar{Y}^2 \\ (Y'Y - n\bar{Y}^2) &= (b'X'Xb - n\bar{Y}^2) + e'e \end{aligned}$$

Let $A = \frac{1}{n}ii'$, an idempotent matrix where i

$$i = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Let X_1 and b_1 as

$$X_1 = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$$

$$b_1 = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$$

Then

$$\begin{aligned} AY &= AXb + Ae \\ &= A(i \quad X_1)\begin{pmatrix} a \\ b_1 \end{pmatrix} + Ae \\ &= (Ai \quad AX_1)\begin{pmatrix} a \\ b_1 \end{pmatrix} + Ae \\ Ai &= \left(I - \frac{1}{n}ii'\right)i \\ &= i - i \\ &= 0 \end{aligned}$$

Therefore

$$\begin{aligned} AY &= AX_1b_1 + Ae \\ \mathcal{Y} &= \mathcal{X}_1 b_1 + Ae \\ \mathcal{Y}'\mathcal{Y} &= (\mathcal{X}_1 b_1 + Ae)'(\mathcal{X}_1 b_1 + Ae) \\ &= b_1'\mathcal{X}_1'\mathcal{X}_1 b_1 + b_1'\mathcal{X}_1'Ae + e'A'\mathcal{X}_1 b_1 + e'A'Ae \\ &= b_1'\mathcal{X}_1'\mathcal{X}_1 b_1 + e'e \end{aligned}$$

In conclusion

$$\underbrace{\mathcal{Y}'\mathcal{Y}}_{\text{Total SS}} = \underbrace{b_1'\mathcal{X}_1'\mathcal{X}_1 b_1}_{\text{Explained SS}} + \underbrace{e'e}_{\text{Residual SS}}$$

Confer that

$$\underbrace{\sum_{t=1}^{n} y_t^2}_{\text{Total SS}} = \underbrace{b^2 \sum_{t=1}^{n} x_t^2}_{\text{Explained SS}} + \underbrace{\sum_{t=1}^{n} e_t^2}_{\text{Residual SS}}$$

For $\mathcal{Y}$

$$\begin{aligned} \mathcal{Y} &= \left(I_n - \frac{1}{n}ii'\right)Y \\ &= Y - \frac{1}{n}in\bar{Y} \\ &= Y - i\bar{Y} \end{aligned}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} - \begin{pmatrix} \bar{Y} \\ \bar{Y} \\ \vdots \\ \bar{Y} \end{pmatrix}$$

4 Classical Assumption

Assumption (Classical Assumption) *For the regression model* $Y = XB + U$

(1) The $(n \times k)$ matrix X is fixed constant.

(2) $\mathbb{E}(U) = 0$

(3) $\mathbb{E}(UU') = \sigma^2 I_n$

Above Assumption (3) can be represented as

$$\mathbb{E}(UU') = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

Remind that

$$\mathbb{E}(u_t u_s) = \begin{cases} \sigma^2, & \forall t = s \\ 0, & \forall t \neq s \end{cases}$$

5 Variance-Covariance Matrix

$$\begin{aligned} \mathbb{E}(UU') &= \mathbb{E}\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \\ &= \begin{pmatrix} u_1^2 & u_1 u_2 & \cdots & u_1 u_n \\ u_2 u_1 & u_2^2 & \cdots & u_2 u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n u_1 & u_n u_2 & \cdots & u_n^2 \end{pmatrix} \end{aligned}$$

For the OLS estimator b

$$\begin{aligned} b &= (X'X)^{-1} X'Y \\ &= (X'X)^{-1} X'(XB + U) \end{aligned}$$

$$= \quad B + (X'X)^{-1}X'U$$

Therefore, the OLS estimator b is unbiased.

$$\begin{aligned}
\mathbb{v}ar - \mathbb{c}ov(b) &= \mathbb{E}[b - \mathbb{E}(b)][b - \mathbb{E}(b)]' \\
&= \mathbb{E}[b - B][b - B]' \\
&= \mathbb{E}[(X'X)^{-1}X'U][(X'X)^{-1}X'U]' \\
&= \mathbb{E}[(X'X)^{-1}X'UU'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'\mathbb{E}(UU')X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{v}ar - \mathbb{c}ov(b) &= \sigma^2(X'X)^{-1} \\
&= \begin{pmatrix} \mathbb{v}ar(b_1) & \mathbb{c}ov(b_2, b_1) & \cdots & \mathbb{c}ov(b_k, b_1) \\ \mathbb{c}ov(b_1, b_2) & \mathbb{v}ar(b_2) & \cdots & \mathbb{c}ov(b_k, b_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{c}ov(b_1, b_k) & \mathbb{c}ov(b_k, b_2) & \cdots & \mathbb{v}ar(b_k) \end{pmatrix} \\
\widehat{\mathrm{S.E.}}(b) &= \sqrt{\mathrm{diag}[s^2(X'X)]^{-1}}
\end{aligned}$$

Code 4.5.1 *Regression model exercise*

```
proc iml;
	reset noprint;

	x={1 1 1,1 2 4,1 3 9,1 4 16,1 5 25};
	y={1,5,27,58,108};

	xtx=x`*x;
	xtxi=inv(xtx);
	b=xtxi*(x`*y);
	yhat=x*b;
	e=y-yhat;
	sse=e`*e;
	n=nrow(x);
	k=ncol(x);
	df=n-k;
	s2=sse/df;
	tss=ssq(y-y[+]/n);
	r2=(tss-sse)/tss;
	stdb=sqrt(vecdiag(xtxi)*s2);
	t=b/stdb;
	prob=1-probf(t##2,1,df);
	print "Parameter Estimates",b stdb t prob;
	print "Regression Results", sse df s2 r2;
	print "y yhat e", y yhat e;
quit;
```

Lecture Note 5

1 Variance Estimator

Theorem *Below s^2 is the unbiased estimator of σ^2.*

$$
\begin{aligned}
s^2 &= \frac{e'e}{n-k} \\
&\neq \frac{e'e}{n} = \hat{\sigma}^2
\end{aligned}
$$

n : The number of observations

k : The number of observations

Proof *Since*

$$
\begin{aligned}
e &= Y - Xb \\
&= Y - X(X'X)^{-1}X'Y \\
&= [I - X(X'X)^{-1}X'I]Y \\
&= [I - X(X'X)^{-1}X'I](XB + U) \\
&= XB + U - XB - X(X'X)^{-1}X'U \\
&= [I - X(X'X)^{-1}X']U
\end{aligned}
$$

Thus

$$e = [I - X(X'X)^{-1}X']U$$

e : The error of sample model which is observed

U : The residual of population model which cannot be observed

$$
\begin{aligned}
e'e &= U'[I - X(X'X)^{-1}X']'[I - X(X'X)^{-1}X']U \\
&= [U' - U'X(X'X)^{-1}X']'[U - X(X'X)^{-1}X'U] \\
&= U'U - U'X(X'X)^{-1}X'U - U'X(X'X)^{-1}X'U + U'X(X'X)^{-1}X'U \\
&= U'U - U'X(X'X)^{-1}X'U \\
&= U'[I - X(X'X)^{-1}X']U
\end{aligned}
$$

Note that, if the any vector A, which can be expressed as

$$A = I - X(X'X)^{-1}X'$$

Then

$$
\begin{aligned}
A^2 &= [I - X(X'X)^{-1}X'][I - X(X'X)^{-1}X'] \\
&= I - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X' \\
&= I - X(X'X)^{-1}X' \\
&= A
\end{aligned}
$$

Thus A is an idempotent matrix. And

$$e'e = U'[I - X(X'X)^{-1}X']U$$

Now

$$
\begin{aligned}
s^2 &= \frac{1}{n-k}\sum_{t=1}^{n} e_t^2 \\
\mathbb{E}(e'e) &= \mathbb{E}\{U'[I - X(X'X)^{-1}X']U\} \\
&= \mathbb{E}[U'U - U'X(X'X)^{-1}X'U] \\
&= \mathbb{E}(U'U) - \mathbb{E}[U'X(X'X)^{-1}X'U] \\
&= n\sigma^2 - \mathbb{E}[U'X(X'X)^{-1}X'U] \\
&= n\sigma^2 - \mathbb{E}\{\mathrm{tr}[(X'X)^{-1}X'UU'X]\}
\end{aligned}
$$

For $U'U$

$$
\begin{aligned}
\mathbb{E}(U'U) &= \mathbb{E}\left(\sum_{t=1}^{n} u_t^2\right) \\
&= \sum_{t=1}^{n} \mathbb{E}(u_t^2)
\end{aligned}
$$

$$= n\sigma^2$$

$$\mathbb{E}(UU') = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

$$\begin{aligned} \text{trace}[\mathbb{E}(UU')] &= n\sigma^2 \\ &= \mathbb{E}(U'U) \end{aligned}$$

Note that

$$\begin{aligned} \text{trace}(AB) &= \text{trace}(BA) \\ \mathbb{E}(e'e) &= n\sigma^2 - \mathbb{E}[U'X(X'X)^{-1}X'U] \\ &= n\sigma^2 - \mathbb{E}\{\text{tr}[(X'X)^{-1}X'UU'X]\} \\ &= n\sigma^2 - \text{tr}\{\mathbb{E}[(X'X)^{-1}X'UU'X]\} \end{aligned}$$

By Classical Assumption (3) that ($\mathbb{E}(UU') = \sigma^2 I_n$)

$$\begin{aligned} \mathbb{E}(e'e) &= n\sigma^2 - \text{tr}[(X'X)^{-1}X'\sigma^2 I_n X] \\ &= n\sigma^2 - \sigma^2\,\text{tr}[(X'X)^{-1}X'X] \\ &= n\sigma^2 - \sigma^2\, tr(I_k) \\ &= (n-k)\sigma^2 \end{aligned}$$

Therefore

$$\begin{aligned} \mathbb{E}(e'e) &= (n-k)\sigma^2 \\ \mathbb{E}\left(\frac{e'e}{n-k}\right) &= \sigma^2 \\ \mathbb{E}(s^2) &= \sigma^2 \end{aligned}$$

This completes the proof. ■

$$\mathbb{E}\left(\frac{e'e}{n}\right) \neq \sigma^2$$

2 Information Criteria

2.1 Akaike Information Criterion

Cf. (Technical issue) From Prof. Kiseok Lee's measure

$$AIC = \ln\left(\frac{e'e}{n}\right) + 2 \times \left(\frac{k}{n}\right)$$

From the measure of SAS

$$AIC = n \times \ln\left(\frac{e'e}{n}\right) + 2 \times k$$

Slightly different, but both work in a same manners.

2.2 Schwartz Bayesian Information Criterion

Cf. (Technical issue) From Prof. Kiseok Lee's measure

$$SBC = \ln\left(\frac{e'e}{n}\right) + \ln(n) \times \left(\frac{k}{n}\right)$$

From the measure of SAS

$$SBC = n \times \ln\left(\frac{e'e}{n}\right) + \ln(n) \times k$$

Note that

$$n \times \ln\left(\frac{e'e}{n}\right) = -\ell^*$$

ℓ^* : Maximized log-likelihood value

For both AIC and SBC, smaller one is better; difference between their penalties.

3 Gauss-Markov Theorem

Theorem (Gauss-Markov Theorem) *Under Classical Assumption, OLS estimator* b *is BLUE; Best Linear Unbiased Estimator.*

Proof *To the contrary, there exists an estimator* $b^* = b + C'Y$ *that attains the smaller variance-covariance than that of* b.

Note that, it is impossible to compare the vectors in a manner of comparing scalar. Which means that, for the inequality $3 < 4$ comparing is applicable but

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad < \quad \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$

This is not applicable. Note that. C is $(n \times k)$ matrix of the constants which does not contain any random variable.

$$\begin{aligned} \mathbb{E}(b^*) &= \mathbb{E}(b + C'Y) \\ &= B + \mathbb{E}[C'(XB + U)] \\ &= B + C'XB + \mathbb{E}(C'U) \end{aligned}$$

By Classical Assumption (2)

$$= B + C'XB$$

If $C'X = 0$, then

$$= B$$

In order for b^* to be unbiased, we need $C'X = 0$ as new assumption. Then

$$\begin{aligned} b^* &= b + C'Y \\ &= b + C'(XB + U) \\ &= b + C'U \end{aligned}$$

Confer that

$$\begin{aligned} \sum_{t=1}^{n} C_t &= 0 \\ \sum_{t=1}^{n} C_t X_t &= 0 \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{v}ar - \mathbb{c}ov(b^*) &= \mathbb{E}(b^* - B)(b^* - B)' \\ &= \mathbb{E}(b + C'U - B)(b + C'U - B)' \\ &= \mathbb{E}(b - B + C'U)(b - B + C'U)' \\ &= \mathbb{E}(b - B)(b - B)' + \mathbb{E}[(b - B)U'C] + \mathbb{E}[C'U(b - B)'] + \mathbb{E}(C'UU'C) \end{aligned}$$

Since $b = B + (X'X)^{-1}X'U$

$$= \quad \mathbb{v}ar - \mathbb{c}ov(b^*) + \mathbb{E}[(X'X)^{-1}X'UU'C] + \mathbb{E}[C'U(X'X)^{-1}X'U] + \sigma^2 C'C$$

$$= \quad \mathbb{v}ar - \mathbb{c}ov(b^*) + \sigma^2(X'X)^{-1}X'C + \sigma^2 C'X(X'X)^{-1} + \sigma^2 C'C$$

Since $X'C = C'X = 0$

$$= \quad \mathbb{v}ar - \mathbb{c}ov(b^*) + \sigma^2 C'C$$

Since $C'C$ is a positive definite matrix

$\mathbb{v}ar - \mathbb{c}ov(b^*) - \mathbb{v}ar - \mathbb{c}ov(b) \quad = \quad$ Positive definite matrix

Therefore

$$\mathbb{v}ar - \mathbb{c}ov(b^*) \quad > \quad \mathbb{v}ar - \mathbb{c}ov(b)$$

This completes the proof. ∎

Let A be an $(n \times n)$ matrix. For a non-zero $(n \times 1)$ vector, if $a'Aa > 0$, then A is called positive definite.

$a'Aa > 0$: Positive definite

$a'Aa < 0$: Negative definite

$a'Aa \geq 0$: Positive semi-definite

For example, for a $(n \times k)$ vector C

CC' : $(n \times n)$ Positive semi-definite

$C'C$: $(k \times k)$ Positive definite with full column rank

For a rectangle vector, if its dimension of self-multiplied square matrix goes to higher, it will be a positive semi-definite. On the other hand, if its dimension goes lower, it will be a positive definite.

For another example, if $\text{rank}(C) \leq k$

$C'C$: $(n \times n)$ Positive semi-definite

$C'C$: $(k \times k)$ Positive semi-definite without full column rank

4 Joint Hypothesis Test

Assumption *Basement by which joint hypothesis test is valid is*

$$U \sim \mathcal{N}(0, \sigma^2 I_n)$$

4.1 General Framework

Ex. Suppose you want to test the null hypothesis $\beta_1 = 0$

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}$$

$$RB = \begin{pmatrix} 0 & 1 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$= \beta_1$$

(Question) *How many parameters restricted?*

(Answer) *One parameter β_2; i.e. the degrees of freedom is equal to 1.*

$$r = 0$$

Continuously, determine another example that

Ex. Suppose that $\beta_1 = \beta_2 = 0$ then

$$R = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \end{pmatrix}$$

$$RB = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

(Question) *How many parameters restricted?*

(Answer) *Two parameters β_1 and β_2; i.e. the degrees of freedom is equal to 2.*

$$r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Continuously

Ex. $\beta_1 + \beta_2 = 1$

$$R = \begin{pmatrix} 0 & 1 & 1 & \cdots & 0 \end{pmatrix}$$

$$RB = \begin{pmatrix} 0 & 1 & 1 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$= \beta_1 + \beta_2$$

(Question) *How many parameters restricted?*

(Answer) *One parameter; i.e.* R *has only one row so the degrees of freedom is 1.*

$$r = 1$$

4.2 Standardization

For the OLS estimator b

$$b \sim \mathcal{N}[B, \sigma^2(X'X)^{-1}]$$

Since the assumption $U \sim N(0, \sigma^2 I_n)$

$$Rb \sim \mathcal{N}[RB, \sigma^2 R(X'X)^{-1}R']$$

$$R(b - B) \sim \mathcal{N}[0, \sigma^2 R(X'X)^{-1}R']$$

Since $(X'X)^{-1}$ is positive definite, $(n \times k)$ vector X has full column rank. $R(X'X)^{-1}R'$ is also positive definite and symmetric. And hence, $[R(X'X)^{-1}R']^{-1}$ exists and it also symmetric and positive definite.

Therefore, there exists a matrix

$$[R(X'X)^{-1}R']^{-1/2} \times [R(X'X)^{-1}R']^{-1/2} = [R(X'X)^{-1}R']^{-1}$$

Then

$$\frac{1}{\sigma}[R(X'X)^{-1}R']^{-1/2}R(b - B) \sim \mathcal{N}(0, I_q)$$

Where q is the row dimension of R.

Note that

$$u_t \sim \mathcal{N}(0, \sigma^2)$$

$$u_t/\sigma \sim \mathcal{N}(0, 1^2)$$

$$u_t/(\sigma^2)^{-1} \sim \mathcal{N}(0, 1^2)$$

$$z \sim \mathcal{N}(0, 1^2)$$

Since

$$z^2 \sim \chi^2(1)$$

$$z_1^2 + z_2^2 \sim \chi^2(2)$$

$$\sum_{i=1}^{n} z_i^2 \sim \chi^2(n)$$

Therefore

$$\frac{1}{\sigma^2}(b-B)'\,R'[R(X'X)^{-1}R']^{-1/2}[R(X'X)^{-1}R']^{-1/2}R(b-B) \sim \chi^2(q)$$

$$\frac{1}{\sigma^2}(b-B)'\,R'[R(X'X)^{-1}R']^{-1}R(b-B) \sim \chi^2(q)$$

Note that

$$\frac{e'e}{\sigma^2} \sim \chi^2(n-k)$$

4.3 F Test

For below ϕ_1 and ϕ_2 are independent χ^2 distributed random variable and hence, the ratio between ϕ_1 and ϕ_2 is

$$\begin{aligned} F &= \frac{\phi_1}{\phi_2} \\ &= \frac{\frac{1}{\sigma^2}(Rb-r)'[R(X'X)^{-1}R']^{-1}(Rb-r)/q}{\frac{e'e}{\sigma^2}/(n-k)} \\ &\sim F_{q,(n-k)} \end{aligned}$$

Therefore

$$\begin{aligned} F &= \frac{(e_*'e_* - e'e)/q}{e'e/(n-k)} \\ &\sim F_{q,(n-k)} \end{aligned}$$

e_* : Residual vector under the null hypothesis (no treatment)

e : Residual vector under the alternative hypothesis (restricted)

Ex. For both H_0 and H_1

$$H_0 \quad : \quad \beta_2 = 0$$

$$Y_t \quad = \quad \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + \varepsilon_t \leftarrow e_*' e_*$$

$$H_1 \quad : \quad \beta_2 \neq 0$$

$$Y_t \quad = \quad \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \varepsilon_t \leftarrow e' e$$

If above H_0 is true, then there is only slight different between $e_*' e_*$ and $e' e$ and hence the value of the numerator of F and also itself will go to zero simultaneously by which H_0 cannot be rejected. On the other hand, if H_0 is not true, then $e_*' e_*$ will be more bigger then $e' e$, which can make the value of the numerator of F and itself go up. Thus, the null hypothesis will be rejected easily with relatively high F value.

LECTURE NOTE 6

1 Chow Forecast Test

If the parameter vector is constant, out of sample prediction will have specified probabilities lying within hands calculated from the sample data range prediction errors therefore cost doubt on the constancy hypothesis, and the converse for small prediction errors. The test of predictive accuracy widely referred to as the Chow[9] Test.

Ex. For below model

$$\begin{aligned} Y_t &= \mu + u_t \\ &= \bar{Y} + e_t \end{aligned}$$

Below right figure show that there is no change in parameter μ, on the other hand left figure shows the parameter break and slight changing its confidence interval.

[9] Gregory C. Chow, 1960, "Tests of Equality between Sets of Coefficients in Two Linear Regressions", *Econometrics* 28 (3)

Figure 6.1.1 *If the parameter is constant*

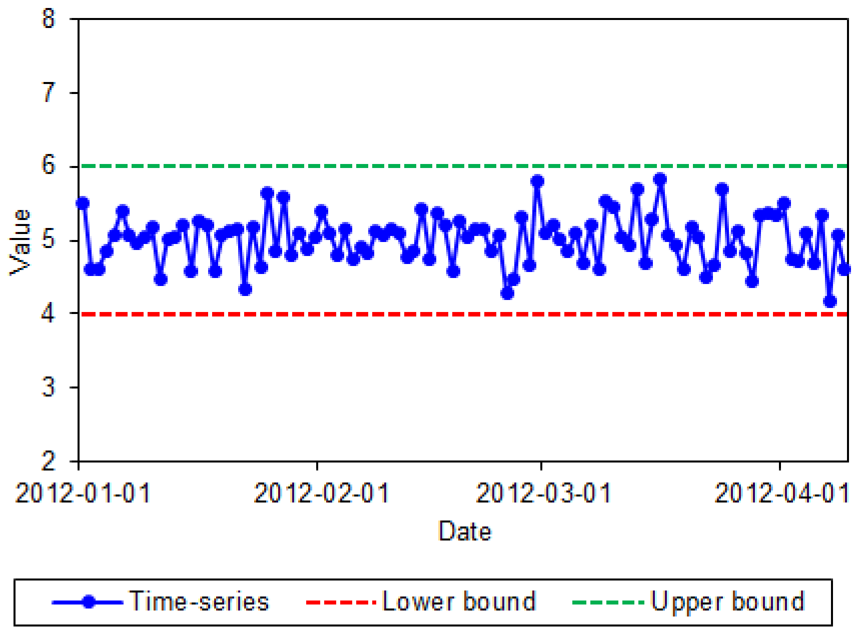
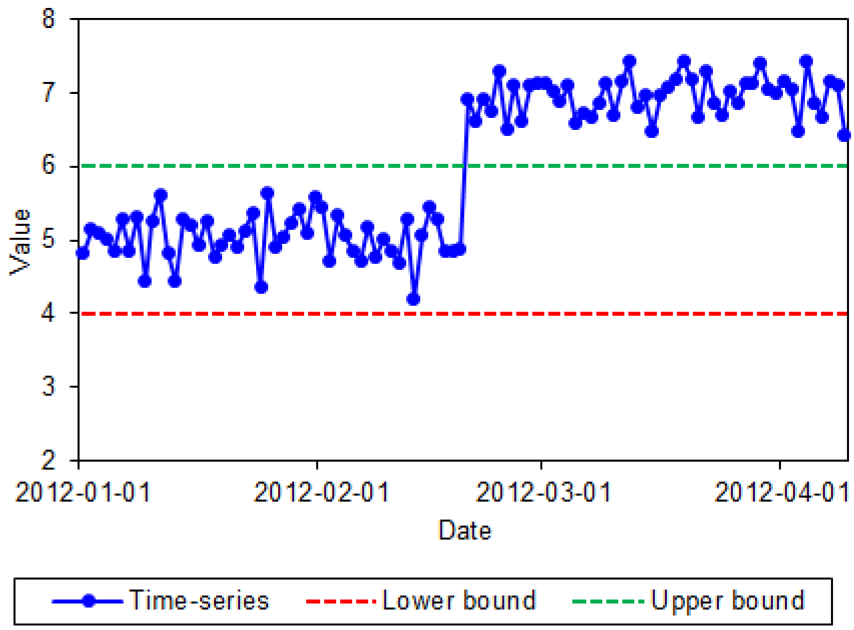


Figure 6.1.2 *If the parameter is broken*



2 Steps of Chow Forecast Test

(1) Estimate the OLS vector from the n_1 observations, obtain

$$b_1 = (X_1'X_1)^{-1}X_1'Y_1$$

Where X_t, Y_t for $t = 1,2$ indicate the partitioning of data into n_1, n_2 observations.

(2) Use b_1 to obtain a prediction of the Y_2 vector.

$$\hat{Y}_2 = X_2 b_1$$

(3) Obtain the vector of prediction error and analyze its sampling distribution under the null hypothesis of the parameter constancy the vector of prediction error is

$$\begin{aligned} d &= Y_2 - \hat{Y}_2 \\ &= Y_2 - X_2 b_1 \\ &= X_2 B + U_2 - X_2 b_1 \\ &= U_2 - X_2(b_1 - B) \end{aligned}$$

Since $\mathbb{E}(U_2) = 0$ and $\mathbb{E}(b_1) = B$

$$\mathbb{E}(d) = 0$$

And

$$\begin{aligned} \mathbb{v}ar(d) &= \mathbb{E}(dd') \\ &= \sigma^2[I_{n2}X_2(X_1'X_1)^{-1}X_2'] \end{aligned}$$

If $u_t \sim \mathcal{N}(0, \sigma^2)$ then

$$d'[\mathbb{v}ar(d)]^{-1}d \sim \chi^2(n_2)$$

And since that

$$\frac{e_1'e_1}{\sigma^2} \sim \chi^2(n_1 - k)$$

And since d and e_1 are independent, under the hypothesis of parameter constancy

$$\begin{aligned} F &= \frac{d'[I_n - X_2(X_1'X_1)^{-1}X_2']d/n_2}{e_1'e_1/(n_1 - k)} \\ &\sim F_{n_2,(n_1-k)} \end{aligned}$$

A simpler way of constructing a test statistic is

(1) Using n_1 observations, regress Y_1 on X_1 and obtain $e_1'e_1$.

(2) Fit the same regression to all $(n_1 + n_2)$ observations and obtain the restricted Residual Sum of Square $e_*'e_*$.

(3) Construct below

$$\begin{aligned} F &= \frac{(e_*'e_* - e_1'e_1)/n_2}{e_1'e_1/(n_1-k)} \\ &\sim F_{n_2,(n_1-k)} \end{aligned}$$

Note that

$$\begin{aligned} \frac{1}{\sigma}\left[I_{n_2} + X_2(X_1'X_1)^{-1}X_2'\right]^{-1/2} d &\sim \mathcal{N}\left(0, I_{n_2}\right) \\ b &\sim \mathcal{N}[B, \sigma^2(X'X)^{-1}] \\ Rb &\sim \mathcal{N}[RB, \sigma^2 R(X'X)^{-1}R'] \\ [\sigma^2 R(X_1'X_1)^{-1}R']^{-1/2}(Rb - r) &\sim \mathcal{N}\left(0, I_q\right) \end{aligned}$$

Since

$$\begin{aligned} u_t &\sim \mathcal{N}(0, \sigma^2) \\ \frac{u_t}{\sigma} &\sim \mathcal{N}(0, 1^2) \\ \left(\frac{u_t}{\sigma}\right)'\left(\frac{u_t}{\sigma}\right) &\sim \chi^2(1) \end{aligned}$$

Thus,

$$\begin{aligned} \left\{[\sigma^2 R(X_1'X_1)^{-1}R']^{-1/2}(Rb - r)\right\}'\left\{[\sigma^2 R(X_1'X_1)^{-1}R']^{-1/2}(Rb - r)\right\} &\sim \chi^2(q) \\ (Rb - r)'[\sigma^2 R(X_1'X_1)^{-1}R']^{-1/2}[\sigma^2 R(X_1'X_1)^{-1}R']^{-1/2}(Rb - r) &\sim \chi^2(q) \\ (Rb - r)'[\sigma^2 R(X_1'X_1)^{-1}R']^{-1}(Rb - r) &\sim \chi^2(q) \\ (Rb - r)'[var(Rb)]^{-1}(Rb - r) &\sim \chi^2(q) \end{aligned}$$

In here, where $(Rb - r)' = d$

Code 6.2.1 *Exercise for joint hypothesis test application*

```
proc iml;
        reset noprint;

        start reg;
                n=nrow(x);
                k=ncol(x);
                xtx=x`*x;
                xty=x`*y;
                xtxi=inv(xtx);
                b=xtxi*xty;
                yhat=x*b;
                resid=y-yhat;
                sse=resid`*resid;
                dfe=n-k;
                mse=sse/dfe;
                rmse=sqrt(mse);
                covb=xtxi#mse;
                stdb=sqrt(vecdiag(covb));
                t=b/stdb;
                probt=1-probf(t##2,1,dfe);
                print name b stdb t probt;
        finish;

        start test;
                dfn=nrow(r);
                rb=r*b;
                vrb=r*xtxi*r`;
                numer=rb`*inv(vrb)*rb/dfn;
                f=numer/mse;
                prob=1-probf(f,dfn,dfe);
                print f dfn dfe prob;
        finish;

        x={1 1 1,
                1 2 4,
                1 3 9,
                1 4 16,
                1 5 25,
                1 6 36,
                1 7 49,
                1 8 64};
        y={4,
                5,
                7,
                10,
                13,
                17,
                23,
                31};
        name={"Intercept",
```

```
                "Decade",
                "Decade*2"};

        run reg;

        print ,"test b2=b3=0";
        r={0 1 0,
                0 0 1};
        run test;

        print ,"test b2+b3=0";
        r={0 1 1};
        run test;

quit;
```

Lecture Note 7

1 Eigenvalue and Eigenvector

1.1 Eigenvalue

Let

$$\begin{aligned} A &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \\ A &: \text{Symmetric positive definite} \end{aligned}$$

The eigenvalues of A can be calculated as

$$\begin{aligned} \left| \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} - \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \right| &= 0 \\ \begin{vmatrix} \lambda - 2 & -1 \\ -1 & \lambda - 2 \end{vmatrix} &= \\ (\lambda - 2)^2 - 1 &= \\ (\lambda - 2)^2 &= 1 \\ \lambda - 2 &= \pm 1 \\ \lambda &= \begin{cases} 3 \\ 1 \end{cases} \end{aligned}$$

Therefore, $\lambda_1 = 3$ and $\lambda_2 = 1$; So we can compute both trace and determinant of matrix A

$$\begin{aligned} \text{trace}(A) &= \sum_{i=1}^{n} \lambda_i \\ &= 4 \end{aligned}$$

$$\begin{aligned}
\det(A) &= \prod_{i=1}^{n} \lambda_i \\
&= 3 \\
A^{-1/2}A^{-1/2} &= A^{-1} \\
A^{-1} &: \text{Symmetric positive definite}
\end{aligned}$$

1.2 Eigenvector

The eigenvector corresponding to eigenvalue 3 is

$$\begin{aligned}
\begin{pmatrix} 3-2 & -1 \\ -1 & 3-2 \end{pmatrix}\begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} &= \\
\begin{pmatrix} v_{11} - v_{12} \\ -v_{11} + v_{12} \end{pmatrix} &=
\end{aligned}$$

So, $v_{11} = v_{12}$ and hence any number for v_{11} and the same number for v_{12} will qualify as the elements of the eigenvector corresponding to eigenvalue 3; the common practice of obtaining eigenvector is to make the norm of the vector being equal to 1.

Thus

$$\begin{aligned}
\sqrt{v_{11}^2 + v_{12}^2} &= 1 \\
\sqrt{\left(1/\sqrt{2}\right)^2 + \left(1/\sqrt{2}\right)^2} &= \\
v_1 &= \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}
\end{aligned}$$

2 CUSUM and CUSUMSQ Tests

2.1 CUSUM Test

One of the weaknesses of Chow test is that there is no scientific reasoning to divide given samples appropriately. Because of this reason Chow test should be applied based on some priori reasons. However, in a manner of Econometrics, intervention of anything non-scientific should be avoided if possible. CUSUM and CUSUMSQ is more scientific way to find the breaking point of parameters than Chow test because they can stand without any intrusion of analysts' subjectivities.

Here, scaled recursive residuals are defined.

$$v_t = \frac{e_{t|t-1}}{\sqrt{1 + x_t (X'_{t-1} X_{t-1})^{-1} x'_t}}$$

$$t = (k+1), (k+2), \cdots, n$$

Where

$$x_t = (1 \quad x_{1t} \quad x_{2t} \quad \cdots \quad x_{kt})$$

$$e_{t|t-1} = y_t - x_t b_{t-1}$$

$$b_{t-1} = (X'_{t-1} X_{t-1})^{-1} X'_{t-1} Y_{t-1}$$

$$\mathbb{v}ar\left(e_{t|t-1}\right) = \sigma^2 [1 + x_t (X'_{t-1} X_{t-1})^{-1} x'_t]$$

Under the Classical and Normality Assumptions

$$v_t \sim \mathcal{N}(0, \sigma^2)$$

Also, the scaled recursive residuals are pairwise uncorrelated. Thus

$$V \sim \mathcal{N}\left(0, \sigma^2 I_{(n-k)}\right)$$

The CUSUM test statistic is

$$w_t = \sum_{j=k+1}^{t} \frac{v_j}{s}$$

$$t = (k+1), (k+2), \cdots, n$$

$$s^2 : \frac{\text{Residual Sum of Square}}{(n-k)}$$

Where w_t is a cumulative sum and it is plotted against t. With constant parameters, $\mathbb{E}(w_t) = 0$ but with non-constant parameters w_t will diverge from zero. The significance of the departure from the zero line may be assessed by reference to a pair of straight lines that pass through the points

$$\left(k, \pm a\sqrt{n-k}\right) \quad , \quad \left(n, \pm 3a\sqrt{n-k}\right)$$

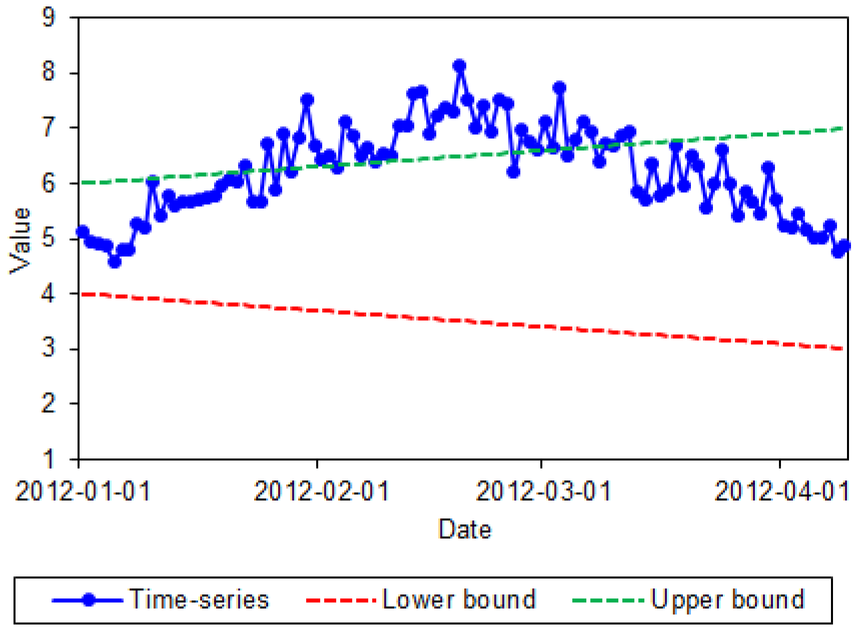
Where a is the parameter depending on the significance level α, chosen from a test. For example

Figure 7.2.1 *99%, 95%, 90% boundary for CUSUM statistics*

$\alpha = 0.01$	$a = 1.143$
$\alpha = 0.05$	$a = 0.948$
$\alpha = 0.10$	$a = 0.850$

Then, based on above numbers, we can construct below plot

Figure 7.2.2 w_t *plot with broken parameters*



In here, it is not intuitive that the model uses the independent variable x_t with the estimator b_{t-1}. Generally, Both the dependent variable y_t and the independent variable x_t are simultaneously observed in the time period t and hence x_t will probably be unknown when it is the point of time to forecast y_t. If, in the case of using the data of $t-1$ as an independent variable, it is reasonable that $e_{t|t-1}$ is One-Step-Ahead-Forecast-Error.

2.2 CUSUMSQ Test

$$s_t = \sum_{j=k+1}^{t} w_j^2 \Big/ \sum_{j=k+1}^{n} w_j^2$$

$$t = (k+1), (k+2), \cdots, n$$

Since both the numerator and the denominator of above fraction follow χ^2 simultaneously, what remained in s_t is just constants peculiarly.

Under the null hypothesis, both w^2s are independent χ^2 variables. Thus

$$\mathbb{E}(s_t) \approx \frac{t-k}{n-k}$$

The significance of departures from the expected value line is assessed by reference to a pain of lines drawn parallel to $\mathbb{E}(s_t)$ line at a distance c_0 above and below.

Refer the charts below.

Figure 7.2.3 *If the parameter is constant.*

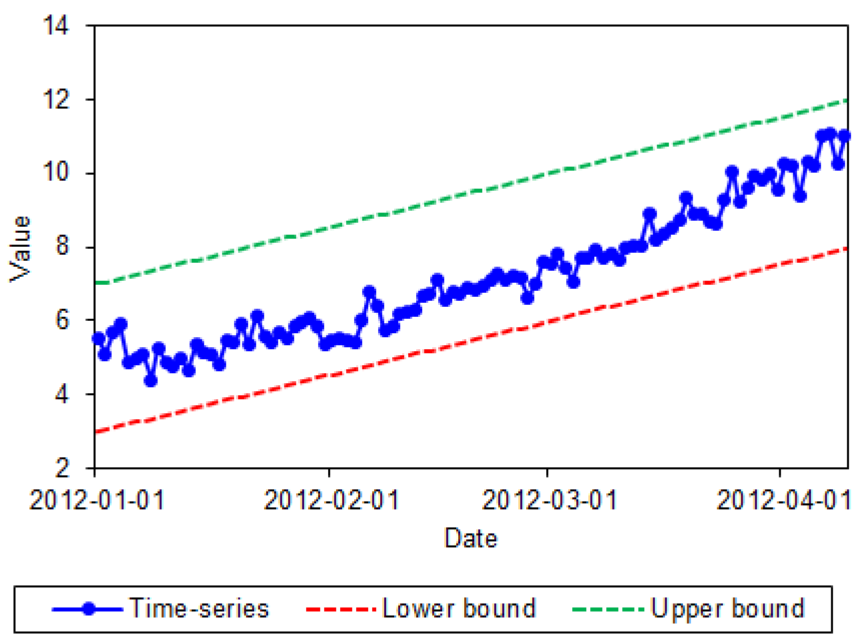
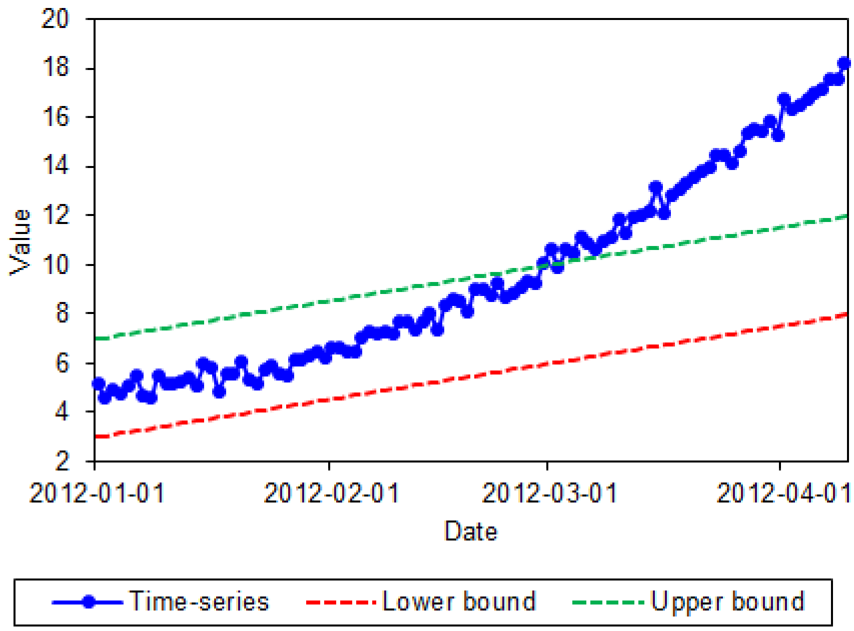


Figure 7.2.4 *If the parameter is broken.*



Code 7.2.1 *Reference code 1; exercise example for CUSUM test and CUSUMSQ test*

```
data cpi;
        infile 'c:₩sas₩cpi.csv' dlm=',' missover firstobs=2;
        input mon yymmdd10. cpi;
run;

data m1sl;
        infile 'c:₩sas₩m1sl.csv' dlm=',' missover firstobs=2;
        input mon yymmdd10. m1sl;
run;

data hw1;
        merge cpi m1sl sr.fedfu(rename=(month=mon));
        by mon;
        cg=dif(log(cpi));
        mg=dif(log(m1sl));
        int=1;
        if mon<input("1959-02-01",yymmdd10.) then delete;
run;

proc iml;
        reset noprint;
        start cusumreg;
                btm1=inv(xtm1`*xtm1)*xtm1`*ytm1;
                yhat=x*btm1;
                e=y-yhat;
                v=e/sqrt(1+x*inv(xtm1`*xtm1)*x`);
                w=v/s;
                w2=w**2;
        finish;
        use hw1;
        read all var{cg} into ytotal;
        use hw1;
        read all var{int mg fedfu} into xtotal;
        n=nrow(xtotal);
        k=ncol(xtotal);
        vtotal=j(n,1,0);
        b=inv(xtotal`*xtotal)*xtotal`*ytotal;
        yh=xtotal*b;
        re=yh-ytotal;
        rss=re`*re;
        s=sqrt(rss/(n-k));
        cus=0;
        cusq=cus;
        cusum=j(n,1,0);
        cusumsq=cusum;
        do t=k+1 to n;
                ytm1=ytotal[1:t-1,];
                xtm1=xtotal[1:t-1,];
                y=ytotal[t,];
                x=xtotal[t,];
```

```
                run cusumreg;
                cus=cus+w;
                cusum[t,]=cus;
                cusq=cusq+w2;
                cusumsq[t,]=cusq;
        end;
        cusumsq=cusumsq/cusq;
        print cusum rss;
        create hw2 from cusum[colname="cusum"];
        append from cusum;
        create hw3 from cusumsq[colname="cusumsq"];
        append from cusumsq;
quit;

data hw4;
        merge hw1(keep=mon) hw2 hw3;
run;

proc gplot data=hw4;
        symbol1 i=join;
        symbol2 i=join;
        plot cusum*mon;
        plot cusumsq*mon;
run;
        symbol1 i=none;
        symbol2 i=none;
quit;
```

Code 7.2.2 *Reference code 2; exercise example for CUSUM test and CUSUMSQ test*

```
data cpi;
        infile 'c:₩sas₩cpi.csv' dlm=',' missover firstobs=2;
        input mon yymmdd10. cpi;
run;

data m1sl;
        infile 'c:₩sas₩m1sl.csv' dlm=',' missover firstobs=2;
        input mon yymmdd10. m1sl;
run;

data hw1;
        merge cpi m1sl fedfu(rename=(month=mon));
        by mon;
        cg=dif(log(cpi));
        mg=dif(log(m1sl));
        int=1;
        if mon<input("1959-02-01",yymmdd10.) then delete;
run;

data hw2;
        set hw1;
        if mon<input("1984-03-01",yymmdd10.) then output hw2;
run;

ods html file='c:₩hw1.html';
proc autoreg data=hw1;
        model cg=mg fedfu/chow=300;
        output out=hw1 cusum=cs cusumsq=csq;
run;

proc gplot data=hw1;
        symbol1 i=spline ci=blue;
        symbol2 i=spline ci=red;
        plot1 cs*mon;
        plot2 csq*mon/overlay;
run;

proc iml;
        reset noprint;
        start reg;
        n=nrow(x);
        k=ncol(x);
        b=inv(x`*x)*x`*y;
        yhat=x*b;
        r=y-yhat;
        rss=r`*r;
        finish;
        use hw2;
        read all var{cg} into y;
        use hw2;
```

```
        read all var{int mg fedfu} into x;
        run reg;
        rss1=rss;
        n1=n;
        use hw1;
        read all var{cg} into y;
        use hw1;
        read all var{int mg fedfu} into x;
        run reg;
        rssall=rss;
        nall=n;
        n2=nall-n1;
        start test;
        df=(n1-k);
        up=(rssall-rss1)/n2;
        down=rss1/df;
        f=up/down;
        p=1-probf(f,n2,df);
        finish;
        run test;
        print f n2 df p;
        ods html close;
quit;
```

3 Test of Structure Change

Structural change in Econometrics is also meant parameter break or structural break. In this chapter, the methodologies which can capture structural change in a way of statistics will be mentioned.

3.1 No Structure Change

$$
\begin{aligned}
\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} &= \begin{pmatrix} i_1 & X_1^* \\ i_2 & X_2^* \end{pmatrix} \begin{pmatrix} \alpha \\ \beta^* \end{pmatrix} + U \\
H_0 &: \alpha_1 = \alpha_2
\end{aligned}
$$

And then, compute Residual Sum of Square (RSS_1)

$$
\begin{aligned}
F &= \frac{RSS_1 - RSS_2/1}{RSS_2/(n-k-1)} \\
&\sim F_{1,(n-k-1)}
\end{aligned}
$$

3.2 Structure Change Only in Intercept

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} i_1 & 0 & X_1^* \\ 0 & i_2 & X_2^* \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta^* \end{pmatrix} + U$$

$$H_0 \quad : \quad \beta_1^* = \beta_2^*$$

And then, compute Residual Sum of Square (RSS_2)

$$F = \frac{RSS_2 - RSS_3/(k-1)}{RSS_3/(n-2k)}$$

$$\sim F_{(k-1),(n-2k)}$$

3.3 Structure Change in Both Intercept and Slope

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} i_1 & 0 & X_1^* & 0 \\ 0 & i_2 & 0 & X_2^* \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1^* \\ \beta_2^* \end{pmatrix} + U$$

$$H_0 \quad : \quad \beta_1 = \beta_2$$

It is not reasonable that the model with structural change just in slopes, except the intercept and hence it should be determined the structural change of all of the coefficients.

$$F = \frac{RSS_1 - RSS_3/k}{RSS_3/(n-2k)}$$

$$\sim F_{k,(n-2k)}$$

Prof. Kiseok Lee (econklee@khu.ac.kr)
TA Junyong Kim (mindhunter@khu.ac.kr)

LECTURE NOTE 8

1 Kalman Filter

Kalman filter is firstly proposed by Kalman[10] and Kalman and Bucy[11] that

$$\theta_t = A\theta_{t-1} + \varepsilon_t$$

A : Assumed known a priori

Evolution of the parameter vector θ_t over time; parameter is time-varying. (Law of Motion)

Thus, the observation equation can be denoted as

$$y_t = x_t'\theta_t + u_t$$

Where y_t and x_t is the column-shaped observation matrix in the time period t and θ_t is also the column-shaped parameter matrix in the same period.

Kalman filter is designed with two steps; the forecasting step and the updating step.

1.1 Forecasting Step

Note that

$$\xi_{1|0} = A\xi_0$$

ξ_0 : Initial estimator of θ_0

$\xi_{1|0}$: Estimator of $\theta_{0|1}$

[10] Rudolf E. Kalman, 1960, "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering* 82 (1)
[11] Rudolf E. Kalman and Richard S. Bucy, 1961, "New Results in Linear Filtering and Prediction Theory", *Journal of Basic Engineering* 83 (1)

And

$$\Phi_{1|0} = A\Phi A' + \Gamma$$

$$y_t = x_1'\xi_{1|0}$$

$$e_{1|0} = y_1 - y_{1|0}$$

$e_{1|0}$: One-Step-Ahead-Forecast-Error

And there is an assumption that the data x_1 is available in the time period zero.

1.2 Updating Step

$$\xi_{1|1} = \xi_{1|0} + K_1\left(y_1 + x_1'\xi_{1|0}\right)$$

$$\Phi_{1|1} = (1 - K_1 x_1')\Phi_{1|0}$$

$$K_1 = \Phi_{1|0}x_1\left(1 + x_1'\Phi_{1|0}x_1\right)^{-1}$$

Since the above $x_1'\Phi_{1|0}x_1$ is a scalar, this process can save huge time because the process of computing any vector's inverse matrix is not only tedious, but also complex. This is the strong point of Kalman filter; it is unnecessary to compute the inverse matrix.

$$y_t = x_t'\theta + u_t$$

Code 8.1.1 *Kalman filter estimation; SAS/ETS version*

```
DATA sp500;
        INFILE "c:₩sp500.prn";
        INPUT mon sp500;
        sp500g=DIF(LOG(sp500));
RUN;

DATA ip;
        INFILE "c:₩ip.prn";
        INPUT mon ip;
        ipg=DIF(LOG(ip));
RUN;

DATA fyff;
        INFILE "c:₩fyff.prn";
        INPUT mon fyff;
RUN;

DATA oilprice;
        INFILE "c:₩oilprice.prn";
        INPUT mon oilprice;
        oilpriceg=DIF(LOG(oilprice));
RUN;

DATA all1;
        MERGE sp500 ip fyff oilprice;
        BY mon;
        IF mon<19590101 THEN DELETE;
        IF mon>20101231 THEN DELETE;
RUN;

PROC AUTOREG DATA=all1;
        MODEL sp500g=ipg fyff oilpriceg;
        OUTPUT OUT=all1 CUSUM=cusum;
RUN;

DATA all1;
        SET all1;
        snosafe=DIF(cusum);
        IF snosafe=. THEN snosafe=cusum;
RUN;

QUIT;
```

Code 8.1.2 *Kalman filter estimation; SAS/IML version*

```
DATA sp500;
	INFILE "c:₩sp500.prn";
	INPUT mon sp500;
RUN;

DATA ip;
	INFILE "c:₩ip.prn";
	INPUT mon ip;
RUN;

DATA fyff;
	INFILE "c:₩fyff.prn";
	INPUT mon fyff;
RUN;

DATA wti;
	INFILE "c:₩wti.prn";
	INPUT mon wti;
RUN;

DATA all1;
	MERGE sp500 ip fyff wti;
	BY mon;
	int=1;
	sg=DIF(LOG(sp500));
	ig=DIF(LOG(ip));
	wg=DIF(LOG(wti));
	IF mon<19590101 THEN DELETE;
	IF mon>20101231 THEN DELETE;
RUN;

PROC IML;
	RESET NOPRINT;
	USE all1;
	READ ALL VAR {sg} INTO y;
	READ ALL VAR {int ig fyff wg} INTO x;
	n=NROW(x);
	k=NCOL(x);
	nosafe_save=J(n,1,0);
	b=((x[1:k,]`*x[1:k,])**-1)*x[1:k,]`*y[1:k,];
	phi=(x[1:k,]`*x[1:k,])**-1;
	DO t=k+1 TO n;
		osafy=x[t,]*b;
		osafe=y[t,]-osafy;
		nosafe=osafe*((1+x[t,]*phi*x[t,]`)**-0.5);
		nosafe_save[t,]=nosafe;
		k=phi*x[t,]`*((1+x[t,]*phi*x[t,]`)**-1);
		b=b+k*osafe;
		phi=phi-k*x[t,]*phi;
	END;
```

```
        PRINT nosafe_save;
        CREATE nosafe FROM nosafe_save;
        APPEND FROM nosafe_save;
QUIT;

DATA nosafe;
        SET nosafe(RENAME=(COL1=nosafe));
        cnosafe+nosafe;
RUN;

QUIT;
```

2 Multicollinearity

Consider the model

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

The multicollinearity problem arises from the fact that the two independent variables X_{1t} and X_{2t} are collinear; they are correlated with each other. This problem is not due to any violation of the three classical assumptions. Instead, the multicollinearity is a data problem. Since the classical assumptions are not violated, OLS estimators of βs are still BLUE.

2.1 Symptom

Following (1), (2), (3), and (4) are the symptoms of multicollinearity.

(1) A large R^2 but few estimators are statistically significant.

(2) A small change in the sample size can induce large changes in parameter estimates; overlapped information.

(3) The estimate of a parameter may be overwhelmingly large, while the estimate of other parameter is unreasonably small.

(4) If one independent variable is regressed upon other independent variables, R^2 will become large.

2.2 Remedy

(1) Add as many observations as possible if you can.

(2) Drop an independent variable if it does not lead to the specification error.

Ex. For a model

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

Assume that Y_t is consumption, X_{1t} is income, and X_{2t} is wealth. Then it is not possible to specify or fix its observations. Probably those two variables will go up or down simultaneously.

Think about radical aspects.

$$\beta_1 = \frac{\partial Y_t}{\partial X_{1t}}$$

And

$$\beta_2 = \frac{\partial Y_t}{\partial X_{2t}}$$

In this model, if there is higher correlation between the variable X_{1t} and X_{2t}, those variables is not partializable; cannot be independently partialized.

For example, if both inflation and the baserate are used, multicollinearity problem will arise when you estimate the model.

Lecture Note 9

If the Classical Assumption is satisfied, it will be good to estimate the model with the OLS estimator. However, if the model which should be treated is non-linear, then the OLS estimation won't be applicable and hence other ways should be determined. The maximum-likelihood estimation is one of the alternatives.

1 Maximum Likelihood Estimation

Let Y' be an $(1 \times n)$ vector of sample values

$$Y' = (Y_1 \quad Y_2 \quad \cdots \quad Y_n)$$

Which dependent on any θ, which is $(1 \times k)$ vector of unknown parameters

$$\Theta' = (\theta_1 \quad \theta_2 \quad \cdots \quad \theta_k)$$

Let the joint density function be written as

$$f(Y;\Theta)$$

This density may be interpreted in 2 different ways. For a given Θ, it indicates the probability of a set of sample outcomes. Alternatively, it may be interpreted as a function of Θ, conditional on a set of sample outcomes. In the latter interpretaion it is referred to as a likelihood function. The formal definition is

$$\mathcal{L}(Y;\Theta) = f(Y_1;\Theta_1)f(Y_2;\Theta_2)\cdots f(Y_n;\Theta_n)$$

Maximizing the likelihood function with respect to Θ amounts to finding a specific value, say $\hat{\Theta}$, which maximizes the probability of obtaining the sample value that actually been observed. Then, $\hat{\Theta}$ is said to be the maximum-likelihood estimator of the unknown parameter vector Θ.

In most applications, it is simpler to maximize the log likelihood function

$$\ell = \ln \mathcal{L}$$

Then

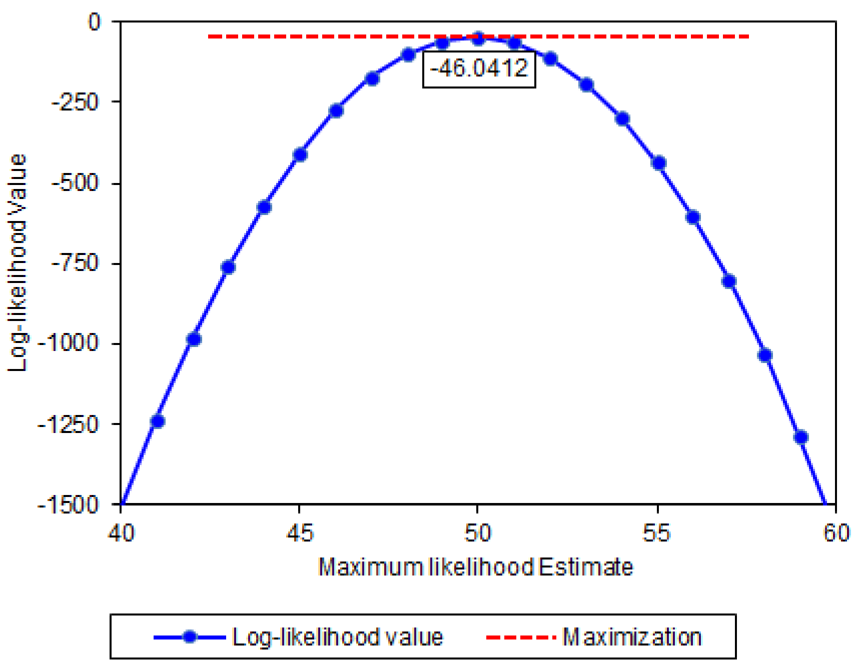
$$\frac{\partial \ell}{\partial \Theta} = \frac{1}{\mathcal{L}} \times \frac{\partial \mathcal{L}}{\partial \Theta}$$

The derivative of ℓ with regarding to Θ is known as the temporary score function $S(\Theta; Y)$. The maximum-likelihood estimator $\hat{\Theta}$ is obtained by setting the score to zero, that is, by finding the value of Θ that solves

$$\begin{aligned} S(\Theta; Y) &= \frac{\partial \ell}{\partial \Theta} \\ &= 0 \end{aligned}$$

2 Properties of MLE

Figure 9.2.1 *Idea of MLE; recall of Figure 7.1.1 of Basic Econometrics*



For the estimator vector $\hat{\Theta}$ of the parameter vector Θ, which is calculated in the manner of maximum-likelihood, below properties are satisfied that

Property (Consistency)

$$\hat{\Theta} \quad \to^{p} \quad \Theta$$

Or

$$\text{plim}\,\hat{\Theta} \quad = \quad \Theta$$

This implies following expression. For all $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\hat{\Theta} - \Theta\right| < \varepsilon\right) \quad = \quad 1$$

Property (Asymptotic Normality)

$$\hat{\Theta} \quad \sim^{a} \quad \mathcal{N}[\Theta, I^{-1}(\Theta)]$$

This states that the asymptotic distribution of $\hat{\Theta}$ is normal with the mean Θ and the variance $I^{-1}(\Theta)$. $I(\Theta)$ is the information matrix and is defined in two equivalent ways by

$$\begin{aligned} I(\Theta) \quad &= \quad \mathbb{E}\left[\left(\frac{\partial \ell}{\partial \Theta}\right) \times \left(\frac{\partial \ell}{\partial \Theta}\right)'\right] \\ &= \quad -\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \Theta \partial \Theta'}\right) \end{aligned}$$

Note that

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \Theta \partial \Theta'}\right) \quad = \quad \begin{pmatrix} \dfrac{\partial^2 \ell}{\partial {\Theta_1}^2} & \dfrac{\partial^2 \ell}{\partial \Theta_1 \partial \Theta_2} & \cdots & \dfrac{\partial^2 \ell}{\partial \Theta_1 \partial \Theta_k} \\ \dfrac{\partial^2 \ell}{\partial \Theta_2 \partial \Theta_1} & \dfrac{\partial^2 \ell}{\partial {\Theta_2}^2} & \cdots & \dfrac{\partial^2 \ell}{\partial \Theta_2 \partial \Theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 \ell}{\partial \Theta_k \partial \Theta_1} & \dfrac{\partial^2 \ell}{\partial \Theta_k \partial \Theta_2} & \cdots & \dfrac{\partial^2 \ell}{\partial {\Theta_k}^2} \end{pmatrix}$$

Property (Asymptotic Efficiency) *If $\hat{\Theta}$ is a maximum-likelihood estimator of a single parameter Θ*

$$\sqrt{n}\left(\hat{\Theta} - \Theta\right) \quad \sim^{d} \quad \mathcal{N}(0, \sigma^2)$$

For some finite constant σ^2, in distribution.

If $\tilde{\Theta}$ denotes any other consistent that is the asymptotically normal estimator of Θ, then $\sqrt{n}\tilde{\Theta}$ has a normal limiting distribution whose variance is greater than or equal to σ^2. The maximum-likelihood estimator has the minimum variance in the class of consistent, asymptotically normal estimators.

This scheme is really close to Gauss-Markov theorem which mentions about the characteristic of several properties of the OLS estimator in the aspects based on the Classical Assumption; if the MLE $\hat{\Theta}$ satisfies consistency and asymptotic normality then it will have smallest variance among those estimators which also satisfies consistency and asymptotic normality.

Property (Invariance) *If $\hat{\Theta}$ is the MLE of Θ and $g(\Theta)$ is a continuous function of Θ, then $g(\hat{\Theta})$ is also the MLE of $g(\Theta)$.*

$$\begin{aligned} f(\Theta) &= \Theta + \Theta^2 \\ f(\hat{\Theta}) &= \hat{\Theta} + \hat{\Theta}^2 \\ &= \hat{f}(\Theta) \end{aligned}$$

Confer that, in the regression model

$$Y_t = \alpha + \beta X_t + u_t$$

For a notation of u_ts' normality, simply it can be denoted as

$$u_t \sim \mathcal{N}(0, \sigma^2) \leftarrow \forall t = 1,2,\cdots,n$$

3 MLE in Linear Model

Suppose below model

$$\begin{aligned} Y &= XB + U \\ U &\sim \mathcal{N}(0, \sigma^2 I_n) \end{aligned}$$

Then *Probability Density Function* of above model can be depicted as

$$\begin{aligned} f(u_t; B) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{u_t}{\sigma}\right)^2} \\ f(U; B) &= \mathcal{L} \\ &= \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{u_t}{\sigma}\right)^2} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{n} u_t^2} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}U'U} \end{aligned}$$

Above *Joint Probability Density Function* can be transformed into logarithm as

$$
\begin{aligned}
\ln f(U;B) &= \ln \mathcal{L} \\
&= \ell \\
&= \ln\left[(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}U'U}\right] \\
&= -\frac{n}{2}\ln(2\pi\sigma^2) + \left(-\frac{1}{2\sigma^2}U'U\right) \\
&= -\frac{n}{2}(\ln 2\pi + \ln\sigma^2) + \left(-\frac{1}{2\sigma^2}U'U\right) \\
&= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}U'U \\
&= -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}(Y-XB)'(Y-XB)
\end{aligned}
$$

Since

$$
\begin{aligned}
(Y-XB)'(Y-XB) &= Y'Y - Y'XB - B'X'Y + B'X'XB \\
&= -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}(Y'Y - Y'XB - B'X'Y + B'X'XB)
\end{aligned}
$$

Then

$$
\max_{\tilde{B},\tilde{\sigma}} \ell \quad \Rightarrow \quad \frac{\partial \ell}{\partial \tilde{B}}
$$

$$
\begin{aligned}
\frac{\partial \ell}{\partial \tilde{B}} &= \frac{\partial}{\partial \tilde{B}}\left[-\frac{1}{2\sigma^2}(Y'Y - Y'XB - B'X'Y + B'X'XB)\right] \\
&= -\frac{1}{2\sigma^2}\left(0 - X'Y - X'Y + 2X'X\hat{B}\right) \\
&= 0
\end{aligned}
$$

Therefore

$$
\begin{aligned}
-2X'Y + 2X'X\hat{B} &= 0 \\
X'X\hat{B} &= X'Y \\
\hat{B} &= (X'X)^{-1}X'Y \\
&= \hat{b}
\end{aligned}
$$

This is exactly same with the OLS estimator. However,

$$
\max_{\tilde{B},\tilde{\sigma}} \ell \quad \Rightarrow \quad \frac{\partial \ell}{\partial \tilde{\sigma}}
$$

$$
\begin{aligned}
\frac{\partial \ell}{\partial \tilde{\sigma}} &= \frac{\partial}{\partial \tilde{\sigma}}\left(-\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}U'U\right) \\
&= -\frac{n}{\hat{\sigma}} + \frac{1}{4\hat{\sigma}^4}U'U4\hat{\sigma} \\
&= 0
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\frac{n}{\hat{\sigma}} &= \frac{1}{4\hat{\sigma}^4}U'U4\hat{\sigma} \\
&= \frac{1}{\hat{\sigma}^3}U'U \\
n\hat{\sigma}^2 &= U'U \\
\hat{\sigma}^2 &= \frac{1}{n}U'U \\
&= \frac{1}{n}\sum_{t=1}^{n} u_t^2 \\
&\neq s^2
\end{aligned}
$$

The estimator of variance of the model is slightly different from the OLS estimator and also the maximum-likelihood estimator of variance is biased because

$$
\mathbb{E}(\hat{\sigma}^2) = \sigma^2
$$

However,

$$
\mathbb{v}ar(\hat{\sigma}^2) < \mathbb{v}ar(s^2)
$$

This implies that the variance of the variance estimator of MLE is much smaller than that of the variance estimator of OLS. But since biased, the estimator is no more the BLUE unfortunately.

$$
\mathbb{v}ar - \mathbb{c}ov(U) = \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix}
$$

For all $t \neq s$, us are pairwise uncorrelated. If u_t and u_s are normally distributed and uncorrelated, then u_t and u_s are independent; then the joint density of u_t and u_s, $f(u_t, u_s)$ is

$$
f(u_t, u_s) = f(u_t)f(u_s)
$$

Note that

(1) If independent, then uncorrelated. (True)

(2) If uncorrelated, then independent. (False)

(3) If uncorrelated and normally distributed, then independent. (True)

4 MLE in Non-linear Model

$$\begin{aligned} Y &= XB + U \\ u_t &= \phi u_{t-1} + \varepsilon_t \end{aligned}$$

In here, for the first

$$\frac{\partial^2 \ell}{\partial B \partial B'} = \frac{-X'X}{\sigma^2}$$

With regarding to

$$-\mathbb{E}\left(\frac{\partial^2 \ell}{\partial B \partial B'}\right) = \frac{X'X}{\sigma^2}$$

And for the second

$$\frac{\partial^2 \ell}{\partial B \partial \sigma^2} = -\frac{1}{\sigma^4}(X'Y - X'XB)$$

With regarding to

$$-\mathbb{E}\left(\frac{\partial^2 \ell}{\partial B \partial \sigma^2}\right) = 0$$

For the third

$$\frac{\partial^2 \ell}{\partial {\sigma^2}^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}(Y - XB)'(Y - XB)$$

With regarding to

$$-\mathbb{E}\left(\frac{\partial^2 \ell}{\partial {\sigma^2}^2}\right) = \frac{n}{2\sigma^4}$$

Thus, information matrix can be depicted as

$$I(\Theta) = \begin{pmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

Take the inverse to $I(\Theta)$ and then

$$I^{-1}(\Theta) = \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}$$

Therefore

$$
\begin{aligned}
\mathcal{L} &= (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{1}{2\hat{\sigma}^2}e'e} \\
&= (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{1}{2\hat{\sigma}^2}n\hat{\sigma}^2} \\
&= (2\pi)^{-\frac{n}{2}} \hat{\sigma}^{-n} e^{-\frac{n}{2}} \\
&= \text{Constant} \times \hat{\sigma}^{-n} \times \text{Constant} \\
&= \text{Constant} \times (e'e)^{-\frac{n}{2}}
\end{aligned}
$$

Hence

$$
\begin{aligned}
\ell^* &= \ln(\text{Constant}) - \frac{n}{2}\ln(e'e) \\
\ell^* &\propto -\frac{n}{2}\ln(e'e)
\end{aligned}
$$

Here, $e'e$ is equal to Residual Sum of Square

5 Hypothesis Test

5.1 Likelihood Ratio Test

$$
H_0 \; : \; RB = r
$$

For this test

$$
\begin{aligned}
\lambda &= \frac{\ell(\tilde{B}, \tilde{\sigma}^2)}{\ell(\hat{B}, \hat{\sigma}^2)} \\
\tilde{B}, \tilde{\sigma}^2 &: \text{Maximum likelihood estimator under } H_0 \text{ (w/ restriction)} \\
\hat{B}, \hat{\sigma}^2 &: \text{Maximum likelihood estimator under } H_1 \text{ (w/o restriction)} \\
LR &= -2\, ln\, \lambda \\
&= n \ln\left(1 + \frac{e_*'e_* - e'e}{e'e}\right) \\
&\sim^a \chi^2(q) \\
e_* &: \text{Under the null} \\
e &: \text{Under the alternative}
\end{aligned}
$$

5.2 Wald Test

Only the unrestricted $\hat{B}$ is calculated as

$$\frac{\left(R\hat{B}-r\right)'[RI^{-1}(B)R']^{-1}\left(R\hat{B}-r\right)}{\hat{\sigma}^2} \quad \sim^a \quad \chi^2(q)$$

Or

$$\begin{aligned} W &= \frac{n(e_*'e_* - e'e)}{e'e} \\ &\sim^a \chi^2(q) \end{aligned}$$

5.3 Lagrange Multiplier Test

$$\frac{ne_*'X(X'X)^{-1}X'e_*}{e_*'e_*} \quad \sim^a \quad \chi^2(q)$$

So, Lagrange Multiplier test only requires restricted $\hat{B}$. This is different from Wald test because Wald test only requires unrestricted $\hat{B}$. Or

$$\begin{aligned} LM &= \frac{n(e_*'e_* - e'e)}{e_*'e_*} \\ &\sim^a \chi^2(q) \end{aligned}$$

Note that $W \geq LR \geq LM$

Code 9.5.1 *Exercise for MLE estimation for non-linear model; GARCH(1,1) model*

```
proc iml;
    reset noprint;
    use temp0;
    read all var {dkospi} into x;
    xmean=mean(x);
    xdeviation=x-xmean;
    xsse=xdeviation`*xdeviation;
    xvariance=xsse/(nrow(x)-1);
    xstandarddeviation=sqrt(xvariance);
    NT=nrow(x);
    RR=x;
    VV=xvariance;
    start garch(parameter) global(NT,RR,VV);
        if parameter[2]<0 then return(-999999);
        if parameter[3]<0 then return(-999999);
        if parameter[4]<0 then return(-999999);
        ht=parameter[2]+parameter[4]*VV;
        et2=(RR[1]-parameter[1])**2;
        LL=-0.5*log(2*constant("pi"))-0.5*log(ht)-0.5*et2/ht;
        do t=2 to NT;
            ht=parameter[2]+parameter[3]*et2+parameter[4]*ht;
            et2=(RR[t]-parameter[1])**2;
            LL=LL-0.5*log(2*constant("pi"))-0.5*log(ht)-0.5*et2/ht;
        end;
        return(LL);
    finish garch;
    opt={1 2};
    parameter={0 0.1 0.1 0.8};
    parameter[1]=xmean;
    call nlpqn(rc,xr,"garch",parameter,opt);
    call nlpfdd(ll,grad,hess,"garch",xr);
    vcv=inv(hess);
    std=sqrt(abs(vecdiag(vcv)));
    xrt=xr`;
    tval=xrt/std;
    dfe=nrow(x)-nrow(xr);
    pval=1-probf(tval##2,1,dfe);
    print xrt std tval pval;
quit;
```

Lecture Note 10

Now Updating, Coming Soon

Now Updating, Coming Soon

Now Updating, Coming Soon

Now Updating, Coming Soon

Now Updating, Coming Soon

LECTURE NOTE 11

1 Regression Model with AR Error

$$
\begin{aligned}
Y &= XB + U \\
u_t &= \alpha u_{t-1} + \varepsilon_t
\end{aligned}
$$

Note that, particularly in the expression of SAS, following expression is adopted that

$$(1 + \alpha L) u_t = \varepsilon_t$$

Therefore, if the outcome is calculated as -0.38, this means 0.38 in very above equation, i.e. opposite sign.

$$
\begin{aligned}
Y_t &= \beta_0 + \beta_1 X_{t-4} + u_t \\
Y_t &: \text{Growth rate of industrial production} \\
X_{t-4} &: \text{Federal Fund Rate}
\end{aligned}
$$

Try X_t, X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, and X_{t-5} as an independent variable and then choose the one that gives the highest absolute t value.

If AR(2), it can be denoted as

$$
\begin{aligned}
Y &= XB + U \\
u_t &= \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \varepsilon_t
\end{aligned}
$$

SAS might express within following manner

$$
\begin{aligned}
Y &= XB + U \\
(1 + \alpha_1 L + \alpha_2 L^2) u_t &= \varepsilon_t
\end{aligned}
$$

Because of this reason, α_1 and α_2 should be interpreted in opposite way.

2 Regression and Time-series

For a regression model

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

Y_t : Dependent variable

X_t : Independent variable

And a time-series model

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t$$

Y_t : Dependent variable

Y_{t-1} : Lagged dependent variable

Their pros and cons can be described as below table.

Figure 11.2.1 *Regression Model and Time-series Model; recall of Figure 6.1.3*

Pros and Cons	Model	
	Regression	Time-series
Pros	Economic theory can be reflected in the model.	Forecasting is very easy.
Cons	Forecasting is very difficult.	Economic theory cannot be reflected in the model.

3 AR(1) Model

For below time-series model

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

Thus, its autocovariance structure is

$$\begin{aligned}\gamma_0 &= \mathbb{v}ar(Y_t) \\ &= \frac{\sigma_\varepsilon^2}{1-\alpha^2} \\ \gamma_1 &= \mathbb{c}ov(Y_t, Y_{t-1})\end{aligned}$$

$$= \alpha\left(\frac{\sigma_\varepsilon^2}{1-\alpha^2}\right)$$

$$\gamma_2 = \mathbb{c}ov(Y_t, Y_{t-2})$$

$$= \alpha^2\left(\frac{\sigma_\varepsilon^2}{1-\alpha^2}\right)$$

And its autocorrelation structure is

$$\rho_0 = \mathbb{c}orr(Y_t)$$

$$= 1$$

$$\rho_1 = \mathbb{c}orr(Y_t, Y_{t-1})$$

$$= \alpha\left(\frac{\sigma_\varepsilon^2}{1-\alpha^2}\right)\Big/\left(\frac{\sigma_\varepsilon^2}{1-\alpha^2}\right)$$

$$= \alpha$$

$$\rho_2 = \mathbb{c}orr(Y_t, Y_{t-2})$$

$$= \alpha^2\left(\frac{\sigma_\varepsilon^2}{1-\alpha^2}\right)\Big/\left(\frac{\sigma_\varepsilon^2}{1-\alpha^2}\right)$$

$$= \alpha^2$$

Autocorrelation function is the collection of autocorrelations and autocorrelation function is a sufficient statistic for identifying the time-series model.

3.1 Order of Time-series Analysis

(1) Model building

(2) Estimation of model parameter

(3) Diagnostic Checking

Look at the charts below narrowly that

Figure 11.3.1 *Stably declining autocorrelation*

```
                              Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      1.859129        1.00000    |                    |********************|            0
  1      1.236349        0.66502    |                   .|*************       |     0.031623
  2      0.803259        0.43206    |                  . |*********           |     0.043411
  3      0.544221        0.29273    |                  . |******              |     0.047517
  4      0.378929        0.20382    |                  . |****                |     0.049287
  5      0.256970        0.13822    |                  . |***                 |     0.050123
  6      0.169530        0.09119    |                  . |**                  |     0.050503
  7      0.078953        0.04247    |                  . |*.                  |     0.050667
  8      0.023710        0.01275    |                  . | .                  |     0.050703
  9      0.025706        0.01383    |                  . | .                  |     0.050706
 10      0.048764        0.02623    |                  . |*.                  |     0.050710

                         "." marks two standard errors
```

Figure 11.3.2 *Non-stably declining*

```
                              Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      1.515984        1.00000    |                    |********************|            0
  1      0.660420        0.43564    |                   .|*********           |     0.031623
  2     -0.100049        -.06600    |                   *|.                   |     0.037142
  3     -0.054779        -.03613    |                   *|.                   |     0.037260
  4    -0.0055517        -.00366    |                   .|.                   |     0.037295
  5      0.012327        0.00813    |                   .|.                   |     0.037295
  6      0.014091        0.00930    |                   .|.                   |     0.037297
  7     -0.040514        -.02672    |                   *|.                   |     0.037299
  8     -0.080350        -.05300    |                   *|.                   |     0.037318
  9     -0.037276        -.02459    |                   .|.                   |     0.037393
 10      0.026316        0.01736    |                   .|.                   |     0.037410

                         "." marks two standard errors
```

Code 11.3.1 *Generation and Estimation of AR process and MA process*

```
data exercise;
        ylag=rannor(1);
        ulag=rannor(2);
        do t=1 to 1000;
                u=rannor(2);
                y=0.7*ylag+u;
                z=u+0.7*ulag;
                output;
                ylag=y;
                ulag=u;
        end;
run;

proc arima data=exercise;
        identify var=y nlag=10;
        identify var=z nlag=10;
run;

quit;
```

In SAS, you can find that there are two factors in autocorrelation plot. First one is asterisk (*) and the second is dot (.) which is distributed on above plot. Simply, asterisks mean correlations among time orders and dots mean its standard errors. So if the length of the kth asterisk bar is enough longer to hide the kth dot on the same level, it can be interpreted that there is autocorrelation between the time period t and $(t-k)$ in the scheme of Statistics and hence, based on this figures, it is possible to forecast the $(t+k)$ future value reasonably.

3.2 AR(p) Model

For AR(2) model, it can be expressed as

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t$$

Autocorrelation function is complicated. But the shape of the autocorrelation function is similar to that of AR(1) model.

Also, for AR(3) model,

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + a_3 Y_{t-3} + \varepsilon_t$$

Because all of AR(p) format model has similar autocorrelation structure, it should not be judged by graphically with autocorrelation plot. Thus, we should check its partial autocorrelation plot instead of autocorrelation plot.

$$\begin{aligned} Y_t &= \phi_{11} Y_{t-1} + \varepsilon_t \\ &= \phi_{21} Y_{t-1} + \phi_{22} Y_{t-2} + \varepsilon_t \\ &= \phi_{31} Y_{t-1} + \phi_{32} Y_{t-2} + \phi_{33} Y_{t-3} + \varepsilon_t \\ &\vdots \end{aligned}$$

Partial autocorrelation function is the collection of ϕ_{11}, ϕ_{22}, ϕ_{33}, $\cdots$

This scheme implies that if the God's model is AR(1), then ϕ_{11} will be statistically significant, otherwise ϕ_{22} and ϕ_{33} will not. Through the function plot of partial autocorrelation, the structure of autocorrelation can be easily captured.

And then, you can find another table that contains

Figure 11.3.3 *Fit statistics given by SAS*

Constant Estimate	-0.00548
Variance Estimate	0.951403
Std Error Estimate	0.975399
AIC	1397.021
SBC	1409.665
Number of Residuals	500

* AIC and SBC do not include log determinant.

From above table, we can find the AIC and SBC statistics, by which we can determine whether the estimated model is accurate or not.

$$\text{AIC} = -2\ell^* + 2k$$

ℓ^* : Maximized log-likelihood value

k : The number of parameters

$2k$: Penalty on the number of parameters

On the other hand,

$$\text{SBC} = -2\ell^* + k\ln n$$

$k\ln n$: Penalty in a manner of SBC

Figure 11.3.4 *Autocorrelation checking table given by SAS*

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.26	4	0.5161	-0.001	-0.014	0.046	-0.045	-0.018	0.042
12	6.70	10	0.7530	0.019	0.014	0.003	0.031	0.033	0.064
18	9.52	16	0.8906	-0.012	0.012	0.052	0.026	0.035	-0.024
24	10.20	22	0.9844	0.008	-0.023	0.012	-0.004	0.013	-0.019
30	15.82	28	0.9684	0.038	-0.030	0.002	0.041	-0.077	0.025
36	18.40	34	0.9866	-0.041	0.003	0.023	0.033	0.007	-0.038
42	24.33	40	0.9760	0.003	-0.078	-0.011	0.031	0.060	0.013
48	32.10	46	0.9402	-0.069	-0.022	-0.059	-0.002	0.022	-0.070

Question *Why do we give some penalty for the number of parameters?*

Answer *Principles of Parsimonious*

We don't want to obtain AR(34) or AR(61) instead of AR(2) because those are neither clear nor efficient. The less the number of parameters we should estimate are, the better the model is. Intuitively, remind Occam's razor.

The final stage of time-series analysis is diagnostic check.

From the very first row of above table, To Lag 6, its DF (degrees of freedom) is 4 because above printed SAS screen comes from AR(2) model experiment. And also all of lag above don't have significant χ^2 statistics, which implies that all of ε_ts are independent and identically distributed, i.e. there is no information which is useful for prediction of futures.

4 Forecasting of AR Model

4.1 Forecasting of AR(1) Model

For the first, Y_t and Y_{t+1} can be expressed as

$$
\begin{aligned}
Y_t &= \alpha Y_{t-1} + \varepsilon_t \\
Y_{t+1} &= \alpha Y_t + \varepsilon_{t+1} \\
\Rightarrow \mathbb{E}(Y_{t+1}|\Omega_t) &= \mathbb{E}(\alpha Y_t + \varepsilon_{t+1}|\Omega_t) \\
&= \alpha Y_t + \mathbb{E}(\varepsilon_{t+1}|\Omega_t) \\
\because \mathbb{E}(\varepsilon_{t+1}|\Omega_t) &= \mathbb{E}(\varepsilon_{t+1}) \\
&= 0 \\
\therefore \mathbb{E}(Y_{t+1}|\Omega_t) &= \alpha Y_t
\end{aligned}
$$

And second, Y_{t+2} can be denoted as

$$
\begin{aligned}
Y_{t+2} &= \alpha Y_{t+1} + \varepsilon_{t+2} \\
\Rightarrow \mathbb{E}(Y_{t+2}|\Omega_t) &= \mathbb{E}(\alpha Y_{t+1} + \varepsilon_{t+2}|\Omega_t) \\
&= \alpha \mathbb{E}(Y_{t+1}|\Omega_t) + \mathbb{E}(\varepsilon_{t+2}|\Omega_t) \\
&= \alpha^2 Y_t + \mathbb{E}(\varepsilon_{t+2}|\Omega_t) \\
\because \mathbb{E}(\varepsilon_{t+2}|\Omega_t) &= \mathbb{E}(\varepsilon_{t+2}) \\
&= 0 \\
\therefore \mathbb{E}(Y_{t+2}|\Omega_t) &= \alpha^2 Y_t
\end{aligned}
$$

And continuously

$$
\begin{aligned}
\mathbb{E}(Y_{t+3}|\Omega_t) &= \alpha^3 Y_t \\
\mathbb{E}(Y_{t+4}|\Omega_t) &= \alpha^4 Y_t \\
&\vdots \\
\mathbb{E}(Y_{t+k}|\Omega_t) &= \alpha^k Y_t
\end{aligned}
$$

In AR(1) model, the unconditional mean of time-series is equal to zero.

$$
\begin{aligned}
Y_t &= \alpha Y_{t-1} + \varepsilon_t \\
\because Y_{t-1} &= \alpha Y_{t-2} + \varepsilon_{t-1} \\
\therefore Y_t &= \alpha(\alpha Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \alpha^2 Y_{t-2} + \alpha\varepsilon_{t-1} + \varepsilon_t \\
&= \alpha^2(\alpha Y_{t-3} + \varepsilon_{t-2}) + \alpha\varepsilon_{t-1} + \varepsilon_t \\
&= \alpha^3 Y_{t-3} + \alpha^2\varepsilon_{t-2} + \alpha\varepsilon_{t-1} + \varepsilon_t \\
&= \alpha^3(\alpha Y_{t-4} + \varepsilon_{t-3}) + \alpha^2\varepsilon_{t-2} + \alpha\varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
\Rightarrow Y_t &= \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \cdots \\
\therefore \mathbb{E}(Y_t) &= \mathbb{E}(\varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \cdots) \\
&= \mathbb{E}(\varepsilon_t) + \alpha\mathbb{E}(\varepsilon_{t-1}) + \alpha^2\mathbb{E}(\varepsilon_{t-2}) + \cdots \\
&= 0 + 0 + 0 + \cdots \\
&= 0
\end{aligned}
$$

4.2 Forecasting of AR(2) Model

In a same manner, Y_t and Y_{t+1} in AR(2) model can be denoted and expanded as

$$
\begin{aligned}
Y_t &= \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t \\
Y_{t+1} &= \alpha_1 Y_t + \alpha_2 Y_{t-1} + \varepsilon_{t+1}
\end{aligned}
$$

To forecast, its distribution-determined expected value should be taken as

$$
\Rightarrow \mathbb{E}(Y_{t+1}|\Omega_t) = \mathbb{E}(\alpha_1 Y_t + \alpha_2 Y_{t-1} + \varepsilon_{t+1}|\Omega_t)
$$

$$
\begin{aligned}
&= \alpha_1 Y_t + \alpha_2 Y_{t-1} + \mathbb{E}(\varepsilon_{t+1}|\Omega_t) \\
&= \alpha_1 Y_t + \alpha_2 Y_{t-1} \\
\Rightarrow \mathbb{E}(Y_{t+2}|\Omega_t) &= \mathbb{E}(\alpha_1 Y_{t+1} + \alpha_2 Y_t + \varepsilon_{t+2}|\Omega_t) \\
&= \alpha_1 \mathbb{E}(Y_{t+1}|\Omega_t) + \alpha_2 Y_t + \mathbb{E}(\varepsilon_{t+2}|\Omega_t) \\
&= \alpha_1(\alpha_1 Y_t + \alpha_2 Y_{t-1}) + \alpha_2 Y_t \\
&= \alpha_1^2 Y_t + \alpha_1\alpha_2 Y_{t-1} + \alpha_2 Y_t \\
&= (\alpha_1^2 + \alpha_2) Y_t + \alpha_1\alpha_2 Y_{t-1} \\
\Rightarrow \mathbb{E}(Y_{t+3}|\Omega_t) &= \mathbb{E}(\alpha_1 Y_{t+2} + \alpha_2 Y_{t+1} + \varepsilon_{t+3}|\Omega_t) \\
&= \alpha_1 \mathbb{E}(Y_{t+2}|\Omega_t) + \alpha_2 \mathbb{E}(Y_{t+1}|\Omega_t) + \mathbb{E}(\varepsilon_{t+3}|\Omega_t) \\
&= \alpha_1[(\alpha_1^2 + \alpha_2) Y_t + \alpha_1\alpha_2 Y_{t-1}] + \alpha_2(\alpha_1 Y_t + \alpha_2 Y_{t-1}) \\
&= \alpha_1(\alpha_1^2 + \alpha_2) Y_t + \alpha_1^2\alpha_2 Y_{t-1} + \alpha_1\alpha_2 Y_t + \alpha_2^2 Y_{t-1} \\
&= \alpha_1^3 Y_t + \alpha_1\alpha_2 Y_t + \alpha_1^2\alpha_2 Y_{t-1} + \alpha_1\alpha_2 Y_t + \alpha_2^2 Y_{t-1} \\
&= (\alpha_1^3 + 2\alpha_1\alpha_2) Y_t + (\alpha_1^2\alpha_2 + \alpha_2^2) Y_{t-1} \\
\Rightarrow \mathbb{E}(Y_{t+4}|\Omega_t) &= \mathbb{E}(\alpha_1 Y_{t+3} + \alpha_2 Y_{t+2} + \varepsilon_{t+4}|\Omega_t) \\
&= \alpha_1 \mathbb{E}(Y_{t+3}|\Omega_t) + \alpha_2 \mathbb{E}(Y_{t+2}|\Omega_t) + \mathbb{E}(Y_{t+3}|\Omega_t) \\
&= \alpha_1[(\alpha_1^3 + 2\alpha_1\alpha_2) Y_t + (\alpha_1^2\alpha_2 + \alpha_2^2) Y_{t-1}] + \alpha_2[(\alpha_1^2 + \alpha_2) Y_t + \alpha_1\alpha_2 Y_{t-1}] \\
&= (a_1^4 + 2a_1^2 a_2) Y_t + (a_1^3 a_2 + a_1 a_2^2) Y_{t-1} + (a_1^2 + a_2^2) Y_t + a_1 a_2^2 Y_{t-1} \\
&= (a_1^4 + 2a_1^2 a_2 + a_1^2 + a_2^2) Y_t + (a_1^3 a_2 + a_1 a_2^2 + a_1 a_2^2) Y_{t-1} \\
&= [a_1^2(a_1^2 + 2a_2 + 1) + a_2^2] Y_t + a_1 a_2(a_1^2 + 2a_2) Y_{t-1}
\end{aligned}
$$

Figure 11.4.1 *Forecasted result given by SAS*

Forecasts for variable c

Obs	Forecast	Std Error	95% Confidence Limits	
501	-0.6461	0.9749	-2.5569	1.2647
502	-0.3944	1.1386	-2.6260	1.8372
503	-0.2426	1.1926	-2.5801	2.0949
504	-0.1510	1.2117	-2.5258	2.2238
505	-0.0958	1.2185	-2.4840	2.2925
506	-0.0624	1.2210	-2.4556	2.3307
507	-0.0423	1.2219	-2.4372	2.3526
508	-0.0302	1.2223	-2.4257	2.3654
509	-0.0228	1.2224	-2.4187	2.3730
510	-0.0184	1.2224	-2.4143	2.3775
511	-0.0158	1.2224	-2.4117	2.3802
512	-0.0142	1.2224	-2.4101	2.3818

5 Stationarity of AR Process

5.1 AR(1) Process

For a time-series model

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t$$

This AR(1) process also can be denoted as

$$(1 + \alpha_1 L) Y_t = \varepsilon_t$$

If the root of the polynomial $1 - az = 0$ line outside the unit circle, then the AR(1) process is stationary. The root of $1 - az = 0$ is $z = 1/a$ and hence, the stationarity condition is $|z| = |1/a| > 1$, i.e. $|a| < 1$

5.2 AR(2) Process

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t$$

$$(1 - \alpha_1 L - \alpha_2 L^2) Y_t = \varepsilon_t$$

The root of above polynomial is equal to

$$1 - a_1 z - a_2 z^2 = 0$$

This should lie outside the unit circle. Remind that

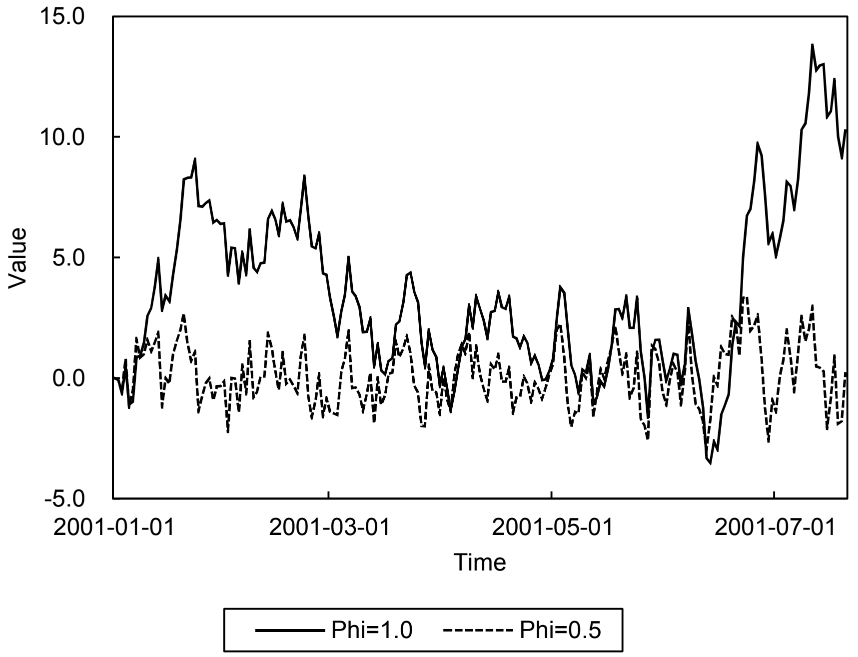
$$ax^2 + bx + c = 0$$

$$\Rightarrow x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Those two roots have to lie the outside of the unit circle.

Similarly, AR(3) process has three roots. Why stationarity is important? Look at the charts below.

Figure 11.5.1 *Comparison of the Series with Different Parameters*



When $\alpha_1 = 1$, AR(1) process can be denoted as

$$Y_t = Y_{t-1} + \varepsilon_t$$

$$\Rightarrow Y_t - Y_{t-1} = \varepsilon_t$$

For this model, it is impossible to forecast the future movement and hence this process can be defined as Random-walk process.[12]

[12] Roughly, this process also can be called as Wiener process or Brownian motion. However, in a serious manner they are different with each other by definition.

Figure 11.5.2 *Autocorrelation plot of Random-walk process* $\alpha_1 = 1$

```
                                   Autocorrelations

Lag   Covariance   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1    Std Error

  0      150.894       1.00000    |                    |********************|          0
  1      149.450       0.99043    |                  . |********************|   0.044721
  2      148.046       0.98113    |                 .  |********************|   0.076966
  3      146.466       0.97065    |                .   |******************* |   0.098865
  4      144.855       0.95998    |               .    |******************* |   0.116374
  5      143.172       0.94883    |               .    |******************* |   0.131260
  6      141.497       0.93773    |              .     |******************* |   0.144327
  7      139.695       0.92578    |              .     |******************* |   0.156037
  8      137.908       0.91394    |             .      |******************  |   0.166661
  9      136.045       0.90159    |             .      |******************  |   0.176400
 10      134.262       0.88978    |             .      |******************  |   0.185388
 11      132.438       0.87769    |            .       |******************  |   0.193740
 12      130.646       0.86581    |            .       |*****************   |   0.201536

                            "." marks two standard errors
```

Figure 11.5.3 *Autocorrelation plot of AR(1) process with* $\alpha_1 = 0.5$

```
                                   Autocorrelations

Lag   Covariance   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1    Std Error

  0     1.586986       1.00000    |                    |********************|          0
  1     0.889252       0.56034    |                  . |***********         |   0.044721
  2     0.566849       0.35719    |                  . |*******             |   0.057061
  3     0.310259       0.19550    |                  . |****                |   0.061370
  4     0.190669       0.12015    |                 .  |**.                 |   0.062603
  5     0.171379       0.10799    |                 .  |**.                 |   0.063062
  6     0.161675       0.10188    |                 .  |**.                 |   0.063431
  7     0.117692       0.07416    |                 .  |* .                 |   0.063758
  8     0.096335       0.06070    |                 .  |* .                 |   0.063930
  9     0.064268       0.04050    |                 .  |* .                 |   0.064045
 10     0.088873       0.05600    |                 .  |* .                 |   0.064096
 11     0.101145       0.06373    |                 .  |* .                 |   0.064194
 12     0.071970       0.04535    |                 .  |* .                 |   0.064321

                            "." marks two standard errors
```

Figure 11.5.2 is garbage. The plot does not give any kind of information. On the other hand, if there is no unit-root, the shape of autocorrelation plot will be changed as Figure 11.5.3; compare the geometric aspects of those charts.

Prof. Kiseok Lee (econklee@khu.ac.kr)
TA Junyong Kim (mindhunter@khu.ac.kr)

LECTURE NOTE 12

1 Autocorrelation Test in Error Term

1.1 Durbin-Watson's d Test

Assumption (Durbin-Watson's d Test) *For any linear regression model to test*

(1) The model has an intercept.

(2) The independent variables are non-random constants.

(3) Conditional mean model does not have lagged dependent variable terms.

(4) The alternative is an AR(1) Model.

Example Suppose the model below

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + u_t \\ u_t &= \alpha u_{t-1} + \varepsilon_t \end{aligned}$$

Then,

$$H_0 : \alpha = 0$$

Hence, we can estimate below

$$\begin{aligned} Y_t &= b_0 + b_1 X_t + e_t \\ e_t &= \phi e_{t-1} + \epsilon_t \end{aligned}$$

Hereby, Durbin-Watson's d statistic can be rewritten with OLS error e_t as

$$d = \sum_{t=2}^{n} (e_t - e_{t-1})^2 \Bigg/ \sum_{t=1}^{n} e_t^2$$

$$= \sum_{t=2}^{n}(e_t^2 - 2e_te_{t-1} + e_{t-1}^2) \Big/ \sum_{t=1}^{n} e_t^2$$

$$= \left(\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 - 2\sum_{t=2}^{n} e_te_{t-1}\right) \Big/ \sum_{t=1}^{n} e_t^2$$

$$\approx 1 + 1 - 2\left(\sum_{t=2}^{n} e_te_{t-1}\right) \Big/ \sum_{t=1}^{n} e_t^2$$

$$= 2\left[1 - \frac{\widehat{cov}(e_t, e_{t-1})}{\widehat{var}(e_t)}\right]$$

Since $\widehat{var}(e_t) = \widehat{var}(e_{t-1})$

$$d \approx 2\left[1 - \frac{\widehat{cov}(e_t, e_{t-1})}{\sqrt{\widehat{var}(e_t)}\sqrt{\widehat{var}(e_{t-1})}}\right]$$

$$= 2[1 - \widehat{corr}(e_t, e_{t-1})]$$

$$= 2(1 - \hat{\rho}_1)$$

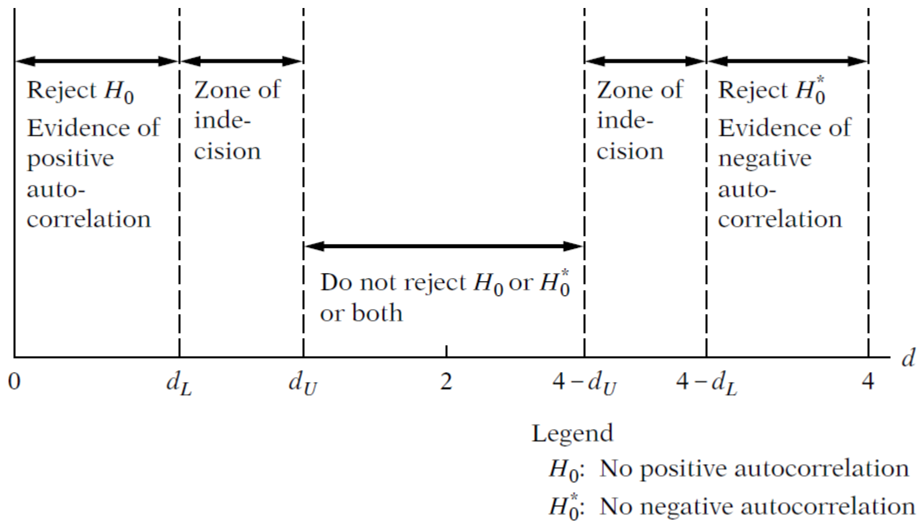
Assumption means that if there is violation in assumption, this implies that it is impossible to use above Durbin-Watson test to check the characteristic of the variable. For example,

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 Y_{t-1} + u_t$$

Since there exist Y_{t-1} as an explanatory variable in above regression model, Durbin-Watson test to check the first order autocorrelation in residual u_t cannot be applicable because this model violates above assumption (3)

Durbin-Watson's d should be judged based on below intervals.

Figure 12.1.1 *Durbin-Watson d statistic*[13]



$0 \sim dL$	: Reject the null in favor of $\alpha > 0$
$dL \sim dU$	: Not sure (This is indeterminancy.)
$dU \sim 2$	: Cannot reject the null
$2 \sim 4 - dU$	: Cannot reject the null
$4 - dU \sim 4 - dL$	: Not sure (indeterminancy)
$4 - dL \sim 4$	: Reject the null in favor of $\alpha > 0$

The reason that the Durbin-Watson test has a complicated (unusual) test procedure is because the test statistic is composed of the OLS residuals where characteristics are affected by the nature of independent variables.

1.2 Durbin's h Test

Procedure

With the model

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 Y_{t-1} + u_t$$

[13] Damodar N. Gujarati, 2004, *Basic Econometrics* (Basic Econometrics Figure 8.2.1 revisitation)

$$u_t = \phi u_{t-1} + \varepsilon_t$$

Follow below step as

(1) Run the regression and obtain residuals $\hat{u}_t = e_t$

(2) Run the following regression as

$$\hat{u}_t = \gamma_0 + \gamma_1 X_{1t} + \gamma_2 X_{2t} + \gamma_3 Y_{t-1} + \nu_t$$

(3) Test the hypothesis $\gamma_3 = 0$ in ordinary manner.

2 Dickey-Fuller Stationarity Test

Remind three form following that

$$\begin{aligned} Y_t &= \rho Y_{t-1} + \varepsilon_t \Rightarrow \hat{\rho} \\ Y_t &= \mu + \rho Y_{t-1} + \varepsilon_t \Rightarrow \hat{\rho}_\mu \\ Y_t &= \mu + \beta t + \rho Y_{t-1} + \varepsilon_t \Rightarrow \hat{\rho}_\tau \\ H_0 &: \rho = 1 \\ \tau &= n(\hat{\rho} - 1) \end{aligned}$$

Under the null, $\rho = 1$, the OLS estimator does not have asymptotic normal distribution. Instead, the asymptotic distribution is a function of Brownian motion process. Thus, the use of t-test is invalid. If statistic τ is larger than critical value, then H_0 cannot be rejected; otherwise, H_0 can be rejected.

Property (Non-stationary process)

(1) Does not have conventional variance, covariance, and correlation.

(2) Permanent shock effect; i.e. the effect of a shock will not diminish as $t \to \infty$.

(3) The probability for a non-stationary process to come back to where it started is 0.

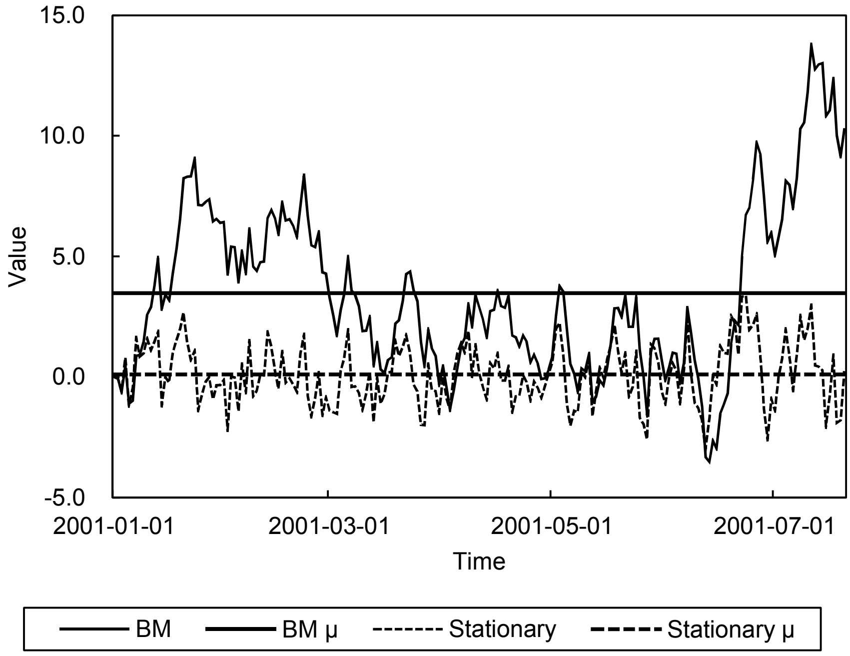
(4) The process is very persistent.

Question *Why should we care about the non-stationary of economic time-series?*

Answer Refer Nelson and Plosser (1982)[14]

[14] C. R. Nelson and C. I. Plosser, 1982, "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", *The Journal of Monetary Economics 10 (2)*, 139-162

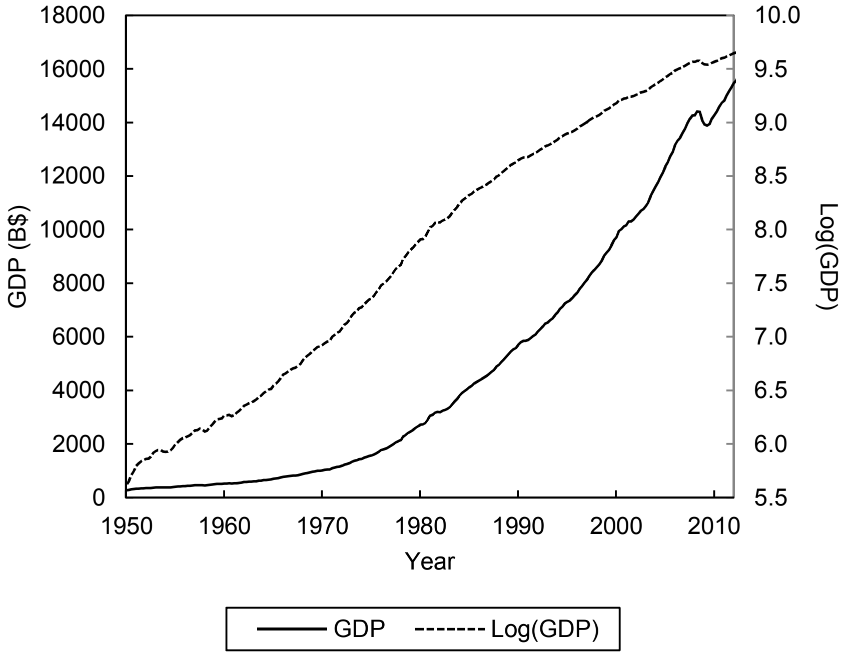
Figure 12.2.1 *Brownian Motion vs. Stationary Series; Convergence vs. Divergence*



From above figures, it is possible to capture the major characteristic of the stationary and non-stationary time-series. For example, for non-stationary time-series like above left figure, there is only rare chance (converges to 0) of returning back to the point of origin, which is expressed as a straight red line on the graph. For stationary process like right one, on the other hand, it can intuitively be captured that the probability of returning back to the point of origin converges to 1.

Following examples will guide you how to treat the random-walking time-series data normally; the GDP data from 1970s to 2010s of the United States is applied.

Figure 12.2.2 *GDP and Log-transformed GDP Series*



Through above left chart, the time-series exponentially growing can be found. Since this time-series data follows Brownian motion, its variance is time-dependent and goes to infinity as time approaches infinity. Thereby, the normal approximation is no longer applicable, even asymptotically.

For data analysis in quantitative manners, its movement should be determined based on relative measure, instead of absolute measure. This implies that the geometric meaning of data, which can be interpreted with multiplication and percentage form, can give us more desirable implication than its arithmetical interpretation, which can be explained with addition and its own level.

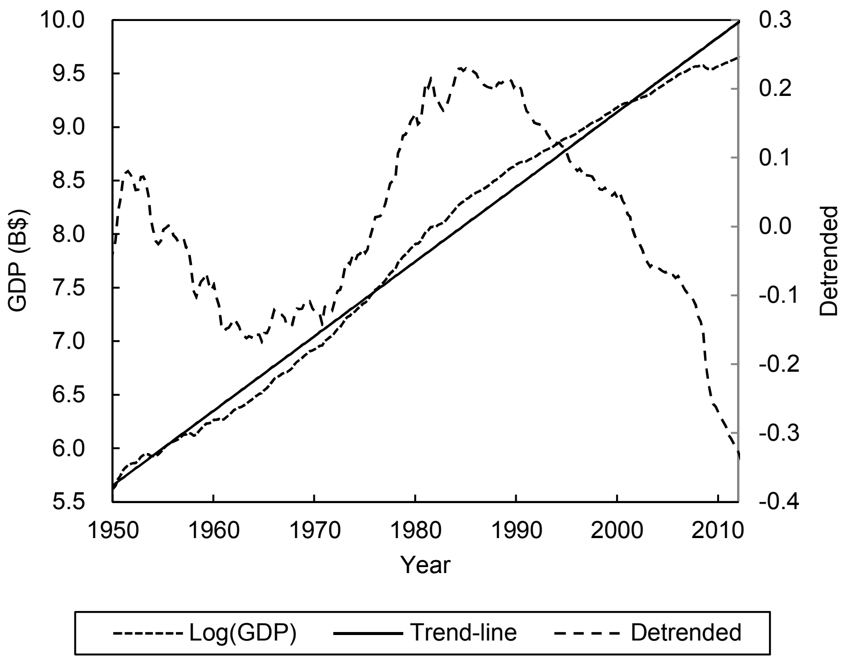
Thus, data transformation with logarithm should be considered first. Compare intuitively three mathematical expressions below that

$$
\begin{aligned}
X_t &= X_0 + x_1 + x_2 + \cdots \\
Y_t &= Y_0(1 + y_1)(1 + y_2) \times \cdots \\
Z_t &= Z_0 e^{z_1} e^{z_2} \cdots \\
&= Z_0 e^{z_1 + z_2 + \cdots}
\end{aligned}
$$

$$\Rightarrow \ln Z_t \quad = \quad \ln Z_0 + z_1 + z_2 + \cdots$$

Therefore, physical movement of level-formed variable can be easily analyzed through log-transformation. The expression of Z_t is the continuous form of Y_t. Refer the Napier's constant also.

Figure 12.2.3 *Elimination of the Trend of the Series*



Its shape changed, but still it is random-walking and also timely dependent. But the manner of time trend elimination expressed on the above charts is not valid. Mean reverting still cannot be found since log-transformed GDP also follows Brownian motion. Thus, the trend elimination through the way of above is till obsolete.

Thereby, difference of log-transformed data should be determined before the process of elimination of time trends. Above left chart shows the movement of differenced logged GDP of the United States and it seems that it has a short-term mean revering with a long-term time trends, which should be eliminated. Thereafter, above right is the result of above processes; there exists mean-reverting characteristic without time trend.

Note that if there is log-transformed variable which is differenced, it will be possibly called as continuous compounding rate of interest in a field of Finance, and also be called as the force of interest in a field of Actuary.

Remind those limit applications below that

Figure 12.2.4 *Log-differencing and Trend-Effect Elimination*

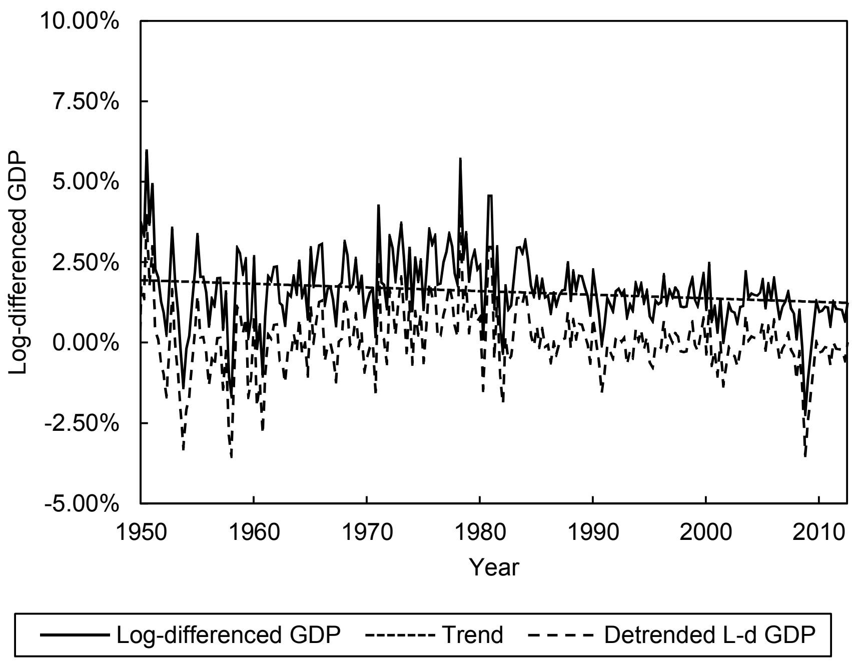


Table 12.2.1 *Heterogeneous Mathematical return expressions*

Simple	Compounding	Continuous compounding
$S_t(1+R)$	$S_t\left(1+\frac{R}{n}\right)^n$	$\lim_{n\to\infty} S_t\left(1+\frac{R}{n}\right)^n = S_t e^{Rn}$

Through the constant e, return can be considered within exponential manners; continuous compounding rate of return. Therefore, with continuous compounded return r and time interval $(T-t)$, it is possible to depict the state with below expression.

$$S_T \quad = \quad S_t e^{r(T-t)}$$

Thus, if $(T-t)$ is substituted by τ, then

$$\begin{aligned}
\frac{dS_T}{d\tau} &= rS_t e^{r\tau} \\
\rightarrow \frac{dS_T/d\tau}{S_T} &= \frac{rS_t e^{r\tau}}{S_t e^{r\tau}} \\
&= r
\end{aligned}$$

Therefore, continuous compound return within time interval τ can be measured as

$$\begin{aligned}
r &= \ln S_T - \ln S_t \\
&= \ln(S_T/S_t)
\end{aligned}$$

LECTURE NOTE 13

1 ARCH Model

1.1 Motivation

ARCH (Auto-regressive Conditional Heteroskedasticity) approaching was proposed by Engle (1982)[15] firstly. Motivation of ARCH model is

(1) Econometric forecasters have found that their ability to predict future varies from one period to another.

(2) The inherent uncertainty or randomness associated with different forecast period seems to vary widely over time; large and small errors cluster together.[16]

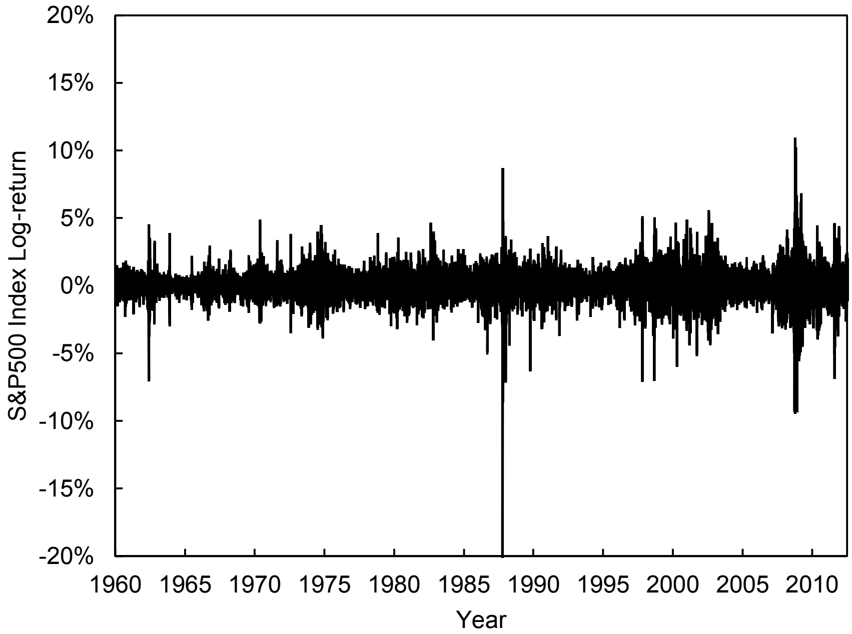
(3) Analysis suggests the underlying forecast variance may change over time and is predicted by past forecast errors.

(4) The existence of an ARCH effect can be interpreted as evidence of misspecification, either by omitted variables or through structural change.

[15] R. F. Engle, 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica 50 (4)*, 987-1008

[16] S. K. McNees, 1979, "The Forecasting Record for the 1970's", *New England Economic Review* September/October, 33-53

Figure 13.1.1 *Volatility Clustering Phenomenon; S&P500 Index Log-return*



1.2 ARCH(1) Model

For instance, ARCH(1) Model can be proposed as

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + u_t \\ \mathbb{E}(u_t | \Omega_{t-1}) &\sim \mathcal{N}(0, h_t) \\ h_t &= \mathbb{E}(u_t^2 | \Omega_{t-1}) \\ &= \alpha_0 + \alpha_1 u_{t-1}^2 \end{aligned}$$

Unconditionally, the distribution of u_t is not normal. Instead, the distribution has fatter tails than the normal.

Because of the presence of the conditional heteroskedasticity, the model is non-linear. Let ℓ_t be the log-likelihood for the tth observation and T the sample size. Then,

$$\ell_t = -\frac{1}{2} \times \ln 2\pi - \frac{1}{2} \times \ln h_t - \frac{1}{2} \times \frac{u_t^2}{h_t}$$

Figure 13.1.2 *Empirical Distribution vs. Normal Distribution; Leptokurtic Distribution?*

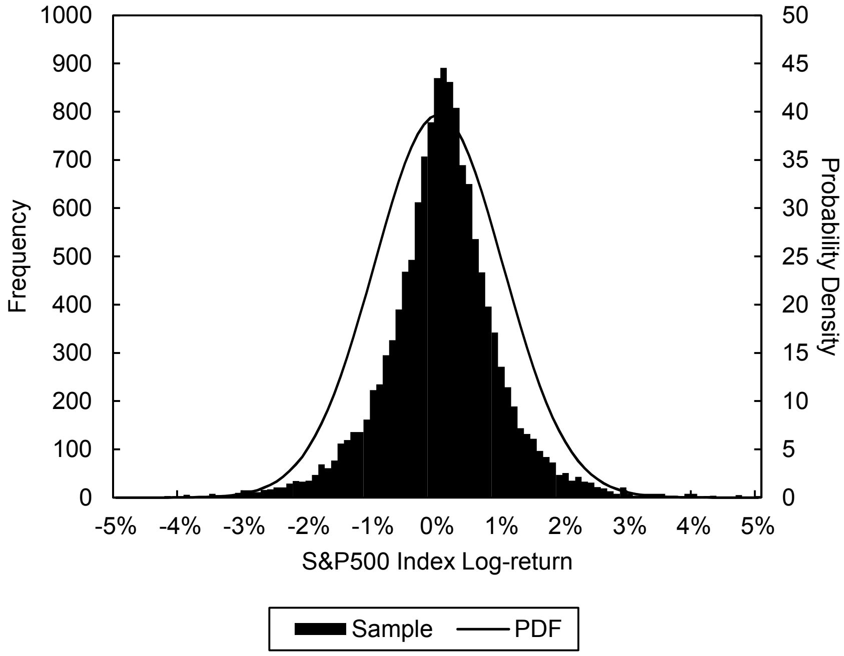
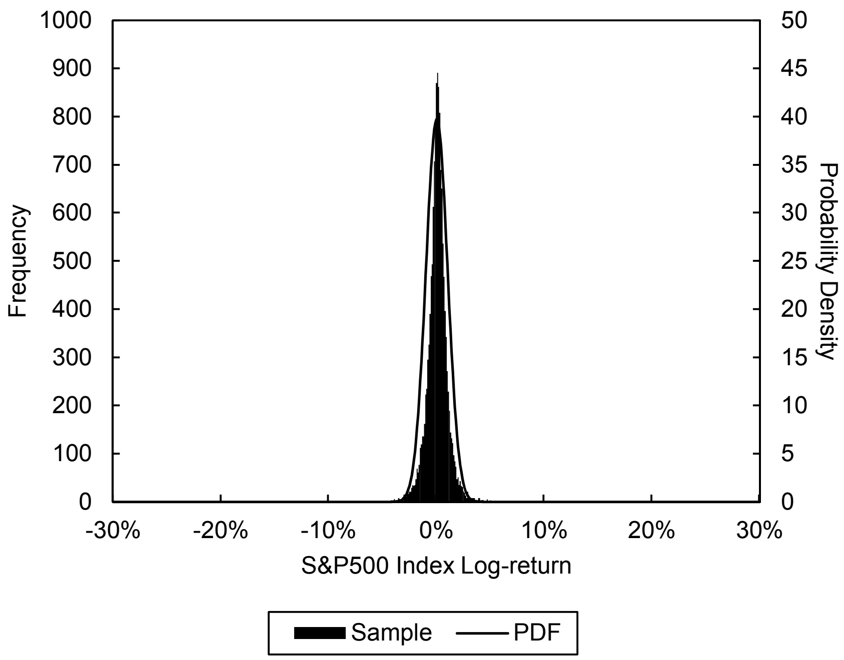


Figure 13.1.3 *Empirical Distribution vs. Normal Distribution; Fat-tailed Distribution?*



The efficiency can be gained in estimating regression model parameters when the ARCH model is applied. The absolute t values of the estimate of β_0 and β_1 will be increased when the ARCH model is used. And h_t represents the scientific uncertainty measure of Y_t; how volatile the market uncertainty is?

Then, apply the Auto-regressive Moving Average scheme into this

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + u_t \\ \mathbb{E}(u_t|\Omega_{t-1}) &\sim \mathcal{N}(0, h_t) \\ h_t &= \mathbb{E}(u_t^2|\Omega_{t-1}) \\ &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \end{aligned}$$

This model is Generalized Auto-regressive Conditional Heteroskedasticity, which was proposed by Bollerslev (1986)[17] as the extension of the ARCH scheme and more widely accepted than it because of its flexible lag structure. Since GARCH, heterogeneous models of conditional heteroskedasticity have been developed.

1.3 ARCH Test in Error Term

Construct following model

$$\begin{aligned} e_t^2 &= \gamma_0 + \sum_{l=1}^{p} \gamma_l e_{t-l}^2 + \eta_t \\ &= \gamma_0 + \gamma_1 e_{t-1}^2 + \gamma_2 e_{t-2}^2 + \cdots + \gamma_p e_{t-p}^2 + \eta_t \end{aligned}$$

And test the null ($H_0: \gamma_1 = \gamma_2 = \cdots = \gamma_p = 0$) through hypothesis testing or

$$\begin{aligned} \tau &= nR^2 \\ &\sim \chi^2(p) \end{aligned}$$

Think intuitively. τ is similar to LM, which is proposed above; if there is no conclusive determination effect, τ will go down because its R^2 is very low and hence H_0 will not be rejected easily since the constant $n \in \mathbb{N}$. Vice versa, H_0 will be rejected with high level of τ and R^2.

[17] T. Bollerslev, 1986, "Generalized Autoregressive Conditional Heteroskedasticity", *The Journal of Econometrics 31 (3)*, 307-327

2 GARCH Family

2.1 GARCH Model

With regarding to the lag p and q, GARCH model is denoted as

$$\begin{aligned} h_t &= \mathbb{E}(\sigma_t^2) \\ &= f\big(u_{t-1}, u_{t-2}, u_{t-3}, \cdots, u_{t-q}, h_{t-1}, h_{t-2}, h_{t-3}, \cdots, h_{t-p}\big) \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \gamma_i h_{t-p} \end{aligned}$$

Specially, GARCH(1,1) model is

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 h_{t-1}$$

Without conditioning

$$\begin{aligned} \hat{\sigma}^2 &= \alpha_0 + \alpha_1 \hat{\sigma}^2 + \gamma \hat{\sigma}^2 \\ \Rightarrow \hat{\sigma}^2 &= \frac{\alpha_0}{1 - \alpha_1 - \gamma_1} \end{aligned}$$

Hence $\alpha_1 + \gamma_1 < 1$ and since $\alpha_1 + \gamma_1 = 1$, it is invalid because of zero denominator.

Where α_0, α_i, and γ_i should be all positive.

$\alpha_0 = [0, \infty)$, $\alpha_i = [0, \infty)$, $\gamma_i = [0, \infty)$, $\forall i \in \mathbb{N}$

The most widely adopted ARCH class model is GARCH(1,1). For a GARCH(1,1) process, the unconditional variance of the error term is

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \gamma_1}$$

Remind that

$$\begin{aligned} h_t &= \hat{\sigma}^2 \\ &= \mathbb{E}(u_t^2) \\ &= \widehat{var}(u_t) \end{aligned}$$

The condition $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \gamma_i < 1$ implies the GARCH process is stationary since the mean, variance, and autocovariance are finite and constant over time.

2.2 EGARCH Model

The GARCH model imposes the non-negativity constraints on the parameters α_i and γ_i. This non-negativity constraint can be considered too restrictive. For EGARCH firstly argued by Nelson (1991)[18], the restriction is not necessary. Also in the EGARCH model, the conditional variance, h_t is an asymmetric function of lagged residuals u_{t-1}.

$$\begin{aligned} \ln h_t &= \ln \hat{\sigma}^2 \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i g(z_{t-i}) + \sum_{i=1}^{p} \gamma_i \ln h_{t-i} \end{aligned}$$

Where

$$\begin{aligned} g(z_t) &= \theta z_t + \lambda[|z_t| - \mathbb{E}(|z_t|)] \\ z_t &= u_t/\sqrt{h_t} \end{aligned}$$

Often, $\gamma = 1$ is assumed and note that if $z_t \sim \mathcal{N}(0,1^2)$ then

$$\mathbb{E}(|z_t|) = (2\pi)^{-1/2}$$

The function $g(z_t)$ is linear in z_t with slope coefficient $(\theta + 1)$ if z_t is positive while $g(z_t)$ is linear in z_t with coefficient $(\theta - 1)$ if z_t is negative.

2.3 Square Root GARCH Model

Square root GARCH(1,1) model is denoted as

$$\hat{\sigma}_t = \alpha_0 + \alpha_1 |u_{t-1}| + \gamma_1 \hat{\sigma}_{t-1}$$

Where

$$\hat{\sigma}_t = \sqrt{h_t}$$

[18] D. B. Nelson, 1991, "Conditional Heteroskedasticity in Asset Returns: a New Approach", *Econometrica 59 (2)*, 347-370

Figure 13.2.1 *Log-return of S&P500 Index and GARCH Volatility*

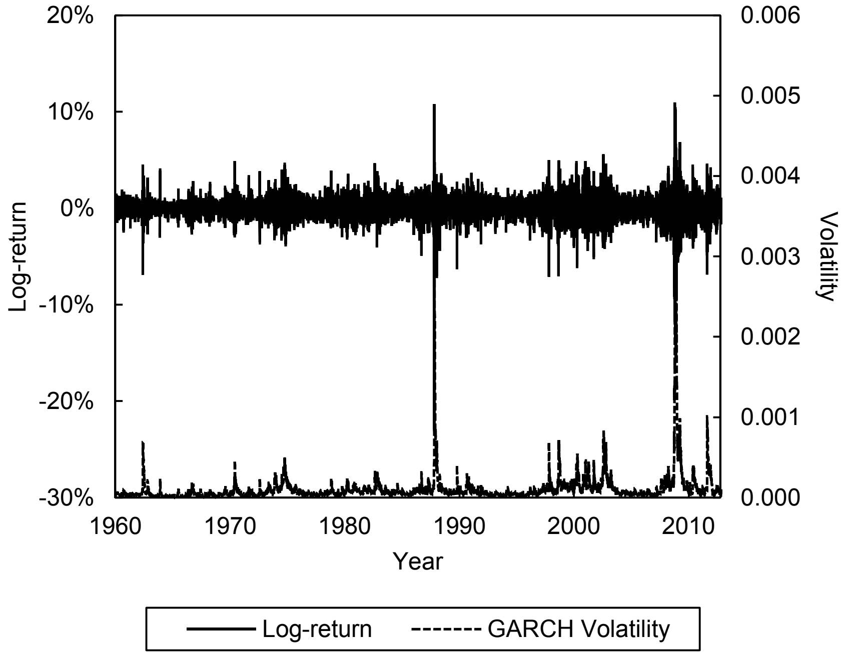
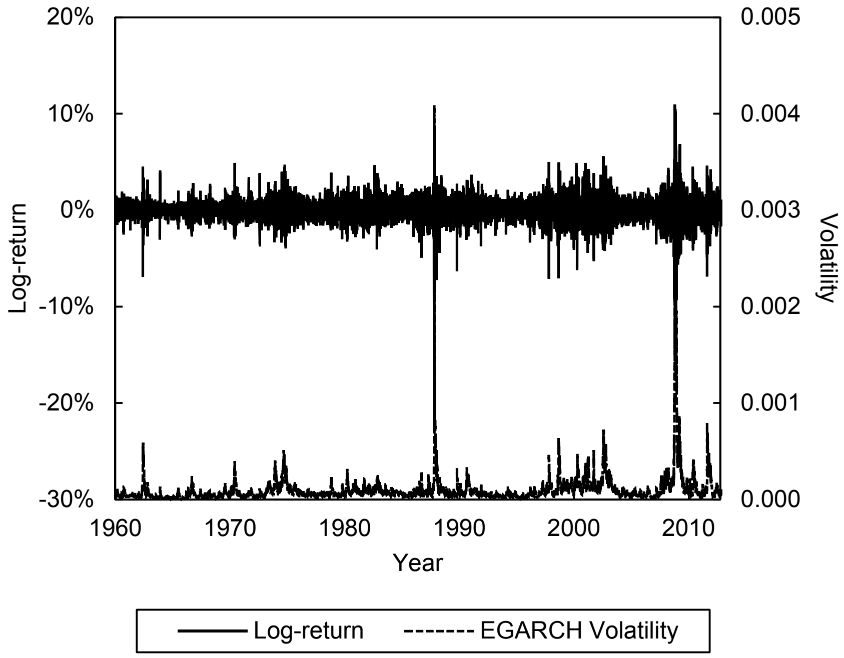


Figure 13.2.2 *Same Series with Different Approach; EGARCH Volatility*



2.4 Quadratic GARCH Model

Quadratic GARCH(1,1) model is denoted as

$$h_t = \alpha_0 + \alpha_1(u_{t-1} - \phi)^2 - \gamma_1 h_{t-1}$$

If $\phi > 0$ then a negative u_{t-1} makes the conditional variance h_t proportionally larger than a positive u_{t-1}; this scheme is similar to EGARCH model but this is more simple.

2.5 Power GARCH Model

Power GARCH(1,1) model is denoted as

$$h_t^{\phi} = \alpha_0 + \alpha_1|u_{t-1}|^{\phi} + \gamma_1 h_{t-1}^{\phi}$$

This process is very general that nests many ARCH type models.

2.6 GARCH-in-Mean Model

This model is also called as GARCH-M model and it reflects its conditional variance to its conditional mean recursively.

In a general manner, GARCH-in-Mean with GARCH(1,1) model is denoted as

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 h_t + u_t$$

$$u_t \sim \mathcal{N}(0, h_t)$$

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 h_{t-1}$$

For the conditional mean, following ideas can be applied also as

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 \ln h_t + u_t$$

Or

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 \sqrt{h_t} + u_t$$

Analysis of Financial Time-series

$\boxed{\text{Now Updating, Coming Soon}}$

$\boxed{\text{Now Updating, Coming Soon}}$

$\boxed{\text{Now Updating, Coming Soon}}$

$\boxed{\text{Now Updating, Coming Soon}}$

$\boxed{\text{Now Updating, Coming Soon}}$

BIBLIOGRAPHY

Domestic

[1] 김동일, 2011, 『SAS를 이용한 계량경제분석』, Philosophy & Art

[2] 김명직 · 장국현, 2011, 『금융시계열분석』, 제2판, 경문사

[3] 김창진 · 황윤재, 2010, 『Econometrics (II)』, 고려대학교 출판부

[4] 이종원, 2007, 『계량경제학』, 박영사

[5] 조인호, 2004, 『SAS 강좌와 통계컨설팅』, 영진닷컴

[6] Keller, Gerald, 2007, 『켈러의 경영경제통계학; 엑셀의 실전적 활용』, 제7판, 이상규 옮김, 센게이지러닝코리아

International

[1] Bollerslev, Tim, 1986, "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 31, 307-327

[2] Chow, Gregory C., 1960, "Tests of Equality between Sets of Coefficients in Two Linear Regressions", *Econometrica* 28 (3), 591-605

[3] Durbin, James, and Geoffrey S. Watson, 1951, "Testing for Serial Correlation in Least Squares Regression", *Biometrika* 38, 159-171

[4] Enders, Walter, 2010, *Applied Econometric Time-series*, Third Edition, John Wiley & Sons Inc., New York, NY

[5] Engle, Robert F., 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica* 50, 987-1008

[6] Gujarati, Damodar N., 2004, *Basic Econometrics*, Fourth Edition, McGraw-Hill Companies, New York, NY

[7] Kalman, Rudolf E., 1961, "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering* 82 (1), 35-45

[8] Kalman, Rudolf E. and Richard S. Bucy, 1961, "New Results in Linear Filtering and Prediction Theory", *Journal of Basic Engineering* 83 (1), 95-108

[9] Lee, Kiseok, 2004, *Econometrics with Applications*, Kyung Hee University Press, Seoul, Korea

[10] Lee, Kiseok, 2004, *Economic Statistics with SAS Applications*, Kyung Hee University Press, Seoul, Korea

[11] Lee, Kiseok, 2005, *Basic Econometrics with SAS Applications*, Kyung Hee University Press, Seoul, Korea

[12] McNees, S. K. 1979, "The Forecasting Record for the 1970's", *New England Economic Review* September/October, 33-53

[13] Nelson, Charles R. and Charles R. Plosser, 1982, "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", *Journal of Monetary Economics* 10 (2), 139-162

[14] Ramanathan, Ramu, 2002, *Introductory Econometrics with Applications*, Fifth Edition, Cengage Learning, Stamford, CT

Curriculum Vitae

Address

E-mail (mindhunter@khu.ac.kr)

Homepage (http://www.junyongkim.com/)

Education

Seoul National University, Seoul, Korea

Master of Science, March 2013-February 2015 (Expected)

Major: Finance

Kyung Hee University, Seoul, Korea

Bachelor of Business Administration, March 2007-February 2013

Major: Business Administration

University of Mississippi, Oxford, MS

Exchange Student, January 2012-May 2012

Major: Banking and Finance