EE325

Introductory Econometrics

Weerawat Phattarasukkumjorn

Semester 1/2021

Course details

› Schedule

Section 2: Tue, Thu 11.00 – 12.30
Moodle class code: **1568**

› Instructor

Weerawat Phattarasukkumjorn, Room no. 437
weerawat@econ.tu.ac.th
Please communicate **in class group.**

› Evaluation

Homework and assignment	30 points
Midterm exam	30 points
Final exam	40 points

› Exam date and time

Midterm: Wed, Sep29 from 15.00 – 17.00 (2 hours)
Final: Mon, Dec 13 from 09.00 – 11.30 (2.5 hours)

What does 'Econometrics' mean?

We, first and foremost, try to understand the topic, analyzing the morphemes.

› Econometrics = economics + metrics

› Economics: you all know what this means.

› Metrics: a system of measurement.

To conclude, *"econometrics is the application of statistical methods to economic data in order to give empirical content to economic relationships. More precisely, it is the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference".*

How can econometrics be used?

First of all, let's introduce how academics work to come up with a conclusion and policy recommendation. You are also to deal with these processes for your seminar.

› Introduction and problem statement (research question)

What question do you want to answer? How and why is it important to answer the question? What is expected to be your contribution to academic world?

› Literature review (research gap)

Have there been any other people trying to answer the same question yet? What and how other people have tried to answer this question? What are their results? What are their methods used and how they are still flawed or lacking?

› Theoretical framework (rationalize your hypothesis)

For economics, what is the rationale behind your speculation or hypothesis? Has there been anyone discover or derive any theory before? What type of explanation will you be using to couple and elaborate your results and implications.

How can econometrics be used?

› Research method (how to find the answer)

How you are going to prove your hypothesis? We have several ways to do, but mostly we can use these methods

- Qualitative methods: descriptive statistics, reviews, deduction, etc.
- Quantitative methods: forecast, quantitative simulation, econometrics, etc.

› Report results and policy implication

What are your results and findings? Moreover, what do those findings suggest? Do those results imply that we should implement which kind of policy? What is the limitation of your work and what can be possibly improved in the future?

Example of econometric process

› Research question

› Economic theory or model

› Empirical data

› Econometric modeling

› Estimation and hypothesis testing

› Forecast or prediction

› Empirical conclusion

› Policy recommendation

Strengths of econometrics

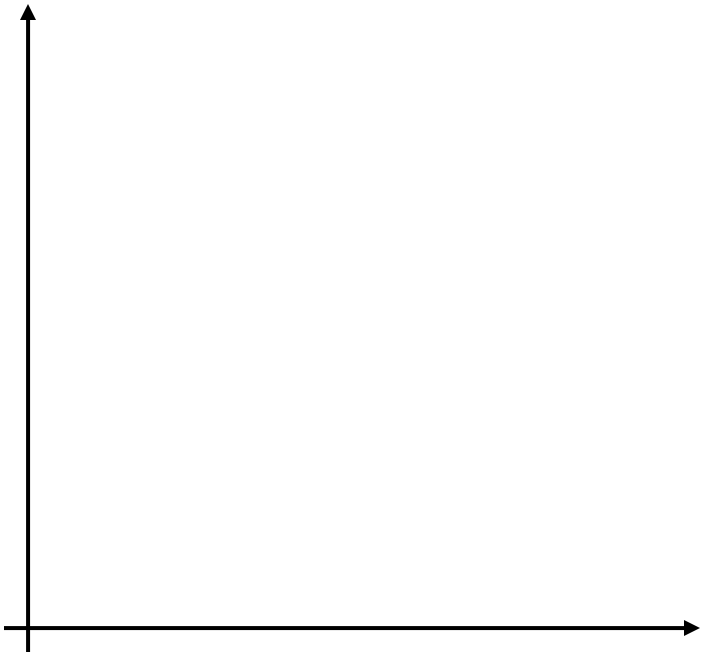
› Ability to quantify direction and magnitude of economic effects

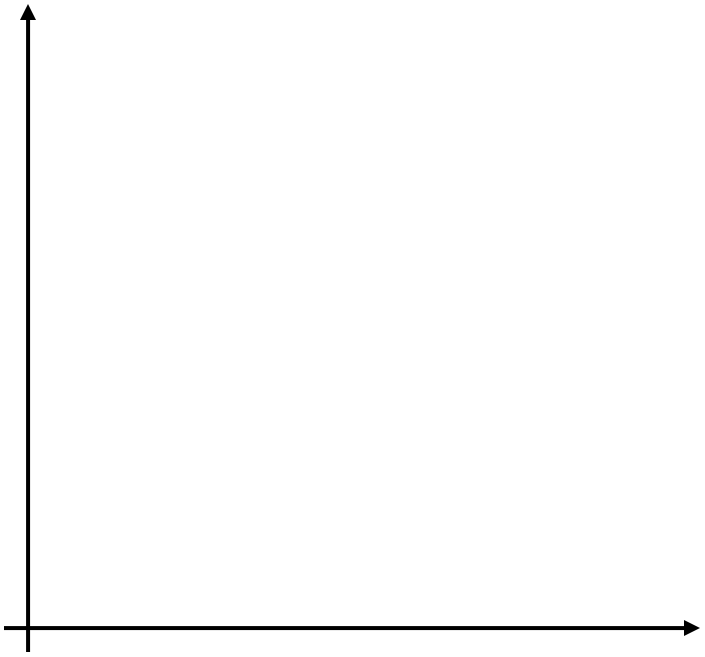
› Statistical testability

› Prediction or simulation (precise, though not always correct)

What are statistical relationships?

As a reminder, when we study mathematics, a function is usually determined. Meanwhile, studying statistics relationship is, most of the time, cannot be captured by a specific function.

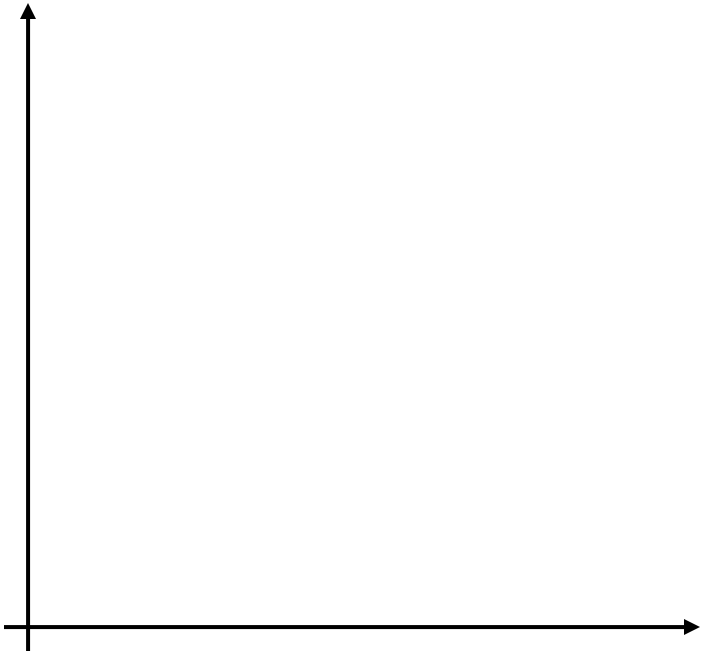


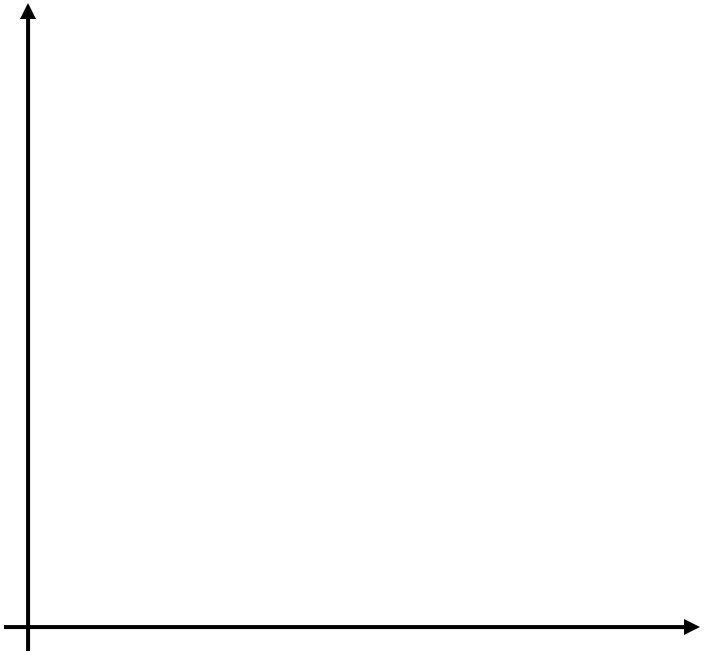


What are statistical relationships?

(1) Correlation relationship is any statistical association or the degree of association between two or more variables (pairwise, linear).

(2) Regression relationship is a relationship derived from a set of statistical processes for estimating the relationships between a dependent and one or more independent variables (linear or nonlinear).





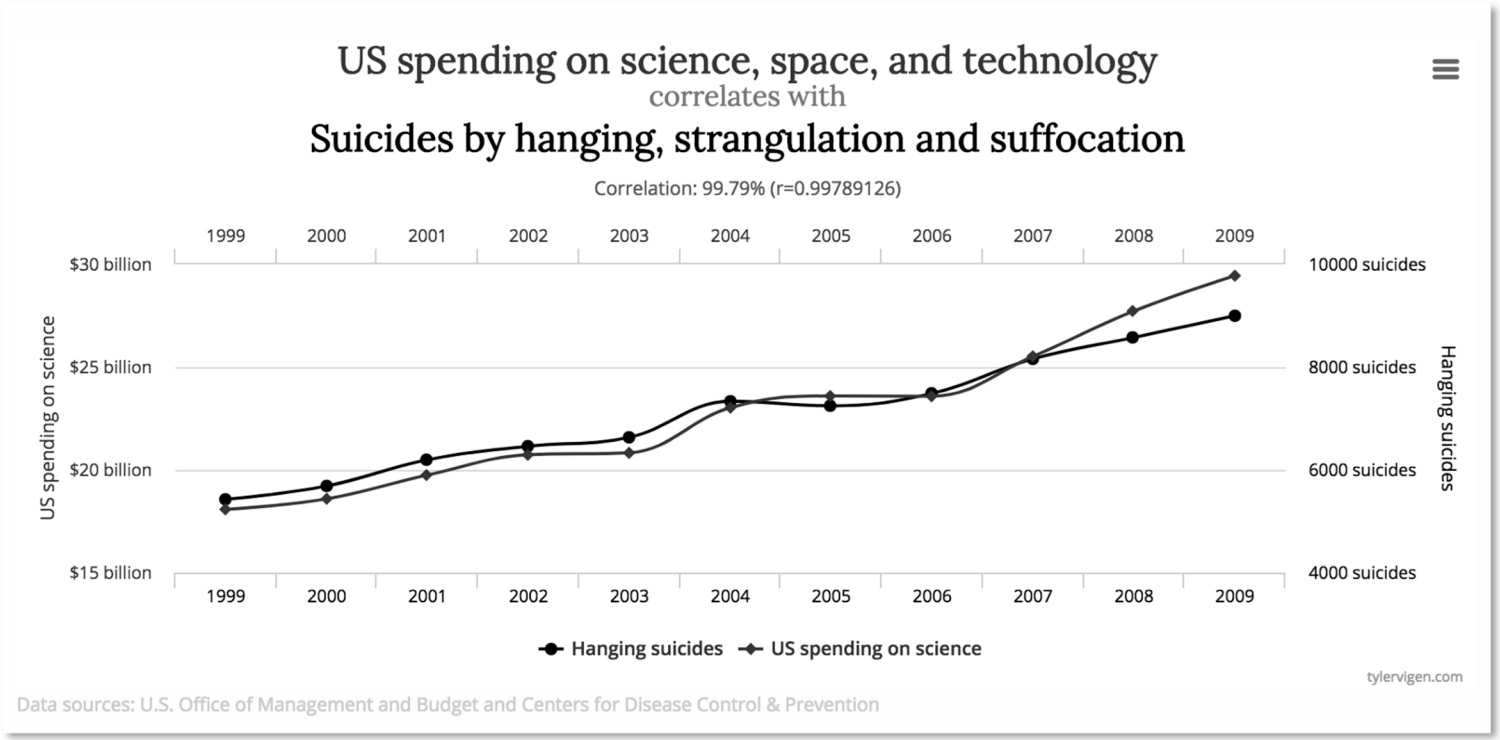
What are statistical relationships?

(3) Causal relationship is influence by which one event, process or state (a cause) contributes to the production of another event, process or state (an effect) where the cause is partly responsible for the effect, and the effect is partly dependent on the cause.

It is very difficult to draw a conclusion that two variables are causal, especially for social science. **Neither** correlation **nor** regression imply causality at all.

Most causality can be base upon economic theory and explanation. Aligning empirical results can also support for theory. Couple both can provide strong evidence, though does not reveal universal truth.

Spurious correlation



Data by collection

(1) Primary data
Collected by a researcher from first-hand sources, using methods like surveys, interviews, or experiments.

(2) Secondary data
Gathered from studies, surveys, or experiments that have been run by other people or for other researches.

Broad categorization

observations

student	GPA
student 1	GPA 1
student 2	GPA 2
student 3	GPA 3
student 4	GPA 4
student 5	GPA 5
student 6	GPA 6

(1) Cross-sectional data
A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at the one point or a period of time. The analysis has no regard to differences in time.

observations

time	GPA
t=1	GPA 1 (t=1)
t=2	GPA 1 (t=2)
t=3	GPA 1 (t=3)
t=4	GPA 1 (t=4)
t=5	GPA 1 (t=5)
t=6	GPA 1 (t=6)

(2) Time series
A series of data points indexed in time order.

Broad categorization

(3) Panel data

A set of data collected over time and over the same individuals.

student	GPA	time
student 1	GPA 1	t=1
student 2	GPA 2	t=1
student 3	GPA 3	t=1
student 1	GPA 1	t=2
student 2	GPA 2	t=2
student 3	GPA 3	t=2
student 1	GPA 1	t=3
student 2	GPA 2	t=3
student 3	GPA 3	t=3

time	student 1	student 2	student 3	student 4
t=1				
t=1				
t=3				

Broad categorization

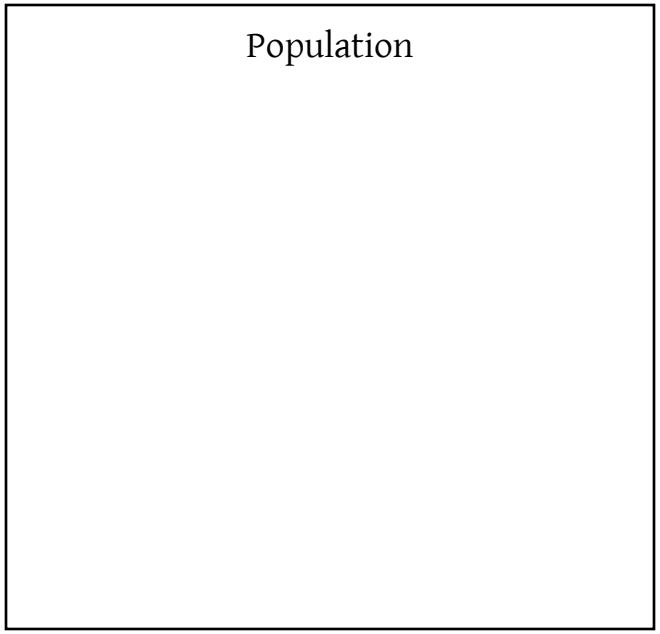
(4) Pooled cross-sectional data

A multiple cross-sectional data pooled without observing the same subject.

student	GPA	time
student 1	GPA 1	t=1
student 2	GPA 2	t=1
student 3	GPA 3	t=1
student 2	GPA 2	t=2
student 3	GPA 3	t=2
student 4	GPA 4	t=2
student 1	GPA 1	t=3
student 3	GPA 3	t=3
student 4	GPA 4	t=3

time	student 1	student 2	student 3	student 4
t=1				
t=1				
t=3				

Inferencing from sample



(1) Population
Refers to a group of observations of interest. If the data cover all the observations of interest, the set of data is called a **census**.

(2) Sample
Refers to a subset of population of interest. Most of the time they are statistically random samples which is called **survey**.

Secondary data in Thailand

(1) Cross-sectional data

Not free most of the time. University students can request from the National Statistics Office (NSO) for their project. Project proposal must be submitted.

› Household Socio-Economic Survey (SES) – income, expenditure, debt, asset (recently added), etc. Unit of analysis is household.

› Labor Force Survey (LFS) – wage, working hour, occupation, job search. Unit of analysis is individual.

› Health and Welfare Survey (HWS) – health insurance, health benefit, partial utilization, etc. Unit of analysis is individual.

› Office of the National Economic and Social Development Council (NESDC) – the NESDC takes NSO data to calculate important statistics. https://www.nesdc.go.th/more_news.php?cid=74

Secondary data in Thailand

(2) Time series data
Widely available because they are not identity specified.

› Bank of Thailand (BOT) – GDP, financial and monetary statistics, currency exchange, etc.
https://www.bot.or.th/Thai/Statistics/Pages/default.aspx

› Ministry of Finance (MOF) and Fiscal Policy Office – tax revenue and expenditure, tax disbursement, national debt, etc.
http://www.fpo.go.th/main/Statistic-Database.aspx

› Others include Ministry of Commerce (MOC), Stock Exchange of Thailand (SET), international organization such as International Monetary Fund (IMF), World Bank, United Nations (UN), OECD, etc.

Secondary data in Thailand

(3) Panel data

Very limited in Thailand. The ones that I know of are

› Panel SES (by the NSO) from 2005, 2006, 2007, 2010, 2012 – not free.

› Townsend Thai data (cooperated with UTCC and TRF) from 1997 to 2017 – request needed.
http://riped.utcc.ac.th/panel/data/townsend-thai-data

Content

› Chapter 2
Statistics revision

› Chapter 3
Simple linear regression

› Chapter 4
Multiple linear regression

› Chapter 5
Dummy variable

› Chapter 6
Relaxing some assumptions

Chapter number

Sub-chapter

How to use this handout?

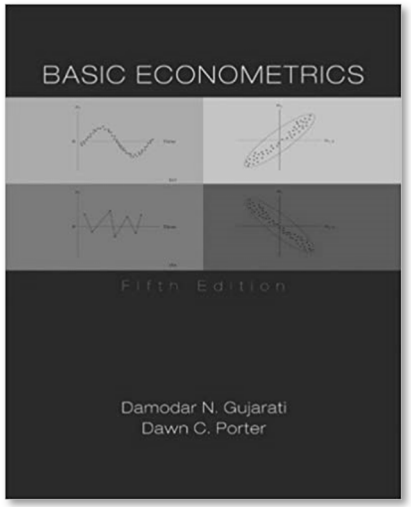
Main topic

Illustration or content

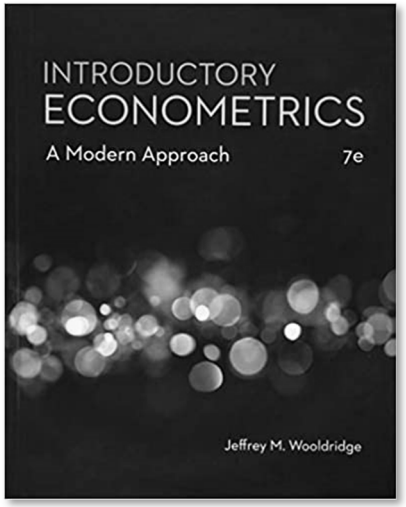
Illustration or content

Content

Main textbook



› Gujarati, D.N., and D.C. Porter, **Basic Econometrics**. 5th ed., N.Y., McGraw-Hill, 2009.



› Wooldridge, J. M. **Introductory Econometrics: A Modern Approach**. 7e ed. Thompson: South-Western, 2019.

Chapter 2

Statistics revision

Flow of study in this chapter

› Probability, random variable and density

Revision the basic concept of probability and its distribution, graphing random variable and its probability.

› Bivariate probability density

When two random variables, or two events, coexist, what aspects of the distribution can be studied.

› Central tendency and dispersion

How to measure, and why, central tendency and dispersion of a distribution for a random variable.

› Common distribution functions

The distributions we are relying on for statistical in this semester.

Further reading can be found in Gujarati and Porter (2009), Appendix A, page 801-837

(1) Event, Sample Space and Probability

Let A be an event of interest, occurring within a given sample space S and $P(A)$ be the probability that A will occur, $P(A)$ is defined as

› $P(A) = \frac{number\ of\ times\ event\ A\ will\ occur}{number\ of\ all\ possible\ outcome\ in\ sample\ space\ S}$

Example tossing 2 fair coins, the sample space is

› $S = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$

If the event of interest is having at least a coin turning head (H) is

› $A = \{\text{HH}, \text{HT}, \text{TH}\}$.

The probability of this event is then

› $P(A) =$

(1) Event, Sample Space and Probability

Probability Axioms

(1) $0 \leq P(A) \leq 1$

(2) $P(S) = 1$

(3) If and A are B mutually exclusive, then $P(A \cup B) = P(A) + P(B)$

(2) Random variable

Definition 2.1

Let X be a **random variable,** the results of an experiment in the form of value, which value is given by one of the results.

Example Tossing 2 fair coins again, let X be 0 if the result shows **at least** a coin turned up head, be 1 otherwise. The sample space was defined as

› $S = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$

Transforming theses events into random variable, we get

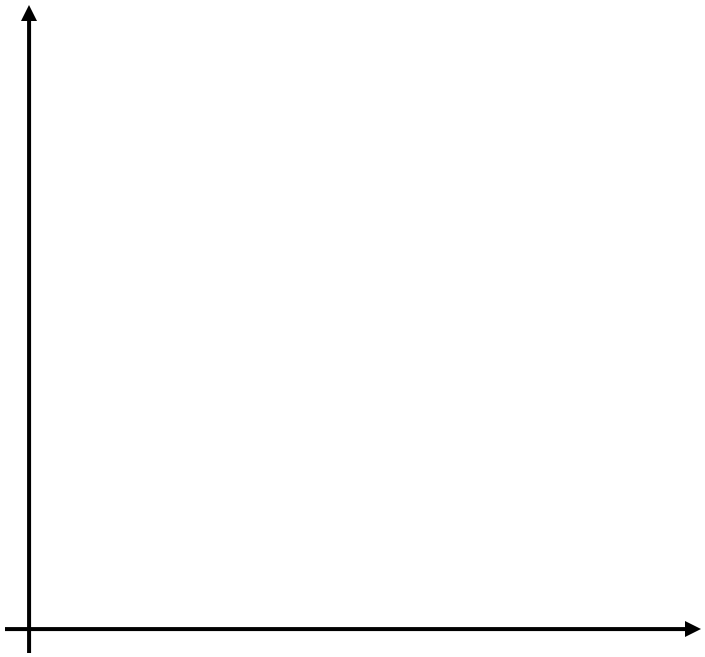
› $S_X = \{0,1\}$

Therefore, if we put probability function with specific value of random variable X, we have

› $P(X = 0) =$

› $P(X = 1) =$

(2) Random variable



Graphing the random variable and its probability here on the left.

› **Discreate random variable** is a random variable that can take specific values of event.

› **Continuous random variable** is a random variable that can takes infinite amounts of value of event.

(3) Probability Density Function (PDF)

Definition 2.2

A function whose value at any given sample (or point) in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample.

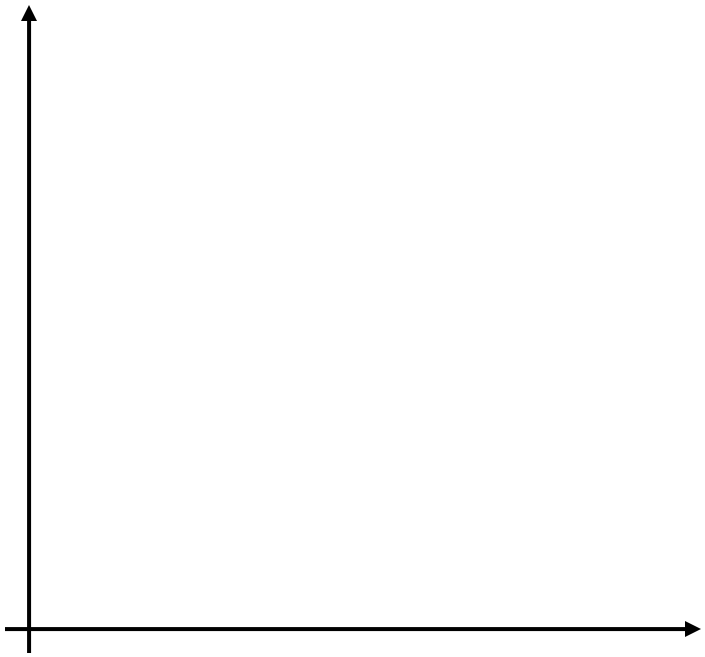
› $f(x_i) = P(X = x_i)$ for $x_i \in S_X$

› $f(x_i) = 0$ for $x_i \notin S_X$

Example Let X be a random variable of total points from rolling 2 fair dices, the sample space would be

› $S_X =$

(3) Probability Density Function (PDF)



Graphing the random variable and its probability here on the left.

Now figure out these probabilities.

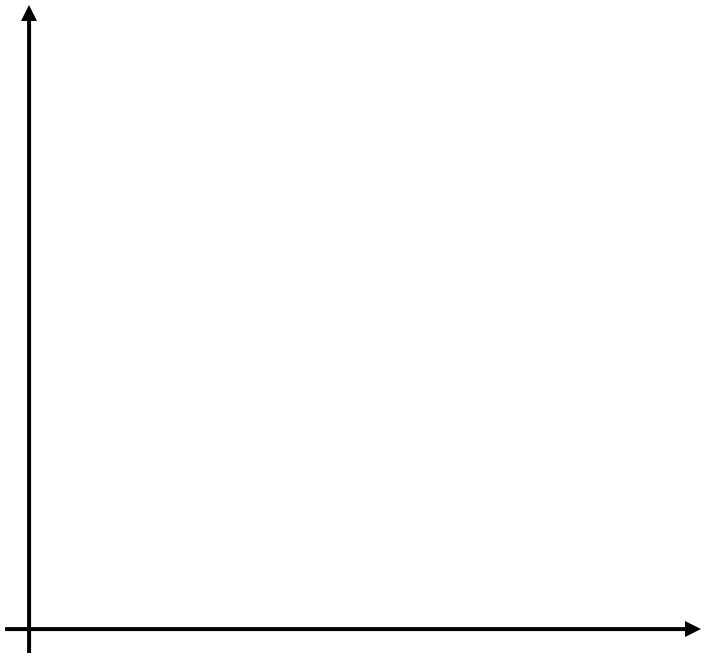
› $P(X = 4) =$

› $P(X = 7) =$

› $P(X < 3) =$

› $P(X \leq 4) + P(X > 9) =$

(3) Probability Density Function (PDF)



PDF can be both discrete and continuous.

Discreate PDF

› $0 \leq f(x) \leq 1$

› $\sum_{-\infty}^{\infty} f(x) = 1$

› $\sum_{a}^{b} f(x) = P(a \leq X \leq b)$

Continuous PDF

› $0 \leq f(x) \leq 1$

› $\int_{-\infty}^{\infty} f(x)dx = 1$

› $\int_{a}^{b} f(x)dx = P(a \leq X \leq b)$

(1) Conditional probability

There are a few basic concepts of bivariate distribution worth revising

› Joint probability density function

› Marginal probability

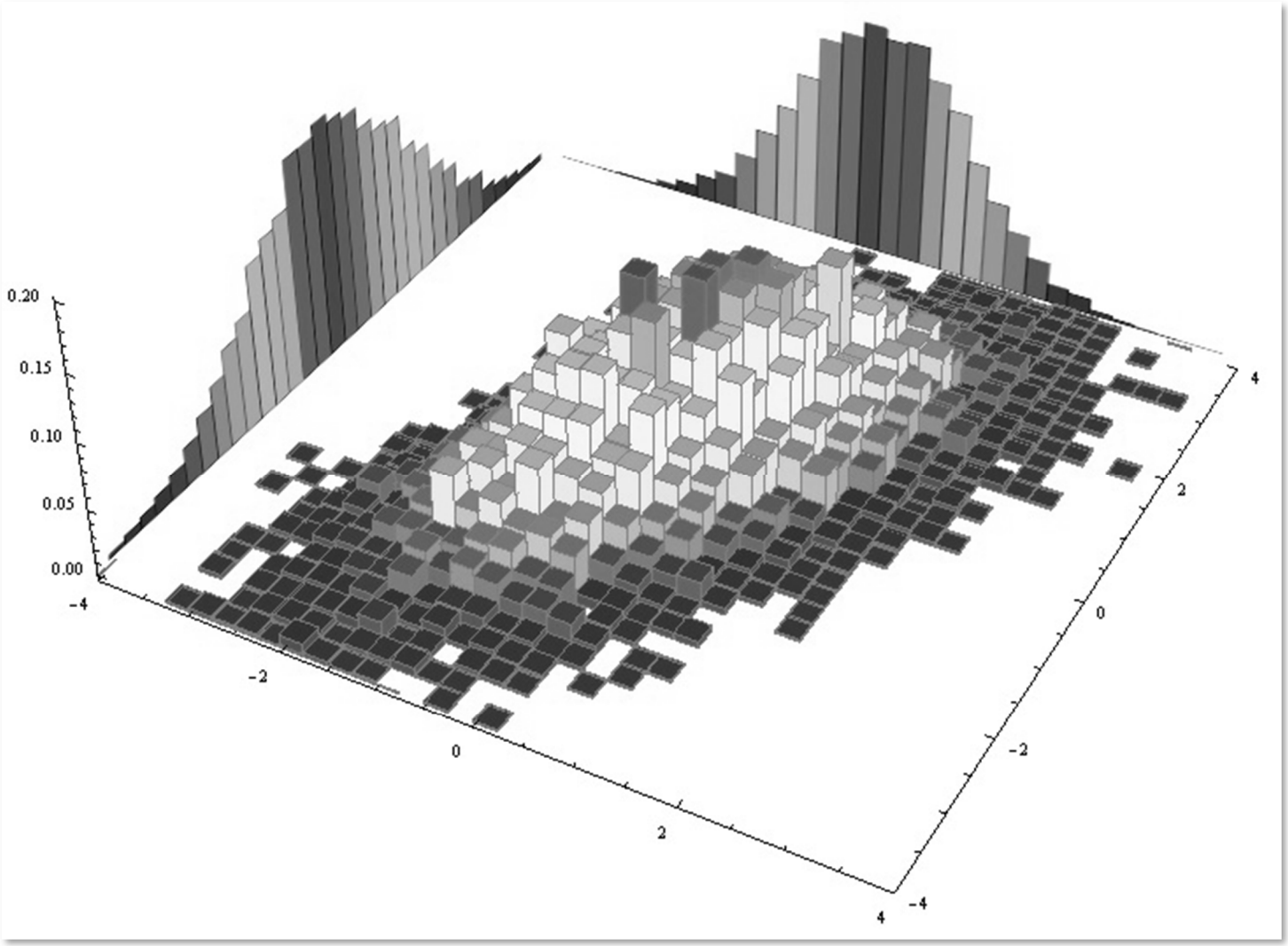
This class we will focus on

Definition 2.3

Conditional probability is a measure of the probability of an event occurring given that another event has (by assumption, presumption, assertion or evidence) occurred. Conditional probability is defined as

› $f(X|Y) = P(X = x|Y = y) = \frac{f(x,y)}{f(y)}$

(1) Conditional probability



(1) Conditional probability

Example Two archers are competing shooting a target for 3 times each.

Let X and Y be number of times archer 1 and archer and archer 2 hit the target respectively.

Thousand of rounds has been competed and the probability is computed as a result in this table.

		X		
		1	2	3
Y	1	0.35	0.2	0.07
	2	0.1	0.05	0.09
	3	0.08	0.04	0.02

› Find $f(X = 1|Y = 2) =$

› Find $f(Y = 2|X = 3) =$

(2) Statistical independence

Definition 2.4

Two random variables are considered **independent** if and only if the condition below is satisfied.

› $f(x,y) = f(x) \cdot f(y)$

Example Using the same archer example, prove that archers' performance is independent.

		X			›
		1	2	3	
Y	1	1/9	1/9	1/9	
	2	1/9	1/9	1/9	
	3	1/9	1/9	1/9	

(1) Expected value

Definition 2.5

Expected value of a random variable is a generalization of the weighted average and intuitively is the arithmetic mean of independent realizations of that variable. Expected value is defined as

› $E(X) = \sum_{i=1}^{n} x_i \cdot f(x_i)$ - discrete

› $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$ - continuous

Grade	Prob
A	0.3
B	0.4
C	0.15
D	0.1
F	0.05

Example A student is trying hard for econometrics class. The probability getting grades are listed in the table.

›

(1) Expected value

Properties of expected value

(1) $E(\mathrm{a}) = \mathrm{a}$ for any constant a

(2) $E(\mathrm{a}X) = \mathrm{a}E(X)$

(3) $E(\mathrm{a}X + \mathrm{b}) = \mathrm{a}E(X) + \mathrm{b}$

(4) $E(X \pm Y) = E(X) \pm E(Y)$

(5) $E(XY) = E(X) \cdot E(Y)$
if and only if X and Y are independent.

(1) Expected value

Example Find the expected value of this distribution

$f(X) = \frac{1}{9}x^2$ for $0 \leq x \leq 3$

›

(2) Conditional expectation

Definition 2.6

Let $f(X,Y)$ be a joint probability density function, the expectation of X conditional on some value of Y is

› $E(X|Y) = \sum_X x_i \cdot f(X|Y=y)$ - discrete

› $E(X|Y) = \int_{-\infty}^{\infty} x \cdot f(X|Y=y)dx$ - continuous

		X			
		-2	0	2	3
Y	3	0.27	0.08	0.16	0
	6	0	0.04	0.1	0.35

Example Find $E(X|Y=3)$ from the PDF given in the table.

›

(3) Variance

Definition 2.7

Variance is a measure of data dispersion from the expected value. Given that μ is the expected value of X, then

› $var(X) = \sum_{i=1}^{n}(x_i - \mu)^2 \cdot f(x_i)$- discrete

› $var(X) = \int_{-\infty}^{\infty}(x - \mu)^2 \cdot f(x)dx$- continuous

Another formula for variance is

› $var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$

(3) Variance

Properties of variance

(1) $\mathrm{var}(\mathrm{a}) = 0$ for any constant a

(2) $\mathrm{var}(\mathrm{a}X + b) = \mathrm{a}^2\mathrm{var}(X)$

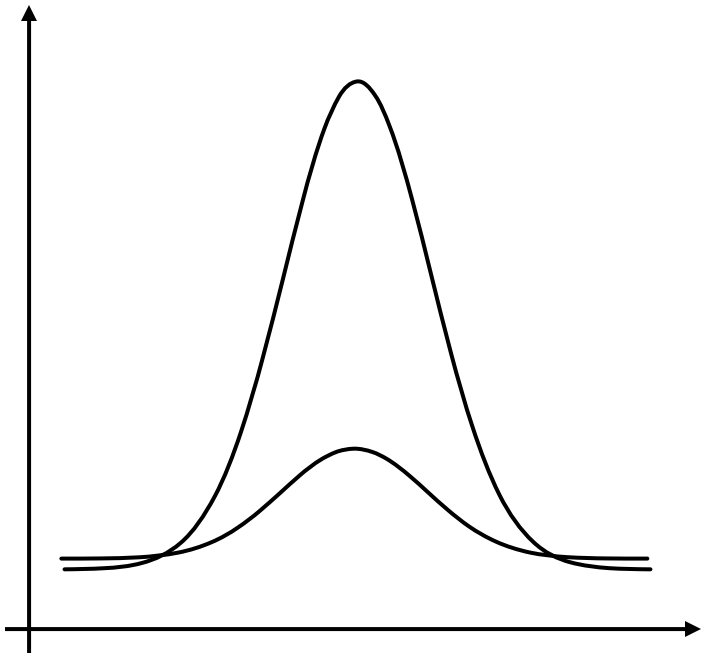
(3) $\mathrm{var}(X \pm Y) = \mathrm{var}(X) \pm \mathrm{var}(Y)$
if and only if X and Y are independent.

X	-2	1	2
$f(X)$	$5/8$	$1/8$	$2/8$

Example Find variance from the PDF given in the table.

›

(3) Variance



Example Find variance from given distribution

› $f(X) = \frac{1}{9}x^2$ for $0 \leq x \leq 3$

(4) Conditional variance

Definition 2.8

Conditional variance is a measure variance, but coupled with a condition on another variable, defined as

› $var(X|Y) = \sum_X [x_i - E(X|Y = y)]^2 \cdot f(X|Y = y)$ - discrete

› $var(X|Y) = \int_{-\infty}^{\infty} [x - E(X|Y = y)]^2 \cdot f(X|Y = y)dx$ - continuous

(5) Covariance

Definition 2.8

Let X and Y be two random variables with expected value of μ_X and μ_Y respectively, the **covariance** is a measure of the joint variability of two random variables.

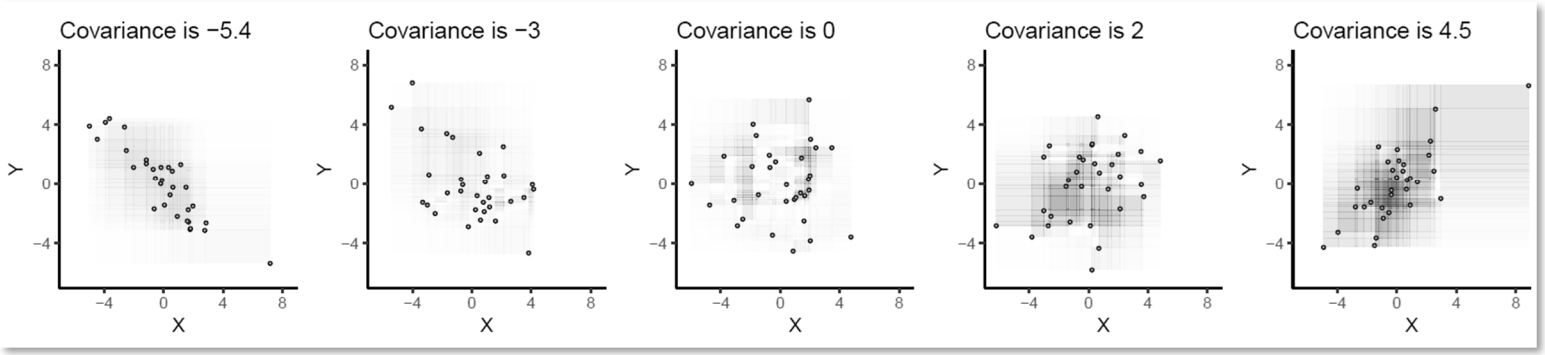
If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values. Defined as

› $cov(X,Y) = E\{(X-\mu_X)(Y-\mu_Y)\} = E(XY) - \mu_X\mu_Y$ - discrete

› $cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X-\mu_X)(Y-\mu_X)\cdot f(x,y)dxdy - \mu_X\mu_Y$

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XYf(x,y)\,dxdy - \mu_X\mu_Y$ - continuous

(5) Covariance



Further properties of variance

If X and Y are ***not*** independent, then

› $var(X \pm Y) = var(X) + var(Y) \pm 2cov(X, Y)$

Problems with interpretation

"A large covariance can mean a strong relationship between variables. However, you can't compare variances over data sets with different scales (like pounds and inches). A weak covariance in one data set may be a strong one in a different data set with different scales.

(6) Correlation coefficient

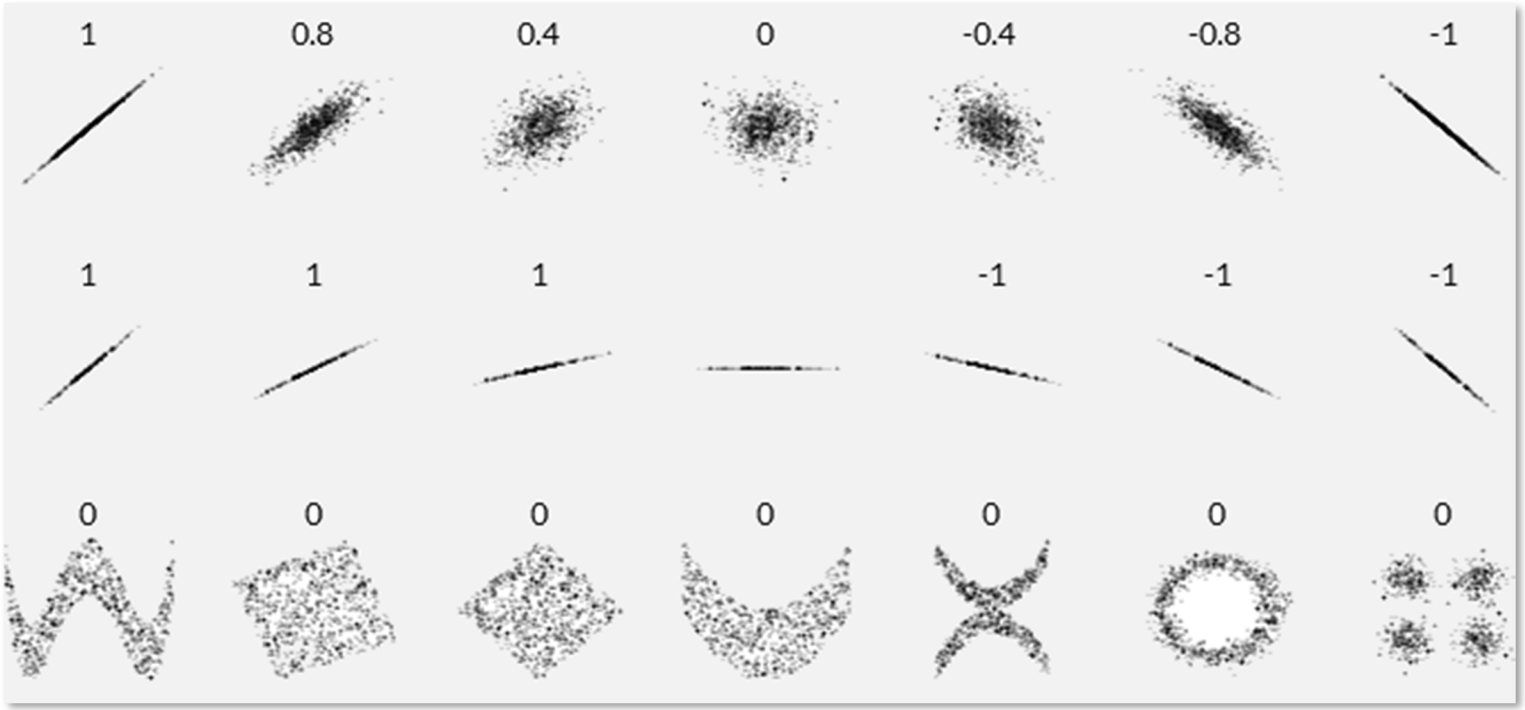
Definition 2.9

Correlation coefficient is a numerical measure of some type of correlation, meaning a statistical relationship between two variables, denoted by ρ_{XY}, r_{XY}, $\mathrm{corr}(X,Y)$.

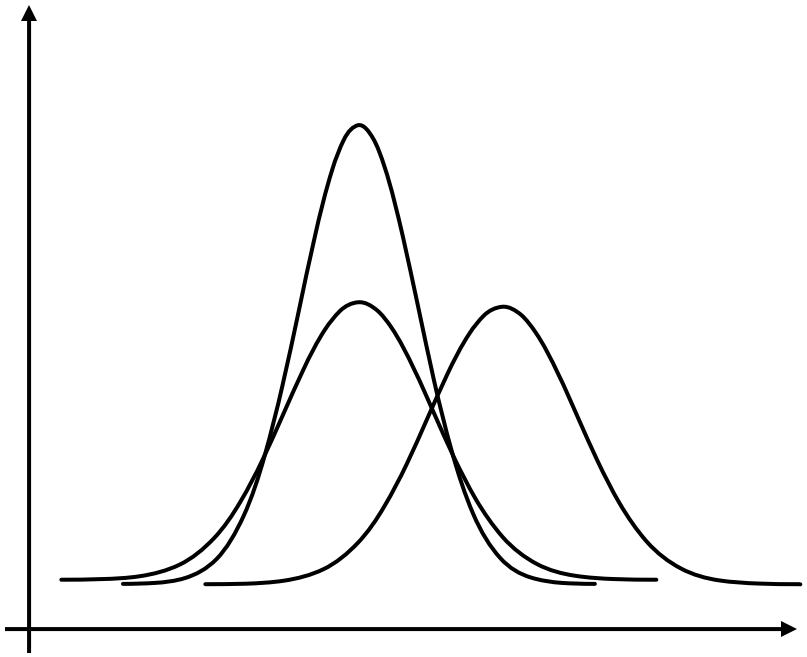
› $\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \epsilon [-1,1]$

where σ_X is standard deviation or $\sigma_X = \sqrt{\mathrm{var}(X)}$

(6) Correlation coefficient



(1) Normal distribution



A continuous random variable X is normally distributed with mean μ and variance σ^2, denoted as $X \sim N(\mu, \sigma^2)$, if the PDF is

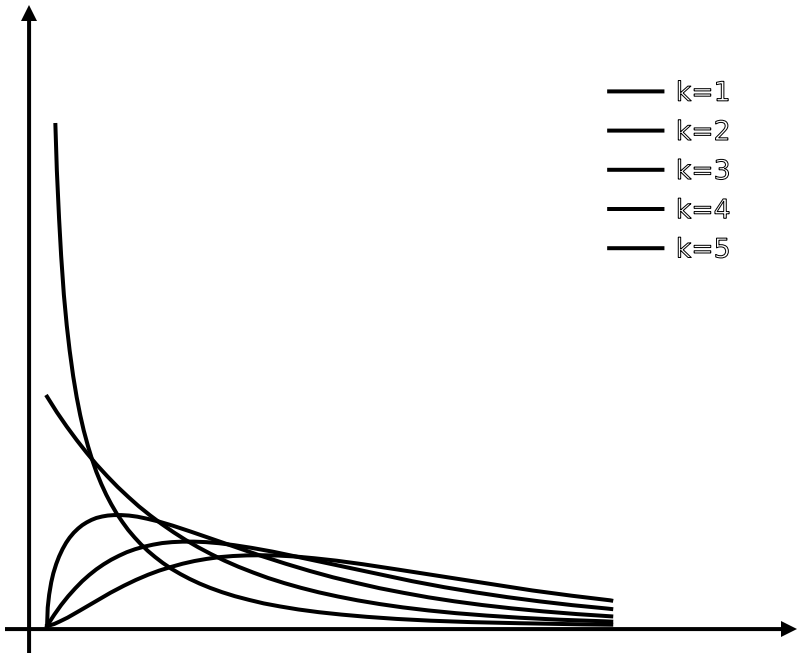
› $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

If the mean or variance changes, the position and shape of the distribution also shift.

We can convert any X that is normally distributed into a **standard normal distribution**, defined as Z, by weighting as follows.

› $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

(2) Chi-square distribution



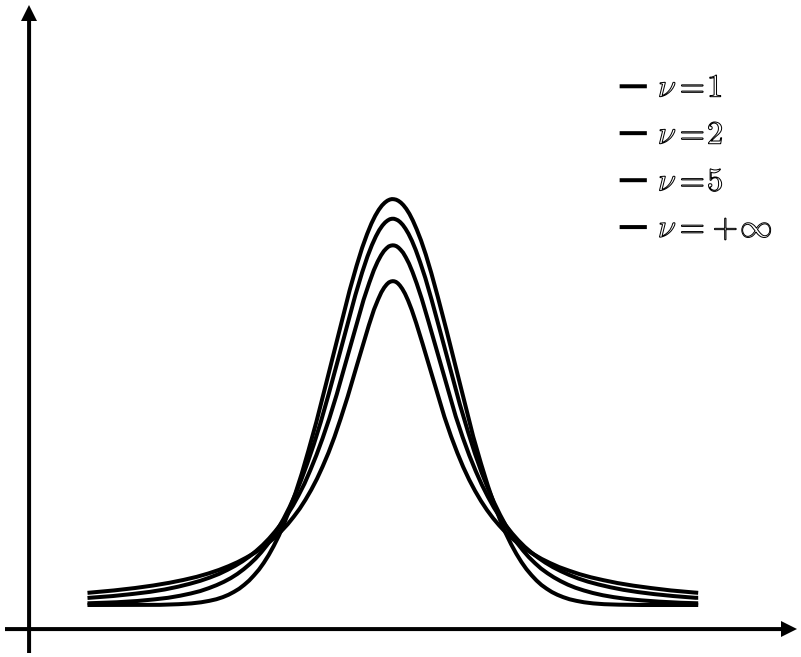
Chi-squared with k degrees of freedom (d.f.) is the distribution of a sum of the squares of k independent standard normal random variables.

› $\chi_k^2 = \sum_{i=1}^{k} Z_i^2$

Propertie of χ_k^2

› Chi-square is skewed depending on d.f. As the d.f. increases it becomes more and more symmetrical.

(3) Student's t-distribution



Let Z and χ_k^2 be random variables distributed as standard normal variable and chi-square respectively and they are independent, the t-distribution with k degrees of freedom can be represented as

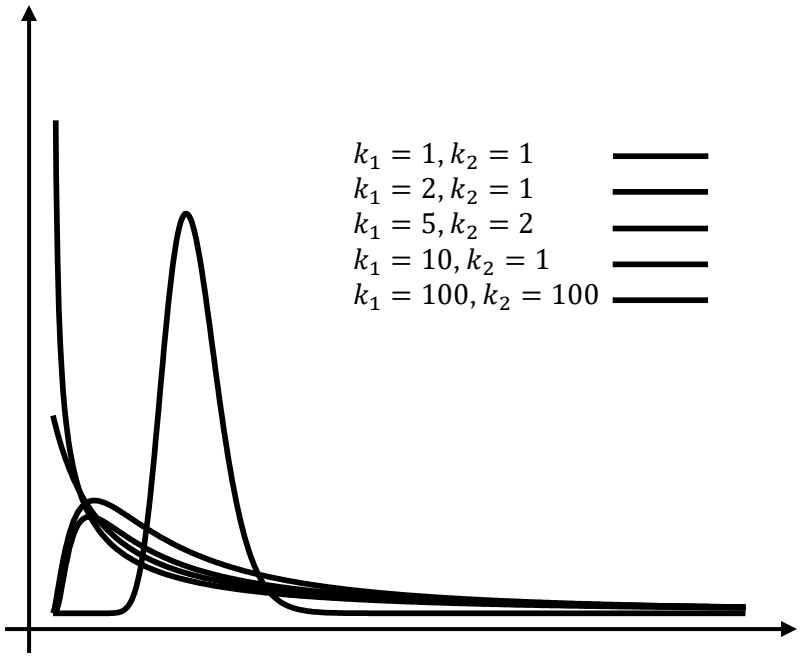
› $t = \frac{Z\sqrt{k}}{\chi_k^2} \sim t_v$

Properties of *t*

› The t-distribution is symmetric but flatter compared to the normal distribution.

› As the d.f. increases, t-distribution is converted to the normal distribution.

(4) F-distribution



Let χ_1^2 and χ_2^2 be random variables distributed as chi-square and they are independent with the d.f. of k_1 and k_2, the F-distribution can be represented as

› $F = \frac{\chi_1^2/k_1}{\chi_2^2/k_2} \sim F(k_1, k_2)$

Properties of F

› The F-distribution is skewed to the right but if k_1 and k_2 becomes larger, the F-distribution becomes normal distribution.

› The square of t-distributed random variable with k degrees of freedom is $t_v^2 = F_{1,v}$

Chapter 3

Simple Linear Regression

Flow of study in this chapter

› Population Regression Function

We first try to understand the meaning of demand, supply and equilibrium. How consumers and producers react in a market and how price can be a signal for both parties.

› Sample Regression Function

Commodities and services can be differently elastic. The implication on many studies forward will also be varied by their elasticity.

› Estimation

How to define what people gain from trade and lay out a framework to study who gains or loses when there is a change in a market.

› Assumptions underlying Classical Linear Regression Model (CLRM)

Learn how a political, economic institution can intervene price in a market, what is the implication and results for those actions.

Further reading can be found in Pindyck and Rubinfeld (2018) Part 2, Chapter 3-4.

(1) Creating a population regression function

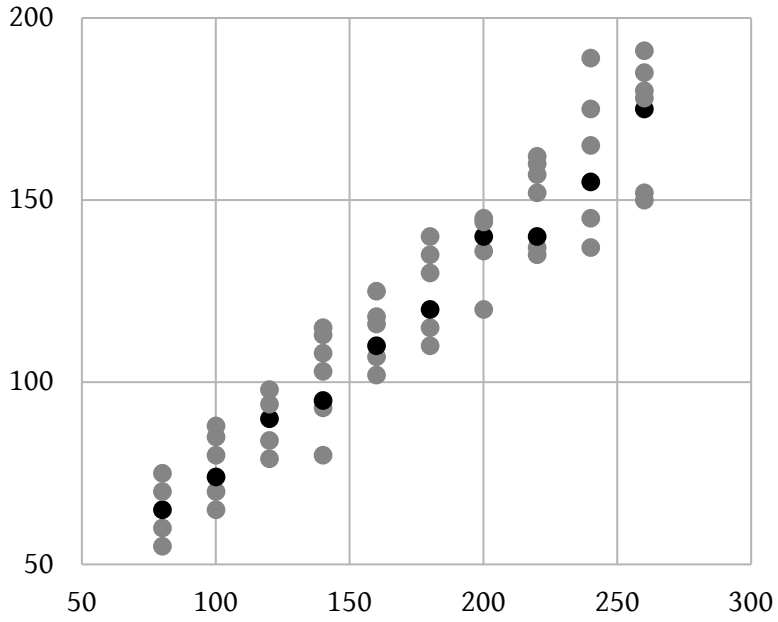
First, let's look at a set of cross-sectional data of household income and consumption. Given that they are both measured in $US.

› X is weekly income

› Y_i is weekly expenditure

X	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
80	55	60	65	70	75	.	.
100	65	70	74	80	85	88	.
120	79	84	90	94	98	.	.
140	80	93	95	103	108	113	115
160	102	107	110	116	118	125	.
180	110	115	120	130	135	140	.
200	120	136	140	144	145	.	.
220	135	137	140	152	157	160	162
240	137	145	155	165	175	189	.
260	150	152	175	178	180	185	191

(1) Creating a regression function

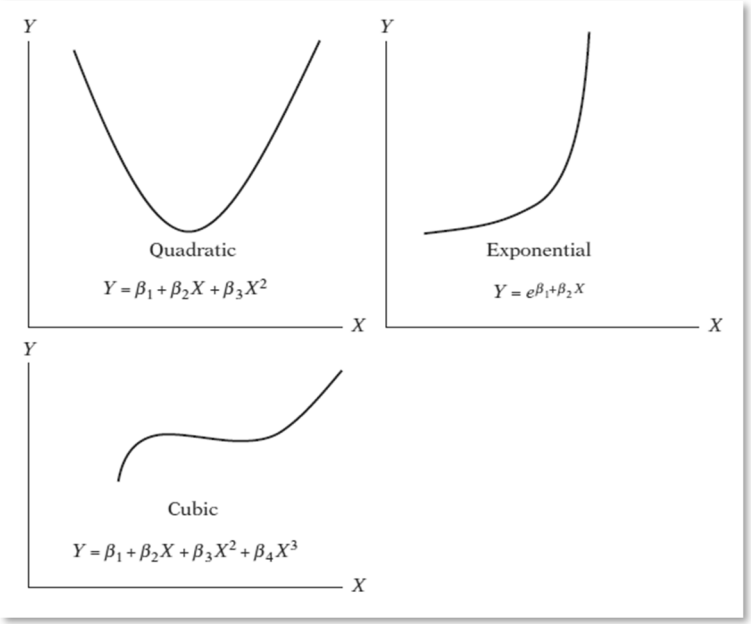


Since we are estimating linear relationship between X and Y, we define a linear **population regression function** (PRF) as

› $E(Y|X_i) = f(X_i) = \beta_1 + \beta_2 X_i$

So, the β_1 and β_2 are the **parameter** or

(2) Notes on linearity

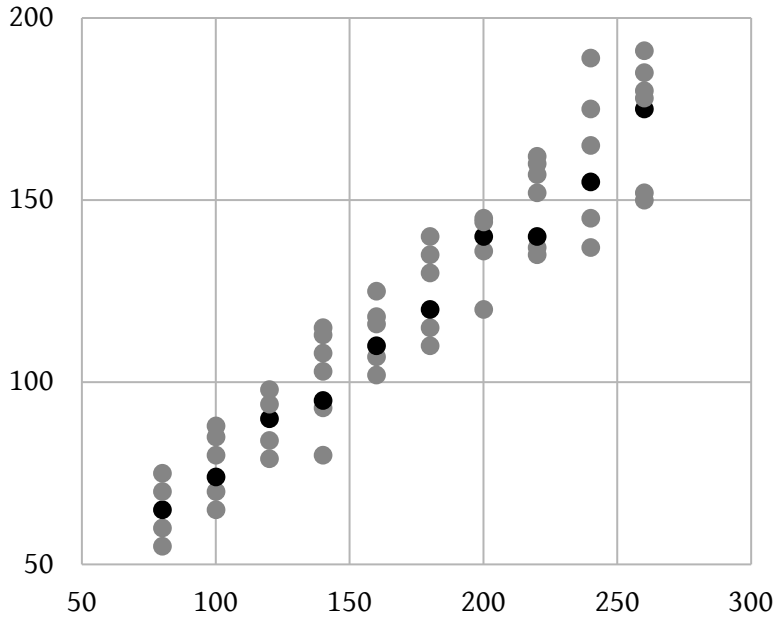


To clear things up, when we mention a linear regression model (**LRM**), we need to specify what are we talking about. There are two types of linearity.

› **Linear in variables**
geometrically, is the function linear or not, considered from the power of X_i.

› **Linear in parameters**
is the function consist of linear parameter or not, considered from the power of β_i.

(3) Stochastic specification



As we know that statistics relation is different from mathematical relation, therefore, we introduce a concept of the **stochastic disturbance** or **stochastic error term** as

› $u_i = Y_i - E(Y|X_i)$ or

› $Y_i = E(Y|X_i) + u_i$

Hence, the stochastic PRF is

› $Y_i = \beta_1 + \beta_2 X_i + u_i$

Now we have two parts of the PRF which are

› **Systematic** or **deterministic** part – which is $\beta_1 + \beta_2 X_i$.

› **Random** or **nonsystematic** part – which is u_i.

(3) Stochastic specification

For instance, let's assume that $\beta_1 = 40$ and $\beta_2 = 0.5$, figure out u_i for the following $E(Y|X_i = 180)$

› $Y_1 = 110 = 40 + 0.5(180) + u_1$ then $u_1 =$

› $Y_3 = 120 = 40 + 0.5(180) + u_3$ then $u_3 =$

› $Y_5 = 135 = 40 + 0.5(180) + u_5$ then $u_5 =$

(3) Stochastic specification

Why do we always have an error term in our equation? Here are some explanations:

› Vagueness of theory

› Unavailability of data
For example, using family wealth to explain consumption behavior but wealth data are usually not available.

› Core variables and peripheral variables
Variable(s) included in our model might be just peripheral ones. Picking a core variable may contribute to our model more.

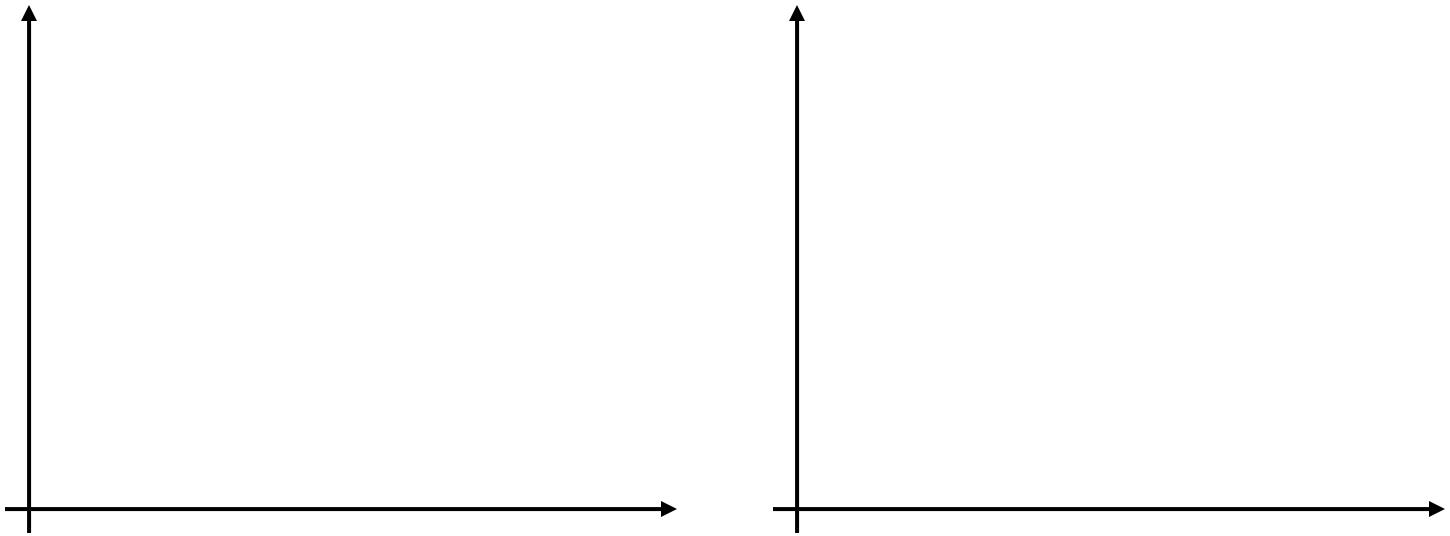
› Intrinsic randomness in human behavior

(3) Stochastic specification

› Poor proxy variables
Some intrinsic variable cannot be observed, such as intelligence and skills. Most of the time we rely on another variable as a proxy, such as test score, GPAX, work experience, etc.

› Principle of parsimony
It is what it is if there is no strong theory suggesting adding more variable, keep our model as simple as possible and let the error terms be as they are.

› Wrong functional form



(4) Expected value of the error term

With the error term included in the PRF, taking the expected through this equation

› $Y_i = E(Y|X_i) + u_i$

›

Since the $E(Y|X_i)$ is a constant, therefore

›

So $E(u_i|X_i) = 0$, or we can say that

› $E(u_i|X_i) = \sum_{i=1}^{n} \left(\frac{u_i|X_i}{n}\right) = 0$

(4) Expected value of the error term

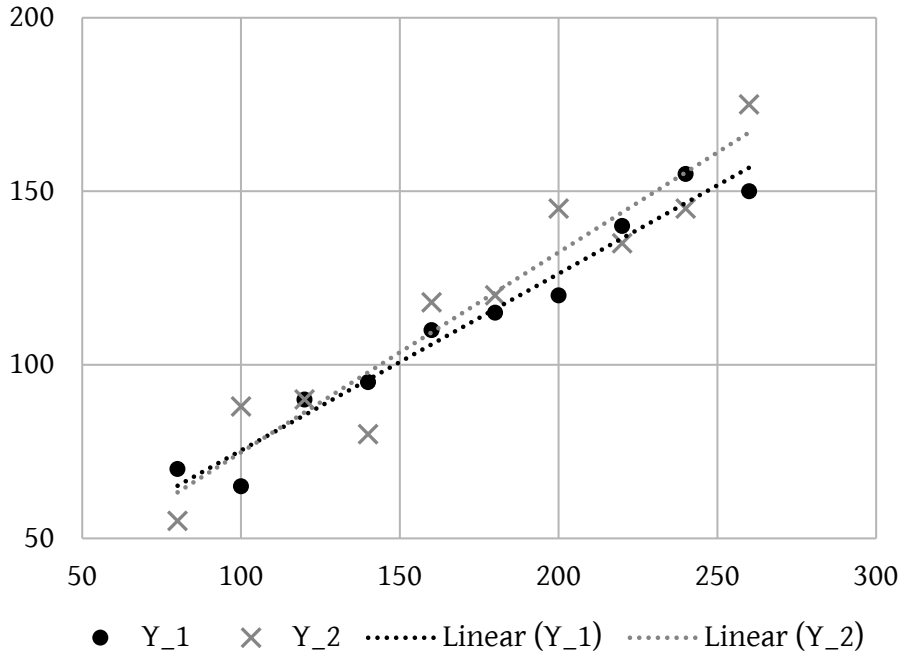
Similar to the property of sum of deviation from the mean, we can see the proof here that it is always zero. For instance,

› $\sum_{i=1}^{n}(x_i - \bar{X}) =$

(1) Creating sample regression functions

In real world scenario, most of the time we cannot collect all the population data. Assumed that we can only collect two sets of data shown in the table.

X	Y_1	Y_2
80	70	55
100	65	88
120	90	90
140	95	80
160	110	118
180	115	120
200	120	145
220	140	135
240	155	145
260	150	175



(1) Creating sample regression functions

Anyway, we still can define our **sample regression function** (SRF) as

› $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

The stochastic form as

› $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$

where $\hat{Y}_i$ is the estimator of $E(Y|X_i)$

$\hat{\beta}_i$ is the estimator of β_i

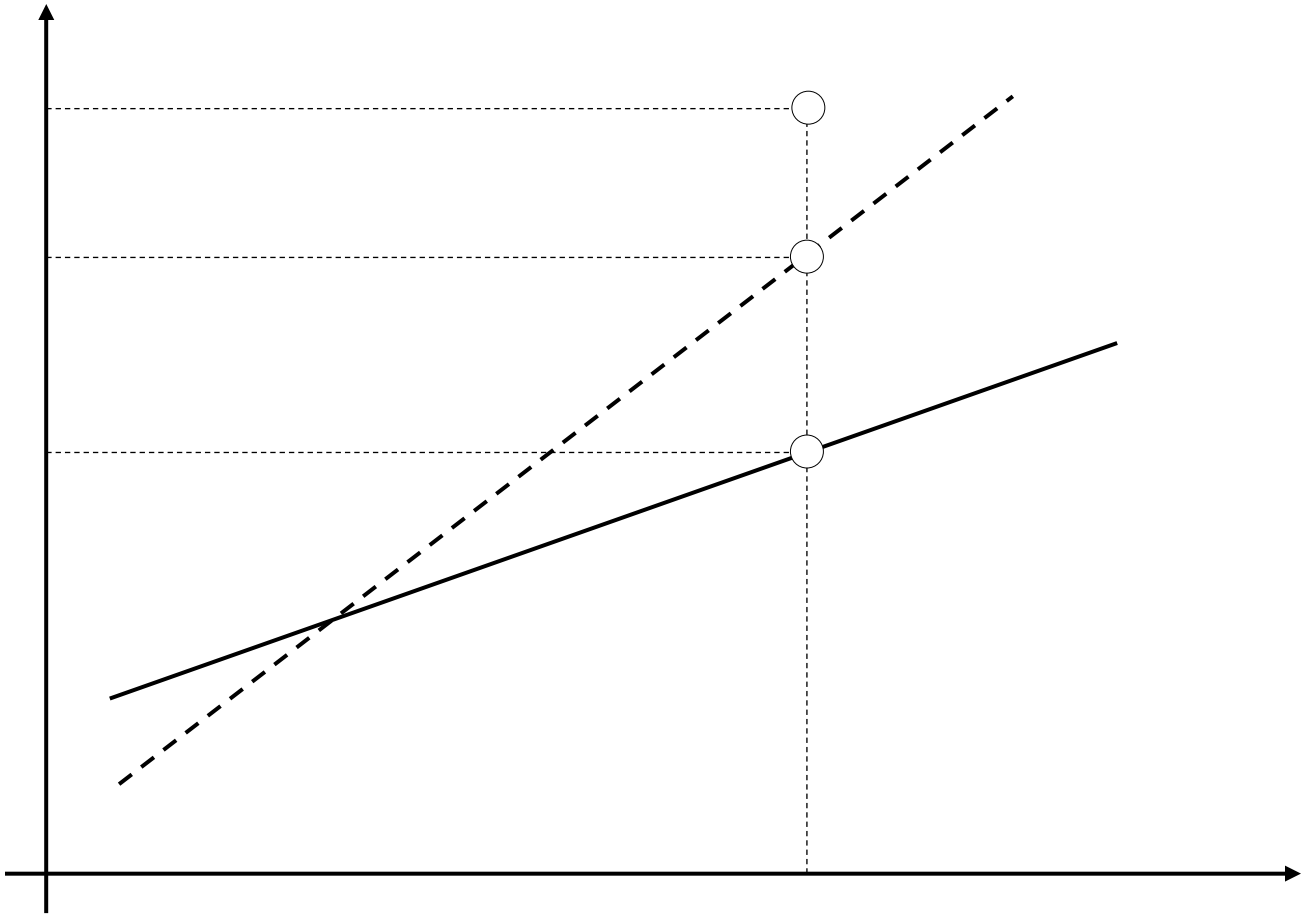
$\hat{u}_i$ is the estimator of u_i

To avoid confusion, since we are referring to β_i quite often

› β_1 and β_2 are called **parameters**.

› $\hat{\beta}_1$ and $\hat{\beta}_2$ are called **estimators**.

(2) Comparing PRF and SRF



Problem statement 1

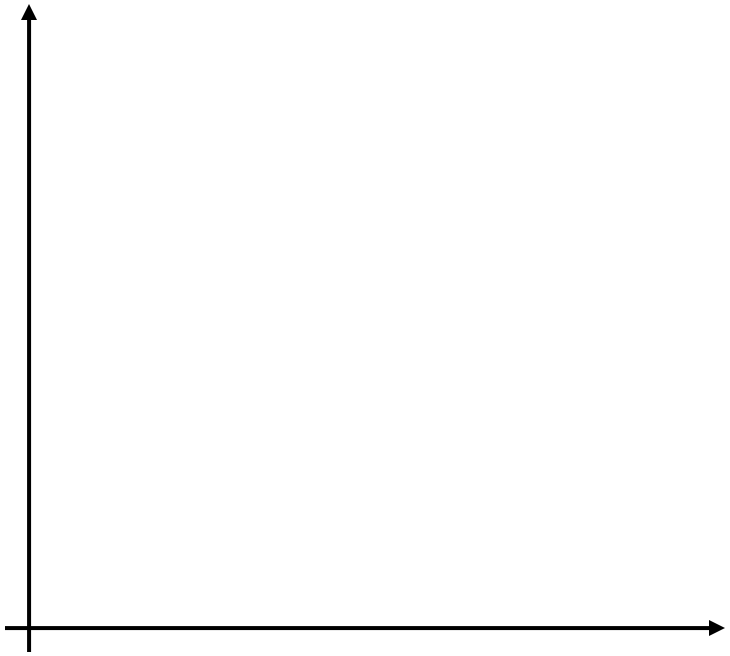
When we obtain a set of data, plot them on a graph, how can we draw a straight line, portraying regression relationship between two variables?

› How to connect each data point with a line that fits best with our data?

› What is/are (a) criteria (ion) that we can rely on?

(1) Ordinary Least Square

This method applies for both PRF and SRF. The intuition is shown here.



Now we try to draw a linear line that minimize sum of the error terms. From

› $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$

Rearranging the equation, we get

›

Setting up the objective function

›

(1) Ordinary Least Square

Solve for $\hat{\beta}_1$.

(1) Ordinary Least Square

Plug in $\hat{\beta}_1$ to solve for $\hat{\beta}_2$.

(1) Ordinary Least Square

$\sum X_i(Y_i - \bar{Y}) = \sum(X_i - \bar{X})(Y_i - \bar{Y})$ because

Eventually, we get

› $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}$

› $\hat{\beta}_2 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$

where $x_i = (X_i - \bar{X}), y_i = (Y_i - \bar{X}), x_i^2 = (X_i - \bar{X})^2$

(2) Properties of OLS estimators

(1) The OLS estimators are expressed solely in terms of the observables.

(2) They are **point estimators**, instead of interval estimators.

(3) They make the SRF passes through the sample mean.

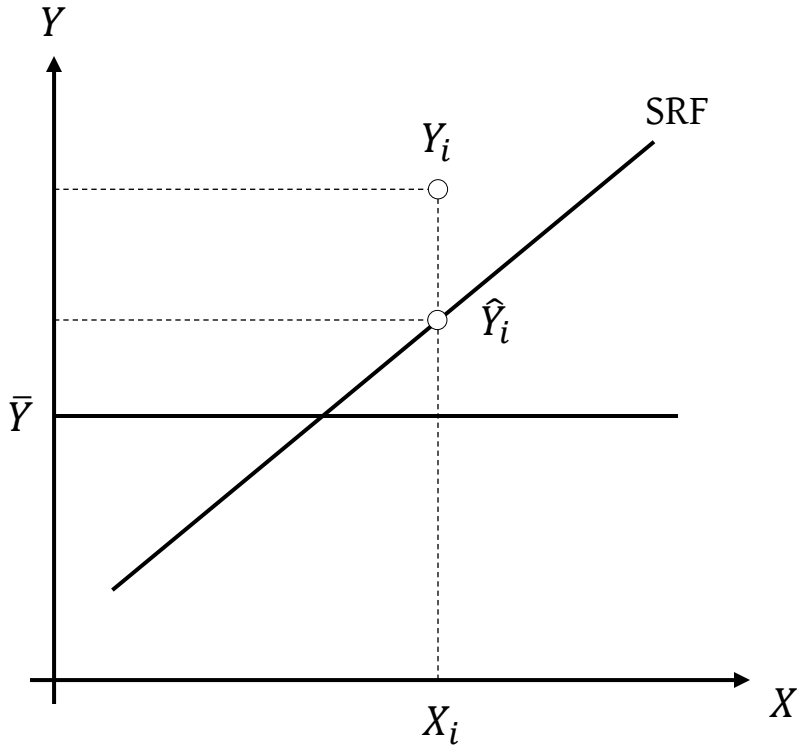
(4) The mean value of $\hat{Y}_i$or $\bar{\hat{Y}} = \bar{Y}$.

(5) The mean value of the residual $\hat{u}_i = 0$.

(6) $\hat{u}_i$ are uncorrelated with both X and $\hat{Y}$.

(3) Coefficient of determination (r^2)

The r^2 is determined by how much the is described by the SRF, or the measurement of **'goodness of fit'** of the fitted regression line comparing to an estimator, $\bar{Y}$.



The intuition is that total sum of squares (TSS) is equal to explained sum of squares (ESS) and residual sum of squares (RSS) or

›

(3) Coefficient of determination (r^2)

Eventually, we get

› $r^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$ or

› $r^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2}$

There are a few more formulae of r^2 in page 76

Properties of r^2

(1) Non-negativity

(2) $0 \leq r^2 \leq 1$

(4) Sample coefficient of correlation (r)

A formula of r^2 is

› $r^2 = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2}$

We can define **sample coefficient of correlation** (r) easily by

› $r = \pm\sqrt{r^2} = \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}}$

Properties of r

(1) Positive or negative depending on the sign of the term.

(2) $-1 \leq r \leq 1$

(3) Independent of the origin and scale.

(4) If X and Y are statistically independent, $r = 0$, but **not** vice versa.

(5) Does not describe non-linear association or causality.

(2) Non-stochastic X values

The **first** assumption, linear in the parameters, is already discussed. It is a starting point and applies throughout the course.

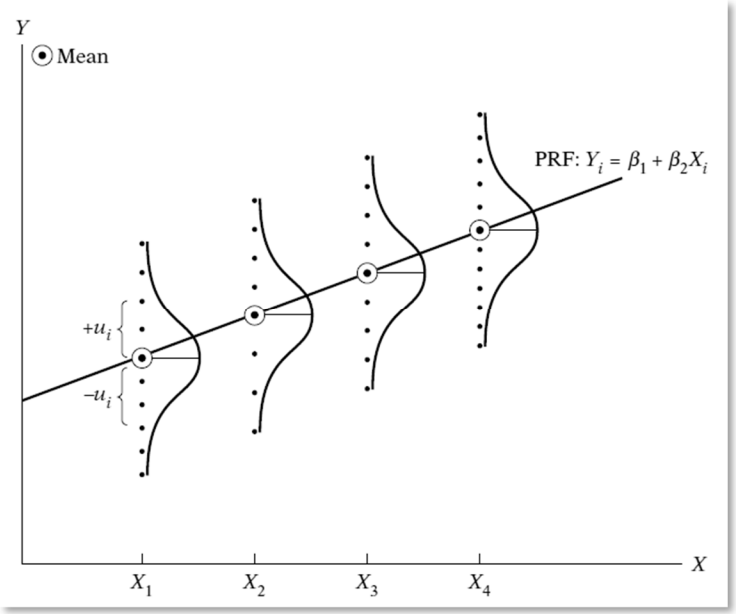
› $Y_i = \beta_1 + \beta_2 X_i + u_i$

The **second** assumption is X_i are independent of the error term, but it is not very important since when this assumption is relaxed, many statistical properties are still valid.

› $cov(X_i, u_i) = 0$

(3) Zero mean value of disturbance u_i

Basically, sum of deviation from the mean is always zero, which also implies that there is no **specification error** or **specification bias**.

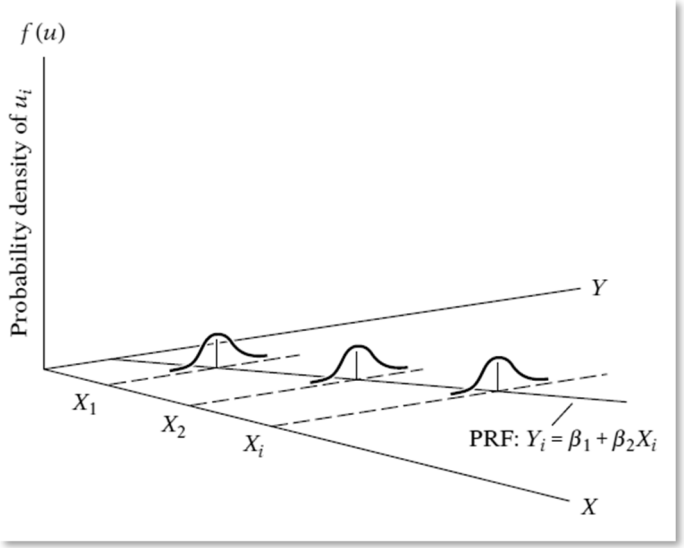


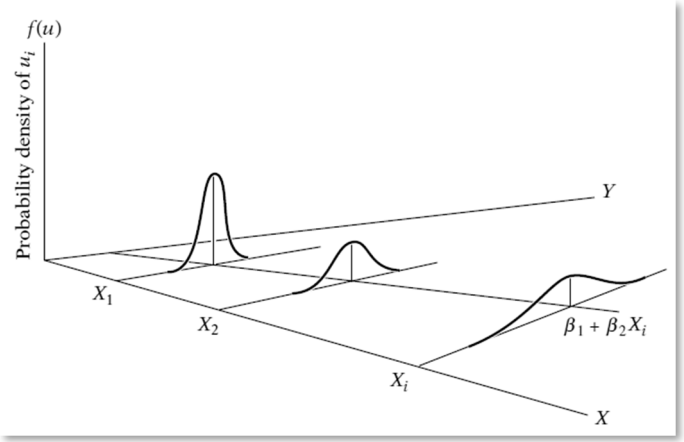
› $E(u_i|X_i) = 0$ or

› $E(u_i) = 0$

(4) Homoscedasticity

Homoscedasticity or constant variance of u_i means that





› $var(u_i) = \sigma^2$

(5) No autocorrelation

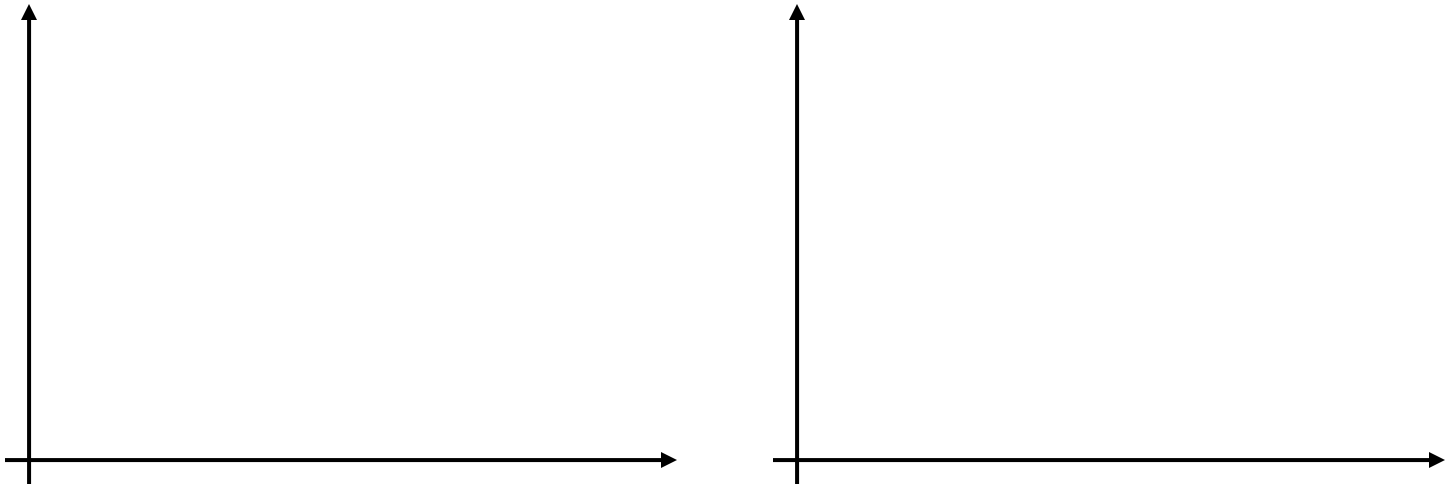
No autocorrelation or serial correlation between disturbances or

› $cov(u_i, u_j | X_i, X_j) = 0$ or

› $cov(u_i, u_j) = 0$ if X is non-stochastic

The **sixth** assumption is number of observations n must be greater than the number of parameters estimated k.

The **seventh** assumption is X value must not all be the same.



(1) Gauss-Markov Theorem

With certain assumptions imposed, OLS estimators is considered **BLUE** (best linear unbiased estimators)

(1) It is **linear**, that is, a linear function of a random variable, such as the dependent variable Y in the regression model.

(2) It is **unbiased**, that is, its average or expected value, $E(\hat{\beta}_2)$ is equal to the true value β_2 (multiple sampling).

› An estimator is considered unbiased when $E(\hat{\theta}) = \theta$

(3) It has minimum variance in the class of all such linear unbiased estimators: an unbiased estimator with the least variance is known as an **efficient estimator**.

› An estimator $\hat{\theta}_i$ is more efficient than $\hat{\theta}_j$ when

› $\text{var}(\hat{\theta}_i) < \text{var}(\hat{\theta}_j)$

(2) Normality assumption for u_i

Assumed that each u_i is distributed normally with

› Mean : $E(u_i) = 0$

› Variance : $E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2$

› Covariance : $E(u_i, u_j) = 0$

We can write it shortly with this notation

› $u_i \sim \text{NID}(0, \sigma^2)$

normally and independently distributed with 0 mean and σ^2 variance

(3) Central Limit Theorem (CLT)

Let $X_1, X_2, \dots, X_n$ denote n independent random variables distributed with the same PDF with mean of μ and variance of σ^2, as n increases indefinitely, then

› $\bar{X}_{n\to\infty} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$ regardless of the form of PDF.

We can then transform into a standard normal distribution as

› $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim \mathrm{N}(0,1)$

(4) Variances

Sampling will not yield the true value of σ^2, we also have no way to know its true value, therefore we can estimate it from

› $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k} = \frac{\sum (Y_i - \hat{Y})^2}{n-k}$

where $\sum \hat{u}_i^2$ is residual sum of squares and n-k is the degrees of freedom. k is number of parameters estimated, in this case it is 2 ($\hat{\beta}_1, \hat{\beta}_2$).

This $\hat{\sigma}^2$ can be found in the estimation result from STATA, known as **residual mean squares**. It will later be used for interval estimation and hypothesis testing, as and estimator of true variance.

(5) Sum up the properties

For the estimators $\hat{\beta}_1, \hat{\beta}_2$ with all the assumptions imposed, we eventually have

› Mean : $E(\hat{\beta}_1) = \beta_1$	› Mean : $E(\hat{\beta}_1) = \beta_1$
› Variance : $\sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2$	› Variance : $\sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2}$
Or more compactly $\hat{\beta}_1 \sim \mathrm{N}(\beta_1, \sigma^2_{\hat{\beta}_1})$	Or more compactly $\hat{\beta}_2 \sim \mathrm{N}(\beta_2, \sigma^2_{\hat{\beta}_2})$

We can also turn these distributions into standard normal as

› $Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim \mathrm{N}(0,1)$ and $Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \sim \mathrm{N}(0,1)$

The covariance between these estimators is

› $cov(\hat{\beta}_1, \hat{\beta}_2) = -\bar{X}\left(\frac{\sigma^2}{\sum x_i^2}\right) = -\bar{X} \cdot \mathrm{var}(\hat{\beta}_2)$

Lastly, since σ^2 is not known, we can plug in $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k}$ instead.

Problem statement 2

Multiple samplings yield different estimators, varied values of $\hat{\beta}_1$ and $\hat{\beta}_2$. If we only have a sample, how reliable the point estimation is?

Once we already set up many assumptions prior to this point, we can utilize them to construct an **interval estimation** by setting up a **confidence interval (CI)**. An example of β_2 is displayed below.

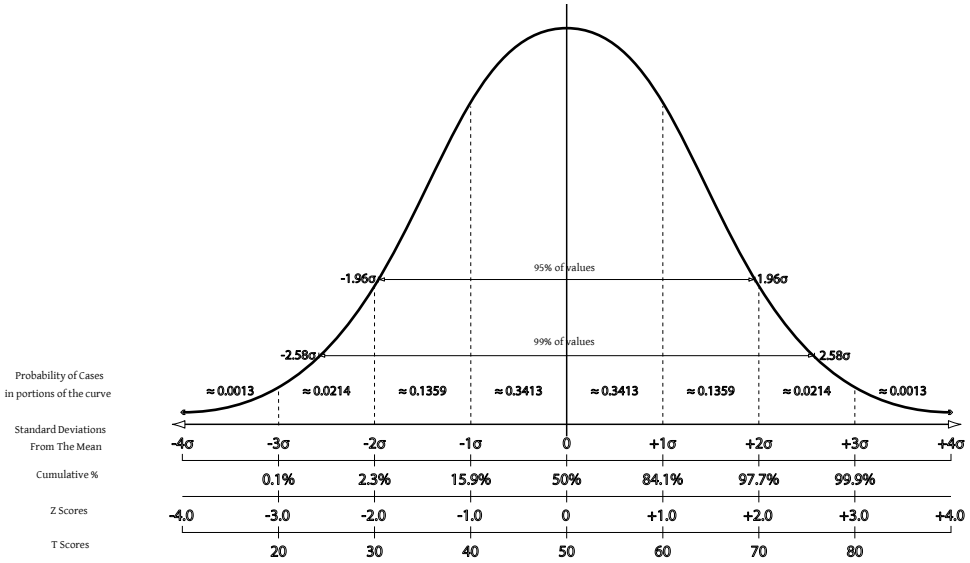
› $P\left[\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta\right] = 1 - \alpha$

where $1 - \alpha$ is confidence coefficient

α is level of significance

$\hat{\beta}_2 - \delta$ is lower limit and $\hat{\beta}_2 + \delta$ is upper limit

(1) Confidence interval



From the normality assumption, we normalize the distribution of $\hat{\beta}_2$ as

› $Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}}$

which means that we can find the lower and upper limit of $\hat{\beta}_2$ at any level of significance.

(1) Confidence interval

Since σ^2 is rarely known, we can t stat instead of Z and replace σ^2 with $\hat{\sigma}^2$

› $t = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}}$

Next, replacing this t into the CI, we get

› $P\left[-t_{\frac{\alpha}{2}} \leq \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \leq t_{\frac{\alpha}{2}}\right] = 1 - \alpha$

Rearranging the term to specify the upper and lower limit

› $P\left[\hat{\beta}_2 - \left(t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}\right) \leq \beta_2 \leq \hat{\beta}_2 + \left(t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}\right)\right] = 1 - \alpha$

In other words, $100(1-\alpha)\%$ CI for β_2 is

› $\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}$ and analogously for β_1, $\hat{\beta}_1 \pm t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_1}$

(2) Report estimation result

Before we move on to see what CI means, we shall consider how an estimation is reported. There can be multiple ways to do so.

Let's first define our model, given that

› $consmp_i = \hat{\beta}_1 + \hat{\beta}_2 inc_i + \hat{u}_i$

where $consmp_i$ is consumption expenditure of student i

inc_i is earned income + received income of student i.

(1) Traditional method

Report the estimators in the equation as follows.

› $\widehat{consmp}_i = 1{,}5690.58 + 0.4395 inc_i$

$se =$	(944.2059)	(0.0827)	$r^2 = 0.3963$
$t =$	(5.31)	(1.66)	$n = 45$
$p =$	(0.104)	(0.000)	$F_{1,43} = 28.23$

(2) Report estimation result

(2) STATA method

Report the estimators and all the essential information.

```
      Source |       SS           df       MS      Number of obs   =        45
-------------+----------------------------------   F(1, 43)        =     28.23
       Model |   208566659         1   208566659   Prob > F        =    0.0000
    Residual |   317733341        43  7389147.46   R-squared       =    0.3963
-------------+----------------------------------   Adj R-squared   =    0.3822
       Total |   526300000        44  11961363.6   Root MSE        =    2718.3

------------------------------------------------------------------------------
      consmp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |    .439518   .0827278     5.31   0.000     .2726815    .6063544
       _cons |   1569.058   944.2059     1.66   0.104    -335.1146    3473.231
------------------------------------------------------------------------------
```

(2) Report estimation result

(3) Modern method

Report the estimators and its significance using asteroids. Other information can be put anywhere as you wish.

(Dependent variable: consumption expenditure) Independent variables	
Income	0.4395*** (0.0827)
constant	1,569.058 (944.2059)
r^2	0.3963
n	45

Note: */**/*** denotes statistical significance at 90, 95 and 99 percent respectively.

(2) Report estimation result

Table 4: Regression Results

(Dependent variable: hours worked) Independent variables	Estimation part: Labor participation – Probit (Marginal effect)	Estimation part: Labor supply – Truncated regression (Coefficient)
1. Log earned income	1.327** (0.331)	12.389** (1.262)
2. Female # Log earned income	1.344 (1.148)	-5.767** (1.775)
3. Married # Log earned income	-0.913* (0.402)	-5.556** (1.363)
4. Female # married # Log earned income	-2.336 (2.012)	5.352* (2.028)
5. Unearned income x 100 [1]	-0.874** (0.150)	-1.080** (0.276)
6. Female # Unearned income x 100	0.119** (0.026)	0.083 (0.065)
7. Married # Unearned income x 100	-0.064 (0.040)	0.054 (0.065)
8. Female # Married # Unearned income x 100	0.081 (0.058)	-0.060 (0.080)
9. 10th decile # unearned income x 100 [2]	0.684** (0.148)	1.029** (0.274)
Region (base case: Bangkok)		
10. Central	0.722 (0.633)	9.928** (1.028)
11. North	1.683* (0.686)	-1.342 (1.247)
12. Northeast	4.662** (0.681)	-3.690** (1.115)
13. South	3.865** (0.716)	-25.091** (1.239)
Municipal area (base case: municipal)		
14. non-municipal	0.850** (0.329)	-13.054** (1.239)
Sex and marital status (base case: single male)		
15. Female	-6.666** (0.525)	58.419** (16.238)
16. Married	11.202** (0.708)	55.425** (12.674)
17. Female # married	-8.445** (0.736)	-53.717** (18.764)
Individual characteristics		
18. Year of education	0.334** (0.047)	-1.219** (0.095)
19. Age	6.032** (0.075)	0.998** (0.240)
20. Age squared	-0.072** (0.001)	-0.019** (0.003)
21. Number of children aged 0-6 in household	0.308 (2.312)	-1.154 (0.906)
22. Female # Number of children aged 0-6 in household	-4.001** (0.383)	0.337 (1.105)
23. Disability	-36.626** (1.730)	2.895 (3.644)
Constant	-4.697** (0.133)	93.248** (11.467)
Classification / Sigma	83.62	51.482** (0.263)

Note: 1) Since the effect is tiny, unearned income is multiplied with 100.

2) Only the tenth decile interacted with unearned income in a significant manner. Other deciles are controlled, but are not shown here for table concision.

*/** denotes statistical significance at 90 and 95 percent, respectively. # sign refers to the interaction term. No serious collinearity is detected and robust standard error is displayed in parentheses.

(3) Finding confidence interval

Once we retrieve estimation result, we can construct confidence interval for both β_1 and β_2.

› From $\widehat{consmp}_i = 1{,}569.058 + 0.4395 inc_i$

$$
\begin{aligned}
se &= (944.2059) \quad (0.0827) \qquad r^2 = 0.3963 \\
t &= (5.31) \qquad (1.66) \qquad n = 45 \\
p &= (0.104) \qquad (0.000) \qquad F_{1,43} = 28.23
\end{aligned}
$$

› Step 1: pick an α

Usually, acceptable levels of significance are 0.01, 0.05 or 0.1, depending on how many percent that you are going to cover.

For example, if we pick $\alpha = 0.05$, the most common value, it means that the CI that we are going to find will indicate that 95% of the time, **the upper and lower limit will contain the true value of β_2.**

(3) Finding confidence interval

› Step 2: look up for $t_{\frac{\alpha}{2}}$

Now it's time that we need to look up on a t stat table, this can be found in Gujarati and Porter, Appendix D, page 877.

Be aware that the degrees of freedom must match our model $(n - k)$.

› Degrees of freedom is

Now we are constructing the upper and lower limit, therefore, t is spread symmetrically to the left and the right of the mean. To look for the 95% (0.95) area limit, the left and the right of that limit is equally 2.5% (0.025).

› $t_{\frac{\alpha}{2}}$ is

(3) Finding confidence interval

› Step 3: calculate the upper and lower limit

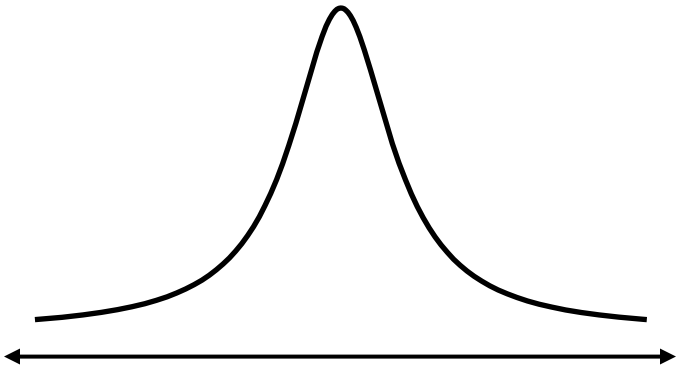
Now we can proceed to calculate the limit, using the formula

› $\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}$

To write it in probability form, on the other hand, we might use this notation.

› $P\left[\hat{\beta}_2 - \left(t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}\right) \leq \beta_2 \leq \hat{\beta}_2 + \left(t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}\right)\right] = 1 - \alpha$

(3) Finding confidence interval



What are we doing actually? Take a look at this graphical representation on the left.

Our calculation shows that where the upper and lower limit for the true β_2 are.

(3) Finding confidence interval

Repeat the process, now for β_1.

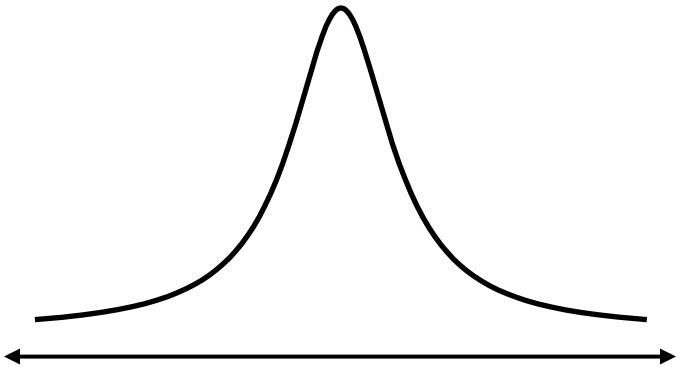
› **Step 1: pick an α**

› **Step 2: look up for $t_{\frac{\alpha}{2}}$**

› **Step 3: calculate the upper and lower limit**

› $\hat{\beta}_1 \pm t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_1}$

(3) Finding confidence interval



Indicate the area where β_1 will be within.

(4) Confidence interval for σ^2

Under the normality assumption, we assume that the error term is normally distributed. Hence, their squares are distributed as chi-square.

› $(n-k)\cdot\frac{\hat{\sigma}^2}{\sigma^2}\sim\chi^2$, we can construct the interval for σ^2 as

› $P\left[(n-k)\cdot\frac{\hat{\sigma}^2}{\chi^2_{\frac{\alpha}{2}}}\leq\sigma^2\leq(n-k)\cdot\frac{\hat{\sigma}^2}{\chi^2_{1-\frac{\alpha}{2}}}\right]=1-\alpha$

Be very careful that chi-square is an asymmetric distribution.

(4) Confidence interval for σ^2

Example: Find the 95% CI of σ^2 given that $\hat{\sigma}^2 = 7{,}389{,}147$ and $n = 45$.

Note that this value is simply mean squares of the residual. If we put this back to calculate lower and upper bound, multiplying $n - k$ back yields RSS

› Step 1: look up for $\chi^2_{\frac{\alpha}{2}}$ and $\chi^2_{1-\frac{\alpha}{2}}$

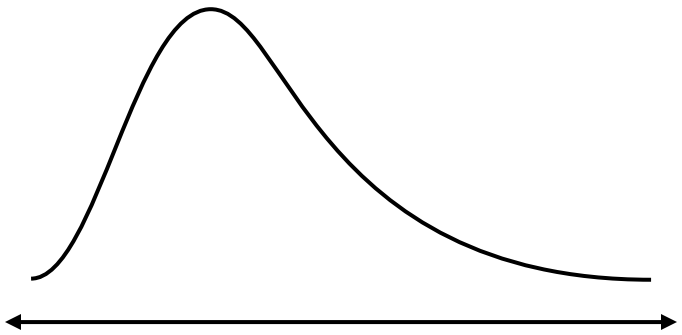
› Step 2: calculate the upper and lower limit

Lower limit:

$$(n-k) \cdot \frac{\hat{\sigma}^2}{\chi^2_{\frac{\alpha}{2}}} =$$

Upper limit:

$$(n-k) \cdot \frac{\hat{\sigma}^2}{\chi^2_{1-\frac{\alpha}{2}}} =$$



(1) Two tails test

Once we can find the confidence interval, we can apply the concept and follow the steps here to test our hypothesis.

Example#1: First of all, we look at the two tails **test against zero.**

› Step 1: State your hypothesis

$H_0: \beta_2 = 0$ – Null hypothesis

$H_a: \beta_2 \neq 0$ - Alternative hypothesis

The reason why we usually test against zero is that we want to make sure that β_2 is not zero. In other words, when β_2 is not zero, our X and Y are said to be related.

(1) Two tails test

› Step 2: Calculate test statistics

Using the same result that we have here

$$\widehat{consmp}_i = 1{,}569.058 + 0.4395 inc_i$$

$se = (944.2059)$	(0.0827)	$r^2 = 0.3963$
$t = (5.31)$	(1.66)	$n = 45$
$p = (0.104)$	(0.000)	$F_{1,43} = 28.23$

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} =$$

where t_{cal} denotes calculated test statistics

(1) Two tails test

› Step 3: State your decision rule

Now we pick an α to have an acceptable probability, most of the time we use $\alpha = 0.05$. Use this information to create CI around $\beta_2 = 0$, based on the degrees of freedom, to see how much the distribution covers.

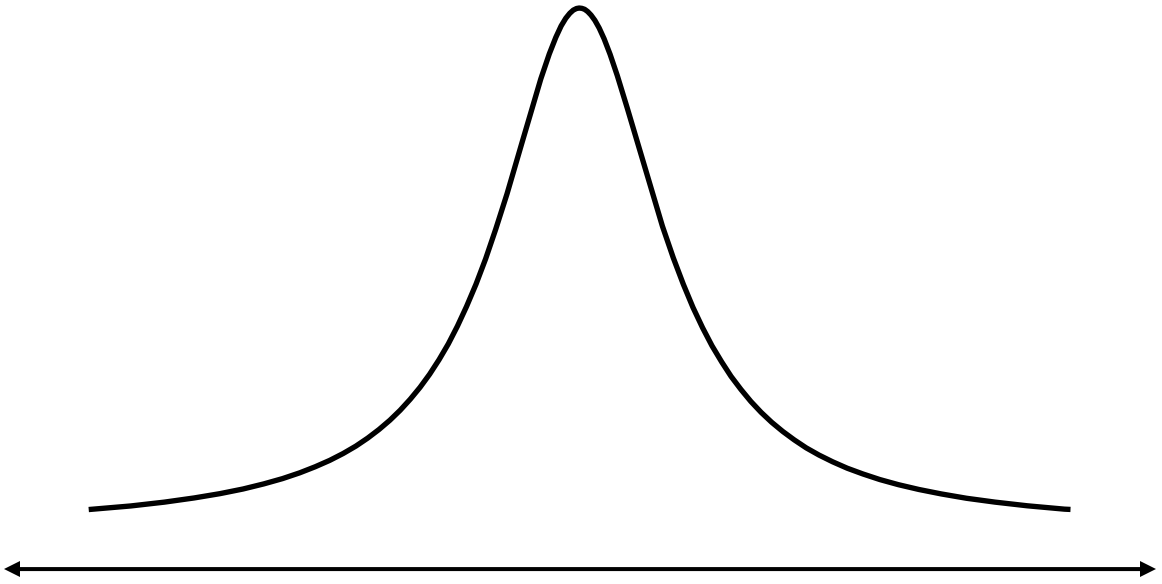
The lower bound : $t_{\frac{\alpha}{2}} =$

The upper bound : $t_{\frac{\alpha}{2}} =$

Note that when we test against zero ($\beta_2 = 0$), the distribution is normalized, therefore, there is no need to transform t from the table value.

(1) Two tails test

› **Step 3: State your decision rule**



(1) Two tails test

› Step 4: Conclude the test

(1) If t_{cal} lies **beyond** any boundary of CI (critical region), we **can reject the null hypothesis**, at the significance level of 95%.

In other words, **we are sure** that β_2 is not zero 95 out of 100 times when we sample.

(2) If t_{cal} lies within any boundary of CI (acceptance region), we **cannot reject the null hypothesis**, at the significance level of 95%.

In other words, we **cannot say for sure** that β_2 is not zero 95 out of 100 times when we sample.

(1) Two tails test

Example#2: Now let's look at another **test against a value that is not zero.**

› Step 1: State your hypothesis

$H_0: \beta_2 = 0.4$ – Null hypothesis

$H_a: \beta_2 \neq 0.4$ - Alternative hypothesis

› Step 2: Calculate test statistics

$$\widehat{consmp}_i = 1{,}569.058 + 0.4395 inc_i$$

$se = (944.2059)$	(0.0827)	$r^2 = 0.3963$
$t = (5.31)$	(1.66)	$n = 45$
$p = (0.104)$	(0.000)	$F_{1,43} = 28.23$

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} =$$

(1) Two tails test

› Step 3: State your decision rule

Again, we pick $\alpha = 0.05$. Use this information to create CI around $\beta_2 = 0.4$, based on the degrees of freedom, to see how much the distribution covers.

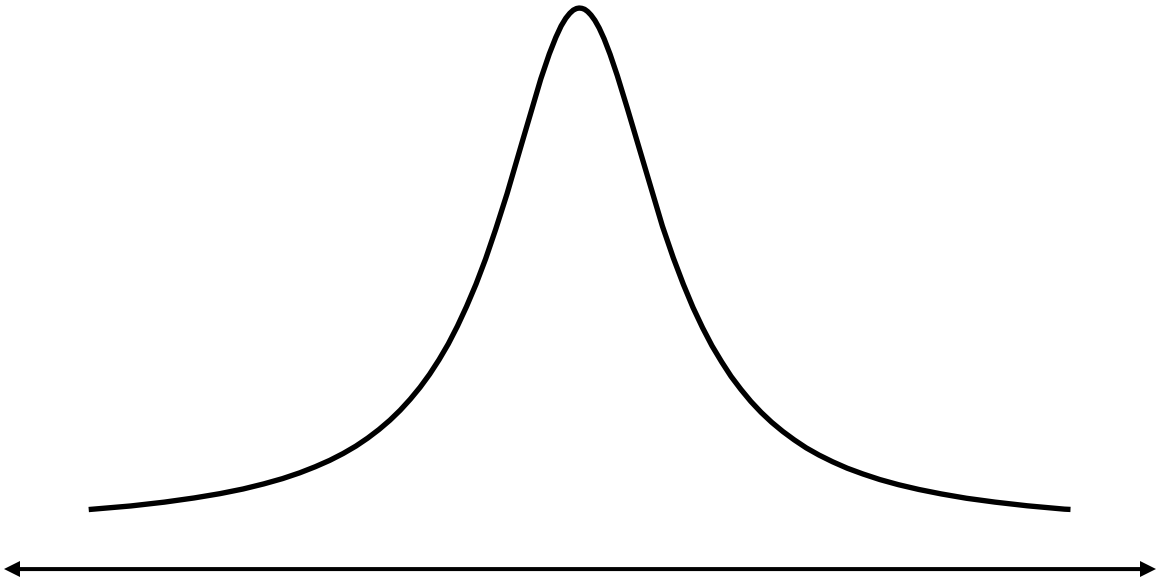
The lower bound : $\beta_2 - t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2} =$

The upper bound : $\beta_2 + t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2} =$

Note that when we test against another value that is not zero (in this case $\beta_2 = 0.5$), the distribution is **not yet** normalized, therefore, we need to transform t from the table value to find the CI around $\beta_2 = 0.4$.

(1) Two tails test

› **Step 3: State your decision rule**



(1) Two tails test

› Step 4: Conclude the test

(1) If t_{cal} lies **beyond** any boundary of CI (critical region), we **can reject the null hypothesis**, at the significance level of 95%.

In other words, **we are sure** that β_2 is not 0.4 95 out of 100 times when we sample.

(2) If t_{cal} lies within any boundary of CI (acceptance region), we **cannot reject the null hypothesis**, at the significance level of 95%.

In other words, we **cannot say for sure** that β_2 is not 0.4 95 out of 100 times when we sample.

(2) One tail test

Example#3: Now we consider a one tail test if the estimator exceeds or below specific value or not

› Step 1: State your hypothesis

$H_0: \beta_2 \geq 0.6$ – Null hypothesis

$H_a: \beta_2 < 0.6$ - Alternative hypothesis

For this test, we want to make sure that β_2 is more than 0.6 or not.

(2) One tail test

› Step 2: Calculate test statistics

Using the same result that we have here

$$\widehat{consmp}_i = 1{,}569.058 + 0.4395 inc_i$$

$$\begin{array}{lll} se = (944.2059) & (0.0827) & r^2 = 0.3963 \\ t = (5.31) & (1.66) & n = 45 \\ p = (0.104) & (0.000) & F_{1,43} = 28.23 \end{array}$$

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} =$$

where t_{cal} denotes calculated test statistics

(2) One tail test

› Step 3: State your decision rule

Again, we pick $\alpha = 0.05$. Use this information to create one tail acceptance region when $\beta_2 \geq 0.6$, based on the degrees of freedom, to see how much the distribution covers. Be careful with t value since this is a one tail test.

The lower bound : $\beta_2 - t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2} =$

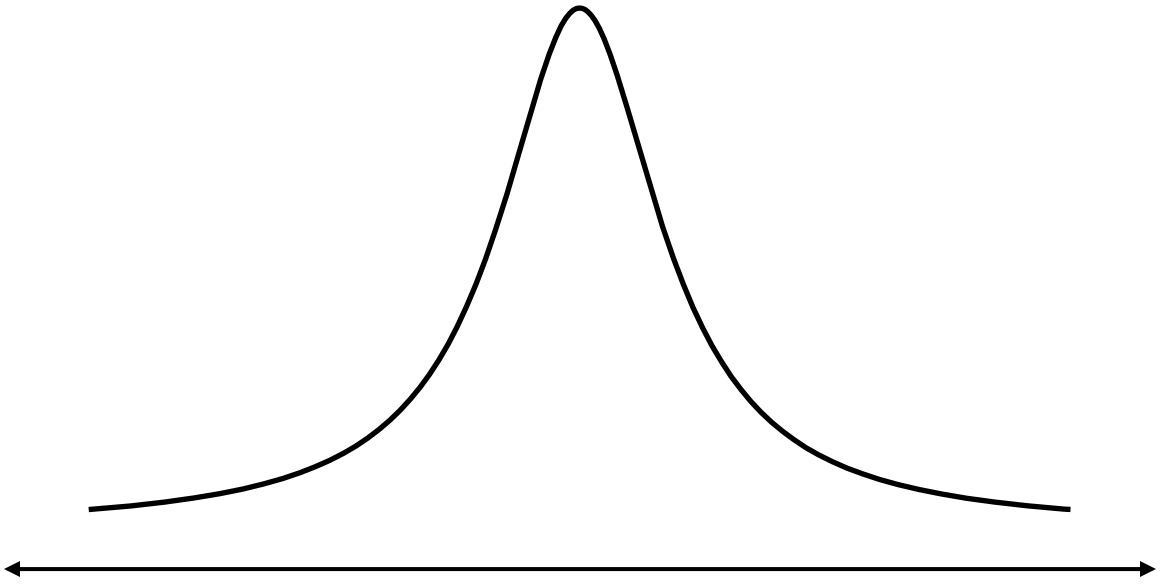
Note that when our null hypothesis is

(1) More than a specific value ($\geq$): acceptance region is on the right, figure out the lower bound of acceptance region.

(2) Less than a specific value ($\leq$): acceptance region is on the left, figure out the upper bound of acceptance region.

(2) One tail test

› **Step 3: State your decision rule**



(2) One tail test

› Step 4: Conclude the test

(1) If t_{cal} lies **beyond** any boundary of CI (critical region), we **can reject the null hypothesis**, at the significance level of 95%.

In other words, **we are sure** that β_2 is more than 0.6 95 out of 100 times when we sample.

(2) If t_{cal} lies within any boundary of CI (acceptance region), we **cannot reject the null hypothesis**, at the significance level of 95%.

In other words, we **cannot say for sure** that β_2 is more than 0.6 95 out of 100 times when we sample.

(3) Additional notes

If we are to make a mistake on our conclusion, there are two types of errors listed here.

› When we **reject** the null hypothesis when it is **true**, we call it a **Type I error.**

› On the other hand, if we **accept** the null hypothesis when it is **false**, we call it a **Type II error**.

(3) Additional notes

According to the STATA report , we can see that P > |t| or P-value is also reported. This is a very useful tool since we do not need to pick a specific level of significance.

› If P > |t| is less than any α, we can make sure that $(1-\alpha)$% of the time β_2 is not zero.

```
      Source |       SS           df       MS      Number of obs   =        45
-------------+----------------------------------   F(1, 43)        =     28.23
       Model |   208566659         1   208566659   Prob > F        =    0.0000
    Residual |   317733341        43  7389147.46   R-squared       =    0.3963
-------------+----------------------------------   Adj R-squared   =    0.3822
       Total |   526300000        44  11961363.6   Root MSE        =    2718.3

------------------------------------------------------------------------------
      consmp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |    .439518   .0827278     5.31   0.000     .2726815    .6063544
       _cons |   1569.058   944.2059     1.66   0.104    -335.1146    3473.231
------------------------------------------------------------------------------
```

(1) Prediction

Note that this is a **historical regression**, we might make use of to 'predict' or 'forecast'.

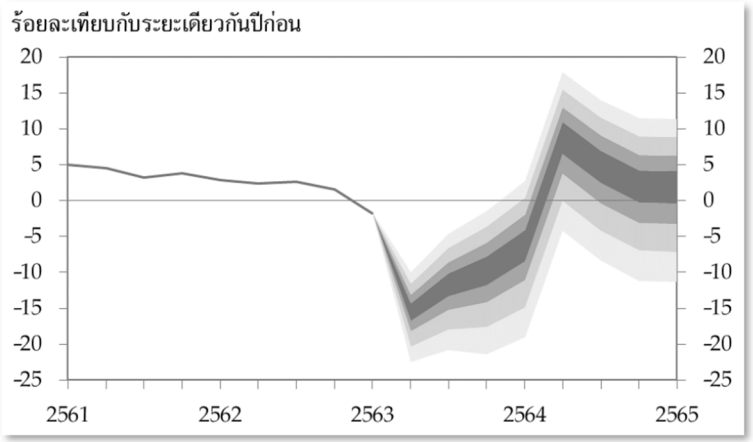
We can actually put our X_0 value of interest into the estimated

› $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$

to get a point estimation. However, it is not a very popular choice.

ตารางประเมินโอกาสที่จะเกิดขึ้นของการขยายตัวทางเศรษฐกิจในอัตราต่าง ๆ

ร้อยละ	2563				2564				2565
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
> 20	0	0	0	0	0	2	0	0	0
16-20	0	0	0	0	0	9	2	0	0
12-16	0	0	0	0	0	21	9	3	3
8-12	0	0	0	0	0	23	20	13	12
4-8	0	0	0	0	2	19	23	22	22
0-4	0	0	0	2	12	13	19	22	22
(-4)-0	100	0	3	12	23	7	13	18	18
(-8)-(-4)	0	1	17	25	23	3	8	11	12
(-12)-(-8)	0	15	32	25	18	1	4	6	6
(-16)-(-12)	0	40	27	18	12	0	1	3	3
(-20)-(-16)	0	30	14	11	6	0	0	1	1
(-24)-(-20)	0	11	5	5	3	0	0	0	0
(-28)-(-24)	0	2	1	2	1	0	0	0	0
< (-28)	0	0	0	1	0	0	0	0	0



Source: Financial Report, June 2020, BOT

(1) Prediction

There are two types of prediction that we can make once we retrieved the estimators.

(1) Mean prediction: Providing that mean estimation follows this equation

› $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$

when $\hat{Y}_0$ is an estimator of $E(Y|X_0)$ while X_0 represents a value of interest. Let's consider an easier example here, given that

› $\hat{Y}_i = -0.0144 + 0.7240 X_i$

If we are interested in out-of-sample $X_0 = 20$, then

› $\hat{Y}_0 = -0.0144 + 0.7240(20) = 14.4656$

(1) Prediction

Given that the variance of $\hat{Y}_0$ is, (no proof provided here)

› $var(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(x_i - \bar{X})^2}\right]$

Again, we do not have the true value of σ^2, so $\hat{\sigma}^2$ is replaced. We can then find the CI at the specific point of X_0 by

› $Pr\left[\hat{Y}_0 - \left(t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{Y}_0}\right) \leq Y_0 \leq \hat{Y}_0 + \left(t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{Y}_0}\right)\right] = 1 - \alpha$

(1) Prediction

Example#1: Given that

› $\hat{Y}_i = -0.0144 + 0.7240X_i$ and $X_0 = 20, \hat{Y}_0 = 14.4656$

› $n = 13, \bar{X} = 12, \sum(x_i - \bar{X})^2 = 182, \hat{\sigma}^2 = 0.8936$

Step 1: Find the $var(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(x_i - \bar{X})^2}\right]$

(1) Prediction

Step 2: Find the $\sigma_{\hat{Y}_0}$

Step 3: Find the 95% CI for $E(Y|X_0 = 20)$

Interpretation: if we create a CI over the mean value, 95 out of 100 times that the CI will cover true value $E(Y|X_0)$.

(1) Prediction

(2) Individual prediction: Contrast to the mean prediction, which estimates the variance around Y_0, individual prediction focuses on forecasting error (fe), defined as

› $fe = \hat{Y}_0 - Y_0$

Therefore, we define the variance of this fe as

› $var(fe) = var(\hat{Y}_0 - Y_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(x_i - \bar{X})^2}\right]$

Similarly, replacing the unknown σ^2 with the unbiased estimator $\hat{\sigma}^2$, we can derive the CI for Y_0 corresponding to X_0

› $Pr\left[\hat{Y}_0 - \left(t_{\frac{\alpha}{2}} \cdot \sigma_{fe}\right) \leq Y_0 \leq \hat{Y}_0 + \left(t_{\frac{\alpha}{2}} \cdot \sigma_{fe}\right)\right] = 1 - \alpha$

(1) Prediction

Example#2: Given that

› $\hat{Y}_i = -0.0144 + 0.7240X_i$ and $X_0 = 20, \hat{Y}_0 = 14.4656$

› $n = 13, \bar{X} = 12, \sum(x_i - \bar{X})^2 = 182, \hat{\sigma}^2 = 0.8936$

Step 1: Find the $var(fe) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(x_i - \bar{X})^2}\right]$

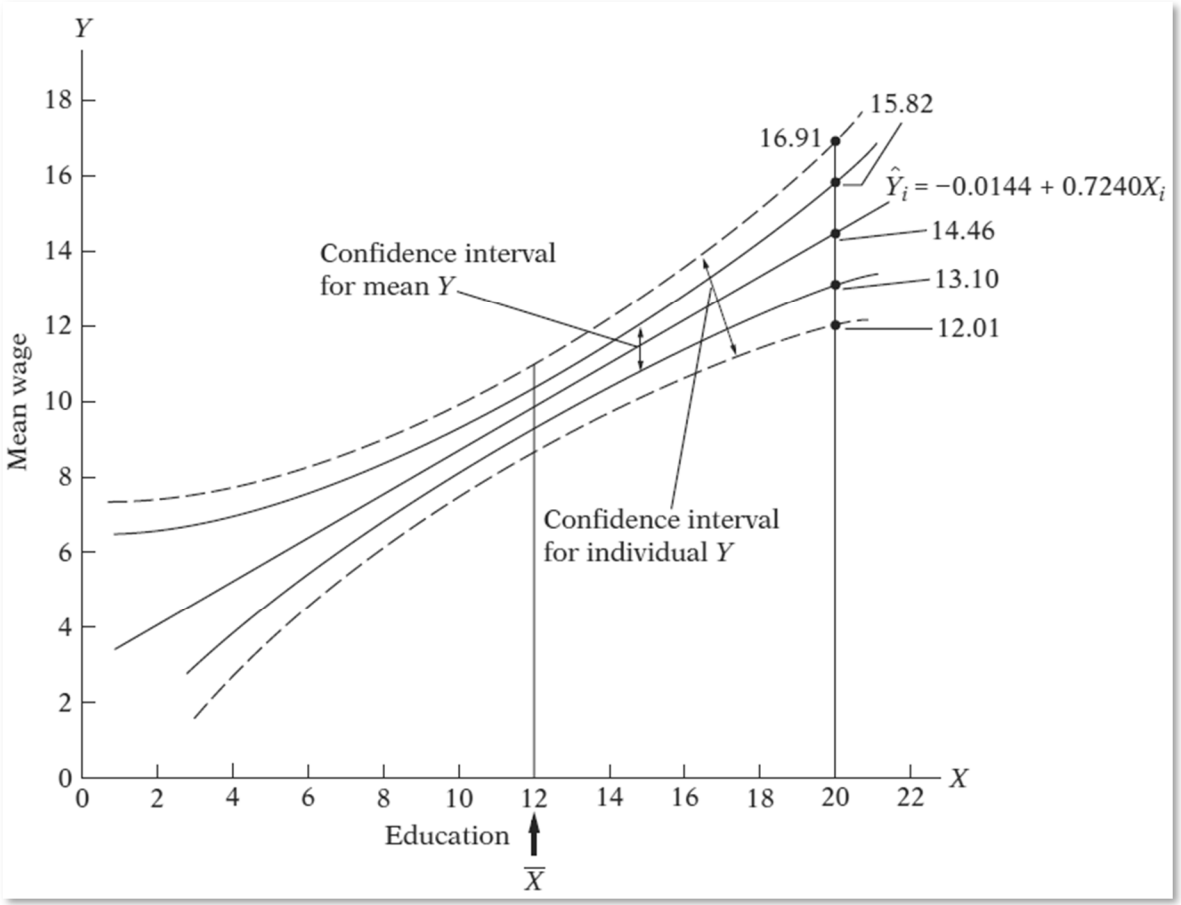
(1) Prediction

Step 2: Find the σ_{fe}

Step 3: Find the 95% CI for Y_0 corresponding to $X_0 = 20$

(1) Prediction

Comparing CI band of mean and individual prediction.



(2) Regression through the origin

There are some occasions that we assume our estimation model without an intercept as

› $Y_i = \hat{\beta}_2 X_i + \hat{u}_i$

Obtaining the estimator from OLS,

› $\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$

› $var(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2}$

› $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1}$

Note that the d.f. is one unit less due to the disappearance of $\hat{\beta}_1$.

(2) Regression through the origin

Differences that should also be noted are

› $\sum \hat{u}_i X_i = 0$ but $\sum \hat{u}_i$ need not be zero

› r^2 can be negative, so we need another coefficient of determination, defined as ***raw* r^2**

› $raw\ r^2 = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$

Unless there is **very strong a priori expectation**, we should avoid zero intercept regression model and stick to the conventional intercept-present model, because it may lead to **specification error** (will be elaborated in the last chapter).

However, if $\hat{\beta}_1$ turns out to be statistically insignificant (from being zero), $\hat{\beta}_2$ is **a lot more precise** when estimated by the regression through the origin model.

(3) Data scaling

Sometimes when the result is reported, scaling can be difficult to make sense of. For example, if $\hat{\beta}_2 = 3.054e^{-15}$ which is not very effective for communication. Thus, data scaling can fix this without affecting the result. See the examples below.

› Both GPDI and GDP in billions of dollars

$$\widehat{GPDI}_t = -926.090 + 0.2535GDP_t$$

$$se = (116.358) \quad (0.0129) \qquad r^2 = 0.9648$$

› Both GPDI and GDP in millions of dollars

$$\widehat{GPDI}_t = -926{,}090 + 0.2535GDP_t$$

$$se = (116{,}358) \quad (0.0129) \qquad r^2 = 0.9648$$

(3) Data scaling

› GPDI in billions of dollars, GDP in millions of dollars

$$\widehat{GPDI}_t = -926.090 + 0.0002535GDP_t$$

$$se = (116.358) \quad (0.0000129) \qquad r^2 = 0.9648$$

› GPDI in millions of dollars, GDP in billions of dollars

$$\widehat{GPDI}_t = -926{,}090 + 253.524GDP_t$$

$$se = (116{,}358) \quad (12.9465) \qquad r^2 = 0.9648$$

(4) Functional forms

There are several functional forms that are linear in parameters and can be estimated.

(1) Log-linear model: Sometimes called **log-log, double-log, or log-linear** models, log-linear model takes a form of

› $Y = \beta_1 X_i^{\beta_2}$ We can linearize the function by taking log on both sides.

We will find that the slope and elasticity has a very interesting properties, by differentiation.

› Slope

› Elasticity

(4) Functional forms

Therefore, we can say that $\beta_2 = \frac{relative\ change\ in\ Y}{relative\ change\ in\ X} =$

A practical use for log-linear model is **Price demand.**

(4) Functional forms

Example of log-log and its interpretation.

› $\ln \widehat{wage}_i \mathrel{+}= 1.5 + 3.30 \ln educ_i$

› If education increase by one year, we expect wages to increase by β_2 percent.

(4) Functional forms

(2) Semi-log models

Respectively, semi-log models consist of **log-lin** and **lin-log model** which take a form of

› $\ln Y = \beta_1 + \beta_2 X_i$ and

› $Y = \beta_1 + \beta_2 \ln X_i$

Again, we are going to find slope and elasticity.

› Slope

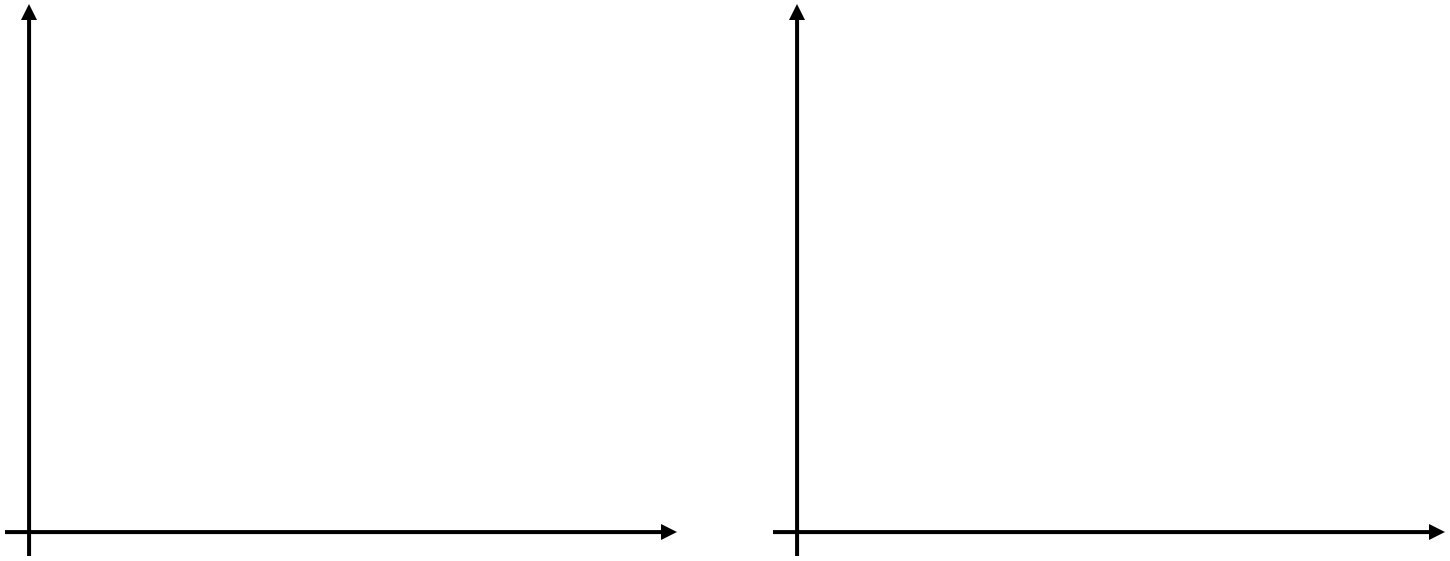
› Elasticity

(4) Functional forms

› $\beta_2 = \frac{relative\ change\ in\ Y}{change\ in\ X}$ for log-lin model

› $\beta_2 = \frac{change\ in\ Y}{relative\ change\ in\ X}$ for lin-log model

Practical examples for log-lin model is **Growth model**, while for the lin-log model is **Engel expenditure** model.



(4) Functional forms

Example of log-lin and its interpretation.

› $\ln \widehat{wage}_i \mathrel{+}= 1.5 + 3.30 educ_i$

› If education increase by one year, we expect wages to increase by $100 * \beta_2$ percent.

(4) Functional forms

Example of lin-log and its interpretation.

› $\widehat{wage}_i \mathrel{+}= 1.5 + 3.30 \ln educ_i$

› If education increase by one year, we expect wages to increase by $\frac{\beta_2}{100}$ unit.

(4) Functional forms

(3) Reciprocal model

Reciprocal model takes a form of

› $Y = \beta_1 + \beta_2(\frac{1}{X_i})$

Let's find the slope and elasticity.

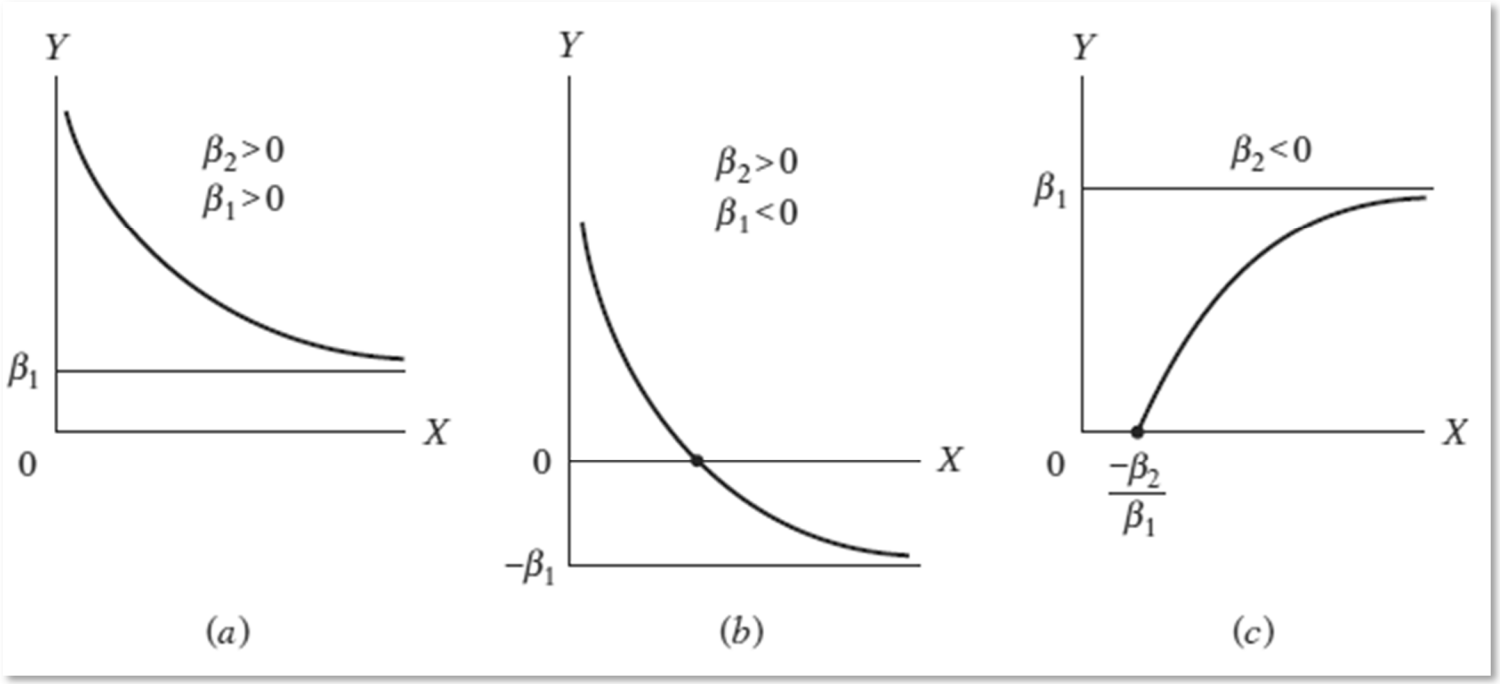
› Slope

› Elasticity

So, if β_2 is positive, the slope is negative, and vice versa.

(4) Functional forms

Practical examples for reciprocal models are **Child mortality rate** (vs. GNP) and **Phillips curve**.



(4) Functional forms

(4) Log-reciprocal model

Respectively, semi-log models consist of log-lin and lin-log model which take a form of

› $\ln Y = \beta_1 - \beta_2(\frac{1}{X_i})$

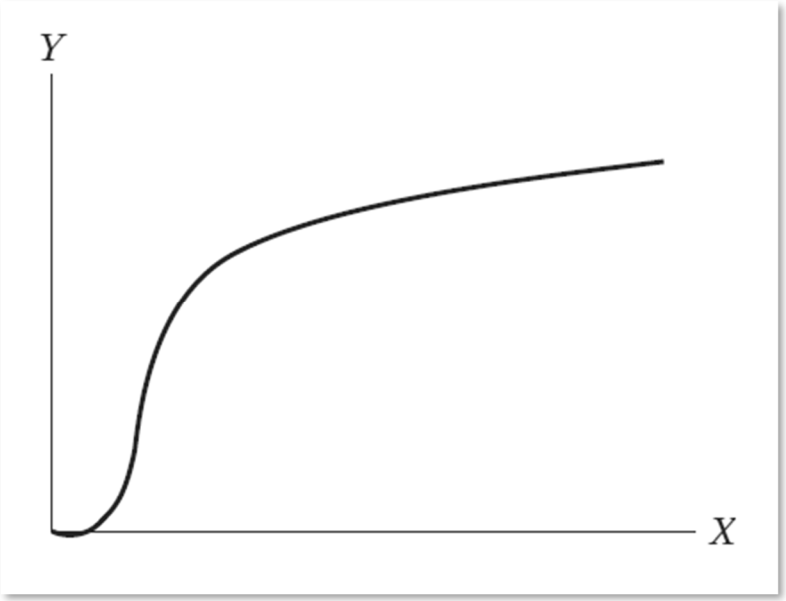
The graph is convex at first, then concave later. Find the slope and elasticity.

› Slope

› Elasticity

(4) Functional forms

Practical example for log-reciprocal model is short-run production.



Chapter 4

Multiple Linear Regression

Flow of study in this chapter

› Multiple Linear Regression Model

Now we add more independent variables into our model. While the estimation method remains the same, the formulae to calculate the estimators are totally different.

› Individual Testing

Like what we studied in the previous chapter, can we still test each coefficients' significance?

› Analysis of Variance (ANOVA)

When we try to jointly test multiple variables at the same time, the F-test is a lot more useful and easier, with some caveats in many situation. This part is to showcase concepts underlying F-test mainly and how we can apply on so many useful test.

Further reading can be found in Gujarati and Porter, Chapter 7-8.

(1) Estimation

If we add more independent variable(s) into our model to increase fitness to the model, the specification of the stochastic form and the SRF become

› $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$: Stochastic form

› $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$: SRF

Note that we are not going to use the notation X_{1i} to make our estimators number corresponded with the variable names.

Now $\hat{\beta}_2$ and $\hat{\beta}_3$ are known as '**partial slope coefficients**' since when we consider

› $\hat{\beta}_2$ represents the change in $\hat{Y}_i$ when X_{2i} increases for 1 unit, holding everything else constant.

› $\hat{\beta}_3$ represents the change in $\hat{Y}_i$ when X_{3i} increases for 1 unit, holding everything else constant.

(1) Estimation

If we add more independent variables into this model, the specification becomes

› $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i$

Let's focus on 2 independent variables model first, we follow the same procedure as when we did, minimizing the term $\sum \hat{u}_i^2$

› $\min_{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3} \sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)^2$

Skipping all the prove because we utilize the same logic of minimization of a function with calculus, we get

(1) Estimation

Coefficients

› $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3$

› $\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i}x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}$

› $\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i}x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}$

Variance

› $\text{var}(\hat{\beta}_1) = \sigma^2\left[\frac{1}{n} + \frac{\bar{X}_2^2\sum x_{3i}^2 + \bar{X}_3^2\sum x_{2i}^2 - 2\bar{X}_2^2\bar{X}_3^2\sum x_{2i}x_{3i}}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}\right]$

› $\text{var}(\hat{\beta}_2) = \sigma^2\left[\frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}\right] = \frac{\sigma^2}{\sum x_{2i}^2(1-r_{23}^2)}$

› $\text{var}(\hat{\beta}_3) = \sigma^2\left[\frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}\right] = \frac{\sigma^2}{\sum x_{3i}^2(1-r_{23}^2)}$

where r_{23} is the coefficient of correlation between X_2 and X_3.

The estimator of σ^2 is $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3}$

(2) Assumptions

(1) Linear in parameters.

(2) All X_i are independent of the error term u_i.

(3) Zero mean of error term $E(u_i|X_{2i}, X_{3i}) = 0$.

(4) Homoscedasticity or $\text{var}(u_i) = \sigma^2$.

(5) No autocorrelation or $\text{cov}(u_i, u_j) = 0$.

(6) $n > k$ and X_i must not all be the same.

(7) No specification bias.

(8) No exact collinearity between X_i.

(2) Assumptions

The 8th assumption becomes more significant in this model. Imagine that

› $X_{2i} = 2X_{3i}$

We then can turn this model into

› $Y_i = \hat{\beta}_1 + \hat{\beta}_2 2X_{3i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$ and then $\hat{\beta}_2 = \hat{\beta}_3$ so

› $Y_i = \hat{\beta}_1 + (\hat{\beta}_2 + 2\hat{\beta}_3)X_{3i} + \hat{u}_i$

which means that either X_{2i} or X_{3i} does not add any more information to this model.

(3) Adjusted Coefficient of Determination

When we add more and more independent variables into the model, the coefficient of determination is likely to increase (at least it will not decrease) from decreasing $\sum \hat{u}_i^2$.

Recall that when we define R^2 as

› $R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$ where $\sum y_i^2 = \sum (Y_i - \bar{Y})^2$

Now we define **adjusted R^2**, denoted as

› $\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2/(n-k)}{\sum y_i^2/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k}$

The word **adjusted** means that the R^2 is adjusted by the degrees of freedom associated with the sum of the squares entering the specification.

(3) Adjusted Coefficient of Determination

Comparison between R^2 and $\bar{R}^2$

(1) $0 \leq R^2 \leq 1$ but $\bar{R}^2$ can be negative (interpreted as 0).

(2) As k increases, R^2 is increasing but $\bar{R}^2$ may not.

(3) $\bar{R}^2 < R^2$

(4) Examples

Cobb-Douglas Production Function

Usual form of the Cobb-Douglas function is

› $Y = AK^{\alpha}L^{\beta}$

where Y is the value of output of an economy,

A is total factor productivity (TFP or sometimes simplified as production technology),

K is number of capital input,

L in number of labor input,

α and β is the output elasticity.

The stochastic form can be expressed as

› $Y_i = AK_i^{\alpha}L_i^{\beta}e^{u_i}$

Taking natural logarithm to enable linear estimation yields

› $\ln Y_i = \ln A + \alpha \ln K_i + \beta \ln L_i + u_i$

(4) Examples

Cobb-Douglas Production Function

The data obtained from manufacturing sector of all states in the US, represented for each observation i. The results of linear regression is as follows.

› $\ln \hat{Y}_i = 3.8876 + 0.5213 \ln K_i + 0.4683 \ln L_i$

$t = (9.8115) \quad (5.3803) \qquad (4.7342)$

$n = 51 \qquad R^2 = 0.9642 \qquad \bar{R}^2 = 0.9627$

We can test each estimator for its significance or we can also jointly test $\alpha + \beta$ to check returns to scale such as

› H_0: $\alpha + \beta = 1$ or constant returns to scale

› H_a: otherwise.

(4) Examples

Polynomial models

There are multiple economic models incorporating polynomial form, such as total cost and marginal cost, an example here is the effect of age to wage in the quadratic form. Given that wage is a function of age as follows (in the stochastic form).

› $w_i = \hat{\beta}_1 + \hat{\beta}_2 age_i + \hat{\beta}_3 age_i^2 + u_i$

where w_i is the value of output of an economy,
age_i is straightforward.
age_i^2 is the squares of age and
$\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ are the estimators.

An example of the estimation is

› $\widehat{w_i} = 3.73 + 0.298 age_i - 0.0061 age_i^2$

(4) Examples

As age_i^2 becomes higher, the square (negative) effect will be dominant.

(1) The t-test

With the normality assumption, we find that estimators are normally distributed with an unknown variance. Therefore, to test statistical significance, we need to rely on t statistics as follows.

› $t_{cal}(\beta_i) = \frac{\hat{\beta}_i - \beta_i}{\text{se}_{\hat{\beta}_i}} \sim t_{n-3}$

where $\hat{\beta}_i$ is the value of estimator,

β_i is the value that we would test against and

$\text{se}_{\hat{\beta}_i}$ is the standard error of the estimator.

Now that the d.f. is $n - 3$ if we have two independent variables in our estimation (3 estimators or 3 unknown). Testing procedures are the same.

(2) Example

Given the data collected at the beginning of this class, we hypothesize that shoe size (Y_i) is determined by both height (X_{2i}) and weight (X_{3i}).

First, we estimate the simple linear regression without including weight. The result is displayed here.

```
      Source |       SS           df       MS      Number of obs   =        66
-------------+----------------------------------   F(1, 64)        =    202.47
       Model |  254.563322         1  254.563322   Prob > F        =    0.0000
    Residual |  80.4669811        64  1.25729658   R-squared       =    0.7598
-------------+----------------------------------   Adj R-squared   =    0.7561
       Total |  335.030303        65  5.15431235   Root MSE        =    1.1213

------------------------------------------------------------------------------
          ss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |   .2369768   .0166543    14.23   0.000      .203706    .2702476
       _cons |   .3862496   2.778896     0.14   0.890    -5.165234    5.937733
------------------------------------------------------------------------------
```

(2) Example

Now we add the additional variable, weight, into the model. Here is the result.

```
      Source |       SS           df       MS      Number of obs   =        66
-------------+----------------------------------   F(2, 63)        =    114.95
       Model |  262.967055         2  131.483527   Prob > F        =    0.0000
    Residual |  72.0632484        63  1.14386109   R-squared       =    0.7849
-------------+----------------------------------   Adj R-squared   =    0.7781
       Total |  335.030303        65  5.15431235   Root MSE        =    1.0695

------------------------------------------------------------------------------
          ss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |   .2051871   .0197458    ______  ______     .1657283    .2446459
         wei |   .0384022   .0141679    ______  ______     .0100898    .0667145
       _cons |   3.384344    2.87211    ______  ______    -2.355109    9.123797
------------------------------------------------------------------------------
```

Let's perform the individual tests. We are going to test all the coefficients, one by one.

(2) Example

(1) State your hypothesis

Stating three hypotheses separately, which are

› For β_1 $H_0: \beta_1 = 0$
$H_a: \beta_1 \neq 0$

› For β_2 $H_0: \beta_2 = 0$
$H_a: \beta_2 \neq 0$

› For β_3 $H_0: \beta_3 = 0$
$H_a: \beta_3 \neq 0$

In this case, we are testing if each coefficient is significantly different from zero or not.

(2) Example

(2) Calculate test statistics

› For β_1 $t_{cal}(\beta_1) = \frac{\widehat{\beta}_1 - \beta_1}{\text{se}_{\widehat{\beta}_1}} =$

› For β_2 $t_{cal}(\beta_2) = \frac{\widehat{\beta}_2 - \beta_2}{\text{se}_{\widehat{\beta}_2}} =$

› For β_3 $t_{cal}(\beta_3) = \frac{\widehat{\beta}_3 - \beta_3}{\text{se}_{\widehat{\beta}_3}} =$

(3) Pick an α and state decision rules

Supposed that we pick $\alpha = 0.05$, now we are testing against zero, we can directly at the t-table for the critical values which are

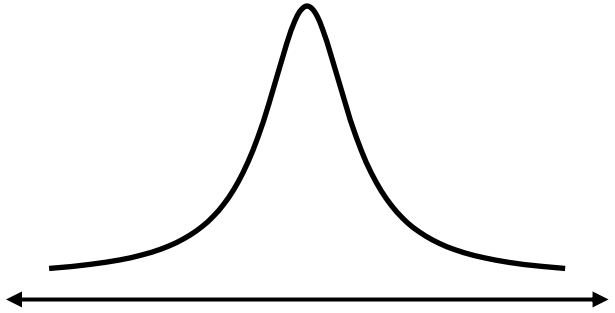
› $t_{lower} =$

› $t_{upper} =$

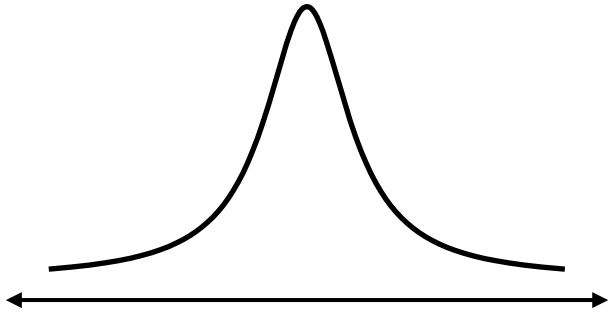
(2) Example

(4) Concluding the test results

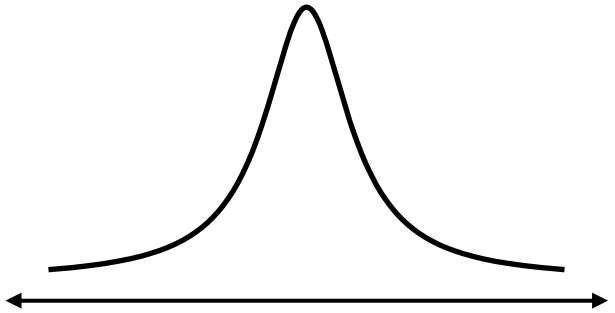
› For β_1



› For β_2



› For β_3



(1) Introduction

Now consider if we want to test parameters jointly under the same hypothesis, for instance,

› $H_0: \beta_2 = \beta_3 = 0$

› H_a: Not all the slope coefficients are simultaneously zero (otherwise).

This hypothesis sates that "β_2 and β_3 are jointly or simultaneously equal to zero". We **cannot** utilize ordinary t-test anymore, but rely on specific joint test or the **Analysis of Variance (ANOVA)**

The ANOVA test is a different approach. The focus is on dispersion of each part, namely the ESS and RSS. According to the study of variation from the mean in the coefficient of determination, we have

› $\sum \hat{y}_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \sum \hat{u}_i^2$

$\text{TSS} = \quad \text{ESS} \quad + \text{RSS}$

(1) Introduction

From the estimation of shoe size, we now look at another information printed from regression table to make use of our joint test.

Source	SS	df	MS
Model	262.967055	2	131.483527
Residual	72.0632484	63	1.14386109
Total	335.030303	65	5.15431235

When we perform the ANOVA, we are comparing between a part that the model can explain (ESS) versus a part it cannot (RSS), therefore the comparison is

$$\rightarrow \frac{\text{ESS}/df}{\text{RSS}/df} = \frac{\widehat{\beta}_2 \sum y_i x_{2i} + \widehat{\beta}_3 \sum y_i x_{3i}/(k-1)}{\sum \widehat{u}_i^2/(n-k)} \sim$$

(1) Introduction

Consider the ratio, from the assumption of $u_i \sim N(0, \sigma^2)$,

› $E\left(\frac{\sum \hat{u}_i^2}{n-k}\right) = \sigma^2$

and if we assume that our stated hypothesis $H_0: \beta_2 = \beta_3 = 0$ is true, then

› $\sum \hat{y}_i^2 = \sum \hat{u}_i^2$ or

› $\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} = 0$

or the effect of X_{2i} and X_{3i} is **trivial to the variation in Y_i, the only source of variation in Y_i is the error term u_i.**

Therefore, the ratio becomes smaller if $\beta_2 = \beta_3 = 0$. The decision rule for this test is

› $F_{cal} > F_\alpha(k-1, n-k)$ then we can reject H_0

(1) Introduction

(1) State your hypothesis

› $H_0: \beta_2 = \beta_3 = 0$

› H_a: otherwise.

(2) Calculate test statistics

Source	SS	df	MS
Model	262.967055	2	131.483527
Residual	72.0632484	63	1.14386109
Total	335.030303	65	5.15431235

› $F_{cal} = \frac{\text{ESS}/df}{\text{RSS}/df} = \frac{\text{ESS}/k-1}{\text{RSS}/n-k} =$

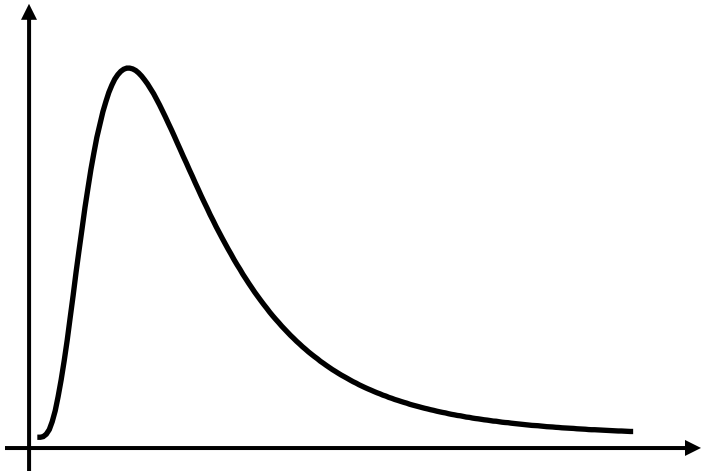
(1) Introduction

(3) Pick an α and state decision rules

› $\alpha =$

› $F_{upper,\alpha}(2,63) =$

(4) Conclude the test result



(1) Introduction

We then can compute the F statistics directly from R^2.

Given that our model has the $R^2 = 0.7849$

› $F_{cal} = \frac{\frac{ESS}{TSS}/(k-1)}{\frac{RSS}{TSS}/(n-k)} = \frac{R^2/(k-1)}{1-R^2/(n-k)} =$

which yields the same result compared to computing from the mean squares.

(2) The marginal contribution

If we estimate SRF with the same context, but we are not sure if we should add X_{3i} into the model or not, we have two different models.

› Excluding: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i}$

› Including: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$

The following test is to measure increment or contribution of added independent variable(s), in this case X_{3i}. Hence, the test highlights a comparison of whole models, instead of just a particular variable's significance.

The answer we are looking for from this test is that "**should the incremental variable(s) be added into the model?**"

(2) The marginal contribution

The F-statistics becomes

› $F_{cal} = \frac{ESS_{new} - ESS_{old}/(number\ of\ new\ regressors)}{RSS_{new}/(n-k_{new})}$

or if we measure from the R^2, it would be

› $F_{cal} = \frac{R^2_{new} - R^2_{old}/(number\ of\ new\ regressors)}{1-R^2_{new}/(n-k_{new})}$

Note that the R^2 method is applicable only when both models have the same Y_i, while the mean squares method can apply to other types of test, more on that later.

(2) The marginal contribution

Comparing between two models that we estimated, the first one with only height variable.

```
      Source |       SS           df       MS
-------------+----------------------------------
       Model |  254.563322         1  254.563322
    Residual |  80.4669811        64  1.25729658
-------------+----------------------------------
       Total |  335.030303        65  5.15431235
```

The second model, weight is added.

```
      Source |       SS           df       MS
-------------+----------------------------------
       Model |  262.967055         2  131.483527
    Residual |  72.0632484        63  1.14386109
-------------+----------------------------------
       Total |  335.030303        65  5.15431235
```

Perform the test to check that weight variable has any contribution.

(2) The marginal contribution

(1) State your hypothesis

› H_0: weight has no marginal contribution to the model.

› H_a: otherwise.

(2) Calculate the test statistics

› $F_{cal} = \frac{ESS_{new} - ESS_{old}/(number\ of\ new\ regressors)}{RSS_{new}/(n - k_{new})} =$

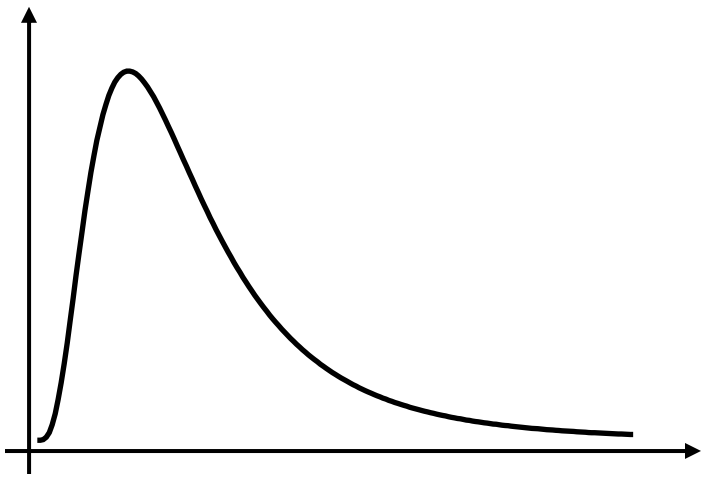
(2) The marginal contribution

(3) Pick an α and state decision rules

› $\alpha =$

› $F_{upper,\alpha}(1,63) =$

(4) Conclude the test result



The addition of variable X_{3i} or *weight*

Note that the F-test for the contribution is most of the time in line with the individual t-test of those added variable(s) since $t^2 \sim F(1,n)$.

(2) The marginal contribution

Selecting the most fitted model

› Choose a model with the highest F value, leading to highest $\bar{R}^2$.

› Absolute term of t value of added variable(s) is more than 1, which will also eventually lead to higher $\bar{R}^2$.

(3) Equality of two regression coefficients

Since the collected data provide interesting insight to this topic, let's shift to another data provided in the book. We have an ordinary demand model represented here.

› $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$

where Y_i is quantity demanded for a commodity

X_{2i} is its price

X_{3i} is consumer income

X_{4i} is consumer wealth

A question arises **if X_{3i} and X_{4i}**, income and wealth, **both are representing affordability in the same way or not.** We can answer this question with two methods

› Using t-test

› Using F-test with Restricted and Unrestricted models

(3) Equality of two regression coefficients

(3.1) Using t-test

› $t_{cal} = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{\text{se}_{(\hat{\beta}_3 - \hat{\beta}_4)}} \sim t_{n-k}$ where

› $\text{se}_{(\hat{\beta}_3 - \hat{\beta}_4)} = \sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2\text{cov}(\hat{\beta}_3, \hat{\beta}_4)}$

Example: Given an estimated of cubic cost function below,

• $\hat{Y}_i = 141.7667 + 63.4777X_i - 12.9615X_i^2 + 0.9396X_i^3$

$(6.3753) \quad (4.7789) \quad (0.9857) \quad (0.0591)$

$cov(\hat{\beta}_3, \hat{\beta}_4) = -0.0576 \quad R^2 = 0.9983 \;\; n = 10$

where Y_i is total cost and X_i is output.

(3) Equality of two regression coefficients

(1) State your hypothesis

› $H_0: \beta_3 = \beta_4$ or $\beta_3 - \beta_4 = 0$

› $H_a: \beta_3 \neq \beta_4$ or $\beta_3 - \beta_4 \neq 0$

(2) Calculate test statistics

› $t_{cal} = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{\text{se}_{(\hat{\beta}_3 - \hat{\beta}_4)}} =$

(3) Equality of two regression coefficients

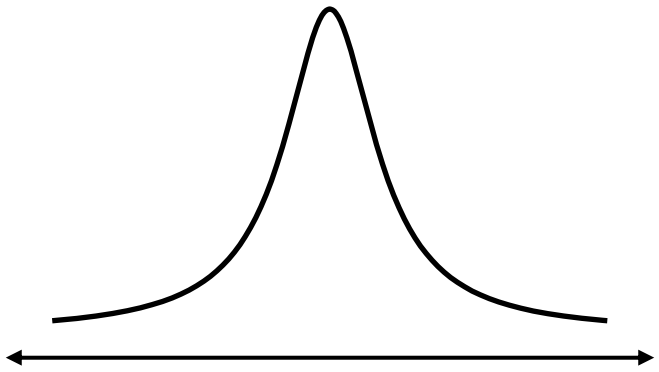
(3) Pick an α and state decision rules

› $\alpha =$

› $t_{lower} =$

› $t_{upper} =$

(4) Conclude the test result



(3) Equality of two regression coefficients

(3.2) Using F-test with Restricted and Unrestricted models

There are some certain models that economic theory suggest that the coefficients in a regression satisfy some linear equality restrictions. Consider a Cobb-Douglas here, where some notations are altered, which is already linearized

› $\ln Y_i = \ln \beta_1 + \beta_K \ln K_i + \beta_L \ln L_i + u_i$

β_K and β_L are the elasticity of capital and labor input respectively. We also know that addition of these two parameters reveals returns to scale.

Supposed that we want to test if there are constants returns to scale or not, we can hypothesize

› $H_0: \beta_K + \beta_L = 1$

› $H_\mathrm{a}: \beta_K + \beta_L \neq 1$

This is a test to see **if we can make sure that the result of addition of two parameters is 1 or not.**

(3) Equality of two regression coefficients

However, we can "**restrict**" the regression or **force** that returns to scale into the regression directly that either

› $\beta_K + \beta_L = 1$ / $\beta_K = 1 - \beta_L$ / $\beta_L = 1 - \beta_K$

If we rewrite the Cobb-Douglas function with the restriction here, it becomes

› $\ln Y_i = \ln \beta_1 + \beta_K \ln K_i + (1 - \beta_K) \ln L_i + u_i$

where $\frac{Y_i}{L_i}$ is output per labor and $\frac{K_i}{L_i}$ is capital per labor.

You may notice that there is one less coefficient to estimate. This is known as "**Restricted Least Square**" (RLS).

(3) Equality of two regression coefficients

Therefore, the test we are about to perform is a comparison between

› Unrestricted model:

$$\ln Y_i = \ln \beta_1 + \beta_K \ln K_i + \beta_L \ln L_i + u_i$$

› Restricted model:

$$\ln \left(\frac{Y_i}{L_i}\right) = \ln \beta_1 + \beta_K \ln\left(\frac{K_i}{L_i}\right) + u_i$$

which is a test to see if **the restriction imposed is valid or not**. We can set up the same hypothesis.

› H_0: $\beta_K + \beta_L = 1$ or the restriction is valid

› H_a: $\beta_K + \beta_L \neq 1$ or the restriction is not valid

(3) Equality of two regression coefficients

After that we can compute the F statistics, which is

› $F_{cal} = \frac{RSS_R - RSS_{UR}/m}{RSS_{UR}/(n-k_{UR})} \sim F_{(m,\, n-k_{UR})}$

where R and UR indicates the value from restricted and unrestricted model respectively, m is the number of linear restriction (1 for this case). We can also derive F_{cal} from R^2 as well.

› $F_{cal} = \frac{(R^2_{UR} - R^2_R)/m}{(1-R^2_{UR})/(n-k_{UR})} \sim F_{(m,\, n-k_{UR})}$

Note that $R^2_{UR} \geq R^2_R$ and the variable on the left-hand side in both models is not the same so **they are not comparable**. In this case, need to rely on calculating F statistics from RSS instead.

(3) Equality of two regression coefficients

Example: Supposed we have a result of

› Unrestricted model:

$$\ln \hat{Y}_i = -1.6524 + 0.8460 \ln K_i + 0.3397 \ln L_i$$

$$R^2_{UR} = 0.9951 \qquad RSS_{UR} = 0.0136$$

› Restricted model:

$$\ln \left(\frac{\widehat{Y_i}}{L_i}\right) = -0.4947 + 1.0153 \ln\left(\frac{K_i}{L_i}\right)$$

$$R^2_R = 0.9777 \qquad RSS_R = 0.0166$$

Both models has 20 observations. Perform the test from both RSS and R^2 with the hypothesis here.

(3) Equality of two regression coefficients

(1) State your hypothesis

› $H_0: \beta_K + \beta_L = 1$ or the restriction is valid

› $H_a: \beta_K + \beta_L \neq 1$ or the restriction is not valid

(2) Calculate test statistics

› $F_{cal} = \frac{RSS_R - RSS_{UR}/m}{RSS_{UR}/(n-k_{UR})} \sim F_{(m,\, n-k_{UR})} =$

Just to prove that R^2 is not comparable, we can also try to calculate the test statistics as follows.

› $F_{cal} = \frac{(R^2_{UR} - R^2_R)/m}{(1-R^2_{UR})/(n-k_{UR})} =$

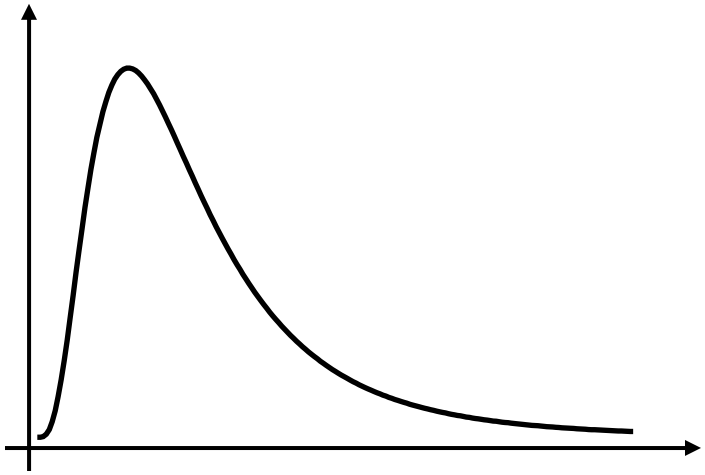
(3) Equality of two regression coefficients

(3) Pick an α and state decision rules

› $\alpha =$

› $F_{upper,\alpha}(1,17) =$

(4) Conclude the test result



(4) General F-testing

As we can see that the F test can be utilize in multiple ways of comparing two competing models, we can also set up the test for "**constrained**" and "**unconstrained**" models as seen in the example below. First, we start from the "unconstrained" model

› $\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \beta_4 \ln X_{4i} + \beta_5 \ln X_{5i} + u_i$

where Y_i is per capita consumption of chicken

X_{2i} is real disposable per capita income

X_{3i} is real retail price of chicken

X_{4i} is real retail price of pork

X_{5i} is real retail price of beef

β_4 and β_5 refer to cross-price elasticity of chicken and pork and chicken and beef.

(4) General F-testing

Thus,

› If $\beta_4 / \beta_5 > 0$, chicken and pork / beef are substitutable products.

› If $\beta_4 / \beta_5 < 0$, chicken and pork / beef are complementary products.

› If $\beta_4 / \beta_5 = 0$, chicken and pork / beef are unrelated products.

If we want to test that how chicken and pork / beef are related, we can set up a hypothesis as such.

› $H_0: \beta_4 = \beta_5 = 0$

› H_a: otherwise

If we cannot reject the null hypothesis, the model become "**constrained**" as

› $\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$

(4) General F-testing

The test is similar to the restricted and unrestricted model. Moreover, we can use the R^2 approach here since the Y_i is the same for both the constrained and unconstrained model. Test statistics can be calculated as follows.

› $F_{cal} = \frac{RSS_R - RSS_{UR}/m}{RSS_{UR}/(n-k_{UR})} \sim F_{(m,\, n-k_{UR})}$

› $F_{cal} = \frac{(R^2_{UR} - R^2_R)/m}{(1-R^2_{UR})/(n-k_{UR})} \sim F_{(m,\, n-k_{UR})}$

Note that we use the same notations compared to the restricted model testing so that we don't have to create another complication.

Hence, m is number of coefficient constrained. R and UR denote values from constrained and unconstrained model respectively.

(4) General F-testing

Example: Given the regression results are as follows (n=23),

› Unconstrained model:

$$\ln \hat{Y}_i = 2.19 + 0.34 \ln X_{2i} - 0.50 \ln X_{3i} + 0.15 \ln X_{4i} + 0.09 \ln X_{5i}$$

$$R^2_{UR} = 0.9823$$

› Constrained model:

$$\ln \hat{Y}_i = 2.03 + 0.45 \ln X_{2i} - 0.38 \ln X_{3i}$$

$$R^2_R = 0.9801$$

(1) State your hypothesis

› $H_0: \beta_4 = \beta_5 = 0$

› H_a: otherwise

(4) General F-testing

(2) Calculate test statistics

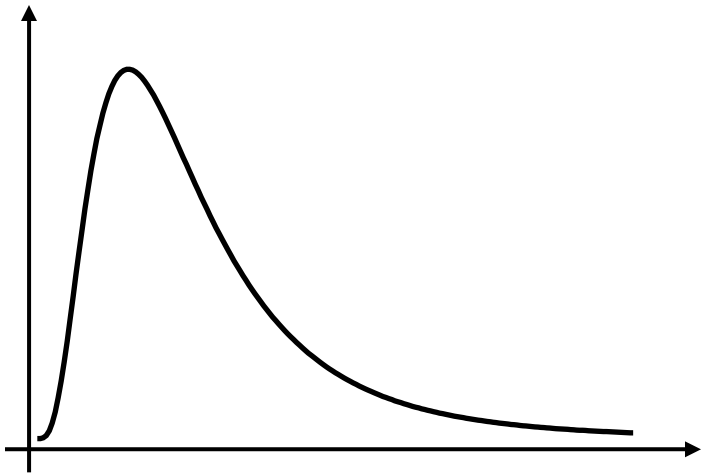
› $F_{cal} = \frac{(R_{UR}^2 - R_R^2)/m}{(1-R_{UR}^2)/(n-k_{UR})} =$

(3) Pick an α and state decision rules

› $\alpha =$

› $F_{upper,\alpha}(2,18) =$

(4) Conclude the test result



(5) Structural change: Chow Test

Many event in our history, we had sometimes faced with "**structural change**" , a major shift in the basic how market works and how agents interact. The shift can be both internal or external.

In Thailand, we may consider opposing party getting elected and alter policy set. Or major flood in 2011 may alter how business decision of location established.

In econometrics term, it refers to a change in parameter, either the intercept or slopes, due to socio-political shift.

Consider when there is a pool of data, ranging a long periods of time, estimation the whole data can be misleading as it reveals the average values of estimator within that range.

(5) Structural change: Chow Test

The Chow Test attempts tackling this problem, to test that **if there is any structural shift sometime in our data or not**.

For this example, we consider saving-income relationship between the second oil shock in the 1980s. Data cover 1970-1995. It is assumed that we can separate between pre and post 1982, when the unemployment rate is at the highest. The setup is as follows.

SRF from 1970-1981 (Eq. 1)

› $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \qquad n_1 = 12$

SRF from 1982-1995 (Eq. 2)

› $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \qquad n_2 = 14$

SRF from pooled data 1970-1995 (Eq. 3)

› $Y_t = \beta_1 + \beta_2 X_t + u_t \qquad n = (n_1 + n_2) = 26$

where Y is savings and X is income. The slope parameter represents

(5) Structural change: Chow Test

Assumptions

(1) $u_{1t} \sim N(0, \sigma^2)$ and $u_{2t} \sim N(0, \sigma^2)$

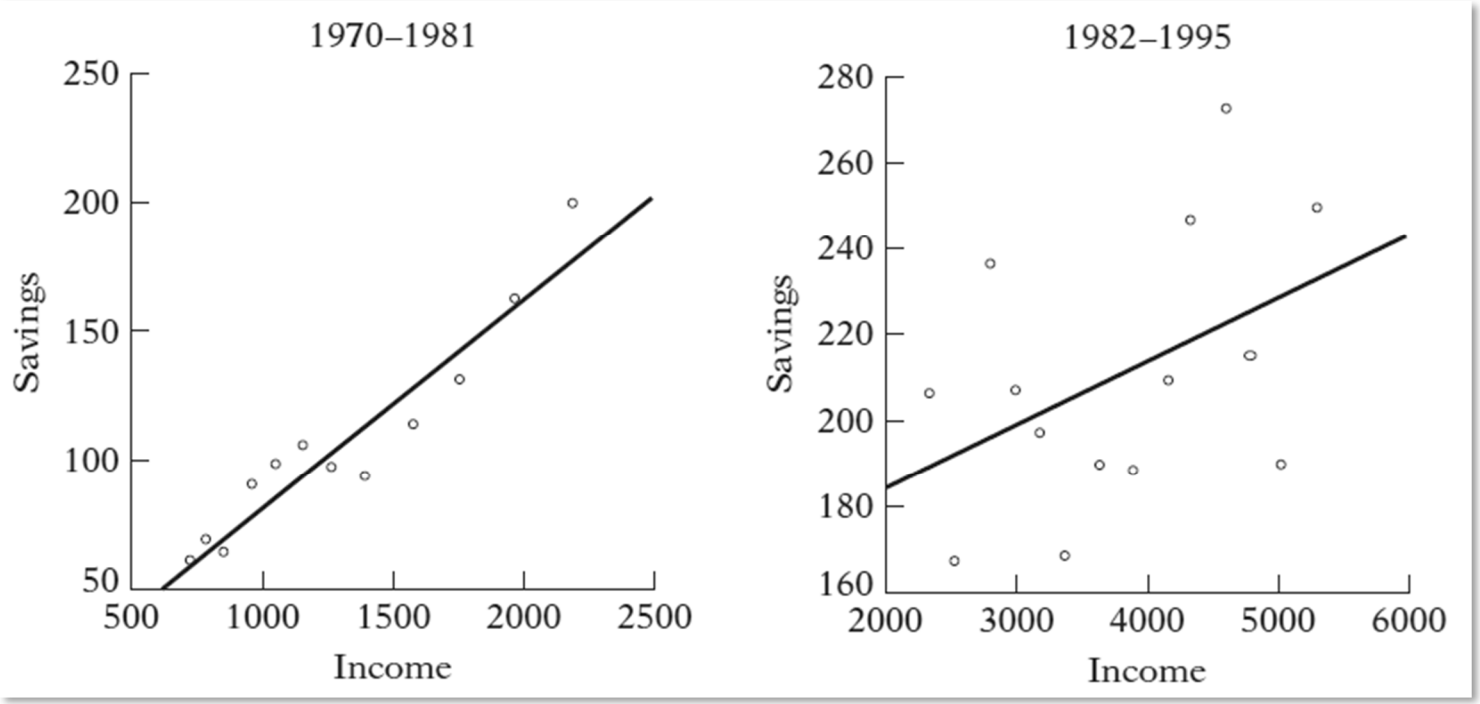
(2) u_{1t} and u_{2t} are independently distributed.

The Chow Test still relies on ANOVA, comparing a restricted model with an unrestricted model.

In this case, the **restricted** one assumes $\lambda_1 = \gamma_1$ and $\lambda_2 = \gamma_2$, or the Eq.3, while the **unrestricted** one allows different values of λ and γ.

(5) Structural change: Chow Test

Plotting the graphs above seems to suggest that there is a structural change in people saving behavior, due to the slope change.



(5) Structural change: Chow Test

Procedures

(1) State a hypothesis

› H_0: $\lambda_1 = \gamma_1$ and $\lambda_2 = \gamma_2$

› H_a: otherwise

(2) Estimate Eq. 3 or the restricted model. Retrieve

› $RSS_R = RSS_3$ or **restricted sum of squares.**

› D.f. is straightforwardly $n_1 + n_2 - k$

(3) Estimate Eq. 1 and Eq.2. Retrieve

› $RSS_{UR} = RSS_1 + RSS_2$ or **unrestricted sum of squares.**

› D.f. for this model is $n_1 + n_2 - 2k$

Note that k is the number of parameter estimated.

(5) Structural change: Chow Test

(4) Calculate F statistics by

› $F_{cal} = \frac{RSS_R - RSS_{UR}/k}{RSS_{UR}/(n_1+n_2-2k)} \sim F_{(k,\, n_1+n_2-2k)}$

(5) Concluding the results,

› We **can** reject the null hypothesis if $F_{cal} > F_{upper}$, meaning that it is either $\lambda_1 \neq \gamma_1$ or $\lambda_2 \neq \gamma_2$ or both. Therefore, it implies a structural change within these two periods.

› We **cannot** reject the null hypothesis if $F_{cal} < F_{upper}$, meaning that $\lambda_1 = \gamma_1$ and $\lambda_2 = \gamma_2$. So, we cannot make sure that there is a structural change within this period.

(5) Structural change: Chow Test

Example Given the regression results are,

› **Eq. 1**: $\hat{Y}_t = 1.0161 + 0.0803X_t$

$R_1^2 = 0.9021$ $RSS_1 = 1{,}785.032$ $n_1 = 12$

› **Eq. 2**: $\hat{Y}_t = 153.4947 + 0.0148X_t$

$R_2^2 = 0.2971$ $RSS_2 = 10{,}005.22$ $n_2 = 14$

› **Eq. 3**: $\hat{Y}_t = 62.4226 + 0.0376X_t$

$R_3^2 = 0.7672$ $RSS_3 = 23{,}248.30$ $n_3 = 26$

(1) State a hypothesis

› H_0: $\lambda_1 = \gamma_1$ and $\lambda_2 = \gamma_2$

› H_a: otherwise

(5) Structural change: Chow Test

(2) Calculate the test statistics

› $RSS_R =$

› $RSS_{UR} =$

Then, the F statistics,

› $F_{cal} = \frac{RSS_R - RSS_{UR}/k}{RSS_{UR}/(n_1+n_2-2k)} =$

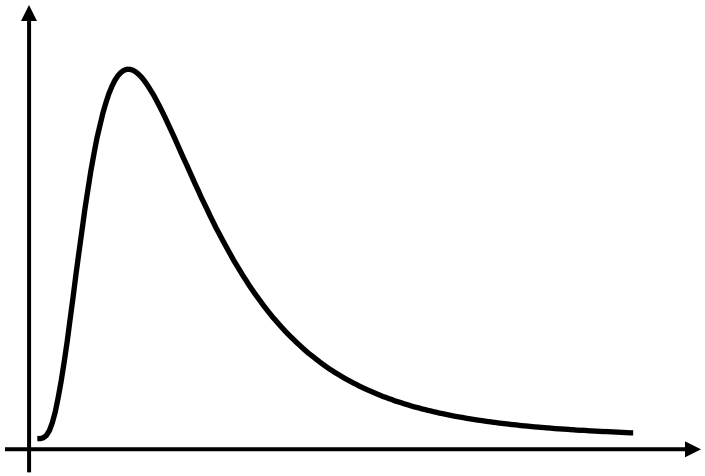
(3) Pick an α and state decision rules

› $\alpha =$

› $F_{upper,\alpha}(2,22) =$

(5) Structural change: Chow Test

(4) Conclude the test result



(5) Structural change: Chow Test

Limitations of The Chow Test

› The assumption of u_{1t} and u_{2t} being independently distributed should (or must) be tested, see the further explanation on the test on page 258.

› We need to assume that we know exactly where or when is the structural break point. If the speculation is not completely correct, the result of the test maybe controversial.

› The Chow Test only reveals difference between two models, which means that there might be a difference either in the intercept, slope or both. (The next topic of dummy variables will address this issue)

Chapter 5

Dummy variable

A problem of Chow test

Recalling the Chow Test, a test for structural change, let's see all the possibilities from ex-ante and post crisis.

› $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}$ $\quad n_1 = 12$

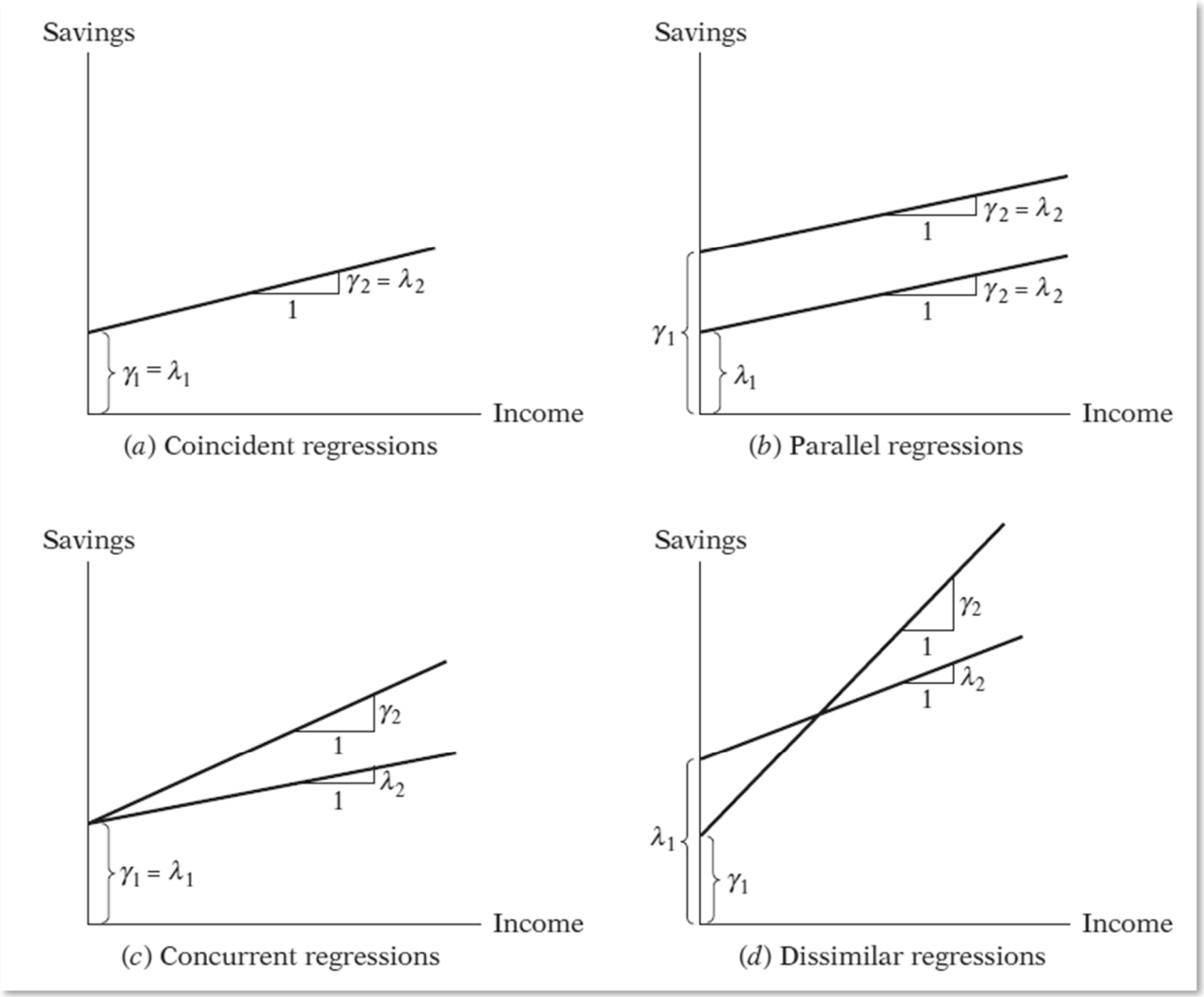
› $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}$ $\quad n_2 = 14$

› $Y_t = \beta_1 + \beta_2 X_t + u_t$ $\quad n = (n_1 + n_2) = 26$

As discussed earlier, the major for overall test is that F-test is usually very general, when a null hypothesis is rejected.

Though a null hypothesis is rejected, we still do not know what and how, in this case, λ_1, γ_1 and λ_2, γ_2 are different. We would know if we keep nesting the F-test, which is way too much work.

A problem of Chow test



A problem of Chow test

If we can include a variable that separates ex-ante and post crisis period in a single equation, that would be ideal because we can see a difference, or no difference between pre and post crisis. We can see its significance with a single t-test.

Not only that, if we can include another variable that can capture the slope for pre and post crisis, that is also very helpful since we can test its significance with a t-test as well.

We are gradually going to implement this concept step-by-step.

(1) ANOVA model

So far, we have only dealt with continuous variables (weight, height, income, price, quantity, temperature, etc.) for both dependent and independent variables.

A natural problem arises since we know that there are so many real-world variables that is '**qualitative**'. A basic and most upfront example is gender. Consider our show size model.

› $ss_i = \beta_1 + \beta_2 sex_i + u_i$

where ss_i is shoe size and sex_i is a binary variable, therefore there are only two possible encodings either

› sex_i = otherwise

› sex_i = female

This model is called ANOVA model, or a model containing only quantitative variables or **dummy variable**.

(1) ANOVA model

Given that the result of this regression model is

› $\widehat{ss}_i = 41.833 - 3.583 sex_i$

First, we look at how this result should be interpreted by considering the expected value. Note that we encode $sex_i = 0$ for otherwise and $sex_i = 1$ for female

› $E(\widehat{ss}_i | sex_i = 0) =$

› $E(\widehat{ss}_i | sex_i = 1) =$

(1) ANOVA model

If we plot each expected value, we see a difference between each group on this graph.

Data for the estimation are in this table. To distinguish gender, we only need one dummy variable to accommodate this difference. To be precise, we need only $n-1$ dummy variable(s) to incorporate n groups of the sample.

(2) Three categories

Now let's assume that we will use chosen gender instead of biological sex, so we have 3 groups which are male, female, and LQBTQ+. Only one option can be chosen among these. Given that

› $D_{2i} = 0$ for otherwise ; $D_{2i} = 1$ for female

› $D_{3i} = 0$ for otherwise ; $D_{3i} = 1$ for LGBTQ+

The estimated model becomes

› $\widehat{ss}_i = \hat{\beta}_1 + \hat{\beta}_2 D_{2i} + \hat{\beta}_3 D_{3i}$

(2) Three categories

Find the expected value and interpretation for

› $E(\widehat{ss}_i | D_{2i} = 0; D_{3i} = 0) =$

› $E(\widehat{ss}_i | D_{2i} = 0; D_{3i} = 1) =$

› $E(\widehat{ss}_i | D_{2i} = 1; D_{3i} = 0) =$

› $E(\widehat{ss}_i | D_{2i} = 1; D_{3i} = 1) =$

(2) Three categories

Given that the result of this regression model is

› $\widehat{ss}_i = 41.833 - 2.751D_{2i} + 0.63D_{3i}$

Plot each group onto this graph.

(3) Two dummy variables

Let's say we now have 2 quantitative variables, sex and area where

› $D_{2i} = 0$ for otherwise ; $D_{2i} = 1$ for female

› $D_{3i} = 0$ for otherwise ; $D_{3i} = 1$ for Bangkokian

The estimated model becomes

› $\widehat{ss}_i = \hat{\beta}_1 + \hat{\beta}_2 D_{2i} + \hat{\beta}_3 D_{3i}$

(3) Two dummy variables

Find the expected value and interpretation for

› $E(\widehat{ss}_i | D_{2i} = 0; D_{3i} = 0) =$

› $E(\widehat{ss}_i | D_{2i} = 0; D_{3i} = 1) =$

› $E(\widehat{ss}_i | D_{2i} = 1; D_{3i} = 0) =$

› $E(\widehat{ss}_i | D_{2i} = 1; D_{3i} = 1) =$

(3) Two dummy variables

Given that the result of this regression model is

› $\widehat{ss}_i = 41.833 - 3.182D_{2i} + 1.54D_{3i}$

Plot each group onto this graph.

(4) ANCOVA model

Regression models containing a mixture of both types of variables are called **ANCOVA models**. (Analysis of covariance)

Let's go back to our real sample of the show size model, now we include height (a quantitative continuous variable) and gender (a qualitative discrete variable) in this model as follows.

› $ss_i = \beta_1 + \beta_2 hei_i + \beta_3 sex_i + u_i$

where ss_i is shoe size, hei_i is height and sex_i is a binary variable as usual.

For this ANCOVA model, we have both quantitative and qualitative variables included. Each of them determines shoe size in a different way. Therefore, we need to interpret both height and gender that coexist in the same model.

(4) ANCOVA model

Given that the result of this regression model is

› $\widehat{ss}_i = 12.017 + 0.172hei_i - 1.456sex_i$

Now consider each case when

› $E(\widehat{ss}_i|sex_i = 0) =$

› $E(\widehat{ss}_i|sex_i = 1) =$

(4) ANCOVA model

If we plot each expected value, we see a difference between each group on this graph.

(1) Dummy and dummy

Dummy variables can be crossed to study differential effect from two or more dummies stacked together. Consider the following example of the same equation, instead we add a cross-product term here.

› $ss_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 (D_{2i} D_{3i}) + u_i$

where

› $D_{2i} = 0$ for otherwise ; $D_{2i} = 1$ for female

› $D_{3i} = 0$ for otherwise ; $D_{3i} = 1$ for Bangkokian

β_4 represents **additional effect**. The term is called **interaction dummy**, effect of the two attributes considered individually.

Therefore, when both dummies or either D_{2i} or D_{3i} is zero, β_4 will be zero. This coefficient can be tested only the case of **Bangkokian female** if there is any additional significant effect.

(2) Dummy and continuous variable

A dummy variable can be crossed with another continuous variable as well. Given that

› $ss_i = \beta_1 + \beta_2 hei_i + \beta_3 sex_i + \beta_4 (hei_i sex_i) + u_i$

Let's find the expected value from estimated model.

› $E(\widehat{ss}_i | sex_i = 0) =$

› $E(\widehat{ss}_i | sex_i = 1) =$

(3) Example

Now all of the concepts are explained, let's take a look at our example from the estimation results. We consider the basic models.

1st model: › $ss_i = \beta_1 + \beta_2 hei_i + u_i$ all observation : n=66

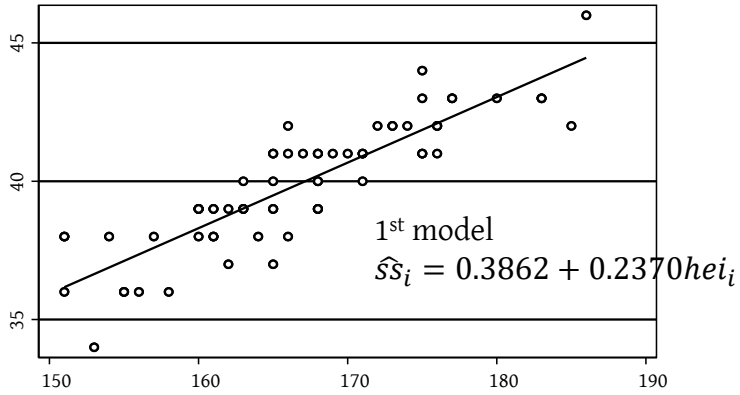
2nd model: › $ss_i = \beta_1 + \beta_2 hei_i + u_i$ male observation only : n=30

3rd model: › $ss_i = \beta_1 + \beta_2 hei_i + u_i$ female observation only : n=36

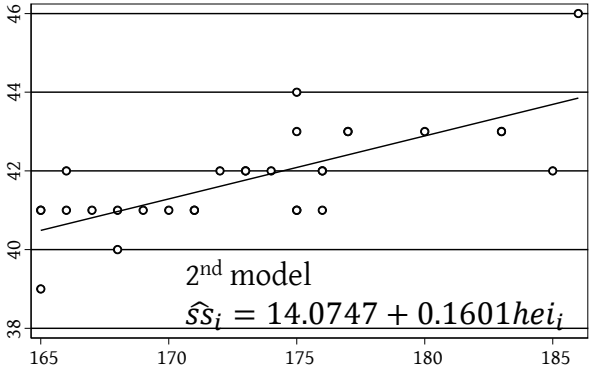
The results are listed here.

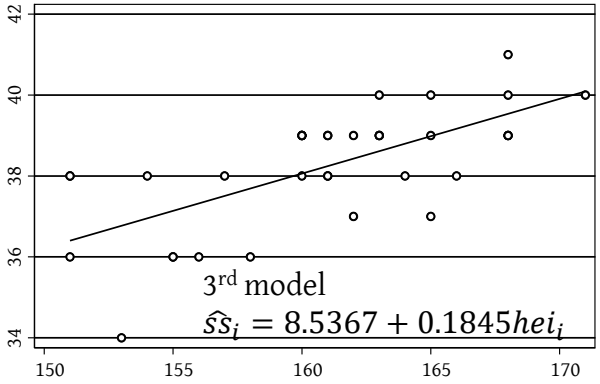
Model		Coefficient	P-value	R^2	$\bar{R}^2$
1st model	$\hat{\beta}_1$	0.3862	0.890	0.7598	0.7561
	$\hat{\beta}_2$	0.2370	0.000		
2nd model	$\hat{\beta}_1$	14.0747	0.008	0.5329	0.5162
	$\hat{\beta}_2$	0.1601	0.000		
3rd model	$\hat{\beta}_1$	8.5367	0.140	0.4487	0.4324
	$\hat{\beta}_2$	0.1845	0.000		

(3) Example



Plotting all the fitted regression line and scatter plot here.





(3) Example

To test if there is any difference between sex, we introduce two more models, incorporating a dummy and an interaction term.

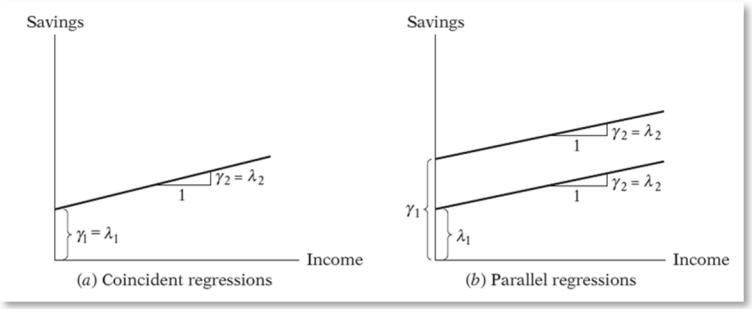
4th model: › $ss_i = \beta_1 + \beta_2 hei_i + \beta_3 sex_i + u_i$

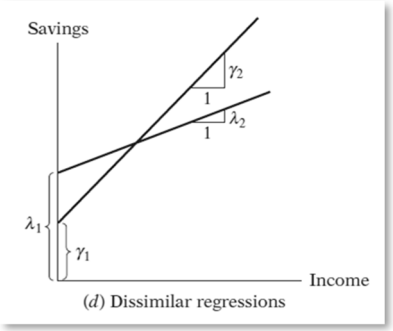
5th model: › $ss_i = \beta_1 + \beta_2 hei_i + \beta_3 sex_i + \beta_4 (sex_i hei_i) + u_i$

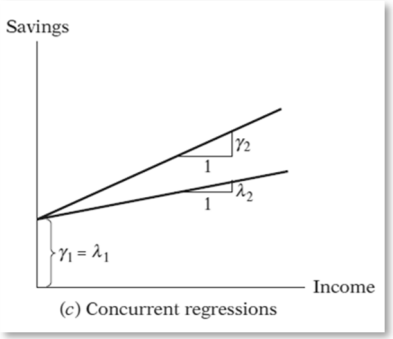
The results are listed here.

Model		Coefficient	P-value	R^2	$\bar{R}^2$
4th model	$\hat{\beta}_1$	12.0167	0.003	0.8061	0.8000
	$\hat{\beta}_2$	0.1720	0.000		
	$\hat{\beta}_3$	-1.460	0.000		
5th model	$\hat{\beta}_1$	14.0747	0.013	0.8070	0.7977
	$\hat{\beta}_2$	0.1601	0.000		
	$\hat{\beta}_3$	-5.5380	0.468		
	$\hat{\beta}_4$	0.0244	0.592		

(3) Example







In the 4th model, the dummy is used to test that the intercept of two groups are significantly different or not.

Meanwhile, in the 5th model we test that **both** the intercepts and slopes are significantly different **simultaneously** or not.
It turns out that they do not.

We can create another model that only test the slopes difference. This will be named as the 6th model.

Note: the pictures' purpose is only to illustrate the difference.

(3) Example

6th model: › $ss_i = \beta_1 + \beta_2 hei_i + \beta_3 (sex_i hei_i) + u_i$

The result is in the table below

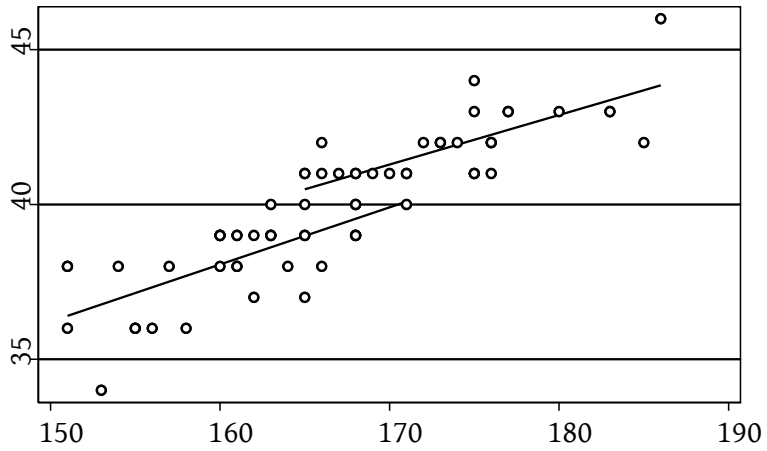
Model		Coefficient	P-value	R^2	$\bar{R}^2$
	$\hat{\beta}_1$	11.1802	0.004		
6th model	$\hat{\beta}_2$	0.1768	0.000	0.8054	0.7992
	$\hat{\beta}_3$	-0.0086	0.000		

This means that for female, the slope is a little bit less steep.

Now in this model, we can see that all the coefficients are statistically significant, as in the 4th model. The question is which model that we rely on.

The first thing to notice is that the 4th model has the highest value of $\bar{R}^2$.

(3) Example



This second thing, which is a lot more important, is that realistically, male and female should have different starting point and height should not affect them differently.

There is no reason to believe that the 6th model is true since there might hardly be any evidence suggesting alternating height affects shoe size differently. Male and female shoe size that are proportionally different makes a lot more sense.

See this plot to further elaborate my argument.

Chapter 6

Relaxing assumptions

Flow of study in this chapter

For this chapter, we are going to relaxing some assumptions we imposed since we estimate the first model. These topics will be covered: (1) multicollinearity (2) heteroscedasticity
(3) autocorrelation. For each section, we will explore

› The nature of assumption

› Effects on the estimated coefficients and variance, also standard error.

› How to detect the problem.

› What are remedial measure(s).

Further reading can be found in Gujarati and Porter, Chapter 10-13.

(1) The nature

This is a simplified version, compared to the book, of multicollinearity problem. Let λ_i be a constant, consider the following argument when X_{2i} and X_{3i} are linearly correlated perfectly, or **perfect multicollinearity**.

› $\lambda_2 X_{2i} + \lambda_3 X_{3i} = 0$

when $\lambda_i \neq 0$ for all i simultaneously. Another relation that describe a non-perfect collinearity, or **multicollinearity**, is

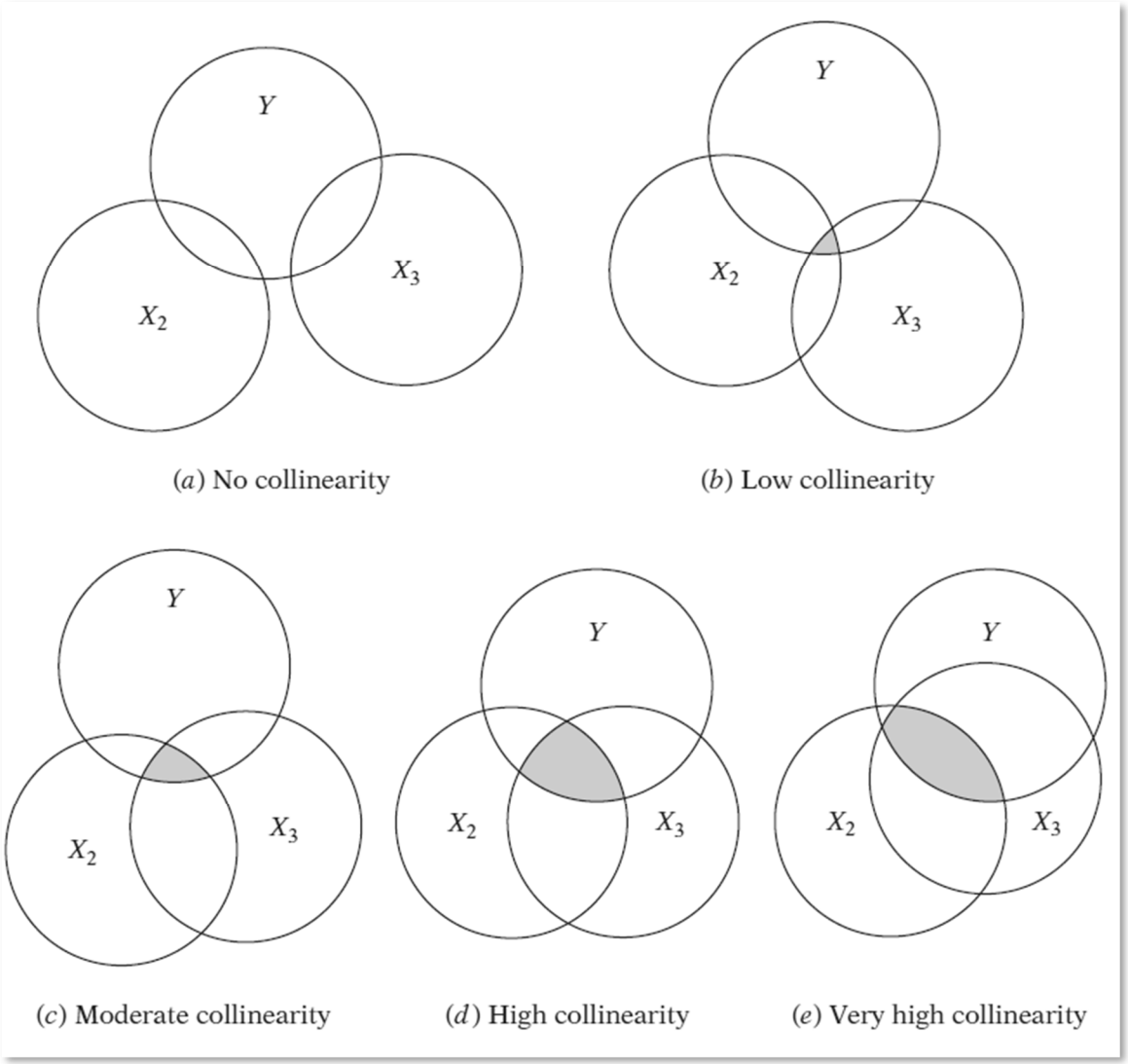
› $\lambda_2 X_{2i} + \lambda_3 X_{3i} + v_i = 0$

where v_i is a stochastic error term. Now assumed that $\lambda_2 \neq 0$ then,

› $X_{2i} = -\frac{\lambda_3}{\lambda_2} X_{3i}$ for the first equation.

› $X_{2i} = -\frac{\lambda_3}{\lambda_2} X_{3i} - v_i$ for the second equation.

(1) The nature


Y
X2
X3
(a) No collinearity
Y
X2
X3
(b) Low collinearity
Y
X2
X3
(c) Moderate collinearity
Y
X2
X3
(d) High collinearity
Y
X2
X3
(e) Very high collinearity

(1) The nature

Precisely taken from the book, here are some causes of multicollinearity.

› ***Data collection method*** may limit range of values taken in the independent variables.

› ***Constraints on the model or in the population being sampled.*** E.g. income and house size tend to be correlated.

› ***Model specification.*** E.g. including a polynomial term especially when the range of X is small.

› ***Overdetermined model*** or when k is larger than n.

› ***Common trend.*** E.g. a time-series data consist of consumption expenditure, income, wealth, and number of population.

(1) The nature

Looking from another perspective apart from stated above, multicollinearity is seen particularly as sampling problem, not a problem on a population, since when we postulate population regression, X variables included in a model have a separate or independent influence on Y.

Meaning that, as Goldberger coined the term, this may be considered as **micronumerosity** problem when our sampling may not be “rich” enough to capture X variability.

By the way, micronumerosity refers to the problem of small sample size.

Another important note is that multicollinearity is **much more common** in cross-sectional data compared to time-series data, in which autocorrelation is much more common.

(2) Effects on estimation

1. Perfect collinearity

Recall that the estimated coefficients are

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

If we assumed that $X_{3i} = \lambda X_{2i}$, replacing this into $\hat{\beta}_2$, we get,

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\lambda^2 \sum x_{2i}^2) - (\lambda \sum y_i x_{2i})(\lambda \sum x_{2i}^2)}{(\sum x_{2i}^2)(\lambda^2 \sum x_{2i}^2) - \lambda^2 (\sum x_{2i}^2)^2} = \frac{0}{0}$$

(2) Effects on estimation

Example: Consider the weight model with height (hei_i) as a regressor, now let's create a perfectly correlated variable of height * 2 (defined as $hei2_i$). The model will be

› $wei_i = \beta_1 + \beta_2 sex_i + \beta_3 hei_i + \beta_4 hei2_i + u_i$

Throwing this model into STATA, we have the regression result as follows. We can see that STATA automatically rejects, or omits, one of these variables immediately because $\hat{\beta}_4$ cannot be estimated.

```
------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -5.383607   3.396954    -1.58   0.118    -12.16779    1.400578
         hei |   .5914144    .206873     2.86   0.006     .1782605    1.004568
        hei2 |           0  (omitted)
       _cons |  -36.09017    35.8754    -1.01   0.318    -107.7383    35.55795
------------------------------------------------------------------------------
```

(2) Effects on estimation

2. Multicollinearity

Given that $X_{3i} = \lambda X_{2i} + \nu_i$, replacing this into $\hat{\beta}_2$, we get,

- $\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\lambda^2 \sum x_{2i}^2 + \sum \nu_i^2) - (\lambda \sum y_i x_{2i} + \sum y_i \nu_i)(\lambda \sum x_{2i}^2)}{(\sum x_{2i}^2)(\lambda^2 \sum x_{2i}^2 + \sum \nu_i^2) - \lambda^2 (\sum x_{2i}^2)^2}$

If ν_i is approaching zero, the more this will be closer to perfect multicollinearity.

We are going to skip lots of proof to get to the conclusion what are affected as follows.

(2) Effects on estimation

1. Coefficients estimated are still BLUE.

2. Large variances and covariances.

Recall that the variance of estimated coefficients can be written into this form.

› $\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2(1-\text{r}_{23}^2)}$

› $\text{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2(1-\text{r}_{23}^2)}$

where r_{23}^2 is the coefficient of correlation between X_2 and X_3. The value is between 0 and 1, as an absolute value.

We can see clearly that when the correlation between X_2 and X_3 gets **higher**, the denominator will be **smaller**, leading to **higher** variance.

(2) Effects on estimation

If we separate a part of $\text{var}(\hat{\beta}_2)$ like this,

› $\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2} \cdot \frac{1}{(1-r_{23}^2)}$

we can define the latter part as **variance-inflating factor** or VIF

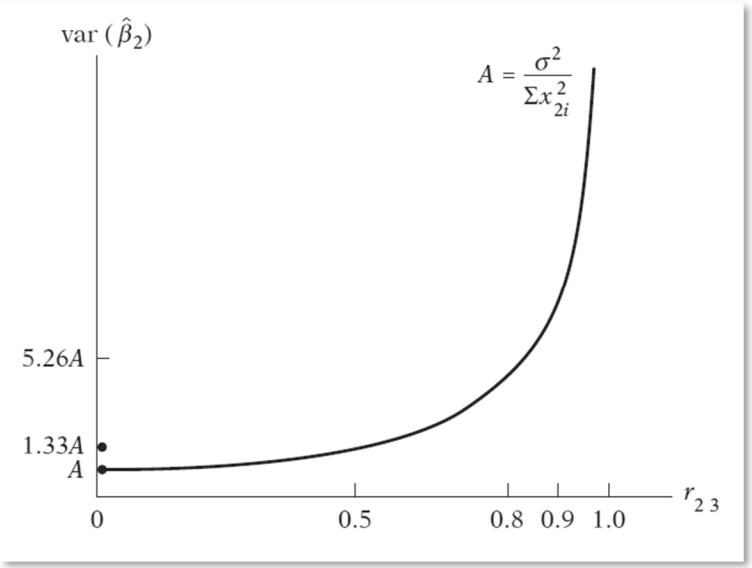
› $\text{VIF} = \frac{1}{(1-r_{23}^2)}$

The higher r_{23}^2, the higher it is for VIF. We can also define the inverse of VIF as **tolerance** or TOL as

› $\text{TOL} = \frac{1}{\text{VIF}} = (1 - r_{23}^2)$

Note that these definition can be generalized for any pair of regressors.

(2) Effects on estimation



3. Coefficients estimated are still BLUE.

Since the variance is used to construct confidence interval, CI is stretched outward and the t-curve is flatter, which will later affect

4. Large variances and covariances.

The acceptance region also becomes larger when variance is high. It is more likely that we would accept the null hypothesis when we should reject, causing more-likely type-II error.

(2) Effects on estimation

Example: Consider the weight model again, the first one take only height (hei_i) as a regressor while the second one include $hein_i$ which is height multiplied by a randomly generated number. The correlation between hei_i and $hein_i$ is exaggerate 0.9928 to the results as follows.

› First model

```
------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -5.383607   3.396954    -1.58   0.118    -12.16779    1.400578
         hei |   .5914144    .206873     2.86   0.006     .1782605    1.004568
       _cons |  -36.09017    35.8754    -1.01   0.318    -107.7383    35.55795
------------------------------------------------------------------------------
```

› Second model

```
------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -5.391169   3.411785    -1.58   0.119    -12.20699    1.424654
         hei |   1.359276   1.180214     1.15   0.254    -.9984733    3.717025
        hein |  -.7848297   1.187455    -0.66   0.511    -3.157043    1.587383
       _cons |  -33.34278   36.27082    -0.92   0.361    -105.8021    39.11651
------------------------------------------------------------------------------
```

(2) Effects on estimation

5. High R^2 but few significant t ratios

The coefficient of determination or R^2 from a model with multicollinearity is likely to be high, also the F-stat of overall model test, since the estimation is 'tricked' to have similar regressors with more explanatory power.

We can also see from the previous example that when we intentionally add a colinear variable into the regression, R^2 is higher.

However, this is not due to more explanatory power of regressors, but multicollinearity. Therefore, each coefficient is not very likely to be significant.

6. Sensitivity of coefficient due to small changes in data.

You can read for this example in page 331. In conclusion, a very slight change in data will affect the value of coefficient tremendously. Some of the direction of coefficient can be different from theoretical speculation.

It would be beneficial to read an illustrative example from page 332 to 337.

(3) Detecting multicollinearity

There are several ways to detect multicollinearity. Some of them are mentioned earlier. Some of them from the book are skipped.

Firstly, the phrase "**Rule of Thumb**" should be introduced. It refers to a specific level of criterion that is usually and mutually accepted as a threshold. More illustrative examples later here.

1. Conflicting test

This is already mentioned that when our estimation reports high R^2 or F value, but rarely coefficients are significant, we should suspect that there might be multicollinearity problem.

2. Pair-wise correlation among regressors

Another easy method to detect multicollinearity is to perform a pair-wise correlation on all regressors. (Easy when using STATA) The rule of thumb suggests that coefficient of correlation **exceeding 0.8** can be problematic and researcher may seek a remedial approach.

(3) Detecting multicollinearity

3. Auxiliary regressions

Regressing X_i on other X variables and obtain R_i^2 from the estimation then calculate

› $F_i = \frac{R^2_{xi \cdot x_2 x_3 \ldots x_k}/(k-2)}{(1-R^2_{xi \cdot x_2 x_3 \ldots x_k})/(n-k+1)}$

where k is the number of independent variables including intercept.

If F_i exceeds critical value from chosen level of significant, it means that X_i is collinear with other X.

Instead of testing all the auxiliary R_i^2, **Klein's rule of thumb** suggests that multicollinearity is troublesome if the R_i^2 is greater than R^2 from another model that we regress Y on these X_i and other X.

4. VIF and TOL

Again, we follow the rule of thumb that VIF should not exceed 10, which will happen when $r_{23}^2 = 0.8$, while TOL should be closer to 1 rather than 0.

(3) Detecting multicollinearity

Example: Using the same weight model of, but add a few more regressors

› $wei_i = \beta_1 + \beta_2 sex_i + \beta_3 hei_i + \beta_4 hein_i + \beta_5 ss_i + \beta_6 exc_i + u_i$

where ss_i is shoe size and exc_i is how many exercising days in a week.

We can ask for a report of VIF and TOL after a regression.

```
------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -2.051504   3.830285    -0.54   0.594    -9.708134    5.605127
         hei |   .7849816   1.206179     0.65   0.518    -1.626135    3.196099
        hein |  -.6364788   1.187903    -0.54   0.594    -3.011063    1.738105
          ss |   2.448109   1.143415     2.14   0.036     .1624558    4.733763
         exc |  -.2271317   .6111166    -0.37   0.711    -1.448737     .994473
       _cons |  -61.16189   38.61168    -1.58   0.118    -138.3455    16.02175
------------------------------------------------------------------------------

    Variable |       VIF       1/VIF
-------------+----------------------
       1.sex |      2.90    0.344756
         hei |     77.56    0.012894
        hein |     73.11    0.013677
          ss |      5.21    0.192038
         exc |      1.10    0.906274
-------------+----------------------
    Mean VIF |     31.98
```

(3) Detecting multicollinearity

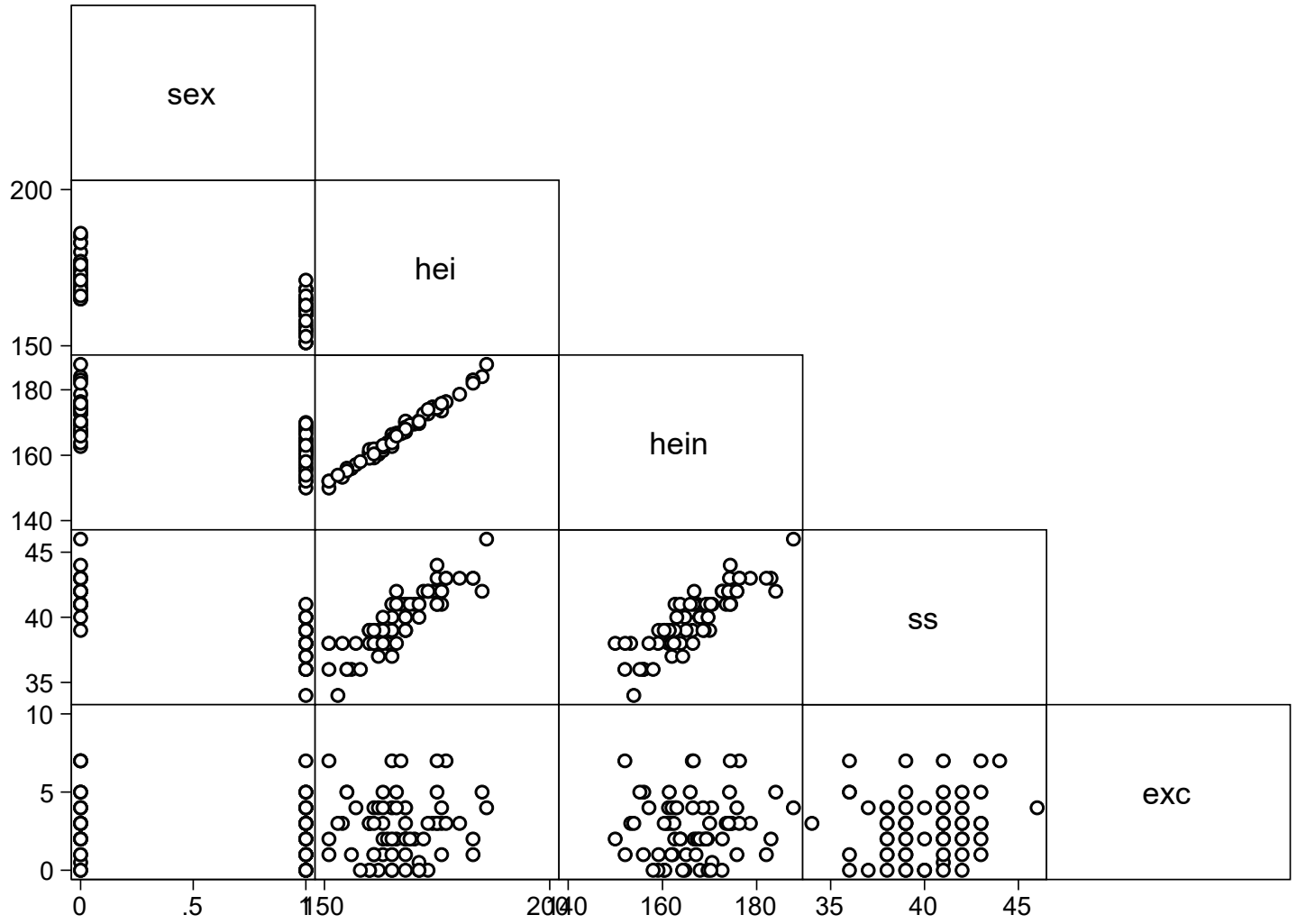
5. Scatter plot

Scatter plot is a good practice revealing linear relationship between two variables, see example below here.

Using the same weight model, we get the correlation matrix here.

	sex	hei	hein	ss	exc
sex	1.0000				
hei	-0.7396	1.0000			
hein	-0.7346	0.9928	1.0000		
ss	-0.7939	0.8708	0.8634	1.0000	
exc	-0.1970	0.0599	0.0831	0.1018	1.0000

(3) Detecting multicollinearity



(4) Remedial measures

1. Do nothing

If and only if when we do not any other choice than using deficient data set. Also, we may not be able to draw any meaningful insight from the regression.

2. Priori information

Supposed we have an income- consumption model as such,

› $Y_i = \beta_1 + \beta_2 income_i + \beta_3 wealth_i + u_i$

we know that income and wealth are highly colinear. If we know that from previous empirical work

› $\beta_3 = 0.1\beta_2$ then

› $Y_i = \beta_1 + \beta_2 income_i + 0.1\beta_2 wealth_i + u_i$ and

› $Y_i = \beta_1 + \beta_2 X_{2i} + u_i$

where $X_{2i} = income_i + 0.1 wealth_i$, we can eliminate one of the variables.

(4) Remedial measures

3. Combining cross-sectional and time-series data

The most completed data would be panel data, repeated samples over time. If that is not possible to obtain, we may use '**pooled-data**', combining multiple waves of data into one large data set.

4. Dropping a variable(s) and specification bias

This is the easiest method, and probably the best if possible. If we are sure which variable should be dropped, according to wrong specification of a model, dropping one of them is very easy and efficient.

Consider dropping $hein_i$ from the show size model, our results would be as follows.

(4) Remedial measures

› First model

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(5, 62)        =      8.97
       Model |  3825.50185         5   765.10037   Prob > F        =    0.0000
    Residual |  5289.56936        62  85.3156348   R-squared       =    0.4197
-------------+----------------------------------   Adj R-squared   =    0.3729
       Total |  9115.07121        67  136.045839   Root MSE        =    9.2366

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -2.051504   3.830285    -0.54   0.594    -9.708134    5.605127
         hei |   .7849816   1.206179     0.65   0.518    -1.626135    3.196099
        hein |  -.6364788   1.187903    -0.54   0.594    -3.011063    1.738105
          ss |   2.448109   1.143415     2.14   0.036     .1624558    4.733763
         exc |  -.2271317   .6111166    -0.37   0.711    -1.448737     .994473
       _cons |  -61.16189   38.61168    -1.58   0.118    -138.3455    16.02175
------------------------------------------------------------------------------
```

› Second model

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(4, 63)        =     11.27
       Model |  3801.00927         4  950.252317   Prob > F        =    0.0000
    Residual |  5314.06194        63  84.3501895   R-squared       =    0.4170
-------------+----------------------------------   Adj R-squared   =    0.3800
       Total |  9115.07121        67  136.045839   Root MSE        =    9.1842

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -2.110053   3.807001    -0.55   0.581    -9.717738    5.497631
         hei |   .1568177   .2819089     0.56   0.580    -.4065322    .7201677
          ss |   2.462676   1.136605     2.17   0.034     .1913511    4.734001
         exc |  -.2933697    .595086    -0.49   0.624    -1.482554    .8958147
       _cons |  -62.84932   38.26466    -1.64   0.105    -139.3152    13.61651
------------------------------------------------------------------------------
```

(4) Remedial measures

› Third model

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(3, 64)        =     12.72
       Model |   3405.0221         3  1135.00737   Prob > F        =    0.0000
    Residual |  5710.04911        64  89.2195173   R-squared       =    0.3736
-------------+----------------------------------   Adj R-squared   =    0.3442
       Total |  9115.07121        67  136.045839   Root MSE        =    9.4456

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -5.769422   3.508961    -1.64   0.105    -12.77938    1.240533
         hei |   .5782139   .2098837     2.75   0.008     .1589229    .9975048
         exc |  -.2961036     .61202    -0.48   0.630    -1.518754     .926547
       _cons |  -32.88176   36.69289    -0.90   0.374    -106.1842    40.42072
------------------------------------------------------------------------------
```

(4) Remedial measures

5) Adding more observations

When possible, more observations lead to more variability in X and therefore, may lead to reduction of severity of multicollinearity problem.

6) Variable transformation

Transforming variables can be complicated, we can either perform

› **Ratio transformation** which will lead us to another problem of heteroscedasticity or

› **First difference form** of variable which is popular in time-series data.

Therefore, transforming is not very much recommended.

(1) The nature

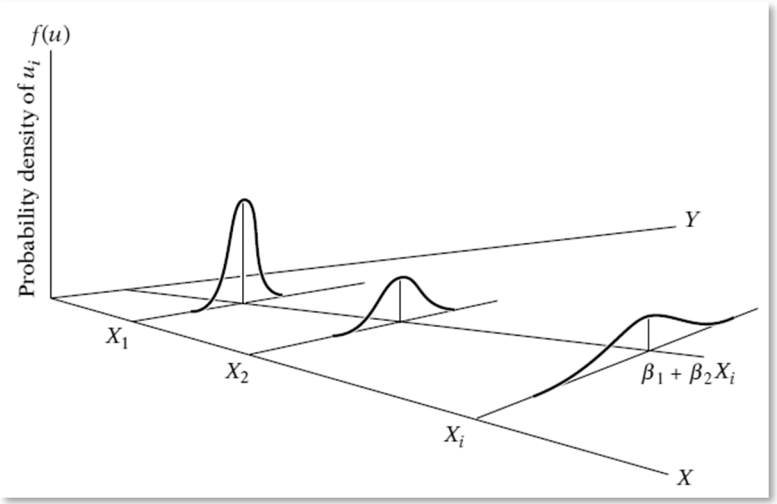
Recall the assumption of homoscedasticity or

› $E(u_i^2|X_i) = \sigma^2$

when this assumption is relaxed, we have

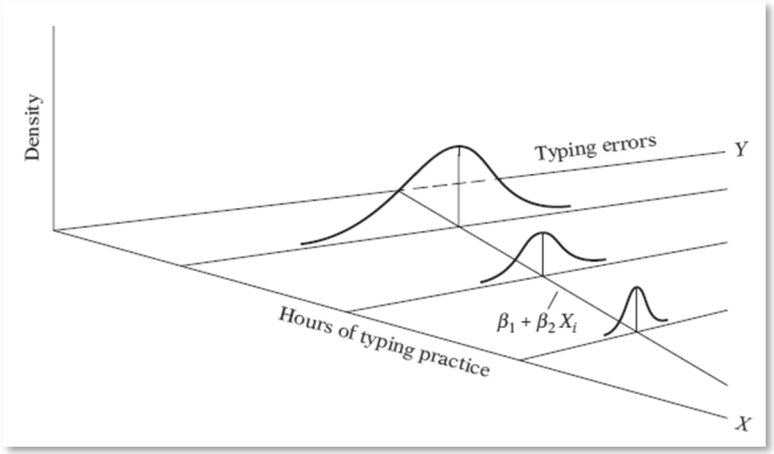
› $E(u_i^2|X_i) = \sigma_i^2$

which means that we allow the error term scattered around each X_i to be different. Two classic examples are income-saving model and error-learning model as follows.



As people get richer, they have more choices over their consumption-saving, leading to a larger variance of saving (Y_i) on larger income (X_i).

(1) The nature



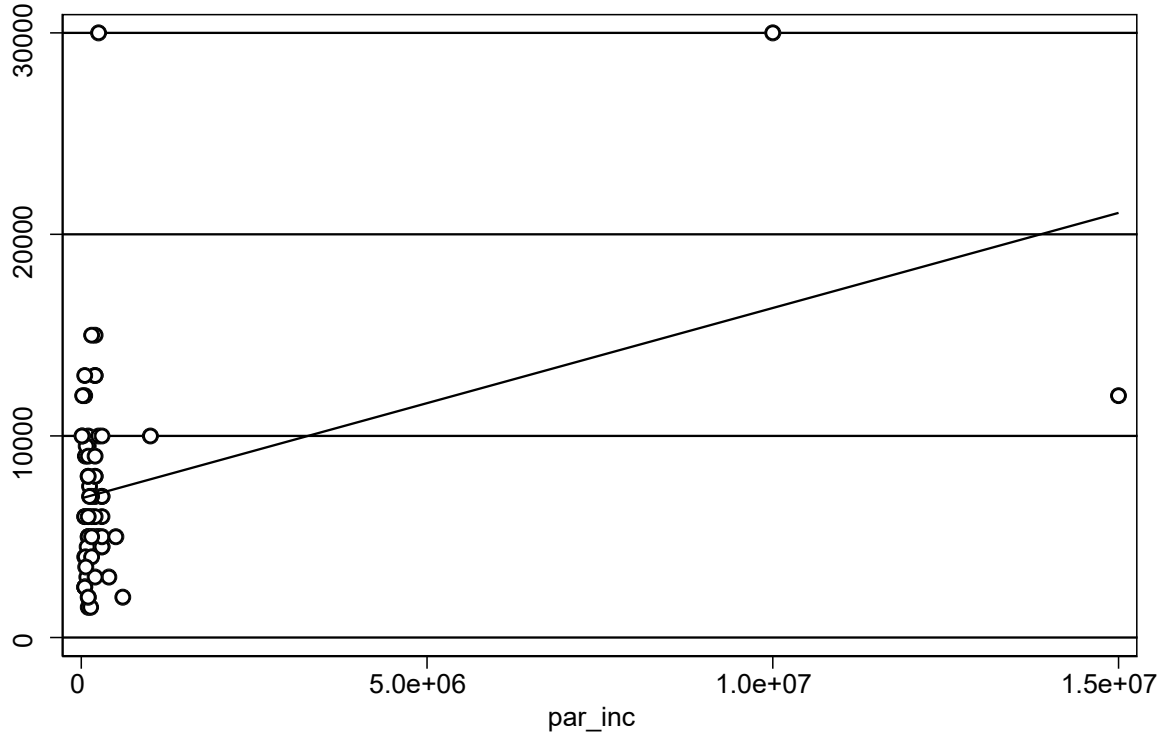
Similarly, people practicing more hours on typing leads to lower typing errors. However, some people maybe pretty good at typing at first and some can be pretty bad due to their familiarity to keyboard layout or language. There are larger difference between people when start practicing but those difference will be minimized as they keep practicing.

Apart from the nature of data, there are some other causes for heteroscedasticity.

(1) The nature

1. Presence of outliers

An outlier is an observation that is much different from the rest, either little or largely different. Inclusion and exclusion of an outlier can alter the result of a regression substantially.



(1) The nature

2. Specification error

For example, a price demand model can be heteroscedastic if we do not include price of complementary or competing commodities. Other commodities' price may be the source of scattered quantity demanded at some level.

3. Skewness of distribution

For example, plotting wealth and education level can show this problem because there a fewer people with high wealth, leading to lower variance in education level, while larger groups of population with lower wealth.

4. Incorrect data transformation and functional form.

Details for this topic will be skipped on this point.

Heteroscedasticity is more likely to be found in cross-sectional data rather than time-series since they deal with members of a population at a given point of time.

(2) Effects on estimation

We begin our examination on simple linear regression, recall that with homoscedasticity assumption yields the variance of the estimator as

› $\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$

Relaxing the assumption, the variance becomes

› $\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$

$\hat{\beta}_2$ is not be BLUE, with heteroscedasticity. It is still linear and unbiased but it is not efficient anymore, comparing to deriving estimator from another method called **generalized least squares** (**GLS**).

(Note that an estimator not being efficient meaning that $\text{var}(\hat{\beta}_i)$ is not the lowest.)

(2) Effects on estimation

To estimate with GLS, we start from the basic model of

› $Y_i = \beta_1 + \beta_2 X_i + u_i$

Then we transform these variables by dividing by σ_i, if this standard error is known.

› $\frac{Y_i}{\sigma_i} = \frac{\beta_1}{\sigma_i} + \frac{\beta_2 X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$

Then figure out the variance, we get

› $\text{var}\left(\frac{u_i}{\sigma_i}\right) = E\left(\frac{u_i}{\sigma_i}\right)^2 = \frac{1}{\sigma_i^2} E(u_i^2)$ and if σ_i is known

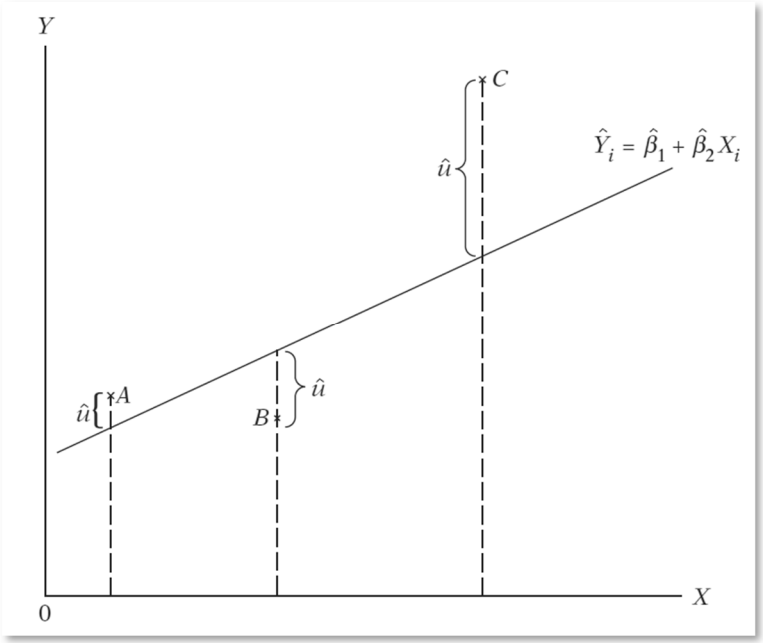
› $\text{var}\left(\frac{u_i}{\sigma_i}\right) = \frac{1}{\sigma_i^2}\left(\sigma_i^2\right) = 1$

which is actually homoscedastic. To retrieve the estimators, we follow least squares method as usual

(2) Effects on estimation

› $\min_{\hat{\beta}_1,\hat{\beta}_2} \sum \left(\frac{u_i}{\sigma_i}\right)^2 = \sum \left(\frac{Y_i}{\sigma_i} - \frac{\hat{\beta}_1}{\sigma_i} - \frac{\hat{\beta}_2 X_i}{\sigma_i}\right)^2$

This method is specifically called **weight least squares (WLS)**, weighing each term especially the error term with the error term itself. Estimators derived from this estimation are called **WLS estimators**. WLS is a class of GLS.



(2) Effects on estimation

WLS estimators derived will have less variance compared to ordinary OLS. Therefore, estimators from OLS is not with the least variance anymore, losing the quality of being efficient.

Consequences of relying on OLS are as follows.

1) OLS estimation allowing heteroscedasticity

› Using OLS while assuming σ_i^2 are known, $\text{var}(\hat{\beta}_2)$ is larger compared to the variance from WLS.

› Larger CI and t value is small, leading to insignificant conclusion.

2) OLS estimation disregarding heteroscedasticity

› Using OLS while assuming homoscedasticity (σ^2) when heteroscedasticity is present, $\text{var}(\hat{\beta}_2)$ will be biased.

› We do not know whether the bias is positive (overestimate: actual variance is lower) or negative (underestimate: actual variance is higher), depending on the relationship between σ_i^2 and X_i.

› Conclusion or inference drawn from hypothesis tests may be misleading.

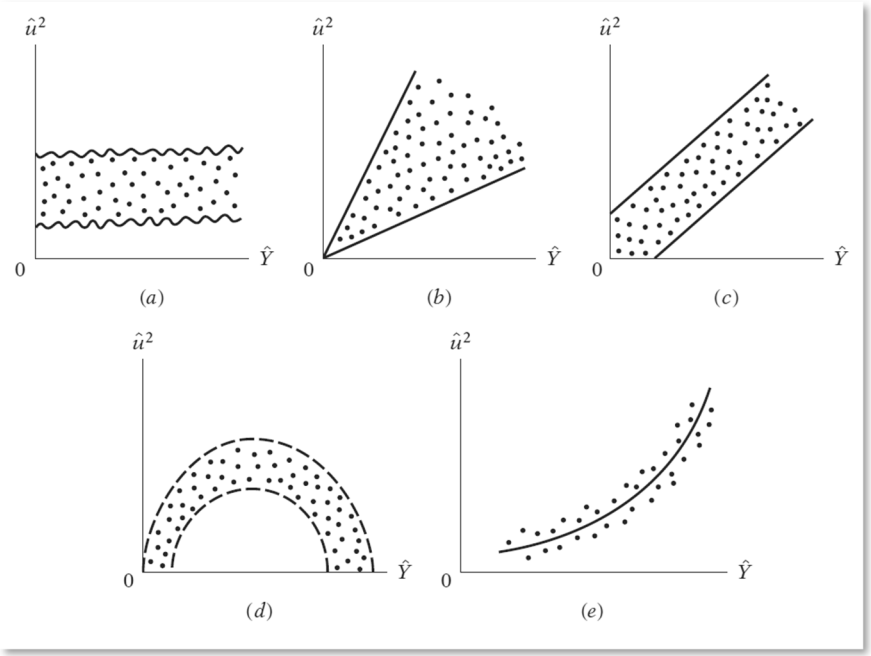
(3) Detecting heteroscedasticity

1. Graphical method

› **Step 1:** Estimate coefficients with OLS.

› **Step 2:** Retrieve $\hat{u}_i^2$. (In STATA, look for predicting residual)

› **Step 3:** Plot $\hat{u}_i^2$ with X_i or $\hat{Y}_i$. There might be multiple X_i in our regression function, so we can rely on $\hat{Y}_i$ as well.



Which of these plots that heteroscedasticity is present?

(3) Detecting heteroscedasticity

2. Park Test

The intuition of Park test assumes that when heteroscedasticity is present, σ_i^2 is a kind of function of X_i. Steps are as follows.

› **Step 1**: Estimate coefficients with OLS and retrieve $\ln \hat{u}_i^2$

› **Step 2**: Estimate $\ln \hat{u}_i^2 = \alpha + \beta \ln X_i + \nu_i$

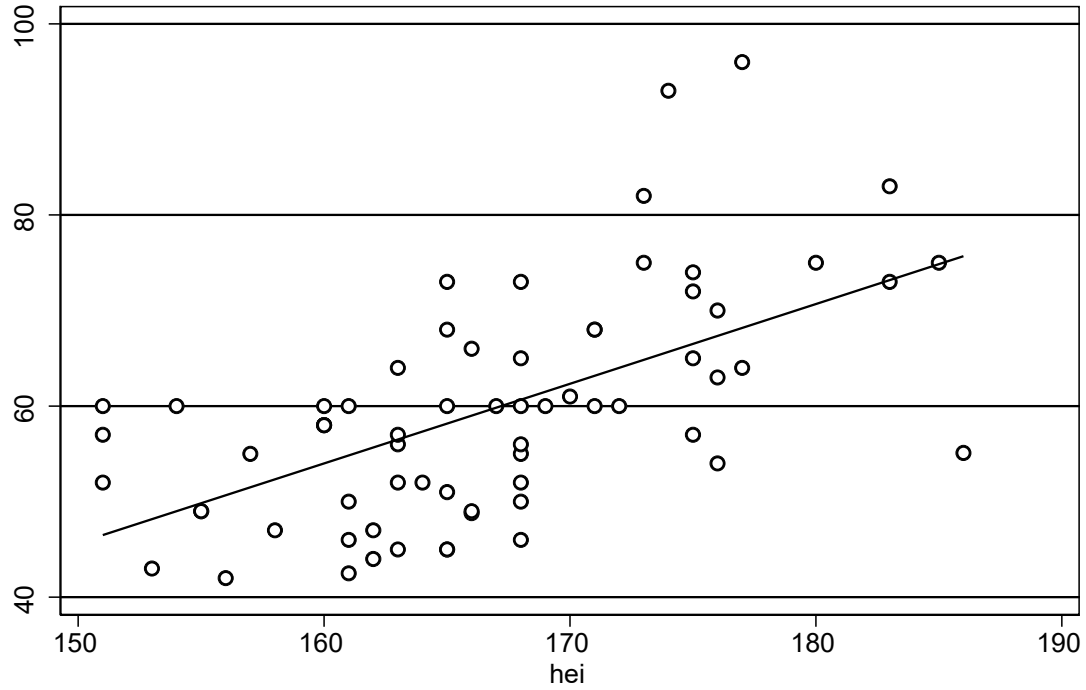
› **Step 3**: Test the significance from zero of β. If it is, it would suggest that heteroscedasticity is present.

Example: Revisit the weight model with only one explanatory variable or height as follows.

› $wei_i = \beta_1 + \beta_2 hei_i + u_i$

We first plot the relationship between weight and height.

(3) Detecting heteroscedasticity



Notice that we do not see significant shift of the variance of $\hat{u}_i$ for each level of X_i.

(3) Detecting heteroscedasticity

› The model

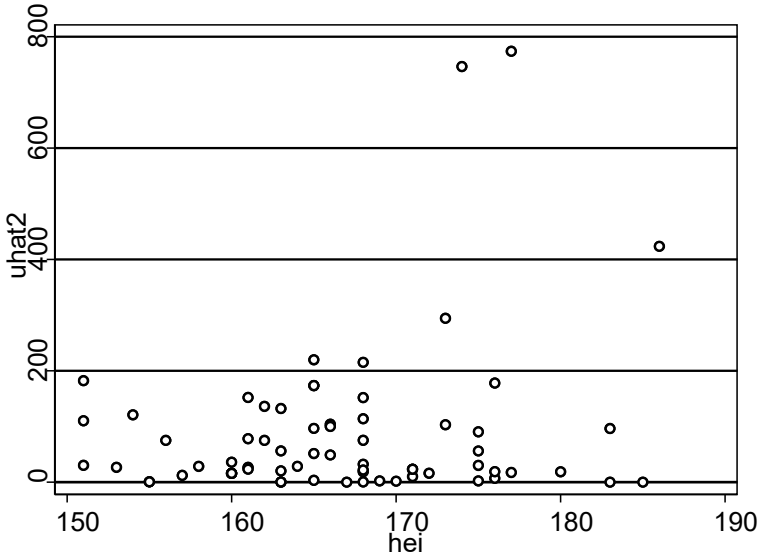
```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =     35.07
       Model |  3162.68606         1  3162.68606   Prob > F        =    0.0000
    Residual |  5952.38514        66  90.1876537   R-squared       =    0.3470
-------------+----------------------------------   Adj R-squared   =    0.3371
       Total |  9115.07121        67  136.045839   Root MSE        =    9.4967

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |    .833902   .1408188     5.92   0.000     .5527483    1.115056
       _cons |   -79.4222   23.49114    -3.38   0.001    -126.3238   -32.52063
------------------------------------------------------------------------------
```

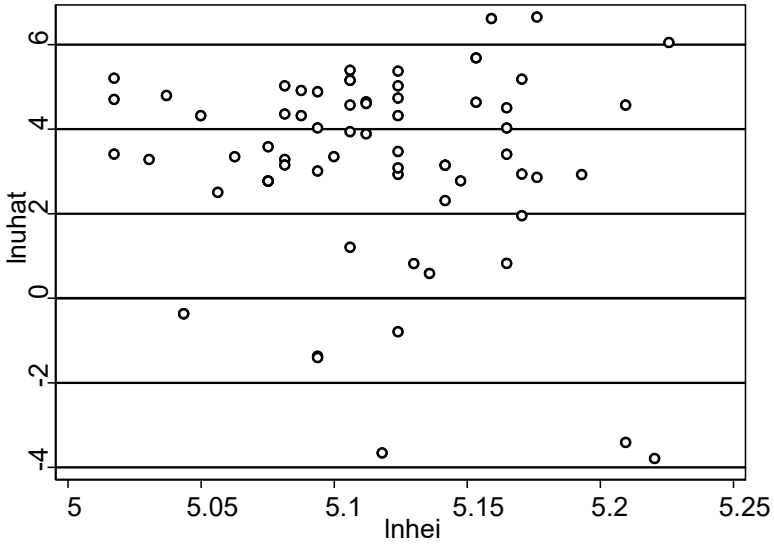
› Park test

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =      0.46
       Model |  2.40426654         1  2.40426654   Prob > F        =    0.5023
    Residual |  348.746061        66  5.28403123   R-squared       =    0.0068
-------------+----------------------------------   Adj R-squared   =   -0.0082
       Total |  351.150328        67  5.24104967   Root MSE        =    2.2987

------------------------------------------------------------------------------
      lnuhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnhei |  -3.842844   5.696973    -0.67   0.502    -15.21722    7.531529
       _cons |   22.80914   29.13852     0.78   0.437    -35.36778    80.98606
------------------------------------------------------------------------------
```

(3) Detecting heteroscedasticity

We can see that when we plot $\hat{u}_i^2$ with height, and $\ln(\hat{u}_i^2)$ with height, there is no correlation with each other.





(3) Detecting heteroscedasticity

3. Breusch-Pagan (BP) Test

This is a general case for other tests that follow Breusch and Pagan's idea (such as the Breusch-Pagan-Godfrey: BPG test in the textbook). This test is taken from Wooldridge page 270.

› **Step 1**: Estimate coefficients with OLS and retrieve $\hat{u}_i^2$.

› **Step 2**: Regress $\hat{u}_i^2 = \delta_1 + \delta_2 X_{2i} + \cdots + \delta_k X_{ki} + \nu_i$ and retrieve $R^2_{\hat{u}_i^2}$.

› **Step 3**: Calculate F-stat by

$$F_{cal} = \frac{R^2_{\hat{u}_i^2}/(k)}{(1 - R^2_{\hat{u}_i^2})/(n-k-1)}$$

› **Step 4**: Test the null hypothesis of homoscedasticity. If we can reject the null hypothesis ($F_{cal} > F_{cri}$) at the selected significant level, heteroscedasticity is present in our model.

(3) Detecting heteroscedasticity

Here is the result of BP-test. ($\hat{u}_i^2 = \delta_1 + \delta_2 hei_i + \nu_i$)

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =      3.45
       Model |  67079.8689         1  67079.8689   Prob > F        =    0.0676
    Residual |  1281878.37        66  19422.3995   R-squared       =    0.0497
-------------+----------------------------------   Adj R-squared   =    0.0353
       Total |  1348958.24        67   20133.705   Root MSE        =    139.36

------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |   3.840458   2.066514     1.86   0.068    -.2854704    7.966387
       _cons |  -552.3531   344.7323    -1.60   0.114    -1240.633    135.9271
------------------------------------------------------------------------------
```

Now let's try calculate F,

› $F_{cal} = \frac{R^2_{\hat{u}_i^2}/(k)}{(1-R^2_{\hat{u}_i^2})/(n-k-1)} =$

And find the $F_{cri} =$

(3) Detecting heteroscedasticity

4. White's test

White's test is also very similar to Breusch-Pagan. The differences are the residual model and test statistics.

The steps here applies for 2 explanatory variables, but it is extendable. White's test has an advantage over BP test as it is not sensitive to the assumption of normality.

› **Step 1:** Estimate coefficients with OLS and retrieve $\hat{u}_i^2$.

› **Step 2:** Regress

$$\hat{u}_i^2 = \delta_1 + \delta_2 X_{2i} + \delta_3 X_{3i} + \delta_4 X_{2i}^2 + \delta_5 X_{3i}^2 + +\delta_6 X_{2i} X_{3i} + \nu_i$$

and retrieve $R^2_{\hat{u}_i^2}$.

Additions of higher power and cross product imply that the error variance is functionally related to regressors, their squares, and their cross product.

(3) Detecting heteroscedasticity

› **Step 3**: Calculate the test stat, which in this case, we use **Lagrange Multiplier** (LM) stat

$$LM_{cal} = n \cdot R^2_{\hat{u}_i^2} \sim \chi^2_{k-1.}$$

LM is very useful in many cases due to its basic calculation and **asymptotically** distributed as Chi-square with k d.f.

› **Step 4**: Test the null hypothesis of homoscedasticity. If we can reject the null hypothesis ($LM_{cal} > \chi^2_{k-1.}$) at the selected significant level, heteroscedasticity is present in our model.

(3) Detecting heteroscedasticity

Here is the result of White's test. ($\hat{u}_i^2 = \delta_1 + \delta_2 hei_i + \delta_2 hei_i^2 + \nu_i$)

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(2, 65)        =      1.96
       Model |  76814.6367         2  38407.3183   Prob > F        =    0.1488
    Residual |   1272143.6        65    19571.44   R-squared       =    0.0569
-------------+----------------------------------   Adj R-squared   =    0.0279
       Total |  1348958.24        67   20133.705   Root MSE        =     139.9

------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |   -42.8244   66.19907    -0.65   0.520    -175.0331    89.38427
        hei2 |   .1392366   .1974249     0.71   0.483    -.2550482    .5335214
       _cons |   3348.114   5541.327     0.60   0.548    -7718.679    14414.91
------------------------------------------------------------------------------
```

Now let's try calculate LM stat

› $LM_{cal} = n \cdot R^2_{\hat{u}_i^2} =$

And find the critical value of $\chi^2_{k-1.} =$

(3) Detecting heteroscedasticity

Note that in STATA, the procedures are far way simpler.

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =     35.07
       Model |  3162.68606         1  3162.68606   Prob > F        =    0.0000
    Residual |  5952.38514        66  90.1876537   R-squared       =    0.3470
-------------+----------------------------------   Adj R-squared   =    0.3371
       Total |  9115.07121        67  136.045839   Root MSE        =    9.4967

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |    .833902   .1408188     5.92   0.000      .5527483    1.115056
       _cons |   -79.4222   23.49114    -3.38   0.001     -126.3238   -32.52063
------------------------------------------------------------------------------
```

```
. estat hettest
Breusch-Pagan / Cook-Weisberg test
for heteroskedasticity

Ho: Constant variance
Variables: fitted values of wei

         chi2(1)      =     4.38
         Prob > chi2  =   0.0364
```

```
. estat imtest, white
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(2)      =     3.87
         Prob > chi2  =   0.1443

Cameron & Trivedi's decomposition of IM-test
---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |       3.87      2    0.1443
            Skewness |       3.10      1    0.0782
            Kurtosis |       0.84      1    0.3608
---------------------+-----------------------------
               Total |       7.81      4    0.0988
---------------------------------------------------
```

(4) Remedial measures

1. Weighted Least Squares (WLS)

It can be estimated when σ_i^2 is known.

2. Data transformation: selecting a class of GLS

Advantage of this method is that we do not need asymptotic property. However, a major drawback is we need to speculate relationship between σ_i^2 and X_i.

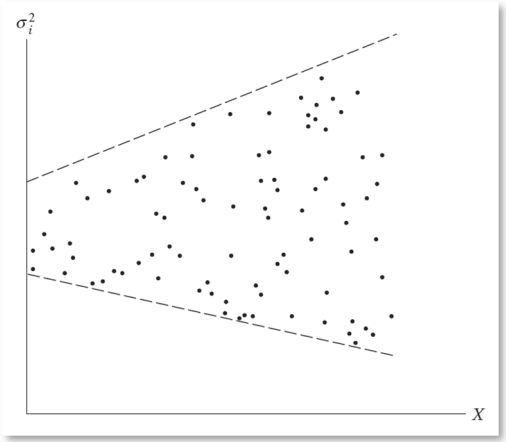
For example, if we assume that σ_i^2 is proportional to X_i so that

› $E(u_i^2) = \sigma^2 X_i$ then $\sigma^2 = \frac{E(u_i^2)}{X_i} = \frac{E(u_i)}{\sqrt{X_i}}$

we can transform our model into

› $\frac{Y_i}{\sqrt{X_i}} = \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}}$ where X_i must be > 0

$E\left(\frac{u_i}{\sqrt{X_i}}\right) = \sigma^2$ or homoscedastic.



(4) Remedial measures

However, we can see another drawback of this method is that the interpretation of coefficients are now different.

Moreover, there are multiple forms of transformation, due to relationship between σ_i^2 and X_i.

3. White's robust standard errors

This method assumes asymptotic property of our data, which is quite common in national cross-sectional data.

We do not cover how it is derived, even in the book Gujarati does not as well, but we compared the result of normal estimation with using White's robust standard errors.

(4) Remedial measures

› OLS model

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =     35.07
       Model |  3162.68606         1  3162.68606   Prob > F        =    0.0000
    Residual |  5952.38514        66  90.1876537   R-squared       =    0.3470
-------------+----------------------------------   Adj R-squared   =    0.3371
       Total |  9115.07121        67  136.045839   Root MSE        =    9.4967

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |    .833902   .1408188     5.92   0.000     .5527483    1.115056
       _cons |   -79.4222   23.49114    -3.38   0.001    -126.3238   -32.52063
------------------------------------------------------------------------------
```

› OLS model with robust standard error

```
Linear regression                               Number of obs     =        68
                                                F(1, 66)          =     27.84
                                                Prob > F          =    0.0000
                                                R-squared         =    0.3470
                                                Root MSE          =    9.4967

------------------------------------------------------------------------------
             |               Robust
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |    .833902   .1580391     5.28   0.000     .5183667    1.149437
       _cons |   -79.4222   25.98683    -3.06   0.003    -131.3066   -27.53781
------------------------------------------------------------------------------
```

(1) The nature

Recall the assumption of no autocorrelation or serial correlation

› $cov(u_i, u_j | x_i, x_j) = E(u_i u_j) = 0$ where $i \neq j$

Put simply, the term means "correlation between members of series of observations ordered in time (as in time series data) or space (as in cross-sectional data)." If the problem exists,

› $E(u_i u_j) \neq 0$ where $i \neq j$

There can be multiple sources of autocorrelation.

1. Inertia

For example, when an economy is in recovery, there is a 'momentum' built into policies driving economic outcome from previous period.

(1) The nature

2. Lags

For instance, a consumption-income model usually includes previous consumption expenditure

› $cons_t = \beta_1 + \beta_2 income_t + \beta_3 cons_{t-1} + u_i$

This is because either psychologically, technologically, or institutionally, people do not change consumption pattern rapidly across periods.

3. Cobweb phenomenon

Most likely to occur with agricultural products, their supply cannot adjust with price instantaneously.

(1) The nature

4. Nonstationarity

Stationarity refers to time-invariant characteristics such as mean, variance, covariance, which is quite rare in time-series data. For example, GDP of an economy is always growing with non-systematic shocks that makes the characteristics time-variant.

5. Specification bias: incorrect functional form

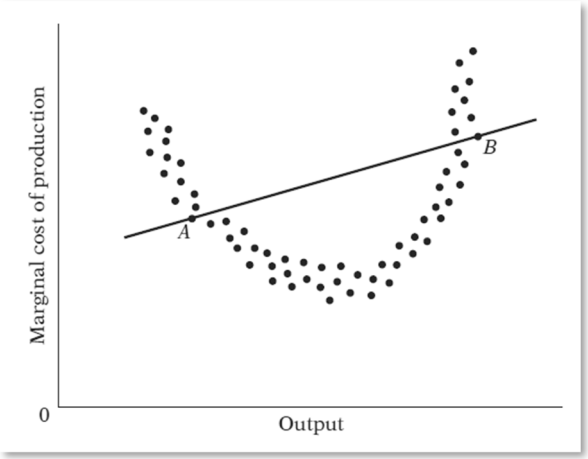
For example, a curved marginal cost which the 'true' model is

› $MC_i = \beta_1 + \beta_2 Q_i + \beta_3 Q_i^2 + u_i$

but if we try to fit our data with a linear model instead

› $MC_i = \beta_1 + \beta_2 Q_i + u_i$

(1) The nature



Between point A and B, linear model underestimates marginal cost while before point A and beyond point B, the model overestimates marginal cost.

If we correlate u_i, u_j we can see clearly that there is a pattern following the curve, hence autocorrelation is present in the linear model.

However, this problem does not surface when we fit the data with polynomial model.

(2) Effects on estimation

From this point on, we will focus on a time-series model, varying Y_t and X_t through time, not across groups of observation in the same period, the model becomes

› $Y_t = \beta_1 + \beta_2 X_t + u_t$

If the error terms are correlated, assumed linearly

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

This equation is called **first-order autoregressive scheme** or shortly **AR(1).** If the error term t also correlates with two period back, the model is called **AR(2)**, and so on.

› $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$

(2) Effects on estimation

Normally, OLS with no autocorrelation will yield the variance of an estimator as

› $\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_t^2}$

If u_t follows AR(1), the variance becomes

› $\text{var}(\hat{\beta}_2)_{AR(1)} = \frac{\sigma^2}{\sum x_t^2}[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + \ldots + 2\rho^{n-1}\frac{\sum x_t x_n}{\sum x_t^2}]$

We cannot say for sure whether $\text{var}(\hat{\beta}_2)$ is more or less than $\text{var}(\hat{\beta}_2)_{AR(1)}$. Though $\hat{\beta}_2$ is still linear and unbiased, variance is not minimum, or not being efficient. See this proof of GLS on page 422.

(2) Effects on estimation

If we run the regressing, disregarding autocorrelation, we will find that

› $\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{(n-2)}$ is likely to underestimate the true σ^2.

› Overestimation of R^2.

› If σ^2 is not underestimated, $\text{var}(\hat{\beta}_2)$ may still underestimate $\text{var}(\hat{\beta}_2)_{AR(1)}$, and the latter is still inefficient compared to $\text{var}(\hat{\beta}_2)_{GLS}$.

› Both t and F tests are no longer valid.

(3) Detecting autocorrelation

1. Graphical method

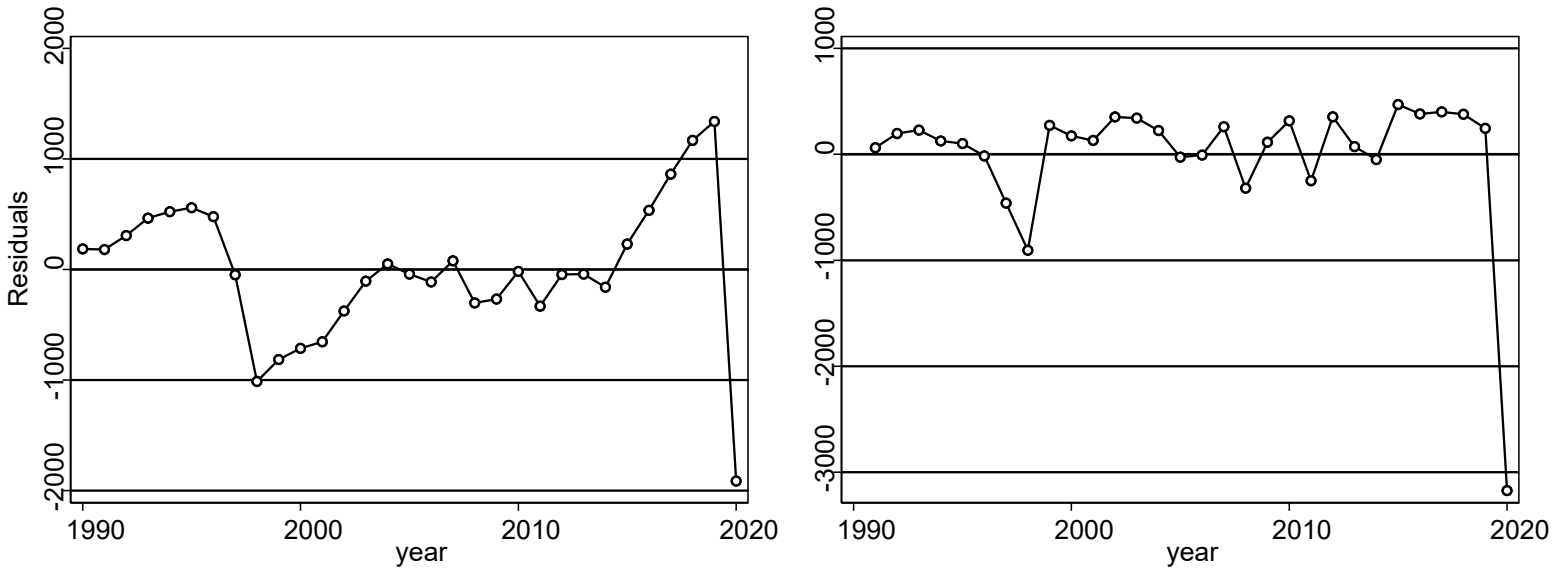
The first one, once again, is not informal but visually telling. If we plot the residuals with time period, we might see interconnection between time period.

Example: Data of GDP (Y_t), in CVM, and headline CPI (X_t) in Thailand from 1990 to 2020 are taken from the Bank of Thailand statistics page. We model it as usual.

› $Y_t = \beta_1 + \beta_2 X_t + u_t$

Now see the result of the estimation on the right-hand side. After that, we can predict $\hat{u}_t$ and plot them over time.

(3) Detecting autocorrelation



On the left-hand side, this is a plot from the model residual. On the other hand, the right-hand side shows a plot from another model that autocorrelation is already resolved.

If autocorrelation is not present, residuals should be randomly distributed around 0 and has no obvious interconnection with the previous period.

(3) Detecting autocorrelation

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =    312.82
       Model |   131519253         1   131519253   Prob > F        =    0.0000
    Residual |  12192677.5        29  420437.155   R-squared       =    0.9152
-------------+----------------------------------   Adj R-squared   =    0.9122
       Total |   143711930        30  4790397.67   Root MSE        =    648.41

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   6.286548    17.69   0.000     98.33017     124.045
       _cons |   -1827.06   507.7518    -3.60   0.001    -2865.529   -788.5911
------------------------------------------------------------------------------
```

We can also see that R^2, t and F are very high, leading to a very significant coefficient, but this is due autocorrelation.

(3) Detecting autocorrelation

2. Durbin-Watson d Test

Durbin-Watson ***d*** statistics is defined as

› $d = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$

implies sum of squared differences to the RSS.

For Durbin-Watson test, we assume

› Regression model includes intercept term.

› Regressors X are stochastic or fixed in repeated sampling.

› u_t are generated by AR(1) and normally distributed.

› The model does not include lagged variable(s) of the dependent variable, such as

$$Y_t = \beta_1 + \beta_2 X_t + \gamma Y_{t-1} + u_t$$

› No missing observations in the data.

(3) Detecting autocorrelation

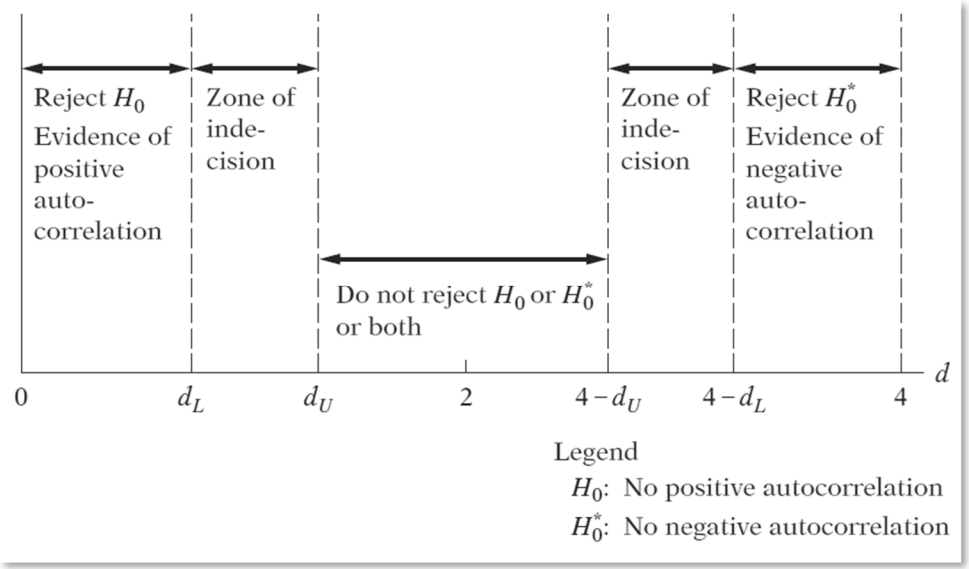
We can also define 'approximate' version of the d stat as

› $d \approx 2(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2})$ now let's define

› $\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$ then

› $d \approx 2(1 - \hat{\rho})$

Since $-1 \leq \hat{\rho} \leq 1$, therefore $0 \leq d \leq 4$



(3) Detecting autocorrelation

Example: using the same dataset, we follow the steps here.

› **Step 1**: State the hypothesis

Since we are testing for general autocorrelation, without any speculation of positive or negative correlation, we are going to set hypothesis for both.

› H_0: No autocorrelation $(d_U < d < 4 - d_U)$

› H_a: Positive autocorrelation $(0 < d < d_L)$ or

negative autocorrelation $(4 - d_L < d < 4)$

Do not forget that we can also reach the inconclusive answer for this test.

› **Step 2:** Run the OLS regression and obtain the predicted residuals.

(3) Detecting autocorrelation

› **Step 3:** Find the critical values. Two 'attributes' that are important for finding d_L and d_U are **number of coefficients** and **number of observation.**

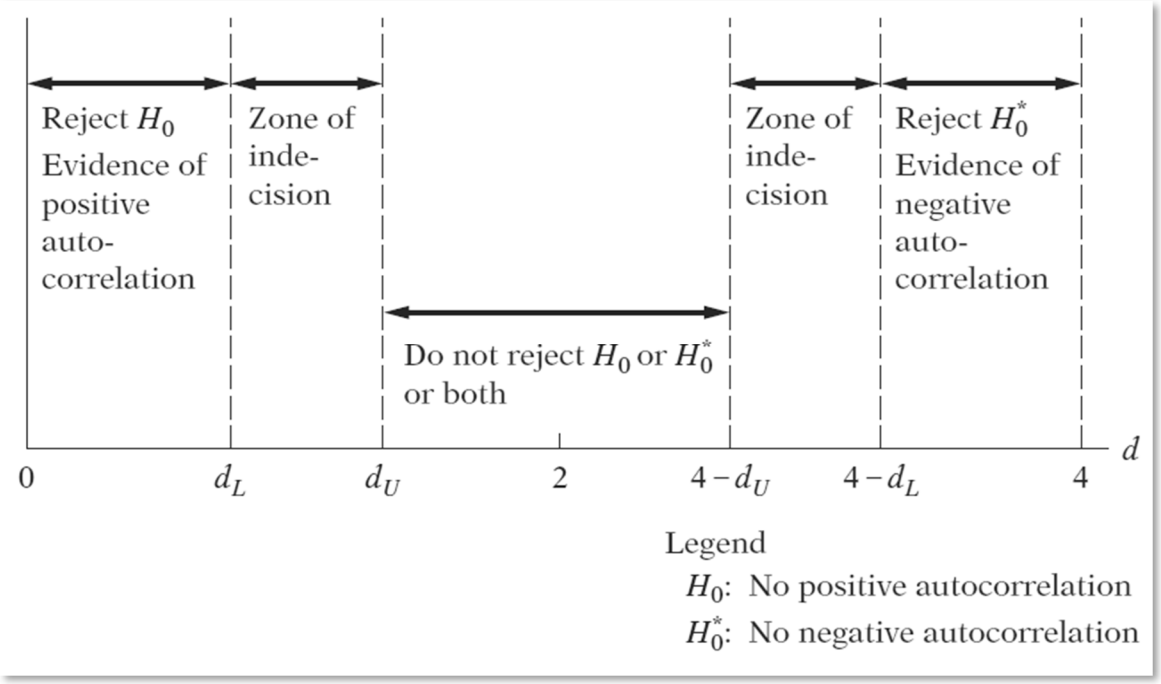
› $d_L =$ and $4 - d_L =$

› $d_U =$ and $4 - d_U =$

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =    312.82
       Model |   131519253         1   131519253   Prob > F        =    0.0000
    Residual |  12192677.5        29  420437.155   R-squared       =    0.9152
-------------+----------------------------------   Adj R-squared   =    0.9122
       Total |   143711930        30  4790397.67   Root MSE        =    648.41

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   6.286548    17.69   0.000     98.33017     124.045
       _cons |   -1827.06   507.7518    -3.60   0.001    -2865.529   -788.5911
------------------------------------------------------------------------------

. estat dwatson

Durbin-Watson d-statistic(  2,    31) =  1.065525
```

(3) Detecting autocorrelation

› **Step 4:** Conclude the test.



(3) Detecting autocorrelation

3. General test of autocorrelation: The Breusch-Godfrey test (BG)

Assume a model of

› $Y_t = \beta_1 + \beta_2 X_t + u_t$

Now if the error term u_t follows the pth-order autoregressive AR(P) as follows

› $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t$

The concept is to test these coefficients simultaneously by following the steps.

› **Step 1:** State the hypothesis

› H_0: No autocorrelation ($\rho_1 = \rho_2 = \cdots = \rho_p = 0$)

› H_a: Autocorrelation (ρ are not simultaneously zero)

Step 2: Run the OLS regression and obtain the residuals.

(3) Detecting autocorrelation

› **Step 3:** Estimate this equation, also including regressor(s) into this one.

› $\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \cdots + \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_t$

Then obtain the R^2 from this model

› **Step 4:** Calculate LM-statistics, if the sample is large, BG have shown that

› $\text{LM} = (n-p)\text{R}^2 \sim \chi_p^2$

› **Step 5:** Look for the critical value in chi-square table and reject the null hypothesis if the LM exceeds the critical value.

Note that, in STATA, the default lag is 1 but there is an option to include more lags.

(3) Detecting autocorrelation

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =    312.82
       Model |   131519253         1   131519253   Prob > F        =    0.0000
    Residual |  12192677.5        29  420437.155   R-squared       =    0.9152
-------------+----------------------------------   Adj R-squared   =    0.9122
       Total |   143711930        30  4790397.67   Root MSE        =    648.41

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   6.286548    17.69   0.000     98.33017     124.045
       _cons |   -1827.06   507.7518    -3.60   0.001    -2865.529   -788.5911
------------------------------------------------------------------------------

. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          4.597               1                   0.0320
---------------------------------------------------------------------------
                       H0: no serial correlation
```

(4) Remedial measures

First of all, make sure that the model is correctly specified (we are going to deal with autocorrelation only). Most of the time we do not know the relationship between u_t and u_{t-1}. In other words, we do not know the value of ρ.

1. First-difference method

The first difference equation takes the form of

› $Y_t - Y_{t-1} = \beta_2(X_t - X_{t-1}) + (u_t - u_{t-1})$ or

› $\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t$ where $\varepsilon_t = \Delta u_t$

The rule of thumb is that we can use this equation to estimate when $d < R^2$. Note that this model has no intercept term. If included, the intercept is interpreted as **time trend**.

(4) Remedial measures

As we can see from the result, when we use the first-difference model, we cannot reject the null hypothesis of BG test any longer because ε_t is a stationary white-noise error term. However, note that we lose 1 observation due to using difference term.

```
      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 28)        =      1.99
       Model |  911983.341         1  911983.341   Prob > F        =    0.1695
    Residual |    12843973        28  458713.323   R-squared       =    0.0663
-------------+----------------------------------   Adj R-squared   =    0.0330
       Total |  13755956.4        29  474343.323   Root MSE        =    677.28
------------------------------------------------------------------------------
       D.cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |
         D1. |   114.3877   81.12535     1.41   0.170    -51.79004    280.5654
             |
       _cons |  -76.02649    196.934    -0.39   0.702    -479.4275    327.3745
------------------------------------------------------------------------------

. estat bgodfrey
Breusch-Godfrey LM test for autocorrelation
----------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+--------------------------------------------------------------
       1     |          0.015               1                   0.9013
----------------------------------------------------------------------------
                       H0: no serial correlation
```

(4) Remedial measures

Now we can try excluding the constant term.

```
      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 29)        =      3.22
       Model |  1432375.26         1  1432375.26   Prob > F        =    0.0833
    Residual |  12912337.4        29  445253.014   R-squared       =    0.0999
-------------+----------------------------------   Adj R-squared   =    0.0688
       Total |  14344712.7        30  478157.089   Root MSE        =    667.27

------------------------------------------------------------------------------
       D.cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |
         D1. |   90.01267   50.18555     1.79   0.083     -12.6283    192.6536
------------------------------------------------------------------------------
```

(4) Remedial measures

2. Estimating ρ

The reason why we want to know the value of ρ is that we can use this value to transform variables, then use the transformed ones in the GLS estimation.

Assume that the error term follows AR(1) scheme,

› $u_t = \rho u_{t-1} + \varepsilon_t$ where $1 < \rho < 1$

we transform $t - 1$ equation by multiply ρ

› $\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1}$

Then, subtract the t with the equation above to remove the coexistence of the element that are the same between time

› $(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t$ or

› $Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t$

There are several methods that we can estimate this value of ρ.

(4) Remedial measures

2.1 ρ based on Durbin-Watson d statistics.

From the Durbin-Watson test, we can derive that

› $\hat{\rho} \approx 1 - \frac{d}{2}$

2.2 ρ estimated from the residuals

We can also estimate another model postestimation,

› $\hat{u}_t = \rho \hat{u}_{t-1} + v_t$

2.3 Cochrane-Orcutt iterative procedure

Iterative method estimates ρ by starting at some value, mostly 0, then successively approximate multiple times until the value of ρ is stable. Then, ρ can be put into the transformation.

(4) Remedial measures

2.4 Prais-Winsten transformation

Using the same concept of iterative ρ, Prais-Winsten fixed losing 1 observation from the first-difference method because of no antecedent by adding

› $Y_1\sqrt{1-\rho^2}$ and $X_1\sqrt{1-\rho^2}$

Note that Prais-Winsten transformation will retain original number of observation, which might be very important especially for a small sample dataset.

Compare the results from the original model to the third and the fourth method in the next page.

(4) Remedial measures

› Original model

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =    312.82
       Model |   131519253         1   131519253   Prob > F        =    0.0000
    Residual |  12192677.5        29  420437.155   R-squared       =    0.9152
-------------+----------------------------------   Adj R-squared   =    0.9122
       Total |   143711930        30  4790397.67   Root MSE        =    648.41

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   6.286548    17.69   0.000     98.33017     124.045
       _cons |   -1827.06   507.7518    -3.60   0.001    -2865.529   -788.5911
------------------------------------------------------------------------------
```

(4) Remedial measures

› Cochrane-Orcutt iterative procedure

```
      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 28)        =     80.97
       Model |  29962164.9         1  29962164.9   Prob > F        =    0.0000
    Residual |  10360869.6        28  370031.057   R-squared       =    0.7431
-------------+----------------------------------   Adj R-squared   =    0.7339
       Total |  40323034.5        29  1390449.47   Root MSE        =     608.3

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   108.1161   12.01497     9.00   0.000     83.50453    132.7276
       _cons |  -1644.838   999.9944    -1.64   0.111    -3693.234     403.558
-------------+----------------------------------------------------------------
         rho |   .4680855
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    1.065525
Durbin-Watson statistic (transformed) 1.298594
```

(4) Remedial measures

› Prais-Winsten transformation

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =     84.10
       Model |  30077374.3         1  30077374.3   Prob > F        =    0.0000
    Residual |  10371295.2        29  357630.871   R-squared       =    0.7436
-------------+----------------------------------   Adj R-squared   =    0.7348
       Total |  40448669.5        30  1348288.98   Root MSE        =    598.02

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   107.0018   10.16464    10.53   0.000     86.21275    127.7908
       _cons |  -1546.283   820.4299    -1.88   0.070     -3224.25     131.685
-------------+----------------------------------------------------------------
         rho |   .4720733
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    1.065525
Durbin-Watson statistic (transformed) 1.297453
```

(4) Remedial measures

3. Newey-West method

This method does not deal with autocorrelation directly, instead, it is very much like White's robust standard error.

The corrected standard errors are known as **HAC standard error.** (heteroscedasticity and autocorrelation).

Newey-West approach is strictly speaking valid in large samples.

```
Regression with Newey-West standard errors      Number of obs     =         31
maximum lag: 1                                  F(  1,        29) =     208.18
                                                Prob > F          =     0.0000

------------------------------------------------------------------------------
             |             Newey-West
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   7.706168    14.43   0.000     95.42672    126.9485
       _cons |   -1827.06   576.8485    -3.17   0.004    -3006.848   -647.2726
------------------------------------------------------------------------------
```