# Simple Linear Regression

## Part 3

**EE325**

**Introductory Econometrics**

Revision Aug 2020

## List of the topics to cover

Let's come up with an example of a hypothesis and models.

- Statement of hypothesis or theory.
- Specification of mathematical model of the theory.
- Specification of statistical or econometric model.
- Obtaining the data.
- Estimation of the parameters of the econometric model.

## List of the topics to cover

- Hypothesis testing
- Forecasting or prediction.
- Using the model for control or policy purpose.

# (1)Simple linear regression

G. 35

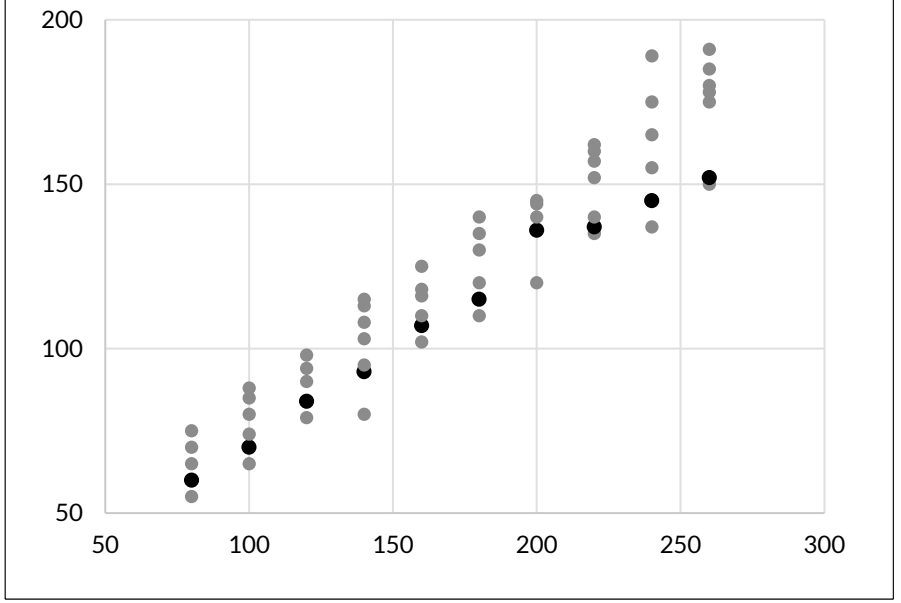
Now let's look at some data. Given that

- $X$ - weekly income of ($) and
- $Y_i$ - weekly expenditure of each household.

There are total of 7 households which the data is collected.

| $X$ | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ | $Y_6$ | $Y_7$ |
|---|---|---|---|---|---|---|---|
| 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 |

*Plotted data*




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

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

So the $\beta_1$ and $\beta_2$ are the **parameter** or

## (2) Notes on linearity

G. 38

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 $\beta_i$.

For example,

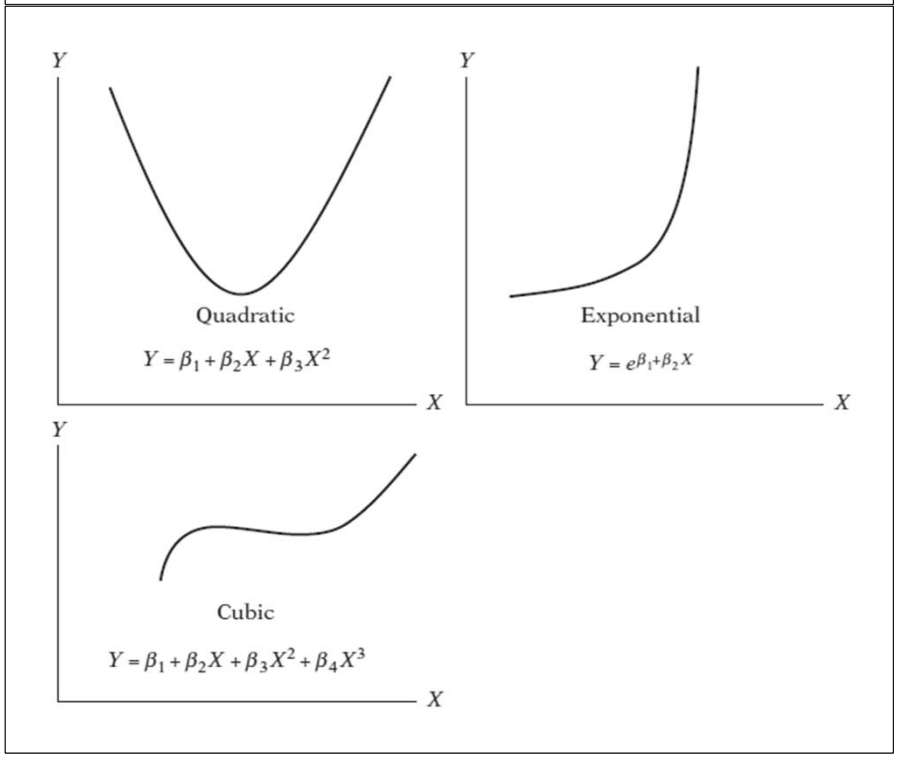
$$E(Y|X_i) = f(X_i) = \beta_1 + \beta_2 X_i^2$$

$$E(Y|X_i) = f(X_i) = \beta_1 + \beta_2^2 X_i$$

This table below sums up how we pick a regression model.

| **LiP?** | LiV? | |
|---|---|---|
| | Yes | No |
| **Yes** | **LRM** | **LRM** |
| No | NLRM | NLRM |

*Linear in parameter and variable*



## (3) Stochastic specification of PRF

G. 39

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,

(1) **Systematic** or **deterministic** part – which is $\beta_1 + \beta_2 X_i$.

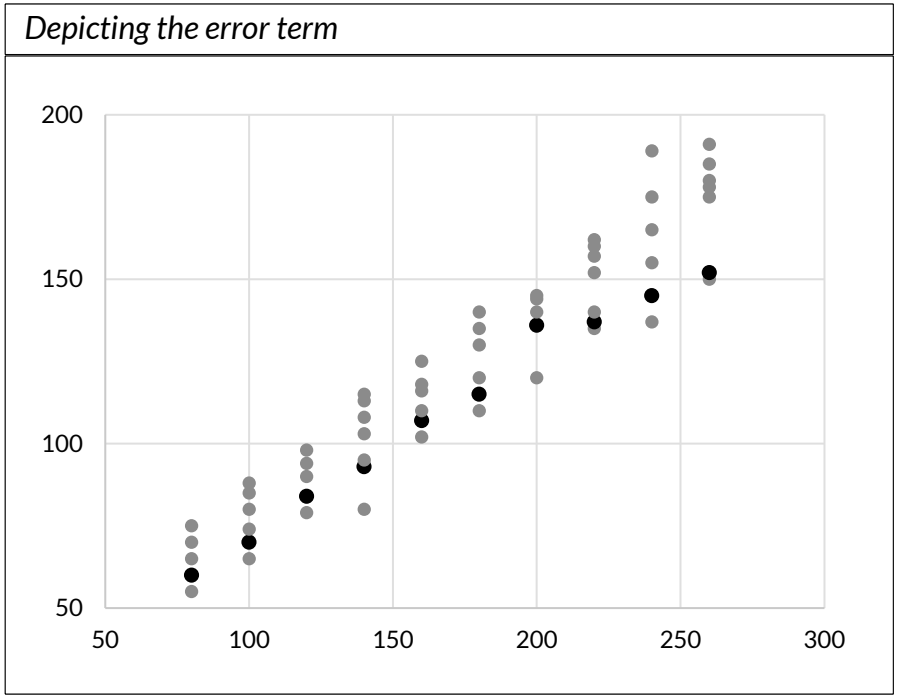
(2) **Random** or **nonsystematic** part – which is $u_i$.

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 $u_5 =$

Please read Gujarati Page 41-42 to understand the importance of the error term.

*Depicting the error term*



## (4) Expected value of the error term



With the error term included in the PRF, taking the expected through the whole equation we get,

- $E(Y_i|X_i) = E[E(Y|X_i)] + E(u_i|X_i)$

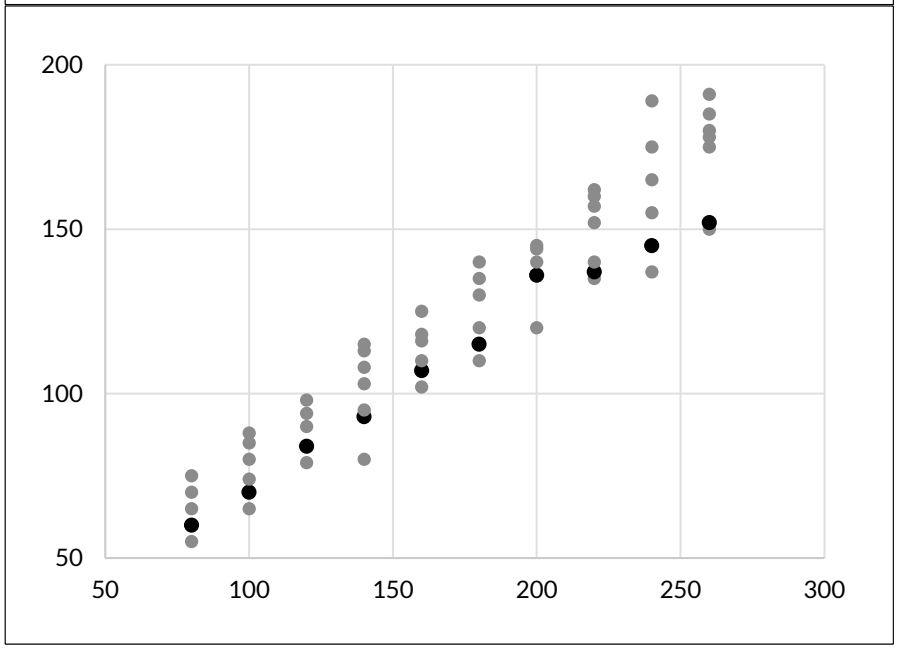
Since the $E(Y|X_i)$ is a constant, therefore,

-

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

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

*Zero expected value of the error term*



# (1) Sampling from a population

G. 42

When we try to imply from a sample to a population, we define our **sample regression function** (SRF), as

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

where $\hat{Y}_i$ is the estimator of $E(Y|X_i)$. To be equivalent, we can write the SRF in stochastic form as

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

To be clear, we call $\hat{\beta}_i$ as **estimated coefficient** (or coefficient in short) or the **estimator**, while $\hat{u}_i$ is called **residual**. The following tables sum the naming scheme.

| **Non-stochastic** $E(Y\|X_i)$ | Population $E(Y\|X_i) = \beta_1 + \beta_2 X_i$ | Sample $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ |
|---|---|---|
| $Y$ | $E(Y\|X_i)$ | $\hat{Y}_i$ (estimator) |
| $\beta$ | $\beta_1, \beta_2$ (parameter) | $\hat{\beta}_1, \hat{\beta}_2$ (estimator, coef) |

| **Stochastic** $Y_i$ | Population $Y_i = \beta_1 + \beta_2 X_i + u_i$ | Sample $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$ |
|---|---|---|
| $Y$ | $Y_i$ | $Y_i$ |
| $\beta$ | $\beta_1, \beta_2$ (parameter) | $\hat{\beta}_1, \hat{\beta}_2$ (estimator, coef) |
| $u$ | $u_i$ | $\hat{u}_i$ (estimator) |

To sum up all the problem we are trying to solve, we are trying to estimate the PRF or

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

on the basis of the SRF or

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

Now the fundamental problem of sampling is consistency. Consider the problem of these two samples of weekly income and expenditure.

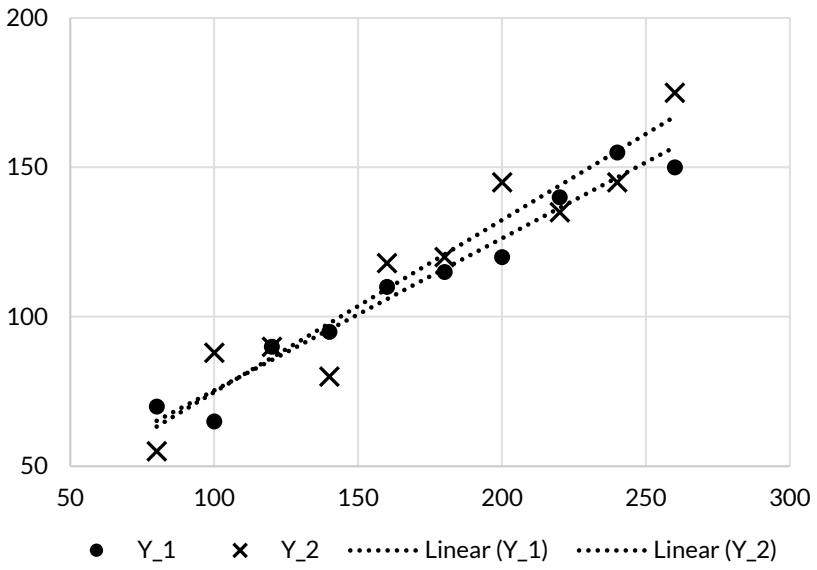
| $Y_1$ | $X$ | $Y_2$ | $X$ |
|---|---|---|---|
| 70 | 80 | 55 | 80 |
| 65 | 100 | 88 | 100 |
| 90 | 120 | 90 | 120 |
| 95 | 140 | 80 | 140 |
| 110 | 160 | 118 | 160 |
| 115 | 180 | 120 | 180 |
| 120 | 200 | 145 | 200 |
| 140 | 220 | 135 | 220 |
| 155 | 240 | 145 | 240 |
| 150 | 260 | 175 | 260 |

# (1) Sampling from a population



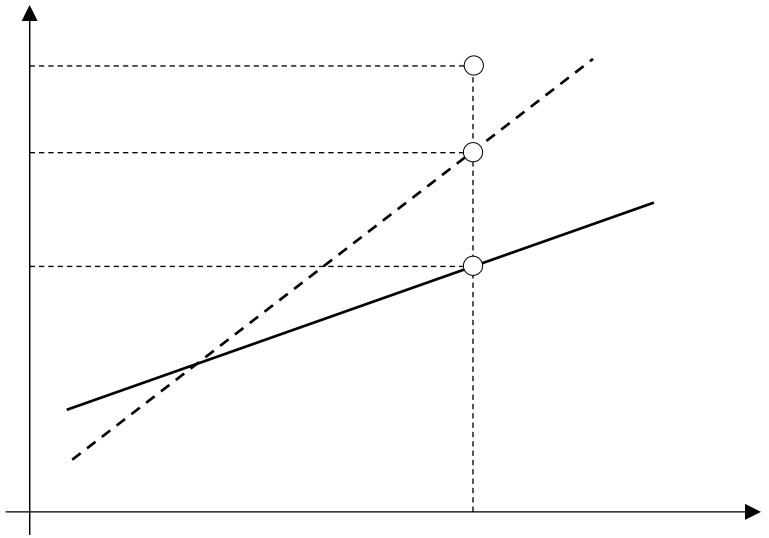
If we plot our two samples here, we get this graph

*Two samples SRF*




If we plot our two samples here, we get this graph

*Comparing PRF and SRF*

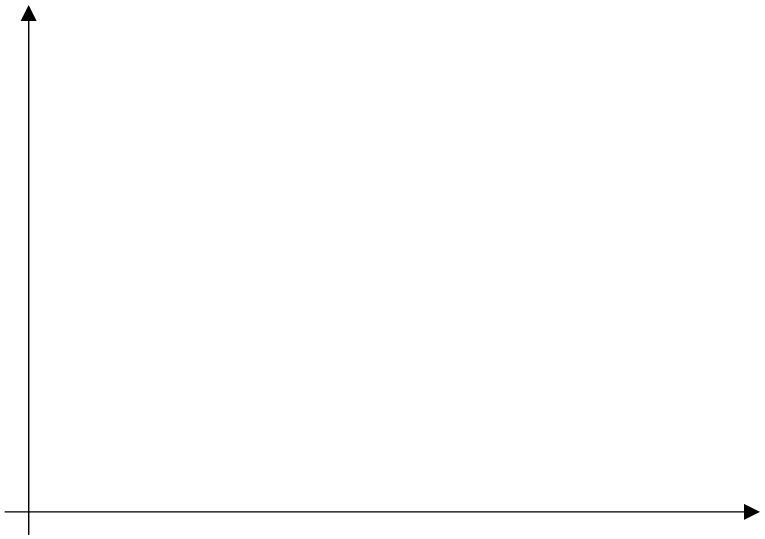


## (2) Estimating coefficients: Ordinary Least Square (OLS)



The intuition of the OLS is shown in the graph below.

*Error terms*



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

Solving the for $\hat{\beta}_1$ and $\hat{\beta}_2$

## (2) Estimating coefficients: Ordinary Least Square (OLS)

G. 59

Eventually, we get

- $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$
- $\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{X})^2}$

# (2) Estimating coefficients: Ordinary Least Square (OLS)



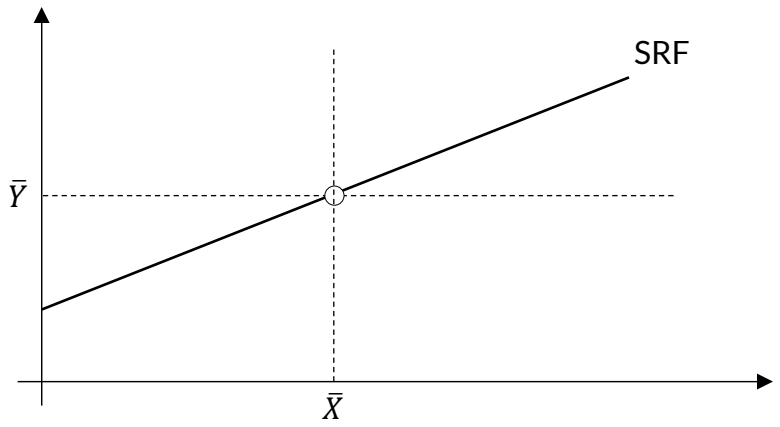
**Properties of OLS estimators**

(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.

*SRF passing 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 $\hat{u}_i = 0$.

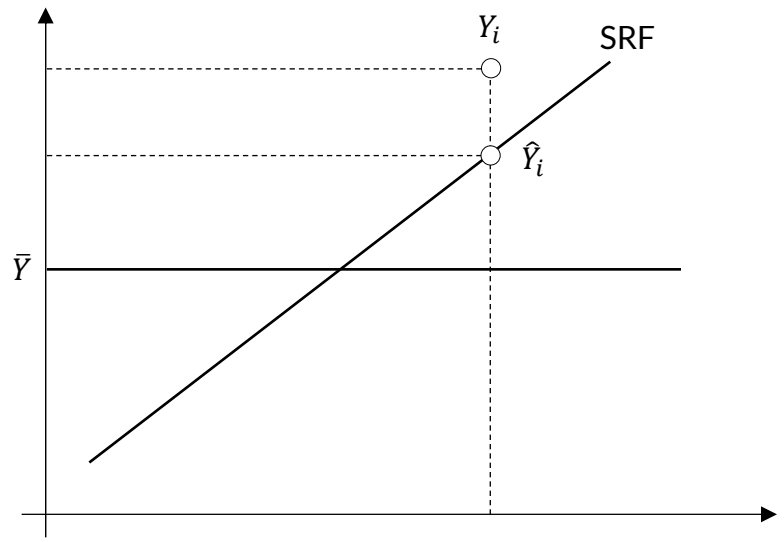
(6) $\hat{u}_i$ are uncorrelated with both $X$ and $\hat{Y}$.

## (3) The 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, $\bar{Y}$.

*Defining the error terms*




$r^2$ can be defined in multiple ways. The intuition is that total sum of squares (TSS) is equal to explained sum of squares (ESS) and residual sum of squares (RSS) or

- $TSS = ESS + RSS$ of if we divide with $TSS$, we have

Eventually, we get

- $r^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(y_i - \bar{Y})^2} = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}$

**Properties of $r^2$**

(1) Nonnegativity

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

## (3) The coefficient of correlation (r)



We can also derive the **sample correlation coefficient** from $r^2$ simply by putting a radical over it as

- $r = \pm\sqrt{r^2}$

An important note is that, even $r$ initiated from the concept of $\sigma$, the former represents sample's association, while the latter represents population's.

Eventually, we get

- $r = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i^2)}}$

**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.

(6) Does not imply causality.

# (1) Assumptions underlying classical linear regression model (CLRM)



(1) Linear in parameter.

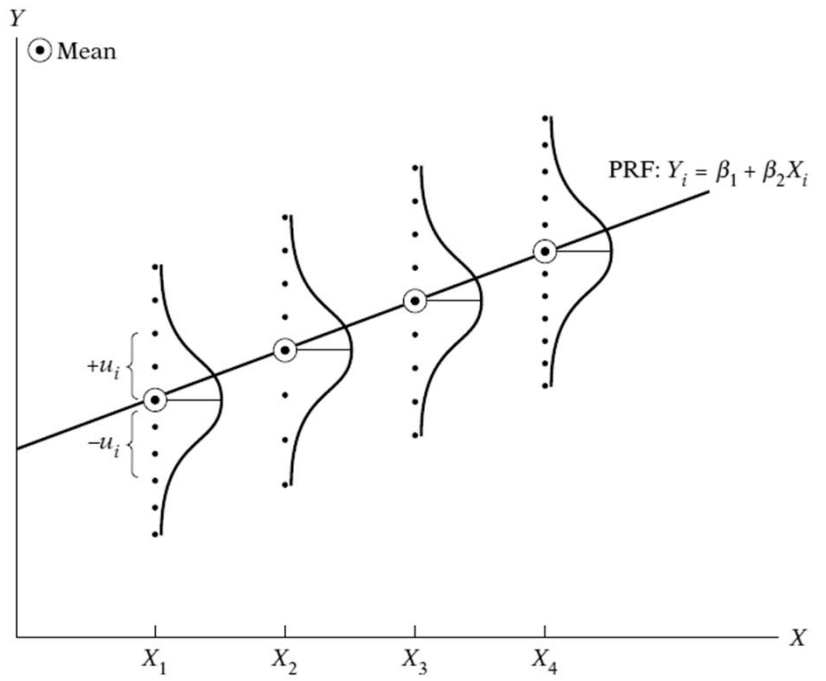
(2) $X$ values are non-stochastic or is independent of $u$ or

- $cov(x_i, u_i) = 0$

(3) No specification error or no omitted variables or

- $E(u_i|x_i) = 0$ or $E(u_i) = 0$

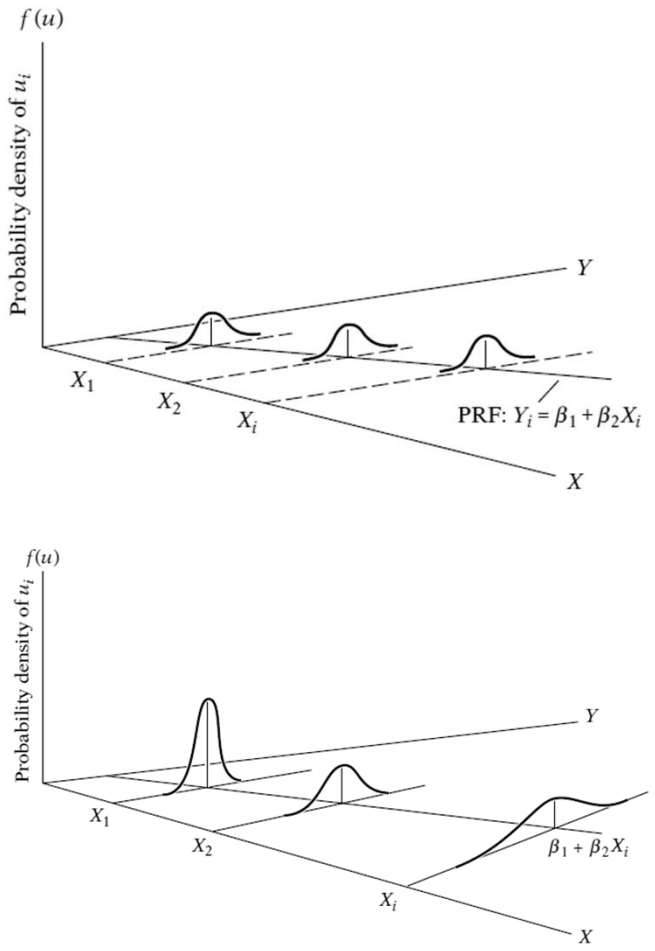
*Zero expected value of $u_i$*




(4) Homoscedasticity or constant variance of $u_i$ or

- $var(u_i) = \sigma^2$

*Homoscedasticity and heteroscedasticity*



# (1) Assumptions underlying classical linear regression model (CLRM)

G. 66 & 98

(5) No autocorrelation or serial correlation between disturbances or

• $cov(u_i, u_j | x_i, x_i) = 0$ or $cov(u_i, u_j) = 0$ if $X$ is non-stochastic

(6) Number of observation $n$ must be greater than the number of parameter estimated $k$

(7) $X$ value must not all be the same.

**Further properties of OLS estimators**

Since the distribution of the estimators is based on how the error term, $u_i$, is distributed, the assumptions of **normality** and **Central Limit Theorem** (CLT) help us simplify the problem.

First, we deal with the **normality** assumption for $u_i$. The CLRM assumes that each $u_i$ is distributed normally with

• Expected value: $E(u_i) = 0$

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

• Covarince: $E(u_i, u_j) = 0$

In short, we can use the notation $u_i \sim NID(0, \sigma^2)$ to represent the normality of $u_i$, or $u_i$ is normally and independently distributed.

Then, the **Central Limit Theorem** (CLT) states that

• Let $X_1, X_2, \ldots, X_n$ denote $n$ independent random variables distributed with the same PDF with mean of $\mu$ and variance of $\sigma^2$, as $n$ increases indefinitely, then

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

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

## (1) Assumptions underlying classical linear regression model (CLRM)

G. 100

We also have the **Gauss-Markov Theorem**, that the OLS estimator $\hat{\beta}_2$ (also for $\hat{\beta}_1$) is said to be a best linear unbiased estimator (**BLUE**) of $\beta_2$ if the following hold

• It is linear, that is, a linear function of a random variable, such as the dependent variable $Y$ in the regression model.

• It is unbiased, that is, its average or expected value, $E(\hat{\beta}_2)$ is equal to the true value $\beta_2$.

• 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.

Proof will be skipped at this point, just to lead you to these properties at last. For $\hat{\beta}_1$, they have

• Mean: $E(\hat{\beta}_1) = \beta_1$

• Variance: $\sigma^2_{\hat{\beta}_1} = \frac{\sum x_i^2}{n \sum x_i^2} \sigma^2$

or we can say shortly that

• $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$

Note that the $\sigma^2$ refers to the variance of error term $u_i$. We can normalize the distribution as the standard normal as follows

• $Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$

The same holds true for $\hat{\beta}_2$, they have

• Mean: $E(\hat{\beta}_2) = \beta_2$

• Variance: $\sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2}$

or we can say shortly that

• $\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})$

in which we can normalize it as well

• $Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \sim N(0,1)$

While the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$ is defined as

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

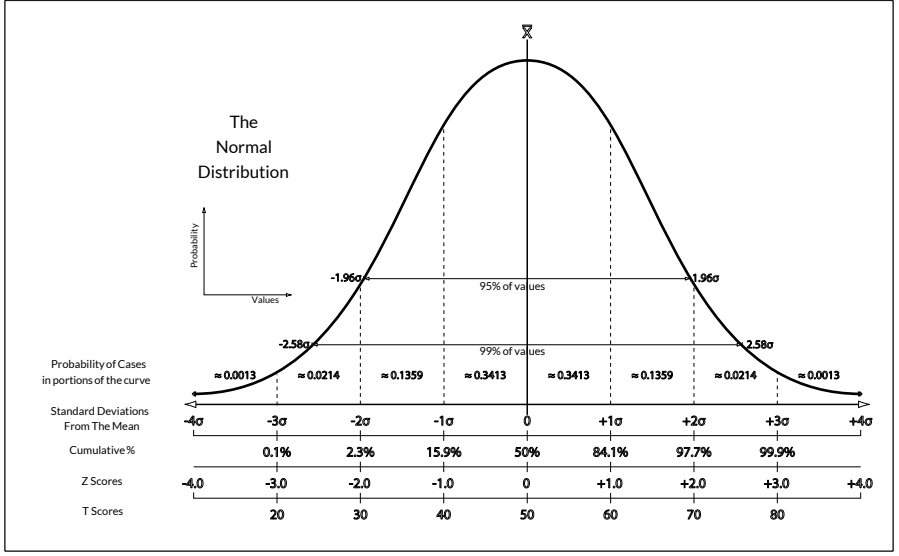
# (2) Interval estimation

G. 108

We have only dealt with **point estimation**, a problem arises therefore, how reliable the point estimate is? In statistics, the reliability of a point estimate is measured by its standard deviation.

Each type of distribution has its own characteristic. For instance, a normal distribution, 95% of the value falls between $\pm\ 2\sigma$ from the mean while 99% of the value falls between $\pm 3\sigma$ from the mean.

*Normally distributed area (standard normal)*




If $\hat{\beta}_1$ and $\hat{\beta}_2$ is normally distributed, we can make sure that how **likely** the mean will be within a specific range. A selected range is called **confidence interval** (CI).

How wide the CI will be depends on at what **level of significance** a researcher is able to accept. Here are some of the definitions

- Level of significance: $\alpha$ for $0 < \alpha < 1$
- Confidence coefficient: $1 - \alpha$

A researcher can make sure, statistically, that the point estimator, here for example is $\hat{\beta}_2$, will be within a range of CI or

- $P\left(\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta\right) = 1 - \alpha$

The most common level of significance are 0.1, 0.05, 0.01 or the confidence coefficient 0.9, 0.95, 0.99 (we sometimes say 90,95,99 percent) respectively.

The larger confidence coefficient $1 - \alpha$ is, the CI will also be larger, on the other hand, the lower level of significance $\alpha$, the CI will be larger.

## (2) Interval estimation

G. 109

**Confidence intervals for $\widehat{\beta}_1$ and $\widehat{\beta}_2$**

Since we know that both $\hat{\beta}_1$ and $\hat{\beta}_2$ is normally distributed with a mean and a variance, the normalized version (for $\hat{\beta}_2$) is

- $Z = \frac{\widehat{\beta}_2 - \beta_2}{\sigma_{\widehat{\beta}_2}} \sim N(0,1)$ where $\sigma^2_{\widehat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2}$ or
- $Z = \frac{(\widehat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\sigma} \sim N(0,1)$

Now we have a major problem because is $\sigma^2$ rarely known and so the $\sigma$, hence we need to instead use the estimated version (from G. page 94) or $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}$. Moreover, when the variance of the error is unknown, the Z-distribution turns into t-distribution as follows.

- $t = \frac{(\widehat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \sim t_{n-2}$

Since $\sigma^2_{\widehat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2}$, then $\sigma_{\widehat{\beta}_2} = \frac{\sigma}{\sqrt{\sum x_i^2}}$. Therefore, $\frac{\sqrt{\sum x_i^2}}{\sigma} = \frac{1}{\sigma_{\widehat{\beta}_2}}$, replacing $\sigma$ with $\hat{\sigma}$ we get

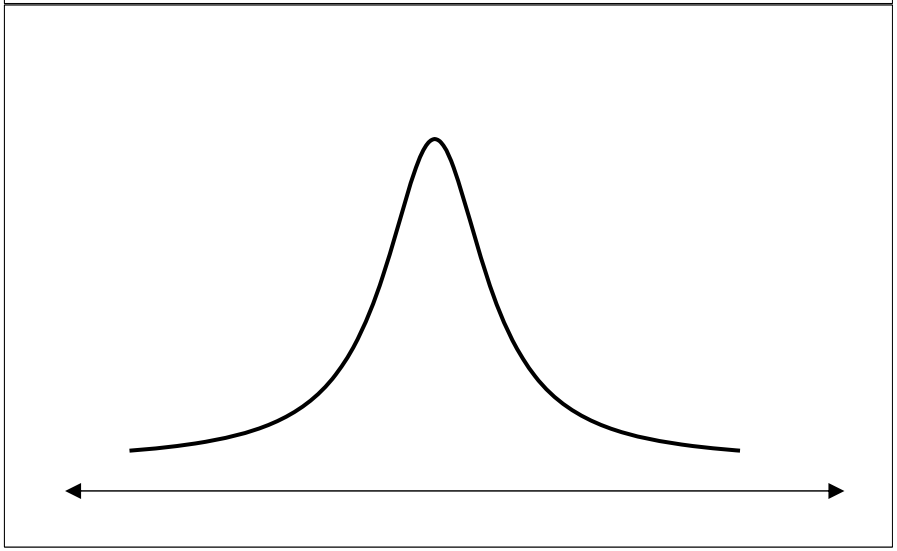
- $t = \frac{\widehat{\beta}_2 - \beta_2}{\sigma_{\widehat{\beta}_2}} \sim t_{n-2}$

If we want to come up with a **two-tail** CI with a level of significance of $\alpha$, bring back the probability function for t-distribution

- $P\left(-t_{\frac{\alpha}{2}} \leq t \leq t_{\frac{\alpha}{2}}\right) = 1 - \alpha$

Supposed we pick $\alpha = 0.05$, it means that we are to test that our estimator $t$ will be within the range between the **critical value** of $-t_{\frac{\alpha}{2}}$ and $t_{\frac{\alpha}{2}}$ 95 percent of the time we sample.

*CI in a t-distribution: two-tail CI*



## (2) Interval estimation

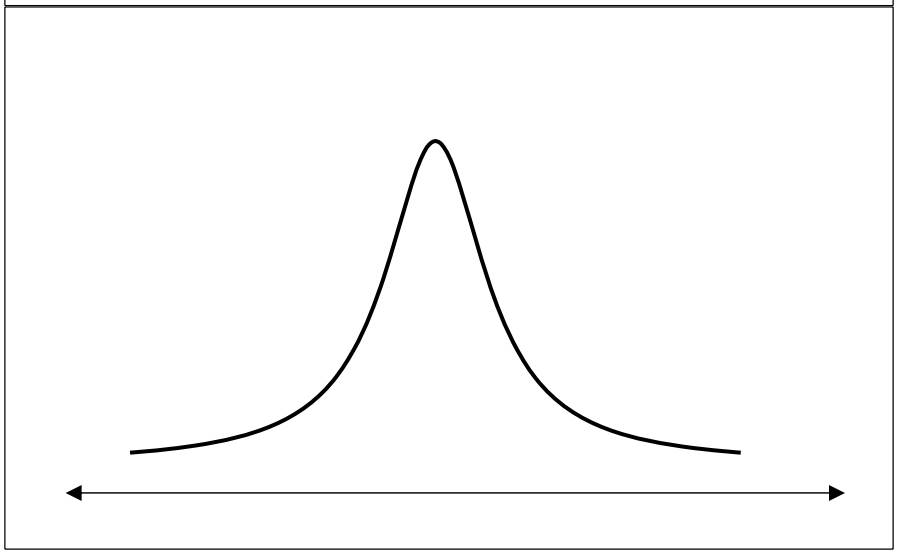
G. 110

On the other hand, a **one-tail** CI with a level of significance of $\alpha$ is

- $P(t \leq t_\alpha) = 1 - \alpha$ or $P(t \geq t_\alpha) = 1 - \alpha$

Supposed we pick $\alpha = 0.05$, it means that we are to test that our estimator $t$ will be lower (or higher) than the critical value $t_\alpha$ 95 percent of the time we sample.

*CI in a t-distribution: one-tail CI*



Bring back the interval estimation of $\hat{\beta}_2$ as

- $t = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \sim t_{n-2}$

and supposed we are testing with two-tail to see if the $\hat{\beta}_2$ falls within an acceptable range of $\beta_2$ plus and minus the error

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

Rearranging to have the central term as $\beta_2$, we have

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

So we have the $100(1-\alpha)\%$ CI for $\hat{\beta}_2$ as

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

Analogously, the $100(1-\alpha)\%$ CI for $\hat{\beta}_1$ is

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

Note that sometimes $\sigma_{\hat{\beta}_1}$ and $\sigma_{\hat{\beta}_1}$ are represented in a term of $se(\hat{\beta}_1)$ and $se(\hat{\beta}_2)$ respectively.

## (2) Interval estimation

G. 111

We can create CI to test for significance of $\sigma^2$ as well. We know that $u_i$ is assumed to be normally distributed, without any proof, we define that

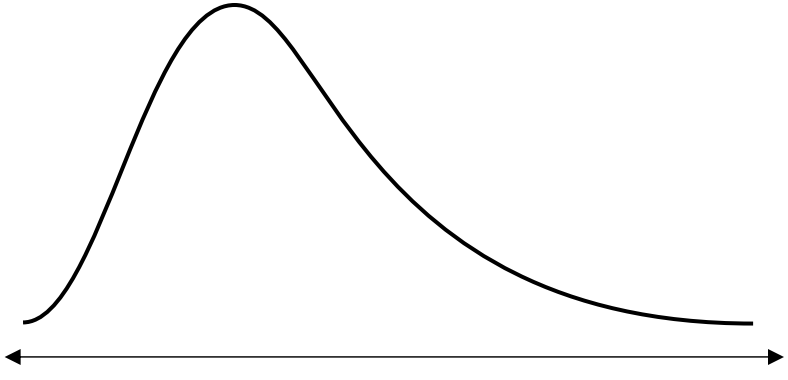
- $(n-2)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$

and the CI for Chi-square is defined as

- $P\left(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}\right) = 1-\alpha$ then replacing the test stat and rearrange terms

- $P\left((n-2)\frac{\hat{\sigma}^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq (n-2)\frac{\hat{\sigma}^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$

*Chi-square distribution and its critical value*



The critical value for both side is different because Chi-square is not a symmetrical distribution. Hence, two-tails testing will have a different critical value. Similarly, one-tail testing is also different when we test on the left and on the right.

The interpretation for this CI is if we establish 95 confidence limit on $\sigma^2$ and if we maintain that these limits will include the true $\sigma^2$, we will be right for 95 percent of the time.

# (1) Testing the coefficients

G. 113

Follow these steps of hypothesis testing

**(1) State your hypothesis**: usually we will have a **null hypothesis** and the **alternative hypothesis**. For example, consider this two-tails test of $\beta_2 = \beta_2^*$ when $\beta_2^* = 0$

- $H_0$: $\beta_2 = 0$ – null hypothesis
- $H_a$: $\beta_2 \neq 0$ – alternative hypothesis

**(2) Calculate the test statistics**: for this case, we rely on t-test. E.g. if $\hat{\beta}_2 = -2.1576$, $\sigma_{\hat{\beta}_2} = 0.1204$ and $n = 10$, plug these value into

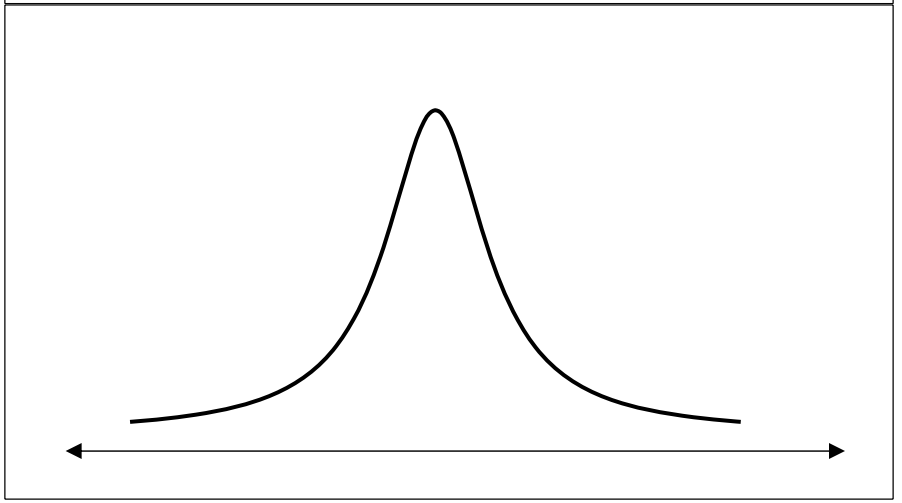
- $t^* = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} =$

where $t^*$ denotes calculated test statistics

**(3) State your decision rule**: picking an $\alpha$ to have an acceptable probability, most of the time we use $\alpha = 0.05$. Use this information to create CI, based on the degrees of freedom. For this case, we test against zero, or $\beta_2^* = 0$, the CI becomes simpler as

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

*Two-tails test for $\beta_2$*



**(4) Conclude the test**: for this case, if

- the $t^*$ lies beyond any boundary of CI, we can **reject the null hypothesis**, at the significance level of 95%. In other words, with 95 percent confidence, $\beta_2$ is not zero.
- the $t^*$ lies within the CI, **we cannot reject the null hypothesis**. In other words, we **cannot say for sure** that 95 percent $\beta_2$ is not zero.

**Important note!** 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**.

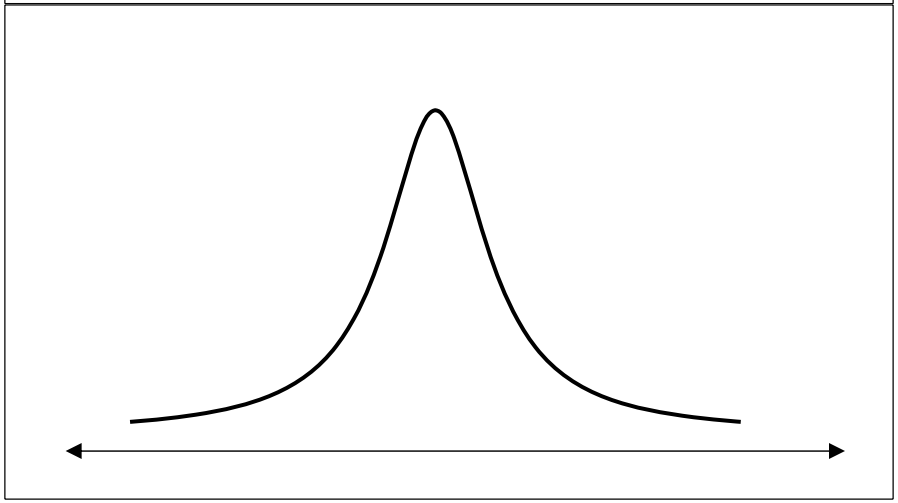
## (1) Testing the coefficients

**#Example**: given that

- $\hat{\beta}_2 = 0.7240$
- $\sigma_{\hat{\beta}_2} = 0.7$
- $n = 13$
- $\alpha = 0.01$

Test that $\beta_2 = 0$ or not.

*Two-tails test for $\beta_2$*



## (1) Testing the coefficients

G. 115

Now try doing the one-tail test.

**(1) State your hypothesis**:

- $H_0$: $\beta_2 \leq 0$
- $H_a$: $\beta_2 > 0$

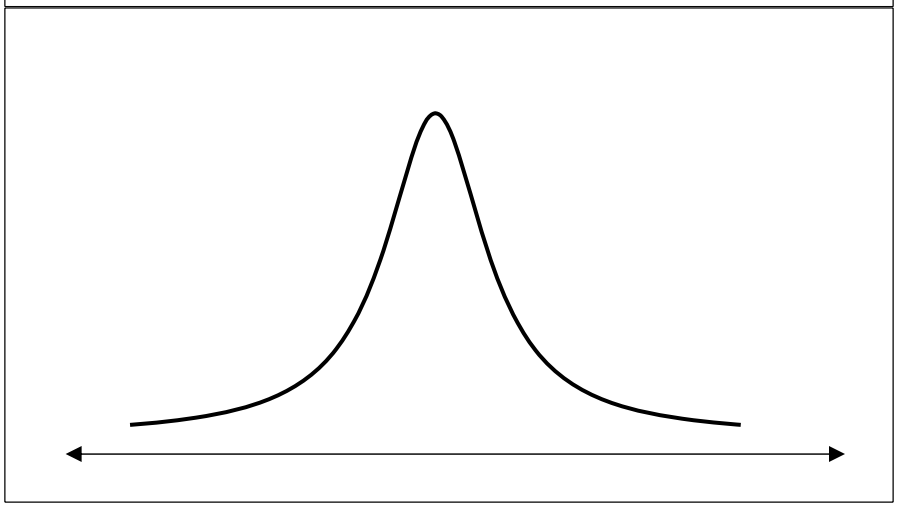
**(2) Calculate the test statistics**: for this case, we rely on t-test. E.g. if $\hat{\beta}_2 = -2.1576$, $\sigma_{\hat{\beta}_2} = 0.1204$ and $n = 10$, plug these value into

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

**(3) State your decision rule**:

- The upper bound: $t_\alpha =$

*One-tail test for $\beta_2$*



**(4) Conclude the test**: for this case, if

- the $t^*$ lies beyond any boundary of CI, we can **reject the null hypothesis**, at the significance level of 95%. In other words, we can make sure that 95 percent of the time that we sample, $\beta_2$ will not be zero.

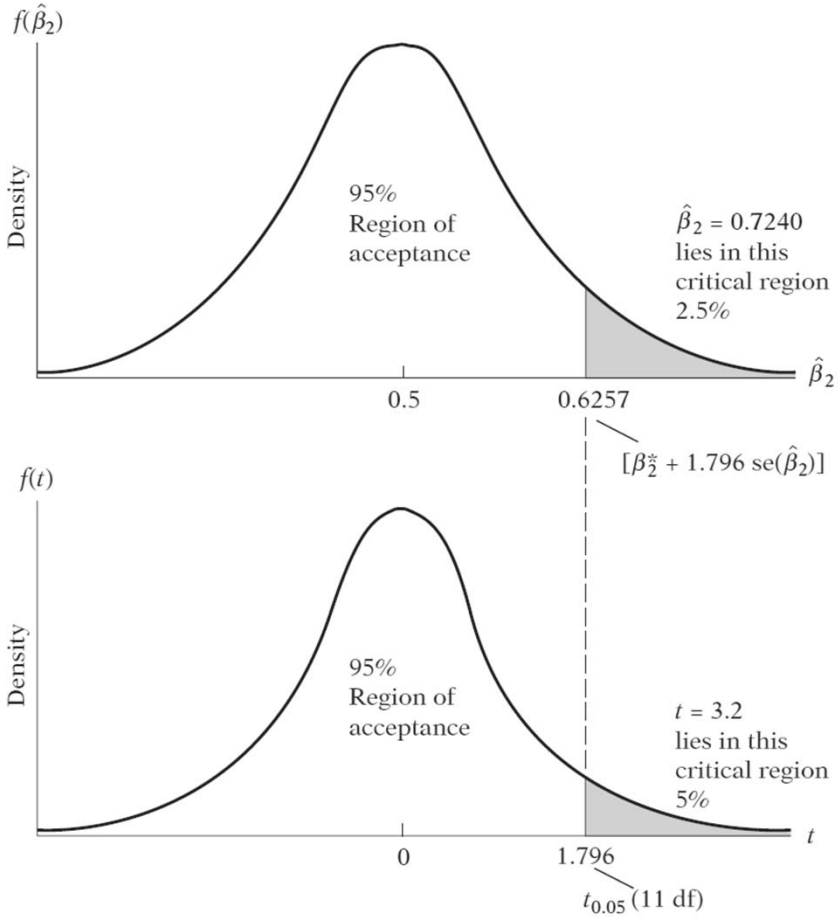
- the $t^*$ lies within the CI, **we cannot reject the null hypothesis**. In other words, we **cannot say for sure** that 95 percent of the time that we sample, $\beta_2$ will not be zero.

**Important note!** 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**.

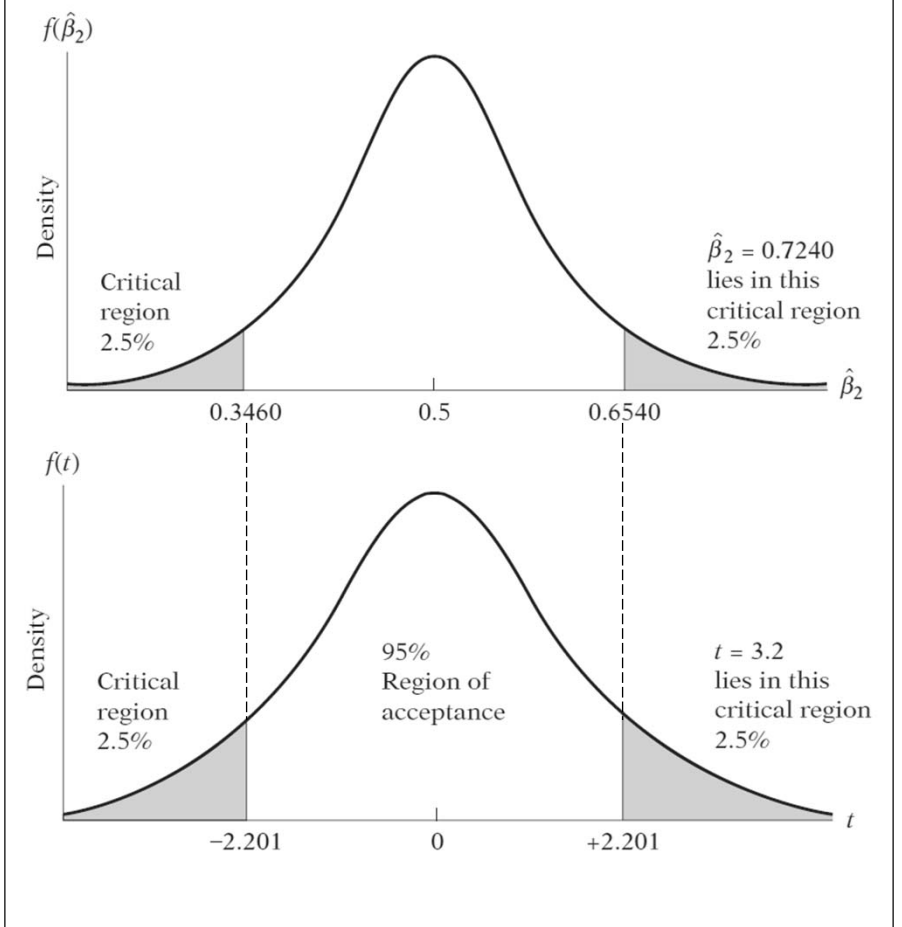
# (1) Testing the coefficients

## Comparing testing against zero and non-zero

*One-tail test*




*Two-tails test*



## (1) Testing the coefficients

**Drawing the test statistics in a regression graph.**

You should also read about statistical significance further in G. page 119-124.

## (1) Reporting regression results

G. 129

Reporting the regression results can be done in multiple ways. However, important statistics should be available. Let's first look at the 'traditional report'.

- $\hat{Y}_i = -0.0144 + 0.7240X_i$

$se = (0.9317) \quad (0.0700) \quad r^2 = 0.9065$

$t = (-0.0154) \quad (10.3428) \quad d.f. = 11$

$p = (0.987) \quad (0.000) \quad F_{1,11} = 108.30$

Below is the result from a statistical program called 'STATA'

*STATA report*

```
. regress cholesterol time_tv

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =   17.47
       Model |  5.04902329     1  5.04902329            Prob > F      =  0.0001
    Residual |  28.3220135    98  .289000137           R-squared     =  0.1513
-------------+------------------------------           Adj R-squared =  0.1426
       Total |  33.3710367    99  .337081179           Root MSE      =  .53759

------------------------------------------------------------------------------
 cholesterol |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     time_tv |   .0440691   .0105434     4.18   0.000     .0231461    .0649921
       _cons |  -2.134777   1.813099    -1.18   0.242    -5.732812    1.463259
------------------------------------------------------------------------------
```

Most modern econometric articles reveal their results in a table with only essential information.

# (1) Reporting regression results

*Regression result table*

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.

## (2) Prediction

G. 126

Supposed we have a sample regression function as

- $\hat{Y}_i = -0.0144 + 0.7240X_i$

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

### (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, if $X_0 = 20$, then

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

We know that there is a variable within the estimator $\hat{Y}_0$, defined as

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

Replacing the unknown $\sigma^2$ with the unbiased estimator $\hat{\sigma}^2$,

- $t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{\sigma_{\hat{Y}_0}} \sim t_{n-2}$

So we can derive the CI for $E(Y_0|X_0)$ as

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

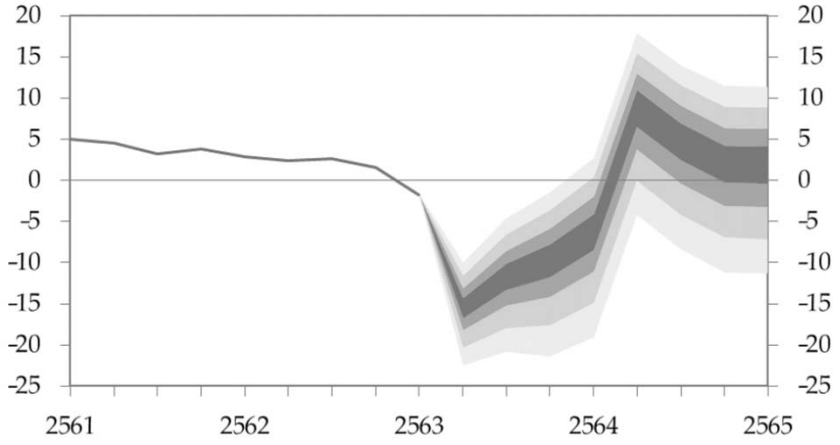
for $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$ and $Y_0 = \beta_1 + \beta_2 X_0$

*BOT Prediction of GDP*

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 |

ร้อยละเทียบกับระยะเดียวกันปีก่อน




*Source: Financial Report, Jun 2020, BOT*

## (2) Prediction

G. 128

**Example**: given that

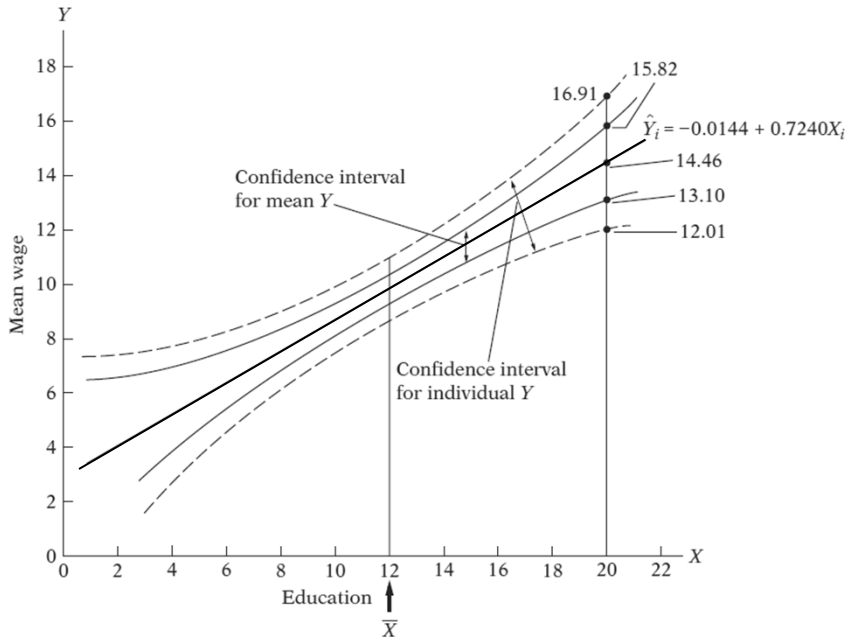
$n = 13, \bar{X} = 12, \sum(x_i - \bar{X})^2 = 182, \hat{\sigma}^2 = 0.8936$

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

- Find the $\sigma_{\hat{Y}_0}$

- Find the 95% CI for $E(Y|X_0 = 20)$

*CI band for mean Y and individual Y values*




**Interpretation**: if we create a CI over the mean value, 95 out of 100 times that the CI will cover true value $Y_0$.

## (2) Prediction

G. 129

### (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 have 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 $\sigma^2$ with the unbiased estimator $\hat{\sigma}^2$,

- $t = \frac{Y_0 - (\hat{\beta}_1 + \hat{\beta}_2 X_0)}{\sigma_{fe}} \sim t_{n-2}$

We can derive the CI for $Y_0$ corresponding to $X_0$

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

**Example**: given that

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

- Find the $Var(fe) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(x_i - \bar{X})^2}\right]$
- Find the $\sigma_{fe}$
- Find the 95% CI for $Y_0$ corresponding to $X_0 = 20$

**Interpretation**: at the confidence level of 95%, when $X_0 = 20$, 95 out of 100 times we will have $Y_0$ between this CI.

## (3) Regression through the origin

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

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

*Regression through the origin*

Obtaining the estimator from OLS, we have

- $\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$
- $Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2}$
- $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1}$

Note that the main difference lies in the d.f., since there is one less term to estimate ($\hat{\beta}_1$).

Differences that should also be noted are

- $\sum \hat{u}_i X_i = 0$ but $\sum \hat{u}_i$ need not to be zero
- $r^2$ can be negative, so we need another coefficient of determination, defined as $\boldsymbol{raw\ r^2}$
- $raw\ r^2 = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$

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.

## (4) Data scaling and units of measurement

G. 156

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

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

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

- Both GPDI and GDP in millions of dollars

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

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

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

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

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

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

$$\widehat{GDPI}_t = -926.090 + 253.524GDP_t$$

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

## (5) Functional forms

G. 159

There are several functional forms that are linear in parameters and can be estimated. Here we will look into:

- Log-linear model
- Semi-log models: log-lin and lin-log
- Reciprocal models
- Logarithmic reciprocal model

**(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.

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

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

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

**(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.

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

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

# (5) Functional forms

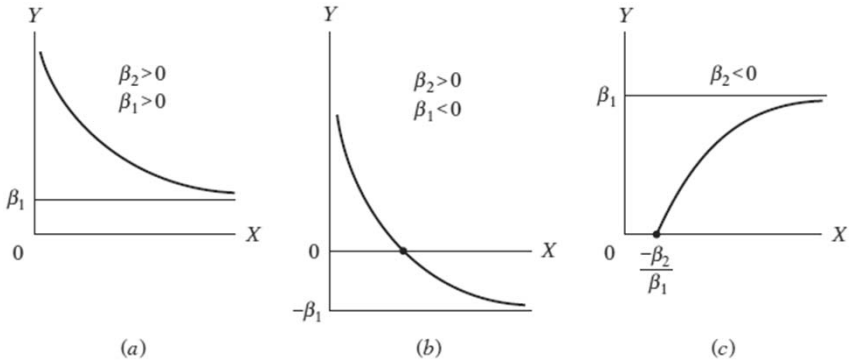
G. 166

### (3) Reciprocal model

Reciprocal model takes a form of

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

*Reciprocal model*




Find the slope and elasticity.

So if $\beta_2$ is positive, the slope is negative, and vice versa.

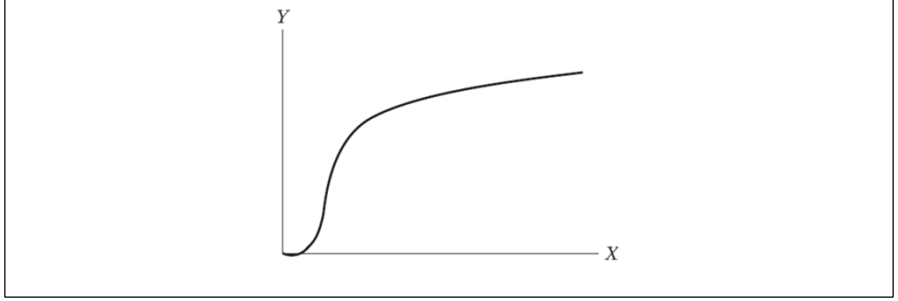
**Example**: Practical examples for reciprocal models are **Child mortality rate** (vs. GNP) and **Phillips curve**.

### (4) Log-reciprocal model

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

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

*Log-reciprocal model*




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

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

The summary of slope and elasticity is in G. page 173. **Be careful** when you interpret $\beta_2$ since it is not straightforward in different functional forms.