(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.

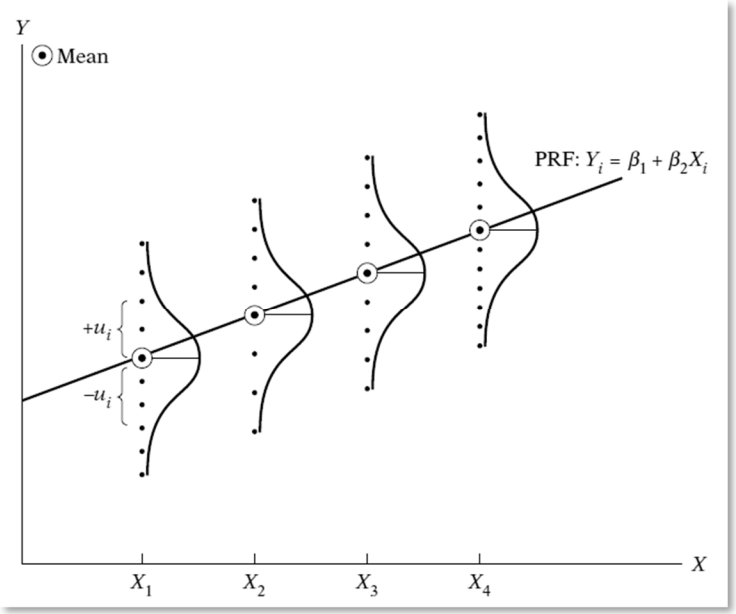
› $Y_i = \beta_1 + \beta_2 X_i + u_i$

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**.

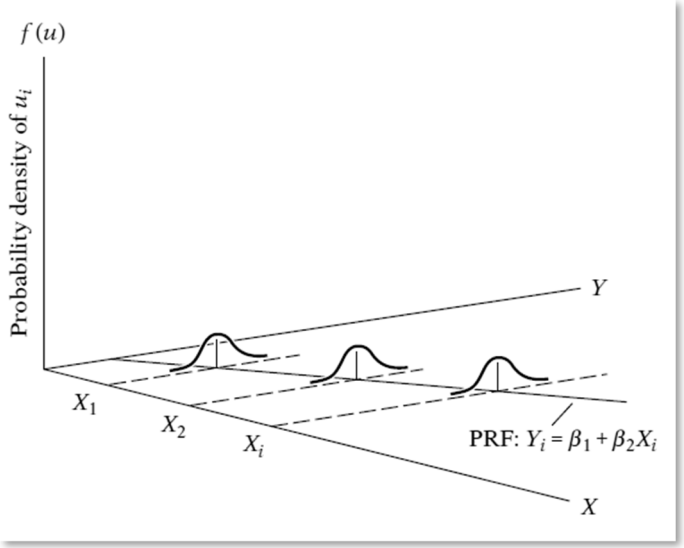


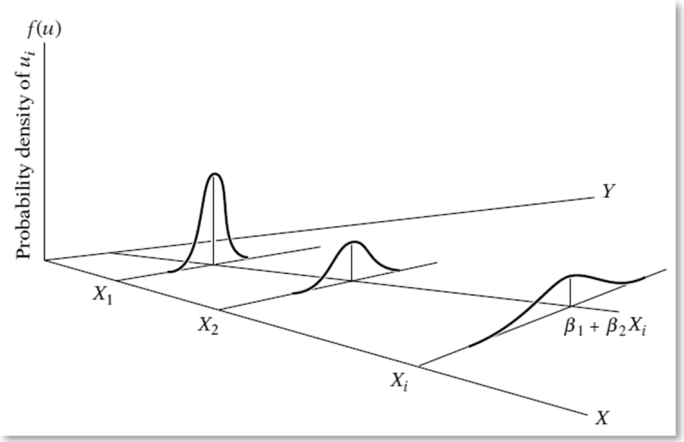
› $E(u_i|X_i) = 0$ or

› $E(u_i) = 0$

(4) Homoscedasticity

Homoscedasticity or constant variance of u_i means that





› $var(u_i) = \sigma^2$

(5) No autocorrelation

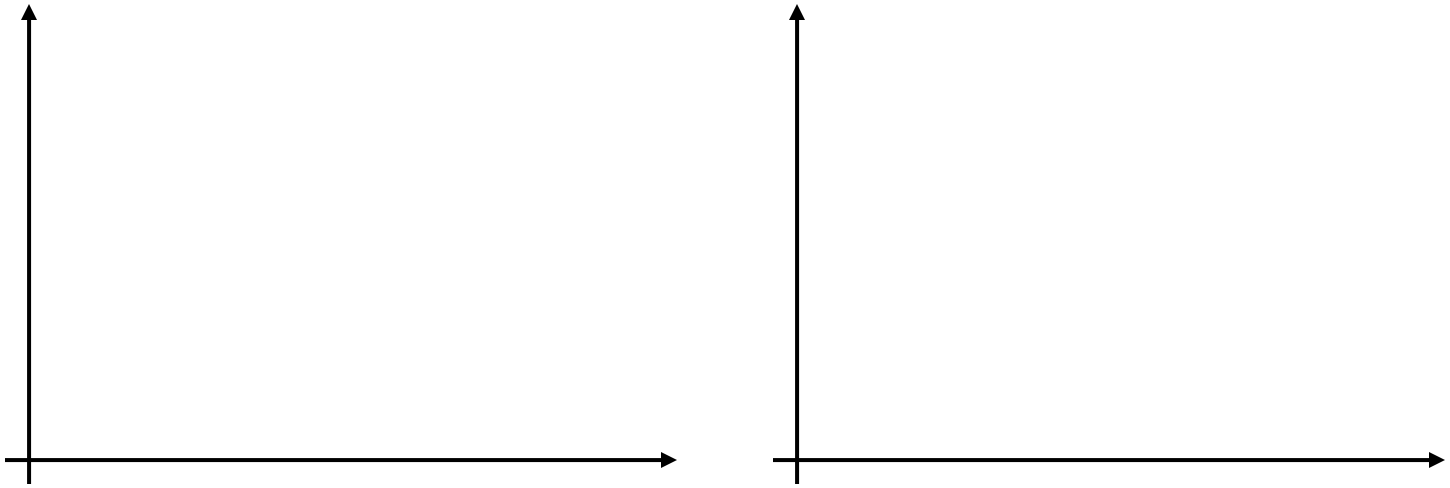
No autocorrelation or serial correlation between disturbances or

› $cov(u_i, u_j | x_i, x_i) = 0$ 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.

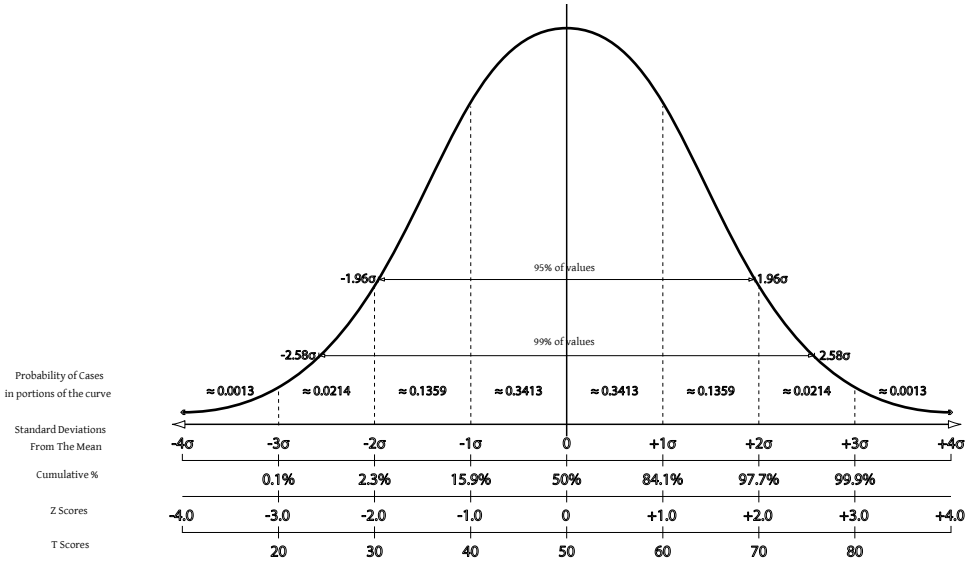
› $P\left[\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta\right] = 1 - \alpha$

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

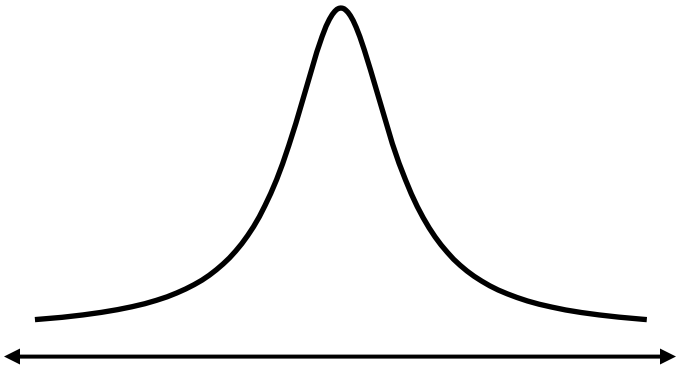
Now we can proceed to calculate the limit, using the formula

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

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.

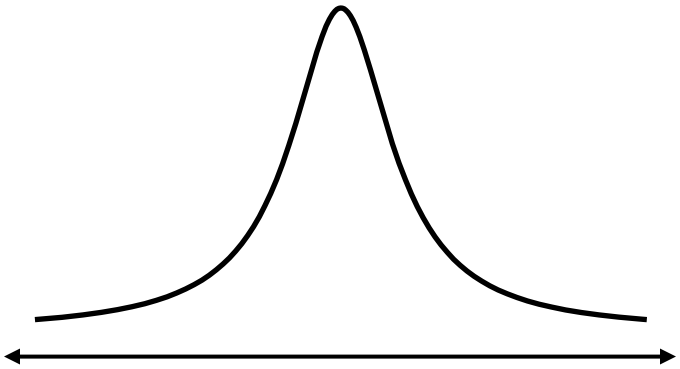
› **Step 1: pick an α**

› **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.

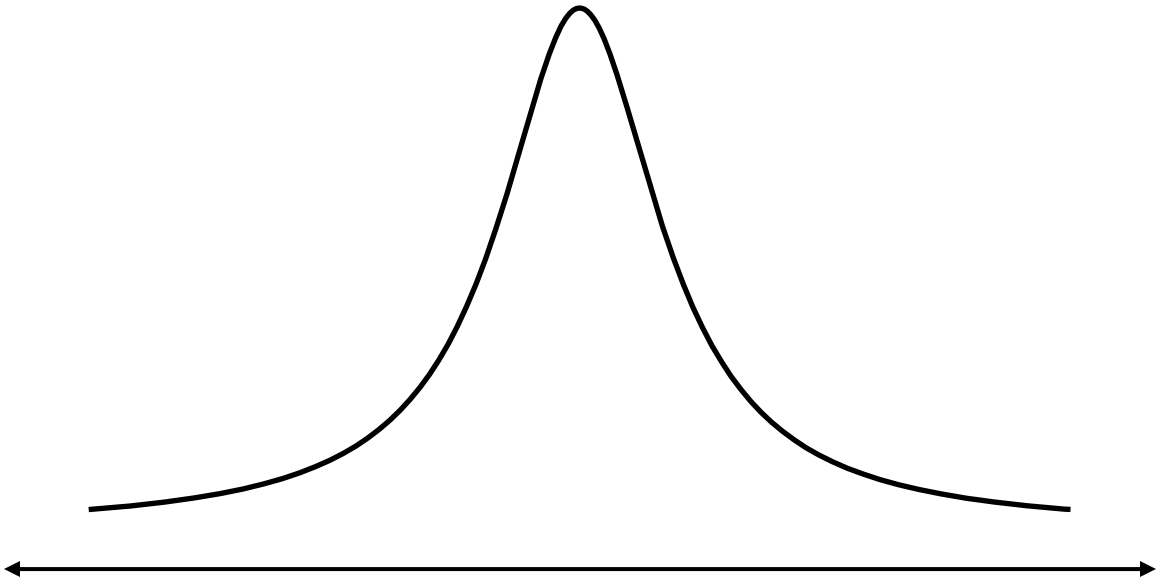
The lower bound : $t_{\frac{\alpha}{2}} =$

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.

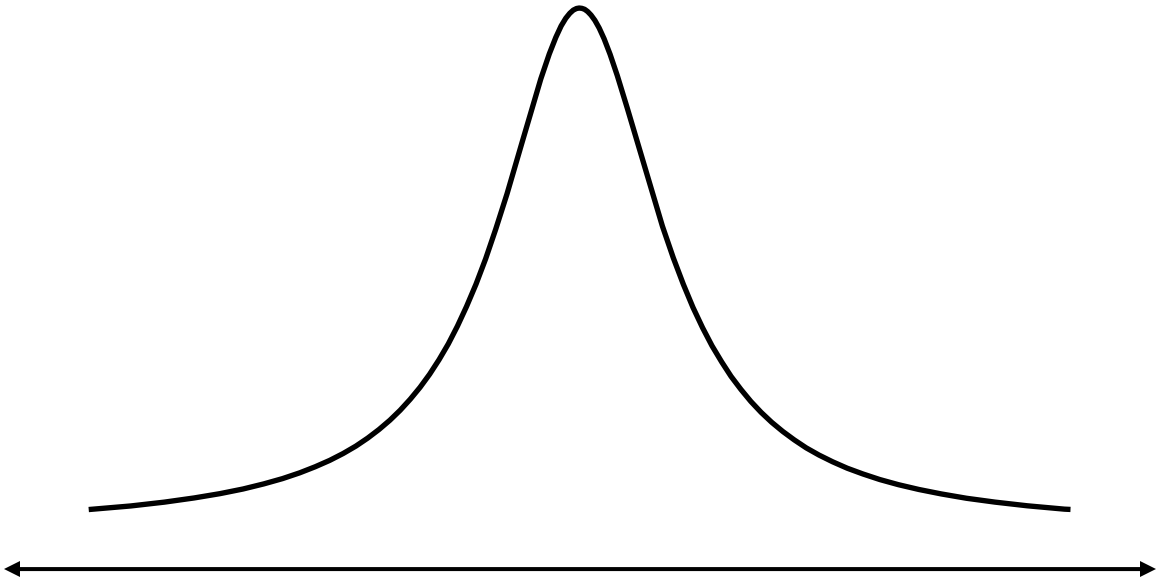
The lower bound : $\beta_2 - t_{\frac{\alpha}{2}} \cdot \sigma_{\widehat{\beta}_2} =$

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

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

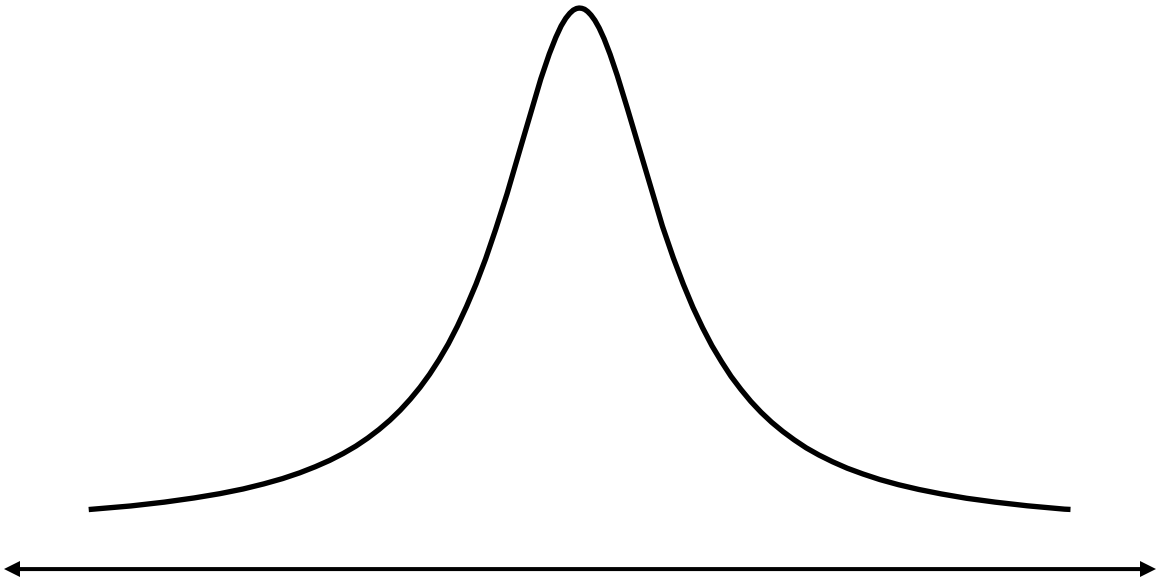
Note that when our null hypothesis is

(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'.

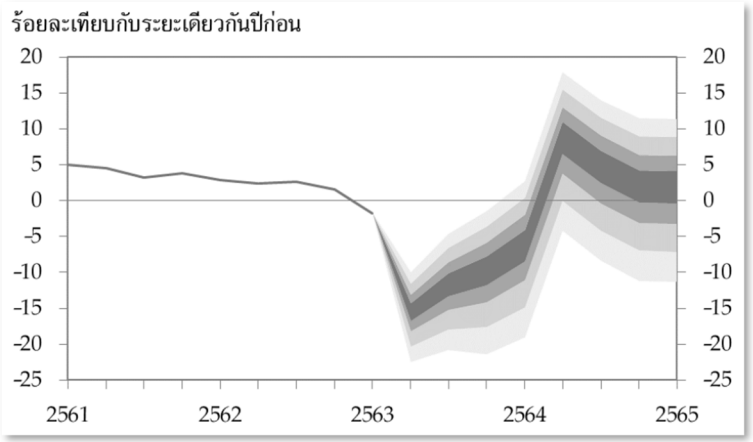
We can actually put our X_0 value of interest into the estimated

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

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

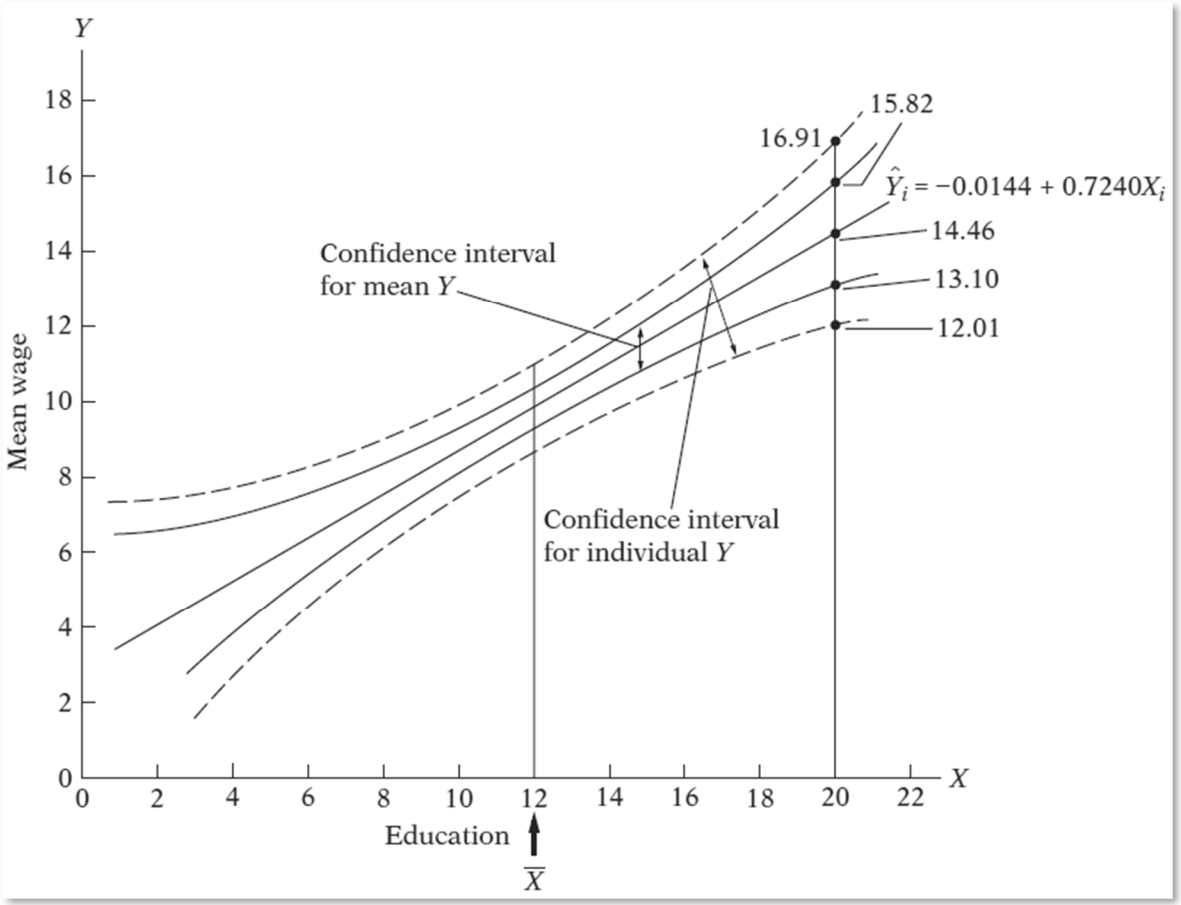
(1) Prediction

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

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

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

› Both GPDI and GDP in millions of dollars

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

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

(3) Data scaling

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

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

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

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

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

$$se = (116.358) \quad (12.9465) \qquad r^2 = 0.9648$$

(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

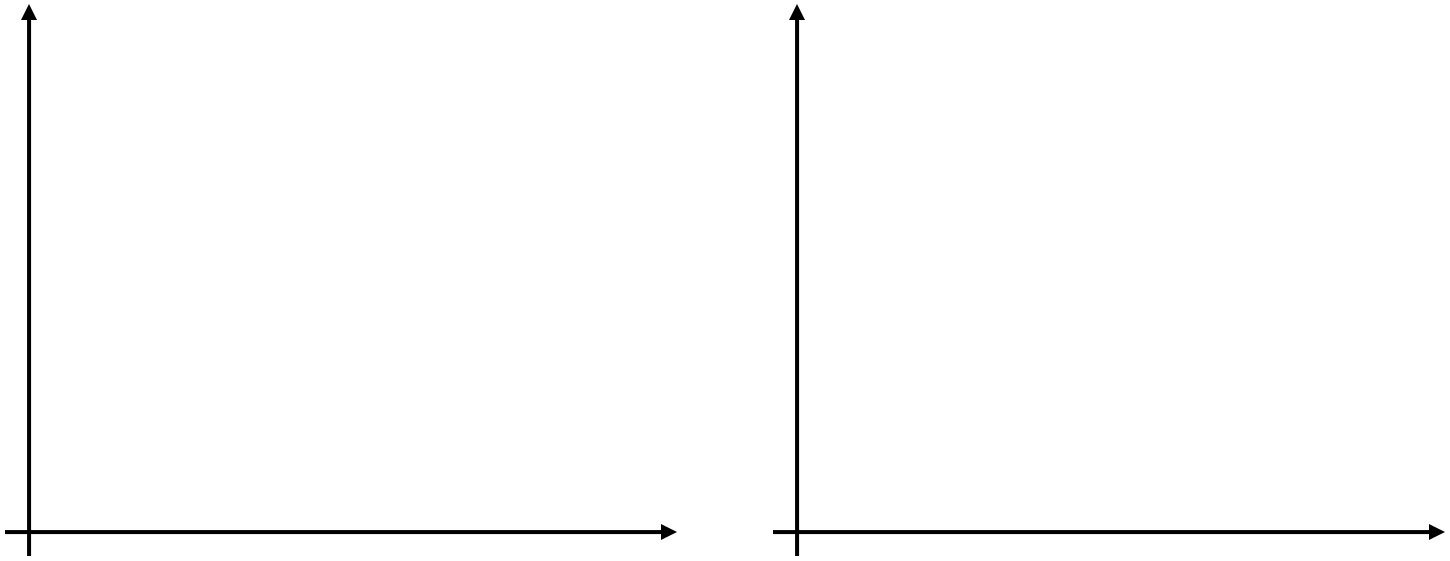
› Elasticity

(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.

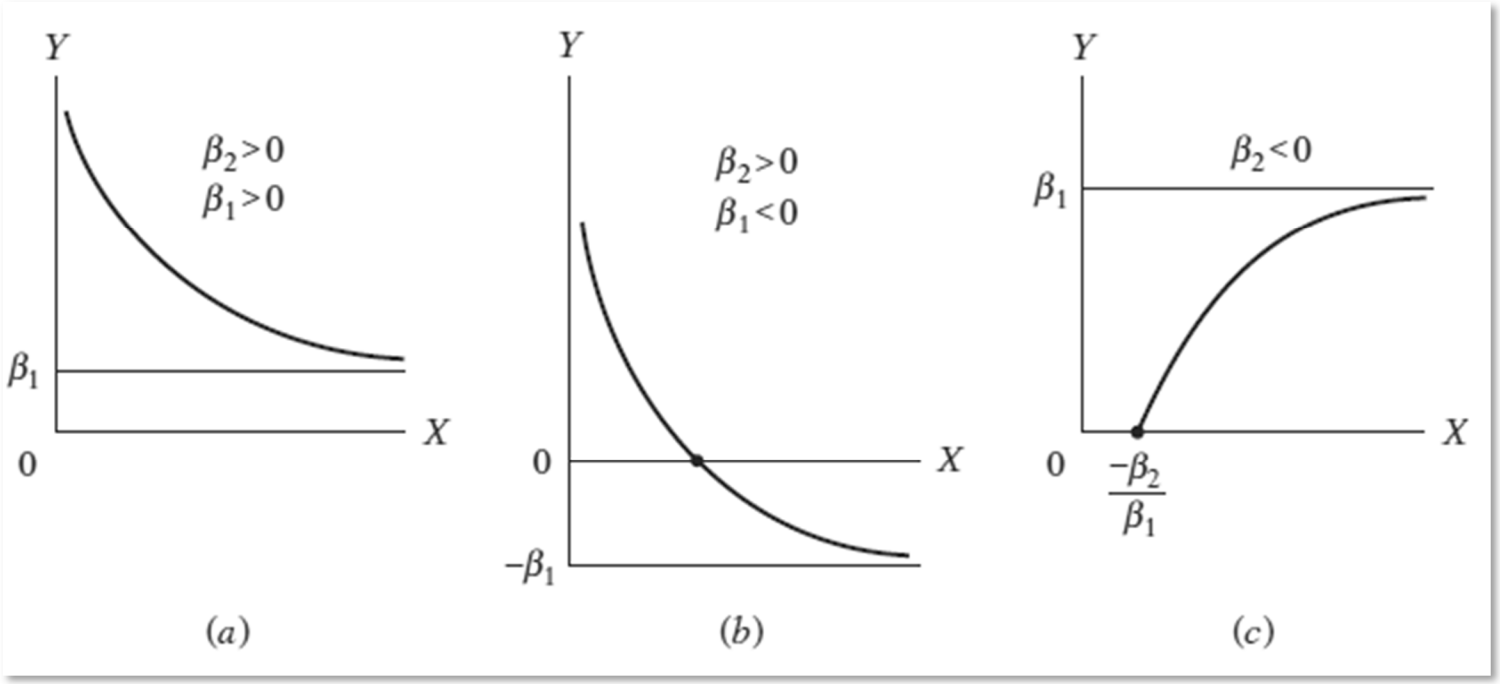
› Slope

› 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.