Ambo University Woliso Campus

School of Business and Economics

Department of Economics

Econometrics –II

Course code: Econ 2062

By: Tesfaye Etensa (Ass.Prof.)

Email:

tesfaye.etansa@ambou.edu.et/tesfayeetensa@gmail.com

Phone No:+251917054131

Ambo University Woliso Campus

School of Business and Economics

Economics Department

Econometrics –II (Econ 2062)

☞ ***(This ppt was prepared from econometrics books of "Tesfaye Etensa (2017): "Introduction to Econometrics: Theory and Practice with Stata"***

Prepared by:

Tesfaye Etensa

E-mail: tesfaye.etansa@ambou.edu.et

tesfayeetensa@gmail.com

May, 2020

Woliso, Ethiopia

CHAPTER ONE

Regression with Qualitative Information

Dummy Variables Regression

Outline of the Chapter

1.1 Introduction

1.2 Dummy as independent variables

- *1.2.1 Cautions in the use of dummy variables*
- *1.2.2 Regression with qualitative variables: Analysis of Covariance (ANOVA)*
- *1.2.3 Regression with a mixture of qualitative and quantitative variables: Analysis of Covariance (ANCOVA)*
- *1.2.4 Interactions between qualitative variables*
- *1.2.5 Interactions between qualitative & quantitative variable (Slope indicators variables)*

1.3 Structural stability

1.4 Dummy as dependent variable: Qualitative Response Model

- *1.4.1 Linear probability Model*
- *1.4.2 Logit model*
- *1.4.3 Probit model*
- *1.4.4Tobit model*

1.1. “Introduction”

1.1 Introduction

- **In regression analysis the dependent variable is not only frequently affected by quantitative (ratio scale variables like price, income, output, etc) but also qualitative variables (nominal scale variables like sex, race, nationality, etc)**
  - **Such variables should be included in the model**
- **Dummy variables are commonly used as proxies for qualitative factors such as sex, religion, etc**
- **Dummy variable is synonymously with non measurable, qualitative in nature, nominal scale, non numeric variable.**

1.1 Introduction

Qualitative information?

- ☞ **It is a non measurable information that we obtain or gather for a given variable.**
- ☞ **It is an indicator variable that is non measurable or non quantify in nature.**
- ☞ **Indicator variable, binary variable, categorical and dichotomous variable are use interchangeable.**

1.1 Introduction

✍ **Not all information can easily be quantified.**

☞ **So, need to incorporate qualitative information.**

Example: 1. Effect of belonging to a certain group:

- 👁 **Gender, location, marital status, occupation**
- 👁 **Beneficiary of a program/policy**

2. Ordinal variables:

- 👁 **Answers to yes/no (or scaled) questions...**

✍ **Effect of some quantitative variable may differ between groups/categories:**

- 👁 **Returns to education may differ between sexes or between ethnic groups …**

1.1 Introduction

Example 2: Suppose the firm utilized two types of production process to obtain its output.

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

Where Y is output obtained

D is the dummy variable

$$D_i = \begin{cases} 0 \text{ if the output is obtained from machine A} \\ 1 \text{ if the output is obtained from machine B} \end{cases}$$

Example 3: Does sex makes any difference in a college teachers salary, assuming that all other variables such as age, education level, and experience etc are held constant.

$$Y_i = \beta_0 + \beta_1 D + u_i$$

1.1 Introduction

✍ **Interest in determinants of belonging to a group**

👁 **Determinants of being poor …**

☞ **Dummy dependent variable (<u>*logit*</u>, <u>*probit*</u>…)**

☞ **<u>Dummy Variable</u>: a variable devised to use qualitative information in regression analysis.**

☞ **A dummy variable takes 2 values: usually 0/1.**

e.g. $Y_i=\beta_0+\beta_1*D+u$; for $\forall i \in$ group 1, and
for $\forall i \notin$ group 1.

➢ $If\ D = 0,\ E(Y) = E(Y|D = 0) = \beta_0$

➢ $If\ D = 1,\ E(Y) = E(Y|D = 1) = \beta_0 + \beta_1$

☞ **Thus, the difference between the two groups (in mean values of Y) is: $E(Y|D=1) - E(Y|D=0) = \beta_1$.**

☞ **The significance of this difference is tested by a t-test of $\beta_1 = 0$.**

Lab session

Use “*lecture_1.xls*” data to practice what we learnt in previous sections

1.2 “Dummy as Independent Variables”

1.2 Dummy as Independent Variables

I. How we include dummy variable as explanatory variable?

☞ **Constructing artificial variable, which take on values of 1 or 0,**

- **0 indicating the absence of the attributes**
- **1 indicating the presence of that attributes**

1.2 Dummy as Independent Variables

II. ANOVA & ANCOVA?

- ☞ **ANOVA: A regression model which contains regressors (explanatory variables) that are all exclusively dummy variables.**
- ☞ **ANCOVA: Regression model which contains quantitatively explanatory variables in addition to dummy variables.**
  - ✍ **A regression model which contains the mixed of both qualitative and quantitative variable.**

1.2 Dummy as Independent Variables

III. Purpose of dummy variable

- ☞ **It allows for difference in intercept**
- ☞ **It allow for difference in slopes**
- ☞ **It help us to estimate equations with cross equation restrictions**
- ☞ **Test for the stability of regression coefficients**

1.2.1 Dummy variable Trap

- ☞ ***It is the Cautions in the use of dummy variables, what we called it as "Dummy variable Trap"***
- ☞ **We should include (introduce) j-1 where j is number of variables**
- ☞ **The general rule is that" *If a qualitative variable has m categories, introduce only (m-1) dummy variables.*"**
- ☞ **Example: Sex has two categories and hence we introduced only a single dummy variable.**
- ☞ **If this is not fulfilled, we faced the problem of perfect multicollinearity (perfect collinearity), which is called "*dummy variable trap.*"**

1.2.1 Dummy variable Trap

- **The category for which no dummy variable is assigned is known as the** ***base, bench mark, control, comparison, reference, or omitted category*****.**
- **Hence, all comparisons are made in relation to the bench mark category (we assigned 0 values)**
- **If there is a constant term in the regression equation the number of dummies defined should always be one less than the number of groupings by that category.**
- **b/c the constant term is the intercept for the base group and the coefficients of the dummy variables measures differences in intercept (the mean difference)**

1.2.1 Dummy variable Trap

- **If the coefficients β_i attached to the dummy variables D, are called differential intercept coefficients.**
  - **Reasons: It tell by how much the value of the intercept of the category that receives that value of 1 differs from the intercept coefficients of the base category.**
- **The intercept value (α) represents the *mean value* of the benchmark category.**
- **If we don’t have constant term, we can’t used j-1 or m-1 because we don’t have dummy variable trap.**

1.2.2 ANOVA Analysis

- *It is regression with qualitative variables.*

A. A single dummy independent variable

☞ **The model specification will be:**

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

$$Y_i = \beta_0 + \beta_1 D + \varepsilon_i$$

$$D_i = \begin{cases} 1 \text{ if male} \\ 0 \text{ if female (otherwise)} \end{cases}$$

β It doesn’t measure the slope rather it measures the mean difference b/n the category

Y is annual salary of a college teacher

α It measure the mean value of the bench category

1.2.2 ANOVA Analysis

- ☞**Example: Suppose a researcher wants to find out whether sex makes any difference in a college teachers' salary, assuming other variables, education level, experiences, and age being constant.**
  - ☞**Assuming the disturbance term satisfy the usual assumptions, of the classical linear regression model, we obtain:**

1.2.2 ANOVA Analysis

Interpretation

- ☞ **The mean (average) annual salary of female college teacher: E(Yi/Di=0)=** β_0
- ☞ **The mean (average) annual salary of male college teacher: E(Yi/Di=1)=** $\beta_0 + \beta_1$
- ☞ **The slope coefficients** β_1 **“tells by how much the mean salary of a male college teacher differs from the mean salary of his female counter part or**
- ☞ **the value of** β_1 **is on average measures the difference between intercept.”**
- ☞ $\beta_0 + \beta_1$ **reflecting the mean salary of the male college professors or the average (mean) salary of non base group (actual value).**

1.2.2 ANOVA Analysis

- **The implication of slope coefficients (β_1) :**
- **The coefficient determines whether there is discriminating on against female.**
- **If then, for the same level of other factors, women earn less than men on average.**
- **A test of the null hypothesis that there is no sex discrimination (H0: $\beta_1 = 0$) can be easily made by running regression by OLS.**
- **Suppose we have the following regression result:**

1.2.2 ANOVA Analysis

Example

Observation	Salary (Y)	Sex (D)
1	27000	1
2	17500	0
3	42500	1
4	29000	1
5	23000	0
6	32000	1
7	18500	1
8	22000	0
9	24000	1
10	18500	0

1.2.2 ANOVA Analysis

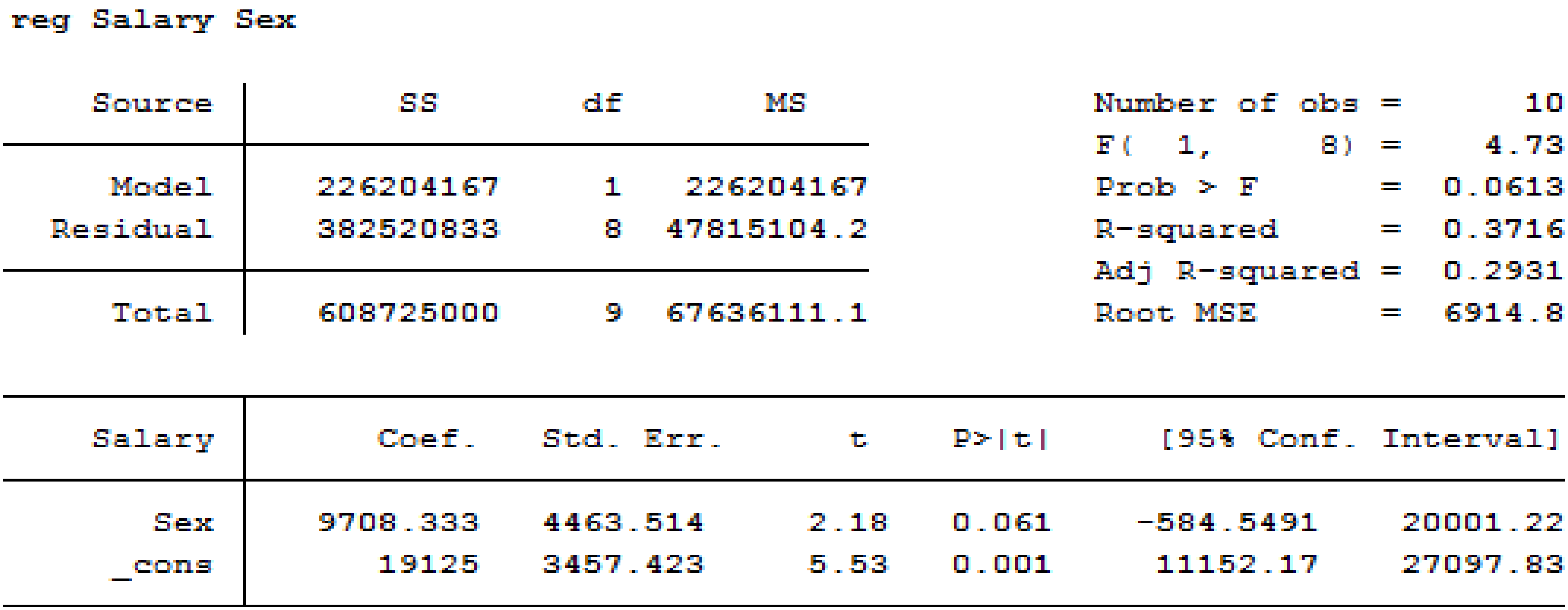
Brainstorm:

Based on the above table:

1. **Estimate the coefficients of the variables**
2. **Square of correlation coefficient**
3. **Standard error and t-statistics**
4. **TSS, ESS and RSS**
5. **What makes this regression different from simple regression you have learnt under Econometrics-I?**
6. **Interpret the result**

1.2.2 ANOVA Analysis

Regression Result of above example



```
reg Salary Sex

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     8) =    4.73
       Model |   226204167     1   226204167           Prob > F      =  0.0613
    Residual |   382520833     8  47815104.2           R-squared     =  0.3716
-------------+------------------------------           Adj R-squared =  0.2931
       Total |   608725000     9  67636111.1           Root MSE      =  6914.8

------------------------------------------------------------------------------
      Salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Sex |   9708.333   4463.514     2.18   0.061    -584.5491    20001.22
       _cons |      19125   3457.423     5.53   0.001     11152.17    27097.83
------------------------------------------------------------------------------
```

$$
\begin{array}{lll}
\hat{Y}_i = 19125 & + 9708.33 D_i & \\
se \quad (4463.514) & (3457.423) & \\
\text{t} \qquad (5.53) & (2.18) &
\end{array}
$$

1.2.2 ANOVA Analysis

☞ ***Interpretation :***

☞ The estimated mean (average) salary of female college teacher is birr 19,125(= $\hat{\beta}_0$).

☞ The mean salary of male college teachers is birr ($(\hat{\beta}_0 + \hat{\beta}_1)$ =28,833.

β_1 =9708 is the mean difference between male and female college teachers.

☞ Since β_1 is statistically significant, the results indicate that the mean salary of two categories are different, actually, the female teacher’s average salary is lower than that of her counterpart.

1.2.2 ANOVA Analysis

- **Interpretation:**
- **If all other variables are held constant, there is sex discrimination in the salaries of the two sexes or the salary of the female is less than male by 9708, on average.**

B. A multiple dummy independent variable

- **It is when more than two distinct values are involved.**
- **Always when there are N variables we develop N-1 dummy variables.**
- **Let, a given output is produced using three methods of production says: Machine A, Machine B , and Machine C.**

$$Y_i = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + u_i$$

1.2.2 ANOVA Analysis

$$D_1 = \begin{cases} 1 \text{ if output is obtained from machine A} \\ 0 \text{ otherwise (if it comes from B \& C)} \end{cases}$$

$$D_2 = \begin{cases} 1 \text{ if output is obtained from machine B} \\ 0 \text{ otherwise (if it comes from A \& C)} \end{cases}$$

Interpretation

☞ β_0 **represents the mean value of output obtained from machine C.**

☞ β_1 **is the mean difference in output associated with a change from machine C to machine A.**

1.2.2 ANOVA Analysis

- ☞ $(\beta_0 + \beta_1)$ **it is the mean value of output obtained from machine A.**
- ☞ β_2 **is the mean difference in output associated with a change from machine C to machine B.**
- ☞ $(\beta_0 + \beta_2)$ **is the mean value of output obtained from machine B.**
- ☞ **Exercise: Interpret the following model**

$$Y_i = \beta_0 + \beta_1 Gender + \beta_2 Instituion + u_i$$

$$Gender = \begin{cases} 1 \text{ male} \\ 0 \text{ female} \end{cases}; \ Institutions = \begin{cases} 1 \text{ government} \\ 0 \text{ private} \end{cases}$$

1.2.2 ANOVA Analysis

Example2.: Wage differential between male and female

☞ **Two possible ways: a male or a female dummy.**

1. Define a *male dummy* (male = 1 & female = 0).

✍ ***reg wage male***

✍ **Result:** $Y_i = 9.45 + 172.84*D + \hat{u}_i$

***p*-value: (0.000) (0.000)**

☞ ***Interpretation the monthly wage of a male worker is, on average, $172.84 higher than that of a female worker.***

☞ **This difference is significant at 1% level.**

2. Define a *female dummy* (female = 1 & male = 0)

✍ **reg wage female**

✍ **Result:** $Y_i = 182.29 – 172.84*D + \hat{u}_i$

p*-value: (0.000) (0.000)** ***Interpretation ??

1.2.3 Analysis of Covariance (ANCOVA)

☞ **Unlike ANOVA, a regression model may contain regressors that are all exclusively dummy, or qualitative, in nature; ANCOVA, is regression with a mixture of qualitative and quantitative independent variables.**

☞ **It is regression on both qualitative and quantitative independent variables.**

A. Single dummy independent variable

☞ **Example: Suppose we identified two variables that affect the salary of a given employee.**

$$Wage = \beta_0 + \beta_1 gend + \beta_2 educ + u_i$$

$$gend = \begin{cases} 1 & \text{female} \\ 0 & \text{male} \end{cases}$$

wage = wage rate of individual

educ= level of education

Gender & level of education is the only observed factors affect wage.

1.2.3 Analysis of Covariance (ANCOVA)

Interpretation

- ☞ β_2 measures the slope.
- ☞ β_1 is the difference in hourly wage between females and males, given the amount of education. Hence, the coefficient determines whether there is discrimination against women.
- ☞ If $\beta_1 < 0$, then for the same level of education, women earns less than men, on average.
- ☞ If we assume the zero conditional mean assumptions E(U)=0, then:

$$\beta_1 = E(wage \mid gend = 1, educ) - E(wage \mid gend = 0, educ)$$

- ☞ **Key: the level of education is the same in both individuals; the difference, β_1 is due to gender only.**

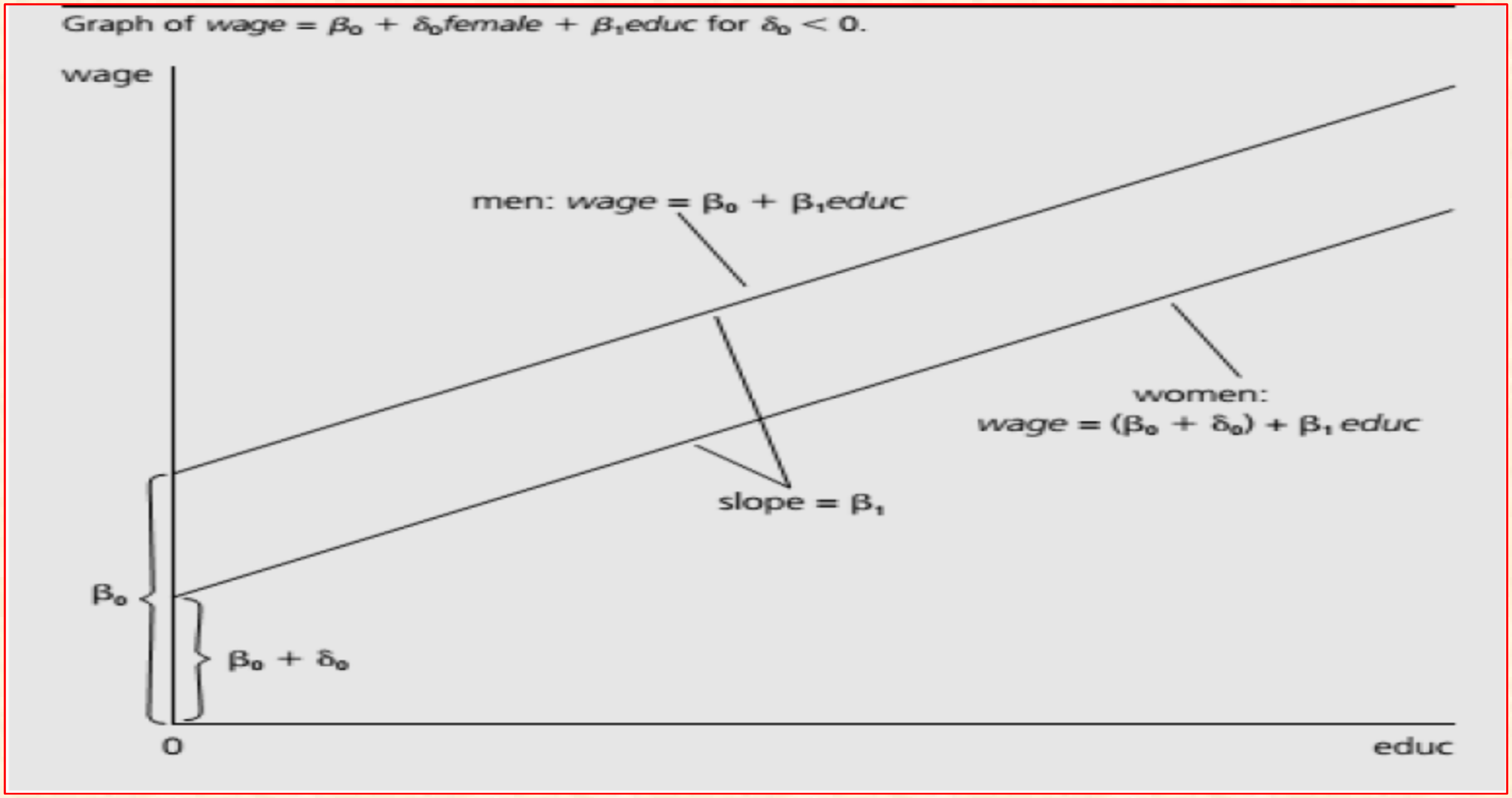
1.2.3 Analysis of Covariance (ANCOVA)

Interpretation

- The intercept for male is β_0
- The intercept for female is $\beta_0 + \beta_1$
- Since there are just two groups, we only need two different intercepts. This means that, in addition to β_0, we need to use only one dummy variable; we have chosen to include the dummy variable for females.
- Using two dummy variables would introduce perfect collinearity because $female + male = 1$, which means that male is a perfect linear function of female.
- Including dummy variables for both genders is the simplest example of the so-called dummy variable trap,

1.2.3 Analysis of Covariance (ANCOVA)

☞ ***Interpretation:***



Graph of $wage = \beta_0 + \delta_0 female + \beta_1 educ$ for $\delta_0 < 0$.

1.2.3 Analysis of Covariance (ANCOVA)

Interpretation

- **Mean salary of female college professor:**

$$E(Y / D = 1, educ) = (\beta_0 + \beta_1) + \beta_2 educ$$

- **Mean salary of male college professor:**

$$E(Y / D = 0, educ) = \beta_0 + \beta_2 educ$$

- **After we run OLS regression, if the *t test shows that it is statistically significant*, we reject the null hypothesis that the male and female college professors' levels of mean annual salary are the same, and we accept the alternative hypothesis.**

1.2.3 Analysis of Covariance (ANCOVA)

Features of the dummy variable regression model

☞ *Introducing dummy variable*

- **The general rule:** ***If a qualitative variable has 'm' categories, introduce only 'm-1' dummy variables*. In our example, sex has two categories, and hence we introduced only a single dummy variable. If this rule is not followed, we shall fall into what might be called the dummy variable trap, that is, the situation of perfect multicollinearity.**

☞ *The assignment of 1 and 0 values to two categories, such as male and female, is arbitrary.*

1.2.3 Analysis of Covariance (ANCOVA)

Features of the dummy variable regression model

- ☞ The group, category, or classification that is assigned the value of 0 is often referred to as the base, benchmark, control, comparison, reference, or omitted category.
  - ☞ It is the base in the sense that comparisons are made with that category.
- ☞ *The coefficient attached to the dummy variable D can be called the differential intercept coefficient.*
  - b/c it tells by how much the value of the intercept term of the category that receives the value of 1 differs from the intercept coefficient of the base category.

1.2.3 Analysis of Covariance (ANCOVA)

B. Dummy Variables for Multiple Categories

- ☞**Example: Suppose that, on the basis of the cross-sectional data, we want to regress the annual expenditure on health care by an individual on the income and education of the individual.**
- ☞**Since the variable education is qualitative in nature, suppose we consider three mutually exclusive levels of education: less than high school, high school, and college.**
- ☞**Now, unlike the previous case, we have more than two categories of the qualitative variable education.**

1.2.3 Analysis of Covariance (ANCOVA)

- ☞ Thus, following the rule that the number of dummies be one less than the number of categories of the variable, we should introduce two dummies to take care of the three levels of education.
- ☞ Assuming that the three educational groups have a common slope but different intercepts in the regression of annual expenditure on health care on annual income, we can use the following model: $Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 X_i + u_i$
- ☞ Where Yi= annual expenditure on heath care

  Xi= annual income

1.2.3 Analysis of Covariance (ANCOVA)

- $D_1 = \begin{cases} 1 \text{ if high school education} \\ 0 \text{ otherwise} \end{cases}$ **;** $D_2 = \begin{cases} 1 \text{ if college education} \\ 0 \text{ otherwise} \end{cases}$
- **We arbitrarily treating the “*less than high school education*” category as the base category.**
- **Therefore, the intercept β_0 will reflect the intercept for this category.**
- **The differential intercepts β_1 and β_2 tell by how much the intercepts of the other two categories differ from the intercept of the base category, which can be readily checked as follows:**
- **Assuming , we obtain from the above specification**

1.2.3 Analysis of Covariance (ANCOVA)

$$E(Y_i \mid D_1 = 0,\ D_2 = 0, X_i) = \beta_0 + \beta_3 X_i$$

- **the mean health care expenditure functions for less than high school**

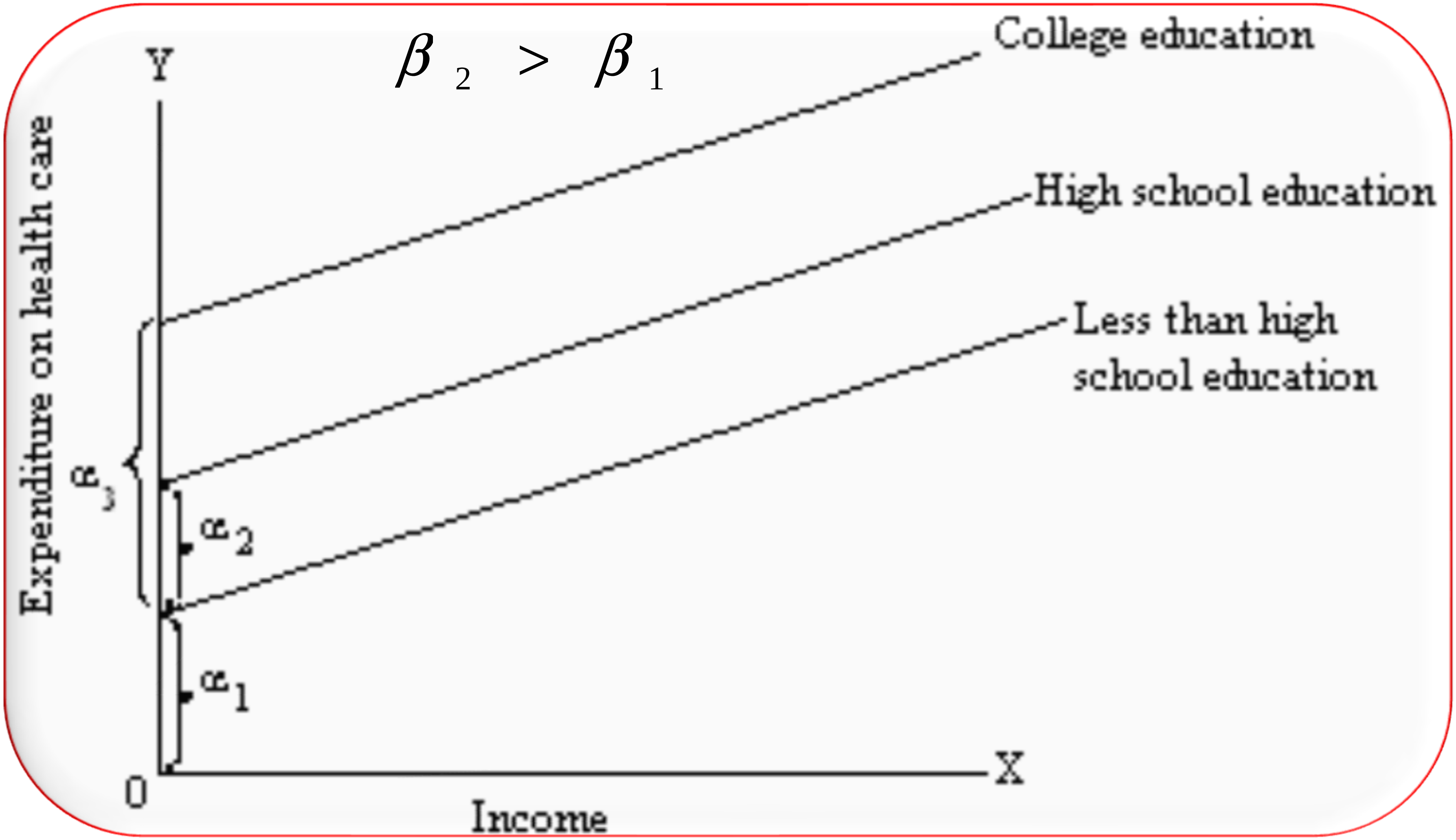
$$E(Y_i \mid D_1 = 1,\ D_2 = 0, X_i) = (\beta_0 + \beta_1) + \beta_3 X_i$$

- **the mean health care expenditure functions for the high school**

$$E(Y_i \mid D_1 = 0,\ D_2 = 1, X_i) = (\beta_0 + \beta_2) + \beta_3 X_i$$

- **the mean health care expenditure functions for the college.**

1.2.3 Analysis of Covariance (ANCOVA)



$\beta_2 > \beta_1$

1.2.3 Analysis of Covariance (ANCOVA)

☞ **If a qualitative variable has more than one category, the choice of the bench mark category is strictly up to the researcher.**

☞ **There is a way suppressed this trap by introducing as *many dummy variable as the number of categorical* of that variable provide we do not introduce the intercept (constant term) in such a model.**

$$Y_i = \beta_1 D_1 + \beta_2 D_2 + \beta_3 X + u_i$$

☞ **When we run regression, we use the non intercept option in your regression packages (suppressed intercept)**

1.2.3 Analysis of Covariance (ANCOVA)

i. Dummy Variables for Multiple Categories :No Intercept Case

- ☞ **If there is no intercept, we have no comparison, base group and we did not omitted one category.**

$$\hat{Y}_i = 13{,}124\,D_1 + 12{,}244\,D_2 + 10{,}453\,D_3$$

se	(546)	(425)	(462)
t	(13)	(14)	(12)

$$R^2 = 0.2546$$

- ☞ **Where Y is salary of teachers**

$$D_1 = \begin{cases} 1 & \text{west} \\ 0 & \text{otherwise} \end{cases} \quad ; D_2 = \begin{cases} 1 & \text{north} \\ 0 & \text{otherwise} \end{cases}$$

$$D_3 = \begin{cases} 1 & \text{south} \\ 0 & \text{otherwise} \end{cases}$$

1.2.3 Analysis of Covariance (ANCOVA)

Interpretation

- **B_1= the mean salary of teachers in west=13124**
- **B_2= the mean salary of teachers in north=12244**
- **B_3= the mean salary of teachers in south=10453**

ii. Dummy Variables for Multiple Categories :Case when constant term is present

- ***Redo the above example, now assume we take west as base category.***

$$\begin{array}{llll} \hat{Y}_i = & 13,456.63 & - 845.23 D_2 & - 1245.14 D_3 \\ se & (234.56) & (125.98) & (262.45) \\ t & (21.26\quad) & (14.84) & (13.68) \end{array}$$

1.2.3 Analysis of Covariance (ANCOVA)

Interpretation

- ☞ **The mean salary of teachers in the west is about 13,456.63.**
- ☞ **The mean salary of teachers in the north is lower by 845.23 and that is the teachers in the south is lower than 1245.15; the mean salary of teachers in the north 13,456.63- 845.23=12611.4**
- ☞ **-845.23 tell us that the mean salary of teachers in the North is smaller by about 845.23 than the mean salary of about 13456 for the bench mark category, west.**

- ☞ **N.B: Model with intercept is more appropriate than no constant term b/c it facilitates comparisons.**

1.2.3 Analysis of Covariance (ANCOVA)

- **Intercept indicators variables:**
- **The above examples we have seen under ANCOVA analysis an example of intercept indicator variables, a regression mixture of qualitative and quantitative variables.**
- **It affects only intercept.**
- **It interact with dummy variables and qualitative variables. This is why it affects only intercepts rather than slope.**

$$Wage = \beta_0 + \beta_1 gend + \beta_1 educ + u_i$$

$$E(Y) = \begin{cases} \beta_0 + \beta_1 + \beta_2 educ \; --- \; when \quad \mathrm{D} = 1 \\ \beta_0 + \beta_2 educ \; ------ \; when \quad \mathrm{D} = 0 \end{cases}$$

1.2.3 Analysis of Covariance (ANCOVA)

- ☞ **Then the difference b/n them is:**

$$\beta_1 = (\beta_0 + \beta_1 + \beta_2 educ) - (\beta_0 + \beta_2 educ)$$

- ☞ **+ve: it is greater than other**
- ☞ **-ve: it is less than other**

1.2.4 Interactions among Dummy Variables

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i$$

☞ Y_i where annual expenditure on clothing

☞ X_i income

$$D_2 = \begin{cases} 1 \text{ if female} \\ 0 \text{ if male} \end{cases}$$

$$D_3 = \begin{cases} 1 \text{ if college graduate} \\ 0 \text{ otherwise} \end{cases}$$

- In many applications there may be interaction between the two qualitative variables D_2 and D_3 therefore their effect on mean Y may not be simply additive but multiplicative as well
- Hence, we re specify the above model as follows:

1.2.4 Interactions among Dummy Variables

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i$$

$$E(Y_i \mid D_2 = 1, D_3 = 1, X_i) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i$$

α_2 = differential effect of being a female

α_3 = differential effect of being a college graduate

α_4 = differential effect of being a female graduate

* The above equation shows that the mean clothing expenditure of graduate females is different by α_4 from the mean clothing expenditure of females or college graduates.

* If α_2, α_3 and α_4 are all +ve, the average clothing expenditure of females is higher (than the base category, which here is male non graduate), but it is much more so if the females also happen to be graduates.

1.2.4 Interactions among Dummy Variables

- **Similarly, the average expenditure on clothing by a college graduate tends to be higher than the base category but much more so if the graduate happens to be *a female.***
  - ***This shows how the interaction dummy modifies the effect of the two* attributes considered individually.**
- **Whether the coefficient of the interaction dummy is statistically significant can be tested by the usual t test.**
- **If it turns out to be significant, the simultaneous presence of the two attributes will attenuate or reinforce the individual effects of these attributes.**
  - ***Omitting a significant interaction term incorrectly will lead to a specification bias.***

1.2.4 Interactions among Dummy Variables

The importance of interactions among dummy variables

☞ **help us to get influential variables**

☞ **to avoid misspecification bias**

1.2.5 Slope indicator variables

- **The interaction between dummy variables and quantitative variables. They affect only slope, i.e, it does not affect intercept.**
- **It help us to captures the interaction effect of dummy and quantitative variables on dependent variables**
- **Look at the following example**
  - **The price of condominium house can be explained as a function of its characteristics such as its size, location, number of bedrooms, age, floor and so on.**

1.2.5 Slope indicator variables

- **For our discussion, let us assume that the number bed room of the house of the measured in numbers, *nbdr*, is the only relevant variable in determining house price.**

$$prhou = \beta_0 + \beta_1 nbdr + u_i$$

- β_1 **is the value of an additional number of bed rooms.**
- β_0 **is the value of land alone**
  - **We can use dummy variable and indicator variable interchangeable.**

1.2.5 Slope indicator variables

$$prhou = \beta_0 + \psi neib + \beta_1 nbdr + u_i$$

$$neib = \begin{cases} 1 \text{ if desirable neibourhood} \\ 0 \text{ if not desirable neibourhood} \end{cases}$$

- ☞ **We make the reference group, non desirable group.**
- ☞ **Instead of assuming that the effect of location on house price causes a change in the intercept.**
- ☞ **Let us assume that the change is in the slope of the relationship.**

1.2.5 Slope indicator variables

- We can allow for a change in a slope by including in the model an additional explanatory variable that is equal to the product of an indicator variable and continuous variable.
- In our model, the slope of the relationship is the value of an additional number of bed rooms.
- If we assume 1 value for homes in desirable neibourhood, and 0 other wise; we can specify our model as follows:

$$prhou = \beta_0 + \beta_1 nbdr + \omega(nbdr * neib) + u_i$$

1.2.5 Slope indicator variables

- **The new variable (nbdr*neib) is the product number of bedroom and the indicator variables, is called an interaction variable as it captures the interaction of location and number of bedroom on condominium house prices.**
- **Or it is called a slope –indicator variable or a slope dummy variable, b/c it allows for the change in the slope of the relationship.**
- **The slope indicator variable takes a value equal to nbdr for houses in the desirable neibourhood, when neib=1, and it is 0 for homes in other neighbourhoods.**

1.2.5 Slope indicator variables

- ☞ A slope indicator variable is treated as just like any other explanatory variable in a regression model.

$$E(prhou) = \begin{cases} \beta_0 + \beta_1 nbdr + \omega nbdr --- when \quad \mathrm{D} = 1 \\ \beta_0 + \beta_1 nbdr -------- when \quad \mathrm{D} = 0 \end{cases}$$

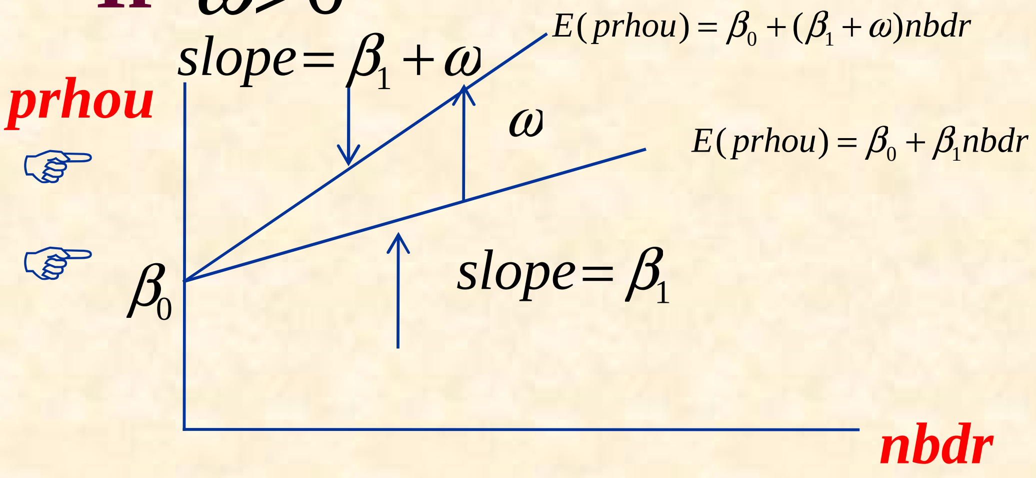
- ☞ In the desirable neighbourhood, the price per additional number of bedrooms of a house is $\beta_1 + \omega$
- ☞ In the non desirable neighbourhood, the price per additional number of bedrooms of a house is β_1.
- ☞ If $\omega > 0$ price per additional number of bedrooms is higher in the more desirable neighbourhood.

1.2.5 Slope indicator variables

- ☞**The effect of including a slope indicator variable also can be see by using calculus.**
- ☞**The partial derivatives of expected house price with respect to number of bed rooms**

$$\frac{\partial E(prhou)}{\partial(nbdr)} = \begin{cases} \beta_1 + \omega & \text{when } D = 1 \\ \beta_1 & \text{w hen } D = 0 \end{cases}$$

- ☞**If** $\omega > 0$



$E(prhou) = \beta_0 + (\beta_1 + \omega)nbdr$
$slope = \beta_1 + \omega$
prhou
ω
$E(prhou) = \beta_0 + \beta_1 nbdr$
β_0
$slope = \beta_1$
nbdr

1.2.5 Slope indicator variables

- ☞ If we assume that house location affects both the intercept and the slope, then both affects can be incorporated into a single model.
- ☞ The model specification will be:

$$prhou = \beta_0 + \psi neib + \beta_1 nbdr + \omega(nbdr * neib) + u_i$$

$$E(prhou) = \begin{cases} (\beta_0 + \psi) + (\beta_1 + \omega) nbdr \quad - - - \; when \quad \mathrm{D} = 1 \\ \beta_0 + \beta_1 nbdr \quad - - - - - - - - - \; when \qquad \mathrm{D} = 0 \end{cases}$$

- ☞ Look numerical example from Principle of Econometrics

1.3 “Structural Stability”

1.3 Structural Stability

- ☞ **Testing for structural stability is help us to *find out whether two or more regressions are different*, where the difference may be in the intercepts or the slopes or both.**
- ☞ **Suppose we are interested in estimating a simple saving function that relates domestic household savings (S) with gross domestic product (Y) for Ethiopia.**
- ☞ **Suppose further that, at a certain point of time, a series of economic reforms have been introduced.**

1.3 Structural Stability

- ☞So far we assumed that the intercept and all the slope coefficients (β_j's) are the same/stable for the **whole** set of observations. $Y = X\beta + e$
- ☞But, **structural shifts** and/or group differences are common in the **real world**. May be:
  - ☞the intercept differs/changes, or
  - ☞the (partial) slope differs/changes, or
  - ☞both differ/change across categories or time period.

1.3 Structural Stability

- **The hypothesis here is that such reforms might have considerably influenced the savings- income relationship, that is, the relationship between savings and income might be different in the post reform period as compared to that in the pre-reform period.**
- **If this hypothesis is true, then we say a structural change has happened.**
  - ***H0: Economic reforms might not have influenced the savings and national income relationship***
  - ***H1: Economic reforms might have influenced the savings and national income relationship***

- **How do we check if this is so?**

1.3 Structural Stability

☞ **We can test structural stability of testing parameter by using two methods.**

1. **Using Dummy variables**
2. **Chow's test**

1. Using dummy variables

*** Write the savings function as:**

$$S_t = \beta_0 + \beta_1 D_t + \beta_2 Y_t + \beta_3 (Y_t D_t) + u_t$$

where S_t *is household saving at time* t, Y_t *is GDP at time* t *and*

$$D_t = \begin{cases} 0 \; if \;\; pre-reform\,(<1991) \\ 1 \; if \;\; post-reform\,(>1991) \end{cases}$$

1.3 Structural Stability

- ☞ β_3 **Is the differential slope coefficient indicating how much the slope coefficient of the pre-reform period savings function differs from the slope coefficient of the savings function in the post reform period.**
- ☞ **Decision rule:**
- ☞ **If** β_1 *and* β_3 **are both statistically significant as judged by the t-test, the pre-reform and post-reform regressions differ in both the *intercept* and the *slope*.**

1.4 Structural Stability

- **If only β_1 is statistically significant, then the pre-reform and post-reform regressions *differ only in the intercept* (meaning the marginal propensity to save (MPS) is the same for pre-reform and post-reform periods).**
- **If only β_3 is statistically significant, then the two regressions differ only in the slope (MPS).**
- **Check structural stability for the f/wing regression result:**

$$\hat{S}_t = -20.76005 + 5.9991\,\hat{D}_t + 2.616285\,\hat{Y}_t - 0.5298177\,(\hat{Y}_t\hat{D}_t)$$

$$s.e \quad (6.04) \qquad (6.4) \qquad (.57) \qquad (.6035149)$$

1.4 Structural Stability

Example 2: Using the DVR to Test for Structural Break:

- **Recall the example of consumption function:**

  period 1: $cons_i = \alpha_1 + \beta_1 * inc_i + u_i$ vs.

  period 2: $cons_i = \alpha_2 + \beta_2 * inc_i + u_i$

- **Let's define a dummy variable D_1, where:**

  for the period 1974-1991, and

  for the period 1992-2006

- **Then, $cons_i = \alpha_0 + \alpha_1 * D_1 + \beta_0 * inc_i + \beta_1 (D_1 * inc_i) + u_i$**

For period 1: **$cons_i = (\alpha_0 + \alpha_1) + (\beta_0 + \beta_1) inc_i + u_i$**

For period 2 (base category): **$cons_i = \alpha_0 + \beta_0 * inc_i + u_i$**

- **Regressing cons on inc, D_1 and (D_1*inc) gives:**

  $cons = 1.95 + 152D_1 + 0.806*inc - 0.056(D_1*inc)$

  p-value: (0.968) (0.010) (0.000) (0.002)

1.4 Structural Stability

- **D_1=1 for i ϵ period-1 & D_1=0 for i ϵ period-2:**

  period 1 (1974-1991): *cons* = 153.95 + 0.75*inc

  period 2 (1992-2005): *cons* = 1.95 + 0.806*inc

- **The Chow test is equivalent to testing $\alpha_1=\beta_1=0$ in:**

  ***cons*=1.95+152D_1+0.806*inc – 0.056(D_1*inc)**

- **This gives: *F(2, 29) = 6.76; p-value = 0.0039.***
- **Then, reject H_0! There is a structural break!**

1.4 Structural Stability

- For a total of *m* categories, use *m–1* dummies!
- Including *m* dummies (1 for each group) results in perfect multicollinearity (dummy variable trap). e.g.: 2 groups & 2 dummies:
- constant = $D_1 + D_2$!!!

$$X = [constant \quad D_1 \quad D_2]$$

$$X = \begin{bmatrix} 1 & X_{11} & 1 & 0 \\ 1 & X_{12} & 1 & 0 \\ 1 & X_{13} & 0 & 1 \end{bmatrix}$$

1.4 Structural Stability

2. *Chow's test*

- **One approach for testing the presence of structural change (structural instability) is by means of Chow's test. The steps involved in this procedure:**
- ***Step 1:*** **Estimate the regression equation for the whole period (pre-reform plus post-reform periods) and find the error sum of squares (ESS$_R$) or RRSS.**
- ***Step 2:*** **Estimate equation (model) using the available data in the pre-reform period (say, of size n 1), and find the error sum of squares (ESS1) or RSS1**
- ***Step 3:*** **Estimate equation (model) using the available data in the pre-reform period (say, of size n 2), and find the error sum of squares (ESS2) or RSS2.**
- ***Step 4:*** **Calculate RSS$_{UR=}$ RSS1+RSS2.**
- ***Step 5:*** **Calculate the Chow test statistic**

$$F_c = \frac{(RSS_R - RSS_U)/k}{RSS_U\ /(n_1 + n_2 - 2k)}$$

- **Where k is number of estimated regression coefficients**

1.4 Structural Stability

- $F^{\alpha}_{(k,n_1+n_2-2k)}$ ***is the critical value from the t-distribution with k (in our case k=2) and n1+n2-2k degrees of freedom from a given significance level,*** α
- ***Decision rule***: **Reject the null hypothesis of identical intercepts and slopes for the pre-reform and post reform periods, that is**

$$H_0 = \begin{cases} \beta_0 = \beta_3 \\ \beta_2 = \beta_4 \end{cases} \text{if } Fc > Fb.$$

- **i.e, Rejecting H0 means there is a structural change.**

1.4 Structural Stability

- **Example: RSS1=64499436.865 (Error sum of squares in the pre-reform period); n1=12; RSS2=2,726,652,790.434 (Error sum of squares in the post-reform period); n2=11;**
- **RSSR=13,937,337,067.461 (Error sum of squares for the whole period)**
- **RSSU=RSS1+RSS2=2,791,152,227.299**
- **The test statistics is:**

$$F_c=\frac{(RSS_R-RSS_U)/k}{RSS_U\;/(n_1+n_2-2k)}=\frac{(13{,}937{,}337{,}067.461-2{,}791{,}152{,}227.2)/2}{(2{,}791{,}152{,}227.299)/(12+11-2(2))}\approx 190$$

- **The tabulated value from the F-distribution with 2 and 19 degrees of freedom at the 5% level of significance is 3.52.**

1.4 Structural Stability

- **Decision: Since the calculated value of F exceeds the tabulated value, we reject the null hypothesis of identical intercepts and slopes for the pre-reform and post reform periods at the 5% level of significance.**
- **Hence, we can conclude that there is a structural break.**

1.4 Structural Stability

Draw backs:

- ☞ **Chow’s test does not tell us whether the difference (change) in the slope only, in the intercept only or in both the intercept and the slope.**

The Chow Tests

- ☞ **Using an F-test to determine whether a single regression is more efficient than two/more separate regressions on sub-samples.**

1.4 Structural Stability

☞ The stages in running the Chow test are:

1. Run 2 separate regressions (say, before & after war or policy reform, …) & save RSS's: RSS_1 & RSS_2.

☞ RSS_1 has $n_1–(K+1)$ *df* & RSS_2 has $n_2–(K+1)$ *df*.

☞ $RSS_1 + RSS_2 = URSS$ with $n_1+n_2–2(K+1)$ *df*.

2. Estimate pooled model (under H_0: β's are stable).

☞ RSS from this model is RRSS with $n–(K+1)$ *df* *where* $n = n_1+n_2$.

3. The test-statistic (under H_0):

$$F_{cal} = \frac{[RRSS - URSS] \big/ (K+1)}{URSS \big/ [n - 2(K+1)]}$$

4. Find the critical value: $F_{K+1,n-2(K+1)}$ from table.

5. If $F_{cal}>F_{tab}$, reject H_0 of stable parameters (and favour H_a: there is structural break).

1.4 Structural Stability

e.g.: we have the ff results from estimation of real consumption from real disposable income:

i. For the period 1974-1991: $\mathbf{cons_i = \alpha_1+\beta_1{*}inc_i+u_i}$

$$\text{Consumption} = \mathit{153.95} + \mathit{0.75}{*}\text{Income}$$

p-value: (0.000) (0.000)

RSS = 4340.26114; R^2 = 0.9982

ii. For the period 1992-2006: $\mathbf{cons_i = \alpha_2+ \beta_2{*}inc_i+u_i}$

$$\text{Consumption} = \mathit{1.95} + \mathit{0.806}{*}\text{Income}$$

p-value: (0.975) (0.000)

RSS = 10706.2127; R^2 = 0.9949

iii. For the period 1974-2006: $\mathbf{cons_i = \alpha+ \beta{*}inc_i+u_i}$

$$\text{Consumption} = \mathit{77.64} + \mathit{0.79}{*}\text{Income}$$

t-ratio: (4.96) (155.56)

RSS = 22064.6663; R^2 = 0.9987

1.4 Structural Stability

1. **URSS = $RSS_1 + RSS_2$ = 15064.474**
2. **RRSS = 22064.6663**

☞ **K = 1 and K + 1 = 2; n_1 = 18, n_2 = 15, n = 33.**

3. **Thus,**

$$F_{cal} = \frac{[22064.6663 - 15064.474] \big/ 2}{15064.474 \big/ 29} = 6.7632981$$

4. **p-value = Prob(F-tab > 6.7632981) = 0.003883**
5. **Reject H_0 at α=1%. Thus, there is structural break.**

☞ **The pooled consumption model is an inadequate specification; we should run separate regressions.**

☞ **The above method of calculating the Chow test breaks down if either $n_1 < K+1$ or $n_2 < K+1$.**

☞ **Solution: use Chow's second (predictive) test!**

1.4 Structural Stability

- If, for instance, $n_2 < K+1$, then the F-statistic will be altered as follows:

$$F_{cal} = \frac{[RRSS - RSS_1] \big/ n_2}{RSS_1 \big/ n_1 - (K+1)}$$

- The Chow test tells if the parameters differ on average, but not which parameters differ.
- Also, it requires that all groups have the same σ^2.
- This assumption is questionable: if parameters can be different, then so can the variances be.
- One way of correcting for unequal σ^2 is to use dummy variable regression with *robust standard errors.*

1.4 Structural Stability

Using Dummy variables vs Chow’s test

☞ **Comparing the two methods, it is preferable to use the method of dummy variables regression.**

☞ **This is because with the method of DVR:**

1. ***We run only one regression.***
2. ***We can test whether the change is in the intercept only, in the slope only, or in both.***

Lab session

Use “*Chowtest.xls*” data to practice what we learnt in previous sections

END OF CHAPTER ONE

THANK YOU FOR BEING WITH ME

BEING @ COMMITTED

Stay Safe!

2.2.“Dummy dependent variable”:

Qualitative Response Model

2.2.1 Introduction

☞ **Qualitative Response Model shows situations in which the dependent variable in a regression equation simply represents a discrete choice assuming only a limited number of values**

☞ **Such a model is called**

- □ **Limited dependent variable**
- □ **Discrete dependent variable**
- □ **Qualitative response**

Categories of Qualitative Response Models

☞ **there are two broad categories of QRM**

2.2.1 Introduction

1. **Binomial Model: it shows the choice between two alternatives**

   e.g: Decision to participate in labor force or not

2. **Multinomial models: the choice between more than two alternatives**

e.g: Y= 1, occupation is farming

=2, occupation is carpentry

=0, government employee

Important terminologies

- **Binary variables: variables that have two categories and used to an event that has occurred or some characteristics present.**

2.2.1 Introduction

- ☞**Ordinal variables: variables that have categories that can be ranked.**
  - ☞**Example: Rank according to education attainment (Y)**

$$Y = \begin{cases} 0 \text{ if primary education} \\ 1 \text{ if secondary education} \\ 2 \text{ if university education} \end{cases}$$

  - ☞**Nominal variables: variables occur when there are multiple outcomes that cannot be ordered.**

2.2.1 Introduction

- **Example: Occupation can be grouped as farming, fishing, carpentry etc.**

$$Y = \begin{cases} 0 \text{ if farming} \\ 1 \text{ if fishermen} \\ 2 \text{ if carpentry} \\ 3 \text{ if government employee} \end{cases}$$

N.B: Numbers are assigned arbitrarily

- **<u>Count variables</u>: indicate the number of times some event has occurred.**
  - **Example: How many years of education you have attend?**
  - **In all of the above situations, the variables are discrete valued.**

2.2.2 Qualitative Choice Analysis

- **In such cases instead of standard regression models, we apply different methods of modeling and analyzing discrete data.**
- **Qualitative choice models may be used when a decision maker faces a choice among:**
  - **The number of choices if finite**
  - **The choices are mutually exclusive (the person chooses only one of the alternatives)**
  - **The choices are exhaustive (all possible alternatives are included)**

Qualitative choice analysis

- ☞ **Throughout our discussion we shall restrict ourselves to cases of qualitative choice where the *set of alternatives is binary*.**
- ☞ **For the sake of convenience the dependent variable is given a value of 0 or 1.**
  - ☞ **Example: Suppose the choice is whether to work or not. The discrete dependent variable we are working with will assume only two values 0 and 1:**

$$Y_i = \begin{cases} 1\, if\ \ i^{th}\ \ individual\ is\ working \\ 0\, if\ \ i^{th}\ \ individual\ is\ notworking \end{cases}$$

where i = 1, 2, …, n.

Qualitative choice analysis

- ☞ The independent variables (called factors) that are expected to affect an individual's choice may be X_1 = age, X_2 = marital status, X_3 = gender, X_4 = education, and the like.
- ☞ These are represented by a matrix X.

Regression Approach

- ☞ The economic interpretation of discrete choice models is typically based on the principle of utility maximization leading to the choice of, say, A over B if the utility of A exceeds that of B.
- ☞ Let U^1 be the utility from working/seeking work and let U^0 be the utility form not working. Then an individual will choose to be part of the labour force if $U^1 - U^0 > 0$, and this decision depends on a number of factors X.

Qualitative choice analysis

- **The probability that the i^{th} individual chooses alternative 1th (i.e. works) given his/her individual characteristics, Xi is:**

$$\mathrm{P}_i = pr(Y_i = 1/X_i) = \Pr[(U^1 - U^o)_i > 0] = G(X_i, \beta)$$

- **The vector of parameters $\beta = (\beta_1, \beta_2, \ldots., \beta_k)$ (measures the impact of changes in X (say, age , marital status, gender, education, occupation, and the like) on the probability of labor force participation.**
- **the probability that the i^{th} individual chooses alternative 0 (i.e. not to work) is given by:**

$$pr(Y_i = 0/X_i) = 1 - \mathrm{P}_i = 1 - \Pr[(U^1 - U^o)_i > 0] = 1 - G(X_i, \beta)$$

Qualitative choice analysis

- **Here *Pi* is called the response probability and *(1-Pi)* is called the non-response probability.**
- **The mean response of the i^{th} individual given his/her individual characteristics X_i is:**

$$E(Y_i \, / \, Xi \; = 1 * \{G(Xi, \beta)\} + 0 * \{1 - G(X_i, \beta)\} = G(X_i, \beta)$$

- **The problem is thus to choose the appropriate form of $G(X_i, \beta)$.**
- **There are several methods to analyze regression models where the dependent variable is binary.**
- **the four most commonly used approaches to estimating binary response models (Type of binomial models). These are:**

Qualitative choice analysis

- **Linear probability models**
- **The logit model**
- **The probit model**
- **The tobit (censored regression) model**

1. The Linear Probability Models (LPM)

- **The term linear probability model is used to denote a regression model in which the dependent variable y is a dichotomous variable taking the value 1 or 0.**

Linear Probability Models

- In the 1960's and early 1970's the **linear probability model** was widely used mainly because it is a model that can be **easily** estimated using multiple regression analysis.
- A "limited dependent variable" y is one which takes a limited set of values. The most common cases are: Binary: $y \varepsilon \{0,1\}$; Multinomial: $y \varepsilon \{0,1,2,\ldots k\}$; ;Integer: Integer: $y \varepsilon \{0,1,2,..\}$ $y \varepsilon \ R^{+}$.
- The traditional approach to the estimation of limited dependent variable models is **parametric maximum likelihood.**

Linear Probability Models

☞A parametric model is constructed, allowing the construction of the likelihood function.

☞A more modern approach is semi-parametric, eliminating the dependence on a parametric distributional assumption. We will limit our discuss to parametric approach.

☞When we use a linear regression model to estimate probabilities, we call the model the linear probability model.

☞The linear probability model is the regression model applied to a binary dependent variable.

Linear Probability Models

☞ **The linear probability model defines** $G(X_i, \beta) = X_i \beta$

☞ **The regression model when Y is a binary variable is thus,** $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon = X\beta + \varepsilon$

☞ **Where y takes only two value: 0 & 1, and** β_j **cannot be interpreted as the change in Y given a one-unit increase in** $X_{j,,}$ **holding all other factors constant rather Y changes either from 0 to 1 or from 1 to 0.**

☞ **If we assume that the zero conditional mean assumption holds, that is,** $E(\varepsilon)=0$ **, then we have, as always**

Linear Probability Models

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k = X\beta$$

☞ $pr(Y=1/X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k = X\beta$ → **it says that the probability of success is a linear function of the Xj; it is called the response probability.**

☞ $\Pr(Y=0/X) = 1 - pr(Y=1/X)$ → **it is the non-response probability, and it is also a linear function of the X_j.**

☞ ***Interpretation of LPM***

- ☞ ***The response probability is linear in the parameters of X_j.***

Linear Probability Models

- Example: Suppose that from hypothetical data of house ownership and income and thus, the LPM estimated by OLS (on home ownership) is given as follows:

$$
\begin{aligned}
\hat{Y}_i &= -0.9457 + 0.1021X_i \\
&\quad\;(0.1228) \qquad\quad (0.0082) \\
t &= (-7.6984) \qquad\; (12.515) \\
&\qquad\qquad R^2 = 0.8048
\end{aligned}
$$

- The above regression is interpreted as follows
- The intercept of –0.9457 gives the “probability” that a family with zero income will own a house. Since this value is negative, and since probability cannot be negative, we treat this value as zero.
- The slope value of 0.1021 means that for a unit change in income, on the average the probability of owning a house increases by 0.1021 or about 10 percent. This is so whether the income level is increased or not. This seems patently unrealistic. In reality one would expect that P_i is non-linearly related to X_i.

Linear Probability Models

☞β_j **measures the change in the probability of success when Xj changes, holding other factors fixed.**

☞***Advantages of LPM*: Easy to estimate and interpret; it is too simple.**

☞***Drawbacks of LPM:***

1. **The dependent variable is discrete while the independent variable is the combination of discrete and continuous variables.**
2. **Usually we arbitrarily (or for convenience) use 0 and 1 for Yi . If we use other values for Yi , say 3 and 4, will also change even if the vector of factors Xi remains unchanged.**

Linear Probability Models

3. Error term assumes only two values.

☞ **If Yi=1 then** $\varepsilon_i = 1 - X_i\beta$ **with the Probability, Pi;**

☞ **If Yi=0 then** $\varepsilon_i = -X_i\beta$ **with Probability, 1-Pi;**

☞ **The variance of the disturbance terms depends on the X's and is thus not constant.; i.e., error term is not normally distributed.**

☞ **Now by definition** $Var(\varepsilon_i) = [E(\varepsilon_i - E(\varepsilon_i)]^2 = E(\varepsilon_i{}^2)$ **since** $E(\varepsilon_{i)} = 0$ **by assumption. Therefore, using the preceding probability distribution of we obtain:**

$$Var(\varepsilon_i) = E(\varepsilon_i{}^2) = (-\ \beta\ X_i)^2\ (1\text{-}P_i) + (1 -\ \beta\ X_i)^2\ (P_i)$$

Linear Probability Models

☞ **This shows that the variance of U_i is heteroscedastic because it depends on the conditional expectation of Y, which, of course, depends on the value taken by X.**

☞ **Thus, the OLS estimator of β is inefficient and the standard errors are biased, resulting in incorrect test.**

4. The expectation (mean) of conditional on the exogenous variables Xi is non sense.

$$E(\varepsilon/X_i)=(1-X_i\beta)P_i+(-X_i\beta)(1-P_i)=P_i-X_i\beta$$

Setting this mean to zero as in →the classical regression analysis

Linear Probability Models

$$E(\varepsilon_i / X_i) = 0 \Rightarrow P_i = X_i\beta$$

$$Y_i = P_i + \varepsilon_i \Rightarrow \varepsilon = Y_i - P_i$$

- **That is, the binary (discrete) disturbance term is equal to the difference between a binary variable Yi and a continuous response probability Pi. Clearly this does not make sense.**

☞**the probability of an event is always a number between 0 and 1 (inclusive). But here we can see that:** $Pi = pr(Yi = 1 | X_i) = X_i\beta$ **, i.e., Pi can take on any value (even negative numbers) leading to nonsense probabilities, the fitting probabilities**

Linear Probability Models

5. Non-Sensical Predictions

- **The LPM produces predicted values outside the normal range of probabilities (0,1). It predicts value of Y that are negative and greater than 1.**
- **This is the real problem with the OLS estimation of the LPM.**

6. Non-normality of U_i

- **Although OLS does not require the disturbance (U's) to be normally distributed, we assumed them to be so distributed for the purpose of statistical inference, that is, hypothesis testing, etc. But the assumption of normality for U_i is no longer tenable for the LPMs because like Y_i, U_i takes on only two values.**

Linear Probability Models

7. *Functional Form*

☞Since the model is linear, a unit increase in X results in a constant change of in the probability of an event, holding all other variables constant.

☞The increase is the same regardless of the current value of X.

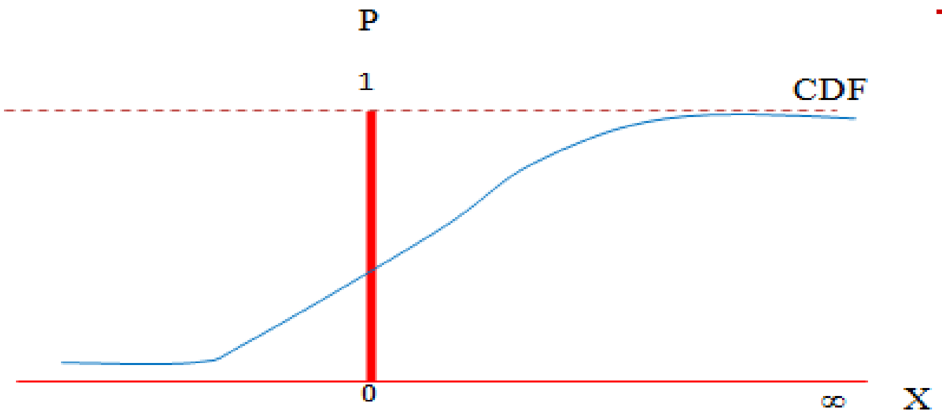
8. *Questionable Value of R^2 as a Measure of Goodness of Fit (lower R^2 values)*

☞ In sum, non-normality of ui; possibility of Yi lying outside the 0-1 range; hetroscedasticity of Ui; lower R2 values; the basic problem is not logically attractive.

Linear Probability Models

- b/c the above mentioned problems, the LPM model is ***not recommend for empirical works***.
- Therefore, what we need is a (probability) model that has the following two features:
  - As X_i increases, $P_i = E(Y = 1/X)$ increases but never steps outside the 0-1 interval.
  - The relationship between P_i and X_i is non-linear, that is, “ one which approaches zero at slower and slower rates as X_i gets small and approaches one at slower and slower rates as X_i gets very large”

Linear Probability Models



- ☞The above S-shaped curve is very much similar with the cumulative distribution function (CDF) of a random variable.
- ☞the CDF of a random variable X is simply the probability that it takes a value less than or equal to x_0, were x_0 is some specified numerical value of X.
- ☞ In short, F(X), the CDF of X, is $F(X = x_0) = P(X \leq x_0)$.

Linear Probability Models

☞ **Therefore, one can easily use the CDF to model regressions where the response variable is dichotomous, taking 0-1 values.**

☞ **The CDFs commonly chosen to represent the 0-1 response models are.**

- □ **the logistic – which gives rise to the logit model**
- □ **the normal – which gives rise to the probit (or normit) model**

2. *Logit model*

- ***Although LPM* is simple to estimate and use, but the two most important disadvantages are:**
  - **the fitted probabilities can be less than zero or greater than one and**
  - **the partial effect of any explanatory variable is constant.**

Logit model

- These limitations of the LPM can be overcome by using more sophisticated binary response models.

$$P(Y = 1/X) = P(Y = 1/X_1, X_2, ..., X_k)$$

- In a binary response model, interest lies primarily in the response probability.
- where we use X to denote the full set of explanatory variables.
- For example, when Y is an employment indicator, X might contain various individual characteristics such as education, age, marital status, and other factors that affect employment status, including a binary indicator variable for participation in a recent job training program.

Logit model

- In the LPM, we assume that the response probability is linear in a set of parameters, β_j .
- To avoid the LPM limitations, consider a class of binary response models of the form.

$$P(Y = 1/X) = G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k) = G(X\beta)$$

- Where G is a function taking on values strictly between 0 & 1: $0 < G(z) < 1,$ for all real numbers z.
- This ensures that the estimated response probabilities are strictly between zero and one.

Logit model

- **Various nonlinear functions have been suggested for the function G in order to make sure that the probabilities are between zero and one.**
- **In the logit model, G is the logistic function:**

$$G(z) = \frac{\exp(z)}{[1 + \exp(z)]} = \Lambda(z) = \frac{e^{X\beta}}{1 + e^{X\beta}}$$

- **which is between zero and one for all real numbers z. This is the cumulative distribution function (cdf) for a standard logistic random variable.**

Logit model

- the response probability P(Y =1/X) is evaluated as:

$$P = P(Y = 1|X) = \frac{e^{X\beta}}{1 + e^{X\beta}}$$

- the non response probability P(Y =0/X) is evaluated as:

$$1 - P = P(Y = 0|X) = 1 - \frac{e^{X\beta}}{1 + e^{X\beta}} = \frac{1}{1 + e^{X\beta}}$$

- Note that: both response and non- response probabilities lie in the interval [0 , 1] , and hence, are interpretable.
- Odd ratio: the ratio of the response probabilities (Pi) to the non response probabilities (1-Pi).

Logit model

- **For the logit model, the odds ratio is given by:**

$$\frac{P}{1-P}=\frac{P(Y=1|X)}{P(Y=0|X)}=\frac{\frac{e^{X\beta}}{1+e^{X\beta}}}{\frac{1}{1+e^{X\beta}}}=e^{X\beta}=e^{\beta_0+\beta_1}+e^{\beta_2 X_2}+e^{\beta_3 X_3}+\ldots e^{\beta_k X_k}$$

- **The natural logarithm of the odds ratio(log-odds ratio) is:**

$$Li = \ln(\frac{P}{1-P}) = \beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k$$

- **L(the log of the odds ratio) is linear in X as well as β (the parameters). L is called the logit and hence the name logit model is given to it.**

Logit model

☞ **Thus, the log-odds ratio is a linear function of the explanatory variables.**

☞ **For the LPM it is Pi, which is assumed to be a linear function of the explanatory variables.**

☞ ***Features of logit model***

☞ **As *P* goes from 0 to 1 (i.e., as *Z* varies from −∞ to +∞), the logit *L* goes from −∞ to +∞. That is, although the probabilities (of necessity) lie between 0 and 1, the logits are not so bounded.**

☞ **Although *L* is linear in *X*, the probabilities themselves are not. This property is in contrast with the LPM model where the probabilities increase linearly with *X*.**

Logit model

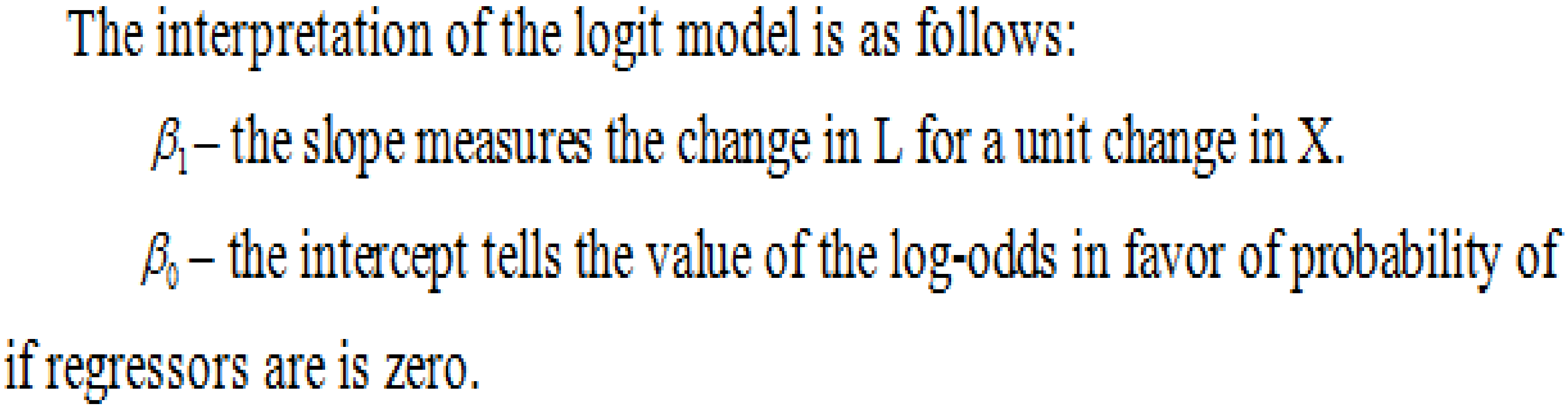
☞If *L*, the logit becomes negative and increasingly large in magnitude as the odds ratio decreases from 1 to 0 and becomes increasingly large and positive as the odds ratio increases from 1 to infinity.

☞LPM assumes that *Pi* is linearly related to X_i, the logit model assumes that the log of the odds ratio is linearly related to X_i .

Interpretation: Be remind that we doesnot directly interpreted the coefficients of the variables rather we interpreted their marginal effects and

Logit model

- *β*2, the slope, measures the change in *L* for a unit change in *X*, that is, it tells how the log-odds in favor of owning a house change as income changes by a unit, say, birr 1000.
- The intercept, *β*1 is the value of the logodds in favor of owning a house if income is zero.



The interpretation of the logit model is as follows:

β_1 – the slope measures the change in L for a unit change in X.

β_0 – the intercept tells the value of the log-odds in favor of probability of

if regressors are is zero.

Probit model

☞The estimating model that emerges from the normal CDF is popularly known as the probit model.

☞In the probit model, G is the standard normal cumulative distribution function (cdf), which is expressed as an integral:

☞ In the probit model, G is the standard normal cumulative distribution function

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v)dv$$

☞Where $\Phi(z)$ is the standard normal density

Probit model

$$\Phi(z) = (2\pi)^{-1/2}\exp(-z^2/2)$$

☞ The estimating model that emerges from the normal CDF is popularly known as the probit model.

☞ Here the observed dependent variable Y, takes on one of the values 0 and 1 using the following criteria.

☞ Define a latent variable Y* such that $Y^* = X_{1i}\beta + \varepsilon_i$

$$Y = \begin{cases} 1 \text{ if } \ \mathrm{Y_i}^* > 0 \\ 0 \text{ if } \ \mathrm{Y_i}^* \le 0 \end{cases}$$

Probit model

- ☞The latent variable Y^* is continuous ($-\infty < Y^* < \infty$).
- ☞It generates the observed binary variable Y.
- ☞An observed variable, Y can be observed in two states:
  - ☞if an event occurs it takes a value of 1
  - ☞if an event does not occur it takes a value of 0
- ☞The latent variable is assumed to be a linear function of the observed X's through the structural model.

Probit model

- However, since the latent dependent variable is unobserved the model cannot be estimated using OLS.
- Maximization of the likelihood function for either the probit or the logit model is accomplished by nonlinear estimation methods. Maximum likelihood can be used instead.
- Most often, the choice is between normal errors and logistic errors, resulting in the probit (normit) and logit models, respectively.

Probit model

- The coefficients derived from the maximum likelihood (ML) function will be the coefficients for the probit model, if we assume a normal distribution.
- If we assume that the appropriate distribution of the error term is a logistic distribution, the coefficients that we get from the ML function will be the coefficient of the logit model.
- In both cases, as with the LPM, it is assumed that $E[\in_i / X_i] = 0$

Probit model

- ☞In the probit model, it is assumed that Var $(\in_i/X_i) = 1$; In the logit model, it is assumed that Var $(\in_i/X_i) = \pi^2/3$.
- ☞Hence, the estimates of the parameters (β's) from the two models are not directly comparable.
- ☞But as Amemiya suggests, a logit estimate of a parameter multiplied by 0.625 gives a fairly good approximation of the probit estimate of the same parameter.

Probit model

- ☞ Similarly the coefficients of LPM and logit models are related as follows:
- ☞ $\beta_{\text{LPM}} = 0.25_{\text{Logit}}$, except for intercept
- ☞ $\beta_{\text{LPM}} = 0.25_{\text{Logit}} + 0.5$ for intercept
- ☞ The standard normal cdf has a shape very similar to that of the logistic cdf.

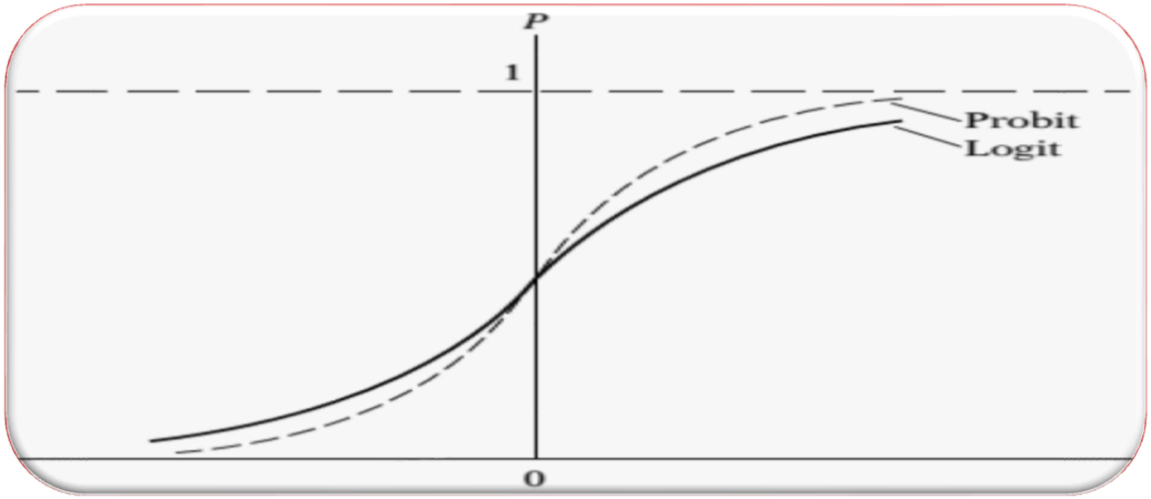


Figure 1: logit and probit cumulative distributions

Probit model

- The estimating model that emerges from the normal CDF is popularly known as the probit model, although sometimes it is also known as the normit model.
- Note that both the probit and the logit models are estimated by *Maximum Likelihood Estimation.*

Probit model

- ***Interpreting the Probit and Logit Model Estimates***
- **The coefficients give the signs of the partial effects of each Xj on the response probability, and the statistical significance of Xj is determined by whether we can reject H0: Bj=0 at a sufficiently small significance level.**
- **However, the magnitude of the estimated parameters (dZ/dX) has no particular interpretation. We care about the magnitude of dProb(Y)/dX.**
- **From the computer output for a probit or logit estimation, you can interpret the statistical significance and sign of each coefficient directly.**

Probit model

- ☞In the *linear regression model*, the slope coefficient measures the change in the average value of the regressand for a unit change in the value of a regressor, with all other variables held constant.
- ☞In the *LPM*, the slope coefficient measures directly the change in the probability of an event occurring as the result of a unit change in the value of a regressor, with the effect of all other variables held constant.

Probit model

- In the *logit model* the slope coefficient of a variable gives the change in the log of the odds associated with a unit change in that variable, again holding all other variables constant.
- But as noted previously, for the logit model the rate of change in the probability of an event happening is given by $\beta j\ Pi(1 - Pi\ )$, where βj is the (partial regression) coefficient of the *j*th regressor. But in evaluating Pi , all the variables included in the analysis are involved.

Probit model

- ☞**In the *probit model*, as we saw earlier, the rate of change in the probability is somewhat complicated and is given by $\beta j\ f\ (Zi\ )$, where $f\ (Zi)$ is the density function of the standard normal variable and $Zi = \beta 1 + \beta 2X2i + \cdot\ \cdot\ \cdot + \beta kXki$, that is, the regression model used in the analysis.**
- ☞**Thus, in both the logit and probit models all the regressors are involved in computing the changes in probability, whereas in the LPM only the j^{th} regressor is involved. This difference may be one reason for the early popularity of the LPM model.**

Probit vs logit model

Is logit or probit model is preferable?

☞ In most applications the models are quite similar, the main difference being that the logistic distribution has slightly fatter tails.

☞ That is to say, the conditional probability *Pi* approaches zero or one at a slower rate in logit than in probit.

☞ Therefore, there is no compelling reason to choose one over the other.

☞ In practice many researchers choose the logit model because of its comparative mathematical simplicity.

☞ The standard normal cdf has a shape very similar to that of the logistic cdf.

Probit vs logit model

- ☞ The probit and logit models differ in the specification of the distribution of the error term u.

- ☞ The difference between the specification and the linear probability model is that in the linaer probability model we analyses the dichotomous variables as they are, where as we assume the existence of an underlying latent variable for which we observe a dichotomous realization.

Probit vs logit model

- The probit model and the logit model are not directly comparable. The reason is that, although the standard logistic (the basis of logit) and the standard normal distributions (the basis of probit) both have a mean value of zero, their variances are different; 1 for the standard normal (as we already know) and $\pi^2/3$ for the logistic distribution, where $\pi \approx 22/7$.

- Therefore, if you multiply the probit coefficient by about 1.81 (which is approximately = $\pi/\sqrt{3}$), you will get approximately the logit coefficient.

Probit vs logit model

- **The R2’s for the linear probability model are significantly lower than those for the logit and probit models. Alternative ways of comparing the models would be:**
  - **To calculate the sum of squared deviations from predicted probabilities**
  - **To compare the percentages correctly predicted**
  - **To look at the derivatives of the probabilities with respect to a particular independent variable.**

Tobit Model

- **An extension of the probit model is the tobit model developed by James Tobin.**
- **Let us consider the home ownership example.**
- **Suppose we want to find out the amount of money the consumer spends in buying a house in relation to his or her income and other economic variables.**
- **If a consumer does not purchase a house, obviously we have no data on housing expenditure for such consumers; we have such data only on consumers who actually purchase a house.**

Tobit Model

- Thus, consumers are divided into two groups, one consisting of say, N_1 consumers about whom we have information on the regressors (say income, interest rate etc) as well as the regresand (amount of expenditure on housing) and another consisting of say, N_2 consumers about whom we have information only on the regressors but on not the regressand.
- A sample in which information on regressand is available only for some observations is known as a censored sample. Therefore, the tobit model is also known as a censored regression model.

Tobit Model

- Mathematically, we can express the tobit model as

$$Y_i = \begin{cases} \beta_0 + \beta_1 X_{1i} + u_i \text{ if RHS} > 0 \\ 0, \text{otherwise} \end{cases}$$

- Where RHS= right hand side
- The method of maximum likelihood can be used to estimate the parameters of such models.

Measuring goodness of fit

- The conventional measure of goodness of fit,R^2 , is not particularly meaningful in binary regressand models. Measures similar to R^2,

Measuring goodness of fit

☞ **Measures based on likelihood ratios: The conventional measure of goodness of fit,R^2 , is not particularly meaningful in binary regressand models.**

☞ **Measures similar to R^2, called pseudo R^2, are available, and there are a variety of them.**

Measures based on likelihood ratios

☞ **Let L_{UR}be the maximum likelihood function when maximized with respect to all the parameters and L_R be the maximum likelihood function when maximized with restrictions $\beta_i = 0$.**

Measuring goodness of fit

$$R^2 = 1 - \left(\frac{L_R}{L_{UR}}\right)^{\frac{2}{n}}$$

- the qualitative dependent variable model, the likelihood function attains an absolute maximum of 1. This means that, $L_R \leq L_{UR} \leq 1$
- Cragg and Uhler (1970) suggested a pseudo R^2 that lies between 0 and 1. $R^2 = \dfrac{L_{UR}^{\frac{2}{n}} - L_R^{\frac{2}{n}}}{(1 - L_R^{\frac{2}{n}})L_{UR}^{\frac{2}{n}}}$
- Mc Fadden (1974) defined R^2 as

$$R^2 = 1 - \frac{\log L_{UR}}{\log L_R}$$

Measuring goodness of fit

- Another goodness-of-fit measure that is usually reported is the so-called *percent correctly predicted*, which is computed as follows. For each i, we compute the estimated probabilit y that Yi takes on the value one, $\hat{Y}i$.

- If $\hat{Y}i \geq 0.5$ the prediction of Yi is unity, and if $\hat{Y}i < 0$ Yi is predicted to be zero. The percentage of times the predicted $\hat{Y}i$ matches the actual Yi (which we know to be zero or one) is the percent correctly predicted.

$$\hat{Y}i^* = \begin{cases} 1 \text{ if } \hat{Y}_i \geq 0.5 \\ 0 \text{ if } \hat{Y}_i < 0.5 \end{cases}$$

$$CountR^2 = \frac{number\ of\ correct\ predictions}{total\ number\ of\ observations}$$

Brainstorm questions

1. **Why LPM is not recommendable for empirical analysis?**
2. **Drive logodds ratio.**
3. **For logit and probit model we can use OLS estimators.**
4. **For logit and probit model we can use maximum likelihood estimators.**

$$\frac{\pi^2}{2}$$

5. **The variance that logit model assume is where as probit model assumes 1.**

Lab session

Use "*lecture_2.xls*" data to practice what we learnt in previous sections

END OF CHAPTER TWO

THANK YOU FOR BEING WITH ME

BEING @ COMMITTED
stay Safe!

CHAPTER THREE

Introduction to Basic Regression Analysis with Time Series Data

3.1 Nature of the Time Series data

3.2 Stationary & non stationary stochastic Processs

3.3 Trend & Difference stationary stochastic process

3.4 Integrated Stochastic Process

3.5 Tests of Stationary

3.1 Nature of the Time Series data

- Time series data have become so frequently and intensively used in empirical research and thus, econometricians have recently begun to pay very careful attention to such data.
- *Time series data* are data collected for a *single entity* (person, firm, and country) collected (observed) at *multiple time* periods.
- A time series is a *sequence of numerical data* in which each item is associated with a particular instant in time.
- Example: Monthly unemployment, weekly measures of money supply, M1 and M2, daily closing prices of stock indices, and so on

3.1 Nature of the Time Series data

- **Thus, the way we collect time series data can be characterised as daily (stock prices, weather report), weekly (gasoline supplied in thousands of barrels), monthly (unemployment rate, CPI), quarterly (GDP), & annual (GDP, Budget).**
- **Quinquennially: Every 5 years (e.g: the census of manufactures)**
- **Decennially (e.g: Census of population)**
- **Exchange rates daily date for 2 years=730 observation**
- **Inflation rate for Ethiopia, quarterly data for 30 years =30*4=120 observation**

3.1 Nature of the Time Series data

- **Gross domestic investment in Ethiopia of annual data for 40 yrs; 40*1= 40 observations.**
- **Time Series data Vs Cross sectional data**

Time Series

- It coming with temporal ordering over period of time on a single entity.
- the past can affect the future, but not vice versa
- We have d/t data for d/t samples
- Also viewed as random varaibles; we do not know what the annual growth in output will be in Ethiopia during the coming year. i.e, the outcomes of these random variables are not foreknown.

Cross sectional

- At a given point of time
- a different sample has drawn from the population will generally yield different values of the independent and dependent variables
- We have d/t data for d/t year
- The OLS estimates computed from d/t random samples will generally differ and this why we consider OLS estimators to be random variables.

3.1 Nature of the Time Series data

- ☞ **A sequence of random variables indexed by time is called a *stochastic process or a time series process*. (“Stochastic” is a synonym for random.)**

- ☞ **When we collect a time series data set, we obtain one possible outcome, or realization, of the stochastic process.**

- ☞ **We can only see a single realization, because we cannot go back in time and start the process over again.**

- ☞ **This is analogous to cross-sectional analysis where we can collect only one random sample.)**

3.1 Nature of the Time Series data

Important terminology :

- **Univariate analysis examines a single data series.**
- **Bivariate analysis examines a pair of series.**
- **The term vector indicates that we are considering a number of series: two, three, or more.**
- **The term ‘‘vector’’ is a generalization of the univariate and bivariate cases.**

3.2 Stationary and non-stationary Stochastic Processes

☞ A random or stochastic process is a collection of random variables ordered in time.

☞ If we let Y denote a random variable, and if it is continuous, we denote it as Y(t), but if it is discrete, we denoted it as Yt.

☞ Example of Yt is GDP, CPI, PDI; since most economic data are collected at discrete points in time, we use Yt notation.

☞ If we let Y represent GDS, for our data we have Y1,Y2,Y3,...,Y21,Y22,Y23, where the subscript 1 denotes the first observation (i.e., GDS of 1991/1992) and the subscript 23 denotes the last observation (i.e.,

3.2 Stationary and non-stationary Stochastic Processes

☞Example of Y(t) is electro cardiogram, record of heart activity.

☞N.B: Each of these Y's is a random variable.

A. Stationary Stochastic Processes

☞ A stochastic process is said to be stationary "*if its mean and variance are constant over time* and the value of the covariance between the two time periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed.

3.2 Stationary and non-stationary Stochastic Processes

- **In the time series literature, such a stochastic process is known as a weakly stationary, or covariance stationary, or second-order stationary, or wide sense, stochastic process.**
- **To explain weak stationarity, let Yt be a stochastic time series with these properties:**

$$Mean \quad : E(Y_t) = \mu$$

$$Variance : Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$$

$$Covariance: \gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)]$$

3.2 *Stationary and non-stationary Stochastic Processes*

☞Where γk, the covariance (or autocovariance) at lag k, is the covariance between the values of Yt and Yt+k, that is, between two Y values k periods apart.

☞If k=0, we obtain γ0, which is simply the variance of Y(=σ2); if k=1,γ1 is the covariance between two adjacent values of Y.

$$\gamma_1 = E[(Y_1 - \mu)(Y_{t+1} - \mu)]$$

☞Suppose we shift the origin of Y from Yt to Yt+m (say, from 1997 to 2002 for our GDS data). Now if Yt is to be stationary, the mean, variance, and autocovariances of Yt+m must be the same as those of Yt.

3.2 Stationary and non-stationary Stochastic Processes

☞ In short, if a time series is stationary, its mean, variance, and autocovariance (at various lags) remain the same no matter at what point we measure them; that is, they are ***time invariant.***

☞ Such a time series will tend to return to its mean (called mean reversion) and fluctuations around this mean (measured by its variance) will have a broadly constant amplitude.

☞ If a time series is not stationary in the sense just defined, it is called a nonstationary time series (keep in mind we are talking only about weak stationarity).

☞ In other words, a nonstationary time series will have a **time-varying mean** or **a time-varying variance** or **both**.

3.2 *Stationary and non-stationary Stochastic Processes*

Why Stationary time series are important?

- **Because if a time series is nonstationary, we can study its behavior only for the time period under consideration.**
- **Each set of time series data will therefore be for a particular episode.**
- **As a consequence, it is not possible to generalize it to other time periods. Therefore, for the purpose of *forecasting*, such (nonstationary) time series may be of little practical value.**

A. Non stationary Stochastic Processes

- **A special type of stochastic process (or time series), is called, a *purely random, or white noise, process*.**

3.2 Stationary and non-stationary Stochastic Processes

- **We call a stochastic process purely random if it has zero mean, constant variance σ2, and is serially uncorrelated.**
- **One of the classical example of non stationary time series is the random walk model (RWM).**
- **It is often said that asset prices, such as stock prices or exchange rates, follow a random walk; that is, they are nonstationary.**
- **We have two types of random walks**
  - **(1) random walk without drift (i.e., no constant or intercept term) and**
  - **(2) random walk with drift (i.e., a constant**

3.2 Stationary and non-stationary Stochastic Processes

1. Random walk without drift

- **Suppose u_t is a white noise error term with mean 0 and variance σ^2.**
- **Then the series Y_t is said to be a random walk if**

$Y_t = Y_{t-1} + u_t$ → shows, the value of Y at time t is equal to its value at time (t−1) plus a random shock; thus it is an AR (1) model.

- **We can think of $Y_t = Y_{t-1} + u_t$ as a regression of Y at time t on its value lagged one period.**

$$Y_1 = Y_0 + u_t$$

$$Y_2 = Y_1 + u_2 = Y_o + u_1 + u_2$$

$$Y_3 = Y_2 + u_3 = Y_0 + u_1 + u_2 + u_3$$

3.2 *Stationary and non-stationary Stochastic Processes*

- **If the process started at some time 0 with a value of Y0, we have**

$$Y_1 = Y_0 + \sum u_t$$

$$E(Y_1) = E(Y_0 + \sum u_t) = Y_0$$

$$Var(Y_t) = t\sigma^2$$

- **the *mean of Y is equal to its initial, or starting, value, which is constant*, but *as t increases, its variance increases indefinitely*, thus violating a condition of stationarity.**

- **In short, the *RWM without drift is a nonstationary stochastic process*. In practice Y0 is often set at zero, in which case E (Yt) =0.**

3.2 *Stationary and non-stationary Stochastic Processes*

- **An interesting feature of RWM is the persistence ~~of random shocks (i.e., random errors), which~~ is clear from** $Y_1 = Y_0 + \sum u_t$**: Yt is the sum of initial Y0 plus the sum of random shocks.**
- **As a result, the impact of a particular *shock does not die away*.**
- **For example, if U_2=2 rather than U_2=0, then all Yt’s from Y2 onward will be 2 units higher and the effect of this shock never dies out.**
- **That is why random walk is said to have an infinite memory. The implication is that, random walk remembers the shock forever; that is, it has infinite memory.**

3.2 *Stationary and non-stationary Stochastic Processes*

☞ **We can rewrite the above equation as:** $Y_1 = Y_0 + \sum u_t$

$$(Y_t - Y_{t-1}) = \Delta Y_t = u_t$$

☞ **Where Δ is the first difference operator. It is easy to show that, while Yt is nonstationary, its *first difference is stationary*. In other words, the first differences of a random walk time series are stationary.**

2. Random Walk with Drift

Let us modify $Y_1 = Y_0 + \sum u_i$ **as follows:**

$$Y_t = \delta + Y_{t-1} + u_t$$

3.2 *Stationary and non-stationary Stochastic Processes*

- **Where δ is known as the drift parameter. The name drift comes from the fact that if we write the preceding equation** $(Y_t - Y_{t-1}) = \Delta Y_t = \delta + u_t$ **.**
- **it shows that Yt drifts upward or downward, depending on δ being positive or negative. It is also an AR(1) model.**
- **Following the procedure discussed for random walk without drift, it can be shown that for the random walk with drift model :**

$$E(Y_t) = Y_0 + t\delta$$

$$Var(Y_t) = t\sigma^2$$

3.2 *Stationary and non-stationary Stochastic Processes*

☞**As you can see for RWM with drift, the *mean* as well as the variance increases over time.**

☞**Thus, it violating the conditions of (weak) stationary. In short, RWM, with or without drift, is a nonstationary stochastic process.**

☞**The random walk model is an example of what is known in the literature as a unit root process.**

$$Y_t = \rho Y_{t-1} + u_t ; \text{-}1 \le \rho \le 1$$

Unit Root Stochastic Process

☞**Let us write the RWM as:**

☞**This model resembles the Markov first-order autoregressive model that we discussed on autocorrelation.**

3.2 *Stationary and non-stationary Stochastic Processes*

☞ **If ρ=1, becomes a RWM (without drift). If ρ is in fact 1, we face what is known as the unit root problem, that is, a situation of nonstationary; we already know that in this case the variance of Yt is not stationary.**

☞ ***The name unit root is due to the fact that ρ=1. Thus the terms nonstationary, random walk, and unit root can be treated as synonymous.* If, however, |ρ|<1, that is if the absolute value of ρ is less than one, then it can be shown that the time series Yt is stationary in the sense we have defined it.**

3.3 Trend Stationary and Difference Stationary Stochastic Processes

If the trend in a time series is completely predictable and not variable, we call it a deterministic trend, whereas if it is not predictable, we call it a stochastic trend.

- ☞ **To make the definition more formal, consider the following model of the time series Y_t.**
- ☞ $Y_t = \beta_0 + \beta_1 t + \beta_2 Y_{t-1} + u_t$ **-------(a)**
- ☞ **Where ut is a white noise error term and where *t* is time measured chronologically.**
- ☞ **Now we have the following possibilities:**

RWM without drift:

Pure random walk: If in (a), $\beta_0 = 0, \beta_1 = 0, \beta_2 = 1$, we get **--------(b)=non stationary**

$$Y_t = Y_{t-1} + U_t$$

$\Delta Y_t = Y_t - Y_{t-1} = U_t$ **………………………….. --------(c)=stationary, Hence, a RWM without drift is a difference stationary process (DSP).**

3.3 Trend Stationary and Difference Stationary Stochastic Processes

RWM with drift: Pure random walk with drift : If in (a), $\beta_0 \neq 0, \beta_1 = 0, \beta_2 = 1$, we get

$Y_t = \beta_0 + Y_{t-1} + U_t$ **-----------(d)-non stationary**

$Y_t - Y_{t-1} = \Delta Y_t = \beta_0 + U_t$**------------(e)—stationary, this means Yt will exhibit a positive (β1>0) or negative (β1<0) trend. Such a trend is called a stochastic trend. Equation (e) is a DSP process because the nonstationarity in *Yt can be eliminated by taking first differences of the time series*.**

Deterministic trend: Pure random walk with drift: If in (a), $\beta_0 \neq 0, \beta_1 \neq 0, \beta_2 = 1$, we get

$Yt = \beta_0 + \beta_1 t + u_t$ **, which is called a trend stationary process (TSP).**

3.3 Trend Stationary and Difference Stationary Stochastic Processes

- ☞**Although the mean of Yt is β0+β1t, which is not constant, its variance (=σ2) is constant.**
- ☞**Once the values of β0 and β1 are known, the mean can be forecasted perfectly. Therefore, if we subtract the mean of Yt from Yt, the resulting series will be stationary, hence the name trend stationary.**
- ☞**This procedure of removing the (deterministic) trend is called *detrending*.**

3.3 Trend Stationary and Difference Stationary Stochastic Processes

Random walk with drift & deterministic trend: If in (a) $\beta_0 \neq 0, \beta_1 \neq 0, \beta_2 = 1$, we get

$Y_t = \beta_0 + \beta_1 t + Y_{t-1} + U_t$ **-----(f) non stationary**

Deterministic trend with stationary AR (1) component:

If in (a) $\beta_0 \neq 0, \beta_1 \neq 0, \beta_2 < 1$, we get

$Y_t = \beta_0 + \beta_1 t + \beta_2 Y_{t-1} + U_t$ **, which is stationary around the deterministic trend.**

3.4 Integrated Stochastic Process

- ☞~~The random walk model is a specific case of~~ a more general class of stochastic processes known as **integrated processes.**
- ☞the RWM without drift is nonstationary, but its **first difference is stationary.**
- ☞the RWM without drift integrated of order 1, denoted as I(1).
- ☞Similarly, if a time series has to be differenced twice (i.e., take the first difference of the first differences) to make it stationary, we call such a **time series integrated of order 2.**

3.4 Integrated Stochastic Process

- ☞ In general, if a (nonstationary) time series has to be differenced d times to make it stationary, that time series is said to be integrated of order d. A time series Yt integrated of order d is denoted as Yt ~I(d).
- ☞ If a time series Yt is stationary to begin with (i.e., it does not require any differencing), it is said to be integrated of order zero, denoted by Yt ~I(0).
- ☞ Thus, we will use the terms “stationary time series” and “time series integrated of order zero” to mean the same thing

3.4 Integrated Stochastic Process

☞Most economic time series are generally I(1); that is, they generally become stationary only after taking their first differences.

Properties of Integrated Series

☞Let Xt, Yt and Zt be three time series

i. If Xt ~I(0) and Yt ~I(1),then Zt =(Xt +Yt)=I(1); that is, a linear combination or sum of stationary and nonstationary time series is nonstationary.

ii. If Xt ~I(d), then Zt =(a+bXt)=I(d), where a and b are constants. That is, a linear combination of an I(d) series is also I(d). Thus, if Xt ~I(0), then Zt =(a+bXt)~I(0)

3.4 Integrated Stochastic Process

iii. If Xt ~I(d1) and Yt ~I(d2), then Zt =(aXt +bYt)~I(d2), where d1<d2.

iv. If Xt ~I(d) and Yt ~I(d), then Zt =(aXt +bYt)~I(d*); d* is generally equal to d, but in some cases d*<d.

3.5 Tests of Stationarity: The Unit Root Test

- A test of stationarity (or nonstationarity) that has become widely popular over the past several years is the **unit root test**.
- $Yt = \rho Yt - 1 + ut; -1 \leq \rho \leq 1$------------------(i)
- where ut is a white noise error term
- We know that if ρ =1, that is, in the case of the unit root, (i) becomes a random walk model without drift, which we know is a nonstationary stochastic process.

3.5 Tests of Stationarity: The Unit Root Test

- Therefore, why not simply regress **Yt** on its (one period) lagged value **Yt −1** and find out if the estimated ρ is statistically equal to 1?
- If it is, then **Yt** is nonstationary. This is the general idea behind the unit root test of stationarit y.
- For theoretical reasons, we manipulate (i) as follows: Subtract **Yt −1** from both sides of (i) to obtain: $Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + u_t$

  $= (\rho - 1)Y_{t-1} + u$, which can be written alternatively $\Delta Y_t = \delta Y_{t-1} + u_t$ ----------------------------------(ii)

3.5 Tests of Stationarity: The Unit Root Test

- **Where δ = (ρ−1) and , as usual, is the first-difference operator.**
- **In practice, therefore, instead of estimating (i), we estimate (ii) and test:**
- **the (null) hypothesis that δ = 0, then ρ = 1, that is we have a unit root (nonstationary).**
    - **the t value of the estimated coefficient of Yt −1 does not follow the t distribution even in large samples; that is, it does not have an asymptotic normal distribution. Thus, we use τ(tau) statistic.**

3.5 Tests of Stationarity: The Unit Root Test

1. Dickey–Fuller (DF) test

☞ **Dickey and Fuller have shown that under the null hypothesis that δ = 0, the estimated t value of the coefficient of Yt −1 follows the τ(tau) statistic.**

☞ **In the literature *the tau statistic or test is known as the Dickey–Fuller (DF) test***

☞ **In conducting the DF test, it was assumed that the error term ut was uncorrelated.**

☞ **the DF test is estimated in three different forms, that is, under three different null hypotheses.**

3.5 Tests of Stationarity: The Unit Root Test

Yt is a random walk : $\Delta Y_{t\text{-}1} + u_t$

Yt is a random walk with drift : $\Delta Y_t = \beta_0 + \delta Y_{t\text{-}1} + u_t$

Yt is a random walk with drift :

around a stochastic trend : $\Delta Y_t = \beta 0 + \beta_1 t + \delta Y_{t-1} + u_t$

- **Where t is the time or trend variable.**
  - **Null hypothesis, δ = 0; that is, there is a unit root—the time series is nonstationary.**
  - **Alternative hypothesis, δ is less than zero; that is, the time series is stationary.**
- **Decision rule: If /tau statistics/> /tau critical value/; we rejected the null hypothesis, it means that Yt is a stationary time series.**

3.5 Tests of Stationarity: The Unit Root Test

2. Augmented Dickey–Fuller (ADF) Test

- In this case the **u_t are correlated**, Dickey and Fuller have developed a test, known as the *Augmented Dickey–Fuller* (ADF) test.
- This test is conducted by “augmenting” the preceding three equations by adding the lagged values of the dependent variable ΔYt -i.
- The ADF test here consists of estimating the following regression:

$$\Delta Y_t = \beta_0 + \beta_1 t + \delta Y_{t-1} + \sum \alpha_t \Delta Y_{t-1} + u_t$$

3.5 Tests of Stationarity: The Unit Root Test

☞ **In ADF we still test whether δ = 0 and the ADF test follows the same asymptotic distribution as the DF statistic, so the same critical values can be used.**

☞ **Hypothesis we use under this test is:**

$H_0 : \delta = 0$; there is a unit root, i.e., the time series is non stationary

$H_1 : \delta \neq 0$; there is no unit root, i.e., the time series is stationary (level stationary)

☞ **Example: the GDP series using one lagged difference of natural log of GDP of Ethiopia; the results were as follows:**

3.5 Tests of Stationarity: The Unit Root Test

$$\Delta Y_t = 0.0145 + 0.0001548t + 0.0125Y_{t-1} + 0.00458\Delta Y_{t-1}$$

$$t(=\tau) \quad (-0.38) \qquad (0.77) \qquad (0.34) \qquad (0.25)$$

$$(1\%\ CV = -4.242;\ 5\%\ CV = -3.540 \qquad)$$

☞ **Decision rule:** **The t(=τ) value of the Yt −1 coefficient (=δ) is 0.34, but this value in absolute terms is much l ess than even the 1% and 5% critical τ value of −4.242 and -3.540 respectively, again suggesting that even after taking care of possible autocorrelation in the error term, the Y series is not stationary.**

3. *The Phillips–Perron (PP) Unit Root Tests*

☞ **An important assumption of the DF test is that the error terms *ut* are independently and identically distributed. The ADF test adjusts the DF test to take care of possible serial correlation in the error terms by adding the lagged difference terms of the regressand.**

3.5 Tests of Stationarity: The Unit Root Test

- **Phillips and Perron use *nonparametric statistical methods* to take care of the serial correlation in the error terms without adding lagged difference terms.**
- **The Phillips-Perron test involves fitting the following regression:** $Y_t = \beta_0 + \beta_1 t + \rho Y_{t-1} + u_t$
- **Under the null hypothesis that ρ = 0, the PP Z(t) and Z(ρ) statistics have the same asymptotic distributions as the ADF t-statistic and normalized bias statistics.**
- **One advantage of the PP tests over the ADF tests is that the PP tests are robust to general forms of heteroscedasticity in the error term ut. Another advantage is that the user does not have to specify a lag length for the test regression.**

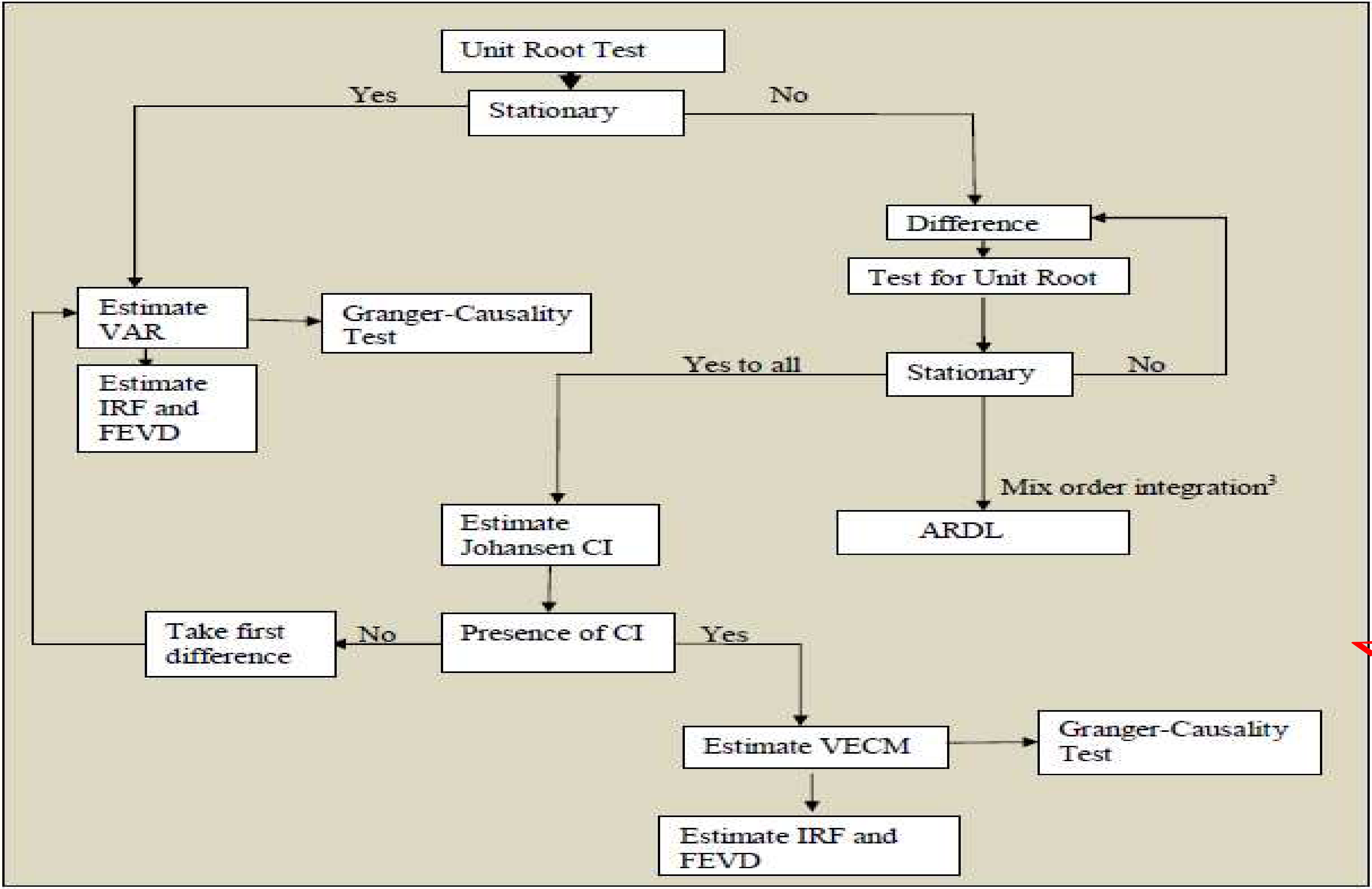
3.5 Tests of Stationarity: The Unit Root Test

☞**~~In some situation, lack of power in both the~~ ADF and PPtests is widely acknowledged,**

☞**Usually ADF yields superior results than PP test, if the data set has no missing observations and structural breaks whilst PP test also yields superior results than ADF test, if the dataset have some missing observations and have structural breaks**

Decision rule: since tests statistic, Z(t) value is greater that critical values we reject the null hypotheses of non stationary.

Next assignment



Unit Root Test

Stationary

Yes

No

Difference

Test for Unit Root

Stationary

Yes to all

No

Mix order integration[3]

ARDL

Estimate VAR

Granger-Causality Test

Estimate IRF and FEVD

Estimate Johansen CI

Presence of CI

No

Take first difference

Yes

Estimate VECM

Granger-Causality Test

Estimate IRF and FEVD

Next assignment

- ☞The series steps we should followed to do with time series analysis:
- ☞Unit root→ stationary→ if all are stationary at a level→ optimum leg length→ we run directly VAR model.
- ☞Unit root→ stationary→ if all stationary at 1st difference → optimum length→ Johanson co integration→ VECM→IRF& VDF→ Granger causality→

Lab Session

Use "*lecture_3.xls*" data to practice what we learnt in previous sections

END OF CHAPTER THREE

THANK YOU VERY MUCH FOR BEING WITH ME

BEING@COMMITTED!

STAY SAFE!

CHAPTER FOUR

INTRODUCTION TO SIMULTANEOUS EQUATION MODELS

4.1 Nature of Simultaneous Equation models

4.2 Simultaneity Bias & Inconsistency of OLS estimators

4.3 Solution to Simultaneous Equations

4.4 Identification problem

4.5 **Formal Rules (Conditions) for Identification**

4.6 **Estimation of Simultaneous Equations Models183**

4.1 Nature of Simultaneous Equation models

- **So far we have been discussed by focusing exclusively on the *problems* and *estimations* of a single equation regression models. In such models, a dependent variable is expressed as a linear function of one or more explanatory variables.**
- **i.e, there was a single dependent variable Y and one or more explanatory variables, X’s.**
- **The *cause-and-effect* relationship in single equation models between the dependent and independent variable is *unidirectional*.**
    - **That is, the explanatory variables are the *cause* and the independent variable is the *effect.***
- **But there are situations where such one-way or unidirectional causation in the function is not meaningful.**

4.1 Nature of Simultaneous Equation models

- This occurs if, for instance, Y (dependent variable) is not only function of X’s (explanatory variables) but also all or some of the X’s are, in turn, determined by Y.
- There is, therefore, a two-way flow of influence between Y and (some of) the X’s which in turn makes the distinction between dependent and independent variables a little doubtful.
- In simultaneous model there is more than one equation –one for each of the mutually, or jointly, dependent or endogenous variables.
- The number of equations in such models is equal to the number of jointly dependent or endogenous variables involved in the phenomenon under analysis.

4.1 Nature of Simultaneous Equation models

- Unlike the single equation models, in simultaneous equation models it is not usually possible (possible only under specific assumptions) to estimate a single equation of the model without taking into account the information provided by other equation of the system.

- If one applies OLS to estimate the parameters of each equation disregarding other equations of the model, the estimates so obtained are *not only biased but also inconsistent,* i.e. even if the sample size increases indefinitely, the estimators do not converge to their true values.

4.1 Nature of Simultaneous Equation models

- ☞ Example: the classic example of simultaneous causality in economics is supply and demand.
- ☞ Both Prices and quantities adjust until supply and demand are in equilibrium.
- ☞ A shock of demand or supply cause both prices and quantities to move.
- ☞ As well known, the prices P of a commodity and quantity Q sold are determined by the intersection of the demand and supply curves for that commodity.
- ☞ Look at the graph of dd and ss from class discussion(???)

4.2 Model specification

- Thus, assuming for simplicity that the demand and supply curves are linear and adding the stochastic disturbance term U1 and U2, we may write the empirical dd and ss function as:

$$\textit{Demand function}:$$

-
- $$Q_t^{\,d} = \beta_0 + \beta_1 P_t + \beta_2 Y_t + U_{1t} --- \beta_1 < 0$$

$$\textit{Supply function}:$$

$$Q_t^{\,s} = \beta_0 + \beta_1 P_t + \beta_2 Y_t + U_{2t} --- \beta_1 > 0$$

- **Equilibrium condition:** $Q_t^{\,s} = Q_t^{\,d}$
- **Where Qtd= quantity demand**
- **Qts=quantity supplied**
- **t=time;** β_0, *and* β_1 *are the parameters*

4.2 Model specification

- **B/C of simultaneous dependence between Q and P, then U1t and Pt, and U2t and pt cannot be independent.**
- **If u1t in above equation changes b/c changes in other variables affecting Qtd such as income, wealth and tastes), the demand curve will shift upward if u1t is +ve and downward if u1t is –ve.**
- **Thus, shift in demand curve changes both P and Q.**
- **Similarly, a change in U2t b/c of weather, import or export restrictions, etc; will shift the ss curves, again affect both P and Q.**
- **B/c of this simultaneous dependence b/c Q and P, u1t and Pt and u2t and pt cannot be independent. Thus, a regression of Q and P as in above equation would violate an important assumptions of the classical linear regression model; namely the assumption of no correlation b/n the explanatory variable(s) and the disturbance term.**

,

- ☞ In simultaneous equation models variables are classified as endogenous and exogenous.
- ☞ ~~Endogenous variables: are variables that~~ are determined by the economic model (within the system) and
- ☞ Exogenous variables: are those determined from outside of the system.
- ☞ Exogenous variables are also called *predetermined*. Since the exogenous variables are predetermined, they are supposed to be independent of the error terms in the model/ non stochastic.
  - ☞ are exogenous variables, lagged exogenous variables and lagged endogenous variables. Predetermined variables are non-stochastic and hence independent of the disturbance terms. ,X_t , X_{t-1} and $Y_{t-1}$are regarded as predetermined (exogenous) variables.

- **Structural models: A structural model describes the complete structure of the relationships among the economic variables.**

- **Structural equations of the model =** ***endogenous variables*+ *exogenous variables* + *disturbances (random variables)*.**

- **The parameters of structural model express the *direct effect* of each explanatory variable on the dependent variable.**

- **The Variables not appearing in any function explicitly may have an *indirect effect* and is taken into account by the simultaneous solution of the system.**

- **Reduced form of the model: The reduced form of a structural model is the model in which the endogenous variables are expressed a function of the predetermined variables and the error term only.**
- **Example: We may write the empirical demand-and-supply functions as**

Demand function: $Qd = \beta_0 + \beta_1 P + \beta_2 Y + u_1$

Supply function: $Qs = \alpha_0 + \alpha_1 P + \alpha_2 F + u_1$

Equilibrium condition: $Qd = Qs$

Where, Qd=quantity demanded; Qs = quantity supplied; Y=income; P=price; F= fertilizer, U_1 & U_2 are error terms. P and Q are endogenous variables and Y and F exogenous variables.

4.3Definitions of Some Concepts

☞ **Example: The following simple Keynesian model of income determination can be considered as a structural model.**

☞ $C = \alpha + \beta Y + U$ **----(1)**

☞ $Y = C + Z$ **----(2)**

for $\alpha > 0$ and $0 < \beta < 1$

☞ **where: C=consumption expenditure; Z=non-consumption expenditure ; Y=national income; C and Y are endogenous variables while Z is exogenous variable.**

☞ **Find the reduced form of the above structural model. Since C and Y are endogenous variables and only Z is the exogenous variables, we have to express C and Y in terms of Z.**

4.3Definitions of Some Concepts

☞To do this substitute Y=C+Z into equation (1).

$$C = \alpha + \beta(C + Z) \; + U$$

$$C = \alpha + \beta C + \beta Z + U$$

$$C - \beta C = \alpha + \beta Z + U$$

$$C(1 - \beta) = \alpha + \beta Z + U$$

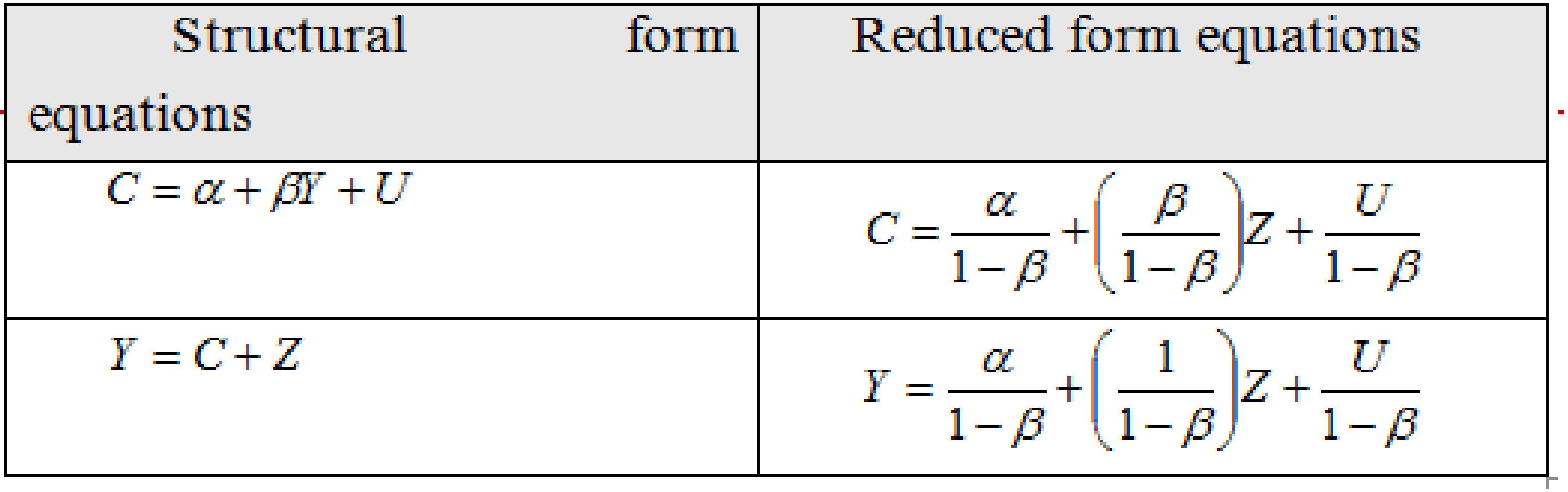
☞ $$C = \frac{\alpha}{1-\beta} + \left(\frac{\beta}{1-\beta}\right)Z + \frac{U}{1-\beta} \quad \text{------(3)}$$

☞Substituting again (3) into (2) we get;

☞ $$Y = \frac{\alpha}{1-\beta} + \left(\frac{1}{1-\beta}\right)Z + \frac{U}{1-\beta} \quad \text{-----------(4)}$$

☞Equation (3) and (4) are called the reduced form of the structural model of the above. We can write this more formally as:

4.3 Definitions of Some Concepts



Structural form equations	Reduced form equations
$C=\alpha+\beta Y+U$	$C=\frac{\alpha}{1-\beta}+\left(\frac{\beta}{1-\beta}\right)Z+\frac{U}{1-\beta}$
$Y=C+Z$	$Y=\frac{\alpha}{1-\beta}+\left(\frac{1}{1-\beta}\right)Z+\frac{U}{1-\beta}$

☞ **Parameters of the reduced form measure the *total effect (direct and indirect)* of a change in exogenous variables on the endogenous variable. For instance, in the above reduced form equation(1), $\left(\frac{\beta}{1-\beta}\right)$ measures the total effect of a unit change in the non-consumption expenditure on consumption. This total effect is β , the direct effect, times $\left(\frac{1}{1-\beta}\right)$,the indirect effect.**

4.4 Inconsistency and Simultaneity Bias of OLS Estimators

☞ Biasedness:

- The two-way causation in a relationship leads to violation of the important assumption of linear regression model, i.e. one variable can be dependent variable in one of the equation but becomes also explanatory variable in the other equations of the simultaneous-equation model.
- In this case E[X_iU_i] may be different from zero. To show simultaneity bias, let's consider the following simple simultaneous equation model.

$$\left.\begin{aligned} Y &= \alpha_0 + \alpha_1 X + U \\ X &= \beta_0 + \beta_1 Y + \beta_2 Z + V \end{aligned}\right\}$$

$$X = f(Y)$$

- $Y = f(X)$ this shows that the 2 way causation in a relationship leads to violations of the important assumptions linear regression model

- **Suppose that the following assumptions hold,**

$$\mathrm{E}(U) = 0\,, \qquad \mathrm{E}(V) = 0$$

$$\mathrm{E}(U^{2}) = \sigma_{u}^{2}\,, \qquad \mathrm{E}(V^{2}) = \sigma_{u}^{2}$$

$$\mathrm{E}(U_{i}U_{j}) = 0\,, \qquad \mathrm{E}(V_{i}V_{j}) = 0, \quad also \quad \mathrm{E}(UiVi) = 0;$$

- **where X and Y are endogenous variables and Z is an exogenous variable.**

- **The *reduced form* of X of the above model is obtained by substituting Y in the equation of X.**

$$X = \beta_0 + \beta_1(\alpha_0 + \alpha_1 X + U) + \beta_2 Z + V$$

$$X = \frac{\beta_0 + \alpha_0\beta_1}{1-\alpha_1\beta_1} + \left(\frac{\beta_2}{1-\alpha_1\beta_1}\right)Z + \left(\frac{\beta_1 U + V}{1-\alpha_1\beta_1}\right)$$

☞ **Applying OLS to the first equation of the above *structural model* will result in biased estimator because** $\text{cov}(X_iU_i) = \text{E}(X_iU_j) \neq 0$ **. Now, let's proof whether this expression.**

$$= \left(\frac{\beta_1}{1-\alpha_1\beta_1}\right)\text{E}(U^2) = \frac{\beta_1\sigma_u^2}{1-\alpha_1\beta_1} \neq 0$$

☞ **That is, covariance between X and U is not zero. As a consequence, if OLS is applied to each equation of the model separately the coefficients will turn out to be biased. Now, let's examine how the non-zero co-variance of the error term and the explanatory variable will lead to biasness in OLS estimates of the parameters.**

Estimators

- **Consistency Problems: An estimator is said to be consistent if its probability limit is equal to its population value.**
- **Inconsistent estimates**

$$p\lim(\hat{\beta}_1) = \beta_1 + \frac{{}^{\beta_2\sigma^2 U}\big/_{1-\beta_1\beta_2}}{\sigma u^2}$$

$$\hat{\beta}_1 = \beta_1 + \frac{\beta_2}{1-\beta_1\beta_2}\left(\frac{\sigma v^2}{\sigma X^2}\right)$$

4.5 Solution to the Simultaneous Equations

- **The obvious solution is to apply other methods of estimation w/c gives better estimates of parameters.**
- **1. the reduced form method or indirect least squares (ISLS)**
- **2. the method of instrumental variables**
- **3. two stage least squares (2SLS)**
- **4. limited information maximum likelihood (LIML)**
- **5. the mixed estimation**
- **6. Three stage least squares**
- **7. Full information maximum likelihood (FIML)**

4.5 Solution to the Simultaneous Equations

☞ **N.B: 1-5---we can applied to one equation at a time, and 6-7----the systems methods b/c they are applied to all equations of the system simultaneously.**

☞ ***How to estimate the reduced form parameters?***

- **The estimates of the reduced from coefficients (π's) may be obtained in two ways.**

☞**1. Direct estimation of the reduced coefficients by applying OLS.**

☞**2. Indirect estimation of the reduced form coefficients.**

4.6 Direct estimation of the reduced form coefficients

- ***Direct Method: Express the three endogenous variables(Ct , It , and Yt) as*** **functions of the two predetermined variables (Gt, andYt-1) directly using π's as the parameters of the reduced form model as follows.**

☞ **$Ct = \pi11Yt-1 + \pi12Gt + V1$**

☞ **$It , =\pi21Yt-1 + \pi22Gt +V2$**

☞ **$Yt =\pi31Yt-1 + \pi32Gt + V3$**

☞ **Note: π11 , π12 , π21 , π22 , π31 , and π32 are reduced from parameters.**

4.6 Direct estimation of the reduced form coefficients

- **The reduced form π ’s may be estimated by the method of least- squares –no restriction (LSNR).**
  - **This means we can apply OLS to reduced form equation because we express all the endogenous variables in terms of exogenous variables.**

4.6 Direct estimation of the reduced form coefficients

- **This method of obtaining the π 's is called least squares no restriction (LSNR) because it doesn't take into consideration any information on the structural parameters.**

- **In this method what required is knowledge of the predetermined variables appearing in the system not about the coefficients of structural questions.**

4.7 Indirect estimation of the reduced form coefficients

- **It is known that there is a relationship between the reduced form coefficients & the structural parameters (explained in the table).**
- **Therefore, to obtain values of coefficients estimate the structural parameters by any appropriate econometric techniques and then substitutes these estimates in to the system of parameters relationships to obtain indirectly.**
- **This indirect method involved three steps.**

4.8 Indirect estimation of the reduced form coefficients

- **1st step: Solve the system of endogenous variables so that each equation contains only predetermined explanatory variables.**
- **2nd step: Obtain the estimates of the structural parameters by any appropriate econometric method.**
- **3rd step: Substitute the estimates of β's and γ's in to the system of parameters relations to find the estimates of the reduced form coefficients.**

☞ ***Advantage of indirect estimation of the reduced-form coefficients***

- **Though it is complicated, it has a very good importance.**

☞ **a) The derivation of parameters like, π's, β's & α's is *more efficient* because in this way we take in to**

4.9 Recursive models

- **A model is called recursive if its structural equations can be ordered in such a way that:**
    - **the first equation includes only the predetermined variables in the right hand side.**
    - **the second equation contains predetermined variables and the first endogenous variable (of the first equation) in the right hand side and so on.**
- **The special feature of recursive model is that its equations may be estimated, one at a time, by OLS without simultaneous equations bias.**

4.9 Recursive models

- **OLS is not applicable if there is interdependence between the explanatory variables and the error term.**
- **In the simultaneous equation models, the endogenous variables may depend on the error terms of the model.**
- **Hence, the OLS technique is not appropriate for estimation of an equation in a simulations equations model.**
- **However, in a special type of simultaneous equations model called *Recursive, Triangular or Causal model, the use of OLS procedure of* estimation is appropriate.**
- **Consider the following three equation system to understand the nature of such models:**

4.9 Recursive models

☞ **Note that:**

$$Y_1 = \alpha_{10} + \beta_{11}X_1 + \beta_{12}X_2 + U_1$$

$$Y_2 = \alpha_{20} + \alpha_{21}Y_1 + \beta_{21}X_1 + \beta_{22}X_2 + U_2$$

$$Y_3 = \alpha_{30} + \alpha_{31}Y_1 + \alpha_{32}Y_2 + \beta_{31}X_1 + \beta_{32}X_2 + U_3$$

- **In the above illustration, the X's and Y's are exogenous and endogenous variables respectively.**
- **The disturbance terms follow the following assumptions.**

$$E(U_1U_2) = E(U_1U_3) = E(U_2U_3) = 0$$

- **The above assumption is the most crucial assumption that defines the recursive model.**

4.9 Recursive models

- **If this does not hold, the above system is no longer recursive and OLS is also no longer valid.**
- **The first equation of the above system contains only the exogenous variables on the *right hand side.***
- **Since by assumption, the exogenous variable is independent of U1 , the first equation satisfies the critical assumption of the OLS procedure.**
- **Hence, *OLS* can be applied straight forwardly to this equation.**

4.9 Recursive models

- Let us build a hypothetical recursive model for an agricultural commodity, say wheat.

- The production of wheat =Y1; , may be assumed to depend on exogenous factors: X2 = climatic conditions; and X3=last season's price. The retail price =Y2 may be assumed to be the function of production level Y1= and exogenous factor X4= disposable income.
- Finally, the price obtained by the producer = Y3 can be expressed in terms of the retail price; Y2 and exogenous factor; Xj= the cost of marketing the producer.
- The relevant equations of the model may be described as under:

4.9 Recursive models

$$Y_1 = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + U_1$$

$$Y_2 = \alpha_4 + \beta_1 Y_1 + \alpha_5 X_4 + U_2$$

$$Y_3 = \alpha_6 + \beta_2 Y_2 + \alpha_7 X_5 + U_3$$

- **In the first equation, there are only exogenous variables and are assumed to be independent of U1.**
- **In the second equation, the causal relation between Y1and Y2 is in one direction.**
- **Also Y1 is independent of U2 and can be treated just like exogenous variable.**
- **Similarly since Y2 is independent of U3 , OLS can be applied to the third equation.**
- **Thus, we can rewrite the above equations as follows:**

4.9 Recursive models

$$Y_1-\alpha_1-\alpha_2X_2-\alpha_3X_3=U_1$$

$$-\beta_1Y_1+Y_2-\alpha_4-\alpha_5X_4=U_2$$

$$-\beta_2Y_2+Y_3-\alpha_6-\alpha_7X_5=U_3$$

- **We can again rewrite this in matrix form as follows:**

$$\underbrace{\begin{bmatrix}1&0&0\\-\beta_1&1&0\\0&-\beta_2&1\end{bmatrix}}_{\text{Coefficient matrix of endogenous variables}}\begin{bmatrix}Y_1\\Y_2\\Y_3\end{bmatrix}+\underbrace{\begin{bmatrix}-\alpha_1&-\alpha_2&-\alpha_3&0&0\\-\alpha_4&0&0&-\alpha_5&0\\-\alpha_6&0&0&0&-\alpha_7\end{bmatrix}}_{\text{coefficient matrix of exogenous variable}}\begin{bmatrix}X_1\\X_2\\X_3\\X_4\\X_5\end{bmatrix}=\begin{bmatrix}U_1\\U_2\\U_3\end{bmatrix}$$

- **The coefficient matrix of endogenous variables is thus a triangular one; hence recursive models are also called as triangular models.**

4.9 Problems of simultaneous equation models

- **Simultaneous equation models create three distinct problems. These are:**
  - ***Identification of each equation of the model***
  - ***Mathematical completeness of the model***
    - **any model is said to be (mathematically) complete only when it possesses as many independent equations as endogenous variables.**
    - **In other words if we happen to know values of disturbance terms, exogenous variables and structural parameters, then all the endogenous variables are uniquely determined.**
  - ***Statistical estimation of each equation of the model***

4.10 The identification problem

☞ **Of the three problems, we are going to discuss the first problem (the *identification problem) in the following section.***

☞ The identification problem

- **In simultaneous equation models, the Problem of identification is a problem of model formulation; it does not concern with the estimation of the model.**
- **The estimation of the model depends up on the empirical data and the form of the model.**
- **If the model is not in the proper statistical form, it may turn out that the parameters may not uniquely estimated even though adequate and relevant data are available.**
- **In a language of econometrics, *a model is said to be identified only when it is in unique statistical form to enable us to obtain unique estimates of its parameters from the sample data.***

4.10 Identification Problem

- **The identical concept concerns with whether the numerical estimates of structural equations can be obtained from the estimated reduced form coefficients.**
  - Look at a simple Keynesian model, to illustrate the problem of identification (look at “Introduction to Econometrics: theory and practice with Stata” by Tesfaye E,)
- **An identification may be either exactly (fully or just identified) or over identified or under identification.**

A. Under identification (SEP>REP)

- **It occurs when the parameters of structural equation is higher than reduced form parameters.**
- **If the coefficients of the structural equations are greater than the coefficients of the reduced form, then we can say that the equation is under identified.**

Under identification (SEP>REP)

$$Qd = \alpha_0 + \alpha_1 P_1 + U_1 \qquad 10.32$$

$$Qd = \beta_0 + \beta_1 P_1 + U_2 \qquad 10.33$$

$$Qd = Qs \qquad 10.34$$

Where: Qd is quantity demand, Qs is quantity supplied and P is price

$$\alpha_0 + \alpha_1 P_1 + U_1 = \beta_0 + \beta_1 P_1 + U_2 \qquad 10.35$$

By rearranging equation (10.35), we obtain the following equations

$$\alpha_1 P_1 - \beta_1 P_1 = \beta_0 - \alpha_0 + U_2 - U_1$$

$$P_1(\alpha_1 - \beta_1) = \beta_0 - \alpha_0 + U_2 - U_1$$

$$P_1 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{U_2 - U_1}{\alpha_1 - \beta_1} \qquad 10.36$$

Let $\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}; \xi_0 = \frac{U_2 - U_1}{\alpha_1 - \beta_1}$

$$P_1 = \pi_0 + \xi_0 \qquad 10.37$$

Substitute equation 10.37 in to equation 10.32

$$Qd = \alpha_0 + \alpha_1\left[\left(\frac{\beta_0-\alpha_0}{\alpha_1-\beta_1}\right)+\left(\frac{U_2-U_1}{\alpha_1-\beta_1}\right)\right]+U_1$$

$$Qd = \frac{\alpha_0\alpha_1 -}{\alpha_1-\beta_1}\frac{\alpha_0\beta_1+\alpha_1\beta_0-\alpha_0\alpha_1}{\alpha_1-\beta_1}+\frac{\alpha_1U_2-\alpha_1U_1+\alpha_1U_1-\beta_1-U_1}{\alpha_1-\beta_1}$$

$$Qd = \frac{\alpha_0\alpha_1 -}{\alpha_1-\beta_1}\frac{\alpha_0\beta_1+\alpha_1\beta_0-\alpha_0\alpha_1}{\alpha_1-\beta_1}+\frac{\alpha_1U_2-\alpha_1U_1+\alpha_1U_1-\beta_1-U_1}{\alpha_1-\beta_1}$$

$$Qd = \frac{\alpha_1\beta_0-\alpha_0\beta_1}{\alpha_1-\beta_1}+\frac{\alpha_1U_2-\beta_1U_1}{\alpha_1-\beta_1} \qquad 10.38$$

Let $\pi_1 = \frac{\alpha_1\beta_0-\alpha_0\beta_1}{\alpha_1-\beta_1};\xi_1 = \frac{\alpha_1U_2-\beta_1U_1}{\alpha_1-\beta_1}$

Then we can write equation number 10.38 as follows

$$Qd = \pi_1 + \xi_1 \qquad 10.39$$

Under identification (SEP>REP)

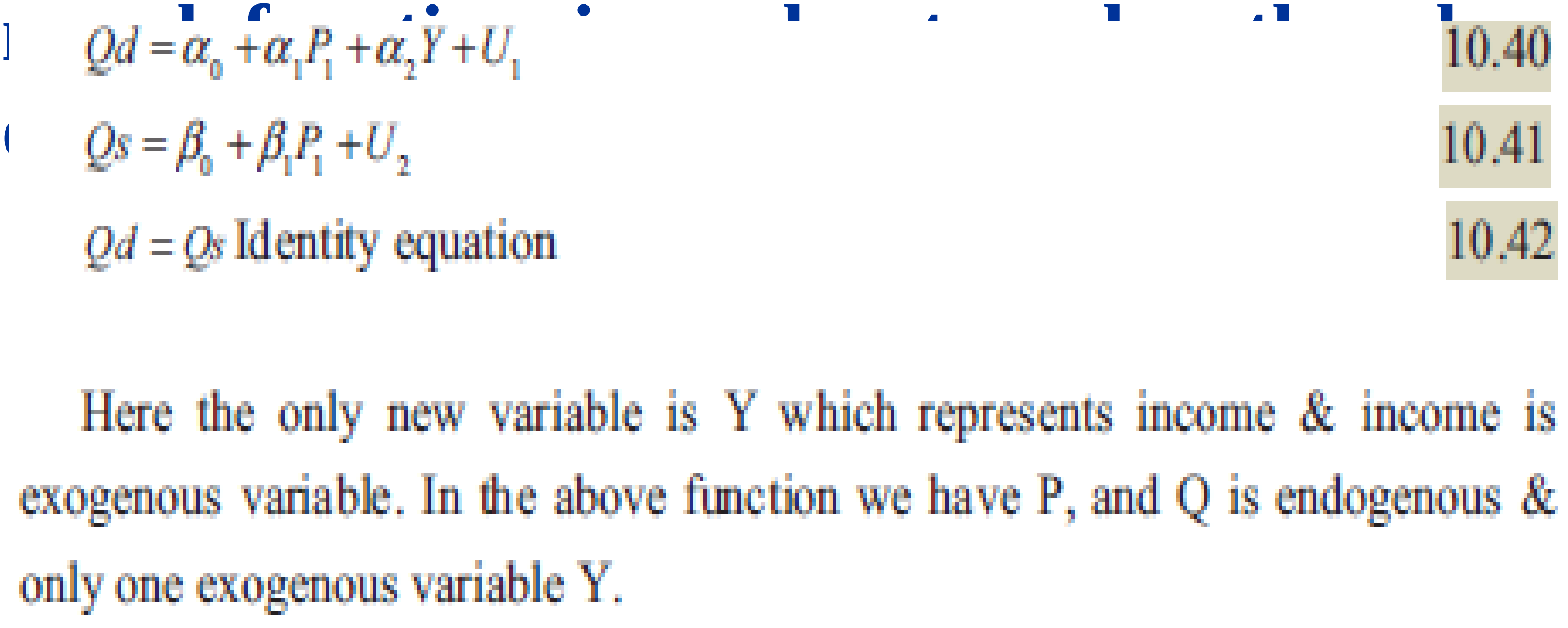
- Equation number 10.37 and 10.39 were the two reduced form equations derived from the structural equations number 10.32 & 10.33.
- Now if you compare the number of structural equation coefficients (α0, α1, β0 and β1) are four where as from the structural equations we have only two coefficients (π0 and π1).
- The coefficients of reduced form contain the coefficients of the structural equations i.e α0, α1, β1 and β2 are found in π0 and π1.
- But how we can find the values of α0, α1, β1 and β2 from π0 and π1. It is an ambiguous question??
- Since it is not possible to find these values from π0andπ1 or the coefficients of the structural equations are greater than the coefficients of the reduced form then we can say that the equation is under identified and we can not compute four structured coefficients from two reduced coefficients.

Why under identification is happened?

- The reason to have under identified function in the previous demand and supply function was that:
  - The same variables P and Q are appearing in both functions (only endogenous variables in both equation)
  - There is no additional information.

Exact /Just/ Identification (SEP=REP)

- **It occurs when structural coefficients are equal to reduced form coefficients.**
- **Now let’s incorporate additional variable in the de e pr**



$$Qd = \alpha_0 + \alpha_1 P_1 + \alpha_2 Y + U_1 \qquad 10.40$$

$$Qs = \beta_0 + \beta_1 P_1 + U_2 \qquad 10.41$$

$$Qd = Qs \text{ Identity equation} \qquad 10.42$$

Here the only new variable is Y which represents income & income is exogenous variable. In the above function we have P, and Q is endogenous & only one exogenous variable Y.

$$\alpha_0 + \alpha_1 P_1 + \alpha_2 Y + U_1 = \beta_0 + \beta_1 P_1 + U_2$$

$$\alpha_1 P_1 - \beta_1 P_1 = \beta_0 - \alpha_0 - \alpha_2 Y + U_2 - U_1$$

$$P_1(\alpha_1 - \beta_1) = \beta_0 - \alpha_0 - \alpha_2 Y + U_2 - U_1$$

$$P_1 = \frac{\beta_0 - \alpha_1}{\alpha_1 - \beta_1} - \frac{\alpha_2 Y}{\alpha_1 - \beta_1} + \frac{U_2 - U_1}{\alpha_1 - \beta_1} \qquad 10.43$$

Let $\pi_0 = \frac{\beta_0 - \alpha_1}{\alpha_1 - \beta_1}; \pi_1 = \frac{\alpha_2}{\alpha_1 - \beta_1}; \psi_0 = \frac{U_2 - U_1}{\alpha_1 - \beta_1}$

$$P_1 = \pi_0 + \pi_1 Y + \psi_0 \qquad 10.44$$

Substituting 10.44 in to equation 10.40.

$$Qd = \alpha_0 + \alpha_1 \left[\frac{\beta_0 - \alpha_1}{\alpha_1 - \beta_1} - \frac{\alpha_2 Y}{\alpha_1 - \beta_1} + \frac{U_2 - U_1}{\alpha_1 - \beta_1} \right] + \alpha_2 Y + U_1$$

$$Qd=\frac{\alpha_0\alpha_1-\alpha_0\beta_1}{\alpha_1-\beta_1}+\alpha_1\left[\frac{\beta_0-\alpha_1}{\alpha_1-\beta_1}-\frac{\alpha_2 Y}{\alpha_1-\beta_1}+\frac{U_2-U_1}{\alpha_1-\beta_1}\right]+\frac{\alpha_1\alpha_2 Y-\beta_1\alpha_2 Y+\alpha_1 U_1-\beta_1 U_1}{\alpha_1-\beta_1}$$

$$Qd=\frac{\alpha_1\beta_0}{\alpha_1-\beta_1}-\frac{\alpha_0\beta_1}{\alpha_1-\beta_1}-\frac{\beta_1\alpha_2 Y}{\alpha_1-\beta_1}+\frac{\alpha_1 U_2}{\alpha_1-\beta_1}-\frac{\beta_1 U_1}{\alpha_1-\beta_1}$$

Let $\pi_2=\frac{\alpha_1\beta_0}{\alpha_1-\beta_1}-\frac{\alpha_0\beta_1}{\alpha_1-\beta_1};\pi_3=\frac{-\beta_1\alpha_2 Y}{\alpha_1-\beta_1};\psi_1=\frac{\alpha_1 U_2}{\alpha_1-\beta_1}-\frac{\beta_1 U_1}{\alpha_1-\beta_1}$

$$Qd=\pi_2+\pi_3 Y+\psi_1 \qquad 10.45$$

Equation number 10.44 and 10.45 are reduced- form equations and OLS can be applied to estimate their parameters. In the structural equations (10.40 and 10.41) contains five structural coefficients α0,α1,α2,β1and ,β2. But there are four reduced form equations coefficients (π0,π1,π2and π3). Since the number of π' are less than (they are four) the structural coefficients (they are five α0,α1, β0,β1,α2, and β2) then we can not find unique solutions. But the supply function is independently identified because

$$Qs = \pi_2 + \pi_3 Y + \psi_1$$

In the supply equation of 10.41 there are two structural parameters (β0 and, β1) again in the reduced form equation of the supply equation we have two reduced form coefficients π2+π3 i.e why the supply function is identified. From equation 10.41 we have

$$Qs = \beta_0 + \beta_1 P_1$$

$$\beta_0 = Qs - \beta_1 P_1$$

Substitute equation 10.45 in place of Qs & equation number 10.44 in place of p and you will get (after simplification)

$$\beta_0 = \pi_2 - \beta_1 \pi_0 \qquad 10.46$$

Again from the same equation number 10.43 you can get the value of β1

$\beta_1 = \dfrac{Qs - \beta_0}{P_1}$ substitute in place of equation number 10.45 and in place of P1 equation number 10.34 and you will get after simplification.

$$\beta_1 = \frac{\pi_3}{\pi}$$

Exact /Just/ Identification (SEP=REP)

- **But in case of the demand function α0, α1, and α2, is 3 structural coefficients but in reduced form of equation the coefficients are two.**
- **Since in the demand function the coefficient of the reduced form (10.45) is less than the coefficients of the structural equation (10.40).**
- **We can concluded that the demand function is under identified (π2,π3) are less than α0,α1,and α2).**
- **But in case of supply function π2,π3 are equal to β0 , β1 then it is just identified.**
- **In conclusion, we can say that the supply function is identified but the demand function is not identified on the basis of this one can say that the system as a whole is not identified.**

Over identification (SEP<REP)

- **It occurs when the coefficients (parameters) of structural equation is less than the coefficients (parameters) of reduced forms.**
- **Let's modify the demand function by incorporating wealth (R) and supply function by incorporating the lagged price we will have the following equation.**

$$Qd = \alpha_0 + \alpha_1 P_1 + \alpha_2 Y + \alpha_3 R + U_1 \qquad 10.47$$

$$Qs = \beta_0 + \beta_1 P_1 + \beta_2 P_{t-1} + U_2 \qquad 10.48$$

Now we will have two endogenous variables (Q and P1) & three exogenous variables (Pt-1, Y and R).

At equilibrium, Qd=Qs

$$\alpha_0 + \alpha_1 P_1 + \alpha_2 Y + \alpha_3 R + U_1 = \beta_0 + \beta_1 P_1 + \beta_2 P_{t-1} + U_2$$

$$\alpha_1 P_1 - \beta_1 P_1 = \beta_0 - \alpha_0 + \beta_2 P_{t-1} - \alpha_2 Y - \alpha_3 R + U_2 - U_1$$

$$P_1(\alpha_1 - \beta_1) = \beta_0 - \alpha_0 + \beta_2 P_{t-1} - \alpha_2 Y - \alpha_3 R + U_2 - U_1$$

$$P_1 = \frac{\beta_0 - \alpha_0 + \beta_2 P_{t-1} - \alpha_2 Y - \alpha_3 R + U_2 - U_1}{\alpha_1 - \beta_1}$$

Let

$$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}; \pi_1 = \frac{-\alpha_2}{\alpha_1 - \beta_1}; \pi_2 = \frac{-\alpha_3}{\alpha_1 - \beta_1}; \pi_3 = \frac{\beta_2}{\alpha_1 - \beta_1}; \frac{\beta_0 - \alpha_0 + \beta_2 P_{t-1} Y - \alpha_3 R + U_2 - U_1}{\alpha_1 - \beta_1}$$

$$\varsigma_0 = \frac{U_2 - U_1}{\alpha_1 - \beta_1}$$

$$P_t = \pi_0 + \pi_1 Y + \pi_2 R + \pi_3 P_{t-1} + \varsigma_0 \qquad 10.49$$

Substitute Pt in the demand or supply function

$$Qd = \alpha_0 + \alpha_1(\pi 0 + \pi_1 Y + \pi_2 R + \pi_3 P_{t-1} + \varsigma_0) + \alpha_2 Y + \alpha_3 R + U_1$$

After simplification you will obtain

Over identification (SEP<REP)

- **From equation number 10.47 and 10.48 we have seven structural coefficients but in equation 10.49 and 10.45 we have eight reduced form coefficients.**
- **Since the coefficients of reduced form coefficients are greater than the reduced form coefficients we can say that the system as a whole is over identified.**
- **A function (an equation) belonging to a system of simultaneous equations is identified if it has a unique statistical form, i.e. if there is no other equation in the system, or formed by algebraic manipulations of the other equations of the system, contains the same variables as the function(equation) in question.**

4.11 Formal Rules (Conditions) for Identification

- **Identification problems do not just arise only on two equation-models.**
- **Using the above procedure, we can check identification problems easily if we have two or three equations in a given simultaneous equation model.**
- **However, for ‘n’ equations simultaneous equation model, such a procedure is very cumbersome.**
- **In general, for any number of equations in a given simultaneous equation, we have *two conditions that need to be satisfied to say that the model is in general* identified or not.**
- **In the following section we will see the formal conditions for identification.**

Formal Rules (Conditions) for Identification

- Actually the term ‘identification’ was originally used to denote the possibility (or impossibility) of deducing the values of the parameters of the structural relations from a knowledge of the reduced form parameters.
- However, we think that the reduced form approach is conceptually confusing and computationally more difficult than the structural model approach, because it requires the derivation of the reduced from first and then examination of the values of the determinant formed form some of the reduced form coefficients.
- The reduced form equation is time consuming process.
- The structural form approach is simpler and more useful.
- Thus, the so called order and rank conditions of identification lighten the task by providing a systematic way.

Formal Rules (Conditions) for Identification

- There are two conditions which must be fulfilled for an equation to be identified. These are:
  - 1. the order condition for identification
  - 2. the rank condition for identification
- The identification of a system means the identification of each question.
- The parameters identification in any equations means there is unique value for each parameter in equations.
- Equation is under identified when its statistical form is not unique/ When one or more of its equation of the model are identified we can say that the system as a whole is under identified.

Formal Rules (Conditions) for Identification

- Equation identified: in this case a system is identified when all the equations are identified.
- In identified system we can have two options:
  - if an equation is under identified it is impossible to estimate all its parameters using any econometric techniques. However, if the equation is identified its coefficients (parameters) can be statistically estimated.
  - If the equation is exactly identified appropriate method for estimation is the method of Indirect Least Square (ILSM).
  - If the equation is over identified, ILS will not give unique estimates of the parameters b/c it will not yield unique estimates of structural parameters.
  - In this case we use various methods. These are:
    - 2SLS (Two Stages Least Squares) or
    - MLM(Maximum Likely hood methods)

A. The order condition for identification

- **This condition is based on a counting rule of the variables included and excluded from the particular equation.**
- **It is a necessary but not sufficient condition for the identification of an equation.**
- **The order condition may be stated as follows.**
  - ***For an equation to be identified the total number of variables (endogenous and exogenous) excluded from it must be equal to or greater than the number of endogenous variables in the model less one.***

☞**Let, G = total number of equations (= total number of endogenous variables)**

☞**K= number of total variables in the model (endogenous and predetermined)**

☞**M= number of variables, endogenous and exogenous, included in a particular equation/ in a specific equation.**

A. The order condition for identification

- **Then** $(K-M) \ge (G-1)$ **tion may be sy** $\begin{bmatrix} Excluded \\ \text{var}\,iable \end{bmatrix} \ge [\text{total number of equations} - 1]$

- **The guidelines is that:**
  - **If (K-M)$\underline{\ge}$ (G-1); the equation is identified.**
  - **If (K-M)= (G-1); the equation is just/exactly identified.**
  - **If (K-M)< (G-1); the equation is under identified.**
  - **If (K-M)>(G-1); the equation is over identified.**

A. The order condition for identification

☞**Example 1:**

$$Q_d = \alpha + \alpha_1 P_1 + \alpha_2 I + U_1 - - - - - (1)$$

$$Q_s = \beta_0 + \beta_1 P_1 + U_2 - - - - - - - - (2)$$

- **<u>Take the dd equation</u>**
- **G= total number of equations/ total number of endogenous variables=2**
- **K=total number of exogenous and endogenous variables in equation (1), i.e., in demand equation=3**
- **The solution is that: (K-M)__________(G-1)**
- **(3-3)_____(2-1)=0<1, we conclude that the demand equation is under identified.**

A. The order condition for identification

☞**Take the ss equation**

- **Given: G=2; K=3; M=2;**
- **Solution:**
- **K-M-------------G-1**
- **(3-2)--------------(2-1)**
- **1=1→ from these we can conclude that the supply function is exactly identified.**

☞**Example 2: Given the structural model and determine whether the equation are identified or under identified.**

A. The order condition for identification

$$y_1 = 3y_2 - 2x_1 + x_2 + u_1 - - - - 1$$

$$y_2 = y_3 + x_3 + u_2 - - - - - - - - 2$$

$$y_3 = y_1 - y_2 - 2x_3 + u_3 - - - - - 3$$

- **Take equation (1);**
- **Given; M (endogenous and exogenous variables) in this specified equation is 4 (y1, y2, x1 and x2); K=6; G=3;**
- **(K-M)----------(G-1)**
- **6-4-------------(3-1)**
- **2=2--→ this equation is identified and it is exactly identified.**

A. The order condition for identification

- **Take equation (2);**
- **Given; M (endogenous and exogenous variables) in this specified equation is 3 (y2, y3, & x3); K=6; G=3;**
- **(K-M)-----------(G-1)**
- **6-3-------------(3-1)**
- **3>2--→ this equation is identified and it is over identified.**

A. The order condition for identification

- **Take equation (3);**
- **Given; M (endogenous and exogenous variables) in this specified equation is 4 (y3, y1, y2 and x3); K=6; G=3;**
- **(K-M)-----------(G-1)**
- **6-4--------------(3-1)**
- **2=2--→ this equation is identified and it is exactly identified.**

- **Example 3: if a system contains 10 equations with 15 variables, ten endogenous and five exogenous, an equation containing 11 variables is not identified, while another containing 5 variables is identified.**

A. The order condition for identification

- **For 1st equation we have:**

☞ **$G=10; K=15; M=11;$**

☞ **Order condition:**

☞ **$K-M \geq G-1$**

☞ **$15-11 \geq 10-1$**

☞ **$4<9$ → that is the order condition is not satisfied.**

- **For the 2nd equation we have:**
- **$G=10; K=15; M=5$**
- **Order condition:**
- **$(K-M) \geq (G-1)$; $10 \geq 9$-----the order conditions satisfied.**

B. The rank condition for identification

- **The rank condition states that: in a system of G equations any particular equation is identified if and only if it is possible to construct at least one nonzero determinant of order (G-1) from the coefficients of the variables excluded from that particular equation but contained in the other equations of the model.**
- **The practical steps for tracing the identifiablity of an equation of a structural model may be outlined as follows.**
- ***Firstly, write the parameters of all the equations of the model in a separate* table, noting that the parameter of a variable excluded from an equation is equal to zero.**

For example let a structural model be:

$$y_1 = 3y_2 - 2x_1 + x_2 + u_1$$

$$y_2 = y_3 + x_3 + u_2$$

$$y_3 = y_1 - y_2 - 2x_3 + u_3$$

- Where y’s are the endogenous variables and

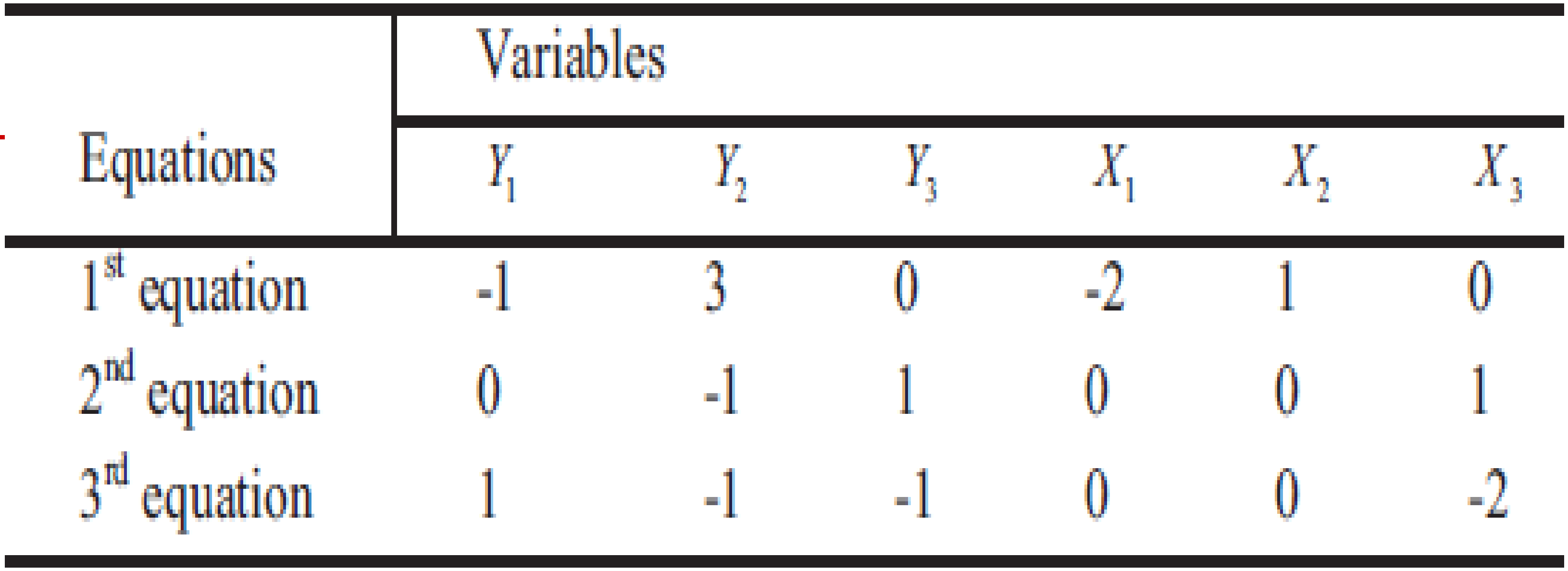
☞ x’s are the exogenous variables

$$-y_1 + 3y_2 + 0y_3 - 2x_1 + x_2 + 0x_3 + u_1 = 0$$

$$0y_1 - y_2 + y_3 + 0x_1 + 0x_2 + x_3 + u_2 = 0$$

$$y_1 - y_2 - y_3 + 0x_1 + 0x_2 - 2x_3 + u_3 = 0$$

- Ignoring the random disturbance the table of the parameters of the model is as follows:



Equations	Variables					
	Y_1	Y_2	Y_3	X_1	X_2	X_3
1st equation	-1	3	0	-2	1	0
2nd equation	0	-1	1	0	0	1
3rd equation	1	-1	-1	0	0	-2

☞ ***Secondly***, ~~***Strike out the row of coefficients of the equation which is being* examined for identification.**~~ **For example, if we want to examine the identifiability of the second equation of the model we strike out the second row of the table of coefficients.**

B. The rank condition for identification

- ***Thirdly*, *Strike out the columns in which a non-zero coefficient of the* equation being examined appears.**
- **Table of structural parameter**

Equations	**Y1**	**~~Y2~~**	**~~Y3~~**	**X1**	**X2**	**~~X3~~**
1st equ.	-1	~~3~~	~~0~~	-2	1	~~0~~
~~2nd equ.~~	~~0~~	~~1~~	~~1~~	~~0~~	~~0~~	~~1~~
3rd equ.	1	~~1~~	~~1~~	0	0	~~2~~

- **By deleting the relevant row and columns we are left with the coefficients of variables not included in the particular equation, but contained in the other equations of the model.**
- **For example, if we are examining for identification the second equation of the system, we will strike out the second, third and the sixth columns of the above table, thus obtaining the following tables.**

Table of structural parameters

–excluded variables

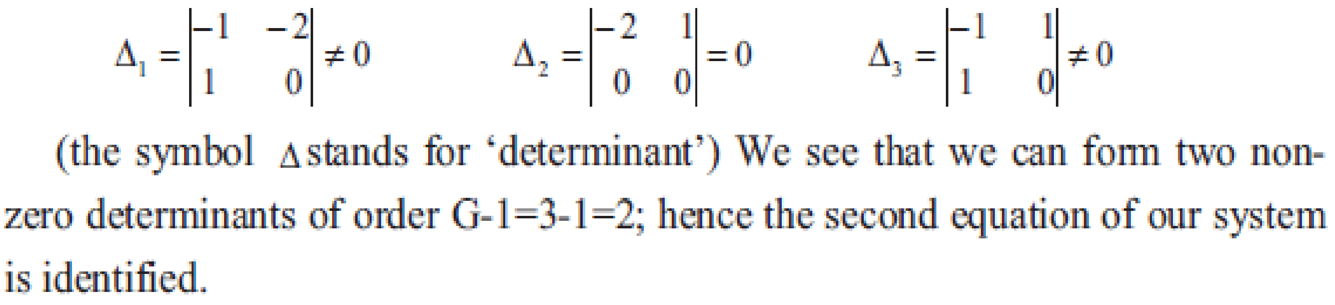
	Y_1	Y_2	Y_3	X_1	X_2	X_3
		↓	↓			↓
1st	-1	~~3~~	~~0~~	-2	1	~~0~~
→2nd	~~0~~	~~1~~	~~1~~	~~0~~	~~0~~	~~1~~
3rd	1	~~1~~	~~1~~	0	0	~~2~~

Table of parameters of

Y_3	X_1	X_2
-1	-2	1
1	0	0

B. The rank condition for identification

- ***Fourthly*, *form the determinant(s) of order (G-1) and examine their value.***
- ***Guide line:***
  - ***If* at least one of these determinants is non-zero, the equation is identified.**
  - **If all the determinants of order (G-1) are zero, the equation is *under identified.***
- **In the above example of exploration of the identifiability of the second structural equation we have three determinants of order (G-1)=3-1=2. They are:**



$$\Delta_1 = \begin{vmatrix} -1 & -2 \\ 1 & 0 \end{vmatrix} \neq 0 \qquad \Delta_2 = \begin{vmatrix} -2 & 1 \\ 0 & 0 \end{vmatrix} = 0 \qquad \Delta_3 = \begin{vmatrix} -1 & 1 \\ 1 & 0 \end{vmatrix} \neq 0$$

(the symbol Δ stands for ‘determinant’) We see that we can form two non-zero determinants of order G-1=3-1=2; hence the second equation of our system is identified.

Fifthly, To see whether the equation is *exactly identified or overidentified* we use the order condition $(K-M)\geq(G-1)$. With this criterion, if the equality sign is satisfied, that is if $(K-M)=(G-1)$, the equation is exactly identified. If the inequality sign holds, that is, if $(K-M)<(G-1)$, the equation is *overidentified*.

In the case of the second equation we have:

G=3 K=6 M=3

And the counting rule $(K-M)\geq(G-1)$ gives

(6-3)>(3-1)

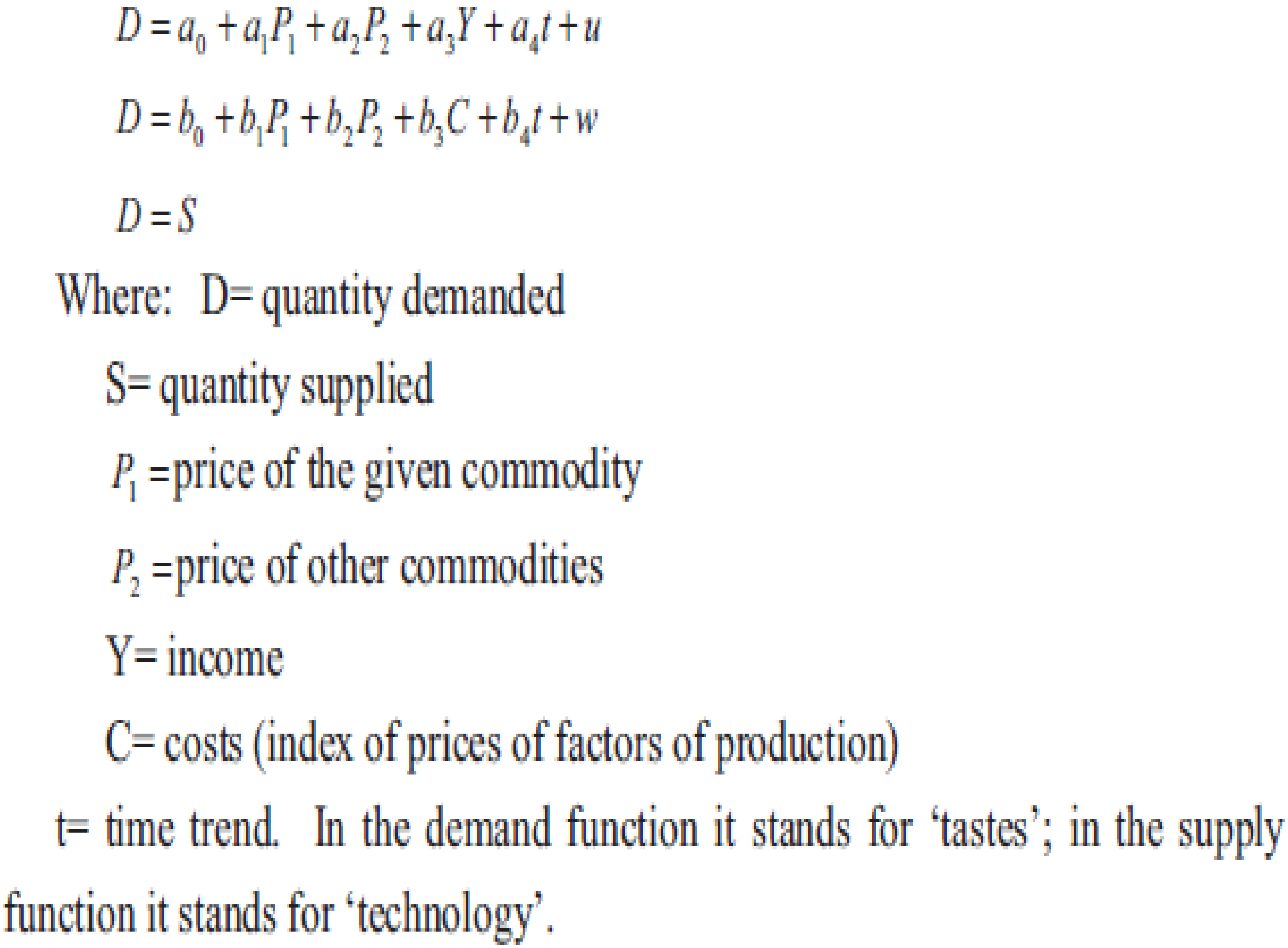
Therefore, the second equation of the model is *overidentified.*

B. The rank condition for identification

- The identification of a function is achieved by assuming that some variables of the model have zero coefficient in this equation, that is, we assume that some variables do not directly affect the dependent variable in this equation.
- This, however, is an assumption which can be tested with the sample data.
- We will examine some tests of identifying restrictions in a subsequent section.
- Some examples will illustrate the application of the two formal conditions for identification.

B. The rank condition for identification

☞**Example:**



$$D = a_0 + a_1P_1 + a_2P_2 + a_3Y + a_4t + u$$

$$D = b_0 + b_1P_1 + b_2P_2 + b_3C + b_4t + w$$

$$D = S$$

Where: D= quantity demanded

S= quantity supplied

P_1 =price of the given commodity

P_2 =price of other commodities

Y= income

C= costs (index of prices of factors of production)

t= time trend. In the demand function it stands for 'tastes'; in the supply function it stands for 'technology'.

Estimation of Simultaneous Equations Models

- To estimate the simultaneous equation models we adopt two approaches.
- The first one is single equation method, also known as limited information method.
  - In this single equation method we estimate each question in the system individually.
- The second one is system methods also known as full information methods.
  - In this case we estimate all equations in the model simultaneously.
- In practice system methods are not commonly used for variety of reasons rather, single equation methods are often used.
- The major single equation methods applied in the estimation of simultaneous equation methods are:
  - □ 1. Ordinary least squares (OLS)
  - □ 2. Indirect least squares (ILS)
  - □ 3. Two stage least squares (2SLS)

1. Ordinary Least Squares

- **We have seen that applying OLS on simultaneous equation produce bias & inconsistent parameters.**
- **But there is one situation OLS can be applied appropriately even in the context of simultaneous equation.**

$$Y_1 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + U_1 \tag{10.51}$$

$$Y_2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U_2 \tag{10.52}$$

$$Y_3 = \mu_0 + \mu_1 X_1 + \mu_2 X_2 + U_3 \tag{10.53}$$

1. Ordinary Least Squares

- **In equation 10.51 the endogenous variables appear in the left & the exogenous variables in the right hand side.**
- **Hence, OLS can apply straight forwardly to this question given all the assumptions of OLS holds true.**
- **In equation 10.52 we can apply OLS provided that Y1 & U2 are uncorrelated.**
- **Again we can apply OLS to the last equation if both Y1 & Y2 are uncorrelated with U3.**
- **In this recursive system OLS can be applied to each equation separately & we do not face a simultaneous equation problem.**
- **The reason for this is that clear, because there is no interdependence among the endogenous variables.**
- **Thus, Y1 affect Y2 influence Y3 without being influenced by Y3.**
- **In other words each equation exhibits a unilateral causal dependence.**

2. Indirect least square (ILS method)

- **ILS is applicable only for just/exact identified equations [(K-M) = (G-1)].**
- **The method of obtaining the estimates of the structural coefficients using OLS of the reduced form coefficients is known as the method of (ILS) indirect least squares & the estimates obtained are known as the indirect least squares estimates.**
- **Indirect Least Square method involves the following Steps**

2. Indirect least square (ILS method)

- ☞ **1st:We first obtain the reduced form equation from the structural equations. i.e. explaining the endogenous variables as a function of explanatory (exogenous variables) & a stochastic term.**
- ☞ **2nd: Apply OLS to the reduced- form equations individually. In this case the exogenous variables are uncorrelated with the stochastic term.**
- ☞ **3rd: we obtain estimates of the original structural coefficients from the estimated reduced-form coefficients obtained in step two. ILS derives from the fact that structural coefficients are obtained indirectly from the OLS estimates of the reduced form coefficients.**

3. Two-Stage Least Squares (2SLS) Method

- **The 2SLS procedure is generally applicable for estimation of over-identified equations as it provides unique estimators.**
- **Two-Stage Least Squares (2SLS) Method involves the following steps.** $\hat{Y}$

☞**1st: Estimate the reduced form equations by OLS and obtain the predicted .**

☞**2nd: Replace the** $\hat{Y}$ **right hand side endogenous variables in the structural equations by the corresponding** ***and estimate them by OLS.***

Lab Session

Use "*lecture_4.xls*" data to practice what we learnt in previous sections

END OF CHAPTER FOUR

Thank you very much for being with me for a while!

Stay Safe!

CHAPTER FIVE

INTRODUCTION TO PANAL DATA REGRESSION MODELS

5.1 Introduction

- **The types of data that are generally available for empirical analysis, namely, time series, cross section, and panel.**
- **In time series data we observe the values of one or more variables over a period of time (e.g., GDP for several quarters or years).**
- **In cross-section data, values of one or more variables are collected for several sample units, or entities, *at the same point in time* (e.g., crime rates for 9 regions in Ethiopia for a given year).**
- ***In panel data, the same cross-sectional unit (say a family or a firm or a state) is surveyed over time.***
- ***In short, panel data have space as well as time dimensions.***

5.1 Introduction

- ☞ A panel of data consists of *a group of cross-sectional units* (people, households, firms, states, countries) who are observed *over time*. We will often refer to such units as individuals, with the term ‘‘*individual*’’ being used generically, even when the unit of interest is not a person.

- ☞ Let us denote the number of cross-sectional units (individuals) by N, and number of time periods in which we observe them as T.

- ☞ Panel data comes in several different ‘‘flavors,’’ each of which introduces new challenges and opportunities.

5.1 Introduction

- **Peter Kennedy1; describes the different types of panel data sets as:**
- **““Long and narrow,”” with ““long”” describing the time dimension and ““narrow”” implying a relatively small number of cross sectional units.**
- **““Short and wide,”” indicating that there are many individuals observed over a relatively short period of time.**
- **““Long and wide,”” indicating that both N and T are relatively large.**
- **A ““long and narrow”” panel may consist of data on several firms over a period of time.**

5.1 Introduction

- **Data on 221 State of Oromia high schools in 2014 and again in 2015, for 442 observations total; Data on 9 states of Ethiopia, each state is observed in 3 years, for a total of 27 observations;**
- **Data on 500 individuals, in five different months, for 2500 observations total.**

5.2 Other names of Panel Data

☞**Pooled data:- pooling of time series and cross-sectional observations.**

☞**Cross-sectional time-series data:- Combination of time series and cross-section data.**

☞**Micro-panel data, longitudinal data: A study over time of a variable or group of subjects.**

☞**Panel data (also called longitudinal data) refers to data for n different entities observed at T different time periods.**

5.3 Balanced and unbalanced data

- ☞ **When describing the cross sectional data it was useful to use a subscript to denote the entity; for instance, *Yi referred to be the variable Yi for the ith entity*.**
- ☞ **When describing panel data, we need some additional notations to keep track of both *the entity and the time period*.**
- ☞ **This is done by using two subscripts rather than one: The first, *i refres to the entity., and the second, t*, *refers to the time period of the observation.***
- ☞ **Thus, Yit denotes the variable Y observed for the ith of n entities in the *i th of T periods*.**

5.3 Balanced and unbalanced data

☞ **Some additional terminology associated with panel data describes weather some observations are missing.**

☞ ***A balanced has all its observations, that is, that variables are observed for each entity and each time period.***

☞ ***A panel that has some missing data for at least*** **one time period** ***for at least one entity is called an unbalanced panel..***

5.4 Why we use panel data?

- ☞Panel data give:
  - ☞ more informative data
  - ☞more variability
  - ☞ less collinearity among the variables
  - ☞ more degrees of freedom and more efficiency. Time-series studies are plagued with multi-collinearity.
- ☞Panel data are better able to:
  - ☞ identify and measure effects that are simply not detectable in pure cross-section or pure time-series data.
  - ☞study the dynamics of adjustment.
    - □ Cross-sectional distributions that look relatively stable hide a multitude of changes.

5.4 Why we use panel data?

☞**Panel data allows you to:**

- **Controlling for individual heterogeneity.**
  - ☞**Control for variables you cannot observe or measure like cultural factors or difference in business practices across companies; or variables that change over time but not across entities (i.e. national policies, federal regulations, international agreements, etc.)**
  - ☞**This is, it accounts for individual heterogeneity.**
  - ☞**Time-series and cross-section studies not controlling this heterogeneity run the risk of obtaining biased results.**

5.4 Why we use panel data?

☞ Panel data models allow us to construct and test more complicated behavioral models than purely cross-section or time-series data.

- ▫ For example, technical efficiency is better studied and modeled with panels.

☞ Micro panel data gathered on individuals, firms and households may be more accurately measured than similar variables measured at the macro level.

- ☞ Biases resulting from aggregation over firms or individuals may be *reduced or eliminated.*

5.6 Limitation of panel data

1. **Design and data collection problems: These include:**

- **problems of coverage (incomplete account of the population of interest)**
- **nonresponse (due to lack of cooperation of the respondent or because of interviewer error)**
- **recall (respondent not remembering correctly)**
- **frequency of interviewing**
- **interview spacing**
- **reference period**
- **the use of bounding and**
- **time-in-sample bias.**

5.6 Limitation of panel data

2. Distortions of measurement errors:

- Measurement errors may arise because of faulty responses due to unclear questions
- Memory errors
- Deliberate distortion of responses (e.g. prestige bias)
- Inappropriate informants
- Misrecording of responses and interviewer effects.

5.6 Limitation of panel data

☝ 3. Selectivity problems. These include:

(a) Self-selectivity: People choose not to work because the reservation wage is higher than the offered wage.

☞In this case we observe the characteristics of these individuals but not their wage.

- **Since only their wage is missing, the sample is censored.**
- **However, if we do not observe all data on these people this would be a truncated sample.**

5.7 Limitation of panel data

(b) **Nonresponse:** This can occur at the initial wave of the panel due to **refusal to participate**, **nobody at home**, **untraced sample unit**, and other reasons.

- ☞Item (or partial) nonresponse occurs when one or more questions are left unanswered or are found not to provide a useful response.

(c) **Attrition:** While nonresponse occurs also in cross-section studies, it is a more serious problem in panels because subsequent waves of the panel are still subject to nonresponse.

- ☞Respondents may **die, or move, or find that the cost of responding is high.**

5.7 Limitation of panel data

- **4. Short time-series dimension:**
  - **Typical micro panels involve annual data covering a short time span for each individual. This means that asymptotic arguments rely crucially on the number of individuals tending to infinity.**
- **Increasing the time span of the panel is not without cost either. In fact, this increases the chances of attrition and increases the computational difficulty for limited dependent variable panel data models.**
- **5. Cross-section dependence: Macro panels on countries or regions with long time series that do not account for cross-country dependence may lead to misleading inference.**

5.8 Notation (Model specification) for panel data

☞ **Panel data consist of observations on the same n entities at two or more time periods T.**

☞ **If the data set contains observations on the variables X and Y, then the data are denoted**

$$(X_{it}, Y_{it}), i = 1,\ldots, n \text{ and } t = 1,\ldots., T, \tag{11.1}$$

☞ **Where the first subscript, i, refers to the entity being observed, and the second subscript, t, refers to the date at which is observed.**

☞ **A *double subscript* is used to distinguish entities (states, family, country, individuals, etc.) and time periods.**

☞ **Consider the following simple panel data regression model:**

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \varepsilon_{it}, i =,\ldots, n, T = 1,\ldots, T \tag{11.2}$$

5.9 Estimation of Panel Data Regression

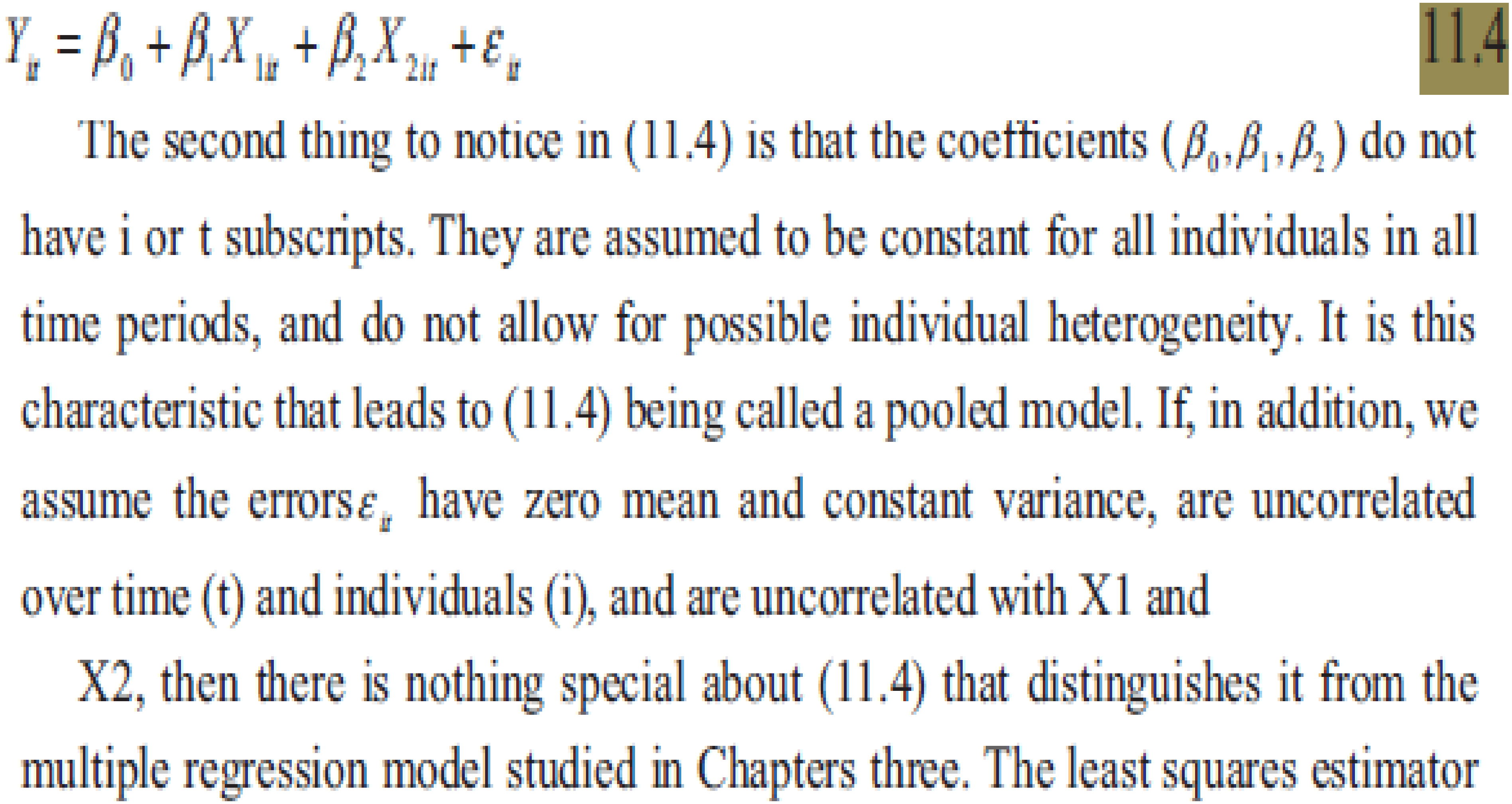
- Where i= entity (state); n=number of entities, so i=1,…,n; t= time period (year, month, quarter, and so on); T= number of time periods, so that t=1,…, T
- Panel data with k regressors:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it}, i =, \ldots, n, T = 1, \ldots, T \tag{11.3}$$

- We have three models to estimate panel data
- 1. Pooled model 2. Fixed Effect model 3. Random effect model
- A pooled model is one where the data on different individuals are simply pooled together with no provision for individual differences that might lead to different coefficients.

5.10 Pooled data

- **For an equation with two explanatory variables X1 and X2, a pooled model can be written as**



$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \varepsilon_{it} \qquad 11.4$$

The second thing to notice in (11.4) is that the coefficients ($\beta_0, \beta_1, \beta_2$) do not have i or t subscripts. They are assumed to be constant for all individuals in all time periods, and do not allow for possible individual heterogeneity. It is this characteristic that leads to (11.4) being called a pooled model. If, in addition, we assume the errors ε_{it} have zero mean and constant variance, are uncorrelated over time (t) and individuals (i), and are uncorrelated with X1 and

X2, then there is nothing special about (11.4) that distinguishes it from the multiple regression model studied in Chapters three. The least squares estimator

5.10 Pooled data

☞ for ($\beta 0$,$\beta 1$,$\beta 2$) has all its desirable properties.

☞ It is **consistent**, and the usual t and F statistics are valid in **large samples** for hypothesis testing and interval estimation.

☞ If we also assume X1 and X2 are nonrandom, the least squares estimator is the minimum variance linear unbiased estimator in finite samples.

☞ We will focus on **large sample properties**, however, because it is typically unrealistic to assume X1 and X2 are nonrandom, and our sample sizes are usually large.

☞ Hence, the main weakness of this model is it doesnot capture the hetrogenity among the entity.

5.11 The Fixed Effects (Entity/Time Fixed) Model

- In the previous section we saw that one way to recognize the existence of individual characteristics in a panel data model is to allow individual errors in different time periods to be correlated.

- A second way is to relax the assumption that all individuals have the same coefficients. Extending the model in (11.4) along these lines, we can write

$$Y_{it} = \beta_{0i} + \beta_{1i} X_{1it} + \beta_{2i} X_{2it} + \varepsilon_{it} \qquad 11.10$$

- An *i subscript has been added to each of the subscripts, implying that ($\beta_0, \beta_1, \beta_2$)* can be different for each individual. This model is a legitimate panel data model, but it is not suitable for panels that are short and wide.

5.11 The Fixed Effects (Entity/Time Fixed) Model

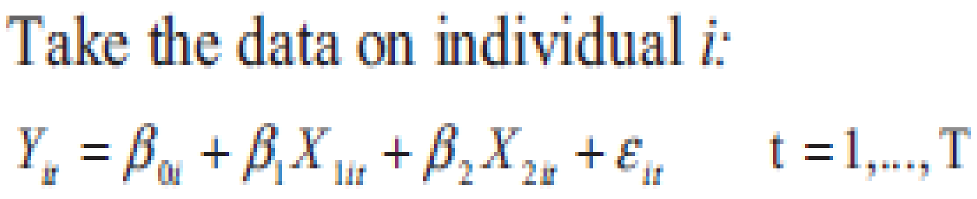
- ☞ **You may apply entity fixed effects regression when you want to control for omitted variables that differ among panels but are constant over time.**
- ☞ **On the other hand, if there are unobserved effects that vary across time rather than across panels, we apply time fixed effects regression model.**
- ☞ **Use fixed-effects (FE) whenever you are only interested in analyzing the impact of variables that vary over time.**
- ☞ **FE explore the relationship between predictor and outcome variables within an entity (country, person, company, etc.).**

5.11 The Fixed Effects (Entity/Time Fixed) Model

- ☞ Each entity has its own individual characteristics that may or may not influence the predictor variables (for example being a male or female could influence the opinion toward certain issue or the political system of a particular country could have some effect on trade or GDP or the business practices of a company may influence its stock price).
- ☞ When using FE we assume that something within the individual may impact or bias the predictor or outcome variables and we need to control for this.
  - ☞This is the rationale behind the assumption of the correlation between entity’s error term and predictor variables.
- ☞ FE removes the effect of those time-invariant characteristics from the predictor variables so we can assess the predictors’ net effect.

5.11 The Fixed Effects (Entity/Time Fixed) Models

- **Another important assumption of the FE model is that those time-invariant characteristics are unique to the individual and should not be correlated with other individual characteristics.**
  - **Each entity is different therefore the entity's error term and the constant (which captures individual characteristics) should not be correlated with the others.**
- **If the error terms are correlated then FE is no suitable since inferences may not be correct and you need to model that relationship (probably using random-effects.)**
- **Think of the following two variables panel regression model in fixed effect form:**



Take the data on individual *i*:

$$Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \varepsilon_{it} \qquad t = 1,\dots,T$$

5.11 The Fixed Effects (Entity/Time Fixed) Models

☞ **Average the data across time, by summing both sides of the equation and dividing by T.**

$$\frac{1}{T}\sum_{t=1}^{T}(Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \varepsilon_{it}) \qquad 11.14$$

Using the fact that the parameters do not change over time, we can simplify this as

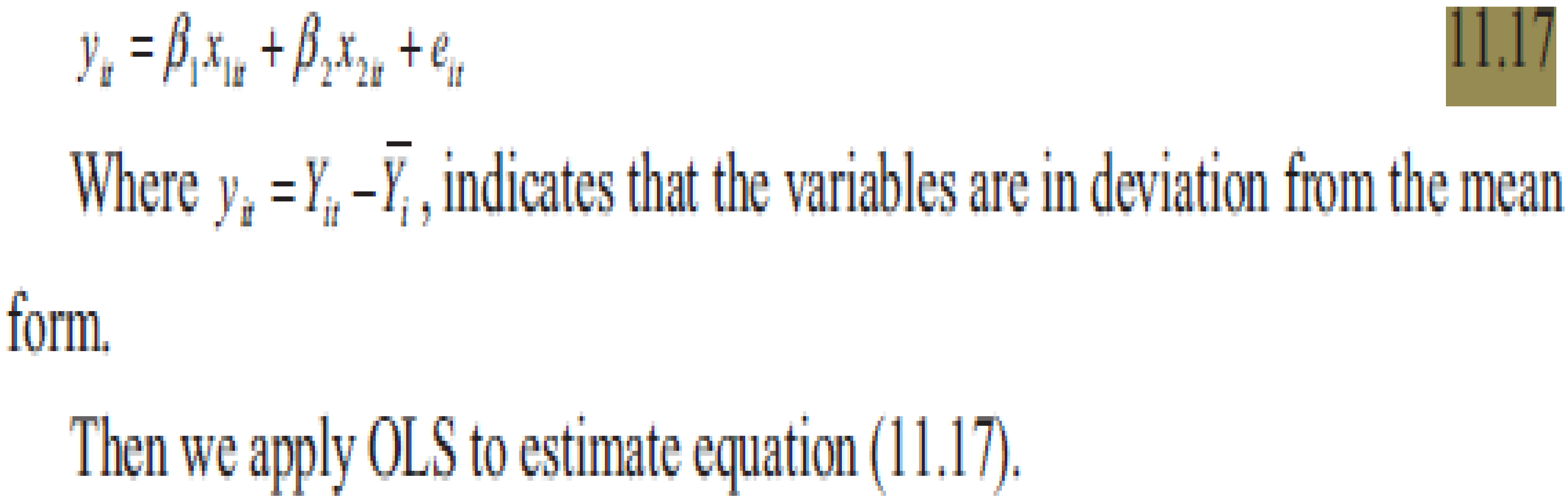
$$\begin{aligned} \bar{Y}_{it} &= \frac{1}{T}\sum_{t=1}^{T} Y_{it} = \beta_{0i} + \beta_1 \frac{1}{T}\sum_{t=1}^{T} X_{1it} + \beta_2 \frac{1}{T}\sum_{t=1}^{T} X_{2it} + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}) \\ \bar{Y}_{it} &= \beta_{0i} + \beta_1 \bar{X}_{1i} + \beta_2 \bar{X}_{2i} + \bar{\varepsilon}_i \end{aligned} \qquad 11.15$$

The ‘‘bar’’ notation $\bar{Y}_i$ indicates that we have averaged the values of Y_{it} over time. Then, subtract (11.15) from (11.14), term by term, to get

$$\begin{array}{l} Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \varepsilon_{it} \\ \underline{-(\bar{Y}_{it} = \beta_{0i} + \beta_1 \bar{X}_{1i} + \beta_2 \bar{X}_{2i} + \bar{\varepsilon}_i\ )} \\ Y_{it} - \bar{Y}_{it} = \beta_1 (X_{1it} - \bar{X}_{1i}) + \beta_2 (X_{2it} - \bar{X}_{2i}) + (\varepsilon_{it} - \bar{\varepsilon}_i) \end{array} \qquad 11.16$$

5.11 The Fixed Effects (Entity/Time Fixed) Models

- In the last line of (10.16) note that the intercept parameter β_{0i} *has fallen* out. These data are said to be in ‘‘deviation from the individual’s mean’’ form, and if we repeat this process for each individual, then we have a transformed model.



$$y_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + e_{it} \qquad 11.17$$

Where $y_{it} = Y_{it} - \bar{Y}_i$, indicates that the variables are in deviation from the mean form.

Then we apply OLS to estimate equation (11.17).

Sum….FEM

☞ **FE:**

- ☞**intercept may differ across entity, but intercepts does not vary overtime, that is it is time invariant.**
- ☞**Error terms are not correlated and each entity is different**
  - the individual differences in the intercept values of each entity are reflected in the error term.
  - One way to take into account the “individuality” of each entity or each cross sectional unit is to let the intercept vary for each entity but still assume that the slope coefficients are constant a cross firms.

□ **If T (is the number of time series data) is large and N (the number of cross- sectional units) is small. The choice is based on computational convince and FEM**

5.12 The Random Effects Model

☞ If you believe that some omitted variables may be constant over time but vary among panels, and others may be *fixed among panels but vary over time*, then you can apply random effects regression model.

☞ Random effects assume that the entity's error term is not correlated with the predictors which allows for time-invariant variables to play a role as explanatory variables.

☞ In random-effects you need to specify those individual characteristics that may or may not influence the predictor variables.

5.12 The Random Effects Model

- What we are essentially saying is that the entities included in our sample are a drawing from a much larger universe of such population and that they have a common mean value for the intercept (= α) and
- the individual differences in the intercept values of each entity are reflected in the error term.
- In random effects model (REM) or error component model (ECM) it is assumed that the intercept of an individual unit is a random drawing from a much larger population with a constant mean value.

5.12 The Random Effects Model

- ☞ **The individual intercept is then expressed as a deviation from this constant mean value.**
- ☞ **One advantage of ECM over FEM is that it is economical in degrees of freedom, as we do not have to estimate N cross-sectional intercepts.**
- ☞ **We need only to estimate the mean value of the intercept and its variance.**
- ☞ **ECM is appropriate in situations where the (random) intercept of each cross-sectional unit is uncorrelated with the regressors.**

5.12 The Random Effects Model

- **The basic idea of random effects model is to start with**

$$Y_{it} = \alpha_i + \beta_1 X_{it} + U_{it} - - - - - (1)$$

- **Instead of treating as fixed, we assume that it is a random variable with a mean value of α_i (no subscript i here).**
- **And the intercept value for individual entity can be expressed as:**

$$\alpha_i = \alpha + \varepsilon_i, \quad \text{i} = 1,2,...,\text{N}-------(2)$$

☞ **Where the random individual differences ε_i is a random error term which are called random effects, are analogous to random error terms,**

5.12 The Random Effects Model

- and we make the standard assumptions about them, namely, that they have zero mean, are uncorrelated across individuals, and have a constant variance σ_u^2, so that; $E(u_i)=0; \mathrm{cov}(u_i,u_j)=0, i \neq j; Var(u_i)=\sigma u^2$
- Substituting equ. (2) into equ. (1), we get

$$Y_{it} = \alpha + \beta_1 X_{it} + \varepsilon_{it} + u_{it} - - - - - (3)$$

$$= \alpha + \beta_1 X_{it} + w_{it}$$

- Where

$$\text{where } w_{it} = \varepsilon_i + u_{it}$$

Random effect

Regression

RE…. Sum

- **RE:**
  - **assume that the entity's error term is not correlated with the predictors which allows for time-invariant variables.**
  - **If it is assumed that the ε_i and X's are uncorrelated, ECM may be appropriate where as if ε_i and X's are correlated, FEM appropriate.**
  - **Each entity have a common mean value for the intercept.**
  - ***If N is large and T is small, and if the assumptions underlying ECM (REM) hold, ECM estimators are more efficient than FEM estimators.***
  - **the individual differences in the intercept values of each entity are reflected in the error term.**

5.13 Choosing between fixed and random effects

- **To check for any correlation between the error component ui and the regressors in a random effects model, we can use a Hausman test.**
- **This test compares the coefficient estimates from the random effects model to those from the fixed effects model.**
- **The idea underlying the Hausman test is that both the random effects and fixed effects estimators are consistent if there is no correlation between ui and the explanatory variables xkit.**

5.13 Hausman test

- If you are not exactly sure, which models, FE or RE you should use, you can do a test called Hausman test.
- To run Hausman test in Stata
- Run panel data→linear models→linear regression→ FE→ statistics→ post estimation→ manage estimation result→store in memory (save as fixed)→ run RE→ statistics→ post estimation→ manage estimation result→store in memory (save as random)→ statistics→ postestimation→tests→Hausman specification test

5.13 Hausman test

- **The hypothesis we use to test Hausman test:**
  - **H0: Random effect is appropriate model**
  - **H1: Fixed effects is appropriate model**
- **Decision Rule: If P-value is less than 5%, we accept the alternative hypothesis and we reject the null hypothesis.**

Example: Test: Ho: difference in coefficients not systematic

chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)

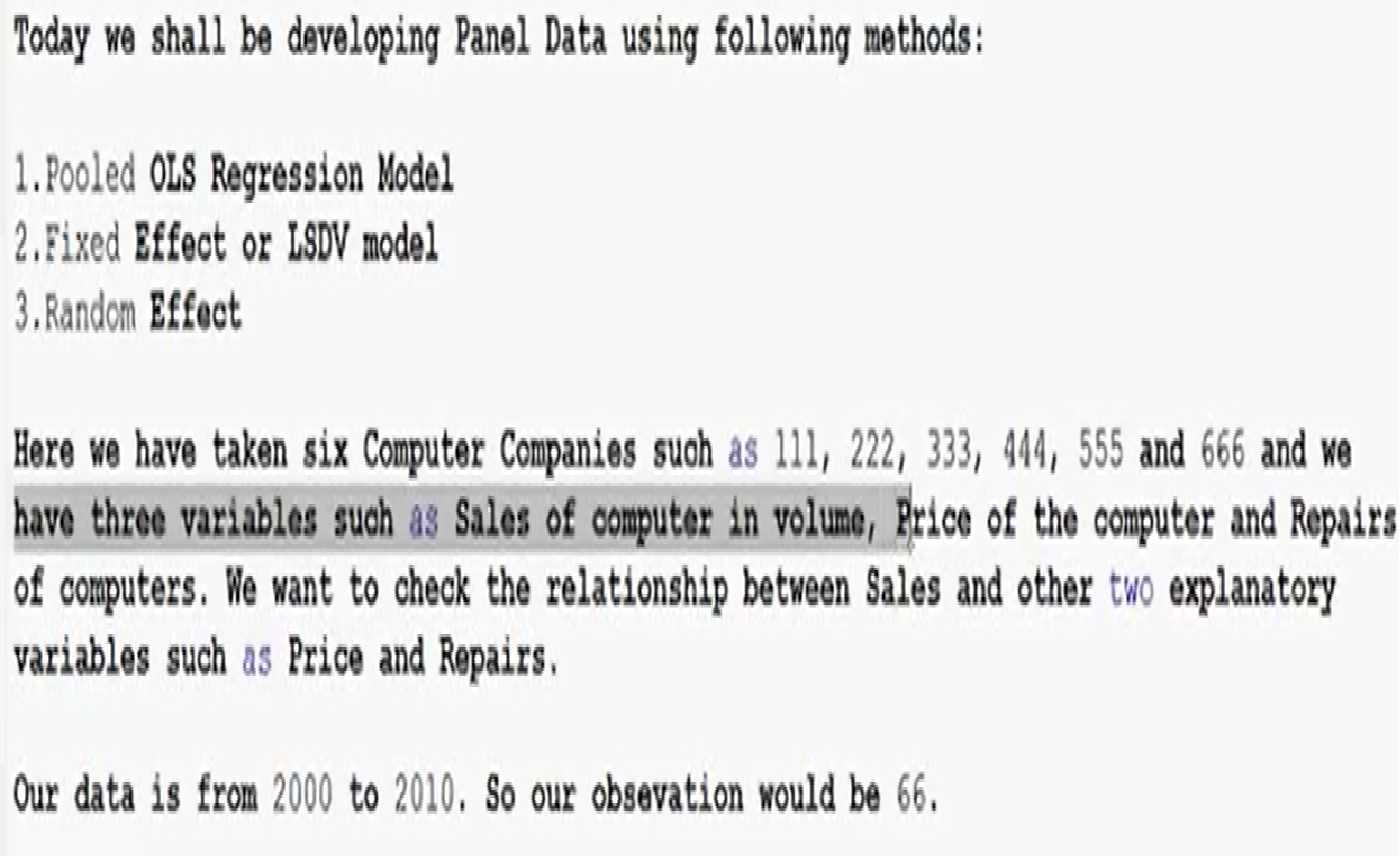
= 18.35

Prob>chi2=0.000

5.14 FE Vs RE

	FE models	RE models
Functional forms	$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}$	$Y_{it} = \alpha + \beta_1 X_{it} + \varepsilon_{it} + u_{it}$
Intercepts	**Varying a cross groups/times**	**Constant**
Error variances	**Constant**	**Varying across groups and/ times**
Slopes	**Constant**	**Constant**
Estimation	**LSDV, within effective model**	**GLS, FGLS**
Hyphothesis test	**Incremental F-test**	**Breush-Pagan LM tests**

5.15 Additional Notes



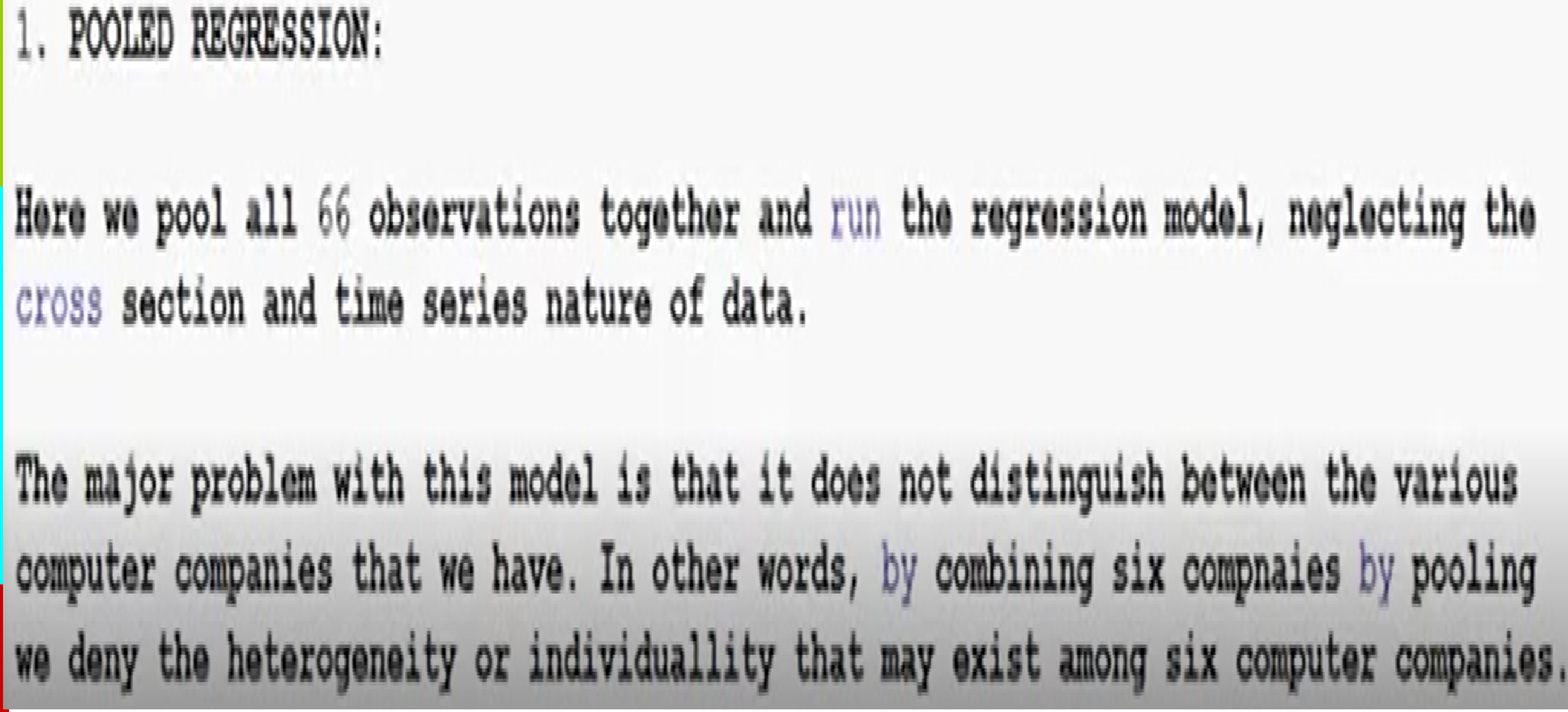
Today we shall be developing Panel Data using following methods:

1. Pooled OLS Regression Model
2. Fixed Effect or LSDV model
3. Random Effect

Here we have taken six Computer Companies such as 111, 222, 333, 444, 555 and 666 and we have three variables such as Sales of computer in volume, Price of the computer and Repairs of computers. We want to check the relationship between Sales and other two explanatory variables such as Price and Repairs.

Our data is from 2000 to 2010. So our obsevation would be 66.

Pooled Regression

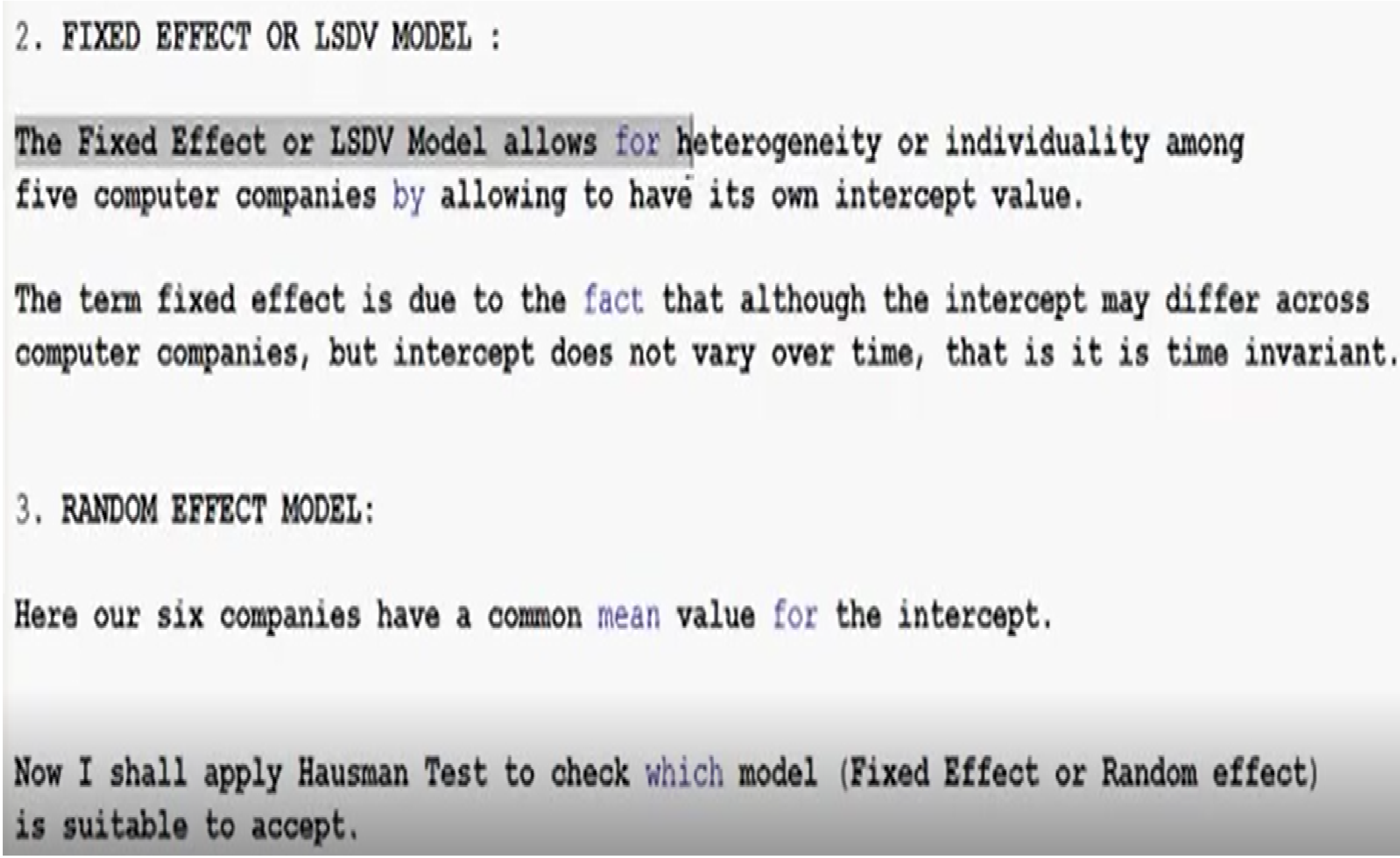


1. POOLED REGRESSION:

Here we pool all 66 observations together and run the regression model, neglecting the cross section and time series nature of data.

The major problem with this model is that it does not distinguish between the various computer companies that we have. In other words, by combining six compnaies by pooling we deny the heterogeneity or individuallity that may exist among six computer companies.

Fixed effect or LSDV model



2. FIXED EFFECT OR LSDV MODEL :

The Fixed Effect or LSDV Model allows for heterogeneity or individuality among five computer companies by allowing to have its own intercept value.

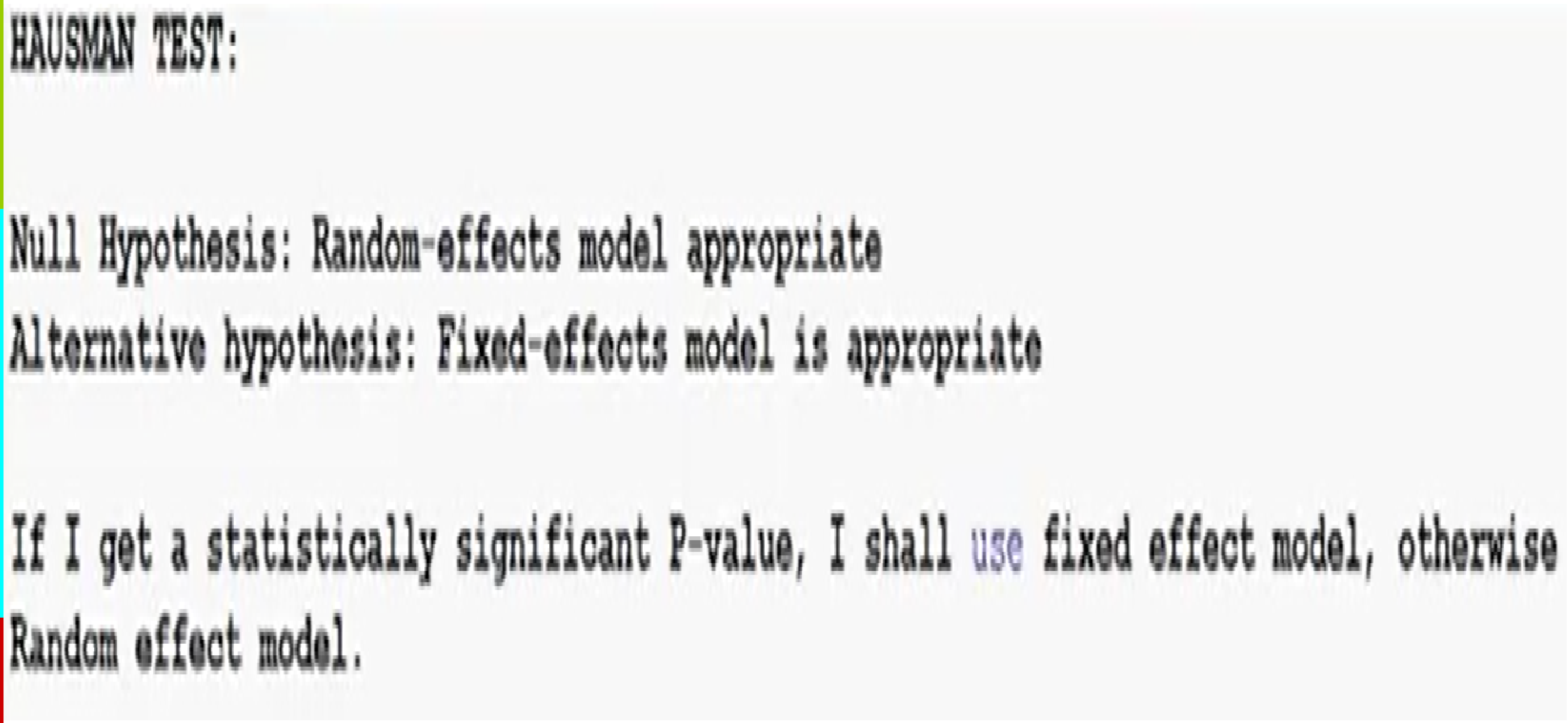
The term fixed effect is due to the fact that although the intercept may differ across computer companies, but intercept does not vary over time, that is it is time invariant.

3. RANDOM EFFECT MODEL:

Here our six companies have a common mean value for the intercept.

Now I shall apply Hausman Test to check which model (Fixed Effect or Random effect) is suitable to accept.

Hausman Test



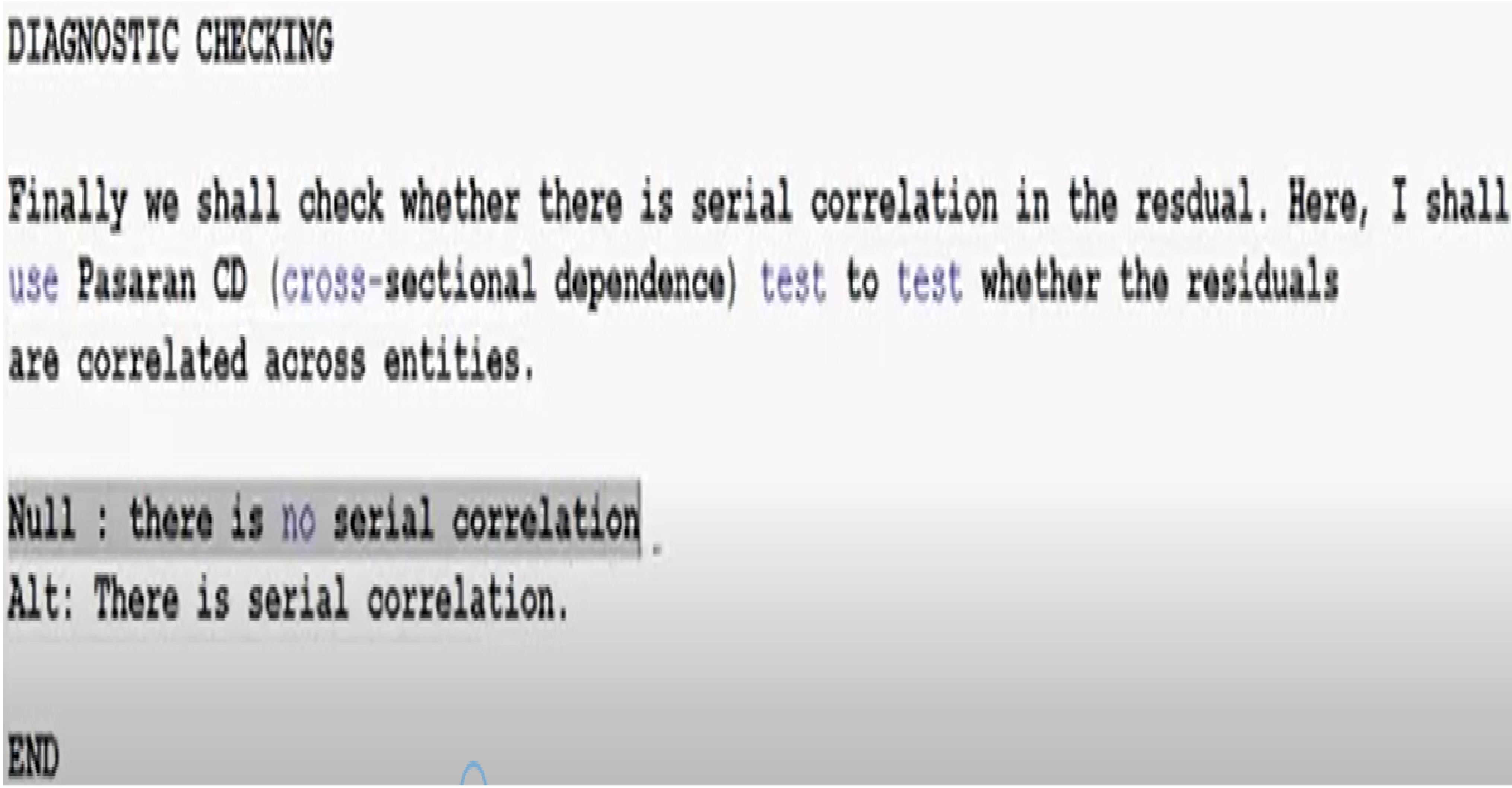
HAUSMAN TEST:

Null Hypothesis: Random-effects model appropriate

Alternative hypothesis: Fixed-effects model is appropriate

If I get a statistically significant P-value, I shall use fixed effect model, otherwise Random effect model.

Diagnostic Checking



DIAGNOSTIC CHECKING

Finally we shall check whether there is serial correlation in the resdual. Here, I shall use Pasaran CD (cross-sectional dependence) test to test whether the residuals are correlated across entities.

Null : there is no serial correlation.
Alt: There is serial correlation.

END

Summary

- ☞Balanced panel: If each cross sections unit has the same number of time series observations.
- ☞Unbanced panel: If the number of observations differ among panel data members. Friends
- ☞Initially, we assume that the X’s are non stochastic and the error term follows the classical assumptions, $E(u_{it}) \sim N(0, \sigma^2)$
- ☞Estimation of panel data regression Models

    1. FEM

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it} \qquad \text{-------A}$$

$$i = 1, 2, 3, \ldots.., n$$

$$t = 1, 2, \ldots..., n$$

FEM

☞Estimation of the above model depends on the assumption we make about the:

- Intercept
- Slope of coefficients
- Error term

➢These possibilities are:

i. Assume that the intercept and slope coefficients are constant a cross time and space and the error term captures differences overtime and individuals.

FEM

ii. The slope coefficients are constant but the intercept varies over individuals.

iii. The slope coefficients are constant but the intercept varies over individuals, and time.

iv. All coefficients are vary over individuals

V. The intercept as well as slope coefficients vary over individuals and time.

- Slope coefficients constant but the intercept varies a cross individuals: the fixed effects or Least Squares Dummy Variables (LSDV) regression on model.

FEM

- One way to take into account the “individuality” of each company or each cross sectional unit is to let the intercept vary for each company but still assume that the slope coefficients are constant a cross firms.

FEM

- **To see this, we write model A as**

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it} - - - - - - - B$$

  - **The subscript i on the intercept term to suggest that the intercepts of say 4 firms may be different. This may be different to special features of each company.**
  - **This may be dude to managerial style or managerial philosophy.**
  - **In the literature model (B) is known as fixed effect due to the fact that although the intercepts may differ across individuals, each individuals's intercept does not vary across individuals; that is time invariant.**

FEM

- ☞ If we write the intercept as β_{1it} , it will suggest that the intercept of each company or individual is time invariant.

- ☞ <u>How do we actually allow for the (FEM) intercept to vary between companies?</u>

  - □ By using dummy variable tecniques (Differential intercept dummies). We can write equation (B) as follows:

$$Y_{it} = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it} ----C$$

FEM

☞Where D2i= 1 if the observation belogs to Say Ethiopia, 0 otherwise; D3i= 1 if the observation belongs to Kenya, 0 otherwise; D4i=1 if the observation belongs to Uganda, 0 otherwise.

☞Since we have 4 dummies, we haves used only three dummies to avoid falling into the dummy variable trap.

☞FEM vs REM

☞The challenge facing researchers is that: Which model is better, FEM or ECM?

FEM vs REM

- **The answer is that:**
- **The assumption that one makes about the likely correlation between the individuals or cross section specific, error component and the X regressors.**
- **If it is assumed that theε_i and X’s are uncorrelated, ECM may be appropriate where as if ε_i and X’s are correlated, FEM appropriate.**
- **As Woodridge “ In many applications, the whole reasons for using panel data is to all the unobserved effect (ε_i) to be correlated with the explanatory variables.**
- **ECM assumptions underlying ECM is that the ε_i are a random drawing from a large population.**

FEM vs REM

1. If T (is the number of time series data) is large and N (the number of cross- sectional units) is small. The choice is based on computational convince and FEM may be preferable.
2. When N is large and T is small, the estimates obtained by the two methods can differ significant.

- Recall, that in ECM, $\beta_{1i} = \beta_1 + \varepsilon_i$, where ε_i is the cross –sectional random component, where as we treat β_{1i} as fixed and not random

FEM vs REM

3. **ECM is appropriate if we strongly believed that the individual or cross sectional units in our sample are not random drawings from a larger sample.**

4. **If the individual error component,** ε_i **and one or more regressors are correlated, then the ECM estimators are biased, where as those obtained from FEM are unbiased.**

5. ***If N is large and T is small, and if the assumptions underlying ECM (REM) hold, ECM estimators are more efficient than FEM estimators.***

Hausman Tests

Hausman Tests

- ☞ **Is there a formal test that will help us to choose between FEM and ECM? Yes, a test was developed by Hausman in 1978.**
- ☞ **The hypothesis of Hausman test is that:**
  - **H0: FEM and ECM estimateors donot differ substationally**
  - **H1: FEM and ECM estimates differ substantially.**
    - Or H0: REM is an appropriate model
    - H1: FEM is an appropriate model
    - Decision Rule: If the null hyphothesis is rejected, the conclusion is that ECM is not appropriate model and we may be better off using FEM, in which case statistical inferences will be conditions on the ε_i in the sample.

1. Panel regression models are based on panel data. Panel data consists of observations on the same cross-sectional, or individual, units over several time periods.
2. Advantages to using panel data
   - They increases the sample size considerably
   - By studying repeated cross sectional observations, panel data are better suited to study the dynamics of change.
   - If enable us to study more complicated

3. Disadvantage of using panel data

-such data involve both cross sectional and time series data, problems that plague cross sectional data (hetroscedasticity) and time series data (autocorrelation) needs to be addressed.

Lab session

Use “*lecture_5.xls*” data to practice what we learnt in previous sections

END OF CHAPTER FIVE

THANK YOU FOR BEING WITH ME

BEING @ COMMITTED

Stay Safe!

◹~~~THE END ~ ~ ~

GOOD LUCK!

Being @ committed!

All rights are reserved!