Innovative Assignment

2HSOE52 - Intro - to - Econometrics

Date
Page 1

Aayush Shah 19BCE245

1.) Discuss assumptions of classical linear regression model

→ Alike many statistical analysis, Ordering Least Squares [OLS] regression has many underlaying assumptions, when these assumptions for linear regression are true; OLS produces the best estimates.

▷ The assumptions are mentioned below:

1.) The regression model is linear in parameters.

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

▷ In this equation, the betas are the parameters that OLS estimates.

▷ The defining characteristic of linear regression is this functional form of the parameters rather than ability to model curvature.

▷ The equation may / may-not be linear in variables and can be non-linear variables by including polynomials & transforming exponential functions, but this equation has to be linear in β_1 & β_2.

2.) Values of X are fixed in repeated sampling.

▷ Values taken by the regressor X are considered fixed in repeated samples.

▷ More technically, X is assumed to be nonstochastic.

▷ Also our regressor analysis is conditional regresiomon analysis. i.e. conditional on the given values of the regressors X.

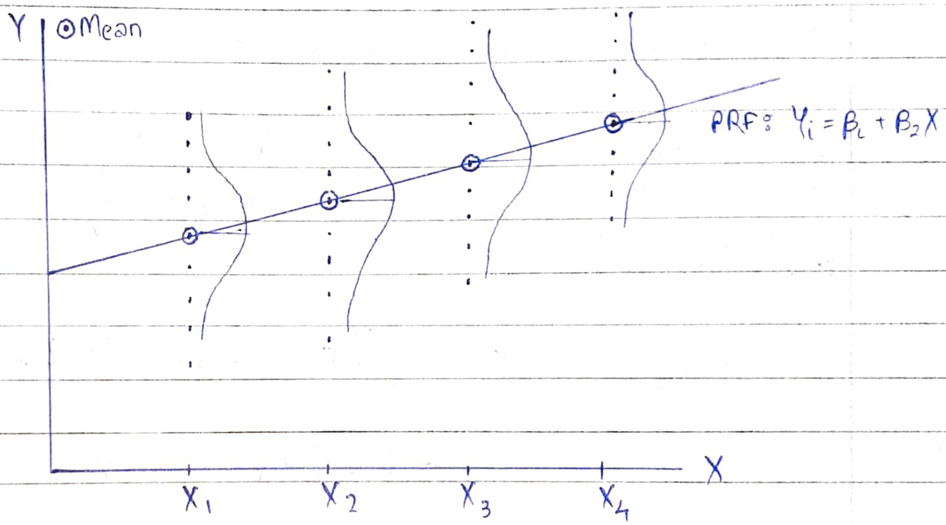
3.) Zero mean value of disturbance u_i

▷ Given the mean value of X, the mean or expected value of the random disturbance term u_i is zero.

- Technically, the conditionally mean value of u_i is zero.
- Symbollically, it is represented as

$$E(u_i / X_i) = 0$$

- ~~Gramm~~ Geometrically, this assumption can be represented as,



- This figure shows values of variable X and the Y populations associated with each of them.
- As shown, each Y population corresponding to a given X is distributed around its mean value with same Y value above the mean & some below it.
- All these assumptions says is that the factors not explicitly included in the model, and therefore & submitted in u_i, do not symmtematically affect the mean value of Y: do the positive u_i values cancel out the negative u_i values so that average or mean effect on mean Y is zero.

4) Homoscedasticity or equal variance of u_i

- Given the value of X, the variance of u_i is the same for all observations. That is, the conditional variances of u_i are identical.

▷ Symbollically, we have

$$var(U_i / X_i) = E[U_i - E(U_i / X_i)]^2$$

$$= E(U_i^2 / X_i) \quad (\because \text{due to assumption } 3)$$

$$= \sigma^2 \quad ; \text{where `var' stands for variance.}$$

5.) No autocorrelation between the disturbances

▷ Given any two X values, X_i & X_j $(i \neq j)$, the correlation between any two U_i & U_j $(i \neq j)$ is zero.

▷ Symobollicaly,

$$cov(U_i, U_j / X_i, X_j) = E\{[U_i - E(U_i)] / X_i\}\{[U_j - E(U_j)] / X_j\}$$

$$= E(U_i / X_i)(U_j / X_j)$$

$$= 0$$

where i and j are two different observations & where cov means covariance.

▷ This assumption means that, given X_i, the deviations of any two Y-values from their mean values do not exhibit patterns.

6.) Zero covariance between U_i & X_i or $E(U_i X_i) = 0$.
Formally,

$$cov(U_i, X_i) = E[U_i - E(U_i)][X_i - E(X_i)]$$

$$= E[U_i (X_i - E(X_i))] \quad (\because \text{since } E(U_i) = 0)$$

$$= E(U_i X_i) - E(X_i) E(U_i) \quad (\because \text{since } E(X_i) \text{ is non-stochastic})$$

$$= E(U_i X_i) - \quad (\because \text{since } E(U_i) = 0) = 0 \text{ by assumption}$$

- This assumption states that the disturbance u & explanotry variable X are uncorrelated.

7.) The number of observations n must be greater than the number of parameters to be estimated.

- Alternatively, the number of observations n must be greater than the number of explanatory variables.

8) Variability in X values.

- The X values in given samples must not be all the same.
- Technically, var(X) must be finite positive number.
- If all the X values are identical, then $X_i = \bar{X}$ & the denominator of that equation will be zero. making it impossible to estimate β_2 & consequtively β_1.

9) The regression model is correctly specified.

- Alternatively, there is no specification bias or error in the model used in emporical analysis.
- Questions such as omitting important variables from the models, or by chossing wrong values of functional form, or by making wrong / worse stochastic assumptions about variables of the model, the validity of interpreting the estimated regression will be highly questionable.
- Unfortunately, in practice only one rarely knows which are the corrected variables & which are to include in model.
- Therefore, One hase to use some judgement in choosing number of variables entering the model & functional form of the model & has to make some assumptions about the stochastic nature of variables included in the model.

▷ This assumption is there to remind us that our regression analysis & therefor the results based on the analysis are conditional upon the chosen model & to warn us that we should give very careful thought in formulating econometrics model especially theories trying to explain phenonenon such as inflation rate, or demand for money etc.

10.) There is no perfect multicollinearity

▷ There are no perfect relationship among the exp-lanotory variables.
▷ Multicollinearity can be checked by making a correlation matrix.
▷ Almost a sure indication of the presence of multi-collinearity is when you get opposite signs ~~fo.~~ for your regression coefficients.
▷ It is highly liked that regression suffers from multi-collinearity.
▷ If the variable is not important then dropping variable or any of the correlated variables can fix the problem.

2.) What is hypothesis ~~&~~ and explain the procedure for hypothesis testing.
→ A hypothesis is a tentative statement about the relation-ship between two or more variables.

▷ Hypothesis testing is an act in analysis statitics where by an analyst tests an assumption regarding a population parameter.
▷ A research hypothesis refers to a tentative solution to a problem which is framed in advanced before the collection & analysis of data for the given objectives.

- Once the researcher identifies & defines the research problem in precise manner, he can make a guess as to possible answers.
- These guess, which are assumed by the researcher for solving the problem or using as guide for further investigation, are called hypothesis.

- Types of Hypotheses :

Six forms of Hypothesis are :
1.) Simple Hypothesis
2.) Complex Hypothesis
3.) Directional Hypothesis
4.) Non-directional Hypothesis
5.) Null Hypothesis
6.) Associative & casual Hypothesis.

- Essential of a good hypothesis :
A good hypothesis is the one which ~~st~~ satisfies the following criteria :
1.) A hypothesis should be hypothetically emparically testable and able to deduce logical inferences.
2.) Hypothesis should be closest to the things observable and it should enable a researcher to reach at correct decision
3.) It should be conceptually clear so as to explain the concept and leaving no scope for ambiguity.
4.) The hypothesis must be specific but not in general terms.

- Procedure for testing a hypothesis :
1.) After having completed collection, processing and analysis of data a test procedure has to be followed for determining if the null hypothesis is to be accepted or rejected.

- The test procedure or the rule is based upon a test statistic and a rejection region.
- The procedure of testing hypothesis is briefly described below.

1.) Setting up a hypothesis.

- Generally there are 2 forms of hypothesis which must be constructed; and if one is accepted, the other one is rejected

i.) Null hypothesis.

- Any hypothesis concerned to a population is called statistical hypothesis
- In the process of statistical test, the rejection or acceptance of hypothesis depends on sample drawn from population.
- The sample hypothesis states that the statistical measures of sample & those of the population under study do not differ significantly
- Similarly it may assume no relationship or association between 2 variables or attributes.
- For example, if we want to find out whether extra coaching has benefitted the students or not the null hypothesis would be.

H_0: The extra coaching class has not benefited the students.

ii.) Alternative Hypothesis.

- As against the null hypothesis, the alternative hypothesis, satisfies/specifies those values that are researcher believes to hold true. and he hopes that the sample data lead to acceptance of this hypothesis as true.
- Rejection of null hypothesis H_0 leads to the acceptance of alternative hypothesis which is denoted by H_1

▷ For some examples, we can write alternative hypothesis as:

H_1: The extra coaching class has benefitted students.

▷ Null and alternative hypothesis can also be written as:

$H_0 : (\sigma_1 - \sigma_2 = 0)$; $H_0 : (\psi_1 - \psi_2 = 0)$

$H_1 : (\sigma_1 - \sigma_2 \neq 0)$; $H_1 : (\psi_1 - \psi_2 \neq 0)$

▷ Type I and Type II errors:

While we ~~we~~ make decisions on the basis of data, analysis & testing of the significant difference, it may lead to wrong conclusions in two ways:

i) Rejecting a true null hypothesis.
ii) Accepting a false hypothesis.

	Decision based on sample.	
	Accept H_0	Reject H_0.
H_0 true (H_a false)	Correct decision	Wrong decision (Type I error) $= \alpha$
H_0 ~~true~~ false (H_a ~~false~~ true)	Wrong decision. (Type II error) $\beta = 1 - \alpha$	Correct decision.

2) Setting up a suitable significant level:

▷ The maximum possibility of commiting type I error are which we use to specify test is known as level of significance.

▷ Generally 5% level of significance is fixed in statistical tests.

▷ This implies we can have 95% confidence in accepting a hypothesis or could be wrong 5% in taking the decision.

▷ The range of variation has 2 regions, acceptance region and rejection or critical region.

19BCE245

Date
Page 9

- The critical region under a normal curve can be divided into 2 ways:
  a) 2 sides under a curve. (2-tailed test)
  b) 1 side under a curve (1-tailed test), either right or left tail.

3.) Setting a test ~~curvature~~ criterion:
- This involves selecting an appropriate probability function for the particular test that is a probability distribution which can be properly ~~aligned~~ applied.
- Some probability distributions that are commonly used in testing procedures are z, t, F & χ^2.

4.) Computation
- This step includes various measures from a random sample of size n, which are necessary for applying the test.
- These calculations includes the test statistics & standard. error of the test statistics.

5.) Making a decision on conclusion.
- The decision is based on the complited value of test statistic.
- If the computed value of the test statistic falls in the acceptance region, the null hypothesis is accepted
- On the contrarey, if the computed value of statistic is greater than critical value, then the computed value falls in rejection region and the null hypothesis is rejected.

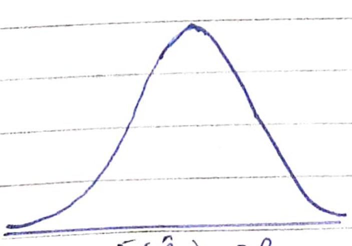
3.) What is guass-markov theorem?

→ The gauss-markov theorem states that if your linear regression model statisfies the classical assumptions, then OLS regression produces unbiased estimates that have the smallest variance of all possible linear estimators.

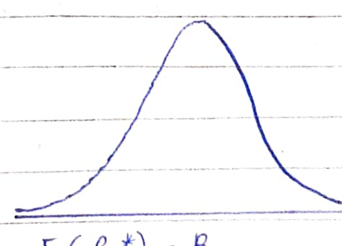
- To comphehend this theorem, let us consider the best linear unbiased property of an estimator.
- We know that an estimator let's say β_2 said to be a best linear unbiased estimator of that particular OLS estimator if the following hold:

i.) It is linear, i.e. a linear function of a random variable such as an independent variable Y in the regression model.
ii.) It is unbiased i.e. average or expected value, $E(\hat{\beta}_2)$ is equal to the true value β_2.
iii.) It has minimum variance in the class of all such linear unbiased estimators; an unbiased with all least variance is known as an efficient estimator.

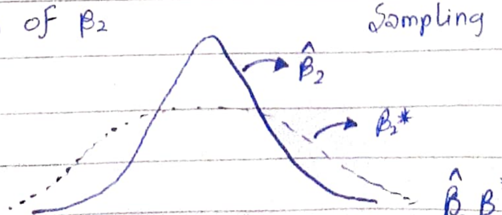
- In the regression context it can be proved that the OLS estimators are BLUE.
- It is sufficient to note here that the theoem has theoritical as well as practical importance.



(a) $E(\hat{\beta}_2) = \beta_2$.
Sampling distribution of β_2



(b.) $E(\beta_2^*) = \beta_2$.
Sampling distribution of β_2^*



(c.) Sampling β_2. distributions.

- In figure (a), Sampling distribution of the OLS estimator $\hat{\beta}_2$ i.e. the distribution of the values taken by $\hat{\beta}_2$ i.e. the distribution of values taken by $\hat{\beta}_2$ in repeated sampling experiments.
- Here for convinemce we have assumed $\hat{\beta}_2$ to be distributed symetrically.
- In figure (b), Sampling distribution of β_2^*, to an alternate-ive estimator of β_2, obtained by using another method.
- Here assume β_2^* is an unbiased & further assume both $\hat{\beta}_2$ & β_2^* are linear estimators.
- Now to resolve the dilemma of which estimator to choose, we have superimposed figures (a) & (b)
- As β_2^* is more diffused than $\hat{\beta}_2$ in figure (c), we can say the variance of $\hat{\beta}_2^*$ is greater than variance $\hat{\beta}_2$.
- So choosing smaller estimater with smaller estimator variance, is more feasible and preferable. In short, we will choose BLUE Estimators.
- The Gauss-Markov theorem is remarkable as it makes no assumptions about the probability distribution of the random variable u_i & therefore of Y_i
- If one or more assumptions of the CLRM do not hold then the theorem is invalid.
- As long as the assumption of CLRM are satisfied, the theorem holds true.