COURSE NAME: ECONOMETRICS

SUBMITTED TO: SIR ZOHAIB ALI

SUBMITTED BY: FATIMA

ID: 14299

Q1;(a) Write down the cautions in using dummy variables in regression equation?

Answer: Dummy variables also called binary, indicator, discrete, or categorical variables) are a way of incorporating qualitative information into regression analysis. To understanding dummy variable there are many social science applications in which dummy variables used to solve problem.. For example, any regression analysis involving information such as race, marital status, political party, age group, or region of residence would use dummy variables.

**cautions in using dummy variables in regression equation:**

dummy variable can be used in regression analysis through quantitative analysis.

**Shift of function:** dummy variable can be used for measuring the shift of function over a specific period of time. Intercept function changes in different period of time when shift in function suggested while other coefficient remains constant

The section in which dummy variable is not allocate in this situation dummy variable called.

1. In Example of average salary of public school teachers in three region. Here we used only two dummy variables, D2 and D3. Why did we not use three dummies to distinguish the three regions? Suppose we do that and write the model.

Yi = α + β1D1i + β2D2i + β3D3i + ui

(b); Explain the reasons and consequences of mullticolinearity?

Answer: multicollinearity is the condition where the independent variables are related to each other. As any two or more variable become more and more closely correlated, the condition worsens, and approaches singularity. Multicolinearity is almost always present, it is problem of degree not merely existence.

**There are certain reasons why multicollinearity occurs:**

* It is caused by an inaccurate use of dummy variables.
* It is caused by the addition of a variable which is computed from other variables in the data set.
* Multicollinearity can also result from the repetition of the same kind of variable.
* Generally occurs when the variables are highly correlated to each other.
* With the increase and decrease in time series data and common trend cause multicollinearity’
  + The data collection method engaged, for example, sampling over a limited range of the values taken by the regressors in the population.
  + Constraints on the model or in the population being sampled. For example, in the regression of electricity consumption on income (X2) and house size (X3) there is a physical constraint in the population in that families with higher incomes generally have larger homes than families with lower incomes.

**Consequences of multicollinearity:**

estimate will remain unbiased. Estimate will be centered around the two values. The variance of estimate will increase we are asking the model to tell us something we know very little about ( e.g what is impact of changing X on Y holding every thing constant)

* 1. Although BLUE, the OLS estimators have large variances and covariances, making precise estimation difficult

2: Because of consequence 1, the confidence intervals tend to be much wider, leading to the acceptance of the “zero null hypothesis” (i.e., the true population coefficient is zero) more readily.

3: Also because of consequence 1, the t ratio of one or more coefficients tends to be statistically insignificant.

4. Although the t ratio of one or more coefficients is statistically insignificant, R2, the overall measure of goodness of fit, can be very high. 5. The OLS estimators and their standard errors can be sensitive to small changes in the data.

(c) time series econometrics, stationarity and integrating variables?

**Time series data:**A set of observations on the values that a variable takes at different times. In time series data we observe the values of one or more variables over a period of time, like year, months. Gross Domestic Product (GDP), Consumer Price Index (CPI), S&P 500 Index, and unemployment rates daily price of stock and shares.

**Stochastic process**: a collection of random variable order in time.

Y represent random variable and t represent time and when random variable Y is continuous we refer to it as Y(t) and if random variable is individual we represent it as Yt. For example stochastic process include exchange rate, GDP and stock market.

**Stationary stochastic process:**

A stochastic process is stationary for every collection of time indices 1<t <….<tm the joint distribution of (x1+h….xtm+h) for h>1

**Integrated variables:**

The integrated variable specifies which part of integrand

The integral sign ∫ represents **integration**. The symbol dx, called the differential of the **variable** x, indicates that the **variable** of **integration** is x. The function f(x) to be **integrated** is called the integrand. The symbol dx is separated from the integrand by a space

Q.2; (a) write down the reasons, consequences and resolving method of autocorrelation in simple ordinary regression square method?

**Autocorrelation:**

Autocorrelation also known as serial correlation or lagged correlation which explain the relationship between the same variable over different periods of time. Autocorrelation is calculated as mean and variance.

The reasons of occurring autocorrelation :

Autocorrelation occur in time series data.

**Inertia:** time series such an GDP, price indexes, production, employment and unemployment exhibit business cycles. Microeconomics data experience cycle/ business cycle.

**Specification of bias\_ Excluded variable**

Specification bia arise when a potential independent variable which is related to the both dependent variable and included independent variable which is omitted from the model. The result is biased estimate of the coefficient of the included variable.

Another reason is systematic measurement.

**Consequences autocorrelation:**

Provide no other classical assumption are simultaneously violated the OLS estimator are still unbiased. However the variance of the parameter estimate will be affected.

The OLS will be inefficient and there for no longer BLUE.

The F-test formulate will also be incorrect.

Forecast based on the OLS regression model will be inefficient.

(b) what are the advantage of panel data over cross sectional pr time series data?

**Panel data;**

Panel data is a method for estimating data which is both time series and cross-sectional

Panel data have more degree of freedom and more sample variability yhan cross sectional data which is a panel wth N = 1

Also controlling the impact of omitted variables.

The ability to control for time invariant individual specific unobserved heterogeneity.

The ability to disentangle the variation over time for given individual and the variation across individuals for given time period.

Q.3; (a) write the steps of Breusch pagan LM test?

**Step#1:** Estimate the model by OLS and obtain the residuals μ^1,μ^2+⋯

2.Estimate the variance of the residuals i.e.. σ^2=∑e2i(n−2)

3. after regression the residuals square on all of the x’s can use the R-square to from an F or LM test.

4. Run the regression e2iσ2^=β1+β2Zi+μiei2σ2^=β1+β2Zi+μi and compute explained sum of squares (ESS) from this regression.

5. Test the statistical significance of ESS/2 by χ2χ2-test with 1 df at appropriate level of significance (α).

(b) explain the Breusch-Godfrey test and write its steps?

The Breusch-Pagan test is one of the most common tests for heteroskedasticity. It begins by allowing the heteroskedasticity process to be a function of one or more of your independent variables, and it’s usually applied by assuming that heteroskedasticity may be a linear function of all the independent variables in the model.

Step 1: Estimate the model by OLS and obtain the residuals

Step 2: Run the following auxiliary regression:

Step 3: Compute *LM=nR2*, where *n* and *R2* are from the auxiliary regression.

Step 4: If LM-stat>χ2p-1 critical reject the null and conclude that there is significant evidence of heteroskedasticity

Q.4; (a) Explain the following concept.

**1. pooled OLS**

Pooled data which is the mixture of time series data and cross-sectional data.

**Cross-sectional data:** Data of one or more variables, collected at the same point in time.

**Pooled data:** A combination of time series data and cross-sectional data. They are independent from each other.

**2.:Fixed effect model:**

When you are observe the same sample of individual, states, cities so fixed effect are employed.

**3.; random effect model:**

The random effects assumption is that the individual specific effects are uncorrelated with the independent variables.

(b) Explain briefly why heteroscedasticity is a problem for economic models and suggest a remedy if heteroscedasticity is detect in your analysis?

Heteroscedasticity is the assumptions of OLS. It is the most important assumption of classical linear regression. Variance of ui is constant. This assumption is known as homoscedasticity. If this assumption is not satisfied we say that ui are heteroscedasticity. The problem heteroscedasticity occur in gross section data. For example data of house hold varying income consumption.

**Reasons;**

1. **Error learning model**
2. **Grouping:** when we survey on a big scale the data in which we collected in survey are in group. When we work in group in many cases heteroscedasticity problem occurs.
3. **Income grow:** in comparison between rich family and poor family. Rich family have high expenditure.so the variance scatter.due to income growth create heteroscedasticity problems.
4. **As data collecting technique improve:**
5. **Violating one of the assumption the regression analysis;**

**Consequences:**

**1),** coefficient of estimation OLS will be unbiased and consistent so automatically OLS will be inefficient.

2),variance are very large and not constant

3), the standard error are biased when hetero is present.

**Solution:**

When variance are known then the most useful method for hetero is weighed least square.

Generalized least square