**Iqra National University**

Name: Khadija Qader

Subject: Econometrics

I.D # 14307

Submitted to: Mr. Zohaib Ali

**Q1: (a) Explain the following concepts.**

1. **Pooled OLS**

In pooled data we assume that the model has common parameters among individuals. Running an OLS technique on panel data is known as pooled OLS estimation. That’s why all individually-specific effects are ignored.

1. **Fixed Effect Model**

In the random effects model, the variables are stable. The individual-specific effect is a random variable that is uncorrelated with the explanatory variables.

1. **Random Effect Model**

In the fixed effects model, the variables are unstable and the individual-specific effect is a random variable that is allowed to be correlated with the explanatory variables.

**(b) Explain, briefly, why Heteroscedasticity is a problem for econometric models and**

**Suggest a remedy if heteroscedasticity is detected in your analysis?**

It is a problem for OLS models because in Heteroscedasticity the parameters will remain unbiased and consistent but it will not remain efficient as we know that for efficiency the variance should be minimum but in hetreroscedasticity the variance doesn’t remain the same it varies from low to high. So in that case if the estimator will not remain same and efficient than the standard error will be biased and due to this all the tests (t,f) result will biased. In short the tests will be inapplicable.

**Remedy and solution:**

Generalized Least Square method, it generalized the things, as we know the variance isn’t on one pattern so we need to generalized it and bring it in one form/pattern.

Consider the model

Yt=β1+β2X2t+β3X3t+β4X4t+…+βkXkt+ut

Where

Var(ut)=σt2

If we divide each term by the standard deviation of the error term, σt we get:

Yt=β1 (1/σt) +β2X2t/σt +β3X3t/σt +…+βkXkt/σt +ut/σt

or

Y\*t= β\*1+ β\*2X\*2t+ β\*3X\*3t+…+ β\*kX\*kt+u\*t

Where we have now that:

Var(u\*t)=Var(ut/σt)=Var(ut)/σt2=1

1 is a variance which is constant. That is the variance of the transformed disturbance term u\*, is now homoscedastic.

**Q2: (a) Write down the cautions in using dummy variables in regression equation?**

Dummy variables usually use in qualitative data where we have variables that are qualitative or nominal scale in nature such as race, color, religion, nationality etc. in such case we assign them dummy variables like 0 for male and 1 for female or may be Yes or No (dichotomous).

Things to keep in mind about dummy variables are

* The total number of dummy variables should be 1 less than the number of levels in that variable.
* One dummy variable cannot be a constant multiple/linear relation of other.
* Do not put over crowd the model with large no of dummy variables with many classes.

**(b) Explain the reasons and consequences of mullticollinearity?**

When more than one independent variables with same motive/purpose applied for same reason/nature of the model then the problem of multicollinearity occurs.

Reasons:

* Large number of variables in the model with the same concept.
* Data collection methods can also be the reason.

Consequences:

* An imperfect multicollinearity can tell that OLS estimators are Blue.
* Due to high correlation among variables the standard errors of the estimates are large.
* Large standard errors means that large confidence intervals.
* Hypothesis canot be concluded which results in disturbing of everything in the model.

**(c) Time series econometrics, stationarity and integrating variables**

**Time series econometrics:**

Any data that we collect over the period of time is known as time series data. Such as the inflation rate from 2000 till 2019, the prices of petrol from 1990 till 2020 etc. it gives a detail and large data to study with less chances of error.

**Stationary stochastic processes:**

Stochastic process is a process that has grabbed a lot of attention in time series analysis. A stochastic process is known as stationary if its variance and mean are constant over the period of time and the value of covariance between the two time periods depends only on the distance between the two time periods and not the actual time at which the covariance is computed. Thereare two forms of stationarity weak and strong.

**Integrating variable:**

If we have unit roots in our time series, a series of successive differences, d, can transform the time series into one with stationarity. The differences are denoted by I(d), where d is the order of integration. Non-stationary time series that can be transformed in this way are called series integrated of order k. Usually, the order of integration is either I(0) or I(1); It’s rare to see values for d that are 2 or more.

**Q3: (a) Write the steps of Breusch pagan LM test?**

One of the most common tests for heteroskedasiticity is Breusch pagan LM test. These are the steps:

1. It starts as that the heteroskedasticity process be the function of one or more then one of the independent variables and its assumed that it is linear function for all the independent variables of the model.
2. Estimate the auxiliary regression using OLS



1. Computing Compute LM=nR2, where n and R2 are from the auxiliary regression.
2. If LM-stat>χ2p-1 critical reject the null and conclude that there is significant evidence of heteroskedasticity

**(b) Explain the Breusch-Godfrey Test and write its steps?**

**The Breusch-Godfrey Test:**

It is a test of autocorrelation to find the errors in the regression model. It makes use of the residuals from the model being considered in regression analysis and test statistic is derived.the null hypothesis that we have generated is that there is no serial correlation of any order up to p.

**Steps of the test are:**

The Breusch-Godfrey test does not rely on the estimated standard errors, hence it does not matter whether you use heteroskedasticity-robust standard errors in your regressions or not.

1. Carry out the OLS regression and compute the residuals.
2. Regress the residuals on the independent variables of your model and on the lagged residuals.
3. Compute the test statistic by multiplying the R-squared of the second regression by your sample size.
4. Compare the test statistic against the relevant Chi-Squared distribution.

**Q4: (a) Write down the reasons, consequences and resolving method of autocorrelation in simple ordinary regression square method?**

**Reasons Autocorrelation:**

1. **Omitted explanatory variables;**

If we have missed any variables in explanatory variable then it goes into error term whos value will be auto correlated.

1. **Interpolation in the statistical observation;**

If we have ten years data (2001-2010) and we are missing 2006 so we take the average for that particular year that causes auto correlation.

1. **Misspecification of true random term ui:**

Random fators such as wars, droughts, strikes had great influence over a long period of time. For example low agriculture cropping in one period of time has influence on several other variables in other period of time.

1. **First-order autocorrelation:**

If we consider the multiple regression model

*Yt=β1+β2X2t+β3X3t+β4X4t+…+βkXkt+ut;*

In this model the the error ut is actualy the error from the periovous observation which is lagging in the current function:

ut=ρut-1+et

The first-order autocorrelation coefficient is represented as ρ and its values varies from -1 to +1. The value of ρ determines the strength of correlation.

**Consequences of autocorrelation:**

1. The estimators are still unbiased and consistent but it remains inefficient there it will no longer BLUE.
2. Since the estimated variance of the regression coefficients will be biased the hypothesis testing will be no longer valid.
3. The value of R2 in many cases is overestimated and therefore the t-statistics will be higher.

**Resolving autocorrelation;**

We have two methods of resolving autocorrelation

* When the ρ is known
* When the ρ is unknown

**When the ρ is known;**

Write the model of t-1:

Yt-1=β1+β2X2t-1+β3X3t-1+β4X4t-1+…+βkXkt-1+ut-1

Multiply both sides by ρ to get

ρYt-1= ρβ1+ ρβ2X2t-1+ ρβ3X3t-1+ ρβ4X4t-1+…+ ρ βkXkt-1+ ρut-1

Subtract those two equations:

Yt-ρYt-1= (1-ρ)β1+ β2(X2t-ρX2t-1)+ β3(X3t-ρX3t-1)+…+ βk(Xkt-ρXkt-1)+(ut-ρut-1)

or

Y\*t= β\*1+ β\*2X\*2t+ β\*3X\*3t+…+ β\*kX\*kt+et

Where now the problem of autocorrelation is resolved because et is no longer autocorrelated.

**(b) What are the advantages of panel data over cross-section or time series data?**

1. We know that panel data are relate to individuals, states, countries etc over the period of time so there is diversity in the data. By using the techniques of panel data we can easily take such diversity in data into the account through individual-specific variables.
2. Panel data gives more information, has more variability and less collinearity between the variables and there is more degree of freedom with high efficiency.
3. It is observedthat panel data gives much better result when studying the dynamics of change such as job turnover, unemployment.
4. When it comes to measure effects of anything panel data shows better results as compare to cross-section and time series data.
5. Complicated behavioral models such as economies of scale, technological change etc can be better studied with the help of panel data.
6. Since panel data gives the availability of large units it minimizes the biasness of the results.