**Subject: Econometrics**

**Instructor: Zohaib Ali**

**Mid exam Summer Semester 2020**

 Student ID 6972

Q1: (a) Write the steps of methodology of econometrics? (5 Marks)

The **methodology of econometrics** is the study of the range of differing approaches to undertaking [econometric analysis](https://en.wikipedia.org/wiki/Econometrics).

Commonly distinguished differing approaches that have been identified and studied include:

* the [Cowles Commission](https://en.wikipedia.org/wiki/Cowles_Commission) approach
* the [vector autoregression](https://en.wikipedia.org/wiki/Vector_autoregression) approach
* the [LSE approach to econometrics](https://en.wikipedia.org/wiki/LSE_approach_to_econometrics) - originated with [Denis Sargan](https://en.wikipedia.org/wiki/Denis_Sargan) now associated with [David Hendry](https://en.wikipedia.org/wiki/David_Forbes_Hendry) (and his general-to-specific modeling). Also associated this approach is the work on integrated and cointegrated systems originating on the work of [Engle](https://en.wikipedia.org/wiki/Robert_F._Engle) and [Granger](https://en.wikipedia.org/wiki/Clive_Granger) and [Johansen](https://en.wikipedia.org/wiki/S%C3%B8ren_Johansen) and [Juselius](https://en.wikipedia.org/wiki/Katarina_Juselius%22%20%5Co%20%22Katarina%20Juselius) (Juselius 1999)
* the use of calibration - [Finn Kydland](https://en.wikipedia.org/wiki/Finn_E._Kydland) and [Edward Prescott](https://en.wikipedia.org/wiki/Edward_C._Prescott)
* the [*experimentalist*](https://en.wikipedia.org/wiki/Experimentalist_approach_to_econometrics) or [difference in differences](https://en.wikipedia.org/wiki/Difference_in_differences) approach - [Joshua Angrist](https://en.wikipedia.org/wiki/Joshua_Angrist) and [Jörn-Steffen Pischke](https://en.wikipedia.org/w/index.php?title=J%C3%B6rn-Steffen_Pischke&action=edit&redlink=1" \o "Jörn-Steffen Pischke (page does not exist)).

In addition to these more clearly defined approaches, [Hoover](https://en.wikipedia.org/wiki/Kevin_Hoover)[]](https://en.wikipedia.org/wiki/Methodology_of_econometrics#cite_note-6) identifies a range of *heterogeneous* or *textbook approaches* that those less, or even un-, concerned with methodology, tend to follow.

1. Difference between causation and regression? (5 Marks)

Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

Regression helps investment and financial managers to value assets and understand the relationships between variables, such as commodity prices and the stocks of businesses dealing in those commodities.

Causation may refer to:

* [Causality](https://en.wikipedia.org/wiki/Causality), in philosophy, a relationship that describes and analyses cause and effect
* [Causality (physics)](https://en.wikipedia.org/wiki/Causality_%28physics%29)

Other uses:

* [Causation (law)](https://en.wikipedia.org/wiki/Causation_%28law%29), a key component to establish liability in both criminal and civil law
* [Causation in English law](https://en.wikipedia.org/wiki/Causation_in_English_law) defines the requirement for liability in negligence
* [Causation (sociology)](https://en.wikipedia.org/wiki/Causation_%28sociology%29), the belief that events occur in predictable ways and that one event leads to another
* [Proximate causation](https://en.wikipedia.org/wiki/Proximate_causation)
* "[Correlation does not imply causation](https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation)", phrase used in the sciences and statistics
* [Proximate cause](https://en.wikipedia.org/wiki/Proximate_cause), the basis of liability in negligence in the United States

Q2: (a) Differentiate between R-square and adjusted R-square? (5 Marks)

R-Squared only works as intended in a simple linear regression model with one explanatory variable. With a multiple regression made up of several independent variables, the R-Squared must be adjusted. The adjusted R-squared compares the descriptive power of regression models that include diverse numbers of predictors. Every predictor added to a model increases R-squared and never decreases it. Thus, a model with more terms may seem to have a better fit just for the fact that it has more terms, while the adjusted R-squared compensates for the addition of variables and only increases if the new term enhances the model above what would be obtained by probability and decreases when a predictor enhances the model less than what is predicted by chance. In an [overfitting](https://www.investopedia.com/terms/o/overfitting.asp) condition, an incorrectly high value of R-squared is obtained, even when the model actually has a decreased ability to predict. [This is not the case with the adjusted R-squared](https://www.investopedia.com/ask/answers/012615/whats-difference-between-rsquared-and-adjusted-rsquared.asp).

(b) Write down the reasons for incorporating the disturbance term? (5 marks)

An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters e, ε, or u.

Q3: (a) What are the properties of least square estimators? (5 Marks)

The method of **least squares** is a standard approach in [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis) to approximate the solution of [overdetermined systems](https://en.wikipedia.org/wiki/Overdetermined_system) (sets of equations in which there are more equations than unknowns) by minimizing the sum of the squares of the residuals made in the results of every single equation.

The most important application is in [data fitting](https://en.wikipedia.org/wiki/Curve_fitting). The best fit in the least-squares sense minimizes *the sum of squared*[*residuals*](https://en.wikipedia.org/wiki/Errors_and_residuals_in_statistics) (a residual being: the difference between an observed value, and the fitted value provided by a model). When the problem has substantial uncertainties in the [independent variable](https://en.wikipedia.org/wiki/Independent_variable) (the *x* variable), then simple regression and least-squares methods have problems; in such cases, the methodology required for fitting [errors-in-variables models](https://en.wikipedia.org/wiki/Errors-in-variables_models) may be considered instead of that for least squares.

Least-squares problems fall into two categories: linear or [ordinary least squares](https://en.wikipedia.org/wiki/Ordinary_least_squares) and [nonlinear least squares](https://en.wikipedia.org/wiki/Nonlinear_least_squares), depending on whether or not the residuals are linear in all unknowns. The linear least-squares problem occurs in statistical [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis); it has a [closed-form solution](https://en.wikipedia.org/wiki/Closed-form_solution). The nonlinear problem is usually solved by iterative refinement; at each iteration the system is approximated by a linear one, and thus the core calculation is similar in both cases.

[Polynomial least squares](https://en.wikipedia.org/wiki/Polynomial_least_squares) describes the variance in a prediction of the dependent variable as a function of the independent variable and the deviations from the fitted curve.

When the observations come from an [exponential family](https://en.wikipedia.org/wiki/Exponential_family) and mild conditions are satisfied, least-squares estimates and [maximum-likelihood](https://en.wikipedia.org/wiki/Maximum_likelihood) estimates are identical. The method of least squares can also be derived as a [method of moments](https://en.wikipedia.org/wiki/Method_of_moments_%28statistics%29) estimator.

The following discussion is mostly presented in terms of [linear](https://en.wikipedia.org/wiki/Linear) functions but the use of least squares is valid and practical for more general families of functions. Also, by iteratively applying local quadratic approximation to the likelihood (through the [Fisher information](https://en.wikipedia.org/wiki/Fisher_information)), the least-squares method may be used to fit a [generalized linear model](https://en.wikipedia.org/wiki/Generalized_linear_model).

The least-squares method is usually credited to [Carl Friedrich Gauss](https://en.wikipedia.org/wiki/Carl_Friedrich_Gauss) (1795), but it was first published by [Adrien-Marie Legendre](https://en.wikipedia.org/wiki/Adrien-Marie_Legendre) (1805).

(b) Write down the assumption of OLS? (5 Marks)

In this section, five assumptions that necessary to derive and use the OLS estimator are presented. The next section will summarize the need for each assumption in the derivation and use of the OLS estimator. You will need to know and understand these five assumptions and their use. Several of the assumptions have already been discussed, but here they are formalized.

***Assumption A - Linear in Parameters***

 This assumption has been discussed in both the simple linear and multiple regression derivations and presented above as a trait. Specifically, the assumption is

*the dependent variable y can be calculated as a linear function of a specific set of independent variables plus an error term.*

Numerous examples of linear in parameters have been presented including in the previous traits section. The equation must be linear in parameters, but does not have to be linear in the x’s. As will be discussed in the model specification reading assignment, the interpretation of the β’s depends on the functional form.

***Assumption B - Random Sample of n Observations***

 This assumption is composed of three related sub-assumptions. Two of these sub-assumptions have been previously discussed; the third is a partially new assumption to our discussion.

***Assumption B1***. *The sample consists of n-paired observations that are drawn randomly from the population.* Throughout our econometric discussion, it has been assumed a dependent variable, y, is associated with a set of the independent variables, x’s. This is often written . Recall, x1 is the variable associated with the intercept.

***Assumption B2*.** *The number of observations is greater than the number of parameters to be estimated, usually written n > k.* As discussed earlier, if n = k, the number of observations (equations) will equal the number of unknowns. In this case, OLS is not necessary, algebraic procedures can be used to derive the estimates. If n < k, the number of observations is less than the number of unknowns. In this case, neither algebra nor OLS provide unique estimates.

***Assumption B3.*** *The independent variables (x’s) are nonstochastic, whose values are fixed.* This assumption means there is a unilateral causal relationship between dependent variable, y, and the independent variables, x’s. Variations in the x’s cause variations (changes) in the y’s; the x’s cause y. On the other hand, variations in the dependent variable do not cause changes in the independent variables. Variations in y do not result in variations in the x’s; y does not cause x. The assumption also indicates that the y’s are random, because of the error terms being random and not because of randomness of the x’s. This can be shown by examining the general equation in matrix form, Y = X + U. In this equation, the X’s and the ’s are nonstochastic (fixed in our previous discussions), but U is a vector of random error terms. With Y being a linear combination of a nonstochastic component and a random component, Y must also be random; Y is random because of the random component.

 Assumption B-3 is a specific statement of the assumption we made earlier of the x’s being fixed.

***Assumption C – Zero Conditional Mean***

*The mean of the error terms has an expected value of zero given values for the independent variables*. In mathematical notation, this assumption is correctly written as . A shorthand notation is often employed and will be used in this class of the following . Here, E is the expectation operator, U the matrix of error terms, and X the matrix of independent variables. This assumption states the distribution each error term, ui, is drawn from has a mean of zero and is independent of the x’s. The last statement indicates there is no relationship between the error terms and the independent variables.

***Assumption D – No Perfect Collinearity***

 The assumption of no perfect collinearity states that *there is no exact linear relationship among the independent variables*. This assumption implies two aspects of the data on the independent variables. First, none of the independent variables, other than the variable associated with the intercept term (recall x1=1 regardless of the observation), can be a constant. Variation in the x’s is necessary. In general, the more variation in the independent variables the better the OLS estimates well be in terms of identifying the impacts of the different independent variables on the dependent variable.

 If you have three independent variables, an exact linear relationship could be represented as follows . This equation states if you know the value for x3,i and x2,i the value for x4,i is also known. For example, let x3,i = 3, x2,i = 2, and 1 = 4 and 2 = .5. Using these numbers, a value for x4,i can be found as follows x4,i = 4 \* 3 + .5 \* 2 = 13. This assumption does not allow for these types of linear relationships. In this example x4,i is not independent of x3,i and x2,i. The value for x4,i is dependent on the values for x3,i and x2.i. The assumption is the relationship cannot be perfect as in this example. A relationship that is close, but not exact does not violate this assumption. As we will see later, close relationships, however, do cause problems in using OLS.

***Assumption E - Homoskedasticity***

 *The error terms all have the same variance and are not correlated with each other.* In statistical jargon, the error terms are independent and identically distributed (iid). This assumption means the error terms associated with different observations are not related to each other. Mathematically, this assumption is written as:

 

where var represents the variance, cov the covariance, 2 is the variance, u the error terms, and X the independent variables. This assumption is more commonly written:

 

**Need / Use of the Five Assumptions**

 In this section, the importance and need for each of the five assumptions is discussed. The discussion continuously adds additional assumptions from these we have made to derive the OLS estimator. Each additional assumption allows statements to be made concerning the OLS estimator. At the same time additional assumptions make the OLS estimator less general.

***Derivation of the OLS Estimator***

 The need for assumptions in the problem setup and derivation has been previously discussed. Only a brief recap is presented. Assumptions A, B1, B2, and D are necessary for the OLS problem setup and derivation. Assumption A states the original model to be estimated must be linear in parameters. Paired observations and the number of observations being greater than k is again part of the original problem set up. This forces the use of an estimator other than algebra. Finally, no perfect collinearity allows the first order conditions to be solved. More on this assumption in a upcoming reading assignment.

***Unbiased Estimator of the Parameters,***

 One of the desirable properties of an estimator discussed earlier was the estimator should be an unbiased estimator of the true parameter values. Because the original problem has a random error term associated with the equation to be estimated, the dependent variable is a random number. Any estimator, which uses the dependent variable to estimate the parameter values, will be a random number because the dependent variable is a random number. Unbiased property is a property of the estimator and not a particular sample. Estimates from a particular sample are just fixed numbers. It makes no sense to discuss unbiased estimates of a particular sample. Rather it means the procedure to obtain the estimates is unbiased, when the procedure is viewed as being applied across all possible samples.

 If assumptions B-3, unilateral causation, and C, E(U) = 0, are added to the assumptions necessary to derive the OLS estimator, it can be shown the OLS estimator is an unbiased estimator of the true population parameters. Mathematically, unbiasedness of the OLS estimators is:

 .

By adding the two assumptions B-3 and C, the assumptions being made are stronger than for the derivation of OLS. However, these two assumptions are intuitively pleasing. Unilateral causation is stating the independent variable is caused by the dependent variables. Assumption C states the mean of the error associated with our equation is zero. KEY POINT: Under fairly unrestrictive and intuitively pleasing assumptions the OLS estimator is unbiased.

***Proof of Unbiased Property of the OLS Estimator.*** Using the equation for the OLS estimator and substituting in for Y, Y = Xβ + U, one obtains:

 

where the capital letters denote matrices. Using the distributive property of matrix algebra and the definitions of inverses and identity matrices, one obtains:

 .

Taking the expectation of both sides of the equation, recall the assumption of the expected value of the error term is zero was added in this section, leads to the following:

 .

The expectation operator can be distributed through an additive equation and a fixed value can be moved to the outside of the expectation operator. The second assumption added was unilateral causation, which means the independent variables, the X’s, are fixed.

We have proved the expected value of the OLS estimator is equal to the true value; therefore, the estimator is unbiased. Note, we have used the assumptions necessary to solve for the OLS estimator and the assumption of the mean of the error term distribution is zero, E(U) = 0. That is, we used assumptions A - D.

***Assumptions A – E***

 If we add in the assumption of homoskedasticity, the error terms all have the same variance and are not correlated with each other, three very important aspects of the OLS estimator can be given. These points are: 1) Gauss-Markov Theorem and extensions; 2) unbiased estimator for the variance of 2; and 3) variance for the . These three points cement the importance of the OLS estimator and provide the background necessary for statistical inference and tests (subject of the next reading assignment).

***Gauss-Markov Theorem.*** Because of the Gauss-Markov Theorem, OLS is one of the strongest and most used estimators for unknown parameters. The Gauss-Markov Theorem is

*Given the assumptions A – E, the OLS estimator is the Best Linear Unbiased Estimator (BLUE).*

Components of this theorem need further explanation. The first component is the linear component. This component is concerned with the estimator and not the original equation to be estimated. To stress, Assumption A is concerned with the original equation being linear in parameters. The Gauss-Markov theorem is concerned with estimators and the use of Assumptions A – E. Within the theorem, linear refers to a class of estimators that are linear in Y. To clarify, consider the OLS estimator . The dependent variable, Y enters this equation linearly. Notice it is a constant (X’X)-1 X’ multiplied by the vector Y. There are no squared terms or inverses associated with the vector, Y. This is the meaning of linear is that in the estimator, Y enters the equation linearly. The theorem states that out of the class of estimators that are linear in Y, OLS is the “Best” where “Best” refers to the smallest variance of the estimated coefficients.

Earlier, one of the desirable properties of estimators was that the estimator has minimum variance. The Gauss-Markov theorem is a very strong statement. The theorem states that any unbiased estimator you can derive, which is linear in Y, will have a larger variance than the OLS estimator.

 Combining the unbiased property with the Gauss-Markov theorem, the OLS estimator has two desirable properties that were discussed in the general problem set-up, unbiasedness and efficiency (within the class of unbiased and linear in Y estimators). This theorem provides a very strong reason to use OLS. It is unbiased and has minimum variance within the class of unbiased and linear in Y estimators. OLS estimator has the minimum mean squared error among unbiased linear in Y estimators. Recall, mean squared error considers both biasness and variance. Because in the class of estimators being considered, the biased component is zero and the variance is at a minimum, the OLS will have the minimum mean squared error in this class of estimators.

***Gauss-Markov Extension***. If it is assumed the error terms are distributed normally,

ui ~ N(0, 2), then the OLS estimator is the Best Unbiased Estimator (BUE). By adding in the addition assumption of normality of the error terms, a stronger statement concerning the variance of the estimated parameters can be given. BUE is a very strong statement. This statement is concerned with all unbiased estimators and not just those estimators that are linear in Y. OLS will have a minimum variance among all unbiased estimators.

 Note, the Gauss-Markov Theorem and its extension do not imply that OLS has the minimum variance among all potential estimators. Estimators that are not unbiased may have a smaller variance. The Gauss-Markov theorem implies nothing about the variance of these biased estimators. Further, nothing can be stated about the mean squared error property discussed earlier between the biased and unbiased estimators.

 Proof of the Gauss-Markov Theorem and its extension requires mathematical concepts that are beyond this class; therefore, the proof is not presented.

***Unbiased Estimator of .*** Assumption E states the error terms have the same variance. An estimator of the variance of the error terms is necessary. The variance of the error term is also the variance of the Y’s net of the influence of the X’s. Recall, the Y’s are random because the U’s are random.

 The simple formula for calculating the variance of a random number is:

 

where E is the expectation operator. Usually, in statistics a sample is taken, which modifies the variance formula to:

