Question No 01

(A)

Pooled OLS:

Pooled OLS is employed when you select a different sample for each year/month/period of the panel data. Fixed effects or random effects are employed when you are going to observe the same sample of individuals/countries/states/cities/etc.

Fixed effects model:

A fixed effects model is a statistical model in which the model parameters are fixed or non-random quantities. This is in contrast to random effects models and mixed models in which all or some of the model parameters are random variables.

Random effect model:

Random-effects models are statistical models in which some of the parameters (effects) that define systematic components of the model exhibit some form of random variation. Statistical models always describe variation in observed variables in terms of systematic and unsystematic components.

Part (B)

Caution in the use of dummy variables:

If there is constant in the regression than the number of dummy variables must be one less than the number of classification of each dummy variable. The coefficient which is attached to dummy variable must be interpreted in base or group.

Create separate equations for each subgroup by substituting the dummy values.

Find the difference between groups by finding the difference between their equations

Dummy variables are dichotomous, quantitative variables. Their range of values is small; they can take on only two quantitative values. As a practical matter, regression results are easiest to interpret when dummy variables are limited to two specific values, 1 or 0.

Question No (2)

(A)

Heteroscedasticity is a problem because ordinary least squares (OLS) regression assumes that all residuals are drawn from a population that has a constant variance (homoscedasticity). To satisfy the regression assumptions and be able to trust the results, the residuals should have a constant variance.

We've detected heteroscedasticity, now what can we do about it? There are various methods for resolving this issue. I'll cover three methods that I list in my order of preference. My preference is based on minimizing the amount of data manipulation. You might need to try several approaches to see which one works best. These methods are appropriate for pure heteroscedasticity but are not necessarily valid for the impure form.

- Redefining the variables
- Weighted regression
- Transform the dependent variable

Part (B)

Advantages of panel data over cross section:

Panel data is a combination of cross-sectional and time series data.

Therefore, using a regression suited to panel data has the advantage of distinguishing between fixed and random effects.

Fixed effects: Effects that are independent of random disturbances, e.g. observations independent of time.

Random effects: Effects that include random disturbances.

Panel data is more informative since it includes more information, but it has to be modelled correctly by taking into account fixed vs. random effects.

Question No (3)

Stochastic process

A stochastic or random process can be defined as a collection of random variables that is indexed by some mathematical set, meaning that each random variable of the stochastic process is uniquely associated with an element in the set.

Stationary process

In mathematics and statistics, a stationary process is a stochastic process whose unconditional joint probability distribution does not change when shifted in time. Consequently, parameters such as mean and variance also do not change over time.

Integrated Variables:

Formally, if (X,Y,Z) are each integrated of order d, and there exist coefficients a,b,c such that aX + bY + cZ is integrated of order less than d, then X, Y, and Z are cointegrated. Cointegration has become an important property in contemporary time series analysis. Time series often have trends—either deterministic or stochastic.

Unit root test:

a unit root test tests whether a time series variable is non-stationary and possesses a unit root. The null hypothesis is generally defined as the presence of a unit root and the alternative hypothesis is either stationarity, trend stationarity or explosive root depending on the test used

Cointegration or error correction model:

An error correction model belongs to a category of multiple time series models most commonly used for data where the underlying variables have a long-run stochastic trend, also known as cointegration.

An interaction variable or interaction feature is a variable constructed from an original set of variables to try to represent either all of the interaction present or some part of it. In exploratory statistical analyses it is common to use products of original variables as the basis of testing whether interaction is present with the possibility of substituting other more realistic interaction variables at a later stage. When there are more than two explanatory variables, several interaction variables are constructed, with pairwise-products representing pairwise-interactions and higher order products representing higher order interactions.

Question No 4
(A)

The binary factor A and the quantitative variable X interact (are non-additive) when analyzed with respect to the outcome variable Y.

Thus, for a response Y and two variables x1 and x2 an additive model would be:

Y = c + a x 1 + b x 2 + error {\displaystyle

Y=c+ax_{1}+bx_{2}+{\text{error}}\,}

In contrast to this,

Y = c + a x 1 + b x 2 + d (x 1 × x 2) + error {\displaystyle

Y=c+ax_{1}+bx_{2}+d(x_{1}\times x_{2})+{\text{error}}\,}

is an example of a model with an interaction between variables x1 and x2 ("error" refers to the random variable whose value is that by which Y differs from the expected value of Y; see errors and residuals in statistics). Often, models are presented without the interaction term d (x 1 × x 2) {\displaystyle d(x_{1}\times x_{2})}   , but this confounds the main effect and interaction effect (i.e., without specifying the interaction term, it is possible that any main effect found is actually due to an interaction).

Part (B)

A log-linear model is a mathematical model that takes the form of a function whose logarithm equals a linear combination of the parameters of the model, which makes it possible to apply (possibly multivariate) linear regression.

in which the fi(X) are quantities that are functions of the variable X, in general a vector of values, while c and the wi stand for the model parameters.

The term may specifically be used for:

- A log-linear plot or graph, which is a type of semi-log plot.
- Poisson regression for contingency tables, a type of generalized linear model.

The specific applications of log-linear models are where the output quantity lies in the range 0 to ∞, for values of the independent variables X, or more immediately, the transformed quantities fi(X) in the range −∞ to +∞. This may be contrasted to logistic models, similar to the logistic function, for which the output quantity lies in the range 0 to 1. Thus the contexts where these models are useful or realistic often depends on the range of the values being modelled.

Question No (5)

(A)

The spurious regression phenomenon

The spurious regression phenomenon in least squares occurs for a wide range of data generating processes, such as driftless unit roots, unit roots with drift, long memory, trend and broken-trend stationarity. Indeed, spurious regressions have played a fundamental role in the building of modern time series econometrics and have revolutionized many of the procedures used in applied macroeconomics. Spin-offs from this research range from unit-root tests to cointegration and error-correction models. This paper provides an overview of results about spurious regression, pulled from disperse sources, and explains their implications.

Part (B)

Harvey-Godfrey LM Test

Step 1: Estimate model by OLS and obtain residuals

Step 2: Run the following auxiliary regression:

Step 3: Compute LM=nR 2 , where n and R 2 are from the auxiliary regression

Step 4: If LM-stat>χ 2 p-1 critical reject the null and conclude that there is significant evidence of heteroskedasticity