BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER – 2022-23

Course No.: ECON F342 | **Course Title: APPLIED ECONOMETRICS**
Date: 13 MAY 2023 | **Max. Marks: 90**

COMPREHENSIVE EXAMINATION

NOTE:

- Attempt all questions. You have 180 minutes to complete the exam.
- Before attempting the paper please check whether all the pages of question paper are printed or not.
- This question paper consists of two parts: PART- A (Closed Book: 50 Marks) and PART-B (Open Book: 40 Marks).
- Attempt PART-A in the same answer sheet and after completing submit this part and attempt PART-B (Open Book) in the separate answer sheet. IT has multiple choice questions and short answer questions.
- Write assumptions if any clearly.
- During this comprehensive exam, students must remain quiet and may not use any external support aids, whether paper or digital (e.g. manuals, lecture notes, personal papers, books, publications, cell phones, handheld computers or other electronic devices), if not expressly authorized by the teacher in Room.
- For PART – B of this paper is OPEN BOOK EXAMINATION. The class notes, prescribed text book, reference books mentioned in the handout are allowed to use for this part. As announced earlier, NO material other than the prescribed, are not permitted to use during the examination time. No exchange of material is permitted.
- In addition, students may not copy or look at other students' exam paper or contact or attempt to contact other people in any way.
- Students must remain in the classroom for the whole of the time and only for the time needed to finish his or her exam, unless teachers in class give other orders.
- Students who have questions for the teacher must raise their hand and wait for the examiner to come to them.
- At the end of the exam, students must return the exam answer script to the invigilating faculty member and leave the room.
- Explain your answer. You need to carefully justify and show your work in the case of "open" questions to get full credit.

PART-A (CLOSED BOOK): **50,.0 MARKS**

Give the correct answer for the following multiple-choice questions. CHOOSE THE CORRECT BEST ANSWER AND PUT THE TICK (√) AGAIST THAT LETTER A/B/C/D/E. *CORRECTIONS/OVERWRITING/ILLEGIBLE ANSWERS ARE STRICTLY INVALID*. No motivation is needed.

1. Sample selection bias
   A. occurs when a selection process influences the availability of data and that process is related to the dependent variable.
   B. is more important for nonlinear least squares estimation than for OLS.
   C. is only important for finite sample results.
   D. results in the OLS estimator being biased, although it is still consistent.
   E. All of the above

2. Misspecification of functional form of the regression function
   A. results in a type of omitted variable bias.
   B. is more serious in the case of Homoscedasticity-only standard error.
   C. is overcome by adding the squares of all explanatory variables.
   D. requires alternative estimation methods such as maximum likelihood.
   E. All of the above

3. Suppose we have estimated the regression model, $y_i = \beta_1 + \beta_2 x_{i2} + \cdot\cdot\cdot + \beta_K x_{iK} + e_i$

Let $\hat{y}_i$ be the fitted value of y_i for each i. Now, we estimate the artificial model,
$y_i = \beta_1 + \beta_2 x_{i2} + \cdot\cdot\cdot + \beta_K x_{iK} + \gamma_1 \hat{y}_i + \gamma_2 \hat{y}_i^2 + v_i$

   A. H_1 can be equivalently rewritten as $H_1 : \gamma_1 \neq 0$ and $\gamma_2 \neq 0$.
   B. An F -test cannot be appropriate for testing H_0.
   C. This test is called the Augmented Dicky-Fuller test.
   D. Rejection of H_0 suggests that there can be omitted variables.
   E. None of the above is correct.

4. Which of the following issues does not cause ordinary least squares estimates to be biased and inconsistent?
   A. The dependent (y) variable is measured with error.
   B. A regressor (x) variable is measured with error.
   C. A regressor, which happens to be correlated with the included regressors, is omitted from the equation.
   D. The dependent variable and a regressor are jointly determined, like price and quantity in a supply-and-demand system.
   E. None of the above

5. Consider the regression model,
$y_i = \beta_1 + \beta_2 x_{i2} + \cdot\cdot\cdot + \beta_K x_{iK} + e_i$

where errors may be heteroskedastic. Choose the most incorrect statement.
   A. The OLS estimators are consistent and unbiased.
   B. We should report the OLS estimates with the robust standard errors.
   C. The Gauss-Markov theorem may not apply.
   D. The GLS cannot be used because we do not know the error variances in practice.
   E. We should take care of heteroscedasticity only if homoscedasticity is rejected.

---

6. Which one of the following would be the most appropriate auxiliary regression for a Ramsey RESET test of functional form?
   A. $\hat{u}_t^{\,2} = \alpha_0 + \alpha_1 \hat{y}_t^{\,2} + v_t$
   B. $u_t = \alpha_0 + \alpha_1 x_{2t} + \alpha_2 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$
   C. $y_t = \alpha_0 + \alpha_1 \hat{y}_t^{\,2} + v_t$
   D. $y_t^{\,2} = \alpha_0 + \alpha_1 x_{2t} + \alpha_2 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$
   E. None of the above

7. Consider the regression model,

$yt = \beta1 + \beta2xt2 + \cdots + \beta K\, xtK \;+ et,$

$et = \rho et{-}1 + vt\; with - 1 < \rho < 1$

where v_t's are independent random error terms with mean zero and variance σ^2_v.
Choose the wrong statement.
   A. OLS is unbiased, consistent, and efficient
   B. We may use the OLS estimates with the HAC standard errors
   C. The errors follow an AR (1) process if ρ /= 0.
   D. Var(et) = $\sigma^2{}_v/(1-\rho^2)$ and Cov(et, et−k) = $\rho k \sigma^2{}_v/(1-\rho 2)$ for k > 0.
   E. The Breusch-Godfrey test considers H_0: ρ = 0.

8. In order to estimate the wage equation, an econometrician regresses the log of wage on individual's observed characteristics including years of schooling. In this problem, however, the error term contains unobserved characteristics such as motivation, and it is likely that the error is correlated with the years of schooling, i.e., a highly motivated person tends to study more and also make more money. Choose the wrong statement.
   A. The years of schooling is endogenous.
   B. The OLS estimator is still consistent and asymptotically normal.
   C. The OLS estimator is not unbiased.
   D. The IV estimator can be used to estimate the coefficients.
   E. None of the above is wrong.

9. The time series y_t satisfies
$E[yt] = \mu < \infty$; $Var(yt) = \sigma^2 < \infty$ and
$Cov(yt, yt{-}s) = \gamma s$, (covariance depends on s, not t) Then, the series yt is said to be:
   A. cross sectional data.
   B. stationary.
   C. reliable.
   D. unreliable.
   E. None of the above.

10. Consider the autoregressive model, $y_t = \theta_0 + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + v_t$ where v_t are independent random error terms with zero means and variances σ_v^2. Choose the wrong statement.

A. This model can be denoted as AR(p) and can be estimated by OLS.
B. We often choose the value of p by hypothesis tests, residual analysis, information criteria, and parsimony.
C. OLS is not appropriate to estimate the autoregressive model given here.
D. The forecast on y_{T+1} is given as

$\hat{y}_{T+1} = \hat{\theta}_0 + \hat{\theta}_1 y_T + \hat{\theta}_2 y_{T-1} + \cdots + \hat{\theta}_p y_{T+1-p}$ where $\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_p$ are estimates of the coefficients.

E. The forecast on y_{T+2} is given as

$\hat{y}_{T+2} = \hat{\theta}_0 + \hat{\theta}_1 \hat{y}_{T+1} + \hat{\theta}_2 y_T + \cdots + \hat{\theta}_p y_{T+2-p}$ where $\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_p$ are estimates of the coefficients.

11. What is the estimator obtained through regression on quasi-demeaned data called?

A. The fixed effects estimator.
B. The mixed effects estimator.
C. The pooled OLS estimator.
D. The random effects estimator.

12. For a white noise error term ε_t, what is the time series process Y_t given by

$Y_t = \varepsilon_t + 0.2\varepsilon_{t-1} - 0.4\varepsilon_{t-2} + 0.6\varepsilon_{t-3}$

A. An ARIMA(0, 0, 3) process.
B. An *MA(4)* process.
C. An *ARMA(3,* 0) process.
D. An *ARIMA(3,* 0, 0) process..
E. None of the above

13. What is the equation obtained by differencing each variable in a single cross sectional equation over time called?

A. The difference-in-differences (DiD) equation.
B. The difference effects equation.
C. The first-differenced equation.
D. The difference-stationary equation.
E. None of the above

14. Demand for seats in a university is at its highest in the fall; demand also tends to grow and fall off in 25-year waves. In time-series forecasting, the former demand characteristic would be called ___________ and the latter would be called ___________.

A. Seasonality: cyclicality
B. Cyclicality; seasonality
C. Randomality; seasonality
D. Seasonality; variability

15. Which of the following cases describes a situation where two-time series are said to be cointegrated?
    A. Both series are 1(1) but a linear combination of them is 1(0).
    B. Both series are 1(0) but a linear combination of them is 1(1).
    C. Both series are 1(1) and a linear combination of them is 1(1).
    D. Both series are 1(0) and a linear combination of them is 1(0).
    E. None of the above

16. For time series data on two variables y and z, where Yt and Zt are dated contemporaneously, what is the model

$$y_t = \beta_0 + \beta_1 z_t + u_t, \quad t = 1, 2, ..., n$$

called?
    A. An autoregressive conditional heteroscedasticity (ARCH) model.
    B. A static model.
    C. A finite distributed lag model.
    D. An infinite distributed lag model.
    E. None of the above

17. Which of the following is a limitation of serial correlation-robust standard errors?
    A. The serial correlation-robust standard errors are smaller than OLS standard errors when there is serial correlation.
    B. The serial correlation-robust standard errors can be poorly behaved when there is substantial serial correlation and the sample size is small.
    C. The serial correlation-robust standard errors cannot be calculated for autoregressive processes of an order greater than one.
    D. The serial correlation-robust standard errors cannot be calculated after relaxing the assumption of homoscedasticity.
    E. None of the above

18. Which of the following statements about dynamically complete models is true?
    A. There is scope of adding more lags to the model to better forecast the dependent variable.
    B. The problem of serial correlation does not exist in dynamically complete models.
    C. All econometric models are dynamically complete.
    D. Sequential endogeneity is implied by dynamic completeness.
    E. None of the above

19. Which of the following statements identifies an advantage of using first differencing on time series data?
    A. First differencing eliminates the possibility of spurious regression.
    B. First differencing eliminates most of the multicollinearity.
    C. First differencing eliminates most of the serial correlation.
    D. First differencing eliminates most of the heteroscedasticity.

20. If T periods of data have been used to fit a model with p parameters and et is the residual from the model-fitting process in period t, what is the value calculated by the following formula called:

$$\ln\left(\frac{1}{T}\sum_{t=1}^{T} e_t^2\right) + \frac{2p}{T}$$

A. The logarithmic mean squared error (LMSE).
B. The log-likelihood (LL) function.
C. The Akaike Information Criteria.
D. The mean logarithmic squared error.
E. None of the above

21. A reduced form equation expresses
A. an exogenous variable as a function of endogenous variables.
B. an endogenous variable as a function of exogenous variables.
C. an exogenous variable as a function of both endogenous and exogenous variables.
D. an endogenous variable as a function of both exogenous and endogenous variables.
E. None of the above

22. Which of the following estimation techniques are available for the estimation of over-identified systems of simultaneous equations?
i) OLS
ii) ILS
iii) 2SLS
iv) IV

A. (iii) only
B. (iii) and (iv) only
C. (ii), (iii), and (iv) only
D. (i), (ii), (iii) and (iv)
E. None of the above is correct

23. For the following regression model $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$
Suppose that a researcher is interested in conducting White's heteroscedasticity test using the residuals from an estimation. What would be the most appropriate form for the auxiliary regression?
A. $u_t^2 = \alpha_0 + \alpha_1 u_{t-1} + v_t$
B. $u_t = \alpha_0 + \alpha_1 u_{t-1} + v_t$
C. $u_t^2 = \alpha_0 + \alpha_1 x_{2t} + \alpha_2 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$
D. $u_t = \alpha_0 + \alpha_1 x_{2t} + \alpha_2 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$
E. None of the above

24. Which of the following four conditions is necessary for calculating the autocorrelation function (ACF) for a time series?
A. The time series has to be stationary.
B. The time series has to be normally distributed.
C. The time series has to be persistent
D. The time series has to be symmetric.
E. None of the above

25. For time series data on two variables y and z, where Yt and Zt are dated contemporaneously, what is the following model called:

$$y_t = \alpha_0 + \beta_0 z_t + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \beta_3 z_{t-3} + u_t, \quad t = 1, 2, ..., n$$

A. An infinite distributed lag model of order 3.
B. An infinite distributed lag model of order 4.
C. A finite distributed lag model of order 3.
D. A finite distributed lag model of order 4.
E. None of the above

26. What is the process represented by the following model called?

$y_t = y_{t-1} + u_t, \quad t = 1, 2, ..., n$

A. An MA (l) process.
B. An AR (2) process.
C. A random walk process.
D. A random walk with drift process.
E. All of the above

27. Which of the following statements is incorrect?

A. Equations that are part of a recursive system can be validly estimated using OLS
B. Unnecessary use of two stage least squares- i.e. on a set of right hand side (RHS) variables that are in fact exogenous, will result in consistent but inefficient coefficient estimates
C. 2SLS is just a special case of instrumental variables (IV) estimation.
D. 2SLS and indirect least squares (ILS) are equivalent for over-identified systems
E. None of the above

28. Which of the following is an advantage of using panel data?

A. The dependent variable *y* can be differenced across time for different cross-sectional units.
B. The dependent variable *y* can be added across time for different cross- sectional units.
C. The dependent variable *y* can be differenced across time for the same cross sectional units.
D. The dependent variable *y* can be added across time for the same cross- sectional units.
E. None of the above

29. Which of the following four statements about spurious regressions is true? (

A. The OLS estimates of the population parameters are efficient and unbiased and the *t* statistic is valid.
B. R^2 may be large even if the explanatory variables and the dependent variable are independent times series processes.
C. Spurious regressions are limited to 1(0) processes, and are not possible for I(l) processes.
D. Spurious regressions are limited to I(l) processes, and are not possible for I(O) processes.
E. None of the above

30. For a white noise error term ε_t, what is the time series process Y_t given by the equation called?

$$y_t = 0.50y_{t-1} - 0.15y_{t-2} + 0.25y_{t-1} + \varepsilon_t$$

A. An ARIMA (2,0,0) process.
B. An AR (3) process.
C. An ARI MA (0,0,3) process.
D. An ARMA (l, 1) process.
E. None

31. Which of the following types of variables cannot be included in fixed effects models?
A. Dummy variables.
B. Discrete dependent variables.
C. Time-varying independent variables.
D. Time-constant independent variables.
E. None of the above

32. In a model based on a weakly dependent time series with serial correlation and strictly exogenous explanatory variables, _____.
A. the feasible generalized least square estimates are unbiased
B. the feasible generalized least square estimates are BLUE
C. the feasible generalized least square estimates are asymptotically more efficient than OLS estimates
D. the feasible generalized least square estimates are asymptotically less efficient than OLS estimates
E. None of the above

33. The fixed effects panel model is also sometimes known as
A. A seemingly unrelated regression model
B. The least squares dummy variables approach
C. The random effects model
D. Heteroscedasticity and autocorrelation consistent
E. None of the above

34. The second stage in twostage least squares estimation of a simultaneous system is:
A. Estimate the reduced form equations
B. Replace the endogenous variables that are onthe RHS of the structural equations with their reduced form fitted values
C. Replace all endogenous variables in the structural equations with their reduced form fi tted values
D. Use the fitted values of the endogenous variablesfrom the reduced forms as additional variables in the structural equations.
E. None of the above

35. Weak instruments are a problem because
A. the TSLS estimator may not be normally distributed, even in large samples.
B. they result in the instruments not being exogenous.
C. the TSLS estimator cannot be computed.
D. you cannot predict the endogenous variables any longer in the first stage.
E. None of the above

36. The conditions for a valid instrument do not include the following:
   A. each instrument must be uncorrelated with the error term.
   B. each one of the instrumental variables must be normally distributed.
   C. at least one of the instruments must enter the population regression of X on the Z's and the W's.
   D. perfect multicollinearity between the predicted endogenous variables and the exogenous variable must be ruled out.
   E. None of the above

Consider the following population model for household consumption, which has been estimated below: consumption = $\beta_0 + \beta_1$income + $\beta_2$education + $\beta_3$household size + u
(Attempt next two questions)

37. Suppose that our variable for consumption is measured with error, so cons = consumption+e_0, where *e0* is uncorrelated with *inc*, *educ* and *size*. We would expect that:
   A. our estimates of β_0, β_1, β_2, and β_3 will all definitely be biased
   B. our estimates of β_0, β_1, β_2, and β_3 will all definitely be unbiased
   C. all our standard errors will be bigger than they would be without the measurement error
   D. both B) and C)
   E. none of the above

38. Now suppose that consumption is fine, but that our variable for income is measured with error, so inc = income + e_1, where e_1 is uncorrelated with true income, educ and size. Assuming that true income is uncorrelated with educ and size, we would expect that:
   A. our estimate of β_1 will likely be less than 0.89
   B. our estimates of β_2 and β_3 will likely be around -143 and -66 respectively
   C. all our standard error will be bigger than they would be without measurement error
   D. all of the above
   E. None of the above

39. Which of the following statements is incorrect?
   A. Equations that are part of a recursive system can be validly estimated using OLS
   B. Unnecessary use of two stage least squares- i.e. on a set of right hand side (RHS) variables that are in fact exogenous, will result in consistent but inefficient coefficient estimates
   C. 2SLS is just a special case of instrumental variables (IV) estimation.
   D. 2SLS and indirect least squares (ILS) are equivalent for over-identified systems
   E. None of the above

40. The conditions for a valid instrument do not include the following:
   A. each instrument must be uncorrelated with the error term.
   B. each one of the instrumental variables must be normally distributed.
   C. at least one of the instruments must enter the population regression of X on the Z's and the W's.
   D. perfect multicollinearity between the predicted endogenous variables and the exogenous variable must be ruled out.
   E. None of the above

----

41. For the following questions answer to the point. Each question carries 2.0 Marks.

(10.0 Marks)

I. Suppose you have a model of the form $Y_i = \beta X_i + u_i$,

$$X_i = \pi Z_i + v_i ,$$

where (Y_i , X_i , Z_i) are i.i.d.; Cov (u_i , v_i) $\neq 0$; $\pi \neq 0$; and, with probability one, E [v_i |Z_i] = 0. In addition, assume that E [Y_i^4] $< \infty$, E [Z_i^4] $< \infty$, and E [X_i^4] $< \infty$. You are not sure, however, whether the condition Cov (Zi , ui) = 0 holds or not. Compute the J-statistic. Can you test the exogeneity assumption using the J-test in this case? Briefly explain.

II. How do you identify the specification error in the model using the Ramsey RESET? Mention the steps. Write the important reason to justify it as a good test?

III. The following model is estimated using a balanced panel of five firms over 20 years:

$I_{it} = \beta_1 F_{it} + \beta_2 C_{it} + u_{it}$, where the regressors are market value (F) and capital (C) and the dependent variable is investment (I).

Suppose that the true error structure of the model is $u_{it} = \alpha_i + \eta_{it}$, where α is uncorrelated with the regressors. Which model will be appropriate?
If the model is estimated as a fixed effects model, what will be the statistical properties, in terms of efficiency and consistency, of the estimates?

IV. Suppose a researcher is interested in whether having a lot of college students in a city affects the price of rental housing. Suppose that the true population model is

$$lrent_{it} = \beta_0 + \beta_1 lpop_{it} + \beta_2 lavginc_{it} + \beta_3 pctstu_{it} + \beta_4 y90_t + a_i + u_{it},$$

```
Regression with robust standard errors                 Number of obs =       128
                                                       F(  4,     60) =    691.38
                                                       Prob > F      =    0.0000
                                                       R-squared     =    0.9827
                                                       Adj R-squared =    0.9633
                                                       Root MSE      =    .06373
------------------------------------------------------------------------------
             |               Robust
       lrent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lpop |   .0722458   .0696803     1.04   0.304    -.0671357    .2116272
     lavginc |   .3099605   .0893101     3.47   0.001     .1313138    .4886072
      pctstu |   .0112033    .002936     3.82   0.000     .0053305    .0170761
         y90 |   .3855214   .0487188     7.91   0.000     .2880693    .4829735
       _cons |   1.409384   1.162338     1.21   0.230    -.9156381    3.734405
-------------+----------------------------------------------------------------
        city |   absorbed                                     (64 categories)
------------------------------------------------------------------------------
```

where lrent is the log of the rental price, lpop is the log of the city's population, lavginc is the log of per capita income, pctstu is the student population as a percent of the city population (during the school year) and y90=1 if the year is 1990. The researcher uses the fixed effect estimator to obtain the following computer software output:

Based on the above results what we can conclude that

A. the researcher thinks the unobserved fixed effect is likely correlated with the observed x variables
B. the researcher thinks these unobserved fixed effects cause serial correlation in the errors
C. the researcher thinks that the error term is heteroskedastic
D. all of the above
E. none of the above

V. Consider a non-stationary process that contains a deterministic trend:

$$z_t = \alpha + \delta t + \epsilon_t \qquad \epsilon_t \text{ IID } \left(0, \sigma_\epsilon^2\right).$$

Which of the following statements is/are correct and also prove your answer?

A The process is made stationary by regressing Z_t on t on and then replacing Z_t with $z_t - \hat{\alpha} - \hat{\delta} t = \hat{\epsilon}_t$ which is white noise under the assumption that the original model was correctly specified

B The process is made stationary by taking its first difference.

C Because it is non-stationary, Zt follows a random walk.

D If the variable Zt is first-differenced, this originates a non-invertible MA(1) that therefore cannot be represented as a stationary autoregressive process.

E None of the above

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER – 2022-23

Course No.: ECON F342 | **Course Title: APPLIED ECONOMETRICS**
Date: 13 MAY 2023 | **Max. Marks: 40**

COMPREHENSIVE EXAMINATION (PART-B)

PART- B (OPEN BOOK)
Attempt all questions. Answer to the point. Write the assumptions if any. NO PARTIAL CREDIT. Show the complete work and highlight the final answer

---

B1 Two Parts A and B.

A. The true relationship of interest is given by: $Y = \beta_0 + \beta_1 X_i^* + \varepsilon_i$. where you can assume that X_i^* and ε_I are independent. However, the researcher only observes X_i^* with error. The observed value is X_{1i} and is related to X_i^* by:
$X_{1i} = X_i^* + u_{1i}$. In this u_{1i} is mean zero and independent of (X_i^*, ε_i).

a) If the researcher runs a regression of Y_i *on* X_{1i} compute the $p\lim \hat{\beta}_1$ and give an intuition for that result.

b) The researcher has available another error-ridden measure of X_i^*. The observed value is X_{2i} and is related to X_i^* by $X_{2i} = X_i^* + u_{2i}$. where u_{1i} is mean zero, independent of $(X_i^*, \varepsilon_I, u_{1i})$ and also $Var(u_{1i}) = Var(u_{2i})$. What does this imply about the correlation between X_{1i} and X_{2i}.

c) One researcher in the department of Economics and Finance suggests defining the average of X_{1i} and X_{2i} as $X_i = (1/2)(X_{1i} + X_{2i})$ and regressing Y_i on X_i. Give an intuition for why the problems of measurement error are less in this case.

d) Another researcher in the department of Economics and Finance suggests running a regression Y_i on X_i, instrumenting X_{1i} by X_{2i} . Explain (give only one valid reason) why this will lead to a consistent estimate of β_1?

(6.0)

B. Suppose that the idiosyncratic errors in the following original unobserved effects model: $y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots\ldots\ldots + \beta_k x_{itk} + a_i + u_{it}$, t = 1, 2, 3..T , are serially uncorrelated with constant variance, σ^2_u , what is the correlation between adjacent differences, $(\Delta u_{it} \cdot \Delta u_{i,t+1})$.

(4.0)

B2 Two Parts A and B.

A. To analyze the effect of a minimum wage increase, a famous study used a quasi-experiment for two adjacent states: Gujarat and Maharashtra. A $\acute{\beta}1$ diffsindiffs estimator –– was calculated by comparing average employment changes per restaurant between to treatment group (Gujarat) and the control group (Maharashtra).

In addition, the authors provide data on the employment changes between "low wage" restaurants and "high wage" restaurants in Gujarat only. A restaurant was classified as "low wage," if the starting wage in the first wave of surveys was at the then prevailing minimum wage of Rs.4.25. A "high wage" restaurant was a place with a starting wage close to or above the Rs.5.25 minimum wage after the increase.

(4.0)

a) Explain why employment changes of the "high wage" and "low wage" restaurants might constitute a quasi-experiment. Which is the treatment group and which the control group?

b) Use the following information

	Low wage	High wage
FTE Employment before	19.56	22.25
FTE Employment after	20.88	20.21

Where FTE is "full time equivalent" and the numbers are average employment per restaurant.

Compute the change in the treatment group and control group, and finally $\acute{\beta}_1$ *diffsindiffs estimator*. Did you expect $\acute{\beta}_1$ *diffsindiffs estimator* to be positive or negative?

B. The following Econometric model is a system of simultaneous equations to study whether the openness of the economy (open) leads to lower inflation rates (inf),

$$\begin{aligned} inf &= \delta_{10} + \gamma_{12} open + \delta_{11} \log(pcinc) + u_1 \\ open &= \delta_{20} + \gamma_{21} inf + \delta_{21} \log(pcinc) + \delta_{22} \log(land) + u_2. \end{aligned}$$

We assume that (the logarithms of) pcinc (per capita income) and land (land for farming) are exogenous in the whole exercise. The following estimations have been obtained by OLS and 2SLS.

a) Discuss the possible identification of each equation of the system, the weakness of the available instruments and perform the correspondent hypothesis tests whenever is possible.
b) Explain how you would perform a test of the exogeneity of the instruments used in the two-stage estimation for an equation and whether it is possible to apply it for the equations of the given system.
c) Test whether the effect of open over inf is lower than -0:2: If open were not a determinant of inf, (but inf is a determinant of open), explain the properties of the estimates of Output 1.

(6.0)

Output 1: OLS estimation using the 114 observations 1–114
Dependent variable: inf

Variable	Coefficient	Standard Dev.	t statistic	p-value
const	25,1040	15,2052	1,6510	0,1016
open	-0,215070	0,0946289	-2,2728	0,0250
lpcinc	0,0175673	1,97527	0,0089	0,9929

Mean of dependent variable	17,2640
Std. dev. of dependent variable	23,9973
Residual sum of squares	62127,5
Residual standard deviation ($\hat{\sigma}$)	23,6581
R^2	0,0452708
$\bar{R}^2$ corrected	0,0280685
$F(2, 111)$	2,63167
p-value for $F()$	0,0764453

Output 2: OLS estimation using the 114 observations 1–114
Dependent variable: open

Variable	Coefficient	Standard Dev.	t statistic	p-value
const	116,226	15,8808	7,3187	0,0000
inf	-0,0680353	0,0715556	-0,9508	0,3438
lpcinc	0,559501	1,49395	0,3745	0,7087
lland	-7,3933	0,834814	-8,8563	0,0000

Mean of dependent variable	37,0789
Std. dev. of dependent variable	23,7535
Residual sum of squares	34865,3
Residual standard deviation ($\hat{\sigma}$)	17,8033
R^2	0,453162
$\bar{R}^2$ corrected	0,438249
$F(3, 110)$	30,3855
p-value for $F()$	$< 0,00001$

Output 3: OLS estimation using the 114 observations 1–114
Dependent variable: inf

Variable	Coefficient	Standard Dev.	t statistic	p-value
const	-12,615	21,0313	-0,5998	0,5498
lpcinc	0,191394	1,98158	0,0966	0,9232
lland	2,55380	1,08049	2,3635	0,0198

Mean of dependent variable	17,2640
Std. dev. of dependent variable	23,9973
Residual sum of squares	61903,2
Residual standard deviation ($\hat{\sigma}$)	23,6154
R^2	0,0487174
$\bar{R}^2$ corrected	0,0315772
$F(2, 111)$	2,84229
p-value for $F()$	0,0625432

Output 4: OLS estimation using the 114 observations 1–114
Dependent variable: open

Variable	Coefficient	Standard dev.	t statistic	p-value
const	117,085	15,8483	7,3878	0,0000
lpcinc	0,546479	1,49324	0,3660	0,7151
lland	-7,5671	0,814216	-9,2937	0,0000

Mean of dependent variable	37,0789
Std. dev. of dependent variable	23,7535
Residual sum of squares	35151,8
Residual standard deviation ($\hat{\sigma}$)	17,7956
R^2	0,448668
$\bar{R}^2$ corrected	0,438734
$F(2, 111)$	45,1654
p-value for $F()$	$<$0,00001

Output 5: 2SLS estimation using the 114 observations 1–114
Dependent variable: inf
Instruments: lland

Variable	Coefficient	Standard dev.	t statistic	p-value
const	26,8993	15,4012	1,7466	0,0807
open	-0,337487	0,144121	-2,3417	0,0192
lpcinc	0,375823	2,01508	0,1865	0,8520

Mean of dependent variable	17,2640
Std. dev. of dependent variable	23,9973
Residual sum of squares	63064,2
Residual standard deviation ($\hat{\sigma}$)	23,8358
$F(2, 111)$	2,62498
p-value for $F()$	0,0769352

Hausman Test –
Null hypothesis: OLS estimates are consistent
Asymptotic test statistic: $\chi^2_1 = 1{,}35333$
with p-value = 0,244697

B3 The government of Bangladesh wants to implement a program where poor families receive food stamps that can be used to purchase prepackaged foods with high nutritional value. The government decides to set up an experiment where 500 families (each with 1 child) are randomly assigned to a treatment group (eligible for food stamps, Ti = 1) and to a control group (ineligible for food stamps, Ti = 0). The government has hired a researcher to investigate the effect of food stamps on the probability that a child has poor health.

After the experiment the researcher performs a regression of H_i (a binary variable that equals 1 if a child has poor health) on Fi (a binary variable that equals 1 if a family received food stamps). The Researcher obtains the following OLS estimation results.

```
. regress H F, robust

Linear regression                               Number of obs =        500
                                                   F(  1,   498) =      20.08
                                                   Prob > F      =     0.0000
                                                   R-squared     =     0.0375
                                                   Root MSE      =     .49141
```

H	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
F	-.2090787	.0466629	-4.48	0.000	-.3007592	-.1173983
_cons	.6538462	.0381663	17.13	0.000	.5788593	.728833

The researcher decides to estimate the effect of food stamps using an instrumental variable approach. She uses assignment to the treatment group as instrument for the actual receipt of food stamps. She obtains the following first stage estimation results

```
. regress F T, robust noheader
```

F	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
T	.624	.0306963	20.33	0.000	.5636897	.6843103
_cons	.376	.0306963	12.25	0.000	.3156897	.4363103

a) Do you think that the instrument relevance condition holds? Is T a weak instrument? Briefly explain.

(3.0)

b) Do you think that the instrument exogeneity condition holds?

(3.0)

B4 Consider the model $y_i = \beta_0 + \beta_1 x_i + u_i$ and data, (y_i , x_i, z_i) i= 1,..., n where i denotes entities, y is the dependent variable, and x is an explanatory variable for each entity and z is an instrument that takes on the value of either 0 or 1 (a dummy variable).

Assume that both x and y are continuous. Note the 2SLS estimator. The following is regression output for some data. Given the information below (leave the most of unnecessary information provided in the output), **compute the $\hat{\beta}_1$?**

Sample Summary Statistics:

$\bar{y} = 54.25$ $\bar{x} = 6.1$ $\bar{z} = 0.5$

stdev(y)= 22.37 stdev(x)=2.31 stdev(z)=0.51

Regression #1Dependent Variable: X

Method: Least Squares

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	4.900000	0.636832	7.694332	0.0000
Z	2.400000	0.900617	2.664840	0.0158

R-squared 0.282908 Mean dependent var 6.100000
Adjusted R-squared 0.243069 S.D. dependent var 2.314713
S.E. of regression 2.013841 Akaike info criterion 4.332604
Sum squared resid 73.00000 Schwarz criterion 4.432177
Log likelihood -41.32604 F-statistic 7.101370
Durbin-Watson stat 1.514521 Prob(F-statistic) 0.015786

Regression #2 Dependent Variable: Y

Method: Least Squares

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	36.48330	14.13098	2.581795	0.0188
X	2.912574	2.172712	1.340524	0.1967

R-squared 0.090772 Mean dependent var 54.25000
Adjusted R-squared 0.040259 S.D. dependent var 22.37686
S.E. of regression 21.92180 Akaike info criterion 9.107479
Sum squared resid 8650.172 Schwarz criterion 9.207052
Log likelihood -89.07479 F-statistic 1.797005
Durbin-Watson stat 0.836087 Prob(F-statistic) 0.196750

Regression #3 Dependent Variable: Y

Method: Least Squares

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	42.00000	6.015027	6.982512	0.0000
Z	24.50000	8.506533	2.880139	0.0100

R-squared 0.315464 Mean dependent var 54.25000
Adjusted R-squared 0.277435 S.D. dependent var 22.37686
S.E. of regression 19.02119 Akaike info criterion 8.823623
Sum squared resid 6512.500 Schwarz criterion 8.923197

(4.0)

B5 Consider a dynamic, linear, cross country, random effects regression model

$$y_{it} = \alpha + \beta x_{it} + \delta z_{it} + \gamma y_{i,t-1} + u_i + \varepsilon_{it},\ t = 2,...,T \text{ (and } y_{i,1} \text{ is observed data).}$$

in which i is a country and t is a year. y_{it} is Petrol consumption per capita, z_{it} is is the price of Petrol and x_{it} is per capita income. You have 18 countries and 50 years of data.

Suppose that δ, the coefficient on z_{it}, is allowed to differ across countries, but it is assumed that $\delta_i = \delta + w_i$ where w_i is random noise uncorrelated with all other variables in the model.

Briefly explain on the following <u>estimation strategies for estimating α, β, δ, γ:</u>

a) **Pooled least squares. Could you confirm that the pooled ordinary least squares estimator is consistent estimator?**

b) **Regression of the 18 country means of y on a constant and country means of x, z and lagged y.**

c) **Regression of 50 period means of y on the period means of the other variables**

d) **18 separate country specific regressions, then average the estimates of α, β, δ and γ.**

(10.0)
