Debre Markos university

Department of economics

Econometrics : assignment for MSc in PPM

Please you have to submit your answer using the email address: demilie604@gmail.com

Your answer should be SMART

1. ***Choose at least 4 questions from part one (from a to h) based on your interest***
2. ***Choose at least 4 questions from part two (from 2 to 8) based on your interest***
3. ***You may consider Part three as a reading assignment (worksheet)***

***Part one:***

1. Carefully read the following statements. Are they true or false? Explain.
2. “econometrics is a science of parameter estimation on the basis of the stated theories via assuming a deterministic association among economic variables
3. logit and probit models provide constant marginal effects
4. In order to apply a *t*-test, the Gauss–Markov conditions are strictly required.
5. A regression of the OLS residual upon the regressors included in the model by construction yields an *R*2 of zero.
6. If the absolute *t*-value of a coefficient is smaller than 1.96, we accept the null hypothesis that the coefficient is zero, with 95% confidence.
7. If a variable in a model is significant at the 10% level, it is also significant at the 5% level.
8. Spherical distribution of the error term assumption is necessary to get unbiased estimators
9. In the model:$y\_{i}=β\_{0}+β\_{1}x\_{1i}+β\_{2}x\_{2i}+u\_{i}$ if $x\_{2i}=x\_{1i}+v\_{i}$ where $v\_{i}$ is white noise satisfying all the CLRM assumptions, muliticolinearity couldn’t be a serious problem

***Part two***

1. Do you think that logit and probit models are facing heteroskedasticity problem? Why?
2. For a sample of 600 married females, we are interested in explaining participation in market employment from exogenous characteristics in *xi* (age, family composition, education). Let *yi* = 1 if person *i* has a paid job and 0 otherwise. Suppose we estimate a linear regression model

$y\_{i}=X\_{i}β+ε\_{i}$ by ordinary least squares.

1. Give two reasons why this is not really an appropriate model. As an alternative, we could model the participation decision by a logit model.
2. Explain the logit model
3. To what extent is a logit model different from a probit model?

Now assume that we have a sample of women that is not working *(yi* = 0*)*, part-timeworking *(yi* = 1*)* or full-time working *(yi* = 2*)*.

1. Is it appropriate, in this case, to specify a linear model as $y\_{i}=X\_{i}β+ε\_{i}$?
2. What alternative model could be used instead, which exploits the information contained in part-time versus full-time working?
3. Assume that the series saving (*St)* and income (*Yt) are nonstationary I* (1) *series*.

***Note: consider regression of saving on income***

1. Are there indications that the relationship between the two variables is ‘spurious’?
2. Explain what we mean by ‘spurious regressions’.
3. Do you thing that there exists a possibility of a cointegrating relationship between *St* and *Yt*?
4. Explain what we mean by a ‘cointegrating relationship’.
5. Describe at least one tests that can be used to test the null hypothesis that *St* and *Yt* are not cointegrated.
6. Assuming that *St* and *Yt* are cointegrated, describe what we mean by an error correction mechanism.
7. Using a sample of 545 full-time workers, a researcher is interested in the question whether women are systematically underpaid compared to men. He estimates the average hourly wages in the sample for men and women, and hence the result is reported in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable  | Estimate | Standard error | t-ratio  |
| Constant | 5.09 | 0.58 | 8.78 |
| Male | 0.82 | 0.15 | 5.47 |

N=545 R-Squared=0.26

1. How can you interpret the coefficient estimate of 0.82? How do you interpret the estimated intercept of 5.09?
2. How do you interpret the *R*2 of 0.26?
3. Explain the relationship between the coefficient estimates in the table and the average wage rates of males and females.
4. A student is unhappy with this model as ‘a female dummy is omitted from the model’. Comment upon this criticism.
5. Test, using the above results, the hypothesis that men and women have, on average, the same wage rate, against the *one-sided* alternative that women earn less. State the assumptions required for this test to be valid.
6. Construct a 95% confidence interval for the average wage differential between males and females in the population.
7. Recall that impact evaluation is a common task for researchers and thus discus briefly at least two impact assessment models?
8. Briefly explain the difference between difference in difference and propensity score matching models?
9. If the model $y\_{i}=β\_{0}+β\_{1}x\_{1i}+β\_{2}x\_{2i}+∪\_{i}$ is to be estimated from a sample of 20 obesrvations using a data given by the following matrix form in which variables are in deviation form

$$\left(x^{I}x\right)^{-1}=\left[\begin{matrix}0.5&-0.08\\-0.08&0.6\end{matrix}\right], x^{I}y=\left[\begin{matrix}100\\250\end{matrix}\right]$$

$\overbar{X\_{1}}=\frac{\sum\_{}^{}X\_{1i}}{20}=10, \overbar{X\_{2}}=\frac{\sum\_{}^{}X\_{2i}}{20}=25, \overbar{Y}=\frac{\sum\_{}^{}Y\_{i}}{20}=30 $ Then: Obtain OLS estimators of the model

***Part three: As work sheet***

1. Which of the following economic variable/s is/are example/s of ratio scale measurements?
2. Income of consumers measured in Birr
3. Household head’s health status identified as good, very good and excellent
4. Gender of respondent identified as male and female
5. Age of respondent classified as less than 15, from 15-20, from 20-25, greater than 25
6. Regression doesn’t mean correlation because regression:
7. solely undertaken on the basis of economic theory
8. Assumes stochastic behavior of both the dependent and independent variables
9. Assumes stochastic pattern of the dependent variable
10. Is undertaken via assuming stochastic nature of independent variables
11. Which of the following is/are example/s of multiple regression model/s?
12. $ lnY=\frac{1}{1+e^{Xβ+ε}}$, where X stands for a vector of explanatory variables
13. $y=β\_{0}+β\_{1}X+ε$, where y and x reprsents level of consumption and income respectively
14. $lnwage=β\_{0}+β\_{1}education+ϵ$
15. $\frac{∆y}{y}=β\_{0}+\sqrt{β\_{1}X}+ε$
16. All except D
17. In the presence of heteroschedasticity problem, OLS estimators:
18. will be biased
19. will be inefficient
20. are efficient
21. Are both inefficient and biased
22. in a regression model a variable is said to significant if:
23. we accept the null hypothesis
24. we reject the alternative hypothesis
25. we don’t accept the null hypothesis
26. we face a difficulty to accept or reject the null hypothesis
27. a stationary time series variable is:
28. I(0) C. I(1)
29. I(2) D. I(3)
30. A random walk model of the form: $y\_{t}=y\_{t-1}+ε\_{t}$, where $ε\_{t}$ is a white noise satisfying CLRM assumptions, is:
31. Stationary at level
32. Stationary at second difference
33. Trend stationary
34. Stationary at first difference
35. Which of the following is/are true in the science of time series econometrics?
36. Different order of integration couldn’t form cointegraion unless data management doesn’t undertaken
37. Engle-Granger two step approach doesn’t restrict the number of cointegrating ranks
38. In the absence of long run relationship among variables, one will have to develop the error correction model
39. The error correction model indicates the short run dynamics only
40. The Durbin- Watson (Dw) test is primarily a test statistics to dictate:
41. Heteroschedasticity
42. Multicollinearity
43. Autocorrelation
44. Functional misspecification error
45. All of the above
46. The random disturbance term is added in econometrics model development to account:
47. Omission of variables
48. Randomness in human behavior
49. Measurement error
50. Functional misspecification problem
51. All of the above
52. Which of the following is/are part of the Gauss-Markov theorem?
53. OLS estimators are efficient and unbiased
54. Expected value of the error term is zero
55. Spherical disturbance of the error terms
56. Absence of perfect Multicollinearity
57. All of the above
58. A model can be regarded as linear regression model if it is:
59. linear in parameters irrespective of manner of variables
60. Linear in variable irrespective of manner of parameters
61. Necessarily linear in both parameters and variables
62. All of the above

***Short answer and software output***

1. Classify the following economic variables into
* ordinal
* nominal and
* ratio scale if any

note: write the right choice from the scale of measurement column on the blank space under the economic variables’ column

|  |  |
| --- | --- |
| economic variables 1. Age measured in years ----------------------------------------
2. Preference towards a good rated as good, very good and excellent ---
3. Geographical location identified as west, east, south and north ---------------
4. Annual income measured in Birr -----------------------------
5. Marital status rated as married and unmarried -----------
6. Gender identified as male and female -----------------------
7. National income measured as GDP of domestic currency
8. Color classified as black and white --------------------------
 | Scale of measurements 1. Ordinal
2. nominal
3. ratio scale
 |

1. Under what condition fixed effect model provides more efficient and consistent estimator than random effect model
2. Consider a panel model: $lwage\_{it}=β\_{0}+β\_{1}black\_{it}+β\_{2}exper\_{it}+β\_{3}married\_{it}+β\_{4}educ\_{it}+β\_{5}union\_{it}+α\_{i}+ε\_{i}$

where, : $lwage\_{it}$ is wage of the ith observation at time t in natural logarism, $black\_{it}$-color of the respondent measured as$ black\_{it}=$ $\left\{\begin{array}{c} 1 if the respondent is black\\0 otherwise\end{array}\right.$, $exper\_{it}$ is working experience of respondent measured in years, $married\_{it}=\left\{\begin{array}{c}1 if married \\o otherwise\end{array}\right.$, $educ\_{it}$- stands for years of schooling and $union\_{it}=\left\{\begin{array}{c}1 if it is member of a social union\\0 otherwise\end{array}\right.$ and $α\_{i}$ - unobserved individual heterogeneity . Suppose the estimated result using fixed effect approach be given as:

****

1. Why the variables black and educ are omitted from the regression result?
2. Do you think that union and married are positively affect wage income? Why?
3. What will you advice to consider the effect of omitted variables?
4. Consider a model: $transport choice=β\_{0}+β\_{1}age\_{i}+β\_{2}income\_{i}+ε\_{i}$ where $transport is transportaion aleternatives of \left\{\begin{array}{c}Bus \\car\\tarin \\foot\end{array}\right.$and age and income represents age and income of respondents measured in years and Birr respectively. Accordingly, the regression result is given as:



1. At 10% level of significance do you think that at higher income level individual will prefer car to on foot alternative? Why?
2. How do you interpret the coefficient of age at 10% level of significance?
3. Do you think that variables age and income are jointly significance to determine transportation choice? Why?
4. Do you think that a variable generated by a random process of the form: $y\_{t}=α+y\_{t-1}+ε\_{t}$ is a non stationary variable? Support your stand with simple algebraic manipulation (where $ε\_{t}$ is white noise with mean 0 and variance $σ^{2}$)
5. Suppose a researcher obtains the following estimates for a firm’s profitability ($π$ )using 120 samples

$$lnπ=0.87-0.54lnwage+0.65lncapital+0.3lnproduct$$

$$ \left(1.06\right) \left(0.24\right) \left(0.3\right) (0.12)$$

*Where numbers stands for standard errors of estimates and all variables in natural logarism (ln)*

Adjusted R-square ($\overbar{R}^{2})=0.34, RSS=1.24 , F\_{116}^{3}=3.75$

1. What is the explanatory power of the regression, $R^{2}$?
2. Test the null hypothesis that each variable has no effect on profitability of the firm at the 5% level of significance
3. Test the overall significance of the model
4. Interpret the result
5. Evaluate the statement “as the sample size gets larger, the variance gets smaller, and each estimate approaches the true value of the coefficient being estimated” is the unbiasedness property of OLS estimates (Hint: argue whether it is true or not)
6. Suppose the following statistical figures have been obtained from a sample of 10 observations on the values of sales (Y) of a firm and the corresponding prices (X).

$$\overbar{X}=2, \overbar{Y}=5, \sum\_{}^{}X^{2}\_{i}=100, \sum\_{}^{}XY=50, \sum\_{}^{}Y^{2}\_{i}=60$$

1. Estimate the regression line of value of sale (Y) on price (X) and interpret the results
2. What is the part of the variation in sales which is not explained by the regression line?
3. Compute the coefficient of determination ($R^{2})$
4. Compute the standard error of the regression coefficients and conduct test of significance at the 5%level of significance.