UNIVERSITY OF WESTMINSTER

WESTMINSTER BUSINESS SCHOOL

7ECON001W DATA ANALYSIS

Empirical Assignment (EA) S1 2024/25

Submission Instructions

1. Only an electronic copy should be handed-in via the blackboard site of the module by **1 p.m. (13:00)** UK time on **NOVEMBER 21 2024**. This copy will automatically be scanned through a text matching system (designed to check for possible plagiarism and collusion).

2. To avoid mark penalisation for late submission ensure you submit your coursework in time according to the deadline above. See module handbook for details on mark penalisation for late submission and step by step on-line submission instructions.

General Instructions

1. This assignment is **INDIVIDUAL.**

2. Provide and explain all calculations and relevant EViews outputs.

3. Use a significance level of 5% for all tests.

4. Presentation is worth 10% of the total coursework marks —note that a good presentation requires clear and concise answers avoiding redundant information.

5. Word limit: 1,500.[1]

6. The EA is worth 40% of the total module mark.

[1] Appendices and footnotes do not count. If necessary 1) Use image files for Tables instead of text files; 2) put tables (Eviews outputs) in an appendix and refer to them in the bodydtext of the paper, and 3) use footnotes for secondary information in your answers.

Questions

The EViews file "data.wf1" (in Blackboard/Assessment) contains data of 9275 individuals from the Survey of Income and Program Participation (SIPP) of 1991. The description of the variables in the file is given below:

$nettfa$	= net financial assets (in thousands of dollars).
age	= age of the survey respondent.
inc	= annual family income (in thousands of dollars).
$fsize$	= family size.
$marr$	= dummy equal to 1 if the respondent is married.
$male$	= dummy equal to 1 if the respondent is male.
$e401k$	= dummy equal to 1 if the respondent is eligible to participate in the retirement pension plan 401k.

Using this data set answer all following questions:

Q1. Regress the variable $nettfa$ on the variables age, inc, $fsize$, $marr$, $male$, and $e401k$ —note that a constant should be included in the model (denote it as β_1). Provide the multiple linear regression model specification including the explanatory variables in the same order as they are presented in the table above. *[10 marks]*

(a) Provide an interpretation to the coefficient estimates of that regression, including the estimated intercept.

(b) Perform tests for the statistical significance of the parameters of the independent variable age using the critical value of the corresponding t-distribution and the test p-value. Interpret the tests results.

ALL questions below refer to the regression in **Q1**:

Q2. Perform a joint significance test for the independent variables of the model using both the p-value and the critical value of the F-distribution. *[5 marks]*

(a) Comment on the goodness-of-fit of the model. What other factors might affect $nettfa$?

(b) What are the consequences of the results of this F-test together with those of the t-tests (from question 1) for the specification of the model?

Q3. Test the hypothesis: two extra years of age has half the effect of one more family member, on $nettfa$. *[10 marks]*

(a) Use the command available in EViews to test for the corresponding coefficient restriction.

(b) Perform the test analytically.

(c) Interpret the test results.

Q4. Answer the subquestions below on multicollinearity. *[8 marks]*

(a) Test for multicollinearity between the independent variables inc and $fsize$ in the model. Explain your answer using EViews outputs.

(b) Assuming that there is multicollinearity between those variables:

i. Explain how you would resolve this problem. Explain your answer using EViews outputs.

ii. What are the consequences of multicollinearity for the OLS estimator?

Q5. Answer all two parts. *[5 marks]*

(a) Perform a graphical analysis to detect the presence of heteroscedasticity in the model using at least two different plots. Do you find evidence of heteroscedasticity? Why?

(b) Explain the consequences of heteroscedasticity on the OLS estimator.

Q6. Perform a White test for heteroscedasticity. *[7 marks]*

(a) Provide the auxiliary regression and explain the meaning of the null hypothesis for this test.

(b) Why is the White test preferred to the Breusch-Pagan test for heteroscedasticity. Explain your answer.

Q7. Assume that there is heteroscedasticity of the form: $\sigma^2_{u_i} = \sigma^2_u \cdot fsize_t^{1/2}$. How would you resolve the problem of heteroscedasticity in this case? Explain your answer analytically. *[5 marks]*

Q8. Answer both parts. *[5 marks]*

(a) Estimate the model using White heteroscedasticity-consistent standard errors. Comment on the results of that estimation in relation to the estimation results in question 1.

(b) When do we use White standard errors?

Q9. Answer both parts. *[5 marks]*

(a) Provide a graphical analysis of the residuals to detect the presence of autocorrelation using at least two different plots. Do you find evidence of autocorrelation? Why?

(b) What are the consequences of autocorrelation on the OLS estimator?

Q10. Test for autocorrelation in the residuals using an appropriate procedure. *[5 marks]*

Q11. Describe step by step how you could test for the best functional form for the model in Q1. Perform the test. *[7 marks]*

Q12. Estimate the semi-logarithmic functional form of the model in Q1 and provide the interpretation of the coefficient estimates. *[6 marks]*

Q13. Test the assumption of normality in the residuals of the selected model in Q1 by using the Jarque-Bera (JB) test. Comment on the implications of your JB test results on the properties of the OLS estimator. *[6 marks]*

Q14. For what purpose can your above analysis be used by the goverment? Explain your answer. *[6 marks]*

[End of coursework]