University of Bath, Department of Economics

ES20069   Introduction to Econometrics

Coursework 2019 / 20

Instructions

1) **Deadline: noon, Friday, 4ᵗʰ May 2020!**

2) The coursework is worth **40%** of your total mark for the course.

3) There are **two parts** to this coursework. Each part carries equal weight.

4) All estimation results, graphs, etc. should be collected in an **Appendix** and clearly referenced in the text (i.e., referring to a table or a graph number in the Appendix).

5) No points will be awarded for output results without **detailed explanations** and careful presentation of arguments.

6) **Appropriate notation:** numbers should be reported with up to 3 decimal points; use Word equation editor if typing formulas; use uniform reference formatting (e.g., Chicago); use at least font size 11.

7) **Report length**: **for each part the word limit is 1500** and thus the word limit in total is 3000. This is the absolute word limit, i.e. it already includes the customary 10% extra allowance described in the student handbook. The word limit does not include the Appendix and References.

8) **Submission**: the coursework must be submitted on **Moodle**: Part I and II must be submitted as a single document in the form of **candidatenumber_ES20069**.

Good luck!

Imran Shah   2020

Some General Guidelines and "Warnings"

1) **Work individually, the coursework is <u>not</u> a group project!** Every year we identify a number of students who "work together" and they are mercilessly punished for their unfair conduct: most of these students end up with 0 mark on their coursework… Trust us, it's not worth the risk.

2) **"Reality check"**: if you get a result that you find strange/unintuitive/implausible, then before drawing the conclusion that you found something striking or there is something wrong with the data, you are strongly advised to think carefully and query the econometric/statistic tool/technique that you employed.

3) Please, do not send us **enquires about the data**: that's the data you have to work with. Datasets in reality are never perfect and it's the empirical analyst's (i.e. your) job to decide how to make the best use of your data.

4) **"Explain your finding carefully"**: when you see this, you should be alarmed: we are asking something that is not obvious, something that requires you to think. Usually a lot of marks are awarded for good explanations. Please note that the explanation does not have to be long at all; it only needs to be clear and correct.

5) Do all **hypothesis tests** properly. For instance, the way you report the significance of a coefficient reveals how well you understand hypothesis testing.

6) Do not be sloppy when you **interpret coefficients**.

7) Do not just cut and paste EViews output into your coursework, create your own tables to fit your needs.

8) **Avoid manual calculations** (e.g. F-statistics), let EViews calculate for you: the likelihood of an error is significantly smaller that way...

9) **Read the question carefully!** Students often answer questions which aren't asked and don't answer questions which are asked.

Imran Shah   2020

Part 1: Cross Sectional Analysis

The data in the EViews file **Injury_part1** contains a random sample of 7150 indemnity claims made in year 1984 from a group of insurance companies based in the states of Kentucky and Michigan in the USA. Each of the claims relates to work-related injuries and benefits ('sick pay') received while unable to work. In this part of the coursework you will investigate how policy changes affect the length of time (number of weeks) for which injured workers claimed benefits.

The dataset consists of information about the claims made in two USA states, which differ in terms of generosity of benefits. Specifically, although both Kentucky and Michigan had limits on the maximum amount of benefit which could be claimed, regardless of salary, for Kentucky the weekly limit was $131 and for Michigan the weekly limit was $181. The implication of these limits is that the loss of wages while injured would be greater for high earners (whose normal salary might exceed the limit) than for low earners. In 1984, both states changed their policy and raised the cap. Kentucky increased the maximum weekly benefit to $217, a 66% increase, and Michigan raised the maximum weekly benefit to $307, a 70% increase. There is a dummy variable in the data set **(afchnge)** which indicates whether a claim was made before or after the policy change. Specifically, it takes the value of 1 if benefits cap has changed and 0 otherwise.

The remaining variables in the dataset are as follows: **durat** = number of weeks claiming benefits ; **highearn** = variable which takes the value of 1 if someone is a high earner and 0 otherwise; **male** = variable which takes the value of 1 if someone is male and 0 otherwise; **married** = variable which takes the value of 1 if someone is married and 0 otherwise; **hosp** = variable which takes the value of 1 if someone's injury required hospital in-patient treatment and 0 otherwise; **age** = age at time of injury; **prewage** = previous weekly wage; **totmed** = total yearly medical costs; **ky** = variable which takes the value of 1 for Kentucky and 0 otherwise; **mi** = variable which takes the value of 1 for Michigan and 0 otherwise; **occdis** = variable which takes the value of 1 if the illness is classified as an occupational disease and 0 otherwise, **manuf** = variable which takes the value of 1 if someone is employed in manufacturing industry and 0 otherwise, **construc** = variable which takes the value of 1 if someone is employed in construction industry and 0 otherwise. Note that all monetary variables are expressed in real terms, as of 1982.

The dataset contains also information about the type of injury sustained: **head** = 1 if someone experienced a head injury and 0 otherwise; **neck** = 1 if someone experienced a neck injury and 0 otherwise; **upextr** = 1 if upper extremities injury and 0 otherwise; **trunk** = 1 if the injury is to the trunk (torso); **lowback** = 1 if the injury is a lower back injury and 0 otherwise; **lowextr = 1** if lower extremities injury and 0 otherwise. All other injuries are unclassified/general.

1) Preliminary data analysis.

   a. Explain what factors may affect duration of benefit claims and why (you may go beyond what is included in the data set). Comment on expected correlations between these variables and the dependent variable (duration of weekly benefits).

   b. Calculate the correlation coefficient between duration of benefits and the benefit policy change dummy. Is there a strong relationship between these two variables? Why or why not? Explain carefully.

   c. Propose a suitable regression model to explain how the reform affected the duration of benefit claims, accounting for any elements you consider important. Do not include quadratic and interaction terms at this stage. Present the exact equation and carefully justify your proposal.

Imran Shah   2020

d. Estimate the regression in part (c). Interpret each coefficient carefully. Do the coefficients on high earners and previous wages have a causal interpretation? Why or why not?

e. Conduct appropriate diagnostic tests for the regression in part (d).

f. Given the results in parts (d) and (e), suggest what adjustments can be made to the specification and why.

2) Consider introducing non-linearities into the regression.

    a. Consider the possibility that high earners are differently affected by the change in benefits cap than the low earners.

        i. Extend your original model in (1) to accommodate for this and estimate it. What do you conclude?

        ii. Given your answer in i), is the coefficient on **afchnge** likely to be causal? Why or why not?

        iii. Test for evidence of functional form misspecification in this regression. Explain your results.

    b. Research suggests that age often has a U-shaped relationship with health outcomes. For that reason, it may also affect the duration of benefit claims in the same way.

        i. Include this non-linearity in your original model in (1) and estimate the regression.

        ii. What is the marginal effect of age? Test whether it is statistically significantly different from zero.

3) We now want to investigate possible differences in the effect of the policy change between Kentucky and Michigan. Although the proportional increases in the limits were similar for both states, the monetary limits were different, which might or might not be offset by differences in the earnings distribution between the two states.

    a. Run the regression of duration of benefits on the explanatory variables (as set out in your answer to Question 2a), separately for Kentucky and Michigan. Are there notable differences in the two estimated regressions? Explain your conclusion carefully.

    b. Test whether there is equality of the parameters in the duration equation for Kentucky and Michigan. What do you conclude and why?

    c. Are there any reasons other than the policy change which can explain your results in (b)? Explain.

End of Part 1

Imran Shah   2020

Part 2: Time Series Analysis

Using US-CPI_part2 dataset answer the Question 1 below.

The data in the EViews file **US-CPI_part2** contain the data series of the quarterly U.S. consumer price index (CPI) (2015=100), after seasonal adjusted. The CPI is expressed in indices. Source: Organization for Economic Cooperation and Development. Sample period: 1980M1 – 2020M2 (T = 242).

1) Perform the ARIMA Box-Jenkins Modelling Approach. When performing the analysis, please make sure that you:

   a) Investigate the unit root properties of the series by selecting appropriate deterministic components (constant and/or linear time trend) for and discuss your results. Explain carefully.

   b) Plot and examine the sample autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the stationary series and discuss their features.

   c) Find the parsimonious ARMA model(s) that adequately describe(s) the process generating the stationary series. Report and explain your findings.

   d) Provide a complete set of diagnostic tests of the parsimonious ARMA model chosen and discuss your results. What do you conclude and why?

Using UK-Macro_part2 dataset answer Questions 2 to 4 below.

The data in the EViews file **UK-Macro_part2** contains quarterly UK data on four macroeconomic variables all in natural logs: consumption (C_0) , investment (I) , GDP in constant prices (Y) and money supply (M) respectively. They have all been seasonally adjusted, and are of time periods between 1963Q1 up to and including 2005Q4, with 172 observations.

2) Test each series for the presence of a unit root. Based on the performed analysis, what do you conclude about the presence of a unit root for each of the series? When performing the analysis, please make sure that you:

   a) Produce and discuss the graphs of the C_0, I, Y and M. What do you conclude?

   b) Test for the presence of a unit root in all four variables. You should also carefully consider whether you should include trends and constants in your tests and consider the appropriate null and alternative hypotheses. Discuss the results. What do you conclude and why?

   c) Do all the variables have the same order of integration? Explain.

3) Test for the presence of cointegration between consumption (C_0) and permanent income (Y). When performing the analysis, please make sure that you:

   a) Test for cointegration between C_0 and Y using the Engle-Granger test. Indicate all the steps performed in your analysis. Explain your results.

   b) Write down the regression that has been estimated for performing the test and discuss the results. Do you find evidence in support of the economic/financial theory? Explain your conclusion carefully.

Imran Shah   2020

4) Assuming the series (C_0, Y) are cointegrated, estimate an error-correction model (ECM). Furthermore, explain the equilibrium that justifies the estimated ECM in light of economic/financial theory and discuss if the estimates obtained in the long-run equilibrium relationship are in accordance with this economic/financial theory. When performing the analysis, please make sure that you:

a) Justify the chosen ECM model and interpret the estimation results (i.e. lags, short-run effects and error-correction term effects etc.).

b) Test whether the adjustment coefficient of the error correction term is zero. Explain your results.

c) Perform the appropriate diagnostic tests for ECM. What do you conclude and why?

End of part 2

Imran Shah   2020