

ECM308 Econometrics

COURSEWORK 2022/23

Answer BOTH PARTS of the assignment.

Part I: Regression Analysis

The data set FERTIL2.dta contains information on a large sample of women living in Botswana in 1988. The variable *children* refers to the number of living children.

1. Undertake preliminary analysis of the data to answer the following questions.
 - (a) Find the smallest and largest values of children in the sample. What is the average number of children? Does any woman have exactly the average number of children? [4 marks]
 - (b) Is number of children discrete or continuous variable? Does this observation matter for the analysis of regressions where number of children is the dependent variable? Explain your answer. [3 marks]
2. The variable *electric* is a binary indicator equal to one if the woman's home has electricity, and zero if not.
 - (a) What percentage of women have electricity in the home? [1 marks]
 - (b) Compute the average number of children for those without electricity and do the same for those with electricity. Comment on what you find. [3 marks]
 - (c) Test the hypothesis that women with electricity have on average the same number of children as those without electricity using a simple regression. Explain your answer carefully. [8 marks]
 - (d) Can you infer that having electricity "causes" women to have fewer children? Explain. [4 marks]
3. Consider the regression model where variable *children* is regressed on *age*, age^2 (that is, *age* in quadratic form), *educ*, *electric*, *urban*, and the three religious affiliation dummies.

- (a) Write down this linear regression model in matrix notation. Estimate the model and report the estimation output. Describe how the software computed the coefficients of the regression. [3 marks]
 - (b) Carefully interpret the estimated coefficient on *educ*. In particular, holding *age*, *electric* and *urban* fixed, what is the estimated effect of another year of education on fertility? If 100 women receive another year of education, how many more (or less) children are they expected to have among them? [3 marks]
 - (c) Carefully interpret the estimated coefficient on *electric*. What does the estimated parameter measure? In your explanation make sure that you identify the reference group in the regression. [5 marks]
 - (d) How does the coefficient on *electric* compare with that in question 2c? Which estimate captures better the effect of electricity on number of children? Which estimate makes more intuitive sense? Explain your answers. [5 marks]
 - (e) In this regression, is the parameter on age on its own of much interest? Explain your answer. [6 marks]
4. In the regression from question 3, explain carefully how you would test the null hypothesis that having electricity does not have a statistically significant effect on the number of children against a two-sided alternative. What is the p-value of the test? With the help of a graph, explain what a p-value means for this test. Last, say if you reject the null at the 1% significance level. [8 marks]
 5. In the regression from question 3, explain carefully how you would test the null hypothesis that the religious affiliation does not have a statistically significant effect on the number of children. Say if you reject the null at the 5% significance level. [6 marks]
 6. To regression in question 3 add an interaction between *electric* and *educ*; that is, add regressor $electric \times educ$.
 - (a) Explain carefully what the coefficient on the interaction term measures. How does the coefficient on *electric* compare with that in question 3? [5 marks]
 - (b) Using THREE criteria, explain whether you prefer regression specification in this question to the smaller model in question 3. [6 marks]

Part II: Time Series

7. Consider an ARMA(1,2) process

$$y_t = \phi y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2},$$

where ϵ_t is white noise with mean 0 and variance σ^2 .

- (a) Under what conditions is this process covariance stationary? Explain your answer. [5 marks]
 - (b) Assuming that the process is covariance stationary, derive the mean and autocovariance function of this process. [12 marks]
 - (c) Write down the theoretical expressions for one-step-ahead, two-step-ahead and three-step-ahead forecasts for this process. Compute the theoretical forecast error and variance of the forecast error. Does the variance of the forecast error converge as forecast horizon h goes to ∞ ? If so, what value does it converge to? Explain your answer. [8 marks]
8. Suppose that $\{y_t\}$ and $\{z_t\}$ are $I(1)$ processes, but $y_t - \beta z_t$ is $I(0)$ for some $\beta \neq 0$. Show that for any $\delta \neq \beta$, $y_t - \delta z_t$ is $I(1)$. [5 marks]