Econometrics 1 - Problem Set 2

LMEC, Fall 2022

Prof. Matteo Barigozzi

TA: Filippo Pavanello

October 31, 2022

In groups (as they have been previously decided), use the software Stata (and your brain), to answer the questions below. Please return one zip-file (one for each group) to filippo.pavanello2@unibo.it by **October 31, 2022 (11:59pm, 23:59)**; write as object of the email PS2-Solutions'. The zip-file must contain (i) a document with answers to each specific question (pdf format); (ii) the Stata log-file; (iii) the Stata do-file. Name the zip-file as "surname1_surname2_surname3.zip", and in any case remember to write name, surname and id number (matricola) of each student in the document.

Concise answers to all questions must be included in the pdf (either theoretical answers, or Stata output) but Stata commands and code can be contained in your do and log files. Good luck!

Question 1

Sapkota and Bastola (EE, 2017) test the validity of environmental Kuznets curve (EKC) hypothesis for 14 Latin American countries. The environmental Kuznets curve (EKC) is a hypothesized relationship between various indicators of environmental degradation - in their case CO_2 emissions - and per capita income. In the early stages of economic growth, pollution emissions increase and environmental quality declines, but beyond some level of per capita income (which will vary for different indicators) the trend reverses, so that at high income levels, economic growth leads to environmental improvement. This implies that environmental impacts or emissions per capita are an inverted U-shaped function of per capita income. The EKC is then estimated by the authors as follows:

$$CO2_i = \beta_0 + \beta_1 GDP_i + \beta_2 GDP_i^2 + \boldsymbol{\beta}_3 X_i + \epsilon_i \tag{1}$$

where $CO2_i$ is the (natural) logarithm of the level of emission per capita in country i; GDP_i is the (natural) logarithm of the Gross Domestic Product per capita; $\mathbf{X}_i$ is a vector of controls, including the natural logarithms of human capital, gross fixed capital formation, unemployment, population density, energy use, foreign direct investment; finally ϵ_i is the error term.

You are provided with a part of the dataset ps2_data1_group##.dta (sent to each group by email) the authors used for their study.

Preliminary Analysis

1. Compute the average value of GDP per capita by country and indicate the countries that registered the highest and lowest variation in the variable (it is enough to look at the average values).

2. Did these countries also registered the highest and lowest variation in the outcome variable?

3. Discuss the average value and the minimum and maximum values of the outcome variable computed at country level.

4. Make a scatter plots with a linear and a quadratic fitted line of change in CO_2 emissions over per capita Gross Domestic Product. Discuss what they suggest.

Regression Analysis

1. Estimate the following univariate regression model through OLS and using heteroskedasticity robust standard errors

$$CO2_i = \beta_0 + \beta_1 GDP_i + \epsilon_i \tag{2}$$

   Comment your results.

2. Estimate now the following multivariate regression model through OLS and using heteroskedasticity robust standard errors

$$CO2_i = \beta_0 + \beta_1 GDP_i + \boldsymbol{\beta}_3 X_i + \epsilon_i \tag{3}$$

   Provide an interpretation for all the parameters. Compare the results of the multivariate model with that obtained at the previous point, motivating your findings.

3. Now add a polynomial of order two of GDP to the previous model and comment the results. Are they in line with the environmental Kuznets curve hypothesis?

4. Decide trough a suitable test if it is optimal to include the quadratic term. Then test whether the marginal effect of GDP per capita in Model 1 is equal to zero. Finally, compute the turning point of the second-degree polynomial.

5. Decide through a suitable test if the marginal effect of GDP per capita in the model estimated in the previous point is equal to zero for a country with GDP per capita equal to 5, 7 and 9.

6. From now on, use the regression model with the quadratic term of GDP per capita. Divide countries in high and low-income based on the median of the per capita Gross Domestic Product (GDP), creating a dummy variable $High_i$. Now include in Model 1 an interaction term between $High_i$ and the two GDP covariates. Is the marginal effect of GDP per capita different between high and low-income countries? Show it with an appropriate test.

7. Perform a Chow test in order to test if there is a structural break between high and low-income countries.

Question 2

Desmet, Ortuño-Ortín and Wacziarg (AER, 2017) study the empirical relationship between ethnicity and culture, defined as a vector of traits reflecting norms, values, and attitudes. Then, they determine whether the relationship between ethnicity and culture is important to understand salient political economy outcomes, such as civil conflict or public goods provision.

For public good provision they estimate the following model:

$$P_i = \gamma_0 + \gamma_1 C_i + \gamma_2 E_i + \gamma_3 O_i + \boldsymbol{\gamma}_4 Z_i + u_i \tag{4}$$

where P_i is a standardised measure of provision of public goods; C_i is a standardised measure of cultural fractionalisation in country i (the greater it is, the more fractionalised culture is); E_i is a standardised measure of ethnolinguistic fractionalisation; O_i is a standardised measure of the overlap between culture and ethnicity (the greater it is,

the greater the overlap); $\mathbf{Z}_i$ is a vector of controls, including regional dummies (Latin America and Caribbean; sub-Saharan Africa; East and Southeast Asia), legal origins dummies (French legal origin, German legal origin, Scandinavian legal origin), and absolute latitude; u_i is the error term.

You are provided with a part of the dataset ps2_data2_group##.dta (sent to each group by email) the authors used for their study.

Regression Analysis

1. Estimate three univariate regressions, one with each standardised covariate (C_i, E_i and O_i), through OLS. Comment your results. [Hint: *Recall that the dependent variable and the three measures are standardised*]

2. Estimate the multivariate Model 4 through OLS. Then, re-estimate using heteroskedasticity robust standard error. Comment your results.

3. Decide trough a suitable test if the joint significance of the three standardised measures in the model estimated in the previous point.

4. Present brief evidence for/against the presence of heteroskedastic errors in your regression.

5. "Huber-White robust standard error should always be used to preserve the validity of your inference in case there is heteroskedasticity". Defend this position against using classical OLS standard errors, with reference to consistency of the estimator and validity of the t-test.

6. Report the results from all the regressions that you performed from points 2-3 in one table that includes a) the estimated coefficients, b) standard errors of the estimators, c) the R-squared, and d) the F-statistic for a test of overall significance of the regression. [Hint: *Use the outreg or esttab command]*

References

[1] Sapkota, P., & Bastola, U. (2017). Foreign direct investment, income, and environmental pollution in developing countries: Panel data analysis of Latin America. Energy Economics, 64, 206-212.

[2] Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2017). Culture, ethnicity, and diversity. American Economic Review, 107(9), 2479-2513.