Problem Set 1 - PANEL

Professor: Chiara Monfardini;
Teaching assistant: Vito Stefano Bramante

- This problem set was sent on **February 24, 2023**
- Due date: **March 3, 2023** at **midnight**
- Submit your assignment on the virtuale page of the Ecometrics 2 course, in the "Assignment 1" dedicated section, no later than **midnight**. Assignments sent after the deadline will not be graded (no exceptions).
- You must attach a single zip file containing: (i) a pdf answer sheet; (ii) the Stata log-file; (iii) the Stata do-file. The pdf file should be **no longer than 5 pages (w/out tables).**
- Each table and graph in the pdf file should be fully reproducible. **By simply running the .do file** I should be able to reproduce the **exact** table or graph you are showing in the pdf, including title, variable names and numbers. **In the .do file you have to signal clearly which chunk of code reproduces which table or which chart.** Doing so will guarantee you up to 2 bonus points on each problem set.
- Name the zip file (and each file) as *surname*1_*surname*2_*surname*3.*zip*; remember to write the name, surname and student number of each student in the answer sheet.
- The grade for the 3 assignments will be 40% of the final grade.
- Please follow carefully the instructions detailed above. **Any misconduct will negatively impact the grading of the assignment.**

Climate change has been shown to affect many aspects of human life and to interact with several phenomena on a global scale. Human migration is, albeit with different sentiments, at the forefront of the current political agenda. In a 2017 paper Anouch Missirian and Wolfram Schlenker explore the relationship between temperature and asylum applications toward EU countries. Their analysis points toward two important findings:

1. **the number of applications of individuals coming from countries that suffer deviations from moderate optimum temperature increase;**
2. **the increase is nonlinear, suggesting an accelerating trend of applications as temperatures deviate more and more from the optimum.**

In this problem set you're asked to replicate these findings and to understand and discuss the reasons behind essential choices that the authors have made. To do so you'll be asked to use an expanded version of the dataset they use; at your disposal you'll find a panel dataset that includes 160 Non-EU source countries for the years 2000-2019. The main dependent variable, `asylums`, contains the number of asylum requests submitted by individuals coming from of the source countries, to the European Union. The main independent variable, `temperature` contains, for each of the source countries and for each year, the yearly average temperature. The authors' main claim, instead, is that unexpected fluctuations in asylum applications, also refferred to as "anomalies", can be explained by corresponding unexpected temperature fluctuations, bringing temperatures as either one of the main determinants of migration decision or as an important stressor to war-torn countries.

In part 1 you'll be asked to define the panel and to explore the data at hand via descriptive statistics. In the second part you'll be asked to run the main specifications and to show the non-linear relationship between temperature and applications.

Part 1: Data Exploration

In this first part we will take a closer look to our variables of interest. This part also allows us to play around with the data and gain more confidence.

1. Load the dataset into Stata and define a panel of countries for the available years. Describe the data: how many countries do you see? how many years? Is the panel balanced?
2. Tabulate the countries and explore which are the countries with very few observations over time. Then, drop from the data all the countries with (strictly) less than 18 observations [1]. From now on you can treat the panel as balanced.
3. For each country compute:
   - the average number of asylum applications
   - the average temperature
   - the standard deviation of asylum applications
   - the standard deviation of temperature

   You can compute these statistics either by collapsing the dataset or by creating new columns; feel free to use the method you prefer. Which are the countries with the highest average number of applicants toward the EU?
4. In a scatterplot called **Graph 1**, place on the x-axis the country-specific standard deviation temperature and on the y-axis the country-specific average of asylum applications (not logged at this stage), sorting on the temperature. Describe the plot. Which are the countries with high standard deviation and high asylum applications? [2] Could the presence of war be a confounding factor?
5. Answer the following questions (it is sufficient to read the first two pages of the paper to do so):
   (a) Explain the following sentence the author use "In other words, we link anomalies in log applications to weather anomalies once common annual shocks are absorbed (e.g., the global financial crisis in 2008)". (Hint: Think about the double-demeaning interpretation)
   (b) Why do the researchers express the asylum applications in log?
   (c) Why do the researchers use a quadratic specification for the temperature?
6. Use the command `xtsum` to obtain summary statistics for the main variable of interest: asylum applications, temperature, rain, minor conflicts and major conflicts. Describe the within and between variation for asylum applications and temperature.

Part 2: Analysis

All the results of the four regression you'll run in Part 2 will have to be shown in a single table, **Table 1**. The table has to report the coefficients, the standard errors and the significance level for each specification. For readability, omit the country-specific and the year-specific coefficients. You do not have to report the output of each regression on your answer sheet, but it has to be in the log file.

1. Run a POLS regression of the log of asylum applications on temperature. Write down the equation that reflects the estimated regression minding the correct use of the indices. Comment the results [3]. Using the command `margins` predict the *log_asylum* at integer values of temperature ranging from 0 to 40 at a 1 degree interval. Plot the predictions using the command `marginsplot`. Please be careful in labeling everything correctly. Briefly describe the plot.
2. Run a POLS regression of the log of asylum applications on temperature and temperature squared. Write down the equation that reflects the estimated regression minding the correct use of the indices. Comment the results. Using the command `margins` predict the *log_asylum* at integer values of temperature ranging from zero to forty at a 1-degree interval. Plot the predictions using the command `marginsplot` [4]. Please be careful in labeling everything correctly. Briefly describe the plot. Is it different from the one of the previous specification?

[1] or alternatively keep only the countries having 18 observations or more. To speed up the process you can look at the command `bysort` combined with the command `egen`

[2] Remember to plot the labels otherwise it will be hard to interpret the graph

[3] Remember you're using a *log-level* specification

[4] When you run this regression you have two options to introduce the square. You can create a new variable by computing the square of temperature or you can, in the `reg` command, use `c.temp##c.temp`. The second option tells Stata to take the interaction of the same variable, hence, squaring it. Try both ways and compare the charts you obtain, to convince yourself that the second option should better suit what you're after.

3. Using the `xtreg` command run a fixed-effect regression of the log of asylum applications on temperature and temperature squared. Comment the coefficients and compare them with the ones obtained at the previous point.

4. Using the `xtreg` command to run a two-way fixed-effect regression of the log of asylum applications on temperature, temperature squared, rain precipitation, rain precipitation squared, minor conflict and major conflict. Comment the coefficients of the new variables introduced. Is the effect of temperatures consistent with previous specifications? Do precipitation matter? Does war matter?

5. Using the command `margins` predict the *log_asylum* at integer values of temperature ranging from zero to forty at a 1-degree interval. Plot the predictions using the command `marginsplot`. Please be careful in labeling everything correctly. Briefly describe the plot. Find and report the optimal temperature [5].

Codebook

The dataset contains a mixture of individual-level survey data and subdistrict-level data. The variables are:

- **year**: the year to which the row refers to
- **country**: string containing the name of the country
- **continent**: string containing the name of the continent
- **asylums**: absolute number of asylum applicants to a EU country
- **temp**: average yearly temperature
- **rain**: total rainfall in the country expressed in meters
- **minor conflict**: dummy containing 1 if country is involved in a minor conflict, 0 otherwise
- **major conflict**: dummy containing 1 if country is involved in a major conflict, 0 otherwise

[5] the **optimal temperature** is the temperature at which we observe the minimum number of asylum applications, i.e. the lowest predicted value of log_asylums which corresponds to the minimum point of the curve shown in marginsplot.