**ECON 453**

Fall 2023

Problem Set 4 – 36 points

Submit by end of day Monday December 11th

Please download the gretl session file “PS4 Session” from Canvas. This is a panel dataset created using the World Bank Development Indicators (<https://data.worldbank.org/indicator>). The panel consists of information on 177 countries from 1999 to 2020. For the United States, the dataset extends back to 1960. The codes for the text variables (country/continent/region) are in the icon view in a text file. The data file has a brief description for each variable, but please let me know if you have any questions as you work with the data.

You will submit one document (Word or PDF) for this problem set. Please copy/paste the relevant regression results or graphs into your document, then add your discussions.

1. (7 points) We will begin with a cross-sectional analysis. Please restrict the sample to include **only the year 2020**.

* Run a regression with fertility rates as the dependent variable and the following independent variables: rural population %, GDP per capita, female labor force population rate, and the % of the population ages 65 and over.
* Report your results and discuss the findings. Interpret at least one of the coefficients numerically. Do the findings make sense? Are they practically significant? How strong is the model, overall?
* Run another regression that uses the same variables as your first regression, but that also includes some dummy variables for continents or regions. Discuss your decision of how/which to include. Present the model that you feel is the best/final choice for the set of dummies to include.
* Discuss what happened to the strength of the model. Summarize what we learn about how the continents/regions differ. Then discuss how (if at all) the coefficients on the explanatory variables changed after you included the dummy variables.

**For Questions 2 and 3, we will work with time-series analysis for the U.S. First, restore the dataset to the full sample (after your cross-sectional analysis). Then, please restrict the sample to only the United States. You should have 61 observations in the restricted dataset.**

1. (6 points) Time-series exploration.

* Create a set of time-series plots for the following variables: fertility rates, GDP per capita, rural population %, and the % of the population ages 65 and over. These should be in 4 different plots (use the **View -> Multiple Graphs** option in gretl).
* Comment on the graphs and what these tell us about the potential relationship between each of these variables and fertility rates in the U.S.
* Run a static model where you use fertility rates as the dependent variable and the other 3 variables (rural pop, GDP, % over 65) as regressors.
* Present your results and discuss your findings. Summarize the strength of the model, whether the estimated coefficients match your expectations, and whether there are any indications of problems in using the static model.

1. (7 points) Time trend analysis

* Run the same regression that you did in question 2 but this time include a linear time trend variable.
* Present your results and discuss: (1) the interpretation of the time trend coefficient, (2) how the estimated coefficients on the other variables changed (from question 2) and, (3) whether this is an improvement to our model.
* Did the trend change during the period examined?
  + Create a dummy variable for the period starting in 1990 (basically, the second half of the dataset
  + Run a regression where the dependent variable is the fertility rate and the explanatory variables are *only:* the time trend, the “since 1990” dummy, and the interaction of these two variables.
  + Report the results. Discuss what we learn about the trends in fertility rates during the period of 1960 to 2020. As part of this, create two equations: one that shows how the fertility rate changes with time between 1960 and 1989, and another that show how it changes with time between 1990 and 2020.

**For Questions 4 and 5, we will work with the panel dataset. First, restore the dataset to the full sample (after your time-series analysis). Then, restrict the sample to years 1999 and forward. Next, add the time trend variable again (it likely was created when you were restricted to the time-series dataset).**

1. (5 points) Run a fixed effects model that uses fertility rates as the dependent variable and the following five independent variables: rural population %, GDP per capita, female labor force population rate, the % of the population ages 65 and over, and the time trend variable.

* Report your results. Summarize what the R-squared statistics tell us about the explanatory power of the model.
* Explain how the interpretation of the coefficients differs in nature from those in question 1 or question 3.
* Discuss the results and whether the estimated coefficients make sense.

1. (7 points) We are going to restrict the sample by continents for this one.

* First, restrict to the continent of Asia. Run the same regression as in question 4. Save these results to the session as an icon. Restore the data to the full sample.
* Next, restrict to the continent of Europe. Run the same regression as in question 4. Save these results to the session as an icon. Restore the full sample.
* Finally, restrict to the continent of Africa. Run the same regression as in question 4. Save these results to the session as an icon.
* Create a model table (in the icon view of gretl) that displays the results from each of the continents (Asia, Europe, Africa) in one table.
* Compare and discuss the results by continent. Do the differences in estimated coefficients make sense? Are there any major differences? Does the predictive power of your model differ by continent?

1. (4 points) Summarize what we should conclude from our exploration of the factors influencing changes in fertility rates in the world. Do you have confidence in declaring the estimated impact of any of the explanatory variables on fertility? If so, which? Overall, which approach to analyzing the data do you think provides the most useful insight for this particular dataset? Explain your reasoning.