ECONOMICS 172: Issues in African Economic Development
Problem Set 2

Due Thursday April 22, 2021 **by 2 pm** on bCourses
[Please turn in your own solutions – do not solve in teams]

1. Slavery and African Economic Development [3 points]

Based on the evidence discussed in the course readings and lecture, are African countries with greater historical exposure to the slave trade richer or poorer than other African countries today? Describe the data and statistical analysis underlying this evidence, mention the magnitude of the impacts, and discuss at least two mechanisms that you believe are most likely to be driving these effects.
Strictly enforced: your answer should be no more than 2 pages, with 12-point font and double-spacing; we will not read past page 2 and you will be penalized for using different font, spacing or margins.

2. Weather and Non-Witch Killing [7 points]

This question builds on the analysis in the poverty and violence paper on the reading list (Miguel 2005). In particular, you will carry out econometric analysis related to the paper.

As was the case for Problem Sets 1 and 2, we will examine whether you have followed best practices in your R script. For instance, you should include detailed comments in your code for easier reproducibility. You may choose to produce an RMarkdown file that integrates your code (set echo=TRUE), your written responses, and tables displaying your regression results into a single PDF (please "knit to PDF" and turn in the resulting PDF file; do not simply turn in your .rmd file). Alternatively, you may choose to write up your answers using a word processor, include copies/screenshots of your tables (note that you may find it helpful to use the type="text" option of the stargazer command in this case) and attach a copy of your R script at the end. In either case, be sure that your submission includes, in some format, (1) your entire code/R script, (2) your written answers, and (3) your regression output. Please merge all documents into a single PDF before submitting.

First, please download the data to be used in the analysis. There are two ways to do this.

You can click on the following link to copy the dataset for this exercise into your repository on U.C. Berkeley's DataHub. RStudio should automatically open, and you should see a new folder in your repository called 'HW2':
https://r.datahub.berkeley.edu/hub/user-redirect/git-pull?repo=https%3A%2F%2Fgithub.com%2Fds-modules%2FEcon-172-Sp21&urlpath=rstudio%2F&branch=main

Alternatively, you can download "Econ172_Spring 2021_ProblemSet3_data.csv" from the bCourses page. Once you have downloaded the data using either method, use the "read.csv" command to open it in R (or RStudio), either on your local computer, or on U.C. Berkeley's DataHub, found on https://r.datahub.berkeley.edu/.

The dataset "Econ172_Spring 2021_ProblemSet3_data.csv" is a partial extract of data from the actual project, where each observation (row) in the dataset represents one village (denoted by "vid") in Meatu district, Tanzania observed during one year ("year"), with years ranging from 1992 to 2002. In other words, this is panel data. The data have also been modified somewhat from the actual data.

Variables in the dataset include: "witch_murders" (the number of witch murders that occurred in a particular village in a particular year), "oth_murders" (the number of non-witch murders), "any_rain" (an indicator variable for whether either a drought or flood occurred in a particular village in a particular year), "any_disease" (an indicator for whether a disease outbreak occurred, e.g., measles, cholera, etc.), "famine" (an indicator for whether there was an extreme shortage of food in a particular village in a particular year), "educat" (average years of schooling in the village, as measured in 2002), and "trad_relig" (the proportion of households in the village practicing traditional religions, also in 2002).

a) Start by constructing a variable for the total number of murders in a village in a given year, namely, witch murders plus non-witch murders. Then, create a table of summary statistics, including the mean, standard deviation, minimum value, maximum value, and number of observations, using the descriptive commands in R you have learned in section and used in previous problem sets. For example, "mean(violence$any_rain)" where "violence" is what we called our data set and we care about computing the mean of the "any_rain" variable. (You can also use stargazer to generate summary statistics table; see ??stargazer for details.). Do this for all variables in the dataset and discuss any noteworthy patterns, paying particular attention to the murder and rainfall variables. [1 point]

b) Now consider the causes of the non-witch murders in the village ("oth_murders")., Regress the number of non-witch murders (in a village in a particular year) on the indicator for whether a drought or flood occurred in that year. Interpret the relationship. Make sure that error terms are allowed to be correlated ("clustered") across years for the same village using the lfe package and the felm command, as discussed in section. Remember that to cluster standard errors, you run "felm(y~x1+x2 |0|0|j, data=violence)", where j is the variable by which you want to cluster and "violence" is the name of your dataset (see ??felm for details).

In a second regression, also include average years of schooling and the proportion of households practicing traditional religions as additional explanatory variables in the regression (while still clustering the error terms at the village level), and again present and interpret the results. [1 point]

c) Next consider the effect of disease epidemics on the number of non-witch number of murders in the village. Run the same two regressions as in part "b" but replace the indicator for whether a drought or flood occurred with the indicator for whether a disease outbreak occurred. Present the results, and contrast them to the results for rainfall shocks. [1 point]

d) Create a figure that has the year on the x-axis (horizontal axis) and the average number of non-witch murders per village (in that year) on the y-axis. Also present the proportion of villages with either drought or flood (again using the variable "any_rain") on the y-axis in the same figure. What patterns become apparent in this figure, if any? You can use the command "summaryBy" in R to create variable averages by year, and then the "plot" command to display the data graphically. Remember that to use the summaryBy command to create a new averaged data set over some variable a from the dataset, you can write:
mydata_avg <- summaryBy(b + c + d ~ a, data=mydata)
where b, c and d are the variables that you want to take the averages of. Describe the main patterns in the figure. Can we interpret these patterns as causal relationships? Explain why or why not? [1 point]

e) Now consider a possible instrumental variables (IV) approach. Economic theory suggests that extreme economic hardship, such as an acute food shortage ("famine") may be associated with more violence, including murders. Famine may be caused by extreme rainfall. Applying the IV example given in Lecture 18 to this context, use equations to write out the first stage regression, the second stage regression, and the reduced form regression. [1 point]

f) In your view, is this likely to be a valid instrumental variables approach? Discuss the plausibility of each of the three IV conditions laid out in Lecture 18 and in Chapter 3 of the Angrist and Pischke Mastering 'Metrics book. What are some specific ways in which each of these assumptions might fail in this context? In your opinion, which estimates are more informative, those from part b. above or the IV estimates proposed in part e., and why? [2 points]