PLSC 503: "Multivariate Analysis for Political Research"

Exercise One
January 27, 2020

Introduction

This homework is an opportunity to demonstrate your mastery of *bivariate* OLS regression. The exercise contains a simulation component and a substantive part, the topic of which is the relationship between political and economic factors and HIV/AIDS.

Exercise: Simulation

For the case of a bivariate regression $Y_i = \beta_0 + \beta_1 X_i + u_i$ where $u_i \sim$ i.i.d. $N(0, \sigma^2)$, show via simulation that:

1. the OLS estimators for $\hat{\beta}_0$ and $\hat{\beta}_1$ yield unbiased estimates of β_0 and β_1, respectively;

2. $\text{Var}(\hat{\beta}_1)$ is increasing in σ^2;

3. $\text{Var}(\hat{\beta}_1)$ is decreasing in N;

4. $R^2_{adj.} \rightarrow R^2$ as $N \rightarrow \infty$.

Hint: One simulated regression isn't sufficient to "show" anything.

Exercise: Data Analysis

The data are from 2007, and contain information on 111 nations in the world system. In addition to identifiers, they contain four variables, and reside in .csv format on the course GITHUB repo. The variables there are:

- `HIVPrevalence` is the percentage of the people aged 15 or over in the country currently infected with the HIV virus (source: WHO / UNAIDS).

- `HIVDeathRate` is the number of people *per 1000 infected* who died from complications due to HIV/AIDS in that country (source: WHO / UNAIDS).

- `GDPPerCap` is the country's gross domestic product per capita, in constant 2005 U.S. dollars.

- `CivilWarLag` is a dichotomous variable, coded 1 if the country experienced a civil conflict in 2006 (the year prior to the HIV/AIDS measures) and 0 if it did not.

Your assignment is to use OLS regression to examine the **(bivariate)** relationships among HIV prevalence or death rates and wealth or civil conflict, and to prepare a short report on that relationship as if you were writing up the results for a paper or journal submission. You should **choose one** of the two HIV/AIDS variables to be your "dependent" variable, and one of the other two as your "independent" variable / covariate. Your report should be in a submission-style format (i.e., no raw output) and contain *at least* the following:

1. A brief summary of the dependent and independent variables, including a graphical presentation of the relationships;

2. Results of your OLS estimate(s), including coefficients, standard errors, SEEs, R^2s, and other relevant statistics;

3. A discussion, in words, of those results/findings, including matters relating to marginal effects, statistical inference, and model fit.

This assignment is due Monday, February 3, 2020 at 5:00 p.m. EDT. You can submit your homework by emailing copies **both** to Dr. Zorn (zorn@psu.edu) <u>and</u> Mr. Kim (txk290@psu.edu). In addition to your responses to the items above, please include all code used to fit models, conduct diagnostics, generate plots, and so forth. This assignment is worth 50 possible points.