PLSC 503: "Multivariate Analysis for Political Research"

Exercise One
January 15, 2025

Introduction

This homework is (once again!) an opportunity to demonstrate your mastery of *bivariate* OLS regression. The exercise contains a simulation component and a substantive part.

Part I: 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. Var$(\hat{\beta}_1)$ is increasing in σ^2;
3. Var$(\hat{\beta}_1)$ is decreasing in N;
4. $R^2_{adj.} \to R^2$ as $N \to \infty$.

Hint: One simulated regression isn't sufficient to "show" anything; think in terms of turning (many) simulated coefficients into data, *then* doing analysis and visualization.

Part II: Data Analysis

The data are from 2019 (the last normal year before the COVID-19 pandemic), and are drawn from the *World Development Indicators* (WDI). The WDI are data collected on around 1,400 demographic, political, social, and economic indicators collected annually by the World Bank. Data are collected at the national level, for each of around 215 countries in the international system. The length of time each variable is collected varies significantly; the longest series extend back to 1960, while others are only available for very recent years or at discrete time points. (The World Bank also gathers and publishes WDI data on regional groupings, but for our purposes we'll focus on data at the national level.) Detailed information on the WDI is available at the WDI website.

For this exercise, we'll consider WDI data[1] on a total of 12 variables:

- `Population` - Population (in, like, people).
- `UrbanPopulation` - Urban Population (percent of total).
- `BirthRatePer1K` - Birth Rate (births per 1K people).

[1] The code to pull and clean these data from the World Bank website is available here.

- `FertilityRate` - Fertility Rate (births per woman).
- `PrimarySchoolAge` - Primary school starting age (years).
- `TotalTrade` - Total Trade (percent of GDP).
- `FDIIn` - Inward Foreign Direct Investment (FDI) (percent of GDP).
- `PublicEdExpend` - Public expenditure on education (percent of GDP).
- `PublicHealthExpend` - Public expenditure on health (percent of GDP).
- `WomenBusLawIndex` - Women Business & the Law Index score.
- `MilitaryExpenditures` - Military expenditures (percent of GDP).
- `CO2Emissions` - CO2 Emissions (metric tons per capita).

Your assignment is to use OLS regression to examine the **(bivariate)** relationship between two variables of your choosing. This part of the exercise is intentionally open-ended; it is up to you to decide what you want to be your "dependent" and "independent" variable. For this part of the exercise, you should include:

1. a brief summary of the dependent and independent variables, including a graphical presentation of the relationship;
2. results of your OLS estimate(s), including coefficients, standard errors, `SEE`s, 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 on or before Wednesday, January 29, 2025 at 11:59 p.m. ET. You can submit your homework by emailing copies **both** to Dr. Zorn (`zorn@psu.edu`) and Ms. Herlihy (`mth5492@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.