

POLS 602 Homework 2

Professor Paul Kellstedt 
Nicholas Ray, Supplemental Instructor (SI) 

2023-09-29

This homework assignment is due October 9th via email to the SI. The assignment is worth 100 points total, with each problem and its subsequent parts labeled with their respective worth. You are allowed to collaborate on this assignment and use online resources, so long as you turn in your own work and clearly attribute others' ideas. In your email, please include the .pdf file with your answers and any code used to produce your answers and the .pdf.

Problem 1 (60 points)

Multivariate Regression

This problem will draw on data from the “Manifesto Project” (Lehmann et al. 2023), which analyzes the platforms of political parties and measures their public stance on a variety of political issues. Suppose that a researcher theorizes that a party’s position on free trade (X) and cultural diversity (Z) will inform their position on civil rights (Y).

1.1 (30 points)

Download the aforementioned data- attached to the email with this assignment- titled “MPDataset_MPDS2023a.csv.” Focusing only on Sweden in 2018, estimate the model implied by the theorized relationship above in R. Present the output in a well-formatted table.

In the dataset, free trade is `per407`, cultural diversity is `per607`, and civil rights is `per201`. Consult the attached codebook if you would like more information.

1.2 (15 points)

Fully interpret the estimated coefficients and their statistical significance.

1.3 (15 points)

Given what has been discussed in class, are there any issues with the above estimation? For example, how may the sample size affect statistical inferences about the theorized coefficients? Additionally, should there be a concern with “where our values of X (and Z) come from” when using this estimation to evaluate the causal relationship of interest?

Problem 2 (40 points)

Omitted Variables Bias

2.1 (10 points)

Given class discussion, under what two conditions will omitting a variable from an estimated model induce bias?

2.2 (30 points)

This part of the problem will rely on simulation to try and show the effects of omitting a variable in such a scenario. Conduct a Monte Carlo experiment to empirically demonstrate that incorrectly omitting a variable will bias the estimated coefficients.

More specifically, simulate the following model:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + \varepsilon_i,$$

where $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 3$, $\varepsilon_i \sim N(0, 1)$, $(X_i, Z_i) \sim N(0, \begin{bmatrix} 1 & p \\ p & 1 \end{bmatrix})$, and $p \in \{0.1, 0.9\}$. For the purposes of this problem, specifying the distribution of (X_i, Z_i) in this manner is to allow for the manipulation of the covariance between these two independent variables (i.e., p). This may be done using the `MASS::mvrnorm()` function in R.

The point of your Monte Carlo experiment will be to vary p , setting it at 0.1 and 0.9, while holding all else constant to examine differences in the bias induced (or not) from omitting Z_i when the relationship between Z_i and X_i varies. You should fix your sample size to be 100 and run 100 trials for each variation you make to the experiment (i.e., the two values of p). Once complete, describe the simulation you conducted and state the results of omitting Z_i at differing levels of p .

Bibliography

Lehmann, Pola, Simon Franzmann, Tobias Burst, Sven Regel, Felicia Riethmuller, Andrea Volkens, Bernhard Webels, and Lisa Zehnter. 2023. “The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2023a.” <https://doi.org/10.25522/manifesto.mpds.2023a>.