POLS 602 Homework 6

Professor Paul Kellstedt
Nicholas Ray, Supplemental Instructor (SI)

2023-11-28

This homework assignment is due December 5th 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: Endogeneity (30 points)

A fundamental assumption of regression is that the disturbances do not covary with the regressors (i.e., the zero conditional mean assumption). Conduct a simulation that demonstrates the consequences of having various levels of correlation between the error term and a single regressor.

More specifically, simulate the data generating process (DGP) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ where $\beta_0 = 1$, $\beta_1 = 2$, and the correlation between x_i and ε_i ($\rho_{x\varepsilon}$) is in the set of -0.99, -0.5, 0, 0.5, and 0.99. You can use `MASS::mvrnorm()`, `faux::rnorm_multi()`, or another function to generate correlated variables. Run 500 trials for each level of correlation you induce and allow your DGP to have 1000 observations.

Describe your simulation and report bias in $\hat{\beta}_1$ as $\rho_{x\varepsilon}$ varies using either **a table or a graph**.

Problem 2: Instrumental Variables (70 points)

2.1 (15 points)

A popular method to accurately estimate parameters despite possible endogeneity is to use an instrumental variable (IV). What are the additional assumptions required by IV analysis?

2.2 (10 points)

Read in `data.csv`. Estimate the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ and display the summary output in your .pdf.

2.3 (25 points)

Now conduct two stage least squares (2SLS) by doing the following.

For the first stage, estimate the model $x_i = \gamma_0 + \gamma_1 z_i + u_i$ where z_i is an IV that meets the requirements for IV analysis and will be used to mitigate against possible endogeneity between y_i and x_i from 2.2.

For the second stage, estimate the model $y_i = \delta_0 + \delta_1 \hat{x}_i + v_i$ where $\hat{x}_i$ are the fitted values from the first stage. Display the summary output in your .pdf and briefly compare $\hat{\beta}_1$ (from 2.2) and $\hat{\delta}_1$. Does there appear to be endogeneity between x_i and y_i?

2.4 (20 points)

Finally, conduct 2SLS in one step by estimating a reduced-form equation which includes both stages. This can be done by using `ivreg::ivreg()`. Display the summary output in your .pdf and briefly comment on any differences between these results and those from 2.3.