PLSC 503: "Multivariate Analysis for Political Research"

Exercise Five
February 19, 2024

Part I

In class on February 19, we used a brief simulation to demonstrate that – with data meeting the assumptions they require – instrumental variable (IV) methods can be used to overcome the bias that results from correlation between predictors $\mathbf{X}$ and regression residuals $\mathbf{U}$. The first part of this exercise asks you to expand upon and extend that analysis, as follows:

1. Begin by repeating the simple analysis from the February 20 class session (slides 39-42)[1] many times, saving the relevant results (in particular, the estimates of $\hat{\beta}_1$, and its associated standard error), and show that the pattern we observed in class is a general one that holds across multiple simulations.

2. Next, repeat step (1), but vary the strength of the association / correlation between $\mathbf{X}$ and $\mathbf{Z}$, from a high of 0.8 down to a low of 0, in several steps. (This is, in essence, varying the "strength" / quality of the instrument.) Report on and discuss how the effectiveness of IV for addressing endogeneity bias changes as this association is varied.

3. Finally, repeat step (1) again, holding Cov($\mathbf{X}, \mathbf{Z}$) constant at 0.8 but increasing the covariation between $\mathbf{Z}$ and $\mathbf{U}$ from 0 to 0.8. (This is essentially violating the assumption that the instrument is uncorrelated with the error term.) Again, report on and discuss how the effectiveness of IV for addressing endogeneity bias changes as this association is varied.

Part II

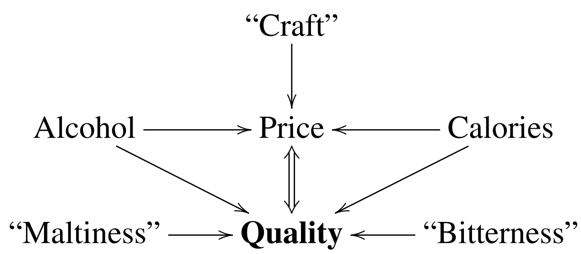
In the spirit of our upcoming spring break, we're taking a break from political science to focus on... beer. We'll use data are from a rating of 69 domestic and imported beers conducted way, way back in June, 1996 by *Consumer Reports* magazine. Our main dependent variable – *quality* – is rated on a 0-100 scale, with higher numbers being better. Explanatory variables are:

- *price* per six 12 ounce servings (in dollars),
- *calories* per 12 ounce serving,
- *alcohol* level (measured as ABV; the percentage of alcohol),

[1] In that example, we had $N = 500$, the association / correlation between the predictor $\mathbf{X}$ and the instrument $\mathbf{Z}$ was 0.8, the association between the predictor $\mathbf{X}$ and the residual $\mathbf{U}$ was 0.4, and the association between the instrument $\mathbf{Z}$ and the residual $\mathbf{U}$ was 0.

- an indicator for whether (=1) or not (=0) the beer was a "craft" beer (i.e., from a microbrewery), and
- two indicators of flavor: a scale (0-100) of the beer's level of *bitterness*, and a similar scale of its level of *maltiness*.

One issue with a standard model of quality is that price may well be endogenous: brewers sell higher quality beer at higher prices (quality affects price), but raters perceive more expensive beers as better (price affects quality). Suppose that after several long, possibly-beer-infused discussions with your graduate school colleagues, you arrive at a model of *quality* that looks like this:



Note that in this formulation, "craft" beers are *only* assumed to be rated as higher-quality because they are (presumably) higher-priced.

Your assignment is straightforward:

1. Start by estimating a single-equation OLS model of beer quality alone, based on the theorized relationships set forth in the figure above. Discuss your results.
2. Next, estimate an instrumental variables (2SLS) model, again comporting with the figure above. Talk about your justification for the choice of instrument(s). Discuss the results as well, with a particular focus on how they differ from the OLS-based findings.

This exercise is worth the usual 50 points, and is due by 11:59 p.m. ET on Wednesday, February 28, 2024. Please submit your exercise via email, in PDF format, to Christopher Zorn (zorn@psu.edu) and Nathan Morse (nam@psu.edu).