

PLSC 503: “Multivariate Analysis for Political Research”

Exercise Four

February 25, 2020

The *subject du jour* is (multi)collinearity.

Part I

Consider a model like:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

where Y , \mathbf{X} , and u meet all the usual assumptions of the classical linear regression model, $\text{corr}(X_1, X_2) \in (-1, 1)$ and $\text{corr}(X_1, X_3) = \text{corr}(X_2, X_3) = 0$. (In other words, there is a non-zero correlation between X_1 and X_2 , but no correlation among the predictors otherwise).

Using simulations, show:

1. The relationship between $\text{corr}(X_1, X_2)$ and $\widehat{\text{s.e.}(\hat{\beta}_1)}$ for $N = 10$,¹
2. How that relationship changes as $N \rightarrow \infty$, and
3. Similarly, the relationship between $\text{corr}(X_1, X_2)$ and $\widehat{\text{s.e.}(\hat{\beta}_3)}$.

Hint: To generate two normal variates that are correlated to a particular degree using R, check out the various `norm2d` commands in the `fMultivar` package.

Part II

As we are perennially reminded, governments experience both deficits and surpluses. What drives deficit spending by governments, and its opposite, government surpluses? That is the question behind the applied part of today’s homework.

To begin to answer this question, you will analyze data on 25 European countries for the years 1950-1994. The data consist of one observation for each country for each year – Austria in 1950, Austria in 1951, etc. – and are “unbalanced” (some countries have larger numbers of years than do others); but, for now, you need not concern yourself (much) with the time-series aspects of the data. In addition to indicators for country (conventionally indexed by i) and year (t), the data contain six variables (all measures are in constant U.S. dollars unless otherwise noted):

¹That is, as $\text{corr}(X_1, X_2)$ varies, what happens to $\widehat{\text{s.e.}(\hat{\beta}_1)}$?

- `exptorev` is a measure of deficits equal to $\frac{\text{Government Expenditures}_{it}}{\text{Government Revenues}_{it}}$; that is, it measures the amount of government spending relative to revenues. A value of 1.0 thus indicates a “balanced budget,” where expenditures equal revenues; values greater than 1.0 indicate deficit spending, while values less than 1.0 indicate surpluses. This is the “dependent” variable of interest.
- `lnpop` is the natural logarithm of that country’s population in year t .
- `lngnp` is the natural logarithm of the per capita GNP for that country in that year.
- `lnimports` is the natural log of per capita imports in that country in that year.
- `lnexports` is (surprise!) the natural log of per capita exports in that country in that year.
- `democracy` is the (by-now familiar) POLITY democracy index, ranging from -10 (fully autocratic) to 10 (fully democratic).

Macroeconomic theory suggests that deficit spending will be positively related to the first two variables, and may have either a positive or a negative relationship with the two trade indicators (Rudra and Haggard 2005). Moreover, some in political science (e.g. Lake and Baum 2001) suggest that democratic governments are better providers of social services for their populations than are autocracies; this suggests a positive association with `democracy` as well.

Please:

1. Estimate a regression model to examine the hypotheses suggested above, and briefly summarize the findings thereof.
2. Assess the extent of multicollinearity, both substantively (that is, *why* might we expect here to be problematic collinearity present?) and statistically, using whatever approaches you deem fit.
3. Discuss what, if any, measures you might take to ameliorate any problematic collinearity among the covariates. Address the potential problems/pitfalls of each approach. If you decide to undertake any of these, show and briefly discuss your work.

This exercise is due in electronic (PDF) form on Wednesday, March 4, 2020 by 5:00 p.m. ET and is worth the customary 50 points.