

PLSC 503: “Multivariate Analysis for Political Research”

Exercise Four

February 23, 2018

Part I

We’re going to investigate what effect “dummying out” a continuous variable has on our conclusions in a linear regression model. Consider a simple model like this:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (1)$$

where $u_i \sim N(0, 1)$ and $X \sim P\phi_0 + (1 - P)\phi_1$, where $P \in (0, 1)$ and ϕ_0 and ϕ_1 are normal densities with means of 0 and 1, respectively, and variances $\sigma_{0,1}^2$ as discussed below.¹ You’ll be considering models where X is replaced by D ,

$$Y_i = \gamma_0 + \gamma_1 D_i + e_i \quad (2)$$

where

$$D_i = \begin{cases} 0 & \text{if } X_i \leq 0 \\ 1 & \text{otherwise.} \end{cases}$$

Your assignment:

1. First, demonstrate (trivially) that for $\sigma_0^2 = \sigma_1^2 = 0$, the estimates from (1) and (2) are identical.
2. For $P = 0.5$, discuss the effect on parameter estimates ($\hat{\beta}_1$ vs. $\hat{\gamma}_1$, and, if relevant, $\hat{\beta}_0$ vs. $\hat{\gamma}_0$), standard errors, model fit, etc. of symmetrically increasing σ_0^2 and σ_1^2 (say, from $\sigma^2 = 0.000001$ to $\sigma^2 = 1$).
3. Similar to part 2, discuss what happens if, for a given value of $\sigma_{0,1}^2$ (say, $\sigma_{0,1}^2 = 0.01$), P is varied from 0 to 1 with D defined as above.

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):

- **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.

¹This is known as a *mixture distribution*, specifically a mixture of two normals. You’ll know you got it right if the histogram of the resulting X is bimodal for small values of $\sigma_{0,1}^2$ and values of P not close to 0 or 1. *Hint*: For any observation i , this implies $X_i = P_i\phi_{0i} + (1 - P_i)\phi_{1i}$, and **rbinom()** is a great way to draw binary values with $\Pr(P_i) = P$.

- `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. Variables like `exptorev` – that is, ratios of positive quantities – often exhibit strong positive skewness, and analysts frequently use logarithmic transformations on those variables as a result. Investigate if, how, and to what extent logging `exptorev` changes the interpretation of the relationships characterized in the model above.
3. A long-standing debate in comparative politics questions whether democracy is best conceptualized as a continuous variable, or as a categorical one (and in particular, a binary indicator). Examine whether and how changing the `democracy` variable to a binary indicator changes the findings and/or interpretation of the model results. In your view, which – continuous or binary – is the better operationalization of `democracy` here, and why? (Note that you need not examine this for both logged and unlogged values of `exptorev`.)

This homework is worth 50 points, and will be due electronically (in PDF format) at or before **5:00 p.m. EST on Friday, March 2, 2018**.