

PS 699: Problem Set 5--Multivariate Regression:

I. Let us assume that we concluded from our last problem set that the proportionality of electoral/parliamentary outcomes (the congruence between the distribution of votes across parties and the distribution of parliamentary seats across parties) as measured by LPROP was (linearly) positively related to the natural log of LMAG (*i.e.*, levels of LPROP relate linearly to proportionate values of LMAG). Now, suppose we consider two extensions of this simple proposition: a) that LPROP is similarly (proportionately) related to the number of electoral districts and b) that it is simply (linearly, in levels) related to the broad electoral-system type, *i.e.*, whether the system is proportional representation or plurality/majority. Theoretically, we consider (a) likely because disproportionality happens when some parties are disproportionately favored and others disadvantaged by some vote distribution in a district given the electoral rules. We might argue, then, that greater number of electoral districts would reduce the likelihood that any particular party(ies) were (dis)advantaged in every district. *I.e.*, more districts might enhance the proportionality national parliamentary election results because advantages in some districts would be more likely to be balanced by disadvantages others. Proposition (b) may arise if proportional representation *per se* increased proportionality beyond its effects through its general relation to district magnitude.

A. Suggest a linear multivariate regression model which you might use to evaluate these theoretical propositions and estimate these relationships empirically.

B. Estimate this regression by OLS (print your results in some appealing tabular form). (Assume for these purposes that we have come to consensus that Germany should be coded LMAG=260, LEFORM=0, and no German dummy variable is to be included.)

C. Interpret your results for the individual coefficients. *I.e.*, what do your coefficients mean? *I.e.*, in substantive terms, what do your coefficient estimates say about the direction and magnitude of these relationships of electoral-system features to proportionality? Report your estimated uncertainty about these coefficients in some fashion; *i.e.*, make some substantively meaningful statement about their standard errors, particular test statistics, statistical significance, and/or confidence intervals.

D. Applying the Wald test-logic and then the loss-of-fit test-logic, test the hypothesis that neither of the new regressors added significant explanatory power to our model from pset4.

E. Applying whichever test-logic you prefer, test the hypothesis that the whole model fails to offer significantly more explanatory power than merely knowing the sample mean of proportionality for this set of countries.

II. A political economist suggests to you some arguments regarding how governments' social-security-spending policies. Specifically, s/he says: "Of course, social-security spending (as a share of GDP) responds to unemployment and to the age distribution of the economy/polity. After all, social security is more or less exactly that: spending on unemployment insurance and old-age 'insurance'. However, social-security spending is also used as a political tool. For example, suppose that coalition governments find staying in power together more difficult the more parties in the coalition. Suppose further that social-security pension-spending help the parties maintain the coalition because sending larger pension checks to retirees earns all governing parties votes, and so it is one of few things on which they can all agree. Smaller coalitions and single-party governments agree on more things, and so social-security spending must share fiscal attention with other budgetary priorities. If all this, social-security spending should be positively related to the number of parties in government. Furthermore, likely, the greater the share of the population that votes, the more the poor will be represented among the voting pool, and so the more the government will spend on redistributive policies like social security, again, *ceteris paribus*." Leaving aside for the moment the varying dates to which our data refer (*i.e.*, assume all data refer to the appropriate period), and leaving aside whatever incompleteness in the logical elaboration of the above arguments we may perceive, I suggest estimating the following CNLR model by OLS regression as a means of empirically evaluating the above theorist's claims:

$$SSPENDG = \beta_c + \beta_a AGE + \beta_u UE + \beta_n NPGOV PW + \beta_v VPART + \varepsilon$$

- A) Estimate this model by OLS (print results in some appealing tabular form).
- B) a) Interpret substantively the estimates of the coefficients β_n and β_v .
- b) Evaluate the statistical evidence on the theorist's hypothesis regarding coalition size and SS spending.
- c) Evaluate the statistical evidence on the theorist's hypothesis regarding voter participation and SS spending.
- d) Using the evidence available from this regression, evaluate the statistical evidence on the theorist's hypothesis that SS spending, beyond its quasi-automatic responsiveness to unemployment rates and the age distribution, is used as a political tool.
- C) Test the hypothesis that, *ceteris paribus*, the effect on SS spending of a 1 percentage-point (p.p.) increase in the over-65 share of the population equals the effect on SS spending of a 1 p.p. increase in the unemployment rate.
- D) Suppose some country experiences a 1 p.p. increase in its population shares that is both over-65 and officially unemployed (this may technically impossible, given official unemployment definitions, but ignore that). What is you estimate of the change in SS spending in that country as a result? Give a 90% confidence interval for this estimate.
- E) (More Difficult) Find the 95% *joint* confidence area for your estimates of β_n and β_v .