

HOMEWORK 3
(due Tuesday, April 15)

1. The file "PS2030.cross-national democracy.dta" contains macro-level information on economic, political, and social variables for some 175 countries from (approximately) 2005. It was gathered by Pippa Norris from Harvard University. The codebook from the data set is posted in the Homework 3 section of the class Canvas site.

In this exercise you will examine the "determinants of democracy". You first need to create a dichotomous variable which equals 1 if the variable "democraticfreedom" equals 1, and 0 otherwise. Call this new variable "freedich".

a. You want to test whether the socio-economic variables HDI (the UN Human Development Index) and "diversesociety" (a dummy variable for whether the country is socially homogenous or heterogeneous) are related to "freedich". Why might estimating a linear probability model for these relationships be inappropriate? Discuss the non-linear logit and probit specifications for this model. What are the assumptions that underlie each model? (Include a discussion of the latent variable Y* in your answer for the probit). What is each model's equation for the probability that freedich=1?

b. Generate **either** the logit or probit estimates for your hypothesized model, produce a nicely-labeled table with the results, and interpret the coefficients, significance levels, and at least TWO relevant goodness of fit statistics.

c. Interpret the effect of both variables in the model on the probability that FREEDICH=1. Explain your reasoning for employing the specific kind of assessments that you are making (e.g. marginal versus discrete change, AME versus MEM, etc.). Which variable has the stronger effect and which the weaker, and on what evidence do you base your answer? Use "margins/marginsplot/mchange/listcoef" commands in Stata (or their R equivalents) as appropriate to estimate and illustrate the effects.

d. If you chose to estimate a logit model, interpret (and compare) the effects of each variable in terms of their changes in the odds that FREEDICH=1. If you chose to estimate a probit model, interpret (and compare) the effects of each variable on latent Y*.

d. Now add two "political" variables to the model: whether the country has a proportional representation system ("PR"), and whether it is a federal system ("fedtype"). Is it "better" than the socio-economic model you estimated in "a"? On what evidence do you base your answer? Which variables are most "important" now in predicting/accounting for "freedich"?

2. A skeptical colleague points out that the model is likely to be plagued by "endogeneity," specifically in the HDI→democracy relationship.

a. What does she mean by this, and why would that be a problem for what you have found in the previous question?

b. You want to estimate a model that controls for this possible endogeneity. Another colleague claims that the Norris data set has two potential "instruments" for HDI – "exports" and "debt". Do you think this colleague is correct on *a priori* grounds?

c. Now estimate two models using the variable "democratization" as a *continuous* measure of a country's level of democracy: 1) an OLS model with HDI and the other variables from question 1; and 2) an alternative model that has the same independent variables, but uses "exports" and "debt" as instruments for HDI. Interpret the estimates (especially for HDI) and summary statistics for the two models.

d. Finally, test whether the instruments satisfy the “exogeneity” assumption, and whether they are “strong enough” to warrant inclusion in the alternative model in the first place. What do these tests tell you (or not tell you) about the adequacy of the alternative model?

EXTRA CREDIT (5 POINTS). Use “Burkina.exercise3.dta”. This is survey data collected from approximately 1400 individuals in the African country of Burkina Faso in 2011, measuring support for violent extremism among ordinary individuals and the factors that predict it.

The main dependent variable for this question is a three-category variable VIO_REL, which are the responses to the question: “Some people think that using arms and violence against civilians in defense of their religion is justified. Other people believe that, no matter what the reason, this kind of violence is never justified. How about you? Do you personally feel that using arms and violence against civilians in defense of your religion can be often justified (3), sometimes justified (2) or never justified (1)?”

The main independent variables are: VIONEIGH, a measure of the extent to which there is violence in the respondent’s neighborhood (this is the average perception on a 1-5 scale of all individuals in the respondent’s area); MALE, YOUTH (under 30 years old), REAL_UNEMPLOY (whether the respondent is unemployed); TRUSTGV (1-3 scale of trust in government), and REL_SERVICE (respondent’s frequency of attending religious services).

- a. Estimate either an ordered logit or ordered probit model predicting VIO_REL with all the independent variables named in the preceding paragraph, produce a nicely labeled table with the model coefficients, standard errors, significance levels, and summary statistics LR and Pseudo-R-squared. Comment on the statistical significance of the individual coefficients and their associated test(s), and the overall fit of the model.
- b. Interpret the variables in terms of their probabilities (and/or odds) of responses being on category 1/2/3. Comment on the relative importance (including the substantive importance) of the independent variables.