Subject: Re: When weights are not supported
Posted by Reduced-For(u)m on Thu, 06 Nov 2014 05:09:13 GMT
View Forum Message <> Reply to Message

Yohannes,

I don't really understand the context of your research well enough to guarantee that any of these suggestions will be helpful, but here are some options (and some comments on the fixed/random effects question).

1 - A very simple solution would be to use a "linear probability model" instead of a logit/probit - meaning an OLS regression on a 0/1 outcome. In big samples where you are interested in the effect of some particular covariate, this should give you a result very similar to the probit (unless you have someth really rare outcome, in which case maybe a Poisson regression). These are easily weighted and can deal with survey effects in any way you choose.

2 - Dis-aggregation: You could estimate each country separately using any method you want, and then weight the coefficient estimates by population size to get a single, overall estimate - or just present the distribution of point estimates. This has a bit of a different interpretation than doing it all at once, and you are implicitly allowing each country to have its own (totally unconstrained relative to other countries) effect of each covariate. It will cost you power (efficiency). There may be a Bayesian hierarchical way to estimate this too all at once, but I don't know it, and I don't think Stata would do it.

3 - Country Fixed Effects. I believe the recommendation for country level FEs had to do with establishing a commonality among countries in "levels". That is - if you use random effects, your model will still identify the co-efficient of interest using both within-country variation and across-country variation. So if the levels of your Y or X variables are majorly different across countries (compared to within-country) you will be estimating you coefficient mostly on differences between countries that may or may not be reasonably comparable. Using country fixed-effects de-means everything so that only the difference from the country level mean will identify the coefficient. In certain cases, this would make the weighting problem less severe (suppose all your observations were from very high or very low level country's for your covariates and outcome - the across-country variation would be terrible, driven by lots of observations in the tails, but the within-country might still be OK).

4 - Depending on what your covariate of interest is, you will likely want to estimate your standard errors in ways that are far more conservative than those recommended for using one survey by the DHS. This depends a great deal on the particular empirical question you are asking, in particular on where the variation in your right-hand-side is (for instance, is it a response to some question in the DHS or some other data you are merging in). The paper linked below gives a good, if somewhat technical, overview of clustered standard errors and the problems you might face. I can offer better suggestions here if I understood your context a little better.

http://cameron.econ.ucdavis.edu/research/Cameron_Miller_Clus ter_Robust_October152013.pdf