
Subject: Re: When weights are not supported
Posted by [Yohannes](#) on Thu, 06 Nov 2014 23:39:41 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for the attention to my question and for the suggestions. Thank you also for the literature and the offer of assistance. Basically, what I am trying to do is look at the covariates of child well being in low and middle income countries in a multi-country context using primarily DHS data, but also supplemented by relevant national level indicators obtained from international databases, namely the World Bank. One of these child well being measures which I am trying to model is primary school attendance which is assumed to be a dichotomous outcome but can also take a count data form if level is taken in to account, and can be fitted using inflated or hurdle variants of Poisson or negative binomial models. I am also assuming that the response variable is affected by four set of factors: (1) child level characteristics (such as sex, age birth order etc); (2) household level characteristics (such as religion, parental education, and wealth status (I will come back to this to show how it is handled); (3) area level characteristics derived from the DHS data itself (such as urban rural residence, and other characteristics derived from the mean values for the strata in which the child lives); and (4) national level covariates (such as GDP, health expenditure per capita, external aid, indices of governance etc..) obtained from the World Bank. Given these, I set out to test two set of models, and in the first wanted to put all the four level variables in the main model and then add two random terms (without covariates) at country and strata level. In the second model, I wanted to keep only the child level and household level factors in the main model, and include the strata level and national level variables in the respective variance functions. For the wealth index variable, given that the DHS generated index is both country and survey specific, I went on and re-calculated the index using the data for all countries combined, and reclassified households based on this new index rather than the original measure developed by the DHS.

So coming back to point 1 of the suggestion:

The data I am using for the study has over 200000 observations (unweighted) and the variable of interest, child schooling, is also not a rare event as such, so I can revert to a "linear probability model" easily. But my challenge is the available multilevel commands (such as the one on STATA which I use extensively for my research) do not support weights for any type of response variable specification. I am yet to go through the attached publication perhaps it may have a way to resolve my problem (??)

Yes, I can also do my analysis on country by country basis but as you noted that will change the research question. Primarily, I will not be able to capture between country effects. Besides, I am not exactly sure how I will be able to capture country level indicators in such analysis.

On the fixed random effects: The countries in the analysis vary widely both in terms of the response variable and explanatory variables as well as in DHS sample sizes. And given that my interest is also to look at both within and between country differences I was thinking that a random effect model that has the ability to correct for sample weights may be the best way to go. And given that the multi-level commands do not allow this, and that the fixed effect approach only partially addresses what I would like to do I am somehow torn between approaches neither of which appear to be ideal.
