Subject: Data analysis using multiple countries different survey years
Posted by busi41 on Thu, 16 Oct 2014 11:15:13 GMT
View Forum Message <> Reply to Message

Hello

I am using a DHS dataset that merges the latest women's individual datasets from 31 different countries. It is therefore a cross-sectional dataset but with survey years ranging from 2004 to 2013 and DHS phases 4 to 6. I would like to use matching to find the best balance of measured covariates before running a regression, and will match using a mixture of individual variables such as the respondent's religion or educational attainment and country level variables such as GDP or Maternal Mortality ratio. I remember reading that cross-sectional datasets should use variables from one point in time, but this is not really the case here. So what's the best approach for choosing country level variables.
1. For example should I choose values from one year e.g. 2011 for all countries
2. Or values from one year before each interview year. Some countries have interviews in two years e.g. some women in the country were interviewed in 2010 and others in 2011 so GDP values would be from 2009 and 2010 depending on the year of the interview.
3. Alternatively, how would one handle lagging in this instance. For example, if I think the Maternal Mortality rate when they were 18 affects the outcome at the time of the survey, then would I use MM values for when each respondent was 18 years old? My age range is 15 to 36 so that would be GDP values going back 18 years for each of the 31 countries?

Also can you think of any issues I need to watch out for from the different phases?

Any ideas would be very welcome

thank you