Data definition of RCS Simulations

* hh.dta
  + hhid       Unique household ID
  + strata    Strata
  + urban    Urban/rural status [1,0]
  + cluster  Cluster
  + weight  Sampling weight
  + hhsize   Number of household members
  + ccons     Per capita consumption for reference and testing (already deflated so that poverty line can be applied) This must be xdurables + sum(fcons) + sum(nfcons)
  + xdurables            Per capita consumption flow from durables (if not needed, just create a variable =0)
  + mcat\_\* Set of categorical explanatory variables for a structural model. These should be variables like roof or floor materials, source of lighting, source of cooking fuel, gender/education/literacy/employment status… of household head, etc. The more you have, the better. Each category for each variable should have a reasonable number of observations. Thus, if there are categories with only very few observations, it’s better to merge them with other categories. None of those variables should have any missing values. If there are missings, it’s best if you can set them either to a default option or mode or something else that makes sense.
  + mcon\_\*               Set of continuous explanatory variables for the structural model, e.g. dependency ratio, number of children, number of adults, number of working adults, number of dependents, number of rooms, number of sleeping rooms, … None of those variables should have any missing values. If there are missings, it’s best if you can set them either to a default option or mode or something else that makes sense.
* fcons.dta
  + hhid (should match m:1 with hh.dta)
  + cluster
  + itemid
  + fcons per capita and deflated, so that it directly yields ccons as described above
* nfcons.dta
  + hhid (should match m:1 with hh.dta)
  + cluster
  + itemid
  + nfcons per capita and deflated, so that it directly yields ccons as described above

You can try running a regression on the log of ccons on i.mcat\_\* mcon\_\* i.strata urban and see whether you can reasonably high R2. Please also make sure in this regression that the number of observations equals to the number of records in the dataset (double-checking that you don’t have any missing data in the explanatory variables).