**How to import the India Agriculture and Climate Data Set into Stata**

The dataset is structured as follows:

1. Don’t be alarmed by the “.cfm” extension – this is a plaintext file
2. Each cfm file is a year’s worth of data, so “india65.cfm” is all the data for 1965, for instance
3. While the files “india50.cfm” through “india55.cfm” exist, they appear to contain little or no data; essentially the dataset begins in 1956 and finishes in 1987
4. The first file however, “india50.cfm”, contains a variables list, used below
5. From the file “india56.cfm” onward, each file contains a contains a continuous list of space-separated values; these are observations for 271 districts and 227 variables per year
6. Refer to the webpage <http://ipl.econ.duke.edu/dthomas/dev_data/datafiles/india_agric_climate.htm> for a codebook

**Importing the data:**

This is the way I went about it, which was a bit laborious – there are surely more efficient methods – but it works:

1. Open each cfm file using TextEdit (Mac) or similar
2. Having opened each cfm file, resave each one individually as a “.txt” file in a new folder
3. Open Stata, and execute the following command from the Do editor for each .txt file you’ve just created (changing the folder locations as necessary):

# delimit ;

infile ID YEAR AGGDPDF CODE POPDEN PRSEED HYVWHEAT HYVRICE HYVMAIZE HYVBAJRA HYVJOWAR AGLABOR CULTIVAT WAGE NCA GCA NIA GIA YWHEAT YRICE YSUGAR YMAIZE YPOTATO YGNUT YBARLEY YTOBAC YGRAM YTUR YRAGI YSESAMUM YRMSEED YBAJRA YCOTTON YJOWAR YOPULS YJUTE YSOY YSUNFLWR ROADS LITERACY FEBXT FEBNT DMS01 DMS02 DMS03 DMS04 DMS05 DMS06 DMS07 DMS08 DMS09 DMS10 DMS11 DMS12 DMS13 DMS14 DMS15 DMS16 DMS17 DMS18 DMS19 DMS20 DMS21 DMAQ3 DMAQ2 DMAQ1 DMSLP4 DMSLP567 DMPH4 DMPH5 DMPH6 DMPH7 DMPH8 DMSLP1 DMSLP2 DMSLP3 COSTAGLB COSTCULT COSTBULL COSTTRAC COSTNITR COSTP2O5 COSTK2O DMTS1 DMTS2 DMTS3 DMTS4 DMTS5 QSUGAR DAYS STATE DISTRICT QBULLOCK QTRACTOR PTRACTOR PUPBULL PBULLOCK QLABOR QNITRO QP2O5 QK2O PNITRO PP2O5 PK2O QLAND ROPUMP RPWPUMP UOPUMP UEPUMP UPWPUMP PHYVWHT PHYVRICE PHYVJOWR PHYVBAJR PHYVMAIZ IROADS AWHEAT QWHEAT ARICE QRICE ASUGAR AMAIZE QMAIZE APOTATO QPOTATO AGNUT QGNUT ABARLEY QBARLEY ATOBAC QTOBAC AGRAM QGRAM ATUR QTUR ARAGI QRAGI ASESAMUM QSESAMUM ARMSEED QRMSEED PWHEAT PRICE PSUGAR PMAIZE PPOTATO PGNUT PBARLEY PTOBAC PGRAM PTUR PRAGI PSESAMUM PRMSEED ABAJRA QBAJRA ACOTTON QCOTTON AJOWAR QJOWAR PJOWAR PBAJRA PCOTTON AOPULS QOPULS POPULS AJUTE QJUTE PJUTE ASOY QSOY ASUNFLWR QSUNFLWR PSOY PSUNFLWR str30 DISTNAME ALT str30 STATENAM DSEA LON LAT RNJAN RNFEB RNMAR RNAPR RNMAY RNJUN RNJUL RNAUG RNSEP RNOCT RNNOV RNDEC TNJAN TNFEB TNMAR TNAPR TNMAY TNJUN TNJUL TNAUG TNSEP TNOCT TNNOV TNDEC DMSTS1 DMSTS2 DMSTS3 DMSTS4 DMSTS5 TOTAREA AHYV QBULLHA QTRACHA DMXS01 DMSX01 DMXS02 DMSX02 DMXS03 DMXS04 DMXS05 DMXS06 DMXS07 DMXTS1 DMSXT1 DMXTS2 DMXTS3

using "/Users/garethnellis/Dropbox/Green Revolution/Data/Ag\_data\_text\_files/india56.txt", clear ;

/\*now save this as a dta file in a different folder\*/

save "/Users/garethnellis/Dropbox/Green Revolution/Data/Ag\_data\_stata\_files/1956.dta", replace

Repeat this for all the .txt files, simply by changing the date at the end of the final two commands (highlighted in red) and re-running the code. That’s it – you should now have all the files in .dta format. These can be appended using the “ap” command to make the full panel.

Gareth Nellis

March 2012