"Readme" File for the Estimation Procedures in

"Selection and Comparative Advantage in Technology Adoption", by Tavneet Suri

This "readme" file documents the estimation procedures and code used in the above paper. It includes a description of the estimation procedures used for each table in the paper and documents what code is included in this supplemental directory. Note that all the code is included as .txt files, please change the file extension appropriately for Gauss.

The data are unfortunately proprietary and cannot be shared. However, you can contact the Director of the policy group that collected the data to request access to the data at the following address and phone number:

Tegemeo Institute of Agricultural Policy and Development

Kindaruma Lane, Off Ngong Road

P.O. Box 20498 00200, Nairobi, Kenya

Telephone: +254-20-2717818/+254-720-895454

Also note that the version of the data used in this paper may be older than any currently available versions (additional periods have been added to the dataset since and there has been additional data cleaning done since). Please feel free contact me at tavneet.suri@gmail.com if you would like more details on how to request access to the data.

The only additional data, aside from the Tegemeo panel survey mentioned above, used in this paper is rainfall data. The rainfall data came from the Climate Prediction Center at the NOAA, part of the USAID/FEWS (Famine Early Warning System) Project. This data is available or can be requested directly from the CPC at http://www.cpc.noaa.gov/products/african_desk/. Tim Love (who is unfortunately no longer with this section of the NOAA) assisted in accessing and extracting the rainfall data for Kenya for me.

Below, I document how all the various tables were created. The code included is for many of the minimum distance estimation procedures. The code included is limited to only the specific versions described below (please feel free to contact me if you would like code for any of the additional robustness checks or descriptions of how exactly to run the robustness checks). The minimum distance code was written in Gauss and the rest of the analysis was performed in Stata (including generating the reduced form estimates for input into the minimum distance procedure). The data extract used imposed few restrictions on the original dataset, except to restrict the sample to a complete panel, to all observations having non-missing data and to households who grew maize across all three periods of the main survey (1997, 2000 and 2004). These restrictions combined eliminated only 6% of the panel sample.

There are a total of the following fifteen tables in the paper. The list below describes what each table is, how the estimates were computed, including the equation numbers from the paper that describe the relevant specifications (where the paper itself does not have equations describing the specifications, the tables have footnotes with detailed empirical specifications). For the cases that require the minimum distance estimation code, I described what code I have included in this directory.

- Tables 1, 2a, 2b, 2c and 2d all report summary statistics.
- Table 3a reports OLS and household level fixed effects regressions, all estimated in Stata (the controls for each column are described in detail in the footnotes to the table).
- Table 3b reports the Correlated Random Effects (CRE) Reduced Forms and Structural Estimates. The reduced form estimates come from the regressions described in the notes to the table (also see footnote 38 in the paper). The structural estimates come from a minimum distance procedure using the reduced form estimates – this code is enclosed (see the Gauss file called md_cre.txt).
- Table 4 reports the Heckit and Treatment Effect Estimates under non-random assignment (i.e. the average treatment effect, the treatment on the treated, the marginal treatment effect and the local average treatment effect), exactly as described in Heckman, Tobias and Vytlacil (2001), all estimated in Stata.

- Table 5 reports OLS regressions of the yields on the different histories of farmers' adoption histories (with and without controls for inputs, the notes to the table describe all the controls), all estimated in Stata.
- Table 6 reports OLS and household fixed effects regressions where the observable inputs are allowed to interact with the hybrid dummy, all estimated in Stata (the notes to the table show the exact specification).
- Table 7 reports the reduced form estimates for the baseline Comparative Advantage Correlated Random Coefficient (CRC) model, all estimated in Stata (see equations (32) and (33) in the paper as well as the notes to the table for the exact specifications of these reduced forms).
- Table 8a reports the structural estimates for the baseline Comparative Advantage Correlated Random Coefficient (CRC) model, calculated by minimum distance (equation (34) in the paper illustrates the minimum distance restrictions) and just for the optimum minimum distance (OMD) case[1]. These structural estimates come from using the reduced form estimates in Table 7. Note that these estimates are for the baseline model (i.e. the full sample). The code for this minimum distance procedure is enclosed (see the Gauss program md_crc.txt). The robustness check included in this table is where I drop two districts, in particular Kisumu and Siaya, with extremely high adult death (mostly likely due to high HIV prevalence rates). The code for this is not included, but is a simple variant of the basic CRC code.
- Table 8b reports the structural estimates for the joint sector Comparative Advantage CRC model (again just for the OMD case) – this code is identical to that used for Table 8a, except that the technology sector is defined not to be the use hybrid alone but the joint use of hybrid and fertilizer as described in Section 4.5.2 of the paper (note that the reduced forms

[1] Footnote 51 in the paper (as well as tables in the earlier working paper versions of the paper) illustrates that the equally weighted minimum distance (EWMD) and the diagonally weighted minimum distance (DWMD) estimates are not different from the OMD estimates. I have not included the code for these EWMD and DWMD estimates but please feel free to request the code. Note that the OMD code for the variance covariance matrices of the structural estimates is written so as to be easily adaptable to these cases of EWMD and DWMD.

also need to be re-estimated using this definition of sector and the coefficients from these new reduced forms used in the minimum distance estimation).

- Table 8c reports the structural estimates for the Comparative Advantage CRC model (only the OMD cases) for the model where both hybrid and fertilizer are endogenous. The reduced form estimates for this minimum distance problem are much more complex – they include the entire history of hybrid, the entire history of fertilizer and all the combinations of all possible interactions between the hybrid histories and fertilizer. The projection used to generate these reduced forms is given by equation 35 in the paper. Using these reduced form estimates. Gauss minimum distance code is used to estimate the structural parameters (see the md_crc_hf.txt file enclosed).
- Table 9 reports the observable correlates of the estimated thetas across the sample. These regressions are run in Stata. The thetas are estimated using the estimates from Table 8c and the projection that is in equation (35).