

The “Out of Africa” Hypothesis, Human Genetic Diversity, and Comparative Economic Development

By QUAMRUL ASHRAF AND ODED GALOR

INSTRUCTIONS TO THE USER FOR REPLICATION OF RESULTS

The accompanying packet contains all the data sets and program files that are necessary for replicating the regression results and scatter plots presented in the main text of the paper as well as in its Online Appendix. Note that the user may require Stata version 10.1 or higher in order to access the data sets and execute the program files. The accompanying data sets and program files are located in the “data” and “programs” folders respectively. Before executing any of the program files in Stata, the user will need to alter the complete directory path, referenced in the “cd” command at the top of each file, to point to the *root* directory containing the accompanying “data,” “programs,” and “results” folders on the user’s machine. Upon successful execution of any given program file in Stata, the corresponding results will be stored in the “results” folder.

Note that, when executing some of the accompanying program files in Stata, the user may encounter an error due to the fact that the program file in question is referencing a Stata “module” that is not included in the user’s installation of Stata. Examples include (but are not necessarily limited to) the **outreg2**, **ivreg2**, **wherext**, **center**, **pcorr2**, **rcspline**, **sppack**, **statsmat**, **outtable**, and **cortex** Stata “modules.” In cases of such error, the user should be able to install the missing “module” either by issuing the “ssc install <module name>” command in Stata’s “command” window or by issuing the “findit <module name>” command and then following the installation instructions provided on the resulting pop-up screen.

country.dta: This file contains the main country-level data set. All variables in the data set are associated with descriptive labels for the user’s benefit. The user may additionally want to refer to Section F of the Online Appendix of the paper for detailed variable definitions and sources.

ethnic.dta: This file contains the ethnicity-level data set corresponding to the HGDP-CEPH sample of 53 ethnic groups referenced in the paper. All variables in the data set are associated with descriptive labels for the user’s benefit.

ethnicpair.dta: This file contains the ethnicity pair-level data set corresponding to all unique ethnic group pairs in the HGDP-CEPH sample of 53 ethnic groups referenced in the paper. All variables in the data set are associated with descriptive labels for the user’s benefit.

table{1-9, A1-A2, A4-A6, D1-D14, D15-D16-[A,B], D17-D18}.do: Each of these is a Stata “do” file that generates the regression results presented in the correspondingly named table appearing in either the main text or the Online Appendix of the paper. With the exception of **table{D2-D3, D13}.do**, each of the other program files will generate correspondingly named results files in both tab-delimited plain text format and LaTeX format (with the latter containing a LaTeX table code snippet). In the case of **table{D2-D3}.do**, the program files will generate correspondingly named Stata log files capturing the program execution output from Stata’s “results” window. In the case of **tableD13.do**, the program file will not

only generate results files in both tab-delimited plain text and LaTeX formats but will also generate a Stata log file capturing the program execution output from Stata's "results" window. For this case, the plain text and LaTeX results files will contain only the standardized betas reported in the Online Appendix version of this table, but the log file may be used to view the partial correlation coefficients that are also reported in the Online Appendix version of the table.

gentables.do: This is a Stata "do" file that executes *all* the aforementioned "do" files, each associated with individual tables appearing in either the main text or the Online Appendix of the paper. This program file is provided for the user's benefit, permitting replication of all tables of regression results in a "one shot" execution. Note that, prior to executing this program file, the user will need to alter not only the directory path referenced in this particular file but also the paths referenced in all the aforementioned individual-table "do" files.

genfigures.do: This is a Stata "do" file that generates image files, in Windows Enhanced Metafile (emf) format, corresponding to all the scatter plots presented in either the main text or the Online Appendix of the paper. The resulting image files are named according to their figure numbering in the main text or Online Appendix of the paper.

gensummstats.do: This is a Stata "do" file that generates output files, each containing a LaTeX table code snippet, corresponding to the tables of descriptive statistics presented in Section G of the Online Appendix of the paper. The resulting output files are named according to their table numbering in the Online Appendix of the paper.

x_ols2.ado: This is a Stata "ado" file that is referenced by **table{D2-D3}.do** to estimate spatial standard errors for OLS regressions presented in Tables D.2 and D.3 in Section D of the Online Appendix of the paper. This program file builds on the original **x_ols.ado** of Timothy G. Conley (available freely on his website at http://economics.uwo.ca/faculty/conley/data_GMMWithCross_99/x_ols.ado). The code in this program file is identical to the original version except that it additionally prints out p-values of the coefficient estimates in the Stata "results" window. Prior to executing **table{D2-D3}.do**, the user will need to install this program file in a relevant location under the Stata root directory (for example, an appropriate folder might be: C:\Program Files\Stata10\ado\base\x) on the user's machine.

x_gmm2.ado: This is a Stata "ado" file that is referenced by **tableD3.do** to estimate coefficients and standard errors for spatial GMM regressions presented in Table D.3 in Section D of the Online Appendix of the paper. This program file builds on the original **x_gmm.ado** of Timothy G. Conley (available freely on his website at http://economics.uwo.ca/faculty/conley/data_GMMWithCross_99/x_gmm.ado). The code in this program file is identical to the original version except that it additionally prints out p-values of the coefficient estimates in the Stata "results" window. Prior to executing **tableD3.do**, the user will need to install this program file in a relevant location under the Stata root directory (for example, an appropriate folder might be: C:\Program Files\Stata10\ado\base\x) on the user's machine.