THIS IS A PDF VERSION OF THE FILE README.txt

README for "The Twin Instrument: Fertility and Human Capital Investment"
Sonia Bhalotra & Damian Clarke
Journal of the European Economic Association

This folder contains four sub-directories:
|_ data: contains all data used in the paper
|_ source: contains all source code used for replication of results
|_ results: tables and figures provided in the paper
|_ log: a log file produced when running each piece of source code

All results in the paper can be replicated by running the programs provided in the "source" sub-directory, simply changing one global variable in each file pointing to the location of this folder on your computer. Additionally, these files replicate the majority of results contained in online appendices. A small number of results from the online appendix are not included here given that they are simply descriptive results from auxiliary datasets. If you wish to generate the precise format of results from the paper, the tex files located in this directory can be compiled, providing the complete rendered output.

All source files are written for Stata version >=15.0.

Three main separate scripts are provided, replicating results using:

 (a) DHS data
 (b) NHIS data, and
 (c) NLSY data.

These are respectively named "analysisDHS.do", "analysisNHIS.do", and "analysisNLSY.do". To replicate all results, each of the scripts should be run from within the "source" sub-directory. These can be run in Stata, after changing the MAIN global at the top of each script. This should be changed to indicate the location of these replication materials on your computer. For example, if the replication materials are located in a folder called "/home/replication", the line in each script defining the MAIN global should be set as:

    global MAIN "/home/replication/"

No other changes are necessary. A number of external user-written ados are required, which can be installed from Stata's SSC. These are installed automatically if not available, provided that internet connectivity is available. If a required ado file is not installed and internet connectivity is not available, the script will fail to run until either the ado has been installed and the scripts are re-run, or internet connectivity is available and the script is re-run. Two slightly modified ado files are additionally provided based on ados from Stata's SSC. These are imperfectivTwins.ado (a modified version of imperfectiv.ado by Clarke and Matta), and plausexogTwins.ado (a modified version of plausexog.ado by Clarke). These are provided in the source directory, and must remain there for replication materials to run correctly.

Tables are all exported as tex files, and figures are all exported as eps files. If you desire to replicate the precise format of these tables and figures as presented in the paper, you can compile the tex file in the main directory using the following command:

```
pdflatex -shell-escape BhalotraClarke_TheTwinInstrument_JEEA.tex
```

These scripts have all been tested on a computer with 32GB of RAM using Stata MP. For a computer with a significantly smaller amount of RAM or Stata SE, the analysisDHS.do script may be somewhat slow to run.

Main Tables and Figures in the paper correspond to the following scripts:

> analysisDHS.do : Tables 2, 4, 5A, 6, 8, 9, 11A. Figures 1A, 3A, 4, 5.
> analysisNHIS.do: Tables 1, 3, 5B, 7, 11B. Figures 1B, 3B.
> analysisNLSY.do: Table 10. Figure 2.

Additionally, results documented in online Appendices are replicated in these three scripts. These can be found below (T refers to Table, F refers to Figure)

> analysisDHS.do : FA2, FA3, FA5b, FA7, FA11, FA13, TA4, TA6, TA9, TA10a, TA11a, TA13, TA16, TA18.
> analysisNHIS.do: FA9, FA10, FA12, FA14, TA5, TA7, TA8, TA10b, TA11b, TA14, TA17, TA19.
> analysisNLSY.do: FA6.

For queries regarding data generation or other replication details, please refer to the document dataSources.pdf also available in this directory, or contact Damian Clarke at

damian.clarke [AT] usach.cl

or

damian.clarke [AT] protonmail.com.