README

Estimating Intergenerational and Assortative Processes in Extended Family Data
Review of Economic Studies

M.D. Collado, I. Ortuño-Ortín and J. Stuhler
Contact: jstuhler@eco.uc3m.es

Overview

This replication package includes all the programs used for the calibrations, graphs, and tables of the article (and most of the tables and figures from its online appendix). Calibration exercises and the main results are performed in Mathematica. The sensitivity tests reported in Tables E1 and E2 are performed in Gauss. The necessary data with all the correlations used in these calibrations are provided in the file "*EmpiricalCorrelations_ColladoOrtunoStuhler.xlsx*". The replicator should expect around 50 hours to run the Mathematica codes (and 1-2 hours for the Gauss program).

Data Availability and Provenance Statements

- ### Statement about Rights

  I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

- ### License for Data

  All program codes are licensed under a Creative Commons/CC-BY-NC license.

Software requirements

- Mathematica (code was run with version 11.0)
- Gauss (code was last run with version 5.0)

Description of programs and instructions for replicators

- \Mathematica folder: The Mathematica notebooks in this folder generate all the results in the main body of the article. The Mathematica notebook "OneFactorModel_ColladoOrtunoStuhler.nb" generates the calibrations and most tables and (raw) figures related to the main model of the paper (the one-factor model). The necessary data with all the correlations used in these calibrations are provided in the file "EmpiricalCorrelations_ColladoOrtunoStuhler.xlsx". The Mathematica notebook "TwoFactorModel_ColladoOrtunoStuhler.nb" generates all the results related to the two-factor model developed in Section 5.2 of the article. These two notebooks are completely independent. The code was last run on a desktop 8-core Intel-7-10700K CPU @ 3.80GHz with 128 GB of RAM. Computation took around 50 hours.

- To run the Mathematica programs, you only need to import the Excel file "EmpiricalCorrelations_ColladoOrtunoStuhler.xlsx". In the introduction of each Notebook, it is clearly explained how to run each program. You can run each section separately or enter all the cells of the notebook at the beginning (each section will run sequentially). The title of each section shows which tables or figures of the article are generated. The results of each section appear as an output on the screen, and it has also been indicated how to export them as excel files with the name of the corresponding figure in the article. Each minimization problem has its random seed specified, therefore the result obtained must always be the same (unless some other parameter of the program is changed).
- \Gauss folder: The GAUSS program "TablesE1_E2.prg" generates Tables E1 and E2 in the appendix. This program is independent of the other programs. We used Gauss in this case as it runs faster than performing the same exercise in Mathematica. To run the Gauss program, you need to import the Gauss data file "EmpiricalCorrelations_ColladoOrtunoStuhler.dat" and ".dht". The program includes comments explaining what is done in each part.
- \Figures folder: The Mathematica program "Plot_Figures_Final.nb" uses the fitted models (as generated by the programs "OneFactorModel_ColladoOrtunoStuhler.nb" and "TwoFactorModel_ColladoOrtunoStuhler.nb") to produce the formatted figures as shown in the article. For convenience, the folder also contains the fitted models as separate Excel files, so that the figures can be produced without running the codes in the \Mathematica folder.
- \Tables folder: The Excel file Tables_Final.xlsx contains the formatted tables as shown in the manuscript (i.e., the formatted version of the tables produced by the Mathematica code) as well as Figure 1.

List of tables, figures and programs

The following table indicates where the code to generate the tables and figures can be found. Abbreviations used in the table:

MathematicaNB1="OneFactorModel_ColladoOrtunoStuhler.nb"
MathematicaNB2="TwoFactorModel_ColladoOrtunoStuhler.nb"
PlotFigures="Plot_Figures_Final.nb"

The tables and the data to generate the figures are in Excel files.

Figure/Table	Program/Source	Section	Note
Table 1	Tables_Final.xlsx		Requires confidential data
Table 2	Tables_Final.xlsx		
Tables 3 and 4	MathematicaNB1	6	
Table 5	MathematicaNB2	3	

Tables E1 and E.2	TablesE1_E2.prg		
Table G.1	Tables_Final.xlsx		Requires confidential data
Tables G.2 and G.3	MathematicaNB1	12	
Tables G4 and G.5	MathematicaNB1	13	
Table G.6	MathematicaNB1	14	
Table H.1	TableH1.do		Requires confidential data
Tables H.2 and H.3	MathematicaNB1	18	
Figure 1	Tables_Final.xlsx		
Figure 2	MathematicaNB1	6	
Figure 3	MathematicaNB1	7	
Figure 4a	MathematicaNB1	8	
Figure 4b	MathematicaNB1	9	
Figure 4c	MathematicaNB1	10	
Figure 4d	MathematicaNB1	11	
Figure 5	Plot_Figures		
Figure 6a	MathematicaNB1	15, 16	
Figure 6b	MathematicaNB1	17	
Figure E.1	MathematicaNB1	19	
Figure F.1	Not provided		Requires confidential data
Figure F.2	Not provided		Requires confidential data
Figure F.3	Not provided		Requires confidential data
Figure G.1	MathematicaNB1	12	
Figure G.2a	MathematicaNB1	13	
Figure G.2b	MathematicaNB1	14	
Figure H.1	MathematicaNB1	18	

Details on confidential microdata sources

The file "EmpiricalCorrelations_ColladoOrtunoStuhler.xlsx" contains the correlations needed to replicate all the calibration exercises, tables, and figures in the paper. These correlations were calculated by the authors from the two databases listed below:

- Data on education, gender, age, income, height, residence, and parent and children's identifiers for individuals in Sweden were obtained from Swedish Multigenerational Registers provided by Statistics Sweden, hosted by the Swedish Institute for Social Research. This data is confidential but may be obtained with Data Use Agreements from Statistics Sweden. Researchers interested in access to the data may visit https://www.scb.se/en/services/guidance-for-researchers-and-universities/. It can take some months to negotiate data use agreements and it may be required that the applicant belongs to a Swedish research institution. The author J. Stuhler will assist with any reasonable replication attempts.
- Data on the name, gender, education, age, province of birth, household and spouse identifiers for individuals in the Spanish region of Cantabria were obtained from the *Instituto Cántabro de Estadística* (ICE). This data is confidential but may be obtained with Data Use Agreements with the ICE or the *Spanish Instituto Nacional de Estadistica*. Researchers interested in access to the data may visit https://www.ine.es/infoine/. It can take some months to negotiate data use agreements and gain access to the data. The author M. D. Collado will assist with any reasonable replication attempts.