

Data and Code for: Wealth, Marriage, and Sex Selection

Two mechanisms have been proposed to explain sex selection in India: son preference in which parents desire a male heir and daughter aversion in which dowry payments make parents worse off with girls. Our model incorporates both mechanisms, providing micro-foundations, based on the organization of the marriage institution, for daughter aversion. Marital matching, sex selection, and dowries are jointly determined in the model, whose implications are tested on a representative sample of rural households. Simulations of the model indicate that existing policies targeting daughter aversion might exacerbate the problem, while identifying other policies that could be effective.

Code

Code is organized into the `\code` folder. The main code files should be run in the following order:

- `1_master.do` : Main do file that runs STATA through code 13
- `2_master.m` : Main code file that runs all MATLAB code
- `3_master.do` : Main do file that runs STATA code to generate outputs

MATLAB

All MATLAB code can be found in the `\matlab` subfolder. The code for running the model simulations, numerical solution of the model structural estimation is written in MATLAB. The bootstrap samples were drawn using a simple python script. Global Optimization Toolbox in MATLAB was used to solve the model most accurately and avoid local minima.

- `sex_simulation_v13.m` : This script generates the data for Matching and dowry figures for the paper. It runs the model on simulated data and puts the output csv files in the `\output` subfolder. It then calls the relevant functions to solve the model for this data and given parameters to get the predicted outcomes from the model.
- `solve_model.m` : Function that takes X, Y and the parameters and returns $H(k^*)$. It solves the model recursively starting from the topmost income class and returns the predicted amount of sex selection in each class for the given distribution of incomes and parameters.
- `set_vin.m` : Function that solves the model and returns the difference in the outside option and utility of the boy's family for a given guess of the boy's family utility at the top.
- `getdata.m` : Function to read in the required data in the appropriate format from the `\data` folder.
- `my_phi.m` : Functions to calculate the utility of the girls family under the corresponding scenarios.
- `estimation_bootstrap.m` : Script to run the structural estimation using the Global Optimization Toolbox in MATLAB to search estimate values of the parameters to get confidence intervals.
- `estimation_sample.m` : Script to run the structural estimation using the Global Optimization Toolbox in MATLAB to search estimate values of the parameters from the data sample.
- `est_all.m` : Function called by the script `estimation_bootstrap.m` to get the overall sum of squared errors across all castes.
- `counter_factuals.m` : Script used to get the results of counterfactual simulations. Uses the files `my_phi_girls.m`, `set_vin_all.m`, `set_vin_girls.m`, `set_vin_top.m`, `solve_model_all.m`, `solve_model_girls.m`, `solve_model_top.m` to get the simulation results for different counterfactual experiments proposed in the paper.
- `plot_preds.m` : Script used to get the data for u_0 decomposition figure in the paper.

STATA

All STATA code can be found in the `\stata` subfolder.

Clean and code raw data

- `1_master_cleaning.do` : Cleans the raw data. The datasets available on the Microdata Library differ slightly from those used in this script, as they exclude variables containing personally identifiable information (PII). As a result, the code may not run fully using only the publicly available data. However, for verification purposes, the version of the dataset containing PII was shared with the verification team.
- `2_identifying_children.do` : Identifies children in the household.
- `3_hh_structure.do` : Identifies household structure.
- `4_adding_historical_data.do` : Cleans historical data.

Prepare data for analysis

- `5_dataprep.do` : Merges family structure variables and historical data to create final data for analysis.

Hypergamy and relative wealth

- `6_hypergamy_analyses_with_histdata.do` : Creates table regressing relative income and education of groom on relative wealth and education of bride.

Sex selection and wealth

- `7_bootstrap_breg_ar.do` : Runs the main bootstrap regressions for all kids under 6. The datasets available on the Microdata Library differ slightly from those used in this script, as they exclude variables containing personally identifiable information (PII). As a result, the code may not run fully using only the publicly available data. However, for verification purposes, the version of the dataset containing PII was shared with the verification team.
- `8_robinson_newpredinc_castefe_1poly_v2.do` : Runs the benchmark robinson regression with control function for a different specification.
- `9_robinson_lowess_hh_withunder6_server.do` : Runs sex selection with all HH and HH with at least one under 6.
- `10_OLS_robinson.do` : Runs the benchmark robinson regression with control function.
- `11_robinson_lowess_over7.do` : Runs sex selection across age groups.
- `12_robinson_firstborn_lowess_server.do` : Runs sex selection across birth order.
- `13_1poly_castebycaste_final.do` : Runs sex selection and relative wealth, caste by caste.

Tables with selected samples

- `14_cross_validation_edit.do` : Runs a cross validation with control function and a mean forecast sum of squared errors.
- `15_marriage_analysis_ar1.do` : Runs marriage patterns. The datasets available on the Microdata Library differ slightly from those used in this script, as they exclude variables containing personally identifiable information (PII). As a result, the code may not run fully using only the publicly available data. However, for verification purposes, the version of the dataset containing PII was shared with the verification team.

DOWRY

- `16_dowry_income_analysis.do` : Runs dowry and relative wealth.

Appendix tables and figures

- `17_table_A1.do` : Compares demographic and socioeconomic characteristics: Table A1.
- `18_age_diff_prop_married.do` : Analyzes (i) Age difference between husband and wife by birth cohorts; (ii) Proportion married by age; (iii) Variation in age at marriage by relative wealth; (iv) Variation in spousal age gap by relative wealth.
- `19_hsc_enroll.do` : Runs and indicator for enrollment in higher secondary school.
- `20_sex_ratios.do` : Creates sex ratio tables.

- 21_structural_figures.do : Runs structural parameter estimates, creates final figures and tables.

Data Availability

- ☐ All data are publicly available.
- ☒ Some data cannot be made publicly available.
- ☐ No data can be made publicly available.

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.
- ☐ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

Data

The raw data, which serves as main input for the project, can be located in the `\raw` subfolder, as its name indicates. Manipulated data in the code will be saved in the `\coded` subfolder.

The empirical analysis uses data from the South India Community Health Study (SICHS). It makes use of two components of the SICHS:

1. A census of all 298,000 households drawn from 57 castes residing in the study area, completed in 2014.
2. A detailed survey of 5,000 representative households, completed in 2016.

The analysis also uses data from the Demographic and Health Surveys (DHS) Program. To access this data, it is necessary to create an account on their [website](#). The databases used are from 2005-06 and 2015-16 for individual recode.

Dataset list

Data file	Source	Notes	License to share data
data/raw/06_8class	Created by authors	Files generated using the Stata code 13_1_smodel_data.do	Yes
data/raw/FPR	Collected by authors	Female primary respondent survey. Available on Microdata Library as Licensed files.	Yes
data/raw/MPR	Collected by authors	Male primary respondent survey. Available on Microdata Library as Licensed files.	Yes
data/raw/ROSTER *	Collected by authors	Household roster survey. Available on Microdata Library as Licensed files. These datasets differ from those used	Yes

Data file	Source	Notes	License to share data
		in the code due to the exclusion of PII variables.	
data/raw/SP0	Collected by authors	Spouse of MPR survey. Available on Microdata Library as Licensed files.	Yes
data/raw/2017-11-23_married_couples_pred_income_1871.dta *	Created by authors	Available on Microdata Library as Licensed files. This dataset differ from the one used in the code due to the exclusion of PII variables.	Yes
data/raw/2017-12-07-Dowry_income.dta *	Created by authors	SICHS survey, uses MPR, FPR and HH roster databases. Available on Microdata Library as Licensed files. This dataset differ from the one used in the code due to the exclusion of PII variables.	Yes
data/raw/Arcot_area.csv	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	No

Data file	Source	Notes	License to share data
data/raw/Arcot_caste.csv	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	No
data/raw/Arcot_occupation.csv	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and	No

Data file	Source	Notes	License to share data
		occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	
data/raw/CompleteData_Rough_10Oct2015.dta *	Collected by authors	NIRT Census data. Available on Microdata Library as Licensed files. This dataset differ from the one used in the code due to the exclusion of PII variables.	Yes
data/raw/IAIR52FL.dta	The DHS Program	To download the data, please register on the website, request the data for India, and download the Individual Recode Stata dataset for 2005–06.	No
data/raw/IAIR74FL.dta	The DHS Program	To download the data, please register on the website, request the data for India, and download the Individual Recode Stata dataset for 2015–16.	No
data/raw/Master_v1.csv	British Library in London	This file contains crosswalk between historical taluks and current villages in Tamil Nadu, done based on names and historical maps. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and	No

Data file	Source	Notes	License to share data
		occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	
data/raw/Predicted1871IncomeForSICH5.dta	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	No
data/raw/tableA1_S1_TN_SICH5_comparison.xlsx	Indiastat online repository.	Compiles data from official administrative sources. Find below the instructions to set up this file.	No
data/raw/TamilNadu-Habitations_Original.dta	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and	No

Data file	Source	Notes	License to share data
		occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	
data/raw/Tirupattu_area.csv	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area,	No

Data file	Source	Notes	License to share data
		caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	
data/raw/Tirupattur_caste.csv	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	No
data/raw/Tirupattur_occupation.csv	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of	No

Data file	Source	Notes	License to share data
		population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432..	
data/raw/Wallajah_area.csv	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	No
data/raw/Wallajah_caste.csv	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press,	No

Data file	Source	Notes	License to share data
		1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras: Lawrence Asylum Press, 1874. Pages: 1 – 432.	
data/raw/Wallajah_occupation.csv	British Library in London	Data collected in 2013. Books: Madras Districts: Census statement of population of 1871 in each village of the North Arcot District arranged according to area, caste and occupation. Madras: Foster Press, 1874. Pages: 1 – 492; Madras Districts: Census statement of population of 1871 in each village of the South Arcot District arranged according to area, caste and occupation. Madras: Adelphi Press, 1874. Pages: 1 – 269; Madras Districts: Census statement of population of 1871 in each village of the Chingleput District arranged according to area, caste and occupation. Madras: Scottish Press, 1874. Pages: 1-199; Madras Districts: Census statement of population of 1871 in each village of the Salem District arranged according to area, caste and occupation. Madras:	No

Data file	Source	Notes	License to share data
		Lawrence Asylum Press, 1874. Pages: 1 – 432.	
data/raw/unweighted_samples_12_castes.csv	Created using Python 3.12	To derive this dataset, start by installing and importing the necessary Python libraries: NumPy (version 2.2.1) and pandas (version 2.2.3). You must set the random seed to 5 (<code>np.random.seed(5)</code>) to ensure reproducibility. Define a list of caste identifiers, which in this case includes 12 specific values (<code>castes = [1, 4, 5, 7, 10, 14, 17, 23, 28, 33, 38, 44]</code>). Then, use <code>np.random.choice()</code> to randomly select values from this list, creating a 500-row by 12-column matrix where each element is a randomly chosen caste identifier. Convert this matrix into a pandas DataFrame and save it as a CSV file (<code>unweighted_samples_12_castes.csv</code>) without including headers or an index. The resulting file contains unweighted random samples of caste identifiers for further analysis.	Yes

Data citation

Data.Name	Citation
“Madras District Village Census, 1871 (North Arcot, South Arcot, Chingleput, Salem)”	British Library (1871)
“Demographic and Socioeconomic Indicators – South India”	Indiastat (2009-2018)
“Demographic and Health Surveys (DHS) 2005-06”	DHS (2005-2006)
“Demographic and Health Surveys (DHS) 2015-16”	DHS (2015-2016)

Notes

(*) The dataset available on the Microdata Library differ slightly from those used by the reproducibility team during verification, as the public versions exclude variables containing personally identifiable information (PII). As a result, some code files may not run in full. Specifically, the following do-files are affected:

- code/stata/1_master_cleaning.do.
- code/stata/subcode/1_1_ind_cleaning_akhil.do

- code/stata/7_bootstrap_breg_ar.do.
- code/stata/15_marriage_analysis_ar1.do.

Instructions to set up Excel file: tableA1_S1_TN_SICHS comparison.xlsx

- To access this information, you must first subscribe to [Indiastat](#).
- Then, on the main page, select Demographics.
- In Geographies, choose the following: Maharashtra, Karnataka, Andhra Pradesh, Tamil Nadu.
- For each of these locations, select:
 - Subcategory: Population
 - Subcategory-1: Population by Age Group
 - Subcategory-2: Age-Group Wise Percentage of Population.
 - Year: 2011 (and also 2018 for Karnataka).
 - In Result: Age-Group Wise Percentage of Population by Sex and Marital Status (Rural and Urban).
- Following the same path, download the following files for India: State-wise Total Population by Residence and Sex, Labour Force Participation Rate (LFPR) by Different Age Groups (in Per 1000 Persons) and Sex in Rural India (2009-2010), State-wise Religious Population by Residence, State-wise Child Sex Ratio (Age Group 0-6 Years) by Residence, State-wise Sex Ratio (Female per 1000 Males), State-wise Literacy Rate by Residence and Sex. 6. Once all files are downloaded, please compile the data into a single Excel file, using separate sheets for each table.

Data included in the Excel sheets of file: tableA1_S1_TN_SICHS comparison.xlsx

Tab name	Information	Notes
Total Population	State-wise Total Population by Residence and Sex in India (As per 2011 Census)	
KR_age_group-wise distribution	Age Group-wise Percent Distribution of Population by Sex and Marital Status (Rural and Urban) in Karnataka - 2018	
MH_age group-wise distribution	Age Group-wise Percent Distribution of Population by Sex and Marital Status (Rural and Urban) in Maharashtra - 2011	
AP_age group-wise distribution	Age Group-wise Percent Distribution of Population by Sex and Marital Status (Rural and Urban) in Andhra Pradesh - 2011	
Age_dist_calculations	Age Group-wise Percent Distribution of Estimated Population by Sex and Residence in Karnataka - 2011 Age Group-wise Percent Distribution of Estimated Population by Sex and Residence in Tamil Nadu - 2011 Age Group-wise Percent Distribution of Estimated Population by Sex and Residence in Maharashtra - 2011 Age Group-wise Percent Distribution of Estimated Population by Sex and Residence in Andhra Pradesh - 2011	Find more details about this tab below.
TN_age group-wise distribution	Age Group-wise Percent Distribution of Population by Sex and Marital Status (Rural and Urban) in Tamil Nadu - 2011	
LFPR-Rural(2009-10)	Labour Force Participation Rate (LFPR) by Different Age Groups (in Per 1000 Persons) and Sex in Rural India (2009-2010)	

Tab name	Information	Notes
religion	State-wise Religious Population by Residence in India (As per 2011 Census)	
ChildSexRatio	State-wise Child Sex Ratio (Age Group 0-6 Years) by Residence in India (As per 2011 Census)	
StateWiseSexRatio	State-wise Sex Ratio (Female per 1000 Males) in India (As per 2011 Census)	
StateWiseLiteracyRateByResidenc	State-wise Literacy Rate by Residence and Sex in India (As per 2011 Census)	
TotalPopulationByResidenceAndSe	State-wise Total Population by Residence and Sex in India (As per 2011 Census)	

Instructions to get Age Group-wise Distribution of Population by Sex and Marital Status in South India

1. Table Structure:
In the first cell of the Age_dist_calculations sheet, create the structure of the table as follows:
- The first column contains the age groups.
 - The second and third columns contain the number of married males and married females, respectively.
 - The fourth and fifth columns contain the total rural male and total rural female populations, respectively.
2. Calculating Columns 4 and 5 (Total Rural Population by Age and Sex):
For the following states: Karnataka, Tamil Nadu, Maharashtra, and Andhra Pradesh:
- a. Use the table "State-wise Total Population by Residence and Sex in India (As per 2011 Census)" to find the rural male and female population for each state.

b. Multiply those values by the age-specific rural male/female distribution from the table "Age Group-wise Percent Distribution of Estimated Population by Sex and Residence - 2011".

c. Sum the resulting age-group-specific values across the four states to get the totals for each age group.
3. Calculating Columns 2 and 3 (Married Rural Population by Age and Sex):
For each state, take the age- and sex-specific rural population from Step 2b and multiply it by the proportion of married males and females from the table "Age Group-wise Percent Distribution of Population by Sex and Marital Status (Rural and Urban)".
Finally, sum the results for Karnataka, Tamil Nadu, Maharashtra, and Andhra Pradesh by age group to fill in Columns 2 and 3.

Shell script files

All `.sh` files inside the `\code` folder are the ones which were used to run the estimation on the SIT cluster at Universitat Pompeu Fabra (UPF). These will have to be modified accordingly based on what cluster the code is being run on.

Computational requirements

Software Requirements

- Stata (code was last run with version 18)
- Matlab (code was run with Matlab 2024b)

Memory, Runtime, Storage Requirements

Summary

Approximate time needed to reproduce the analyses on a standard (CURRENT YEAR) desktop machine:

- ☐ <10 minutes
- ☐ 10-60 minutes
- ☐ 1-2 hours
- ☐ 2-8 hours
- ☐ 8-24 hours
- ☐ 1-3 days
- ☐ 3-14 days
- ☒ > 14 days

Approximate storage space needed:

- ☐ < 25 MBytes
- ☐ 25 MB - 250 MB
- ☐ 250 MB - 2 GB
- ☐ 2 GB - 25 GB
- ☒ 25 GB - 250 GB
- ☐ > 250 GB

Details

Portions of the code were reproduced on a Windows Server 2019 Standard machine with an Intel Xeon E7-4850 v4 CPU @ 2.10GHz (16 processors), 64 GB of RAM, running Stata 18.0 MP.

Portions of the code were reproduced on a Windows 10 Enterprise machine with an Intel Xeon Gold 6132 CPU @ 2.60GHz (16 cores, 16 logical processors), 128 GB of RAM, running Matlab 2024b.

Figures and Tables

`\figures` and `\tables` subfolders inside the `\output` folder are shell folders that will populate with figures and tables produced by the code.

List of Exhibits

Exhibit name	Output filename	Script	Note
Figure 1	structural_matching.pdf, structural_dowry.pdf	21_structural_figures.do (line 315, line 81)	Found in output/figures
Figure 2	fig_dowries_relwealth.pdf, fig_dowry_male_female.pdf	16_dowry_income_analysis.do (line 542, 457)	Found in output/figures

Exhibit name	Output filename	Script	Note
Figure 3	robinson_all_0to6_lpoly.pdf, predincome_controlfunction_lpoly.pdf	10_robinson_hh_withunder6_server.do (line 356), 9_robinson_newpredinc_castefe_lpoly_v2.do (line 172)	Found in output/figures
Figure 4	structural_ubarea.pdf, structural_counterfactual.pdf	21_structural_figures.do (line 253, line 37)	Found in output/figures
Table 1	Table_marriage_analysis.tex	15_marriage_analysis_ar1.do (line 525)	Found in output/tables
Table 2	Table_marital_matching.tex	15_marriage_analysis_ar1.do (line 558)	Found in output/tables
Table 3	2018-02-01_Table_Reg_Hypergamy_EducCNS.tex	6_hypergamy_analyses_with_histdata.do (line 967)	Found in output/tables
Table 4	d_girl6_bmark_table4.tex	7_bootstrap_breg_ar.do (line 422)	Found in output/tables
Table A1	table_A1.tex	17_table_A1.do (line 417)	Found in output/tables
Table A2	Table_Sex_Ratios_upd_v2.tex	20_sex_ratios.do (line 191)	Found in output/tables
Table A3	table_A3.tex	19_hsc_enroll.do (line 25)	Found in output/tables
Table A4	cross_validation_mse.tex	14_cross_validation_edit.do (line 507)	Found in output/figures
Table A5	robinson_ols_np.tex	8_OLS_robinson.do (line 47)	Found in output/tables
Table A6	structural_estimates_2.tex	21_structural_figures.do (line 289)	Found in output/tables
Figure A1	age_diff.pdf	18_age_diff_prop_married.do (line 70)	Found in output/figures
Figure A2	age_at_marr_rank.pdf	18_age_diff_prop_married.do (line 206)	Found in output/figures
Figure A3	age_gap_rank.pdf	18_age_diff_prop_married.do (line 86)	Found in output/figures
Figure A4	prop_marr_age.pdf	18_age_diff_prop_married.do (line 162)	Found in output/figures
Figure A5	robinson_dgirl713_castefe_vfe.pdf, birthorder.pdf	11_robinson_over7.do (line 99) 12_robinson_firstborn_server.do (line 372)	Found in output/figures

Exhibit name	Output filename	Script	Note
Figure A6	all_specifications.pdf	14_cross_validation_edit.do (line 463)	Found in output/figures
Figure A7	csr_castescombined_desc_sort.pdf, csr713_castescombined_desc_sort.pdf	13_ipoly_castebycaste.do (line 378, line 412)	Found in output/figures

To replicate analysis

It is possible to reproduce all the analysis using only the codes available in GitHub.

1. Download the `\sex-selection-india` .zip folder from [World Bank Reproducibility Research Repository](#) and unzip it.
2. Inside the `\sex-selection-india` folder, create a new folder called "output". Inside this new folder, create two new folders with the names "figures" and "tables".
3. Following the Dataset list notes download the `\data` and save it inside the `\sex-selection-india` folder. This way, the `\sex-selection-india` folder must contain the folders `\code` , `\excel` , `\output` and `\data` , the `1_master.do` code file, the `2_master.m` script, the `3_master.do` do file and the `README.md` .
 - The dataset downloaded from the Microdata Library should be saved in the folder specified in the "Data file" column, and the file names should be changed to uppercase (CAPS) to match the code requirements.
4. In the `1_master.do` code file, the `2_master.m` script and the `3_master.do` do file, please update the main path (global `sexselection`) to match the path on your computer that leads to the `\sex-selection-india` folder.
5. Run the `1_master.do` code file, the `2_master.m` script, and then, the `3_master.do` do file.

Everything should run smoothly. If any inconvenience is encountered, please don't hesitate to reach out to the authors.

References

British Library. 1871. Madras District Village Census [dataset]. British Library.

DHS. 2005-2006. "India: Standard DHS, 2005-06 [dataset]." https://dhsprogram.com/data/dataset/India_Standard-DHS_2006.

DHS. 2015-2016. "India: Standard DHS, 2015-16 [dataset]." https://dhsprogram.com/data/dataset/India_Standard-DHS_2015.

Indiastat. 2009-2018. "Demographics [dataset]." <https://www.indiastat.com/data/demographics>.