README and Guidance for Replication of:

Bad Taste: Gender Discrimination in Consumer Lending
(Journal of Political Economy: Microeconomics)

Ana María Montoya, Eric Parrado, Alex Solís, and Raimundo Undurraga

July 5th, 2024

Overview

The code in this replication package generates the results from the two main data sources: observational data and experimental data on loan applications. The software used is Stata version 17. The master do File run all of the codes to generate the tables in the main body of the paper. The replicator should expect the code to run for about 1-2 hours. For questions regarding the replication code, please email Raimundo Undurraga to `undurraga.raimundo@gmail.com`.

Data Availability and Provenance Statements

1. Observational Data

SBIF data contains the universe of consumer loan transactions between banks and borrowers for the period 2013-2016. It provides information at the transaction level, including whether the loan application was approved or not, dummies by loan amount, dummies by loan length, dummies by bank, loan application date, as well as applicant level characteristics including gender, dummies by age, if married, wage dummies, debt dummies, if client of the bank, and dummies by credit score category.

The data is available in the public use data folder (**\Replication\Data\),** and it is structured as follows:

- **ObservationalData_Applications.dta**: it includes the full sample of observational loan applications
- **ObservationalData_Applications_randomsample15000_A4.dta**: it includes a sample of 15,000 observations randomly sampled from ObservationalData_Applications.dta

- **ObservationalData_Applications_randomsample15000_A8.dta**: it includes a sample of 15,000 observations randomly sampled from ObservationalData_Applications.dta, conditional on loan applications requested by non-client applicants
- **ObservationalData_Applications_randomsample15000_B4.dta**: it includes a sample of 15,000 observations randomly sampled from ObservationalData_Applications.dta, conditional on 1,500-13,500 USD loan applications requested by applicants aged 25-35
- **ObservationalData_Applications_randomsample15000_B8.dta**: it includes a sample of 15,000 observations randomly sampled from ObservationalData_Applications.dta, conditional on 1,500-13,500 USD loan applications requested by non-client applicants aged 25-35

2. Experimental Data

We conducted a correspondence study with real loan applicants. We randomly assigned loan requests to gender-balanced prospective borrowers (i.e., men and women with identical risk profiles), who then sent the assigned requests to randomly assigned loan officers. The experiment worked based on two sides: loan applications (submitted by loan applicants) and loan officers.

Loan applications data includes information on whether the loan application was submitted by the applicant, if it was responded by the assigned loan officer, if the officer asked for additional information, and if the application was approved. The data also include dummies by loan amount-term, dummies by week of submission of loan application, and dummies by region-bank where the assigned loan officer works. It also include loan officer covariates, including gender, age dummies, experience dummies, if has higher education, dummies by gender preferences, dummies by gender beliefs, a dummy for whether the officer donated the second ticket or not, the gender and last name type of the assigned ticket for donation, a dummy for whether the loan officer was treated or not by the information intervention, if their gender beliefs are aligned or not with the treatment message, and a dummy for whether the officer agrees with the treatment message or not. Finally, the data also includes loan applicant covariates, including gender, age dummies, if married, wage dummies, if self-employed, and whether the applicant is client of the bank or not.

The data is available in the public use data folder (**\Replication\Data\),** and it is structured as follows:

- **ExperimentalData_Applications_All.dta**: it includes all the experimental loan applications (N=1,616), regardless of whether the applicant submitted the assigned application (non-attriter applications) or not (attriter applications).

- **ExperimentalData_Applications_NonAttrit.dta**: it includes only experimental loan applications that were submitted by the applicants, i.e., non-attriter applications (N=1,313).

- **ExperimentalData_Applicants.dta**: it includes data on all the experimental applicants (N=404).
- **ExperimentalData_Officers_All.dta**: it includes data on all the experimental loan officers (N=629).
- **ExperimentalData_Officers_NonAttrit.dta**: it includes only data on experimental loan officers receiving at least one loan application on the part of loan applicants (N=546).

Note that region-bank dummies and week dummies are specific for each sample of analysis, which avoids estimation problems of statistics associated to multiple hypothesis testing when using the mhtreg command.

Computational requirements

Software Requirements

All analyses in the paper are run using Stata MP, version 17.

Memory and Runtime Requirements

Approximate time needed to reproduce the analyses on a standard desktop machine: 1.5 hours. The code was last run on a on a Windows 10 Home 64-bit Operating System, X64 based server with 12 GB of RAM.

Instructions to Replicators

To replicate, run the Master do File (`master_do.do`) changing the path in line 9. Running the Master do File will run all the commands to generate the tables in the paper. The only figure of the paper (Figure I) does not requires data analysis to be computed and is provided externally.

Description of programs/code

- Program in `code/0_stata_setup` will add required packages from SSC.
- Programs in `code/1_analysis will` generate all tables 1 through 8 in the paper. Each program identifies the table it creates (e.g., `table1.do`). Output files are called appropriate names (`Results/Tables/Table1_PanelA_123.xls`) and should be easy to correlate with the manuscript.

Notes:

- Output files computing regression results clustering the standard errors at the region bank level have names ending in "`_cse`". Output files computing regression results that

bootstrap the standard errors have names ending in "_bse". Output files computing regression results with Huber-White (robust) standard errors have names ending in "_r".

- Bootstrapped standard errors are generally computed by using the `vce(boot)` option from the `regress` command. However, sometimes the command warns that there are "`insufficient observations to compute bootstrap standard errors`", in which case we proceed using Frisch-Waugh method. In those cases, the output is directly observed in the log file (no exported table is created).
- Multiple hypothesis testing is conducted using the `mhtreg` command, and output is directly observed in the log file (no exported table is created).
- Descriptive statistics to compute table 4 are directly observed in the log file (no exported table is created).

List of programs and tables

master_do.do: this do file automatizes the full replication of Tables 1 through 8.

stata_setup.do: this do file adds the required packages from SSC.

table1.do: this do file replicates results in Table 1. Output is observed in the following excel files in `Results/Tables/`: `Table1_PanelA_123, Table1_PanelA_4, Table1_PanelA_567, Table1_PanelA_8, Table1_PanelB_123, Table1_PanelB_4, Table1_PanelB_567, Table1_PanelB_8`

table2.do: this do file replicates results in Table 2. Output is observed in the following excel files in `Results/Tables/`: `Table2_PanelA_Unadjusted_Mean_Diff_cse , Table2_PanelA_Unadjusted_Mean_Diff_bse, Table2_PanelA_Model1_cse, Table2_PanelA_Model1_bse, Table2_PanelA_Model2_cse, Table2_PanelA_Model2_bse, Table2_PanelB_Unadjusted_Mean_Diff, Table2_PanelB_Model1, Table2_PanelB_Model2`

table3.do: this do file replicates results in Table 3. Output is observed in the following excel files in `Results/Tables/`: `Table3_Unadjusted_Mean_Diff_cse, Table3_Unadjusted_Mean_Diff_bse, Table3_Model1_cse, Table3_Model1_bse, Table3_Model2_cse, Table3_Model2_bse.` Output for p-values associated with multiple hypothesis testing is directly observed in the log file (column `pthm3_1` in the corresponding MHT table).

table4.do: this do file replicates results in Table 4. Output is observed directly in the log file

table5.do: this do file replicates results in Table 5. Output is observed in the following excel files in `Results/Tables/`: `Table5_col_1_r, Table5_col_2_r, Table5_col_3_r, Table5_col_4_r, Table5_col_5_r, Table5_col_6_r.` Output for regressions with bootstrapped standard errors (estimated through Frisch-Waugh) is directly observed in the log file. Output for p-values associated with multiple hypothesis

testing is directly observed in the log file (column `pthm3_1` in the corresponding MHT table).

table6.do: this do file replicates results in Table 6. Output is observed in the following excel files in `Results/Tables/`: `Table6_cols_1_2_3_cse, Table6_cols_1_2_3_bse, Table6_cols_4_cse, Table6_cols_5_cse, Table6_cols_5_bse, Table6_cols_6_cse, Table6_cols_6_bse, Table6_cols_7_cse, Table6_cols_7_bse`. Output for regression column 4 with bootstrapped standard errors (estimated through Frisch- Waugh) is directly observed in the log file. Output for p-values associated with multiple hypothesis testing is directly observed in the log file (column `pthm3_1` in the corresponding MHT table).

table7.do: this do file replicates results in Table 7. Output is observed in the following excel files in `Results/Tables/`: `Table7_bse`. Output for regression column 3 with bootstrapped standard errors (estimated through Frisch-Waugh) is directly observed in the log file. Output for p-values associated with multiple hypothesis testing is directly observed in the log file (column `pthm3_1` in the corresponding MHT table).

table8.do: this do file replicates results in Table 8. Output is observed in the following excel files in `Results/Tables/`: `Table8_cols_1_2_3_cse, Table8_cols_1_2_3_bse, Table8_cols_4_5_6_cse, Table8_cols_4_5_6_bse, Table8_cols_7_8_9_cse, Table8_cols_7_8_9_bse.` Output for p-values associated with multiple hypothesis testing is directly observed in the log file (column `pthm3_1` in the corresponding MHT table).