Replication Package for
"The Long-Term Impact of High School Financial Education:
Evidence from Brazil"

Miriam Bruhn, Gabriel Garber, Sergio Koyama, and Bilal Zia

This replication package contains the publicly available data and code for replicating all results in the paper. To gain access to the data that is not publicly available, researchers interested in replicating the study can do so at BCB´s premises by submitting a research project as explained in BCB external collaboration page (https://www.bcb.gov.br/en/publications/externalcollaborationresearch).

A. Data sources and abbreviations

short-term IE - Data collected during the short run impact evaluation
SRF - Internal Revenue Secretariat Registry of individuals, firms, firm ownership and Simples/MEI option
CCS - Registry of the clients of the financial system, Central Bank of Brazil
SCR - Credit Registry System, Central Bank of Brazil
RAIS - Annual Report of Social Information, Ministry of Labor
AE - Resgitry of Auxílio Emergencial payments, Ministry of Social Development
CU - Cadastro Único, Ministry of Social Development
IBGE - National Instituto of Geography and Statistics

B. Main replication code: LTimpactFE.do

Ran using Stata 19, required packages: estout, ftools, outreg2, reghdfe, unique, winsor2
OS: Windows Server 2016 Standard (version 1607)

The main replication code uses 4 data files:

1. "intervention_data.dta" [not public due to variables from the SRF]

This file includes variables from the short-term IE and a dummy variable indicating whether the student is in the long run sample or not. It is used to generate Tables 1, A1, A2, A3, and A4.

Variable description [source]:

id_geral: Student ID [short-term IE]
treatment: School randomly assigned to treatment group [short-term IE]
cd_escola: School code [short-term IE]
pair_all: Matched school pair for randomization [short-term IE]
round: Follow-up 2 [short-term IE]
female_coded: Student is female (coded based on name) [short-term IE]

miss_f_coded: Gender could not be coded based on name [short-term IE]
female: Student is female, sourced from the taxpayer registry for long-term sample [SRF]
matriculas: Number of students in school (2008 data) [short-term IE]
docentes: Number of teachers in school (2008 data) [short-term IE]
abandonona1sriemdio: Grade-level dropout rate (2009 data) [short-term IE]
aprovaona1sriemdio: Grade-level passing rate (2009 data) [short-term IE]
vl_proficiencia_bl: Student financial literacy score [short-term IE]
bl_test: Student took fin lit test at baseline [short-term IE]
bl_aluno: Student answered baseline questionnaire [short-term IE]
vl_proficiencia_fup: Financial Proficiency Score [short-term IE]
autonomia_final2_bl: Student financial autonomy index [short-term IE]
poupar_final2_bl: Student intention to save index [short-term IE]
dumm_rp_08_bl: Education of student's mother: At least some secondary [short-term IE]
dumm_rp_09_bl: Education of student's father: At least some secondary [short-term IE]
dumm_rp_14_bl: Student's family receives Bolsa Familia cash transfer [short-term IE]
dumm_rp_23_bl: Student has computer with internet at home [short-term IE]
dumm_rp_24_bl: Student has failed at least one school year [short-term IE]
dumm_rp_49_bl: Student is not working at the moment [short-term IE]
dumm_rp_50_bl: Q50 Receives income [short-term IE]
dumm_rp_53B_bl: Q53 Pct of income saved is non-zero [short-term IE]
dumm_rp_55_bl: Q55 Has borrowed money (any source) [short-term IE]
dumm_rp_56_bl: Q56 Is behind on payments (unconditional) [short-term IE]
dumm_rp_64A_bl: Q64 I make a list of all monthly expenses [short-term IE]
dumm_rp_65A_bl: Q65 Saves money for future purchases [short-term IE]
dumm_negotiates_bl: Q93 Q95: Negotiates prices or payment methods [short-term IE]
dumm_rp_53B_fup: Q53 Pct of income saved is non-zero [short-term IE]
dumm_rp_55_fup: Q55 Has borrowed money (any source) [short-term IE]
dumm_rp_56_fup: Q56 Is behind on payments (unconditional) [short-term IE]
short_sample: 1 for short-term sample, missing otherwise [short-term IE]
baseline_sample: 1 for short-term baseline sample, missing otherwise [short-term IE]
fup1_sample: 1 for short-term follow-up 1 sample, missing otherwise [short-term IE]
fup2_sample: 1 for short-term follow-up 2 sample, missing otherwise [short-term IE]
sample: 1 for long-term sample, missing otherwise [short-term IE + SRF]

2. "main_panel.dta" [not public]

This file contains background and treatment variables form the short-term IE, as well as the panel data from administrative datasets. It is used to generate Figures 1 and 2, as well as Tables 2, 3, 4, 5, A6, A7, A8, A9, A10, A11, A12, A14, and A15.

Variable description [source]:

cpf_hash: student unique identifier [SRF]
month: month in format YYYYMM (integration of all datasets)
treatment: School randomly assigned to treatment group [short-term IE]

cd_escola: School code [short-term IE]
pair_all: Matched school pair for randomization [short-term IE]
cd_uf: State code [short-term IE]
female: Student is female [short-term IE]
relationship: 1 if student has account, 0 otherwise [CCS]
d_balance 1: for balance>0, 0 for balance=0 [SCR]
d_creditcardpurch: 1 if student uses credit card purchases, 0 othewise [SCR]
d_creditcarddebt: 1 if student uses credit debt, 0 othewise [SCR]
d_overdrafts: 1 if student uses overdrafts, 0 othewise [SCR]
d_nonpayrollloans: 1 if student uses nonpayroll loans, 0 othewise [SCR]
d_auto: 1 if student uses auto loans, 0 othewise [SCR]
d_payrollloans: 1 if student uses payroll loans, 0 othewise [SCR]
balance: student's total credit balance [SCR]
valccpurch: student's balance in credit card purchases [SCR]
valcreditcardebt: student's balance in credit card debt [SCR]
valoverdrafts: student's balance in overdrafts [SCR]
valnonpayrollloans: student's balance in nonpayroll loans [SCR]
valauto: student's balance in auto loans [SCR]
valpayrollloans: student's balance in payroll loans [SCR]
maxdelay maximum delay in days [SCR]
maxdelay2 maximum delay in days, excluding losses from before June 2016 [SCR]
loss201606 1 if student had losses in June 2016, 0 otherwise [SCR]
mei: 1 if student owns an individual microenterprise (MEI), 0 otherwise [SRF]
firm_owner 1 if student owns a larger firm (i.e. non MEI), 0 otherwise [SRF]
d_employed 1 if employed, 0 otherwise [RAIS]
ae_dummy 1 if student received Auxílio Emergencial, 0 otherwise [AE]
marc_pbf_all 1 if student received Bolsa Familia, 0 otherwise [CU]
factor_ipca_201912 multiplicative factor to adjust for inflation to BRL from Decemeber 2019, using [IBGE]
avg_wage sum of mean wages, measured in minimum wages [RAIS]
d_emp 1 if student is an employer, 0 otherwise [SRF+RAIS]

3. "coarser_panel.dta" [not public]

This file contains background and treatment variables from the short-term IE, as well as the panel data from administrative datasets using coarsely identified students. It is used to generate Table A13.

Variable description [source]:

cpf_hash: student unique identifier [SRF]
month: month in format YYYYMM (integration of all datasets)
treatment: School randomly assigned to treatment group [short-term IE]
cd_escola: School code [short-term IE]
pair_all: Matched school pair for randomization [short-term IE]

cd_uf: State code [short-term IE]
female: Student is female [short-term IE]
relationship: 1 if student has account, 0 otherwise [CCS]
d_balance 1: for balance>0, 0 for balance=0 [SCR]
maxdelay maximum delay in days [SCR]
mei: 1 if student owns an individual microenterprise (MEI), 0 otherwise [SRF]
d_employed 1 if employed, 0 otherwise [RAIS]

4. "income.xlsx"

This file contains income estimates, generated using the PNAD code described below. Results with this income variable are reported in Table A12.

C. PNAD data replication code: PNAD_LTFE.R

Ran using R-4.2.1/ RStudio 1.2.1335, packages listed in the code
OS: Windows 10 Enterprise (22H2)

This code uses two data files from the publicly available National Household Sample Survey (PNAD), 2019 Q4, conducted by IBGE. First, "input_PNADC_trimestral.txt" is necessary to load data in R. Second, the data are in "PNADC_042019." The data is used to generate income.xlsx, as well as Figure A1.

Variable description [source]:

UF: State
V1028: Weight of household and people (weighted by population projection)
V2007: Gender
V2009: Age
V2010: Race
V4012: Occupation description (In that job I was...)
V4019: Is the main occupation in a formal firm? (i.e. having a taxpayer inscription number)
VD3004: Highest level of education achieved
VD4009: Occupational position and job category (derived variable)
VD4020: Effective income from all jobs