



Higher Education Expansion and Firm Organization

Second Submission: RR_BRA_2025_429

Luis Eduardo San Martin
reproducibility@worldbank.org

October 7, 2025

This review verifies the reproducibility of the exhibits included in the paper “*Higher Education Expansion and Firm Organization*”.

Contents in this review:

1. Main findings
2. List of exhibits and reproducibility status
3. Reproduction Environment

Main findings

- **Every exhibit has been reproduced accurately.**
- The reproducibility verification was performed via a virtual meeting on September 23, 2025, as the restricted-access data could not be shared with reviewers. The code outputs and logs were sent to the reviewers afterward to complete the reproducibility verification.
- The code was successfully executed after:
 1. Changing the global path in *o. Main.do*.
- The code takes approximately 24 hours to run.
- We conducted our reproducibility analysis based on the paper shared by the author via OneDrive on October 7, 2025.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on 4 types of data: open, accessible (public but not included in the reproducibility package), limited-access (available after purchase), and restricted data (available only to the authors).
 - Open data is included in the public reproducibility package.
 - *data_hash_report.csv* lists the SHA256 hashes of all files in the Data folder to support data integrity checks. For the *BADEPI* dataset, it includes the hash for the intermediate dataset that compiled all data from multiple data files into a single data table (the file “*BADEPI.dta*”). Users who acquire the accessible, limited-access, and restricted data can use the hashes to verify that the data has not been altered.
- **Reproducibility Summary:**

- **Data:** Some data is restricted and has not been included in the reproducibility package.
- **Code:** All code files (from cleaning to analysis) are included in the reproducibility package.
- **Outputs:** All outputs are generated by code included in the reproducibility package.
- **Dependencies environment:** The authors created a new environment using the latest versions of dependencies available at the moment of the review.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1** Reproduced.
- **Table 1** Does not apply. The exhibit does not show analysis results.
- **Figure 2** Reproduced.
- **Figure 3 (a)** Reproduced.
- **Figure 3 (b)** Does not apply. The exhibit displays findings from another paper.
- **Figure 4 (a)** Reproduced.
- **Figure 4 (b)** Reproduced.
- **Table 2** Reproduced.
- **Figure 5** Reproduced.
- **Figure 6** Reproduced.
- **Figure 7 (a)** Reproduced.
- **Figure 7 (b)** Reproduced.
- **Figure 8 (a)** Reproduced.
- **Figure 8 (b)** Reproduced.
- **Figure 9** Reproduced.
- **Table 3** Reproduced.
- **Table 4** Reproduced.

Results in the Annex

For the Appendix, we did not review every exhibit. Instead, we randomly selected 10 exhibits from the remaining datasets to assess the appendix. Our review was based on those 10 exhibits. Since they were chosen randomly, we are operating under the assumption that if all randomly selected exhibits are reproducible, then the rest should be as well. The seed used to generate the random selection was [SEED], the Stata version used was 18.0, and the exhibits selected were: Figure A1, Figure A2 (a) Figure A3 (a), Figure A4 (a), Figure A4(b), Table B1, Figure B1 (a), Figure B1 (b), Figure B2 (a), Figure B2 (b).

- **Figure A.1** Reproduced.
- **Figure A.2 (a)** Reproduced.
- **Figure A.3 (a)** Reproduced.
- **Figure A.4 (a)** Reproduced.
- **Figure A.4 (b)** Reproduced.
- **Table B.1** Reproduced.
- **Figure B.1 (a)** Reproduced.
- **Figure B.1 (b)** Reproduced.
- **Figure B.2 (a)** Reproduced.
- **Figure B.2 (b)** Reproduced.

Reproduction Environment

Paper exhibits were reproduced on a computer with the following specifications:

- OS: Windows Server 2016 Standard
- Processor: Intel(R) Xeon(R) CPU E7-4890 v2 @ 2.80GHz 2.79 GHz (8 processors)
- Memory available: 128 GB
- Software version: Stata 18.0 MP