



Reproducible Research Repository

(In)Formalizing Jobs in Latin America and the Caribbean

Second Submission: FR_LAC_2025_380

Maria Reyes Retana, Kanika Shokeen

reproducibility@worldbank.org

2025-07-11

This review verifies the reproducibility of the exhibits included in the paper "*(In)Formalizing Jobs in Latin America and the Caribbean*".

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 code was successfully executed on a new computer after:
 1. Change the path in the main do file and run the code.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 20 minutes to run.
- We conducted our reproducibility analysis based on the paper shared by the authors by mail on August 5th.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on three types of data: open, accessible (public but not included in the reproducibility package), and restricted data (available only from the authors).
 - Open data is included in the public reproducibility package.
 - Reviewers used data provided directly by the authors to conduct the reproducibility verification, and this is not all 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. Users who acquire the accessible/restricted data can use this file 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. For more details, please refer to the README file.

- **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.
- **Reproducibility verification:** Reviewers used data provided directly by the authors to conduct the reproducibility verification, and this is not included in the public reproducibility package. The reviewers did not verify if publicly available data matches the data provided by the authors. A hash report that contains all the datasets used in the analysis is included in the reproducibility package.
- **Dependencies environment:** The reviewers created a new environment for dependencies using specific versions defined by the authors in the README file.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

Chapter 2

- **Figure 2.1 Reproduced, but table or figure includes manual changes from code output.** Size of the dots in the scatter plot were changed to make the observations more distinguishable. Compared to 'Figure_1_1.png'.
- **Figure 2.2 Reproduced, with manual changes.** The colours of the dots in the scatter plot were changed from red to green, and green to dark blue. Smaller units were also added to the x axis. Compared to 'Figure_1_2_.png'.
- **Figure 2.3 Does not Apply.** No analytical outputs are created in this figure.
- **Figure 2.4 Reproduced, with manual changes.** The outputs produced by the code was the total number of observations for each variable. They were manually calculated as a percentage for each category per country, and all decimal points were rounded up to whole numbers. Compared to 'Figure1.4_all'.
- **Figure 2.5a Reproduced, with manual changes.** The underlying data used to create this figure was matched from the file generated by the code Figure1.5_BRA.xlsx to the one used by the authors, F1.5.xlsx. The figure was then created in Excel (this file is included in the package), and the numbers match exactly. The authors manually cleaned the final excel for rows, and the figure was reproduced.
- **Figure 2.5b Reproduced, with manual changes.** The underlying data used to create this figure was matched from the file generated by the code textbf'Figure1.5_COL.xlsx' to the one used by the authors, 'F1.5.xlsx'. The authors manually cleaned the final excel for rows with missing entries for 'sector'. The figure was then created in Excel (this file is included in the package), and the numbers match, as well as the reproduced figure.
- **Figure 2.5c Reproduced, with manual changes.** The underlying data used to create this figure was matched from the file generated by the code textbf'Figure1.5_SLV.xlsx' to the one used by the authors, 'F1.5.xlsx'. The authors manually cleaned the final excel for rows with missing entries for 'sector'. The figure was then created in Excel (this file is included in the package), and the numbers match, as well as the reproduced figure.

- **Figure 2.5d Reproduced, with manual changes.** The underlying data used to create this figure was matched from the file generated by the code textbf'Figure1.5_MEX.xlsx' to the one used by the authors, 'F1.5.xlsx'. The authors manually cleaned the final excel for rows with missing entries for 'sector'. The figure was then created in Excel (this file is included in the package), and the numbers match, as well as the reproduced figure.
- **Figure 2.5e Reproduced, with manual changes.** The underlying data used to create this figure was matched from the file generated by the code textbf'Figure1.5_PER.xlsx' to the one used by the authors, 'F1.5.xlsx'. The authors manually cleaned the final excel for rows with missing entries for 'sector'. The figure was then created in Excel (this file is included in the package), and the numbers match, as well as the reproduced figure.
- **Figure 2.5f Reproduced, with manual changes.** The underlying data used to create this figure was matched from the file generated by the code textbf'Figure1.5_URY.xlsx' to the one used by the authors, 'F1.5.xlsx'. The authors manually cleaned the final excel for rows with missing entries for 'sector'. The figure was then created in Excel (this file is included in the package), and the numbers match, as well as the reproduced figure.
- **Table 2.1 Reproduced, with manual changes.** The code produced four excel sheets, each representing one of the four characteristics (gender, age, area, level of education). Manually, percentages were created for each category for the columns 'Mandated & complying' and 'Mandated & non-complying'. Separately, each category of values for the columns 'Exempt & participating' and 'Exempt & non-participating' were also totalled to create percentages. All decimals were rounded to whole numbers. After manual calculations in excel, the exhibits were deemed reproduced.
- **Figure 2.6a Reproduced, with manual changes.** The output was produced as 'Figure 1.6_BRA.xlsx'. Percentages for each of the four values were manually calculated. A line chart was produced in excel from the percentages, which matched the values in the paper.
- **Figure 2.6b Reproduced, with manual changes.** The output was produced as 'Figure 1.6_COL.xlsx'. Percentages for each of the four values were manually calculated. A line chart was produced in excel from the percentages, which matched the values in the paper.
- **Figure 2.6c Reproduced, with manual changes.** The output was produced as 'Figure 1.6_MEX.xlsx'. Percentages for each of the four values were manually calculated. A line chart was produced in excel from the percentages, which matched the values in the paper.
- **Figure 2.6d Reproduced, with manual changes.** The output was produced as 'Figure 1.6_URL.xlsx'. Percentages for each of the four values were manually calculated. A line chart was produced in excel from the percentages, which matched the values in the paper.
- **Figure 2.6e Reproduced, with manual changes.** The output was produced as 'Figure 1.6_SLV.xlsx'. Percentages for each of the four values were manually calculated. A line chart was produced in excel from the percentages, which matched the values in the paper.
- **Figure 2.6f Reproduced, with manual changes.** The output was produced as 'Figure 1.6_PER.xlsx'. Percentages for each of the four values were manually calculated. A line chart was produced in excel from the percentages, which matched the values in the paper.

- **Figure 2.7a Reproduced, with manual changes.** The output was produced as 'Figure 1.7_BRA.xlsx'. Sum total across the four categories (*Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating*) were calculated for each of the 10 percentiles. A stacked box chart was created in excel, which matched the values in the paper.
- **Figure 2.7b Reproduced, with manual changes.** The output was produced as 'Figure 1.7_COL.xlsx'. Sum total across the four categories (*Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating*) were calculated for each of the 10 percentiles. A stacked box chart was created in excel, which matched the values in the paper.
- **Figure 2.7c Reproduced, with manual changes.** The output was produced as 'Figure 1.7_MEX.xlsx'. Sum total across the four categories (*Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating*) were calculated for each of the 10 percentiles. A stacked box chart was created in excel, which matched the values in the paper.
- **Figure 2.7d Reproduced, with manual changes.** The output was produced as 'Figure 1.7_URY.xlsx'. Sum total across the four categories (*Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating*) were calculated for each of the 10 percentiles. A stacked box chart was created in excel, which matched the values in the paper.
- **Figure 2.7e Reproduced, with manual changes.** The output was produced as 'Figure 1.7_SLV.xlsx'. Sum total across the four categories (*Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating*) were calculated for each of the 10 percentiles. A stacked box chart was created in excel, which matched the values in the paper.
- **Figure 2.7f Reproduced, with manual changes.** The output was produced as 'Figure 1.7_PER.xlsx'. Sum total across the four categories (*Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating*) were calculated for each of the 10 percentiles. A stacked box chart was created in excel, which matched the values in the paper.
- **Figure 2.8a Reproduced.**
- **Figure 2.8b Reproduced.**
- **Figure 2.8c Reproduced.**
- **Figure 2.8d Reproduced.**
- **Figure 2.8e Reproduced.**
- **Figure 2.8f Reproduced.**
- **Figure 2.9a Reproduced, with manual changes.** The figure is compared to 'Figure 1.9_BRA.xlsx'. A line graph was created manually using the excel output. All observations reproduced.
- **Figure 2.9b Reproduced, with manual changes.** The figure is compared to 'Figure 1.9_COL.xlsx'. A line graph was created manually using the excel output. All observations reproduced.
- **Figure 2.9c Reproduced, with manual changes.** The figure is compared to 'Figure 1.9_MEX.xlsx'. A line graph was created manually using the excel output. All observations reproduced.

- **Figure 2.9d Reproduced, with manual changes.** The figure is compared to 'Figure 1.9_URY.xlsx'. A line graph was created manually using the excel output. All observations reproduced.
- **Figure 2.9e Reproduced, with manual changes.** The figure is compared to 'Figure 1.9_SLV.xlsx'. A line graph was created manually using the excel output. All observations reproduced.
- **Figure 2.9f Reproduced, with manual changes.** The figure is compared to 'Figure 1.9_PER.xlsx'. A line graph was created manually using the excel output. All observations reproduced.

Chapter 3

- **Figure 3.1 Does not Apply.** No analytical output was created in this figure.
- **Figure 3.2a Reproduced, with manual changes.** The figure is compared to 'Figure 2.2.xlsx'. A scatter with only markers was generated in excel using the 'Diff_hiring' and 'LM_rigidity_index' columns. Data labels were also added manually from the 'economycode' column.
- **Figure 3.2b Reproduced, with manual changes.** The figure is compared to 'Figure 2.2.xlsx'. A scatter with only markers was generated in excel using the 'Rig_hiring' and 'LM_rigidity_index' columns. Data labels were also added manually from the 'economycode' column.
- **Figure 3.2c Reproduced, with manual changes.** The figure is compared to 'Figure 2.2.xlsx'. A scatter with only markers was generated in excel using the 'Diff_dismiss' and 'LM_rigidity_index' columns. Data labels were also added manually from the 'economycode' column.
- **Figure 3.2d Reproduced, with manual changes.** The figure is compared to 'Figure 2.2.xlsx'. A scatter with only markers was generated in excel using the 'Diss_costs' and 'LM_rigidity_index' columns. Data labels were also added manually from the 'economycode' column.
- **Figure 3.3 Does not Apply.** No analytical output was created in this figure.

Chapter 4

- **Figure 4.1 Reproduced, with manual changes.** This figure was manually compiled using information from the OECD, and it was manually reproduced using instructions in the README file using data from Augmented_tax_wedge.xlsx.
- **Table 4.1 Does not Apply.** No analytical outputs were created in this figure.
- **Figure B4.1.1 Reproduced, with manual changes.** This figure was manually compiled using information from the OECD, and it was manually reproduced using instructions in the README file using data from Tax_wedge_minimum_wage_BR.xls.
- **Figure 4.2 Reproduced, with manual changes.** This figure was manually compiled using information from countries' labor laws, and it was manually reproduced using instructions in the README file using data from Tax_wedge_minimum_wage_single_earner.xls.
- **Figure 4.3 Reproduced, with manual changes.** This figure was manually compiled using information from the countries' teams, and it was manually reproduced using instructions in the README file using data from Tax_wedge_minimum_wage_adults&child.xls.

- **Figure 4.4** *Reproduced, with manual changes*. This figure was manually compiled using information from the countries' teams, and it was manually reproduced using instructions in the README file using data from Tax_wedge_minimum_wage_fullvaluation.xls.

Chapter 5

- **Figure 5.1a** *Reproduced*. The exhibit was compared to 'Boxplots_dependent.gph'.
- **Figure 5.1b** *Reproduced*. The exhibit was compared to 'Boxplots_independent.gph'.
- **Figure 5.2a** *Reproduced*. The exhibit was compared to 'Figure 4.3_all.png'.
- **Figure 5.2b** *Reproduced*. The exhibit was compared to 'Figure 4.3_all.png'.
- **Figure 5.3** *Reproduced*. The exhibit was compared to 'Figure 4.4_all.png'.
- **Figure B5.4.1** *Reproduced, with manual changes*. The exhibit was compared to 'Figure 4.6_all.png'. Colors representing each of the countries were changed between the output and the final paper. There were also changes to the size of the dots in the final plot to improve visibility.
- **Figure 5.4** *Reproduced, with manual changes*. The exhibit was compared to 'Figure 4.5_all.png'. The line plots for the independent values were changed from straight lines to dotted lines in the final output.

Chapter 6

- **Table 6.1** *Does not Apply*. No analytical outputs were contained in this figure.
- **Figure 6.1a** *Reproduced*.
- **Figure 6.1b** *Reproduced*.
- **Figure 6.2a** *Reproduced*.
- **Figure 6.2b** *Reproduced*.
- **Figure 6.3a** *Reproduced*.
- **Figure 6.3b** *Reproduced*.
- **Figure 6.3c** *Reproduced*.
- **Figure 6.4a** *Reproduced*.
- **Figure 6.4b** *Reproduced*.
- **Figure 6.5a** *Reproduced*.
- **Figure 6.5b** *Reproduced*.
- **Figure 6.6a** *Reproduced*.
- **Figure 6.6b** *Reproduced*.

- Figure 6.7a Reproduced.
- Figure 6.7b Reproduced.
- Figure 6.8a Reproduced.
- Figure 6.8b Reproduced.
- Figure 6.8c Reproduced.
- Figure 6.8d Reproduced..
- Figure 6.9a Reproduced.
- Figure 6.9b Reproduced.
- Figure 6.9d Reproduced.
- Figure 6.10a Reproduced.
- Figure 6.10b Reproduced.
- Figure 6.11a Reproduced.
- Figure 6.11b Reproduced.
- Figure 6.11c Reproduced.
- Figure 6.12a Reproduced.
- Figure 6.12b Reproduced.
- Table 6.2 Does not show analysis results.
- Figure 6.13a Reproduced.
- Figure 6.13b Reproduced.
- Table 6.3 Does not show analysis results.
- Figure 6.14a Reproduced.
- Figure 6.14b Reproduced.
- Figure 6.14c Reproduced.
- Figure 6.14d Reproduced.
- Figure 6.15a Reproduced.
- Figure 6.15b Reproduced.
- Figure 6.15c Reproduced.
- Figure 6.15d Reproduced.
- Figure 6.16a Reproduced.
- Figure 6.16b Reproduced.

Chapter 7

- **Figure 7.1 Reproduced, with manual changes.** This figure was manually compiled using information from the OECD (Report here), and it was manually reproduced using instructions in the README file using data from `Augmented_dejur_tax.xlsx`.
- **Figure 7.2** Does not apply Exhibit is sourced from another paper.
- **Table 7.1** Does not show analysis results.
- **Figure 7.3** Does not apply Exhibit is sourced from another paper.
- **Table 7.2** Does not show analysis results.
- **Table 7.3 Reproduced, with manual changes.** The code produces each observation as a unique excel file. The table was manually formatted and compiled by the authors. The exhibit is a compilation from `'Table7_3_a.xls'`, `'Table7_3_b.xls'`, `'Table7_3_c.xls'`, `'Table7_3_d.xls'`, `'Table7_3_e.xls'`, `'Table7_3_f.xls'`
- **Figure 7.4a Reproduced.** The exhibit was compared to `'Figure 7_4.png'`.
- **Figure 7.4b Reproduced.** The exhibit was compared to `'Figure 7_4.png'`.
- **Figure 7.4c Reproduced.** The exhibit was compared to `'Figure 7_4.png'`.
- **Figure 7.4d Reproduced.** The exhibit was compared to `'Figure 7_4.png'`.
- **Figure 7.4e Reproduced.** The exhibit was compared to `'Figure 7_4.png'`.

Chapter 8

- **Figure B8.1.1** Does not show analysis results
- **Figure 8.1 Reproduced with manual changes.** This exhibit was manually created in Excel (see Tables&Figures sheet, cell reference F8.1), using data generated by the code (`Figure8.1.xlsx`). The replicators verified that the underlying data used in the figure matches exactly the data produced by the code.
- **Figure 8.2 Reproduced with manual changes.** This exhibit was manually created in Excel (see Tables&Figures sheet, cell reference F8.2). The replicators verified that the underlying data used in the figure matches exactly the data produced by the code.
- **Figure 8.3** Does not apply Exhibit is sourced from another paper.
- **Figure 8.4** Does not apply Exhibit is sourced from another paper.
- **Figure 8.5a Reproduced with manual changes.** This exhibit was manually created in Excel (see Tables&Figures sheet, cell reference F8.5a). The replicators verified that the underlying data used in the figure matches exactly the data produced by the code (`Figure8.5a.xlsx`).
- **Figure 8.5b Reproduced with manual changes.** The replicators verified that the underlying data used in the figure matches exactly the data produced by the code (`Figure8.5b.xlsx`).

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 147160, the Stata version used was 18.5, and the exhibits selected were: Figure 2A.3, Table 2A.2, Figure 3A.1, Table 4A.3, Table 4B.1, Table 4B.3, Table 4B.6, Table 4B.7, Figure 4C.2, and Figure 4C.4.

- **Figure 2A.3a Reproduced, with manual changes.** The output was produced as **Figure A1.3_BRA.xlsx**'. Sum total across the four categories (Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating) were calculated for each of the 10 percentiles. A stacked box chart was created in Excel, which matched the values in the paper
- **Figure 2A.3b Reproduced, with manual changes.** The output was produced as **Figure A1.3_COL.xlsx**'. Sum total across the four categories (Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating) were calculated for each of the 10 percentiles. A stacked box chart was created in Excel, which matched the values in the paper
- **Figure 2A.3c Reproduced, with manual changes.** The output was produced as **Figure A1.3_SVL.xlsx**'. Sum total across the four categories (Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating) were calculated for each of the 10 percentiles. A stacked box chart was created in Excel, which matched the values in the paper
- **Figure 2A.3d Reproduced, with manual changes.** The output was produced as **Figure A1.3_MEX.xlsx**'. Sum total across the four categories (Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating) were calculated for each of the 10 percentiles. A stacked box chart was created in Excel, which matched the values in the paper
- **Figure 2A.3e Reproduced, with manual changes.** The output was produced as **Figure A1.3_PER.xlsx**'. Sum total across the four categories (Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating) were calculated for each of the 10 percentiles. A stacked box chart was created in Excel, which matched the values in the paper
- **Figure 2A.3f Reproduced, with manual changes.** The output was produced as **Figure A1.3_URY.xlsx**'. Sum total across the four categories (Mandated and complying, mandated and noncomplying, exempt and participating, exempt and nonparticipating) were calculated for each of the 10 percentiles. A stacked box chart was created in Excel, which matched the values in the paper
- **Table 2A.2 Reproduced, with manual changes.** The code produced four Excel sheets, each representing one of the four characteristics (gender, age, area, level of education). Manually percentages were created for each category for the columns 'Mandated & complying' and 'Mandated & non-complying'. Separately, each category of values for the columns 'Exempt & participating' and 'Exempt & non-participating' were also totalled to create percentages. All decimals were rounded to whole numbers. After manual calculations in Excel, the exhibits were deemed reproduced.
- **Figure 3A.1 Does not Apply.** This item does not represent analytical outputs.
- **Table 4A.3 Does not Apply.** This item does not represent analytical outputs.
- **Table 4B.1 Does not Apply.** This item does not represent analytical outputs.

- **Table 4B.3** Does not Apply. This item does not represent analytical outputs.
- **Table 4B.6** Does not Apply. This item does not represent analytical outputs.
- **Table 4B.7** Does not Apply. This item does not represent analytical outputs.
- **Figure 4C.2a** Reproduced.
- **Figure 4C.2b** Reproduced.
- **Figure 4C.2c** Reproduced.
- **Figure 4C.2d** Reproduced.
- **Figure 4C.2e** Reproduced.
- **Figure 4C.2f** Reproduced.
- **Figure 4C.2g** Reproduced.
- **Figure 4C.4** Reproduced, with manual changes. This figure was manually compiled using information from the countries' teams, and it was manually reproduced using instructions in the README file using data from Tax_wedge_minimum_wage_fullvaluation.xls.

Reproduction Environment

- Paper exhibits were reproduced in a computer with the following specifications:
- OS: Windows 11 Enterprise
- Processor: Intel(R) Core(TM) i5-1145G7 CPU @ 2.60GHz
- Memory available: 15.7 GB
- Software version: Stata version 18 MP