

## READ ME

The following file explains how to get the regression results of the paper “Explaining Gender Differences in Economic Outcomes in Burkina Faso.” The paper analyzed two Burkinabe datasets. Both the surveys were conducted by the National Institute of Statistics and Demography (INSD), the national statistical agency of Burkina Faso:

1. **“Enquête Harmonisée sur les Conditions de Vie des Ménages 2018-2019” (EHCVM):** The EHCVM are a multi-country nationally representative surveys. The countries are members of the West African Economic Monetary Union (WAEMU). In this round, about 7,000 households were sampled in two stages. Five hundred eighty five enumeration areas (EAs) were randomly chosen in the first then, and consequently, 12 households were randomly chosen from each EA. More information can be obtained from the following link:  
<https://microdata.worldbank.org/index.php/catalog/4290>
2. **The 2014 (LSMS) Enquête Multisectorielle Continue:** The 2014 LSMS is a nationally representative survey implemented by INSD with funding from the Swedish Cooperation. A total of 10,860 households were surveyed in this survey. In this survey, a two-stage sampling method was used to identify households and collect data. In the first stage, a total of 905 EAs were randomly chosen. Subsequently, 12 households were selected from each EA in the second stage. More information can be obtained from the following link:  
<https://microdata.worldbank.org/index.php/catalog/2538>

The do file “**0.Master File.do**” also has instructions on how to obtain the regression results.

The user must open the”**0.Master File.do**” and change the path names in line 2. The path name tell Stata where the reproducibility package files are stored.

Once the path name is changed, the user can run the do command on “**0.Master File.do**” and it will execute the other do files, namely,

- **Step 1 - Cleaning EHCVM.do** - this do file cleans the ehcvms files and then saves them in the folder specified in the reproduce\_files pathname
- **Step 2 - Cleaning LSMS.do** - this do file cleans the lsms files and then saves them in the folder specified in the reproduce\_files pathname
- **Step 3 – Figure 1 Tables 1A 2 and App Table 1.do** - runs the regressions comparing the raw gender gap in ehcvms, and compares the gender gap between lsms and ehcvms

Additionally, it creates a number of summary stats tables:

- Excel file Table 1a.xls, which is the Table 1a in the paper. Note: the values of overall lfp and paid lfp are already converted to percentage points using mfx, but the values for business revenues, harvest value and crop sold value need to be converted to percents using "exp(coefficient) - 1" transformation. Since the wage gap proportion is found using the formula "coefficient on female dummy / constant \* 100"
- Table 2 - the wage gap coefficients and their t-values for the year 2014, and those of 2018. The pvalue of the equality of coefficients between 2014 and 2018 are also calculated and reported.
- Appendix table 1 - this shows the urbanicity of provinces in 2014 and 2018.

- Figure 1 - the numbers used to draw the venn diagram are calculated and stored in the excel file 'figure 1.' the user can draw the venn diagram in word.
- **Step 4. KOB Regressions.do** - runs the Kitagawa-Oaxaca-Blinder decomposition of EHCVM and finds the summary statistics. This do file creates the following:
  - Table 1b - the summary statistics (separate files created for paid only lfp, overall lfp, wages, business revenues, harvest value, crop sales. In each excel file, the first three column is for the full population, then next three (4-6) is for the male population, and then the next three (7-9) is for the female population. The next three (10-12) is for all 15-29 population, then (13-15) is for the male 15-29 year olds, and then (columns 16-18) is for the females aged 15-29. The subsequent columns (19-21) is for all population aged 30-64, then (columns 22-24) is for males aged 30-64, and then (columns 25-27) is for females aged 30-64)
  - Table 3 - paid lfp and overall lfp decompositions
  - Table 4 - wage decomposition
  - Table 5 - business revenues decomposition
  - Table 6 - harvest value and crop sales decompositions
  - Appendix table 2 - overall lfp decomposition by age
  - Appendix table 3 - overall lfp decomposition by urban/rural
  - Appendix table 4 - wage decomposition by age
  - Appendix table 5 - wage decomposition by urban/rural
  - Appendix table 6 - business revenues decomposition by age
  - Appendix table 7 - business revenues decomposition by urban/rural
  - Appendix table 8 - harvest value decomposition by age
  - Appendix table 9 - crop sales decomposition by age

The user does not need to change any of the other do files in order to get the regression results. The user can take a look at the four other do files to understand how the results were obtained, and if the user wishes to make any necessary changes.

Once this file is run, all the tables and all the appendix tables that were used in this paper will be generated and stored in the folder specified in the master file.do. The files are in XLS format, and they are numbered in the same way as their location in the paper.