Replication package for "Do bilateral labor agreements increase migration? Global evidence from 1960 to 2020"

Samik Adhikari, Narcisse Cha'ngom, Heidi Kaila, and *Maheshwor Shrestha*

1. Data

This paper uses data on bilateral migration stocks, covering the period from 1960 to 2020 in ten-year intervals. The data on bilateral labor agreements (BLAs) come from Chilton and Woda (2022), comprising 1,222 agreements signed between countries since World War II including information on the year of each BLA and the countries involved. These data are publicly available from the following sources:
- Bilateral Labor Agreements data 1945-2020 available at:
https://dataverse.harvard.edu/dataverse/BLA - Bilateral Migration data 1960-2000 available at: https://databank.worldbank.org/source/global-bilateral-migration - Bilateral Migration data 2010-2020 available at:
https://www.un.org/development/desa/pd/content/international-migration-1

We follow the data cleaning process outlined below to create the combined data used for analysis:

Data cleaning process for BLA data

The raw dataset, "BLAs_Formation_Data," is cleaned in four steps.

- First, ISO3 codes are assigned to origin and destination countries following World Bank standards using the excel file "countries_match.xlsx" to ensures that the country spelling the BLA that, in some cases differ with the World Bank spelling does not prevent the merge.

- Second, country names are harmonized to reflect historical changes, such as the breakup of Yugoslavia (now Serbia, Montenegro, Croatia, Slovenia, North Macedonia, and Bosnia and Herzegovina) and Czechoslovakia (now the Czech Republic and Slovakia).

- Third, overseas territories are recoded as their metropolitan counterparts for consistency with migration data (e.g., Guam and Puerto Rico as the USA, Greenland as Denmark, and Guadeloupe as France). Territories linked to the Netherlands, the UK, and New Zealand are similarly adjusted.

- The coding norms used for this process is identical to what was used for the Bilateral Migration data (see Ozden et al, 2011 for details on this).

- Fourth, to match the migration data format, we transform the BLA data by identifying the country of origin, the country of destination, and aligning each BLA with the corresponding year in the migration data. For instance, a BLA signed

between 1970 and 1979 is assigned to t=1980 in the migration dataset. For consistency, we classify the country with the lower GDP per capita as the "origin" country and the other as the "destination" country. Origin-destination-year cells without a recorded BLA are assumed to have no agreement in place.

Data clenaning process for migration data

Migration stock data for 1960, 1970, 1980, 1990, and 2000 are sourced from the World Bank (2022), while data for 2010 and 2020 are drawn from UNDESA (2020), which employs a similar methodology. They use information collated from censuses and surveys at destination and origin countries to construct a matrix of bilateral migration stock data from each of the origin countries to each potential destination country.

Analysis data

The combined data used for most of the analysis in the paper, "BLA_Migration_data.dta", is provided with this repository under the folder "DATA".

- In addition to the migration and BLA variables from the above mentioned datasets, the data also includes the government effectiveness score, gee, from the Worldwide Governance Indicators, 2023 Update. Full documentation and access to the aggregate governance indicators, and the underlying source data, is available at www.govindicators.org. Source: Daniel Kaufmann, Aart Kraay and Massimo Mastruzzi (2010). "The Worldwide Governance Indicators : A Summary of Methodology, Data and Analytical Issues". World Bank Policy Research Working Paper No. 5430, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682130

List of variables included, souce and construction details

Below you will find two tables with details on variables with their storage type, display format, labels, original source, and information on how they are computed.

Table 1: Variable Names, Labels, and Types

Variable	Storage Type	Variable Label
d_iso3	str3	Destination ISO3
o_iso3	str3	Origin ISO3
BLAs_signed	byte	Number of signed Bilateral Labor Agreements
year	int	Year
o_region	str3	Origin Region
d_region	str3	Destination Region
mig	float	Migrant stock
dum_BLAs_10	byte	BLA in the past decade
dum_5	byte	BLA in the past 5 years
AFR_o	byte	Origin in Africa
AFR_d	byte	Destination in Africa

Variable	Storage Type	Variable Label
dum90	byte	After 1990
gee	float	Government Effectiveness Score
devcat	byte	Origin income category
devcatd	byte	Destination income category
GCC_d	byte	Destination in the GCC

Table 2: Details on Construction

Variable	Details on Construction
d_iso3	Country codes + destination
o_iso3	Country codes + origin
BLAs_signed	From `BLAs_Formation_Data.dta` (raw data) as described in the construction section
year	Year (1960,1970,1980,1990,2000,2010 or 2020)
o_region	Region of the origin country
d_region	Region of the destination country
mig	Data sourced from multiple sources, e.g., UNDESA for Uganda (destination: UGA, origin: CON, mig: 159519)
dum_BLAs_10	Dummy variable indicating if BLA was signed in the past decade, source: `BLAs_signed`
dum_5	Dummy variable indicating if BLA was signed in the past 5 years, source: `BLAs_signed`
AFR_o	Dummy variable indicating if the origin country is in Africa (1)
AFR_d	Dummy variable indicating if the destination country is in Africa (1)
dum90	Dummy variable indicating if the year is after 1990 (1)
gee	World Development Indicators
devcat	Income category of the origin country (WB country groups)
devcatd	Income category of the destination country (WB country groups)
GCC_d	Dummy variable indicating if the destination is one of the GCC countries (Bahrain, Kuwait, etc.)

Example: Migrant Stock for Uganda

The `mig` variable represents the migrant stock. For example:

- **Destination Country**: Uganda (ISO3: UGA)

- **Origin Country**: Democratic Republic of the Congo (ISO3: CON)

- **Year**: 2010

- **Migrant Stock Value**: 159,519 (For this year available at DATA/undesa_pd_2020_ims_stock_by_sex_destination_and_origin.xslx)

In this case, the region for both countries is **Sub-Saharan Africa (SSF)**. The data for 2010 and 2020 is sourced from UNDESA. For earlier years (1960, 1970, 1980, 1990, and 2000), similar data is sourced from the other two sources described in the documentation.

The intermediate data includes this processed information, and the same methodology is applied consistently across all years and data sources. Detailed information on these processes can be found in the README file.

Example: Migrant Stock for Uganda 2010

The `mig` variable represents the migrant stock. For example:

- **Destination Country**: Uganda (ISO3: UGA)
- **Origin Country**: Democratic Republic of the Congo (ISO3: CON)
- **Year**: 2010 (dummy after 1990 == 1)
- **Migrant Stock Value**: 159,519
- No Bilateral Labor Agreements signed == 0 (dummys related == 0)
- In this case, the region for both countries is **Sub-Saharan Africa (SSF)**. The data for 2010 and 2020 is sourced from UNDESA. For earlier years (1960, 1970, 1980, 1990, and 2000), a similar process is follow from the other two sources described in the documentation.
- Income category for both **LLMIC**.

As the year is 2010 (and for 2020) the migrant stock data is sourced from UNDESA (DATA/undesa_pd_2020_ims_stock_by_sex_destination_and_origin.xlsx), but the method is exactly the same for the rest of the years 1960,1970,1980,1990,2000 with WB data (DATA/P_Data_Extract_From_Global_Bilateral_Migration.xlsx) For instance, migrant stock for UGA and AGF, 1960 is 1.

2. Instructions for Replication

i. Download the entire package

Download the entire content of this package "Replication File BLA-Migration" to a location in your computer. The package contains four folders:

- DATA: this folder contains the dataset used throughout the paper (BLA_Migration_data.dta)

- DOFILES: this folder contains seven .do files (program to run) and one excel sheet that accompany welfare calculations reported in section five of the paper.

- FIG: this folder will contain after the replication of the code the nine figures of the results section (main text + appendix) named accordingly.

- REGOUT: this folder will contain 10 tables reported in the result section of the paper (main text + appendix).

ii. Set directory reference and install user-written packages
- Open "Master_file.do" and change the directory path in line 22 to be the corresponding location in your computer where you store the "Replication File BLA-Migration" folder.
- Ensure all the user written Stata Packages needed for the analysis are installed and are up to date. The file "master_file.do" automatically checks and installs all required user-written commands from SSC that are not installed in your machine. Note that some of the user-written packages are not automatically updated to the latest version in SSC (e.g., `ivreghdfe`). If you run into any issues with the SSC versions of these commands, you may want to install the latest stable version that is actively maintained by the authors.

iii. Execute the "Master_file.do"
- Run the "Master_file.do" in Stata that will replicate all tables and figures in the paper. The file will in turn call the following do-files:
  - "Header.do" to: Load the analysis dataset and create high dimensional fixed effects and macros.
  - "Descriptives_Tab1_Tab_B1.do.do" to: Create Table 1 and Table B.1.
  - "Main_analysis.do" to: Create Table 2, Table B.2, Table B.3 and Table 3.
  - "Heterogeneity_analysis.do" to: Create Table 4, Table 5, Table B.4, Table B.5, and Table B.6.
  - "Event_study.do" to: Create Figure 1 and Figure 2 (panels Figure_2A and Figure_2B).
  - "Heterogeneity_robust_DiD.do" to: Create Table C.1, Figure C.1, Figure C.2, Figure C.3 (panels Figure_C3A, Figure_C3B, Figure_C3C, and Figure_C3D)
  - "welfare calculations.do" to produce the necessary descriptive statistics for welfare numbers reported in the section 5 of the paper.
  - "Appendix_A.do" to: creates Figures in appendix A (Figure A.1-A.6) using raw data "BLAs_Formation_Data.dta" and "BLAs_Coding_Data.dta" provided in "DATA" folder.

3. Output
- 3.1 Tables
  - Table 1: "REGOUT/Descs.tex"
  - Table 2: "REGOUT/Table_2.tex"
  - Table 3: "REGOUT/Table_3.tex"
  - Table 4: "REGOUT/Table_4.tex"
  - Table 5: "REGOUT/Table_5.tex"

- – Table B.1: "REGOUT/Descriptives_decadal.tex"
- – Table B.2: "REGOUT/Table_B2.tex"
- – Table B.3: "REGOUT/Table_B3.tex"
- – Table B.4: "REGOUT/Table_B4.tex"
- – Table B.5: "REGOUT/Table_B5.tex"
- – Table B.6: "REGOUT/Table_B6.tex"
- – Table C.1: "REGOUT/Table_C1.tex"
- 3.2 Figures
  - – Figure 1: "FIG/Figure_1.png"
  - – Figure 2: "FIG/Figure_2A.png"; "FIG/Figure_2B.png"
  - – Figure A.1 : "FIG/Figure_A1.png"
  - – Figure A.2 : "FIG/Figure_A2.png"
  - – Figure A.3 : "FIG/Figure_A3.png"
  - – Figure A.4 : "FIG/Figure_A4.png"
  - – Figure A.5 : "FIG/Figure_A5.png"
  - – Figure A.6 : "FIG/Figure_A6.png"
  - – Figure C.1: "FIG/Figure_C1.png"
  - – Figure C.2: "FIG/Figure_C2.png"
  - – Figure C.3: "FIG/Figure_C3A.png"; "FIG/Figure_C3B.png"; "FIG/Figure_C3C.png"; "FIG/Figure_C3D.png"
- 3.3 Welfare calculations in Section 5
  - – The excel file provided "welfare calculations.xlsx" under the folder "DOFILES" generate the relevant welfare numbers used in section 5 of the paper. The excel sheet uses the results of the regressions and the descriptive statistics produced by "welfare calculations.do" for the calculations. The regression results and descriptive statistics are manually copied into the relevant cells in the excel sheets.

4. Runtime

Machine details:

Processor: 12th Gen Intel(R) Core (TM) i5-1245U 1.60 GHz

Installed RAM: 16.0 GB

System type: 64-bit operating system, x64-based processor

Software: STATA MP 18.0

Runtime: 8 to 10 minutes.