

## Creation of the ladder\_cross\_country.dta file

The ladder\_cross\_country.dta file is extracted from two sets of processes run by the team.

The first is the creation of the third version of the Jobs Indicators Database (JOIN V3). JOIN is a World Bank effort to leverage all available harmonized household surveys to create a comprehensive and comparable database of indicators, centred on labour market information. It draws from harmonized surveys from the three World Bank collections: the International Income Distribution Database (I2D2), the Global Monitoring Database (GMD), and the Global Labor Database (GLD).

The second is the creation of parallel database to JOIN that would include indicators structured around the concept of ladders or tiers of labour markets (of varying levels of prestige and formality) for a World Bank flagship report.

The codes included have been taken from these efforts but numbered to indicate the process. Each is described in the following by the blocks denoted by the first number (i.e., 01 to 07).

- Block 01 – Creation of a list of surveys to include

This block is taken from the efforts to create JOIN V3 and consists of four files. The paths are set to extract information from the GLD server (accessible in its entirety only to the GLD leads) and the I2D2 server (accessible to former collaborators of the I2D2 harmonization effort), and GMD via datalibweb. The particular files do the following:

- Create GLD List

The code reads through the GLD server (as of May 2023) and lists all surveys to be used.

- Compare GLD to JOIN V2

The code compares the list of GLD surveys to what was in JOIN V2 to highlight differences in naming conventions (e.g., Kosovo listed as KOS in one XKX in other) to ensure later efforts can correctly identify a unique set of surveys.

- Create GMD survey list

The code lists the datalibweb repository for GMD and reduces it to the set of useable GMD surveys (e.g., surveys with only wealth aggregates lack labour information and thus are not listed).

- Compare GMD and GLD to JOIN V2

The code compares both GLD and GMD lists to JOIN V2 to (a) ensure a unified spelling and (b) denote where GLD and GMD overlap for internal controls.

- Block 02 – Accessing and storing listed surveys

Block 02 is taken from the code to produce JOIN V3 and serves to store in a separate folder, grouped by collection, a copy of all surveys to be used. It consists of five files.

- Download GMD surveys

The code goes through the GMD list and downloads all surveys if (a) they are accessible to the user and (b) have at a minimum the “lstatus” and “empstat” variables.

Note that process (a) is dependent on the user, that is, the number of surveys a user has access to is different and depends on which surveys they have requested and been given access to. A user who works for the GMD harmonization effort will likely have access to many more surveys than the average user. However, the user running the code (UPI 529028) has not requested more than a handful of surveys and thus the number of surveys downloaded should be similar for any other World Bank user.

Note further that the GMD team may have changed the access rights for certain surveys and thus some surveys downloaded may not be accessible anymore, while others that could not be downloaded are accessible now. The set of surveys stored represent the state of possible surveys for user 529026 in May of 2023.

- Transfer GLD surveys

The code stores all GLD surveys in a separate folder.

- Transfer I2D2 surveys

The code stores all I2D2 servers in a separate folder.

- Transfer missed I2D2 surveys

The code identifies surveys that were included from I2D2 in JOIN V2 but that are not in the list previously made, given the lack of a unified naming convention on the I2D2 server. It identifies missing files manually and transfers them to a separate folder.

- Unite missing I2D2 surveys into a single group

The code corrects the previous file’s process of creating separate folders and unifies the missing (or better: overlooked) surveys into a single folder.

- Block 03 – Appending stored surveys in groups

Block 03 is taken from the code to produce JOIN V3 and serves to append surveys by collection into grouped data files. It consists of six folders.

- Append GLD Add files

The code takes all GLD surveys for country/ year combination that did not exist in JOIN V2 and appends them into a single file.

- Append GLD Replace files

The code takes all GLD surveys for country/year combinations that did exist in JOIN V2 and appends them into a single file.

- Append GMD files

The code appends all GMD surveys files into a single file.

- Append I2D2 additions

The code appends all I2D2 surveys not in JOIN V2 into a single file.

- Append I2D2 older surveys

The code appends all I2D2 surveys in JOIN V2 into a series of larger files (single file would be too large).

- Append I2D2 missing surveys

The code appends all I2D2 surveys that the previous step had identified as missing and that needed to be manually transferred.

- Block 04 – Reduce appended surveys to a set of useable surveys

Block 04 builds on the JOIN V3 creation effort. JOIN V3, after creating indicators goes through a quality check procedure. The indicators created are compared to ILO (International Labour Organisation) and WDI (World Development Indicators) data and scored based on the difference to these indicator databases. Based on these scores, the sample of surveys is reduced to a set of the most robust surveys.

Additionally, the results undergo a process of manual inspection of key labour market results. The surveys with implausible information (e.g., because WDI also has implausible information and thus the difference score was low) are dropped.

The code (single code in the block) uses this list of acceptable surveys to drop from the surveys appended in the previous step all those deemed of insufficient quality.

Both the file containing the acceptable surveys as well as the quality check code are available on request from [mgronert@worldbank.org](mailto:mgronert@worldbank.org).

- Block 05 – Create ladder variables for each group

Block 05 is taken from the efforts to create jobs ladder information for a Flagship report. The single code in the block creates the relevant ladder assignment (i.e., assign each individual to which ladder they belong) based on the ladder definitions present in the working paper.

Note that this code reads in the Global Macro Indicator Database created by the Jobs Group. The file is stored here along the codes.

- Block 06 – Create indicators out of each group

Block 06 is taken as well from the efforts to create the Flagship report. The single code in the block collapses the individual level information from the previous step into a set of indicators, such that each row represents a unique country/year combination.

- Block 07 – Include IMF CPI and PPP information

Block 07 was created specifically to incorporate the IMF World Economic Outlook Database. The data (as the single code in the block notes) is taken from <https://www.imf.org/en/Publications/WEO/weo-database/2023/October/download-entire-database> and was downloaded as an Excel file on the 22<sup>nd</sup> of March 2024. It was manually converted into a CSV file, which the code reads in.

For any further details or questions, please reach out to Mario Gronert ([mgronert@worldbank.org](mailto:mgronert@worldbank.org)).