

Readme File for Replication Package for 2024 Update of Worldwide Governance Indicators

August 21, 2024

Download contents of .zip archive into desired folder on your computer. The folder will contain:

1. **readme.docx** – this file.
2. **wgicode1.do** – this is the master .do file that needs to be executed to generate aggregate indicators. An overview of the code structure is at the top of this file.
3. **estimate1.do, importmerge1.do, ucmcode1.do, nonrepcode1.do, rankcalc1.do, exporter1.do** – these are all subsidiary .do files called by wgicode.do. Descriptions of these files are included at the top of each file.
4. **countrycodes.xlsx** – this file contains the country codes and country names as used in the WGI.
5. **sourcedata** – this folder contains 35 Excel files **XXX.xlsx** where XXX is a three-letter code corresponding to each of the 35 WGI data sources. The Excel files contain the data from each source and are posted separately on the WGI website as well. This folder also contains a copy of **countrycodes.xlsx**.
6. **sourcedofiles** – this folder contains 34 Stata .do files **XXX_merge.do**. These files are called by **importmerge1.do** and are used to retrieve the data from each Excel file into Stata format.
7. **reproducibilitycheck.pdf** – this file contains the summary findings of a code review of the replication package conducted by DIMEAnalytics at the World Bank

For a complete technical description of the Unobserved Components Model used to construct the aggregate WGI indicators, please refer to Section 3 of [Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobaton \(1999\). “Aggregating Governance Indicators”. World Bank Policy Research Working Paper No. 2195.](#)

To replicate the WGI aggregation calculations:

1. At the beginning of wgicode1.do, set the global \$filepath to match the folder to which you have downloaded the replication package on your computer.
2. Open Stata and run wgicode1.do.

This will generate the following output files that will be stored in the same folder to which you downloaded the replication package:

1. **wgidataset.dta, wgidataset.xlsx** – Stata and Excel versions of dataset containing the six aggregate indicators, standard errors, number of sources, and percentile ranks. This dataset is in “long” format with countries and time periods stacked vertically. These files are posted on the WGI homepage at www.govindicators.org.
2. **wgidataset_with_sourcedata.dta, wgidataset_with_sourcedata.xlsx** – Stata and Excel versions of dataset containing the six aggregate indicators, standard errors, number of sources, and percentile ranks, as well as the source data for each data source. This dataset is in “long” format with countries, time periods and WGI aggregate indicators periods stacked vertically. These files are posted on the WGI homepage at www.govindicators.org.

3. **parameterestimates.dta, parameterestimates.xlsx** – Stata and Excel versions of a dataset containing the parameter estimates of the unobserved components model (α, β, σ) for all indicators, sources, and years. Data is in “long” format with indicators, sources and years stacked vertically. These files are linked in the “Documentation” tab of www.govindicators.org.
4. **wgisourceweights.xlsx** – this is an Excel file containing the average (across countries) of the weights assigned to each data source for each source and WGI aggregate indicator. This dataset is in “long” format with years and aggregate indicators stacked vertically. This file is linked in the “Documentation” tab of www.govindicators.org.
5. **sourcedata.dta** -- this is a Stata dataset containing all the individual indicators used to calculate the aggregate indicators. This dataset is in “wide” format with countries in rows and separate columns for each variable and period. Subscripts “1” and “2” appended to variable names indicate whether the data sources is treated as globally representative or non-representative in the aggregation procedure. This file does not contain variable labels and is used primarily as an intermediate step in the calculations.
6. **resultsdata.dta** – this is a Stata dataset containing all the individual indicators, all the aggregate indicators, and all the estimated parameters of the unobserved components model used to generate the aggregates. This dataset is in “wide” format with countries in rows and separate columns for each variable and period. This file does not contain variable labels and is used primarily as an intermediate step in the calculations.
7. **individualindstableau.xlsx** and **aggindstableau.xlsx** are two Excel files containing all the individual indicators and all of the aggregate indicators. These contain no new information relative to the previous files but are used as inputs to the Tableau dashboards on the interactive portion of www.govindicators.org.

For questions, please email wgi@worldbank.org.