

The Worldwide Governance Indicators, 2024 Update: Readme File for Downloadable Datasets

www.govindicators.org

November 5, 2024

The Worldwide Governance Indicators (WGI) project constructs aggregate indicators of six broad dimensions of governance: *Voice and Accountability (va)*, *Political Stability and Absence of Violence/Terrorism (pv)*, *Government Effectiveness (ge)*, *Regulatory Quality (rq)*, *Rule of Law (rl)*, and *Control of Corruption (cc)*.

The six aggregate indicators are based on 35 underlying existing data sources reporting the perceptions of governance of many survey respondents and experts worldwide. Details on the underlying data sources, the aggregation method, and the interpretation of the indicators, can be found in the WGI methodology paper:

[*Daniel Kaufmann and Aart Kraay \(2024\). "The Worldwide Governance Indicators: Methodology and 2024 Update". World Bank Policy Research Working Paper No. 10952.*](#)

Full documentation and interactive access to the aggregate indicators and the underlying source data, is available at www.govindicators.org. Note that this Worldwide Governance Indicators update incorporates revisions to data for previous years, and so this data release supersedes data from all previous releases. For a summary of the revisions to the historical data, click [here](#).

This document serves as a “readme” file for all the downloadable datasets available on the WGI website. Please refer to Table of Contents below to find the documentation for each downloadable dataset.

Contents

1. wgidataset.dta, wgidataset.xlsx	2
2. wgidataset_with_sourcedata.dta, wgidataset_with_sourcedata.xlsx	3
3. wgiparameterestimates.dta, wgiparameterestimates.xlsx	4
4. wgisourceweights.dta, wgisourceweights.xlsx	5
5. wgicalculator.xlsx	6
Annex 1: WGI Data Source Codes	7
Annex 2: Demonstration Stata Code to Calculate Aggregate WGI Indicators	8

For questions about the 2024 WGI update please email wgi@worldbank.org.

The Worldwide Governance Indicators (WGI) are a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy, completeness, or currency of the data included in this work and does not assume responsibility for any errors, omissions, or discrepancies in the information, or liability with respect to the use of or failure to use the information, methods, processes, or conclusions set forth.

1. [wgidataset.dta](#), [wgidataset.xlsx](#)

These Stata and Excel files contain the six aggregate WGI indicators. The data are reported in “long” format with countries, years, and aggregate indicators stacked vertically. The database contains the following variables:

codeindyear: Unique identifier for each row in dataset (concatenation of code, indicator, and year)

code: ISO 3-letter country code

countryname: Name of country

year: Year

indicator: Identifier for WGI aggregate indicator (va/pv/ge/rq/rl/cc)

estimate: Estimate of governance in standard normal units ranging from approximately -2.5 (weak) to 2.5 (strong) governance performance.

stddev: standard deviation of governance estimate, reflecting uncertainty around the point estimate of governance. A 90 percent confidence interval is the estimate of governance +/- 1.64 times the standard deviation.

nsource: Number of data sources on which estimate of governance is based.

pctrank: Percentile rank among all countries, ranging from 0 (lowest) to 100 (highest) rank

pctranklower: Lower bound of 90% confidence interval for governance, in percentile rank terms

pctrankupper: Upper bound of 90% confidence interval for governance, in percentile rank terms

2. wgidataset_with_sourcedata.dta, wgidataset_with_sourcedata.xlsx

These Stata and Excel files contain the WGI aggregate indicators contained in wgidataset.dta and wgidataset.xlsx, as well as the data from all the underlying data sources, and some variables to facilitate reproducing the calculation of the aggregate indicators given the source data and parameter estimates. The data are in “long” format with countries, years, and the six aggregate indicators stacked vertically. The dataset includes the following variables:

codeindyear: Unique identifier for each row in dataset (concatenation of code, indicator and year)

code: ISO 3-letter country code

countryname: Name of country

year: Year

indicator: Identifier for WGI aggregate indicator (va/pv/ge/rq/rl/cc)

estimate: Estimate of governance in standard normal units ranging from approximately -2.5 (weak) to 2.5 (strong) governance performance.

stddev: standard deviation of governance estimate, reflecting uncertainty around the point estimate of governance. A 90 percent confidence interval is the estimate of governance +/- 1.64 times the standard deviation.

nsource: Number of data sources on which estimate of governance is based.

pctrank: Percentile rank among all countries, ranging from 0 (lowest) to 100 (highest) rank

pctranklower: Lower bound of 90% confidence interval for governance, in percentile rank terms

pctrankupper: Upper bound of 90% confidence interval for governance, in percentile rank terms

XXX: 35 variables named XXX=(adb/afr/asd/bps/bti/ccr/ebr/eiu/eqi/frh/gcb/gcs/gii/gwp/her/hum/hrm/ifd/ijt/ipd/irp/lbo/msi/obi/pia/prc/prs/rsf/tpv/vab/vdm/wbs/wcy/wjp/wmo) containing data from the 35 WGI component data sources, identified by three-letter source codes XXX (see Annex 1). Data are scaled to run from 0 (weak) to 1 (strong) and are identical to those reported in the source-specific Excel files available on the WGI website.

scalemean: mean across countries of estimate of governance *before* estimates are rescaled to have zero mean and unit standard deviation in each year. Same value repeats for all countries within a given indicator/year pair. *This variable is useful only for users who wish to reproduce the calculation of the aggregate indicators for a single country in Excel, which can be done using the file wgicalculator.xlsx available on the WGI website [here](#).*

scalesd: standard deviation across countries of estimate of governance *before* estimates are rescaled to have zero mean and unit standard deviation in each year. Same value repeats for all countries within a given indicator/year pair. *This variable is useful only for users who wish to reproduce the calculation of the aggregate indicators for a single country in Excel, which can be done using the file wgicalculator.xlsx available on the WGI website [here](#).*

3. [wgiparameterestimates.dta](#), [wgiparameterestimates.xlsx](#)

These Stata and Excel files contain the estimated parameters from the Unobserved Components Model used to construct the aggregate indicators. The data are in “long” format with years, indicators, and data sources stacked vertically. The dataset contains the following variables:

srcindyr: unique identifier for each row in the dataset (concatenation of source, indicator, and year)

year: Year

indicator: Identifier for WGI aggregate indicator (va/pv/ge/rq/rl/cc)

source: Identifier for WGI data source (adb/afr/asd/bps/bti/ccr/ebr/eiu/eqi/frh/gcb/gcs/gii/gwp/her/hum/hrm/ifd/ijt/ ipd/irp/lbo/msi/obi/pia/prc/prs/rsf/tpv/vab/vdm/wbs/wcy/wjp/wmo). See Annex 1 for definitions of data source codes.

alpha: intercept parameter in unobserved components model $y(j, k) = \alpha(k) + \beta(k)(g(j) + \varepsilon(j, k))$ where j indexes countries, k indexes data sources, $y(j, k)$ is the observed data on governance, and $g(j)$ is the true unobserved level of governance.

beta: slope parameter in unobserved components model $y(j, k) = \alpha(k) + \beta(k)(g(j) + \varepsilon(j, k))$ where j indexes countries, k indexes data sources, $y(j, k)$ is the observed data on governance, and $g(j)$ is the true unobserved level of governance.

sigma: standard deviation of error term $\varepsilon(j, k)$ in unobserved components model $y(j, k) = \alpha(k) + \beta(k)(g(j) + \varepsilon(j, k))$ where j indexes countries, k indexes data sources, $y(j, k)$ is the observed data on governance, and $g(j)$ is the true unobserved level of governance.

4. [wgisourceweights.dta](#), [wgisourceweights.xlsx](#)

These Stata and Excel files provide a summary of the weights applied to each data source. The data is organized in “long” format by aggregate WGI indicator and year. The dataset contains the following variables:

year: Year

indicator: Identifier for WGI aggregate indicator (va/pv/ge/rq/rl/cc)

For each of the 35 data sources identified by the three-letter code

XXX=(adb/afr/asd/bps/bti/ccr/ebr/eiu/eqi/frh/gcb/gcs/gii/gwp/her/hum/hrm/ifd/ijt/ipd/irp/lbo/msi/obi/pia/prc/prs/rsf/tpr/vab/vdm/wbs/wcy/wjp/wmo), the dataset contains:

XXX_weight: 35 variables with XXX=(adb/afr/asd/bps/bti/ccr/ebr/eiu/eqi/frh/gcb/gcs/gii/gwp/her/hum/hrm/ifd/ijt/ ipd/irp/lbo/msi/obi/pia/prc/prs/rsf/tpr/vab/vdm/wbs/wcy/wjp/wmo) containing the average across all countries in the indicated year and aggregate indicator of the weight assigned to source XXX. See Annex 1 for definitions of the source codes. The weight assigned to each data source k for a given indicator and year is

$w(k) = \frac{\sigma(k)^{-2}}{1 + \sum_k \sigma(k)^{-2}}$, where $\sigma(k)$ is the standard deviation of measurement error in data source k obtained from the Unobserved Components Model. These weights differ across countries because different sets of data sources are available for different countries. The variable XXX_weight reports the average across all countries of these weights for each data source XXX.

5. [wgicalculator.xlsx](#)

This file can be used to reproduce the WGI aggregate indicators using the provided data from the individual component data sources and the parameter estimates of the Unobserved Components Model. This is provided as a tool for users who would like to better understand the mechanics of the WGI calculations for a single country, taking as given the parameter estimates from the Unobserved Components Model.

This file contains three tabs:

data: This tab contains the WGI aggregate indicators and the component data, exactly as reported in `wgidata_with_sourcedata.xlsx`.

params: This tab contains the parameter estimates from the unobserved components model, exactly as reported in `wgiparameterestimates.xlsx`.

filters: This tab contains country names and country codes.

calculations: This tab reproduces the WGI calculations. To use this tab, select a country, WGI aggregate indicator (va/pv/ge/rq/rl/cc), and year using the drop-down menus in cells B2, B3 and B4. The spreadsheet retrieves the source data and corresponding parameter estimates and calculates the WGI aggregate indicator, standard error, number of data sources, percentile rank, and percentile rank of upper and lower bounds of 90 percent confidence interval in cells C9:C15. The same data as reported directly in the WGI website is reproduced in cells B9:B15 for comparison purposes.

The details of the calculations are shown in rows 19 to 25.

- Row 19 retrieves the available source data for $y(j, k)$ for selected country j and source k .
- Rows 20-22 retrieve the UCM parameter estimates $\alpha(k)$, $\beta(k)$, $\sigma(k)$ for each source k .
- Row 23 calculates the rescaled data for each source, $(y(j, k) - \alpha(k))/\beta(k)$.
- Row 23 calculates the precision of each source, $\sigma(k)^{-2}$.
- Row 24 calculates the weight assigned to each data source, $w(k) = \sigma(k)^{-2}/(1 + \sum_k \sigma(k)^{-2})$.

Finally, the estimate of governance is calculated as the weighted average of the rescaled data from each source,

$$\sum_k \left(\frac{y(j, k) - \alpha(k)}{\beta(k)} \right) w(k)$$

This number is then rescaled by the mean and the standard deviation of the estimates of governance for all other countries to ensure that the global distribution of governance estimates has zero mean and unit standard deviation. This number is reported in cell C9 and matches the corresponding number as reported in the WGI website (in cell B9). Similar calculations are done for the standard deviation, number of sources, and percentile ranks.

Please note that this entire file is password protected to prevent accidental edits. Users can only change the selection of country, aggregate WGI indicator, and year.

Annex 1: WGI Data Source Codes

ADB	African Development Bank Country Policy and Institutional Assessments
AFR	Afrobarometer
ASD	Asian Development Bank Country Performance Assessments
BPS	Business Enterprise Environment Survey
BTI	Bertelsmann Transformation Index
CCR	Freedom House Countries at the Crossroads
EBR	European Bank for Reconstruction and Development Transition Report
EIU	Economist Intelligence Unit Riskwire & Democracy Index
EQI	European Quality of Governance Survey
FRH	Freedom House
GCB	Transparency International Global Corruption Barometer Survey
GCS	World Economic Forum Global Competitiveness Report Survey
GII	Africa Integrity Indicators
GWP	Gallup World Poll
HER	Heritage Foundation Index of Economic Freedom
HRM	Human Rights Measurement Initiative
HUM	Cingranelli Richards Human Rights Database and Political Terror Scale
IFD	IFAD Rural Sector Performance Assessments
IJT	Crisis24 Country Security Assessment Ratings
IPD	Institutional Profiles Database
IRP	African Electoral Index
LBO	Latinobarometro
MSI	International Research and Exchanges Board Vibrant Information Barometer
OBI	International Budget Project Open Budget Index
PIA	World Bank Country Policy and Institutional Assessments
PRC	Political Economic Risk Consultancy Corruption in Asia Survey
PRS	Political Risk Services International Country Risk Guide
RSF	Reporters Without Borders Press Freedom Index
TPR	US State Department Trafficking in Persons Report
VAB	Vanderbilt University Americas Barometer
VDM	Varieties of Democracy Project
WBS	World Bank Enterprise Surveys
WCY	Institute for Management and Development World Competitiveness Yearbook
WJP	World Justice Project Rule of Law Index
WMO	S&P Global Country Risk Service

Annex 2: Demonstration Stata Code to Calculate Aggregate WGI Indicators

This annex provides Stata code that can be used to calculate the aggregate WGI indicators given the WGI source data and parameter estimates reported on the WGI website. This is provided as a tool for users who would like to better understand the basic WGI calculations and/or simulate different WGI scores with different weights or data inputs.

```

/*****
// Demonstration code to calculate WGI aggregate indicators using provided
// data on sources and parameter estimates for the 2024 WGI update.
// Copy files wgiparameterestimates.dta and wgidataset_with_sourcedata.dta
// into desired folder and set filepath below to point to that folder.
// Then run this code.
*****/
clear
set more off

global filepath "INSERT YOUR DESIRED FILEPATH HERE"
global sourcelist "adb afr asd bps bti ccr ebr eiu eqi frh gcb gcs gii gwp her"
global sourcelist "$sourcelist hum hrm ifd ijt ipd irp lbo msi obi pia prc prs"
global sourcelist "$sourcelist rsf tpr vab vdm wbs wcy wjp wmo"
cd "$filepath"

/*****
// Load and merge data and parameters
*****/
// Load parameter estimates and reshape into long format
use "wgiparameterestimates.dta"
gen indyr=substr(srcindyr,4,6)
drop srcindyr
reshape wide alpha beta sigma, i(indyr) j(source) string
drop indyr
order indicator year
save temp.dta, replace
// Load WGI dataset with source data, merge in parameter estimates
clear
use "wgidataset_with_sourcedata.dta"
merge m:1 indicator year using "temp.dta"
drop _merge
// Set parameter estimates to missing if corresponding source data are missing
foreach src in $sourcelist{
    foreach param in alpha beta sigma{
        replace `param'`src'=. if `src'==.
    }
}

/*****
// Calculate aggregate indicators
*****/
// Generate rescaled indicators  $(y(j,k) - \alpha(k)) / \beta(k)$ 
foreach src in $sourcelist{
    qui gen `src'resc=(`src'-alpha`src')/beta`src'
}
// Generate precision  $(\sigma(j,k)^{-2})$ 
foreach src in $sourcelist{
    qui gen `src'prec=sigma`src'^-2
}
// Generate weight for source  $(\sigma(j,k)^{-2} / (1 + \sum_k (\sigma(j,k)^{-2})))$ 
egen sumprec=rowtotal(*prec), missing
foreach src in $sourcelist{
    qui gen `src'weight=`src'prec/(1+sumprec)
}
// Product of rescaled sources times weights
foreach src in $sourcelist{
    qui gen `src'rescweight=`src'resc*`src'weight
}
// Calculate estimate, standard deviation, and number of sources
```



```

egen estimate_new=rowtotal(*rescweight), missing
gen stddev_new=(1+sumprec)^-0.5
egen nsource_new=rownonmiss($sourcelist)
// Rescale estimate and standard deviation so that estimates have mean zero and
// standard deviation of 1 across countries
gen indyr=substr(codeindyr,4,6)
egen mean=mean(estimate_new), by(indyr)
egen sd=sd(estimate_new), by(indyr)
replace estimate_new=(estimate_new-mean)/sd
replace stddev_new=stddev_new/sd

/*****/
// Confirm _new aggregate indicators are identical to those in dataset
/*****/
gen estimate_test=estimate-estimate_new
gen stddev_test=stddev-stddev_new
gen nsource_test=nsource-nsource_new
su *_test

```