**README**

**Cross-Country Empirical Analysis of GovTech Platforms on Citizen Engagement**

This folder contains the reproducibility package for the Working Paper titled “Cross-Country Empirical Analysis of GovTech Platforms on Citizen Engagement” by Maimouna Diakite and Abdoul-Akim Wandaogo. If you encounter any issues running the codes or reproducing the results, please get in touch with the authors at mdiakite2@worldbank.org and awandaogo@worldbank.org.

## I. Computational Requirements

### 1. Software Requirements

- Stata (code last executed with version 18)
- Packages:
  - ⇒ ebalance
  - ⇒ Psmatch
  - ⇒ psmatch2
  - ⇒ covbal
  - ⇒ pbalchk
  - ⇒ outreg2
  - ⇒ estout
  - ⇒ grc1leg
  - ⇒ rbounds

  - All of these packages are included in the reproducibility package.
  - The "MASTER.do" program will load the packages from the folder “ado” in line 66.

### 2. Memory and Runtime Requirements

- Paper exhibits were generated on a MacBook Pro M1 laptop with 16GB of RAM.
- The Stata analysis code takes approximately 3 hours minutes to run. This duration may vary depending on the operating system, RAM, and Stata version.

- The post-estimation test of PSM of line 289 can be skipped if you wish to save time. This has been provided for transparency purposes to reproduce all results in Table 6.

## I. Instructions for Replicators

### 1. Dataset construction

The dataset was constructed using Excel and Stata. The data extracted for the study are available in the folder Dataset_construction, and the raw data (original versions) are available in the subfolder Raw_data located in the folder Dataset_construction. Details on data availability and downloading are available in the Data Availability Statement section. The data preparation was done by following the steps below:

**a. Extraction and manipulation of the GovTech Maturity Index (GTMI) 2022 data**: the study used GTMI 2022 data to construct the digital platforms' implementation variables. The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the Excel file named WBG_GovTech Dataset_Mar2023 (sheet: CG_GTMI_Data). The data were extracted by year (the variable *Year* was sorted "Sort Smallest to Largest") in Excel, and the suffix "_22" was added for 2022 data). For each year, the values of the following cells were extracted and used: *Economy* (renamed ***country*** in Excel), *Code* (renamed ***code*** in Excel), *Year* (renamed ***year*** in Excel), and the variables *I-30* (renamed ***I30/I30_22*** in excel) and *I-31* (renamed ***I31/I31_22***). After that, these data were saved in the Excel file named gtmi (available in the folder Dataset_construction). Then, the Excel file gtmi was imported into Stata (all the Stata codes and formulas executed to construct the dataset are available in the Stata dofile named Dataset_construction.do located in the folder Dataset_construction). The variable ***GovTech*** was constructed in Stata through 2020 ***I30*** and ***I31*** data; it takes the value 1 if ***I30*** and/or ***I31*** is/are equal to 1 and 0; otherwise; ***GovTech_22*** was constructed through 2022 ***I30_22*** and ***I31_22*** data; it takes the value 1 if ***I30_22*** and/or ***I31_22*** is/are equal to 1 and 0 otherwise. The GTMI 2022 dataset was used as the reference dataset in the study, and the GTMI codes/country names were employed as a reference. Thus, the remaining datasets were merged through the codes/country names to gtmi.dta using Stata. Therefore, for datasets for which the country codes/ country names (for the same country) were different from those employed in GTMI 2022, the study replaced the codes/names provided in the raw databases (available in the subfolder Raw_data) with the GTMI codes/names in the dta files located in the Dataset_construction folder. For example, the code for Kosovo in GTMI 2022 is KSV; however, the code for Kosovo in the World Development Indicators (WDI) dataset is XKX. Thereby, for WDI and

the remaining datasets, the study ensured that the codes were harmonized before merging the different datasets using Stata.

b. **Extraction and manipulation of World Development Indicators (WDI) data**: The study uses 2020 WDI data for several variables. The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the Excel files named P_Data_Extract_From_World_Development_Indicators (with suffixes _1 to _2; sheets: Data). These variables were extracted in two Excel files (wdi_1 and wdi_2 located in the folder Dataset_construction, where *Country Name* was replaced with ***country***, *Country Code* was replaced with ***code***, and labels were created for the *Series Names*). The list of sub-Saharan African countries (including the variables **country** and **code**) inserted in the Excel file wdi_3 and the list of developed countries/High-income countries inserted in the Excel file wdi_4 were also downloaded from WDI. These WDI data were then imported and merged in Stata (dofile: Dataset_construction.do). After that, the WDI data were merged with those of gtmi.dta, and the merged data were manipulated and saved as dataset1.dta.

c. **Extraction and manipulation of Varieties of Democracy (V-Dem) data**: The study uses 2020 V-Dem data for the variables *Public sector corruption* (renamed ***v2x_pubcorr_20***) and *Equal distribution of political power* (renamed ***v2pepwrsoc_ord_20***). The 2021 and 2022 data for the *variables Civil Society Participation Index* (renamed ***e_v2x_cspart_5C_21/22***) and *Engaged Society* (renamed ***v2dlengage_ord_21/22***) were also extracted. The raw data are available in the subfolder Raw_data (in the file V-Dem-CY-Full+Others-v13.dta) and the folder Dataset_construction (in the file vdem.dta). The 2020, 2021, and 2022 data were extracted and manipulated using Stata in three different files, which were merged into a common file named vdemdcep.dta. Then, the file vdemdcep.dta was merged with dataset1.dta to create dataset2.dta.

d. **Extraction and manipulation of Polity V data**: The study uses 2018 Polity V data for the variable ***polity2*** "*Democracy*". The raw data are available in the subfolder Raw_data (in the file p5v2018.xls) and the folder Dataset_construction (in the file polity.xls). The 2018 data were extracted and manipulated using Stata. Then, these data were merged with dataset2.dta to create dataset3.dta.

e. **Extraction and manipulation of the International Telecommunication Union (ITU) data**: The study uses 2020 ITU data for the variable *Fixed broadband subscriptions per 100 inhabitants* (***broadband_inhab_subsc***). The raw data are available in the subfolder Raw_data (in the file FixedBroadbandSubscriptions_2000-2021.xlsx) and the folder Dataset_construction (in the file

broadband.xlsx). The 2020 data were extracted manually in Excel and then manipulated using Stata. Then, these data were merged with dataset3.dta to create dataset4.dta.

f. **Extraction and manipulation of UCDP/PRIO Armed Conflict data**: The study uses 2020 PRIO data for the conflict variables ***intensity_level*** and ***cumulative_intensity***. The 2020 data were first extracted manually and manipulated using Excel and then imported and treated in Stata. The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the Excel file named ucdp-prio-acd-221. The data preparation was done in Excel by following the steps below:

- Step 1: the raw data (namely the variables *location*, ***year***, ***intensity_level****, and cumulative_intensity) were first copied and pasted into a new Excel file named prio_treatment,* available in the folder Dataset_construction.
- Step 2: the Excel command Sort was used to sort the year (with the option to expand the selection) and keep the 2020 data by deleting the remaining observations.
- Step 3: For rows for which the variable location includes more than a country (when two countries are at war), an additional row was created for the second country (with the same information for the conflict variables as for the first country), and the variable *location* was renamed ***country***.
- Step 4: the variable **country** was sorted, and for countries with more than one conflict/row, only the more intensive conflict/row was kept by removing the remaining conflicts. This process allowed to keep one observation per country, and the command COUNTIF was used to ensure there was only one observation per country.
- Step 5: the observations for the dummy variable ***cumulative_intensity*** were changed. The original data include two values: 0 (low cumulative intensity) or 1 (high intensity). However, there are values for only countries at war. Thus, to cover the remaining sample countries, the study replaced 1 by 2 for high intensity and 0 by 1 for low intensity. This allowed the study to assign (in Stata) the value 0 for both the *intensity and cumulative intensity* variables to countries that are not at war. The values of the ***intensity_level*** variable, which were already 1 or 2, were not changed. Finally, Step 5 data were copied and pasted into a new Excel file named prio.xlsx (available in the Dataset_construction folder), and the remaining manipulation needed for merging the PRIO data with dataset4.dta was conducted using Stata. Then, the merged data was saved as dataset5.dta.

g. **Extraction and manipulation of the World Bank FY20 List of Fragile and Conflict-affected Situations**: The study uses the World Bank 2020 List of Fragile and Conflict-affected Situations data

for the variables *Conflict-affected situations* (***conflictwbg***) and *Fragile and Conflict-affected situations* (***conflictnfragwbg***). The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the pdf file named FCSFY20. These data were extracted and treated in the Excel file wbglist2020 located in the folder Dataset_construction. The data preparation was done in Excel by following the steps below:

- Step 1: The first step of the data treatment consisted of reporting only the raw data.
- Step 2: the variables were created in this step. For ***conflictwbg***, a value of 2 was assigned to countries with high-intensity conflict; and 1 was assigned to countries with medium-intensity conflict. For ***conflictnfragwbg***, a value of 3 was assigned to countries with a high-intensity conflict; 2 was assigned to countries with a medium-intensity conflict; and 1 was assigned to countries with an absence of conflict but high institutional and social fragility.

  Then, Step 2 data were copied and put in a new Excel file named wbgl.xlsx (available in the Dataset_construction folder). The remaining manipulation needed for merging the data with dataset5.dta was conducted using Stata, and the merged data were saved as dataset6.dta.

**h. Extraction and manipulation of Worldwide Governance Indicators (WGI) data**: The study uses 2020 WGI data for the variable Government effectiveness (**goveffect_wgi**). The raw data are in the subfolder Raw_data in the Excel file named P_Data_Extract_From_Worldwide_Governance_Indicators (sheet: Data). The data were extracted in a new Excel file named wgi (where *Country Name* was replaced with ***country***, *Country Code* was replaced with ***code***, and the label ***goveffect_wgi*** was created for *Government Effectiveness: Estimate*) located in the folder Dataset_construction. Finally, the data were imported into Stata and then manipulated and merged with those of dataset6.dta to create **Lastdataset_DCE_WP.dta**.

**2. Econometric Analysis**

The codes provided reproduce the results included in the working paper. To recreate the outputs, follow these steps:

1. Open the downloaded folder, navigate to the "MASTER.do" file, and go to line 3.
2. Edit the global path in line 3 to the corresponding folder in your computer.

The "MASTER.do" file creates all the tables and graphs included in the working paper. The outputs will be recreated in the first directory folder. Please note that the only output that cannot be reproduced is Annex

A, for which you may need to collect information externally. For Figure 1, the version in the paper was created in Excel, but the data was processed in Stata; the code for a Stata version is also provided.
The PSM estimation is performed using the Bootstrap method, which provides robust results. Additionally, there are three statistics verifying the quality of the Matching and these statistics are not automatically exported with the results table. They must, therefore, be included manually in the table. Another detailed do file (entitled PSM post_estimation test) is included in the package with permutation codes to estimate these statistics. These statistics are estimated and extracted as follows:

i. Estimate the treatment effects on the treated from the MASTER do file (Lines 247-285, after importing the dataset and creating the logarithm of the GDP per capita (Lines 72-76)).
ii. Then, using the codes from the "PSM post_estimation test" do file, evaluate the estimation quality by executing the codes block by block (The blocks are numbered from 1 to 6 for each of the two explained variables).
iii. The statistics will be registered in the log file generated when running the MASTER do file.
iv. In the penultimate result table displayed after executing a block, the value of "Ps R2" represents the pseudo-R2, while the value of "p>chi2" represents the Standardized bias. These two values can sometimes change, but the most important thing is that the Pseudo-R2 should be lower than, at least, the 10 percent level, while the standardized bias should be higher than all the conventional levels.
v. The Rosenbaum upper bound sensitivity test (gamma) will be in the last table of the estimation. This value (gamma) is determined by comparing the values of "sig+" to the defined threshold value (we have defined 5% in our study). The R-Baum value will be the first gamma value for which the sig+ value is higher than the defined threshold value (5% in our case). It is desirable for this gamma value to be at least greater than 1.

Note 1: You can increase the higher bound and the interval until you find the gamma for which sig+ is higher than the defined level (which is 5% in our case).

Note 2: Except for the estimation coefficients, which remain intact, changes in values, especially Pseudo R2 and Standardized bias, may sometimes be observed. However, they must adhere to the aforementioned conditions (PseudoR2 lower than the threshold, Rosenbaum upper bound ideally greater than 1, and Standardized bias higher than the 10% threshold).

## II. Data Availability Statement

This section lists the sources of raw data (used to extract the data needed for the model simulations). All the data employed by the study are publicly available. Below is a description of where one can download the data.

1. **GovTech Maturity Index (GTMI) 2022 dataset** (downloaded on August 24, 2023)

The study uses 2020 and 2022 data extracted from the 2022 version of the GovTech Maturity Index (GTMI) dataset published on March 22, 2023, which can be downloaded from https://datacatalog.worldbank.org/int/search/dataset/0037889/govtech-dataset. The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the Excel file named WBG_GovTech Dataset_Mar2023 (sheet: CG_GTMI_Data).

2. **World Development Indicators (WDI) dataset** (downloaded on July 26, 2023)

The study uses 2020 WDI data, which can be downloaded from https://databank.worldbank.org/source/world-development-indicators. The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the Excel file named P_Data_Extract_From_World_Development_Indicators (sheet: Data).

3. **Varieties of Democracy (V-Dem) v13 dataset** (downloaded on February 22, 2023)

The study uses 2020, 2021 & 2022 V-Dem data downloaded from https://v-dem.net/data/the-v-dem-dataset/. The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the dta file named V-Dem-CY-Full+Others-v13.

4. **Polity V dataset** (downloaded on July 26, 2023)

The study uses 2018 Polity V data (the most recent found) downloaded from https://www.systemicpeace.org/inscrdata.html. The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the Excel file named p5v2018. Data files for this dataset are not included in the reproducibility package, and must be manually downloaded from the data URL and converted into “polity.xls” following the instructions of page 3 for the code to run.

5. **International Telecommunication Union (ITU) dataset** (downloaded on August 27, 2023)

The study uses 2020 ITU data downloaded from https://datahub.itu.int/. Original data was downloaded by selecting the indicator “Fixed-broadband subscriptions” and unit “per 100 people”. The original data was manually reshaped to a wide format by country with year columns for the indicator. The resulting data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the Excel/CSV file FixedBroadbandSubscriptions_2000-2021 (sheet i992b).

6. **UCDP/PRIO Armed Conflict dataset** (downloaded on July 27, 2023)

The study uses 2020 UCDP/PRIO Armed Conflict (version 22.1) data downloadable from https://ucdp.uu.se/downloads/index.html#armedconflict. The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the Excel file named ucdp-prio-acd-221.

7. **The World Bank FY20 List of Fragile and Conflict-affected Situations** (downloaded on July 27, 2023)

The study uses the World Bank 2020 List of Fragile and Conflict-affected Situations data, which can be downloaded from https://pubdocs.worldbank.org/en/179011582771134576/FCS-FY20.pdf. The data was manually constructed from the PDF file available in the data URL and saved in the file "wbgl.xlsx".

8. **Worldwide Governance Indicators (WGI) dataset** (downloaded on August 8, 2023)

The study uses 2020 WGI data, which can be downloaded from https://databank.worldbank.org/source/worldwide-governance-indicators. The raw data are available in the subfolder Raw_data (located in the folder Dataset_construction) in the Excel file named P_Data_Extract_From_Worldwide_Governance_Indicators (sheet: Data).

**Thank you.**