

README and Guidance
for
“Revisiting Public Capital Needs: An Analysis of Growth-Maximizing Investment with
Efficiency and Congestion Effects”
Santiago Herrera, Christophe Hurlin, and Hironobu Isaka
November 7, 2025

Overview

This replication package contains the code and data used for the analysis in the paper *“Revisiting Public Capital Needs: An Analysis of Growth-Maximizing Investment with Efficiency and Congestion Effects.”* The package provides R scripts to reproduce all main tables and figures.

Datasets and Availability

The data we use are described in the appendix table. The datasets provided there are required to run the replication code. Please ensure they are stored locally before execution, and specify the correct folder path in the scripts.

Description of Programs and Instructions for Replicators

This section describes the steps required to replicate the results, from data preparation through estimation and post-estimation (organized by the corresponding sections of the paper).

Before running the code, please ensure the following:

- All datasets listed in Table A1 are available in the file paths specified there.
- File paths in the scripts are correctly updated (these are typically specified in the first few lines of each script).

1. Data Collection

The final output from this step is

“IMFInvestmentandCapitalStockDataset2021_extended_step4_1.xlsx.” To replicate the exact results presented in the paper, please use the dataset available in the Data Development Hub (forthcoming). Because the WDI API (used in the code) always provides the most recent data—which may differ from what was available when we originally queried it—it might not be possible to reproduce the dataset exactly. Nevertheless, we include the underlying code for the sake of transparency.

This step prepares the datasets to be used in the estimation. Please run the following scripts in the “01. Data Collection” folder in the order listed below. Running them sequentially ensures proper preparation of all datasets for estimation.

File name	Description
Step 0 - PotentialGrowth&Delta.ipynb ¹	Prepare datasets used in this section.
Step 1 – Data Preparation.R	Cleans and organizes the raw data for subsequent analysis.
Step 2.1 - OECD Extension of Public Capital Series.R	Extends the OECD public capital series data for replication.
Step 2.2 - Non-OECD Merge of sources.R	Merges non-OECD data sources into a unified dataset.
Step 2.3 - Non-OECD Extension of public capital series.R	Extends the public capital series for non-OECD countries.
Step 3.1 - OECD Extension of Private Capital Series.R	Extends the OECD private capital series data.
Step 3.2 - Extension of other series - GDP.R	Prepares and extends GDP series for consistency with capital data.
Step 3.3 - Non-OECD Extension of Private Capital Series.R	Extends the private capital series for non-OECD countries.
Step 4 - OECD Extension of other series - Employment - Human Capital.R	Extends OECD datasets on employment and human capital.

2. Estimation²

Now move on to the “02. Estimation” folder and execute the following R scripts (or main.R) in the order listed below. These scripts generate the results presented in the paper, with each script corresponding to specific tables and figures.

File name	Description
01. Descriptive Statistics.R	Table 2 and 3: Descriptive Statistic.
02. Model_All_countries.R	Table 4: Estimation Results for All Countries
03. Model By Regions.R	Table 5: Panel Estimation Results by Country Group and Specification

¹ This code retrieves data from APIs such as WDI and WEO, which always return the most recent version of the datasets. As a result, if the data is queried at different points in time, the returned values may differ. For replication purposes, we will add the output dataset “Depreciation and Potential Growth per Worker.xlsx” to the Development Data Hub (DDH).

² The scripts generate all figures and tables in the main text, except for Figure 4. Figure 4 is available in the Excel files—see the “Figures” tab in Excess_Inv.xlsx (located in the “03. Post-estimation” folder).

File name	Description
04. Mixed_Effect_Model By Regions.R	Table 6: Mixed-effect model results by income group and specification
05. Optimal_Investment_All_countries.R	Table 7: Observed and Optimal Ratios of Public Capital and Investment in All Countries, 2014-2024
06. Optimal_Investment_by_Region_Spec3.R	Table 8: Observed and Optimal Ratios of Public Capital and Investment by Income Group under the Full Congestion Specification (OCRS) with Observed GDP per Worker Growth, 2014-2024
07. Optimal_Investment_by_Region_Spec3_Potential_Growth.R	Table 9: Observed and Optimal Ratios of Public Capital and Investment by Income Group under the Full-Congestion Specification (OCRS) with Potential GDP Growth, 2014–2024
08. Optimal_Investment_by_Region_Spec4.R	Table 9A: Observed and Optimal Ratios of Public Capital and Investment by Income Group under the Endogenous Growth Specification, 2014-2024
09. Distribution of the ratio of actual-to-optimal public investment over the period 1990–2024, by	Figure 1: Distribution of the Ratio of Actual-to-Optimal Public Investment by Country Group, 1990–2024
10. Excess public investment per countries	- Figure 2: Gaps between Actual and Optimal Public Investment Ratios by Country, 2014-2024 - Figure 3: Gaps between Actual and Optimal Public Investment Ratios by Country Group, 2014-2024, Assuming Full Public Investment Efficiency (left column) and Accounting for Inefficiency
11. Optimal_Investment_per_Country_Endo.R	Excess investment estimates under the endogenous growth model.

3. Post-estimation

Finally, move on to the “03. Post-estimation” folder. After estimating excess investment, we conduct a post-estimation analysis using the estimated level of optimal investment. In particular, we correlate the results with other variables, as reported in Table 10. To reproduce Table 10 (Mean_Test.xlsx), please run the following files in the order listed below:

File name	Description
00. Synthesis_PEFA2011.ipynb ³	Construct a variable used in the mean testing.

³ This code generates the file “Imputed_results_PEFA2016.xlsx”, but the output may differ from the file with the same name stored in “03. Post-estimation\Imputed_results_PEFA2016.xlsx”. This is because the code includes random elements in the parameter selection process, and the chosen parameters may vary across

File name	Description
01. Posttest_Ig_cross.ipynb	Data Preparation for Table 10.
02. Excess_Inv.do	Backend data for Table 10: Comparison between excess-investing and under-investing countries (means test, 2014-2024).
Mean_Test.xlsx ⁴	Table 10.

runs. Therefore, for replication purposes, please use the dataset already stored in that folder as the fixed reference. The code is provided solely for transparency regarding how the data were constructed.

⁴ In Excel Query, if you want to revise the estimation and retrieve the results, make sure the file path is correct, and then refresh the query.

Appendix 1.

Table A1. List of Datasets

Dataset	File name	Description	Publicly available?	Version	URLs
Human Development Index	01. Data Collection /Human_Capital.xlsx	Composite index published by UNDP capturing key dimensions of human development—health, education, and standard of living.	Yes	2023–24 Human Development Report	https://hdr.undp.org/data-center
Investment and Capital Stock Dataset (ICSD), IMF	01. Data Collection /IMFInvestmentandCapitalStockDataset2021	IMF database providing data on investment flows and capital stock across countries; extended version includes author-constructed series.	Yes.	ICSD 2021	https://infrastructuregovern.imf.org/content/dam/PIMA/Knowledge-Hub/dataset/IMFInvestmentandCapitalStockDataset2021.xlsx
World Economic Outlook Database, October 2024, IMF	01. Data Collection /WEOOct2024all.xlsx	IMF's flagship macroeconomic database providing forecasts and historical data on GDP, fiscal, external, and other key indicators.	Yes	WEO, October 2024	https://www.imf.org/en/Publications/WEO/weo-database/2024/October
Infrastructure Efficiency Index	01. Data Collection /Infrastructure-Efficiency-Index-IEI-forsharing.xlsx	Composite indicator measuring how effectively countries convert public capital stock spending into infrastructure outcomes.	Yes	–	https://thedocs.worldbank.org/en/doc/a2071d190cf4c3cf96636aed353a61c0-0050062021/original/Infrastructure-

					Efficiency-Index-IEI-forsharing.xlsx
OECD Economic Outlook	• 01. Data Collection /OECD_ECO_GDP.xlsx • 01. Data Collection /OECD_ECO_GDP_PPP.xlsx • 01. Data Collection / • 01. Data Collection /OECD_ECO_IGV.xlsx • 01. Data Collection /OECD_ECO_ITV.xlsx • 01. Data Collection /OECD_ECO_PIT.xlsx	OECD's semiannual dataset providing projections and historical data on GDP, purchasing power parity, investment, government accounts, taxation, and other macro-fiscal variables.	Yes	OECD Economic Outlook (Retrieved on April 30, 2024)	https://www.oecd.org/en/publications/serials/oecd-economic-outlook_g1ghgh13.html
World Bank Fiscal Datasets	01. Data Collection /World Bank Fiscal Datasets.xlsx	Collection of several fiscal policy datasets, primarily from internal surveys of country economists, covering government revenues, expenditures, and budget structures.	No, only internally available.	Retrieved on February 2025	https://worldbankgroup.sharepoint.com/sites/MFMfpag/SitePages/_World-Bank-

					Fiscal-Datasets.aspx ⁵
Penn World Table version 10.01	01. Data Collection/pwt1001.xlsx	An internationally comparable database providing national accounts data, including GDP, capital stock, and productivity measures, standardized across countries and over time.	Yes	PWT 10.01	https://www.rug.nl/ggdc/productivity/pwt/?lang=en
Macro Poverty Outlook Data	01. Data Collection\series_2_28_2024.csv	World Bank dataset containing medium-term macroeconomic and poverty projections for low- and middle-income countries.	No, only internally available.	–	https://mtimodelling.worldbank.org/livempodata/mpodata.html ⁶
A Cross-Country Database of Potential Growth	01. Data Collection\Pot_HP_23Q4.csv 01. Data Collection\Pot_MVF_23Q4.csv 01. Data Collection\Pot_UCM_23Q4.csv	Dataset constructed by World Bank staff providing potential growth estimates using different methodologies (HP filter, MVF, UCM) across countries.	Yes, but shared directly by the author (slightly different format). ⁷ Will be archived to Development Data Hub (DDH).	23Q4 vintage	https://documents1.worldbank.org/curated/en/099426303102336406/pdf/IDU0f9f2613e0a18e0486c0a2710e80b2c21a128.pdf
Depreciation rate and Potential	01. Data Collection\Depreciation and Potential Growth per Worker	Retrieves depreciation data from PWT and calculates potential growth per worker based on WDI, WEO, and the Database of Potential Growth.	Will be archived to DDH.	-	-

⁵ See Appendix 2 as to how to download the data.

⁶ See Appendix 3 as to how to download the data.

⁷ To request access to the raw data used in this study for future updates, please contact one of the authors of this paper: <https://openknowledge.worldbank.org/entities/publication/c8106c9c-98a2-4ef4-8a28-403bfd1c9852>

Growth per Worker					
WEO	03. Post-estimation \\WEO_Data.xlsx	Extract of IMF WEO used specifically for GDP per capita series.	Yes	WEO October 2023	https://www.imf.org/en/Publications/WEO/Issues/2023/10/10/world-economic-outlook-october-2023
ICRG Dataset	/03. Post-estimation/qog_std_ts_jan23_stata14.dta	Quality of Government (QoG) Standard Time-Series dataset, containing ~400 variables from 80 sources covering institutions, governance, and policy.	Yes	QoGStdT Sjan23 (Jan 2023)	https://www.gu.se/en/quality-government/qog-data/data-downloads/basic-dataset
Resource Rich	03. Post-estimation\\Resource_rich_countries.xlsx (Tab: Sheet1)	List of 32 natural resource-dependent countries as defined in a World Bank study on fiscal procyclicality.	Yes	From paper, Appendix E	https://documents1.worldbank.org/curated/en/510041565102137832/pdf/Why-Some-Countries-Can-Escape-the-Fiscal-Pro-Cyclicality-Trap-and-Others-Cannot.pdf
IMF resource rich	03. Post-estimation\\Resource_rich_countries.xlsx (Tab: IMF (2012))	List of resource-rich countries as classified in IMF (2012) policy paper.	Yes	From paper, Appendix 1	https://www.imf.org/external/np/pp/eng/2012/082412.pdf

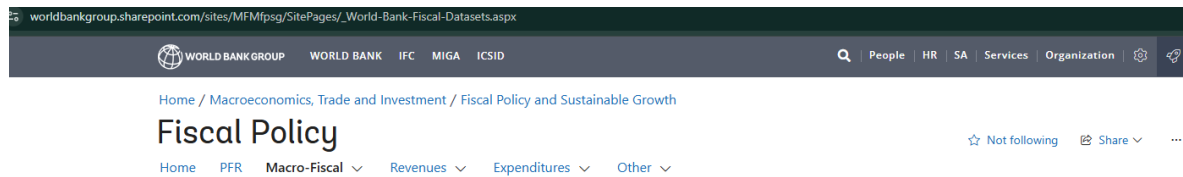
PEFA ⁸	03. Post-estimation\Imputed_results_PEFA2016.xlsx	Public Expenditure and Financial Accountability (PEFA) assessments, extended using author imputations to align 2011 and 2016 frameworks.	Yes for initial dataset, but No for the extended version. Will be archived to DDH.	Retrieved on Feb 2024	https://www.pefa.org/assessments/batch-downloads
Quality of Expenditure Heatmaps	03. Post-estimation\Quality of Expenditure Heatmaps.xlsx	Heatmap dataset summarizing public spending quality along three pillars (growth, stabilization, redistribution) and subcomponents, using multiple indicators.	No, only internally available. Will be archived to DDH.	Version 1.2	https://worldbankgroup.sharepoint.com/:r/sites/KnowledgeProduct/Shared%20Documents/PFR/Tools/Expenditure%20heatmap/Quality%20of%20Expenditure%20Heatmaps.xlsx?download=1
IMF Income Classification	03. Post-estimation\FP_Classification_Table.dta	IMF classification of economies into Advanced, Emerging Market, and Developing categories.	Yes	–	https://www.imf.org/en/Publications/WEO/weo-database/2025/April/groups-and-aggregates
Estimated Excess Investment	03. Post-estimation\ Excess_Investment.xlsx	An estimation result.	No. Only Internally Available.	-	-

⁸ The extended PEFA dataset was constructed by imputing missing scores for the PEFA 2016 framework using machine learning techniques trained on the PEFA 2011 framework. Since this imputed version is generated internally for research purposes, it is not publicly available.

			Will be archived to DDH.		
PEFA Assessments	03. Post-estimation /00. Synthesis PEFA2011/PEFA_data.xlsx	Public Expenditure and Financial Accountability (PEFA) assessments, extended using author imputations to align 2011 and 2016 frameworks.	Open data	Retrieved on Feb 2024	https://www.pefa.org/assessments/batch-downloads

Appendix 2. World Bank Fiscal Datasets

- Access this link: https://worldbankgroup.sharepoint.com/sites/MFMfpstg/SitePages/_World-Bank-Fiscal-Datasets.aspx
- Click either of the datasets highlighted in red (in .dta or .csv format).
- Save the data in an Excel file named “World Bank Fiscal Datasets.xlsx”, and name the worksheet containing the data “Data” to ensure the code runs without any path errors.



The **World Bank Fiscal Datasets** (WBFD) combines the fiscal data semi-annually submitted by country economists with other data sources to improve and supplement it. This results in a rich set of databases that could be used not only to evaluate the information provided by the country economists but also to evaluate other data sources, provide information to the fiscal enhanced MFMOD, inform analytical tools such as the PFR Data Visualization Tool, and be the main data source for advances analysis carried out by EMFTX or other World Bank users.

✓ Presentation on databases, variable codes, and tool

^ Databases

The databases are available for download via the links provided in the next two columns. Alternatively, they can be accessed from a shared folder that is available to all World Bank employees. To access the database from your computer, use the following address in your file explorer or preferred statistical software.

\\fiscalpolicy\data\World_Bank_Fiscal_Datasets

dta (Stata) files, variables in columns

- Core
- Harmonized
- Supplemented
- Retropolated**

CSV, variables in columns

- Core
- Harmonized
- Supplemented
- Retropolated**

Appendix 3. MPO data

- Access the link: <https://mtimodelling.worldbank.org/livempodata/mpodata.html>
- Select all the countries and all series as shown below.
- Click “Download” button

Save

Country Available 157 | Selected 157

Series Available 581 | Selected 581

User Saved Queries

Series Countries Summary

Table Chart Download

Graph may contain a maximum of 5 series. To change selection, cli can be displayed.

x Gross Domestic Product at Market Price, Volume, Millions 2015

- Select “Series” and submit the query to start downloading.

Download

Series Countries Overview

Series Download Format

	1990	1991	1992
Series1			
Country1	1234	23	345
Country2	1234	23	345
Series2			
Country1	1234	23	345
Country2	1234	23	345

Selected Countries

Deselect all

- South Asia (MPO countries)
- Sub-Saharan Africa (MPO countries)
- Middle East & North Africa (MPO countries)
- Latin America & Caribbean (MPO countries)
- Europe & Central Asia (MPO countries)
- East Asia & Pacific (MPO countries)
- Zimbabwe
- Zambia

Selected Series

Deselect all

- Upper middle-income poverty rate (\$6.85 in 2017 PPP)
- Lower middle-income poverty rate (\$3.65 in 2017 PPP)
- International poverty rate (\$2.15 in 2017 PPP)
- Emissions: Waste (ktCO2e), % change
- Emissions: Total GHG emissions (ktCO2e), % change
- Emissions: Energy intensity of GDP (ktOE / mn LCU), % change
- Emissions: Land-use change and Forestry (ktCO2e), % change
- Emissions: Industrial processes (ktCO2e), % change
- Emissions: GHG intensity of energy (ktCO2e / ktOE), % change
- Emissions: Energy, fugitive (ktCO2e), % change
- Emissions: Energy, fuel combustion activities (ktCO2e), % change

Time 1980 2027

Submit