

ETHIOPIA FDI SYNTHETIC CONTROL ANALYSIS - README

DESCRIPTION

This README has two parts.

-The first offers an overview of the technical details of the project including the research questions and the method employed. -The second part describes the do files that are deployed in the analysis with the syntax used for the visualization of results including all the Tables and Figures. -The data file is one file that is interchangeably labelled as either `final_master_data.dta` or `masterV2.dta` as two authors were working on these files simultaneously.

PROJECT OVERVIEW

This repository contains the complete synthetic control analysis examining the impact of Ethiopia's 2015 policy reforms on Foreign Direct Investment (FDI) flows. The study employs SCM to construct counterfactual scenarios and estimate causal treatment effects.

RESEARCH QUESTION

What was the causal impact of Ethiopia's 2015 policy intervention on FDI inflows?

ANALYTICAL FRAMEWORK

- Method: Synthetic Control Method (Abadie, Diamond & Hainmueller, 2010)
- Treatment Unit: Ethiopia
- Treatment Period: 2015
- Pre-treatment Window: 2000-2015
- Post-treatment Window: 2016-2023
- Outcome Variable: Annual FDI flows as a share of GDP

DATA STRUCTURE

- Panel Data Format: Country-Year observations
- Time Coverage: 2000-2023 (24 years)
- Cross-sectional Units: 48 countries
- Total Observations: 960 country-year pairs

IDENTIFICATION STRATEGY

The synthetic control method constructs a weighted combination of donor countries that best approximates Ethiopia's pre-treatment characteristics and

FDI trajectory.

Key Assumptions:

1. No interference between units (SUTVA)
2. No anticipation effects before 2015
3. Donor pool countries unaffected by Ethiopia's treatment

DONOR POOL SPECIFICATIONS

PRIMARY ANALYSIS (All Donors):

- 47 potential donor countries
- Excludes: Countries with major policy changes 2000-2019
- Selection Criteria: Data availability and comparability

ROBUSTNESS CHECK (Alternative Donor Pool Specification):

- Based on ChatGPT deep search, country selection
- 9 carefully selected donors
- Countries: Azerbaijan, Cape Verde, Chad, Comoros, El Salvador, Equatorial Guinea, Kiribati, Nigeria, Morocco
- Selection based on: Similar development patterns, no major FDI reforms

PREDICTOR VARIABLES

ECONOMIC FUNDAMENTALS:

- `log_gdp`: Logarithm of GDP (constant 2015 USD)
- `gdp_growth`: Annual GDP growth rate (%)
- `gdp_per_capita`: GDP per capita (constant 2015 USD)
- `annual_inflation`: Consumer price inflation rate (%)

INSTITUTIONAL QUALITY (World Bank Governance Indicators):

- `control_corruption`: Control of corruption percentile rank
- `government_effectiveness`: Government effectiveness percentile rank
- `political_stability`: Political stability percentile rank
- `regulatory_quality`: Regulatory quality percentile rank
- `voice_accountability`: Voice and accountability percentile rank

CONFLICT & SECURITY:

- `conflict_intensity_index`: Armed conflict intensity measure

OUTCOME HISTORY:

- `fdi_flow(t)`: Annual FDI flows for each pre-treatment year $t \in \{2000, \dots, 2015\}$ as a share of GDP

DATA SOURCES

1. The main outcome variable is FDI inflow as a percentage of GDP, measured in constant 2015 U.S. dollars, representing direct investment equity flows, including equity capital, reinvestment of earnings, and inter-company debt as a share of GDP. This data is obtained from United Nations Conference on Trade and Development (UNCTAD). Accessed from <https://unctadstat.unctad.org/datacentre/dataviewer/US.FdiFlowsStock> Approximate date of download 28.03.2025
2. The macroeconomic indicators including GDP, GDP growth, GDP per capita and inflation drawn from the World Development Indicators (WDI) Accessed from <https://unctadstat.unctad.org/datacentre/dataviewer/US.FdiFlowsStock> Approximate date of download 28.03.2025
3. Conflict Intensity Index obtained from Uppsala Conflict Data Program Accessed from Uppsala Data Conflict Programme Armed dataset Conflict Dataset version 25.1 Note: Methodology for creating the conflict intensity index used in the dataset is described in Table A2 in the paper Approximate date downloaded: 28/08/2025
4. Control of Corruption obtained from Worldwide Governance Indicators (WGI) Worldwide Governance Indicator Accessed from <https://www.worldbank.org/en/publication/worldwide-governance-indicators/documentation#4> Approximate date downloaded: 28/08/2025
5. Government effectiveness obtained from Worldwide Governance Indicators (WGI) Accessed from <https://www.worldbank.org/en/publication/worldwide-governance-indicators/documentation#4> Approximate date downloaded: 28/08/2025
6. Regulatory Quality obtained from Worldwide Governance Indicators (WGI) Accessed from <https://www.worldbank.org/en/publication/worldwide-governance-indicators/documentation#4> Approximate date downloaded: 28/08/2025
7. Voice and Accountability obtained from Worldwide Governance Indicators (WGI) Accessed from <https://www.worldbank.org/en/publication/worldwide-governance-indicators/documentation#4> Approximate date downloaded: 28/08/2025

OPTIMIZATION PROCEDURE

The synthetic control weights are determined by solving:

$$\min_{\Sigma} \sum_w ||X - XW||^2$$

Where:

- X : Ethiopia's predictor characteristics

- X : Donor countries' predictor matrix
- W : Vector of donor weights ($w_1, \dots, w_n$)
- V : Diagonal matrix of predictor importance weights

CUSTOM V-MATRIX ESTIMATION:

V-matrices were pre-estimated using cross-validation procedures to optimize pre-treatment fit quality (RMSPE minimization).

STATISTICAL INFERENCE

Inference conducted via placebo tests (permutation-based):

1. Run synthetic control for each donor country as placebo treatment
2. Calculate treatment effect ratios
3. Assess statistical significance through ranking of actual vs. placebo effects

ROBUSTNESS CHECKS

1. Alternative donor pool specifications
2. Different pre-treatment periods
3. Placebo-in-time tests (false treatment dates)
4. Leave-one-out jackknife sensitivity analysis

Stata DO FILE STRUCTURE

This section describes the Stata .do files used for data cleaning, analysis, and visualization. Each do file is fully annotated with table and figure names with further descriptions to clarify its purpose and expected outputs.

1. The do files labelled "Master_Merge_Clean" is for data construction (see below)
2. The do file labelled "Main_Analysis_Tables_Figures" produces Figures 2-7, A4, A6 and Tables 1 and A1 (in the Appendix). It also produces the outputs referenced in the paper's discussion.
3. The do file labelled "Main_Alternative_Pool_Final" produces Table A4 and Figure A5. This uses a custom V matrix based on R code with the restricted raw data 'filtered.csv' as the input, it will output the "Alternative_Donor_Vweight.xlsx".
4. The do file labelled "Figure 1" produces Figures 1 and A1 (in the Appendix)
5. The do file labelled "Figure A2" produces Figure A2 (in the Appendix)
6. The do file labelled "Figure A3" produces panels A and B in Figure A3 (in the Appendix)
7. The do file labelled "Figure A7" produces Figure A7 (in the Appendix)
8. The do file labelled "Figure A8" produces Figure A8 (in the Appendix)

9. The R file “Alternative_donor_pool” produces Figure A5, and the custom V matrix used in the Stata do file “Main_Alternaitve_Pool_Final”. See below for the detailed description of the R file structure
10. The deep search AI result can be found [here](#)

R FILE STRUCTURE

This section describes the R files and how its output is used in Stata. It shows how the custom weighting matrix for the alternative pool and Figure A5 are generated in R and subsequently incorporated into the Stata do-file. Figure A5 is produced by both R and Stata with the use of the custom weighting matrix

1. The R script “Alternative_Donor_Pool.R” generates Figure A5 and produces the custom V matrix used in the Stata do-file “Main_Alternaitve_Pool_Final.do”.
2. The R script requires the restricted raw dataset “filtered.csv” as input. This input is drawn from the masterV2 by removing countries that have introduced FDI reforms according to our AI deep search (see line 191 below)
3. Running the R script with “filtered.csv” as input will create Figure 5 and the Excel file “Alternative_Donor_Vweight.xlsx”.
4. For consistency Figure A5 is reproduced using the Stata do-file “Main_Alternaitve_Pool_Final.do”, which uses these values from R output as hard-coded inputs. Each number was manually copied from the “Alternative_Donor_Vweight.xlsx” file and matched to its corresponding variable.

To run these .do files on your own system: 1. Update the working directory: change the file path at the top of each .do file to match your local directory structure. This is the only modification most users will need to make. 2. Install required add-ons. A few scripts call the scmrunner package. If you don’t already have it installed, you may want to install it before running the code.

Raw Data description

- Data files labelled “masterV2” and “final_master_data” are precisely the same data sets that are only different in names. They are produced by running “Master_Merge_Clean” do file. Please see description above in line 205-303 in this file.
- Date file labelled “UNCTAD” is raw data downlonad from UNCTAD’s website, see line 91-93
- Data file labelled “filtered.csv” is the file that is used in the alternative pool analysis in the R file. This is simply obtained by restricting the list of countries to those which have not introduced FDI reforms during the study period based on the deep AI search indicated in line 168.

MASTER__MERGE__CLEAN.DO

This section documents the data construction process for the primary raw dataset used in the analysis, as implemented in the master_merge_clean.do file.

PURPOSE

The first section of the .do file sets up the directory structure so that multiple users can easily access datasets and related .do files. It then calls a series of data cleaning scripts, each responsible for cleaning raw data from different sources and preparing them for merging.

The final section merges all the cleaned datasets into a single master dataset (masterV2.csv), which is ready for analysis.

How to change user:

The first section of the .do file is designed to easily switch between users. Each user needs to update the global path variable, which automatically adjusts all directory references throughout the project.

To set up a new user:

- Assign a user number: Edit lines 44–46 to assign yourself a unique user number.
- Define your path Add your file paths under the corresponding user number in the global path section (lines 62–72).

Once the user switch is updated, run the master_merge_clean.do file. This will clean all raw datasets, merge them, and output the final masterV2.csv file for analysis.

PURPOSE OF EACH .DO FILE CALLED:

CLEANING .DO FILES:

- Original/raw_master.do:
 - Input: finaldataset30_COPY.xlsx
 - Input Source:
 - Output: master.dta
 - Core functions of the code: rename variable names iso_code, country and year
- Armed Conflict/raw_armedconflict.do
 - Input:
 - * UcdpPrioConflict_v25_1.dta
 - * country_codes_ref.csv
 - Input Source:
 - * Output:
 - armed_conflict_cleaned.dta
 - country_codes_ref.dta

- Core functions of the code: Clean up location string, drop missing location, and clean country names
- Corruption Index/raw_v2.do
 - Input: CPI from year 2000-2011
 - Input Source:
 - Output: corruption_index_final.dta
 - Core functions of the code: clean and append annual CPI data from year 2000-2011
- Economic Freedom Indicator/raw.do
 - Input:
 - * heritage-index.csv
 - * country_codes_ref.dta
 - Input Source:
 - Output: heritage_index_cleaned.dta
 - Core functions of the code: Clean country names and merge with country codes
- Global Competitiveness/raw.do
 - Input: WEF_GCI.csv
 - Input Source:
 - Output: WEF_GCI.csv
 - Core functions of the code: No changes
- Government Stability Measures/raw_gov_stab.do
 - Input: wgidataset.dta
 - Input Source:
 - Output: gov_stab_measures.dta
 - Core functions of the code: Reshape the data, and clean variable names for consistency across all datasets
- Trade Open/trade_open_raw.do
 - Input: trade_open_raw.csv
 - Input Source:
 - Output: trade_open.dta
 - Core functions of the code: Reshape data from wide to long
- Transport Infra/raw.do
 - Input: transport_infra.csv
 - Input Source:
 - Output: transport_infra.dta
 - Core functions of the code: Reshape data from wide to long

MERGE .DO FILE:

- Merge Code/master_merge.do
 - Inputs:
 - * master.dta
 - * armed_conflict_cleaned.dta
 - * corruption_index_final.dta
 - * heritage_index_cleaned.dta

- * gov_stab_measures.dta
- * trade_open.dta
- * transport_infra.dta
- Output:
 - * masterV2.dta
 - * masterV2.csv
- Core functions of the code:
 - * Created intensity index for armed conflicts
 - * Merges the data with other indicators based on country_code, iso, and year