

DIME Analytics

Reproducibility Package Checklist

v1.4

Required Content and Practices

The reproducibility package must be shared as a .zip file, GitHub link or shared folder containing the items listed below.

1. Files

A filled version of this checklist.

A README file describing the folders and datasets included in the package, and any general or ad-hoc instructions about the directory structure or for the code to run.

A PDF version of the research output being reviewed (paper, brief, report).

2. Code

Software used (including version):

Operating system used:

All scripts needed to reproduce the analysis is included in the reproducibility package.

The reproducibility package includes a **master script**.

Master script path and name:

The items below only apply if the reproducibility package includes a master script.

Approximate time it takes to recreate all results from the the master script:

RAM size of computer used to calculate time:

The master script re-creates all raw outputs (e.g. tables and figures) when executed.

All raw outputs are re-created **as they appear in the paper**, requiring no manual changes.

All raw outputs are re-created in the same folder structure as they were submitted in the reproducibility package.

The master script **installs any commands required** (e.g. from SSC in Stata, CRAN in R, etc.).

In Stata, the master do file sets **critical configurations** such as version, matsize, and varabbrev (either sets directly or through a wrapper command such as ieboilstart from ietoolkit).

In R, all necessary packages to run the code are loaded in the master script.

All file paths in the code use only forward slashes, i.e. C:/users/....

Select only one of the following two options.

Results are reproducible using the **most recent version** of the user-written commands (as available in CRAN/SSC) at the time of submitting the reproducibility package.

The master script installs the **exact version** of each package required, either from within the reproducibility package itself, or from a remote location.

3. Data

The reproducibility package includes:

Fully de-identified data. **Any exception where DIME Analytics needs to access restricted or sensitive data sets will require additional ethics and security measures, and will be handled on a case-by-case basis.**

All data necessary to run the master script that reproduces the analysis.

Indicate the stage of the data this package begins reproducibility from (select only one):

Raw data

Cleaned data

Cleaned data with constructed variables

Indicate if the data used in this project or reproducibility package has been published:

Yes, it has been published

No, it has not been published

If it has been published, please indicate the URL or DOI of the data publication:

URL / DOI 1:

URL / DOI 2:

URL / DOI 3:

4. Outputs

The reproducibility package includes:

All raw outputs used for the paper (e.g. tables and figures), as created by the code.

Files that clearly correspond by name to an exhibit in the paper, and vice versa.



The items below may be left blank if they don't apply to the outputs in the package:

If any formatting to a particular output is needed after running the code that creates it, all additional steps needed to create this output **as it appears in the paper** are indicated in the **master script** or a **readme** file. (DIME Analytics recommends against this practice.)

If any tables or figures included in the paper are not created through code (for example, a table with variable definitions), this is indicated in a **readme**.

Access to Data and Documentation (recommended)

Code repository

URL:

Code repository is public

Study registration

URL:

Study registration is public

Pre-analysis plan

URL:

Pre-analysis plan is public

Other document:

URL:

Document is public

Other document:

URL:

Document is public

Recommended Practices

These practices are recommended but not required:

The package includes all the necessary files to recreate the final analysis data set from the raw data (e.g., scripts for data cleaning and indicators construction are also provided).

Scripts for data cleaning, variable creation, and analysis are separate; and analysis scripts do not include any data processing, unless necessary for the creation of a table or graphic.

Variable construction scripts include detailed comments about each variable created.

Analysis scripts do not depend on having the results of other scripts in memory, except for the Master script.

Separate scripts are provided for each exhibit.

All scripts are well-commented and formatted, such that one can easily identify functional chunks of code and evaluate whether they correctly implement the econometric or statistical process described.

Graphics are output as .eps files or other vector images when possible.

Tables are output as .csv or .tex files or other raw text files when possible.

The submission package includes code to create all in-text numerical citations that are not drawn directly from tables and figures.

In-text numerical citations (other than those drawn directly from tables and figures) are computed and recorded in a dynamic document format like .ipynb, .stmd using markstat in Stata, or .Rmd using R.