

Overview

This reproducibility package relates to the study **“Which Data Do Economists Use to Study Corruption? A Cross-Section of Corruption Research”** by James Anderson and Akanksha Baidya of the World Bank. For more information, please contact James Anderson janderson2@worldbank.org .

The package consists of:

- Two Stata data files containing the data compiled as part of the research.
 - o “Use of Corruption Data by Economists.dta”. Includes data on 339 articles which were in search results and include JEL codes. This is the main sample.
 - o “Use of Corruption Data by Economists_Longlist.dta” . Includes some basic information on 882 articles which were in search results, whether or not they included JEL codes.
- A Stata .do file for reproducing findings in the report. The .do file accesses and works with both .dta files.
 - o “Use of Corruption Data by Economists.do”

Data availability

All data used in the study are publicly available.

Data sources

There are five sources of data:

1. **Search results for journal articles indexed in the EBSCO and EconLit databases with the word “corruption” in the abstract.**
 - a. **Use:** The search results provided basic information about journal articles with corruption mentioned in the abstract. For each paper, additional information was added based on the content of the articles.
 - b. **Citation:**
 - i. For references to the compilation of papers, please cite:
James Anderson and Akanksha Baidya, (2024). “Which Data Do Economists Use to Study Corruption? A Cross-Section of Corruption Research”. World Bank.
 - c. **Filenames:**
 - i. “Use of Corruption Data by Economists.dta” . Includes data on 339 articles which were in search results and include JEL codes. This is the main sample.
 - ii. “Use of Corruption Data by Economists_Longlist.dta” . Includes some basic information on 882 articles which were in search results, whether or not they included JEL codes.
 - d. **Source:** The data was compiled by: (i) Searching EBSCOhost platform, including EconLit, for the keyword “corruption” within the abstracts of journal articles, limited to publications from 2022. (ii) From those search results, we narrowed the sample to papers with Journal of Economic Literature (JEL) codes. (iii) We subsequently searched EconLit directly with full text using the same parameters. (The JEL codes come from either the articles themselves (e.g., on the first page of the published version), or from search results in the case of the EconLit search. Note that many articles had differing

sets of JEL codes in the search results and on the cover of the published article. This is discussed further in the paper.) (iv) We examined the papers to make judgments about the features of the papers, for example, which data source for “corruption” was used, whether it is a country-focused or regional study, and others. These assessments are included in the data file “Use of Corruption Data by Economists.dta”. (v) For the longlist of papers that came up in searches, we identified whether the publications were in the Top-20 lists of journals according to Google Scholar for each of three categories (described below), and for that subset of papers we identified whether they used data on corruption.

- e. **URL:** The EBSCOhost search engines are available for subscribers here: <https://www.ebsco.com/>
- f. **Access year:** Search parameters focused on publications in 2022. The searches were undertaken in 2023 and 2024.

2. Journal ranks based on Combes & Linnemer (2010).

- a. **Use:** This was used for Table 1 and Table 6.
- b. **Citation:**
Combes, P.-P., & Linnemer, L. (2010). Inferring Missing Citations: A Quantitative Multi-Criteria Ranking of All Journals in Economics. HAL Open Science.
- c. **Filename:** The journal rank information is included in “Use of Corruption Data by Economists.dta”.)
- d. **URL:** <https://shs.hal.science/halshs-00520325>
- e. **Access year:** 2023, although the paper was published in 2010.

3. Journal ranks based on IDEAS/RePEc H-Index for Journals.

- a. **Use:** This was used for Table 1 and Table 6.
- b. **Citation:**
IDEAS. (2023). IDEAS/RePEc H-Index for Journals. Retrieved December 18, 2023.
- c. **Filename:** The journal rank information is included in “Use of Corruption Data by Economists.dta”.)
- d. **URL:** <https://ideas.repec.org/top/top.journals.hindex.html> .
- e. **Access year:** 2024.

4. Journal of Economic Literature codes.

- a. **Use:** This was used for Table 2 and 3, providing the definitions of the JEL codes associated with the papers in the sample.
- b. **Citation:**
American Economic Association. (2023). JEL Classification System / EconLit Subject Descriptors. Retrieved December 1, 2023.
- c. **Filename:** The JEL codes associated with the papers are included in “Use of Corruption Data by Economists.dta”.
- d. **URL:** <https://www.aeaweb.org/econlit/jelCodes.php> .
- e. **Access year:** 2023.

5. Journal ranks based on Google Scholar.

- a. **Use:** The Top 20 journals according to Google Scholar for economics, political science, and public policy and administration were identified and used for Table 8.
- b. **Citations:**
 - Google Scholar Top 20 Publication for Economics (subcategory of Business, Economics & Management).
 - Google Scholar Top 20 Publication for Political Science (subcategory of Social Sciences).
 - Google Scholar Top 20 Publication for Public Policy & Administration (subcategory of Social Sciences).
- c. **Filename:** The journal rank information is included in “Use of Corruption Data by Economists_Longlist.dta”.
- d. **URLs:**
 - https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=bus_economics ;
 - https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=soc_politicalscience ;
 - https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=soc_publicpolicyadministration .
- e. **Access year:** 2024 (May).

Instructions for Replicators

1. Save the data files and the .do file to a folder on your computer.
2. Open the .do file and replace the name of the working folder with the one where you have placed the datafiles and will use as your workspace.
3. Run the .do file “Use of Corruption Data by Economists.do”.
4. The output exports tables as Excel files and also generates all of the information necessary to reproduce tables and text references in the paper.

List of Exhibits

The provided Stata code and Excel file reproduces:

- All numbers provided in text in the paper
- All tables and figures in the paper

Exhibit name	Output filename	Script	Note
Table 1. Articles published in highly ranked economics journals	Table1.xlsx	“Use of Corruption Data by Economists.do”	
Table 2. Frequencies of JEL Categories for papers in the full sample	Table2.xlsx	“Use of Corruption Data by Economists.do”	

Table 3. Frequencies of the most common JEL Codes for the full sample	Table3.xlsx	“Use of Corruption Data by Economists.do”	
Table 4. Numbers of articles by geographic focus	Table4Top.xlsx Table4Bottom.xlsx	“Use of Corruption Data by Economists.do”	
Table 5. Data sources for studies on corruption, by geographical category	Table5sources.xlsx Table5lastrow.xlsx	“Use of Corruption Data by Economists.do”	
Table 6. Data sources for studies on corruption, for ranked journals	Table6sources.xlsx Table6lastrow.xlsx	“Use of Corruption Data by Economists.do”	
Table 7. Data sources for studies on corruption, by direction of causality	Table7sources.xlsx Table7lastrow.xlsx	“Use of Corruption Data by Economists.do”	
Table 8. Use of data on corruption in economics, political science, and public administration	Table8top.xlsx Table8middle.xlsx Table8bottom.xlsx	“Use of Corruption Data by Economists.do”	

Requirements

Computational Requirements

Computational requirements are modest and easily managed on a PC.

Software Requirements

- Stata version 18.
- Microsoft Excel

Memory and Runtime and Storage Requirements

Data files are small and easily managed on a PC.

Code Description

1. “Use of Corruption Data by Economists.do” accesses the two data files to perform all of the necessary calculations. The .do file is organized by the Tables as they appear in the paper.

Folder Structure

Stata code and files

- └─ “Use of Corruption Data by Economists.do”
- └─ “Use of Corruption Data by Economists.dta”
- └─ “Use of Corruption Data by Economists_Longlist.dta”