

DATA DESCRIPTION SHEET

Corporate Tax System Complexity and Investment Sensitivity to Tax Policy Changes

Harald Amberger
Vienna University of
Economics and Business

John Gallemore
University of North Carolina

Jaron Wilde
University of Iowa

December 17, 2025

This data description sheet accompanies our manuscript accepted for publication in the *Journal of Accounting Research* and provides the following information required under *JAR*'s data and code sharing policy.¹ Data handling and programming procedures are described in detail below. All authors vouch for the stated sources of the raw data.

1. Authors (Item #1):

All three authors were fully informed about the data collection process and had continuous access to all data and code through a shared Dropbox folder. Harald Amberger was responsible for handling the raw data, preparing and merging datasets to construct the final sample for the analyses, and conducting all empirical analyses. Two research assistants assisted with collecting statutory corporate income tax rate data and identifying tax rate change events used in the stacked difference-in-differences design. A third research assistant assisted with collecting information on reduced corporate tax rates.

2. Data Sources (Items #2 and #4):

Most of primary data collections, including financial statement and ownership data from Orbis and corporate tax system complexity data from World Bank, were conducted in March 2022. Additional data collections occurred in February 2023 (country-level macroeconomic data from the World Bank), April 2023 (OECD Tax administration Survey), August 2023 (political risk data), September 2023 (Worldwide Governance Indicators), October 2023 (country-level effective tax rate data), February 2025 (Tax Complexity Index and data from Compustat Global), and July 2025 (country-level fiscal condition data from IMF and World Bank). Statutory corporate tax rate data, information on tax rate change events, and data on reduced corporate tax rates were collected during September 2022, April 2023, January 2025, and March 2025.

The specific data sources used in the analysis are described below:

a. Orbis data

- Source: Moody's (formerly Bureau van Dijk) Orbis database; accessed via the Orbis web interface.
- Download: To obtain the initial sample reported in Table 1, we applied the following search criteria in Orbis:
 - Status: Active companies, Unknown situation
 - World region/Country/Region in country: OECD, European Union [27], EFTA, United Kingdom

¹ In line with the Data and Code Sharing Policy for the *Journal of Accounting Research* as of December 2014 (updated April 2024). The policy can be access at <https://www.chicagobooth.edu/research/chookaszian/-/media/a08a88aaca645989a7d1e62eb928161.ashx>

- Operating revenue (Turnover): All companies with a known value, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, for at least one of the selected periods
- Consolidation code: U1 (unconsolidated accountings with no consolidated companion), U2 (unconsolidated accounts with a consolidated companion)
- P/L before tax: All companies with a known value, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, for at least one of the selected periods
- Tangible fixed assets: All companies with a known value, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, for at least one of the selected periods
- Entity type: Corporate
- Status: Active companies
- Subsidiaries with Ultimate Owners by profile: UO, GUO and DUO; Def. of the UO: min. path of 50.01%, known or unknown shareholder

We selected the required variables capturing financial statement and ownership information and downloaded the data in March 2022 as 504 individual .txt files due to size restrictions imposed by the Orbis web interface. The datasets included firm identifiers such as firm name and the Bureau van Dijk (BvD) ID number. A second data extraction was conducted in March 2022 to obtain depreciation expense and long-term debt variables that were not included in the initial download.

b. Primary Tax Complexity Measure:

- Source: World Bank's Doing Business Survey; obtained from the Paying Taxes component of the survey ([Methodology for Paying Taxes](#)).
- We downloaded the data in March 2022 in machine-readable format. We then merged these data with the Orbis data. No manual data entry was involved.

c. Statutory corporate income tax rate data:

- Sources: Hand-collected from public sources, including tax rate tables and publications from the Tax Foundation, OECD, KMPG, PWC, and EY, as well as news articles, government announcements, and domestic law.
- We manually collected statutory tax rate data in February and March 2022 and entered it into Excel. Information on tax rate change events and data on reduced corporate tax rates was subsequently collected in September 2022, April 2023, January 2025, and March 2025. These data were then merged with the Orbis data.

d. Country-level macroeconomic controls:

- Source: World Bank's World Development Indicators; obtained from the public database ([World Development Indicators | DataBank](#)).
- We downloaded the data in February 2023 in machine-readable format. We then merged these data with the Orbis data. No manual data entry was involved.

e. Country-level institutional controls:

- Source: World Bank's Worldwide Governance Indicators; obtained from the public database ([Home | Worldwide Governance Indicators](#)).
- We downloaded the data in September 2023 in machine-readable format. We then merged these data with the Orbis data. No manual data entry was involved.

f. EATR data:

- Source: Effective Tax Levels using the Devereux/Griffith Methodology and as reported in Spengel et al. (2020); obtained from the report published by the European Commission ([final_report_2020_effective_tax_levels_revised_en.pdf](#)).
- We manually extracted the data from the publicly available PDF report in October 2023 and entered it into Excel. These data were then merged with the Orbis data.

g. Compustat Global data:

- Source: Compustat Global Database; obtained via the WRDS web access.
- We downloaded data in February 2025 for fiscal years 2012-2019 in machine-readable format. We then applied the relevant code to compute aggregate effective tax rates and merged the resulting data with the Orbis data. No manual data entry was involved.

h. Political risk data:

- Source: Tax policy risk data from Hassan et al. (2019); obtained from the public website (www.firmlevelrisk.com).
- We downloaded firm-level data in August 2023 in machine-readable format. We then applied the relevant code to compute country-year averages and merged the resulting data with the Orbis data. No manual data entry was involved.

i. Tax enforcement data:

- Source: 2015 OECD Tax Administration Survey; obtained from the OECD website ([Tax Administration 2015 | OECD](#)).
- We manually extracted the data from the publicly available PDF report in April 2023 and entered it into Excel. These data were then merged with the Orbis data.

j. Tax Complexity index:

- Source: Tax Complexity Index published by Hoppe et al. (2023); obtained from the public website ([Tax Complexity Index](#)).
- We downloaded the data in February 2025 in machine-readable format. We then applied the relevant code to backfill missing values of the index and merged the resulting data with the Orbis data. No manual data entry was involved.

k. Additional country-level fiscal controls:

- Sources: Data on government interest was obtained from the World Bank ([Interest payments \(% of revenue\) | Data](#)); data on central government gross debt and government primary balance was obtained from the IMF ([IMF DataMapper](#)).
- We downloaded the data in July 2025 in machine-readable format. We then merged these data with the Orbis data. No manual data entry was involved.

3. Proprietary Data (Item #3):

No proprietary data were used in the analyses. All data are accessible either through subscriptions to the relevant databases or via the publicly available sources listed above. No survey instrument was used and no experiment was conducted.

4. Data Processing (Item #5):

All data were obtained directly from the sources described under item #2. Data processing involved preparing the raw data and merging multiple sources into a firm-year panel, as described in the manuscript. All steps were implemented using the program code described below.

5. Code, Identifiers, and Comprehensive Log File (Items #6 and #7):

We provide a set of Stata codes with detailed comments, along with a corresponding comprehensive log file (AGW_comprehensive_log_file.log). The do-files reproduce all tables and figures reported in the manuscript and the Online Appendix.

The replication package includes the following files (see AGW_Replication_Package_ReadMe.txt):

- 0_master.do: Master file that executes all remaining do-files in the correct sequence.
- 0_1_additional_orbis_data_prep: Prepares supplemental ORBIS firm-level data that are later merged with the primary sample.
- 0_2_country_level_data_prep: Prepares country-level data for merging with primary sample.
- 0_3_tax_rate_data_prep_announced_enacted: Prepares country-level data on announced/enacted tax rates for merging with the primary sample.
- 0_4_tax_rate_data_prep_aseffective: Replicates 0_3_tax_rate_data_prep_announced_enacted but prepares tax rates measured on an as-effective basis (i.e., tax rates that are effective in a given year).
- 1_data_sample_prep: Constructs the primary sample, including processing raw Orbis data, merging country-level data, constructing dependent and independent variables, applying sample restrictions, and cleaning the final dataset.
- 2_primary_analysis: Conducts the primary analyses and reproduces the corresponding tables and figures reported in the manuscript and the Online Appendix.
- 3_stacked_DiD.do: Conducts the secondary analyses using a stacked difference-in-differences research design and reproduces the corresponding tables and figures reported in the manuscript and the Online Appendix.

In addition, the file `primary_sample_BVD_IDs.csv` contains Bureau van Dijk (BvD) firm identifiers for all firms included in the primary sample (417,161 unique firms).

6. Retention of Data and Code (Item #8):

We commit to maintaining the data and program files for at least six years, consistent with National Science Foundation guidelines.