This document accompanies the SAS program and Stata .do file used to generate the final samples and analyses used in “Do Foreign Component Auditors Harm Financial Reporting Quality? A Subsidiary-Level Analysis of Foreign Component Auditor Use.”

The programs used are:

* DGLM\_Sample.sas
	+ Imports data from various sources, combines datasets, calculates variables, and exports file to Stata.
* DGLM\_SampleFinalization.do
	+ Finalizes the samples, aggregates subsidiary-level variables for use in multinational corporation (MNC) analyses

DGLM\_Sample.sas consists of 14 steps to generate the MNC Sample (Steps 1 through 9) and the Subsidiary Sample (Steps 10 through 14). This document briefly summarizes the data sources and output of each step.

MNC Sample:

1. Obtain Compustat Data from WRDS
	* Connects to WRDS, then merges and downloads data from Compustat Fundamentals Annual, Names, and Segments
2. Calculate Compustat Segment variables
	* Inputs: Compustat Segments data (from Step 1)
	* Calculates the total number of geographic segments, business segments, foreign segments, and foreign sales by company-year.
	* Outputs: dataset “Compmerged”
3. Calculate Compustat control variables
	* Inputs: dataset Compmerged (from Step 2)
	* Calculates Compustat control variables used in the MNC analyses and winsorizes continuous variables at the 1 and 99 percent levels. See SAS program for list of variables and the Appendix in the study for detailed variable definitions, sources.
	* Outputs: dataset “C1”
4. Import Form AP Data from Excel
	* Inputs: PCAOB Form AP Data. Form AP is downloaded directly from <https://pcaobus.org/Pages/AuditorSearch.aspx>. Prior to importing, the Excel function “text-to-columns” is used to break out information about individually identified component auditors into separate columns.
	* Outputs: dataset “FormAP”
5. Calculate MNC-level and country-level component auditor variables from Form AP
	* Inputs: dataset FormAP (from Step 4)
	* Calculates Form AP-based variables used in the MNC analyses and winsorizes continuous variables at the 1 and 99 percent levels. See SAS program for list of variables and the Appendix in the study for detailed variable definitions, sources.
	* Drop observations based with audit report dates before June 30, 2017 (the date on which component auditor disclosures became mandatory) and with fiscal year-ends prior to March 31, 2017 or with Form AP filings more than 60 days after the audit report date (observations with abnormally late Form AP filings)
	* Transposes data for individually identified component auditors on Form AP and generate dataset at the client/audit firm/component auditor-year level. This dataset is used to match individual component auditors to MNC subsidiaries in Step 12.
	* Outputs: dataset “F1” (MNC-level Form AP data), dataset “cmpcountries” (individual component auditor data)
6. Import Audit Analytics Data
	* Inputs: Audit Analytics (“AA”) Opinions data and Restatements data, downloaded directly from <https://www.auditanalytics.com> and saved as Excel files
	* Outputs: dataset “AA\_data” (Opinions dataset), dataset “restatements” (restatements data)
7. Calculate AA control variables
	* Inputs: dataset AA\_data (from Step 6), dataset “cbsa” (hand-collected Core based statistical area (CBSA) codes by city and state for summation of fees by city)
	* Calculates AA-based variables used in the MNC analyses. Sums audit fees by CBSA for calculation of industry expertise. See SAS program for list of variables and the Appendix in the study for detailed variable definitions, sources.
	* Outputs: dataset “A0”
8. Merge MNC-level datasets and calculate additional controls
	* Inputs: dataset “C1” (from Step 3), dataset “restatements” (from Step 6), dataset “A0” (from Step 7), dataset “F1” (from Step 5), dataset “arc” (from Hoitash and Hoitash [2018]), downloaded directly from <http://www.xbrlresearch.com>)
	* Merges MNC datasets based on year and company identifier.
	* Calculates variables used in the MNC analyses and winsorizes continuous variables at the 1 and 99 percent level. See SAS program for list of variables and the Appendix in the study for detailed variable definitions, sources.
	* Drop observations missing Form AP data (from Step 5), missing assets or sales data, or with assets less than or equal to $1 million.
	* Outputs: dataset “MNC”
9. Export dataset to Stata dta file
	* Inputs: dataset “MNC” (from Step 8)
	* Exports data to Stata format.
	* Outputs: Stata dataset “MNC”

Generating the Subsidiary Sample:

1. Import Orbis Subsidiary Data
	* Inputs: Orbis subsidiary financial information and MNC parent-ID from the Global Ultimate Owner (GUO) (downloaded from <https://orbis.bvdinfo.com>, see the full study for additional information)
		1. Orbis GUOs are matched to observations on Form AP based on CIK code
	* Renames variables and converts variables measured in thousands to millions
	* Outputs: dataset “orbis\_subs\_det”
2. Calculate Subsidiary-level control variables
	* Inputs: dataset “orbis\_subs\_det” (from Step 10), dataset “countrycode” (country-level data on IFRS use [from <https://www.ifrs.org/use-around-the-world/use-of-ifrs-standards-by-jurisdiction>], English Proficiency [from <https://www.ef.com/wwen/epi>], and Tax Haven [<https://sites.google.com/site/scottdyreng/Home/data-and-code/EX21-Dataset>] – see Appendix for definitions), dataset “wgi” (rule of law data collected from World Governance Indicators [<https://info.worldbank.org/governance/wgi/>])
	* Transposes Orbis data so that the unit of observation is the MNC-subsidiary-year level.
	* Calculates control variables used in the subsidiary-level analyses and winsorizes continuous variables at the 1 and 99 percent levels. See SAS program for list of variables and the Appendix in the study for detailed variable definitions, sources.
	* Output: dataset “orbis\_detail\_final”
3. Merge subsidiary data with Form AP component auditor data
	* Inputs: dataset “orbis\_detail\_final” (from Step 11), “cmpcountries” (from Step 5), dataset “F1” (from Step 5), dataset “C1” (from Step 3), dataset “orbis\_lastavail” (dataset downloaded from Orbis containing most recent data collection date), “orbis\_auditors” (dataset downloaded from Orbis containing statutory auditor data for each subsidiary), dataset “sub\_country\_mw” (hand-collected material weakness data at the country level)
	* Merges Orbis data with most recent data collection date (“orbis\_lastavail”), statutory auditor data (“orbis\_auditors”), Form AP data (“cmpcountries” and “F1”), MNC parent-level control variables (“C1”), and country-level material weakness data (“sub\_country\_mw”).
	* Drops observations based on the following criteria:
		1. Observations with most recent Orbis data collection date before 2017 (to limit observations with outdated information prior to merge with Form AP data)
		2. Observations that do not merge with MNC parent companies on Form AP or with principal auditors outside the U.S.
		3. Observations missing assets data, with assets less than $1 million, missing sales data, or missing MNC parent-level sales or MNC parent-level assets less than $1 million
	* Calculates additional control variables used in the subsidiary analyses. See SAS program for list of variables and the Appendix in the study for detailed variable definitions, sources.
	* Output: dataset “orbis\_variables”
4. Calculate component auditor variables and Jones (1991) abnormal accruals
	* Input: dataset “orbis\_variables” (from Step 12)
	* Calculates abnormal accruals for subsidiary analyses based on Jones(1991), controlling for ROA (Kothari et al. 2005), and winsorizes abnormal accruals and total accruals at the 1 and 99 percent levels. Calculates additional variables used in the subsidiary analyses. See SAS program for list of variables and the Appendix in the study for detailed variable definitions.
	* Merges abnormal accruals data into the final dataset, and keeps only the largest subsidiary within each MNC-country-year by total assets
	* Output: dataset “orbis\_final\_allsubs”
5. Export to Stata dta file
	* Input: dataset “orbis\_final\_allsubs”
	* Exports dataset to Stata for analyses
	* Output: Stata dataset “Subsidiary”

DGLM\_SampleFinalization.do imports the dataset “Subsidiary” and performs data finalization steps, including the following:

* Aggregate the total number of subsidiary-countries to the MNC parent-level as control variable (lnSubsidiary)
* Drop observations for which component auditor use is collinear with country fixed effects
* Aggregate subsidiary data for use in the MNC Coverage Analyses and calculate *Sales Coverage* as the percentage of total subsidiary sales by MNC Parent audited by component auditors.
* Merge aggregate subsidiary data into MNC dataset (“MNC.dta”) and calculate final variables, including *Any Coverage*, *MNC Ln Subsidiaries*, and *sub\_saleperc* (used to restrict the sample to observations with Orbis subsidiaries contributing at least 10% of total MNC sales).

See Stata code for specific steps and the Appendix in the study for detailed variable definitions and sources.

References

Dyreng, S. D., and B. P. Lindsey. 2009. Using Financial Accounting Data to Examine the Effect of Foreign Operations Located in Tax Havens and Other Countries on U.S. Multinational Firms’ Tax Rates. Journal of Accounting Research 47 (5): 1283–1316, https://doi.org/10.1111/j.1475-679X.2009.00346.x.

Hoitash, R., and U. Hoitash. 2018. Measuring Accounting Reporting Complexity with XBRL. The Accounting Review 93 (1): 259–87, https://doi.org/10.2308/accr-51762.

Jones, J. J. 1991. Earnings Management During Import Relief Investigations. Journal of Accounting Research 29 (2): 193, <https://doi.org/10.2307/2491047>.

Kothari, S. P., A. J. Leone, and C. E. Wasley. 2005. Performance matched discretionary accrual measures. Journal of Accounting and Economics 39 (1): 163–97, https://doi.org/10.1016/j.jacceco.2004.11.002.