Earnings Management and Investor Protection:

Accounting Reading Group - Assignment III

Melisa Mazaeva

Aug 9, 2024

Abstract

This project uses the TRR 266 Template for Reproducible Empirical Accounting Research (TREAT) to provide an infrastructure for open science-oriented empirical projects. Leveraging Worldscope data on financial information, this repository showcases a reproducible workflow that integrates Python scripts for data analysis. The project's output demonstrates a comprehensive application of skills to replicate and extend the findings from the seminal paper by Leuz, Nanda, and Wysocki (2003), particularly in providing descriptive statistics for four individual earnings management measures, as well as an aggregate earnings management score across various countries. In doing so, it documents and discusses the research design choices made and the variations between the original and reproduced results. This code base, adapted from TREAT, provides an overview of how the template can be used for this specific project and how to structure a reproducible empirical research project.

1 List of Abbreviations

EM: Earnings Management

IDE: Integrated Development Environment

UK: United Kingdom

US: United States

WRDS: Wharton Research Data Services

2 Introduction

The aim of this paper is to illustrate the use of open science tools in empirical accounting research.

The original study by Leuz, Nanda, and Wysocki (2003) examines systematic differences in earnings management (EM) across 31 countries and proposes that these differences are influenced by the level of investor protection. Specifically, the paper argues that stronger investor protection reduces earnings management by limiting insiders' ability to acquire private control benefits.

This paper presents the replication of Table 2 Panel A from the seminal study by Leuz, Nanda, and Wysocki (2003). The original table provides descriptive statistics for four individual earnings management measures and the aggregate earnings management score across various countries, highlighting significant differences in EM practices between Continental Europe, Asia, and Anglo-American countries.

The project uses data from the Worldscope Database via Wharton Research Data Services (WRDS), focusing on fiscal years 1990 to 1999 and covering multiple countries. The replication aims to follow the research design by Leuz, Nanda, and Wysocki (2003) as closely as possible with the available data. It employs four individual earnings management measures: earnings smoothing using accruals (EM1), the correlation between changes in accruals and cash flows (EM2), the magnitude of accruals (EM3), and the ratio of small profits to small losses (EM4). These measures are combined to create an aggregate earnings management score, providing a comprehensive overview of EM practices.

The analysis includes pulling and filtering the relevant financial data, identifying relevant companies and countries based on the authors' requirements, calculating the EM measures for each country, and examining the relationship between each country's aggregate earnings management score and its ranking. Explicit assumptions were made whenever the paper by Leuz, Nanda, and Wysocki (2003) was unclear on how to proceed. The complete computation code for EM measures is available in `code/python/do_analysis.py` for detailed review if necessary. Code snippets will be presented only in exceptional cases to highlight specific points.

The replicated table is shown below, followed by the research design assumptions in Section 3, documentation of the replication steps in Section 4, a detailed comparison in **?@sec-results**,

and concluding remarks in **?@sec-conclusion**.

3 Research Design Choices and Assumptions

The aim of Assignment III is to replicate a specific empirical table (Table 2 Panel A) from the seminal paper by Leuz, Nanda, and Wysocki (2003). This table involves calculating the EM measures for firms across various countries over a defined period and examining the relationship between these measures and investor protection. The replication process includes data loading, preparation, cleaning, and normalization, followed by the application of statistical methods to compute and interpret financial metrics. For Assignment III, I pulled data from the Worldscope database through WRDS and used the Python programming language to carry out the empirical analysis. Visual Studio Code was used as the Integrated Development Environment (IDE) for writing, debugging, and optimizing the Python code.

The replication is based on data pulled from the Worldscope database, specifically from the `wrds_ws_company` and `wrds_ws_funda` tables, which were merged for the analysis. The first table provides company profile information, including items such as ISIN, Worldscope Identifiers, company name, and the country where the company is domiciled (Wharton Research Data Services 2024a). The latter table contains Fundamentals Annual data at the company-year level, including items such as total assets, net income, and other relevant financial variables (Wharton Research Data Services 2024b).

Following Leuz, Nanda, and Wysocki (2003), I focus the analysis on companies across various countries, ensuring that the data accurately reflects the fiscal years 1990 to 1999 as specified in the original study. The replication aims to mirror the research design as closely as possible with the available data.

In addition, I impose the following assumptions to ensure clarity and consistency where the paper by Leuz, Nanda, and Wysocki (2003) does not provide explicit guidance:

1. The original paper references the November 2000 version of the Worldscope Database. However, the data used for this analysis represents the latest available version, updated in July 2024, with quarterly frequency updates (Wharton Research Data Services 2024c). Due to potential adjustments and updates made to the database since November 2000, there may be

differences between the databases that could affect the results. For example, companies may restate financials after the original reporting period, so these restatements are reflected in the later database version rather than the historical one. Moreover, the data vendor Refinitiv regularly updates its databases to correct errors and add new information, which may be included in the later data but not in the November 2000 snapshot.

2. The original paper outlines key terms that will be used in this project to ensure consistency and accuracy in the replication. Leuz, Nanda, and Wysocki (2003) define earnings management as the manipulation of a firm's reported economic performance by insiders to deceive certain stakeholders or to affect contractual outcomes. The authors describe investor protection as a key institutional factor that limits insiders' acquisition of private control benefits, thereby reducing their incentives to manage accounting earnings by ensuring strong and well-enforced rights for outside investors. Finally, private control benefits are the advantages that insiders can gain from controlling a firm, which can include financial gains or other benefits not shared with other stakeholders (Leuz, Nanda, and Wysocki 2003).

3. While pulling the data for analysis, I encountered negative values for some key financial metrics such as operating income (`item1250`) and net income before preferred dividends (`item1651`). The paper by Leuz, Nanda, and Wysocki (2003) does not explicitly specify how to handle negative values in key financial metrics. For the purpose of this replication study, I included negative values in the analysis. Including these values ensures that the analysis captures the full spectrum of earnings management activities across different countries.

4. Another potential source of discrepancies between the original and replicated tables may be the choice of variables pulled from Worldscope. For example, in the `wrds_ws_funda table`, both `item1151` and `item4051` are labeled as "DEPRECIATION, DEPLETION AND AMORTIZATION" (Wharton Research Data Services 2024b). I chose `item1151` from the Income Statement rather than `item4051` from the Cash Flow Statement, based on the *Worldscope Balancing Model - Industrials* by Wharton Research Data Services (2024d). Since the authors do not specify the choice of variables used from the database, this could cause differences in the results.

5. The EM measures are based on scaled variables (e.g., operating cash flow scaled by lagged total assets). As such, the currency of the relevant data items should not affect the results as long

5

as the same currency is used consistently for both the numerator and the denominator. This approach ensures comparability across different countries, regardless of their local currencies. Additionally, according to a document by (Thomson Financial 2007, 20), all Worldscope data is consistently reported in the local currency of each firm's country of domicile, eliminating the need for currency conversions in this project.

By following the steps provided in Section 4 and adhering to the assumptions made, I successfully replicated the analysis and produced the required table. A thorough step-by-step approach, with each step clearly documented, helped to understand and verify the outputs.

4 Replication Steps

4.1 Step 1: Pulling the Data and Managing the Databases

Unlike Assignment I, where the data was provided externally, Assignment III involves pulling data directly from the Worldscope database, merging relevant tables, and preparing the data for further analysis from raw data to final output.

To ensure data integrity, I filtered out rows with empty `item6105` (Worldscope Permanent ID) values, as this identifier is crucial for firm/year-level filtering in the data preparation step. A total of 125 observations from the dynamic data and 10,306 observations from the static data were removed. The dynamic and static datasets were then merged on the `item6105` identifier, representing the unique Worldscope Permanent ID. According to Wharton Research Data Services (2024e), this identifier should be used consistently within Worldscope data, as it remains stable over time. An inner join was used for this merge to ensure that only complete and consistent data from both tables were retained.

Additionally, a filter was applied to retain only company rows. This was achieved by selecting rows where the `item6100` field equals 'C', indicating that the Worldscope Identifier represents a company. This step ensures that the analysis includes only company data, excluding averages, exchange rates, securities, or stock indices, as advised by Wharton Research Data Services (2024e).

In line with the methodology used by Leuz, Nanda, and Wysocki (2003), financial institutions were removed from the analysis based on their SIC codes. This was done by filtering out rows where the `item7021` identifier, representing the SIC code, falls within the range of 6000 to

6999. This ensures that the dataset focuses solely on non-financial companies.

Further filtering ensured that only data from the 31 countries specified in the paper were included. These steps were applied to reduce the dataset size and improve the workflow.

Moreover, the configuration file utilized additional refinements as proposed by Wharton Research Data Services (2024e). Specifically, using "A" (Annual) in the `freq` variable ensured that the data represents financial information reported on an annual basis, consistent with the methodology outlined in the paper by Leuz, Nanda, and Wysocki (2003). This excludes data reported on a current, daily, or quarterly basis.

After retaining only relevant company data, filtering out financial institutions, and focusing on the specified countries, the processed data was saved to a CSV file in the `data/pulled` directory, as specified in the configuration file `config/pull_data_cfg`.

4.2 Step 2: Data Preparation

To verify the pulled data, I checked the dataset for duplicate firm-year observations based on the combination of the Worldscope Permanent ID (`item6105`) and the fiscal year (`year_`). I confirmed that no duplicates were present, ensuring the accuracy of the data for further analysis. As the original study does not specify the net income measure, I used `item1651` (Net Income before Preferred Dividends) as the variable for net income. This choice aligns with the final net income figure reported in the income statement, based on *Worldscope Balancing Model - Industrials* by Wharton Research Data Services (2024d).

To prepare the data sample in line with the methodology outlined in the paper, it was essential to follow the requirements set by Leuz, Nanda, and Wysocki (2003). First, countries with sufficient firm-year observations were filtered. Each country needed at least 300 firm-year observations for key accounting variables, including total assets, sales, net income, and operating income. In this step, no countries were eliminated, as all countries met the requirement, which aligns with the paper's overview of countries (31 countries in total).

In the second filtration step, firms with adequate consecutive data were identified. Each firm had to have income statement and balance sheet information for at least three consecutive years, with all key accounting variables mentioned above present. If a firm had at least three consecutive years of complete data at any point, all its data entries were retained in the final

dataset. Therefore, only those countries and firms that met these criteria for all specified variables were retained in the dataset.

During the preparation step, 8,265 firms and 20,521 firm-year observations (all due to the second filtration step) were dropped, resulting in a final dataset with 18,040 non-financial firms and 123,469 firm-year observations. The differences in the numbers between the prepared dataset and the figures mentioned in the paper (70,955 firm-year observations and 8,616 non-financial firms) could be due to assumptions outlined in Section 3, such as variations in initial datasets, data updates, and filtering criteria. However, the original paper might have included additional data cleaning steps not explicitly mentioned, such as handling outliers, specific industry exclusions, or other criteria, which could affect the final counts.

Given that the number of observations in this project is significantly higher than in the original study, Table 1 from Leuz, Nanda, and Wysocki (2003) was partially replicated (focusing only on columns related to countries and firm-year observations) to highlight specific discrepancies that could arise for certain countries. The results in the table below show a higher number of firm-year observations for almost all countries compared to the paper by Leuz, Nanda, and Wysocki (2003), likely due to database updates leading to a more comprehensive dataset. The trend in the number of firm-year observations across countries is consistent with the original study, with larger capital markets like Japan, the United Kingdom, and the United States having the highest numbers. The distribution of observations in my dataset shows higher mean and median values compared to the paper by Leuz, Nanda, and Wysocki (2003), indicating a greater overall number of firm-year observations, while the minimum values are relatively close, suggesting that the overall distribution of observations across countries remains broadly similar.

However, there are significant discrepancies, particularly for the United States, which has 41,585 observations compared to 3,792 in the original study, possibly due to more comprehensive data or changes in dataset scope. Leuz, Nanda, and Wysocki (2003) focused only on S&P 500 companies for the U.S., while this analysis uses a more comprehensive dataset from the Worldscope database, which includes detailed financial statement data and profile data on public companies globally. This broader scope results in a higher number of firm-year observations for the United States. Despite the differences in absolute numbers, the relative trends across countries align with the paper.

[Table 1 about here.]

This final assignment has required the comprehensive application of all skills learned and feedback received from previous assignments, making it a more complex project workflow. In the future, this repository can be cloned or forked (if made public) to kickstart further projects on earnings management measures analysis. Thanks for reading!

References

Leuz, Christian, Dhananjay Nanda, and Peter D. Wysocki. 2003. "Earnings Management and Investor Protection: An International Comparison." *Journal of Financial Economics* 69 (3): 505–27. https://doi.org/10.1016/S0304-405X(03)00121-1.

Thomson Financial. 2007. "Worldscope Database: Datatype Definitions Guide." Issue 6. https://www.tilburguniversity.edu/sites/default/files/download/WorldScopeDatatypeDefinitionsGuide_2.pdf.

Wharton Research Data Services. 2024a. "Table: WRDS - Worldscope - Company Information (Wrds_ws_company)." https://wrds-www.wharton.upenn.edu/data-dictionary/tr_worldscope/wrds_ws_company/.

———. 2024b. "Table: WRDS - Worldscope - Fundamentals Annual (Wrds_ws_funda)." https://wrds-www.wharton.upenn.edu/data-dictionary/tr_worldscope/wrds_ws_funda/.

———. 2024c. "Thomson Reuters - Worldscope." https://wrds-www.wharton.upenn.edu/data-dictionary/tr_worldscope/.

———. 2024d. "Worldscope Balancing Model - Industrials." https://wrds-www.wharton.upenn.edu/documents/526/Worldscope_Balancing_Model_-_Industrials.xls.

———. 2024e. "WRDS Overview of Worldscope Data." https://wrds-www.wharton.upenn.edu/pages/support/manuals-and-overviews/thomson-reuters/worldscope/wrds-overview-worldscope-data/.

index	Country	# Firm-years
1	AUSTRALIA	2099
2	AUSTRIA	686
3	BELGIUM	932
4	CANADA	4215
5	DENMARK	1326
6	FINLAND	1058
7	FRANCE	5515
8	GERMANY	5307
9	GREECE	1034
10	HONG KONG	1862
11	INDIA	2068
12	INDONESIA	978
13	IRELAND	625
14	ITALY	1591
15	JAPAN	20528
16	KOREA (SOUTH)	2038
17	MALAYSIA	2283
18	NETHERLANDS	1782
19	NORWAY	1210
20	PAKISTAN	598
21	PHILIPPINES	574
22	PORTUGAL	577
23	SINGAPORE	1291
24	SOUTH AFRICA	1334
25	SPAIN	1221
26	SWEDEN	1766
27	SWITZERLAND	1553
28	TAIWAN	1259
29	THAILAND	1664
30	UNITED KINGDOM	12910
31	UNITED STATES	41585
32	Mean	3983
33	Median	1553
34	Min	574
35	Max	41585

Table 1: **Replicated table - Number of Firm-Year Observations per Country**