How Much New Information Is There in Earnings?
Corporate Decision-Making and Quantitative Analysis
Summer 2025 - Individual Report

Mel Mzv

May 4, 2025

Abstract

This project uses the TRR 266 Template for Reproducible Empirical Accounting Research (TREAT) to provide infrastructure for open science-oriented empirical projects. Leveraging data from the CRSP and Compustat databases, alongside Worldscope and Datastream via WRDS, this repository showcases a reproducible workflow integrating Python scripts for data preparation, analysis, and visualization. Integrating multiple databases adds complexity, requiring a detailed understanding of their structures and careful scripting to extract, align, and analyze data effectively. The project replicates and extends Ball and Shivakumar (2008) to analyze the informativeness of quarterly earnings announcements and their contribution to share price movements, highlighting their critical role in investment decisions and impact on investors and analysts. Key tasks include replicating and comparing original results, extending the analysis to 2007–2023, and applying the methodology to a non-U.S. country. The project also documents research design choices, discusses variations between original and reproduced results, and provides insights into earnings informativeness across different timeframes and jurisdictions. Additionally, it sketches a research design for a non-archival study to evaluate the paper's findings. This code base, adapted from TREAT, demonstrates how the template applies to this project and serves as a structured guide for reproducible empirical research in accounting.

Table of contents

List of Figures

List of Tables

1 List of Abbreviations

BHR: Buy-and-Hold Returns

CRSP: Center for Research in Security Prices

DS: Datastream

IDE: Integrated Development Environment

TREAT: TRR 266 Template for Reproducible Empirical Accounting Research

WRDS: Wharton Research Data Services

WSCP: Worldscope

2 Introduction

This project serves as a template-based example for conducting and documenting reproducible empirical accounting research. It leverages the TRR 266 Template for Reproducible Empirical Accounting Research (TREAT) and builds upon the methodological foundations of the Corporate Decision-Making and Quantitative Analysis course, as well as the Accounting Reading Group course, which explored archival and non-archival research methods. To support future empirical thesis projects, this template was adapted and further developed specifically for students at the School of Business and Economics at Humboldt-Universität zu Berlin—particularly those affiliated with the Institute of Accounting and Auditing and the Finance Group.

Following prior assignments on audit market concentration and financial reporting effects (e.g., Graham, Harvey, and Rajgopal (2005)), this project focuses on the informativeness of quarterly earnings announcements. Specifically, it replicates and extends the study by Ball and Shivakumar (2008), which examines whether earnings announcements contribute new information to markets or merely confirm what is already known—contributing to the long-standing discussion on market efficiency and the valuation versus contracting roles of accounting.

Using updated U.S. data and applying the methodology to a second country, this thesis combines multiple WRDS datasets and applies an event-study design to evaluate how much of the annual stock return variation can be explained by earnings announcement windows. The approach is designed to foster transparent research workflows, with a fully scripted pipeline for data pulling, preparation, analysis, and reporting.

The paper is organized into four main parts. In Section 3, I replicate the core findings of the original study using U.S. data from 1972 to 2006. In Section 4, I apply the same methodology to U.S. data from 2007 to 2023 to explore temporal robustness. In Section 5, I generalize the findings to a non-U.S. country to assess cross-country variation and data integration requirements. Finally, Section 6 sketches a non-archival research design using a survey-based approach to evaluate the underlying assumptions of the original paper. The thesis concludes with reflections and key insights in Section 7.

3 Task 1 - Replication of Key Tables and Figures

Task 1 replicates key tables and figures from Ball and Shivakumar (2008) using daily US stock return data (1972–2006) to establish baseline results. By reconstructing Table 1 (Panel A), Table 2, and Figure 1, it evaluates the robustness of the original findings and highlights any discrepancies. Daily CRSP data ensures accurate measurement of stock reactions around earnings announcements. Replication results are discussed in Section 3.3, with extended analysis in Section 4.

3.1 Research Design Choices and Assumptions

In line with Ball and Shivakumar (2008), this replication examines U.S. earnings announcements and stock returns from 1972 to 2006, following the original study's sample design, event window (-1 to +1), and return computations. Unlike the original, I omit analyst expectations, focusing solely on returns and announcement dates for simplicity.

To address gaps in methodological detail, several assumptions are applied. Event windows follow CRSP trading calendar conventions, shifting dates when earnings fall on non-trading days (Center for Research in Security Prices (CRSP) 2024). While Ball and Shivakumar (2008) require a minimum of 240 trading days, I relax this constraint, as per course guidance, resulting in a broader sample. All firms in Compustat meeting the quarterly announcement and return data criteria are included, regardless of activity status. I use PERMNO to uniquely track securities, avoiding aggregation across multiple share classes (Gow and Ding 2025). The CRSP/Compustat Merged (CCM) database is used to link identifiers, noting that differences from the original study may arise due to updated data (WRDS version: January 2025). As in Ball and Shivakumar (2008), the analysis is return-based (RET in CRSP), excluding bid-ask pricing. Earnings are grouped by announcement date (RDQ), not fiscal period, to align directly with CRSP trading dates.

These assumptions, together with the procedural details in Section 3.2, guide the replication of Ball and Shivakumar (2008) and ensure transparency in design and implementation.

3.2 Replication Steps

This section outlines the structured replication process: data pulling, preparation, and analysis, aligned with Ball and Shivakumar (2008).

Step 1: Pulling the Data and Managing the Databases

Data from WRDS combines daily CRSP stock returns (dsf) and Compustat quarterly earnings announcements (fundq), linked via the CRSP-Compustat link table (ccmxpf_lnkhist). The `permno` and `gvkey` identifiers ensure accurate firm-year alignment. Key variables are `ret` (returns) and `rdq` (announcement dates), pulled for the full 1972–2023 period used in Tasks 1 and 2.

Step 2: Data Preparation

The merged dataset includes CRSP, Compustat, and the link table. After removing irrelevant linking variables, the dataset is trimmed to key variables: `ret`, `rdq`, `fyr`, `permno`, and `gvkey`. Only firms with four earnings announcements per calendar year are retained. Firms are also classified by fiscal year-end (`fyr`) to match Table 2 Panels C–D. Duplicate or incomplete firm-years are excluded.

Step 3: Analysis Implementation and Reproduction of Tables and Figure

BHRs are computed using the cumulative return formula per Gundersen (2022). Cross-sectional regressions are then run for each firm-year to replicate Table 1, Table 2, and Figure 1 from the original paper.

3.3 Results

This section presents the replication results.

4 Task 2 - Extending the Analysis to 2007–2023

This section extends the analysis to 2007–2023, a period shaped by events like the 2008 financial crisis and the COVID-19 pandemic. Although the programming involves only a sample period update, the task requires thoughtful economic interpretation of how structural shifts and evolving market conditions may have affected the informativeness of earnings announcements.

4.1 Research Design Choices and Assumptions

Assumptions from Section 3.1 still hold.

4.2 Replication Steps

The pull, preparation, and analysis step would remain the same as designed in Section 3.2 with the only change in data range.

4.3 Results

Shifting the sample to 2007–2023 introduces potential changes in earnings announcement informativeness driven by key economic and regulatory events. The 2008 financial crisis and the 2020 COVID-19 pandemic likely heightened market uncertainty, increasing reliance on earnings disclosures to assess firm stability. Moreover, Regulation (EU) No 537/2014 may have enhanced financial reporting transparency and audit quality in EU countries, potentially affecting how earnings news is reflected in stock prices and contributing to cross-country variation in informativeness.

5 Task 3 - Cross-Country Replication

Task 3 generalizes the analysis to a non-U.S. market by replicating Figure 1 using 1972–2023 data from Canada. This involves aligning CRSP/Compustat variables with Worldscope and Datastream equivalents while accounting for differences in reporting standards, market liquidity, and institutional context. Though the core methodology is retained, adapting it to international data demands careful mapping and interpretation.

5.1 Research Design Choices and Assumptions

Since Task 3 applies a generalization and extension approach to a non-U.S. market, it requires adapting the methodology to alternative databases and market structures. Like in the original study, no analyst forecasts or earnings surprises are used—only stock returns and earnings announcement dates are required. Canada was selected for its comparable quarterly reporting frequency and institutional similarity to the U.S., making it a suitable choice for cross-country replication (Short 2025). Canadian firms reporting under U.S. GAAP further enhance comparability.

To address database-specific issues not covered in Section 3.1, I make the following adjustments: First, Worldscope/Datastream workflows are kept separate from CRSP/Compustat to prevent interference and facilitate debugging, given the scale of data involved. Second, I exclude

fiscal-year-end-based subgroup analysis (Table 2 Panels C–D), as Figure 1 uses the full sample without splitting by fiscal period. Third, I use Worldscope items 5901–5904 to define earnings dates, though limited availability before 1992 may reduce the sample size (Thomson Financial 2007, 281). Fourth, differences in update frequency between quarterly Worldscope and weekly Datastream can lead to timing mismatches when aligning fundamentals and returns. Fifth, I rely on Datastream's `ret` variable for percentage returns, ignoring bid/ask spreads. Finally, currency effects are ignored, as percentage returns are unaffected by currency denomination.

5.2 Replication Steps

The process is same as in Section 3.

Step 1: Pulling the Data and Managing the Databases

Following Dai and Drechsler (2021), I link Worldscope and Datastream using code and infocode. Since Worldscope's year_ starts in 1980 (Wharton Research Data Services 2025), Datastream data is also restricted to 1980–2023 to ensure consistent coverage. I extract quarterly earnings announcement dates (items 5901–5904) from Worldscope and daily stock returns from Datastream, focusing on Canadian firms and filtering for common equity (typecode = EQ). This step yields three datasets: 27.9M (Datastream), 77.9K (link table), and 1.94M (Worldscope) observations.

Step 2: Data Preparation

Worldscope is first merged with the link table via code, producing 747,760 rows. This is then joined with Datastream via infocode, resulting in 10.2M observations. I exclude firms missing any earnings date fields, removing 3.6M rows, and drop 6M rows where earnings do not span all four quarters within the same calendar year. To retain more valid windows, event days (-1, 0, 1) are dynamically shifted within ±3 days when needed. After cleaning, I compute 3-day BHRs and extract full-year stock return data from Datastream. All retained infocode entries have around 250 trading days per year, confirming data completeness for annual BHR computation.

Step 3: Analysis Implementation and Reproduction of Tables and Figure

The analysis replicates Table 1 (summary statistics), Table 2 (yearly regressions), and Figure 1 (Abnormal R^2 and slope trends). Using `BHR_3day` and `BHR_Annual`, I run yearly regressions of annual returns on earnings-window returns and compute Abnormal R^2 (adjusted R^2 – 4.8%), handling missing values appropriately. The resulting trends are plotted and saved as both PNG

and pickle files for use in the paper.

5.3 Results

Comparing the replication results with those from Ball and Shivakumar (2008) reveals both similarities and discrepancies. The original figure shows a general upward trend in abnormal R^2 , peaking in the early 2000s, whereas the replicated figure captures greater volatility in the later years (2010–2023). Additionally, Ball's figure suggests a smoother long-term dynamic, while the replication exhibits more pronounced fluctuations, particularly post-2010. Panel B in replicated figure exhibits higher volatility and larger coefficient magnitudes, suggesting increased earnings informativeness post-2006, while Ball's original figure has more stable and lower-magnitude coefficients.

The post-2000 period in the replication shows more inconsistent abnormal R^2 values, possibly reflecting evolving market structures, shifts in disclosure practices, and increased earnings informativeness following regulatory changes (e.g., IFRS adoption, post-SOX adjustments, different financial market structure). The pronounced fluctuations could also be driven by global financial crisis, Canada's higher market concentration, fewer publicly traded firms, and potential regulatory differences, which may affect earnings informativeness and stock return patterns. The variability in slope coefficients further suggests that earnings announcements' impact on stock returns is less stable than in prior decades, potentially due to differences in investor response.

6 Task 4 - Research Design for a Non-Archival Study

To extend the analysis beyond archival methods, this section proposes a survey to assess whether investors perceive earnings announcements as informative or merely confirmatory, following the triangulation principle in empirical research (Bloomfield, Nelson, and Soltes 2016). While Ball and Shivakumar (2008) measure market reactions via stock returns, surveys can capture investor beliefs directly, offering complementary evidence on earnings informativeness.

A structured survey targeting 300 institutional and retail investors will be distributed through financial networks and platforms (e.g., LinkedIn, CFA societies), with a pilot study refining its clarity. Participants will receive a €20 honorarium, and responses will remain anonymous. The

survey measures perceived informativeness (DV) using Likert-scale and multiple-choice questions, with IVs capturing factors such as investor preparedness and reaction timing.

Follow-up interviews with selected respondents will add qualitative depth, enabling a richer understanding of investor perspectives. This mixed-method approach enhances validity and supports generalization of findings beyond price-based evidence.

6.1 Survey Questions

1. **How do you typically prepare for earnings announcements in your investment decisions?**
   - I conduct in-depth research and adjust my positions in advance
   - I monitor but rarely adjust positions pre-announcement
   - I rely on market consensus, analyst forecasts, and AI-driven insights
   - I do not make investment decisions based on earnings announcements
2. **How do you typically react to earnings announcements?**
   - I adjust my investment strategy immediately based on the announcement.
   - I wait for further analysis before making changes.
   - I rarely make investment decisions based on earnings announcements.
3. **To what extent do you believe that earnings announcements provide new information beyond what is already reflected in market prices?**
   - Always
   - Most of the time
   - Sometimes
   - Rarely
   - Never
4. **Rank the following factors in order of importance when evaluating earnings announcements (1 = most important, 5 = least important):**
   - Earnings per share (EPS) compared to analyst forecasts
   - Revenue growth
   - Management guidance and commentary
   - Market reaction on the day of the announcement

- Industry trends and macroeconomic conditions

5. **Do you consider earnings announcements to be more confirmatory or informative?**
   - Primarily confirmatory (reinforce existing expectations)
   - Primarily informative (provide new insights)
   - A mix of both

6.2 Follow-Up Interview Questions

1. Can you describe a recent instance where an earnings announcement significantly influenced your investment decision?
2. In your experience, are there specific industries where earnings announcements are more informative than confirmatory?
3. Do you use earnings announcements differently depending on market conditions (e.g., economic downturn vs. growth periods)?

7 Conclusion

This thesis template demonstrates how the TRR 266 framework can be applied to structure a reproducible and transparent empirical accounting study. Using Quarto proved to be a particularly effective and user-friendly solution for structuring and rendering the thesis. It facilitates seamless integration of code, results, and narrative, making it an ideal tool for students at the Institute of Accounting and Auditing or the Finance Group at the School of Business and Economics, HU Berlin.

Beyond replication, this project integrates programming, data processing, and visualization skills, offering a foundation for future extensions—whether through archival analysis or survey-based designs. As an open and modular template, this repository is designed to support ongoing student and academic work in empirical research. Thanks for reading!

References

Ball, Ray, and Lakshmanan Shivakumar. 2008. "How Much New Information Is There in Earnings?" *Journal of Accounting Research* 46 (5): 975–1016. https://doi.org/10.1111/j.1475-679X.2008.00299.x.

Bloomfield, Robert, Mark W. Nelson, and Eugene Soltes. 2016. "Gathering Data for Archival, Field, Survey, and Experimental Accounting Research." *Journal of Accounting Research* 54 (2): 341–95. https://doi.org/10.1111/1475-679X.12104.

Center for Research in Security Prices (CRSP). 2024. "CRSP US Stock & Index Databases - Data Descriptions Guide." Graduate School of Business, University of Chicago. https://www.crsp.org/wp-content/uploads/guides/CRSP_US_Stock_&_Indexes_Database_Data_Descriptions_Guide.pdf.

Dai, Rui, and Qingyi (Freda) Song Drechsler. 2021. "Research Guide for Datastream and Worldscope at Wharton Research Data Services." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3843551.

Gow, Ian D., and Tongqing Ding. 2025. *Empirical Research in Accounting: Tools and Methods*. Chapman and Hall/CRC Series on Statistics in Business and Economics. Boca Raton, FL: Taylor; Francis. https://iangow.github.io/far_book/.

Graham, John R., Campbell R. Harvey, and Shiva Rajgopal. 2005. "The Economic Implications of Corporate Financial Reporting." *Journal of Accounting and Economics* 40 (1): 3–73. https://doi.org/https://doi.org/10.1016/j.jacceco.2005.01.002.

Gundersen, Gregory. 2022. "Returns and Log Returns." https://gregorygundersen.com/blog/2022/02/06/log-returns/.

Short, Christine. 2025. "The Short List - Reporting Frequency Isn't the Only Difference with How Earnings Results Are Shared Worldwide." *Wall Street Horizon*. https://www.wallstreethorizon.com/blog/Reporting-Frequency-Isnt-the-Only-Difference-with-How-Earnings-Results-Are-Shared-Worldwide.

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. 2025. "Worldscope - Stock Data." *WRDS*. https://wrds-www.wharton.upenn.edu/pages/get-data/thomson-reuters/worldscope/stock-data/.