How Much New Information Is There in Earnings? Evidence from Earnings Releases in the U.S. and Canada

Master's thesis

for acquiring the degree of
Master of Science (M.Sc.)

in Business Administration

at the School of Business and Economics
of Humboldt-Universität zu Berlin

submitted by
Max Mustermann
Student no. xxxxxx

First Examiner: « name of first examiner »
Second Examiner: « name of second examiner »

Berlin, May 20, 2025

Table of contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my professors, colleagues, and friends for their support and insights throughout the preparation of this thesis. Special thanks go to [Name], whose guidance was invaluable during the project.

Abstract

This thesis template explores global development patterns using the publicly available Gapminder dataset, with a specific focus on the year 2007. Leveraging a reproducible workflow adapted from the TRR 266 Template, the project presents descriptive statistics by continent and a visual analysis of the relationship between life expectancy, income per capita, and population. The findings reveal clear disparities in health and wealth outcomes across world regions. A comparison with Gapminder's 2021 global development chart highlights persistent global inequalities, while illustrating notable progress in certain regions. The analysis serves as a modern, transparent example of public data storytelling in empirical research.

List of Abbreviations

GDP Gross Domestic Product

IDE Integrated Development Environment

TREAT TRR 266 Template for Reproducible Empirical Accounting Research

1 Introduction

This project serves as a template-based example for conducting and documenting reproducible empirical data analysis using publicly available datasets. 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, as well as the Accounting Reading Group course. To support future empirical thesis projects, this template was adapted and 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. The template was developed in adherence to the formal formatting and content guidelines of the School of Business and Economics, and in particular, those of the Institute of Accounting and Auditing. It is not a completed thesis but a demonstrative project hosted in the template repository.

This template project explores global development patterns using the Gapminder dataset, focusing on the year 2007. The project illustrates how reproducible research workflows can be applied beyond traditional financial datasets by using public and open-source data. Specifically, it presents a summary statistics table by continent and visualizes the relationship between GDP per capita and life expectancy, offering a static snapshot of world development during 2007. These descriptive results are presented in Section 2.

The results are further contrasted with Gapminder's interactive global development visualizations as of 2021 (Gapminder Foundation 2021), highlighting the potential and limits of static data analysis. This comparative discussion is provided in Section 3. The project concludes with reflections on transparency, reproducibility, and the importance of making research accessible to a broader audience, as discussed in Section 4.

2 Replication of Key Tables and Figures

This section replicates key tables and figures from (**ball_how_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 2.3, with extended analysis in Section 3.

2.1 Research Design Choices and Assumptions

In line with (**ball_how_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** While (**ball_how_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. 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_how_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 2.2, guide the replication of (**ball_how_2008?**) and ensure transparency in design and implementation.

2.2 Replication Steps

This section outlines the structured replication process: data pulling, preparation, and analysis, aligned with (**ball_how_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_returns_20**
Cross-sectional regressions are then run for each firm-year to replicate Table 1, Table 2, and Figure 1 from the original paper.

2.3 Results

This section compares the static 2007 snapshot with Gapminder's dynamic global development chart Figure 1.

[Figure 1 about here.]

3 Cross-Country Replication

This section 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.

3.1 Research Design Choices and Assumptions

Since this secition 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_short_2025?**). Canadian firms reporting under U.S. GAAP further enhance comparability.

To address database-specific issues not covered in Section 2.1, 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_worldscope_2007?**). 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.

3.2 Replication Steps

The process is same as in Section 2.

Step 1: Pulling the Data and Managing the Databases
Following (**dai_research_2021?**), I link Worldscope and Datastream using code and infocode. Since Worldscope's `year_` starts in 1980 (**wharton_research_data_services_worldscope_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.

Specifically, buy-and-hold return over the three-day event window, that measures the stock's reaction only to earnings news, is computed as shown in Equation 1:

$$BHR_{\text{event}} = (1 + R_{t_1})(1 + R_{t_2}) \dots (1 + R_{t_T}) - 1, \tag{1}$$
where R_t is the daily return and T is the total trading days in a year.

3.3 Results

Comparing the replication results with those from (**ball_how_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.

4 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. Thanks for reading!

References

Gapminder Foundation. 2021. "World Health Chart." https://www.gapminder.org/fw/world-health-chart/.

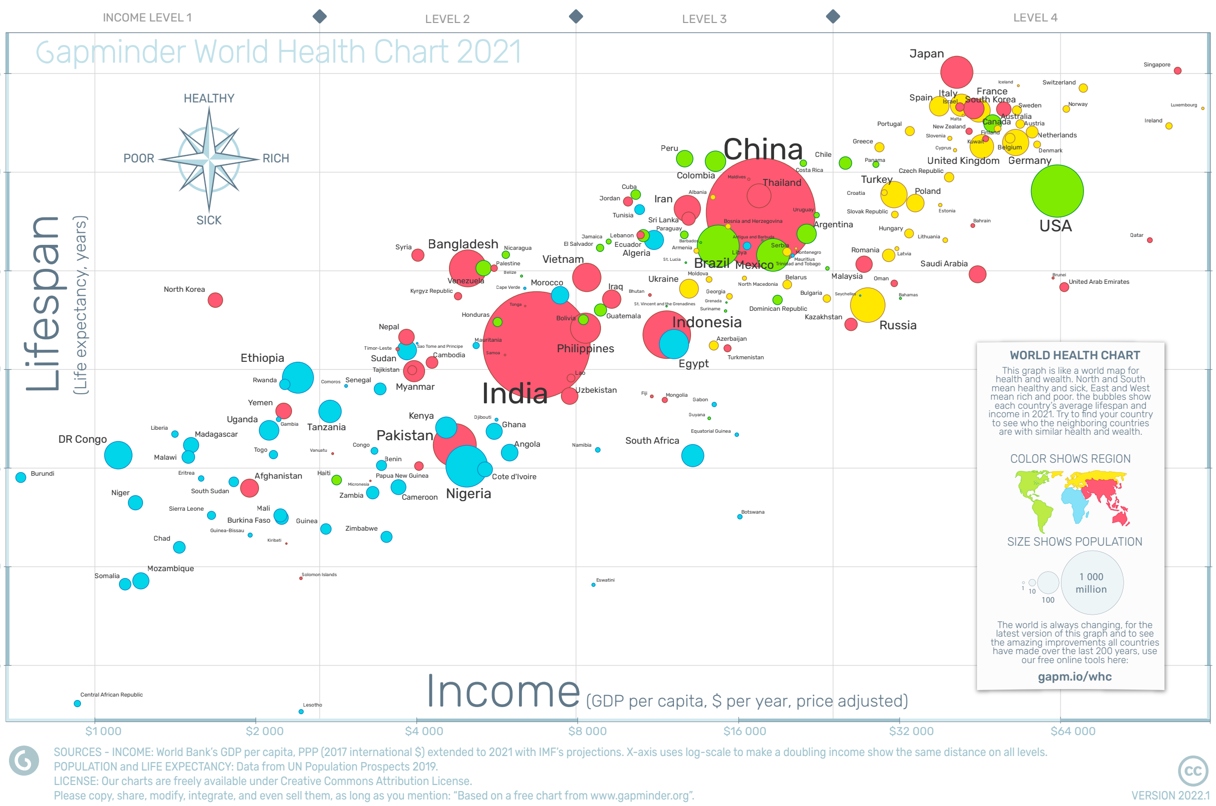

Figure 1: Original World Health Chart 2021 from Gapminder Foundation

Declaration of Academic Honesty

I, Max Mustermann, hereby declare that I have not previously submitted the present work for other examinations. I wrote this work independently. All sources, including sources from the Internet, that I have reproduced in either an unaltered or modified form (particularly sources for texts, graphs, tables and images), have been acknowledged by me as such.

I understand that violations of these principles will result in proceedings regarding deception or attempted deception.

Max Mustermann

Berlin, May 20, 2025