Investigating the Lagged Impact of Sectoral GDP Growth on Stock Price Performance

Gregory M Pina

Western Governors University

Table of Contents

A. Project Highlights

- **Research Question or Organizational Need**
    - This research explores whether quarterly GDP fluctuations in the Tech, Energy, Materials, and Industrial sectors lead or lag corresponding changes in stock prices in subsequent quarters—a question of particular importance to investment firms, financial analysts, and corporate strategists. By examining if official sector-level GDP releases significantly correlate with next-quarter stock returns, the study aims to provide timely insights that can help organizations optimize portfolio strategies, enhance risk management, and improve forecasts of sector performance.

- **Scope of the Project:**
    - The project entails analyzing a decade-long time series (approximately 2010 to 2024/2025) for four key sectors—Technology, Energy, Materials, and Industrials—using public data sources from the BEA (sector-level GDP) and Yahoo Finance (daily stock/ETF prices). It excludes explicitly additional macroeconomic factors, such as interest rates and inflation, as well as global or non-U.S. GDP data.

- **Overview of the Solution:**
    - This study examined how quarterly GDP changes relate to subsequent stock returns in Technology, Energy, Materials, and Industrials sectors (2010-2025). Daily stock prices were aggregated into quarterly averages using BEA GDP and Yahoo Finance data to match GDP reporting periods. Analysis through EDA, correlation assessment, and regression modeling found moderate positive correlations for Technology and Industrials sectors. However, regression results showed statistically insignificant relationships, indicating GDP fluctuations alone have limited predictive value for stock returns. Future research should incorporate additional economic factors for better investment insights.

- **Tools and Methodologies:**
  - **Data Analytics Tools and Techniques:**
    - The project leveraged Python-based analytics, including Pandas for data wrangling, Matplotlib and Seaborn for visualization (box plots, line charts), and Statsmodels for correlation and OLS regression analysis, to examine the relationship between quarterly sector-level GDP fluctuations and subsequent stock returns.
  - **Data Source:**
    - Utilized sector-level GDP data from the Bureau of Economic Analysis (BEA), stock and ETF historical price data from Yahoo Finance, and supplementary ticker-level stock market data from Kaggle datasets.
  - **Programming Language:**
    - Python (Jupyter Notebooks), incorporating Pandas, Statsmodels, Matplotlib, Seaborn, Requests, yfinance, and Kaggle API.
  - **Methodology:**
    - Conducted time-series data preprocessing, merging sector GDP with stock price data by quarter-end dates, calculated correlation coefficients, and performed Ordinary Least Squares (OLS) regression analyses to test relationships between GDP growth and subsequent quarterly stock returns.

B. Project Execution

The execution closely matched the initial project plan, though some minor adjustments were necessary, primarily due to external factors related to data availability.

- **Project Plan:**
  - The execution followed the initial plan closely. Minor variances occurred due to issues accessing Kaggle datasets, necessitating manual data downloads as a workaround.

- **Project Planning Methodology:**
  - An iterative, agile-style approach was consistently employed, effectively accommodating challenges such as technical difficulties in accessing Kaggle data through automated scripts.

- **Project Timeline and Milestones:**
  - The project is significantly ahead of schedule according to the original timeline, with slight delays in the data-gathering phase offset by faster-than-anticipated analysis, resulting in overall adherence to the planned project schedule.

Milestone	Start Date	End Date	Duration
1) Data Collection	Feb 18, 2025	Feb 22, 2025	4 days
2) Data Cleaning & Preparation	Feb 23, 2025	Feb 27, 2025	4 days
3) Exploratory Data Analysis (EDA)	Mar 3, 2025	Mar 6, 2025	3 days
4) Modeling (Correlation & Regression)	Mar 6, 2025	Mar 10, 2025	4 days
5) Evaluation & Visualization	Mar 10, 2025	Mar 12, 2025	2 days
6) Final Report & Presentation	Mar 12, 2025	Mar 13, 2025	1 days

C. Data Collection Process

- **Data Selection and Collection:**
    - The data selection and collection process primarily aligned with the initial plan. However, due to unforeseen technical issues accessing Kaggle via automated methods, manual downloads replaced some originally intended automated retrieval processes.

- **Handling Data Quality and Completeness:**
    - The main obstacle encountered was the intermittent inability to access Kaggle datasets through the Kaggle API. To mitigate this, data was downloaded manually and integrated into the project workflow, enabling continued progress with minimal disruption.

- **Data Governance, Privacy, and Security:**
    - No unplanned data governance issues arose during this project. All utilized data sources were publicly available and required no additional compliance actions or special permissions beyond standard attribution.

1. **Advantages and Limitations of Data Set**

- **Advantages**:
    - **Comprehensive Metrics:**
        - The datasets provided detailed economic and financial metrics, such as quarterly GDP by sector (BEA) and extensive historical stock market data (Yahoo Finance and Kaggle), allowing for thorough sector-specific analyses.

    - **Relevance to Research Question:**
        - The selected datasets closely aligned with the core research question, facilitating a detailed examination of correlations between quarterly

sector GDP and subsequent stock performance.

- **Suitability for Predictive Modeling:**
  - With their extensive temporal coverage and rich granularity, the datasets effectively supported the predictive modeling process, enabling accurate forecasting and insightful trend analyses.

Limitations:

- **Data Cleaning and Preprocessing Required:**
  - Substantial data cleaning was necessary, especially for addressing missing and inconsistent stock price entries within the Kaggle dataset, which added complexity and extended data preparation time.

- **Handling Mixed Data Types:**
  - Multiple data formats, including Excel (.xlsx) for BEA data and CSV for stock prices, required additional preprocessing efforts, complicating the integration process and demanding careful transformations.

- **Focus on Engineered Features:**
  - The need for extensive feature engineering—such as manually calculating quarterly returns and GDP growth lags—required meticulous preprocessing and validation, potentially introducing minor delays and complexity into the analytical workflow.

The selected datasets provided relevant sector-specific economic and financial data suitable for addressing the research question. However, adjustments were necessary, including converting BEA's quarterly GDP data from string formats (e.g., '2020QI') into accurate quarter-end timestamps and restructuring datasets through data frame melting to facilitate meaningful analysis. Additionally, extensive data cleaning was required due to missing stock records in the

Kaggle dataset, posing a challenge but ultimately enabling effective data integration and predictive modeling.

D. Data Extraction and Preparation

The project's analytical foundation required precise extraction and cleaning of GDP and stock data to manage inconsistencies and align datasets. Essential transformations included assigning sectors, converting quarter-end dates, and calculating returns, which enhanced predictive relevance. Python-based libraries efficiently streamlined these tasks, ensuring accurate alignment for robust correlation and predictive modeling analyses.

- **Data Extraction:**
  - GDP data was extracted via API calls from the Bureau of Economic Analysis (BEA) GDP-by-Industry dataset.
  - ETF composition data was retrieved through direct downloads of Excel files from State Street Global Advisors (SSGA).
  - Historical stock prices were sourced from the Kaggle financial dataset and Yahoo Finance API.

- **Data Cleaning and Transformation:**
  - Cleaned GDP data involved melting wide-format BEA data into long format and converting quarter strings to accurate end-of-quarter timestamps.
  - Handling missing values and inconsistent records in Kaggle's stock data was addressed via filtering, aggregation, and date alignment.

- **Feature Engineering:**
  - Creation of lagged GDP variables to analyze predictive relationships.
  - Calculated percentage returns and sector-wise GDP aggregates to facilitate direct comparison between economic indicators and stock performance.

- **Tools and Techniques:**
    - Python with various libraries: P*andas*, *NumPy*, *Matplotlib*, *Seaborn*, *requests*, *yfinance*, *statsmodels*, and Excel management with *Openpyxl*.
    - Data wrangling, merging, and time-series processing techniques leveraging pandas' powerful data frame manipulation features.

The extraction and preparation processes effectively aligned the GDP and stock datasets, enabling robust analyses such as correlation testing and linear regression modeling to explore economic indicators' predictive strength on stock returns across sectors.

E. Data Analysis Process

E.1 Data Analysis Methods

- **Ordinary Least Squares (OLS) Regression Analysis:**
    - OLS regression was employed to quantify the linear relationship between sector-level GDP growth and corresponding stock returns. This method helps identify whether variations in GDP can predict future stock returns, supporting the hypothesis testing and predictive modeling goals.

- **Correlation Analysis:**
    - Pearson correlation analysis was applied to initially explore the strength and direction of linear associations between sector GDP and stock performance, guiding variable selection for subsequent modeling.

- **Data Cleaning and Preprocessing:**
    - The data required substantial preprocessing steps, including handling missing values, merging datasets by quarter-end timestamps, and transforming wide-format GDP data into a long format suitable for analysis. These processes were critical to aligning and standardizing data from multiple sources to support accurate statistical modeling.

E.2 Advantages and Limitations of Analytical Tools

- **Pandas**
  - *Advantage:*
    - Provides intuitive, efficient data manipulation, reshaping, and integration capabilities ideal for handling financial and economic datasets.
  - *Limitation:*
    - Large-scale data transformations can become computationally expensive, requiring careful optimization for performance.
- **Statsmodels (OLS Regression):**
  - *Advantage:*
    - Delivers interpretable regression outputs, including significance tests (*p-values*) and goodness-of-fit metrics (*R-squared*), vital for statistical inference.
  - *Limitation:*
    - Assumes linearity and homoscedasticity in data relationships, potentially oversimplifying complex real-world interactions.
- **Seaborn and Matplotlib (Visualization):**
  - *Advantage:*
    - Provide clear, insightful visualization of trends and correlations to aid exploratory analysis.
  - *Limitation:*
    - Visualizations alone do not quantify causality and can mislead interpretation if underlying data assumptions are violated.

E.3 Application of Analytics Methods:

- **OLS Regression Analysis:**
  - **Feature Engineering and Selection:**
    - Selected and created lagged GDP growth features to assess predictive capability on future stock returns. Sector labels and quarterly timestamps were engineered to align GDP data with stock performance data for analysis.

  - **Preprocessing and Data Cleaning:**
    - Cleaned and merged data from BEA, Yahoo Finance, and Kaggle datasets. Transformed BEA data from wide to long format, standardizing quarterly dates for alignment with stock returns.

  - **Model Implementation (insert model here):**
    - Developed Ordinary Least Squares (OLS) regression models for each sector individually:

$$Stock\ Return_t = \beta_0 + \beta_1\,(GDP\ Growth_{t-1}) + \epsilon$$

  - **Model Validation and Testing:**
    - Assessed model accuracy and predictive performance using *R-squared* values and statistical significance (*p-values*). Analyzed regression outputs to interpret relationships.

  - **Assumption and Requirement Verification:**
    - Validated assumptions of linearity, homoscedasticity, normality, and independence of residuals through residual plots, Durbin-Watson tests, and Jarque-Bera statistics. Adjustments were made where necessary to ensure model validity.

F Data Analysis Results

F.1 Statistical Significance

- **Statistical Test Summary:**
  - **Null Hypothesis (H_0):**
    - No statistically significant relationship exists between sector-level GDP growth in one quarter and the subsequent quarter's stock returns within each sector.
  - **Statistical Test:**
    - Ordinary Least Squares (OLS) Regression.
  - **Test Metrics:**
    - ***R-squared* values**: Technology (0.004), Industrials (0.005), Materials (0.000), Energy (0.001)
    - ***t-statistics* and *p-values*** from regression results for GDP_growth_lag1 coefficients:
      - **Technology:** t = -1.485, p = 0.138
      - **Industrials:** t = -1.570, p = 0.117
      - **Materials:** t = 0.471, p = 0.638
      - **Energy:** t = -0.618, p = 0.537
    - **Alpha:** 0.05 (standard threshold for statistical significance).
  - **Conclusion:**
    - Based on the *p-values* (> 0.05), the null hypothesis cannot be rejected for any sector analyzed. Thus, the analysis suggests no statistically significant predictive relationship exists between lagged sector-level GDP growth and next-quarter stock returns at the 5% significance level.

- **Model Analysis:**
    - **Model:**
        - Supervised regression model predicting quarterly stock returns from previous-quarter GDP growth.

    - **Algorithm:**
        - Ordinary Least Squares (OLS) regression analysis.

    - **Performance:**
        - *R-squared* values ranged from negligible (0.000–0.005), indicating low explanatory power.
        - *P-values* were high (all >0.05), demonstrating non-significant predictors.

    - **Benchmark:**
        - A minimum *R-squared* value of 0.25 and statistically significant *p-value* ($\alpha = 0.05$) as outlined in Task 2.

    - **Conclusion:**
        - The analysis did not achieve the defined performance benchmark, showing that GDP growth did not significantly predict next-quarter stock returns.

- **Reasoning Behind Choices:**
    - **OLS Regression:**
        - Chosen to quantify the linear relationship clearly and interpret coefficients.

- **Pearson analysis:**
  - Selected to quantify linear correlation strength directly, useful for preliminary relationship insights.

- **Metrics:**
  - **Accuracy:**
    - Accuracy was assessed using *R-squared*, evaluating how well GDP growth explained variations in stock returns, and *p-values*, to measure the statistical significance of GDP as a predictor. Both metrics indicated low predictive accuracy and nonsignificant relationships, suggesting limited explanatory power of GDP for quarterly stock returns.

F.2 Practical Significance

The analysis's practical significance lies in its ability to guide strategic investment decisions based on the relationship between sector GDP performance and stock returns. Although statistical relationships were generally weak and not statistically significant, the results still provide meaningful insights into the complex dynamics between economic indicators and market performance. Clients can use these findings to enhance decision-making related to portfolio diversification, sector-based investment strategies, and risk assessment.

- **Findings and Their Implications:**
  - **Limited GDP Impact on Stock Returns:**
    - Weak statistical relationships suggest that sector-level GDP alone cannot predict quarterly stock returns reliably, prompting investors to integrate additional market indicators.

- **Sector Variability:**
    - Stronger correlations in Technology and Industrials highlight potential areas for deeper economic analysis, aiding sector-specific investment strategies.

- **Enhanced Risk Management:**
    - Recognition of weak predictive signals helps refine risk models, encouraging caution and diversification rather than overreliance on economic indicators alone.

- **Application for the Client:**
    - **Sector-focused Portfolio Optimization:**
        - Investment managers can refine portfolio allocations by leveraging identified correlations between GDP indicators and sector returns.

    - **Informed Risk Assessment:**
        - Analysts can utilize the insights on GDP and return relationships to improve risk management strategies, especially in sectors sensitive to economic fluctuations.

    - **Strategic Forecasting:**
        - The analysis provides strategic foresight, allowing corporate strategists to anticipate sector performance trends, thus enhancing market positioning and investment timing.

F.3 Overall Success

- **Statistical Outcomes:**
    - The analysis revealed moderate correlations between quarterly GDP changes and sector stock returns, notably stronger in Technology (0.59) and Industrials (0.62), indicating sector-dependent sensitivity to economic indicators. Regression results showed limited statistical significance, suggesting that GDP alone isn't a strong standalone return predictor.

- **Real world Relevance:**
    - Findings provide practical insights for investment banks aiming to refine sector-specific portfolio strategies. Understanding sector responsiveness to GDP changes enhances forecasting accuracy and strategic investment timing.

- **Meeting Success Criteria:**
    - **Data Integrity:**
        - Ensured through comprehensive data preprocessing, cleaning, and rigorous verification procedures, providing reliable datasets for modeling.

    - **Statistical Significance:**
        - Although significance levels were limited, analysis offered nuanced insights into sector sensitivity, meeting analytical exploration objectives.

    - **Documentation:**
        - Robustly documented data processes, methods, and codebase ensured project transparency and reproducibility.

- **Conclusion:**
    - The project successfully provided practical insights and validated analytical processes. Despite modest statistical significance, its findings offer valuable guidance for strategic sector-focused decision-making in investment banking contexts.

G. Conclusion

G.1 Summary of Conclusions

The study analyzed the relationship between quarterly sector GDP and subsequent stock returns, revealing moderate correlations, especially in the Technology and Industrials sectors. However, regression analyses showed minimal statistical significance, suggesting GDP alone is insufficient for reliably predicting sector returns.

- **Statistical Insights:**
    - Moderate correlations between GDP growth and stock returns were identified, strongest in technology (r=0.59) and industrials (r=0.62).
    - OLS regression indicated weak predictive capability (low R^2 values), implying GDP is a limited predictor in isolation.

- **Real-world Impact:**
    - Provides actionable insights for sector-based investment strategy, enhancing market-timing decisions.
    - Reinforces the importance of a multi-factor analytical approach over reliance on a single economic indicator.

- **Data Approach:**
    - Data from BEA and Kaggle was systematically cleaned, standardized, and integrated, ensuring compatibility across sources. Feature engineering, particularly converting quarterly GDP data into usable timestamps, enhanced

model interpretability. Despite extensive preprocessing, findings suggest incorporating broader economic factors would improve predictive performance.

G.2 Effective Storytelling:

- **Rationale:**
  - The chosen visualizations illustrate sector-specific GDP and stock return relationships, clearly highlighting correlations and limitations through intuitive scatter and line plots. Regression plots visually communicated the lack of strong linear predictive relationships, while time-series visualizations effectively conveyed historical trends.
- **Tools and Graphical Representation:**
  - **Correlation Heatmap:**
    - **Description:**
      - Color-coded matrix showing correlations between variables (GDP, Close, Volume, Dividends, Weight).
    - **Purpose:**
      - Quickly assesses the strength and direction of relationships among key variables.
    - **Tool:**
      - Seaborn
  - **Scatterplots (GDP vs. Close Price):**
    - **Description:**
      - Scatter plots displaying GDP against stock prices by sector, including regression lines.

- **Purpose:**
  - Visually explores linear relationships, complementing regression analysis.
- **Tool:**
  - Seaborn *regplot* and Matplotlib.

- **Boxplots (Closing Prices):**
  - **Description:**
    - Summarize distribution, median, variability, and outliers of sector-specific stock prices.
  - **Purpose:**
    - Allows rapid sector comparison and identification of outlier stocks.
  - **Tool:**
    - Matplotlib and Seaborn

- **Time Series and Dual-Axis Line Charts (GDP vs. Price Trends):**
  - **Description:**
    - Visualize quarterly trends in sector-level GDP and corresponding average stock prices on dual axes.
  - **Purpose:**

    Provides intuitive visual insights into sector performance trends over time, facilitating stakeholder understanding of temporal patterns.

        - **Tool:**
            - Matplotlib

G.3 Recommended Courses of Action:

- **Sector-Specific Investment Strategies:**
    - **Rationale:**
        - The analysis indicated sector-specific differences in the relationship between GDP changes and stock returns. Industries like Technology and Industrials exhibited higher correlation values than Materials and Energy, suggesting differentiated sensitivity to economic indicators. Investors should therefore consider customizing strategies for each sector, emphasizing those with higher correlation.
    - **How it Relates:**
        - A focused strategy leveraging insights from the Industrials and Technology sectors, given their stronger correlations, could enhance portfolio returns by timing sector-specific investments more effectively based on economic signals.

- **Enhanced Predictive Modeling:**
    - **Rationale:**
        - The weak predictive power of GDP alone (low R^2 values across sectors) suggests the need for multivariate models that integrate additional economic indicators such as inflation, employment, or interest rates, potentially increasing prediction accuracy.
    - **How it relates:**
        - Incorporating these broader macroeconomic factors can improve forecasting capabilities, offering investment managers better predictive accuracy, refined risk management strategies, and improved responsiveness to market conditions.

H Panopto Presentation

References

- Andrei, D., Mann, W., & Peyer, U. (2020). Macroeconomic announcements and stock market reactions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3640322

- Caporale, G. M., Spagnolo, F., & Spagnolo, N. (2021). Sectoral stock prices and GDP: International evidence from wavelet analysis. *Journal of International Financial Markets, Institutions & Money, 75*, Article 101457. https://doi.org/10.1016/j.intfin.2021.101457

- Choi, S., & Dou, W. (2018, October). *Market efficiency and news release lags: Empirical evidence from GDP data* [Conference presentation]. Financial Management Association Annual Meeting, San Diego, CA, United States.

- Kaggle. (n.d.). *9000 tickers of stock market data (entire history)* [Dataset]. Retrieved from https://www.kaggle.com/datasets/jakewright/9000-tickers-of-stock-market-data-full-history

- State Street Global Advisors. (n.d.). *ETF composition data*. Retrieved from https://www.ssga.com/us/en/intermediary/etfs/

- U.S. Bureau of Economic Analysis. (n.d.). *GDP by industry*. Retrieved from https://www.bea.gov/data/gdp/gdp-industry

- Yahoo Finance. (n.d.). Historical stock prices. Retrieved from https://finance.yahoo.com