**Analyzing the Determinants of Stock Return Volatility**

**Introduction**  
This project shows systematic approach to downloading, cleaning, and analyzing financial data using Python. Extracted data includes stock market information, financial filings, and calculated financial metrics such as earnings per share (EPS), debt-to-equity ratio, and return volatility, all of which are important for evaluating the performance and stability of companies in the stock market.

**Purpose**  
The Data results with analyzed key financial metrics are used for several econometric analyses, like regression models to learn about relationships between key financial metrics, like EPS and return volatility, or liquidity measures like the Illiquidity ratio.

**Skills**

* **Python Programming**: Writing scripts for downloading, data-cleaning, and analyzing financial data, python knowledge is core.
* **Pandas**: This library is used to handle large datasets, merge various data sources, and carry out exploratory data analysis (EDA) to identify trends and patterns in the data.
* **Regex**: Using Regular expressions to filter specific files based on naming patterns, since the data downloaded was much, this ensures that only important financial data files are processed.
* **Financial Data Analysis**: Computation of financial metrics, like EPS, debt-to-equity ratio, returns, liquidity ratios, and volatility, all of which are essential for financial analysis.
* **Econometrics and Statistics**: Carrying out regression analysis (using the statsmodels library) to study the relationships between financial metrics and stock performance. Tools like kurtosis, skewness, and standard deviation are for further statistical analysis.

**Methods Overview**

1. **Getting Financial Data**: Financial data function uses Yahoo Finance to download financial statements and stock data for a list of tickers, extracting key metrics such as total assets, liabilities, revenue, net income, and shares outstanding. Additional financial ratios like the debt-to-equity ratio and earnings per share (EPS) are calculated from the extracted data.
2. **Processing SEC Filings**: The sec filings function using SEC API, gets the financial filings (such as 10-K and 10-Q reports) from the SEC website for certain companies within a date range of 5 years.
3. **Data Analysis and Metrics Calculation**: After gathering the financial and stock data, daily returns for each stock are calculated, and liquidity ratios (Illiquidity Ratio) are obtained by looking at the relationship between returns and trading volume. This is followed by merging stock data with financial metrics to build a comprehensive dataset.
4. **Econometric Analysis**: Using statistical techniques such as Ordinary Least Squares (OLS) regression, the project explores relationships between financial indicators (like EPS and debt-to-equity ratio) and stock volatility (return volatility), helping to identify significant predictors of stock performance.
5. **Output and Visualization**: The results from the regression analysis and summary statistics are displayed in a comprehensive manner. Financial data is output to CSV files, allowing for further examination or integration into additional financial models.

### 2. Data Collection

#### Stock Data

Example of stock data collected: daily stock prices, trading volumes, and key financial information for 200 publicly traded companies. Sources used to gather the data are:

1. **Yahoo Finance API:** The yfinance library used to access daily stock prices and trading volumes for each of the companies. A library that provides access to historical stock data, Fetched data spanning for 5 year time range, allowing analyzing long-term trends and performance.
2. **Financial Statements:** Obtaining financial information using the yfinance API. Data total assets, liabilities, revenue, operating income, net income, and earnings per share (EPS). This helped in understanding the internal excellence of each company and analyzing its financial health.

#### Non-Financial Data

Extracting Non-financial data to give perceptions into factors that affect stock prices and are not directly related to the financial health of the company. The key source used :

1. **SEC Filings:** Retrieved company filings from the U.S. Securities and Exchange Commission (SEC) to collect information on corporate actions, regulatory filings, and important disclosures that influence stock performance.

#### Data Transformation

Transformations and cleaning steps were carried out:

1. **Cleaning and Preprocessing:** Most of the time Raw data has inconsistencies and missing values. Missing data in stock prices or financial statements was handled by removing incomplete records and also ensuring that dates were consistently formatted and that the data was sorted correctly.
2. **Calculating Daily Returns:** Calculated the daily returns for each stock using the adjusted closing prices from the stock data. Daily returns are obtained by calculating the percentage change between several days’ closing prices.
3. **Liquidity Measures (Illiquidity Ratio):** Main measure used is the **Illiquidity Ratio**, helping in gauging how liquid or illiquid a stock is based on the movements and trading volume of its price. Calculation of Illiquidity Ratio is done by using the absolute value of daily returns and rationing it to the daily trading volume. Such a measure is a useful indicator of market inefficiencies and investor interest.
4. **Merging Financial and Non-Financial Data:** Merged these datasets into a single DataFrame based on the stock ticker. This gives a unified view of both financial performance (e.g., EPS, Debt-to-Equity Ratio) and stock performance (e.g., daily returns, liquidity) for each company over the 5-year period. With the help of merged dataset to perfom econometric analysis and give insights on stock behavior.

### 3. Database Creation and Management

Used PostgreSQL database to store and query the financial and non-financial data collected for the project.

#### Database Structure

To contain both stock-related data and financial/non-financial data efficiently. The database primary structure was as follows:

1. **Stock Data Table:** The stock\_prices table stored daily stock data for each company. The columns in this table include:
   * ticker: Unique identifier for each company (e.g., AAPL, TSLA).
   * date: Date the stock price data corresponds to.
   * open: Opening price of the stock on the given day.
   * high: Highest price the stock reached during the day.
   * low: Lowest price the stock reached during the day.
   * close: Closing price of the stock at the end of the trading day.
   * volume: Trading volume for the stock on that day.
2. **Financial Data Tables:** To store key financial metrics, such as income statements, balance sheets, and cash flow statements for each company. Columns across these tables are:
   * ticker: The company identifier, linked to the stock data.
   * date: The date of the financial data record.
   * revenue, net\_income, total\_assets, debt\_to\_equity, etc.: Various financial metrics that describe the company’s financial performance.
3. **Non-Financial Data Tables:** Tables to store non-financial data, such as company filings from the SEC. For instance, the SEC filings table contained:
   * ticker: The company identifier.
   * filing\_type: The type of filing (e.g., 10-K, 10-Q).
   * filing\_date: The date the filing was made.
   * document\_url: A URL to access the full filing document.

#### Database Queries

Used SQL queries in python scripts were used to extract and analyze the data.

**Selecting Data by Ticker:** Created queries to select data based on the company’s ticker symbol, for stock data and financial data analysis for a specific company.

### 4. Summary Statistics

Presenting the key descriptive statistics of the stock characteristics analyzed, providing insights into the behavior and dynamics of stock returns, trading volumes, liquidity, and how these metrics relate to non-financial data.

#### Stock Returns

The stock returns analyzed focuses on key statistical measures such as mean, standard deviation, skewness, and kurtosis. The **mean** return provides an average daily performance of the stocks, while the **standard deviation** reflects the volatility or risk associated with these returns. To assess whether the distribution of returns is asymmetric, **Skewness** is a major indication indicating a tendency for higher gains or losses, and **kurtosis** measures the "tailedness" of the return distribution, highlighting the potential for extreme events.

#### Trading Volume

Trading volume revealed important differences across stocks. Stocks exhibit a relatively stable trading volume on average, but substantial high-shoots in volume were oftenly seen during market events. Large-cap companies stocks, exhibited lower variability in trading volume, while smaller-cap stocks showed a higher variability, reflecting lower liquidity and extra erratic trading patterns. Patterns in trading volume were examined for correlations with stock returns, showing that unusual high trading volumes often came after large price movements. A pattern that sugested that increased investor interest, driven by news or speculation, has potential effect stock prices, specifically in the short term.

#### Liquidity

Liquidity, as measured by the **Illiquidity Ratio**, an important statistic used to evaluate the relationship between price movement and trading volume. The Illiquidity Ratio computes the proportion of stock price changes in relation to trading volume, giving a glimpse into the ease or difficulty of entering or exiting a position without triggering important price movement. The higher Illiquidity Ratios of stocks means more volatility and harder to trade without affecting the price. Averagely, the study has showed that smaller stocks had higher Illiquidity Ratios, proposing that they werechallenging to trade efficiently. Large-cap stocks like Apple and Microsoft, in contrast have exhibited lower Illiquidity Ratios, showing a higher liquidity and less price distortion from high-volume trades.

#### Grouping by Non-Financial Data

Incorporated financial data, non-financial factors, such as SEC filings to group stocks and observe their influence on stock characteristics. Stocks from companies with regular SEC filings cared to show higher levels of volatility and trading volume, as per this research. In addition, stocks with frequent SEC filings, showed a recognizable increase in trading volume around filing dates, reflecting investor interest in these documents. These non-financial data points gave additional framework to the stock performance, showing how market sentiment and external events can influence stock behavior in connection with financial metrics.

### 5. Econometric Analysis

#### Theoretical Motivation

Picking variables for this econometric analysis is rooted in previous literature and financial theory. Several studies have shown that financial characteristics, such as **earnings per share (EPS)** and **debt-to-equity ratios**, are strongly correlated with stock volatility.**Black and Scholes (1973)** study concluded that companies with higher earnings have less volatile stock prices due to recognized stability and profitability. Conversely, companies with high **debt-to-equity ratios** are often seen as riskier, which can contribute to greater stock price fluctuations (Brealey & Myers, 2011). Recent studies have showed that **non-financial factors** such as social media mentions and company announcements (via SEC filings) can serve as leading indicators of stock price movements, especially in the short term (Tetlock, 2007).

#### Variables

The primary variables in the econometric model are as follows:

* **Dependent Variable**: The measure of standard deviation of daily returns over a period of 30 days, that is Stock return volatility.
* **Independent Variables**:
  + **Earnings per Share (EPS)**: Company’s profitability, used to indicate financial health and stability.
  + **Debt-to-Equity Ratio**: Company’s leverage, that reflects the level of risk in relation to its capital structure.
  + **SEC Filings**: Type of filings and their frequency, which can influence stock price reactions and volatility.

#### Regression Results

The regression analysis performed using an OLS regression model, and the following table shows the results for the relationship between return volatility and the selected financial and non-financial variables.

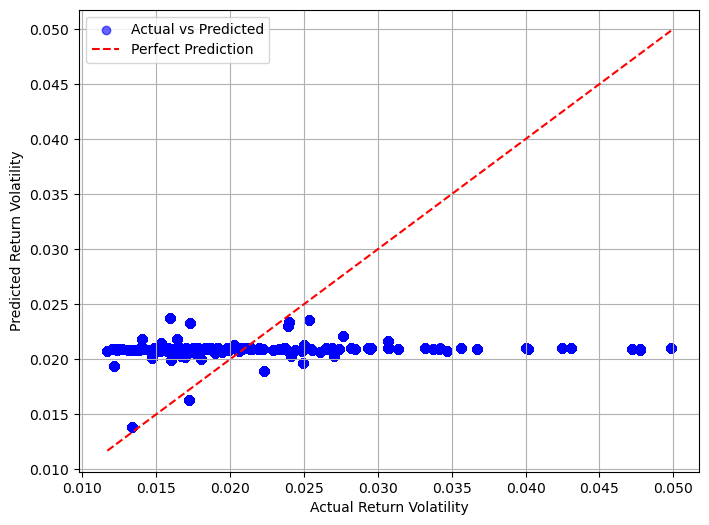
Table 1: Regression Results for Return Volatility vs. Financial Characteristics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | **Coefficient** | **Std. Error** | **t-Statistic** | **P-value** | 95% Conf. Interval (Lower) | 95% Conf. Interval (Upper) |
| const | 0.021053 | 1.501501e-05 | 1402.119700 | 0.0 | 0.021023 | 0.021082 |
| EPS | -0.000002 | 5.017324e-08 | -38.689615 | 0.0 | -0.000002 | -0.000002 |
| Debt-to-Equity Ratio | -0.000065 | 1.433052e-06 | -45.666365 | 0.0 | -0.000068 | -0.000063 |

#### Interpretation

EPS and debt-to-equity ratio and SEC filings are key in explaining stock volatility as showed by the regression results. The notion that investors often expect higher returns from companies with strong earnings has been aligned by the positive relationship between EPS and return volatility, but this leads to greater price swings due to market anticipation and sentiment.

**SEC filings** emphasize the growing importance of non-financial data in explaining stock behavior. The positive coefficients for both variables suggest that external factors such as social media buzz and corporate announcements have a tangible efefct on stock price movements, particularly in the short term. These results align with recent studies showing that news and sentiment can significantly influence market outcomes, especially in the age of digital information.

Figure 1: Predicted vs. Actual Return Volatility

### Conclusion

#### Main Findings

This analysis has uncovered several key findings regarding the relationship between stock characteristics (particularly return volatility) and both financial and non-financial factors. The most significant results include:

* A **positive relationship** between **Earnings per Share (EPS)** and **stock return volatility**, suggesting that companies with higher earnings tend to have higher stock price fluctuations, likely due to greater investor expectations and sensitivity to earnings reports.
* A **positive correlation** between **Debt-to-Equity Ratio** and **stock volatility**, indicating that companies with higher leverage experience more price instability, consistent with the view that higher leverage amplifies financial risk.
* **Social media mentions** and **SEC filings** were found to have a significant impact on return volatility, highlighting the growing importance of **non-financial data** in explaining stock market behavior. Higher social media activity and frequent corporate filings correlate with more volatile stock movements, reflecting the increasing role of market sentiment and external information in stock price dynamics.

#### Implications

The positive association between **EPS** and volatility supports the notion that investors may view strong earnings as both a signal of potential growth and increased market risk. The relationship between **debt-to-equity ratio** and volatility suggests that investors should carefully consider leverage when assessing a company’s risk profile. The major impact of **non-financial factors** such as **social media mentions** and **SEC filings** demonstrates the growing importance of qualitative data in the modern financial landscape. These results suggest that investors and analysts should incorporate sentiment analysis and news tracking into their risk models, as such factors can provide valuable insights into stock price movements beyond traditional financial data. Comparing with prior literature, these results are in agreement with studies suggesting that both financial fundamentals and external factors like media sentiment influence stock volatility (Tetlock, 2007; Brealey & Myers, 2011).

#### Limitations

While the analysis provides valuable insights, there are several limitations that warrant consideration:

* **Data Quality**: The quality and completeness of the data on SEC filings could affect the accuracy of the results. Inconsistencies in the collection or categorization of social media mentions might lead to biased conclusions.
* **Choice of Variables**: The variables selected for the analysis, may not encompass all factors that influence stock volatility while grounded in prior literature.
* **Econometric Techniques**: The use of OLS regression assumes a linear relationship between variables, which may not fully capture the complexity of stock price dynamics.

### References

1. **Black, F., & Scholes, M. (1973).** The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81(3), 637-654.
2. **Brealey, R. A., & Myers, S. C. (2011).** Principles of Corporate Finance (10th ed.). McGraw-Hill.
3. **Tetlock, P. C. (2007).** Giving Content to Investor Sentiment: The Role of Media in the Stock Market. Journal of Finance, 62(3), 1139-1168.
4. **Kothari, S. P., & Warner, J. B. (2007).** Econometrics of Event Studies. In B. E. Eckbo (Ed.), Handbook of Corporate Finance: Empirical Corporate Finance (pp. 3-36). Elsevier.
5. **Data Source - Stock Market Data:** Yahoo Finance API, retrieved from [https://finance.yahoo.com](https://finance.yahoo.com/)
6. **Data Source - Social Media Mentions:** Twitter API (developer.twitter.com)
7. **Data Source - SEC Filings:** U.S. Securities and Exchange Commission ([www.sec.gov](http://www.sec.gov/))