AALBORG UNIVERSITET

Impact of Forecast Dispersion as a Measure of Information Asymmetry on Stock Volatility Post-Earnings Announcements

An Empirical Study Over Three Years on 75 S&P 500 Stocks

Raymond Fennjan - Study No: 20165427 –
Supervisor – Frederik Steen Lundtofte
AALBORG UNIVERSITY BUSINESS SCHOOL

Contents

Abstract

This study will investigate the influence of information asymmetry on stock volatility around earnings announcements. There is a great amount of literature on the longer effects of earnings reports and how the market response slowly or quickly to the information. This kind of literature is generally categorized as Post-Earnings Announcement Drift (PEAD). But the study of the immediate effect of earnings reports is not seen commonly in financial literature and research. Given the small amount of research in this area, this study tries to fil the gap on this missing literature. This research tries to use historical prices of 75 S&P 500 stocks over the last three years and all their 900 earnings announcements and calculate it as idiosyncratic abnormal volatility, this abnormal volatility will be used against information asymmetry which is measured through analyst forecast dispersion. The use of forecast dispersion as a measure of information asymmetry is due to limited access to other data observation that are normally used, like bid-ask spread, volatility spreads in option trading, insider trading information etc. This study will use Ordinary Least Squares (OLS) regression to analyze the relationship between information asymmetry and abnormal stock volatility during a 20-day window around earnings announcement. The study also tries to analyze the immediate effect post earnings on day 0 and1. The findings of this study are made to improve the understanding of how the differences in information availability for analysts and investors have an impact on stock price reactions and could give insights that would help short-term trading strategies. This study not only helps the investors trading strategies but offers an extension to the literature of Post-Earnings Announcement Drift (PEAD).

Introduction

The literature on financial markets has its limitations and studies on the earnings announcements immediate influence on investor behavior and stock prices is limited. This study tries to analyze the relationship between the information asymmetry and its impact on stock volatility, but also includes other variables that may also have an impact on stock volatility. There have been many studies in financial literature about the Post-Earnings Announcement Drift (PEAD), where the stock market has been giving abnormal returns after earnings announcement. The PEAD studies have shown that the market can continue to its trend after earnings announcements, this trend can last for weeks and months. This drift happens because the market doesn't immediately incorporate the news and information into the stock market right after the earnings announcement. This can be because of information asymmetry where all the investors don't have access to information at the same time or limited access. The PEAD studies show that the market does not fully incorporate the information into the stock prices right after earnings and therefore the market is not efficient. It takes time for the stock prices to fully incorporate the earnings announcement information. This can give investors an investment opportunity before the market fully incorporates information into the stock prices (Shaun Bond, 2023). This is especially the case for companies with less institutional investors and less attention from financial analysts, where information asymmetry is higher. Literature shows that the effect of PEAD is different for different types of companies. Analysts that have less institutional investors or financial analyst tend to have more PEAD, this is more significant in smaller, less liquid, and more volatile stocks (Josef Fink, 2021). The PEAD seems to be less for larger, more liquid companies.

Literature and research on the options markets can also have an importance in the understanding of this topic. An increase in the activity of the option markets has been linked to market reactions, this indicates that option markets can give a predictions or early indication of how the market would react the earnings announcements (Lei Qin, 2020). Implied volatility is the market's expectations of how much the stock price will move in the future, it also affects the price of the option. When the implied volatility increases it often suggests that the investors expected high volatility on the stock market. There is a relationship between the stock prices and the option market before earnings announcements. This relationship could show the strategies made by investors and give some indications of asymmetrical distribution. This could be due to some investors who possess information that is not available for others, or they got the information faster, it could also be insider trading. This does not mean the study will cover the option market, this is difficult due to limited access to option data. But this could be evidence, reason, and confirmation for studying information asymmetry. Despite the studies that can be found on PEAD, there are very limited studies on the impacts of information asymmetry on stock volatility around earnings announcements, in a shorter observation period, and its immediate effects of the earnings announcement right after its release. It could be interesting to investigate if there are any variables that could affect or predict the volatility right after the release of earnings, and if there is any abnormality that could be found here. This could be beneficial for investing strategies and short-term investing, especially in the option market, there could be some strategic possibilities. Because of the limitations on bid and ask spreads on the stock market, this study tries to employ the analyst forecast dispersion as a measure for information asymmetry and analyzes the effects on stock volatility in a narrower perspective. This study would not only contribute to the understanding of how information asymmetry affects market behavior but also help the literature on this topic. This study investigates the immediate effect on earnings releases and tries to understand how variations in information accessibility among different investors can lead to volatile market behaviors. Doing analysis by using an OLS regression model, this study will provide insights into the effects of information asymmetry, which is a missing topic in the literature on market reaction to earnings announcements.

Literature Review

The efficient market hypothesis suggests that stock prices fully reflect all available information, which means that stocks are traded at their fair value. The efficient market hypothesis is in three forms, weak, semi-strong and strong, each is based on how much information incorporated into the market prices (Eugene F. Fama, 1970). Information asymmetry challenges the semi-strong and strong form of efficient market hypothesis, because of the difference in access to or understanding of information which can delay its incorporation in stock prices, leading to market inefficiencies. Studies have shown that information asymmetry can lead to increased volatility as the investors react to the slow availability or understanding of new information. For example, Lee, Mucklow, and Ready (Charles M. C. Lee, n.d.) write about how asymmetric information before block transactions leads to increased price volatility because of uncertainty with less informed investors. Easley, Hvidkjaer, and O'hara (David Easley, 2002) finds that periods of high trading activity, that consist of private information where some traders have information that is not public, are associated with high volatility. These findings suggest that information asymmetry could be an evident and significant reason for volatility on the stock market. Regulations that try to reduce information asymmetry, like Sarbanes- Oxley Act in the U.S, can impact market efficiency by making the corporations information more available for the public. Studies by Zhang (Jin E. Zhang. Jinghong Shu. Menachem Brenner, n.d.) suggest that regulations like this can reduce volatility by lowering information asymmetry and supporting stronger market efficiency, but the effectiveness of these regulations varies by market and firm size. Behavioral finance gives importance to the EMH by suggesting that markets are not always efficient due to irrational behaviors and psychological bias. Hirshleifer (David Hirshleifer, n.d.) describes how psychological biases can lead to stock prices that deviate from their true values for periods, thereby creating market inefficiencies that are caused by asymmetric information. In the past years there has been more and more algorithmic trading and there has been found a new behavior in the relationship between information asymmetry and market efficiency. Brogaard, Hendershott, and Riordan (Jonathan Brogaard, n.d.) have made findings on how high frequency trading, that are made of algorithms, can process information faster than human traders, which can impact market efficiency. Their findings suggest that while these algorithms can reduce the time it takes for prices to incorporate new information, they may also lead to increased intra-day volatility. This literature highlights the significance of information asymmetry in influencing stock market volatility. The Kyle model (Albert S. Kyle, 1985) is important for understanding how private information is incorporated into market prices through informed traders. In this model, an informed trader, also known as an insider, has private information about the intrinsic value of an asset, which is not yet known of the market. Trades on this information lead to prices movements, the trades there are being done because of this information, are steadily incorporating the information into the market. When other investors see these price movements they also trade on these movements. So, the asymmetric information where the informed traders have the advantage, the market will over time adjust because of uninformed investors follows the trend. But this still gives evidence of the information asymmetry in the market (Albert S. Kyle, 1985). The Glosten-Milgrom model helps to show how the markets adjust prices when there is information asymmetry in the market. The market maker, which is sets the buying price and selling price, changes these prices if they think that the traders know something the rest of the market don't know. This gives a higher difference between the prices, which is also called bid-ask spread, which are bigger if the market maker things there are investors who have better information than others. This bigger spread compensates the market maker for the risk of trading with someone that for example has insider information. The model shows that this information asymmetry can lead to larger bid-ask spreads which thereby leads to more volatile markets. A larger spread can indicate unequal information on the stocks true value (Lawrence R. Glosten, n.d.). Post-Earnings Announcement Drift (PEAD) refers to stock prices continuing to move in the same direction as the earnings surprises for several weeks and even months after the announcement. This is due to evidence of information asymmetry, where the information may not reach everyone on the market at the same time. Some investors may have the information immediate and other investors may take longer to reach this information. This can cause a slow adjustment of the stock price or overreaction when some investors see the information finally incorporated into the market. There can also be psychological factors that influence the market. Investors may have overconfidence or are more risk averse and this can also affect how they react to news. When all the investors receive the information at the same time, but react differently, this can also cause a later price adjustment in the market. The PEAD challenges the strong form of the Efficient Market Hypothesis, where the PEAD shows evidence of a semi-efficient market because of the markets longer period to adjust the information into the stock prices. Stocks with lower liquidity and volume after earnings announcements can cause a PEAD. Because of the lower number of trades there is for lower volume and liquidity stocks, the trades can change the stock price more significantly, but at the same time, the stocks that has less traders in the market can give a slower adjustment for the stock price. The institutional investors, like pension or hedge funds, can react differently to earning announcements. If these big institutions need more time to decide or analyze the announcement report, it can cause a drift in the stock, because they have so much influence on the stock prices movements. (Josef Fink, 2021)(Shaun Bond, 2023). Hong and Stein discuss a model that analyze how different types of agents only have limited access to information from public news, this can cause the agents to underreact to new information and overreact to price changes when the information finally incorporate to the market, which are more seen in markets with high information asymmetry. Smaller firms with fewer institutional investors and those who have fewer financial analyst are prone to stronger PEAD effects due to higher levels of information asymmetry and investors don't fully covers and follow these smaller firms (Hong & Stein, 1999). Companies with large positive earnings surprises experience significant positive abnormal returns. The PEAD anomaly is strong in smaller, less liquid stocks that are followed by less skillful investors (Bernard and Thomas, 1989). The option market can be used for prediction of volatility around earnings announcements. The market reacts quickly to changes in investors' expectations about a company. The implied volatility, which is an indicator of the option market, shows how volatile investors expected the market to be. So, if the implied volatility is increasing for the upcoming earnings announcement, this could indicate that the investors expected an earnings report that would give a significant reaction. This increases the demand for options and pushes the price on options, so higher prices mean higher implied volatility. The difference in implied volatility between options that expire before and after the earnings announcement can give some indications of how the market could react to some events. A higher difference could mean that the investors are uncertain about what or which way the price will go but still expect a large volatility. The options market can thereby reveal those investors who have information that is not available for others. There are several studies in this area which discuss how the option market can reveal insider trading, when there is a sudden increase in the implied volatility. This can indicate that some investors now something others don't, this is typically when there is a merger or acquisition (Lei Qin, 2020). Stocks that have high trading volumes in options before the earnings announcements are more likely to adjust the price correctly post earnings announcements. This is because of pre-earnings announcements under reaction, which are caused by information asymmetry in the option market (Patrick Augustin a b, n.d.). As mentioned before, implied volatility can give expectations of future stock price movements around earnings announcements. When there is an increase in implied skewness and kurtosis in the implied volatility, it could imply that the investors are expecting extreme price movements. Skewness in this case is an indication of the investor's expectations of which way the stock price will move. Kurtosis is an indication of the probability of extreme price changes (Dean Diavatopoulos a, n.d.). Studies have shown that when options markets are more liquid, the stock market tends to be less liquid, and the predictability of returns is higher for stock with greater information asymmetry and higher liquidity ratios (Cameron Truong, n.d.). This could be due to more investment in the option market versus the stock market. The reason why the predictably is higher when the stock has higher information asymmetry is because those who have better information can make better investment decisions and thereby take advantage of this information asymmetry. The research question for this study is not fully related to this literature review but it is still important to understand the broader studies that have been made in this area on information asymmetry and volatility. The literature explains information asymmetry, stock market volatility and market efficiency. It explains how different models, like the kyle and Closten-Milgrom models, can explain private information's influence on market prices and how it can lead to market inefficiencies. Additionally, empirical studies on topics like Post-Earnings Announcements Drift and the behavior of option markets before earnings announcements shows how asymmetric information can lead to market reactions and abnormal returns, especially in less liquid and smaller stocks. Regulatory measures like the Sarbanes-Oxley Act aim to reduce the effects by making information more available.

Information asymmetry

For the measurement of information asymmetry, the analyst forecast dispersion is used as an indicator and a variable. The forecast dispersion is calculated as the difference between the highest and lowest analyst forecast on Earnings Per Share. This difference shows the uncertainty and information asymmetry withing the market. The idea of using the analyst forecast dispersion as a measure for asymmetry is because of the variations in the expectations of future earnings. Studies have implied that significant differences in EPS predictions are correlated with the amount of access to information among the analysts (Jeffery S. Abarbanell, n.d.). A high level of dispersion would suggest a high level of uncertainty about future earnings. Studies have used different methods to measure information asymmetry in the market. The bid-ask spread of stock prices and the implied volatility spread in options trading are the two most normal measures. The bid-ask spread demonstrates the levels of uncertainty of the private information between traders, where a greater spread indicates higher transaction costs and more information asymmetry (Lei Qin, 2020). The studies made for volatility spreads (implied volatilities for put and call options) discuss how movements in these spreads can predict earnings announcements and can be an indication for asymmetrical information for some investors. There are some investors that might have some information that others don’t, and it can be detected in the option market. Increased implied volatility spreads can indicate that some investors have an information advantage, this can be seen in the position they take in option market pre-earnings announcements (Lei Qin, 2020). While these measures are effective to use, they have limited public access.

Abnormal stock volatility

The choice of abnormal stock volatility as the dependent variable in this study is important because of its sensitivity to new information and its ability to mirror the markets immediate reactions to earnings announcements. Abnormal volatility excesses the general market volatility by isolating the fluctuations that are directly linked to specific event dates and excluding broader effects of market trends. There have been used a couple of methods to isolate the abnormal volatility. These methods calculate the expected normal performance based on historical volatility and a measurement of deviations from this normal performance during the earnings announcements. This approach makes sure that the observed volatility is directly related to the earnings announcement. Other measurements like trading volume or beta can also be used as measurement for market behavior, but abnormal volatility is a better choice for immediate reactions on stock prices. In this study, 'abnormal volatility' is defined as idiosyncratic or firm-specific volatility, which captures the response of a company's stock price to its earnings announcements, excluding overall market movements. Idiosyncratic volatility is important to our analysis as it allows us to investigate how effects of an earnings report are understood differently due to different levels of information asymmetry between investors, and directly affect the stock's price. This kind of volatility shows the intrinsic behavior of a stock in reaction to new, firm-specific information that is not influenced by the broader economic or market conditions. In this study idiosyncratic volatility is calculated by adjusting the stock returns for market movements using the be coefficient from a 40-days control period. This beta shows the stock sensitivity to market changes and is used to abnormal returns to make sure that the volatility is only measured by company by company-specific events, instead of overall market movements. This makes the effects rely more on how the earnings announcement impacts on the stocks without the effect of other market factors.

Controlled variables

To have a more robust analysis, this study includes other variables which are used as control variables in the regression. The first controlled variable is the 'Number of estimates', a higher number of estimates can possibly lead to better information, a greater number typically means that the analyst and investors are well informed, and the information is well reflected on the stock market. This can reduce the difference between investors with more and less information (Kevin K. Li and Haifeng You, n.d.). 'Total debt to equity ratio' is used as a variable that measures the company's leverage, this can influence the risk the investors a willing to take before investing. Companies with a higher leverage can be seen as riskier and affect the overall volatility. 'Institutional Ownership' measures the percentage of the shares owned by institutions. A higher percentage can possibly be associated with more stable and stable stock prices. Institutions have more resources to analyze the companies before making the investment, also when these large institutions have a significant amount of ownership in these companies, they tend to be more active in the management and strategical operations. They have enough influence to be included in some decision making (Elyas Elyasiani, 2007). 'Insider stakeholders', this includes the number of shares owned by insiders like executives and directors. Insiders having shares in their own company can signal confidence in the company's futures which can affect how others view the coming earnings announcements. The insiders trading activity can give some trustworthiness to the company's fundamentals and future. When insiders have a high number of shares it can give investors more confidence about the company. A high percentage of insider stock owners can stabilize the stock prices and give a more adjusted response to earnings announcements (R. Shruti, n.d.). The 'Market capitalization' variable is used as a measurement for size of the companies. Large companies are associated with more stable stock prices, and they may have a different reaction to information than smaller companies (Chongyu Dang a, n.d.).

Research Question

How does forecast dispersion as a measure for information asymmetry, affect abnormal stock volatility post earnings announcements?

Objectives:

1. Reviewing existing literature on information asymmetry and volatility and their relationship with the stock market.
2. To examine the influence of forecast dispersion (information asymmetry) and other variables on abnormal stock volatility within a 20-day window.
3. To examine the immediate effect of earnings announcement on abnormal volatility and if information asymmetry or other variables influence this.
4. To analyze the effectiveness of existing models in predicting effects of information asymmetry on stock volatility and identify improvements.

Hypotheses

Based on the literature review, there has been made some hypotheses, such as:

1. H1: High information asymmetry increases volatility – There is a positive correlation between analyst forecast dispersion and abnormal stock volatility around earnings announcements.
2. H2: Control variables – Control variables significantly affects the abnormal volatility, including market capitalization, number of analysts covering the stock, institutional ownership, insider stakeholders, the companies leverage.

Methodology

Data Collection

We use a dataset of 75 stocks from the S&P 500 index. 25 stocks have been picked from the top, middle, and bottom of the S&P 500, this is to include different sectors and company sizes. Each stock has 12 earnings announcement dates over a period of three years, resulting in a total of 900 earnings announcements. To analyze the immediate market reaction to these announcements, we gathered stock prices within a 20-day window surrounding each announcement date—10 days before and 10 days after. This allows us to observe a total of 18,000 data points reflecting short-term price fluctuations and potential information asymmetry caused by the announcements. The same has been done with the S&P 500 index, where the prices were collected for the same periods to have a baseline for market returns. To control normal price fluctuations and make a baseline, the stock prices were also collected for the same stock and periods, but starting 40 days before each earnings announcement, creating a 40-day control period. This ensures to cover the market behavior before the announcement influences the stock price. All the other variables (analyst forecast dispersion, number of analyst estimates, total debt/equity ratio, institutional ownership, insider/stakeholders) were constructed to match the 10-day pre- and post-announcement windows. These other variables where extracted to match the daily volatility and where constant over time. For example, some variables were repeating themselves through a whole earnings announcement period or was the same for the specific stock in all periods. This could be market capitalization. All the data was sourced from FactSet.

OLS

To analyze the impact of asymmetry on stock volatility, the OLS regression model has been chosen. This model is simple and has a straightforward approach and is capable enough for this study's purpose. We just want to find out if there is any relationship at all and if there is evidence for such statistical significance. This model's simplicity is also easy to understand for readers. This method assumes linearity, no perfect multicollinearity, independence of residuals, homoscedasticity, and normally distributed error terms. Each of these assumptions will be tested in the analysis to ensure the validity of the regression results. OLS is preferred over other complex models because it does not require large sample sizes to achieve statistical significance and remains the best linear unbiased estimator if the assumptions are met. The data that is being analyzed is categorized as time series data, it is an analysis of price volatility around earnings announcements, which involved change over time. When doing time series analysis, it is important to check for stationarity, this will make the regression modelling more reliable. This is done by using the Dickey-Fuller test. To do Ordinary Least Squares (OLS) we must follow some assumptions. These assumptions are taken from the book "A Guide to Modern Econometrics." The assumptions are referred to as the Gauss-Markov assumptions and give the most accurate coefficients of a linear regression model if the assumption is met accordingly.

1. Linearity: The dependent and independent variables must be linear.
2. Random sampling: The data must be a random sample from the population.
3. No Perfect Multicollinearity: Independent variables should not be perfectly correlated.
4. Zero Conditional Mean: The mean of the error terms should be zero.
5. Homoscedasticity: The variance of the error terms must be constant across all levels of the independent variable.
6. No Autocorrelation: There should not be any correlation between error from observations.

Data Structure Challenge

The dataset includes a range of variables, most of which, except for abnormal volatility, remain constant over time. Because some variables do not change across different periods, any change in the dependent variable (abnormal volatility) cannot be directly connected to changes in these predictors within the same unit over time. We considered many models for this analysis, such as the Fixed Effect Model, which is suitable because it relies on variations within units to estimate the impact of predictors on the outcome variables. The Random Effects Model was also considered, although it can handle non-constant variables, it requires that these variables be uncorrelated with the unit-specific effects. This assumption may not always hold true, and if violated, it leads to biased estimates. Given the structural limitations for the dataset, the Ordinary Least Squares regression is more suitable for several reasons: Inclusion of constant variables, the OLS does not differentiate between non-constant and constant variables, allowing for the direct inclusion of all predictors in the model. This is beneficial given the constant variables in the dataset. When using OLS with a dataset that has constant variables, the robust and clustered standard error helps this problem by solving different potential error variance incorrect estimations. While the robust standard error adjusts for heterogeneity of variance across all observations, clustered standard error accounts for non-independence of observations within cluster over time. Applying this together helps the reliability of the estimates from OLS regression in datasets where some predictors do not vary over time. For datasets like this, traditional assumptions about error variance and observation independence do not hold, this can skew results and lead to incorrect estimates. The use of both robust and clustered standard errors helps the analysis and estimates with both constant and non-constant variables are as accurate as possible under the given data constraints.

Software

The statistical analysis for this study is done by using R studio, a well-known and reliable platform for statistical computing and graphics. R studio supports a range of statistical methods and provides many different packages for data manipulation, making it a good choice for analysis tasks like this data study. For this research, the packages 'lm()' for the OLS regression is used and 'stargazer' for making regression tables and 'ggplot2' for data visualization are also used to present the data effectively.

Log Returns

The log returns for the stock and S&P 500 index are made to standardize the returns and it helps when dealing with compounded returns over many periods. Log return a preferred for financial data because it is more symmetric and normalizes the returns. Log returns also slow down the effect of extreme values, which gives stability and makes it easy to use as a percentage changes.

$$r_t = \ln(\frac{P_t}{P_{t-1}})$$

$r_t$is the log return at time t, P_t is the price at time t, and $P_{t-1}$is the price at time $t - 1$. (Hull, 2012)

Abnormal Returns

We want to isolate the effect of the earnings announcements from the market movements, and therefore we calculate the abnormal returns. We subtract the Log return of the stock prices with the log returns of the S&P 500.

$$Abnormal\ Return = Log\ return_{Stock} - Log\ Return_{Market}$$

This model shows the excess returns of the stock returns to the market returns.

Beta and Control period

A 40-day control period has been chosen to make a baseline for normal stock market movements. The 40-day control period is prior to the earnings announcements and is used to calculate the beta coefficient. Beta is used to measure the stocks' sensitivity to market movements and is done through a regression, which predicts the stock price based on the S&P 500 index (market returns).

$$Stock\ Return_{i,t} = \alpha_i + \beta_i * Market\ Return_t + \epsilon_t$$

$\alpha_i$is the intercept.

β_i is the beta.

ϵ_t is the residual error term (Sharpe, n.d.).

Using Beta in Abnormal Returns

The beta from the control period is used to adjust the abnormal returns for the 10-event window to make sure the returns match the isolated response to the earnings announcement and not the whole market.

$$Ajusted\ Abnormal\ Returns = Abnormal\ Return_{i,t} - (\alpha_i + \beta_i * Market\ Return_t)$$

The equation gives a more isolated measure of the events effect by taking away the market change in the stock price (Ball, n.d.).

Volatility

The volatility is calculated by the standard deviation of the daily returns. It is calculated for the 10-day event window and the 40-day control period.

$$Volatility = \sqrt{\frac{\sum(R_{i,t} - \bar{R}_i)^2}{n-1}}$$

$R_{i,t}$ is the return.

$\bar{R}_i$ average return over the period.

This equation measures the movement of the price fluctuations (Mandelbrot, n.d.).

Abnormal Volatility

The abnormal volatility is calculated by a comparison between the volatility from the 10-day period and the 40-day control period.

$$Anormal\ Volatility = Volatility_{10day} - Volatility_{40day}$$

The results show an increase or decrease in volatility around earnings announcements.

Analyst Forecast Dispersion

To measure the information asymmetry, the analyst forecast dispersion is used. It is the difference between the highest and lowest earnings per share.

$$Dispersion = \max(EPS\ Forecast) \text{ - min (EPS forecast)}$$

This equation gives the difference between expectations between analysts and is an indicator of the uncertainty (asymmetric information) before the earnings announcements. Higher dispersion gives a greater uncertainty and higher reaction to news (according to our hypothesis) and lower uncertainty gives lower reactions.

Model Definition

Stationarity

Stationarity in a time series refers to the statistical properties of the mean, variance and autocovariance is constant over time and do not depend on the time the data is observed. It also implies that the series will tend to return to a long-term meaning over time. The variance does not depend on time, the spread around the mean remains constant. It has limited memory of the past; the impact of shocks will not persist and will allow the series to return to its usual pattern (Marno Verbeek, n.d.-a).

Non-Stationarity

Non-Stationarity means the time series has a unit root, which means the statistical properties change over time. A unit root in a time series means the series does not return to a long-term mean but instead follow shocks and trends, which has a lasting effect. This kind of model has no mean reversion, but instead has properties acting more like random walk (Marno Verbeek, n.d.-a).

The Dickey-Fuller test

The Dickey-Fuller test is a model used for testing stationarity in time series. It determines whether it has a unit root, which indicates non-stationarity. The unit root would mean that unexpected changes have a lasting effect, preventing the series to return to its mean over time. The model test statistic to critical values from a distribution under the null hypothesis of a unit root (Marno Verbeek, n.d.-b).

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \emptyset_i \Delta y_{t-i} + \epsilon_t$$

y_t Is the time series at time t.

Δ The change from one period to the next-

α is the intercept (constant).

βt is a coefficient for time trend variable t.

γ The coefficient on y_{t-1}(lagged values), $\gamma = 0$ suggest the time series is random walk. If $\gamma \neq 0$ means stationarity, it indicates the series reverts to mean over time.

$\emptyset_i$is the coefficient on the lagged first difference Δy_{t-i}. This is for autocorrelation.

ϵ_t is the error term. It is randomness not explained by the model.

(James D. Hamilton, 1994)

Anderson-Darling Test

The test is made to check the goodness of fit for normal distribution against a set of sample data. It tests if the data matches normal distribution, focusing on the ends of the distribution to detect outliers and how data behaves in the tails. Its helps identify if the data follows a normal distribution, and this is one of the assumptions for the OLS.

$$A^2 = n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log\left(\Phi(Y_i)\right) + Log(1 - \Phi(Y_{n+1-i}))\right]$$

n is the sample size.

Y_i are the ordered sample data.

Φ is the cumulative distribution function of normal distribution (D.B Owen, n.d.).

OLS regression

Ordinary Least Squares Regression is a way to estimate relationships between a dependent and independent variable. This statistical model tries to find the best-fitting line that describes the relationship between these variables.

$$Y = \beta_0 + \beta_1 x1 + \beta_2 x2 + \cdots + \beta_k xk + \epsilon$$

Y is the dependent variable.

$x1, x2 \ldots xk$ are the dependent variables.

$\beta_0, \beta_1 \ldots \beta_k$ are the estimations (coefficients)

ϵ are the error term, this measure the variability in Y that is not explained by the independent variable.

The coefficients are estimated by least squares, this involves minimizing the sum of the residuals, between the observed values and the predicted values. The OLS has assumptions to provide the best linear unbiased results. This includes linearity between the dependent and independent variables. Independence and observation must be uncorrelated in terms of residuals. Homoscedasticity, the residuals must be constant across all levels of the independent variable. No multicollinearity, the independent variable must not be highly correlated with each other. Normal, distribution of errors, the residuals must be normally distributed (Marno Verbeek, n.d.-c).

Linear Regression Equation

This is used for the purpose of testing the relationship between the independent variable and the dependent variable.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

Y is the dependent variable.

$X_1 \ldots X_2 \ldots X_n$ are the independent variables. In this case there is only one X_1.

$\beta_1 \ldots \beta_2 \ldots \beta_n$are the coefficients, this is what the regression estimates. In this case only one β_1.

ϵ is the error term, this is the portion of the dependent variable that is not explained by the predictors. It is normally distributed with a mean of zero and constant variance (Marno Verbeek, n.d.-d).

Robust Regression

The Robust Regression is used to estimate models when they are outliers, or the assumptions are violated for an OLS. The Robust Regression reduces the influence of outliers by dividing them into weights which in a way minimize the influences of outliers. This approach improves the model against normality of error or homoscedasticity.

$$\underset{\beta}{min} \sum_{i=1}^{n} \rho(y_{i-x_i^T}\beta)$$

This is the equation for estimating the coefficients that minimized the influence of the outliers.

ρ Minimize the sum of residuals. It is increasing less quickly than the square of residuals.

β These are the parameters we are trying to estimate.

$(y_{i-x_i^T}\beta)$ Difference between the observed values and the predicted values.

$\sum_{i=1}^{n}$ This indicates that the ρ is applied to all the residuals and the observations are summed up. It tries to find the coefficients that's give the smallest possible total, which is the best fit (Marno Verbeek, n.d.-e).

Clustered Regression

Clustered regression is used to adjust standard errors in regression models where observations within the same cluster might not be independent. This could be due to factors like time periods. Such cluster can lead to error terms that are correlated within a cluster but independent between clusters. It is calculated by using a covariance matrix of the regression coefficients.

$\hat{V} = (X'X)^{-1}(\sum_{c=1}^{C} X_c'\hat{u}_c\hat{u}_c'X_c)\,(X'X)^{-1}$

X Is the matrix of independent variables for all observations.

$X'X$ The product of X transposed and X summarizing the variance and covariance for the independent variables.

$(X'X)^{-1}$ The inverse of $X'X$ is important for estimating regression coefficients.

X_c Shows the matrix of independent variables for observations within the cluster c.

$\hat{u}_c$ The residuals from the regression model for each cluster c.

$\sum_{c=1}^{C} X_c{}'\hat{u}_c\hat{u}_c{}'X_c$ is a summation over all clusters, measuring the covariance of residuals within each cluster (Marno Verbeek, n.d.-f).

The Breusch-Pagan test

This is used to test for heteroskedasticity, which means that the variability for the residuals is not constant across all levels of the independent variable. Heteroskedasticity can make the significance of the coefficients incorrect because the standard errors are biased. We want the regression to have homoskedasticity, which means that the variance of the residuals is consistent across all levels of the independent variable.

BP Statistic $= n * R^2$

n is the number of observations.

R^2 is the coefficient that measures how well the independent variables explain the variation in the squared residuals from the original regression model, indicating whether the variability of errors changes with these variables.(Marno Verbeek, n.d.-g)

Weighted Least Squares (WLS)

WLS is a variation of the OLS regression, that makes different weights of residuals. It gives less weight to observations with higher variance that are considered less reliable. This improves efficiency of the parameter estimates, making it more reliable when the OLS contains heteroscedasticity.

$$\beta_w = \left(\sum_{i=1}^{N} w_i x_i x_i^T\right)^{-1} \sum_{i=1}^{N} w_i x_i y_i$$

w_i this represents the weight for each observation.

$x_i$is the vector for the independent variable w_i .

$w_i$is the weight for each observation i.

y_i is the dependent variable.

β_w $is\ the$ estimated coefficient.

$x_i x_i^T$ is the matrix made by multiplying the vector x_i by its transpose.

$\sum_{i=1}^{N} w_i x_i x_i^T$ aggregates the outer products across the observations, each scaled by its weight. This sum gives the weighted version of the covariance matrix of the independent variable.

$\sum_{i=1}^{N} w_i x_i y_i$ this sums up all these weighted products across all observations (Marno Verbeek, n.d.-h).

Durbin-Watson Test

The test is used to check for autocorrelation in the residuals of a regression.

$$DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Where e_t represents the residuals at time t, T is the number of observations. If the number of the statistics is near 2, there is no autocorrelation. Values near 0 suggest that there is a positive correlation. Values greater than 2 suggest negative autocorrelation (Marno Verbeek, n.d.-i).

Variance Inflation Factor (VIF)

This is a measure that quantifies the correlation between predictors in a regression model. It is used to identify multicollinearity where the higher values indicate greater multicollinearity.

$$VIF = \frac{1}{1 - R^2}$$

Where R^2 measures the percentage of variation in one predictor that is explained by the others, a higher R^2 means that the predictor is highly dependent on the other variables (Marno Verbeek, n.d.-j).

Analysis

In this section of the study, we make a statistical analysis to examine the volatility effects on stock prices and S&P 500 indices around earnings announcements. Our analysis begins with an examination of the distributions of stock and index prices, focusing on skewness and kurtosis to understand their deviations from the normal distribution. To build a robust foundation for our regression analysis, we first make the stationarity of our datasets through the Augmented Dickey-Fuller (ADF) test. The results confirm the absence of unit roots across all variables, which supports the reliability of the statistical modeling. Because of the complexity of the data, regression needs to account for potential non-linear relationships. Scatter plots employed show that linearity assumptions required for ordinary least squares (OLS) regression do not hold in our data. To solve this, a robust regression is made, this regression is less sensitive to outliers. We also employ clustered standard errors to handle correlations within the group, to reduce the effects of heteroscedasticity and autocorrelation. Autocorrelation in the residuals, identified through the Durbin-Watson test, gives challenges, indicating that OLS assumptions about independent errors are violated. We solve this problem by using methods that adjust the standard errors to improve the reliability of the estimates, ensuring that the findings are not biased. Moreover, to measure the immediate effects of earnings announcements on stock price volatility, there have been introduced dummy variables representing the event day. This approach allows us to isolate the impact of earnings releases from the regular market movements, giving a better view of the announcement's influence on market behavior.

Descriptive statistics

Looking at the histograms and the statistical results for the stock prices and S&P 500 prices used for this analysis, we can give an insight into the distribution of these data. The stock prices skewness (1.991662) implies that it is a right-skewed distribution. This means that there is a long tail on the right side of the distribution. Most of the stock prices are clustered around the lower end, with not many extreme prices. The kurtosis (9.411465) is significantly higher than 3 which is normally the normal distribution. This indicates leptokurtic distribution. This suggests that stock prices have a sharp peak and fat tail compared to normal distribution. This can mean a higher probability of extreme values. The skewness (0.3890543) of the S&P 500 prices has relatively low skewness, which means the distribution of the S&P 500 prices is more symmetric. There is a small right skew, but it is minimal, this means that most of the data is around the mean. The kurtosis (3.003058) is close to 3 which gives a mesokurtic distribution and is a normal distribution. This implies that the S&P 500 prices do not have heavy tails and are less likely to have outliers compared to stock prices.

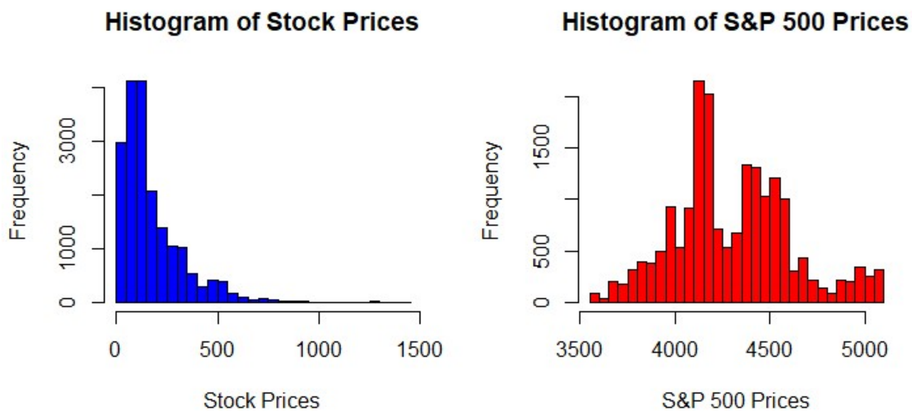

Figure 1 Histogram for data distribution on Stock prices and S&P 500 index prices.

The histogram of stock prices shows a right skew with large clustering of stocks at lower prices and progressive decline as prices increase, this supports the analysis of skewness and kurtosis. The histogram of the S&P 500 prices gives a more central peak around 4500 with symmetric tails on either side, also confirms the normal kurtosis and lower skewness. The distribution centers around the middle value with fewer extreme points. The stock prices high kurtosis and skewness can indicate a dataset with outliers where the prices are either very high or very low, which is more than it would be in a normal distribution. The S&P 500's almost normal distribution suggests less risk of extreme values, this can mean it may be a more predictable market for investors based on these historical prices. The stock prices range from minimum (0.0468) to maximum (1407.01), which indicates a wide range of prices. The first quartile (25%), (73.5575) is the stock prices that are below this value. Median (50%), (125.3766) means that half of the stock prices are below this level, which is the midpoint of the dataset. Third quartile (75%), (230.1550) means that 75 percent of the stock prices are below this value. The mean (171.7951) which is significantly higher than the median, can imply presence of higher value outliers affecting the average. The S&P 500 prices range from 3577 to 5096 which means it has less variations compared to the stock prices. First quartile (25%), (4117). Median (50%), (4272). Third quartile (75%), (4499). The mean (4301) is close to the median which means a more symmetrical distribution. The significant variations in the prices of 75 stocks in normal due different companies and sectors. Each stock behaves according to its own set of market conditions, this influences the overall distributions shape. Different industries have different price behavior, tech stocks may have higher growth and high volatility, which gives high skewness and kurtosis. Larger companies normally have more stable stock prices, which gives less skewness and kurtosis, while the smaller companies' stock might have volatility price movements. Some stocks may have higher trading and speculative interest which can lead to price spikes and asymmetrical distribution. The S&P 500 index appears more normally distributed, because it is an index where the large 500 companies er from different sectors which smooth out the volatility through diversification. This makes extreme values and asymmetries less seen and makes distributions more closely to normal curve. The index gives a broader market price which averages out the behavior of price compared to individual stock prices. Descriptive statistics are normally a summary of the observations and their distributions. The problem with this dataset occurs because of the 75 different stocks, all the stocks have different price ranges. The stock price data is observed over different sectors and has significant variability, for example, the histograms shows that while individual stock prices show a right-skewed distribution with extreme outliers and high kurtosis, the S&P 500 prices shows more symmetric and mesokurtic patterns, which is normal distribution with fewer extreme values. This shows the challenges of stock price distribution. Applying the same statistical measures across different stocks can lead to misleading results. The non-normality of stocks directly influences the calculation of abnormal returns. This can lead to underestimated or overestimated volatility if one assumes normality. Higher moments of distribution can affect the variance and standard deviation which lead to incorrection of the risk and returns calculated. Statistical rest and confidence intervals based on normal distribution may not be valid for example, the t-test and z-test that assume normal distribution for calculating p-values, may be misleading. When prices show significant deviations from normality, the results of abnormal returns from prices may lead to challenges in modeling a standard regression that assumes normal error distributions. Abnormal volatility in stocks often comes from the idiosyncratic movements of individual stocks which may not be normally distributed. When making abnormal volatility using statistical methods is necessary when we are considering these problems. The Anderson darling test is used to check for normality. When using this test and it shows non-normality in stock prices it confirms the need to take this into account when analyzing the abnormal volatility. To solve this issue there has been used a robust regression.

The 40-day control period is chosen to make a baseline for normal stock behavior prior to the earnings announcements. Comparing the stock movements with this extra period, we can make a more accurate analysis of abnormal volatility.

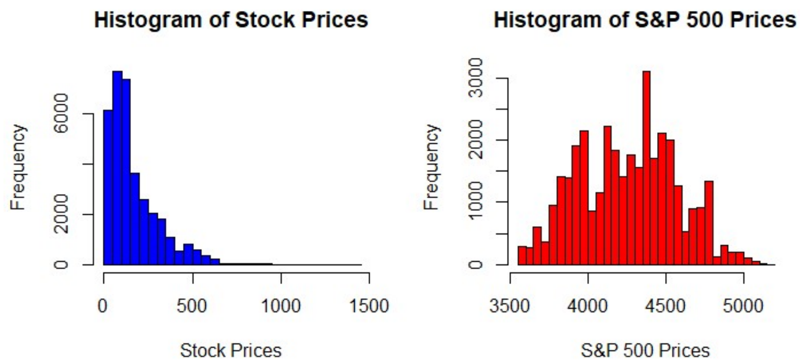

Figure 2 Histogram for data distribution on Stock prices and S&P 500 index prices on the 40-day control period.

The histogram here shows a left skew (1,843508), this means most of the data is clustered around the right side with a tail towards the lower side. This means that most of the stock with a very high price is pulling the mean up. The high kurtosis (8.34213) values indicate a leptokurtic distribution, this means the data has a heavy tail and are sharp peaked compared to the normal distribution. The S&P 500 prices histogram is more symmetric as it reflects a low skewness (0.674357). This means that the S&P 500 index prices are distributed around the mean with fewer extreme values compared to stock prices. The kurtosis value (2.335775) is close to the kurtosis of a normal distribution, this indicates a platykurtic distribution. This means that the S&P 500 prices have less outliers compared to stock prices. The right skewness and high kurtosis observed in both periods indicates stock prices are generally prone to extreme values, but the 40-day control period is slightly less prone to this. This suggests that while extreme price movements are normal, they are less likely to happen when observing long periods. This is of course before earnings announcements which gives fewer extreme values compared to the dataset which has earnings announcements included. The S&P 500 prices low skewness and kurtosis in both periods confirms the stability to use this for a benchmark. The 40-day periods allow for more precis estimation of abnormal returns by giving stock normal performance without the immediate effects of earnings. Having a longer window slows down the effect of outliers and reduces the skewness and kurtosis, this gives a more normalized dataset that could help the robustness of the regression models. The right skewness and high kurtosis in both periods stock prices even during the control period, can give extreme values which could skew the results and be misleading. The S&P 500 shows a near-normal distribution and the individual stock does not, this could mean that using the index as a benchmark could overlook individual stock volatility. This could affect the accuracy of individual stock reaction to earnings announcements. It has been chosen not to perform the same distribution analysis on the control variables as the observations used for the variables (stock prices and S&P 500 prices), because they are more directly related to the research question and the dependent and independent variables is most important for the analysis.

Abnormal Volatility

The beta coefficient from the 40-control periods is used to adjust the stock returns observed in the 20-day window around earnings announcements. Beta is a measure of the stock volatility to the market, is calculated during the control period to capture stocks typical market movements. Beta is then used to adjust the returns in the 20- day period for normal market movements, which isolates the excess return directly. This adjustment makes sure that the abnormal returns calculated reflect only the effects of the earnings announcement. The control period is also used as a baseline for normal volatility of stock prices. By measuring the standard deviation of the stock returns during the 40-day control period, we make a benchmark for typical volatility for each stock. This normal volatility is then compared to the volatility observed in the 20-day window to calculate abnormal volatility. Is done by subtracting the normal volatility from the observed volatility.

$$Anormal\ Volatility = Volatility_{10day} - Volatility_{40day}$$

The challenge of using a 40-day control period to calculate abnormal returns and abnormal volatility is that the investors often already expect higher volatility around earnings announcements. These expectations can influence their behavior. As a result, the stock prices and volatility observed during the event window may already include these expectations, making it difficult to isolate the effect of the earnings announcements themselves. The market reaction to the earnings announcement is not only reflected in earnings announcements but also the expectations built up in the days leading up to the announcement. This can result in overestimation of abnormal returns and abnormal volatility. This could be solved by modeling the expected volatility in the earnings announcements period, using historical data. This involves analyzing past earnings announcements and their impact on volatility to develop an expected volatility benchmark for similar futures events. By adjusting the observed volatility in the event window for this expected volatility, we can calculate an excess abnormal volatility, that isolates the market reactions to earnings announcement itself and excess what was expected of the investors. This can give a better view of the information asymmetry and investor sentiment around the earnings announcement. The GARCH model (Generalized Autoregressive conditional heteroskedasticity) can be sued to estimate expected volatility. These models can provide more accurate estimates of expected volatility (Ruey S. Tsay, n.d.). While the 40-day control period is useful as a benchmark, it is important to account for the higher volatility expected by the investor during the earnings announcements. Having another approach like using another model for expected volatility could provide a more accurate measure of abnormal volatility, better showing the true effect of earnings announcements.

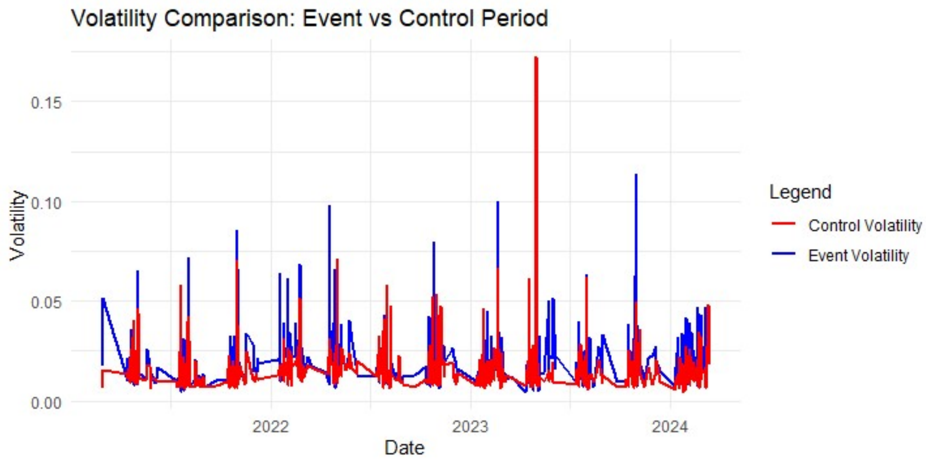

Figure 3 Volatility comparison for the 20 days event period and the 40-days control period.

The plot shows the comparison of volatility for the event period (blue line) and control period (red line) across different dates. The ‘Event-Date’ is on the X-axis and the volatility values (‘Event_vol’ and ‘Control_vol) are on the Y-axis. The event volatility is the volatility during the 20-day window around the earnings announcements. There are many peaks, which implies high volatility during earnings announcements. The control volatility is the baseline for volatility during the 40-day control period but before the earnings announcements and generally the control volatility is lower than the event volatility, which is expected since it covers period without influence of earnings announcements. As expected, the earnings announcements (event periods) show higher volatility compared to the control periods. This confirms the hypothesis that earnings announcements significantly impact the stock price volatility. The peak in event volatility is an indication of the market is highly reactive to new information linked to earnings reports. There is one instance where control volatility is higher than event volatility, suggesting unusual market behavior, this could be caused by external factors or market events unrelated to earnings announcements.

The next graph illustrates the abnormal volatility around the earnings announcement dates, from 10 days before and 10 days after the announcement day, which is day 0. The abnormal volatility here is defined as the difference between the stock prices volatility and the control periods volatility. The vertical axis is the measure of abnormal volatility, and each red dot shows the range of volatility observed on each day. This gives day-to-day abnormal volatility movements. The blue line on the horizontal axes connects the average values of abnormal volatility for each day and shows the overall trend and the volatility response to earnings announcement.

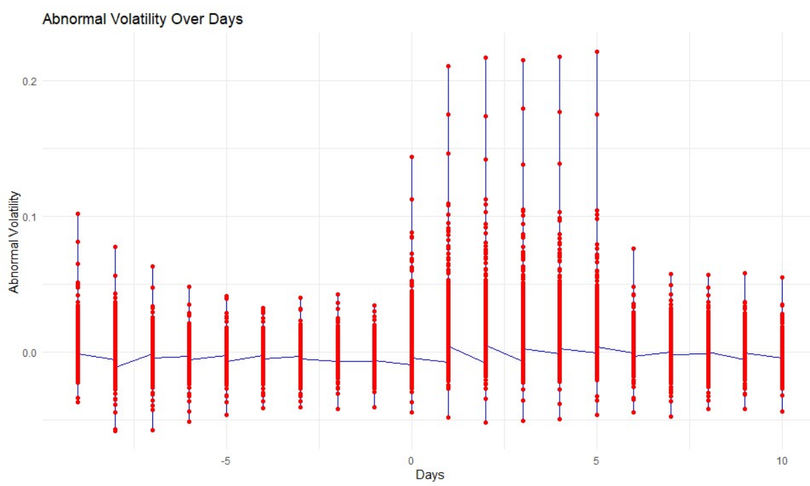

Figure 4 Abnormal Volatility over the 20 days period.

The graph shows that the abnormal volatility decreases the days before earnings announcement. This could mean that the investors are reducing their activities because of uncertainty as the announcement approaches. After the announcement on day 0, there is a significant peak in the abnormal volatility, especially on day 1,2,3,4 and 5. After these peak days, their volatility starts to adjust to normal again. Overall, the abnormal volatility confirms the hypothesis about earnings announcements influence the abnormal volatility significantly. This is also evidence of the calculation of the abnormal volatility is made correctly as it shows traditional assumptions that suggest lower volatility before earnings, and a reaction after earnings, and market adjustments.

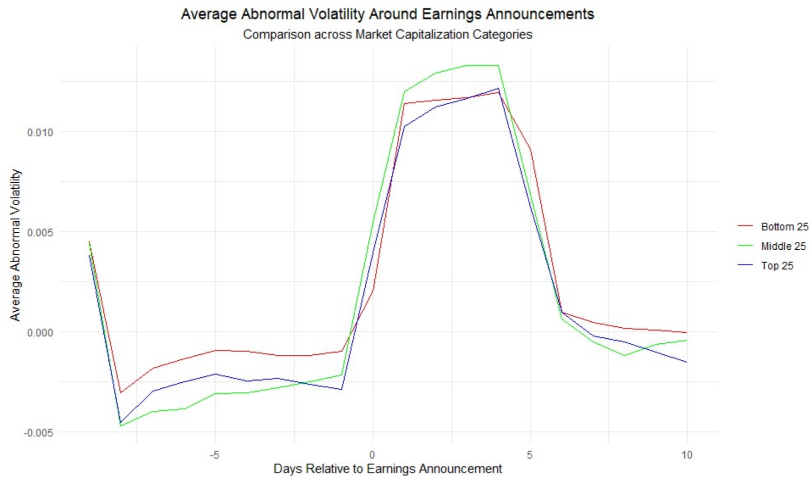

Figure 5 Abnormal volatility around earnings announcement, categorized into market capitalizations.

The graph above illustrates the average abnormal volatility of stocks around their earnings announcements, analyzed across three market capitalization categories: the bottom 25, middle 25, and top 25 percentiles of the S&P 500. The x-axis is the days relative to the earnings announcement, from 10 days before to 10 days after the announcement, with day 0 which is the actual day of the earnings release. The y-axis shows average abnormal volatility, reflecting deviations from typical stock price volatility levels expected under normal market conditions. In the days leading up to the earnings announcement, there is a trend across all categories. Volatility begins to decrease as the announcement day approaches. Investors potentially reduce trading activity as they wait for new financial data, leading to lower volatility. This behavior shows the market's uncertainty and the high stakes associated with the upcoming earnings information. On the announcement day itself, there is a peak in abnormal volatility across all three categories, with the middle 25 group having the most significant spike. This higher volatility in the mid-cap segment could be because of their position within the market. Mid-cap companies are often in a phase of growth, potentially leading to more significant disparities between expected and actual earnings, which in turn gives stronger market reactions. These companies may not be as closely followed as the larger companies, meaning any unexpected news can cause extreme reactions. Following the earnings announcement, the graph shows a period of abnormal volatility that extends through five days, reflecting the market's reaction to the new information. This may be caused by investors trying to reposition portfolios based on the earnings outcomes and their implications. The abnormal volatility begins reducing, indicating that the market is reaching a new equilibrium after the earnings data. The volatility observed among the three groups also confirms the diverse investor behaviors and liquidity typically associated with each category. Large-cap stocks, represented by the top 25, normally attract more stable, long-term investors, which might explain the relatively lower peaks in volatility compared to the mid-caps. The bottom 25, which includes smaller companies, shows less volatility peaks may be because of less market attention and lower trading volumes. This analysis suggests that while all market segments react to earnings announcements, the level of reaction is significantly different.

Correlation Matrix

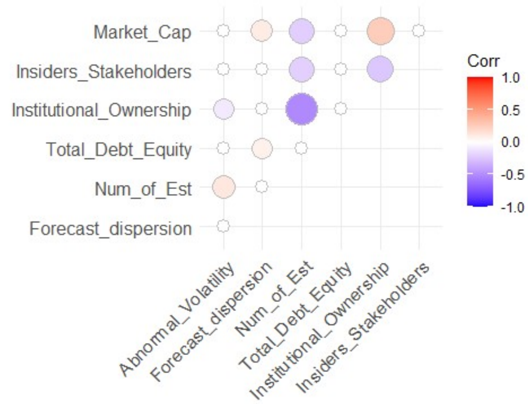

Figure 6 Correlation matrix of all the variables.

The correlation matrix shows the relationship between abnormal volatility (abnormal_volatility) and other variables. (forecast_dispersion), (num_of_est), (total_debt_equity), (institutional_ownership), (insiders_stakeholders), and (market_cap). The correlation coefficient ranges from -1 to 1. 1 indicates a perfect positive correlation. -1 indicates a perfect negative correlation. 0 indicates no correlation. The circles show the correlation coefficients, larger circles indicate stronger correlations. Positive correlations are shown by red circles. Negative correlations are shown as blue circles. Only significant correlations are shown in the figure. The plot does not show insignificant correlations, the spots will be blank. Abnormal volatility has very weak correlation with the other variables. This suggests that abnormal volatility does not have a strong linear relationship with all the other variables. Forecast dispersion shows weak correlation with other variables. There are no large circles, indicating forecast dispersion does not have strong linear relationship with other variables. The number of estimates has a negative correlation with institutional ownership, this implies that when the number of estimates increases institutional ownership decreases and vice versa. The number of estimates also show a small positive correlation with the total debt/equity ratio. Total debt/equity has a moderate negative correlation with institutional ownership. This could imply that the companies with higher debt to equity ratios tend to have lower institutional ownership. There is also a small positive correlation with the number of estimates. Institutional ownership has a strong negative correlation with insider stakeholders. This could imply that as institutional ownership increases, insider stakeholder ownership decreases. There is a small positive correlation between institutional ownership and market capitalization. Insider's stakeholders have negative correlation with institutional ownership, this is the strongest negative correlation in the matrix. It also has a small negative correlation with market cap. Market cap has a small positive correlation with institutional ownership. Overall, the abnormal volatility is independent of the other variables, showing no significant correlations. Institutional ownership and insider stakeholders are the most correlated pair. This overall may suggest non-linear relationship that is not captured by correlation. The number of estimates' negative correlation with institutional ownership suggests that the companies with more analyst coverage tend to have lower institutional ownership. This could be because that more analyst coverage might be associated with higher information availability and the institutional investors don't necessarily need to invest in large companies with more information availability because they have more access to detailed information than the rest of the public. The number of estimates has a small positive correlation with total debt/equity. This could mean that companies with more debt might attract more analysts. Total debt/equity has negative correlation with institutional ownership, this might suggest that institutions prefer investing in companies with more stable capital structures. The total debt/equity has positive correlation with the number of estimates and might indicate that companies with higher debt are more favorable and interesting for analysts. Institutional ownership has a strong negative correlation with insider/stakeholders, this suggests that higher institutional ownership means a decrease in insider ownership. This reason for this may be because when the institutions buy more shares, they often buy them from inside owners. Institutional ownership has a positive correlation with market cap, this implies larger companies tend to have higher institutional ownership, due to greater stability and less risky companies. Insiders/stakeholders have negative correlation with institutional ownership, this supports the correlation before, but there is a slight negative correlation with market capitalization. Smaller companies tend to have higher insider ownership, this is normal because these smaller companies may be at their earlier stage of growth and insiders maintain more control. The market cap has positive correlation with institutional ownership, which means that larger companies a more likely to have institutional ownership which reflects their stability, lower risk and attract large investors. Institutional investors seem to prefer larger and less leverage companies with lower insider control (insider stakeholders). The number of estimates is influenced by financial structures like debt but not significantly by ownership structures, this implies that analysts focus on financial health. Abnormal volatility seems to be influenced by other factors that are not related to this study.

The Dickey-Fuller

The first step is to test our variables for stationarity, using the ADF test. The Dickey-Fuller test results are overall positive for the purpose of this research. The ADF Statistic is negative in all the datasets, which rejects the null hypothesis of the unit root and indicates stationarity. This includes the dataset for Abnormal volatility, Forecast dispersion, Total debt/equity ratio, Institutional Ownership, Insiders Stakeholders, Market Capitalization, Number of Estimates. This suggests all the data is stationary and implies that mean, variance does not change over time, this is important for the reliability in the modeling. This helps the OLS regression for estimate unbiased and consistent results. It is important that the error term has constant variance and does not depend on series correlation (Marno Verbeek, n.d.-k). Non-stationary variables could lead to high R-squared and t-statistics that implies good fit even when the coefficients are not correct as the models suggest, this could lead to a Spurious regression. The successful Dickey-Fuller test also confirms the OLS model assumption about the t-test and F-test error terms are independent and identical distributed with a constant variance and a zero mean.

Variable	ADF Stats	Lag Order	p-value	Alternative Hypothesis
Abnormal Volatility	-21.115	25	0.01	stationary
Number of Estimates	-6.2733	25	0.01	stationary
Forecast Dispersion	-9.1736	25	0.01	stationary
Institutional Ownership	-5.5067	25	0.01	stationary
Insiders/Stakeholders	-6.6813	25	0.01	stationary
Total Debt/Equity (%)	-16.509	25	0.01	stationary
Market Cap	-6.3421	25	0.01	stationary

Linearity

The scatter plots show the linearity between the dependent variable abnormal volatility and the other variables (independent and control variables). The results indicate that the assumption of linearity is not met, because of the flat and non-linear relationships in the scatterplots. Even when the analysis shows that the linearity assumptions are not met, we still decide to proceed with the robust regression and clustered standard error.

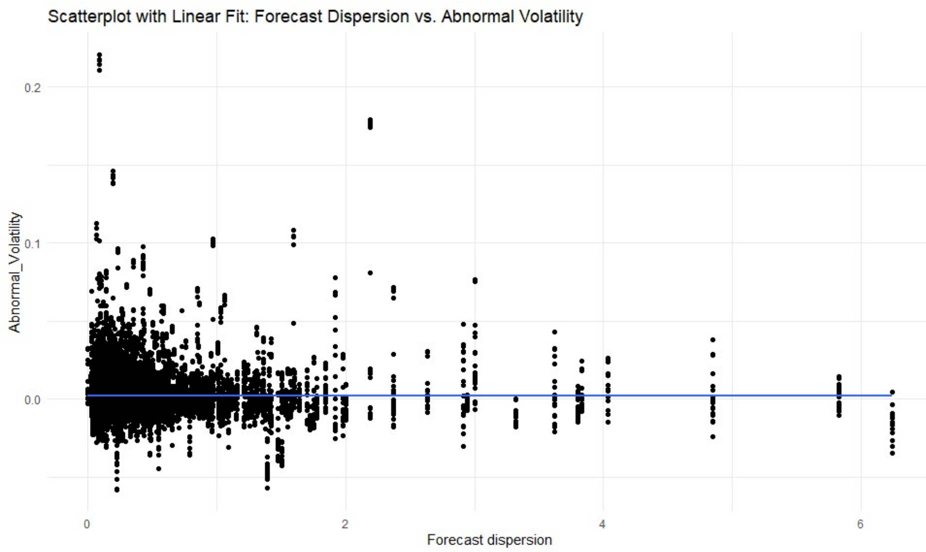

Figure 7 Linear Fit between forecast dispersion and abnormal volatility.

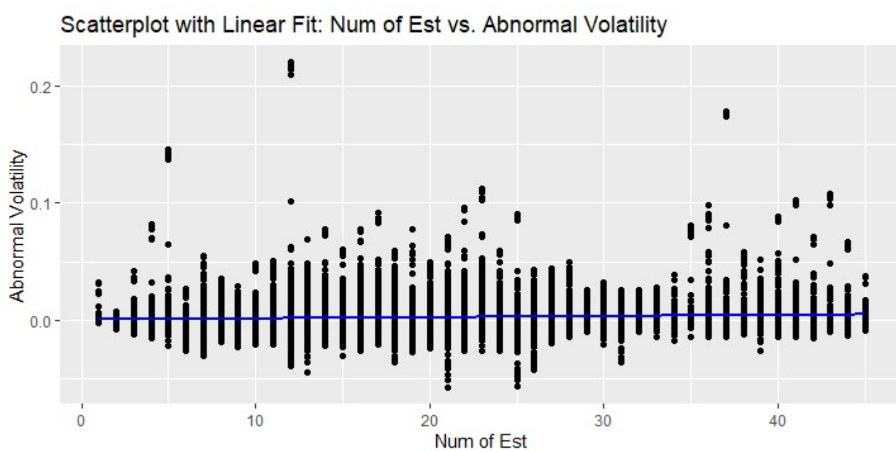

Figure 8 Linear fit between number of estimates and abnormal volatility.

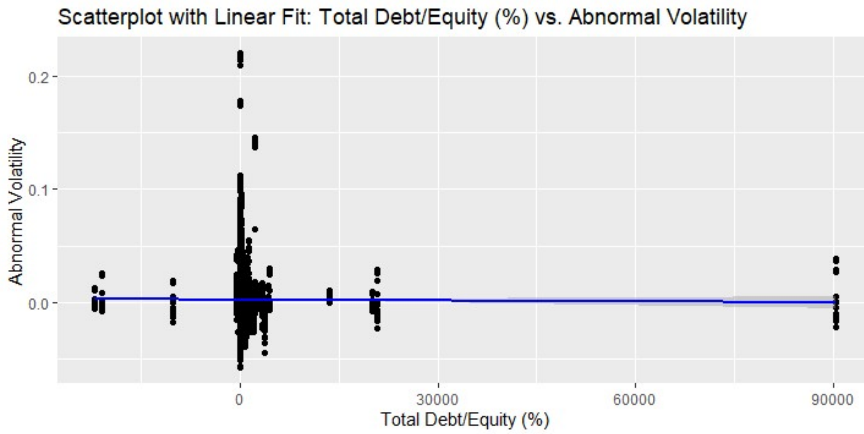

Figure 9 Linear fit between the total debt/equity ratio and abnormal volatility.

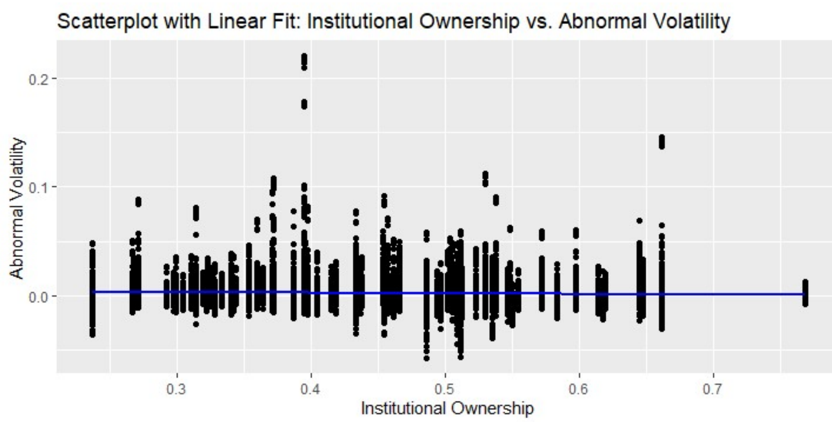

Figure 10 Linear fit between institutional ownership and abnormal volatility.

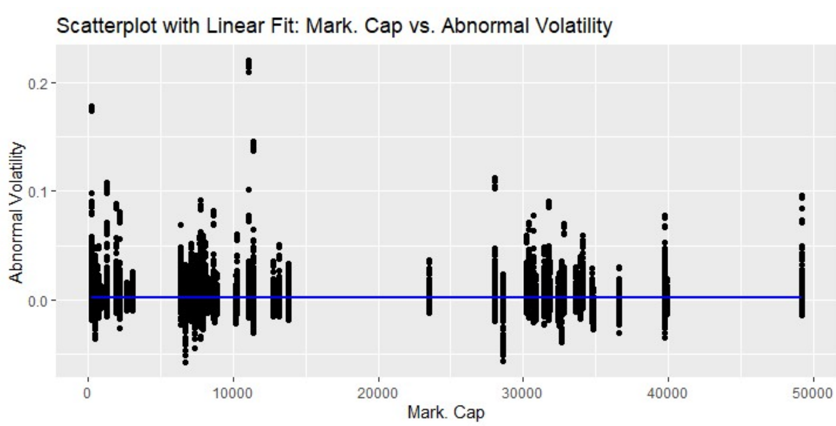

Figure 11 Linear fit between market capitalization and abnormal volatility.

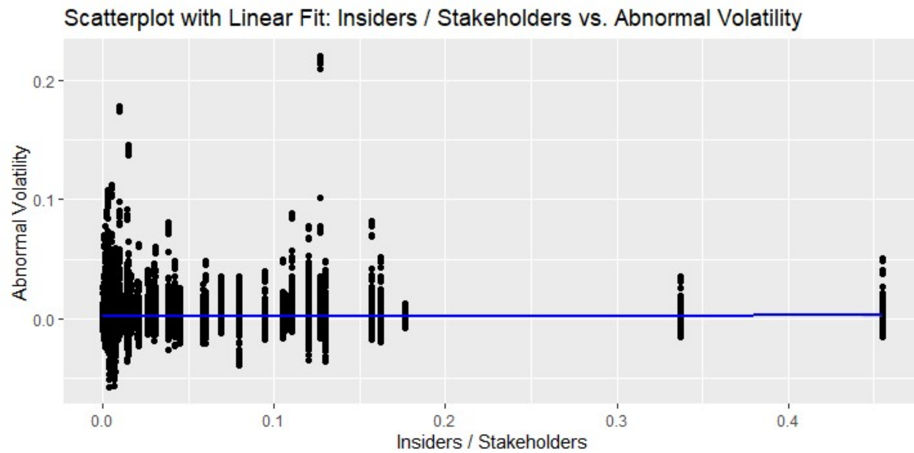

Figure 12 Linear fit between insider/stakeholder and abnormal volatility.

Even when the linearity is missing, we continue the analysis for the following reasons. Robust regression is less sensitive to violations of the linearity assumption. It can handle outliers and provide more reliable estimates even when the assumptions of ordinary least squares regression are not met. The clustered standard errors can account for within-group correlations and heteroscedasticity. This can help reduce the impact of violations in the assumptions of OLS regression, including linearity and gives more reliable standard error and test. The primary focus is to understand the impact of information asymmetry on stock price volatility, robust regression and clustered standard error are appropriate methods. This method will be executed later in the regression, but first we check for autocorrelation.

Durbin-Watson

The Durbin-Watson test indicates significant autocorrelation in residuals at lag 1. Durbin-Watson statistic (0.4542602) is far from 2 and implies strong positive autocorrelation. Lag 1 autocorrelation (0.7728615) This mean the residuals from on observation to another one is autocorrelated. P-value=0 indicates that the test rejects the null hypothesis.

Metric	Value
Durbin-Watson statistic	0.45564
p-value	< 2.2e-16
Alternative hypothesis	True autocorrelation is greater than 0

Residuals vs. Index

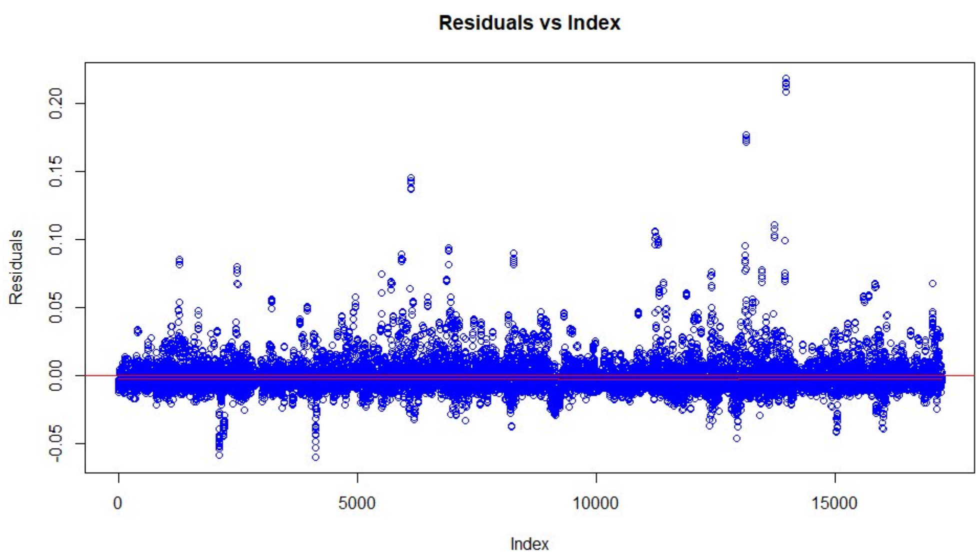

Figure 13 residuals against index.

This figure illustrates residuals plotted against their index, showing patterns that suggest there is autocorrelation. Most of the residuals are centered around zero, indicating the model generally fits the data well. However, there are clusters of residuals that are above or below the zero line, suggesting a non-random pattern. This clustering indicates autocorrelation. The variance of the residuals appears to be consistent across the index, suggesting homoscedasticity. But the presence of some extreme residuals above and below the zero line further supports the possibility of autocorrelation, which aligns with the Durbin-Watson test result indicating positive autocorrelation. Autocorrelation can lead to biased parameter estimates and make hypothesis tests unreliable. This issue needs to be addressed as it violates one of the key assumptions of OLS regression, which requires no autocorrelation residuals. Solutions such as using robust standard errors can solve this problem.

Robust Standard Error

To solve this problem, we have done a robust standard error using the ‘sandwich’ packaged in R. This is adjusting the standard error of the coefficients to account for autocorrelation and the non-linearity we tested for earlier, this will allow for more reliable models, even when the autocorrelation OLS assumption is not met.
It is a way to correct the standard errors of the estimates, to make them robust for homoscedasticity and autocorrelation. The results of the Standard Error Testing will be analyst in the following section:
t test of coefficients:

Variable	Estimate	Std. Error	t value	Pr(>|t|)	Significance
Intercept	0.00075945	0.00066159	1.1479	0.25102	
Forecast Dispersion	-0.00019611	0.00024962	0.7856	0.43209	
Number of Estimates	0.000096387	0.000014887	6.4747	9.76E-11	***
Total Debt/Equity	-1.0043E-08	3.9221E-08	0.2561	0.79791	
Institutional Ownership	-0.0014597	0.0010183	1.4335	0.15174	
Insiders/Stakeholders	0.0026289	0.0011799	2.2281	0.02589	*
Market Cap	3.1798E-09	7.6438E-09	0.416	0.67742	

The intercept is the expected value of abnormal volatility when all the independent variables are zero. This intercept (7.5945e-04) is not significant and implies that it does not differ from zero.
Forecast dispersion (-1.9611e-04) is negative, which means that higher forecast dispersion is linked with a small decrease in abnormal volatility. This relationship is not statistically significant, indicating that forecast dispersion does not have meaningful impact on abnormal volatility in this model. The number of estimates (9.6387e-05) is positive and significant. This indicates that an increase in the number of estimates is linked to an increase in abnormal volatility. This relationship is robust and suggests that more estimates lead to greater market volatility. Total debt/equity ratio (-1.0043e-08) is negative but not statistically significant ($p>0.05$). This implies that the debt/equity ratio does not have an impact on abnormal volatility. Institutional ownership (-1.4597e-03) is negative and indicates that higher Institutional ownership is associated with lower abnormal volatility. But this relationship is not statistically significant ($p>0.05$). Insiders/Stakeholders (2.6289e-03) is positive and statistically significant. This means that the higher insider/stakeholder ownership is associated with an increase in abnormal volatility. This relationship suggests that insider activities may lead to greater market volatility. Market capitalization (3.1798e-09) is positive but not significant. This means that market capitalization does not have a significant impact on abnormal volatility. The R-squared value is very low which means that the model explains almost zero of the variability in abnormal volatility. The F-statistic is also not significant ($P>0.05$), suggesting that the model does not provide a better fit than a model with predictors.
The regression analysis indicates that only number of estimates and insider/stakeholder ownership show significant relationship with abnormal volatility. Forecast dispersion, total debt/equity ratio, institutional ownership and market capitalization do not have an impact on abnormal volatility. Given the low explanatory power, it suggests that other factors that are not included in the model may influence the abnormal volatility.

To account for the possibility of correlations within entities over time, the clustered standard errors are used. This approach improves reliability but correctly estimating the standard error in the presence of within-entity correlation.

t test of coefficients:

Variable	Estimate	Std. Error	t value	Pr(>|t|)	Significance
Intercept	0.00075945	0.00066159	1.1479	0.25102	
Forecast Dispersion	-0.00019611	0.00024962	0.7856	0.43209	
Number of Estimates	0.000096387	0.000014887	6.4747	9.76E-11	***
Total Debt/Equity (%)	-1.0043E-08	3.9221E-08	0.2561	0.79791	
Institutional Ownership	-0.0014597	0.0010183	1.4335	0.15174	
Insiders / Stakeholders	0.0026289	0.0011799	2.2281	0.02589	*
Market Cap	3.1798E-09	7.6438E-09	0.416	0.67742	

The intercept is not significant and indicates that when all predictions are zero, the abnormal volatility is not different from zero. The coefficient for forecast dispersion (-0.00019611) with p-value (0.43209) indicates it is not significant. This means that the forecast dispersion does not have a significant impact on abnormal vol atility. The coefficient for number of estimates (0.000096387) and is statistically significant. This means that an increase in the number of estimates is associated with a small but significant increase in abnormal volatilit y. The debt/equity coefficient (-0.000000010043) with a p-value (0.79791) indicates that it is not significant and suggests that total debt/equity ratio does not have impact on abnormal volatility. The coefficient for instit utional ownership (-0.0014597) with p-value (0.15174) is not significant and suggests that institutional owne rship does not have an impact on abnormal volatility. Insider/stakeholders coefficient (0.0026289) and a p-va lue less than 0.05 implies that an increase in insider/stakeholder ownership is linked with an increase in abno rmal volatility. Market caps coefficient (0.0000000031798) and a p-value (0.67742) indicates that it is not sta tistically significant and suggests that market capitalization does not have an impact on abnormal volatility. The clustered standard error regression shows that of all the variables tested, the number of estimates and ins ider/stakeholder ownership does affect abnormal volatility.

Breusch-Pagan

Now we use the Breusch-Pagan test for homoscedasticity. This test tries to figure out if the variance of the re siduals in the regression is constant with the independent variable. It is one of the assumptions for the OLS re gression not to have heteroscedasticity (non-constant variance), this could lead to biased estimates.

The Breusch-Pagan test (37.834) with a p-value (7.704e-10) shows strong evidence against the null hypothes is of homoscedasticity, this is shows by the low p-value.

Test	BP Statistic	Degrees of Freedom	p-value
Studentized Breusch-Pagan Test	37.834	1	7.7E-10

Residuals vs. Fitted

The residuals vs. fitted values plot is a diagnostic tool used to show the assumptions of linear regression. It h elps to identify non-linearity, heteroscedasticity, and outliers. The plot shows the residuals on the y-axis and t he fitted values on the x-axis. This plot includes all variables.

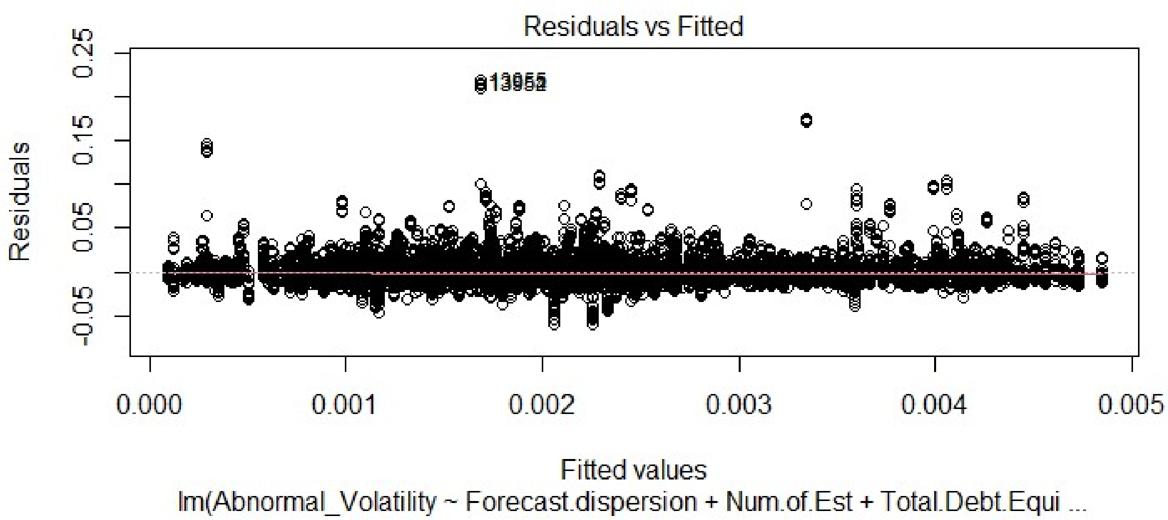

Figure 14 Residuals vs Fitted values on all the variables.

The plots residuals seem to be scattered randomly around the horizontal line at zero. This is a positive sign as it suggests that there is no pattern or trend, which indicates that the linearity assumption of the regression mo del might be met. However, earlier scatter plots of the dependent variable against all the other variables, indi cated non-linearity. This suggests that while the residuals vs. fitted values plot does not show clear patterns, i t does not fully capture all non-linear relationships. Therefore, it is important to acknowledge that the residua ls plot alone does not guarantee that the linearity assumption is not violated. The residuals seem to have cons tant variance across all fitted values, indicating heteroskedasticity is not an issue. This supports the use of W

LS to solve any remaining heteroskedasticity. There are a few points that lie far from the rest of the data. The se outliers need to be examined as they could influence the regression results.

While the residuals vs. fitted values plot suggests that the residuals are randomly distributed without patterns, our earlier diagnostics indicated potential non-linearity in the relationships between the dependent variable a nd some independent variables. This inconsistency shows the importance of using robust regression methods, such as WLS and clustered standard errors, to solve the impact of non-linearity and heteroscedasticity on our regression. To solve the issues, the WLS should be used to adjust for any heteroscedasticity that is found, giv ing different weights to different observations based on their variance. The clustered standard errors should b e used. Clustered standard errors to account for potential correlations within groups such as stocks or time pe riods, to make more reliable standard error estimates.

Weighted Least Squares

Heteroscedasticity violates the assumptions and therefore must be addressed. To solve this issue, we will use a WLS (Weighted Least Squares). OLS assumes that all observations have homoscedasticity. When this assu mption is violated, as we found out in the Breusch-Pagan test, using the WLS regression gives different weig hts to different data points based on the variance of their residuals. By doing this we lower the heteroscedasti c issues. The observation is weighted based on dividing1 by its variance, so an observation with a high varia nce leads to a small weight. This assumes that the variability of the residuals increases and the fitted value fr om the model gets larger. By adding more weight to the observations with smaller residuals and less weight t o those with larger, it helps evening out.

Statistic	Value
Min	-7.2863
1Q	-0.7958
Median	-0.2685
3Q	0.4133
Max	27.1739

Variable	Estimate	Std. Error	t value	Pr(>|t|)	Significance
Intercept	0.001125	0.0007451	1.509	0.1312	
Forecast Dispersion	-0.0003225	0.0001766	-1.826	0.0679	.
Number of Estimates	0.00008452	0.00001425	5.929	3.11E-09	***
Total Debt/Equity	-6.872E-09	2.731E-08	-0.252	0.8013	
Institutional Ownership	-0.001701	0.001149	-1.48	0.1389	
Insiders / Stakeholders	0.00264	0.001566	1.685	0.0919	.
Market Cap	5.179E-09	7.623E-09	0.679	0.4969	

Metric	Value
Residual Standard Error	1.643
Degrees of Freedom	17212
Multiple R-squared	0.004206
Adjusted R-squared	0.003859

The intercept is not significant at the 0.05 level, this means that when all the independent variables are zero, t he abnormal volatility is not different from zero. Forecast dispersion (-3.225e-04) is not significant (p=0.067 9). The negative coefficient means that as the forecast dispersion increases, abnormal volatility decreases a s mall bit. But this is not significant, indicating that this relationship is not robust. The number of estimates (8. 452e-05) is significant and the positive coefficient indicates that an increase in the number of estimates is ass ociated with an increase in abnormal volatility. This means that more analyst coverage is associated with hig her volatility around earnings announcements. The debt/equity ratio (-6.872e-09) is not significant, which me ans that variations in debt/equity do not have an impact on abnormal volatility. Institutional ownership (-1.70 1e-03) is not significant at the 0.05 level, but the negative coefficient suggests that a higher institutional own ership is associated with lower abnormal volatility. This is not significant enough to confirm a relationship. I nsiders/stakeholders (2.640e-03) are significant at the 0.1 level (p=0.0919). The positive coefficient suggests that greater insider/stakeholder ownership may be associated with higher abnormal volatility. The market cap italization (5.179e-09) is not significant and indicates that the size of companies does not impact the abnorma l volatility. The overall fit of the model suggests that the low R-squared indicates that the model explains onl y a small percentage (0.42%) of the variance in abnormal volatility. This suggests that other factors not inclu ded in the model may be a better choice for variables. The WLS regression model suggests that the number o f estimates shows a significant positive relationship with abnormal volatility, the other variables such as fore cast dispersion and insider/stakeholder ownership show even higher significance. The overall low R-squared suggests the model captures only a small percentage of the variability in abnormal volatility.

Multicollinearity

Another OLS assumption is that the data should not have multicollinearity. This means that the two or more predictor variables that er very correlated, lead to increased standard errors in the coefficients, making them unstable. This can also reduce the precision of the coefficients, which disturbs the significance of the predict ors. If we can make sure the regression model doesn't have multicollinearity, this can make our regression m odel more reliable and valid. We check for multicollinearity by using a VIF test in r studio. If a VIF value is greater than 10 indicates significant multicollinearity that could disturb the validity of the model by increasin g the variance of the coefficient. The VIF values for all the variables were under 10. This means there is no c oncern of level of multicollinearity among the predictors.

	Value
Forecast_dispersion	1.028535
Num_of_Est	1.610434
Total_Debt_Equity	1.009661
Institutional_Ownership	1.651031
Insiders_Stakeholders	1.254599
Market_Cap	1.100003

Q-Q plot

The Q-Q plot for normality of residuals tries to show if the residuals follow a normal distribution by compari ng the quantiles of the residuals to the quantiles of a normal distribution. It follows mostly the line in the mid dle portion of the plot, but there is deviation in the right tail where the residuals deviate upward. This indicat es a heavier tail than normal distribution suggesting non-normality. This can affect the reliability of the hypot hesis.

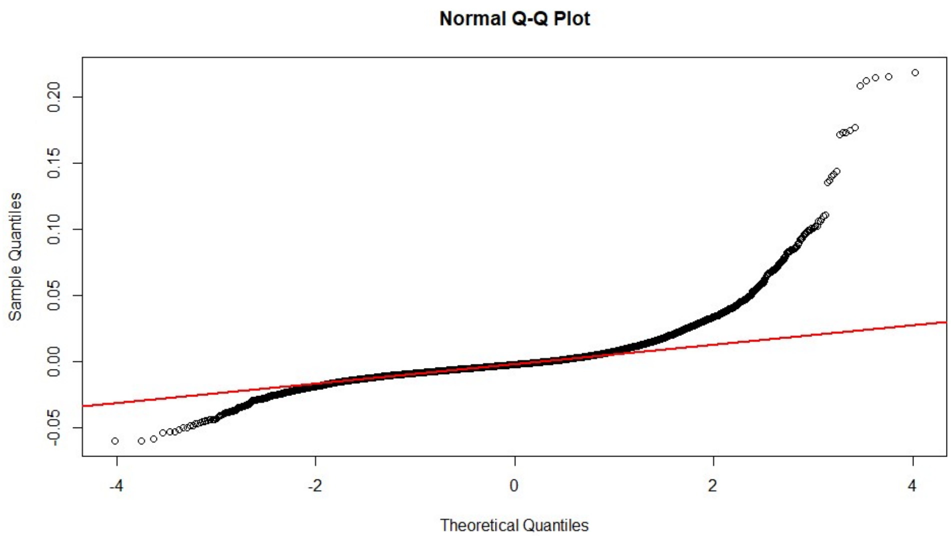

Figure 15 Normal Q-Q plot for normality of residuals.

Anderson-Darling

We also do an Anderson-Darling test for asses the normality of the residuals for the OLS model. This test evaluates whether the residuals of the model follow normal distribution, normality of this is necessary for having a correct hypothesis testing, this ensures the correct p-values and confidence intervals. The results of the Andersson-Darling test (A=822.7 and P-value <2.2.e-16) shows very low P-value and indicates that the residuals does not follow normal distribution, which violates on of the OLS assumptions about non-normality. To solve this problem, we use a robust regression method that is less sensitive to outliers and non-normal distributed residuals. This regression does not normalize the residuals but minimizes the influence of outliers, that may reduce the skewness and kurtosis the causes non-normal distribution. Overall, this does not solve the distribution problem but improves the overall robustness.

Statistic	Residual Value
Min	-0.0584499
1Q	-0.0048963
Median	-0.0005444
3Q	0.0049798
Max	0.2204697

Variable	Value	Std. Error	t value
Intercept	0.0691	0.0005	149.3457
Forecast Dispersion	0.0006	0.0001	-5.7307
Number of Estimates	0	0	4.5679
Total Debt/Equity	0	0	-2.3292
Institutional Ownership	0.0027	0.0007	-3.6795
Insiders/Stakeholders	0.0002	0.001	-0.1625
Market Cap	0	0	0.1828

Metric	Value
Residual Standard Error	0.007313
Degrees of Freedom	17212

Intercept (0.0691) is significant with a high t-value (149.2457), this means when all the predictors are zero the abnormal volatility is (0.0691). Forecast dispersion (-0.0006) is significant (t = -5.7307) suggesting that an increase in forecast dispersion makes a small decrease in abnormal volatility. The reason for this can be that higher dispersion which means higher uncertainty could make investors more conservative and leads to lover volatility. The number of estimates (0.0000) indicating no effect on abnormal volatility, but the t-value (4.5679) is significant which can indicate very small effects that are not visible. Total debt equity (0.0000) this also implies no impact on abnormal volatility, the t.-value here is also significant but only means there could be very small effects on the model. Institutional ownership (-0.0027) a significant effect (t = -3.6795) implies th at stocks with higher institutional ownership may experience less volatility, this is maybe due to a more disciplined investing strategi. Insider stakeholders (-0.0002) are not significant (t = -0.1625), this variable does not have effect on abnormal volatility. Market capitalization (0.0000) is not significant (t=0.1828), this indicates that the variables do not have any effect on abnormal volatility. Residual Standard Error (0.007313) indicates the general difference between the regression line and the observed values. A lower value means that predictions are close to the actual data points. VIF scores are below 5, this indicates that there is no multicollinearity among the predictors. This model shows a significant small negative relationship between forecast dispersion and abnormal volatility, this means that a forecast uncertainty increase will make abnormal volatility slightly decrease. Institutional ownership also significantly reduces abnormal volatility, this can mean more stable investment decisions from these owners. This is important to consider when predicting stock market volatility.

Dummy variable model 1

Here we are trying to make an analysis of the immediate effects of earnings announcements on abnormal vol atility. The core of this research is to capture the immediate effects of abnormal volatility following the earni ngs announcement. This is done by using a dummy variable for days 0 and 1, aimed at isolating the effects of these event days.

Variable	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	0.000557	0.0007877	0.707	0.479502
Forecast Dispersion	-0.0002246	0.000192	-1.17	0.242105
Day Dummy	0.002027	0.002491	0.814	0.415734
Number of Estimates	0.0000895	0.00001459	6.135	8.69E-10
Total Debt/Equity	-4.37E-08	3.136E-08	-1.393	0.163555
Institutional Ownership	-0.001951	0.001233	-1.583	0.113506
Insiders/Stakeholders	0.003095	0.001627	1.902	0.057141
Market Cap	4.392E-10	8.038E-09	0.055	0.956424
Forecast Dispersion:Day Dummy	0.0002856	0.0006072	0.47	0.638187
Day Dummy:Number of Estimates	0.00006882	0.00004613	1.492	0.135762
Day Dummy: Total Debt/Equity	3.366E-07	9.918E-08	3.394	0.000692
Day Dummy:Institutional Ownership	0.00491	0.003899	1.259	0.207898
Day Dummy:Insiders/Stakeholders	-0.004667	0.005146	-0.907	0.364474
Day Dummy:Market Cap	2.743E-08	2.542E-08	1.079	0.280468

Metric	Value
Residual Standard Error	0.01346
Degrees of Freedom	17205
Multiple R-squared	0.02378
Adjusted R-squared	0.02304
F-statistic	32.23
F-statistic DF	13 and 17205
p-value	< 2.2e-16

The 1st and 3rd quartiles of the residuals (-0.006513 and 0.003419) indicate that 50 percent of the residuals li e within this range, suggesting that most prediction errors of the model are small. The median of the residual s is near zero, indicating that the model's predictions are centered around the actual values, which is a positiv e sign for the model's reliability. The intercept coefficient (Estimate: 5.570e-04, Std. Error: 7.877e-04) is not statistically significant (p = 0.479502), indicating that when all predictors are zero, the baseline level of abno rmal volatility is not significantly different from zero. The forecast dispersion (Estimate: -2.246e-04, Std. Err or: 1.920e-04) is not statistically significant (p = 0.242105), suggesting that forecast dispersion does not have a significant effect on abnormal volatility on non-event days. The dummy variable for event days (Estimate:

2.027e-03, Std. Error: 2.491e-03) is also not statistically significant ($p = 0.415734$), indicating that the event day itself does not significantly affect abnormal volatility. The number of estimates (Estimate: 8.950e-05, St d. Error: 1.459e-05) is statistically significant ($p = 8.69e-10$), showing that the number of estimates positivel y impacts abnormal volatility on non-event days. The total debt/equity ratio (Estimate: -4.370e-08, Std. Error : 3.136e-08) is not statistically significant ($p = 0.163555$), suggesting no impact on abnormal volatility on no n-event days. Institutional ownership (Estimate: -1.951e-03, Std. Error: 1.233e-03) is also not statistically sig nificant ($p = 0.113506$), indicating no significant impact on abnormal volatility on non-event days. The coeff icient for insiders/stakeholders (Estimate: 3.095e-03, Std. Error: 1.627e-03) is significant ($p = 0.057141$), sug gesting a potential positive effect on abnormal volatility on non-event days. Market capitalization (Estimate: 4.392e-10, Std. Error: 8.038e-09) is not statistically significant ($p = 0.956424$), indicating no impact on abnor mal volatility on non-event days. The relationship between forecast dispersion and the day dummy (Estimate : 2.856e-04, Std. Error: 6.072e-04) is not statistically significant ($p = 0.638187$), suggesting that forecast disp ersion does not significantly affect volatility specifically on days 0 and 1 compared to other days. The relatio nship between the day dummy and the number of estimates (Estimate: 6.882e-05, Std. Error: 4.613e-05) is al so not statistically significant ($p = 0.135762$), indicating no effect between the number of estimates and the e vent day dummy on abnormal volatility. The relationship between the day dummy and total debt/equity ratio (Estimate: 3.366e-07, Std. Error: 9.918e-08) is statistically significant ($p = 0.000692$), indicating that the rela tionship between the total debt/equity ratio and abnormal volatility changes on event days. The interaction be tween the day dummy and institutional ownership (Estimate: 4.910e-03, Std. Error: 3.899e-03) is not statistic ally significant ($p = 0.207898$), suggesting no relationship effect between institutional ownership and the eve nt day dummy on abnormal volatility. The relationship between the day dummy and insiders/stakeholders (E stimate: -4.667e-03, Std. Error: 5.146e-03) is not statistically significant ($p = 0.364474$), indicating no relatio nship effect between insiders/stakeholders and the event day dummy on abnormal volatility. The relationship between the day dummy and market capitalization (Estimate: 2.743e-08, Std. Error: 2.542e-08) is not statisti cally significant ($p = 0.280468$), suggesting no relationship effect between market capitalization and the even t day dummy on abnormal volatility. The residual standard error is 0.01346 on 17205 degrees of freedom. Th e multiple R-squared value of 0.02378 indicates that the model explains only about 2.4% of the variability in abnormal volatility, suggesting low explanatory power. The adjusted R-squared value of 0.02304 supports th is conclusion. The F-statistic of 32.23 ($p < 2.2e-16$) indicates that the overall model is statistically significant , meaning that at least one predictor variable has a significant relationship with abnormal volatility.

In summary, days 0 and 1 are associated with increased abnormal volatility, reflecting market reactions to ea rnings announcements. However, the model's low R-squared value indicates that there are other factors not c aptured by the model that influence abnormal volatility. The significant relationship between the total debt/e quity ratio and the event day dummy suggests that financial fundamentals impact volatility changes during e vent days. The non-significant relationship for other predictors suggests that most do not significantly impact abnormal volatility on these event days.

Dummy variable model 2

When studying the immediate effects of earnings announcements on abnormal volatility, a simpler model wa s used to focus directly on two variables: the dummy variable for event days (days 0 and 1) and the relations hip with forecast dispersion. The reason for this approach is to make an analysis of the direct impact of earni ngs announcements without the noise from other variables. Using just these two variables allows for a more s traightforward approach for regression. By excluding other variables, we avoid the dilution of these effects, making it easier to measure any observed changes in market volatility directly to the earnings announcement s and the expectations by forecast dispersion.

Statistic	Residual Value
Min	-0.059334
1Q	-0.006513
Median	-0.002084
3Q	0.003419
Max	0.219191

Variable	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	0.004324	0.0004552	9.499	< 2e-16
Forecast Dispersion	0.0005246	0.0005998	0.875	0.382
Day Dummy	0.005782	0.0004254	13.591	< 2e-16

Metric	Value
Residual Standard Error	0.01348
Degrees of Freedom	17211
Multiple R-squared	0.01998
Adjusted R-squared	0.01958
F-statistic	50.12
F-statistic DF	7 and 17211
p-value	< 2.2e-16

The coefficient for the day dummy variable (Estimate: 5.782e-03, Std. Error: 4.254e-04) is highly significant ($p < 2e-16$). This indicates that abnormal volatility increases significantly on days 0 and 1, reflecting market reactions immediately after earnings announcements. This result aligns with the expectation that earnings an nouncements will increase market volatility. The relationship between forecast dispersion and the day dumm y is not statistically significant ($p = 0.382$). This suggests that forecast dispersion does not significantly impa ct abnormal volatility specifically on event days compared to other days. When comparing these results with the full model, several points stand out. In the simpler model, the day dummy variable is highly significant, i ndicating that abnormal volatility is significantly higher on event days. This contrasts with the full model, wh ere the day dummy was not significant, suggesting that the inclusion of additional variables might have dilut ed the effect of event days. This highlights that the simpler model is better at capturing the immediate impact of event days on abnormal volatility. The simpler model has a low R-squared value (0.01998), indicating that it explains only about 2% of the variability in abnormal volatility. This is expected given the model's simplici ty. However, this simplicity allows for a clearer interpretation of the key variable of interest, the day dummy.

Analysis Resume

In this analysis, there have been used many statistical methods and models to investigate the market behavior around earnings announcements. The analysis began by identifying the distribution of the individual' stock p rices and the S&P 500. There was found right skew and leptokurtic patterns in stock prices, which suggest ou tliers and extreme values. In contrast, the S&P 500 had a more symmetrical, mesokurtic distribution, supporti ng its role as a more stable and predictable benchmark. The study further calculated the beta coefficient from a 40-day control period to adjust abnormal stock returns around earnings announcements, which tried to isola te the impact of these events. Even when there was an expected increase in volatility around earnings announ cements, the use of historical volatility models like GARCH would have provided a deeper insight into mark et reactions and giving a more precis anticipation of expected behaviors to accurately evaluate the impact of earnings announcements. So, we could have used GARCH expected volatility to make the abnormal volatilit y even more accurate. We have so far only calculated the abnormal volatility by using the 40 days volatility a s expected volatility and these 40 periods do not account for event dates where the volatility might by higher. Through graphical analysis, it was confirmed that volatility peaks around earnings announcements across mo st of the stock and index prices. The approach not only supported the significant impact of these announceme nts on market volatility but also illustrated the markets' immediate reaction, anticipation, and post earnings st abilization. The correlation matrix analysis gave information about the weak relationship between abnormal volatility and other variables. But correlations were identified between institutional ownership and insider sta keholders, with institutional ownership inversely related to insider holdings and directly to larger, less risky c ompanies. The application of the Augmented Dickey-Fuller test confirmed the stationarity of all key variable s, supporting the credibility of the regression model. But there were challenges such as autocorrelation and h eteroscedasticity were identified. Therefore, the implementation of robust regression and clustered standard e rrors was made to reduce these issues and improve the model. The regression analysis showed a limited influ ence of forecast dispersion, total debt/equity ratio, and market capitalization on abnormal volatility while the number of estimates and insider/stakeholder directly influence market volatility. The Durbin-Watson test indi cated significant positive autocorrelation in the residuals, with a statistic of 0.454, indicating that the residual s from one observation to the next are correlated. This autocorrelation violates one of the key assumptions of ordinary least squares (OLS) regression, which requires no autocorrelation among residuals, as it can lead to biased parameter estimates and unreliable hypothesis tests. To address this issue, robust standard errors using the "sandwich" package in R were employed to adjust for autocorrelation, improving the reliability of the model despite this violation. The analysis of residuals showed no clear trends supporting the model's reliability, although some points suggested influential outliers. The regression analysis, with adjustments for autocorrelation, tested relationships between abnormal volatility and several predictors. Significant relationships were observed only for the number of estimates and insider/stakeholder ownership, indicating that these factors are associated with changes in abnormal volatility. Other variables like forecast dispersion, total debt/equity ratio, institutional ownership, and market capitalization showed no significant impact. The Breusch-Pagan test for homoscedasticity indicated a strong presence of heteroscedasticity, suggesting that the variance of residuals is not constant. This violation of one of the OLS assumptions might lead to biased estimates. The Q-Q plot for the normality of residuals showed deviations, especially in the right tail, indicating a heavier than normal distribution, which could compromise the reliability of hypothesis tests in the model. The residuals vs. fitted values plot, used as a diagnostic tool in linear regression, shows the residuals scattered around the zero line, suggesting no patterns or trends. This random distribution is a positive indication that the model might meet the linearity assumption. However, this observation contrasts with earlier scatter plots that indicated non-linearity between the dependent and independent variables, showing that the residuals plot alone cannot fully dismiss the possibility of non-linearity. Furthermore, the plot suggests a constant variance across fitted values, indicating no major issues with heteroskedasticity. A few outliers identified in the plot could impact the regression outcomes and should be carefully considered. Given these findings, we used robust regression methods like Weighted Least Squares (WLS) and clustered standard errors is recommended. WLS can adjust for any heteroskedasticity by assigning different weights to observations based on variance, while clustered standard errors can handle potential correlations within groups, such as across different stocks or over time periods, providing more accurate and reliable standard error estimates. To solve heteroscedasticity found by the Breusch-Pagan test, we did a Weighted Least Squares (WLS) regression, which uses different weights to data points based on the variance of their residuals. This approach minimizes the impact of heteroscedasticity by giving less weight to observations with higher variance, thereby balancing the influence of each point on the regression results. The results from the WLS regression show that most predictors are not significantly correlated with abnormal volatility at traditional significance levels. But the number of estimates showed a significant positive correlation, suggesting that increased analyst coverage correlates with higher volatility, especially around earnings announcements. Insider/stakeholder ownership also showed a positive correlation with abnormal volatility, though less significantly. But the overall model has a low R-squared value (0.42%), indicating it explains only a small percentage of the variance in abnormal volatility. This suggests that additional variables not included in the model might be more influential in explaining abnormal volatility. The use of WLS helped in solving heteroscedasticity but also indicated the need for possibly expanding the model to capture more abnormal volatility effectively. The analysis of the regression model involved tests for multicollinearity and normality of residuals to ensure the reliability and validity of the results. The Variance Inflation Factor (VIF) tests showed no significant multicollinearity among predictors, indicating that the model coefficients are stable and precise. But the Anderson-Darling test showed that the residuals did not follow a normal distribution, suggesting a potential violation of OLS assumptions. To address the issue of non-normal residuals, a robust regression method was employed, improving the model's resilience to outliers and non-normality. This approach confirmed that forecast dispersion negatively impacts abnormal volatility, but just a small amount, implying that greater uncertainty in forecasts leads to lower volatility. But higher institutional ownership correlates with reduced abnormal volatility, suggesting that institutional investors contribute to market stability. Overall, the robust regression approach provided a more reliable analysis of the factors affecting abnormal volatility. We also tried to understand the immediate effects of earnings announcements on abnormal volatility, by making a dummy variable on day 0 and 1. This approach was meant to isolate and direct the impact of these event days on market behavior. The results indicated that while the overall model, had a variety of predictors from institutional ownership to market capitalization, indicated some relationships, the impact of the dummy variable itself on abnormal volatility was not statistically significant. This suggests that the earnings announcement, represented by days 0 and 1, does not automatically mean an increase in volatility when controlling for other variables. But the total debt/equity ratio and the event day dummy were significant, indicating that the financial conditions of a company might affect how its stock price reacts to earnings announcements. The number of estimates was a significant predictor, indicating that greater analyst attention might increase volatility, meaning higher market sensitivity. Other variables like forecast dispersion and the broader market capitalization did not have a direct impact on the volatility during these event days. The simpler model, focusing only on the dummy variable for event days and the forecast dispersion, showed that the day dummy variable was significantly impacted by abnormal volatility, meaning that event days are associated with increased market activity. Both models accounted for a relatively low percentage of the variability in abnormal volatility, as indicated by R-squared. This suggests that while certain factors like the timing of earnings announcements and the number of analyst estimates play a role, much of the volatility observed around these events can be caused by other factors or market conditions not captured by the variables used. The findings suggest that while earnings announcements can lead to volatility, their impact is reduced by some factors, from corporate financial health to other market conditions, all of which can be considered in future analyses.

Results compared with literature.

The Efficient Market Hypothesis (EMH) made by Fama suggests that stock prices reflect all available information, implying that markets are efficient in processing information. The literature explains three forms of EMH: weak, semi-strong, and strong. The results of this study challenge the semi-strong form of EMH, which suggests that prices adjust to public information immediately. The expected relationship between forecast dispersion and abnormal stock volatility was weak, suggesting that information may not be as integrated into stock prices as EMH suggests. Literature such as Lee, Mucklow, and Ready, and Easley, Hvidkjaer, and O'hara discurss that information asymmetry can cause significant market volatility, the difference or asymmetry of information access or understanding can delay its incorporation into market prices, leading to inefficiencies. This is linked with our observations where the weakness of forecast dispersion on stock volatility. This could suggest that other factors may cause this, like the possibility of external market shocks or complex, non-linear relationships. Regulatory like the Sarbanes-Oxley Act try to reduce information asymmetry and reduce volatility and improve market efficiency. Studies by Zhang, Shu, and Brenner confirm that such regulations can stabilize volatility, but their effectiveness is different across markets and firm sizes. This difference could explain the inconsistencies in this research findings about the impact of forecast dispersion on volatility, which confirms the influence of regulatory initiatives on market behavior. The literature also discusses how trading technologies such as algorithmic trading (Brogaard, Hendershott, and Riordan) may affect information incorporation into market prices. While these technologies are designed to improve market efficiency by reducing the time it takes for information to be reflected in stock prices, they may also contribute to intra-day volatility. This is shown in my analysis where the increased integration of information does not reduce volatility. Models like those developed by Kyle and Glosten-Milgrom offer understanding of how private information is processed in markets. These models suggest that informed trading leads to price adjustments that incorporate asymmetric information. This theoretical perspective supports this study's findings where the impact of information asymmetry on stock volatility is seen in a more subtle way than the direct correlations that is expected under EMH. The concept of Post-Earnings Announcement Drift (PEAD) further illustrates that information asymmetry can lead to longer adjustments in stock prices, in contrast to the immediate adjustments by strong forms of EMH. My study's observations on the weak impacts of earnings announcements support this, suggesting that different investor reactions and the time taken to process the information can lead to varied stock price behavior. While the EMH provides an understanding of market behaviors, this research shows the complexities and limitations of this hypothesis in fully explaining the information processing on the markets. The result from this study recommends more complex models that account for non-linearity, psychological factors, and the new trading technologies, offering a better view of market efficiency and volatility.

Alternative Methods

The analysis could benefit from including a broader range of variables. That could be macroeconomics factor s like GDP growth rates, inflations rates, interest rates and unemployment report releases. These variables ar e examples of what could help the model to get better results on the hypothesis about influence of abnormal volatility. The psychological and behavioral factors that affect the investor could also be investigated. Behavi or finance as indicator or if investors decision making process under uncertainty could be measured, this coul d also add value to the models. It could also be beneficial to make a comparative analysis across different ma rkets. Comparing the influence of variables on volatility in developed versus emerging markets or different a ssets classes, like bonds and commodities. Espacielly commodities, there could be some regulatory measures that could affect the market behavior. Using models like GARCH could be useful in this analysis. GARCH models are helpful at modeling financial time series data because they can capture the clusters of volatility often observed in financial markets. This would allow for a better estimation of expected volatility, especially during earnings announcements, providing a clearer understanding of how announcements influence market volatility. The expected volatility from the GARCH model could be used as expected volatility from the investors and thereby used in the abnormal volatility. The incorporation of market liquidity measures could provide insights into the market’s volatility spikes. Market liquidity, often reflected through the volume of trades or the bid-ask spread, can significantly influence how quickly prices adjust to new information, such as earnings announcements. Incorporating panel data analysis could help control individual heterogeneity and examine time-specific effects across entities. Cointegration analysis would be valuable for exploring long-term relationships between key economic indicators and market volatility. Using event study methodology could precisely measure the market's response to specific economic or news events, improving our understanding of markets volatility. Further to improve the model, could involve incorporating data on financial derivatives, such as options and futures, to investigate market expectations of volatility through measures like implied volatility. Integrating the Economic Policy Uncertainty Index could provide insights into how policy fluctuations impact market stability and investor behavior. Examining the effects of technological innovations in trading on markets, especially around earnings announcements, could also offer valuable perspectives on how modern trading practices influence market behavior.

Discussion on non-significant results

The lack of a significant relationship between forecast dispersion and abnormal stock volatility can be caused by many factors. Data limitations, such as the quality of the data, may restrict the model's ability to capture the effects of forecast dispersion on stock volatility. The exclusion of certain market segments or variables might be missing. If important predictors that influence volatility are not included in the model, this can lead to variable bias. Furthermore, the assumption of linearity in the relationship between forecast dispersion and stock volatility might not be valid. Financial markets often have complex, non-linear relationships that standard linear models cannot capture. External shocks or market trends, such as financial crises or significant regulatory changes during the period analyzed, might also be the reason. The non-significant coefficient for forecast dispersion suggests that with our dataset and model, forecast dispersion does not linearly predict abnormal stock volatility. This might imply that the market efficiently incorporates dispersed forecasts into stock prices, reflecting a form of market efficiency where available information is already priced in. It could also suggest that other factors, perhaps macroeconomic or firm-specific, play a more dominant role in influencing volatility during the earnings announcements period. Improving the model with additional variables or using different modeling approaches could also be an opportunity. Integrating non-linear models like Generalized Additive

Models (GAM) might give different aspects of the relationship between forecast dispersion and volatility. These approaches could help the potentially non-linear and complex nature of financial data, providing deeper insights into how information is processed in financial markets. The exploration of non-significant results is as important as significant findings, as it helps accuracy for the research approach. By solving these limitations and considering alternative explanations and model structures, future research can more effectively help explain the relationship of forecast dispersion and stock market volatility.

Reflections

The findings violate the assumption of high information asymmetry leading to increased stock volatility, some research has suggested. This observation could lead to reconsideration of market efficiency theories, especially the semi-strong form of the Efficient Market Hypothesis (EMH). According to the EMH, stock prices reflect all publicly available information at any given time, making it impossible to earn higher returns. However, the weak correlation between forecast dispersion (information asymmetry) and abnormal stock volatility suggests that the market may not always integrate information in a straightforward or immediate way. This could indicate that markets are not always efficient, or that their efficiency varies depending on specific conditions or types of information. If information asymmetry does not consistently lead to increased volatility, it suggests that other factors might be at play in helping the market to react to new information. This could involve behavioral finance, where investor psychology and decision-making processes play a role in how information is processed, which traditional market efficiency theories may not account for. For investors, understanding that not all forms of publicly available information, like analyst forecasts, can lead to impact on stock prices and help for more strategic investment decisions, especially around earnings announcements. Investors might benefit from a more traditional approach that considers a variety of factors, including analyst coverage, insider activities, and broader economic indicators, rather than relying on forecast dispersion. This research has robust statistical methods to solve issues like autocorrelation and heteroscedasticity, improving the reliability of the results. But to rely on linear models and some financial variables as measures for factors like information asymmetry could limit the accuracy of results. Linear assumptions may not capture complex, non-linear relationships in financial data. The study's focus on specific market conditions and data from a particular period may affect the results. Future research could expand on this by incorporating non-linear models, more diverse datasets, and real-time analysis to capture a broader range of market behaviors. The choice of a 40-day control period is randomly chosen and may not fully show normal market conditions. Market behavior can be influenced by many factors not captured by this timeframe and may be skewing the baseline against abnormal volatility. The approach that is used in this study may not capture significant market related events, this may not be captured in the 20 days window. This may cause not accounting for delayed market reactions.

Conclusion

H1 Testing:

The hypothesis that high information asymmetry, measured through analyst forecast dispersion, increases stock volatility was not supported. The analysis showed that forecast dispersion does not significantly predict abnormal stock volatility around earnings announcements. This suggests that the market might already incorporate dispersed forecasts into stock prices efficiently, or other factors play a higher role in the impact of forecast dispersion during the earnings announcements period. The investigation into the immediate impact of earnings announcements showed that stock volatility peaks around these events, confirming their significant influence on market behavior. The dummy variable for days 0 and 1 (immediate days surrounding the earnings announcements) did not significantly impact abnormal volatility when controlling for other variables.

H2 Testing:

Control variables such as market capitalization, the number of analysts, institutional ownership, insider stakes, and leverage were investigated for their influence on the relationship between information asymmetry and stock volatility. Results were mixed, with only the number of estimates and insider/stakeholder ownership showing significant relationships with market volatility. This suggests that while certain control variables impact volatility, many do not significantly affect the relationship between information asymmetry and volatility as hypothesized. The use of robust regression methods helped to solve issues like autocorrelation and heteroscedasticity in the data, improving the model's predictive power. But these models explained only a small bit of the variance in abnormal volatility, indicating the need for integrating better variables or more complex models to capture the market behavior fully.

Conclusion

The literature review discusses how information asymmetry affects stock markets volatility and the efficiency on the stock market. It discusses the Efficient Market Hypothesis and how information asymmetry within investors causes market inefficiencies in the market and increases volatility. The literature review finds evidence on information asymmetry as a significant factor in causing stock market volatility and its impact on market efficiency, which can be profitable for skilled investors. The distribution of the stock prices and S&P 500 prices shows the market behavior by statistical measuring skewness and kurtosis. The stock prices show a right-skewed leptokurtic distribution which means that the value is at lower levels with a long tail of high values, this means high probability of extreme values. The distribution of the stock prices gives high abnormal volatility due to the non-normal distribution. The S&P 500 are more symmetric with a mesokurtic distribution, which means less outliers. A 40-day control period is used as a baseline for normal stock behavior, which isolates the abnormal volatility more precis and gives a more firm-specific volatility. The correlation matrix shows that the abnormal volatility is independent of variables like forecast dispersion, number of estimates, total debt to equity ratio, institutional ownership, insider stakeholders, and market capitalization, indicating no lin ear relationships. But the institutional ownership and insider stakeholders have a negative correlation, which can be an indication of trade-off between these two. Institutional ownership is shown to be higher in larger c ompanies, this makes sense as they have a more stable capital structure, this is also shows be the positive cor relation with market capitalization and negative correlation with total debt/equity ratio. The analysis using ro bust standard errors, by the 'sandwich' package in R, it was made to correct standard errors for autocorrelatio n and non-linearity problems. This allowed for a more reliable estimation. The regression results showed that most predictors, including forecast dispersion, total debt/equity ratio, institutional ownership, and market cap italization, did not significantly influence abnormal volatility. But the number of estimates and insider/stakeh older ownership were significant predictors, suggesting that higher analyst coverage and insider ownership ar e associated with increased market volatility. The analysis also did a Weighted Least Squares (WLS) to solv e the problem of heteroscedasticity in the data. The regression results indicated that most variables, including forecast dispersion, total debt/equity, institutional ownership, and market cap, did not significantly affect abn ormal volatility. But the number of estimates showed a significant positive relationship. Increased analyst co verage correlates with higher market volatility. Insiders/stakeholders have a nearly significant positive impac t on volatility. The Anderson-Darling test used to analyze the normality of the residuals in an OLS model sho ws significant deviation from normality, suggesting a violation of one of the key OLS assumptions necessary for reliable hypothesis testing. To solve this problem, robust regression methods were used to reduce the infl uence of outliers and reduce non-normal distribution effects like skewness and kurtosis. The intercept and for ecast dispersion showed impacts on abnormal volatility, with forecast dispersion indicating that increased un certainty (higher dispersion) makes a small decrease in abnormal volatility and may be due to more conserva tive behavior by investors. Institutional ownership was found to significantly decrease volatility, suggesting t hat institutional investors might help to stability in the stock prices they are involved with due to conservativ e investment strategies. The analysis also did a regression with a dummy variable and by this, tried to analyz e the immediate effects of earnings announcements on abnormal volatility. Two models were made, and each had dummy variables which represent day 0 and 1. Model 1 had many variables like forecast dispersion, the number of estimates, total debt/equity ratio, institutional ownership, insiders/stakeholders, and market capital ization. Most of the predictors had no significant impact on abnormal volatility on the announcement days, e xpect total/equity. This could maybe suggest that financial fundamentals may influence market reactions on a nnouncement days. Model 2 is a simpler version of the model, it only uses the dummy variable for event day s and forecast dispersions impact on abnormal volatility. This model found a significant impact of the dumm y variable where an increase in abnormal volatility on day 0 and 1. Which means that the earnings announce ment has a great impact on abnormal volatility, but this was also expected. Besides this there was no relation ship between forecast dispersion and abnormal volatility. All the models had a low R-square. This suggests t hat even when we use different variables and different statistical methods, the models cannot measure all the variance in abnormal market volatility. This low explanatory power indicates that there are other factors affe

cting the market that are not captured by our variables. The analysis needs to use a broader set of variables to
improve the understanding of factors influencing abnormal volatility. This may include macroeconomic indic
ators like GDP growth, inflation rates, interest rates, and unemployment numbers. Additionally, behavioral fi
nance could offer a better understanding of how psychological factors influence investor decisions under unc
ertainty. The suggestion to use models like GARCH to better capture and estimate volatility, during earnings
announcements, could be a useful approach to calculate the abnormal volatility. It could be a good idea to us
e the impact of market liquidity and suing methods for panel data analysis and cointegration. The discussion
also mentions a need to improve models by using variables like financial derivatives and the Economic Polic
y Uncertainty Index to better understand investor behavior and market reactions to policy changes. Technolo
gical innovations in trading and their impacts on the market are also relevant for this kind of study. The analy
sis suggests that the limitations in data quality, model assumptions, and the exclusion of relevant variables co
uld be leading to biased results of the effects of forecast dispersion on volatility. We discussed the benefits of
using non-linear models. The findings of this analysis also gave a perspective on market efficiency. The resul
ts challenge the hypothesis that information asymmetry leads to increased market volatility. This is the oppos
ite to what might be expected under the semi-strong form of the Efficient Market Hypothesis, which states th
at all publicly available information is already reflected in stock prices, our results did not show a significant
increase in volatility due to information asymmetry. This could indicate that the markets may be more efficie
nt than thought and reflects available information in stock prices. This suggests that markets might be good a
t maintaining efficiency even when information between investors is uneven. To end the conclusion: the imp
act of forecast dispersion on abnormal volatility was not found in this analysis.

Bibliography

Albert S. Kyle. (1985). *Continuos Auctions and Insider Trading*.

Ball, R. , & B. P. (n.d.). *An empirical evaluation of accounting income numbers.*

Bernard and Thomas. (1989). *Post-earnings announcement drift*.

Cameron Truong, C. C. Y. C. (n.d.). *The options market response to accounting earnings announcements*.

Charles M. C. Lee, B. M. and M. J. R. (n.d.). *Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis*.

Chongyu Dang a, Z. (Frank) L. b, C. Y. (n.d.). *Measuring firm size in empirical corporate finance*.

David Easley, S. H. M. O. (2002). *Is Information Risk a Determinant of Asset Returns?*

David Hirshleifer. (n.d.). *Investor Psychology and Asset Pricing*.

D.B Owen. (n.d.). *Goodness-of-Fit Techniques* .

Dean Diavatopoulos a, J. S. D. b, A. F. D. R. P. (n.d.). *The information content of implied skewness and kurtosis changes prior to earnings announcements for stock and option returns*.

Elyas Elyasiani, J. J. J. (2007). *Institutional ownership stability and BHC performance*.

Eugene F. Fama. (1970). *Efficient Capital Markets*.

Hong & Stein. (1999). *Market underreaction and overreaction*.

Hull, J. C. (2012). *Options, Futures, and Other Derivatives*.

James D. Hamilton. (1994). *Time Series Analysis*. 501–503.

Jeffery S. Abarbanell, W. N. L. R. E. Verrecchia. (n.d.). *Analysts' forecasts as proxies for investor beliefs in empirical research*.

Jin E. Zhang. Jinghong Shu. Menachem Brenner. (n.d.). *The New Market for Volatility Trading* .

Jonathan Brogaard, T. H. R. R. (n.d.). *High-Frequency Trading and Price Discovery*.

Josef Fink. (2021). *A review of the Post-Earnings-Announcement Drift*.

Kevin K. Li and Haifeng You. (n.d.). *What is the value of sell-side analysts? Evidence from coverage initiation* .

Lawrence R. Glosten, P. R. M. (n.d.). *Bid, ask and transaction prices in a specialist market with heterogeneously informed traders*.

Lei Qin, W. X. W. Y. Z. (2020). *Volatility spread and stock market response to earnings announcements*.

Mandelbrot, B. (n.d.). *The Variation of Certain Speculative Prices*. 394–419.

Marno Verbeek. (n.d.-a). *A guide to modern econometrics, Fifth Edition*. 299–301.

Marno Verbeek. (n.d.-b). *A guide to modern econometrics, Fifth Edition*. 301–304.

Marno Verbeek. (n.d.-c). *A guide to modern econometrics, Fifth Edition*. 6–13.

Marno Verbeek. (n.d.-d). *A guide to modern econometrics, Fifth Edition*. 9–12.

Marno Verbeek. (n.d.-e). *A guide to modern econometrics, Fifth Edition*.

Marno Verbeek. (n.d.-f). *A guide to modern econometrics, Fifth Edition*.

Marno Verbeek. (n.d.-g). *A guide to modern econometrics, Fifth Edition*.

Marno Verbeek. (n.d.-h). *A guide to modern econometrics, Fifth Edition*. 7–8.

Marno Verbeek. (n.d.-i). *A guide to modern econometrics, Fifth Edition*.

Marno Verbeek. (n.d.-j). *A guide to modern econometrics, Fifth Edition*.

Marno Verbeek. (n.d.-k). *A guide to modern econometrics, Fifth Edition* . 7–16.

Patrick Augustin a b, M. B. G. G. P. O. M. G. S. (n.d.). *Informed options strategies before corporate events*.

R. Shruti, M. T. (n.d.). *Foreign institutional ownership stability and stock price crash risk*.

Ruey S. Tsay. (n.d.). *Analysis of Financial Time Series, third edition*.

Sharpe, W. F. (n.d.). *Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*.

Shaun Bond, W. W. S. Z. (2023). *Seasonal patterns of earnings releases and post-earnings announcement drift*.

Ticker codes of 75 chosen stocks				
AAPL	ABBV	ACGL	ADM	AMD
AMZN	AVGO	BEN	BF.B	BIIB
BIO	BKR	BRK.B	BWA	BXP
CDW	CHRW	CMA	COST	CPB
CRM	CTSH	CZR	DD	DFS
DG	DVA	EA	EFX	EL
ETSY	FANG	FMC	FOXA	FRT
GNRC	GOOG	GOOGL	GPN	HAL
HAS	HD	IVZ	JNJ	JPM
KMI	LLY	MA	META	MHK
MKTX	MPWR	MRK	MRNA	MSFT
NCLH	NFLX	NVDA	ON	PAYC
PCG	PG	PNW	PPG	RCL
RHI	RL	ROK	TSLA	UNH
V	VRSK	XOM	XRAY	

Figure 16 Chosen stocks from the S&P 500.