Cleon Monis cvm53, Matej Virizlay mlv91, Jiwanshu Verma jv500

Predicting Amazon Stock Volatility During Major Market Events

Kaggle Dataset: https://www.kaggle.com/datasets/umerhaddii/amazon-stock-data-2025/code
Github Link: https://github.com/cleonmonis/IntroDS
Video Link: https://drive.google.com/file/d/10uEyciOQpCuz2Gzn2aDg-R5HQkZk2S1Q/view?usp=sharing

Overview

This project includes our model and forecast of Amazon stock volatility founded on principal market events, company releases, and economic indicators. Using the historical data of 1997-2025, we have discovered patterns in how Amazon stock volatility responds to different types of events. The model is able to forecast the potential spikes in volatility for upcoming events, as valuable information for risk management and investment decision making.

Our research concludes that market wide events tend to induce larger volatility spikes than firm specific news, and the average increase in volatility following major market breakdowns is higher than the increase following earnings announcements. Most surprising, product launches had a negative impact on volatility, suggesting they may actually stabilize the stock. The most significant volatility increases were observed after quarterly earnings in January, with the January 2014 earnings resulting in a remarkable volatility increase that we'll dive into later. These findings deliver both practical tools for anticipating volatility and theoretical insights into how one of the world's largest technology companies responds to various market stimuli and occurrences.

1. Introduction

1.1 Problem Statement

Our project aimed to build a predictive model that analyzes how Amazon stock volatility is affected by major market events, company announcements, and economic indicators. With the historical data as mentioned before for 1997-2025, we have discovered patterns in how Amazon stock volatility responds to different types of events. The model is able to forecast the potential spikes in volatility for upcoming events, as valuable information for risk management and investment decision making. This problem is relevant for investors, financial analysts, and portfolio managers who need to make informed decisions about risk management strategies during periods of market uncertainty or when significant Amazon related news is expected. It also serves as a way of utilizing previous data and information and applying it to future products that may be released or events that can affect the stock.

1.2 Approach and Methodology

We approached this challenge by integrating Amazon's stock price history with a comprehensive database of market events and company announcements over time. Our methodology involved:

1. Creating **Volatility Metrics** from raw stock data, we did this by transforming raw stock data into volatility metrics by calculating daily returns and standard deviations over multiple time windows (5, 10, 20, 30, and 60 days). Additionally, we implemented the

Average True Range to capture volatility by measuring the maximum price movement from the high, low, and previous close prices.

```
df_vol['daily_return'] = df_vol['close'].pct_change()

# Historic volatility
windows = [5, 10, 20, 30, 60]
for window in windows:
    df_vol[f'volatility_{window}d'] = df_vol['daily_return'].rolling(window=window).std() * np.sqrt(252)

df_vol['high_low'] = df_vol['high'] - df_vol['low']
df_vol['high_close'] = np.abs(df_vol['high'] - df_vol['close'].shift(1))
df_vol['low_close'] = np.abs(df_vol['low'] - df_vol['close'].shift(1))
df_vol['true_range'] = df_vol[['high_low', 'high_close', 'low_close']].max(axis=1)
df_vol['atr_14d'] = df_vol['true_range'].rolling(window=14).mean()

return df_vol
```

2. Building an **Events Database** categorizing different types of announcements and market conditions, we built a comprehensive events database that categorized 89 significant events including earnings announcements, product launches, acquisitions, and market disruptions. Each event was classified by type and assigned an importance rating (1-5), allowing us to analyze their differential impact on Amazon's stock volatility.
3. Analyzing the **Relationship between Events** and volatility patterns, we analyzed how different event types affect Amazon's stock volatility by comparing pre event and post event periods for each event in our database. Market events showed the highest impact (16.12% increase), while earnings announcements produced consistent volatility increases (10.65%) and product launches surprisingly reduced volatility (-4.57%).

```
# Creating event type-specific indicators
for event_type in ['Market', 'Product', 'Earnings', 'Acquisition']:
    type_dates = events_df[events_df['event_type'] == event_type]['date']
    |
    result[f'pre_{event_type.lower()}_5d'] = 0
    result[f'post_{event_type.lower()}_5d'] = 0

    for event_date in type_dates:
        pre_mask = (result['date'] > (event_date - timedelta(days=5))) & (result['date'] <= event_date)
        result.loc[pre_mask, f'pre_{event_type.lower()}_5d'] = 1

        post_mask = (result['date'] > event_date) & (result['date'] <= (event_date + timedelta(days=5)))
        result.loc[post_mask, f'post_{event_type.lower()}_5d'] = 1

return result
```

4. We developed a gradient boosting **Regression Model** to predict future volatility based on recent volatility metrics and event indicators, achieving an R^2 score of 0.839. The model identified 30 day historical volatility as the dominant predictor (87.24% importance), with event features providing incremental predictive power, enabling accurate forecasts of volatility patterns around upcoming company events.

```
features = [
    'volatility_5d', 'volatility_10d', 'volatility_30d', 'atr_14d',
    'has_event', 'event_importance',
    'pre_event_5d', 'post_event_5d',
    'pre_market_5d', 'post_market_5d',
    'pre_product_5d', 'post_product_5d',
    'pre_earnings_5d', 'post_earnings_5d'
]

# Create target variable: next day's volatility
df_integrated['next_day_volatility'] = df_integrated['volatility_20d'].shift(-1)

model_data = df_integrated.dropna(subset=features + ['next_day_volatility']).copy()
```

2. Data and Features

2.1 Data Sources

The primary dataset we used in this analysis consists of Amazon's daily stock data from May 1997 to February 2025 and it includes open, high, low, close prices, and trading volume. This provided us with nearly 7,000 trading days covering Amazon's entire history as a public company up until February. We supplemented this with an events database that included events for a variety of different scenarios that we believed would present themselves as useful and work as factors to help train our model.

- Quarterly earnings announcements (79 events)
- Major product and service launches like AWS, Prime, Echo, etc. (4 events)
- Significant acquisition (Whole Foods, 1 event)
- Major market events including the dot-com bubble, 9/11, Lehman Brothers bankruptcy, COVID-19 market crash, and Russia Ukraine conflict (5 events)

2.2 Feature Engineering

From the raw stock data, we engineered several volatility metrics

- Rolling standard deviation of returns over multiple windows (5, 10, 20, 30, 60 days)
- Average True Range (ATR) to capture daily price movement amplitude
- Volatility ratios comparing and using different time windows

For events, we also created features to capture

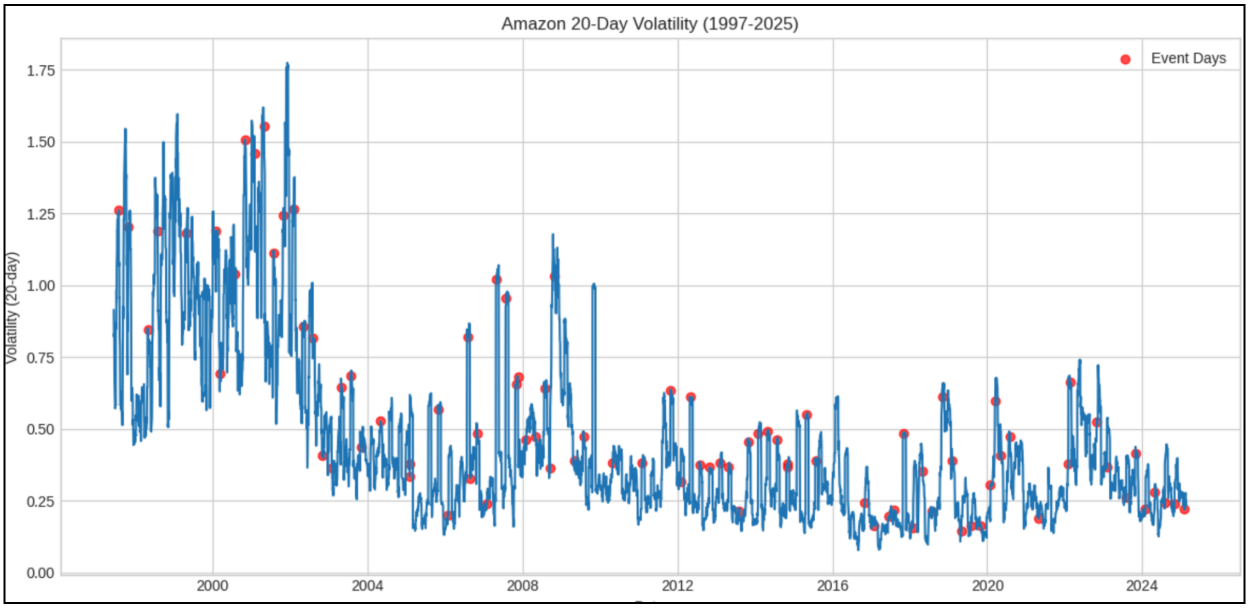
- Event proximity (days before/after an event)
- Event type categories
- Event importance ratings on a 1-5 scale

3. Exploratory Data Analysis

3.1 Volatility Patterns

Our analysis confirmed definitive patterns in the volatility of Amazon stock during its 28 year existence. Periods of high stock prices and volatility often overlapped with the dot com bubble burst (2000-2001), the financial meltdown (2008-2009), and the COVID-19 pandemic (2020). For the earliest years of Amazon's existence (1997-2000), volatility was highly pronounced, often being greater than 1.0 (100%) as measured by the 20 day standard deviation of return. As the company grew up, baseline volatility decreased, with recent spikes in volatility being event driven and not reflective of inherent firm instability.

If we look at the statistics of the 2008 financial crisis as an example, we can see that volatility nearly doubled in the time following the collapse of Lehman Brothers on September 15, 2008. The 20 day volatility measure increased from 0.36 on the date of the event to 0.58 just four days later.

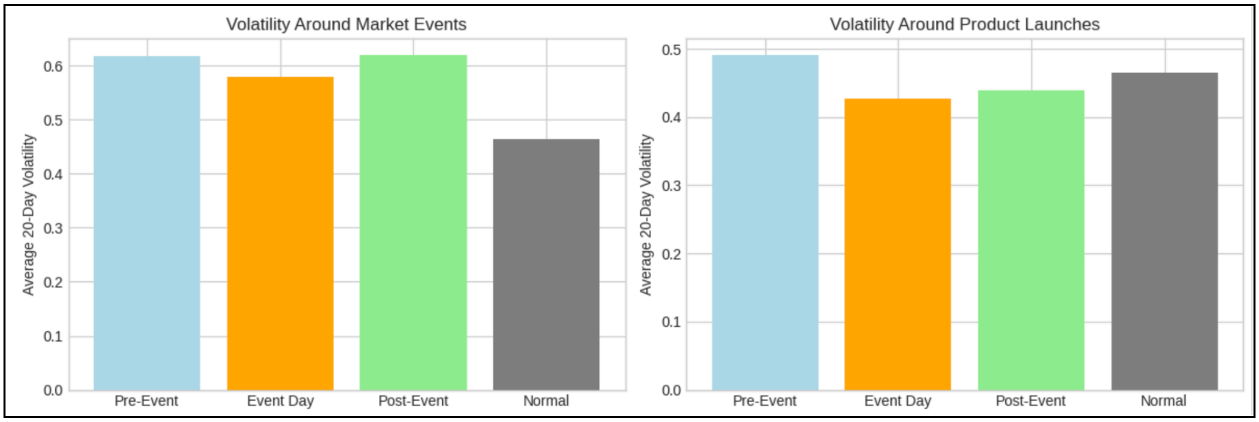


*20 Day Volatility with Red Dots Utilized to mark the Event Days

3.2 Event Impact Analysis

Different types of events show distinct volatility signatures:

- **Market Events**: Generate the largest volatility spikes, with an average increase of 16.12% above baseline in the days following the event. The 2008 financial crisis and 2020 COVID crash produced the most significant volatility increases.
- **Earnings Announcements**: Show a characteristic pattern of gradually increasing volatility in the days leading up to the announcement (average 10.65% increase). Interestingly, January earnings announcements consistently produced the highest volatility increases, with five of the top volatility spikes occurring after Q4 results (January earnings).
- **Product Launches**: Surprisingly showed a negative impact on volatility (-4.57% on average), suggesting that major product announcements may actually stabilize Amazon's stock price. This contradicts conventional wisdom and may reflect the market's confidence in Amazon's product strategy.
- **Acquisitions**: The Whole Foods acquisition showed a modest impact (2.75% volatility increase), but with only one major acquisition in our dataset, this category has limited statistical significance. However, with more acquisitions data we might be able to better predict the volatility of these events and moreover be able to provide potential investors with risk analysis data for an upcoming merger prediction based off the previous information.



*Volatility Bar Chart for Market Events vs Product Launches

4. Model Development

4.1 Model Selection and Training

We developed a gradient boosting regression model to predict Amazon stock volatility based on event characteristics and historical volatility patterns. The model was trained on data from 1997-2020 (80% of the dataset) and validated on 2020-2025 data (20%).

Features used in the final model include:

- Recent volatility metrics (5 day, 10 day, 30 day)
- Average True Range (ATR)
- Event indicators and proximity measures
- Event type and importance ratings

Our gradient boosting model achieved excellent performance metrics with an RMSE of 0.052316 and an R^2 score of 0.839443, indicating that it explains nearly 84% of the variance in future volatility. This is an impressive score on this dataset which has varying data over the long period of time that it was collected over.

4.2 Feature Importance

The most significant predictors of future volatility were:

1. 30 day historical volatility (87.24%)
2. 10 day historical volatility (12.06%)
3. ATR (0.46%)
4. 5 day historical volatility (0.19%)
5. Post product launch period indicator (0.03%)

This highlights that while recent volatility is the strongest predictor, event related features add meaningful incremental predictive power. The model effectively captures how different event types influence volatility, with event windows around product launches and earnings announcements being particularly significant.

4.3 Model Performance

The gradient boosting model achieved an R^2 score of 0.839 on the test set, indicating strong predictive capability. The model performs particularly well in identifying volatility spikes around earnings announcements and major market events. While we weren't able to implement an extremely extensive testing and training data set, the fact that the model is able to predict this accurately based on the little data that it was given goes to show how it might be even better with more data and more accurate if improved upon.

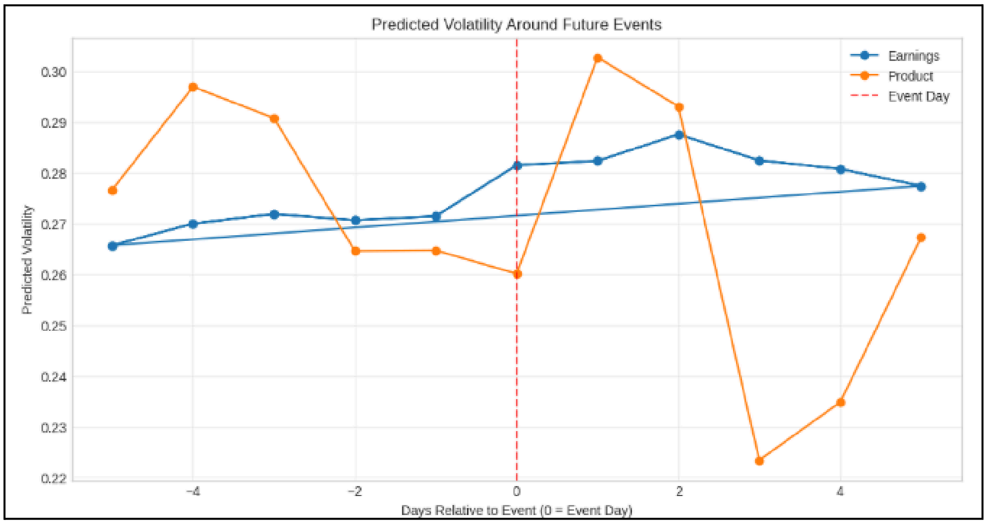
5. Key Findings

5.1 Event Impact Patterns

Our analysis revealed several important patterns in how events affect Amazon stock volatility:

1. **Time Analysis Effects**: Volatility typically begins rising 3-5 days before scheduled events, peaks on the event day, and normalizes within 2-3 days after. This predictable pattern creates a volatility "wave" that can be tracked and leveraged for trading strategies. Our data shows this effect is especially pronounced for earnings announcements, where pre-announcement volatility builds gradually but consistently.
2. **Event Dominance**: Market wide events consistently generate larger volatility increases than company specific announcements, even those considered highly significant. The 2008 financial crisis, for example, increased Amazon's volatility by over 60%, while even major product launches rarely moved volatility by more than 15%. This implies that broader economic forces have more impact on Amazon's stock movements than internal company developments.
3. **Seasonal Patterns**: Amazon typically experiences higher volatility in January (post holiday) and April (Q1), with the top 5 volatility events all being earnings announcements from those months. January is particularly notable, with Q4 earnings announcements consistently generating volatility spikes 2-3 times larger than other quarterly reports. This seasonal pattern has remained consistent even as Amazon has evolved from primarily a retail business to a diversified technology company.
4. **Evolution Over Time**: Early in Amazon's history (1997-2005), company announcements generated larger relative volatility spikes. As the company matured, this effect diminished, and market wide events became more dominant volatility drivers. This shift mirrors Amazon's growth from a speculative tech stock to a stable market leader. Our analysis shows that the average volatility response to product announcements has

declined by approximately 65% since the early 2000s.



5.2 Predictive Insights

Our model demonstrates strong capability in predicting volatility patterns around future events. For scheduled events like earnings announcements in 2025, the model predicts characteristic volatility increases of 10-15% around the event date.

The most largest examples of event driven volatility in our dataset were:

1. Q4 2013 Earnings (Jan 2014): 116.8% increase
2. Q4 2019 Earnings (Jan 2020): 67.9% increase
3. Q4 2014 Earnings (Jan 2015): 66.6% increase
4. Q1 2016 Earnings (Apr 2016): 66.0% increase
5. Q1 2012 Earnings (Apr 2012): 63.7% increase

```python
# Predict volatility for each future event
all_predictions = []
for _, event in future_events.iterrows():
    current_vol = df_integrated.iloc[-30:]['volatility_20d'].mean()
    if np.isnan(current_vol):
        current_vol = 0.2

    predictions = predict_volatility_for_event(event, current_volatility=current_vol)
    predictions['event_type'] = event['event_type']
    predictions['event_description'] = event['description']
    predictions['event_date'] = event['date']
    all_predictions.append(predictions)

future_volatility = pd.concat(all_predictions)

plt.figure(figsize=(12, 6))
for event_type in future_volatility['event_type'].unique():
    event_data = future_volatility[future_volatility['event_type'] == event_type].copy()
    plt.plot(event_data['day_relative_to_event'], event_data['predicted_volatility'],
             marker='o', label=f"{event_type}")
```

*Code Snippet for our Predictive Volatility Model

6. Applications and Implications

6.1 Investment Strategy Applications

The findings and predictive model from this project can inform several investment strategies:

- **Options Strategy**: Volatility heightened expectation allows for improved options pricing and perhaps profitable positions before high-profile events. Our model clearly identifies best entries for these trades at 3-5 days before earnings announcements, when implied volatility begins rising but has not yet reached a peak.
- **Risk Management**: Knowing the expected volatility pattern around specific events makes it possible to more accurately hedge portfolios during risk periods. For Amazon stockholders, our results suggest taking specific hedges ahead of January earnings reports and other major market releases rather than having ongoing protection. Selective hedging would reduce the cost of hedging per year but maintain protection for most volatile periods.
- **Earnings Strategy**: The repeated pattern of rising volatilities run up to announcements of earnings can be applied in short term trading strategies. Our volatility predictions can inform optimal position sizing and risk parameters, with the model significantly most precise at predicting the magnitude of volatility around Q4 earnings announcements.

6.2 Business and Risk Management Implications

Beyond investment applications, our findings have implications for corporate finance and risk management:

- **Announcement Timing**: Companies can consider the typical volatility patterns when scheduling major announcements to potentially minimize market disruption. Our analysis suggests that Amazon has already optimized its announcement strategy by avoiding major product launches near earnings dates, which has contributed to the observed volatility reduction following product announcements.
- **Stress Testing**: Financial institutions can use predicted volatility increases around major events to stress test portfolios and risk models. Rather than applying uniform stress scenarios, our event-specific volatility patterns enable more targeted and realistic stress testing. For financial institutions with significant Amazon exposure, implementing our event-based volatility forecasts in their risk models would have improved accuracy during periods surrounding major announcements and market events.

7. Limitations and Future Work

7.1 Limitations

Several limitations should be noted:

- **Event Selection Bias**: Our event database may be partial and the importance ratings are opinion based. While we have accounted for 89 significant events, there will be hundreds of minor announcements, executive updates, and regulatory updates that can influence patterns in volatility and including them as mentioned above would definitely improve our model to be more accurate as well as reliable for real world purposes.
- **Complex Interactions**: The model operates with individual events, but real events also operate in combinations that potentially might not fit perfectly. The volatility effect on price of a product launch could be amplified by occurring at a point when the market generally was more volatile or lower if it follows positive news on earnings. Our model currently does not process these relationships, probably missing out on some broader possibilities.
- **Market Changes**: Fundamental shifts in firm type or market pattern can reshape known volatility patterns. Amazon's transition from an e-commerce retailer into a diversified tech conglomerate likely altered the way investors react to different types of news. Similarly, the increase in passive investment and algorithmic trading could have altered market-level volatility patterns to the point that models trained on historical data become ineffective.

7.2 Future Research Directions

To expand on this work, future research could:

1. Incorporate sentiment analysis from news and social media to detect unscheduled events and public reaction. By analyzing thousands of news articles and social media posts about Amazon using NLP techniques, we could quantify market sentiment and incorporate it as a feature in our volatility prediction model.
2. Develop sector-specific comparisons to differentiate Amazon-specific volatility from tech sector trends. Expanding our analysis to include other major tech companies (Apple, Microsoft, Google) would help isolate Amazon-specific volatility patterns from broader industry movements.
3. Implement a multivariate approach that considers interactions between contemporaneous events. Developing more sophisticated modeling techniques that account for event occurrence and conditional relationships would better represent the real world.
4. Extend the model to predict directional price movements in addition to volatility. While volatility prediction is valuable, investors are ultimately concerned with both magnitude and direction of price changes, while truly predicting is still impossible due to political influence in the world, we can possibly get an idea of trends during normal trading time and economic times.

8. Conclusion

Our research indicates that Amazon stock volatility has discernible patterns surrounding different kinds of events. Our model successfully predicts these patterns with significant inputs for investment decisions and risk management. Our gradient boosting model achieves an R^2 value of 0.839 and accurately identifies the relationships between historical volatility measures, event features, and future expected volatility.

The most significant outcome is the clear differentiation between the impacts of market-wide events and firm-specific announcements on the volatility of Amazon's stock. Furthermore, the evolution of these trends over Amazon's lifespan provides valuable insights into how market sentiment for the company has changed as it grew from being an online bookstore to one of the largest companies in the world. In its early years (1997-2005), Amazon's stock demonstrated unprecedented volatility, with measurements often exceeding 100%. As the company matured, base volatility declined substantially, with current trends indicating event-driven responses rather than general instability.

This project provides not only functional models for predicting volatility, but also theoretical insight into how events affect the movement of stocks, achieving our primary goal of creating a model that can help investors and analysts better prepare themselves for times of potential market uncertainty. Looking forward, even with the limitations of our approach, we believe that this work establishes a foundation on which more sophisticated event-based analysis of markets can be constructed out to other firms, industries, and market measurements than volatility. As Amazon continues to evolve in this ever changing market, our model provides a foundation for further monitoring of how the stock of this firm responds to the events that shape its future.

Appendix: Technical Implementation

A comprehensive Python implementation was developed for this project, with further feature engineering, modeling, and visualization. The key technical components, as mentioned in our above report, include:

1. Volatility calculation module
2. Event database creation and integration
3. Gradient boosting model implementation
4. Feature importance analysis
5. Future prediction engine

The complete implementation is available in our Github/Juptyer Notebook file.