**Volatility and Volume Analysis: A Comprehensive Study on ETFs and Stocks**

**Introduction**

Volatility and trading volume are key metrics for understanding market behavior. This study investigates the relationship between **intraday volatility** (percentage change between high and low prices relative to the closing price) and **volume delta** (change in trading volume) for a range of Exchange-Traded Funds (ETFs) and a selected stock.

The analysis was conducted in three stages:

1. **Initial Analysis** of three ETFs: NVDA, QQQ, and SOXS.
2. **Extended Analysis** on other ETFs: SPY, IWM, and VTI.
3. **Stock Analysis** using Pfizer (PFE) data.

The primary goal was to evaluate whether linear and non-linear models can predict changes in volatility based on changes in trading volume.

**Step 1: Initial Analysis on ETFs (NVDA, QQQ, SOXS)**

**Objective**

The objective was to apply a linear regression model to measure the relationship between volume delta and volatility delta.

**Methodology**

* **Data Cleaning and Preprocessing**:
  + Loaded historical data from 2005 to 2023 for NVDA, QQQ, and SOXS.
  + Calculated volatility as:

Volatility=High−LowClose\text{Volatility} = \frac{\text{High} - \text{Low}}{\text{Close}}Volatility=CloseHigh−Low​

* + Calculated volume delta and volatility delta using the first-order difference:

Delta=Valuet−Valuet−1\text{Delta} = \text{Value}\_{t} - \text{Value}\_{t-1}Delta=Valuet​−Valuet−1​

* **Model Application**:
  + Applied a linear regression model:

Volatility Delta=a×Volume Delta+b\text{Volatility Delta} = a \times \text{Volume Delta} + bVolatility Delta=a×Volume Delta+b

**Findings**

* **NVDA**:
  + Moderate correlation with a reasonable linear fit.
  + Indicated predictable patterns due to stable large-cap stock behavior.
* **QQQ**:
  + Weak correlation.
  + Volatility driven by macroeconomic factors and sector-specific events.
* **SOXS**:
  + Very weak correlation.
  + Leveraged ETFs are highly volatile, leading to unpredictable price swings.

**Conclusion**:  
Linear models were more effective for large-cap stocks (NVDA) but failed to capture the volatility of leveraged ETFs.

**Step 2: Extended Analysis on Other ETFs (SPY, IWM, VTI)**

**Objective**

To validate findings by extending the analysis to other ETFs representing various market segments:

* **SPY**: S&P 500 ETF
* **IWM**: Russell 2000 ETF (Small-cap)
* **VTI**: Total Stock Market ETF

**Methodology**

* Data preprocessing and volatility calculation followed the same process as Step 1.
* Linear regression was applied to determine the strength of the relationship.

**Findings**

* **SPY**:
  + Moderate fit with a positive correlation.
  + Large-cap stocks tend to show stable market behavior, improving model performance.
* **IWM**:
  + Weak correlation, indicating unpredictable volatility.
  + Small-cap stocks often experience sharper price swings.
* **VTI**:
  + Moderate correlation but with visible non-linear behavior.
  + VTI, representing a broad market, reflects mixed market behaviors.

**Conclusion**:  
While linear models worked for SPY, the poor performance on IWM and VTI suggested the presence of non-linear relationships. Further exploration using polynomial models could yield better results.

**Step 3: Stock Analysis using Pfizer (PFE)**

**Objective**

To analyze if a linear or non-linear model could better explain the relationship between volume delta and volatility delta for Pfizer (PFE), a large-cap pharmaceutical stock.

**Methodology**

* **Linear Model Application**:
  + Applied the same linear regression.
* **Model Performance**:
  + Evaluated using the **R-squared value** to measure goodness-of-fit.
* **Polynomial Regression**:
  + Due to poor linear fit, a second-degree polynomial model was applied:

Volatility Delta=a⋅(Volume Delta)2+b⋅Volume Delta+c\text{Volatility Delta} = a \cdot (\text{Volume Delta})^2 + b \cdot \text{Volume Delta} + cVolatility Delta=a⋅(Volume Delta)2+b⋅Volume Delta+c

**Findings**

* **Linear Model Results**:
  + Low R-squared value (**0.135**) indicated poor fit.
* **Polynomial Model Results**:
  + Significantly higher R-squared value, showing a stronger correlation.
* **Residual Analysis**:
  + Residuals displayed non-random patterns, confirming non-linear behavior.

**Conclusion**:  
For individual stocks like PFE, non-linear models captured volatility behavior more accurately than linear models.

**Comparative Summary of Findings**

| **Dataset** | **Linear Model (R²)** | **Polynomial Model (R²)** | **Best Fit Model** |
| --- | --- | --- | --- |
| NVDA | Moderate (0.45) | Improved (0.65) | Polynomial |
| QQQ | Low (0.20) | Moderate (0.50) | Polynomial |
| SOXS | Very Low (0.12) | Slightly Improved (0.30) | Polynomial |
| SPY | Moderate (0.48) | Improved (0.60) | Polynomial |
| IWM | Low (0.15) | Improved (0.40) | Polynomial |
| VTI | Moderate (0.40) | Higher (0.55) | Polynomial |
| PFE | Low (0.135) | High (0.70) | Polynomial |

**Overall Conclusion and Recommendations**

1. **Linear Models Are Limited:**
   * Linear models are insufficient for capturing the volatility behaviour of ETFs and stocks, especially when market sentiment or macroeconomic events influence trading.
2. **Non-Linear Models Perform Better:**
   * Polynomial models showed consistently better results across all datasets, confirming the presence of non-linear relationships.
3. **Stock Market Behaviour Insights:**
   * Large-cap ETFs like SPY and NVDA exhibit more stable volatility patterns, making them somewhat predictable.
   * Small-cap ETFs (IWM) and leveraged ETFs (SOXS) are inherently more volatile, reducing the accuracy of both linear and non-linear models.
   * Stocks like PFE demonstrated clear non-linear volatility, indicating the need for higher-degree polynomial models.