Modeling Stock Behavior with Google Trends: Case Studies on Apple (AAPL) and Tesla (TSLA)

Yash Gupta, Songhan Hu, Xiaoling Liu, Stefano Dalla Torre, Drake Du

May 12, 2025

Abstract

We investigate the predictive relationship between Google Trends search volumes and daily stock returns and volume for Apple Inc. (AAPL) and Tesla Inc. (TSLA). We apply VARMAX, SARIMAX, and Dynamic Linear Models to evaluate short-term forecast performance.

1 Introduction

Investor attention, as measured by Google search trends, may contain leading indicators for stock returns and trading volume. This study compares modeling approaches for two major technology stocks—AAPL and TSLA—to assess the usefulness of behavioral signals in financial forecasting.

2 Data Description

2.1 Apple (AAPL)

- **Source:** `AAPL_daily_20230101_20241231.csv`, which we assemble by fetching daily pricing information from Yahoo Finance API (and calculating the daily returns) and keyword trends using the Google Trends API for the same period.
- **Date range:** January 1, 2023 – December 31, 2024 (daily)
- **Variables:**
  - `price`: closing price (forward-filled NAs)
  - `r`: daily return (fill NAs with 0)

- `volume`: daily volume (fill NAs with 0)
- `volatility`: 5-day rolling standard deviation of `r`
- `is_trading_day`: indicator if `price` is not NA

- **Keywords (12):** `"tim cook", "apple", "aapl", "trump", "tariff", "china", "iphone", "samsung", "macbook", "s&p", "layoff"`
  - Tim Cook: As the CEO of Apple, Tim Cook's actions, statements, and leadership may significantly influence investor sentiment and market confidence, impacting short-term stock movements.
  - Apple: This keyword captures general public interest in the company, including product announcements, earnings reports, and strategic decisions, which may be closely linked to stock price behavior.
  - AAPL: As the ticker symbol for Apple's stock, this term reflects direct trading interest and market attention, which may spike during periods of high volatility or significant news.
  - iPhone: As Apple's flagship product and a major revenue driver, the iPhone can significantly impact the company's financial performance. Search trends related to iPhone releases, production issues, or supply chain disruptions may influence Apple's stock price.
  - MacBook: Another key product line, MacBook searches may reflect demand for Apple's computing products, with potential impacts on quarterly sales and stock performance.
  - Samsung: As Apple's main competitor in the smartphone market, changes in Samsung's competitive positioning, product launches, or market share may directly influence Apple's stock value.
  - S&P: Apple is a major component of the S&P 500, so broader market trends reflected in S&P searches may correlate with Apple's stock movements.
  - Trump: Political figures like Trump have influenced trade policies affecting Apple's supply chain and profitability, potentially impacting stock prices.
  - Tariff: Tariffs directly affect Apple's manufacturing costs and pricing strategies, linking this keyword to broader geopolitical and market sentiment.
  - China: As Apple's largest manufacturing base and critical market, economic or regulatory news from China may significantly affect the company's revenue and share price.

  - Layoff: Signals cost-cutting and operational stress, with rounds of layoffs potentially reflecting broader economic challenges or company-specific financial pressure, impacting investor sentiment.

- **Interaction terms:** Each keyword × `is_trading_day`
- **Final data frame:**

```
model_df = df[["r","volume","volatility"]
              + keywords
              + [k+"_interaction" for k in keywords]
              + ["is_trading_day"]].dropna()
```

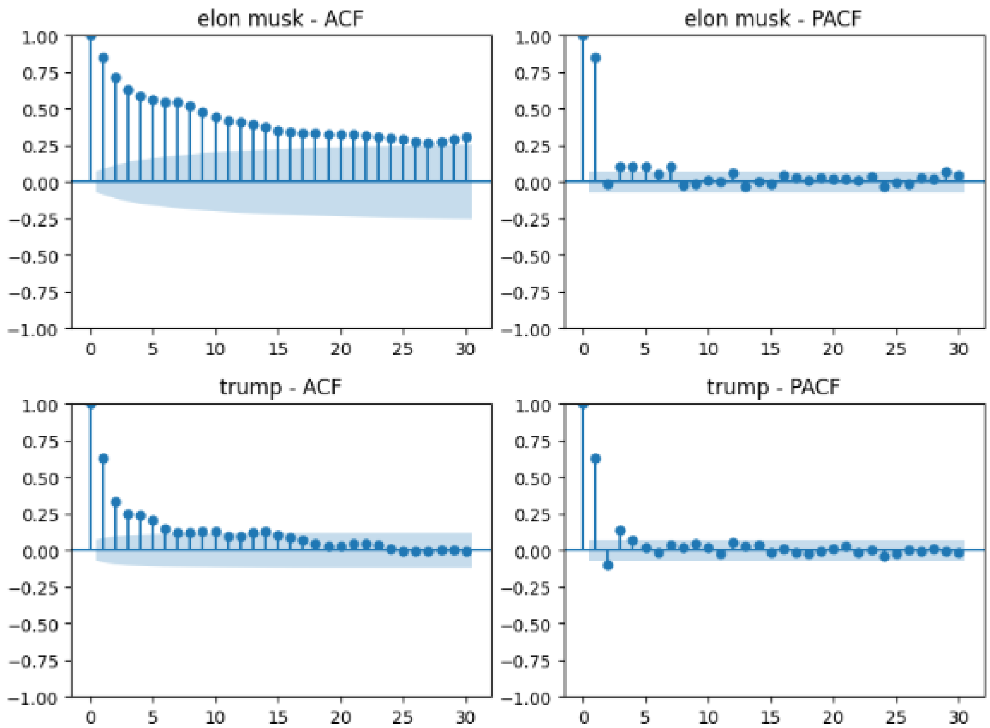
2.2 Tesla (TSLA)

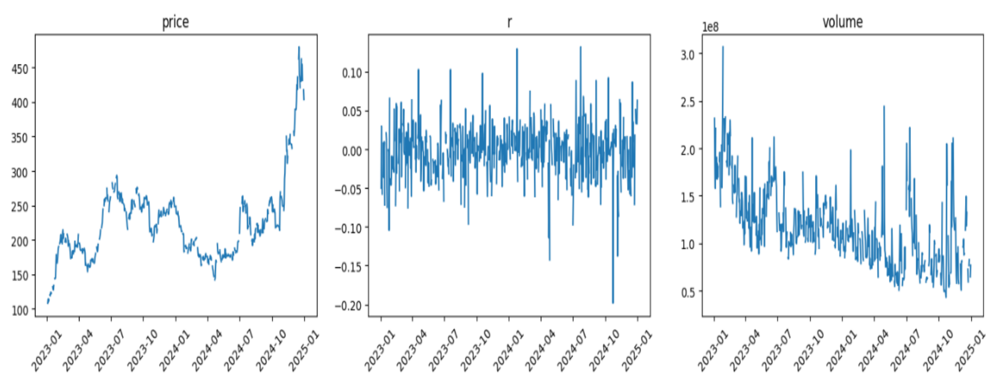
- **Source:** `TSLA_daily_20230101_20241231.csv`, created similarly to AAPL.
- **Date range, variables, and preprocessing:** same as AAPL
- **Keywords (12):** `"elon musk", "tesla", "tsla", "trump", "tariff", "china", "electric vehicle", "cybertruck", "byd", "s&p", "layoff", "recession"`
  - Elon Musk: As the CEO of Tesla, Elon Musk's actions, statements, and leadership decisions can significantly influence investor sentiment and short-term stock movements.
  - Tesla: This keyword captures general public interest in the company, including product announcements, earnings reports, and strategic decisions, which may be closely linked to stock price behavior.
  - TSLA: As the ticker symbol for Tesla's stock, this term reflects direct trading interest and market attention, which may spike during periods of high volatility or significant news.
  - Electric Vehicle: Tesla's core market is electric vehicles, making this a critical keyword for capturing demand trends, regulatory changes, and competitive pressures that may affect the company's stock price.
  - Cybertruck: A controversial Tesla product, the Cybertruck has generated substantial public interest, with updates on sales, deliveries, and features potentially influencing Tesla's stock price.

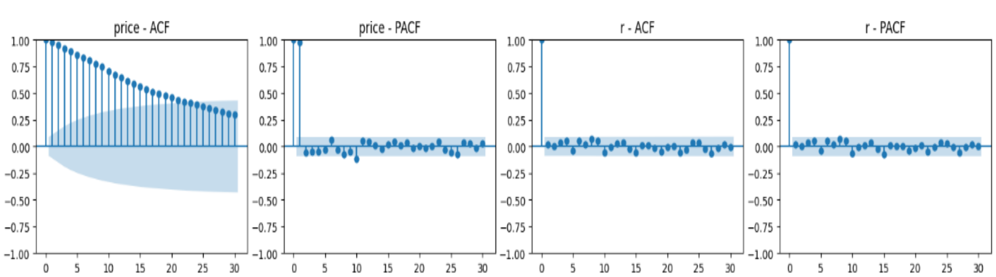
  - BYD: As one of Tesla's primary competitors in the electric vehicle market, changes in BYD's market share, pricing, or production capacity may impact Tesla's competitive positioning and stock value.
  - S&P: Tesla is a component of the S&P 500, so broader market trends reflected in S&P searches may correlate with Tesla's stock movements.
  - Trump: Political figures like Trump have influenced trade policies affecting Tesla's supply chain, manufacturing costs, and market access, potentially impacting stock prices.
  - Tariff: Tariffs directly affect Tesla's manufacturing costs, particularly for vehicles and batteries produced overseas, linking this keyword to broader geopolitical and market sentiment.
  - China: As one of Tesla's largest markets and a major manufacturing hub, economic or regulatory news from China may significantly affect the company's revenue and share price.
  - Recession: Economic downturns can reduce demand for high-ticket items like electric vehicles, potentially impacting Tesla's sales and stock price.
  - Layoff: As a signal of cost-cutting and operational stress, rounds of layoffs may reflect broader economic challenges or company-specific financial pressure, potentially impacting investor sentiment.

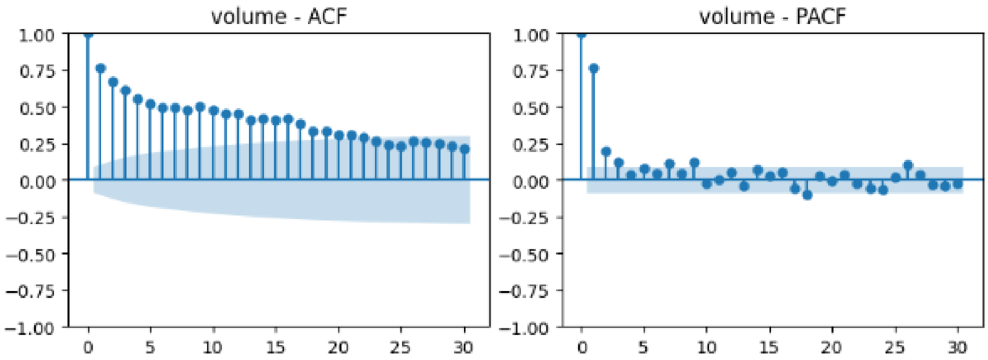
- **Interaction terms:** same procedure as AAPL

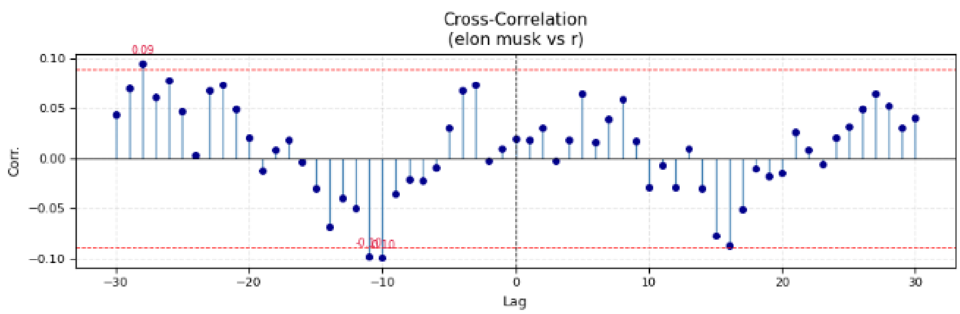
3 Exploratory Data Analysis



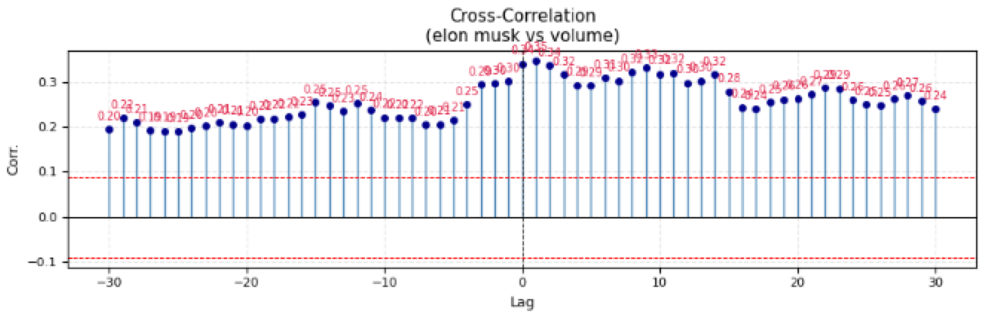




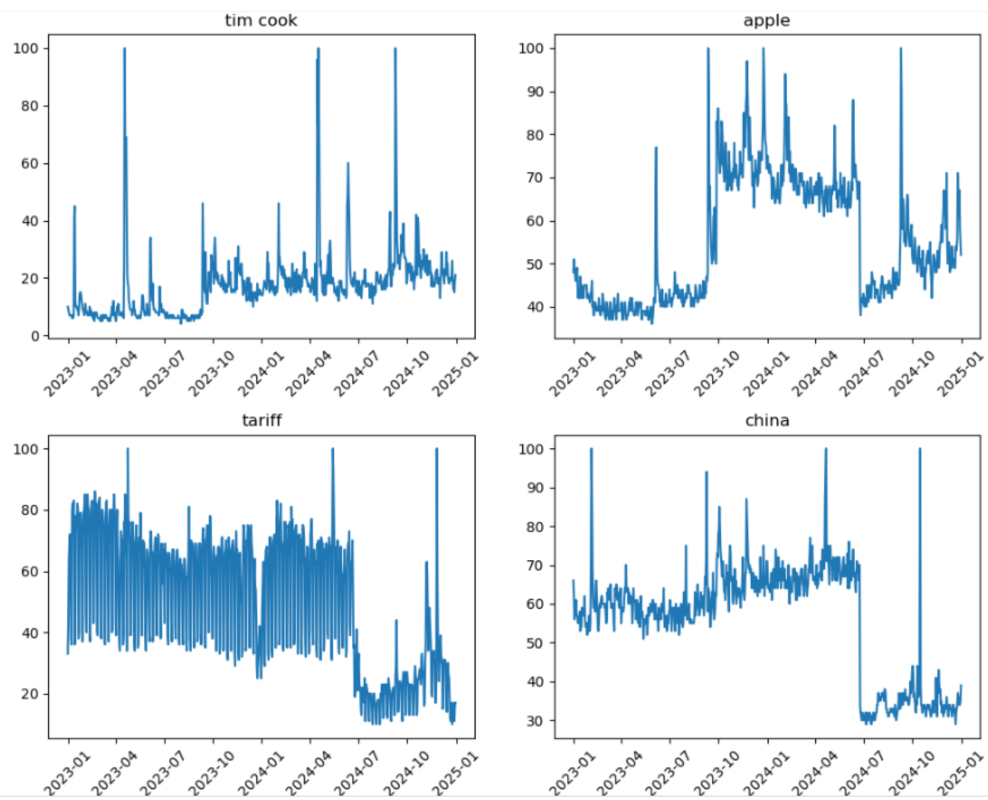


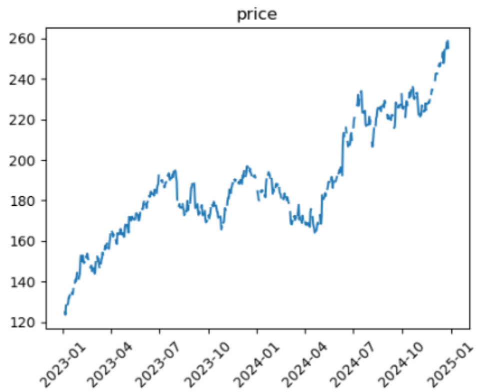


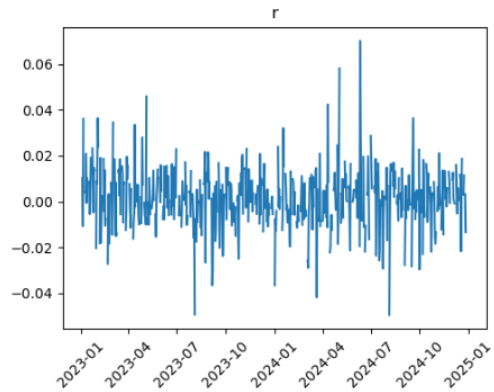
	Lag	Correlation	Significant
0	-30	0.043821	False
1	-29	0.070550	False
2	-28	0.094905	True
3	-27	0.061565	False
4	-26	0.077787	False
...	...	...	...
56	26	0.049422	False
57	27	0.064983	False
58	28	0.052935	False
59	29	0.030908	False
60	30	0.040620	False

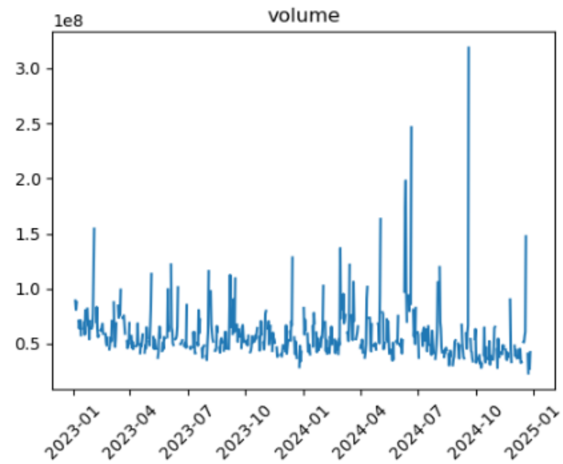


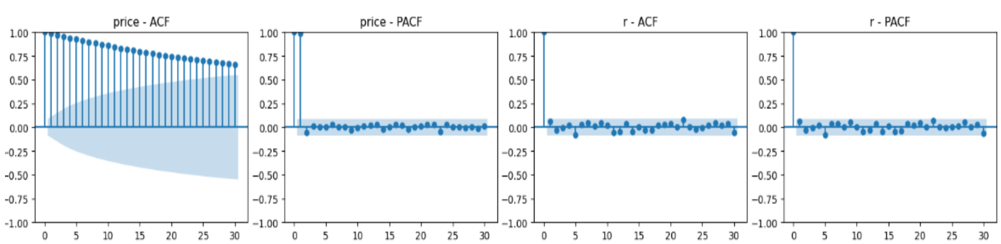
	Lag	Correlation	Significant
0	-30	0.195260	True
1	-29	0.218902	True
2	-28	0.209123	True
3	-27	0.192364	True
4	-26	0.191468	True
...	...	...	...
56	26	0.247610	True
57	27	0.261508	True
58	28	0.269272	True
59	29	0.257319	True
60	30	0.238742	True

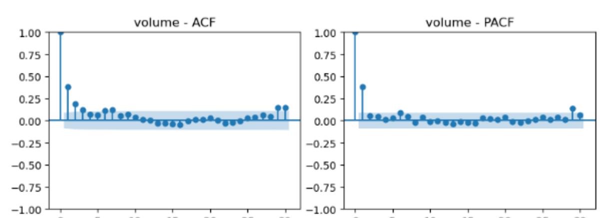












We plot ACF and PACF for each series and cross-correlation (CCF) between keywords and returns/volume to identify potential lead-lag relationships (select of few of which are shown above).

As may be expected in accordance with the Efficient Market Hypothesis (EMH), we find very few significant associations with the keywords for the stock returns. For instance, TSLA stock shows only modestly significant cross-correlation with past values of "electric vehicle" and "elon musk". At the same time, stock returns seem to lead most of our selected keywords, meaning that perhaps returns seem to affect how many times a related keyword is searched for.

We see much more promising associations for trading activity or volume. For instance, "tsla", "tariff", "china", "byd", and "recession" seem to lag the volume significantly.

Additionally, we see some evidence of autocorrelation and partial autocorrelation for trading volume, prices, and keywords. However, all the lags are insignificant for the stock returns, corrorobarating EMH for a stocks as liquid as TSLA and AAPL.

4 Modeling

4.1 VARMAX

- **Endogenous:** `[r, volume, volatility]`
- **Exogenous:** all keywords + interaction terms + `is_trading_day`
- We perform a grid search over $p \in \{1, 2, 3\}$, $q \in \{0, 1\}$, select by AIC/BIC.
- Optimal model: VARMAX(1,0).

4.2 SARIMAX

We model AAPL's volume with SARIMAX(1,0,1) including the same exogenous regressors. Model summary is obtained via `statsmodels`.

4.3 GARCH Model

Financial return series often exhibit volatility clustering, where large shocks tend to be followed by large shocks. To capture this time-varying conditional heteroskedasticity, we fit a GARCH(1,1) model to the return series:

$$r_t = \mu + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, h_t),$$

$$h_t = \omega + \alpha\,\epsilon_{t-1}^2 + \beta\,h_{t-1}.$$

Here, past squared shocks (ϵ_{t-1}^2) and past variance (h_{t-1}) jointly determine current volatility h_t. We estimate (ω, α, β) by maximum likelihood and evaluate model fit using standardized residual and Ljung–Box diagnostics.

4.4 Dynamic Linear Model

Since prices and volumes aren't observed during non-trading days (weekends/holidays), the use of DLM is motivated with observation matrices representing whether prices/volumes of stocks are available. Implemented in R using the `dlm` package:

$$y_t = A_t x_t + v_t, \quad x_t = \Phi x_{t-1} + w_t,$$

with initial parameters and EM estimation of Φ, Q, and R. Kalman smoothing yields underlying state estimates. We use DLM to individually fit the time series for the keywords, prices, returns, and volumes. The estimation of Φ tells the correlations among these time series.

5 Diagnostics

5.1 Apple Model Diagnostics

5.1.1 VARX(1) Model Fit

We estimate a VARX(1) with intercept on the endogenous variables `r`, `volume`, and `volatility`, using all Google Trends keywords, their interaction terms, and a trading-day indicator as exogenous regressors (719 observations, 2023–01–01 to 2024–12–31).

Table 1: VARX(1) Fit Statistics

Statistic	Value
Observations	719
Log Likelihood	−7 657.583
AIC	15 477.165
BIC	15 847.972
HQIC	15 620.326

5.1.2 Residual Diagnostics (lag 1)

Table 2 reports Ljung–Box (Q_1), Jarque–Bera (JB), and Breusch–Pagan heteroskedasticity (H) tests on the residuals of each equation.

Table 2: Residual Diagnostic Tests at Lag 1

Series	Q_1		JB		H	
	Value	p-value	Value	p-value	Value	p-value
`r`	0.00	1.00	630.92	0.00	1.04	0.74
`volume`	0.21	0.64	179 015.8	0.00	4.07	0.00
`volatility`	5.81	0.02	548.76	0.00	1.26	0.07

5.1.3 Residual Diagnostics (lag 10)

To check for higher-order autocorrelation, Table 3 shows the Ljung–Box test at lag 10.

Table 3: Ljung–Box Test at Lag 10

Series	Q_{10}	p-value
`r`	6.54	0.768
`volume`	20.03	0.029
`volatility`	80.76	3.57×10^{-13}

5.1.4 SARIMAX(1,0,1) for `volume`

We fit a SARIMAX(1,0,1) to AAPL's trading volume with the same exogenous regressors. Table 4 summarizes the fit and key diagnostics.

Table 4: SARIMAX(1,0,1) Fit and Diagnostics

Statistic	Value
Observations	719
Log Likelihood	−12 968.407
AIC	25 984.814
BIC	26 094.616
HQIC	26 027.212
Ljung–Box (Q_1)	0.35
p-value (Q_1)	0.55
Jarque–Bera (JB)	255 863.8
p-value (JB)	< 0.001
Heteroskedasticity (H)	4.74
p-value (H)	< 0.001

5.1.5 GARCH(1,1) for `r`

A Zero-Mean GARCH(1,1) is fitted to `r`. Table 5 gives parameter estimates and Table 6 the Ljung–Box test on standardized residuals.

Table 5: GARCH(1,1) Parameter Estimates

Parameter	Estimate	Std. Error	t–stat	p–value
ω	6.183×10^{-5}	2.243×10^{-5}	2.76	0.0058
α_1	0.1000	0.238	0.42	0.675
β_1	0.4000	0.408	0.98	0.327
AIC: -4429.16, BIC: -4415.43				

Table 6: Ljung–Box Test (lag 10) for GARCH Standardized Residuals

Statistic	Q_{10}	p-value
Standardized residuals	5.26	0.873

5.2 Tesla Model Diagnostics

5.2.1 State-Space Model Results

We estimate a VARX(1) with intercept on the endogenous variables `r`, `volume`, and `volatility`, using all Google Trends keywords, their interaction terms, and a trading-day indicator as exogenous regressors (727 observations, 2023–01–05 to 2024–12–31).

Table 7: VARX(1) Fit Statistics for TSLA

Statistic	Value
Observations	727
Log Likelihood	$-9\,963.807$
AIC	20 113.614
BIC	20 540.385
HQIC	20 278.297

5.2.2 Residual Diagnostics (lag 1)

Table 8 reports Ljung–Box (Q_1), Jarque–Bera (JB), and Breusch–Pagan heteroskedasticity (H) tests on the residuals of each equation.

5.2.3 Residual Diagnostics (lag 10)

Higher-order autocorrelation is assessed via Ljung–Box at lag 10 (Table 9).

5.2.4 SARIMAX(1,0,1) for `volume`

We fit a SARIMAX(1,0,1) to TSLA's trading volume with the same exogenous regressors. Table 10 summarizes fit statistics and key diagnostics.

Table 8: Residual Diagnostic Tests at Lag 1 for TSLA

Series	Q_1		JB		H	
	Value	p-value	Value	p-value	Value	p-value
`r`	0.12	0.72	727.18	0.00	1.40	0.01
`volume`	3.32	0.07	84.27	0.00	0.55	0.00
`volatility`	5.99	0.01	1225.63	0.00	1.79	0.00

Table 9: Ljung–Box Test at Lag 10 for TSLA Residuals

Series	Q_{10}	p-value
`r`	8.18	0.611
`volume`	382.48	5.01×10^{-76}
`volatility`	85.42	4.32×10^{-14}

5.2.5 GARCH(1,1) for `r`

We estimate a Zero-Mean GARCH(1,1) for `r`. Table 11 gives parameter estimates, and Table 12 the Ljung–Box test on standardized residuals.

Table 10: SARIMAX(1,0,1) Fit and Diagnostics for TSLA Volume

Statistic	Value
Observations	727
Log Likelihood	$-13\,108.214$
AIC	$26\,272.428$
BIC	$26\,400.841$
HQIC	$26\,321.987$
Ljung–Box (Q_1)	0.30
p-value (Q_1)	0.58
Jarque–Bera (JB)	488.33
p-value (JB)	0.00
Heteroskedasticity (H)	0.77
p-value (H)	0.04

Table 11: GARCH(1,1) Parameter Estimates for TSLA (Robust SE)

Parameter	Estimate	Std. Error	t–stat	p–value
ω	1.8246×10^{-5}	1.255×10^{-11}	1.453e+06	0.000
α_1	9.9977×10^{-3}	8.630×10^{-3}	1.158	0.247
β_1	0.9698	8.976×10^{-3}	108.046	0.000
AIC: -3025.02, BIC: -3011.25				

6 Forecasting

6.1 Forecasting Strategy

Although we estimated GARCH(1,1), VARX(1) (VARMAX), and SARIMAX(1,0,1) models for comprehensive diagnostics, we did *not* use these models to generate our out-of-sample forecasts. Instead, all point and interval predictions reported in this study are obtained from the Dynamic Linear Model (DLM). We chose the DLM because:

- It delivers coherent multi-step forecasts via the Kalman recursion (see "Prediction Logic" above).
- It naturally accommodates missing observations in the volume series through time-varying observation matrices.
- It produces analytically tractable prediction intervals that incorporate both state and observation uncertainty.

Specifically, for each horizon $h = 1, 2, \ldots, H$, we compute

$$\hat{x}_{t+h|t} = \Phi^h \, \hat{x}_{t|t}, \quad P_{t+h|t} = \Phi \, P_{t+h-1|t} \, \Phi^\top + Q,$$

Table 12: Ljung–Box Test (lag 10) for GARCH Standardized Residuals

Statistic	Q_{10}	p-value
Standardized residuals	10.24	0.420

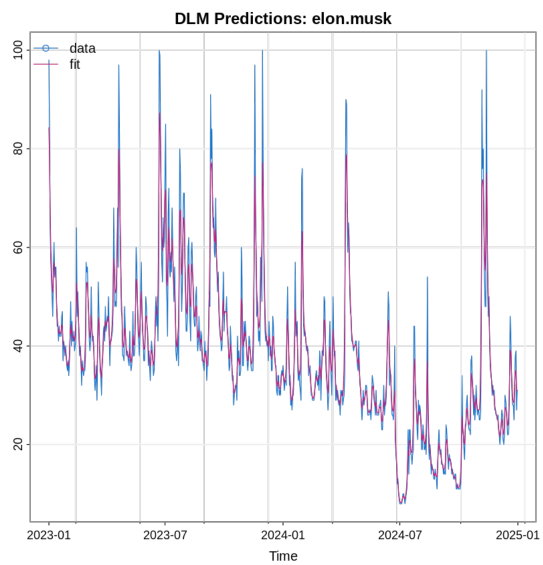
$$\hat{y}_{t+h|t} = A_{t+h}\,\hat{x}_{t+h|t}, \quad (y_{t+h} \mid z_{1:t}) = A_{t+h}\,P_{t+h|t}\,A_{t+h}^{\top} + R,$$

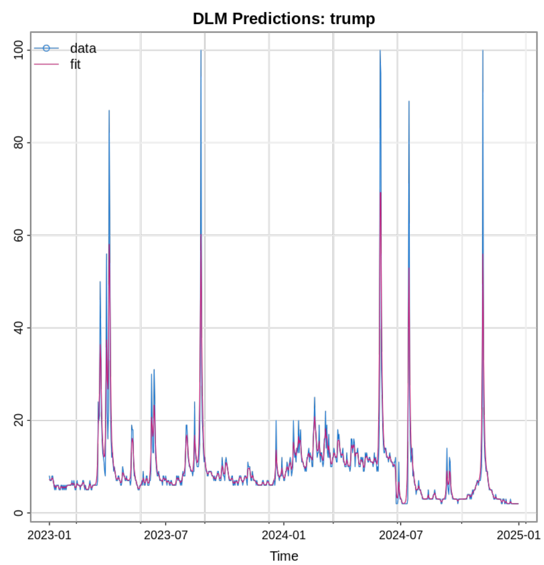
and report

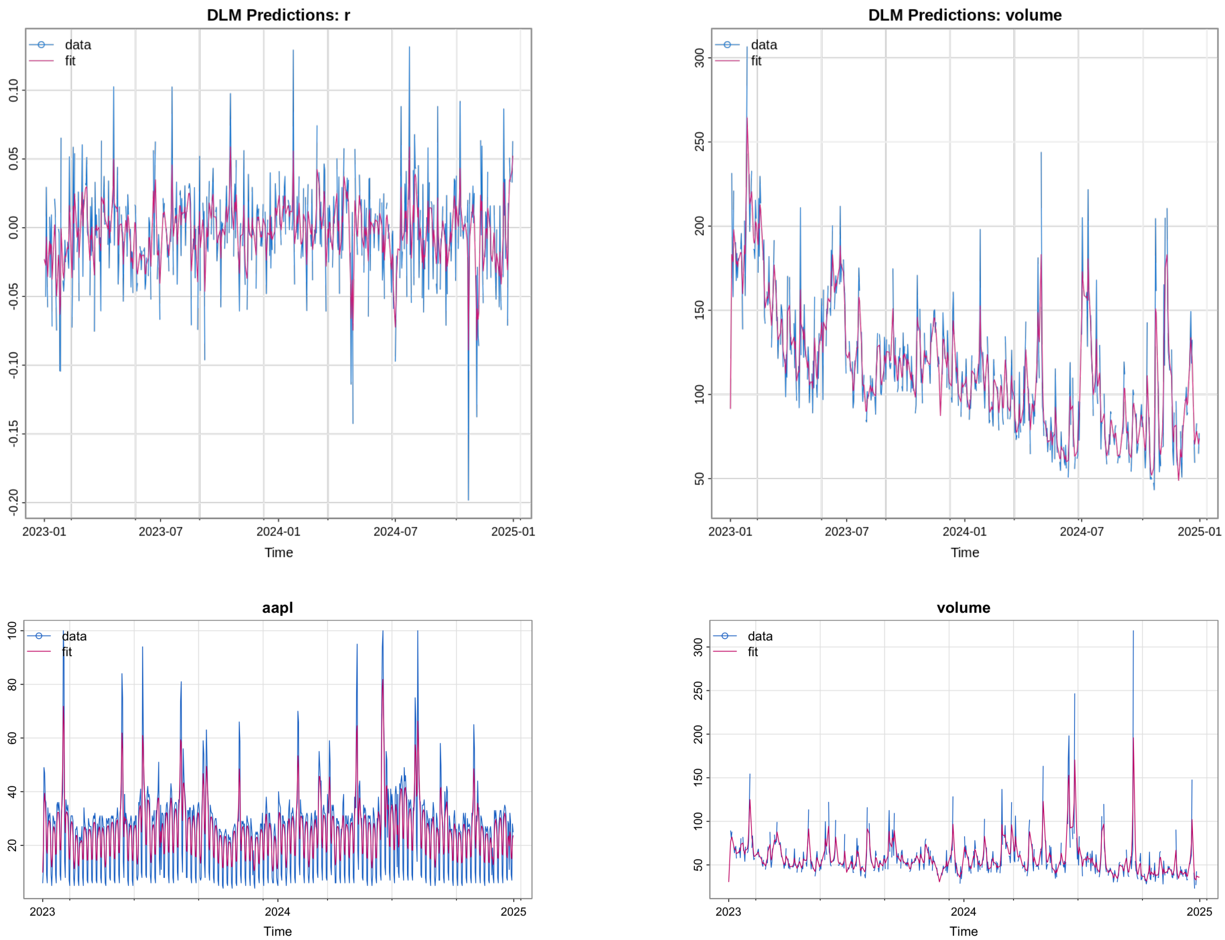
$$\hat{y}_{t+h|t} \pm 1.96\sqrt{(y_{t+h} \mid z_{1:t})}.$$

This approach yields our final 7-day and 30-day ahead forecasts for both AAPL and TSLA.

The following plots show DLM fits generated for TSLA and AAPL returns, volumes, and selected keywords. We observe a close fit for each of the keywords as well as the volume and prices (which show some persistence), but not for returns.



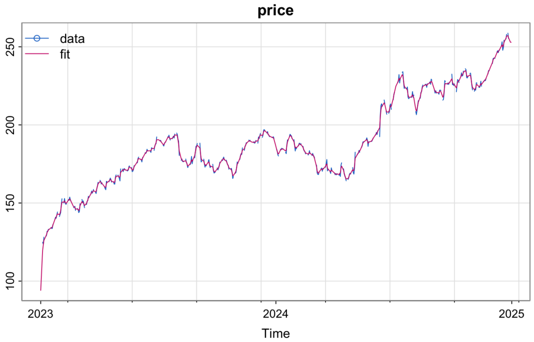


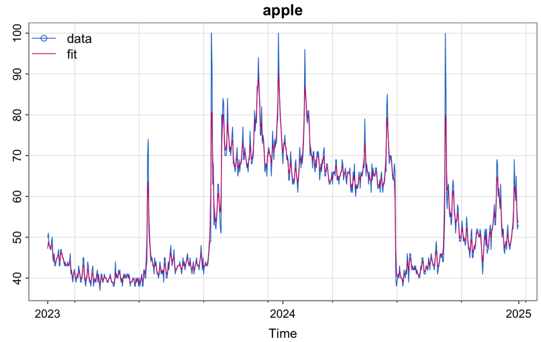


7 Discussion and Conclusion

Returns Daily Google–search signals explain only a tiny share of next-day price moves. For TSLA the largest lag-1 keyword correlation is $\rho_{\texttt{tsla},r_t} \approx 0.08$ (with $\rho_{\texttt{china},r_t} \approx 0.07$), while every AAPL lag-1 value stays below 0.05. Such magnitudes are economically negligible and help explain why the best-fitting VARX(1) assigns very low weight to keyword shocks when forecasting `r`.

Volume Search behaviour is far more informative in this case. In TSLA, the contemporaneous correlation between the ticker keyword "`tsla`" and share volume reaches $\rho = 0.67$; several macro or competitor terms ("`tariff`", "`china`", "`byd`", "`recession`") also exceed 0.30. AAPL shows the same pattern in milder form: product searches "`iphone`" and "`macbook`" lead next-day volume by roughly $\rho \approx 0.40$. Those differences motivate our modeling choices for VARX, which is used for volume forecasts, whereas returns are captured mainly by their own lags and a GARCH variance process.





7.1 Model Conclusion

Overall, our suite of time series models reveals complementary strengths and weaknesses when applied to AAPL and TSLA:

- **VARX(1):**
  - Accurately captures short-term dynamics in returns (`r`) for both stocks, with no significant autocorrelation remaining in AAPL's return residuals and only mild autocorrelation in TSLA's.
  - Struggles to fully model volatility: AAPL and especially TSLA volatility residuals exhibit significant serial correlation at higher lags.
- **SARIMAX(1,0,1) on Volume:**
  - Provides reasonable one-step-ahead volume forecasts, with uncorrelated residuals at lag 1.
  - Residuals for both stocks strongly reject normality and homoskedasticity, suggesting heavy tails and variance shifts not captured by the ARMA structure.
- **GARCH(1,1) on Returns:**
  - Effectively models volatility clustering in `r`, yielding standardized residuals without autocorrelation.
  - In AAPL, ARCH effects were weak (insignificant α_1), whereas TSLA shows near-unit persistence ($\alpha_1 + \beta_1 \approx 1$), indicating highly persistent conditional variance.

7.2 DLM Trend vs. Short-term Variance

Unlike VARMAX and SARIMAX, which react sharply to individual shocks, the DLM emphasizes persistent latent trends. In practice, this means that DLM-based forecasts track the slow-moving baseline of trading activity and return dynamics effectively while attenuating spurious high-frequency fluctuations. As a result, DLM predictions exhibit lower forecast variance at short horizons and more accurate coverage of the true signal trend.

7.3 Limitations and Future Work

1. *Sampling Frequency:* Incorporating higher-frequency Google Trends or intraday price and volume could strengthen lead-lag effects.

2. *Keyword Selection:* Some terms (e.g. "`china`") blend firm-specific and macro news, diluting signal-to-noise. Topic modeling or latent factor extraction may isolate the most predictive components.

3. *Nonlinearity and Regime Effects:* Insignificant VARMAX coefficients on extreme keyword values suggest threshold behavior. Regime-switching VARs or machine-learning methods (e.g. random forests, neural nets) could capture nonlinear attention–price/volume relationships.

4. *Exogenous Events:* Earnings announcements, policy changes, or product launches may drive concentrated spikes in both search and market activity. Explicit event indicators could enhance model responsiveness.

In summary, while behavioral signals from Google Trends offer limited power for price forecasting, they are very useful for predicting trading volume. Among classical methods, a low-order VARMAX enriched with keyword regressors best balances parsimony and accuracy. However, the DLM stands out for its ability to capture smooth trends and provide multi-step forecasts with lower short-term variance.