Time Series Analysis of Google Realized Volatility

Huy Le and Tyler Yee

February 2026

Abstract

Stock prices are driven by a mix of economic fundamentals, investor sentiment, and external shocks, making them inherently difficult to predict. While forecasting price direction remains elusive, volatility exhibits strong temporal persistence that time series models can exploit. We study the weekly log realized volatility (Log RV) of Google (GOOG) from 2019 to 2025 using ARIMA and SARIMAX frameworks. An extensive AIC-based grid search identifies ARIMA(3,0,3) as the best univariate model. Incorporating the VIX index as an exogenous regressor yields $\text{SARIMAX}(3,0,3)(0,0,1)_4$ with substantially lower AIC. One-step-ahead rolling forecasts on a held-out test set confirm that the SARIMAX model outperforms the ARIMA baseline on RMSE, MAE, and MAPE, demonstrating that market-wide fear captured by VIX provides genuine predictive power for individual stock volatility.

1 Introduction

Every trading day, financial markets produce a stream of prices, volumes, and derived indicators that unfold in sequence. Because this data is naturally ordered by time, time series analysis offers a natural lens for studying it. ARIMA models, for instance, describe how a variable's recent values and recent forecast errors carry forward into the future, while SARIMAX models build on this by allowing for seasonal rhythms and the influence of outside variables (1; 2). These frameworks fit financial data well because many market quantities tend to be persistent and mean-reverting, patterns that methods ignoring temporal order would overlook.

Predicting whether a stock will go up or down is famously difficult. The efficient market hypothesis holds that prices already reflect available information, so returns behave roughly like a random walk (3). Volatility tells a different story. It clusters, where calm periods tend to follow calm periods, and turbulent ones tend to persist. A trader who can anticipate a spike in volatility can hedge a portfolio with options, whose prices are tied directly to expected volatility, or scale down positions before the storm hits. Conversely, when implied volatility rose above what actually realized, strategies like selling straddles can profit from the gap (4). Realized volatility (RV), computed from observed returns, measures how much a stock actually moved, and forecasting it well is critical to option pricing, risk budgeting, and portfolio construction.

We focus on Google (Alphabet, ticker GOOG), one of the most liquid large-cap equities globally. Its volatility is shaped by both company-specific events like earnings reports and regulatory actions and by broader market forces. The question we set out to answer is: **can market-wide volatility, as measured by the VIX index, improve forecasts of Google's realized volatility beyond what the stock's own history provides?**

The VIX is often called the "investor fear gauge" because it captures the market's expectation of near-term S&P 500 volatility as implied by option prices (5). Prior work has shown that realized volatility exhibits long memory and strong persistence (6; 7), which motivates fitting ARMA-type models to log-transformed RV. We take this a step further by adding VIX as an exogenous regressor within a SARIMAX framework and testing whether it actually improves out-of-sample forecasts.

Our data runs from January 2019 through December 2025, a window that covers the COVID crash, the post-pandemic recovery, and the 2022 rate-hike cycle, giving us a wide range of volatility regimes to work with.

2 Data and Preprocessing

2.1 Realized volatility construction

We download daily GOOG closing prices and compute daily log returns $r_t = \log(P_t/P_{t-1})$. Realized volatility over a 5-day (weekly) window is given by

$$\mathrm{RV}_t = \hat{\sigma}_t\sqrt{252}, \quad \hat{\sigma}_t = \mathrm{std}(r_{t-4}, \dots, r_t),$$

where the $\sqrt{252}$ annualizes the estimate. We log-transform to obtain $\mathrm{Log\,RV}_t = \log(\mathrm{RV}_t)$, which produces a more symmetric, approximately stationary series (6).

To avoid mechanical autocorrelation from overlapping windows, we retain every 5th observation, yielding 351 non-overlapping weekly Log RV values.

2.2 Exogenous variables

We also download the VIX index and extract GOOG daily trading volume over the same period. After merging onto the non-overlapping dates (forward-filling VIX for any missing days), each observation carries three predictors: lagged Log RV values, the VIX level, and trading volume.

Figure 1 shows the resulting Log RV series. The COVID spike in early 2020 is the dominant feature, followed by elevated volatility during the 2022 tightening cycle. Figure 2 displays the VIX and volume series for context.
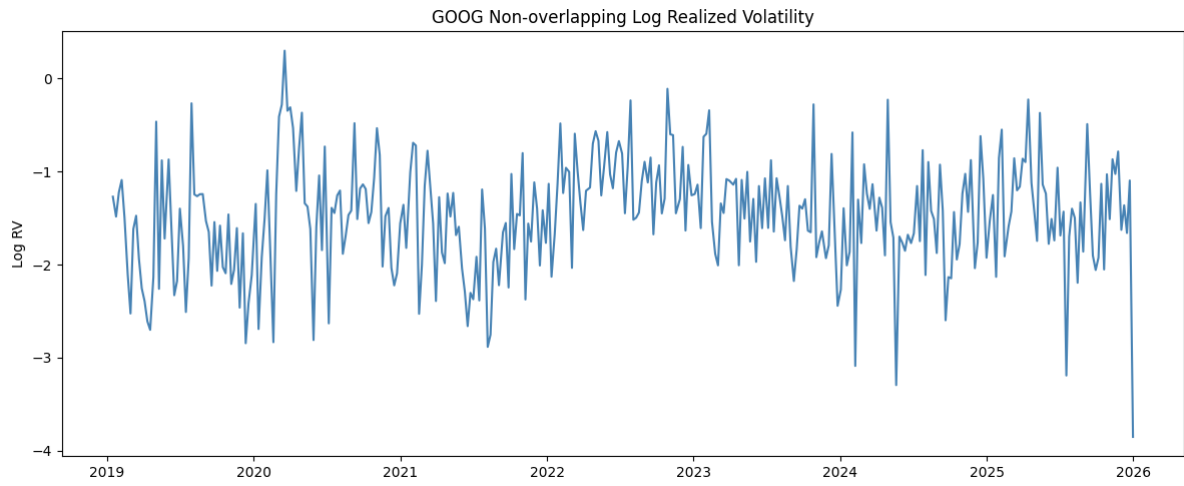
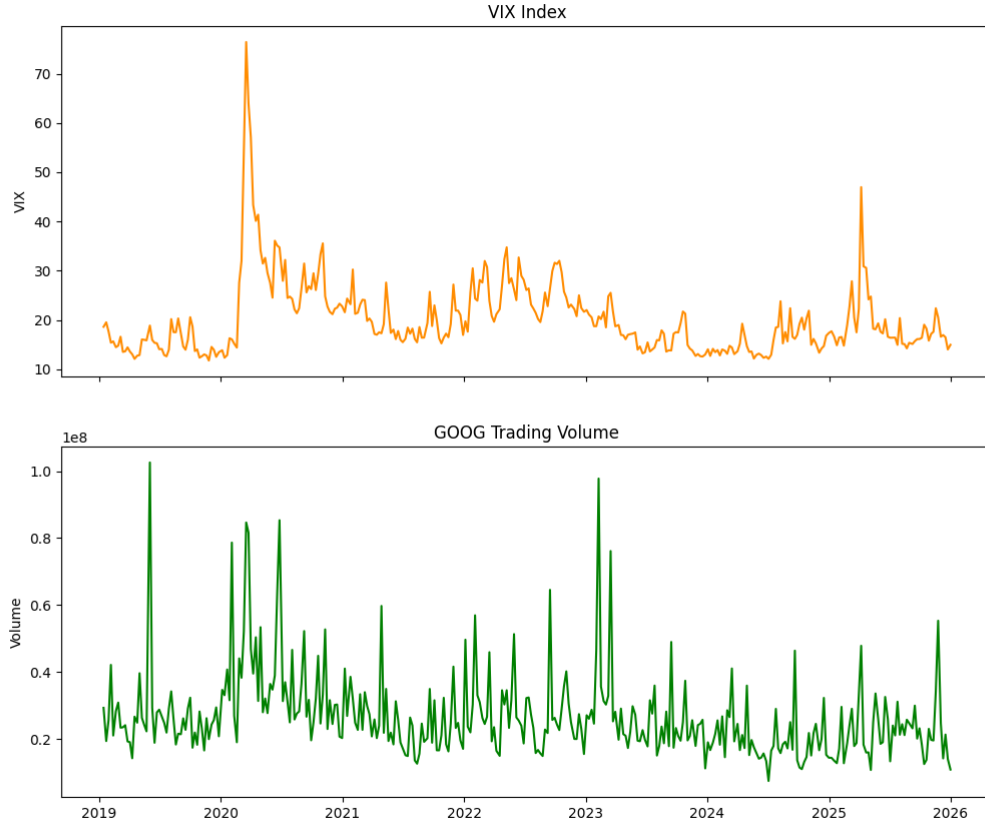Figure 1: Non-overlapping weekly Log Realized Volatility of GOOG, January 2019 to December 2025.

Figure 2: VIX index and GOOG daily trading volume over the sample period.

2.3 Train-test split

We split the data chronologically: the first 80% (280 observations, Jan 2019 – Aug 2024) form the training set, and the remaining 20% (71 observations, Aug 2024 – Dec 2025) form the test set. All model fitting and selection use only the training set; the test set is reserved for out-of-sample evaluation. Figure 3 shows the partition.
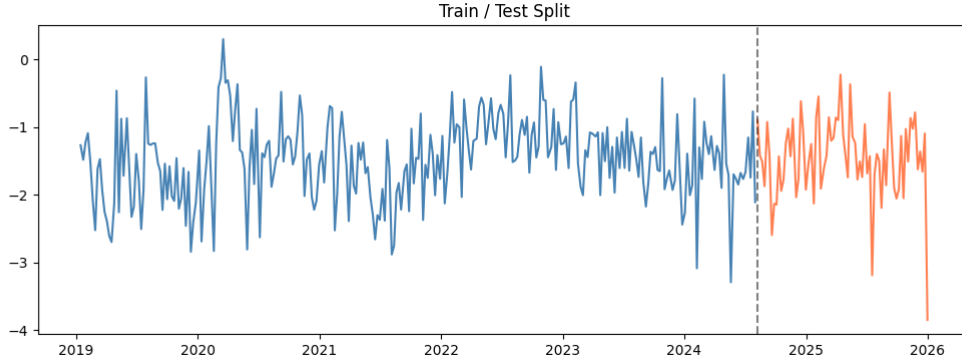
Figure 3: Train/test split of the Log RV series. The dashed line marks the boundary.

3 Exploratory Analysis

3.1 Stationarity

We apply two complementary tests to assess stationarity. The Augmented Dickey-Fuller (ADF) test, whose null hypothesis is that the series contains a unit root, yields a test statistic of -5.32 ($p < 10^{-5}$), strongly rejecting the unit root null ([2]). However, rejecting a unit root does not by itself establish stationarity. It only implies the series does not follow a random walk. To confirm stationarity from the opposite direction, we run the KPSS test ([8]), whose null hypothesis is that the series is stationary. The KPSS test returns $p = 0.309$, so we fail to reject stationarity at any conventional significance level. Taken together, the two tests agree that the Log RV series is stationary and no differencing is needed.

3.2 Autocorrelation structure

Figure 4 shows the ACF and PACF of training Log RV. The ACF is generally noisy, but there is a slight decreasing trend, consistent with the well-documented long-memory behavior of volatility ([6]; [7]). The PACF shows significant spikes at lags 1-2, suggesting AR components of order at least 2. Both functions show mild structure around lag 4, motivating a seasonal component with period $s = 4$ (roughly monthly in weekly data).
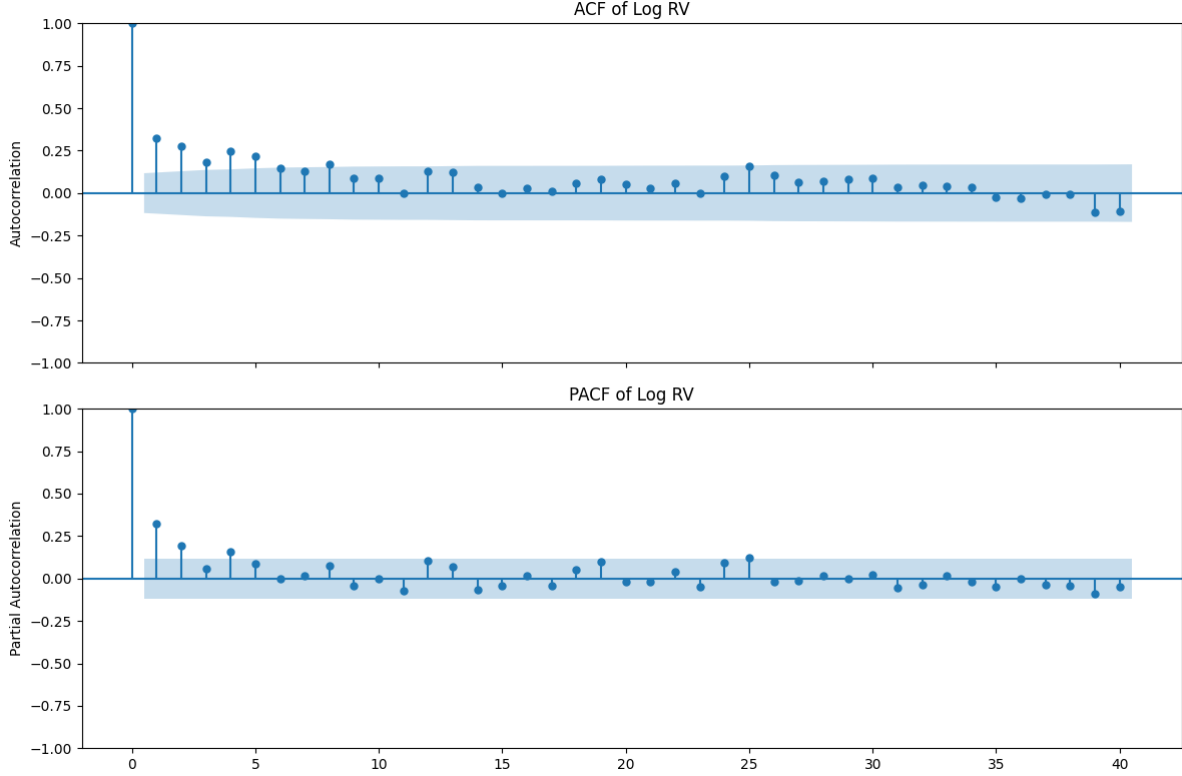
Figure 4: ACF and PACF of training Log RV.

3.3 Spectral analysis

Figure 5 presents the raw periodogram and a smoothed spectral density estimate. The raw periodogram, computed via the discrete Fourier transform, is an inconsistent estimator of the true spectral density because its variance does not decrease with sample size (1). To obtain a more stable estimate, we apply Welch's method (9; 10), which divides the series into overlapping segments, computes a periodogram for each segment, and averages them. This trading of frequency resolution for reduced variance aligns with our dataset since there are few, dominant periodicities.

The raw periodogram shows elevated power near frequency zero, but this should not be interpreted as a genuine long-period cycle. With only about 280 training observations, the lowest resolvable frequencies correspond to periods of 140 observations (weeks), meaning the series contains too few complete cycles at those frequencies for the estimate to be reliable. This kind of low-frequency inflation is a well-known artifact in short financial time series and typically reflects the slow mean-reversion and persistence of volatility rather than any true periodicity (7).

The more meaningful feature is a peak near frequency 0.24 cycles per observation, corresponding to a period of approximately 4.2 weeks, which we round to 4 weeks. This aligns with the seasonal period $s = 4$ that we adopt in our SARIMA specifications, and it likely reflects the monthly cadence of earnings-related news flow and options expiration cycles that periodically elevate stock volatility.
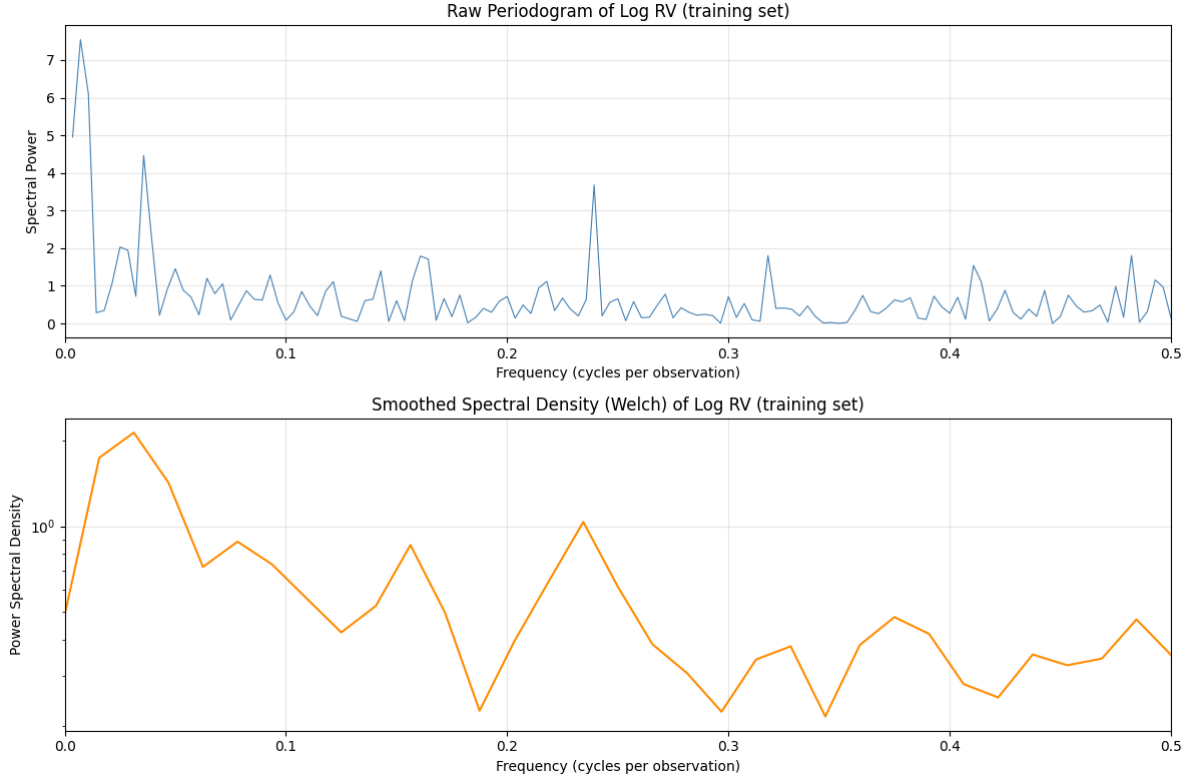
Figure 5: Raw periodogram (top) and Welch-smoothed spectral density (bottom).

4 Model Selection

We search over three model families, all fit on the training set with AIC as the selection criterion.

4.1 ARIMA and SARIMA

We fit ARIMA(p, 0, q) for $p, q \in \{0, \ldots, 5\}$ (36 models) and SARIMA(p, 0, q)(P, 0, Q)$_4$ with $p, q \in \{0, \ldots, 3\}$ and $P, Q \in \{0, 1\}$ (64 additional models). Table 1 shows the top models from both families, ranked by AIC. ARIMA(3,0,3) achieves the lowest AIC at 459.31, consistent with the ACF/PACF evidence for three AR and three MA lags. The seasonal extensions do not improve AIC enough to justify their extra parameters.

Table 1: Top (S)ARIMA models by AIC. ARIMA(3,0,3) leads despite having no seasonal component.

Model	AIC
ARIMA(3,0,3)	459.31
SARIMA(1,0,3)(0,0,1)$_4$	463.20
SARIMA(3,0,3)(0,0,1)$_4$	463.34
ARIMA(4,0,4)	463.92
SARIMA(1,0,3)(1,0,1)$_4$	464.16

Model	AIC
SARIMA(3,0,3)(1,0,0)$_4$	464.28

4.2 SARIMAX grid search

For each SARIMA specification, we additionally include three exogenous configurations: VIX only, Volume only, and VIX + Volume. This yields $64 \times 3 = 192$ additional fits. Table 2 summarizes the top models.

Table 2: Top SARIMAX models by AIC. All top entries use VIX as the sole exogenous variable.

Model	Exogenous	AIC
SARIMAX(3,0,3)(0,0,1)$_4$	VIX	403.28
SARIMAX(3,0,3)(1,0,0)$_4$	VIX	404.23
SARIMAX(3,0,3)(0,0,0)$_4$	VIX	407.83
SARIMAX(1,0,1)(0,0,1)$_4$	VIX	409.92
SARIMAX(1,0,3)(0,0,1)$_4$	VIX	410.81
SARIMAX(1,0,1)(1,0,0)$_4$	VIX	410.86

VIX-only models uniformly outperform Volume-only (best AIC ≈ 721) and VIX+Volume (best AIC ≈ 621) configurations. VIX directly measures market-wide volatility expectations, providing a clean signal. Volume is noisier and adds parameters without commensurate AIC improvement. Including both VIX and Volume worsens AIC relative to VIX alone, suggesting Volume introduces noise when VIX is already present.

4.3 Selected models

We carry forward two models for diagnostics and forecasting:

- **Model A**: ARIMA(3,0,3) — best pure time series model (AIC = 459.31)
- **Model B**: SARIMAX(3,0,3)(0,0,1)$_4$ with VIX — best exogenous model (AIC = 403.28)

The AIC gap of ≈ 56 strongly favors Model B.

5 Model Interpretation

5.1 ARIMA(3,0,3)

This model says the current Log RV depends on the past 3 values (AR terms) and the past 3 forecast errors (MA terms). The AR(3) component captures the short-term persistence: if volatility was elevated recently, it tends to remain so. The MA(3) component allows the model to correct for recent surprises. All AR and MA coefficients except ar.L2 are significant at the 5% level.

5.2 SARIMAX(3,0,3)(0,0,1)$_4$ [VIX]

This model adds two ingredients. First, VIX enters as a linear regressor with coefficient $\hat{\beta} \approx 0.041$ ($p < 0.001$), meaning each 1-point increase in VIX raises predicted Log RV by about 0.04 — a sensible relationship since elevated market fear corresponds to higher realized stock volatility. Second, the seasonal MA(1) at lag 4 captures monthly-scale corrections: if the model's forecast error from 4 weeks ago was large, it still influences today's prediction.

6 Diagnostics

6.1 Residual tests

Table 3 summarizes the diagnostic tests on both fitted models.

Table 3: Residual diagnostic tests for both models.

Test	ARIMA(3,0,3)	SARIMAX(3,0,3)(0,0,1)$_4$ [VIX]
Ljung-Box (lag 10)	$Q = 3.10$, $p = 0.979$	$Q = 6.09$, $p = 0.807$
Jarque-Bera	$JB = 2.93$, $p = 0.231$	$JB = 26.20$, $p < 0.001$
ARCH LM (5 lags)	$LM = 1.97$, $p = 0.854$	$LM = 0.96$, $p = 0.966$

Both models pass the Ljung-Box test, confirming no significant residual autocorrelation (11). Neither shows ARCH effects, suggesting that residual heteroskedasticity is not a major concern. The ARIMA model also passes the Jarque-Bera normality test, while the SARIMAX model shows mild non-normality (excess kurtosis of 3.87), likely driven by a few large residuals associated with the VIX regressor during extreme market events.

6.2 Residual plots

Figure 6 and Figure 7 show ACF, QQ, and histogram plots of the residuals. The ARIMA residuals are well-behaved. The SARIMAX residuals show slightly heavier tails in the QQ plot, consistent with the Jarque-Bera rejection, but the ACF is clean.
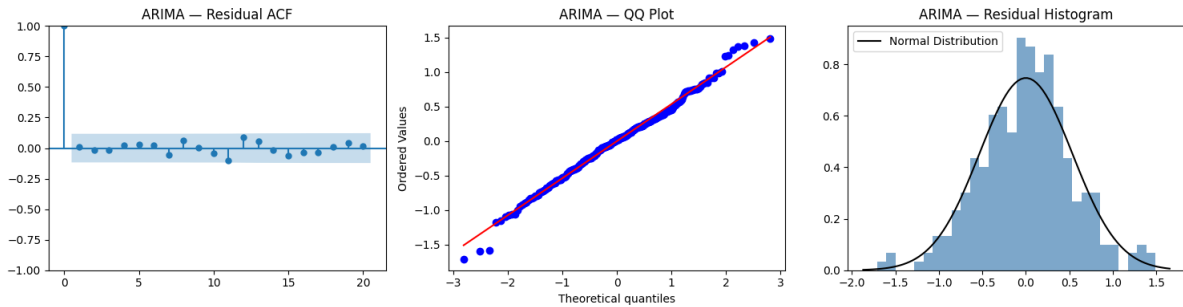

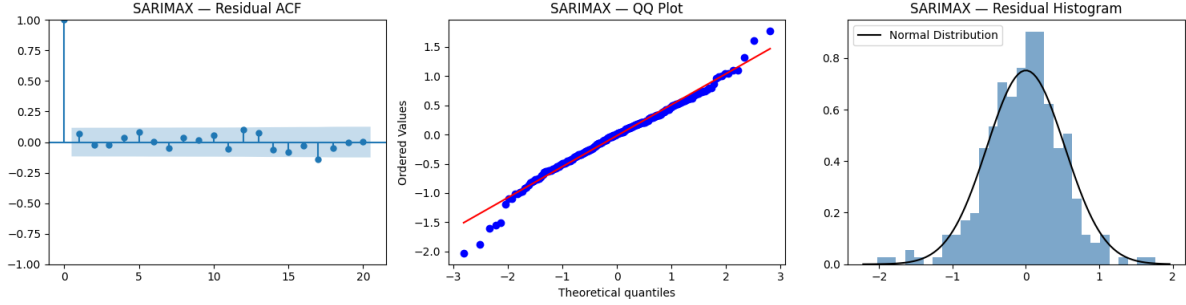Figure 6: Residual diagnostics for ARIMA(3,0,3): ACF, QQ plot, and histogram.

Figure 7: Residual diagnostics for SARIMAX$(3,0,3)(0,0,1)_4$ with VIX.

7 Out-of-Sample Evaluation

We produce one-step-ahead rolling forecasts on the 71-observation test set. At each step, the model is refit on all data observed so far (training + previously revealed test observations) and a single forecast is issued for the next period.

7.1 Forecast accuracy

Table 4 compares the two models on the test set.

Table 4: Out-of-sample forecast metrics. Bold indicates the better model.

Metric	ARIMA(3,0,3)	SARIMAX [VIX]
RMSE	0.6222	**0.5723**
MAE	0.4662	**0.4238**
MAPE (%)	43.96	**39.67**
Test R^2	-0.093	**0.076**
Train R^2	0.199	**0.213**

The SARIMAX model reduces RMSE by 8% and MAE by 9% relative to the pure ARIMA. The ARIMA model achieves a negative test R^2, meaning it performs worse than simply predicting the test-set mean, while the SARIMAX model achieves a small but positive test R^2 of 0.076. These low R^2 values are not an anomaly but rather a direct implication of standard volatility models, as Andersen and Bollerslev (12) demonstrated. Realized volatility contains a large unpredictable component driven by news arrivals, earnings surprises, and sudden shifts in sentiment, so even well-specified models typically achieve out-of-sample R^2 in the range of 0.05 to 0.15 at weekly horizons. In that context, the SARIMAX result sits comfortably within the expected range and confirms that VIX provides genuine out-of-sample predictive content.

7.2 Forecast plots

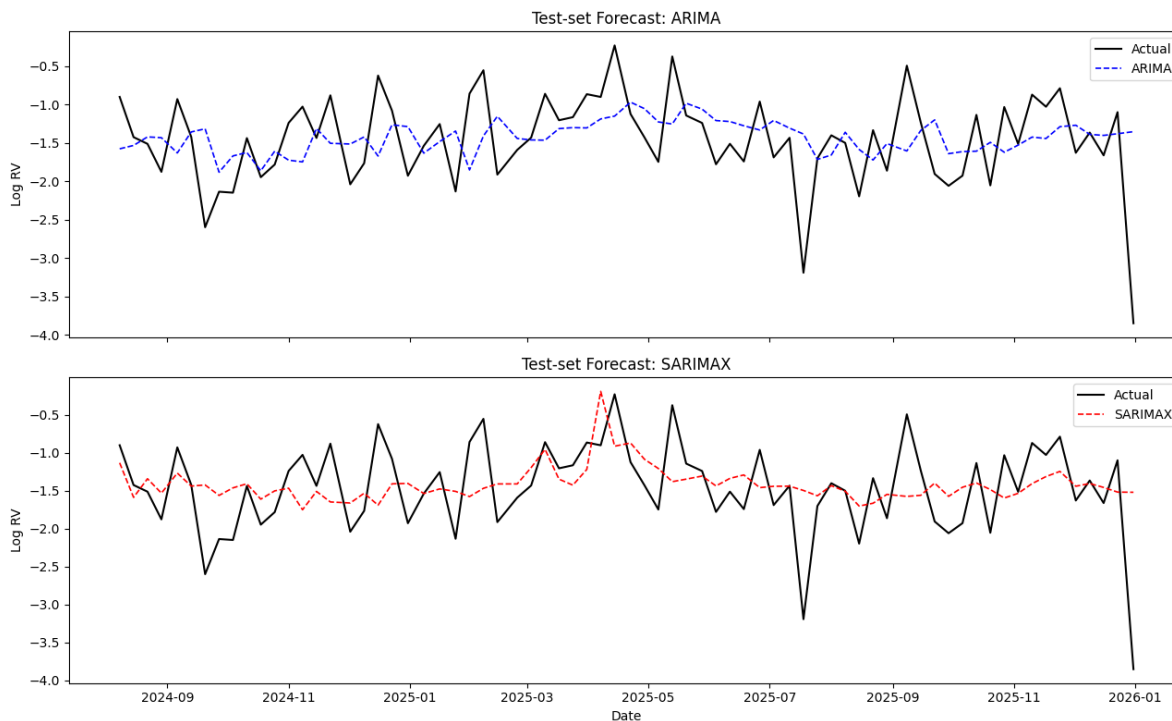Figure 8 overlays the actual and predicted Log RV on the test period for both models.

Figure 8: One-step-ahead rolling forecasts on the test set. Top: ARIMA(3,0,3). Bottom: SARIMAX$(3,0,3)(0,0,1)_4$ with VIX.

Both models track the general level and broad swings, but the SARIMAX model is visibly more responsive to sharp movements — precisely because VIX provides a contemporaneous signal of market stress that the ARIMA model, relying solely on past Log RV, cannot access.

8 Conclusions

We draw three main conclusions from this analysis.

First, **ARIMA(3,0,3) provides a reasonable baseline** for weekly GOOG Log RV. The ACF/PACF structure, AIC selection, and clean residual diagnostics all support this specification. However, its out-of-sample performance is weak: it barely improves upon — and by R^2 actually underperforms — a naive mean forecast.

Second, **VIX is a powerful exogenous predictor**. Among all exogenous configurations tested, VIX-only models consistently achieved the lowest AIC by a wide margin. Volume provided negligible additional information. This aligns with the interpretation that market-wide fear, not idiosyncratic trading activity, drives the forecastable component of Google's volatility.

Third, **the SARIMAX(3,0,3)(0,0,1)$_4$ model with VIX is the best overall model**, achieving the lowest AIC (403.28), the best test-set RMSE (0.572), and the only positive out-of-sample R^2. The seasonal MA term at lag 4 contributes a modest monthly correction. The VIX coefficient of ≈ 0.04 is highly significant and economically interpretable.

These results suggest that for practical volatility forecasting of individual stocks, incorporating a market-wide volatility measure like VIX is more valuable than increasing the complexity of the univariate time series model. Future work could explore GARCH-type models for the conditional variance of residuals, or test whether intraday realized volatility measures (using 5-minute returns) improve upon our daily-close-based estimates.

Acknowledgments

Bibliography

[1] Robert H Shumway and David S Stoffer. *Time Series Analysis and Its Applications: With R Examples.* Springer, 4th edition, 2017.

[2] James D Hamilton. *Time Series Analysis.* Princeton University Press, 1994.

[3] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[4] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[5] Robert E Whaley. The investor fear gauge. *The Journal of Portfolio Management*, 26(3):12–17, 2000.

[6] Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.

[7] Clive WJ Granger and Zhuanxin Ding. Varieties of long memory models. *Journal of Econometrics*, 73(1):61–77, 1996.

[8] Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54 (1–3):159–178, 1992.

[9] Edward L. Ionides. Stats 531 (winter 2026) course notes: Analysis of time series. https://ionides.github.io/531w26/, 2026.

[10] Peter Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[11] Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[12] Torben G Andersen and Tim Bollerslev. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905, 1998.