Can Machine Learning Predict Rivian Returns?

ARIMA · ARIMAX · GARCH Family · Random Forest

Theophile Muller

ECO508

Prof. Jin Man Lee | DePaul University | Spring 2026

Sample: Jan 3, 2024 - Jun 12, 2026 | 613 daily observations

1. Introduction

Rivian Automotive (RIVN) completed its IPO in November 2021. Following a post-IPO repricing that drove the stock from its peak of around 130$ to around 19$ by early 2024, RIVN entered a stabilized trading range from January 2024 onward fluctuating between 25$ and 7.5$. This paper focuses on that post-stabilization period starting on January 3, 2024 through June 12, 2026, this means that we have 613 daily observations.

In this paper we look at: *does a Random Forest model with engineered macro features outperform ARIMA, ARIMAX, and a naive random walk benchmark in forecasting daily RIVN log returns out-of-sample?* A secondary question asks whether any model predicts the direction of daily returns better than chance (while this wasn't what we started with, it ended up giving some of the most interesting results). Over the 613-day sample, RIVN had a daily standard deviation of 4.33% and an annualized volatility of 68.8%, this is around 4.6 times more than the S&P 500. Excess kurtosis of 7.29 confirms heavy-tailed, non-normal returns.

Four methods were used: ARIMA is used as the classical benchmark. ARIMAX with TSLA returns and VIX changes as exogenous regressors. GARCH family gives us the volatility structure and gives us volatility estimates for the Random Forest model. A Random Forest trained on a lag-based feature. All models were ranked based on a 126-day out-of-sample RMSE/MAE and directional accuracy.

The main findings are: ARIMA selected ARIMA(0,0,0) this means that there are no own-return autocorrelation; ARIMAX with TSLA and VIX saw AIC become better by 79.89 units, with TSLA coefficient being .294 and VIX coefficient being -.090; Random Forest got the best RMSE and the best directional accuracy, this is very interesting results considering that the test set is nearly 50/50 up and down days; GARCH(1,1) estimates persistence of .996 with a half-life of 183 trading days, and EGARCH got the lowest AIC.

2. Literature Review with Paper Summaries

2.1 Engle (1982) : ARCH

Research question and motivation: Before Engle, most standard econometric models thought that the variance of the error term was constant over time. This was a problem because in many economic and financial series, large errors tend to cluster together, meaning that the variance of the error term was indeed not constant over time. Engle asked whether this changing conditional variance could be modeled in a formal way.

Data and methodology: Engle introduced the ARCH(q) model and applied it to quarterly United Kingdom inflation data. He showed how the model could be estimated using maximum likelihood and also introduced the ARCH-LM test, which tests whether past squared errors help explain current variance.

Main findings: Engle found strong evidence that the variance of UK inflation changed over time. The ARCH model was useful because it allowed the forecast variance to depend on past information instead of assuming constant variance.

Relevance: This paper is important for my model because it gives the foundation for testing and modeling volatility grouping. In my 2024 to 2026 RIVN sample, the ARCH-LM test gives chi-sq = 6.80 and p = 0.236, meaning it does not reject the null of no ARCH effects at the 5% level. However, I still estimate GARCH-type models because the returns show excess kurtosis of 7.29 and visual signs of volatility grouping.

2.2 Bollerslev (1986) : GARCH

Research question and motivation: Engle’s ARCH model was useful, but it most of the time needed many lags to capture volatility persistence. Bollerslev asked whether volatility could be modeled in a simpler and more flexible way by including past conditional variances.

Data and methodology: Bollerslev extended ARCH into the GARCH(p,q) model:

$$\sigma^2_t = \omega + \Sigma\alpha_i\varepsilon^2_{t-i} + \Sigma\beta_j\sigma^2_{t-j}$$

This allowed current volatility to depend on both past squared shocks and past conditional variance estimates.

Main findings: GARCH provides a more flexible and more parsimonious version of ARCH. Instead of needing many ARCH lags, in fact a simple GARCH(1,1) model can capture a majority of the volatility in financial and economic time series.

Relevance: GARCH(1,1) is the main volatility model used in this project. The estimated persistence is very high, with an $\hat{\alpha}$ of .009 and a $\hat{\beta}$ of .987, giving $\hat{\alpha} + \hat{\beta} = .996$. This means volatility shocks decay very slowly. Using the rounded coefficients, the volatility half-life is about 183 trading days. The estimated conditional volatility $\hat{\sigma}_t$ is then used as a feature in the Random Forest, which connects the GARCH model to the machine learning part of the project.

2.3 Nelson (1991) : EGARCH

Research question and motivation: Standard GARCH models treat positive and negative shocks the same way because they use squared errors. They also require non-negative parameter restrictions to make sure that the variance continues to be positive. Nelson proposed EGARCH to solve these issues.

Data and methodology: EGARCH models the log of variance instead of variance directly. This means the model does not need the same non-negative restrictions as standard GARCH. It also includes an asymmetry term, allowing negative and positive shocks to have different effects on future volatility.

Main findings: EGARCH is useful for asset returns because it can capture asymmetric volatility responses. This matters because bad news often increases volatility more than good news like with oil prices, which is typically called the leverage effect.

Relevance: EGARCH has the lowest AIC among the three GARCH-type models estimated in this project, with an AIC of -3.6486. The EGARCH asymmetry coefficient is statistically significant, with a γ_1 of .312 and a p of .005. This shows that RIVN volatility responds differently to positive and negative shocks. Under the sign convention used in this model, this is interpreted as evidence of a leverage effect for RIVN.

2.4 Breiman (2001) : Random Forests

Research question and motivation: Decision trees can be useful, but individual trees have a tendency to overfit. Bagging improves this by using many trees and then averaging them out, but the trees can still be highly correlated. Breiman asked whether randomly selecting predictors at each split could reduce correlation between trees and improve prediction.

Data and methodology: Random Forest builds many decision trees using bootstrapped samples of the data. At each split, the model only considers a random subset of predictors instead of all predictors. This helps make the trees less correlated with each other. The model also produces out-of-bag error estimates and variable importance measures.

Main findings: Random Forests often perform very well because they combine many weak or moderate trees into one stronger model. Breiman also showed that as more trees are added, the generalization error converges instead of overfitting in the usual way. Random Forests performed well compared to AdaBoost and were more robust to noise.

Relevance: Random Forest is the best-performing model in this project. It is trained on 298 observations and achieves an OOB R^2 of 12.5%, an out-of-sample RMSE of .0305, and directional accuracy of 64.3%.

2.5 Meese & Rogoff (1983) : Random Walk Benchmark

Research question and motivation: Meese and Rogoff asked whether exchange rate models could are better than a simple random walk forecast out of sample. This was very important because it tested whether more complicated models actually better forecasting performance.

Data and methodology: They compared several structural exchange rate models with a random walk benchmark using dollar exchange rates from the 1970s and early 1980s. The key part of the study was that the models were judged based on out-of-sample forecasting performance.

Main findings: The random walk performed as well as, or better than, the structural models in out-of-sample forecasts. This became known as the Meese-Rogoff puzzle and showed how difficult it is to beat a simple benchmark in financial forecasting.

Relevance: This paper is important because it shows why my models need to be compared against a naive benchmark, not just judged by their own performance. In this project, both ARIMAX and Random Forest beat the naive benchmark on RMSE. ARIMAX improves RMSE by 9.5%, while Random Forest improves RMSE by 33.2%. The directional accuracy results add another layer. Since the sample mean is slightly negative, it predicts negative returns every day and only achieves 49.2% directional accuracy. Random Forest reaches 64.3%, which is a very large improvement.

2.6 Muth (1960) : Optimal Forecasting

Research question and motivation: Muth asked a theoretical question: under what conditions is the simple practice of giving more weight to recent observations actually the mathematically optimal thing to do? This basically connects the intuitive idea of exponential smoothing to formal statistics.

Data and methodology: Muth showed that exponentially weighted forecasts can be interpreted as optimal when the series has both a permanent component and a transitory noise component. This is closely related to the IMA(1,1) process.

Main findings: Exponential smoothing works best when the data follow the type of permanent-transitory structure that Muth describes. The paper helps explain why simple smoothing methods can work well in some forecasting problems, but not all.

Relevance: Muth is useful for this project because it connects simple forecasting models to the ARIMA framework. In my results, the selected ARIMA model is ARIMA(0,0,0), which means the ARIMA-only model does not find useful autoregressive or moving-average structure beyond the sample mean. However, ARIMAX improves AIC by 79.89, which means that the macro and market covariates give us the main predictive structure. In other words, the improvement does not come from the time-series structure alone, but from the outside variables added to the model.

3. Data

3.1 Variables and Sources

Four daily time series were retrieved from Yahoo Finance using quantmod. The main series being RIVN daily closing prices from January 3, 2024 through June 12, 2026. Three exogenous series are included: TSLA daily log returns representing the EV sector in general, ^VIX log changes as a market fear measure, and ^TNX first representing as an interest rate.

3.2 Transformations and Stationarity

The log returns were created for RIVN, TSLA, and VIX. For TNX, the first difference ΔTNX is taken. All three unit root tests show that they are stationarity: ADF (-8.48, $p < 0.01$), KPSS (.154, $p > 0.10$), and PP (-554.6, $p < 0.01$).

3.3 Summary Statistics

The mean daily return is -.04%. Daily SD of 4.33% annualizes to 68.8%, that is 4.6 times more volatile than the S&P 500 (if we had used all of RIVN's stock history it would have been 7x more, however we chose not to include the first two years as RIVN had a rough start after their IPO). Excess kurtosis of 7.29 confirms heavy-tailed non-normal returns. The Anderson-Darling test rejects normality at $p < 0.001$.

Table 1: Summary Statistics RIVN Daily Log Returns (Jan 2024 - Jun 2026)

Statistic	Value	Interpretation
Mean daily return	-.04%	Near-zero drift in stabilized period
Std deviation	4.33%	Ann. vol = 68.8% (4.6× S&P 500)
Minimum (worst day)	-29.57%	Feb 22, 2024 this was production guidance cut
Maximum (best day)	+23.62%	Feb 13, 2026 this was after very good Q4 results
Skewness	.32	Slight right skew
Excess kurtosis	7.29	Very heavy tails
Anderson-Darling p	≈ .000	Strongly non-normal
Ann. volatility	68.8%	4.6× S&P 500

Figure: RIVN Daily Closing Price (Jan 2024 - Jun 2026)

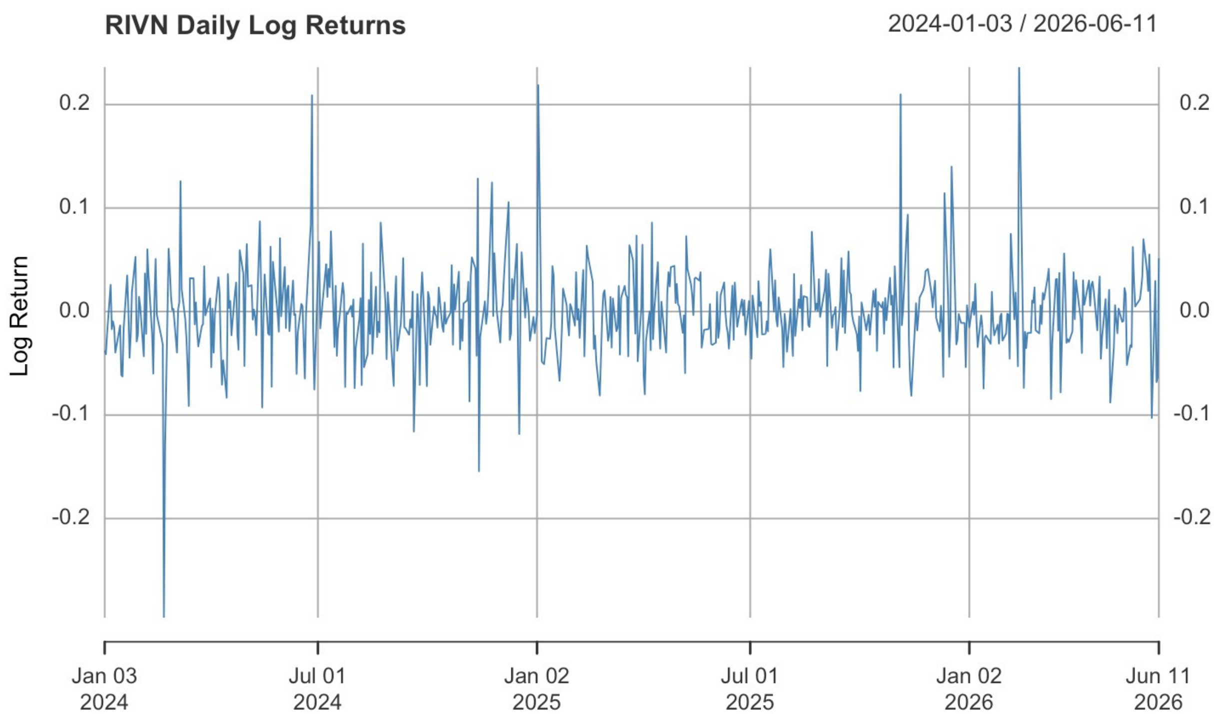

Figure: RIVN Daily Log Returns (Jan 2024 - Jun 2026)

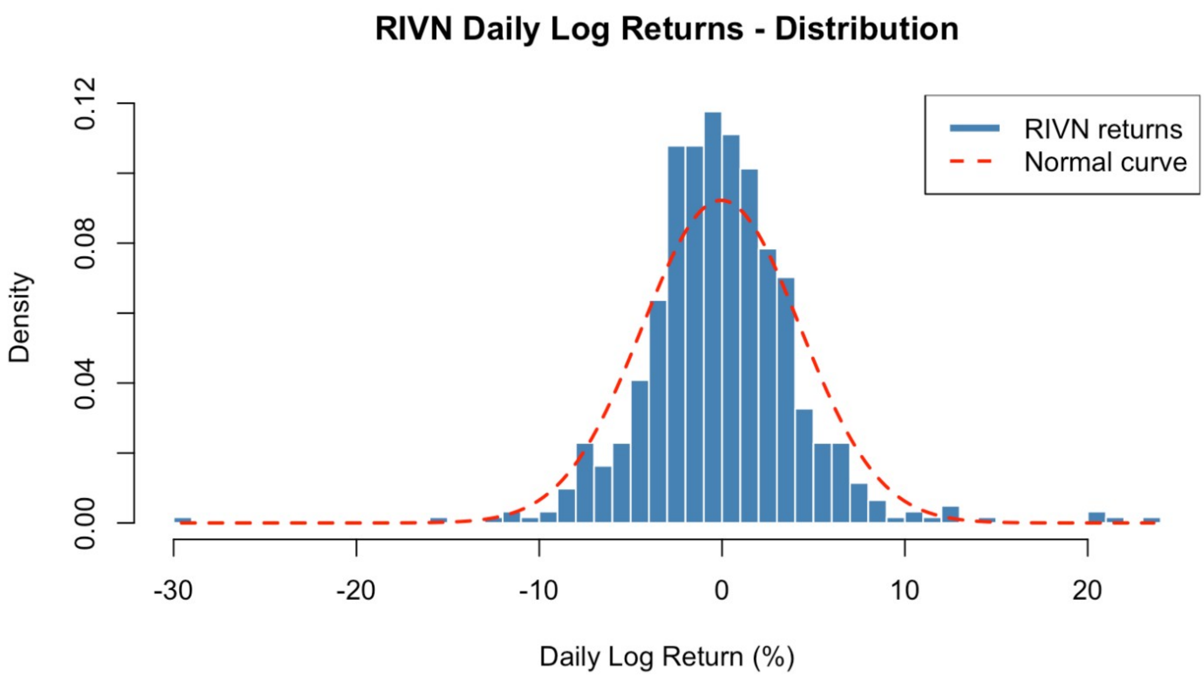

Figure: RIVN Daily Return Distribution with Normal Overlay

3.4 Correlation Structure

RIVN-TSLA correlation is .36 (this makes sense as there are many EV focused ETFs with TSLA being a bigger part of the pie). RIVN-VIX is -.33 (market fear hurts RIVN). TNX yield changes show weak negative correlation (-.09).

Table 2: Correlation Matrix

	RIVN	TSLA	VIX	ΔTNX
RIVN	1.00	0.36	-0.33	-0.09

TSLA	0.36	1.00	-0.50	+0.01
VIX	-.33	-.50	1.00	-0.02
ΔTNX	-0.09	+0.01	-0.02	1.00

3.5 Train/Test Split

We ended up using 487 training days (Jan 2024 - Dec 2025) and 126 test days (Jan-Jun 2026). After lag engineering the Random Forest feature matrix contains 424 rows, of which 298 train the model and 126 form the test set.

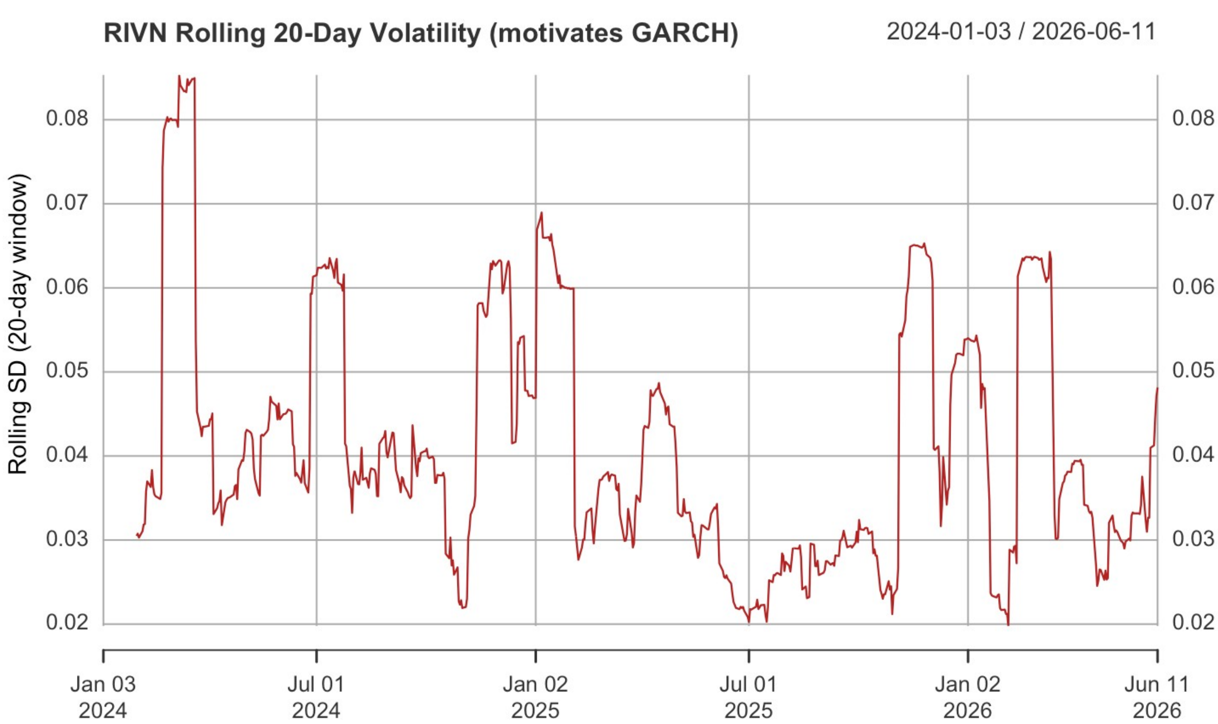

Figure: RIVN Rolling 20-Day Volatility

4. Methodology

4.1 ARIMA

The ARIMA(p, d, q) model: $\Phi(B)(1-B)^d y_t = \theta_0 + \Theta(B)\varepsilon_t$. This model uses ACF/PACF plots and six models ranked by AIC. The naive random walk benchmark sets $\hat{y}_{t+1} = y_t$. The performance is then calculated based on the 126-day test set using RMSE, MAE, and directional accuracy.

4.2 ARIMAX

ARIMAX goes deeper than the simple ARIMA through the addition of exogenous regressors: $y_t = \beta_1 \, TSLA_t + \beta_2 \, VIX_t + N_t$, where N_t is the ARIMA noise model. TSLA daily log returns and VIX daily log changes are used as regressors. The model is selected by auto.arima() with xreg.

4.3 GARCH Family

GARCH(1,1): $\sigma^2_t = \omega + \alpha_1\varepsilon^2_{t-1} + \beta_1\sigma^2_{t-1}$. Persistence = $\alpha_1 + \beta_1$.

TGARCH: $\sigma^2_t = \omega + \alpha_1\varepsilon^2_{t-1} + \gamma_1\varepsilon^2_{t-1}\mathbb{1}[\varepsilon_{t-1}<0] + \beta_1\sigma^2_{t-1}$. Positive γ_1 = leverage effect.

EGARCH: $\log \sigma^2_t = \omega + \alpha \, g(z_{t-1}) + \beta \log \sigma^2_{t-1}$. Log-variance guarantees positivity without constraints.

The Engle-Ng sign bias test is applied to GARCH(1,1) standardized residuals. GARCH(1,1) conditional volatility $\hat{\sigma}_t$ is incorporated as a Random Forest feature.

4.4 Random Forest

We used own lags (lag5, lag21, lag63), volatility and momentum (rolling 20-day SD, MA20, MA63), macro covariates (TSLA return, VIX change, ΔTNX, GARCH sigma), and a weekday dummy. After the addition of the lag we now have 424 rows, 298 training, 126 testing. ntree = 500, m = 3. Rolling-window CV (minimum 180 days, 244 folds).

4.5 Directional Accuracy Analysis

Alongside the RMSE and MAE, all four models are ranked based on directional accuracy, this is the percentage of test days where the predicted sign of the return matches the actual sign. We show this as: overall hit rate, up-day accuracy (precision when predicting positive returns), and down-day accuracy (precision when predicting negative returns). The baseline is 50% (random guessing). The actual test set is 50.8% up days, so a model always predicting ‘up’ would achieve 50.8% hit rate.

4.6 Model Comparison Framework

All four models are compared on the same 126-day test set. RMSE and MAE measure magnitude accuracy of the daily returns. Directional accuracy measures sign prediction. Naive random walk serves as the baseline on both dimensions as shown in Meese and Rogoff (1983).

5. Results

5.1 Stationarity Tests

Table 3: Stationarity Tests RIVN Daily Log Returns

Test	Statistic	p-value	Conclusion
ADF (H_0: unit root)	-8.48	< 0.01	Stationary
KPSS (H_0: stationary)	.154	> 0.10	Stationary
PP (H_0: unit root)	-554.6	< 0.01	Stationary

5.2 ARIMA Results

ARIMA(0,0,0) got the lowest AIC (-1677.85) out of the other six models. Ljung-Box got Q(252) = 275.9 (p = .144) this confirms white noise residuals. The ARIMA forecast is a flat line at the sample mean (-.04%), predicting negative returns every single test day because the mean is slightly negative. Because of this ARIMA gets 0% up-day accuracy and 49.2% overall hit rate as this was the percentage of down days, basically no directional forecasting capabilities.

Table 4: ARIMA Model Selection (Training Set, n = 487)

Model	AIC	BIC

ARIMA(0,0,0) (later selected)	-1677.85	-1669.47
ARIMA(0,0,2)	-1676.89	-1660.13
ARIMA(2,0,0)	-1676.85	-1660.10
ARIMA(0,0,1)	-1676.83	-1664.27
ARIMA(1,0,0)	-1676.70	-1664.14
ARIMA(1,0,1)	-1675.73	-1658.98

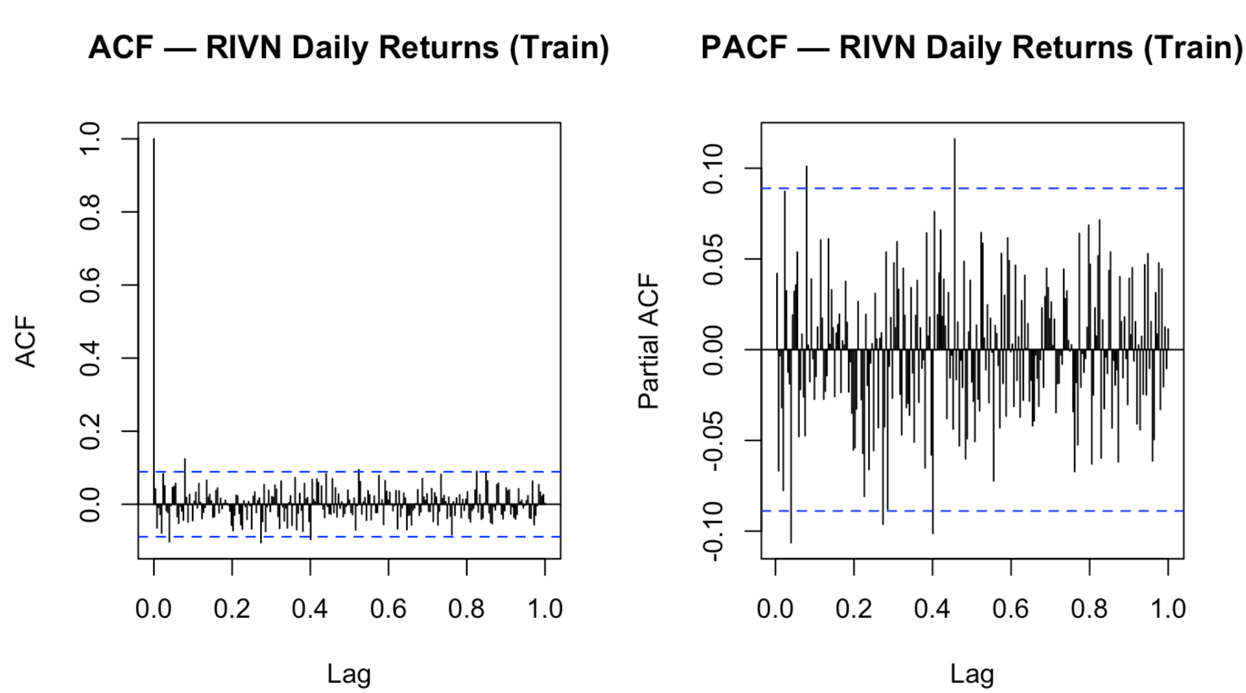

Figure: ACF and PACF RIVN Daily Returns (Training Set)

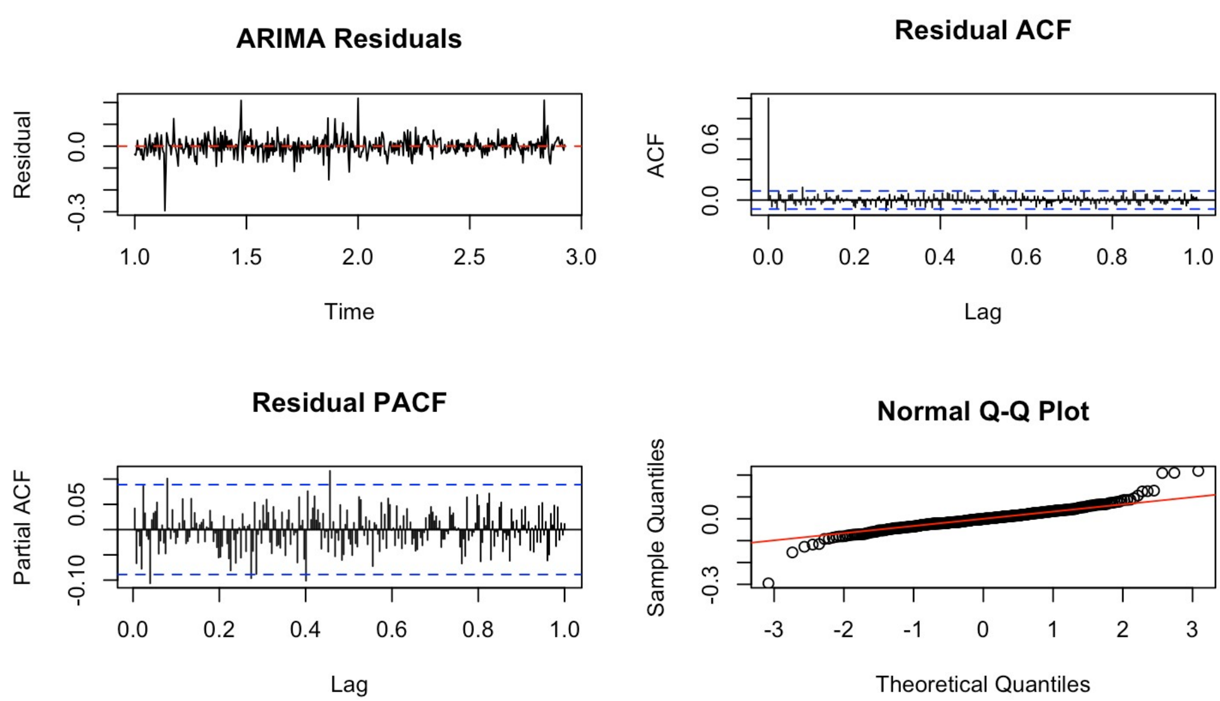

Figure: ARIMA(0,0,0) Residual Diagnostics

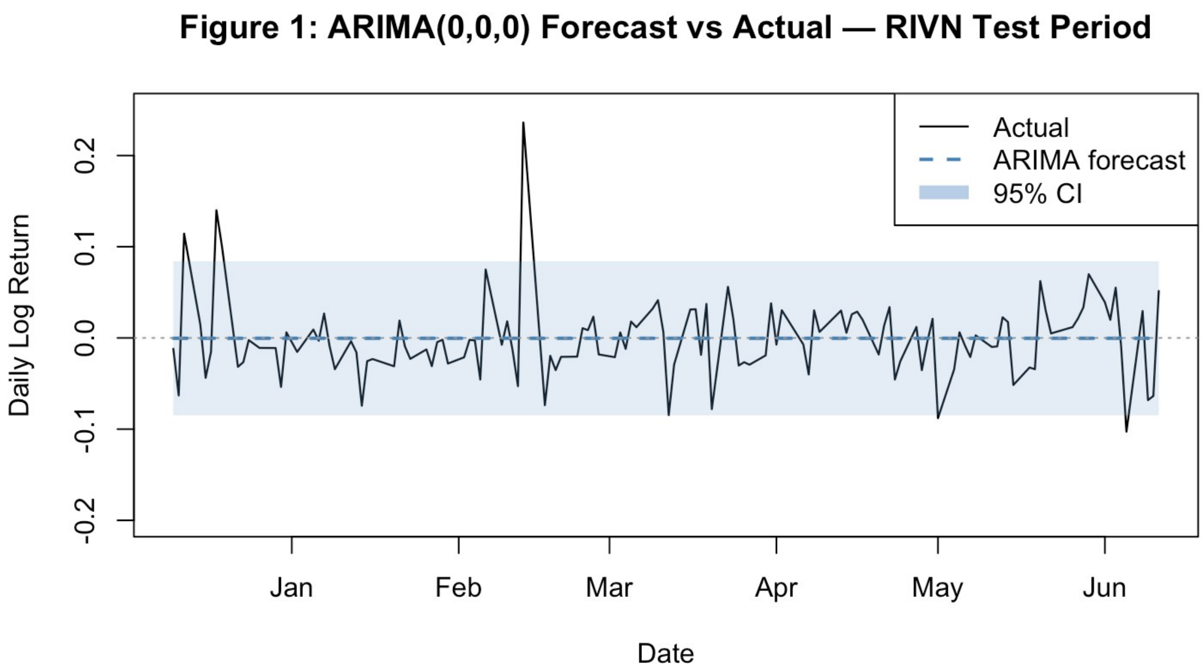

Figure 1: ARIMA(0,0,0) Forecast vs Actual RIVN Test Period

5.3 ARIMAX Results

ARIMAX selects ARIMA(0,0,0) errors with TSLA and VIX as exogenous regressors. The AIC improves by 79.89 units going from -1677.85 to -1757.74, this tells us that the variables bring a significant linear predictive signal, much better than what any ARIMA structure on its own returns.

Table 5: ARIMAX Exogenous Coefficient Estimates

Variable	Estimate	Std.	t-	Interpretation

		Error	value	
TSLA return	.2941	.0518	5.67	EV sector co-movement and is highly significant
VIX change	-.0896	.0242	-3.70	Market fear hurts RIVN and is highly significant

Both coefficients are significant at the one percent level. TSLA coefficient of .294: on a day TSLA rises 10%, RIVN is predicted to rise around 2.94%. VIX coefficient of -.090: a 10% increase in VIX is associated with an around .90% decrease in RIVN. Ljung-Box residuals are white noise (p=.364). ARIMAX predicts both up and down days (65 predicted up, 61 predicted down), achieving 51.6% overall hit rate and 52.3% up-day accuracy, this is slightly above the 50.8% always-up baseline, meaning that the directional accuracy isn't that much better than a simple coinflip.

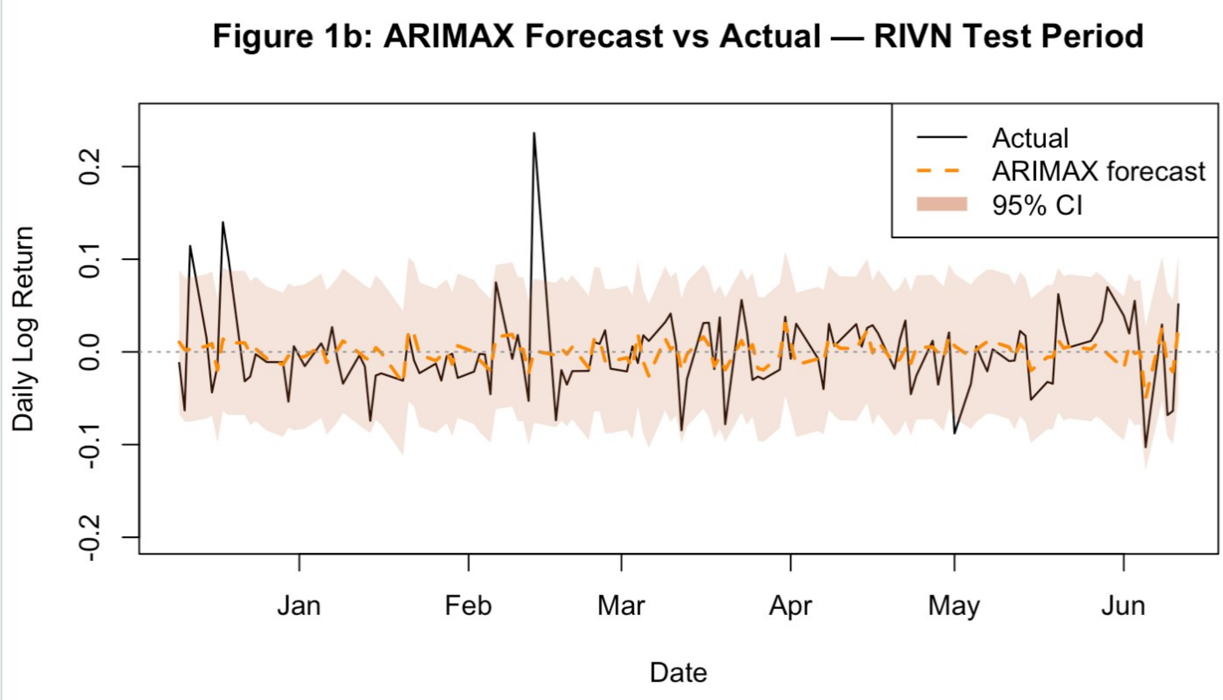

Figure 1b: ARIMAX Forecast vs Actual RIVN Test Period

The out-of-sample had an RMSE of .0404 and an MAE of .0275. ARIMAX outperforms both ARIMA (-8.6% RMSE) and the naive benchmark (∑9.5% RMSE) which shows that while the predictions aren’t the best they are still better.

5.4 GARCH Results

Although the ARCH test does not show volatility grouping at the five percent significance level, we still estimate GARCH models because the return distribution has extremely heavy tails as the excess kurtosis is 7.29 and the rolling volatility plot shows clearly that there are periods of calm and periods of turbulence alternating throughout the sample (this can be explained by the fact that these heavy tails come from the news such as quarterly expectation being beat or new contracts such as with Amazon). The GARCH(1,1) model finds that volatility shocks are highly persistent, once RIVN becomes volatile, it stays volatile for a long time. Specifically, a shock to volatility takes approximately 183 trading days, or about nine calendar months, to decay halfway back to its normal level.

Table 6: GARCH(1,1) Parameter Estimates (Training Set, n = 487)

Parameter	**Estimate**	**Std. Error**	**t-statistic**	**p-value**
μ (mean)	-.000880	.001576	-.56	.577
ω (const.)	.000004	.000006	0.78	.438
α_1 (ARCH)	.008781	.002997	2.93	.003
β_1 (GARCH)	.987438	.002315	426.5	< .001
shape (ν)	4.246761	.720530	5.89	< .001

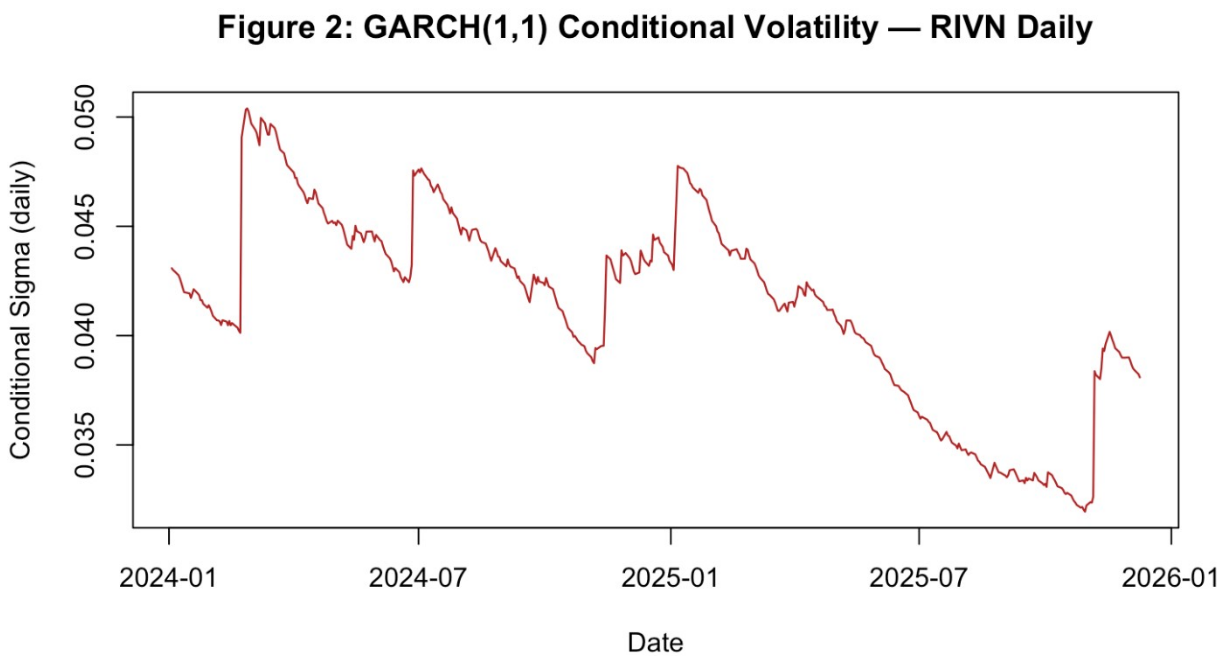

Figure 2: GARCH(1,1) Conditional Volatility RIVN Daily (Training Period)

Table 7: GARCH Family Model Comparison

Model	AIC	BIC	Persistence (α+β)
GARCH(1,1)	-3.6456	-3.6026	.9962
TGARCH (GJR)	-3.6426	-3.5910	.9953
EGARCH	-3.6486	-3.5970	N/A

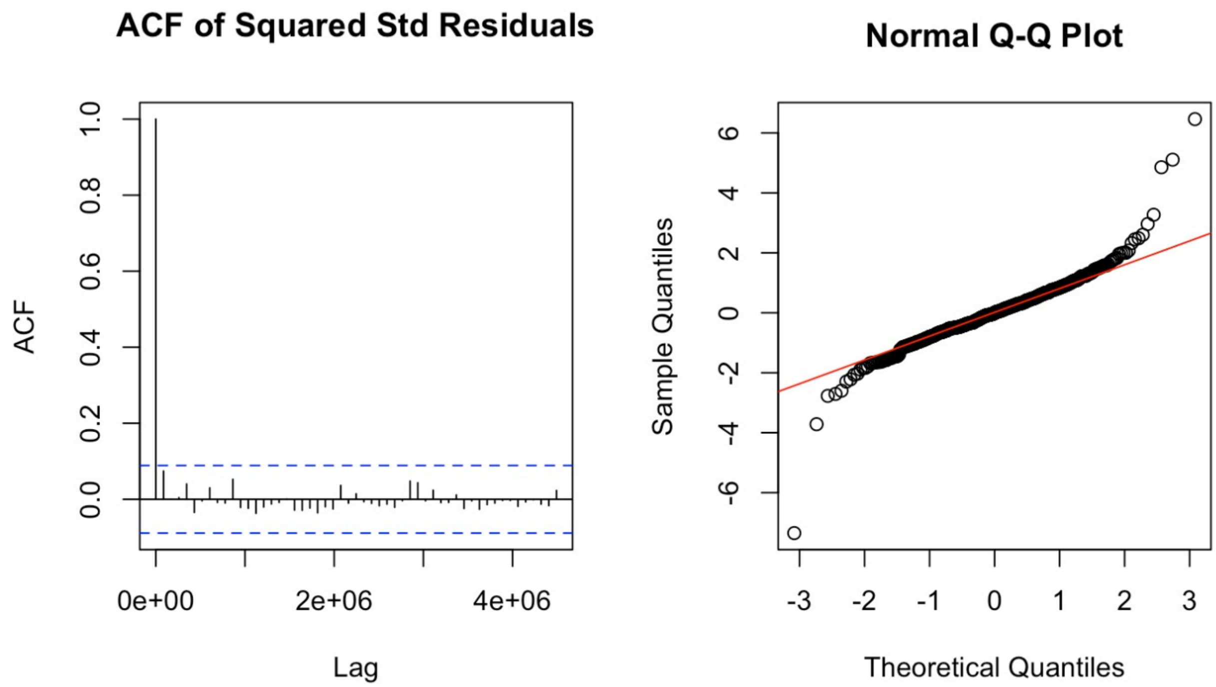

Figure: GARCH(1,1) Residual Diagnostics

EGARCH has the lowest AIC (-3.6486) and EGARCH γ_1 = .312 (p = .005) this tells us that bad news hits RIVN harder than good news. The 95% VaR from GARCH averages approximately -6.2% per day over the test period.

5.5 Random Forest Results

The Random Forest explains 12.5% of OOB training variance (R^2 = .125). Table 8 shows variable importance. VIX changes and TSLA returns are the top two predictors, this is consistent with the ARIMAX coefficient table. GARCH sigma ranks third, meaning that we do benefit by connecting GARCH and RF.

Table 8: RF Variable Importance (%IncMSE)

Feature	%IncMSE	Interpretation
vix_ret	13.58	Market fear is the top predictor
tsla_ret	13.21	EV sector co-movement nearly tied
garch_sigma	6.91	GARCH vol regime carries signal
ma20	6.81	Short-term momentum informative
ma63	5.62	Quarterly momentum also useful
roll_vol20	3.76	Realized vol adds modest value
lag5	1.72	1-week lag marginally useful
lag21	.79	1-month lag near irrelevant
lag63	-1.33	1-quarter lag counterproductive
weekday	-2.06	No day-of-week effect
dgs10	-2.70	Rate changes not informative

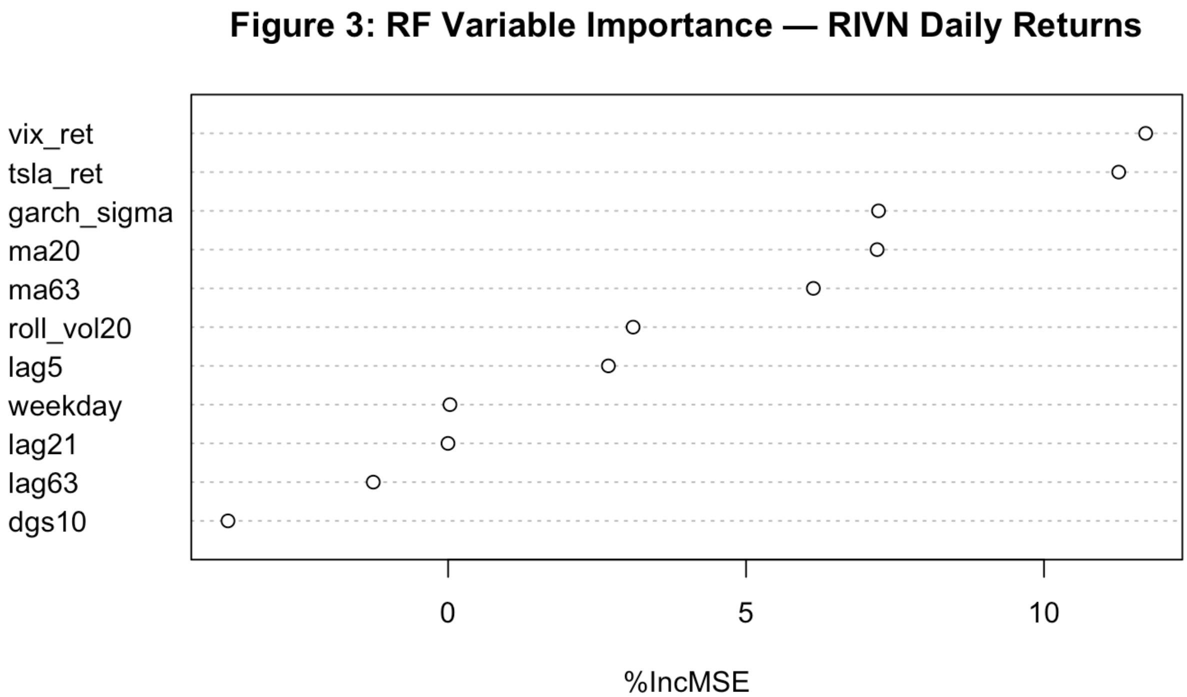

Figure 3: RF Variable Importance RIVN Daily Returns

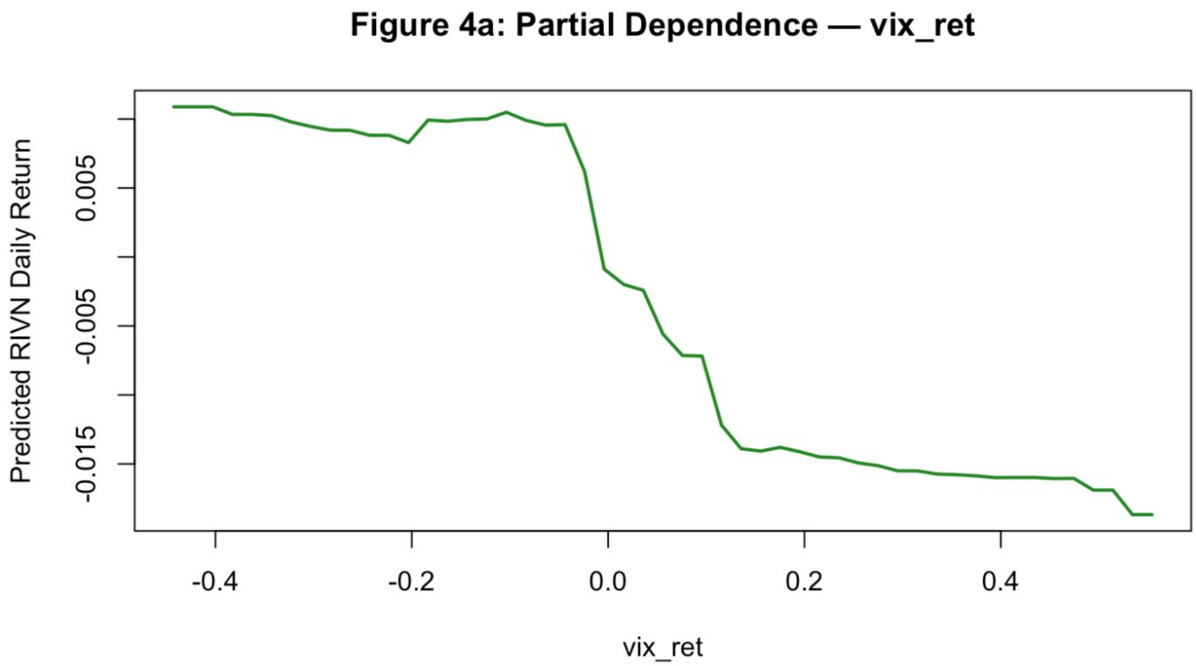

Figure 4a: Partial Dependence vix_ret

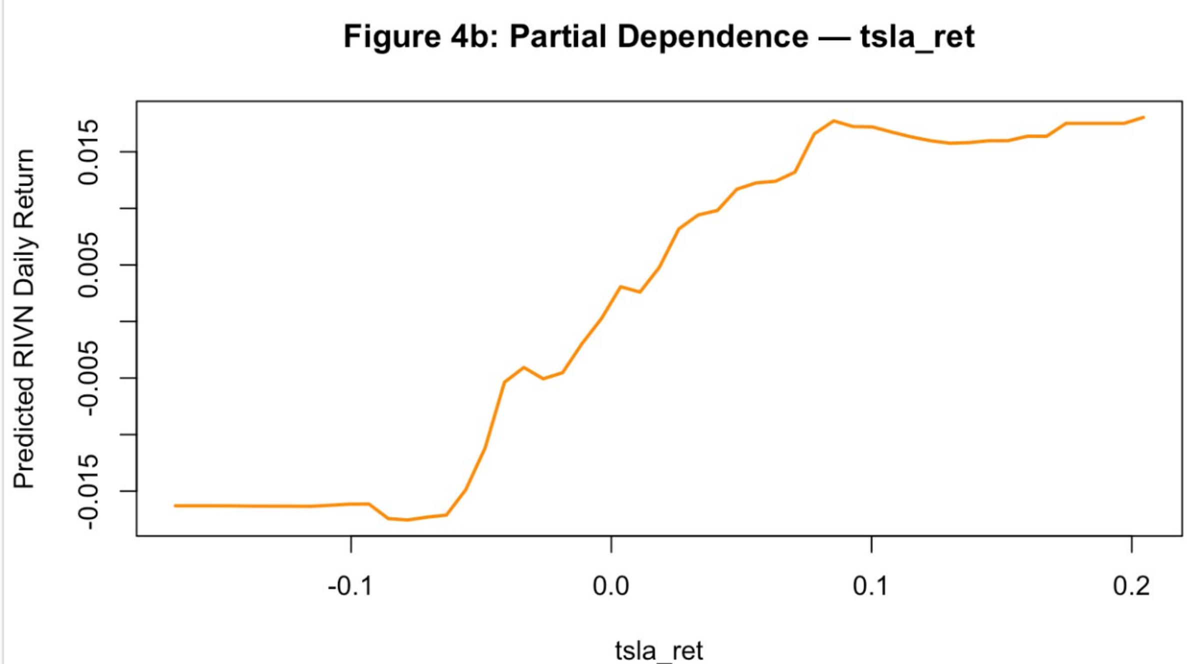

Figure 4b: Partial Dependence tsla_ret

The rolling-window CV has a CV-RMSE of .0336 and a CV-MAE of .0234 across 244 folds. The out-of-sample has an RMSE of .0305 and a MAE of .0207.

5.6 Magnitude Accuracy Model Comparison

Table 9 shows RMSE and MAE across all four models on the 126-day test set. Random Forest is the all around best model, it does better than the naive benchmark by 33.2% on RMSE. ARIMAX ranks second with a 9.5% improvement compared to the naive benchmark.

Table 9: Out-of-Sample Magnitude Accuracy 126-Day Test Set

Model	RMSE	MAE	vs. Naive RW (RMSE)
Naive Random Walk	.04570	.03175	Benchmark
ARIMA(0,0,0)	.04421	.03138	-3.3%
ARIMAX	.04037	.02746	∑9.5%
Random Forest	.03055	.02071	-33.2%

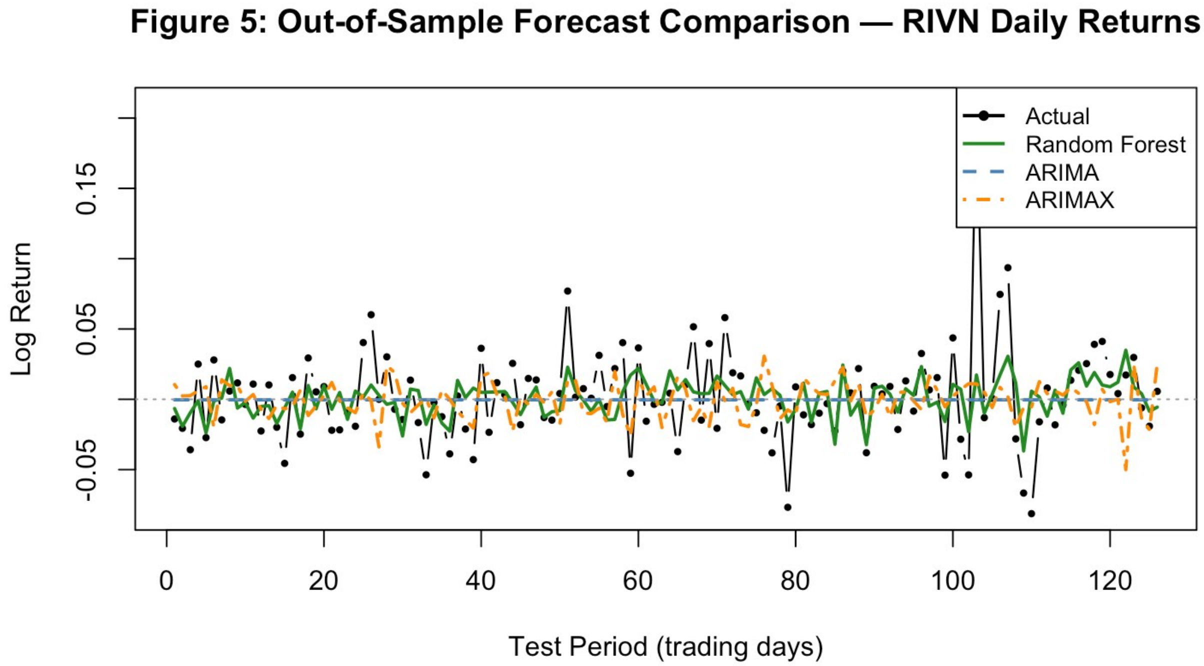

Figure 5: Out-of-Sample Forecast Comparison

5.7 Directional Accuracy

In Table 10 we see the directional accuracy results. The test set contains 50.8% up days and 49.2% down days which is basically a coin flip. The naive benchmark and ARIMA both predict ‘down’ every single test day (because the sample mean is slightly negative), achieving 49.2% hit rate and 0 predicted up days. This means their directional accuracy is simply the percentage of actual down days in the test set, not an actually good prediction that can lead to actual returns in a portfolio.

Table 10: Directional Accuracy 126-Day Test Set (50.8% of days were actually up)

Model	Overall Hit Rate	Up-Day Accuracy	Down-Day Accuracy	Days Predicted Up
Naive RW	49.2%	Never Predicted Up	49.2%	0
ARIMA	49.2%	Never Predicted Up	49.2%	0
ARIMAX	51.6%	52.3%	50.8%	65 of 126
Random Forest	64.3%	63.0%	66.0%	73 of 126

The Random Forest achieves 64.3% overall hit rate, 63% on up-day accuracy, and 66% on down-day accuracy. This is the paper's most interesting and useful result. To put it in context: a 64.3% directional hit rate on daily stock returns, in a test period that is nearly 50/50, is far above what would be expected by chance. For example, lets take a look at a very popular casino game roulette, you can bet on red or black making it relatively 50%, but there are not just 36 numbers you can fall on as there is a green 0 (and depending on the type of roulette there could be two zeros). The reason why the casino ultimately wins is because of that zero as it makes the player's hit accuracy to go from 50% to 47.37%, that 2.63% over thousands if not millions of rolls is what in the long run is able to make this a viable game for the casino. The model correctly identifies the direction of RIVN's daily return nearly two out of every three days.

ARIMAX got a 51.6% hit rate, this is marginally better than the 50.8% always-up baseline and above the 50% random baseline, showing that the contemporaneous TSLA and VIX regressors provide very modest directional impact when combined. The Random Forest's much higher hit rate suggests that the nonlinear combination of these same variables especially along with GARCH sigma, momentum features, and lagged returns we are able to get a directional signal that a simple linear model cannot capture.

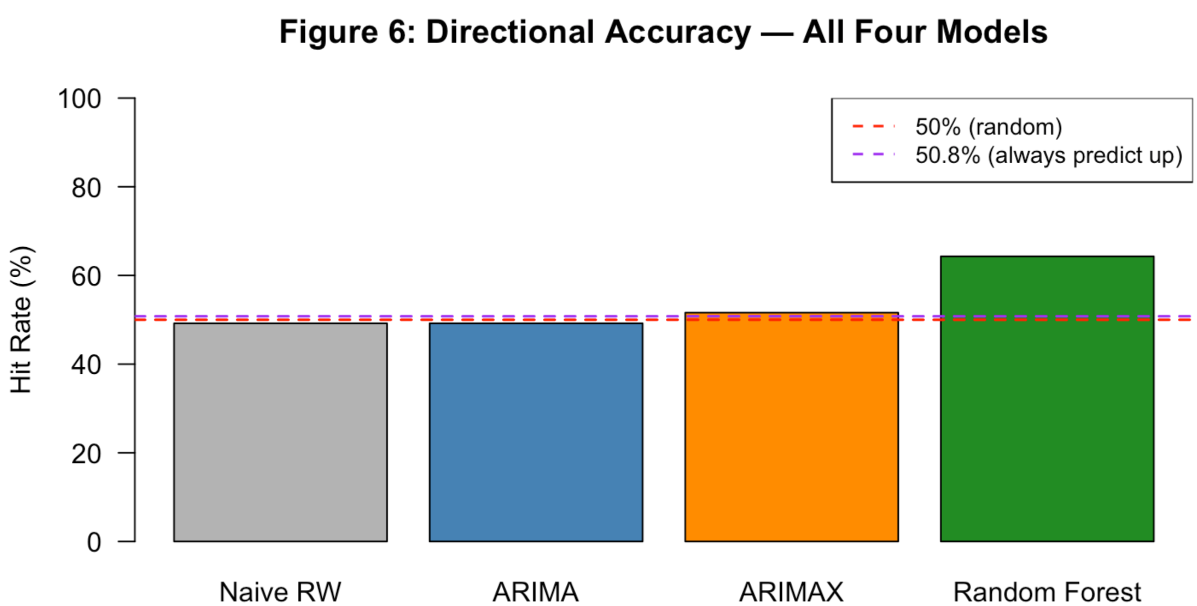

Figure 14: Directional Accuracy

6. Conclusion

This paper used four forecasting methods: ARIMA, ARIMAX, GARCH family, and Random Forest to daily RIVN log returns from January 2024 through June 2026. The results show a clear winner across both magnitude accuracy and directional accuracy. The Random Forest is the best model with an RMSE of .0305 and the directional hit rate is 64.3%. ARIMAX ranks second with an RMSE of .0404 and hit rate of 51.6%. ARIMA and the naive random walk both fail to predict direction at all, this is because both of the models predict negative every day.

The directional accuracy finding is the paper's most interesting and useful results. A 64.3% hit rate on daily RIVN returns in a test period that is basically 50/50 up and down is very good prediction. This comes from the same two variables identified by ARIMAX: VIX changes and TSLA returns. The Random Forest is able to use these variables nonlinearly, combined that with GARCH-estimated conditional volatility and momentum features, we are able to produce a much stronger directional signal than the linear ARIMAX could capture.

The GARCH results show a volatility structure with persistence of .996 and a half-life of 183 trading days. EGARCH achieves the lowest information criterion and gave us a statistically significant asymmetry coefficient ($\gamma_1 = .312$, $p = .005$), this confirms that negative shocks raise RIVN's volatility more than positive shocks of equal magnitude. The GARCH-estimated volatility, when incorporated as a Random Forest feature is the third most important predictor.

We do need to point out one limitation of this paper. The 298-observation RF training set is smaller than what was originally wanted; out-of-sample performance should be validated over longer out-of-sample windows.

7. Bibliography

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 31(3), 307-327.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987-1007.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? Journal of International Economics, 14(1-2), 3-24.

Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. Journal of the American Statistical Association, 55(290), 299-306.

Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. Econometrica, 59(2), 347-370.