

VolForecast: Equity Portfolio Risk Forecasting

EE 344: Final Project Report

Oorjit Chowdhary (oorjitic@uw.edu)

1. Background

Financial volatility forecasting estimates the magnitude of future price fluctuations rather than predicting prices themselves. I developed and evaluated multiple time series models to forecast realized volatility for the “Magnificent 7” (MAG7) technology stocks: Apple (AAPL), Microsoft (MSFT), Google (GOOG), Amazon (AMZN), Tesla (TSLA), Meta (META), and NVIDIA (NVDA).

I collected daily open-high-low-close-volume (OHLCV) data from January 2015 to December 2024 using the Yahoo Finance API. From this raw price data, I computed daily log returns and rolling 21-day realized volatility (annualized) as the primary forecasting target. I also constructed an equal-weighted MAG7 portfolio to analyze how diversification affects volatility predictability.

Financial volatility exhibits volatility clustering (high volatility periods cluster together) and mean reversion, making it challenging to forecast with standard time series methods.

2. Approach

I followed a systematic progression from classical time series methods to domain-specific financial econometric models, applying concepts from class including model selection, cross-validation, and performance evaluation.

Phase 1: Data Preprocessing and Feature Engineering

I began with exploratory data analysis to understand price dynamics, return distributions, and cross-asset correlations. My preprocessing pipeline included: (1) handling missing values and aligning dates across all tickers, (2) computing log returns for statistical stationarity, (3) calculating 21-day rolling realized volatility as the annualized standard deviation of returns, (4) constructing an equal-weighted portfolio by averaging returns across all seven stocks, and (5) generating multiple volatility windows (21-day and 63-day) for comparative analysis. This phase applied curve fitting knowledge, transforming raw price data into stationary return series suitable for time series modeling.

Phase 2: Classical Time Series Modeling

I implemented three fundamental forecasting approaches: Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing methods (Simple,

Holt's, and Holt-Winters), and Facebook Prophet. For ARIMA, I performed grid search over hyperparameters (p , d , q) to identify optimal model orders, testing stationarity with the Augmented Dickey-Fuller test. Prophet decomposes the series into trend and seasonality components.

I split the data into 80% training (2015-2022) and 20% testing (2022-2024) sets. I evaluated model performance using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the R^2 score.

Phase 3: Cross-Validation

To ensure my findings were robust and not artifacts of a single train-test split, I implemented rolling window cross-validation with three temporal splits. Each split used 70% of data for training and 30% for testing, with windows shifted forward in time to evaluate model performance across different market regimes. I computed mean and standard deviation of performance metrics across folds, providing confidence intervals for model comparison. This approach applied the principles discussed in class about avoiding overfitting and ensuring model generalization.

Phase 4: Financial Econometric Models

Recognizing the limitations of generic time series methods, I implemented two specialized volatility models: Exponentially Weighted Moving Average (EWMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH). EWMA estimates volatility using the recursive update $\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1 - \lambda)r_{t-1}^2$, where $\lambda = 0.94$ is the decay factor recommended by JPMorgan's RiskMetrics. GARCH(1,1) models conditional variance as $\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2$, capturing both volatility shocks (α) and persistence (β).

3. Results

Classical Model Performance

My initial experiments with classical time series models revealed fundamental limitations in volatility forecasting. On a single train-test split, ARIMA models achieved R^2 scores of 0.0009 for AAPL, -0.0057 for NVDA, and -0.0040 for the portfolio. Simple Exponential Smoothing (SES) performed similarly with R^2 scores of -0.0315 for AAPL, -1.252 for NVDA, and -0.158 for the portfolio. Prophet significantly underperformed, achieving R^2 scores of -0.102 for AAPL, -1.055 for NVDA, and -1.044 for the portfolio.

Cross-validation analysis confirmed these findings were consistent across different time periods. For AAPL, ARIMA(2,0,5) achieved $MAE = 0.078 \pm 0.0001$, $RMSE = 0.113 \pm 0.00002$, $MAPE = 33.84\%$, and $R^2 = -0.0097$. For NVDA, ARIMA(5,0,5) achieved $MAE = 0.138 \pm 0.0001$, $RMSE = 0.176 \pm 0.0002$, $MAPE$

$= 28.82\%$, and $R^2 = -0.0258$. For the portfolio, ARIMA(3,0,2) achieved $MAE = 0.069 \pm 0.00004$, $RMSE = 0.105 \pm 0.00006$, $MAPE = 25.29\%$, and $R^2 = 0.0166$. Prophet showed $MAE = 0.266 \pm 0.008$, $MAPE = 123.15\% \pm 3.55\%$, and $R^2 = -7.2472 \pm 0.5253$ for the portfolio. The small standard deviations across folds indicated stable but consistently poor performance.

Model	AAPL MAE	AAPL R^2	NVDA MAE	NVDA R^2	Portfolio MAE	Portfolio R^2
ARIMA	0.075	0.001	0.145	-0.006	0.072	-0.004
SES	0.069	-0.032	0.203	-1.252	0.072	-0.158
Prophet	0.073	-0.102	0.222	-1.055	0.126	-1.044

Residual diagnostics revealed systematic patterns (trends and volatility clusters) rather than random noise, indicating model misspecification.

Financial Model Performance

In stark contrast, domain-specific financial models demonstrated transformational predictive power. EWMA achieved R^2 scores of approximately 0.85 for AAPL, 0.84 for NVDA, and 0.86 for the portfolio—representing orders of magnitude improvement over classical methods. GARCH(1,1) models achieved R^2 scores of 0.79-0.80, slightly below EWMA but still vastly superior to generic approaches.

For the portfolio, EWMA achieved $MAE = 0.028$ and $RMSE = 0.039$, while GARCH(1,1) achieved $MAE = 0.030$ and $RMSE = 0.045$. These represented 59% and 58% error reductions compared to the best classical model (ARIMA). The GARCH parameter estimates ($\alpha \approx 0.1$, $\beta \approx 0.88$, $\alpha + \beta \approx 0.98$) confirmed high volatility persistence.

Model	AAPL MAE	AAPL R^2	NVDA MAE	NVDA R^2	Portfolio MAE	Portfolio R^2
EWMA	0.030	0.85	0.033	0.84	0.028	0.86
GARCH	0.032	0.79	0.036	0.80	0.030	0.80

Portfolio vs. Individual Stocks

Across all models, the equal-weighted MAG7 portfolio demonstrated consistently higher predictability than individual stocks. This diversification benefit arose from the cancellation of idiosyncratic volatility shocks across assets. Classical models showed relative performance improvements on portfolio data (ARIMA $R^2 = 0.0166$ in cross-validation vs. -0.0097 and -0.0258 for individual stocks), while financial models maintained their superiority with slightly better R^2 scores for the portfolio.

My correlation analysis revealed moderate positive correlations (0.4-0.7) among MAG7 stocks, confirming diversification benefits. This finding suggested that portfolio-level volatility forecasts are more reliable than aggregating individual stock forecasts.

4. Conclusion

This project demonstrated that financial volatility forecasting requires domain-specific econometric models. Classical time series methods (ARIMA, Exponential Smoothing, Prophet) achieved near-zero or negative R^2 scores across all assets, failing to capture volatility clustering and mean reversion. These models treated volatility as a generic time series with independent observations.

In contrast, financial econometric models (EWMA and GARCH) achieved R^2 scores of 0.79-0.86, representing transformational predictive power. EWMA efficiently captured volatility persistence through a single decay parameter ($\lambda = 0.94$), while GARCH modeled both volatility shocks and persistence simultaneously.

Key Findings

1. **Domain-specific models were essential:** Generic time series methods could not capture financial volatility dynamics, regardless of hyperparameter tuning or cross-validation strategies.
2. **Volatility clustering drove predictability:** Both EWMA and GARCH exploited the observation that high volatility periods cluster together, providing short-term forecasting power.
3. **Portfolio diversification improved forecasts:** Equal-weighted MAG7 portfolio volatility was more predictable than individual stocks across all models.
4. **Cross-validation validated findings:** Rolling window analysis confirmed that model performance was consistent across different market regimes, not artifacts of a single train-test split.

Challenges and Limitations

The primary challenge was the inherently noisy nature of financial volatility. Even my best models (EWMA/GARCH with $R^2 \approx 0.85$) explained only 85% of variance, leaving substantial unpredictable components.

Classical models struggled fundamentally with volatility's non-constant variance structure. ARIMA assumes constant variance errors, while volatility is inherently time-varying. No amount of hyperparameter tuning could overcome these architectural limitations.

Computational constraints limited my GARCH implementation to rolling one-step-ahead forecasts, which was computationally expensive (30+ minutes per asset for 500-day test periods). Multi-step forecasts would require more sophisticated approaches.

My equal-weighted portfolio construction, while simple and interpretable, ignored optimization for minimum variance or maximum Sharpe ratio. Risk-parity or mean-variance optimized portfolios might exhibit different volatility dynamics.

Finally, my 21-day volatility window represented a specific time horizon (approximately one trading month). Shorter windows might be too noisy, while longer windows might sacrifice responsiveness.

References

1. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.
2. RiskMetrics Group. (1996). *RiskMetrics Technical Document* (4th ed.). J.P. Morgan/Reuters.
3. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37-45.
4. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice* (3rd ed.). OTexts.