Volatility Forecasting and Option Pricing on Gold ETF

TKSS ASHOK

15 Aug 2025

1 Introduction

Gold has historically been considered a safe-haven asset and plays a central role in financial markets. The aim of this project is to:

1. Forecast the daily volatility of SPDR Gold Shares (GLD) returns using statistical and machine-learning approaches.
2. Apply the forecasted volatility within the Black–Scholes (BS) framework to price European call options for selected expiries on a single valuation date.

2 Data and Preprocessing

2.1 Data Source

We use historical daily adjusted closing prices of SPDR Gold ETF (GLD) from Yahoo Finance.

2.2 Log Returns

Given price series P_t, log returns are:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right).$$

These returns are the basis for volatility estimation and modeling.

3 Volatility Forecasting

Let σ_{t+1} denote the (annualized) next-day volatility. We consider six approaches:

3.1 Naïve

$$\widehat{\sigma}^2_{t+1} = r_t^2.$$

3.2 Rolling Average

For window m,

$$\widehat{\sigma}_{t+1}^2 = \frac{1}{m} \sum_{i=1}^{m} r_{t-i}^2.$$

3.3 GARCH(1,1)

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \qquad \alpha_0 > 0,\ \alpha_1, \beta_1 \geq 0,\ \alpha_1 + \beta_1 < 1.$$

3.4 Random Forest (RF)

$$\widehat{\sigma}_{t+1} = f_{\mathrm{RF}}\big(x_t;\, \theta\big),$$

with x_t lagged features and θ tree hyperparameters.

3.5 XGBoost

Volatility as an additive ensemble of trees:

$$\widehat{y}_t = \sum_{k=1}^{K} f_k(x_t), \quad f_k \in \mathcal{F}.$$

Objective:

$$\mathcal{L} = \sum_t \ell(y_t, \widehat{y}_t) + \sum_{k=1}^{K} \Omega(f_k), \quad \Omega(f) = \gamma T + \frac{\lambda}{2} \sum_{j=1}^{T} w_j^2.$$

3.6 LSTM

With lookback window L,

$$h_t = \mathrm{LSTM}\big(r_{t-L+1}, \ldots, r_t\big), \quad \widehat{\sigma}_{t+1} = W h_t + b.$$

3.7 Evaluation of Forecast Accuracy

We compare model forecasts to realized volatility using error metrics:

$$\mathrm{RMSE} = \sqrt{\tfrac{1}{n} \sum_{t=1}^{n} (\sigma_t - \widehat{\sigma}_t)^2}, \quad \mathrm{MAE} = \tfrac{1}{n} \sum_{t=1}^{n} |\sigma_t - \widehat{\sigma}_t|,$$

$$R^2 = 1 - \frac{\sum_t (\sigma_t - \widehat{\sigma}_t)^2}{\sum_t (\sigma_t - \bar{\sigma})^2}.$$

Model	RMSE	MAE	R^2
Naïve	0.009039	0.005705	0.9733
Rolling Avg	0.010591	0.007576	0.9634
GARCH(1,1)	0.023179	0.017012	0.8247
Random Forest	0.009707	0.006589	0.9693
XGBoost	0.002478	0.001341	0.9981
LSTM	0.006973	0.005229	0.9811

Table 1: Comparison of volatility forecast models by error metrics.

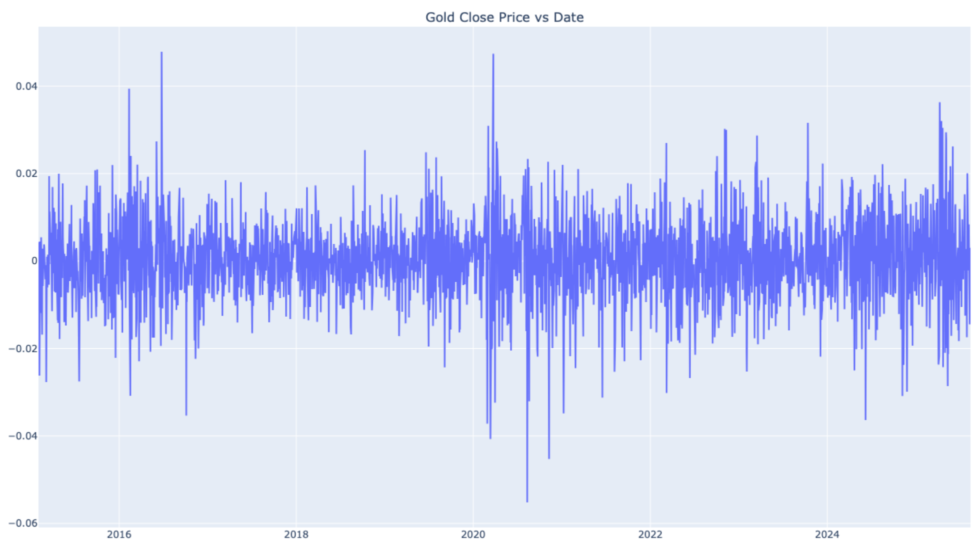

Figure 1: GLD log returns over time.

4 Option Pricing (Single-Day Application)

4.1 Black–Scholes Framework

European call price:

$$C = S_0\, N(d_1) - Ke^{-rT} N(d_2), \quad d_1 = \frac{\ln(S_0/K) + \left(r + \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T}.$$

4.2 Comparison Across Models

5 Conclusion

We compared statistical (Naïve, Rolling, GARCH) and ML (RF, XGBoost, LSTM) volatility models for GLD. Error metrics (RMSE, MAE, R^2) guided model selection, and forecasted

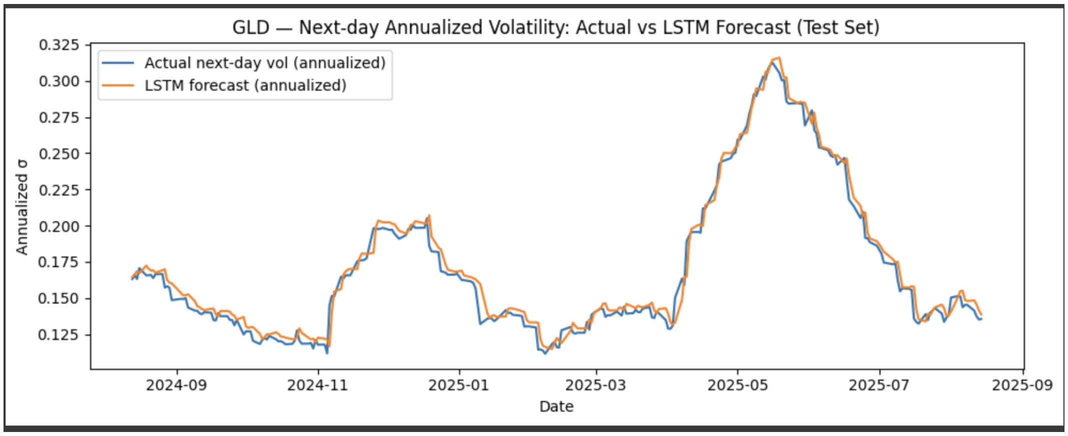

Figure 2: Actual vs. LSTM-forecasted volatility on the test set.

Model	σ_T	Strike Price	Sep 2025	Feb 2026
Naïve	0.152039	300	10.78	20.831
Rolling Avg	0.151225	300	10.757	20.769
GARCH(1,1)	0.142754	300	10.514	20.115
Random Forest	0.155018	300	10.865	21.028
XGBoost	0.141204	300	10.464	19.953
LSTM	0.138648	300	10.39	19.752

Table 2: BS call prices using forecasted volatilities on valuation date(15th Aug 2025, Spot Price = 307.5.

volatilities were applied in Black–Scholes to obtain option values for Sep–Feb maturities.