Predicting Volatility with Regression

Cindy Xu & Rohan Shukla

1 What is Volatility?

Volatility is...

- A measure of how much an asset's price moves over time
- Higher volatility means larger and more frequent price swings
- Lower volatility means smaller and more stable price movements
- Captures risk and uncertainty in financial markets

There are 2 distinct measures of volatility

VIX / Implied Volatility

- Represents the market’s expectation of future volatility.
- Based on prices of S&P 500 options.
- Higher VIX = market expects bigger future moves.
- Multiple VIX indices exist, based on different option expiration dates:
  - VIX: 30-day (standard)
  - VIX3M: 3-month
  - VIX6M: 6-month
  - VIX1Y: 1-year
- Called "implied" because it's inferred from option prices using models like Black-Scholes.

Realized Volatility

- Measures the actual historical volatility observed in asset returns.
- Calculated using past daily returns over a fixed time window (e.g., 1 month, 1 year).
- No market expectations — purely reflects how much asset prices have already moved.

2 What Did We Do?

Project Overview

Goal: Predict future volatility of the stock market using statistical models.

Research Questions:

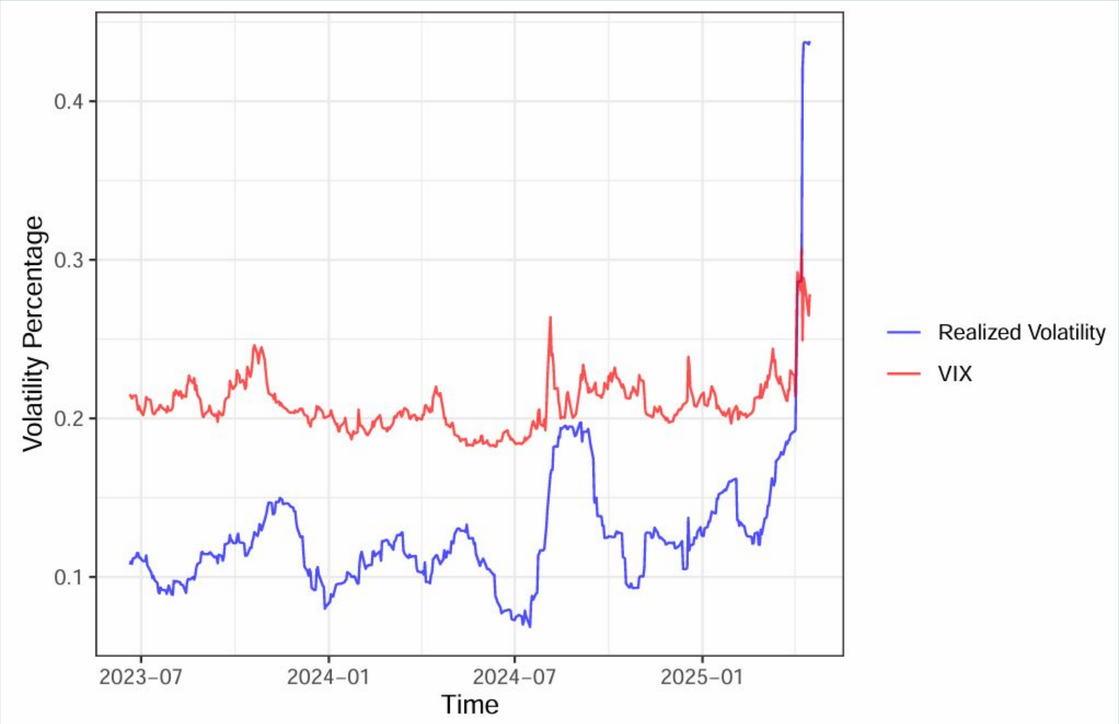
Can we predict future VIX based on only today's realized volatility?

Can we predict future realized volatility based on today's VIX and trailing volatilities?

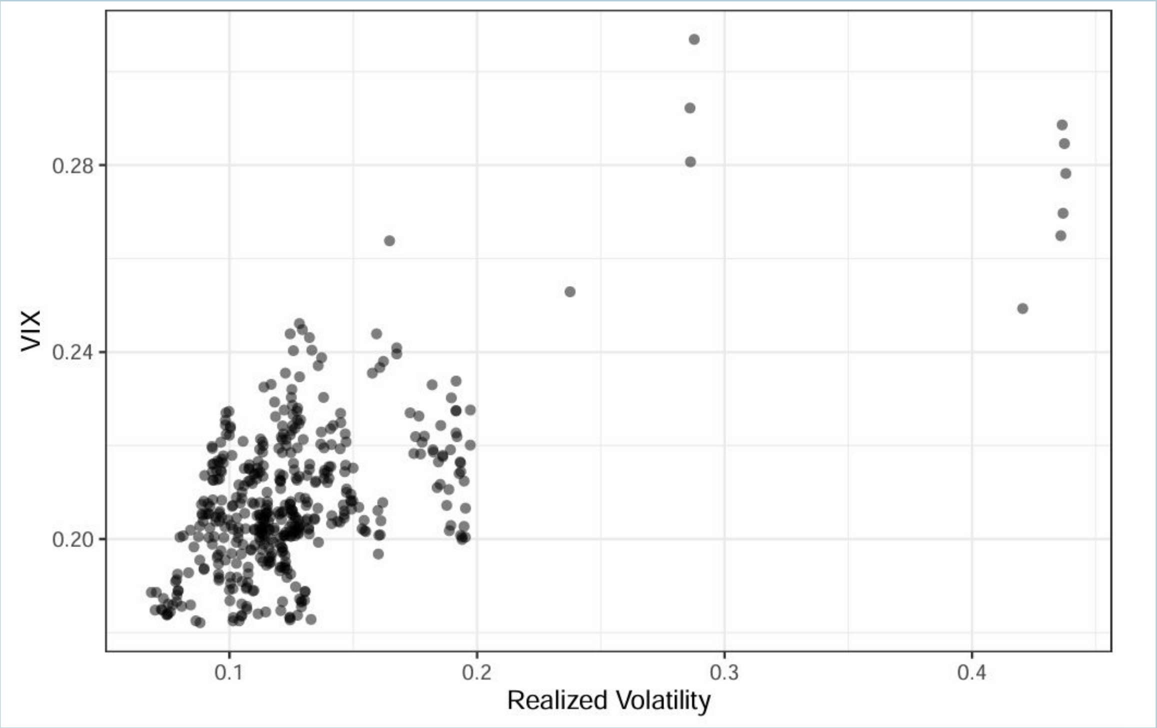
3 VIX vs Realized Volatility

Volatility Over Time

Spans from 6/20/2023 - 4/17/2025



VIX vs Realized Volatility



Regress VIX on Realized Volatility

- Linear Regression

$$\widehat{\mathbb{E}}[Y \mid X] = \widehat{\beta}_0 + \widehat{\beta}_1 X$$

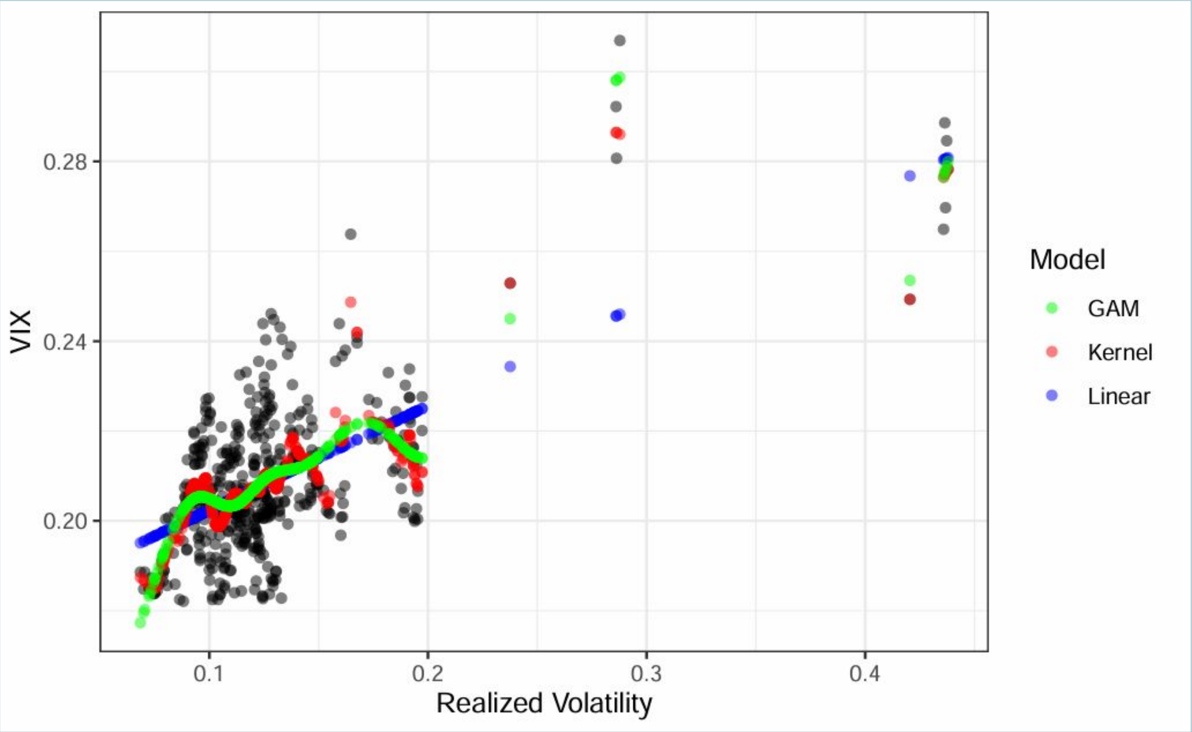
- Generalized Additive Model (GAM)

$$\widehat{\mathbb{E}}[Y \mid X] = \widehat{f}(X)$$

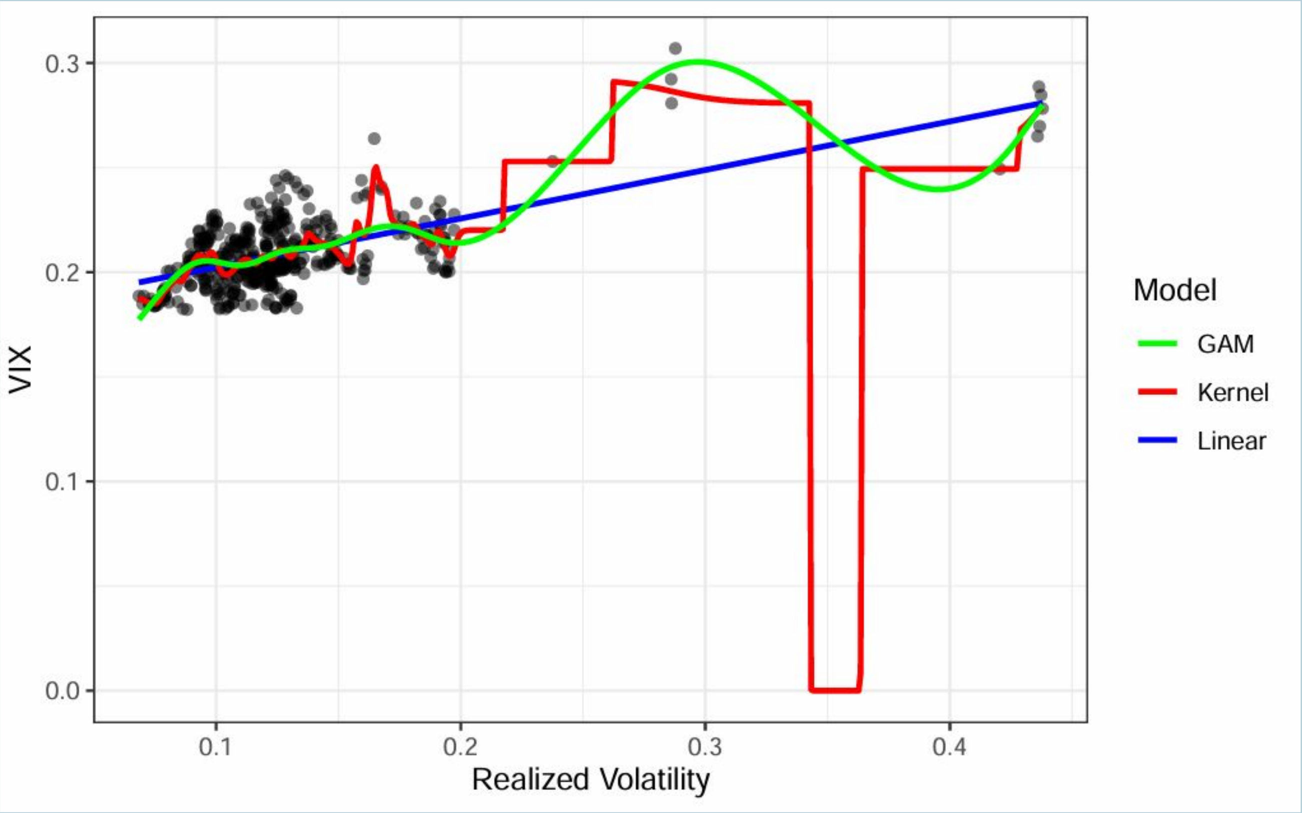
- Kernel Smoothing

$$\begin{aligned}\widehat{\mathbb{E}}[Y \mid X = x] &= \sum_{i=1}^{n} y_i w(x_i, x) \\ &= \sum_{i=1}^{n} y_i \frac{K(\frac{x-x_i}{h})}{\sum_{j=1}^{n} K(\frac{x-x_j}{h})}\end{aligned}$$

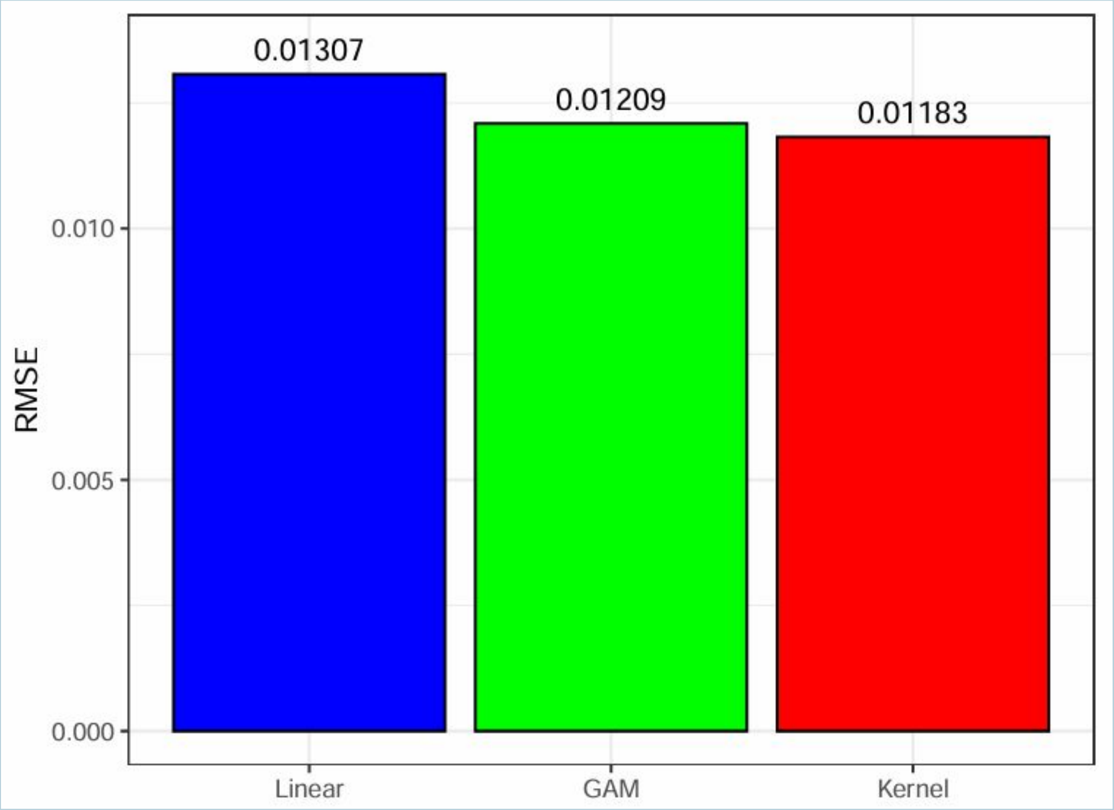
Fitted VIX From Each Model



Fitted Curves From Each Model



RMSE Comparison



4 VIX 1Y, EWMA vs 1 yr Realized Volatility

SARIMA and Regression Models

- Linear Regression

$$\widehat{\mathbb{E}}[Y \mid X] = \widehat{\beta}_0 + \widehat{\beta}_1 X$$

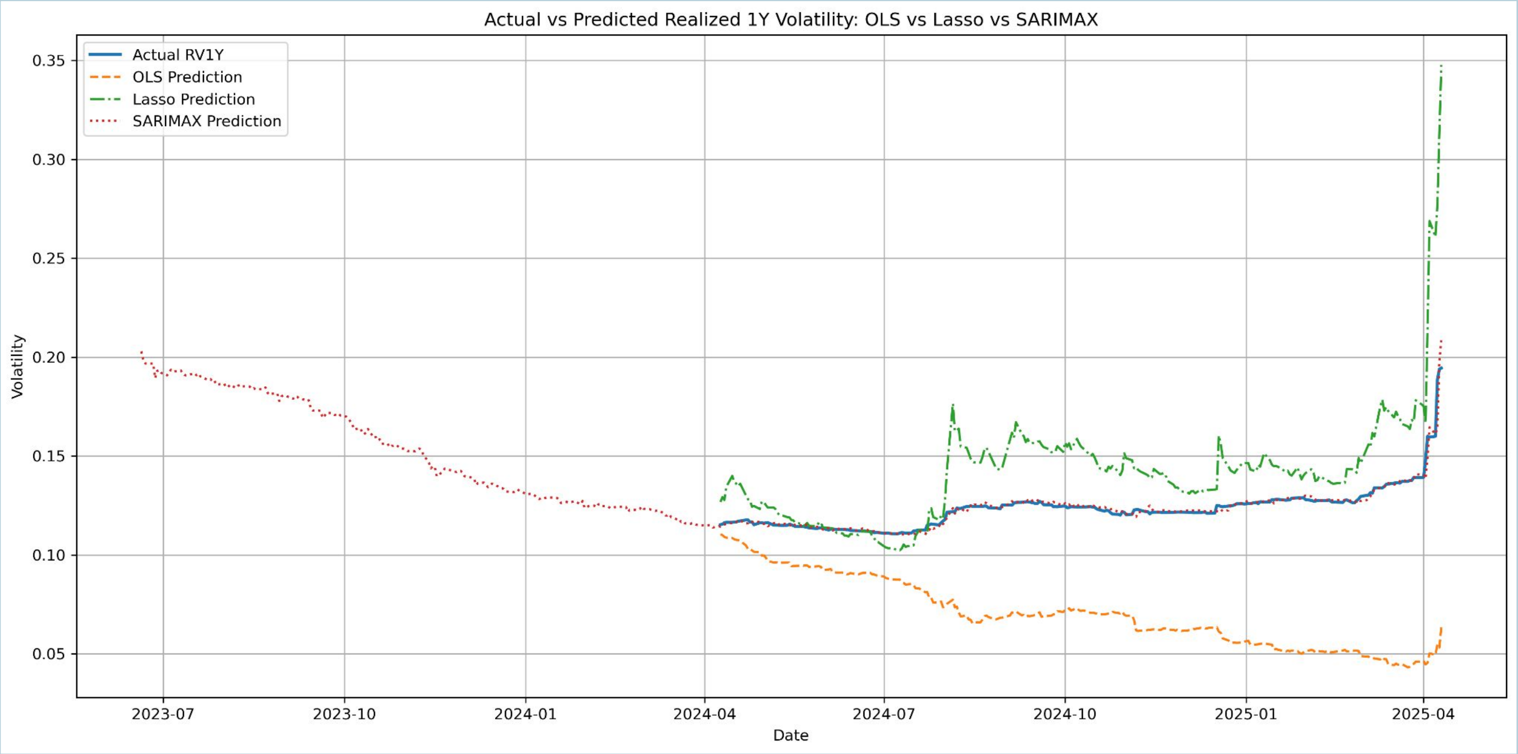
- Lasso Regression

$$\widehat{\beta} = \arg\min_{\beta} \left(\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right)$$

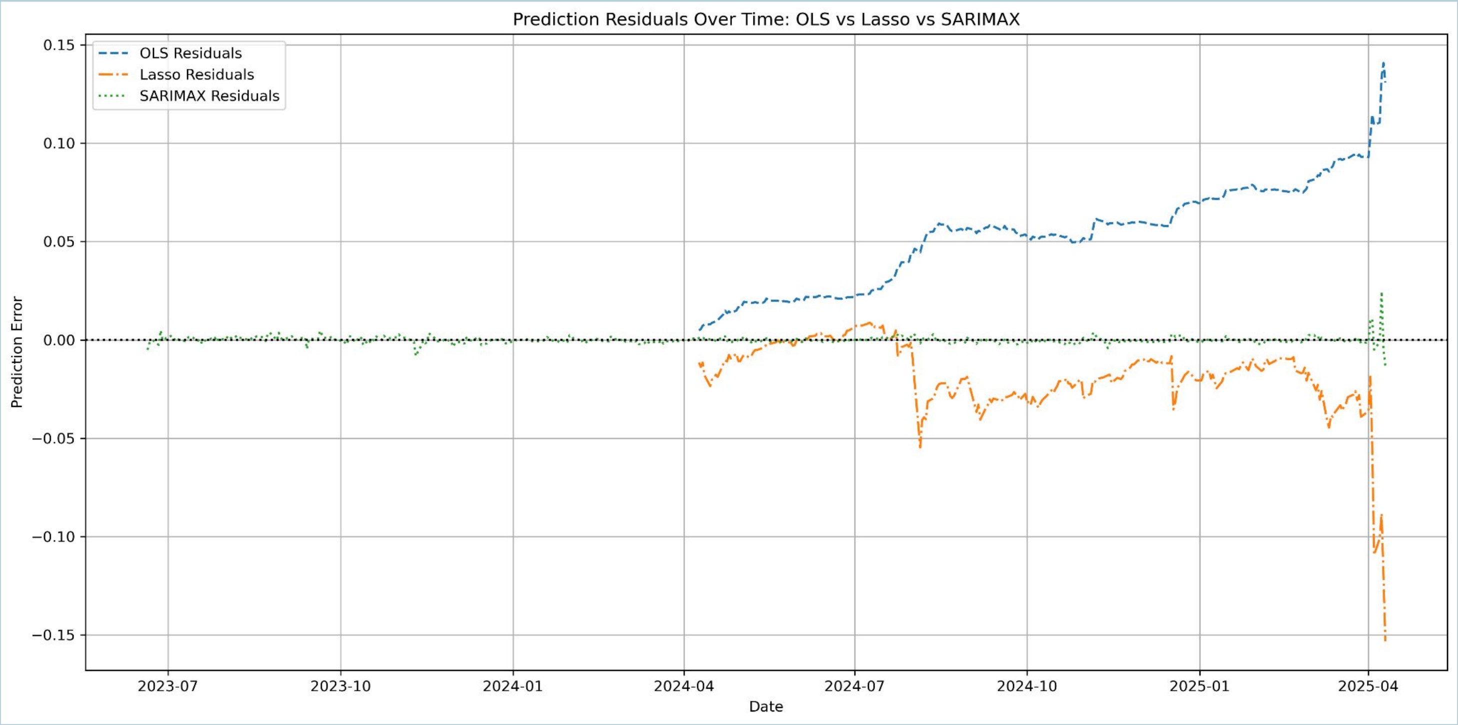
- SARIMAX

$$\Phi(B^s)\phi(B)(X_t - \mu) = \Theta(B^s)\theta(B)\epsilon_t$$

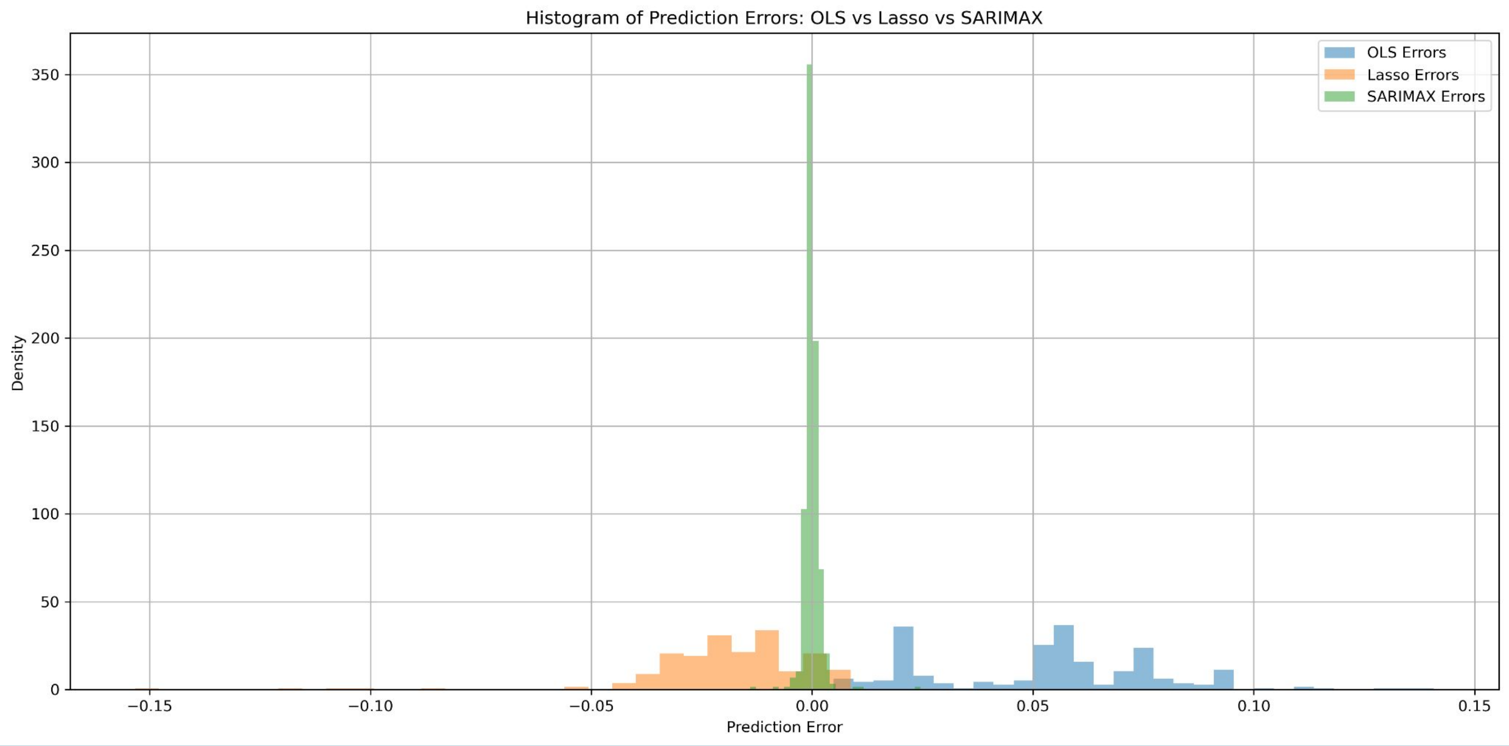
Actual vs OLS, Lasso, SARIMAX Predicted Volatility



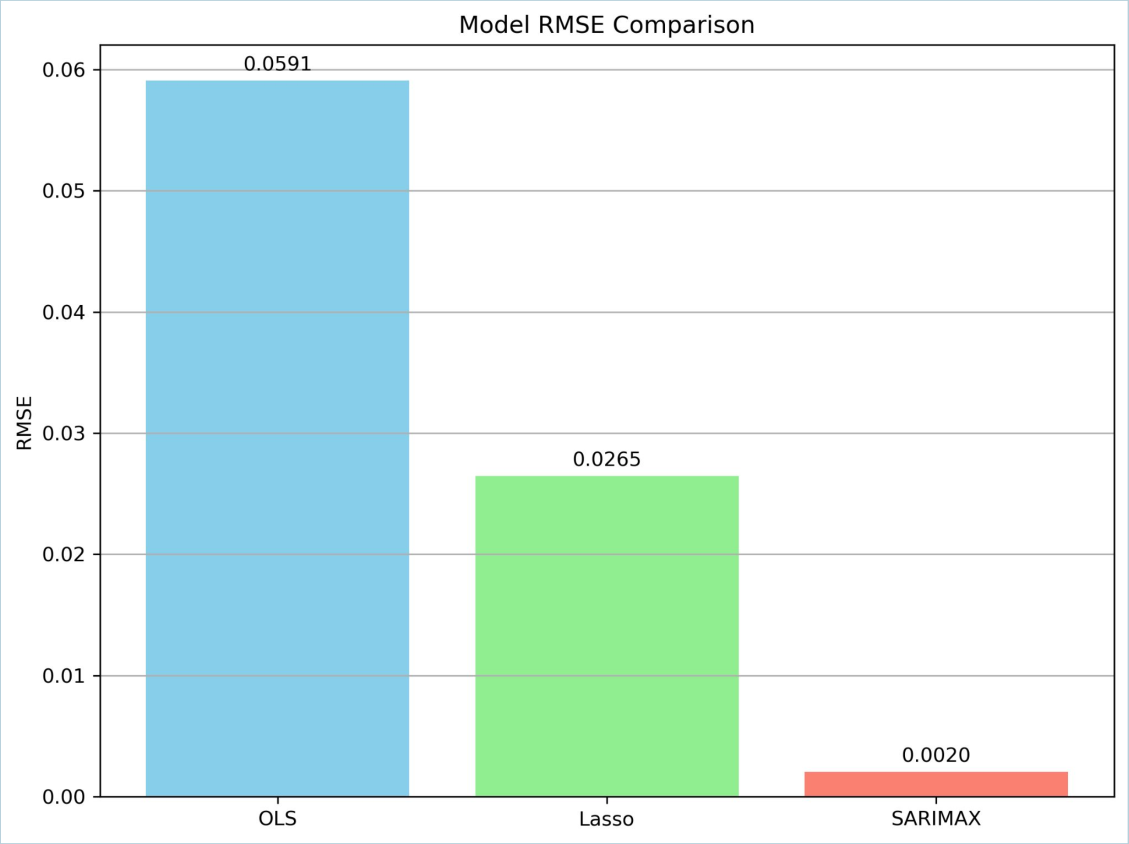
Errors: Residuals



Errors: Histograms



RMSE Comparison



Key Takeaways

- SARIMAX best captures the dynamic behavior of realized 1-year volatility.
- Lasso improves over standard OLS by reducing multicollinearity and shrinking unnecessary predictors.
- Simple OLS regression fails to capture volatility patterns due to ignoring time dependency.
- We used Time Series Cross-Validation (TimeSeriesSplit) to train the models, ensuring the temporal order was respected and avoiding lookahead bias.

Thanks!

Any questions?