ARCH-BASED VOLATILITY MODELING

Aidan Olson, Arzaan Singh, Leo Small

AGENDA

Motivation and Objectives

Data & Exploratory Analysis

ARCH Model

Model Selection and ARCH(M) Results

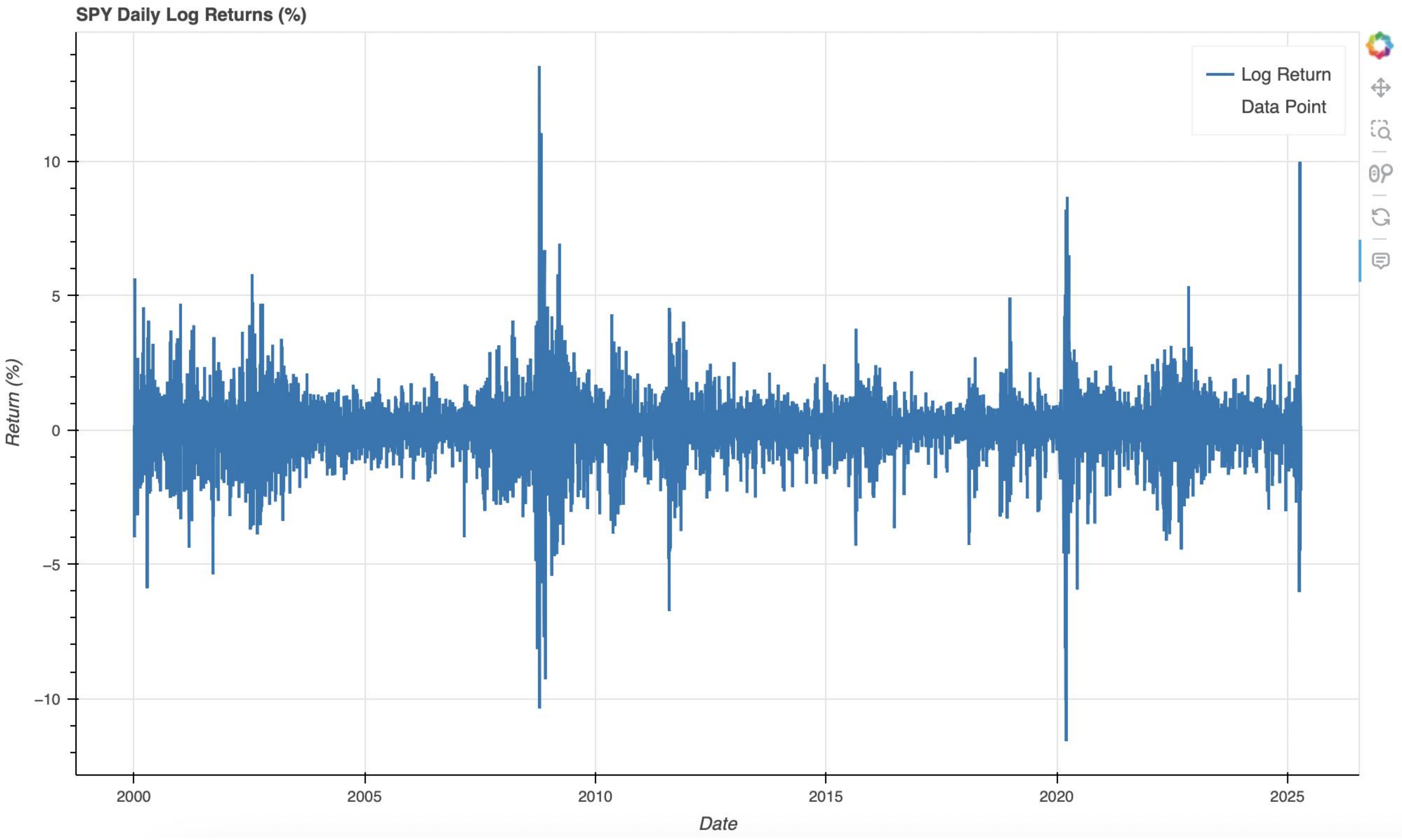
Volatility Forecasting (Deterministic & Monte-Carlo)

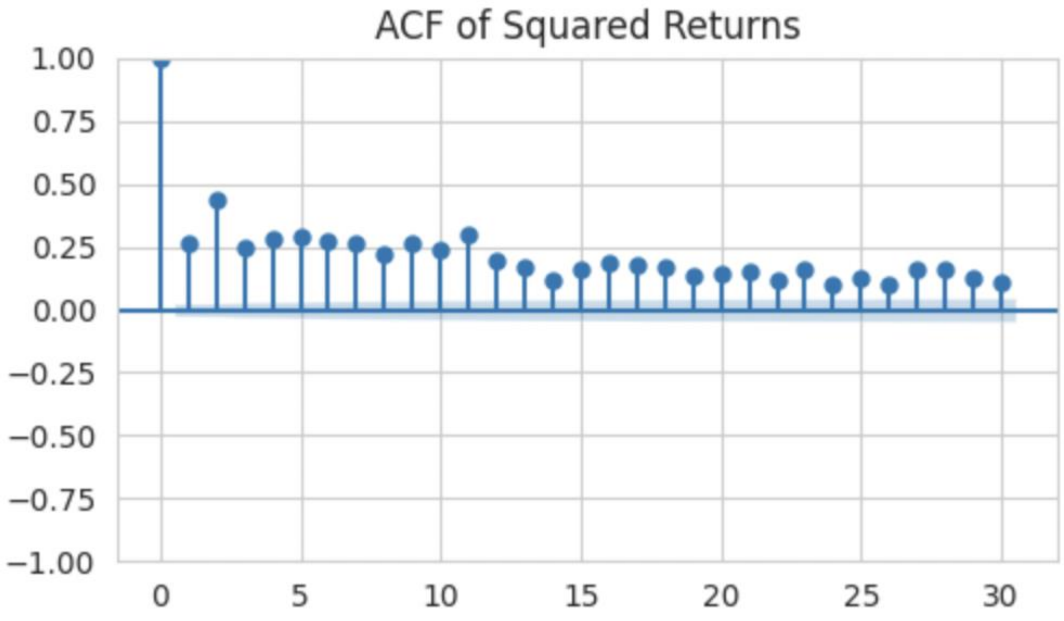
Back testing, Validation, and Library Implementation

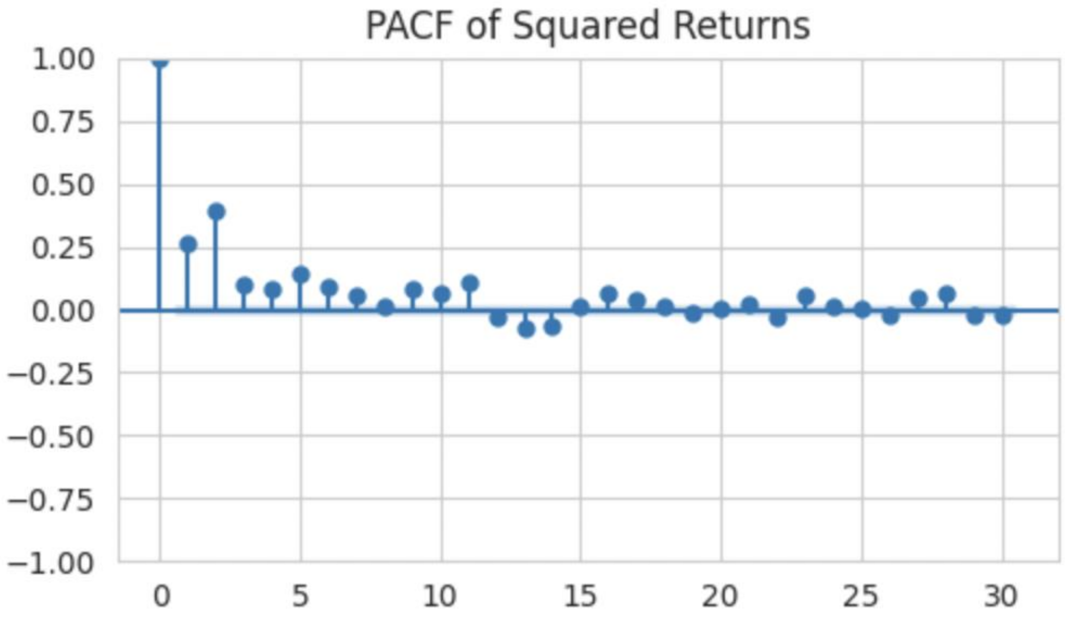
Options Pricing

Conclusion & Next Steps

WHY DO WE NEED THE ARCH MODEL?







DATA & EXPLORATORY ANALYSIS

- Loaded in data from yfinance API
- Created summary statistics and plots to understand underlying trends/distribution
- Analyzed Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

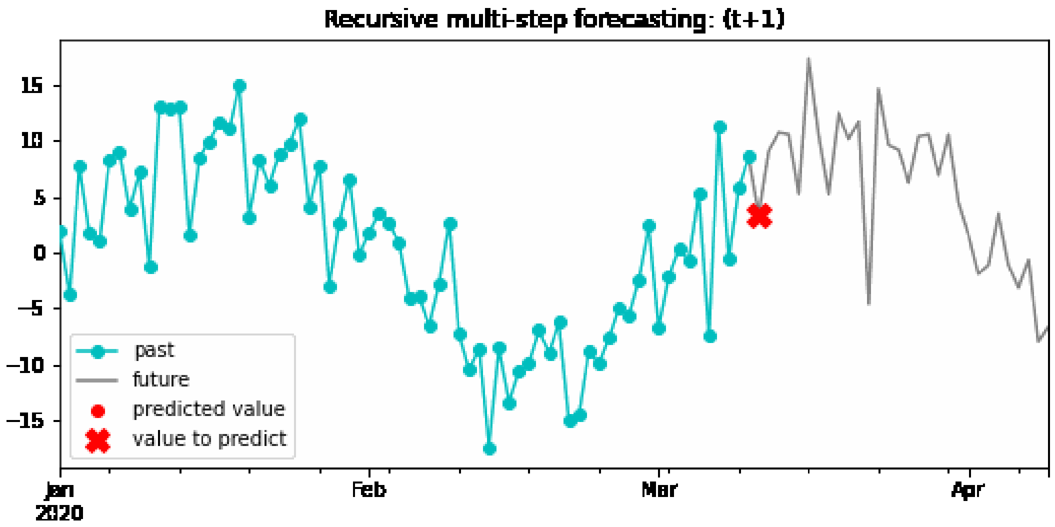
THE MODEL

AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY MODEL

- A model created in 1982 ARCH forecasts how volatility will change over time. ARCH uses the previous a weighted average of previous days' volatility to predict future volatility
- Volatility is a key input in options pricing. However, unlike what we have seen in class, volatility is far from constant. ARCH helps model that variance in volatility by considering past volatility.
- In essence, current volatility is **conditional** on past **volatility**

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$$

CONDITIONAL AUTOREGRESSION AND VOLATILITY CLUSTERING



- The first main concept of the ARCH Variance Model
- Present data is regressed on previous data which is then used to predict future data

HETEROSKEDASTICITY

- The second main concept of the ARCH Variance Model
- Refers to the fact that variance of the return is not constant, changing in response to shocks.
- Homoscedastic Return Modeling:

$$r_t \sim N(\mu, \sigma^2)$$

- Heteroskedastic Return Modeling:

$$r_t = \mu + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_t^2)$$

MODEL FORMULATION

MATH

- If we model our returns using the following formula to capture time-varying volatility, we need the volatility for time t:

$$r_t = \mu + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_t^2)$$

- This is where ARCH(q) helps us find the volatility using past volatility

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{\{t-1\}}^2 + \alpha_2 \epsilon_{\{t-2\}}^2 + \cdots + \alpha_q \epsilon_{\{t-q\}}^2$$
$$\alpha_0 > 0, \qquad \alpha_i \geq 0 \;\; for\ i = 1, \ldots q$$

- Where t is the day, α is the feature weight (what we are finding!), and q is the number of features (or days we are considering in our formula for current volatility

MORE MATH

$$\sigma_t^2 = \omega + \sum_{i=1}^{m} \alpha_i \, r_{t-i}^2 .$$

$$-\ell(\omega, \alpha_1, ..., \alpha_m) = \frac{1}{2} \sum_{t=1}^{T} [\ln(2\pi) + \ln(\sigma_t^2) + \frac{r_t^2}{\sigma_t^2}].$$

So, how do we solve for these features (α)?

- **Gaussian log-likelihood parameter estimation:**
- We maximize the likelihood of a parameters for our given dataset using gradient descent
- We also add a positivity constraint for the parameters to avoid cases like negative variance
- Use

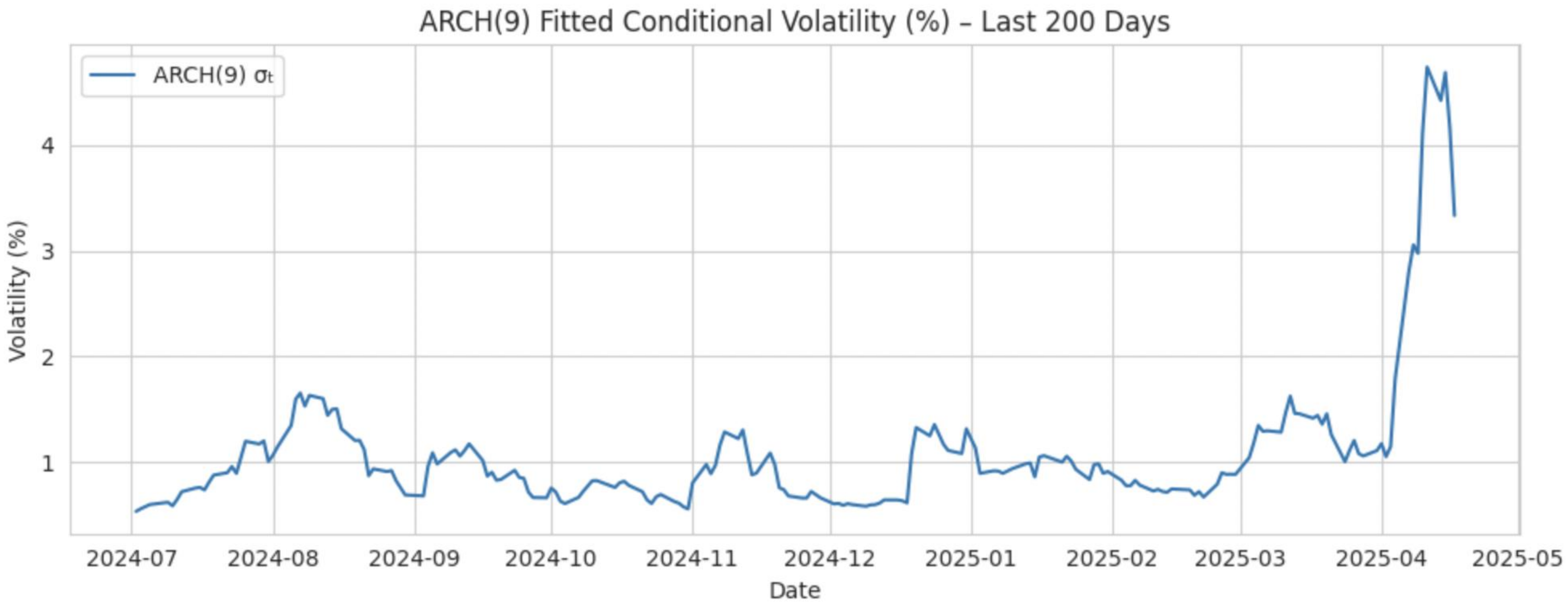
CHOOSING OUR ARCH(Q) MODEL

ARCH(m) Order Comparison (AIC & BIC)

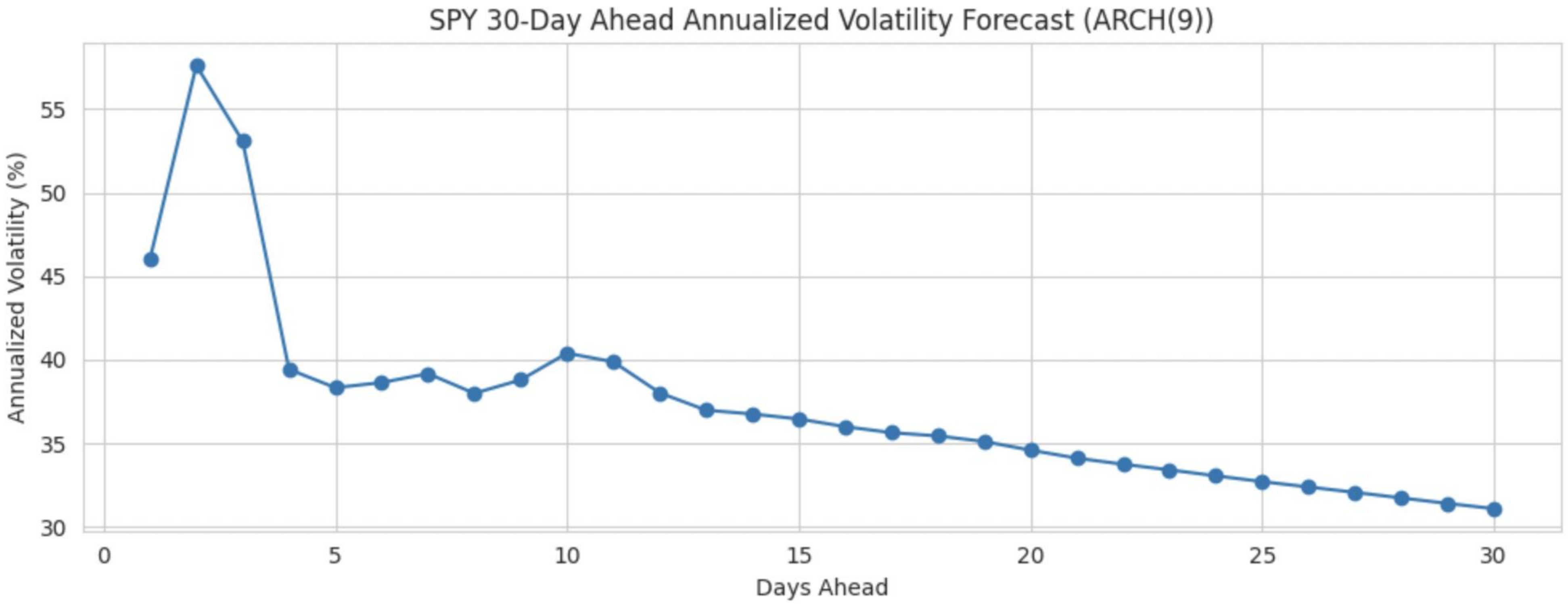
Order m	AIC	BIC	Converged
1	19,621.53	19,635.05	Yes
2	18,750.98	18,771.25	Yes
3	18,304.14	18,331.17	Yes
4	18,001.29	18,035.08	Yes
5	17,882.74	17,923.29	Yes
6	17,808.78	17,856.09	Yes
7	17,770.52	17,824.59	Yes
8	17,703.78	17,764.60	Yes
9	**17,683.99**	**17,751.57**	Yes
10	Diverged	Diverged	No

RESULTS

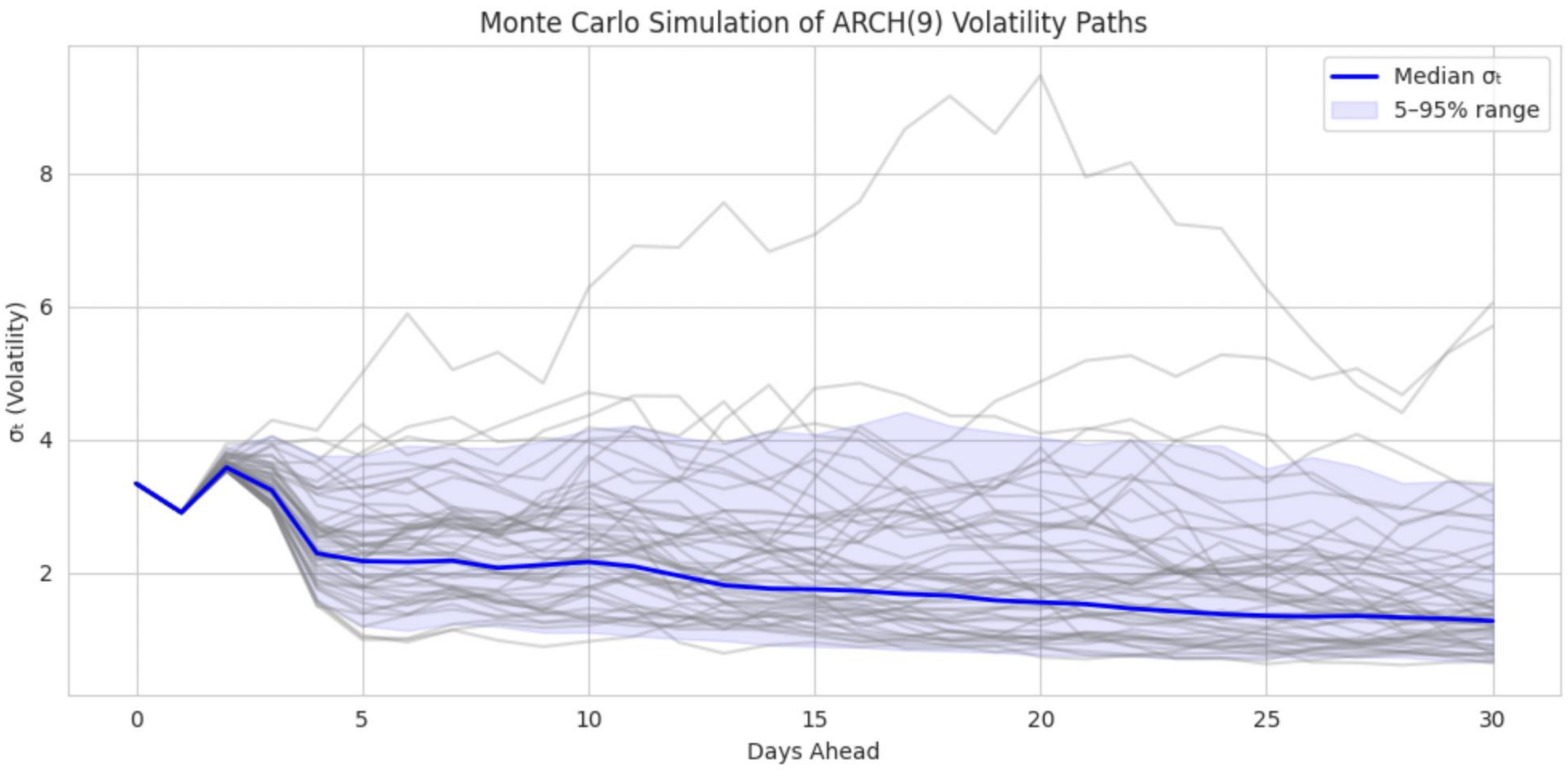
ARCH(9) FITTED VOLATILITY



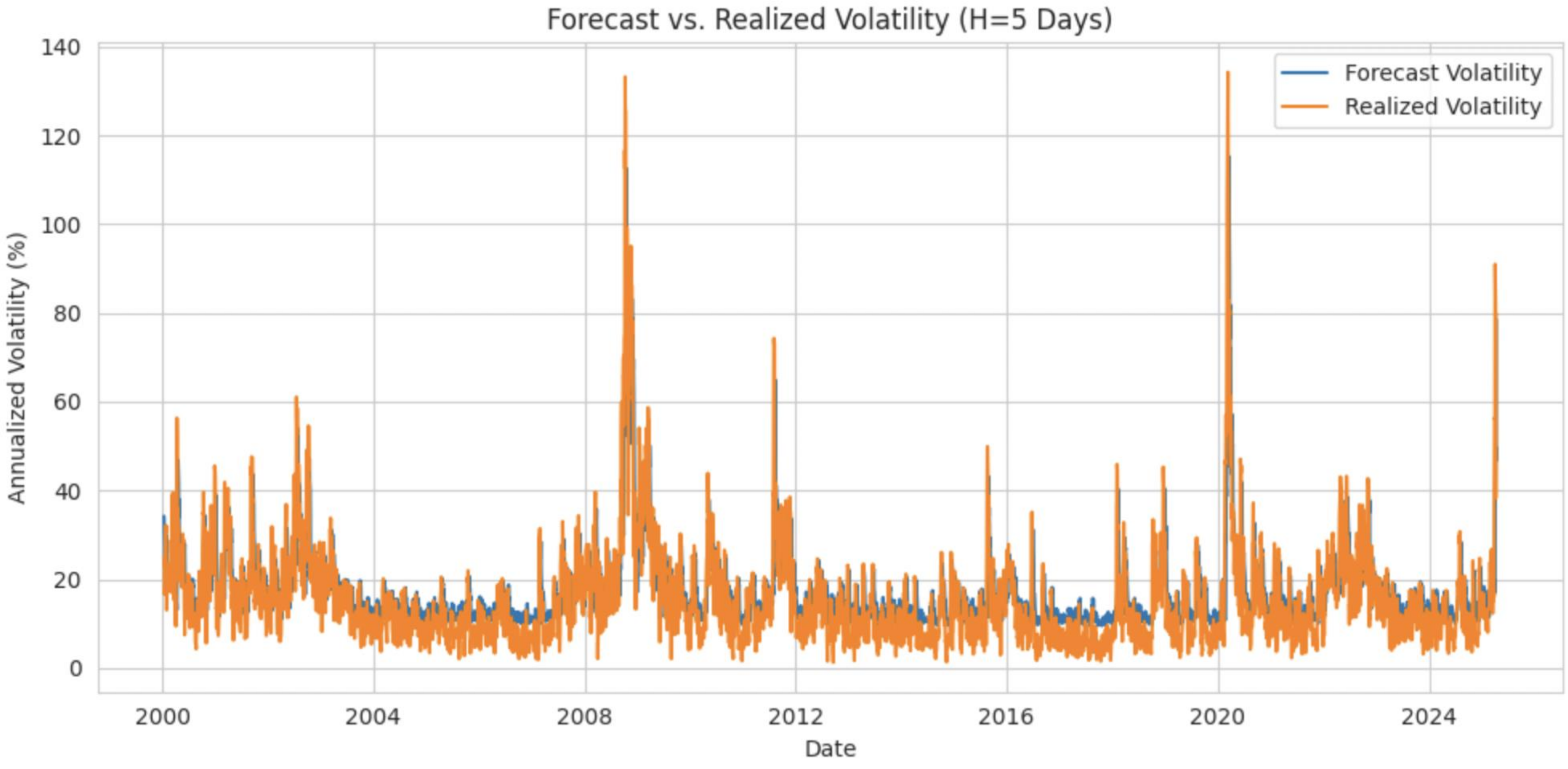
ARCH(9) FUTURE VOLATILITY PREDICTION



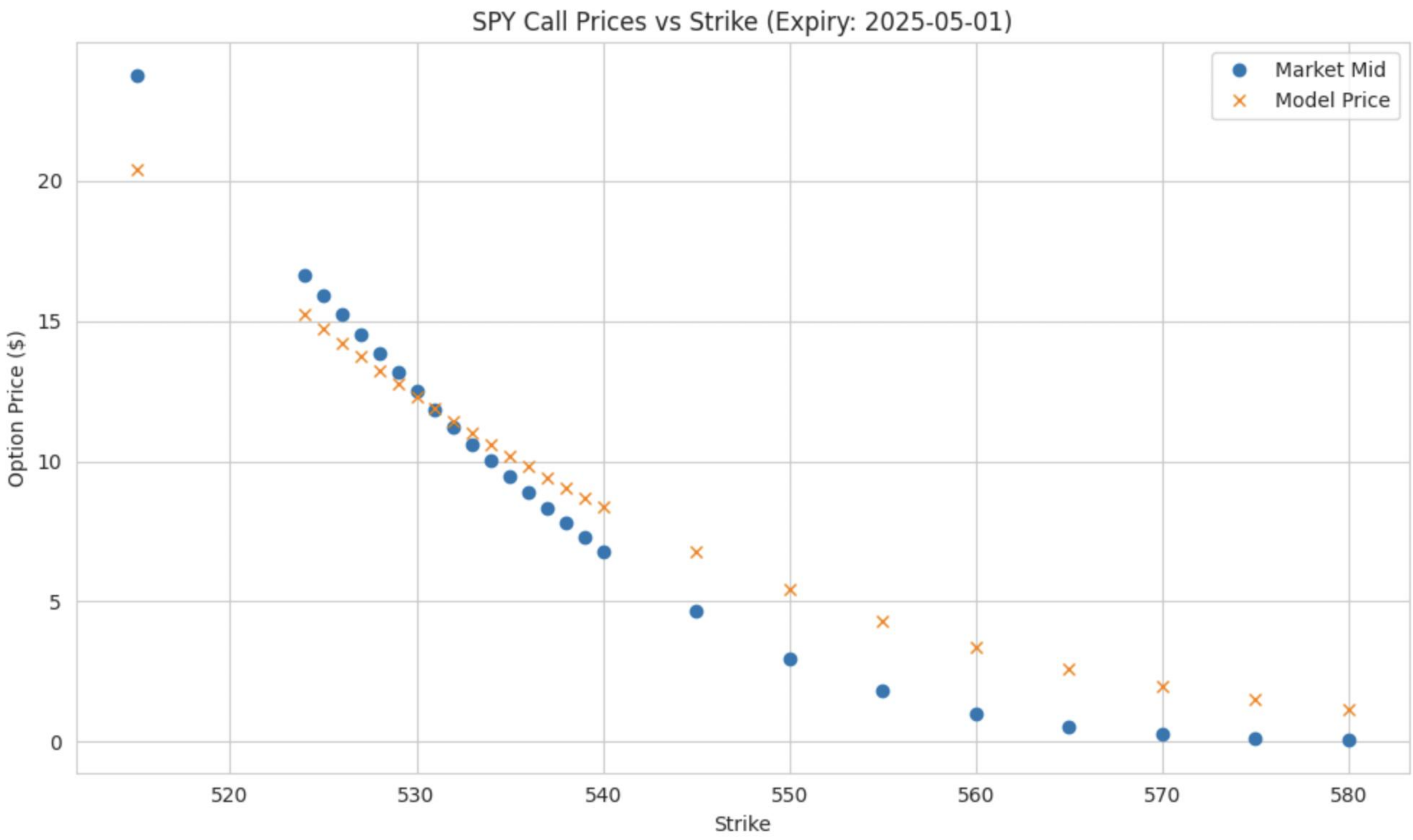
ARCH(9) VOLATILITY WITH RANDOMNESS



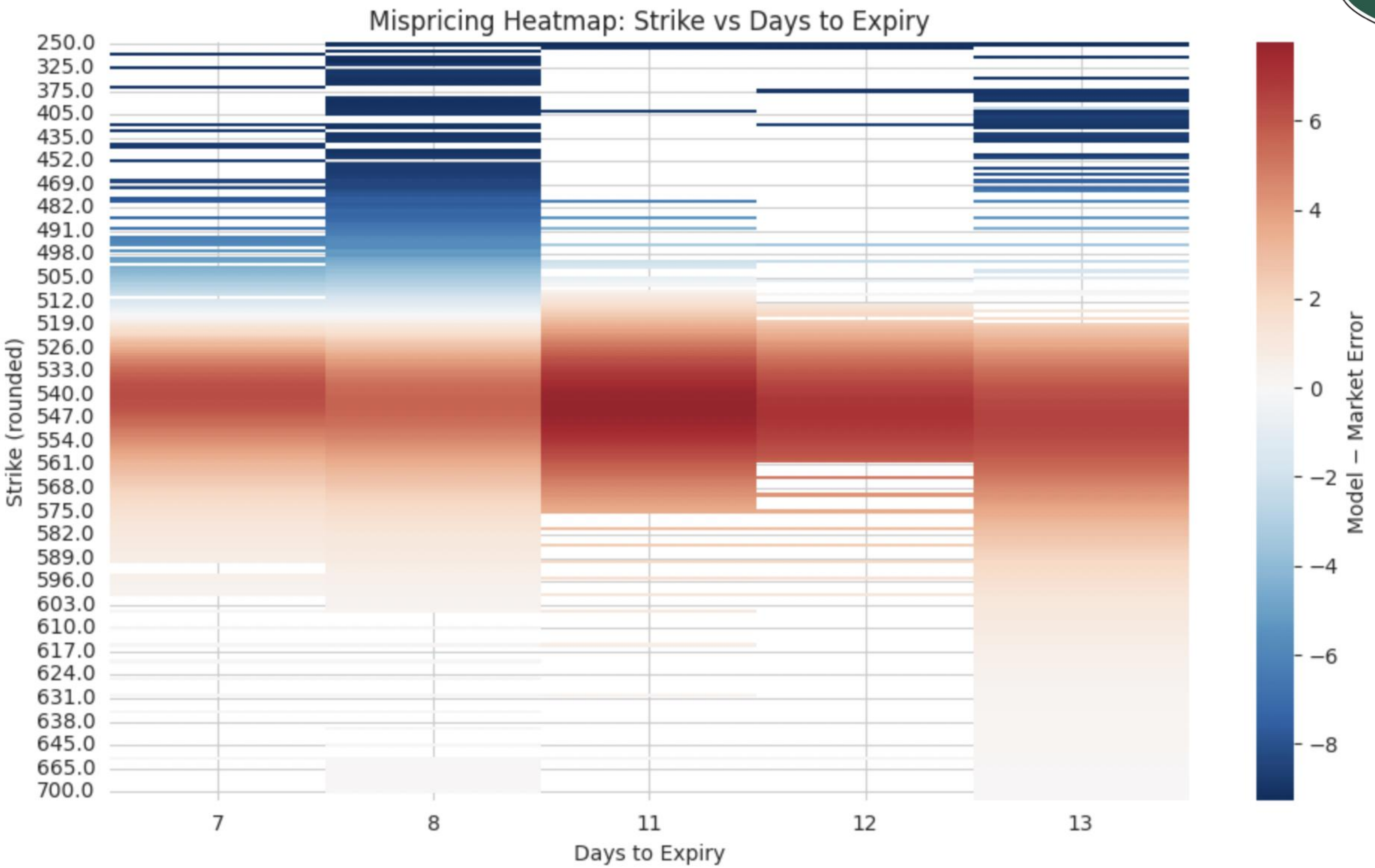
FORECASTING VOLATILITY



FORECASTED VS ACTUAL OPTIONS PRICE



MISPRICING



Mispricing Heatmap: Strike vs Days to Expiry
Strike (rounded)
250.0
325.0
375.0
405.0
435.0
452.0
469.0
482.0
491.0
498.0
505.0
512.0
519.0
526.0
533.0
540.0
547.0
554.0
561.0
568.0
575.0
582.0
589.0
596.0
603.0
610.0
617.0
624.0
631.0
638.0
645.0
665.0
700.0
7
8
11
12
13
Days to Expiry
Model − Market Error
6
4
2
0
−2
−4
−6
−8

ARCH MODEL'S LIMITATIONS

- ARCH treats positive and negative return shocks identically—cannot capture the "leverage effect" (the tendency for volatility to rise more after negative shocks).
- While modeling, ARCH looks at past errors and not past variances. This oversight causes it to sometimes miss long-term volatility trends. Standard ARCH assumes Gaussian innovations; it struggles with fat tails unless paired with heavy-tailed error distributions (e.g., Student's t).
- Since ARCH also uses past data, huge sudden swings in the market due to external events can lead to its predictions to be off.
- Overfitting vs. Training Error

FURTHER RESEARCH

GARCH/EGARCH Model to capture persistence and leverage effects

Add in more representative (fat-tailed distributions) to represent black-swan events in model

Fractal Distributions to simulate volatility clustering and not just a lagged indicator

THANK YOU