Hypothesis: GB/RF on MACD+RSI improves predictions of ARIMA on grouped price & volatility

ARIMA

- Big ‘q’ = Very ‘unstable’
- Big ‘p’ = Length; High (linear) autocorrelation; Long ‘simple memory’
- Big ‘q’ = Length (shock); High (linear) autocorrelation of shocks; Long ‘simple shock memory’
- Low MSE = Strength; Strong and consistent
    - Hence Accurate ARIMA (high p/q, low MSE) = Long and strong linear memory/shock memory

MACD/RSI

- Accurate MACD/RSI = Longer, linear/non-linear memory

GB/RF

- GB better→ SImilar structural relationships between features→variable
- RF better→ Different structural relationships or none

Prices

Price, Stacked

- d=1→ Fairly stable
- p=0→ Price and shocks are memoryless→ RW
- Extremely high MSE→ **Dramatic RW**

Decreased MSE for rolling

- **Short-term** forecasts follow a much **less dramatic RW** than **longer-term**→ ARIMA **predicts next-day** prices well but **poor in year-long** predictions
- Mathematical reason: Errors ‘stacks’ up over time; Error from each day ‘passed’ on

High p,q and low MSE for rolling groups:

- **Price trends** tend to have **strong momentum** and **shock-diluting** in the **short-term**

Grouping changed MSE by:

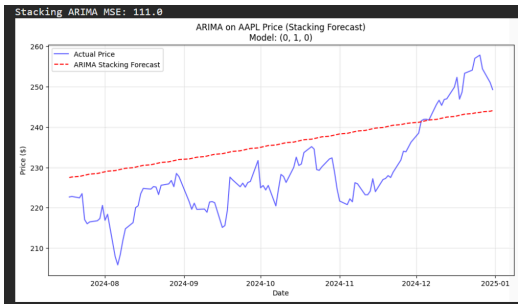
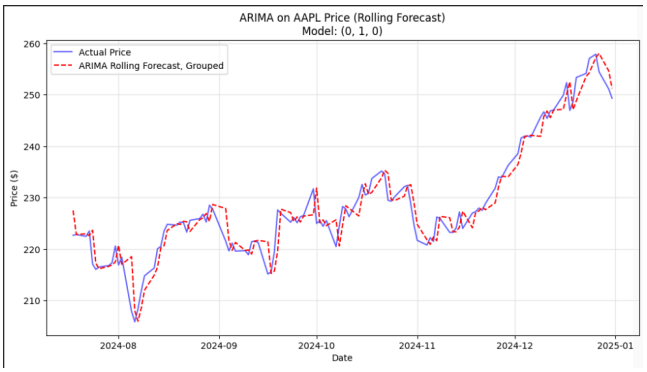
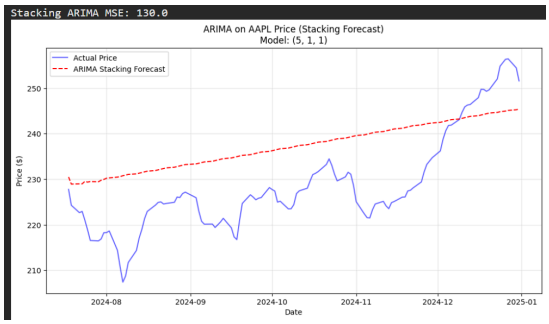
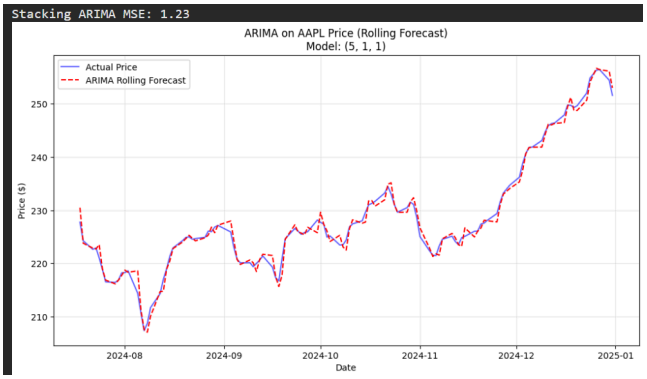
Increased (rolling)

- Mathematical reason: Grouping→ Smoother data→ Lower noise→ Lower MSE

Decreased (stacking)

- Same smoothing effect occurs, BUT
- Higher p→ More days that the error from each prediction gets ‘passed’ on→ Higher MSE
- Evidently, latter outweighs the former

<u>Side-lesson:</u> The more AC the data, the better it is for short-term forecasting BUT worse for longer-term

	Stacking	Rolling
Prices	Ungrouped: 111.0 	7.97 
Price 3	130.0 	1.23 

Volatility

VOL 10

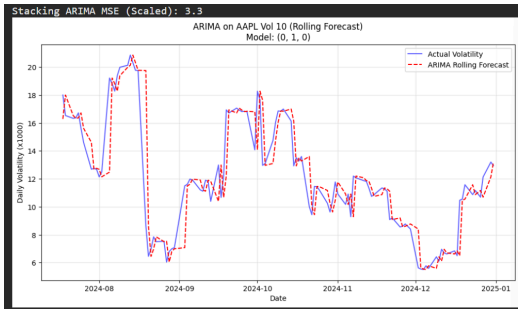
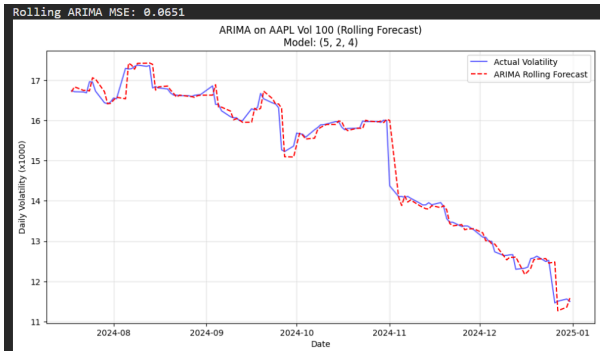
SIMILAR: Identical p, d, q values = Random walk
DIFFERENCE: Low MSE→ **Quieter RW**

Vol 100

d=2→ Big accelerations→ Highly ‘unstable’
LOW MSE +

- Big p→ Long, strong memory
- Big q→ Shocks last long but are predictable

Overall: In terms of short-term predictions, volatility itself is only somewhat predictable, but trends in volatility are both very predictable and very momentum-reliant

	Rolling
Vol 10	3.9 
Vol 100	0.0651 

With GB/RF

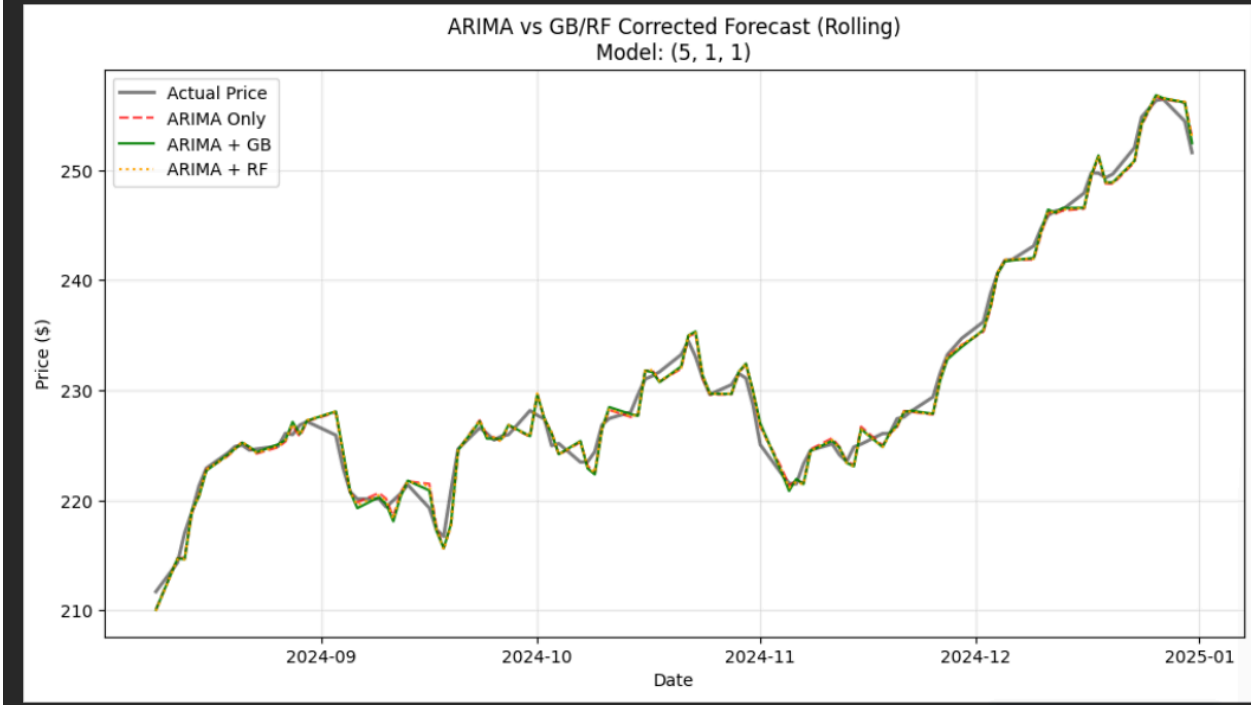
Selection criteria: Those with somewhat autocorrelative patterns (p,q>0) but not necessarily the greatest accuracy with ARIMA (MSE high)

Price 3, Rolling

Reason: Auto-correlation (p, q>0) but imperfect predictions
Result:

- <1% MSE change→ No significant improvement→ **Hypothesis false**
- Accurate ARIMA, inaccurate GB/RF on residuals→ **No medium-term, non-linear trends**; Only the one above

```
ARIMA Only:           1.2347
ARIMA + GB Corrected: 1.2367
ARIMA + RF Corrected: 1.2452
```



Vol 100, Rolling

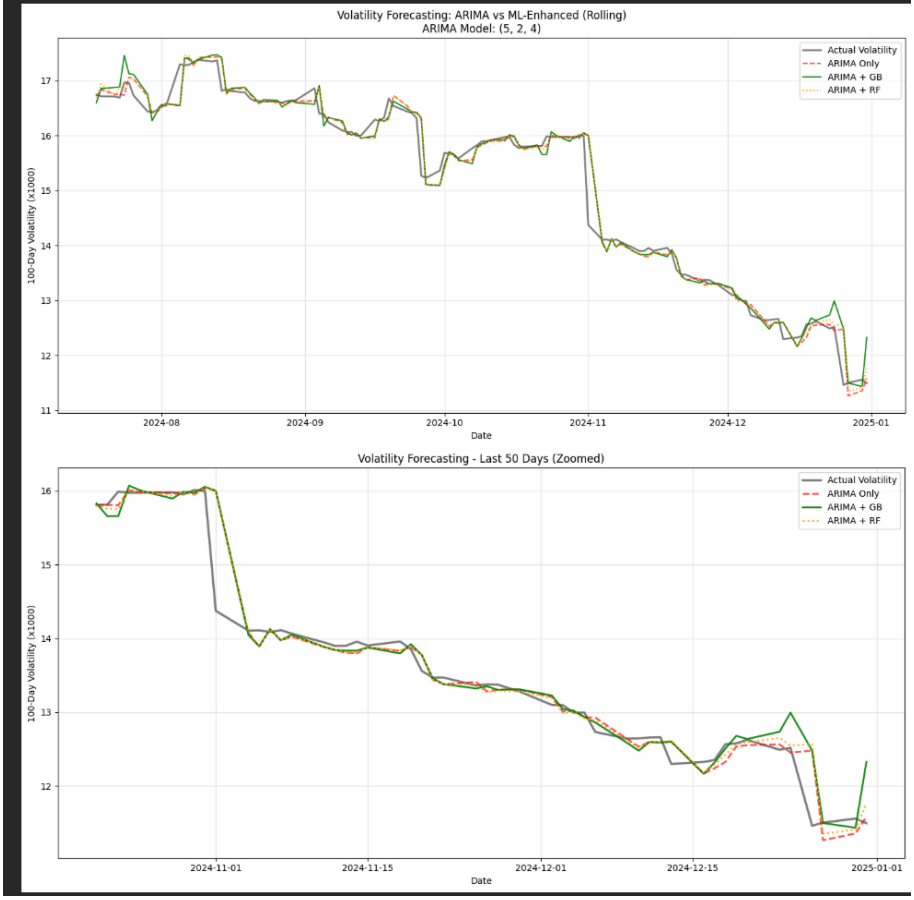
Reason: Same
Result:

- >1% decrease→ Significantly worse performance→ **Hypothesis false**
- Much greater % decrease→ Much worse overfitting→ **Volatility tends to have isolated clustering** of **different trends**

- Volatility has both:
    - The behaviour above
    - Occasional longer-term, non-linear behaviour- ones that remain for a while but change over time

```
--- Results (MSE x 1,000,000) ---
ARIMA Only:              0.065
ARIMA + GB Corrected: 0.077
ARIMA + RF Corrected: 0.067
```

```
Improvement vs ARIMA:
GB: -18.45%
RF: -3.12%
```



Price, Stacked

Reason: Want to try forecast, not just theoretical understanding
Also higher MSE→ More informative changes