Intro + Investment Thesis

Forecasting volatility is often the biggest discrepancy between different market agents in options pricing. Differences in forecasted volatility can result in dramatically different price estimates, which is the focus of this pitch. Using two different time series models, we will construct price intervals for options contracts, and based off of discrepancies with the forecasted price interval and current option prices we will trade to find out if:

a) The two time series modeling methods create returns that are statistically significant from null hypothesis of zero
b) If one of the methods produces substantially different results than the other, and the reason for the difference

Procedure

1) The first step I took was getting daily SPX Index prices for the past 5 years on Bloomberg, and training both the ARIMA - GARCH and HARX models on the daily data from 08/29/16 to 04/29/21.
2) Using the remaining test data I generated one step ahead daily forecasts for the SPX (upper and lower bounds), updating the model after each day with the day's SPX price. I then tested the efficacy of the model by running a success rate calculation for spot one step ahead, and found that my results were pretty poor (forecast interval was too small)
3) From there I decided I should be using a modifier for my bounds (some multiple of the error term and standard deviation), and so I had to optimize a modifier term based on the number of contracts it yielded (that fall outside my bounds).
4) After creating decent forecasts for both models, I had to download options chain data. I manually downloaded a month's worth of SPX options chain data (April) from Bloomberg, cleaned it up, and tried to import it, but found that I was having memory issues since the df was quite large. Instead, I decided to import a week's worth of data from the chain, and trade off of that
5) I filtered and cleaned the options chain dataframe, adding various Black Scholes inputs to it. After I added the model forecast bounds to the options chain df, I went ahead and used the Black Scholes formula to calculate the IV of the option, as well as convert the spot forecasts I generated into an IV range. I ran a filter on the options chain to find contracts that I wanted to trade on (expiring within 30 days so I could get the full PNL), and chose 5 positions from that that I wanted to trade from. I created separate dataframes for all 5 of the positions, delta hedging every day at market open by buying and selling shares of the theoretical "underlying."
6) I calculated the PNLS of each of the positions, adding them up at the end to get a list of cumulative PNLS for each of the portfolios. From there I generated the PNL graph for both time series models.

Black Scholes + Delta Hedge Refresher

The BS model is a mathematical model for pricing an options contract. The model predicts that the price of heavily traded assets follows a geometric Brownian motion with constant drift and volatility. Inputs for the BS model are the current underlying / spot price (S), strike price (K), risk free rate (r), time to maturity (t), and volatility (sigma). Furthermore, the implied volatility (IV) of a stock can be calculated by solving for the sigma value (given you have the option price). IV plays a major role in options pricing, and a higher IV implies a higher option price and vice versa.

$$C = S_t N(d_1) - K e^{-rt} N(d_2)$$

where:

$$d_1 = \frac{ln\frac{S_t}{K} + (r + \frac{\sigma_v^2}{2})\, t}{\sigma_s \sqrt{t}}$$

and

$$d_2 = d_1 - \sigma_s \sqrt{t}$$

where:

$C =$ Call option price

$S =$ Current stock (or other underlying) price

$K =$ Strike price

$r =$ Risk-free interest rate

$t =$ Time to maturity

$N =$ A normal distribution

For the basis of the pitch, I'm trading on IV. If the IV of an options contract is outside of the range that I forecast, then I either buy or sell the contract (depending on if it's above or below the range). Since my thesis is based on the IV of a contract, I need to delta hedge to limit my directional exposure. Delta is one of the options greeks, and is the ratio that compares the change in the price of an underlying asset with the change in the price of a derivative or option. Long market positions have positive deltas, and short market positions have negative deltas. For the purpose of the pitch, I sold and bought the "theoretical" underlying stock to hedge my directional exposure at market open each day. Buying underlying results in positive delta, and selling underlying results in negative delta. The sum of each position's delta should be as close to zero as possible in a delta neutral portfolio. To give an example of delta hedging in action, if I buy an apple call with an IV of 0.5, I would sell 50 shares of AAPL to have a net zero delta.

Model 1: ARIMA - GARCH Forecasting

The first time series model I utilized for this project is the ARIMA-GARCH forecasting model, where the ARIMA component forecasts a conditional mean and the GARCH component forecasts the error term.

ARIMA (AutoRegressive Integrated Moving Average)

The autoregressive (AR) component comes from the following mathematical model:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \varepsilon_t$$

where {ε } is a zero-mean white noise. We use the term autoregression since this model is actually a linear regression model for x_t in terms of the explanatory variable x_{t-1}. Basically, x is being modeled as regression on its own past. In the AR model, the ε is uncorrelated with past values of the AR series x . We can think of {ε} as a series of random shocks or innovations. The above equation is the general AR(p) form of the model, where the number of terms in the equation is the term p (excluding error term). The moving average (MA) component comes from the following mathematical model:

$$x_{n+1} = \varepsilon_{n+1} + \beta_1 \varepsilon_n + \beta_2 \varepsilon_{n-1} + \cdots + \beta_q \varepsilon_{n-q+1} \quad ,$$

where the forward prediction of x is derived through weighting the past forecast errors. If β is positive, then adjacent terms of x will be positively correlated, so the above average x will tend to be followed by a further above average value. For a white noise series, β = 0 , an above average value is equally likely to be followed by an above average a b or a below average value. If β is negative, then an above average term is likely to be followed by below average term. In this model there is no correlation between the present value of x_t and all previous values apart from the most recent. The combined AR and MA model (ARIMA) model follows the following general equation:

$$x_t = [\alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p}] + [\varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \cdots + b_q \varepsilon_{t-q}] \quad .$$

Generalized AutoRegressive Conditional Heteroskedasticity (GARCH)

The GARCH portion models the error term {ε} that we previously just considered to be forecast error. In this model, the variance error is believed to be serially autocorrelated. Unlike the linear arima process, the GARCH model does not assume conditional volatility is constant, which is a significant improvement over the base ARIMA model. The GARCH models assume that the variance of the error term follows an autoregressive moving average process. The ARCH (q) model for the series {ε } is defined by specifying the conditional distribution of εt , given the information available up to time t −1. It consists of the knowledge of all available values of the series, and anything which can be computed from these values, e.g., innovations, squared observations, etc. The GARCH(p,q) adds on to this model by allowing for an extra set of terms to to allow σ_t^2 (variance) to have an additional autoregressive structure within itself. The equation can be described as the following:

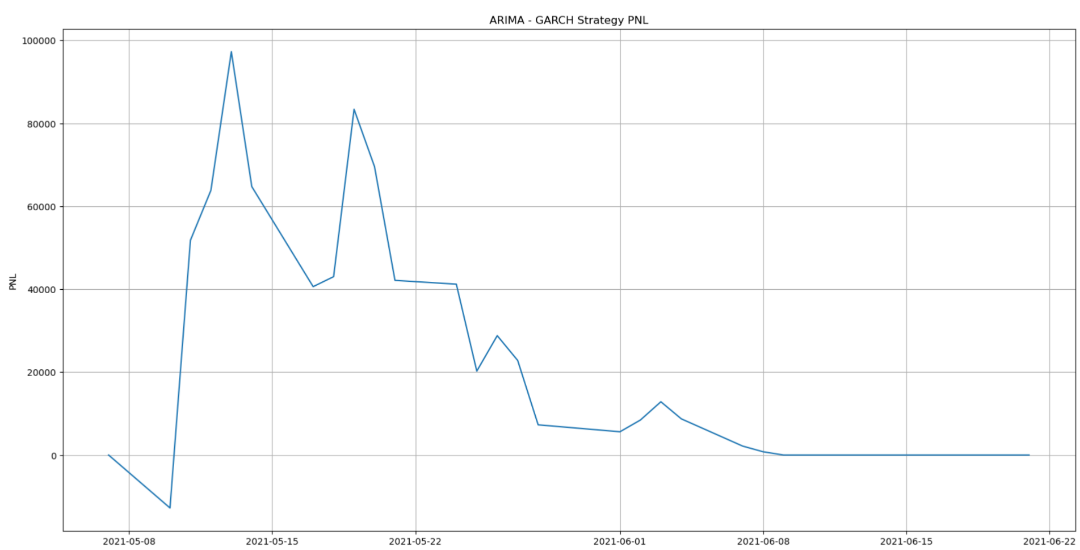
$$X_t = e_t \sigma_t$$

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2.$$

Results and Commentary

So in general, I seem to have made a solid PNL near the beginning of the strategy, however, towards the end it seems to taper off. This could be the result of the fact that the underlying was quite volatile during the period in which I was trading, but this could be a quirk of the model in selection. If I had to hypothesize on why the gains seemed more front ended, it would most likely be due to the fact that the spot went from 4210 to 4074 in only 5 days (just one example), and my strategy was able to cash in on the high volatility early. To speak a bit on the pure forecasting side of things, the model had pretty poor initial predictive power (less than 2.20% success), which when the modifier of 34 was included shot up to around 49.3%. However, this didn't end up being a significant problem, as almost every position ended with positive PNL. The benefits of the GARCH error forecasting are that it creates volatility clusters, and that shocks have heavy tails, but a tradeoff with this is that the model assumes that positive and negative shocks have the same effects on volatility as well as that ARCH models are likely to over-predict the volatility because they respond slowly to large isolated shocks in the time series. It so happens that the period in which I was forecasting had a large degree of volatility in the spot, which means that a shorter term memory model like the one I used was not able to capture those dramatic price swings well.

	Dates	m1_p1_PNL	m1_p2_PNL	m1_p3_PNL	m1_p4_PNL	m1_p5_PNL	Total PNL
0	2021-05-07	0.00	0.00	0.00	0.00	0.00	0.00
1	2021-05-10	-2746.35	-2692.50	-3410.50	-3051.50	-861.60	-12762.45
2	2021-05-11	12142.10	11728.25	12023.60	12304.65	3581.55	51780.15
3	2021-05-12	14972.07	14558.22	14893.15	15174.20	4214.83	63812.47
4	2021-05-13	22917.15	22503.30	22893.79	23174.84	5770.51	97259.59
5	2021-05-14	15110.78	14696.93	15032.83	15313.88	4569.53	64723.95
6	2021-05-17	9382.50	8928.31	9223.87	9504.92	3561.03	40600.63
7	2021-05-18	9951.64	9493.47	9800.97	10082.02	3680.43	43008.53
8	2021-05-19	19535.22	19077.05	19519.53	19800.58	5435.17	83367.55
9	2021-05-20	16195.38	15737.21	16156.17	16437.22	5011.81	69537.79
10	2021-05-21	9572.50	9114.33	9486.65	9767.70	4172.29	42113.47
11	2021-05-24	9353.95	8895.78	9263.45	9546.05	4139.74	41198.97
12	2021-05-25	4308.97	3850.80	4146.91	4429.51	3495.70	20231.89
13	2021-05-26	6332.32	5874.15	6270.71	6524.61	3782.70	28784.49
14	2021-05-27	4883.32	4425.15	4780.31	5044.56	3668.85	22802.19
15	2021-05-28	0.00	0.00	3508.79	3773.04	0.00	7281.83
16	2021-06-01	0.00	0.00	2669.29	2939.29	0.00	5608.58
17	2021-06-02	0.00	0.00	4095.19	4345.79	0.00	8440.98
18	2021-06-03	0.00	0.00	6295.96	6546.56	0.00	12842.52
19	2021-06-04	0.00	0.00	4234.54	4485.14	0.00	8719.68
20	2021-06-07	0.00	0.00	950.65	1201.25	0.00	2151.90
21	2021-06-08	0.00	0.00	222.04	562.04	0.00	784.08



Model 2: HARX Forecasting

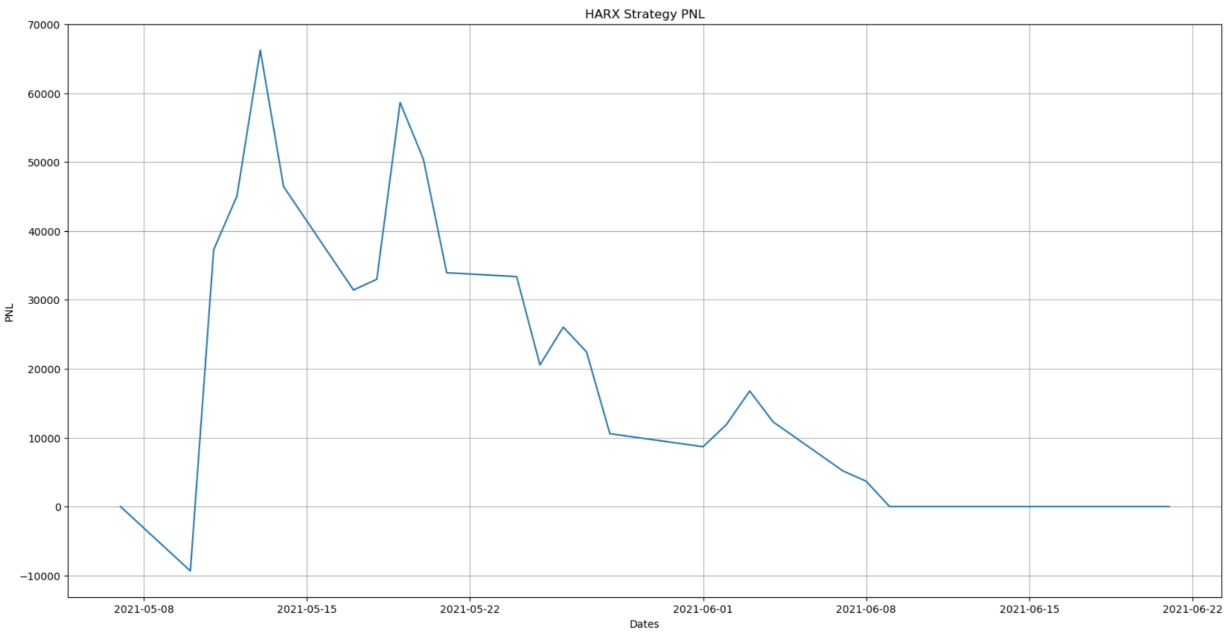
The HARX (Heterogeneous Autoregressive Model) is a natural extension to the AR model discussed before. We know that volatility exhibits a high degree of persistence and it's likely that a model is better forecast by using more lags, but for the sake of parsimony we usually resort to having 3 or fewer terms in our AR models. However, the creator of the HAR model hypothesized that for forecasting purposes all that might matter is the average level of volatility over the previous week and month. This results in HAR models being composed of three terms, a measure of daily volatility, weekly volatility, and monthly volatility, which are calculated as the AR average of 1 day, 5 days, and 22 days prior to the forecast. The equation for the HAR model can be described as the following:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_w RV_{t-1}^{w} + \ \beta_m RV_{t-1}^{m} + u_t$$

Results and Commentary:

The results of the HARX strategy were remarkably similar to that of the ARIMA GARCH strategy. In fact, both models seemed to capture almost the exact same contracts as each other. In the HAR model, we see almost exactly the same shape for PNL as we saw with the ARIMA GARCH model, and only one contract directly yielded a different result from both models. This could be explained by the fact that forecasts were only one step ahead (model was updated each day), and we would most likely need to forecast several steps ahead to see a notable difference in prediction intervals. From a pure forecasting perspective, the HARX generally performed better than the ARIMA GARCH at the base level, hitting 4.70% at the first time and 45.2% with the modifier. Again, forecasting power did not have much of an affect on the PNL of this strategy, which was also positive as well. The HAR model is a solid model for forecasting because it parsimoniously captures the strong persistence typically observed in RV, and has consistently good forecasting performance. However, it is highly sensitive to outliers, and suboptimal in the presence of conditional heteroskedasticity (which is a very real factor in options pricing). These drawbacks were not too much of an issue for me because there were not any outlier per say., but the PNL's were more or less the same for both models.

	Dates	m2_p1_PNL	m2_p2_PNL	m2_p3_PNL	m2_p4_PNL	m2_p5_PNL	Total PNL
0	2021-05-07	0.00	0.00	0.00	0.00	0.00	0.00
1	2021-05-10	-3410.50	-3051.50	-1023.15	-861.60	-1005.20	-9351.95
2	2021-05-11	12023.60	12304.65	5212.85	3581.55	4139.50	37262.15
3	2021-05-12	14893.15	15174.20	5885.71	4214.83	4911.31	45079.20
4	2021-05-13	22893.79	23174.84	7552.51	5770.51	6855.91	66247.56
5	2021-05-14	15032.83	15313.88	6242.35	4569.53	5272.80	46431.39
6	2021-05-17	9223.87	9504.92	5193.51	3561.03	3941.58	31424.91
7	2021-05-18	9800.97	10082.02	5324.85	3680.43	4096.80	32985.07
8	2021-05-19	19519.53	19800.58	7282.06	5435.17	6593.93	58631.27
9	2021-05-20	16156.17	16437.22	6835.18	5011.81	5958.89	50399.27
10	2021-05-21	9486.65	9767.70	5902.38	4172.29	4606.33	33935.35
11	2021-05-24	9263.45	9546.05	5865.18	4139.74	4553.63	33368.05
12	2021-05-25	4146.91	4429.51	5113.80	3495.70	3372.89	20558.81
13	2021-05-26	6270.71	6524.61	5472.55	3782.70	3989.94	26040.51
14	2021-05-27	4780.31	5044.56	5327.65	3668.85	3617.34	22438.71
15	2021-05-28	3508.79	3773.04	0.00	0.00	3290.63	10572.46
16	2021-06-01	2669.29	2939.29	0.00	0.00	3072.13	8680.71
17	2021-06-02	4095.19	4345.79	0.00	0.00	3450.43	11891.41
18	2021-06-03	6295.96	6546.56	0.00	0.00	3927.52	16770.04
19	2021-06-04	4234.54	4485.14	0.00	0.00	3576.64	12296.32
20	2021-06-07	950.65	1201.25	0.00	0.00	3017.68	5169.58
21	2021-06-08	222.04	562.04	0.00	0.00	2879.11	3663.19



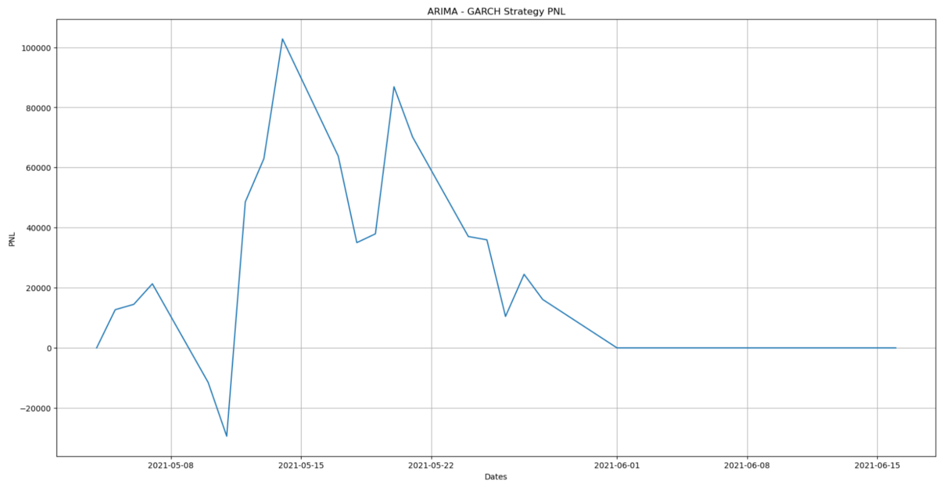
Manipulating Forecast Horizons

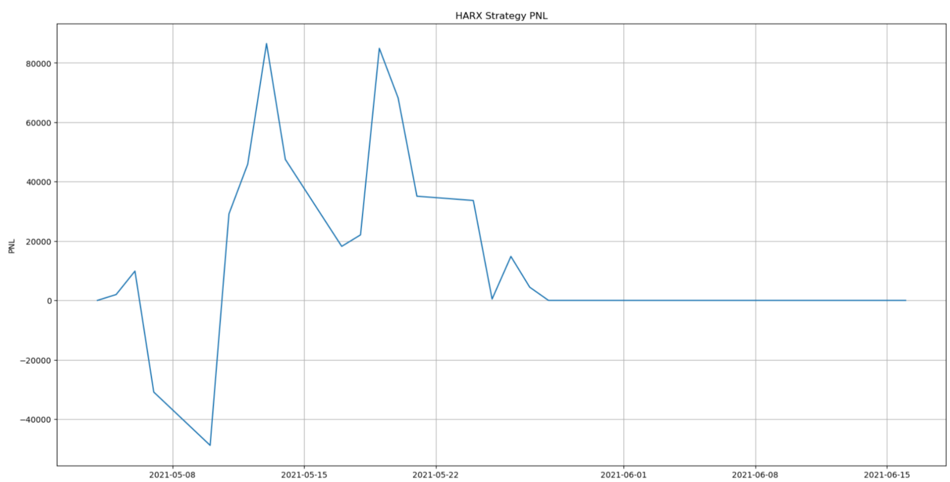
In order to create a more clear distinction between the models, I decided to expand the forecast from one step ahead to 5 and 7 steps ahead. The results were as follows:

5 Steps Ahead Results

	Dates	m1_p1_PNL	m1_p2_PNL	m1_p3_PNL	m1_p4_PNL	m1_p5_PNL	Total PNL
0	2021-05-03	0.00	0.00	0.00	0.00	0.00	0.00
1	2021-05-04	2575.06	2575.06	2549.18	2523.30	2484.48	12707.08
2	2021-05-05	2963.14	2955.22	2919.44	2853.96	2793.36	14485.12
3	2021-05-06	4499.62	4452.10	4250.00	4113.24	4013.04	21328.00
4	2021-05-07	-3039.98	-2098.70	-2094.80	-2107.96	-2125.76	-11467.20
5	2021-05-10	-6629.98	-5688.70	-5684.80	-5680.01	-5697.81	-29381.30
6	2021-05-11	8960.02	9901.30	9905.20	9909.99	9892.19	48568.70
7	2021-05-12	11869.15	12790.64	12774.75	12779.54	12741.95	62956.03
8	2021-05-13	19869.79	20735.72	20775.39	20724.62	20687.03	102792.55
9	2021-05-14	12118.01	12929.35	12969.02	12918.25	12880.66	63815.29
10	2021-05-17	6349.39	7160.73	7200.40	7189.97	7112.04	35012.53
11	2021-05-18	6954.35	7749.77	7781.48	7767.07	7685.16	37937.83
12	2021-05-19	16875.38	17603.31	17567.53	17485.63	17336.23	86868.08
13	2021-05-20	13535.54	14263.47	14227.69	14145.79	13996.39	70168.88
14	2021-05-21	6912.66	7640.59	7604.81	7522.91	7373.51	37054.48
15	2021-05-24	6687.91	7418.94	7383.16	7302.81	7153.41	35946.23
16	2021-05-25	1535.59	2302.40	2302.40	2222.05	2108.43	10470.87
17	2021-05-26	4405.59	5172.40	5158.05	5034.65	4705.78	24476.47
18	2021-05-27	2356.29	3299.05	3646.95	3564.95	3246.43	16113.67

	Dates	m2_p1_PNL	m2_p2_PNL	m2_p3_PNL	m2_p4_PNL	m2_p5_PNL	Total PNL
0	2021-05-04	0.00	0.00	0.00	0.00	0.00	0.00
1	2021-05-05	396.00	396.00	396.00	396.00	394.02	1978.02
2	2021-05-06	1980.00	1980.00	1980.00	1972.08	1962.18	9874.26
3	2021-05-07	-6260.00	-6218.80	-6218.80	-6144.32	-6030.62	-30872.54
4	2021-05-10	-9850.00	-9808.80	-9808.80	-9734.32	-9620.62	-48822.54
5	2021-05-11	5740.00	5781.20	5781.20	5855.68	5969.38	29127.46
6	2021-05-12	9539.68	9462.14	9026.76	8923.13	8957.67	45909.38
7	2021-05-13	17818.12	17629.46	17138.52	16979.33	16958.31	86523.74
8	2021-05-14	10011.75	9823.09	9332.15	9172.96	9151.94	47491.89
9	2021-05-17	4081.77	3933.45	3482.85	3364.00	3342.98	18205.05
10	2021-05-18	4877.77	4725.47	4266.91	4128.16	4075.30	22073.61
11	2021-05-19	18308.28	18021.00	17224.99	16478.83	14873.70	84906.80
12	2021-05-20	14968.44	14681.16	13885.15	13138.99	11533.86	68207.60
13	2021-05-21	8345.56	8058.28	7262.27	6516.11	4910.98	35093.20
14	2021-05-24	8035.56	7751.38	6964.67	6237.11	4670.73	33659.45
15	2021-05-25	879.56	631.16	23.35	-239.07	-839.39	455.61
16	2021-05-26	3749.56	3501.16	2893.35	2630.93	2030.61	14805.61
17	2021-05-27	1679.56	1431.16	823.35	560.93	-39.39	4455.61


ARIMA - GARCH Strategy PNL
100000
80000
60000
40000
20000
0
−20000
PNL
2021-05-08
2021-05-15
2021-05-22
2021-06-01
2021-06-08
2021-06-15
Dates


HARX Strategy PNL
80000
60000
40000
20000
0
−20000
−40000
PNL
2021-05-08
2021-05-15
2021-05-22
2021-06-01
2021-06-08
2021-06-15
Dates

7 Steps Ahead Results

	Dates	m1_p1_PNL	m1_p2_PNL	m1_p3_PNL	m1_p4_PNL	m1_p5_PNL	Total PNL
0	2021-05-03	0.00	0.00	0.00	0.00	0.00	0.00
1	2021-05-04	2575.06	2575.06	2549.18	2523.30	2484.48	12707.08
2	2021-05-05	2963.14	2955.22	2919.44	2853.96	2793.36	14485.12
3	2021-05-06	4499.62	4452.10	4250.00	4113.24	4013.04	21328.00
4	2021-05-07	-3039.98	-2098.70	-2094.80	-2107.96	-2125.76	-11467.20
5	2021-05-10	-6629.98	-5688.70	-5684.80	-5680.01	-5697.81	-29381.30
6	2021-05-11	8960.02	9901.30	9905.20	9909.99	9892.19	48568.70
7	2021-05-12	11869.15	12790.64	12774.75	12779.54	12741.95	62956.03
8	2021-05-13	19869.79	20735.72	20775.39	20724.62	20687.03	102792.55
9	2021-05-14	12118.01	12929.35	12969.02	12918.25	12880.66	63815.29
10	2021-05-17	6349.39	7160.73	7200.40	7189.97	7112.04	35012.53
11	2021-05-18	6954.35	7749.77	7781.48	7767.07	7685.16	37937.83
12	2021-05-19	16875.38	17603.31	17567.53	17485.63	17336.23	86868.08
13	2021-05-20	13535.54	14263.47	14227.69	14145.79	13996.39	70168.88
14	2021-05-21	6912.66	7640.59	7604.81	7522.91	7373.51	37054.48
15	2021-05-24	6687.91	7418.94	7383.16	7302.81	7153.41	35946.23
16	2021-05-25	1535.59	2302.40	2302.40	2222.05	2108.43	10470.87
17	2021-05-26	4405.59	5172.40	5158.05	5034.65	4705.78	24476.47
18	2021-05-27	2356.29	3299.05	3646.95	3564.95	3246.43	16113.67

	Dates	m2_p1_PNL	m2_p2_PNL	m2_p3_PNL	m2_p4_PNL	m2_p5_PNL	Total PNL
0	2021-05-04	0.00	0.00	0.00	0.00	0.00	0.00
1	2021-05-05	396.00	396.00	396.00	396.00	394.02	1978.02
2	2021-05-06	1980.00	1980.00	1980.00	1972.08	1962.18	9874.26
3	2021-05-07	-6260.00	-6218.80	-6218.80	-6144.32	-6030.62	-30872.54
4	2021-05-10	-9850.00	-9808.80	-9808.80	-9734.32	-9620.62	-48822.54
5	2021-05-11	5740.00	5781.20	5781.20	5855.68	5969.38	29127.46
6	2021-05-12	9539.68	9462.14	9026.76	8923.13	8957.67	45909.38
7	2021-05-13	17818.12	17629.46	17138.52	16979.33	16958.31	86523.74
8	2021-05-14	10011.75	9823.09	9332.15	9172.96	9151.94	47491.89
9	2021-05-17	4081.77	3933.45	3482.85	3364.00	3342.98	18205.05
10	2021-05-18	4877.77	4725.47	4266.91	4128.16	4075.30	22073.61
11	2021-05-19	18308.28	18021.00	17224.99	16478.83	14873.70	84906.80
12	2021-05-20	14968.44	14681.16	13885.15	13138.99	11533.86	68207.60
13	2021-05-21	8345.56	8058.28	7262.27	6516.11	4910.98	35093.20
14	2021-05-24	8035.56	7751.38	6964.67	6237.11	4670.73	33659.45
15	2021-05-25	879.56	631.16	23.35	-239.07	-839.39	455.61
16	2021-05-26	3749.56	3501.16	2893.35	2630.93	2030.61	14805.61
17	2021-05-27	1679.56	1431.16	823.35	560.93	-39.39	4455.61

Commentary

So as you could probably see above, both models seemed to select the exact same first five options, which is what I traded off of. Furthermore, when I compared capture rates and the number of contracts each model said it would trade, they were basically the same, but it was a significant jump from the one step ahead forecasts to the 5 and 7 step ahead forecasts (from around 20 contracts being traded to ~120-140). However, this time around the models did choose significantly different contracts to trade, and the PNL varied considerably. There might be two reasons for this difference, and it is worth noting in this case that the ARIMA-GARCH PNL is around 3.6x that of the HARX strategy. The first reason is more circumstantial, and I think that since there was a lot of vol in the underlying that the HARX model might've overpredicted some moves and missed out on mispriced options during those periods of high vol. This could also be a result of the fact that generally in more volatile times, ARIMA GARCH's lesser sensitivity to vol spikes might allow it to be better at picking up a few good opportunities for mispriced options.

Possible Future Steps

- **Using a stochastic volatility model for forecasting** might be interesting to see the difference in forecasts, as they model the error terms as completely random processes, and simulate results. Totally different prediction method, and would be interesting to see if it chooses different contracts based on its generated intervals

- **Comparing results in times of higher volatility with lower volatility**, since the models have pretty short term memory, I think that they would perform better in periods of high volatility, but it would be interesting to see what could happen
- **Change time interval**, since both these models treat data in discrete time intervals, there is a degree of interpolation between data points, and in daily close data (as I used for this pitch) we might miss significant opportunities in daily price swings. It would be interesting to feed the models on intraday data and compare results.