

## **Capstone Final Project: Model Selection and Justification**

Objective: The goal of this project is to explore different technical analysis strategies to evaluate their performance against a passive buy and hold investment strategy. By incorporating a number of investment strategies and machine learning optimization techniques such as hypertuning this should give the foundation of building my own investment strategy.

The project's problem: There are currently so many different ways to use technical analysis techniques and it can be confusing to understand which performs best. It has been traditional to use algorithmic trading but with the increase in power of computing and the use of machine learning/AI the power in optimizing investment strategies are not much further in depth than some trading rule. That is where that project is going to try and bridge the gap between traditional static strategies with dynamic and intelligent strategies.

The goal of this is to explore a variety of different technical analysis techniques which are

- Bollinger bands
- Relative Strength Index
- Moving Average Convergence Divergence
- Ichimoku Clouds

### **How do different technical analysis techniques combine to create a more powerful investment strategy?**

Each of these will be performed on Sector ETF's during different time periods. This is done so that the model can be built off of a variety of performing sectors which is meant to be most generalized to prevent overfitting. Another advantage is that you are able to see which sectors are best performing during different invest strategies and in different periods.

For example you might have a portfolio with 10 stocks  
Stock A might be AAPL

AAPL is a part of the technology (XLK) sector and if this sector has performed particularly well during an expansion then the strategy should hopefully be generalized to perform better when the stock is moving in a certain direction (suggestive of a business cycle). The stocks will have buy/sell signals from a variety of different strategies as listed above and this can also be incorporated into giving a specific weight of investment power. All of this combined should be able to create a powerful investment strategy.

By combining a number of these strategies whilst optimizing returns using a variety of machine learning techniques a final model should be created.

By utilizing the foundations of stochastic modeling where different start dates are going to be used to prevent overfitting a distribution of returns is going to deliver an expected return value.

### **Model 1. Buy and Hold Passive Strategy**

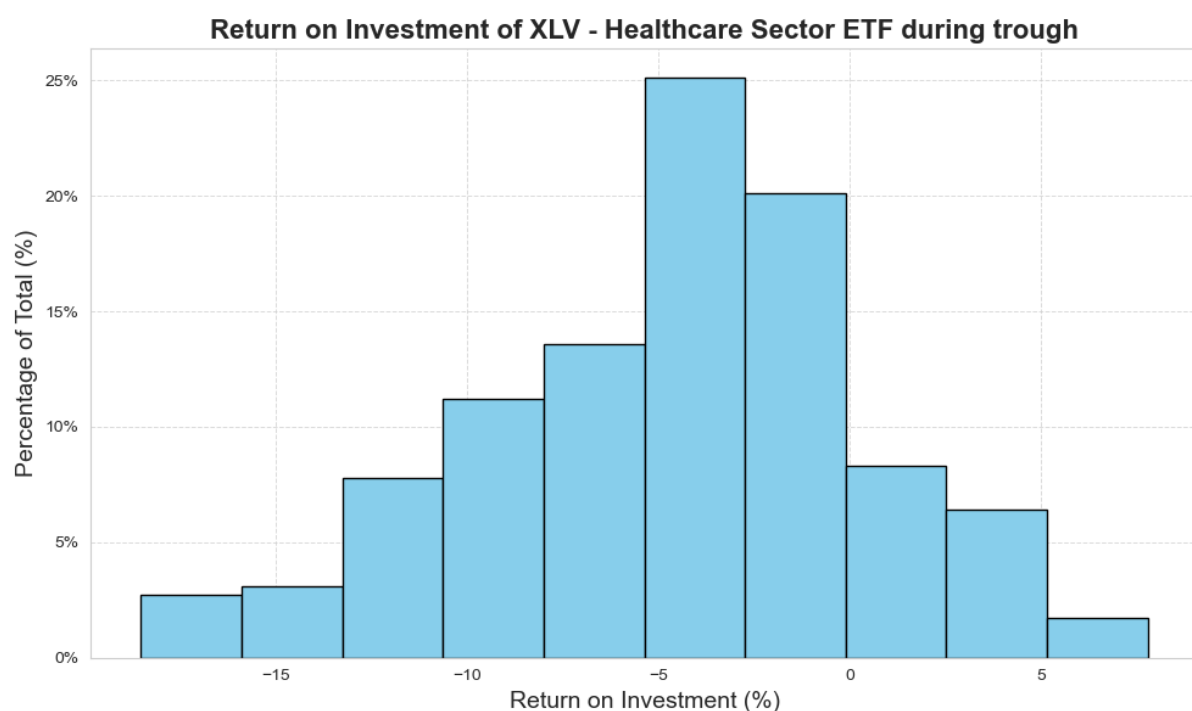
The buy-and-hold strategy is going to be used as the baseline investment strategy. It is a passive and low-stress strategy that any successful model should be able to outperform.

Stochastic Modeling Methodology for buy and hold strategy:

The following is going to occur for each Sector ETF within each period:

1. Choose 1000 start dates
2. Purchase a single stock
3. Hold onto the stock
4. Sell the stock at the end of the 90 day period
5. Record the ROI as a list in a nested dictionary
6. Repeat 2-5 for each 1000 iterations

The following is an example of a 90-day buy-and-hold distribution of the ROI of the healthcare sector ETF, XLV during a trough



Observe that the distribution is almost normal which suggests that for XLV during a trough you can expect almost a -5% return with a 5% return being extremely unlikely, dependent on the start date of investing.

### Sector Performance

The following is the table of returns using a buy-and-hold strategy during different macroeconomic cycles. These returns took place over 90 days.

	trough	expansion	peak	contraction	all_data
XLB	-0.801716	5.152597	4.193117	-0.124475	3.138665
XLI	-13.286162	6.518785	6.927373	-4.351198	3.68415
XLF	-21.345726	7.250742	9.636204	-15.203804	2.387638
XLK	3.413614	5.719647	12.687111	-2.766366	5.360743
XLY	-1.217808	6.99843	3.33871	-3.987359	4.34363
XLP	-8.142633	5.851742	5.118839	0.511474	3.093067
XLE	-4.607492	3.858138	-0.368145	3.699586	2.492215
XLV	-4.760442	9.183555	8.714912	-3.956509	3.334249
VOX	4.097503	5.555871	6.201991	-5.513514	2.878921
XLU	-8.70834	4.980092	6.263849	-2.757124	2.39118
IYR	-17.694321	4.359444	4.293125	0.28201	2.20345

The average result of this investment period combining all of the sector ETFs is:

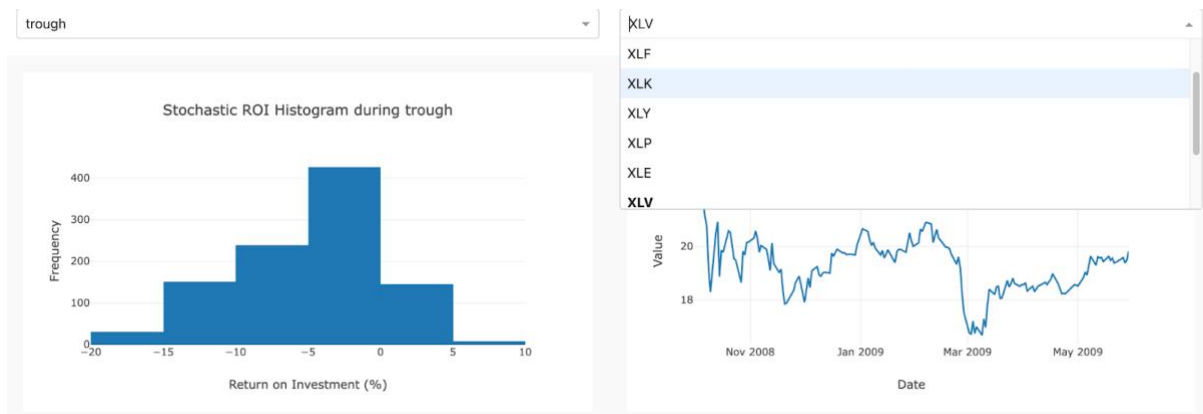
trough	-6.897239
expansion	5.779927
peak	6.101778
contraction	-3.130004
all_data	4.306979

Observe that the trough period performs the worst with sectors such as financials and real estate returning as low as -22% and 18% respectively. Meanwhile during a peak the financial returns 10% and technology a 12% this clearly shows that this investment strategy is heavily dependent on the time of investment. If you are able to understand the macroenvironment and only invest during an expansion and a peak ,expect a 6% return during a 90 day period however this is extremely hard to do. Making a model that tries to perform better when the overall economy decreases is going to be the challenge!

### Dashboard Creation

See video link to dashboard here - <https://www.youtube.com/watch?v=adQphpCyl4w>

The dashboard as shown below has a dropdown for time period and stock to show the distribution function of returns.



The dashboard would be extremely useful for individuals who are looking to invest into a stock. Again using AAPL as a stock you could look at the XLK technology sector ETF and see the likely return based on the macroeconomic cycle you are in.

The standard deviation is a metric used to measure risk and volatility. It indicates the dispersion from the mean, which can be seen in the distribution of results. As the standard deviation increases, there is a higher risk because the returns have a wider spread. When the standard deviation is lower, the expected return is much more precise.

	trough	expansion	peak	contraction	all_data
XLB	17.772145	5.720613	3.15482	10.688353	11.461349
XLI	12.273994	5.925943	3.270425	6.929078	11.035735
XLF	24.479161	5.923939	4.811564	6.487851	14.090727
XLK	12.862374	4.615319	5.402516	7.787906	10.552899
XLY	14.984387	5.309021	3.261831	6.008357	10.720121
XLP	5.924415	4.288111	1.481321	2.67845	5.954582
XLE	9.71644	8.998463	6.441554	15.193391	15.135068
XLV	5.166795	4.493787	5.778321	6.436054	7.309347
VOX	10.039646	5.412558	3.574633	6.992901	10.394956
XLU	4.699424	5.912746	1.966758	6.745153	7.500404
IYR	18.09904	6.454825	2.187071	8.149909	12.36354

The average standard deviation across the buy and hold strategy is

trough	12.365256
expansion	5.732302
peak	3.757347
contraction	7.645218
all_data	10.592611

### **Model 2. Bollinger Bands**

Bollinger bands use momentum indicators to create a channel for price of stock to move between, if it falls below a buy signal is triggered and if it rises above a sell signal is indicated. This can be used as a technique to invest in stocks.

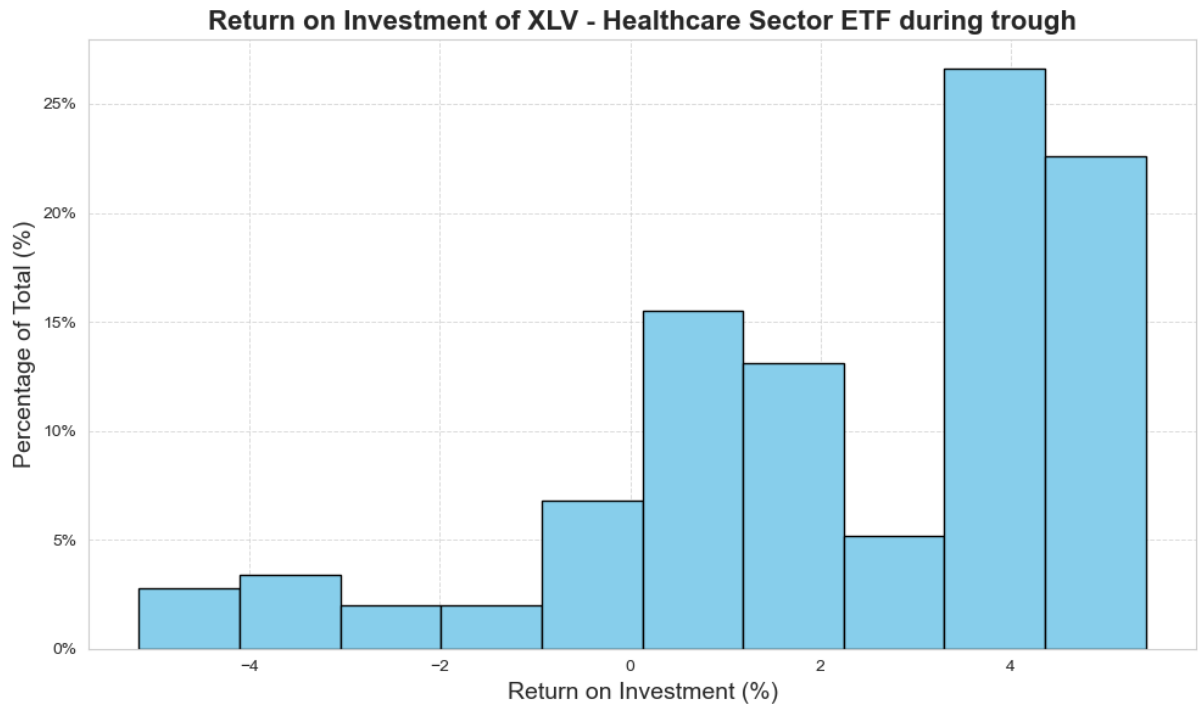
The parameters that go into bollinger bands are the rolling average period and the standard deviation or the confidence interval that is associated.

### **Model 2. Baseline Bollinger Band Model**

The baseline model is what John Bollinger himself suggested which was to use a 20 day buy with a confidence interval of 95% or about 2 standard deviations above and below the 20 day rolling average.

Baseline Bollinger Bands Stochastic Modeling Methodology:

1. Choose 1000 start dates
  2. Iterate through each ticker of each day
  3. Perform action
    - Buy: purchase 20% of account balance (starts at \$100)
    - Sell: Sell 20% of current stock holding
    - Hold: Go onto the next ticker/day
  4. Update the current holdings through multiplying stock price by number of shares owned
  5. Continue for the 90 day investment period
  6. Record ROI after the 90 day period
  7. Store ROI for each stock in each period as a nested dictionary
- The below is the return of a healthcare sector stock (XLV) during a trough time period



The following is a return on investment for each stock in each period

	trough	expansion	peak	contraction	all_data
XLB	4.545175	1.244462	2.090428	0.66712	0.690705
XLI	3.794377	1.547458	1.971617	0.044681	0.783543
XLF	2.526989	1.402378	3.090899	0.686678	0.806997
XLK	2.160918	1.33556	1.45019	-0.793992	0.960723
XLY	5.072016	1.396591	1.603604	1.09939	0.981233
XLP	1.518364	1.194837	0.632582	0.569678	0.810053
XLE	3.447952	0.872342	1.148966	0.72667	0.832747
XLV	2.242881	1.499922	1.070105	-0.262099	0.821963
VOX	6.012704	0.866472	1.444071	-1.174485	0.740857
XLU	-0.341954	0.561168	0.842866	-1.186619	0.656775
IYR	1.824177	0.525455	0.677302	1.835057	0.900798

The following is the overall return for each period using the baseline bollinger band model, observe that during a trough it performs quite well.

```
trough      2.982145
expansion   1.131513
peak        1.456603
contraction  0.201098
all_data     0.816945
dtype: float64
```

### Model 3. Parameter Optimized Bollinger Band Model

The rolling average [10,15,20,25,30]

The confidence interval would be [0.85,0.90,0.95]

The grid search is an optimization technique which will brute force test each different combination to try and find the highest ROI

	<b>rolling_window</b>	<b>confidence_interval</b>	<b>average_roi</b>
13	30	0.90	1.169210
12	30	0.85	1.125352
9	25	0.85	1.108568
10	25	0.90	0.978441
6	20	0.85	0.928941
3	15	0.85	0.876273
4	15	0.90	0.763554
7	20	0.90	0.762654
0	10	0.85	0.727782
14	30	0.95	0.726472
11	25	0.95	0.634777
1	10	0.90	0.594852
8	20	0.95	0.529623
5	15	0.95	0.337114
2	10	0.95	0.275850

The highest average ROI is a rolling window of 30 and a confidence interval of 0.90 which means that the higher returns take place where there are more buy and sell signals that occur.

Perform stochastic modeling with optimized values:



	<b>trough</b>	<b>expansion</b>	<b>peak</b>	<b>contraction</b>	<b>all_data</b>
XLB	8.042618	1.721901	2.207053	1.008785	0.575778
XLI	5.657168	2.356069	1.806753	-0.359526	0.801396
XLF	10.230909	2.297216	1.871583	-1.983297	0.722754
XLK	8.846397	1.319866	1.646746	-0.523981	1.240709
XLY	10.940854	2.210262	2.074202	1.387191	1.035198
XLP	1.503228	1.550296	0.476947	1.56653	1.010457
XLE	6.020594	0.274044	1.673338	-1.289695	0.33185
XLV	4.63952	1.787505	1.802676	-0.197021	0.995615
VOX	10.523264	1.614777	2.057181	-1.336171	0.935801
XLU	0.682048	0.906088	1.568088	-1.58423	1.002208
IYR	9.941065	0.929345	1.103506	2.388579	0.697984

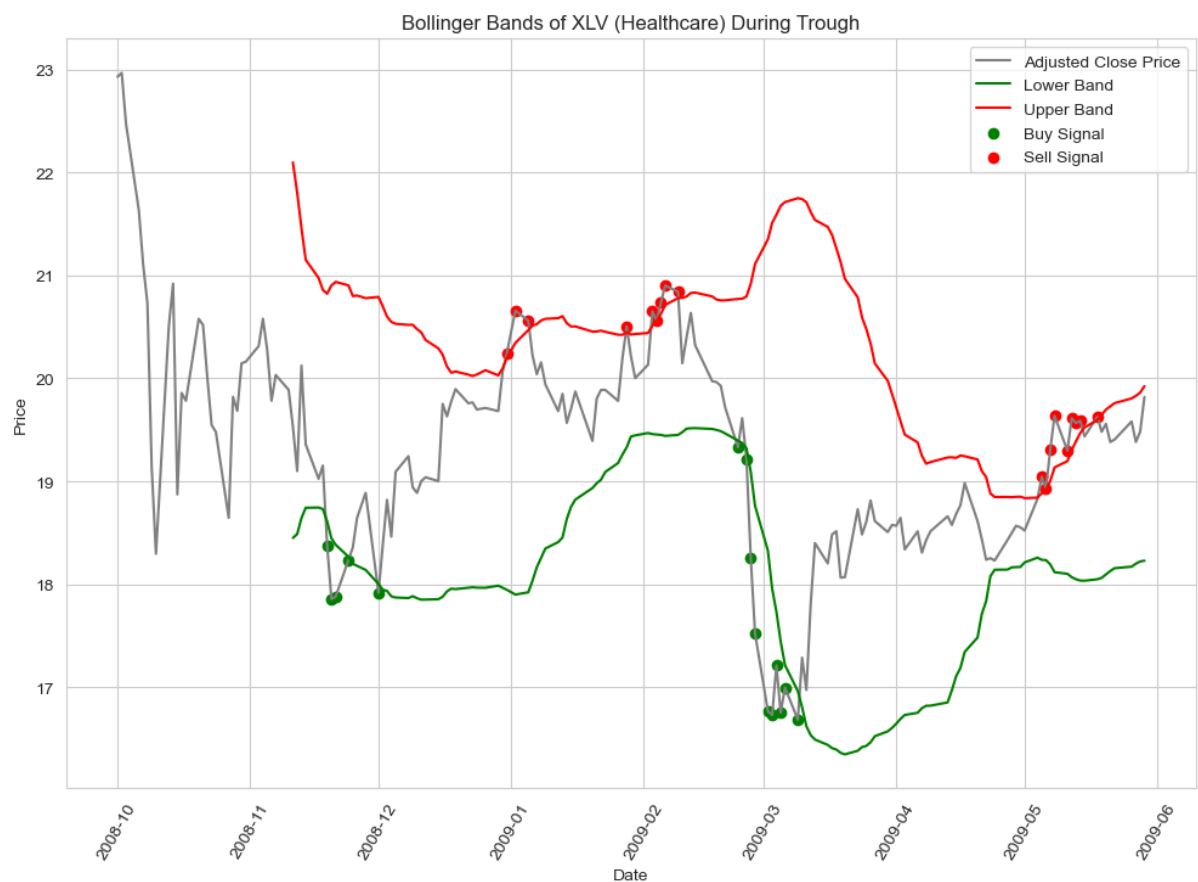
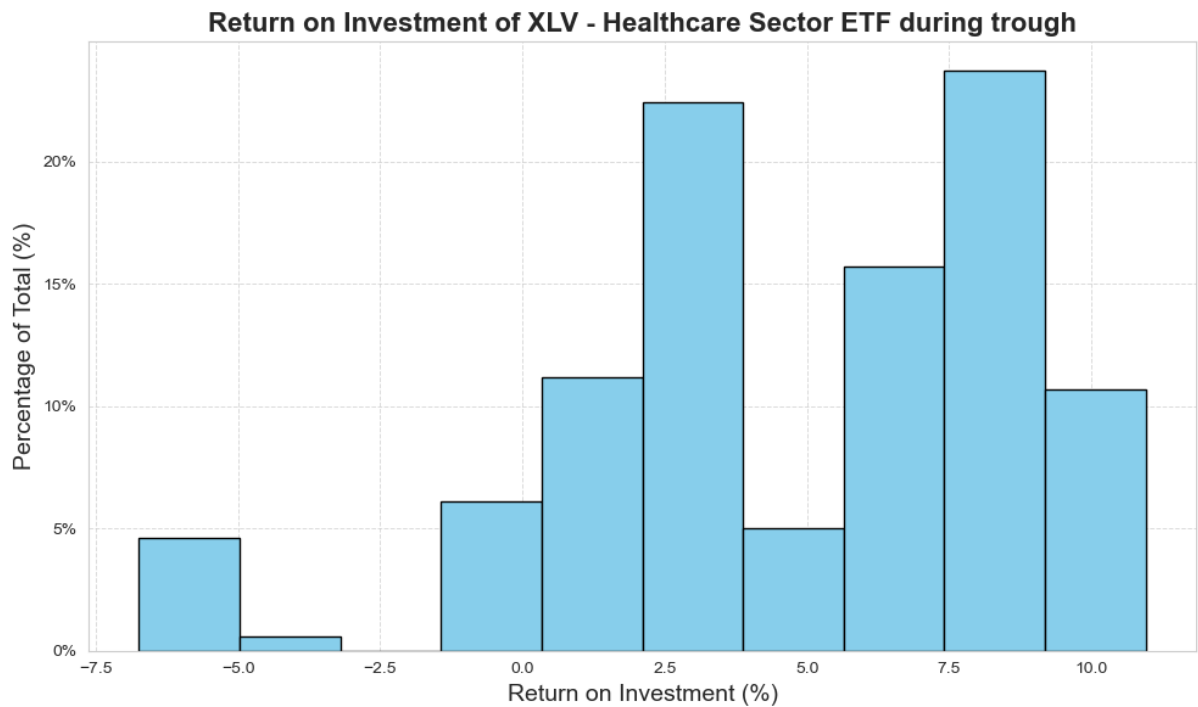
There is a greater improvement to the model with some sectors such as financial seeing a 10% return on a 90 day investment.

The average return across the different sectors

<b>trough</b>	<b>7.002515</b>
<b>expansion</b>	<b>1.542488</b>
<b>peak</b>	<b>1.662552</b>
<b>contraction</b>	<b>-0.083894</b>
<b>all_data</b>	<b>0.849977</b>

It is clear that Bollinger Bands do well during a trough and the optimized parameters boost that to be an even stronger indication.

The following is the healthcare sector ETF return during a trough time period



Observe that the investment period takes longer to begin, so you are only actually getting 60 days of availability to trade. However during that time there are more trades that take place because the confidence level is only at 90%. I believe that there can be a multiplier effect so that during these cluster periods of buy and sells there is more investment with each incremental signal.

### Evaluation of base model against benchmarks (buy and hold strategy)

The base model showed that during the trough time period the success rate was far better, over 10% better with some industries showing returns of 25% on average over 1000 iterations which is quite impressive for a 90 day investment period.

An investigation can be done to understand why the trough and contraction showed better performance, it is likely because of the pace of the improvement of peak and expansion outpaces the profits made from an investment strategy such as bollinger bands.

	trough	expansion	peak	contraction	all_data
XLB	7.470721	-3.813003	-1.612373	0.028536	-2.5022
XLI	18.20149	-5.080882	-4.480696	4.017265	-3.041776
XLF	25.485514	-5.705638	-6.017982	15.934575	-1.986317
XLK	-0.196842	-4.346143	-10.801574	1.681744	-4.211055
XLY	7.725168	-5.524039	-1.588255	5.30898	-2.876046
XLP	10.204989	-4.470547	-4.361884	0.266263	-2.239774
XLE	9.357896	-2.994891	1.524977	-3.943486	-2.613981
XLV	7.294864	-7.405278	-7.267788	3.954364	-2.446141
VOX	3.266331	-4.451902	-4.380335	4.233821	-2.101935
XLU	8.338845	-4.468636	-5.349021	1.04563	-2.45464
IYR	20.047647	-3.595702	-3.559509	1.869503	-1.574561

This shows the average return

trough	10.654238
expansion	-4.714242
peak	-4.354040
contraction	3.127018
all_data	-2.549857

### Evaluation of optimized model against benchmark (buy and hold strategy)

By using the optimized parameters as seen above in the model creation it performs well but it performs even better when you compare it to the benchmark.

	trough	expansion	peak	contraction	all_data
XLB	9.535226	-3.291987	-1.981974	0.266731	-1.920474
XLI	19.707817	-4.47671	-5.055674	3.727432	-2.444517
XLF	33.409749	-5.093585	-7.795042	13.859548	-1.719924
XLK	5.721678	-4.41776	-10.95206	2.292937	-3.62285
XLY	12.487448	-5.0023	-1.30426	5.696816	-2.332732
XLP	10.007532	-4.405705	-4.714826	1.072344	-1.977947
XLE	11.250417	-3.250985	1.88486	-6.254872	-2.394736
XLV	9.49858	-7.560319	-6.896231	4.12736	-2.199595
VOX	6.510152	-3.877776	-4.089092	4.537263	-1.674742
XLU	9.587605	-4.24575	-4.80078	0.835445	-1.9285
IYR	28.135532	-3.507482	-3.321833	2.197681	-1.273847

You can again see that during the trough period the stocks all perform well with Financial, Real Estate and Industrial Sectors all performing 20-33% better than passive investment strategies. Meanwhile it again performs about 5% worst during peaks and expansions.

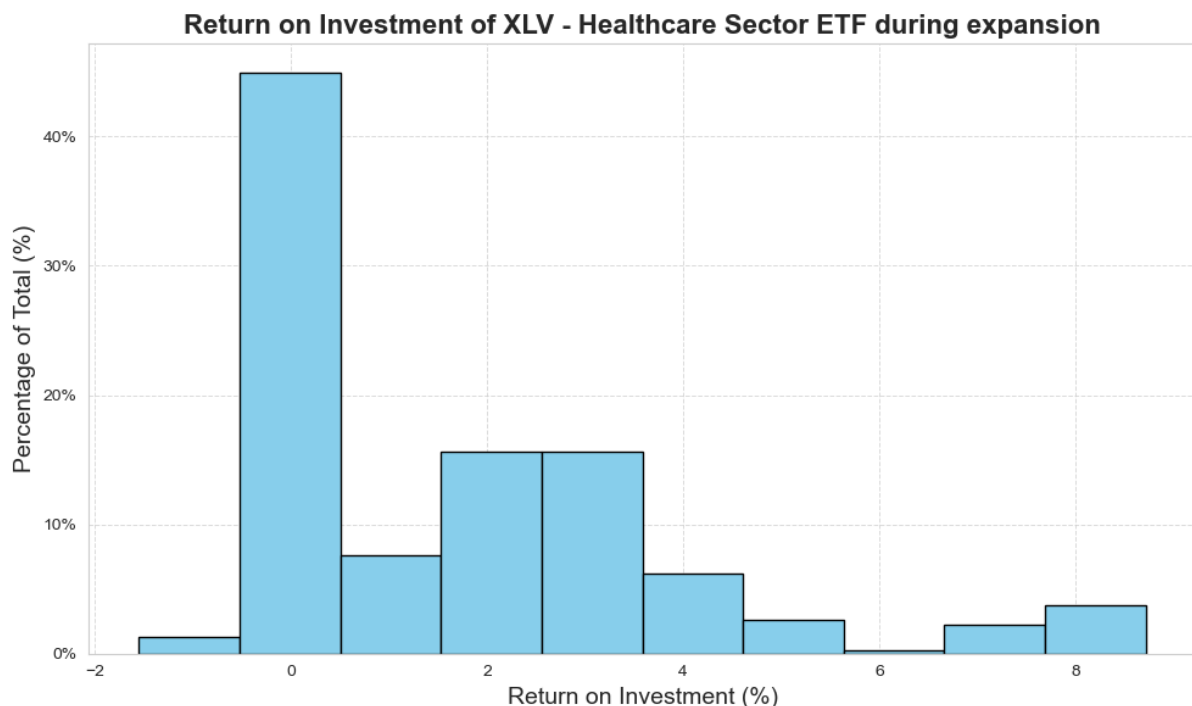
trough	14.168340
expansion	-4.466396
peak	-4.456992
contraction	2.941699
all_data	-2.135442

It performs 14% better during a trough and overall you can see in some industries such as Finance it performs 33% better than a traditional buy and hold strategy. It is clear that the performance during peak and expansion periods (where economic activity is higher) the technical analysis strategy cannot keep up with the performance of the overall market.

#### Model 4. Optimizing investment amounts Bollinger Band Model

The amount that is invested for each buy/sell signal may greatly impact the profits. If you take bigger investments when the market is showing a higher likelihood of a rebound then you are likely to return a higher profit. Using grid search again, the bollinger band model is going to use the optimized 30 day rolling average period with a 90% confidence interval for investing.

Another reason this is important because as you can see below, the end of the investment period most of the values of returns are close to 0, this is because the number of sell signals are about twice as high as the buy signals.



Where there was 25 buy signals and 57 sell signals so it is likely that whenever a stock is bought it is not held on for long enough to see returns.

To try to over come, using different percentage multipliers to the amount to invest can be a good way to overcome this. Perhaps during an expansion, a higher amount to buy is going to perform better as the stock continues to rise with a lower sell rate.

The following will the parameters tested using the brute force grid search method

Amount\_to\_buy = [0.10,0.20,0.25,0.30]

Amount\_to\_sell = [0.10,0.20,0.25,0.30]

Gridseach results

	<b>amount_to_buy</b>	<b>amount_to_sell</b>	<b>average_roi</b>
13	0.30	0.20	2.738156
6	0.20	0.25	2.478402
11	0.25	0.30	2.354784
9	0.25	0.20	2.269526
5	0.20	0.20	2.211863
8	0.25	0.10	2.163808
14	0.30	0.25	2.130303
12	0.30	0.10	2.118689
10	0.25	0.25	2.065724
15	0.30	0.30	1.984020
4	0.20	0.10	1.935747
7	0.20	0.30	1.922692
3	0.10	0.30	1.807983
2	0.10	0.25	1.780429
0	0.10	0.10	1.778946
1	0.10	0.20	1.775225

The higher ROI are all quite similar and the amount invested does not seem to make too large of a difference when looking at the average ROI. However, a slightly higher buy amount compared to sell amount is going to return the highest returns, with the combination of having a lower buy amount and a higher sell amount returning less favorable returns. This is likely because there are too many sell signals with too high of a sell amount which does not allow for the value of shares to increase.

Performing stochastic modeling using amount to buy as 0.30 and amount to sell as 0.20 with a 0.90 confidence interval and a 30 day rolling average

	<b>trough</b>	<b>expansion</b>	<b>peak</b>	<b>contraction</b>	<b>all_data</b>
XLB	8.084809	1.939313	2.175693	0.736057	1.47482
XLI	6.109358	2.344157	1.869727	-0.34065	1.427816
XLF	10.675616	2.2804	1.927837	-1.372914	1.054825
XLK	8.818159	1.395846	1.739324	-0.412845	1.687891
XLY	11.086948	2.198826	2.174299	1.760213	1.49431
XLP	1.607423	1.506292	0.529383	1.609285	1.199448
XLE	6.104101	0.72266	1.615231	-1.864918	1.230688
XLV	4.719168	1.703973	1.713492	-0.092934	1.290925
VOX	10.555388	1.592997	2.077507	-1.083539	1.218523
XLU	0.848035	0.836308	1.516674	-1.812094	1.201043
IYR	10.341867	0.814036	0.997426	2.633443	1.034854

The following were the mean returns on the investment

<b>trough</b>	<b>7.177352</b>
<b>expansion</b>	<b>1.575892</b>
<b>peak</b>	<b>1.666963</b>
<b>contraction</b>	<b>-0.021900</b>
<b>all_data</b>	<b>1.301377</b>

### RSI Model

The goal of the RSI is to create a dataframe of signals based on the thresholds as explained above. This can then be used to analyze the performance of incorporating RSI signals in comparison to a passive buy and hold strategy.

Use n days for the time period where you are looking at average gain/average loss. It is typical to use 14 days as recommended by J. Welles Wilder but longer and shorter time periods are also utilized.

Another parameter that can be looked at is the minimum RSI value before a buy/sell signal is created.

## Methodology

1. Difference the data
2. Track average gain/loss where gain is the positive movement from day to day and loss is the negative movement
3. Create the RS via the average gain/average loss
4. Create the RSI value by  $100 - (100/(1+RS))$

### Model 5. RSI Base Model

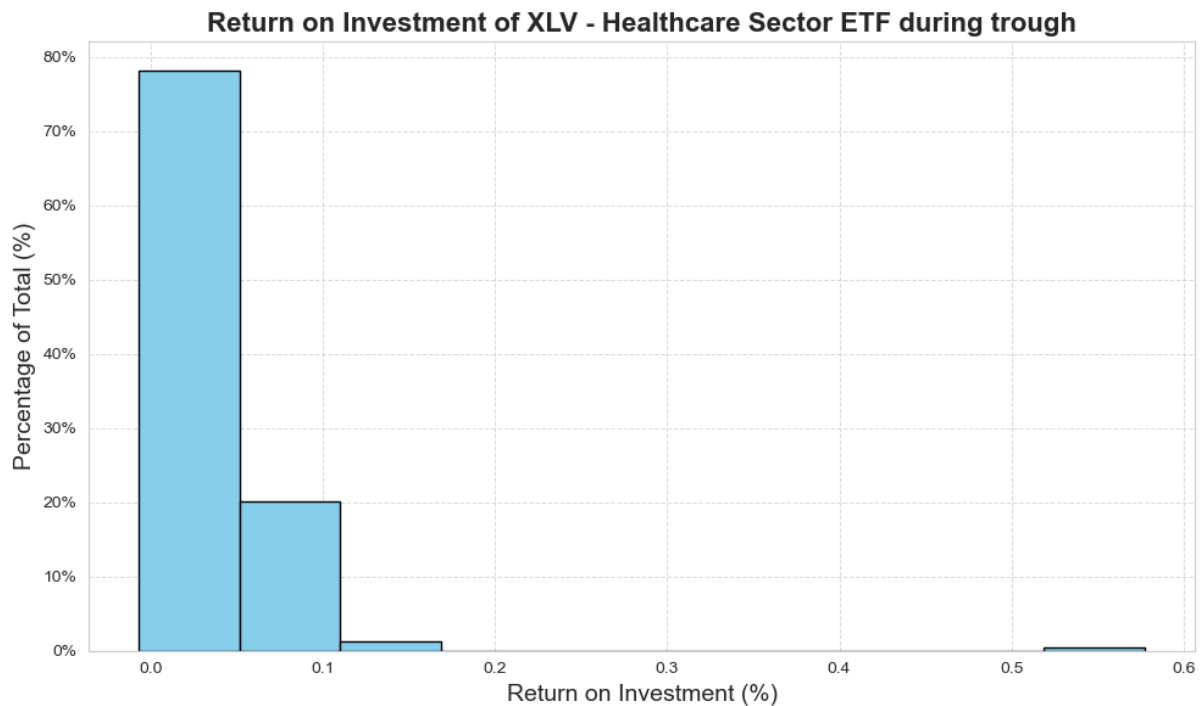
- 14 day investment period
- 0.20 buy/sell amount
- 0.30 as the minimum to trigger a buy signal
- 0.70 to trigger a sell signal

	trough	expansion	peak	contraction	all_data
XLB	-6.062995	-0.760002	-0.068952	-0.072808	-0.938557
XLI	-1.164835	-1.501483	1.264356	-0.486026	-0.594141
XLF	-94.866598	-3.142823	-0.881535	-42.454886	-8.297766
XLK	-0.569589	-0.949457	1.882154	-1.226761	-1.088577
XLY	-0.264764	-0.340079	10.654978	-3.970506	-0.397608
XLP	-0.882723	-0.785248	0.0	0.085674	-0.226629
XLE	0.088464	0.92247	2.026147	-0.105323	0.164995
XLV	0.023359	0.008633	0.63537	-9.30539	-0.569769
VOX	0.769874	0.527419	1.956271	0.163558	1.318814
XLU	-0.260447	-2.969168	-0.173927	-0.119009	-1.41444
IYR	-6.754216	-1.448055	-3.922092	0.62573	-1.122947

The average roi for different time periods using the RSI signals.

trough	-9.994952
expansion	-0.948890
peak	1.215706
contraction	-5.169613
all_data	-1.196966





## MACD

The MACD technical analysis will create the following parameters:

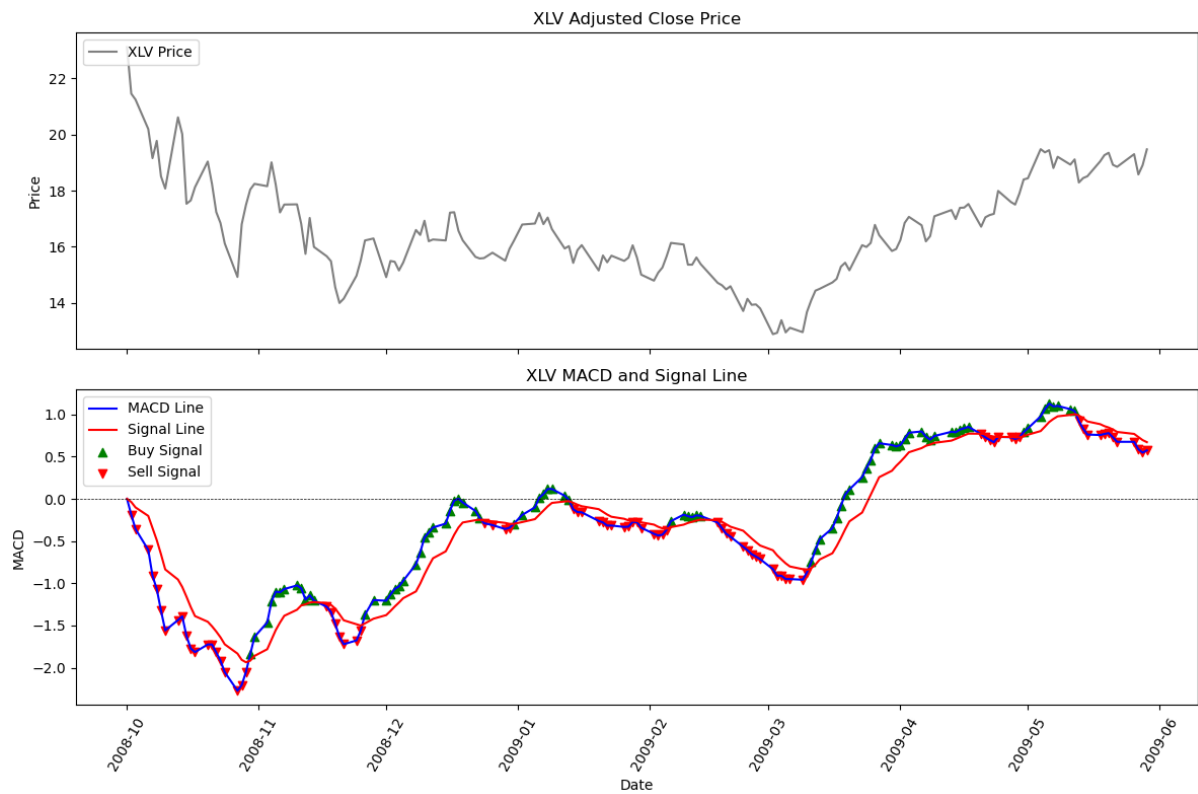
- Short Exponential Moving Average (12-day)
- Long Exponential Moving Average (26 days)
- MACD Line: The difference between Short EMA and Long EMA
- Signal Line: EMA of the MACD line

By utilizing an exponential moving average it gives a greater weight to the more recent values, this makes it more sensitive to recent changes in price. The MACD line measures the convergence or divergence between fast and slow moving averages. When the MACD line is positive (above 0), the short term trend is stronger than the long term trend which suggests an upwards trend. Where as when the MACD line is negative (below 0) the short term is weaker than the long term trend which suggests downwards momentum.

The signal line is used as the buy and sell signal by smoothing the MACD line using a 9 day ema period. When the MACD line goes above the signal line it indicates a buy signal and the MACD line goes below the signal line it indicates a sell signal.

	trough	expansion	peak	contraction	all_data
XLB	2.223146	4.371161	4.709639	2.74568	5.39825
XLI	1.1145	4.499344	4.33136	3.763293	4.605288
XLF	-2.203351	2.30887	3.787912	1.815807	3.51773
XLK	3.134058	4.65809	6.649532	5.700821	5.353619
XLY	1.680266	5.324184	7.169371	3.768419	4.934382
XLP	1.587954	5.14465	5.333487	3.635571	4.829118
XLE	4.860537	4.570936	2.54982	3.675122	5.645957
XLV	4.311102	4.934449	7.516143	6.23347	5.377811
VOX	2.529552	5.964052	5.482569	3.625926	5.457352
XLU	4.41519	5.930813	8.22497	5.135808	5.643733
IYR	0.869297	6.673431	6.518015	2.941624	5.603821

trough	2.229296
expansion	4.943635
peak	5.661165
contraction	3.912867
all_data	5.124278



## Ichimoku Cloud

	trough	expansion	peak	contraction	all_data
XLB	-11.511676	-2.601211	4.557582	1.081471	-1.761156
XLI	-6.950527	-2.232283	4.186573	-1.472503	0.867622
XLF	-11.901469	-12.936917	-2.654943	-5.685049	-8.280768
XLK	-18.03106	-1.312423	5.544916	-1.559594	-2.122587
XLY	-8.457993	2.459957	10.208738	-6.233795	1.155507
XLP	-3.90847	-3.241483	9.837462	-8.036019	-1.041676
XLE	-3.279311	4.331577	1.04828	1.18059	2.028666
XLV	-3.709393	0.852362	-2.34638	-6.848324	0.316075
VOX	2.259706	2.802487	0.676973	1.806382	1.86525
XLU	-21.907585	-3.560861	3.113038	-1.404659	-3.249921
IYR	-1.00838	-1.017075	10.985952	-5.084883	2.509452

trough	-8.036923
expansion	-1.495988
peak	4.105290
contraction	-2.932398
all_data	-0.701230

