Time Series Modelling effects on Pairs Trading Profitability Among US Sector SPDR ETFs

Team 5

April 2018

Members

- Yuanyuan Liu
- Wyatt Marciniak
- Jimit Sanghvi
- Muye Zhang

Objective

Our project analyses time series modeling effects on pairs trading strategies between highly correlated US sector Exchange Traded Funds (ETF). Our hypothesis is that modeling a time series' trends and behaviors away will offer a less biased correlation effect between ETFs and allow us to create pairs trading strategies that will yield a greater return. This will be observed in the model residuals, which when modelled correctly will show no correlation or discernible trend and resemble a white noise series.

Data

We have collected 10 years of daily data from the SPDR ETFs for the 10 sectors of the US economy. This excludes Real Estate, as it's much more recent inception date offers a much smaller time frame of data to analyze. We have written a data mining script in R to collect data through the 'quantmod' package and utilize the daily prices as well as calculating the daily returns.

Methodology

Our project consists of multiple modelling and testing stages.

- (1) Analyzing Correlations between ETFs using market price and daily return data
- (2) Analyzing Correlations between ETFs using mean time series modelled residuals
- (3) Analyzing Correlations between ETFs using mean-variance time series modelled residuals
- (4) Optimizing pairs trading backtests for each highly correlated pair per analysis process
- (5) Creating and testing trading portfolios utilizing the optimized pairs trading strategies

For each step of the analysis, we will be closely monitoring how the correlations between the ETFs change and how the strongest correlated pairs are selected. We will also be analyzing how changing pair relationships based on the modeling techniques do or do not improve the profitability of the strategies.

Methodology: Testing Market Price and Return Correlations

To test the correlations between the ETFs, we rank the relationship of 9 ETFs to the target ETF (running this procedure for all 10 ETFs). We take the absolute value of their correlation from a correlation matrix to identify which relationships are the strongest regardless of direction. The correlations are then ranked and plotted, labeled accordingly with their data set name and actual correlation coefficient (Figures 1 & 2). These plots offer a strong and simple way of selecting the most highly correlated pairs for trading. They also allow us to see the correlation behavior across the ETFs list as well as observe groups of potential candidates for an optimized pair.

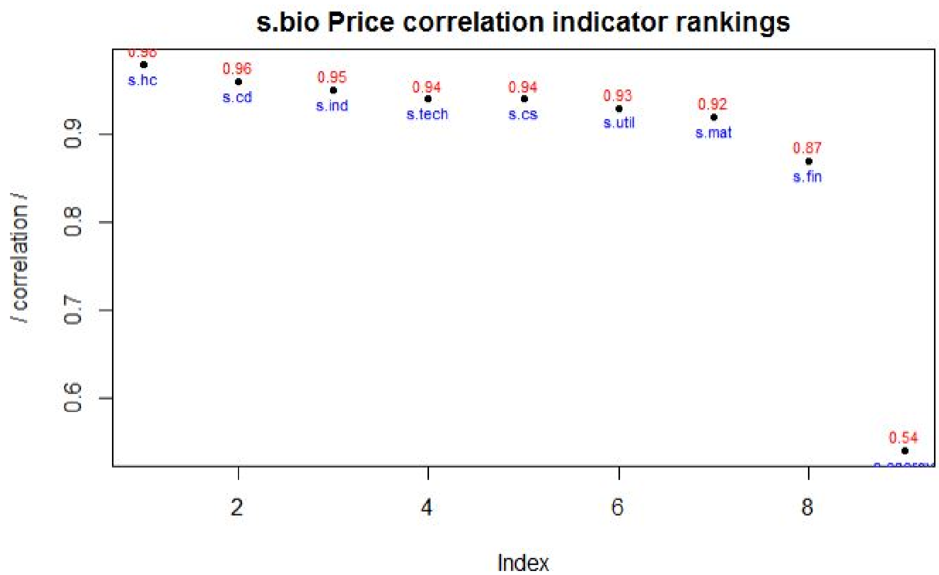

Figure 1: Biotech Sector ETF: Ranked Correlations based on market share prices

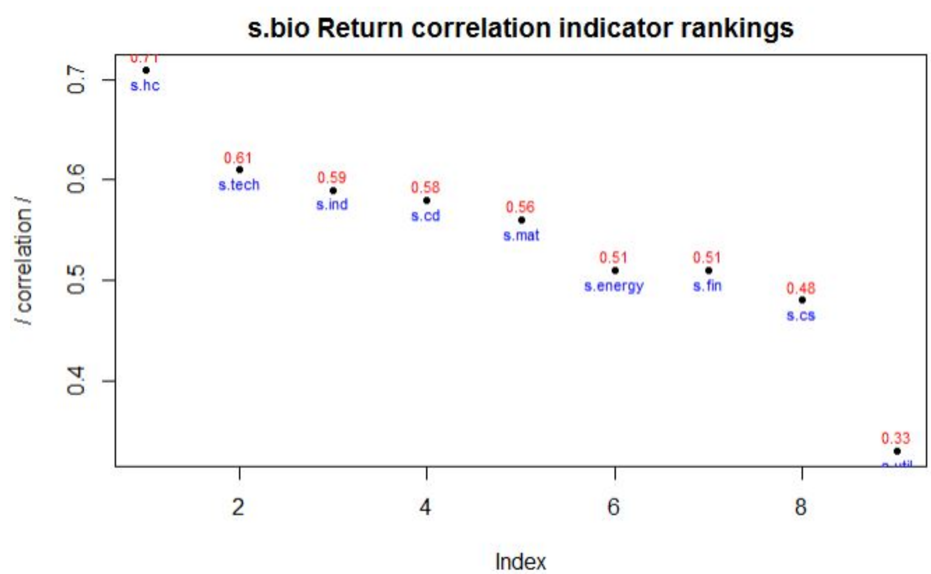

Figure 2: Biotech Sector ETF: Ranked Correlations based on market daily returns

Current Results: After completing this analysis, we have identified the optimal pairs to be tested in a pairs trading strategy for both price and return (Figure 3). Interestingly, while price correlations tend to be higher than return correlations among the ETFs, for the Energy sector ETF, this behavior is reversed. In addition, when cross analyzing the price and return correlation rankings, most ETFs show that their most correlated sector is consistent across the data type in all except for the Energy Sector. We will need to do some further analysis as to why this behavior is occurring, however in terms of our project objective at this time this observation will not affect our final results. What is really interesting here is that, when comparing by price correlation, most ETFs are correlated with the Consumer Discretionary sector (s.cd, 3 selections), Consumer Staples (s.cs, 2 selections) and Industrials (s.ind, 3 selections). When comparing that to the returns correlations, we see that Industrials are most highly correlated with 6 of the 10 ETFs. We should also note that the correlations in terms of returns are lower than those of the price correlations. This makes sense as price movements swing much more greatly then actual prices and those movements appear to reveal a stronger correlation indication.

```
- Pairs Chosen:
     OBS            Price         /  Return

(1)  s.bio    ~ s.hc  (0.98)    /  s.hc   (0.71)
(2)  s.cd     ~ s.cs  (0.99)    /  s.ind  (0.86)
(3)  s.cs     ~ s.cd  (0.99)    /  s.ind  (0.75)
(4)  s.energy ~ s.mat (0.60)    /  s.mat  (0.83)
(5)  s.fin    ~ s.ind (0.95)    /  s.ind  (0.80)
(6)  s.hc     ~ s.cd  (0.99)    /  s.ind  (0.76)
(7)  s.ind    ~ s.cd  (0.98)    /  s.tech (0.86)
(8)  s.mat    ~ s.ind (0.97)    /  s.ind  (0.86)
(9)  s.tech   ~ s.ind (0.98)    /  s.ind  (0.86)
(10) s.util   ~ s.cs  (0.95)    /  s.cs   (0.64)
```

Figure 3: ETF Market Price and Daily Return best pairs selection)

Methodology: Analyzing ARMA modelled Time Series Residuals

We have fit ARMA(p,q,d) models to the 10 sector ETFs. The goal of doing this step separately from the the full modeling step (with volatility) is to increase the granularity of our analysis with respect to modeling effects. We have constructed these models and tested them for pairs creations as well as testing their residuals to identify models that will need volatility modeling as well. To construct the ARMA models, we have constructed functions to automate the ACF and PACF analysis by the following steps (figures 4 & 5):

- (1) Calculating the Significance Level for lag spikes
- (2) Extracting all lags that touch or cross this threshold
- (3) Validate their significance with Ljung-Box dependency tests
- (4) Create final order lists based on lags with dependency
- (5) Construct all AR(p), MA(q) and ARMA(p,q) models with significant lag orders
- (6) Select the best model by AIC criterion
- (7) Visualize all plots and outputs for manual analysis checking

Once the ETFs are modeled properly and their residuals are uncorrelated, we test the ETF correlations using the residuals data as the time series data. This process is the same as the one we followed for prices and market returns, and will remain a constant analysis tool. These modeled residuals series should offer a less biased correlation analysis and we should be able to see how the best pairs selection has changed as well as test those strategies for profitability.

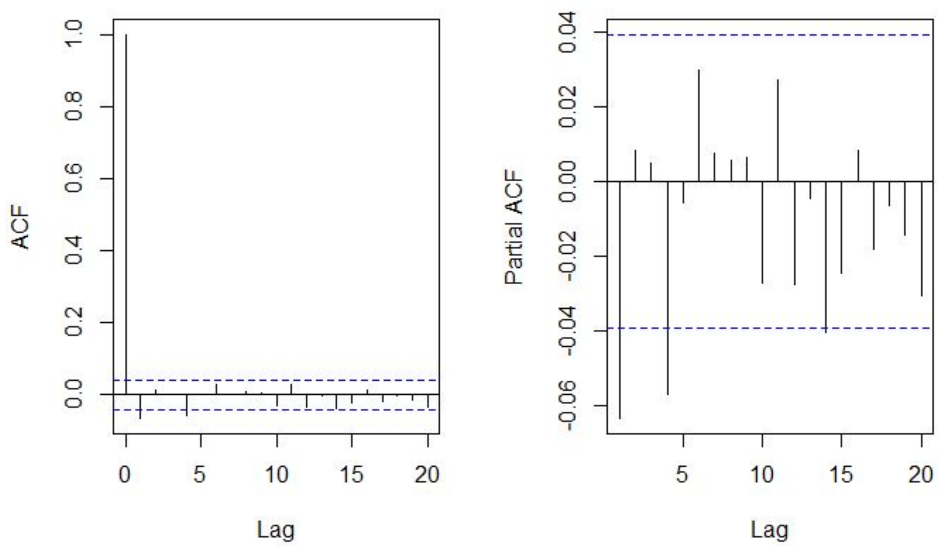

Figure 4: Biotech Sector ETF: ACF and PACF plots (Lag = 20)

```
[1] Creating AR models...
[1] ARMA model (4,0,0) has AIC = -13562.9914873678
[1] ARMA model (11,0,0) has AIC = -13556.4033550399
[1] ARMA model (14,0,0) has AIC = -13558.4881269211
[1] ARMA model (18,0,0) has AIC = -13562.8095264218
[1] Creating MA models...
[1] ARMA model (0,0,5) has AIC = -13561.7605849569
[1] ARMA model (0,0,12) has AIC = -13555.9851937926
[1] ARMA model (0,0,15) has AIC = -13555.3541701606
[1] ARMA model (0,0,19) has AIC = -13559.6017741949
[1] Creating ARMA models...
[1] ARMA model (4,0,1) has AIC = -13561.7852753458
[1] ARMA model (4,0,5) has AIC = -13554.5546003424
[1] Auto-ARMA model (1,0,1) has AIC = -13561.7574134945
```

Figure 5: Biotech Sector ETF: Dependant Lags used to create and evaluate models

Current Results: After modelling the time series with fitted ARMA models, we identified the pairs (Figure 6) and they are the same pairs as we identified for market return data. The correlations have changed minimally and we believe that the ARMA modelling will not be enough to see major changes in correlation or pairs trading results. The residuals also show slight correlation possibilities that we will be investigating further (Figure 7). The spikes at larger lags are small, as is the significance level, but we will need to be sure that the ARMA models are properly fitted with a few dependency tests at these significant lags. This ACF pattern is consistent across the ETF data sets. We expect that when the strategies are aggregated into a portfolio, we will be able to see if a clearer distinction of improvement is present.

```
- Pairs Chosen:
       OBS               Residuals

(1)  s.bio    ~ s.hc   (0.71)
(2)  s.cd     ~ s.ind  (0.86)
(3)  s.cs     ~ s.ind  (0.74)
(4)  s.energy ~ s.mat  (0.82)
(5)  s.fin    ~ s.ind  (0.80)
(6)  s.hc     ~ s.ind  (0.75)
(7)  s.ind    ~ s.tech (0.86)
(8)  s.mat    ~ s.ind  (0.85)
(9)  s.tech   ~ s.ind  (0.86)
(10) s.util   ~ s.cs   (0.64)
```

Figure 6: ARMA residuals modelled Residuals correlation testing results

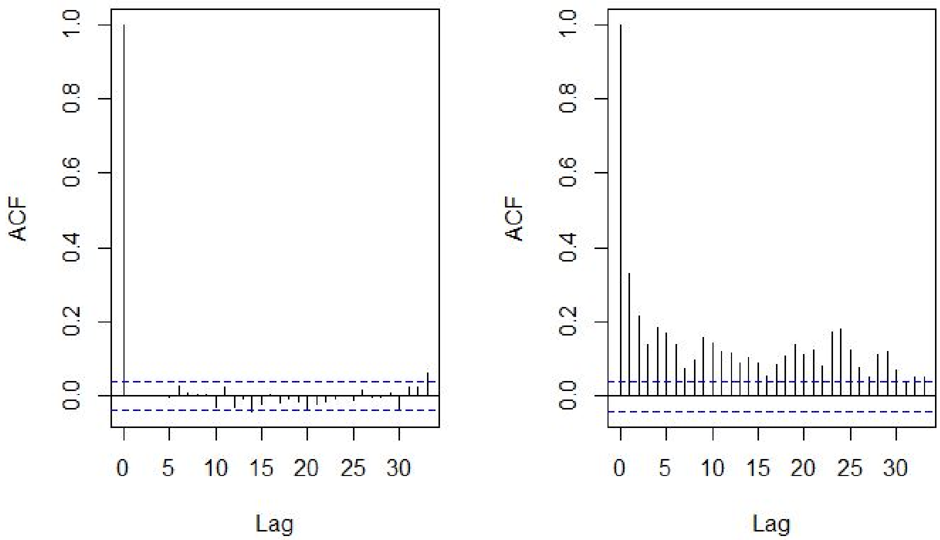

Figure 7: ARMA residual and squared residuals plots

Methodology: Analyzing Volatility Modeled Time Series Residuals

We will attempt to use ARCH/GARCH modeling techniques to model the volatility of the ETF time series. To identify the models that have volatility effects, we will observe the ACF and behavior of the squared residuals. We will then need to model the time series and the volatility simultaneously because modeling the time series based off of the residuals from the ARMA model means that we are modeling estimated parameters. We will only use the squared residuals of the ARMA models as indicators of possible GARCH effects. Once these models are constructed, we will test the residuals and squared residuals to ensure that our models are complete. We will then use the new residuals to test correlations between the ETFs, generate new pairs for trading and test the strategies.

Current Results: We are currently working on how to determine the order of GARCH models and use R packages to construct these models. This analysis will be our primary focus over the next 2 weeks once we have finished the ARMA models and finish the backtesting procedures for portfolio strategies. At this point, we have identified by the ARMA model squared residuals that volatility modelling is necessary (Figure 7).

Methodology: Optimizing pairs trading backtests for optimal pairs

The main testing procedure we are using is a pairs trading strategy we have constructed by hand. We use a ratio of the stock price, return or model residual as the distance or relationship series and calculate the mean of that series. We then set the signal lines to be 'k' standard deviations away from that mean line. These act as signals to open and close positions by denoting when the distance (spread) of the pairs prices/returns/residuals is abnormally large. This is possible because by picking 2 assets that are highly correlated, we know that their price and return movements in the markets will be closely related. This means up and down movements, as well as trends, are very similar and when the assets' relationship diverges from its normal behavior (the spread increases) we can capitalize on the hypothesis that this relationship will return to normal. We long/short the winner/loser depending on which asset moves too high and which moves too low. When their spread normalizes (come within the signal lines), we reverse our trade and take our profit. We have also included a stop loss parameter for a more realistic test when considering an investors risk tolerance. In addition to this, we have created an optimizer function that optimizes the back test strategy 'k' and stop loss percentage to give the best parameters for maximized return. We will use these tools to test the strategies and as well as optimize them for a portfolio backtest in our final conclusion.

Current Results: We have constructed the pairs trading function and optimizer. The pairs trading function runs the back test with the user specified 'k' and stop loss parameters and returns the profit, return and tracking list of all actions taken in the back test. The strategy is also plotted with green dots denoting opens, blue dots denoting closes and red crosses denoting stopped out positions. These outputs allow us to check the testing procedure for accuracy as well as analyze its effectiveness. The optimizer returns a list of results for a sequence of 'k' standard deviations with a fixed stop loss and a matrix where 'k' and the stop loss threshold are sequenced to identify the best parameters to maximize return. We have tested market price, return and ARMA residuals relationships and conducted optimized backtesting on the selected pairs in each category. As an example, we will show the progression of the Biotechnology and Health Care sectors pair trading through the 3 analysis procedures conducted so far. The standard deviation 'k' is 0.8 and the stop loss 's' is -0.05. We will also show the optimization results for the 3 backtests:

- (Figure 11) Market Price correlation - Return of 55%
- (Figure 12) Market Daily Return Correlation - Return of 28%
- (Figure 13) ARMA model residuals correlation - Return of 50%
- (Figure 14) Optimized parameter results (In order of models generated)

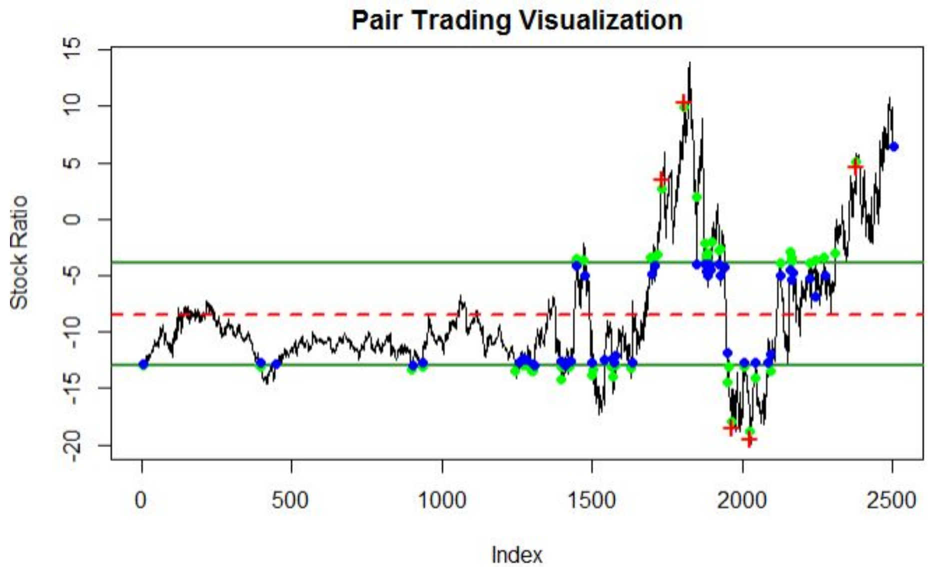

Figure 8: Biotech and Health Care Pairs Trade Backtest - Market Price Correlation

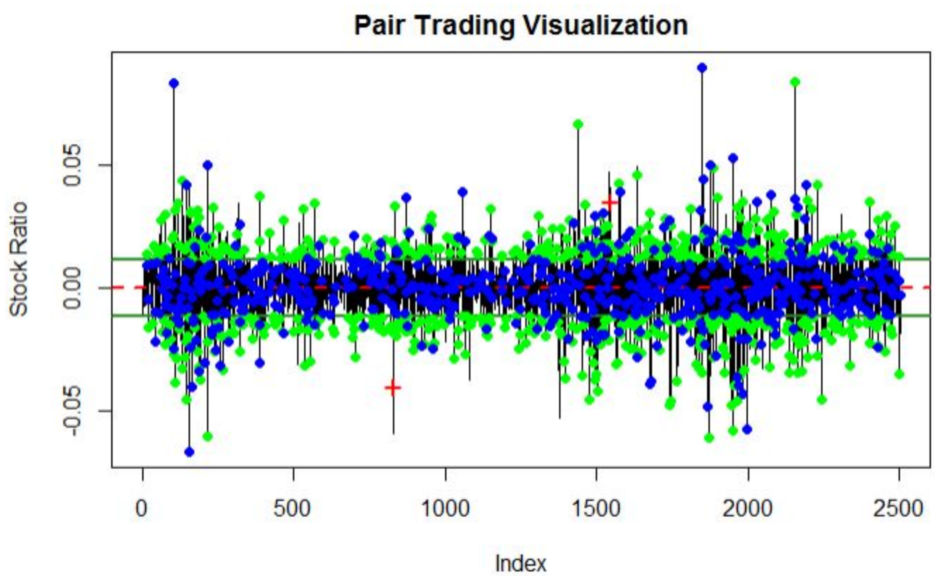

Figure 9: BBiotech and Health Care Pairs Trade Backtest - Market Return Correlation

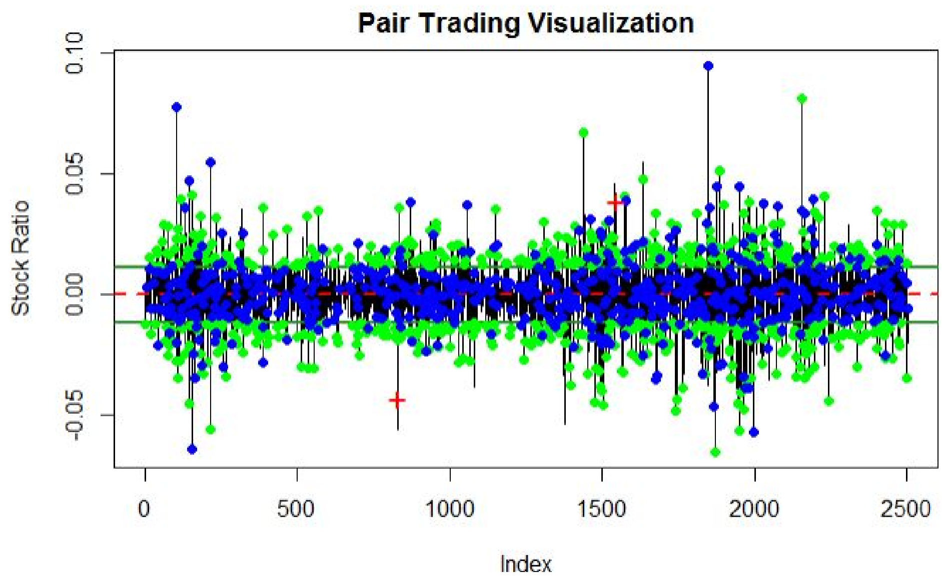

Figure 10: Biotech and Health Care Pairs Trade Backtest - ARMA residuals Correlation

```
Computing matrix of optimal k and stop losses
Optimized Params for strategy are k = 0.5 and s = -0.01 with return = 110.52 %
Computing matrix of optimal k and stop losses
Optimized Params for strategy are k = 0.4 and s = -0.01 with return = 91.96 %
Computing matrix of optimal k and stop losses
Optimized Params for strategy are k = 0.5 and s = -0.01 with return = 103.15 %
```

Figure 11: Biotech Sector ETF: Optimized Parameter returns

Optimizing the parameters made a significant difference in the profitability of the pairs trading strategy. Currently, we have observed that the market prices seem to produced a larger profit than returns, however the ARMA model residuals perform well when compared to returns. We expect that when we model the GARCH effects we may be able to see the residuals performing very well and potentially provide more profitability than the price relationships.

Methodology: Testing trading portfolios utilizing the optimized pairs trading strategies

In our final test we will use the 10 optimal pairs for each modelling process in a 'portfolio' backtest, where we will allocate an even amount of capital to each strategy. They will run them independently and the returns will be aggregated to simulate a portfolio return (to avoid multithreading and complex back-end capital account management). This is being done to show the possible applicability of our research in the real world. We will then compare multiple portfolios to see which perform the best as well as how much difference there is between the performances. We will compare the following portfolios:

- Holding the SP500
- Holding the DJIA
- Holding the 10 Sector ETFs
- Pairs trading with market price ratio
- Pairs trading with ARMA modeled residuals ratio
- Pairs trading with ARMA + GARCH modeled residuals ratio

Current Results: We are working on this part of the project currently and do not have any concrete results to report at this time

Next Steps

We are working to finish the ARMA modeling procedures as well as the backtesting functionality for portfolio returns. We must still conduct GARCH modelling and we will be expanding our analysis of the correlation-driven pairs to include the testing of multiple pair possibilities by level of correlation. This means that if multipl ETFs are correlated very highly to a target ETF then we will need to analyze them all through an optimized backtest to identify the proper pairs. We are on track complete this project in full and produce concise results on the viability of utilizing time series modeling to improve pairs trading results.