Statistical Arbitrage

Shane Barratt Russell Clarida Mert Esencan Francesco Insulla
Cole Kiersznowski Andrew Perry*

June 8, 2020

1 Introduction

According to Samuelson and Fama's Efficient Market Hypothesis (EMH) - namely, that prices of assets contain all available information - it is impossible to profit by trading on the basis of an information set. Eugene Fama is largely credited for assembling and extending upon pieces of theory established by Bachelier, Mandelbrot, and Samuelson in a 1970 book entitled "Efficient Capital Markets". Since agents are utility-maximizing, rationally motivated and update their expectations whenever new information appears, it assumes a now common no-arbitrage condition.

Statistical Arbitrage (StatArb) strategies, on the other hand, are based on the belief that such opportunities do exist. By assuming that risky assets get mispriced, one can then assemble long and short positions on these assets, which are assumed to be mean-reverting in a random walk formulation. In this fashion, StarArb strategies attempt to generate a positive expected payoff (alpha).

To the financial journalism community, StarArb often goes hand in hand with High Frequency Trading (HFT), which is a type of quantitative trading based on holding periods on the magnitude of milliseconds. However, we would note this is not always true. HFT strategies often marry computational power, speed, and statistical models to exploit micro-mispricings within the market. HFT can lead to Sharpe Ratios that are multiples of those of traditional trading strategies in these short time frames. We do not employ HFT strategies, the most commonly understood of which is Latency Arbitrage, in this work.
Ip

Most StatArb strategies take advantage of mispricing based on a pair of stocks that move together by taking short and long positions simultaneously when they diverge from the mean. By assembling "signals", one is able to, with some range of certainty, determine which stocks do indeed move together and what constitutes as divergence from the mean. In all forms of StatArb, the researcher is looking for co-integrated baskets, as we will demonstrate.

*Authors listed in alphabetical order

In this paper, a StarArb strategy is proposed, and back tested. As we discuss later, we use a universe containing 21 ETFs that contain US equities. We mainly use baskets that are simply pairs of these ETF's. These baskets are chosen from ETF's that are correlated naturally by their formulation. Then regression techniques are employed to construct baskets. These baskets are later analyzed with an altered version of the AD Fuller test to investigate if the choices were reasonable.

In this paper, we will first go into how we collected our data and the important decisions we made regarding which data to use. We will then go into the methods we used in model selection, focusing on our use of a Ornstein-Uhlenbeck process combined with an AR(1) process. Next, we present our results of the backtest of our model on the period 2007-2020 and analyze the strengths and limitations that we find, including what environments in which our model performed well and what environments in which our model performed poorly. Finally, we present our conclusions.

2 Data Collection

As mentioned above, we decided to use ETF data from the United States equity market for our project. We actually initially started by gathering a larger universe of data including equity data from the SP 1500, but we found that accounting for complexities such as delistings and other corporate actions became exceedingly challenging. Thus, we downloaded data for the following 21 US ETF's: SPY, MDY, DIA, XLK, XLV, XLF, XLP, XLY, XLU, XLE, XLI, XLB, QQQ, IVV, IWB, IWF, IJH, IJR, IWN, IWD, IVW, and IVE. Several of these ETF's have a lot of overlap and are thus highly correlated, so we would expect the spread of pairs including two such ETF's to be mean reverting. We collected data from the NYSE TAQ WRDS on each of these ETF's on a minute-by-minute basis for the period of 2003-2020 (we have data available up to and including present day). Unfortunately, since several of the ETF's have somewhat low volume at the beginning of this period, we actually start our trading strategy in 2007, as you will see in the analysis and results section below. Finally, we used a Yahoo Finance API to get corporate actions and to adjust for splits (several of the ETF's in our universe did actually split in the time period of our data). The entire process resulted in over a million price points for our model to use, which was definitely one of our paper's major strengths.

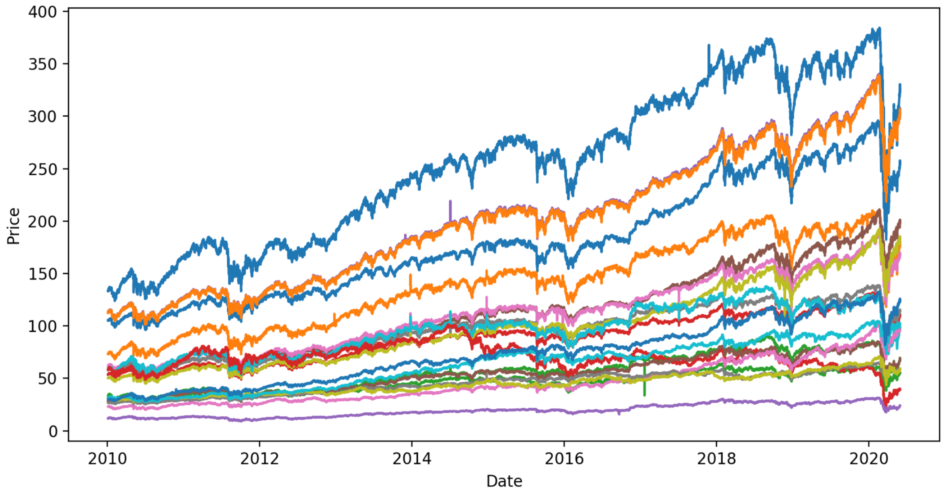

Figure 1: Price data of all ETFs

Above is a graph of the price of each of the ETF's in our dataset over the entire period we collected data. It is easy to see the correlation in the returns of many of the ETF's. There are also several large jumps in the data; we think these come from flash crashes or just general outliers in the data. We attempted to remove as many of these as possible, but, as we will discuss later in the paper, there were still some significant outliers in our data.

3 Model and Assumptions

Let R_t be the returns of an ETF, and F_t, be the returns of a market following index such as SPY. By regressing R_t on F_t, we get the residuals, $X_t = R_t - \hat{\beta} F_t$. Our strategy relies on the key assumption that the residuals can be modelled as an Ornstein-Uhlenbeck (OU) process. An OU process is analogous to a random walk with a tendency to move back to its center and it can be expressed as SDE:

$$dX_t = \rho(\mu - X_t)dt + \sigma dB_t, \quad \rho, \sigma > 0 \tag{1}$$

In Equation (1), ρ is the speed of mean reversion, μ is the long run average, σ is the instantaneous volatility, and B_t is a standard Brownian motion. This model is both Markov and Gaussion in nature so its constants can be estimated with an autoregressive process.

For the model a lag of 1 time period was assumed, so that the current term was dependent only on the previous term in the time series plus some Gaussian noise. That means that an AR(1) process can be used to calculate the parameters of the OU model.

An AR(1) process is given by

$$X_{t+1} = \lambda + \phi X_t + \epsilon_t \tag{2}$$

The interpretation of λ, ϕ, and ϵ as it relates to the OU process is

$$\begin{array}{rcl} \lambda & = & \mu(1-e^{-\rho\delta t}) \\ \phi & = & e^{-\rho\delta t} \\ \epsilon_t & \sim & \mathcal{N}\left(0, \frac{\sigma^2}{2\rho}(1-e^{-2\rho\delta t})\right) \end{array} \tag{3}$$

For 21 ETFs, one was used (SPY) to neutralize our baskets against the market. To find the weights of our portfolio of ETFs, we regressed all the $\binom{20}{2} = 190$ possible baskets, and uses the individuals residuals as a collection of signals.

While this approach is quite rudimentary, we found it to have a relatively good performance, as is shown in the next section. A more complete approach to finding optimal baskets can be developed using the Augmented Dickey–Fuller test, which we now discuss.

Suppose we have N assets, over a time T, and we would like to find the optimal linear combination of assets that results in a mean reverting signal. Let $X \in \mathbb{R}^{T\times N}$ be a matrix of returns, and $\alpha \in \mathbb{R}^N$ be the vector of weights. The signal has the form $Z = X\alpha$. Given an α, we would like to know how mean reverting Z is. This can be done using a ADF test. Suppose $Z \sim$ AR, meaning

$$Z = \rho BZ + \varepsilon$$
$$\Delta Z = \underbrace{\delta}_{\rho-1} BZ + \varepsilon$$

The ADF test has a null hypothesis of $\delta = 0$, corresponding to a random walk process. The t-statistic is $t = \frac{\hat{\delta}}{se(\hat{\delta})}$ and the p-value is then

$$p = 2\left(1 - \int_0^{t_0} t_{T-N}(t)\mathrm{d}t\right)$$

Since $\int_0^{t_0} t_{T-N}(t)\mathrm{d}t$ is a monotonically increasing function in t_0, to minimize p, it suffices to maximize t_0. We can therefore define the optimal basket, α^*, as

$$\alpha^* = \underset{1^T\alpha=1}{\operatorname{argmin}} -t_0(\alpha)$$

Thus we can optimize the objective function to get the most co-integrated baskets.

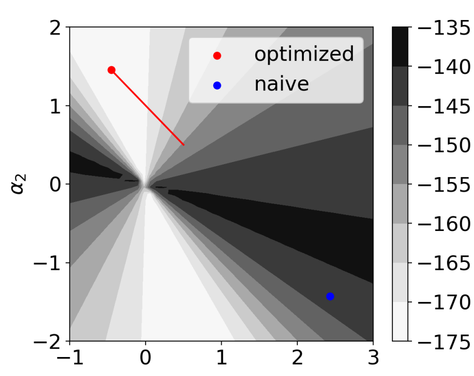

Figure 2: Contour plot and values of objective function

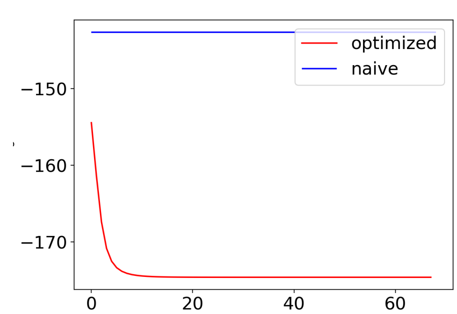

Figure 3: Objective function value by iteration

In Figures 2 and 3, we show an example of the optimization a two asset basket, comparing it to the naive regression approach. The optimization is done using projection gradient descent, implemented in jax. This shows that the regression residuals are not necessarily most mean reverting signal we can produce.

Optimizing one basket is relatively fast (0.3-1 seconds), however computing all the two asset optimal baskets would take approximately 3 minutes, which would be too slow. Thus, an alternative, is to use a greedy approach: optimize baskets of size one, then consider baskets of size two that add to the previously found one and so on. To evaluate the validity of this method, we take a sample in the data and check.

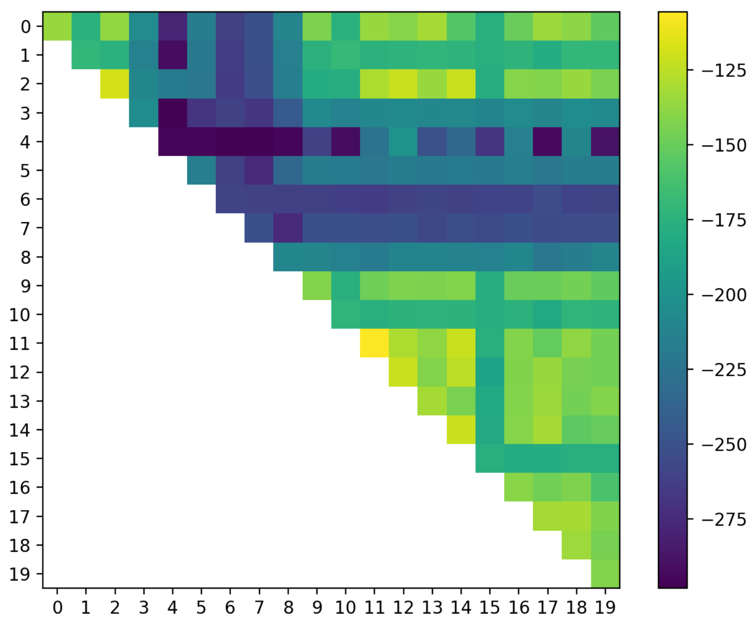

Figure 4: $-t_0$ for all baskets of size 1 and 2

As shown in Figure 4, it appears that the greedy approach will find the global minimum by first finding the minimum along the diagonal, of index 5, and then finding the minimum within the L shape containing the point (5,5). This more complete approach can be developed into a strategy, by computing all baskets or weighting them based on $-t_0$, however, due to lack of time, we weren't able to implement this.

4 Analysis and Results

The strategy we employed consisted of constructing baskets every trading day (total of 350 minutes). Trading signals were calculated and trades were executed each minute, from 2007 to present. Approximately seven fold returns were achieved.

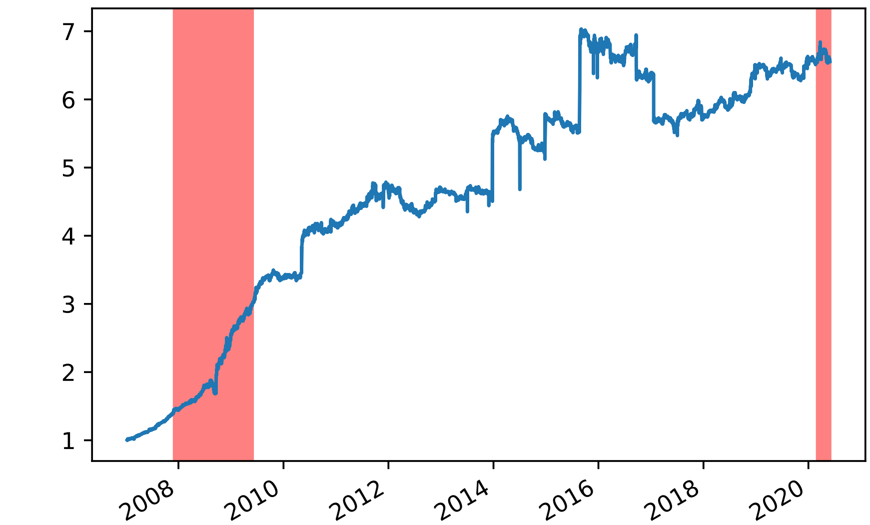

Figure 5: Cumulative relative returns

The Sharpe ratio for the whole execution is calculated to be 1.244. The relatively low value of the ratio compared to other HFT strategies comes from the high standard deviation resulting from the re-calibration of trading signals each trading day.

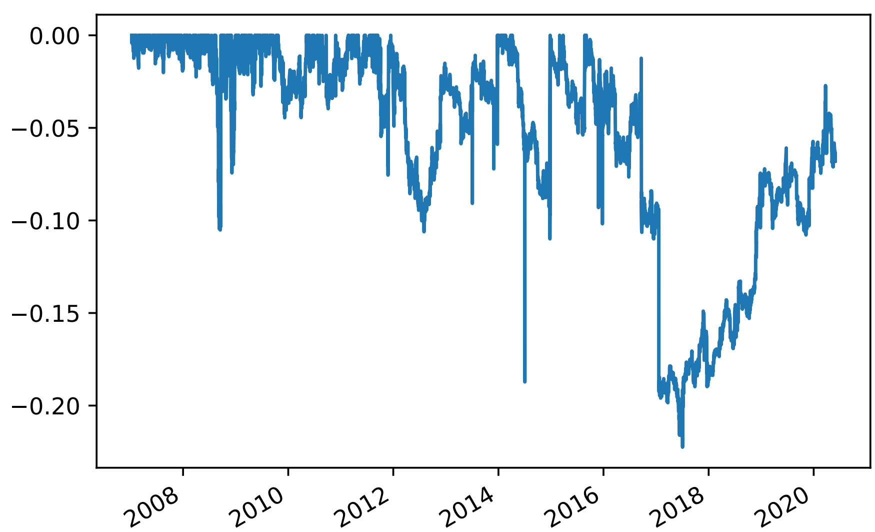

Figure 6: Drawdown

We also achieved a Max Drawdown of 20% When compared with the returns plot it is clear that the jumps in the returns plot correspond to the local optima of the drawdown plot. These jumps are due to missing values in the CRSP data. The data could be further cleaned to achieve a more continuous returns plot which would result in a decrease of the drawdown plot.

We also regress the returns against various sector indicators to understand in what environments the strategy performs best. Figure 7 displays that there are no significant effects on the returns.

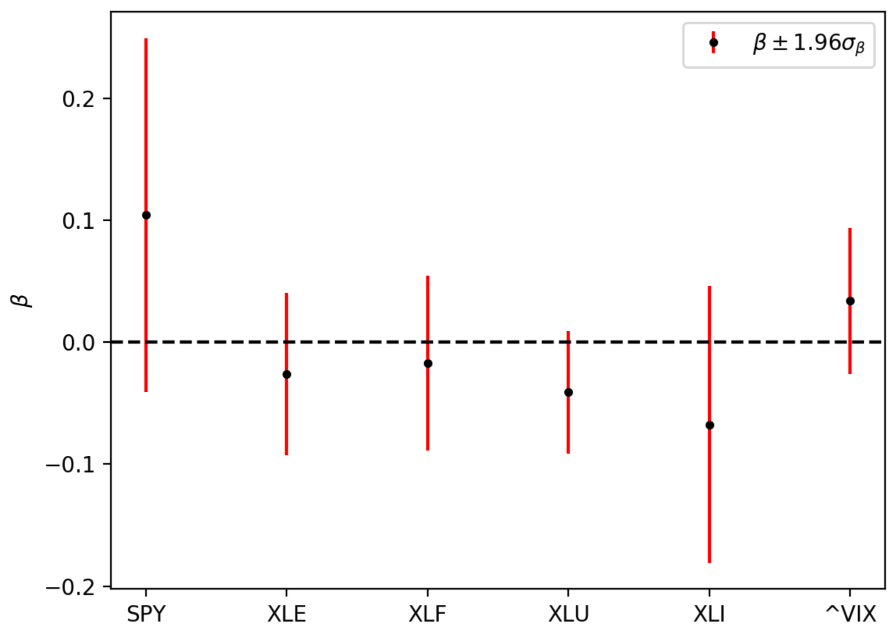

Figure 7: Regression coefficients and 95% confidence intervals

5 Conclusion

A critical insight of the workflow we developed is the necessity to construct robust and continuous data collection and storage methods, back-testing methodology, and accurate signal discovery and quantification. We will touch on these points in further detail.

While we are proud of the model we constructed in regards to its in-sample and back-tested performance, none of us is confident we have accounted for the main complexities that remove consistent alpha. Namely, transaction costs and size limitations. It is well known, especially given interviews with practitioners in the space, conversations with Dr. Lisa, and a good read of Zuckerman's "The Man Who Solved the Market" these considerations can become the entire game, in a way. On T-costs, for example, Robert Mercer, who took over

Renaissance from Jim Simons, said what the firm did best was model the cost of trades. On size, understanding how much money can be poured into a strategy before it exposes the signal to the broader market or slippage takes away the alpha we have not even attempted.

That being said, we did assemble a strong pipeline from CRSP, into a python and R basket assembly and back-testing, to out-of-sample testing and performance evaluation. Though we had one main signal , we did fully begin to realize how we could likely integrate others and check performance, too.

References

[1] George J. Miao *High Frequency and Dynamic Pairs Trading Baed on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach.* International Journal of Economics and Finance; 6(3) 96-110, 2014

[2] Xin Guo, Tze Leung Lai, Howard Shek, Samuel Po-Shing Wong *Quantitative Trading: Algorithms, Analytics, Data, Models, Optimization.* Chapman & Hall, CRC (2017)

[3] Zuckerman, Gregory. *T*he Man Who Solved the Market. New York: Portfolio/Penguin, 2019.