Statistical Arbitrage in Federal Bond Markets

By Daniel Yu and Nezam Jazayeri

Abstract

By nature, arbitrage opportunities are transient and concentrated in underlooked sectors. This research was motivated by the prevalence and history of trading models built on statistical arbitrage techniques in more traditional, liquid financial markets namely public equities, commodities, foreign exchange, credit markets, and other derivative markets compared to the perceived lack of such exploration in the Federal Bond market. Our paper examines Federal Bond markets and indices for statistical arbitrage opportunities across different timescales. We outline the theoretical framework of Bond markets and their differentiation from equity and other financial markets given their unique characteristics such as correlation to federal interest rate, maturity date, yield, etc. Finally, we apply pairs trading techniques and construct basic trading strategies to try and generate alpha from statistical arbitrage methods. We find that ultimately, while the federal bond market is generally efficient, opportunities for statistical arbitrage exist.

Glossary and Terms

Statistical Arbitrage

Arbitrage: simultaneously buying and selling of assets in different markets to take advantage of pricing differences i.e. a risk-free exploit of inefficient market pricing to generate profit after transaction costs.

Statistical Arbitrage: “probabilistic” arbitrage where one attempts to trade on the underlying which are statistical distributions of the assets instead of the assets themselves. In practice, this is often various forms of mean reversion and pairs trading.

Bonds

$$P_{bond} = NPV_{interest\ cash\ flow} + NPV_{par} = \sum_{i=1}^{N} \frac{C}{(1+r)^{i}} + \frac{F}{(1+r)^{N}}$$

$N = periods\ until\ maturity$ (annual/ semi-annual)
$C = coupon\ payment\ per\ period$ (fixed interest paid on the bond, annually or semi-annually)
$r = yield\ to\ maturity\ (YTM)\ or\ discount\ rate$
$F = face\ value\ or\ principal\ (aka\ par)$

$$Yield_{bond} = \frac{annual\ coupon\ payments}{current\ market\ price} \times 100\% = \frac{C}{P_{bond}} \times 100\%$$

Trading at a premium = the bond is trading above the face value of the bond reflecting a higher coupon rate than the market interest rate

Note the inverse relationship between the market interest rate and the price of a bond. When given a higher market interest rate, the price of the bond adjusts downward since given a bond with the same principal, new bonds with a higher coupon rate (and thus higher expected future returns) will be issued. This is finally reflected in a change in the YTM (yield to maturity) which increases to reflect the higher rate of return.

Stationarity and Cointegration

Stochastic process: a sequence of random variables which can be thought of as the "population" that a time series samples from.

Weakly Stationary: a process is weakly stationary, if its mean, variance, and covariance are unchanged by time shifts, i.e. for a process $Y_{1,} Y_2, \ldots , Y_m$.

- $E(Y_i) = \mu\ \forall\ i$, μ a constant
- $Var(Y_i) = \sigma^2\ \forall\ i$
- $Corr(Y_i, Y_j) = p(|i - j|)\ \forall\ i, j$ for some function ρ(h)

Note: when working with log-normal time series data, weak stationarity implies strong stationary, as a normal distribution is entirely defined by its first two moments..

Orders of Integration: the number of differencing operations to transform a non-stationary time series into a stationary time series.

- I(0) : already stationary time series Y_t
- I(1) : $\Delta Y_t = Y_t - Y_{t-1}$
- I(n) : $\Delta^n Y_t = \Delta(\Delta^{n-1} Y_t) = \sum_{k=0}^{n} (-1)^k (n\ choose\ k)(Y_{t-k})$

Note: in practice, one rarely goes beyond 2nd order of integration due to loss of information.

Strongly stationary: a stochastic process is considered strongly stationary when the joint distribution of any finite collection of random variables remains the same regardless of shifts in time (i.e. all moments not just mean and variance but also skewness, kurtosis, etc.).

- $\forall\ m, n,$ the distribution of $Y_1, Y_2, \ldots , Y_m$ and $Y_{1+m}, Y_{2+m}, \ldots , Y_{n+m}$ are equivalent

Cointegration: a set of assets is cointegrated if for some set of time series $(X_1, X_2, \dots, X_k)$ where each X_i is I(1), there exists a linear combination such that the new time series Y is I(0). $Y = b_1 X_1 + b_2 X_2 + \dots + b_k X_k$, where $b_1 \dots b_k \in R$.

Autoregressive (AR) Model: a regression model that predicts future values of a time series based on past values. A simple AR(1) model, i.e. considering only 1 lagged term is:
$Y_t = \phi Y_{t-1} + e_t$ where e_t is assumed to be some white noise. For higher order AR(p) model where $Y_{t-1}, \dots, Y_{t-p}$ feed back into the current value :
$Y_t = \beta_0 + \phi_1 Y_{t-1} + \phi_2 \Delta Y_{t-2} + \dots + \phi_p \Delta Y_{t-p} + e_t$. We will consider AR models in the context of stationarity and cointegration.

Given the equation AR(1) for simplicity, since any AR process can be rewritten as a MA equation: $Y_t = \phi Y_{t-1} + e_t = \phi^t Y_0 + \sum_{k=0}^{t-1} \phi^k e_{t-k}$, we can simplify the variance and expected value:

- $var(Y_t) = \sigma^2[\phi^0 + \phi^2 + \dots + \phi^{2(t-1)}]$
- $E(Y_t) = \phi E(Y_{t-1}) = \dots = \phi^t Y_0$

$|\phi| < 1$ is stationary:

- As $t \to \infty$: $E(Y_t) = 0$ a constant, satisfying the first condition of stationarity
- As $t \to \infty$: $var(Y_t) = \sigma^2[\phi^0 + \phi^2 + \dots + \phi^{2(t-1)}] = \frac{\sigma^2}{1-\phi^2}$ as it becomes an infinite geometric series. The variance is also constant, satisfying the second condition of stationarity.

$|\phi| = 1$ known as a **unit root** is non-stationary:

- As $t \to \infty$: $E(Y_t) = \phi^t Y_0 = Y_0$, a constant satisfying the first condition of stationarity
- However, as $t \to \infty$: $var(Y_t) = \sigma^2[\phi^0 + \phi^2 + \dots + \phi^{2(t-1)}] = \sigma^2 t$, failing the second condition of stationarity since the variance changes depending on time t.

$|\phi| > 1$ is non-stationary:

- As $t \to \infty$: $E(Y_t) = \infty$ as ϕ^t increases to ∞, failing stationarity first condition as the mean is not constant.

Background

Our research was inspired by MARKOWITZ ALLOCATION-FIXED INCOME SECURITIES by Tom Barnes which introduced the general differentiation of working with bonds vs stocks and explored how to extend stationary time series (crucial to many statistical techniques) to bonds [1].

The difference between bond time series and stock time series is that bonds have a fixed time horizon, their maturity, after which expected future cash flows decrease to 0. Stocks on the other hand, are supposedly valued on the net present value of cash flows to infinity. This introduces the so-called "maturity bias" whereas when a bond nears its maturity date, it converges to its face value.

In other words, as the number of periods $N \rightarrow 0$:

$P_{bond} = \sum_{i=1}^{N} \frac{C}{(1+r)^{i}} + \frac{F}{(1+r)^{N}} \rightarrow 0 + \frac{F}{(1+r)^{0}} = F$, the face value.

Looking at duration, a metric measuring the sensitivity of a bond's price to changes in interest rates, we see that empirically as a bond nears its maturity, returns become less volatile. This often means the distribution of a bond's returns from inception to maturity is not stationary. In mathematical terms, the mean and variance of the random variables can't be assumed to follow a standard distribution. This problem is solved in the federal bond market by forming expectations for each maturity date rather than for each individual bond. Note that for corporate bonds, the limited sample means it is far more difficult to curate an appropriate sample to generate expectations which is why they are generally avoided..

In our paper, this first meant examining 10-year treasury bonds with similar maturity dates across governments and bond indices whose set of bonds should, cerberus paribus, eliminate maturity bias. Secondly, we examined the 10-year treasury bonds and bond indices across varying timescales and time intervals from the past 20 years at daily price intervals to past 3 months at 15 minute intervals with the idea that the shorter the time frame, the more likely the asset is stationary.

We chose 10-year treasury bonds because out of all bonds, the treasury bond is (generally) the most liquid in secondary markets and government bonds are, in general, much less volatile than corporate and municipal bonds, so seemed the safest, low-hanging fruit to conduct statistical analysis. Government bonds also have the interesting property of reflecting the wider bond market.

Hypothesis

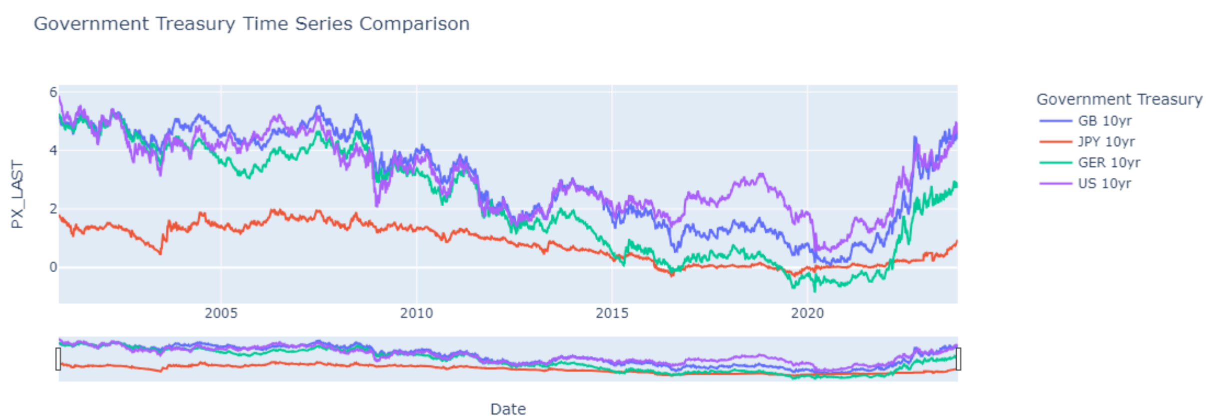
Our hypothesis is that Central Banks across the developed world tend to move in tandem, following the US Federal Reserve's lead, especially demonstrated during the COVID-19 pandemic. The interest rate environment and treasury yields of different countries should move in parallel with macroeconomic trends [3]. Thus, there should exist arbitrage opportunities between different government treasuries. For example, for pair trading two government treasury assets, we can expect an optimal combination of treasury bond 1 and treasury bond 2 such that if the prices diverge, we can short one and long the other under the expectation that the pair should mean-revert.

Data Collection

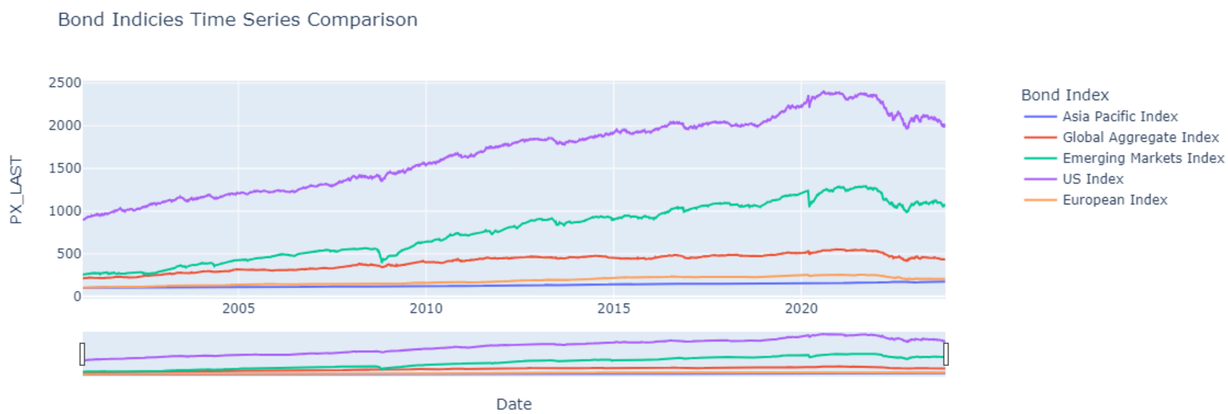
We used the Bloomberg terminal to retrieve our data. The excel and csv files can be found under the data folder in `ThGreatExplorer/DRP_Statistical_Arbitrage repository` [2].

We arbitrarily selected 10yr Treasury Bonds for the United States (US), Great Britain (GB), Japan (JPY), Germany (GER) from the G7 which further included France, Italy, and Canada. Bloomberg already adjusts for maturity bias, correcting one of the major potential issues.

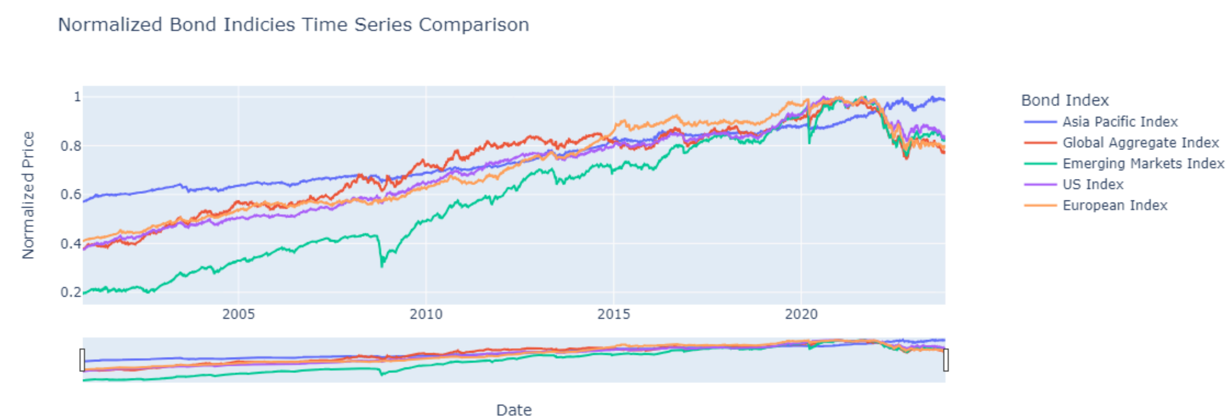
We started looking at over a 20 year timeframe at daily intervals. A cursory analysis reveals apparent strong correlation, suggestive of potential cointegration.



Then we selected the top bond indices on Bloomberg for the Global Aggregate, US Aggregate, Asia-Pacific Aggregate, Pan-European Aggregate, and Emerging Markets Aggregate again over a 20 year timeframe.

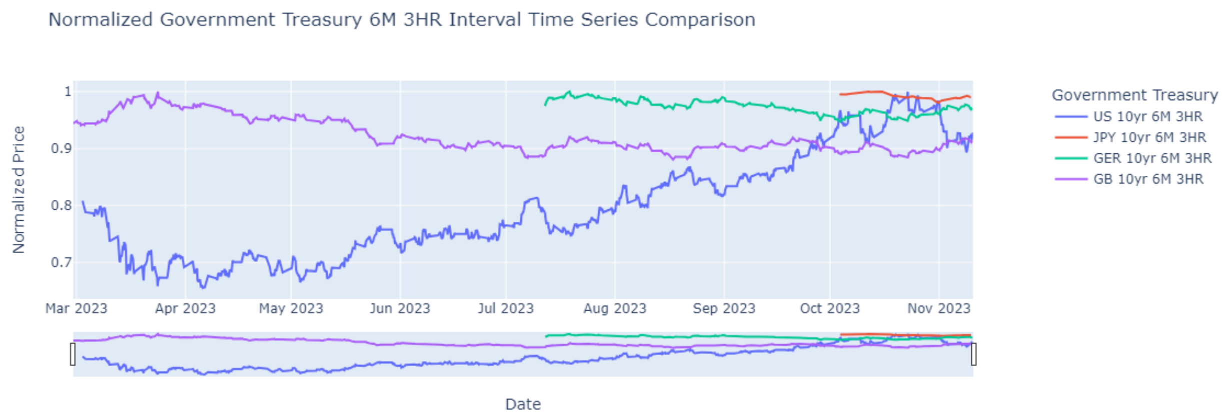


After normalizing the data, we see a much stronger correlation. Henceforth, the graphics will be normalized for bond indices due to the pricing differences.

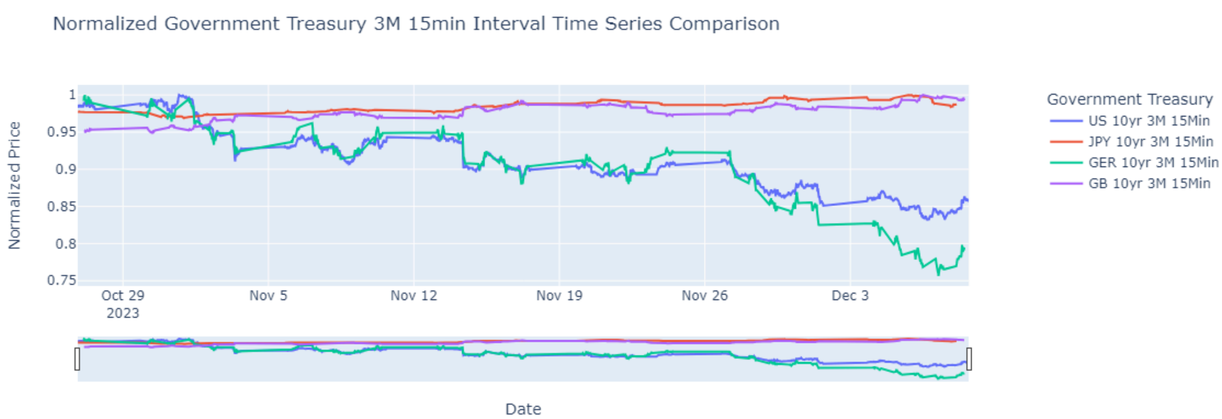


We then collected data from 20 year daily intervals to 5 year daily intervals to 3 year daily intervals to 2 year daily intervals and finally to 1 year daily intervals for Government Treasuries, essentially subsections of the 20 year data. This was for ease of processing shorter timeframes.

Finally, we took higher frequency trading data, collecting data at 6 months at 3 hour intervals and 3 months at 15 minute intervals. Note: Bloomberg had incomplete data for Japan and Germany. Japan had only about 1 month's worth of data and Germany had about 3 Months worth.



Notice the distinct correlation between US-JPY and GER-GB.



The data cutoff for all of our data was late november to early december 2023 depending on when the data was pulled. We then performed statistical tests on the data for the appropriate trading strategies.

Methodology

We used the following methodology for our two hypothesis test:

1. Dicky Fuller (DF) Test: a hypothesis test which takes the time series Y_t and fits a AR(1) model, $Y_t = \mu + \phi Y_{t-1} + e_t$ where ϕ represents the coefficient of the lagged time variable, μ a constant, and e_t represents the noise. Remember for an AR(1) model, where the null hypothesis H_0: $\phi = 1$ (unit root) represents non-stationarity, since over time, $Y_t = Y_{t-1} + e_t$ (when $e_t = -1$ or 1, this is a random walk) grows without bound

shifting mean, variance, and autocovariance. The Alternate hypothesis is $H_1: \beta < 1$, if $\beta = 0$ then strongly stationary, otherwise weakly stationary. Here we assume e_t is not so large that Y_t experiences significant variability. Manipulating the expression, we derive the mathematical formula:

$Y_t = \mu + \phi Y_{t-1} + e_t \rightarrow Y_t - Y_{t-1} = \mu + (\phi - 1)Y_{t-1} + e_t \rightarrow \Delta Y_t = \mu + \delta Y_{t-1} + e_t$

a. Null Hypothesis becomes $\delta = 0$: This transformation allows us to assume $\Delta Y_t = \mu + e_t$, i.e. I(1), which would be stationary (but Y_t would still be non-stationary under null hypothesis).
b. Alternative Hypothesis becomes $\delta < 0$: reject non-stationarity of Y_t. We perform a t-test against a special Dicky Fuller distribution and if the t-statistic < $DF_{critical\ value}$, we reject the null hypothesis.

2. Augmented Dicky Fuller (ADF) Test: expansion of the Dicky Fuller test which is an AR(1) test, to train the time series to higher order AR(p) model where $Y_{t-1}, \dots, Y_{t-p}$ feed back into the current value : $Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + e_t \rightarrow \Delta Y_{t-i} = \mu + \delta Y_{t-i} + \sum_{i=1}^{p} \beta_i \Delta Y_{t-i} + e_t$. Then the same hypothesis test above ($H_0: \delta = 0$ and $H_1: \delta < 1$) is applied against a Dicky Fuller Distribution with an additional step of t-statistics of each regression term B_i to check if they are each statistically significant.

3. Engle-Granger Two-step Cointegration Test: Remember that two series X_t and Y_t cointegrate if there exists a linear combination of them that is I(0).
   a. The first of the Engle-Granger Test is to estimate a linear relationship between X_t, Y_t using Ordinary Least Squares (OLS) regression: $X_t = \mu + \beta Y + e_t$ where e_t represents the “residuals” (errors) from the regression.
   b. The second step is to perform ADF tests on the residuals: e_t where the null hypothesis H_0 is the residuals are not stationary and the alternate hypothesis is the residuals are stationary meaning the two time series have a long-run equilibrium relationship where the difference between the two reverts to e_t. A more detailed treatment of this test can be found in the **References** section [3].

Pairs Trading Strategy

Pairs Trading is the simplest statistical arbitrage tool, where two assets are cointegrated and trades are placed betting that the cointegration relationships will hold in the future. Generally, a linear relationship between two assets is predicted. Then bets are placed that whenever the two asset prices diverge from that relationship, they will converge in the future.

We constructed a basic algorithm which takes in data containing two time series for asset1 and asset2, assumed to be O(1) stationary. The default parameters are a rolling period of 30 days, a standard deviation threshold of 1, and a portfolio size of 1 million. The algorithm attempts to find cointegration in some training set, up until $20 \cdot rolling\ period$ or the end of the data. If cointegration is found, the dataset is split into train and test sets where the linear relationship from the cointegrated training set is used as the initial beta for the test set. If no cointegration is found, an error is thrown.

Once the dataset is split and cointegration is found, the portfolio is allocated based on the initial beta where if $px_{asset1} = 2 \cdot px_{asset2}$, then the allocation would be two-thirds of the portfolio for asset1 and one-third for asset2. The main method is the backtest method where at each timestep, a signal is generate based on if the cointegration relationship holds and the beta is within $\pm\ threshold\ \cdot standard\ deviation$ of the mean of the historical beta over the rolling period. The signal then is passed to a helper method which updates the portfolio position based on the signal. If there is no cointegration relationship or the beta is within $\pm\ threshold\ \cdot standard\ deviation$, then we exit from the positions and liquidate our portfolio to cash. The portfolio and trading_signals_history fields track the performance of our strategy over time.

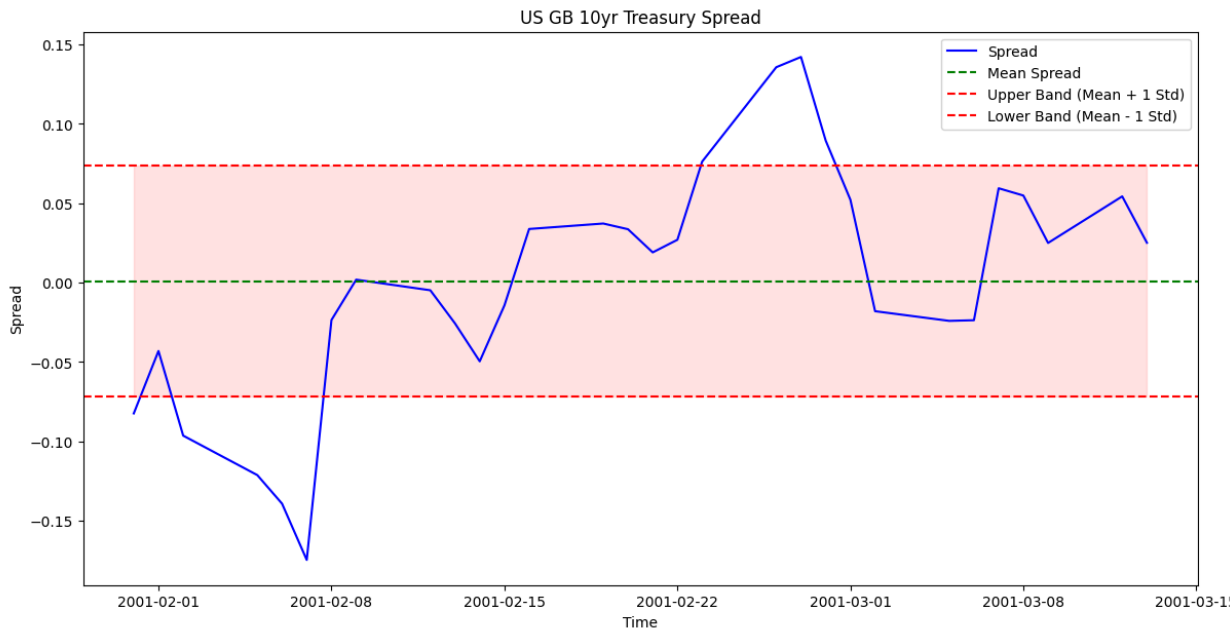
Finally, there are various plotting methods supported for plotting the portfolio value, the time series, the trading signals, and the position sizes.

Example

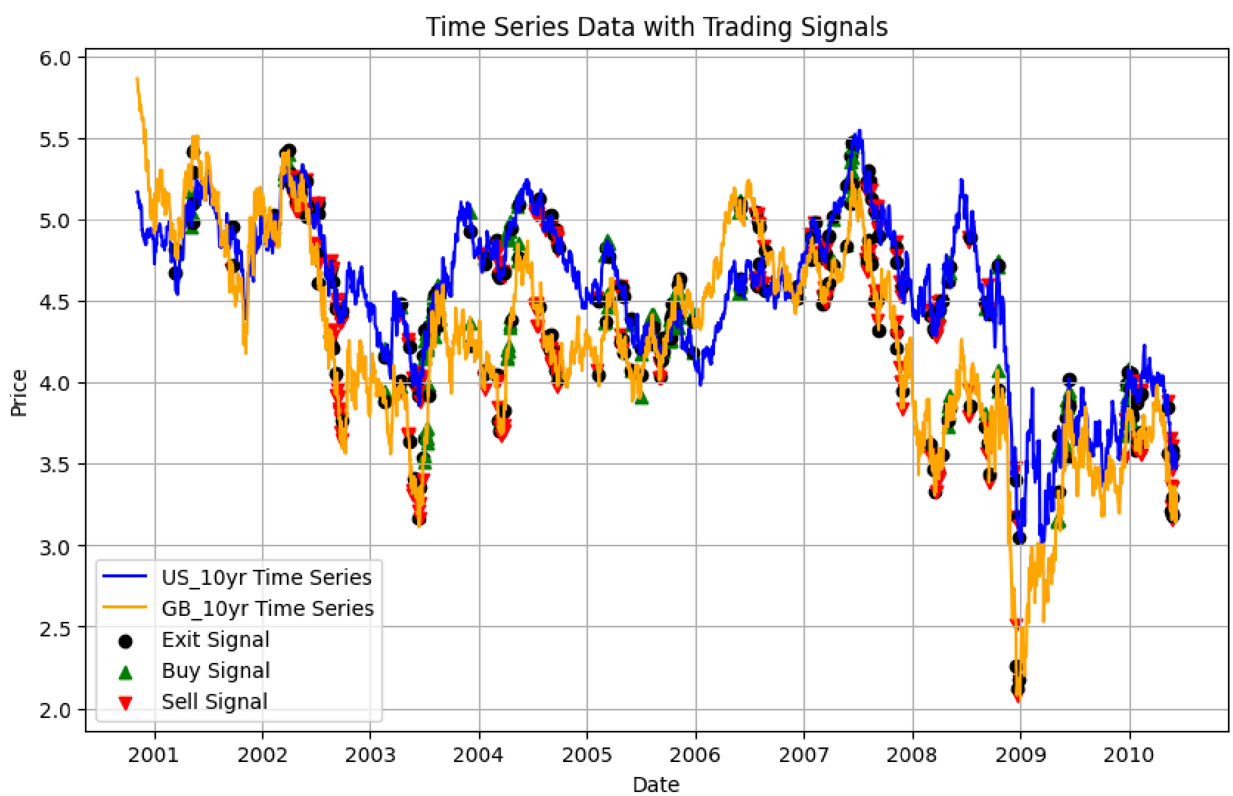
Let's demonstrate with an example with the US and GB 10-year Treasury Bonds with default parameters, discovered to be cointegrated.

The algorithm finds that the assets are cointegrated after $3 \cdot rolling\ period\ = 90$ days. The dataset from 0 to 90 is now the training set and everything after becomes the test set.

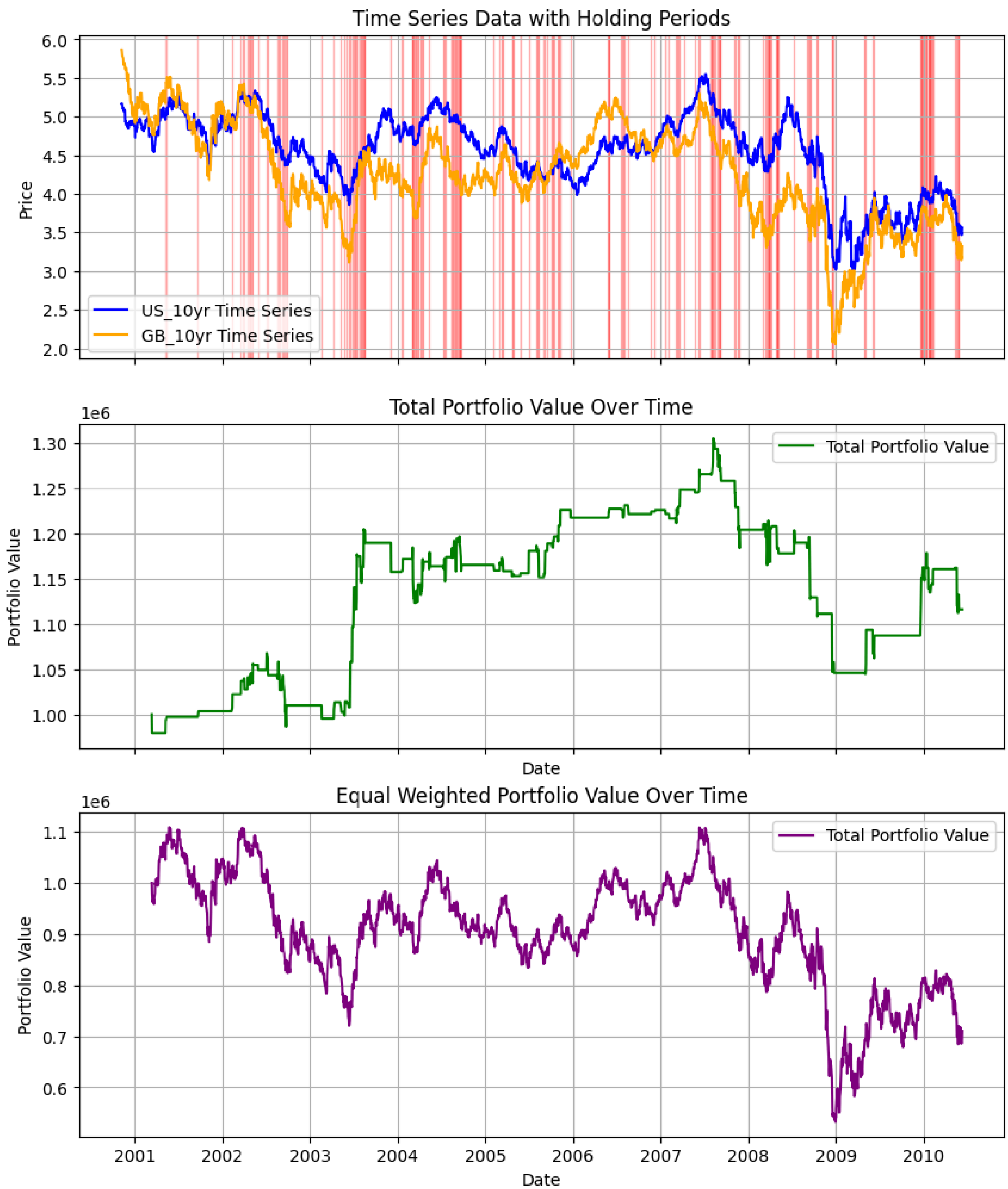
To generate a trading signal, we take the linear relationship between the two assets and trade off whenever the spread exceeds 1 standard deviation to long one asset and short the other. In this example, a rolling period of 30 days was used for calculating the mean, beta, and spread. Theoretically, we can bet that 68% of the time, at any timestep, the spread would fall between $\pm\ 1$ standard deviation and only a 32% chance to fall outside that range. On average, these trades should generate a profit over time.



Performing the backtest over about 10 years, this operation is repeatedly at every timestep and the generated signals are used to inform the trading strategy. The plot below plots the signals over the time series.

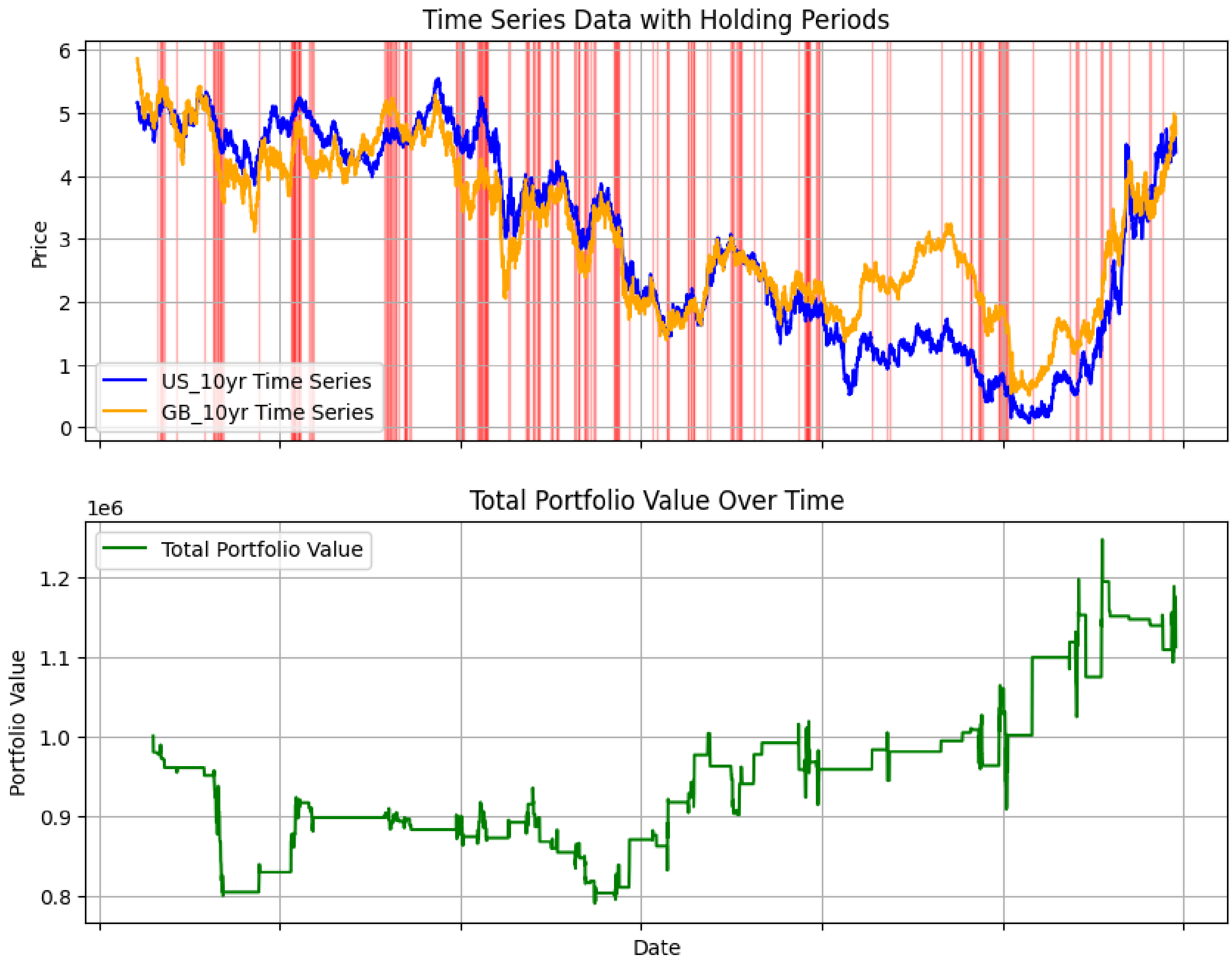


The holding period of a trade is the period when we buy (i.e. long/short assetA/assetB) and then either sell/exit once we rebound within the standard deviation or the cointegrating relationship fails to hold. As the pair remain outside the band, we continue to allocate to, essentially “buying” more into the trade. The holding periods are generally extremely brief, rarely more than a few days.



With the given parameters, the pair trading strategy has a sharpe ratio of 0.042 against a 5% risk free rate, so performs very poorly. The equal weighted portfolio has a sharpe ratio of -0.018. There is a correlation of about 33% with the equal weighted portfolio.

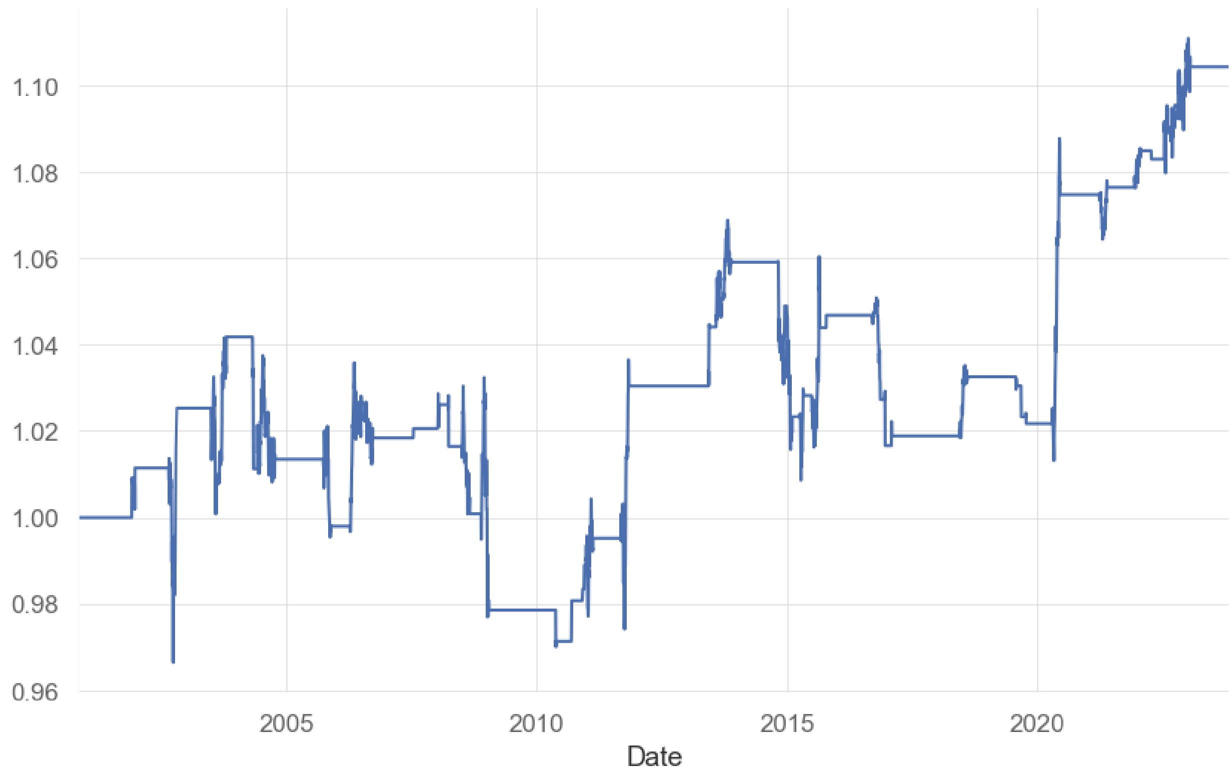
Adjusting the hyperparameters gives varying performance. When working with a rolling period of 90 days and a threshold of 1, we can see the holding periods tend to be thicker but less frequent, expected behavior since given more data points, we would expect that a pair would more easily approach cointegration.

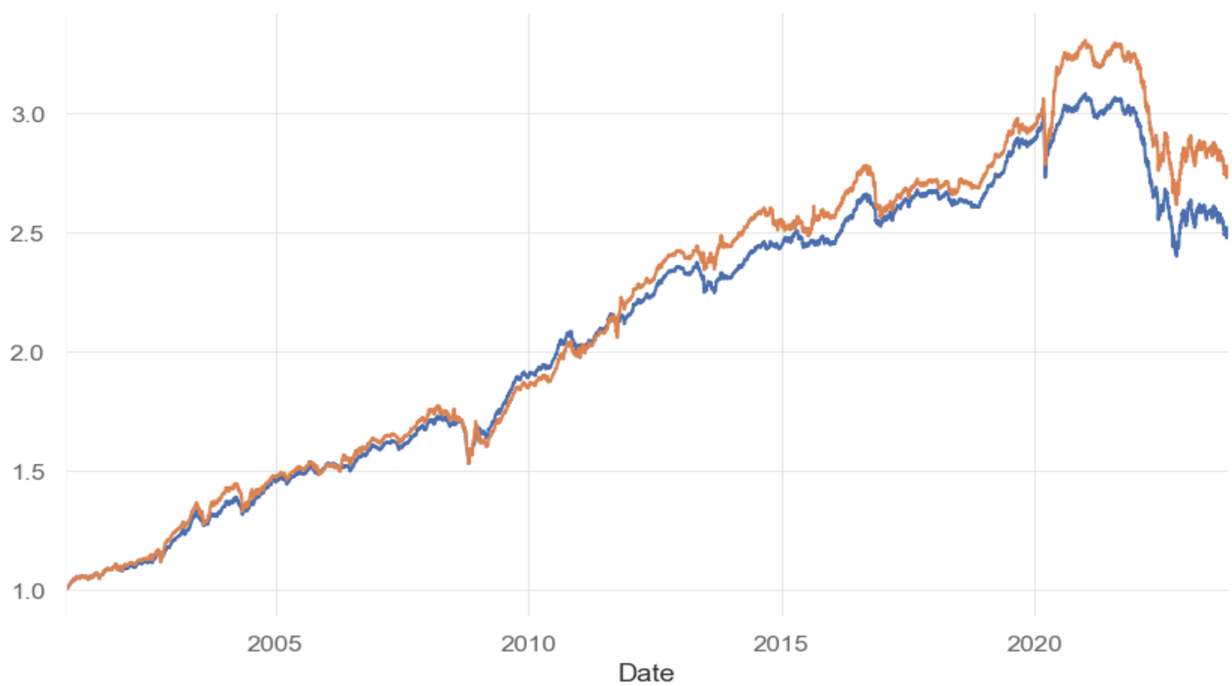


The strategy now has a sharpe ratio of 0.1 and is -43% correlated with the equal weight portfolio.

Volatility Trading

Take monthly annual volatility of an equal weighted portfolio, and for each date, make note of whether in the past year, whether we are > 80th percentile. If so, we are in “high volatility environment.” We predict meaningful statistical arbitrage will occur between the assets in these environments and trade on them accordingly. After backtesting the strategy on `'us_agg',` `'global_agg', 'em_agg', 'euro_agg'` bond indices over the past 20 years, we find that the strategy under performs against equal weighted portfolio.





Next Steps

So far we have only explored basic statistical arbitrage techniques on Treasury Bonds. Further steps that could be taken to improve the performance of the algorithms is to:

Pair Trading:

1. Implement stop signals where after the standard deviation exceeds a certain amount or that after a certain number of timesteps of staying above the standard deviation, to exit the trade and liquidate. The idea is that at that point the cointegration relationship may no longer be valid.
2. Implement Kalman Filter as a variant of the pair trading strategy to help with the rolling periods and improve on simply using a moving average.
3. Implement a basic Machine Learning algorithm to help predict the beta values given past data.
4. Experiment with position sizing on trades perhaps based on the strength of the signal.
5. Experiment with different hyperparameters such as the rolling period and threshold.

Volatility Trading:

1. Use GARCH modeling to better predict future volatility so that it's not just based on past volatility.

Additional Features:

1. Implement portfolio allocation/ management for managing and placing multiple pair trades at a given time.
2. Experiment with testing algorithm on higher frequency data which has more variance and less probability of cointegrating relationship.
3. Try coding different algorithms!

References

[1] Markowitz Allocation Fixed Income Securities
https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6803.1985.tb00401.x

[2] Repository
https://github.com/ThGreatExplorer/DRP_Statistical_Arbitrage

[3] Engle-Granger Cointegration Test:
https://warwick.ac.uk/fac/soc/economics/staff/gboero/personal/hand2_cointeg.pdf

[4] US Monetary Policy on Global Interest Rates:
https://www.imf.org/en/Publications/WP/Issues/2016/12/31/U-S-44315

[5] Quantopia Lecture Series
https://github.com/quantopian/research_public/blob/master/notebooks/lectures/Kalman_Filters/notebook.ipynb

[6] Trading Strategies Repository
https://github.com/je-suis-tm/quant-trading

[7] Arbitrage Strategies Repository
https://github.com/JerBouma/AlgorithmicTrading

Bloomberg Index Codes:

US 20yr Treasury Bonds

Germany 20yr Treasury Bonds

Japan 20yr Treasury Bonds

Great Britian 20yr Treasury Bonds

LEGATRUU – Global Aggregate

LBUSTRUU – US Aggregate

LAPCTRJU – Asian Pacific Aggregate

LP06TREU – Pan Euro Aggregate

EMUSTRUU – EM USD Aggregate