Pairs trading strategy

Design & Backtest

Santiago Garcia

1 Contents

List of tables.

List of figures.

2 Data and pairs selection.

In practice, it is very difficult to find tradeable assets that present a mean-reverting behavior given that assets like equities, currencies and commodities behave like GBMs, a process that we demonstrate in **Section 2.1**. Nevertheless, once we find that a pair of assets comply with the set of rules to implement the arbitrage trading, the opportunities of generating profits increase significantly.

For the purpose of analyzing the short and long memory properties of the assets, we take data from July 2019 to July 2023, which is a period characterized by an unprecedented drop in the global markets due to the COVID-19 pandemic (around 2020'Q2), and then one of the fastest recoveries in the aftermath of the contagion, where the S&P500 reached its all-time high around later 2021 and earlier 2022 thanks to an injection of economic stimulus (and COVID vaccines) by most of the important governments, that later will result in an spike of inflation reaching year-over-year numbers above 7% in countries like US, UK and most of Europe. Consequently, the central banks have been forced to combat the increase of cost prices of basic goods and services, among others, by raising their reference rate, something that historically have worried the securities' market due to the impact in the economy and potential recession that higher interest rates might cause. Therefore, it will be interesting to study the performance of our strategies during distressed and greedy periods.

Another consideration is that we will concentrate on stochastic non-stationary assets (unit root case), therefore we selected two futures contracts that possibly will present a cointegrating nature. In the case of Engle-Granger procedure (further discussed in **Section 3**), and unlike Halls-Moore (2015) whom study is directed to same sector equities that are likely to be exposed to similar market factors, our selection is based on global events that could open the window to arbitrage trading.

It is worth noting that the Engle-Granger procedure works best to estimate a cointegrating relationship and error correction between a pair of time series. Working with n > 2 variables is out of the scope of this strategy, although further adaptions can be done using the Johansen MLE procedure in a multivariate setting.

We make use of Yahoo Finance as data source, as it provides historical data for continuous futures contracts, including their Opening, Closing, Highest and Lowest price of a given period.

Important: For code ONE we store the historical data downloaded from the Yahoo Finance API in a data frame, therefore there is no need for external files. On the other hand, code TWO requires the historical data for both Platinum (PL=F) and Copper (HG=F) to be stored as csv files in the folder, as well as the *spx_data.csv* file.

2.1 Confirming the integrated process with Dickey-Fuller Test for unit root.

In this section we describe the process of confirming the integrated process, also known as unit root. We use the term "confirming" because the assumption is that the time series of the assets that we analyze in this report are known to have an integrated process of order one $Y_t \sim I(1)$

and therefore they are non-stationary. However, it will be helpful to define the numerical workings of this process for further investigation when confirming the stationarity of time series.

Therefore, we start with the following time series equation:

$$y_t = \epsilon_t + \beta y_{t-1} \tag{1}$$

Decomposing the above equation renders:

$$y_t = \epsilon_t + \beta(\epsilon_{t-1} + \beta y_{t-2})$$

$$y_t = \epsilon_t + \beta\epsilon_{t-1} + \beta^2(\epsilon_{t-2} + \beta y_{t-3})$$

$$y_t = \epsilon_t + \beta\epsilon_{t-1} + \beta^2\epsilon_{t-2} + \cdots \tag{1.1}$$

By following the decomposition method, we are left with just the residuals as shown in (1.1). Furthermore, we can note that the impact of past residuals fades as $i \sim \infty$:

$$y_t = \sum_{i=0}^{p} \beta^i \, \epsilon_{t-j} \tag{1.1.1}$$

The null hypothesis is that time series has a unit root. Therefore:

$$H_0: Y_t = Y_{t-1} + \epsilon_t \quad implies \quad \beta = 1, \Delta Y_t = \epsilon_t$$

$$H_1: Y_t = \beta Y_{t-1} + \epsilon_t \quad substract \quad Y_{t-1}$$

We test for significance of $ø = \beta - 1$ by comparing the t-statistic to a critical value taken from the Dickey-Fuller distribution. Hence:

$$\Delta Y_t = \emptyset Y_{t-1} + \epsilon_t$$

Failing to reject the Null Hypothesis (H_0) confirms that the time series has a unit root given:

$$\emptyset = 1 - \beta = 0 \;\rightarrow\; \beta = 1 \;\therefore\; \Delta Y_t = \epsilon_t$$

3 Engle-Granger process.

As shown in the previous section, we can examinate whether a time series is non-stationary by running a simple test. In practice, we can find that several financial time series are *I(1)*, but the question relies in whether these financial variables are cointegrated so that one or more linear combinations of them becomes a stationary *I(0)* processes. For this reason, we carry on this investigation using the Platinum and Copper futures contracts.

First, it is very common to find asset prices that range from one to five digits, rendering huge residual values when running the cointegration process. For instance, Platinum Futures trade in the COMEX (the reference that Yahoo Finance uses) for a value between U.S$600 and U.S$1,200 as far as our data concerns; on the other hand, Copper Futures trade at the same exchange for a price between U.S$2 and $5. Therefore, it is a “stylized fact” that working with standardized futures help to better linearize the features and hence produce a better model fit for the linear regression, in other words, to make the residuals look more normal, as well as the hedge ratio. For this purpose, we make use of the function *StandardScaler* from the *sklearn.preprocessing* library.

The spread time series is tested for stationarity by the Cointegrated Augmented Dickey-Fuller (CADF) test in **Section 3.1**. In other words, if the pair of assets are cointegrated, it suggests that the mean and variance of this correlation remains constant over time. However, there is a major issue which makes the strategy difficult to implement and it is encountered in the long-term relationship, as it can break down, rendering the spread to move further from one equilibrium to another. Hence, in **Section 3.2** we perform the second step of Engle-Granger to estimate the Equilibrium Correction Model, or error correction equations.

In **Section 3.3** we assess the quality of the mean reversion once we define that the pair of assets are suitable for trading. Arbitrage is realized by using cointegrating coefficients β_{Coint} as allocations ω. In this case, we make use of the Ornstein-Uhlenbeck (OU) process for signal generation and to assess the quality of mean reversion.

Finally, we implement and backtest the strategy in **Section 4.1** and **Section 4.2** respectively. The strategy triggers new orders when the spread e_t deviates from the mean μ. For this purpose, we must pre-define the bounds to enter the trades as $\mu_e \pm Z_{opt}\sigma_{eq}$ where wider bounds might give the highest P&L and low trade signals, while narrower bounds produce signals more frequently but tighter P&L. Therefore, we will test if varying Z would yield better strategy performance in **Section 4.3**.

3.1 Step 1. Cointegrated residual.

The pairs trade works by using a linear model for a relationship between the asset prices:

$$y_t = \beta x_t + e_t \tag{2}$$

If y_t and x_t are levels, then:

$$\Delta y = \beta_g \Delta x \tag{2.1}$$

The ECM equation (which we will discuss in detail in the following sections) addresses both the short-run correlation-like $\beta_1 \Delta x$ and correction to equilibrium over the long-run. In fact, if there is a common factor, it must affect the changes in y_t, which is the endogenous variable.

We define the common factor as:

$$\hat{e}_t = y - \hat{a} - \hat{b}x \tag{2.1.1}$$

Therefore:

$\Delta y \approx \Delta x \quad ; \quad \Delta y \approx (y - \hat{b}x)$

We essentially use a linear model for a relationship between Platinum (PL) and Copper (HG) as described in (2), where PL is the dependent variable y_t. Then, reordering (2) we can obtain the naïve cointegrated residual:

$$e_t = y_t - \beta x_t - \epsilon_t \tag{2.1.2}$$

The linear regression is done using the *sklearn's LinearRegression* function. Then, we extract the linear model's coefficient β and the intercept ϵ_t, and replace in (2.1.2):

$\hat{e}_t = y_t - (0.6708)x_t - (0)$

Hence, $\beta'_{Coint} = [1 - 0.6708]$ and the equilibrium level is $\mathbb{E}[\hat{e_t}] = \hat{a} = \mu_e \approx 0$

We then run the CADF test to confirm whether the series of the equation above (including the values that we extracted from the linear model) is stationary. The results should comply with:

a. $e_t \sim I(0)$
b. $e_t \rightarrow$ AR(1)

Therefore, using the *statmodels.tsa.stattools* library and *adfuller* function we obtain:

p-value	T-statistic	ADF Critical Values
0.0118	-3.3747	{'1%': -3.4370, '5%': -2.8644, '10%': -2.5683}

Table 1. Residuals CADF Test results

From the table above we can see that the calculated test statistic of -3.3747 is smaller than the 5% critical value of -2.8644, which indicates that we can reject the null hypothesis that there is not a cointegrating relationship at the 5% level; therefore, points *a* and *b* above are true.

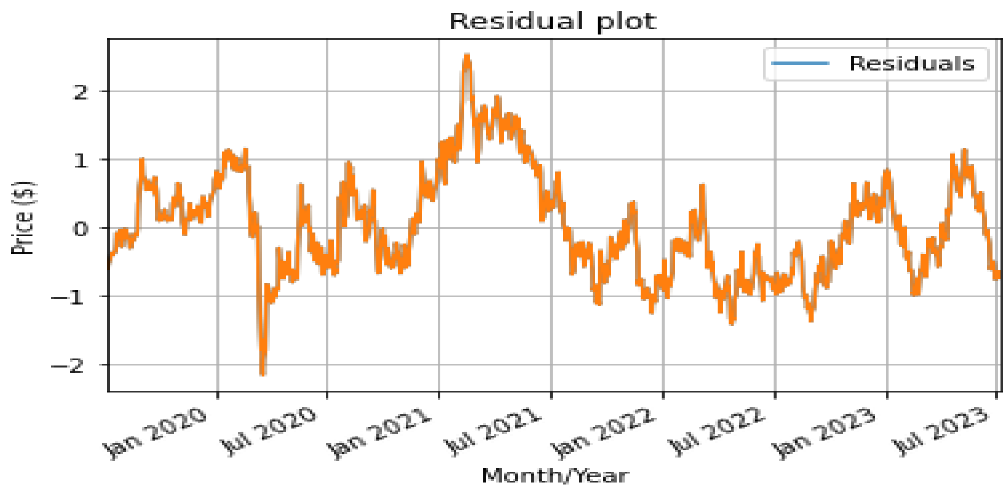


Figure 1. Platinum - Copper residual plot

3.2 Step 2. Error correction equation.

When we plot the residuals between *y* (platinum) and *x* (copper), $\hat{e}_t$ will look like a stationary process such that [1, b] is a cointegrating vector (see **Figure 1**). Moreover, the term is significant if cointegration exists between the time series (defined by the CADF test), but its expectation of the equilibrium-correction term is equal to zero (Diamond R. V., 2014): $\mathbb{E}[y_{t-1} - a_e - b_e x_{t-1}] = \mathbb{E}[e_{t-1}] = 0$ → given the ECM equation.

For a robust regression model of two non-stationary series like *y* and *x*, we utilize the growth rate defined as Δy_t. Recall from **Section 3.1** that we derived the naïve cointegrated residual from a simple linear regression equation (equations 2 and 2.1), but for this case we require a dynamic regression model with several lagged values as:

$$y_t = \alpha y_{t-1} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \epsilon_t$$

Re-specifying the model to equilibrium-correction form:

$$y_t - y_{t-1} = \alpha y_{t-1} - y_{t-1} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} - \beta_1 x_{t-1} + \beta_1 x_{t-1} + \epsilon_t$$

$$\Delta y_t = -(1-\alpha) y_{t-1} + \beta_0 + \beta_1 \Delta x_t + (\beta_1 + \beta_2) x_{t-1} + \epsilon_t$$

$$= \beta_1 \Delta x_t - (1-\alpha)\left(y_{t-1} - \frac{\beta_0}{1-\alpha} - \frac{\beta_1 + \beta_2}{1-\alpha} x_{t-1}\right) + \epsilon_t$$

$$= \beta_1 \Delta x_t - (1-\alpha)(y_{t-1} - a_e - b_e x_{t-1}) + \epsilon_t$$

Rearranging:

$$\Delta y_t = \beta_1 \Delta x_t - (1-\alpha)\hat{e}_{t-1} + \epsilon_t \tag{2.2}$$

Where the disequilibrium $e_{t-1} \neq a$ is corrected at the speed of $-(1-\alpha)$ in the long run. The speed of correction is inevitably small but not zero; in fact, it must be significant for cointegration to exist and ensure correction in the long run equilibrium.

The second step of Engle-Granger procedure consists of estimating the Equilibrium Correction Model (ECM). As such, we check the significance of $-(1-\alpha)$ and confirm that the p-value is less than 5%, thus we will reject the null hypothesis that a regression coefficient is zero.

Note that the *sklearn.linear_model.LinearRegression* fails to provide the significance of $(1-\alpha)$ when running the linear regression for the ECM. Therefore, we employ the *statsmodels.api.OLS* function to run the linear regression and obtain β_1, as well as $(1-\alpha)$ and its significance. Using either LR or OLS should yield the same coefficients.

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   PL=F   R-squared (uncentered):                   0.204
Model:                            OLS   Adj. R-squared (uncentered):              0.203
Method:                 Least Squares   F-statistic:                              127.5
Date:                Tue, 01 Aug 2023   Prob (F-statistic):                    5.21e-50
Time:                        17:21:48   Log-Likelihood:                          400.41
No. Observations:                 995   AIC:                                     -796.8
Df Residuals:                     993   BIC:                                     -787.0
Df Model:                           2
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
HG=F               1.0308      0.066     15.682      0.000       0.902       1.160
(Lag 1, e_hat)    -0.0267      0.007     -3.849      0.000      -0.040      -0.013
==============================================================================
```

Figure 2. E-G Step 2. Error correction equation

As shown in **Figure 2,** the coefficient of e_{t-1} is significant given that the p-value is less than 0.05. An influential and undesired drawback of the Ordinary Least Square is that it is asymmetric, meaning that by switching the dependent and independent variables it will result in a different hedging ratio. A common solution is to do the same process switching *y* and *x*; therefore, our endogenous variable will be *Copper*.

Appendix 5.1 shows that after running the EG process in the other way around, we encounter that the cointegrated residual is non-stationary, suggesting that we should continue the process with Copper being the exogenous variable and Platinum the endogenous one.

3.3 Assessing the quality of reversion.

In this section we perform the Ornstein-Uhlenbeck process to assess the quality of reversion, also known as the Engle-Granger "step 3". We will see through the report that the optimal parameters really depend on the aim of each strategy in terms of P&L and the frequency of trades that we can obtain in a determined time frame. Therefore, OU acts as a baseline to build our strategies:

$$de_t = -\theta(e_t - \mu_e)dt + \sigma_{OU}dW_t \tag{2.3}$$

Equation (2.3) is the Ornstein-Uhlenbeck SDE[1], where θ is the speed of mean-reversion and μ_e is the equilibrium level. The speed of diffusion under the OU scheme is defined by σ_{OU}.

The continuous-time solution to the OU SDE has a mean-reverting, autoregressive, residual terms respectively:

$$de_{t+\tau} = -\theta(e_t - \mu_e)dt + \sigma_{OU}dW_t \tag{2.3.1}$$

The regression-like solution in continuos time $t + \tau$

$$de_{t+\tau} = (\mathbf{I} - e^{-\theta\tau})\mu_e + e^{-\theta\tau}e_t \ + \epsilon_{t,\tau} \tag{2.3.2}$$

And the Vector Autoregression equivalent for a small time period τ:

$$e_{t+\tau} = \ \beta_0 + \beta_1 e_t + \epsilon_t \tag{2.3.3}$$

The above is estimated using the *statsmodels.tsa.api* function VAR(p) with p=1. The results of the OU process are given at the end of this chaper.

The coefficients β_0 and β_1 in equation (2.3.3) are obtained from the same VAR(p) process. Then, we can solve for the square matrix θ and vector μ_e as follow:

$$e^{-\theta\tau} \ = \ \beta_1 \quad \therefore \ \theta = -\frac{ln\beta_1}{\tau}$$

Where τ represents our data frequency (daily) and it equals 1/252. Moreover, we define the halflife as $\tilde{\tau}$ and it is calculated as follow:

$$\tilde{\tau} \ \propto \ \frac{ln(2)}{\theta}$$

The half-life offers an indication of the holding time required to generate the PnL. This is the time between the equilibrium situation of the spread e_t being at μ_e .

[1] We refer to Diamon (2014) to obtain the full Ornstein-Uhlenbeck process.

Then, μ_e is obtained from the following equation:

$$(1 - e^{-\theta\tau})\mu_e = \beta_0 \therefore \mu_e = \frac{\beta_0}{1-\beta_1} \neq 0$$

At this point, it just left to calculate the diffusion part of equation (2.3). From Diamon (2014), the σ_{OU} is parameter of the SDE over each small dt denoted as:

$$\sigma_{OU} = \sqrt{\frac{2\theta SSE}{1 - e^{-2\theta\tau}}} \tag{2.3.4}$$

Another consideration regarding the PnL, is that it is proportional to the standard deviation σ_{eq} of equilibrium $e_{t,\tau} \rightarrow \infty \sim N(\mu_e, {\sigma_{eq}}^2)$. Therefore:

$$\sigma_{eq} \approx \frac{\sigma_{OU}}{\sqrt{2\theta}} = \sigma_{OU}\sqrt{\frac{\tau}{2}}$$

$$\sigma_{eq} = \sqrt{\frac{Var[\epsilon_{t,\tau}]}{1 - e^{-2\theta\tau}}} \tag{2.3.5}$$

Where $Var[\epsilon_{t,\tau}] = \frac{1}{T} \times SSE \times 252$ and it is the annualised covariance-multivariate also denoted as $\boldsymbol{\sum_{\tau}}$. Then, replacing into equation (2.3.5) we obtain:

$$\sigma_{eq} = \sqrt{\frac{SSE \times \tau}{1 - e^{-2\theta\tau}}} \tag{2.3.6}$$

```
Results for equation e_hat
==========================================================================================
                   coefficient       std. error           t-stat              prob
------------------------------------------------------------------------------------------
const                -0.000095         0.005219           -0.018             0.985
L1.e_hat              0.972766         0.031772           30.617             0.000
L1.e_hat_t-1          0.002616         0.031774            0.082             0.934
==========================================================================================
```

Figure 3. E-G Step 3. Vector Autoregression (VAR)

Table 2. OU process results

Parameter	**Result**
$\boldsymbol{\beta_1}$	0.9727
$\boldsymbol{\theta}$	6.9582
Half-life $\tilde{\boldsymbol{\tau}}$	0.0996
Working days	25
$\boldsymbol{\beta_0}$	-9.54514e-05
μ	-0.00350483
$\boldsymbol{\sigma_{eq}}$	1.4078

From the table above, $\tilde{\tau}$ gives us an indication of holding time (fraction of a year). While θ indicates that the speed of mean-reversion is not fast given that instruments like futures contracts are linked by a term structure, therefore we do not expect as many trades when trading this pair. Consequently, we must adapt our strategy to what we are trading and choose the correct parameters to define our entry and exit points.

4 Pairs trading strategy, design & backtest.

4.1 Strategy implementation.

We first start by defining our entry boundaries using the μ and *sigma_eq* in **Table 2**. The variable *Z* is a calibrated factor that will determine the number of trades and thus the frequency of PnL:

$$\mu_e \pm Z\sigma_{eq} \tag{2.4}$$

At a first glance, we see that higher *Z* produces wider bounds and thus greater profit potential, but it comes with some risk as the profits will take longer to materialize or the residual could not reach the equilibrium level μ_e and reverse beyond the entry point, creating undesired losses. Later on, we implement a looping process that will indicate the *Z* value that would yield a desired Sharpe Ratio and Maximum Drawdown, all else being equal. But for the moment, we present the results when we work with a default Z = 0.7:

Replacing the OU parameters and *Z* into equation (2.4) we get:

$$High\ bound\ (HG): (-0.0035) + (0.7)(1.4078)\ =\ 0.9819$$

$$Low\ bound\ (LB): (-0.0035)\ -\ (0.7)(1.4078)\ =\ -0.9890$$

Now that we have defined our entry points, we move onto defining the signal generation. Recall from **Section 3.1** and **Section 3.2**, where we ran the Engle-Granger process, that cointegrated prices have a mean-reverting spread $e_t = \beta'_{Coint} P_t$, and when e_t goes above-below the boundaries, it gives an entry signal (see **Figure 4**). Furthermore, our exit signal is triggered when $e_t \approx \mu_e$.

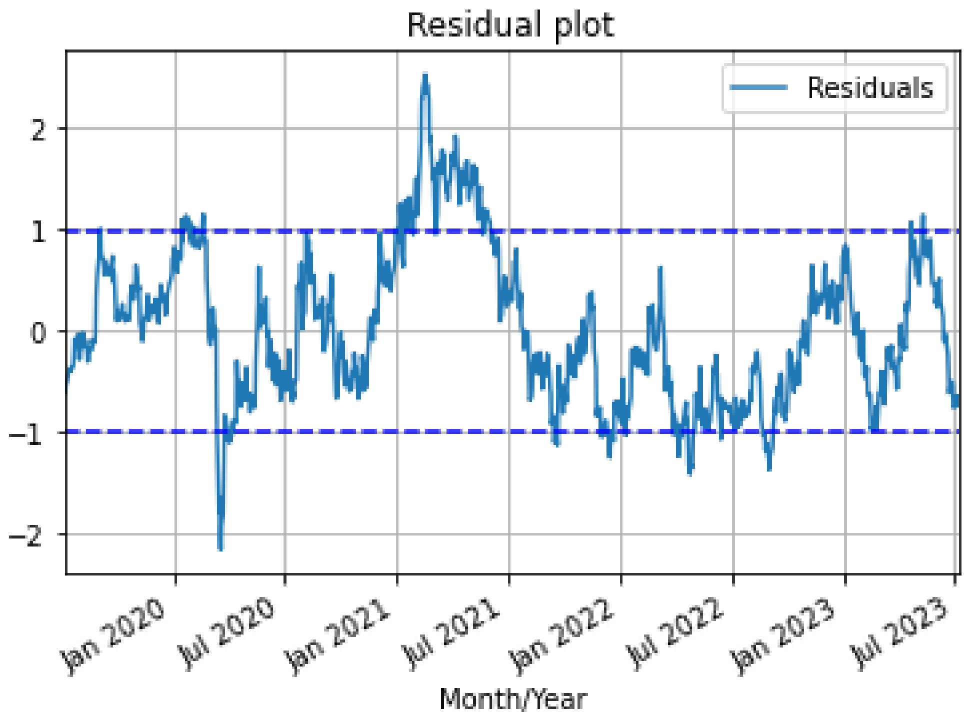


Figure 4. Residual plot with Z = 0.7.

Loadings β'_{Coint} give the weights for each trade (as calculated in **Section 3.1)**. Therefore, the positions for assets A and B are:

$$If\ e_t \gg\ HG\ ;\ Sell\ \ -100\%P^A\ and\ Buy\ \ +67.08\%P^B$$

$$If\ e_t \ll\ LG\ ;\ Buy\ \ +100\%P^A\ and\ Sell\ \ -67.08\%P^B$$

It is important to note that we keep β'_{Coint} stable through time, as recursive re-estimation of the cointegration residual simply contradicts the idea of long-term error correction.

Further, for trading implementation in Python we make us of an already-made code exposed in Halls-Moore (2015) for data handling and portfolio tracking, although some changes were made to adapt the trading rules to our strategy. As explained before, we keep two codes that are zipped in the folder:

a. Code ONE file contains the calculations up to **Section 3.3**.
b. Code TWO is where we take the results from code "a" and put them into the strategy. Note that the parameters β'_{Coint}, μ_e and σ_{eq} will remain fixed as the optimization process is just made to *Z*.

Code file TWO is amended to handle position sizing according to the Engle-Granger process, whilst maintaining the author's proposed trading orders irrespective of cash held, contract and tick size (most common for futures trading) and margin requirements[2]; as well, we do not consider any commissions charged by the broker, the depth of the order book nor the bid/ask spreads as the liquidity of some commodities futures like Copper and Platinum contracts is rather low. Therefore, the fill cost is set to the current market price which is the closing price of each day and the number of contracts (set to 100 as default).

4.2 Backtesting and performance.

In this section we present the strategy's performance over different periods but maintaining constant β'_{Coint}, μ_e and σ_{eq}. In this case, *Z* is set to be 0.7, but we will present a summary of results for a range of different *Z*.

Table 3. Summary of results with Z = 0.7

Total Return	94.42%
Sharpe Ratio	1.05
Max Drawdown	25.06%
Drawdown Duration	185
Signals: 36	
Orders: 36	
Fills: 36	

As shown in the results **Table 3**, the strategy with a random Z=0.7 delivered total return of 94.42% with a Sharpe Ratio of 1.05, which is around our desired value ~ 1. We also had 36 entry and exit signals, for which 36/36 orders were filled. As well, when we developed the OU process, we mentioned that this strategy combined with these assets would not deliver as many

[2] These are unrealistic assumptions and the more factors inherent in trading we consider, the more accurate it would be to assess the performance of the strategy.

trades. Another reason for such a lack of entries is that the bounds are located far away from the cluster (see **Figure 4**).

In the calculation of the Sharpe Ratio, we ignore the use of a benchmark or just the risk-free rate as a measure that our strategy must overcome, or the comparison of its performance against a simply buy and hold strategy using for instance a market index. Therefore, our choice when evaluating the strategy is not subtract the benchmark rate from the strategy's returns because it is self-financing, since we gain a credit interest (R_f) from holding a margin, the actual calculation for returns is: $(R_a + R_f) - R_f = R_a$. Thus, there is no actual subtraction of the risk-free rate for market neutral strategies like the cointegration pairs trade.

Arguably, the strategy should be adapted to the trader's capacity to stomach losses and the ability of the user to drive turbulent periods by hedging or adding capital to his account to comply with margin requirements (if required). Here is where the Maximum Drawdown and the Value at Risk (VaR) analysis take relevance as both indicators can help to gauge the risk of the strategy. In fact, drawdown could be considered as the most relevant performance indicator because if the trader's equity is wiped out then none of the other performance metric matter. Therefore, the strategy must be able to survive without running into a close-out, and it makes sense to pre-define a Maximum Acceptable Drawdown (MADD), which again, is subject to the trader's willpower to continue with the strategy.

Going back to the summary of results, adjusting for higher Z seems to be a wise option for those who would not accept a MADD > 25%. Nevertheless, they must be willing to see their equity in "red" for longer as the drawdown duration could be longer as **Table 4** suggests. Here is where it makes sense to estimate the risk of loss to a strategy by estimating the VaR. The given degree of confidence is settled at 95% which indicates the probability of losing no more than "x" USD in the following day. For a robust analysis, we highlight the periods of which VaR exceeded MADD – DD, or $VaR_t^{95\%} > MADD - DD_t$, where DD_t is the current drawdown.

Table 4. VaR_95% against MADD = 25%

Datetime	PL=F	HG=F	Cash	Commission	Total	Returns	Equity_curve	Drawdown	DD_$	MADD	RHS	VaR_95
16/02/2021	-$127,700.00	$257.82	$245,752.91	$0.00	$118,310.73	-1.71%	1.18310726	23.72%	$28,064.46	$29,577.68	$1,513.22	$2,048.62
19/02/2021	-$129,060.00	$274.03	$245,752.91	$0.00	$116,966.94	-1.52%	1.169669424	25.06%	$29,317.48	$29,241.74	-$75.75	$2,025.35
22/02/2021	-$128,000.00	$278.22	$245,752.91	$0.00	$118,031.13	0.91%	1.180311275	24.00%	$28,328.15	$29,507.78	$1,179.64	$2,043.78

* RHS refers to MADD - DD

Of course, VaR should not be used in isolation to evaluate the risk level of one portfolio is exposed to, as it has it disadvantages such as misleading magnitude of the expected loss beyond the level of VaR (tail risk), since VaR tells us that we are likely to see a loss exceeding a value, but not how much exceeds it. Nevertheless, VaR is straightforward to calculate and given the low frequency of the strategy, we can use VaR as a baseline for risk management, but it must be accompanied with other factors that the trader might consider deciding whether to hold or exit the strategy (MADD for instance). Also note that we are not utilizing Stop Loss orders as the aim is to evaluate the performance of the strategy based on the entry and exit signals coming from the OU process, and we fix a MADD of 25% to give enough room to the strategy to perform. The reason we can deal with a drawdown of this magnitude is that we realized that this behavior may me optimal for the long-term PnL growth of the strategy.

We present the plot of portfolio value, returns and drawdowns in **Figure 5** to observe the evolution of these parameters through the entire data. Finally, **Figure 6** shows the strategy's 30-days rolling Beta to the SPX Index and **Figure 7** compares the equity evolution of a buy and hold investment in the SPX index and our pairs strategy, when Z = 0.7.

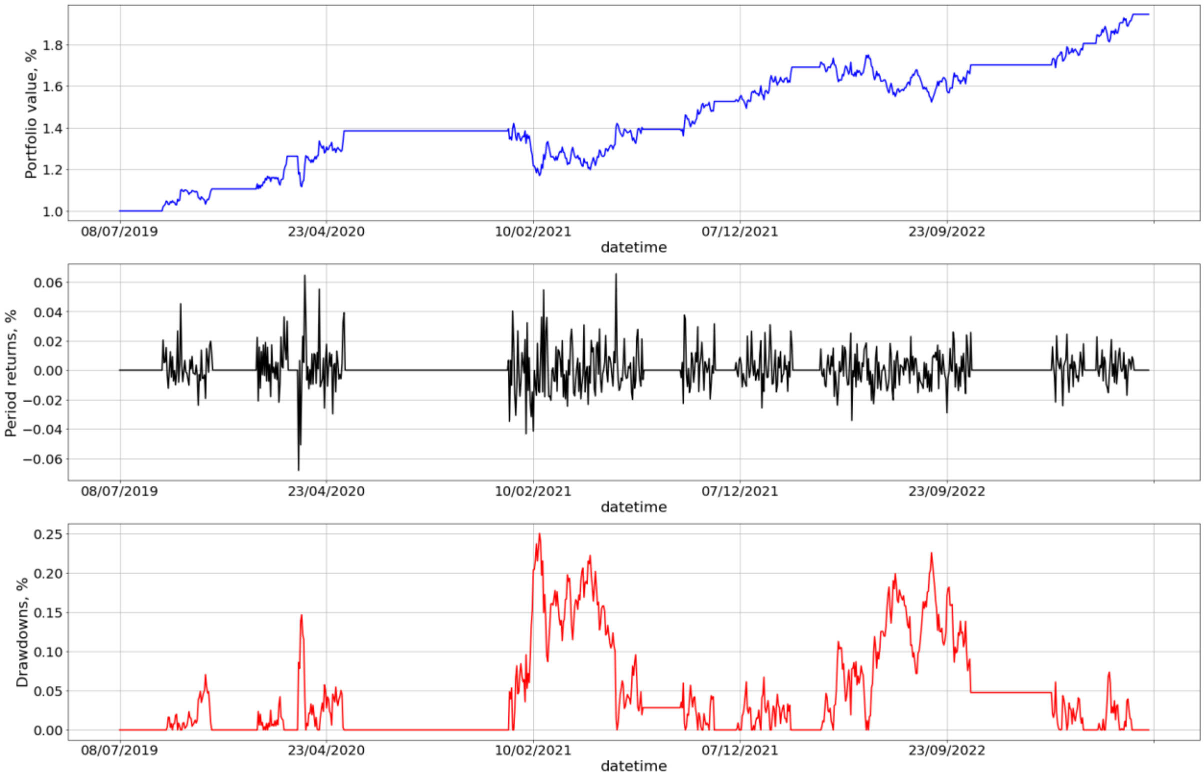


Figure 5 Plot of results with Z = 0.7.

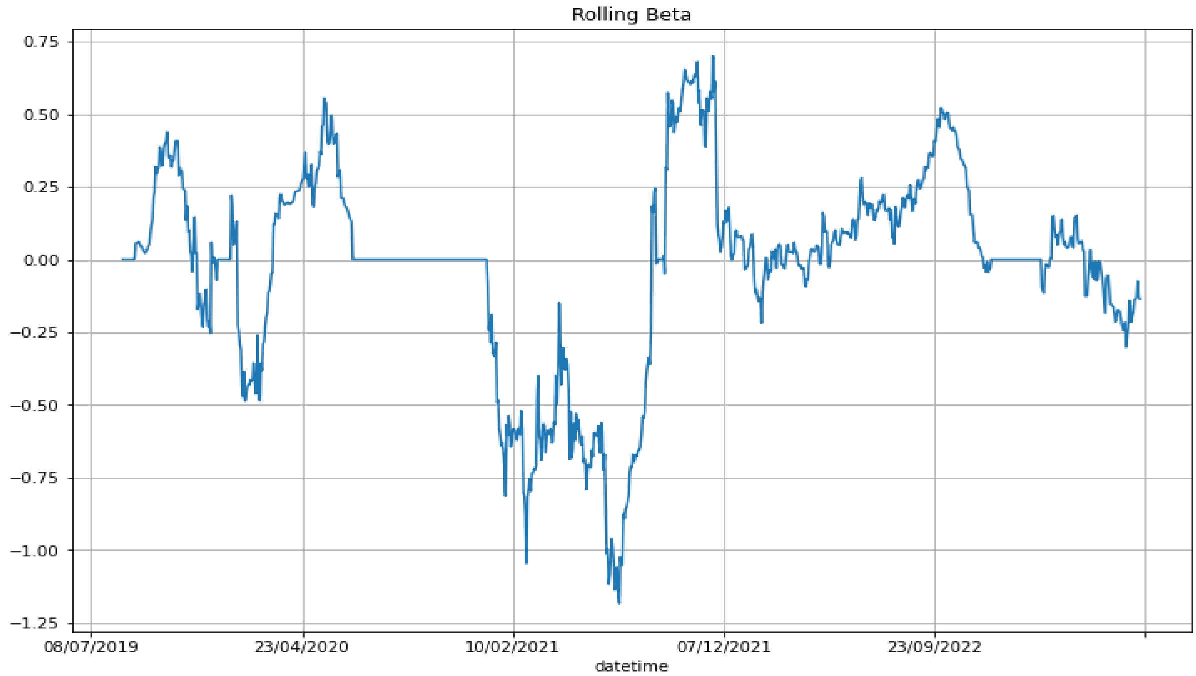


Figure 6. 30-days Rolling Beta with Z = 0.7

The Rolling Beta measures the strategy's movements against a benchmark, which is selected to be the SPX Index, but adjusted to the risk-free rate[3]. As shown in **Figure 6**, during the COVID-19 outbreak and almost the whole 2021 year our strategy presented a negative Beta, as it did not manage to keep it up with the overall bullish market (see **Figure 7**); in contrast, 2021 was the period where we accumulated the maximum drawdowns and the VaR was breached three times, whilst the index bounced really fast from the bottom, we can also observe that there is a big gap between February 2021 and October 2021 when the index overlapped the strategy's equity. Later in 2022, the benchmark almost erased all the gains capitalized in the previous year, showing a giant correction during the period due to major central banks increasing their reference rate to combat inflation. Although it was a bear market overall, our strategy moved less than one to one with the return of the SPX, as the rolling Beta did not surpass the 1 threshold, which in general means good news for the strategy.

Figure 7. SPX investment vs Z=0.7 strategy

[3] For this calculating, we consider the USGG10YR Index yield which for the period between July 2019 and July 2023, presented an average yield of 2% (Bloomberg).

The figure above shows that overall, our strategy with a random *Z* value performed better than a buy and hold investment in the SPX Index during the whole period. Should we have invested $100,000 in the index in July 2019 and cashed it up in July 2023, it would have returned roughly 50%[4], whereas the strategy with *Z* = 0.7 yielded 94.42%. Moreover, at first glance, the maximum drawdown in the SPX Index occurred during the COVID-19 outbreak, where the index suffered a 40.98% drawdown, and it took the benchmark 377 days to break even from that drop. Other than that, there are some improvements to make to the strategy, given that there were times when the SPX index performed better than the cointegration pairs trade, for example during the year 2021. For this reason, we ran the strategy with a range of different *Z* using a simple for loop process in Python.

It is worth noting that there exists more complex process for parameter optimization like Random and Grid Search, however, these conventional methodologies require calculating the model's cost function multiple times to find the most optimized parameters. In fact, calculating the cost function in Random Search turns out to be computationally expensive, and that is why Grid Search seems more appropriate, as can save us time, effort, and resources. Nevertheless, running a Grid Search process requires a pre-defined list of parameters and the model will choose from them the most optimal value, and this would be more appealing when we deal with multiple parameter optimization. We show below that we can obtain acceptable results using a simpler model, with a pre-defined list of *Z* values as well.

4.3 Parameter optimization.

Using our Python code, we interact through a list of *Z* values from 0.5 to 1.5 to find the combination that presents a better Sharpe Ratio, total return, and Maximum drawdown.

Table 5. Adjusting Z values

Z	Total Returns	Sharpe Ratio	Max DD	DD duration
0.5	82.43%	0.88	28.17%	189
0.6	**105.55%**	**1.14**	28.17%	189
0.7	94.42%	1.05	25.06%	**185**
0.8	75.35%	0.99	25.06%	234
0.9	31.96%	0.52	22.61%	272
1	35.62%	0.58	22.61%	272
1.1	31.23%	0.62	**13.55%**	515
1.2	31.23%	0.62	**13.55%**	515
1.3	38.37%	0.78	**13.55%**	515
1.4	43.65%	0.9	**13.55%**	515
1.5	44.47%	0.92	**13.55%**	515

[4] Refer to **Appendix 5.2** for details of the SPX investment.

Table 5 shows a lower return accompanied by a lower Sharpe ratio when the bounds are wider, for instance, when the bounds are adjusted to be at {-2 , 2}, i.e., with a $Z \approx 1.40$. As shown in the figure above, a higher *Z* works good for risk averse traders, given that the Max DD is far smaller and the total return when Z = 1.5 is just below the return of the SPX index. However, we can argue that a more conservative *Z* will not beat the benchmark, and therefore it is necessary to go for a *Z* between 0.5 and 0.8, being the 0.6 the most optimal value as it delivers the best Sharpe Ratio (1.14), with a total return of 105.55%. The maximum drawdown and the 30-days rolling beta (**Figure 8** and **Figure 9** respectively) barely moved, although the strategy with Z = 0.6 suffered far more VaR breakouts when the MADD is fixed at 25% (**Table 7**). Other than that, we can observe in **Figure 10** that the recovery from the Max DD in early 2021 happened much faster than the strategy with Z = 0.7, in fact, the equity value for the optimal strategy moved in tandem and the gap narrowed to the SPX index during 2021, delivering a significant improvement when *Z* is adjusted to 0.6.

Table 6. Summary of results with Z=0.6

Total Return	105.55%
Sharpe Ratio	1.14
Max Drawdown	28.17%
Drawdown Duration	189
Signals: 44	
Orders: 44	
Fills: 44	

Table 7. VaR 95% against MADD = 25% in Optimal Strategy

Datetime	PL=F	HG=F	Cash	Commission	Total	Returns	Equity_curve	Drawdown	DD_$	MADD	RHS	VaR_95
10/02/2021	-$124,420.00	$253.63	$256,099.43	$0.00	$131,933.07	-3.85%	1.319330662	$0.24	$31,072.97	$32,983.27	$1,910.30	$2,254.99
11/02/2021	-$124,440.00	$253.76	$256,099.43	$0.00	$131,913.19	-0.02%	1.319131929	$0.24	$31,094.51	$32,978.30	$1,883.79	$2,254.65
12/02/2021	-$125,640.00	$254.73	$256,099.43	$0.00	$130,714.16	-0.91%	1.307141644	$0.25	$32,379.17	$32,678.54	$299.37	$2,234.15
16/02/2021	-$127,700.00	$257.82	$256,099.43	$0.00	$128,657.25	-1.57%	1.286572488	$0.27	$34,516.03	$32,164.31	-$2,351.71	$2,199.00
17/02/2021	-$125,530.00	$257.01	$256,099.43	$0.00	$130,826.44	1.69%	1.308264399	$0.25	$32,260.10	$32,706.61	$446.51	$2,236.07
18/02/2021	-$127,240.00	$262.41	$256,099.43	$0.00	$129,121.84	-1.30%	1.29121836	$0.26	$34,040.78	$32,280.46	-$1,760.32	$2,206.94
19/02/2021	-$129,060.00	$274.03	$256,099.43	$0.00	$127,313.47	-1.40%	1.273134653	$0.28	$35,866.33	$31,828.37	-$4,037.97	$2,176.03
22/02/2021	-$128,000.00	$278.22	$256,099.43	$0.00	$128,377.65	0.84%	1.283776504	$0.27	$34,799.96	$32,094.41	-$2,705.54	$2,194.22
24/02/2021	-$125,560.00	$288.17	$256,099.43	$0.00	$130,827.60	-1.37%	1.308276023	$0.25	$32,258.86	$32,706.90	$448.04	$2,236.09
26/04/2021	-$124,260.00	$298.02	$256,099.43	$0.00	$132,137.45	-0.90%	1.321374512	$0.23	$30,851.04	$33,034.36	$2,183.32	$2,258.48
27/04/2021	-$124,690.00	$317.78	$256,099.43	$0.00	$131,727.21	-0.31%	1.317272114	$0.24	$31,295.66	$32,931.80	$1,636.15	$2,251.47
06/05/2021	-$125,550.00	$309.47	$256,099.43	$0.00	$130,858.91	-2.16%	1.308589058	$0.25	$32,225.62	$32,714.73	$489.11	$2,236.63
07/05/2021	-$125,230.00	$319.32	$256,099.43	$0.00	$131,188.75	0.25%	1.311887499	$0.24	$31,874.13	$32,797.19	$923.06	$2,242.26
10/05/2021	-$126,300.00	$317.14	$256,099.43	$0.00	$130,116.58	-0.82%	1.301165773	$0.25	$33,008.70	$32,529.14	-$479.56	$2,223.94
17/05/2021	-$124,250.00	$316.34	$256,099.43	$0.00	$132,165.77	-1.61%	1.321657733	$0.23	$30,820.22	$33,041.44	$2,221.22	$2,258.96

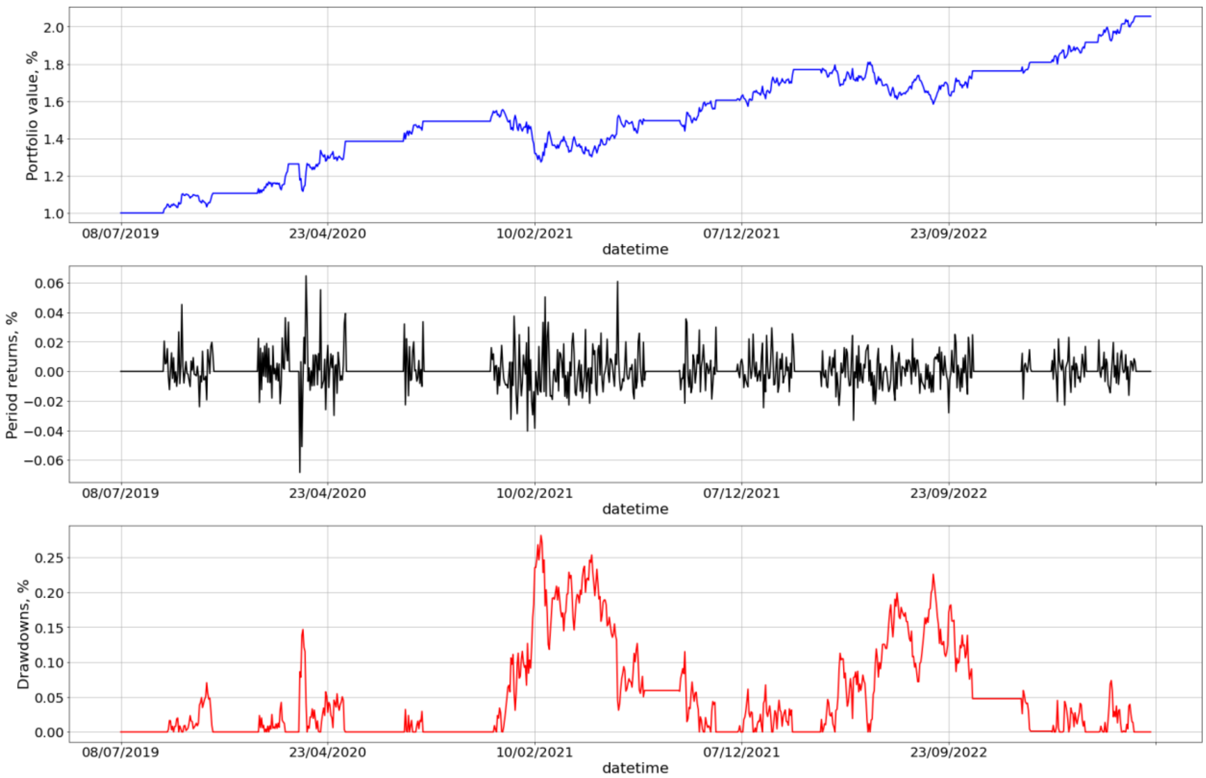


Figure 8. Plot of results with Z = 0.6

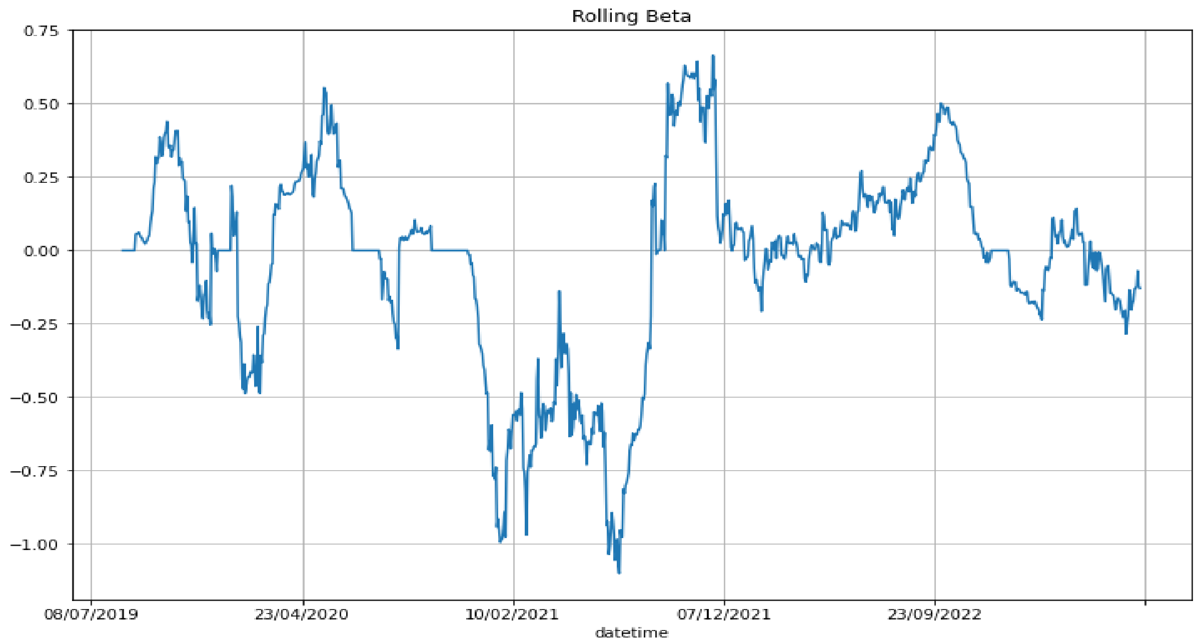


Figure 9. 30-days Rolling Beta with Z=0.6

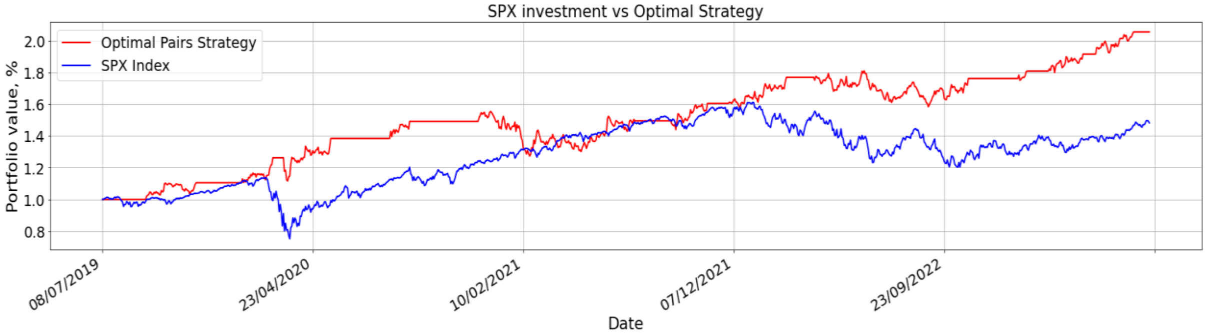


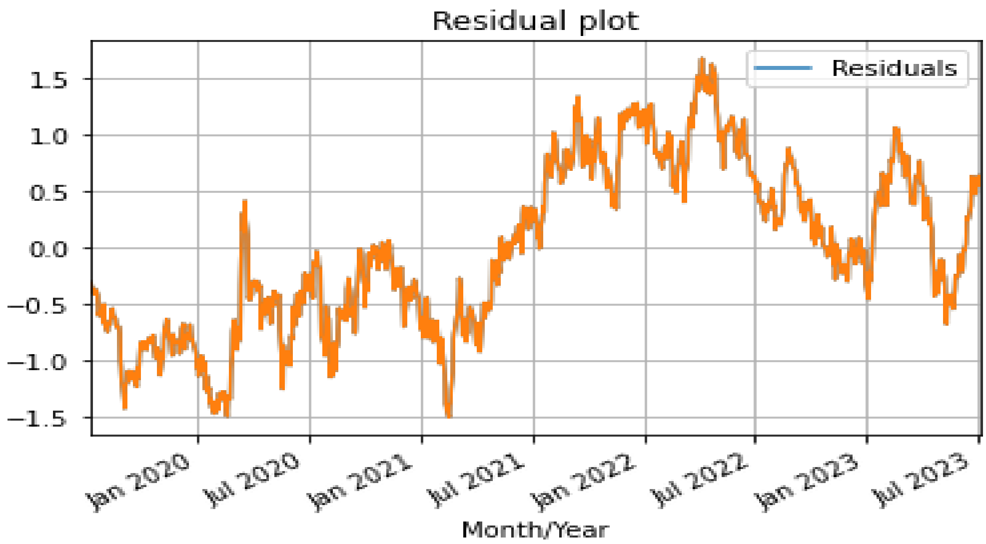
Figure 10. Equity evolution of SPX investment vs Optimal strategy

Figure 11. Drawdown of SPX investment vs Optimal Strategy

5 Appendix.

5.1 Switching variables in the Engle-Granger process.

As observed in the figure below, the residual plot gives us a free insight into whether the residuals are stationary. Clearly, there is a trend from July 2021 and onwards, suggesting a non-stationary process.



We can confirm the aforementioned hypothesis with the following ADF test:

p-value	T-statistic	ADF Critical Values
0.18	-2.2478	{'1%': -3.4370, '5%': -2.8644, '10%': -2.5683}

Therefore, it is unnecessary to run the step 2 of the EG procedure.

5.2 SPX investment.

Total Return	48.24%
Max Drawdown	40.98%
Drawdown Duration	377
Signals: 2	
Orders: 2	
Fills: 2	

6 References

Diamond, R. V. (2014). Learning and Trusting Cointegration in Statistical Arbitrage. *Social Science Research Network*, 27. doi:https://dx.doi.org/10.2139/ssrn.2220092

Halls-Moore , M. L. (2015). *Successful Algorithmic Trading.* Retrieved 07 2023, from https://www.quantstart.com/successful-algorithmic-trading-ebook/

7 Bibliography

Diamond, R. (n.d.). *Cointegration and statistical arbitrage*. Retrieved from CQF Life long learning: https://lifelonglearning.cqf.com

Hilpisch, Y. (2015). *Python for finance.* O'Reilly. doi:1492024333

Pipis, G. (2021, 01 29). *Stocks Market Beta With Rolling Regression*. Retrieved from predictivehacks.com: https://predictivehacks.com/stocks-market-beta-with-rolling-regression/

Zhou, S. (n.d.). *Cointegration: Pairs Trading*. Retrieved from CQF Life long learning: https://lifelonglearning.cqf.com/