

1 Implemented Strategy

The implemented strategy is a simple pairs trading strategy, in which we trade on the mean reversion of a time series, given by the difference of two cointegrating assets.

Let us suppose that the daily closing price of two assets X and Y at time t is denoted with x_t and y_t , respectively. We define the spread z_t as

$$z_t := y_t - \beta \cdot x_t \quad (1)$$

where β is the *hedge ratio* of Y against X (given by the coefficient of the linear regression of Y against X).

If we denote with μ and σ the mean and the standard deviation of $\{z_t\}$ respectively, we obtain the z_{score} of $\{z_t\}$, defined as

$$z_{score,t} = \frac{z_t - \mu}{\sigma}. \quad (2)$$

We define the entry and exit signals in the following way:

1) **Entry signal:** If no position is taken:

- if $z_{score,t} \geq 2$: take SHORT position in z_t , i.e. SHORT y_t and BUY $\beta \cdot x_t$
- if $z_{score,t} \leq -2$: take LONG position in z_t , i.e. BUY y_t and SHORT $\beta \cdot x_t$

2) **Exit signal:** If position has been taken at time t_0 for any time $t > t_0$:

- if SHORT position has been taken in z_{t_0} and $z_{score,t} \leq 0$: CLOSE position, i.e. BUY y_t and SHORT $\beta \cdot x_t$
- if LONG position has been taken in z_{t_0} and $z_{score,t} \geq 0$: CLOSE position, i.e. SHORT y_t and BUY $\beta \cdot x_t$
- if the current drawdown exceeds 20%, CLOSE position and do not open another one before $|z_{score,t}| \leq 0.5$

A portfolio of 15 pairs (30 assets) has been considered, where each pair has been selected among a set of assets from the same type (financial sector, technology, currencies, futures etc). In order to generate a pair of assets, a cointegration test has been performed between all the assets in a particular sector, and only those with significant statistical evidence have been selected. Furthermore, a portfolio consisting of only one pair at a time has been tested during a time interval of 2 years (referred as *in-sample period*). Only pairs having in-sample Sharpe ratio higher than 0.5 have been included in the final portfolio. Finally, a slippage of 5% has been included, in order to model trading costs.

The in-sample performance of the final portfolio can be seen in Figure 1, with overall Sharpe ratio of 2.7214.

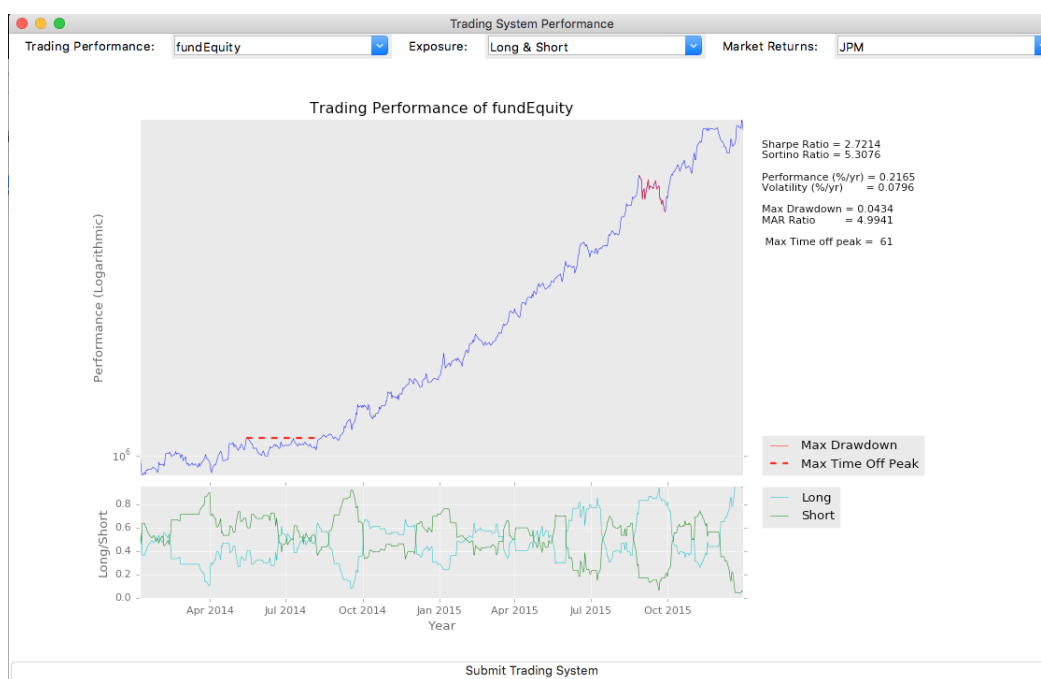


Fig. 1: In-sample performance of the portfolio of pairs

The performance of the same portfolio during the period January-August 2016 (referred as *out-of-sample period*) can be found in Figure 2 with Sharpe ratio 2.1301.

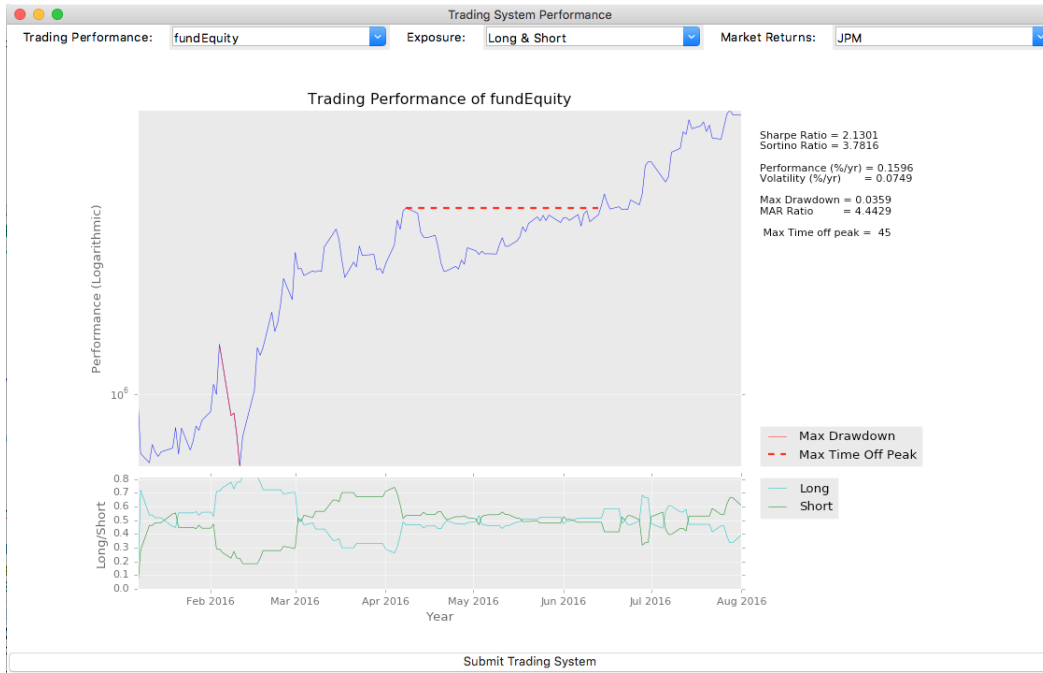


Fig. 2: Out-of-sample performance of the portfolio of pairs

2 Detailed description

In this section, we provide a detailed description of the research and implementation of the trading algorithm.

2.1 Pairs selection

In order to form a pair of assets, which could be a suitable candidate of the final portfolio, we consider different market sectors: *Financial, IT services, Currencies, Agricultural, Aerospace, Oil and Gas, Futures, Chemical, Tobacco, Clothing, Real Estate, Healthcare, Commodities, Industry and Energy*. Assets within each sector are tested for cointegration and candidate pairs are formed.

The cointegration test can be performed in **python** via

```
statsmodels.tsa.stattools.coint(x,y)
```

The p-value of the cointegration test is returned and we consider pairs with p-value in the range 0.001 – 0.005 (hence, high statistical evidence of cointegration).

Finally, a pairs trading strategy on each of the cointegrating pairs is tested on the in-sample period and a pair is included in the final portfolio only if its Sharpe ratio is above 0.5.

2.2 Half-life of a pair and lookback period

In order to update certain parameters (the mean and standard deviation of $\{z_t\}$ as well as the hedge ratio β), we need a lookback period, which is suitable for each pair. This can be done by exploiting the half-life of a mean reverting process.

A mean-reverting process have to satisfy the Ornstein-Uhlenbeck stochastic differential equation:

$$dz_t = (\lambda z_{t-1} + \mu)dt + d\varepsilon$$

with $\lambda < 0$ (hence, a positive value of λ provides an additional test for mean reverting series). We can estimate λ by a linear regression of $\Delta z_t = z_t - z_{t-1}$ against z_{t-1} . The half-life of the process (i.e. the expected time for the process to revert) is given by

$$half-life = -\frac{\log(2)}{\lambda}.$$

It is a good practice to set a lookback period equal to a small multiple of the half-life. In our strategy, we set

$$lookback = \min\{1.5 * half-life, 30\}$$

we estimate the mean and standard deviation of z_t as well as the hedge ratio β on the last *lookback* days.

2.3 Dynamic hedge ratio

In order to make the trading system more robust, we dynamically estimate the hedge ratio β once a month based on the *lookback* period of time. This is done by a linear regression of $\{x_t\}$ against $\{y_t\}$, hence β is given by

$$\beta = \arg \min_{\alpha_0, \beta_0} \sum_{s=t-lookback}^t |y_s - (\alpha_0 + \beta_0 \cdot x_s)|^2.$$

2.4 Pairs evaluation

Pairs have been selected based on the following two criteria:

- 1) p-value less then 0.001 from the cointegration test
- 2) pairs satisfying 1) have been tested during the *in-sample* period (2014 – 2016) and only those which achieved Sharpe ratio above 0.5 have been included into the final portfolio.

Assets from the following markets have been tested and the Sharpe ratio of each pair has been obtained for the in-sample period:

- **Financial sector:** BAC, BK, C, GS, JPM, MS, USB, WFC, BBT, CMA, ETFC, FITB, HBAN, KEY, MTB, NTRS, PBCT, PNC, RF, SCHW, STI, STT, ZION.

Pair	Sharpe ratio	Included
JPM-SCHW	0.7553	Yes
MS-STT	0.7094	Yes
ETFC-HBAN	1.4802	Yes

- **IT Services:** AAPL, ACN, AMZN, T, CSCO, CMCSA, EBAY, EMC, FB, GOOG, GOOGL, HPQ, INTC, IBM, MSFT, ORCL, QCOM, TXN, TWX, FOXA, VZ, DIS, ADBE, ADI, ADSK, AKAM, AMAT, APH, AVGO, CA, CBS, CERN, CRM, CSC, CTL, CTSH, CTXS, DISCA, DISCK, DNB, EA, EXPE, FFIV, FIS, FISV, FLIR, FTR, GCI, INTU, IPG, JNPR, KLAC, LLTC, LRCX, LVL, MCHP, MSI, MU, NFLX, NTAP, NVDA, NWSA, OMC, PAYX, PBI, PCLN, RHT, STX, SYMC, TDC, TEL, TRIP, VIAB, VRSN, WDC, WIN, XLNX, XRX, YHOO.

Pair	Sharpe ratio	Included
AAPL-APH	0.6581	Yes
FB-QCOM	1.2283	Yes
FB-FLIR	0.9742	
HPQ-FOXA	1.1262	Yes
HPQ-CTL	-0.1811	
QCOM-VRSN	-0.3518	
CRM-PBI	0.5599	Yes
FLIR-NWSA	0.7793	Yes
FLIR-VIAB	-1.2582	
JNPR-XRX	0.2037	

- **Currencies:** F_AD, F_BP, F_CD, F_DX, F_EC, F_JY, F_MP, F_SF.

No cointegrating pairs found.

- **Agricultural Sector:** F_B0, F_C, F_CC, F_CT, F_FC, F_KC, F_LB, F_LC, F_LN, F_NR, F_O, F_OJ, F_S, F_SB, F_SM, F_W, ADM, DE.

No cointegrating pairs found.

- **Aerospace Sector:** BA, GD, HON, LMT, RTN, UTX, COL, HRS, LLL, NOC, TXT.

Pair	Sharpe ratio	Included
HON-NOC	-0.5076	

- **Oil and Gas Sector:** MPC, PSX, TSO, VLO, APC, APA, COP, DVN, OXY, CHK, COG, DNR, EQT, HES, NBL, NFX, PXD, QEP, RRC, SWN, XEC.

Pair	Sharpe ratio	Included
DNR-NBL	0.1246	

- **Futures:** F_AD, F_BO, F_BP, F_C, F_CC, F_CD, F_CL, F_CT, F_DX, F_EC, F_ED, F_ES, F_FC, F_FV, F_GC, F_HG, F_HO, F_JY, F_KC, F_LB, F_LC, F_LN, F_MD, F_MP, F_NG, F_NQ, F_NR, F_O, F_OJ, F_PA, F_PL, F_RB, F_RU, F_S, F_SB, F_SF, F_SI, F_SM, F_TU, F_TY, F_US, F_W, F_XX, F_YM.

No cointegrating pairs found.

- **Chemical Sector:** DOW, DD, MON, APD, CF, ECL, EMN, FMC, IFF, LYB, MOS, PPG, SHW.

No cointegrating pairs found.

- **Tobacco Sector:** MO, PM, RAI.

No cointegrating pairs found.

- **Clothing Sector:** NKE, COH, GPS, KORS, LB, PVH, RL, ROST, UA, URBN, VFC.

No cointegrating pairs found.

- **Real Estate Sector:** SPG, AIV, AMT, AVB, BXP, CBG, EQR, ESS, GGP, HCP, HST, IRM, KIM, LUK, MAC, PLD, PSA, VNO, VTR.

Pair	Sharpe ratio	Included
GGP-MAC	1.9376	Yes

- **Healthcare Sector:** ABT, ABBV, AMGN, BAX, BMY, CVS, GILD, JNJ, LLY, MDT, MRK, PFE, UNH, WBA, ABC, AET, AGN, ALXN, BCR, BDX, BSX, CAH, CELG, CI, DGX, DHR, DVA, ESRX, EW, HUM, ISRG, LH, MCK, MJN, MNK, MYL, PDCO, PRGO, REGN, STJ, SYK, THC, TMO, UHS, VAR, VRTX, XRAY, ZTS.

Pair	Sharpe ratio	Included
MDT-ZTS	-0.9783	
MRK-MJN	0.6890	Yes
AET-HUM	0.6733	Yes
AGN-ZTS	-0.4765	
CAH-ZTS	-0.6124	
VRTX-ZTS	-0.6292	

- **Industrial Sector:** BA, CAT, EMR, FDX, F, GD, GM, HON, LMT, NSC, RTN, UNP, UPS, UTX, A, ALLE, AME, BLL, BMS, BWA, CHRW, CMI, COL, CTAS, DE, DLP, DOV, ETN, EXPD, FLR, FLS, GLW, GRMN, GT, HOG, HRS, IR, ITW, JCI, JEC, JOY, KSU, LLL, MAS, MLM, NLSN, NOC, OI, PCAR, PH, PKI, PNR, PWR, R, RHI, ROK, ROP, SEE, SNA, SWK, TXT, TYC, URI, VMC, WAT, XYL.

Pair	Sharpe ratio	Included
EMR-MAS	-0.0769	
HON-BMS	-0.4541	
LMT-FLR	0.2849	
AME-GRMN	-0.3688	
AME-GT	-0.2889	
AME-HRS	0.5482	Yes
AME-KSU	0.0899	
AME-MLM	-0.1210	
AME-PKI	0.3429	
AME-TYC	0.0938	
AME-VMC	-0.0856	

- **Energy Sector:** F_CL, F_HO, F_NG, F_RB, AEP, APC, APA, CVX, COP, DVN, EXC, XOM, HAL, NOV, OXY, SLB, SO, AEE, AES, BHI, CHK, CMS, CNP, CNX, COG, D, DNR, DO, DTE, DUK, ED, EIX, EOG, EQT, ESV, ETR, FE, FSLR, FTI, HES, HP, KMI, MPC, MRO, MUR, NBL, NBR, NE, NEE, NFX, NI, NRG, OKE, PCG, PEG, PNW, PPL, PSX, PXD, QEP, RIG, RRC, SCG, SE, SRE, SWN, TSO, VLO, WEC, WMB, XEC, XEL.

Pair	Sharpe ratio	Included
COP-EOG	0.2611	
XOM-CHK	-0.7106	
HAL-NBR	0.2698	
CMS-XEL	1.6298	Yes
D-DUK	0.0612	
D-FE	0.1039	
D-PSX	0.9998	Yes
DUK-FE	1.1402	Yes
HP-NE	-0.0634	
NEE-EOG	-0.2656	

2.5 Risk management

In order to limit the potential losses of the algorithm, we implement a risk management logic, which closes the current taken position whenever the current drawdown exceeds 20% (this usually happens for high values of $|z_{score,t}|$). Once a position is closed due to exceeded maximum drawdown, the algorithm is not allowed to trade until $z_{score,t}$ returns under a reasonable values ($|z_{score,t}| \leq 0.5$).

We compute the drawdown in the following manner: let us enter a (long or short) position at time t_0 with allocated capital p_x and p_y (the portion of the available capital, invested in X and Y respectively). The value of the portfolio at time t_0 is defined as

$$P_{t_0} := \frac{p_x}{|p_x| + |p_y|} x_{t_0} + \frac{p_y}{|p_x| + |p_y|} y_{t_0}.$$

Let us denote with $\hat{P}_{[t_0, t]}$ the maximum value, achieved by the portfolio over the time interval $[t_0, t]$, i.e.

$$\hat{P}_{[t_0, t]} := \max_{s \in [t_0, t]} P_s.$$

Then we compute the drawdown of the portfolio at time t as

$$DD_t := \frac{P_t - \hat{P}_{[t_0, t]}}{\hat{P}_{[t_0, t]}}.$$

A previously opened position is closed due to exceeded maximum drawdown, whenever DD_t goes below -0.2 .

2.6 Trading logic

The trading logic can be summarized in the following algorithm:

Algorithm 1 Trading logic

```

1: Set POSITION = NONE and  $p = [0, 0]$ 
2: for each day in trading period do
3:   get CLOSE price of  $X, Y$ 
4:   if  $day == 1$  (first day of the month) then
5:     update beta and lookback
6:   end if
7:   compute  $z_{score}$ 
8:   if POSITION == NONE then
9:     if  $z_{score,t} \geq 2$  then
10:      BUY  $\beta \cdot X$ , SELL  $Y$ 
11:      set POSITION = ENTERED_ABOVE,  $p = [\beta, -1]$ 
12:    end if
13:    if  $z_{score,t} \leq -2$  then
14:      SELL  $\beta \cdot X$ , BUY  $Y$ 
15:      set POSITION = ENTERED_BELOW,  $p = [-\beta, 1]$ 
16:    end if
17:  end if
18:  if POSITION == MAX_DD_EXCEEDED and  $|z_{score,t}| \leq 0.5$  then
19:    set POSITION = NONE (return to normal state)
20:  end if
21:  if POSITION  $\notin \{\text{NONE}, \text{MAX\_DD\_EXCEEDED}\}$  then
22:    if POSITION == ENTERED_ABOVE and  $z_{score,t} \leq 0$  then
23:      set POSITION = NONE,  $p = [0, 0]$  (close position)
24:    else if POSITION == ENTERED_BELOW and  $z_{score,t} \geq 0$  then
25:      set POSITION = NONE,  $p = [0, 0]$  (close position)
26:    else
27:      compute DRAWDOWN
28:      if DRAWDOWN  $\geq$  MAX_DRAWDOWN then
29:        set POSITION = MAX_DD_EXCEEDED,  $p = [0, 0]$ 
30:        (close position due to exceeded maximum drawdown)
31:      end if
32:    end if
33:  end if
34:  return  $p$ 
35: end for

```

Note that the parameter p in Algorithm1 is the portion of capital to be allocated. In order to make the whole algorithm reproducible, it has been implemented into a separate python class.

2.7 Code execution

We provide all the source code of the trading algorithm as well as the research done. In order to reproduce the results stated in the previous sections, the

reader should have running **python** and the following libraries: **numpy**, **scipy**, **pandas** and **quantiacsToolbox**.

The evaluation of the single pairs of stocks is implemented in the **pairs.py** file. Each pair can be executing by commenting it out in the source code (lines 260 – 317). The file can be executed by

```
python pairs.py
```

The implementation of the in-sample and out-of-sample portfolio evaluation is held in **portfolio-IS.py** and **portfolio-OOS.py** respectively. Both the files can be run by

```
python portfolio-IS.py
```

and

```
python portfolio-OOS.py
```

The research done is forming the pairs of securities can be found in **research/identify-pairs.ipynb**. In order to run the code, **ipython notebook** is required, but we provide also an *html* file containing all the results (**research/identify-pairs.html**).

Finally, we provide all the data downloaded from Quantiacs (for sake of time saving), which is contained in the **tickerData** folder.

2.8 Comments and further ideas

In order to improve the performance of the algorithm, the following ideas could be implemented:

- 1) More robust cointegration research could be done on the assets. Although the cointegration test performed better than correlation and minimum distance between assets, I believe there is still a lot of space for a better pairs selection.
- 2) Additional tests, such as the Hurst exponent and p-value of ADF test on the residuals of $y_t - \beta \cdot x_t$ could be used as additional entry (and exit signals). I did not put a lot of effort in that direction, but a better research could lead to improved performance.
- 3) Although pairs trading strategies are usually market neutral, I did not monitored the beta of the portfolio with respect to the overall market (or the SPY index).
- 4) The allocation of capital is not optimized (the capital is allocated equally between pairs). I believe there is a lot of space for improvement in this direction (simple idea could be capital proportional to the Sharpe ratio obtained during the *in-sample* period).
- 5) The risk management I am using limits the drawdown on a single trades, but does not take in consideration sequence of losing trades. Hence the overall loss might exceed 20%.

- 6) Some parameters (such as the mean and standard deviation of $y_t - \beta \cdot x_t$) are estimated using simple moving average. More sophisticated methods such as exponential moving average or Kalman filter (widely suggested in literature) will almost surely improve the performance of the algorithm.
- 7) Finally, I decided to update the parameters once a month, in order to prevent overfitting the training data, nevertheless I did not optimized this parameter (hence twice a month might improve the performance).

References

- [1] Caldeira, J. and Moura, G. V.: *Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy*, Available at SSRN: <http://ssrn.com/abstract=219639>, (2013)
- [2] Chan, E. P.: *Algorithmic Trading: Winning Strategies and Their Rationale*, Wiley Finance, (2013)
- [3] Gundersen, R.J.: *Statistical Arbitrage: High Frequency Pairs Trading*, Master thesis, Norwegian School of Economics, (2014)
- [4] Quantopian Lectures : <https://www.quantopian.com/lectures>
- [5] QuantStart: <https://www.quantstart.com/>
- [6] Triantafyllopoulos, K. and Montana, G.: *Dynamic modeling of mean-reverting spreads for statistical arbitrage*, Available at <https://arxiv.org/pdf/0808.1710.pdf>, (2009)