WOLFE RESEARCH

Internship Project: Pairs Trading

Kai Wu

August 24, 2020

Outline

- Overview
- The Strategy
- Programming
- Analysis
- Next Step

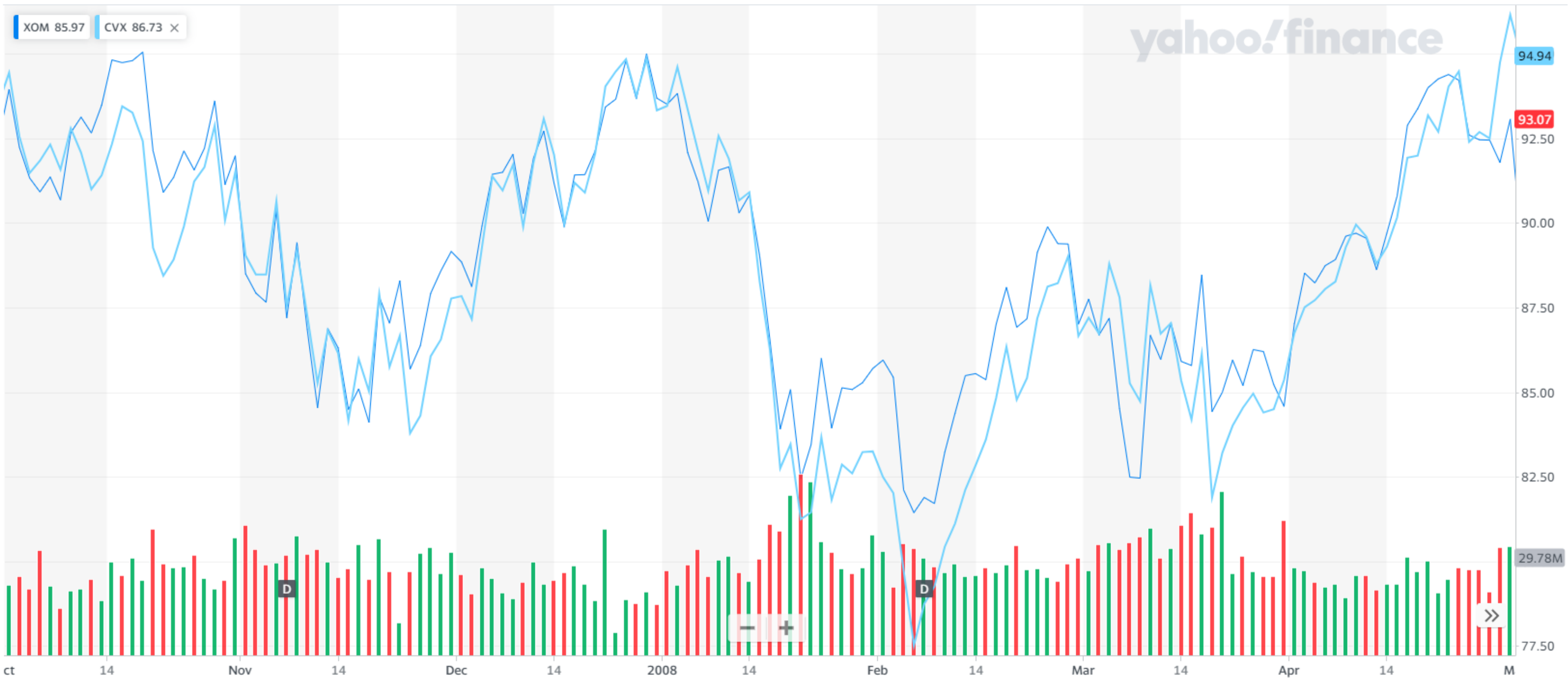
Overview

- Pairs trading is a statistical arbitrage strategy that profits from the temporary divergence of the prices of two or more securities that are correlated/sharing similar characteristics/affected by same factors.

- Pairs Trading is market neutral, and is evidenced to perform well in turbulent markets (Deutsche Bank, 2012 & 2015)

Sources: Google Images

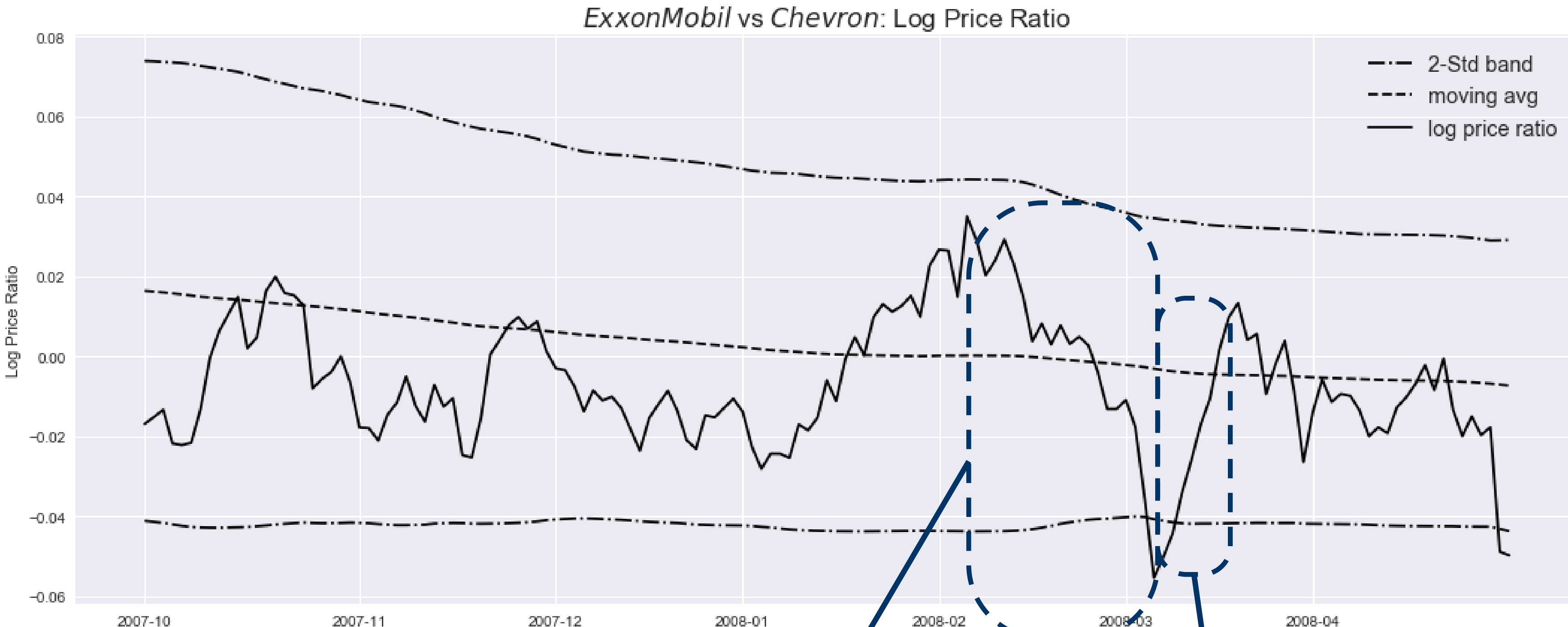
Exxon Mobil vs ***Chevron***, Historic Stock Prices, Oct 2007 – Apr 2008



Sources: Yahoo Finance

Overview

Exxon Mobil vs ***Chevron***, Logarithm of Price Ratio, Oct 2007 – Apr 2008



Two things happened:

1. Dow Jones Index added Chevron
2. Exxon Mobil and Venezuela were in dispute

3.61% in 4 days

The Strategy

- **Two stages**: <u>Selection</u> and Trading

1. Pairs Selection
How to identify pair candidates

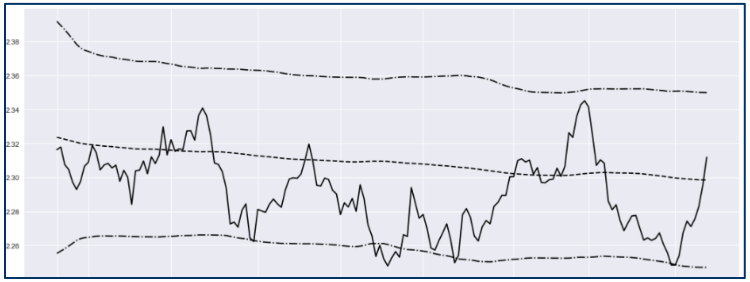
- Beta Neutral: difference of beta ≤ 0.3
- Industry Neutral: same GICS industry
- High Return Correlation
- Strong Mean Reversion of Relative Price

Following Deutsche Bank 2015 Report

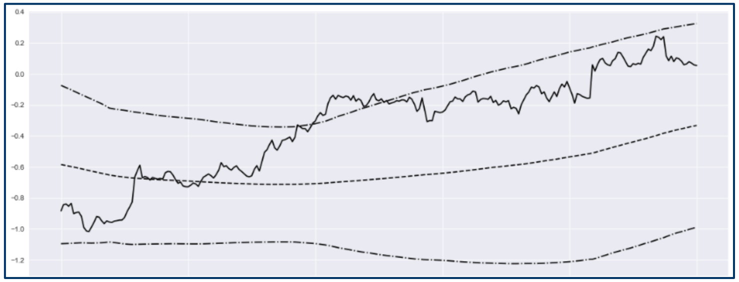
The Strategy

- **Two stages**: Selection and Trading

Stationary Time-series



Non-stationary Time-series



➢ Strong Mean Reversion of Relative Price

- ADF Test
  - Test whether a time series is stationary
- Cointegration Test
  - Test whether there exists a stationary linear combination of multiple time-series
- Hurst Exponent:
  - Hurst Exponent < 0.5 indicates a mean-reverting time series
- Half-life based on calibration of OU process
  - A shorter half life indicates a faster reversion to mean

The Strategy

- **Two stages**: Selection and Trading

2. Pair Trading
How to profit from identified pairs

Following Deutsche Bank 2015 Report

Open
- 2 Std Cross-over & Cross-back
- Refresh candidate pool at each month end

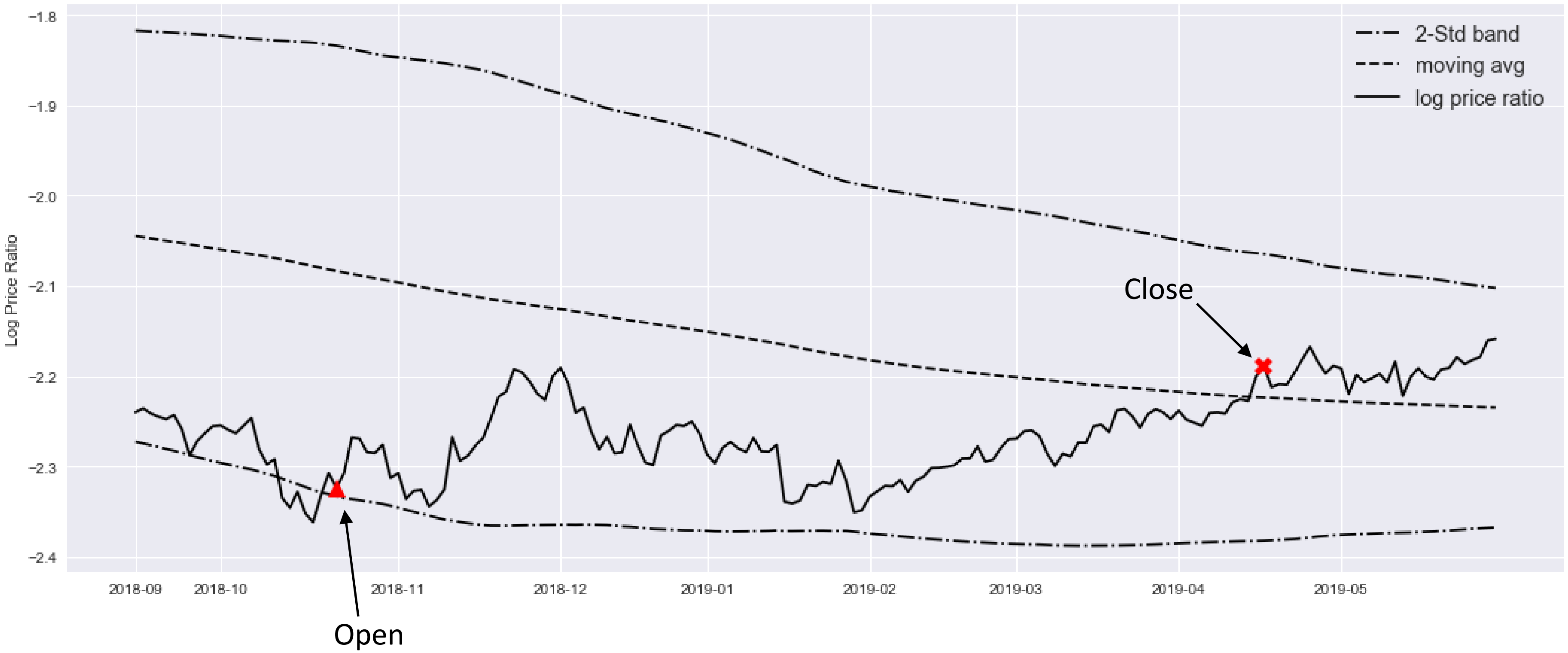
Allocation
- Dollar Neutral

Close
- Mean Reversion
- Maximum Holding Period (9 months)
- Stop Loss (-60%)
- Delisting

The Strategy

- **Two stages**: Selection and <u>Trading</u>

Trading Rule: an Example



Programming

- **Python Implementation:**

class **PairBacktesting**
backtesting & analytics of the pairs trading strategy

class **PairFilter**
flexible & efficient pair selection

class **Pair**
backtesting of an arbitrary pair

class **DataLoader**
foundational data pipeline

Programming

- **Challenges:**
  - Pair selection process can be quite involved
    - A series of filtering conditions
    - Each condition has multiple tunable parameters
  - Pair selection is computationally expensive
    - Statistical computations like cointegration tests are time-consuming
    - 80k~105k potential pairs cross-sectionally

Programming

Challenge 1: Pair selection process can be quite involved

Solution: specify the selection process in a chain before computation starts

```
DATA = DataLoader()  # Initialize the data pipeline
pair_filter = PairFilter(DATA)  # Create a PairFilter object
conditions = pf_filter.new_filter()
conditions.start()
conditions.same_industry(gics_level=3)
conditions.industry_subset(subset=40)
conditions.beta_diff().less_than(0.3)
conditions.correlation(corr_type='residual').top(quantile=0.2)

conditions.ADF(max_lag=4).pvalue(cutoff=0.05)

conditions.end()
```

Programming

Challenge 1: Pair selection process can be quite involved

Solution: specify the selection process in a chain before computation starts

```
DATA = DataLoader()
pair_filter = PairFilter(DATA)
conditions = pf_filter.new_filter()
conditions.start()  # start composing conditions
conditions.same_industry(gics_level=3)  # stocks should be of the same GICS level 3
conditions.industry_subset(subset=40)  # only focus on the Financials sector (GICS code: 40)
conditions.beta_diff().less_than(0.3)  # difference of beta should be less than 0.3
conditions.correlation(corr_type='residual').top(quantile=0.2)
# residual correlation should be among the top 20% percentile
conditions.ADF(max_lag=4).pvalue(cutoff=0.05)
# ADF test with a lag of 4 should meet the 0.05 significance level
conditions.end()  # end composition

Pair_filter.filter_given_dates(dates=['2019-05-31', '2019-06-30'])
```

Programming

Challenge 2: Pair selection is computationally expensive

Solutions:

1. Use vectorized computation other than for-loops whenever possible
2. Use Multi-processing to speed up time-consuming computations
   - Single CPU core: more than 24 hours
   - Multiple cores: under 4 hours
3. Save intermediate results to reduce redundant computation
   - Using different cutoffs/thresholds/percentiles requires no new computation

   Example:

   old: conditions.ADF(max_lag=4).pvalue(cutoff=0.05)

   new: conditions.ADF(max_lag=4).pvalue(cutoff=0.01)
   - After expanding the backtesting interval, results corresponding to the older interval will be reused

Analysis

Baseline model

(1)

- Industry Neutrality
- ↓
- Beta Neutrality
- ↓
- Correlation Top 20%
- ↓
- ADF test Top 5%

↓

(2) Half-life x 2 < holding days limit

↓

(3) Hurst Exponent < 0.5

↓

(4) Johansen Cointegration Test Top 5%

Other Settings

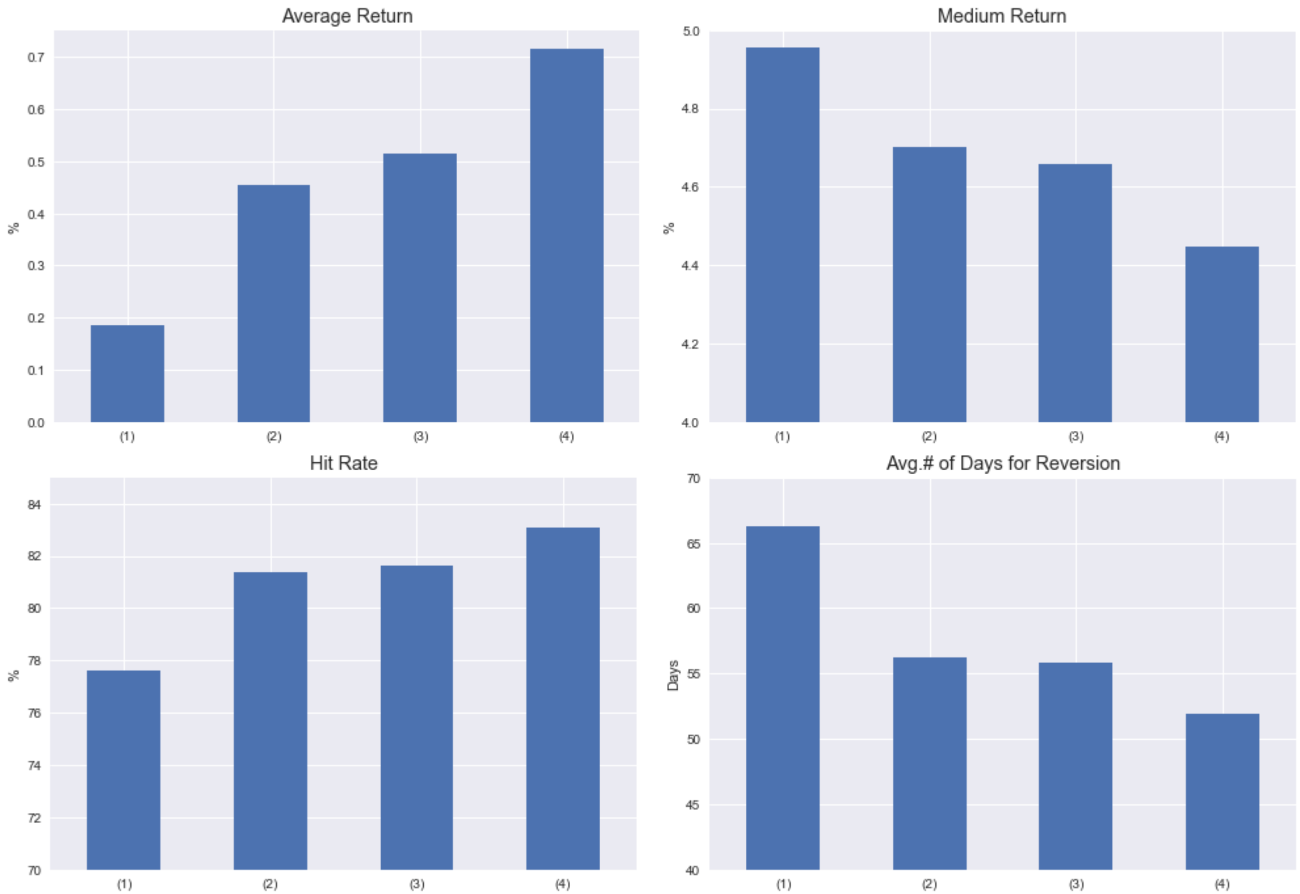
Interval: 2003 to 2015
Universe: Russell 3000
Maximum Holding Period: 9 months
Stop loss: -60%
Transaction cost: 0%
Pairs are identified at each month end

Analysis

Backtesting Results Summary

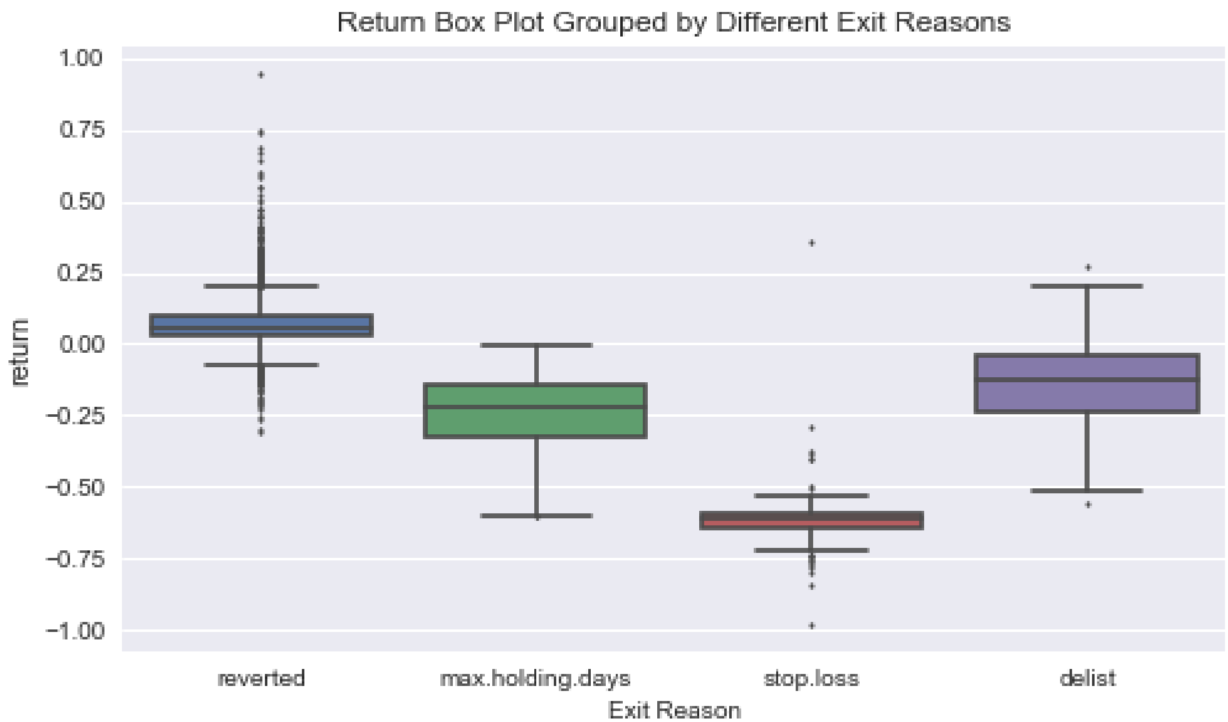
	(1)	(2)	(3)	(4)
Baseline Settings	Yes	Yes	Yes	Yes
Half-life Condition	/	Yes	Yes	Yes
Hurst Exponent	/	/	< 0.5	< 0.5
Johansen Cointegration Test	/	/	/	Top 5%
Num. of Pairs Traded per Month	218.70	134.67	125.06	43.24
Avg. Return	0.185%	0.455%	0.515%*	0.716%*
Med. Return	4.957%	4.703%***	4.659%***	4.447%***
Hit Rate	77.60%	81.37%***	81.64%***	83.06%***
Avg. Holding Period for Reverted Pairs	66.29	56.19***	55.82***	51.90***

Analysis



- Introducing more Time-series statistics tends to improve the average return and the hit rate, and reduces the number of days it takes for a pair to close

Analysis



Return Box Plot Grouped by Different Exit Reasons
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
return
reverted
max.holding.days
stop.loss
delist
Exit Reason

Next Step

- **Next step:**
  - Incorporate fundamental information
    - Customer-client relationship
    - Industry exposure
  - Eliminate negative impact of news/big corporate events
  - Explore the driving factors behind pair performance
    - Machine learning is likely to help
  - Form a portfolio

WOLFE RESEARCH

Q&A