



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich



Institut für  
Technische Informatik und  
Kommunikationsnetze

# Make Us Richer!

Bachelor Thesis  
March 8, 2013

**Severin Wischmann**  
wiseveri@student.ethz.ch

**Supervisor:**

Philipp Brandes  
philipp.brandes@tik.ee.ethz.ch

**Professor:**

Prof. Dr. Roger Wattenhofer  
wattenhofer@tik.ee.ethz.ch

Distributed Computing Group – DCG  
Computer Engineering and Networks Laboratory – TIK  
Department of Information Technology and Electrical Engineering – ITET  
Swiss Federal Institute of Technology – ETH



---

## **Abstract**

In the area of algorithmic trading, high-frequency trading (HFT) was the main focus of research and industry. In recent years major incidents with those algorithms, like the Knight Capital disaster, led to political debates about regulations, which would destroy that business model. Therefore systematic algorithmic trading got back into focus.

This thesis extends the previously developed framework by Thomas Buerli with strategies that use dependencies between stocks to make more profitable and reliable decisions as to when to buy and sell. Those strategies were profitably evaluated on historical data of half a year, which indicates that traditional algorithmic trading leads to profits, too.



---

---

## Contents

---

<b>1. Introduction</b>	<b>1</b>
1.1. Motivation . . . . .	1
1.2. Related Work . . . . .	2
1.3. Contributions . . . . .	2
<b>2. Background</b>	<b>3</b>
2.1. Stock Market . . . . .	3
2.2. Statistical Tools . . . . .	3
2.2.1. Linear Regression . . . . .	4
2.2.2. Bollinger Bands . . . . .	4
2.2.3. Augmented Dickey-Fuller Test . . . . .	6
2.2.4. Cointegration . . . . .	6
<b>3. Implementation</b>	<b>9</b>
3.1. The framework . . . . .	9
3.2. Complex Strategies . . . . .	9
3.2.1. Linear Regression . . . . .	10
3.2.2. Cointegration . . . . .	11
<b>4. Analysis and Evaluation</b>	<b>13</b>
4.1. Assumptions . . . . .	13
4.2. Testing Set . . . . .	13
4.3. Measures . . . . .	13
4.3.1. Mean-Variance Measures . . . . .	14
4.3.2. Profitability Measures . . . . .	14
4.3.3. Risk Measures . . . . .	15
4.4. Analysis . . . . .	15
4.4.1. SimpleLinReg . . . . .	15

4.4.2. Cointegration Agents . . . . .	15
4.5. Evaluation . . . . .	17
<b>5. Conclusion</b>	<b>21</b>
5.1. Outlook . . . . .	21
<b>A. Appendix</b>	<b>23</b>
A.1. Installation . . . . .	23
A.2. How To Use . . . . .	23
A.2.1. CLI . . . . .	24
A.2.2. iPython . . . . .	24
A.3. Statistics . . . . .	24
A.4. Original Problem . . . . .	33

---

---

## List of Figures

---

2.1. Linear approximation of data points . . . . .	4
2.2. Google Stockprices and Linear Combination . . . . .	5
2.3. Bollinger bands give an upper and a lower bound to asset prices[dailyfx.com, 2012] . . . . .	5
2.4. Cointegrated time series are tied together by asset prices . . . . .	7
3.1. Structure of a strategy[Bürli, 2012] . . . . .	10
4.1. Skewness moves the mean[Wikipedia, 2012] . . . . .	14
4.2. Assets of <i>SimpleLinReg</i> <sub>15,3</sub> and the reference index . . . . .	16
4.3. Weekly Log and Excess Return of <i>SimpleLinreg</i> <sub>17,3</sub> . . . . .	19

---

---

## List of Tables

---

4.1. The best Linear Regression Strategies . . . . .	17
4.2. The best Cointegration Strategies . . . . .	18
4.3. Analysed Strategies on new Data . . . . .	18
A.1. Complete Linear Regression Analysis . . . . .	25
A.1. Complete Linear Regression Analysis . . . . .	26
A.1. Complete Linear Regression Analysis . . . . .	27
A.2. Complete SimpleCointMiM analysis . . . . .	29
A.3. Complete Analysis of SimpleCoint . . . . .	30
A.4. Complete SimpleCointBoll Analysis . . . . .	31
A.5. Complete Evaluation data . . . . .	32



---

# CHAPTER 1

---

## Introduction

---

Algorithmic trading has seen enormous growth and is widely used by investment banks, pension funds, market makers, and hedge funds. It is responsible for more than 70% of the trades in the United States. Out of these trades, High-Frequency Trading (HFT) covers about half of all trades. Profits by these algorithms peaked at 4.9 billion in 2009[Times, 2012]. These algorithms make thousands of trades a day, holding shares just for a very short amount of time and profit mainly from slower investors.

### 1.1. Motivation

The downside to these HFT algorithms is that their complexity cannot be grasped anymore by traders. Small mistakes lead to big losses. The "flash crash" in May 6 2010, when the Dow Jones dropped over 1000 points during one day[Lauricella and STRASBURG, 2010], or when Knight Capital lost over 400 Million in August 2012 in only 45 minutes, because a HFT algorithm went awry, are examples of this. That is why HFT algorithms became the target of financial regulations[BOWLEY, 2011].

Systematic algorithmic trading on the other hand is not regulated and can be freely used to trade with stocks and their derivated products. With mathematical models of the market, algorithmic trading tries to translate historical data to future trading advice. It tries to do this with a maximum of profitability and a minimum of risk.

Systematic trading uses historical data to base its decision on, compared to fundamental trading for example which bases its decision on business health, their competitors, their markets, their management, and their competitive advantage.

## 1.2. Related Work

Algorithmic Trading is a widely discussed topic and lots of work has been done in the field. General models of the stock market can be found in Alexander [2001]. Autocorrelation, cointegration, and other market models used by investment analysts are explained. The mathematics used in econometrics is explained by Vince [1992]. The book covers many practices used by professional money managers, such as risk management and modern portfolio theory.

Some applications of linear regression in financial models are described in Christian L. Dunis [2003], such as modeling a time series as a linear model or interpreting it as neural networks with no hidden layers. João F. Caldeira [2012] and Schmidt [2008] use cointegration in pairs trading. They both estimate long-run equilibria and model the mean-reverting residuals, but the latter shows that cointegrated stocks can be linearly combined such that the resulting portfolio is governed by a stationary process.

## 1.3. Contributions

This bachelor thesis uses different market models to define, implement and evaluate strategies that generate trading orders for the stock market.

---

## CHAPTER 2

---

### Background

---

This chapter provides the reader with the information about the stock market and the mathematical tools that were used to understand this thesis.

#### **2.1. Stock Market**

The stock or equity market is a place where shares of companies are issued and traded. The new shares are first offered in the primary market on an exchange. For companies this is one of the most important sources for money. The secondary markets, as the New York Stock Exchange, Swiss Exchange or NASDAQ, deal with trades, where investors buy assets from other investors rather than from the issuing companies themselves. Another possibility are over-the-counter trades, where trades occur between two parties without supervision from e.g. a stock exchange.

Stock exchanges provide traders with information, such as opening and closing prices, the prices at the beginning and at the end of a trading day. The opening price of a stock does not have to be the same as the closing price of the day before. This is due to after-hour trading or changing expectations of investors. Stock exchanges also publish the traded volume and the lowest and the highest price of each stock that is traded on their exchange.

#### **2.2. Statistical Tools**

To model the stock market various approaches have been used, each based on different statistical properties of time series.

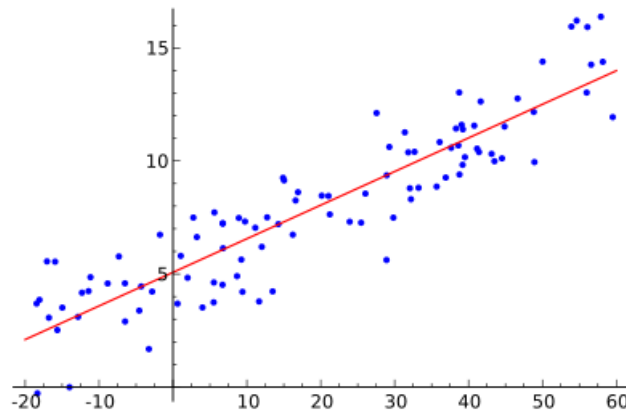


Figure 2.1.: Linear approximation of data points

### 2.2.1. Linear Regression

Linear regression finds the best fitting linear function for a set of data points. Given a set of points  $(\mathbf{x}_i, y_i)$ <sup>1</sup> estimate the parameters  $\mathbf{a}$ ,  $c$  such that  $y = \mathbf{a}^T \mathbf{x} + c$ . This can be generalized to explaining a dependant variable  $y$  with one or multiple explanatory variables  $X$  ( $X$  is a matrix), such that  $\mathbf{y} = \alpha + X\beta$ .

In general, the least squares approach is used to estimate the parameters  $\beta$

$$S(\mathbf{b}) = \sum_{i=1}^n (y_i - \mathbf{x}_i \cdot \mathbf{b}^T)^2 = (\mathbf{y} - X\mathbf{b})^T (\mathbf{y} - X\mathbf{b}) \quad (2.1)$$

where  $\mathbf{b}$  is the estimator for  $\beta$ . Now we have to find the minimum

$$\hat{\beta} = \arg \min S(\mathbf{b}) = \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \cdot \mathbf{x}_i^T \right)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \cdot y_i = (X^T X)^{-1} X^T \mathbf{y}$$

which always exists [Hayashi, 2000]. As a measure of how good the estimation fits, the value  $r^2$  is used.

As an example choosing the asset price of Google as the dependent variable and the asset prices of Microsoft and Yahoo as the explanatory variables and estimating  $\beta$  and  $\alpha$  results in the picture in Figure 2.2. The values of the estimated dependent variable are calculated day by day using the estimated values of  $\beta$  and  $\alpha$  and the daily data of the explanatory variables.

### 2.2.2. Bollinger Bands

Bollinger Bands [Bollinger, 2011] are a technical analysis tool for financial data. They provide a relative measure for high and low prices and are used in pattern recognition. They are defined as

---

<sup>1</sup>bold letters represent a vector

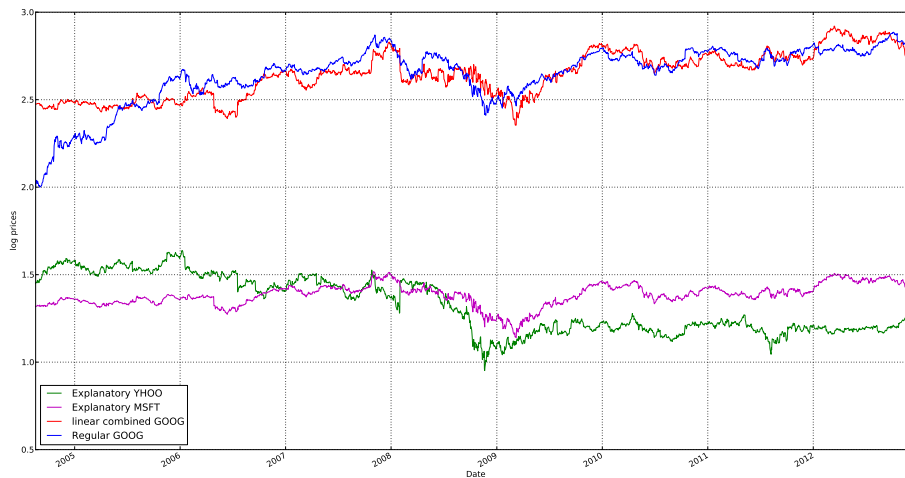


Figure 2.2.: Google Stockprices and Linear Combination

- **middle** - moving average over N-periods
- **lower** - middle band minus K times an N-period standard deviation
- **upper** - middle band plus K times an N-period standard deviation

Typically N and K are chosen as 20 and 2, respectively. Typically the simple moving average is used, but other moving averages can be used too. The exponential moving average is another common choice.

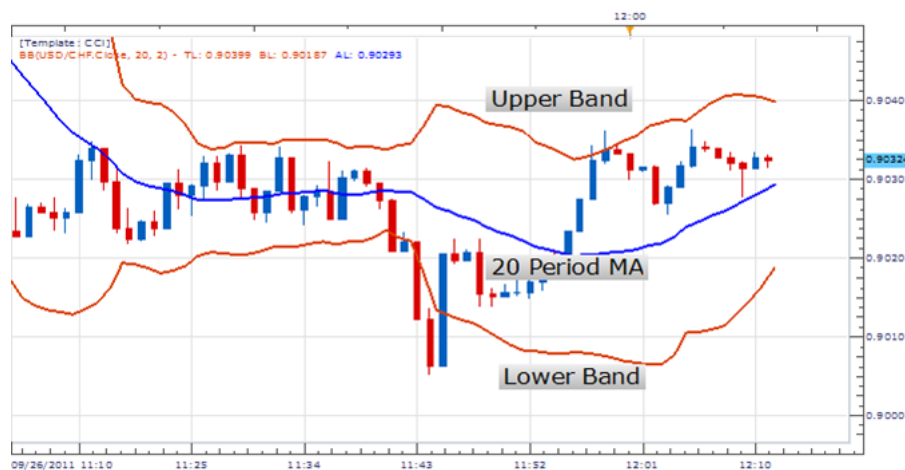


Figure 2.3.: Bollinger bands give an upper and a lower bound to asset prices[dailyfx.com, 2012]

### 2.2.3. Augmented Dickey-Fuller Test

The Augmented Dickey-Fuller Test (ADF) is used in the test for cointegration. ADF is a unit root test for stationarity.

An autoregressive process of the form

$$\begin{aligned}y_t &= \gamma y_{t-1} + \delta y_{t-2} + \varepsilon_t \\y_t &= (\gamma + \delta)y_{t-1} - \delta(y_{t-1} - y_{t-2}) + \varepsilon_t\end{aligned}\tag{2.2}$$

is called stationary if in the long run  $y_t$  reaches a fixed value or has a deterministic trend. In more mathematical terms, a process is stationary if its statistical properties, such as mean and variance, do not change over time.

Subtracting  $y_{t-1}$  from both sides of Equation (2.2) results in

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \varepsilon_t\tag{2.3}$$

where  $\phi_1 = \gamma + \delta - 1$  and  $\phi_2 = -\delta$ . To test whether or not Equation (2.2) is stationary, we test if 1 is among the roots of Equation (2.3). This is called testing for an unit root.

The null hypothesis is that a unit root exists  $H_0 : \phi_1 = 0$  and is tested against the alternative hypothesis that no such root exists  $H_1 : \phi_1 < 0$ . After calculating the test statistic it is compared to precomputed and well known critical values, the 't'-statistics of the  $\phi_1$  coefficient [David A. Dickey, 1979]. If the test statistic is smaller than the critical value,  $H_0$  is rejected and autoregressive process is stationary.

### 2.2.4. Cointegration

Cointegration [Robert F. Engle, 1987] is used to model the co-movements of asset prices that are tied together in the long run by a common stochastic trend. Looking at the asset prices of the NYSE Euronext and comparing it to Devon Energy Corporation and the General Mills Inc as an example, Figure 2.4 shows that they are closely tied to the former and move differently than the latter. Cointegration allows to reflect the complex dependencies in the stock market much better than autocorrelation and should therefore be used when modeling a stock market [Alexander, 2001].

#### Testing for Cointegration

To test for cointegration the Engle-Granger Cointegration test was used. The cointegration test includes two steps: First the data is estimated using ordinary least squares (OLS) and then the residuals, the errors between the actual data points and the estimation, are tested for stationarity, e.g. using ADF. If they are not stationary, the time series are not cointegrated.

For the test to be conclusive a sufficiently large timeframe has to be considered as cointegration is meant to reveal long running trends in the variables.

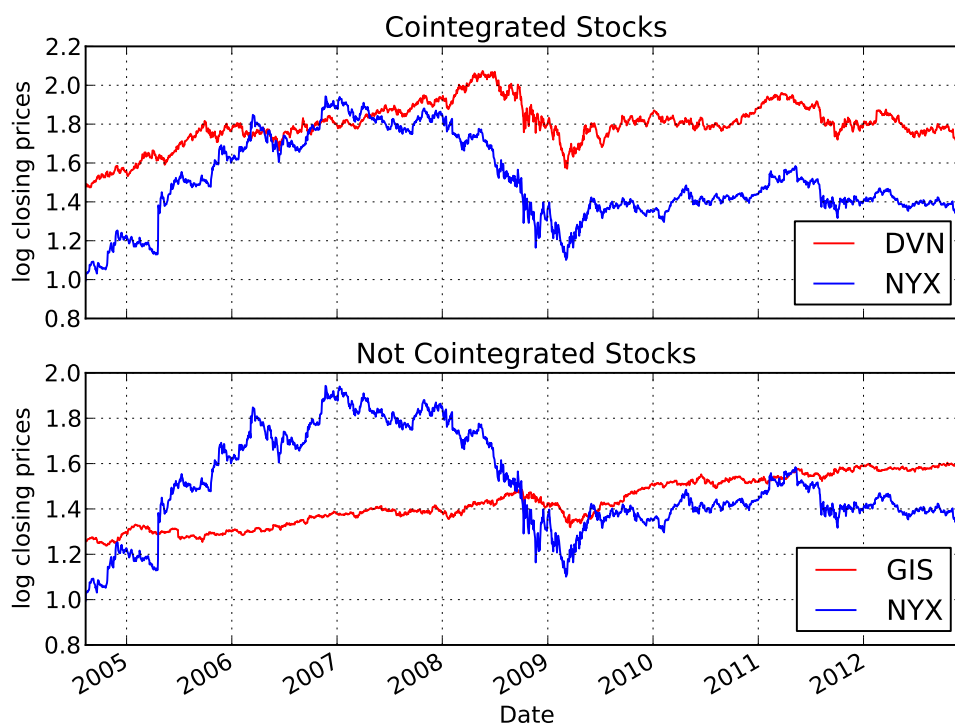


Figure 2.4.: Cointegrated time series are tied together by asset prices





---

## CHAPTER 3

---

### Implementation

---

This thesis tried to model the stock market using the relationships of the traded stocks and implement trading algorithms that use the information gained to make informed decisions. To evaluate the strategies some parameters were introduced to the given framework, such as a balance.

This chapter describes the implementation of the strategies and the additions to the framework developed by Thomas Bürli[Bürli, 2012].

#### 3.1. The framework

The framework consist of four layers relevant to implementing a strategy:

1. **Strategy** - A strategy uses a council to make its decision.
2. **Council** - A council gathers the advice of agents. The advice consists of a signal, which states whether to buy an asset, and the confidence, which states how certain the agent is about its decision.
3. **Agent** - An agent uses indicators to get signals whether to buy or sell an asset.
4. **Indicator** - Indicators apply statistical tools on historical data.

Figure 3.1 shows how a strategy uses the different layers.

#### 3.2. Complex Strategies

Two different models were used to generate strategies that decide whether to buy or sell a stock. The strategies use the above described layout of the framework.

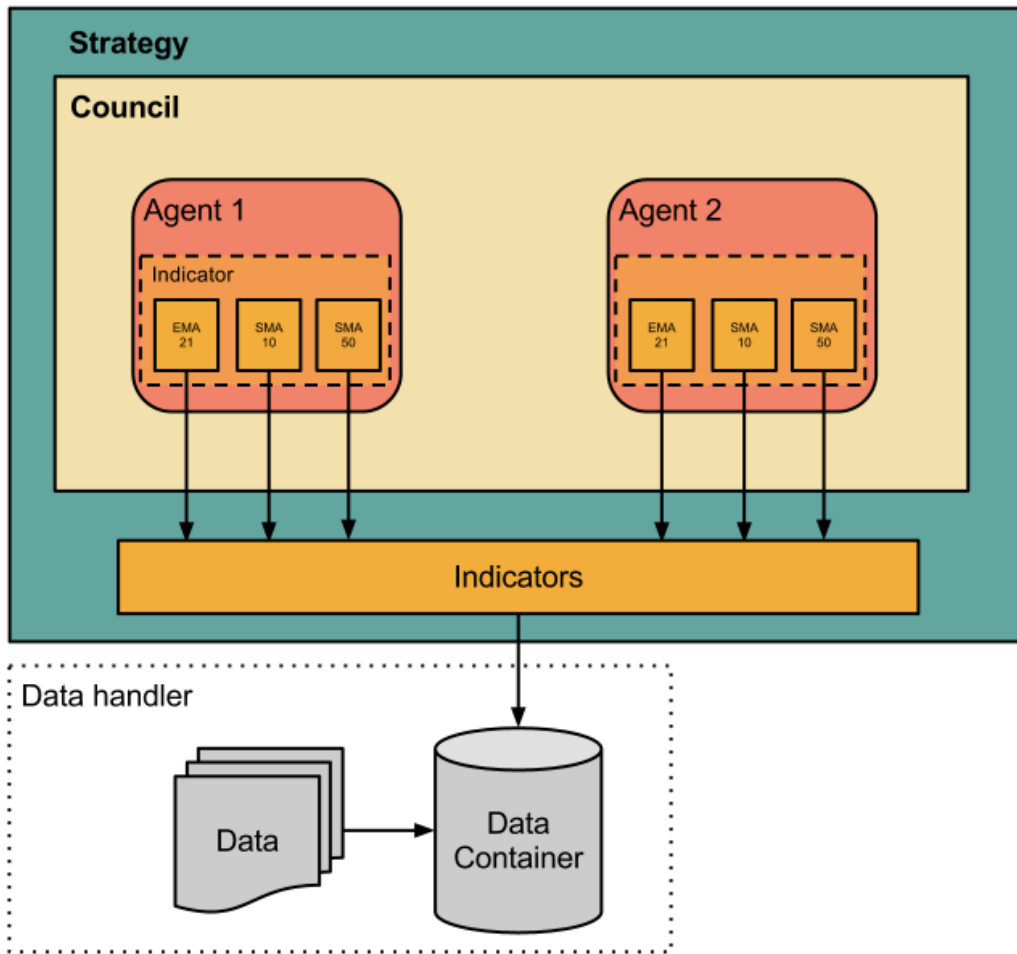


Figure 3.1.: Structure of a strategy[Bürli, 2012]

### 3.2.1. Linear Regression

For the linear regression strategy each stock is modeled as a linear combination of a random subset of the other stocks in the market.

#### SimpleLinReg

This agent calls the  $linreg_{x,y}$  indicator, where  $x$  is the window size for the rolling ordinary least squares and  $y$  the number of randomly chosen stocks to consider as explanatory variables. The window size determines how many days are considered when fitting the time series of the stock to be traded, the dependent variable, to the stocks used as explanatory variables. Therefore the OLS are calculated for each day separately using  $x$  days of data to do so.

The agent then buys a stock when yesterday's closing price is higher than today's predicted value. The confidence of this decision is based on the  $r^2$  value returned by the linear regression.

### 3.2.2. Cointegration

There are three cointegration agents implemented. They all work only on two stocks simultaneously and only trade if the asset prices are cointegrated. They differ though in the way they use the fact that the prices are tied together.

#### SimpleCoint

This simplecoint agent assumes that asset prices follow a long running trend and that positive and negative deviations from it are only a short term effect. Based on this assumption it is best to buy a stock when it is as close as possible to a negative peak and sell it when it is as close to the maximum peak deviation from its simple moving average.

For this some lower and upper threshold based on the maximum and minimum of the asset prices in the timeframe is introduced to serve as a boundary for when to buy and sell.

The fact that two stocks are cointegrated is used when no decision could be made based on the threshold values. In this case the signal of the other stock is used, if any exists, to make a decision. To mitigate the fact that only statistical properties of another time series are used, the confidence is set to only half of the one of the other signal.

#### SimpleCointBoll

This agent does the same thing as the agent described above, but uses Bollinger Bands as the thresholds. The agent buys stocks of an asset if the price hits the lower band and sells the stocks he is holding if the price hits the upper band.

Cointegration is again used when no individual decision could be made. The confidence again is set to half the confidence of the individual decision.

#### SimpleCointMiM

This agent assumes that cointegrated time series evolve like the averaged daily prices of the two time series  $asset_1$  and  $asset_2$ . Buy and sell signals are therefore based on the asset price being below or above the average and react accordingly:

- If e.g. the price of  $asset_1$  is below the average the price should be higher based on our assumption. Therefore we buy stocks of  $asset_1$ .
- If the price of e.g.  $asset_2$  is above the average the price should be lower. Therefore we sell our stocks of  $asset_2$ , if we hold any.

The confidence is based on changes in the signal, according to this formula:

$$confidence = \left| \frac{|signal_t + signal_{t-1}| - 2}{2} \right|$$



---

## CHAPTER 4

---

### Analysis and Evaluation

---

The evaluation of the performance of the implemented algorithms was done in two steps. First, a variety of different parameters was tested on a testing set and the performance of the algorithms was analysed on different parameters. Then the best configurations for each algorithm were selected to perform the evaluation on the new data set.

#### 4.1. Assumptions

To evaluate the algorithms some assumptions have been made regarding the market, namely that trading does not impact the market, therefore trades cannot affect the quoted prices. The algorithms also only engage in long strategies, meaning that they follow the pattern buy-hold-sell when trading stocks.

#### 4.2. Testing Set

The analysis and evaluation of the algorithms strongly depends on the testing set they are tested on. That is why the common Standard and Poor 500 (S&P 500) index. Stocks in the index are chosen based on size, liquidity and industry grouping, among other factors. It is the most commonly used index to benchmark the U.S. stock market[Investopedia, 2012]. The index used in this thesis is based on the data from mid february.

#### 4.3. Measures

To analyse the implemented algorithms and tested parameters three types of measures were used: Mean-variance measures, profitability measures, and risk measures.

### 4.3.1. Mean-Variance Measures

The mean-variance measures are used for the statistical analysis of the asset-returns.

#### Mean

The mean is the first moment of a distribution and describes its central tendency. The mean is directly related to the profitability of the algorithms, therefore a high mean is preferred.

#### Standard Deviation

The standard deviation is a measure for risk and uncertainty of an algorithm, as it tells us how much it tends to deviate from the expected return. It is calculated as the square root of the variance, which measures the deviation of the data points from the mean.

#### Skewness

The third moment of a distribution measures the amount of asymmetry of a distribution. Negative skewness leads to a higher mean as in Figure 4.1 and therefore to more profit, but Holton [2003] argues that a negative skewness leads to more risk.

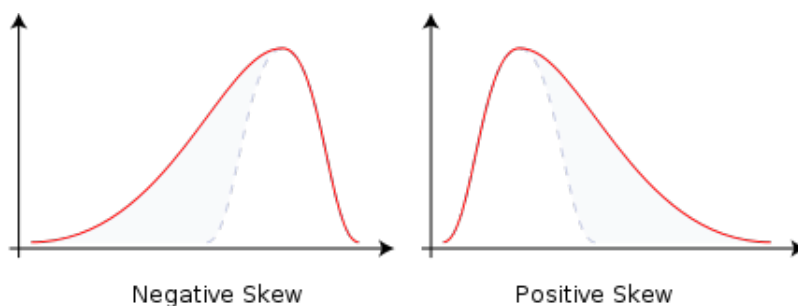


Figure 4.1.: Skewness moves the mean[Wikipedia, 2012]

#### Kurtosis

Kurtosis describes the "peakedness" relative to the normal distribution. The standard normal distribution has a value of three[Brown, 2011]. A kurtosis smaller than the normal distribution will lead to a flatter curve, which means that values close to the mean are more likely to occur, and values bigger than the normal distribution will lead to a more peaked curve, which means that the probability of values close to the mean drop more rapidly. A strategy with a negative kurtosis is therefore more stable around the mean.

### 4.3.2. Profitability Measures

The profitability is measured using the logarithmic and excess returns. The logarithmic return is calculated with

$$return = \ln(return).$$

The excess return is the difference of the monthly return of the asset and the return of the reference index (S&P 500):

$$excess\_return = asset\_return - index\_return$$

### 4.3.3. Risk Measures

As a measure of risk, the monthly and yearly volatility is used. The volatility describes how much the values of an asset change over a period of time. A strategy with a high volatility is therefore riskier than a strategy with a smaller one.

To put this number into perspective we compare it to the return that an algorithm produces. For this the Risk Return Ratio (RRR) is used. It is defined as:

$$RRR = \frac{1 - \frac{return_{max} - return}{return_{max}}}{1 - \frac{risk_{min} - risk}{risk_{min}}}$$

An ideal strategy has a RRR of 1, having the highest return and the lowest risk associated with it.

## 4.4. Analysis

The strategies described in Section 3.2 were analysed over the period of three and a half years starting on January 1, 2009 and ending on June 1, 2012. This period was chosen to exclude the year of the financial crisis, which occurred in the year 2008. The complete analysis with all the parameters chosen can be found in Appendix A.3.

### 4.4.1. SimpleLinReg

The linear regression strategy was analysed with different window sizes for the rolling ordinary least squares method and different number of dependent variables, which were chosen randomly. Due to the random nature of the selection, each configuration was ran 30 times and the evaluation then used the average values of all these runs. Because this generates a huge number of iterations when run on the S&P 500, which would not have finished in time the analysis was performed on the S&P 100 only.

In Table 4.1 the ten best strategies have been listed. As one can see *SimpleLinReg*<sub>15,3</sub> has the best RRR combined with the skewness closest to zero. The still very small RRR can be explained with the very small volatility some less performant strategies produced.

The strategies *SimpleLinReg*<sub>15,3</sub>, *SimpleLinReg*<sub>17,3</sub>, *SimpleLinReg*<sub>19,3</sub> and *SimpleLinReg*<sub>27,5</sub> have been chosen for evaluation.

### 4.4.2. Cointegration Agents

The cointegration agents all trade only two stocks simultaneously. To analyse the performance of these agents, 300 unique pairs of stocks were chosen randomly at first and

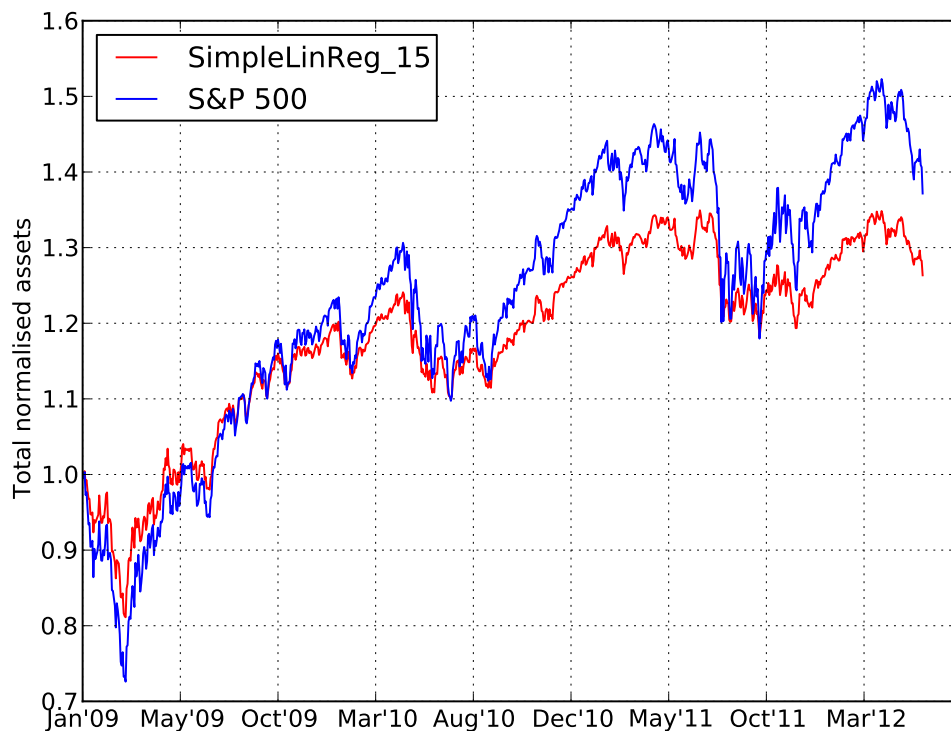


Figure 4.2.: Assets of *SimpleLinReg*<sub>15,3</sub> and the reference index

then from the set of stocks that are cointegrated over the complete timeframe the data is available. Both approaches generated results for only about 1% of the pairs, as we do not trade if the assets are not cointegrated. Therefore averaging the results does not provide significant results, but gives only an idea about the performance of the implemented strategies.

### SimpleCoint

The performance of this agent was analysed with different thresholds  $t$ . The upper and lower threshold was chosen relative to the simple moving average (SMA) as  $(1+t) \cdot SMA$  and  $(1-t) \cdot SMA$  respectively. No value of  $t$  led to profit, but strategies *SimpleCoint*<sub>0,1</sub> and *SimpleCoint*<sub>0,2</sub> had the best combination of RRR, skewness and kurtosis, so they were chosen for analysis on the new data. The complete analysis can be found in Table A.3.

### SimpleCointBoll

The bollinger bands have two parameters, the number of periods  $N$  and the number of standard deviations  $K$ . To analyse this agent,  $N$  was fixed at the typical value and  $K$  was varied. Table A.4 shows the complete data from the analysis. Some exponential profits as well as losses were encountered which on average result in the numbers shown. For anal-



name	return (%)	$\bar{x}$ ( $10^{-5}$ )	$\sigma$ ( $10^{-3}$ )	skew.	kur.	vol. yearly	RRR
S&P 500	37.15	36.74	13.9	-0.24	3.13	0.097	0.049
<i>SimpleLinReg</i> <sub>15,3</sub>	26.35	26.99	9.10	-0.23	3.85	0.088	0.038
<i>SimpleLinReg</i> <sub>15,5</sub>	16.14	17.31	7.58	-0.28	3.92	0.068	0.030
<i>SimpleLinReg</i> <sub>17,3</sub>	25.59	26.17	9.17	-0.20	3.94	0.09	0.036
<i>SimpleLinReg</i> <sub>19,3</sub>	21.86	22.76	9.27	-0.29	3.84	0.069	0.041
<i>SimpleLinReg</i> <sub>21,3</sub>	19.46	20.51	9.44	-0.26	4.23	0.075	0.033
<i>SimpleLinReg</i> <sub>21,5</sub>	12.50	13.61	8.13	-0.35	4.1	0.058	0.027
<i>SimpleLinReg</i> <sub>25,3</sub>	20.96	21.74	9.75	-0.31	3.98	0.07	0.038
<i>SimpleLinReg</i> <sub>25,5</sub>	11.78	12.84	8.12	-0.35	3.99	0.052	0.029
<i>SimpleLinReg</i> <sub>27,3</sub>	19.95	21.02	9.69	-0.29	4.06	0.067	0.038
<i>SimpleLinReg</i> <sub>27,5</sub>	13.01	14.15	8.25	-0.28	4.5	0.049	0.034

Table 4.1.: The best Linear Regression Strategies

ysis the strategies *SimpleCointBoll*<sub>1,0</sub> and *SimpleCointBoll*<sub>1,5</sub> were chosen. This was based on a first run of analysis which produced astronomically big numbers. The values were appended to the table in the appendix.

### SimpleCointMiM

As this agent relies on the fact, that cointegrated assets move as their average, strongly tied asset prices are required. Therefore it is analysed how many days need to be considered when testing for cointegration, to get a meaningful result. Due to lots of losses occurring during testing, an agent was implemented and evaluated that inverts the signals of the regular SimpleCointMiM agent. The results are shown in Table A.2. The RRR could not be calculated for the Meet-in-the-Middle strategies as the framework did not provide any yearly volatility measures. The reason for this could not be determined.

For the evaluation the strategies *SimpleCointMiM*<sub>180</sub>, *SimpleCointMiM*<sub>1260</sub>, *SimpleCointMiMInv*<sub>180</sub>, and *SimpleCointMiMInv*<sub>1080</sub> were chosen.

## 4.5. Evaluation

The selected strategies from Section 4.4 are evaluated on new data, to see if the parameters that performed best on the reference data still perform well on new data. The timeframe is chosen from June 1, 2012 until December 17, 2012. For the linear regression strategies again the S%P 100 index was used for evaluation. The cointegration strategies used 500 randomly selected pairs of stocks that are cointegrated in the timeframe from January 2, 2001 until December 17, 2012.

Even though the *SimpleCoint* agents used the same timeframe as the *SimpleCointBoll* agents, the *SimpleCoint*<sub>0,1</sub> agent did not find any cointegrated pair of assets among his set and the *SimpleCoint*<sub>0,2</sub> agent found only one. In comparison, the *SimpleCointBoll* agents both found more than 50 pairs to be cointegrated.

#### 4.5. Evaluation

name	return (%)	$\bar{x}$ ( $10^{-5}$ )	$\sigma$ ( $10^{-3}$ )	skew.	kur.	vol. yearly	RRR
S&P 500	37.15	36.74	13.9	-0.24	3.13	0.097	0.85
<i>SimpleCoint</i> <sub>0,1</sub>	-5.06	-83.76	5.39	-1.34	6.33	0.083	-0.14
<i>SimpleCoint</i> <sub>0,2</sub>	-10.31	-29.2	599	-0.23	5.34	0.21	-0.11
<i>SimpleCointBoll</i> <sub>1,0</sub>	-0.72	-259	9.43	-1.87	8.95	0.38	-0.026
<i>SimpleCointBoll</i> <sub>1,5</sub> -0.63	-162	8.12	-2.60	13.94	0.27	-0.032	
<i>SimpleCointMim</i> <sub>180</sub>	-1.48	-344	6.24	-0.46	0.34	0.0	0.0
<i>SimpleCointMim</i> <sub>900</sub>	-4.51	562	20.2	-0.22	0.0054	0.0	0.0
<i>SimpleCointMim</i> <sub>1260</sub>	-4.37	49.02	18.04	-0.26	-0.53	0.0	0.0
<i>SimpleCointMim</i> <sub>1620</sub>	-1.85	-315	6.46	-0.37	0.63	0.0	0.0
<i>SimpleCointMimInv</i> <sub>180</sub>	9.13	-7.23	32.71	-0.42	1.12	0.0	0.0
<i>SimpleCointMimInv</i> <sub>540</sub>	-7.29	-879	30.19	-0.49	1.25	0.0	0.0
<i>SimpleCointMimInv</i> <sub>1080</sub>	1.66	543	13.55	0.82	0.0	0.0	0.0
<i>SimpleCointMimInv</i> <sub>1440</sub>	-8.66	-1057	17.72	-0.63	1.07	0.0	0.0

Table 4.2.: The best Cointegration Strategies

The Table 4.3 shows the results of the evaluation on the new data. The RRR has been calculated with the monthly volatility this time, as the timeframe is much shorter. The complete evaluation can be found in Table A.5.

Even though the return of the *SimpleCointMimInv*<sub>1080</sub> strategy is the highest, this value is based on one very high value and not on a good average performance. Also the volatility can not be calculated, which does not allow a direct comparison with the other strategies. Based on the RRR, the overall best strategy is *SimpleLinReg*<sub>17,3</sub>. As can be seen in Figure 4.3 this strategy does not outperform the S&P 500 index.

name	return (%)	$\bar{x}$ ( $10^{-5}$ )	$\sigma$ ( $10^{-3}$ )	skew.	kur.	vol. monthly	RRR
S&P 500	11.92	82.79	8.23	0.14	1.12	0.03	0.13
<i>SimpleLinReg</i> <sub>15,3</sub>	7.29	51.69	5.83	-0.015	1.89	0.018	0.14
<i>SimpleLinReg</i> <sub>17,3</sub>	7.75	54.85	5.86	-0.058	1.84	0.017	0.15
<i>SimpleLinReg</i> <sub>19,3</sub>	7.72	54.60	5.98	-0.065	1.83	0.018	0.15
<i>SimpleLinReg</i> <sub>27,5</sub>	5.18	37.09	5.03	-0.15	2.52	0.013	0.13
<i>SimpleCoint</i> <sub>0,2</sub>	-21.02	-173	5.16	-1.37	1.91	0.058	-0.12
<i>SimpleCointBoll</i> <sub>1,0</sub>	-40	-559	13.47	-1.6	4.71	0.11	-0.12
<i>SimpleCointBoll</i> <sub>1,5</sub>	-30.45	-347	11.54	-2.35	9.85	0.086	-0.12
<i>SimpleCointMim</i> <sub>180</sub>	-4.32	-476	9.70	-0.39	0.35	0.0	0.0
<i>SimpleCointMim</i> <sub>1260</sub>	-6.22	-919	17.25	-0.5	1.22	0.0	0.0
<i>SimpleCointMimInv</i> <sub>180</sub>	-0.23	-208	4.14	-0.032	-0.21	0.0	0.0
<i>SimpleCointMimInv</i> <sub>1080</sub>	39.25	1115	59.92	-0.71	0.01	0.0	0.0

Table 4.3.: Analysed Strategies on new Data

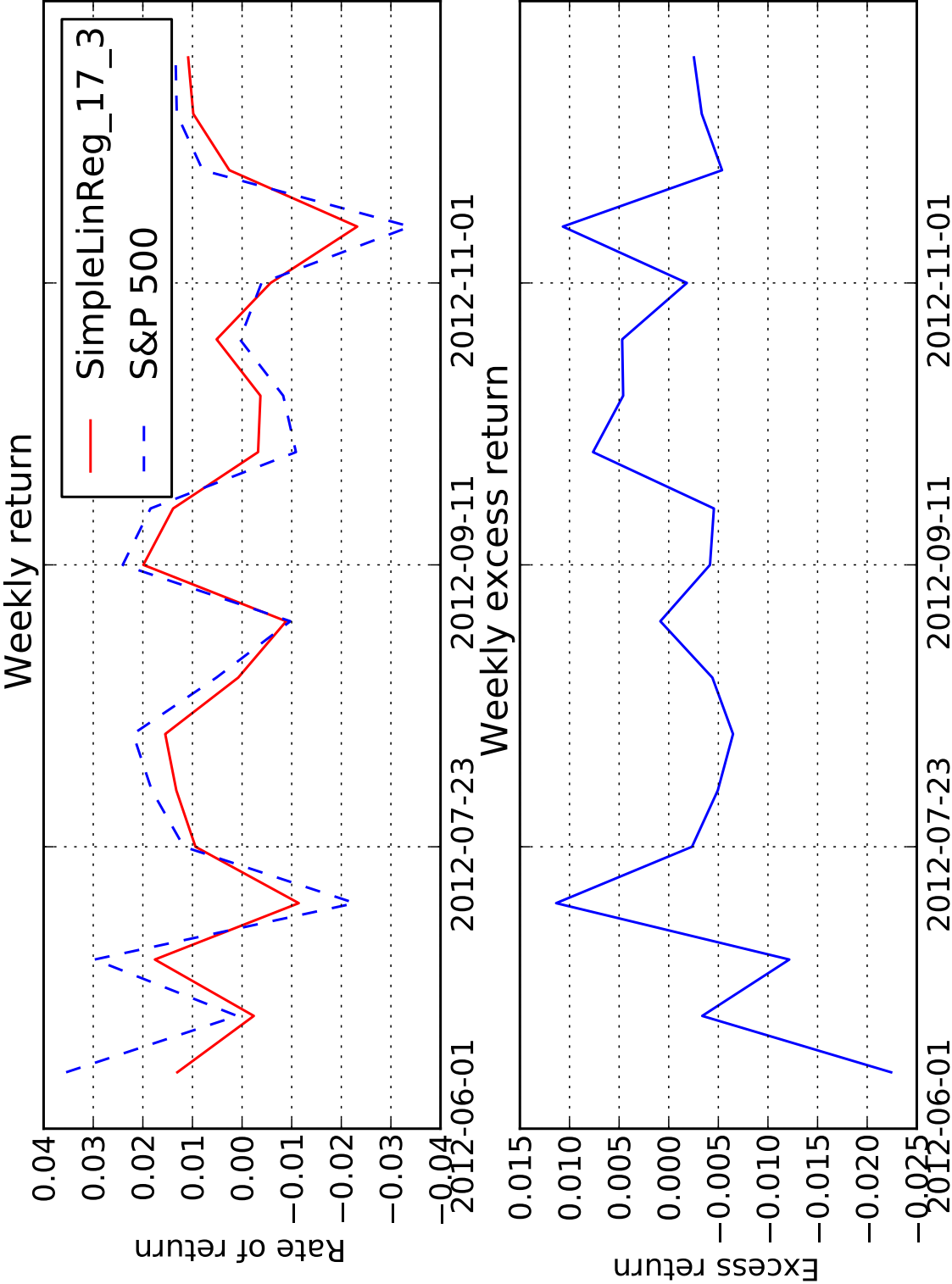


Figure 4.3.: Weekly Log and Excess Return of SimpleLinReg<sub>17,3</sub>



---

## CHAPTER 5

---

### Conclusion

---

The goal of this thesis was to make the already developed framework more realistic by introducing transaction costs and budget constraints and also to implement more complex market models to make trading decisions. Evaluating these strategies led to the conclusion that something something something

#### **5.1. Outlook**

Even though the framework was improved, there is still room for more.

It is still only possible to hold long positions, with which a certain risk is associated. Implementing short selling would therefore be a desirable addition, to also be able to profit from falling prices.

Further work might want to look into the Black-Scholes model[Fischer Black, 1973] as it implies that there is only one right price for an option. Another direction would be implementing other artificial intelligence techniques like artificial neural networks or support vector machines.



---

# APPENDIX A

---

## Appendix

---

### A.1. Installation

To use the framework some python packages need to be installed. The following command installs all the required software on an Ubuntu system using the package manager:

```
1 sudo apt-get install python python-numpy python-scipy python-matplotlib
2 python-dateutil
```

There are some additional python specific packages that are needed to be installed. This can be done with a python package manager, such as pip:

```
1 sudo apt-get install python-pip
```

With the package manager now the additional packages can be installed:

```
1 sudo pip install nose pandas pytz
```

With all this packages installed, the framework can now be used. For easier usage it is recommended to install iPython, an interactive python shell:

```
1 sudo apt-get install ipython-notebook
```

### A.2. How To Use

There are two different ways to use the framework. One way is to use the command line interface (CLI), the other is to use the iPython shell.

### A.2.1. CLI

The framework provides a command line interface that allows an easy evaluation of a strategy. The CLI offers a help menu, that explains all the options:

```
1 python makeUsRicher-cli.py -h
```

There are default values set for all the options except the stocks. The stocks can be set individually or load the S&P 100 or the S&P 500 directly.

```
1 python makeUsRicher-cli.py -t 'BHI' -t 'JNJ'  
2 python makeUsRicher-cli.py -t 'load_100'  
3 python makeUsRicher-cli.py -t 'load_all'
```

### A.2.2. iPython

To use the framework with iPython, run the following command in the python subfolder:

```
1 ipython run.py -i
```

In the interactive shell that opens, the following command executes the simulation and provides an object with the data for further analysis and plotting:

```
1 analysis = run('SimpleLinReg', ['BHI', 'JNJ'], '2012-01-01', '2012-06-01',  
2 100000, 0.2)  
3 analysis.plot_asset()  
4 analysis.plot_excess()  
5 plt.show()
```

This will simulate the SimpleLinReg strategy on the stocks of Baker Hughes Incorporation and Johnson & Johnson during the five months from January 1, 2012 to June 1, 2012 with a starting balance of 100000 and using 20% of the remaining money in each trade.

## A.3. Statistics

This chapter contains the complete data of all the simulations.



Table A.1.: Complete Linear Regression Analysis

name	return (%)	mean ( $10^{-5}$ )	var. ( $10^{-5}$ )	std. ( $10^{-3}$ )	skew.	kur.	vol. monthly	vol. yearly	RRR
S&P 500	37.15	36.74	19.35	13.9	-0.24	3.13	0.069	0.097	0.049
<i>SimpleLinReg</i> <sub>15,3</sub>	26.35	26.99	8.30	9.10	-0.23	3.85	0.047	0.088	0.038
<i>SimpleLinReg</i> <sub>15,5</sub>	16.14	17.31	5.76	7.58	-0.28	3.92	0.04	0.068	0.030
<i>SimpleLinReg</i> <sub>15,7</sub>	11.79	12.86	4.79	6.92	-0.25	4.35	0.036	0.06	0.024
<i>SimpleLinReg</i> <sub>15,9</sub>	9.74	10.74	3.99	6.31	-0.24	4.48	0.032	0.055	0.023
<i>SimpleLinReg</i> <sub>15,11</sub>	6.51	7.30	3.53	5.94	-0.25	4.41	0.031	0.043	0.019
<i>SimpleLinReg</i> <sub>15,13</sub>	6.61	7.41	3.02	5.49	-0.14	4.69	0.028	0.04	0.021
<i>SimpleLinReg</i> <sub>15,15</sub>	-0.17	-0.39	0.89	1.51	-0.05	2.15	0.0073	0.015	-0.0015
<i>SimpleLinReg</i> <sub>17,3</sub>	25.59	26.17	8.43	9.17	-0.20	3.94	0.049	0.09	0.036
<i>SimpleLinReg</i> <sub>17,5</sub>	15.17	16.3	6.09	7.8	-0.25	4.05	0.041	0.072	0.027
<i>SimpleLinReg</i> <sub>17,7</sub>	10.87	11.93	4.85	6.96	-0.26	4.09	0.036	0.059	0.024
<i>SimpleLinReg</i> <sub>17,9</sub>	9.08	9.99	4.19	6.47	-0.21	4.26	0.038	0.062	0.019
<i>SimpleLinReg</i> <sub>17,11</sub>	6.96	7.78	3.62	6.01	-0.24	4.13	0.031	0.05	0.018
<i>SimpleLinReg</i> <sub>17,13</sub>	6.67	7.47	3.29	5.73	-0.20	4.22	0.029	0.045	0.019
<i>SimpleLinReg</i> <sub>17,15</sub>	4.4	4.95	2.83	5.32	-0.19	4.9	0.027	0.037	0.015
<i>SimpleLinReg</i> <sub>17,17</sub>	0.73	0.656	1.92	3.34	0.05	6.68	0.017	0.028	0.0033
<i>SimpleLinReg</i> <sub>17,19</sub>	-0.48	-0.598	0.166	0.0325	-0.07	1.08	0.0015	0.0047	-0.013
<i>SimpleLinReg</i> <sub>19,3</sub>	21.86	22.76	8.61	9.27	-0.29	3.84	0.048	0.069	0.041
<i>SimpleLinReg</i> <sub>19,5</sub>	11.98	13.05	6.32	7.94	-0.33	4	0.041	0.064	0.024
<i>SimpleLinReg</i> <sub>19,7</sub>	8.05	8.95	4.96	7.04	-0.31	4.16	0.037	0.058	0.018
<i>SimpleLinReg</i> <sub>19,9</sub>	6.48	7.27	4.15	6.44	-0.3	4.15	0.034	0.056	0.015
<i>SimpleLinReg</i> <sub>19,11</sub>	5.57	6.23	3.65	6.04	-0.24	4.43	0.031	0.049	0.014
<i>SimpleLinReg</i> <sub>19,13</sub>	4.56	5.15	3.30	5.74	-0.21	4.41	0.03	0.044	0.013
<i>SimpleLinReg</i> <sub>19,15</sub>	4.90	5.52	3.11	5.57	-0.19	4.70	0.028	0.043	0.015
<i>SimpleLinReg</i> <sub>19,17</sub>	3.77	4.27	2.8	5.28	-0.20	4.78	0.027	0.037	0.013
<i>SimpleLinReg</i> <sub>19,19</sub>	0.24	0.0565	1.48	2.7	0.01	3.98	0.014	0.031	0.00099

Table A.1.: Complete Linear Regression Analysis

name	return (%)	mean ( $10^{-5}$ )	var. ( $10^{-5}$ )	std. ( $10^{-3}$ )	skew.	kur.	vol. monthly	vol. yearly	RRR
<i>SimpleLinReg</i> <sub>21,3</sub>	19.46	20.51	8.93	9.44	-0.26	4.23	0.048	0.075	0.033
<i>SimpleLinReg</i> <sub>21,5</sub>	12.50	13.61	6.63	8.13	-0.35	4.1	0.042	0.058	0.027
<i>SimpleLinReg</i> <sub>21,7</sub>	8.16	9.07	4.86	6.97	-0.31	3.72	0.036	0.052	0.02
<i>SimpleLinReg</i> <sub>21,9</sub>	4.69	5.27	4.21	6.49	-0.36	4.09	0.033	0.047	0.013
<i>SimpleLinReg</i> <sub>21,11</sub>	3.12	3.53	3.69	6.07	-0.35	4.33	0.032	0.043	0.0092
<i>SimpleLinReg</i> <sub>21,13</sub>	3.94	4.47	3.42	5.84	-0.28	4.56	0.030	0.046	0.011
<i>SimpleLinReg</i> <sub>21,15</sub>	3.71	4.21	3.14	5.6	-0.22	4.69	0.029	0.044	0.011
<i>SimpleLinReg</i> <sub>21,17</sub>	4	4.52	2.95	5.43	-0.20	4.75	0.028	0.038	0.014
<i>SimpleLinReg</i> <sub>21,19</sub>	3.6	4.06	2.74	5.23	-0.19	5.35	0.027	0.036	0.013
<i>SimpleLinReg</i> <sub>25,3</sub>	20.96	21.74	9.53	9.75	-0.31	3.98	0.049	0.07	0.038
<i>SimpleLinReg</i> <sub>25,5</sub>	11.78	12.84	6.62	8.12	-0.35	3.99	0.042	0.052	0.029
<i>SimpleLinReg</i> <sub>25,7</sub>	8.07	8.94	5.26	7.24	-0.29	4.24	0.038	0.048	0.022
<i>SimpleLinReg</i> <sub>25,9</sub>	4.31	4.87	4.35	6.59	-0.33	4	0.035	0.042	0.013
<i>SimpleLinReg</i> <sub>25,11</sub>	3.78	4.25	3.78	6.15	-0.26	4.19	0.032	0.042	0.012
<i>SimpleLinReg</i> <sub>25,13</sub>	3.04	3.44	3.42	5.84	-0.25	4.07	0.029	0.041	0.0095
<i>SimpleLinReg</i> <sub>25,15</sub>	3.61	4.08	3.2	5.65	-0.18	4.38	0.029	0.040	0.011
<i>SimpleLinReg</i> <sub>25,17</sub>	3.34	3.8	2.95	5.43	-0.16	4.36	0.028	0.037	0.011
<i>SimpleLinReg</i> <sub>25,19</sub>	2.31	2.63	2.84	5.34	-0.2	4.61	0.027	0.034	0.0088
<i>SimpleLinReg</i> <sub>27,3</sub>	19.95	21.02	9.4	9.69	-0.29	4.06	0.047	0.067	0.038
<i>SimpleLinReg</i> <sub>27,5</sub>	13.01	14.15	6.83	8.25	-0.28	4.5	0.043	0.049	0.034
<i>SimpleLinReg</i> <sub>27,7</sub>	7.25	8.01	5.21	7.21	-0.33	3.98	0.037	0.042	0.022
<i>SimpleLinReg</i> <sub>27,9</sub>	4.62	5.17	4.29	6.54	-0.30	3.90	0.034	0.040	0.015
<i>SimpleLinReg</i> <sub>27,11</sub>	4.42	4.99	3.86	6.21	-0.28	4.00	0.031	0.039	0.015
<i>SimpleLinReg</i> <sub>27,13</sub>	3.46	3.91	3.56	5.96	-0.25	4.21	0.030	0.038	0.012
<i>SimpleLinReg</i> <sub>27,15</sub>	3.20	3.59	3.28	5.73	-0.21	4.31	0.029	0.040	0.010
<i>SimpleLinReg</i> <sub>27,17</sub>	2.32	2.63	3.01	5.49	-0.16	4.23	0.028	0.037	0.0081

Table A.1.: Complete Linear Regression Analysis

name	return (%)	mean ( $10^{-5}$ )	var. ( $10^{-5}$ )	std. ( $10^{-3}$ )	skew.	kur.	vol. monthly	vol. yearly	RRR
<i>SimpleLinReg</i> <sub>27,19</sub>	3.21	3.65	2.88	5.36	-0.14	4.42	0.027	0.038	0.011



name	return (%)	mean ( $10^{-5}$ )	var. ( $10^{-5}$ )	std. ( $10^{-3}$ )	skew.	kur.	vol. monthly	vol. yearly	RRR
S&P 500	37.15	36.74	19.35	13.9	-0.24	3.13	0.069	0.097	1.0
<i>SimpleCoIntMim</i> <sub>180</sub>	-1.48	-344	10.35	6.24	-0.46	0.34	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>360</sub>	-8.65	-4748	898.87	67.04	0.0	0.0	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>720</sub>	-18.34	-380	412.00	45.39	-0.25	-1.17	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>900</sub>	-4.51	562	109.44	20.2	-0.22	0.0054	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>1080</sub>	-5.64	703	136.80	25.25	-0.27	0.0068	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>1260</sub>	-4.37	49.02	89.92	18.04	-0.26	-0.53	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>1440</sub>	-5.7	-485	35.28	9.17	-0.26	0.092	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>1620</sub>	-1.85	-315	22.08	6.46	-0.37	0.63	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>1800</sub>	-7.03	-926	69.4	16.31	-0.32	0.77	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>1980</sub>	-5.97	-862	63.72	15.9	-0.31	0.89	0.0	0.0	0.0
<i>SimpleCoIntMim</i> <sub>2160</sub>	-6.48	-1108	79.75	18.22	-0.28	0.66	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>180</sub>	9.13	-7.23	163	32.71	-0.42	1.12	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>360</sub>	-19.97	-1162	78.19	22.67	-1.13	2.95	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>540</sub>	-7.29	-879	91.65	30.19	-0.49	1.25	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>720</sub>	-9.42	131	49.11	19.04	0.78	1.21	0.00074	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>900</sub>	-11.59	-3258	553	67.32	-0.12	1.18	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>1080</sub>	1.66	543	36.74	13.55	0.82	0.0	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>1260</sub>	-18.56	-2335	220	40.39	-1.44	2.3	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>1440</sub>	-8.66	-1057	60.61	17.72	-0.63	1.07	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>1620</sub>	-9.05	-1604	112	21.14	-0.55	-0.15	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>1800</sub>	-10.54	-1488	100	21.69	-0.24	1.12	0.0	0.0	0.0
<i>SimpleCoIntMimInv</i> <sub>1980</sub>	-15.32	-1601	114	24.74	-0.61	1.87	0.015	0.18	-0.22
<i>SimpleCoIntMimInv</i> <sub>2160</sub>	-13.56	-1596	133	27.62	-0.59	1.83	0.013	0.16	-0.23

Table A.2.: Complete SimpleCoIntMim analysis

name	return (%)	mean ( $10^{-5}$ )	var. ( $10^{-5}$ )	std. ( $10^{-3}$ )	skew.	kur.	vol. monthly	vol. yearly	RRR
S&P 500	37.15	36.74	19.35	13.9	-0.24	3.13	0.069	0.097	0.85
<i>SimpleCoint</i> <sub>0,1</sub>	-5.06	-83.76	3.19	5.39	-1.34	6.33	0.044	0.083	-0.14
<i>SimpleCoint</i> <sub>0,2</sub>	-10.31	-29.2	5.19	599	-0.23	5.34	0.043	0.21	-0.11
<i>SimpleCoint</i> <sub>0,3</sub>	-14.13	-48.58	5.33	5.73	-2.44	34.69	0.054	0.21	-0.15
<i>SimpleCoint</i> <sub>0,4</sub>	-24.48	-117	10.05	8.64	-0.27	9.29	0.085	0.2	-0.28
<i>SimpleCoint</i> <sub>0,5</sub>	-24.93	-45.82	3.48	5.52	-0.51	16.66	0.054	0.20	-0.27

Table A.3.: Complete Analysis of SimpleCoint

name	return (%)	mean ( $10^{-5}$ )	var. ( $10^{-5}$ )	std. ( $10^{-3}$ )	skew.	kur.	vol. monthly	vol. yearly	RRR
S&P 500	37.15	36.74	19.35	13.9	-0.24	3.13	0.069	0.097	0.052
<i>SimpleCointBoll</i> <sub>1,0</sub>	-259	9.33	9.43	-1.87	8.95	0.071	0.38	-0.026	
<i>SimpleCointBoll</i> <sub>1,5</sub>	-162	6.86	8.12	-2.60	13.94	0.051	0.27	-0.032	
<i>SimpleCointBoll</i> <sub>2,0</sub>	452	-33.12	12.63	9.65	-0.92	30.28	0.066	0.25	0.24
<i>SimpleCointBoll</i> <sub>2,5</sub>	31.93	-1.65	6.44	6.61	-1.95	83.40	0.033	0.13	0.033
<i>SimpleCointBoll</i> <sub>3,0</sub>	-0.80	-4.46	3.64	0.0045	-6.25	310	0.019	0.062	-0.0018
<i>SimpleCointBoll</i> <sub>1,0</sub>	15343700	-34.16	31.12	14.46	-0.02	8.60	0.12	0.45	
<i>SimpleCointBoll</i> <sub>1,5</sub>	19600	-55.75	21.86	12.52	-0.56	15.41	0.093	0.37	

Table A.4.: Complete SimpleCointBoll Analysis

name	return (%)	mean ( $10^{-5}$ )	var. ( $10^{-5}$ )	std. ( $10^{-3}$ )	skew.	kur.	vol. weekly	vol. monthly	RRR
S&P 500	11.92	82.79	6.77	8.23	0.14	1.12	0.017	0.03	0.13
<i>SimpleLinReg</i> <sub>15,3</sub>	7.29	51.69	3.41	5.83	-0.015	1.89	0.012	0.018	0.14
<i>SimpleLinReg</i> <sub>17,3</sub>	7.75	54.85	3.46	5.86	-0.058	1.84	0.012	0.017	0.15
<i>SimpleLinReg</i> <sub>19,3</sub>	7.72	54.60	3.59	5.98	-0.065	1.83	0.012	0.018	0.15
<i>SimpleLinReg</i> <sub>27,5</sub>	5.18	37.09	2.54	5.03	-0.15	2.52	0.011	0.013	0.13
<i>SimpleCoim</i> <sub>0,2</sub>	-21.02	-173	2.66	5.16	-1.37	1.91	0.021	0.058	-0.12
<i>SimpleCoimBoll</i> <sub>1,0</sub>	-40	-559	25.04	13.47	-1.6	4.71	0.049	0.11	-0.12
<i>SimpleCoimBoll</i> <sub>1,5</sub>	-30.45	-347	18.86	11.54	-2.35	9.85	0.039	0.086	-0.12
<i>SimpleCoimMim</i> <sub>1,80</sub>	-4.32	-476	40.08	9.70	-0.39	0.35	0.018	0.0	0.0
<i>SimpleCoimMim</i> <sub>1,260</sub>	-6.22	-919	106	17.25	-0.5	1.22	0.00026	0.0	0.0
<i>SimpleCoimMimInv</i> <sub>1,80</sub>	-0.23	-208	3.48	4.14	-0.032	-0.21	0.002	0.0	0.0
<i>SimpleCoimMimInv</i> <sub>1,080</sub>	39.25	1115	1374	59.92	-0.71	0.01	0.0083	0.0	0.0

Table A.5.: Complete Evaluation data



## A.4. Original Problem

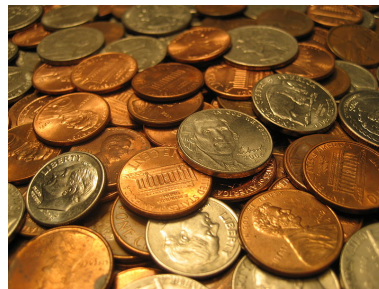
**Lab/BA/SA/Group:**

**Make Us Richer!**

### Motivation and Informal Description

In the recent past the algorithmic trading has seen enormous growth and a good place to make lots of money. It is now responsible for more than 70% the trades in the US. A very important subclass are the high frequency trading (HFT) algorithms. These algorithms usually hold stocks or certificates only for a brief time, sometimes only for a few seconds or even milliseconds and earn money by making thousands of trades a day. But since these algorithms increase the volatility of the market, they are becoming the target of a financial regulations which would destroy that business model.

Therefore, we want to return to systematic algorithmic trading to get rich. We already developed a simple framework and implemented some basic strategies with it. We want you to extend this framework. This includes, but is not limited to, making it more realistic by adding a budget constraint and transaction costs, and creating more powerful strategies, e.g., by being able to short sell stocks. But of course, your own ideas are also welcome.



### Requirements

Good programming skills (preferably in Python) and a genuine interest in the financial markets. The student(s) should be able to work independently on this topic!

**Interested? Please contact us for more details!**

### Contact

- Philipp Brandes: [philipp.brandes@tik.ee.ethz.ch](mailto:philipp.brandes@tik.ee.ethz.ch), ETZ G64.2



---

---

## Bibliography

---

- Alexander, C. (2001). *Market Models: A Guide to Financial Data Analysis*. Wiley.
- Bollinger, J. (2011). Bollinger Bands.
- BOWLEY, G. (2011). Clamping Down on Rapid Trades in Stock Market. <http://www.nytimes.com/2011/10/09/business/clamping-down-on-rapid-trades-in-stock-market.html>.
- Brown, S. (2011). Measures of Shape: Skewness and Kurtosis. <http://www.tc3.edu/instruct/sbrown/stat/shape.htm>.
- Bürli, T. (2012). Make Us Rich!
- Christian L. Dunis, J. L. P. N. (2003). *Applied Quantitative Methods for Trading and Investment*. Wiley.
- dailyfx.com (2012). Bollinger Bands.
- David A. Dickey, W. A. F. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root.
- Fischer Black, M. S. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81:637–654.
- Hayashi, F. (2000). *Econometrics*. Princeton University Press.
- Holton, G. A. (2003). Negatively Skewed Trading Strategies. *Derivatives Week*.
- Investopedia (2012). Standard & Poor's 500 Index - S&P 500. <http://www.investopedia.com/terms/s/sp500.asp>.
- João F. Caldeira, C. V. M. (2012). Selection of a Portfolio of Pairs Based on Cointegration: The Brazilian Case. *Economics Bulletin*.
- Lauricella, T. and STRASBURG, J. (2010). "How a Trading Algorithm Went Awry". *The Wall Street Journal*.

- Robert F. Engle, C. G. (1987). Co-Integration and Error Corerection: Representation, Estimation and Testing. *Econometrics*, 55(2):251–276.
- Schmidt, A. D. (2008). Pairs Trading: A Cointegration Approach. *Times*, N. Y. (2012). High-Frequezy Trading. Article.
- Vince, R. (1992). *Mathematics of Money Management: Risk Analysis Techniques for Traders*. John Wiley & Sons, Inc.
- Wikipedia (2012). Skewness Statistics. [http://en.wikipedia.org/wiki/File:Skewness\\\_Statistics.svg](http://en.wikipedia.org/wiki/File:Skewness\_Statistics.svg).