Singular Value Decomposition as a Method for Analyses and Forecasts of Financial Data

Author: Axel Gustavsson

Supervisor: Thomas Fischer
Bachelor's Thesis - Finance
NEKH01
Lund University School of Economics and Management
May 2021

Abstract

This paper examines the sufficiency of a trading method based on singular value decomposition (SVD) of past stock prices. The SVD method is frequently used as a tool to reduce data noise, compress big-data, and analyse data components. Hence, the method is well suited to form a ground for a predictive tool of price developments. From the predicted pattern, a strategy was formed by construction of a portfolio of two business sectors concluded to be negatively correlated in one of the price movement components. An algorithm was programmed to receive buy- and sell signals when the difference in the price gradient exceeded a fixed value. The active portfolio was bench-marked against a buy-and-hold strategy for a portfolio consisting of the same stocks and weights.

A set of 93 stocks from the NYSE were selected and divided into groups to represent a variety of business sectors in the market. The active strategy showed, in the simulation period 2014-2018, to at best have an average annual excess return to the passive portfolio of $13,66\%$. The strategy performance was improved with stronger negatively correlated sector pairs in price components of sizeable significance.

It should be emphasised that the paper does not account for trading costs and market risks. However, the general conclusion is assumed not to be affected by the absence of incorporated trading costs. In addition, the strategy assumes an investor would follow the strategy over the entire simulation period, and not interfere with the trading algorithm, hence only be affected by the results on the last trading day.

SPECIAL THANKS TO:

Thomas Fischer

Mats Gustavsson

Carl Möllerström

Contents

1 Introduction **6**
1.1 Background 6
1.2 Problem 7
1.3 Disposition 8

2 Theory **9**
2.1 The Efficient Market Hypothesis 9
2.1.1 The "fair game" model and expected returns 10
2.1.2 The Random Walk model 10
2.2 Singular Value Decomposition 11
2.2.1 Algebraic definitions 12
2.2.2 Properties 12
2.2.3 Applications 12
2.2.4 Strengths and limitations 14

3 Previous research **16**
3.1 Stock market anomalies 16
3.1.1 Calendar anomalies 16
3.1.2 Fundamental anomalies 17
3.1.3 Chartist anomalies 17
3.1.4 Psychological heuristics and biases 18
3.1.5 Practical use of trading strategies based on anomalies 18
3.2 Singular value decomposition with entropy 19
3.3 Market interdependence 19

4 Empirical data **20**
4.1 Data-set 20
4.2 Data adjustments 21

5 Methodology **22**
5.1 Normalizing data and constructing data matrix 22
5.2 Applying SVD 22
5.3 Basis vector analysis 23
5.4 Building a trading strategy 23
5.5 Simulation 25

6 Results **26**
6.1 Original matrix and decomposition 26
6.2 Basis vector identification 26
6.3 Cross correlations 28
6.4 Simulation of the trading strategy 30

7 Analysis **34**
7.1 Basis vectors and anomalies 34
7.2 Correlations 34
7.3 Strategy performance and market efficiency 35

8 Concluding remarks **36**
8.1 Conclusion 36
8.2 Further research 36

A Data **39**

B Strategy and Matlab code **44**

1. Introduction

1.1 Background

The efficient market hypothesis (EMH) is an economic theory developed by the American economist Eugene Fama in 1965 that laid the foundation for the modern financial theory. The theory proposes that the price of an asset perfectly reflects all available and relevant information at any given time (Fama 1970). The EMH usually is divided into strong-, semi-strong and weak market efficiency. If a market is weakly efficient, investors are unable to profit from analysing historical data, as it is assumed the historical data reflects the prevailing price accurately.

In contrast, a tool routinely used by practitioners is technical analysis. Technical analysis is based purely of mathematics applied to historical data. The purpose is to produce a predictive tool that enables the investor to forecast future prices. Many of the methods are based of finding market anomalies, that is, trying to identify regular fluctuations in prices over a certain period. Strategies based on moving averages uses daily fluctuations, whilst methods based on Fourier analysis uses longer time periods. In this paper, singular value decomposition (SVD) is applied using a yearly time-frame. The method fragments prices, enabling a deeper analysis of the price data. The strategy will be formed by creating an algorithm that identifies patterns in the fragmented price data. Sectors of stocks with a negative correlation in one of the fragmented price patterns will then be actively traded, based on the predicted future prices. The performance of the active portfolio will be bench-marked against a buy-and-hold strategy.

The usefulness of the strategy can also provide some evidence in the controversies of the EMH. If the passive investment strategy advocated by the EMH are to be beaten by an active portfolio, the market (in this case the New York Stock Exchange) ought not to be efficient. Moreover, the tool, if satisfactory, would naturally be useful for the profit seeking investor.

The paper presents evidence that an actively managed portfolio strategy based of the SVD approach typically performs superior to the buy-and-hold strategy over the period of 2014-2018. The portfolios based of negatively correlated stock sectors in the most significant basis vectors gave the highest average annual excess returns.

1.2 Problem

Difficulty arises when researching the applicability and advantages of chartist analysis[1] on financial markets. This is namely, and certainly intuitively, that if any investor would find the mathematical "secret recipe" of the stock market, the possible excess returns would be removed in an instant if the investor were to share these findings. Alas, the successful investor would not risk to lose his or her profits, therefore rather keep the information to self. Consequently, the look for evidence has been widened not to *how* the chartist approach is applied, but rather *if* it is applied. If practitioners in real markets apply chartist analysis with successive excess returns, that alone is considered proof that the chartist way of investing is beneficial.

Taylor and Allen (1992) conclude how approximately 90% of practitioners in foreign exchange markets rely to some extent on non-fundamental analysis when predicting future market behavior. They also conclude a skewness towards technical analysis by practitioners with shorter investment time horizons. The paper specifically exemplifies by highlighting how clients of the financial firms ask not only if a currency will decrease or increase in value (which normally would be answered with a fundamental view), but also by *how much*. The latter would normally need a technically oriented response from the firm, and therefore a need of a chartist method.

In addition, Menkhoff (2010) found similar results suggesting this was not only the case in exchange markets, but also by fund managers operating in different investment domains. Evidence was concluded from 692 fund managers in five countries, where the vast majority replied they rely on technical analysis. Many of which also stated that they strongly believe in heterogeneous reactions to information, and thus that prices are determined largely by psychological influences. Menkhoff additionally concludes that 20% of the fund managers *prefer* technical analysis to fundamental analysis in their work. A possible explanation as to why this is the case, Menkhoff argues, is that the fundamental approach generally is more time consuming, and therefore less attractive, especially to smaller firms.

It is suggested that firms with chartist trading strategies, that perform better than the market in instances, might do so due to extreme luck. However, for each year the over-performers keep over-performing (some has accomplished this almost 30 years in succession), the probability of it being due to pure luck gets closer and closer to zero. Nevertheless, the studies highlights a potential gap between economic theory and real-world financial markets. Markets are generally considered to at least be weakly and semi-strongly efficient, yet, practitioners still uses technical methods - some with substantial results many years in succession. By applying singular value decomposition to financial data, using the results to form a prediction of future price movements and implementing a trading strategy, this paper aims at closing this evident gap in financial theory.

[1] Chartist referring to the procedure of analysing past data using mathematical methods, synonym to the term technical analysis.

1.3 Disposition

Firstly, the efficient market hypothesis, with applications, will be presented. Thereafter, the SVD method will be introduced. Followed by this, chapter 3 provides an insight into previous research. Further, chapter 4 will present empirical data, followed by chapter 5 which gives an insight into the methodology of the thesis. Chapter 6 will present the results, followed by an analysis of these in chapter 7. Chapter 8 ends on concluding remarks.

2. Theory

2.1 The Efficient Market Hypothesis

The origin of the efficient market hypothesis can be traced back to a paper by Fama (1970), even though his paper initially was intended to present an overview on already accomplished research. The EMH concludes that any investor is unable to, in successive style, perform excess market returns. This suggests that interpretation and usage of known information, chartist analysis and even insider information is unable to result in greater profits than a broad market index. To classify a market as efficient, Fama (1970) describes "sufficient conditions" of an efficient market. These are stressed however not as mandatory conditions, rather as circumstances that favors prices to adjust efficiently. The circumstances are:

- No transaction costs in the trading procedure, which implies availability of short selling along with no tax distortions.
- All available information is available at no cost to any party of the market.
- Homogeneous beliefs amongst market participants who act as rational agents.

The EMH is traditionally divided into different forms of efficiency, reflecting the different "strengths" of efficiency, reaching from *weak-form* efficient to *strong-form* efficient markets.

Firstly, a **weak-form efficient** market is one where the price of an asset reflects all past and publicly accessible information. Thus, analysing historical data will not give predictive power of future prices, since prices hitherto have incorporated the information in the current price. If markets are weakly efficient, the application of chartist investment strategies ought to be unable to generate excess returns. Technical analysis therefore is a waste of time, since any profits that could have been made with algorithms or similar tools already have been collected. Under this premise, prices should follow a random walk[2].

A **semi-strong-form efficient** market is one where prices incorporates all historically available information together with *new* publicly accessible information. Prices adjust rapidly to new information regarding fundamentals (such as earnings and dividend announcements, stock splits, management quality, and other corporate actions). This state of efficiency declares fundamental analysis as pointless, since the price of an asset quickly changes to the "correct" value whenever news occur.

Lastly, a **strong-form efficient** market asserts that (encompassing the weak-form and semi-strong-form efficient market conditions) prices fully reflects publicly and *privately* accessible information at any given time. Hence, insider information will not enable excess returns for an investor.

The definitions above help to distinguish markets into different categories of efficiency. However, controversies regarding the efficiency of global stock markets have risen with recurrent research suggesting divergent results, further discussed in Chapter 3.

[2] Discussed more thoroughly later in this chapter.

2.1.1 The "fair game" model and expected returns

In a try to quantify the statement of an efficient market, Fama (1970) suggests that a possible direction is to incorporate expected returns. Assume Φ_t is the information set at time t. The expected return of an asset can thus be described as follows:

$$E(\tilde{p}_{j,t+1}|\Phi_t) = [1 + E(\tilde{r}_{j,t+1}|\Phi_t)]p_{jt}, \tag{2.1}$$

where E_t is the expected value operator, p_{jt} is the price of an asset j at a time t and $p_{j,t+1}$ the asset price at time $t+1$. Also, $\tilde{r}_{j,t+1}$ is the percentage return on the asset, i.e equal to $(p_{j,t+1} - p_{jt})/p_{jt}$. Note that the tildes of $p_{j,t+1}$ and $r_{j,t+1}$ reflects that these variables are random at t. The application of expected returns to asset prices has the significant implication that the set of information Φ_t is unable to generate excess returns in equilibrium, since the equilibrium expected return depends on the same Φ_t. The statement leads to a "fair game" sequence:

$$z_{j,t+1} = p_{j,t+1} - E(p_{j,t+1}|\Phi_t), \tag{2.2}$$

where

$$E(p_{j,t+1}|\Phi_t) = 0. \tag{2.3}$$

Above, $z_{j,t+1}$ is the return at time $t+1$, excess to the expected equilibrium return of time t i.e $p_{j,t+1}$. Equation (2.3) demonstrates that the expected return at time $t+1$ from usage of past information Φ_t is zero. The result manifests the impotence of above market performance when incorporating the efficient market hypothesis to expected returns.

2.1.2 The Random Walk model

A special case of the "fair game" efficient market model is that of the random walk theory (Fama 1970). This extended application of the efficient market model has some further assumptions, namely that;

- Price changes are independent. Thus, the price changes of day $t+1$ is unaffected by price changes of day t.
- Prices are identically distributed, the standard deviation is constant over time.

The assumptions suggests a times-series in which prices only deviate in the case of random events. Furthermore, it suggests that deviations occur independently of past behavior in the price. Let P_n denote the state of the stock price at time n, we state that:

$$P_n = P_0 + \sum_{j=1}^{n} x_j, \tag{2.4}$$

Above, P_0 is the stock price at time $t = 0$, whilst x_j is the random walk generating variable, identically distributed, returning either a positive or a negative value. It is important to stress that the "fair game" model simply implies that asset prices can be expressed in terms of expected returns. But the random walk model on the other hand is applicable only in an economic environment when the distribution of returns are independent and repetitive, that is, constantly added new information produce instant equilibrium prices, that only deviates by random influences. Therefore, in an efficient market, prices solely deviate from new information that flows randomly, hence, prices move randomly. Also, since new information usually come at any time a day there are no price clusters. The new information of today is therefore argued to be independent of the news from yesterday.

One could question the assumptions that prices only deviate randomly considering that stock prices on average tend to increase with time. This might seem somewhat contradictory to the EMH, however, random in the EMH context is not equivalent to trend-less. Investing in a stock is a risk taking action, therefore the investor expects compensation (Byström 2014). The random walk model implies that profits only occurs as compensation for the risk exposure of the investor. Thus, profits are not related to timing of purchase, nor the instinct value of the stock calculated by the lone investor.

To conclude the section above, the EMH has several assumptions as well as implications that makes the theory a subject of discussion. Firstly, transaction costs and taxes are present at almost any financial market. However, one could argue these are rather small and not influential in market decisions. Information is suggested to be free to all investors at all times. The presence of high frequency trading, referring to investors trading with rapid frequency, paying exchanges to obtain trading information milliseconds before others (Lewis 2015) and expensive terminals points to the opposite. Also, prices of financial assets tend to be extra volatile in periods of financial instability such as during crises, questioning the stationary assumption of the random walk model. Possibly the most prominent reason to question the market efficiency conditions are the assumed homogeneous beliefs amongst investors. It seems rather peculiar that in a world of varying individual opinions in more or less every aspect of life (politics, monetary policy and sports to name a few), information regarding the stock market would be interpreted and acted upon in precisely the same fashion by every investor. Having said that, one could argue that individuals may not be fully rational, but as a group they are, and those who are rational will cover the arbitrages from the actions of those who are not.

2.2 Singular Value Decomposition

> "Things are seldom what they seem. Skim milk masquerades as cream."
>
> - Gilbert and Sullivan, *H.M.S Pinafore* (Malkiel 2019)

The trading strategy developed in this paper is based of an analysis of historical prices using singular value decomposition (SVD). The SVD method enables price data to be structured into different layers, and thus enables identification of price movements otherwise disguised by noise. With the layers identified from the historical data, the method could be used to predict future movements in the prices, thus to form a strategy.

The SVD method is of great use in many fields of research, since data often can appear to be random, or follow a certain pattern, when in fact the appearance is mainly due to noise. The important information in the data, containing causal interrelationships, can be examined by investigating each layer of information at a time. The following chapter will present the SVD method, its applications in other fields, and lastly how the method will be utilized in this paper - to develop a trading strategy.

2.2.1 Algebraic definitions

Before this section, a clarification of matrices and linear algebra terminology is in order.

- **Matrix** - A set of values (or symbols and expressions) ordered in a system of rows and columns.
- **Dimension** - Referring to the number of rows and columns of a matrix. A matrix with 3 rows and 2 columns has the dimension 3×2 (generally denoted as $m \times n$).
- **Vector** - A geometric object of a certain length and direction.
- **Basis** - If a vector $\overline{u}$ can be written as a linear combination of vectors $\overline{v}$ and $\overline{w}$, we can denote $\overline{v}$ and $\overline{w}$ a basis for $\overline{u}$.
- **Orthogonality** - If two vectors are perpendicular to each other, they are orthogonal vectors. If the vectors also have the length equal to 1, they are *orthonormal*.
- **Transposition** - If the matrix A has the dimension 3×2, the transposed matrix of A, A^T, has the dimension 2×3. Transposing a matrix substitutes the rows for the columns, and the columns for the rows.

2.2.2 Properties

Singular value decomposition is a linear technique whereupon a matrix is decomposed to left-singular vectors, right singular vectors and a diagonal matrix. The theory states that each matrix, regardless of dimension, can be written as:

$$M = U \Sigma V^T \tag{2.5}$$

Where M is the original matrix, U is a matrix containing the left singular vectors, Σ is a diagonal matrix with the singular values, where $\Sigma = diag(\lambda_1, \lambda_2, ..., \lambda_k)$ and lastly V^T is a matrix with the right side singular vectors. $\lambda_1, \lambda_2, ..., \lambda_k$ is denoted as the singular values of the matrix M. Note that the order of the singular values are such that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k \geq 0$. Further, the column vectors in U are orthonormal, just as the row vectors in V^T.

Equation (2.5) has the important and useful implication that every matrix can be decomposed into component matrices. These components represents, depending of the original data set, different structures of the original data. For a more in depth explanation of the mathematics behind the method, Henry and Hoffrichter (1992) is recommended.

2.2.3 Applications

The SVD method is frequently used in fields such as image processing (dimensional reduction), machine learning and statistics. In many of these applications, the objective is to accomplish a *truncation* of an original matrix. This is the process of reducing the size of a matrix, often with the purpose to decrease file size. But it is also of important use when a set of data contains substantial noise, as SVD filters and suppresses the noise whilst identifying the significant components of the data set. The truncation of a matrix includes selecting some of the subsets of columns U and V and corresponding singular values that is of greater significance, whilst eliminating those of less significance to rebuild the matrix. By doing so, the noisy data is separated from the original matrix, hence decreasing its size. The number of chosen significant singular values specifies the *rank* of the truncated matrix. In figure 1, a picture is decomposed by the SVD method. The condensed picture does a good job replicating the original image, using less data. In some instances, the condensation in fact improves the quality of the picture, since noisy elements are removed. The application has shown to be useful for instance in fingerprint analysis.

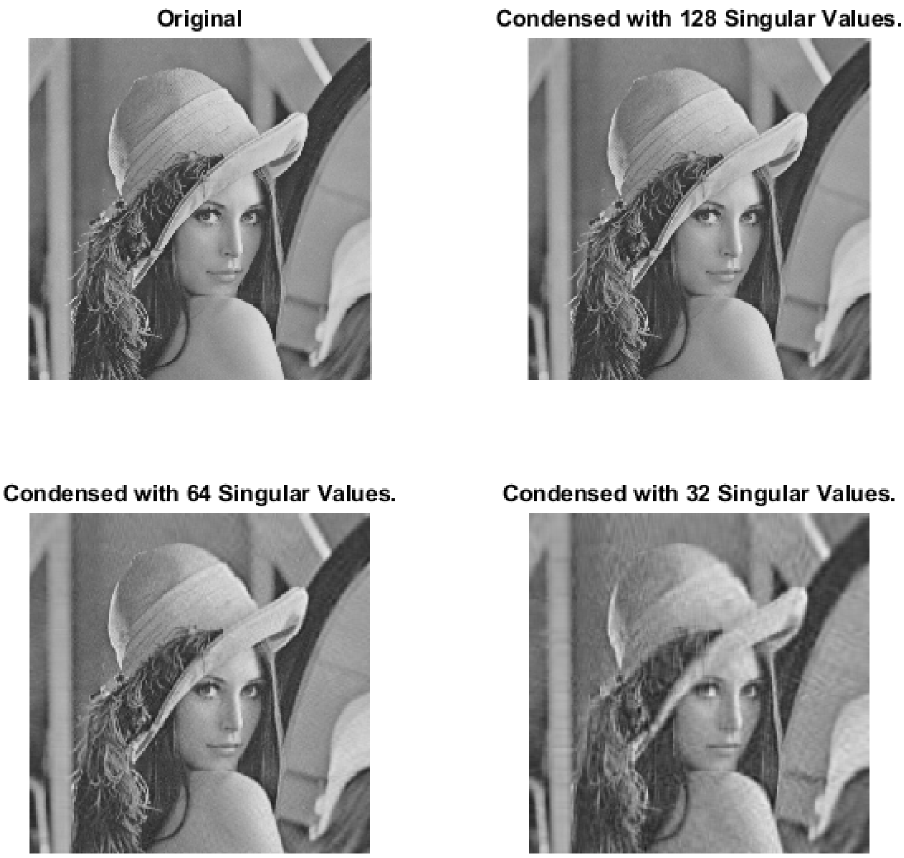

Figure 1: A picture being decomposed by the SVD method. Own computations.

Baker (2005) summarizes the most frequently used applications of the method; To transform correlated variables into uncorrelated ones (to identify relationships in the data set), to discover the rank of the matrix (and determine the variables with higher and lower variation), as a tool for data reduction. In the application used in this paper, the focus is to find relationships, i.e price patterns, and determine their significance for each sector of business. By filtering the price information for the selected stocks/sectors, one can identify the different concepts in the price movements otherwise disguised by general market movements. These concepts can be used to predict future prices, and to find market anomalies. Hence, the SVD method is in this paper implemented to find concepts and anomalies in the price data, which forms the basis for a trading strategy.

A few recent publications have applied the SVD method to the entropy concept of financial markets. Entropy is a measurement of the level of uncertainty or disorder is a system. The procedure is usually performed by computing the SVD of a raw correlation matrix. Results highlights that SVD has predictive power over financial market entropy (Caraiani 2014; Kenett and al. 2011; Gu, Xiong, and X. Li 2015).

2.2.4 Strengths and limitations

The SVD method in itself has a broad spectrum of areas in which it can be of use. The method tends to work well on a wide spread of data sets, including financial data[3]. In addition, the method is efficient even when applied to large matrices. With yearly trading days as rows, and stock/year as columns, the resulting matrix is of dimension 254 by 1942 (a total of 493 268 values, more on this in chapter 4). This is a rather small matrix in the SVD context. Finally, the method automatically orders the significant variables of the data-set since the singular values of the Σ matrix decreases on the diagonal. This makes identification of relevant variables straight-forward.

In contrast, the method undeniably has some limitations. Firstly, if the columns in the data matrix are of significantly different magnitudes, the method might perform poorly. Hence, normalizing data is of importance. Further, the SVD results can be difficult to display, partly since the derived matrices often are of large dimension. Moreover, the SVD-procedure can come about as troublesome to interpret, especially when used on original matrices with complex structures. Lastly, as to date, no scientific publication has been discovered where the SVD method has been applied in the same fashion as in this paper. Whether it is due to the lack of practical value of the particular application will naturally be examined in this paper. This section will be concluded with an illustration of a simple SVD computation. Note that M can be any arbitrary matrix.

$$M = \begin{bmatrix} -4 & 0 \\ 2 & 1 \\ 3 & -2 \end{bmatrix} \tag{2.6}$$

The SVD is given by,

$$\mathbf{SVD}(M) = \underbrace{\begin{bmatrix} -0.725 & -0.307 \\ 0.333 & 0.627 \\ 0.603 & -0.716 \end{bmatrix}}_{U}, \underbrace{\begin{bmatrix} 5.445 & 0 \\ 0 & 2.086 \end{bmatrix}}_{\Sigma}, \underbrace{\begin{bmatrix} 0.987 & 0.16 \\ -0.16 & 0.987 \end{bmatrix}}_{V} \tag{2.7}$$

[3]Historically proven mainly when investigating data entropy, described in the section above.

We can confirm that the column vectors of U are orthonormal. Likewise, the column vectors of V are orthonormal (i.e. the row vectors of V^T are orthonormal). By definition we can reestablish the M matrix using matrix multiplications.

$$M = \underbrace{\begin{bmatrix} -0.725 & -0.307 \\ 0.333 & 0.627 \\ 0.603 & -0.716 \end{bmatrix}}_{U} \times \underbrace{\begin{bmatrix} 5.445 & 0 \\ 0 & 2.086 \end{bmatrix}}_{\Sigma} \times \underbrace{\begin{bmatrix} 0.987 & -0.16 \\ 0.16 & 0.987 \end{bmatrix}}_{V^T} \tag{2.8}$$

Efficient computing methods of SVD are broadly established. In software like Matlab, SVD of matrices are computed in a fraction of a second (even for matrices with millions of values).

3. Previous research

3.1 Stock market anomalies

A trading strategy is generally based of the finding of an irregularity in the price movements of financial assets. In the case of this paper, the SVD method is used for this purpose. The irregularities, or sometimes regularities, are referred to as market anomalies. The following chapter will present a literature review of the field of research in market anomalies and trading strategies. For if the strategy developed in this paper is useful, it can likely be explained by one or a few of the phenomena described in this section.

The construct of anomalies in stock markets was thoroughly articulated in a book by Elroy Dimson on the issue, released in 1988 (Dimson 1988). The author presented anomalies that can be divided into three categories. Firstly, the calendar anomalies, referring to irregularities in the pattern of prices occurring with respect to seasonality. Those patterns can be found in daily, weekly, yearly or even longer time frames. Secondly, fundamental anomalies, which include the over-returns of smaller firms, firms with a lower price per earnings ratio (P/E-ratio) and firms with higher dividend yields. Lastly, there are anomalies suggested for instance by Brock, Lakonishok, and LeBaron (1992) to occur due to chartist phenomena, such as moving averages and trading range breaks. In addition to these, are the anomalies with occurrence due to psychological biases, see for example Thaler and Bondt (1984). The field of behavioral economics had its breakthrough in the early 2000's, hence not included in Dimson (1988). The section below will present some empirical evidence to the above described market abnormalities, and briefly discuss their practical value.

3.1.1 Calendar anomalies

Possibly one of the better known anomalies, but heavily debated in the regard of whether it is of use or not, is the *January effect*, referring to the phenomena that prices tend to decrease in December to increase in January. It is suggested that the phenomena is due to increased selling of poorly performing stocks to avoid or effectively manage tax costs (Dimson 1988). In addition, other empirical research suggests prices tend to be higher on Fridays, and lower on Mondays (an example of the *day of the weak* effect). This has been shown to be the case for instance by Keim and Stambaugh (1986), who examined the phenomena in the S&P Composite index. It should be stated, however, that these findings not necessarily points to the fact that stock prices fall due to the trading on the Friday and Monday, but could also be explained by the non-trading days separating the two.

In addition, anomalies over longer periods seems to be apparent in financial markets. The "Sell in May" phenomena (or the *SIM* phenomena) accords for prices suggested to be on average higher in winter months than in summer months. This is in line with the findings of Dimson (1988). The anomaly has been known to economists for some time, however in a recent article by Degenhardt and Auer (2017) new empirical evidence was presented, to conclude the phenomena seems to still be apparent - even in today's liquid stock markets. It is however suggested that stock markets, in contrast to commodity markets, have less SIM effects today than some decades ago. This point to the fact that profits from well-known anomalies diminishes as the phenomena is spread amongst investors, and with the liquidity of the market. If investors get a hint that a stock price will rise tomorrow, one can count on prices rising today.

The *United States presidential election cycle* is another observable anomaly identified in stock markets, especially in the US. Gärtner and Wellershoff (1995) finds evidence that stock prices tend to fall during the first haft of the presidency period, and rise during the second - thereby following a four-year cycle. The results are similar for a variety of indexes (over the last 30 years). An investment strategy exploiting the anomaly would according to Gärtner and Wellershoff (1995) have resulted in a real rate excess return to a buy-and-hold strategy of 35,86 percentage points. Even with transaction costs included, the excess returns would still be rather impressive.

3.1.2 Fundamental anomalies

Several fundamental anomalies have been found by a variety of researchers. Since these primarily are relevant to those practicing fundamental analysis, and since this paper is an application of a technical or chartist method, only a chosen set of these anomalies will be briefly introduced below.

- The low P/E anomaly: Suggests that stocks with lower P/E ratios have higher returns than high P/E stocks (Goodman and Peavy 1983).

- High dividend yield: Stocks with higher dividend yield outperform the market (Fama and French 1988).

- Stock neglecting: Suggested by Thaler and Bondt (1984) as a phenomena where neglected stocks from the past generate higher returns in the future.

- Small firm effect: For instance described by Barry and Brown (1984), straightforwardly states that the stocks of small firms perform superior to those of larger firms.

3.1.3 Chartist anomalies

Continuing with technical anomalies, Jegadeesh and Titman (1993) highlights one particular example by concluding that the technique of buying "past winners" and selling "past losers" would have rewarded its practitioners with a significant abnormal return (excess 12,01% annually compared to index) from 1965 to 1989. That is, buying stocks that have performed well in the past (and selling stocks with poor past performance) is suggested to give the investor higher returns than a passive portfolio. In addition, Brock, Lakonishok, and LeBaron (1992) highlights two alternative approaches: Trading on moving-averages, and trading on range breaks. To trade on moving averages is a technique in which the investor buys stocks when it trades with a higher short-period average price than long-period average price. Naturally, the investor sells stock who appear with a short-period average below the long-period average. Brock, Lakonishok, and LeBaron (1992) presents evidence that the general excess return of a moving average strategy could be roughly 0.8% over a 10-day period of trading.

The strategy of trading on range breaks is one were the investor receives a buy signal on a stock when it breaks the so-called resistance level (a calculated global maximum level) of the stock price. Under the same rule, a sell signal is received by the investor when a stock falls though the global minimum price. An investor applying the range break method usually claims that if a price "breaks" the positive/negative barrier, it ought to continue to increase/decrease. This enables the investor to gain quick returns if one can buy the stocks at the right moment, i.e when the barrier is broken through. The strategy is proposed by Brock, Lakonishok, and LeBaron (1992) to enable for excess returns.

3.1.4 Psychological heuristics and biases

Kahneman (2011), a Nobel prize winner in economics, concludes over several years of clinical research biases and heuristics in the minds of humans. As many of the findings are beyond the scope of the purpose of this thesis, only a selection of the results that ought to be relevant to financial markets are presented below.

- Humans tend to be impulsive and act on automatic signals from our brains. In financial markets, this could imply we sell/buy stocks on impulses and not necessarily due to thoughtful reasoning.

- The brain is lazy, hence, we make clumsy and "easy" mistakes more often than we might think we do.

- We are loss averse, meaning we are a lot more afraid to lose what we already have, as we are keen on gathering more. Conversely, humans generally bet on losses, but not on wins.

- As a whole, humans are overconfident in their ability to control and affect the outcome of their lives. For instance, we believe we know more about financial markets, and their behavior, than we do. The heuristic is most apparent amongst men.

3.1.5 Practical use of trading strategies based on anomalies

It is important to state that the empirical evidence for the variety of anomalies in this section needs to be interpreted rather cautiously. When transaction costs, analyst costs and taxes are included, the profits generally are too small to be of any use to a lone investor. Also, if profits actually can be made, it seems reasonable to believe that those profits have disappeared due to the publication of the empirical evidence. In addition, one should be aware that there could be a general bias in the evidence on market anomalies. Studies that discover anomalies presumably receive more attention than those who do not.

Nevertheless, if transaction costs are low (say, if you operate in a large firm), and if you have millions of dollars to invest, small percentages can be of use. The fact that *some* anomalies exist might be enough for such an investor. As the saying goes, small streams make great rivers.

3.2 Singular value decomposition with entropy

With research on market anomalies introduced and discussed, this section instead focuses on singular value decomposition and financial data. Though SVD indeed has several benefits for data analysis, limited research has been made on financial data. To date, the SVD method has mainly been used when determining the entropy of a system in a financial context. For instance, Caraiani (2014) examined the predictive power of the SVD entropy for dynamics of the stock market. By applying correlation matrices, the paper found that entropy has predictive power when examining market dynamics. Also, the paper suggests that a drastic change in the entropy occurs especially in times of financial disorder, such as when the dot-com bubble bust in 2001, and in the Sub-prime mortgage crisis of 2008. In similar fashion, Sandoval and Franca (2011) concluded how markets tend to behave ”as one” during crashes of the financial market. Though Sandoval and Franca did not use the SVD method, a similar approach was applied, namely using eigenvalues of the correlation matrix constructed of market indices. In addition Gu, Xiong, and X. Li (2015) evaluated the predictive power of the SVD entropy in the Chinese Shenzen stock market. In similarity to Caraiani (2014), the paper presents notable evidence that the singular value decomposition entropy has predictive power. However, this phenomena is concluded only to be apparent in the time period after the reformation of the non-tradable share environment in the Shenzen market.

3.3 Market interdependence

Since the strategy in this paper depends on findings of negatively correlated stocks, a few words on the development of correlations in financial markets seems in order.

As financial markets become more liquid, and as economies have been radically globalized and more interlinked, the appearances of negative correlations between areas of business have decreased. Still, a few business sectors can still act as hedges to each other. Baur and McDermott (2009), with many others, show that gold continues to be a ”safe haven” for investors in European and US stock markets. That is, gold tends to be negatively correlated to the stock market in general. If one expects a bearish stock market, gold is a suggested to be a good investment. Furthermore, commodities have been considered to be a good hedge against the stock market. On the contrary, Olson, Vivian, and Wohar (2017) concluded that commodities generally are an ineffective hedge for investors in the S&P index, thus a bad hedge to American stocks (stocks with a larger capitalization tends to have strong correlation to the market in general). Nonetheless, research has shown that stock markets do have negative correlations, for instance by Maheshwari, Gupta, and Li (2017). Finding negative correlations is of great value when implementing an investment strategy, since it enables to decrease the risk of an investment without compromising the expected returns. Disappearances of negative correlations in the markets makes it more difficult for investors to find ways to implement clever strategies, and to hedge against extensive losses. Therefore, research in this field is of great priority for risk management, especially with the rise in globalization. The SVD method is an alternative approach to general correlation approaches, since it investigates correlations in separate fractions of the data.

4. Empirical data

4.1 Data-set

A total of 94 stocks from the NYSE was initially selected (all listed in Appendix 1), chosen to represent the biggest corporations in each sector of business on the NYSE. The data was obtained from Wharton Research Data Services (WRDS). The data-set contained daily closing prices for each stock in the period of 1990-2018. A few of the chosen stocks had been listed on the exchange later than 1990, and thus contained less data points. These were also used in the computations, but only contributes to the result for the years in which data existed. To find functioning results, the stocks were divided into areas of business. Below is listed the areas of business for the stocks, as well as the number of stocks in each sector.

Table 4.1: List of areas of business (defined by ICE data services) for selected stocks as well as the sample size in each sector.

Area of business	**Number of stocks**
Financial Services	15
Retail Trade	10
Consumer Non-Durables	9
Electronic Technology	8
Technology Services	7
Non-Energy Minerals	6
Health Technology	4
Producer Manufacturing	4
Energy Minerals	4
Transportation	4
Health Services	3
Consumer Services	3
Process Industries	3
Consumer Durables	2
Utilities	1

4.2 Data adjustments

Some of the selected stocks have been involved in mergers, stock splits and acquisitions. In such circumstances, stock prices diverge by giant leaps over the course of one or a few trading days. As an example, if a stock is split with a 2:1 ratio (often an action taken to improve stock liquidity) the stock price would fall by 50% since the number of stocks available for purchase has doubled. The company will still *ceteris paribus* have the same market value, but in an increased amount of shares. The splits in the stock data had to be manually nullified, accomplished by multiplying the price by the split quotient on the day of the split, the as seen in the chart for the Apple Inc. stock below.

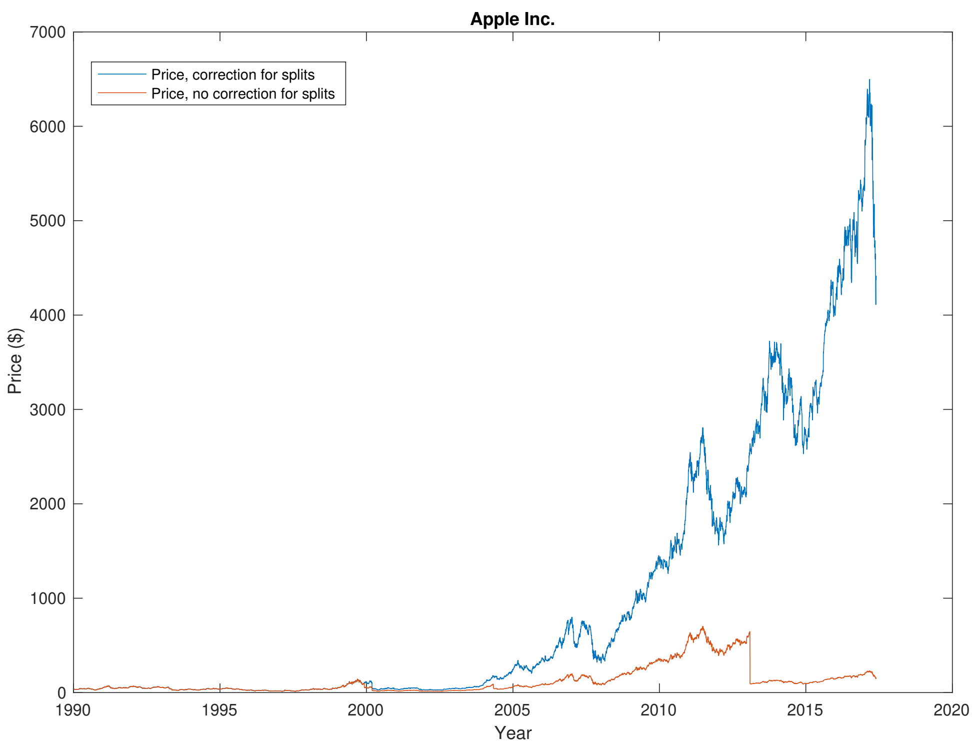

Figure 2: The price of Apple stock adjusted for splits (blue), and not adjusted for splits (red). Data obtained from WRDS.

In the figure above, note that the price per the end of the time period seems to be unrealistically high ($\approx \$4,500$). This is due to the fact that the correction for splits is constructed with a forward method. In common charts, found on various platforms such as Yahoo Finance, the splits usually are nullified with a backwards method, for it satisfies the current price of the stock. The reason behind using the forward method is to simplify the comparison and the SVD analysis, since it returns a continuous graph. The aim of this thesis is to investigate similarities and differences in historical price changes, hence, the relevant measure is the price development (the return over the simulation period) rather than the actual price in dollar and cents in today's market.

5. Methodology

5.1 Normalizing data and constructing data matrix

Firstly, all stock prices was corrected for splits according to the procedure in chapter 4.1. When corrected for splits, the prices showed a somewhat exponential behavior (with some exceptions). To normalize data, the *norm* function in Matlab was used. Prices were normalized by using Euclidean norm (or Euclidean length) of the annual price vectors.

$$||v|| = \sqrt{\sum_{k=1}^{N} |v_k|^2}. \tag{5.1}$$

Above, v_k denotes the price for all elements k, with the total amount of N elements. To appropriately scale the prices, the following computation was made:

$$P_{sc} = \frac{1000}{||v||} \tag{5.2}$$

P_{sc} is the scaled price whilst $||v||$ was derived from 5.1. A variety of normalization methods can be used, yet the described method was concluded to give the most appropriate values and convenient scale for the matrix.

Once the normalized prices was calculated, the original matrix could be constructed. The rows of the matrix denotes the 254 trading days of each year. Each column represents each stock, for one year at the time. To demonstrate with an example, let the first row and first column (i.e element (1×1)) of the matrix be the normalized closing price of the Apple stock on the first trading day of 1990. The following column, on the same row, (1×2), then represents the price of the Apple stock on the first trading day of 1991. Intuitively, the second row, first column (2×1) represents the closing price on the second trading day of 1990 for the Apple stock, and so on. Because of the amount of stocks and years in the data, the matrix contained 1942 rows. The dimensions of the matrix hence was 254×1942.

Generally, financial economics investigates returns rather than prices. However, in this research the price is used, since the strategy is based of long term price developments rather than daily fluctuations.

5.2 Applying SVD

When the matrix was established, the SVD procedure was applied. The *svd* function was used for computation. The decomposition process was followed by an analysis of the identified basis vectors, i.e the left diagonal vectors in U. All of the orthogonal vectors are, if put together, able to reconstruct the original data matrix, hence able to reconstruct the yearly closing prices for all of the stocks in the data-set. By identifying the magnitude of the singular values provided by the Σ matrix, and by displaying the contribution of each basis vector graphically, the amount of basis vectors to be used for further analysis was determined.

5.3 Basis vector analysis

Once the significant U vectors were found, these could be displayed to identify some of the underlying price movement behavior of the stocks. One would expect the first basis vector, the one related to the largest singular value, to show a positive trend with no lucid period, comparable to a broad market index. The less significant vectors, however, provide an insight into behavior in the price that would be difficult or impossible to find without the decomposition process. The time window, or the period of the function of U, is 254 days, i.e the number of trading days each year on average.

To continue the analysis, it needed to be distinguished how much each stock contributed to the significant need of each basis vector to reconstruct the original matrix. The SVD computation is made on the M matrix of *all* of the prices for *all* of the stocks. Since the data consisted of price data from a large sample of stocks, the stocks were divided into different business sectors (figure 4.1). Once divided into categories of business, the average contribution of each business sector to the importance of each basis vector, for each year, was computed using the mean function[4]. With the average dependence of the basis vectors for each sector computed, it followed to compare the sectors to one another. This could have been done by simply over viewing the graphs, and seeking interesting vector behavior for the averages, but was more accurately done by calculating the cross correlation between the averages through the Pearson's correlation formula:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}. \tag{5.4}$$

The correlations in each basis vector, for each sector, could then been analyzed.

By analysing the contribution from each stock sector to the basis vector it is possible to distinguish which sectors to use in the portfolio. The aim is to find two sectors, both following the same pattern, but with different directions. By doing so, one can decide on which stocks that is assumed to be easiest to trade against each other. This, since the algorithm would be programmed to trade whenever the slope of the predicted price vectors would differ from each other, more thoroughly discussed in the following section.

5.4 Building a trading strategy

With negatively correlated sectors of stocks identified, the portfolio could be constructed. Two sectors of stocks, with the same pattern, but negatively correlated, formed the strategy. The basis vectors from the SVD would form a predictive pattern for how the stock prices are suggested to move with time. The strategy would be built on trading whenever the difference between the slope, or gradient, in the predicted price vector would reach a certain value. That is, to create an algorithm

[4]The mean is derived from:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} A_i, \tag{5.3}$$

where A_i is the value (or length) of each vector and N the number of values in the set.

that would predict future price movements, derived from the basis vector analysis, and program the algorithm to trade when the difference of the slopes of the predicted prices for the two sectors exceeds a certain point. At that point, the sector of stocks with a predicted upside would be bought, and the negatively correlated sector would be sold. The predicted price pattern of the stock used by the algorithm would be derived exclusively on past data.

The first step to implement the strategy is to select a period of historical data for establishing the basis vectors. This time period is typically 5-15 years. All price estimates and price movements will be based on this fixed set of basis vectors. At every trading day, the algorithm computes a new best estimate of the scaling factor for each basis vector, by a least mean square error method. The scaling factor determine how much each basis vector contributes to the historical price development for a specific sector of stocks. With the basis vectors and the scaling factors the gradients of the stock sector prices are derived. By evaluating the difference in the gradients, the algorithm would pursue profit opportunities. For a more in depth explanation of the strategy implementation, see Appendix B.

Figure 3, below, illustrates the strategy of using gradients for initiating trading activities. The figure uses simplified mathematical functions, rather than actual price data, for a clearer display. In practice, the price curves would be estimated from the SVD analysis. To be able to maximize the profits in the portfolio, the algorithm would seek to buy the stock (or sector of stocks) in which the price is estimated to increase, whilst selling the other. In doing so, the portfolio ought to over-perform the passive strategy.

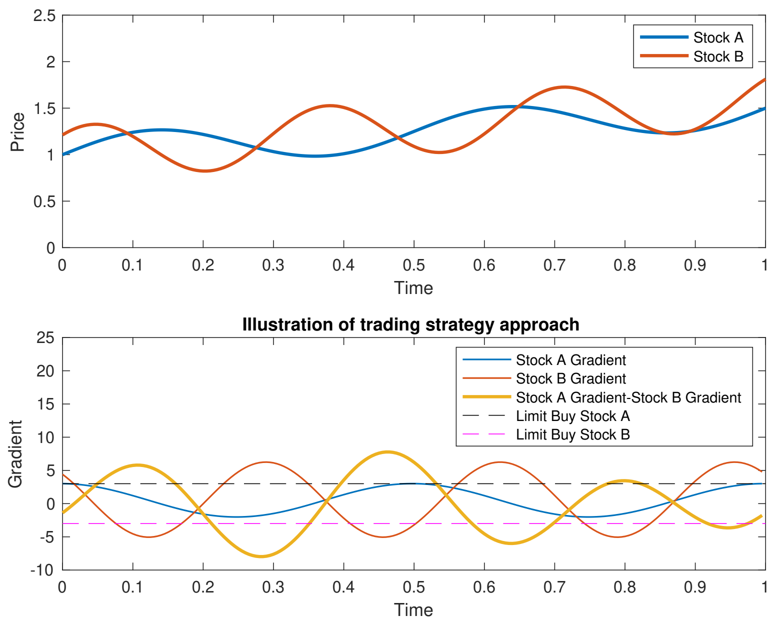

Figure 3: Illustrative example of the portfolio strategy. When the algorithm identifies signals above the limit for the stock (A or B) it buys the stock until the gradient difference decreases below the limit.

The actively sold and bought sectors would form a portfolio that would be benchmarked with a buy-and-hold strategy over a determined trading period. Initially, the portfolio would be constructed so that the same amount of capital would be invested into each of the sectors of stocks. The buy-and-hold strategy would then hold the same portfolio weights over the entire investment horizon, whilst the active portfolio managed by the algorithm would alternate the weights if considered favourable.

5.5 Simulation

The algorithm behind the active portfolio is working on existing data, updating every trading day, to forecast the trends of future prices. Intuitively, the buy-and-hold portfolio will be the actual prices for the simulated years, whilst the active strategy will be the outcome if the portfolio instead followed the predicted prices by the algorithm. For the active portfolio, the profits will be reinvested into the portfolio. To compare the performances, the annual excess return will be computed.

To end this chapter, the method will be listed in short, to clarify the steps latter to the SVD computation and component analysis.

- Identify negative correlations for business sectors in basis vectors one concludes to be significant.
- Use the sectors where the negative correlation is found to test the strategy. Set up the initial portfolio weights.
- Create an algorithm that uses only the significant basis vectors to predict future prices for the two sectors.
- If the difference in the gradient in the predicted price vector for the two sectors exceeds a decided limit, program the algorithm to buy the bullish stock and sell the bearish (in line with to figure 3).
- Repeat the process over the entire simulation period.
- Compare the returns of the strategy to a buy-and-hold strategy over the same time frame.

6. Results

6.1 Original matrix and decomposition

The original data matrix, after correcting for splits and normalizing, is displayed in figure 4. The graph shows the complete data-set of stock prices for each of the years in which closing prices subsisted. Also, the stocks are in ordered so that the first 28 values of the y-axis (labeled "Each stock, each year") represents the closing price of the first stock in the list, from 1990-2018 (28 years). Hence, y-value number 29 accounts for the closing prices of the next stock in the list, in year 1990.

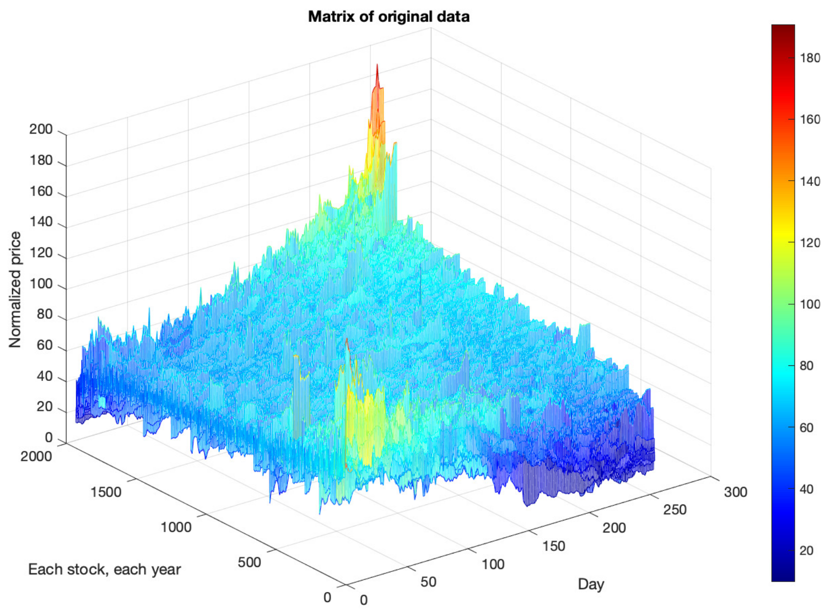

Figure 4: The original matrix of the normalized stock prices (each stock, each year), over the trading days each year. The color bar displays the price range (normalized).

6.2 Basis vector identification

The SVD computation resulted in several identified basis vectors. By manual inspection, the first sixteen was concluded to be of significance when reconstructing the original matrix. This was further concluded by viewing the singular values in the Σ matrix. The deviations between each graph (each containing one or a few added basis vectors) and the original matrix was calculated and displayed to highlight the significance of addition of the vector/vectors (deviation calculated by the euclidean norm of the true price minus the price estimate for the projection, divided by the euclidean norm of the true prices). The divergence from the original matrix with the sixteenth U vector added was less than 1% from the original matrix, and therefore no further basis vectors were added. Vectors of a smaller rank would not give a useful contribution to the compressed matrix, i.e they do not contribute substantially to describe the movement of the actual stock prices. The procedure is the same as in the computations in figure 1, but this time with stock data rather than pixels from a picture.

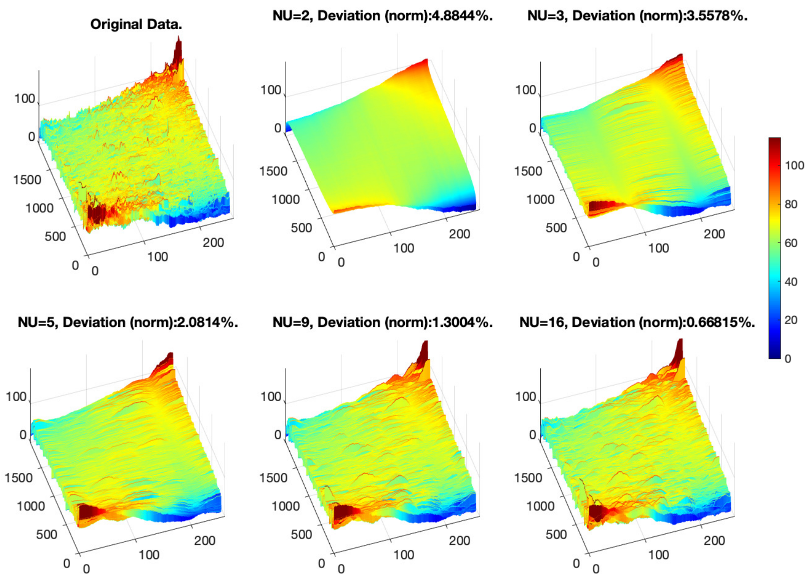

Figure 5: Matrix reconstruction by adding basis vectors. Note that the axis properties are the same as in figure 3. The color bar displays the price range (normalized).

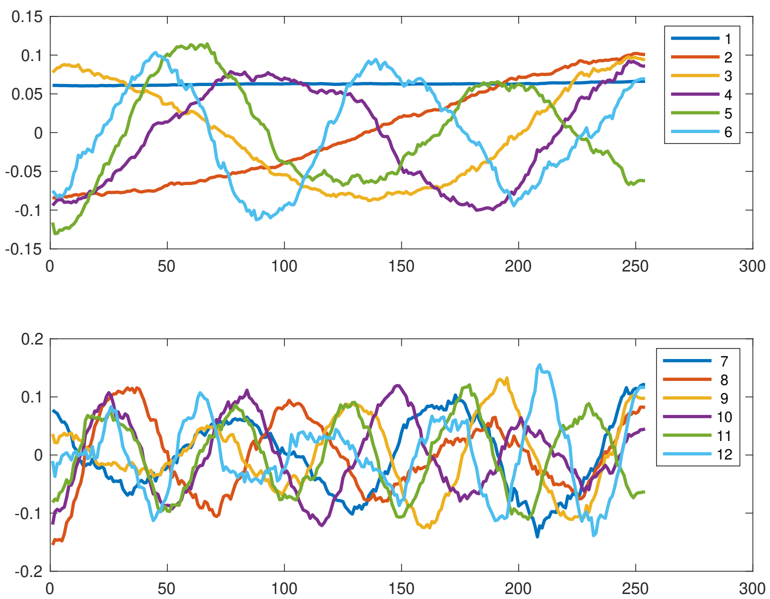

Figure 6: The basis vectors, also denoted as the U-vectors, and their appearance. The y-axis is the magnitude (normalized price increase/decrease) and the x-axis denotes the yearly trading days, starting at the first trading day of January and ending on the last trading day of December.

The basis vectors displayed above all represent a certain movement or behavior in all of the stock prices over time. That is, the most significant basis vector (the vector connected to the largest singular value) represent the most crucial vector to add to the decomposed matrix to describe the over-all behavior of all of the stocks. This is the U vector denoted number 1 in figure 5. Continuing, the second most essential basis vector (connected to the next-to largest singular value) represents the second most necessary price-cycle needed to reconstruct the entire data set. Notably, the U vectors have been arranged so that all of the vectors are increasing over the period. To note is that the frequency of the basis vectors decreases with significance. The most crucial price movements are hence of greater wavelength. Thus, prices are suggested to move in a yearly cycle, in line with previous research (*SIM*).

6.3 Cross correlations

The matrix correlations between each sector of business, for each basis vector, is shown below. As seen, the first basis vector shows a positive and equal to 1 correlation between all of the business sectors for the first basis vector. This makes intuitive sense, since one would expect all of the stocks to have a strong correlation to the general market movement. However, in the slightly less significant basis vectors, negative correlations between the sectors are recognized. Some sectors of stocks hence has followed the same fragmented price pattern, however by opposite direction. As expected when constructing a correlation matrix, the matrix has a symmetric appearance. This follows from the axiom of correlations, namely that the correlation between A and B is equivalent to the correlation between B and A, i.e we can state that corr(A,B) = corr(B,A).

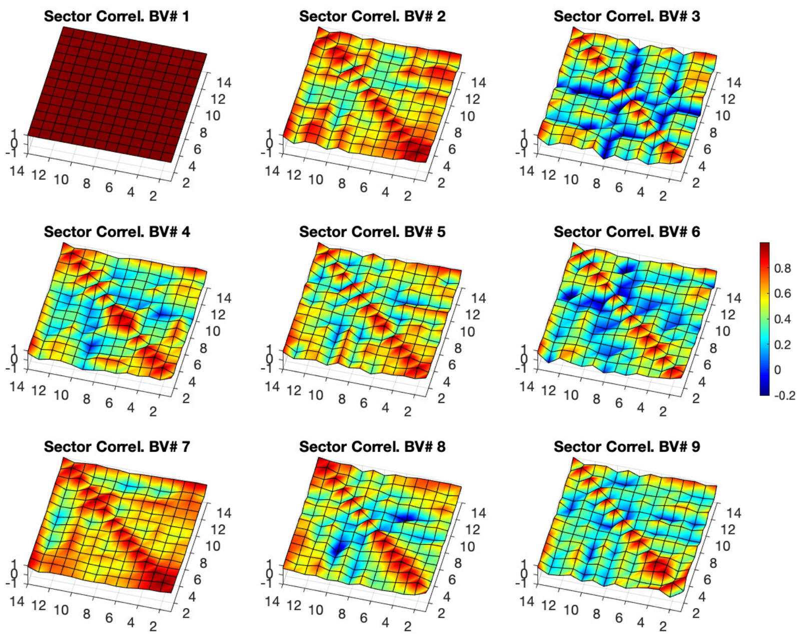

Figure 7: The cross correlation in each basis vector U for all of the sectors. The correlations ranges according to the color bar in the figure.

In the figure above, each sector has been given a number. The following lists each sector and its given number.

1. Electronic Technology
2. Financial Services
3. Producer Manufacturing
4. Transport
5. Retail Trade
6. Consumer Non-Durables
7. Health Technology
8. Health Services
9. Technology Services
10. Energy Minerals
11. Consumer Services
12. Non-Energy Minerals
13. Producer Industries
14. Consumer Durables

The most notable correlations found in the cross correlation table were those of basis vector 3,6,8 and 9. As expected, negative correlations are sparse in the first 2 basis vectors. For previous research has suggested that stocks from different sectors generally are positively correlated. Nevertheless, negative correlations were found in less weighty basis vectors. Those of the two most significant basis vectors are displayed in the following table.

Table 6.1: List of identified negative correlations in important basis vectors.

Basis vector (U)	**Sectors**	**Correlation**
3	Health Technology(7) and Health Services(8)	-0.3692
3	Transport(4) and Energy Minerals(10)	-0.3341
3	Health Services(8) and Non-Energy Minerals (12)	-0.3262
3	Electronic Technology(1) and Health Services(8)	-0.2366
6	Technology Services(9) and Producer Industries(13)	-0.4003

6.4 Simulation of the trading strategy

To evaluate the usefulness of the strategy, it was compared to a standard buy-and-hold strategy over the last four years in the data set (2014-2018). The portfolio included two sectors at a time, each consisting of a number of stocks. The portfolio weights for the two sectors were derived so that an equivalent amount of capital was invested in both of the sectors at $t = 0$. In addition, the stocks in the sectors were normalized (value weighting), due to some stocks being particularly more expensive than others, thus overriding the other stocks impact on the portfolio. The algorithm reprogrammed the estimate of the scaling factor for each basis vector, and derived the gradient of the asset prices, at every new trading day - to evaluate if a redistribution would be favourable for the portfolio.

In the simulation, a trade rate of 50% was used, i.e for each buy/sell signal from the algorithm, 50% of the invested capital in the disposed asset was reallocated. If the stocks had been exceptionally volatile, a larger gradient difference was chosen. All of the simulations are displayed below. Note that the gain measurement is derived as the difference between the portfolio values at t. For a comparison of returns, see table 6.2. The value of the active and the passive portfolio is always equal at $t = 0$, demonstrating that the same amount of money is invested in the passive and active portfolio at the beginning of the simulation period. To note is that the parametric values chosen in the code (for instance, the amount of past trading days used for future prediction) was chosen depending on the isolated portfolio. The simulation hence can be modified, for better or worse. Below, some results of the five simulations, with slightly different parametric values, are displayed.

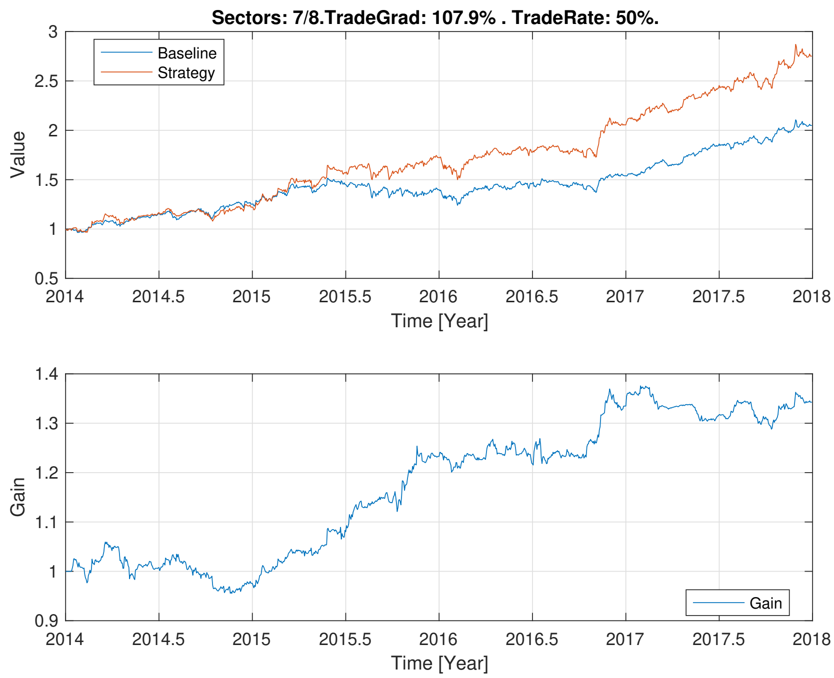

Figure 8: Portfolio value and strategy gain over the simulated period. Sectors: Health Technology (7) and Health Services (8).

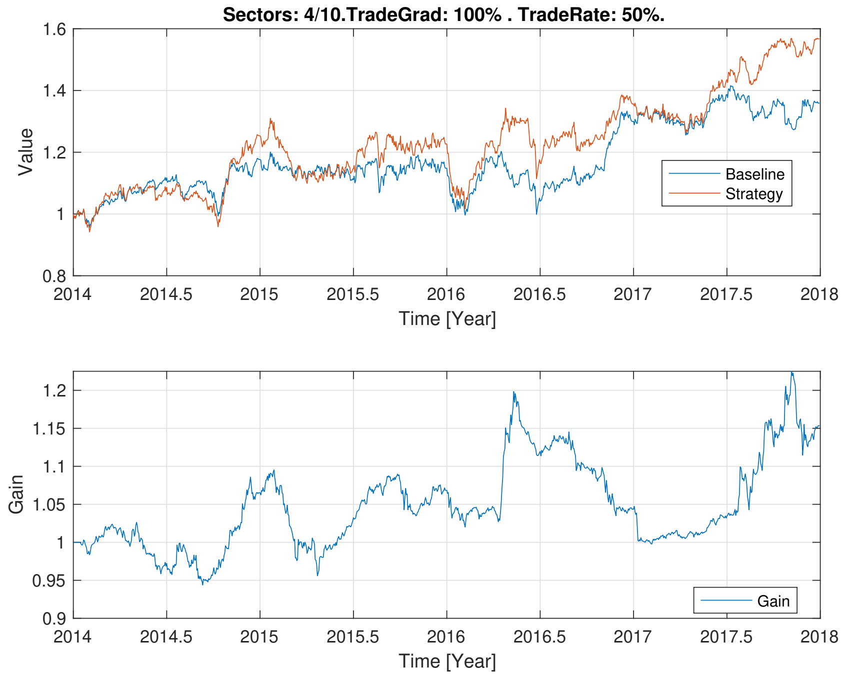

Figure 9: Portfolio value and strategy gain over the simulated period. Sectors: Transport (4) and Energy Minerals (10).

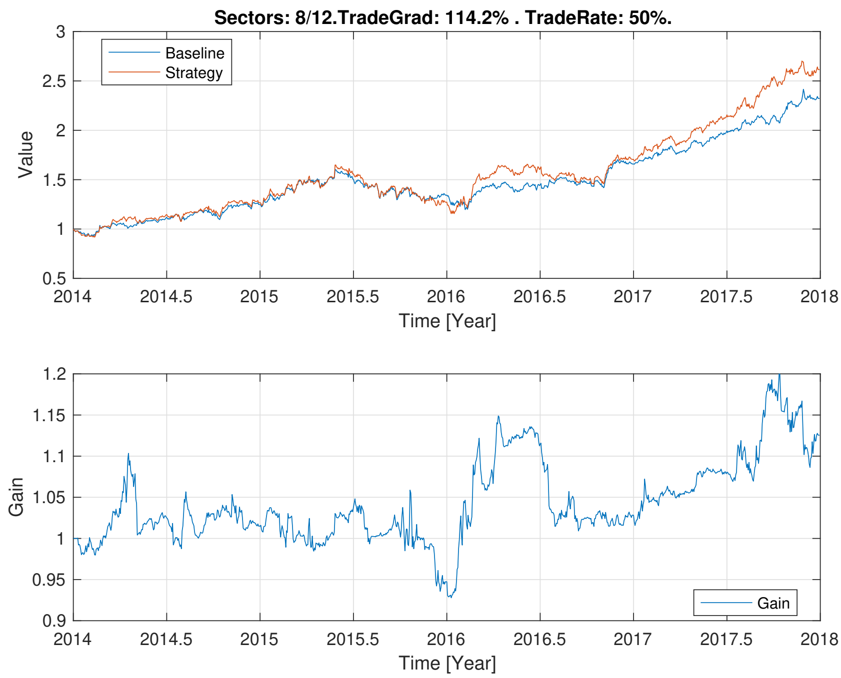

Figure 10: Portfolio value and strategy gain over the simulated period. Sectors: Health Services (8) and Non-Energy Minerals (12).

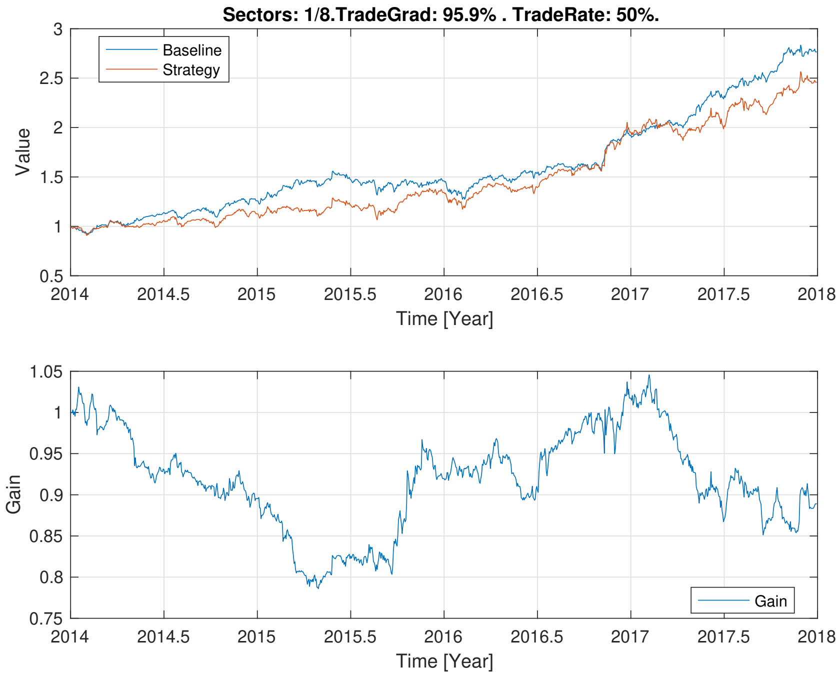

Figure 11: Portfolio value and strategy gain over the simulated period. Sectors: Electronic Technology (1) and Health Services (8).

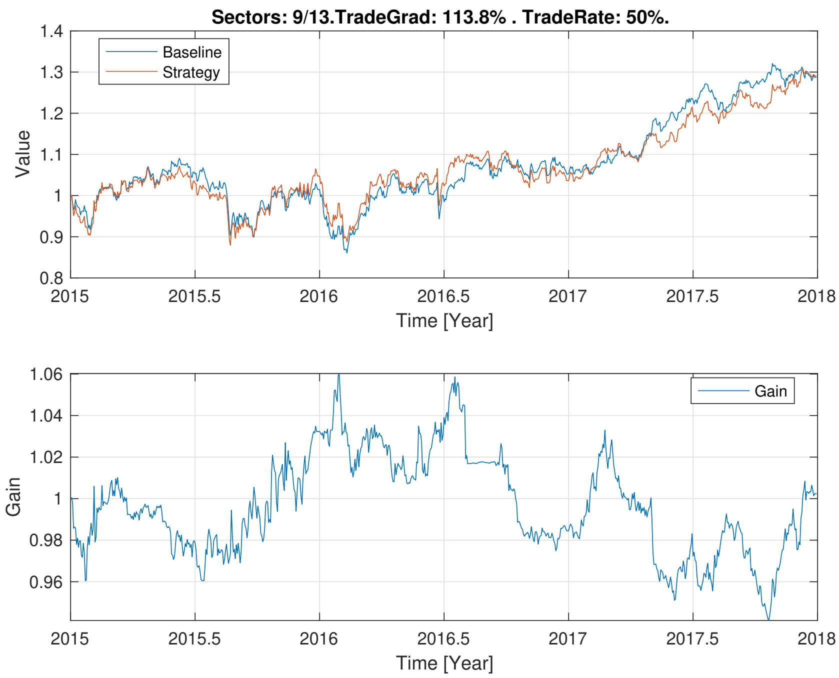

Figure 12: Portfolio value and strategy gain over the simulated period. Sectors: Technology Services (9) and Producer Industries (13). Note that the simulation period is three years due to missing data for one of the stocks for 2014.

Table 6.2: Average annual excess returns for the active portfolios versus the passive portfolios.

Sector pair used in simulation	**Average annual excess return of active portfolio**
Portfolio(7,8)	$+13,66\%$
Portfolio(4,10)	$+12,10\%$
Portfolio(8,12)	$+5,11\%$
Portfolio(1,8)	$-4,66\%$
Portfolio(9,13)	$+0,46\%$

Average annual excess returns (AAER) for the chosen amount of simulated years, T, calculated as:

$$\text{AAER} = [(\frac{R_s}{R_p})^{\frac{1}{T}}] - 1. \tag{6.1}$$

R_s denotes the return of the strategy portfolio, and R_p the return of the passive portfolio, over the entire simulation period.

In figure 8, the active portfolio outperforms the passive portfolio, but the results from the first year and a half of trading are not weighty. However, in the end the algorithm betters its performance and top the passive strategy by some margin. The same tendency is apparent in figure 9 and figure 10, where the active strategy performs better with time compared to the buy-and-hold portfolio. In figure 11 and figure 12, the passive portfolio performs better than the active trading algorithm. Since the passive portfolio is more comfortable for the investor, and since it is low-cost, the buy-and-hold strategy would have been preferable in these instances.

7. Analysis

7.1 Basis vectors and anomalies

In figure 6, it can be concluded from basis vector 1 that the entire data set shows a general trend of a yearly increase. That is, the most significant pattern in the price is an up-going trend with a slight increase in prices every year. This reflects the fact that stock prices over the last decades have shown a slight increase annually. It is however important do recognize that some of the sectors of stocks might show the same pattern but in the opposite direction. But from figure 7, it is established that there seems to be no negative correlations in the first basis vector. Moreover, basis vector 2 highlights that the historical prices tend to be higher during the winter months, at least at the end of the year. This is in line with previous findings on market anomalies, presented in chapter 3 (Dimson 1988; Degenhardt and Auer 2017).

Basis vector 3, concluded important in shaping the trading strategy, show a periodicity where prices increase in the beginning of the year, to later decrease until approximately the 150:th trading day. Thereafter prices increase until the end of December, where they show a slight dip. Basis vector 6, also used as a tool of indication of negative correlations, show a high frequency periodicity. However, this basis vector is of modest significance, due to the small corresponding singular value.

Lastly, no evidence was found of longer and shorter time period anomalies. This is due to the decided period of the basis vectors U. If the aim of the paper was to solely identify anomalies, one could change the time frame from yearly trading days to for instance time at day, or a four year period.

7.2 Correlations

The basis vector correlations did not collect evidence to underpin previous results regarding sectors valuable for hedging. Possibly, the negative correlation between the Health Services sector and Non-Energy Mineral sector provides some evidence that gold can be a hedge to the general market. The conclusion is debatable for health services may not represent the general stock market accurately.

The most significant negative correlation found in the basis vectors was that between Health Technology and Health Services in basis vector 3. Possibly, better medicines and treatments results in a decrease in demand for doctor visits and check-ups. No matter for what reason, a negative correlation was found between the two sectors, which contributed to the picking of sectors for strategy implementation.

To conclude, the SVD method is suggested as a possible tool to use in the search for negative correlations. As stated formerly, stock markets as a whole seems to become more interlinked with time. Finding stocks, or sectors of stocks, with negative correlations can be a challenge of itself. The SVD approach can be helpful in this regard.

7.3 Strategy performance and market efficiency

Table 6.2 provides evidence that the active strategy based of the predictions from the SVD algorithm enables excess annual returns compared to a passive strategy. The most notable excess returns are found when the sectors used in the active portfolio exhibit strong negative correlations in significant basis vectors. When applying the strategy to sectors with negative correlations in less significant basis vectors, the results were either inferior to the passive portfolio, or modest in magnitude. Thus, the method should be implemented only to strongly negatively correlated stocks in U vectors with significance.

The excess returns suggests that historical data can be of use to an investor. Thus, this paper can not conclude that the NYSE is a weakly-efficient market. However, it should be stated that the simulation in this research consists of maximally 4 years, and that a longer period, and a broader bench-mark, could display other results. Also, the simulation does not take into account the transaction costs of trading financial assets. Ordinarily, a few points of the share price is payed for each transaction as a fee. Since the strategy developed in this paper trades approximately every second trading day (depending on the buy/sell signal frequency), for 4 years, commissions would indeed reduce the profit. On the other hand, for the top performing portfolios, the trading fees are assumed not to change the general verdict of the results, which is a generous average annual excess return of the strategy.

In summary, to implement the strategy in real life markets, further research on the strategy needs first to be done. In this paper, the focus was on the implementation of a gradient strategy from SVD-filtered data - only one of the many possible angles one can use from the SVD approach. Nevertheless, the paper can be a first step towards a reasonably reliable and accurate predictive tool. If used correctly, it can please investors with excess returns, prompted from the results in figure 8.

8. Concluding remarks

8.1 Conclusion

This research concludes that excess returns on the NYSE can be obtained from an active strategy based on the predictions from an SVD analysis. A simulation of an active trading strategy based of an SVD-algorithm compared to a buy-and-hold benchmark shows that the active portfolio performed at best an average annual excess return of 13,66% compared to the passive portfolio. The strategy was actualized using gradient differences in predicted prices. Results demonstrated that substantial negative correlations between stock sectors, in significant basis vectors, gave the most convincing excess returns. No transaction costs were incorporated in the simulation, yet the costs are assumed not to exceeded the profits of the strategy for the best performing portfolio, hence not affect the general verdict of the paper.

The results of the study underlines the usefulness of the SVD method in the financial context. Thus, the paper shows that regardless of which trading strategy is implemented, basing it of the filtered data ought to give more accurate forecasts.

8.2 Further research

The active strategy has been shown to be useful in this paper by comparison to a buy-and-hold strategy consisting of two sectors of stocks. To further investigate the strategy, one should perform the same procedure for a larger amount of sectors, and bench-mark against a broad index. Preferably, by using a wide spectrum of negatively correlated individual stocks in only significant basis vectors. Also, strategies derived from the predictive pattern of the SVD analysis using other methods than gradient differences can give improved results, thus should be investigated.

Although this study focuses on the bench-mark of the buy-and-hold strategy, the simulation could as well be compared to other strategies such as an optimal portfolio strategy based on for instance Sharpe ratio maximization, volatility minimization, et cetera. One could also compare the predictive power of the SVD analysis to other predictive tools for an accuracy benchmark. Furthermore, additional research would be needed to conclude if the SVD method can prove useful for identification of hedging possibilities as well as to find anomalies for different time periods.

Bibliography

Baker, Kirk (2005). "Singular Value Decomposition Tutorial". In: *Rough Draft*, pp. 1–24.

Barry, Christopher and Stephen Brown (1984). "Differential information and the small firm effect". In: *Journal of Financial Economics* 13 (2), pp. 283–294.

Baur, Dirk and Thomas McDermott (2009). "Is gold a safe haven? International evidence". In: *Journal of Banking & Finance* 34, pp. 1886–1898.

Brock, William, Josef Lakonishok, and Blake LeBaron (1992). "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns". In: *The Journal of Finance* 47 (5), pp. 1731–1764.

Byström, Hans (2014). *Finance*. 3rd ed. Lund: Studentlitteratur AB, pp. 200–201.

Caraiani, Petre (2014). "The predictive power of singular value decomposition entropy for stock market dynamics". In: *Physica A* 393, pp. 571–578.

Degenhardt, Thomas and Benjamin Auer (2017). "The "Sell in May" effect: A review and new empirical evidence". In: *North American Journal of Economics and Finance* 43, pp. 169–205.

Dimson, Elroy (1988). *Stock market anomalies*. 1st ed. Cambridge University Press.

Fama, Eugene F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The American Finance Association* 25 (2), pp. 383–417.

Fama, Eugene F. and Kenneth R. French (1988). "Dividend Yields and Expected Stock Returns". In: *Journal of Financial Economics* 22 (1), pp. 3–25.

Gärtner, Manfred and Klaus W. Wellershoff (1995). "Is there an Election Cycle in American Stock Returns?" In: *International Review of Economics and Finance* (4), pp. 387–410.

Goodman, David A. and John W. Peavy (1983). "Industry Relative Price-Earnings Ratios as Indicators of Investment Returns". In: *Financial Analyst Journal* 39 (4), pp. 60–66.

Gu, Rongbao, Wei Xiong, and Xinjie Li (2015). "Does the singular value decomposition entropy have predictive power for stock market? — Evidence from the Shenzhen stock market". In: *Physica A*, pp. 103–113.

Henry, E. R. and J. Hoffrichter (1992). "Singular Value Decomposition: Application to Analysis of Experimental Data". In: *Methods in Enzymology* 210 (8), pp. 138–139.

Jegadeesh, Narasimham and Sheridan Titman (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency". In: *The Journal of Finance* 48 (1), pp. 65–91.

Kahneman, Daniel (2011). *Thinking fast and slow*. 1st ed. New York: Farrar, Straus and Giroux.

Keim, Donald B. and Robert F. Stambaugh (1986). "Predicting returns in the stock and bond markets". In: *The Journal of Financial Economics* 17, pp. 357–390.

Kenett, Dror Y. and et. al. (2011). "Index Cohesive Force Analysis Reveals That the US Market Became Prone to Systemic Collapses Since 2002". In: *PLoS ONE* 6 (4).

Lewis, Michael (2015). *Flash Boys: A Wall Street Revolt*. 1st ed. New York: W.W Norton Company.

Maheshwari, Suneel, Rakesh Gupta, and Li (2017). "A Comparative Analysis of Sector Diversification in Australia, India and China". In: *Financial Planning Research Journal*, pp. 76–100.

Malkiel, Burton G. (2019). *A Random Walk Down Wall Street*. 14th ed. New York: W.W Norton & Company Inc., pp. 136–137.

Menkhoff, Lukas (2010). "The use of technical analysis by fund managers: International evidence". In: *Journal of Baking & Finance* 34, pp. 2573–2586.

Olson, Eric, Andrew Vivian, and Mark E. Wohar (2017). "Do commodities make effective hedges for equity investors?" In: *Research in International Business and Finance* 42, pp. 1274–1288.

Sandoval, Leonidas Jr. and Italo De Paula Franca (2011). "Correlation of financial markets in times of crisis". In: *Physica A* 391, pp. 187–208.

Taylor, Mark P. and Helen Allen (1992). "The use of technical analysis in the foreign exchange market". In: *Journal of International Money and Finance* 11, pp. 304–314.

Thaler, Richard and Werner F.M De Bondt (1984). "Does the Stock Market Overreact?" In: *The Journal of Finance* 40 (3), pp. 793–805.

A. Data

Table A.1: Complete list of stocks, acquired from WRDS.

Stock	Ticker	Sector	Subsector
Apple computer Inc	AAPL	Electronic Technology	Telecommunications Equipment
American Express Co	AXP	Financial Services	Finance/Rental/Leasing
Applied Materials Inc	AMAT	Producer Manufacturing	Industrial Machinery
Berkshire Hataway Inc	BRK.B	Financial Services	Multi-Line Insurance
Federal Express Corp	FDX	Transportation	Air Freight/Couriers
Gap Inc	GPS	Retail Trade	Apparel/Footwear Retail
Intel Corp	INTC	Electronic Technology	Semiconductors
McCormick and Co Inc	MKC	Consumer Non-Durables	Candy
Nike Inc	NKE	Customer Non-Durables	Apparel/Footwear
Nordstrom Inc	JWN	Retail Trade	Apparel/Footwear Retail
Thermo Electron Corp	TMO	Health Technology	Medical Specialities
Home Depot Inc	HD	Retail Trade	Home Improvement Chains
United Healthcare Corp	UNH	Health Services	Managed Health Care
Ross Stores Inc	ROST	Retail Trade	Apparel/Footwear
Costco Wholesale Corp	COST	Retail Trade	Speciality Stores
Oracle Systems Corp	ORCL	Technology Services	Packaged Software
Microsoft Corp	MSFT	Technology Services	Packaged Software
Newscope Rescources Ltd[5]	DEN	Energy Minerals	Integrated Oil

Yahoo Inc	YOJ	Technology Services	Internet Software/Services
Amazon Com Inc	AMZN	Retail Trade	Internet Retail
Ryanair Holdings Plc	RYAAY	Transportation	Airlines
Ebay Inc	EBAY	Consumer Services	Other Consumer Services
Nvidia Corp	NVDA	Electronic Technology	Semiconductors
Alaska Airgroup Inc	ALK	Transportation	Airlines
American Barrick Resources Corp	GOLD	Non-Energy Minerals	Precious Metals
Ball Corp	BLL	Process Industries	Containers/Packaging
Boeing Co	BA	Electronic Technology	Aerospace and Defense
Caterpillar Inc	CAT	Producer Manufacturing	Trucks/Construction/Farm Machinery
Chemical Banking Corp[6]	JPM	Financial Services	Major Banks
Chevron Corp	CVX	Energy Minerals	Integrated Oil
Coca Cola Co	KO	Consumer Non-Durables	Non-Alcoholic Beverages
Colgate Palmolive Co	CL	Consumer Non-Durables	Household/Personal Care
Crown Cork and Seal Co Inc	CCK	Process Industries	Containers/Packaging
Disney Walt Co	DIS	Consumer Services	Cable/Satellite TV
Dun and Bradstreet Corp	DNB	Technology Services	Packaged Software
Exxon Mobil Corp	XOM	Energy Minerals	Integrated Oil
F P L Group[7]	NEE	Utilities	Electric Utilities
Ford Motor Co Del	F	Consumer Durables	Motor Vehicles
General Electric Co	GE	Electronic Technology	Aerospace and Defense
Humana Inc	HUM	Health Services	Managed Health Care

Johnson and Johnson	JNJ	Health Technology	Pharmaceuticals
McDonalds Corp	MCD	Consumer Services	Restaurants
Newmont Mining Corp	NEM	Non-Energy Minerals	Precious Metals
Norwest Corp[8]	WFC	Financial Services	Major Banks
Pfizer Inc	PFE	Health Technology	Pharmaceuticals
Sony Corp	SONY	Consumer Durables	Electronics/Appliances
Texas Instruments	TXN	Electronic Technology	Semiconductors
Unilever PLC	UL	Consumer Non-Durables	Household/Personal Care
Walmart Stores Inc	WMT	Retail Trade	Food Retail
Allied Signal Inc[9]	HON	Producer	Manufacturing Industrial Conglomerates
Abercrombie and Fitch Co	ANF	Retail Trade	Apparel/Footwear Retail
Grand Metropolitan Plc (Diageo)	DEO	Consumer Non-Durables	Alcoholic Beverages
Autozone Inc	AZO	Retail Trade	Speciality Stores
Allstate Corp	ALL	Financial Services	Proptery/Causalty Insurance
AstraZeneca PLC	AZN	Health Technology	Pharmaceuticals
Realty Income Corp	O	Financial Services	Real Estate Investment Trusts
CEMEX S A DE C V	CX	Non-Energy Minerals	Construction Materials
AMB Property[10]	PLD	Financial Services	Real Estate Investment Trusts
Steelcase Inc and	SCS	Producer Manufacturing	Office Equipment/Supplies
American Tower Systems Corp	AMT	Financial Services	Real Estate Investment Trusts

Trex Inc	TREX	Non-Energy Minerals	Forest Products
Goldman Sachs Group Inc	GS	Financial Services	Investment Banks/Brokers
United Parcel Service Inc	UPS	Transportation	Air Freight/Couriers
Metlife Inc	MET	Financial Services	Life/Health Insurance
Companhia Vale De Rio Doce	VALE	Non-Energy Minerals	Steel
Entegris Inc	ENTG	Electronic Technology	Semiconductors
Kraft Foods Inc	KHC	Consumer Non-Durables	Major Diversified Food
Accenture Plc	ACN	Technology Services	Information Technology Services
Credit Suisse Group	CS	Financial Services	Major Banks
Anthem Inc	ANTM	Health Services	Managed Health Care
Advance Auto Parts Inc	AAP	Retail Trade	Speciality Stores
Aluminium Corp China Ltd	ACH	Non-Energy Minerals	Aluminium
Prudential Financial Inc	PRU	Financial Services	Financial Conglomerates
China Life Insurance Co Ltd	LFC	Financial Services	Life/Health Insurance
Morningstar Inc	MORN	Financial Services	Investment Managers
Alliance Resource Partners LP	ARLP	Energy Minerals	Coal
Philip Morris International Inc	PM	Consumer Non-Durables	Tobacco
Visa inc	V	Financial Services	Finance/Rental/Leasing
Aneheuser Busch Inbev SA NV	ABI	Consumer Non-Durables	Alcoholic Beverages
Renewable Energy Group Inc	REGI	Process	Industries Chemicals

Facebook Inc	FB	Technology Services	Internet Software/Services
Twitter	TWTR	Technology Services	Internet Software/Services
Santander Consumer USA Holdings Inc	SC	Financial Services	Finance/Rental/Leasing
Alcoa Coirp	AA	Non-Energy Minerals	Aluminium
Kimbell Royalty Partners LP	KRP	Miscellaneous	Investment Trusts/Mutual Funds
Dowdupont Inc	DD	Process Industries	Chemicals

[5] The name was changed in 1985 to become Denbury Resources Inc.
[6] Merged with Chase Manhattan Corporation in 1996. Later J.P Morgan Chase and Co.
[7] Changed name to NextEra Energy Inc. in 2010.
[8] Merged with Wells Fargo and Co. in 1998.
[9] Adopted the name Honeywell in 1999 after a merger.
[10] Became Prologis after a merger in 2011.

B. Strategy and Matlab code

This section presents a more detailed description of the trading strategy implemented in this paper. First, the general steps in data processing, followed by the daily activities by the algorithm.

1. Select period for historical data to be used in SVD procedure.

2. Compute the SVD to derive the Basis Vectors (U-Vectors).

3. Select the basis vectors to be used for the process. These vectors are in the present analysis not changed over the simulation period. However, a natural updating would be on an annual basis.

4. Select the historical time period for basis vector amplitude estimates, and apply a time window on the stock price data (if desired). If using a time window, it is advisable to use the same time window also for the U-vectors.

The algorithm would additionally, for each trading day, perform the following steps:

1. Make best estimate of scaling factor for each basis vector (by the use of e.g., least mean square error method). The scaling factor determines the contribution of each basis vector to the historical price data for the specific sector of stocks in the study.

2. Derive the gradients of asset prices from the basis vectors and scaling factors.

3. Evaluate the gradient to identify if trade should be made (asset re-distribution), i.e. if the difference between the gradients are significant enough depending on the chosen limit.

4. If asset re-distribution is favorable according to the method; trade the stocks.

Matlab code may be shared on request for academic purposes. Contact `Ax2000gu-s@student.lu.se`.