Stockholm School of Economics
Department of Economics
5350 Master's Thesis in Economics
Academic Year 2017–2018

Risk-Managed Momentum Strategy Using Support Vector Machines

Patrick Daniel Schneeberger (41214)*

Abstract

Investment decisions are difficult to make, given the uncertainty about the future. For the purpose of reducing that uncertainty, I investigate, for one, how the consumer price index and the return on the 3-month US Treasury bill can be used by support vector machines to make monthly directional trend predictions of a value-weighted portfolio of stocks traded at AMEX, NYSE and NASDAQ. For another, I examine to which extent the risk-managed momentum strategy proposed by Barroso and Santa-Clara (2015) can be improved when those predictions are incorporated. I find that the monthly stock market prediction accuracy lies at 61.1 percent over the time horizon from 1965 to 2017. A prediction-based flexible volatility target in the risk-managed momentum strategy achieves an improvement in the higher order moments as well as in the Sharpe ratio which in turn reduces the crash risk of momentum.

Keywords: Financial Market, Investment Decisions, Momentum Strategy, Support Vector Machines

JEL: C38, G01, G10, G11

Supervisor: Federica Romei

Date submitted: 9th May 2018

Date examined: 29th May 2018

Discussant: Jonas Nauerz

Examiner: Maria Perrotta Berlin

*41214@student.hhs.se

Acknowledgements

I would like to thank my thesis supervisors Prof. Federica Romei, Ph.D. of the Economics Department at the Stockholm School of Economics, and Prof. Paul Söderlind, Ph.D. of the Swiss Institute of Banking and Finance at the University of St.Gallen, for always having made time for me when I experienced difficulties – despite their various commitments.

I must also express my gratitude to my family for providing me with tireless support throughout my years of study and through the process of researching and writing this thesis. Particular thanks go to my brother Dominik, who endured countless hours of discussions and always motivated me when needed.

Contents

List of Figures

List of Tables

List of Abbreviations

AMEX	American Stock Exchange
CPI	Consumer Price Index
EMH	Efficient Market Hypothesis
KURT	Excess Kurtosis
NASDAQ	National Association of Securities Dealers Automated Quotations
NYSE	New York Stock Exchange
SKEW	Skewness
SR	Sharpe Ratio
SVC	Support Vector Classifier
SVM	Support Vector Machine
WML	Winners-Minus-Losers

1 Introduction

Investment decisions would be straightforward if an investor had complete information about the future. Unfortunately, investment decisions must be made under uncertainty, for the future has yet to unfold. Accurate predictions would reduce that uncertainty but, as Niels Bohr ironically states, "It's tough to make predictions, especially about the future."[1] Despite this problematique, it is not a matter of impossibility. Indeed, when it comes to financial time series applications, the classification method of support vector machines has proven to provide very promising results (Theofilatos, Georgopoulos, Likothanassis & Mavroudi, 2014).

This leads to the two-fold research question: First, to which extent are support vector machines capable of predicting a stock market's directional trend when using macroeconomic input features on a monthly basis? Second, how much can an investment strategy, namely the risk-managed momentum strategy proposed by Barroso and Santa-Clara (2015), be improved when those predictions are incorporated? I will develop a model along these two questions but in reverse order; doing so elucidates more clearly the momentum strategy's link with the support vector machine predictions.

Those questions will be answered by investigating the US stock market. Using support vector machines, I predict the directional trend of the value-weighted portfolio constructed out of all stocks traded at AMEX, NYSE and NASDAQ as these are the stocks which are used to build the momentum portfolios. The monthly percentage change in the consumer price index as well as the monthly percentage change in the return on the 3-month US Treasury bill are used as macroeconomic input features for the support vector machines. Subsequently, I incorporate these findings into the risk-managed momentum strategy proposed by Barroso and Santa-Clara (2015), which in turn is based on the findings of Jegadeesh and Titman (1993). Barroso and Santa-Clara found that around 77 percent of the risk of momentum is attributed to specific risk and even though it is highly variable over time, it can be predicted to some degree. Their model uses monthly volatility forecasts of momentum returns to scale the exposure to the momentum risk towards a certain target level in order to have a constant risk over time. They chose a target corresponding to an annualised volatility of 12 percent, whereas I let this target to be flexible and dependent on the stock market's directional trend prediction. When the market is predicted to flourish, I allow for a higher volatility target and for a lower if otherwise.

Thereafter, I analyse the performance of the following investment strategy. At first an investor

[1] It is unclear who deserves the laurels. Niels Bohr, Samuel Goldwyn, K. K. Steincke, Robert Storm Petersen, Yogi Berra, Mark Twain and Nostradamus are worth considering.

puts exactly one dollar into the risk-free asset. Simultaneously, he invests a certain percentage of that risk-free investment in the momentum portfolio.[2] Each month, this strategy reinvests the accumulated wealth in the risk-free rate and puts anew a certain percentage of this investment in the momentum portfolio.

Given the availability of data, I am able to apply the above mentioned strategy on the time horizon beginning in February 1965 and ending in December 2017. Predictions of the support vector machines prove to be accurate 61.1 percent of the time, which reduces the uncertainty about the future. The incorporation into the risk-managed momentum strategy results in an increase in the Sharpe ratio from 0.752 to 0.832 when compared to the model proposed by Barroso and Santa-Clara (2015). Furthermore, the higher order moments improve as well, which in turn leads to a lower crash risk of momentum.

The remainder of this thesis is as follows. Chapter two contains a review of the existing literature in the field of momentum strategy as well as of support vector machines. Chapter three elaborates on all components of my model on the risk-managed momentum strategy with a flexible target volatility. Chapter four discusses the data. In chapter five, the results are presented and discussed while chapter six investigates the robustness of these results. Chapter seven closes the thesis with a conclusion.

[2] This percentage depends on whether the standard momentum strategy by Jegadeesh and Titman (1993), the risk-managed momentum strategy by Barroso and Santa-Clara (2015) or my strategy using different scaling methods is applied.

2 Economic Background and Motivation

My model makes a two-fold contribution to the academic literature. On the one hand, it demonstrates how macroeconomic data can be used by support vector machines [SVMs] in order to make one-month ahead forecasts of the stock market's trend direction. On the other hand, it shows that the incorporation of these findings into a risk-managed momentum strategy leads to an improvement in the Sharpe ratio as well as in the higher order moments.

I build on the momentum strategy proposed by Barroso and Santa-Clara (2015), which has, in contrast to the standard momentum strategy of Jegadeesh and Titman (1993), some sort of risk-management involved. The aim of this risk management is to steer the portfolio's volatility to a previously determined target in order to achieve a constant risk over time. This is achieved by scaling the weights invested in the winners-minus-losers [WML] portfolio by the fraction of the target over the portfolio's monthly volatility forecast. Barroso and Santa-Clara opted for a 12 percent annualised volatility as their target level. This is exactly the point where I enhance their strategy by using a flexible instead of a constant volatility target level. The trend prediction of the stock market movement determines the level of the target: more volatility is allowed if the prediction is positive and less if the forecast is negative.

Kim (2003) showed that support vector machines give promising results when it comes to financial time series predictions. His empirical findings on the daily Korean composite stock market index are that SVMs outperform other prediction methods such as back-propagation neural networks and case-based reasoning. Likewise, I take the SVM approach to get stock market predictions but on a monthly basis. The aim is to predict the trend direction of a value-weighted portfolio, which is constructed out of all the stocks traded at AMEX, NYSE and NASDAQ stock exchanges. I use the monthly percentage change in both the consumer price index [CPI] and the return on the 3-month US Treasury bill as input features for the SVMs. The output of the model is either an *up* or *down* movement prediction of the value-weighted portfolio. Confident predictions of the future trend-direction of these stocks – which potentially are part of the momentum strategy –, allow an investor to appropriately adjust the volatility target from the momentum strategy and thereby to optimise his profits.

All in all, my empirical research shows that predictions made by SVMs can be incorporated into a risk-managed momentum strategy, and doing so helps to improve it in terms of lowering an investor's risk. The following two subchapters provide a broader overview on the current state of literature regarding momentum strategy as well as support vector machines.

2.1 Momentum Strategy

Momentum is a measure of the increase or decrease of a stock price over a given time horizon. Stocks that grow fast over a certain period are considered to have a positive momentum or a negative momentum if otherwise. Jegadeesh and Titman (1993, 2001) and Chan, Jegadeesh and Lakonishok (1996) demonstrated that stocks having performed well in the past continued to do so over the next year. Additionally, they found that stocks with a higher momentum underperform in the long term, that is, in the subsequent 24 to 60 months. These findings led to a so-called momentum strategy of buying past winners and selling past losers which has proven to generate good returns.

While Jegadeesh and Titman exclusively investigated the US stock market, other researchers examined in-depth the European stock market. They find, even tough Europe differs from the United States in its social, cultural and economic environment, that momentum can be observed in the European stock market as well. Research in this field has shown that returns to momentum can be found across asset classes, time and country boarders. (Schiereck, De Bondt & Weber, 1999; van Dijk & Huibers, 2002; Rouwenhorst, 1998; Hou, Karolyi & Kho, 2011)

Furthermore, Herberger, Kohlert and Oehler (2011) found that the momentum trading strategy is nothing that belongs to the past, for it is nowadays still possible to generate superior returns by applying it. But why do higher momentum stocks generate superior returns? The literature splits on this question into two fields. On the one hand, there is the risk-based explanation stating that those higher returns come along with higher risk. For instance, Chan et al. (1996) argued that those profits are generated through informational asymmetries in the financial markets. Chordia and Shivakumar (2002) asserted that momentum does not *per se* represent a market risk in itself but potentially correlates to an unobserved source of risk and thus serves as a proxy of this risk. They found that many macroeconomic factors are correlated with momentum and regard this as a plausible explanation for those higher returns on momentum.

On the other hand, behavioural economists argue that it is the investment decision relying on some sort of behavioural bias that leads to higher returns. That said, investors are viewed as irrational and thus fail to take into account lagging macroeconomic effects. Some behavioural observations on this issue have been made. For one, people are slow to incorporate new information and, for another, tend to jump on the bandwagon. Furthermore, they are likely to over- or underreact to new information, which leads to selling winning stocks too early or holding on to losing stocks for too long. (Frazzini, 2006; Kahneman & Tversky, 1982; Shiller, 1981; De Bondt & Thaler, 1985)

Despite the disparate opinions on the cause of superior returns of momentum, its effectiveness is undoubted. But what happens when we take trading costs into account? Sagi and Seasholes (2007) as well as Ammann, Moellenbeck and Schmid (2010) investigated on how the profitability of the momentum strategy changes when transaction costs are considered. After doing so and even additionally adjust for risk measures, they still found significantly high returns.

Furthermore, Barroso and Santa-Clara (2015) investigated on how much risk is attributed to both market and strategy. It turned out that only 23 percent of the risk involved in the momentum strategy is explained by the former while the rest is specific to the latter. Furthermore, this specific risk (77 percent) is persistent and predictable; therefore, an investor may indeed take this risk into account. They also found that the monthly volatility forecast for the WML portfolio is very well approximated by the average monthly volatility over the previous six months which is calculated on the basis of daily returns. This volatility forecast can then further be used in order to scale the exposure to the WML portfolio which in turn allows an investor to get a constant level of volatility over time. I will exclusively enhance the approach of Barroso and Santa-Clara (2015) for two reasons. On the one hand, their strategy serves as a good baseline model, as it implements a scaling of the exposure to the momentum risk. On the other hand, this leaves room for improvement by allowing the target level to be dependent on the stock market's directional trend predictions. Support vector machines, for instance, constitute such a prediction method.

2.2 Support Vector Machines

The financial sector relies increasingly heavily on machine learning algorithms when it comes to modelling and trading of financial indices. Theofilatos et al. (2014) showed that the conventional artificial neural networks reach their limits when solving problems with multiple inputs. The technique of support vector machines, which was originally developed by Vapnik (1995), remedies these problems, for SVMs were precisely designed to solve classification problems in the n-dimensional space. The goal is to find a structural model, which has little risk of miss-specifying out-of-sample data.

Even though support vector machines were already developed in the early 1990s, only as early as in the year 1997 did SVMs find their way into time series applications (Moskowitz, Ooi & Pedersen, 2012, pp. 999–1004). Cao and Tay (2001), for example, used them to forecast future contracts from the Chicago Mercantile Exchange while Kim (2003) was able to predict the direction of the Korean composite stock index with the help of SVMs. They both independently showed that SVMs outperform back-propagation neural networks. Their results indicate that support vector machines are considerably valuable when it comes to financial time series forecasting. In 2005, Huang, Nakamori and Wang found strong results by applying SVMs on the Nikkei 225 index in order to predict its weekly up and down movement.

Now we proceed with the construction of a model that combines the risk-managed momentum strategy of Barroso and Santa-Clara (2015) with a prediction-based flexible target volatility.

3 Model

In this chapter, a model of risk-managed momentum strategy combined with a flexible target volatility will be developed. I will start with the basic momentum strategy which has been widely discussed by Jegadeesh and Titman (1993). Thereafter, relying on the findings of Barroso and Santa-Clara (2015), I will elaborate on the implementation of the risk-management into the momentum strategy. After a discussion of all the relevant parts of the momentum strategy which are necessary to build my own model, I will finally approach the machine learning technique of support vector machines to predict the directional trend of the stocks of interest – there, I rely mainly on James, Witten, Hastie and Tibshirani (2017, chapter 9).

3.1 Basic Momentum Strategy

As mentioned in the literature review, people are likely to either over- or underreact to information.[3] Based on this behaviour, it is possible to develop profitable trading strategies that are built on the history of stock returns. In particular, Jegadeesh and Titman (1993) investigated such strategies in-depth. One of their main findings is that there exists a way of creating a portfolio such that significant positive returns over a holding period of three to twelve months are generated. The key lies in not only holding long positions in those stocks, which have performed well over the past few months, but also holding short positions in those that have performed poorly. Such portfolios which are long in the past winners and short in the past losers are well-known in the finance literature as momentum portfolios or winners-minus-losers portfolios. They are zero-cost investments.

Jegadeesh and Titman (1993) analysed such momentum portfolios on stocks that are traded at NYSE and AMEX for the time period from January 1965 to December 1989. But how are those WML portfolios constructed? In each month t, all the stocks, which are traded at NYSE and AMEX, are listed based on their returns over the last J-months in ascending order, whereas J can either be 3, 6, 9, or 12. The next step is to form ten decile portfolios that are equally weighted. We call the top decile *losers,* for they had the least returns over the past J-months, and the bottom decile *winners,* for they had the largest returns. At time t, the strategy then takes a long position in the winner portfolio and a short position in the loser portfolio. These long and short positions will subsequently be held for K-months, where K can also either be 3, 6, 9 or 12. With the mentioned formation and holding periods ($J = 3, 6, 9, 12$; $K = 3, 6, 9, 12$), one can build 16 different momentum portfolios. For example, the strategy that was built on the

[3] See Shiller (1981), Kahneman and Tversky (1982) as well as De Bondt and Thaler (1985).

returns over the previous quarter can be held for either one, two, three or four quarters. This already yields four different strategies.

Because the holding periods are longer than one month, an investor using such a strategy holds at every point in time t multiple portfolios, namely the one selected at time t and all the ones that had been selected in the previous $(K-1)$-months. Therefore, the strategy does not only select a new WML portfolio at time t, it also closes out old positions, namely those initiated in month $t-K$. Interestingly, Jegadeesh and Titman (1993) found that the most successful zero-cost strategy out of those 16 strategies is the one with a formation and holding period of 12 and 3 months, respectively.

In addition to those 16 strategies, they analysed the same portfolios again but then implementing a one-week lag between the formation period and the holding period in order to avoid some of the bid-ask spread as explained by Lehmann (1990). Doing so improved the afore-mentioned portfolio's return from 1.31 percent to 1.49 percent.

In this thesis, I will not weigh up the performance of momentum portfolios of different formation and holding periods against each other, for my contribution lies on the implementation of SVMs. The reason for pointing to Jegadeesh and Titman (1993) is that their research created the foundation for the momentum strategy on which Barroso and Santa-Clara (2015) further built up and eventually developed the risk-managed strategy.

3.2 Risk-Managed Momentum Strategy

Having conveyed the basics of the momentum strategy, I will take a further step forward by introducing the concept of risk-management into the momentum strategy. For this purpose, we shall start by answering the following three questions: how can the risk of a portfolio be measured? What is the advantage of controlling this risk? And perhaps the most interesting question is, *how* can it be controlled? In the financial literature, a portfolio's variance indicates its risk. The variance describes the strength of the amplitude around the data's mean – its square root (standard deviation) is also well known under the term *volatility.* An increase in the portfolio's variance implies higher risk, as the probability of large gains as well as large losses are more likely. In short, a high (low) variance implies high (low) fluctuations around the mean and therefore more (less) uncertainty about the future path of that portfolio. To control the risk of an investment, one needs to predict the expected future volatility of the portfolio of interest. With such a forecast, one can scale the exposure to the strategy and thereby control the risk.[4] A

[4] Such volatility-scaling methods were already investigated by Moskowitz et al. (2012) as well as Baltas and Kosowski (2013), whereas they mainly applied these methods on asset-specific volatilities.

higher volatility increases the exposure whereas a lower one reduces the risk of the investment. I will build up on the findings of Barroso and Santa-Clara (2015), who used such a scaling method for WML portfolios in order to have a constant volatility over time.

They decomposed the risk of momentum into market and specific risk and found that the market component accounts only for 23 percent of the total risk and a large part of the risk of momentum is specific, that is, depending on the strategy. Furthermore, they found that the risk of momentum is highly variable over time and is predictable to some degree.

But why do we even need such a volatility-scaling despite the findings of Jegadeesh and Titman (1993) showing that WML portfolios are quite profitable? Barroso and Santa-Clara (2015) demonstrated in their research paper, where they analysed the same stocks as Jegadeesh and Titman, that WML portfolios mainly generated considerable profits but once in a while the strategy was fatally hit by crashes (Daniel & Moskowitz, 2016). Worth mentioning are the ones that occurred in 1932 and 2009, where the WML strategy delivered a -91.59 and a -73.42 percent cumulative return over two and three months, respectively. Imagine you had invested one dollar in the WML portfolio in July 1932, you would have recovered from that crash only by April 1963, which is a time-horizon of 31 years (Barroso & Santa-Clara, 2015, pp. 1–6). The aim of the volatility-scaling is to bring down exactly this risk to get hit by such huge crashes. Although WML portfolios are profitable, they occasionally suffer severely from recessions. Given that this risk is only partly market-related and mainly strategy-related, it can actually be controlled for by means of volatility-scaling.

Now let me reveal the *open secret* behind the risk-management of momentum. Barroso and Santa-Clara (2015, pp. 10–12) estimated the monthly variance forecast on the basis of the daily returns from the previous six months. This variance forecast looks as follows.[5]

$$\hat{\sigma}^2_{WML,t} = \frac{21}{126} \sum_{j=0}^{125} (r_{WML,d_{t-1}-j} - \bar{r}_{WML})^2, \tag{1}$$

where

- $\hat{\sigma}^2_{WML,t}$ is the monthly variance forecast,
- $\{r_{WML,d}\}_{d=1}^{D}$ are the daily returns,
- $\bar{r}_{WML}$ is the mean of the WML returns of the previous 126 days[6] and
- $\{d_t\}_{t=1}^{T}$ is the time series of the dates of the last trading session of each month.

[5] I follow the notation of Barroso and Santa-Clara (2015).

[6] It is calculated as follows: $\bar{r}_{WML} = \frac{1}{126} \sum_{j=0}^{125} r_{WML,d_{t-1}-j}$.

According to Barroso and Santa-Clara (2015): "As WML is a zero-investment and self–financing strategy we can scale it without constraints" (p. 11). We can now use the square root of equation (1) in order to scale the returns of the momentum portfolio the following way.

$$r_{WML^*,t} = \frac{\sigma_{target}}{\hat{\sigma_t}} r_{WML,t}, \tag{2}$$

where

- $r_{WML,t}$ is the unscaled monthly return of the momentum,
- $r_{WML^*,t}$ is the scaled/risk-managed momentum and
- σ_{target} is a constant number towards which the volatility shall be steered.

Barroso and Santa-Clara (2015, p. 11) chose a constant target of an annualised volatility of 12 percent when scaling the momentum portfolios.[7] The fraction $\frac{\sigma_{target}}{\hat{\sigma_t}}$ then provides a number that determines the positions in the long-short portfolio.

By applying this technique, Barroso and Santa-Clara (2015, p. 12) found that for the time period 1927 to 2011, the volatility-scaled momentum portfolio improved in comparison to the standard momentum strategy in the following way. First, the Sharpe ratio went up from 0.53 to 0.97. Second, the excess kurtosis dropped from 18.24 to 2.68 and the left skewness improved from -2.47 to -0.42, which clearly indicates that huge crashes affect this type of improved momentum strategy much less strongly. Third, the lowest one-month return was only -45.20 percent, whereas it was -96.69 percent for the standard momentum portfolio. Although these results demonstrate a high performance of the volatility-scaled momentum strategy, there is still room for improvement.

For instance, my own model allows a flexible target volatility, which is solely determined by the stock market movement predictions made by support vector machines.[8] I will test four different possibilities of incorporating the stock market predictions into the volatility target in order to make the latter flexible. The 12 percent target volatility decided by Barroso and Santa-Clara (2015, p. 11) serves as a basis. Then, if the stock market's prediction is positive, x percentage points are added to the 12 percent target volatility and if the stock market's prediction is negative, x percentage points are subtracted from the target level. For example, if x is one, then the target volatility is 13 when the stock market is predicted to go up and 11 if it is predicted to go down. My first three scaling methods work exactly as described above, whereas the first scaling method uses $x = 1$, the second $x = 2$ and the third $x = 3$. However,

[7] No reasoning for choosing a 12 percent annualised target volatility was provided.

[8] How the stock market movement will be predicted via SVMs will be explained shortly.

this method is flawed in that it treats a single down prediction in the same way as it treats multiple consecutive down predictions, because the target level is completely independent from the previous ones. The same holds true for up predictions.

My last scaling method fixes this issue in the following way. I either add or subtract now a certain value from the one of the last month's target volatility. The start value at the initial period is again set to 12 percent. If the stock market's prediction is positive, I add one percentage point and if it is negative, I subtract one percentage point. Furthermore, I set an upper limit of 15 percent and a lower bound of 9 percent. I additionally include a constraint correcting the current months target level to 12 percent if the prediction is down and the previous months target level is above 12 percent. This shall help to react faster to down predictions.[9]

We reach the preliminary conclusion that holding on those momentum strategies of Jegadeesh and Titman (1993) is profitable but suffers severely during economic crashes. Barroso and Santa-Clara (2015) implemented some sort of risk-management by introducing volatility scaling into the strategy and demonstrated that doing so clearly improves the performance of WML portfolios.

Having now completed the part about the momentum strategy itself, I will subsequently approach the machine learning part. With the help of support vector machines, for which I use macroeconomic indicators as input features, I will create a prediction model for the trend direction of the value-weighted portfolio of the stocks of interest, namely the ones traded at AMEX, NYSE and NASDAQ. This trend direction will then be used to adjust the level of the target volatility according to the prediction of the market movement. Depending on whether the SVM's output predicts the stocks to go up or down next month, I allow for more or less volatility, respectively.

3.3 Support Vector Machines

Recall that financial time series are very noisy, non-stationary and chaotic. Therefore, it is very difficult to make an accurate prediction of the future behaviour of a financial time series by using information only of its past.[10] Vapnik (1995) revolutionised the field of classification techniques by inventing the concept of support vector machines which has gained in popularity since the 1990s. According to James et al. (2017, chapter 9), SVMs are among the best methods when it comes to classification and pattern recognition problems. Subsequently, we will see in more

[9] Admittedly, the upper and lower bound of 15 and 9 percent, respectively, are arbitrarily chosen but so would be any other.

[10] See Deboeck (1994) as well as Abu-Mostafa and Atiya (1996).

detail that even the complex stock market prediction problem can be broken down into a binary classification problem. A benefit of using SVMs over other methods is that any kind of input feature can be used. Applied on the prediction of the future path of a specific stock market, one could potentially use input features such as technical, inter-market or macroeconomic indicators.

As my intention in this chapter is mainly to convey the concept of SVMs rather than to reinvent the wheel, I will follow closely the explanations made by James et al. (2017, chapter 9).[11] Before delving into the world of SVMs, I start with an example illustrating how a support vector classifier [SVC] works.[12]

Linearly Separable Data

Imagine that while you and your friends are walking down the streets, an animal suddenly appears. Unfortunately, none of you is certain about the type of animal. However, it seems to be either a cat or a dog. In order to identify which one it is, you may make use of support vector classifiers. By chance, one of your friends has a data set available, which contains information about dogs, cats and some of their characteristics such as body size and nose shape. The idea is to use this data set in order to assign the unknown animal to either class (cat or dog). The procedure involves three steps. First, you plot your data and mark the animals that are classified as dogs with a blue dot and those which are characterised as cats with a red dot. Second, you search for the specific line which perfectly separates those two classes – this line is called maximal margin hyperplane. (The tiny dots in Figure 1 on the following page indicate which side of the hyperplane belongs to which classification while the larger dots are generated from that data set about cats and dogs.) Third, you take the measurements from that unknown animal about its size and the shape of its nose and plot it as well into the graph. Depending on which side of the hyperplane the point lies on, the animal is expected to belong to this or that species. If it falls into the blue area, it would be classified as a dog; and as a cat if it falls into the red area. Ultimately, you and your friends would come to the conclusion that the animal is, given its nose shape and size, a dog (see the blue dot that is surrounded by a black square in Figure 1 on the next page).

This example illustrates how SVCs are used to solve a binary classification problem. Let us

[11]If not indicated otherwise, the subsequent explanations in the balance of this chapter are based on James et al. (2017, chapter 9).

[12]Support vector classifier is the precursor of support vector machines. SVCs are used in the linearly separable case such as in the following cat/dog example. When we deal with non-separable data sets, kernels are used in order to enlarge the feature space; in this case, where a SVC is combined with a kernel, we speak of SVMs.

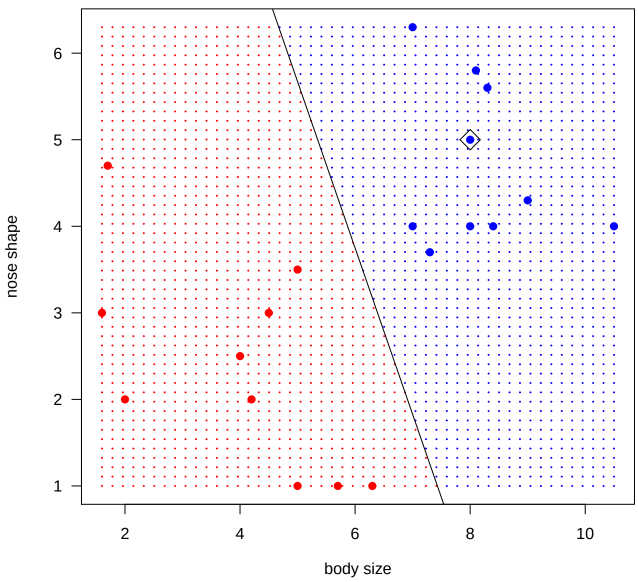


Figure 1: Classification Example

now begin to develop the SVM theory more or less from scratch by starting with the definitions of input features and classifiers.[13] An input feature is basically just an observation, which is numerically listed in a matrix X that consists of

$$x_1 = \begin{pmatrix} x_{1,1} \\ \vdots \\ x_{1,p} \end{pmatrix}, \ldots, x_n = \begin{pmatrix} x_{n,1} \\ \vdots \\ x_{n,p} \end{pmatrix},$$

whereas n stands for the number of observations and p describes how many different input features are used. Those input features are then usually linearly re-scaled into a range of $[-1, 1]$. This normalisation process is necessary to get all input features into the same range, which avoids that large value inputs would dominate low value inputs. (Kim, 2003, p. 310) As I use two macroeconomic input features, which both describe a monthly change in percentages and thus are already on the same scale, I do not need to apply such a linear scaling method. These data of the input features are listed in an $n \times p$ matrix in the p-dimensional space.[14] These observations are then further assigned to one out of two classes. Mathematically, this is written the following way: $y_1, \ldots, y_n \in \{-1, 1\}$ whereas -1 as well as 1 describes numerically one class, respectively.

[13] In this part, I follow closely the explanations made by James et al. (2017, chapter 9) as well as their labels.

[14] In our cat and dog classification example, we had $p = 2$ as there were two input features, namely body size and nose shape.

Referring to the cat/dog example, this would be equivalent to assign all the cats to class -1 and all the dogs to class 1. This data set X, which is called training data, is used to derive the parametrisation of the maximal margin hyperplane. This parametrisation is then used for classifying the so-called hold-out data, which looks mathematically as follows: $x^* = (x_1^*, \ldots, x_p^*)^T$. It is evident that the hold-out data contains only data of the input features, yet no information of its classification, for it is simply not available. With the help of the SVCs, the hold-out data will then be assigned to either of the two classes.

In this thesis, the monthly data of the CPI growth and of the change in the 3-month US Treasury bill rate will serve as such input features that classify the up or down movement of the stocks traded at AMEX, NYSE and NASDAQ. Notwithstanding the great advantage of SVMs to work in a p-dimensional space, I limit myself to two input features for four reasons. First, macroeconomic input features only are considered in order to use non-financial data. As a downside, this reduces the range of possible input features. Second, their availability is further restricted as data on at least a monthly basis is required.[15] Third, the limited computational capacity constitutes a practical obstacle. Finally, adding more input features would not help to convey the idea of SVMs any better and so, for the sake of parsimony, two suffice. As a pleasant side effect, using only two input features and thereby restricting the model to a two-dimensional space makes it possible to visualise the results.

But how are those two macroeconomic inputs assigned to a class? Returning to the example with cats and dogs, body size and nose shape are replaced by macroeconomic indicators, whereas the classification changes now from cats and dogs to up and down movements of the stock market. However, the classification in this work is slightly more demanding than the one from the cat/dog example, because we find ourselves dealing with time series data. This gives rise to the following idea. For each month t, a value-weighted portfolio of all the stocks traded at AMEX, NSYE and NASDAQ is constructed. Then one applies a simple difference method to those portfolios where the portfolio price from month t is subtracted from the portfolio's price from month $t+1$. The classification procedure then assigns at time t a one to the input features if the portfolio price from month t to month $t+1$ went up or remained at the exact same level. In the case that the portfolio price dropped from month t to month $t+1$, a minus one is assigned to the input features. The reason why the input features are classified with the stock market movement from month t to $t+1$ is that this allows us to use input features at time t to get a prediction of where the stock market is likely to be at time $t+1$. This classification can be done in the explained

[15] GDP, for instance, is reported on a quarterly basis and thus is too infrequent.

manner because all the data from the past for deriving the parametrisation is available. In other words, one is actually able to assign these up and down movements to the input features, which are older than t.[16] The clue is to apply at time t only the newest input features to the model which is parametrised with older data.

Now having discussed what an input feature and a classification are, we can proceed to the next obstacle – how to find the best hyperplane. Having p input features means that we are in a p-dimensional space, where a hyperplane is a flat affine subspace with a dimension of $p - 1$. As mentioned earlier, I will use only two input features. Accordingly, we are working in a two-dimensional space with a one-dimensional subspace (a line). Now imagine having two input features that are categorised into one out of two classes like in Figure 2.

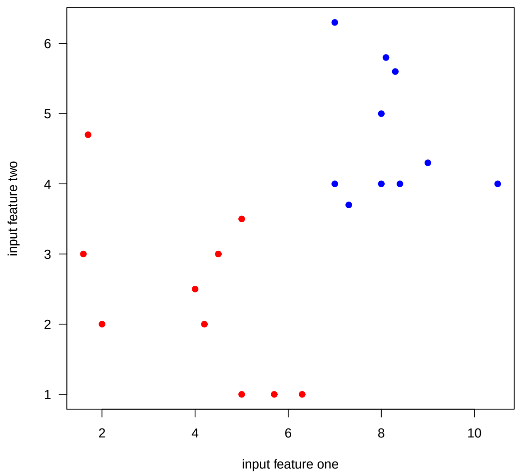


Figure 2: Linearly Separable Data in the Two-Dimensional Space

Obviously, the red points can be linearly separated from the blue ones by drawing a straight line through Figure 2. But what is the best way to separate those two classes? In Figure 3 on the following page, the left panel shows four out of infinite possible straight lines being separating hyperplanes. The right panel depicts the best separating hyperplane, which is also coined as the *maximal margin hyperplane.*

But why is this maximal margin hyperplane considered to be the best choice? Imagine you shift this hyperplane as far to the left and to the right until it touches another point on either side (dashed lines in Figure 4 on the next page), then the shortest distance from the hyperplane

[16]For example, at time $t - 10$, the up or down movement of the mentioned stocks is calculated as follows: portfolio price at $t - 9$ minus the portfolio price at $t - 10$; if this results in a positive difference, one assigns a 1 to the input features at $t - 10$ and otherwise a -1.

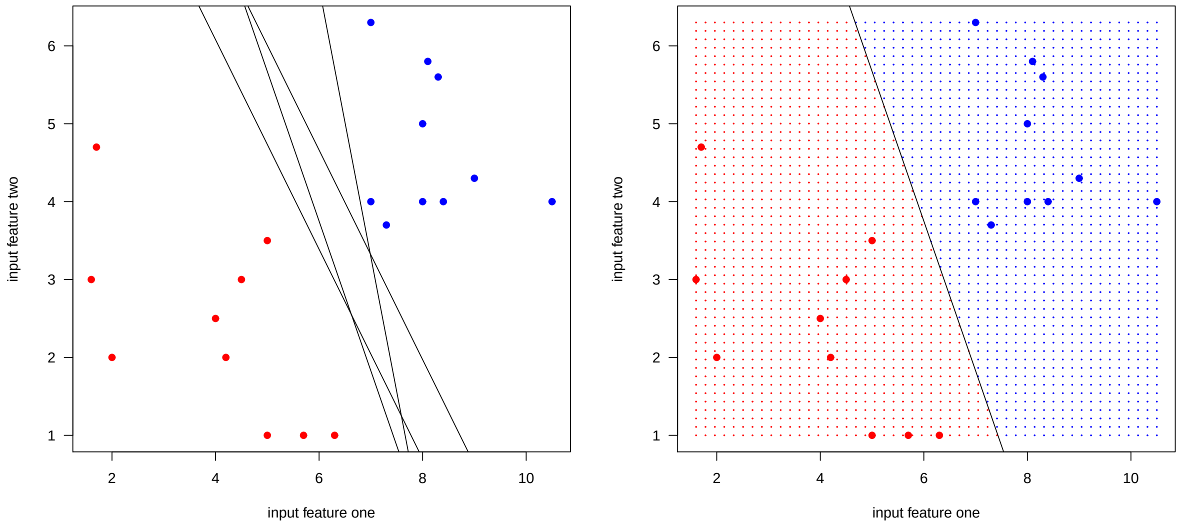


Figure 3: Different Possibilities of Separating Hyperplanes

to one of the dashed lines is the margin. Maximising this margin simultaneously maximises the distance between the hyperplane and the nearest data points to it. Doing so minimises the risk of misclassifying some of the hold-out data. Those dots surrounded by black squares in Figure 4 are called support vectors. The maximal margin hyperplane can be derived solely by the use of these support vectors. They are also called *support vector classifiers* as they determine to which side of the hyperplane hold-out data is assigned.

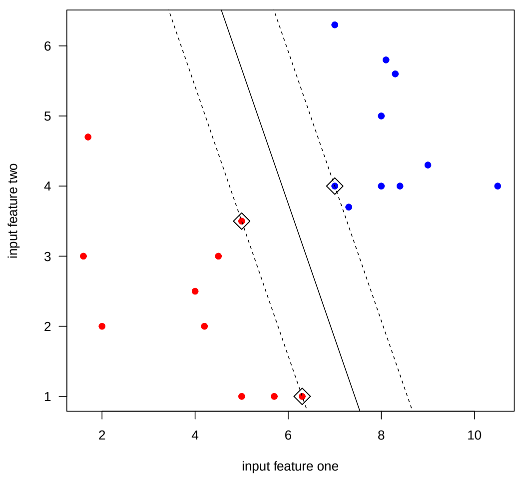


Figure 4: The Maximal Margin Hyperplane

Graphically it is very intuitive what a hyperplane is and why the margin has to be maximised.

But how is it calculated mathematically? In the p-dimensional space, a *normal hyperplane* is defined as follows.[17]

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p = 0 \tag{3}$$

All points $X = (X_1, \ldots, X_p)^T$ that fulfil equation (3) are said to lie exactly on the hyperplane. For reasons already mentioned, only two input features are used and so we are dealing with a hyperplane in the two-dimensional space of the following form

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0, \tag{4}$$

whereas all points $X = (X_1, X_2)^T$ fulfilling equation (4) lie on the hyperplane itself. There are points X, which do not fulfil equation (4) and thus are separated to either side of the hyperplane. All points X, for which

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 > 0 \tag{5}$$

holds, lie on one side of the hyperplane and for all points X, for which

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 < 0 \tag{6}$$

holds, lie on the other side of the hyperplane. Referring to the introductory example with cats and dogs, equation (5) mathematically describes the side where all the dogs are located and equation (6) describes the side where all cats are located.

In order to find the parametrisation of this optimal hyperplane, one uses the training data, which consists of $x_1, ..., x_n \in \mathbb{R}^2$ and the associated classifications $y_1, ..., y_n \in \{-1, 1\}$. Using this notation, one can rewrite equation (5) as

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} > 0 \qquad \text{if } y_i = 1 \tag{7}$$

and equation (6) as

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} < 0 \qquad \text{if } y_i = -1. \tag{8}$$

The maximal margin hyperplane can then be found as the solution to the following optimisation problem.

$$\max_{\beta_0,\beta_1,\beta_2} M \tag{9}$$

$$\text{subject to } \sum_{j=1}^{2} \beta_j^2 = 1 \tag{10}$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) \geq M \qquad \forall i = 1, ..., n \tag{11}$$

[17] The term *normal hyperplane* is used to differentiate it from the maximal margin hyperplane.

This optimisation problem states that we aim to maximise the margin M over the parameters $\beta_0, \beta_1, \beta_2$. The three β_i parameters are weights, which are learned by the algorithm and they fully determine the hyperplane. Furthermore, the sum of squares of all beta parameters have to add up to one as the distance between the separating hyperplane and an observation x is described by $\frac{1}{||\beta||}$, where the length of β is scaled to one, which forces the nearest observations to the separating hyperplane to have an inner product of 1. Maximising $\frac{1}{||\beta||}$ means minimising $||\beta||$, where one could also minimise $||\beta||^2 = \sum_{j=1}^{2} \beta_j^2 = 1$. Equation (11) is a constraint which requires that each observation be on the correct side of the hyperplane. Later on in this chapter, we will see that the maximal margin hyperplane can also be represented by a simple equation which solely makes use of the inner products of the support vectors with each observation. Due to the limited scope of this thesis, I shall not deal with every mathematical derivation.

Now the only remaining question is: how will this model be applied on the macroeconomic data for predicting the trend direction of the stocks? As mentioned before, I divide the data into a formation sample (in-sample) and a hold-out sample (out-of-sample). The goal of the SVM is to develop a classifier based on the in-sample data that can correctly classify the hold-out observation on the basis of its input features. As this thesis deals with the prediction of time series, I will use a rolling window for this estimation process. That means, I will use at any point in time the data from the previous 60 months ($t - 60$ to $t - 1$) in order to get the parametrisation for the maximal margin hyperplane. This estimated model will then be used to classify the input features at time t, which gives us a prediction of the stock market movement at time $t + 1$. This procedure, which is called walk-forward, was already applied by Żbikowski (2015) on financial time series and was described in detail by Ładyżyński, Żbikowski and Grzegorzewski (2013).[18]

The classification process looks mathematically as follows. After having created a hyperplane on the basis of the in-sample data, we use the hold-out observation x^* as input to the equation $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^*$. If $f(x^*)$ is positive, then I assign a one (up movement) to the hold-out observation and otherwise I will assign a minus one (down movement) to it.

Linearly Non-Separable Data

None of this is of any concern provided that the data is perfectly linearly separable by a hyperplane. However, this is most certainly not the case, which is why a situation is conceivable where the data set of interest is not linearly separable as illustrated in the Figure 5 on the next page.

A possible solution to this non-separating problem is to use soft-margins.[19] A hyperplane with

[18]See also Colby (2003, pp. 10–12).

[19]This technique is also often applied when one outlier distorts the maximal margin hyperplane from an

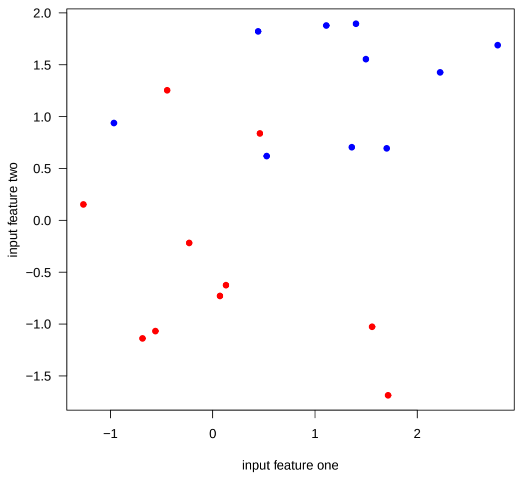


Figure 5: Non-Separable Data

soft-margins still tries to linearly separate the two classes as good as possible but it additionally allows for some observations of the training data to lie on the wrong side of the margin or even on the wrong side of the hyperplane. This has two benefits: first, we get greater robustness to individual observations and second, we achieve a better classification for the majority of the training observations. Mathematically, we only need to slightly adjust the maximisation problem, which was stated for the maximal margin hyperplane (this will be explained afterwards). For the time being, it is sufficient to know that two new parameters are implemented. On the one hand, there is a slack variable $\epsilon_1, \ldots, \epsilon_n$ that allows single observations to be on the wrong side of either the margin or hyperplane or both. On the other hand, there is a parameter C, which is called tuning parameter. This tuning parameter sums up all the ϵ_i and serves as a boundary, meaning that it determines how many violations to the margin and to the hyperplane are tolerated. Figure 6 on the following page shows two such hyperplanes with soft-margins, whereas in the left panel a smaller value for C is used than in the right panel.[20] As the package e1071 in the software R is used to derive all the upcoming results in this thesis, it is worth mentioning that this R package uses the value of C the other way around. Hence, low cost values allow for wide margins while high values of C tighten the margin. In the remainder of the theoretical part I use

hyperplane that would obviously fit the data better under the condition that this one allows the outlier to be misclassified (James et al., 2017, chapter 9).

[20]Because C is the sum of all ϵ_i, the tolerance to violation increases in C and therefore, the margin for a large value of C is wider than for a low value of C. Furthermore, wider margins use more data points as support vectors than small margins do.

C as defined in the literature while I use C in the results part as defined by this package.

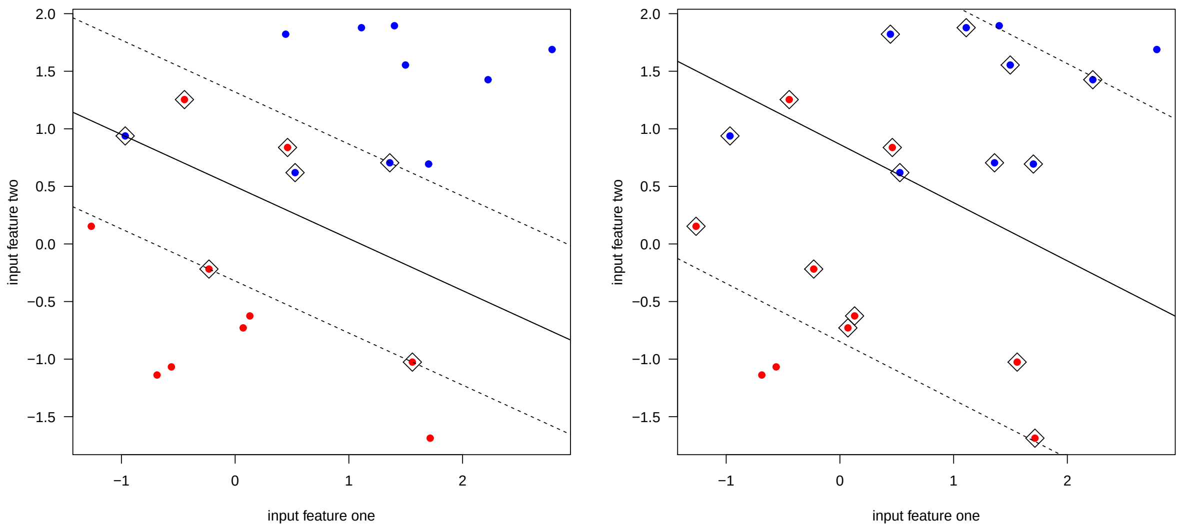


Figure 6: Tuning Parameter C – Tolerance for Violations

Given the practicality of using soft-margins, I will now proceed to the mathematical part, where this tuning parameter C will be implemented into the already known maximisation problem. The maximisation problem under soft-margins looks now as follows.

$$\max_{\beta_0,\beta_1,\beta_2} M \tag{12}$$

$$\text{subject to } \sum_{j=1}^{2} \beta_j^2 = 1 \tag{13}$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) \geq M(1 - \epsilon_i) \tag{14}$$

$$\epsilon_i \geq 0, \qquad \sum_{i=1}^{n} \epsilon_i \leq C \tag{15}$$

Remember that M is the margin's width and we aim at having it as large as possible. C is a non-negative parameter which sums up all ϵ_i, which are so-called slack variables. They allow some individual observations to lie on the wrong side of the margin or even on the wrong side of the hyperplane. The values of the epsilons can be interpreted as follows:

- If $\epsilon_i = 0$, then the i^{th} observation is on its correct position.
- If $\epsilon_i > 0$, then the i^{th} observation is on the wrong side of the margin.
- If $\epsilon_i > 1$, then the i^{th} observation is on the wrong side of the hyperplane.

Which role does the tuning parameter C exactly take? It is a value that we can freely choose in order to bound the sum of ϵ_i. In other words, it tells something about our tolerance of violations of observations which are located on the wrong side. The tuning parameter C is usually chosen via the grid search method, that means a set of different values for the parameters C are tested and then the best performing one is chosen. In addition to the grid search method, I use also the so-called k-fold cross-validation technique. In practice, I use data of 60 months in order to get a parametrisation of the model with a specific cost value C, which is then further used to classify the hold-out data.(Jung, 2018)

I will apply a ten-fold cross-validation on the SVM, wherefore I randomly split those 60 inputs into ten folds. Then the model of interest will be fitted ten times on this data set, where nine folds are used as training data and one fold is used as test data, which differs from the hold-out data in that the latter is used for predictions only. This procedure enables me to fit the model on a total of 540 inputs and test it also on all 60 inputs. Having fitted this model 10 times for the 60 inputs, I thereafter average the values to produce the final model. One could say that this method allows me to use 50 years of data by actually only using 5 years of data as input in order to get the parametrisation of the SVM. The cost of doing so is that one needs to fit the model eleven times instead of only once, but this minimises the out-of-sample error (test data). As I aim to evaluate different cost values, this ten-fold cross-validation will be repeated for every cost value I want to test for and then the grid search method chooses the model that had the smallest out-of-sample error. (Jung, 2018) This model is then further used to classify the hold-out data for which only the input features are available.

An alternative approach to deal with non-separable data, with non-linear decision boundaries, is to enlarge the feature space of the predictors by applying a quadratic, cubic or even higher-order polynomial function to it. This enlargement would solve some of the non-linearity problems but is inefficient when it comes to practical applications due to its computational inefficiency. Instead of enlarging the feature space by applying polynomial functions on the input features, one can make use of so-called *kernels,* which is mathematically extremely efficient because only inner products are needed to solve the previously mentioned maximisation problem from the support vector classifier. The property of the inner product in the feature space having a corresponding kernel in its input space allows it to do calculations, via the use of kernel functions, in the input space rather than in the high dimensional feature space. The use of kernels also helps to solve highly non-linear classification problems. This method of using kernels is called support vector machines and is an extension of the support vector classifier. I will only briefly discuss the topic of inner products and kernels here, as it is, for one, well explained in mathematical textbooks and, for another, the focus of this thesis lies in the practical application of the model.[21] The maximisation problem stated in the equations (12) to (15) involves in principle only the inner products of the observations (input features).[22] The inner product of two observations x_i and x_i' is given by the following equation.

$$\langle x_i, x_i' \rangle = \sum_{j=1}^{p} x_{ij} x_{i'j} \tag{16}$$

For instance, a linear support vector classifier can be described by using inner products as follows.

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x, x_i \rangle \tag{17}$$

From equation (17) one needs to estimate n alpha parameters and β_0. In order to do so, only the inner products of all pairs of training observations are needed. If equation (17) has to be evaluated for some out-of-sample data, one needs to compute the inner product of the new points from the out-of-sample data (x) with each point from the training data (x_i). If a training observation is not a support vector, then its α_i value is zero; in contrast, if α_i is non-zero, the training observation is in the solution for the support vectors.

$$f(x) = \beta_0 + \sum_{i \in S} \alpha_i \langle x, x_i \rangle \tag{18}$$

represents any solution function from equation (17), where S is the collection of all α_i, of which those that are non-zero constitute support vectors. Now, one can replace the inner product in equation (18) with the following generalisation of the inner product

$$K(x_i, x_i'), \tag{19}$$

where K is a function, which we call *kernel*.[23] Equation (18) can then be re-written as

$$f(x) = \beta_0 + \sum_{i \in S} \alpha_i K(x, x_i). \tag{20}$$

James et al. (2017) describe a kernel as "a function that quantifies the similarity of two observations" (p. 352). The mathematical expression of the linear kernel, where the support vector classifier is linear in its features, is given by equation (16). In contrast, a polynomial kernel would look like

$$K(x_i, x_i') = (1 + \sum_{j=1}^{p} x_{ij} x_{i'j})^d, \tag{21}$$

[21] Aronszajn (1950), Girosi (1997) and Heckman (2012) elaborate more on the topic of kernels.

[22] Remember that the inner product of two k-vectors m and n is: $\langle m, n \rangle = \sum_{i=1}^{k} m_i n_i$.

[23] For a thorough discussion on kernels, the reader is referred to Bishop (2006, chapter 6).

where d describes the order of the polynomial. A SVM using a polynomial kernel allows for much more flexible decision boundaries.

In this thesis, I will, on the one hand, investigate whether it is possible to fit a SVM with a linear kernel using different cost functions and macroeconomic input features. On the other hand, I will examine a model with a radial kernel, as this is quite a popular choice in the literature and renders promising results. The radial kernel takes the mathematical form of

$$K(x_i, x_i') = exp\left(-\gamma \sum_{j=1}^{p}(x_{ij}x_{i'j})^2\right), \tag{22}$$

where $\gamma = \frac{1}{2\sigma^2}$ is the free parameter of the Gaussian basis function. More precisely, γ is the inverse of the standard deviation of the Gaussian basis function. The radial basis function kernel is used to make a statement about the similarity of two points, in this case about a support vector and an out-of-sample point. The size of γ determines the variance of the Gaussian function. If γ is chosen to be small, the Gaussian function gets a large variance, which in turn leads to the case that two points are considered to be similar even if they are far apart (in Euclidean distance). But if γ is chosen to be large, the Gaussian function gets a small variance and then two points are considered to be similar only if they are extremely close to each other. In order to find the best value for γ, the grid search method combined with the k-fold cross-validation method can be applied in the same way as for the cost value. (Bishop, 2006, Chapter 6)

The use of such a radial kernel is shown in Figure 7 on the next page on the same non-separable data, which has been used with soft-margins in the linear separation case (compare Figure 6 on page 20).

In conclusion, this section about the support vector machines has shown how perfectly linearly separable data can be classified into two classes by using support vector classifiers. We then proceeded to a case, where the data was no longer linearly separable and we introduced soft-margins, which allow some of the training observations to lie on the wrong side of the margin or even the wrong side of the hyperplane. Finally, we went from the support vector classifiers to the support vector machines, where the main idea is to use the computationally efficient concept of kernels in order to get non-linear decision boundaries. I will investigate in the practical part whether support vector machines with a linear kernel and different cost values can be applied on the macroeconomic input features in order to classify the up and down movement of the value-weighted portfolio of interest. Furthermore, I will also examine SVMs using a radial kernel with different values for the cost and gamma by means of a grid search combined with a ten-fold cross-validation.

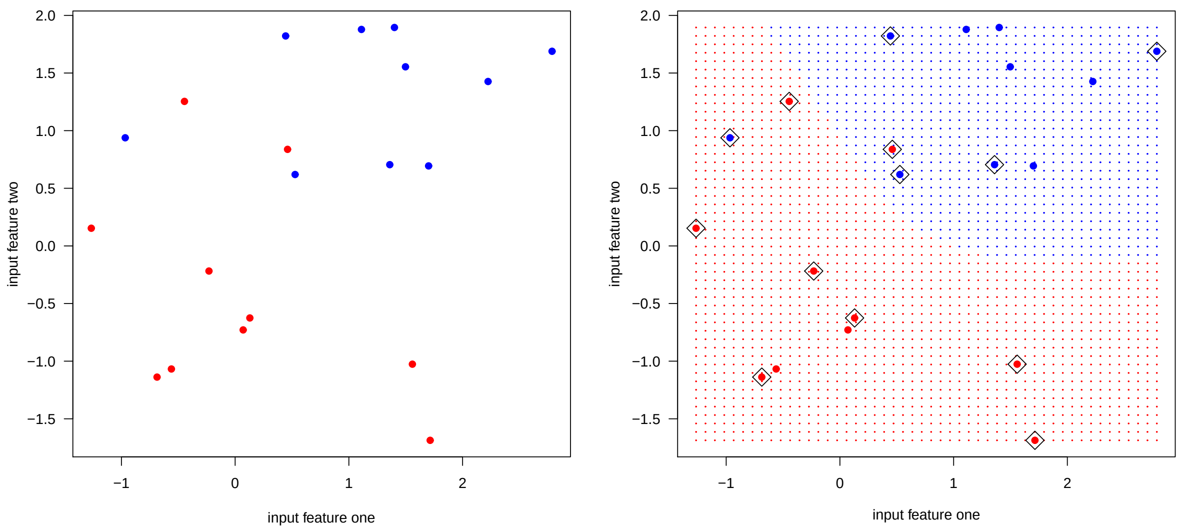


Figure 7: SVM Using a Radial Kernel

Overall, this chapter has shown that momentum profits can be made due to the fact that people over- and underreact to information (Jegadeesh & Titman, 1993). Furthermore, it has been shown that the basic momentum strategy can be improved by implementing some sort of risk-management in order to reduce the loss due to a financial crash (Barroso & Santa-Clara, 2015). Lastly, I introduced the concept of SVM to predict the trend direction of the stocks, which are potentially part of the momentum strategy. These predictions optimise the flexible target volatility.

4 Data

The aim of this chapter is mainly to establish transparency behind the data gathering process but also to provide a reasoning for why some of the data has been chosen. For that purpose, summary statistics and few figures will visualise the data set. The data for this thesis is of two sorts. First, data on macroeconomic fundamentals as well as data on a value-weighted portfolio containing all stocks traded at AMEX, NYSE and NASDAQ is used. This data is needed to fit the prediction model by means of support vector machines. Second, data on the daily and monthly momentum portfolios as well as on the risk-free rate is gathered. This data, in combination with the outcome of the prediction model, is used to construct the risk-managed momentum strategy with a variable target volatility.

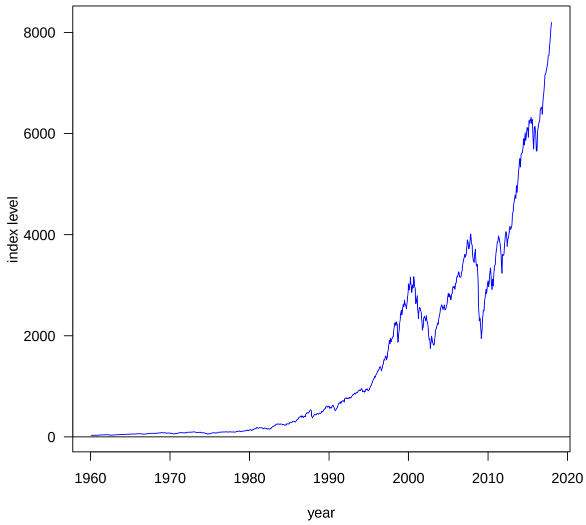
4.1 Support Vector Machines

The objective of SVMs is to predict the up and down movement of all stocks traded at AMEX, NYSE and NASDAQ. For this purpose, monthly data on the value-weighted portfolio of all stocks traded at those stock exchanges is gathered from the Wharton Research Data Services (2018).[24] The data set contains the monthly value of the value-weighted index as well as the corresponding returns for the time period from December 1925 to December 2017. This data is reported on the last trading day of each month. The up and down movement of this index from time t to $t+1$ decides whether the classifier variable gets assigned a 1 or a -1 at time t. For example, we are interested in the classification for February, then one subtracts the reported index value from January from the one reported in February. If this difference is positive, then the stock market is expected to go up and therefore the classifier is assigned a 1 and if it is negative it gets assigned a -1.[25] Figure 8 on the next page illustrates in the left panel how the index has evolved over the time horizon of interest while the right panel shows the corresponding return series. One can see that this return series is indeed very noisy as already indicated by Abu-Mostafa and Atiya (1996) and Deboeck (1994).

In this thesis two macroeconomic variables are used as input features for the SVM, namely the monthly change in the consumer price index as well as the monthly change in the 3-month US Treasury bill return. Given that the three stock exchanges of interest register a broad array of different securities (5,729 in the end of 2017), input features with a wide economic impact are required, for the securities are distributed over a great variety of industries. The CPI is

[24] The Wharton Research Data Services obtained these data from the Center for Research and Security Prices.

[25] In the unlikely event that the difference is zero, I would treat it like an up movement.



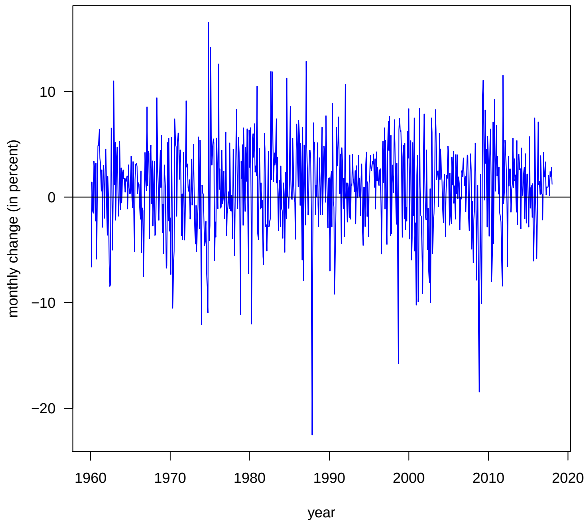


Figure 8: Time and Return Series of the Value-Weighted Index
Source: Author's rendering of CRSP data (2018).

commonly used as a measure of the current inflation level in an economy (Amstad, Potter & Rich, 2017). In addition, inflation plays an important role across all industry sectors as it impacts the interest rate, which in turn affects investment decisions in any business.

The data on the monthly CPI change is gathered from the Federal Reserve Bank of St. Louis (2018) from January 1960 to February 2018.[26] The current level of the CPI is officially reported in the middle of a month. This means that when the stock market prediction for the next month is made (at the end of the month), the input data on the CPI is already two weeks old. Nevertheless, the CPI is still seen as one of the most adequate macroeconomic indicator and therefore used as input feature.

Figure 9 on the following page shows how the consumer price index has changed on a monthly basis for the time period of interest. Especially the crisis of 2009 stands out with a significant drop of 1.915 percent. Table 1 also shows that the monthly change in the CPI was around 0.3 percent, which means that there is a positive trend over the whole time interval.

Min.	Median	Mean	Max.
-1.915	0.291	0.307	1.806

Table 1: Summary Statistics: CPI
Source: Author's rendering of OECD data (2018).

[26] In the time series of the CPI, the year 2010 is indexed as 100.

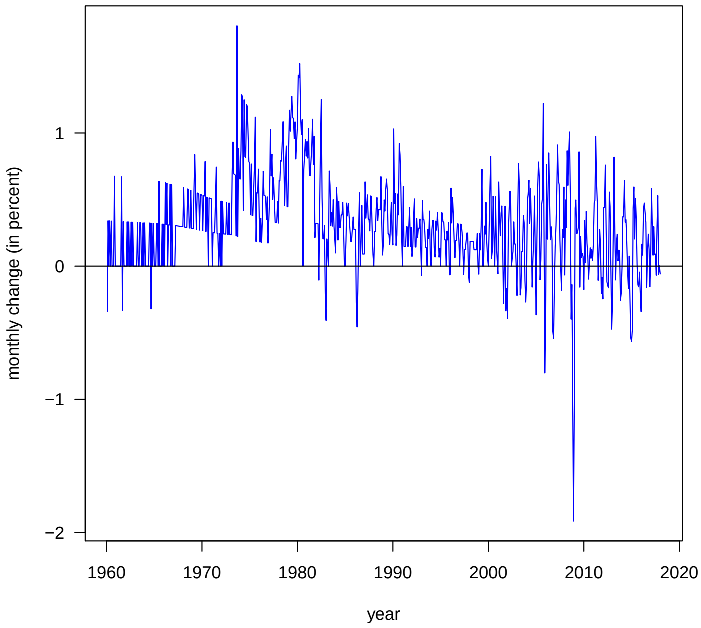


Figure 9: Monthly Change in the Consumer Price Index
Source: Author's rendering of OECD data (2018).

The 3-month US Treasury is gathered from the Federal Reserve Bank of St. Louis (2018). Fama and Schwert (1977) and Fama (1981) showed that the 3-month US Treasury bill is negatively related to future stock market returns and serves very well as a proxy of future economic activity. It has been chosen as an input feature because these securities are held from a multitude of different investors such as domestic financial intermediaries, institutional investors or the Federal Reserve System. Furthermore, the US Treasury securities take on a pivotal role in the world's financial markets. As the payments of the principal as well as the interest of US Treasury securities is fully secured by the credit of the US government, it also reveals something about the health of the state. The great diversity of the different types of investors, the financial influence and its safety mean that these securities are not only widely spread across the economy but also potentially a good macroeconomic input feature. (Dupont & Sack, 1999, pp. 791–792)

Figure 10 on the following page illustrates the return in the 3-month US Treasury bill during the time horizon of interest. Using the monthly percentage changes in these returns as input features yields to both negative and positive changes.[27]

Given the data of those two macroeconomic input features and the value-weighted index

[27] Figure 16 on page V in the appendix illustrates the monthly percentage change in the 3-month US Treasury bill.

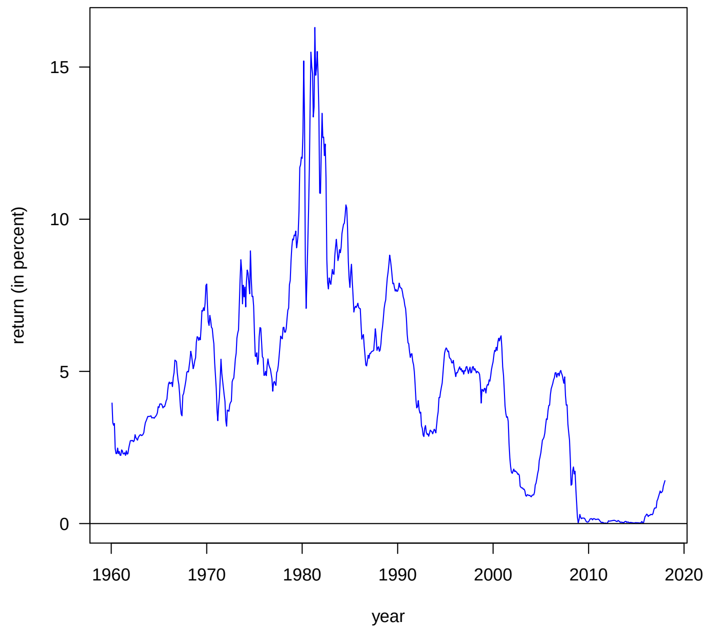


Figure 10: Return on the 3-Month US Treasury Bill
Source: Author's rendering of data from the Board of Governors of the Federal Reserve System (2018).

as well as the constraint that 60 months of data are required to make predictions, the stock market's directional trend was predicted the first time for February 1965. The time interval ends in December 2017.

4.2 Risk-Managed Momentum Strategy

We will now discuss the data used to build the risk-managed momentum strategy with a flexible target volatility. From the data library provided by Kenneth R. French (2018), I gathered data on the daily and monthly ten-momentum sorted portfolios formed on the stocks traded at AMEX, NYSE and NASDAQ. From there, I also gathered data on the monthly risk-free rate.[28,29]

French (2018) constructed the ten-momentum sorted portfolios the following way. For each month, all stocks traded at AMEX, NYSE and NASDAQ with prior return data are sorted in ascending order. The ten deciles are constructed by using the breakpoints of only those stocks that are listed at the NYSE. In other words, each decile has the same amount of NYSE stocks included. For a stock to be part of the portfolio for month t which is constructed at the end of month $t-1$, it must have a price for the end of the month $t-13$ and a good return for $t-2$.

[28]The one-month US Treasury bill return has been chosen as the risk-free rate.

[29]The risk-free rate will be used later in order to show how different portfolios could have evolved over time.

Evidently, there is a time lag between the formation period and the holding period in order to avoid some of the bid-ask spread, which has been mentioned in the previous chapter.

The monthly return on the WML portfolios is illustrated in Figure 11 showing the range in which those returns lie. The first WML portfolio (February 1965) contains 229 securities in the winner and 339 in the loser portfolio, respectively, whereas in the last WML portfolio (December 2017) 463 securities were held in the winner and 492 in the loser portfolio. Table 2 provides further information about the firm size and the monthly returns of the WML portfolios. The firm size of the winners was on average three times as large as the one of the losers. Furthermore, the huge losses of -42.02 percent in 2001 and -45.59 percent in 2009 catch the eye. Especially crashes of such magnitude will be mitigated by using my risk-managed momentum strategy with a flexible target volatility.

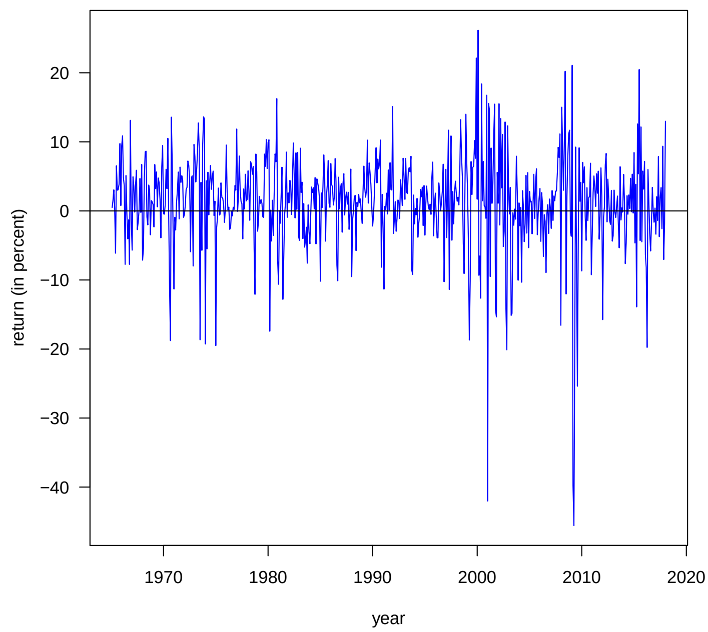


Figure 11: Monthly Return on the WML Portfolios
Source: Author's rendering of data from the Kenneth R. French Data Library (2018).

	Min.	Median	Mean	Max.
firm size losers (in millions)	18.01	151.00	362.740	4,717.67
firm size winners (in millions)	32.02	500.73	1,327.040	9,493.07
monthly returns (in percent)	-45.58	1.72	1.279	26.16

Table 2: Summary Statistics: WML portfolios
Source: Author's rendering of data from the Kenneth R. French Data Library (2018).

5 Results and Discussion

The aim of this chapter is to illustrate and discuss the results for both the stock market predictions from the SVMs and the implementation of a flexible target volatility to the risk-managed momentum strategy. These findings shall then be used to answer the two-fold research question. How accurate can support vector machines fed with macroeconomic input features predict the future stock market movement? How far can those new findings from the SVMs be used to improve the risk-managed momentum strategy by applying those insights to scale the target volatility?

5.1 Support Vector Machine Predictions

In order to predict the up and down movements of the value-weighted portfolio, I investigated on support vector machines with two different types of kernels, namely linear and radial. Due to the fact that the used input data is not perfectly linearly separable, it was not possible to obtain meaningful results by applying the linear kernel.[30] As matter of fact, many cases exist where a linear kernel fails to differentiate between the two classes and thus assigns all test data and therefore also all the hold-out data (out-of-sample data) to the same class. This is the reason why I decided not to further investigate in the linear kernel. In contrast, the SVMs with a radial kernel were capable of differentiating the two classes. Figure 12 on the next page illustrates a SVM with radial kernel applied on the given data set. To remember, the dots that are surrounded by black squares are support vectors and play a crucial role when it comes to the classification of the hold-out data. As mentioned in the chapter on the model, I use the grid search method combined with the ten-fold cross-validation method in order to find the best cost value as well as the best γ parameter. Table 3 on the following page contains information on how often which cost and γ value has been chosen in the best fitting model. In 464 out of 635 times, a γ value of 0.4 has been chosen as best performing value. The reason why I did not allow for values of $\gamma < 0.4$ is that the variance of the Gaussian function would become larger and this would, in turn, lead to the case where two points are considered as similar even if they are far apart. In short, too small a value for γ would lead to the same problem that is present with the linear kernel, where it became impossible to consistently differentiate between both classes.

Using the given data, it is possible to predict the stock market movement for the time horizon starting in February 1965 and ending in December 2017. Table 4 on page 32 displays a confusion

[30] Figure 17 on page VI in the appendix illustrates two examples, where a linear kernel has been applied on the given data.

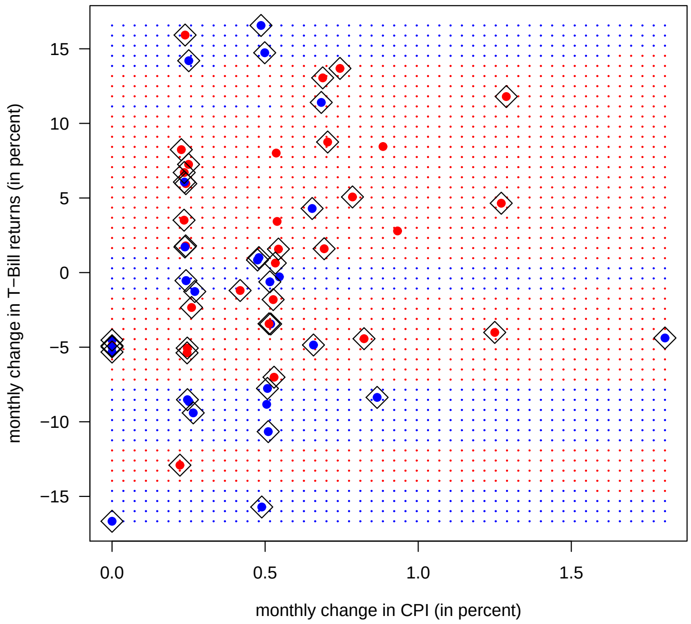


Figure 12: SVM Using a Radial Kernel

Cost Value	Occurrence	Gamma Value	Occurrence
10^{-1}	301	0.40	464
10^{0}	108	0.45	37
10^{1}	70	0.50	48
10^{2}	90	0.55	48
10^{3}	66	0.60	38

Table 3: Count of Used Cost and Gamma Values

matrix as well as the average in- and out-of-sample accuracy of the SVMs using a radial kernel. It can be seen that the out-of-sample prediction is correct in 61.1 percent of the time. Already this result answers the first part of the two-fold research question. Indeed, it is possible to predict the stock market movement to some degree by using the monthly change in the 3-month US Treasury bill returns as well as the monthly change in the CPI as input features. Furthermore, it outperforms that strategy, in which the stock market continues its trend from last month, by 4.9 percentage points in prediction accuracy.

As an interim result, SVMs with a radial kernel are able to predict the stock market movement with an accuracy of 61.1 percent. The confusion matrix also shows that the SVMs with a radial kernel perform much better in predicting the up than the down movement.

prediction		truth -1	truth 1
	-1	52	56
	1	191	336

in-sample accuracy	out-of-sample accuracy
65.677%	61.102%

Table 4: Predicted vs. True Stock Market Movements

5.2 Risk-Managed Momentum Strategy

Having shown the strength of the SVM, we take now the next step and incorporate the obtained results to the risk-managed momentum strategy via the four different scaling methods for the target volatility. Imagine the following investment strategy. An investor initially puts exactly one dollar into the risk-free asset. Simultaneously, he invests a certain percentage (dependent on the strategy) of that risk-free investment in the WML portfolio. Each month, this strategy reinvests the accumulated wealth in the risk-free rate and again spends a certain percentage of this investment on the WML portfolio. Figure 13 on the following page illustrates how such a one-dollar investment in the beginning of 1965 would have developed until the end of December 2017 using different strategies. On the ordinate are the cumulative returns on a logarithmic scale so that the performance in the beginning of the strategy, too, is recognisable.[31] The standard momentum strategy reaches a cumulative return of 6,795.50 US dollar at the end of the year 2017. The model of Barroso and Santa-Clara (2015), which I use as a baseline model, clearly outperforms the standard momentum strategy by achieving a cumulative return of 84,632.29 US dollar. My strategy of the risk-managed momentum strategy with a flexible target volatility is able to outperform the baseline model independently of the scaling method applied on the target volatility. The scaling method where I allow 15 percent annualised volatility when the SVMs have predicted an up movement for the stock market and 9 percent annualised volatility if the SVMs have predicted a down movement, turns out to perform best.[32] It achieves an astonishing return of 184,143.30 US dollar by the end of 2017, which is roughly more than the double of the baseline model. Figure 18 on page VII in the appendix shows the weights in the WML portfolios.

[31] Figure 19 on page VIII in the appendix displays the growth of this one dollar investment on a non-log scale.

[32] The model that uses the scaling method four, which reacts faster to down movements, also shows strong results. Unfortunately, due to too many wrong down predictions, the model is not able to skim all the gains when the market is flourishing.

The average weight from my strategy in the WML portfolio is 98.56 percent, whereas it ranges from 13.27 percent to 229.23 percent over the entire time horizon.

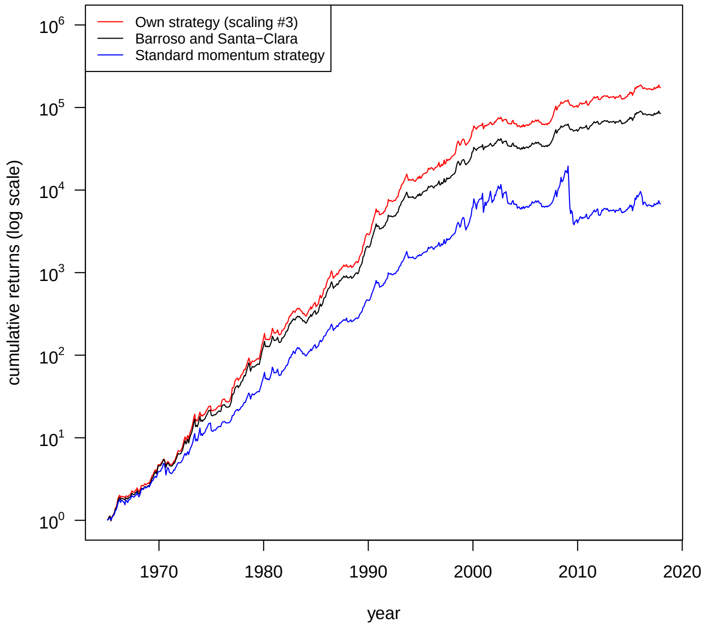


Figure 13: Development of Momentum Strategies from 1965:02 to 2017:12 (Log Scale)

This higher return does not *per se* mean that I am able to outperform the baseline model in all respects. Good measures of comparison between the performance of my model and the one proposed by Barroso and Santa-Clara (2015) are the higher order moments such as skewness and excess kurtosis. An also often-used indicator is the so-called Sharpe ratio [SR], which sets the excess return of a portfolio and its risk in relation to each other. The higher the SR, the more excess return is being generated at the same level of risk. (Sharpe, 1994) The SR is calculated as follows.[33]

$$SR = \frac{\mathbb{E}(r_{portfolio} - r_f)}{\sigma_{portfolio}} \tag{23}$$

Table 5 on the following page displays the SR, the skewness, the excess kurtosis and the cumulative return for all strategies. Getting straight to the point, I will only compare my best performing strategy with the model from Barroso and Santa-Clara (2015). In addition, I will juxtapose its performance with the standard momentum strategy. Barroso and Santa-Clara (2015) were able to decrease the skewness from -1.39 to -0.15 and reduce the excess kurtosis from 7.47

[33]The explanation of Figure 20 on page IX in the appendix elaborates more on how the realised variance is obtained in order to calculate the SR.

to 1.34 when compared to the standard momentum strategy. These substantial improvements in the higher order moments almost eliminate the crash risk of momentum. My model with the scaling method three on the target volatility is still able to decrease the left skewness further from -0.15 to -0.01. The excess kurtosis declines from 1.34 to 1.18. As mentioned above, the Sharpe ratio serves as a good measure to compare two portfolios' returns as it sets the excess return in relation its the risk. In other words, the SR brings the performance of two portfolios to the same scale making them comparable. By using my strategy instead of the one proposed by Barroso and Santa-Clara (2015), the SR increases from 0.75 to 0.83. Hence, my strategy generates more excess return at the same level of risk. All in all, one can conclude that my strategy outperforms the baseline model by not only lowering the crash risk of momentum but also by generating higher excess returns at the same risk. My strategy with scaling method

Strategy	SR	SKEW	KURT	Cumulative Return
Scaling method 1	0.780	-0.100	1.244	111,332.63
Scaling method 2	0.806	-0.053	1.197	144,213.36
Scaling method 3	0.832	-0.008	1.189	184,143.30
Scaling method 4	0.827	-0.021	1.226	173,623.88
Barroso and Santa-Clara	0.752	-0.151	1.339	84,632.29
Barroso and Santa-Clara (13.98 target)	0.752	-0.151	1.339	159,327.58
Standard momentum	0.601	-1.391	7.472	6,785.80

Table 5: Comparison of the Different Strategies

three has on average an annualised target volatility of 13.98 percent. If the strategy proposed by Barroso and Santa-Clara (2015) were applied on the same target (13.98 percent), it would generate a cumulative return of 159,327.58 US dollar without altering the distribution of the return series. The fact that I am able to increase the SR demonstrates the superiority of my model.

Before closing this chapter, the efficient market hypothesis [EMH] provides a possible explanation for why my strategy had superior returns compared to Barroso and Santa-Clara (2015). Imagine the setting of the weak form of the EMH as elaborated by Bodie, Kane and Marcus (2014, pp. 353–375), where investors cannot outperform the market by analysing the history of prices and returns. Superior returns might still be achieved through a fundamental analysis such as macroeconomic models, which make use of public information.[34] The obtained information

[34]For more information on the efficient market hypothesis, see also Fama (1970, pp. 383–417) and Burton

can then be used to improve investment decisions. As a matter of fact, my strategy trumps their strategy in terms of excess returns because the former takes additional information, which have turned out to be relevant, into consideration.

Put figuratively, applying a risk-managed momentum strategy with a flexible target volatility resembles a poker player who, knowing when to raise or fold, increases or decreases accordingly his bet and thus gains more and loses less. He plays as if he could see one card of every opponent's hand. Although he might lose in some instances, he still wins overall. The same applies on the stock market, where we use the information on the stock market's directional trend as advantage over others. This information edge allows us to accept more of the good risk during flourishing market times and less in recessions. There, too, will be instances of losses due to erroneous predictions (type I error) but this strategy is, generally speaking, lucrative. Certainly, the incorporation of this information leads to superior returns as long as it is rarely used among other investors. If all the poker players saw one card of their opponent's hand, then this card would become commonly shared information and one would lose his edge. In the stock market, superior returns are only delivered as long as the information advantage remains unique.

Now the second part of the two-fold research question can be answered. Allowing for a flexible target level in the risk-managed momentum strategy, whereas this target level is determined on the basis of the stock market predictions made by the SVMs, generates higher returns at the same risk level.

In conclusion, it has been shown that SVMs with a radial kernel can predict the stock market movement with an accuracy of 61.1 percent, when using the monthly change in the 3-month US Treasury bill returns and the monthly change in the CPI as input features. However, SVMs fail to accurately predict on a consistent basis the down movements. Nevertheless, the gained information on the stock market movement improves the risk-managed momentum strategy by allowing a flexible target level. The added value comes in the form of both a lower crash risk of momentum and a higher Sharpe ratio.

(2003, pp. 59–82).

6 Robustness

It has been shown in the previous chapter that my model outperforms the risk-managed momentum strategy invented by Barroso and Santa-Clara (2015) when analysed over the time horizon from February 1965 to December 2017. Now the question arises whether my strategy shows such a good performance simply because it survives the crash in 2009 better than other strategies; or may it also outperform the model proposed by Barroso and Santa-Clara (2015) and the standard momentum strategy during a period without a financial crisis? The aim of this chapter is to compute all the models for two different time horizons: one without crises and one where a financial crisis occurs. If my strategy is still able to dominate the others during both time intervals, we can be sure that the obtained results in the previous chapter were neither pure luck nor due to a single event. The first time horizon is chosen from February 1965 to December 1999. The second time begins in January 2000 and ends in December 2009, where a crash occurred in the final year.

6.1 Non-Crisis

The confusion matrix reported in Table 6 shows that the out-of-sample accuracy of the stock market predictions using the SVM with a radial kernel is slightly higher than it was over the whole time horizon. Figure 14 on the following page depicts the growth of an initial one-dollar investment when following my strategy with the scaling method three for the variable target volatility.[35] In addition, the evolution of that investment following the strategy of Barroso and Santa-Clara (2015) as well as the standard momentum strategy is illustrated. During this non-crisis period, my strategy achieved a cumulative return of 53,279.29 US dollar, which is roughly twice the amount of money that would have been accumulated by the strategy of Barroso and Santa-Clara (2015) and nine times when compared to the standard momentum strategy.

		truth		in-sample accuracy	out-of-sample accuracy
		-1	**1**		
prediction	**-1**	41	40	66.881%	61.337%
	1	122	216		

Table 6: Predicted vs. True Stock Market Movements (Non-Crisis 1965:02 to 1999:12)

Table 7 on the next page reports that my strategy with the flexible target improves the

[35] The investment strategy is the same as in the previous chapter.

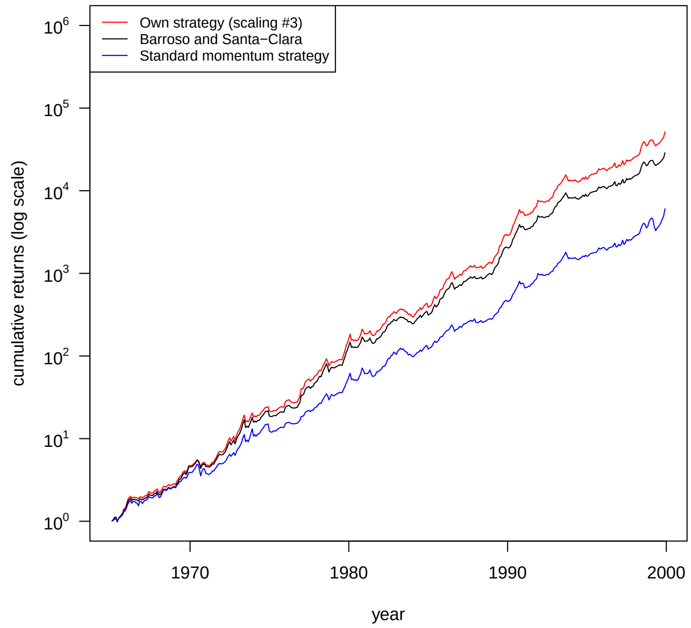


Figure 14: Performance During Non-Crisis Times

higher order moments. It shortens, so to say, the left skew and lowered the excess kurtosis, which in turn means that the crash risk of momentum is reduced. The increase in the Sharpe ratio means that my strategy is able to generate more excess returns for a certain amount of risk when compared to the other strategies. As an interim conclusion, I hold that my strategy with the third scaling method (allowing for an additional 3 percent annualised volatility on top of the 12 percent target level in up predictions and 3 less in down predictions) is able to outperform the baseline model during non-crisis times as well.

Strategy	SR	SKEW	KURT	Cumulative Return
Scaling method 1	1.085	-0.236	1.034	35,808.46
Scaling method 2	1.118	-0.179	0.966	43,970.61
Scaling method 3	1.149	-0.125	0.937	53,279.29
Scaling method 4	1.146	-0.142	0.979	51,691.74
Barroso and Santa-Clara	1.050	-0.296	1.150	28,753.09
Standard momentum	0.797	-0.674	2.096	6,051.05

Table 7: Performance During a Non-Crisis Period (1965:02 to 1999:12)

6.2 Crisis

It remains to investigate how the model performs during a time of crisis. For that purpose, I analyse the strategies over a time horizon starting in January 2000 and ending in December 2009, which includes the crash of that year. A sharp fall in the out-of-sample accuracy can be observed in Table 8. This result emphasises the finding from the previous chapter and underscores that the SVMs exhibit a deficient accuracy when predicting stock market downturns. One reason is that the prediction is based on too low a frequency. In other words, the monthly data does not completely capture such fast market movements. Nevertheless, the obtained out-of-sample accuracy is still 53.33 percent.

prediction	truth -1	truth 1
-1	11	14
1	42	53

in-sample accuracy	out-of-sample accuracy
60.194%	53.333%

Table 8: Predicted vs. True Stock Market Movements (Crisis 2000:01 to 2009:12)

Figure 15 on the next page displays how a one-dollar investment in the beginning would have evolved following different strategies. It is evident that the standard momentum strategy fluctuates much more than the other two strategies. In particular, the sharp drop in 2009 draws special attention. Over this ten year time horizon, an investor following the standard momentum strategy would have recorded a loss of 27.5 percent, which would have mainly been caused by the crash in 2009. The strategy proposed by Barroso and Santa-Clara (2015) is still able to gain 87.7 percent over its initial investment. This is mainly due to the fact that their strategy survives the crash by being exposed to the momentum risk by only 15.9 percent at this time.[36] My own strategy with the flexible target volatility and scaling method three is able to prevent most of the losses occurred by the momentum crash by merely having a minor exposure of 13.7 percent at that time. Over the decade, my strategy would have generated a total return of 108.9 percent and therefore would have registered the best return among the tested strategies.

Here, we take also a look at the higher order moments and the Sharpe ratio (see Table 9 on page 40). The Sharpe ratio of my model is only slightly higher than that of the others. The reason might be that I report the average Sharpe ratio over a time horizon of ten years. Hence, the higher SR of the standard momentum strategy in the beginning signifies that, in average, it

[36]In March 2009, the WML portfolio had a negative return of 39.76 percent and in April it was even worse when it gave a negative return of 45.58 percent.

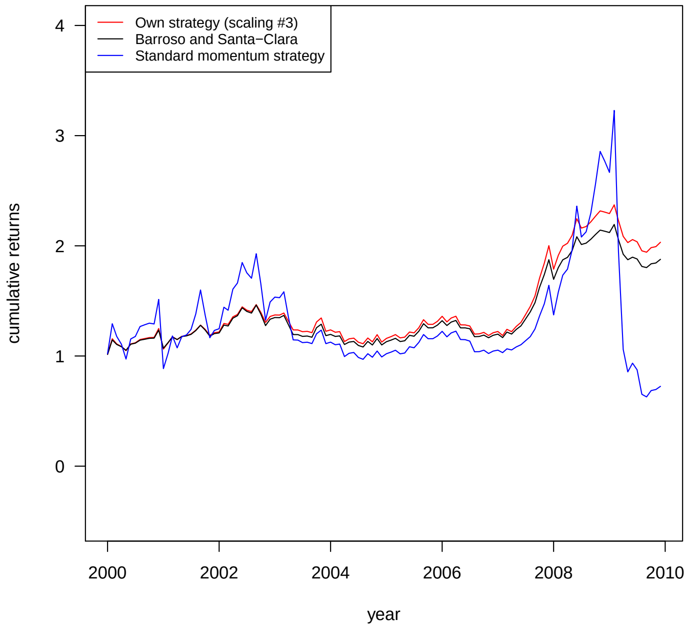


Figure 15: Performance During Crisis Times

outperforms the model of Barroso and Santa-Clara (2015). An investigation of the mean SR for the time window including the ten months before and after the shock in the year 2009 reveals that the SR would change as follows. The standard momentum strategy would result in an average SR of -0.305, Barroso and Santa-Clara's strategy would register an average SR of 0.065 and my strategy would reach an average SR of 0.081. A closer look around the crash elucidates the performance of all three strategies: the standard momentum performs extremely poorly while mine is with a small margin the best performing strategy. The skewness of my strategy is slightly better than the one of Barroso and Santa-Clara; in contrast, I report a higher excess kurtosis than they do. As the excess kurtosis makes a statement about the tail distribution, which in turn includes both large positive and large negative returns, a slightly higher excess kurtosis is not *per se* detrimental. Finally, one may reach two conclusions. First, the standard momentum strategy's performance over this time period is inferior to the remaining two strategies that managed to avoid the crash in 2009 to some extent. Second, all Sharpe ratios are close to each other but my model is able to perform best when taking a even closer look around the crash. Third, Barroso and Santa-Clara (2015) and I achieve major improvements compared to the standard momentum strategy when considering the higher order moments. All in all, my model outperforms the others in several ways despite the margin occasionally being very small.

Strategy	SR	SKEW	KURT	Cumulative Return
Scaling method 1	0.128	-0.330	0.948	1.946
Scaling method 2	0.144	-0.316	1.078	2.017
Scaling method 3	0.160	-0.306	1.214	2.089
Scaling method 4	0.146	-0.344	1.198	2.032
Barroso and Santa-Clara	0.112	-0.348	0.826	1.877
Standard momentum	0.140	-1.298	27.567	0.725

Table 9: Performance During a Crisis Period (2000:01 to 2009:12)

In conclusion, this chapter has demonstrated that my model with a flexible target level shows better results in times of crisis and non-crisis as well as when narrowing down the time horizon to time intervals in both scenarios. Because my model exhibits a higher performance in either one, it trumps the other two strategies overall.

7 Conclusion

The aim of this thesis was two-fold. On the one hand, I have demonstrated that macroeconomic data may very well serve as input features to SVMs when it comes to one-month ahead forecasts of a stock market's directional trend. On the other hand, I have shown that the risk-managed momentum strategy proposed by Barroso and Santa-Clara (2015) can be improved by allowing more volatility during times where the stock market is flourishing and less volatility if otherwise.

My contribution to academia is the conflation of two otherwise disparately used concepts – momentum strategy and SVMs – for the purpose of constructing a generally enhanced model. In order to get the one-month ahead forecasts of a value-weighted portfolio constructed out of all the stocks traded at AMEX, NYSE and NASDAQ, the technique of SVMs was applied. The monthly percentage change in both the CPI and the 3-month US Treasury bill rate served as input features to the SVMs. Making predictions over a time horizon beginning in February 1965 and ending in December 2017 resulted in a 61.1 percent prediction accuracy of the stock market's directional trend. These findings were then incorporated into the target level of the risk-managed momentum portfolio of Barroso and Santa-Clara (2015), which scales the exposure to the momentum risk by means of its monthly volatility forecast. A prediction of an up movement corresponded to a higher volatility target while a prediction of a down movement corresponded to a lower volatility target. The momentum portfolios are constructed on a 12 month formation period and are held for one month, whereas a time lag between the formation an holding period is needed to avoid some of the bid-ask spread.

The implementation of a flexible volatility target increased the Sharpe ratio from 0.752 to 0.832 when compared to the baseline model from Barroso and Santa-Clara (2015) which used a annualised target volatility of 12 percent. Furthermore, the skewness has improves from -0.151 to -0.008 and the excess kurtosis went from 1.339 to 1.190. This means a reduction of the overall crash risk of momentum. Furthermore, the risk-managed momentum strategy with a flexible target level outperformed the other models in both non-crisis and crisis times as well.

It is necessary to briefly discuss one caveat: transaction costs. They were ignored because the main objective of this thesis was the incorporation of predictions made by SVMs into the risk-managed momentum strategy. This allowed me to compare my results with those obtained by Barroso and Santa-Clara (2015). The findings of Ammann et al. (2010) suggest that including trading costs in the momentum strategy still renders significant high returns. Hence, my results are expected to alter at best negligibly when taking transaction costs into account.

In conclusion, SVMs are able to predict a stock market's monthly directional trend to a certain degree when using macroeconomic variables as input features. Furthermore, the use of those predictions allows to reduce the risk of an investment strategy, as shown on the example of the momentum strategy. This begs the question: in which direction may the strategy be further enhanced? I see the greatest potential in the improvement of the prediction accuracy of the down movements. For this purpose, one could investigate different input features for the SVMs, be they either technical (e.g., a simple moving average) or macroeconomic (e.g., the unemployment rate or the dividend price ratio). An alternative approach is to apply my model while predicting the stock market's directional trend on an hourly basis, whereas those predictions would serve as an early warning system: after a certain number of consecutive down predictions on an hourly basis, the investment strategy should sell off its investment in the WML portfolio instead instead of idly observing an entire month of down movements.

Coming full circle, predictions – especially those about the future – remain a daunting endeavour; nevertheless, this uncertainty may partially be reduced when applying the risk-managed momentum strategy with a flexible target volatility.

References

Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. *Applied Intelligence*, *6*(3), 205–213. doi: 10.1007/BF00126626

Ammann, M., Moellenbeck, M. & Schmid, M. (2010). Feasible momentum strategies in the US stock market. *Journal of Asset Management*, *11*(6), 362–374. doi: 10.1057/jam.2010.22

Amstad, M., Potter, S. & Rich, R. (2017). The New York Fed staff underlying inflation gauge (UIG). *Economic Policy Review*, *23*(2), 1–32. Retrieved from `https://ssrn.com/abstract=3090327`

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, *68*(3), 337–404. doi: 10.2307/1990404

Baltas, A.-N., & Kosowski, R. (2013). Momentum strategies in futures markets and trend-following funds. *Paris December 2012 Finance Meeting EUROFIDAI-AFFI Paper*. doi: 10.2139/ssrn.1968996

Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, *116*(1), 111–120. doi: 10.1016/j.jfineco.2014.11.010

Bishop, C. M. (2006). Pattern recognition and machine learning. In *Information science and statistics.* Secaucus, NJ: Springer-Verlag New York, Inc.

Board of Governors of the Federal Reserve System (US). (2018). *3-month treasury bill: Secondary market rate [tb3ms].* Retrieved from `https://fred.stlouisfed.org/series/TB3MS`

Bodie, Z., Kane, A. & Marcus, A. J. (2014). *Investments.* New York, NY: Maidenhead McGraw-Hill Education.

Burton, G. M. (2003). The efficient market hypothesis and its critics. *The Journal of Economic Perspectives*, *17*(1), 59–82. doi: 10.1257/089533003321164958

Cao, L., & Tay, F. E. (2001). Financial forecasting using support vector machines. *Neural Computing & Applications*, *10*(2), 184–192. doi: 10.1016/S0925-2312(03)00372-2

Chan, L. K. C., Jegadeesh, N. & Lakonishok, J. (1996). Momentum strategies. *The Journal of Finance*, *51*(5), 1681–1713. doi: 10.1111/j.1540-6261.1996.tb05222.x

Chordia, T., & Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. *The Journal of Finance*, *57*(2), 985–1019. doi: 10.1111/1540-6261.00449

Colby, R. W. (2003). *The encyclopedia of technical market indicators* (2nd ed.). New York, NY: McGraw-Hill Education.

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, *122*(2), 221–247. doi: 10.2139/ssrn.2371227

Deboeck, G. J. (1994). *Trading on the edge: Neural, genetic, and fuzzy systems for chaotic financial markets.* New York, NY: John Wiley & Sons, Inc.

De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, *40*(3), 793–805. doi: 10.2307/2327804

Dupont, D., & Sack, B. P. (1999). The treasury securities market: overview and recent development. *Federal Reserve Bulletin*, 785–806. Retrieved from `https://ideas.repec.org/a/fip/fedgrb/y1999idecp785-806nv.85no.12.html`

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. Retrieved from `http://www.jstor.org/stable/2325486`

Fama, E. F. (1981). Stock returns, real activity, inflation, and money. *The American Economic Review*, *71*(4), 545–565. Retrieved from `http://www.jstor.org/stable/1806180`

Fama, E. F., & Schwert, G. (1977). Asset returns and inflation. *Journal of Financial Economics*, *5*(2), 115–146. doi: 10.1016/0304-405X(77)90014-9

Frazzini, A. (2006). The disposition effect and underreaction to news. *Journal of Finance*, *61*(4), 2017–2046. doi: 10.1111/j.1540-6261.2006.00896.x

French, K. R. (2018). *Data library.* Retrieved from `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Girosi, F. (1997). An equivalence between sparse approximation and support vector machines. *Neural Computation*, *10*(6), 1455–1480. doi: 10.1162/089976698300017269

Heckman, N. (2012). The theory and application of penalized methods or reproducing kernel Hilbert spaces made easy. *Statistics Surveys*, *6*(1), 113–141. doi: 10.1214/12-SS101

Herberger, T., Kohlert, D. & Oehler, A. (2011). Momentum and industry-dependence: An analysis of the Swiss stock market. *Journal of Asset Management*, *11*(6), 391–400. doi: 10.1057/jam.2010.23

Hou, K., Karolyi, G. A. & Kho, B.-C. (2011). What factors drive global stock returns? *Review of Financial Studies*, *24*(8), 2527–2574. doi: 10.1093/rfs/hhr013

Huang, W., Nakamori, Y. & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers and Operations Research*, *32*(10), 2513–2522. doi: 10.1016/j.cor.2004.03.016

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2017). *An introduction to statistical learning: With applications in R.* New York, NY: Springer.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, *48*(1), 65–91. doi: 10.2307/2328882

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance*, *56*(2), 699–720. Retrieved from http://www.jstor.org/stable/222579

Jung, Y. (2018). Multiple predicting k-fold cross-validation for model selection. *Journal of Nonparametric Statistics*, *30*(1), 197–215. doi: 10.1080/10485252.2017.1404598

Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 414–421). Cambridge University Press. doi: 10.1017/CBO9780511809477.031

Kim, K.-J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1), 307–319. doi: 10.1016/S0925-2312(03)00372-2

Ładyżyński, P., Żbikowski, K. & Grzegorzewski, P. (2013). Stock trading with random forests, trend detection tests and force index volume indicators. In L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh & J. M. Zurada (Eds.), *Artificial intelligence and soft computing* (pp. 441–452). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-38610-7_41

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, *105*(1), 1–28. doi: 10.2307/2937816

Moskowitz, T. J., Ooi, Y. H. & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, *104*(2), 228–250. doi: 10.1016/j.jfineco.2011.11.003

Organization for Economic Co-operation and Development. (2018). *Consumer price index: Total all items for the united states [cpaltt01usm657n].* Retrieved from https://fred.stlouisfed.org/series/CPALTT01USM657N

Rouwenhorst, K. G. (1998). International momentum strategies. *The Journal of Finance*, *53*(1), 267–284. doi: 10.2139/ssrn.4407

Sagi, J. S., & Seasholes, M. S. (2007). Firm-specific attributes and the cross-section of momentum. *Journal of Financial Economics*, *84*(2), 389–434. doi: 10.1016/j.jfineco.2006.02.002

Schiereck, D., De Bondt, W. & Weber, M. (1999). Contrarian and momentum strategies in Germany. *Financial Analysts Journal*, *55*(6), 104–116. Retrieved from http://www.jstor.org/stable/4480212

Sharpe, W. F. (1994). The sharpe ratio. *The Journal of Portfolio Management*, *21*(1), 49–58. doi: 10.3905/jpm.1994.409501

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *The American Economic Review*, *71*(3), 421–436. doi: 10.3386/w0456

Theofilatos, K., Georgopoulos, E., Likothanassis, S. & Mavroudi, S. (2014). Computational intelligence: recent advances, perspectives and open problems. In *Computational intelligence techniques for trading and investment* (1st ed., pp. 3–21). London, UK: Routledge.

van Dijk, R., & Huibers, F. (2002). European price momentum and analyst behavior. *Financial Analysts Journal*, *58*(2), 96–105. Retrieved from `https://ssrn.com/abstract=305407`

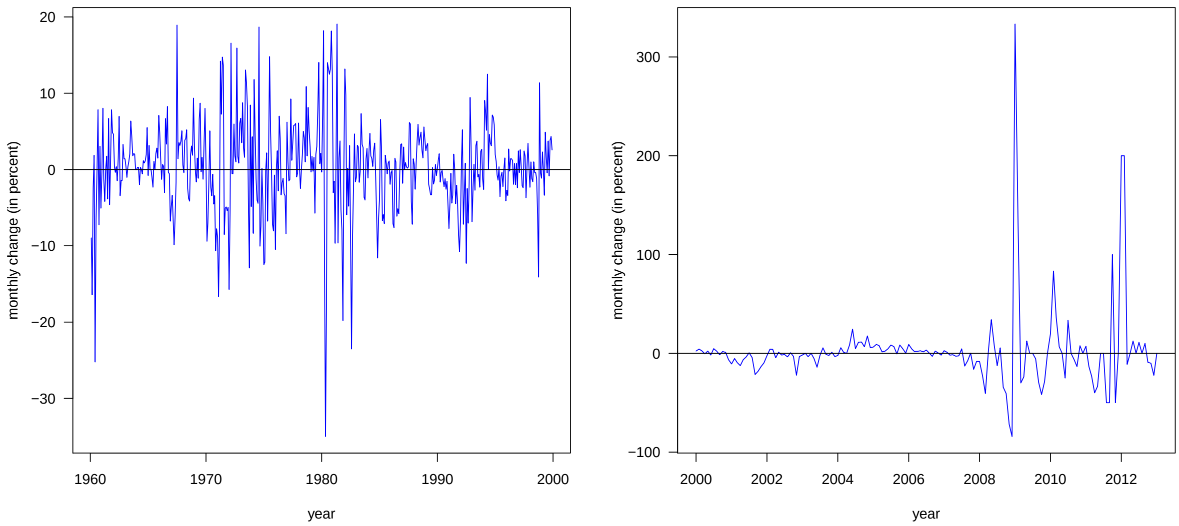
Vapnik, V. N. (1995). The nature of statistical learning theory. *IEEE Transactions on Neural Networks*, *8*(6), 1564–1564. doi: 10.1109/TNN.1997.641482

Wharton Research Data Services. (2018). *CRSP stock file indexes – monthly index built on market capitalization.* Retrieved from `https://wrds-web.wharton.upenn.edu/wrds/ds/crsp/indexes_a/mktindex/cap_m.cfm?navId=124`

Żbikowski, K. (2015). Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. *Expert Systems with Applications*, *42*(4), 1797–1805. doi: 10.1016/j.eswa.2014.10.001

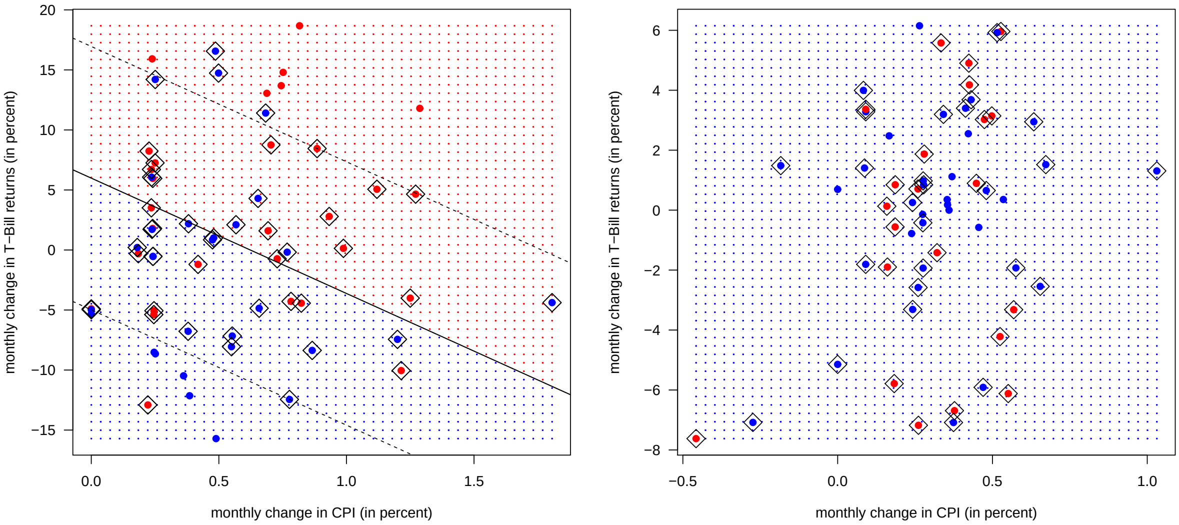
Appendix

Figure 16: Monthly Percentage Change in the 3-Month US Treasury Bill Returns
Source: Author's rendering of data from the Board of Governors of the Federal Reserve System (2018).



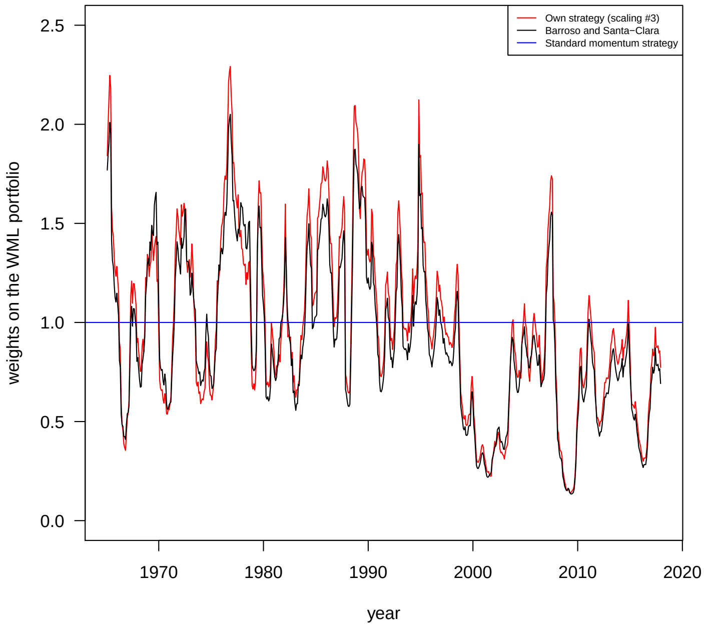
This figure displays the monthly percentage change in the 3-month US Treasury bill returns. When looking at Figure 10 on page 28, one notes that this security has huge swings during the time horizon of interest. It reached a peak of 16.3 percent return in the 1980s and a low of 0.01 percent in 2010. By solely observing the monthly percentage change (Figure 16), one might get the impression that after 2008 the swings became worse. But Figure 10 on page 28 illustrates that the security returns reached its lower bound and subsequently flattened out. The huge spikes in the monthly percentage change are therefore due to the fact that the proportional changes from month to month increased despite lower levels of return.

Figure 17: SVM Using a Linear Kernel



This figure illustrates some of the difficulties which arise – using the monthly change in both the CPI and the 3-month US Treasury bill returns as input features – when it comes to the classification of the stock market's up and down movement. Both panels display a SVM with a linear kernel. In the left panel it was feasible to find a maximal margin hyperplane that divides the up movements (blue dots) from the down movements (red dots). The support vectors are indicated with squares around the data point. In the right panel the SVM with a linear kernel would assign all points in the given data range to the same class. Such classification is useless because it predicts an up movement of the stock market regardless of which values the input features take on. This is why I discarded the possibility of getting accurate estimations for the stock market movement when using a SVM with linear kernel.

Figure 18: Weights in the WML Portfolio for 1965:02 to 2017:12



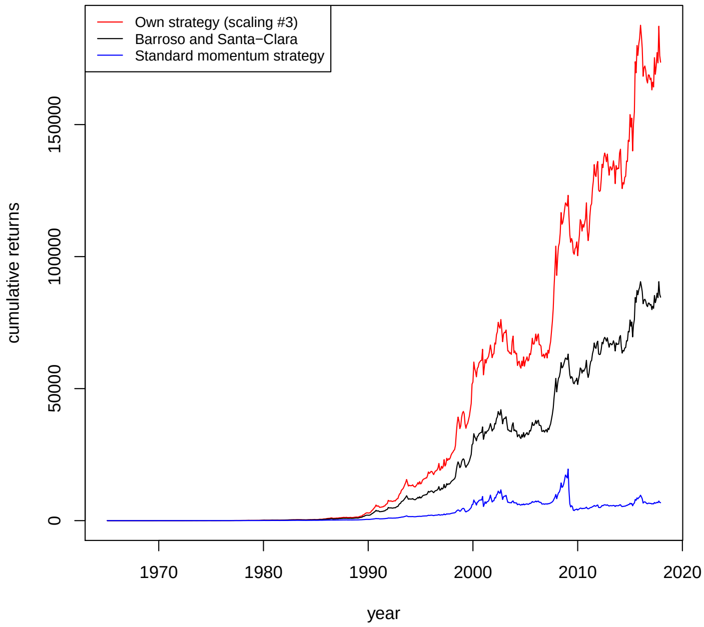
This graph shows the weights for three different strategies in the WML portfolio. The standard momentum strategy always goes long in the winners decile and short in the losers decile for the same amount of money that already had been invested in the risk-free rate. Therefore, the weight of the WML portfolio is always 100 percent for the standard momentum strategy. Both the model proposed by Barroso and Santa-Clara (2015) and my own adjust the exposure to the WML portfolio by scaling its weights (cf. chapter 3 on the model). Table 10 provides

Strategy	Min.	Median	Mean	Max.
Scaling method 3	0.133	0.933	0.986	2.293
Barroso and Santa-Clara	0.134	0.849	0.916	2.050
Standard momentum	1	1	1	1

Table 10: Summary Statistics: Weights of the Different Strategies

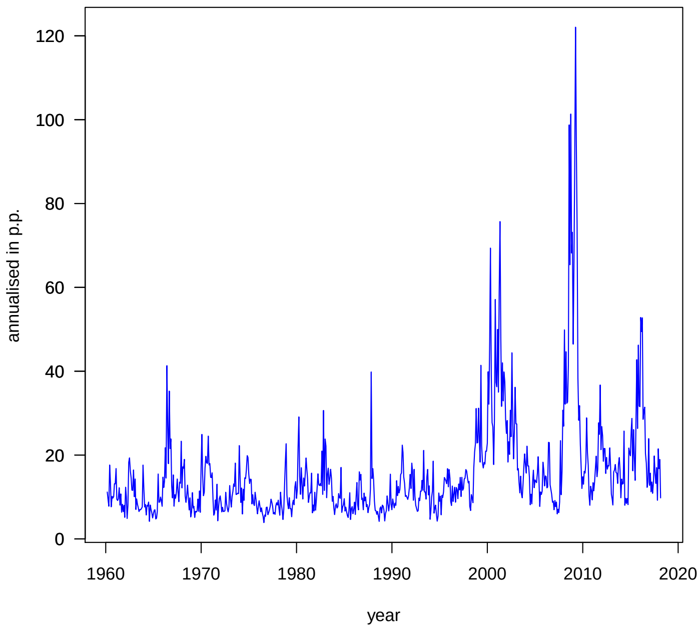
ancillary information on the weights of the strategies. On average, my strategy's exposure to the WML portfolio is slightly below 100 percent while the exposure from Barroso and Santa-Clara (2015) is on average 91.6 percent. Over the entire time horizon, my strategy's exposure to the momentum portfolio ranges between 13 and 229 percent.

Figure 19: Development of Momentum Strategies from 1965:02 to 2017:12



This figure shows how a one-dollar investment in February 1965 would have evolved until December 2017 following the different strategies. The investment strategy is as follows. At start, an investor puts exactly one dollar into the risk-free asset while simultaneously investing a certain percentage (dependent on the strategy) of that risk-free investment in the WML portfolio. Each month, this strategy reinvests the accumulated wealth in the risk-free rate and again takes a certain percentage of this investment in the WML portfolio. Even though, the investment strategy started in February 1965, one can recognise a growth only from the year 1990 onwards as the scale is very huge. This figure shows nicely the difference in the cumulative returns at the end, whereas Figure 13 on page 33 shows better the evolution over the whole time horizon.

Figure 20: Realised Monthly Volatility of Momentum
Source: Author's rendering of data from the Kenneth R. French Data Library (2018).



This figure shows the annualised realised monthly volatility for the time horizon of interest. It is used to calculate the Sharpe ratio, which helps to compare the success of the different portfolios. The realised variance is calculated as follows.

$$\hat{\sigma}^2_{RV,t} = \frac{1}{21} \sum_{j=0}^{20} (r_{WML,d_t-j} - \bar{r}_{WML})^2,$$

where

- $\hat{\sigma}^2_{RV,t}$ is the monthly realised variance,
- $\{r_{WML,d}\}_{d=1}^{D}$ are the daily returns,
- $\bar{r}_{WML}$ is the mean of the WML returns of the previous 21 days including day t.[37]

Particularly noticeable is the high level of annually realised monthly volatility around the dotcom bubble and the financial crisis.

[37] They are calculated as follows: $\bar{r}_{WML} = \frac{1}{21} \sum_{j=0}^{20} r_{WML,d_t-j}$.