Algorithmic Finance 2 (2013) 45–58
DOI 10.3233/AF-13016
IOS Press

Nonlinear support vector machines can systematically identify stocks with high and low future returns

Ramon Huerta[a], Fernando Corbacho[b,*] and Charles Elkan[c]
[a]*Biocircuits Institute, University of California San Diego, San Diego, CA, USA*
Tel: 1-858-534-1942, Fax: 1-858-534-1892; E-mail: rhuerta@ucsd.edu
[b]*Computer Science Department, Universidad Autonoma de Madrid, Madrid, Spain*
E-mail: fernando.corbacho@cognodata.com
[c]*Computer Science Department, University of California San Diego, San Diego, CA, USA*
E-mail: elkan@ucsd.edu

Abstract. This paper investigates the profitability of a trading strategy based on training a model to identify stocks with high or low predicted returns. A tail set is defined to be a group of stocks whose volatility-adjusted price change is in the highest or lowest quantile, for example the highest or lowest 5%. Each stock is represented by a set of technical and fundamental features computed using CRSP and Compustat data. A classifier is trained on historical tail sets and tested on future data. The classifier is chosen to be a nonlinear support vector machine (SVM) due to its simplicity and effectiveness. The SVM is trained once per month, in order to adjust to changing market conditions. Portfolios are formed by ranking stocks using the classifier output. The highest ranked stocks are used for long positions and the lowest ranked ones for short sales. The Global Industry Classification Standard is used to build a model for each sector such that a total of 8 long-short portfolios for Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, and Information Technology are formed. The data range from 1981 to 2010. Without measuring trading costs, but using 91 day holding periods to minimize these, the strategy leads to annual excess returns (Jensen alpha) of 15% with volatilities under 8% using the top 25% of the stocks of the distribution for training long positions and the bottom 25% for the short ones.

Keywords: Support vector machines, sector neutral portfolios, long-short portfolios, technical analysis, fundamental analysis.

1. Introduction

The question explored in this paper is whether there are features in accounting data and in historical price information that can help to predict stock price changes of companies. To address this question, we train predictive models on sets of stocks that undergo significant price changes. For instance, a 5% quantile selection means that we take those stocks whose positive (negative) volatility-adjusted price returns are in the top (bottom) 5% among all stocks. These 10% of all stocks are used for training a nonlinear support vector machine (SVM) to learn correlations between the features of a stock and the class it belongs to (top or bottom). Which quantile threshold best captures significant correlation between future changes of the stock price and fundamental and technical data is a key issue that we investigate.

In the empirical finance literature, portfolios are typically formed based on a set of equities ranked by a particular scalar that is believed to reflect an inefficiency in the market; see Chopra et al. (1992), Jegadeesh and Titman (2012), and Sloan (1996). Our work is similar in that a scalar provided by a classification function is used to rank the stocks. The stocks with the highest score are used for long

*Also at Cognodata Consulting, S. L.

positions, while the stocks with the lowest score are used for short sales. A straightforward approach to create the classification function is to use one of the most successful and convenient methods developed by machine learning researchers, support vector machines (SVMs).

An important innovation in our approach is the selection of the data used to train the classifier. We do not use all the available data. Omitting stocks that are in the middle of the distribution of returns improves performance. Equities with mid-ranking volatility-adjusted returns tend to follow the trend of the market, or to be idiosyncratic, and there tend to be no strong correlations to be identified by a classifier. A useful consequence of this observation is that one can train the classifier faster, leading to significant reduction in computational time.

2. What is an SVM?

In this section we provide a brief explanation of what an SVM classifier is. The reader can skip this section without losing the ability to understand the rest of the paper. We chose a nonlinear SVM classifier because it works well in multiple applications, is convenient to use and fast to train; see Muller et al. (2001), Shevade et al. (2000), and Vapnik (1999). There is a myriad of alternative methods that might to do as good a job as an SVM, but the simplicity of the mathematical functions, and the theory that frames the training of the model as a convex optimization problem (Boyd & Vandenberghe, 2004) make SVMs a preferred option. An important feature of convex optimization problems is a guarantee that there is a single optimal model to fit the data. In addition, we also have our own SVM implementation that allows a great flexibility in inserting the code into a forward testing algorithm.

Another important point of discussion is the choice of the type of SVM. There are two alternative paradigms: linear versus nonlinear. Linear SVMs are fast to train and execute, but they tend to underperform on complex datasets with many training examples and not too many features. Nonlinear SVMs can be more consistent in performance across different problems, and are the preferred option in many applications, albeit losing explanatory power.

For the sake of simplicity and for the purposes of this article, let us explain how an SVM function classifies a stock that has M features or input variables. In other words, the feature vector $\mathbf{x}$ has M components. Typically, the number of features varies from 7 to 51 depending on whether we use technical data, fundamental data, or both. An SVM classification function is

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i \, y_i \, K(\mathbf{x}, \mathbf{x}_i) - b$$

where

- $\mathbf{x}_i$ is all the vectors of the training set. Given a training history of tail sets, the SVM keeps them in memory because they will be used for prediction purposes.
- N is the number of training examples used to fit the SVM parameters, which varies from 10,000 to 100,000 examples depending on the history, the sector, and the quantile threshold on the distribution of volatility-adjusted returns.
- α_i is a scalar, that is a real number, that takes values between 0 and C. The value of C is important because it indicates how much emphasis we give to fitting the model closely to the training data. If we make it high we might incur overfitting: even though all data points might be well classified in the training set, the model might lose its ability to generalize in out-of-sample tests.
- y_i identifies whether the feature vector $\mathbf{x}_i$ of the stock i belongs to the tail set $+1$ or not -1. We use $+1$ to label positive stock returns and -1 to label the opposite ones. The stocks that lie in the middle of distribution are disregarded for training.
- b is obtained by training the SVM, and is a scalar that shifts the output of the SVM by a constant.
- Finally, $K(\mathbf{x}, \mathbf{x}_i)$ is the kernel function. A kernel is a function that takes two vectors as inputs and produces a single scalar value which is positive. The kernel function has a series of interesting properties that are required to make the SVM work (Muller et al., 2001). For our investigation, we use the Gaussian kernel

$$K(\mathbf{x}, \mathbf{x}_i) = e^{-\gamma \|\mathbf{x}-\mathbf{x}_i\|^2/M}$$

such that when $\mathbf{x}$ is identical to $\mathbf{x}_i$ the kernel value is 1, and when they are far apart the kernel value is negligible.

The SVM function is trained such that $f(\mathbf{x})$ is larger or equal than 1 if $\mathbf{x}$ belongs to class $+1$, and smaller or equal than -1 when it belongs to class -1. The α_i values and the b value are selected to match these requirements.

The meta parameters are C and γ. An important area of active research in machine learning is meta parameter search. These two meta parameters are chosen according to past performance on a training window, which emulates the way this system would be operated in real conditions today: evaluate the system performance in a given history for different pairs C, γ, choose the best set of parameters and use them to train the model for tomorrow's opening. One needs to replicate this procedure carefully to validate the model without incurring forward-looking bias. The software used to train the SVMs is based on Muezzinoglu et al. (2010), and is made available by Huerta (2010).

3. Previous applications of SVMs to financial data

The application of SVMs to financial data has been mostly focused on index prediction; see Kim (2003) and Sewell (2010). For example, Van Gestel et al. (2001) used and SVM for one step ahead prediction of the 90-day T-bill rate in secondary markets and the DAX 30. SVMs were used for regression purposes instead of classification, and the feature vector was based on lagged returns of the index, bond rates, S&P500, FTSE and CAC40. The paper showed that a rolling approach for meta parameter selection yielded better performance. The rolling approach selects the meta parameters using all currently available past information. A variation of a rolling approach is what we apply here. Other examples of SVM regression for futures index prediction are found in Tay and Cao (2001, 2002) and Cao and Tay (2003) that tend to beat artificial neural network approaches. Index prediction by Huang et al. (2005) and Kim (2003) was also used in the context of SVM classification to predict the directional changes of the markets. In all these cases the importance of the meta parameter search is emphasized in the performance of the prediction. According to the review by Sewell (2010), SVM and its variations outperformed other methods by a significant margin.

An interesting approach that uses fundamental data to predict credit ratings was developed by Huang et al. (2004). SVMs are used to classify with a significant level of success the ratings of the companies. It was pointed out that different markets and sectors have different subset of factors for classification. This is an aspect that we take into consideration in our work, by building different models for different sectors within the U.S. market. Similarly, in Hu et al. (2009) and Ding et al. (2008), SVMs based on fundamental data predict quite successfully stock crises and the financial conditions of the companies in the Chinese market.

The most related work to our contribution is perhaps due to Fan and Palaniswami (2001) and Ruei-Shan (2008), where SVM classifier outputs are used to rank the best long stocks in the Australian and Taiwanese markets. The authors used the whole database for training, defining the positive class to be the top 25% of the stocks; the remaining 75% are assigned to the negative class. The authors achieve significant excess returns in long-only equally weighted portfolios, but they do not control for liquidity and size in the portfolios.

Many of the published papers using SVMs in financial data emphasize the impact on the results of the meta parameter selection. It is crucial to make sure that the selection of meta parameters has no forward-looking bias. In particular, if a method tries several values of C, γ with the training set, and reports only the ones that do best on the test set, t then the method is incurring forward-looking bias. To avoid this common problem, the meta parameters must be chosen based only on past information. A distinct contribution of our work is the selection of meta parameters by a type of reinforcement learning.

4. Data description

The data are downloaded from the merged CRSP/ Compustat database and range from 1981 to 2010. This period of time is interesting because it includes the crash of 1987, the Long-Term Capital Management induced crash of 1998, the technology bubble of the 1990s, its bursting in 2000, and the major crisis of 2008. It also incorporates the growth in automatic trading in recent years.

4.1. Data filters

Three filters are applied to form a database of tradable stocks. One filters by a proxy for liquidity (LIQ), another one by dollar trading volume (DTV), and the last one by stock price only. For a given asset, the liquidity calculation involves regressing the market returns on the dollar volume, taking into account the sign of the order flow; see Kyle (1985) and Pastor and Stambaugh (2001). Specifically, LIQ is calculated following Khandani and Lo (2011), which is based on Kyle (1985). The daily returns $r(t)$ are regressed on the prices $p(t)$ and volume $v(t)$ according to the equation

$$r(t) = c + \lambda \operatorname{sign}(t) \log v(t)p(t)$$

where $\operatorname{sign}(t)$ takes values $+1, -1$ depending on the direction of the trade: $+1$ for net buyer initiated trades and -1 for net seller initiated trades. For our purposes, in the absence of order book data, we assume that if $r(t) \geq 0$ then $\operatorname{sign}(t) = 1$ otherwise $\operatorname{sign}(t) = -1$.[1] The regression coefficient λ is used as an inverse proxy of liquidity. We estimate it using all the trading days in a 91-day window.[2] The inverse of liquidity factor quantifies how much impact the dollar traded amount has on the price change of the day. The logarithm captures the observation that for high values of dollar traded amounts, the expected price changes are not so large. Figure 1 shows an example of the median of λ for all the stocks in some sectors. The S&P 500 index is also presented to show how the liquidity proxy characterizes different market periods. The most recent spike occurred after the Lehman Brothers bankruptcy.

The DTV is formed by multiplying the daily volume by the price of the stock, $p(t)v(t)$. This filter eliminates stocks that do not have sufficient capacity to be traded by mutual funds. For example, if a stock trades one million dollars in a day, one cannot expect to open a position of that size. The daily values of DTV are smoothed using a daily exponential moving average as $e(t) = \alpha p(t)v(t) + (1-\alpha)e(t-1)$ with $\alpha = 2/(91+1)$.

Since we want to simulate real trading conditions as much as we can, every trading day before opening positions, the filters are run. If we train using a stock that was below a threshold in the past, but is above at present, we will be introducing spurious training examples in the SVM model. The cutoffs of the filters are the bottom 50% for DTV and price, and the top

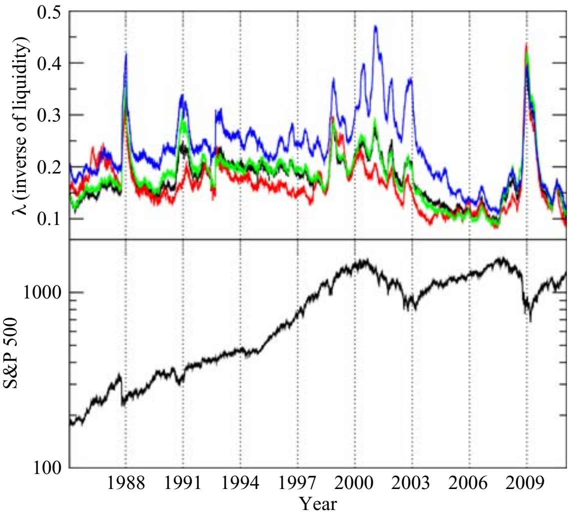

Fig. 1. (Top panel) Inverse of liquidity factor for four sectors: Energy (red), Industrials (black), Consumer Discretionary (green), and Information Technology (blue). For comparison, the bottom panel shows the S&P 500. One can detect big spikes in low liquidity for the Black Monday of 1987, the Long-Term Capital Management crash of September 1998 (see Pastor & Stambaugh, 2001, pp. 643–644), the dotcom bubble burst, and the recent collapse of Lehman Brothers. It is striking to see the current levels of high liquidity compared with recent history.

50% for λ on the LIQ filter.[3] If a stock that belongs to the tradable list falls below the cutoff mark during the holding period in the portfolio, the stock is kept until the position is closed. Doing otherwise would introduce a significant forward survival bias in the long portfolio, by only keeping in the tradable list stocks that are doing better than average. We apply the DTV and the LIQ filter because we do not want to learn correlations of stocks which are difficult to trade.

In order to align the fundamental data to the time when the data is actually visible to investors, we primarily use the Final Date (FDATEQ) from Compustat. If that is not available, we use the Report Date of Earnings (RDQ) with a delay of 45 days. This is a conservative approach, according to CapitalIQ (2009). If FDATEQ and RDQ are both unknown, then we take the current quarter and add 3 months. We want to avoid look-ahead bias as much as possible. To optimize the method further, a close look at when the fundamental data is visible should be paid. This is a full research topic by itself, and we will not further elaborate on

[1] Another approach consists of comparing the mid-value of bid-ask with the last sale. If the mid-value is larger than the last sale then $\operatorname{sign}(t) = 1$.

[2] The choice of 91 is because it is a multiple of 7 that lies close to a 3-month cycle.

[3] Stocks filtered by LIQ, DTV and price are not mutually exclusive. Real price and volume values, not adjusted ones, are used to calculate the filters.

Table 1
Number of stocks per GICS sector from 1981 to 2010 in the CRSP/Compustat merged database

Sector	GICS label	Number of stocks	Percentage
Energy	10	944	5.5%
Materials	15	858	5%
Industrials	20	2231	13%
Consumer Discretionary	25	2930	17.1%
Consumer Staples	30	639	3.7%
Health Care	35	1855	10.8%
Financials	40	2873	16.8%
Information Technology	45	2920	17%
Telecommunication Services	50	336	2%
Utilities	55	290	1.7%
Unclassified		1221	7.14%

the subject in this paper. Since the U.S. markets trade foreign companies, we use the Incorporation Code field (FIC) to select U.S. companies only.

4.2. Sector separation

Different fundamental data fields can have a different impact in each sector.[4] We build a model for each sector defined by the Global Industry Classification Standard (GICS): Energy (10), Materials (15), Industrials (20), Consumer Discretionary (25), Consumer Staples (30), Health Care (35), Financials (40), Information Technology (45), Telecommunication Services (50), and Utilities (55). Table 1 shows the number of stocks per sector. Some sectors do not have sufficient data to build a model. Specifically, sectors 50 and 55 have very few stocks, so we omit them. A fairly large number of stocks are unclassified. In this paper we disregard them, but a study via correlation analysis should be able to classify them into the known sectors. Methods are known for identifying sectors automatically; see Doyle and Elkan (2009).

5. Constructing the training data

One of the key issues in this investigation is how to form the tail sets that constitute the positive and negative classes of the training data. We investigate three types of metrics: i) real returns, ii) residual α after regression on the sector index, and iii) returns divided by a volatility estimate. Option i) that uses real returns makes the system focus on stocks with high volatility and ignores stocks with larger market cap, which usually have much smaller volatility. Strategies that use real returns tend to generate larger drawdowns and volatility. Option ii) that regresses returns on the sector index makes the system focus on examples with excess returns, discounting correlation to the index. This choice carries a higher computational load, because it requires the α values for each day and for each stock. Option iii) creates an ordered list of stocks with volatility-adjusted returns. The estimate of the volatility is an exponential moving average

$$\mathrm{vol}(t) = \kappa|R(t-t',t)| + (1-\kappa)\mathrm{vol}(t-1)$$

where $R(t-t',t)$ is the return of the stock from time $t-t'$ to t and $\kappa = 2/(180+1)$. Note that if one calculates volatility using standard deviations, the computational load is higher than for the method we propose. Moreover, Huffman and Moll (2011) show that risk measured as the mean absolute deviation has more explanatory power for future expected returns than standard deviation. Preliminary experiments show no significant difference between both methods of estimating volatility. The volatility-adjusted return is

$$r(t-t',t) = \frac{R(t-t',t)}{\mathrm{vol}(t)}.$$

Option iii) is our preferred choice. Note that our t' of choice to calculate the risk adjusted returns is 1 that refers to the previous trading day. For illustration, we show an example for the Industrials sector in Figure 2. In the bottom panel we show the performance on the Industrials sector for the three methods of creating the training labeled examples. All the methods are capable of delivering excess returns, but regressed alphas and volatility-adjusted returns lead to better performance.

[4] GSECTOR variable in the CRSP/Compustat Merged Database - Fundamentals Quarterly.

Table 2
Technical indicators used

Feature	Type	Reference
α_{3m}	Momentum 3 months	Jegadeesh and Titman (2012) and Rouwenhorst (2002)
α_{1y}	Momentum 1 year	Jegadeesh and Titman (2012) and Rouwenhorst (2002)
ΔV_{3m}	Volume change 3 months	Lee and Swaminathan (2000) and Chordia and Swaminathan (2002)
ΔV_{1m}	Volume change 1 month	Lee and Swaminathan (2000) and Chordia and Swaminathan (2002)
$N_{high,low}$	Contrarian	Mizrach and Weerts (2009)
$maxR$	Contrarian	Bali et al. (2011)
RL	Resistance levels	

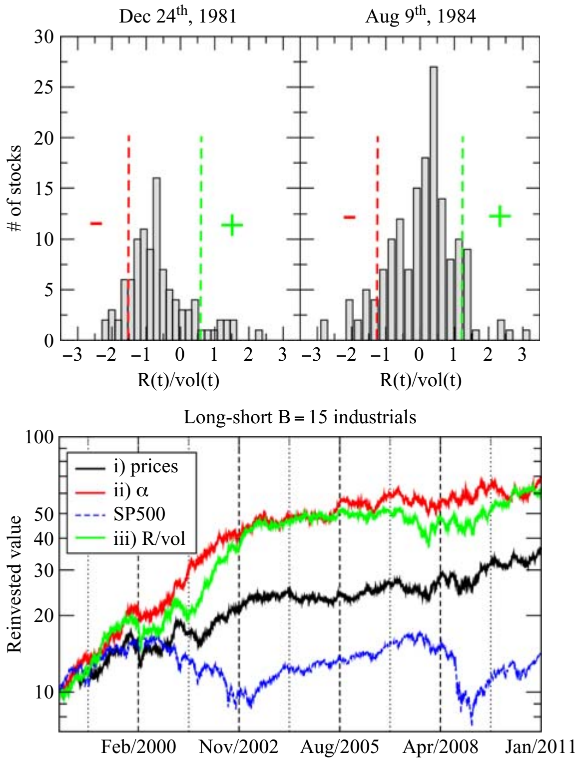

Fig. 2. (Top panel) Distributions of 91-day volatility-adjusted returns for the Industrials sector, starting on two arbitrarily chosen days. The positive tail sets are the 10% most positive volatility-adjusted returns, and the negative tail sets are the 10% most negative. The vertical dotted lines represent the decile cut. The $+$ and $-$ regions are the ones used for SVM training. (Bottom panel) SVM model performance based on the three options to define examples of excess returns on Industrials, using a quantile of $B = 15\%$ at the top and the bottom of the distribution, using the fundamental features.

In the rest of the paper we denote by B the quantile of volatility-adjusted returns used in designing the training data sets.

6. The features

Each stock on day t is characterized by a vector of technical and fundamental features $\mathbf{x}_i(t)$. The technical features are calculated from CRSP and the fundamental features are obtained from Compustat. Features are chosen based on their being mentioned in the academic literature. Considering a broader range of candidate indicators, and applying a feature selection algorithm such as the QPFS method proposed by Rodriguez-Lujan et al. (2010) is a topic for future research.

6.1. *Technical features*

Momentum has been one of the dominant factors covered in the literature; see for example Jegadeesh and Titman (2012) and Rouwenhorst (2002). It has been reported that stocks with high (low) returns over periods of three to 12 months continue to have high (low) returns over subsequent three to 12 month periods. Thus, we use two features embodying momentum as the excess returns calculated with respect to the S&P 500 over the previous three months, α_{3m}, and another one calculated over the previous one year, α_{1y}.

It is believed that if a price change occurs on heavy volume, the movement is more significant than if it happens on low volume; see Chordia and Swaminathan (2002) and Lee and Swaminathan (2000). Volume is a way to characterize underreactions and overreactions in stock price changes. To capture this phenomenon, we include a percentage change of the average daily trading volume over the last three months compared to the daily volume of the last year, ΔV_{3m}. We also include the volume change in the last month, ΔV_{1m}, to capture recent movements.

We include the number of highs and lows as suggested by Mizrach and Weerts (2009). They observed a contrarian indicator for excess returns based on the number of n-day highs and n-day lows, $N_{high,low}$. A similar contrarian indicator is proposed by Bali et al. (2011), where the maximum of daily returns, $maxR$, is considered an indicator of the interest of traders with few open positions.

Table 3
Fundamental indicators used

Feature	Formula or variable	Reference
Total Revenue	TR	
Gross Profit		
Operating Income		
Income before Tax		
Income after Tax		
Net Income before extraordinary items		
Net Income	NI	
Dividends	D	
Diluted normalized Earnings per Share		
Cash and Equivalents	CE	
Short Term Investments		
Accounts Receivable	AR	
Total Inventory	TI	
Total Current Assets	TCA	
Total Assets	TA	
Short Term Liabilities		
Total Current Liabilities	CL	
Total Long Lerm Debt		
Total Debt	TD	
Total Liabilities		
Total Equity	TE	
Total Shares	SO	
Depreciation		
Cash from operating activities		
Capital expenditures		
Cash from investing activities		
Cash from financing activities		
Net change in cash		
Snapshot accrual	SAC=TCA-CE-CL+TD	Sloan (1996)
Accrual based on balance sheet	SAC(quarter)-SAC(quarter-4)	Sloan (1996)
Accrual based on cash flow		Bradshaw et al. (2002)
Financial Health		Piotroski (2000)
Working capital	TCA-CL	
Quick ratio	(TCA-TI)/CL	
Dividend payout ratio	D/NI	
Book value	BV	
Book value -Total Debt	BV-TD	
Receivables to sales	AR/TR	
Debt to Assets	TD/TA	
Debt to Equity	TD/TE	
Cash to Assets	CE/TA	
Liabilities to Income	TL/NI	
Return on Equity	ROE = NI/TE	
Sales per shares	TR/SO	

A proxy for the resistance level, RL, is also used because, at least psychologically, it can be an important factor for traders. The RL is calculated as the percentage difference from the closest peak in the past. One can also run feature selection algorithms (Rodriguez-Lujan et al., 2010) over a wide range of features (Kim & Han, 2000), but due to the high correlation between many features, the list of most informative features is not large. In addition, increasing the number of features raises the risk of overfitting.

In summary, we consider seven technical indicators mostly based on the academic literature. There is an abundant academic literature about the existence of correlations between additional technical indicators and future returns, but in general the correlations are expected to fade away in the future, because of the competition among agents attempting to profit from strategies based on them.

6.2. *Fundamental data*

The fundamental data is organized by blocks. A block has 44 features and is a snapshot of one stock at a particular time. As already mentioned, we primarily use the Final Date (FDATEQ) from Compustat to create snapshots.

The performance of the fundamental indicators is analyzed independently of that of the technical ones. Later both feature sets are merged, to provide a more complete snapshot of a stock.

7. Training protocol

We form portfolios of 10 equally weighted long and 10 equally weighted short positions. Each position is closed at the end of the last trading day in the following 91 days. Every 28 days we open an additional 20 positions. Therefore, on most days we have 60 total positions for each sector. The reason we take 20 new positions every 28 days, instead of opening 60 every 91 days, is to decorrelate temporally the natural fluctuations of the SVM classifier. Positions could be opened more frequently, to decrease the correlations between portfolios further, but that would require extra effort in maintaining the portfolios through time.

An important characteristic of our approach is that a fresh model is trained every time a new portfolio is to be formed, in order to adapt to changing environments. The SVM is trained over tail sets in the period of time $[\tau - h, \tau]$ and tested on the period $[\tau + d, \tau + r]$ with $d \geq 1$. The parameter h is the length of history used to build the model, d identifies the nearest trading day after τ, and r is the rebalancing period of the portfolio. For example, if today is Friday, June 15, 2012, $d = 3$ because the next trading day is Monday, June 18, 2012. We train with the the preset history $[\tau - h, \tau]$ and test from Monday to the following $\tau + r$ day. To make sure that there is no look-ahead during training, only data within times $[\tau - h, \tau]$ is loaded by the program during execution. This is not a fast way to implement the method, but it increases confidence that there are not leaks in the training set of future data. If one trains with data from even just one day after τ, there is already a significant forward-looking bias. The classifier then mostly concentrates on that bias, which is essentially the correlation with the movement of the overall market.

We denote by $SVM_{\tau,h,r}$ the model trained at time τ with history h and rebalancing period r. A ranking of stocks provided by the ordered values of $SVM_{\tau,h,r}(\mathbf{x}_i(\tau+d))$ is formed where d is the number of days to the following trading day and r is usually 91 days. The stocks with the highest values are chosen as long positions and the stocks with the lowest values are short sales. Note that, even though the system is trained for a particular quantile B, the SVM evaluation for time $\tau + d$ is run over the list of all tradable stocks obtained after running the filters. The net dollar value of the long positions is equal to the dollar value of the short ones. Since the trend of the longs is to increase in value and the trend of the shorts is to decrease in value, the portfolios become slightly unbalanced at the end of the rebalancing period. However, the correlation of the portfolio with the S&P 500 index is almost zero. Thus, it is not critical to actively manage the portfolio to keep the long and short positions balanced.

8. Selection of the meta parameters

The meta parameters of the SVM model are the C and γ values. The question of how to choose these parameters without incurring overfitting is a major one in the machine learning community (see Cawley and Talbot (2010) for a detailed examination of the problem). Often, in a standard classification problem, a dataset is separated into two mutually exclusive sets. One of them is used for training and the other one is used for testing. The training set is further split into sets for training and validation using a technique called cross-validation. The parameters C and γ are chosen to maximize performance on the validation subsets, over all the partitions into training and validation subsets.

In our problem there is an additional difficulty due to the lack of stationarity of the underlying probability distributions of each of the classes. The best one can hope for is that the underlying statistics in the past will be in the vicinity, in the near future. Obviously there are no guarantees that this will be the case. Moreover, once significant correlations are observed by a large number of market participants, then the market is expected to price them in.

To adjust the meta parameters C and γ, we define an action a as a pair of values of C and γ used to build a portfolio at time t. We use a grid of 16 values with $C = 0.5, 1, 2, 4$ and $\gamma = 0.5, 1, 2, 4$. Borrowing terms from the field of reinforcement learning (see Sutton & Barto, 1998), we denote the quality of an action at time t as $Q(t, a)$, which is updated using an exponential moving average as

$$Q(t, a) = (1 - \alpha)Q(t - 1, a) + \alpha R(t - 1, a)$$

where α is a learning rate that we typically choose to average over three years, and $R(t - 1, a)$ is the reward obtained by using the action a in the immediate past. For our application we choose pure return as the reward. The meta parameter values used to form the portfolio at time t are obtained from

$$a_t = \arg\max_a \, Q(t, a).$$

We essentially take the parameters that worked best on average in the recent past. Table 4 shows an example of running the algorithm on the Industrials sector in the last 30 years. Almost all meta parameter values are picked occasionally, but two cells dominate. A drawback of nonlinear SVMs is that meta parameter values affect performance strongly, but it is difficult to explain why some values are better. In the rest of the paper the meta parameters are chosen using the algorithm just described.

Table 4
Meta parameter choices from 1982 to 2010. The entries in the table are the number of occurrences during training. Although most values are picked at least once, [0.5, 0.5] and [2, 0.5] are the most common

	γ			
	0.5	1	2	4
$C = 0.5$	**40**	5	10	0
$C = 1$	11	9	12	9
$C = 2$	**52**	8	1	0
$C = 4$	11	0	15	5

9. Results and analysis on the Industrials sector

Let us start the results section by showing yearly returns using the SVM models trained on randomized labels. These serve as control for the rest of the results. Figure 3 shows a histogram corresponding to 100 realizations of random models on the Industrials sector. The maximum value is 6.45%, which means that any alphas over 6.45% have a p-value of 0.01 or less.

The analysis is separated into three different experiments: i) technical features only, ii) fundamentals only, and iii) combined technicals and fundamentals. Figure 4 shows the result for the best B for each feature set. Each series is based on 60 positions from 3 subportfolios within the sector, with an investment horizon of 91 days separated by 28 days.

Let us compare the yearly returns and the volatility for each of the three different sets of features: technicals, fundamentals, and combined (see Table 5). The performance of the technical set is better than the fundamental one, with both displaying similar levels of volatility. The best performance is achieved for the joint feature space of technicals and fundamentals, with yearly returns that are well above the maximum value of 100 replications of random SVMs, as shown in Figure 3. It is interesting to note that while the fundamental and combined feature spaces have an optimal quantile threshold in the domain $B \in [20, 40]$, the model using technical features should use almost all the data for training, with $B = 45$.

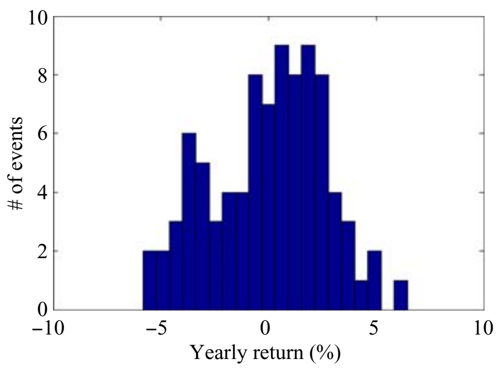

Fig. 3. Histogram of the number of observations versus yearly returns on the Industrials sector using random models. The correlation, β, with the S&P 500 is nearly 0.

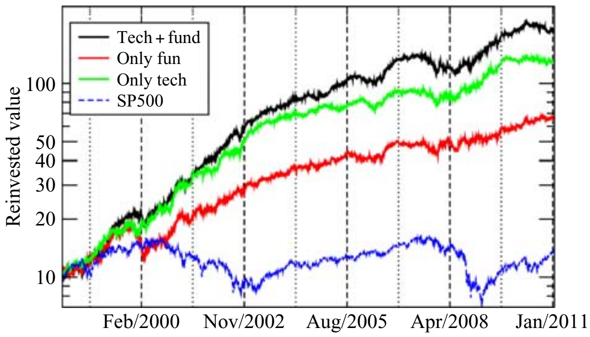

Fig. 4. Comparison of the three feature sets. The blue curve shows the S&P 500 index while the black, green, and red lines are the total value of portfolios of 60 positions in the Industrials sector. Transaction costs are not included. The green line is the result of using the optimal $B = 45$ value for the technical features, and the red one is using only the fundamental features with $B = 20$. The combined set leads to the best performance overall, using $B = 25$.

We also run a direct comparison of linear SVMs and nonlinear ones. The results are shown in Table 6. Linear SVMs lead to returns inferior to those yielded by nonlinear SVMs. It is interesting to note that the performance of the linear SVMs is better for quantiles $B = 10, 15, 20, 25$, consistent with the nonlinear SVMs. This result is a corroboration that the information present in the tails is sufficient to learn a classifier that detect patterns that deviate from the norm. An advantage of linear SVMs is that one can begin to isolate the importance of each individual factor in each sector. This may provide additional information for a portfolio manager who wants to understand the characteristics for which the model selects. This is a line of work left for the future.

10. Results on the aggregated sectors

Once we build the portfolio for each sector we can integrate them, excluding Utilities and Telecommunication Services. These two sectors represent no more than 2% of the whole universe of stocks. Thus, the portfolios are formed by 8 sectors with 60 open positions in each sector at any particular time, which leads to 240 long and 240 short positions in total for an equally weighted aggregate portfolio. Health Care is a high risk sector, as shown by the dramatic drawdowns in Table 7 regardless of the quantile. Nevertheless, there is no reason to leave this sector out, since its volatility is similar to that of the Information Technology sector.

Table 8 shows the Jensen alpha, the Sharpe ratio, and the maximum drawdown. Both the Jensen alpha and the Sharpe ratio require a risk free interest rate, which we obtain from the CRSP database from 4-week and 3-month Treasury bills.[5] The Jensen alpha is a

[5]The 4-week and 3-month Treasury bill rates from the secondary market were downloaded from the CRSP database *Interest Rates (Federal Reserve, H15 report)*. The 4-week rate was taken as the default risk-free rate unless it was empty. If it was empty the 3-month rate was chosen.

Table 5

Yearly returns with their volatilities and the ratios between them for the Industrials sector, for different quantiles B of the distribution as described in Section 5. Returns do not include transaction costs. Not surprisingly, the best performance is achieved for the combined technical and fundamental features. Horizontal bars delimit the region of B values that leads to better performance. The rightmost column shows the largest observed drawdown, D, i.e. percentage change from peak to trough

B	Technicals			Fundamentals			**Combined**			
	r	σ	r/σ	r	σ	r/σ	r	σ	r/σ	D
5	15.14%	11.39%	1.32	13.05%	12.86%	1.01	20.51%	12.37%	1.65	−24.24%
10	15.44%	11.40%	1.35	12.99%	13.01%	0.99	18.74%	12.70%	1.47	−33.19%
15	15.84%	12.39%	1.27	14.23%	12.72%	1.11	20.53%	12.59%	1.63	−36.57%
20	13.12%	12.46%	1.05	15.13%	12.48%	1.21	21.92%	12.79%	1.71	−24.19%
25	14.51%	12.55%	1.15	14.78%	12.75%	1.15	23.15%	12.67%	1.83	−22.83%
30	14.30%	12.50%	1.14	14.45%	12.96%	1.11	21.22%	12.42%	1.70	−22.03%
35	17.33%	11.95%	1.44	11.30%	12.85%	0.87	21.55%	12.28%	1.75	−17.42%
40	19.48%	11.95%	1.62	13.19%	13.28%	0.99	21.85%	12.28%	1.77	−17.57%
45	20.22%	11.96%	1.69	11.93%	13.28%	0.89	20.26%	12.98%	1.60	−14.42%
50	16.53%	12.62%	1.30	9.69%	13.61%	0.71	20.00%	12.65%	1.58	−21.18%

Table 6
Performance comparison between linear SVMs and nonlinear SVMs. Yearly returns with their volatilities, ratios, and drawdowns for the Industrials sector, for different quantiles B as described in Section 5. Yearly returns do not include transaction costs. Nonlinear SVMs outperform linear SVMs, but linear SVMs also achieve excess returns

B	Linear SVMs				Nonlinear SVMs			
	r	σ	r/σ	D	r	σ	r/σ	D
5	12.91%	13.80%	0.93	−23.79%	20.51%	12.37%	1.65	−24.24%
10	12.66%	13.62%	0.93	−26.41%	18.74%	12.70%	1.47	−33.19%
15	14.27%	13.21%	1.08	−28.18%	20.53%	12.59%	1.63	−36.57%
20	13.78%	13.23%	1.04	−23.82%	21.92%	12.79%	1.71	−24.19%
25	12.17%	13.28%	0.91	−24.99%	23.15%	12.67%	1.83	−22.83%
30	11.44%	13.20%	0.86	−22.74%	21.22%	12.42%	1.70	−22.03%
35	11.95%	13.05%	0.91	−20.80%	21.55%	12.28%	1.75	−17.42%
40	11.09%	13.01%	0.85	−21.29%	21.85%	12.28%	1.77	−17.57%
45	11.79%	12.93%	0.91	−22.28%	20.26%	12.98%	1.60	−14.42%
50	11.20%	13.10%	0.85	−25.11%	20.00%	12.65%	1.58	−21.18%

simple regression that discounts the risk-free rate as follows:

$$R(t-1,t) - R_{rf}(t-1,t) = \alpha_{Jensen} + \beta\,(R_{SP500}(t-1,t) - R_{rf}(t-1,t))$$

where $R(t-1,t)$ denotes the daily returns. The Sharpe ratio is computed from the monthly return minus the risk-free rate,

$$u(t) = R_{\text{monthly}}(t) - R_{\text{monthly risk free}}(t),$$

and the monthly volatility of the portfolio, $\text{vol}(R_{\text{monthly}}(t))$. The Sharpe ratio is then

$$S = \frac{12\mathbf{E}_t u(t)}{\sqrt{12}\,\text{vol}(R_{\text{monthly}}(t))}.$$

We also calculate the information ratio following Goodwin (1998) as the ratio between the average returns of the portfolio minus the S&P500 index and its corresponding standard deviation, namely

$$IR = \frac{12\,\mathbf{E}_t(R_{\text{monthly}}(t) - R_{SP500}(t))}{\sqrt{12}\,\sigma(R_{\text{monthly}}(t) - R_{SP500}(t))}.$$

11. Discussion

We have shown that one can train a classifier to predict returns in eight GICS sectors, and obtain excess annual returns of 15% (not counting transaction costs) with volatilities under 8%. Training uses the tails of the volatility-adjusted returns distribution corresponding to the top 25% of stocks as long positions and the bottom 25% as short positions; see Table 8. The stocks that lie in the middle of the distribution do not contribute to the performance of the system, suggesting that these stocks are driven by noise and do not have correlations with the technical and fundamental features that we analyze. The Industrials sector appears to lead to the best results.

In this investigation we focus on the importance of the tails of the distribution of historical volatility-adjusted returns. The recommendation is to use the smallest tails that lead to good performance, because the training and execution times of the SVM depend on the amount of data. However, the parameter that controls the size of the tails is another meta parameter of the system that should be trained using the reinforcement learning algorithm described above. Another issue for future work is that the strategy proposed above requires repeated training in order to let the SVM adjust to the most recent market conditions. A question that remains open is how often one really needs to train the classifier. Currently, we conservatively retrain the SVM the day prior to each trading day

The results shown in Table 8 are obtained using well-known indicators from the literature. This is public information available to all the agents trading in the markets. Thus, one cannot be too optimistic in expecting high levels of excess returns in the future, because machine learning tools and computing power are becoming widespread and available to all market

Table 7

Comparative table for nine GICS sectors using the combined technical and fundamental features. As explained in the text, each sector consists of a portfolio of 60 equally weighted positions, 30 long and 30 short. The Utility sector is shown for illustration purposes, but it has too few stocks for reliable training compared to the others. We show yearly returns with their volatilities, a proxy for the Sharpe ratio (the risk-free rate is not subtracted), and the largest drawdown, D, for the period of 20 years; all for different quantiles B as described in Section 5. Note that the yearly returns do not account for the transaction costs. A conservative estimate is about 1% for each rebalance, which leads to a 4% discount a year to the returns shown above. The volatility values hardly change with the quantile and the drawdowns do not have a significant trend. Nevertheless, they are helpful to identify risky sectors like Health Care and Consumer Discretionary

B	Energy				Materials				Industrials			
	r	σ	r/σ	D	r	σ	r/σ	D	r	σ	r/σ	D
5	10.5%	13.8%	0.7	−39.3%	9.6%	12.3%	0.7	−37.1%	20.5%	12.3%	1.6	−24.2%
10	10.2%	13.1%	0.7	−33.6%	9.7%	13.3%	0.7	−58.4%	18.7%	12.7%	1.4	−33.1%
15	8.7%	13.4%	0.6	−43.8%	7.7%	12.2%	0.6	−51.2%	20.5%	12.5%	1.6	−36.5%
20	7.8%	13.5%	0.5	−57.0%	9.2%	11.8%	0.7	−32.1%	21.9%	12.7%	1.7	−24.1%
25	9.0%	13.1%	0.6	−38.3%	9.9%	12.1%	0.8	−26.2%	23.1%	12.6%	1.8	−22.8%
30	9.4%	12.9%	0.7	−30.9%	7.5%	12.6%	0.5	−24.5%	21.2%	12.4%	1.7	−22.0%
35	9.6%	13.2%	0.7	−37.8%	7.5%	11.7%	0.6	−18.7%	21.5%	12.2%	1.7	−17.4%
40	10.0%	12.9%	0.7	−38.6%	4.7%	11.1%	0.4	−29.8%	21.8%	12.2%	1.7	−17.5%
45	10.8%	12.9%	0.8	−44.2%	8.6%	11.8%	0.7	−24.0%	20.2%	12.9%	1.6	−14.4%
50	9.8%	12.8%	0.7	−38.5%	6.4%	11.8%	0.5	−31.3%	20.0%	12.6%	1.5	−21.1%

B	Consumer Discretionary				Consumer Staples				Health Care			
	r	σ	r/σ	D	r	σ	r/σ	D	r	σ	r/σ	D
5	10.1%	14.7%	0.7	−72.6%	11.5%	11.6%	0.9	−26.7%	25.6%	17.2%	1.5	−48.3%
10	15.0%	14.0%	1.0	−55.7%	14.3%	11.3%	1.2	−26.6%	29.3%	17.0%	1.7	−49.7%
15	16.5%	13.9%	1.1	−42.4%	12.6%	11.4%	1.1	−27.2%	33.7%	17.8%	1.9	−60.9%
20	14.1%	14.3%	0.9	−58.7%	13.1%	11.4%	1.1	−31.7%	29.1%	18.6%	1.5	−73.2%
25	14.9%	13.4%	1.1	−42.0%	13.5%	11.3%	1.1	−25.5%	29.6%	18.7%	1.6	−67.6%
30	16.4%	13.8%	1.1	−52.8%	12.8%	11.3%	1.1	−18.9%	29.0%	19.5%	1.5	−70.1%
35	17.3%	14.2%	1.2	−44.1%	10.8%	11.3%	0.9	−20.3%	31.8%	19.7%	1.6	−52.8%
40	13.9%	13.8%	1.0	−62.0%	9.9%	11.0%	0.9	−22.9%	29.9%	21.3%	1.4	−73.4%
45	15.0%	14.4%	1.0	−60.0%	12.2%	11.3%	1.0	−20.0%	26.8%	20.6%	1.3	−63.6%
50	15.9%	14.5%	1.0	−55.9%	10.5%	11.2%	0.9	−22.8%	28.8%	20.2%	1.4	−54.5%

B	Financials				Information Technology				Utilities (too few stocks)			
	r	σ	r/σ	D	r	σ	r/σ	D	r	σ	r/σ	D
5	6.4%	13.1%	0.4	−33.9%	15.7%	19.0%	0.8	−41.7%	−1.7%	8.1%	−0.2	−35.3%
10	8.3%	12.8%	0.6	−23.6%	19.3%	19.2%	1.0	−36.8%	1.2%	7.2%	0.0	−19.7%
15	8.5%	12.2%	0.6	−20.9%	21.0%	18.4%	1.1	−27.2%	1.7%	6.5%	0.2	−22.1%
20	8.2%	11.9%	0.7	−21.2%	25.7%	18.7%	1.3	−31.7%	0.4%	6.8%	0.0	−26.3%
25	10.0%	12.6%	0.8	−18.5%	26.3%	18.4%	1.4	−35.9%	0.2%	6.3%	0.0	−25.2%
30	6.6%	12.4%	0.5	−23.1%	24.5%	17.4%	1.4	−25.8%	0.0%	6.3%	0.0	−29.5%
35	7.7%	12.7%	0.6	−19.1%	21.7%	17.4%	1.2	−33.3%	−2.2%	11.7%	−0.2	−68.3%
40	5.9%	12.1%	0.4	−25.0%	21.5%	17.7%	1.2	−34.6%	2.9%	11.5%	0.2	−27.9%
45	7.3%	12.2%	0.5	−29.1%	26.3%	17.3%	1.5	−33.8%	0.0%	11.4%	0.0	−45.6%
50	5.6%	12.9%	0.4	−33.8%	18.8%	17.5%	1.0	−48.4%	0.0%	11.9%	−0.0	−39.2%

participants. Ideally one wants to use indicators and information not widely available. A future research direction is to use a large set of potential indicators, and then to identify automatically subsets of features that have the strongest influence on excess returns. An approach that is capable of this type of feature selection is the QPFS method of Rodriguez-Lujan et al. (2010).

Table 8
Overall performance of the combined sectors, for increasing values of the threshold quantile of the distribution. The correlation to the index, β, is shown for completeness although it is close to zero. IR is Information ratio which is the ratio between the average returns of the portfolio minus the S&P 500 index and its corresponding standard deviation. The maximum drawdown is calculated within the range January 1998 to December 2010. The largest two drawdowns start in November 1999 and July 2007. The correlation to the unwinding hypothesis formulated by Khandani and Lo (2011) is striking; the middle of 2007 was a time when hedge funds and banks reduced leverage by selling their "good" stocks and buying back their short sales. Further work should analyze whether drawdowns are generally associated with low-liquidity situations as explained by Pastor & Stambaugh (2001). In any case, the SVM long and short choices appear to be not free from systemic risk

B	All sectors except Utilities and Telecommunications					
	Jensen α	β	Sharpe ratio	volatility	max drawdown	IR
5	11.31%	0.0041	1.26	8.96%	−26.32%	0.63
10	13.06%	−0.0047	1.64	8.14%	−21.73%	0.73
15	13.96%	−0.0035	1.74	7.97%	−18.62%	0.73
20	13.86%	−0.0096	1.80	7.71%	−20.02%	0.69
25	14.86%	−0.0167	2.06	7.29%	−12.27%	0.75
30	13.69%	−0.0173	1.89	7.18%	−12.54%	0.64
35	14.06%	−0.0236	1.88	7.41%	−11.18%	0.67
40	12.70%	−0.0210	1.77	7.27%	−14.86%	0.59
45	13.76%	−0.0222	1.79	7.73%	−14.08%	0.62
50	12.76%	−0.0199	1.81	7.01%	−9.03%	0.56

References

Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. J. Financ. Econ. 99 (2), 427–446.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, New York, NY, USA.

Bradshaw, M., Richardson, S., Sloan, R., 2002. Do analysts and auditors use information in accruals? J. Accounting Res. 39 (1), 45–74.

Cao, L., Tay, F., 2003. Support vector machine with adaptive parameters in financial time series forecasting. Neural Netw. IEEE Trans 14 (6), 1506–1518.

CapitalIQ 2009. Data navigator white paper: Exploring lagging error in quantitative modeling. http://www.google.com/search?gcx=c&sourceid=chrome&ie=UTF-8&q=CapitalIQ-Compustat+Data+Navigator+White+Paper%3A+Exploring+Lagging+Error+in+Quantitative+Modeling

Cawley, G., Talbot, N., 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. J. Mach. Learn. Res. 11, 2079–2107.

Chopra, N., Lakonishok, J., Ritter, J., 1992. Measuring abnormal performance: Do stocks overreact? J. Financ. Econ. 31 (2), 235–268.

Chordia, T., Swaminathan, B., 2002. Trading volume and cross-autocorrelations in stock returns. J. Financ. 55 (2), 913–935.

Ding, Y., Song, X., Zen, Y., 2008. Forecasting financial condition of chinese listed companies based on support vector machine. Expert Syst. Appl. 34 (4), 3081–3089.

Doyle, G., Elkan, C., 2009. Financial topic models. Working Notes of the NIPS-2009 Workshop on Applications for Topic Models: Text and Beyond Workshop.

Fan, A., Palaniswami, M., 2001. Stock selection using support vector machines. Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on, Vol. 3, IEEE, pp. 1793–1798. Washington, DC.

Goodwin, T., 1998. The information ratio. Financ. Anal. J. 54, 34–43.

Hu, Y., Su, P., Hao, X., Tang, F., 2009. The long-term predictive effect of svm financial crisis early-warning model. Proceedings of the 2009 First IEEE International Conference on Information Science and Engineering, pp. 4573–4576. ICISE '09, IEEE Computer Society, Washington, DC, USA.

Huang, W., Nakamori, Y., Wang, S.-Y., 2005. Forecasting stock market movement direction with support vector machine. Comput. Oper. Res. 32 (10), 2513–2522.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., Wu, S., 2004. Credit rating analysis with support vector machines and neural networks: A market comparative study. Decis. Support Syst. 37 (4), 543–558.

Huerta, R., 2010. Support vector classification and regression software. http://biocircuits.ucsd.edu/huerta/svmr.tar.gz

Huffman, S., Moll, C., 2011. The impact of asymmetry on expected stock returns: An investigation of alternative risk measures. Algorithmic Financ. 1 (2), 79–93.

Jegadeesh, N., Titman, S., 2012. Returns to buying winners and selling losers: Implications for stock market efficiency. J Financ. 48 (1), 65–91.

Khandani, A., Lo, A., 2011. What happened to the quants in august 2007? evidence from factors and transactions data. J. Financ. Markets 14 (1), 1–46.

Kim, K., 2003. Financial time series forecasting using support vector machines. Neurocomputing 55 (1), 307–319.

Kim, K., Han, I., 2000. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert Syst. Appl. 19 (2), 125–132.

Kyle, A., 1985. Continuous auctions and insider trading. Econometrica: J. Economet. Soc. 53, 1315–1335.

Lee, C., Swaminathan, B., 2000. Price momentum and trading volume. J. Financ. 55 (5), 2017–2069.

Mizrach, B., Weerts, S., 2009. Highs and lows: A behavioural and technical analysis. Appl. Financ. Econ. 19 (10), 767–777.

Muezzinoglu, M., Vergara, A., Huerta, R., 2010. A unified framework for volatile organic compound classification and regression. Neural Networks (IJCNN), The 2010 International Joint Conference on, IEEE, pp. 1–7. Barcelona, Spain.

Muller, K., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B., 2001. An introduction to kernel-based learning algorithms. IEEE Neural Network. 12 (2), 181–201.

Pastor, L., Stambaugh, R., 2001. Liquidity risk and expected stock returns. *Technical report*, National Bureau of Economic Research.

Piotroski, J., 2000. Value investing: The use of historical financial statement information to separate winners from losers. J. Accounting Res. 38, 1–41.

Rodriguez-Lujan, I., Huerta, R., Elkan, C., Cruz, C., 2010. Quadratic programming feature selection. J. Mach. Learn. Res. 11, 1491–1516.

Rouwenhorst, K., 2002. International momentum strategies. J. Financ. 53 (1), 267–284.

Ruei-Shan, L., 2008. A study on application of smooth support vector classification to stock selection in taiwan's stock market. Joint Conference of WEHIA & CIEF. Taoyuan, Taiwan.

Sewell, M., 2010. *The Application of Intelligent Systems to Financial Time Series Analysis*, PhD thesis, PhD dissertation, Department of Computer Science, University College London, University of London.

Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., Murthy, K. K., 2000. Improvements to the smo algorithm for svm regression. IEEE Trans. Neural Netw. 11 (5), 1188–1193.

Sloan, R., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? Account. Rev. 71, 289–315.

Sutton, R., Barto, A., 1998. Reinforcement learning: An introduction, vol. 1(1). Cambridge Univ Press, Cambridge, UK.

Tay, F., Cao, L., 2001. Application of support vector machines in financial time series forecasting. Omega 29 (4), 309–317.

Tay, F., Cao, L., 2002. Modified support vector machines in financial time series forecasting. Neurocomputing 48 (1), 847–861.

Van Gestel, T., Suykens, J., Baestaens, D., Lambrechts, A., Lanckriet, G., Vandaele, B., et al., 2001. Financial time series prediction using least squares support vector machines within the evidence framework. Neural Netw., IEEE Trans. 12 (4), 809–821.

Vapnik, V., 1999. The Nature of Statistical Learning Theory. springer, Heidelberg, Germany.