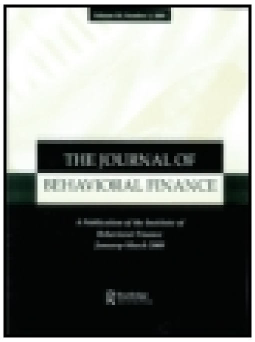

Journal of Behavioral Finance

ISSN: 1542-7560 (Print) 1542-7579 (Online) Journal homepage: https://www.tandfonline.com/loi/hbhf20

Exuberance in Financial Markets: Evidence from Machine Learning Algorithms

Jan Viebig

JOURNAL OF BEHAVIORAL FINANCE
https://doi.org/10.1080/15427560.2019.1663849

Routledge
Taylor & Francis Group

Exuberance in Financial Markets: Evidence from Machine Learning Algorithms

Jan Viebig

Goethe University Frankfurt

ABSTRACT
Motivated by *Campbell and Shiller (1998)*, we show that the probability that abnormally low returns over long-term investment horizons occur in the future is disproportionately high when equity markets trade at extremely high valuation levels. Support vector machines are able to learn "clustering patterns" from fundamental data with high precision rates. Decision boundaries calculated with machine learning algorithms can help investors to detect irrational exuberance in financial markets followed by abnormally low returns.

KEYWORDS
Exuberance; Machine learning; Fundamental data

Introduction

The perception that extreme boom and bust periods occur regularly in financial markets due to shifts in investor sentiment ranging from euphoria to panic was widespread long before academics formulated the efficient market hypothesis. The tulip bubble (1636-1637), the South Sea Company bubble (1711-1720), and the Mississippi Company bubble (1719-1720) are well documented in Charles MacKay's book *Extraordinary Popular Delusions and the Madness of Crowds*. Behavioral economists acknowledge that financial markets are subject to financial fashions and fads described by MacKay (1841) and view stock markets as complex, dynamic, social systems driven not only by fundamental, economic factors but also by investor sentiment. They point out that asset prices may diverge from fundamental values due to limits of arbitrage and cognitive biases (Shleifer, Vishny (1997), Thaler (1999), Kent et al. (2002), Shiller (2003), Subrahmanyam (2007), De Bondt et al. (2008), Hirshleifer (2015)).

As Shiller (1981a, 1981b, 1988, 1989) suggested that stock market movements are too volatile to be explained by new fundamental information and drift time and again far away from long-term fundamental valuation levels, John Campbell and Robert Shiller asked whether long-horizon stock returns are predictable when markets trade at extreme valuation levels (Campbell, Shiller (1987, 1988, 1989, 1998, 2001)). In their influential paper "Valuation Ratios and the Long-Run Stock Market Outlook" published in 1998, Campbell and Shiller analyzed a U.S. data set extending the S&P 500 index back until 1872 and an international data set including MSCI data series for 12 developed markets starting in 1970. Applying linear regression analysis, they found that dividend-to-price ratios and price-to-smoothed earnings ratios explain a large part of the long-term return variability in equity markets.

If equity markets systematically deviate from levels supported by fundamentals, machine learning tools should be able to recognize and learn these patterns. In this paper, we show that support vector machines are able to detect "clustering patterns" in long-term return series. Decision boundaries calculated with machine learning algorithms are able to detect irrational exuberance in financial markets followed by abnormally low long-term returns. Machine learning algorithms are able to learn "rules" from large datasets which can help investors to predict abnormally low long-term returns with high accuracy rates and high precision rates.

The remainder of this paper is organized as follows. The second section discusses the data. The third section introduces the probabilistic support vector machine approach used in this study. In the fourth section, we apply support vector machines with different kernels to analyze whether abnormally low returns can be predicted by support vector machines and

CONTACT Jan Viebig janviebig@global.t-bird.edu Professor, Goethe University Frankfurt, Auf dem Seif 10, 61462 Königstein, Germany.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hbhf.

Table 1. Key statistics price-to-book ratios and dividend yields for sample 1 "MSCI World" and sample 2 "Developed Markets".

				Explanatory Variable 1: Price-to-Book Ratio								Explanatory Variable 2: Dividend Yields							
IndexName	StartDate	EndDate	No_Obs	MIN	MEDIAN	MAX	MEAN	STD	SKEW	KURT	JBTest	MIN	Median	MAX	MEAN	STD	SKEW	KURT	JBTest
Sample 1: MSCI World											(5% sig.)								(5% sig.)
MSCI World	12/31/1974	8/31/2016	501	1.010	2.099	4.232	2.086	0.663	0.694	3.416	rej. H0	1.27%	2.51%	5.70%	2.90%	1.07%	0.868	2.659	rej. H0
Sample 2: Dev. Mkts.																			
MSCI Australia	8/30/1996	8/31/2016	241	1.443	2.066	3.191	2.174	0.384	0.843	2.936	rej. H0	2.769%	3.836%	7.436%	3.972%	0.807%	1.333	6.100	rej. H0
MSC Austria	8/30/1996	8/31/2016	241	0.636	1.515	2.823	1.497	0.514	0.300	2.070	rej. H0	1.126%	2.265%	7.896%	2.420%	1.020%	1.958	8.824	rej. H0
MSCI Belgium	8/30/1996	8/31/2016	241	0.617	2.040	3.511	1.976	0.533	0.130	3.202	not rej	1.927%	3.019%	10.456%	3.286%	1.212%	3.319	17.834	rej. H0
MSCI Canada	8/30/1996	8/31/2016	241	1.373	2.050	3.290	2.151	0.397	0.739	2.655	rej. H0	0.915%	2.037%	4.074%	2.221%	0.610%	0.424	2.510	rej. H0
MSCI Denmark	8/30/1996	8/31/2016	241	1.376	2.685	5.364	2.845	0.811	0.689	3.139	rej. H0	0.794%	1.627%	2.572%	1.622%	0.383%	−0.065	2.510	not rej.
MSCI Finland	8/30/1996	8/31/2016	241	1.207	2.472	16.000	3.319	2.626	2.871	11.585	rej. H0	0.462%	3.109%	7.634%	3.191%	1.448%	0.367	2.784	rej. H0
MSCI France	8/30/1996	8/31/2016	241	1.002	1.962	4.601	2.058	0.833	1.270	4.359	rej. H0	1.402%	3.098%	6.380%	3.058%	0.877%	0.565	3.640	rej. H0
MSCI Germany	8/30/1996	8/31/2016	241	0.966	1.712	4.458	2.014	0.808	1.234	3.482	rej. H0	1.632%	2.518%	6.319%	2.695%	0.741%	1.424	6.222	rej. H0
MSCI Hong Kong	8/30/1996	8/31/2016	241	0.826	1.452	2.492	1.521	0.340	0.523	2.733	rej. H0	1.950%	2.921%	6.635%	3.100%	0.723%	2.056	8.157	rej. H0
MSCI Ireland	8/30/1996	8/31/2016	241	0.585	2.148	3.696	2.188	0.690	0.125	2.497	not rej.	1.245%	2.280%	9.944%	2.370%	1.204%	3.975	22.394	rej. H0
MSCI Italy	8/30/1996	8/31/2016	241	0.634	1.825	4.311	1.798	0.914	0.731	2.776	rej. H0	1.188%	3.463%	9.078%	3.489%	1.405%	1.229	5.873	rej. H0
MSCI Japan	8/30/1996	8/31/2016	241	0.899	1.547	2.462	1.581	0.405	0.185	1.947	rej. H0	0.589%	1.134%	2.936%	1.411%	0.601%	0.608	2.296	rej. H0
MSCI Netherlands	8/30/1996	8/31/2016	241	1.198	2.206	4.239	2.329	0.733	0.569	2.464	rej. H0	1.726%	2.807%	6.897%	2.943%	0.913%	1.782	7.485	rej. H0
MSCI New Zealand	8/30/1996	8/31/2016	241	1.346	1.908	3.459	2.030	0.512	1.045	3.048	rej. H0	2.903%	4.627%	7.646%	4.728%	0.848%	1.002	4.847	rej. H0
MSCI Norway	8/30/1996	8/31/2016	241	1.000	1.600	2.940	1.744	0.441	0.923	3.067	rej. H0	1.709%	2.878%	6.903%	3.213%	1.119%	0.646	2.752	rej. H0
MSCI Portugal	8/30/1996	8/31/2016	241	0.974	1.942	3.873	2.058	0.632	0.436	2.288	rej. H0	1.618%	3.366%	8.788%	3.655%	1.343%	1.469	5.718	rej. H0
MSCI Singapore	8/30/1996	8/31/2016	241	0.772	1.613	2.558	1.632	0.344	0.323	2.858	not rej.	0.858%	2.503%	6.217%	2.630%	1.009%	0.697	3.741	rej. H0
MSCI Spain	8/30/1996	8/31/2016	241	0.797	2.126	3.606	2.104	0.752	0.190	1.723	rej. H0	1.447%	3.234%	10.645%	3.919%	1.954%	0.835	3.109	rej. H0
MSCI Sweden	8/30/1996	8/31/2016	241	1.243	2.260	6.184	2.571	0.889	1.381	5.054	rej. H0	0.944%	2.674%	6.348%	2.759%	0.989%	0.698	4.036	rej. H0
MSCI Switzerland	8/30/1996	8/31/2016	241	1.595	2.699	5.349	2.937	0.747	1.040	3.424	rej. H0	0.974%	1.851%	4.157%	2.181%	0.807%	0.375	1.717	rej. H0
MSCI United Kingdom	8/30/1996	8/31/2016	241	1.193	2.073	4.180	2.315	0.684	0.981	3.014	rej. H0	2.096%	3.348%	6.005%	3.387%	0.694%	0.731	4.548	rej. H0
MSCI United States	8/30/1996	8/31/2016	241	1.463	2.866	5.812	3.080	0.949	1.139	3.530	rej. H0	1.109%	1.845%	3.343%	1.841%	0.363%	0.478	4.870	rej. H0

Reported are the start date, end date, number of observations, the minimum, medium, maximum, standard deviation, skewness and kurtosis of the price-to-book ratios and the dividend yields of the MSCI World Index and 22 market indices which are classified as developed markets by MSCI. The valuation ratios are calculated by MSCI monthly. The MSCI Israel is not included in Sample 2 as the fundamental time series of this index start after August 30, 1996. The price-to-book ratios and dividend yields are used as explanatory variables in this study. The Jarque Bera Test (JBTest) tests for the null hypothesis H0 that the ratios come from a normal distribution. The result is "rej. H0" if the test rejects the null hypothesis at the 5% significance level (5% sig.).

assess the robustness of the SVM-methodology to changes in model specifications and changes in the cross-validation approach. The fifth section concludes.

Data

We examine the relationship between stock market returns and fundamental data globally. Our data set consists of 23 international equity indices: the MSCI World index and 22 equity market indices classified as "developed markets" by MSCI. The MSCI World index currently includes 1,654 constituents in 23 countries and covers approximately 85% of the global investable market capitalization in each country according to MSCI. We received the total return indices and the fundamental valuation ratios used in this study directly from MSCI. MSCI's total return indices measure the price performance of equity markets with the income from dividends. MSCI calculates total return time series with gross dividends and net dividends. For the MSCI World index, we apply the net total return series in US-Dollars. For the 22 country indices, net total return indices are used in local currency. The MSCI methodology for calculating fundamental data at an index level is described in MSCI (2005, 2011, 2015).

MSCI computes book values per share (BVPS) and dividends per share (DVPS) on an index level and provides monthly price-to-book ratios and dividend yields for each index used in this study. The total return index time series and the dividend yield time series of the MSCI World Index go back to 1969 when MSCI started licensing equity index products. In December 1974, MSCI began reporting BVPS and price-to-book ratios for the MSCI indices. For the MSCI World index, our analysis comprises 501 monthly observations between December 1974 and August 2016.

Our second sample includes 22 equity markets classified as "developed markets" by MSCI. Each time series comprises 241 monthly observations between August 1996 and August 2016. Table 1 documents key statistics for Sample 1 (MSCI World) and Sample 2 (Developed Markets). We perform Jarque/Bera (1987) tests to test the null hypothesis *H0* that the price-to-book ratios and dividend yields come from a normal distribution. We can reject the null hypothesis for most indices at the 95% level of confidence and conclude that the ratios are not normally distributed for most markets. Table 1 documents that price-to-book ratios and dividend yields drift time and again far away from their long-term averages.

Methodology: SVM and probabilistic outputs

Boser, Guyon and Vapnik (1992) and Cortes and Vapnik (1995) introduced support vector machines (SVM) as a new machine learning tool for binary data classification building on earlier work of Vladimir Vapnik in the 1960s. In this paper, we analyze if support vector machines are capable to learn long-term return patterns from fundamental data.

SVM search for the optimal separating hyperplane $f(x) = w \bullet x + b$ with the largest margin between the grouped outcomes. The points which fall on the hyperplane satisfy $w \bullet x_i + b = 0$, where w is the normal vector perpendicular to the hyperplane. The perpendicular distance of the hyperplane to the origin is given by $\frac{|1-b|}{||w||}$, where $||w||$ is the Euclidean norm or magnitude of w. For linearly separable data, the SVM will search for the maximum hyperplane which satisfies the two inequality constraints $H_1: \; w \bullet x_i + b \geq +1$ for $y_i = +1$ and $H_2: \; w \bullet x_i + b \leq -1$ for $y_i = -1$. Introducing a variable γ, the two inequality constraints can be combined into $\gamma_i(w \bullet x + b) - 1 \geq 0$.

As the distance between H_1 and H_2 equals $2/||w||$, the margin can be maximized by minimizing $||w||$. The optimization problem subject to the given inequality constraints H can be solved computationally by minimizing the following Lagrange function with respect to w and b (Burges (1998), Bishop (2006)):

$$L = \frac{1}{2} ||w||^2 - \sum_i \alpha_i \left\{ \gamma_i (w \bullet x + b) - 1 \right\} \rightarrow min \tag{1}$$

where $\alpha_i \geq 0$ are the Lagrange multipliers. Instead of minimizing L, a solution to the optimization problem can be found by solving the dual problem: maximize L subject to the constraints that the derivatives of L with respect to w and b vanish to zero. If we take the derivatives of L and require them to vanish to zero, we get $\omega = \sum_i \alpha_i \; \gamma_i \; x_i$ and $\sum_i \alpha_i \; \gamma_i = 0$.

By substituting the two solutions to the derivatives in equation (1) above, we obtain the Lagrange function L_D in the dual form (Webb, Copsey (2011)):

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \; \gamma_i \gamma_j \; x_i \cdot x_j \longrightarrow max \tag{2}$$

Cortes and Vapnik (1995) showed in their seminal paper that the SVM-approach can be generalized by transforming the input vector in a higher dimensional separating space and argued that any symmetrical function k of two vectors u and v can be formulated as $k(u,v) = \sum_{l=1}^{\infty} \lambda_l \phi_l(u)\ \phi_l(v)$, where λ_l and ϕ_l are eigenvalues and eigenfunctions. As equation (2) depends only on the dot product $x_i \cdot x_j$ of two vectors, a generalized version of support vector training can be obtained by substituting $\gamma_i\ \gamma_j\ x_i \cdot x_j$ in equation (2) by $\gamma_i\ \gamma_j\ k(x_i \cdot x_j)$, where k is a symmetrical kernel function. The fact that support vector training only requires to define a kernel function k without explicitly transforming the vectors is known as "kernel trick" and is one reason why SVM have recently gained increasing attention in the scientific community.

Support vector machines compute decision functions $f(x)$ which can be used to predict the group of new data. Platt (2000) proposed a method to approximate the posterior probability given a set of data x by using a sigmoid function:

$$P(y = +1\ f(x)) = \frac{1}{1 + \exp(A\ f + B)} \tag{3}$$

The two parameters A and B can be estimated using maximum likelihood estimation. In the next section, we apply SVM with two different kernels to estimate the optimal separating hyperplane and apply Platt's sigmoid approach to estimate the probabilities that future returns are abnormally low.

Results

As standard model, we apply SVM-model with a linear kernel $\boldsymbol{k}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{x_i}^T \boldsymbol{x_j}$, where $\boldsymbol{x_i}^T$ denotes the transpose of a vector $\boldsymbol{x_i}$. In addition, we use SVM-models with a Radial Basis Function (RBF) kernel, $k(x_i, x_j) = \exp\left(-\left\|x_i - x_j\right\|\right)^2$, and a kernel scale ς of 1.4 to analyze whether our results are robust to alternative model specifications (Bishop (2006), James et al. (2013)). Before implementing the SVM-models, we standardize our two fundamental predictor variables. We apply 10-fold cross-validation as resampling method before assessing the quality of the SVM-models. Of the ten sets, one partition is retained as the validation set. The other nine partitions are used to train the SVM-models which are evaluated on the validation set. Two criteria, the accuracy rate and the precision rate, are calculated to assess how well the two SVM-models learn if abnormally low returns will occur in the future. We use the lowest $\tau = 25\%$ return percentile as threshold variable τ. Returns which are equal to or larger than the lowest 25% return percentile are classified as positive events $\{+1\}$, returns which are smaller than the threshold variable are classified as negative events {-1} or "abnormally low returns".

In a binary classification problem, four classification outcomes are possible, "true positive (TP)", "true negative (TN)", "false positive (FP)" and "false negative (FN)", depending whether the models predict positive and negative events correctly. Accuracy rates $AR = (TP + TN)/(TP + TN + FP + FN)$ and precision rates $\mathrm{PR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$ are common criteria to evaluate the quality of machine learning algorithms. The accuracy rate measures the number of events predicted correctly relative to all events. The precision rate measures the proportion of positive (negative) events predicted correctly relative to the number of all events predicted as positive (negative) events.

We analyze how well support vector machines predict "abnormally low returns" over three different investment horizons. The returns are calculated over rolling 1-month, 12-month and 60-month periods. Figure 1 reflects the results of a SVM with a linear kernel used to predict abnormally low future 60-month returns of the MSCI World index from current price-to-book ratios and current dividend yields.

Figure 1 shows that SVM are very precise machine learning algorithms to separate abnormally low long-term returns from the complete return sample. Figure 1 panel 1 illustrates that SVM are capable to learn that abnormally low future 60-month returns are clustered when price-to-book ratios are high and dividend yields are low. Figure 1 panel 2 shows that the linear SVM is able to learn that the probability of abnormally low returns increases substantially when markets trade at very high valuation levels.

Figure 2 graphs the 10 posterior probability regimes for the MSCI World shown in Figure 1, Panel 2, between December 1974 and August 2011 over time. The time-series reflects the posterior probability as regimes between 1 and 10 where 1 indicates a low posterior probability and 10 a high posterior probability of abnormally low long-term returns.

We apply Platt's (2000) sigmoid approach to estimate the posterior probability that large negative events {-1} occur in the future 60-month period. Predicting the current market regime is straightforward if the posterior probability regimes are estimated with Platt's (2000) sigmoid approach as each combination of a price-to-book ratio and a dividend yield at time t can be attributed to exactly one posterior probability regime. The graph shows the ten different regimes (which are shown in Figure 1, Panel 2 in different colors) over time given the price-to-book ratio and the dividend yield at time t.

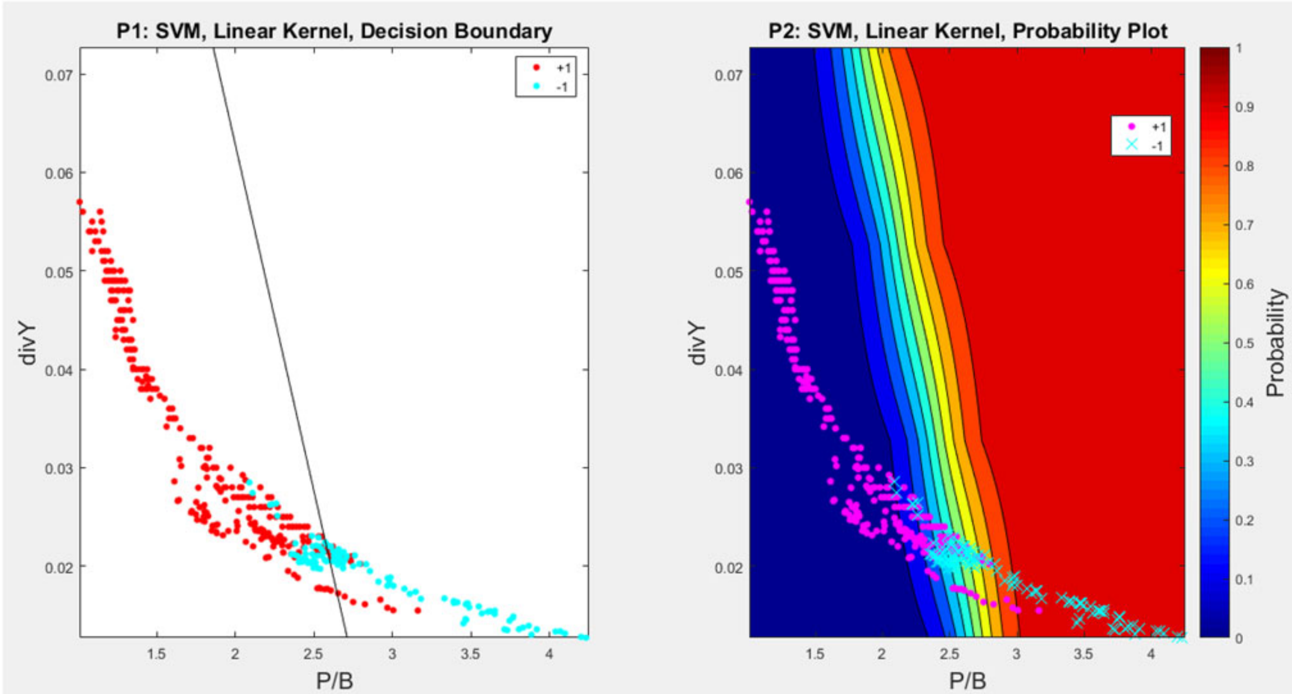

Figure 1. Decision Boundary and Probability Plot of an SVM with a linear kernel to predict abnormally low future 60-month returns from current price-to-book ratios and dividend yields.

Panel 1 (P1): Plotted are the abnormally low 60-month returns $\{-1\}$ of the MSCI World index which are smaller than the lowest 25% return percentile, the 60-month returns exceeding the $\tau = 25\%$ threshold $\{+1\}$ and the decision boundary estimated with a SVM with a linear kernel. Panel 2 (P2) shows the probability plot estimated by a probabilistic SVM with a linear kernel. Both graphs show that abnormally low 60-month returns $\{-1\}$ of the MSCI World index are clustered in the lower right corner of the graph, i.e. when price-to-book ratios are high and dividend yields are low.

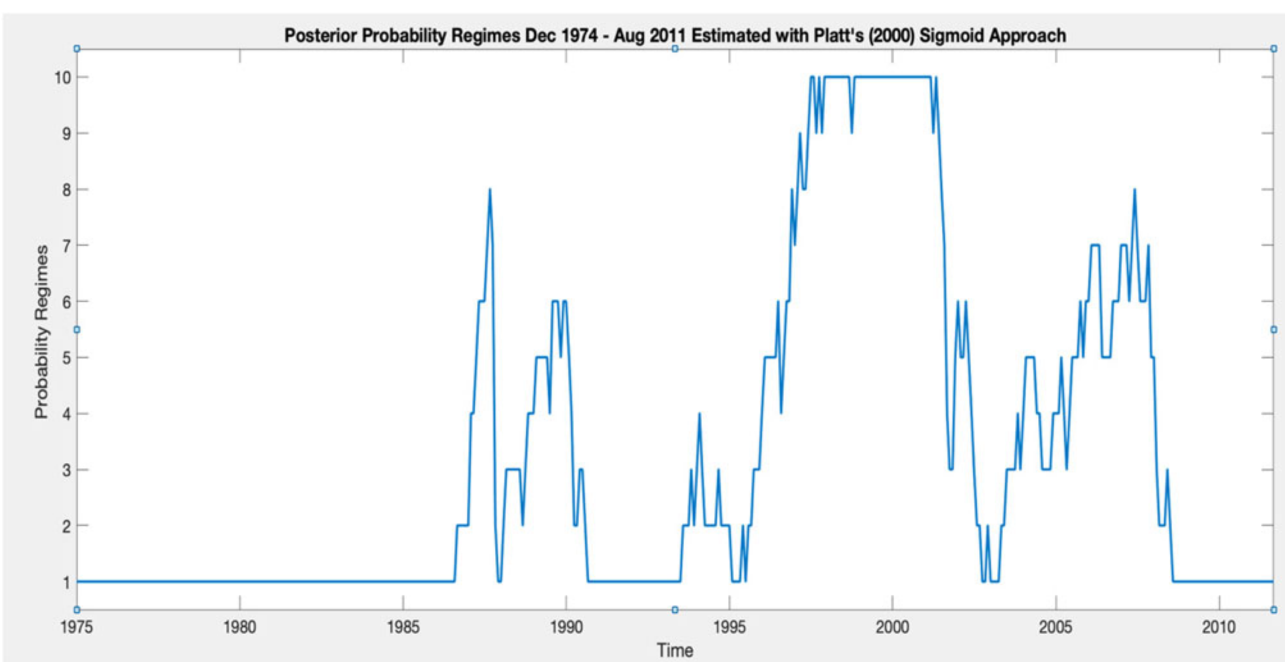

Figure 2. Posterior probability regimes estimated with Platt's Sigmoid Approach between December 1974 and August 2011.

Figure 2 shows that the market remained 226 months in the lowest posterior probability regime (1) over the entire observation period. Table 2 reveals that the ex-post probability of abnormally low long-term returns was zero when the posterior probability was in its lowest regime (1), but 95.35% when the

Table 2. Ex-post probability of negative events of the MSCI World Index in ten different posterior probability regimes estimated with Platt's Sigmoid approach

Posterior Probability Regimes ("Exuberance Levels")	Ex-Post Probability of Large Negative Events Over Following 60-month period	Number of Total Realizations
1	0.00%	226
2	11.11%	36
3	25.81%	31
4	42.86%	21
5	45.45%	33
6	52.17%	23
7	85.71%	14
8	42.86%	7
9	71.43%	7
10	95.35%	43

Reported are the ex-post probabilities of large negative events in the following 60-month periods for the MSCI World index in each of the ten different posterior probability regimes. The posterior probability regimes are estimated with Platt's (2000) sigmoid approach. The table shows that a strong relationship exists between posterior probability regimes and the probability of large negative ex-post events.

posterior probability was in its highest regime (10). We conclude that SVM are useful tools to detect irrational exuberance in financial markets followed by abnormally low long-term returns. Abnormally low long-term returns occur much more often when equity markets trade at extremely high valuation levels than in other periods.

The posterior probability regime reached an "exuberance level" of 6 or higher in April 1987 to September 1987, July 1989 to September 1989, November 1989 to December 1989, June 1996, September 1996 to July 2001, December 2001, March 2002, September 2005, November 2005 to April 2006 and in September 2006 to October 2007. These periods were regularly followed by large market corrections as shown in Table 2 reporting the ex-post probabilities of large negative 60-month events for the MSCI World in each of the 10 different posterior probability regimes estimated with Platt's (2000) sigmoid approach. We conclude that a strong relationship exists between posterior probability regimes which represent the degree of exuberance in financial markets and abnormally low long-term returns in the future.

Combining SVM and Platt's (2000) sigmoid approach is an efficient and mathematically elegant way to help financial researchers and investors to analyze the degree and the length of exuberance in financial markets. Table 3 reports the accuracy rates and precision rates of the two SVM-models for the MSCI World index and 22 developed markets over three different investment horizons. As discussed above, we apply 10-fold cross-validation before computing

Table 3. Assessment of two different SVM-models to predict abnormally low returns of the MSCI World Index (Sample 1) and 22 developed markets (Sample 2) over different investment horizons

Sample 1: MSCI World	1-month		12-month		60-month	
MSCI World	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	87.3%	**62.7%**
SVM, RBF-kernel, 1.4	75.0%	**0.0%**	77.1%	**8.2%**	87.8%	**71.8%**
Sample2: Dev. Mkts.	**1-month**		**12-month**		**60-month**	
MSCI Australia	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	90.6%	**68.9%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	77.7%	**14.0%**	91.2%	**71.1%**
MSCI Austria	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	85.6%	**55.6%**
SVM, RDF-Kernel, 1.4	75.8%	**5.0%**	75.1%	**0.0%**	85.1%	**57.8%**
MSCI Belgium	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	75.1%	**0.0%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	74.2%	**5.3%**	75.1%	**0.0%**
MSCI Canada	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	85.1%	**66.7%**
SVM, RDF-Kernel, 1.4	74.6%	**0.0%**	83.0%	**31.6%**	89.5%	**64.4%**
MSCI Denmark	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.5%	**14.0%**	75.7%	**22.2%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	82.1%	**56.1%**	86.7%	**71.1%**
MSCI Finland	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	79.5%	**22.8%**	84.0%	**35.6%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	79.0%	**21.1%**	84.0%	**35.6%**
MSCI France	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	76.9%	**14.0%**	80.1%	**31.1%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	79.5%	**21.1%**	80.7%	**31.1%**
MSCI Germany	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	87.8%	**82.2%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	75.5%	**5.3%**	90.1%	**80.0%**
MSCI Hong Kong	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	83.0%	**49.1%**	86.7%	**96.3%**
SVM, RDF-Kernel, 1.4	74.2%	**0.0%**	82.1%	**42.1%**	92.3%	**98.5%**
MSCI Ireland	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	75.1%	**0.0%**
SVM, RDF-Kernel, 1.4	75.0%	**5.0%**	76.0%	**8.8%**	92.3%	**71.1%**
MSCI Italy	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	75.1%	**0.0%**
SVM, RDF-Kernel, 1.4	73.8%	**0.0%**	75.1%	**0.0%**	88.4%	**80.0%**
MSCI Japan	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	84.5%	**62.2%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	79.5%	**21.1%**	90.1%	**60.0%**
MSCI Netherlands	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	87.3%	**75.6%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	75.1%	**0.0%**	88.4%	**73.3%**
MSCI New Zealand	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	85.1%	**66.7%**
SVM, RDF-Kernel, 1.4	74.6%	**0.0%**	78.2%	**12.3%**	85.6%	**64.4%**
MSCI Norway	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	90.6%	**77.8%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	74.2%	**8.8%**	93.9%	**88.9%**
MSCI Portugal	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	75.1%	**0.0%**
SVM, RDF-Kernel, 1.4	74.6%	**0.0%**	76.4%	**7.0%**	84.5%	**48.9%**
MSCI Singapore	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.5%	**26.3%**	79.6%	**89.0%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	78.6%	**52.6%**	81.2%	**92.6%**
MSCI Spain	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	80.7%	**46.7%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	75.1%	**0.0%**	80.1%	**46.7%**
MSCI Sweden	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	88.4%	**77.8%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	77.7%	**14.0%**	87.8%	**75.6%**
MSCI Switzerland	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	75.1%	**0.0%**	75.1%	**0.0%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	75.1%	**8.8%**	87.3%	**66.7%**
MSCI United Kingdom	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	82.1%	**36.8%**	95.6%	**88.9%**
SVM, RDF-Kernel, 1.4	74.2%	**0.0%**	84.7%	**40.4%**	95.0%	**82.2%**

(*Continued*)

Table 3. Continued.

Sample2: Dev. Mkts.	1-month		12-month		60-month	
MSCI United States	**Accuracy**	**PR**	**Accuracy**	**PR**	**Accuracy**	**PR**
SVM, Linear Kernel	75.0%	**0.0%**	77.7%	**19.3%**	89.5%	**66.7%**
SVM, RDF-Kernel, 1.4	75.0%	**0.0%**	86.5%	**50.9%**	89.5%	**66.7%**

Reported are the precision rates PR and the accuracy rates of two different SVM-models. The SVM-models are used to predict abnormally low returns over different investment horizons (1-month, 12-month, 60-month). Abnormally low returns are defined as returns which are smaller than the lowest 25% return percentile over a given investment horizon. The precision rates and accuracy rates are calculated after a 10-fold cross-validation to partition the data randomly in ten sets. The models are trained on the training sets and evaluated on the cross-validation sets. The precision rate is the proportion of correct negative classifications (TN) relative to the number of all events predicted as negative (TN + FN). The accuracy rates measure the percentage of events predicted correctly by the SVM.

accuracy rates and precision rates which are used to assess the quality of the two SVM-models. The classification involves that 75% of the events are positive events and 25% are negative events. If SVM-classifiers are not able to separate data consisting of 75% positive events and 25% negative events, the models predict that all events are positive which leads to an accuracy rate of 75% and an undefined precision rate ("0.0").

The results presented in Table 3 show that SVM are capable to learn that abnormally low 60-month returns are clustered. For the MSCI World, the precision rates are as high as 62.7% and 71.8% for the two SVM-models used in this study. For the 22 country indices, we find that the precision rates of abnormally low long-term returns, which occur by definition with a chance of only 25%, are on average above 60%. For one country, the precision rate of abnormally low 60-month returns is as high as 98.5%. For most markets, the precision rates are higher if SVM with RBF-kernels are applied instead of SVM with linear kernels. We conclude that applying kernels which are able to detect nonlinear patterns in return data may lead to (slightly) higher precision rates.

Table 3 documents that the precision rates increase on average from less than 1% to over 60% if the investment horizon increases from 1 month to 60 months. For longer return periods, SVM are able to learn return patterns with high accuracy rates and high precision rates as shown in Table 3.

We conducted several tests to assess if our methodology is robust to changes in its underlying assumptions. Above we applied 10-fold cross validation (KFOLD = 10) as resampling method before assessing the quality of the SVM-models. In addition to our standard approach (KFOLD = 10), we conducted cross-validations with KFOLD = 5 and KFOLD = 3. Table 4 shows that the results are highly robust to changes in the resampling method. Table 4 also reports additional measures to assess the quality of the SVM-models. The F1-measure $F1 = 2\ PR\ RR/(PR + RR)$ is a harmonic mean combining the precision rate PR and the recall rate RR to a single number. We also measure the accuracy of the model by the area under the ROC curve (AUC).

Table 4. Robustness of the SVM-methodology to Changes in Model Specifications and Changes in the Cross-validation Approach

Panel 1: KFOLD 10	Accuracy	F1	AUC
SVM, Linear Kernel	0.873	0.919	0.944
SVM, RBF-kernel, 0.4	0.923	0.948	0.938
SVM, RBF-kernel, 1.4	0.878	0.919	0.923
SVM, Polynomial Kernel	0.880	0.921	0.950
Panel 2: KFOLD 5	**Accuracy**	**F1**	**AUC**
SVM, Linear Kernel	0.878	0.922	0.946
SVM, RBF-kernel, 0.4	0.921	0.946	0.945
SVM, RBF-kernel, 1.4	0.873	0.917	0.925
SVM, Polynomial Kernel	0.882	0.922	0.951
Panel 3: KFOLD 3	**Accuracy**	**F1**	**AUC**
SVM, Linear Kernel	0.859	0.909	0.938
SVM, RBF-kernel, 0.4	0.907	0.937	0.938
SVM, RBF-kernel, 1.4	0.866	0.912	0.914
SVM, Polynomial Kernel	0.866	0.911	0.941

Reported are three different measures (accuracy rates, F1-measures, AUC) to assess the quality of four different SVM models with different kernels and kernel scales. The table reports the three quality measures for three different cross-validations (KFOLD = 3, 5, 10), separately. The table shows that the SVM-methodology is robust to changes in model specifications (four different models) and changes in the cross-validation method (three different resampling methods) for all three quality measures (accuracy rates, F1-measures, AUC).

Table 4 shows that the accuracy rates, the F1-measures and the AUC remain consistently high when the resampling method is altered. In addition to the SVM with a linear kernel and the SVM with a Radial Basis Function (RBF) kernel and a kernel scale of 1.4 used above, we introduce two additional SVM-models, a SVM with a RBF and a kernel scale of 0.4 and a SVM with a polynomial kernel $\boldsymbol{k}(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{x_j} + 1)\boldsymbol{d}$ of the order of $d = 3$, to analyze whether our results are robust to alternative model specifications. Table 4 reflects that the SVM-approach is robust to changes in model specifications and KFOLD-assumptions for all three quality measures (accuracy, F1-score, AUC). We conclude that the SVM-methodology presented in this article is a robust approach to analyze the length and the degree of exuberance in financial markets.

Combining SVM with Platt's sigmoid approach can help investors to predict the probability of abnormally low long-term returns with high accuracy rates, high precision rates, high F1-measures and high AUC. We conclude that SVM are useful tools to learn long-term return patterns from datasets.

Conclusions

Overly optimistic earnings and risk expectations drive equity markets time and again far away from levels justified by fundamentals. The probability that investors generate abnormally low returns over long investment horizons is disproportionately high when equity markets trade at extremely high valuation levels. Machine learning tools are designed to learn "rules" from large datasets. Support vector machines are capable to learn from datasets that abnormally low long-term returns occur much more frequently when equity markets trade at extremely high valuation levels with high precision rates and high accuracy rates.

References

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. New York: Springer Science + Business Media LLC.

Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. 1992. *A Training Algorithm for Optimal Margin Classifiers*, COLT '92 Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 144–152.

Burges, Christopher J. C. 1998. "A Tutorial on Support Vector Machines for Pattern Recognition." *Data Mining and Knowledge Discover* 2, no. 2: 121–167.

Campbell, John Y., and Robert J. Shiller. 1987. "Cointegration and Tests of Present Value Models," *Journal of Political Economy* 95, no. 5: 1062–1088.

Campbell, John Y., and Robert J. Shiller. 1988. "Stock Prices, Earnings, and Expected Dividends." *The Journal of Finance* 43, no. 3: 661–676.

Campbell, John Y. and Robert J. Shiller. 1989. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *The Review of Financial Studies* 1, no. 3: 195–228.

Campbell, John Y., and Robert J. Shiller. 1998. "Valuation Ratios and the Long-Run Stock Market Outlook." *The Journal of Portfolio Management* 24, no. 2: 11–26.

Campbell, John Y., and Robert J. Shiller. 2001. "Ratios and the Long-Run Stock Market Outlook: An Update." NBER Working Paper 8221.

Cortes, Corinna, and Vladimir Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20: 273–297.

De Bondt, Werner, Gulnur Muradoglu, Hersh Shefrin, and Sotiris K. Staikouras. 2008. "Behavioral Finance: Quo Vadis?" *Journal of Applied Finance* 18, no 2: 7–21.

Hirshleifer, David. 2015. "Behavioral Finance." *Annual Review of Financial Economics* 7: 133–159.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer Science + Business Media.

Jarque, Carlos M., and Anil K. Bera. 1987. "A Test for Normality of Observations and Regression Residuals." *International Statistical Review* 55, no. 2: 163–172.

Kent, Daniel, David Hirshleifer, and Siew Hong Teoh. 2002. "Investor Psychology in Capital Markets: Evidence and Policy Implications." *Journal of Monetary Economics* 49, no. 1: 139–209.

MacKay, Charles. 1995. *Extraordinary Popular Delusions and the Madness of the Crowds*. First Published 1841. Ware: Wordsworth Reference. (Edition Published 1995).

MSCI. 2005. *MSCI Methodology for Calculating Fundamental Data Ratios at an Index Level* July, 2005.

MSCI. 2011. MSCI Fundamental Data Methodology, May 2011.

MSCI. 2015. MSCI Fundamental Data Methodology, January 2015.

Platt, John C. 2000. "Probabilities for SV Machines." In *Advances in Large Margin Classifiers*, edited by Alexander J. Smola, Peter Bartlett, Bernhard Schölkopf, Dale Schuurmans, 61–74. Cambridge: The MIT Press.

Shiller, Robert. 1981a. "The Use of Volatility Measures in Assessing Market Efficiency." *The Journal of Finance* 36, no. 2: 291–301.

Shiller, Robert. 1981b. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71, no. 3: 421–436.

Shiller, Robert. 1988. "The Volatility Debate." *American Journal of Agricultural Economics* 70, no. 5: 1057–1063.

Shiller, Robert. 1989. "Comovements in Stock Prices and Comovements in Dividends." *The Journal of Finance* 44, no. 3: 719–729.

Shiller, Robert. 2003. "From Efficient Markets Theory to Behavioral Finance." *The Journal of Economic Perspectives* 17, no. 1: 83–104.

Shleifer, Andrei, and Robert W. Vishny. 1997. "The Limits of Arbitrage." *The Journal of Finance* 52, no. 1: 35–55.

Subrahmanyam, Avanidhar. 2007. "Behavioural Finance: A Review and Synthesis." *European Financial Management* 14, no. 1: 12–29.

Thaler, Richard H. 1999. "The End of Behavioral Finance." *Financial Analysts Journal* 55, no. 6: 12–17.

Viebig, Jan. 2015. "Are Emerging Market Investors Overly Pessimistic in Extreme Risk-Off Periods?" *The Journal of Behavioral Finance* 16, no. 2: 163–172.

Webb, Andrew R., and Keith D. Copsey. 2011. *Statistical Pattern Recognition*, Third Edition, Chichester: John Wiley & Sons Ltd.