

Equity Style Timing using Support Vector Regressions ^{*}

G. Nalbantov¹, R. Bauer², and I.G. Sprinkhuizen-Kuyper³

¹ ERIM, Erasmus University Rotterdam,
Postbus 1738, 3000 DR Rotterdam, The Netherlands
`nalbantov@few.eur.nl`

² LIFE, Universiteit Maastricht and ABP Investments
`r.bauer@berfin.nl`

³ IKAT, Universiteit Maastricht
`kuyper@cs.unimaas.nl`

Abstract. The disappointing performance of value and small cap strategies shows that style consistency may not provide the long-term benefits often assumed in the literature. In this study we examine whether the short-term variation in the U.S. size and value premium is predictable. We document style-timing strategies based on technical and (macro-)economic predictors using a recently developed artificial intelligence tool called Support Vector Regressions (SVR). SVR are known for their ability to tackle the standard problem of overfitting, especially in multivariate settings. Our findings indicate that both premiums are predictable under fair levels of transaction costs and various forecasting horizons.

1 Introduction

There is no doubt about the importance of investment styles in modern portfolio management. The underlying rationale for this relates to a series of influential studies documenting the potential benefits of investing in stocks with fundamental commonalities or “styles”. In the past two decades, substantial evidence surfaced suggesting that investing in portfolios of stocks with a small market capitalization and value orientation provides a premium in the long run. The “size premium” has been first reported by Banz [5], who found a negative relation between a firm’s market capitalization and its stock performance in the U.S. The extensively researched “value premium” has been documented most prominently by Fama and French [19, 21] and Lakonishok et al. [27]. These studies showed that stocks with typical value features such as low market-to-book (M/B), low price-to-earnings (P/E) and low price-to-cash (P/C) ratios provided

^{*} We would like to thank ABP Investments for providing the data, Evgueni Smirnov (IKAT), Roderick Molenaar (ABP Investments) and Borislav Pavlov (LIFE) for the many fruitful discussions, and Evgueni Smirnov also for the many hours of using his computer.

higher average returns than so-called “growth” stocks, with high M/B, P/E and P/C ratios. These empirical findings induced a discussion on the source and magnitude of the value and size premium. Some studies argued that this premium is a compensation for holding stocks under relative distress, see for instance [12, 20]. Another view, put forward in [27, 23], is that stock markets lack efficient pricing ability. A third possible explanation suggested in [30] is that the obtained results are due to data snooping biases. A recent review and update for the U.S. by Chan and Lakonishok [13] shows that value investing still generates promising returns in the long run. Dimson et al. [18] make similar statements for the U.K. value premium.

The rather disappointing performance of small cap and value strategies during the nineties has however pointed out that style consistency may not necessarily provide superior returns in any economic regime. A relatively small body of literature has explicitly addressed the potential benefits of style timing strategies over a style consistent approach. Although most of these papers may differ in methodology, they all rely on the notion that the cyclical behavior of investment styles is correlated with systematic economical and technical forces, which could make the value and size premium partially predictable. Cooper et al. [16] find sufficient predictability for size-sorted strategies in the U.S., but weaker results for value-sorted strategies.¹ Levis and Liodakis [28] find moderate evidence in favor of small/large rotation strategies, but less evidence for value/growth rotation in the United Kingdom. Bauer et al. [6] find evidence for the profitability of style rotation strategies in Japan, but point out that moderate levels of transaction costs can already make these results less interesting in a practical context. The majority of rotation studies employ technical (or market-based) and (macro-)economic indicators. The dependent variables, either the value or the size premium, are constructed using well-known style index series.

In this study we will use a similar approach by constructing the value and size premium in the U.S. based on S&P style indices. The sign and magnitude of both premiums will then subsequently be forecasted using a broad set of (macro-)economic and technical predictors. In contrast to the studies mentioned above, we will not apply a standard multifactor model framework. Factor models in general suffer from deficiencies intrinsic to multiple regression techniques. Most of the studies based on this methodology *ex ante* decide to construct parsimonious models to avoid the problem of overfitting. Increasing the number of factors at some point will deteriorate the out-of-sample prediction ability of the rotation models. Levis and Liodakis [28] for instance report empirical results based on six factors for the size spread and eight factors for the value spread. Although their regression window is expanding, thereby updating the relevance of the factors through time, it does not provide the ability to add or delete economically viable factors. In most cases the “optimal” choice of independent variables is based on a set of statistical criteria, like adjusted R^2 , the Akaike information criterion or the Schwarz criterion. These criteria are designed to correct the inclusion of fac-

¹ Other related work includes [1–4, 17, 24, 26, 31].

tors for the increased model complexity. Potentially, numerous relevant variables are bound to be excluded as predictors.

A further complication arises from the fact that individual factors in a model are usually assumed to be independent. Most linear regression models however are likely to suffer from multi-collinearity, especially when forecasting variables are numerous and closely related. One could therefore argue that factor models face two pivotal challenges: first, how to employ a large set of potentially relevant variables in a factor model without jeopardizing its predictive power, and second, how to incorporate possible interactions between individual variables in the course of the model-building process without deteriorating the quality of the model.²

Support Vector Regressions (SVR) have become a popular analytical tool following a series of successful applications in fields ranging from optical character recognition to DNA analysis [35, 41]. In essence, the SVR technique is used for function estimation based on an observed finite number of input-output relations, just like the linear multiple-regression technique. Numerous potential applications of SVR in finance have been reported elsewhere.³ The combination of three key features can justify a priori the utilization of the SVR tool in financial forecasting modeling. First, SVR behave robustly even in high-dimensional feature problems [32], or in other words, where the explanatory variables are numerous, and in noisy, complex domains [10]. Second, SVR achieve remarkable generalization ability by striking a balance between a certain level of model accuracy on a given training data-set, and model complexity.⁴ And third, SVR always find a global solution to a given problem [40, 44], in sharp contrast with neural networks for instance. A general limitation of SVR is that they produce point estimates rather than posterior probability distributions of the obtained results, which follows from the fact that SVR are a nonparametric tool. Some parameters however have to be estimated in advance via a standard procedure called “cross-validation”. This procedure, though quite computationally extensive, additionally ensures that model selection is based on out-of-sample rather than in-sample performance.

Using SVR we will construct models in order to predict the value and size premium in the U.S. stock market. Our aim is to test on a preliminary level the performance of SVR, and not to engage in an extensive data-mining exercise. For that reason, we build our models on historical data of 60 months, which is a quite common horizon in the literature. Obviously, other model-building horizons can be explored, but in such a way artificially good results could emerge, falling pray to the data-mining critique. We compare the results of the rotation strategies with so-called style consistent passive strategies. Furthermore, we vary the forecast horizon (one-, three- and six-month signals), which serves as a model-stability test, and measure the impact of a wide range of transaction costs. The

² See [9] and [37] for a discussion on these and related issues.

³ See, e.g. [34, 36, 38, 41, 43].

⁴ Note that in real-world applications the presence of noise in regression estimation necessitates the search for such a balance, see [44, 46].

empirical section shows that style rotation strategies using signals created by SVR produce outstanding results for both the value and the size premium.

The remainder of the paper is organized as follows. Section 2 discusses the choice of explanatory variables and the nature of the explained variables (the proxies for the value and size premium). Section 3 deals extensively with Support Vector Regressions as an analytical tool and how it can be used to predict the value and size premiums. Section 4 presents our main empirical findings, and finally section 5 concludes.

2 Data

2.1 Construction of the value and size premium series

The choice of an appropriate measure to determine the value premium is crucial. Our main goal is to come up with a trading strategy, which can be easily implemented in a practical context⁵. In principle, long time series data from the Center for Research in Security Prices (CRSP) can be used. Following this venue is not well suited for a low transaction cost strategy however, since there are no readily-available instruments (e.g. futures) to exercise such a trading strategy in practice. As we expect the rotation strategies to have a considerable turnover, we conduct our analysis on the S&P Barra Value and Growth indices (the value premium). Transaction costs are expected to be relatively low as we are able to buy and sell futures on these indices.⁶ Both indices are constructed by dividing the stocks in the S&P 500 index according to just one single attribute: the book-to-market ratio. This procedure splits the S&P 500 index into two, mutually exclusive groups of stocks and is designed to track these accepted investment styles in the U.S. stock market. The Value index contains firms with high book-to-market ratios and conversely the Growth index firms with lower book-to-market ratios. The combination of both (market cap weighted) indices adds up to the (market cap weighted) S&P 500.

Figure 1 and table 1 show that a strategy purely based on the value premium would have witnessed some highly volatile periods. These series are the returns of a long position in the Value index and a short position in the Growth index throughout the entire sample period ranging from January 1988 to December 2002. Monthly maximum and minimum returns of this strategy are considerably high: 9.74% and -12.02%. Summary statistics (see table 1) reveal that the spread series exhibits excess kurtosis. The number of negative performance months of this passive value strategy is approximately 47%. The average return on an annualized basis is -0.86% with a standard deviation of 9.64%. We therefore

⁵ In the case of for instance the High book-to-market minus Low book-to-market (HML) series of [20], we can expect relatively high transaction costs as portfolios generally exhibit unacceptable liquidity features, particularly in a monthly long/short setting.

⁶ In practice the maximum exposure of the trading strategy is still restricted by the liquidity features of this future.

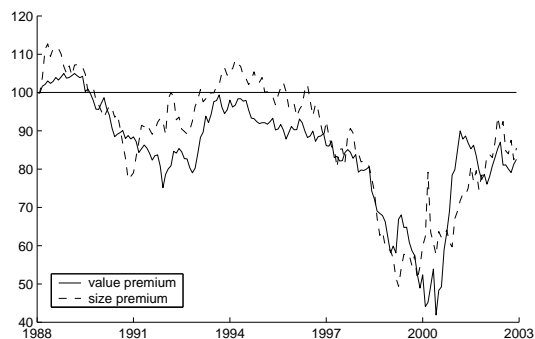


Fig. 1. Cumulative performance of the value and size premiums (1988:01–2002:12).

Table 1. Summary statistics for the value and size premiums (1988:01 - 2002:12). All numbers are annual data (in %) unless stated otherwise. The spread series for the value premium are computed as returns of a long/short portfolio (long S&P Barra Value index and short S&P Barra Growth index). The spread series for the Size premium are computed as returns of a long/short portfolio (long S&P 500 index and short S&P SmallCap 600 index). Prior to the introduction of the S&P SmallCap 600 index in January 1994, the Frank Russell 1000 and Frank Russell 2000 indices have been used as inputs for the small-large calculations.

	Value premium	Size premium
Mean	-0.86	-0.91
Standard deviation	9.64	12.04
Information ratio	-0.09	-0.08
Minimum (monthly)	-12.02	-15.71
Maximum (monthly)	9.74	16.78
Skewness (monthly)	0.06	0.27
Excess kurtosis (monthly)	3.15	4.31
% negative months	47.22	51.11

conclude that pure and unconditional value investing in this particular sample period has not been a very attractive trading strategy. Furthermore, we indeed observe a cyclical pattern in the behavior of the premium. In some periods, like for instance in the last years of the previous decade, growth stocks persistently outperformed value stocks and in other periods value stocks clearly outperformed growth stocks. A good example of the latter is the crisis in Technology (and hence “growth”) stocks in the beginning of this century. A possible explanation for this phenomenon could be that the sign of the value premium is strongly connected with the business cycle and the economic regime. It is likely that value stocks — relative to growth stocks — gain more from a surge of economic activity and a sharp upward revision of sentiment, see e.g. [39]. As profit expectations turn sharply and broadly positive at the bottom of the economic cycle, profitability and earnings growth become a less scarce resource. In such an environment port-

folio managers start looking for stocks with typical value features. This largely explains why value stocks generally belong to cyclical industries. Moreover, value companies tend to belong to mature sectors of the economy. These sectors generally grow and shrink with the economy, whereas growth companies can offer protection during weaker periods in the economy.

Analogously, the size premium series is created by comparing the S&P 500 index (large cap) and the S&P Small Cap 600 index.⁷ The passive small-large strategy has not performed satisfactorily during our sample period as well: a mean return of 0.91% (see figure 1 and table 1). Investors that have followed this strategy have experienced even greater fluctuations than those opting for the passive value-growth strategy, as revealed by the maximum (16.78%) and minimum (-15.71%) monthly returns and the higher standard deviation (12.04%). All of these findings cast serious doubt on the wisdom of persistently favoring small stocks over large stocks in the past two decades.

2.2 Choice of the forecasting variables

We will introduce two classes of forecasting variables in this section. First, we give a brief overview of potential technical variables. Subsequently, we will address several macro-economic variables, which might shed some light on the behavior of the spread series. There appears to be a striking similarity between the chronological cumulative performance of the value and size premiums (see figure 1), which suggests that the behavior of both premiums might be subject to the same cyclical effects. We aim to provide a wide range of relevant forecasting variables, but we restrict ourselves to those claimed to be economically interpretable in the literature on this subject.

Good examples of technical factors are the lagged value and small cap spreads used by Levis and Liodakis[28]. Asness et al. [4] propose two other variables of this class: the spread in valuation multiples and expected earnings growth between value portfolios and growth portfolios. Other candidates are changes in the implied volatility of the market, see [17], and price and earnings momentum in the market, see for instance [33] and [8].

The class of economic variables is mainly related to economic fundamentals, the business cycle and trends in corporate earnings. Examples of macro-economic series can be found in a variety of papers on style rotation. Kao and Humaker [26] document the influence of industrial production, the yield-curve spread, inflation (CPI) and the corporate credit spread on the value premium. In their view, industrial production reflects the corporate earnings cycle. In periods of high corporate earnings growth, the often highly leveraged value (and small) companies profit disproportionately. The composite leading indicators (CLI) can serve as an alternative to measure the same relationship. The interest rate environment can also have a substantial impact on the sign of the value premium. A yield

⁷ Prior to the introduction of the S&P Small Cap 600 index in January 1994, the Frank Russell 1000 and Frank Russell 2000 indices have been used as input for the small-large calculations.

spread widening between long government bonds and short term T-bills will probably hurt growth companies more than value companies as their profits are based further into the future. Growth stocks have longer durations than value stocks and are therefore more interest rate sensitive. These companies will underperform most likely in a setting with steep yield curves, which implies rising interest rates in the future. In the study of [28] the spread series are explained by the level of inflation, changes in the short-term interest rate and the equity risk premium respectively.⁸

In table 2 we list the variables actually used in our empirical analysis. In the next section we describe and discuss the nonparametric modeling tool used: Support Vector Regressions.

Table 2. Variables used in the style timing models based on Support Vector Regressions.

<i>Technical Variables</i>	
LagVmG	Lagged Value/Growth spread
LagSmL	Lagged Small/Large spread
VOL	Volatility of the S&P 500
FPE	12-month Forward P/E of the S&P 500
MOM	6-month Momentum of the S&P 500
Profit cycle	Year on Year change in earnings per share of the S&P 500
PE dif	Price/Earnings difference between Value and Growth indices, or between S&P 500 and Small cap indices
DY dif	Difference between dividend yields on Value and Growth indices, or S&P 500 and Small cap indices
<i>Economic Variables</i>	
Corporate Credit Spread	The yield spread of (Lehman Aggregate) Baa over Aaa
Core inflation	The 12-month trailing change in the U.S. Consumer Price Index
Earnings-yield gap	The difference between the forward E/P ratio of the S&P 500 and the 10-year T-bond yield
Yield Curve Spread	The yield spread of 10-year T-bonds over 3-month T-bills
Real Bond Yield	The 10-year T-bond yield adjusted for the 12-month trailing inflation rate
Ind. Prod	U.S. Industrial Production Seasonally Adjusted
Oil Price	The 1-month price change
ISM (MoM)	1-month change of U.S. I.S.M. Purchasing Managers Index (Mfg Survey)
Leading Indicator	The 12-month change in the Conference Board Leading Indicator

⁸ Liew and Vassalou [29] claim that past style performance can actually function as a *forecast* for economic growth, which brings a new dimension to this literature.

3 Methodology

This section describes the model-building tool (Support Vector Regression) and the construction of the SVR style rotation models. Alongside, we focus on the qualities of SVR that justify their employment as a factor model tool.

3.1 Function estimation with SVR

Support Vector Regressions (SVR), and Support Vector Machines (SVM) in general, are rooted in Statistical Learning Theory, pioneered by Vapnik [44]. In essence, SVR are just functions, named “learning machines”, of which the basic task is to “explore” data (input-output pairs⁹) and provide optimally accurate predictions on unseen data. Extensive descriptions of SVR and SVM can be found, for example, in [11, 40, 41]. Here we present a complete, but still compact and accessible representation of the basic SVR tool. The technical exposition follows mostly the descriptions in the abovementioned papers.

First, it should be mentioned that the standard loss function employed in SVR is the ϵ -insensitive loss function, which has the following form:

$$|y - f(\mathbf{x})|_\epsilon \equiv \max\{0, |y - f(\mathbf{x})| - \epsilon\} \quad (1)$$

Here ϵ is predetermined and nonnegative, y is the true target value, \mathbf{x} is a vector of input variables and $f(\mathbf{x})$ is the estimated target value. If the value of the estimate $f(\mathbf{x})$ of y is off-target by ϵ or less, then there is no “loss”, and no penalty will be imposed. However, if $|y - f(\mathbf{x})| - \epsilon > 0$, then the value of the loss function rises linearly with the difference between y and $f(\mathbf{x})$ above ϵ . In practice, the actual loss associated with a given training error is equal to $C(|y - f(\mathbf{x})|_\epsilon)$, where C is a nonnegative constant. The term $|y - f(\mathbf{x})|_\epsilon$ is denoted by ξ if $y \leq f(\mathbf{x}) - \epsilon$, and by ξ^* if $y \geq f(\mathbf{x} + \epsilon)$.

Let us consider the simplest case of function estimation first, where there is only one input variable, x_1 , one output variable, y , and l training data points, and a linear relationship between the input and output variables (see figure 2).

Notice that in the case of figure 2, the total amount of loss is equal to $C(\xi + \xi^*)$, since there are two training errors. The SVR algorithm estimates the parameters w_1 and b of the linear function $y = w_1x_1 + b$ for prespecified values of ϵ and C , ensuring that the resulting regression function achieves good generalization ability. It should not be too “complex”, but, at the same time, it should not make too many training errors. Complexity here is defined in terms of “flatness” of the line, i.e. the smaller the slope of the line, the lower the complexity. By striking a balance between the function’s complexity and accuracy on the training data in the model-construction phase, the SVR offers a solution to the common problem of overfitting.

⁹ The terms “inputs” and “outputs” in the machine learning domain stand for the “independent variables” and the “dependent variables” in the finance domain.

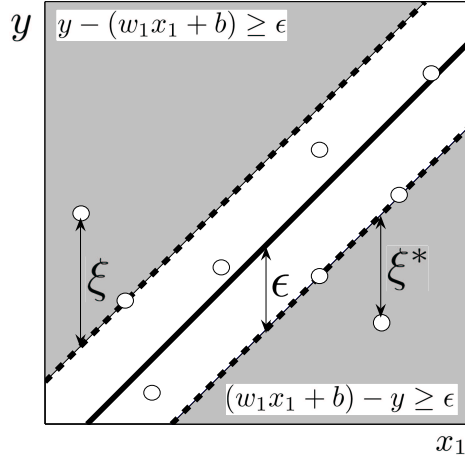


Fig. 2. An SVR solution to the problem of estimating a relation between x_1 and y . All points inside the white region in the figure are within ϵ distance from the solid, optimal regression line $y = w_1x_1 + b$, and therefore are not penalized. However, penalties ξ_i and ξ_i^* are assigned to the two points that lie inside the shaded areas (given by $y - (w_1x_1 + b) \geq \epsilon$ and $(w_1x_1 + b) - y \geq \epsilon$). The optimal regression line is as flat as possible, and strikes a balance between the area of the white region and the amount of points that lie outside this region.

Figure 2 considers a one-dimensional input space, i.e. there is only one independent variable. If the dimension of the input space equals n , we are looking for the optimal regression function $f(\mathbf{x}) = (\mathbf{w} \cdot \mathbf{x}) + b$, with a vector of input variables $\mathbf{x} = (x_1, x_2, \dots, x_n)$, “weight” vector $\mathbf{w} = (w_1, w_2, \dots, w_n)$, and the inner product $(\mathbf{w} \cdot \mathbf{x}) = w_1x_1 + w_2x_2 + \dots + w_nx_n$. Flatness in that case is defined in terms of the Euclidean norm of the weight vector: $\|\mathbf{w}\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$. The parameters of the linear SVR $f(\mathbf{x}) = (\mathbf{w} \cdot \mathbf{x}) + b$, i.e. \mathbf{w} , b , ξ_i and ξ_i^* , $i = 1, 2, \dots, l$, can be found as the unique solution of the (convex quadratic) optimization problem:

$$\begin{aligned}
 & \text{Minimize} \\
 & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \tag{2} \\
 & \text{Subject to} \\
 & y_i - (\mathbf{w} \cdot \mathbf{x}) - b \leq \epsilon + \xi_i \\
 & (\mathbf{w} \cdot \mathbf{x}) + b - y_i \leq \epsilon + \xi_i^* \\
 & \xi_i, \xi_i^* \geq 0 \\
 & \text{for } i = 1, 2, \dots, l
 \end{aligned}$$

The first term of the objective (minimization) function in equation 2 deals with the complexity, and the second term deals with the accuracy (or, amount of training errors) of the model. In general, both terms cannot be minimal (or, close to zero) at the same time. The nonnegative parameter C determines the trade-off between the flatness of $f(\mathbf{x})$ and the amount of tolerated deviations. If C is large, some flatness could be lost in order to achieve greater training accuracy.

All points on the boundary of the ϵ -insensitive region together with the points outside that region (the training errors) are called “support vectors”. The computation of the regression is solely based on the support vectors.

The minimization problem of equation 2 can be represented in dual form, as a maximization problem:

$$\begin{aligned}
& \text{Maximize} \\
& -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(\mathbf{x}_i, \mathbf{x}_j) + \\
& + \sum_{i=1}^l (\alpha_i - \alpha_i^*)y_i - \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) \\
& \text{Subject to} \\
& 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, l \text{ and} \\
& \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0
\end{aligned} \tag{3}$$

where $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)$.

Here we introduced with the kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ instead of the inner product $(\mathbf{x}_i \cdot \mathbf{x}_j)$ the possibility to replace the inner product with a different kernel function (see below).

In SVR, the regression estimates, resulting from solving equation 3 take the form of:

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i^* - \alpha_i)k(\mathbf{x}, \mathbf{x}_i) + b \tag{4}$$

The value of b can be found from the so-called Karush-Kuhn-Tucker (KKT) conditions associated with the dual optimization problem (equation 3).

The training points in the series in equation 4 with coefficient $(\alpha_i^* - \alpha_i)$ unequal to zero are exactly the support vectors. For each training point \mathbf{x}_i atmost one of the two numbers α_i and α_i^* is unequal to zero. For the training points on the boundary of the ϵ -insensitive region holds either $0 < \alpha_i < C$ or $0 < \alpha_i^* < C$ and for the training errors outside the ϵ -insensitive region holds either $\alpha_i = C$ or $\alpha_i^* = C$.

Application of a kernel function transforms the original input space implicitly into a higher-dimensional input space where an optimal linear decision surface

(corresponding to a nonlinear decision surface in the original input space) is found. One of the most frequently applied kernels is the so-called Radial Basis Function kernel (RBF). The dimension of the feature space for the RBF is infinite, which on first sight is counterintuitive from a complexity perspective: that should lead to overfitting. However, the literature reports very good performance of SVR using the RBF kernel (see, e.g. [11, 15, 35]). Possible theoretical explanations thereof have been suggested in [11]. Therefore, it appears that SVR with a RBF kernel are able to tackle the problems of overfitting effectively. For this reason we apply this kernel in our research.

The RBF kernel is defined as $k(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$, where γ is a manually adjustable parameter. The Radial Basis Function kernel is equal to 1 if $\mathbf{x}_i = \mathbf{x}_j$ and drops monotonically to zero with the Euclidean distance $\|\mathbf{x}_i - \mathbf{x}_j\|$ between the vectors \mathbf{x}_i and \mathbf{x}_j . The greater the value of γ , the faster the function $k(\mathbf{x}_i, \mathbf{x}_j)$ decreases. So, for large values of γ the influence of a training point will be only local and the risk of overfitting will be large. So, the larger γ , the more “complex” the radial basis function is, and the smaller the number of training errors.

Summarizing, we have three parameters ϵ , C , and γ , which have to be tuned in order to find the optimal trade-off between complexity and training accuracy of the SVR. One of the ways to find the best trade-off between these parameters is via the standard cross-validation technique, which will be explained in section 3.2.

3.2 SVR style timing models

We will present the construction of the value rotation model only, since the size rotation model is constructed analogically.¹⁰

The input vectors for the SVR consist of the (historical) values for all 17 candidate explanatory factors as described in table 2. The outputs are the corresponding differences in returns between the S&P 500 Barra Value and Growth indices. Each SVR model is trained on the data for months $t - 60$ till month $t - 1$ in order to predict the output for month t . In order to find the optimal model parameters ϵ , C and γ we applied 5-fold cross-validation, a standard technique in machine learning (see e.g., [42, 45]) on the training data sets of 60 months. A k -fold cross-validation procedure is utilized as follows: a given dataset is divided into k folders of equal size; subsequently, a model is built on all possible (k) combinations of $k - 1$ folders, and each time the remaining one folder is used for validation. The model that achieves minimum mean sum of squared errors on average (over the k validation folders) is considered to be the best. This best model is said to achieve minimum cross-validation mean squared error, and the parameters of this model are used in the final model for the prediction of month t .

The advantage of using a cross-validation procedure is that it ensures that model selection is based entirely on out-of-sample rather than in-sample performance. The disadvantage however is that the procedure is rather time-consuming.

¹⁰ The software program used throughout the analysis is LIBSVM 2.4, developed by Chang and Lin [14].

A tiny part of the cross-validation procedure is visualized in figure 3, where the vertical axis shows the cross validation minimal squared errors for $C \in (0, 32)$, while keeping ϵ and the kernel function parameter γ fixed at 1.0 and 0.007, respectively. As suggested by the figure, the value for the minimum cross-validation mean squared error is well defined.

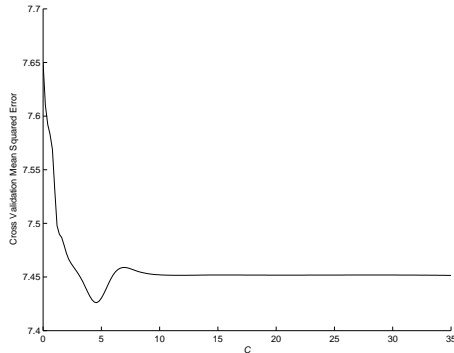


Fig. 3. Five-fold cross validation mean squared errors associated with the penalty-on-error parameter $C \in (0, 32)$ and fixed ϵ -insensitive loss function parameter (ϵ) at 1.0 and Radial Basis Function parameter at 0.007. The to-be-predicted month here is April 2000. The “best” model is the one for which the combination of the three parameters over suitable parameter ranges produces minimal cross validation mean squared error.

The predicted output, i.e. the value premium for month t is used to decide on our timing rotation strategy. A positive output will result in a signal “Value” in which case we will buy the Value index and sell the Growth index, while a negative output will result in a signal “Growth” with the opposite effect. In order to avoid taking decisions based on noise, we treat an output value close to zero as a “no signal” signal.¹¹

The SVR small-large strategy is defined analogically, using S&P SmallCap 600 and S&P 500 indices.

4 Empirical Results

In this section we will present the main results from value-growth and small-large rotation strategies using SVR with different levels of transaction costs and varying forecast horizons (one-month, three-month and six-month). Additionally, we will show the output of an equally weighted combination of both strategies. Throughout this empirical section we show returns that could be achieved when we would have been able to forecast the signal correctly each month: MAX_VG

¹¹ We used a range of $(-0.05, 0.05)$ standard deviations relative to the average of the estimates over the training period.

(value-growth rotation) and MAX_SL (small-large rotation). The input for the SVR model consists of 60 months of data on the whole set of 17 predetermined factors. The passive style strategies are constructed in accordance with what is expected in the literature: each month a long position is taken in the Value index and a short position in the Growth index. The passive small-large strategy consistently buys the S&P SmallCap 600 index and sells the S&P 500 index.

4.1 Value-growth rotation strategies

Detailed results of the SVR value-growth strategy can be found in table 3. What strikes most at first sight, is that this strategy has produced much better results than the passive strategy in the out-of-sample period starting January 1993 and ending December 2002.

Under the assumption of zero transaction costs, the SVR strategy achieves an annualized mean return of 10.30%, against a modest 0.24% respective return of the passive strategy.

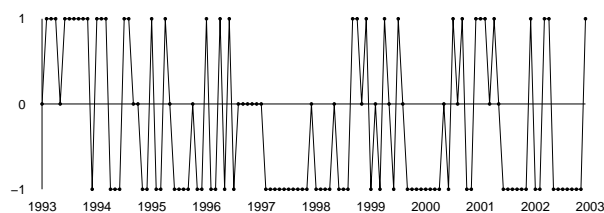


Fig. 4. Investment signals (“value” = 1, “growth” = -1, “no signal” = 0) produced by the SVR value-growth model investment strategy.

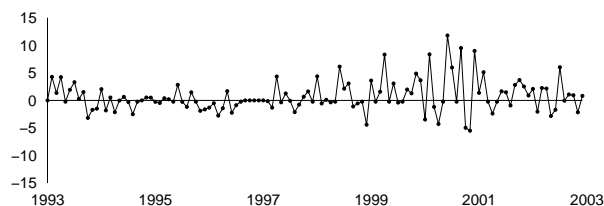


Fig. 5. Realized excess returns forecasted by the SVR value-growth investment strategy for the 0 bp transaction costs scenario.

Combining these results with the standard deviations of returns yields (annualized) information ratios of 1.04 and 0.02, respectively. Besides, even when high transaction costs of 50 bp (single trip) are added into the calculations, the realized SVR-model information ratio remains quite high (0.64), and statistically significant at the 5% two-tail level. When compared to other studies on the

Table 3. Passive and Support Vector Regression value-growth rotation strategies. Summary statistics for the passive value-growth and the Support Vector Regression value-growth one-month-ahead return forecasting models for the period 1993:01 to 2002:12. The predicted variables are the monthly return differences between S&P Barra Value and Growth indices. VmG denotes the passive value-growth strategy. MAX_VG denotes the perfect foresight style rotation strategy. CV denotes the timing strategy based on Support Vector Regression Cross Validation Mean Squared Error. All numbers are annualized data unless stated otherwise. All strategies are long/short monthly positions on the S&P Barra Value and Growth indices. The overall position for month $t + 1$ is based on the signal produced by the optimal model based on 60 months of historical data of all explanatory factors. Transaction costs are assumed to be 0 bp, 25 bp, and 50 bp single trip.

	value-growth rotation				
	VmG	CV, 0 bp	CV, 25 bp	CV, 50 bp	MAX_VG, 50 bp
Mean	0.24	10.30	8.30	6.30	21.54
Standard deviation	10.95	9.95	9.90	9.90	7.83
Information ratio	0.02	1.04***	0.84***	0.64**	2.75***
Z(equality)		2.15***	1.73*	1.30	5.00***
Median	-0.11	0.33	0.31	0.30	0.50
Minimum (monthly)	-12.02	-5.51	-5.51	-5.51	-0.98
Maximum (monthly)	9.74	12.02	11.77	11.52	11.02
Skewness (monthly)	0.01	1.21	1.18	1.13	1.61
Excess kurtosis (monthly)	2.40	2.66	2.52	2.34	3.41
% negative months	45.83	32.50	49.17	50.00	19.17
Largest 3-month loss	-11.55	-5.90	-6.40	-6.90	-1.99
Largest 12-month loss	-22.86	-8.07	-11.51	-15.26	2.21
% months in Growth	0.00	53.33	53.33	52.33	45.83
% months in Value	100.00	28.33	28.33	28.33	54.17
% months no position	0.00	18.33	18.33	18.33	0.00

* indicates significance at the (2-tail) 10% level

** indicates significance at the (2-tail) 5% level

*** indicates significance at the (2-tail) 1% level

subject, for example [7] in the U.S. and [28] in the U.K., the SVR results seem to demonstrate a significant improvement. The calculated Z(equality)-scores¹² provide further evidence (in the 0 bp and 25 bp transaction-cost environment) of a significant performance difference. In addition, the SVR strategy is able to capture 37.7%, 34.0%, and 29.3% of the return from the MAX_VG strategy under 0 bp, 25 bp and 50 bp transaction costs, respectively. Table 3 further reveals that the largest three-month and twelve-month losses associated with the SVR value-growth model are substantially less than the respective losses incurred by

¹² Z(equality) measures the risk-adjusted performance difference between a switching Support Vector Regression strategy and the passive value-growth strategy. The Z(equality)-score is computed in a standard way (in line with, e.g. [22]).

the passive strategy. Summarizing, all of these findings can serve as an indication of robustness of the SVR strategy.

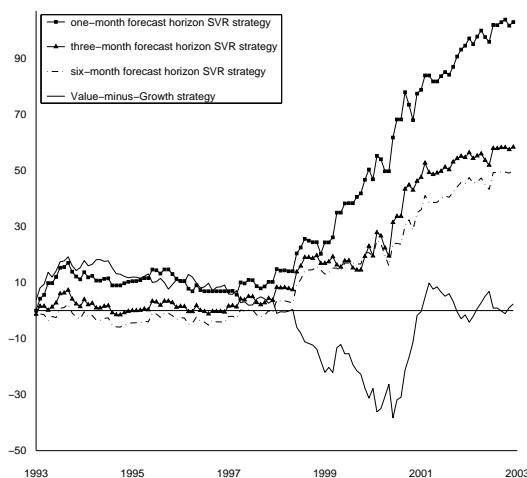


Fig. 6. Accrued cumulative returns from the passive value-growth strategy and the Support Vector Regression (SVR) one-, three-, and six-month horizon strategies for the period January 1993 – December 2002, under no transaction costs. The one-month horizon strategy performs best, gaining most of its accumulated profits during turbulent times on the financial market. In such periods, the three- and six-month horizon models follow suit with a time lag, as logically expected. During relatively calmer periods, all strategies perform similarly.

Figure 4 shows style signals associated with the SVR value-growth rotation strategy. The predominant style signal during this period is “Growth”, with some notable exceptions however. “Value” signals have been produced mostly in 1993, in the beginning of 1994, and in the first half of 2001. Almost no “Value” signals have been given during the periods stretching from June 1996 till August 1998, and from June 1999 till November 2000.

Figure 5 presents the realized excess returns forecasted by the basic SVR style timing strategy in the 25 bp transaction-cost scenario. It can be seen from the figure that most of the accrued returns come out of the last four years of the sample period, which actually appears to be the most volatile.

A number of further conclusions can be drawn by examining figure 6. Next to the cumulative returns from the passive strategy and the SVR strategy that predicts the one-month-ahead return difference under zero transaction costs, the figure reveals the cumulative returns from two more strategies: the three- and six-month-horizon SVR strategies. The latter two strategies are constructed simply by taking the (unweighted) average of the signals produced by models constructed up to three and up to six months before any predicted month, and

investing according to this combined signal. Our procedure is equal to that used in [25].

Table 4. Passive and Support Vector Regression small-large rotation strategies. Summary statistics for the passive small-large and the Support Vector Regression small-large one-month-ahead return forecasting models for the period 1993:01 to 2002:12. The predicted variables are the monthly return differences between S&P 500 and S&P SmallCap 600 indices, respectively. SmL denotes the passive small-large strategy. MAX_SL denotes the perfect foresight size rotation strategy. CV denotes the timing strategy based on Support Vector Regression Cross Validation Mean Squared Error. All numbers are annualized data unless stated otherwise. All strategies are long/short monthly positions. The overall position for month $t + 1$ is based on the signal produced by the optimal model based on 60 months of historical data of all explanatory factors. Transaction costs are assumed to be 0 bp, 25 bp, and 50 bp single trip.

	small-large rotation				
	SmL	CV, 0 bp	CV, 25 bp	CV, 50 bp	MAX_SL, 50 bp
Mean	-1.26	10.71	9.11	7.51	27.04
Standard deviation	13.00	10.92	10.92	10.96	8.87
Information ratio	-0.10	0.98***	0.83***	0.69**	3.05***
Z(equality)		2.23**	1.93*	1.63	5.69***
Median	0.05	0.45	0.44	0.43	0.79
Minimum (monthly)	-15.71	-7.70	-7.70	-7.70	-0.96
Maximum (monthly)	16.78	16.78	16.53	16.28	16.78
Skewness (monthly)	0.20	0.77	0.76	0.74	2.63
Excess kurtosis (monthly)	4.49	4.74	4.55	4.30	11.74
% negative months	51.67	33.33	44.07	45.83	10.00
Largest 3-month loss	-21.63	-8.84	-8.84	-9.21	-1.18
Largest 12-month loss	-31.85	-3.21	-5.46	-7.71	8.03
% months in Large	0.00	45.00	45.00	45.00	51.67
% months in Small	100.00	45.00	45.00	45.00	48.33
% months no position	0.00	10.00	10.00	10.00	0.00

* indicates significance at the (2-tail) 10% level

** indicates significance at the (2-tail) 5% level

*** indicates significance at the (2-tail) 1% level

The first striking feature is that most of the cumulative returns are accrued in times of relatively higher volatility, and especially during 1993, in the beginning of 1994, and between 1999 and 2002. The magnitude of the volatility of returns can be observed by tracking the (monthly) changes in the cumulative returns of the passive strategy. Larger shocks in these series correspond to greater volatility of the value premium. A second interesting feature is that the basic one-month-horizon SVR strategy performs better than in the case of three- and six-month forecast horizons. Apparently, it is worthwhile to time investment styles on a very short horizon.

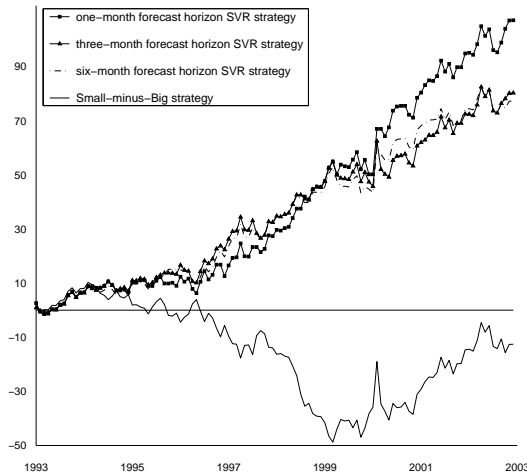


Fig. 7. Accrued cumulative returns from the passive small-large strategy and the Support Vector Regression (SVR) one-, three-, and six-month horizon strategies for the period January 1993 – December 2002, under no transaction costs. The one-month horizon strategy performs best, gaining the predominant part of its accumulated profits during turbulent times on the financial market. The three- and six-month horizon models follow a very similar pattern, but perform slightly worse. During relatively calmer periods, all strategies perform similarly.

4.2 Small-large rotation strategies

Detailed information on the small-large SVR strategy, the passive small-large strategy and the maximum attainable MAX_SL strategy can be found in table 4.

In the out-of-sample period the passive small-large rotation strategy achieves an annual return of -1.26%. The optimal MAX_SL strategy provides an annual return of 27.04% in the 50 bp transaction-cost scenario, which is 5.50% more than the corresponding result for the MAX_VG strategy. This reveals that the potential benefit from size rotation seems to be much greater than the one from the corresponding value-growth rotation. Table 4 shows that this extra potential can indeed be captured. For the zero-transaction-cost regime, for example, the one-, three- and six-month forecast horizon small-large strategies produce 10.71%, 8.03% and 7.73% annual returns, while the respective results from the SVR value-growth strategies are 10.30%, 5.84% and 5.02% respectively. Moreover, as in the value-growth case, the SVR size model is able to capture roughly one third of the maximum attainable cumulative returns under all considered transaction cost regimes.¹³

As it turns out, the results from the robustness checks that were performed on the SVR value-growth model are also valid for SVR size rotation. Under the assumption of 50 bp transaction costs, the realized information ratio of 0.69 from

¹³ Results are available upon request.

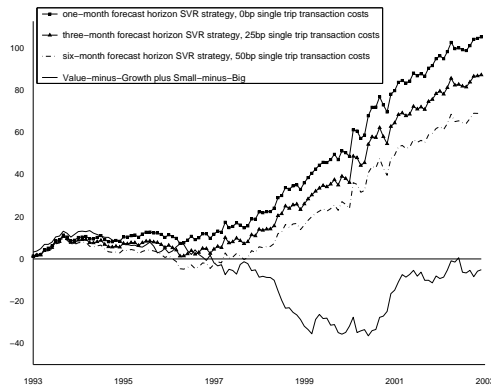


Fig. 8. Accrued cumulative returns from investing simultaneously according to the Support Vector Regression (SVR) one-month horizon value-growth and small-large strategies on the one hand, and from investing simultaneously according to the passive value-growth and small-large strategies during the period January 1993 – December 2002, under different transaction costs.

the SVR size model is significant at the (two-tail) 5% level. The SVR size strategy produces significantly different results from the passive size strategy. The largest three-month and, especially, twelve-month losses from the SVR model are drastically more bearable than the ones from the passive size strategy: -8.84% vs. -21.63% and -3.21% vs. -31.85%, respectively, under zero transaction costs. Notice, additionally, that the one-month strategy again outperforms the longer horizon alternatives, consistent with the findings of the SVR value-growth strategy, see figure 7.

4.3 Simultaneous value-growth and size timing

In case investors have decided to follow both the value and size SVR strategies simultaneously at the beginning of the sample period, they would have witnessed even greater relative gains as compared to sticking only to single-style timing (see table 5 and figure 8 for details).

Indicative of this are the realized information ratios of simultaneous style timing: 1.27, 1.06 and 0.84 under 0 bp, 25 bp and 50 bp single trip transaction cost regimes, all significant at the (two-tail) 1% level. Not surprisingly, these information ratios are higher than the ones associated with either value or size timing individually, as investors actually diversify the risk associated with each timing strategy. The information ratio of the simultaneous timing strategy is negative (-0.06). Interestingly, the largest three-month and twelve-month losses associated with simultaneous investing turn out to be quite tolerable: -4.10% and -3.96%, assuming zero transaction costs. It appears that it pays to diversify the market timing strategies, at least as far as value and size timing are concerned.

Admittedly, we expect all of our findings to be dependent on the historical model-building horizon and on the length of the trading period. Choosing to

Table 5. Simultaneous passive and Support Vector Regression value-growth and size rotation strategies. Summary statistics for following simultaneously passive value-growth (VmG) and passive small-large (SmL) strategies on the one hand, and Support Vector Regression value-growth and small-large rotation strategies using a one-month forecast horizon, on the other, for the period 1993:01 to 2002:12. The explanatory variables are listed in table 2. MAX denotes the perfect foresight value-growth and size rotation combined strategy. CV denotes the timing strategy based on Support Vector Regression cross validation mean squared error. All numbers are annualized data unless stated otherwise. All strategies are long/short monthly positions on the style and size indices. The overall position for month $t + 1$ is based on the signal produced by the optimal model based on 60 months of historical data of all explanatory factors. Transaction costs are assumed to be 0 bp, 25 bp, and 50 bp single trip.

	simultaneous rotation				
	VmG plus SmL	CV, 0 bp	CV, 25 bp	CV, 50 bp	MAX, 50 bp
Mean	-0.51	10.51	8.71	6.91	24.29
Standard deviation	8.71	8.25	8.23	8.23	6.76
Information ratio	-0.06	1.27***	1.06***	0.84***	3.59***
Z(equality)		2.90***	2.43**	1.96*	7.11***
Median	0.06	0.44	0.41	0.39	0.56
Minimum (monthly)	-7.26	-3.86	-4.24	-4.61	-0.94
Maximum (monthly)	8.43	12.56	12.44	12.31	12.06
Skewness (monthly)	0.13	1.51	1.49	1.45	1.87
Excess kurtosis (monthly)	1.12	5.29	5.30	5.23	6.05
% negative months	49.17	33.33	37.50	40.83	7.50
Largest 3-month loss	-13.20	-4.10	-5.10	-6.10	-1.17
Largest 12-month loss	-26.71	-3.96	-6.34	-8.82	6.77

* indicates significance at the (2-tail) 10% level

** indicates significance at the (2-tail) 5% level

*** indicates significance at the (2-tail) 1% level

trade for a longer period could come at the expense of incurring formidable transaction costs, as noted in the Data section. Additionally, varying the length of the model-building horizon might yield a “best” horizon that would be difficult to justify. Thus, future research could concentrate on both of these issues.

5 Conclusion

This paper examines whether short-term directional variations in the size and value premium in the U.S. stock market are sufficiently predictable to be exploited by means of a tactical timing strategy. As a forecasting tool, we employ so-called Support Vector Regressions (SVR). SVR have only recently been developed in the artificial intelligence field and have been rarely applied in a financial context. Using SVR, we are able to circumvent the well-known problems of overfitting, especially in multivariate settings, in an elegant way.

Our empirical findings clearly show that both premiums are highly predictable during the trading period. This comes at odds with the mainstream literature that provides evidence for the long-term superiority of returns to value vis-a-vis growth and small vis-a-vis large stocks. After adjustment for fair levels of transaction costs this result still holds. Under high transaction cost levels, expected to be relevant in a dynamic economic environment, it is difficult in practice to obtain incremental benefits over style consistent strategies. That is why it is critical to develop timing strategies that can be implemented using index futures or low-cost trading baskets like exchange traded funds. In terms of realized information ratios, a combination of both value-growth and small-large timing produces most interesting results.

References

1. P. Ahmed, L. Lockwood, and S. Nanda. Multistyle rotation strategies. *Journal of Portfolio Management*, 28:17–29, 2002.
2. R.D. Arnott, J.L. Dorian, and R. Macedo. Style management: The missing element in equity portfolios. *Journal of Investing*, 1:13–21, 1992.
3. R.D. Arnott, D.L. Rice, C.M. Kelso, S. Kiscadden, and R. Macedo. Forecasting factor returns: An intriguing possibility. *Journal of Portfolio Management*, 16:28–35, 1989.
4. C. Asness, J. Friedman, R. Krail, and J. Liew. Style timing: Value versus growth. *Journal of Portfolio Management*, 26:51–60, 2000.
5. R.W. Banz. The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9:3–18, 1981.
6. R. Bauer, J. Derwall, and R. Molenaar. The real-time predictability of the size and value premium in Japan. *Pacific-Basin Finance Journal*, 2004. . Forthcoming.
7. R. Bauer and R. Molenaar. Is the value premium predictable in real time? LIFE working Paper 02-003, 2002.
8. R. Bernstein. *Navigate the noise: Investing in the new age of media and hype*. John Wiley and Sons, New York, 2001.
9. P. Bossaerts and P. Hillion. Implementing statistical criteria to select return forecasting models: What do we learn? *Review of Financial Studies*, 12:405–428, 1999.
10. R. Burbidge and B. Buxton. An introduction to support vector machines for data mining. In M. Sheppee, editor, *Keynote Papers, Young OR12*, pages 3–15, University of Nottingham, March 2001. Operational Research Society.
11. C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.
12. K.C. Chan and N.-F. Chen. Structural and return characteristics of small and large firms. *Journal of Finance*, 46:1467–1484, 1991.
13. L.K. Chan and J. Lakonishok. Value and growth investing: Review and update. *Financial Analysts Journal*, 60(1):71–86, 2004.
14. Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines, 2002.
15. M.-W. Chang, B.-J. Chen, and C.-J. Lin. EUNITE network competition: Electricity load forecasting, 2001. Winner of EUNITE world wide competition on electricity load prediction.
16. M. Cooper, H. Gulen, and M. Vassalou. Investing in size and book-to-market portfolios using information about the macroeconomy: some new trading rules, 2001. mimeo, Columbia University, New York.

17. M.M Copeland and T.E. Copeland. Market timing: style and size rotation using the VIX. *Financial Analysts Journal*, 55:73–81, 1999.
18. Elroy Dimson, Stefan Nagel, and Garrett Quigley. Capturing the value premium in the United Kingdom. *Financial Analysts Journal*, 59(6), 2003.
19. E.F. Fama and K.R. French. The cross-section of expected stock returns. *Journal of Finance*, 47:427–465, 1992.
20. E.F. Fama and K.R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–53, 1993.
21. E.F. Fama and K.R. French. Value versus growth: the international evidence. *Journal of Finance*, 53:1975–1999, 1998.
22. S.A. Glantz. *Primer of biostatistics*. McGraw-Hill, third edition, 1992.
23. R.A. Haugen and N.L. Baker. Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41:401–439, 1996.
24. B.I. Jacobs and K.N. Levy. High definition style rotation. *Journal of Investing*, 5:14–23, 1996.
25. N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: Implications for market efficiency. *Journal of Finance*, 48:65–91, 1993.
26. D.-L. Kao and R. Shumaker. Equity style timing. *Financial Analysts Journal*, 55:37–48, 1999.
27. J. Lakonishok, A. Schleifer, and R.W. Vishny. Contrarian investment, extrapolation and risk. *Journal of Finance*, 49:1541–1578, 1994.
28. M. Levis and M. Liodakis. The profitability of style rotation strategies in the united kingdom. *Journal of Portfolio Management*, 25(1):73–86, 1999.
29. J. Liew and M. Vassalou. Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics*, 57:221–245, 2000.
30. A. Lo and A.C. MacKinlay. Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3:431–468, 1990.
31. A. Lucas, R. Van Dijk, and T. Kloek. Stock selection, style rotation, and risk. *Journal of Empirical Finance*, 9:1–34, 2002.
32. M. Maragoudakis, K. Kermanidis, N. Fakotakis, and G. Kokkinakis. Combining bayesian and support vector machines learning to automatically complete syntactical information for HPSG-like formalisms. In *LREC 2002, 3rd International Conference on Language Resources and Evaluation, Las Palmas, Spain*, volume 1, pages 93–100, 2002.
33. K.L. Miller, H. Li, and D.E. Cox. U.S. style rotation model, 2001. Industry Note, Salomon, Smith Barney.
34. A. Monteiro. Interest rate curve estimation: A support vector regression application to finance. Working paper, 2001.
35. K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.
36. K.-R. Müller, A.J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V.N. Vapnik. Predicting time series with support vector machines. In W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Proceedings of the International Conference on Artificial Neural Networks*, volume 1327 of *Springer Lecture Notes in Computer Science*, pages 999–1004. Springer, 1997.
37. M.H. Pesaran and A. Timmermann. Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, 50:1201–1228, 1995.

38. C.M. Rocco S. and J.A. Moreno. A support vector machine model for currency crises discrimination. In S. Chen and P. Wang, editors, *Computational Intelligence in Economics and Finance*, Advanced Information Processing, pages 171–81. Springer-Verlag, 2003.
39. R. Schwob. Style and style analysis from a practitioners perspective: What is it and what does it mean for european equity investors. *Journal of Asset Management*, 1:39–59, 2000.
40. A.J. Smola. Regression estimation with support vector learning machines. Master’s thesis, Technische Universität München, 1996.
41. A.J. Smola and B. Schölkopf. A tutorial on support vector regression. NeuroCOLT2 Technical Report NC-TR-98-030, University of London, UK, 1998.
42. M. Stone. Asymptotics for and against cross-validation. *Biometrika*, 64:29–35, 1977.
43. T. Van Gestel, B. Baesens, J. Garcia, and P. Van Dijcke. A support vector machine approach to credit scoring. *Bank en Financiewezen*, 2:73–82, 2003.
44. Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995. 2nd edition, 2000.
45. S.M. Weiss and C.A. Kulikowski. *Computer systems that learn*. Morgan Kaufman, 1991.
46. Brendon J. Woodford. Comparative analysis of the EFuNN and the support vector machine models for the classification of horticulture data, 2001.