

## Research Article

# A New Kernel of Support Vector Regression for Forecasting High-Frequency Stock Returns

Hui Qu<sup>1</sup> and Yu Zhang<sup>2</sup>

<sup>1</sup>*School of Management and Engineering, Nanjing University, No. 22 Hankou Road, Nanjing 210093, China*

<sup>2</sup>*School of Physics, Nanjing University, No. 22 Hankou Road, Nanjing 210093, China*

Correspondence should be addressed to Hui Qu; [linda59qu@nju.edu.cn](mailto:linda59qu@nju.edu.cn)

Received 22 December 2015; Accepted 31 March 2016

Academic Editor: Yannis Dimakopoulos

Copyright © 2016 H. Qu and Y. Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper investigates the value of designing a new kernel of support vector regression for the application of forecasting high-frequency stock returns. Under the assumption that each return is an event that triggers momentum and reversal periodically, we decompose each future return into a collection of decaying cosine waves that are functions of past returns. Under realistic assumptions, we reach an analytical expression of the nonlinear relationship between past and future returns and introduce a new kernel for forecasting future returns accordingly. Using high-frequency prices of Chinese CSI 300 index from January 4, 2010, to March 3, 2014, as empirical data, we have the following observations: (1) the new kernel significantly beats the radial basis function kernel and the sigmoid function kernel out-of-sample in both the prediction mean square error and the directional forecast accuracy rate. (2) Besides, the capital gain of a simple trading strategy based on the out-of-sample predictions with the new kernel is also significantly higher. Therefore, we conclude that it is statistically and economically valuable to design a new kernel of support vector regression for forecasting high-frequency stock returns.

## 1. Introduction

Although the efficient market hypothesis is one of the most influential theories in the past few decades, researchers have never given up examining the predictability of stock returns. The complexity of the market makes the relationship between past and future financial data nonlinear [1, 2]. Linear statistical models, such as the autoregressive (AR) model, the autoregressive moving average (ARMA) model, and the autoregressive integrated moving average (ARIMA) model, are apparently powerless when compared to nonlinear approaches such as the generalized autoregressive conditional heteroskedasticity (GARCH) model, the artificial neural network (ANN), and the support vector machine for regression (SVR). Atsalakis and Valavanis [3] provide a very comprehensive review of the nonlinear models used in stock market forecasting.

With its remarkable generalization performance, support vector machine (SVM), firstly designed for pattern recognition by Vapnik [4], has gained extensive applications

in regression estimation (in which it is called SVR) and is thus introduced to time series forecasting problems. SVR is compared with multiple other models such as the backpropagation (BP) neural network [5–7], the regularized radial basis function (RBF) neural network [6], the case-based reasoning (CBR) approach [7], the GARCH-class models [8] and shows superior forecast performance. The kernel function used in SVR plays a crucial role in capturing the nonlinear dynamics of the time series under study. Several commonly used kernels, for example, the radial basis function kernel, are firstly derived mathematically and are widely applied in time series forecasting problems. Parameters are empirically tuned by researchers to achieve good performance of prediction. In addition, several researchers argue that using a single kernel may not solve a complex problem satisfactorily and thus propose the multiple-kernel SVR approach [9–11]. Yeh et al. [11] show that multikernel SVR outperforms single-kernel SVR in forecasting the daily closing prices of Taiwan capitalization weighted stock index. Besides, Huang et al. [12] linearly combine the predicting result of SVR with those of other

classifiers trained with the same data set and realize better forecast performance of the weekly movement direction of NIKKEI 225 index. However, current applications seldom touch the inner structure of the existing kernels, which depicts the nonlinear relationship between past and future data. Thus, it is reasonable to argue that if the kernel is designed according to the specific nonlinear dynamics of the series under study, improvement in forecast accuracy can be expected.

In addition, the above-mentioned studies mostly use daily (or lower frequency) data in their empirical experiments. Since high-frequency trading has gained its popularity in recent years, the ability to forecast intraday stock returns is becoming increasingly important. Thus in this study, we instead consider the forecasting of high-frequency stock returns. Matías and Reboledo [13] and Reboledo et al. [14] empirically show the forecast ability of SVR for high-frequency stock returns by directly using the radial basis function kernel. Instead of directly applying some conventional kernel or some combination of conventional kernels, we design a kernel for the specific forecasting problem. Specifically, under the assumption that each high-frequency stock return is an event that triggers momentum and reversal periodically, we decompose each future return into a collection of decaying cosine waves that are functions of past returns. After taking several realistic assumptions, we reach an analytical expression of the nonlinear relationship between past and future returns and design a new kernel accordingly. One-minute returns of Chinese CSI 300 index are used as empirical data to evaluate the new kernel of SVR. We show that the new kernel significantly beats the conventional radial basis function and sigmoid function kernels in both the prediction mean square error and the directional forecast accuracy rate. Besides, the capital gain of a practically simple trading strategy based on the predictions with the new kernel is also significantly higher.

The remainder of this paper is organized as follows. Section 2 introduces the basic idea of SVR. Section 3 presents our basic assumptions and designs the new kernel. Section 4 determines the newly introduced kernel parameters and compares the new kernel with two commonly used kernels in terms of the forecast performance of the SVR. Finally, the conclusions are drawn in Section 5.

## 2. Support Vector Machine for Regression

First designed by Vapnik as a classifier [4], SVR is featured with the capability of capturing nonlinear relationship in the feature space and thus is also considered as an effective approach to regression analysis. The following sketches the basic idea of SVR. For more detailed illustration of SVR, please refer to Burges [15].

**2.1. SVR for Linear Regression.** In a regression problem, given a finite data set  $F = \{(\mathbf{x}_k, y_k)\}_{k=1}^n$  derived from an unknown function  $y = g(\mathbf{x})$  with noise, we need to determine a function  $y = f(\mathbf{x})$  solely based on  $F$  and to minimize the difference between  $f$  and the unknown function  $g$ . For linear

regression,  $g$  is assumed to be a linear relationship between  $\mathbf{x}$  and  $y$

$$y = g(\mathbf{x}, \mathbf{w}, b) = \mathbf{w} \cdot \mathbf{x} + b = \sum_{j=1}^m w_j x_j + b, \quad (1)$$

where  $\mathbf{x}$  is called feature vector and the space  $\mathbf{X}$  it lives in is named as feature space.  $m$  is the dimension of the feature vector  $\mathbf{x}$  and the feature space  $\mathbf{X}$ .  $y$  is referred to as the label for each  $(\mathbf{x}, y)$ . Now that the relationship to be determined is assumed linear, our goal is to find a hyperplane  $y = f(\mathbf{x})$  in the  $m + 1$  dimension space, where  $\{(\mathbf{x}_k, y_k)\}_{k=1}^n$  are plotted and to minimize the fitting errors by adjusting the parameters. As is proven by Vapnik, the hyperplane is given as

$$y = f(\mathbf{x}, \boldsymbol{\alpha}, b) = \sum_k \alpha_k y_k \mathbf{x}_k \cdot \mathbf{x} + b, \quad (2)$$

where  $\mathbf{x}_k$ 's are support vectors in the given data set  $F$  and  $y_k$ 's are the corresponding labels. “ $\cdot$ ” represents the inner product in the feature space  $\mathbf{X}$ . Finding the support vectors and determining the parameters  $\boldsymbol{\alpha}$  and  $b$  turn out to be a linearly constrained quadratic programming problem that can be solved in multiple ways (e.g., the sequential minimal optimization algorithm [16]). Such a process conducted on the given data set  $F$  is called learning. Once the learning phase is done, the model built can be used to predict the corresponding label  $y$  from any feature vector  $\mathbf{x}$  in the feature space  $\mathbf{X}$ .

**2.2. SVR for Nonlinear Regression.** However, the linear relationship assumption is often too simple to characterize the dynamics of the time series, and thus it is necessary to consider the case when  $g$  is nonlinear. The idea of SVR for nonlinear regression is to build a mapping  $\mathbf{x} \rightarrow \boldsymbol{\phi}(\mathbf{x})$  from the original  $m$  dimension feature space  $\mathbf{X}$  to a new feature space  $\mathbf{X}'$  whose dimension depends on the mapping scheme and is not necessarily finite. In the new space  $\mathbf{X}'$ , the relationship between the new feature vector  $\boldsymbol{\phi}(\mathbf{x})$  and label  $y$  is believed to be in a linear form. By building a proper mapping, the nonlinear relationship can be approximated by doing in the new feature space  $\mathbf{X}'$  exactly the same thing as is done for the linear case, and it can be proven that the nonlinear version of (2) is

$$y = f(\mathbf{x}, \boldsymbol{\alpha}, b) = \sum_k \alpha_k y_k K(\mathbf{x}_k, \mathbf{x}) + b, \quad (3)$$

where  $K(\mathbf{x}_k, \mathbf{x}) = \boldsymbol{\phi}(\mathbf{x}_k) \cdot \boldsymbol{\phi}(\mathbf{x})$  is the kernel function and “ $\cdot$ ” represents the inner product in the new feature space  $\mathbf{X}'$ .

The new feature  $\boldsymbol{\phi}(\mathbf{x})$ , which can be an infinite dimension vector, is usually not necessary to be computed explicitly, since we normally work with the kernel function in the training and forecasting phases. Accordingly, the kernel function is essential to the performance of SVR. Any function satisfying Mercer's condition can be used as the kernel function. Commonly used kernels include the radial basis function kernel  $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|/2\sigma^2)$  and the sigmoid function kernel  $K(\mathbf{x}, \mathbf{y}) = \tanh(\kappa \mathbf{x} \cdot \mathbf{y} - \delta)$ , where  $\sigma$ ,  $\kappa$ , and  $\delta$  are kernel parameters that can be tuned.

### 3. A New Kernel for Forecasting High-Frequency Stock Returns

Most applications of SVR directly apply the commonly used kernels and tune the kernel parameters for improved forecast performance. However, we argue that a specifically designed kernel function, which builds on the properties of the underlying data, can enable the SVR to better capture the nonlinear relationship between the original feature vector  $\mathbf{x}$  and label  $y$ . Thus, in this study, we develop a new kernel specifically for forecasting high-frequency stock returns from some basic assumptions about the stock market.

**3.1. High-Frequency Stock Return Series Forecasting Problem.** High-frequency stock return series refers to the return time series with one-minute or comparatively small time intervals. Given a return time series  $I_n = \{r_n, r_{n-1}, r_{n-2}, \dots\}$  from present  $t = n$  to some time point in the history, a forecasting problem is to find the one-step-ahead return  $r_{n+1}$  based on the knowledge of  $I_n$ .

To put it in another way, we need to determine the function  $r_{n+1} = f(r_n, r_{n-1}, r_{n-2}, \dots)$  that best fits the given return series  $I_n$ . (Some studies introduce exogenous variables in stock return series forecasting. However, since we focus on the high-frequency return series where the inefficiency of the market is obvious [17, 18], it is reasonable to believe that the return series itself can provide enough information for its forecasting.) The vector  $\mathbf{r}_n = (r_n, r_{n-1}, r_{n-2}, \dots)$  is thus the feature vector, and  $r_{n+1}$  is the label. The training data set  $F$  includes every single return as label and all the returns before it as the elements of the corresponding feature vector. According to former studies,  $f$  is believed to be nonlinear considering the complexity of the financial market.

**3.2. The New Kernel for SVR.** It is straightforward that every single return has impact on the market. Due to behavioral effects such as overreaction and underreaction, such impact is not unidirectional during its remaining period. In this study, we assume that each high-frequency stock return is an event that triggers momentum and reversal periodically and thus express the impact generated by the return  $r_i$  as

$$A(r_i, t_p) = r_i \cos[\omega(i, t_p)|r_i|t_p + \varphi(i)] e^{-\lambda t_p}, \quad (4)$$

where  $i$  is the time point when the return  $r_i$  first occurs,  $t_p$  is the time past since  $t = i$ , and  $\lambda$  is the parameter that controls the decay rate.  $\omega(i, t_p)|r_i|$  is the frequency of the cosine wave where  $\omega(i, t_p)$  is a nonzero factor, and  $\varphi(i)$  is the phase factor. Such an expression ensures that the same level of return occurring at different time points can have different impact waves.

According to (4), a larger return leads to greater (amplitude) and faster (frequency) price change in the future, and as time goes by, such an impact wave will gradually fade to vanity. This exactly meets the basic intuition about how

the market reacts to events. For subsequent deduction, (4) is transformed as follows:

$$A(r_i, t_p) = r_i \left\{ a_i \cos[\omega(i, t_p)|r_i|t_p] + (1 - a_i^2)^{1/2} \sin[\omega(i, t_p)|r_i|t_p] \right\} e^{-\lambda t_p}, \quad (5)$$

where  $a_i \in [-1, 1]$ .

It is natural to assume that every return is comprised of all the impact waves generated by the past returns. (We only consider the predictable part of every future return and ignore the innovation part.) Thus, we decompose every return  $r_n$  into a collection of decaying cosine waves

$$r_n = \sum_{i=1}^{\infty} A(r_{n-i}, i\Delta t), \quad (6)$$

where  $\Delta t$  denotes the time interval length, which is 1 minute in this study, and thus the time passed since event  $r_{n-i}$  is  $t_p = i\Delta t$ .

Now we substitute (5) into (6). Taylor serial expansion is used on the “sin” and “cos” terms and the coefficients are rearranged to generate the following:

$$\begin{aligned} r_n &= \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} C_{i,k} r_{n-i} [\omega(n-i, i\Delta t)|r_{n-i}|i\Delta t]^k e^{-\lambda i\Delta t} \\ &= \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} C'_{i,k} r_{n-i} (|r_{n-i}|i\Delta t)^k e^{-\lambda i\Delta t}, \end{aligned} \quad (7)$$

where  $\{C'_{i,k}\}_{i,k}$  is the coefficient set to be determined.

Therefore, based on the assumptions above, a nonlinear relationship between past returns and the one-step-ahead future return is derived. It is easy to see that  $r_n$  is a linear combination of  $\{r_{n-i}(|r_{n-i}|i\Delta t)^k e^{-\lambda i\Delta t}\}_{i,k}$ . Thus, it is possible to map the original feature vectors to a proper form to make the relationship between label and feature vector linear. However, we once again fall into the dilemma that (7) is a collection of infinite series, which makes the mapped feature vectors have infinite dimension and hard to compute. It is also unlikely to derive the kernel function instead like before.

To solve the above-mentioned problem, we first consider the decay property of the impact waves, which is presented by the factor  $e^{-\lambda i\Delta t}$ . Since the decay rate  $\lambda$  is constant, we believe that the impact of an event that is  $m\Delta t$  before the time being is negligible, and  $m$  is the minimum integer that satisfies the condition  $e^{-\lambda m\Delta t}/e^{-\lambda\Delta t} < \delta$ , where  $\delta \rightarrow 0$ . Since we do not know how little  $\delta$  should be,  $m$  is determined through experiments in Section 4.2.

It is also necessary to consider the high-frequency property of the data. Such a property ensures that the scale of the time interval  $\Delta t$  is subtle compared to the time scale of the fluctuation of the market. Thus, it is reasonable to assume  $\Delta t \ll 2\pi/\omega(i, t_p)|r_i|$ ,  $\forall i$ , where the right hand side is the period of the impact wave of return  $r_i$  and “ $\ll$ ” represents at least 2 orders of magnitude smaller. Furthermore, since the Taylor serial expansion reserved till order  $q$  has an error

$\text{err}(x) = (f^{(q+1)}(\xi)/(q+1)!)x^{q+1}$ , the error from truncating the Taylor expansions in (7) satisfies

$$\begin{aligned} \text{err} &\leq \frac{[\omega(n-i, i\Delta t) |r_{n-i}| i\Delta t]^{q+1}}{(q+1)!} \\ &= \left( \frac{\omega(n-i, i\Delta t) |r_{n-i}| \Delta t}{\ll 1} \right)^{q+1} \cdot \frac{(2\pi i)^{q+1}}{(q+1)!}. \end{aligned} \quad (8)$$

Thus, setting  $q = 3$  can well ensure the accuracy of the approximation.

Now, a much more accessible approximation of (7) is derived

$$r_n = \sum_{i=1}^m \sum_{k=0}^q C'_{i,k} r_{n-i} (|r_{n-i}| i\Delta t)^k e^{-\lambda i\Delta t}. \quad (9)$$

Equation (9) explicitly states how the mapping is constructed. For any past return series  $\{r_{n-i}\}_{i=1}^m$ , the original feature vector is  $\mathbf{r}_{n-1} = (r_{n-1}, r_{n-2}, \dots, r_{n-m})$ , and the mapping  $\phi$  is defined as

$$\phi(\mathbf{r}_{n-1}) = \left\{ r_{n-i} (|r_{n-i}| i\Delta t)^k e^{-\lambda i\Delta t} \right\}_{i=1, k=0}^{m, q}. \quad (10)$$

It is important to understand that the new feature vector  $\phi(\mathbf{r})$  is  $m \times (q+1)$  dimension, with each element  $\phi_{i,k}$  given as  $\phi_{i,k} = r_{n-i} (|r_{n-i}| i\Delta t)^k e^{-\lambda i\Delta t}$ . Now that  $\phi(\mathbf{r})$  has finite dimension, the dimension of the new feature space  $\mathbf{X}'$  is also finite, and the kernel function is naturally built as the inner product in such a space

$$K(\mathbf{r}^{(1)}, \mathbf{r}^{(2)}) = \sum_{i=1}^m \sum_{k=0}^q \phi_{i,k}^{(1)} \phi_{i,k}^{(2)}. \quad (11)$$

Although it is theoretically important to examine whether the new kernel satisfies Mercer's condition or not, considering that some kernels which fail to meet the condition still lead to perfectly converged results [15], we will examine the appropriateness of the new kernel through experiments.

## 4. Empirical Experiments

**4.1. Data.** One-minute prices of Chinese CSI 300 index from January 4, 2010, to March 3, 2014 (1000 trading days), are used as empirical data in this study. The data are obtained from Wind Financial Terminal, and days with missing prices are deleted. The official trading hours are from 9:30 to 11:30 and from 13:00 to 15:00, and thus we have 240 one-minute returns per day. The returns within the same trading day are used for learning and forecasting, since the continuousness of the time is important in the above derivation. (Although there is a 1.5-hour break in each trading day, buy and sell orders submitted in the morning are still valid and new orders can still be submitted during this period; thus we deem the time as continuous in each trading day.) Specifically, the first  $100 + m$  returns within each trading day are set as the in-sample data ( $m \leq 30$ ), which are used for learning, and the

last 110 returns are set as the out-of-sample data, which are used for prediction and evaluation. The average performance of the SVR during the first 500 trading days, that is, from January 4, 2010, to February 1, 2012, is used for determining the kernel parameters. The last 500 trading days, that is, from February 2, 2012, to March 3, 2014, is used for performance comparison against the commonly used kernels. To improve the performance, the logarithmic returns are normalized to  $[-1, 1]$  before being input into the SVR.

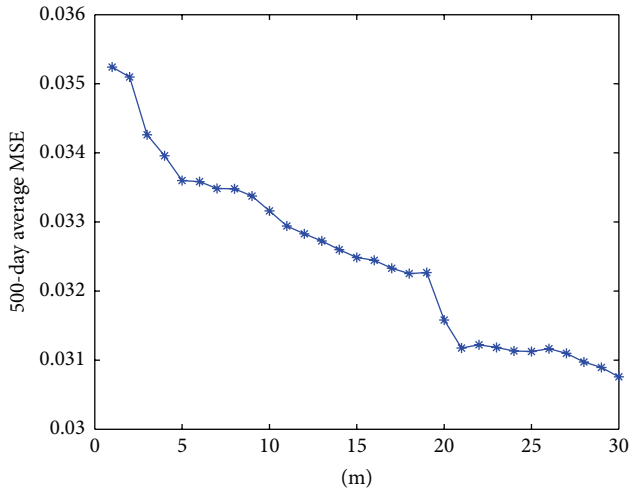
**4.2. Determining the Kernel Parameters.** Before evaluating the performance of the new kernel, multiple parameters still need to be determined to make best use of the kernel. The undetermined parameters are  $m$ , which measures how many historical data are used,  $\lambda$ , which measures how fast the impact of one event decays with time, and  $\Delta t$ , which indicates how we represent 1 minute numerically. We optimize the parameters according to the out-of-sample forecast performance of the corresponding SVR, which is evaluated by MSE and hit rate, respectively. MSE is the mean square error of the predictions and is computed using the normalized returns. Hit rate is the proportion of predicted returns that have the same sign with the actual ones, that is, the directional forecast accuracy rate.

To determine  $m$ , all the other parameters are fixed, and  $m$  varies from 1 to 30. The average MSE and hit rate in the first 500 trading days are plotted in Figures 1(a) and 1(b), respectively. (Figures 1(a) and 1(b) are plotted with  $\lambda$  and  $\Delta t$  set to the optimal values. Actually, the trends of the plots do not vary with the other kernel parameters.) The average MSE decreases sharply with  $m$  when  $m \leq 21$ , is relatively constant when  $m \in [21, 26]$ , and decreases slowly with  $m$  when  $m \in [26, 30]$ . The average hit rate increases sharply with  $m$  when  $m \leq 21$ , is relatively constant when  $m \in [21, 25]$ , and gets smaller afterwards. Considering that smaller MSE and greater hit rate are preferred,  $m = 30$  and  $m \in [21, 25]$  are all appropriate choices suggested by the experiments. In the following comparative experiments,  $m$  is set to 25. (We have also done comparative experiments with  $m$  set to the other suggested values, and the results are consistent.)

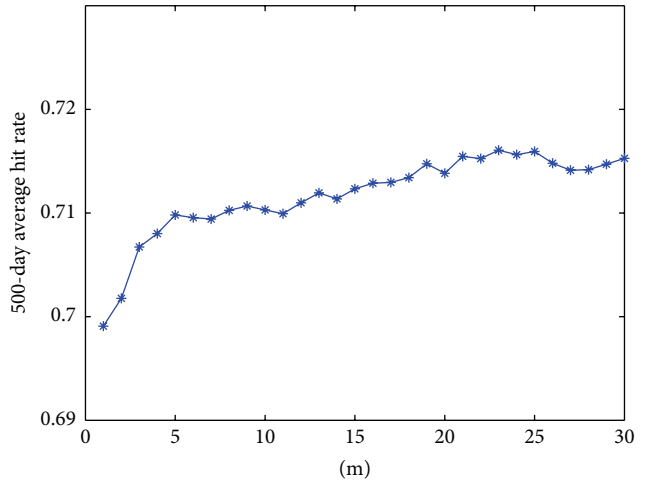
To determine  $\lambda$ , all the other parameters are fixed, and  $\lambda$  varies from 1 to 50. The average MSE and hit rate in the first 500 trading days are plotted in Figures 2(a) and 2(b), respectively. It is quite clear that the minimum average MSE and maximum average hit rate are both achieved at  $\lambda = 10$ , and thus we set  $\lambda$  to 10 in the following comparative experiments.

The same method is used to optimize  $\Delta t$ , and the average MSE and hit rate are plotted in Figures 3(a) and 3(b), respectively. We can see that the minimum average MSE and maximum average hit rate are both achieved at  $\Delta t = 0.05$ , and thus we set  $\Delta t$  to 0.05 in the following comparative experiments.

**4.3. New Kernel versus Commonly Used Kernels.** The new kernel is compared with the commonly used kernels in terms of the out-of-sample forecast performance of the corresponding SVR. Although any function satisfying Mercer's condition

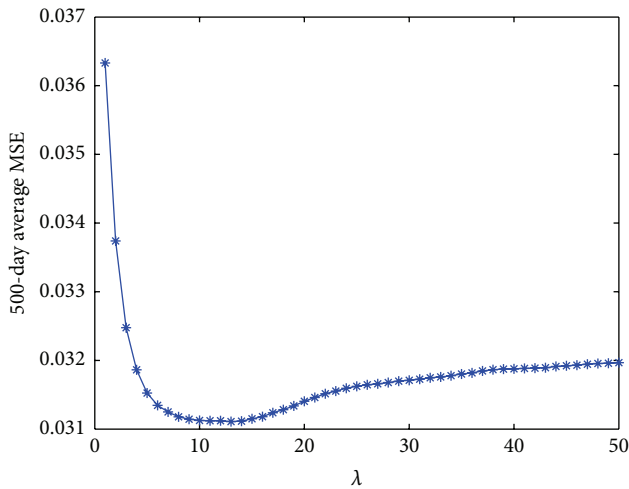


(a) Average MSE versus  $m$

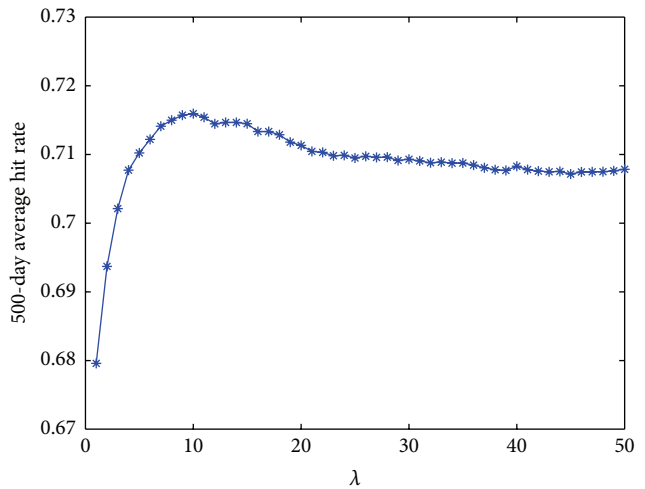


(b) Average hit rate versus  $m$

FIGURE 1: 500-day average MSE and hit rate achieved with different values of  $m$ .



(a) Average MSE versus  $\lambda$



(b) Average hit rate versus  $\lambda$

FIGURE 2: 500-day average MSE and hit rate achieved with different values of  $\lambda$ .

can be used as the kernel function, radial basis function and sigmoid function are two widely used kernels. The former tends to outperform others under general smoothness assumption [12], and the latter gives a particular kind of two-layer sigmoidal neural network [15]. Thus, these two kernel functions are used for performance comparison. We realize each SVR by LIBSVM-3.20 [19], and the related codes are modified for the new kernel.

The out-of-sample MSE and hit rate are computed for each of the three kernels on each of the last 500 trading days. The results of the new kernel are plotted against those of the radial basis function kernel and the sigmoid function kernel in Figures 4 and 5, respectively.

In Figure 4(a), the vertical and horizontal axes represent the MSE of the new kernel and the radial basis function kernel, respectively, while, in Figure 4(b), the vertical and horizontal axes represent the hit rate of the new kernel and the

radial basis function kernel, respectively. There are 500 points corresponding to the 500 trading days plotted in each subplot, and the diagonal line representing  $y = x$  is for reference. We can see that most points lie below the line  $y = x$  in Figure 4(a) and lie above the line  $y = x$  in Figure 4(b). This indicates that the new kernel leads to smaller MSE and greater hit rate than the radial basis function kernel in most of the 500 trading days. Similarly, Figures 5(a) and 5(b) indicate that the new kernel leads to smaller MSE and greater hit rate than the sigmoid function kernel in most of the 500 trading days. Therefore, the SVR with the new kernel has obviously better forecast performance in terms of both the MSE and the hit rate.

In addition, a simple trading strategy is carried out based on the out-of-sample forecasts of the SVR with different kernel specifications. The initial capital is set as 100 on each trading day. The index is bought if the one-step-ahead

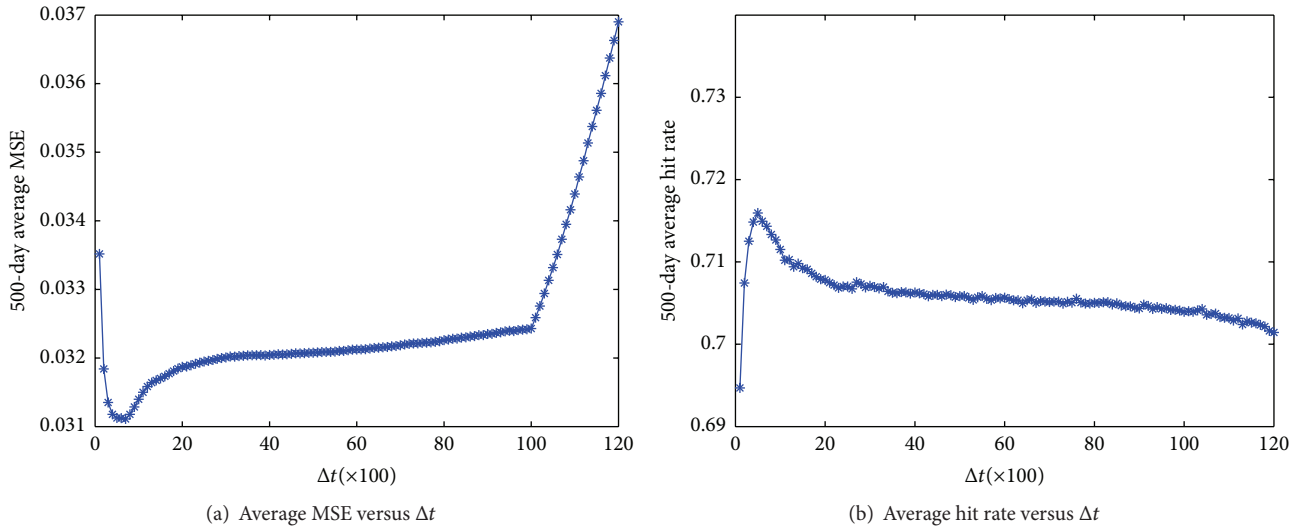


FIGURE 3: 500-day average MSE and hit rate achieved with different values of  $\Delta t$ .

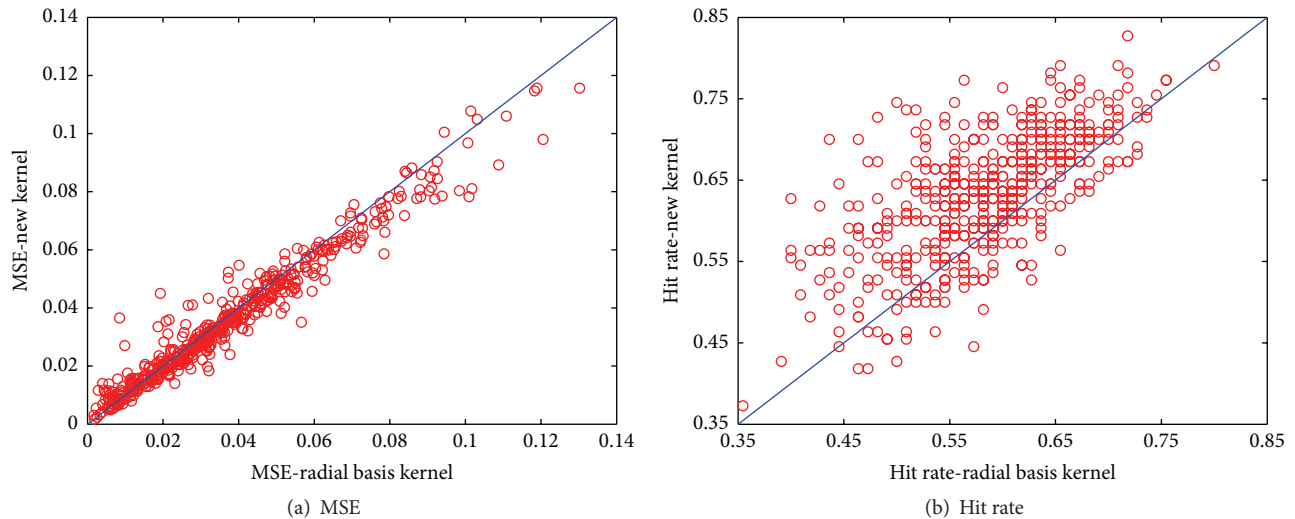


FIGURE 4: New kernel versus radial basis function kernel from February 2, 2012, to March 3, 2014.

predicted return is positive and exceeds a threshold and sold if the one-step-ahead predicted return is negative and below a threshold, and no action is performed otherwise. The threshold is set as the average of the  $(100 + m)$  normalized returns in the training period divided by the scale coefficient of  $\ln(1000)$ . (The scale coefficient controls the trading strategy's sensitivity to index price change, and a higher value leads to more frequent trading. The value of  $\ln(1000)$  is arbitrarily set and can be adjusted. The results are consistent as the scale coefficient varies.)

The variation of capital under such a strategy in the 110-minute out-of-sample period on February 2, 2012, is plotted in Figure 6(a). We can see that the new kernel leads to higher capital gain than the other two kernels most of the time. Also, the average capital variation in the last 500 trading days is plotted in Figure 6(b). Unlike the fluctuant plots in Figure 6(a), the capital increases steadily when it is averaged

over the 500-day period, no matter which kernel is used. This confirms the effectiveness of SVR in forecasting high-frequency stock returns. Once again, the new kernel leads to the highest capital gain and the advantage gets more obvious as the trading period is prolonged. At the end of a day, the new kernel leads to a return at about 0.6%/110 min on average, while the resulting returns of the other two kernels are both less than 0.4%/110 min on average.

Furthermore, Student's  $t$ -test is used to test whether the new kernel significantly outperforms the commonly used kernels. Specifically, we calculate the differences of MSE, hit rate, and 110-minute capital gain between the forecasts with the new kernel and those with each comparative kernel, respectively, on each of the last 500 trading days. Table 1 reports the mean ( $\bar{x}_d$ ) and the standard deviation ( $s_d$ ) of these differences in these 500 trading days. We test the null hypothesis that "the forecasts with the new kernel and those

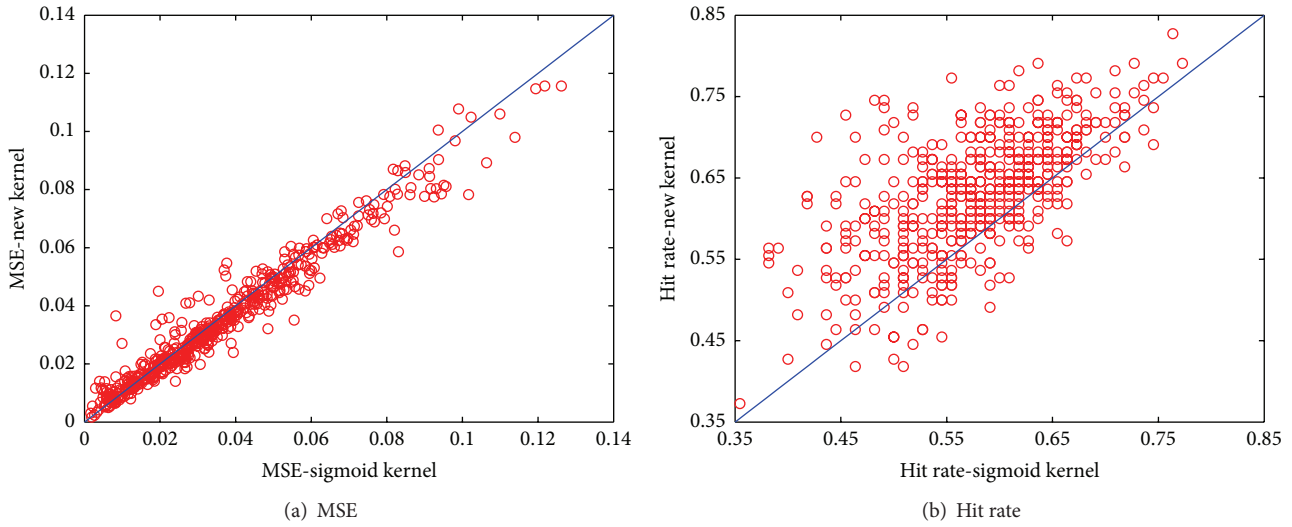


FIGURE 5: New kernel versus sigmoid function kernel from February 2, 2012, to March 3, 2014.

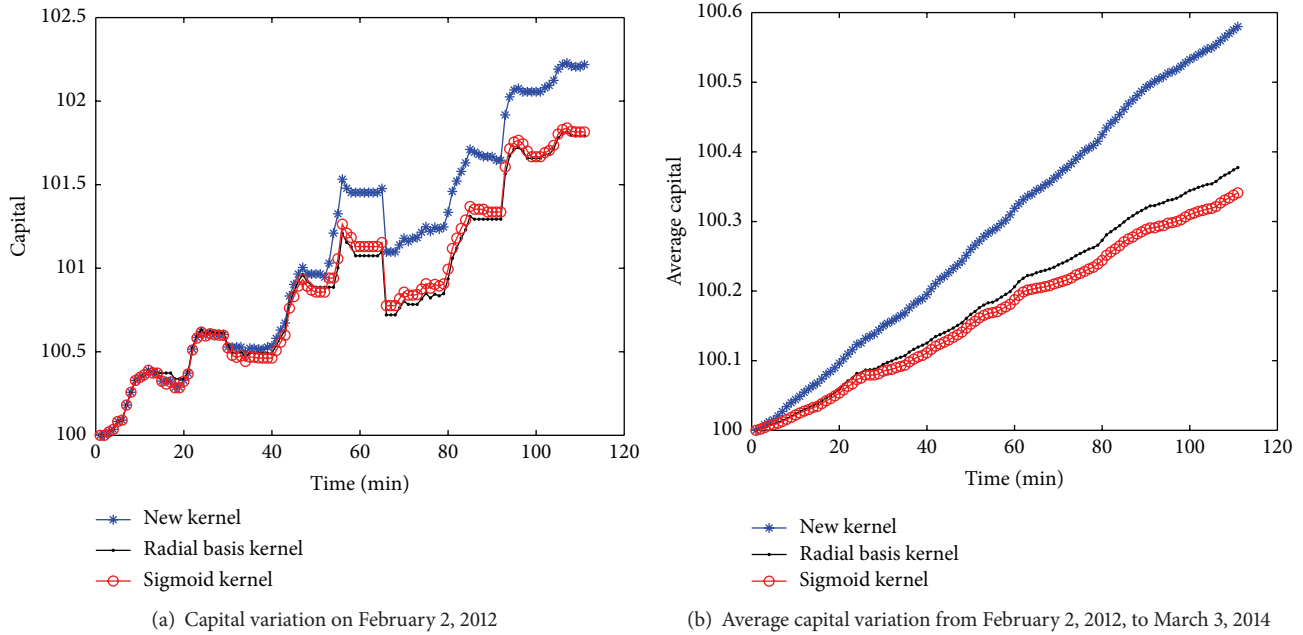


FIGURE 6: Capital variation of a simple trading strategy.

with the comparative kernel have the same accuracy in terms of the specified criterion,” with the comparative kernel and the criterion specified in the first row and the second row of Table 1, respectively. The  $t$ -statistics are reported in the last row of Table 1.

We can easily see that the six null hypotheses are all significantly rejected. The new kernel has smaller MSE, greater hit rate, and higher 110-minute capital gain than both the radial basis function kernel and the sigmoid function kernel. And the differences are all significant at the 1% level. Therefore, the results of Student’s  $t$ -test indicate that the improvement in out-of-sample forecast accuracy brought about by the new kernel is significant both statistically and economically, and

thus the new kernel is preferred in forecasting high-frequency stock returns.

### 5. Summary and Conclusion

Support vector machine for regression is now widely applied in time series forecasting problems. Commonly used kernels such as the radial basis function kernel and the sigmoid function kernel are first derived mathematically for pattern recognition problems. Although their direct applications in time series forecasting problems can generate remarkable performance, we argue that using a kernel designed according

TABLE 1: Out-of-sample forecast performance comparison results from February 2, 2012, to March 3, 2014 ( $n = 500$ ).

Criterion	New kernel versus (minus) radial basis kernel			New kernel versus (minus) sigmoid kernel		
	MSE	Hit rate	110-min gain	MSE	Hit rate	110 min gain
$\bar{x}_d$	$-1.5201e - 3$	$4.8709e - 2$	$2.0250e - 3$	$-1.3917e - 3$	$5.3945e - 2$	$2.3879e - 3$
$s_d$	$5.3684e - 3$	$6.0785e - 2$	$2.8530e - 3$	$5.3658e - 3$	$6.3875e - 2$	$3.2202e - 3$
$t = \frac{\bar{x}_d \sqrt{n}}{s_d}$	$-6.33^{***}$	$17.92^{***}$	$15.87^{***}$	$-5.80^{***}$	$18.88^{***}$	$16.58^{***}$

Note: \*\*\* indicates significance at the 1% level.

to the specific nonlinear dynamics of the series under study can further improve the forecast accuracy.

Under the assumption that each high-frequency stock return is an event that triggers momentum and reversal periodically, we decompose each future return into a collection of decaying cosine waves that are functions of past returns. Under realistic assumptions, we reach an analytical expression of the nonlinear relationship between past and future returns and thus design a new kernel specifically for forecasting high-frequency stock returns. Using high-frequency prices of Chinese CSI 300 index as empirical data, we determine the optimal parameters of the new kernel and then compare the new kernel with the radial basis function kernel and the sigmoid function kernel in terms of the SVR's out-of-sample forecast accuracy. It turns out that the new kernel significantly outperforms the other two kernels in terms of the MSE, the hit rate, and the capital gain from a simple trading strategy.

Our empirical experiments confirm that it is statistically and economically valuable to design a new kernel of SVR specifically characterizing the nonlinear dynamics of the time series under study. Thus, our results shed light on an alternative direction for improving the performance of SVR. Current study only utilizes past returns to predict future returns. A natural extension is to introduce intraday trading volumes and intraday high/low prices into the feature vector of the SVR and develop kernels characterizing the corresponding nonlinear relationship between future return and feature vector. Another possible extension is to apply the SVR with new kernel in the energy markets where trading is continuous 24 hours. We leave them for future work.

## Competing Interests

The authors declare that they have no competing interests.

## Acknowledgments

The authors are grateful for the financial support offered by the National Natural Science Foundation of China (71201075) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (20120091120003).

## References

- [1] D. A. Hsieh, "Testing for nonlinear dependence in daily foreign exchange rates," *The Journal of Business*, vol. 62, no. 3, pp. 339–368, 1989.
- [2] J. A. Scheinkman and B. LeBaron, "Nonlinear dynamics and stock returns," *The Journal of Business*, vol. 62, no. 3, pp. 311–337, 1989.
- [3] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques-part II: soft computing methods," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5932–5941, 2009.
- [4] V. N. Vapnik, *The Nature Of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [5] L. Cao and F. E. H. Tay, "Financial forecasting using support vector machines," *Neural Computing & Applications*, vol. 10, no. 2, pp. 184–192, 2001.
- [6] L. J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1506–1518, 2003.
- [7] K.-J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1-2, pp. 307–319, 2003.
- [8] V. V. Gavrishchaka and S. Banerjee, "Support vector machine as an efficient framework for stock market volatility forecasting," *Computational Management Science*, vol. 3, no. 2, pp. 147–160, 2006.
- [9] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large scale multiple kernel learning," *Journal of Machine Learning Research*, vol. 7, pp. 1531–1565, 2006.
- [10] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "Simple MKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.
- [11] C.-Y. Yeh, C.-W. Huang, and S.-J. Lee, "A multiple-kernel support vector regression approach for stock market price forecasting," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2177–2186, 2011.
- [12] W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," *Computers and Operations Research*, vol. 32, no. 10, pp. 2513–2522, 2005.
- [13] J. M. Matías and J. C. Reboredo, "Forecasting performance of nonlinear models for intraday stock returns," *Journal of Forecasting*, vol. 31, no. 2, pp. 172–188, 2012.
- [14] J. C. Reboredo, J. M. Matías, and R. Garcia-Rubio, "Nonlinearity in forecasting of high-frequency stock returns," *Computational Economics*, vol. 40, no. 3, pp. 245–264, 2012.
- [15] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [16] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods: Support Vector Learning*, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds., chapter 12, pp. 185–208, MIT Press, Cambridge, Mass, USA, 1999.



- [17] T. Chordia, R. Roll, and A. Subrahmanyam, "Evidence on the speed of convergence to market efficiency," *Journal of Financial Economics*, vol. 76, no. 2, pp. 271–292, 2005.
- [18] D. Y. Chung and K. Hrazdil, "Speed of convergence to market efficiency: the role of ECNs," *Journal of Empirical Finance*, vol. 19, no. 5, pp. 702–720, 2012.
- [19] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Tech. Rep., Department of Computer Science, National Taiwan University, 2003.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

