

# NON-FIXED AND ASYMMETRICAL MARGIN APPROACH TO STOCK MARKET PREDICTION USING SUPPORT VECTOR REGRESSION

*Haiqin Yang, Irwin King and Laiwan Chan*

Department of Computer Science and Engineering  
The Chinese University of Hong Kong  
Shatin, N.T. Hong Kong

{hgyang, king, lwchan}@cse.cuhk.edu.hk

## ABSTRACT

Recently, Support Vector Regression (SVR) has been applied to financial time series prediction. Typical characteristics of financial time series are non-stationary and noisy in nature. The volatility, usually time-varying, of the time series is therefore some valuable information about the series. Previously, we had proposed to use the volatility to adaptively change the width of the margin of SVR. We have noticed that upside margin and downside margin do not necessarily be the same, and we have observed that their choice would affect the upside risk, downside risk and as well as the overall prediction result. In this paper, we introduce a novel approach to adapt the asymmetrical margins using momentum. We applied and compared this method to predict the Hang Seng Index and Dow Jones Industrial Average.

## 1. INTRODUCTION

SVR is a recently introduced approach to regression problems [5]. It is a variation of Support Vector Machine (SVM), which was developed by Vapnik and his co-workers [8, 9]. Nowadays, SVR has been successfully applied to time series prediction [3, 2] and financial forecasting [7, 6].

In general, SVR uses the  $\varepsilon$ -insensitive loss function to measure the empirical risk and minimizes the regression error based on the Structural Risk Minimization (SRM) principle [8]. The  $\varepsilon$ -insensitive loss function contains a fixed and symmetrical margin and it is insensitive and non-adaptive to the input data which may result in less-than-optimal performance in the testing data.

For financial data, due to the embedded noise, one must set a suitable margin in order to improve the accuracy of prediction. In [10], we have extended the standard SVR with adaptive margin and classified it as four cases: Fixed and Symmetrical Margin (FASM), Fixed and Asymmetrical Margin (FAAM), Non-fixed and Symmetrical Margin

(NASM) and Non-fixed and Asymmetrical Margin (NAAM). Comparing FASM with FAAM, we find that the downside risk can be reduced by employing asymmetrical margins. While comparing NASM with FASM, FAAM, a good predictive result is obtained by exploiting the standard deviation to calculate the margin.

However, NAAM requires the adaptation of the margin width and the degree of asymmetry, and no exact algorithm for such margin setting has been introduced.

In this paper, we propose to use NAAM which combines two characteristics of the margin; non-fixed and asymmetry, to reduce the downside risk while improving the accuracy of the financial prediction. More specially, we add the momentum term to achieve this. The width of the margin is controlled by the standard deviation [10]. The asymmetry of the margin is determined by the momentum. This momentum term can trace the up trend and down tendency of the stock prices. Since the financial time series are often following a long term trend but with small short term fluctuations, we exploit a larger up margin and a smaller down margin to under-predict the stock price when the momentum is positive and we use a smaller up margin and a larger down margin to over-predict the stock price while the momentum is negative. We show a simple illustration about this margin setting in Figure 1.

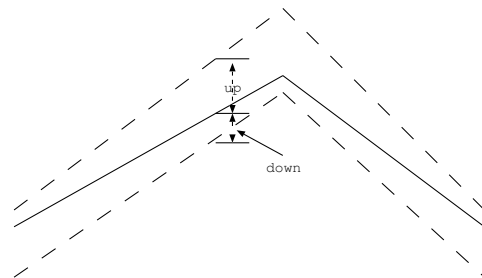


Figure 1: Margin Setting

We organize the paper as follows. Firstly, we give a brief

introduce of SVR with a general type of  $\varepsilon$ -insensitive loss function and the concept of momentum in Section 2. The accuracy metrics and experimental results are elucidated in Section 3. Lastly, we conclude the paper with a brief discussion and final remarks in Section 4.

## 2. SUPPORT VECTOR REGRESSION WITH MOMENTUM

### 2.1. Support Vector Regression

Suppose we are given a training data set,  $(x_1, y_1), \dots, (x_N, y_N)$ , where  $x_i \in X, y_i \in R, N$  is the size of training data, and  $X$  denotes the space of the input samples—for instance,  $R^n$ . Now the problem is to find a function which can estimate all these data well. SVR is one of the methods to perform the above regression task [8, 5].

In general, the estimation function in SVR takes the following form,

$$f(x) = (w \cdot \phi(x)) + b, \quad (1)$$

where  $(\cdot)$  denotes the inner product in  $\Omega$ , a feature space of possibly different dimensionality such that  $\phi : X \rightarrow \Omega$  and  $b \in R$ .

Now the question is to determine  $w$  and  $b$  from the training data by minimizing the regression risk,  $R_{reg}(f)$ , based on the empirical risk,

$$R_{reg}(f) = C \sum_{i=1}^N \Gamma(f(x_i) - y_i) + \frac{1}{2}(w \cdot w), \quad (2)$$

where  $C$  is a pre-specified value,  $\Gamma(\cdot)$  is a cost function that measures the empirical risk. In general,  $\varepsilon$ -insensitive loss function is used as the cost function [8, 5]. For this function, when the data points are in the range of  $\pm\varepsilon$ , they do not contribute to the output error. Thus, it leads to the sparseness of the solution. The function is defined as,

$$\Gamma(f(x) - y) = \begin{cases} 0, & \text{if } |y - f(x)| < \varepsilon \\ |y - f(x)| - \varepsilon, & \text{otherwise} \end{cases} \quad (3)$$

In [10], we have introduced a general type of  $\varepsilon$ -insensitive loss function,  $\Gamma'(f(x_i) - y_i), i = 1, \dots, N$ , which is given as,

$$\begin{cases} 0, & \text{if } -d(x_i) < y_i - f(x_i) < u(x_i) \\ y_i - f(x_i) - u(x_i), & \text{if } y_i - f(x_i) \geq u(x_i) \\ f(x_i) - y_i - d(x_i), & \text{if } f(x_i) - y_i \geq d(x_i) \end{cases}, \quad (4)$$

where  $d(x), u(x) \geq 0$ , are two functions to determine the down margin and up margin respectively. When  $d(x)$  and  $u(x)$  are both constant functions and  $d(x) = u(x)$ , Eq. (4)

amounts to the  $\varepsilon$ -insensitive loss function in Eq. (3) and we labeled it as FASM (Fixed and Symmetrical Margin). When  $d(x)$  and  $u(x)$  are both constant functions but  $d(x) \neq u(x)$ , this case is labeled as FAAM (Fixed and Asymmetrical Margin). In the case of NASM (Non-fixed and Symmetrical Margin),  $d(x) = u(x)$  but are varied with the data. The last case is with a non-fixed and asymmetrical margin (NAAM) where  $d(x)$  and  $u(x)$  are varied with the data and  $d(x) \neq u(x)$ .

We use the standard method to find the solution of Eq. (2) with the cost function of Eq. (4) as [8]. At first we construct a corresponding Lagrange function, then find a saddle point of the Lagrange function. After applying the dual theorem, we obtain the following Quadratic Programming (QP) problem:

$$\begin{aligned} \min Q(\alpha^{(*)}) = & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(\phi(x_i) \cdot \phi(x_j)) \\ & + \sum_{i=1}^N (u(x_i) - y_i)\alpha_i + \sum_{i=1}^N (d(x_i) + y_i)\alpha_i^* \end{aligned} \quad (5)$$

subject to

$$\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C], \quad (6)$$

where  $\alpha_i$  and  $\alpha_i^*$  are corresponding Lagrange multipliers used to push and pull  $f(x_i)$  towards the outcome of  $y_i$  respectively.

Solving the above QP problem of Eq. (5) with constraints of Eq. (6), we determine the Lagrange multipliers  $\alpha$  and  $\alpha^*$  and obtain  $w = \sum_{i=1}^N (\alpha_i - \alpha_i^*)\phi(x_i)$ . Therefore the estimation function in Eq. (1) becomes

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*)(\phi(x) \cdot \phi(x_i)) + b. \quad (7)$$

So far, we have not considered the computation of  $b$ . In fact, this can be solved by exploiting the Karush-Kuhn-Tucker (KKT) conditions. These conditions state that at the optimal solution, the product between the Lagrange multipliers and the constraints has to vanish. In this case, it means that

$$\begin{aligned} \alpha_i(u(x_i) + \xi_i - y_i + (w \cdot \phi(x_i)) + b) &= 0 \\ \alpha_i^*(d(x_i) + \xi_i^* + y_i - (w \cdot \phi(x_i)) - b) &= 0 \end{aligned} \quad (8)$$

and

$$\begin{aligned} (C - \alpha_i)\xi_i &= 0 \\ (C - \alpha_i^*)\xi_i^* &= 0. \end{aligned}$$

Since  $\alpha_i \cdot \alpha_i^* = 0$  and  $\xi_i^{(*)} = 0$  for  $\alpha_i^{(*)} \in (0, C)$ ,  $b$  can be computed as follows:

$$b = \begin{cases} y_i - (w \cdot \phi(x_i)) - u(x_i), & \text{for } \alpha_i \in (0, C) \\ y_i - (w \cdot \phi(x_i)) + d(x_i), & \text{for } \alpha_i^* \in (0, C) \end{cases}.$$

Using the trick of kernel function, Eq. (7) can be written as,  $f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x, x_i) + b$ , where the kernel function,  $K(x, x_i) = (\phi(x) \cdot \phi(x_i))$ , is a symmetric function and satisfies the Mercer's condition. In this paper, we select a common kernel function, e.g., RBF function,  $K(x, x_i) = \exp(-\beta|x - x_i|^2)$ , as the kernel function.

## 2.2. Momentum

In [10], we have considered the case of FASM, FAAM and NASM. In this paper, we will focus on the case of NAAM. More specially, we add a momentum term in the margin setting. The margin is a linear combination of the standard deviation and the momentum. The up margin and down margin are set in the following forms:

$$\begin{aligned} u(x_i) &= \lambda_1 \cdot \sigma(x_i) + \mu \cdot \Delta(x_i) i = 1, \dots, N, \\ d(x_i) &= \lambda_2 \cdot \sigma(x_i) - \mu \cdot \Delta(x_i) i = 1, \dots, N. \end{aligned} \quad (9)$$

where  $\sigma(x_i)$  is the standard deviation of input  $x_i$  which determines the width of margin and  $\Delta(x_i)$  is the momentum at point  $x_i$ .  $\lambda_1, \lambda_2$  are two positive constants called the coefficients of the margin width.  $\mu$  is a non-negative constant called the coefficient of momentum. Thus, when  $\mu = 0$ , it becomes the NASM case. When  $\mu \neq 0$  and  $\Delta(x) > 0$ , the up margin is larger than the down margin and we use them to under-predict the stock price. While  $\mu \neq 0$  and  $\Delta(x) < 0$ , the up margin is smaller than the down margin and we use them to over-predict the stock price.

As a matter of fact, there are many ways to calculate the momentum. For example, it may be set as a constant. In this paper, we will consider using the Exponential Moving Average (EMA), which is time-varying and can reflect the up trend and down tendency of the financial time series data. An  $n$ -day's EMA is calculated by

$$EMA_i = EMA_{i-1} \times (1 - r) + y_i \times r,$$

where  $r = 2/(1 + n)$  and it begins from the first day,  $EMA_1 = y_1$ . The current day's momentum is set as the difference between the current day's EMA and the EMA in the previous  $k$  day, i.e.  $\Delta(x_i) = EMA_i - EMA_{i-k}$ .

## 3. EXPERIMENTS

### 3.1. Accuracy Metrics

We use the following statistical metrics to evaluate the prediction performance in Table 1, including Mean Absolute

Error (M-AE), Up side Mean Absolute Error (UMAE), Down side Mean Absolute Error (DMAE). MAE is the measure of the discrepancy between the actual and predicted values. The smaller the value of MAE, the closer are the predicted values to the actual values. UMAE is the measure of up side risk. DMAE is the measure of down side risk. The smaller the values of them, the smaller the corresponding risks are.

Table 1: Accuracy Metrics and their calculations

Metrics	Calculation
MAE	$MAE = 1/m * \sum_{i=1}^m  a_i - p_i $
UMAE	$UMAE = 1/m * \sum_{i=1, a_i \geq p_i}^m (a_i - p_i)$
DMAE	$DMAE = 1/m * \sum_{i=1, a_i < p_i}^m (p_i - a_i)$

$a_i$  and  $p_i$  are the actual values and predicted values.  $m$  is the number of testing data.

### 3.2. Experimental Procedure and Results

In this section, we first use the SVR algorithm which is modified from LibSVM [1] and conduct experiments on two data sets to illustrate the effect of NASM, NAAM, next we compare them with AutoRegressive(AR) model and RBF network.

We use the daily closing prices of Hang Seng Index (HSI) and Dow Jones Industrial Average (DJIA) as the data sets in our experiments. The corresponding time periods are from January 2nd, 1998 to December 29, 2000. Furthermore, we model the system as  $p_t = f(a_{t-4}, a_{t-3}, a_{t-2}, a_{t-1})$  and set the ratio of the number of training data and the number of testing data to 5:1. Therefore the corresponding initial training time periods are listed as in Table 2.

In the first experiment, we firstly do a cross-validation on the initial training data to determine the parameters that are needed in SVR, they are  $C$ , the cost of error;  $\beta$ , the parameter of kernel function. The corresponding parameters are also listed in Table 2. With these parameters, we begin to build the model by SVR from the initial training data. After obtaining the predictive value, we shift the input window one day's step and train the model again to predict the next day's price. This one-step ahead prediction is done as the window shifted for the remaining data.

Table 2: Indices, time periods and parameters

Indices	Initial Training time periods	$C$	$\beta$
HSI	02/01/1998 - 04/07/00	16000	$2^{-27}$
DJIA	02/01/1998 - 29/06/00	8000	$2^{-22}$

The experiments are conducted on a Pentium 4, with 1.4 GHZ, 512M RAM and Windows2000. With these configurations, the predictive results are obtained within few hours.

We set the margins as Eq. (9). Concretely, in the case of NASM, we set  $\lambda_1 = \lambda_2 = \frac{1}{2}$  and  $\mu = 0$ , thus the overall

margin widths are equal to the standard deviation of input  $x$ . In the case of NAAM, we also fix  $\lambda_1 = \lambda_2 = \frac{1}{2}$ , therefore we have a fair comparison of NASM case. In addition, we set  $k = 1$ ,  $\mu = 1$  and use 10, 30, 50, 100 as the length of EMA. From the result of Table 3 and Table 4, we can see that the DMAE values in all cases of NAAM are smaller than that in NASM case, thus we have a smaller downside risk in NAAM case. We also see that the MAE gradually decreases with the length of EMA increases and when the length equals 100, the MAE and the DMAE are the smallest in all case of NAAM. In Table 4, the MAE decreases and then increases again. When the length equals 30, the MAE and the DMAE are also the smallest in all cases of NAAM.

We plot the price of HSI with 100-days' EMA and the price of DJIA with 30-days' EMA in Figure 2, Figure 3 respectively and list the average of the standard deviation of HSI and DJIA, the average of the absolute momentum in Table 5. We can observe that the volatility of HSI is higher than that of DJIA and the ratio of momentum and standard deviation is smaller for HSI than that for DJIA.

Table 3: Effect of the length of EMA on HSI

type	$n$	MAE	UMAE	DMAE
NASM		216.78	104.58	112.20
NAAM	10	222.43	115.64	106.79
	30	218.18	114.04	104.14
	50	217.93	113.38	104.55
	100	<b>216.50</b>	113.04	<b>103.46</b>

Table 4: Effect of the length of EMA on DJIA

type	$n$	MAE	UMAE	DMAE
NASM		85.33	40.29	45.04
NAAM	10	85.68	43.13	42.55
	30	<b>84.12</b>	41.82	<b>42.30</b>
	50	84.57	42.12	42.45
	100	84.80	42.41	42.39

Table 5: Std and EMA

data set	$n$	$\sigma$	$\Delta$	ratio
HSI	100	182.28	20.80	0.114
DJIA	30	79.95	15.64	0.196

In the second experiment, we use the AR model with order 4 to predict the price of HSI and DJIA, hence we can compare the AR model with NASM, NAAM in SVR with the same order. The results are listed in the Table 6, and we can see that NASM and NAAM are superior to AR model with same order.

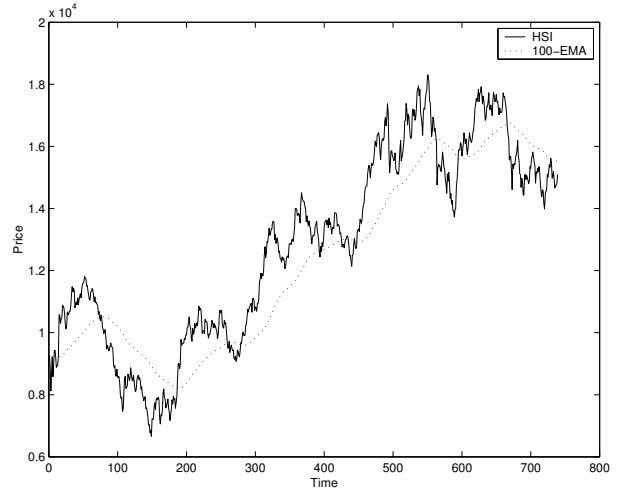


Figure 2: HSI and 100 days' EMA

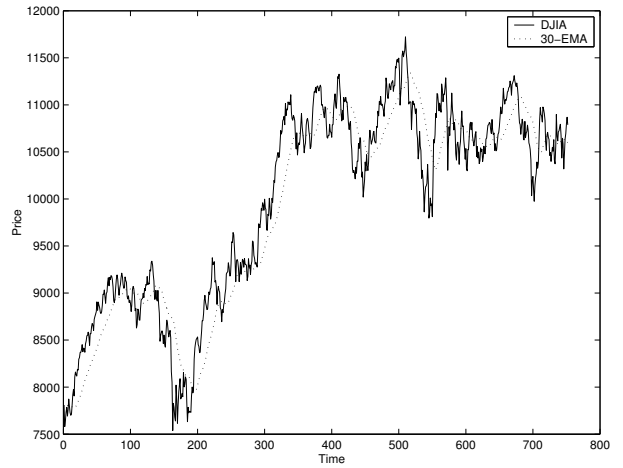


Figure 3: DJIA and 30 days' EMA

Table 6: Results on AR(4)

data set	MAE	UMAE	DMAE
HSI	217.75	105.96	111.79
DJIA	88.74	46.36	42.38

In the third experiment, we use the RBF network which was implemented in NETLAB [4] and perform the one-step ahead prediction to predict the price of HSI and DJIA. Concretely, we set the number of hidden units to 3, 5, 7, 9 to train the RBF network and get the results in Table 7 for HSI, in Table 8 for DJIA respectively. Comparing these two tables with Table 3 and Table 4, we can see that NASM and NAAM are also better than RBF network. Moreover an uncertain thing is that there are more parameters need to be determined in RBF network than that in SVR and here we just test one parameter, the number of hidden units, others are default.

Table 7: Effect of number of hidden units on HSI

# hidden	MAE	UMAE	DMAE
3	386.65	165.08	221.57
5	277.83	128.92	148.91
7	219.32	104.15	115.17
9	221.81	109.46	112.35

Table 8: Effect of number of hidden units on DJIA

# hidden	MAE	UMAE	DMAE
3	88.31	44.60	43.71
5	98.44	48.46	49.98
7	90.53	46.22	44.31
9	87.23	44.09	43.14

#### 4. DISCUSSION AND CONCLUSION

In this paper, we propose to use non-fixed and asymmetrical margin (NAAM) in the prediction of HSI and DJIA. From the experiments, we make the following observations and discussions:

1. In comparison NAAM with the case of NASM which just uses the standard deviation, we find that adding the momentum to set the margin we can reduce the the downside risk. We may also improve the accuracy of our prediction significantly by selecting a long term EMA for higher volatility financial data.
2. The SVR algorithm using NASM and NAAM outperforms the AR model with same order.
3. The third observation is that applying NASM and NAAM in SVR, we will obtain better results than using RBF network and, in practice, SVR needs to determine fewer parameters than RBF network.

#### 5. ACKNOWLEDGEMENT

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administration Region, China.

#### 6. REFERENCES

- [1] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines (version 2.31), 2001.
- [2] S. Mukherjee, E. Osuna, and F. Girosi. Nonlinear prediction of chaotic time series using support vector machines. In J. Principe, L. Giles, N. Morgan, and E. Wilson, editors, *IEEE Workshop on Neural Networks for Signal Processing VII*, page 511. IEEE Press, 1997.
- [3] K. R. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik. Predicting time series with support vector machines. In *ICANN*, pages 999–1004. In W. Gerstner, A. Germond, M. Hasler, and J. D. Nicoud, editors, *Artificial Neural Networks - ICANN'97*, Springer, 1997.
- [4] Ian T. Nabney. *Netlab: Algorithms for Pattern Recognition*. Springer, London; New York, 2002.
- [5] A. Smola and B. Schölkopf. A tutorial on support vector regression. Technical Report NC2-TR-1998-030, NeuroCOLT2, 1998.
- [6] E. H. Tay and L. J. Cao. Application of support vector machines to financial time series forecasting. *Omega*, 29:309–317, 2001.
- [7] T.B. Trafalis and H. Ince. Support vector machine for regression and applications to financial forecasting. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN2000)*, volume 6, pages 348–353. IEEE, 2000.
- [8] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
- [9] V. N. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.
- [10] Haiqin Yang, Laiwan Chan, and Irwin King. Support vector machine regression for volatile stock market prediction. In *Third International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'02)*, pages 391–396. Springer, 2002.