JETEM

Journal of Economic Theory and Econometrics

The Journal of the Korean Econometric Society

(계 량 경 제 학 보)

June 2009

Volume 20. Issue 2. pp.65-81

Predicting Exchange Rates by Support Vector Regression*

Shiyi Chen† · Kiho Jeong ‡

Abstract In recent years, support vector regression (SVR), a novel neural network technique, has been successfully used for financial forecasting. This paper deals with the application of SVR in exchange rate forecasting. Based on SVR, a nonparametric autoregressive (AR) model is applied to forecasting the daily exchange rates of two currencies (South Korea Won and Singapore Dollar) against the US dollar. The empirical results show that under various forecasting horizons, SVR performs better than the random walk model, parametric AR model and the nonparametric AR model estimated by neural network, based on the criteria of two evaluation metrics and three encompassing tests. No structured way being available to choose the free parameters of SVR, the sensitivity of the forecasting performance is also examined to the free parameters

Keywords Support Vector Regression, nonparametric AR model, forecasting exchange rates

JEL Classification C45, G17

---

* The first author is sponsored by Shanghai Leading Academic Discipline Project (No. B101) and State Innovative Institute of Project 985 at Fudan University.

† Associate Professor, China Center for Economic Studies (CCES), Department of Economics, Fudan University, Shanghai 200433, China; (Email) shiyichen@fudan.edu.ch; (TEL) 86-21-65642050

‡ (Corresponding author) Professor, School of Economics and Trade, Kyungpook National University, Korea, E-mail address: khjeong@mail.knu.ac.kr

Received September 2, 2007, Revised December 3, 2008, Accepted April 1, 2009

1. Introduction

Exchange rate forecasting is one of the most challenging applications of modern time series forecasting. It has been widely accepted that exchange rates are inherently noisy, non-stationary and nonlinear. In recent years, Neural Network model (NN) has been successfully used for forecasting financial time series including exchange rates; for recent work, see Kuan and Liu (1995), Leung et al. (2000) and Chen and Leung (2005). The main appeal of NNs is their flexibility in approximating any non-linear function arbitrarily well without a priori assumptions about the properties of the data; see Hornik et al. (1989) for a discussion of the NN's universal approximation property. However, NN suffers from a number of weaknesses including the need for a large number of controlling parameters, difficulty in obtaining a global solution and the danger of over-fitting (Tay and Cao, 2001). The over-fitting problem is a consequence of the optimization algorithms used for parameter selection and the statistical measures used to select the best model.

Recently, a novel neural network algorithm, called support vector machine (SVM), was developed by Vapnik and his co-workers (1995, 1997) and is gaining popularity due to many attractive features. While the traditional NN implements the empirical risk minimization (ERM) principle, SVM implements the structural risk minimization (SRM) principle which seeks to minimize an upper bound on the Vapnik-Chervonenkis (VC) dimension (generalization error), as opposed to ERM that minimizes the training error on the in-sample estimating data; refer to Gunn (1998) for a good introduction to SVM and related concepts. Based on SRM principle, SVM achieves a balance between the training error and generalization error, leading to better forecasting performance than traditional NN. Selecting the best model in SVM is equivalent to solving a quadratic programming, which gives SVM another merit of a unique global solution.[1]

SVM was originally developed for classification problems (SVC) and then extended to regression problems (SVR). Both SVC and SVR have been successfully applied to financial variables classification (e.g., Härdle *et al.*, 2005, 2006; Chen *et al.*, 2006; Lee *et al.*, 2006) and financial time series forecasting (see Trafalis and Ince, 2000; Cao and Tay, 2001; Yang *et al.*, 2002, to name a few). However, the literatures of SVR above mostly focus on stock market prediction and, to the best of our knowledge, none in forecasting exchange rates. In this paper, we apply SVR to forecasting the daily exchange rates of two currencies, South Korea Won (KRW) and Singapore Dollar (SGD) against the US dollar in the framework of a nonparametric univariate Autoregressive (AR) model. The data period is from January 2, 2003 to December 29, 2006. To examine the sensitivity of SVR results to kernel

[1] See Haykin (1999) pp344 for more explanation on this issue and Deng and Tian (1994) for prove of the unique solution of SVR

functions and free parameters, we experiment with three kernel functions (namely, linear, polynomial, Gaussian) and also with several values of the free parameters. The forecasting performance between SVR and three competing approaches, random walk model (RW), parametric AR model and nonparametric AR model estimated by NN, are compared using two evaluation metrics and three encompassing tests at 1-day, 1-week and 1-month forecasting horizons[2].

This paper is organized as follows. Section 2 introduces the theory of SVR briefly. Section 3 describes the real data, specifies the empirical modeling and forecasting scheme. Section 4 evaluates the forecasting performance of SVR for KRW and SGD exchange rates. The conclusion is presented in Section 5.

2. Support Vector Regression (SVR)

SVR originates from Vapnik's statistical learning theory (Vapnik, 1995, 1997), which has the design of a feedforward network with an input layer, a single hidden layer of nonlinear units and an output layer (Haykin, 1999). Unlike the traditional NN, SVR estimates a function by nonlinearly mapping the input space into a high dimensional feature space and then running the linear regression in the output space (see Figure 1). Thus, the linear regression in the output space corresponds to a nonlinear regression in the low dimensional input space.

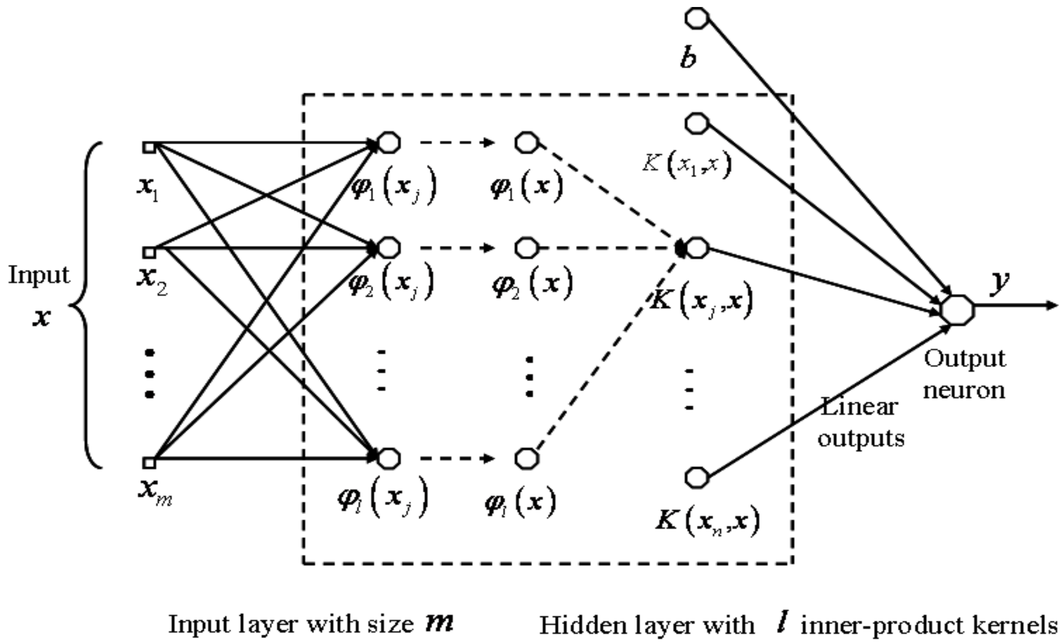

Figure 1 Architecture of Support Vector Machines for Regression (SVR)

Given a set of data points $\{(x_i, y_i)\}_i^n$ where $x_i \in R^m$ forms the input vector with

[2] Equivalent to be 1-, 5-, 20-day-ahead forecast.

m-dimension and $y_i \in R^1$ is a scalar output, SVR approximates the unknown function by $f(x) = w'\phi(x) + b = \sum_{i=1}^{l} w_i \varphi_i(x) + b$, where $\varphi(x) = (\varphi_1(x), \cdots, \varphi_l(x))'$ is the nonlinear transfer function from the input layer to the feature layer with the dimensions l;. $w = (w_1, \cdots, w_l)'$ is a set of linear weights that connect the feature space to the output space and b is the threshold.

To estimate the coefficients w and b, SVR formulates the regression problem by using a specific loss function called the ε-insensitive loss function $L_\varepsilon(x, y) = \max\{0, |y - f(x)| - \varepsilon\}$, which ignores errors smaller than a certain threshold $\varepsilon > 0$. The regression problem becomes

$$\min \quad C\frac{1}{n}\sum_{i=1}^{n} L_\varepsilon(x_i, y_i) + \frac{1}{2}\|w\|^2 \tag{1}$$

where the first term $C\frac{1}{n}\sum_{i=1}^{n} L_\varepsilon(x_i, y_i)$ is the empirical error measured by the ε-insensitive loss function; the second term $\frac{1}{2}\|w\|^2$ indicates the Euclidean norm of the weight vector w; C controls the penalizing extent on the sample data which exceeds ε. Both C and ε are free parameters which must be selected by the user.

In SVR, equation (1) is transformed to the primal problem given by equation (2) by introducing the positive slack variables ξ_i and ξ_i^* as follows:

$$\min \quad \frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{2}$$

$$s.t. \quad w'\varphi(x) + b - y_i \le \varepsilon + \xi_i \quad i = 1, 2, \cdots, n$$

$$y_i - w'\varphi(x) - b \le \varepsilon + \xi_i^* \quad i = 1, 2, \cdots, n$$

$$\xi_i \ge 0, \xi_i^* \ge 0 \quad i = 1, 2, \cdots, n$$

The formulation of the cost function in equation (2) of primal problem is in perfect accord with the principle of structural risk minimization, which is illustrated in Figure 2 (in which the dark circles are data points extracted as support vectors). In equation (2), the first term indicates the Euclidean norm of the weight vector w and measures the function flatness, the minimization of which is related to the maximization of the margin of separation ($2/\|w\|$), i.e., maximizing the generalization ability. The second term represents the empirical risk loss determined by the ε-insensitive loss function and is similar to the sum of residual squares in the objective function of MLE and NN. Finally, SVR obtains the tradeoff between the two terms; as a result, it not only captures the historical data well but forecasts the future data excellent. It is the special design of minimizing the structural risk that endows SVR with the strongest forecasting ability among all methods. Evgeniou et al. (2002) also denoted that minimization of an empirical error only is both ill-posed and not necessarily leading to models with good predictive capabilities, thus, one needs to minimize a structural risk.

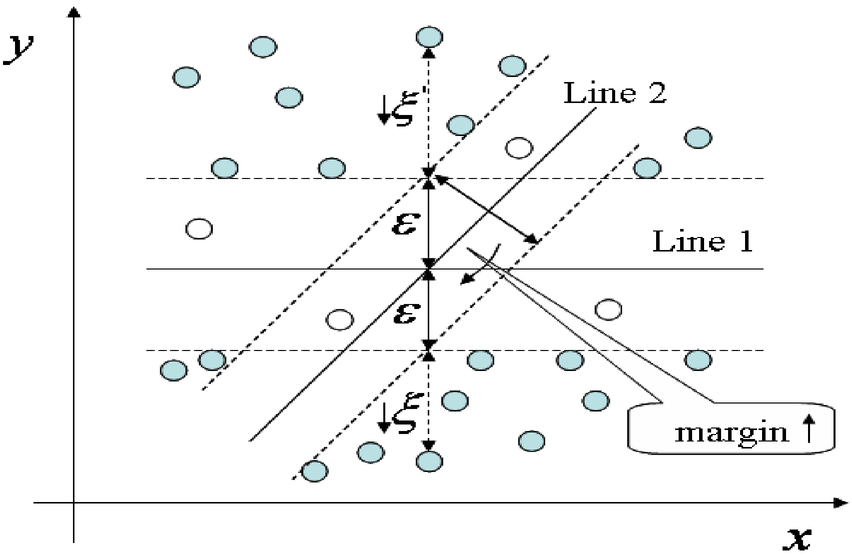

Figure 2 Principle of Structural Risk Minimization of SVR

The corresponding dual problem of the SVR can be derived from the primal problem by using the Karush-Kuhn-Tucker conditions as follows: [3]

$$\min \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\alpha_i^*-\alpha_i\right)\left(\alpha_j^*-\alpha_j\right)K\left(x_i\cdot x_j\right)+\varepsilon\sum_{j=1}^{n}\left(\alpha_j^*+\alpha_j\right)-\sum_{j=1}^{n}y_j\left(\alpha_j^*-\alpha_j\right) \quad (3)$$

$$s.t. \quad \sum_{j=1}^{n}\left(\alpha_j-\alpha_j^*\right)=0$$

$$0\le \alpha_j,\alpha_j^*\le C \quad i,j=1,2,\cdots,n$$

where α and α^* are the Lagrange multipliers and $K\left(x_j,x\right)=\varphi\left(x_j\right)'\varphi\left(x\right)$ is called as the kernel function. In fact, the SVR theory considers only the form of $K\left(x_j,x\right)$ in the feature space without specifying $\varphi\left(x\right)$ explicitly and without computing all corresponding inner products. Therefore, the kernel function

[3] The dual problem can be solved more easily than the primal problem (Scholkopf and Smola, 2001; Deng and Tian, 2004).

becomes the crucial part of SVR and greatly reduces the computational complexity of SVR. This paper experiments with three different kernels to investigate the sensitivity of kernel types.

Linear: $K\left(x_j, x\right) = x_j\,'x$

Polynomial: $K\left(x_j, x\right) = \left(x_j\,'x + 1\right)^d$

Gaussian: $K\left(x_j, x\right) = \exp\left(\frac{-\left\|x - x_j\right\|^2}{2\sigma^2}\right)$

The solution of the dual problem given by equation (3) can be used to obtain the solution of primal problem:

$$w = \sum_{j=1}^{n}\left(\alpha_j^* - \alpha_j\right)\varphi\left(x_j\right)$$

$$b = \frac{1}{2}\left[y_s + y_t - \left(\sum_{j=1}^{n}\left(\alpha_j^* - \alpha_j\right)K\left(x_j \cdot x_s\right) + \sum_{j=1}^{n}\left(\alpha_j^* - \alpha_j\right)K\left(x_j \cdot x_t\right)\right)\right]$$

Finally, the decision function can be obtained as follows,

$$\begin{aligned} f\left(x\right) &= w'\varphi\left(x\right) + b \\ &= \sum_{j=1}^{n}\left(\alpha_j^* - \alpha_j\right)\varphi\left(x_j\right)'\varphi\left(x\right) + b \\ &= \sum_{j=1}^{n}\left(\alpha_j^* - \alpha_j\right)K\left(x_j, x\right) + b \end{aligned}$$

3. Empirical modeling and forecasting scheme

3.1. Data Description

Exchange rates used for the empirical analysis in the current study are the daily nominal bilateral exchange rates. The data were obtained from a database provided by Policy Analysis Computing and Information Facility in Commerce (PACIFIC) at University of British Columbia, which contains the closing rates for a total of 81 currencies and commodities. T wo currencies were selected for our study: South Korea Won (KRW) and Singapore Dollar (SGD) against the U.S. dollar for the period of January 2, 2003 to December 29, 2006. The ADF unit root tests reveal that both variables are integrated of order one. To avoid the issue of possible nonstationarity, this paper considers the rates in first-difference logarithm terms, y_t. It is also called returns of the level of exchange rates, I_t. The transformation formula is as below:

$$y_t = 100 \times \left(\log I_{t+1} - \log I_t \right)$$

3.2. Empirical Modeling

In terms of theoretical view, the daily exchange rate might follow a random-walk process and thus it is difficult to forecast such a random-walk process (Taylor (2003) for the exchange rate, and Hamilton (1996) and Sarno, Thornton, Valente (2005) for the daily Federal funds rate). From this point of view, the paper incorporates the simple random walk model (RW) without drift, specified as below, into the comparison of forecasting performance.

$$y_t = y_{t-1} + \varepsilon_t$$

The linear univariate Autoregressive Integrated (AR) model estimated by MLE is also the benchmark described as below:

$$y_t = \mu + \phi_1 y_{t-1} + \cdots + \phi_P y_{t-p} + \varepsilon_t$$

According to Akaike information criterion, the appropriate lag lengths of AR model for KRW and SGD returns are 1 and 3, respectively.

As the basic forecasting framework, the AR model estimated by SVR and NN (ARSVR and ARNN for short in the remaining part of this paper) has the nonlinear form different with the linear one above, which is described as follow:

$$y_t = f(y_{t-1}, \cdots, y_{t-p}) + \varepsilon_t .$$

The lag orders are set the same as the linear ARI model for the purpose of comparison.

For SVR approach, we choose the fixed value of $\sigma^2 = 0.2$ for Gaussian kernels and $d = 2$ for polynomial kernels which are determined by the sensitivity analysis and are not provided here for saving space. The choice of the values of two free parameters are C=0.1, ε=0.3 for both series will also be determined according to the sensitivity analysis which is discussed in detail in Section 4. The benchmark neural network is feedforward back-propagation network. We specify it with the following architectures: one nonlinear hidden layer with 4 neurons in which tangent-sigmoid differentiable transfer function is used to generate their output, and one linear output layer with 1 neuron. The fast training Levenberg-Marquardt algorithm is chosen and designed into the training function. The value of learning rate parameter used in this training process is 0.05. These specification and choices are quite standard in the neural network literatures.

3.3. Forecasting Scheme

In this paper, a recursive forecasting scheme is employed. The estimating and forecasting process is carried out recursively by updating the sample with one observation each time. The notations are as follows. The total number of returns y_t is denoted as T and the number of observations used for the first in-sample estimation is T_1. Then, $T - T_1$ observations are retained as a hold-out sample. Let the actual return at period $t + j$ and the j-step-ahead forecast of the returns made at period t be written as y_{t+j} and $\hat{y}_{t+j}$, respectively. Then, we can write

$$\hat{y}_{t+j|t} = E\left(y_{t+j} \mid y_t, y_{t-1}, \ldots, y_1\right) \tag{4}$$

so that the j-step-ahead forecast of the series made at time t is the expected value of the series j periods in the future, given all information available at time t. In equation (4), $t = T_1, \ldots, T - j$ and $j = 1, 5, 20$. The forecast horizon is fixed at j steps ahead, and the starting point t is varied. Therefore, we can estimate and forecast the model for $n = T - j - T_1 + 1$ times.

In this paper, we set $n = 60$ and obtain sixty 1-, 5- and 20-day-ahead forecasts which are employed to evaluate the out-of-sample forecasting performance. The recursive estimation is carried out from the former 923 observations through the former 982 observations of the exchange rate returns and the sixty 1-, 5- and 20-day-ahead forecasts correspond to the 924th, 928th and 943rd actual return through 983rd, 987th and the last returns data.

3.4. Evaluation Metrics and Encompassing Tests

To compare forecasting performance between SVR and two competing methods, we use two approaches: evaluation metric approach and encompassing test approach. The first approach tells us which method is the dominant one in forecasting the error magnitude and turning points; particularly, the correctness of turning points forecasts may yield important information for financial decisions such as market timing. The results should be tested statistically which can be done using the second approach, encompassing tests.

For the evaluation metric approach, two statistical criteria are considered: normalized mean square error (NMSE) and correct sign predictions (sign) (Pesaran and Timmerman, 1990; Moosa, 2000). The NMSE measures the magnitude of the forecasting error and the sign metrics estimates the correctness in predicted directions. Their formulas are

$$NMSE(\%) = 100 \times \frac{MSE}{\mathrm{var}(y)} = 100 \times \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 / n}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2 / (n-1)}$$

$$sign(\%) = \frac{100}{n}\sum_{i=1}^{n} a_i \text{ , where } a_i = \begin{cases} 1 & (y_{i+1} - y_i)\left(\hat{y}_{i+1} - \hat{y}_i\right) > 0 \\ 0 & otherwise \end{cases}$$

The second approach, encompassing tests, can be described as follows. When a forecast does not carry any additional information compared to a target forecast, it is said to be encompassed (Sollis, 2005). In this study, the null hypothesis of encompassing test is that the SVR based forecasts encompass those by the benchmark methods. We consider three encompassing tests: OE statistic by Ericsson (1992), HLN test by Harvey *et al.* (1998), and CM test by Clark and McCracken (2001). Assume that e_{1t} and e_{2t} are sets of forecast errors of SVR and the benchmark method, respectively, and that there are n of these forecasts. Three statistics and their distribution can be expressed as

$$OE = \sqrt{n}\frac{n^{-1}\sum_{t=1}^{n} e_{1t}(e_{1t} - e_{2t})}{\sqrt{n^{-1}\sum_{t=1}^{n}(e_{1t} - e_{2t})^2 n^{-1}\sum_{t=1}^{n} e_{1t}^2 - \bar{c}^2}} \sim N(0,1)$$

$$HLN = \frac{\bar{c}}{\sqrt{\mathrm{var}(c_t)}} = \sqrt{n}\frac{n^{-1}\sum_{t=1}^{n}(e_{1t}^2 - e_{1t}e_{2t})}{\sqrt{n^{-1}\sum_{t=1}^{n}(e_{1t}^2 - e_{1t}e_{2t})^2 - \bar{c}^2}} \sim N(0,1)$$

$$CM = \frac{n\bar{c}}{MSE_2} = \frac{\sum_{t=1}^{n}(e_{1t}^2 - e_{1t}e_{2t})}{n^{-1}\sum_{t=1}^{n} e_{2t}^2}$$

where $c_t = e_{1t}(e_{1t} - e_{2t})$, $\bar{c} = n^{-1}\sum_{t=1}^{n} c_t$, and MSE_2 is the mean squared forecasting error of the competing method. The CM statistic does not follow the standard normal distribution, their critical values at different significance level are given in Clark and McCracken (2001).

4. Forecasting Performance Evaluation

In this section, we investigate the forecasting performance of all candidates for two financial returns: KRW/USD and SGD/USD exchange rates.

4.1. Comparing SVR Based Forecasts with the Candidates

Table 1 presents the forecasting performance of ARSVR and three competing approaches at 1-, 5- and 20-day horizons, in which the upper panel displays the results of two quantitative metrics (NMSE and sign) and the lower panel shows the corresponding average rankings.

Table 1 Measures of forecasting performance of various methods for exchange rate returns

samples	horizons	metrics	RW	AR	ARNN	ARSVR		
						linear	polynomial	Gaussian
KRW/USD	1	NMSE	215.72	103.20	100.73	100.59	100.80	100.07
		sign	50.85	47.46	62.71	62.71	59.32	61.02
	5	NMSE	191.24	105.53	104.21	102.42	102.97	100.72
		sign	55.93	42.37	66.10	64.41	67.80	67.80
	20	NMSE	195.14	100.83	100.43	100.04	100.06	100.37
		sign	50.85	55.93	61.02	59.32	61.02	62.71
SGD/USD	1	NMSE	168.94	104.59	102.71	102.66	100.06	103.31
		sign	52.54	57.63	54.24	55.93	57.63	61.02
	5	NMSE	189.46	103.91	102.04	102.36	101.98	102.38
		sign	52.54	59.32	55.93	55.93	61.02	59.32
	20	NMSE	207.32	106.97	105.96	105.05	104.43	104.24
		sign	50.85	49.15	61.02	61.02	61.02	57.63
Rankings Averaged Across Two Samples	1	NMSE	6	5	4	1	2	2
		sign	6	5	3	2	3	1
	5	NMSE	6	5	4	2	1	2
		sign	6	4	3	4	1	2
	20	NMSE	6	5	4	1	1	1
		sign	5	5	1	3	1	3
	average	NMSE	6	5	4	1	1	3
		sign	6	5	3	4	1	2

Note: "average" is the average ranking over two samples and over all the forecast horizons.

First, we compare the forecasting performance between ARSVR and AR. Obviously, the ARSVR consistently behaves better than AR estimated by MLE in forecasting exchange rates magnitude in terms of NMSE metric. For example, the values of NMSE in 1-period-ahead forecast for KRW are about 100% for three kinds of ARSVR but reach to 103.2% for AR. The sign metric reveals that ARSVR also has a superior ability to simple AR in forecasting the turning points just with two exceptions: 1- and 5-day-ahead direction forecast for SGD, in which ARSVR with linear kernel are all inferior to AR.

Then, we evaluate the forecasting ability between ARSVR and ARNN. Again, based on the NMSE rankings averaged across two returns, ARSVR with different kernels generally produces more accurate post-sample forecasts than ARNN at different horizons in forecasting the returns magnitude. However, according to sign metric, ARNN also predicts the directions of change not bad. It even dominates some kinds of ARSVR in 1- and 5-day-ahead forecast for KRW and 20-day-horizon for both series. Of course, in other cases, ARSVR still performs better than ARNN in forecasting the turning points.

As Table 1 denoted, the simple random walk model (RW) always performs the worst in any case. In a word, according to the average rankings of two samples and three horizons, three kinds of ARSVR are top three methods on NMSE grounds. This is followed by ARNN, while RW is ranked last. The situation is the same as that according to sign metric except that ARSVR with linear kernel and ARNN change the ranking turn.

Table 2 Encompassing tests in forecasting exchange rates

Horizons	Statistic	RW vs. ARSVR		AR vs. ARSVR		ARNN vs. ARSVR	
		KRW/USD	SGD/USD	KRW/USD	SGD/USD	KRW/USD	SGD/USD
1	OE	0.0095	0.0992	0.9367	−1.5898	0.9512	−1.6359
	HLN	0.0154	0.0898	0.0661	−0.2423	0.1187	−0.1964
	CM	0.3969	0.1833	0.7055	−0.1032	0.3698	−0.6332
5	OE	−0.0096	0.0889	1.8110*	−1.7559*	1.1100	1.9058*
	HLN	−0.0092	0.1090	0.0272	−0.1459	0.0287	−0.0101
	CM	−0.2390	0.8511	0.3951	−0.4681	0.3708	−0.0092
20	OE	0.0219	−0.0359	−1.2880	1.7346*	−1.0380	−0.5726
	HLN	0.0388	0.0544	0.1241	−0.2032	0.1225	0.1655
	CM	1.2432**	0.7142	1.6603**	−0.6288	1.4414**	0.1417

Note: * and ** denote significance at the 10% and 5% levels, respectively.

The encompassing tests for the forecasts of exchange rate returns are displayed in Table 2, in which the values of ARSVR adopt the polynomial kernel for SGD and Gaussian kernel for KRW, the best one among three kernels. As for RW and ARSVR, the hypothesis that ARSVR forecasts encompass the random walk model is only rejected by the CM test at-20-horizon for KRW. In the case of AR vs. ARSVR, the null hypothesis that ARSVR forecasts outperform the AR is only weakly rejected by OE statistic in 5-day-ahead forecast for both series, 20-day-ahead for SGD and rejected by CM at 20-day-horizon for KRW. In other situations, the null hypothesis of encompassing test is not rejected by any of the tests. For ARNN vs. ARSVR, the null hypothesis is not rejected by HLN statistic but is only weakly rejected by OE at 5-day-horizon for SGD and rejected by CM test for KRW at 20-day-horizon. In a word, the statistical encompassing tests also indicate that ARSVR significantly outperforms three competing methods in forecasts, which further justifies the conclusions obtained in Table 1 as correct.

We plot the actual and 1-day-ahead forecasting returns by AR, ARNN and ARSVR in Figure 3, in which ARSVR uses the best kernel, i.e. polynomial kernel for SGD and Gaussian kernel for KRW - described before. The forecasts by random walk model were not drawn in this figure because they were just one lag of actual returns of exchange rates. The 60 one-day-ahead forecasts correspond to the out-of-sample period from September 6, 2006 through November 30, 2006 for both exchange rate returns. As anticipated, ARSVR captures the actual returns more accurately.

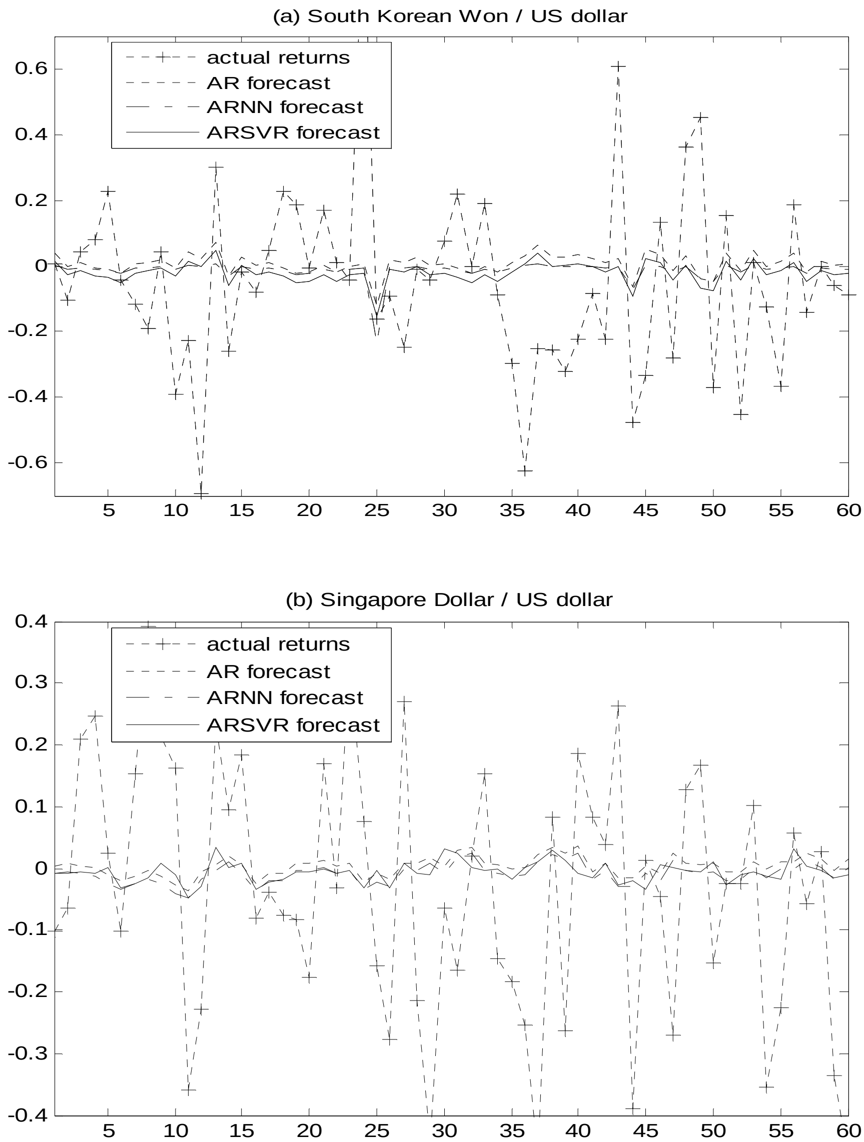

Figure 3 Actual and One-Step-Ahead Forecasted Returns of Exchange Rates

4.2. Investigating the Sensitivity of SVR Forecasts

In fact, Table 1 also describes the effect of kernel functions on the forecasting performance of SVR. Both metrics generally select the polynomial kernel as the best one for SGD and the Gaussian kernel for KRW. However, the linear kernel also behaves well in some situations (turning points predicting at 1-day horizon for KRW and 20-day horizon for SGD, and 20-day-ahead forecast of error magnitude for KRW). No clear winner emerges and the preferred kernel for each case is different. Furthermore, the three kinds of SVR almost have similar performance each other and behave better than the competing methods.

Figure 4 describes the effect of the values of C and ε (epsilon) in the case of the one-day-ahead forecasts. The polynomial kernel is exemplified for SGD and the Gaussian kernel for KRW. Figure 4 (a) illustrates the NMSE (implying the generalization error) of SVR versus C. The regularization parameter C varies between values of 0.001 and 1000 for both returns, with ε being arbitrarily chosen to be 0.3. Apparently, C is found to have almost no impact on the generalization error, NMSE, within a quite broad range. Figure 4 (b) depicts the NMSE of SVR versus ε. Parameter ε varies between values of 0.01 and 0.8 for two samples, with C being fixed to be 0.1 both. As illustrated in Figure 4 (b), the generalization error was not influenced greatly by ε in some intervals. Of course, the insensitivity interval of ε varies from the different samples. In this study, the better values of free parameters are C=0.1 and ε=0.3 for both samples. Thus, the researchers can easily choose the appropriate values of the free parameters according to the sensitivity analysis.

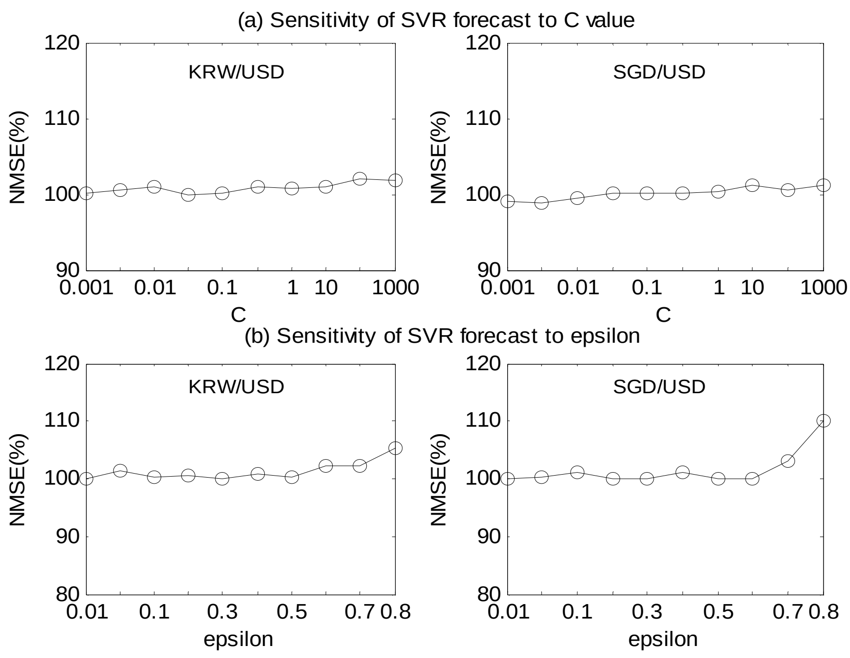

Figure 4 Sensitivity of ARSVR Forecasts with respect to Free Parameters

5. Concluding Remarks

In this paper we have investigated the post-sample forecasting ability of SVR for daily exchange rates of two currencies (South Korea Won and Singapore Dollar) against the US dollar for the period from January 2, 2003 to December 29, 2006. The forecasting performance of SVR at different horizons has been compared with a linear random walk and AR model and a nonlinear ARNN by using two evaluation metrics and three encompassing tests. NMSE metric reveals that ARSVR with linear, polynomial and Gaussian kernel are top three in forecasting the exchange rates magnitude. This is followed by ARNN, while AR is ranked last. In terms of sign metric, ARNN defeats AR and linear ARSVR but is still inferior to other two kinds of ARSVR. Empirical analysis is in favor of the theoretical advantage of SVR for forecast, and the encompassing tests significantly confirm all these conclusions in statistical view. The sensitivity to kernel type and two free parameters of SVR results is also examined. It is found that the kernel types can affect the SVR's forecasting performance but the values of free parameters are not very sensitive to the generalization error of SVR. Thus future research topic is either how to choose a suitable kernel function or to find a method insensitive to the choice of kernels. Furthermore, the exchange rate forecasting performance of nonparametric SVR model can also be investigated as opposed to the newly developed parametric models such as Exponential Smoothing Transitional Autoregressive Model (ESTAR) proposed by Taylor and Sarno (1998) and Taylor and Peel (2000). Finally, statistical comparison of forecasting results by bootstrap or other hypothesis tests proposed by, say, Diebold and Mariano (1995), is worth studying in the future for the nonparametric SVR approach.

References

Database of Exchange Rates: http://pacific.commerce.ubc.ca/xr, Policy Analysis Computing and Information Facility In Commerce (PACIFIC) at University of British Columbia

Adya, M. & F., Collopy (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*, 17: 481-495.

Ashok, K.N. & A. Mitra (2002). Forecasting daily foreign exchange rates using genetically optimized neural networks. *Journal of Forecasting*. 21, 501-511.

Balkin, S. D., & J.K., Ord (2000). Automatic neural network modelling for Univariate time series. *International Journal of Forecasting*, 16, 509–515.

Box, GEP, GM Jenkins & GC, Reinsel (1994). *Time Series Analysis: Forecasting and Control*, Englewood

Cliffs, Prentice Hall, New Jersey.

Cao, L. & F., Tay (2001). Financial forecasting using support vector machines. *Neural Computation and Application*, 10, 184-192.

Chen, A. and M. Leung (2005), "Performance Evaluation of Neural Network Architectures: The Case of Predicting Foreign Exchange Correlations," *Journal of Forecasting*, Vol 24, 403-420.

Chen, S.Y., W.K. Härdle, R.A. Moro (2006) Estimation of Default probabilities with Support Vector Machines, A discussion paper published by the SFB 649, Humboldt University with the number 2006-077 (http://sfb649.wiwi.hu-berlin.de/fedc/discussionPapers_de.php).

Clark, T. & M. McCracken (2001). Tests of Equal Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics*, 105, 85-110.

Darbellay, G. A., & M. Slama (2000). Forecasting the short-term demand for electricity? Do neural networks stand a better chance? *International Journal of Forecasting*, 16, 71–83.

Deng, N.Y. & Y.J. Tian (2004). *New Methods in Data Mining: Support Vector Machine*, Science Press, Beijing.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy" *Journal of Business and Economic Statistics*, 13, No.3, 253 – 263.

Ericsson, N. (1992). Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: an Exposition, Extensions, and Illustration. *Journal of Policy Modelling*, 4,465-495.

Evgeniou, T., T. Poggio, M. Pontil & A. Verri (2002). Regularization and statistical learning theory for data analysis, *Computational Statistics & Data Analysis*, 38(4), 421-432.

Ghiassia, M., H., Saidaneb& D.K., Zimbrac (2005). A dynamic artificial neural network model for forecasting time series events, *International Journal of Forecasting*, 21, 341–362.

Gunn, S. (1998), "Support Vector Machines for Classification and Regression," Technical Report, University of Southampton.

Hamilton, JD. (1996), "The Daily Market for Federal Funds" *Journal of Political Economy*, 104, No.1, 26 – 56

Hamilton, JD. (1997). *Time Series Analysis*, Princeton University Press.

Härdle, W.K., R.A. Moro & D. Schäfer (2005). Predicting Bankruptcy with Support Vector Machines, *Statistical Tools for Finance and Insurance*, Springer Verlag, Berlin.

Härdle, W.K., R.A. Moro & D. Schäfer (2006). Graphical Data Representation in Bankruptcy Analysis, *Handbook for Data Visualization*, Springer Verlag, Berlin.

Harvey, D., S. Leybourne & P. Newbold (1998). Tests for Forecast Encompassing. *Journal of Business and Economic Statistics*, 16, 254-259.

Haykin, S.(1999). *Neural Networks: a comprehensive foundations*. Prentice Hall, New Jersey.

Heravi, S., D. R., Osborn, & C. R., Birchenhall (2004). Linear versus neural network forecasts for European industrial production series. *International Journal of Forecasting*, 20, 435–446.

Hippert, H.S., D.W., Bunn & R.C., Souza (2005). Large neural networks for electricity load forecasting: Are they overfitted? *International Journal of Forecasting*, 21(3), 425-434.

Hornik, K., M. Stinchcombe, and H. White (1989), "Multi-layer Feedforward Networks Are Universal Approximators," *Neural Networks*, Vol 2, 359-366.

Kuan, CM & T. Liu (1995). Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*, 10, 347-364.

Kamruzzaman, J. & R. Sarker (2004). ANN-Based Forecasting of Foreign Currency Exchange Rate. *Neural Information Processing: Letters and Reviews*, 3(2), May.

Lee, T.S., C.C. Chiu, Y.C. Chou & C.J. Lu (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines, *Computational Statistics & Data Analysis*, 50(4), 1113-1130.

Leung, M., A. Chen and H. Daouk (2000), "Forecasting Exchange Rates Using General Regression Neural Networks," *Computers & Operations Research*, Vol 27, 1093-1110.

Lisi, F. & R. A. Schiavo (1999). A comparison between neural networks and chaotic models for exchange rate prediction, *Computational Statistics & Data Analysis*, 30(1), 87-102.

Moosa, I., (2000). *Exchange Rate Forecasting: Techniques and Applications*, Macmillan Press LTD, Lonton.

Olson, D., & C., Mossmanb (2003). Neural network forecasts of Canadian stock returns using accounting ratios. *International Journal of Forecasting*, 19, 453-465.

Pesaran, M. & A. Timmerman (1990). The Statistical and Economic Significance of the Predictability of Excess Returns on Common Stocks, Department of Applied Economics, University of Cambridge, Working Paper No. 9022.

Priestley, MB. (1988). *Nonlinear and Non-stationary Time Series Analysis*. Academic Press: London.

Sarno, Lucio, Daniel Thornton, and Giogio Valente (2005), "Federal Funds Rate Prediction" *Journal of Money, Credit and Banking* 37, No. 3, 449 – 471.

Scholkopf, B. & A. Smola (2001). *Learning with Kernels*, Cambridge, MA: MIT Press.

Sollis R. (2005). Predicting Returns and Volatility with Macroeconomic Variables: Evidence from Tests of Encompassing. *Journal of Forecasting*. 24, 221-231.

Tay, F. and L. Cao (2001), "Application of Support Vector Machines in Financial Time Series Forecasting," *Omega*, Vol 29, 309-317.

Taylor, Mark and Lucio Sarno (1998), "The Behavior of Real Exchange Rates During the Post Bretton Woods Period" *Journal of International Economics*, 46, 261 – 312.

Taylor, Mark and David Peel (2000), "Non-linear Adjustment, Long Run Equilibrium and Exchange Rate Fundamentals" *Journal of International Money and Finance*, 19, 33 – 53.

Taylor Mark (2003), "Why is it so Difficult to Beat the Random Walk Forecast of Exchange Rates?" *Journal of International Economics*, 60, 85 – 107.

Tian, J., M. Juhola & T. Grönfors (1997). AR parameter estimation by a feedback neural network, *Computational Statistics & Data Analysis*, 25(1), 17-24.

Tkacz, G. (2001). Neural network forecasting of Canadian GDP growth. *International Journal of Forecasting*. 17, 57–69.

Trafalis, T. & H. Ince (2000). Support Vector Machine for Regression and Applications to Financial Forecasting. *International Joint Conference on Neural Networks*, 348-353.

Vapnik, VN. (1995). *The Nature of Statistical Learning Theory*, Springer, New York.

Vapnik, VN. (1997). *Statistical Learning Theory*, Wiley, New York.

Yang, H, Chan L. & I. King (2002). Support vector machine regression for volatile stock market prediction. *Proceedings of the Third International Conference on Intelligent Data Engineering and Automated Learning*. 391-396.