Forecasting Exchange Rates using Support Vector Machine Regression

SHIYI CHEN AND KIHO JEONG *

Kyungpook National University, the Republic of Korea

Abstract

This paper applies Support Vector Regression (SVR) to estimate and forecast nonlinear autoregressive integrated (ARI) model of the daily exchange rates of four currencies (Swiss Francs, Indian Rupees, South Korean Won and Philippines Pesos) against U.S. dollar. The forecasting abilities of SVR are compared with linear ARI model which is estimated by OLS. Sensitivity of SVR results are also examined to kernel type and other free parameters. Empirical findings are in favor of SVR. SVR method forecasts exchange rate level better than linear ARI model and also has superior ability in forecasting the exchange rates direction in short test phase but has similar performance with OLS when forecasting the turning points in long test phase.

KEY WORDS: Support Vector Machine Regression; OLS; ARI model; forecasting; exchange rates

BIOGRAPHY:

Shiyi Chen: Graduate Student in Ph.D. Course at the Department of Economics, Kyungpook National University, the Republic of Korea

Kiho Jeong: Professor of Econometrics, the Department of Economics, Kyungpook National University, the Republic of Korea

---

Correspondence to: Kiho Jeong, Department of Economics, Kyungpook National University, Daegu 702-701; the Republic of Korea; Email:khjeong@mail.knu.ac.kr;

TEL: +82 53-950-5416; FAX: +82 53-950-5407

1. Introduction

In this paper, univariate Autoregressive Integrated (ARI) model will be devised as a basic forecasting framework, the linear version of which is usually used to forecast nonstationary time series (Box et al., 1994; Niemira and Klein, 1994). The linear ARI model is generally estimated via ordinary least squares (OLS) method or maximum likelihood estimation (MLE) method assuming that the data are normally distributed. This paper concentrates on the nonlinear ARI model estimated by Support Vector Machine (SVM), a novel and computationally powerful class of supervised learning network developed by Vapnik (1995, 1997). SVM is theoretically the best in estimating finite sample without invoking a probabilistic distribution (Vapnik, 1995). The main advantage of SVM is its ability to minimize structural risk as opposed to empirical risk minimization as employed by OLS.

This paper applies SVR to estimate and forecast nonlinear ARI model of the daily exchange rates of four currencies (Swiss Francs, Indian Rupees, South Korean Won and Philippines Pesos) against U.S. dollar. To examine sensitivity of SVR results to kernel type and other free parameters, we experiment with three different kernel functions, namely, linear, polynomial, Gaussian, and also with several values of free parameters. Three quantitative measures are used to compare the forecasting abilities between SVR and linear ARI model. It is shown that SVR can outperform linear ARI in forecasting exchange rates.

2. Support Vector Machines for Regression

Given a training data set $\{(x_t, y_t)\}_{t=1}^{N}$, with vector inputs $x_t \in \mathbb{R}^{m_0}$ and a scalar outputs $y_t \in \mathbb{R}$, realizing some unknown function $g(x)$, we need to estimate a decision function $f(x)$ that approximates $g(x)$ as below.

$$f(x) = \sum_{j=1}^{m_1} w_j \varphi_j(x) + b = w^T \varphi(x) + b \qquad (1)$$

where $\varphi(x) = \left[\varphi_1(x), \cdots, \varphi_{m_1}(x)\right]^T$, $w = \left[w_1, \cdots, w_{m_1}\right]^T$

In order to get the decision function, coefficients w_i and b have to be estimated from data. Firstly, we define a linear ε-insensitive loss function, $L_\varepsilon(x, y, f(x))$, originally proposed by Vapnik (1995),

$$L_{\varepsilon}\left(x, y, f\left(x\right)\right) = \begin{cases} \left|y - f\left(x\right)\right| - \varepsilon & for \left|y - f\left(x\right)\right| \geq \varepsilon \\ 0 & otherwise \end{cases} \tag{2}$$

The primal constrained optimization problem of ε-SVR is obtained below.

$$\min_{w \in \mathbb{R}^N, \xi^{(')} \in \mathbb{R}^{2N}, b \in \mathbb{R}} \quad \Phi\left(w, b, \xi_t, \xi_t^{'}\right) = \frac{1}{2}\left\|w\right\|^2 + C\left(\frac{1}{N}\sum_{t=1}^{N}\left(\xi_t + \xi_t^{'}\right)\right) \tag{3}$$

$$s.t. \qquad w^T\varphi\left(x\right) + b - y_t \leq \varepsilon + \xi_t \quad t = 1, 2, \cdots, N \tag{4}$$

$$y_t - w^T\varphi\left(x\right) - b \leq \varepsilon + \xi_t^{'} \quad t = 1, 2, \cdots, N \tag{5}$$

$$\xi_t \geq 0, \xi_t^{'} \geq 0 \quad t = 1, 2, \cdots, N \tag{6}$$

The corresponding dual problem of nonlinear ε-SVR can be derived by using the Karush-Kuhn-Tucker conditions as follows:

$$\min_{\alpha^{(')} \in \mathbb{R}^{2N}} \quad \frac{1}{2}\sum_{s=1}^{N}\sum_{t=1}^{N}\left(\alpha_s^{'} - \alpha_s\right)\left(\alpha_t^{'} - \alpha_t\right)K\left(x_s \cdot x_t\right) + \varepsilon\sum_{t=1}^{N}\left(\alpha_t^{'} + \alpha_t\right) - \sum_{t=1}^{N} y_t\left(\alpha_t^{'} - \alpha_t\right) \tag{7}$$

$$s.t. \quad \sum_{t=1}^{N}\left(\alpha_t - \alpha_t^{'}\right) = 0 \tag{8}$$

$$0 \leq \alpha_t, \alpha_t^{'} \leq \frac{C}{N} \quad s, t = 1, 2, \cdots, N \tag{9}$$

where, $\alpha_t^{'}$ and α_t are the Lagrange multipliers.

Let's denote the optimizing solutions of the dual problem as $\overline{\alpha}_t^{(')}$. Then we can use them to obtain the solution of primal problem;

$$\overline{w} = \sum_{t=1}^{N}\left(\overline{\alpha}_t^{'} - \overline{\alpha}_t\right)\varphi\left(x_t\right) \tag{10}$$

$$\overline{b} = \frac{1}{2}\left[y_j + y_k - \left(\sum_{t=1}^{N}\left(\overline{\alpha}_t^{'} - \overline{\alpha}_t\right)K\left(\mathrm{x}_t \cdot \mathrm{x}_j\right) + \sum_{t=1}^{N}\left(\overline{\alpha}_t^{'} - \overline{\alpha}_t\right)K\left(\mathrm{x}_t \cdot \mathrm{x}_k\right)\right)\right] \tag{11}$$

, for $\overline{\alpha}_j, \overline{\alpha}_k^{'} \in \left(0, \frac{C}{N}\right)$

Substituting Eq. (10) and Eq. (11) into Eq. (1), the decision function can be obtained:

$$f(x)=\sum_{t=1}^{N}\left(\overline{\alpha}_t^{'}-\overline{\alpha}_t\right)\varphi^T(x_t)\varphi(x)+\overline{b}=\sum_{t=1}^{N}\left(\overline{\alpha}_t^{'}-\overline{\alpha}_t\right)K(x_t,x)+\overline{b}, \tag{12}$$

where $K(x_t,x)=\varphi^T(x_t)\varphi(x)$ is the kernel function. This paper experiments with three different to investigate the effect of kernel type.

Linear: $$K(x_t,x)=x_t^T x \tag{13}$$

Polynomial: $$K(x_t,x)=\left(x_t^T x+1\right)^d \tag{14}$$

Gaussian: $$K(x_t,x)=\exp\left(\frac{-\left\|x-x_t\right\|^2}{2\sigma^2}\right) \tag{15}$$

3. Empirical Modeling

3.1 Data Description

Daily exchange rates used in this study are the nominal bilateral exchange rate and 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 closing rates for a total of 81 currencies and commodities. Four currencies are selected for our study: Swiss Francs (CHF), Indian Rupees (INR), South Korean Won (KRW) and Philippines Pesos (PHP) against U.S. dollar for the period of beginning of 2001 to the end of 2004. After the first-order differencing, the total 1004 daily observations of the exchange rates are divided in training and test data set. The first 752 daily training data points cover the time period from 2001/01/02 up to 2003/12/31, while in the remaining data the 127 daily data points starting from 2004/01/02 up to 2004/06/30 are used in test phase 1 and the 252 daily data points starting from 2004/01/02 to 2004/12/31 are used in test phase 2.

3.2 Model Specification

For SVR, this paper focuses on the following simple nonlinear ARI model:

$$\Delta y_t = g(\Delta y_{t-1},\Delta y_{t-p})+e_t,\ for\ p=2,3,\cdots, \tag{17}$$

where the input in SVR is $x_t=\left(y_{t-1},y_{t-p}\right)$. To determine the lag order of the nonlinear ARI model, we calculate the cost function values of SVR, provided by Eq. (3), with the lag minimizing cost values to be chosen as the p in SVR. The cost functions values with different lag length, p, are described in table 1. The appropriate lag lengths of nonlinear ARI model used in SVR are 3, 3, 3 (linear, polynomial and Gaussian kernels, respectively) for CHF, 2, 1, 2 for INR, 6, 8, 7 for KRW and 10, 3, 9 for PHP, respectively. To compare adequately, we fix the number of training points to be 742.

Table 1 Determination of lag order of ARI model using SVR

Currency		CHF			INR			KRW			PHP		
SVR		linear	poly	Gaussian	linear	poly	Gaussian	linear	poly	Gaussian	linear	poly	Gaussian
Cost function values with different lag length	1	3.4734	3.4872	3.4213	1.7619	1.7242	1.7405	3.2015	3.142	3.215	2.8835	2.8587	2.8289
	2	3.4676	3.4858	3.3768	1.752	1.7417	1.694	3.207	3.1505	3.093	2.8795	2.894	2.8021
	3	3.4576	3.4388	3.3004	1.7728	1.7634	1.7641	3.2272	3.1756	3.0995	2.8923	2.8467	2.7958
	4	3.4659	3.4733	3.3709	1.7618	1.7716	1.7003	3.2084	3.1592	3.1309	2.8893	2.8914	2.8155
	5	3.4746	3.4837	3.3315	1.7795	1.7708	1.7031	3.1981	3.1682	3.1237	2.9104	2.8776	2.8113
	6	3.4752	3.491	3.3644	1.7652	1.7723	1.7444	3.1922	3.1466	3.1185	2.9037	2.8555	2.8473
	7	3.4931	3.4877	3.4051	1.7709	1.782	1.7235	3.1952	3.1437	3.09	2.8799	2.8863	2.7819
	8	3.4801	3.458	3.3118	1.7622	1.7562	1.7227	3.2279	3.1203	3.1042	2.8895	2.8691	2.7751
	9	3.4927	3.5049	3.3124	1.7618	1.7687	1.7532	3.2047	3.1835	3.0902	2.8823	2.885	2.772
	10	3.4733	3.4944	3.3563	1.7645	1.766	1.7744	3.2088	3.1742	3.166	2.8756	2.8542	2.8432
lag order with minimum		3	3	3	2	1	2	6	8	7	10	3	9

Note: The numbers of training data points are all 742.

For linear ARI model , this paper uses AIC and BIC as model selection criteria, the results of which is given in table 2. The appropriate lag lengths for the linear ARI model of CHF, INR, KRW and PHP are 1, 1, 1 and 3, respectively.

Table 2 Determination of lag length of ARI model using OLS

Currency		CHF		INR		KRW		PHP	
		AIC	SBC	AIC	SBC	AIC	SBC	AIC	SBC
lag length	1	1530.116	1534.739	-1005.54	-1000.92	1127.764	1132.387	1438.475	1443.098
	2	1531.384	1540.63	-1006	-996.756	1129.764	1139.01	1437.774	1447.02
	3	1533.284	1547.152	-1004.66	-990.794	1130.908	1144.776	1415.36	1429.228
	4	1534.82	1553.311	-1002.73	-984.239	1131.838	1150.329	1415.651	1434.142
	5	1536.193	1559.307	-1002.13	-979.015	1133.234	1156.348	1411.549	1434.663
	6	1537.974	1565.71	-1000.33	-972.596	1135.214	1162.95	1413.407	1441.144
	7	1539.399	1571.758	-1000.09	-967.726	1136.885	1169.244	1414.615	1446.974
	8	1541.374	1578.356	-998.744	-961.762	1137.96	1174.942	1416.504	1453.486
	9	1538.859	1580.464	-998.194	-956.589	1139.682	1181.286	1418.503	1460.108
	10	1540.431	1586.658	-996.236	-950.009	1140.129	1186.356	1420.301	1466.528
lag length of AR by OLS		1		1		1		3	

Note: The number of residuals are all 752.

4. Forecasting Results

4.1 Quantitative measures of forecasting accuracy

Table 3 Measures of forecasting accuracy used in the study

Measures	Calculation
NMSE	$NMSE(\%) = 100 \times \frac{MSE}{Var(y)} = 100 \times \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 / n}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2 / (n-1)}$
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert y_i - \hat{y}_i \right\rvert$
DA	$DA(\%) = \frac{100}{n}\sum_{i=1}^{n} a_i \quad \text{where} \quad a_i = \begin{cases} 1 & if \ (y_{i+1} - y_i)(\hat{y}_{i+1} - \hat{y}_i) \geq 0 \\ 0 & otherwise \end{cases}$

4.2 Forecasting Performance Comparison

Different choice of parameter values will lead to different estimating results. For the convenience of comparison, the fixed values of parameters, C=0.1, ε =0.05, a width value 0.2 for Gaussian kernels and d=2 for polynomial kernels, are kept for all SVR approach of four currency series. The SVR results in section 3.2 also use such fixed parameters. Table 4 and 5 report the corresponding forecasting performance revealed by quantitative measures of the forecasting rates obtained from OLS and SVR (three kernels) over two testing phases, respectively.

Table 4 Measures of out-of-sample forecasting accuracy for test phase 1: 127 days

Currency	Method		Quantitative Measures		
			NMSE(%)	MAE	DA(%)
CHF	MLE		18.761	0.0087031	42.857
	SVR	linear	18.708	0.0087203	42.857
		poly	18.716	0.0087428	43.651
		Gaussian	18.75	0.0088658	42.857
INR	MLE		5.3933	0.090969	47.619
	SVR	linear	5.2842	0.091475	48.413
		poly	5.3755	0.090934	47.619
		Gaussian	5.2187	0.090582	46.032
KRW	MLE		15.368	3.5906	44.444
	SVR	linear	15.277	3.5528	46.032
		poly	15.258	3.5706	46.825
		Gaussian	15.386	3.5751	44.444

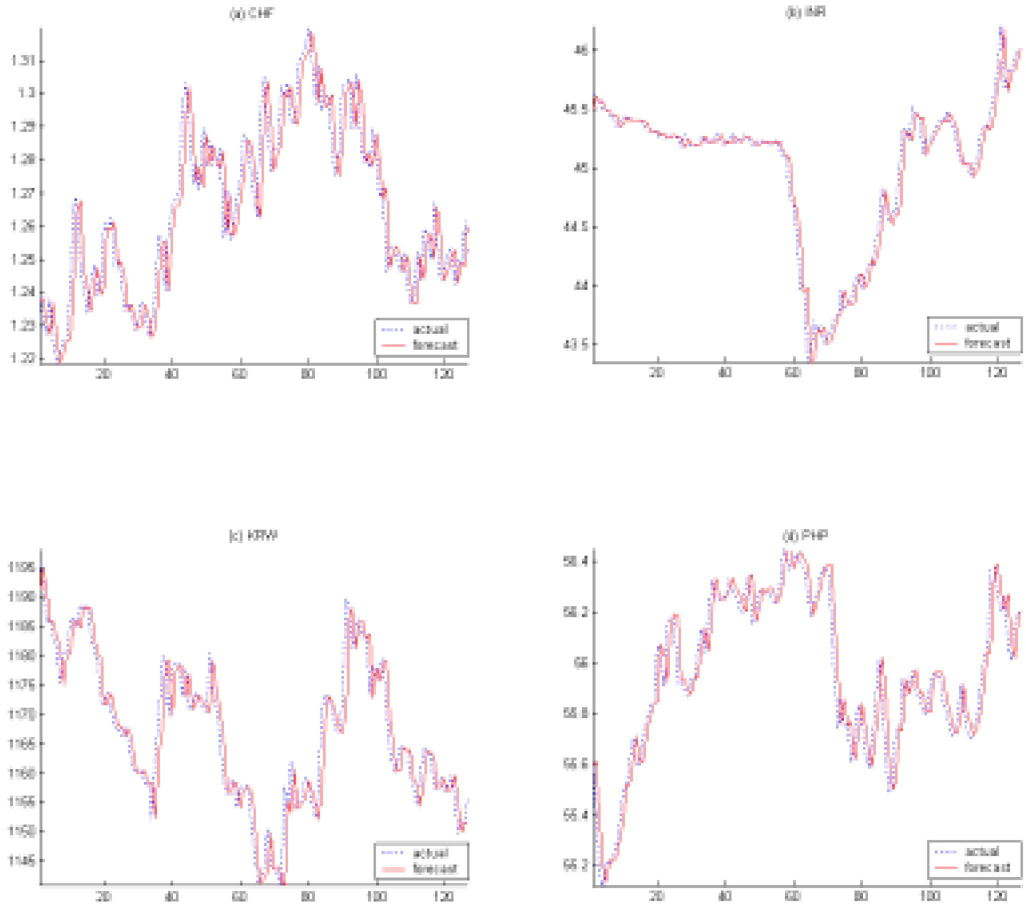
PHP	MLE		11.885	0.082898	55.556
	SVR	linear	11.564	0.081242	53.968
		poly	11.522	0.080795	57.143
		Gaussian	11.588	0.081279	58.73

Table 5 Measures of out-of-sample forecasting accuracy for test phase 2: 252 days

Currency	Method		Quantitative Measures		
			NMSE(%)	MAE	DA(%)
CHF	MLE		4.6935	0.0074222	41.833
	SVR	linear	4.6747	0.0074491	41.833
		poly	4.6892	0.007478	41.434
		Gaussian	4.6774	0.007565	41.434
INR	MLE		3.0545	0.086726	51.793
	SVR	linear	3.0047	0.087659	51.793
		poly	3.0243	0.086524	51.394
		Gaussian	2.987	0.087877	50.996
KRW	MLE		1.4957	3.3278	47.012
	SVR	linear	1.4679	3.2756	47.41
		poly	1.4614	3.2827	47.41
		Gaussian	1.483	3.2903	43.825
PHP	MLE		10.923	0.069607	53.785
	SVR	linear	10.732	0.068898	50.996
		poly	10.676	0.068517	52.59
		Gaussian	10.619	0.06869	54.183

Figure 1 (a)-(d) illustrate the actual and forecasting rates of four currencies over 127 test periods using SVR method correspond to the better records in above process. Namely, SVR forecasts of CHF result from linear kernel, KRW and PHP from polynomial kernel and INR from Gaussian kernel. It is obvious that forecasting levels obtained from better SVR are very close to the actual exchange rates for all four currencies.

Figure 1 Forecasting of four currencies by using SVR over 127 test periods



Because of being normalized, for example, NMSE (%) summarized in table 4 and 5 can be used as the generalization error to investigate the forecasting accuracy between OLS and SVR method. The short testing phase is exemplified again. Comparing to result of OLS, generalization error of SVR from above better records has been reduced by 0.053% for CHF, 0.175% for INR, 0.11% for KRW and 0.363% for PHP rates, respectively.

As a whole, SVR approach produces more accurate out-of-sample forecasts and should become the preferential choice in forecasting exchange rates.

4.3 Sensitivity of SVR to parameters

In absence of a principled way to choose free parameter values of SVR, therefore, other aim of this study is to investigate the sensitivity of SVR, represented by the generalization error (NMSE) and numbers of support vector, with respect to these free parameters. These parameters we analyze are regularization parameter C and ε (epsilon). South Korean Won forecasts using SVR with polynomial kernel are exemplified below. Similar behavior is observed for linear and Gaussian kernels also.

Figure 2 Sensitivity of SVR versus C with polynomial kernel for forecasting KRW rates

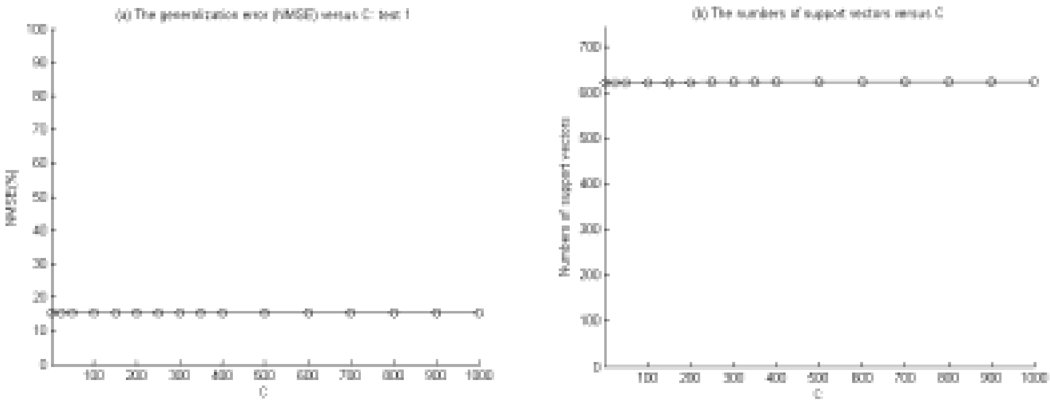


Figure 2 illustrates the sensitivity of SVR versus regularization parameter C. Here, C is varied from a very small value 0.01 to very large value 1000 and ε is arbitrarily chosen to be 0.05. From fig. 2(a), C is found to almost have no impact on the generalization error, which is only between 15.254% and 15.277%. The numbers of support vector, as demonstrated in fig. 2(b), also remain constant in the range [622,624] when C increases from 0.01 to 1000.

Figure 3 Sensitivity of SVR versus epsilon with polynomial kernel for forecasting KRW rates

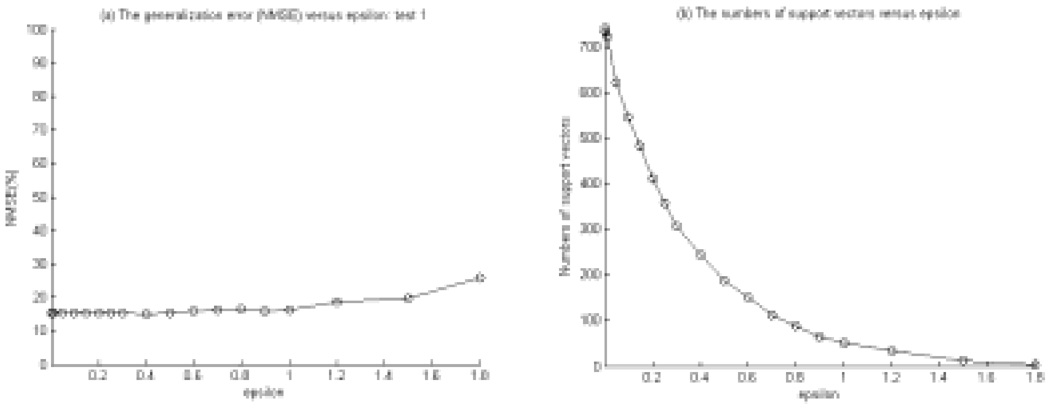


Figure 3 illustrates the forecasting performance versus ε with fixed C=0.1. The generalization error is not influenced greatly by epsilon, whose value firstly keeps almost fixed and later increases little from 16% to about 25%. However, the numbers of support vector decreases rapidly from 740 to near 0 when ε rises from 0.001 to 1.8. This indicates that epsilon has a strong effect on the numbers of support vector.