

SVM Based Models for Predicting Foreign Currency Exchange Rates

Joarder Kamruzzaman
GSCIT, Monash University
Churchill, Victoria, Australia

Ruhul A Sarker
School of Computer Science
University of NSW, Canberra, Australia

Iftekhar Ahmad
GSCIT, Monash University
Churchill, Victoria, Australia

Abstract

Support vector machine (SVM) has appeared as a powerful tool for forecasting forex market and demonstrated better performance over other methods, e.g., neural network or ARIMA based model. SVM-based forecasting model necessitates the selection of appropriate kernel function and values of free parameters: regularization parameter and ε -insensitive loss function. In this paper, we investigate the effect of different kernel functions, namely, linear, polynomial, radial basis and spline on prediction error measured by several widely used performance metrics. The effect of regularization parameter is also studied. The prediction of six different foreign currency exchange rates against Australian dollar has been performed and analyzed. Some interesting results are presented.

1. Introduction

Exchange rate prediction is one of the challenging applications of modern time series forecasting and very important for the success of many businesses and financial institutions. The rates are inherently noisy, non-stationary and deterministically chaotic [12]. One general assumption is made in such cases is that the historical data incorporate all those behaviour. As a result, the historical data is the major input to the prediction process. There are several forecasting techniques available; those are developed mainly based on different assumptions, mathematical foundations and specific model parameters. However, for better result, it is important to find the appropriate technique for a given forecasting task.

For more than two decades, Box and Jenkins' Auto-Regressive Integrated Moving Average (ARIMA) technique [1] has been widely used for time series forecasting. However, ARIMA is a general univariate model and it is developed based on the assumption that the time series being forecasted are linear and stationary [2]. In recent years, Neural Networks has found useful applications in financial time-series analysis and forecasting [9], [11-14]. In several applications, Tang and Fishwick [9], Jhee and Lee [5], Wang and Leu [11], Hill *et al.* [4], Kamruzzaman and Sarker [6] and many other researchers have shown that ANNs perform better than ARIMA models, specifically, for more irregular series and for multiple-period-ahead forecasting. Zhang and Hu [14] analysed backpropagation neural networks' ability to forecast an exchange rate. Recently, Yao *et al.* [12] evaluated the capability of a backpropagation neural-network model as forecasting tool of Singapore forex market. Medeiros [8] proposed neuro-coefficient smooth transition autoregression to model and forecast monthly exchange rate time series. Zimmermann *et al.* [13] developed a neural network modelling to analyse the decisions and interaction of multiple agents to capture the price dynamics of multiple forex market. In [6], Kamruzzaman and Sarker investigated three

ANN based forecasting models using Standard Backpropagation, Scaled Conjugate Gradient and Backpropagation with Bayesian Regularization with six moving average technical indicators to predict six major currencies against Australian dollar. However, there are several disadvantages for NN based model: (i) dependency on a large number of parameters, e.g, network size, learning parameters and initial weight chosen, (ii) possibility of being trapped into local minima resulting in a very slow convergence, and (iii) over-fitting on training data resulting in poor generalization ability.

Recent research has been directed to Support Vector Machine (SVM) which has emerged as a new and powerful technique for learning from data and in particular for solving classification and regression problems with better performance [2], [3], [10]. Study in [7] with Australian forex data showed better forecasting of exchange rate by SVM based model compared to NN based model. The main advantage of SVM is its ability to minimize structural risk as opposed to empirical risk minimization as employed by the NN system. This paper reports a further investigation of the study in [6-7]. One aim of this paper is to investigate the effect of various kernel functions on predicting forex market. Forecasting performance is evaluated in term of five commonly used metrics measuring accuracy and upward/downward trend of the market. The effect of selecting free parameters of SVMs on prediction error is also investigated.

2. Support Vector Machine Regression Model

SVMs originate from Vapnik's statistical learning theory [10], and their major advantage over NN is that they formulate the regression problem as a quadratic optimization problem. SVMs perform by nonlinearly mapping the input data into a high dimensional feature space by means of a kernel function and then do the linear regression in the transformed space. The whole process results in nonlinear regression in the low-dimensional space.

Given a data set $G = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ realizing some unknown function $g(\mathbf{x})$, we need to determine a function f that approximates $g(\mathbf{x})$, based on the knowledge of G as below.

$$f(\mathbf{x}) = \sum_i \omega_i \phi_i(\mathbf{x}) + b$$

where $\phi_i(\mathbf{x})$ are the features, coefficients ω_i and b have to be estimated from data. In SVM regression, Vapnik proposed the use of ε -insensitive loss function where error below is not penalized, ε being known *a priori*. Then the unknown coefficient is estimated by minimizing the following risk function.

$$R = C \frac{1}{N} \sum_{i=1}^N |y_i - f(\mathbf{x}_i)|_\varepsilon + \frac{1}{2} \|\omega\|^2$$

The first term describes the ε -insensitive loss function and the second term is a measure of function flatness. The constant $C > 0$ is a regularized constant determining the trade-off between the training error and model flatness. Using the slack variables ζ and ζ^* , the optimization problem solved by SVM under appropriate constraints depends on the finite number of parameters and has the following form.

$$f(\mathbf{x}_i, \alpha_i, \alpha_i^*) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{x}_i) + b$$

$K(\mathbf{x}, \mathbf{x}_i)$ is called kernel function. Coefficients α, α^* are obtained by maximizing the following quadratic form:

$$W(\alpha, \alpha^*) = \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j)$$

The constraints $0 \leq \alpha_i^*, \alpha_j \leq C$, $\sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0$ and the above

equations define the optimization problem. Only a number of coefficients satisfy $\alpha_i^* - \alpha_i \neq 0$ and the associated points to those coefficients are called support vectors.

Kernel Function: The value of the kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ is equal to the inner product of two vectors \mathbf{x}_i and \mathbf{x}_j in the feature space $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$ satisfying Mercer's condition [3]. The kernel functions used in this study are the followings.

Linear : $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$
 Polynomial : $K(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + 1)^d$, $d = \text{degree}$
 Radial Basis : $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$, $\sigma = \text{width}$
 Spline : $K(\mathbf{x}_i, \mathbf{x}_j) = 1 + \langle \mathbf{x}_i, \mathbf{x}_j \rangle + (1/2) \langle \mathbf{x}_i, \mathbf{x}_j \rangle \min(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, -1/6) \min(\langle \mathbf{x}_i, \mathbf{x}_j \rangle)^3$

3. Forecasting Model and Evaluation

Technical and fundamental analyses are the two major financial forecasting methodologies. In recent times, technical analysis has drawn particular academic interest due to the increasing evidence that markets are less efficient than was originally thought. In this study, we used time delay moving average as technical data. The advantage of moving average is its tendency to smooth out some of the irregularity that exists between market days [12]. We used moving average values of past weeks to build SVM model to predict following week's rate. The indicators are MA5 (moving average on 5 days), MA10, MA20, MA60, MA120 and X_i , last week's closing rate. Exchange rate for the period $i+1$ is predicted. These six indicators are used as inputs to construct support vectors.

Forecasting performance of the above model is evaluated against six widely used statistical metrics [7], namely, Normalized Mean Square Error (NMSE), Mean Absolute Error (MAE), Directional Symmetry (DS), Correct Up trend (CU) and Correct Down trend (CD). NMSE and MAE measure the deviation between actual and forecasted value. Smaller values of these metrics indicate higher accuracy in forecasting. DS measures correctness in predicted directions. CU and CD measure the correctness of predicted upward

trend and downward trend, respectively. Higher values of DS, CU and CD indicate correctness in trend prediction. A detail formulation of these metrics are presented in [6].

4. Experimental Results and Analysis

In this study, we experimented with foreign currency exchange rate of US dollar, British Pound, Japanese Yen, Singapore dollar, New Zealand dollar and Swiss Franc against Australian dollar from January 1991 to July 2002. The available data was divided in training and test set. In total 565 weekly data was considered of which first 500 weekly data was used as training set and the remaining 56 weekly data as test set to evaluate the model. Since no analytical method is available to determine the most suitable kernel for a particular data set, kernel type is arbitrarily chosen by the user. We experimented with four different kernel functions, namely, linear, polynomial, radial basis and spline, on the same data set to investigate the effect of kernel type. Other aim of this study is to investigate the variation of performance with respect to regularization parameter C . In this experiment C was varied from a very small value (0.1) to very large value (10^5) and ε was arbitrarily chosen to be 10^{-3} .

In the first set of our experiment, NSME, MAE, DS, CU and CD were calculated using the forecast obtained by SVM and the available historical (/test) data over 35 weeks and 65 weeks, for each of six different currency rates against Australian dollars, using four different kernel functions (linear, polynomial, radial Basis, and spline). Figure 1(a)~(e) illustrates these performance metrics over 35 weeks and Figure 2(a)~(e) illustrates those over 65 weeks. As found in Fig. 1 & 2, no single kernel function dominates all currency prediction. This outcome encourages us to analyze each currency independently.

First of all, we consider JPY which is the most fluctuating currency. The actual exchange rate of JPY is much lower at the end compared to the beginning of the period with many sharp jumps and drops in between. Based on NSME and MAE, Spline kernel seems performing better though polynomial kernel looks also promising for NSME only. A plot of historical data shows that the actual end of period rates of other currencies are also lower, but not as low as JPY compared to their start rates. The gap between the start rates of USD and GBP is higher than their gap between the end rates. However the change in rates for both USD and GBP looks like slow but random fluctuation with no sharp up and down. Based on the predicted error, the linear kernel is the best for USD. On the other hand, the performances of linear and polynomial kernels are very close for GBP. The actual start rates of NZD and SGD are very close, and the CHF rate is lower than these two. The end rates of SGD and CHF are somewhat close, but the NZD rate is much higher than these two currencies. Interestingly, the rates of these three currencies are almost same for 30-40 weeks at the middle of the time horizon. These three rates favor polynomial kernel based on the predicted error. That means, the polynomial kernel can be recognized as the suitable kernel in predicting exchange rate for all currencies except US dollar in this

experiment. This trend in predicted error over 35 weeks (Fig. 1(a)~(b)) is very much consistent with the same over 65 week prediction (Fig. 2(a)~(b)).

Figure 1(c)~(e) show that, in term of trend prediction, linear kernel is not clearly superior in any of the cases. With respect to directional change, Fig. 1(c) shows that radial basis kernel yields best result for JPY, NZD and CHF whereas polynomial kernel produces better result for USD and SGD. The maximum difference in term of correct directional changes produced by radial basis and polynomial kernel is about 2.8%. Fig. 1(d)~(e) show that radial basis and polynomial kernels performs equally in case of three currencies in term of CU and four currencies in term of CD. Similar trend is also observed for 65 weeks prediction as illustrated in Fig. 2(c)~(d). From the directional accuracy point of view, it appears that radial basis or polynomial would be a better choice of kernel function.

Based on all the performance metrics, there is no single winner for kernel function. This can not be a convincing conclusion for practitioners. Ideally, a SVM that minimizes the prediction errors while maximizing the coverage of directional matching would be the most attractive method. However, in practice, it may happen one of the following: 1) The minimum possible prediction errors but lower coverage of directional matching than one or more kernel functions, 2) The maximum possible directional accuracy but higher prediction errors than one or more kernel functions.

In addition, it could be more complicated when one prediction error is minimum and the other is not. Similar situation can be thought for the directional indicators too. The directional indicators check only the matching of trends in each every time unit/week irrespective of their deviations. Suppose the actual trend in a given week is upward and the forecast trend is also upward with a deviational error E_1 . As per the deviational indicators, this is a 100% coverage/success for the given week. Now suppose the forecast trend is downward with a deviational error E_2 . That means there is no success (0%) in terms of deviational indicators. If $E_2 < E_1$, the predicted error indicators would favour the later case. However, from the practical point of view, the forecast in the first case is still showing gain whereas the second case is showing loss. This is not a good situation for short term forecasting. In the long term forecasting, if the predicted error is minimum and the directional indicators are accurate for most time units, then the forecast can be accepted. In our case study, the coverage of DS, CU and CD is more than 80% for most cases and no one is less than 73%.

Although the polynomial kernel performs better than the other three kernels for most currencies considered in this study, we believe the kernel function should be chosen based on the pattern of individual currency rate. In other words, no single kernel is suitable for forecasting all currency rates.

Figure 3 (a)~(c) illustrates prediction error of Japanese Yen vs regularization error C with linear kernel. Fig. 3(a) shows that NMSE drops from 0.06 to 0.035 when C is increased from 0.5 to 50. For any value larger than 50, NMSE remains the same. MAE exhibits the similar behaviour in Fig. 3(b). Fig. 3(c) shows that DS varies from

75.4% to 83.0% when C is increased from 0.5 to 7.0, later DS remains 81.54% as C increases further. Similar behaviour is observed for other kernels also. This indicates little impact on variation of prediction error with respect to regularization parameter.

5. Conclusions

This paper investigates the performance of support vector machine for Australian forex forecasting in terms of kernel type and sensitivity of free parameters selection. In reducing total prediction error (MNSE and AME) polynomial kernel produced the best result while in predicting trend (DS, CU and CD) radial basis and polynomial kernel produced equally good results. In general radial basis and polynomial kernel appeared to be a better choice in forecasting forex market in this study. However, kernel function should be chosen based on the behavior of the pattern of historical rates of individual currency. Over a wide range of values for regularization parameter C, support vectors remain exactly the same and hence do not affect the performance at all, however, showed some sensitivity over a small range of value. Further improvement in performance may be achieved by selecting complex or mixed kernel function and is currently under investigation.

6. References

- [1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, CA.
- [2] L. Cao and F. Tay, "Financial forecasting using support vector machines," *Neural Comput & Applic*, vol. 10, pp.184-192, 2001.
- [3] T. Gestel *et al.*, "Financial time series prediction using least squares support vector machines within the evidence framework," *IEEE Trans. Neural Network*, vol. 12, no.4, 2001.
- [4] T. Hill, M. O'Connor and W. Remus, "Neural network models for time series forecasts," *Management Science*, vol. 42, pp 1082-1092, 1996.
- [5] W. C. Jhee and J. K. Lee, "Performance of neural networks in managerial forecasting," *Intelligent Systems in Accounting, Finance and Management*, vol. 2, pp 55-71, 1993.
- [6] J. kamruzzaman and R. Sarker, "Forecasting of currency exchange rate: a case study," to appear in *Proc. IEEE International Conference on Neural Networks & Signal Processing (ICNNSP'03)*, Nanjing., 2003.
- [7] J. kamruzzaman and R. Sarker, "Application of support vector machine to Forex monitoring," *submitted to 3rd Int. Conf.on Hybrid Intelligent Systems HIS03, Melbourne, 2003*.
- [8] C. Medeiros, A. Veiga and C. Pedreira, "Modeling exchange rates: smooth transitions, neural networks, and linear models," *IEEE Trans. Neural Networks*, vol. 12, no. 4, pp. 755-764, 2001.
- [9] Z. Tang and P. A. Fishwick, "Back-Propagation neural nets as models for time series forecasting," *ORSA Journal on Computing*, vol. 5(4), pp 374-385, 1993.
- [10] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [11] J. H. Wang and J. Y. Leu, "Stock market trend prediction using ARIMA-based neural networks," *Proc. of IEEE Int. Conf. on Neural Networks*, vol. 4, pp. 2160-2165, 1996.
- [12] J. Yao and C.L. Tan, "A case study on using neural networks to perform technical forecasting of forex," *Neurocomputing*, vol. 34, pp. 79-98, 2000.
- [13] H. Zimmermann, R. Neuneier and R. Grothmann, "Multi-agent modeling of multiple FX-markets by neural networks," *IEEE Trans. Neural Networks*, vol. 12, no. 4, pp. 735-743, 2001.
- [14] G. Zhang and M. Y. Hu, "Neural network forecasting of the British Pound/US Dollar exchange rate," *OMEGA: Int. Journal of Management Science*, 26, pp 495-506, 1998.

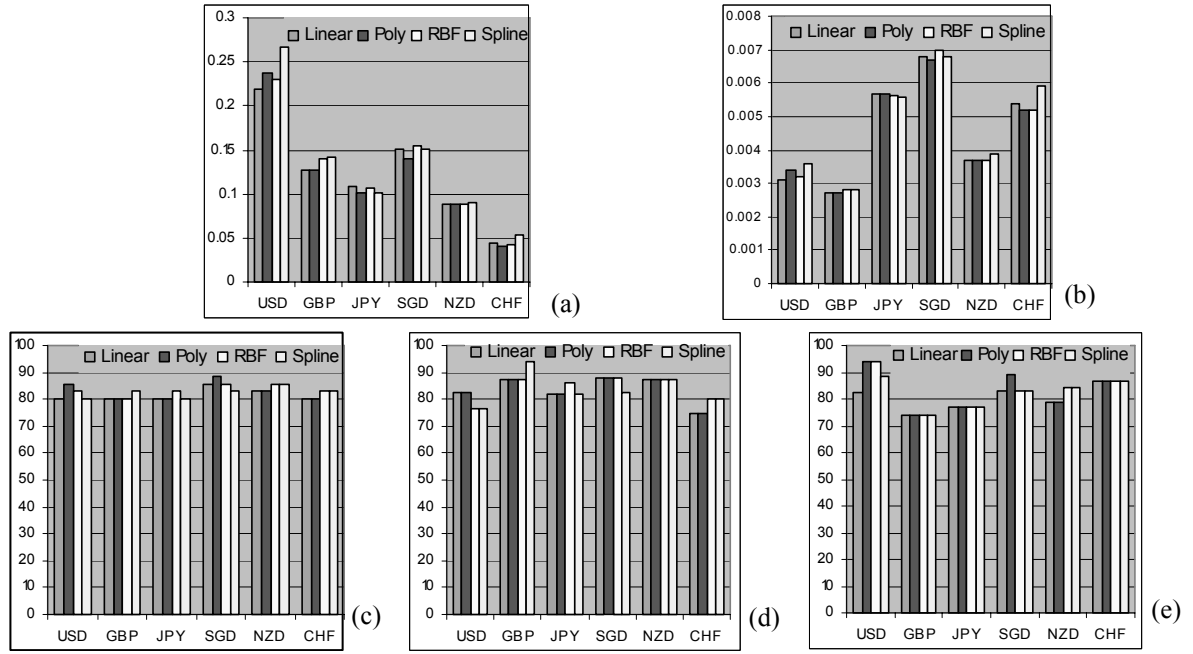


Fig. 1. Prediction performance in terms of (a) NSME, (b) MAE, (c) DS, (d) CU and (e) CD of six currencies over 35 weeks using different kernels.

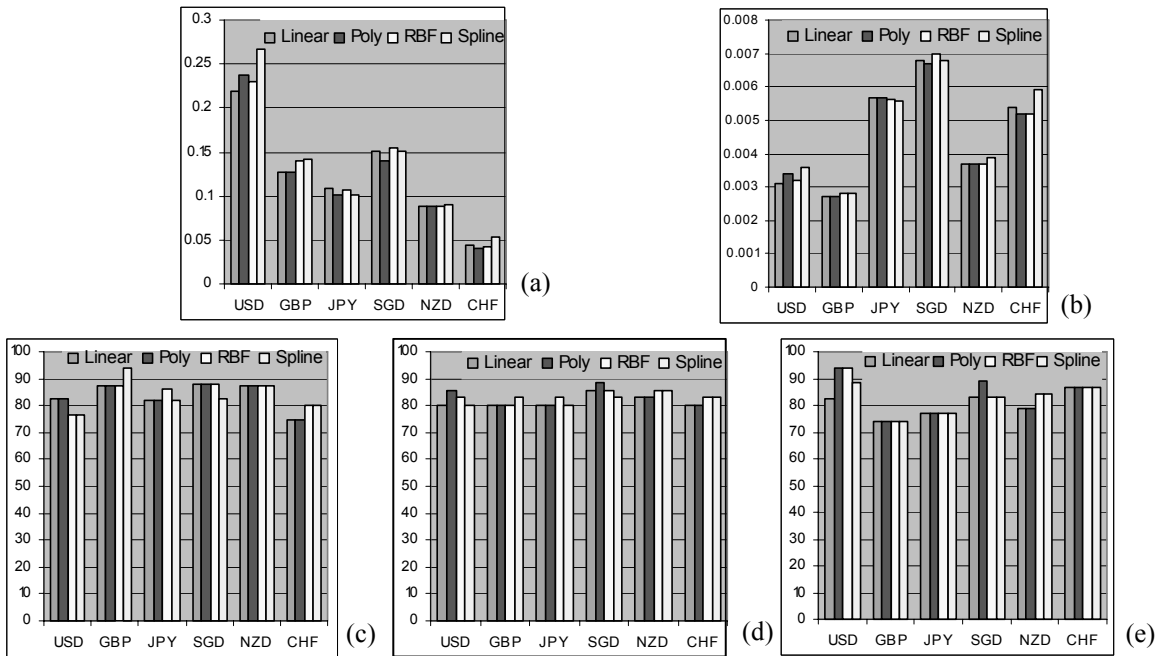


Fig. 2. Prediction performance in terms of (a) NSME, (b) MAE, (c) DS, (d) CU and (e) CD of six currencies over 65 weeks using different kernels.

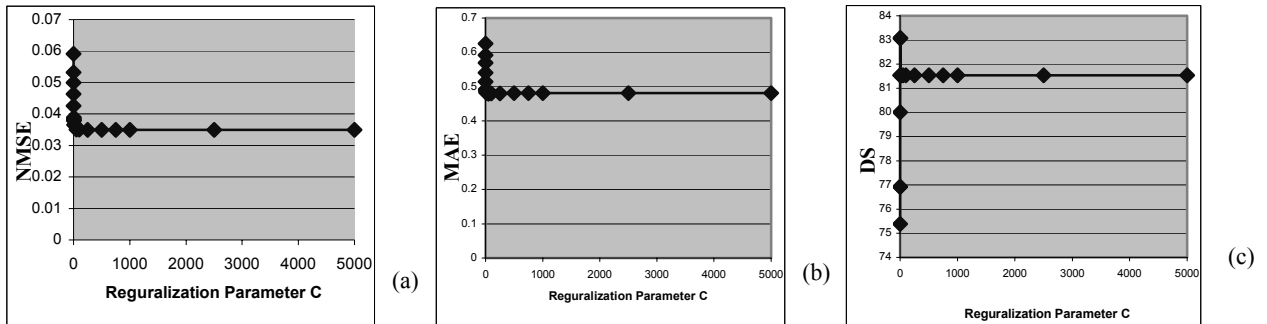


Fig. 3. Prediction performance versus regularization parameter C with linear kernel for predicting Japanese Yen exchange rate.