Control and Cybernetics

vol. **38** (2009) No. 3

Application of SVR with improved ant colony optimization algorithms in exchange rate forecasting*

by

Wei-Mou Hung[1] **and Wei-Chiang Hong**[2]

[1] Department of Industrial Engineering and Technology Management
Da Yeh University
168, University Rd., Dacun, Changhua 51591, Taiwan
[2] Department of Information Management, Oriental Institute of Technology
58 Sec. 2, Sichuan Rd., Panchiao, Taipei, 220, Taiwan
e-mail: samuelhong@ieee.org

Abstract: Traditional time series forecasting models, like ARIMA and regression models, can hardly capture nonlinear patterns. Support vector regression (SVR), a novel neural network technique, has been successfully used to solve nonlinear regression and time series problems. The SVR model applies the structural risk minimization principle to minimize the upper bound of the generalization error, instead of minimizing the training error, employed by most conventional neural network models. Thus, parameter determination for an SVR model is appropriate for achieving high forecasting accuracy. Several evolutionary algorithms, such as genetic algorithms and simulated annealing algorithms have been used in parameter selection, but these algorithms often suffer from the possibility of being trapped in local optimum. This study used an improved ant colony optimization algorithm in an SVR model, called SVRCACO, for selecting suitable parameters, with encouraging local search in areas where forecasting accuracy improvement continues to be made, then, autocatalytically converge to promising regions. Numerical examples of exchange rate forecasting from the existing literature are employed to assess the performance of the proposed model. Experimental results show that the proposed model outperforms other approaches from the literature.

Keywords: support vector regression (SVR), continuous ant colony optimization algorithms (CACO), exchange rates, financial forecasting.

1. Introduction

Exchange rate forecasting is one of the most important and challenging issues in the modern financial forecasting fields (Beran and Ocker, 1999; Fernandez-Rodriguez, Sosvilla-Rivero and Andrada-Felix, 1999). Numerous models were

*Submitted: June 2007; Accepted: January 2009.

expected to provide the investors with more precise predictive information. Traditionally, researchers and practitioners employed economic theory to understand the structural relations between exchange rate and other variables. They also applied some statistical approaches to identify the structure of correlation in historical tendencies. The famous representatives are ARIMA model, regression model, and random walk model. However, due to linear or specific assumptions, these models can hardly ensure accurate forecasting (Brooks, 1997; Diebold, Gardeazabal and Yilmaz, 1994; Kilian and Taylor, 2003; Trapletti, Geyer and Leisch, 2002). To overcome this drawback, Engle (1982) proposed the ARCH (autoregressive conditional heteroscedasticity) model, followed by the GARCH generalization of Bollerslev (1986). Henceforth, a series of successors (Kilian and Taylor, 2003; Fernandes, 1998; Vilasuso, 2002) extended the GARCH model to capture the salient data patterns of the exchange rate volatility. The superior predictability of those succeeding models is still controversial.

Recently, due to their nonlinear mapping capabilities and data processing characteristics, artificial neural networks (ANNs) attracted increasing attention in financial forecasting. Many researchers applied ANN concepts to construct appropriate forecasting models, including the original ANN models (Yao and Tan, 2000; Zhang and Hu, 1998; Hann and Steurer, 1996; Lisi and Schiavo, 1999; Chen and Leung, 2004; Chun and Kim, 2003; Davis, Episcopos and Wettimuny, 2001; Nag and Mitra, 2002), hybrid models of fuzzy logic (Kodogiannis and Lolis, 2002), multi-layer feed-forward network (Qi and Wu, 2003; Wu, 1995), general regression neural networks (GRNN), Leung, Chen and Daouk (2000), etc. These studies demonstrated that ANN-based models outperform the econometric forecasting models. ANN-based models seem to achieve improved and acceptable performance in financial forecasting, but the training procedure of ANN models is not only time consuming but can also get trapped in local minima and behave subjectively in selecting model architecture (Suykens, 2001).

Support vector machines (SVMs) are a significant development in overcoming the shortcomings of ANN, mentioned above. SVMs are based on statistical learning theory and kernel functions, i.e. the so-called kernel based neural networks. Rather than implementing the empirical risk minimization (ERM) principle to minimize the training error, SVMs employ the structural risk minimization (SRM) principle to minimize an upper bound on the generalization error, and allow learning any training set without error. Thus, SVMs might theoretically guarantee achieving global optimum, instead of being trapped in local optima like ANN models. In addition, the learning algorithm of SVMs may decide on suitable architecture, i.e. the number of units in the hidden layer. Furthermore, SVMs employ kernel functions to implicitly map the data into a higher, possibly infinite, dimensional space (Fig. 1). Thus, the solution of a non-linear problem in the original lower dimensional input space could be linear in a higher dimensional feature space. This makes SVMs a feasible choice for solving a variety of problems in many fields which are non-linear in nature. For more detailed introduction to SVMs, a reader is referred to, e.g., Vapnik

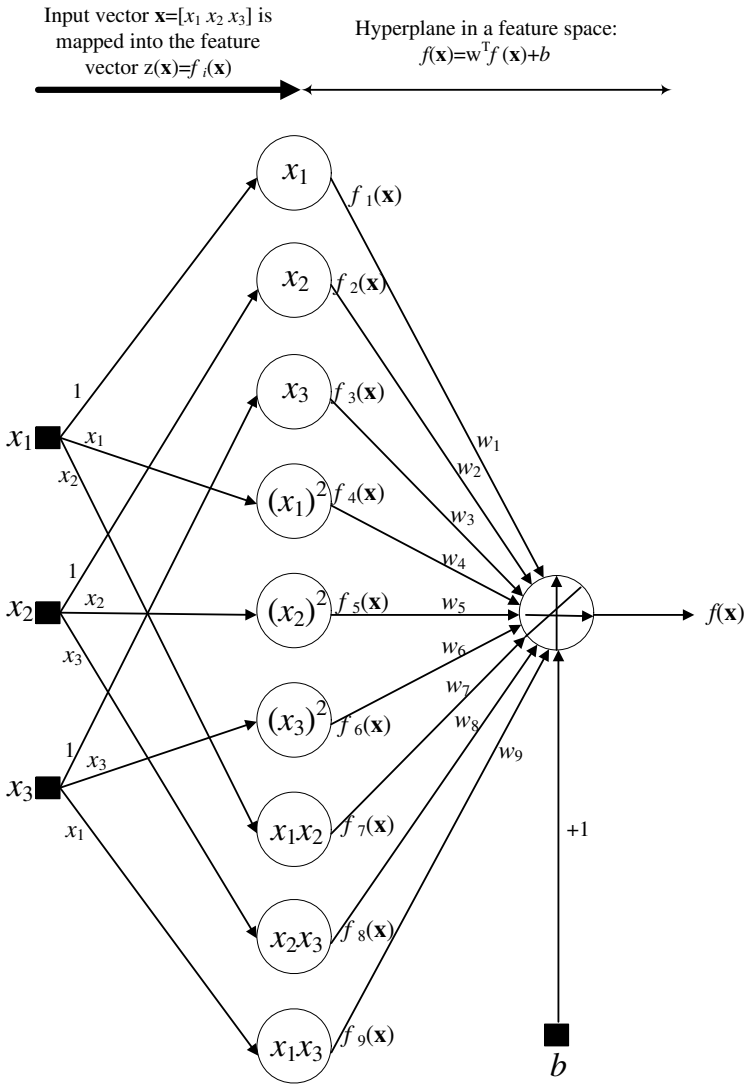

Figure 1. Transformation of the second-order polynomial hypersurface over a three-dimensional original space in a SVR model

(1995, 1998), Cortes and Vapnik (1995), Cristianini and Shawe-Taylor (2000), and Scholkopf and Smola (2002).

Although originally designed to solve pattern recognition problems, SVMs also found wide application in the fields of time series/regression, bio-informatics, and other artificial intelligence relevant applications. Particularly, along with the introduction of Vapnik's ε-insensitive loss function, SVMs also have been extended to solve nonlinear regression estimation problems, with the so-called support vector regression (SVR). Thus, SVR have been successfully employed to solve forecasting problems in many fields. These include financial time series (stock index and exchange rate) forecasting (Cao, 2003; Huang, Nakamori and Wang, 2005; Pai and Lin, 2005a; Pai et al., 2006; Tay and Cao, 2001, 2002), engineering and software (production and reliability) forecasting (Pai and Lin, 2005b; Hong and Pai, 2006; Hong et al., 2005; Pai and Hong, 2006), atmospheric science forecasting (Hong and Pai, 2007; Lu et al., 2002; Mohandes et al., 2004), and so on. The SVR model was also successfully applied to forecast electric load (Pai and Hong, 2005a,b). The practical results indicated that poor forecasting accuracy is due to inadequate selection of three parameters (C, ε and σ) in the SVR model. These parameters are often hard to determine directly, due to time consuming and conceptual constraints. It is possible to employ an optimization procedure to obtain a suitable parameter combination, e.g. by minimizing the objective function describing the structural risk, mentioned above. The present authors conducted a series of relevant studies, employing different optimization algorithms (genetic, simulated annealing, immunological, and Tabu search) for parameter determination, in order to identify empirical rules as to which algorithms are best suited for specific data patterns.

Thus, in this study, the ant colony optimization algorithm (ACO) is tried out to determine the values of three parameters in an SVR model. In addition, as ACOs were developed for discrete optimization, their application to continuous optimization problems requires transformation of a continuous search space to a discrete one by discretization of the continuous decision variables, the entire procedure referred to as CACO. Thus, the proposed SVRCACO model is applied to forecast exchange rates. A numerical example from the literature (Pai et al., 2006) is employed to demonstrate the forecasting accuracy improvement of the proposed model. The remainder of this paper is organized as follows. In Section 2, we explain the methodology. Section 3 gives experimental results. Finally, Section 4 concludes the paper.

2. Forecasting methodology

2.1. Structural risk minimization

As mentioned above, traditional AI approaches tended to be based on functions minimizing training errors, i.e., empirical risk minimization (ERM). However, ERM does not guarantee good generalization to new data sets. To separate

the classes with a surface (hyperplane) maximizing the margin in the training data set, SVMs employ the structural risk minimization (SRM) principle that aims to minimize a bound on the generalization error, rather than minimizing the mean square error over the training data set. SRM provides a well-defined quantitative measurement for the capacity of a learned function to capture the true structure of the data distribution and generalize over unknown test data set. The Vapnik-Chervonenkis (VC) dimension (Vapnik, 1995) has been applied to measure this capacity.

We are given a training data set of N elements $\{(\mathbf{x}_i, y_i), i = 1, 2, ..., N\}$, where $\mathbf{x}_i$ is the ith element in n-dimensional space, i.e., $\mathbf{x}_i = [x_{1i}, ..., x_{ni}] \in \Re^n$, and $y_i \in \{-1, +1\}$ is the label of $\mathbf{x}_i$. We wish to define a deterministic function $f : \mathbf{x} \to \{-1, +1\}$ for a given input data $\mathbf{x}$ and adjustable weights $\mathbf{w}$ ($\mathbf{w} \in \Re^n$), according to the same but unknown probability distribution ($P(\mathbf{x}, y)$). The weights $\mathbf{w}$ would be adjusted during the training stage. Since the underlying probability distribution $P(\mathbf{x}, y)$ is unknown, the upper bound for the probability of classification errors on the test data set (i.e., expected error of f), $R(f)$, cannot be minimized directly. Thus, one can estimate an approximate function of $R(f)$, i.e., empirical risk, denoted as $R_{emp}(f)$, that is close to the optimal one based on the training data pairs $(\mathbf{x}, y)$. Then, according to the SRM principle (Vapnik, 1995, 1998), $R(f)$ and $R_{emp}(f)$ are expressed as:

$$R(f) \leq R_{emp}(f) + \varepsilon_1(N, h, \eta, R_{emp}) \tag{1}$$

$$R_{emp}(f) = \frac{1}{N} \sum_{i=1}^{N} |y_i - f(\mathbf{x}_i)|_{\text{loss function.}} \tag{2}$$

$$\varepsilon_1(N, h, \eta, R_{emp}) = 2\varepsilon_0^2(N, h, \eta) \left(1 + \sqrt{1 + \frac{R_{emp}(f)}{\varepsilon_0^2(N, h, \eta)}}\right) \tag{3}$$

$$\varepsilon_0(N, h, \eta) = \sqrt{\frac{h\left(\ln\left(\frac{2N}{h}\right) + 1\right) - \ln\left(\frac{\eta}{4}\right)}{N}} \tag{4}$$

Equation (1) holds with probability $1 - \eta$ for $0 \leq \eta \leq 1$; $\varepsilon_0(N, h, \eta)$ is the so-called VC confidence interval. The values of $\varepsilon_0(N, h, \eta)$ depend on the number of training data N, the VC dimension h, and the value of η.

For a small empirical risk $R_{emp}(f)$, e.g. close to 0, (1) would approximately get reduced to $R_{emp}(f)+4\varepsilon_0^2(N, h, \eta)$, in contrast, for a large empirical risk, close to 1, (2) would approximately get reduced to $R_{emp}(f) + \varepsilon_0(N, h, \eta)$ (Haykin, 1999).

Thus, there are two strategies for minimizing the upper bound, $R(f)$. The first one is to keep the VC confidence ($\varepsilon_0(N, h, \eta)$) fixed and to minimize the empirical risk, and most of ANN models employ it. However, it does not perform well, because dealing with $R_{emp}(f)$ alone could not guarantee of reduction VC confidence. The second one is to fix the empirical risk to a small value and to minimize the VC confidence, this being the so-called SRM principle. Although

SVMs implement this principle, their training algorithm that aims to minimize the VC dimension is still based on a hierarchy that depends on the data (Vapnik, 1995; Shawe-Taylor et al., 1998).

2.2. Support Vector Machines for regression

As mentioned above, SVMs were originally used for classification purposes, but their principles can be extended easily to the task of regression and time series prediction. The concept of SVMs for the case of regression will be briefly introduced. A nonlinear mapping $\varphi(\cdot) : \Re^n \rightarrow \Re^{n_h}$ is defined to map the input data (training data set) $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ into a so-called high dimensional feature space (which may have infinite dimensions), $\Re^{n_h}$ (Fig. 2 (a) and (b)). Then, in the high dimensional feature space, there theoretically exists a linear function, f, corresponding to the nonlinear relationship between input and output data. Such a linear function, namely SVR function, is

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}) + b \tag{5}$$

where $f(\mathbf{x})$ denotes the forecasting values; the coefficients $\mathbf{w}$ ($\mathbf{w} \in \Re^{n_h}$) and b ($b \in \Re$) are adjustable. As mentioned above, SVM aims at minimizing the empirical risk,

$$R_{emp}(f) = \frac{1}{N}\sum_{i=1}^{N} \Theta_\varepsilon(y_i, \mathbf{w}^T\varphi(\mathbf{x}_i) + b) \tag{6}$$

where $\Theta_\varepsilon(\mathbf{y}, f(\mathbf{x}))$ is the ε-insensitive loss function (thick line in Fig. 2(c)), defined as

$$\Theta_\varepsilon(\mathbf{y}, f(\mathbf{x})) = \begin{cases} |f(\mathbf{x}) - \mathbf{y}| - \varepsilon, & \text{if} \quad |f(\mathbf{x}) - \mathbf{y}| \geq \varepsilon \\ 0, & otherwise \end{cases} . \tag{7}$$

$\Theta_\varepsilon(\mathbf{y}, f(\mathbf{x}))$ is employed to find out an optimum hyperplane in the high dimensional feature space (Fig. 2 (b)) so as to maximize the distance separating the two subsets of the training data. Thus, SVR focuses on finding the optimum hyperplane and minimizing the training error between the training data and the ε-insensitive loss function.

Then, the SVR minimizes the overall error

$$\underset{\mathbf{w},b,\xi^*,\xi}{\mathrm{Min}} \; R_\varepsilon(\mathbf{w}, \xi^*, \xi) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + C\sum_{i=1}^{N}(\xi_i^* + \xi_i) \tag{8}$$

with the constraints

$$\begin{aligned}
&\mathbf{y}_i - \mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i^*, i = 1, 2, ..., N \\
&-\mathbf{y}_i - \mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i, i = 1, 2, ..., N \\
&\xi_i^* \geq 0, i = 1, 2, ..., N \\
&\xi_i \geq 0, i = 1, 2, ..., N.
\end{aligned}$$

The first term of (8) employs the concept of maximizing the distance of two separated training data and is used to regularize weight sizes, to penalize large weights, and to maintain regression function flatness. The second term penalizes training errors of $f(\mathbf{x})$ and $\mathbf{y}$ by using the ε-insensitive loss function. C is a trade-off parameter of these two terms. *Training errors above ε are denoted as ξ_i^*, while training errors above ε are denoted as ξ_i (Fig. 2 (b)).*

After the quadratic optimization problem with inequality constraints is solved, the parameter vector w in Eq. (6) is obtained,

$$\mathbf{w} = \sum_{i=1}^{N} (\beta_i^* - \beta_i)\, \varphi(\mathbf{x}_i) \tag{9}$$

where β_i^*, β_i are obtained by solving the quadratic program and are the Lagrangian multipliers. Finally, the SVR regression function is obtained in the dual space,

$$f(\mathbf{x}) = \sum_{i=1}^{N} (\beta_i^* - \beta_i)\, K(\mathbf{x}_i, \mathbf{x}) + b \tag{10}$$

where $K(\mathbf{x}_i, \mathbf{x}_j)$ is a kernel function, and the value of the kernel equals the inner product of two vectors, $\mathbf{x}_i$ and $\mathbf{x}_j$, in the feature space, $\varphi(\mathbf{x}_i)$ and $\varphi(\mathbf{x}_j)$, respectively; i.e., $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \circ \varphi(\mathbf{x}_j)$. Any function that meets Mercer's condition (Vapnik, 1995) can be used as the kernel function.

There are several types of kernel functions. The most often used kernel functions are the Gaussian one with the width of $\sigma : K(\mathbf{x}_i, \mathbf{x}_j) = \exp(\frac{-0.5\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2})$, and the polynomial kernel of order d and constants a_1 and a_2: $K(\mathbf{x}_i, \mathbf{x}_j) = (a_1 \mathbf{x}_i \mathbf{x}_j + a_2)^d$. If the value of σ is very large, the Gaussian kernel approximates the linear kernel (polynomial of order 1). It is hard to determine the proper type of kernel functions for specific data patterns (Amari and Wu, 1999; Vojislav, 2001). However, the Gaussian kernel is not only easier to implement, but also capable of nonlinear mapping of the training data into an infinite dimensional space. Thus, it is suitable to deal with nonlinear problems. Therefore, the Gaussian kernel function is used in this study.

The selection of the three parameters, σ, ε and C, of a SVR model influences the accuracy of forecasting. However, structural methods for confirming efficient selection of parameters are lacking. Recently, Pai and Hong (2005a,b) introduced the traditional approach for parameter determination. They applied a series of search algorithms to test the potentiality and the suitability involved in the parameter selection of an SVR model. However, as mentioned above, GA and SA lack memory functions, which leads to time consuming search for the suitable parameters of an SVR model. Therefore, continuous ant colony optimization (CACO) is used in the proposed SVR model to optimize parameter selection.

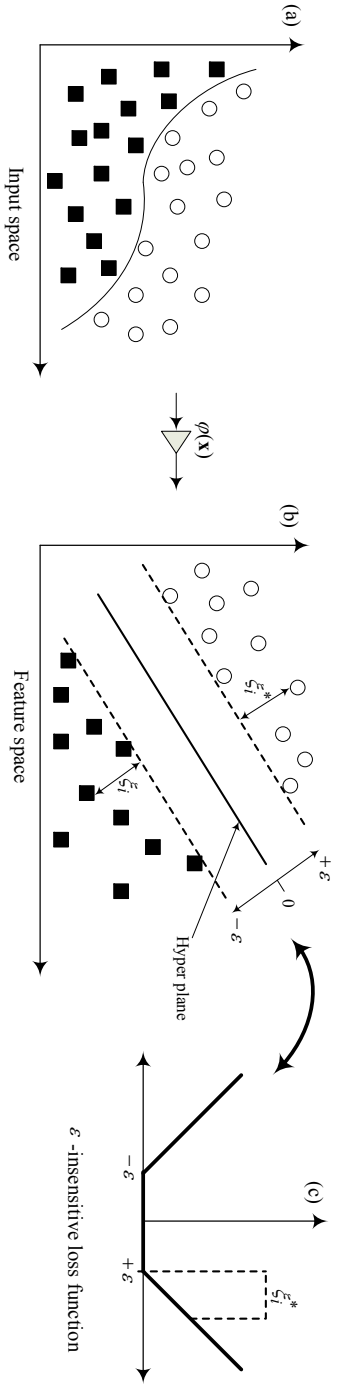

Figure 2. Illustration of the transformation process of an SVR model

2.3. CACO in selecting parameters of the SVR model

Ant colony optimization algorithms were first proposed by Dorigo (1992) and Dorigo, Maniezzo and Colorni (1996). The process by which ants can establish the shortest path between ant nest and food is illustrated in Fig. 3. Initially, ants leave their nest in random directions to search for food. Roaming around, ants deposit some amount of pheromone trails, which can be detected by other ants. For example, assume Ant 1 finds a food source. It will pick up some food and go back to the nest by following its previous pheromone trail, laying *additional* pheromone on the same path while other ants (Ant 2, Ant 3,...) are still roaming about randomly. When the second ants group leaves the nest to look for food, those ants can detect higher level of pheromone on Path 1 than on other paths. Since the probability for a path to be followed is determined by pheromone amount, more ants will follow Path 1 in this second round of looking for food. In this way, the ants can establish the shortest paths from their colony to the food sources. Obviously, even if an isolated ant roams randomly, it can, informed by pheromones, follow the collective behavior of ant colony.

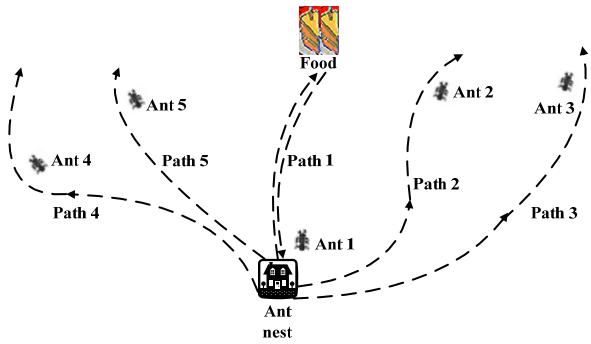

Figure 3. Description of establishing the shortest path between ant nest and food

Due to their learning and search capabilities, ACO algorithms have been successfully used to deal with different combinatorial optimization problems including job-shop scheduling (Colorni et al., 1994), traveling salesman problem (Dorigo and Gambardella, 1997), space-planning (Bland, 1999), quadratic assignment problems (Maniezzo and Colorni, 1999), and data mining (Parpinelli, Lopes and Freitas, 2002). Application of ACO algorithms to traffic flow prediction problems, however, is quite rare. These algorithms were originally proposed for discrete optimization and their application to continuous optimization problems requires some specific transformation techniques. In the literature, only few approaches for continuous optimization have been proposed, such as continuous ACO (Bilchev and Parmee, 1995; Mathur et al., 2000; Woodrich and

Bilchev, 1997), API algorithm (Monmarche, Venturini and Slimane, 2000), and continuous interacting ACO (Dreo and Siarry, 2002). However, these algorithms added some operational mechanisms that are mostly beyond the regular essence of ACO. Recently, Socha and Dorigo (2006) proposed an extension of ACO to continuous domain by applying the continuous probability density function to decide on probabilistic pheromone choice. *In this model, though, other external parameters should be determined in advance, so that one would be solving a continuous technological issue instead of appropriate SVR parameter determination.*

Hence, the concept of transforming continuous search space to a discrete one by discretizing the continuous decision variables (Abbaspour, Schulin and van Genuchten, 2001) can be employed, the so-called continuous ACO algorithm (CACO). In this study, the CACO for the traveling salesman problem is modified to determine the parameters of an SVR model in the discrete search space. The probability, $P_k(i,j)$, that an ant k moves from city i to city j is expressed as

$$P_k(i,j) = \begin{cases} \arg \max\limits_{S \in M_k} \left\{ [\tau(i,S)]^{\alpha} [\eta(i,S)]^{\beta} \right\}, & \text{if } q \le q_0 \\ \text{Eq.(12)}, & \textit{otherwise} \end{cases} \tag{11}$$

$$P_k(i,j) = \begin{cases} [\tau(i,j)]^{\alpha} [\eta(i,j)]^{\beta} \Big/ \sum\limits_{S \in M_k} [\tau(i,S)]^{\alpha} [\eta(i,S)]^{\beta}, & j \notin M_k \\ 0, & \textit{otherwise} \end{cases} \tag{12}$$

where $\tau(i,j)$ is the pheromone level between cities i and j, $\eta(i,j)$ is the inverse of distance between cities i and j. In this study, forecasting error is represented by the distance between cities. Parameters α and β determine the relative importance of pheromone level and M_k is a set of cities in the next column of the city matrix for ant k; q is a random uniform variable [0,1] and the value q_0 is a constant between 0 and 1. The values of α, β and q_0 are set to be 8, 5 and 0.2 respectively.

Once ants have completed their tours, the pheromone deposited by ants on the visited paths is considered as the information on the quality of paths from the nest to the food sources. Therefore, dynamic updating of pheromone plays the main role in real ant colony search behavior. The local and global pheromone updating rules are expressed as (13) and (14), respectively:

$$\tau(i,j) = (1-\rho)\tau(i,j) + \rho\tau_0 \tag{13}$$

$$\tau(i,j) = (1-\delta)\tau(i,j) + \delta\Delta\tau(i,j) \tag{14}$$

where ρ is the local evaporation rate of pheromone, $0 < \rho < 1$; τ_0 is the initial amount of pheromone deposited on each of the paths. In this work, the value of ρ is set to be 0.01. In addition, the initial amount of pheromone, τ_0, generated

conform to the approach of Dorigo and Gambardella (1997), is expressed as

$$\tau_0 = \frac{1}{nL_{nn}} \tag{15}$$

where n is the number of cities and L_{nn} is the tour length produced by the nearest neighbor heuristic.

Global trail updating is accomplished according to (14); δ is the global pheromone decay parameter, $0 < \delta < 1$, set equal 0.2 for this study; $\Delta\tau(i,j)$, expressed as in (16), is used to increase the pheromone on the path of the solution:

$$\Delta\tau(i,j) = \begin{cases} 1/L\,, & \text{if } (i,j) \in \textit{global best route} \\ 0\,, & \textit{otherwise} \end{cases} \tag{16}$$

where L is the length of the shortest route.

The details of the CACO procedure used in this study are as follows:

(1) Initialization: Set upper bounds of the three SVR positive parameters, σ, C and ε. In this study, to discretize these continuous parameters, each digit of the parameters is represented by ten cities. Thus, each digit contains 10 possible values from 0 to 9. Assume the limits of parameters σ, C and ε are 1, 10 000, and 1 correspondingly. The numbers of digits that represent each parameter are all set as four. Hence, three ant colonies are defined as σ-ant-colony, C-ant-colony, and ε-ant-colony in the search for three parameter values. The numbers of cities for each ant colony are 40, the total number of cities is 120.

(2) Assigning tasks to each ant colony: On step one, pathway-structure list of each ant-colony would be generated. Fig. 4 shows the parameters represented by the CACO algorithms and pathway-structure list in this study. Each ant will randomly select a pathway from the pathway list in its colony and remember the values of the represented parameters (σ, C or ε). At the end of the pathway, the three parameter values enter the SVR model (i.e., objective function) and the forecasting error is calculated. The shortest travel pathway in each search loop would be determined based on the smallest forecasting error. In this work, the normalized mean square error (NMSE) is used as the forecasting error index as given by

$$NMSE = \frac{\sum_{i=1}^{n}(a_i - f_i)^2}{\sum_{i=1}^{n}(a_i - \bar{a})^2} \tag{17}$$

where n is the number of forecasting periods; a_i is the actual exchange rate at period i; $\bar{a}$ denotes the mean of the actual exchange rate; and f_i is the forecasting exchange rate at period i.

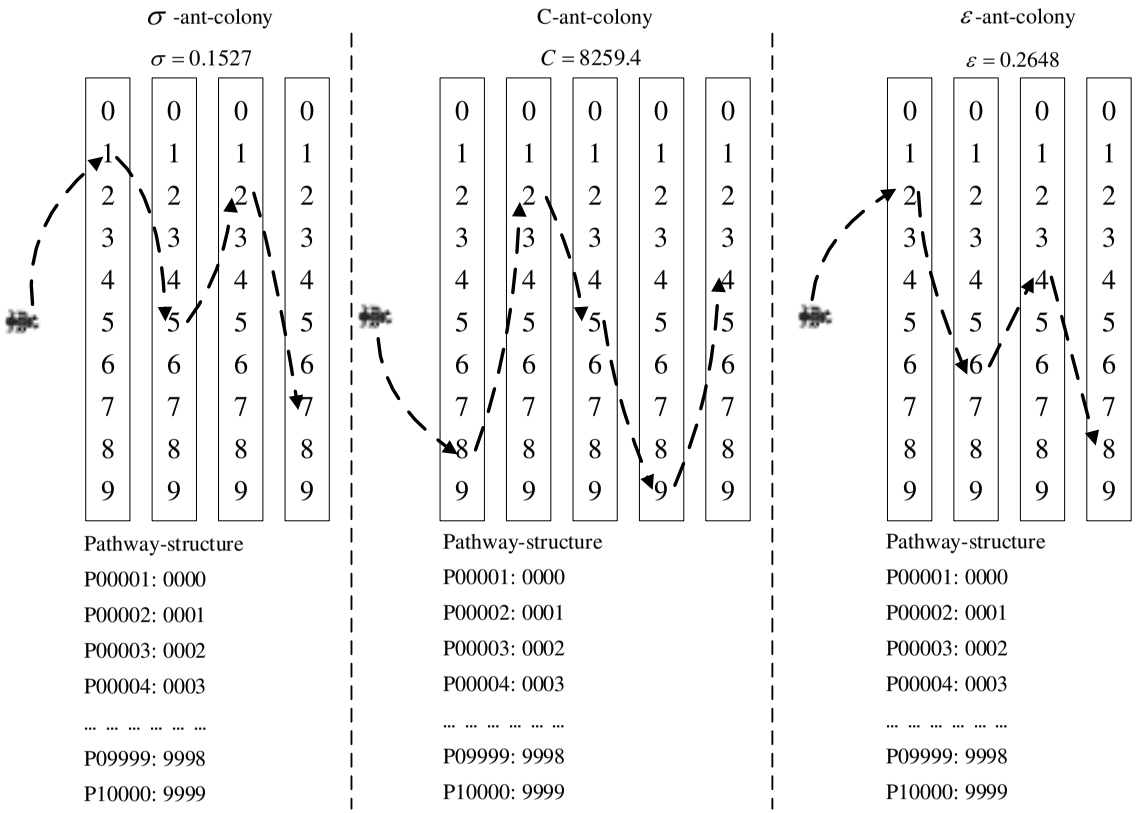

Figure 4. SVR parameter representation by the CACO algorithms

(3) Determining the numbers of ants and calculating distance between cities: The numbers of ants are set to be 10 in each ant colony, i.e., total of 30 ants for each search iteration. The maximum number of iterations is set to 20 000.

(4) Stop criterion I: When the maximum number of iterations is reached, the algorithm stops and the shortest travel path of the ant colony is an approximate optimal solution. Otherwise, continue to the next step.

(5) Calculating probability of visiting: If the maximum number of iterations has not been reached, calculate the probability that an ant k in city i moves to city j in accordance with (11). Repeat preceding steps.

(6) Stop criterion II: If each ant finished its pathway-structure list from the nest to the food source, passing through all cities, then the shortest path is an approximate optimal solution. Otherwise, conduct the pheromone updating process of (13) and (14) to renew the reinforcement of pheromone. Then, go back to Step 3.

Notice that, in any iteration, when the shortest path is determined, the appropriate solution is obtained, and for those three parameters, a new search space is then re-discretized. The framework of the proposed SVRCACO model is depicted in Fig. 5. CACO is used to find a better combination of the three parameters in the SVR so that a smaller NMSE is obtained during forecasting iteration.

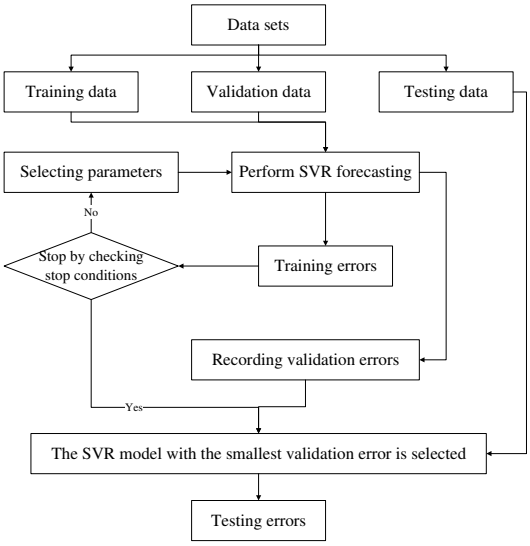

Figure 5. Framework of SVRCACO

3. Numerical example and experimental results

The exchange rate data of four European currencies, used by Lisi and Schiavo (1999) and Pai et al. (2006), and the forecasting results therein, are employed in this study to assess accuracy improvement with the proposed SVRCACO model. There are monthly exchange rates of four currencies, namely French Franc (FF), Deutschmark (DM), Italian Lira (LIT), and British Pound (GBP), all against the American Dollars (USD). The time period of the data is from January 1973 to October 1995. Then, the data are divided into three sets: the training set (from January 1973 to July 1987), the validation set (from August 1987 to September 1991), and the testing set (from October 1991 to October 1995). In this study, a rolling-based forecasting procedure is carried out and three-step-ahead forecasting policy is implemented. Four other approaches, namely the hybrid support vector machine with genetic algorithms (HSVMG), the neural networks model (NN), the chaotic model (vector-valued, local linear approximation; VLLR) and the random walk model (RW), are used to compare the forecasting performance with the proposed model.

Values of the three parameters obtained by CACO for four currencies are listed in Table 1. Due to lack of forecasting values indicated in Lisi and Schiavo (1999), Figs. 6-9 show point-to-point comparisons of actual values, HSVMG (Pai et al., 2006), and values forecasted by the proposed SVRCACO model.

Table 1. Values of three parameters for four currencies according to SVRCACO models

Currencies	Parameters			NMSE* of testing
exchange rates	σ	$\mathbf{C}$	ε	(%)
FF/USD	0.5339	33.4960	0.2099	0.3921
DM/USD	1.3467	13.8350	0.0365	0.4467
LIT/USD	0.6815	3186.1	19.7100	0.0620
GBP/USD	2.8125	9.8047	0.0117	0.4123

* – the values of *NMSE* are based on levels of exchange rates.

Lisi and Schiavo (1999) used the proportional error reduction (*per*) to compare the improvement of forecasting accuracy. The *per* is defined as

$$per = 1 - \left(\frac{NMSE_{ml}}{NMSE_{RW}} \right) \tag{18}$$

where $NMSE_{ml}$ is the *NMSE* of the proposed model, the SVRCACO model, the HSVMG model, the NN model, or the VLLR model; $NMSE_{RW}$ refers to the *NMSE* of the random walk model. To compare the forecasting performance of different models on the same basis, the proportional error reduction (*per*) is employed in this study. However, comparison should be based on the same conditions, for example, in Lisi and Schiavo (1999), due to the value of

$NMSE_{RW} = 0.999$, the forecasting results are referred to returns of exchange rates instead of levels of exchange rates. Therefore, the forecasting results from Pai et al. (2006) and this research should be transferred from levels to returns by the following relation:

$$R_i = \Delta \log a_i \tag{19}$$

After transformation by (19), the values of $NMSE_{HSVMG}$ and $NMSE_{SVRCACO}$ are calculated, as shown in Table 2. The experimental results compared with the other four approaches are also illustrated in Table 2. These results indicate that the SVRCACO model provides more accurate forecasting results than the HSVMG, NN, VLLR and RW models.

Table 2. The comparison of forecasting performances for different models

Model	Index	FF/USD	DM/USD	LIT/USD	GBP/USD
SVRCACO	NMSE*	0.7831	0.7239	0.7530	0.6452
	per	21.6%**	26.3%**	23.6%**	34.4%**
HSVMG	NMSE*	0.8489	0.7579	0.7757	0.6530
	per	15.0%	22.9%	21.3%	33.6%
NN	NMSE*	0.864	0.808	0.891	0.658
	per	13.5%	17.8%	9.6%	33%
VLLR	NMSE*	0.864	0.845	0.780	0.792
	per	13.5%	14.0%	20.9%	19.4%
RW	NMSE*	0.999	0.983	0.986	0.983

* – the values of $NMSE$ are based on returns of exchange rates.

** – indicating the highest improvement rate (Per)

Furthermore, due to the same forecasting theoretical background of the HSVMG and SVRCACO models, it is necessary to verify the contributions from the employed heuristics, i.e., genetic algorithms versus continuous ant colony optimization algorithm. To assess the statistical significance of the forecasts, we used a statistical test proposed by Diebold and Mariano (1995). First, the actual and forecasted changes of four currencies are shown in Figs. 10-13, in which the forecasted values of HSVMG and SVRCACO move over time, and, particularly, *the movements of HSVMG are more sensitive than those of SVRCACO in terms of exchange rates*. Secondly, the loss-differential series of the two models for the four currencies are shown in Figs. 14 to 17. The sample autocorrelation function (shown in Figs. 18 to 21), which decays quickly, indicates that the approximate stationarity is supported in the four currency cases. In addition, because the forecasts in this paper are three-step-ahead, thus, at least *two-dependent* forecast errors should be allowed. Based on the autocorrelation function of the loss differential for the four currencies (shown in Figs. 18 to 21),

in which significant autocorrelations appear at lags 1 and 2, the intuition of *two-dependence* is confirmed.

Based on Diebold and Mariano's suggestion, the asymptotic test (S_1) is applied, because one or more assumptions are violated necessary for the simple F test, the Morgan-Granger-Newbold test, and the Meese-Rogoff test. Therefore, the S_1 statistic, defined as below, is employed in this study,

$$S_1 = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}} \tag{20}$$

where d_i is the loss-differential series of HSVGM and SVRCACO models, defined as

$$d_i = e_{1i}^2 - e_{2i}^2 \tag{21}$$

e_1 and e_2 denoting the errors of the HSVMG and SVRCACO models, respectively; $2\pi \hat{f}_d(0)$ is the weighted sum of the available sample autocovariances:

$$2\pi \hat{f}_d(0) = \sum_{\tau=-(T-1)}^{T-1} 1 * \left(\frac{\tau}{S(T)}\right) \hat{\gamma}_d(T) \tag{22}$$

where T is the sample size; $\hat{\gamma}_d(T)$ is defined as below:

$$\hat{\gamma}_d(T) = \frac{1}{T} \sum_{t=|\tau|+1}^{T} (d_t - \bar{d})(d_{t-|\tau|} - \bar{d}) \tag{23}$$

and $1 * \left(\frac{\tau}{S(T)}\right)$ is the lag window, defined as:

$$1 * \left(\frac{\tau}{S(T)}\right) = \begin{cases} 1 & \text{if } \left|\frac{\tau}{S(T)}\right| \leq 1 \\ 0 & otherwise \end{cases} \tag{24}$$

obviously, $S(T) = k - 1$, where k denotes as the number of forecasting steps ahead.

The test was performed at the 0.05 and 0.10 significance levels in two-tail-tests under the null hypothesis of equal forecast accuracy for the HSVGM model and the SVRCACO model. The test results (Table 3) showed that the SVRCACO model only yields significantly improved forecast accuracy in comparison with the HSVMG model in FF/USD and DM/USD currencies. Altogether, with the forecasting results at hand, we could only conclude that the CACO may be a useful alternative evolutionary algorithm to determine more suitable parameter combination in a SVR model in terms of FF/USD and DM/USD currencies exchange rates forecasting.

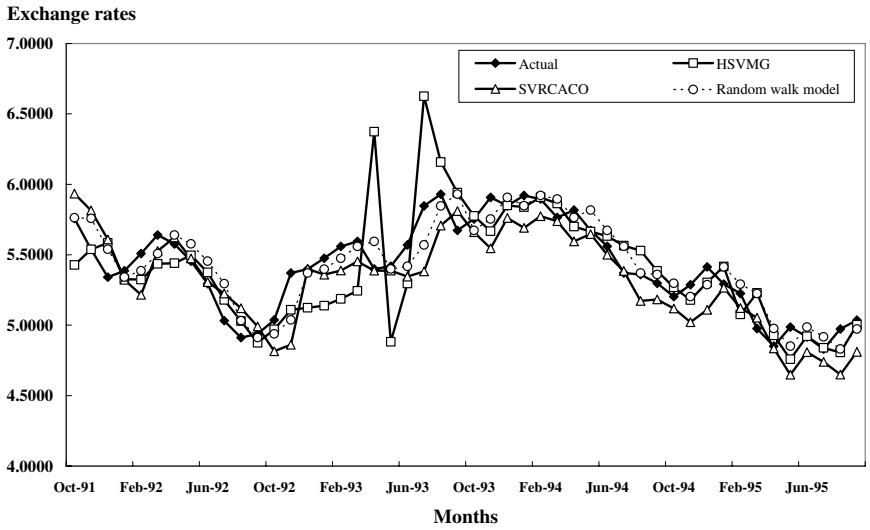

Figure 6. Monthly exchange rates of French Franc (FF)

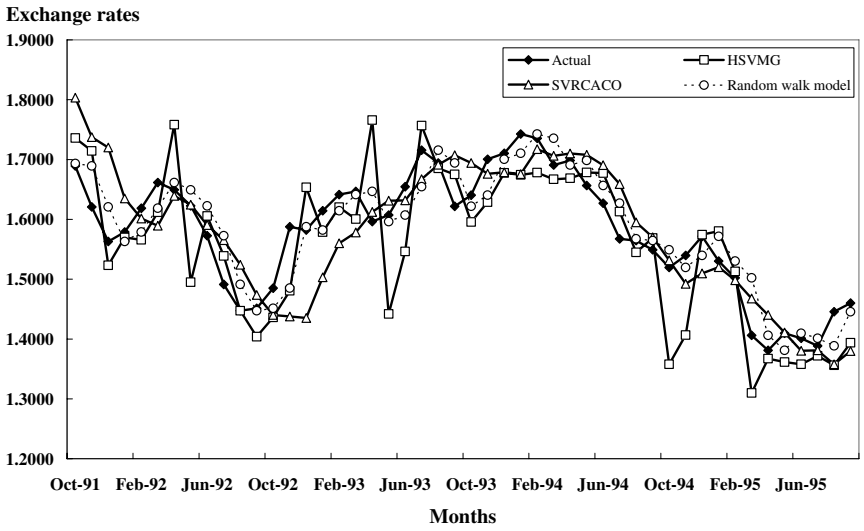

Figure 7. Monthly exchange rates of Deutschmark (DM)

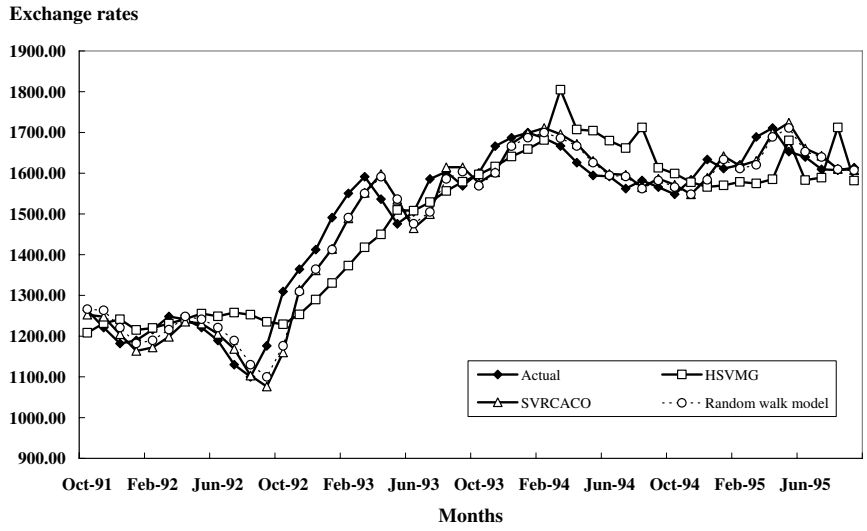

Figure 8. Monthly exchange rates of Italian Lira (LIT)

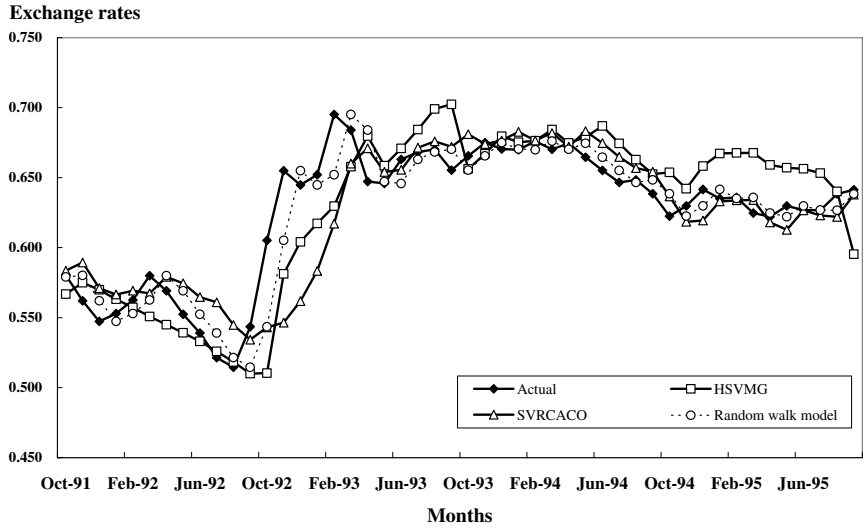

Figure 9. Monthly exchange rates of British Pound (GBP)

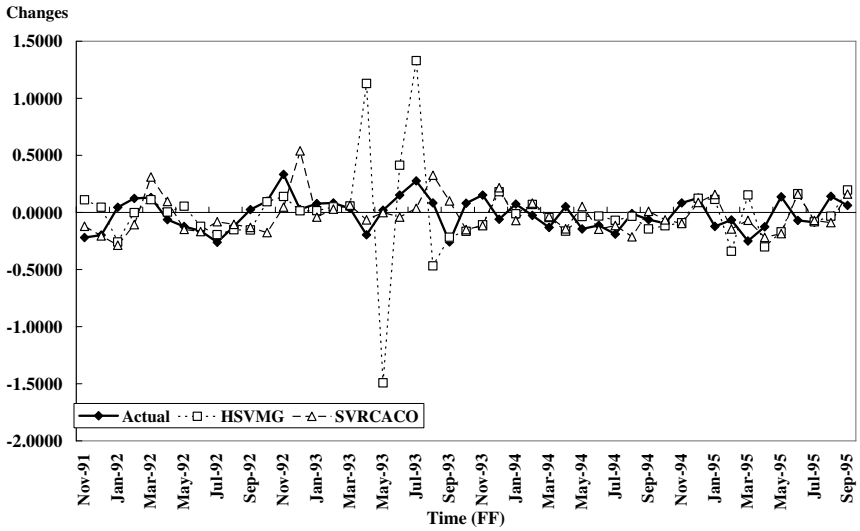

Figure 10. Actual and forecasted changes of the exchange rate of French Franc (FF)

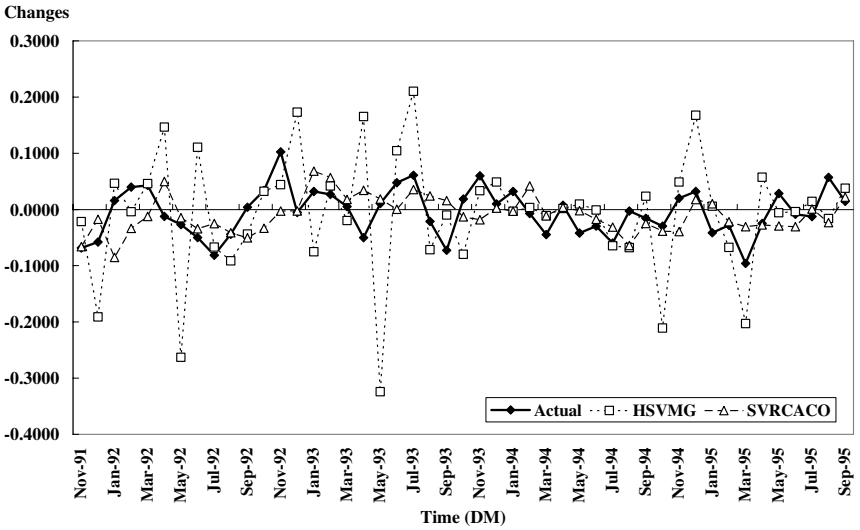

Figure 11. Actual and forecasted changes of the exchange rate of Deutschmark (DM)

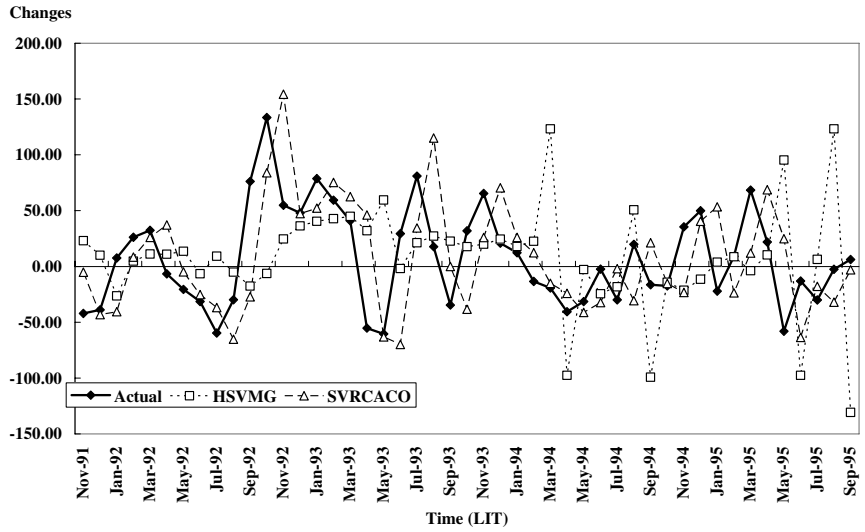

Figure 12. Actual and forecasted changes of the exchange rate of Italian Lira (LIT)

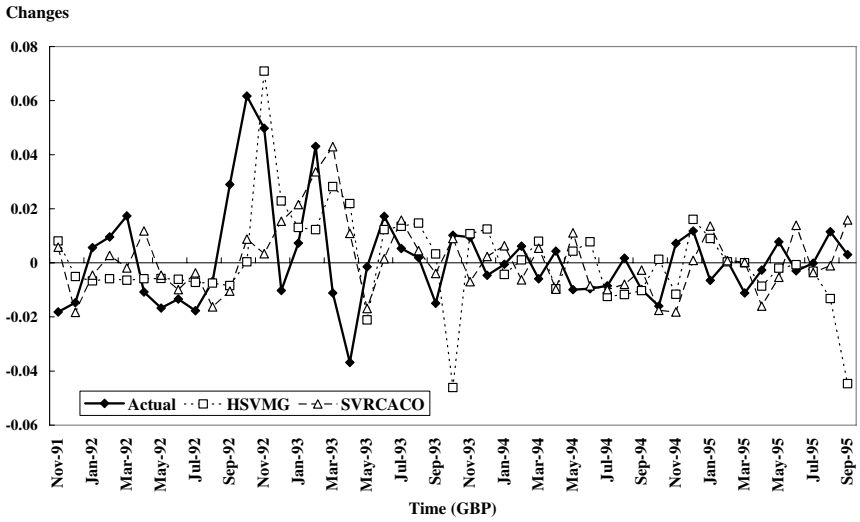

Figure 13. Actual and forecasted changes of the exchange rate of British Pound (GBP)

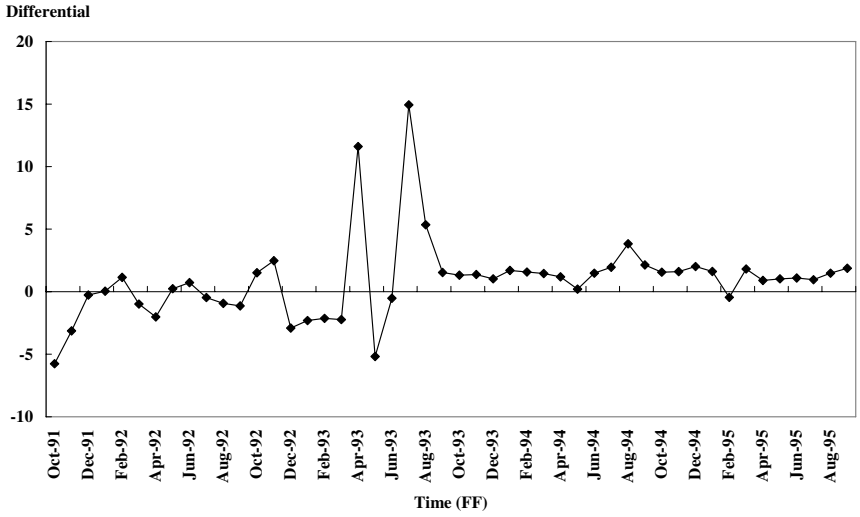

Figure 14. Loss differential (HSVMG-SVRCACO) of French Franc (FF)

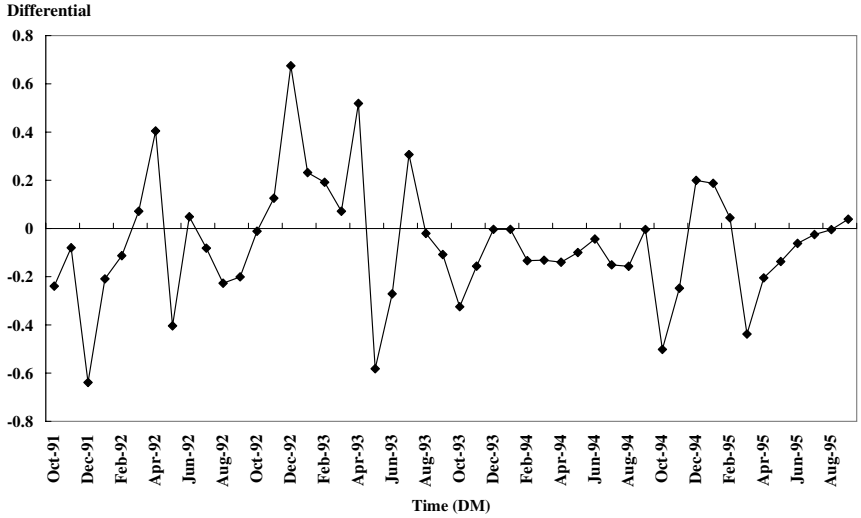

Figure 15. Loss differential (HSVMG-SVRCACO) of Deutschmark (DM)

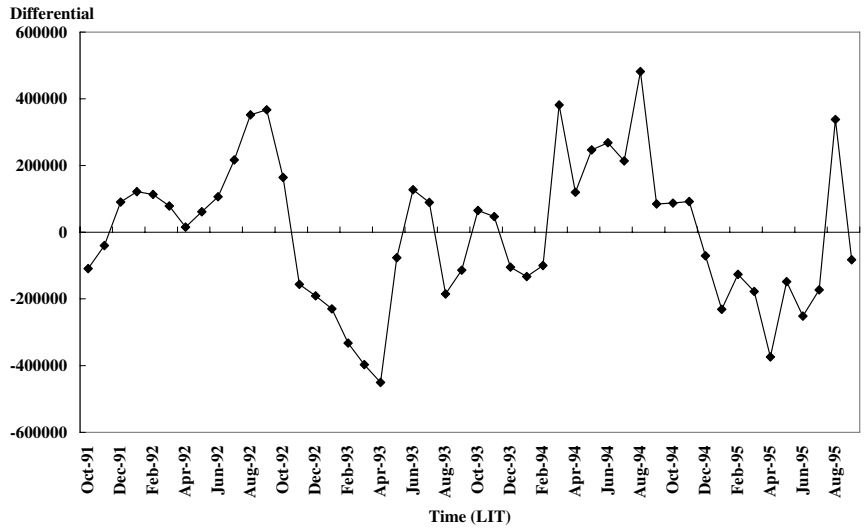

Figure 16. Loss differential (HSVMG-SVRCACO) of Italian Lira (LIT)

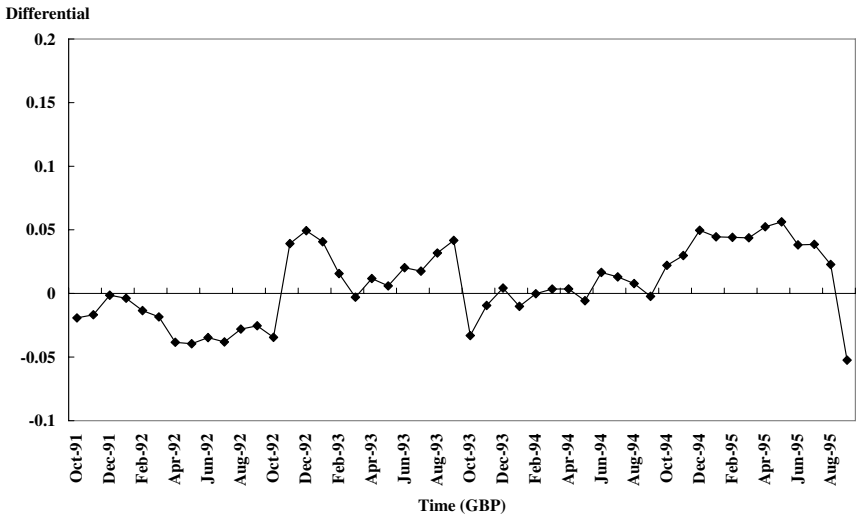

Figure 17. Loss differential (HSVMG-SVRCACO) of British Pound (GBP)

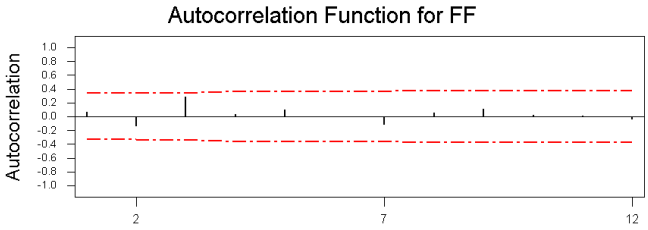

Figure 18. Loss differential autocorrelations of French Franc (FF)

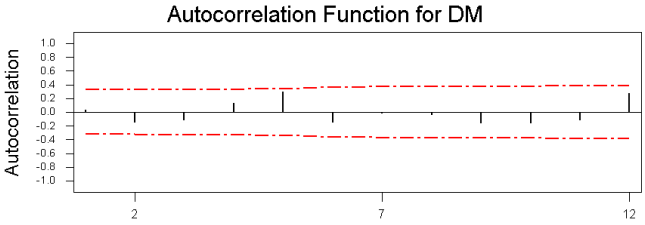

Figure 19. Loss differential autocorrelations of Deutschmark (DM)

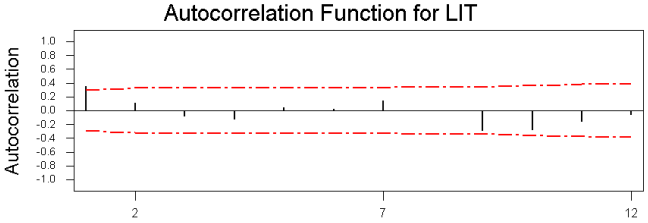

Figure 20. Loss differential autocorrelations of Italian Lira (LIT)

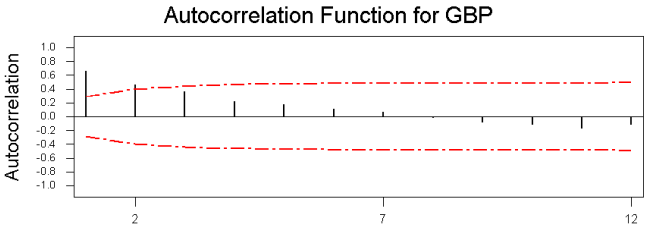

Figure 21. Loss differential autocorrelations of British Pound (GBP)

Table 3. Asymptotic test

Currencies exchange rates	Asymptotic (S_1) test	
	α=0.05	α=0.10
FF/USD HSVMG v.s SVRCACO	$H_0 : e_1 = e_2$ S_1=1.913; p-value=0.0277 (reject H_0)	$H_0 : e_1 = e_2$ S_1=1.913; p-value=0.0277 (reject H_0)
DM/USD HSVMG v.s SVRCACO	$H_0 : e_1 = e_2$ S_1=1.707; p-value=0.0440 (not reject H_0)	$H_0 : e_1 = e_2$ S_1=1.707; p-value=0.0440 (reject H_0)
LIT/USD HSVMG v.s SVRCACO	$H_0 : e_1 = e_2$ S_1=1.126; p-value=0.1296 (not reject H_0)	$H_0 : e_1 = e_2$ S_1=1.126; p-value=0.1296 (not reject H_0)
GBP/USD HSVMG v.s SVRCACO	$H_0 : e_1 = e_2$ S_1=0.936; p-value=0.1748 (not reject H_0)	$H_0 : e_1 = e_2$ S_1=0.936; p-value=0.1748 (not reject H_0)

4. Conclusions

Accurate exchange rate forecasting is crucial for researchers and practitioners to reduce the fluctuation *clustering* effect that worsens the efficiency and effectiveness of time series prediction, particularly for avoiding large investment risk of an enterprise during international business purchasing. The historical exchange rate data of exemplary four currencies in this paper show a fluctuation trend, which occurs in many currency exchange rate data. Therefore, overshoot or undershoot of exchange rates influences the hedging decisions of an enterprise. This study introduces the application of a forecasting technique, SVRCACO, to investigate its feasibility for forecasting currency exchange rates. The experimental results indicate that the SVRCACO model has better forecasting performance than the HSVMG, NN, VLLR, and RW models. The superior performance of the SVRCACO model is firstly due to the generalization ability of SVR model for forecasting and the proper selection of SVR parameters by the CACO algorithm. Secondly, the SVR method employs the quadratic programming technique which is based on the assumptions of convex set and existence of global optimum solution. Therefore, SVR could theoretically approximate global optimum solution if superior search algorithms were employed. As it is known, genetic algorithms lack knowledge memory or storage functions, *previous knowledge of the problem being associated with the population.* Thus, this drawback of GA would lead to time consuming and inefficient search for suitable parameters of an SVR model, which also often suffers from being trapped in local optimum (Angeline, 1998; Liu et al., 2005). ACO is designed with knowledge memory (pheromone updating) to efficiently learn and quickly search, and, in this paper, three homogeneous ant colonies (for determination of three pa-

rameters) could independently and simultaneously establish the shortest path between ant nest and food. Thus, the algorithm is capable of avoiding being trapped in local optimum and of maintaining forecasting accuracy.

Concerning forecasting accuracy improvement, based on the accuracy improvement significance test, it was shown that the SVRCACO model could significantly outperform the HSVMG model only if minor changes of the actual exchange rate occurred (like for FF/USD and DM/USD), because minor changes often generate several local minima; on the other hand, while changes of the actual exchange rate are important (like for LIT/USD and GBP/USD), the HSVMG model could significantly outperform the SVRCACO model as there were less local minima.

This study is the first application of SVR with CACO for forecasting currency exchange rates. Many forecasting methodologies have been proposed to deal with the fluctuation clustering effects. However, most models are time consuming in verifying the suitable time-phase divisions, particularly when the sample size is large. In this investigation, the SVRCACO model was shown to provide a convenient and valid alternative for exchange rates forecasting. The SVRCACO model directly uses historical observations from currency exchange rate data and then determines suitable parameters by efficient optimization algorithms. The next step would be to develop strategies to involve other factors and meteorological control variables during investment period, such as critical social events (e.g., terrorist attacks), the percentage of foreign direct investment of the target country, and the hedging financial goods selection can be included in the exchange rate forecasting model. In addition, as the proposed SVRCACO model is a hybrid forecasting model, other advanced optimization algorithms for parameter selection can be applied for the SVR model to satisfy the requirement of real-time currency exchange rate data. The goal of the authors was to show that combination of novel techniques is at least as good as the pure techniques.

Acknowledgment

The authors thank the anonymous reviewers for pointing out several analytical issues that have improved the study considerably, and for patiently guiding authors in improving the quality of the manuscript and in synthesizing the message from this study. This research was conducted with the support of National Science Council, Taiwan, project numbers NSC 97-2410-H-161-001, NSC 98-2410-H-161-001 and NSC 98-2811-H-161-001.

References

Abbaspour, K.C., Schulin, R. and van Genuchten, M.T. (2001) Estimating unsaturated soil hydraulic parameters using ant colony optimization. *Advances in Water Resources* **24**, 827-841.

Amari, S. and Wu, S. (1999) Improving support vector machine classifiers by modifying kernel functions. *Neural Networks* **12**, 783-789.

Angeline, P.J. (1998) Evolutionary optimization versus particle swarm optimization: philosophy and performance differences. *Evolutionary Programming* **6**, 601-610.

Beran, J. and Ocker, D. (1999) SEMIFAR forecasts, with applications to foreign exchange rates. *Journal of Statistical Planning and Inference* **80**, 137-153.

Bilchev, G. and Parmee, I.C. (1995) The ant colony metaphor for searching continuous design spaces. *Lecture Notes in Computer Sciences* **993**, 25-39.

Bland, J.A. (1999) Space-planning by ant colony optimization. *Int. J. of Computer Applications in Technology* **12**, 320-328.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-327.

Brooks, C. (1997) Linear and non-linear (non-)forecastability of high frequency exchange rates. *Journal of Forecasting* **15**, 125-145.

Cao, L. (2003) Support vector machines experts for time series forecasting. *Neurocomputing* **51**, 321-339.

Chen, A.S. and Leung, M.T. (2004) Regression neural network for error correction in foreign exchange forecasting and trading. *Computers and Operations Research* **31**, 1049-1068.

Chun, S.H. and Kim, S.H. (2003) Impact of momentum bias on forecasting through knowledge discovery techniques in the foreign exchange market. *Expert Systems with Applications* **24**, 115-122.

Colorni, A., Dorigo, M., Maniezzo, V. and Trubian, M. (1994) Ant system for job-shop scheduling. *JORBEL-Belgian Journal of Operations Research, Statistics and Computer Science* **34**, 39-53.

Cortes, C. and Vapnik, V. (1995) Support vector networks. *Machine Learning* **20**, 273-297.

Cristianini, N. and Shawe-Taylor, J. (2000) *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, Cambridge, Massachusetts.

Davis, J.T., Episcopos, A. and Wettimuny, S. (2001) Predicting direction shifts on Canadian–US exchange rates with artificial neural networks. *Int. J. of Intelligent Systems in Accounting* **10**, 83-96.

Diebold, F.X., Gardeazabal, J. and Yilmaz, K. (1994) On cointegration and exchange rate dynamics. *Journal of Finance* **49**, 727-735.

Diebold, F. X. and Mariano, R. S. (1995) Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13** (3), 134-144.

Dorigo, M. (1992) *Optimization, Learning, and Natural Algorithms* (Doctoral dissertation). Dip. Elettronica e Informazione, Politecnico di Milano, Italy.

Dorigo,M. and Gambardella, L. (1997) Ant colony system: a cooperative

learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation* **1**, 53-66.

Dorigo, M., Maniezzo, V. and Colorni, A. (1996) Ant system: optimization by a colony of cooperating ants. *IEEE Transactions on System, Man and Cybernetics Part B* **26**, 29-41.

Dreo, J. and Siarry, P. (2002) A new ant colony algorithm using the hierarchical concept aimed at optimization of multiminima continuous functions. *Lecture Notes in Computer Sciences* **2463**, 216-221.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**, 987-1008.

Fernandes, M. (1998) Non-linearity and exchange rates. *Journal of Forecasting* **17**, 497-514.

Fernandez-Rodriguez, F., Sosvilla-Rivero, S. and Andrada-Felix, J. (1999) Exchange rate forecasts with simultaneous nearest-neighbor methods: evidence from the EMS. *Int. J. of Forecasting* **15**, 383-392.

Hann, T.H. and Steurer, E. (1996) Much ado about nothing? Exchange rate forecasting: neural networks vs. linear models using monthly and weekly data. *Neurocomputing* **10**, 323-339.

Haykin, S. (1999) *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, Upper Saddle River, NJ.

Hong, W.C. and Pai, P.F. (2006) Predicting engine reliability by support vector machines. *Int. J. of Advanced Manufacturing Technology* **28**, 154-161.

Hong, W.C. and Pai, P.F. (2007) Potential assessment of the support vector regression technique in rainfall forecasting. *Water Resources Management* **21**, 495-513.

Hong, W.C., Pai, P.F., Chen, C.T. and Chang, P.T. (2005) Recurrent support vector machines with genetic algorithms in reliability prediction. *Lecture Notes in Computer Science* **3610**, 619-629.

Huang, W., Nakamori, Y. and Wang, S.Y. (2005) Forecasting stock market movement direction with support vector machine. *Computers & Operations Research* **32**, 2513-2522.

Kilian, L. and Taylor, M.P. (2003) Why is it so difficult to beat random walk forecast of exchange rates? *Journal of International Economics* **60**, 85-107.

Kodogiannis, V. and Lolis, A. (2002) Forecasting financial time series using neural network and fuzzy system-based techniques. *Neural Computing and Applications* **11**, 90-102.

Leung, M.T., Chen, A.S. and Daouk, H. (2000) Forecasting exchange rates using general regression neural networks. *Computers & Operations Research* **27**, 1093-1110.

Lisi, F. and Schiavo, R.A. (1999) A comparison between neural networks and chaotic models for exchange rate prediction. *Computational Statistics*

& Data Analysis **30**, 87-102.

Liu, B., Wang, L., Jin, Y.H., Tang, F. and Huang, D.X. (2005) Improved particle swarm optimization combined with chaos. *Chaos, Solitons & Fractals* **25** (5), 1261-1271.

Lu, W., Wang, W., Leung, A.Y.T., Lo, S.M., Yuen, R.K.K., Xu, Z. and Fan, H. (2002) Air pollutant parameter forecasting using support vector machines. *Proceedings of the 2002 International Joint Conference of Neural Networks*, IEEE, 12-17.

Maniezzo, V. and Colorni, A. (1999) The ant system applied to the quadratic assignment problem. *IEEE Transactions on Knowledge and Data Engineering* **11**, 769-778.

Mathur, M., Karale, S.B., Priye, S., Jyaraman, V.K. and Kulkarni, B.D. (2000) Ant colony approach to continuous function optimization. *Industrial & Engineering Chemistry Research* **39**, 3814-3822.

Mohandes, M.A., Halawani, T.O., Rehman, S. and Hussain, A.A. (2004) Support vector machines for wind speed prediction. *Renewable Energy* **29**, 939-947.

Monmarche, N., Venturini, G. and Slimane, M. (2000) On how Pachycondyla apicalis ants suggest a new search algorithm. *Future Generation Computer Systems*, **16**, 937-946.

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

Pai, P.F. and Hong, W.C. (2005a) Forecasting regional electric load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research* **74**, 417-425.

Pai, P.F. and Hong, W.C. (2005b) Support vector machines with simulated annealing algorithms in electricity load forecasting. *Energy Conversion and Management* **46**, 2669-2688.

Pai, P.F. and Hong, W.C. (2006) Software reliability forecasting by support vector machines with simulated annealing algorithms. *Journal of Systems and Software* **79**, 747-755.

Pai, P.F. and Lin, C.S. (2005a) A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* **33**, 497-505.

Pai, P.F. and Lin, C.S. (2005b) Using support vector machines to forecast the production values of the machinery industry in Taiwan. *The Int. J. of Advanced Manufacturing Technology* **27**, 205-210.

Pai, P.F., Lin, C.S., Hong, W.C. and Chen, C.T. (2006) A hybrid support vector machine regression for exchange rate prediction. *Int. J. of Information and Management Sciences* **17**, 19-32.

Parpinelli, R.S., Lopes, H.S. and Freitas, A.A. (2002) Data mining with an ant colony optimization algorithm. *IEEE Transactions on Evolutionary Computation* **6** 321-332.

Qi, M. and Wu, Y. (2003) Nonlinear prediction of exchange rates with mon-

etary fundamentals. *Journal of Empirical Finance* **10**, 623-640.

Scholkopf, B. and Smola, A.J. (2002) *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, Cambridge, Massachusetts.

Shawe-Taylor, J., Bartlett, P.J., Williamson, R.C. and Anthony, M. (1998) Structural risk minimization over data-dependent hierarchies. *IEEE Transactions on Information Theory* **44**, 1926-1940.

Socha, K. and Dorigo, M. (2006) Ant colony optimization for continuous domains. *EJOR*, doi:10.1016/j.ejor.2006.06.046.

Suykens, J.A.K. (2001) Nonlinear modelling and support vector machines. In: *Proceedings of IEEE Instrumentation and Measurement Technology Conference*, 287-294.

Tay, F.E.H. and Cao, L. (2001) Application of support vector machines in financial time series forecasting. *Omega* **29**, 309-317.

Tay, F.E.H. and Cao, L. (2002) Modified support vector machines in financial time series forecasting. *Neurocomputing* **48**, 847-861.

Trapletti, A., Geyer, A. and Leisch, F. (2002) Forecasting exchange rates using cointegration models and intra-day data. *Journal of Forecasting* **21**, 151-166.

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

Vapnik, V. (1998) *Statistical Learning Theory*. Wiley, New York.

Vilasuso, J. (2002) Forecasting exchange rate volatility. *Economic Letters* **76**, 59-64.

Vojislav, E. (2001) *Learning and Soft Computing - Support Vector Machines, Neural Networks and Fuzzy Logic Models*. The MIT Press, Cambridge, Massachusetts.

Wodrich, M. and Bilchev, G. (1997) Cooperative distributed search: the ant's way. *Control and Cybernetics* **26**, 413-446.

Wu, B. (1995) Model-free forecasting for nonlinear time series (with application to exchange rates). *Computational Statistics & Data Analysis* **19**, 433-459.

Yao, J. and Tan, C.L. (2000) A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing* **34**, 79-98.

Zhang, G. and Hu, M.Y. (1998) Neural network forecasting of the British Pound/US Dollar Exchange Rate. *Omega* **26**, 495-506.