Grey Support Vector Regression Model with Applications to China Tourists Forecasting in Taiwan

doi:10.6186/IJIMS.2014.25.2.3

International Journal of Information and Management Sciences, 25(2), 2014

作者/Author： Ruey-Chyn Tsaur;Shu-Feng Chan

頁數/Page： 121-138

出版日期/Publication Date：2014/07

International Journal of Information and Management Sciences
25 (2014), 121-138

International Journal of
Information
and
Management
Sciences

Grey Support Vector Regression Model with Applications to China Tourists Forecasting in Taiwan

Ruey-Chyn Tsaur and Shu-Feng Chan

Tamkang University

Abstract

Support vector regression (SVR) has been successful in function approximation for forecasting analysis based on the idea of structural risk minimization. SVR has perfect forecasting performance by employing in large sample size for training and solving its parameters, where the SVR is difficult to be applied in limited time series data with some fluctuated points; in contrast, grey model has better forecasting performance in limited time series data. In order to cope with this problem, we use both of the advantages of support vector regression model and grey theory to construct a new grey support vector regression (GSVR) model for solving limited data with some fluctuations. Finally, we demonstrate an application for planning China tourism demand for improving the tourism infrastructure in Taiwan with a better forecasting performance.

Keywords: Grey support vector regression, grey theory, support vector regression, tourism demand forecasting.

1. Introduction

Forecasting is an important part of decision-making for sales planning, marketing research, pricing, and production planning. In the historical development, exponential smoothing methods and decomposition methods were the first forecasting approaches to be developed in the mid-1950s, and then time series models as the autoregressive integrated moving average models for increasing the forecasting accuracy whereas the complexity of forecasting approaches is not always guaranteed to obtain the desired predictive accuracy [12][16]. Support vector machines (SVM) developed by Vapnik in 1995 based on the idea of structural risk minimization, then many researchers devoted themselves to research on SVM over the past few years. Next, the functional dependence of the dependent variable and independent variables are estimated, and a loss function [6] is applied to formulate support vector regression (SVR) model, as long as the actual data values lie within a distance of from the predicted values which assumed that there is no prediction error. Therefore, the predictions from an SVR can be regarded as an -tube and extends the traditional linear regression to nonlinear regression by introducing kernel functions which maps the input vectors into higher feature space, and thus a linear regression hyperplane is considered in the feature space for nonlinear forecasting analysis [1]. Because SVR can fit linear or nonlinear problems, thereafter, many researchers devote themselves into SVR forecasting. Tay and Cao [15] used support vector machine in financial time series forecasting; Hong and Hwang [7] introduced the multivariate fuzzy linear and nonlinear regression SVM models which give computational efficiency of getting solutions; Levis and Papageorgiou [10] proposed three-step algorithm comprises both nonlinear programming and linear programming model to determine the SVR function and employs a recursive methodology to perform customer demand forecasting; Chen and Wang [3] proposed a genetic algorithm to construct the SVR models to forecast the tourism demand; Hong [9] proposed a hybrid evolutionary algorithm to the SVR model whose empirical results indicate that the SVR model has better forecasting performance; Lu et al. [11] used independent component analysis and support vector regression for financial time series forecasting which showed that the proposed model outperforms the SVR model with non-filtered forecasting variables and a random walk model; Farguad et al. [5] used hybrid rule extraction methods to forecast city-cycle fuel consumption; Hong et al. [8] used genetic algorithmsimulated annealing hybrid SVR algorithm model for 3G mobile phones demand forecasting in Taiwan with dominated performance; Shieh and Kuo [14] proposed a reduced data set method for support vector regression to reduce the computational time of SVR and greatly mitigates the influence of data noise and outliers; Che and Wang [2] proposed a hybrid model that combines both SVR and Auto-regressive integrated moving average models whose experimental results demonstrate that the model proposed outperforms the existing neural-network approaches, the traditional ARIMA models and other hybrid models.

However, for a specified event with limited time series, it is usually difficult to use such limited information to forecast the future trend. Deng [4] developed grey theory whose systems characterized by poor information and/or for which information is lacking, and its fields include systems analysis, data processing, modeling, prediction, as well as decision making and control. From many literatures, we believe that the grey theory is an effective mathematical means for resolving problems containing uncertainty and indetermination whereas grey model is worse to forecast with some fluctuated datum. By contrast, SVR can work when there are fluctuated datum but in vain to derive the upper vector regression equation because of no support vectors when the collected data is limited. In this paper, a new grey support vector regression model (GSVR) is proposed to overcome the above problems. We take the advantage of grey model $GM(1,1)$ using a one-order accumulated generating operation (AGO) to assemble the collected time series data and obtain internal regularity to manage the disorganized original data, combine into SVR model to minimize the structural risk, and then propose a grey support vector regression model (GSVR). An experiment is illustrated with some sensitivity analyses, and their results have shown the proposed GSVR model has better performance.

The rest of this paper is organized as follows. Section 2 includes reviews on the grey model $GM(1,1)$ and SVR model. Section 3 introduces our proposed GSVR model.

In Section 4, we illustrate the use of this model for tourism forecasting by taking an example. Section 5 includes concluding remarks.

2. Reviewing Models

2.1. Grey Model $GM(1,1)$

The general form of the grey model $GM(k,N)$ [4], defined as a linear differential equation, where k stands for the kth order derivative of the AGO series of dependent variables, and N stands for the number of variables (i.e. one dependent variable and $N-1$ independent variables), is as follows

$$\frac{d^k F_t^1}{dt^k}+a_1\frac{d^{k-1}F_t^1}{dt^{k-1}}+\cdots+a_{k-1}\frac{dF_t^1}{dt}+a_kF_t^1 = b_1X_1^1(t)+b_2X_2^1(t)+\cdots+b_{N-1}X_{N-1}^1(t) \quad (2.1)$$

where F_t^1 is the dependent variable, and $X_1^1(t), X_2^1(t), \ldots, X_{N-1}^1(t)$ are independent variables which are obtained by inputting the AGO input values from the time series set $f^0 = \{f_t^0 \mid t \in 1,2\ldots,n\}$ and $x_i^0 = \{x_i^0(t) \mid t \in 1,2,\ldots,n\}$, $\forall\ i = 1,2,\ldots,N-1$, respectively (the AGO formulation is shown later, see Eq. (2.3)), and $a_1, a_2, \ldots, a_k$ and $b_1, b_2, \ldots, b_{N-1}$ are unknown parameters. If $k = 1$ and $N = 1$, then the grey model $GM(1,1)$ with one order differential equation and one dependent variable model can be constructed. In order to construct a grey model, Deng applied an one order accumulated generating operation (AGO) to assemble the collected time series data and obtain internal regularity in order to manage the disorganized original data. This means that if an original time series set f^0 is defined as

$$f^0 = \{f_t^0 \mid t \in 1,2\ldots,n\} \quad (2.2)$$

where t denotes the number of data observed in period t, then the input value for the dependent variable F_t^1 is obtained from AGO value of original time series f_t^0 as

$$f_t^1 = \left(\sum_{k=1}^{t} f_k^0\right) \qquad t = 1,2,\ldots,n \quad (2.3)$$

Therefore, the new set of the AGO series f^1 is monotonically increasing obtained as $f^1 = \{f_t^1 \mid t \in 1,2\ldots,n\}$. Then, the grey model $GM(1,1)$ is constructed as

$$\frac{dF_t^1}{dt} + dF_t^1 = b, \qquad t = 1,2,\ldots,n \quad (2.4)$$

where a represents the unknown developed parameter, b represents the unknown grey controlled parameter, and F_t^1 is the dependent variable with AGO input value f_t^1. For solving model (2.4), the derivative $\frac{dF_t^1}{dt}$ for dependent variable is represented as

$$\frac{dF_t^1}{dt} = \lim_{h\to 0}\frac{F_{t+h}^1 - F_t^1}{h}, \qquad \forall t \geq 1. \quad (2.5)$$

Because the collected data is a set of time series, we assume the sampling time interval between period t and $t+1$ to be one unit. Then, the derivative $\frac{dF_t^1}{dt}$ is approximated to an inverse accumulated generating operation (IAGO) variable F_{t+1}^0 as

$$\frac{dF_t^1}{dt} \approx \frac{F_{t+1}^1 - F_t^1}{1} = F_{t+1}^1 - F_t^1 = F_{t+1}^0, \qquad \forall\, t \geq 1 \tag{2.6}$$

for the original $(t+1)$th time series data f_{t+1}^0, $\forall\, t \geq 1$. In order to have a more steady value for the dependent variable F_t^1, $\forall\, t \geq 1$, the second part of model (2.4) is suggested as the average of two successive periods of F_t^1 and F_{t+1}^1, $\forall\, t \geq 1$. Then, we can rewrite model (2.4) as

$$F_{t+1}^0 = -a\Big[\frac{1}{2}(F_{t+1}^1 + F_t^1)\Big] + b, \qquad \forall\, t \geq 1. \tag{2.7}$$

If $t = 1, 2, \ldots, n$, then (2.7) can be rewritten into matrix form as

$$\begin{bmatrix} F_2^0 \\ F_3^0 \\ \vdots \\ F_n^0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}(F_2^1 + F_1^1) & 1 \\ -\frac{1}{2}(F_3^1 + F_2^1) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(F_n^1 + F_{n-1}^1) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}. \tag{2.8}$$

By applying the least square method with input data sets f^1 and f^0, the parameters of a and b in matrix $\hat{\boldsymbol{a}}$ can be solved as

$$\hat{\boldsymbol{a}} = \begin{bmatrix} a \\ b \end{bmatrix} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{F}^0 \tag{2.9}$$

where matrices $\boldsymbol{F}^0 = \begin{bmatrix} F_2^0 \\ F_3^0 \\ \vdots \\ F_n^0 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} -\frac{1}{2}(F_2^1 + F_1^1) & 1 \\ -\frac{1}{2}(F_3^1 + F_2^1) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(F_n^1 + F_{n-1}^1) & 1 \end{bmatrix}$ and $\boldsymbol{B}^T$ are the transpose of matrix $\boldsymbol{B}$. Then, the differential equation of model (2.4) can be solved to obtain the estimated value $\hat{f}_{t+1}^1$ for the dependent variable F_t^1, $\forall\, t \geq 1$ as

$$\hat{f}_{t+1}^1 = (f_1^0 - (b/a))e^{-at} + (b/a), \qquad \forall\, t \geq 1. \tag{2.10}$$

Finally, the estimated value $\hat{f}_{t+1}^0$ for the IAGO variable F_{t+1}^0 is obtained as

$$\hat{f}_{t+1}^0 = \hat{f}_{t+1}^1 - \hat{f}_t^1, \qquad \forall\, t = 1, 2, \ldots, n. \tag{2.11}$$

Therefore, by inputting the time series $f_1^0, f_2^0, \ldots, f_n^0$ into the grey model $GM(1,1)$, it can obtain the extrapolative value of $\hat{f}_2^0, \hat{f}_3^0, \ldots, \hat{f}_n^0$, and $\hat{f}_{n+1}^0$.

2.2. Support vector regression model

The SVR formulation follows the principle of structural risk minimization, trying to minimize an upper bound of the generalization error rather than minimize the empirical risk. Therefore, SVR can generalize the input-output relationship learned during its training process for deriving good forecasting with any new input data. In addition, in order to make a good forecasting for nonlinear data type, the SVR model is developed to map the input data into a high-dimensional feature space and solve a linear regression problem in the feature space. The regression approximation estimates a function using a given data set $G = \{(x_i, y_i) \mid i = 1, 2, \ldots, n\}$ where x_i denotes the input vector; y_i denotes the output value and n denotes the total number of data patterns. The modelling aim is to identify a regression function, $y = f(x)$, that accurately predicts the outputs $\{y_i \mid i = 1, 2, \ldots, n\}$. Using mathematical notation, the linear regression function is approximated using the following function:

$$f(x_i) = w\phi(x_i) + b, \qquad i = 1, 2, \ldots, n, \tag{2.12}$$

where w and b are coefficients; $\phi(x_i)$ denotes the high dimensional feature space, which is nonlinearly mapped from the input space x. Therefore, the linear regression in the high-dimensional feature space responds to nonlinear regression in the low-dimension input space, disregarding the inner product computation between w and $\phi(x_i)$. The unknown parameters w and b in Eq. (2.12) are estimated using the training set in the high-dimensional feature space by Vapnik's linear loss function with ε- insensitivity interval defined as follows:

$$|y_i - f(x_i, w)|_\varepsilon = \begin{cases} 0 & \text{if } |y_i - f(x_i, w)| \le \varepsilon, \\ |y_i - f(x_i, w)| - \varepsilon, & \text{otherwise}, \end{cases} \tag{2.13}$$

where $\frac{1}{2}\|w\|^2$ is minimized to ensure minimum ε deviation. Then, the problem can be written as a convex optimization problem:

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\|w\|^2 \\ \text{Subjected to} \quad & y_i - w\phi(x_i) - b \le \varepsilon, \qquad i = 1, 2, \ldots, n, \\ & w\phi(x_i) + b - y_i \le \varepsilon, \qquad i = 1, 2, \ldots, n. \end{aligned} \tag{2.14}$$

The assumption in model (2.14) is that such a SVR function $f(x)$ actually exists that approximates all pairs (x_i, y_i) with ε precision. However, in the case where constraints are infeasible in model (2.14), slack variables ξ_i and ξ_i^* were introduced. This case is called soft margin formulation as shown as Figure 1, model (2.14) is revised as follows:

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i, \xi_i^*) \\ \text{Subjected to} \quad & y_i - w\phi(x_i) - b \le \varepsilon + \xi_i, \\ & w\phi(x_i) + b - y_i \le \varepsilon + \xi_i^*, \\ & \xi, \xi^* \ge 0, \ \forall\, i = 1, 2, \ldots, n. \end{aligned} \tag{2.15}$$

The constant $C > 0$ denotes a cost function measuring the empirical risk which deviations larger than ε are tolerated. It is clearly that in nonlinear optimization problem (2.15) can be solved more easily in its dual formulation.

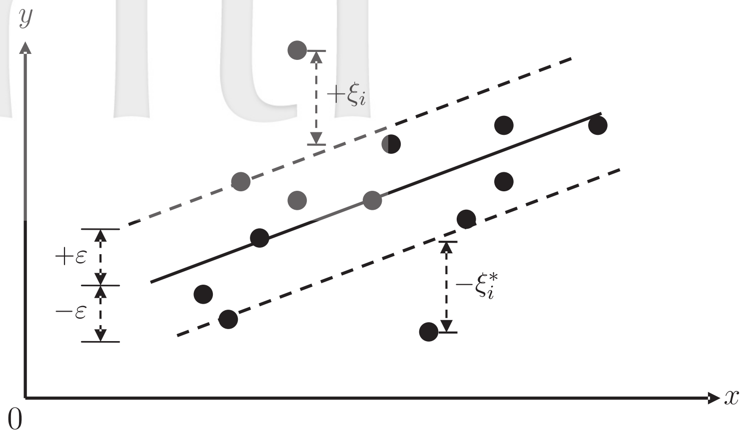

Figure 1: The soft margin loss for a linear SVR.

Therefore, Lagrange multipliers are used to get the dual formulation as described as follows:

$$\begin{aligned}
\text{Max} \quad & \sum_{i=1}^{n} y_i(\alpha_i-\alpha_i^*) - \varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x_j)\\
\text{s.t.} \quad & \sum_{i=1}^{n} y_i(\alpha_i-\alpha_i^*)=0, \qquad \begin{matrix} 0\le\alpha_i\le C\\ 0\le\alpha_i^*\le C\end{matrix} \qquad i=1,2,\ldots,n.
\end{aligned} \tag{2.16}$$

Based on the KarushKuhnTucker's (KKT) conditions in model (2.16), the approximation errors of data point on non-zero coefficient will be equal to or larger than ε , and are referred to as the support vector. That is, these data points lie on or outside the ε bound of the decision function. Generally, the larger that of ε value, the fewer the number of support vectors. Nevertheless, increasing ε decreases the approximation accuracy of training data. In this sense, the parameter $w=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)\phi(x_i)$, and the SVR is $f(x)=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)K(x_i,x)+b$.

3. Grey Support Vector Regression Model

For a time series set $x^0=\{x_t^0 \mid t\in 1,2,\ldots,n\}$, an AGO time series set $x^1=\{x_t^1 \mid t\in 1,2,\ldots,n\}$ is obtained as $x_t^1=\left(\sum_{t=1}^{n} x_t^0\right)$, $t=1,2,\ldots,n$. Then a grey model $GM(1,1)$ can be defined as follows:

$$X_{t+1}^0 = wX_t^1 + b, \qquad \forall\ t\ge 1 \tag{3.1}$$

where w represents the unknown developed parameter, b represents the unknown grey controlled parameter, and AGO value X_t^1 is the independent variable used for forecasting the period $(t+1)$-th value X^0_{t+1}. Then, a grey support vector regression can be defined as follows:

$$f(x^1_t) = wX^1_t + b. \tag{3.2}$$

Following the principle of structural risk minimization, and trying to minimize an upper bound of the generalization error, the grey support vector regression (GSVR) is aimed to identify a regression function, $f(x^1_t) = wX^1_t + b$, that accurately predicts the outputs $\{x^0_{t+1}\}$. For nonlinear type data, the GSVR can also maps the high-dimensional feature space responds to the low-dimension space by using the concept of kernel function. The unknown parameters w and b in Eq. (3.2) is estimated by the training set in the high-dimensional feature space using Vapnik's linear loss function with ε - insensitivity interval. Since $\frac{1}{2}\|w\|^2$ is minimized to ensure minimum ε deviation, the GSVR problem can be written as a convex optimization problem:

$$\begin{array}{ll} \text{Min} & \frac{1}{2}\|w^2\| \\ \text{s.t.} & \begin{array}{ll} x^0_{t+1} - wx^1_t - b \le \varepsilon, & t = 1,2,\ldots,n-1 \\ wx^1_t + b - x^0_{t+1} \le \varepsilon^*_t, & t = 1,2,\ldots,n-1 \end{array} \end{array} \tag{3.3}$$

where $\phi(x_i)$ is the high-dimensional feature space. The assumption in model (3.3) is that such a GSVR function $f(x^1_t)$ actually exists that approximates all pairs (x^1_t, x^0_{t+1}) with ε precision. However, in the case where constraints are usually infeasible in model (3.3) when we set a smaller ε value, or there are any fluctuated points, and then some slack variables ξ_i and ξ^*_i were introduced in order to cope with infeasible solution, then model (3.3) can be revised as follows:

$$\begin{array}{lll} \text{Min} & \frac{1}{2}\|w^2\|C\sum_{t=1}^{n-1}(\xi_t + \xi^*_t) & \\ \text{s.t.} & x^0_{t+1} - wx^1_t - b \le \varepsilon + \xi_t, & t = 1,2,\ldots,n-1, \\ & wx^1_t + b - x^0_{t+1} \le \varepsilon + \xi^*_t, & t = 1,2,\ldots,n-1, \\ & \xi_t \ge 0, \xi^*_t \ge 0, & t = 1,2,\ldots,n-1. \end{array} \tag{3.4}$$

Then, model (3.4) can be solved by using well-known Lagrange Multiplier method as follows:

$$\begin{aligned} L(w,b,\xi,\xi^*;\alpha,\alpha^*,\eta,\eta^*) &= \frac{1}{2}w^2 + c\sum_{t=1}^{n-1}(\xi_t + \xi^*_t) - \sum_{t=1}^{n-1}\alpha_t(\varepsilon + \xi_t - x^0_{t+1} + wx^1_t + b) \\ &\quad - \sum_{t=1}^{n-1}\alpha^*_t(\varepsilon + \xi^*_t + x^0_{t+1} - wx^1_t - b) - \sum_{t=1}^{n-1}(\eta_t\xi_t + \eta^*_t\xi^*_t) \end{aligned} \tag{3.5}$$

where $\alpha_t, \alpha^*_t, \eta_t, \eta^*_t \ge 0$ are Lagrange multipliers for every constraints. The optimization problem is solved by finding the saddle point of the Lagrange function, by taking the derivative of the Lagrange with respect to the unknown parameters, i.e., w, b, ξ, ξ^*, as follows:

$$\frac{\partial L}{\partial w} = 0 \quad \rightarrow \quad w = \sum_{t=1}^{n-1} (\alpha_t - \alpha_t^*) x_t^1 \tag{3.6}$$

$$\frac{\partial L}{\partial b} = 0 \quad \rightarrow \quad \sum_{t=1}^{n-1} (\alpha_t - \alpha_t^*) = 0 \tag{3.7}$$

$$\frac{\partial L}{\partial \xi_t} = 0 \quad \rightarrow \quad C - \alpha_t - \eta_t = 0 \tag{3.8}$$

$$\frac{\partial L}{\partial \xi_t^*} = 0 \quad \rightarrow \quad C - \alpha_t^* - \eta_t^* = 0 \tag{3.9}$$

Substituting (3.6)$-$(3.9) into (3.5), we can derive the following dual problem:

$$\begin{aligned}
\operatorname*{Max}_{\alpha,\alpha^*} \quad & -\frac{1}{2} \sum_{k,l=1}^{n-1} (\alpha_k - \alpha_k^*)(\alpha_l - \alpha_l^*)(x_k^1, x_l^1) - \varepsilon \sum_{t=1}^{n-1} (\alpha_t + \alpha_t^*) + \sum_{t=1}^{n-1} x_{t+1}^0 (\alpha_t - \alpha_t^*) \\
\text{s.t.} \quad & \sum_{t=1}^{n-1} (\alpha_t - \alpha_t^*) = 0 \\
& \alpha_t, \alpha_t^* \in [0, C], \qquad t = 1, 2, \ldots, n-1.
\end{aligned} \tag{3.10}$$

When there is a non-linear relationship between data pairs (x_t^1, x_{t+1}^0), a linear GSVR model cannot derive a feasible solution. Therefore, a GSVR using kernel function which maps the input vectors into a higher-dimensional feature space can solve the problem with a linear relationship in the feature space. In practice, the most popular kernel functions are polynomial function $K(x_i^1, x_j^1) = [(x_i^1)^T x_j^1 + 1]^d$, $d > 0$, and radius basis function (RBF) $K(x_i^1, x_j^1) = \exp(-v\|x_i^1 - x_j^1\|^2)$, $v > 0$ where v and d are kernel parameters defined by the users. Then, using the kernel functions, model (3.10) can be rewritten as a nonlinear GSVR model as follows:

$$\begin{aligned}
\operatorname*{Max}_{\alpha,\alpha^*} = & -\frac{1}{2} \sum_{k,l=1}^{n-1} (\alpha_k - \alpha_k^*)(\alpha_l - \alpha_l^*)(x_k^1, x_l^1) - \varepsilon \sum_{t=1}^{n-1} (\alpha_t + \alpha_t^*) + \sum_{t=1}^{n-1} x_{t+1}^0 (\alpha_t - \alpha_t^*) \\
\text{s.t.} \quad & \sum_{t=1}^{n-1} (\alpha_t - \alpha_t^*) = 0 \\
& \alpha_t, \alpha_t^* \in [0, C], \qquad t = 1, 2, \ldots, n-1.
\end{aligned} \tag{3.11}$$

If model (3.11) is maximized, the parameters α_t, α_t^*, $t = 1, 2, \ldots, n-1$ which are the Lagrange multipliers for the first two constraints in the primal model (3.4) can be solved. Then, the new forecasting value for a GSVR can be obtained as follows:

$$f(x_{\text{new}}^1) = \sum_{t=1}^{n-1} (\alpha_t - \alpha_t^*) K(x_t^1, x_{\text{new}}^1) + b, \qquad \forall\, t = 1, 2, \ldots, n-1 \tag{3.12}$$

where x^1_{new} is the new input vector whose output $f(x^1_{\text{new}})$ is to be forecasted, and the bias term b is an estimated value computed from the support vector.

By solving model (3.11), we can find that a nonlinear regression is performed by using kernel function to carry the mapping into the feature space and consider a linear regression hyperplane in the feature space. Therefore, a limited data with any type of distribution such as linear, nonlinear, or any fluctuated points can be efficiently solved when a well-fitted kernel function is chosen in modelling the GSVR model. In other words, the GSVR model takes the advantages of grey model $GM(1,1)$ and SVR model and proposes a new GSVR model for solving limited and fluctuated data in time series analysis. In practice, the forecasting process for GSVR can be described as follows:

Step 1. Define the universal discourse U for the historical data. In this step, the range of the collected time series data can be clearly defined.

Step 2. Scale the collected time series data. In this step, all the collected data is scaled into $[0, 1]$ in order to reduce the influence of the scale for the collected data.

$$S = \frac{L - L_{\min}}{L_{\max} - L_{\min}}, \tag{3.13}$$

where S is the scaling value for the collected data L, $L_{\min}$ is the lowest value in the collected data, and $L_{\max}$ is the uppermost value in the collected data.

Step 3. Calculate the AGO time series. We can derive the AGO time series data being the input values for the independent variable in the GSVR model.

Step 4. Find a well-fitted kernel function. A good kernel function can make the nonlinear data for linear analysis in the feature space.

Step 5. Forecast using the GSVR model. Using the dual model (3.11), we can derive the GSVR, and then use it to forecast the trend of the collected time series data.

Step 6. Rescaled the forecasting values. The forecasting values obtained from step 5 are rescaled to the original scale with the following function.

$$Rs = \left\{\frac{D - D_{\min}}{D_{\max} - D_{\min}}\right\}(L_{\max} - L_{\min}) + L_{\min}, \tag{3.14}$$

where Rs is the rescaling value for the forecasting value D obtained in step 5 in which $D_{\max}$ is the maximum forecasting value and $D_{\min}$ is the minimum forecasting value; $L_{\min}$ is the lowest value and $L_{\max}$ is the uppermost value in the collected data.

Step 7. Forecasting error analysis. After operating the above steps, the forecasting performance is measured in terms of the mean absolute percentage error (MAPE) which is defined as follows:

$$MAPE = \Big(\frac{1}{n-1}\sum_{t=1}^{n-1}\frac{|f(x^1_t) - x^0_{t+1}|}{x^0_{t+1}}\Big)100\% \tag{3.15}$$

where $n-1$ is the number of collected data, x^0_{t+1} is the collected data, and $|f(x^1_t)-x^0_{t+1}|$ is the difference between forecasting value and the collected data.

4. An Illustrated Example for Forecasting Mainland Tourists to Taiwan

4.1. Tourists forecasting using GSVR

Relations between mainland China and Taiwan have been complicated for decades when the KMT government fled to Taiwan. After election in 2008, Taiwan President Ma tried hard to reduce the tensions between mainland China and Taiwan, and then many tourists from the mainland have made the trip to Taiwan. The official with the Taiwan Strait Tourism Association said that the number of mainland tourists to Taiwan in 2010 is expected to reach 1 million. Since, mainland tourists to Taiwan can contribute tourist market rapidly expand which is not only conducive to the development of travel and hotel industries, but also transport and retail industries. Then, the government hopes enormous employee opportunities can effectively stimulate the domestic market during the financial crisis. Because the more that of tourists from the mainland China, the more developing to the tourism industry and its derivatives. Therefore, an accuracy forecasting for the tourists from mainland China is an important official work for the Tourism Bureau, ROC. However, a new system, the collected data is limited and fluctuated, the traditional statistical method is difficult to be applied. In order to cope with such problem, in this paper, we use the proposed GSVR for solving such kind of problem.

In this example, the forecasting process using GSVR can be described as follows:

Step 1. Define the universal discourse U for the historical data. The China tourists from September 2008 to December 2009 are summarized as a series $x^{(0)}$ as follows:

$$x^{(0)} = (10176, 11797, 12728, 11782, 19420, 16249, 45739, 79292, 64458, 26135, 37890, 43614, 31620, 42840, 74296, 57573),$$

where the minimum value of this series is 10, 176, and the maximum value of this series is 79, 292.

Step 2. Scale the collected time series data. The collected series is scaled into $[0, 1]$ by Eq. (3.13), and we can obtained the scaled series as follows:

$$x^{(0)} = (0.0000, 0.0235, 0.0369, 0.0232, 0.1338, 0.0879, 0.5147, 1.0000, 0.7856, 0.2310, 0.4011, 0.4839, 0.3104, 0.4727, 0.9280, 0.6860).$$

Step 3. Calculate the AGO time series. The AGO series obtained from the scaling series is

$$x^{(1)} = (0.0000, 0.0235, 0.0604, 0.0836, 0.2174, 0.3053, 0.8200, 1.8200, 2.6056, 2.8366, 3.2377, 3.7216, 4.0320, 4.5047, 5.4327, 6.1186).$$

Step 4. Find a well-fitted kernel function. The most popular kernel function named as Gaussian radial basis function (RBF) is chosen to make the nonlinear data for linear analysis in the feature space. The RBF is defined as $K(x_i, x_j) = \exp(-\|(x_i - x_j)\|^2/2\sigma^2)$, and for simplicity, we set $\sigma^2 = 1$.

Step 5. Forecast using the GSVR model. In this example, the parameters C and ε are set as 1.3835 and 0.1, respectively. Then, using the dual model (3.11), we can derive the GSVR, and then use it to forecast the trend of the collected time series data. The solution for the dual variables α_t and α_t^* are obtained in the second and third column of Table 1, fourth and fifth are the bias term b^+ and b^- for the upper GSVR and lower GSVR. Then the forecasting values with upper bias and lower bias terms are derived in the 6$^{\text{th}}$ and 7$^{\text{th}}$ columns. The average forecasting values are obtained by averaging the last two columns of Table 1.

Step 6. Rescaled the forecasting values. The forecasting values obtained from step 5 are rescaling to the original scale, and then we can obtain the forecasting values in the last column of Table 2. In addition, the extrapolative value for January 2010 is also derived for comparison.

Table 1: The solving results for the GSVR.

t	α_t	α_t^*	b^+	b^-	$f_{b^+}(x)$	$f_{b^-}(x)$
2008/10	0	0			0.485221	1.097938
2008/11	0	0			0.504435	1.117152
2008/12	0	1.308452		1.23224	0.535704	1.148421
2009/1	0	0			0.555975	1.168692
2009/2	0	1.3835			0.679795	1.292512
2009/3	1.3835	0			0.765493	1.37821
2009/4	1.3835	0			1.216569	1.829286
2009/5	4.24E-02	0	0.843217		1.098339	1.711056
2009/6	0	0			0.623278	1.235995
2009/7	0	1.3835			0.576388	1.189105
2009/8	1.3835	0			0.620302	1.233019
2009/9	0	2.13E-02		1.572692	0.82305	1.435767
2009/10	0	1.381722		1.811976	0.985267	1.597985
2009/11	1.3835	0			1.19029	1.803007
2009/12	0	9.79E-02		1.206826	1.198702	1.811419

Step 7. Forecasting error analysis. After operating the above steps, the forecasting performance is measured in terms of MAPE in the last three and last rows of Table 1.

For comparison, the forecasting values between GSVR with some forecasting models are also list in Table 2 which shows that the GSVR model has smaller MAPE than the others models, especially, in the training data and testing data sets. The new open tourism market for mainland Chain is limited and fluctuated, it is difficult to use the traditional forecasting methods such as simple regression, single exponential smoothing, and grey model $GM(1,1)$, but GSVR provides good forecasting values.

Table 2: The forecasting results with some models.

Time	Collected data	Simple regression	Single Exponential Smoothing	$GM(1,1)$	GSVR
2008/9	10176				
2008/10	11797	15215	9474.9	22211.91	10176
2008/11	12728	18505	13155.4	23937.08	11991.35
2008/12	11782	21795	12477.9	25796.24	14945.55
2009/1	19420	25085	11374.9	27799.80	16860.68
2009/2	16249	28375	24126.4	29958.97	28558.87
2009/3	45739	31665	11640.7	32285.84	36655.4
2009/4	79272	34955	65686.5	34793.44	79272
2009/5	64458	38245	87219.5	37495.80	68101.94
2009/6	26135	41535	51142.5	40408.04	23219.31
2009/7	37890	44825	11505.6	43546.48	28789.3
2009/8	43614	48115	53324.9	46928.67	32938.12
2009/9	31620	51405	37933.2	50573.56	42093.25
2009/10	42840	54695	27926.8	54501.53	47419.19
2009/11	74296	57985	51564.2	58734.59	76789.21
2009/12	57573	61275	87594.1	63296.42	67583.92
MAPE (%)		**39.7775**	**38.0638**	**46.4863**	**19.0025**
2010/1 (F)	59348	64565	40010.7	68212.57	58776.04
MAPE (%)		**8.7905**	**32.5829**	**14.9366**	**0.9637**

4.2. Discussion for parameters

In this subsection, an in-depth discussion for the parameters ε, C, σ^2 are surveyed. First, the decreasing function for parameter ε is shown in Figure 3 which implies that the larger that of value ε, the smaller rate of "support vectors per sample size" [13]. Therefore, we find that $\varepsilon = 0.1$ is better in this illustrated example because that too little value ε with large rate of "support vectors per sample size" might make the forecasting

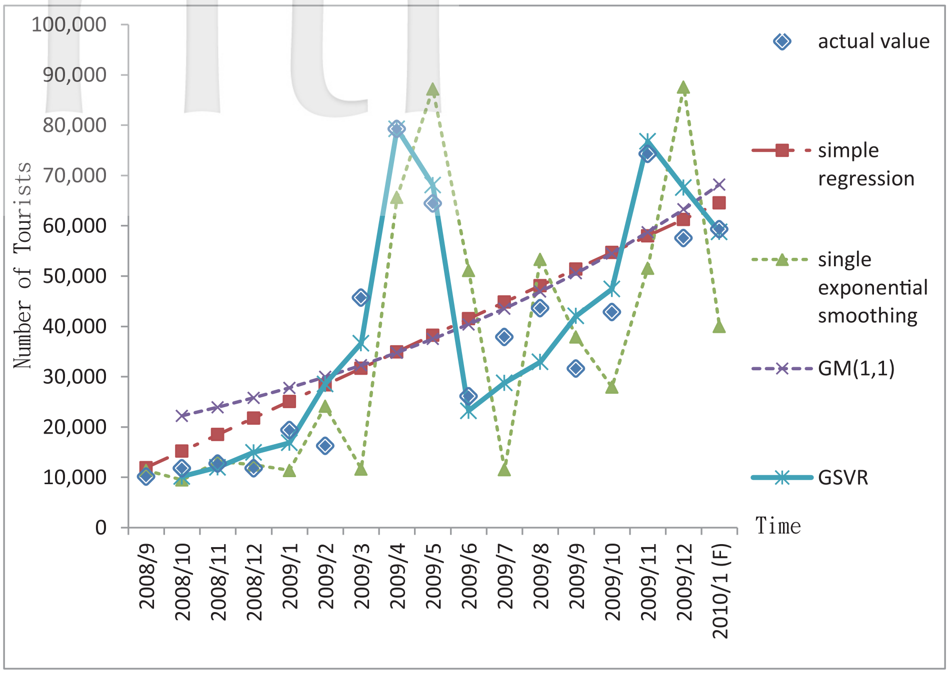

Figure 2: A comparison between GSVR and the other forecasting models.

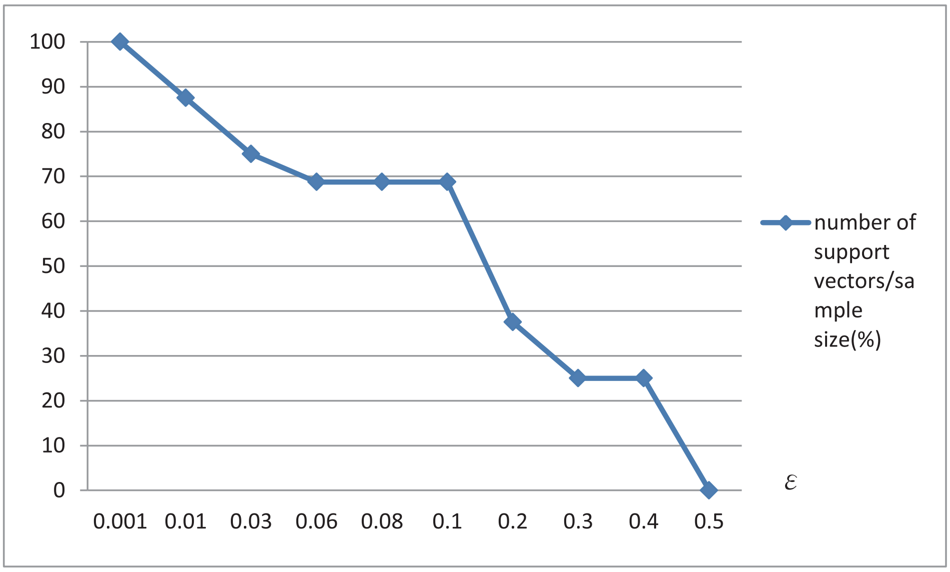

Figure 3: The number of support vectors.

values over-fitting; on the contrary, too large value ε with small rate of "support vectors per sample size" might make under-fitting [1]. Second, the variation of C is tested and shown in Figure 4, we can find out that too large value $C = 0.5$ might make the

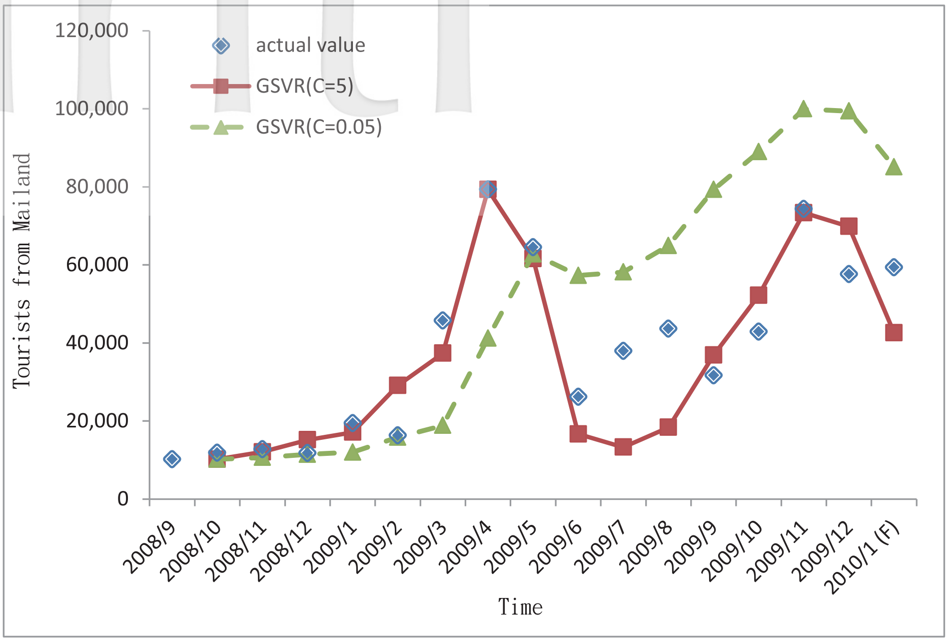

Figure 4: The Mainland tourists forecasting values with different calues C.

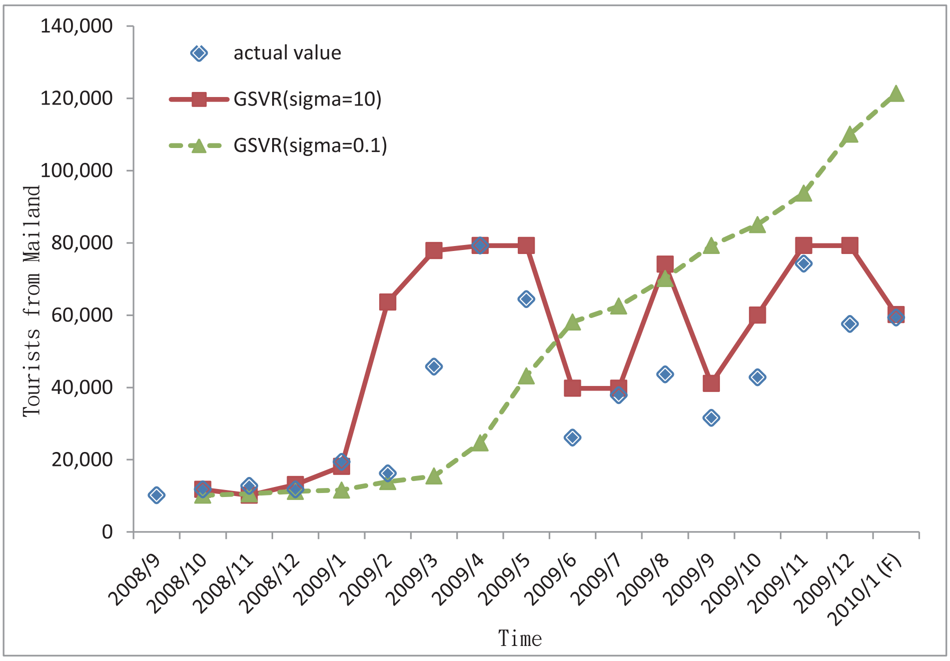

Figure 5: The Mainland tourists forecasting values with different values σ^2.

forecasting values over-fitting; on the contrary, too little value $C = 0.05$ might make under-fitting [3]. Third, as in Figure 5, we can find out that too large value $\sigma^2 = 10$

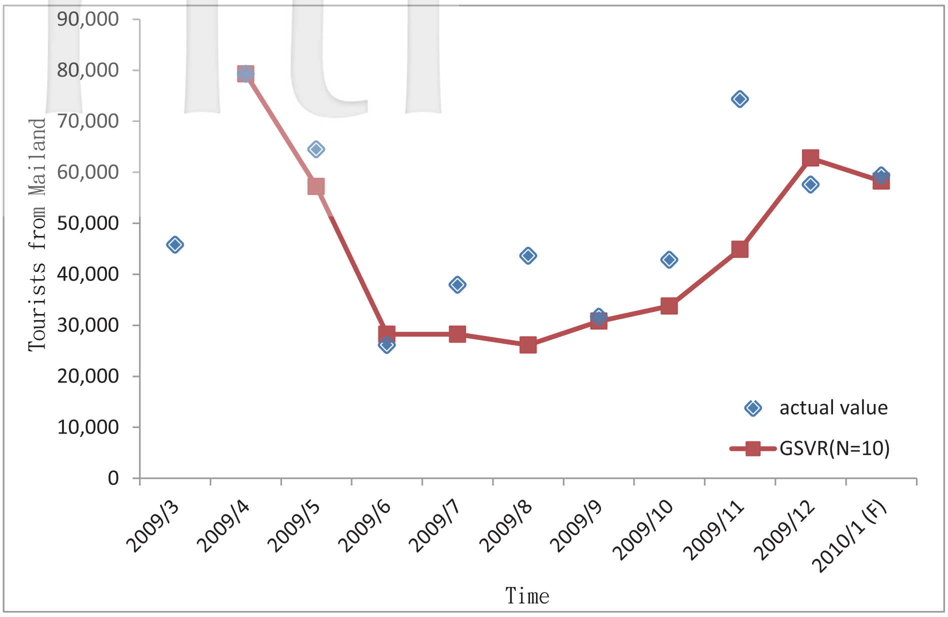

Figure 6: The tourists forecasting using GSVR with sample size $N = 10$.

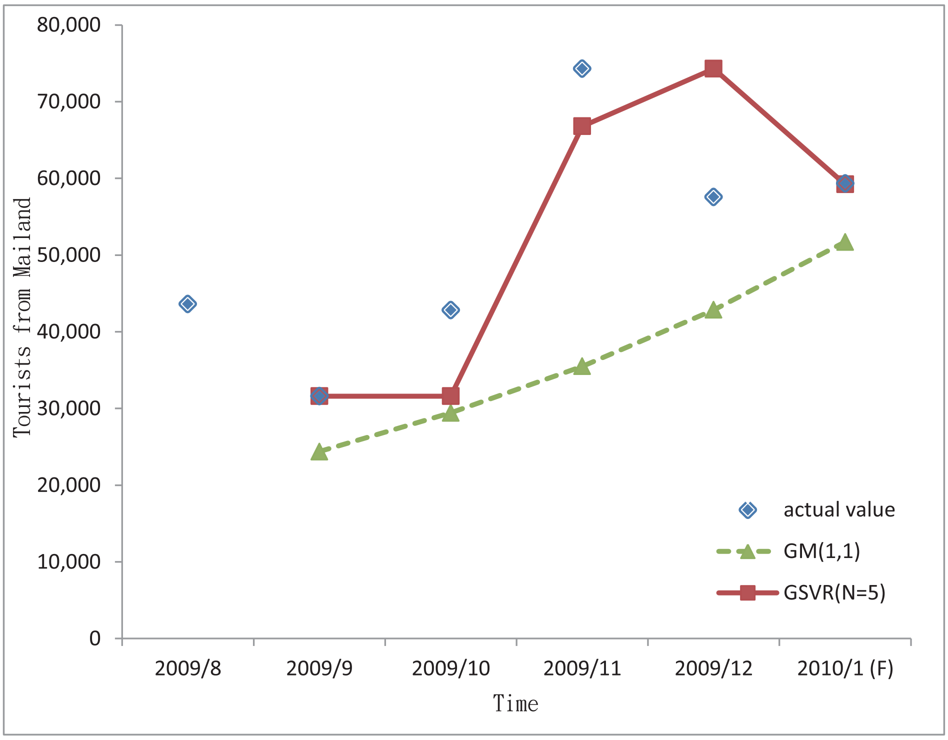

Figure 7: The tourists forecasting using GSVR and $GM(1,1)$ with sample size $N = 5$.

Table 3: The tourists forecasting using GSVR with sample size $N = 10$.

Time	Actual value	GSVR ($N = 10$)	MAPE(%)
2009/3	45739		
2009/4	79272	79272	0
2009/5	64458	57204.71	11.2527
2009/6	26135	28237.28	8.04392
2009/7	37890	28237.278	25.4756
2009/8	43614	26135	40.0766
2009/9	31620	30783.74	2.64471
2009/10	42840	33758.325	21.1991
2009/11	74296	44876.635	39.5975
2009/12	57573	62774.682	9.0349
MAPE (%) (2009/4∼2009/12)			**17.4806**
2010/1 (F)	59348	58251.92	**1.8469**

Table 4: The tourists forecasting using GSVR with sample size $N = 5$.

Time	Actual value	GSVR ($N = 5$)	MAPE(%)
2009/8	43614		
2009/9	31620	31620	0
2009/10	42840	31620	26.1905
2009/11	74296	66800.98	10.0881
2009/12	57573	74296	29.0466
MAPE (%) (2009/9∼2009/12)			**16.3313**
2010/1 (F)	59348	59236.88	**0.1872**

might make the forecasting values over-fitting; on the contrary, too little value $C = 0.1$ might make all the kernel values to be near zero which derives worse forecasting values.

Next, different sample sizes are used for testing the proposed model. First, ten collected data from March to December 2009 are used for modelling GSVR where its forecasting values are listed in Table 3 and shown in Figure 6. Obviously, the forecasting results are still better to fit the collected data with good forecasting performance MAPE = 17.4806%, and the extrapolation error smaller for January 2010 is still smaller as 1.8469%. Second, five data collected from August to December 2009 are used for modelling GSVR, as in Table 4 and Figure 7, the forecasting results are still better than $GM(1,1)$ model and fit the collected data with good forecasting performance MAPE = 16.3313%, and the extrapolation error smaller for January 2010 is 0.1872%.

5. Conclusion

In this study, a new approach GSVR for analyzing the limited time series with some fluctuated points has been proposed. The results indicated the GSVR model can be used to derive considerable forecasting value by AGOing the time series data and using kernel function to transfer the nonlinear time series data into feature space. Both the experimental results and the analytical forecasting confirm the potential benefits of the new approach in terms of limited time series forecasting. Most importantly, the illustrated experiments were archived with a good MAPE on different of sample sizes, and showed a better forecasting trend on the out-of-sample. If the GSVR model meets its expectations, then this approach will be an important tool in non-statistical forecasting.

Acknowledgements

The authors gratefully acknowledge the financial support from National Science Foundation with project No. NSC 98-2221-E-032-014.

References

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

[2] Che, J. and Wang, J. (2010). *Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling*, Energy Conversion and Management, Vol.51, 1911-1917.

[3] Chen, K. Y. and Wang, C. H. (2007). *Support vector regression with genetic algorithms in forecasting tourism demand*, Tourism Management, Vol.28, 215-226.

[4] Deng, J. L. (1986). Grey Forecasting and Decision, Huazhong of Society and Technology Press, Wuhan.

[5] Farquad, M. A. H., Ravi, V. and Raju, S. B. (2010). *Support vector regression based hybrid rule extraction methods for forecasting*, Expert Systems with Applications, Vol.37, 5577-5589.

[6] Gunn, S. R. (1998). Support Vector Machine for Classification and Regression, University of Southampton, U.S.A.

[7] Hong, D. H. and Hwang, C. (2003). *Support vector fuzzy regression machines*, Fuzzy Sets and Systems, Vol.138, 271-281.

[8] Hong, W. C., Dong, Y., Chen, L. Y. and Lai, C. Y. (2010). *Taiwanese 3G mobile phone demand forecasting by SVR with hybrid evolutionary algorithms*, Expert Systems with Applications, Vol.37, 4452-4462.

[9] Hong, W. C. (2009). *Hybrid evolutionary algorithms in a SVR-based electric load forecasting model*, Electrical Power and Energy Systems, Vol.31, 409-417.

[10] Levis, A. A. and Papageorgiou, L. G. (2005). *Customer demand forecasting via support vector regression analysis*, Chemical Engineering Research and Design, Vol.83, 1009-1018.

[11] Lu, C. J., Lee, T. S. and Chiu, C. C. (2009). *Financial time series forecasting using independent component analysis and support vector regression*, Decision Support Systems, Vol.47, 115-125.

[12] Makridakis, S. and Hibon, M. (1979). *Accuracy of forecasting: An empirical investigation*, Journal of the Royal Statistical Society, Series A, Vol.142, 97-125.

[13] Suykens, J. K. A., Gestel, T. V., Brabanter, D., Moor, B. D. and Vandewalle, J. (2002). Least Squares Support Vector Machines. World Scientific, Singapore.

[14] Shieh, H. L. and Kuo, C. C. (2010). *A reduced data set method for support vector regression*, Expert Systems with Applications, Vol.37, 7781-7787.

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

[16] Tsaur, R. C. (2009). *Insight of a fuzzy grey autoregressive model*, Soft Computing, Vol.13, 919-931.

[17] Vapnik, V. (1995). The Nature of Statistical Learning Theory. New York: Springer.

Department of Management Sciences, Tamkang University, New Taipei City, Taiwan, R.O.C.

E-mail: rctsaur@yahoo.com.tw

Major area(s): Multicriteria decision making, fuzzy set theory, operations research.

Department of Management Sciences, Tamkang University, New Taipei City, Taiwan, R.O.C.

E-mail: cmann0918@yahoo.com.tw

Major area(s): Multicriteria decision making, fuzzy set theory, operations research.

(Received March 2014; accepted June 2014)