Communications of the Korean Statistical Society
2009, Vol. 16, No. 2, 335–348

Fuzzy Semiparametric Support Vector Regression for Seasonal Time Series Analysis

Jooyong Shim[a], Changha Hwang[1,b], Dug Hun Hong[c]

[a]Dept. of Applied Statistics, Catholic Univ. of Daegu, [b]Dept. of Statistics, Dankook Univ., [c]Dept. of Mathematics, Myongji Univ.

Abstract

Fuzzy regression is used as a complement or an alternative to represent the relation between variables among the forecasting models especially when the data is insufficient to evaluate the relation. Such phenomenon often occurs in seasonal time series data which require large amount of data to describe the underlying pattern. Semiparametric model is useful tool in the case where domain knowledge exists about the function to be estimated or emphasis is put onto understandability of the model. In this paper we propose fuzzy semiparametric support vector regression so that it can provide good performance on forecasting of the seasonal time series by incorporating into fuzzy support vector regression the basis functions which indicate the seasonal variation of time series. In order to indicate the performance of this method, we present two examples of predicting the seasonal time series. Experimental results show that the proposed method is very attractive for the seasonal time series in fuzzy environments.

Keywords: Fuzzy regression, seasonal time series, semiparametric model, support vector regression.

1. Introduction

Time series analysis is a very important task in many practical circumstances such as monitoring, diagnosis, control and dynamic decision making (Box and Jenkins, 1976; Franses, 1998; Haykin and Kosko, 2001; West and Harrison, 1997). Most time series patterns can be described in terms of two basic classes of components: trend and seasonality (Ghysels and Osborn, 2001; West and Harrison, 1997). The former represents a general systematic linear or nonlinear component that changes over time and does not repeat or at least does not repeat within the time range captured by the observed data. The latter may have a formally similar nature, however, it repeats itself in systematic intervals over time. These two general classes of time series components may coexist in real-life data. Regression methods can be used to deal with time series with seasonality. A collection of recent papers dealing with seasonal time series can be found in Burman and Shumway (1998), Dominici *et al.* (2004) and Ghysels and Osborn (2001).

Various parametric and nonparametric methods have been proposed and applied for the time series data. Semiparametric model specifications for seasonal time series analysis have been extensively discussed in the literature, *e.g.* Burman and Shumway (1998), Dominici *et al.* (2004). Semiparametric model is used for characterizing seasonal time series model, which consists of a common trend

This work was supported by grant No. R01-2006-000-10226-0 from the Basic Research Program of the Korea Science and Engineering Foundation.

[1] Corresponding author: Department of Statistics, Dankook University, Gyeonggido 448-160, Korea. Email: chwang@dankook.ac.kr

function over periods and additive individual trend functions that are specific to each season within periods. For the seasonal times series the seasonal variation can be described by linear combinations of components of the basis function of time, which leads to construct the semiparametric approach to the seasonal times series.

Conventional methods are based on statistical and probabilistic approaches which may not be suitable for applying purely mathematical models to the data generated by human activities. Conventional methods require that at least 50 and preferably 100 observations or more should be used (Box and Jenkins, 1976). In particular, these methods require a large amount of data to show the underlying pattern of seasonal time series. However, when we consider the rapid change of current socio-economic situations, it is almost impossible to collect 50 observations or more for time series analysis. Fuzzy regression is used as a complement or an alternative to represent the relation between variables among the forecasting models when the data is insufficient to evaluate the relation (Watada, 1992). However, some fuzzy regression models can be utilized when the data is sufficient.

Based on the works by Zadeh (1965), fuzzy set theory has been increasingly recognized as a useful tool in handling the vagueness of human knowledge. In many practical circumstances, real intervals or fuzzy numbers are used to model imprecise observations derived from uncertain measurements or linguistic assessments. Tanaka *et al.* (1982), Tanaka (1987), Tanaka and Ishibuchi (1992) suggested the use of fuzzy regression to solve the fuzzy environment problem and avoid modeling error. This model is an interval prediction model based on possibility theory. Song and Chissom (1993a, 1993b, 1994) proposed the concepts of fuzzy time series to deal with the forecasting problem in which the historical data are linguistic values. They applied fuzzy time series models to forecasting enrollments of the University of Alabama. Song *et al.* (1995) proposed a new fuzzy time series model by means of defining some new operations on fuzzy numbers. Chen (1996) presented a fuzzy time series method based on the concept of Song and Chissom. Hwang *et al.* (1998) proposed a new method to forecast university enrollments, which is more efficient than the ones presented in Song and Chissom (1993b, 1994), Song *et al.* (1995) in that the proposed method simplifies the arithmetic operation process.

An application of fuzzy regression to fuzzy time series analysis was found by Watada (1992). Chang (1997) presented a fuzzy forecasting technique for seasonality in the time series data by incorporating fuzzy seasonality into fuzzy regression. Tseng *et al.* (2001) developed the fuzzy ARIMA model by combining the works of ARIMA model and fuzzy regression. Tseng and Tzeng (2002) proposed a fuzzy seasonal ARIMA(FSARIMA) forecasting model, which combines the advantages of the seasonal ARIMA(SARIMA) model and the fuzzy regression model. Tsaur *et al.* (2002) proposed a fuzzy regression model for solving time series problem with seasonal data by considering within-cyclic and between-cyclic patterns.

In this paper, we propose a fuzzy semiparametric support vector regression(FSSVR) model for the seasonal times series by combining the possibility estimation formulation (Tanaka *et al.*, 1982; Tanaka, 1987; Tanaka and Ishibuchi, 1992) integrating the property of central tendency with the principle of semiparametric support vector regression(SSVR) (Smola *et al.*, 1998). The rest of this paper is organized as follows. Section 2 illustrates the fuzzy support vector regression(FSVR) for crisp input and output data, which combines the possibility estimation formulation of fuzzy regression with the principle of support vector regression(SVR) (Vapnik, 1995, 1998). Section 3 proposes a FSSVR model. Section 4 applies the FSSVR model to the seasonal time series and illustrates the efficacy of the proposed approach through real examples of seasonal time series. Finally, Section 5 presents the conclusions.

2. Fuzzy Support Vector Regression

In this section, we illustrate fuzzy support vector regression(FSVR) to evaluate fuzzy linear and nonlinear regression models for crisp input and output data by means of combining the possibility estimation formulation (Tanaka *et al.*, 1982; Tanaka, 1987; Tanaka and Ishibuchi, 1992) integrating the property of central tendency with the principle of SVR (Vapnik, 1995, 1998). In the case of noisy learning data, the use of traditional neural network, often leads to poor generalization and overfitting. The SVR, whose foundations have been established by Vapnik, has been designed to overcome these problems. For the estimation of central tendency, the SVR uses ϵ-insensitive loss function defined by

$$|x|_\epsilon = \begin{cases} 0, & |x| < \epsilon, \\ |x| - \epsilon, & |x| > \epsilon. \end{cases}$$

FSVR uses the standard SVR approach to yield the fuzzy estimate of output.

We now look at how to derive FSVR using the SVR approach. It is very similar to Hong and Hwang (2005) and Hwang *et al.* (2006). Suppose that we are given training data $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\} \subset R^m \times R$. Let x_{ij} be element of $\mathbf{x}_i$. Then, it is no loss of generality that we assume $x_{ij} \geq 0$, because we usually deal with positive observations in time series analysis. For pedagogical reasons, we begin by describing the case of fuzzy linear regression functions $Y(\mathbf{x})$, taking the form

$$Y(\mathbf{x}) = A_0 + A_1 x_1 + \cdots + A_m x_m = \mathbf{A}^T \mathbf{x}, \tag{2.1}$$

where $\mathbf{x} = (1, x_1, \ldots, x_m)^T$ is a crisp input vector, $\mathbf{A} = (A_0, A_1, \ldots, A_m)^T$ is a coefficient vector of symmetric triangular fuzzy numbers, and $Y(\mathbf{x})$ is the corresponding estimated fuzzy output. Here the superscript ^{T} denotes vector or matrix transpose. From now on we use such a reexpressed input vector $\mathbf{x} = (1, x_1, \ldots, x_m)^T$ instead of $\mathbf{x} = (x_1, \ldots, x_m)^T$ for our purpose. A fuzzy coefficient A_i is denoted as $A_i = (a_i, c_i)$ where a_i is a center and c_i is a spread. By fuzzy arithmetic, the regression model (1) can be expressed as

$$\begin{aligned} Y(\mathbf{x}_i) &= (a_0, c_0) + (a_1, c_1)x_{i1} + \cdots + (a_m, c_m)x_{im} \\ &= (a_0 + a_1 x_{i1} + \cdots + a_m x_{im},\ c_0 + c_1 x_{i1} + \cdots + c_m x_{im}) \\ &= \left(\mathbf{a}^T \mathbf{x}_i,\ \mathbf{c}^T \mathbf{x}_i\right), \end{aligned} \tag{2.2}$$

where $\mathbf{a} = (a_0, a_1, \ldots, a_m)^T$ and $\mathbf{c} = (c_0, c_1, \ldots, c_m)^T$. By the Extension Principle (Zadeh, 1965) the membership function of the estimated fuzzy output $Y(\mathbf{x}_i)$ can be expressed as

$$\mu_Y(y_i) = \max\left\{1 - \frac{|y_i - \mathbf{a}^T \mathbf{x}_i|}{\mathbf{c}^T \mathbf{x}_i},\ 0\right\}. \tag{2.3}$$

As in SVR (Vapnik, 1995, 1998), the fuzzy linear regression model (2.2) can be extended to the following nonlinear model

$$Y(\mathbf{\Phi}(\mathbf{x}_i)) = \left(\mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i),\ \mathbf{c}^T \mathbf{\Phi}(\mathbf{x}_i)\right) \tag{2.4}$$

by using the nonlinear function $\mathbf{\Phi} : R^{m+1} \rightarrow \mathcal{F}$ which maps the input space to a so-called higher dimensional feature space. It is important to note that the dimension of space $\mathcal{F}$ is only defined in an implicit way (it can be infinite dimensional) and that the identity map $\mathbf{\Phi}$ leads the nonlinear model

(2.4) to the linear model (2.2). Here, we use the same notations **a** and **c** as ones in the linear case to avoid the abuse of notation, although they are in general different.

In this paper we focus on explaining how to derive the nonlinear FSVR since the linear FSVR can be obtained straightforwardly from the nonlinear FSVR by using the identity map $\mathbf{\Phi}$. For FSVR we basically use the way of constructing objective function in SVR and the criterion of minimizing the total vagueness and the sum of ϵ-insensitive distances between the estimated output centers and the observed output, which reflects both the property of central tendency and the possibility estimation formulation (Hong and Hwang, 2005; Hwang *et al.*, 2006; Tanaka and Lee, 1998). Furthermore, we also take into account the condition that the membership degree of each observation y_i is greater than an imposed threshold possibility as h, $h \in [0, 1]$. This criterion requires that each observation y_i has at least h degree of belonging to $Y(\mathbf{\Phi}(\mathbf{x}_i))$ as $\mu_Y(y_i)$, which is equivalent to

$$\mu_Y(y_i) \geq h, \quad i = 1, \ldots, n.$$

Then, similar to Hong and Hwang (2005), Hwang *et al.* (2006), the problem finding the fuzzy regression parameters is formulated as the following quadratic programming problem:

$$\min_{\mathbf{a},\mathbf{c}} \frac{1}{2}\left(\|\mathbf{a}\|^2 + \|\mathbf{c}\|^2\right) + \gamma_1 \sum_{i=1}^{n} \xi_{1i} + \gamma_2 \sum_{i=1}^{n} (\xi_{2i} + \xi_{2i}^*) \tag{2.5}$$
$$\text{st} \begin{cases} \mathbf{c}^T \mathbf{\Phi}(\mathbf{x}_i) \leq \xi_{1i}, \\ y_i - \mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i) \leq \xi_{2i} + \epsilon, \\ \mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i) - y_i \leq \xi_{2i}^* + \epsilon, \\ \mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i) + (1-h)\mathbf{c}^T \mathbf{\Phi}(\mathbf{x}_i) \geq y_i, \\ \mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i) - (1-h)\mathbf{c}^T \mathbf{\Phi}(\mathbf{x}_i) \leq y_i, \\ i = 1, \ldots, n. \end{cases}$$

The weight coefficient $\gamma_1 > 0$ determines the trade-off between $\sum_{i=1}^n \mathbf{c}^T \mathbf{\Phi}(\mathbf{x}_i)\mathbf{\Phi}(\mathbf{x}_i)^T \mathbf{c}$ and the flatness of the estimated spread of $Y(\mathbf{\Phi}(\mathbf{x}))$, and $\gamma_2 > 0$ determines the trade-off between $\sum_{i=1}^n |y_i - \mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i)|_\epsilon$ and the flatness of the estimated center of $Y(\mathbf{\Phi}(\mathbf{x}))$. Here, ξ_{1i} represent spreads of the estimated outputs, and ξ_{2i}, ξ_{2i}^* are slack variables representing upper and lower constraints on the outputs of the model. Now, we can construct a Lagrange function as follows:

$$\begin{aligned} L = & \frac{1}{2}\left(\|\mathbf{a}\|^2 + \|\mathbf{c}\|^2\right) + \gamma_1 \sum_{i=1}^{n} \xi_{1i} + \gamma_2 \sum_{i=1}^{n} \left(\xi_{2i} + \xi_{2i}^*\right) \\ & - \sum_{i=1}^{n} \alpha_{1i} \left(\xi_{1i} - \mathbf{c}^T \mathbf{\Phi}(\mathbf{x}_i)\right) \\ & - \sum_{i=1}^{n} \alpha_{2i} \left(\xi_{2i} + \epsilon - y_i + \mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i)\right) \\ & - \sum_{i=1}^{n} \alpha_{2i}^* \left(\xi_{2i}^* + \epsilon - \mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i) + y_i\right) \\ & - \sum_{i=1}^{n} \alpha_{3i} \left(\mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i) + (1-h)\mathbf{c}^T \mathbf{\Phi}(\mathbf{x}_i) - y_i\right) \end{aligned} \tag{2.6}$$

$$
\begin{aligned}
&- \sum_{i=1}^{n} \alpha_{3i}^{*} \left(y_i - \mathbf{a}^T \mathbf{\Phi}(\mathbf{x}_i) + (1-h)\mathbf{c}^T \mathbf{\Phi}(\mathbf{x}_i) \right) \\
&- \sum_{i=1}^{n} \left(\eta_{1i}\xi_{1i} + \eta_{2i}\xi_{2i} + \eta_{2i}^{*}\xi_{2i}^{*} \right).
\end{aligned}
$$

Here, α_{1i}, α_{2i}, α_{2i}^{*}, α_{3i}, α_{3i}^{*}, η_{1i}, η_{2i} and η_{2i}^{*} are Lagrange multipliers. It follows from the saddle point condition that the partial derivatives of L with respect to the primal variables $\mathbf{a}, \mathbf{c}$, ξ_{1i}, ξ_{2i} and ξ_{2i}^{*} have to vanish for optimality.

$$
\frac{\partial L}{\partial \mathbf{a}} = \mathbf{0} \quad \rightarrow \quad \mathbf{a} - \sum_{i=1}^{n} \left(\alpha_{2i} - \alpha_{2i}^{*} \right) \mathbf{\Phi}(\mathbf{x}_i) - \sum_{i=1}^{n} \left(\alpha_{3i} - \alpha_{3i}^{*} \right) \mathbf{\Phi}(\mathbf{x}_i) = \mathbf{0}, \tag{2.7}
$$

$$
\frac{\partial L}{\partial \mathbf{c}} = \mathbf{0} \quad \rightarrow \quad \mathbf{c} + \sum_{i=1}^{n} \alpha_{1i} \mathbf{\Phi}(\mathbf{x}_i) - (1-h) \sum_{i=1}^{n} \left(\alpha_{3i} + \alpha_{3i}^{*} \right) \mathbf{\Phi}(\mathbf{x}_i) = \mathbf{0}, \tag{2.8}
$$

$$
\frac{\partial L}{\partial \xi_{1i}} = 0 \quad \rightarrow \quad \gamma - \alpha_{1i} - \eta_{1i} = 0, \tag{2.9}
$$

$$
\frac{\partial L}{\partial \xi_{2i}^{(*)}} = 0 \quad \rightarrow \quad \gamma - \alpha_{2i}^{(*)} - \eta_{2i}^{(*)} = 0. \tag{2.10}
$$

The algorithm would only depend on the data through inner products in $\mathcal{F}$, *i.e.* on functions of the form $\mathbf{\Phi}(\mathbf{x}_i)^T \mathbf{\Phi}(\mathbf{x}_j)$. Hence it suffices to know and use $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{\Phi}(\mathbf{x}_i)^T \mathbf{\Phi}(\mathbf{x}_j)$ instead of defining $\mathbf{\Phi}(\cdot)$ explicitly. Notice that the identity map $\mathbf{\Phi}$ leads nonlinear model to linear model. The use of a kernel function is an attractive computational short-cut. Hence, replacing $\mathbf{\Phi}(\mathbf{x})^T \mathbf{\Phi}(\mathbf{y})$ with $K(\mathbf{x}, \mathbf{y})$ and substituting (2.7)–(2.10) into (2.6) yields the following dual optimization problem:

$$
\max \left\{
\begin{aligned}
&-\frac{1}{2} \sum_{i,j=1}^{n} \left(\alpha_{2i} - \alpha_{2i}^{*} \right) \left(\alpha_{2j} - \alpha_{2j}^{*} \right) K\left(\mathbf{x}_i, \mathbf{x}_j \right) \\
&-\frac{1}{2} \sum_{i,j=1}^{n} \left(\alpha_{3i} - \alpha_{3i}^{*} \right) \left(\alpha_{3j} - \alpha_{3j}^{*} \right) K\left(\mathbf{x}_i, \mathbf{x}_j \right) \\
&- \sum_{i,j=1}^{n} \left(\alpha_{2i} - \alpha_{2i}^{*} \right) \left(\alpha_{3j} - \alpha_{3j}^{*} \right) K\left(\mathbf{x}_i, \mathbf{x}_j \right) \\
&-\frac{1}{2}(1-h)^2 \sum_{i,j=1}^{n} \left(\alpha_{3i} + \alpha_{3i}^{*} \right) \left(\alpha_{3j} + \alpha_{3j}^{*} \right) K\left(\mathbf{x}_i, \mathbf{x}_j \right) \\
&-\frac{1}{2} \sum_{i,j=1}^{n} \alpha_{1i}\alpha_{1j} K\left(\mathbf{x}_i, \mathbf{x}_j \right) \\
&+(1-h) \sum_{i,j=1}^{n} \alpha_{1i} \left(\alpha_{3j} + \alpha_{3j}^{*} \right) K\left(\mathbf{x}_i, \mathbf{x}_j \right) \\
&+ \sum_{i=1}^{n} \left(\alpha_{2i} - \alpha_{2i}^{*} \right) y_i + \sum_{i=1}^{n} \left(\alpha_{3i} - \alpha_{3i}^{*} \right) y_i \\
&-\epsilon \sum_{i=1}^{n} \left(\alpha_{2i} + \alpha_{2i}^{*} \right).
\end{aligned}
\right. \tag{2.11}
$$

Here the constraints are $0 \le \alpha_{1i} \le \gamma_1$, $0 \le \alpha_{2i}, \alpha_{2i}^* \le \gamma_2$ and $\alpha_{3i}, \alpha_{3i}^* \ge 0$. Solving (2.11) with the above constraints determines the Lagrange multipliers, $\alpha_{1i}, \alpha_{ki}, \alpha_{ki}^*$, $k = 2, 3$.

The expansions of $\mathbf{a}$ and $\mathbf{c}$ can be written as

$$\mathbf{a} = \sum_{i=1}^{n} \left[\left(\alpha_{2i} - \alpha_{2i}^*\right) + \left(\alpha_{3i} - \alpha_{3i}^*\right) \right] \mathbf{\Phi}(\mathbf{x}_i),$$

$$\mathbf{c} = \sum_{i=1}^{n} \left[-\alpha_{1i} + (1-h)\left(\alpha_{3i} + \alpha_{3i}^*\right) \right] \mathbf{\Phi}(\mathbf{x}_i).$$

We notice that $\mathbf{a}$ and $\mathbf{c}$ are no longer explicitly given except the linear kernel function. However, they are uniquely defined in the weak sense by the inner products $\mathbf{a}^T\mathbf{\Phi}(\mathbf{x})$ and $\mathbf{c}^T\mathbf{\Phi}(\mathbf{x})$. It is noted that $\mathbf{c}^T\mathbf{\Phi}(\mathbf{x})$ is nonnegative. Therefore, the nonlinear FSVR function is given as follows:

$$Y(\mathbf{x}) = \left[\sum_{i=1}^{n} \left\{ \left(\alpha_{2i} - \alpha_{2i}^*\right) + \left(\alpha_{3i} - \alpha_{3i}^*\right) \right\} K(\mathbf{x}_i, \mathbf{x}),\ \sum_{i=1}^{n} \left\{ -\alpha_{1i} + (1-h)\left(\alpha_{3i} + \alpha_{3i}^*\right) \right\} K(\mathbf{x}_i, \mathbf{x}) \right]. \quad (2.12)$$

3. Fuzzy Semiparametric Support Vector Regression

In this section, we propose a fuzzy semiparametric support vector regression(FSSVR) for the purpose of employing for seasonal time series analysis by combining the possibility estimation formulation (Tanaka, 1987; Tanaka *et al.*, 1982; Tanaka and Ishibnchi, 1992) integrating the property of central tendency with the principle of semiparametric support vector regression(SSVR) (Smola *et al.*, 1998). In general, semiparametric models are useful techniques in the case where additional domain knowledge about the problem is available and emphasis is put onto understandability of the model. If the major properties of data can be described by a linear combination of a small set of basis functions $\{\varphi_1(\cdot), \ldots, \varphi_d(\cdot)\}$, we can construct semiparametric model, which is easy to understand and performs well. Smola *et al.* (1998) extended SVR to SSVR. Following the principle of constructing SSVR, we will extend FSVR to FSSVR. In this paper we focus on explaining how to derive the nonlinear FSSVR since the linear FSSVR can be obtained straightforwardly from the nonlinear FSSVR by using the identity map $\mathbf{\Phi}$.

Let us define the vector of basis functions for $\mathbf{x}$ and the coefficient vector as $\boldsymbol{\varphi}(\mathbf{x}) = (\varphi_1(\mathbf{x}), \ldots, \varphi_d(\mathbf{x}))^T$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^T$, respectively. Then, similar to (2.5) the objective function can be formulated as the following quadratic function:

$$\min_{\mathbf{a},\mathbf{c},\boldsymbol{\beta}} \frac{1}{2}\left(\|\mathbf{a}\|^2 + \|\mathbf{c}\|^2\right) + \gamma_1 \sum_{i=1}^{n} \xi_{1i} + \gamma_2 \sum_{i=1}^{n} \left(\xi_{2i} + \xi_{2i}^*\right) \quad (3.1)$$

$$\text{st} \begin{cases} \mathbf{c}^T\mathbf{\Phi}(\mathbf{x}_i) \le \xi_{1i}, \\ y_i - \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) - \boldsymbol{\beta}^T\boldsymbol{\varphi}(\mathbf{x}_i) \le \xi_{2i} + \epsilon, \\ \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) + \boldsymbol{\beta}^T\boldsymbol{\varphi}(\mathbf{x}_i) - y_i \le \xi_{2i}^* + \epsilon, \\ \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) + \boldsymbol{\beta}^T\boldsymbol{\varphi}(\mathbf{x}_i) + (1-h)\mathbf{c}^T\mathbf{\Phi}(\mathbf{x}_i) \ge y_i, \\ \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) + \boldsymbol{\beta}^T\boldsymbol{\varphi}(\mathbf{x}_i) - (1-h)\mathbf{c}^T\mathbf{\Phi}(\mathbf{x}_i) \le y_i. \end{cases}$$

Hence, we can construct a Lagrange function as follows:

$$L = \frac{1}{2}\left(\|\mathbf{a}\|^2 + \|\mathbf{c}\|^2\right) + \gamma_1 \sum_{i=1}^{n} \xi_{1i} + \gamma_2 \sum_{i=1}^{n} \left(\xi_{2i} + \xi_{2i}^*\right) \quad (3.2)$$

$$
\begin{aligned}
&-\sum_{i=1}^{n} \alpha_{1i}\left(\xi_{1i} - \mathbf{c}^T\mathbf{\Phi}(\mathbf{x}_i)\right)\\
&-\sum_{i=1}^{n} \alpha_{2i}\left(\xi_{2i} + \epsilon - y_i + \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) + \boldsymbol{\beta}^T\varphi(\mathbf{x}_i)\right)\\
&-\sum_{i=1}^{n} \alpha_{2i}^*\left(\xi_{2i}^* + \epsilon - \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) - \boldsymbol{\beta}^T\varphi(\mathbf{x}_i) + y_i\right)\\
&-\sum_{i=1}^{n} \alpha_{3i}\left(\mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) + \boldsymbol{\beta}^T\varphi(\mathbf{x}_i) + (1-h)\mathbf{c}^T\mathbf{\Phi}(\mathbf{x}_i) - y_i\right)\\
&-\sum_{i=1}^{n} \alpha_{3i}^*\left(y_i - \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) - \boldsymbol{\beta}^T\varphi(\mathbf{x}_i) + (1-h)\mathbf{c}^T\mathbf{\Phi}(\mathbf{x}_i)\right)\\
&-\sum_{i=1}^{n} \left(\eta_{1i}\xi_{1i} + \eta_{2i}\xi_{2i} + \eta_{2i}^*\xi_{2i}^*\right).
\end{aligned}
$$

Here, α_{1i}, α_{2i}, α_{2i}^*, α_{3i}, α_{3i}^*, η_{1i}, η_{2i} and η_{2i}^* are Lagrange multipliers.

By using the partial derivatives of Lagrangian function (3.2) with respect to $\mathbf{a}, \mathbf{c}$, β, ξ_1, ξ_2 and ξ_2^*, respectively, we have the same dual optimization problem as (2.11). The objective function and the box constraints on $\alpha_{1i}, \alpha_{2i}, \alpha_{2i}^*, \alpha_{3i}, \alpha_{3i}^*$ remain unchanged. The only modification comes from the additional unregularized basis functions as follows:

$$
\sum_{i=1}^{n}\left(\alpha_{2i} - \alpha_{2i}^* + \alpha_{3i} - \alpha_{3i}^*\right)\varphi_j(\mathbf{x}_i) = 0, \quad \text{for } j = 1, \ldots, d. \tag{3.3}
$$

The only difficulty remaining is how to determine β_i. These β_i can be determined by employing the Karush-Kuhn-Tucker(KKT) optimality conditions as follows:

$$
\begin{aligned}
\boldsymbol{\beta}^T\varphi(\mathbf{x}_i) &= y_i - \epsilon - \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i), && \text{for } i \in \{\, i : 0 < \alpha_{2i} < \gamma_2\}, && (3.4)\\
\boldsymbol{\beta}^T\varphi(\mathbf{x}_i) &= y_i + \epsilon - \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i), && \text{for } i \in \left\{\, i : 0 < \alpha_{2i}^* < \gamma_2\right\}, && (3.5)\\
\boldsymbol{\beta}^T\varphi(\mathbf{x}_i) &= y_i - \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) - (1-h)\mathbf{c}^T\mathbf{\Phi}(\mathbf{x}_i), && \text{for } i \in \{\, i : \alpha_{3i} > 0\}, && (3.6)\\
\boldsymbol{\beta}^T\varphi(\mathbf{x}_i) &= y_i - \mathbf{a}^T\mathbf{\Phi}(\mathbf{x}_i) + (1-h)\mathbf{c}^T\mathbf{\Phi}(\mathbf{x}_i), && \text{for } i \in \left\{\, i : \alpha_{3i}^* > 0\right\}. && (3.7)
\end{aligned}
$$

Therefore, the nonlinear FSSVR function is given as follows:

$$
\begin{aligned}
Y(\mathbf{x}) = \Bigg(&\sum_{i=1}^{n}\left[(\alpha_{2i} - \alpha_{2i}^*) + (\alpha_{3i} - \alpha_{3i}^*)\right] K(\mathbf{x}_i, \mathbf{x}) + \boldsymbol{\beta}^T\varphi(\mathbf{x}), \qquad (3.8)\\
&\sum_{i=1}^{n}\left[-\alpha_{1i} + (1-h)(\alpha_{3i} + \alpha_{3i}^*)\right] K(\mathbf{x}_i, \mathbf{x})\Bigg).
\end{aligned}
$$

4. Applications of FSSVR to Seasonal Time Series

In this section we illustrate how to apply FSSVR to seasonal time series and compare FSSVR with SARIMA and FSARIMA models in terms of the forecasting performance. We consider two real data sets quoted from Montgomery *et al.* (1990). One is regarding the monthly sales volume of soft drinks. The other is regarding the monthly demand for a carpet.

4.1. Additive decomposition method

Suppose the series $\{Z_t\}$ is seasonal with seasonal period s. Seasonal time series have been conventionally thought to consist of an mixture of trend(T_t), seasonal(S_t) and irregular components(e_t). We use an additive structural decomposition (West and Harrison, 1997) of a seasonal time series Z_t in terms of three components:

$$Z_t = T_t + S_t + e_t. \tag{4.1}$$

In the decomposition (4.1), the trend component T_t can be written as a function of $\mathbf{x}_t = (1, Z_{t-1}, \ldots, Z_{t-k})^T$ for some k. Here, k is determined from the training data. In this paper we assume a linear model for both data sets. Thus, the linear kernel is used for both data sets. The seasonal component S_t is written as

$$S_t = t + \sum_{j=1}^{q} \left[\beta_{1j} \sin\left(\frac{2\pi jt}{s}\right) + \beta_{2j} \cos\left(\frac{2\pi jt}{s}\right) \right], \tag{4.2}$$

where $q = [s/2]$ with $[x]$ equal to the integer portion of x.

Accordingly, the basis function vector $\boldsymbol{\varphi}(\mathbf{x}_t)$ and the coefficient vector $\boldsymbol{\beta}$ of FSSVR are written as

$$\boldsymbol{\varphi}(\mathbf{x}_t) = \left(t, \sin\left(\frac{2\pi t}{s}\right), \cos\left(\frac{2\pi t}{s}\right), \ldots, \sin\left(\frac{2\pi qt}{s}\right), \cos\left(\frac{2\pi qt}{s}\right) \right)^T$$

$$\boldsymbol{\beta} = \left(1, \beta_{11}, \beta_{21}, \ldots, \beta_{1q}, \beta_{2q}\right)^T.$$

In the applications of FSSVR to seasonal time series the associated membership degree h is set to 0.

4.2. Model selection

The functional structure of FSSVR is characterized by γ_1, γ_2 and ϵ or the kernel parameter. In this subsection, we describe a method for selecting these important parameters of FSSVR. There could be several parameter selection methods. In this paper we use a kind of cross-validation(CV) method. If data is not scarce, then the set of available input-output measurements can be divided into two parts. One of them is used to train a model while the other, called the test set, is used for testing the model. In this way several different models, all trained on the training set, can be compared on the test set. According to their performance on the test set, we try to infer the proper values of parameters. This is the basic form of cross-validation. A better method is to partition the original set in several different ways and to compute an average score over the different partitions.

In this paper, we do not use the above CV methods, since we do not have the training data large enough to divide. As illustrated, the linear kernel is used in this paper. Thus, we do not need to consider selecting the kernel parameter. We choose the parameter values $(\gamma_1, \gamma_2, \epsilon)$ which for the training data minimize

$$\frac{1}{n}\left[r(p,n) \sum_{t=1}^{n} \left(Z_t - \hat{Z}_t\right)^2 + \sum_{t=1}^{n} \left(|\hat{Z}_t^U - Z_t| + |Z_t - \hat{Z}_t^L| \right)^2 \right], \tag{4.3}$$

where $r(p,n) = (1 - \sqrt{p - p\ln p + \ln n/(2n)})_+^{-1}$ for $p = v/n$ with VC-dimension (for Vapnik-Chervonenkis dimension) v = number$\{\alpha_{2i}^{(*)} : 0 < \alpha_{2i}^{(*)} < \gamma_2\} - 1$ and the number of the training data n and

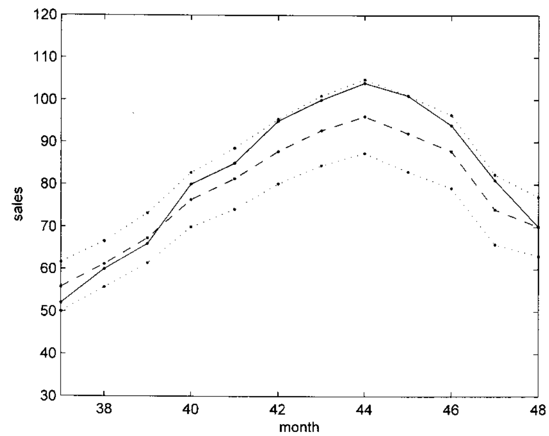

Figure 1: *One-step ahead forecasting results of FSSVR for the test data of soft drink data: forecasted bounds(dotted line), forecasted sales(dashed line) and actual sales(solid line).*

$(u)_+ = \max(u, 0)$. The VC-dimension is a measure of the capacity of a statistical learning algorithm (Cherkassky *et al.*, 1999; Vapnik, 1995, 1998). Here, $\hat{Z}_t^L$ and $\hat{Z}_t^U$ represent the forecasted lower and upper bounds of FSSVR, respectively. The first term of (4.3) is VC-based model selection criterion and it is more powerful than classical model selection criteria especially for small crisp samples (Chang, 1997). Thus, it is reasonable to use (4.3) as the model selection criterion for FSSVR.

4.3. Numerical studies

In this subsection, we use two real examples from Montgomery *et al.* (1990) to verify the effectiveness of FSSVR for seasonal time series.

4.3.1. The soft drink time series data

In order to demonstrate the performance the FSSVR model, we first consider soft drink data set (Montgomery *et al.*, 1990), which consists of 48 monthly sales volume of a 32-oz soft drink in hundreds of cases from January 1972 to December 1975. We divide the whole data into 2 partitions - the training data (t = 1 to 36) and the test data (t = 37 to 48). Tseng and Tzeng (2002) preprocessed soft drink data using logarithmic transformation and then acquired using SAS package software the best model of the training data which is SARIMA$(1, 1, 0)(0, 1, 0)_{12}$. In this paper we use original soft drink data without being preprocessed. We also obtain the best model as SARIMA$(1, 1, 0)(0, 1, 0)_{12}$ using SPSS package software. The best model is actually determined using Akaike information criterion(AIC) and Bayesian information criterion(BIC) in this package software. Thus, we recognize that $\mathbf{x}_t = (1, Z_{t-1})^T$, $s = 12$ and $q = 6$.

Using the model selection criterion (4.3), we select $(\gamma_1, \gamma_2, \epsilon)$ as (3, 10, 0). The most of experiments are conducted in MATLAB environment over Pentium IV at 2.0GHz. It takes CPU time 2.0469 in seconds to train FSSVR with already adaptively tuned parameters. Thus, it does not take long to train FSSVR. The mean absolute error(MAE) and the mean squared error(MSE) for the test data are $(4.84, 31.50)$ for FSSVR and $(6.96, 68.41)$ for SARIMA and FSARIMA. Figures 1, 2 and 3 show the actual sales, one-step ahead forecasts, upper and lower bounds of FSSVR, SARIMA and FSARIMA for the test data from January 1975 up to December 1975. Table 1 summarizes the actual sales, the upper and lower bounds of three models for the test data. As seen from Figures 1, 2 and 3 and Table 1,

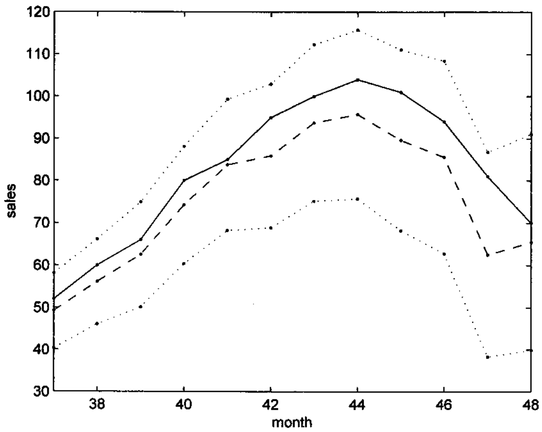

Figure 2: *One-step ahead forecasting results of SARIMA for the test data of soft drink data: forecasted bounds(dotted line), forecasted sales(dashed line) and actual sales(solid line).*

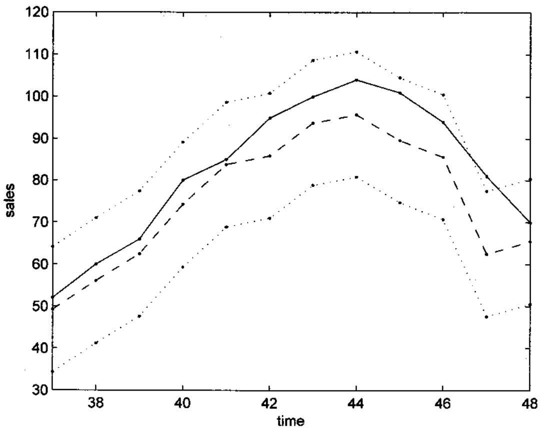

Figure 3: *One-step ahead forecasting results of FSARIMA for the test data of soft drink data: forecasted bounds(dotted line), forecasted sales(dashed line) and actual sales(solid line).*

the bounds of FSSVR and SARIMA contain all of test data points, while a bound of FSARIMA does not contain the corresponding data point. The average length of the prediction intervals is 14.71 for FSSVR, 35.08 for SARIMA and 29.80 for FSARIMA. Thus, we see the prediction interval of FSSVR model is narrower than the 95% prediction interval of SARIMA model and the prediction interval of FSARIMA model. Based on our experiment, we can see that FSSVR is an effective method to forecast the monthly sales of soft drink.

4.3.2. The carpet time series data

In this example we consider carpet data set from Montgomery *et al.* (1990), which consists of 48 monthly demand for a carpet from January 1970 to December 1973. We divide the whole data into 2 partitions - the training data (t = 1 to 36), and the test data (t = 37 to 48). We obtain the best model of the training data as SARIMA$(1,0,0)(1,1,0)_{12}$ using AIC and BIC in SPSS package software. As

Table 1: Actual values, forecasted lower and upper bounds of FSSVR, SARIMA and FSARIMA for the test data of soft drink data.

Date	Actual sales	FSSVR	SARIMA	FSARIMA
Jan-75	52	(49.98, 61.61)	(40.32, 58.12)	(34.32, 64.12)
Feb-75	60	(55.75, 66.57)	(46.06, 66.16)	(41.21, 71.01)
Mar-75	66	(61.41, 73.21)	(50.15, 74.89)	(47.62, 77.42)
Apr-75	80	(69.96, 82.74)	(60.39, 88.10)	(59.34, 89.14)
May-75	85	(74.14, 88.57)	(68.21, 99.34)	(68.87, 98.67)
Jun-75	95	(80.12, 95.44)	(68.87, 102.97)	(71.02, 100.82)
Jul-75	100	(84.47, 100.99)	(75.18, 112.32)	(78.85, 108.65)
Aug-75	104	(87.38, 104.82)	(75.73, 115.75)	(80.84, 110.64)
Sep-75	101	(82.99, 101.06)	(68.21, 111.09)	(74.75, 104.55)
Oct-75	94	(79.16, 96.45)	(62.76, 108.44)	(70.70, 100.50)
Nov-75	81	(65.86, 82.35)	(38.31, 86.74)	(47.63, 77.43)
Dec-75	70	(63.10, 77.02)	(39.88, 91.05)	(50.57, 80.37)
Average length		14.71	35.08	29.80

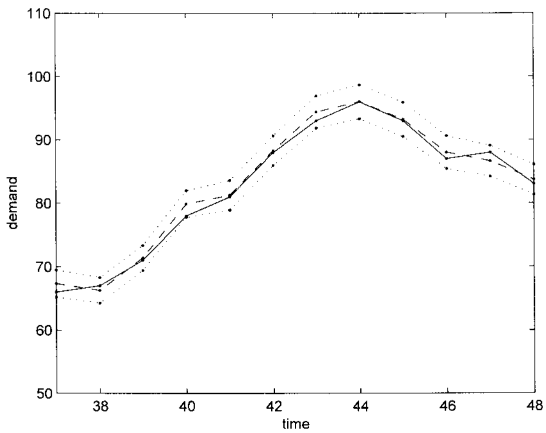

Figure 4: *One-step ahead forecasting results of FSSVR for the test data of carpet data: forecasted bounds (dotted line), forecasted demand (dashed line) and actual demand (solid line).*

before, we recognize that $\mathbf{x}_t = (1, z_{t-1})^T$, $s = 12$ and $q = 6$. Using the model selection criterion (4.3), we select $(\gamma_1, \gamma_2, \epsilon)$ as (40, 190, 0). The most of experiments are also conducted in MATLAB environment over Pentium IV at 2.0GHz. It takes CPU time 1.8594 in seconds to train FSSVR with already adaptively tuned parameters and thus does not take long.

The mean absolute error(MAE) and the mean squared error(MSE) for the test data are $(0.80, 0.96)$ for FSSVR and $(1.03, 1.61)$ for SARIMA and FSARIMA. Figures 4, 5 and 6 show the actual demand, one-step ahead forecasts, upper and lower bounds of FSSVR, SARIMA and FSARIMA for the test data from January 1973 up to December 1973. Figure 4 summarizes the actual demand, the upper and lower bounds of three models for the test data. As seen from Figures 4, 5 and 6 and Table 2, all bounds of FSSVR, SARIMA and FSARIMA contain their corresponding data points. The average length of the prediction intervals is 4.70 for FSSVR, 5.59 for SARIMA and 7.21 for FSARIMA. Thus, we notice the prediction interval of FSSVR model is narrower than the 95% prediction interval of SARIMA model and the prediction interval of FSARIMA model. Based on our experiment, we recognize that FSSVR is a convincing method to forecast the monthly demand of the carpet.

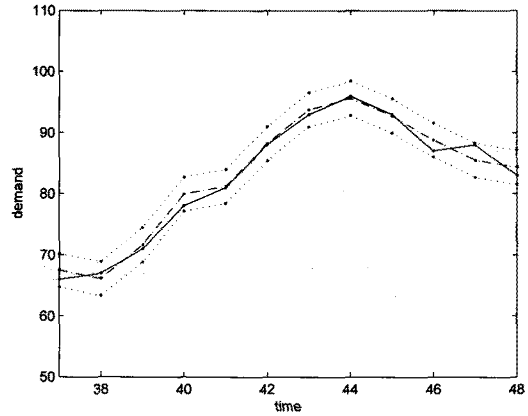

Figure 5: *One-step ahead forecasting results of SARIMA for the test data of carpet data: forecasted bounds(dotted line), forecasted demand(dashed line) and actual demand(solid line).*

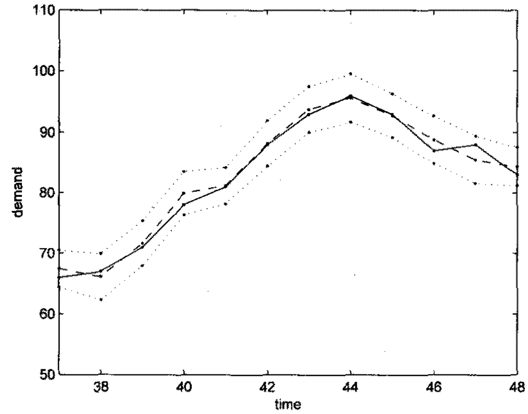

Figure 6: *One-step ahead forecasting results of FSARIMA for the test data of carpet data: forecasted bounds(dotted line), forecasted demand(dashed line) and actual demand(solid line).*

5. Conclusion

In this paper we propose the FSSVR model for seasonal time series analysis by combining the advantages of the SSVR model and Tanaka's fuzzy regression model, and apply it to forecast the monthly sales volume of the soft drink and the monthly demand of the carpet. One advantage of the FSSVR model is that the seasonal variation is easily described by the semiparametric term which is a simple linear combination of sine and cosine waves, and that the related parameters are easily estimated by a linear programming. We also deal with the problem of selecting the parameters of the FSSVR model which have a lot of influence on the forecasting accuracy. These parameters are tuned using VC-based model selection criterion.

From the empirical results of two examples, we find all of the three models have the capacity to treat growth trends and seasonal cycles. However, we notice that the FSSVR model performs better than SARIMA and FSARIMA in terms of MAE, MSE and the average length of the prediction intervals. That is, the MAE and the MSE of the FSSVR model are appreciably smaller than SARIMA and

Table 2: Actual demand, forecasted lower and upper bounds of FSSVR, SARIMA and FSARIMA for the test data of carpet data.

Date	actual demand	FSSVR	SARIMA	FSARIMA
Jan-73	66	(65.22, 69.49)	(64.74, 70.23)	(64.41, 70.55)
Feb-73	67	(64.29, 68.30)	(63.36, 68.93)	(62.32, 69.97)
Mar-73	71	(69.42, 73.35)	(68.85, 74.46)	(67.93, 75.38)
Apr-73	78	(77.78, 81.98)	(77.11, 82.70)	(76.33, 83.48)
May-73	81	(78.94, 83.61)	(78.42, 84.01)	(78.18, 84.25)
Jun-73	88	(85.95, 90.66)	(85.42, 91.01)	(84.47, 91.96)
Jul-73	93	(91.85, 96.93)	(90.99, 96.59)	(90.05, 97.54)
Aug-73	96	(93.32, 98.69)	(92.84, 98.44)	(91.72, 99.56)
Sep-73	93	(90.52, 95.93)	(89.99, 95.59)	(89.22, 96.36)
Oct-73	87	(85.43, 90.62)	(85.99, 91.59)	(84.87, 92.71)
Nov-73	88	(84.20, 89.05)	(82.69, 88.28)	(81.57, 89.41)
Dec-73	83	(81.32, 86.06)	(81.57, 87.17)	(81.15, 87.58)
Average length		4.70	5.59	7.21

FSARIMA. In addition, the interval of the FSSVR model is narrower than SARIMA and FSARIMA. We also observe it does not take much CPU time to train the FSSVR model with already adaptively tuned parameters and to forecast the future. Thus, we realize the FSSVR model provides a promising alternative to the analysis of seasonal time series consisting of limited amount of data. In addition, our FSSVR model has the following advantages over FSARIMA:

1. Our FSSVR model can capture more complex pattern of seasonal time series than FSARIMA, since the FSSVR model is inherently nonlinear model, whereas FSARIMA is linear model.

2. The FSSVR model can forecast better the central tendency than FSARIMA, since it takes over the forecasting capability of SVR for time series.

Acknowledgement

The authors wish to thank the anonymous reviewers for their comments which led to improvements of this paper.

References

Box, G. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Burman, P. and Shumway, R. (1998). Semiparametric modeling of seasonal time series, *Journal of Time Series Analysis*, **19**, 127–145.

Chang, P. T. (1997). Fuzzy seasonality forecasting, *Fuzzy Sets and Systems*, **90**, 1–10.

Chen, S. M. (1996). Forecasting enrollments based on fuzzy time series, *Fuzzy Sets and Systems*, **81**, 311–319.

Cherkassky, V., Shao, X., Mulier, F. M. and Vapnik, V. N. (1999). Model complexity control for regression using VC generalization bounds, *IEEE Transactions on Neural Networks*, **10**, 1075–1089.

Dominici, F., McDermott, A. and Hastie, T. J. (2004). Improved semi-parametric time series models of air pollution and mortality, *Journal of the American Statistical Association*, **99**, 938–948.

Franses, P. H. (1998). *Time Series Models for Business and Economic Forecasting*, Cambridge University Press, Cambridge.

Ghysels, E. and Osborn, D. R. (2001). *The Econometric Analysis of Seasonal Time Series*, Cambridge University Press, Cambridge.

Haykin, S. and Kosko, B. (2001). *Intelligent Signal Processing*, Wiley-IEEE Press, New York.

Hong, D. H. and Hwang, C. (2005). Interval regression analysis using quadratic loss support vector machine, *IEEE Transactions on Fuzzy Systems*, **13**, 229–237.

Hwang, C., Hong, D. H. and Seok, K. H. (2006). Support vector interval regression machine for crisp input and output data, *Fuzzy Sets and Systems*, **157**, 1114–1125.

Hwang, J. R., Chen, S. M. and Lee, C. H. (1998). Handling forecasting problems using fuzzy time series, *Fuzzy Sets and Systems*, **100**, 217–228.

Montgomery, D. C., Johson, L. A. and Gardiner, J. S. (1990). *Forecasting and Time Series Analysis*, McGraw-Hill, New York.

Smola, A. J., Frieβ, T. T. and Schölkopf, B. (1998). Semiparametric support vector and linear programming machines, In *Proceedings of the 1998 conference on Advances in Neural Information Processing Systems*, 585–591.

Song, Q. and Chissom, B. S. (1993a). Fuzzy time-series and its models, *Fuzzy Sets and Systems*, **54**, 269–277.

Song, Q. and Chissom, B. S. (1993b). Forecasting enrollments with fuzzy time series - part I, *Fuzzy Sets and Systems*, **54**, 1–9.

Song, Q. and Chissom, B. S. (1994). Forecasting enrollments with fuzzy time series - part II, *Fuzzy Sets and Systems*, **62**, 1–8.

Song, Q., Leland, R. P. and Chissom, B. S. (1995). A new fuzzy time-series model of fuzzy number observations, *Fuzzy Sets and Systems*, **73**, 341–348.

Tanaka, H. (1987). Fuzzy data analysis by possibility linear models, *Fuzzy Sets and Systems*, **24**, 363–375.

Tanaka, H. and Ishibuchi, H. (1992). Possibility regression analysis based on linear programming, *Fuzzy Regression Analysis*, 47–60.

Tanaka, H. and Lee, H. (1998). Interval regression analysis by quadratic programming approach, *IEEE Transactions on Fuzzy Systems*, **6**, 473–481.

Tanaka, H., Uejima, S. and Asai, K. (1982). Linear regression analysis with fuzzy model, *IEEE Transactions on Systems Man Cybernetics*, **12**, 903–907.

Tsaur, R. C., Wang, H. F. and Yang, J. C. O. (2002). Fuzzy regression for seasonal time series analysis, *Internatioanl Journal of Information Technology & Decision Making*, **1**, 165–175.

Tseng,F. M. and Tzeng, G. H. (2002). A fuzzy seasonal ARIMA model for forecasting, *Fuzzy Sets and Systems*, **126**, 367–376.

Tseng, F. M., Tzeng, G. H., Yu, H. C. and Yuan, B. J. C. (2001). Fuzzy ARIMA model for forecasting the foreign exchange market, *Fuzzy Sets and Systems*, **118**, 9–19.

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

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

Watada, J. (1992). Fuzzy time series analysis and forecasting of sales volume, *Fuzzy Regression Analysis*, Omnitech Press, Warsaw and Physica-Verlag, Heidelberg, 211–227.

West, M. and Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York.

Zadeh, L. A. (1965). Fuzzy sets, *Information and Control*, **8**, 338–353.

Received October 2008; Accepted November 2008