



Trends in
**Applied Sciences
Research**

ISSN 1819-3579



Academic
Journals Inc.

www.academicjournals.com

A Comparative Study of Neural Networks and Non-Parametric Regression Models for Trend and Seasonal Time Series

Dursun Aydın

Department of Statistics, Muğla University, 48000, Muğla, Turkey

Abstract: In this study, we will investigate and compare the performance of some forecasting methods for time series with both trend and seasonal patterns. The forecasting performance has been compared with six models and these include: Auto Regressive Integrated Moving Average (ARIMA), Smoothing Spline Model (SSM), Regression Spline Model (RSM), Additive Regression Model (ARM), Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) network models. The SSM, RSM and ARM are called as non-parametric regression models, whereas MLP and RBF are known as artificial neural network models. For these models, we conducted a comparison based on actual data sets, the number of tourist coming to Turkey. The empirical results obtained have shown that MLP performed better than other models. In addition, the SSM can be considered as an alternative to MLP.

Key words: Time series, neural networks, non-parametric regression, smoothing spline, regression spline, additive regression model

INTRODUCTION

The ARIMA models is also called as Box-Jenkins models, are the most general class of models for forecasting a time series (Box and Jenkins, 1976). A restrictive aspect of these models is that, they have linear structured models. The relationship between the variables is never linear in real life in general for most problems (Granger and Terasvirta, 1993) and using linear models is not efficient for such problems.

Artificial Neural Networks (ANN) are a class of flexible nonlinear models that can discover patterns adaptively from the data. Theoretically, it has been shown that given an appropriate number of nonlinear processing units, ANN can learn from experience and estimate any complex functional relationship with high accuracy. Recently, numerous successful applications have established their role time series forecasting problems such as prediction of electric demand, prediction of breast cancer survivability, future evolution of electricity markets, modeling rainfall-runoff process (Abraham and Nath, 2001; Delen *et al.*, 2005; Gareta *et al.*, 2006; Srinivasulu and Jain, 2006). One of the main reasons that ANN models can produce more effective results in classification, sample recognition and forecasting problems than linear models. The ANN also does not require any knowledge nor prior information about systems of interest. Therefore, using of the ANN models affecting the nonlinear structure in data may be helpful (Zhang *et al.*, 2001).

Latterly, non-parametric regression methods have become a very useful tool for non-linear data such as time series (Ferreira *et al.*, 2000). However, these approaches do not perform well when trend and seasonality is present. To overcome this problem, we considered two alternatives methods proposed in study. In both approaches the trend is specified as non-parametric, but the seasonal component specification is different. First, we take into account a partial linear model where the parametric part is a dummy-variable specification for the seasonality. Secondly, we consider the seasonal component to be a smooth function of time and, the model falls within the class of additive

models. The non-parametric regression models are discussed in detail by Hastie and Tibshirani (1999), Wahba (1990), Hardle (1991), Green and Silverman (1994) and Hardle *et al.* (2004).

There are some forecasting techniques used to forecast data time series including trend and seasonality. We considered the ones which is more flexible than linear models. One of these techniques is non-parametric regression models analyzed by Ferreira *et al.* (2000). The other technique is ANN models used for time series forecasting. The purpose of this study is to compare some forecasting models and to examine the issue whether a more complex model work more effectively in forecasting a trend and seasonal time series.

MATERIALS AND METHODS

The monthly arrival data of tourist coming to Turkey has been used for model estimation and evaluation. The artificial neural networks and non-parametric regression called as semi-parametric and additive regression models are discussed follows, respectively.

ARTIFICIAL NEURAL NETWORK (ANN) APPROACH

Multi-Layer Perceptrons (MLPs)

The MLP models used in this study are the standard three-layer MLP called as the feed-forward neural networks models with only one output node and with nodes in adjacent layers fully connected. The activation function for hidden nodes is the logistic function $f_h(x) = 1/(1+\exp(-x))$ and for the output node, the linear function. Bias (or intercept) terms are used in both hidden and output layers. MLP model can be expressed by:

$$y_t = f_o \left(b_k^o + \sum_{j=1}^p w_{jk}^o f_h \left(\sum_{i=1}^n w_{ij}^h x_i + b_j^h \right) \right) \quad (1)$$

where, y_k represents the kth output value, $\{w_{ij}^h\}$ denote the weights for the connections between the ith input and jth hidden units, $\{w_{jk}^o\}$ denotes the weights for the connections between the jth hidden and kth output units, b_j^h denote the bias for jth hidden unit, b_k^o denote the bias for kth output unit, $f_h(\cdot)$ is activation function applied to the hidden units and $f_o(\cdot)$ is the activation function applied to the output units.

Back-Propagation (BP) is the widespread approximation for training of the multi-layer feed-forward neural networks based on Widrow-Hoff training rule (Bishop, 1995; Haykin, 1999). The main idea here is to adjust the weights and the biases that minimize the sum of square error by propagation the error back at each step, namely:

$$S = \sum_{k=1}^m [y_k - \hat{y}_k]^2$$

over the first of time series, called the training set in neural networks. To minimize the sum of square error, different BP algorithms are constructed by applying different numeric optimization algorithms among gradient and Newton methods class. Conjugate Gradient (CG) algorithms provided by the Statistica Neural Network (SNN) is also employed in training of MLP networks. Another algorithm of CG algorithms is Scaled Conjugate Gradients (SCG) algorithm (Moller, 1993). The basic idea of SCG is to combine the model trust region approach with the CG approach (Bishop, 1995; Nocedal and Wright, 1999).

Radial Basis Function Networks (RBF)

The RBF is composed of three layers: an input layer, a hidden layer and an output layer. The hidden layer of an RBF is non-linear, whereas the output layer is linear. In the RBF, one hidden layer with required number of units is enough in order to model a function. The activations of hidden (radial) units are defined depending on the distance of the input vector and the center vector. Typically, the radial layer has exponential activation functions and the output layer a linear activation function. Appropriate y output vector for the x input vector is calculated as follows for n input, m output and p radial units for the RBF:

$$y_k(x) = \sum_{j=0}^p w_{kj} \phi_j(x), \quad k = 1, 2, \dots, m \tag{2}$$

where, w_{kj} , $j = 1, 2, \dots, p$ are the appropriate weights for kth output unit, $\phi_j(x)$, $j = 1, 2, \dots, p$, is the basis function of jth radial unit, w_{k0} , $k = 1, 2, \dots, m$ are the appropriate deviations for kth output unit, ϕ_0 is an extra basis function with activation value fixed at $\phi_0 = 1$. Usually, more attention paid for the following Gaussian basis function:

$$\phi_j(x) = \exp(-\|x - \mu_j\|^2 / 2\sigma_j), \quad j = 1, 2, \dots, p \tag{3}$$

where, $\mu_j = (\mu_{j1}, \mu_{j2}, \dots, \mu_{jn})$ vector is center for $\phi_j(x)$ and σ_j is deviation (or width) parameters of that function. The basis function of the unit is defined using those two parameters. Equation 2 can be written in matrix notation as:

$$y(x) = W\phi \tag{4}$$

where, $W = (w_{kj})$ and $\phi = (\phi_j)$. As can be seen from Eq. 2, the linear activation function is used in the RBF for output layer. Education is made in three stages in the RBF. In the first stage, by unsupervised education, radial basis function centers (in other words μ_j) are optimized using all $\{x^{(i)}\}$, $i = 1, 2, \dots, N$, education data. Centers can be assigned by a number of algorithms: Sub-sampling, K-means, Kohonen training, or Learned Vector Quantization. In the second stage σ_j , $j = 1, 2, \dots, p$, parameters can be assigned by algorithms explicit, isotropic and K-nearest neighbor. In the third stage of education, the basis functions that are obtained for adjusting the appropriate weights for output units are taken as fixed and deviation parameters are added to linear sum. Optimum weights are obtained by minimization of the sum of square errors:

$$E(w) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^m [y_k(x^{(i)}) - t_k^{(i)}]^2 \tag{5}$$

In Eq. 14, $t_k^{(i)}$ is the target value for output unit k when the network is presented with input vector $x^{(i)}$, $i = 1, 2, \dots, N$. Since, the equality in Eq. 4 is the quadratic function of the weights, optimum weights can be found as the solution of the linear equations system. The output layer is usually optimized using the pseudo-inverse technique.

The MLP with a defined architecture, is given by the appropriate weights and the biases of the units, but in the RBF, it is given by the center and the deviation of the radial units and by the weights and biases of the output units. As the point is given by n coordinates in n dimensional space, the number of the coordinates are equal to the linear input units n. Hence, in SNN software, the coordinates of the center radial unit are taken as weights and the deviation of the radial unit is taken as bias. As a result, radial weights denotes the center point, radial bias denotes the deviation.

THE ANN APPROACH TO TIME SERIES MODELING

The feed-forward ANN method is usually used for time series modeling and forecasting (Zhang *et al.*, 1998). For one hidden layer network architecture n:p:1 (n, number of inputs, p, number of hidden units and 1, number of outputs), inputs are the observed values of nth previous time points and outputs (targets) are (n+1)th observed value. When the network square error function is examined, it can be seen that ANN are a nonlinear functions of previous observations ($y_{t-1}, y_{t-2}, \dots, y_{t-n}$) to future observations y_t (Zhang, 2003):

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-n}, w) + \varepsilon_t$$

where, ($y_{t-1}, y_{t-2}, \dots, y_{t-n}$) denote input values, y_t denotes target (or output) value, w denote the weights of the network, ε_t denote the vector of network error at time point t . The predicted \hat{y}_t is calculated as follows:

$$\hat{y}_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-n}, w)$$

If N number of y_1, y_2, \dots, y_N observations are used for a time series and 1-step forward forecast is made, the number of training samples are $N-n$. (y_1, y_2, \dots, y_n) is taken as first input training sample and y_{n+1} is accepted as the target. The second training pattern will contain y_2, y_3, \dots, y_{n+1} as inputs and y_{n+2} as the second target output. Finally, ($y_{N-n}, y_{N-n+1}, \dots, y_{N-1}$) and y_N will be the last inputs pattern and target correspondingly. In training procedure, with the help of different BP algorithms, the parameters (weights and biases) of the network is obtained by getting closer to the minimum value of Sum of Square Error (SSE):

$$SSE = \sum_{t=n+1}^N (y_t - \hat{y}_t)^2$$

THE NON-PARAMETRIC APPROACH IN TIMES SERIES PREDICTION

The following basic model form has been considered as:

$$y(t_i) = s(t_i) + z(t_i) + \varepsilon(t_i), \quad i = 1, \dots, n \quad (6)$$

where, t_i 's are spaced in $[a, b]$, $s(t_i)$ denotes the seasonal component, $z(t_i)$ represents the trend and $\varepsilon(t_i)$ indicates the terms of error with zero mean and common variance σ_ε^2 . The model 6 can be also written as:

$$y_i = s_i + z_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (7)$$

It is assumed that the following model structure for the trend:

$$z_i = f(t_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (8)$$

where, f is a smooth function in $[a, b]$ and ε_i 's are assumed to be with zero mean and common variance σ_ε^2 , which is different from ε_i 's.

The main idea is to estimate the functions f and s . The function f is estimated as a smooth function, but the estimation of the function s is different due to seasonality (Ferreira *et al.*, 2000).

Therefore, it is considered two alternative models for the estimation of the s . Firstly, it is treated a semi-parametric model where parametric component is dummy variable for the seasonality. Secondly, it is discussed the seasonal component to be a smooth function of time and use a non-parametric method.

Semi-Parametric Regression Model

It is assumed that the seasonality is built as follows:

$$s_i = s(t_i) = \sum_{k=1}^{r-1} \beta_k D_{ki}^* + v_i, \quad i = 1, \dots, n \tag{9}$$

where, r is the number of annual observations ($r = 12$) and v_i 's are assumed to be with zero mean and common variance σ_v^2 and different from the errors in Eq. 7 and 8. D_{ki}^* 's are dummy variable that denotes the seasonal effects and β_k 's are parametric coefficients. Dummy variables are denoted by $D_{ki}^* = D_{ki} - D_{ri}$ (where, $D_{ki} = 1$ if i observation correspond to the k th month of year and $D_{ki} = 0$ otherwise) for cancels the seasonal effects when a year is completed (Ferriera *et al.*, 2000). By substitution Eq. 9 and 8 in Eq. 7, it is obtained as:

$$y_i = \sum_{k=1}^{r-1} \beta_k D_{ki}^* + f(t_i) + u_i \tag{10}$$

where, u_i 's are the sum of the random errors with zero means and constant variance $\sigma_u^2 = \sigma_\epsilon^2 + \sigma_s^2 + \sigma_v^2$.

Model 10 is called as a semi-parametric model due to consist of a parametric linear component and only a non-parametric component. The main purpose is to estimate the parameter vector β and function f at sample points t_1, \dots, t_n . For this aim, two estimation methods, called as smoothing spline and regression spline, have been considered.

Estimation with Smoothing Spline Method (SSM)

Estimation of the parameters of interest in Eq. 10 can be performed using smoothing spline. Mentioned here the vector parameter β and the values of function f at sample points t_1, \dots, t_n are estimated by minimizing the penalized residual sum of squares:

$$PSS(\beta, f) = \sum_{i=1}^n \{y_i - d_i^T \beta - f(t_i)\}^2 + \lambda \int_0^1 (f''(u))^2 du \tag{11}$$

where, $f \in C^2[0, 1]$ and d_i is the i th row of the matrix D . When the $\beta = 0$, resulting estimator has the form $\hat{f} = (\hat{f}(t_1), \dots, \hat{f}(t_n)) = S_\lambda y$, where, S_λ a known positive-definite (symmetric) smoother matrix that depends on λ and the knots t_1, \dots, t_n (Wahba, 1990; Eubank, 1999).

For a pre-specified value of λ the corresponding estimators for f based on Eq. 10 can be obtained as follows (Wahba, 1990). Given a smoother matrix S_λ , depending on a smoothing parameter λ , construct $\tilde{D} = (I - S_\lambda)D$. Then, by using penalized least squares, mentioned here estimator are given by:

$$\hat{\beta} = (D^T \tilde{D})^{-1} \tilde{D}^T y \tag{12}$$

$$\hat{f} = S_\lambda (y - D \hat{\beta}) \tag{13}$$

Evaluate some criterion function (such as cross validation, generalized cross validation) and iterate changing λ until it is minimized.

Estimation with Regression Spline Method (RSM)

Smoothing spline become less practical when sample size n is large, because it uses n knots. Regression spline is a more general approach to spline fitting. Smoothing spline require many parameters to be estimated, typically at least as many parameter as observations. A regression spline is a piecewise polynomial function whose highest order nonzero derivative takes jumps at fixed knots. Usually, regression splines are smoothed by deleting nonessential knots. When the knots are selected, regression spline can be fitted by ordinary least squares. For further discussion on selection of knots, see study of Ruppert and Carrol (2000).

f(t) in equality (Eq. 10) is approximated by:

$$f(t_i) = f(t_i, \gamma) = \gamma_0 + \gamma_1 t_i + \dots + \gamma_p t_i^p + \sum_{k=1}^K b_k (t_i - \kappa_k)_+^p, \quad i = 1, \dots, n \tag{14}$$

where, $p \geq 1$ is an integer (order of the regression spline and usually chosen a priori), b_1, \dots, b_K are independently and identically distributed (i.i.d) with $N(0, \sigma_b^2)$, $(t)_+ = t$ if $t > 0$ and 0 otherwise and $\kappa_1 < \dots, \kappa_K$ are fixed knots ($\min(t_i) < \kappa_1, \dots, \kappa_K < \max(t_i)$).

Using vector and matrix notation model 10 can be expressed as:

$$y = D\beta + Zb + \eta \tag{15}$$

Where:

$$D^T = \begin{bmatrix} 1 & 0 & \dots & 0 & -1 & 1 & 0 & \dots & 1 & t_1 & \dots & t_n \\ 0 & 1 & \dots & 0 & -1 & 0 & 1 & \dots & 1 & t_1^{p-1} & \dots & t_n^{p-1} \\ & & \cdot & & & & & & & \cdot & & \cdot \\ & & & \cdot & & & & & & \cdot & & \cdot \\ & & & & \cdot & & & & & \cdot & & \cdot \\ 0 & 0 & \dots & 1 & -1 & 0 & 0 & \dots & 1 & t_1^p & \dots & t_n^p \end{bmatrix}$$

and

$$Z = \begin{bmatrix} (t_1 - \kappa_1)_+^p & \dots & (t_1 - \kappa_K)_+^p \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ (t_n - \kappa_1)_+^p & \dots & (t_n - \kappa_K)_+^p \end{bmatrix}$$

where, $b = (b_1, \dots, b_K)^T$ is the vector of coefficients and $\eta = (\eta_1, \dots, \eta_n)^T$ is a vector of the random error. Predicted value of \hat{y} in Eq. 10 is given by:

$$\hat{y} = \hat{\gamma}_0 + \hat{\gamma}_1 t_i + \dots + \hat{\gamma}_p t_i^p + \hat{\beta}_1 D_1 + \dots + \hat{\beta}_{K-1} D_K + (Z_1, \dots, Z_K) (\hat{b}_1, \dots, \hat{b}_K)^T \tag{16}$$

Regression spline estimators

$(\hat{\beta} = (\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_p, \hat{\beta}_1, \dots, \hat{\beta}_{K-1})^T, \hat{f} = (\hat{b}_1, \dots, \hat{b}_K))$ of (β, f) are defined as the minimizer of:

$$PSS(\beta, f) = \sum_{i=1}^n \{y_i - d_i^T \beta - f(t_i)\}^2 + \lambda \sum_{k=1}^K b_k^2 \tag{17}$$

where, $\lambda > 0$ is a smoothing parameter such as in Eq. 11. As $\lambda \rightarrow \infty$, the regression spline converges to a pth degree polynomial fit. As $\lambda \rightarrow 0$, the regression spline converges to the Ordinary Least Squares (OLS) fitted spline. For a pre-specified value of λ the corresponding estimators for β and f based on Eq. 15 can be obtained as follows (Ruppert *et al.*, 2003):

$$\hat{\beta} = (D^T \hat{\Sigma}^{-1} D)^{-1} D \hat{\Sigma}^{-1} y \tag{18}$$

where, $\hat{\Sigma} = ZZ^T \hat{\sigma}_0^2 + \text{diag}(\hat{\sigma}_{\eta_i}^2)$, $i = 1, 2, \dots, n$.

$$\hat{f} = (\hat{b}_1, \dots, \hat{b}_k)^T = \hat{\sigma}_0^2 Z^T \hat{\Sigma}^{-1} (y - D \hat{\beta}) \tag{19}$$

The smoothing parameter λ and the number of knots K must be selected in implementing the regression spline. However, λ plays a more essential role. See for a detailed discussion of the knot selection (Ruppert, 2002). The solution can be obtained in S-Plus.

Additive Regression Model (ARM)

The semi-parametric model has been used for estimation of the parameters in Eq. 10. However, there are situations in which a dummy variable specification does not capture all fluctuations because of existing any seasonal effect. Therefore, here, a more general case for seasonal component has been considered as follows:

$$s_i = g(t_i) + v_i, \quad i = 1, \dots, n \tag{20}$$

where, $g \in C^2[a, b]$, v_i 's denote the terms of random error with zero mean and common variance σ_v^2 . By substitution of the Eq. 8 and 20 in Eq. 7, y_i is obtained as:

$$y_i = g(t_i) + f(t_i) + u_i, \quad i = 1, \dots, n \tag{21}$$

where, u_i 's are the terms of random error with zero mean and constant variance $\sigma_u^2 = \sigma_s^2 + \sigma_e^2 + \sigma_v^2$.

The model presented in Eq. 21 has a fully non-parametric model because parametric component is missing. These models are called as additive non-parametric regression models. In order to estimate the model in Eq. 21, the criterion Eq. 11 and 21 can be generalized in an obvious way. Estimator of the model 21 is based on minimum of the penalized residual sum of squares (Hastie and Tibshirani, 1999):

$$\text{PSS}(f, g) = \sum_{i=1}^n \{y_i - f(t_i) - g(t_i)\}^2 + \lambda_1 \int_0^1 (f''(u))^2 du + \lambda_2 \int_0^1 (g''(u))^2 du \tag{22}$$

where, the first term in Eq. 22 denotes the Residual Sum of the Squares (RSS) penalizing the lack of fit, the second term multiplicand by λ_1 denotes the roughness penalty for the f and the third term multiplicand by λ_2 denotes the roughness penalty for g . Firstly, Eq. 22 can be written as:

$$\text{PSS}(f, g) = (y - f - g)^T (y - f - g) + \lambda_1 f^T K_f f + \lambda_2 g^T K_g g \tag{23}$$

where, K_f is a penalty matrix for f and K_g is a penalty matrix for g . Then, by differentiating according to f and g and afterwards, by setting to zero, the estimators of f and g are defined as:

$$\hat{f} = (I + \lambda_1 K_f)^{-1} (y - g) = S_{\lambda_1} (y - g) \quad (24)$$

$$\hat{g} = (I + \lambda_2 K_g)^{-1} (y - f) = S_{\lambda_2} (y - f) \quad (25)$$

MEASURING MODEL PERFORMANCE METHODS

In order to evaluate the predictions of a model with observations, the following statistical performance measures, which include the Mean Square Error (MSE), the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) have been used (Carey and Rob, 2002). Forecast evaluation measurements are defined as following:

$$\begin{aligned} \text{MSE} &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, & \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \\ \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \text{and} & \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|} (100\%) \end{aligned}$$

where, y_i is represent the observed values, \hat{y}_i is indicate the forecasted values and Mean is the arithmetic mean value.

A perfect model would have MSE or RMSE, MAE and MAPE ≈ 0.00 . Of course, because of the influence of random errors, there is no such thing as a perfect model in time series modeling.

EXPERIMENTAL EVALUATIONS

Here, a real data sets occurred in Turkey is discussed as experimental. Appropriate non-parametric regression and ANN models were chosen by doing experiments to forecast and these models are also compared with each other. To conduct these experiments, SNN, S-Plus and R-Programs are used.

Data Set

The real data set is from Turkish Statistical Institute (TÜİK) analyzed by Aslanargun *et al.* (2007). The data can be found in www.tuik.gov.tr and represent the number of monthly tourist coming to Turkey between January 1984 and December 2003. The data set is divided into two parts for the use in training and forecasting. In the first part, 216 monthly data are taken into account for the period of the January 1984-December 2001 period. These data are used in training to construct the models. In the second part, with the help of the models constructed in the first part, the performances of those models are calculated using the 24 monthly data for the January 2002-December 2003 period.

Selection of the Appropriate ANN Models

The 216 monthly data were used in training of the network. An evaluation of the model was made depending on the forecasts of the 24 monthly data. The best model have been selected by scores obtained from MSE, RMSE, MAE and MAPE. As the initial weight and bias values of the network were random, experiments with 150 replications were made for the same network structure and the models giving the best forecasts were determined. Because of the monthly tourist arrival data included seasonality, the number of input units was determined as 12. During these experiments, various neural network algorithms with single layer, with one or two hidden layers, MLP and the RBF models were applied on the data set. As the initial 12 data were lost because of the seasonality, 204 from the 216 data were used to adjust the weights. In the training stage of the network, data were divided into

two parts: 132 of the 204 data were used for training and 72 data were used for validation. This division was used to restrict memorization of the network and provided for better forecasts (Bishop, 1995; Haykin, 1999).

Among the ANN models, the MLP (12:3:1) model have indicated the best performance. The CG algorithm indicated the best performance on the 51th epoch. A hyperbolic tangent function is applied in the hidden unit and the linear activation function is applied in the output unit. The weights and biases of the MLP (12:3:1) model are shown in Table 1.

Constructing the Appropriate Non-Parametric Models

The 216 monthly data set including January 1984-December 200 period were used in training the models called as semi-parametric and additive regression model. For estimation of these models, we need to select smoothing parameter λ . In general, the λ can be selected by using automatic selection methods such as Cross Validation (CV), Generalized Cross Validation (GCV). In practice, it is reasonable to select the λ by specifying degrees of freedom ($df = \text{trace}(S_\lambda)$) for the non-parametric components (Hastie and Tibshirani, 1999). Therefore, the df is used to select the smoothing parameter λ in smoothing spline. On the other hand, both the smoothing parameter λ and the number of knots K must be selected in implementing the regression spline. The solution is obtained by S-Plus and R

Table 1: The weights and biases of the MLP (12:3:1)

	2.1	2.2	2.3	3.1
Thresh	0.685052	0.651214	0.544668	-0.57875
1.1	-0.069968	0.085128	0.186845	
1.2	0.021359	0.019630	0.072838	
1.3	-0.026189	0.054980	0.029329	
1.4	-0.016201	0.035831	-0.004167	
1.5	0.000860	0.009109	0.006610	
1.6	0.008219	0.032057	0.026170	
1.7	0.089622	-0.009947	-0.052460	
1.8	0.073210	0.027442	0.113376	
1.9	0.093421	-0.030547	-0.012583	
1.10	0.037213	-0.002316	-0.010518	
1.11	-0.013186	0.011851	0.093191	
1.12	-0.166940	0.139571	0.238188	
2.1				-1.30809
2.2				1.47106
2.3				0.93885

The row and column header numeric terminology first lists the layer, then the unit number within the layer. For example, 2.1 stands for unit 1 in layer 2

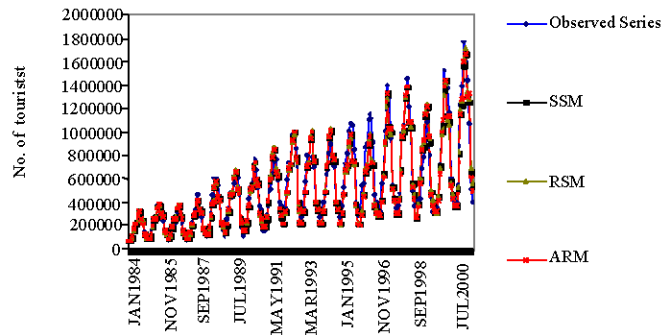


Fig. 1: Observed number of tourists and their estimation values obtained by appropriate SSM, RSM and ARM

programs. Secondly, we consider an additive regression model where both the seasonal and the trend components are unknown smooth functions of time. So, we have chosen λ_1 and λ_2 parameters by specifying the df. Observed number of tourist for the January 1984-December 2001 period and the tourist estimation results obtained by appropriate non-parametric models are shown in Fig. 1.

Figure 1 shows that the data have an upward trend together with seasonal variation. In these series, the main advantage of the non-parametric regression models that perform prediction by means of the models without data loss. In other words, as the initial 216 data are used to estimate the model, 216 residuals are obtained for each of non-parametric model. Consequently, 216 data was used to estimate non-parametric regression models called as smoothing spline, regression spline and additive regression models.

COMPARISONS OF THE MODELS

An evaluation of the non-parametric regression models obtained by using S-Plus and R programs was made depending on the forecasts for the January 2002-December 2003 period. Namely, we calculated the performance values of the non-parametric regression and ANN models for the mentioned period. Table 2 shows for each model, the values of the MSE, RMSE, MAE and MAPE called as performance indicators. The scores of the MSE, RMSE, MAE and MAPE belong to MLP model are lower than the ones of the concurrent specifications. In that sense, the MLP model outperforms the other formulations. Furthermore, as shown Table 2, the SSM model has indicated better performance than the RSM and ARM since SSM has had the lowest scores of the model evaluation criteria among the non-parametric models.

For test data composed of the 24 values, the observed and forecasted values obtained by different models are calculated, but they are only given as graphically since they would occupy very much place. Observed and their forecasted values by the models are shown in Fig. 2.

As shown in Fig. 2, the output from the MLP model shows that predicted and observed values very close to each other. However, the output from the RBF model presents that there is big

Table 2: Performance values for the selected models

Models	MSE	RMSE	MAE	MAPE
ARIMA (1,1,1) (1,1,0) ₁₂	3.06E+10	174926.3	137589.4	13.74
SSM	2.45E+10	156520.6	117409.6	14.05
RSM	2.55E+10	159783.1	119122.2	14.35
ARM	3.41E+10	184594.6	141483.7	15.85
MLP (12:3:1)*	2.12E+10	145575.7	123480.7	13.73
RBF (12:7:1)	1.80E+11	424314.6	352104.6	34.02

*The model having best performance

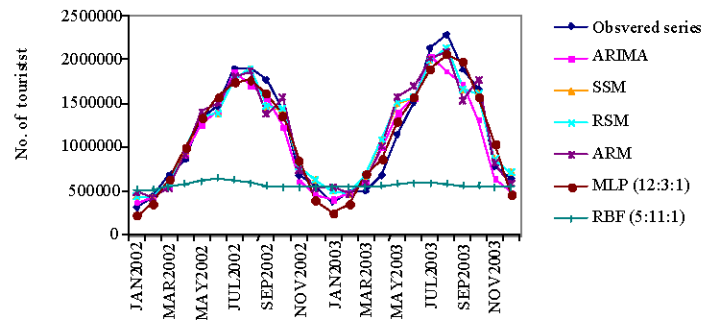


Fig. 2: Observed and their forecasted values by the models

difference between actual and predicted observations. The outputs shown in Fig. 2 support the ideas advanced in the Table 2.

RESULTS AND DISCUSSION

It is proved that the ANN model gives a comparable performance to the ARIMA model for longer time horizons (Zhang, 2003). There have a great deal of studies comparing various neural network and some non-parametric methods for time series. For example, Sallur-Ruiz *et al.* (2008) indicated that MLP model outperformed than other methods. Different ANN and non-parametric methods have been applied to time series data set in the study. However, most of them are designed to predict the time series. In this study, these methods mentioned here are considered as a comparative on aspect of forecasting performance.

It is known that neural networks very good performance in time series forecasting problems. As can be seen from Table 2, the MLP (12:3:1) have performed very good performance, while the RBF have not performed well enough. This case supports the idea that the RBF is usually unsuccessful in extrapolation problems (Bishop, 1995). On the other hand, as shown Fig. 2, the values forecasted by ARIMA, SSM, RSM, ARM and MLP (12:3:1) are closely following the real observed values, whereas the values forecasted by the RBF is not proper the observed series. It is also seen that RBF is not good predictor for such time series.

Non-parametric regression models can be considered as alternative method to ANN models. The SSM indicated a good empirical performance among the non-parametric models, while another non-parametric model called as ARM showed the worst performance. As a result, our opinion is that MLP can be useful in time series forecasting problems included seasonality and trend. We propose to use the MLP, especially on these type series. Furthermore, the SSM model can be used as an alternative method to the MLP.

REFERENCES

- Abraham, A. and B. Nath, 2001. A neuro-fuzzy approach for modeling electricity demand in Victoria. *Applied Soft Comput. J.*, 1: 127-138.
- Aslanargun, A., M. Mammadov, B. Yazici and S. Yolacan, 2007. Comparison of ARIMA, neural networks and hybrid models in time series: Tourist arrival forecasting. *J. Stat. Comput. Simulat.*, 77: 29-53.
- Bishop, C., 1995. *Neural Networks for Pattern Recognition*. 1st Edn., Oxford University Press, USA., ISBN-13: 978-0198538646.
- Box, G.E.P. and G.M. Jenkins, 1976. *Time Series Analysis: Forecasting and Control*. Chapter 2, Holden Day, Oakland, California, ISBN: 0-8162-1104-3.
- Carey, G. and L. Rob, 2002. Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Manage.*, 23: 499-510.
- Delen, D., G. Waller and A. Kadam, 2005. Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intell. Med.*, 34: 113-127.
- Eubank, R.L., 1999. *Nonparametric Regression and Smoothing Spline*. 2nd Edn., Marcel Dekker, Inc., New York.
- Ferreira, E., V. Núñez-Antóna and J. Rodríguez-Póob, 2000. Semi-parametric approaches to signal extraction problems in economic time serie. *Comput. Stat. Data Anal.*, 33: 315-333.
- Gareta, R., L.M. Romeo and A. Gil, 2006. Forecasting of electricity prices with neural networks. *Energy Convers. Manage.*, 47: 1770-1778.
- Granger, C.W.J. and T. Terasvirta, 1993. *Modeling Nonlinear Economic Relationships*. Oxford University Press, Oxford.

- Green, P.J. and B.W. Silverman, 1994. *Nonparametric Regression and Generalized Linear Models* 1st Edn., Chapman and Hall, London.
- Hardle, W., 1991. *Applied Nonparametric Regression*. 1st Edn., Cambridge University Press, Cambridge.
- Hardle, W., M. Müller, S. Sperlich and A. Werwatz, 2004. *Nonparametric and Semiparametric Models*. 1st Edn., Springer, New York.
- Hastie, T. and R.J. Tibshirani, 1999. *Generalized Additive Models*. 1st Edn., Chapman and Hall, London.
- Haykin, S., 1999. *Neural Networks: A Comprehensive Foundation*. 2nd Edn., Prentice-Hall International Inc., New Jersey, ISBN: 0-13-908385-5.
- Moller, F.M., 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6: 525-533.
- Nocedal, J. and S.J. Wright, 1999. *Numerical Optimization*. Springer Verlag, New York.
- Ruppert, D. and R. Carroll, 2000. Spatially-adaptive penalties for penalized spline. *N. Z. J. Stat.*, 42: 205-224.
- Ruppert, D., 2002. Selection the number of knots for penalized spline. *J. Comput. Graphical Stat.*, 11: 735-757.
- Ruppert, D., M.P. Wand and R.J. Carroll, 2003. *Semi-Parametric Regression*. 1st Edn., Cambridge University Press, Cambridge.
- Sallur-Ruiz, E., B.J. Ordiers, E.P. Vergara and S.F. Capuz-Rizo, 2008. Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence-based models in Mexicali, Baja California (Mexico) and Calexico, California (US). *Environ. Modell. Software*, 8: 1056-1069.
- Srinivasulu, S. and A. Jain, 2006. A comparative analysis of training methods for artificial neural network rainfall-runoff models. *Applied Soft Comput.*, 6: 295-306.
- Wahba, G., 1990. *Spline Models Of Observational Data*. 1st Edn., University of Wisconsin, Madison, Pennsylvania.
- Zhang, G.P., B.E. Patuwo and M.Y. Hu, 1998. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.*, 14: 35-62.
- Zhang, G.P., G.E. Patuwo and M.Y. Hu, 2001. A simulation study of artificial neural networks for nonlinear time-series forecasting. *Comput. Operat. Res.*, 28: 381-396.
- Zhang, G.P., 2003. Time-series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50: 159-175.