Comparison Studies of Hybrid and Non-hybrid Forecasting Models for Seasonal and Trend Time Series Data*

Chulwoo Jeong
Korea Institute for Defense Analyses
(jchuroo@naver.com)

Myung Suk Kim
Global Service Management Department, Sogang Business School
(myungkim@sogang.ac.kr)

In this article, several types of hybrid forecasting models are suggested. In particular, hybrid models using the generalized additive model (GAM) are newly suggested as an alternative to those using neural networks (NN). The prediction performances of various hybrid and non-hybrid models are evaluated using simulated time series data. Five different types of seasonal time series data related to an additive or multiplicative trend are generated over different levels of noise, and applied to the forecasting evaluation. For the simulated data with only seasonality, the autoregressive (AR) model and the hybrid AR-AR model performed equivalently very well. On the other hand, if the time series data employed a trend, the SARIMA model and some hybrid SARIMA models equivalently outperformed the others. In the comparison of GAMs and NNs, regarding the seasonal additive trend data, the SARIMA-GAM evenly performed well across the full range of noise variation, whereas the SARIMA-NN showed good performance only when the noise level was trivial.

Received : January 29, 2013 Revision : February 27, 2013 Accepted : March 13, 2013
Type of Submission : English Fast-Track Corresponding author : Myung Suk Kim

1. Introduction

Recently, data are exponentially increased with the emergence of various types of information channels like Social Networking Services. As the data storage techniques are developed, high capacity data or big data analysis becomes available. Some meaningful implications are often

* This work has been supported by the Sogang Univeristy Research Grant (201010024.01) and the Sogang Business School's World Class University Project (R31-20002) funded by Korea Research Foundation.

derived from the big data analysis. In particular, versatile time series type big data are cumulated in numerous institutions. In this study, we consider five different types of time series data simulation, and try to find appropriate forecasting models regarding each type of simulated time series data. If some real time series data have a similar pattern with one of these five simulated time series data, we may consider applying the recommended models from the simulation results.

There have been many attempts to model seasonal and trend time series data. In terms of modeling the patterns of the underlying time series, two types of approach, namely linear and nonlinear, can be categorized. Firstly, a type of linear model known as the seasonal autoregressive integrated moving average (SARIMA) was proposed by Box and Jenkins (1976), and this has been commonly applied over the past several decades. After this, nonlinear models such as neural networks (NN or NNs) have received considerable attention due to their ability to capture nonlinear patterns in the underlying series (Adya and Collopy, 1998; Bodyanskiy and Popov, 2006; Freitas and Rodrigues, 2006; Barbounis and Teocharis, 2007; Celik and Karatepe, 2007).

On the other hand, hybrid modeling approaches combining linear and nonlinear models have been proposed to utilize the merits of both types of model. Tseng et al. (2002) suggested combining NNs and SARIMA in such a way that the forecasts and residuals from the SARIMA model are applied to input variables of NNs. Similarly, Zhang (2003) suggested a hybrid approach using NNs and ARIMA, wherein NNs are applied to model the residuals from the ARIMA model. Although these studies have commonly argued for the superiority of a hybrid model over the individual models, extensive comparison of the prediction accuracy between such models has not been previously studied. Furthermore, existing hybrid approaches are limited in that they only combine NNs and ARIMA type models.

Recently, a nonparametric additive model has been widely applied in nonlinear time series data forecasting (Prada-Sánchez and Febrero-Bande, 1997; Dominici et al., 2002; Berg, 2007). Since in this model the response variable is allowed to have many types of distribution (e.g. Normal, Poisson, Logistic), it tends to be referred to as a generalized additive model (GAM). GAMs differ from linear models in that they are capable of adjusting for the nonlinear confounding effects of seasonality and trend. Moreover, GAMs are unlike NNs in that they do not require data preprocessing approaches. Nelson et al. (1999) argue that the forecasting error of NNs can be reduced via detrending or deseasonalization. However, many researchers have recognized the difficulty in distinguishing seasonality from non-seasonal components and in modeling trend patterns (Nelson and Plosser, 1982; Ittig, 1997). Unless the trend or seasonality is specified correctly, preprocessing may not guarantee better prediction results. Due to these strengths of GAMs over NNs and linear models, hybrid models employing a GAM are newly suggested for forecasting seasonal and trend time series data in this article.

Several types of seasonal time series data with additive or multiplicative trends are generated over different levels of noise, which are then applied to an evaluation of forecasting accuracy, comparing between the suggested hybrid models and the other competing models. For the simulated data with only seasonality without any trend, the AR model and the hybrid AR-AR model equivalently performed very well, whereas the SARIMA model and the SARIMA related hybrid models showed poor performances. On the other hand, when using a time series data set employing a linear or a quadratic trend, the SARIMA model and some of its hybrid models both outperformed the other competitors. In the comparison between GAMs and NNs, both the SARIMA-GAM and SARIMA-NN hybrid models tended to show very good prediction performances for the multiplicative trend models. Regarding the additive trend model, the SARIMA-GAM performed well for the full range of noise variation, whereas the SARIMA-NN showed good performance only when the noise level was trivial.

The remainder of this article is arranged as follows. Section 2 briefly defines the mathematical concepts of the linear, nonlinear and hybrid models. Likewise, the newly proposed hybrid models using the GAM are introduced. The design of the simulation studies is explained in Section 3. A summary of the simulation analysis results is also reported in bulletin form. Some concluding remarks are given in Section 4.

2. Forecasting Models

In this article, two linear models (AR and SARIMA) and two nonlinear models (NN and GAM) are compared and evaluated for various seasonal and trend time series data. Furthermore, a total of 16 hybrid models combining these four models will be also examined. In the following sub-sections, the concepts and mathematical expressions of the linear, nonlinear and the hybrid models are briefly explained. The new hybrid models using GAMs are also introduced.

2.1 Linear models

The SARIMA model (Box and Jenkins, 1976) is a popular linear forecasting scheme for seasonal and trend time series data. For time series data $\{y_t\}$, this model can be represented as

$$\phi_p(B)\Phi_p(B^S)(1-B)^d(1-B^S)^D(y_t-c) \quad (1)$$
$$=\theta_q(B)\Theta_Q(B^S)\varepsilon_t,$$

where ε_t is usually a Gaussian white noise; c is a constant; p, d, q, P, D and Q are integers; $\phi(B)$ and $\theta(B)$ of order p and q are the polynomials representing autoregressive and moving average components, respectively; $\Phi_p(B^S)$ and $\Theta_Q(B^S)$ of orders P and Q are the polynomials representing seasonal autoregressive and moving average components, respectively; d and D are orders of differencing; $(1-B)$ and $(1-B^S)$ are difference operators and S is the length of the seasonal cycles. This model is usually known as the SARIMA $(p, d, q) \times (P, D, Q)_s$.

The SARIMA model is reduced to a seasonal autoregressive moving average (SARMA) if d and D are at zero in (1). In particular, the SARIMA model becomes autoregressive (AR) if d, q, P, and D are all at zero. The AR model can be written as

$$y_t = c + \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t, \tag{2}$$

where ϕ_j is the j th weight and p is the order (or the number of lags). This model is often expressed as $AR(p)$.

Estimation and prediction will be carried out using the functions in the package of *R* program. Under the Normal distribution assumption on the error term, the parameters of the SARIMA model can be estimated using maximum likelihood methods. The Akaike information criterion (AIC) has been applied to the model selection in this study. Regarding the AR model, the 'Yule-Walker' estimation method has been applied using a function in *R* program, which is the default method. More detailed explanations on the utilized *R* program functions are given in subsection 3.2.

2.2 Nonlinear Models

Among various types of NNs models, the model of three-layer perceptron NN is employed in this study, which has been widely applied in empirical studies (e.g., Adya and Collopy, 1998). The model is given by the equation

$$y_t = \alpha_0 + \sum_{j=1}^{k} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{m} \beta_{ij} y_{t-i}\right) + \varepsilon_t, \tag{3}$$

where $a_j (j = 0, 1, 2, \cdots, k)$ and $\beta_{ij} (i = 1, 2, \cdots, m;\ j = 0, 1, 2, \cdots, k)$ are the path coefficients; m is the number of lags and k is the number of hidden nodes and $g(\cdot)$ is the activation function in each of hidden nodes.

The performance of an NN model is known to depend on several components, including the number of lags, hidden layers and hidden nodes, and its activation function (Hansen et al., 1999). In this paper, the number of hidden layers is set to be one and the logistic function is applied as the activation function, which is given by $g(x) = 1/(1+e^{-x})$. Regarding the other two components, the number of lags and hidden nodes will be selected for a given data set for better prediction performance via the AIC criterion.

Another competing nonlinear model considered in this study is the GAM. Since it was initially proposed by Hastie and Tibshirani (1986), it has been widely applied to numerous fields, including time series data forecasting (Dominici et al., 2002). The popularity of GAMs in the area of time series forecasting is due to their flexibility, in that they allow for nonparametric adjustments for the nonlinear confounding effects of seasonality and trend. GAMs also require no assumption of relationships between variables, unlike the other linear alternatives. Furthermore, data preprocessing approaches such as detrending or deseasonalization are not necessary in GAMs, unlike in NNs.

The GAM for time series forecasting is mathematically expressed by the equation

$$y_t = w_0 + \sum_{j=1}^{m} s_j(y_{t-j}) + \varepsilon_t, \tag{4}$$

where w_0 is a constant; $s_j(\cdot)$ is the j th additive function weight; m is the number of lags; ε_t is the t th error term, and ε_t for all t are uncorrelated.

The function $s_j(\cdot)$ can be fitted using parametric or non-parametric methods, which provides a potentially better fit to data than do the other methods due to its strength of capturing both the linear and nonlinear patterns of underlying processes. In this study, the spline function is applied for $s_j(\cdot)$. For further details about the additive functions in GAMs, refer to Hastie and Tibshirani (1986) and Wood (2006). The AIC will be again employed for model selection.

2.3 Hybrid Models

Recently, hybrid models combining the strengths of linear and nonlinear schemes have been suggested. Tseng et al. (2002) proposed a hybrid model combining NNs and SARIMA models in which an NN model is applied to forecast y_{t+1} by using residuals from SARIMA and observations; $y_t, \cdots, y_{t-m}$. Zhang (2003) suggested a hybrid model from NN and ARIMA model. The forecasted value $\hat{y}_{t+1}$ at time $t+1$ can be obtained as follows :

$$\hat{y}_{t+1} = \hat{A}_{t+1} + \hat{N}_{t+1}, \tag{5}$$

where $\hat{A}_{t+1}$ is the fitted value from ARIMA model; $\hat{N}_{t+1}$ is an estimates via NN model, where residuals from ARIMA fits are used as inputs of NN. The linear model fit is first applied, and the nonlinear model fit follows. Tseng, et al. (2002) and Zhang (2003) empirically showed that applying the linear and nonlinear models in sequence can improve the forecasting performance.

The suggested hybrid models in this paper are similar to those in the previous methods, but are also novel in some aspects. First, we initially suggest applying a GAM as the nonlinear model, as an alternative to the NNs. Second, the order in which the linear model and nonlinear model are applied can be switched. For example, the GAM can be estimated first, and its residuals can be used to fit the model. The notation 'GAM-AR' will be employed in this case. Finally, we also allow applying linear models together or nonlinear models together. For example, the GAM is estimated first, and its residuals are then used to fit another GAM; for this case, the notation 'GAM-GAM' will be used.

The suggested GAM related hybrid models can be expressed as follows.

$$\begin{aligned} y_t = & f(y_{t-1}, y_{t-2}, \cdots, y_{t-m}) \\ & + g(a_{t-1}, a_{t-2}, \cdots, a_{t-m}) + \varepsilon_t \end{aligned} \tag{6}$$

where $f(\cdot)$ is either a linear model (of AR, SARIMA) or an additive function in the GAM; $g(\cdot)$ is also either a linear model (of AR, SARIMA) or an additive function in the GAM; a_t's are residuals from $f(y_{t-1}, y_{t-2}, \cdots, y_{t-m})$ and ε_t is the error term at t, and ε_t for all t are uncorrelated. Here, at least one of $f(\cdot)$ or $g(\cdot)$ should include the additive function in the GAM. Therefore, a total of seven models (GAM-AR, GAM-SARIMA, GAM-NN, GAM-GAM, AR-GAM, SARIMA-GAM, NN-GAM) will be considered.

For extensive comparison of the various types of hybrid models, we have also considered additional nine hybrid models combining two linear AR and SARIMA models and a nonlinear NN model. Prediction performances between the total 16 hybrid models and the four pure models are evaluated with different five types of simulated data in the following sections.

3. Simulation Analysis

The forecasting performance for the four pure models (AR, SARIMA, NNs and GAM) and the 16 hybrid models are compared and evaluated via simulation data in this section. Time series data are generated via combining the three important components of seasonality, trend, and irregularity. The specific data generating process (DGP) and simulation analysis results are explained below.

3.1 Data Generating Process and Prediction

Various types of seasonal and trend time series data are generated using the following five approaches : (i) a seasonality (ii) an additive combination of a linear trend and a seasonality (iii) a multiplicative combination of a linear trend and a seasonality (iv) an additive combination of a quadratic function trend and a seasonality, and (v) a multiplicative combination of a quadratic function trend and a seasonality. These five approaches are realized via the following five data generating processes, which are written as

$$\text{DGP 1}:\ y_t = 10 + 3\cos\left(2\pi t\frac{10}{300} + 8\right) + \varepsilon_t \tag{7}$$

$$\text{DGP 2}:\ y_t = (20 + 0.1t) + \left(10 + 3\cos\left(2\pi t\frac{10}{300} + 8\right)\right) + \varepsilon_t \tag{8}$$

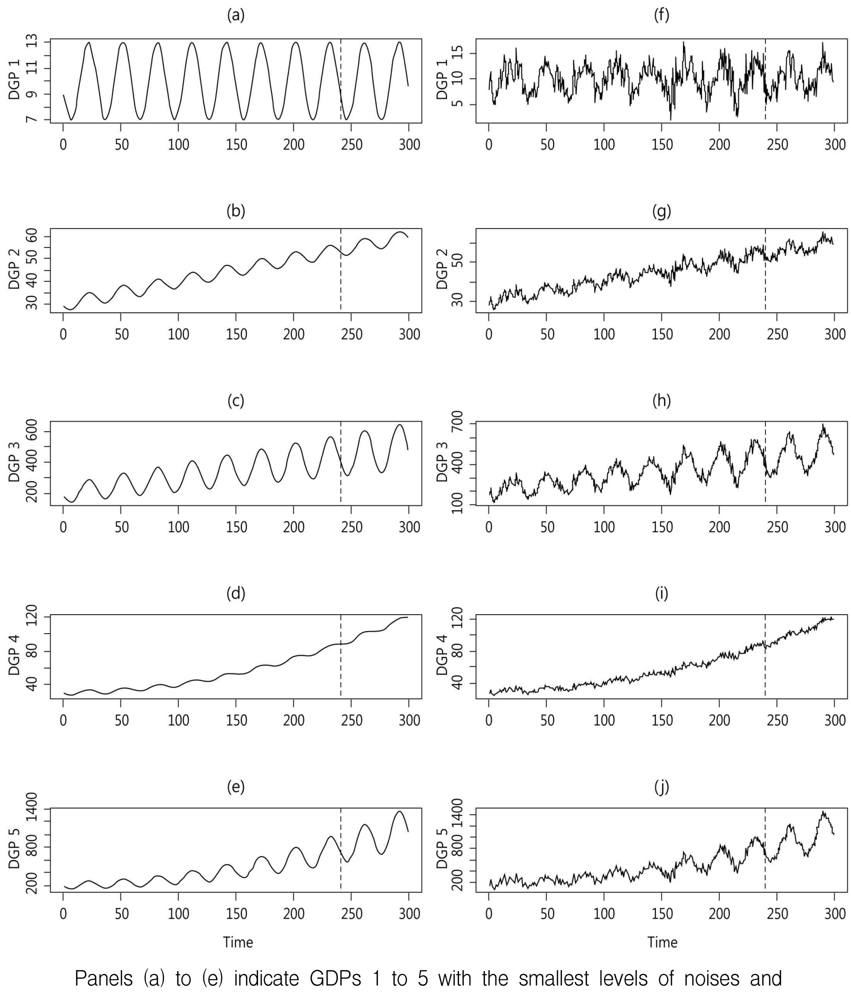
$$\text{DGP 3}:\ y_t = (20 + 0.1t)\left(10 + 3\cos\left(2\pi t\frac{10}{300} + 8\right)\right) + \varepsilon_t \tag{9}$$

$$\text{DGP 4}:\ y_t = (200 + 0.001t^2) + \left(10 + 3\cos\left(2\pi t\frac{10}{300} + 8\right)\right) + \varepsilon_t \tag{10}$$

$$\text{DGP 5}:\ y_t = (20 + 0.001t^2)\left(10 + 3\cos\left(2\pi t\frac{10}{300} + 8\right)\right) + \varepsilon_t \tag{11}$$

where $\varepsilon_t \sim N(0, \sigma^2)$. To examine the effects of irregularity, various levels of innovation variance are allowed. In particular, σ is set to be 0.01, 0.1, 0.5, 1 and 2 for DGPs 1, 2 and 4, 1, 5, 10, 20 and 30 for DGP 3, and 1, 10, 30, 40 and 50 for DGP 5 to reflect the more explicit effects of irregularity. Panels (a) to (e) in <Figure 1> describe the realized DGPs 1 to 5 with the smallest level of noise. For the purpose of comparing the effects of irregularity, the DGPs 1 to 5 with the largest level of noise are depicted in panels (f) to (j) in <Figure 1>. The amplitude of vibration due to the seasonal factor tends to show a broadening trend in the multiplicative models, whereas it seems to be parallel in the additive models. Regardless of the different levels of noise, the patterns of the five DGP counterparts seem to be still evident in terms of seasonality and trend.

There are some real data demonstrating these five types of data processes. For example, dynamics of 5-minute call arrivals at a US North American bank call center appear to follow the pattern of DGP 1. The monthly US retail sales series at department stores (from US Census Bureau) seem to be like DGP 2. The monthly sales of soft drinks (from Montgomery et al., 1990) seem to have a similar pattern with DGP 3. Some monthly observed US consumer good production (from Federal Reserve Board) appears to demonstrate DGP 4. The monthly US national monetary aggregates are similar to DGP 5.



Panels (a) to (e) indicate GDPs 1 to 5 with the smallest levels of noises and panels (f) to (j) indicate DGPs 1 to 5 with the largest levels of noise.

<Figure 1> Description of Simulation Data

To evaluate the forecasting performance between models, the following procedures are considered. For each DGP, 300 points are generated, which are then divided into two data sets: one works as an in-sample data set with 240 data points from the initial point, whereas the other works as an out-of sample data set with the remaining 60 data points.

The first one time ahead (the 241^{st} data point) forecasting is carried out using the in-sample data (the initial 240 data points). Then, the original in-sample data set is updated by excluding the 1^{st} data point and including the 241^{st} data point, which maintains the total number of in-sample data set equally 240. The next one time ahead (the 242^{nd} data point) forecasting is done using this updated in-sample data. In this way, the one time ahead prediction procedures continue up to the last 300^{th} data point. By comparing the predicted data points and actual data points over the total forecasting period (the 241^{st}~300^{th} data points), the prediction performances are measured.

The root mean square error (RMSE) statistic using the out-of-sample data set is computed to evaluate the forecasting performance of the models in this study. The RMSE is a popularly applied standard prediction performance measure and can be written as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}(y_i - \hat{y_i})^2}{M}} \tag{12}$$

where $\hat{y}_i$ is the predicted point for actual point y_i and M is the number of predicted points. In this study, $M = 60$.

We need to note that there has been a data preprocessing in our analysis. If there are seasonal variations, they tend to be removed from the original time series before forecasting in the literature. So, we first estimate the seasonal effect from each DGP using the function 'decompose' in *R* program. Then the estimated seasonal effects are deleted from the original time series. Lastly, the estimated effects are scaled back for forecasting.

3.2 Results

Regarding the optimal model selection, various statistics packages for *R* program are utilized in this study. The function 'ar' in the package 'stats', which is installed in *R* by default, is used for estimation of the AR models, and the function 'auto. arima' in the package 'forecast' by Hyndman (2012), is employed for the estimation of the SARIMA models. To determine the number of lags for GAMs, we calculated the AIC values for all the considered lags, and the minimum AIC value is used when the optimal lag number is selected. Similarly, the minimum AIC values are applied to the selection of the number of lags and hidden nodes for the NN models. The maximum number of lags is set to be five across all models. The maximum numbers of hidden nodes for the NN models are set to be 32, which is considered an appropriate level for NN models with the maximum of five inputs. The estimation for NNs and GAMs is conducted using functions of the R package 'tsDyn'; the functions 'nnetTs' and

<Table 1> RMSE of the Models for DGP 1 (Seasonality)

	Level of noise				
	.01	.1	.5	1	2
AR	**0.010074(1)**	**0.10074(1)**	**0.503699(1)**	**1.007398(2)**	**2.014797(3)**
SARIMA	0.010182(16)	0.101823(16)	0.509054(15)	1.019475(15)	2.036954(15)
NN	0.010092(8)	0.100817(4)	**0.503721(3)**	**1.006184(1)**	2.028162(11)
GAM	0.010091(6)	0.100905(7)	0.504526(7)	1.009052(7)	2.018104(5)
AR-AR	**0.010074(1)**	**0.10074(1)**	**0.503699(1)**	**1.007398(2)**	**2.014797(3)**
AR-SARIMA	0.010181(15)	0.101777(15)	0.508158(11)	1.016325(11)	2.0326(14)
AR-NN	0.01013(12)	0.101273(11)	0.517992(19)	1.025508(17)	2.031552(13)
AR-GAM	0.010091(4)	0.100905(6)	0.504526(8)	1.009052(5)	2.018104(7)
SARIMA-AR	0.010184(17)	0.101835(17)	0.509033(14)	1.019433(14)	2.03711(16)
SARIMA-SARIMA	3.000785(20)	3.009286(20)	3.077646(20)	3.228921(20)	3.712402(20)
SARIMA-NN	0.010241(19)	0.10236(18)	0.516577(18)	1.035142(18)	2.07281(19)
SARIMA-GAM	0.010237(18)	0.102371(19)	0.511716(17)	1.024842(16)	2.047835(17)
NN-AR	**0.010078(3)**	**0.100805(3)**	0.504081(4)	1.008323(4)	**2.009448(1)**
NN-SARIMA	0.010155(13)	0.101572(13)	0.508296(12)	1.015865(10)	2.060087(18)
NN-NN	0.010113(10)	0.101263(10)	0.504134(5)	1.039319(19)	2.020224(9)
NN-GAM	0.010171(14)	0.101613(14)	0.508338(13)	1.018919(12)	**2.0142(2)**
GAM-AR	0.010091(5)	0.100905(8)	0.504526(6)	1.009052(6)	2.018104(6)
GAM-SARIMA	0.010124(11)	0.10154(12)	0.509782(16)	1.019139(13)	2.030546(12)
GAM-NN	0.010094(9)	0.100889(5)	0.504767(10)	1.011186(9)	2.022236(10)
GAM-GAM	0.010091(7)	0.100909(9)	0.504541(9)	1.009081(8)	2.0183(8)

'aar' are used for the estimation of the NNs and GAMs, respectively. The package is provided by Di Narzo et al. (2012).

Prediction performances between 20 models are compared in this study. Tables 1 to 5 report RMSE results for DGPs 1 to 5 with different levels of noise. The number in parentheses indicates the rank of the corresponding model's RMSE in increasing order. The characters in bold style indicate the top three in each rank.

The RMSEs of several forecasting models for the generated data with only seasonality are reported in <Table 1>, where the AR model appears to be superior to other models. A pure NN model also outperformed several competing models when the noises variations were 0.5 and 1. Among the hybrid models, the AR-AR ranked at the top level, and the NN-AR also showed good performance. However, the AR-NN model did not perform very well. These results seem to indicate that the order of model fitting can help to significantly improve the forecasting performances. Meanwhile, the SARIMA model and the SARIMA-based hybrid models ranked bottom. The SARIMA model did not seem to be appropriate to the seasonality pattern only data, since this model is constructed to catch some non-stationary patterns due to the existence of trend. Note again that the DGP 1 is a generated data set with only seasonality without any trend. On the other hand, the GAM and the hybrid model AR-GAM ranked at a slightly higher level than the middle.

<Table 2> RMSE of the Models for DGP 2 (Additive Linear Trend and Seasonality)

	Level of noise				
	0.01	0.03	0.05	1	2
AR	0.015352(19)	0.117614(7)	0.567743(8)	1.06691(6)	2.066311(3)
SARIMA	**0.010175(1)**	**0.103607(2)**	**0.518935(3)**	**1.040322(2)**	**2.068004(6)**
NN	0.012145(7)	0.119904(11)	0.704267(18)	1.901565(18)	2.956209(13)
GAM	0.012841(12)	0.129529(17)	0.581949(10)	1.073619(9)	2.069018(8)
AR-AR	0.013134(16)	0.116179(6)	0.566903(7)	1.06691(6)	**2.066311(3)**
AR-SARIMA	0.010385(4)	0.116086(5)	0.534448(5)	**1.040322(2)**	2.068004(6)
AR-NN	0.013858(17)	0.118847(9)	0.691305(14)	1.938472(19)	3.947892(20)
AR-GAM	0.01421(18)	0.11785(8)	0.569102(9)	1.073618(8)	2.069019(10)
SARIMA-AR	3.256782(20)	3.251754(20)	3.275902(20)	3.440062(20)	2.6971(11)
SARIMA-SARIMA	**0.01021(2)**	**0.103968(3)**	0.521061(4)	1.046467(4)	2.066812(5)
SARIMA-NN	**0.010262(3)**	**0.103086(1)**	0.618533(11)	1.743788(14)	3.578652(18)
SARIMA-GAM	0.010576(5)	0.106683(4)	**0.512207(1)**	**1.034958(1)**	**2.065271(2)**
NN-AR	0.012276(9)	0.120163(12)	0.696138(16)	1.873189(17)	3.490935(16)
NN-SARIMA	0.012939(15)	0.125451(15)	0.698652(17)	1.749196(15)	3.918667(19)
NN-NN	0.01216(8)	0.1232(14)	0.690407(13)	1.598648(12)	3.358867(15)
NN-GAM	0.012336(11)	0.122845(13)	0.692675(15)	1.806272(16)	3.528814(17)
GAM-AR	0.012116(6)	0.133381(19)	0.564461(6)	1.073619(9)	2.069018(8)
GAM-SARIMA	0.012284(10)	0.11957(10)	**0.518795(2)**	1.056138(5)	**2.056273(1)**
GAM-NN	0.012893(14)	0.130186(18)	1.378949(19)	1.662809(13)	3.116038(14)
GAM-GAM	0.012846(13)	0.129389(16)	0.657358(12)	1.312947(11)	2.737856(12)

We could observe that the prediction on AR and AR-AR are same in DGP 1. Note that residuals from AR fits are used as inputs of another AR in AR-AR. Since DGP 1 consists of only seasonality without any trend, if the first AR is appropriately fitted, the residuals may be left as random error term without any dependence structure. In this case, there will be no significant lag term in the second AR fits, and of which prediction results may be same as those of pure AR.

<Table 2> summarizes the results for prediction performance of models for the simulated data using DGP 2, which has the pattern of an additive combination of a linear trend and seasonality. Unlike in the case of DGP 1, the SARIMA model outperformed many competing models, the rankings of which were also very high. Some SARIMA based hybrid models also showed very good performances. In particular, the combination of SARIMA and GAM performed very well under several levels of noises (σ = 0.05, 1, 2). The hybrid model SARIMA-SARIMA ranked within the top three when the variation of noise was relatively small (σ = 0.01, 0.03). The AR-SARIMA model also showed good performance, recording moderately high ranks over all the considered levels of noise variation.

In contrast, the SARIMA-AR ranked bottom. The SARIMA-NN also showed good performance under small variation of noises, but its performance worsened as the variation became larger. These results appear to indicate that the GAM is

<Table 3> RMSE of the Models for DGP 3 (Multiplicative Linear Trend and Seasonality)

	Level of noise				
	1	5	10	20	30
AR	1.8133(19)	6.8922(15)	13.2986(10)	26.192(10)	39.6263(11)
SARIMA	1.1848(5)	**5.8611(3)**	**11.9528(3)**	23.5143(4)	**36.0216(3)**
NN	1.6881(15)	6.6698(9)	16.5101(19)	29.5686(16)	40.7554(17)
GAM	1.5638(12)	6.8661(14)	14.9659(15)	29.6492(17)	40.8046(18)
AR-AR	1.3877(9)	6.9219(16)	13.1571(9)	26.1677(9)	39.5139(10)
AR-SARIMA	1.2594(6)	6.372(6)	12.4228(7)	25.1221(6)	37.1172(7)
AR-NN	1.7681(18)	7.227(20)	13.3043(11)	26.6354(11)	40.1207(14)
AR-GAM	1.933(20)	7.1102(19)	13.4726(12)	26.1373(8)	38.8931(9)
SARIMA-AR	**1.1721(2)**	**5.8372(2)**	11.9577(4)	**23.4687(2)**	**35.9196(2)**
SARIMA-SARIMA	**1.1632(1)**	**5.6509(1)**	**11.6063(1)**	**23.4692(3)**	**35.7107(1)**
SARIMA-NN	1.1845(4)	5.9915(5)	**11.9345(2)**	23.8508(5)	36.6398(6)
SARIMA-GAM	**1.1751(3)**	5.9255(4)	11.9638(5)	**23.4067(1)**	36.5971(5)
NN-AR	1.4545(11)	6.8404(12)	13.8051(13)	29.0211(13)	39.6769(12)
NN-SARIMA	1.3178(7)	6.7086(10)	12.3625(6)	32.4608(20)	37.6408(8)
NN-NN	1.697(16)	6.9902(17)	16.9325(20)	30.8026(18)	40.0346(13)
NN-GAM	1.7148(17)	6.9945(18)	16.066(18)	29.1209(14)	40.4714(15)
GAM-AR	1.3998(10)	6.6128(8)	14.3384(14)	29.5542(15)	40.8046(18)
GAM-SARIMA	1.3478(8)	6.4734(7)	12.9812(8)	25.9979(7)	36.5155(4)
GAM-NN	1.5702(13)	6.826(11)	15.9407(17)	31.6959(19)	40.6085(16)
GAM-GAM	1.6145(14)	6.8642(13)	15.5459(16)	27.4761(12)	41.1674(20)

superior to NNs in the combination of a SARIMA-type hybrid model, in terms of providing stable good performance. On the other hand, the NN-SARIMA ranked below the middle. The prediction performances of the pure models of NN and GAM were not very good. Likewise, the hybrid models based on the combinations of NNs and GAM also did not show good performances.

As seen in <Table 3>, for DGP 3 the SARIMA model tended to perform very well compared with the other models, which is a similar result to the case for DGP 2. The other pure models, AR, GAM and NN, showed relatively poor performances. Regarding the hybrid models, the SARIMA-AR, SARIMA-SARIMA, SARIMA-NN and the SARIMA-GAM showed parallel outstanding performances. When the SARIMA was estimated first, the related hybrid models performed well. However, if the SARIMA was fitted later, the related hybrid models did not show good performances.

<Table 4> shows the forecasting results for the simulation data using DGP 4, which includes an additively related quadratic trend with seasonality. Here, the SARIMA model still performed well except in the case of very small noise ($\sigma = 0.1$). The hybrid models AR-SARIMA, SARIMA-AR, and SARIMA-GAM appeared to provide parallel good performances over the whole range of noise variation. Meanwhile, the SARIMA-NN showed good performance only under small variation of noises, similar to the results for DGP 2.

<Table 4> RMSE of the Models for DGP 4 (Additive Quadratic Trend and Seasonality)

	Level of noise				
	0.1	0.5	1	1.5	2
AR	0.0158(20)	0.1195(10)	0.5728(9)	1.0813(6)	**2.0965(1)**
SARIMA	0.0153(19)	0.1129(5)	0.5478(6)	1.0813(6)	**2.0965(1)**
NN	0.0122(7)	0.1184(8)	0.8085(15)	1.804(18)	3.3336(16)
GAM	0.013(14)	0.1321(18)	0.591(11)	1.0884(9)	2.1065(9)
AR-AR	0.0129(13)	0.1184(9)	0.5721(8)	1.0813(6)	**2.0965(1)**
AR-SARIMA	**0.0108(1)**	0.1156(6)	0.5382(5)	**1.0554(3)**	2.0979(5)
AR-NN	0.014(17)	0.1205(13)	0.9376(18)	1.7295(16)	3.5446(17)
AR-GAM	0.015(18)	0.1195(11)	0.5743(10)	1.0884(11)	2.1065(11)
SARIMA-AR	0.0118(4)	**0.1073(1)**	**0.5182(1)**	**1.0513(2)**	2.1032(7)
SARIMA-SARIMA	**0.0116(2)**	**0.1076(2)**	**0.5296(3)**	1.0652(4)	2.0969(4)
SARIMA-NN	**0.0116(3)**	0.108(3)	0.7403(12)	1.717(15)	4.2334(20)
SARIMA-GAM	0.0119(5)	0.1098(4)	**0.5193(2)**	**1.0454(1)**	2.1036(8)
NN-AR	0.0123(8)	0.1196(12)	0.843(16)	2.0556(20)	3.6932(18)
NN-SARIMA	0.012(6)	0.1209(14)	0.9125(17)	1.5431(13)	3.0392(14)
NN-NN	0.0125(10)	0.1215(15)	0.806(14)	1.743(17)	3.8882(19)
NN-GAM	0.0124(9)	0.1216(16)	0.7953(13)	1.9722(19)	3.0965(15)
GAM-AR	0.0126(11)	0.1345(20)	0.5717(7)	1.0884(9)	2.1065(9)
GAM-SARIMA	0.0127(12)	0.117(7)	0.531(4)	1.0794(5)	2.0992(6)
GAM-NN	0.013(15)	0.1326(19)	0.9391(19)	1.6231(14)	2.7224(13)
GAM-GAM	0.013(16)	0.132(17)	1.2504(20)	1.2056(12)	2.5747(12)

<Table 5> RMSE of the Models for DGP 5 (Multiplicative Quadratic Trend and Seasonality)

	Level of noise				
	1	5	10	20	30
AR	2.8537(19)	17.8441(11)	52.2656(13)	68.1616(16)	82.7149(16)
SARIMA	**1.8103(3)**	15.9457(5)	47.3936(4)	62.6604(4)	78.112(7)
NN	2.1462(15)	25.0637(19)	52.7886(14)	68.2284(17)	80.549(11)
GAM	1.9992(11)	20.7432(16)	55.9308(19)	69.7498(19)	84.3723(18)
AR-AR	2.0548(13)	17.7995(10)	51.7819(12)	67.6431(14)	82.0754(14)
AR-SARIMA	1.9872(10)	17.2787(8)	48.1017(5)	64.1948(8)	79.0541(9)
AR-NN	3.0317(20)	17.5493(9)	50.902(10)	64.5542(9)	76.8408(5)
AR-GAM	2.7768(18)	16.614(7)	50.4985(9)	67.0034(12)	82.0685(13)
SARIMA-AR	1.8114(4)	**15.3686(2)**	**46.4445(3)**	**61.233(2)**	**76.455(3)**
SARIMA-SARIMA	**1.6896(1)**	15.8129(4)	48.2172(6)	63.4718(7)	80.5887(12)
SARIMA-NN	1.8389(6)	**15.5966(3)**	**45.8388(2)**	**62.0935(3)**	**76.2773(1)**
SARIMA-GAM	**1.7848(2)**	**15.2735(1)**	**45.7927(1)**	**60.8941(1)**	76.551(4)
NN-AR	1.914(8)	19.1036(13)	51.1936(11)	67.319(13)	80.3705(10)
NN-SARIMA	2.127(14)	16.4955(6)	49.3004(8)	64.8044(10)	**76.4207(2)**
NN-NN	2.1726(16)	26.7244(20)	54.9529(16)	66.899(11)	82.1915(15)
NN-GAM	2.3636(17)	22.8876(18)	53.4925(15)	63.1403(6)	77.8108(6)
GAM-AR	1.857(7)	19.5359(14)	55.0056(17)	68.3076(18)	83.0325(17)
GAM-SARIMA	1.8223(5)	18.883(12)	48.9484(7)	63.1402(5)	78.9818(8)
GAM-NN	2.0337(12)	21.6489(17)	55.8115(18)	67.7275(15)	84.8106(20)
GAM-GAM	1.9601(9)	20.5745(15)	56.4315(20)	79.0112(20)	84.6583(19)

For DGP 5, the forecasting results in <Table 5> were very similar to those for DGP 3. The SARIMA model performed relatively well. The SARIMA-type hybrid models, that is, the hybrid models in which the SARIMA was estimated first, performed better than did the other hybrid models.

We could not find a model that conclusively outperformed all the other models based on the outputs in Tables 1 to 5. However, we could find some important results, which are summarized as follows.

- For the data with only seasonality without any trend, the AR, AR-AR, and NN-AR showed relatively better performances than did the other competing models. In contrast, the SARIMA and SARIMA related hybrid models showed poor performances.
- If a linear trend is additively employed with seasonality, the SARIMA model tended to outperform many other competitors. The hybrid models SARIMA-SARIMA, SARIMA-GAM, and AR-SARIMA also showed relatively good performances.
- When the linear trend and seasonality are multiplicatively related, the SARIMA, SARIMA-AR, SARIMA-SARIMA, SARIMA-NN, SARIMA-GAM showed parallel good performances. Note that the SARIMA is estimated first in these hybrid models.
- Prediction performances for the case of a quadratic trend were very similar to those for the case of a linear trend.
- In the comparison of GAMs and NNs, both the hybrid models SARIMA-GAM and SARIMA-NN tended to be in parallel, showing very good prediction performances for the multiplicative trend models.
- Regarding data of the additive trend model, the SARIMA-GAM performed well for the full range of noise variation, whereas the SARIMA-NN provided a good performance only when the noise level was trivial.

4. Conclusions

This paper suggests novel hybrid forecasting models using GAMs as an alternative to those using NNs. Versatile hybrid prediction models combining four non-hybrid models (AR, SARIMA, GAM, and NN) are also suggested. The prediction performances of the suggested hybrid models were compared with those of the non-hybrid models via simulation studies. Data are simulated using five different types of seasonality and trend with various levels of noise.

Our study was restricted to only finite cases of simulation, but some important results were nevertheless obtained. According to the simulation analysis results, the prediction performances between models seem to depend on the type of time series data. In particular, for the data set with only seasonality, the AR (a non-hybrid model) and AR-AR (a hybrid model) models equivalently outperformed many other competing models. On the other hand, once the time series data included any trend, the SARIMA and some SARIMA related hybrid models performed well. The SARIMA-GAM seemed to be better than did the SARIMA-NN for the time series data with an additive trend, given that the former showed overall good prediction performances for the full range of noise levels, whereas the latter did so only under small levels of noise.

These results can be applied to the corresponding real data, and some related business implications seem to be available. For example, the AR and AR-AR are expected to show relatively better performances than other models in predicting the data like call arrivals which are with strong seasonality, but without any increasing or decreasing trend. The accurately forecasted call volumes may be helpful in optimizing the schedules of call agents. The SARIMA and some SARIMA related hybrid models are recommended to predict any trend related series like monthly US retail sales at department stores, monthly sales of soft drinks, monthly US series of some consumer good production, and monthly observed US national monetary aggregates. Inventory management of department stores will be improved if retail sales amounts are accurately forecasted. Soft drink companies may be able to make an effective production plan according to the predicted demand. Accurately predicted consumer good production can be utilized in building the consumer price index, which will be used as a measure of inflation. Effective monetary policies via governments may be available using the predicted money demand.

References

Adya, M. and F. Collopy, "How effective are neural networks at forecasting and prediction? A review and evaluation", *Journal of Forecasting*, Vol.17(1998), 481~495.

Barbounis, T. G. and J. B. Teocharis, "Locally recurrent neural networks for wind speed prediction using spatial correlation", *Information Science*, Vol.177(2007), 5775~5797.

Berg, D., "Bankruptcy prediction by generalized additive models", *Applied Stochastic Models in Business and Industry*, Vol.23(2007), 129~143.

Bodyanskiy, Y. and S. Popov, "Neural network approach to forecasting of quasiperiodic financial time series", *European Journal of Operational Research*, Vol.175(2006), 1357~1366.

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

Celik, A. E. and Y. Karatepe, "Evaluating and forecasting banking crises through neural network models : an application for Turkish banking sector", *Expert Systems with Applications*, Vol.33(2007), 809~815.

Di Narzo, A. F., J. L. Aznarte, and M. Stigler, *Nonlinear time series models with regime switching*, R package version (http://cran.us.r-project.org/web/packages/tsDyn/), 2012.

Dominici, F., A. McDermott, S. L. Zeger, and J. M. Samet, "On the use of generalized additive models in time-series studies of air pollution and health", *American Journal of*

Epidemiology, Vol.156(2002), 193~203.

Freitas, P. S. A. and A. J. L. Rodrigues, "Model combination in neural-based forecasting", *European Journal of Operational Research*, Vol. 173(2006), 801~814.

Hansen, J. V., J. B. McDonald, and R. D. Nelson, "Time series prediction with genetic-algorithm designed neural networks : an empirical comparison with modern statistical models", *Computational Intelligence*, Vol.15(1999), 171~184.

Hastie, T. and Tibshirani, R., "Generalized additive models", *Statistical Science*, Vol.1(1986), 297~310.

Hyndman, R. J., Forecasting functions for time series and linear models, R package version (http://cran.r-project.org/web/packages/forecast/), (2012).

Ittig, P. T., "A seasonal index for business", *Decision Sciences*, Vol.28(1997), 335~355.

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

Nelson, C. R. and C. I. Plosser, "Trends and random walks in macroeconomic time series: Some evidence and implications", *Journal of Monetary Economics*, Vol.10(1982), 139~162.

Nelson, M., T. Hill, W. Renus, and M. O'Connor, "Time series forecasting using neural networks : should the data be deseasonalized first?", *Journal of Forecasting*, Vol.18(1999), 359~367.

Prada-Sánchez, J. M. and M. Febrero-Bande, "Parametric, non-parametric and mixed approaches to prediction of sparsely distributed pollution incidents : a case study", *Journal of Chemometrics*, Vol.11(1997), 13~32.

Tseng, F. M., H. C. Yu, and G. H. Tzeng, "Combining neural network model with seasonal time series ARIMA model", *Technological Forecasting and Social Change*, Vol.69(2002), 71~87.

Wood, S. N., *Generalized additive models : An introduction with R*. Florida : Chapman and Hall/CRC, 2006.

Zhang, G. P., "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomputing*, Vol.50(2003), 159~175.

Abstract

트렌드와 계절성을 가진 시계열에 대한 순수 모형과 하이브리드 모형의 비교 연구

정철우* · 김명석**

본 연구에서는 시계열 예측을 위해 선형 모형과 비선형 모형의 하이브리드 모형 및 순수 모형의 성과를 비교 · 평가하였다. 이를 위해 5가지 서로 다른 패턴을 가지는 데이터를 생성하여 시뮬레이션을 진행하였다. 본 연구에서 고려한 선형 모형은 AR(autoregressive model)과 SARIMA(seasonal auto-regressive integrated moving average model)이고, 비선형 모형은 인공신경망(artificial neural networks model)과 GAM(generalized additive model)이다. 특히, GAM은 여러 장점에도 불구하고 시계열 예측을 위한 비선형 모형으로 기존 연구들에서는 거의 쓰이지 않았던 모형이다. 시뮬레이션 결과, seasonality를 가지는 시계열에 대해서는 AR 및 AR-AR 모형이, trend를 가지는 시계열에 대해서는 SARIMA 및 SARIMA와 다른 모형의 하이브리드 모형이 다른 모형에 비해 높은 성과를 보였다. 한편, 인공신경망과 GAM을 비교하면, 트렌드와 계절성이 더해진 시계열에 대해 SARIMA와 GAM의 하이브리드 모형이 거의 모든 노이즈(noise) 수준에 대해 높은 성과를 보인 반면, 노이즈 수준이 미미한 경우에 한해 SARIMA와 인공신경망의 하이브리드 모형이 높은 성과를 보였다.

Keywords : Forecasting, Generalized Additive Models, Seasonal ARIMA, Neural Networks, Hybrid Models

* 한국국방연구원 국방운영연구센터
** 서강대학교 경영전문대학원 글로벌 서비스 경영학과, 교신저자

저 자 소 개

Chulwoo Jeong

Chulwoo Jeong is an Associate Research Fellow at Korea Institute for Defense Analyses. He received B.S., M.S. and Ph.D. in business administration from Sogang University. His research interest is focused on developing methodologies based on management science and statistics and applying them to problems of business and national defense management.

Myung Suk Kim

Myung Suk Kim is an Associate Professor at School of Business, Sogang University, Seoul, Korea. He received his M.S. and Ph.D.in Statistics from the Texas A&M University. He earned his B.A. in Business Administration from Sogang University. His research interests include forecasting, nonparametric model specification, computational statistics, and so on. He has published articles in the Computational Statistics and Data Analysis, Expert Systems with Applications, Environmetrics and many others.