Combining neural network model with seasonal time series ARIMA model

Fang-Mei Tseng[a,*], Hsiao-Cheng Yu[b], Gwo-Hsiung Tzeng[c]

[a]*Department of Finance, Hsuan Chuang University, Hsinchu, Taiwan*
[b]*Institute of Management of Technology, College of Management, National Chiao Tung University, Hsinchu, Taiwan*
[c]*Energy and Environmental Research Group, Institute of Management of Technology and Institute of Traffic and Transportation, College of Management, National Chiao Tung University, Hsinchu, Taiwan*

Abstract

This paper proposes a hybrid forecasting model, which combines the seasonal time series ARIMA (SARIMA) and the neural network back propagation (BP) models, known as SARIMABP. This model was used to forecast two seasonal time series data of total production value for Taiwan machinery industry and the soft drink time series. The forecasting performance was compared among four models, i.e., the SARIMABP and SARIMA models and the two neural network models with differenced and deseasonalized data, respectively. Among these methods, the mean square error (MSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) of the SARIMABP model were the lowest. The SARIMABP model was also able to forecast certain significant turning points of the test time series. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: ARIMA; Back propagation; Machinery industry; Neural network; SARIMA; SARIMABP; Time series

1. Introduction

Dorfman and McIntosh [1] suggested that structural econometrics may not be superior to time series techniques, even when the structural modelers are given the elusive true model.

* Corresponding author. No. 12, Lane 10, Ming Yeo 2nd Street, Hsinchu 300, Taiwan. Tel.: +886-3-5784531; fax: +886-3-5784531.

E-mail address: cflin@hc.ethome.com.tw (F.-M. Tseng).

Wheelwright and Makridakis [2] concluded that time series methods are preferable to explanatory approaches, at least for short-term horizons, for short-term forecasting that applies to control production, post working capital, etc. These papers emphasize short-term forecasting so time series forecasting methods are suitable. Therefore, this paper focuses on time series models. Traditional time series methods were based on the concept of probabilistic statistics, although recently, the concept of neural networks has also been integrated into time series forecasting. Neural networks and traditional time series techniques have been compared in several studies. Sharda and Patil [3] used 75 out of the 111 time series from the well-known M-competition [4] as test cases and found that the neural networks performed as well as the automatic Box–Jenkins procedure. Ansuj et al. [5], Kohzadi et al. [6], Chin and Arthur [7], Hill et al. [8] and Caire et al. [9] all proved neural networks to be a superior method for forecasting in their test cases. Maier and Dandy [10] suggested that the ARIMA model is better suited for short-term forecasts and that neural networks are better suited for longer-term forecasts.

There has been three studies suggesting hybrid models, combining the ARIMA model and neural networks. Voort et al. [11] used this combination to forecast short-term traffic flow. Their technique used a Kohonen self-organizing map as an initial classifier; with each class having an individually tuned ARIMA model associated with it. Wang and Leu [12] used a hybrid model to forecast the mid-term price trend of the Taiwan stock exchange weighted stock index. This was a recurrent neural network trained by features extracted from ARIMA analyses. Their results showed that the neural networks trained by differenced data produced better predictions than otherwise trained by raw data. Su et al. [13] used the hybrid model to forecast a time series of reliability data with growth trend. Their results showed that the hybrid model produced better forecasts than either the ARIMA model or the neural network by itself. So, the combined model has produced promising results in these studies, although no seasonal time series were analyzed in those studies. In this paper, we combine the seasonal time series ARIMA (SARIMA) model and the neural network back propagation (BP) model to forecast time series with seasonality. The results show that the SARIMABP model outperforms the SARIMA model, the BP with deseasonalized data (input the deseasonalized data generated by the moving-to-ratio-average method to the input layer) and the BP with differenced data (input the differenced data generated by the SARIMA model to the input layer).

The remainder of this paper is organized as follows. In Section 2, the SARIMA and the neural network BP models are described. In Section 3, details of the combined model (SARIMABP) are discussed. Section 4 discusses the evaluation methods used for comparing the forecasting techniques. Test data set and the forecasting procedures of the four models are discussed in Section 5. Section 6 compares the results obtained from the SARIMABP model against the SARIMA model, the neural network with deseasonalized data, and the neural network with differenced data. Section 7 provides concluding remarks.

2. Methodology

The SARIMA and the neural network BP models are summarized in the following as foundation to describe the SARIMABP model.

2.1. SARIMA model

A time series $\{Z_t | t = 1, 2, \ldots, k\}$ is generated by SARIMA $(p, d, q)(P, D, Q)_s$ process with mean μ of Box and Jenkins [14] time series model if

$$\varphi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D(Z_t - \mu) = \theta(B)\Theta(B^s)a_t \tag{1}$$

where p, d, q, P, D, Q are integers; s is periodicity; $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p$, $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{P^2}$, $\theta(B) = 1 - \theta_2 B - \theta_2 B^2 - \ldots - \theta_q B^q$, and $\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs}$ are polynomials in B of degree p, q, P, and Q; B is the backward shift operator; d is the number of regular differences, D is the number of seasonal differences; Z_t denotes the observed value at time t, $t = 1, 2, \ldots, k$; and a_t is the estimated residual at time t (Eq. (1)).

The SARIMA model involves the following four-step iterative cycles:

(a) Identification of the SARIMA $(p, d, q)(P, D, Q)_s$ structure;
(b) Estimation of the unknown parameters;
(c) Goodness-of-fit tests on the estimated residuals;
(d) Forecast future outcomes based on the known data.

The a_t should be independently and identically distributed as normal random variables with mean $= 0$ and variance σ^2. The roots of $\varphi(Z) = 0$ and $\theta(Z) = 0$ should all lie outside the unit circle. It was suggested by Box and Jenkins [14] that at least 50 or preferably 100 observations should be used for the SARIMA model.

2.2. Neural network approach

The BP neural network consists of an input layer, an output layer and one or more intervening layers also referred to as hidden layers. The hidden layers can capture the nonlinear relationship between variables. Each layer consists of multiple neurons that are connected to neurons in adjacent layers. Since these networks contain many interacting nonlinear neurons in multiple layers, the networks can capture relatively complex phenomena [8,15].

A neural network can be trained by the historical data of a time series in order to capture the characteristics of this time series. The model parameters (connection weights and node biases) will be adjusted iteratively by a process of minimizing the forecast errors. For each training iteration, an input vector, randomly selected from the training set, was submitted to the input layer of the network being trained [16]. The output of each processing unit (or neuron) was propagated forward through each layer of the network, using the equation

$$\text{NET}_t = \sum_{i=1}^{N} w_{ti} x_i + b_t \tag{2}$$

where NET_t is an output of unit t; w_{ti} is the weight on connection from the ith to the tth unit; x_i is an input data from unit i (input node) to t; b_t denotes a bias on the tth unit; and N is the total

number of input units. A bias or activation threshold of a proper magnitude can affect output activation in the same manner as imposing a limit on the network mapping function.

A sigmoid transformation was then applied to the summation of Eq. (2) for each unit in a hidden layer (see Eq. (3)),

$$y_t = f(\text{NET}_t) = 1/(1 + e^{-\text{NET}_i}) \tag{3}$$

The activity of each output unit was also computed by Eq. (2), using the weights on the connections from the last hidden layer. But, unlike any output from the hidden unit activity, NET_j was not transformed by the sigmoid function. An error $\delta_j^{(L)}$ for the jth output unit was calculated by

$$\delta_j^{(L)} = (T_j - \text{NET}_j) \tag{4}$$

where L denotes the number of the output layer (Eq. (4)). For example, $L = 3, j = 1, 2, \ldots, k, k$ is the number of output unit, and T is the target or the desired activity of the output unit. This error was propagated back to the lower hidden layers as follows:

$$\delta_t^{(l)} = \sum_{i=1}^{N} \delta_i^{l+1} w_{ti}^{(l)} f'(\text{NET}_t^{(l)}) \tag{5}$$

where $w_{ti}^{(l)}$ is the weight from the ith unit in layer l to the tth unit in layer $(l+1)$, $l = 1, 2, \ldots, L - 1$, and $f'(\cdot)$ is the first derivative of the sigmoid function.

In order for the network to learn, the value of each weight had to be adjusted in proportion to each unit's contribution to the total error in Eq. (5). The incremental change in each weight for each learning iteration was computed by (Eqs. (6) and (7))

$$\Delta w_{ti}^{(l)} = c_1 \delta_t^{(i+1)} f(\text{NET}_t^{(l)}) + c_2 m_{ti}^{(l)} \tag{6}$$

where c_1 is a learning constant that controls the rate of learning; c_2, a positive constant that, being less than 1.0, is the momentum term to smooth out the weight changes, and

$$m_{ti}^{(l)} = \Delta w_{ti}^{(l-1)} \tag{7}$$

The procedures for developing the neural network BP model are as follows [13]:

(a) Normalize the learning set;
(b) Decide the architecture and parameters: i.e., learning rate, momentum, and architecture. There are no criteria in deciding the parameters except on a trial-and-error basis;
(c) Initialize all weights randomly;
(d) Training, where the stopping criterion is either the number of iterations reached or when the total sum of squares of error is lower than a pre-determined value;
(e) Choose the network with the minimum error;
(f) Forecast future outcome.

Hansen and Nelson [17] who found the portfolio of forecasts of the forecasting methods were more accurate than the sum of the individual parts in the case of the Utah revenue. Maier and Dandy [10] suggested that the SARIMA model is suited for short-term forecasts,

although the neural network model is better suited for long-term forecasts. The current authors propose to combine these two models into a hybrid model to forecast seasonal time series data.

3. The combined SARIMA with neural network model

The BP neural network model, containing many interacting nonlinear neurons in multiple layers, can capture relatively complex phenomena, adjusting the parameters iteratively to minimize the forecast errors. However, when sample size is small, Liu et al. [18] showed that the neural network model performs better than probability model for the clustering problem. The BP model demonstrated good forecasting performance in previous studies, although it ignores forecasting errors when the model is built. Their forecasting errors provide feedback to revise the weights, but they do not provide feedback to modify the input variables. Thus, this piece of information lost by the BP model can be very important [13]. Very often, quarterly or monthly seasonal cycles exist in real-world time series. The SARIMA model has been shown to make good forecasts for seasonal time series especially for short-term periods, but it is limitation by the large amount of historical data (at least 50 and preferably 100 or more) that is required. However, in modern society, due to factors of uncertainty from the integral environment and rapid development of new technology, we usually have to forecast future situations in a short span of time using limited amounts of data. The data insufficiency will sometimes limit its application when used with the ARIMA model. The SARIMABP model combines the advantages of the SARIMA and the BP models that input the forecasts and residuals generated by a SARIMA model to the input layer of a BP model (i.e., the SARIMABP is a nonlinear model). The latter attempts, after the training process, to minimize the residuals.

The SARIMABP method benefits from the forecasting capability of the SARIMA model and the capability of the neural network model to further reduce the residuals. Thus, the training set will have lower errors. We used Hill et al.'s [8] concept of deseasonalizing a seasonal time series (input the deseasonalized data generated by the moving-to-ratio-average method to the input layer) to formulate a neural network model. We also used Wang and Leu's [12] approach of taking the differences of a seasonal time series data to build another neural network model (input the differenced data generated by the SARIMA model to the input layer). However, both of the two neural network models have difficulties in deciding how many input variables to use. We chose to use the estimated values $\hat{Z}_t$ and residuals $\hat{a}_t$ of the SARIMA model as input variables to eliminate this problem. There exists a problem when a neural network model is used to forecast future outcomes, because future residuals are not yet known. This issue was not addressed by Su et al. [13]. The authors were proposed to use the weighted average of past residuals from the same period in the past as a first guess of the residuals in the forecast period. For example, the first guess of the residual of January 1996 could be a weighted sum of the residuals of January 1995, January 1994, etc. The closer is the residual to present, the larger is its weight.

The following pieces of the information are input variables to the neural network model: (a) the forecasts generated by SARIMA model; and (b) the residuals of the SARIMA model.

4. Forecast evaluation methods

Yokum and Armstrong [19] conducted an expert opinion survey about evaluation criteria to select forecasting techniques. Clearly, accuracy was the most important criterion, with the other one being the cost savings generated from improved decisions. In addition, execution issues such as ease of interpretation and ease of use were also highly rated. In this study, four criteria of forecasting accuracy were used to make comparisons of the forecasting capabilities among the SARIMA model, the BP model with differenced data, the BP with deseasonalized data, and the SARIMABP model.

The first measurement is the mean square error (MSE; Eq. (8)),

$$MSE = \sum_{t=1}^{T}(P_t - Z_t)^2/T \tag{8}$$

where P_t is the predicted value at time t; Z_t is the actual value at time t; and T is the number of predictions.

The second criterion is the mean absolute error (MAE; Eq. (9)),

$$MAE = \sum_{t=1}^{T}|P_t - Z_t|/T \tag{9}$$

The third criterion is the mean absolute percentage error (MAPE; Eq. (10))

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{P_t - Z_t}{Z_t} \tag{10}$$

While the above criteria are good measures of the deviations of the predicted values from the actual values, they can not reflect a model's ability to predict turning points. For example, in monetary markets, it is critical for traders and analysts to forecast the change of market direction. Therefore, the turning points are as important as the forecast value itself. A correct turning point forecast requires that (Eq. (11)):

$$sign(P_t - P_{t-1}) = sign(Z_t - Z_{t-1}) \tag{11}$$

The ability of a model to forecast turning points can be measured by a fourth evaluation method developed by Cumby and Modest [20]. This model defines a forecast direction variable F_t and an actual direction variable A_t such that

$$A_t = 1 \text{ if } \Delta Z_t > 0 \text{ and } A_t = 0 \text{ if } \Delta Z_t \le 0 \tag{12}$$

$$F_t = 1 \text{ if } \Delta P_t > 0 \text{ and } F_t = 0 \text{ if } \Delta P_t \le 0 \tag{13}$$

where ΔZ_t is the amount of change in actual variables between time $t-1$ and t; and ΔP_t is the amount of change in forecasting variables between time $t-1$ and t (Eqs. (12) and (13)). Cumby and Modest [20] suggest the following regression equation:

$$F_t = \alpha_0 + \alpha_1 A_1 + \varepsilon_t \tag{14}$$

where ε_t is error term; and α_1 is the slope of this linear equation. Here, α_1 should be positive and significantly different from 0 in order to demonstrate those F_t and A_t have a linear

relationship. This reflects the ability of a forecasting model to capture the turning points of a time series.

5. Experimental results

In the following, the performance of the SARIMABP model is compared with other models using two seasonal time series: one is the total production value of the Taiwan machinery industry, and the other is the sales volume of soft drinks quoted from Montgomery et al. [21]. In Section 5.1, the characteristics of the total production value of the Taiwan machinery industry are described. Sections 5.1.1, 5.1.2, and 5.1.3 report the results obtained from the SARIMA, the three neural network, and the SARIMABP models. Section 5.2 describes the soft drink time series. Sections 5.2.1 and 5.2.2 report the forecasting results of the soft drink time series.

5.1. The Taiwan machinery industry time series

The machinery industry in Taiwan has made steady progress over the past decade, playing a critical supporting role as the foundation to the overall manufacturing industry in Taiwan. Nevertheless, it is a major exporting industry by itself. The time series data of the total production revenues of Taiwan machinery industry for the period from January 1991 to December 1996 showed strong seasonality and growth trends, as shown in Fig. 1 and Table 1. The sharp drop each year that generally occurs in January or February is due to plants closing for the Chinese New Year holiday. The data set was used in two different ways in an attempt to experiment with the amount of historical data required for generating better forecasts. The first experiment used a 5-year data set from January 1991 to December 1995 as the training data, and a forecast for the 12 months of 1996. The
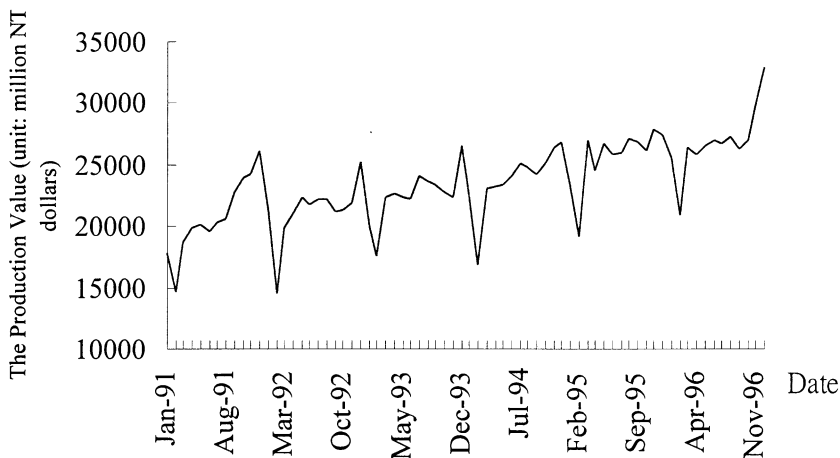

Fig. 1. The production value of Taiwan's machinery industry.

Table 1
The machinery production time series

Date	Actual value	Date	Actual value	Date	Actual value	Date	Actual value	Date	Actual value	Date	Actual value
Jan 91	17798.5	Jan 92	21372	Jan 93	17598	Jan 94	22466	Jan 95	23370	Jan 96	25493
Feb 91	14671.5	Feb 92	14555	Feb 93	20006	Feb 94	16794	Feb 95	19229	Feb 96	20969
Mar 91	18688.5	Mar 92	19824	Mar 93	22380	Mar 94	23087	Mar 95	26947	Mar 96	26488
Apr 91	19832.5	Apr 92	21115	Apr 93	22589	Apr 94	23249	Apr 95	24520	Apr 96	25783
May 91	20182.5	May 92	22414	May 93	22301	May 94	23373	May 95	26644	May 96	26537
Jun 91	19626.5	Jun 92	21805	Jun 93	22204	Jun 94	24110	Jun 95	25822	Jun 96	26938
Jul 91	20366.5	Jul 92	22172	Jul 93	24110	Jul 94	25138	Jul 95	25930	Jul 96	26687
Aug 91	20630.5	Aug 92	22217	Aug 93	23702	Aug 94	24885	Aug 95	27096	Aug 96	27273
Sep 91	22719.5	Sep 92	21255	Sep 93	23404	Sep 94	24185	Sep 95	26810	Sep 96	26339
Oct 91	23898.5	Oct 92	21333	Oct 93	22821	Oct 94	25083	Oct 95	26103	Oct 96	27073
Nov 91	24206.5	Nov 92	21884	Nov 93	22390	Nov 94	26494	Nov 95	27821	Nov 96	29705
Dec 91	26075.5	Dec 92	25191	Dec 93	26502	Dec 94	26821	Dec 95	27459	Dec 96	33021

second experiment used only the 3-year data set from January 1993 to December 1995 as the training data, and forecast for the same time horizon. It was assumed that the seasonality and the trend existed in the historical data extended to the future with the same pattern.

5.1.1. Building SARIMA models

The test time series data was processed by taking the first-order regular difference and the first seasonal difference in order to remove the growth trend and the seasonality characteristics. The authors used SCA and SAS statistical packages to formulate the SARIMA model. Akaike Information Criterion (AIC) was used to determine the best model. The model generated from the 5-year data set is ARIMA$(0,1,1)(1,1,1)_{12}$. The equation used is presented in Eq. (15):

$$(1 + 0.309B^{12})(1 - B)(1 - B^{12})Z_t = (1 - 0.9907B)(1 - 0.7159B^{12})a_t \qquad (15)$$

The model generated from the 3-year data set is ARIMA$(0,1,1)(0,1,0)_{12}$. The equation used is presented in Eq. (16):

$$(1 - B)(1 - B^{12})Z_t = (1 - 0.88126B)a_t. \qquad (16)$$

5.1.2. Building neural network models

Hill et al. [8] chose to deseasonalize a seasonal time series data before formulating his neural network model. Wang and Leu [12] suggested that neural networks trained by differenced data (from ARIMA difference) can produce better predictions than otherwise trained by raw data. The current authors used the Neural Networks Application Development Software, Qnet97, to build neural network models using the deseasonalized, the differenced, and the original seasonal time series.

The implementation of the BP model involved three stages [22]:

Stage 1. Choose the best architecture for a neural network model. One hidden layer was set up. The input and the hidden neurons of the test time series were examined. The learning rate was initially set at 0.2, and the momentum was set at 0.8. The output layer had one neuron, which was the forecast value. The authors used RMS (root of MSE) and the product–moment correlation's coefficient as references to select the architecture.

Stage 2. Readjust the learning rate and momentum in the training stage. The transformation function used was the sigmoid function, and the learning rule used was the generalized Delta rule. There were two steps in this stage. First, adjusting the learning rates, which include 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8. Second, adjusting the momentum values, which include 0.95, 0.9, 0.85, 0.8, 0.7, 0.6, and 0.5. RMS and correlation coefficients were the evaluation criteria.

Stage 3. The authors tested the goodness-of-fit of this forecast model for the period from January to November 1996. The data for December 1996 was exceptionally high. The authors skipped this data point to avoid possible distortion caused by an outlier. The three forecast neural network models are shown in Table 2.

The BP model with deseasonalized time series (adjust the seasonal factor by using the ratio-to-moving-average method) generated better forecasts than the one with raw data. The BP with differenced data did not, however, generate a better forecast than the one with raw data.

5.1.3. Building the SARIMABP model

The SARIMABP hybrid model integrated the SARIMA model with the neural network model and tested with the raw seasonal data. There were four neurons in the input layer of the

Table 2
The best results of neural networks

	Hill et al. [8] (deseasonalized data, w_t)	Wang and Leu [12] (differenced data, y_t)	Raw data (Z_t)
5-year data set			
Input nodes	w_{t-1}, w_{t-2}	y_{t-1}, y_{t-2}	Z_{t-1}, Z_{t-12}
The number of hidden neurons	2	4	3
Learning rate	0.5	0.3	0.3
Momentum	0.9	0.8	0.8
RMS of training data	0.086	0.083	0.088
Correlation of training data	0.808	0.533	0.750
3-year data set			
Input nodes	w_{t-1}, w_{t-2}	y_{t-1}, y_{t-2}	Z_{t-1}, Z_{t-12}
The number of hidden neurons	2	2	4
Learning rate	0.5	0.3	0.2
Momentum	0.9	0.8	0.7
RMS of training data	0.061	0.117	0.064
Correlation of training data	0.889	0.538	0.812

SARIMABP model: Z_{t-1}, Z_{t-12}, $\hat{Z}_t$, and residual value a_t (the results from the SARIMA model). Su et al. [13] used only the forecasts $\hat{Z}_t$ by an ARIMA model as input nodes, although their forecast results were not good. We built four models with the following input layers:

(a) Z_{t-1}, Z_{t-12}, $\hat{Z}_t$, and residual value a_t (the results from the SARIMA model);
(b) Z_{t-1}, Z_{t-12}, and $\hat{Z}_t$ (the forecasts of SARIMA model);
(c) $\hat{Z}_t$ and residual value a_t (the results from the SARIMA model);
(d) $\hat{Z}_t$ (the forecasts of SARIMA model).

The authors experimented with a number of different neurons in the hidden layer and a number of different settings of learning rate and momentum. The best results are reported in Table 3. Computing the weighted average of the future estimates, the weight of 1995 is higher by 0.1 than the other year (e.g., estimate residual value a_t of February 1993, February 1994, and February 1995 are higher by 0.3, 0.3, and 0.4, respectively. Table 3 indicates that the RMS of training data has the largest number of input neurons. Model II (input nodes $\hat{Z}_t$, residual value a_t) has lowest out-of-sample errors. For the 3-year data set, Model I (input nodes include Z_{t-1}, Z_{t-12}, $\hat{Z}_t$, and residual value a_t) has lower RMS, MAE, and MAPE than the other methods.

Table 3
The best results of SARIMABP model

	Model I	Model II	Model III	Model IV
Input nodes	Z_{t-1}, Z_{t-12}, $\hat{Z}_t$, $\hat{a}_t$	Z_{t-1}, Z_{t-12}, $\hat{Z}_t$	$\hat{Z}_t$, $\hat{a}_t$	$\hat{Z}_t$
5-year data set				
The number of hidden neurons	7	3	3	2
Learning rate	0.3	0.2	0.2	0.3
Momentum	0.95	0.95	0.85	0.85
RMS of training data	0.002	0.034	0.0035	0.040
Correlation of training data	0.9998	0.9924	0.9995	0.926
Out-of-sample errors				
MSE	539061.8	623463.2	535639.7	653491.1
MAE	595.83	626.01	591.89	612.5
MAPE	2.26	2.36	2.25	2.29
3-year data set				
The number of hidden neurons	7	2	3	2
Learning rate	0.5	0.5	0.4	0.2
Momentum	0.9	0.95	0.85	0.75
RMS of training data	0.0023	0.059	0.0028	0.075
Correlation of training data	0.9998	0.889	0.9998	0.815
Out-of-sample errors				
MSE	1005412	755190.60	536793.20	1693124
MAE	535.73	644.94	546.75	872.81
MAPE	2.08	2.40	2.13	3.29

The first model is more complex than the third model. Razor principle is guided by the fact that for a given performance level, the smaller network will be the superior generalizer. The fewer weights in the network, the greater the confidence that over-training has not resulted in noise being fitted [23]. Considering the training time and the complexity of Models I and III, the authors suggest Model III to be a preferred model.

5.2. The soft drinks time series

In order to demonstrate the performance of the SARIMABP model, the authors applied the models to another time series, which was the monthly sales volume of soft drinks from the Montgomery et al.'s [21] book "Forecasting and Time Series Analysis," p. 364. This time series demonstrates growth trend and seasonality, as is shown in Fig. 2.

5.2.1. Building SARIMA model

The time series data was pre-processed using the logarithmic transformation, first-order regular differencing, and first-order seasonal differencing in order to stabilize the variance and remove the growth trend and seasonality. The authors used the SAS statistical package to formulate the SARIMA model. Akaike Information Criterion [11,12] was used to determine the best model. The derived model is ARIMA$(1,1,0)(0,1,0)_{12}$, and the equation used is presented in Eq. (17):

$$(1 + 0.73B)(1 - B)(1 - B^{12})Z_t = a_t \tag{17}$$

5.2.2. Building neural network models and the SARIMABP model

The authors built the SARIMABP model, the BP using the deseasonalized time series (adjust the seasonal factor by using the ratio-to-moving-average method), and the BP with
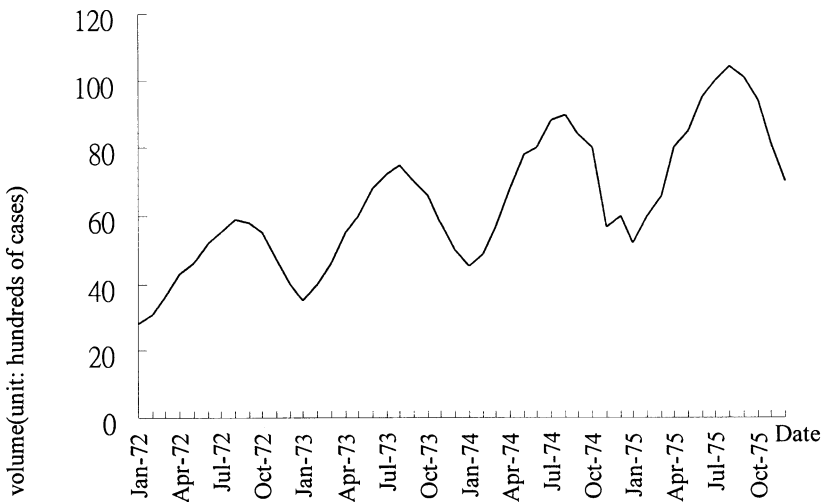

Fig. 2. Monthly sales volume of soft drinks.

Table 4
The best results of neural networks

	Hill et al. [8] (deseasonalized data, w_t)	Wang and Leu [12] (differenced data, y_t)	SARIMABP
Input nodes	w_{t-1}, w_{t-2}	y_{t-1}, y_{t-2}	$\hat{Z}_t, \hat{a}_t$
The number of hidden neurons	5	4	3
Learning rate	0.4	0.5	0.3
Momentum	0.95	0.95	0.85
RMS of training data	0.029999	0.084685	0.007579
Correlation of training data	0.9788	0.717738	0.9998

differenced time series. The differenced data were acquired from Section 5.2.1. The original time series was pre-processed using the logarithmic transformation, first-order regular differencing, and first-order seasonal differencing in order to stabilize the variance and remove the trend and seasonality.

The implementation of the neural network BP model involved three stages [22], which are the same as in Section 5.1.2. The SARIMABP model used the SARIMA model as input nodes. The authors experimented with a number of different neurons in the hidden layer, a number of different settings of learning rate and momentum. The best results are reported in Table 4.

The SARIMABP model outperforms the BP using the deseasonalized time series and the BP with differenced data in terms of predictions.

6. Evaluation and comparison

In this section, the performances of the previous models built for forecasting the outputs of Taiwan's machinery industry as well as the sales volume of soft drinks are reported. The measurements evaluated include MAPE, MAE, MSE, and turning point evaluation method.

The machinery production value of December 1996 is significantly bigger than the past value. In order to effectively evaluate the performance of the models, we deleted December 1996 when calculating MSE, MAE, and MAPE of the out-of-sample data set.

6.1. Quantitative evaluation

In additional to MAPE, MAE, and MSE measurements, the t value was also used to test the hypothesis that SARIMABP and SARIMA models, as well as the BP with deseasonalized data and the BP with differenced data, have the same means of absolute forecast errors. If this hypothesis were statistically significant, we would have demonstrated a better model. The results are shown in Tables 5, 6, and 7.

For in-sample error comparison of the machinery production time series, Table 5 indicates that SARIMABP model outperformed the SARIMA model, the BP with deseasonalized data, and with differenced data for both 5- and 3-year data sets. For the 5-year data set, the MSE of the SARIMA model, BP with deseasonalized data, and BP with differenced data were 1,022,546.00, 1,086,283.00, and 2,431,096.00, respectively. Whereas, the MSE of SAR-IMABP model was considerably lower at 6728.37. The MAPE of the SARIMABP model

Table 5
In-sample errors (training set) of the machinery production time series

| Method | SARIMA | Neural network | | Combined model (SARIMABP) |
		Deseasonalized data	Differenced data	
5-year data set				
MSE	1022546.00	1086283.00	2431096.00	6728.37
MAE	750.38	835.23	1236.50	62.13
MAPE	3.18	3.76	5.37	0.27
T value	6.49*	9.99*	5.94*	–
3-year data set				
MSE	2735484.00	488765.30	1548039.00	1722.40
MAE	1213.73	542.05	1058.17	31.89
MAPE	4.70	2.23	4.13	0.13
T value	4.87*	5.14*	6.95*	–

Null hypothesis of the existence of the same means of the forecasted absolute errors generated by the combined model and other models.
* To be rejected at 10% significance level.

was only 0.27, which is a full percentage point better than the other models. The MAE of the SARIMABP model is also an order of magnitude better than those of the other models. For the 3-year data set, the SARIMABP model still has the lowest MAE, MSE, and MAPE. *t* tests also indicated a rejection of the hypothesis that the MAEs of the SARIMABP model are the same as those of the other models.

Table 6
Out-of-sample errors (Jan–Nov 1996) of the machinery production time series

| Method | SARIMA | Neural network | | Combined model (SARIMABP) |
		Deseasonalized data	Differenced data	
60 historical data				
MSE	1197387.00	3046768.00	1028880.00	535639.00
MAE	816.72	1477.67	899.28	591.89
MAPE	2.88	5.72	3.45	2.25
T value	0.66	3.26**	2.10*	–
36 historical data				
MSE	5303143.00	1742084.00	1497008.00	536793.20
MAE	1874.25	968.23	1109.19	546.75
MAPE	6.86	3.72	4.21	2.13
T value	3.54**	2.06*	2.14*	–

Null hypothesis of the existence of the same means of the forecasted absolute errors generated by the combined model and other models.
* To be rejected at 10% significance level.
** To be rejected at 5% significance level.

Table 7
Performance measurements of the forecast of the soft drink time series

Method	SARIMA	Neural network		Combined model (SARIMABP)
		Deseasonalized data	Differenced data	
In-sample errors				
MSE	37.69	4.28	9.79	0.50
MAE	3.67	1.34	1.76	0.35
MAPE	5.43	2.15	2.85	0.75
T value	− 3.36**	− 2.98**	− 2.07*	–
Out-of-sample errors				
MSE	172.96	114.23	108.84	14.49
MAE	12.20	8.81	9.12	2.32
MAPE	14.33	10.48	12.54	5.74
T value	− 3.92**	− 1.98*	− 3.65**	–

Null hypothesis of the existence of the same means of the forecasted absolute errors generated by the combined model and other models.
* To be rejected at 10% significance level.
** To be rejected at 5% significance level.

For out-of-sample error comparisons of the machinery production time series, Table 6 indicates that the SARIMABP model outperformed the SARIMA model, the BP with deseasonalized data, and the BP with differenced data for both the 5- and 3-year data sets. The MSE of the SARIMA model, BP with deseasonalized data, and BP with differenced data were 1,197,387, 3,046,768, and 1,028,880, respectively, Whereas, the MSE for the SAR-IMABP model was clearly lowest at 535,639. The MAPE of the SARIMABP model was only 2.25, which is a full percentage point better than the other three models. The MAE of the SARIMABP model was also better than those of the other models. For the 3-year data set, the SARIMABP model had the lowest MAE, MSE, and MAPE. *t* tests also showed rejection of the hypothesis that the MAE of the SARIMABP model is the same as those of the other models except the SARIMA model.

Tables 5 and 6 also indicate that when the number of historical data is reduced from 60 historical data (5 years) to 36 historical data (3 years), the performance of the SARIMABP model is still stable. However, the performance of the SARIMA model drops off significantly. Interestingly enough, the goodness-of-forecast measurements for the SARIMABP model and the BP with deseasonalized data should be better for the 3-year than the 5-year data case. One explanation for this phenomenon may be that the Taiwan machinery production revenues in 1996 resemble more those in 1995, 1994, and 1993 than with those in 1992 and 1991.

The in-sample errors in Table 7 indicate that the SARIMABP model outperformed the SARIMA model, the BP with deseasonalized data, and the BP with differenced data for the soft drink time series. The MAE, MSE, and MAPE of the SARIMABP model are the lowest among the four test models. In out-of-sample comparisons of the soft drink time series, the SARIMABP model outperformed the SARIMA model, the BP with deseasonalized data, and the BP with differenced data.

For the two time series, the neural network models are better than the SARIMA model for a small sample size. This was especially true for the SARIMA model for the machinery production time series when the sample size was reduced from 60 to 36, and the MAE, MSE, and MAPE were increased. However, the neural network models did not display this phenomenon. The reason for this is that the a_t of the SARIMA model should be independently and identically distributed, so the sample size of the SARIMA model is at least 50. However, the parameters of the neural network models are adjusted iteratively by a process of minimizing the forecast errors, for which the sample sizes are less than required by the ARIMA model.

6.2. Turning point evaluations

The turning point evaluation method using regression Eq. (14) is shown in Table 8 for both the machinery production and the soft drink time series.

The t ratio of the slope coefficient α_1 of the SARIMABP model shows that it is statistically different from zero for both the machinery production time series and the soft drink time series. This implies that the SARIMABP model has good turning point forecasting ability. On the other hand, for the SARIMA model, the α_1 parameter is not significantly different from zero for both the machinery production and the soft drink time series. For the machinery production time series, when the number of historical data is reduced from 60 to 36, α_1 even turned out to be negative. This reflects the fact that SARIMA could not forecast turning points well, especially when the number of observations is small.

Table 8
Results of turning point forecasting capability of the four models

Method	SARIMA	Neural network		Combined model (SARIMABP)
		Deseasonalized data	Differenced data	
5-year data set of the machinery production time series, 1996.1–1996.12				
α (*t*-ratio)	0.40 (1.97)	0.33 (1.75)	0.43 (2.26*)	0.20 (1.10)
α (*t*-ratio)	0.46 (1.72)	0.50 (1.86*)	0.37 (1.27)	66 (2.76**)
3-year data set of the machinery production time series, 1996.1–1996.12				
α (*t*-ratio)	0.63 (3.30**)	43 (2.26*)	0.33 (1.75)	0.33 (1.75)
α (*t*-ratio)	− 0.13 (− 0.38)	37 (1.26)	0.50 (1.87*)	0.50 (1.87*)
Soft drink time series				
α (*t*-ratio)	0.80 (6.33**)	0.20 (0.89)	0.20 (0.95)	0 (–)
α (*t*-ratio)	0.20 (1.21)	0.37 (1.27)	0.51 (1.86*)	1 (**)

Null hypothesis of the α_0 or α_1 existence of the is equal to zero. The t ratio of the slope coefficient α_1 is statistically different from zero. This implies that for the 12-month out-of-sample period, the model had turning point forecasting power.
 * To be rejected at 10% significance level.
 ** To be rejected at 5% significance level.

7. Conclusions

The authors proposed to use a hybrid model (SARIMABP) that combines the time series SARIMA model and the neural network BP model to predict seasonal time series data. The results showed that SARIMABP is superior to the SARIMA model, the BP with deseasonalized data, and the BP with differenced data for the test cases of machinery production time series and soft drink time series. The MSE, MAE, and MAPE were all the lowest for the SARIMABP model. The SARIMABP also outperformed other models in terms of overall proposed criteria, including MSE, MAE, MAPE, and turning points forecasts. For the machinery production time series, the SARIMABP model remained stable even when the number of historical data was reduced from 60 to 36. On the other hand, the accuracy of the SARIMA model reduced when the number of historical data points decreased.

References

[1] J.H. Dorfman, C.S. McIntosh, Results of a price forecasting competition, Am. J. Agric. Econ. 72 (1990) 804–808.

[2] S.C. Wheelwright, S. Makridakis, Forecasting Methods for Management, Wiley, New York, 1985.

[3] R. Sharda, R. Patil, Neural networks as forecasting exports: an empirical test, Proc. IJCNN Meet., 1990 2 (1990) 491–494.

[4] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, R. Winkler, The accuracy of extrapolation (time series) methods: results of a forecasting competition, J. Forecast. 1 (1982) 111–153.

[5] A.P. Ansuj, M.E. Camargo, R. Radharamanan, D.G. Petry, Sales forecasting using time series and neural networks, Comput. Ind. Eng. 31 (1) (1996) 421–424.

[6] N. Kohzadi, M.S. Boyd, B. Kermanshahi, I. Kaastra, A comparison of artificial neural network and time series model for forecasting commodity prices, Neurocomputing 10 (2) (1996) 169–181.

[7] K. Chin, R. Arthur, Neural network vs. conventional methods of forecasting, J. Bus. Forecast. 14 (4) (1996) 17–22.

[8] T. Hill, M. O'Connor, W. Remus, Neural network models for time series forecasts, Manage. Sci. 42 (7) (1996) 1082–1092.

[9] D. Caire, G. Hatabian, C. Muller, Progress in forecasting by neural networks, Int. Conf. Neural Networks II (1992) 540–545.

[10] H.R. Maier, G.C. Dandy, Neural network models for forecasting univariate time series, Neural Networks World 6 (5) (1996) 747–772.

[11] V.D. Voort, M. Dougherty, M. Watson, Combining Kohonen maps with ARIMA time series models to forecast traffic flow, Transp. Res. Circ. (Emerge Technol.) 4C (5) (1996) 307–318.

[12] J.H. Wang, J.Y. Leu, Stock market trend prediction using ARIMA-based neural networks, IEEE Int. Conf. Neural Networks, 1996 4 (6) (1996) 2160–2165.

[13] C.T. Su, L.I. Tong, C.M. Leou, Combination of time series and neural network for reliability forecasting modeling, J. Chin. Inst. Ind. Eng. 14 (4) (1997) 419–429.

[14] G.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, New Jersey, 1976.

[15] W.C. Chiang, T.L. Urban, G.W. Baldridge, A neural network approach to mutual fund net asset value forecasting, Omega, Int. J. Manage. Sci. 24 (2) (1996) 205–215.

[16] M.C. Liu, W. Kuo, T. Sastri, An exploratory study of a neural network approach for reliability data analysis, Qual. Reliab. Eng. Int. 11 (1995) 107–112.

[17] J.V. Hansen, R.D. Nelson, Neural networks and traditional time series methods: a synergistic combination in state economic forecasts, IEEE Trans. Neural Networks 8 (4) (1997) 863–873.

[18] M.C. Liu, W. Kuo, T. Sastri, An exploratory study of a neural network approach for reliability data analysis, Qual. Reliab. Eng. Int. 11 (1995) 107–112.

[19] J.T. Yokum, J.S. Armstrong, Beyond accuracy: comparison of criteria used to select forecasting methods, Int. J. Forecast. 11 (4) (1995) 591–597.

[20] R.E. Cumby, D.M. Modest, Testing for market timing ability: a framework for forecast evaluation, J. Financ. Econ. 19 (2) (1987) 169–189.

[21] D.C. Montgomery, L.A. Johnson, J.S. Gardiner, Forecast. Time Ser. Anal. 364, McGraw-Hill, New York, 1990.

[22] C.S. Huang, D.H. Shih, C.L. Liu, Neural network prediction of risk premiums on Taiwan life insurer stocks, J. Inf. Manage. 3 (1) (1996) 63–80.

[23] E.M. Azoff, Neural Network Time Series Forecasting of Financial Markets, Wiley, New York, 1994.