Forecasting energy consumption in Taiwan using hybrid nonlinear models

H.T. Pao*

Department of Management Science, National Chiao Tung University, 1001 Ta Hsueh Road, 30010 Hsinchu, Taiwan, ROC

ABSTRACT

The total consumption of electricity and petroleum energies accounts for almost 90% of the total energy consumption in Taiwan, so it is critical to model and forecast them accurately. For univariate modeling, this paper proposes two new hybrid nonlinear models that combine a linear model with an artificial neural network (ANN) to develop adjusted forecasts, taking into account heteroscedasticity in the model's input. Both of the hybrid models can decrease round-off and prediction errors for multi-step-ahead forecasting. The results suggest that the new hybrid model generally produces forecasts which, on the basis of out-of-sample forecast encompassing tests and comparisons of three different statistic measures, routinely dominate the forecasts from conventional linear models. The superiority of the hybrid ANNs is due to their flexibility to account for potentially complex nonlinear relationships that are not easily captured by linear models. Furthermore, all of the linear and nonlinear models have highly accurate forecasts, since the mean absolute percentage forecast error (MAPE) results are less than 5%. Overall, the inclusion of heteroscedastic variations in the input layer of the hybrid univariate model could help improve the modeling accuracy for multi-step-ahead forecasting.

© 2009 Published by Elsevier Ltd.

1. Introduction

Worldwide energy consumption is rising sharply, owing to increasing human population, continuing pressures for better living standards and emphasis on large-scale industrialization in developing countries, thus sustaining positive economic growth rates. Taiwan's energy consumption increased sharply from 49.67 million kiloliters of oil equivalent (KLOE) in 1990 to 113.85 million KLOE in 2007. The annual growth rate was 5.00% during this period. Among the various forms of energy consumed in 2007, electricity accounted for 51.18%, petroleum 38.35%, and the others 10.47% (Bureau of Energy, Ministry of Economic Affairs in Taiwan). Total electricity consumption rose sharply from 82.65 billion (kwh) in 1990 to 229.20 billion (kwh) in 2007 with an annual growth rate of 6.18%. Petroleum consumption increased from 22.97 KLOE in 1990 to 43.66 KLOE in 2007 with an annual growth rate of 3.85%. Given this fact, the accuracy of energy demand forecasting is important not only for energy utilities themselves but also for consumers.

A sound forecasting technique is essential for accurate investment planning in energy production/generation and distribution. Multivariate modeling along with co-integrated techniques or regression analysis has been used in a number of studies to analyze and forecast energy consumption [1–6]. One limitation of multivariate models is that they depend on the availability and reliability of data on independent variables over the forecasting period, which requires further efforts in data collection and estimation. On the other hand, univariate time series analysis provides another modeling approach, which only requires the historical data of the variable of interest to forecast its future evolution behavior. The univariate Box–Jenkins autoregressive integrated moving average (ARIMA) [7] analysis has been widely used for modeling and forecasting many medical, environmental, financial, and engineering applications [8–11]. In addition, Zhou et al. [12] presented a univariate trigonometric grey prediction approach for forecasting electricity demand in China.

Recently, artificial neural network (ANN) techniques have also gained popularity in energy demand and load forecasting. For short-term forecasting, Gonzalez and Zamarreno [13] proposed specifications for a self-exciting neural network (NN) model to forecast energy consumption in buildings. Lauret [14] proposed the use of Bayesian regularization as a technique to estimate the parameters of a NN in order to forecast load. Since the Bayesian methods provide an explicit handling of uncertainty in the modeling, Lauret [14] concluded that the Bayesian NN approach to modeling offers significant advantages over classical NN learning methods for short-term load forecasting. Hipper et al. [15] made a literature review and evaluation in forecasting load using NNs. Amjady and Keynia [16] proposed a hybrid method composed of wavelet transform, NN and evolutionary algorithm for load

* Tel.: +886 3 5131578; fax: +886 3 5710102.
E-mail address: htpao@cc.nctu.edu.tw

forecasting. Additionally, some mid-term forecast models have also been proposed and implemented in this field [17–20]. For long-term energy forecasting, Padmakumari et al. [21] used fuzzy NNs for long-term load forecasting. Kermanshahi and Iwamiya [22] used back-propagation networks and Jordan recurrent networks to forecast Japan's electricity energy consumption until 2020. Pao [23] concluded that the forecasting performance of ANN for Taiwan's energy consumption is higher than that of the other linear models. Moreover, Jebaraj and Iniyan [24] made a literature survey and gave a brief overview of the different types of energy modeling and forecasting.

For a univariate time series forecasting problem, the inputs of the NN are the past lagged observations of the data series, and the outputs are the future values. For multi-step-ahead forecasting, however, one or more output nodes can be used. If one output node is employed, then the iterative forecasting approach is assumed, and each forecast value is used iteratively as input for the next forecasts. In contrast, if the number of output nodes is equal to the length of the forecasting horizon, then the direct forecasting approach is used, in which the future values can be predicted directly from the network outputs [25]. The iterative forecasting approach may generate more prediction errors, because the forecast values are iteratively used as inputs for the next forecasts. The direct forecasting approach can raise serious round-off errors, because the number of output nodes is equal to the length of the forecasting horizon [17].

The aim of this paper is to focus on multi-step-ahead forecasts for energy consumption in Taiwan using univariate modeling. In order to avoid excessive round-off and prediction errors, taking heteroscedastic variations into account, a new hybrid univariate nonlinear network is proposed with two input nodes generated by a linear model: level forecasts $(\widehat{y}_t)$ and volatility forecasts $(\widehat{\sigma}_t)$ and a single output node y_t. The forecast encompassing tests and three different statistical measures are used to assess the out-of-sample performance of the proposed techniques to prevent data mining-induced overfitting.

The rest of this paper is organized as follows. In Section 2, two proposed new hybrid ANN models are described. Section 3 presents the performance evaluation methods by using statistical measures and forecast encompassing tests. The model construction and model comparisons are explained in Sections 4 and 5, respectively. The last section summarizes and concludes the paper.

2. Methodology

This section describes several linear models: the exponential smoothing model (Winters), the exponential form of the generalized autoregressive conditional heteroscedasticity (EGARCH) and seasonal EGARCH (SEGARCH) models, the combined Winters with volatility EGARCH model (WARCH), and an artificial neural network (ANN) nonlinear model. They are briefly described below as the basis on which to present two new hybrid nonlinear models: SEGARCH–ANN and WARCH–ANN. Both hybrid models are formed by combing a linear model with a NN to predict Taiwan's consumption of electricity and petroleum.

2.1. Winters models

Exponential (EXPO) smoothing methods are often useful for forecasting a time series whose parameters change slowly over time. These methods can be implemented by using the Box–Jenkins methodology [7]. For seasonal data, an exponential smoothing approach is the Winters method. In particular, an ARIMA $(0,1,1) \times (0,1,1)_S$ model may be a good alternative to the additive

Winters method, where S is the seasonal periodicity. The additive Winters ARIMA$(0,1,1) \times (0,1,1)_S$ model can be written as

$$\left(1 - B\right)\left(1 - B^S\right)z_t = (1 - \theta_1 B)\left(1 - \Theta_1 B^S\right)v_t, \tag{1}$$

where B is a backward shift operator and v_t is a random error. Let $\widehat{z}_t$ be the forecasting time series; then the residual $v'_t (= z_t - \widehat{z}_t)$ time series is both detrended and deseasonalized.

2.2. EGARCH and seasonal EGARCH (SEGARCH) models

The generalized autoregressive conditional heteroscedasticity (GARCH) model is an approach to modeling time series with heteroscedastic errors [26]. Nelson and Cao [27] argued that the nonnegative constraints on the parameters α_i and γ_i in the linear GARCH model are too restrictive. There are no restrictions on these parameters in the exponential form of the GARCH model (EGARCH). In this model, the conditional variance σ_t^2 is an asymmetric function of the lagged disturbances ε_{t-i}. The EGARCH regression model can be written as

$$z_t = x'_t \beta + \varepsilon_t,$$

$$\varepsilon_t = \sigma_t e_t,$$

$$\ln\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{q} \alpha_i g(e_{t-i}) + \sum_{j=1}^{p} \gamma_j \ln\left(\sigma_{t-j}^2\right), \tag{2}$$

where $g(e_t) = \theta e_t + |e_t| - E|e_t|$, and $e_t \sim N(0,1)$.

Note that $E|e_t| = \sqrt{2/\pi}$ if $e_t \sim N(0,1)$. The function $g(e_t)$ is linear in e_t with slope $\theta + 1$ if e_t is positive, and with slope $\theta - 1$ if e_t is negative.

The seasonal intervention model employs dummy variables to forecast the time series. The model with autoregressive errors and EGARCH variances (SEGARCH) is expressed as follows:

$$z_t = \alpha_0 + \alpha_1 t + d_1 x_{s1,t} + d_2 x_{s2,t} + \cdots + d_{10} x_{s10,t} + d_{11} x_{s11,t} + v_t,$$

$$v_t = \varepsilon_t - \varphi_1 v_{t-1} - \cdots - \varphi_m v_{t-m},$$

$$\varepsilon_t = \sigma_t e_t,$$

$$\ln\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{q} \alpha_i g(e_{t-i}) + \sum_{j=1}^{p} \gamma_j \ln\left(\sigma_{t-j}^2\right), \tag{3}$$

where $g(e_t) = \theta e_t + |e_t| - E|e_t|$, $e_t \sim IN(0,1)$, and

$$x_{s1,t} = \begin{cases} 1 & \text{if period } t \text{ is January} \\ 0 & \text{otherwise} \end{cases} \cdots$$

$$x_{s11,t} = \begin{cases} 1 & \text{if period } t \text{ is November} \\ 0 & \text{otherwise} \end{cases}.$$

This model could be called an AR(m)-SEGARCH(p,q) regression model (henceforth SEGARCH (m,p,q)). The optimal lag length m is determined based on the information criteria, AIC and SBC, and the Durbin Watson (DW) statistic. Both the Portmanteau Q statistic [28] and the Lagrange multiplier (LM) test [29] are used to determine the lag lengths p and q of the ARCH model. These tests are significant ($p < 0.0001$) for lags between 1 and 12, which indicate that a very high order ARCH process is needed to model the heteroscedasticity. Both the forecasted values of $\widehat{z}_t$ and $\widehat{\sigma}_t$ are used as inputs of the SEGARCH–ANN model, discussed in the following sections.

2.3. The combined Winters with volatility EGARCH model (WARCH)

Let $v'_t = z_t - \widehat{z}_t$ be the t-th residual, where z_t is the observed value and $\widehat{z}_t$ is the predicted value given by the Winters model. Therefore, $\{v'_t\}$ is a detrended and deseasonalized time series. The Ljung–Box Q^* statistics are used to test the autocorrelation. If the p-values of Q^* are less than 0.05, this is an evidence that the $\{v'_t\}$ is highly autocorrelated. To construct the EGARCH model for $\{v'_t\}$, the three statistical tests, DW, AIC and SBC, are used for the autocorrelation to determine the lag length m, and both the Q and LM tests are used for the ARCH process to determine the lag lengths p and q. Once these tests indicate heteroscedasticity with $p < 0.05$ for lag between 1 and 12, the EGARCH model can be used to produce a forecasted conditional error variance $\widehat{\sigma}_t^2$ by modeling the residuals $\{v'_t\}$ with heteroscedastic errors. The proposed two-step WARCH (m,p,q) model combines a Winters model in the first step to obtain the detrended and deseasonalized residuals $\{v'_t\}$ with the AR(m)-EGARCH (p,q) model in the second step to produce the estimated heteroscedastic error variance $\widehat{\sigma}_t$ for the historical and forecast periods. The WARCH (m,p,q) model is expressed as

$$\text{Step 1}: \left(1 - B\right)\left(1 - B^S\right)z_t = (1 - \theta_1 B)\left(1 - \Theta_1 B^S\right)$$
$$v_t, v'_t = z_t - \widehat{z}_t,$$

$$\text{Step 2}: v'_t = \varepsilon_t - \varphi_1 v'_{t-1} - \cdots - \varphi_m v'_{t-m},$$
$$\varepsilon_t = \sigma_t e_t,$$
$$\ln\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{q} \alpha_i g(e_{t-i}) + \sum_{j=1}^{p} \gamma_j \ln\left(\sigma_{t-j}^2\right),$$
$$\text{where} \quad g(e_t) = \theta e_t + |e_t| - \sqrt{2/\pi}, \text{ and } e_t \sim IN(0,1).$$

Both forecasted values of $\widehat{z}_t$ and $\widehat{\sigma}_t$ from the Winters and ARCH steps, respectively, are used in the WARCH–ANN model discussed below.

2.4. Artificial neural network (ANN) model

NNs can be described as an attempt by humans to mimic the functioning of the human brain. The models are analytical techniques modeled after the processes of learning in the cognitive system and the neurological functions of the brain and are capable of predicting new observations (of specific variables) from other observations (of the same or other variables) after executing a process of so-called learning from existing data [20]. The models can be built without explicitly formulating the possible relationship that exists between variables. Theoretical results show that NNs are also able to sufficiently approximate arbitrary mappings to the desired accuracy if given a large enough network [30]. In this sense, NNs may be seen as multivariate, nonlinear and nonparametric methods, and they should be expected to model complex nonlinear relationships much better than the traditional linear models.

Fig. 1 shows a popular three-layer feedforward NN model. It consists of one input layer with m input variables, one hidden layer with h hidden nodes, and one output layer with a single output node. The hidden layers perform nonlinear transformations on the inputs from the input layer and feed the transformed values to the output layer. The connection weights and node biases are the model parameters. The model estimation process is called network training. Usually in applications of ANNs, the total available data are split into a training set and a test set. The training set is used to calibrate the network model, while the test set is used to evaluate its forecasting ability. During the training procedure, an overall error measure is minimized to get the estimates of the parameters
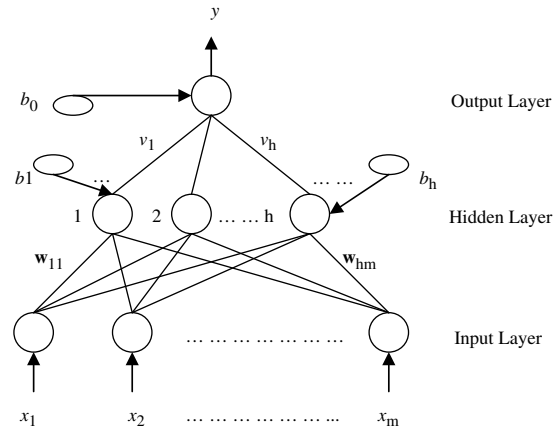

Fig. 1. Neural network model.

of the models. More detailed materials about NN learning can be found in Bishop [31].

For m-step-ahead forecasting ($m > 1$), both iterative and direct forecasting approaches can be used. The iterative forecasting approach with p input nodes has a mapping function of the form

$$y_{t+1} = f\left(y_t, y_{t-1}, \ldots, y_{t-p+1}\right),$$
$$y_{t+2} = f\left(\widehat{y}_{t+1}, y_t, \ldots, y_{t-p+2}\right),$$
$$y_{t+3} = f\left(\widehat{y}_{t+2}, \widehat{y}_{t+1}, y_t, \ldots, y_{t-p+3}\right),$$
$$\vdots$$
$$y_{t+m} = f\left(\widehat{y}_{t+m-1}, \widehat{y}_{t+m-2}, \ldots, \widehat{y}_{t+1}, y_t, \ldots, y_{t-p+m}\right). \quad (5)$$

The direct forecasting approach has a mapping function of the form

$$\left(y_{t+m}, y_{t+m-1}, \ldots, y_{t+1}\right) = f\left(y_t, y_{t-1}, \ldots, y_{t-p+1}\right). \quad (6)$$

The iterative forecasting approach may generate more prediction errors, because the forecast values are iteratively used as inputs for the next forecasts. The direct forecasting approach, however, is subject to serious round-off errors, because the number of output nodes is equal to the length of the forecasting horizon.

In order to avoid excessive round-off and prediction errors, taking heteroscedastic variations into account, a new hybrid univariate network with two input nodes, $\widehat{\sigma}_t$ and $\widehat{y}_t$, and a single output node is proposed. The form of the mapping function can be expressed as

$$y_{t+1} = f\left(\widehat{\sigma}_{t+1}, \widehat{y}_{t+1}\right), y_{t+2} = f\left(\widehat{\sigma}_{t+2}, \widehat{y}_{t+2}\right), \ldots, y_{t+m}$$
$$= f\left(\widehat{\sigma}_{t+m}, \widehat{y}_{t+m}\right), \quad (7)$$

where $(\widehat{\sigma}_{t+1}, \widehat{y}_{t+1}), (\widehat{\sigma}_{t+2}, \widehat{y}_{t+2}), \ldots, (\widehat{\sigma}_{t+m}, \widehat{y}_{t+m})$ can be predicted by using a linear model. This hybrid model, using a univariate modeling approach for multi-step-ahead forecasting, is described in the next section.

2.5. The hybrid SEGARCH–ANN and WARCH–ANN models

The practical advantage of ANN models is that the relationships between input and output variables do not need to be specified in advance, since the method itself establishes these relationships

through a training process. Also, ANNs do not require any assumptions on the underlying population distributions.

Both the SEGARCH and WARCH linear approaches are outlined above. The two-step WARCH (p,q) model produces $\hat{z}_t$ and the detrained and deseasonalized residual v'_t in the first Winters step, and the estimated heteroscedastic error variance $\hat{\sigma}_t$ in the second ARCH step for the historical and forecast periods. The SEGARCH model generates predicted values $\hat{z}_t$ and its conditional standard deviation estimates $\hat{\sigma}_t$ for the historical and forecast periods. The step proposed here takes only the two values $\hat{z}_t$ and $\hat{\sigma}_t$ as inputs to an ANN model.

The new hybrid nonlinear univariate model is constructed by using a two-step process. In step 1, a linear model with an error volatility component (SEGARCH or WARCH) is estimated to generate values for both the level forecasts $\hat{z}_t$ and the volatility forecasts $\hat{\sigma}_t$. In step 2, both of the values estimated in step 1, $\hat{\sigma}_t$ and $\hat{z}_t$, are plugged into an ANN model with the corresponding output target z_t. These models are called the WARCH–ANN and SEGARCH–ANN models. All of the proposed hybrid ANN models are formed by combining a linear model with a NN to develop an adjusted forecast for Taiwan's electricity and petroleum consumption levels.

In order to prevent data mining-induced overfitting, this paper uses out-of-sample tests to compare the multi-step-ahead forecasting capabilities of the WARCH, SEGARCH, WARCH–ANN, and SEGARCH–ANN univariate models, where WARCH and SEGARCH are the benchmark models.

3. Forecasting evaluation methods

For the purpose of evaluating out-of-sample forecast capability, two different testing approaches are used. The first test associates the three evaluation statistics, root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage forecast error (MAPE), to each model. They are expressed as below:

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n}(P_i - A_i)^2 \Big/ n},$$

$$\mathrm{MAE} = \sum_{i=1}^{n}|P_i - A_i|\Big/ n, \tag{8}$$

$$\mathrm{MAPE} = \sum_{i=1}^{n}\left|(P_i - A_i)\Big/ A_i\right|\Big/ n \times 100.$$

where P_i and A_i are the i-th forecasting and actual values, respectively, and n is the total number of predictions. Lewis [32] interprets the MAPE results as a means to judge the accuracy of the forecast:

less than 10% is a highly accurate forecast, 10–20% is a good forecast, 20–50% is a reasonable forecast, and more than 50% is an inaccurate forecast.

The second test for forecast encompassing was introduced by Chong and Hendry [33]. This test formalizes the intuition that model i should be preferred to model j if model i can explain what model j cannot explain, without model j being able to explain what model i cannot explain. Granger and Newbold [34] argued that forecast encompassing was more stringently required than forecast accuracy. Clements and Hendry [35] proposed the argument that the encompassing test is implemented through testing the significances of the α_1 and β_1 coefficients in the following two regression equations:

$$E_i = \alpha_0 + \alpha_1 D_{ij} + \omega_t, \tag{9}$$

$$E_j = \beta_0 + \beta_1 D_{ij} + v_t,$$

where E_i and E_j denote the forecast errors for model i and model j ($E_i = P_i - A_j$, $E_j = P_j - A_j$), respectively; D_{ij} denotes the differences between the forecast results i and j models ($D_{ij} = P_i - P_j$), and ω_t and v_t are random errors. The null hypothesis is that neither model encompasses (outperforms) the other. If α_1 is significantly different from zero and β_1 is not, then the null hypothesis is rejected in favor of the alternative hypothesis that model j encompasses model i. Conversely, if β_1 is significant but α_1 is not, then this is an evidence that model i encompasses model j. If neither α_1 nor β_1 is significant, or conversely if both α_1 and β_1 are significant, then we fail to reject the null hypothesis and conclude that neither model encompasses the other. Table 4 reports the results of the forecast encompassing tests.

4. Experimental results

In this section, the performance of the alternative modeling approaches is compared using two seasonal time series: electricity consumption and petroleum consumption in Taiwan. The period under examination extends from January 1993 to December 2007 with a total of 180 observations for each series. The period from January 1993 to December 2005 is treated as the estimation (or training) period for the models. The subsequent period, from January 2006 through December 2007, is the testing or out-of-sample period.

4.1. Electricity consumption series

As shown in Fig. 2, the time series data of Taiwan's electricity consumption show strong seasonality and growth trends. The peak
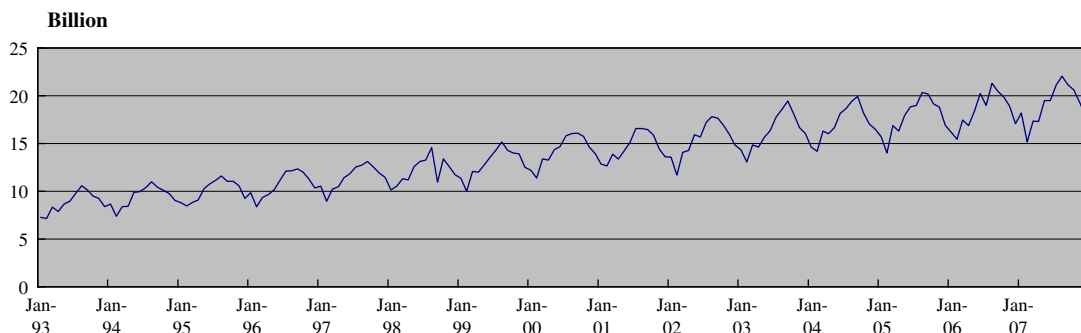

Fig. 2. Taiwan's electricity consumption from January 1993 to December 2007.

	WARCH–ANN	SEGARCH–ANN
Input nodes	2	2
The number of hidden neurons	3	4
Learning rate	0.04	0.1
Momentum	0.1	0.1
RMSE of training data	0.019	0.016

season for each year generally occurs in June or July, because energy use is greatest in the summer.

4.1.1. Winters and WARCH models for electricity

The electricity consumption series z_t, given in Fig. 2, assumes that the seasonality and the growth trend exist in the historical data and extend to the future with the same pattern. The statistical properties can be examined by using the autocorrelation function (acf) and the partial autocorrelation function (pacf). The result reveals that z_t is non-stationary. The first regular differences and first seasonal differences are calculated in order to remove the growth trend and the seasonality characteristics. In this step, the first 13 observations are lost. The stationary time series thus acquired can be used to identify the Winters model. The estimated equation is presented as follows:

$$\left(1 - B\right)\left(1 - B^{12}\right)z_t = \left(1 - 0.820B\right)\left(1 - 0.671B^{12}\right)v_t. \tag{10}$$

Let $\widehat{z}_t$ be the point prediction of z_t and $v'_t = z_t - \widehat{z}_t$ be the t-th residual, where $\{v'_t\}$ is the detrended and deseasonalized time series with 143 observations. The Augmented Dickey-Fuller (ADF) unit root test can be rejected for the residuals $\{v'_t\}$ at the 5% level of significance, since the ADF test statistic -12.27 is lower than the critical value -2.88. This indicates that the series is stationary. The p-values for the Ljung–Box test statistics are less than 0.05, this means that the $\{v'_t\}$ is highly autocorrelated. Both the Q and LM tests are significant with $p < 0.05$ for lags between 1 and 12, which indicate that an ARCH process is needed to model hetero-scedasticity. The conditional error variance $\widehat{\sigma}_t$ for $\{v'_t\}$ can be forecasted by estimating the parameters of EGARCH process. The new WARCH model is constructed by combining the Winters with the AR(13,18,23,24)-EGARCH $(q = (1,24))$ model describing the error variance, which can be expressed as

$$\text{Step 1}: \left(1 - B\right)\left(1 - B^{12}\right)z_t = \left(1 - 0.820B\right)\left(1 - 0.671B^{12}\right)v_t,$$
$$v'_t = z_t - \widehat{z}_t,$$

$$\text{Step 2}: v'_t = -39783 - 0.148v'_{t-13} + 0.183v'_{t-18} - 0.330v'_{t-23}$$
$$+ 0.041v'_{t-24} + \varepsilon_t, \varepsilon_t = \sigma_t e_t,$$

$$\ln\left(\sigma_t^2\right) = 25.623 + 0.541g(e_{t-1}) + 0.725g(e_{t-24}),$$

$$\text{where} \quad g(e_t) = 0.042e_t + |e_t| - \sqrt{2/\pi},$$
$$e_t \sim IN(0, 1). \tag{11}$$

Once estimated, Eq. (11) can be used to compute $\widehat{z}_t$ and $\widehat{\sigma}_t$ in the historical and forecast periods from the Winters and ARCH steps, respectively. Both the values of $\widehat{z}_t$ and $\widehat{\sigma}_t$ are used as the input variables in the WARCH–ANN model whose corresponding output value is z_t.

4.1.2. SEGARCH model

The electricity consumption series z_t given in Fig. 2 exhibits a reasonable deterministic linear trend and monthly seasonal variation. Seasonal intervention models are employed to forecast this time series. The derived SEGARCH model with autoregressive error AR(1,2,8)-SEGARCH $(q = (1,24))$ is

$$z_t = 5609782 + 58818t - 243591x_{s1,t} - 1049180x_{s2,t}$$
$$- 40242x_{s3,t} + 25837x_{s4,t} + 788991x_{s5,t} + 1284619x_{s6,t}$$
$$+ 1902389x_{s7,t} + 2334958x_{s8,t} + 1684728x_{s9,t}$$
$$+ 1498304x_{s10,t} + 818588x_{s11,t} + v_t,$$

$$v_t = \varepsilon_t - 0.230v_{t-1} - 0.196v_{t-2} - 0.102v_{t-8},$$

$$\varepsilon_t = \sigma_t e_t, \tag{12}$$

$$\ln\left(\sigma_t^2\right) = 25.716 + 0.360g(e_{t-1}) + 0.533g(e_{t-24}),$$

$$\text{where} \quad g(e_t) = 0.477e_t + |e_t| - \sqrt{2/\pi}, \ e_t \sim IN(0, 1) \quad \text{and}$$
$$x_{s1,t} = \begin{cases} 1 & \text{if period } t \text{ is January} \\ 0 & \text{otherwise} \end{cases} \dots$$
$$x_{s11,t} = \begin{cases} 1 & \text{if period } t \text{ is November} \\ 0 & \text{otherwise} \end{cases}.$$

The estimated parameters and estimated values of $\widehat{z}_t$ and $\widehat{\sigma}_t$ are obtained simultaneously for the sample and forecast periods. The values of both $\widehat{z}_t$ and $\widehat{\sigma}_t$ are used as input variables in SEGARCH–ANN model where the corresponding output value is z_t.

4.1.3. WARCH–ANN and SEGARCH–ANN models

In this step, $\widehat{z}_t$ and $\widehat{\sigma}_t$ values from the WARCH and SEGARCH estimation steps are regarded as input variables to WARCH–ANN and SEGARCH–ANN models in which z_t is used in the output layer. All networks are trained with the forecasting data from January 1993 to December 2005 and forecast for 24 months from January 2006 to December 2007. A back-propagation learning algorithm is used in the training process. More than 50 experiments are conducted to
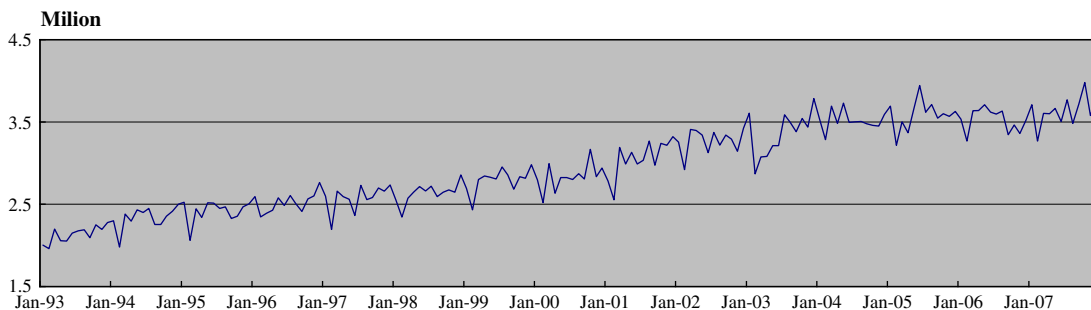

Fig. 3. Taiwan's petroleum consumption from January 1993 to December 2007.

Table 2
The best results of neural networks for petroleum consumption data.

	WARCH–ANN	SEGARCH–ANN
Input nodes	2	2
The number of hidden neurons	5	5
Learning rate	0.4	0.06
Momentum	0.1	0.05
RMSE of training data	0.029	0.023

determine the best combination of the learning rates, momentum, and number of hidden nodes. Throughout the training, the Neural-Ware [36] utility 'SAVEBEST' is used to monitor and save the lowest RMSE from the training set. For WARCH–ANN, the best RMSE result is obtained by using a learning rate of 0.08, a momentum rate of 0.1, and 3 nodes in a single hidden layer that uses the generalized data learning rule and a sigmoid transfer function ($y = 1/(1 + e^{-x})$). The best architecture for the network is {2:3:1}. For the SEGARCH–ANN, the data period and the estimation details for the ANN are the same as those discussed above. The best architecture of the network is {2:4:1}. The results are reported in Table 1.

4.2. Petroleum consumption series

The petroleum consumption time series is recorded monthly wise from 1993 to 2007. These data include trend and seasonal varia tions, as shown in Fig. 3. With their electric data, the values from 1993 to 2005 (156 observations) are used for estimating the models, and the monthly data from 2006/1 to 2007/12 are used for testing.

4.2.1. Winters, WARCH, and SEGARCH models

Investigation of the *acf* for the petroleum data reveals that it is non-stationary. The first regular differences and first seasonal differences are calculated in order to remove the growth trend and the seasonality characteristics. This process loses the first 13 observations. The resulting stationary time series (143 observations) can be used to identify the Winters model. The estimated equation is presented below:

$$(1 - B)\left(1 - B^{12}\right)y_t = (1 - 0.804B)\left(1 - 0.668B^{12}\right)v_t \qquad (13)$$

The ADF unit root test can be rejected for the residuals $\{v'_t\}$ at the 5% level of significance, since the ADF test statistic -11.02 is lower than the critical value -2.88. This indicates that the series is stationary. Both the Q and LM tests are significant with $p < 0.05$ for lags between 1 and 12. This enables us to estimate the conditional

error variances $\widehat{\sigma}_t^2$ by using EGARCH process in historical and forecast periods. The final WARCH((16),0,5) model, combining the Winters with the AR(16)-EGARCH ($q = (5)$) model, is expressed as follows:

$$\text{Step 1}: \left(1 - B\right)\left(1 - B^{12}\right)y_t = (1 - 0.804B)\left(1 - 0.668B^{12}\right)$$
$$v_t, v'_t = y_t - \widehat{y}_t, \qquad (14)$$

$$\text{Step 2}: v'_t = 3935 + 0.215v'_{t-16} + \varepsilon_t,$$
$$\varepsilon_t = \sigma_t e_t,$$
$$\ln\left(\sigma_t^2\right) = 23.061 + 0.356g(e_{t-5}),$$

where $g(e_t) = -0.339e_t + |e_t| - \sqrt{2/\pi},\ \ e_t \sim IN(0, 1)$.

For the SEGARCH model with autoregressive errors, the final specification for the petroleum series is AR(1,5)-SEGARCH ($q = (1,24)$), and the estimated model is as follows:

$$y_t = 1919088 + 8615t - 97131x_{s1,t} - 374934x_{s2,t}$$
$$- 59440x_{s3,t} - 134773x_{s4,t} - 54148x_{s5,t} - 112180x_{s6,t}$$
$$- 54368x_{s7,t} - 85199x_{s8,t} - 176545x_{s9,t} - 62519x_{s10,t}$$
$$- 88123x_{s11,t} + v_t,$$

$$v_t = -0.208v_{t-1} - 0.302v_{t-5} + \varepsilon_t,$$

$$\varepsilon_t = \sigma_t e_t, \qquad (15)$$

$$\ln\left(\sigma_t^2\right) = 22.682 + 0.2293g(e_{t-1}) - 0.4362g(e_{t-24}), \text{where}$$

$$g(e_t) = 0.458e_t + |e_t| - \sqrt{2/\pi}, \quad e_t \sim IN(0, 1),$$
$$x_{s1,t} = \begin{cases} 1 & \text{if period } t \text{ is January} \\ 0 & \text{otherwise} \end{cases} \quad \ldots$$
$$x_{s11,t} = \begin{cases} 1 & \text{if period } t \text{ is November} \\ 0 & \text{otherwise} \end{cases}.$$

4.2.2. WARCH–ANN and SEGARCH–ANN models

The predicted values and the conditional variance estimates from the WARCH and SEGARCH models for petroleum are used to estimate the ANN models in this section. The back-propagation learning algorithm is used in the training process. For the
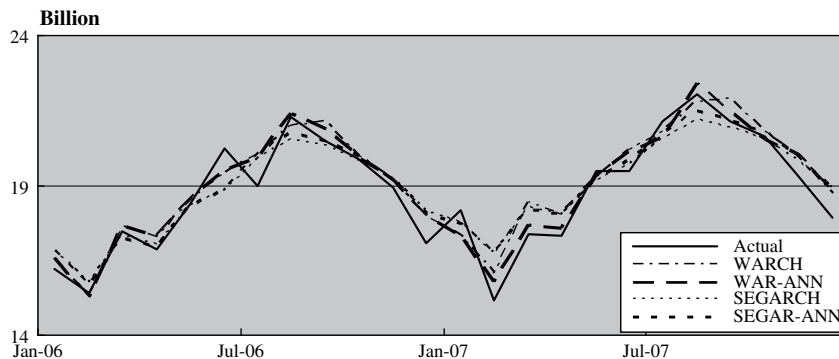

Fig. 4. Actual and model values for electricity consumption data.
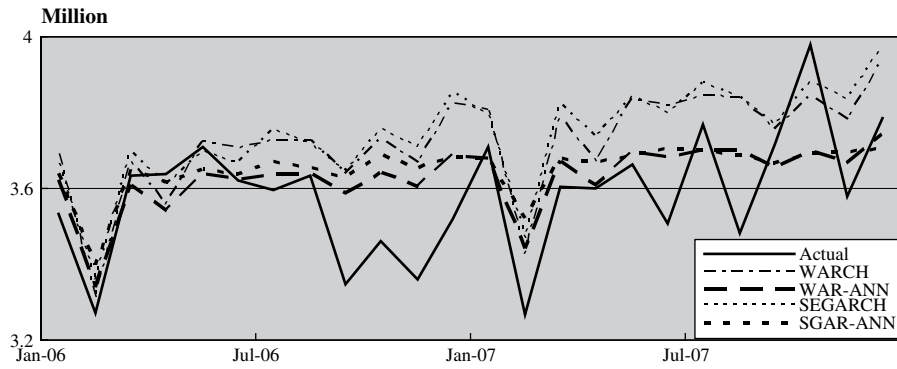
Million

Fig. 5. Actual and model values for petroleum consumption data.

WARCH–ANN model, the best RMS error is obtained using a learning rate of 0.4, a momentum of 0.1, and 5 nodes in a single hidden layer ({2-5-1}). For the SEGARCH–ANN model, the best RMS error result is obtained by using a learning rate of 0.06, a momentum of 0.05, and 5 hidden nodes ({2-5-1}). The results are reported in Table 2.

5. Out-of-sample forecasting performance comparison

In this section, the out-of-sample forecasting ability of four models (WARCH, SEGARCH, WARCH–ANN and SEGARCH–ANN) is evaluated over a 24-month forecast period, where WARCH and SEGARCH are the benchmark models. For the years 2006 and 2007, the forecasting values given by the proposed four models as well as the actual values for both types of energy are shown in Figs. 4 and 5. Clark [37] showed that out-of-sample forecast comparisons can help prevent data mining-induced overfitting. While the hybrid ANN would clearly be expected to dominate in the sample, since it nests the linear model, there is in fact no a prior guarantee that the hybrid ANNs will dominate with out-of-sample data. Indeed, it is possible that ANN could overfit the data in the sample and thus produces out-of-sample forecasts that are inferior to forecasts from the linear model [38]. Thus, three different statistical measures and forecast encompassing tests are employed to evaluate the out-of-sample forecast capability of each of the linear and hybrid nonlinear models.

5.1. Root mean square, mean absolute and mean absolute percentage forecast errors

Table 3 reports the RMSE, MAE and MAPE for each model for the out-of-sample period from January 2006 to December 2007. The results show that the WARCH model is better than the SEGARCH model, SEGARCH–ANN is better than the SEGARCH model, and WARCH–ANN is the best of the four models for both electricity and petroleum consumption. However, SEGARCH–ANN is better than the WARCH model on petroleum consumption only. All of the models have highly accurate forecasts, because the MAPE results are less than 5% [32].

Furthermore, none of the comparisons, neither by RMSE nor by MAE and MAPE, can provide any indication of whether any one model's performance is significantly better than that of other models in a formal statistical sense [38]. Therefore, in the next section we present an additional means of comparison between forecasting models, namely comparison by forecast encompassing, which allows us to test whether one model has significantly better performance than another.

5.2. Forecast encompassing tests

The forecast encompassing test was applied to the out-of-sample comparison for nested models by Clark and McCracken [39]. The results from the encompassing tests reported in Table 4 paint a picture similar to that in Table 3. This table reports the t-statistics of the estimated coefficients $\widehat{\alpha}_1$ and $\widehat{\beta}_1$ from Eq. (9) and the corresponding p-values. As clearly seen from the Table 4 (Panel A for the electricity consumption and Panel B for the petroleum consumption), the differences D_{14}, D_{24}, and D_{34} between the WARCH–ANN and the other three models (Winters, SEGARCH, SEGARCH–ANN) explain the forecasting errors E_1, E_2, and E_3 well from each of the alternative models, respectively. Moreover, the forecasting errors (E_4) of the WARCH–ANN model cannot be accounted for by any of the differences, D_{14}, D_{24}, or D_{34}, for either fuel. Pairwise comparisons for the encompassing reveal that WARCH–ANN is the only model whose forecast is not encompassed by the other models, and WARCH–ANN significantly encompasses the other models. Thus, WARCH–ANN can be considered the dominant forecasting device for both energy consumptions. Table 4 also reveals that SEGARCH–ANN significantly encompasses the SEGARCH model. The graphical representation of encompassing tests is shown in Fig. 6 for both the types of energy.

As a result, it should be clear that the ANN steps reproduce the predicted values from the initial linear model and the ANN step encompassing step 1. Moreover, a poor linear model may produce a poor hybrid nonlinear model. The output of the SEGARCH model is poorer than that of the Winters one, so the WARCH–ANN results are better than the SEGARCH–ANN ones. The significantly superior performance of the hybrid nonlinear ANN models compared with other conventional linear models in

Table 3
Comparative forecasting performance of energy consumption.

	WARCH	WARCH–ANN	SEGARCH–ANN	SEGARCH
Panel A: electricity energy consumption				
Input nodes		$\widehat{z}_t, \widehat{\sigma}_t$	$\widehat{z}_t, \widehat{\sigma}_t$	
RMSE	643744.33	**531545.14**	596013.96	824500.08
MAE	474189.18	**404184.25**	464632.42	606629.27
MAPE	2.90%	**2.56%**	2.98%	3.65%
Panel B: petroleum energy consumption				
Input nodes		$\widehat{z}_t, \widehat{\sigma}_t$	$\widehat{z}_t, \widehat{\sigma}_t$	
RMSE	165753.68	**134832.21**	148234.91	204369.84
MAE	134300.13	**112542.53**	122320.08	167031.13
MAPE	4.08%	**3.51%**	3.71%	4.88%

The forecasting period is from January 2006 to December 2007.

Table 4
Encompassing tests of forecasting performance of alternative models.

| Dependent variable: forecasting errors	Independent variable: forecasting difference from two models					
	D_{12} [b]	D_{13} [b]	D_{23} [b]	D_{14} [b]	D_{24} [b]	D_{34} [b]
Panel A: electricity consumption (forecast period: from January 2006 to December 2007)						
E_1 [a] (WARCH)	−1.52*(−2.98)	0.52*(1.26)		**0.81*(2.84)**		
E_2 [a] (SEGARCH)	−2.52*(−4.95)		**1.34*(3.57)**		**0.97*(4.73)**	
E_3 [a] (SEGARCH–ANN)		−0.48*(−1.15)	0.34(0.90)			**0.70*(3.08)**
E_4 [a] (WARCH–ANN)				−0.19(−0.69)	−0.03(−0.16)	−0.30(−1.33)
Panel B: petroleum consumption (forecast period: from January 2006 to December 2007)						
E_1 [a] (WARCH)	0.90(1.17)	1.66(1.24)		**0.82*(2.71)**		
E_2 [a] (SEGARCH)	−0.10(−0.21)		**1.00*(4.55)**		**−0.10*(−3.02)**	
E_3 [a] (SEGARCH–ANN)		0.66(1.03)	−0.13(−0.21)			**−0.65*(−2.90)**
E_4 [a] (WARCH–ANN)				−0.12(−0.09)	−1.18(−1.05)	−1.65(−1.08)

Which indicates that statistical significance is at the 0.1 level.
The values in parentheses are *t*-statistics.
  [a] 1 = WARCH, 2 = SEGARCH, 3 = SEGARCH–ANN, 4 = WARCH–ANN and E_i denotes the forecast error for model *i*.
  [b] D_{ij} denotes the difference between the forecast from the model *i* and model *j*.
  * Which indicates that statistical significance is at the 0.05 level.
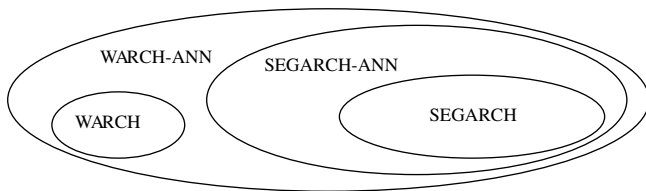
Fig. 6. Results of encompassing tests for both types of energy.

out-of-sample test suggests that flexible hybrid ANNs may be able to account for potentially complex nonlinear relationships not easily captured by linear models.

6. Conclusion

Forecasting energy consumption is of utmost importance to the reconstruction process going on in Taiwan, specifically in the energy generation systems. It is a challenge for us to develop forecast tools with the data obtained in Taiwan due to her rapid economic growth and increasing demand for both electricity and petroleum. Taking the heteroscedastic variation into account, this study proposes two hybrid univariate nonlinear models, SEGARCH–ANN and WARCH–ANN for multi-step-ahead forecasting. Both hybrid ANNs can decrease round-off and prediction errors for multi-step-ahead forecasting.

The out-of-sample forecasting performance of each model is assessed by three statistical measures: RMSE, MAE, MAPE, and encompassing tests. The results of the statistical measures suggest that the WARCH model is better than the SEGARCH model, SEGARCH–ANN is better than the SEGARCH model, and WARCH–ANN is the best of the four models for both electricity and petroleum consumption. However, SEGARCH–ANN is better than the WARCH model on petroleum consumption only. All of the models have highly accurate forecasts, since the MAPE results are less than 5%. Furthermore, the WARCH–ANN model significantly encompasses the other three models, while SEGARCH–ANN considerably encompasses the SEGARCH model in the consumption of both the types of energy. So it should be clear that the ANN steps would reproduce the predicted values from the initial linear model, and the ANN step would encompass the linear step. The WARCH–ANN model is the dominant forecasting approach for both electricity and petroleum.

In summary, the preponderance of the statistical evidence presented in this paper suggests that the proposed hybrid ANNs forecasts generally outperform the forecasts from a variety of linear models in predicting the data on Taiwan's energy consumption. The practical significance of this result is evident from the out-of-sample nature of the tests employed. Although the hybrid ANN nests the linear model as a special case and would therefore be expected to dominate this model in the sample, assuming there was no a priori guarantee that hybrid ANNs would dominate out-of-sample, especially if the ANNs overfit the in-sample data. The fact that the proposed hybrid ANNs did outperform the conventional linear models in the out-of-sample tests therefore reveals that flexible hybrid ANNs may be to account for potentially complex nonlinear relationships not easily captured by linear models. Furthermore, the information on the interactions and nonlinear integrating effects between $(\widehat{z}_t, \widehat{\sigma}_{z_t})$ and z_t are important because both types of energy consumption are well accommodated by the hybrid nonlinear algorithm. Overall, the inclusion of heteroscedastic variations in the input layer of the hybrid univariate model could help improve the modeling accuracy of multi-step-ahead forecasts.

References

[1] Himanshu AA, Lester CH. Electricity demand for Sri Lanka: a time series analysis. Energy 2008;33:724–39.
[2] Fatai K, Oxley L, Scrimgeour FG. Modeling and forecasting the demand for electricity in New Zealand: a comparison of alternative approaches. The Energy Journal 2003;24(1):75–102.
[3] Hamzacebi C. Forecasting of Turkey's net electricity energy consumption on sector bases. Energy Policy 2007;35:2009–16.
[4] Mohamed Z, Bodger P. Forecasting electricity consumption in New Zealand using economic and demographic variables. Energy 2005;30(10):1833–43.
[5] Gorucu FB, Gumrah F. Evaluation and forecasting of gas consumption by statistical analysis. Energy Sources 2004;26:267–76.
[6] Yang M, Yu X. China's rural electricity market – a quantitative analysis. Energy 2004;29(7):961–77.
[7] Box GEP, Jenkins GM. Time series analysis: forecasting and control. San Francisco: Holden-Day; 1976.
[8] Bowden N, Payne JE. Short term forecasting of electricity prices for MISO hubs: evidence from ARIMA–EGARCH models. Energy Economics 2008;30(6):3186–97.
[9] Ediger VS, Akar S. ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy 2007;35(3):1701–8.

[10] Pappas SS, Ekonomou L, Karamousantas DC, Chatzarakis GE, Katsikas SK, Liatsis P. Electricity demand loads modeling using AutoRegressive moving average (ARMA) models. Energy 2008;33(9):1353–60.

[11] Saab S, Badr E, Nasr G. Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon. Energy 2001;26:1–14.

[12] Zhou P, Ang BW, Poh KL. A trigonometric grey prediction approach to forecasting electricity demand. Energy 2006;31(14):2839–47.

[13] Gonzalez PA, Zamarreno JA. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. Energy and Buildings 2005;37(6):595–601.

[14] Lauret P, Fock E, Randrianarivony RN, Manicom-Ramasamy JF. Bayesian neural network approach to short time load forecasting. Energy Conversion and Management 2008;49(5):1156–66.

[15] Hippert HS, Pedreira CE, Souza RC. Neural networks for short-term load forecasting: a review and evaluation. IEEE Transactions on Power Systems 2001;16(1):44–55.

[16] Amjady N, Keynia F. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. Energy 2009;34(1):46–57.

[17] Pao HT. Forecasting electricity market pricing using artificial neural networks. Energy Conversion and Management 2007;48:907–12.

[18] Mirasgedis S, Sarafidis Y, Georgopoulou E, Lalas DP, Moschovits A, Karagiannis F, et al. Models for mid-term electricity demand forecasting incorporating weather influences. Energy 2006;31:208–27.

[19] Amjady N, Keynia F. Mid-term load forecasting of power systems by a new prediction method. Energy Conversion and Management 2008;49(10):2678–87.

[20] Tso GKF, Yau KKW. Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural network. Energy 2007;32(9):1761–8.

[21] Padmakumari K, Mohandas KP, Thiruvengadam S. Long term distribution demand forecasting using neuro fuzzy computations. International Journal of Electrical Power and Energy Systems 1999;21(5):315–22.

[22] Kermanshahi B, Iwamiya H. Up to year 2020 load forecasting using neural nets. International Journal of Electrical Power and Energy Systems 2002;24:789–97.

[23] Pao HT. Comparing linear and nonlinear forecasts for Taiwan's electricity consumption. Energy 2006;31:1793–805.

[24] Jebaraj S, Iniyan S. A review of energy models. Renewable and Sustainable Energy Reviews 2006;10:281–311.

[25] Hu MY, Zhang G, Jiang CX, Patuwo BE. A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting. Decision Sciences 1999;30(1):197–216.

[26] McKenzie E. General exponential smoothing and the equivalent ARMA process. Journal of Forecasting 1984;3:333–44.

[27] Nelson DB, Cao CQ. Inequality constraints in the univariate GARCH model. Journal of Business and Economic Statistics 1992;10(2):229–35.

[28] McLeod AI, Li WK. Diagnostic checking ARMA time series models using squared-residual autocorrelations. Journal of Time Series Analysis 1983;4:269–73.

[29] Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 1982;50:987–1007.

[30] Zhang GP. An investigation of neural networks for linear time-series forecasting. Computer and Operations Research 2001;28:1183–202.

[31] Bishop CM. Neural networks for pattern recognition. Oxford: Oxford University Press; 1995.

[32] Lewis CD. Industrial and business forecasting method. London: Butterworth; 1982.

[33] Chong YY, Hendry DF. Econometric evaluation of linear macroeconomic models. Review of Economic Studies 1986;53:671–90.

[34] Granger CWJ, Newbold P. Forecasting economic time series. Orlando, FL: Academic Press; 1986.

[35] Clements MP, Hendry DF. Forecasting economic time series. UK: Cambridge University Press; 1998.

[36] NeuralWare. Neural computing: neural works professional II/PLUS and neuralworks explorer. NeuralWare 1993;.

[37] Clark TE. Can out-of-sample forecast comparisons help prevent overfitting. Journal of Forecasting 2004;23:115–39.

[38] Donaldson RG, Kamstra M. Forecast combining with neural networks. Journal of Forecasting 1996;15:49–61.

[39] Clark TE, McCracken MW. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 2001;105:85–110.