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

A R T I C L E   I N F O

Article history:
Received 9 September 2008
Received in revised form
25 March 2009
Accepted 22 April 2009
Available online 15 August 2009

Keywords:
Energy consumption
Artificial neural networks
SEGARCH models
Multi-step-ahead forecasting

A B S T R A C T

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 forecasting.
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, Jebrair and Injyan [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 model is proposed with two input nodes generated by a linear model: level forecasts \( y_t \) and volatility forecasts \( \hat{y}_t \) 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 (GARCH) and seasonal GARCH (SEGARCH) models, the combined Winters with volatility GARCH 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 \((0,1,1) \times (0,1,1)_{12}\) model may be a good alternative to the additive Winters method, where \( S \) is the seasonal periodicity. The additive Winters ARIMA\((0,1,1) \times (0,1,1)\) model can be written as

\[
(1 - B)(1 - B^S)z_t = (1 - \theta_1 B)(1 - \Theta_1 B^S)r_t,
\]

where \( B \) is a backward shift operator and \( v_t \) is a random error. Let \( \hat{z}_t \) be the forecasting time series; then the residual \( r_t = z_t - \hat{z}_t \) 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 \( a_0 \) and \( \gamma \) 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 \( \sigma_t^2 \) is an asymmetric function of the lagged disturbances \( e_{t-1} \).

EGARCH regression model can be written as

\[
z_t = \mu + \eta_t + \epsilon_t,
\]

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

\[
z_t = \alpha_0 + \alpha_1 t + d_1 x_{S1.t} + d_2 x_{S2.t} + \cdots + d_{10} x_{S10.t} + d_{11} x_{S11.t} + r_t,
\]

\[
r_t = \epsilon_t - \varphi_1 r_{t-1} - \cdots - \varphi_m r_{t-m}.
\]

\[
\ln \left( \sigma_t^2 \right) = \omega + \sum_{i=1}^{q} a_i g(\epsilon_{t-i}) + \sum_{j=1}^{p} \gamma_j \ln \left( \sigma_{t-j}^2 \right),
\]

where \( g(\epsilon_t) = \theta |\epsilon_t| - E(\epsilon_t), \quad \epsilon_t \sim \text{N}(0, 1) \), and \( x_{S1.t} \) is defined as follows:

\[
x_{S1.t} = \begin{cases} 1 & \text{if period } t \text{ is January} \\ 0 & \text{otherwise} \end{cases},
\]

\[
x_{S11.t} = \begin{cases} 1 & \text{if period } t \text{ is November} \\ 0 & \text{otherwise} \end{cases}.
\]

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 \( \hat{z}_t \) and \( \hat{\eta}_t \) 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 \( r'_t = z_t - \hat{z}_t \) be the t-th residual, where \( z_t \) is the observed value and \( \hat{z}_t \) is the predicted value given by the Winters model. Therefore, \( \{r'_t\} \) 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 \( \{r'_t\} \) is highly autocorrelated. To construct the EGARCH model for \( \{r'_t\} \), 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 \( \hat{\sigma}^2_t \) by modeling the residuals \( \{r'_t\} \) 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 \( \{r'_t\} \) with the AR(m)–EGARCH \((p,q)\) model in the second step to produce the estimated heteroscedastic error variance \( \hat{\sigma}^2_t \) for the historical and forecast periods. The WARCH \((m,p,q)\) model is expressed as

\[
\begin{align*}
\text{Step 1:} & \quad (1 - B) (1 - B^2) z_t = (1 - \theta_1 B) (1 - \theta_2 B^5) \\
& \quad r_t, r'_t = z_t - \hat{z}_t, \\
\text{Step 2:} & \quad \phi_t = \gamma + \theta_t \gamma_t - \ldots - \phi_m \gamma_{t-m}, \\
& \quad \epsilon_t = \sigma_t \epsilon_t, \\
& \quad \ln(\hat{\sigma}^2_t) = \omega + \sum_{i=1}^{q} \alpha_i g(\epsilon_{t-i}) + \sum_{j=1}^{p} \gamma_j \ln(\hat{\sigma}^2_{t-j}).
\end{align*}
\]

where \( g(\epsilon_t) = \theta \epsilon_t + |\epsilon_t| - \sqrt{2/i} \), and \( \epsilon_t \sim \text{IN}(0, 1) \).

Both forecasted values of \( \hat{z}_t \) and \( \hat{\sigma}_t \) 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 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

\[
\begin{align*}
Y_{t+1} &= f(y_t, y_{t-1}, \ldots, y_{t-p+1}), \\
Y_{t+2} &= f(y_{t+1}, y_t, \ldots, y_{t-p+2}), \\
Y_{t+3} &= f(y_{t+2}, y_{t+1}, y_t, \ldots, y_{t-p+3}), \\
& \quad \vdots \\
Y_{t+m} &= f(y_{t+m-1}, y_{t+m-2}, \ldots, y_{t+1}, y_t, \ldots, y_{t-p+m}).
\end{align*}
\]

The direct forecasting approach has a mapping function of the form

\[
\begin{align*}
\{Y_{t+m}, Y_{t+m-1}, \ldots, Y_{t+1}\} &= f(y_t, y_{t-1}, \ldots, y_{t-p+1}).
\end{align*}
\]

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, \( \hat{\sigma}_t \) and \( \hat{y}_t \), and a single output node is proposed. The form of the mapping function can be expressed as

\[
\begin{align*}
Y_{t+1} &= f(\hat{\sigma}_{t+1}, \hat{y}_{t+1}), \\
Y_{t+2} &= f(\hat{\sigma}_{t+2}, \hat{y}_{t+2}), \ldots, Y_{t+m} \\
& = f(\hat{\sigma}_{t+m}, Y_{t+m}).
\end{align*}
\]

where \( (\hat{\sigma}_{t+1}, \hat{y}_{t+1}), (\hat{\sigma}_{t+2}, \hat{y}_{t+2}), \ldots, (\hat{\sigma}_{t+m}, Y_{t+m}) \) 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
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:

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - A_i)^2},
\]

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |P_i - A_i|,
\]

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{P_i - A_i}{A_i}\right| \times 100.
\]

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,
- 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 \(\alpha_1\) and \(\beta_1\) coefficients in the following two regression equations:

\[
E_i = \eta_0 + \eta_1 D_{ij} + \omega_i,
\]

\[
E_j = \eta_0 + \eta_1 D_{ij} + \nu_j,
\]

where \(E_i\) and \(E_j\) denote the forecast errors for model \(i\) and model \(j\) \((E_i = P_i - A_i, E_j = P_j - A_j)\), respectively; \(D_{ij}\) denotes the differences between the forecast results \(i\) and \(j\) models \((D_{ij} = P_i - P_j)\), and \(\omega_i\) and \(\nu_j\) are random errors. The null hypothesis is that neither model encompasses (outperforms) the other. If \(\alpha_1\) is significantly different from zero and \(\beta_1\) is not, then the null hypothesis is rejected in favor of the alternative hypothesis that model \(j\) encompasses model \(i\). Conversely, if \(\beta_1\) is significant but \(\alpha_1\) is not, then this is an evidence that model \(j\) encompasses model \(i\). If neither \(\alpha_1\) nor \(\beta_1\) is significant, or conversely if both \(\alpha_1\) and \(\beta_1\) 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
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:

\[
(1 - B)(1 - B^12)z_t = (1 - 0.820B)(1 - 0.671B^{12})\varepsilon_t. \tag{10}
\]

Let \( \tilde{z}_t \) be the point prediction of \( z_t \) and \( \varepsilon'_t = z_t - \tilde{z}_t \) be the \( t \)-th residual, where \( \{\varepsilon'_t\} \) is the detrended and deseasonalized time series with 143 observations. The Augmented Dickey-Fuller (ADF) unit root test can be rejected for the residuals \( \{\varepsilon'_t\} \) 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. 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 heteroscedasticity. The conditional error variance \( \tilde{\sigma}_t \) for \( \{\varepsilon'_t\} \) 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 \( (q = 124) \) model describing the error variance, which can be expressed as

**Step 1**: \( (1 - B)(1 - B^{12})z_t = (1 - 0.820B)(1 - 0.671B^{12})\varepsilon_t, \)

\( \varepsilon'_t = z_t - \tilde{z}_t. \)

**Step 2**: \( \varepsilon'_t = -39783 - 0.148\varepsilon'_{t-13} + 0.183\varepsilon'_{t-18} - 0.330\varepsilon'_{t-23} + 0.041\varepsilon'_{t-24} + \varepsilon_t = \sigma_t\varepsilon_t, \)

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

where \( g(e_t) = 0.042e_t + |e_t| - \sqrt{2/\pi}, \)

\[ e_t \sim IN(0,1). \tag{11} \]

Once estimated, Eq. (11) can be used to compute \( \tilde{z}_t \) and \( \tilde{\sigma}_t \) in the historical and forecast periods from the Winters and ARCH steps, respectively. Both the values of \( \tilde{z}_t \) and \( \tilde{\sigma}_t \) 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(12,8)-SEGARCH \( (q = 1,24) \) is

\[ z_t = 5609782 + 58818t - 243591x_{11,t} - 1049180x_{12,t} - 40242x_{13,t} + 25837x_{14,t} + 788991x_{15,t} + 1284619x_{16,t} + 1902389x_{17,t} + 2334958x_{18,t} + 1684728x_{19,t} + 1498304x_{20,t} + 818588x_{11,t} + \varepsilon_t. \]

\[ \varepsilon_t = \varepsilon_t - 0.230\varepsilon_{t-1} - 0.196\varepsilon_{t-2} - 0.102\varepsilon_{t-8}. \]

\[ \varepsilon_t = \sigma_t\varepsilon_t. \tag{12} \]

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

where \( g(e_t) = 0.477e_t + |e_t| - \sqrt{2/\pi}, \)

\[ e_t \sim IN(0,1) \]

and

\[ x_{11,t} = \begin{cases} 1 & \text{if period } t \text{ is January } \ldots \\ 0 & \text{otherwise} \end{cases} \]

\[ x_{11,t} = \begin{cases} 1 & \text{if period } t \text{ is November } \\ 0 & \text{otherwise} \end{cases} \]

The estimated parameters and estimated values of \( \tilde{z}_t \) and \( \tilde{\sigma}_t \) are obtained simultaneously for the sample and forecast periods. The values of both \( \tilde{z}_t \) and \( \tilde{\sigma}_t \) 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, \( \tilde{z}_t \) and \( \tilde{\sigma}_t \) 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
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 variations, 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:

\[
(1 - B)(1 - B^{12})y_t = (1 - 0.804B)(1 - 0.668B^{12})\rho_t \quad (13)
\]

The ADF unit root test can be rejected for the residuals \(\rho_t^n\) 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 \(\sigma_t^2\) 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:

\[
\begin{align*}
\nu_t, \nu'_t &= y_t - \tilde{y}_t, \\
\ln\left(\sigma_t^2\right) &= 23.061 + 0.356g(e_{t-5}),
\end{align*}
\]

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:

\[
y_t = 1919088 + 8615\rho_t - 97131x_{11.t} - 374934x_{32.t} - 59440x_{33.t} - 134773x_{44.t} - 54148x_{55.t} - 112180x_{66.t} - 54368x_{77.t} - 85199x_{88.t} - 17654x_{99.t} - 62519x_{10.t} - 88123x_{11.t} + r_t,
\]

\[
r_t = -0.208\nu_{t-1} - 0.302\nu_{t-5} + \epsilon_t, \quad \epsilon_t = \sigma_t e_t, \quad (15)
\]

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

\[
g(e_t) = 0.458e_t + |e_t| - \sqrt{2/\pi}, \ e_t \sim IN(0, 1),
\]

\[
x_{11.t} = \begin{cases} 1 & \text{if period } t \text{ is January} \\ 0 & \text{otherwise} \end{cases}
\]

\[
x_{11.t} = \begin{cases} 1 & \text{if period } t \text{ is November} \\ 0 & \text{otherwise} \end{cases}
\]

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
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. The forecasting period is from January 2006 to December 2007. 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 priori 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 \( \tilde{\alpha}_1 \) and \( \tilde{\beta}_1 \) 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 \( (E_3) \) 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

![Fig. 5. Actual and model values for petroleum consumption data.](image)

Table 3

<table>
<thead>
<tr>
<th>Panel A: electricity energy consumption</th>
<th>WARCH</th>
<th>WARCH–ANN</th>
<th>SEGARCH–ANN</th>
<th>SEGARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input nodes</td>
<td>( z_t, \tilde{\alpha}_1 )</td>
<td>( z_t, \tilde{\alpha}_1 )</td>
<td>( z_t, \tilde{\alpha}_1 )</td>
<td>( z_t, \tilde{\alpha}_1 )</td>
</tr>
<tr>
<td>RMSE</td>
<td>643744.33</td>
<td>531545.14</td>
<td>596013.96</td>
<td>824500.08</td>
</tr>
<tr>
<td>MAE</td>
<td>474189.18</td>
<td>404184.25</td>
<td>464632.42</td>
<td>606629.27</td>
</tr>
<tr>
<td>MAPE</td>
<td>2.96%</td>
<td>2.56%</td>
<td>2.98%</td>
<td>3.65%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: petroleum energy consumption</th>
<th>WARCH</th>
<th>WARCH–ANN</th>
<th>SEGARCH–ANN</th>
<th>SEGARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input nodes</td>
<td>( z_t, \tilde{\alpha}_1 )</td>
<td>( z_t, \tilde{\alpha}_1 )</td>
<td>( z_t, \tilde{\alpha}_1 )</td>
<td>( z_t, \tilde{\alpha}_1 )</td>
</tr>
<tr>
<td>RMSE</td>
<td>165753.68</td>
<td>134832.21</td>
<td>148234.91</td>
<td>204369.84</td>
</tr>
<tr>
<td>MAE</td>
<td>134300.13</td>
<td>112542.53</td>
<td>122320.08</td>
<td>167031.13</td>
</tr>
<tr>
<td>MAPE</td>
<td>4.08%</td>
<td>3.71%</td>
<td>3.71%</td>
<td>4.88%</td>
</tr>
</tbody>
</table>

The forecasting period is from January 2006 to December 2007.
Table 4
Encompassing tests of forecasting performance of alternative models.

<table>
<thead>
<tr>
<th>Dependent variable: forecasting errors</th>
<th>Independent variable: forecasting difference from two models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D_{ij}^b$</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Panel A: electricity consumption (forecast period: from January 2006 to December 2007)</td>
<td></td>
</tr>
<tr>
<td>$E_i^a$ (WARCH)</td>
<td>$-1.52^* (-2.98)$</td>
</tr>
<tr>
<td>$E_i^a$ (SEGARCH–ANN)</td>
<td>$-0.48^* (-1.15)$</td>
</tr>
<tr>
<td>$E_i^a$ (WARCH–ANN)</td>
<td></td>
</tr>
<tr>
<td>Panel B: petroleum consumption (forecast period: from January 2006 to December 2007)</td>
<td></td>
</tr>
<tr>
<td>$E_i^a$ (WARCH)</td>
<td>$0.90^* (1.17)$</td>
</tr>
<tr>
<td>$E_i^a$ (SEGARCH)</td>
<td>$-0.10 (-0.21)$</td>
</tr>
<tr>
<td>$E_i^a$ (SEGARCH–ANN)</td>
<td></td>
</tr>
</tbody>
</table>
| Which indicates that statistical significance is at the 0.1 level.
| $D_{ij}^b$ denotes the difference between the forecast from the model $i$ and model $j$.
| $D_{ij}^a$ denotes the forecast error for model $i$.
| $^*$ Which indicates that statistical significance is at the 0.05 level.

In summary, the preponderance of the statistical evidence presented in this paper suggests that the proposed hybrid ANNs forecast generally outperform the four models 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 $\{z_t, \vartheta_s\}$ 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.

Acknowledgments

The author would like to thank the anonymous referees for their valuable suggestions and useful comments.

References


