Prediction of energy's environmental impact using a three-variable time series model

Jeh-Nan Pan\textsuperscript{a,}\textsuperscript{*}, Pin Kung\textsuperscript{a}, Abraham Bretholt\textsuperscript{b}, Jia-Xiang Lu\textsuperscript{a}

\textsuperscript{a}Department of Statistics, National Cheng Kung University, No.1, University Road, Tainan 70101, Taiwan, ROC
\textsuperscript{b}National Dong Hwa University, Hualien 97401, Taiwan, ROC

\textbf{A B S T R A C T}

From the Vienna Convention for the Protection of the Ozone Layer (VCPOL) held in 1985 to the joint United Nations Convention of Climate Change held in 2010, environmental protection has become increasingly urgent. In 2011, the International Standards Organization launched a new energy management system ISO50001 for improving energy efficiency. As an adjunct to these increasingly stringent international agendas, a three-variable time series model is proposed to improve the prediction accuracy of related data. Using a 29 year (1982 to 2010) panel data in Taiwan, the model shows that the environmental impact of increases in CO$_2$ emission is highly correlated to the three leading impact factors, i.e. GDP per capita, renewable energy supplies and coal consumption. The proposed method integrates these three leading impact factors into a highly accurate stationary prediction model. Then, the future environmental impact is evaluated by preforming the trend analysis of CO$_2$ emissions with different growth rate combinations of coal consumption, renewable energy supplies and GDP per capita. Finally, a comparative analysis is performed between our proposed method and the backpropagation neural network (BPN). The comparison results show that the prediction accuracy of our proposed method outperforms BPN in terms of mean absolute percentage error (MAPE) and mean absolute scaled error (MASE).

\textsuperscript{*}Corresponding author. Tel.: +886 6 2757575x53637.
E-mail address: jnpan@mail.ncku.edu.tw (J.-N. Pan).
into products of several factor indices and sets up various weights

to calculate portions of each index. It is widely used in policy

taking to mitigate environmental concern (Ang, 2004; Ang & Choi,

Laspeyres (1864) proposed the Laspeyres Index as a weighted

aggregate index, which used a fixed quantity index for weights in

base period. In the environmental research aspect, Ehrlich and

Holdren (1971) presented the first IPAT formula which includes

impact, population, affluence and technology. It is widely used

for analyzing human activities with regard to their environmental

impacts. Moreover, Kaya (1990) provided a decomposition index

\[ \text{CO}_2 = \frac{(\text{CO}_2 / \text{EF}) \times (\text{TEC/} \text{GD}) \times (\text{GD/P}) \times P}{\text{FSIGP}}, \] (1)

such that FEC denotes fossil energy consumption, and TEC denotes

the main energy consumption. They found that GDP growth in the

developed countries was higher than the other countries and that

\text{CO}_2 emission was decreasing in APEC countries from 1980 to

1998. Recently, Liu and Wu (2011) proposed the indices for energy

use efficiency and \text{CO}_2 emission control efficiency to evaluate the

economic development using developed countries’ \text{CO}_2 emission

percentage data in 2008. Various relationships among GDP per ca-
pita, energy use efficiency and \text{CO}_2 emission control efficiency have

also been explored.

2.2. Time series analysis

Box and Jenkins (1970) suggested an autoregressive moving

average (ARMA) model for reducing time series data for prediction

analysis. The model has been widely used in economics, engineer-
ing, and the social sciences. In this paper, the model is adapted from

ARMA’s close cousin, ARIMA.

Univariate time series analysis uses an ARIMA model. If a variable

\(Y_t\) has \(n\) observations, ARIMA \((p,d,q) (P,D,Q)^\text{d}\) model is given by

\[ \Phi(B)^p \Phi_d(B) (1 - B)^D Y_t = C + \Theta(B)^q \theta_d(B) a_t, \] (2)

where \(C\) is a constant term, \(S\) is the cycle in the seasonal data, \(B\) is a

backward shift operator, \(1 - B\) is the difference, \(\{a_t\}\) is a white noise

process which is a series of independent variables with expected

value 0 and with normally distributed variance \(\sigma^2\). \(P\) and \(p\) belong

to AR, \(D\) and \(d\) belong to I, and \(Q\) and \(q\) belong to MA. The values of \(p,\)

\(P, D, Q, q\) and \(q\) are nonnegative. The difference is used in a non-sta-
tionary time series to transfer the model to a stationary series. Then

the AR and the MA models are identified as the autocorrelation

function (ACF) or the partial autocorrelation function (PACF) deter-

mined by the order of \(p\) or \(q\).

2.3. Transfer function

Let \(X_t\) be the input and \(Y_t\) be the output; thus the relationships

between the input and output variables are shown in Fig. 1.

The input and output variables can be written as

\[ Y_t = \nu(B) X_t + N_t, \] (3)

where the input and output variables are stationary time series pro-

cesses, \(\nu(B) = \nu_0 + \nu_1 B + \cdots + \nu_k B^k\), and the parameter \(\nu(B)\) is an

impulse response function at each time lag. A rational lag structure is

used with the transfer function, such that

\[ \nu(B) = \frac{\omega(B)}{\Phi(B)} = \frac{\omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_k B^k}{1 - b_1 B - b_2 B^2 - \cdots - b_s B^s}. \] (4)

In Eq. (4), the \((r,s,b)\) order is the transfer model, where \(\omega(B)\) is s or-

der for the moving average operator, \(\delta(B)\) is \(r\) order for the auto-

regression operator, and \(b\) is a backshift time lag.

In general, the disturbance term is non-stationary, so that it fits the

ARIMA \((p,d,q)\) model:

\[ N_t = \frac{\Phi(B)}{\phi(B)} a_t, \] (5)

where \(a_t\) is white noise and follows NID \((0, \sigma^2)\). Thus, the transfer

function and the disturbance model can be rewritten as

\[ Y_t = C + \frac{\omega(B)}{\Phi(B)} X_{t-r} + \frac{\Phi(B)}{\phi(B)} a_t. \] (6)

2.4. Artificial neural networks

Artificial neural networks (ANNs) are one of the most popular and

accurate forecasting methods that have a wide range of

\[ \frac{\Phi(B)}{\phi(B)} \]

\[ N_t \]

\[ \text{Disturbance Term} \]

\[ \Phi(B) \]

\[ \phi(B) \]

\[ a_t \]

\[ X_t \]

\[ Y_t \]

\[ \text{Explained Output Part by } X_t \]

\[ \text{White Noise } a_t \]

\[ \text{Disturbance Model } \frac{\Phi(B)}{\phi(B)} \]

\[ \text{Transfer Model } \frac{\omega(B)}{\delta(B)} \]

\[ \text{Input } X_t \]

\[ \text{Output } Y_t \]

Fig. 1. Relationship between input and output variables in transfer function.
applications in economic, engineering and social problems. Warren McCulloch and Walter Pitts developed the first conceptual model of an ANN in 1943. ANN is a mathematical model for simulating the structure of biological neural networks, which are considered a complex nonlinear dependence between the inputs and outputs. It is also an interconnected network of nodes (neurons) that have billions of simple processing units. The inputs are weighted, thus every input value is multiplied with individual weight. The artificial neuron is function sum that sums all weighted inputs and a bias. The output of a neuron is a function of the weighted sum of the inputs plus a bias. The activation functions (or transfer function) are used to the weighted sum of the inputs of a neuron to produce the outputs, thus the artificial neuron passes the processed information via outputs. According to the above fundamental concepts, various researchers have developed different approaches for improving the system performance (Krenker, Bešter, & Kos, 2011; Jain, Mao, & Mohiuddin, 1996; Laguna & Martí, 2002; Nuchitprasit & Srisai and Bijari (2010) further proposed an ANN model using ARIMA, and their outputs showed that hybrid model is more accurate than ANN. However, they did not consider the causal relationship between input (GDP per capita, coal and new energy supplies, etc.) and output (CO₂ emission) variables in measuring the forecasting performance of their data sets.

3. Research methodology

3.1. Cross correlation function (CCF)

To improve the prediction accuracy of the model developed here, two well-known forecasting methods are employed: the cross correlation function (CCF) and the linear transfer function (LTF) (Lin, 2006). The idea is to combine the techniques of time series and regression analysis to obtain a dynamic regression model which contains more information about the explainable variables. The CCF, developed by Box and Jenkins (1970), is very useful in establishing the transfer function model and the dynamic regression model. The CCF model construction procedure is described as follows.

Step 1. CCF model construction

(1) Build an ARIMA model and consider an input variable \(X_t\), and retain the residual series \(\{e_t\}\). The prewhitened input series can be expressed as

\[ x_t = \frac{\phi(B)}{\theta(B)} X_t. \]  

(2) Use the white noise input variable ARIMA model. The filtered output series can be expressed as

\[ \beta_t = \frac{\phi(B)}{\theta(B)} Y_t. \]  

(3) Calculate the sample CCF \( \hat{\rho}_{xy}(k) \) of \( \{x_t\} \) and \( \{\beta_t\} \), and estimate the impulse response function \( \hat{\gamma}_k \)

\[ \hat{\gamma}_k = \frac{\hat{\sigma}_y}{\hat{\sigma}_x} \hat{\rho}_{xy}(k), \]  

where \( \hat{\rho}_{xy}(k) = E[(x_t - \bar{x})(\beta_{t+k} - \bar{\beta})]/\hat{\sigma}_x \hat{\sigma}_\beta = \hat{\gamma}_y(k)/\hat{\sigma}_x \hat{\sigma}_\beta, \) \( k = 0, 1, \ldots, 2, \ldots, \gamma_{xy}(k) \) is the cross variance of \( \{x_t\} \) and \( \{\beta_t\} \) series with time lag \( k \).

(4) Use the \( \hat{\gamma}_k \) form to fit the theoretical graphs and to determine a suitable \( r \) and \( s \) values, and time lag \( b \) value.

(5) Fit the ARIMA model with disturbance term using the univariate model method.

Step 2. Model estimation

If the temporary model is

\[ y_t = C + \frac{\omega(B)}{\phi(B)} X_{t-b} + \frac{\theta(B)}{\phi(B)} e_t, \]  

where the estimates include \( \hat{\omega} = \{\omega_0, \ldots, \omega_s\}, \hat{\phi} = \{\phi_1, \ldots, \phi_s\}, \hat{\theta} = \{\theta_1, \ldots, \theta_s\} \) and \( \sigma_e^2 \). The conditional likelihood function of \( e_t \) is

\[ L(\hat{\gamma}, \hat{\omega}, \hat{\phi}, \hat{\theta}, \sigma_e^2 | x, y) = (2\pi\sigma_e^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma_e^2} \sum_{t=1}^{n} e_t^2 \right]. \]  

The parameters can be obtained by nonlinear least squares method, that is

\[ \min S(\hat{\gamma}, \hat{\omega}, \hat{\phi}, \hat{\theta}, \sigma_e^2) = \sum_{t=1}^{n} e_t^2. \]  

where \( t_0 = \max(p + r + 1, b + p + s + 1) \).

Step 3. Model diagnosis

In the transfer function model, suppose that the white noise series \( \{e_t\} \) and the input variable \( X_t \) are mutually independent. After obtaining parameter estimates, two tests are needed:

(1) The self-autocorrelation test determines a suitable model for the temporary model of disturbance term. For an appropriate model, the residual term of the sample correlation function has no patterns.

(2) The cross correlation test determines \( \{x_t\} \) and \( X_t \) which should be mutually independent. For an appropriate model, \( \{x_t\} \) and \( \{x_t\} \) of the sample correlation coefficient \( \rho_{xy}(k) \) lies within 2 standard deviation, and the residual term has no pattern.

3.2. Linear transfer function (LTF)

LTF was proposed by Liu and Hanssens (1982) to solve the drawbacks of the CCF. If the input has many variables, then LTF can be written as

\[ Y_t = C + \upsilon(B) X_t + \frac{\theta(B)}{\phi(B)} e_t, \]  

where \( \upsilon(B) \) can be expressed as

\[ \upsilon(B) = \upsilon_0 + \upsilon_1 B + \cdots + \upsilon_b B^b. \]  

LTF estimates the impulse response weights using the least squares method. Since the whole roots of \( \delta(B) \) are out of the unit circle, \( \upsilon(B) \) can be approximated as \( \omega(B)/\delta(B) \). A decision flowchart for constructing the LTF model is given below in Fig. 2.

The LTF model construction procedure is described as follows:

Step 1. LTF model construction

(1) Choose an efficient order \( k \) value and a reasonable disturbance term \( N_t \), and estimate a linear transfer model. The general economic series has 5 orders, since Liu (1991) pointed out that general time series data has autocorrelation. Thus, it is more efficient that the disturbance term is given as AR (1) or AR (2) to estimate the impulse response weight. If the data is a non-seasonal series, the disturbance term can be temporarily given as AR (1),

\[ N_t = \frac{1}{1 - \phi B} e_t. \]  

If the data is a seasonal series, the disturbance term can be temporarily written as

\[ N_t = \frac{1}{1 - \phi B}. \]
According to Johnston (1984), the estimated result of the asymptotic bias is

$$\lim(\hat{\beta} - \beta) = \frac{1}{\sum_{i=1}^{p} \sum_{j=1}^{q} \gamma_{ij}}$$

(19)

where $$\sum_{ij} = \gamma_{ij}(0) + \gamma_{ij}(1) + \cdots + \gamma_{ij}(k)$$, and $$\gamma_{ij}(k)$$ is the cross variance of $$\{x_i\}$$ and $$\{x_j\}$$ series with $$k$$ time lag

$$\sum_{ij} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{bmatrix}$$

(20)

where $$\gamma_i$$ is the cross covariance of $$\{x_i\}$$ and $$\{x_{i-1}\}$$. Step 3. Model diagnosis

The diagnosis step is the same as CCF.

3.3. Model assessment criteria

Besides ACF and PACF, many well-known criteria can be used for assessing the model fit. Akaike (1974) provided his information criterion (AIC). However, Hurvich and Tsai (1989) described AIC as having a serious over-fit problem due to small sample sizes. Hurvich and Tsai (1989) thus devised a bias-corrected Akaike Information Criterion (AICc). Schwarz (1978) also introduced a bayesian criterion (SBC). These three equations are given below:

1. AIC criteria:

$$\text{AIC}(k) = -2 \ln(L) + 2k$$

(21)

2. AICc criteria:

$$\text{AICc}(k) = \text{AIC}(k) + \frac{2k(k+1)}{n-k-1}$$

(22)

3. SBC criteria:

$$\text{SBC}(k) = -2 \ln(L) + k \ln(n)$$

(23)

where $$n$$ denotes observation, $$k$$ denotes parameter in the model, and $$L$$ denotes method of maximum likelihood estimate. The mean absolute percentage error (MAPE) shown in Eq. (24) can be used as an evaluation criterion as discussed by Render, Stair, and Hanna (1999) and Pan and Chen (2010).

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\%$$

(24)

where $$Y_i$$ is the real value, $$\hat{Y}_i$$ is the predicted value, and $$n$$ is the predicted period. In addition, the following mean absolute scaled error (MASE) proposed by Hyndman and Koehler (2006) is used as another evaluation criterion since it has been well-recognized as the standard measure for comparing prediction accuracy across multiple time series.

$$\text{MASE} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - F_i}{\frac{1}{n-1} \sum_{j=2}^{n} |Y_{i-j} - Y_{i-1}|} \right)$$

(25)

where $$Y_i$$ denotes the actual value at time $$t$$, $$F_i$$ denotes the forecast value, and $$F_{i-1}$$ equals $$Y_{i-1}$$.

Note that the denominator is the average forecast error of the one-step “naive forecast method”, which uses the actual value from the prior period as the forecast. The prediction accuracy is considered adequate if the MASE value is less than one.
4. The state of energy usage and CO₂ emission in Taiwan

The data source is the International Energy Agency website and IEA Statistics (IEA, 2011). According to IEA report, the Global CO₂ emissions reached the highest record in 2010. For a global temperature increase of less than 2 Celsius degree by 2020, worldwide CO₂ emissions per capita must remain less than 32 gigatons per year. CO₂ emissions per capita in Taiwan are proportionally higher than this target since its primary energy consumption is oil followed closely by coal. In 2009, CO₂ emissions were 10.89 tons/capita, much greater than the world production level of 4.29 tons/capita. On the other hand, the Taiwan Bureau of Energy (2010) reported that the Energy Density (energy consumption/GDP) has decreased from 10.14 LEO (liter of oil equivalent) per thousand New Taiwan Dollars (NT$1000) in 2001 to 8.46 in 2010. Since energy productivity is the inverse of Energy Density, energy productivity is increasing.

Energy productivity = 1/Energy density
\[ = \text{GDP/Energy consumption}. \quad (26) \]

This is a conundrum for sustainable development since it is a positive indicator, even though the environment is deteriorating. Hence, a time series model of energy consumption can reveal its future effect on GDP. This approach has been lacking in the research and can be a useful adjunct to the ISO specification and in policy making.

Using Lee and Oh (2006) decomposition formula, the rate of change of CO₂ emissions is given by:
\[ \Delta C_t = C_t - C_{t-1} = F_t S_t I_t G_t P_t - F_{t-1} S_{t-1} I_{t-1} G_{t-1} P_{t-1}, \]
\[ \Delta C_t = \Delta C_{ \text{eff}} + \Delta C_{ \text{S-eff}} + \Delta C_{ \text{I-eff}} + \Delta C_{ \text{G-eff}} + \Delta C_{ \text{P-eff}}. \quad (27) \]

where C denotes CO₂ emission, F denotes CO₂ emission per fossil energy consumption (FEC), S denotes fossil energy consumption per total energy consumption (TEC), I denotes energy consumption per GDP or Energy Intensity, G denotes GDP per capita, and P denotes population. This notation is consistent with Eq. (1) as the factor effects extend and standardized that idea. The five factor effects are F-effect, S-effect, I-effect, G-effect and P-effect are given by:
\[ \Delta C_{ \text{F-effect}} = L(C_t, C_{t-1}) \ln(F_t/F_{t-1}), \quad \Delta C_{ \text{S-effect}} = L(C_t, C_{t-1}) \ln(S_t/S_{t-1}), \]
\[ \Delta C_{ \text{I-effect}} = L(C_t, C_{t-1}) \ln(I_t/I_{t-1}), \quad \Delta C_{ \text{G-effect}} = L(C_t, C_{t-1}) \ln(G_t/G_{t-1}), \]
\[ \Delta C_{ \text{P-effect}} = L(C_t, C_{t-1}) \ln(P_t/P_{t-1}), \]

where
\[ L(C_t, C_{t-1}) = (C_t - C_{t-1})/\ln(C_t/C_{t-1}). \]

From Table 1, from 1995 to 2010, the following effects were decreasing: CO₂ from fossil fuels, \( \Delta C_{ \text{F-effect}} \), the proportion of fossil fuels in the energy mix, \( \Delta C_{ \text{S-effect}} \), and Energy Intensity, \( \Delta C_{ \text{I-effect}} \). That is, even though the efficiency of using fossil fuel is increasing and its proportional use is decreasing, CO₂ emission increase is still rampant. Thus, CO₂ emission increase is due to the effect of GDP and population increase.

The LMDI approach can also be used to show the annual structure of CO₂ emission as shown in Fig. 3(a). Carbon emission decreases from 2007 to 2010 largely due to the lagged GDP effect shown in Fig. 3(b). Hence, the time series model can give appropriate additional information for energy policy makers.

5. Data analysis

5.1. Data information

The Taiwan energy supply relies heavily on importation of high carbon fuels. Similar to other developed countries, Taiwan’s CO₂ emission per capita is still increasing, and therefore retreating from any world CO₂ emission standard. Hence, this research attempts to build key sustainability indicators (KSI) by exploring the relationship between CO₂ emission and the use of both traditional and green energy sources. According to Kaya (1990), the decomposition formula for the environmental impact of new and old energy sources is
\[ I = \frac{P \times A \times E_1}{E_2}, \quad (28) \]

where \( I \) represents CO₂ emission and is a function of \( P \), population, \( A \), GDP, \( E_1 \), old energy, and \( E_2 \), new energy (i.e. renewable energy).

Here oil and coal are considered old energy. Though hydro resources account for most of Taiwan’s green energy supply it is already fully exploited with little growth potential. On the other hand, biomass energy is hardly used at all. Therefore, only solar energy and wind power are considered as new energy in this research. Based on the prediction from World Energy Council (WEC), the renewable energy needs to reach to 30% of the world energy supplies in 2020 if we want to achieve the goal of environmental sustainability. The past 29 years (from 1982 to 2010) panel data was used with the first 19 years (1982 to 2000) used to “train” the model. The last 10 years are reserved for testing the prediction accuracy using mean absolute percentage error (MAPE). The data include CO₂ emission (million tons), GDP per capita, oil, coal and the new energy supplies (kiloliter of oil equivalent, KLOE) series.

Generally, time series analysis is assumed to be stationary. The stationary null hypothesis (H₀: time series is stationary) was tested using a new unit root test proposed by Phillips and Perron (1988), called the Phillips–Perron (PP) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (1992). The result of the unit root test shows that the initial variables are non-stationary (Table 2). After performing the second-order difference (\( \nabla^2 = (1 – B^2) \)), the results show that all variables were stationary at \( \alpha = 0.1 \) level of significance.

5.2. Autocorrelation analysis

This study uses ten years data (2001–2010) for testing, which is a subset of the original 29 year dataset. MAPE and MASE are used to measure the accuracy of the predicted model. First, CO₂ emission was tested for autocorrelation. The initial series had a trend term. By fitting the MA (1) model after second-order difference, the temporary CO₂ prediction model follows ARIMA (0,2,1), as given by
\[ \nabla^2 Y_t = (1 + 0.6988) \alpha_t, \quad t = 1.2, \ldots, 19, \quad (29) \]
\[ \alpha_t = 5.22, \quad \text{MAcc} = 107.34 \quad \text{and SBC} = 108.04. \]
The residual model of ACF and PACF plots (Fig. 4(a)) are 2 standard deviations apart, thus the residuals are random. The Box-Pierce Q test is 0.414 with \( p \)-value = 0.55 indicating the residuals are mutually independent. The Shapiro–Wilk test is 0.308 so the residuals follow the normality assumption. Fig. 4(b) shows that the predicted value has an upward trend. That is, when the predicted period is increasing, the difference between the predicted and actual values is increasing. However, the MAPE value of 16.53% indicates that the prediction accuracy is poor using ARIMA (0,2,1) model.

### 5.3. Transfer function

#### 5.3.1. Univariate model

Since GDP per capita is considered as an input variable for predicting CO₂ emission, we use CCF to identify the model. After pre-whitening, GDP per capita follows ARIMA (1,2,0). In Fig. 5, the ACF value falls more than 2 standard deviations away when the time lag is 2, and there is no special pattern. This indicates that GDP per capita lags 2 periods behind CO₂ emission after taking the second-order difference. Thus, we temporarily set up the model as

\[
\nabla^2 Y_t = \alpha_0 \nabla^2 X_{t-2} + N_t, \tag{30}
\]

where \( N_t \) is a disturbance term. Using the nonlinear least squares method, the transfer function is
\( \nabla^2 Y_t = 6.56 \times \nabla^2 X_{g(t-2)} + \alpha_t \) \hspace{1cm} (31)

\( \hat{\sigma}_g = 4.82, \text{ AICc} = 91.97 \text{ and SBC} = 92.66. \)

In Fig. 6, the residuals of ACF and PACF fall within 2 standard deviations, thus the disturbance term is accepted. All CCF values are not significant in Fig. 7, so the transfer function is accepted as well. For predicting CO₂ emission using Eq. (31), the predicted values are close to the actual values in the first three years and then these values diverge (see Fig. 8 for details). This prediction model is also considered not appropriate since its MAPE equals 15.5% and its MASE equals 4.26 (see Table 8 for details).

5.3.2. Multivariate model

We first determine the impulse response order using LTF. The disturbance term is assumed to be an AR (1) process. New energy supplies \( \{X_n\} \) plus the sum of oil and coal supplies forms the aggregated fossil energy supply variables \( \{X_{all}\} \). The unit is “year” and the sample size is small. To avoid lack of sufficient degrees of freedom, we use a third-order impulse response function:

\[
\nabla^2 Y_t = C + (t_{10} + t_{11}B + t_{12}B^2 + t_{13}B^3) \nabla^2 X_{all} + (t_{20} + t_{21}B + t_{22}B^2 + t_{23}B^3) \nabla^2 X_n + \frac{1}{1 - \phi_1 B} \alpha_t.
\] \hspace{1cm} (32)

The estimates of the impulse response function for the aggregated fossil and new energy supply series \( \{X_{all}\} \) is listed in Table 3.

The transfer function and its theoretical graph are shown in Table 4 in which \((r,s,b) = (2,1,2)\). The first period parameters \( (t_{1i}) \) of \( \{X_{all}\} \) and the current period \( (t_{2i}) \) of \( \{X_n\} \) are not significant \((|t| < 2)\). The parameter estimates after excluding non-significant values are

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard deviation</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_{10} )</td>
<td>2.24</td>
<td>0.66</td>
<td>3.41*</td>
</tr>
<tr>
<td>( t_{11} )</td>
<td>2.46</td>
<td>0.84</td>
<td>2.93*</td>
</tr>
<tr>
<td>( t_{12} )</td>
<td>2.38</td>
<td>0.66</td>
<td>3.64*</td>
</tr>
<tr>
<td>( t_{13} )</td>
<td>-3.24</td>
<td>0.74</td>
<td>-4.38*</td>
</tr>
<tr>
<td>( t_{20} )</td>
<td>5.64</td>
<td>0.79</td>
<td>7.18*</td>
</tr>
<tr>
<td>( t_{21} )</td>
<td>-1.86</td>
<td>0.87</td>
<td>-2.15*</td>
</tr>
<tr>
<td>( t_{22} )</td>
<td>4.10</td>
<td>1.10</td>
<td>3.72*</td>
</tr>
<tr>
<td>( t_{23} )</td>
<td>-1.34</td>
<td>0.87</td>
<td>-1.54</td>
</tr>
<tr>
<td>( \phi_1 )</td>
<td>0.54</td>
<td>0.14</td>
<td>3.73</td>
</tr>
</tbody>
</table>

* Denotes \(|t| > 2\).
listed in Table 5. Since the new energy supply series \( \{X_n\} \) is significant from period 0 to period 2 with no special patterns, we temporarily set up the model as

\[
\nabla^2 Y_t = \left( \frac{\omega_{10} B}{1 - \delta_1 B - \delta_2 B^2} \right) \nabla^2 X_{all} + \left( \omega_{20} + \omega_{21} B \right) \nabla^2 X_{a} + \frac{1}{1 - \theta_1 B} a_t,
\]

where the disturbance term is determined to be AR (1).

Moreover, \( \delta_1 \) is 3.18, AICc is 104.66 and SBC is 98.14 for this prediction model. After performing model diagnosis, ACF and PACF fall within 2 standard deviations. The residuals are calculated using CCF, and they are not significant. Thus, this prediction model is considered appropriate.

Next, we consider three input variables, GDP per capita, coal supplies and new energy supplies in the impulse response function. Given a second-order impulse response function, the estimates and statistics for this impulse response function are listed in Table 6.

The statistics in Table 6 indicate that the parameters of current and second period impulse response function for GDP per capita series \( \{X_t\} \) are significant. Moreover, the parameters of current, first and second period impulse response function for the coal supply series \( \{X_t\} \) are significant, and the parameter of the current impulse response function for new energy supplies series \( \{X_n\} \) are also significant. Thus, the temporary model is:

\[
\nabla^2 Y_t = (\omega_{10} + \omega_{12} B) \nabla^2 X_{c} + (\omega_{20} + \omega_{21} B + \omega_{22} B^2) \nabla^2 X_{n} + \frac{1}{1 - \theta_2 B} a_t.
\]

In Eq. (33), the disturbance term is determined as AR (2). The parameter estimates and their statistics for LTF are listed in Table 7.

Moreover, \( \delta_2 \) is 2.14, AICc is 104.26, and SBC is 97.46 for this prediction model. After performing the model diagnosis, ACF and PACF fall within 2 standard deviations and the residuals for CCF are not significant. Thus, this prediction model is considered appropriate.

The predicted results of the two-variable time series model are shown in Fig. 9. The aggregated fossil energy (oil and coal) and renewable energy supply data are the input variables. Although its MAPE value has been reduced to 8.8% and the difference between the predicted and actual values of carbon emission is getting smaller, the prediction model is not considered to be adequate since its MASE value equals 2.40 (>1) as indicated in Table 8.

The predicted result of the three-variable time series is shown in Fig. 10. Given MAPE value equals 2.8% and MASE value equals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard deviation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega_{10} )</td>
<td>3.64</td>
<td>1.61</td>
<td>2.26*</td>
</tr>
<tr>
<td>( \omega_{11} )</td>
<td>-1.18</td>
<td>1.07</td>
<td>-1.10</td>
</tr>
<tr>
<td>( \omega_{12} )</td>
<td>3.69</td>
<td>1.66</td>
<td>2.23*</td>
</tr>
<tr>
<td>( \omega_{20} )</td>
<td>0.18</td>
<td>0.05</td>
<td>3.66*</td>
</tr>
<tr>
<td>( \omega_{21} )</td>
<td>-0.11</td>
<td>0.04</td>
<td>-3.00*</td>
</tr>
<tr>
<td>( \omega_{22} )</td>
<td>0.12</td>
<td>0.05</td>
<td>2.47*</td>
</tr>
<tr>
<td>( \omega_{30} )</td>
<td>-0.64</td>
<td>0.31</td>
<td>-2.06*</td>
</tr>
<tr>
<td>( \omega_{31} )</td>
<td>-0.15</td>
<td>0.45</td>
<td>-0.33</td>
</tr>
<tr>
<td>( \omega_{32} )</td>
<td>0.63</td>
<td>0.32</td>
<td>1.96</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>-1.64</td>
<td>0.79</td>
<td>-2.02</td>
</tr>
</tbody>
</table>

* Denotes \( |t| > 2 \).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard deviation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega_{10} )</td>
<td>2.46</td>
<td>1.19</td>
<td>2.07</td>
</tr>
<tr>
<td>( \omega_{12} )</td>
<td>3.74</td>
<td>1.31</td>
<td>2.86</td>
</tr>
<tr>
<td>( \omega_{20} )</td>
<td>0.17</td>
<td>0.04</td>
<td>4.23</td>
</tr>
<tr>
<td>( \omega_{21} )</td>
<td>0.13</td>
<td>0.03</td>
<td>4.33</td>
</tr>
<tr>
<td>( \omega_{22} )</td>
<td>-0.12</td>
<td>0.03</td>
<td>-4.00</td>
</tr>
<tr>
<td>( \omega_{30} )</td>
<td>-0.39</td>
<td>0.19</td>
<td>-2.05</td>
</tr>
<tr>
<td>( \omega_{32} )</td>
<td>0.62</td>
<td>0.29</td>
<td>2.14</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>0.63</td>
<td>0.23</td>
<td>2.74</td>
</tr>
</tbody>
</table>

Fig. 9. \( \text{CO}_2 \) emission forecasting using LTF (Considering aggregated fossil and new energy supplies).

Fig. 10. \( \text{CO}_2 \) emission forecasting using LTF (Considering GDP per capita, coal and new energy supplies).
Emissions in Taiwan for 2011 were 249.2 million tons, and our proposed method exhibited lower emissions in Taiwan.

Thus, a similar reduction method is required and the training/testing data set is small. Hence, it can be used as an alternate model for GDP per capita, coal and new energy supplies. Therefore, this multiple time series model can be served as a useful reference in predicting coal's environmental impact in other countries.

| Table 8 | The comparison of MAPE and MASE values for predicting CO2 emissions using BPN and our proposed method. |
|---|---|---|
| Considering GDP per capita | Considering aggregated fossil and new energy supplies | Considering GDP per capita, coal and new energy supplies |
| BPN | MAPE (%) | 12.6% | 14.37% | 11.9% |
| | MASE | 3.47 | 3.97 | 3.30 |
| Our proposed method | MAPE (%) | 15.5% | 8.8% | 2.8% |
| | MASE | 4.26 | 2.40 | 0.71 |

Table 9: Carbon emission predictions for combinations of R_c, R_n, R_g.

<table>
<thead>
<tr>
<th>(R_c, R_n, R_g)</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3%, 10%, 1%)</td>
<td>263.76</td>
<td>300.94</td>
<td>309.10</td>
<td>319.02</td>
<td>321.68</td>
<td>320.06</td>
</tr>
<tr>
<td>(3%, 10%, 2%)</td>
<td>266.73</td>
<td>306.29</td>
<td>312.16</td>
<td>337.17</td>
<td>346.65</td>
<td>352.23</td>
</tr>
<tr>
<td>(3%, 10%, 3%)</td>
<td>269.38</td>
<td>311.69</td>
<td>322.98</td>
<td>355.73</td>
<td>372.46</td>
<td>385.72</td>
</tr>
<tr>
<td>(3%, 15%, 1%)</td>
<td>260.91</td>
<td>293.71</td>
<td>303.02</td>
<td>313.49</td>
<td>317.45</td>
<td>317.60</td>
</tr>
<tr>
<td>(3%, 15%, 2%)</td>
<td>263.56</td>
<td>299.11</td>
<td>314.71</td>
<td>331.73</td>
<td>342.55</td>
<td>349.91</td>
</tr>
<tr>
<td>(3%, 15%, 3%)</td>
<td>266.17</td>
<td>304.47</td>
<td>326.43</td>
<td>350.14</td>
<td>368.18</td>
<td>383.18</td>
</tr>
<tr>
<td>(3%, 20%, 1%)</td>
<td>257.34</td>
<td>286.17</td>
<td>294.92</td>
<td>305.9</td>
<td>310.21</td>
<td>311.01</td>
</tr>
<tr>
<td>(3%, 20%, 2%)</td>
<td>260.31</td>
<td>291.53</td>
<td>306.97</td>
<td>324.04</td>
<td>335.18</td>
<td>343.18</td>
</tr>
<tr>
<td>(3%, 20%, 3%)</td>
<td>262.96</td>
<td>296.93</td>
<td>318.79</td>
<td>342.59</td>
<td>360.97</td>
<td>376.66</td>
</tr>
<tr>
<td>(3%, 25%, 1%)</td>
<td>254.49</td>
<td>278.30</td>
<td>286.57</td>
<td>316.5</td>
<td>324.61</td>
<td>331.77</td>
</tr>
<tr>
<td>(3%, 25%, 2%)</td>
<td>257.14</td>
<td>283.70</td>
<td>298.26</td>
<td>314.04</td>
<td>324.61</td>
<td>331.77</td>
</tr>
<tr>
<td>(3%, 25%, 3%)</td>
<td>259.75</td>
<td>289.06</td>
<td>309.98</td>
<td>332.81</td>
<td>350.24</td>
<td>365.04</td>
</tr>
<tr>
<td>(3%, 30%, 1%)</td>
<td>252.12</td>
<td>276.09</td>
<td>283.88</td>
<td>302.03</td>
<td>309.58</td>
<td>313.46</td>
</tr>
<tr>
<td>(3%, 30%, 2%)</td>
<td>255.54</td>
<td>280.87</td>
<td>299.97</td>
<td>320.59</td>
<td>335.36</td>
<td>346.94</td>
</tr>
</tbody>
</table>

R_c, R_n, R_g denote growth rates of coal energy, new energy, GDP per capita respectively. Unit: million tons.

0.71 (<1) as indicated in Table 8, this prediction model is considered the most appropriate one. It is more effective to use three input variables, i.e. GDP per capita, coal supplies and renewable energy supplies in predicting CO2 emission.

5.4. The comparison of MAPE and MASE values using BPN and our proposed method

Using MATLAB (2012b)’s neural network toolbox (i.e. nntool), we consider 10 neurons with tangent sigmoid transfer function for hidden layer, 1000 epochs for training parameters and a Levenberg–Marquardt training function in the backpropagation neural network (BPN). The comparison results of MAPE and MASE values for predicting CO2 emissions using BPN and our proposed method are summarized in Table 8. Apparently, the prediction accuracy of BPN is not appropriate since its MAPE and MASE values are too high and unstable (while the MAPE and MASE values of our proposed method show a downward trend). Due to the fact that the MASE value of the three-variable time series model (considering GDP per capita, coal and new energy supplies) equals 0.71, the prediction accuracy of our proposed method is considered the most appropriate one. Hence, it can be used as an alternate model for forecasting task especially when the prediction accuracy is required and the training/testing data set is small.

6. Discussions and conclusions

A key sustainability indicator (KSI) for worldwide environmental impact is the trend of CO2 emission under various energy and economic inputs. The prediction accuracy for this trend is greatly enhanced when a three-variable time series model is employed. In contrast to the dependent residual terms occurred in traditional multivariate regression analysis, the residual terms for the three-variable time series model becomes independent. The impact of coal, new energy supplies, and GDP are reduced to a stationary, integrated movement in the prediction of CO2 emissions in Taiwan. The poor prediction accuracy of BPN, one-variable and two-variable time series models as demonstrated suggest this to be the best forecast method.

Here are some important statistics that support the three-variable prediction model in Taiwan for the study period (1995 to 2010):

1. Oil consumption has decrease from 55.2% to 43%.
2. Coal consumption has increased from 24.5% to 33%.
3. Coal pricing was less per BTU (British thermal unit) than oil.
4. Coal pricing was less volatile than oil.
5. Coal supply increased about 8%, and
6. Oil supply decreased about 5%.

Also supporting the three-variable time series model was the observation that oil supply did not improve prediction accuracy whereas coal supply was more significant than oil supply in predicting CO2 emission change. Thus, a similar reduction method for time series analysis could be used for various other countries as an adjunct to the deployment of the new energy management system ISO 50001.

Policy makers can also benefit by more accurate predictions. For example, the Taiwan government suggests that CO2 emission should be reduced to the 2008 level (252 million tons) between 2016 and 2020 whereas emissions by the year 2025 should be reduced to the 2000 level (215.5 million tons). An intermediate prediction timeframe is illustrated below to forecast carbon emissions from 2011 to 2016 wherein the growth rate of coal, R_c, is fixed at 3%, GDP per capita, R_g, varies from 1% to 3% and new energy, R_n, ranges from 10% to 30%. These values are based on recent statistics for 2010–2011: GDP per capita increased by 0.85%, new energy supplies increased by 26%, and coal usage by increased 2.7%. The predicted results are shown in Table 9.

Actual CO2 emissions in Taiwan for 2011 were 249.2 million tons, very close to the predicted value at (3%,25%,1%) of 254.49 million tons. The 98% prediction accuracy is considered adequate. Although the best case scenario by 2016 (281.29 million tons at 3%,30%,1%) will not meet the Kyoto Protocol standard, the government of Taiwan can provide incentives, such as levying carbon tax or subsidizing low carbon energy to the enterprises for improving their energy utilization efficiency.

In this paper, we have added BPN as a competing model and also shown that the prediction accuracy of our proposed method outperforms BPN in terms of MAPE and MASE especially when the testing data set is small. Similar approach can be applied to other energy time series in other parts of the world if both the input and output data sets are available. Thus, this multiple time series approach can be served as a useful reference in predicting energy’s environmental impact in other countries.

Acknowledgements

The authors would like to gratefully acknowledge financial supports from the National Science Council of Taiwan (NSC 101-2410-H-006-016).