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

Jeh-Nan Pan, Pin Kung, Abraham Bretholt, Jia-Xiang Lu

Abstract

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 CO2 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 CO2 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 that the prediction accuracy of our proposed method outperforms BPN in terms of mean absolute percentage error (MAPE) and mean absolute scaled error (MASE).

1. Introduction

Due to population growth and today’s high-tech expectations, humanity is confronted with the contentious issue of limiting energy consumption especially since most countries are facing scarcity of oil supplies and serious environmental pollution. Fossil fuels are the main type used worldwide, but produce about 70% of greenhouse gases. In the past twenty years, world climate change has become increasingly more serious; hence, a period of multilateral environmental agreements has stretched from 1985 to the present, but these have been ignored by the major polluting entities. The International Standards Organization has also launched its latest energy standard ISO50001 in 2011 with the purpose to establish a global standardized energy management system, the three-variable time series model proposed here could be a useful adjunct to the new ISO 50001 specification.

Increased environmental awareness has made its protection an international concern. Thus, green energy consciousness has shifted awareness to the elimination of carbon emissions altogether through the development of new energy sources. Declining oil supplies has resulted in higher oil prices since the invasion of Iraq and oil prices will tend to rise more steeply as the end of the global oil supply is recognized as an urgent concern. Of course, these price hikes are harmful to economic development in most countries that are not oil producers and this makes it necessary for countries like Taiwan to reduce its reliance on fossil fuels. Consequently, the rise and development of renewable energy supplies to reduce price vulnerability as well as carbon emissions is tantamount to lifestyle sustainability in a highly developed nation such as Taiwan.

Hence, the purpose of this research is to predict CO2 emission based on the existing energy usage patterns in Taiwan and to explore the relationship between renewable (new) energy and old energy (oil and coal) in detail. The study considers different combinations of growth rate in coal consumption and the development of new energy sources while assuming a modest GDP increase. The results are indicative of other industrial nations in the region and the proposed forecasting method can be similarly applied to any other country. It can also serve as a standardized model for certain aspects of the ISO50001 energy management system or for governmental policy creation.

Keywords: Multivariate time series analysis, GDP per capita, Renewable energy supplies, CO2 emission, Backpropagation neural network, MAPE, MASE

© 2013 Published by Elsevier Ltd.
2. Literature review

2.1. Index decomposition method

An indexed decomposition method is the earliest approach for analyzing CO2 emission. This method separates CO2 emission into products of several factor indices and sets up various weights to calculate portions of each index. It is widely used in policy making to mitigate environmental concern (Ang, 2004; Ang & Choi, 2003; Ang & Liu, 2007; Ang & Pandiani, 1997; Ang & Zheng, 2000).

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 (CO2 = [CO2 Eaton] × [E/GDP] × [GDP/P]) using the IPAC formula, where GDP denotes gross domestic product, P denotes population and E denotes energy consumption. The later CO2 decomposition methods are based on Kaya’s decomposition model.

Boyd, Hanson, and Sterner (1988) provided a Simple Average Divisia (SAD) used to average the energy consumption (E) in year 0 and year t given certain decomposing factor weights. It also uses a logarithm method for computing the annual increment of input factor (X). Thus, the index change can be computed as (E0 - E1) / ln(E1 / E0). This formula can also extend to compute the change of Energy Intensity (I = E/GDP). However, zero values in the data can lead to discontinuities, later solved by Ang and Lee (1994). Liu, Ang, and Ong (1992) provided an adaptive weighting divisia (AWD) that uses a discrete differentation function for decomposing factor rates of changes.

The AWD is the preferred method to reduce residuals, whereas the Laspeyres Index, using equal weights, results in huge residual values. Fan, Liu, Wu, Tsai, and Wei (2007) used the AWD method to analyze the relationship between energy consumption and CO2 emission intensity (CO2/GDP) in China from 1980 to 2003. They revealed that decreasing CO2 emission intensity results from decreasing the real Energy Intensity (Total energy consumption/GDP).

Greening, Davis, Schipper, and Kruschwitz (1997) considered energy structure, intensity and residual indices for the comparison of ten OECD countries using six decomposition methods: Laspeyres-fixed base year, Laspeyres-rolling base year, SAD-fixed base year, SAD-rolling base year, AWD-fixed base year, and AWD-rolling base year. The result showed that Laspeyres-fixed base year and Laspeyres-rolling base year underestimated Energy Intensity while the SAD-fixed base year overestimated Energy Intensity and their residual terms increase over time. The AWD and SAD-rolling base year showed the best results. Ang, Zhang, and Choi (1998) provided a logarithmic mean weight divisia index (LMDI) that uses a logarithmic function instead of the arithmetic mean weight function. This decomposition is popular because it does not include residual terms. Wang, Chen, and Zou (2005) considered fossil fuel energy consumption, total energy consumption, GDP and population using the LMDI in their CO2 emission study for China. Lee and Oh (2006) also used the LMDI for analyzing CO2 emissions in the APEC countries. They redefined the decomposition formula

\[
\text{CO2} = \left(\text{CO2/FEC}\right) \times \left(\text{FEC/TEC}\right) \times \left(\text{TEC/GDP}\right) \times \left(\text{GDP/P}\right) + P = FSIGP.
\]

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 CO2 emission was decreasing in APEC countries from 1980 to 1998. Recently, Liou and Wu (2011) proposed the indices for energy use efficiency and CO2 emission control efficiency to evaluate the economic development using developed countries’ CO2 emission percentage data in 2008. Various relationships among GDP per capita, energy use efficiency and CO2 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, engineering, 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 \( (p, d, q) \) is a white noise process which is a series of independent variables with expected value \( \mu \) 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 \( p \), \( P \), \( d \), \( Q \) and \( q \) are nonnegative. The difference is used in a non-stationary 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) determined 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.

\[ Y_t = v(t)X_t + N_t, \]

where the input and output variables are stationary time series processes, \( v(t) = v_0 + v_1B + \cdots + v_kB^k \), and the parameter \( v(B) \) is an impulse response function at each time lag. A rational lag structure is used with the transfer function, such that

\[ v(B) = \frac{\omega(B)}{\delta(B)}B^p = \frac{\omega_0 - \omega_1B - \omega_2B^2 - \cdots - \omega_pB^p}{1 - \delta_1B - \delta_2B^2 - \cdots - \delta_pB^p}. \]

In Eq. (4), the \((r,s,b)\) order is the transfer model, where \( \omega(B) \) is \( s \) order for the moving average operator, \( \delta(B) \) is \( r \) order for the autoregression 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:

\[ N_t = \frac{\theta(B)}{\phi(B)} \epsilon_t, \]

where \( \epsilon_t \) is white noise and follows \( \text{NID}(0, \sigma^2) \). Thus, the transfer function and the disturbance model can be rewritten as

\[ N_t = \frac{\theta(B)}{\phi(B)} \epsilon_t. \]

Fig. 1. Relationship between input and output variables in transfer function.
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 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 sum function 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; Nuchitprasitichai & Cremaschi, 2012; Zhang, Eddy Patuwo, & Y Hu, 1998). Khassei and Bijiari (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 an ordinary 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 (Xₜ), and retain the residual series (zₜ). The prewhitened input series can be expressed as

   \[ X_t = \frac{\partial(B)}{\partial(B)} X_t \]  

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

   \[ Y_t = \frac{\partial(B)}{\partial(B)} Y_t \]

3. Calculate the sample CCF \( \hat{\rho}_{xy}(k) \) of \( \{x_t\} \) and \( \{y_t\} \) and estimate the impulse response function \( \theta_k \)

   \[ \hat{\theta}_k = \frac{\sigma_y}{\sigma_x} \hat{\rho}_{xy}(k) \]

   Where \( \hat{\rho}_{xy}(k) = \text{E}(\{x_t - \mu\}(\{y_{t-k} - \mu\})) / \sigma_x \sigma_y \), \( k = 0, \pm 1, \pm 2, \ldots \)

   Cross correlation variance of \( \{x_t\} \) and \( \{y_t\} \) series with time lag \( k \).

4. Use the \( \hat{\theta}_k \) 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

\[ Y_t = C + \frac{\partial(B)}{\partial(B)} X_{t-b} + \frac{\partial(B)}{\partial(B)} a_t \]

Where the estimates include \( \hat{a}_t = (a_1, a_2, \ldots, a_n) \), \( \hat{\theta}_t = (\theta_1, \theta_2, \ldots, \theta_n) \) and \( \sigma_n^2 \). The conditional likelihood function of \( a_t \) is

\[ L(\hat{a}, \hat{\theta}, \sigma_n^2) = (2\pi \sigma_n^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma_n^2} \sum_{t=1}^{n} a_t^2 \right] \]

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

\[ \min L(\hat{a}, \hat{\theta}, \sigma_n^2) = \sum_{t=1}^{n} a_t^2 \]

Where \( \tau_0 = \max(p + r + 1, b + p + s + 1) \).

**Step 3. Model diagnosis**

In the transfer function model, suppose that the white noise series \( \{a_t\} \) 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 \( \{z_t\} \) and \( X_t \) which should be mutually independent. For an appropriate model, \( \{z_t\} \) and \( \{X_t\} \) of the sample correlation coefficient \( \rho_{xz}(k) \) 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

\[ Y_t = C + v(B)X_t + \frac{\theta(B)}{\phi(B)} a_t \]

Where \( v(B) \) can be expressed as

\[ v(B) = v_0 + v_1 B + \cdots + v_b B^b \]

LTF estimates the impulse response weights using the least squares method. Since the whole roots of \( \phi(B) \) are out of the unit circle, \( v(B) \) can be approximated as \( \text{c}(\phi(B)) \). 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 \( a_t \), 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
\[ N_t = \frac{1}{1 - \phi B} a_t. \] (15)

If the data is a seasonal series, the disturbance term can be temporarily written as
\[ N_t = \frac{1}{(1 - \phi_1 B)(1 - \phi_s B)} a_t. \] (16)

(2) Test the disturbance term \( N_t \) with an autocorrelation regression parameter estimate value. If the parameter value is close to 1, the input and output series should be used by difference or transfer.

(3) Re-estimate the impulse response weight, and observe the estimator of impulse response weight. If the impulse response weight belongs to a cut-off, a simple regression model with time lag is considered. On the contrary, a corner method provided by Liu and Hansen (1982) can be used to determine \( r, s, b \) values of polynomial \( \omega_m(B)B^b/\phi(B) \).

Step 2. Model estimation
\[ Y_t = C + (v_0 + v_1 B + \ldots + v_k B^k)X_t + \varepsilon_t \] (17)
and
\[ \hat{\beta} = \left[ \begin{array}{cccc} v_0 & v_1 & \ldots & v_k \end{array} \right], \]
\[ \hat{Y}_t = [X_{t-1}, X_{t-k}]^T, \]
\[ \hat{X}_t = [X_t^0, X_t^1, \ldots, X_t^k]. \]

where \( \hat{X} = B^0 \hat{X}^0 \) and \( \hat{X}^0 = [X_{t-1}, X_{t-k+1} \ldots X_{t(n)}] \), then \( \hat{\beta} \) can be estimated using least squares method
\[ \hat{\beta} = (\hat{X}^T \hat{X})^{-1} \hat{X}^T \hat{Y}. \] (18)

According to Johnston (1984), the estimated result of the asymptotic bias is
\[ \lim(\beta - \hat{\beta}) = \frac{1}{n} \sum_{k=1}^{n} \gamma_k \] (19)
where \( \gamma_k = \left[ \hat{\gamma}_{n+1}(0) \hat{\gamma}_{n+1}(1) \hat{\gamma}_{n+1}(2) \ldots \hat{\gamma}_{n+1}(k) \right]' \), and \( \hat{\gamma}_{n+1}(k) \) is the cross variance of \( \{x_i\} \) and \( \{X_i\} \) series with \( k \) time lag
\[ \sum_{k=1}^{n} \left[ \begin{array}{cccc} \hat{\gamma}_{0} & \hat{\gamma}_{1} & \ldots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_{1} & \hat{\gamma}_{0} & \ldots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \ldots & \hat{\gamma}_{0} \end{array} \right] \] (20)

where \( \gamma_1 \) 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:
\[ AIC(k) = -2 \ln(L) + 2k, \] (21)

2. AICc criteria:
\[ AICc(k) = AIC(k) + \frac{2k(k + 1)}{n - k - 1}, \] (22)

3. SBC criteria:
\[ SBC(k) = -2 \ln(L) + k \ln(n). \] (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 (2009) and Pan and Chen (2010).

\[ MAPE = \frac{\sum_{i=1}^{n} |Y_i - \hat{Y}_i| / n}{Y_i} \times 100\%, \] (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.

\[ MASE = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{|Y_i - \hat{F}_i|}{\frac{1}{n} \sum_{i=2}^{n} |Y_i - Y_{i-1}|} \right) = \frac{\sum_{i=1}^{n} |Y_i - \hat{F}_i|}{\frac{1}{n} \sum_{i=2}^{n} |Y_i - Y_{i-1}|}. \] (25)

where \( Y_i \) denotes the actual value at time \( t \), \( \hat{F}_i \) denotes the forecast value, and \( \hat{F}_i \) equals \( Y_{i-1} \).

Note that the denominator is the average forecast error of the one-step "naïve 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 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 LcE (liter of 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 = GDP/Energy consumption. (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:

\[
\Delta C = C_t - C_{t-1} = F_t S_t I_t G_t P_t - F_{t-1} S_{t-1} I_{t-1} G_{t-1} P_{t-1},
\]

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 given by:

\[
\Delta C = \Delta C_F \text{-effect} + \Delta C_S \text{-effect} + \Delta C_I \text{-effect} + \Delta C_G \text{-effect} + \Delta C_P \text{-effect},
\]

where \( \Delta C_F \text{-effect} = L(C_t, C_{t-1}) \ln(F_t/F_{t-1}), \) \( \Delta C_S \text{-effect} = L(C_t, C_{t-1}) \ln(S_t/S_{t-1}), \)

\[
\Delta C_I \text{-effect} = L(C_t, C_{t-1}) \ln(I_t/I_{t-1}), \quad \Delta C_G \text{-effect} = L(C_t, C_{t-1}) \ln(G_t/G_{t-1}),
\]

\[
\Delta C_P \text{-effect} = L(C_t, C_{t-1}) \ln(P_t/P_{t-1}),
\]

From Table 1, from 1995 to 2010, the following effects were decreasing: CO2 from fossil fuels, \( \Delta C_F \text{-effect} \), the proportion of fossil fuels in the energy mix, \( \Delta C_S \text{-effect} \) and Energy Intensity, \( \Delta C_I \text{-effect} \). 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

\[
I = \frac{P \times A \times E_1}{E_2},
\]

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 as second-order difference, the temporary CO2 prediction model follows ARIMA (0,2,1), as given by

\[
\nabla^2 Y_t = (1 + 0.6988) a_t, \quad t = 1, 2, \ldots, 19,
\]

\[\hat{\alpha}_0 = 5.22, \quad \hat{\alpha}_1 = 107.34 \text{ and } \beta = 108.04.\]

<table>
<thead>
<tr>
<th>Table 1: Item analysis of CO2 emission from 1995 to 2010.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO2 emission decomposition term</td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>ΔC = C - C_1 = FSIIGP - F_1S_1I_1G_1P_1</td>
</tr>
</tbody>
</table>

Unit: million tons.

Theoretical increment = ΔC_F-effect + ΔC_S-effect; Theoretical decrement = ΔC_F-effect + ΔC_S-effect + ΔC_I-effect.

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

\[
\nabla^2 Y_t = \omega_0 \nabla^2 X_{g(t-2)} + N_t, \tag{30}
\]

where \( N_t \) is a disturbance term. Using the nonlinear least squares method, the transfer function is

\[
\nabla^2 Y_t = 6.56 \times \nabla^2 X_{g(t-2)} + a_t, \tag{31}
\]

\( \sigma_a = 4.82 \), AICc = 91.97 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

\[
\begin{array}{cccc}
\text{CO₂} & \text{GDP per capita} & \text{Oil} & \text{Coal} & \text{New energy} \\
\text{PP-test} & -1.99 & -1.29 & -6.34 & -2.40 & -5.73 \\
\text{p-Value} & 0.96 & 0.92 & 0.71 & 0.95 & 0.76 \\
\text{KPSS-test} & 1.03 & 1.02 & 1.01 & 1.01 & 0.99 \\
\text{p-Value} & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\
\text{\nabla^2 CO₂} & \nabla^2 \text{GDP per capita} & \nabla^2 \text{Oil} & \nabla^2 \text{Coal} & \nabla^2 \text{New energy} \\
\text{PP-test} & -21.12 & -20.82 & -23.04 & -22.31 & -18.75 \\
\text{p-Value} & 0.02 & 0.02 & 0.01 & 0.01 & 0.04 \\
\text{KPSS-test} & 0.07 & 0.06 & 0.03 & 0.03 & 0.23 \\
\text{p-Value} & 0.1 & 0.1 & 0.1 & 0.1 & 0.08 \\
\end{array}
\]

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.

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:

\[
\nabla^2 Y_t = C + (v_{10} + v_{11}B + v_{12}B^2 + v_{13}B^3)\nabla^2 X_{all} + (v_{20} + v_{21}B + v_{22}B^2 + v_{23}B^3)\nabla^2 X_n + \frac{1}{1 - \phi_1B}a_t. \tag{31}
\]

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 \((r,s,b)\) is \((2,1,2)\). The first period parameters \( (\delta_1) \) of \( \{X_{all}\} \) and the current period \( (\omega_0) \) of \( \{X_n\} \) are not significant \((|t| < 2)\). The parameter estimates after excluding non-significant values are 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

\[
\nabla^2 Y_t = \left( \frac{\omega_{11}B}{1 - \delta_{11}B - \delta_{12}B^2} \right)\nabla^2 X_{all} + (\omega_{20} + \omega_{21}B + \omega_{22}B^2)\nabla^2 X_n + \frac{1}{1 - \phi_1B}a_t. \tag{32}
\]

Table 3

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard deviation</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(v_{10})</td>
<td>2.24</td>
<td>0.66</td>
<td>3.41*</td>
</tr>
<tr>
<td>(v_{11})</td>
<td>2.46</td>
<td>0.84</td>
<td>2.93*</td>
</tr>
<tr>
<td>(v_{12})</td>
<td>2.38</td>
<td>0.66</td>
<td>3.64*</td>
</tr>
<tr>
<td>(v_{13})</td>
<td>-3.24</td>
<td>0.74</td>
<td>-4.38*</td>
</tr>
<tr>
<td>(v_{20})</td>
<td>5.64</td>
<td>0.79</td>
<td>7.18*</td>
</tr>
<tr>
<td>(v_{21})</td>
<td>-1.86</td>
<td>0.87</td>
<td>-2.15*</td>
</tr>
<tr>
<td>(v_{22})</td>
<td>4.10</td>
<td>1.10</td>
<td>3.72*</td>
</tr>
<tr>
<td>(v_{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 \(|t| > 2\).
Moreover,\( \sigma^2 = 3.18, AICc = 104.66 \) and SBC is 98.14 for this emission. The statistics in Table 6 indicate that the parameters of current and second period impulse response function for GDP per capita series \( X_g \) are significant. Moreover, the parameters of current, first and second period impulse response function for the coal supply series \( X_c \) 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:

\[
\begin{align*}
\nabla^2 Y_t &= (\omega_{10} + \omega_{12} B^2) \nabla^2 X_g + (\omega_{20} + \omega_{21} B + \omega_{22} B^2) \nabla^2 X_c \\
&+ (\omega_{30} + \omega_{32} B^2) \nabla^2 X_n + \frac{1}{1 - \theta_2 B} \alpha_t.
\end{align*}
\]

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, \( \sigma^2 = 2.14, AICc = 104.66 \) and SBC is 98.14 for this prediction model. After performing model diagnosis, ACF and PACF fall within 2 standard deviations. The residuals for CCF and PACF 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). The predicted result of the three-variable time series is shown in Fig. 10. Given MAPE value equals 2.8% and MASE value equals 0.71 (<1), this prediction model is considered the most appropriate one. Therefore, it is more effective to use three input variables, i.e. GDP per capita, coal supplies and renewable energy supplies in predicting CO\(_2\) emission.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard deviation</th>
<th>t 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>4.25</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>

Table 6

Estimates and statistics for the impulse response function of GDP per capita, coal and new energy supplies.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard deviation</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_{11} )</td>
<td>1.07</td>
<td>0.31</td>
<td>3.53</td>
</tr>
<tr>
<td>( \delta_{12} )</td>
<td>-1.01</td>
<td>0.09</td>
<td>-11.00</td>
</tr>
<tr>
<td>( \omega_{21} )</td>
<td>-1.31</td>
<td>0.64</td>
<td>-2.04</td>
</tr>
<tr>
<td>( \phi_1 )</td>
<td>-0.61</td>
<td>0.16</td>
<td>-3.89</td>
</tr>
</tbody>
</table>
5.4. The comparison of MAPE and MASE values using BPN and our proposed method.

<table>
<thead>
<tr>
<th>Considered factors</th>
<th>BPN</th>
<th>Our proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP per capita</td>
<td>MAPE (%)</td>
<td>12.6%</td>
</tr>
<tr>
<td></td>
<td>MASE</td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.30</td>
</tr>
</tbody>
</table>

5.5. The comparison of MAPE and MASE values for predicting CO2 emissions using BPN and our proposed method.

<table>
<thead>
<tr>
<th>Region</th>
<th>MAPE (%</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPN</td>
<td>4.26</td>
</tr>
<tr>
<td></td>
<td>Our proposed method</td>
<td>2.8%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Region</th>
<th>MAPE (%)</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPN</td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td>Our proposed method</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Table 8

<table>
<thead>
<tr>
<th>Applications</th>
<th>MAPE (%)</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPN</td>
<td>4.26</td>
</tr>
<tr>
<td></td>
<td>Our proposed method</td>
<td>2.8%</td>
</tr>
</tbody>
</table>

Table 9

<table>
<thead>
<tr>
<th>Region, Year</th>
<th>MAPE (%)</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPN</td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td>Our proposed method</td>
<td>2.00</td>
</tr>
</tbody>
</table>

6. Discussions and conclusions

A key sustainability indicator (KSI) for worldwide environmental and economic 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 decreased 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%.
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 from 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 2001 to 2016 wherein the growth rate of coal, R is fixed at 3%, GDP per capita, Rg varies from 1% to 3% and new energy, Rn 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 used as a useful reference in predicting energy's environmental impact in other countries.

7. Uncited references

References


Box and Pierce (1970), Lu (2012) and Shapiro and Wilk (1965).