A Descriptive Time Series Analysis Applied to the Fit of Carbon-Dioxide (CO₂)

Chijioke Joel Nweke¹,a*, George Chinanu Mbaeyi¹,b, Kelechi Charity Ojide¹,c, Ogbonnia Elem-Uche²,d and Onyekachi Stanley Nwebe³,e

¹Department of Mathematics/Computer Science/Statistics, Alex-Ekwueme Federal University Ndifu-Alike, Nigeria
²Department of Mathematics and Statistics, Federal Polytechnic Unwana, Nigeria
³Department of Mathematics and Statistics, Institute of Management Science and Technology, Enugu, Nigeria

a*cj_nweke@yahoo.com, bgeorge.chinanu@funai.edu.ng, ckelechiojide@yahoo.com, ducogbonna@gmail.com, enwebeos@gmail.com

Keywords: spectrum, coefficient of variation, seasonal variation, multiplicative model, additive model, CO₂

Abstract. The study examined the use of population spectrum in determining the nature (deterministic and stochastic) of trend and seasonal component of given time series. It also adopts the use of coefficient of variation approach in the choice of appropriate model in descriptive time series technique. Illustrations were carried out using average monthly atmospheric Carbon dioxide (CO₂) from 2000-2017 with 2018 used for forecast. Spectrum analysis showed that the descriptive technique of time series is more appropriate for analysis of the study data. The coefficient of variation revealed that the multiplicative model was appropriate for the CO₂ data while the forecast and the actual values showed no significant mean difference at 5% level of significance.

1. Introduction

Atmospheric Carbon-dioxide (CO₂)is one of the greenhouse gases and it is known for its heat absorption and sunlight warming. In addition to this, CO₂ in particular dissolves into the ocean thereby producing carbon acid and lowering the ocean’s pH value [1]. It has been reported that higher concentration of CO₂ increases the growth of cereal crops but reduces the nutritional value of key staple crops [2]. Its occurrence has been traced to some socio-economic characteristics, income, car ownership [2], poor rail accessibility and urban density [3], human activity and natural system [4]. The impact has been projected to be greater in the southern Asia and African region [1,5] and may likely result to inability to keep global average temperature increase below 2ºc [6]. With these in view, it suffices to consider the fundamental processes that are believed to generate the subject matter and then build a detailed model that can be used for control and also make predictions. Since the study data is a time series data, such model process will be a stochastic process indexed with time. To build such model, we adopt a time series technique. Commonly used time series techniques include; descriptive technique, probability models technique and spectral density technique [7,8]. Any appropriate technique to be adopted depends on the composition of the time series data. If a descriptive technique is to be applied, the time series is assumed to comprise of four components namely: trend, seasonal variation, cyclical variation and irregular component [9]. Apart from comprising all or some of these components, the trend and seasonal component has to be deterministic in nature. When the nature of trend and seasonal component has been ascertained, then an appropriate model can be chosen among the competing model of descriptive time series.

Time Series decomposition involves the estimation, isolations and removal of various components of time series using descriptive techniques [10,11]. To choose among competing models for decomposition, a graphical approach [12] which involves the use of a time series plot has been recommended. If the seasonal variation stays roughly the same size regardless of the mean level, then additive model is more appropriate, but if it increases in size in direct proportion to the
mean level, then the multiplicative model is considered more appropriate[12]. Another graphical approach to the choice of model is the use of Buys-Ballot table[10] in which the relationship between the seasonal means and seasonal standard deviations provides guide to the desired model.

All the above have some degree of subjectivity and thus may not be dependable for selection of appropriate model, hence an objective approach is required. One objective approach is to use both additive and multiplicative model and choose any that has minimum mean square of its residual [10]. However, this is time consuming and often, it has been observed that the mean square error and other accuracy measures may be misleading [13]. Therefore, this study is aimed at recommending an objective approach to ascertain the nature of trend and seasonal component of a given time series as well as determining the appropriate choice among competing models in the application of descriptive technique of time series analysis.

2. Methodology

Descriptive techniques assume that a series is the combination of four major components which could be represented thus:

\[ X_t = f(T_t, S_t, C_t, e_t) \]  

(1)

where, \( X_t \) is the original value of the series at time \( t \), \( T_t \) is the Trend component, \( S_t \) is the Seasonal variation, \( C_t \) is the Cyclical component, and, \( e_t \) the Irregular component.

2.1 Determination of the Nature of Trend and Seasonal Components of a Given Series

For a given series \( X_t \) with its assumed various components, the trend component could be deterministic and stochastic. To ascertain the nature of trend and seasonal component, we will adopt spectral analysis approach. The method will be to evaluate the estimate of the population spectrum of the first-differenced series at zero frequency.

Let \( X_t, t \in T \) be an observed time series at time \( t \), let \( \Delta X_t = X_t - X_{t-1} \) be the first difference of the series. If \( \Delta X_t = X_t - X_{t-1} \) is covariance stationary with constant mean and summable autocovariance given by \( r_k \), then the autocovariance generating function is given by:

\[ g_{\Delta X_t}(z) = \sum_{k=-\infty}^{\infty} r_k z^k = \psi(z) \sigma^2 \psi(z^{-1}) \]  

(2)

where \( z \) denotes a complex scalar which could be represented as \( e^{-i\omega} \) for \( i = \sqrt{-1} \).

Therefore, the population spectral is given by

\[ S_{\Delta X_t}(\omega) = \frac{1}{2\pi} g_{\Delta X_t}(z) = \frac{1}{2\pi} g_{\Delta X_t}(e^{-i\omega}) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} r_k e^{-i\omega k} \]  

(3)

where \( \omega \) is a real scalar.

Putting \( z = 1 \) in equation (2), we have

\[ g_{\Delta X_t}(1) = \sum_{k=-\infty}^{\infty} r_k = [\psi(1)]^2 \sigma^2 \]  

(4)

The population spectrum in equation (3) evaluated at \( z = 1 \) and zero frequency becomes

\[ S_{\Delta X_t}(0) = \frac{1}{2\pi} g_{\Delta X_t}(1) = \frac{1}{2\pi} [\psi(1)]^2 \sigma^2 \]  

(5)

From equation (5), when a series is difference stationary, the population spectrum of the first difference at zero frequency is positive.

On the other hand, if the process is trend stationary, it has been shown that the auto-covariance generating function is given by

\[ g_{\Delta X_t}(z) = (1-z)\psi(z) \sigma^2 \psi(z^{-1})(1-z^{-1}) \]  

(6)
Therefore, with $z = 1$ in equation (6)
\[ g_{\Delta X_t}(z) = 0 \]  

(7)

Hence, the population spectrum of the first difference of a trend stationary series at zero frequency is zero.

### 2.2 Choice of Appropriate Model

The representation of the model in equation (1) depends on its various functional forms. They include:

- **Additive Model**
  \[ X_t = T_t + S_t + C_t + e_t \]  
  (8)

- **Multiplicative Model**
  \[ X_t = T_t \cdot S_t \cdot C_t \cdot e_t \]  
  (9)

- **Mixed or Pseudo Additive Model**
  \[ X_t = T_t \cdot S_t \cdot C_t + e_t \]  
  (10)

Mixed or Pseudo Additive model is mostly applied when the series comprises of small values. The two competing model are the additive and multiplicative model.

To choose the appropriate model for $X_t$, we adopt a coefficient of variation approach [14] outlines below:

(i) Calculate the seasonal difference of the series $X_t$, given by
\[ d_{t,i} = X_{t,i} - X_{t-1,i} \]  
and its corresponding seasonal quotient using
\[ q_{t,i} = \frac{X_{t,i}}{X_{t-1,i}} \]  

(11)

Where, $d_{t,i}$ is the $i^{th}$ seasonal differential at time $t$, $q_{t,i}$ is the $i^{th}$ seasonal quotient at time $t$, $X_{t,i}$ is the observed series at season $i$ and time $t$.

Note that seasonal difference is the difference ($d_{t,i}$) between a certain season of a year and same season from the previous year while seasonal quotient ($q_{t,i}$) is the quotient of certain season of a year and the same season from the previous year.

(ii) Calculate the coefficient of variation (CV) of the seasonal difference and seasonal quotient given:
\[ CV(d_{t,i}) = \frac{\text{Standard Deviation of } d_{t,i}}{\text{Mean of } d_{t,i}} \]  

(13)

and
\[ CV(q_{t,i}) = \frac{\text{Standard Deviation of } q_{t,i}}{\text{Mean of } q_{t,i}} \]  

(14)

(iii) Apply the following decision rule:
- If $CV(q_{t,i}) > CV(d_{t,i})$, Additive model is appropriate.
- If $CV(q_{t,i}) \leq CV(d_{t,i})$, Multiplicative model is appropriate.

### 2.3 Time Series Decomposition

In fitting the trend, the least squares method is considered appropriate, the seasonal component are isolated using ratio-to-trend method while the cyclical component, where present, are estimated using moving average approach.

### 3. Results and Discussion

To illustrate the above methodology, we use a monthly atmospheric CO2 data from 2000 to 2017 as our study data while 2018 will be use to examine the predictive performance.
The series in figure 1 shows an evidence of upward trend with regular periodic movement at almost equal interval indicating presence of seasonal variation.

### 3.1 Determination of the Nature of the Trend and Seasonal Component

The application of spectrum analysis in the determination of the nature of trend and seasonal component leads to the result presented in Figure 2.

**Figure 2: Spectrum of the first differenced series of the CO2 Data**

The spectrum represented in Figure 2 is the periodogram of the first differenced series of the study data, the value of the spectrum at zero frequency is zero. This is based on equations 3-7 which indicate that the trend and seasonal component in the study data is deterministic. Hence, the descriptive technique of time series is more appropriate for analysis of the study data.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Seasonal Difference ($d_{t,i}$)</th>
<th>Seasonal Quotient ($q_{t,i}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.1753</td>
<td>1.0052</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.5808</td>
<td>0.0035</td>
</tr>
<tr>
<td>Coefficient of Variation</td>
<td>0.2670</td>
<td>0.0035</td>
</tr>
</tbody>
</table>

Table 1 are the values for the choice of appropriate model. Based on the decision rule stated in section 2, since $\langle CV(d_{t,i}) \rangle > \langle CV(q_{t,i}) \rangle$, it suffices that the multiplicative model is more appropriate. Hence, further analysis shall be based on the multiplicative model.
3.2 Estimation and Isolation of Trend Component

Using the least square method to fit the trend on the original series of CO$_2$, we have the result as presented in Table 2.

Table 2: Trend Analysis of the original series

<table>
<thead>
<tr>
<th>Trend Line</th>
<th>Fitted Line</th>
<th>RMSD</th>
<th>MAD</th>
<th>MAPE</th>
<th>Adj-R$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>$\hat{T}_t = 367.908 + 0.177t$</td>
<td>2.260</td>
<td>1.914</td>
<td>0.495</td>
<td>96.0%</td>
</tr>
<tr>
<td>Quadratic</td>
<td>$\hat{T}_t = 368.875 + 0.150t + 0.000123t^2$</td>
<td>2.220</td>
<td>1.891</td>
<td>0.489</td>
<td>96.1%</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\hat{T}_t = 368.230(1.00046)^t$</td>
<td>2.238</td>
<td>1.899</td>
<td>0.491</td>
<td>96.0%</td>
</tr>
</tbody>
</table>

RMSD- root mean square deviation, MAD-mean absolute deviation, MAPE- mean absolute percentage deviation, Adj-R$^2$- adjusted coefficient of determination.

The trend analysis shows common similarities on the accuracy measures of the likely competing trend lines (Linear, quadratic and exponential model). Hence, based on theory of parsimony the linear trend is chosen to fit the trend line. The fitted trend on the original series is displayed in figure 3 while the trend equation is shown in Table 1.

3.3 Seasonal Index Estimation and Isolation

The result of application ratio-to-trend on the de-trended series for estimation of seasonal index is as displayed in table 4.

Table 3: The Seasonal Indices for Carbon-dioxide (CO$_2$)

<table>
<thead>
<tr>
<th>Season</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sept</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index</td>
<td>1.001</td>
<td>1.002</td>
<td>1.004</td>
<td>1.007</td>
<td>1.008</td>
<td>1.001</td>
<td>0.995</td>
<td>0.991</td>
<td>0.991</td>
<td>0.995</td>
<td>0.998</td>
<td></td>
</tr>
</tbody>
</table>

3.4 Removal of Cyclical Variation and Diagnosis of Residual

The pattern of the detrended-deseasonalized series suggests that there is a presence of cyclical variation in the series.

To extract the cyclical component, the detrended-deseasonalized is fitted to a moving average and the result of the fit is as shown in Figure 4.
Examining the residual for normality and serial correlation, the result is as displayed in figure 5 and 6. The result on figure 5 is on normality test, which show Anderson-Darling (AD) of 0.48 with p-value greater than the level of significance. This suggests that the residual is normally distributed. This is also evident with the histogram having almost a bell shape of the normality curve. Figure 6 is the autocorrelation function of the residual. The result showed that none of the residual is significantly different from zero, indicating that the residual is a white noise. This implies that the model was appropriately fitted.
In this study, the average monthly atmospheric Carbon-dioxide CO₂ from 2000 to 2018 was analyzed using descriptive time series technique. The population spectrum of the first difference of the series showed that the nature of the trend and seasonal component inherent in the CO₂ data is deterministic which prompted the use of descriptive technique. Also, the coefficient of variation approach applied for choice of model revealed that the multiplicative model was appropriate for the CO₂ data. Finally, the forecasted values and the actual values showed no significant mean difference at 5% level of significance.

### References


### Table 4: Forecast for 2018

<table>
<thead>
<tr>
<th>Season</th>
<th>Trend</th>
<th>Seasonal Index</th>
<th>Forecast</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>406.317</td>
<td>1.001</td>
<td>406.561</td>
<td>407.710</td>
</tr>
<tr>
<td>February</td>
<td>406.494</td>
<td>1.002</td>
<td>407.348</td>
<td>408.320</td>
</tr>
<tr>
<td>March</td>
<td>406.671</td>
<td>1.004</td>
<td>408.298</td>
<td>409.410</td>
</tr>
<tr>
<td>April</td>
<td>406.848</td>
<td>1.007</td>
<td>409.696</td>
<td>410.240</td>
</tr>
<tr>
<td>May</td>
<td>407.025</td>
<td>1.008</td>
<td>410.322</td>
<td>411.240</td>
</tr>
<tr>
<td>June</td>
<td>407.202</td>
<td>1.006</td>
<td>409.645</td>
<td>410.790</td>
</tr>
<tr>
<td>July</td>
<td>407.379</td>
<td>1.001</td>
<td>407.868</td>
<td>408.710</td>
</tr>
<tr>
<td>August</td>
<td>407.556</td>
<td>0.995</td>
<td>405.681</td>
<td>406.990</td>
</tr>
<tr>
<td>September</td>
<td>407.733</td>
<td>0.991</td>
<td>404.226</td>
<td>405.510</td>
</tr>
<tr>
<td>October</td>
<td>407.910</td>
<td>0.991</td>
<td>404.402</td>
<td>406.000</td>
</tr>
<tr>
<td>November</td>
<td>408.087</td>
<td>0.995</td>
<td>405.924</td>
<td>408.020</td>
</tr>
<tr>
<td>December</td>
<td>408.264</td>
<td>0.998</td>
<td>407.447</td>
<td>409.070</td>
</tr>
</tbody>
</table>

### Variable Mean ± std error Mean Diff. t-value p-value

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean ± std error</th>
<th>Mean Diff.</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td>407.28±0.52</td>
<td>-1.216</td>
<td>-1.560</td>
<td>0.132</td>
</tr>
<tr>
<td>Forecast</td>
<td>408.50±0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


