University of North Carolina at Chapel Hill

Department of Mathematical Decision Science

STOR 456

Group Project

Forecasting Mauna Loa CO_2 Level for April, 2015

Team TarHeels:

Runke, Liao
Yingfang, Wu
Liziqiu, Yang
Junyan Yao

I. Abstract

This paper predicted the monthly average carbon dioxide level (ppm) of Mauna Loa, Hawaii in April 2015. Using data extracted from National Oceanic and Atmospheric Administration (NOAA) Mauna Loa Observatory, we projected the April average CO_2 level to be 402.7285 ppm with 95% confidence interval of (402.1101, 403.3469) ppm and 99% confidence interval of (401.9161, 403.541) ppm. The ARMA (1, 1) model was used to forecast the CO_2 level from previously recorded data. (The classical decomposition model fails to pass the residual diagnosis.) Series of data from different time periods were chosen for analysis by the ITSM package of R software. Results were compared and suggested data from March 1999 to February 2015 provided the most reasonable estimation.

II. Introduction

Located at 11,141 feet on Mauna Loa volcano, the Mauna Loa Observatory is the world's earliest air monitoring station. It continuously monitors and collects data related to atmospheric change, particularly the level of CO_2. Recorded concentrations of CO_2 from observatory dating from 1958 gave rise to the famous Keeling Curve shown below.

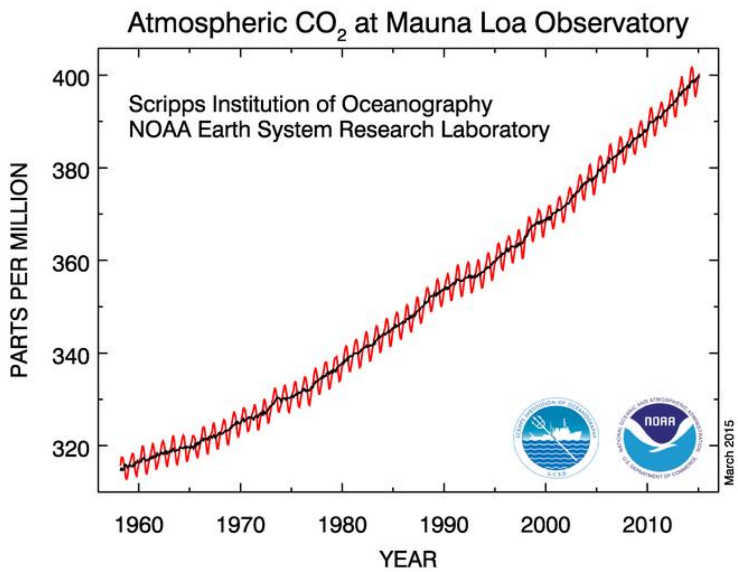

Figure 1. Monthly CO_2 fluctuations are captured by the red curve. The black curve demonstrates the data after removal of seasonality. Concentrations of CO_2 are measured in a "dry air mole fraction" defined as the number of CO_2 molecules over a given number of air molecules after removal of water vapor, multiplied by one million (ppm).

Source: National Oceanic and Atmospheric Administration

III. Observation of Data

The Mauna Loa Observatory provides the average monthly carbon dioxide data starting from March 1958. C. David Keeling first developed the idea to record CO_2 level in 1958. The NOAA started its own CO_2 observation after 1972. For consistency, we focused on the time frame after 1972. In addition, time prior to 1999, the year of the most recent significant El Nino, was discarded from the analysis due to the disrupting effects of El Nino on the level of CO_2. According to the NOAA analysis, the CO_2 fluxes during a typical El Niño period are reduced by more than 100% compared with the non-El Niño condition (National Oceanic and Atmospheric Administration, 2015). While El Niño sometimes triggers substantial changes in the weather, numerous El Niño events subsequent to that in 1998 were not strong enough to contribute to the global weather change in the near future (Conners, D. 2015). According to "RealClimate", a popular commentary blog on climatology, the El Nino events are particularly strong around 1984, 1988, 1996 and 1998. In this case, it is justifiable to consider only the period after 1998 for analysis to eliminate the potential effect of El Nino on prediction.

a. Trend

According to Figure 2, the current level of atmospheric CO_2 is significantly higher than at any time in the past 10,000 years. Continuous increase in CO_2 level has occurred since the beginning of the industrial era (around 1750) and the most substantial elevation, recorded by the Keeling Curve, occurred five decades ago (Cook J, 2015). The most prevailing explanation for the skyrocketing CO_2 level is the increase in burning of fossil fuels, half of which occurred since the mid-1970s (Cook J, 2015). While land-use changes and cement manufacturing also contribute to the CO_2 rise, another intriguing reason is about the atmospheric temperature. As the level of CO_2 increases, the greenhouse effect raises global temperature. Higher temperature in turn reduces the solubility of CO_2 in ocean. Therefore, more CO_2 remains in the atmosphere in recent 50 years than before.

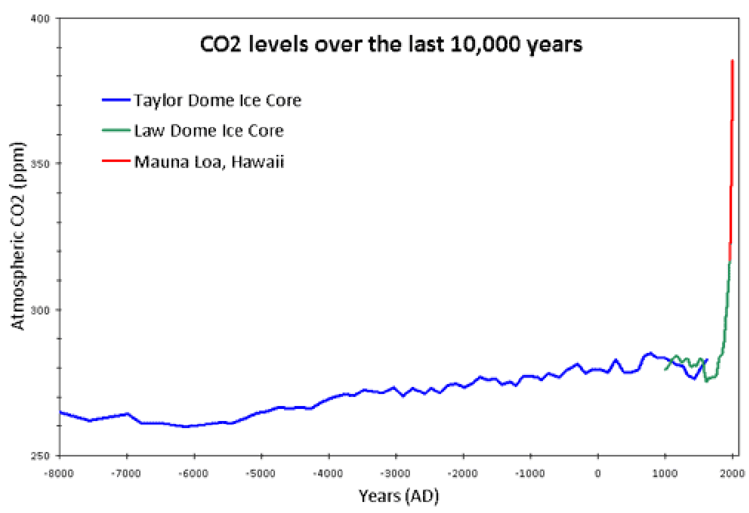

Figure 2. *CO_2 levels over the last 10,000 years*
Source: Cook, J. (2015). *Are CO2 levels increasing?*

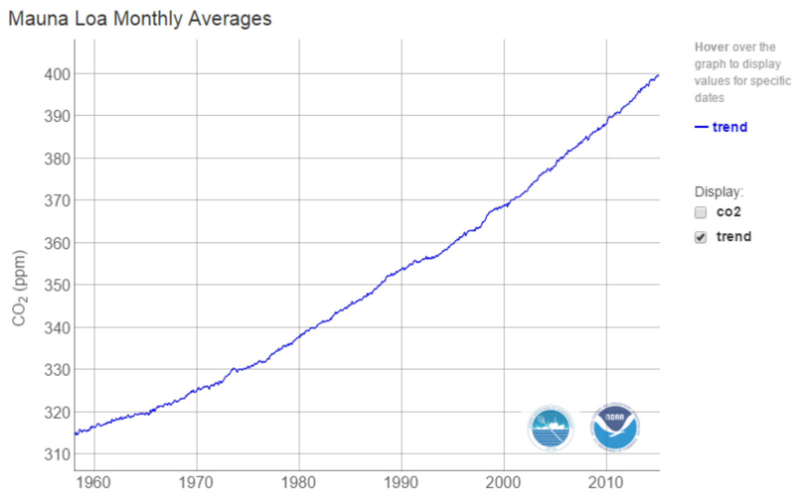

Figure 3. *Monthly Average CO_2 level from Mauna Loa Observatory Station at Mauna Loa since 1958*
Source: National Oceanic and Atmospheric Administration (2015)

b. Seasonality

Carbon dioxide level observations for Mauna Loa clearly demonstrate a seasonal cycle, which s predicted from the original C. Keeling's research. As the study indicates, they conclude, "The concentration of atmospheric carbon dioxide at Mauna Loa Observatory exhibits a seasonal pattern that repeats with striking regularity from year to year (Bacastow, R. B, 1985)." In their study, they point out that the CO_2 Concentration increases at an average rate of about 0.7% per year since 1958, which is also statistically significant. This is also corresponding to our analysis for trend. The seasonal cycle of CO_2 in Mauna Loa is very likely to be associated with metabolic activity of terrestrial vegetation. Precisely, the increasing plant activity has contributed to at least part of the upward rising in the CO_2 concentration (Bacastow, R. B, 1985).

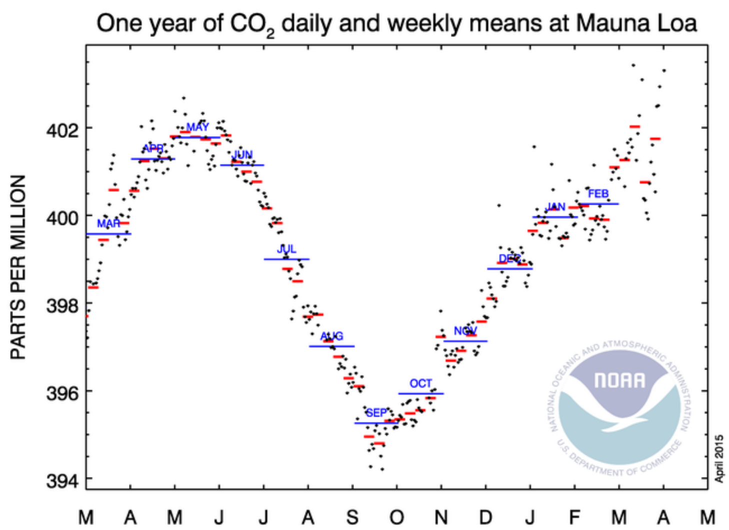

Figure 4. This figure shows a strong seasonal effect. CO_2 concentration reaches peak regularly in May and declines until September.

Source: National Oceanic and Atmospheric Administration (2015)

Ⅳ. Model Specification:

We have already confirmed the general rising trend and seasonality for the level of CO_2, both of which interferes with data analysis. Therefore, the classical decomposition model is used to remove trend and seasonality for analysis.

Data analysis with classical decomposition model

Denote $$X_t = S_t + M_t + Y_t$$

Where $\{X_t\}$, t=1, 2, …, 192 is the monthly average levels of CO_2 data from March 1999 (t=1) to February 2015 (t=192)
S_t is the seasonal component of the time series of a period 12, i.e. $\sum_{St=0}^{12} i=1$ *and* $S_t + 12 = S_t$
M_t is the trend component of the time series and Y_t is the stationary process with mean zero.

We first followed the classical decomposition process to remove the seasonal component of the data. The appearance of deseasonalized data d_s ($d_s=X_t-S_t$) indicates that it fits a quadratic trend. Then we remove the quadratic trend component M_t from the data.

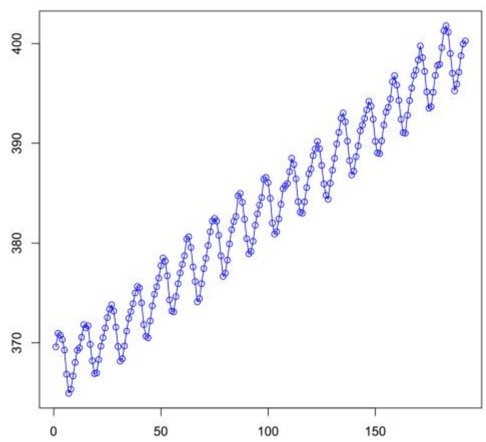

Figure 5: Plots of Original Data: By input the times series data in R and plot the data, it generate figure 1 with obvious seasonality.

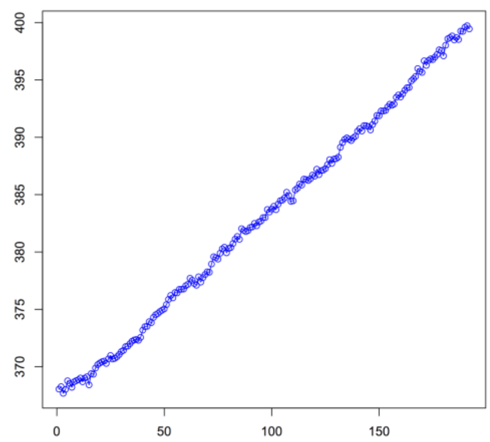

Figure 6: Plots of Deseasonalized Data: After removing the seasonal component of original data and it left a rising trend.

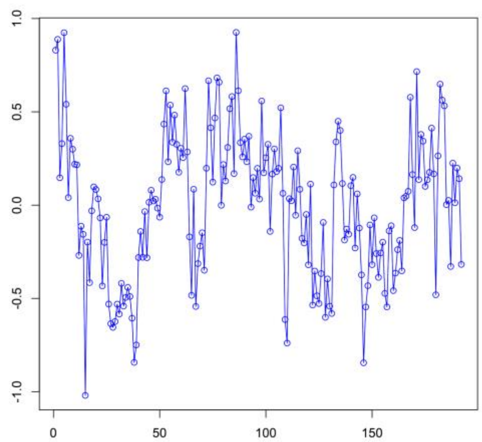

Figure 7: Plots of Residual after removing both seasonal and trend component, the data only left with residuals

Figure 5 presents the CO_2 data has trend and seasonal pattern with a 12-month cycle. Figure 6 shows the rising trend after removing the seasonal component of CO_2 level. Figure 7 is the residual plot.

It is unlikely to determine from the plots whether the residual is stationary, we therefore perform the diagnosis of e_t by employing the following tests:

Test of Null Hypothesis: Residuals are IID Noise

Test	Distribution	Statistics	P-value
Ljung-Box	Q~Chisq(20)	297.65	0
McLeod-Li	Q~Chisq(20)	33.67	0.0284
Turning Points	(T-126.7)/5.8~N(0,1)	117	0.0964
Diff signs S	(S-95.5)/4~N(0,1)	94	0.7084
Rank P	(P-9168)/445.1~N(0,1)	9144	0.957

Table 1. Diagnosis results of the residuals

The null hypothesis assumes that the residuals are independently and identically distributed noise. Although SACF, Turning Points Test, Difference Sign Test, and Rank Test do not reject the null hypothesis, the diagnosis table indicates that both Ljung Box test and McLeod-Li test reject the null hypothesis. We therefore conclude the classical decomposition estimation is not an appropriate model to predict the CO_2 level in April 2015.

Therefore, we decide to use another model to simulate the data to see whether we can get a better fit and prediction.

By looking at the ACF and sample PACF of AR (1) model, we decide that AR (1) is appropriate. Then, we examine the MA (1) model, and it also fits the model, which indicates that an MA (1) model now seems appropriate. The best model for the data then would seem to be an ARMA model.

Model ARMA:

An ARMA (1, 1) process of $\{X_t\}$ presents itself as

$$X_t = \Phi_1 X_{t-1} + \Theta_1 Z_{t-1} + Z_t, \text{ where } Z_t \sim WN(0, \sigma^2).$$

The ITSMR package in R automatically generates the number of p and q to be (1, 1). By plugging p and q, we are able to obtain the value of phi and theta as following:

- Φ_1=0.848564
- Θ_1=-0.3452003
- AICC=65.38236

Thus, the exact formula of the ARMA (1, 1) is

$$X_t=0.848564X_{t-1}-0.3452003Z_{t-1}+Z_t$$

With white noise variance σ^2=0.0794

ACF and PACF:

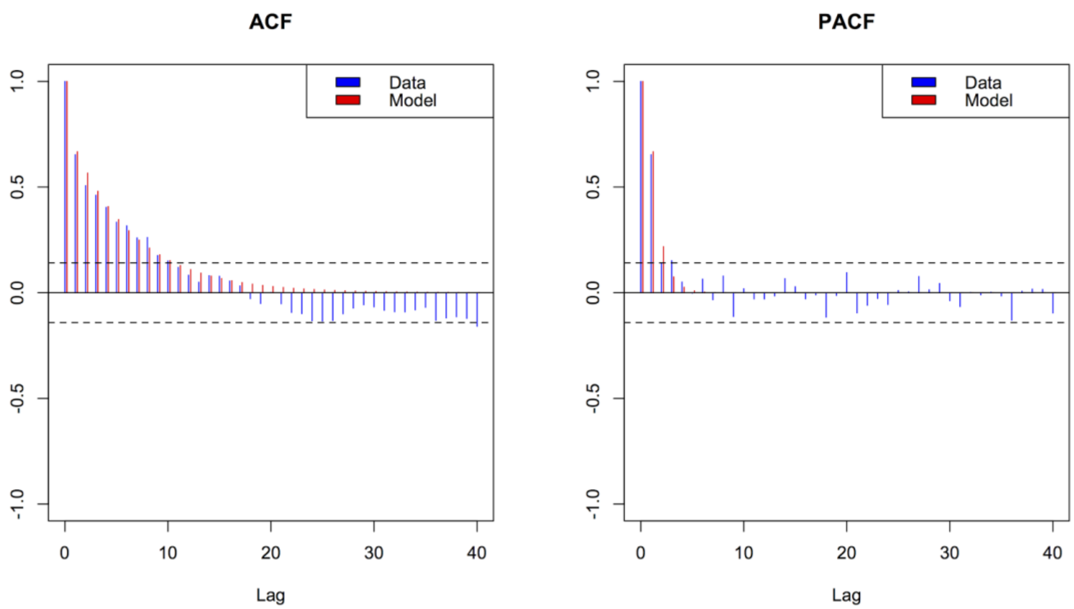

Figure 8. This figure shows a good fit of ARMA (1, 1) process and residuals. Together with a low AICC, ARMA (1, 1) is believed to be a model with a good combination with simplicity and precision.

Prediction and Confidence Interval

We also check the causality and invertibility of the model, which is valid. Therefore, we predict that the CO_2 level in April 2015 is 402.7285 ppm.

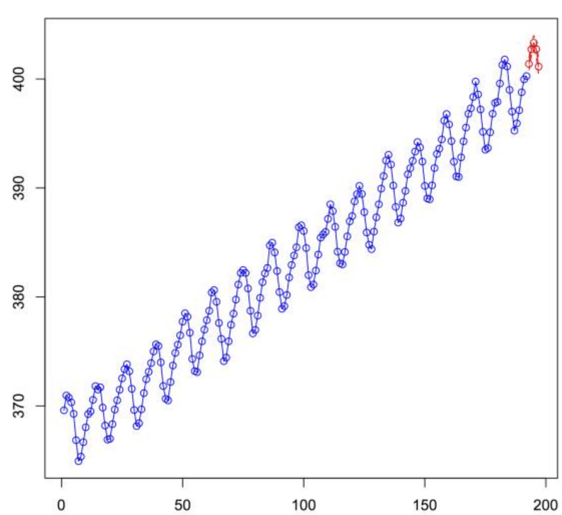

Figure 9. 95% Confidence Interval of ARMA (1,1) prediction

Step	Prediction	Sqrt(MSE)	Lower Bound	Upper Bound
1	401.3826	0.2818247	400.8302	401.935
2	402.7285	0.3155147	402.1101	403.3469
3	403.3214	0.3376985	402.6595	404.9833
4	402.737	0.3528094	401.0455	403.4285
5	401.1292	0.3633012	400.4172	401.8413

Table 2. 95% Confidence Interval of ARMA (1,1) prediction

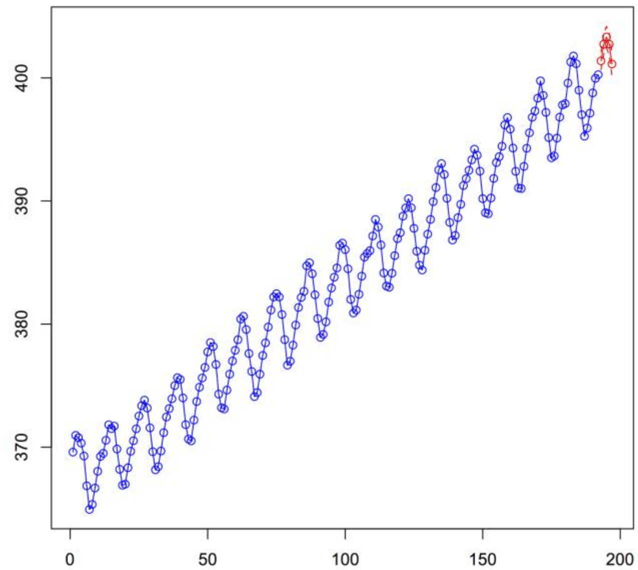

Figure 6. 99% Confidence Interval of ARMA (1,1) prediction

Step	Prediction	Sqrt (MSE)	Lower Bound	Upper Bound
1	401.3826	0.2818247	400.6569	402.1083
2	402.7285	0.3155147	401.9161	403.541
3	403.3214	0.3376985	402.4519	404.191
4	402.737	0.3528094	401.8285	403.6455
5	401.1292	0.3633012	400.1937	402.0647

Table 3. 99% Confidence interval of ARMA (1,1) prediction

Discussion:

We decided not to utilize the classical decomposition model to predict the CO_2 level because the residual e_t failed to pass the Ljung-Box test and McLeod-Li test. In contrast to the classical decomposition model, the ACF and PACF of ARMA (1, 1) demonstrated a good fit to the data. As a result, we decided to employ ARMA (1, 1) to estimate the CO_2 level in April 2015. The output of R showed that the estimated CO_2 level in April 2015 is 402.7285 ppm. In addition, the 95% confidence interval is (402.1101, 403.3469), and the 99% confidence interval is (401.9161, 403.541).

The exact value of CO_2 level in March 2015 has not been released. Although Dr. Pieter Tans from NOAA provided us with rough estimates for the March 2015 data, the accuracy of the data needed further confirmation from quality control procedures. Thus, we decide not to include the recent measurements.

We tried to predict the CO_2 level in 2014 from our model to check the feasibility of the model, and it indicated the time period we chose yielded best fit. That is why we chose CO_2 level from March 1999 to February 2015 as data input.

Record shows that the El Nino effect was very strong in 1998. Since the exact relationship between El Nino and the CO_2 concentration was still uncertain, we removed the time period before 1999 to minimize interference to analysis. El Nino events in the years after 1999 were in smaller scale and were not expected to strongly affect the accuracy of the result. Thus, the time period we chose was reasonable and could generate sound predictions.

We tested both lag-2 prediction and lag-1 prediction in April 2014, and the lag-1 prediction gave the better fit. However, since we did not have the accurate data in March 2015, the accuracy of the prediction might be compromised.

Conclusion:

Based on the ARMA (1, 1) process, the CO_2 level in April 2015 was predicted to be 402.7285 ppm. The 95% confidence interval is (402.1101, 403.3469), and the 99% confidence interval is (401.9161, 403.541).

According to the previous analysis, we rejected the classical decomposition model because it did not pass the Ljung Box test or the McLeod-Li test.

According to the results from MA (1) and AR (1), we decided to use the ARMA (1, 1) model to fit this process. After running the model, we found out that the residuals ACF and PACF of the ARMA model gradually tails off to zero, which captures the feature of the ARMA model. Thus, we believed that the ARMA (1, 1) model is a good candidate to fit the data and to make reasonable predictions.

The trend of the CO_2 levels from our research shows a continuous increase in concentration, which indicates the severe global warming that human beings should be aware of.

References

1. National Oceanic and Atmospheric Administration. (2015). *Trends in Atmospheric Carbon Dioxide*. Retrieved from http://www.esrl.noaa.gov/gmd/ccgg/trends/graph.html
2. National Oceanic and Atmospheric Administration. (2015). *The Impact of El Nino on Equatorial Pacific CO2 Fluxes*. Retrieved from http://www.pmel.noaa.gov/co2/elnino.html
3. Conners, D. (2015, March 15). El Niño has arrived, says NOAA. *EarthSky*.
4. Cook, J. (2015). *Are CO_2 levels increasing?* Retrieved from https://www.skepticalscience.com/co2-levels-airborne-fraction-increasing.htm
5. Bacastow, R. B., Bacastow, R. B., Keeling, C. D., & Whorf, T. P. (1985). *Journal of geophysical research. D, atmospheres: Seasonal amplitude increase in atmospheric CO_2 concentration at mauna loa, hawaii, 1959 - 1982* American Geophysical Union.

Appendix

R Codes:

```
#Read in data
Data<-scan("test.tsm")
#Seasonality estimation
s=season(Data,12)
#Remove seasonality
y=Data-s
#Trend estimation
m=trend(y,2)
#Remove seasonality and trend
e=Data-s-m
#Test residuals for stationarity and randomness
test(e)
#Find the best models from a range of possible ARMA models
autofit(e,p=0:7,q=0:7)
#Check causality and invertibility of ARMA
Check(arma(e,1,1))
#Estimate the residuals of ARMA model
p=Resid(Data,xv,arma(e,1,1))
#Test residuals for stationarity and randomness of ARMA model
test(p)
#Define a new function to predict the result with 99% Confidence Interval
#To predict 10 more steps, and the display plot and tableau
#This is a two-sided test, so we choose the score as 2.575
forecast99 <-function (x, xv, a, h = 10, opt = 2)
```

```
{
        f = .forecast.transform(x, xv, a, h, 1)
        if (!is.null(f$phi)) {
                psi = ma.inf(list(phi = f$phi, theta = a$theta), h)
                g = function(j) sum(psi[1:j]^2)
                se = sqrt(a$sigma2 * sapply(1:h, g))
                l = f$pred - 2.575 * se
                u = f$pred + 2.575 * se
                f = list(pred = f$pred, se = se, l = l, u = u)
  }
if (opt > 0) {
        if (is.null(f$se))
                cat(" Step Prediction Lower Bound Upper Bound\n")
        else cat(" Step Prediction sqrt(MSE) Lower Bound Upper Bound\n")
        for (i in 1:h) {
                cat(format(i, width = 5))
                cat(format(f$pred[i], width = 15))
                if (!is.null(f$se))
                        cat(format(f$se[i], width = 15))
                cat(format(f$l[i], width = 15))
                cat(format(f$u[i], width = 15))
                cat("\n")
      }
   }
   if (opt > 1)
      .plot.forecast(x, f)
   return(invisible(f))
}
#Transform data: remove seasonality and quadratic trend
xv=c("season",12,"trend",2)
#Prediction the results with default of 95% Confidence Interval
forcast(Data, xv, arma(e,1,1), h=12, opt=2)
#Prediction the results the newly defined function with 99% Confidence Interval
forcast99(Data, xv, arma(e,1,1), h=12,opt=2)
```

Output:

```
$phi
[1] 0.848564

$theta
[1] -0.3452003

$sigma2
[1] 0.07942515

$aicc
[1] 65.38236

$se.phi
[1] 0.00370301

$se.theta
[1] 0.01318803
```