ANALYSIS AND FORECAST OF CO2 FROM VOLCANIC ACTIVITY

CORY NGUYEN, LESLIE MEDINA, ADRIAN GONZALEZ, CHRISTOPHER COBAR, KEVIN AYALA

ABSTRACT

The data analyzed comes from the Mauna Loa data set which measures CO2 composition parts per million from a volcano on the island of Mauna Loa. This paper seeks to model this data set using time series analysis. The data was split into a training set (1965-1979) and a test set (1980) which will be used to validate forecasting using the model. First, the data was checked to see if a transformation was necessary using Box-Cox but it was found that no transformation was necessary. The data had visible trend and seasonality and was thus differenced at lags 1 and 12 to form a SARIMA model. After selecting preliminary models based on the ACF and PACF plots, a model was chosen based on the Akaike Information Criterion Corrected for Bias (AICC). The model most suitable for our data was a SARIMA (0,1,3) x (0,1,1)12 model; however, due to the law of parsimony, a SARIMA (0,1,1) x (0,1,1)12 model was instead chosen given it had the second best AICC value. The model passed all diagnostic checks and was used to forecast CO2 composition parts per million of Mauna Loa for the year 1980. The forecast predicted the values well and the true data points fell within the 95% confidence interval.

PSTAT 174 – DR. SUDEEP BAPAT
JUNE 6, 2018

TABLE OF CONTENTS

SECTION 1– INTRODUCTION

The data for our dataset was taken at the Mauna Loa Observatory (MLO). The location of the MLO was chosen by scientist Charles David Keeling in 1957. Keeling was the first to make accurate measurements of CO_2 in the atmosphere. He chose the location of the MLO high on the slopes of the Mauna Loa volcano because he wanted to measure CO_2 in air masses that represented most of the Northern Hemisphere, and possibly the entire globe. The MLO has rigorous data selection criteria including series of flags placed on certain hours, based on wind flow patterns, hourly average measurements, and ten second average measurements within an hour.

How is CO_2 measured at the MLO? Air is slowly pumped through a small cylindrical cell with flat windows on both ends. Infrared light is transmitted on one end, where it goes through the cell and out the opposite end, hitting a light detector. In the atmosphere, CO_2 absorbs infrared radiation. In the cell, the same process occurs, where CO_2 absorbs infrared light. When there is more CO_2 in the cell, more light is absorbed and less light is detected.

CO_2 is measured in parts per million, which is the number of carbon dioxide molecules in every million molecules of dry air. Water vapor is removed from the air collected and the amount of CO_2 is measured in parts per one million (ppm) molecules of dry air. Water vapor is removed because the mole fraction (or ppm) in dry air doesn't change when air expands upon heating or upon ascending to locations with lower pressure.

We forecasted ppm of CO_2 using a time series approach. The sample CO_2 levels were collected at monthly intervals from 1960-1985. We concluded the data is best modeled using a SARIMA $(0, 1, 1) \times (0, 1, 1)_{12}$ approach to account for the seasonality component. The details of the steps taken to determine this model and forecast future CO_2 levels are recorded in the following sections. [1]

SECTION 2– EXPLORATORITORY ANALYSIS

The data was collected at monthly intervals for 15 years with 180 observations. A brief overview of the data includes the minimum value of carbon dioxide levels at 317.2 ppm with a maximum of 341.2 ppm. The average was 328.5 ppm carbon dioxide with a variance of 30.402. *figure 2.1* displays the raw time series plot of the data. After verifying the data was approximately normal and testing other transformations, it was found the data would be best used untransformed.

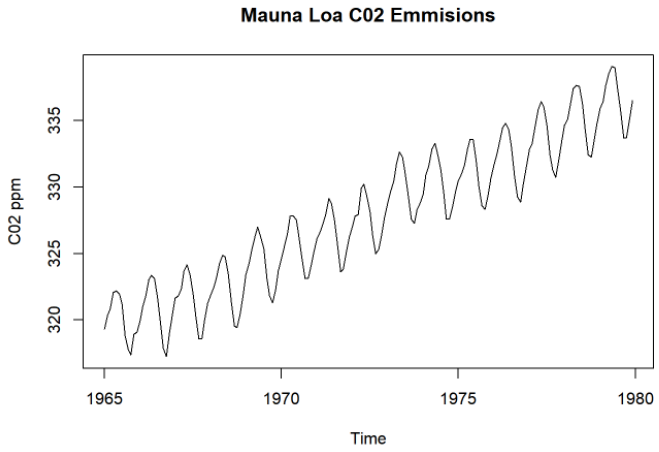

figure 2.1: The Raw Time Series Plot

At first glance of the plot of the time series, we observed a clear positive trend. A unique feature of this data is a presence of seasonality. The data showed carbon dioxide levels rising during the first half of each year, declining during the summer, and climbing again in the fall. It may not seem intuitive to most people, but volcanoes produce different levels of CO_2 depending on seasons. One may assume volcanoes have an equal probability of erupting anytime of the year. According to the Journal of Geophysical Research, "An analysis of volcanic activity during the last three hundred years revealed that volcanic eruptions exhibit seasonality to a statistically significant degree" [4]. This is remarkable and the article later goes on to declare this seasonality is perhaps explained due to annual movements of water mass.

SECTION 3– DATA TRANSFORMATIONS

SECTION 3.1– EXAMINING TRANSFORMATIONS:

Which type of transformation would be best? To determine the transformation needed a Box-Cox plot was created in top of *figure 3.1.2*. The formula for said transformation is given by *figure 3.1.1*, where U_t is the original data set and lambda is a value in the Box-Cox plot. By looking at the graph, it can be observed that the 95% confidence interval encompasses lambda values ranging from about -5 and 5, with the suggested transformation being approximately lambda= -0.1010101.

$$f_\lambda(U_t) = \begin{cases} \ln U_t, & \text{if } U_t > 0,\ \lambda = 0; \\ \lambda^{-1}(U_t^\lambda - 1), & \text{if } U_t \geq 0,\ \lambda > 0 \end{cases}$$

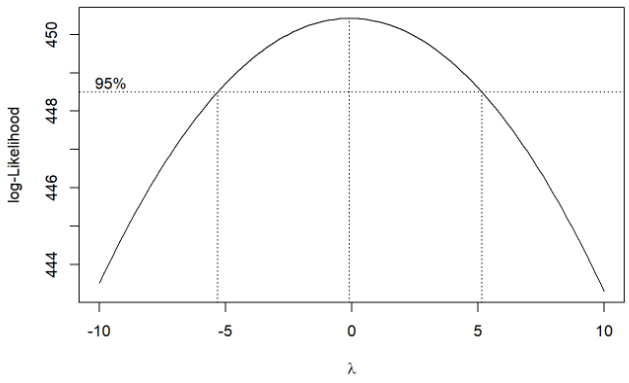

figure 3.1.1

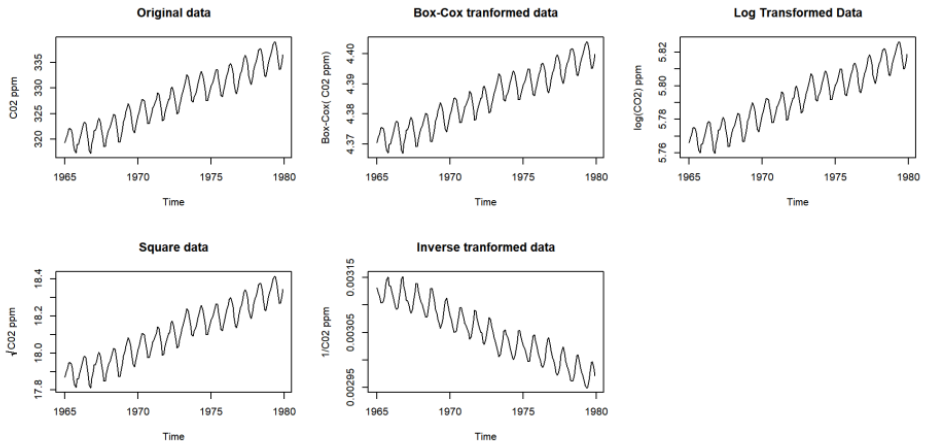

figure 3.1.2: Transformation Comparison

SECTION 3.2– REMOVING SEASONALITY AND TREND:

Differencing is an important method used to remove seasonality and trend to make data stationary. Stationarity of time series data must be assumed before one can do proper forecasting. Differencing can be defined mathematically as:

$$\nabla Y_t = Y_t - Y_{t-1}$$

We can also difference multiple times. Differencing using this method is typically used when removing trend. The equation for differencing *d* times is formulated below:

$$\nabla^d Y_t = (Y_t - Y_{t-1})^d$$

The last way we can use differencing to make data stationary is by differencing at lag *s*. When given a data set that is seasonal every *s* points, we use the formula below to remove this seasonality:

$$\nabla_s Y_t = Y_t - Y_{t-s}$$

Because it was unnecessary to transform the original data, we plotted the ACF and PACF of the original data. In *figure 3.2.1* an oscillating ACF and PACF can be observed. This oscillation is a sign of seasonality. Because data points in the data set

were recorded monthly, twelve times a year, we differenced at lag 12 to remove this seasonality ($\nabla_{12}Y_t$). When differenced at lag 12, the variance of the data decreases greatly (from 30.402 to 0.373). It is apparent that our data has already become much more stationary.

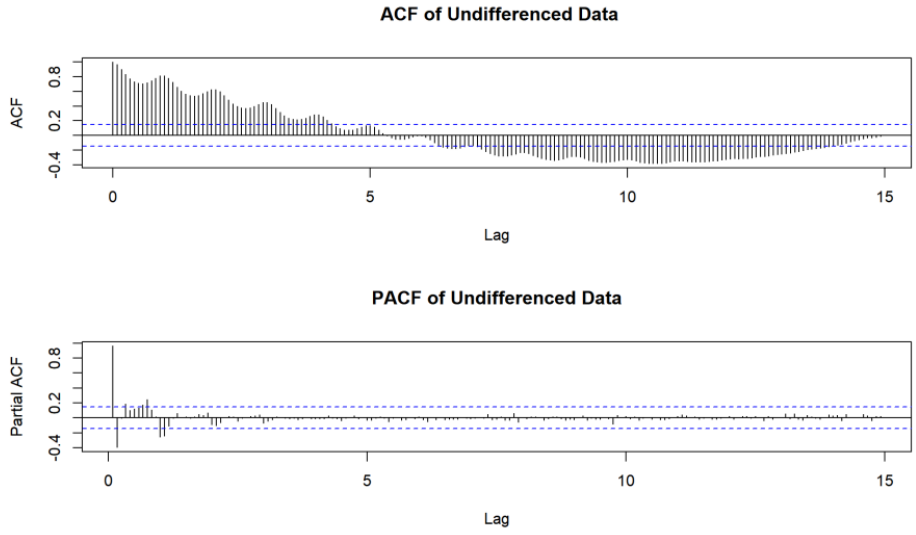

figure 3.2.1

In *figure 3.2.2*, the time series of the newly differenced data ($\nabla_{12}Y_t$) is presented. In this data, we observed a clear upward trend in the data. We differenced the data one more time ($\nabla\nabla_{12}Y_t$) to remove this seasonality. When we differenced the data, we saw a decrease in variance from 0.373 to 0.163. In addition, there appeared to be no trend this differenced time series data. To confirm that we differenced the appropriate amount of times, we checked if differencing an additional time ($\nabla^2\nabla_{12}Y_t$) would reduce variance. The variance increased from 0.163 to 0.386. Because the variance increased, our group decided to use the data differenced as follows:

$$\nabla\nabla_{12}Y_t$$

The differenced time series plot is presented in *figure 3.2.*3 and the ACF and PACF corresponding to this differenced data is presented in *figure 3.2.4*.

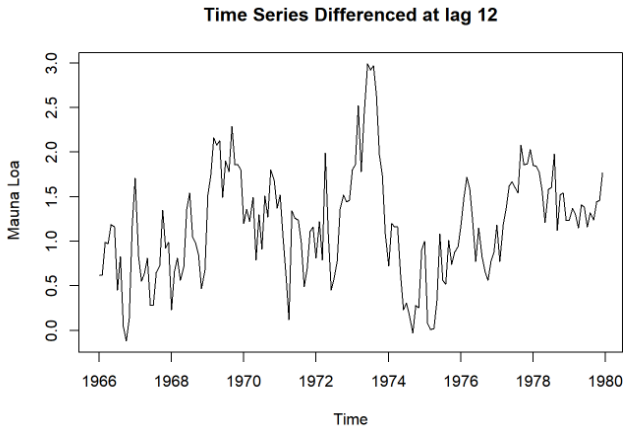

figure 3.2.2

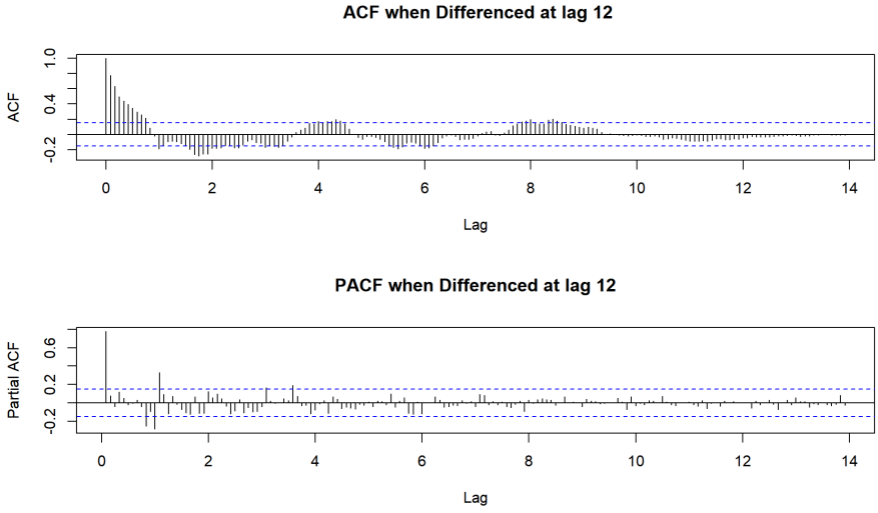

figure 3.2.3

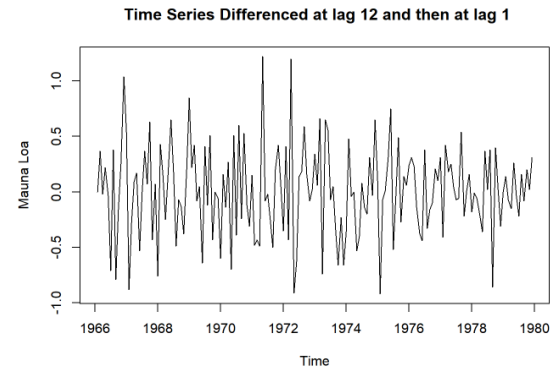

figure 3.2.4

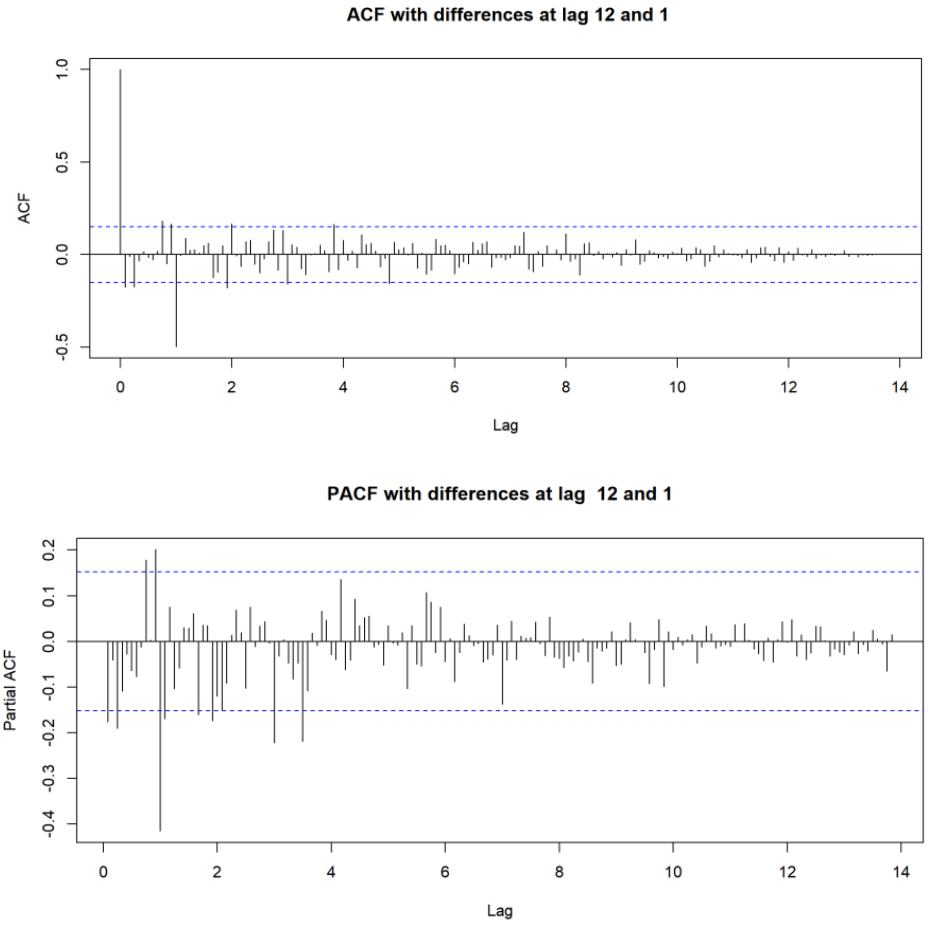

figure 3.2.5

Relative to our original time series graph in *figure 3.2.6*, we have successfully de-trended as well as removed the seasonality of the data. After establishing our desired data, we can fully dissect the ACF and PACF plot and are now ready to move to preliminary model

selection. We suspect a SARIMA model would be appropriate given the need of differencing to get rid of seasonality.

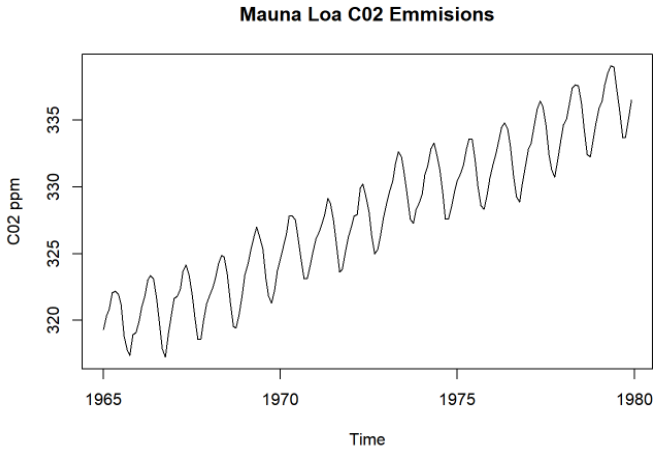

figure 3.2.6

SECTION 4– MODEL BUILDING AND SELECTION

The first step in choosing potential SARIMA (p,d,q) x $(P,D,Q)_{12}$ models was to observe points along the ACF and PACF plots having values outside the 95% Confidence interval. From observing *figure 4.1*, we noticed at lags 1/12 and 3/12 both ACF and PACF plots are outside the 95% confidence interval. This can suggest models with p=0 to 3 which would be an ARIMA (0 → 3, 1, 0) model, as well as models with q=0 to 3 ARIMA (0, 1, 0 → 3) or perhaps a mixed model with p and q taking on an integer value from 0 to 3. Lag 1 is outside the 95% confidence interval for the ACF and PACF, meanwhile Lag 3 is outside the confidence interval for the PACF plot. This suggests a Q=1 model, or a mixed model with P=1 and Q=1. From the last step, it is certain the SARIMA model has values of d=1 and D=1 because of the number of differences with each lag 1 and 12 respectively.

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2k^2 + 2k}{n - k - 1}$$

figure 4.1

To assess the best capable model, we ran a Akaike's Information Criterion Corrected for Bias (AIC_C). AIC_C is a method which employs the use of maximum likelihood estimators (MLE) and the number of parameters used for the model and the number of data points in the data set. Ideally, the model with the lowest AICC values are considered the best fit.

According to the table of AICc table created the best two models are represented by a SARIMA (1,1,3) x $(0,1,1)_{12}$ and SARIMA (0,01,1) x $(0,1,1)_{12}$. Both models will go through diagnostic checking to see if any assumptions are violated. The method of the AIC model selection assumes the data at be normally distributed with a mean of 0, have constant variance, and be independent. Using the AICc in the figure below we have the following:

```
##           p q P Q     aicc
## AiccRow   0 3 0 1 91.47572
## AiccRow.1 0 1 0 1 91.53033
## AiccRow.2 1 1 0 1 92.32254
## AiccRow.3 1 0 0 1 92.45189
## AiccRow.4 2 3 0 1 92.66945
## AiccRow.5 2 1 0 1 92.93756
```

To choose the optimum model, we chose the model with the minimum AIC as our best model, but due to the rule of parsimony we selected the model with the second lowest AICc because it had fewer parameters. The model chosen was SARIMA (0,01,1) x $(0,1,1)_{12}$.

$$X_t = (1 - 0.3154B)(1 - 0.9991B^{12})Z_t$$

SECTION 5 – DIAGNOSTICS

With our final model of SARIMA (0, 1, 1) x $(0, 1, 1)_{12}$ from the model selection, we conducted diagnostic checks to determine whether our forecasted estimates were valid or not. We conducted analysis on the model residuals to verify normality, independence, and constant variance.

We checked the normality assumption by employing the use of a Q-Q plot, where theoretical quantiles were plotted against observed quantiles. Once we plotted the residuals, we compared expected deviations from sample deviations at a 45-degree line. In *figure 5.1* below, we found that the residuals possessed a smooth-positive trend along the Q-Q line. Except for one outlier, we concluded that for the most part, the residuals were normally and identically distributed since there was not an extremity of outliers.

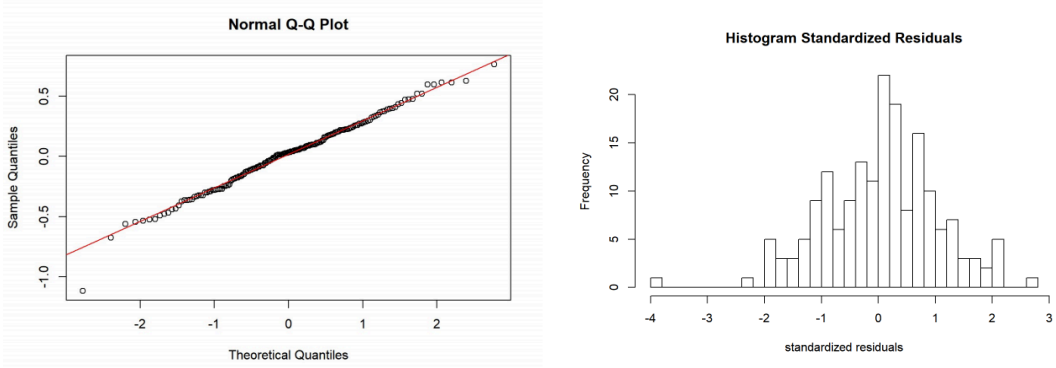

figure 5.1

A Shapiro-Wilk test was also conducted to cross validate the normality assumption. We observed a p-value of .2322 associated with the W-test statistic seen in *figure 5.2*, this was significant and therefore we failed to reject the null hypothesis of the data being normally distributed. This cross validated the outcome the Q-Q plot produced.

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

figure 5.2

Next, we checked independence by conducting a modified Box-Pierce test, also known as the Ljung-Box test. The test worked to check the autocorrelation structure of residuals with the null hypothesis in *figure 5.3*. The alternative hypothesis was that at least one autocorrelation structure was not zero, meaning at least one structure was dependent.

$$H_0 : \rho_1 = \rho_2 = \Lambda = \rho_K = 0$$

figure 5.3

The Ljung-Box test yielded a p-value of .3226 at a level of confidence of 95%. We concluded there was not enough evidence to reject the null the hypothesis. Autocorrelation structure of residuals was then deemed independent based on the evidence of the Ljung-Box test.

$$Q = n(n+2)\sum_{k=1}^{k}(n-k)^{-1}\hat{\rho}_k^2 \;\; ; \; Q \overset{approx.}{\sim} \chi^2_{K-m}$$

With normality and independence diagnostics being passed, a heteroscedasticity check was used to diagnose the constant variance assumption. A violation of this assumption would mean variance of the ordinary least squares estimator is biased for the true variance. We tested this by referring to the ACF and PACF plots of the residuals and constructing a confidence interval.

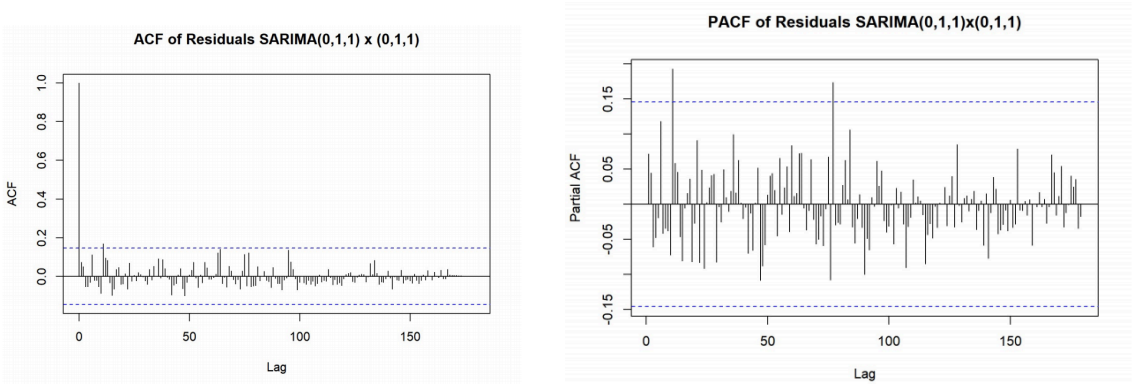

figure 5.4

In *figure 5.4* we observed that most lags were within the 95% confidence intervals. The PACF of residuals had two lags outside the interval, however the lags did not follow a trend of repeatedly moving outside the interval and we concluded these occurrences were due to random

chance. The ACF also checked out except for one lag insignificantly going out of bounds. In addition, the variance was constant. We then concluded the assumptions of normality, constant variance, and independence were accounted for in the model. With all diagnostics passed, we then concluded the SARIMA (0, 1, 1) x $(0, 1, 1)_{12}$ model was well equipped to handle forecasting with no distorted estimates.

SECTION 6 – FORECASTING

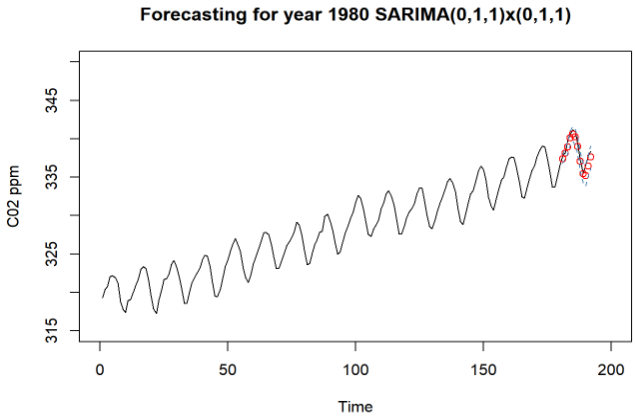

figure 6.1

The SARIMA (0,1,1) x (0,1,1) model was created using monthly data from 1965 to the year 1979. Using the formula from the estimated coefficients the following forecast is made for the year 1980. The black line represents the true values found in the data set, while the red dots represent the predicted values based on the SARIMA model and the blue lines represent a 95% around the predicted values. As can be seen in *figure 6.1* the model did well in predicting future values as the predicted values were very close to the actual values. Furthermore, by looking at *figure 6.2*, one can see all true data points for the year 1980 fall in the 95% confidence interval or very well near it.

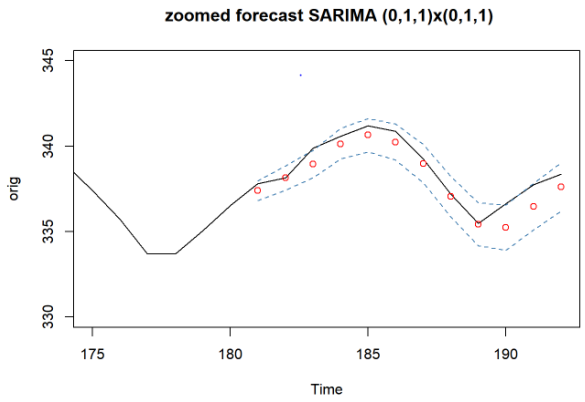

figure 6.2

SECTION 7 – CONCLUSION

With our model of SARIMA (0, 1, 1) x $(0, 1, 1)_{12}$ we estimated CO_2 emission levels for the last 12 data points by comparing forecasted values to observed values. We analyzed the initial raw time series by using the ACF and PACF plot of the data. We then transformed the data into possible transformed sets and concluded that no transformation was necessary as the raw data was already in an optimal state. We then differenced the data once to remove the trend component although there were still signs of the apparent seasonality. We differenced the data at lag 12 and, this time, the seasonality along with trend was fully removed, leading to a stationary data set. Further analysis of the ACF and PACF plots led to suspecting models that can best be used to fit the data and forecast accurately. Using a model selection process by method of AICC, we concluded two models that would be accurate. Instead of choosing the model with the lowest AICC, we opted with the model with the lowest assumption by law of parsimony. Running this model through diagnostics, we found evidence that the model did indeed follow a normal distribution by tests, such as Shapiro-Wilk and Q-Q plot. The model also passed tests for identicality by Ljung-Box test and constant variance by constructing an interval on the ACF and PACF of the model residuals.

We concluded the model had a constant variance by the previous tests, and thus the model had passed the diagnostics; forecasting was sure to yield undistorted estimates. Figures 3.8 and 3.9 in the forecast section proved that our model was indeed correct as our estimated values were within the 95% confidence interval compared to actual values in 1980. Further analysis that can be conducted in this model for future study is the incorporation of spectral analysis as well as a periodogram.

8 REFERENCES

https://datamarket.com/data/set/22v1/co2-ppm-mauna-loa-1965-1980#!ds=22v1&display=line

The dataset used was “CO2 (ppm) Mauna Loa, 1965-1980.”

[1] https://www.esrl.noaa.gov/gmd/ccgg/about/co2_measurements.html

Detailed information about the Mauna Loa Observatory and how it collects CO_2 in the atmosphere.

[2] https://www.nps.gov/articles/volcano-monitoring.htm

An article by the National Park Service on volcanic monitoring, topics include ecosystems and methods of measurements.

[3] https://volcanoes.usgs.gov/vhp/gas.html

United States Geological Survey on C02 effects on the environment.

[4] https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2002JB002293

Journal of Geophysical Research, study on why volcanoes exhibit seasonality.

[5] http://www.stat.pitt.edu/stoffer/tsa4/

E-Text book: “Time Series Analysis”

9 APPENDIX (R CODE)

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(MASS)
library(forecast)
library(astsa)
library(dse)
library(readr)
library(dplyr)
library(qpcR)
ExRate.csv = read.table("Maunaloa.csv", sep=",", header=FALSE, skip=1, nrows=192)
size= length(ExRate.csv[,2])
Train=as.data.frame(ExRate.csv[1:(size-12),2])
hist(ExRate.csv[1:180,2], main=" Mauna Loa C02 Emissions ppm", xlab="C02 ppm")
# size of data set is 192
Test=as.data.frame(ExRate.csv[180:192,2])
ExRate = ts(Train, start = c(1965,1), frequency = 12)
ts.plot(ExRate, main="Mauna Loa C02 Emmisions",xlab="Time",  ylab="C02 ppm")
#checking for distribution
t = 1:length(ExRate)
fit = lm(ExRate ~ t)
bcTransform = boxcox(ExRate ~ as.numeric(t),plotit=T,lambda=c(-10,10,0.1))
qqnorm(ExRate, main=" Normal Q-Q plot for original data")
qqnorm(log(ExRate), main="Normal Q-Q plot for log(data) ")
# checking box-cox transform
lambda = bcTransform$x[which(bcTransform$y == max(bcTransform$y))]
lambda
ExRate.bc = (1/lambda)*(ExRate^lambda-1)
qqnorm(ExRate.bc, main="Normal Q-Q plot Box-Cox")
# cheching other transforms
op <- par(mfrow = c(2,3))
```
```

```
ExRate.Sq = sqrt(ExRate)

ExRate.Iv = 1/ExRate

ts.plot( ExRate,main     = "Original data",ylab="C02 ppm")

ts.plot( ExRate.bc ,main = "Box-Cox tranformed data", ylab="Box-Cox( C02 ppm)")



ts.plot(log(ExRate),main="Log Transformed Data", ylab="log(CO2) ppm")

ts.plot( ExRate.Sq,main     = "Square data", ylab=expression(sqrt("C02 ppm")))

ts.plot( ExRate.Iv ,main = "Inverse tranformed data", ylab = "1/C02 ppm")

par(op)

BOXTransforms=c("Original", "BoX-COX Best", "Log","Sqrt","Inverse")

TransformVariances=c(var(ExRate),var(ExRate.bc), var(log(ExRate)), var(sqrt(ExRate)),var(1/ExRate))

VarianceAfterBox=tibble(Name=BOXTransforms,Variance=TransformVariances)

VarianceAfterBox

orig.ap =  par(mfrow = c(2,1))

acf(ExRate,lag.max = 182,main = " ACF of Undifferenced Data")

pacf(ExRate,lag.max = 182,main = "PACF of Undifferenced Data")

par( orig.ap)

Exrate.dif12 <- diff(ExRate,lag=12)

ts.plot(Exrate.dif12,main = "Time Series Differenced at lag 12", ylab=" Mauna Loa")

abline(Exrate.dif12~as.numeric(1:182), col="red")

var(ExRate)

var(Exrate.dif12)

dif12 <- par(mfrow = c(2,1))

acf (Exrate.dif12, lag.max = 192,main = " ACF when Differenced at lag 12")

pacf(Exrate.dif12, lag.max = 192,main = "PACF when Differenced at lag 12")

par(dif12)

Exrate.dif13 <- diff(Exrate.dif12,lag=1)

ts.plot(Exrate.dif13,main=" Time Series Differenced at lag 12 and then at lag 1" ,ylab="Mauna Loa")

var(Exrate.dif13)

Exrate.dif14=diff(Exrate.dif13)

var(Exrate.dif14)
```

```
var(Exrate.dif13)>=var(Exrate.dif14)
op2 <- par(mfrow = c(2,1))
acf(Exrate.dif13 ,lag.max = 192,main = " ACF with differences at lag 12 and 1")
pacf(Exrate.dif13,lag.max = 192,main = "PACF with differences at lag  12 and 1")
par(op2)
ocloser <- par(mfrow = c(1,2))
acf(Exrate.dif13 ,lag.max = 48,main = "ACF with differences at lag  12 and at lag 1")
pacf(Exrate.dif13,lag.max = 48,main = "PACF with differences at lag  12 and at lag 1")
par(ocloser)
var(Exrate.dif13)
var(diff(Exrate.dif13))
DiffVariances=c(var(ExRate),var(Exrate.dif12), var(Exrate.dif13), var(Exrate.dif14))
DifferenceName=c("Original","diff(1) @ lag12","diff 1 @ lag 1&12", "diff 1 @ lag12 and  diff(2)@ lag1")
VarianceAfterBox=tibble(Name=DifferenceName,Variance=DiffVariances)
VarianceAfterBox
AICctable=c()
for (s in c(0,1))
 {
  for (r in c(0,1,3))
  {
    for(j in c(0:3))
    {
     for (i in c(0:3))
     {
      AICcValue=(try(AICc(arima(Train, order = c(i,1,j),seasonal = list(order = c(r, 1, s), period = 12))), silent =
TRUE))
     if(class(AICcValue)=="try-error"){
      AICcValue=NA
     }
     AiccRow=c(i,j,r,s,AICcValue)
     AICctable=rbind(AICctable,AiccRow)
```

```
        }
      }
    }
  }

  AICtablesort=as.data.frame(AICctable)


  names(AICtablesort)=c("p","q", "P","Q","aicc")
  AICtablesort=AICtablesort[order(AICtablesort["aicc"]),]
  head(AICtablesort)
  plot.roots <- function(ar.roots=NULL, ma.roots=NULL, size=2, angles=FALSE, special=NULL,
sqecial=NULL,my.pch=1,first.col="blue",second.col="red",main=NULL)
  {xylims <- c(-size,size)
  omegas <- seq(0,2*pi,pi/500)
  temp <- exp(complex(real=rep(0,length(omegas)),imag=omegas))
  plot(Re(temp),Im(temp),typ="l",xlab="x",ylab="y",xlim=xylims,ylim=xylims,main=main)
  abline(v=0,lty="dotted")
  abline(h=0,lty="dotted")
  if(!is.null(ar.roots))
  {
   points(Re(1/ar.roots),Im(1/ar.roots),col=first.col,pch=my.pch)
   points(Re(ar.roots),Im(ar.roots),col=second.col,pch=my.pch)
  }
  if(!is.null(ma.roots))
  {
   points(Re(1/ma.roots),Im(1/ma.roots),pch="*",cex=1.5,col=first.col)
   points(Re(ma.roots),Im(ma.roots),pch="*",cex=1.5,col=second.col)
  }
  if(angles)
  {
   if(!is.null(ar.roots))
```

```
 {
  abline(a=0,b=Im(ar.roots[1])/Re(ar.roots[1]),lty="dotted")
  abline(a=0,b=Im(ar.roots[2])/Re(ar.roots[2]),lty="dotted")
 }
 if(!is.null(ma.roots))
 {
  sapply(1:length(ma.roots), function(j) abline(a=0,b=Im(ma.roots[j])/Re(ma.roots[j]),lty="dotted"))
 }
}
if(!is.null(special))
{
 lines(Re(special),Im(special),lwd=2)
}
if(!is.null(sqecial))
{
 lines(Re(sqecial),Im(sqecial),lwd=2)
}
}


spec.arma <- function(ar=0,ma=0,var.noise=1,n.freq=500, ...)
{
 # check causality
 ar.poly <- c(1, -ar)
 z.ar <- polyroot(ar.poly)
 if(any(abs(z.ar) <= 1)) cat("WARNING: Model Not Causal", "\n")
 # check invertibility
 ma.poly <- c(1, ma)
 z.ma <- polyroot(ma.poly)
 if(any(abs(z.ma) <= 1)) cat("WARNING: Model Not Invertible", "\n")
 if(any(abs(z.ma) <= 1) || any(abs(z.ar) <= 1) ) stop("Try Again")
 #
```

```
 ar.order <- length(ar)
 ma.order <- length(ma)
 # check (near) parameter redundancy [i.e. are any roots (approximately) equal]
 for (i in 1:ar.order) {
  if ( (ar == 0 & ar.order == 1) || (ma == 0 & ma.order ==1) ) break
  if(any(abs(z.ar[i]-z.ma[1:ma.order]) < 1e-03)) {cat("WARNING: Parameter Redundancy", "\n"); break}
 }
 #
 freq <- seq.int(0, 0.5, length.out = n.freq)
 cs.ar <- outer(freq, 1:ar.order, function(x, y) cos(2 *
                                        pi * x * y)) %*% ar
 sn.ar <- outer(freq, 1:ar.order, function(x, y) sin(2 *
                                        pi * x * y)) %*% ar
 cs.ma <- outer(freq, 1:ma.order, function(x, y) cos(2 *
                                        pi * x * y)) %*% -ma
 sn.ma <- outer(freq, 1:ma.order, function(x, y) sin(2 *
                                        pi * x * y)) %*% -ma
 spec <- var.noise*((1 - cs.ma)^2 + sn.ma^2)/((1 - cs.ar)^2 + sn.ar^2)
 spg.out <- list(freq=freq, spec=spec)
 class(spg.out) <- "spec"
 plot(spg.out, ci=0, main="", ...)
 return(invisible(spg.out))
}

fit_sarima01.01 = arima(Train, order = c(0,1,1), seasonal=list(order=c(0,1,1),period=12))
AICc(fit_sarima01.01)
summary(fit_sarima01.01)

acf (residuals(fit_sarima01.01)^2,main = "ACF of Residuals SARIMA(0,1,1) x (0,1,1)",lag.max = 192)

pacf(residuals(fit_sarima01.01)^2,main = "PACF of Residuals SARIMA(0,1,1)x(0,1,1)",lag.max = 192)


shapiro.test(residuals(fit_sarima01.01))
```

```
qqnorm(residuals(fit_sarima01.01))

qqline(residuals(fit_sarima01.01),col="red")

hist(residuals(fit_sarima01.01)/(var(residuals(fit_sarima01.01))^0.5), breaks=30,xlab="standardized residuals",
main="Histogram Standardized Residuals")


Box.test(residuals(fit_sarima01.01),lag = 13, type = "Box-Pierce",fitdf = 2)

Box.test(residuals(fit_sarima01.01),lag=13, type = "Ljung-Box",fitdf=2)

oroots <- par(mfrow = c(1,2))

plot.roots(NULL,polyroot(c(1,-.3154))  , main="roots of nonseasonal MA  ")

polyroot(c(1,-0.3154))

plot.roots(NULL,polyroot(c(1,-0.991)), main="roots of  seasonal MA")

polyroot(c(1,-0.991))

par(oroots)

myforecast= predict(fit_sarima01.01,n.ahead=12)

original=as.data.frame(ExRate.csv[1:192,2])

testset=(ExRate.csv[181:192,2])

orig=ts(original)

ts.plot(orig, xlim=c(0,200), ylim=c(315,350) ,main=" Forecasting for year 1980 SARIMA(0,1,1)x(0,1,1) ", ylab="C02
ppm")

points(181:192,myforecast$pred,col="red")

lines(181:192,myforecast$pred+1.96*myforecast$se,lty=2, col="steelblue")

lines(181:192,myforecast$pred-1.96*myforecast$se,lty=2,col="steelblue")

# zoomed in version of data

ts.plot(orig, xlim=c(175,192), ylim=c(330,345), main=" zoomed forecast SARIMA (0,1,1)x(0,1,1)")

points(181:192,myforecast$pred,col="red")

lines(181:192,myforecast$pred+1.96*myforecast$se,lty=2, col="steelblue")

lines(181:192,myforecast$pred-1.96*myforecast$se,lty=2,col="steelblue")
```