Justin Nichols[1]

Time Series and Its Applications—MATH 598

Dr. Soutir Bandyopadhyay

Due: November 16^{th}, 2021

Midterm 2

"Don't gain the world and lose your soul, wisdom is better than silver and gold."

-Bob Marley

[1]jnichols2@mymail.mines.edu

1. **(All students must attempt Problem 1.)**

Concentration of CO2 in the atmosphere, which is the primary cause of global warming, has now reached an unprecedented level. In March 2015, the average of all of the global measuring sites showed a concentration above 400 parts per million (ppm). This follows the individual observatory high points of 400 ppm in 2012 at the Barrow observatory in Alaska, and the 2013 high of 400 ppm at the Mauna Loa observatory in Hawaii. Mauna Loa has been running consistently above 400 ppm since late 2015. Scientists advising the United Nations recommend the world should act to keep the CO2 levels below 400-450 ppm in order to prevent even more irreversible and disastrous climate change effects.

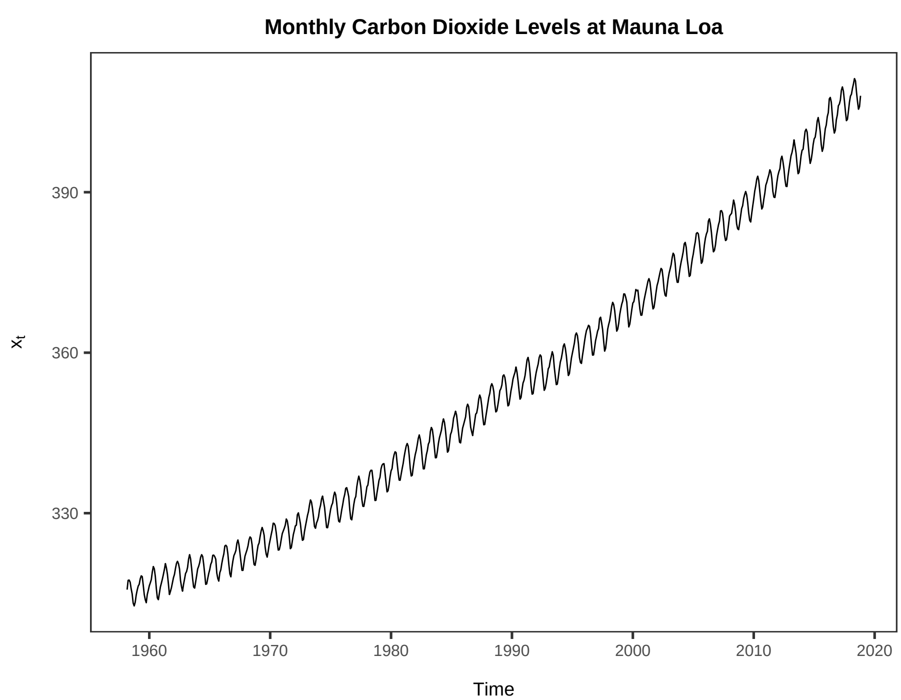
In this problem, you need to work with the `cardox` data set in the R package `astsa`. The data are the CO2 readings, say x_t, from March 1958 to November 2018 at the Mauna Loa observatory, which is the oldest continuous monitoring station of carbon dioxide.

Run a thorough time domain analysis of the data by following necessary steps (e.g., including, but not limited to, plotting the data, examining the ACF/PACF plots, identifying any seasonality, fitting a model, and running model diagnostics etc.) as discussed in class.

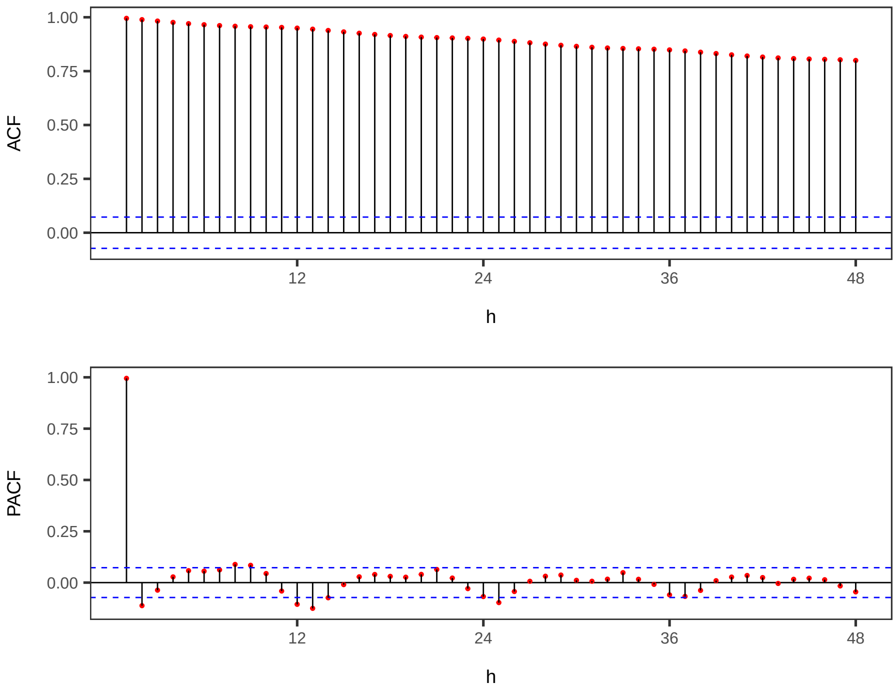
With your fitted model provide forecast out five years and finally provide a scientific conclusion based on your results.

As we should always do initially to get an idea of the evolution of the data we plot it over time.

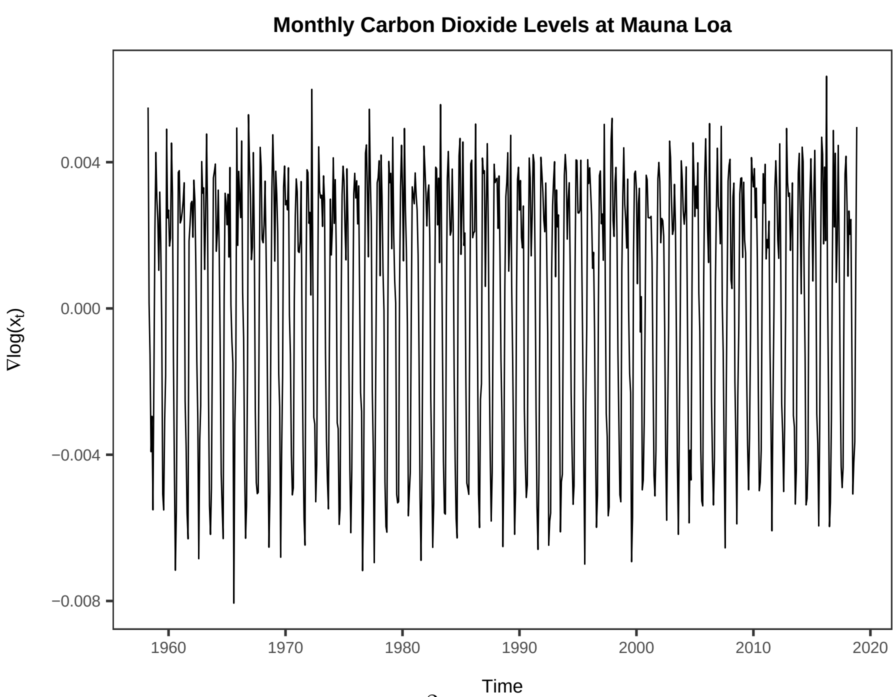
```
xt <- cardox
```

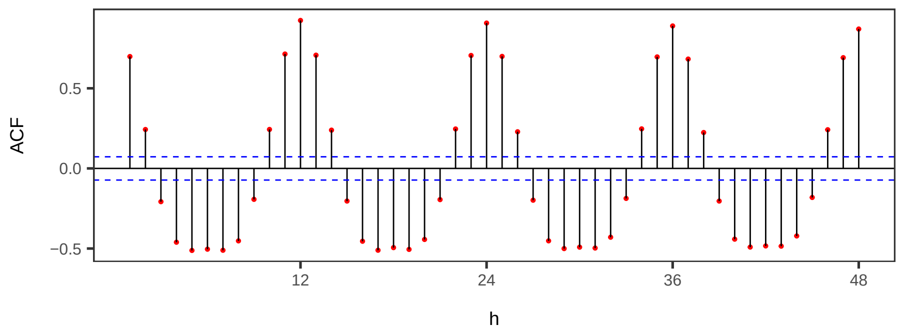


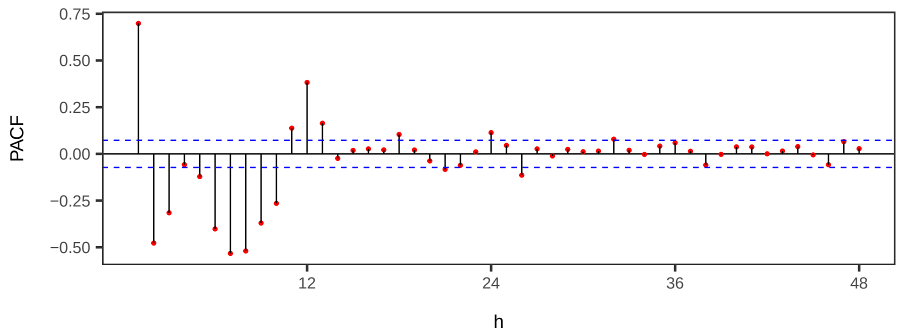
Let's take a look at the ACF and PACF for x_t.



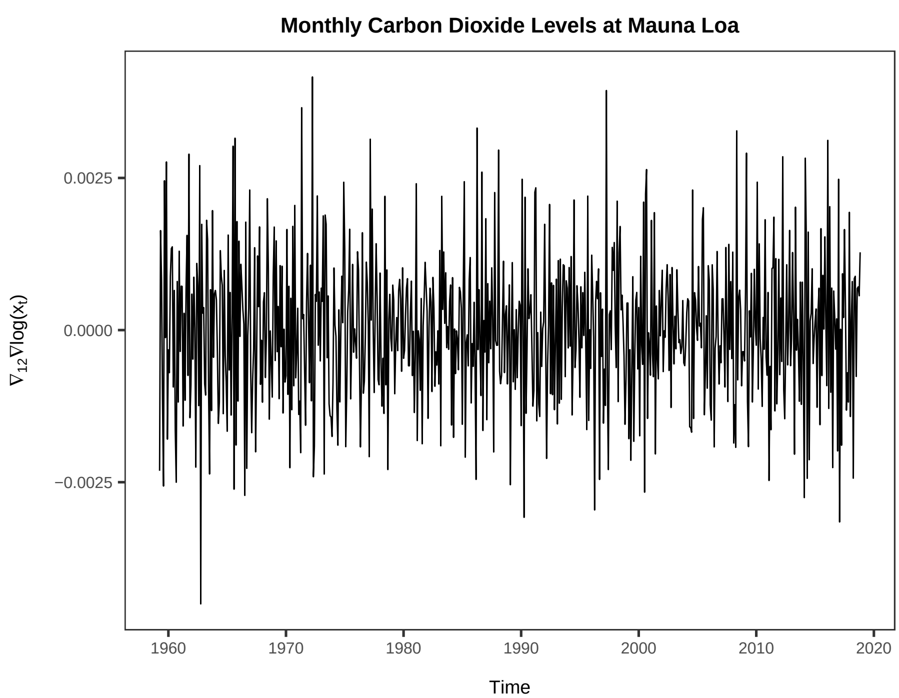
Neither cut off at a particular value. Also, there doesn't look to be changes to the variance over time, but there does seem to be a linear trend and a large range. So, we take the first difference to remove the trend, take the logarithm to scale, and then plot the result. Then, we will look at the ACF and PACF.

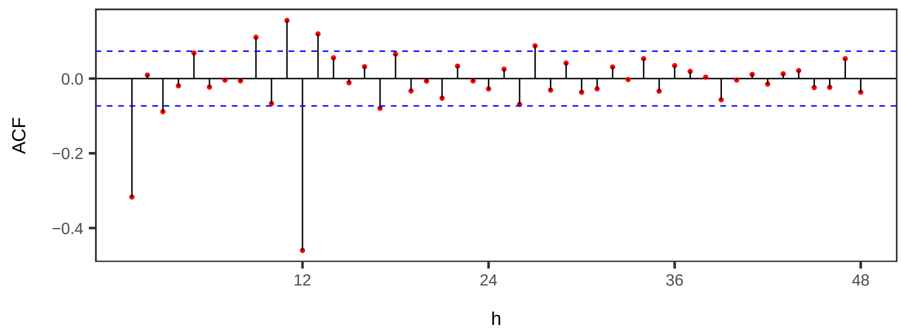


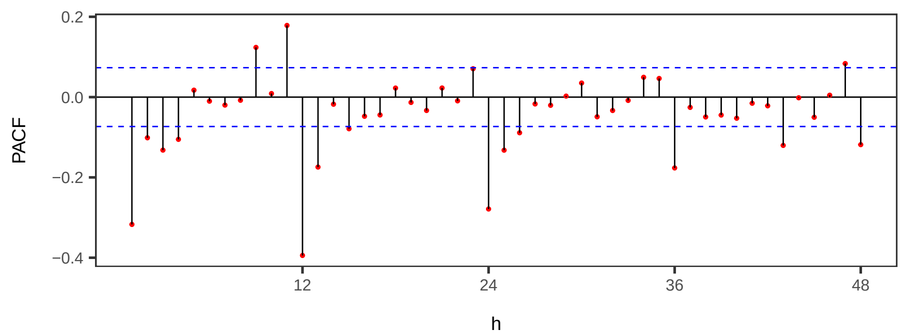




We see that there tends to be similar values according to s and increments of multiples of s. Therefore, we take $D = 1$ and similarly plot the resulting time series and ACF and PACF.







We also can look at the Dickey-Fuller test to confirm this provides us the stationarity we are looking for.

```
adf.test(diff(diff(lxt, lag=12)))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(diff(lxt, lag = 12))
## Dickey-Fuller = -8.7698, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary
```

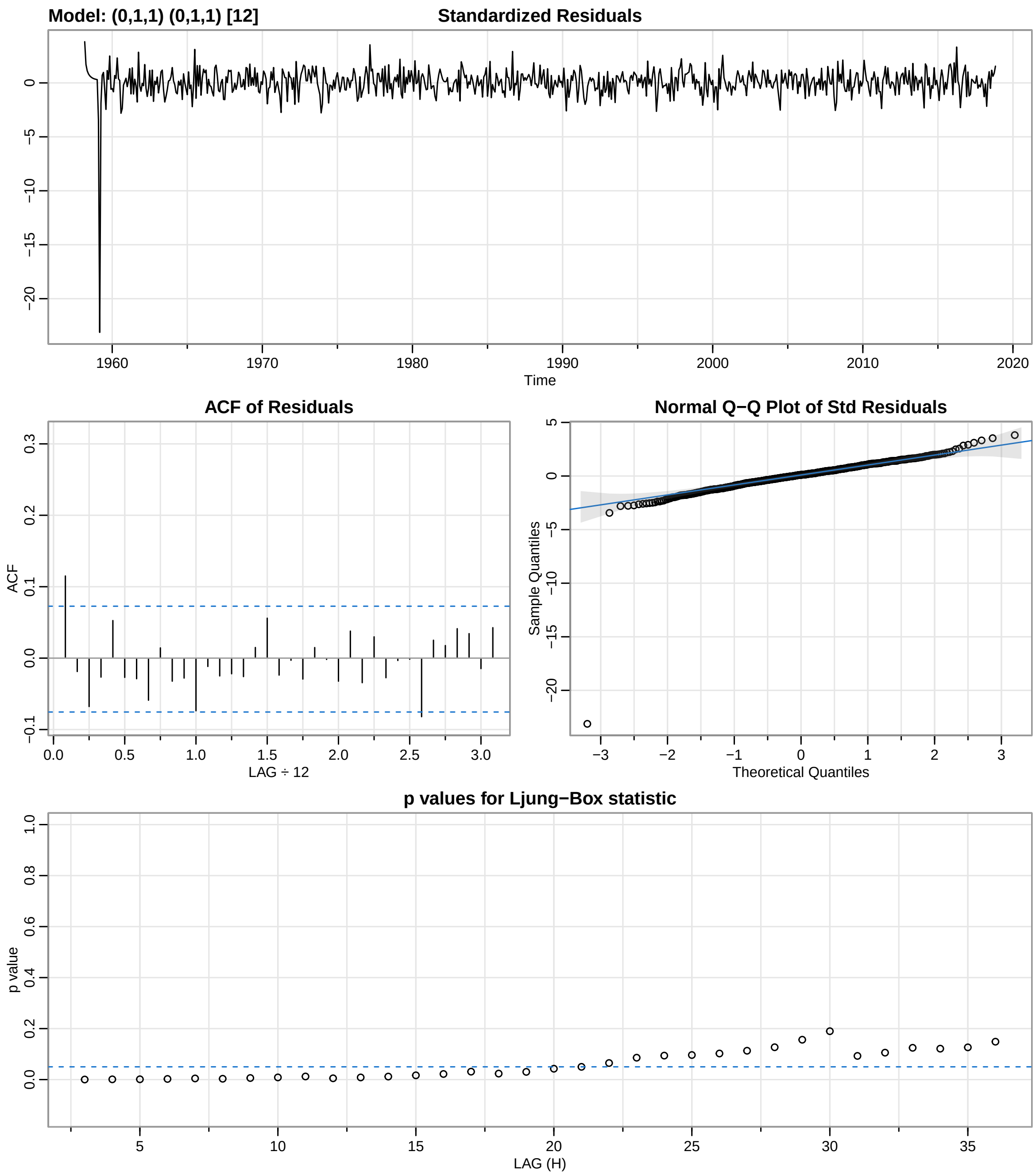
Referencing the ACF and PACF above, we now are looking at an ARIMA$(0,1,0) \times (0,1,0)_{12}$ model. We now want to analyze for the seasonal (P,Q) and non-seasonal components (p,q). Observing the two plots at $1s, 2s, 3s, 4s$ with $s = 12$ we see that the ACF seems to cut off at $1s$, but the PACF just slowly decays. Therefore, it seems reasonable that this would imply $P = 0$ and $Q = 1$. In the non-seasonal part it's not definitive so we will assume $P = 0$ and $Q = 1$ and run a grid search in order to determine which variation of p, q goes best with this assumption. To do this, we will use the BIC metric and find the minimum value.

```
s <- 12; d <- 1; D <- 1; P <- 0; Q <- 1
BICtable <- matrix(NA, 5, 5) %>%
  set_colnames(paste('q =', 0:4)) %>%
  set_rownames(paste('p =', 0:4))
for (p in 0:4) {
  for (q in 0:4) {
    temp_arima <- sarima(lxt, p, d, q, P, D, Q, s, details=FALSE)
    BICtable[p+1, q+1] <- temp_arima$BIC
  }
}
# subtract 1 from index to match with p and q
(p_q <- c(which(BICtable==min(BICtable), arr.ind=TRUE)) - 1)
```

```
## [1] 0 1
```

We now plot the overall diagnostics of the overall chosen model ARIMA$(0,1,1) \times (0,1,1)_{12}$.

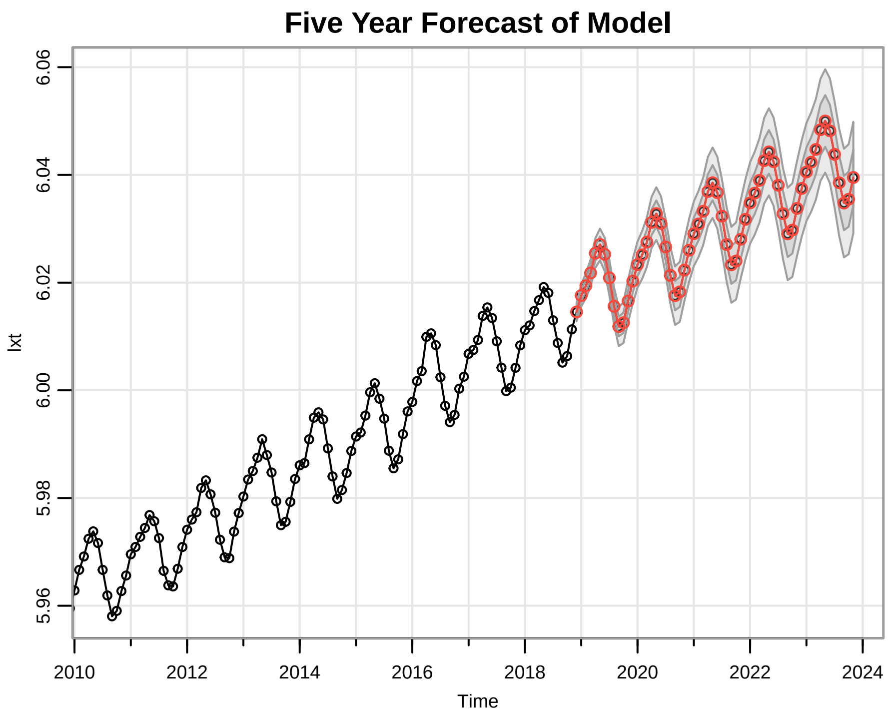
```
p <- p_q[1]; q <- p_q[2]
final_model <- capture.output(sarima(lxt, p, d, q, P, D, Q, s))
```



Judging by the diagnostics above, the model does well except small errors such as a weird anomaly at the beginning of the data set, which may have to do with the issues at small lag in the ACF plot. We now use this model to forecast 5 years into the future. In other words, $5s = 5(12) = 60$ months.

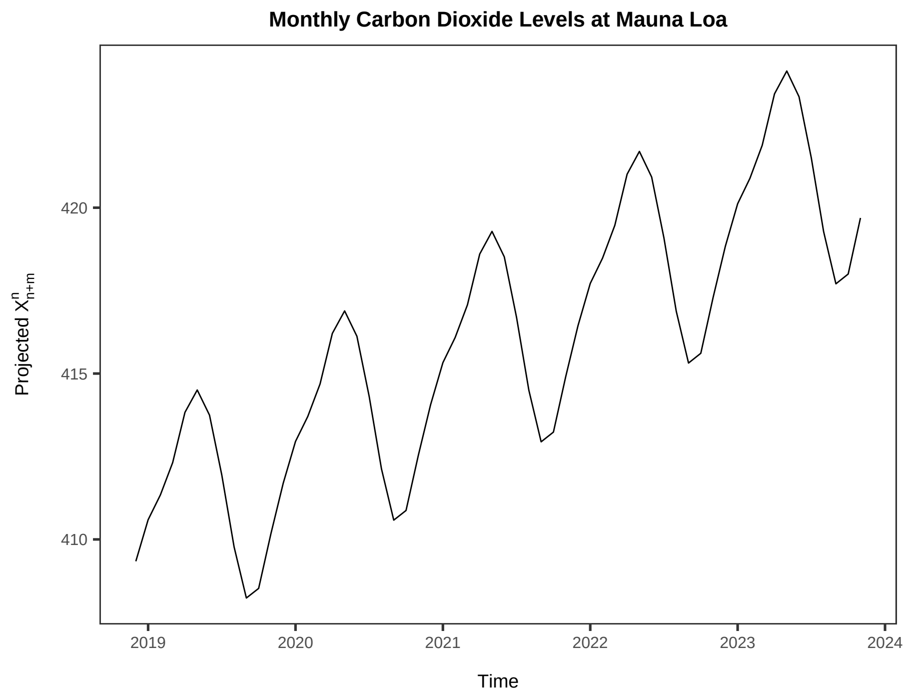
In order to do this, we use `sarima.for`, where we add in the number of months we would like to forecast, m.

```
m <- 60
par(cex=0.75, cex.sub=0.75, cex.main=0.35, cex.lab=0.75, cex.axis=0.75)
lxt_forecast <- sarima.for(lxt, m, p, d, q, P, D, Q, s, main='Five Year Forecast of Model')
```



Now, lets just plot the prediction but use it as the power of e, so that we get the non-transformed values. We won't look at the prediction errors, just the predicted values. This will give us an idea of where we project the ppm to be in about 5 years. With this information, we can give an educated guess to whether we will be encroaching on the danger region of being closer to 450.

Here is that plot of projected non-transformed average CO2.



My ultimate, and what should be obvious conclusion, would be that according to this projection, if not in the next 5 years we will start to get too close to that 450 danger zone we will be in 10-20 years most likely. There doesn't seem to be any stopping and even in this 5 year projection we are attaining a peak of 424.1207 ppm, which is right in the middle of the weary 400-450 range.