Stat 463 Final Project

Gustav Vu and Peidong Liu

2024-11-12

Load Data

Clean Data

Introduction

For decades, the scientific community, policymakers, and global organizations have been grappling with the pressing issue of global warming, aiming to identify, develop, and implement the most effective strategies to mitigate its effects. One of the most critical steps in addressing this challenge is to thoroughly investigate its underlying causes, particularly the factors contributing to the increased concentration of carbon dioxide (co_2) in the atmosphere. As a primary greenhouse gas, co_2 plays a pivotal role in the intensification of the greenhouse effect, which has directly and indirectly led to a steady rise in global average temperatures. Understanding the historical trends and sources of co_2 emissions is crucial to comprehending the broader dynamics of climate change.

This research specifically focuses on analyzing the significant boom in co_2 emissions that occurred during the Second Industrial Revolution, a period marked by rapid industrialization, technological advancements, and fossil fuel exploitation. By evaluating the socio-economic activities and technological developments of this transformative era, we aim to uncover the extent to which these factors contributed to the expansion of greenhouse gas levels and how they set the stage for the modern challenges of global warming. Through this analysis, we hope to provide insights into historical emission trends and their implications for current and future efforts to combat climate change.

Background

Global warming common issue

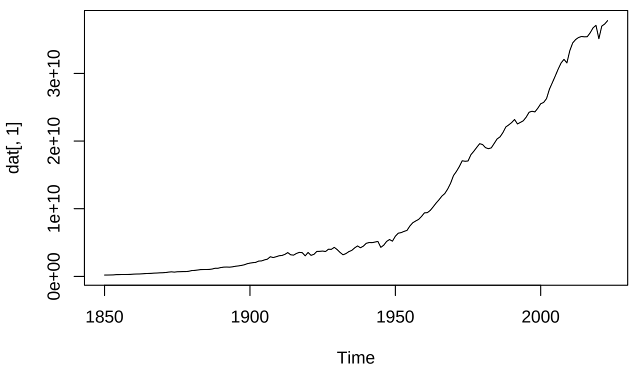
The Intergovernmental Panel on Climate Change (IPCC) issued a report in 1990 utilizing global mean near surface temperature which raised concern on the rate of increasing. [5]. This concern raised the question of further research on greenhouse effect of the

Anthropological contribution on the emission of carbon dioxide that mainly draws from the effect of human activity [3] on a large scale acts as the main motive of the extreme spur especially after the second industrial revolution era.[2] Predominantly starts from western europe and north america where both regions exceeded 10 ton per capita as from the region report (Ritchie, Rosado, and Roser 2023). A combination of boost in different industries which is facilitated by the second industrial revolution including chemical and transportation domain [4] after the mid 1880s followed by the expansion boom leads to the official start of boosting the aggregate global carbon dioxide emission. This major event placed a major factor on the co_2 emission which reflected on the overall data trend in terms of human intervention.

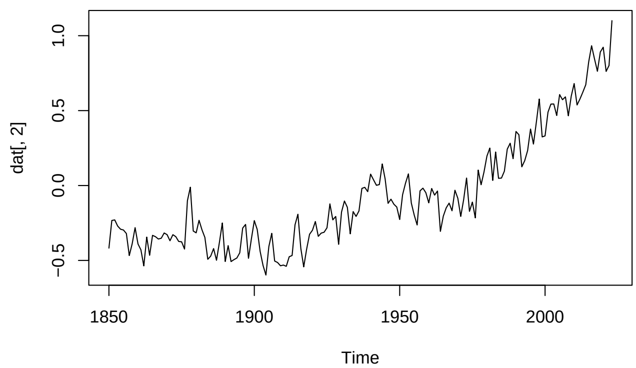
Method

In this study, we determine the effect of co_2 on temperature and compare the trend found through constructing a time series model for forecasting and the Arch-Garch method for taking care of the variance change aspect as time progresses. We use linear regression and Auto-Aggressive-Moving-Average(ARIMA) models for stationary and stochastic respectively. The model takes in consideration the data behavior change due to real world events.

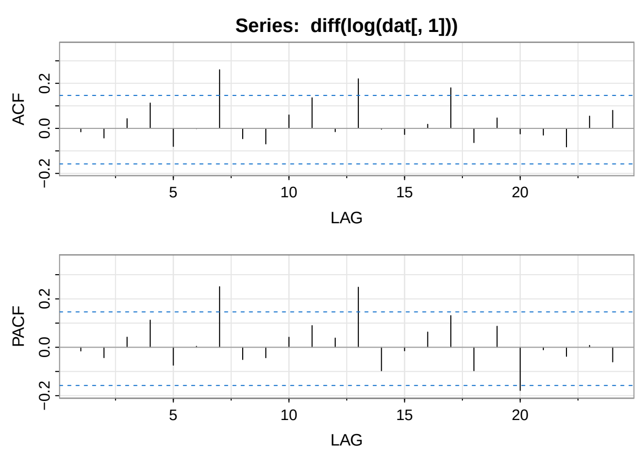
```
##       [,1]  [,2] [,3] [,4]  [,5] [,6] [,7]  [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF  -0.02 -0.04 0.04 0.11 -0.08    0 0.26 -0.05 -0.07  0.06  0.14 -0.01  0.22
## PACF -0.02 -0.04 0.04 0.11 -0.07    0 0.25 -0.05 -0.04  0.04  0.09  0.04  0.25
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
## ACF    0.0 -0.03  0.02  0.18 -0.06  0.05 -0.02 -0.03 -0.08  0.05  0.08
## PACF  -0.1 -0.01  0.06  0.13 -0.10  0.09 -0.18 -0.01 -0.04  0.01 -0.06
```



(a) Univariate outlier detection for Co-variate



(b) Univariate outlier detection for Response

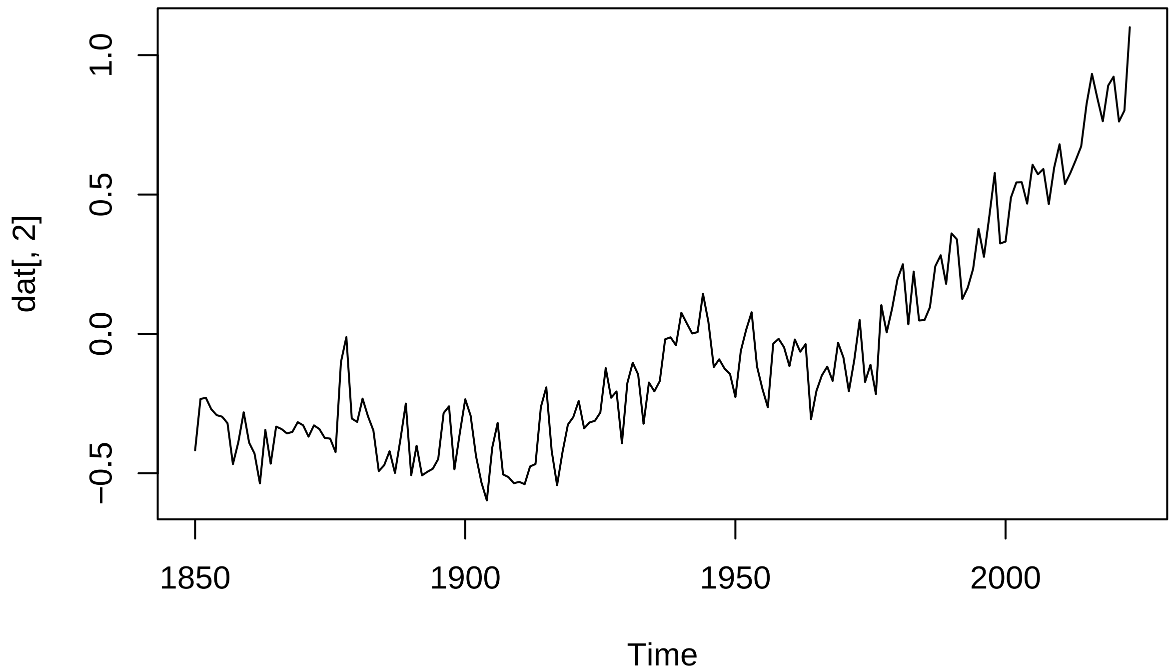


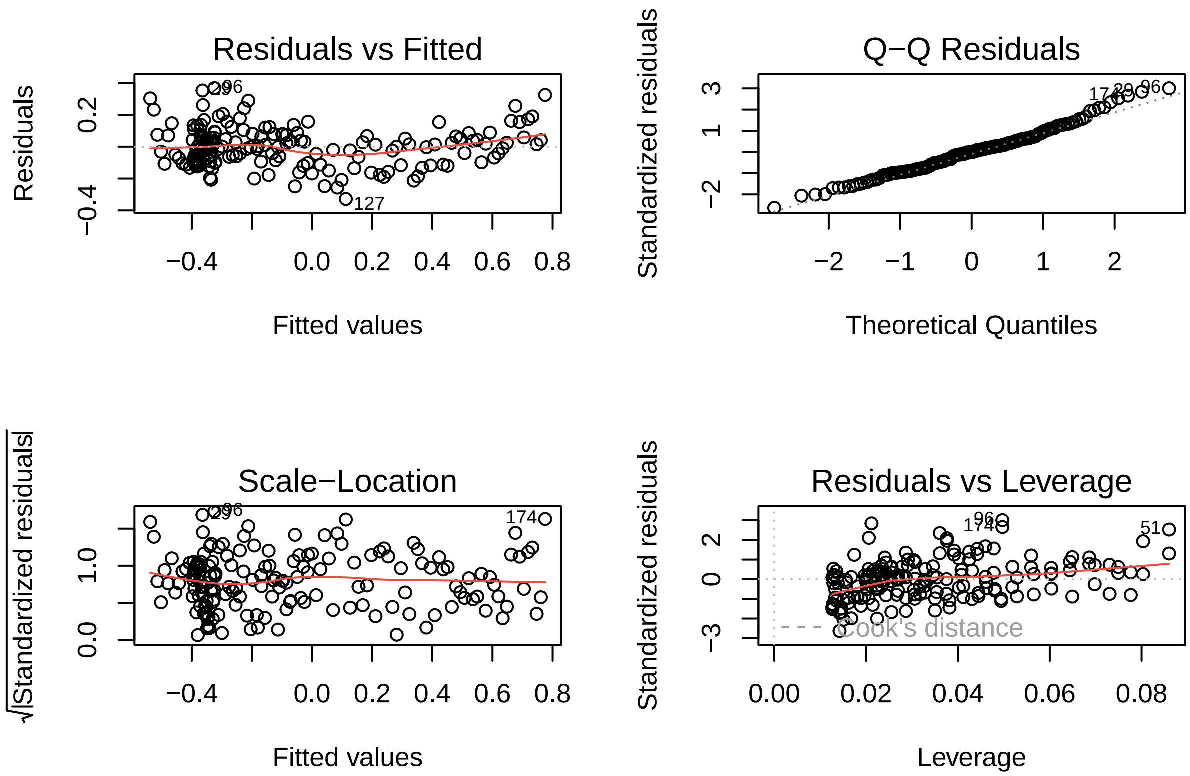
(c) Multivariate Boxplot Outlier Dection

Figure 1: Outlier Dection

Model Deterministic Part.

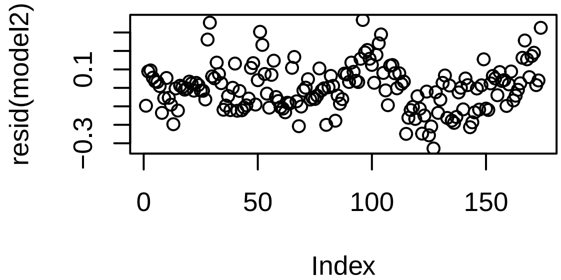
Looking at our plot It seems as if events occurred around 1950, and 1900 start of second industrial revolution and end of World War II.

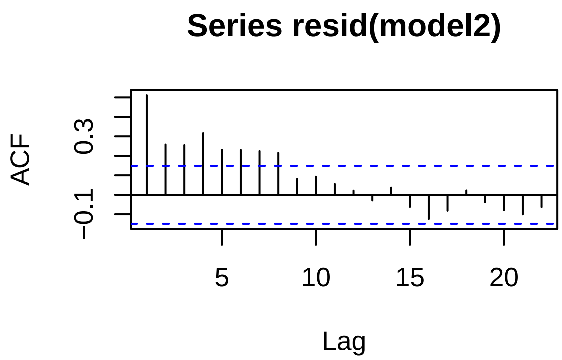


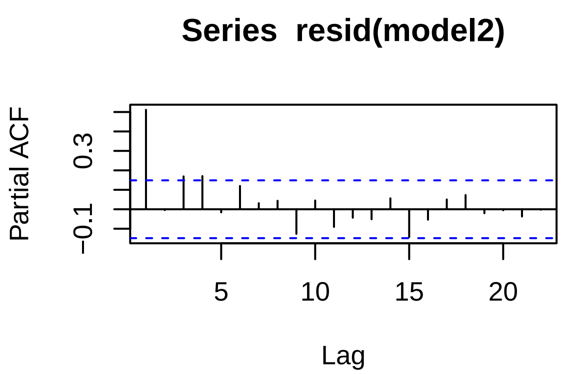



```
##
## Call:
## lm(formula = dat[, 2] ~ time(dat) + time(dat) * Afterww2 + time(dat) *
##     IndustryRev, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

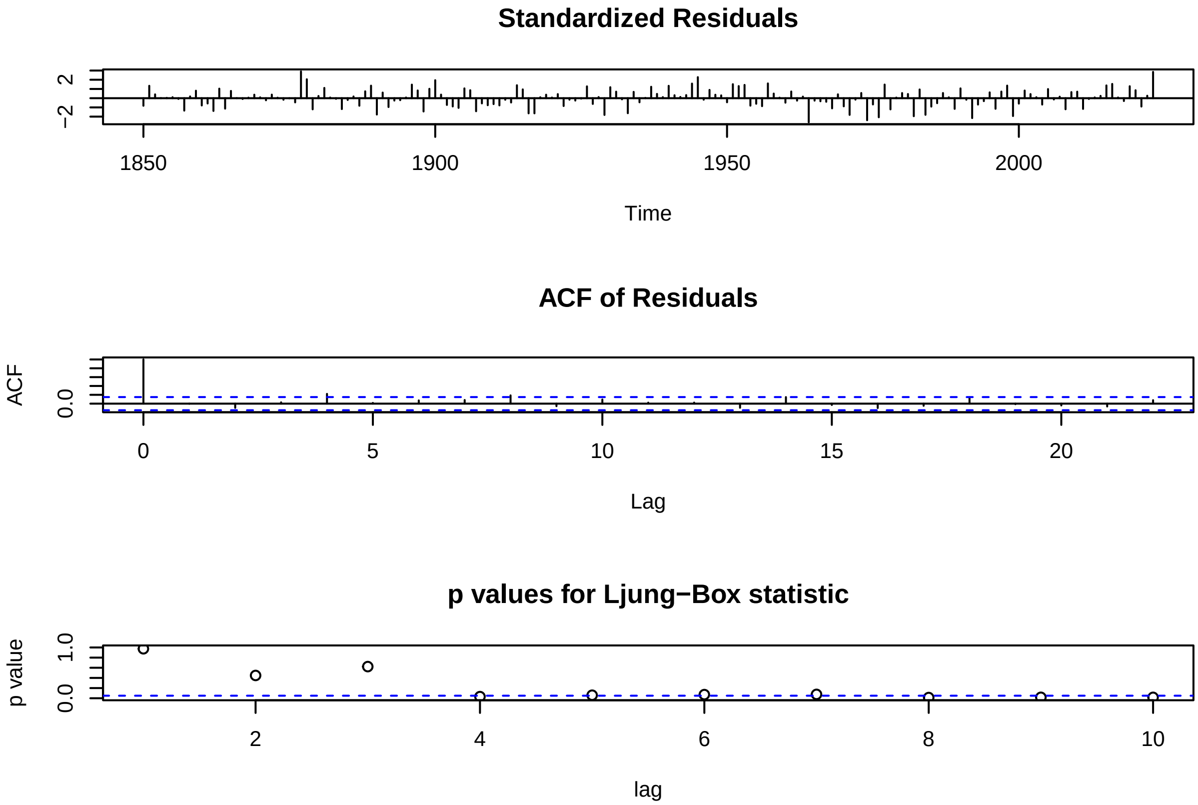
```
## -0.32823 -0.09304  0.00055  0.06970  0.36771
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             2.551782   2.306113   1.107    0.270
## time(dat)              -0.001553   0.001230  -1.262    0.209
## Afterww2               -4.543077   3.029908  -1.499    0.136
## IndustryRev           -25.753919   3.603931  -7.146 2.60e-11 ***
## time(dat):Afterww2      0.002170   0.001568   1.383    0.168
## time(dat):IndustryRev   0.013481   0.001895   7.115 3.08e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1255 on 168 degrees of freedom
## Multiple R-squared:  0.8936, Adjusted R-squared:  0.8904
## F-statistic: 282.1 on 5 and 168 DF,  p-value: < 2.2e-16
```

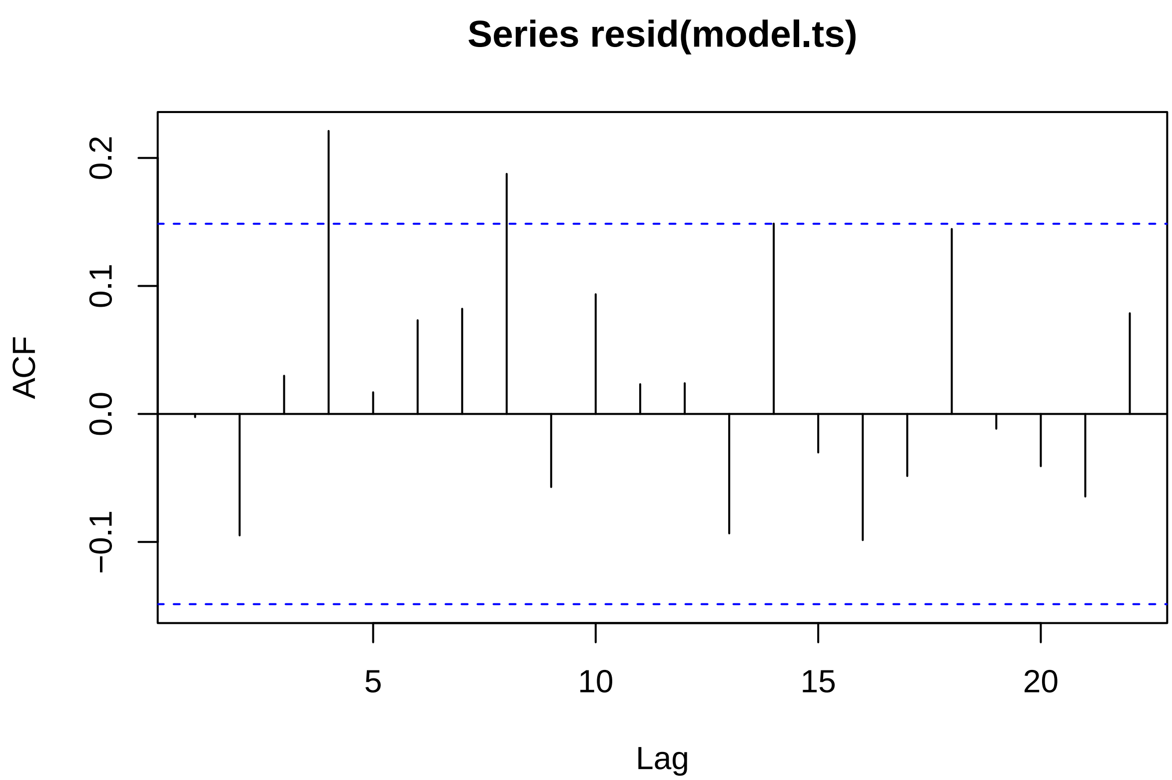


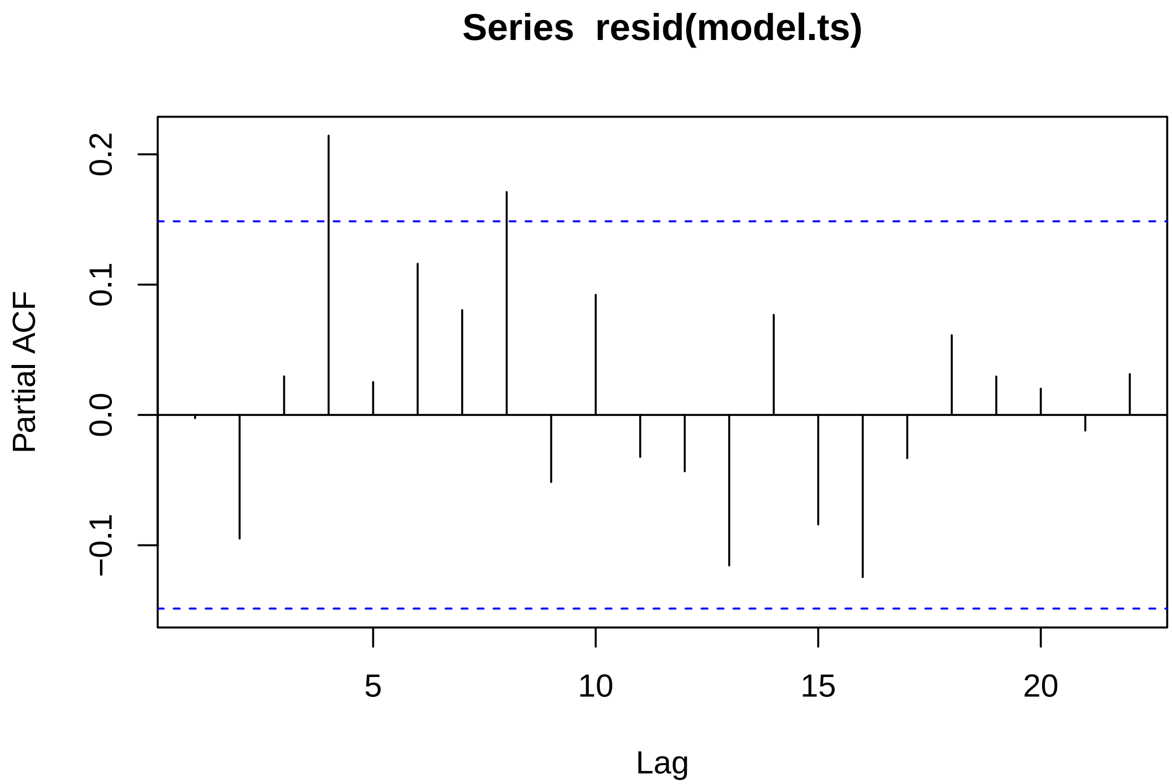




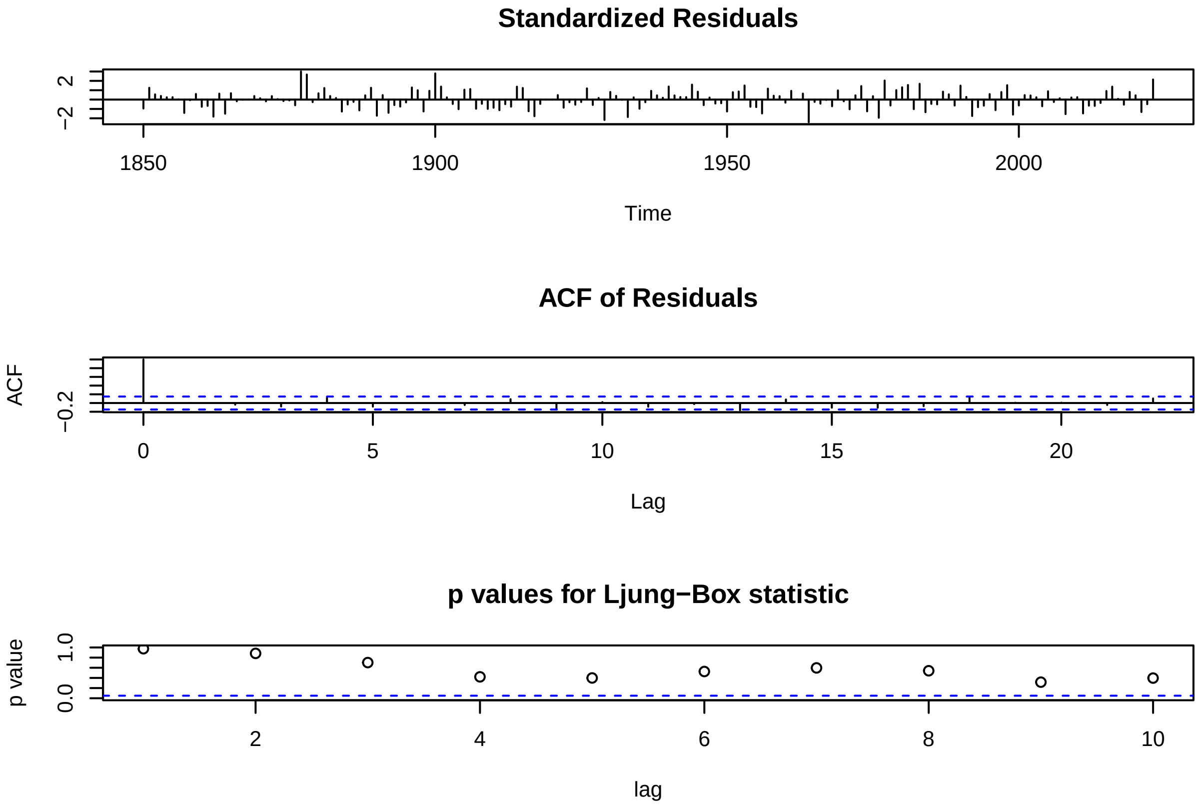
ACF tails off or cuts off after lag 2. PACF cuts off after lag 1. I Will fit an arima(1,0,0) and arima(2,0,2) to see which fits better.

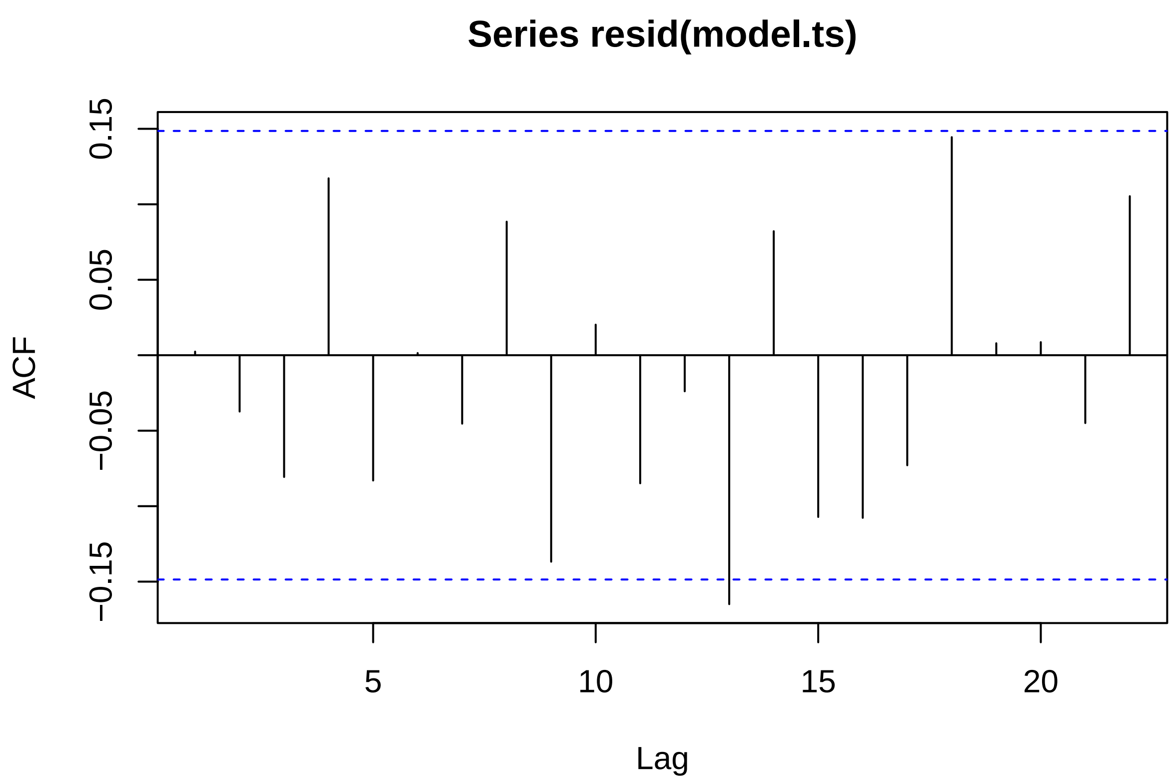

Standardized Residuals
2
−2
1850
1900
1950
2000
Time
ACF of Residuals
ACF
0.0
0
5
10
15
20
Lag
p values for Ljung−Box statistic
p value
0.0
1.0
2
4
6
8
10
lag

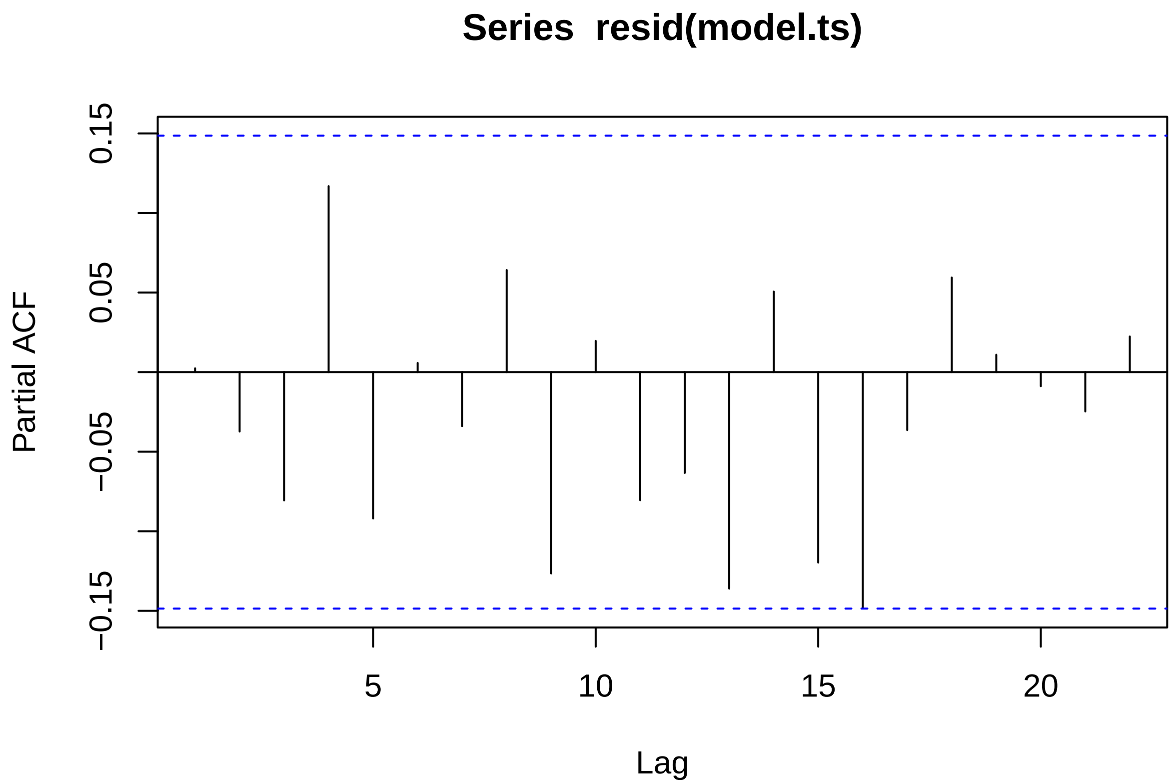

Series resid(model.ts)
ACF
−0.1
0.0
0.1
0.2
5
10
15
20
Lag



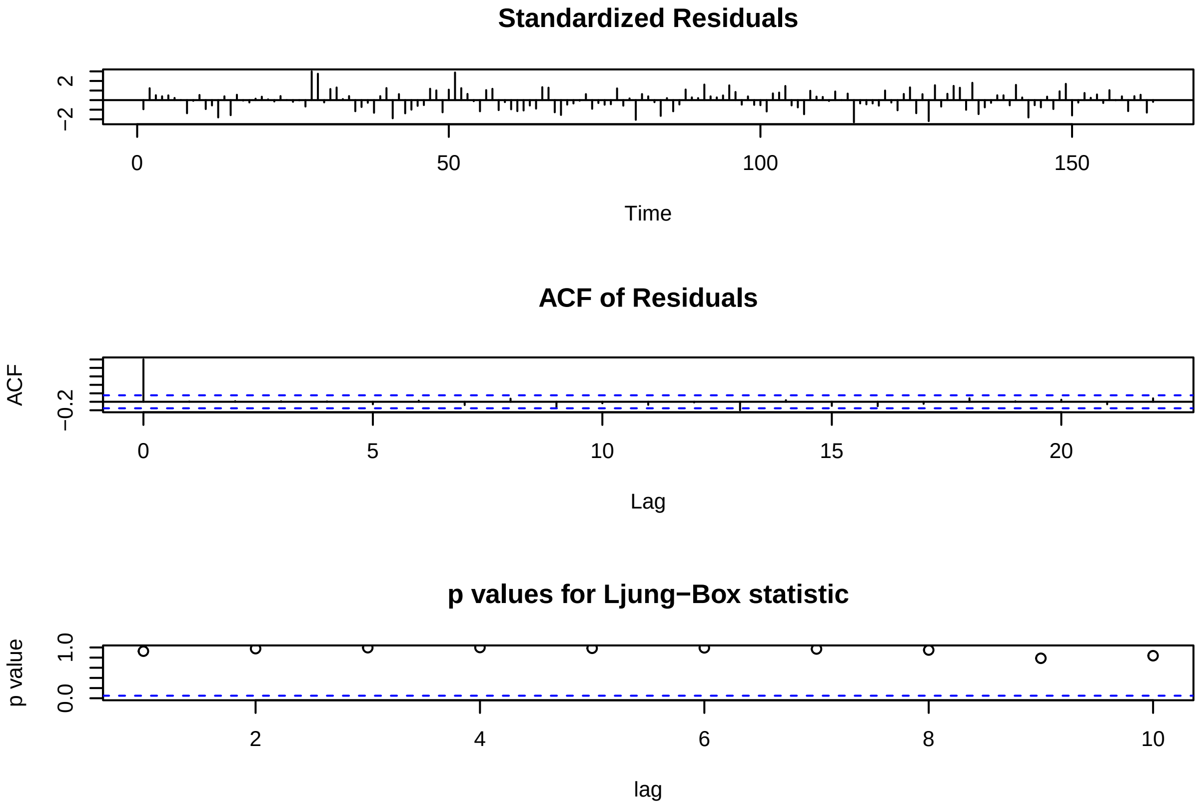

```
##
## z test of coefficients:
##
##                        Estimate  Std. Error z value  Pr(>|z|)
## ar1                   0.5561333   0.0801484  6.9388 3.955e-12 ***
## intercept             4.0105058   3.9954799  1.0038 0.3154940
## time(dat)            -0.0023375   0.0021384 -1.0931 0.2743316
## Afterww2             -9.4970649   5.4416527 -1.7453 0.0809407 .
## IndustryRev         -22.2203007   6.4850219 -3.4264 0.0006116 ***
## time(dat):Afterww2    0.0047519   0.0028103  1.6909 0.0908609 .
## time(dat):IndustryRev   0.0116694   0.0034080  3.4241 0.0006168 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

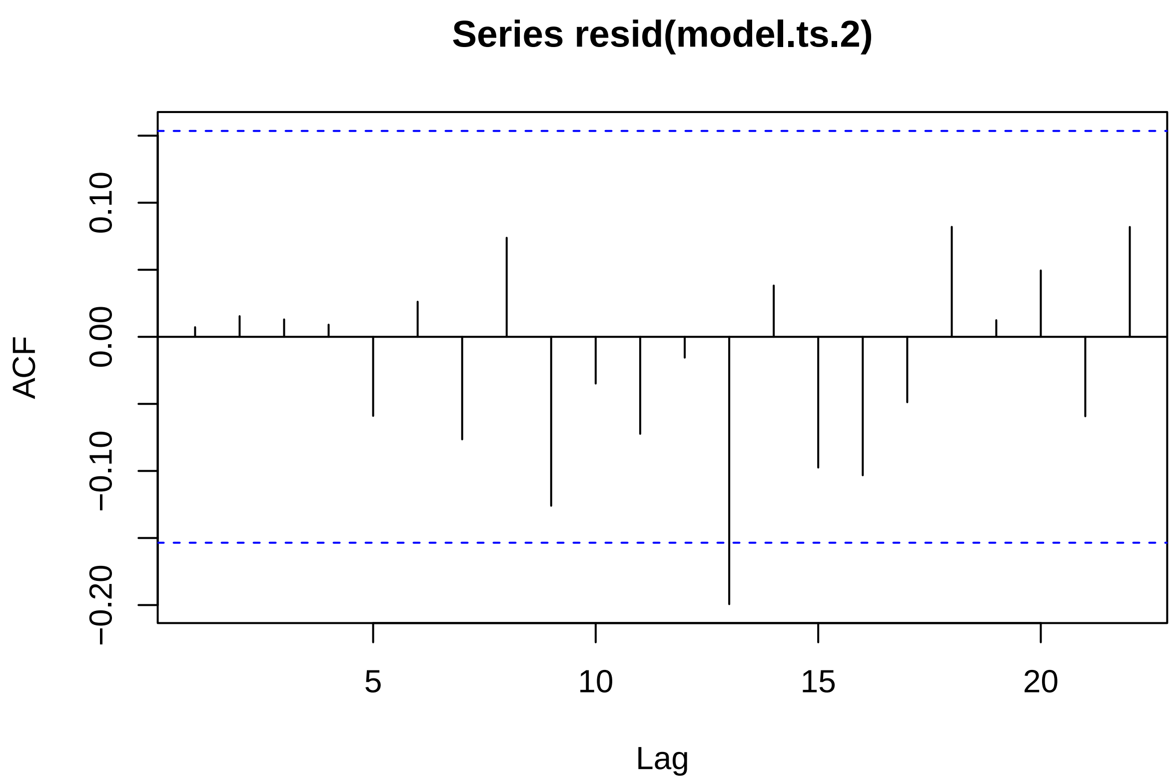

Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung−Box statistic
p value
lag

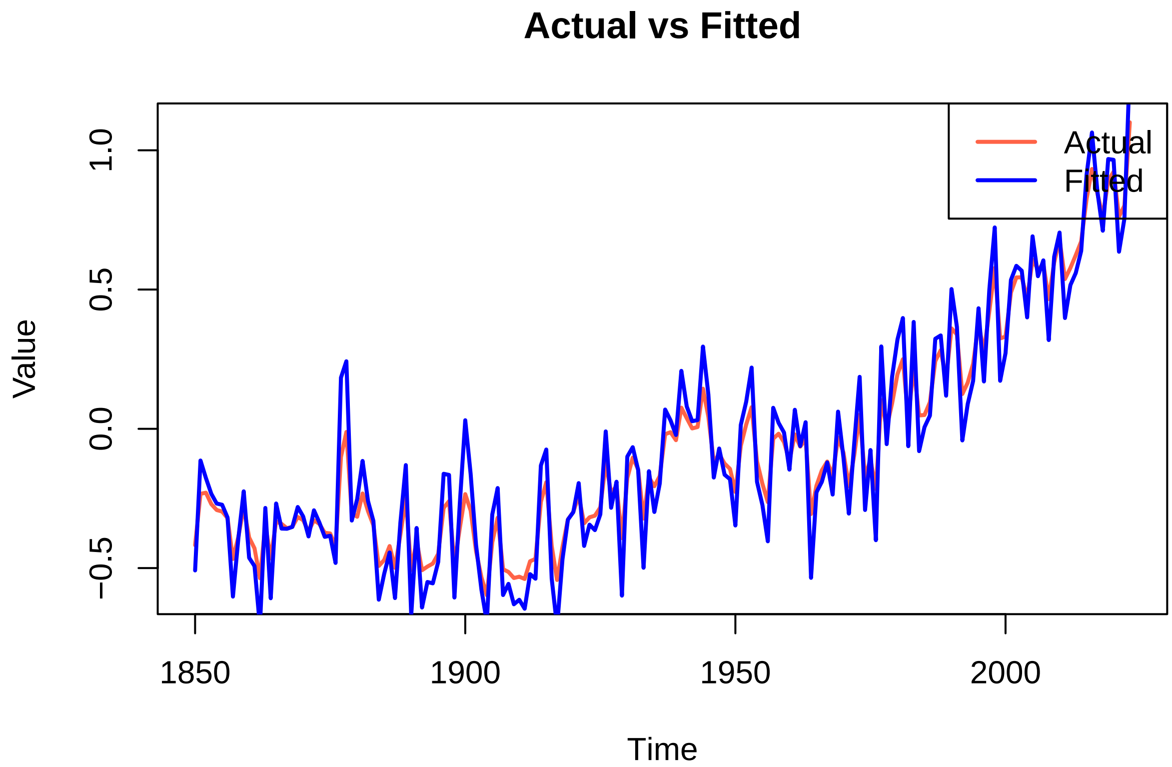

Series resid(model.ts)
ACF
Lag



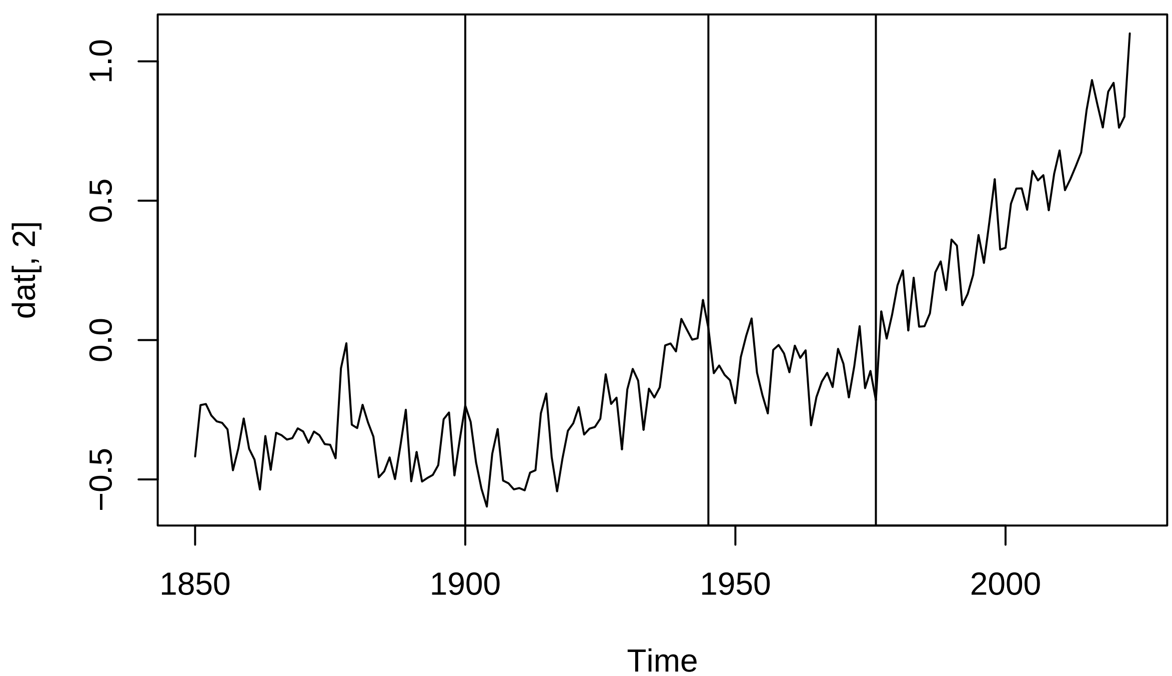

```
##
## z test of coefficients:
##
##                          Estimate  Std. Error z value  Pr(>|z|)
## ma1                     0.3577942   0.0699044  5.1183 3.082e-07 ***
## intercept               2.6897378   2.3217003  1.1585    0.2467
## time(dat)              -0.0016276   0.0012435 -1.3090    0.1905
## Afterww2               23.4711228   5.7243811  4.1002 4.128e-05 ***
## IndustryRev           -25.4781150   3.6554503 -6.9699 3.172e-12 ***
## I1976                 -40.9890374   5.6387633 -7.2692 3.617e-13 ***
## time(dat):Afterww2     -0.0121119   0.0029390 -4.1211 3.770e-05 ***
## time(dat):IndustryRev   0.0133418   0.0019239  6.9347 4.072e-12 ***
## time(dat):I1976         0.0207733   0.0028617  7.2591 3.896e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Standardized Residuals
2
−2
0
50
100
150
Time
ACF of Residuals
ACF
−0.2
0
5
10
15
20
Lag
p values for Ljung−Box statistic
p value
0.0
1.0
2
4
6
8
10
lag


Series resid(model.ts.2)
ACF
0.10
0.00
−0.10
−0.20
5
10
15
20
Lag

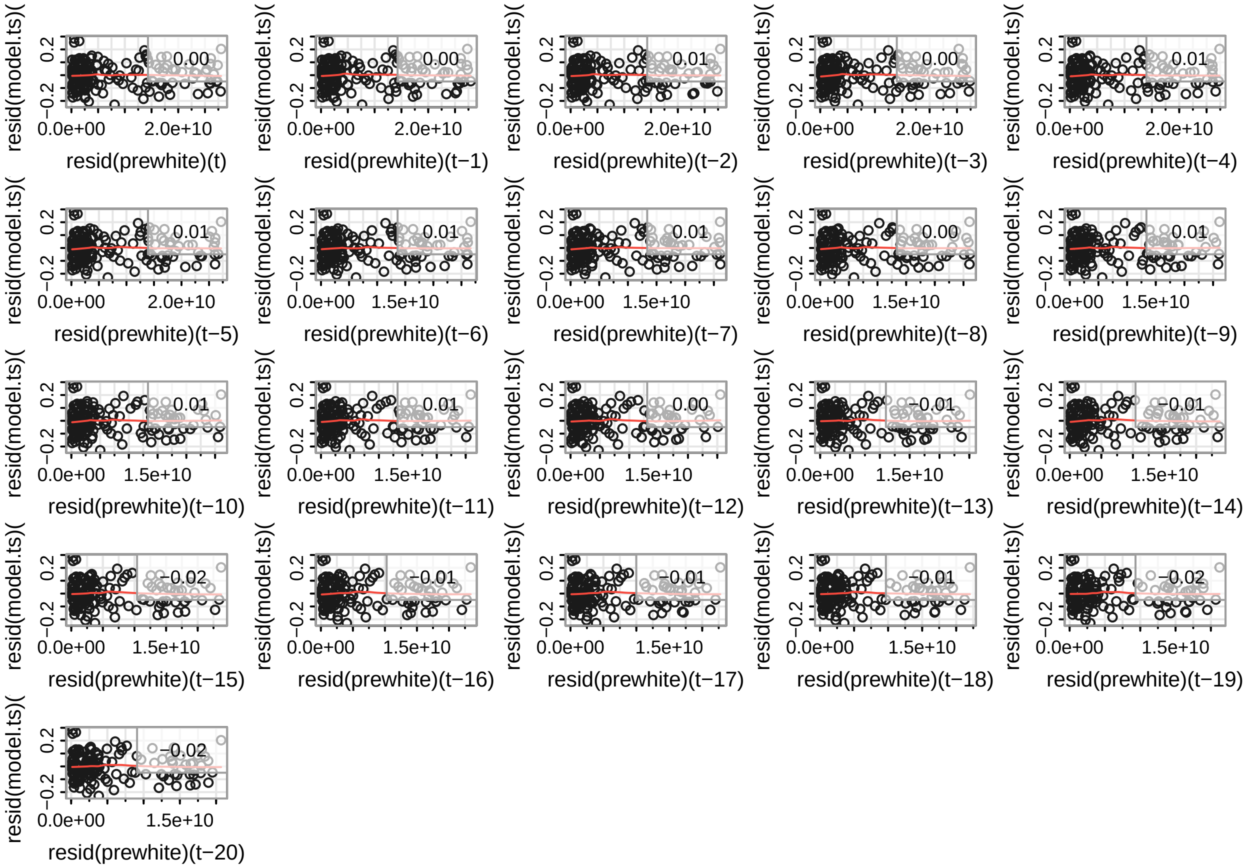


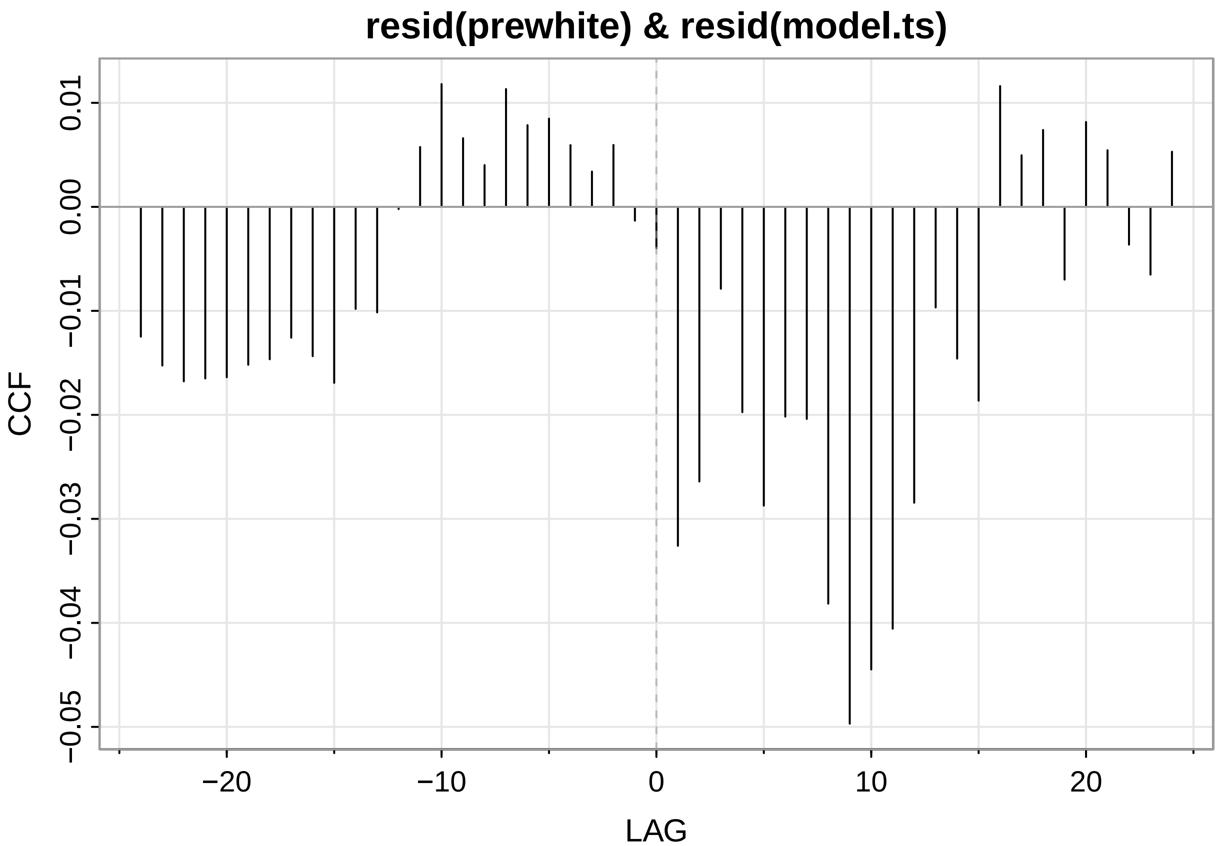
Actual vs Fitted
Actual
Fitted
Value
Time
1850
1900
1950
2000
−0.5
0.0
0.5
1.0



dat[, 2]
Time
1850
1900
1950
2000
−0.5
0.0
0.5
1.0

usuing c02 to predict





Everything from c02 is captured in our model for tempature.

Code Appendix

```
# Setting Document Options ----
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.align = "center"
)


library(TSA)
library(forecast)
library(astsa)
library(dplyr)
library(rugarch)
library(tseries)
library(lmtest)
library(jsonlite)





# Fetch the data
co2Data <- read.csv("https://ourworldindata.org/grapher/annual-co2-emissions-per-country.csv?v=1&csvType


YearlyTemp <- read.csv("Yearly.csv")
YearlyTemp <- YearlyTemp[YearlyTemp$Source == "gcag", ] # keep source constant.
c02_TS <- co2Data[, c("Year", "Annual.CO..emissions")]
c02_TS <- c02_TS[order(c02_TS$Year), ]
c02_TS <- ts(c02_TS$Annual.CO..emissions,
                   start = min(c02_TS$Year),
                   end = max(c02_TS$Year),
                   frequency = 1)

temp_TS <- YearlyTemp[, c("Year", "Mean")]
temp_TS <- temp_TS[order(temp_TS$Year), ]
temp_TS <- ts(temp_TS$Mean,
                   start = min(temp_TS$Year),
                   end = max(temp_TS$Year),
                   frequency = 1)

dat <- ts.intersect(c02_TS, temp_TS)

plot(dat[,1])
plot(dat[,2])

acf2(diff(log(dat[,1])))
plot(dat[,2])
Afterww2 <- as.numeric(time(dat)>= 1945)
IndustryRev <- as.numeric(time(dat)>= 1900)
```

```r
I1976 <- as.numeric(time(dat)>= 1976)

model2 <- lm(dat[,2] ~ time(dat)+time(dat)*Afterww2+ time(dat)*IndustryRev, data = dat)
#plot(model2)
par(mfrow=c(2,2))
plot(model2, whixh = 1:4)
summary(model2)
plot(resid(model2))

acf(resid(model2))
pacf(resid(model2))
model.matrix = model.matrix(object= ~ time(dat)+time(dat)*Afterww2+ time(dat)*IndustryRev-1)

model.ts <- arima(x = dat[,2], order = c(1,0,0),  xreg = model.matrix, method = "ML")

tsdiag(model.ts)

acf(resid(model.ts))
pacf(resid(model.ts))
coeftest(model.ts)
#model.matrix = model.matrix(object= ~ time(tmp)+time(tmp)*Afterww2+ time(tmp)*IndustryRev-1)




model.matrix = model.matrix(object= ~ time(dat)+time(dat)*Afterww2+ time(dat)*IndustryRev + I1976*time(

model.ts <- arima(x = dat[,2], order = c(0,0,1),  xreg = model.matrix, method = "ML")

tsdiag(model.ts)

acf(resid(model.ts))
pacf(resid(model.ts))

coeftest(model.ts)
#model.matrix = model.matrix(object= ~ time(tmp)+time(tmp)*Afterww2+ time(tmp)*IndustryRev-1)

tmp <- dat[,2][1:163]

Afterww2.2<-Afterww2[1:163]
IndustryRev.2<- IndustryRev[1:163]
I1976.2 <- I1976[1:163]
model.matrix.2 = model.matrix(object= ~ time(tmp)+time(tmp)*Afterww2.2+ time(tmp)*IndustryRev.2+ time(t

model.ts.2 <- arima(x = tmp, order = c(2,0,2),  xreg = model.matrix.2, method = "ML")

tsdiag(model.ts.2)

acf(resid(model.ts.2))
#pacf(resid(model.ts.2))

#coeftest(model.ts.2)
fitted_values <- resid(model.ts) + dat[,2]
```

```r
plot(dat[, 2], type = "l", col = "tomato", lwd = 2, main = "Actual vs Fitted", ylab = "Value")
lines(fitted_values, col = "blue", lwd = 2)
legend("topright", legend = c("Actual", "Fitted"), col = c("tomato", "blue"), lty = 1, lwd = 2)
plot(dat[,2])
abline(v=1900)
abline(v = 1945)
abline(v = 1976)

prewhite <- Arima(dat[,1], model = model.ts, xreg = model.matrix)


lag2.plot(resid(prewhite), resid(model.ts),max.lag = 20)

ccf2(resid(prewhite), resid(model.ts))
```

Ritchie, Hannah, Pablo Rosado, and Max Roser. 2023. "CO and Greenhouse Gas Emissions." *Our World in Data*. https://ourworldindata.org/co2-and-greenhouse-gas-emissions.