Global CO_2 Emissions 1995 and Pressent

true

Abstract

Global average temperatures have increased by more than 1°C from pre-industrial times to now, and it's common scientific belief that the increase of carbon emissions worldwide has played a critical role in this global warming. Scientists believe that one of the primary drivers of climate change is that continued, and increasing, human emissions of carbon dioxide and other greenhouse gases into the atmosphere. This rapid global warming can have and already has had significant impacts on global and local climates as well as natural and industrial systems across the world. Studying CO2 presense in the atmosphere is important in this context then in order to better understand not only one of the drivers of climate change, and in order to ascertain what steps we might be able to take to predict and manage the effects that will arise as a result of it.

Understanding a changing climate, and what it means for the earth's inhabitants is of growing interest to the scientific and policy community. Although, at this point in 1997 it is not entirely clear what the consequences of this growing awareness will be, in this report we present likely outcomes under "business-as-usual" scenarios. In doing so, our hope, is to establish a series of possible futures, and, as evidence, technology, and policy develop over the coming decades, our goal is that we can theb weigh the impacts that carbon-emission reduction efforts might take.

Background

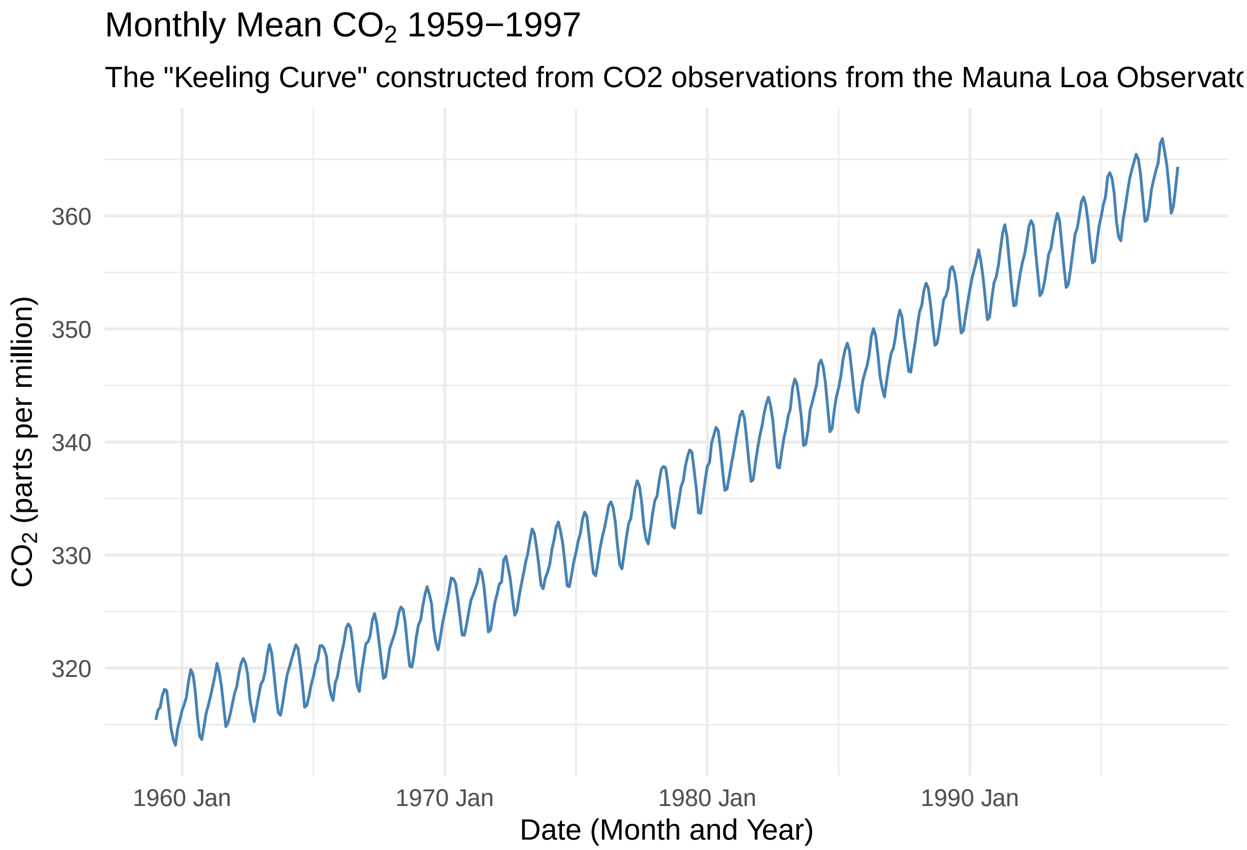
Carbon Emissions

In this report we seek to understand if atmospheric CO2 levels have increased since 1997 and if we anticipate that this trend will continue. To do this we seek to answer 2 main questions: > What is the best model we can use to model CO2 levels over time? > Using this model, what do we forecast atmospheric CO2 levels will be? And how confident are we in these predictions?

High CO2 levels have been shown to negatively play into climate change, and the increase in atmospheric CO2 has cascading effects from global warming to ocean acidification. Due to these effects, we may notice food supply chain disruptions, natural disasters of increasing intensity, and habitat disruption of all ecosystems and the animal species living there. In an economical sense, understanding these patterns is important so we can prepare for any shifts and uncertainties in global supply chains that might arise for these new climate patterns. In total understanding the patterns in atmospheric CO2 concentrations is important both so that we can mitigate the existing effects of this increase and prevent continuing upwards trends in CO2. In addition if our report findings are significant, the findings could persuade other governmental organizations to increase their CO2 monitoring efforts and global preparedness efforts.

Historical Trends in Atmospheric Carbon

In 1958 Charles Keeling began continuous monitoring of atmospheric carbon dioxide concentrations from the Mauna Loa Observatory in Hawaii and soon observed a trend increase carbon dioxide levels in addition to the seasonal cycle. He was able to attribute this trend increase to growth in global rates of fossil fuel combustion. This trend has continued to the present, and is known as the "Keeling Curve,", this curve is plotted below. And we can see in this curve both an overall rising trend in CO2 levels as well as some sort of cyclic or seasonal pattern of atmospheric CO2 fluctuations each year.

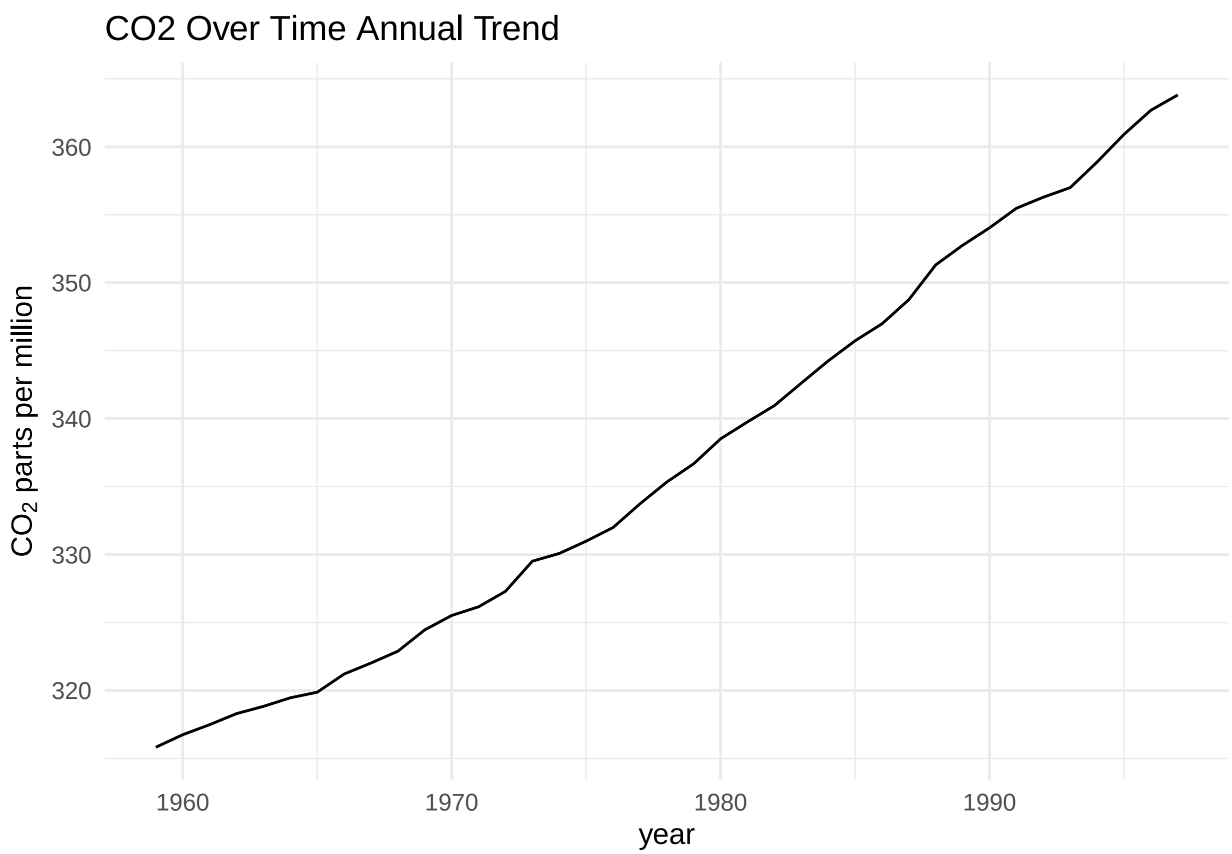


Models and Forecasts

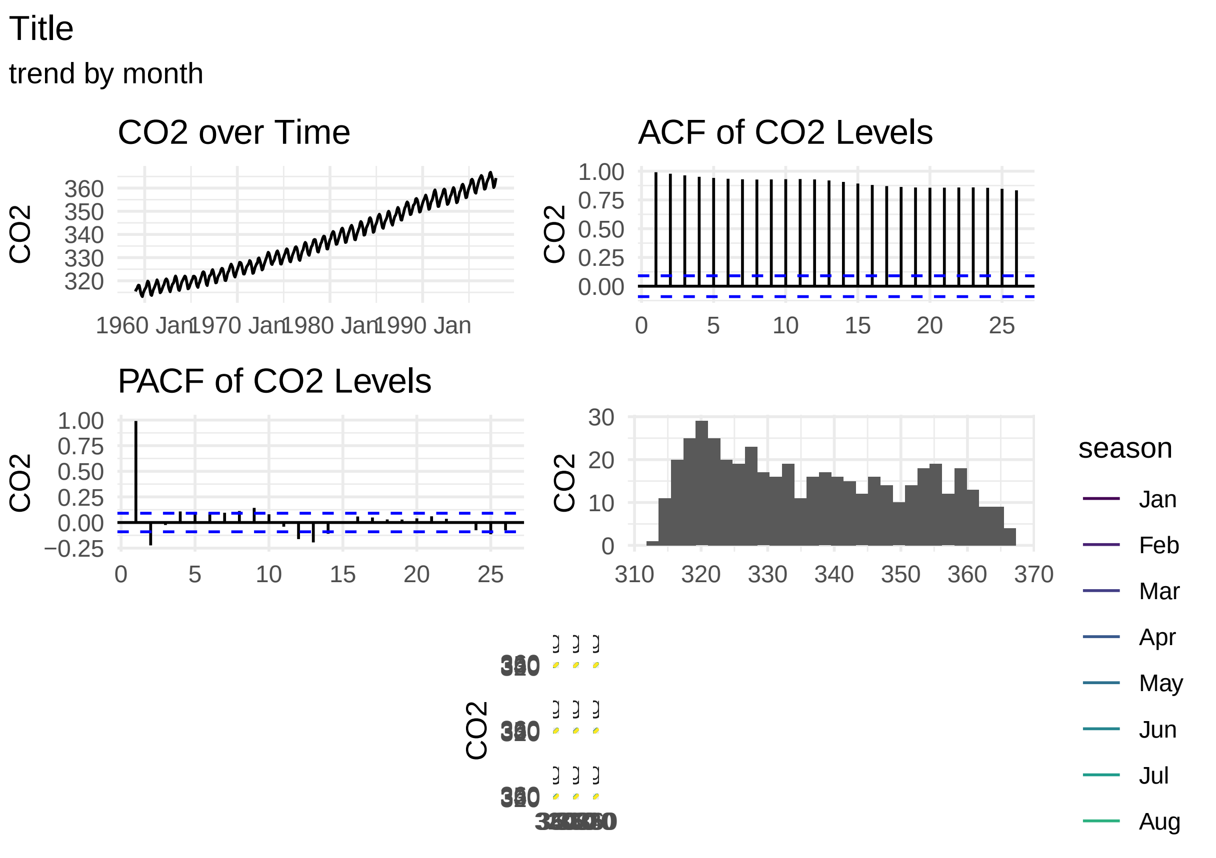
In this section, we evaluate two classes of models for answering our questions – a linear time model and and an ARIMA model to assess which time series model is most appropriate to use to model CO2 levels over time.

Data EDA

As a background on the data we are using it is important to note that this data measures the mean atmospheric CO2 concentration and was collected at the Mauna Loa Observatory in Hawaii. The CO2 levels in the data range from 313 parts per million by volume (ppmv) in March 1958 to 406 ppmv in November 2018. The data was also normalized to remove any influence from local contamination. Carbon dioxide measurements at the Mauna Loa Observatory in Hawaii are made with a type of infrared spectrophotometer, now known as a nondispersive infrared sensor, that is calibrated using World Meteorological Organization standards.




```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```



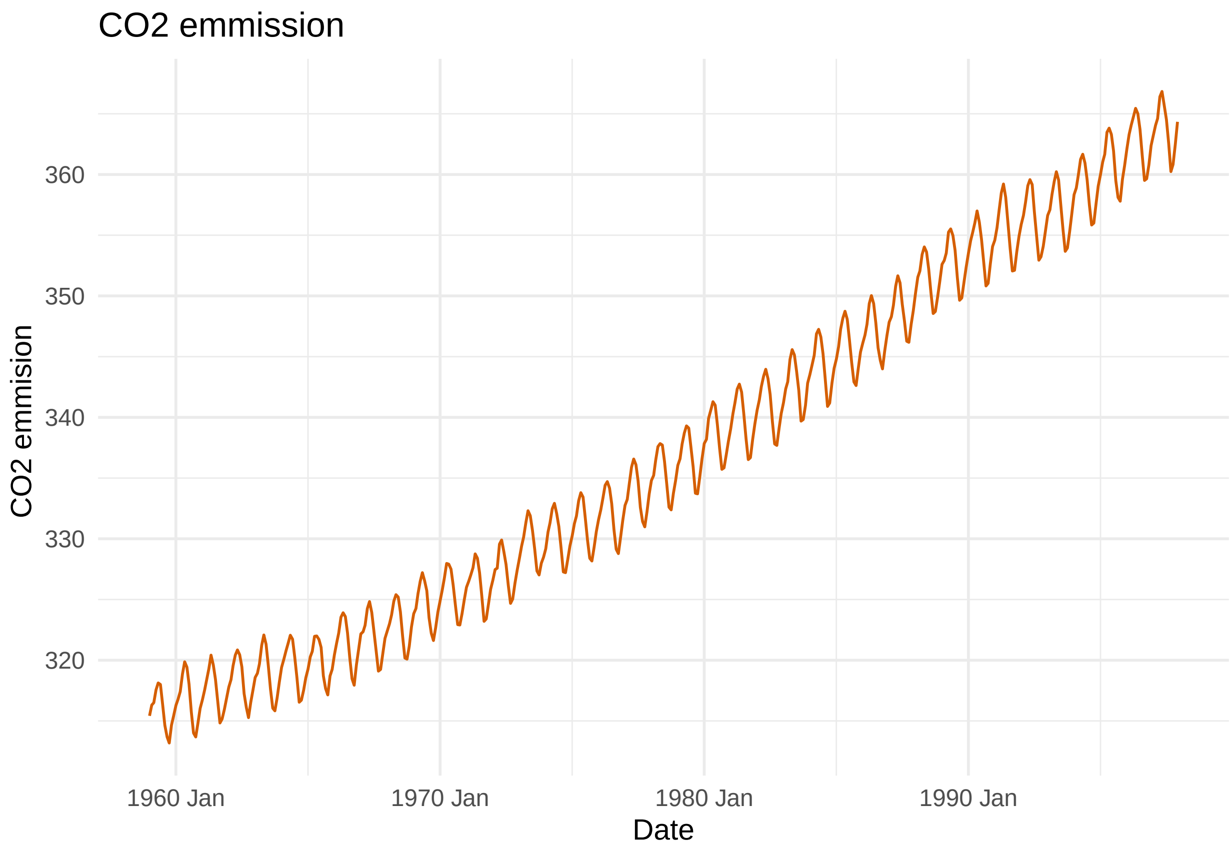
Observations: - Time Series: From the CO2 Over Time plot, there is an obvious trend of increasing levels throughout time, and we can some regular oscillations which could mean there is seasonality in the data. - ACF: There's a gradual decay in the ACF values over the lags, perhaps indicating there is a trend in the data. - PACF: PACF cuts off to zero after lag 2 and stays that way. However, at lags 12 and 13, the values become significant again. This suggests that although the model exhibits AR model-like behavior, there's some deviances in the data. - Distribution:

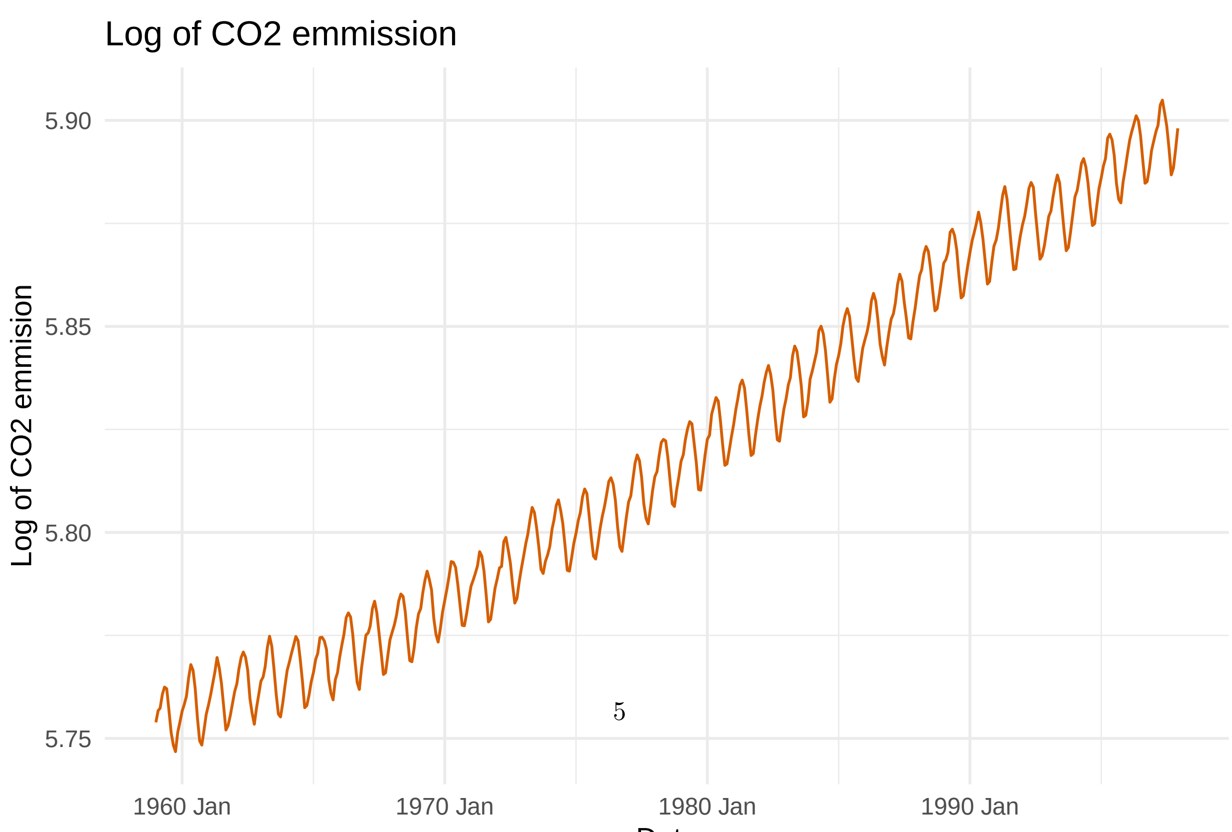
> Observations: - Time Series: From the yearly CO2 Over Time plot, there is an obvious trend of increasing levels throughout time, but we can see that there is no seasonality now. This points to the period of the seasonal trend being a year. - ACF: There's a gradual decay in the ACF values over the lags, indicating there is a still a trend in the data.

Need to fix the EDA plots to better look at the trend, decompose seasonality, and to look at the trend og growth (acceleration)

Linear Models

Time series decomposition



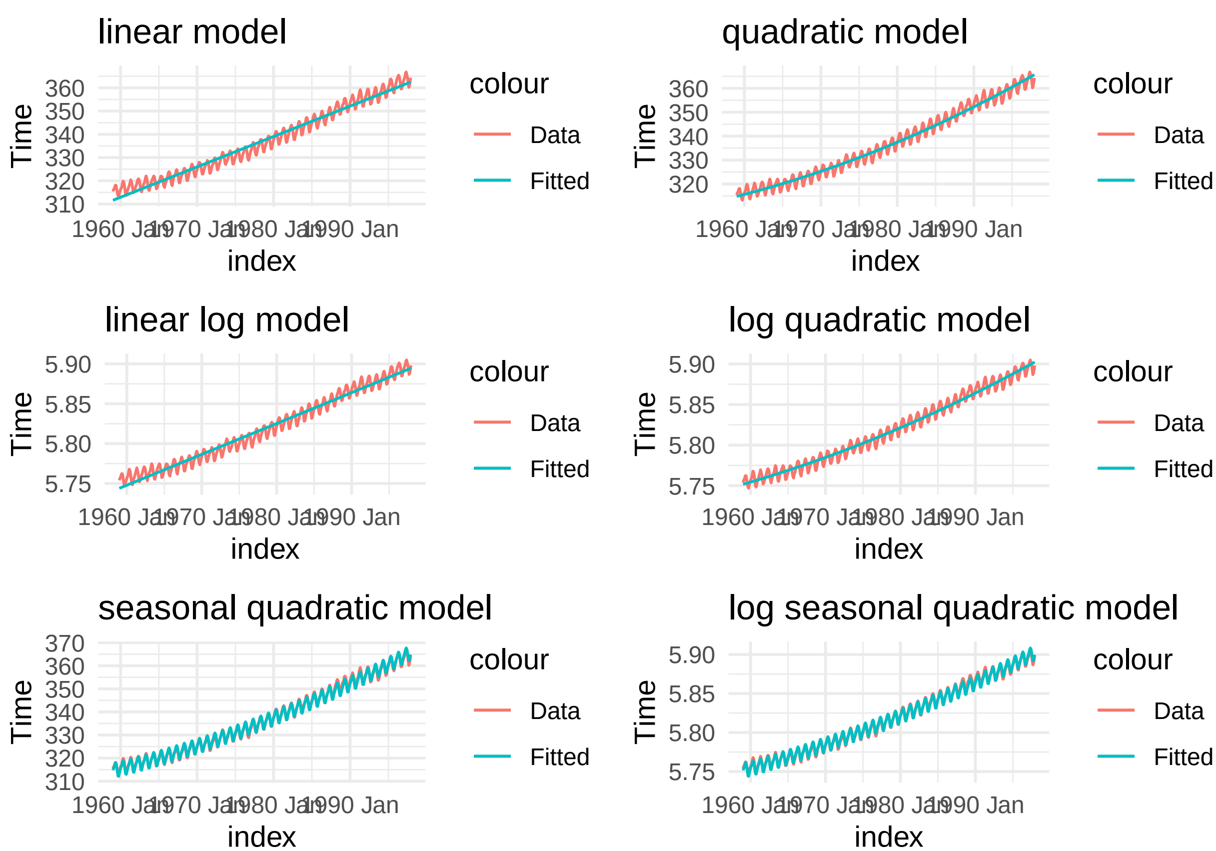


To fit linear time trend model to the `co2` series we will compare a regular time trend linear model to a quadratic time trend model. We will also fit a polynomial time trend model that incorporates seasonal dummy variables, and use this model to generate forecasts to the year 2020. Note that we will be evaluating using both the unscaled CO2 and the log(CO2) value for constructing our models.

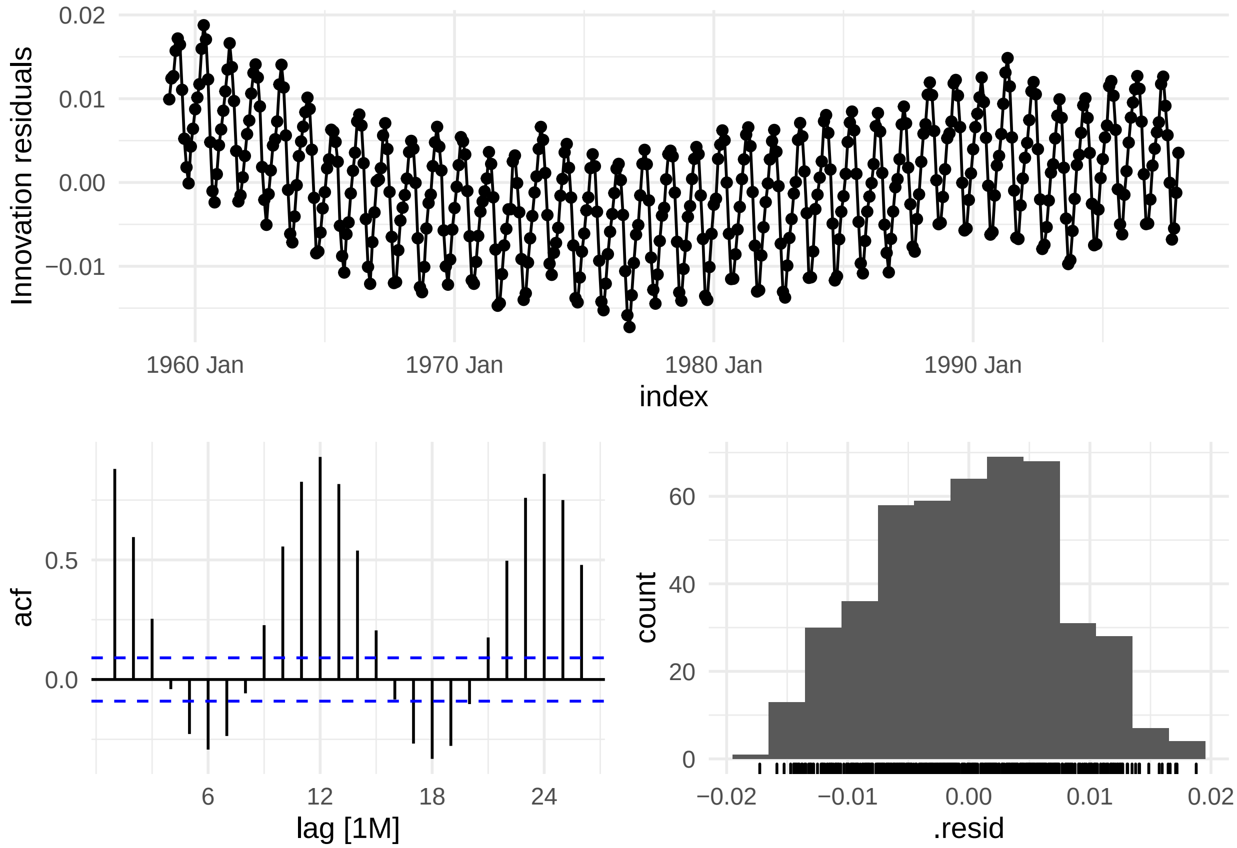
To begin our analysis, we fit a linear model of the form to our data:

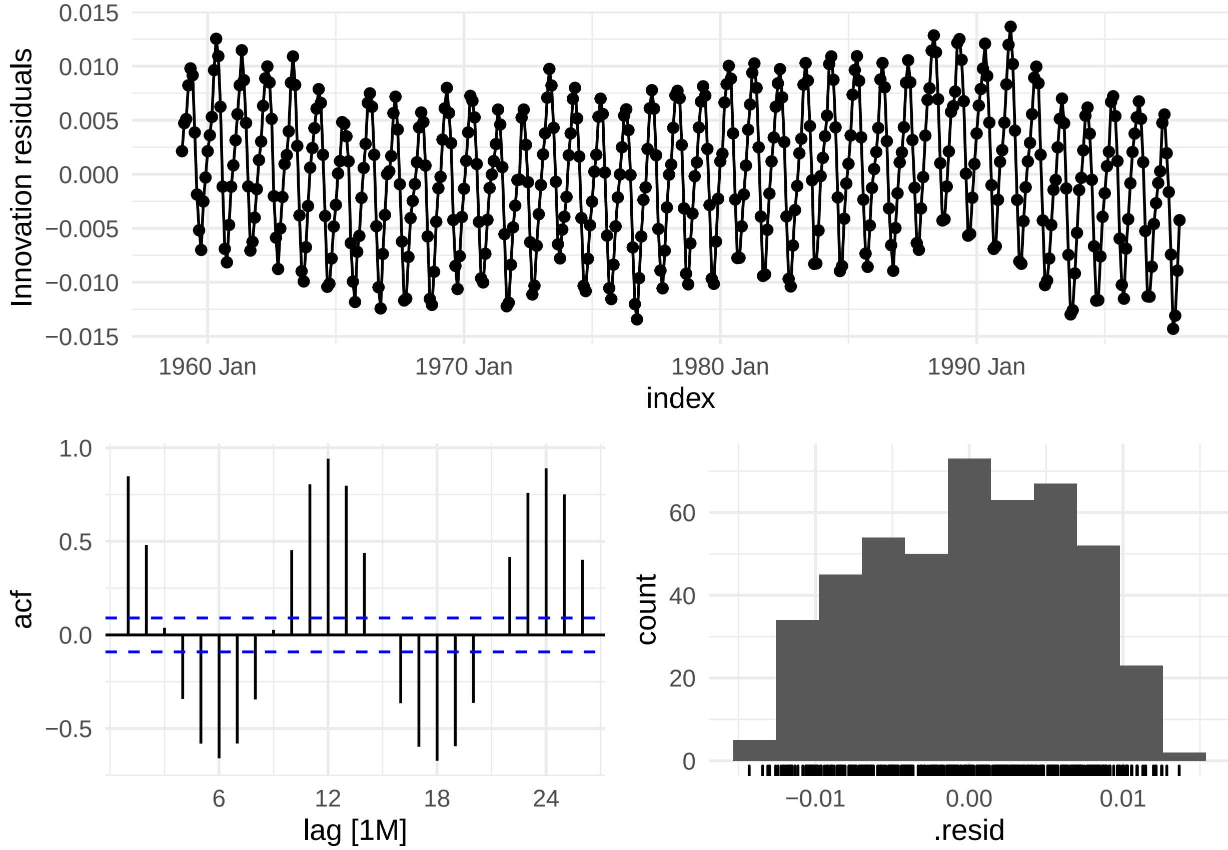
$$\mathrm{CO}_2 = \phi_0 + \phi_1 + \epsilon_{eit} \tag{1}$$

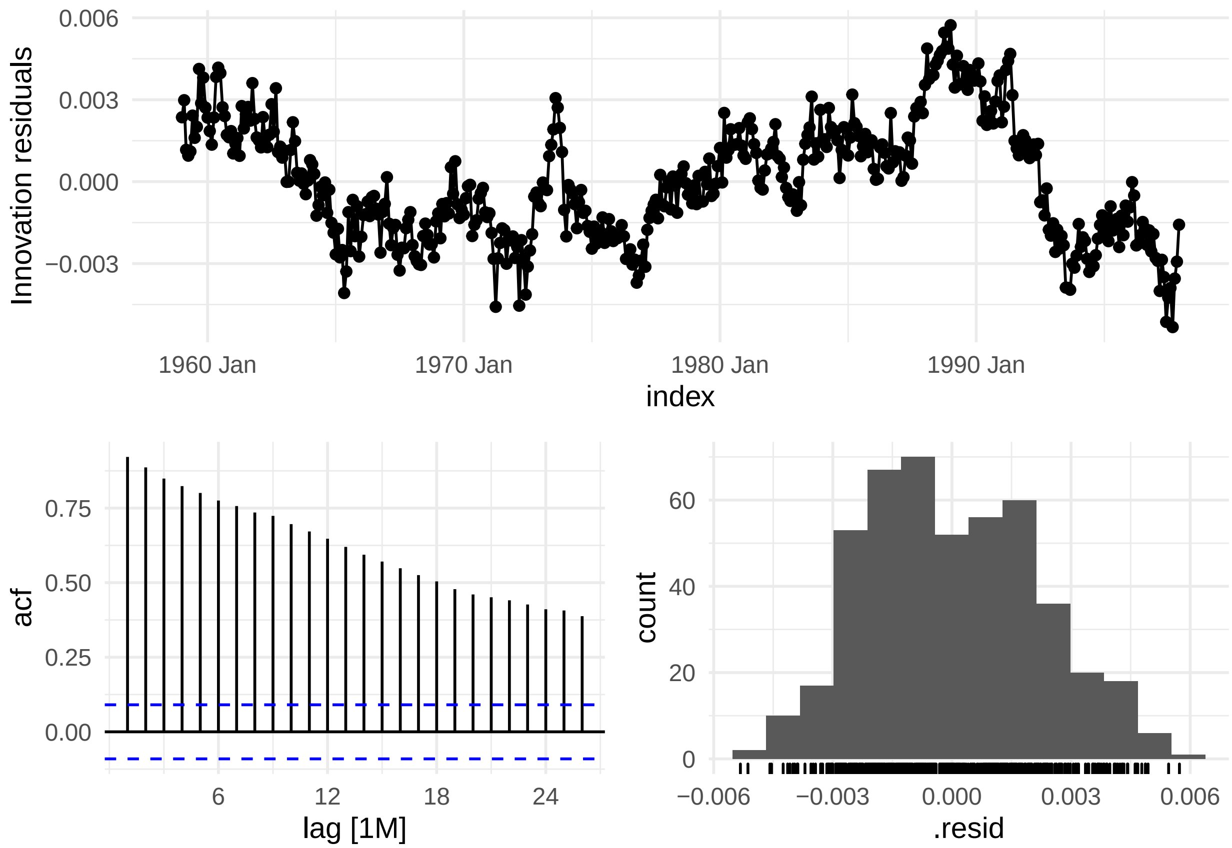
```
##
## My TSLM Models
## =================================
## Statistic N Mean St. Dev. Min Max
## =================================
```



We can see that the log models appear to fit the data slightly better (some overshoot in the linear model around 1980s, whereas less in log linear model). Same with the quadratic models, appears like it's slightly overestimating in the quadratic model and is overestimating less (ie. in the middle of the seasonal variation) in the log quad model. zcan see this in the seasonal model too. Moving forward with working on the log(co2) data.

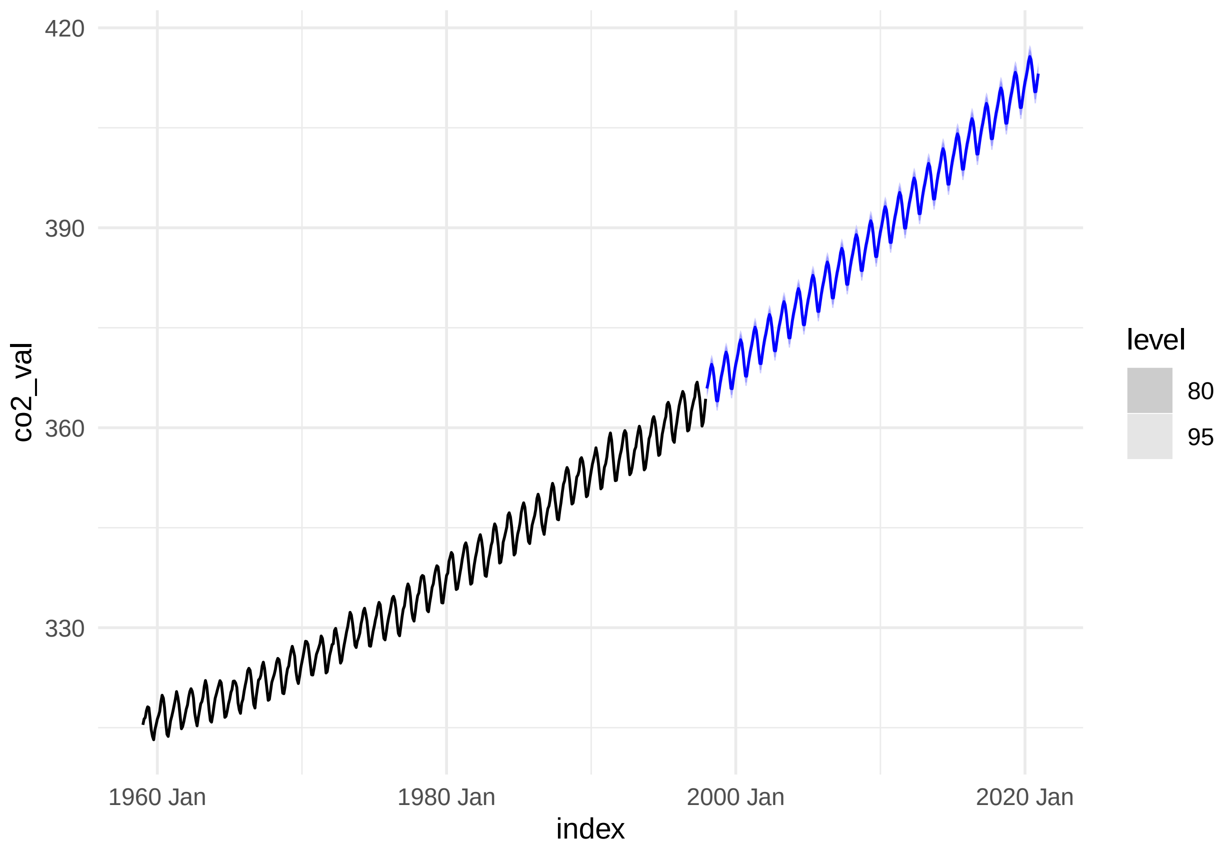

Innovation residuals
0.02
0.01
0.00
−0.01
1960 Jan
1970 Jan
1980 Jan
1990 Jan
index
acf
0.5
0.0
6
12
18
24
lag [1M]
count
60
40
20
0
−0.02
−0.01
0.00
0.01
0.02
.resid


Innovation residuals
0.015
0.010
0.005
0.000
−0.005
−0.010
−0.015
1960 Jan
1970 Jan
1980 Jan
1990 Jan
index
acf
1.0
0.5
0.0
−0.5
6
12
18
24
lag [1M]
count
60
40
20
0
−0.01
0.00
0.01
.resid



Innovation residuals
0.006
0.003
0.000
−0.003
1960 Jan
1970 Jan
1980 Jan
1990 Jan
index
acf
0.75
0.50
0.25
0.00
6
12
18
24
lag [1M]
count
60
40
20
0
−0.006
−0.003
0.000
0.003
0.006
.resid

Model Forecast



ARIMA Models

We will also generate a few ARIMA models and compare them to our linear model. To select our ARIMA model we use the AIC, the AIC selects against models with too many parameters. We know that there is a seasonal component to our data, so we will also select for seasonal AR and MA values and check if seasonal differencing will make the data stationary. I am choosing to use the ARIMA() function to select our data. This method will determine the optimal p,d,q,P,D,Q values as well as the optimal levels of differencing for us to have stationary data and will then select a model based on the metric we have chosen as our criteria (AIC). We can see interpret the resulting ARIMA model in terms of the number of (seasonal) AR terms, the level of differencing to acheive stationary data, and the (seasonalWe are including seasonality in our model because we can see in the decomposition as well as the time series plots that there is a strong possibility of a seasonal component in the CO2 levels. If adding a seasonal component improves the AIC score then such a model will be chosen.

```
## Series: co2_val
## Model: ARIMA(1,0,1)(4,1,0)[12] w/ drift
##
## Coefficients:
##          ar1      ma1     sar1     sar2    sar3     sar4  constant
##       0.9819  -0.3751  -0.7144  -0.5526  -0.334  -0.1504    0.0630
## s.e.  0.0098   0.0495   0.0488   0.0581   0.058   0.0497    0.0081
##
## sigma^2 estimated as 0.09338:  log likelihood=-101
```

```
## AIC=218   AICc=219   BIC=251
```

With the AIC as our selection criteria we have estimated the model to be an ARIMA(1,0,1)(4,1,0)[12] with drift. This model has an AIC of 218, since this was the model that was chosen, this must be the smallest AIC value in the models that we have compared. This ARIMA(1,0,1)(4,1,0)[12] model can be interpreted by saying that there is 1 AR term, 1 MA term, 4 seasonal AR terms and the data had to be seasonally differenced once to be made stationary. the [12] at the end of the model also indicates

To be thorough we will also repeat the above process but instead using BIC and AICC as our selection metrics. In general the BIC is stricter than the AIC in penalizing additional parameters in our model, so it is possible that this selection process will result in a different model. When we repeated the above process using the bic or the aicc as our criteria, we found that we were sitll selecting the exact same ARIMA model, so we will move forwards with this model.

Forecasting Atmospheric CO2 Growth

Now that we have a few different models that do a decent job at modeling our CO2 data, we want to use these models to generate predictions for us.

When will we reach 420 and 500 ppm CO2?

First we want to see when these models forecast that the atmospheric CO2 levels will reach 420ppm and 500ppm. Since we know there is a overall upwards trend as well as a seasonal component to the CO2 levels, when we generate these predictions we will actually want to look at both the first and the last times that the model predicts the atmospheric CO2 levels to be at these values. We know from our previous predictions for the linear model that even by 2020, the predicted CO2 ppm is not 420 ppm. Of course, there may be some variance in the predictions between the ARIMA and linear model but this does serve as a good guideline.

```
## # A fable: 0 x 4 [?]
## # Key:     .model [0]
## # ... with 4 variables: .model <chr>, index <mth>, co2_val <dist>, .mean <dbl>
```

```
## # A fable: 0 x 4 [?]
## # Key:     .model [0]
## # ... with 4 variables: .model <chr>, index <mth>, co2_val <dist>, .mean <dbl>
```

What do we predict CO2 levels will be in 2100 (80 years from now)?

We also want to look at the predictions for our models for the more distant future. Here we will look at what the model predicts will be the CO2 levels in the year 2100. We will also include the confidence intervals for these predictions.

```
## # A tsibble: 12 x 7 [1M]
## # Key:       .model [1]
##    .model         index    co2_val .mean        `90%` `95%_lower` `95%_upper`
##    <chr>          <mth>     <dist> <dbl>       <hilo>       <dbl>       <dbl>
##  1 trend_model 2100 Jan N(683, 19)  683. [676, 691]90        675.        692.
##  2 trend_model 2100 Feb N(684, 19)  684. [677, 692]90        676.        693.
##  3 trend_model 2100 Mar N(685, 19)  685. [678, 693]90        677.        694.
##  4 trend_model 2100 Apr N(687, 20)  687. [680, 694]90        678.        696.
##  5 trend_model 2100 May N(688, 20)  688. [680, 695]90        679.        696.
```

```
## 6 trend_model 2100 Jun N(687, 20)  687. [680, 695]90        679.         696.
## 7 trend_model 2100 Jul N(686, 20)  686. [679, 694]90        678.         695.
## 8 trend_model 2100 Aug N(685, 20)  685. [677, 692]90        676.         693.
## 9 trend_model 2100 Sep N(683, 20)  683. [676, 690]90        674.         692.
## 10 trend_model 2100 Oct N(683, 20)  683. [676, 691]90        675.         692.
## 11 trend_model 2100 Nov N(685, 20)  685. [678, 692]90        676.         694.
## 12 trend_model 2100 Dec N(686, 20)  686. [679, 694]90        678.         695.
```

Report from the Point of View of the Present

One of the very interesting features of Keeling and colleagues' research is that they were able to evaluate, and re-evaluate the data as new series of measurements were released.

Introduction

In this introduction, you can assume that your reader will have **just** read your 1997 report. In this introduction, **very** briefly pose the question that you are evaluating, and describe what (if anything) has changed in the data generating process between 1997 and the present.

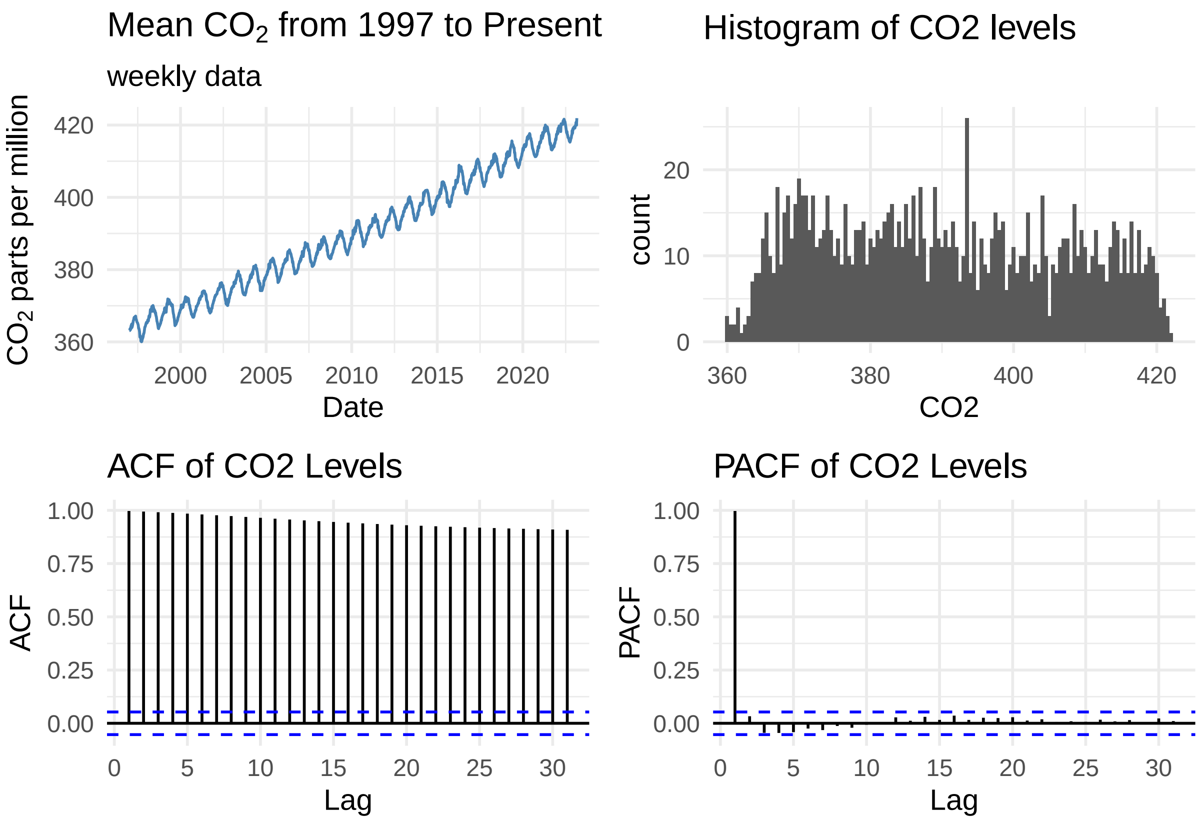
(3 points) Task 1b: Create a modern data pipeline for Mona Loa CO2 data.

The most current data is provided by the United States' National Oceanic and Atmospheric Administration, on a data page [here]. Gather the most recent weekly data from this page. (A group that is interested in even more data management might choose to work with the hourly data.)

Create a data pipeline that starts by reading from the appropriate URL, and ends by saving an object called `co2_present` that is a suitable time series object.

Conduct the same EDA on this data. Describe how the Keeling Curve evolved from 1997 to the present, noting where the series seems to be following similar trends to the series that you "evaluated in 1997" and where the series seems to be following different trends. This EDA can use the same, or very similar tools and views as you provided in your 1997 report.

```
## # A tibble: 6 x 11
##    year month   day decimal average ndays X1.year.ago X10.years.ago
##   <int> <int> <int>   <dbl>   <dbl> <int>       <dbl>         <dbl>
## 1  1997     1     5   1997.    363.     7        362.          349.
## 2  1997     1    12   1997.    363.     7        362.          349.
## 3  1997     1    19   1997.    364.     7        362.          349.
## 4  1997     1    26   1997.    363.     7        363.          349.
## 5  1997     2     2   1997.    364.     7        363.          349.
## 6  1997     2     9   1997.    364.     7        363.          348.
## # ... with 3 more variables: increase.since.1800 <dbl>, time_index <dttm>,
## #   month_index <mth>
```



(1 point) Task 2b: Compare linear model forecasts against realized CO2

Descriptively compare realized atmospheric CO2 levels to those predicted by your forecast from a linear time model in 1997 (i.e. "Task 2a"). (You do not need to run any formal tests for this task.)

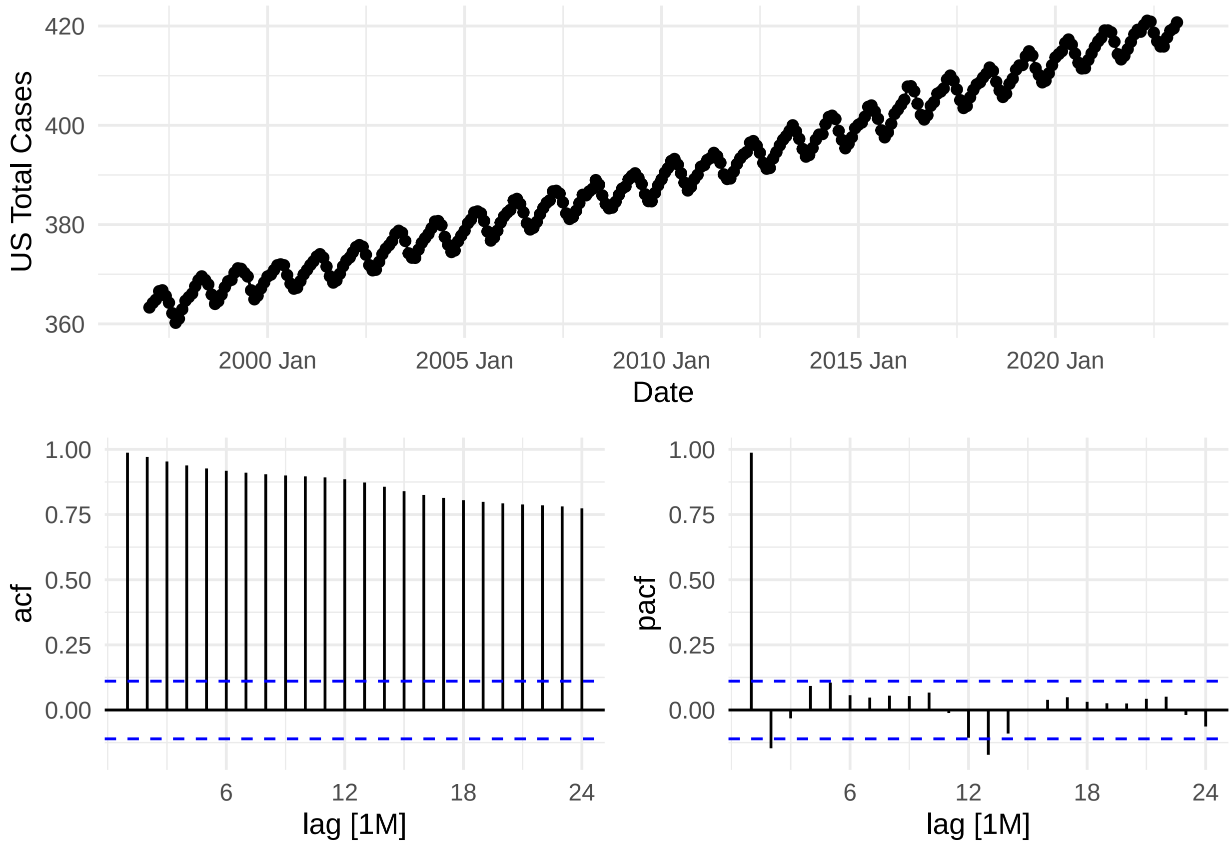
(1 point) Task 3b: Compare ARIMA models forecasts against realized CO2

Descriptively compare realized atmospheric CO2 levels to those predicted by your forecast from the ARIMA model that you fitted in 1997 (i.e. "Task 3a"). Describe how the Keeling Curve evolved from 1997 to the present.

(3 points) Task 4b: Evaluate the performance of 1997 linear and ARIMA models

In 1997 you made predictions about the first time that CO2 would cross 420 ppm. How close were your models to the truth?

After reflecting on your performance on this threshold-prediction task, continue to use the weekly data to generate a month-average series from 1997 to the present, and compare the overall forecasting performance of your models from Parts 2a and 3b over the entire period. (You should conduct formal tests for this task.)



(4 points) Task 5b: Train best models on present data

Seasonally adjust the weekly NOAA data, and split both seasonally-adjusted (SA) and non-seasonally-adjusted (NSA) series into training and test sets, using the last two years of observations as the test sets. For both SA and NSA series, fit ARIMA models using all appropriate steps. Measure and discuss how your models perform in-sample and (psuedo-) out-of-sample, comparing candidate models and explaining your choice. In addition, fit a polynomial time-trend model to the seasonally-adjusted series and compare its performance to that of your ARIMA model.

(3 points) Task Part 6b: How bad could it get?

With the non-seasonally adjusted data series, generate predictions for when atmospheric CO2 is expected to be at 420 ppm and 500 ppm levels for the first and final times (consider prediction intervals as well as point estimates in your answer). Generate a prediction for atmospheric CO2 levels in the year 2122. How confident are you that these will be accurate predictions?

Conclusions

What to conclude is unclear.

While the most plausible model that we estimate is reported in the main, "Modeling" section, in this appendix to the article we examine alternative models. Here, our intent is to provide a skeptic that does not accept our assessment of this model as an ARIMA of order (1,2,3) an understanding of model forecasts under alternative scenarios.