Time Series Project

Jacob Polomsky

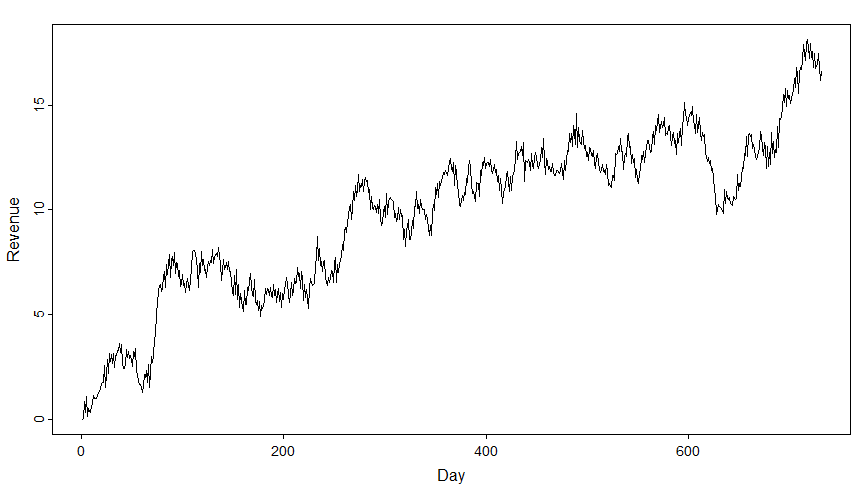
August 17, 2024

A1. Can we forecast our revenue for the upcoming year?

A2. Our goal is to build a model that can predict the revenue for the upcoming year by looking at trends in the dataset.

B. Assumptions of time series are that the data is stationary, that the mean, variance, and autocorrelation are consistent over time, and that no major outliers will throw off the analysis. Autocorrelation measures the correlation of current data points to previous data points; this can help predict models if they have a high autocorrelation to help predict future values by looking at the previous values.

C1



C2.



There were no null or missing values in the dataset, and each value in the day column went up by one, with no skipped day.  The dataset goes from 1 to 731, and each value represents a day. For setting the dates, since we don't have any gaps or missing data, we are measuring two years, so we have 731 points, and each point represents a day, so we set the second part of the expression to 24 for the number of hours in a day.

C3. After running the Augmented Dickey-Fuller, the P-value is 0.02, which is lower than 0.05, so we reject the null hypothesis and go with the alternative hypothesis that the data is stationary.C3. After running the Augmented Dickey-Fuller, the P-value is 0.02, which is lower than 0.05, so we reject the null hypothesis and go with the alternative hypothesis that the data is stationary.

A white background with black text

Description automatically generated

C4



I used this code the read the csv file



This code is used to plot the data.



This Code checks to see if there are any NA values or missing values that might cause a gap there are not any.



Next Link checks to see if there are any duplicates and there aren’t any.



This is to look at the data manually see if there are any missing days which there aren’t any.



This is used to check for any outliers in the data.



This code converts the data frame to a time series data frame.



Ran the Adfuller test and found that the data was stationary



ACF found that is an AR model

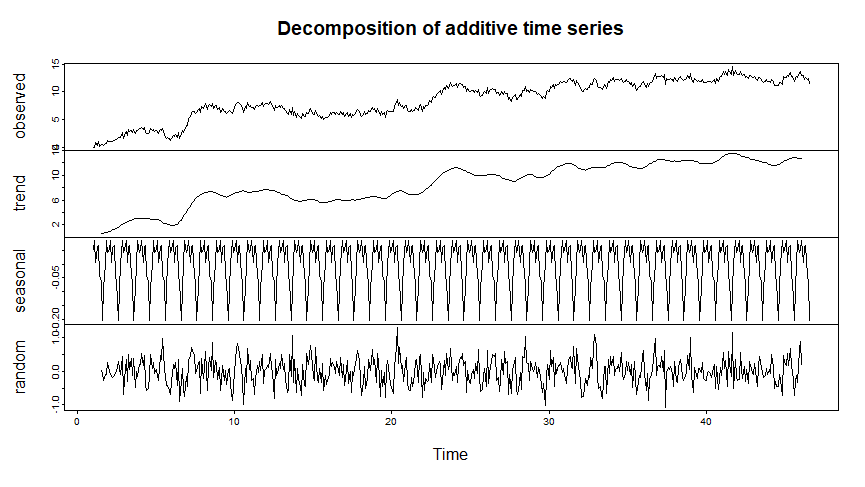




I used this code to split the data into a 75/25 split.

C5. Timeseries.csv, train.csv, test.csv

D1

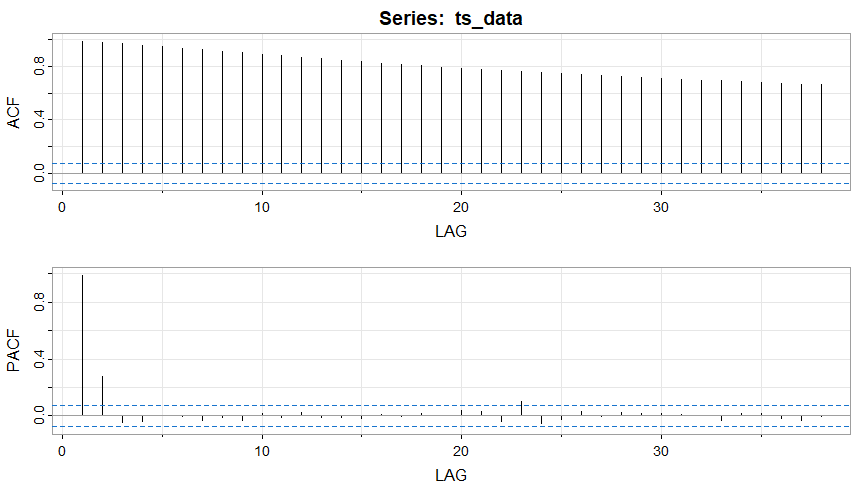


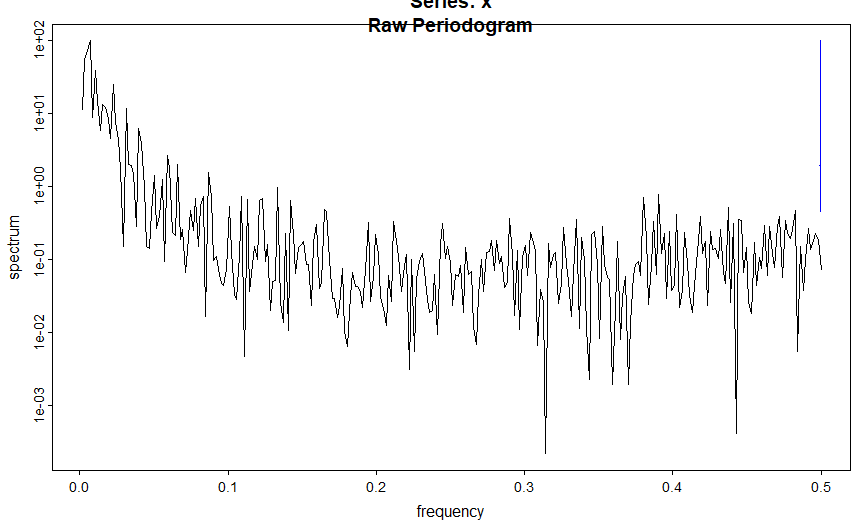
The trend component of the decomposition chart shows that there has been a positive trend over time. As time increases, the price increases

Seasonality: The seasonality of the decomposition chart shows that there is a solid seasonal trend because the pattern constantly repeats itself.

The random plot residual plot shows that the residuals are distributed relatively evenly.

The decomposition chart shows that there is a strong trend that also has a vital seasonal component.



The ACF slows tails off over the lags, while the PACF cuts off over a couple of lags. These two together show that we are dealing with an AR model, not an MA model. If it also had an MA model, it would trail off.

The spectral density plot shows signs of the data not being stationary. Still, when you factor in the ADfuller test being stationary and seasonality showing a repeating pattern, this shows signs of the data being seasonality nonstationary.

D2.

A black text with black letters

Description automatically generated with medium confidence

A black and white text

Description automatically generated

 I picked this model p = 1 d = 0 q = 0 because the ACF and pacf show that there is an AR model but not an MA model, so p = 1 and q= 0. After running the ADFuller, the test showed that the data was stationary. The seasonal decompose plot showed a strong seasonal pattern, and the spectral density plot showed signs of the data being nonstationary, while ADFuller was stationary; I made a seasonal difference of D = 1.

D3.

A graph with a red line

Description automatically generated

D4.

Stationarity



A white background with black text

Description automatically generated

ACF and PACF

A close up of a computer code

Description automatically generated

A white background with black and white text

Description automatically generated

A screen shot of a graph

Description automatically generated

Model

A black text with blue and white text

Description automatically generated with medium confidence

A screenshot of a computer code

Description automatically generated

Summary



A white background with black text

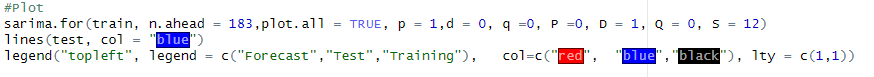
Description automatically generated

Check Residuals

A screenshot of a graph

Description automatically generated

Prediction



A graph of a graph of a wave

Description automatically generated with medium confidence

* to see the plot made D3

D5. Timeseries.R

E1. The best model based on autocorrelation that looks at AIC is a (1,1,0) (0,1,0,12) model. With an ACF that trails down and a PACF that cuts off, we are looking at an AR model. Now, with our Augmented Dickey-Fuller, we have stationary data with the Decompose plot showing a repeating trend when you use a difference of 12. The prediction interval for the data is one day, and our entire dataset is two years. So, looking at the data, over 50% prediction is considered extreme, so it is best to have a prediction for a maximum offer of one year, which would be half our dataset. Looking at the RMSE MAE, we are off by around .5 for our predictions and RMSE of 0.7, showing that the data stays close to what we predicted.

A white background with black text

Description automatically generated

E2

A graph showing a graph

Description automatically generated with medium confidence

A graph with a red line

Description automatically generated

E3. This model performs very well on the test set, with most of the revenue falling within the first confidence interval. The model follows the pattern of the data well, but it did struggle when it had the big downturn. This dataset, being close to two years, would be a good model for predicting what revenues are projected for the next year. I think the model performs better than what is being represented with RMSE and MAE because the model follows the trend very well, but the price going way down, I think, makes it worse than how it projects because, after the downturn, it goes back to being right on point. I think the model will perform relative accuracy for the trend over the course of a year.

F. Timeseries.pdf