Katherine Olson

998992204

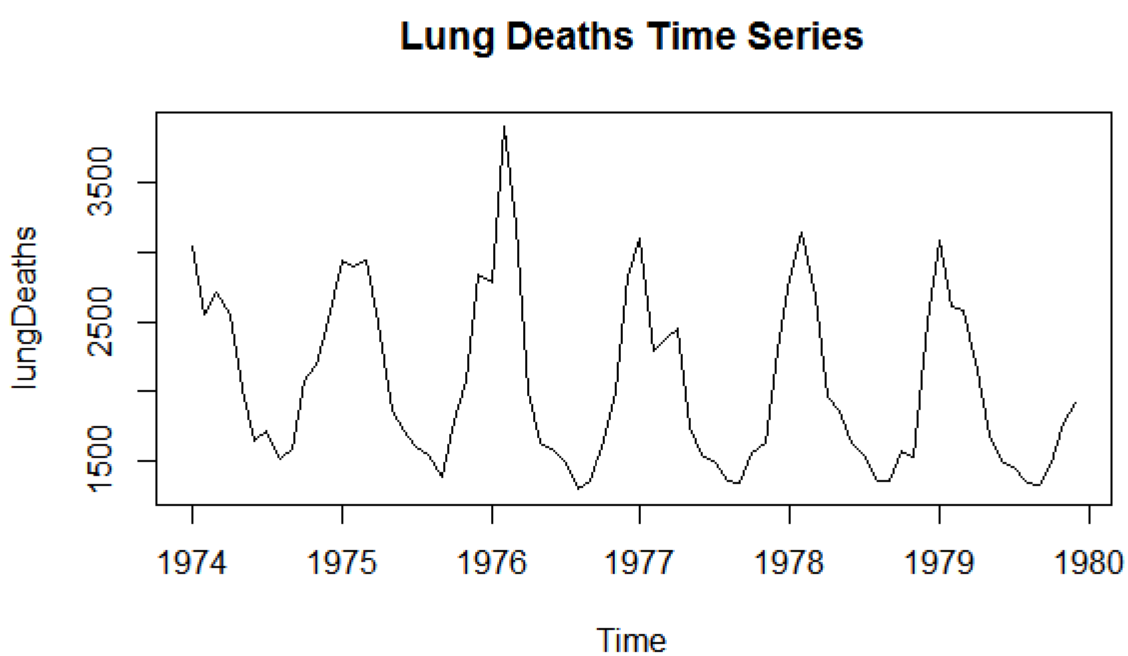
STA 137 Project II

Introduction

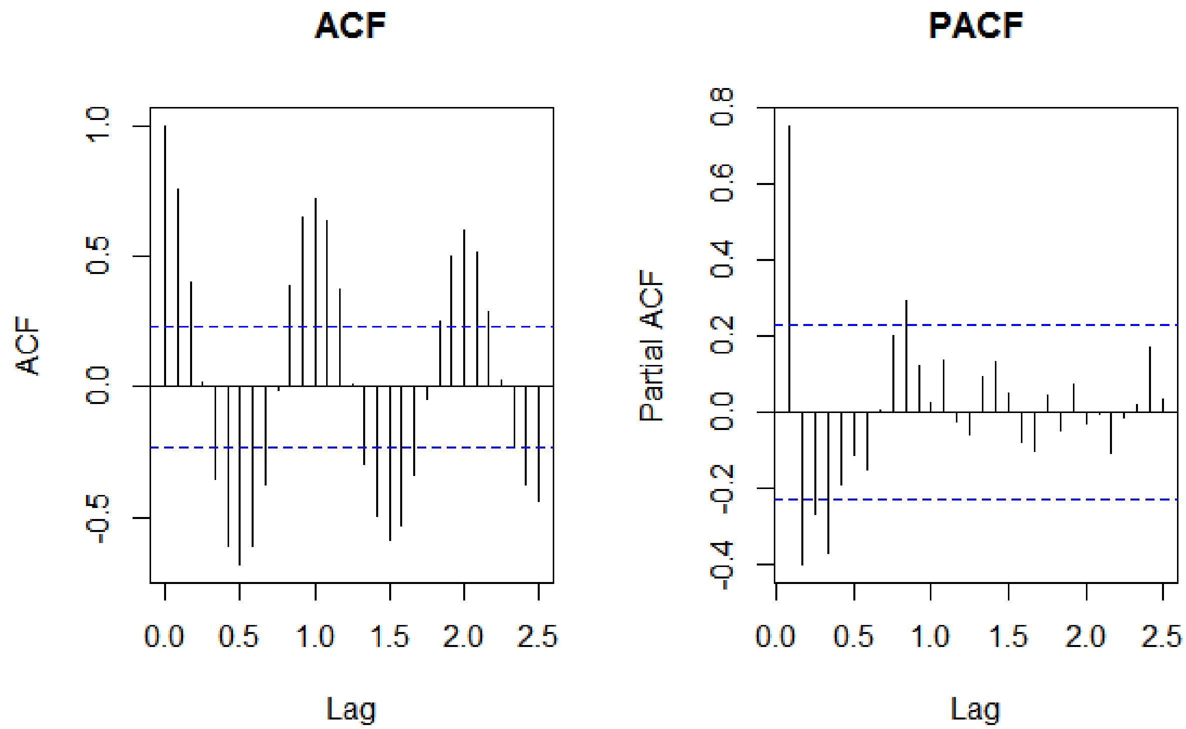
I am working with a time series from the dataset package called ldeaths that contains the number of monthly deaths from bronchitis, emphysema and asthma in the UK from 1974–1979. This is a time series dataset because it is taken over a period of time, in this case monthly for a few years. This dataset is important to analyze because we want to do all we can to stop lung deaths. It will be particularly useful to see if there is a seasonal pattern to help pinpoint the cause and prevent future lung deaths. If this dataset was more recent, it would also be of interest to predict future values, but those future values are still interesting to look at.

Materials, Methods, and Results

I started by plotting the series:



Looking at the series, there is clearly a seasonal component every year, which we will come back to later in the analysis. The series actually looks stationary for once, so I am not applying any transformations or differences. The mean looks constant and so does the variance, so there is no need. So next I looked at the ACF and PACF:

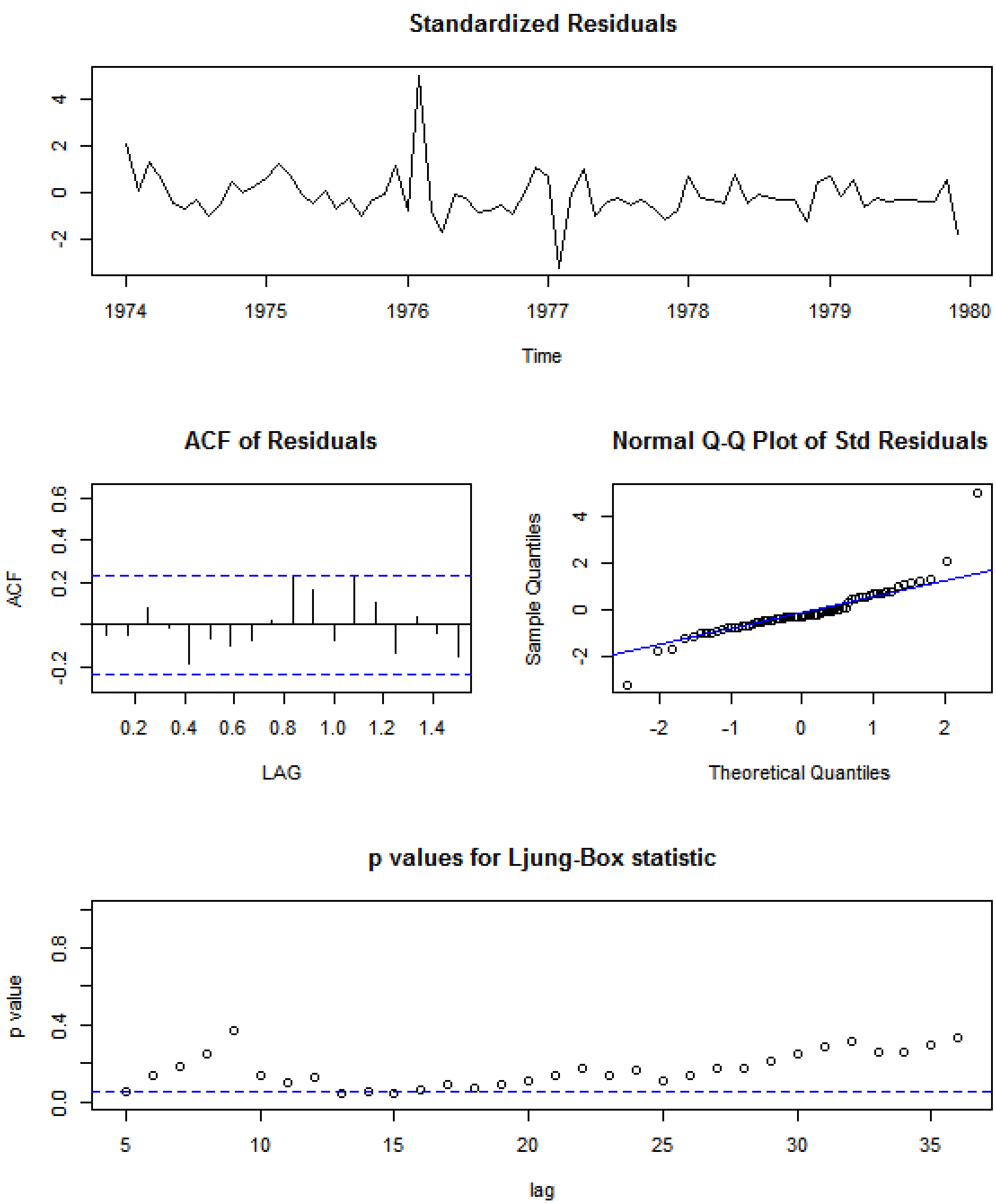


Looking at the ACF, the function decays slowly in a repeated pattern that switches from positive to negative. The PACF kind of cuts off after one lag, but it also kind of looks like it decays slowly. So the model is AR, but the order is not super clear. I will try a few such as ARIMA(1, 0, 0), ARIMA(2, 0, 0) and a few others and compare AIC values to decide. For the seasonal element, it is clear that s = 12 from the plot of the time series. The PACF mostly cuts off after 12 lags, but it could go until 24 lags. So P could equal 1 or 2. The ACF has an up and down pattern that is 6 lags long and the repeated pattern is 12 lags long. This cuts off way past what the ACF shows at k * s where k = 1, 2, ..., so the seasonal element is either $(1, 0, 0)_{12}$ or $(2, 0, 0)_{12}$.

Now I need to try all of these model possibilities and pick the best one using AIC:

Model	**AIC**
ARIMA(1, 0, 0)X$(1, 0, 0)_{12}$	1056.37
ARIMA(1, 0, 0)X$(2, 0, 0)_{12}$	1040.13
ARIMA(2, 0, 0)X$(1, 0, 0)_{12}$	1056.31
ARIMA(2, 0, 0)X$(2, 0, 0)_{12}$	<u>1039.93</u>
ARIMA(3, 0, 0)X$(1, 0, 0)_{12}$	1055.63
ARIMA(3, 0, 0)X$(2, 0, 0)_{12}$	1040.87

This tells us that the best model is ARIMA(2, 0, 0)X$(2, 0, 0)_{12}$. The results of the fit if this model is:

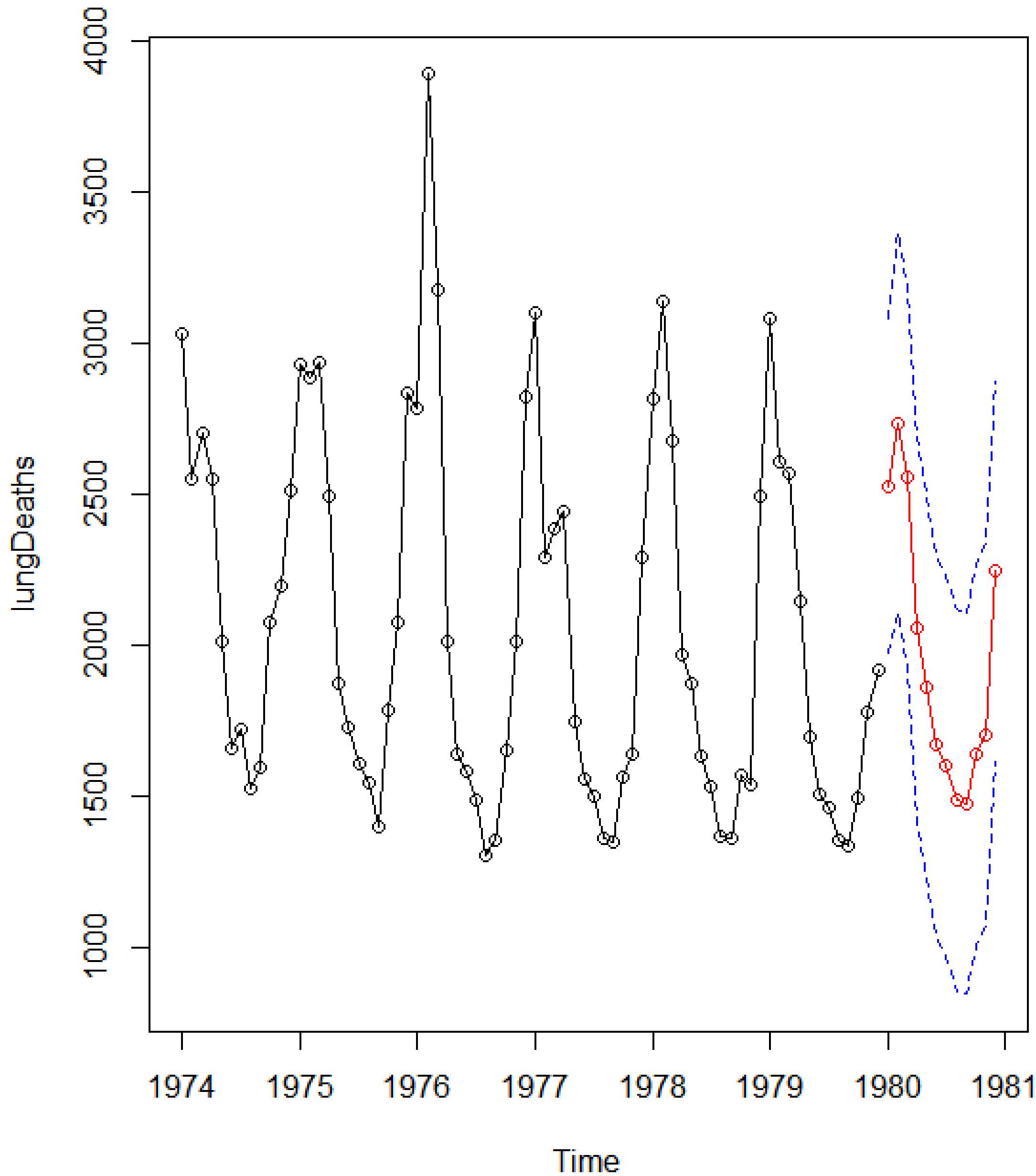


The residuals look mostly white, but not perfect. Expressing the final model in terms of the estimated parameters:

AR(1)	**AR(2)**	**sAR(1)**	**sAR(2)**	**Mean**
0.5388	-0.1903	0.2955	0.5244	2062.8113

Sigma^2 estimate = 76590

Now I am going to forecast future values for the next 12 months:



These values are: 2527.886, 2736.960, 2554.078, 2056.083, 1859.235, 1672.410, 1603.836, 1487.306, 1476.937, 1635.798, 1702.838, and 2243.678.

The graph shows us that the forecast continues the seasonal pattern with a bit of a lower peak and drop.

Conclusion and Discussion

I used a time series of lung deaths in the UK from 1974–1979 for seasonal ARIMA analysis. Right away there was a clear seasonal pattern with s = 12. The data was already stationary so I went right into looking at the ACF and PACF. These functions told us that the model had an AR component with an order not super clear but probably 1 or 2. The functions also told us that the seasonal portion of the model was AR with order either 1 or 2. Fitting all the models and comparing AIC, the best model was ARIMA(2, 0, 0)X$(2, 0, 0)_{12}$. Using that model to forecast future values showed that the seasonal trend continues. From this analysis, the seasonal component makes forecasting especially useful. Seeing that this seasonal trend continues, if it were 1979, would be very useful. This would have been useful for people of the UK to use to try and fight against the number of lung deaths and figure out why there are more deaths during certain months.

Code Appendix

```
library(datasets)
library(astsa)
lungDeaths = ldeaths
plot.ts(lungDeaths, main = "Lung Deaths Time Series")
par(mfrow = c(1, 2))
acf(lungDeaths, main = "ACF", lag.max = 30)
pacf(lungDeaths, main = "PACF", lag.max = 30)
sarima(lungDeaths, 1, 0, 0, 1, 0, 0, 12) # AIC = 1056.37
sarima(lungDeaths, 1, 0, 0, 2, 0, 0, 12) # AIC = 1040.13
sarima(lungDeaths, 2, 0, 0, 1, 0, 0, 12) # AIC = 1056.31
sarima(lungDeaths, 2, 0, 0, 2, 0, 0, 12) # AIC = 1039.93
sarima(lungDeaths, 3, 0, 0, 1, 0, 0, 12) # AIC = 1055.63
sarima(lungDeaths, 3, 0, 0, 2, 0, 0, 12) # AIC = 1040.87
fit = sarima(lungDeaths, 2, 0, 0, 2, 0, 0, 12)
par(mfrow = c(1, 1))
forecast = sarima.for(lungDeaths, 12, 2, 0, 0, 2, 0, 0, 12)
preds = forecast$pred[1:12]
```