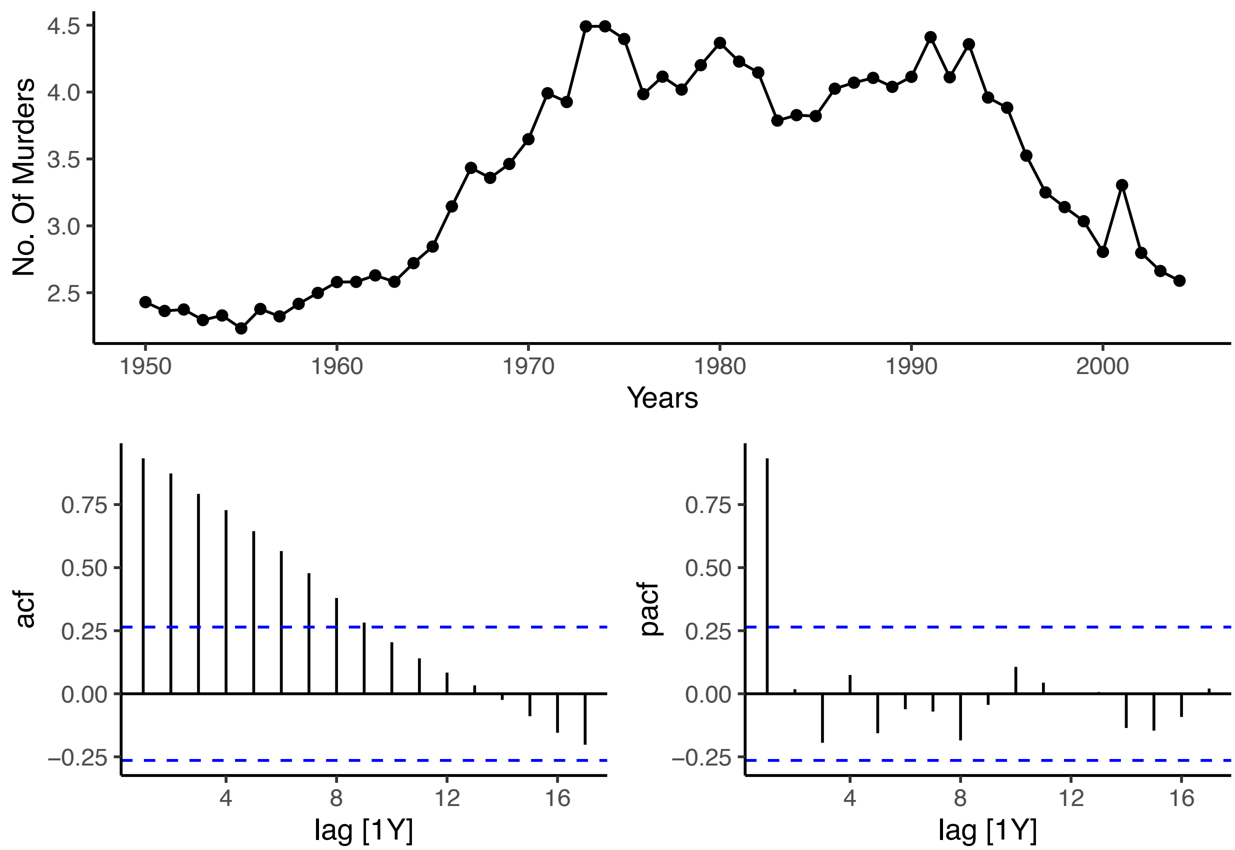
BUAN6357_HomeWork-2_Peddisetty

Tharun Peddisetty

9/24/2020

```
pacman::p_load(fpp3, patchwork, purrr,fpp2,knitr, forecast, ggplot2, rmarkdown)
theme_set(theme_classic())
```

```
data <- as_tsibble(wmurders)
data %>% gg_tsdisplay(value, plot_type = 'partial') + xlab("Years") + ylab("No. Of Murders")
```




```
data %>% features(value, unitroot_kpss)
```

```
## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1     0.633      0.0196
```

1) Choice of Model

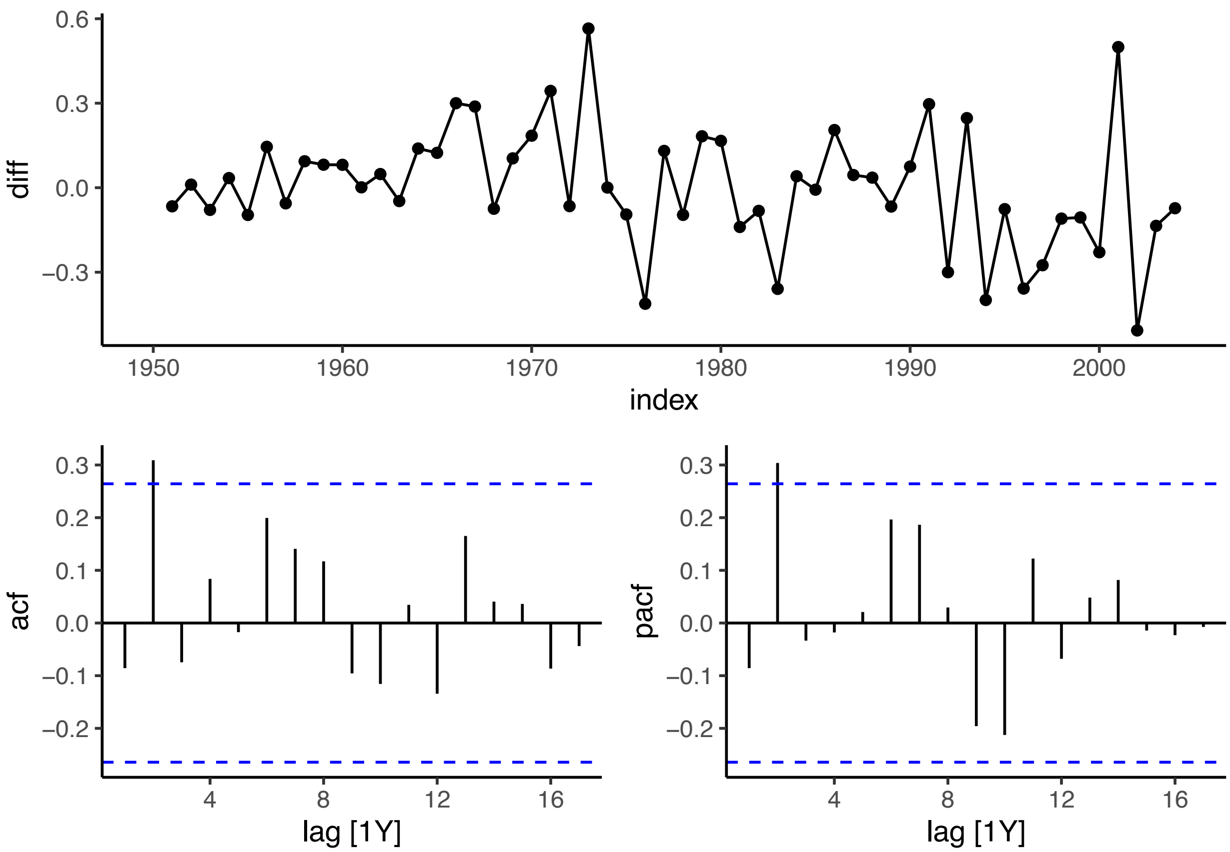
The KPSS test reveals that the series is not stationary with a p-value < 0.05. Therefore, we have considered first difference of the series. After studying the below model plots, there seems to be a sinusoidal pattern in both acf and pacf plots. Therefore both Auto Regressive and Moving Average models are possible. From the plots below we can decide on p and q values: For AR model: p=2 (from pacf plot) For MA model: q=2 (from acf plot)

Since the AR and MA models choices are equally likely candidate models, I'm choosing MA as per the specification. Final Model choice: ARIMA (0,1,2)

```
data%>% mutate(diff = difference(value)) %>%
  gg_tsdisplay(diff, plot_type = 'partial')
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



2) Should we include a constant ?

I don't think a constant needs to be added to the final model because the average of the first difference of the series is constant around zero. Adding a constant will shift the mean upward or downward depending on the sign of the constant.

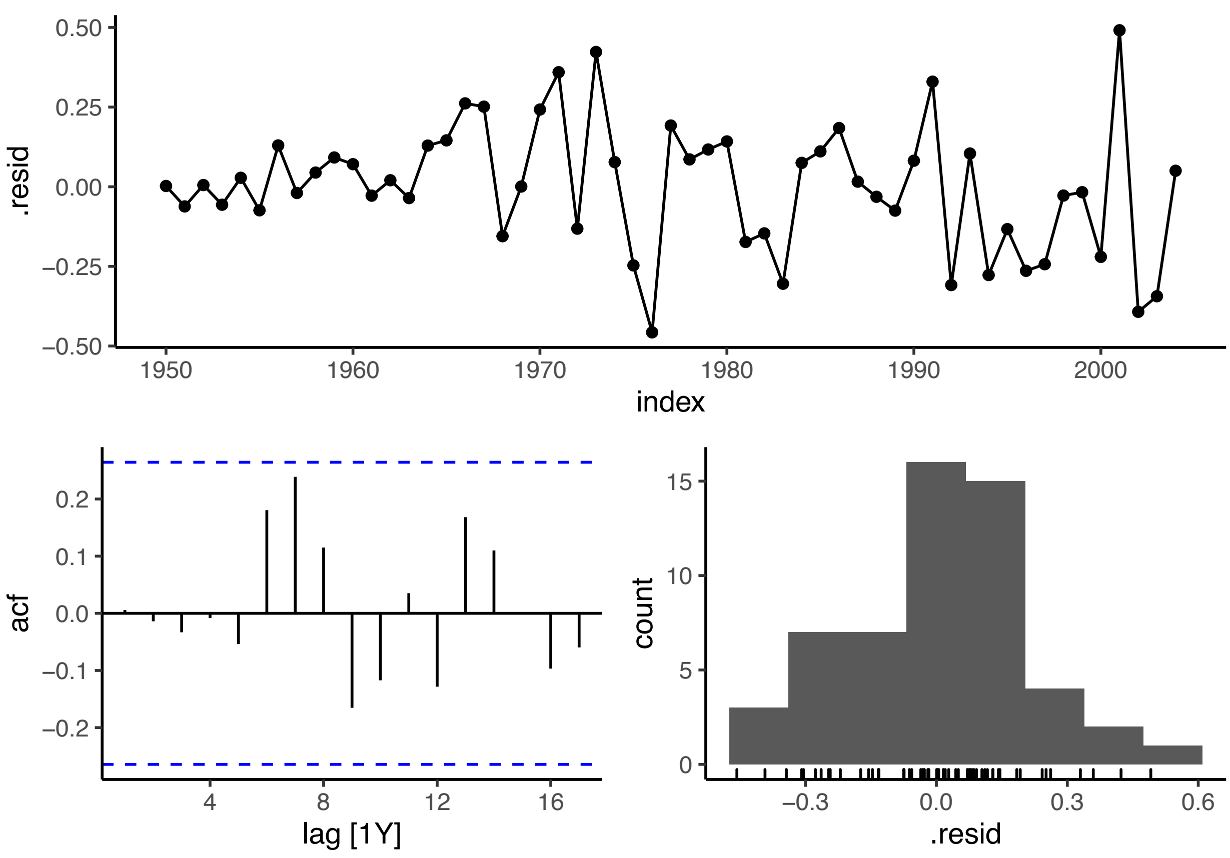
3) Examining ARIMA(0,1,2) model

The model is descent enough to be considered because the residual plot looks almost normal with mean =0 signifying the residual series has white noise. Furthermore, the Ljung-Box test reveals that the residuals are uncorrelated. Satisfying the above mentioned criteria means that there is no more information left in the first difference series to be extracted. However, it is also a good idea to check the AR model before coming to a final conclusion if this model is satisfactory or not.

```
fit1 <- data %>% model(arima = ARIMA(value ~ pdq(0, 1, 2))) %>% report
```

```
## Series: value
## Model: ARIMA(0,1,2)
##
## Coefficients:
##           ma1     ma2
##       -0.0660  0.3712
## s.e.   0.1263  0.1640
##
## sigma^2 estimated as 0.0422:  log likelihood=9.71
## AIC=-13.43   AICc=-12.95   BIC=-7.46
```

```
gg_tsresiduals(fit1)
```




```
#Checking for autocorrelation using Ljung-Box Test
augment(fit1) %>% features(.resid, ljung_box, lag=2, dof =2)
```

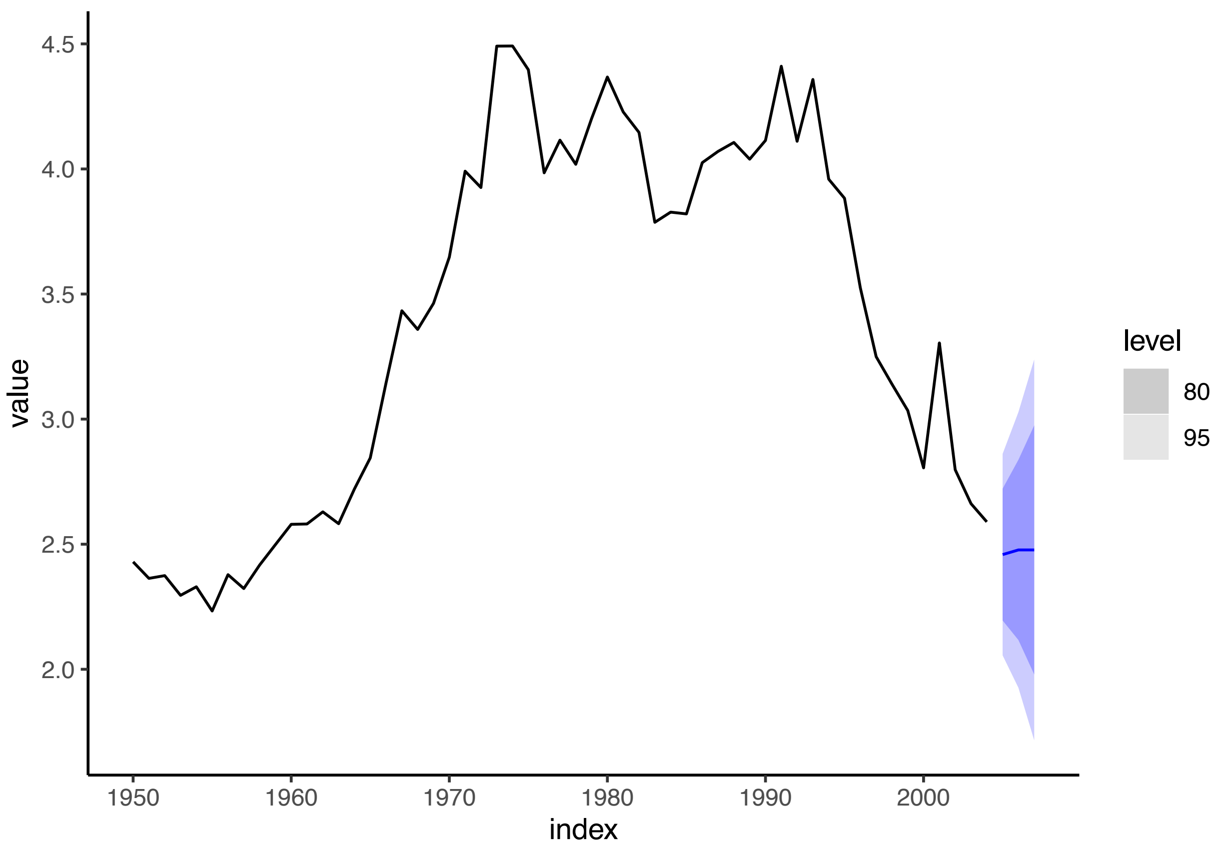
```
## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
##   <chr>    <dbl>     <dbl>
## 1 arima   0.0135         0
```

4) Forecasting 3 times ahead

```
fit1 %>% forecast(h = 3)
```

```
## # A fable: 3 x 4 [1Y]
## # Key:     .model [1]
##   .model index         value .mean
##   <chr>  <dbl>        <dist> <dbl>
## 1 arima   2005 N(2.5, 0.042)  2.46
## 2 arima   2006 N(2.5, 0.079)  2.48
## 3 arima   2007  N(2.5, 0.15)  2.48
```

```
fit1 %>% forecast(h = 3) %>% autoplot(data)
```



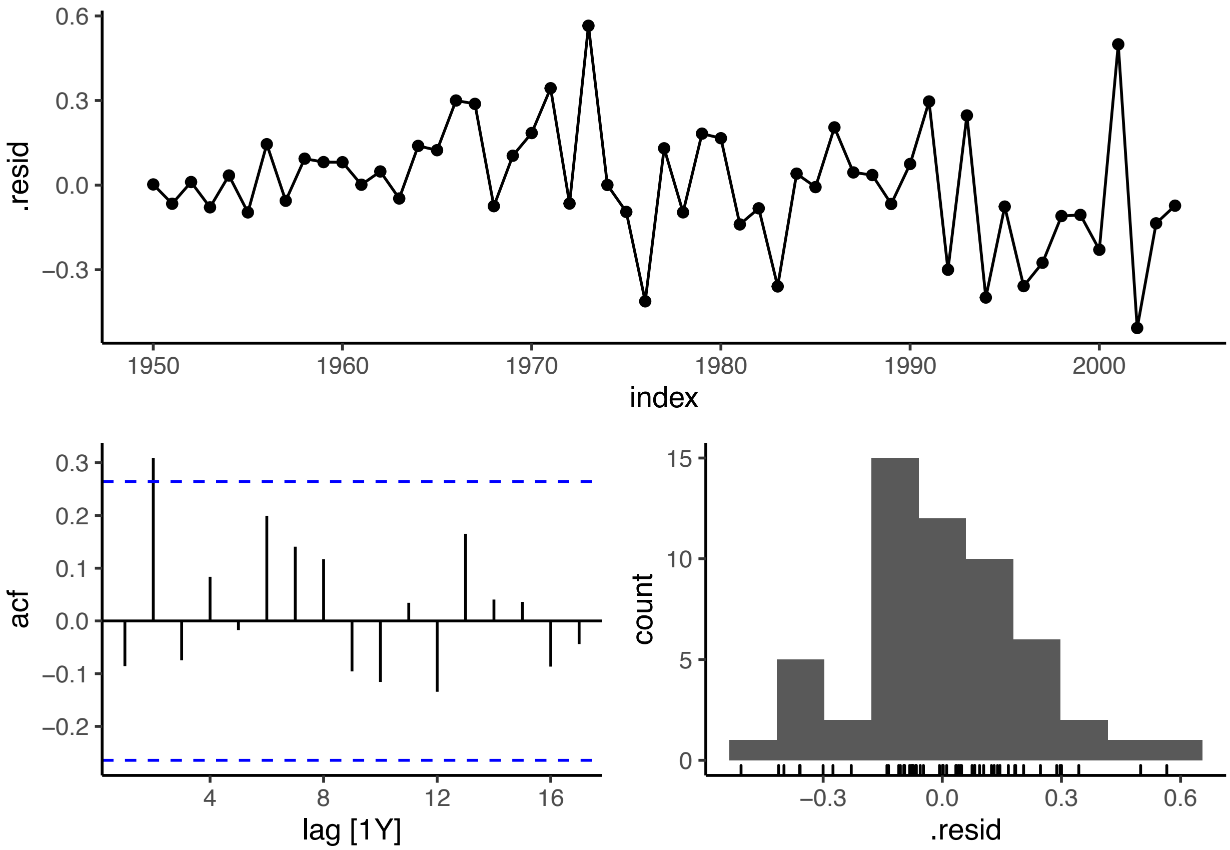
5) ARIMA Model Choice

No, ARIMA doesn't give the same model as I chose. It gives ARIMA(0,1,0) as the best model. However, I believe that my model is better than this one because the information criterion AICc value of ARIMA(0,1,2) model is lesser than ARIMA(0,1,0). [AICc=-12.95 < AICc=-11.38]

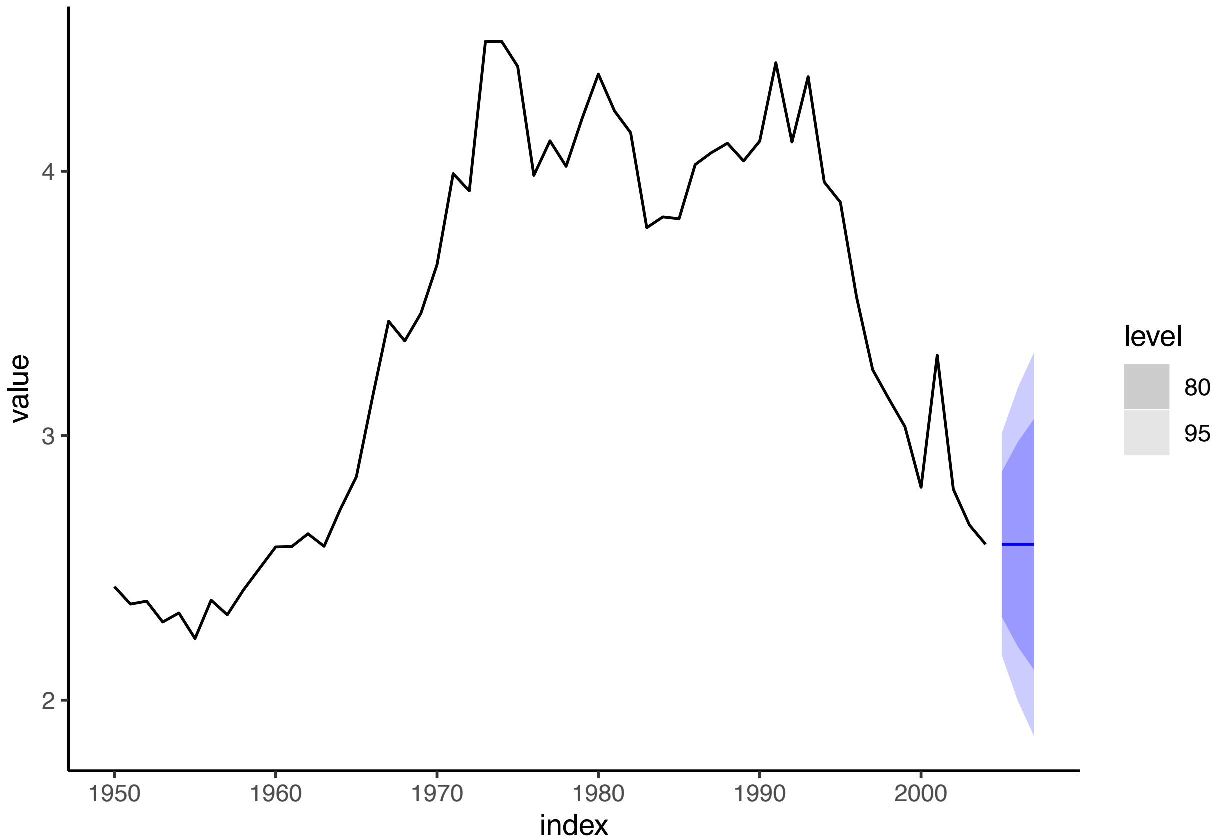
```r
fit2 <-data %>% model(ARIMA(value ~ pdq(d=1))) %>% report()
```

```
## Series: value
## Model: ARIMA(0,1,0)
##
## sigma^2 estimated as 0.04563:  log likelihood=6.73
## AIC=-11.46   AICc=-11.38   BIC=-9.47
```

```r
gg_tsresiduals(fit2)
```




```r
fit2 %>% forecast(h = 3) %>% autoplot(data)
```


value
4
3
2
1950
1960
1970
1980
1990
2000
index
level
80
95