Arima Assignment

Sujeeth Shetty

2/23/2020

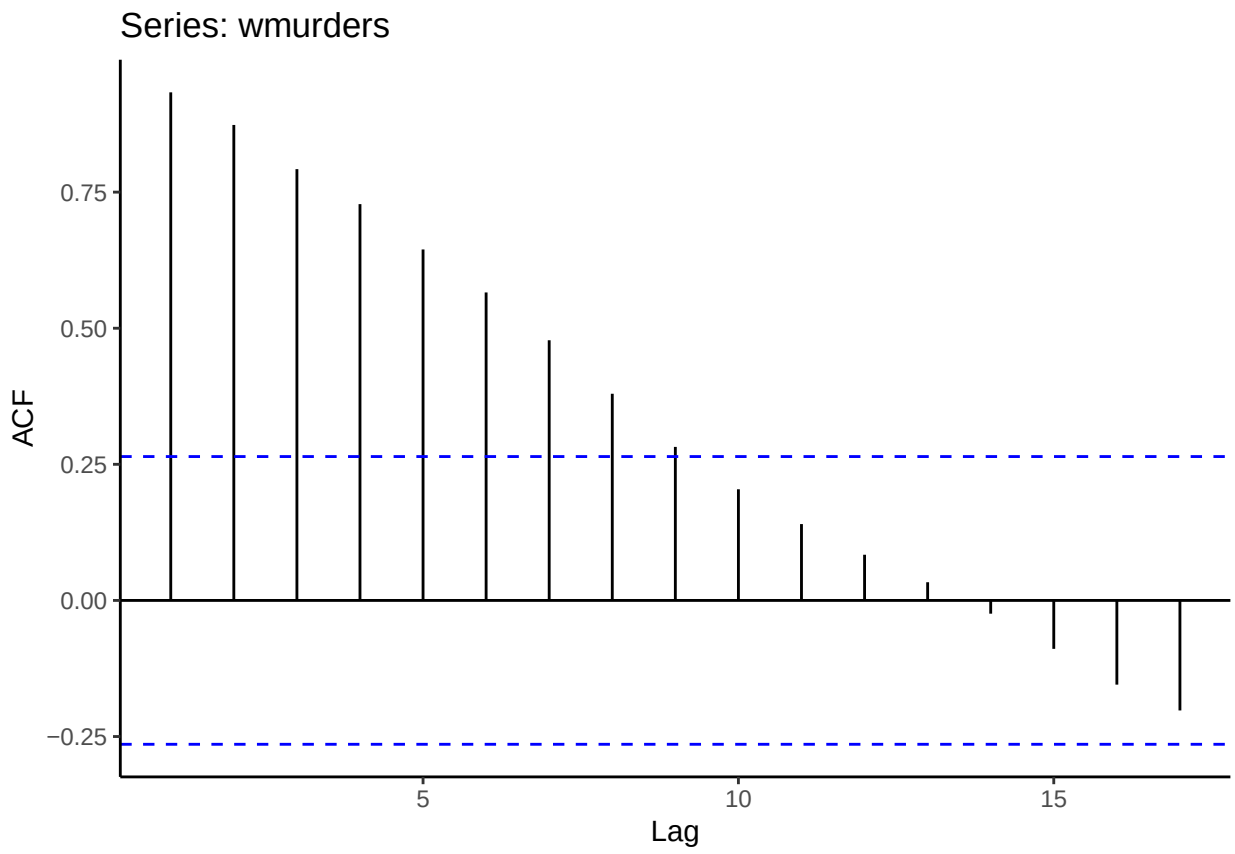
R Markdown

##wmurders data set from the fpp2 package

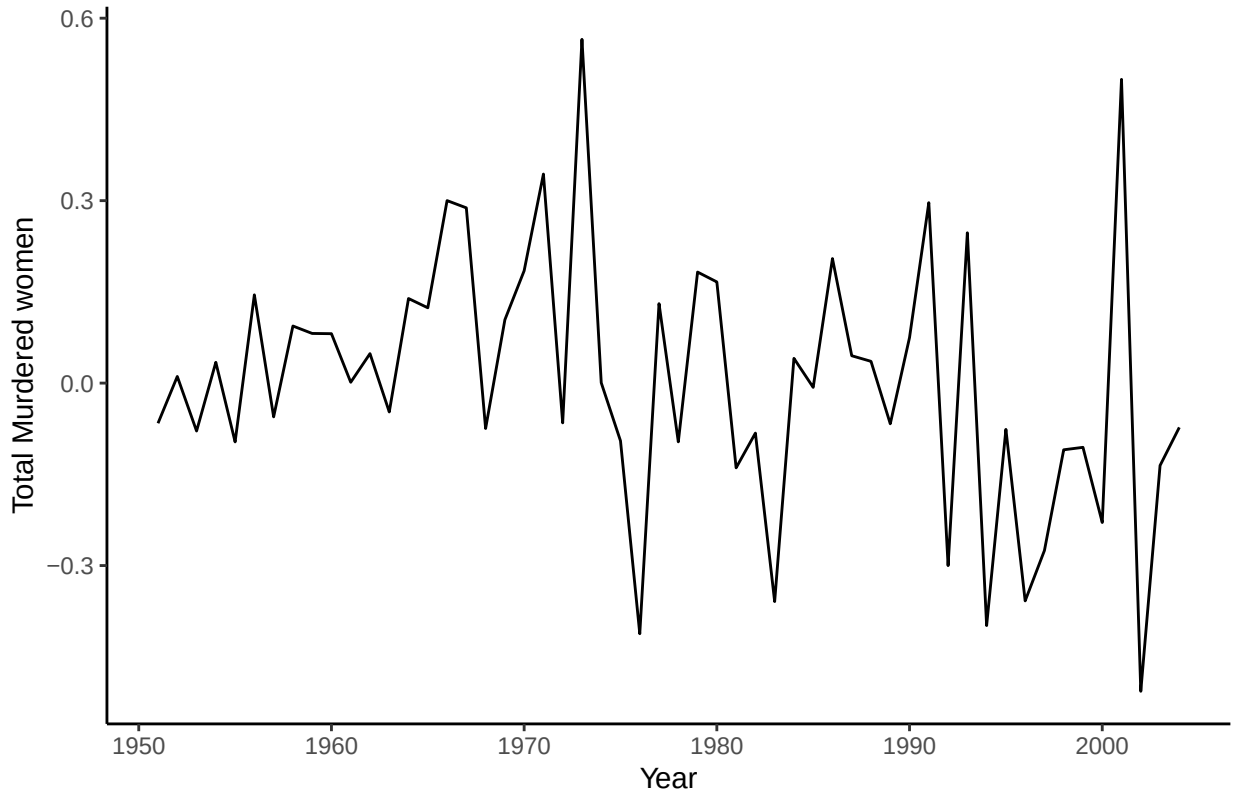
```
pacman::p_load(fpp2, urca, gridExtra)
theme_set(theme_classic())
```

```
#before difference
autoplot(wmurders) +
  ylab("Total Murdered women") + xlab("Year")
```

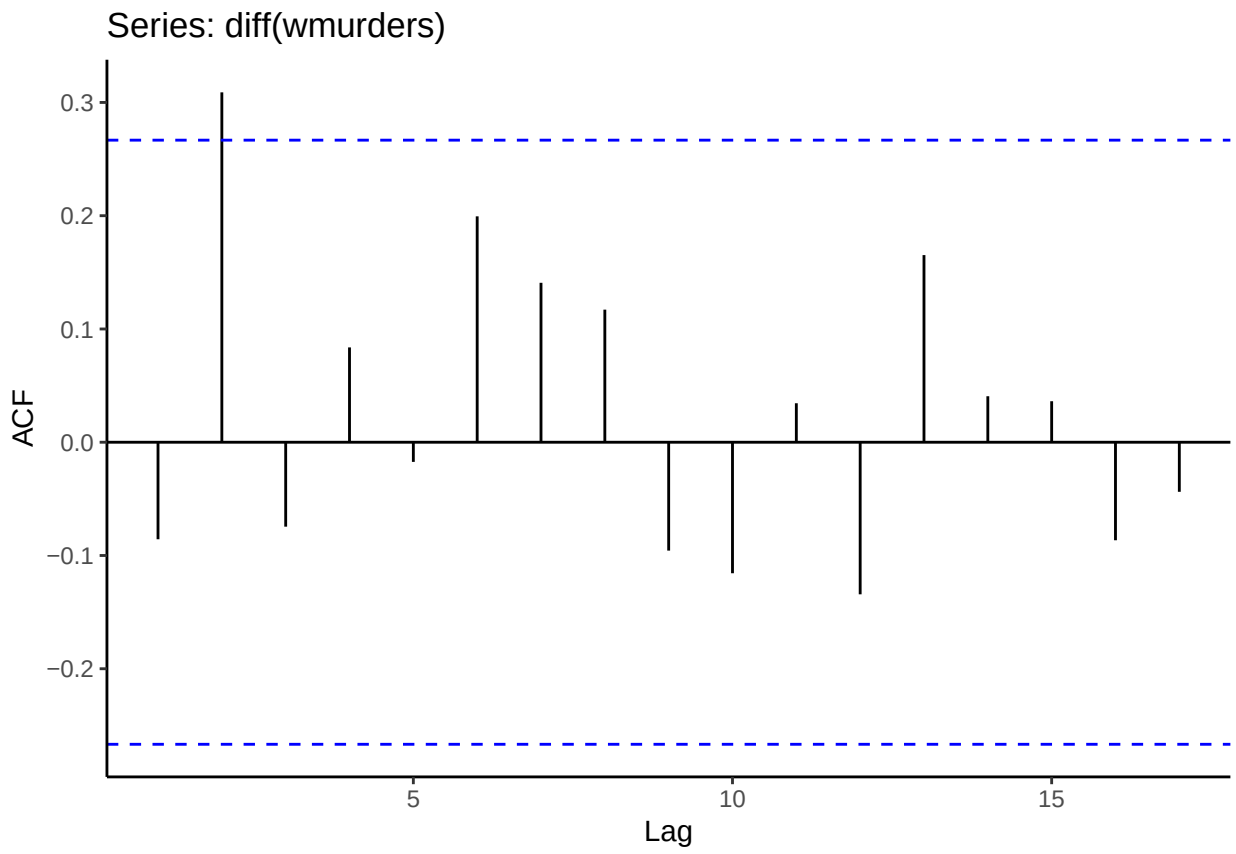

```
ggAcf(wmurders)
```




```
#After Difference
autoplot(diff(wmurders)) + ylab("Total Murdered women") + xlab("Year")
```




```
ggAcf(diff(wmurders))
```




```
#Since there is no seasonality i wont be performing Seasonal difference.
```

There is trend in the data, hence the data might not be stationary. Also, looking at the ACF plot, it appears we need differencing to reduce trend

```
##Null Hypothesis the data is stationary
summary(ur.kpss(wmurders))
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.6331
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

```
#test-statistic is higher, we gonna reject null hyp that it is a stationary

# Number of differences needed to make the data stationary
ndiffs(wmurders)
```

```
## [1] 2
```

```
wmurders %>% diff(differences=2) %>% ur.kpss() %>% summary()
```

```
##
## ########################
## # KPSS Unit Root Test #
## ########################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0458
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

```
#after taking two differences, the series seems to be stationary
```

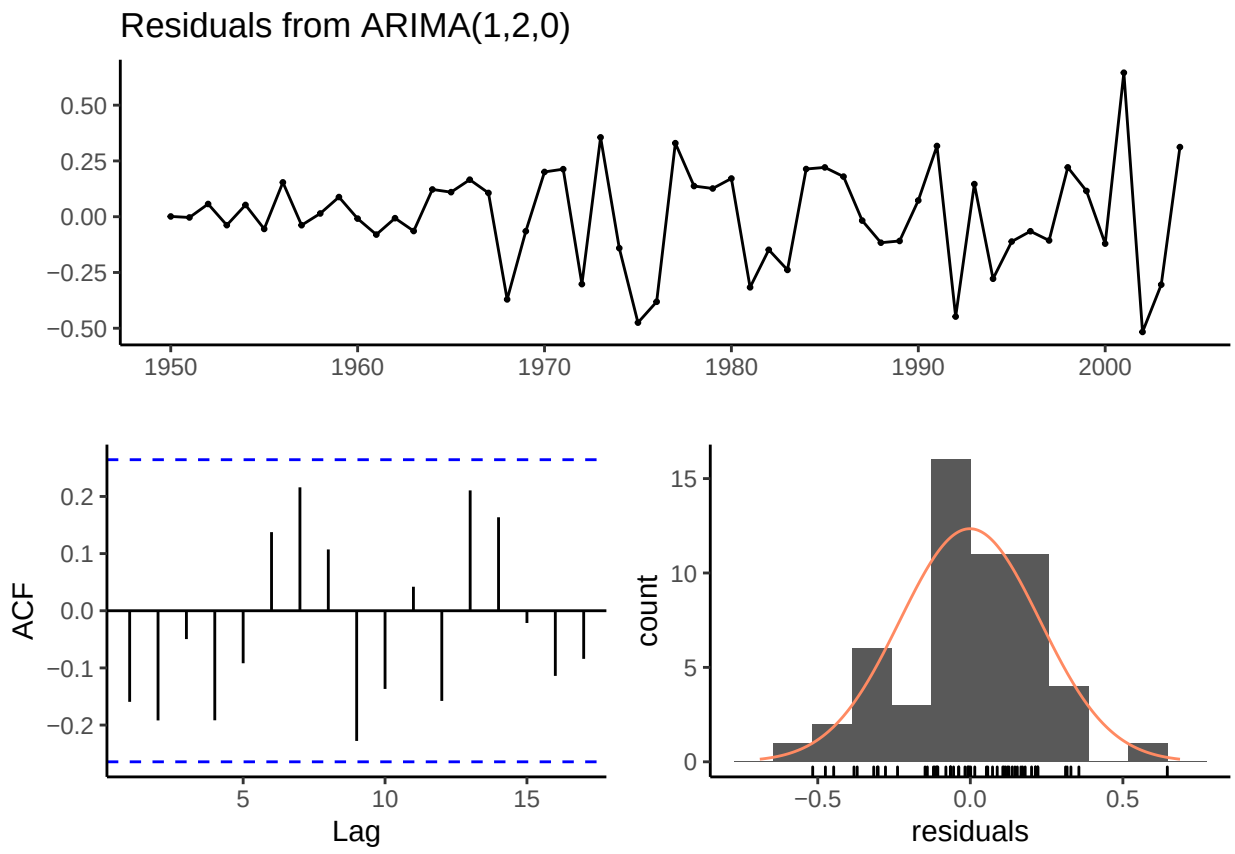
I would not include a constant since we are computing second-order differences.

```
# Specify (p, d, q) - (1,2,0)

(fit1 <- Arima(wmurders, order = c(1,2,0)))
```

```
## Series: wmurders
## ARIMA(1,2,0)
##
## Coefficients:
##           ar1
##       -0.6719
## s.e.   0.0981
##
## sigma^2 estimated as 0.05471:  log likelihood=2
## AIC=0   AICc=0.24   BIC=3.94
```

```
checkresiduals(fit1)
```



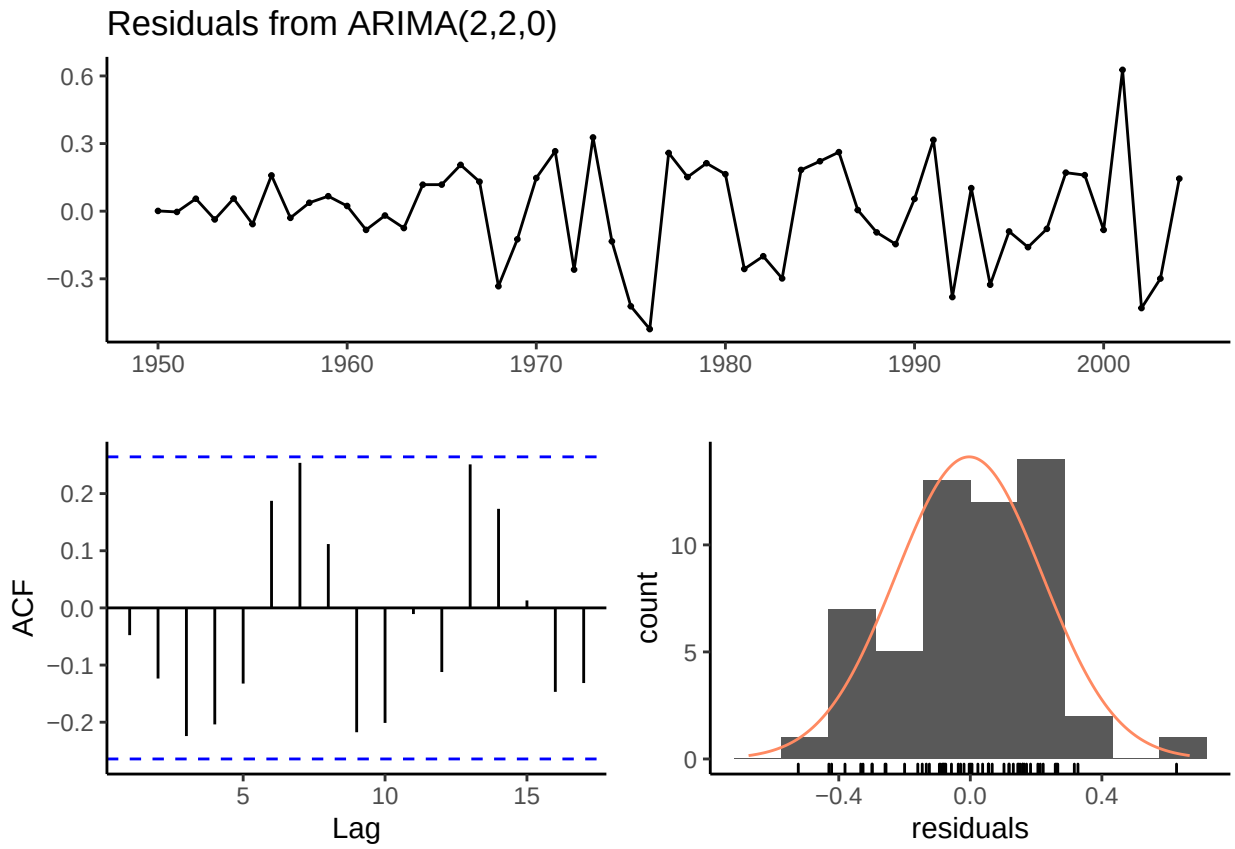

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,2,0)
## Q* = 16.452, df = 9, p-value = 0.05802
##
## Model df: 1.   Total lags used: 10
```

```
# Specify (p, d, q) - (2,2,0)

(fit2 <- Arima(wmurders, order = c(2,2,0)) )
```

```
## Series: wmurders
## ARIMA(2,2,0)
##
## Coefficients:
##           ar1      ar2
##       -0.8289  -0.2246
## s.e.   0.1346   0.1353
##
## sigma^2 estimated as 0.05292:  log likelihood=3.34
## AIC=-0.68   AICc=-0.19   BIC=5.23
```

```
checkresiduals(fit2)
```

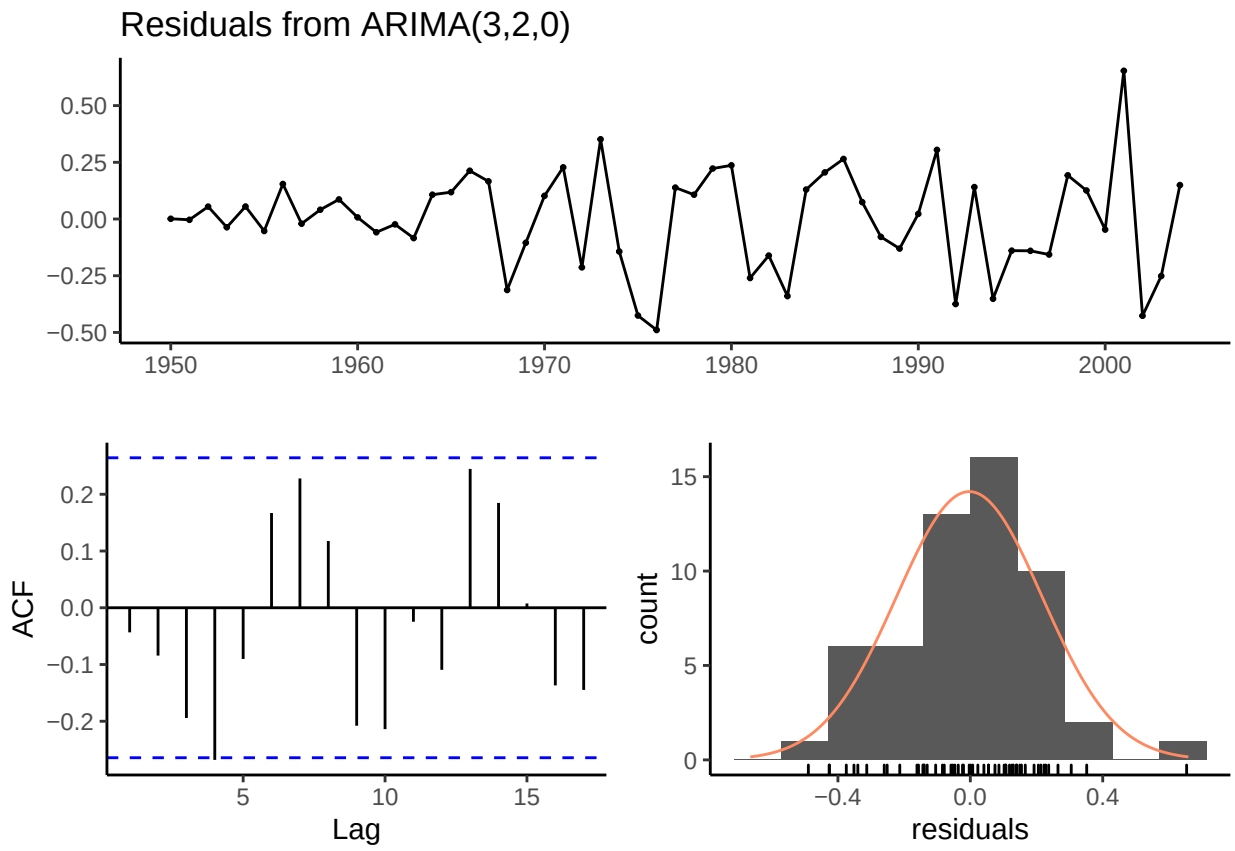



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,2,0)
## Q* = 21.05, df = 8, p-value = 0.007015
##
## Model df: 2.   Total lags used: 10
```

```
# Specify (p, d, q) - (3,2,0)
(fit3 <- Arima(wmurders, order = c(3,2,0)) )
```

```
## Series: wmurders
## ARIMA(3,2,0)
##
## Coefficients:
##          ar1      ar2      ar3
##       -0.855  -0.3561  -0.1753
## s.e.   0.135   0.1747   0.1502
##
## sigma^2 estimated as 0.05256:  log likelihood=4.01
## AIC=-0.02   AICc=0.81   BIC=7.86
```

```
checkresiduals(fit3)
```



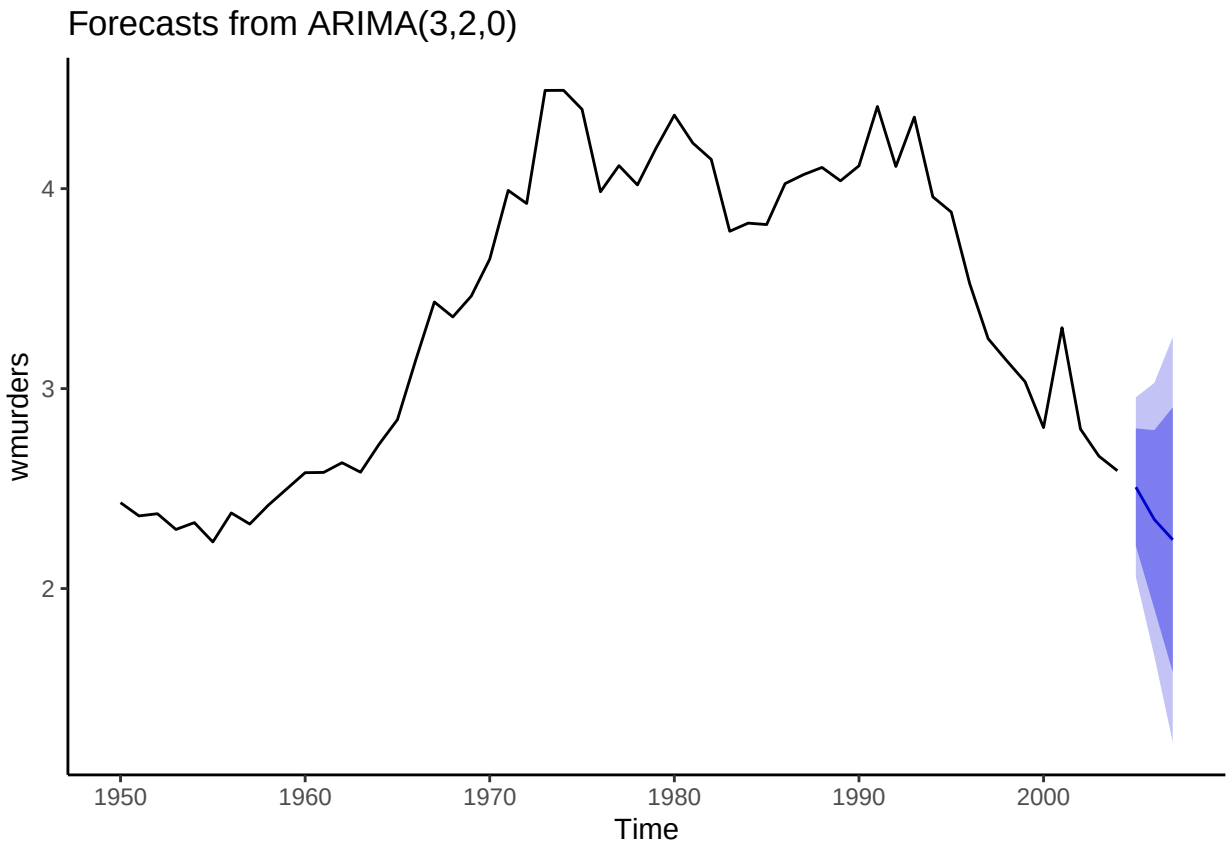

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,2,0)
## Q* = 19.964, df = 7, p-value = 0.005649
##
## Model df: 3.   Total lags used: 10
```

I have selected model3 with (p,d,q) value (2,2,0) based on the BIC value & p-value. For the model 3, all lags within ACF plot appear to be within boundaries, but the residuals could be more normal. Overall, I think it is satisfactory.

```
forecast(fit3,h=3)
```

```
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2005       2.507201 2.213401 2.801002 2.057873 2.956530
## 2006       2.345607 1.898976 2.792239 1.662543 3.028671
## 2007       2.244257 1.581865 2.906650 1.231215 3.257299
```

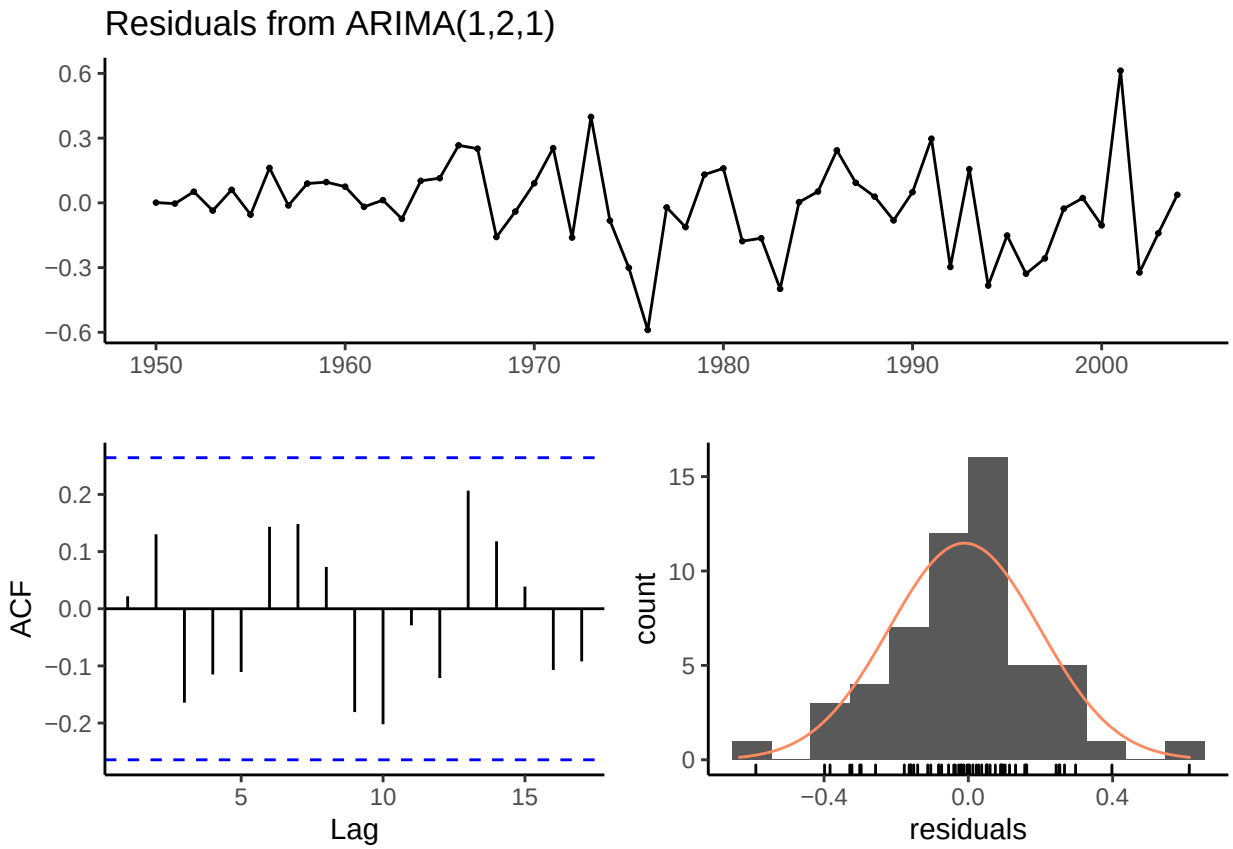
```
autoplot(forecast(fit3,h=3))
```




```
(fit4<-auto.arima(wmurders))
```

```
## Series: wmurders
## ARIMA(1,2,1)
##
## Coefficients:
##           ar1      ma1
##       -0.2434  -0.8261
## s.e.   0.1553   0.1143
##
## sigma^2 estimated as 0.04632:  log likelihood=6.44
## AIC=-6.88   AICc=-6.39   BIC=-0.97
```

```
checkresiduals(fit4)
```

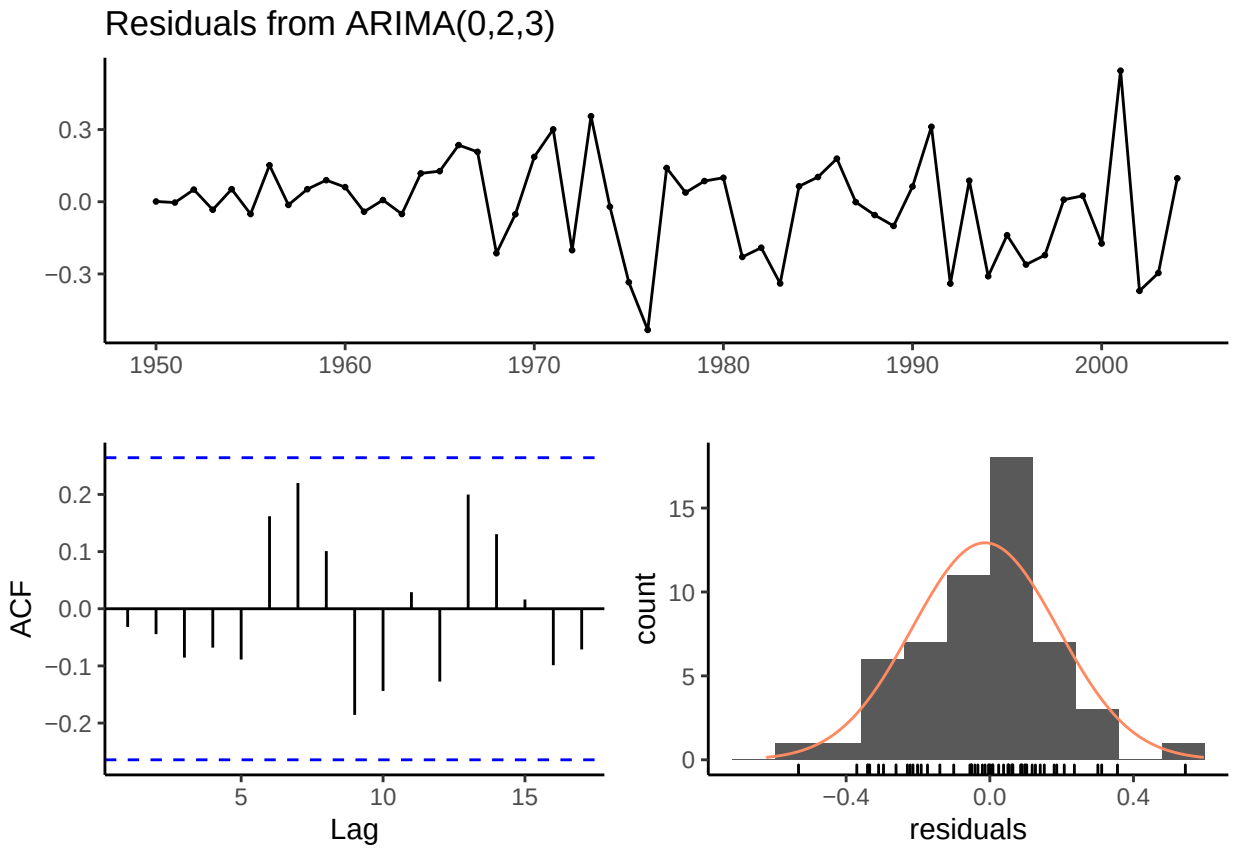



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,2,1)
## Q* = 12.419, df = 8, p-value = 0.1335
##
## Model df: 2.   Total lags used: 10
```

```
(fit5<-auto.arima(wmurders, stepwise=FALSE,
          approximation=FALSE, seasonal = FALSE))
```

```
## Series: wmurders
## ARIMA(0,2,3)
##
## Coefficients:
##           ma1     ma2      ma3
##       -1.0154  0.4324  -0.3217
## s.e.   0.1282  0.2278   0.1737
##
## sigma^2 estimated as 0.04475:  log likelihood=7.77
## AIC=-7.54   AICc=-6.7   BIC=0.35
```

```
checkresiduals(fit5)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,2,3)
## Q* = 10.706, df = 7, p-value = 0.152
##
## Model df: 3.   Total lags used: 10
```

auto.arima() didn't give the same model. Even though the BIC value of auto.arima() ARIMA(1,2,1) is lower and all the lags are within boundaries compared to the model I have choosen, the p-value is higher than the 5% significant level. Hence the model isn't significant. So I'll pick my model ARIMA(2,2,0) for forecasting.