TIME SERIES ANALYSIS OF MUTTON PRODUCED IN AUSTRALIA

Table of Contents

1 Introduction

Australia is one of the largest exporters of red meat in the world. However, over the past two decades, the meat production has significantly reduced mostly due to the persisting dry conditions particularly in the interior regions of Australia. This report analyses the monthly production of mutton in Australia since 1972 and forecasts the amount of mutton that would be produced in the near future using time series analysis techniques.

1.1 Data Description

The data is sourced from ABS(Australian Bureau of Statistics) and TABLE 10. Red Meat Produced - Mutton: All series (tonnes) is used in particular. It consists of monthly data from July 1972 to March 2020 and has 573 observations of multiple features of certain degree of abstraction. Column with all Australian mutton produced is selected for this task.

```
mutton.data <- read_excel("7218010.xls", sheet = "Data1", skip = 9)
mutton.data <- mutton.data %>% select("Series ID","A3484648J")
names(mutton.data)[1] <- "Year"
names(mutton.data)[2] <- "Tonnes"
mutton.data
```

```
## # A tibble: 573 x 2
##    Year                Tonnes
##    <dttm>               <dbl>
##  1 1972-07-01 00:00:00  39814
##  2 1972-08-01 00:00:00  37682
##  3 1972-09-01 00:00:00  33739
##  4 1972-10-01 00:00:00  41975
##  5 1972-11-01 00:00:00  48027
##  6 1972-12-01 00:00:00  44494
##  7 1973-01-01 00:00:00  48338
##  8 1973-02-01 00:00:00  43000
##  9 1973-03-01 00:00:00  46553
## 10 1973-04-01 00:00:00  24575
## # … with 563 more rows
```

```
mutton.ts.data <- as.ts(read.zoo(mutton.data, FUN = as.yearmon))
class(mutton.ts.data)
```

```
## [1] "ts"
```

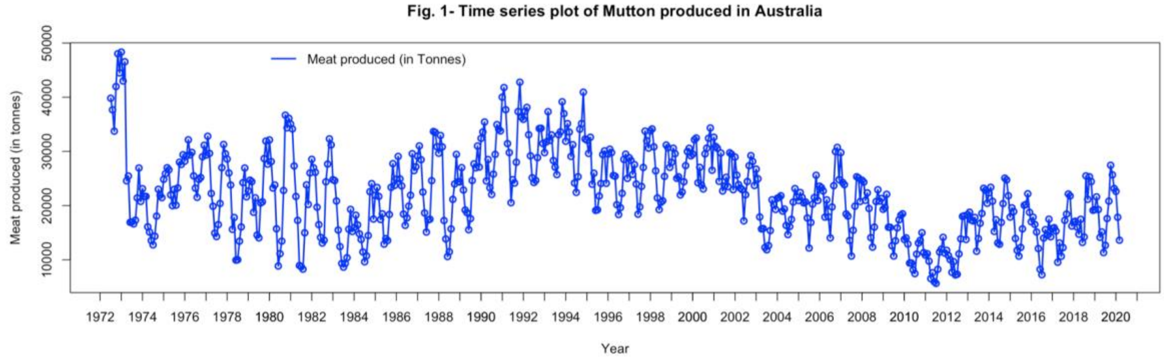
2. Data Exploration

2.1 Time series plot

```
#plot of time series data
plot(mutton.ts.data,
     ylab='Meat produced (in tonnes)',
     xlab='Year',
     type='o',
     col = c("blue"),
     lwd=2,
     main = "Fig. 1- Time series plot of Mutton produced in Australia")

axis(side=1, at=c(1972:2021))
legend("topright",lty=1, bty = "n" ,text.width = 40, col=c("blue"),lwd=2,
       c("Meat produced (in Tonnes)"))
```

Fig. 1- Time series plot of Mutton produced in Australia



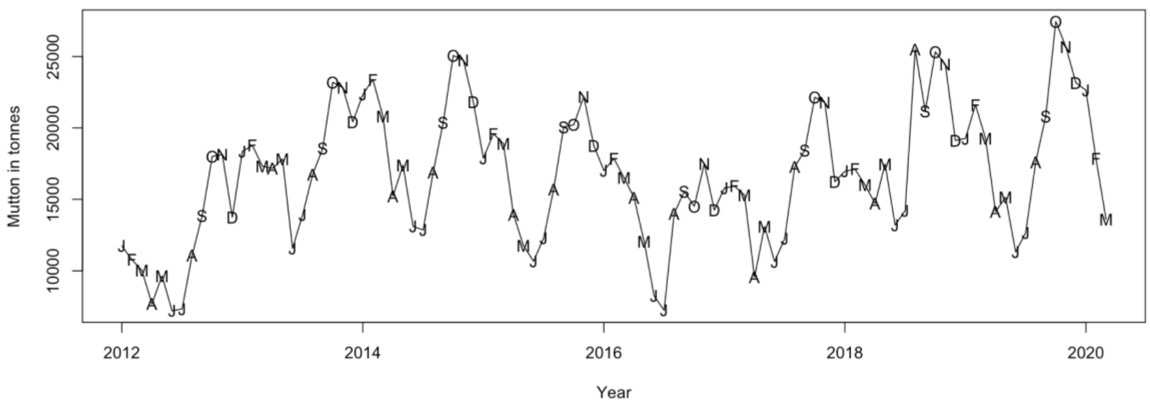
In the time series plot above (Fig. 1), below are the characteristics

- ***Trend**:* There is no trend pattern observed. However, sub- trends can be seen with a slight positive trend between 1984 and 1992, and slight negative trend between 1993 and 2010.
- ***Changing variance**:* It is hard to determine changing variance as there is seasonality associated.
- ***Seasonality**:* It can be seen clearly as there seems to be repeating pattern at regular intervals.
- ***Intervention**:* There is an intervention point around 1973 needs further investigation to look in at government records for further information.
- ***Behaviour:*** Seasonality makes it difficult to determine the Auto-regressive/Moving Average behaviour.

2.2 Time series plot with monthly label

```
plot(window(mutton.ts.data,start=c(2012,1)),ylab='Mutton in tonnes', main='Fig. 2- Time series plot pf mutton series with monthly symbols', xlab='Year')
Month=c('J','F','M','A','M','J','J','A','S','O','N','D')
points(window(mutton.ts.data,start=c(2012,1)),pch=Month)
```

Fig. 2- Time series plot pf mutton series with monthly symbols

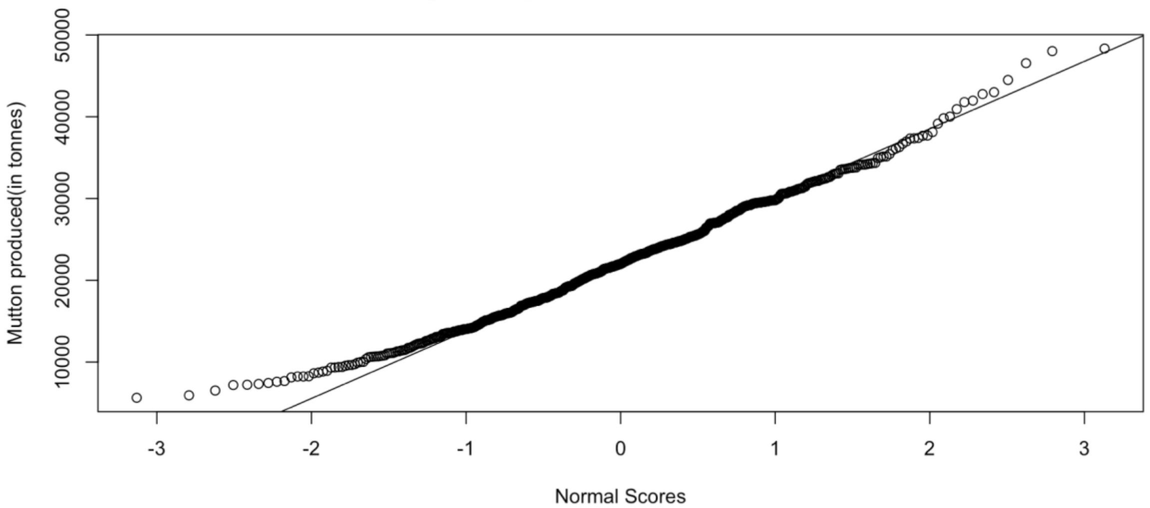


Above plot is the concentrated version from 2012 to 2020 of the complete series. In this plot, we can observe that October's usually have highest amount of mutton production and least being in June and July.

2.3 Normality of the series

```
##QQ plot for normality
qqnorm(mutton.ts.data, ylab="Mutton produced(in tonnes)", xlab="Normal Scores",main="Fig. 3- QQ plot for the time series data")
qqline(mutton.ts.data)
```

Fig. 3- QQ plot for the time series data



The series normally distributed with very minimal points tailing off towards both the ends.

2.4 ACF and PACF

```r
par(mfrow=c(1,2))
# acf: wave like pattern; non-stationary pattern
acf(mutton.ts.data, lag.max=36,main="Fig. 4- ACF plot for mutton meat produced")
pacf(mutton.ts.data, lag.max=36,main="Fig. 5- PACF plot for mutton meat produced")
```

Fig. 4- ACF plot for mutton meat produced

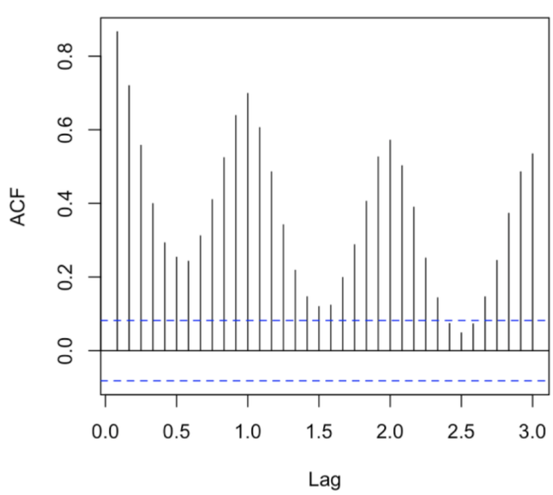
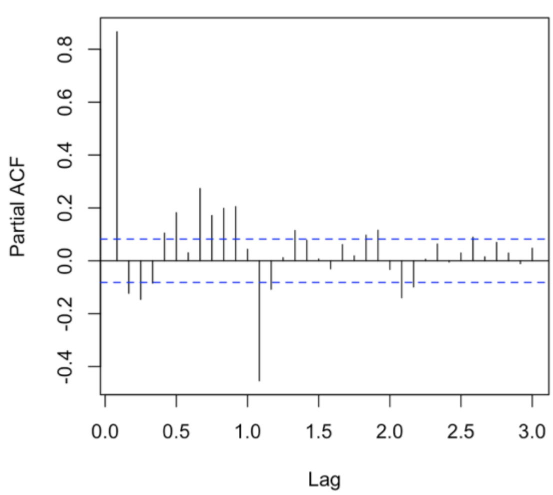


Fig. 5- PACF plot for mutton meat produced



- **ACF**(Fig. 4) - The plot shows a wave like pattern signifying the presence of seasonality in the series. Also, there is slowly decaying pattern observed at the seasonal lags implying seasonal trend as well.
- **PACF**(Fig. 5) – There are no apparent trend or pattern observed in the seasonal lags as well as the ordinary lags.

3. Model specification

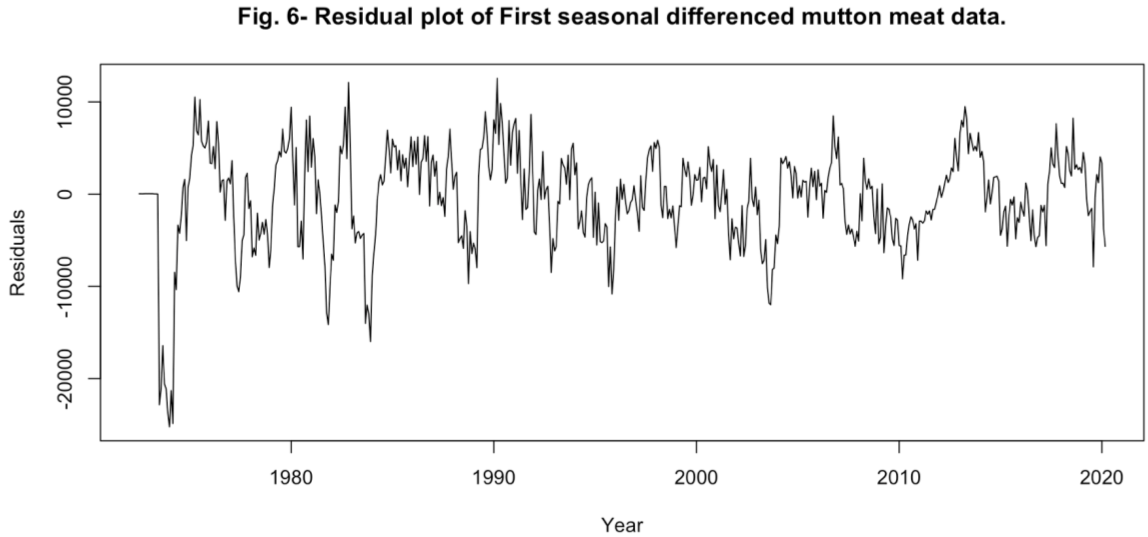
As we are dealing with seasonal models with seasonal trends, we need to exercise SARIMA modelling approach to obtain a feasible model by taking principle of parsimony into account while deciding the parameters.

3.1 Residual Approach

In this section, we try to get the orders for the model by considering the lags in ACF/PACF of the residuals. To achieve this, seasonal orders have to be taken care until there is White Noise observed in the residuals. This is followed by handling ordinary part of the lags with same process as the latter.

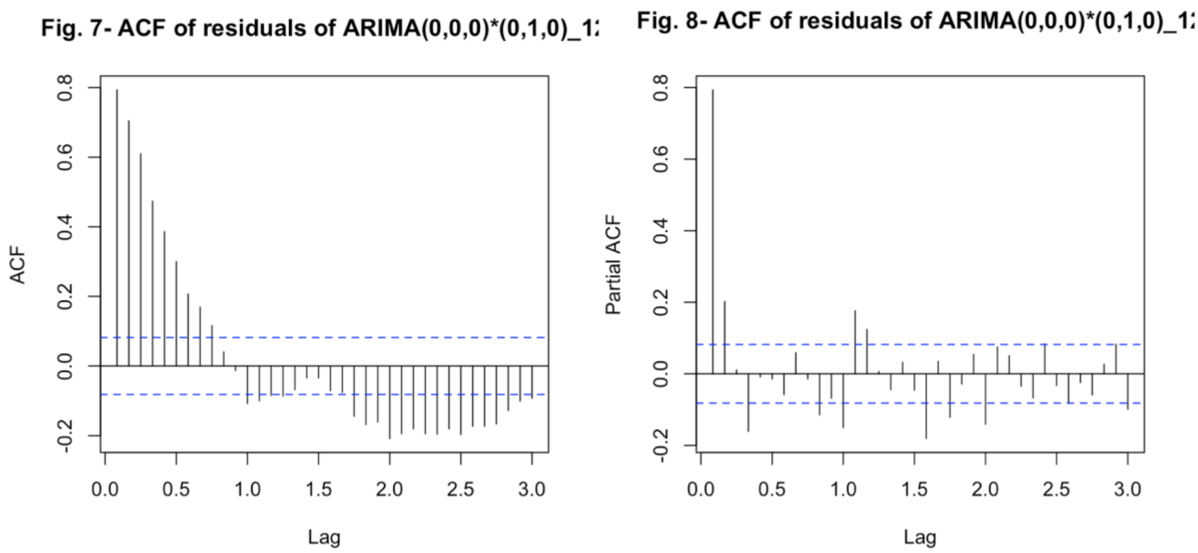
3.1.1 SARIMA(0,0,0)*(0,1,0)_12

```
m1.mutton.ts.data = arima(mutton.ts.data,order=c(0,0,0),seasonal=list(order=c(0,1,0), period=12))
res.m1 = residuals(m1.mutton.ts.data);
plot(res.m1,xlab='Year',ylab='Residuals',main="Fig. 6- Residual plot of First seasonal differenced mutton meat da
ta.")
```



In the above Fig. 6, the residuals of the first seasonal differenced data are normally distributed across the mean. However, the intervention can still be observed.

```
par(mfrow=c(1,2))
acf(res.m1, lag.max=36,main="Fig. 7- ACF of residuals of ARIMA(0,0,0)*(0,1,0)_12", cex=1)
pacf(res.m1, lag.max=36,main="Fig. 8- ACF of residuals of ARIMA(0,0,0)*(0,1,0)_12", cex=1)
```



ACF - In the above Fig. 7, seasonal lags at 1 and 2 have significance and it can be considered for the seasonal order.(Q=2) and slowly decaying pattern is observed for the ordinary part of the plot signifying trend

PACF - In the plot above(Fig. 8), it can be seen that there are no patterns observed and only two significant seasonal lags at 1 and 2.(P=2)

3.1.2 SARIMA(0,0,0)*(2,1,2)_12

```
m3.mutton.ts.data = arima(mutton.ts.data,order=c(0,1,0),seasonal=list(order=c(2,1,2), period=12))
res.m3 = residuals(m3.mutton.ts.data);
plot(res.m2,xlab='Year',ylab='Residuals',main="Fig. 12- Residual plot of the ARIMA(0,1,0)*(2,1,2)_12.")
```

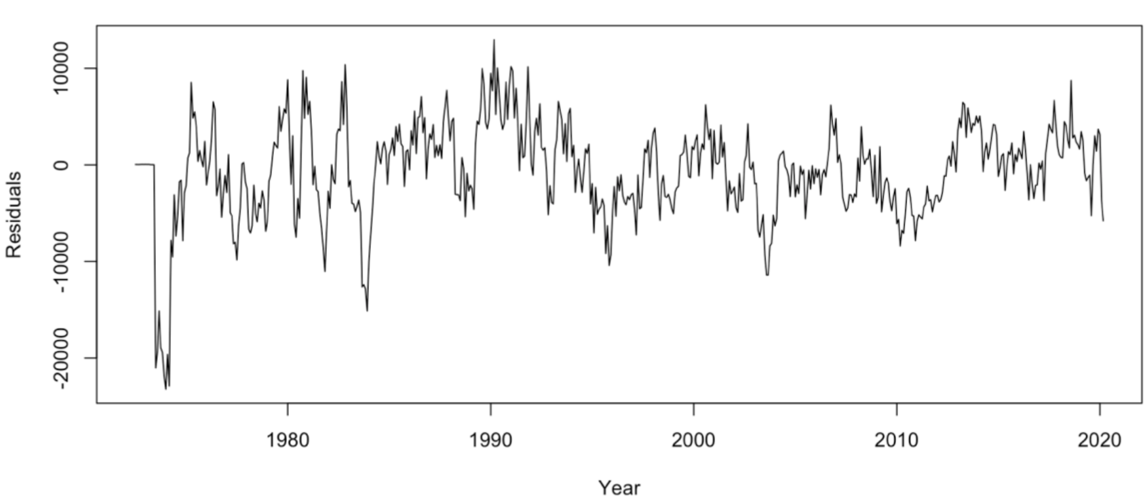


Fig. 12- Residual plot of the ARIMA(0,1,0)*(2,1,2)_12.

In the residual plot, we can observe the stationarity in the series, but intervention still remains.

```
par(mfrow=c(1,2))
acf(res.m2, lag.max=36,main="Fig. 10- ACF of residuals of ARIMA(0,0,0)*(2,1,2)_12.")
pacf(res.m2, lag.max=36,main="Fig. 11- PACF of residuals of ARIMA(0,0,0)*(2,1,2)_12.")
```

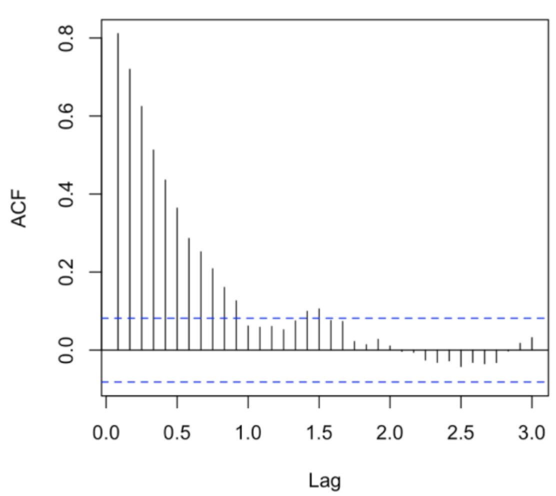


Fig. 10- ACF of residuals of ARIMA(0,0,0)*(2,1,2)_1

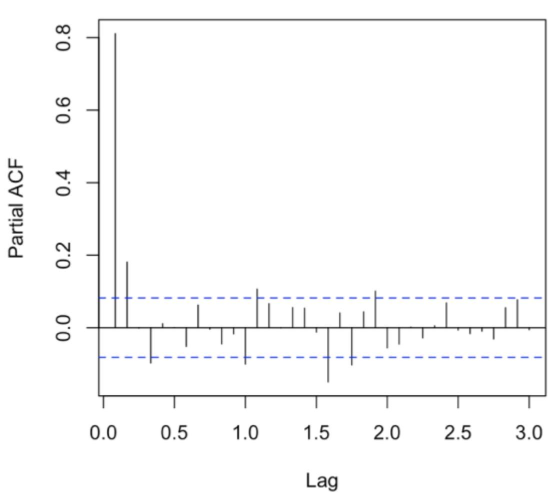


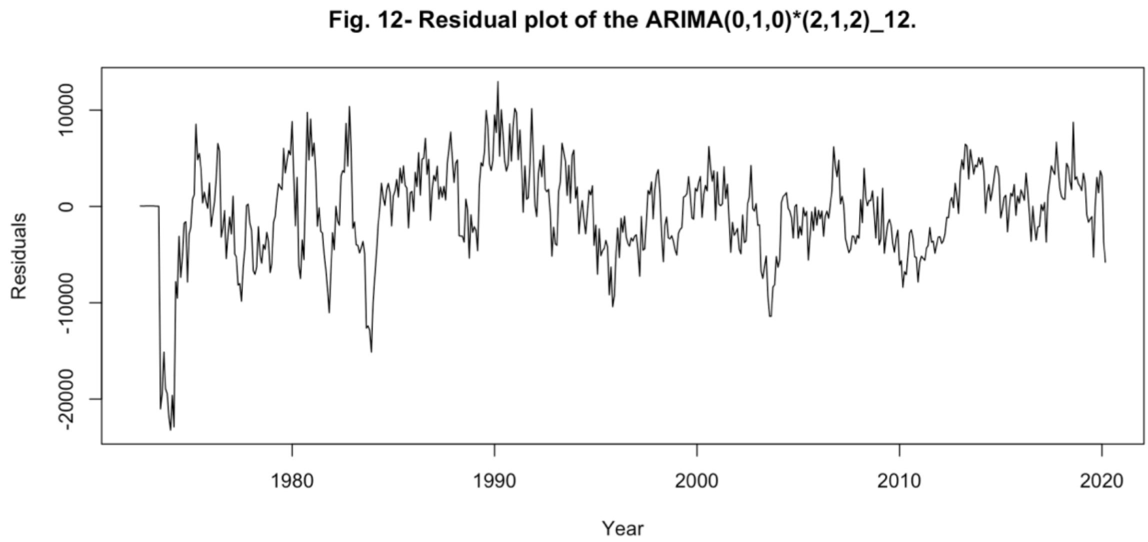
Fig. 11- PACF of residuals of ARIMA(0,0,0)*(2,1,2)_

ACF(Fig. 10) – Seasonal lags have been captured signifying white noise like behaviour at seasonal lags, however ordinary part still has slowly decaying pattern.

PACF(Fig. 11) – No seasonal lags are significant signifying white noise like behaviour at the seasonal lags

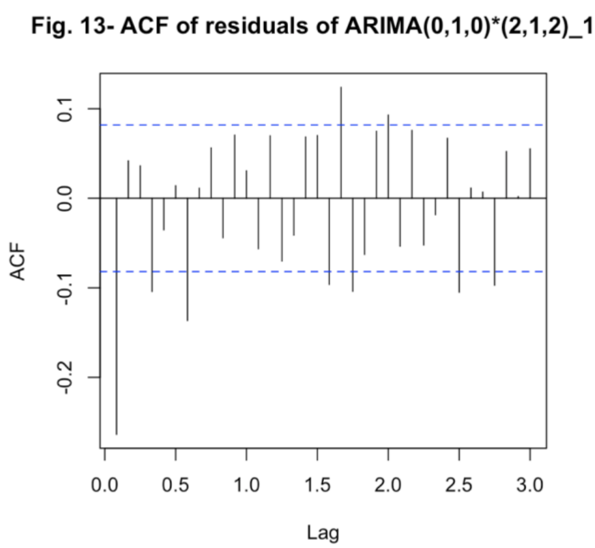
3.1.3 SARIMA(0,1,0)*(2,1,2)_12

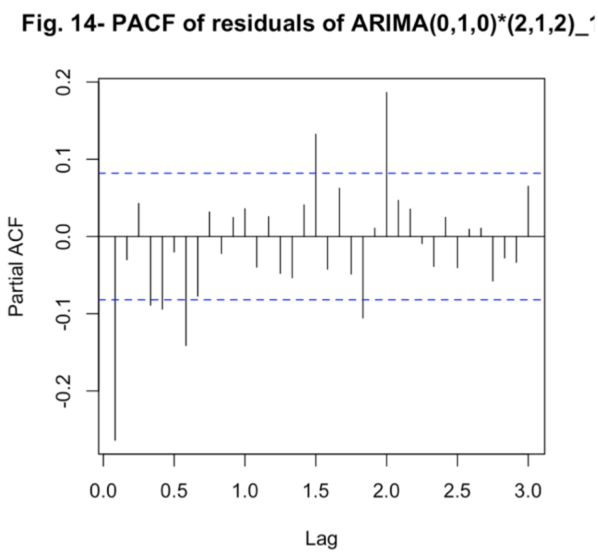
```
m3.mutton.ts.data = arima(mutton.ts.data,order=c(0,1,0),seasonal=list(order=c(2,1,2), period=12))
res.m3 = residuals(m3.mutton.ts.data);
plot(res.m2,xlab='Year',ylab='Residuals',main="Fig. 12- Residual plot of the ARIMA(0,1,0)*(2,1,2)_12.")
```



The series are fluctuating around the zero mean but the intervention still remains.

```
par(mfrow=c(1,2))
acf(res.m3, lag.max=36,main="Fig. 13- ACF of residuals of ARIMA(0,1,0)*(2,1,2)_12.")
pacf(res.m3, lag.max=36,main="Fig. 14- PACF of residuals of ARIMA(0,1,0)*(2,1,2)_12.")
```





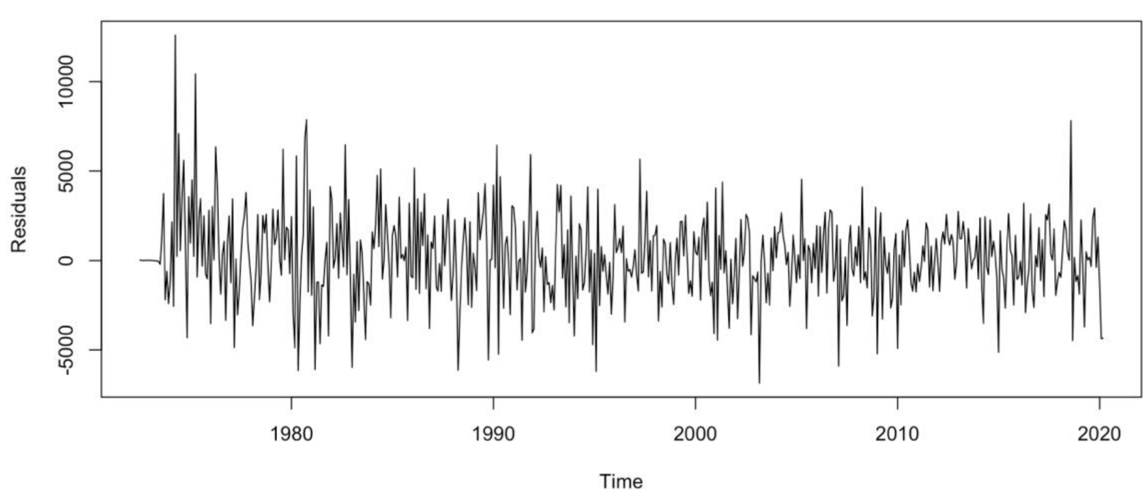
ACF(Fig. 13) – No significant pattern is observed at the ordinary part of the plot, three significant lags can be considered(q=3)

PACF(Fig. 14) – No pattern is observed to disregard the lags. Three lags can be considered from this PACF plot(p=3)

3.1.4 SARIMA(3,1,3)*(2,1,2)_12

```
m4.mutton.ts.data = arima(mutton.ts.data,order=c(3,1,3),seasonal=list(order=c(2,1,2), period=12))
res.m4 = residuals(m4.mutton.ts.data);
plot(res.m4,xlab='Time',ylab='Residuals',main="Fig. 15- Residual plot of the ARIMA(3,1,3)*(2,1,2)_12.")
```

Fig. 15- Residual plot of the ARIMA(3,1,3)*(2,1,2)_12.



In this residual plot, we can see that the intervention is captured by the model and series look stationary with constant variance.

```
# acf: wave like pattern
par(mfrow=c(1,2))
acf(res.m4, lag.max=36,ci.type='ma',main="Fig. 16- ACF of residuals of ARIMA(3,1,3)*(2,1,2)_12.")
pacf(res.m4, lag.max=36,ci.type='ma',main="Fig. 17- PACF of residuals of ARIMA(3,1,3)*(2,1,2)_12.")
```

Fig. 16- ACF of residuals of ARIMA(3,1,3)*(2,1,2)_1

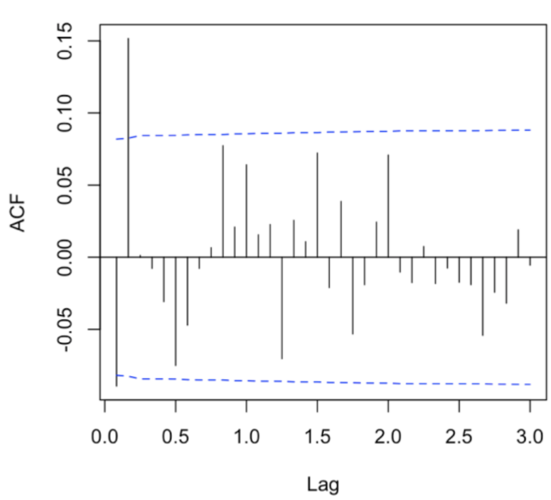
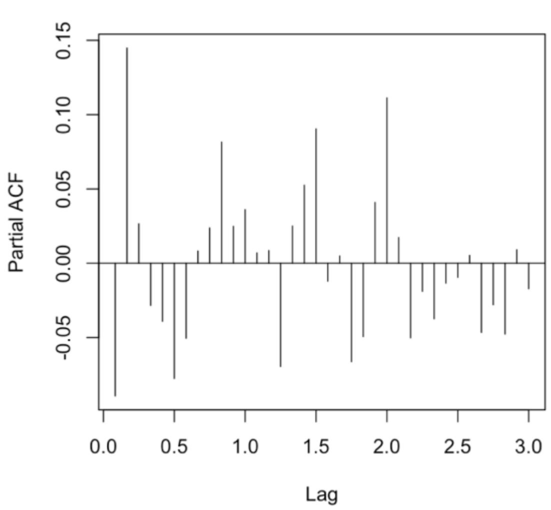


Fig. 17- PACF of residuals of ARIMA(3,1,3)*(2,1,2)_



ACF(Fig. 16) – There are no seasonal lags signifying white noise for the seasonal part. For the ordinary part, first lag is not so significant, however the second lag is significant even after specifying the order for it as the reason could be the presence of intervention in the series. This can be modelled using other techniques, but it is not in the scope of this report. Here, we assume white noise for the ACF plot.

PACF(Fig. 17) – There are no significant co-relations at various lags signifying white noise like behaviour.

3.1.5 Residual approach summary

- Seasonal Difference was considered first for the residual approach.(D=1)
- Orders of P=2,Q=2 was observed in seasonal co-relations.
- Difference for the ordinary part was considered(d=1).
- Orders of p=3,q=3 was observed for the ordinary co-relations.

3.2 EACF

```
eacf(diff(diff(mutton.ts.data), lag=12))
```

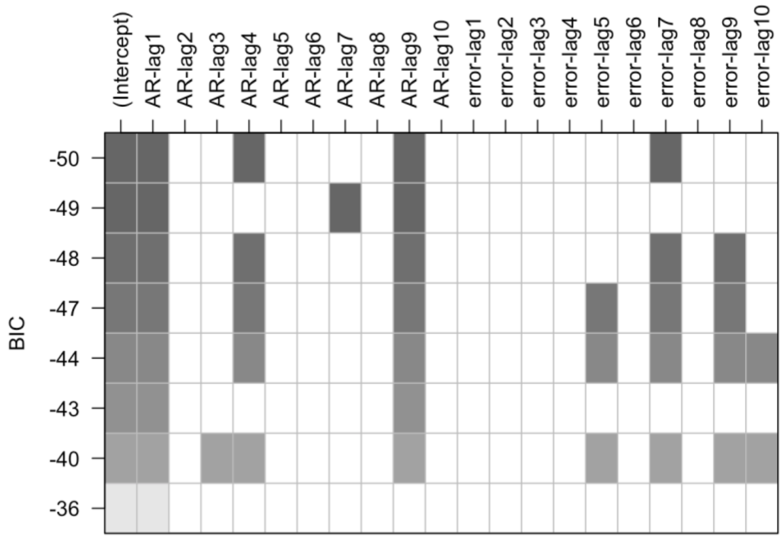
```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o x x o o x o x o x  x  o  o
## 1 x x o x x o x o x o o  x  o  o
## 2 x o x x o o x o x o o  x  x  o
## 3 x x x o o o o o o o o  x  x  x
## 4 x x x o o o o o o o o  x  x  x
## 5 x x x x x x o o o o o  x  x  o
## 6 x x x x x o o o o o o  x  o  o
## 7 x x x x o x x o o o o  x  o  x
```

Next, we plot the EACF table of the seasonal and ordinary differenced mutton time series data as these differences make the series stationary. From the above table, we can see that the vertex can be found at the intersection of (3,3). The possible set of candidate models, we can use based on the EACF table are {ARIMA(3,1,3), ARIMA(3,1,4), ARIMA(4,1,5)} *(2,1,2)_12.

3.3 BIC Table

Next, we plot the BIC table to get some more candidate models.

```
res = armasubsets(y=diff(diff(mutton.ts.data), lag=12),nar=10,nma=10,y.name='AR',ar.method='ols')
plot(res)
```



From the above BIC table, the shaded regions correspond to AR(1), AR(4), AR(9), MA(7) coefficients, so the set of possible models, we obtain are {ARIMA(1,1,7),ARIMA(4,1,7), ARIMA(9,1,7)} *(2,1,2)_12.

4. Model Fitting

From the above analysis, we have concluded that the possible set of models are {ARIMA(3,1,3), ARIMA(3,1,4), ARIMA(4,1,5), ARIMA(4,1,7), ARIMA(9,1,7)} *(2,1,2)_12. Next, we apply fitting methods CSS and ML.

CSS - Minimises the sum of squared residuals.

ML - Maximises the log-likelihood function of the model.

4.1 SARIMA(3,1,3)*(2,1,2)_12

```
model_313_css = arima(mutton.ts.data,order=c(3,1,3),seasonal=list(order=c(2,1,2), period=12), method = 'CSS')
coeftest(model_313_css)
```

```
##
## z test of coefficients:
##
##         Estimate  Std. Error   z value  Pr(>|z|)
## ar1  -0.28897472  0.00076580 -377.3506 < 2.2e-16 ***
## ar2   0.01141498  0.00087143   13.0991 < 2.2e-16 ***
## ar3   0.86992735  0.00054619 1592.7233 < 2.2e-16 ***
## ma1   0.14987734  0.00198294   75.5834 < 2.2e-16 ***
## ma2  -0.16600597  0.00199041  -83.4029 < 2.2e-16 ***
## ma3  -1.02552971  0.00141671 -723.8793 < 2.2e-16 ***
## sar1 -0.57316424  0.01020893  -56.1434 < 2.2e-16 ***
## sar2 -0.13733289  0.01015072  -13.5294 < 2.2e-16 ***
## sma1 -0.17058080  0.03969292   -4.2975 1.727e-05 ***
## sma2 -0.48826338  0.04292527  -11.3747 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above model fitting using the CSS method, we can conclude that all the coefficients are significant.

```
model_313_ml = arima(mutton.ts.data,order=c(3,1,3),seasonal=list(order=c(2,1,2), period=12), method = 'ML')
coeftest(model_313_ml)
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error z value  Pr(>|z|)
## ar1  -0.564165   0.455890 -1.2375   0.21590
## ar2  -0.312538   0.528163 -0.5917   0.55402
## ar3   0.593219   0.457570  1.2965   0.19482
## ma1   0.388443   0.375549  1.0343   0.30098
## ma2   0.121943   0.441673  0.2761   0.78248
## ma3  -0.753766   0.409756 -1.8396   0.06583 .
## sar1 -0.656257   0.143903 -4.5604 5.105e-06 ***
## sar2  0.072069   0.070955  1.0157   0.30977
## sma1 -0.047682   0.138909 -0.3433   0.73140
## sma2 -0.737579   0.122186 -6.0365 1.574e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

While using the ML method only two coefficients are significant. This result conflicts with that of the CSS method. Hence, this model cannot be used for further diagnostics.

4.2 SARIMA(3,1,4)*(2,1,2)_12

```
model_314_css = arima(mutton.ts.data,order=c(3,1,4),seasonal=list(order=c(2,1,2), period=12), method = 'CSS')
coeftest(model_314_css)
```

```
##
## z test of coefficients:
##
##         Estimate Std. Error   z value Pr(>|z|)
## ar1  -0.2533846  0.0010767 -235.3402  < 2e-16 ***
## ar2   0.0521134  0.0011813   44.1163  < 2e-16 ***
## ar3   0.9038950  0.0005637 1603.5115  < 2e-16 ***
## ma1  -0.0153454         NA        NA       NA
## ma2  -0.1952619         NA        NA       NA
## ma3  -0.9951434         NA        NA       NA
## ma4   0.1749644         NA        NA       NA
## sar1 -0.6284800  0.0028005 -224.4176  < 2e-16 ***
## sar2 -0.1310157  0.0059685  -21.9512  < 2e-16 ***
## sma1 -0.1186715  0.0360335   -3.2934  0.00099 ***
## sma2 -0.5076209  0.0419407  -12.1033  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Due to the presence of NA values in the standard error, z value and p value, we decide not to consider this model for further analysis.

```r
model_314_ml = arima(mutton.ts.data,order=c(3,1,4),seasonal=list(order=c(2,1,2), period=12), method = 'ML')
coeftest(model_314_ml)
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error  z value  Pr(>|z|)
## ar1  -0.1636832         NA        NA        NA
## ar2   0.1509835         NA        NA        NA
## ar3   0.9921657         NA        NA        NA
## ma1  -0.0675865  0.0403146   -1.6765   0.09364 .
## ma2  -0.1853969  0.0120638  -15.3681 < 2.2e-16 ***
## ma3  -0.9548741  0.0067009 -142.4988 < 2.2e-16 ***
## ma4   0.2079744  0.0391690    5.3097 1.098e-07 ***
## sar1 -0.6623695  0.1542209   -4.2949 1.747e-05 ***
## sar2  0.0845087  0.0740713    1.1409   0.25391
## sma1 -0.0297953  0.1463207   -0.2036   0.83864
## sma2 -0.7411521  0.1351279   -5.4848 4.139e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Presence of NA values in the z test is also found while using the ML method clearly showing that this model is not fit for modelling.

4.3 SARIMA(4,1,5)*(2,1,2)_12

```r
model_415_css = arima(mutton.ts.data,order=c(4,1,5),seasonal=list(order=c(2,1,2), period=12), method = 'CSS')
coeftest(model_415_css)
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error  z value  Pr(>|z|)
## ar1  -0.0895288  0.0140127  -6.3891 1.668e-10 ***
## ar2   0.0831837  0.0078908  10.5418 < 2.2e-16 ***
## ar3   0.8812197  0.0057831 152.3792 < 2.2e-16 ***
## ar4  -0.1602305  0.0107165 -14.9518 < 2.2e-16 ***
## ma1  -0.1572549         NA       NA        NA
## ma2  -0.0805709         NA       NA        NA
## ma3  -0.9395155  0.0146234 -64.2474 < 2.2e-16 ***
## ma4   0.2907076         NA       NA        NA
## ma5  -0.1425708         NA       NA        NA
## sar1 -0.5355559  0.0207191 -25.8484 < 2.2e-16 ***
## sar2 -0.0608029  0.0070186  -8.6631 < 2.2e-16 ***
## sma1 -0.1699245  0.0517131  -3.2859  0.001017 **
## sma2 -0.5245410  0.0414800 -12.6456 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It can be seen that NA values are present in the z test, therefore we decide to disregard this model for further diagnostics.

```
model_415_ml = arima(mutton.ts.data,order=c(4,1,5),seasonal=list(order=c(2,1,2), period=12), method = 'ML')
coeftest(model_415_ml)
```

```
##
## z test of coefficients:
##
##         Estimate Std. Error  z value  Pr(>|z|)
## ar1   0.7557499  0.0415086  18.2071 < 2.2e-16 ***
## ar2   0.2956255  0.0099159  29.8132 < 2.2e-16 ***
## ar3   0.8474756  0.0055750 152.0125 < 2.2e-16 ***
## ar4  -0.9205010  0.0397892 -23.1344 < 2.2e-16 ***
## ma1  -0.9992108  0.0563640 -17.7278 < 2.2e-16 ***
## ma2  -0.1375567  0.0367600  -3.7420 0.0001825 ***
## ma3  -0.7875740  0.0206792 -38.0853 < 2.2e-16 ***
## ma4   1.1215725  0.0512262  21.8945 < 2.2e-16 ***
## ma5  -0.1909337  0.0435179  -4.3875 1.147e-05 ***
## sar1 -0.7179688  0.1186385  -6.0517 1.433e-09 ***
## sar2  0.1235530  0.0752998   1.6408 0.1008362
## sma1  0.0338875  0.1050975   0.3224 0.7471207
## sma2 -0.8056019  0.1007959  -7.9924 1.323e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that the coefficients of the seasonal AR2 and MA1 are insignificant due their p values being greater than 0.05. Hence, we cannot consider this model for modelling.

4.4 SARIMA(4,1,7)*(2,1,2)_12

```
model_417_css = arima(mutton.ts.data,order=c(4,1,7),seasonal=list(order=c(2,1,2), period=12), method = 'CSS')
coeftest(model_417_css)
```

```
##
## z test of coefficients:
##
##         Estimate Std. Error  z value  Pr(>|z|)
## ar1   0.2233252  0.0247051   9.0396 < 2.2e-16 ***
## ar2  -0.1513179  0.0094692 -15.9800 < 2.2e-16 ***
## ar3   0.4995914  0.0091317  54.7094 < 2.2e-16 ***
## ar4  -0.7533103  0.0241349 -31.2125 < 2.2e-16 ***
## ma1  -0.4840573  0.0517972  -9.3452 < 2.2e-16 ***
## ma2   0.2307699  0.0526454   4.3835 1.168e-05 ***
## ma3  -0.6418659  0.0531379 -12.0792 < 2.2e-16 ***
## ma4   0.8861134  0.0454768  19.4850 < 2.2e-16 ***
## ma5  -0.3301621  0.0517599  -6.3787 1.786e-10 ***
## ma6  -0.0324326  0.0579899  -0.5593 0.5759704
## ma7  -0.1114392  0.0441764  -2.5226 0.0116492 *
## sar1 -0.3014680  0.1584217  -1.9029 0.0570475 .
## sar2 -0.1283820  0.0367469  -3.4937 0.0004764 ***
## sma1 -0.4735226  0.1555802  -3.0436 0.0023377 **
## sma2 -0.2791434  0.1416127  -1.9712 0.0487039 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above z test, we can see that all the coefficients of the model are significant except for MA6.

```
model_417_ml = arima(mutton.ts.data,order=c(4,1,7),seasonal=list(order=c(2,1,2), period=12), method = 'ML')
coeftest(model_417_ml)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error  z value  Pr(>|z|)
## ar1   0.122187   0.024875   4.9120 9.015e-07 ***
## ar2  -0.423066   0.027056 -15.6367 < 2.2e-16 ***
## ar3   0.228151   0.027939   8.1662 3.184e-16 ***
## ar4  -0.903889   0.028900 -31.2765 < 2.2e-16 ***
## ma1  -0.405856   0.047729  -8.5034 < 2.2e-16 ***
## ma2   0.548124   0.055262   9.9187 < 2.2e-16 ***
## ma3  -0.404288   0.056779  -7.1204 1.076e-12 ***
## ma4   1.033331   0.035574  29.0471 < 2.2e-16 ***
## ma5  -0.332640   0.056073  -5.9323 2.988e-09 ***
## ma6   0.035306   0.054819   0.6441 0.5195430
## ma7  -0.141985   0.049833  -2.8492 0.0043830 **
## sar1 -0.589901   0.165214  -3.5705 0.0003563 ***
## sar2  0.035821   0.072302   0.4954 0.6202951
## sma1 -0.117555   0.159397  -0.7375 0.4608193
## sma2 -0.672559   0.146862  -4.5795 4.660e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The coefficients from ML method prove significance for most of the parameters except for a few. This model can be considered for further diagnostics.

5. Model Diagnostics

SARIMA(4,1,7)*(2,1,2)_12 proved most coefficient significance among the other models. Next, we carry out model overfitting by increasing the model parameters for both AR and MA of the model.

5.1 Model Overfitting

Next, we will try overfitting the SARIMA(4,1,7)*(2,1,2)_12 model. This is done by checking whether the SARIMA(4,1,8) and SARIMA(5,1,7) models are significant or not. First, we conduct the z test of coefficients for the model SARIMA(4,1,8) using the CSS method.

```
model_418_css = arima(mutton.ts.data,order=c(4,1,8),seasonal=list(order=c(2,1,2), period=12), method = 'CSS')
coeftest(model_418_css)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error  z value  Pr(>|z|)
## ar1   0.137132   0.022251   6.1628 7.146e-10 ***
## ar2  -0.395006   0.017754 -22.2485 < 2.2e-16 ***
## ar3   0.245770   0.017643  13.9300 < 2.2e-16 ***
## ar4  -0.893968   0.021671 -41.2513 < 2.2e-16 ***
## ma1  -0.419641   0.046301  -9.0634 < 2.2e-16 ***
## ma2   0.476666   0.047860   9.9595 < 2.2e-16 ***
## ma3  -0.406847   0.050794  -8.0097 1.150e-15 ***
## ma4   0.905440   0.054997  16.4635 < 2.2e-16 ***
## ma5  -0.333853   0.054591  -6.1156 9.622e-10 ***
## ma6  -0.102498   0.052317  -1.9592  0.050093 .
## ma7  -0.119441   0.049244  -2.4255  0.015288 *
## ma8  -0.141964   0.048554  -2.9238  0.003457 **
## sar1 -0.757654   0.071508 -10.5954 < 2.2e-16 ***
## sar2 -0.109036   0.042935  -2.5396  0.011099 *
## sma1 -0.033879   0.071830  -0.4717  0.637170
## sma2 -0.641810   0.069445  -9.2419 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Looking at the above z test, we can see that all the coefficients except for ma6 and sma1 are significant as their p values are less than 0.05. Next, we conduct the same test using the ML method.

```
model_418_ml = arima(mutton.ts.data,order=c(4,1,8),seasonal=list(order=c(2,1,2), period=12), method = 'ML')
coeftest(model_418_ml)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error  z value  Pr(>|z|)
## ar1   0.106179   0.029200   3.6362 0.0002767 ***
## ar2  -0.436896   0.026543 -16.4601 < 2.2e-16 ***
## ar3   0.214227   0.027251   7.8612 3.804e-15 ***
## ar4  -0.900936   0.032720 -27.5344 < 2.2e-16 ***
## ma1  -0.388077   0.051490  -7.5370 4.810e-14 ***
## ma2   0.555074   0.051709  10.7345 < 2.2e-16 ***
## ma3  -0.377824   0.055648  -6.7896 1.125e-11 ***
## ma4   0.964332   0.061319  15.7265 < 2.2e-16 ***
## ma5  -0.307325   0.057096  -5.3826 7.343e-08 ***
## ma6   0.010862   0.054990   0.1975 0.8434184
## ma7  -0.123529   0.049971  -2.4720 0.0134353 *
## ma8  -0.073607   0.047251  -1.5578 0.1192814
## sar1 -0.588505   0.182280  -3.2286 0.0012441 **
## sar2  0.020998   0.073371   0.2862 0.7747328
## sma1 -0.131686   0.178904  -0.7361 0.4616859
## sma2 -0.653968   0.163558  -3.9984 6.378e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above test, we can see that ma6, ma8, sar2 and sma1 are insignificant since their p values are above the significance level. From the tests, we conclude there is no improvement in the SARIMA(4,1,8) compared to the selected model SARIMA (4,1,7). Therefore, we decide not to include the model in the analysis. Next, we conduct the z test for SARIMA(5,1,7) model using the CSS method.

```
model_517_css = arima(mutton.ts.data,order=c(5,1,7),seasonal=list(order=c(2,1,2), period=12), method = 'CSS')
coeftest(model_517_css)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error z value Pr(>|z|)
## ar1   0.212445   0.865975  0.2453 0.806205
## ar2  -0.135218   0.634270 -0.2132 0.831181
## ar3   0.562286   0.204739  2.7464 0.006026 **
## ar4  -0.681003   0.649206 -1.0490 0.294188
## ar5   0.037723   0.827320  0.0456 0.963632
## ma1  -0.455202   1.346063 -0.3382 0.735233
## ma2   0.198995   0.799734  0.2488 0.803495
## ma3  -0.693189   0.372635 -1.8602 0.062852 .
## ma4   0.808751   0.766509  1.0551 0.291375
## ma5  -0.326295   1.320016 -0.2472 0.804761
## ma6  -0.016660   0.285039 -0.0584 0.953393
## ma7  -0.087895   0.099784 -0.8808 0.378399
## sar1 -0.188473   0.541293 -0.3482 0.727696
## sar2 -0.139634   0.247569 -0.5640 0.572742
## sma1 -0.564005   0.574945 -0.9810 0.326607
## sma2 -0.192926   0.365894 -0.5273 0.598004
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Almost all the coefficients are shown to be insignificant meaning that this model is not a good model for this series. Next, we conduct z test using the ML method.

```r
model_517_ml = arima(mutton.ts.data,order=c(5,1,7),seasonal=list(order=c(2,1,2), period=12), method = 'ML')
coeftest(model_517_ml)
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error z value  Pr(>|z|)
## ar1   0.0344898         NA      NA        NA
## ar2  -0.3974731         NA      NA        NA
## ar3   0.1802756         NA      NA        NA
## ar4  -0.8739404         NA      NA        NA
## ar5  -0.1023877         NA      NA        NA
## ma1  -0.3077536         NA      NA        NA
## ma2   0.4665241         NA      NA        NA
## ma3  -0.2994306         NA      NA        NA
## ma4   0.9409512         NA      NA        NA
## ma5  -0.1724921         NA      NA        NA
## ma6  -0.0500797  0.0024572 -20.381 < 2.2e-16 ***
## ma7  -0.1141833         NA      NA        NA
## sar1 -0.5144944         NA      NA        NA
## sar2  0.0625762         NA      NA        NA
## sma1 -0.1554357         NA      NA        NA
## sma2 -0.6487390         NA      NA        NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Almost all the coefficients are found to contain NA values, thereby confirming that this model is a bad model for this series.

```r
#ARIMA(3,1,3),ARIMA(3,1,4),ARIMA(4,1,4),ARIMA(4,1,7),ARIMA(9,1,7)
sort.score(AIC(model_313_ml,model_314_ml,model_415_ml,model_417_ml,model_917_ml), score = "aic")
```

```
##              df      AIC
## model_417_ml 16 10354.19
## model_917_ml 21 10357.62
## model_415_ml 14 10372.89
## model_314_ml 12 10376.89
## model_313_ml 11 10380.07
```

```r
sort.score(BIC(model_313_ml,model_314_ml,model_415_ml,model_417_ml,model_917_ml), score = "bic" )
```

```
##              df      BIC
## model_417_ml 16 10423.44
## model_313_ml 11 10427.67
## model_314_ml 12 10428.83
## model_415_ml 14 10433.48
## model_917_ml 21 10448.51
```

By overfitting, the SARIMA(4,1,7) model, we conclude that the model SARIMA(4,1,8) has no improvement over SARIMA(4,1,7) while the model SARIMA(5,1,7) is a bad fit for this series.

6. Residual Analysis

Here we will do the residual analysis of our predictive model which is SARIMA(4,1,7). This residual analysis includes time series plot of the standardised residuals, histogram of the standardised residuals, QQ Plot of the standardised residuals, ACF of the standardised residuals, Shapiro-Wilk test of the standardised residuals and the Ljung-Box test.

Fig. 18- Time series plot of standardised residuals

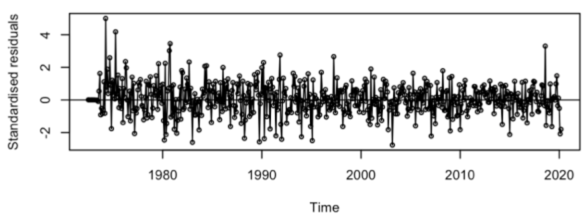


Fig. 19- Histogram of standardised residuals

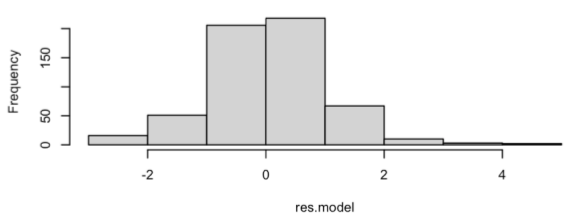


Fig. 20- ACF of standardised residuals

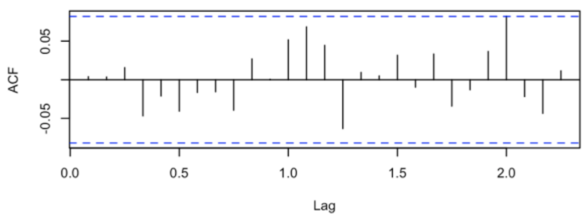


Fig. 21- PACF of standardised residuals

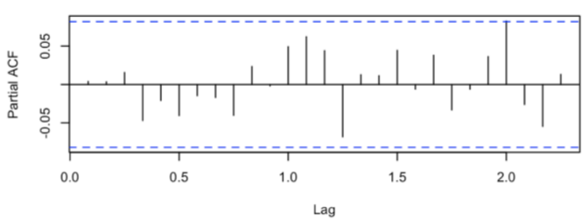
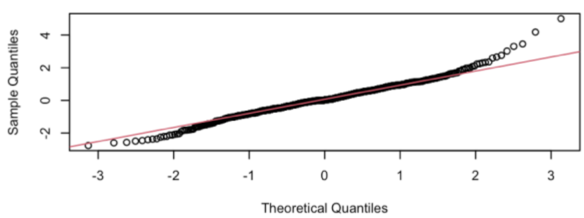
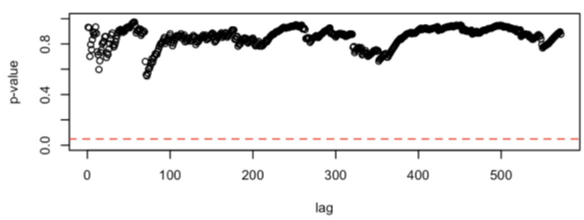


Fig. 22- QQ plot of standardised residuals



Ljung-Box Test



Results of Residual Analysis

- **Time Series Plot**(Fig. 18) - No trend to be found. It seems the residuals are randomly distributed.
- **Histogram**(Fig. 19) - We can see that the residuals appear to be normally distributed in the graph.
- **ACF & PACF**(Fig. 20) - Both do not show any significant lag. The patterns in ACF and PACF implies existence of white noise behaviour. So, residuals are uncorrelated.
- **QQ-plot**(Fig. 21) - The data points are close to the normality line for most part with exception of some points in the extreme ends.
- **Ljung-Box Test** - We can see that all the p values are above the significance level indicating the absence of co-relations in the residuals.

7. Forecasting

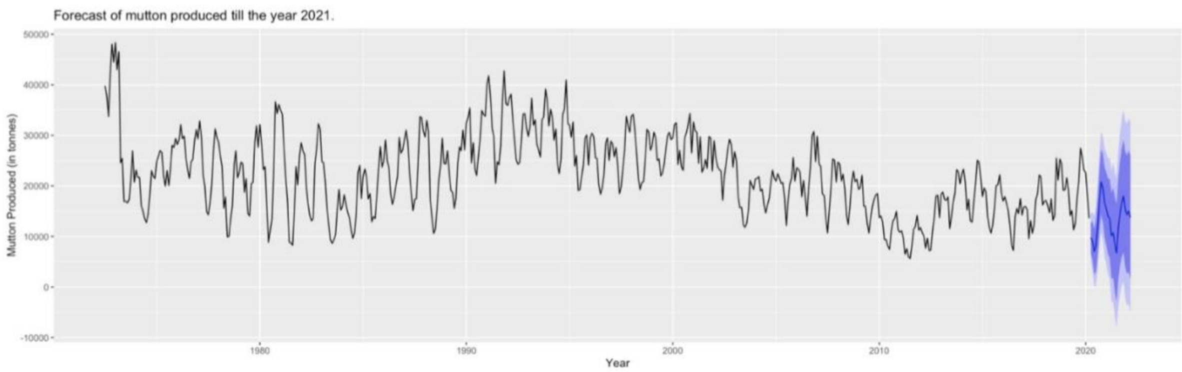
```r
fit = Arima(mutton.ts.data,order=c(4,1,7),seasonal=list(order=c(2,1,2), period=12))
Forecast = forecast(fit,h=24)
Forecast
```

```
##          Point Forecast     Lo 80    Hi 80        Lo 95    Hi 95
## Apr 2020       9775.191  6681.593 12868.79   5043.94016 14506.44
## May 2020       9008.498  5199.720 12817.28   3183.47455 14833.52
## Jun 2020       7019.843  2459.394 11580.29     45.23813 13994.45
## Jul 2020       8084.017  2963.529 13204.50    252.90650 15915.13
## Aug 2020      12167.690  6498.766 17836.61   3497.81837 20837.56
## Sep 2020      17127.469 11001.067 23253.87   7757.94613 26496.99
## Oct 2020      20788.833 14321.347 27256.32  10897.66674 30680.00
## Nov 2020      19545.639 12828.877 26262.40   9273.23778 29818.04
## Dec 2020      16667.170  9721.357 23612.98   6044.46575 27289.87
## Jan 2021      15513.757  8309.354 22718.16   4495.57289 26531.94
## Feb 2021      13970.067  6462.533 21477.60   2488.28390 25451.85
## Mar 2021      13557.193  5706.109 21408.28   1549.99641 25564.39
## Apr 2021      10142.855  1658.103 18627.61  -2833.45264 23119.16
## May 2021      10695.215  1759.793 19630.64  -2970.33333 24360.76
## Jun 2021       8306.301 -1036.282 17648.88  -5981.94663 22594.55
## Jul 2021       6718.400 -2968.617 16405.42  -8096.61380 21533.41
## Aug 2021      11772.085  1743.025 21801.15  -3566.03956 27110.21
## Sep 2021      14650.157  4258.916 25041.40  -1241.87449 30542.19
## Oct 2021      16768.218  6012.513 27523.92    318.78711 33217.65
## Nov 2021      17963.892  6861.099 29066.68    983.63565 34944.15
## Dec 2021      15325.344  3893.310 26757.38  -2158.44283 32809.13
## Jan 2022      14300.120  2568.854 26031.39  -3641.30353 32241.54
## Feb 2022      15026.970  3015.557 27038.38  -3342.90131 33396.84
## Mar 2022      13760.172  1465.576 26054.77  -5042.79042 32563.13
```

```r
par(mfrow=c(1,1))

#Forecast
autoplot(Forecast,

     ylab='Mutton Produced (in tonnes)',
     xlab='Year',
     type='o',
     col = c("blue"),
     lwd=2,
     main = "Forecast of mutton produced till the year 2021.")
```



From the above forecast, we can see that the production of mutton is likely to decrease in the upcoming 2 years. Looking at the confidence interval, we can see that the probability of the values exceeding the interval is higher for increase in meat production compared to decrease of meat production.

8. Conclusion

We started off the analysis by examining the time series plot for the mutton produced in Australia from 1972 to 2020. Due to the presence of seasonality in the series, we decide to go with a SARIMA model. Following the residual approach, ACF, PACF, EACF and BIC table, we ended up with a set of potential candidate models - {ARIMA(3,1,3), ARIMA(3,1,4), ARIMA(4,1,5), ARIMA(4,1,7), ARIMA(9,1,7)} *(2,1,2)_12. The z test of coefficients was conducted to test the fit of these models using both CSS and ML methods. From the tests, we concluded that the best fit for this series is the SARIMA(4,1,7) *(2,1,2)_12 model. Following this, model overfitting was carried out, resulting in both the overfitted models being insignificant. Next, residual analysis was carried out for the selected model. Finally, the forecast for the next 24 months was plotted. Judging by the forecast, we have concluded that the mutton production in Australia is likely to have a decreasing trend in the upcoming months.