BPDA_Time series

Weixiao Li

2022-11-17

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(tseries)
library(ggplot2)
library(ggfortify)
```

```
## Registered S3 methods overwritten by 'ggfortify':
##   method                 from
##   autoplot.Arima         forecast
##   autoplot.acf           forecast
##   autoplot.ar            forecast
##   autoplot.bats          forecast
##   autoplot.decomposed.ts forecast
##   autoplot.ets           forecast
##   autoplot.forecast      forecast
##   autoplot.stl           forecast
##   autoplot.ts            forecast
##   fitted.ar              forecast
##   fortify.ts             forecast
##   residuals.ar           forecast
```

```
library(fpp)
```

```
##      fma
```

```
##      expsmooth
```

```
##      lmtest
```

```
##      zoo
```

```
##
##    'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(vars)
```

```
##      MASS
```

```
##
##    'MASS'
```

```
## The following objects are masked from 'package:fma':
##
##     cement, housing, petrol
```

```
##      strucchange
```

```
##      sandwich
```

```
##      urca
```

```
library(readxl)
BPDA_Q1_1_ <- read_excel("C:/Users/16932/Desktop/BPDA Q1(1).xlsx")
```

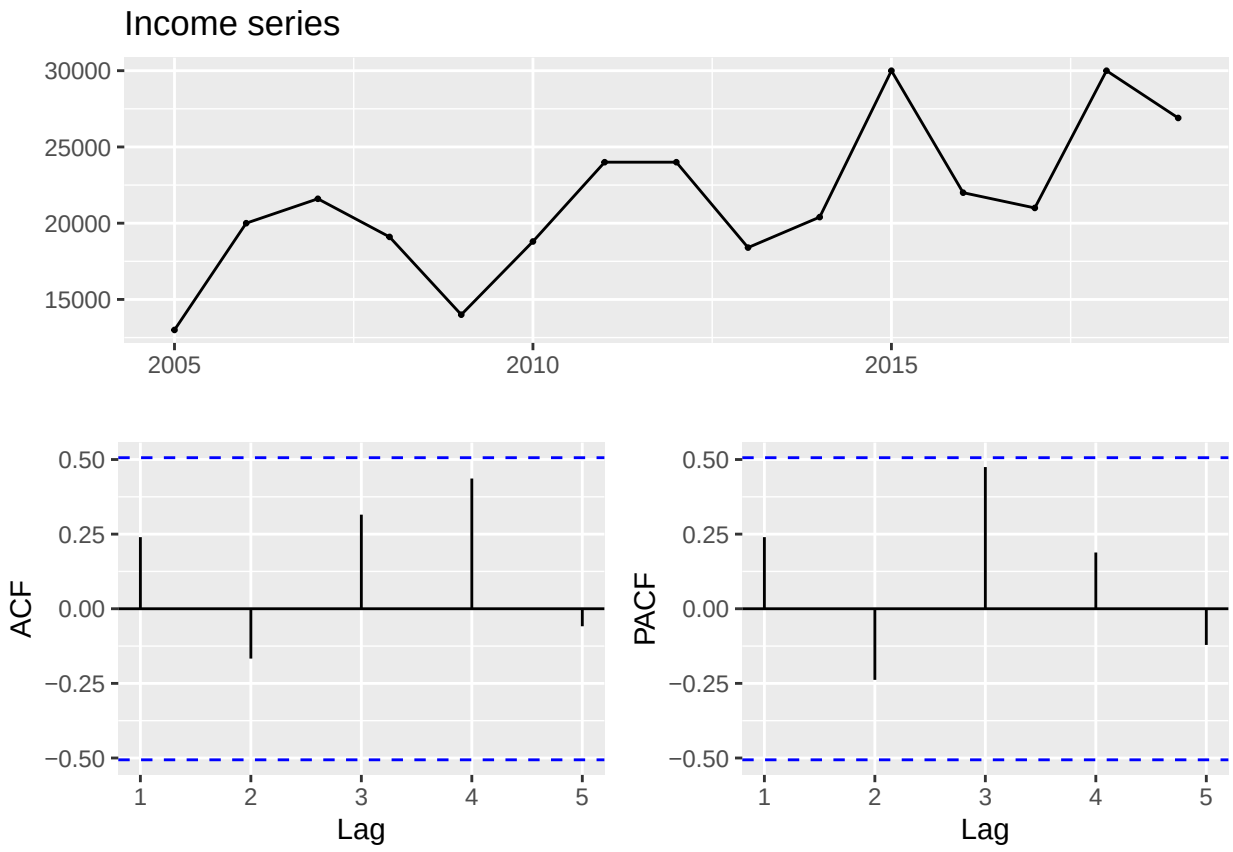
```
## New names:
## * `` -> `...3`
```

```
BPDA <-BPDA_Q1_1_
BPDA <- ts(BPDA,start=2005,end=2019,frequency=1)
```

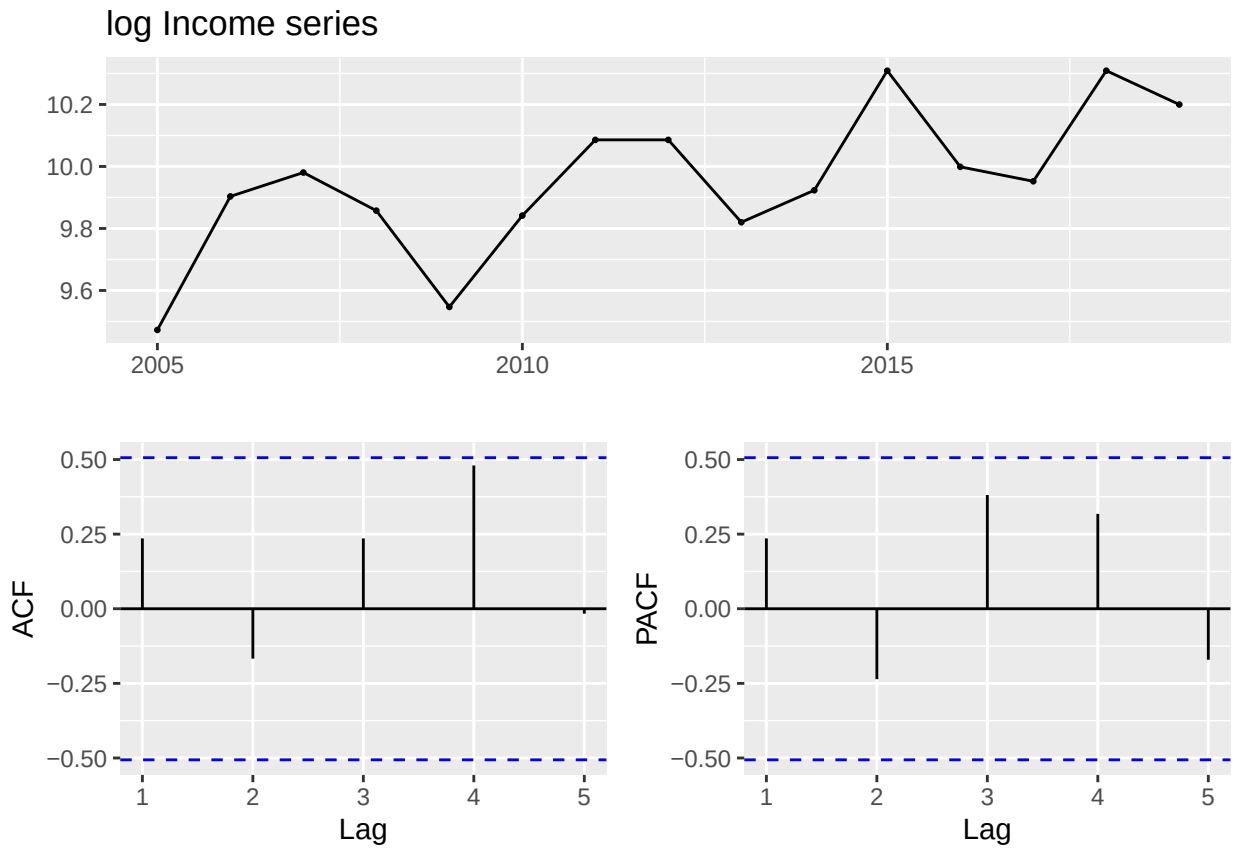
```
#1
Income <- BPDA[, "Income"]
Income
```

```
## Time Series:
## Start = 2005
## End = 2019
## Frequency = 1
##  [1] 13000 20000 21600 19100 14000 18800 24000 24000 18400 20400 30000 22000
## [13] 21000 30000 26900
```

```
ggtsdisplay(Income, main="Income series")
```




```
# log transformation
log_Income <- log(Income)
ggtsdisplay(log_Income, main="log Income series")
```




```
log_Income <- log(Income)
```

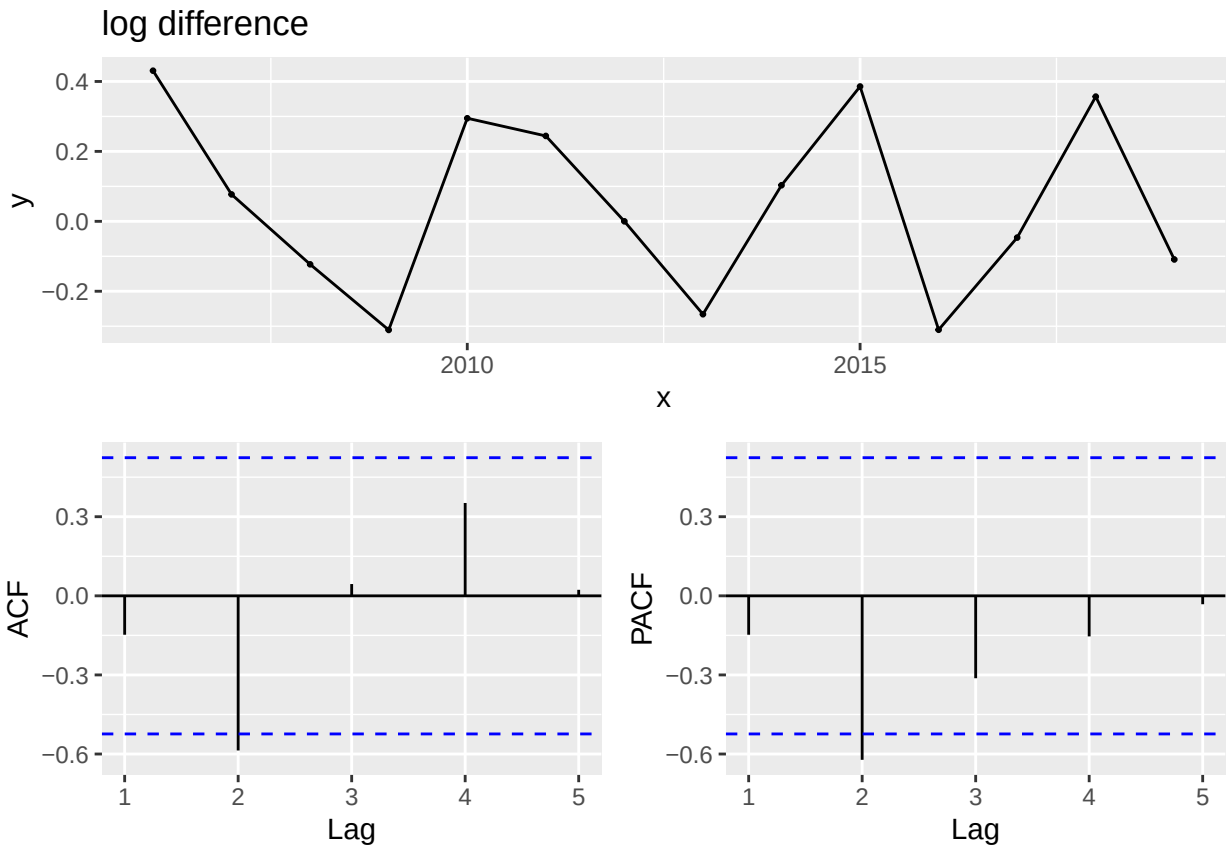
```
#Lag
Income <- as.integer(Income)
lag(Income, k=-1)
```

```
##  [1] 13000 20000 21600 19100 14000 18800 24000 24000 18400 20400 30000 22000
## [13] 21000 30000 26900
## attr(,"tsp")
## [1]  2 16  1
```

```
Income <- lag(Income, k=-1)
diff(Income, lag=1)
```

```
##  [1]  7000  1600 -2500 -5100  4800  5200     0 -5600  2000  9600 -8000 -1000
## [13]  9000 -3100
```

```
log_diff <- diff(log_Income)
ggtsdisplay(log_diff, xlab="x",ylab="y",main="log difference")
```



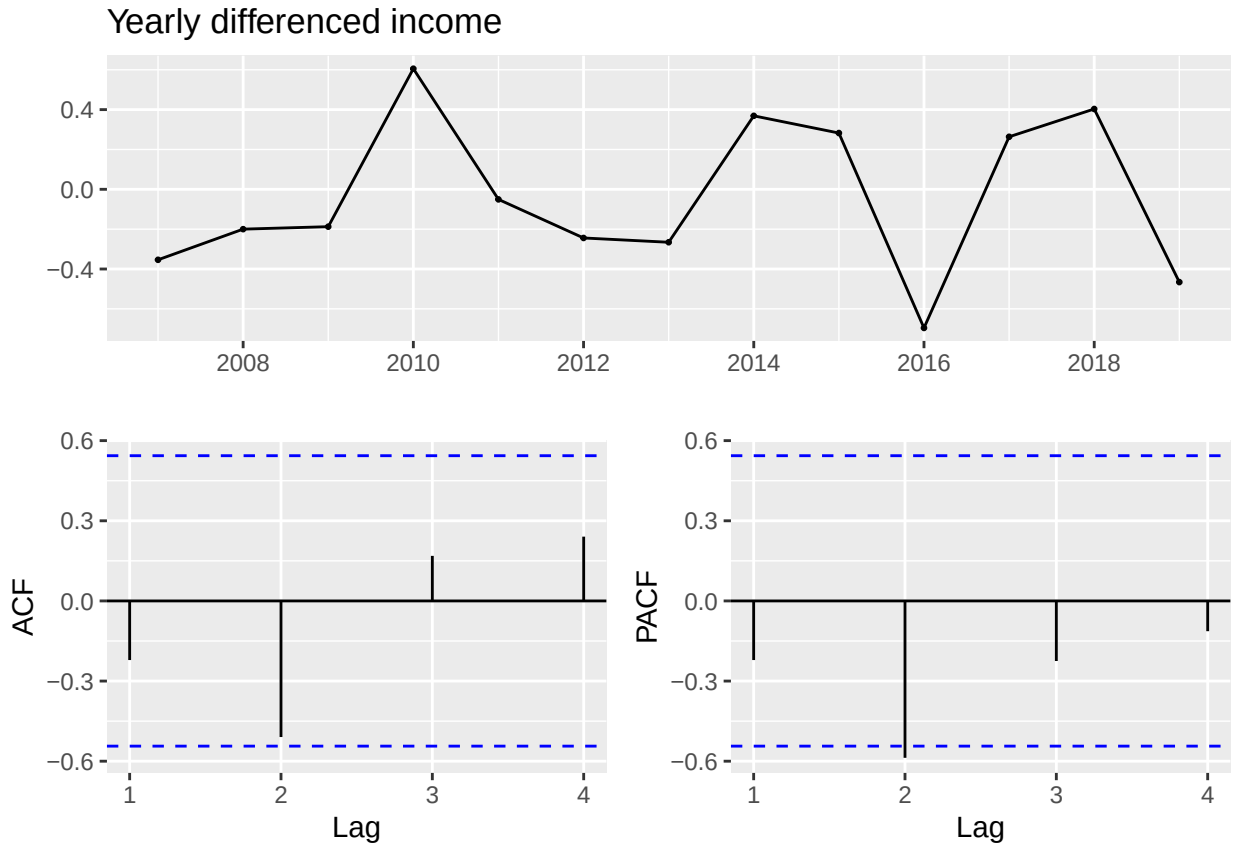

```
#Frequency
frequency(Income)
```

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

```
diff(Income, lag=frequency(Income))
```

```
##  [1]  7000  1600 -2500 -5100  4800  5200     0 -5600  2000  9600 -8000 -1000
## [13]  9000 -3100
```

```
seas_log_diff <- diff(log_diff, lag=frequency(log_diff))
ggtsdisplay(seas_log_diff, main="Yearly differenced income")
```



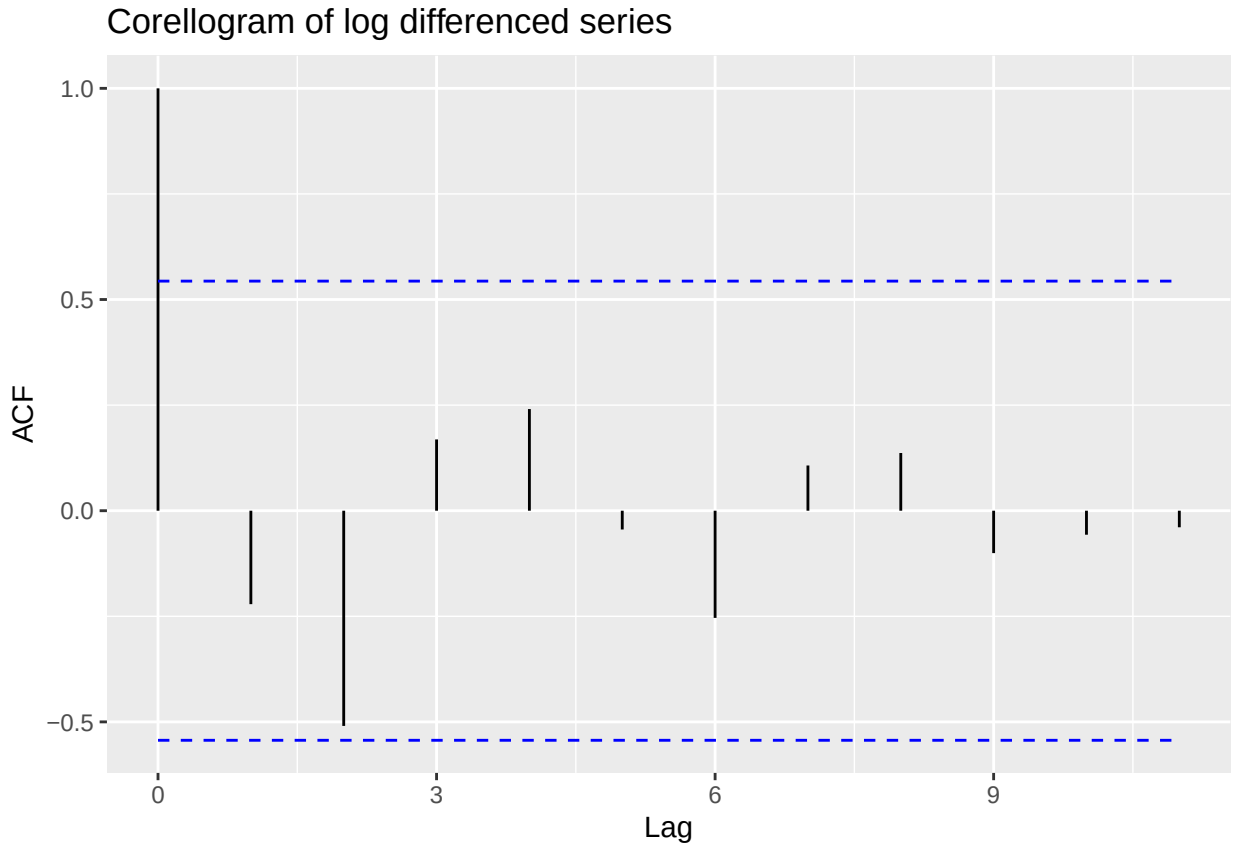

```
# Auto correlation function and partial auto correlation function
acf(seas_log_diff, plot=F, lag.max=12)
```

```
##
## Autocorrelations of series 'seas_log_diff', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000 -0.221 -0.510  0.169  0.241 -0.044 -0.254  0.107  0.137 -0.100 -0.057
##     11     12
## -0.039  0.073
```

```
pacf(seas_log_diff, plot=F, lag.max=12)
```

```
##
## Partial autocorrelations of series 'seas_log_diff', by lag
##
##      1      2      3      4      5      6      7      8      9     10     11
## -0.221 -0.587 -0.225 -0.113  0.061 -0.156  0.007 -0.066 -0.031 -0.032 -0.178
##     12
## -0.137
```

```
autoplot(
  acf(seas_log_diff, plot=F),
  main="Corellogram of log differenced series"
)
```

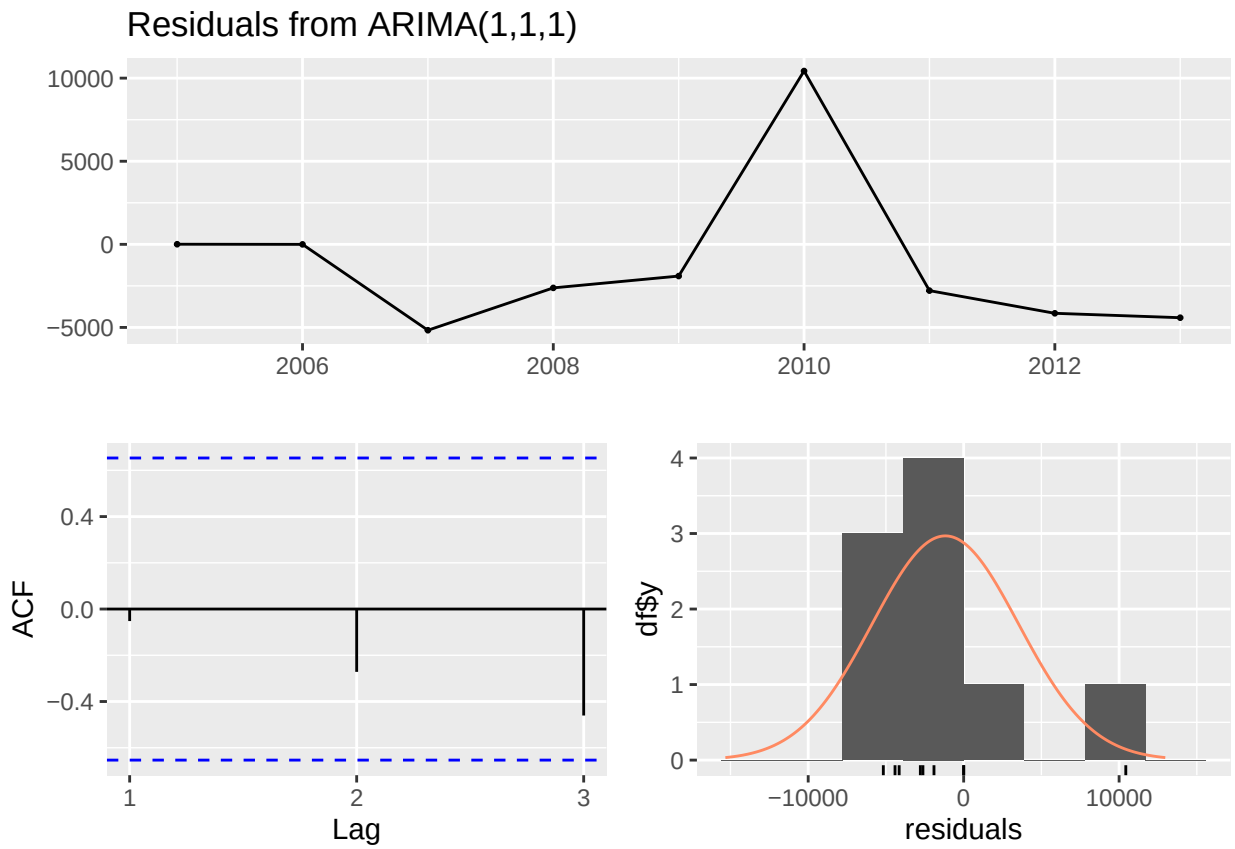



```
## ARIMA
BPDA_FLO <- BPDA[,c("Income", "Citizen", "Not a Citizen")]
train <- window(BPDA, end=c(2013))
test <- window(BPDA, start=c(2014))
Citizen_Not<- train[, c("Citizen", "Not a Citizen")]
model_arima <- Arima(
  y = train[, "Income"],
  order = c(1, 1, 1),
  seasonal = list(order = c(0, 1, 0))
)

model_arima
```

```
## Series: train[, "Income"]
## ARIMA(1,1,1)
##
## Coefficients:
##           ar1     ma1
##       -0.0654  0.3676
## s.e.   0.7259  0.6099
##
## sigma^2 = 38095692:  log likelihood = -69.9
## AIC=145.8   AICc=153.8   BIC=145.64
```

```
checkresiduals(model_arima)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)
## Q* = 5.624, df = 3, p-value = 0.1314
##
## Model df: 2.   Total lags used: 5
```

```
#Auto ARIMA
I_C_N <- auto.arima(
  y = train[, "Income"],
  xreg = Citizen_Not,
  ic = "aic",
  max.order = 7,
  stepwise = F,
  approximation = F,
  parallel = T,
  num.cores = 1
)
I_C_N
```

```
## Series: train[, "Income"]
```

```
## Regression with ARIMA(0,0,0) errors
##
## Coefficients:
##       Citizen  Not a Citizen
##        0.7757         0.5172
## s.e.   0.5161         0.3239
##
## sigma^2 = 8576861:  log likelihood = -83.48
## AIC=172.96   AICc=177.76   BIC=173.55
```