STATS531 Midterm Project \n Time Series Analysis on Fatal Car Accidents in Michigan

Geovanna Caballero, Annie Tai, Paul Zhang

2025-02-11

Introduction

Data Description

The data for this study is obtained from the National Highway Traffic Safety Administration (NHTSA) and covers the years 2008 to 2022 [1]. The data is aggregated on a monthly basis and includes the number of fatal crashes and the number of people killed in fatal crashes in both Michigan and the overall United States.

Exploratory Analysis

```
describe(crashes_MI_ts)
```

Some Descriptive Statistics

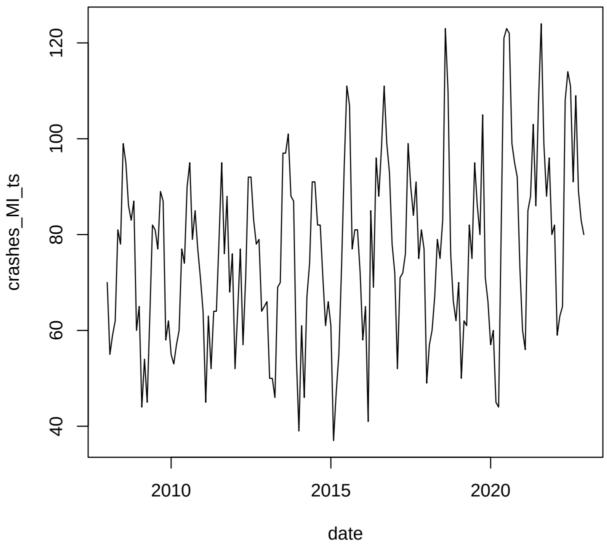
```
##    vars   n  mean    sd median trimmed   mad min max range skew kurtosis  se
## X1    1 180 76.54 18.77     77   75.92 20.02  37 124    87 0.28    -0.33 1.4
```

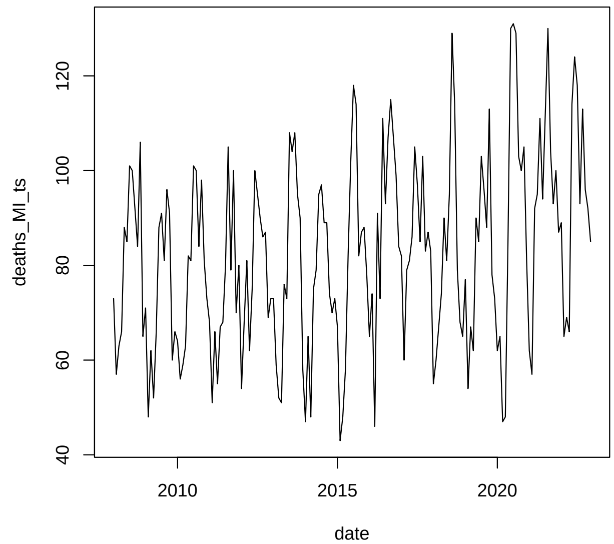
```
describe(deaths_MI_ts)
```

```
##    vars   n  mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 180 82.47 19.85     82   81.93 22.24  43 131    88 0.23    -0.51 1.48
```

```
# psych::describe(crashes_US_ts)
# psych::describe(deaths_US_ts)
```

```
par(mfrow = c(1, 2))
plot(date, crashes_MI_ts, type = "l")
plot(date, deaths_MI_ts, type = "l")
```






```r
# plot(date, crashes_US_ts, type = "l")
# plot(date, deaths_US_ts, type = "l")
```

Model Selection

Stationarity Tests **Updates** I have found another statistical test called the KPSS Test[1]. It seems to be an advanced version of stationary test since it takes trends into consideration. The trend doesn't have to be linear.

The KPSS test may be helpful since our data apparently have nonlinear trends.

The null hypothesis, the alternative hypothesis for the test are as follows:

- H_0: The time series is a trend-stationary process (A stochastic process from which an underlying trend (function solely of time) can be removed, leaving a stationary process[2].)
- H_1: The time series is a unit root process.

```r
kpss.test(crashes_MI_ts, null = "Trend")
```

```
## Warning in kpss.test(crashes_MI_ts, null = "Trend"): p-value greater than
## printed p-value

##
##  KPSS Test for Trend Stationarity
##
## data:  crashes_MI_ts
## KPSS Trend = 0.043981, Truncation lag parameter = 4, p-value = 0.1
```

```r
kpss.test(deaths_MI_ts, null = "Trend")
```

```
## Warning in kpss.test(deaths_MI_ts, null = "Trend"): p-value greater than
## printed p-value

##
##  KPSS Test for Trend Stationarity
##
## data:  deaths_MI_ts
## KPSS Trend = 0.035296, Truncation lag parameter = 4, p-value = 0.1
```

[1] https://en.wikipedia.org/wiki/KPSS_test
[2] https://en.wikipedia.org/wiki/Trend-stationary_process

Both of p-values exceed 0.05, indicating trend stationary.

ARMA Models, Original Data The following code block generates the AIC table, given a stationary time series. It's borrowed from the lecture notes[3].

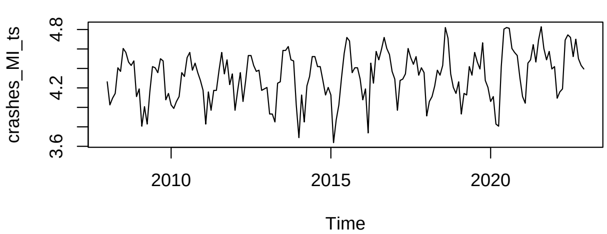
```
aic_table <- function(data,P,Q){
table <- matrix(NA,(P+1),(Q+1))
for(p in 0:P) {
    for(q in 0:Q) {
    # table[p+1,q+1] <- arima2::arima(data,order=c(p,0,q))$aic
    table[p + 1, q + 1] = tryCatch({
        arima(data, order = c(p, 0, q))$aic},
        error = function(e) {NA}
        )
    }
}
dimnames(table) <- list(paste("AR",0:P, sep=""),
paste("MA",0:Q,sep=""))
table
}
require(knitr)
```

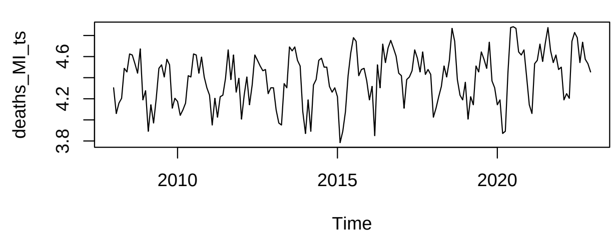
```
## Loading required package: knitr
```

take the log transformation

```
# crashes_MI_diff = diff(crashes_MI_ts, differences = 2)
crashes_MI_ts = log(1 + crashes_MI_ts)
deaths_MI_ts = log(1 + deaths_MI_ts)
```

```
par(mfrow = c(1, 2))
plot(crashes_MI_ts)
plot(deaths_MI_ts)
```






```
crashes_table = aic_table(crashes_MI_ts, 4, 3)
kable(crashes_table, digits=2)
```

Crashes

	MA0	MA1	MA2	MA3
AR0	13.48	-43.85	-79.50	-82.71
AR1	-77.44	-75.44	-85.07	-83.62
AR2	-75.44	-75.95	-137.67	-144.57
AR3	-94.74	-105.79	-148.04	-135.10

[3] https://ionides.github.io/531w25/05/slides.pdf, pp.29

	MA0	MA1	MA2	MA3
AR4	-100.98	-105.20	-102.80	-144.04

ARMA(3, 2)

```
crashes_arma = arima(crashes_MI_ts, order = c(3, 0, 2))
summary(crashes_arma)
```

```
##
## Call:
## arima(x = crashes_MI_ts, order = c(3, 0, 2))
##
## Coefficients:
##          ar1      ar2     ar3      ma1     ma2  intercept
##       1.9978  -1.4632  0.2678  -1.7184  0.9997     4.3219
## s.e.  0.0756   0.1307  0.0755   0.0237  0.0236     0.0159
##
## sigma^2 estimated as 0.02264:  log likelihood = 81.02,  aic = -148.04
##
## Training set error measures:
##                        ME      RMSE      MAE        MPE     MAPE      MASE
## Training set -0.001296803 0.1504754 0.118653 -0.1654812 2.784281 0.7192993
##                     ACF1
## Training set -0.04115989
```

Diagnostics:

- Check the AR roots (the code also comes from the lecture notes[4]):

```
AR_roots <- polyroot(c(1,-coef(crashes_arma)[c("ar1","ar2", "ar3")]))
abs(AR_roots)
```

```
## [1] 1.000133 1.000133 3.732854
```

- Ljung-Box Test

```
Box.test(crashes_arma$residuals, lag = 20, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  crashes_arma$residuals
## X-squared = 45.283, df = 20, p-value = 0.00101
```

```
adf.test(crashes_arma$residuals)
```

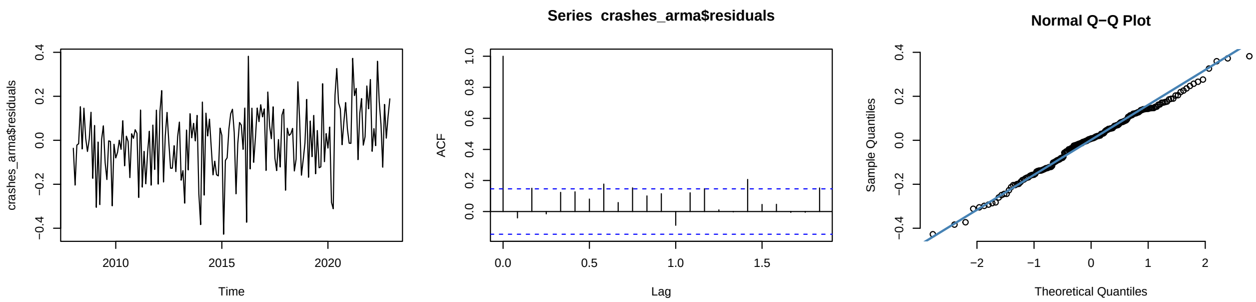
```
## Warning in adf.test(crashes_arma$residuals): p-value smaller than printed
## p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  crashes_arma$residuals
## Dickey-Fuller = -5.2483, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

- Residual Plots

[4] https://ionides.github.io/531w25/05/slides-annotated.pdf, p.32

```r
par(mfrow = c(1, 3))
plot(crashes_arma$residuals)
acf(crashes_arma$residuals)
qqnorm(crashes_arma$residuals, pch = 1, frame = FALSE)
qqline(crashes_arma$residuals, col = "steelblue", lwd = 2)
```



positive autocorrelation; right skew

Deaths The deaths series.

```r
deaths_table = aic_table(deaths_MI_ts, 4, 3)
kable(deaths_table, digits=2)
```

	MA0	MA1	MA2	MA3
AR0	7.21	-44.07	-82.69	-83.76
AR1	-75.71	-73.76	-85.68	-83.74
AR2	-73.83	-76.30	-135.87	-141.12
AR3	-95.66	-104.16	-144.01	-133.04
AR4	-98.59	-102.43	-101.39	-140.03

ARMA(3, 2)

```r
deaths_arma = arima(deaths_MI_ts, order = c(3, 0, 2))
summary(deaths_arma)
```

```
##
## Call:
## arima(x = deaths_MI_ts, order = c(3, 0, 2))
##
## Coefficients:
##          ar1      ar2     ar3      ma1     ma2  intercept
##       1.9724  -1.4192  0.2424  -1.7203  0.9995     4.3964
## s.e.  0.0753   0.1302  0.0752   0.0234  0.0258     0.0155
##
## sigma^2 estimated as 0.02317:  log likelihood = 79.01,  aic = -144.01
##
## Training set error measures:
##                        ME      RMSE       MAE        MPE     MAPE      MASE
## Training set -0.001239843 0.1522095 0.1238735 -0.1604316 2.850103 0.7302796
##                     ACF1
## Training set -0.03679909
```

Diagnostics

- Check AR roots

```r
AR_roots <- polyroot(c(1,-coef(deaths_arma)[c("ar1","ar2","ar3")]))
abs(AR_roots)
```

```
## [1] 1.000094 1.000094 4.124917
```

- Ljung-Box Test

```r
Box.test(deaths_arma$residuals, lag = 20, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  deaths_arma$residuals
## X-squared = 46.677, df = 20, p-value = 0.0006503
```

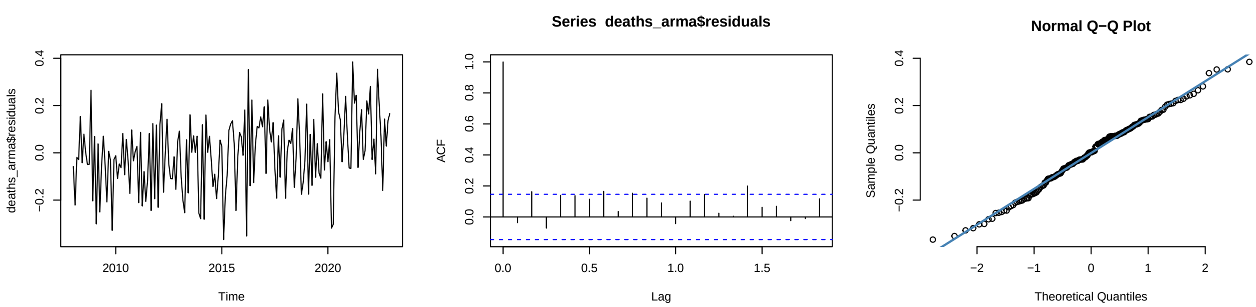
```r
adf.test(deaths_arma$residuals)
```

```
## Warning in adf.test(deaths_arma$residuals): p-value smaller than printed
## p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  deaths_arma$residuals
## Dickey-Fuller = -5.1643, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

- Residual Plots

```r
par(mfrow = c(1, 3))
plot(deaths_arma$residuals)
acf(deaths_arma$residuals)
qqnorm(deaths_arma$residuals, pch = 1, frame = FALSE)
qqline(deaths_arma$residuals, col = "steelblue", lwd = 2)
```

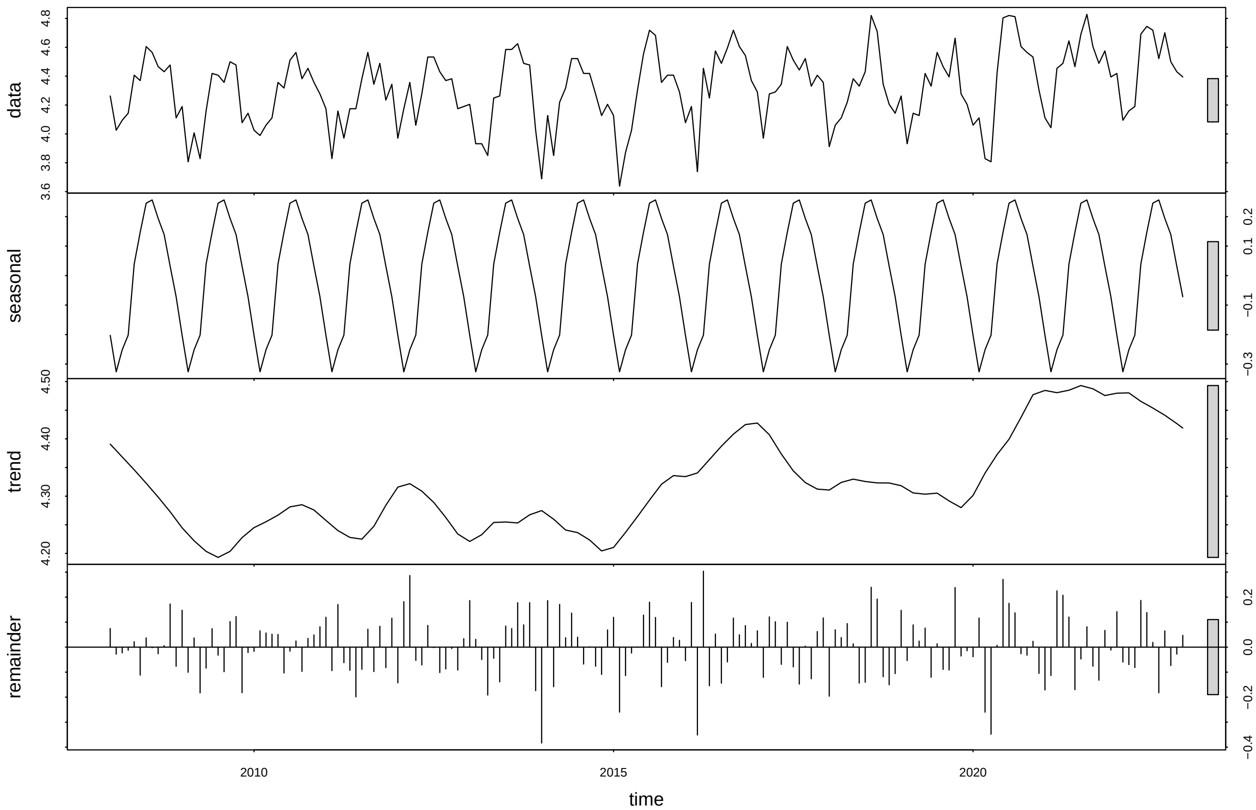


ARMA models, After Detrending Let's see what happens after we detrend the data. (I read some docs[5] be4 writing the following code.)

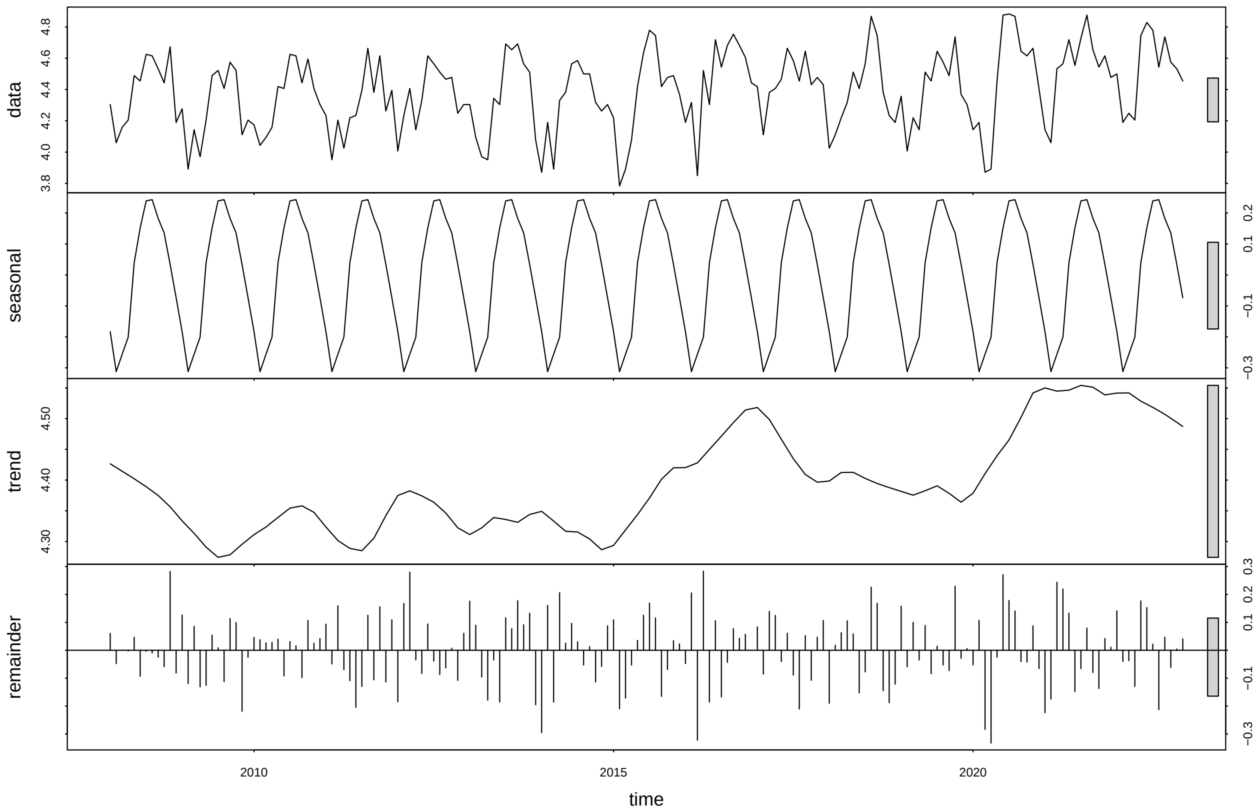
```r
crashes_MI_decomposed = stl(crashes_MI_ts, s.window = "periodic")
crashes_MI_detrended = crashes_MI_ts - crashes_MI_decomposed$time.series[, "trend"]

deaths_MI_decomposed = stl(deaths_MI_ts, s.window = "periodic")
deaths_MI_detrended = deaths_MI_ts - deaths_MI_decomposed$time.series[, "trend"]
```

[5] https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/stl

```r
plot(crashes_MI_decomposed)
```




```r
plot(deaths_MI_decomposed)
```



Crashes, detrended ARMA(3, 3)

```
crashes_detrend_table = aic_table(crashes_MI_detrended, 4, 3)
kable(crashes_detrend_table, digits=2)
```

	MA0	MA1	MA2	MA3
AR0	-10.45	-59.53	-89.71	-90.81
AR1	-83.20	-81.31	-91.40	-89.91
AR2	-81.46	-81.04	-224.02	-226.50
AR3	-107.54	-196.71	-225.72	-240.89
AR4	-121.61	-205.94	-223.99	-222.50

```
crashes_detrend_arma = arima(crashes_MI_detrended, order = c(3, 0, 3))
```

Diagnostics

- Check the AR roots

```
AR_roots <- polyroot(c(1,-coef(crashes_detrend_arma)[c("ar1","ar2", "ar3")]))
abs(AR_roots)
```

```
## [1] 1.000019 1.000019 1.744419
```

- Ljung-Box Test

```
Box.test(crashes_detrend_arma$residuals, lag = 20, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
```

```
##
## data:  crashes_detrend_arma$residuals
## X-squared = 37.535, df = 20, p-value = 0.01009
```

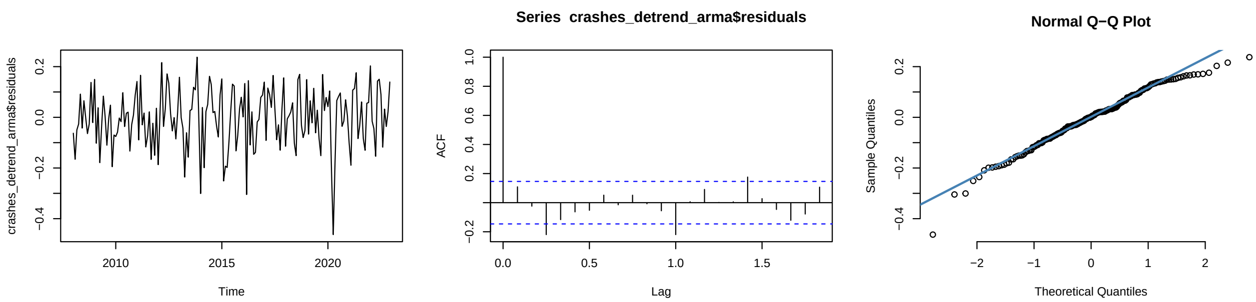
```
adf.test(crashes_detrend_arma$residuals)
```

```
## Warning in adf.test(crashes_detrend_arma$residuals): p-value smaller than
## printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  crashes_detrend_arma$residuals
## Dickey-Fuller = -7.0768, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

- Residual Plots

```
par(mfrow = c(1, 3))
plot(crashes_detrend_arma$residuals)
acf(crashes_detrend_arma$residuals)
qqnorm(crashes_detrend_arma$residuals, pch = 1, frame = FALSE)
qqline(crashes_detrend_arma$residuals, col = "steelblue", lwd = 2)
```



Deaths, detrended ARMA(2, 3)

```
deaths_detrend_table = aic_table(deaths_MI_detrended, 4, 3)
kable(deaths_detrend_table, digits=2)
```

	MA0	MA1	MA2	MA3
AR0	-16.35	-59.65	-92.53	-91.96
AR1	-81.88	-79.89	-92.50	-90.57
AR2	-79.90	-81.56	-220.72	-222.88
AR3	-108.87	-193.67	-222.44	-220.76
AR4	-117.91	-198.84	-190.69	-218.56

```
deaths_detrend_arma = arima(deaths_MI_detrended, order = c(2, 0, 3))
```

Diagnostics

- Check the AR roots

```
AR_roots <- polyroot(c(1,-coef(deaths_detrend_arma)[c("ar1","ar2")]))
abs(AR_roots)
```

```
## [1] 1.018732 1.018732
```

- Ljung-Box Test

```
Box.test(deaths_detrend_arma$residuals, lag = 20, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  deaths_detrend_arma$residuals
## X-squared = 31.035, df = 20, p-value = 0.05474
```

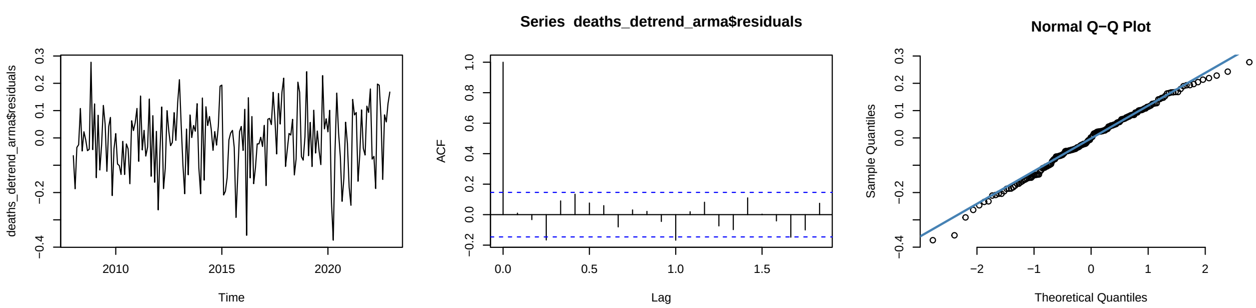
```
adf.test(deaths_detrend_arma$residuals)
```

```
## Warning in adf.test(deaths_detrend_arma$residuals): p-value smaller than
## printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  deaths_detrend_arma$residuals
## Dickey-Fuller = -4.5201, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

- Residual Plots

```
par(mfrow = c(1, 3))
plot(deaths_detrend_arma$residuals)
acf(deaths_detrend_arma$residuals)
qqnorm(deaths_detrend_arma$residuals, pch = 1, frame = FALSE)
qqline(deaths_detrend_arma$residuals, col = "steelblue", lwd = 2)
```



Conclusion

References

[1] https://cdan.dot.gov/query [2] https://ionides.github.io/531w25/08/slides.pdf, p.6