EXAM2submission

Michael Pena

2024-06-08

Question 1.

1.

A B F

2a.

notice that when we take the expected value of X_t

$$
\begin{aligned}
E[X_t] &= E[\beta_0 + \beta_1 t + \beta_2 sin(\frac{\pi}{2}t) + Z_t] \\
&= E[\beta_0 + \beta_1 t + \beta_2 sin(\frac{\pi}{2}t)] + E[Z_t] \\
&= \beta_0 + \beta_1 t + \beta_2 sin(\frac{\pi}{2}t) + 0
\end{aligned}
$$

The above expectation results in a function that is dependent on t, thus it fails the first requirement to be a stationary system.

2b.

given the definition $\nabla_4 X_t = X_t - X_{t-4}$ we can assume that

$$
\begin{aligned}
\nabla_4 X_t &= \beta_0 + \beta_1 t + \beta_2 sin(\frac{\pi}{2}t) + Z_t - \left(\beta_0 + \beta_1(t-4) + \beta_2 sin(\frac{\pi}{2}(t-4)) + Z_{t-4}\right) \\
&= \beta_0 - \beta_0 + \beta_1 t - \beta_1 t + 4\beta_1 + \beta_2 sin(\frac{\pi}{2}t) - \beta_2 sin(\frac{\pi}{2}(t-4)) + Z_t - Z_{t-4}
\end{aligned}
$$

notice that

$$
\begin{aligned}
sin(\frac{\pi}{2}(t-4)) &= sin(\frac{\pi}{2}t - 2\pi) \\
&= sin(\frac{\pi}{2}t)cos(-2\pi) + sin(-2\pi)cos(\frac{\pi}{2}t) \\
&= sin(\frac{\pi}{2}t) \cdot 1 + 0 = sin(\frac{\pi}{2}t)
\end{aligned}
$$

thus

$$
\begin{aligned}
&\beta_0 - \beta_0 + \beta_1 t - \beta_1 t + 4\beta_1 + \beta_2 sin(\frac{\pi}{2}t) - \beta_2 sin(\frac{\pi}{2}(t-4)) + Z_t - Z_{t-4} \\
&= \beta_0 - \beta_0 + \beta_1 t - \beta_1 t + 4\beta_1 + \beta_2 sin(\frac{\pi}{2}t) - \beta_2 sin(\frac{\pi}{2}t) + Z_t - Z_{t-4} \\
&= 0 + 0 + 4\beta_1 + 0 + Z_t - Z_{t-4} \\
&= 4\beta_1 + Z_t - Z_{t-4}
\end{aligned}
$$

moving forward to the expected value

$$
\begin{aligned}
&E[4\beta_1 + Z_t - Z_{t-4}] \\
&E[4\beta_1] + E[Z_t] - E[Z_{t-4}] \\
&= 4\beta_1 + 0 + 0 \\
&= 4\beta_1
\end{aligned}
$$

this expectation is depenedent on t; $\nabla_4 X_t$ does not have a time-dependent mean.

3.

using Bayes

$$
\begin{aligned}
&f(x_n|x_{n-1}, ..., x_1) f(x_{n-1}, ..., x_1) \\
&= f(x_n|x_{n-1}, ..., x_1) f(x_{n-1}|x_{n-2}, ..., x_1) f(x_{n-2}, ..., x_1) \\
&= f(x_n|x_{n-1}, ..., x_1) f(x_{n-1}|x_{n-2}, ..., x_1) f(x_{n-2}|x_{n-3}, ..., x_1) f(x_{n-3}, ..., x_1) \\
&\text{...keep extracting...} \\
&= \prod_{i=2}^{n} f(x_i|x_{i-1}, ..., x_1) f(x_1)
\end{aligned}
$$

note that

$X_i|X_{i-1},...,X_1 = X_i|X_{i-1} \sim f(X_i|X_{i-1})$

$$E[X_i|X_{i-1}] = \phi X_{t-1}$$

$$Var(X_i|X_{i-1}) = \sigma^2$$

$$\text{express function}$$

$$\implies f(X_i|X_{i-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{-(x_i - \phi X_{i-1})^2}{2\sigma^2}\}$$

$$\implies \ell(X_i|X_{i-1}) = \sum_{i=2}^{n} ln(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{-(x_i - \phi X_{i-1})^2}{2\sigma^2}\})$$

$$\sum_{i=2}^{n} \frac{-1}{2} ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}(x_i - \phi x_{i-1})$$

$$\text{deriving with respect to } \phi$$

$$\frac{\partial \ell}{\partial \phi} = \sum_{i=2}^{n} \frac{1}{2\sigma^2} 2(x_i - \phi x_{i-1})(-x_{i-1})$$

$$\text{set partial to 0}$$

$$\implies 0 = \sum_{i=2}^{n} \frac{1}{2\sigma^2} 2(x_i - \phi x_{i-1})(-x_{i-1})$$

$$0 = \sum_{i=2}^{n} \frac{1}{\sigma^2}(-x_i x_{i-1} + \phi x_{i-1}^2)$$

$$0 = \sum_{i=2}^{n} -x_i x_{i-1} + \phi x_{i-1}^2$$

$$\text{isolate } \phi$$

$$\phi \sum_{i=2}^{n} x_{i-1}^2 = \sum_{i=2}^{n} x_i x_{i-1}$$

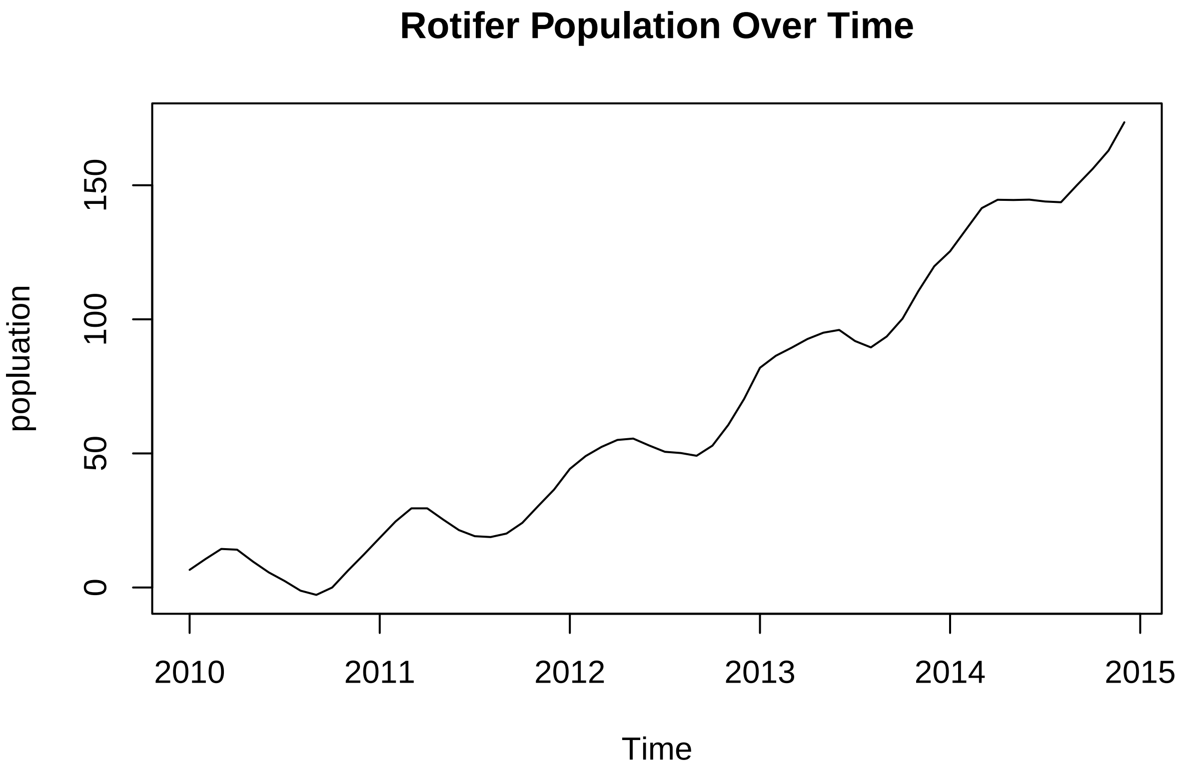
$$\phi = \frac{\sum_{i=2}^{n} x_i x_{i-1}}{\sum_{i=2}^{n} x_{i-1}^2}$$

MLE for $\hat{\phi}$ is $\frac{\sum_{i=2}^{n} x_i x_{i-1}}{\sum_{i=2}^{n} x_{i-1}^2}$.

Question 2.

part (a).

```r
plot.ts(series, ylab = "popluation", main = "Rotifer Population Over Time")
```



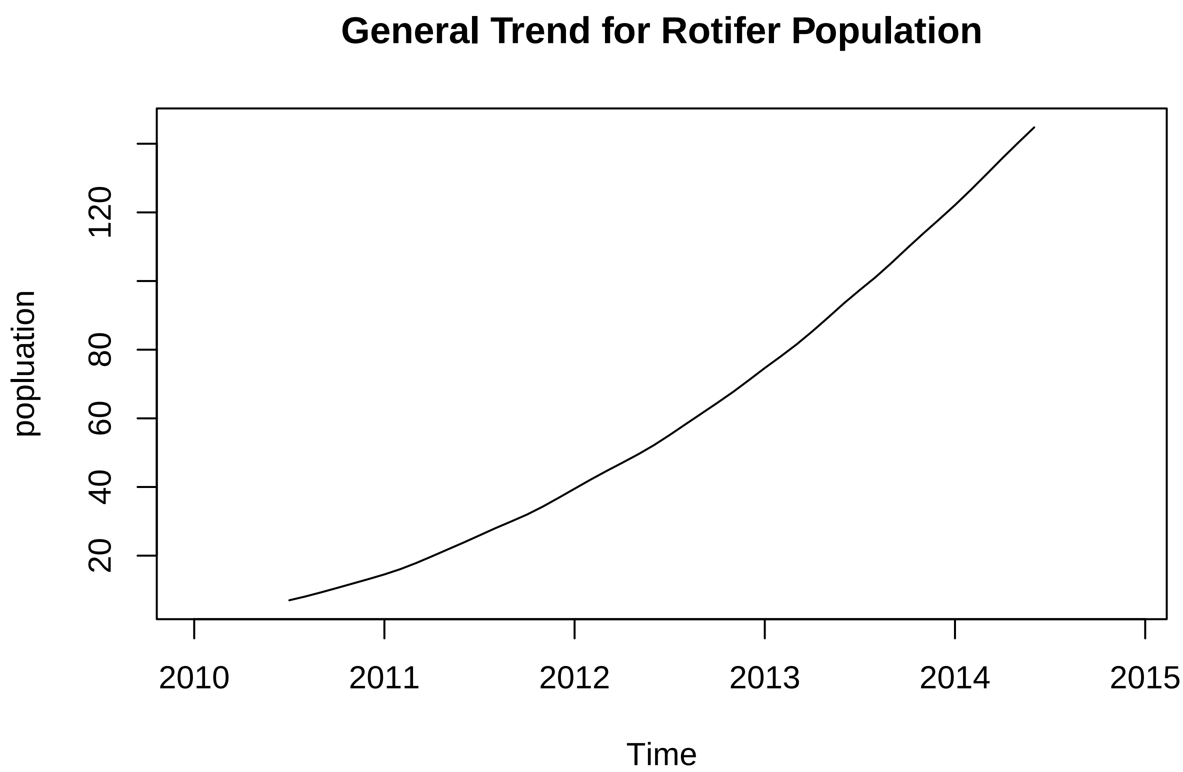
by simply looking at the graph, there seems to be peaks around the 4 month and troughs around the 8th month of every cycle. This could be interpreted as an expansion in population during April and a recession in population in August.

part (b).

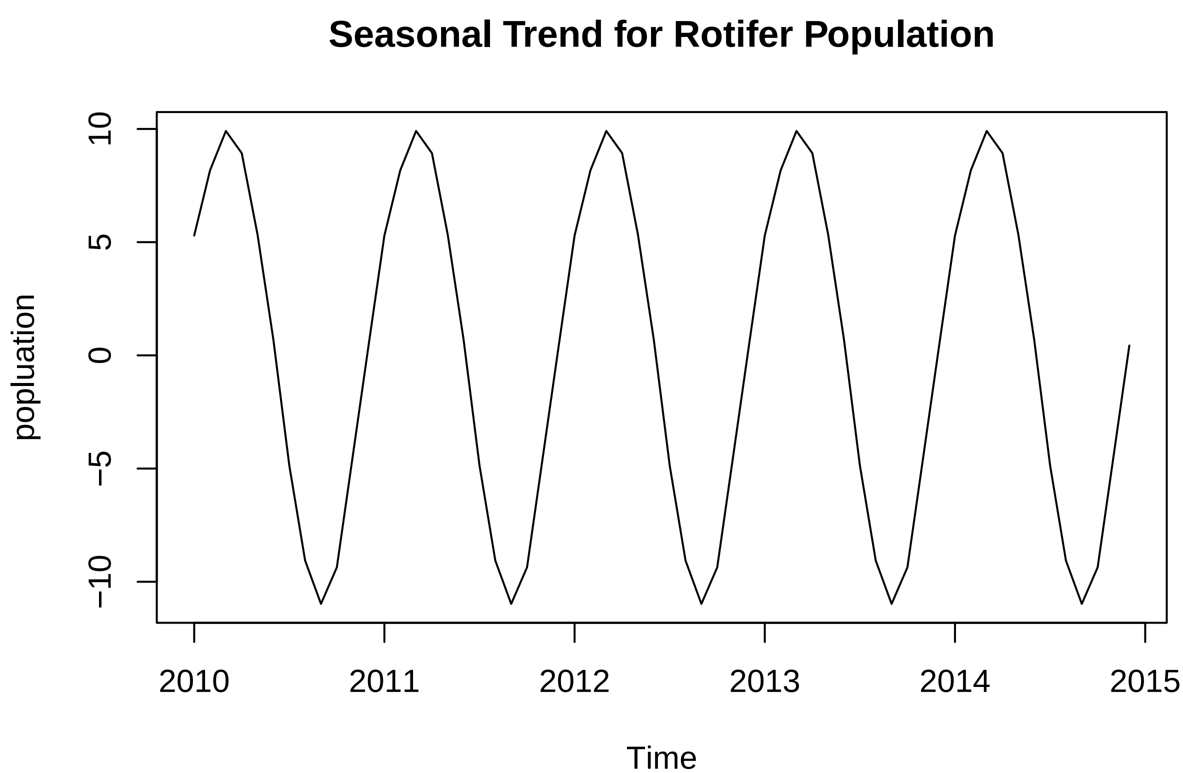
```
m <- decompose(series, type = "additive")
m$figure
```

```
## [1]  5.2915233  8.1712939  9.9105755  8.9306827  5.3224497  0.6762617
## [7] -4.8604657 -9.0656164 -10.9770474 -9.3627326 -4.4704093  0.4334846
```

```
plot(m$trend,
     ylab = "popluation",
     main = "General Trend for Rotifer Population")
```

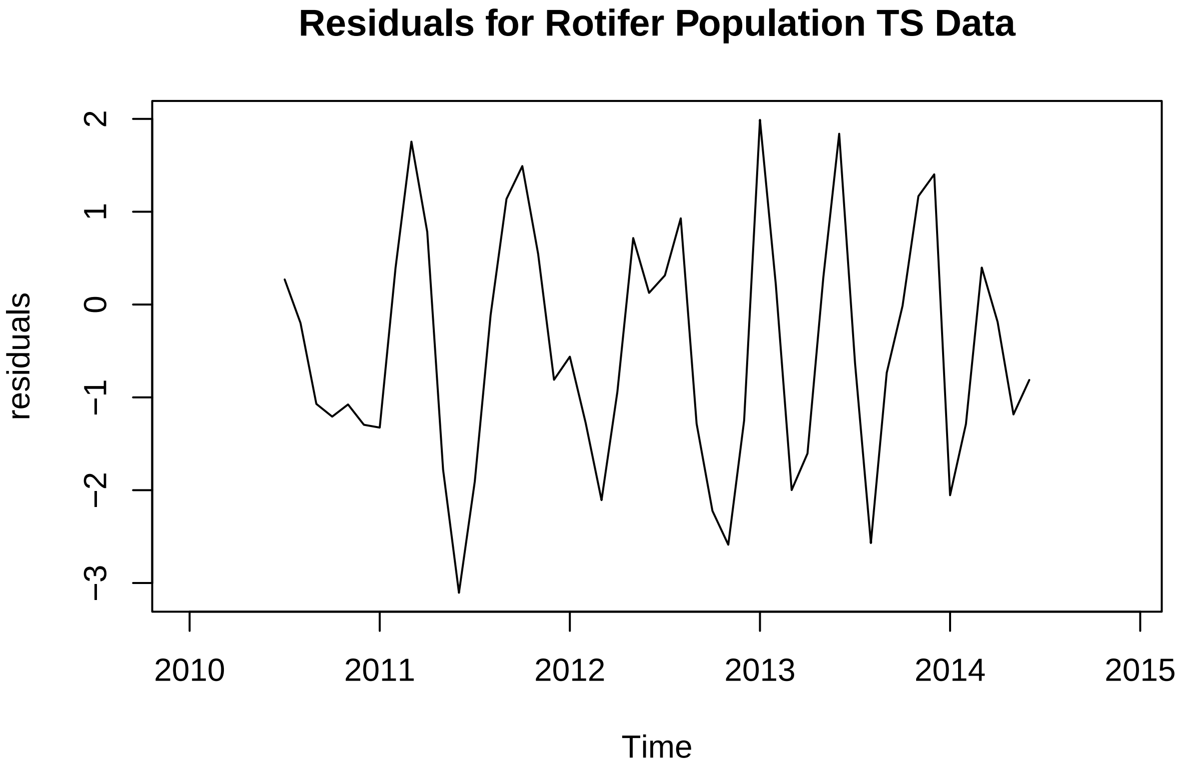



```r
plot(m$seasonal,
     ylab = "popluation",
     main = "Seasonal Trend for Rotifer Population")
```



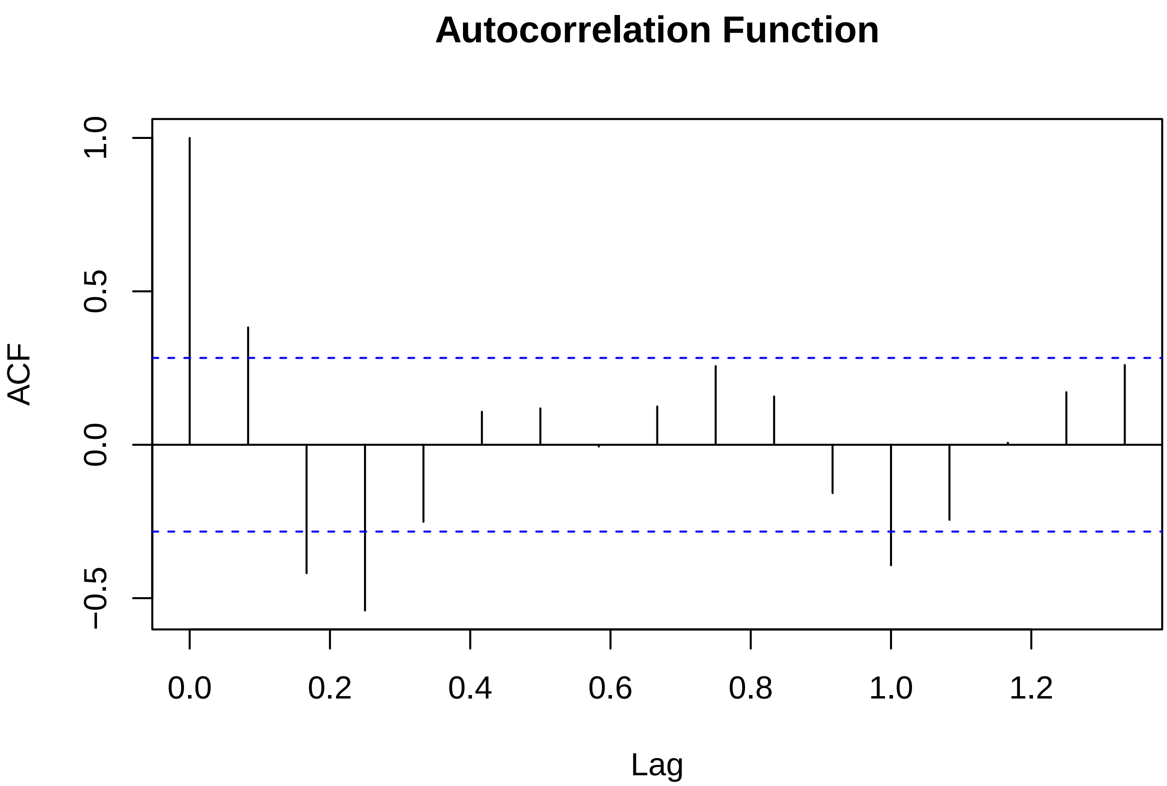

```r
plot(m$random,
     ylab = "residuals",
```

```
    main = "Residuals for Rotifer Population TS Data")
```



part (c)

```
acf(na.omit(m$random), main = "Autocorrelation Function")
```



Residuals seem to not look stationary as there is a linear trend in height downward. They also seem to peak out of the confidence interval cyclically.

part (d)

because ARMA only works with stationary systems, we need to do some referencing

```r
# differencing
seriesD <- diff(series,diff = 1)
# ARMA1
arma1 <-arima(seriesD,order = c(1,0,1))
arma1
```

```
##
## Call:
## arima(x = seriesD, order = c(1, 0, 1))
##
## Coefficients:
##          ar1     ma1  intercept
##       0.5532  0.7496     3.0659
## s.e.  0.1212  0.1017     1.1562
##
## sigma^2 estimated as 5.402:  log likelihood = -134.42,  aic = 276.84
```

```r
# ARMA2
arma2 <-arima(seriesD,order = c(2,0,0))
arma2
```

```
##
## Call:
## arima(x = seriesD, order = c(2, 0, 0))
##
```

```
## Coefficients:
##          ar1      ar2  intercept
##       1.1672  -0.4911     3.0132
## s.e.  0.1141   0.1140     0.9414
##
## sigma^2 estimated as 5.54:  log likelihood = -134.97,  aic = 277.95
```

```r
# ARMA3
arma3 <-arima(seriesD,order = c(0,0,2))
arma3
```

```
##
## Call:
## arima(x = seriesD, order = c(0, 0, 2))
##
## Coefficients:
##          ma1     ma2  intercept
##       1.2078  0.3528     2.9268
## s.e.  0.1118  0.0965     0.8027
##
## sigma^2 estimated as 5.943:  log likelihood = -137.22,  aic = 282.45
```

```r
# ARMA4
arma4 <-arima(seriesD,order = c(2,0,1))
arma4
```

```
##
## Call:
## arima(x = seriesD, order = c(2, 0, 1))
##
## Coefficients:
##          ar1     ar2     ma1  intercept
##       0.4946  0.0775  0.7824     3.0980
## s.e.  0.1683  0.1639  0.1018     1.2246
##
## sigma^2 estimated as 5.382:  log likelihood = -134.31,  aic = 278.62
```

```r
# ARMA5
arma5 <-arima(seriesD,order = c(2,0,2))
arma5
```

```
##
## Call:
## arima(x = seriesD, order = c(2, 0, 2))
##
## Coefficients:
##          ar1     ar2     ma1     ma2  intercept
##       0.1878  0.2658  1.0796  0.2205     3.1055
## s.e.  0.4787  0.2817  0.4719  0.3625     1.2334
##
## sigma^2 estimated as 5.356:  log likelihood = -134.15,  aic = 280.31
```

After a single iteration of differencing, the arima() function was able to accept the model.

```r
# Make table
fr <- data.frame(
Model = c("ARMA(1,0,1)","ARMA(2,0,0)","ARMA(0,0,2)","ARMA(2,0,1)","ARMA(2,0,2)"),
ar1 = c(0.5532,1.1672,'-',0.4946,0.1878),
```

```r
ar2 = c('-',-0.4911,'-',0.0775,0.2658),
ma1 = c(0.7496,'-',1.2078,0.7824,1.0796),
ma2 = c('-','-',0.3528,'-',0.2205),
intercept = c(3.0659,3.0132,2.9268,3.0980,3.1055),
AIC = c(276.84,277.95,282.45,278.62,280.31))
# render table
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```r
kable(fr,caption = "Parameter Estimates and AIC for ARMA models")
```

Table 1: Parameter Estimates and AIC for ARMA models

Model	ar1	ar2	ma1	ma2	intercept	AIC
ARMA(1,0,1)	0.5532	-	0.7496	-	3.0659	276.84
ARMA(2,0,0)	1.1672	-0.4911	-	-	3.0132	277.95
ARMA(0,0,2)	-	-	1.2078	0.3528	2.9268	282.45
ARMA(2,0,1)	0.4946	0.0775	0.7824	-	3.0980	278.62
ARMA(2,0,2)	0.1878	0.2658	1.0796	0.2205	3.1055	280.31

Our lowest AIC is coming from the model ARMA(1,0,1) which would make sense as it is one of the models with the least parameters. Based on AIC, ARMA(1,0,1) seems to fit the best out of all models; however, I think it would be in our best interest if we did not use AIC alone to determine the model that fits the best.