Residual Diagnostics and Model Selection

Relevant libraries

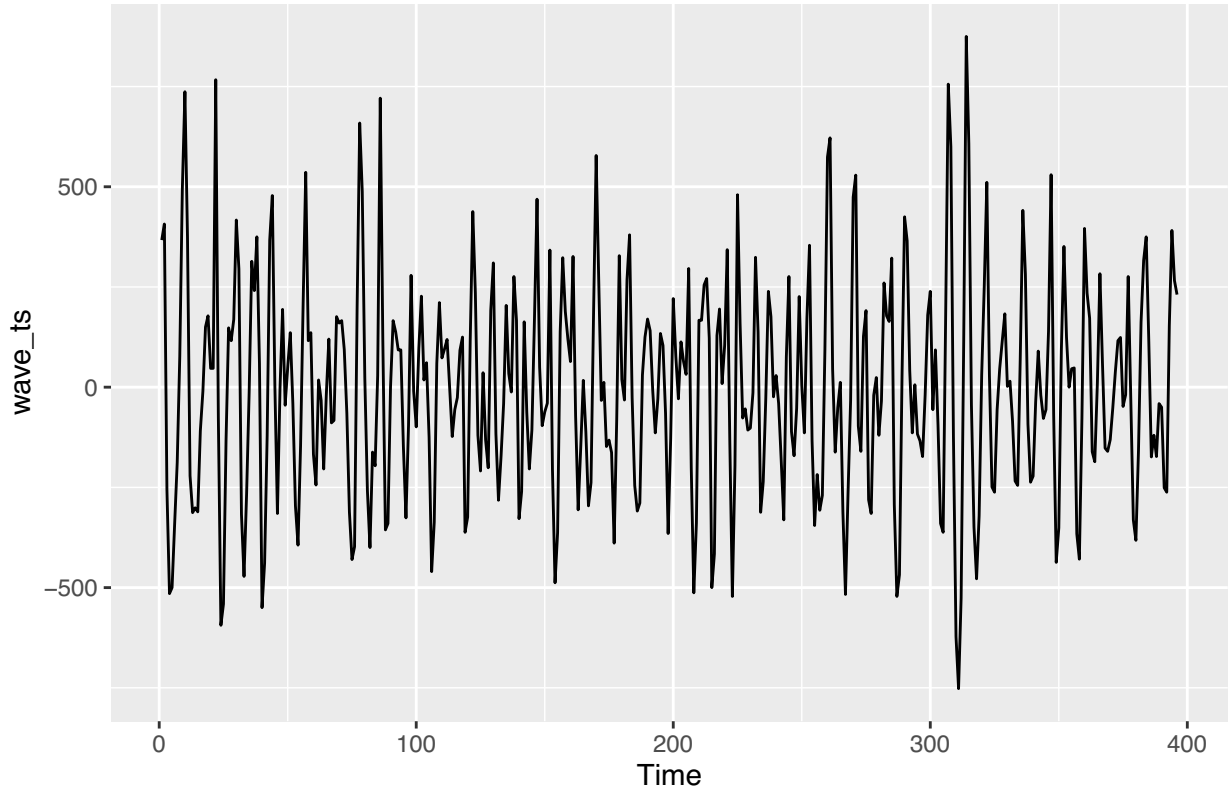
```
library(tidyverse)
library(tidyquant)
library(gridExtra)
library(tibbletime)
library(forecast)
library(itsmr)
library(here)
library(bbmle)
knitr::opts_chunk$set(comment=NA,tidy=FALSE)

#library(future) Not needed yet
#library(doFuture) Not needed yet
#library(rbenchmark) Not needed yet
```

Wave tank data: Wave height data collected from a wave tank

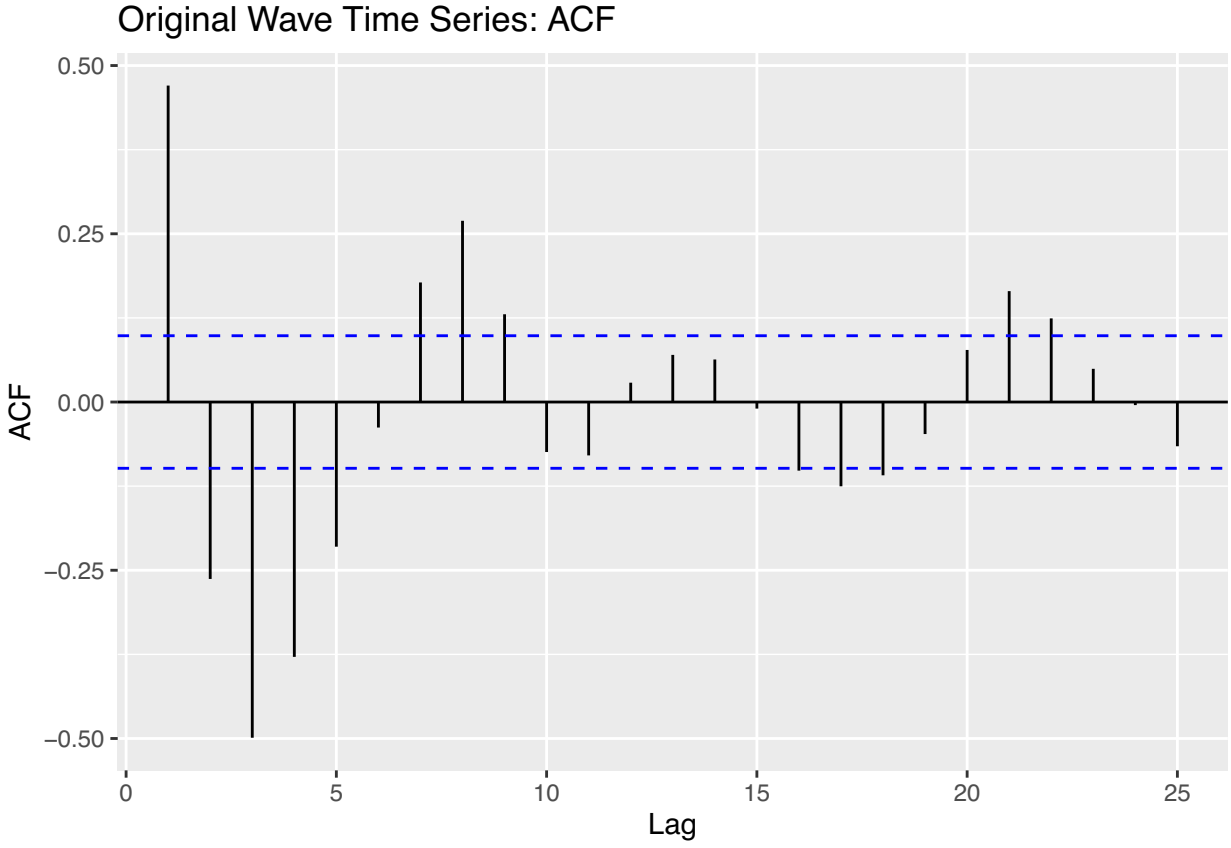
```
#wave_data = read_csv(here("data/wave_data.csv"))
wave_data = read_csv("wavedat_new.csv")
```

```
Parsed with column specification:
cols(
  waveht = col_double()
)
```

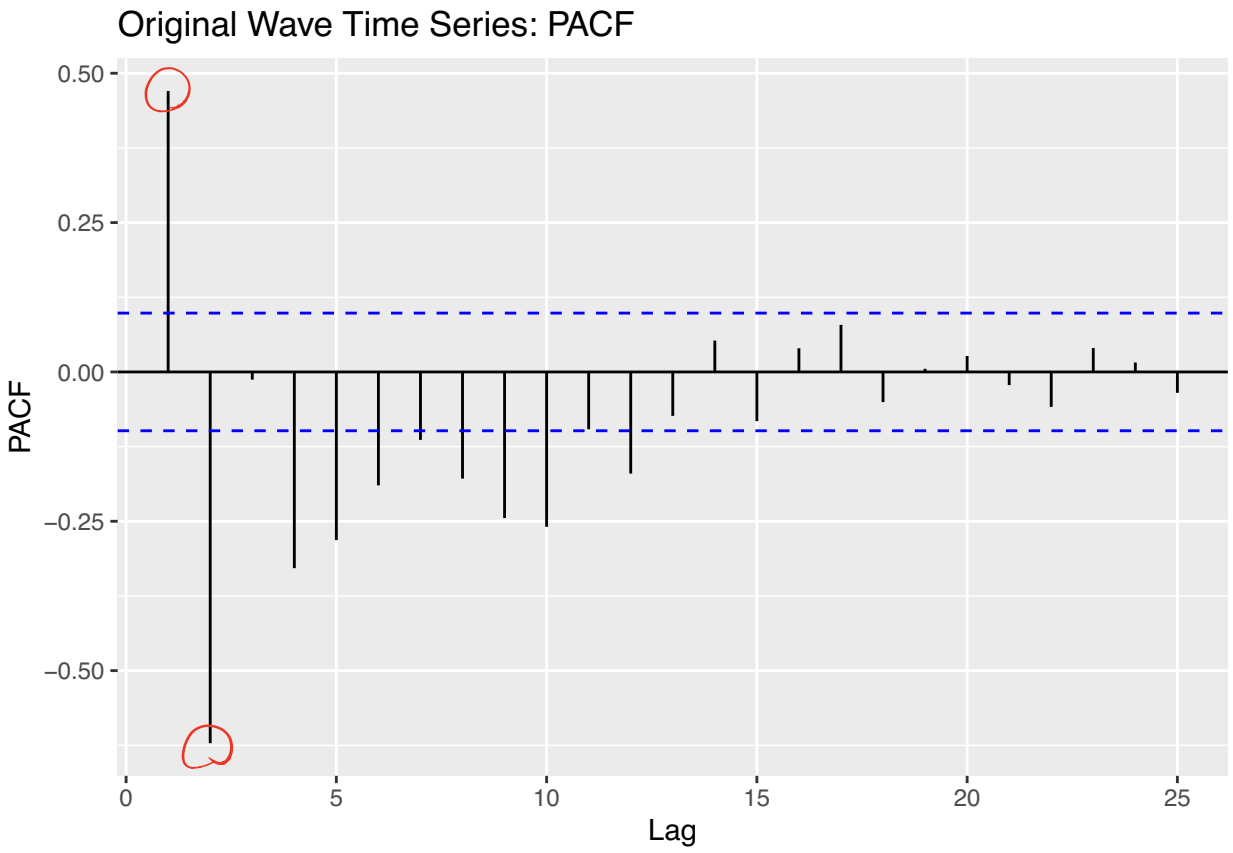
```
wave_ts = wave_data %>% pull(waveht) %>% ts(.,frequency=1)

autoplot(wave_ts)
```




```
ggAcf(wave_ts) + ggtitle("Original Wave Time Series: ACF")
```



```
ggPacf(wave_ts) + ggtitle("Original Wave Time Series: PACF")
```

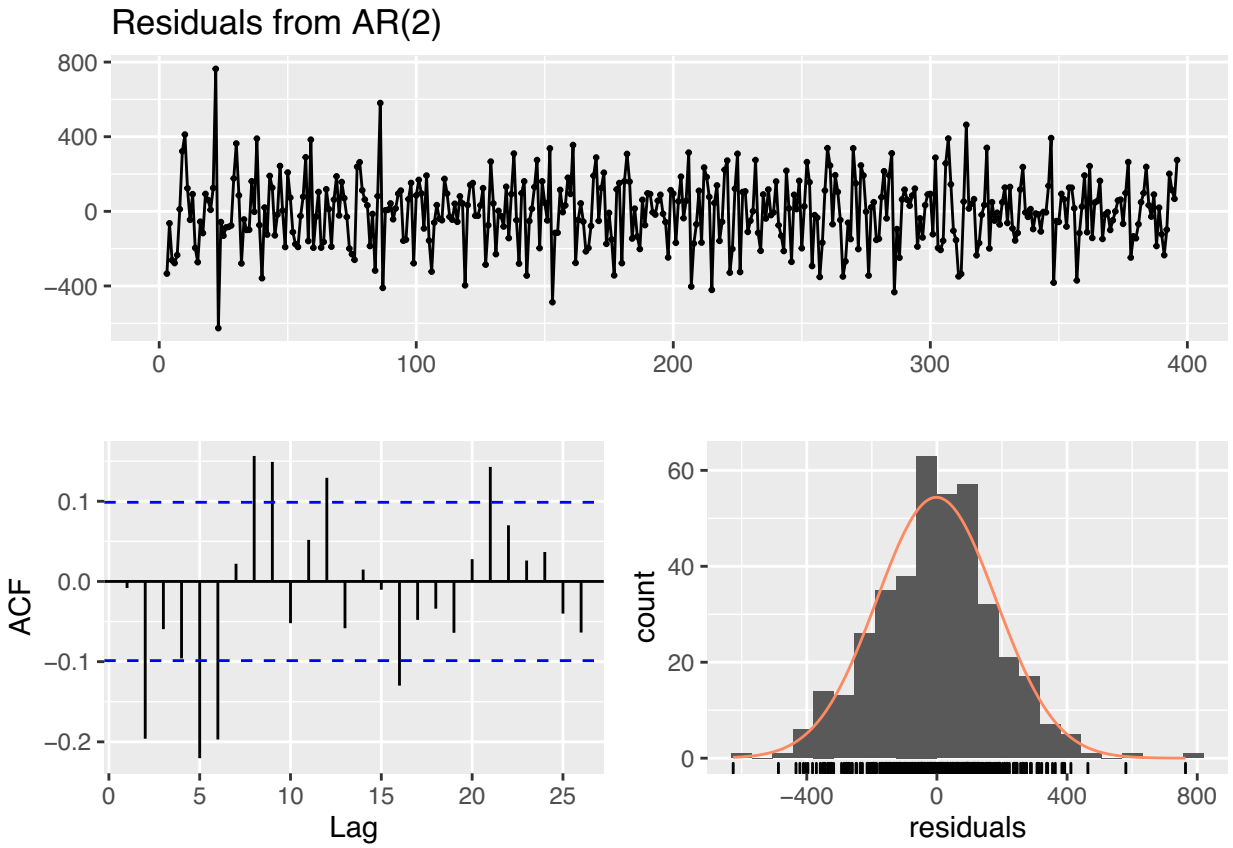



```
wave_ar2 = ar(wave_ts,order.max = 2)
wave_ar2$ar
```

```
[1]  0.7625158 -0.6214895
```

```
checkresiduals(wave_ar2, main="Residuals for AR(2) Model")
```

```
Warning in modeldf.default(object): Could not find appropriate degrees of
freedom for this model.
```

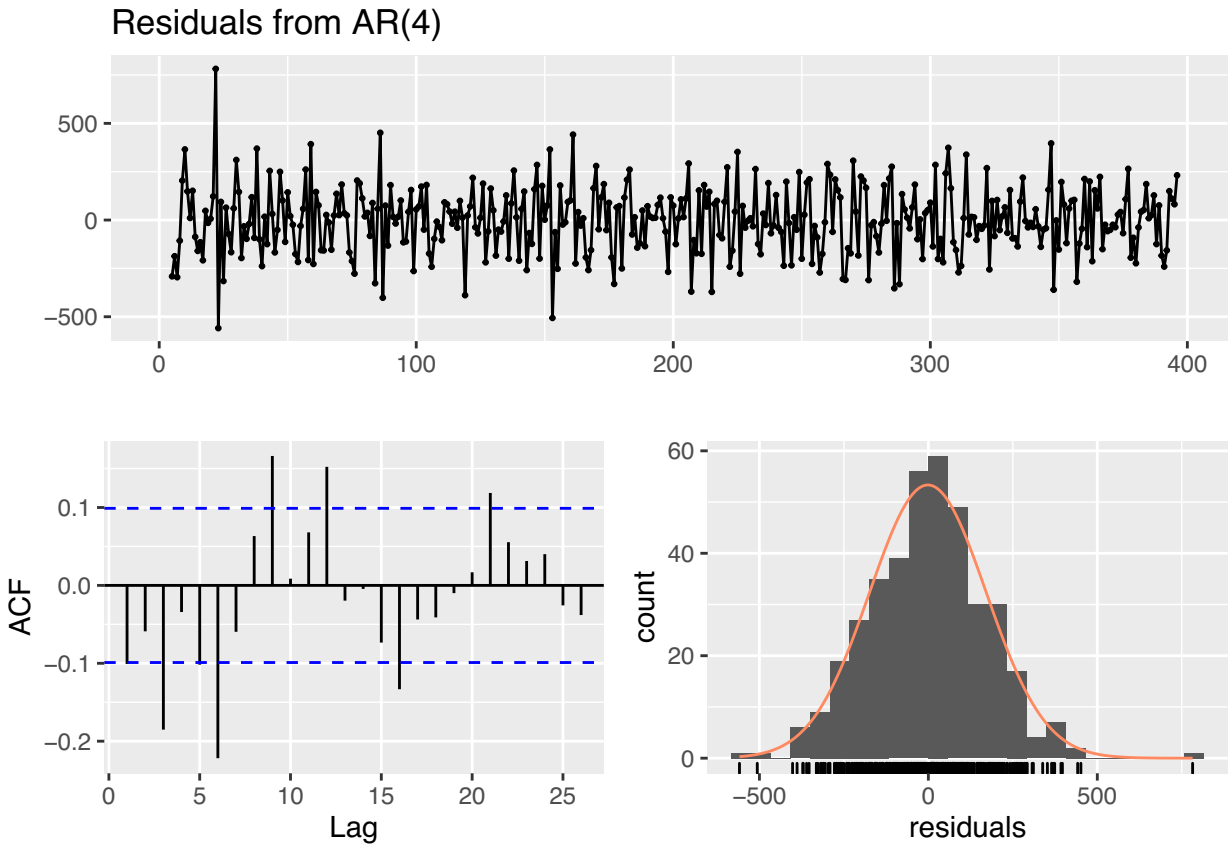



```
wave_ar4 = ar(wave_ts,order.max = 4)
wave_ar4$ar
```

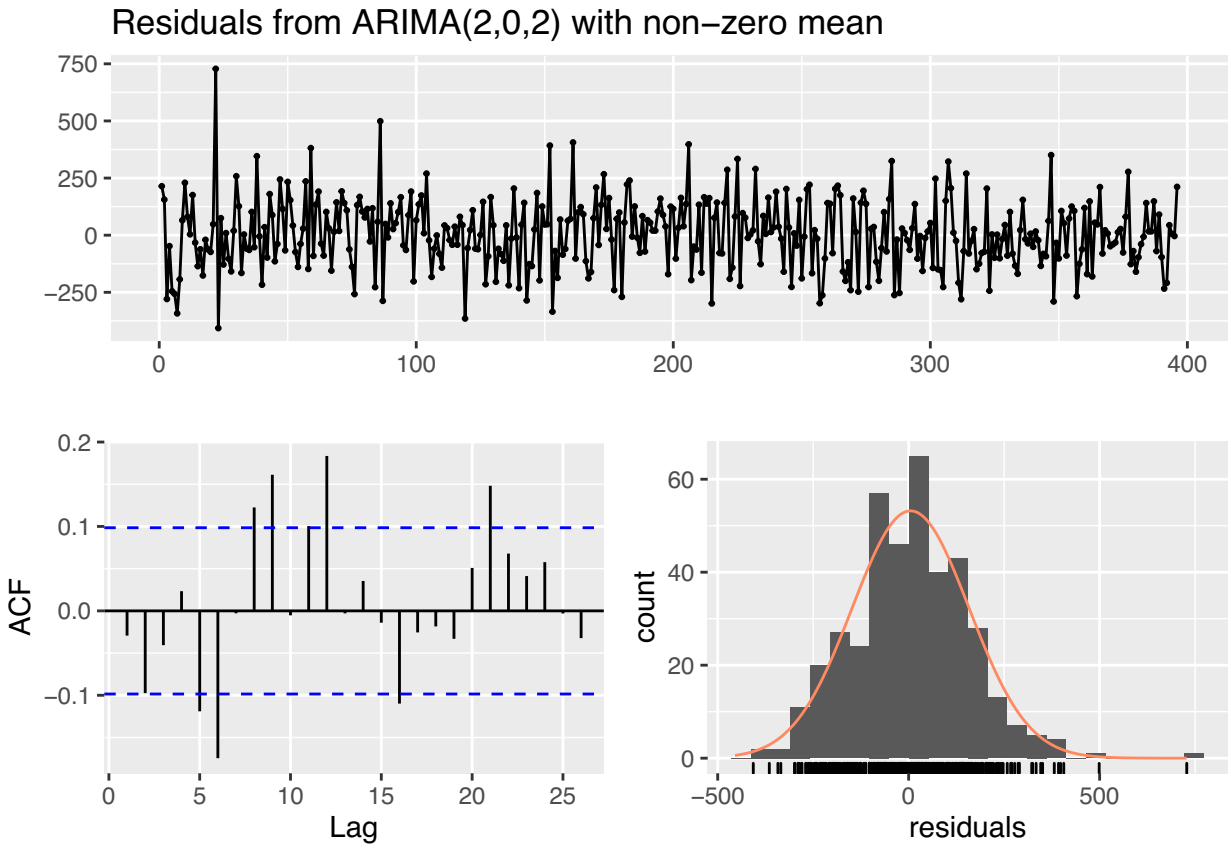
```
[1]  0.7502273 -0.8125464  0.2349117 -0.3285020
```

```
checkresiduals(wave_ar4, main="Residuals for AR(4) Model")
```

```
Warning in modeldf.default(object): Could not find appropriate degrees of
freedom for this model.
```

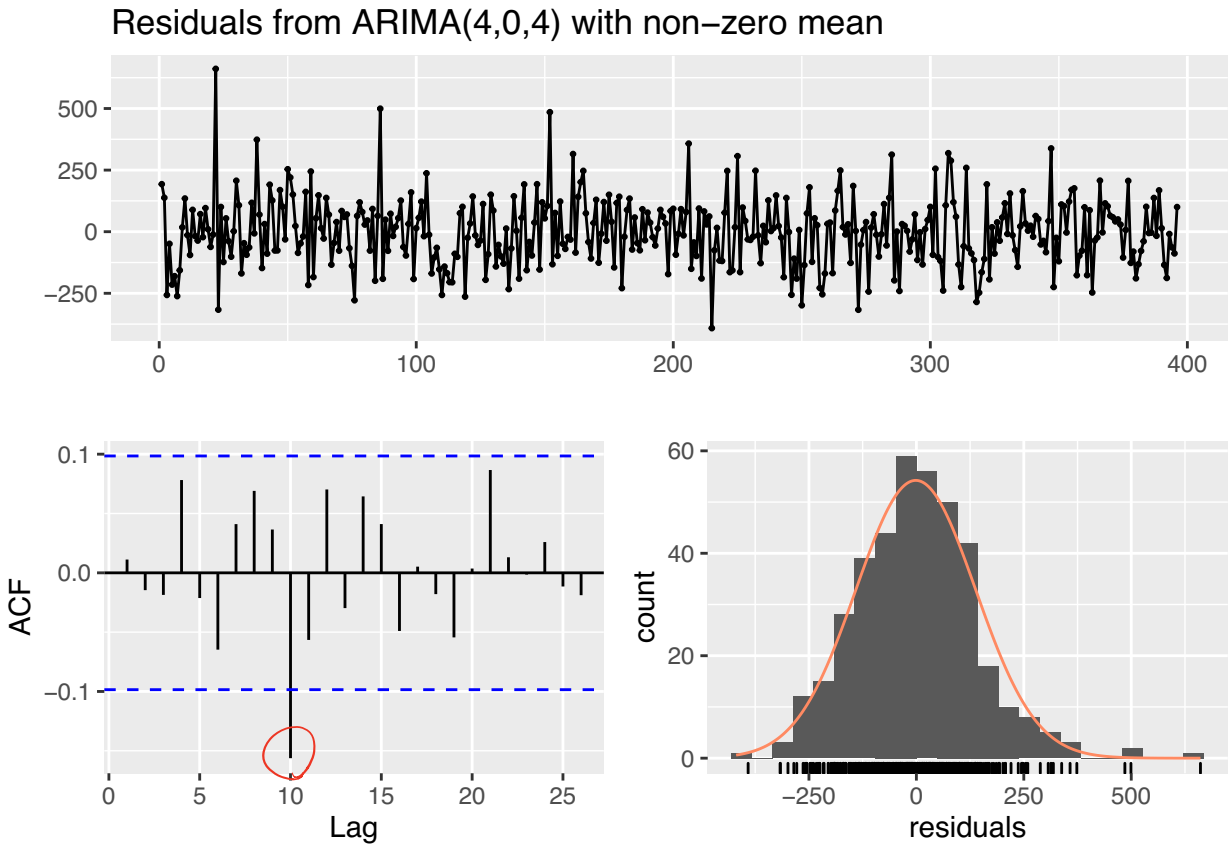



```
wave_arma22 = arima(wave_ts, order=c(2,0,2))
checkresiduals(wave_arma22, main="Residuals for ARMA(2,2) Model")
```




```
    Ljung-Box test

data:  Residuals from ARIMA(2,0,2) with non-zero mean
Q* = 39.73, df = 5, p-value = 1.692e-07

Model df: 5.   Total lags used: 10
```

```
wave_arma44 = arima(wave_ts, order=c(4,0,4))
checkresiduals(wave_arma44,lag=9, main="Residuals for ARMA(4,4) Model")
```

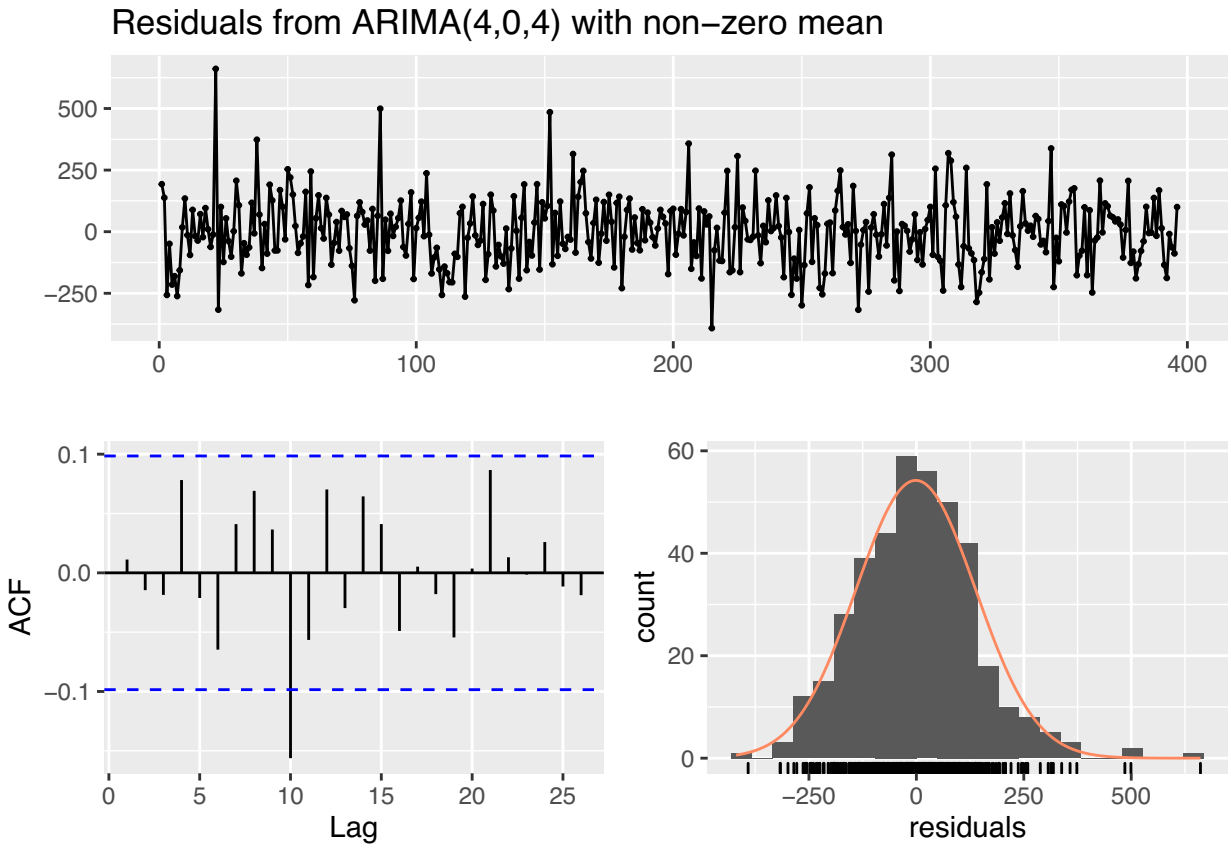



```
    Ljung-Box test

data:  Residuals from ARIMA(4,0,4) with non-zero mean
Q* = 7.7709, df = 0, p-value < 2.2e-16

Model df: 9.   Total lags used: 9
```

```
logLik(wave_arma44)
```

```
'log Lik.' -2519.419 (df=10)
```

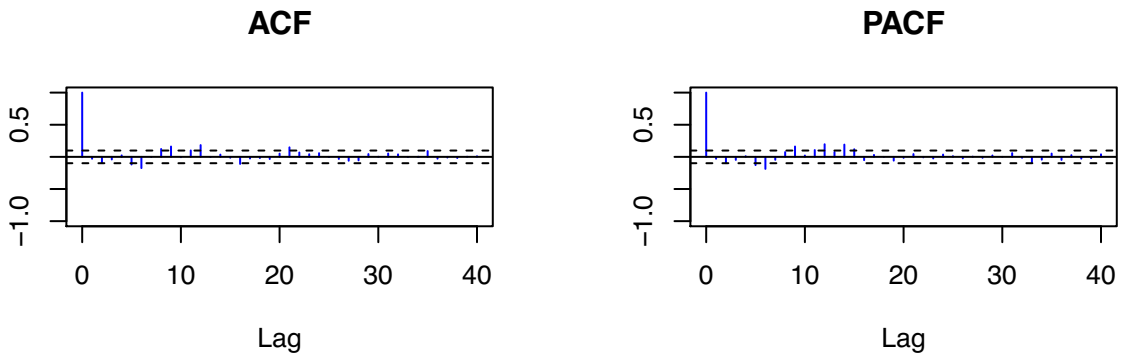
```
### With larger lag to get df for test
wave_arma44 = arima(wave_ts, order=c(4,0,4))
checkresiduals(wave_arma44,lag=20, main="Residuals for ARMA(4,4) Model")
```

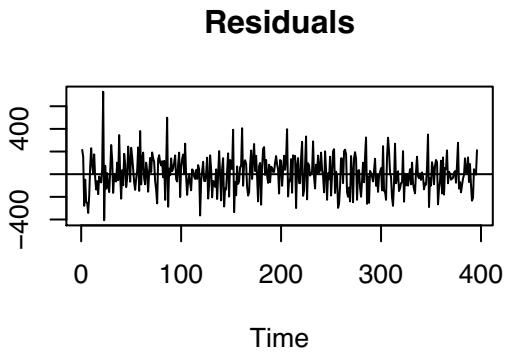


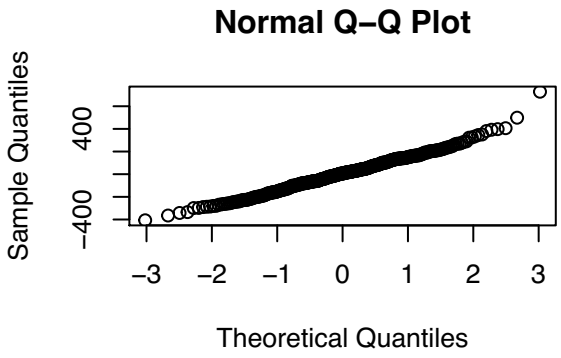

```
    Ljung-Box test

data:  Residuals from ARIMA(4,0,4) with non-zero mean
Q* = 26.231, df = 11, p-value = 0.005995

Model df: 9.   Total lags used: 20
```

```
test(resid(wave_arma22))
```

```
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)     65.23         0 *
McLeod-Li Q                Q ~ chisq(20)     25.24    0.1926
Turning points T  (T-262.7)/8.4 ~ N(0,1)       245    0.0348 *
Diff signs S      (S-197.5)/5.8 ~ N(0,1)       200    0.6638
Rank P         (P-39105)/1315.9 ~ N(0,1)     37552    0.2379
```








```
1-pchisq(65.23,20-5)
```

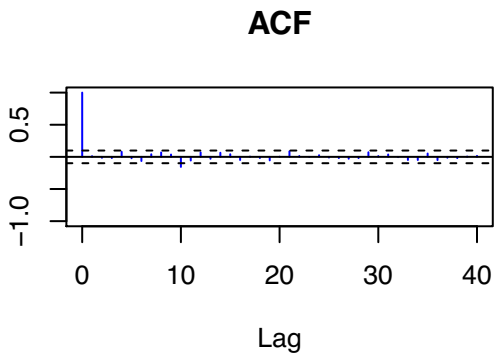
```
[1] 3.113963e-08
```

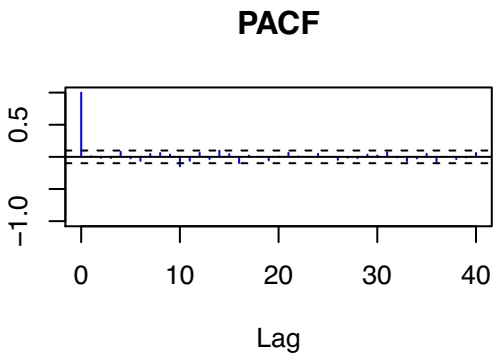
```
1-pchisq(25.24,20-5)
```

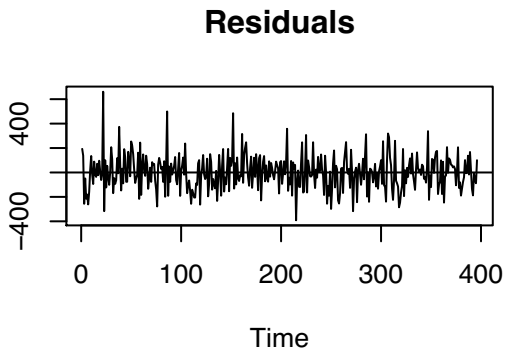
```
[1] 0.04681165
```

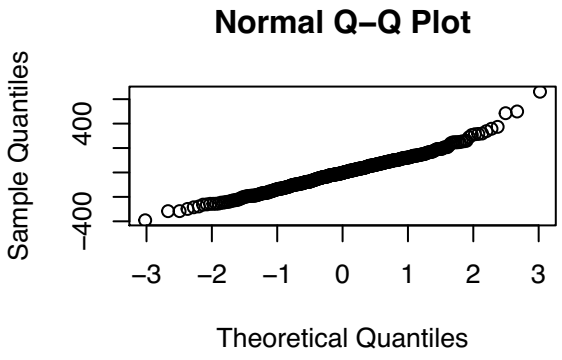
```
test(resid(wave_arma44))
```

```
Null hypothesis: Residuals are iid noise.
Test                         Distribution Statistic   p-value
Ljung-Box Q                 Q ~ chisq(20)     26.23    0.1583
McLeod-Li Q                 Q ~ chisq(20)      14.8    0.7875
Turning points T  (T-262.7)/8.4 ~ N(0,1)       247    0.0613
Diff signs S        (S-197.5)/5.8 ~ N(0,1)       201    0.5429
Rank P         (P-39105)/1315.9 ~ N(0,1)     38569    0.6838
```










```
1-pchisq(26.23,20-9)
```

```
[1] 0.005997302
```

```
1-pchisq(14.8,20-9)
```

```
[1] 0.1918373
```

```
wave_ar2_b = arima(wave_ts, order=c(2,0,0))
wave_ar4_b = arima(wave_ts, order=c(4,0,0))

AIC(wave_ar2_b)
```

```
[1] 5261.28
```

```
AIC(wave_ar4_b)
```

```
[1] 5219.029
```

```
AIC(wave_arma44)
```

```
[1] 5058.839
```

```
BIC(wave_ar2_b)
```

```
[1] 5277.206
```

```
BIC(wave_ar4_b)
```

```
[1] 5242.918
```

```
BIC(wave_arma44)
```

```
[1] 5098.653
```

Calculate $l(\underset{\sim}{x}; \hat{\underset{\sim}{\phi}}_m, \hat{\underset{\sim}{\theta}}_m, \hat{\sigma}^2)$

but it is an optimistic measure of fit

⇒ Choose with highest likelihood means choosing more hyperparameter

```
models = list(wave_ar2_b, wave_ar4_b,wave_arma22,wave_arma44)
names(models)=c("AR(2)","AR(4)","ARMA(2,2)", "ARMA(4,4)")
```

```
models %>% map(~logLik(.)) %>% AICtab(., mnames=names(models),base=TRUE)
```

```
          AIC    dAIC   df
ARMA(4,4) 5058.8    0.0 10
ARMA(2,2) 5127.3   68.5 6
AR(4)     5219.0  160.2 6
AR(2)     5261.3  202.4 4
```

```
models %>% map(~logLik(.)) %>% AICctab(., mnames=names(models),base=TRUE)
```

```
          AICc   dAICc  df
ARMA(4,4) 5059.4    0.0 10
ARMA(2,2) 5127.6   68.1 6
AR(4)     5219.2  159.8 6
AR(2)     5261.4  202.0 4
```

```
models %>% map(~logLik(.)) %>% BICtab(., mnames=names(models),base=TRUE)
```

```
          BIC    dBIC   df
ARMA(4,4) 5098.7    0.0 10
ARMA(2,2) 5151.2   52.6 6
AR(4)     5242.9  144.3 6
AR(2)     5277.2  178.6 4
```

```
wave_auto_select_AIC = auto.arima(wave_ts,stepwise=FALSE,seasonal=FALSE,ic="aic",trace=TRUE)
```

```
 Fitting models using approximations to speed things up...

 ARIMA(0,0,0) with zero mean     : 5548.674
 ARIMA(0,0,0) with non-zero mean : 5550.582
 ARIMA(0,0,1) with zero mean     : 5334.636
 ARIMA(0,0,1) with non-zero mean : 5336.596
 ARIMA(0,0,2) with zero mean     : 5316.739
 ARIMA(0,0,2) with non-zero mean : 5318.729
 ARIMA(0,0,3) with zero mean     : 5218.578
 ARIMA(0,0,3) with non-zero mean : Inf
 ARIMA(0,0,4) with zero mean     : 5185.9
 ARIMA(0,0,4) with non-zero mean : Inf
 ARIMA(0,0,5) with zero mean     : 5187.523
 ARIMA(0,0,5) with non-zero mean : Inf
 ARIMA(1,0,0) with zero mean     : 5449.941
 ARIMA(1,0,0) with non-zero mean : 5451.884
 ARIMA(1,0,1) with zero mean     : 5333.608
 ARIMA(1,0,1) with non-zero mean : 5335.598
 ARIMA(1,0,2) with zero mean     : 5326.179
 ARIMA(1,0,2) with non-zero mean : 5328.157
 ARIMA(1,0,3) with zero mean     : 5212.332
 ARIMA(1,0,3) with non-zero mean : 5202.621
 ARIMA(1,0,4) with zero mean     : 5202.095
 ARIMA(1,0,4) with non-zero mean : 5189.405
```

```
 ARIMA(2,0,0) with zero mean     : 5257.361
 ARIMA(2,0,0) with non-zero mean : 5259.053
 ARIMA(2,0,1) with zero mean     : 5203.155
 ARIMA(2,0,1) with non-zero mean : 5195.208
 ARIMA(2,0,2) with zero mean     : 5171.268
 ARIMA(2,0,2) with non-zero mean : 5163.619
 ARIMA(2,0,3) with zero mean     : 5156.341
 ARIMA(2,0,3) with non-zero mean : 5151.446
 ARIMA(3,0,0) with zero mean     : 5256.836
 ARIMA(3,0,0) with non-zero mean : 5258.614
 ARIMA(3,0,1) with zero mean     : 5256.42
 ARIMA(3,0,1) with non-zero mean : 5258.171
 ARIMA(3,0,2) with zero mean     : 5141.022
 ARIMA(3,0,2) with non-zero mean : Inf
 ARIMA(4,0,0) with zero mean     : 5213.393
 ARIMA(4,0,0) with non-zero mean : 5214.895
 ARIMA(4,0,1) with zero mean     : 5111.274
 ARIMA(4,0,1) with non-zero mean : 5090.584
 ARIMA(5,0,0) with zero mean     : 5178.428
 ARIMA(5,0,0) with non-zero mean : 5179.615

 Now re-fitting the best model(s) without approximations...



 Best model: ARIMA(4,0,1) with non-zero mean
```

```
wave_auto_select_AIC
```

```
Series: wave_ts
ARIMA(4,0,1) with non-zero mean

Coefficients:
         ar1      ar2     ar3      ar4      ma1     mean
      1.3460  -1.3009  0.6457  -0.3901  -0.9465  -5.0761
s.e.  0.0477   0.0756  0.0759   0.0479   0.0170   0.5925

sigma^2 estimated as 21740:  log likelihood=-2538.54
AIC=5091.08   AICc=5091.37   BIC=5118.95
```

```
wave_auto_select_BIC = auto.arima(wave_ts,stepwise=FALSE,seasonal=FALSE,ic="bic",trace=TRUE)
```

```
 Fitting models using approximations to speed things up...

 ARIMA(0,0,0) with zero mean     : 5552.656
 ARIMA(0,0,0) with non-zero mean : 5558.544
 ARIMA(0,0,1) with zero mean     : 5342.599
 ARIMA(0,0,1) with non-zero mean : 5348.54
 ARIMA(0,0,2) with zero mean     : 5328.684
 ARIMA(0,0,2) with non-zero mean : 5334.655
 ARIMA(0,0,3) with zero mean     : 5234.503
 ARIMA(0,0,3) with non-zero mean : Inf
 ARIMA(0,0,4) with zero mean     : 5205.807
```

```
 ARIMA(0,0,4) with non-zero mean : Inf
 ARIMA(0,0,5) with zero mean     : 5211.412
 ARIMA(0,0,5) with non-zero mean : Inf
 ARIMA(1,0,0) with zero mean     : 5457.903
 ARIMA(1,0,0) with non-zero mean : 5463.828
 ARIMA(1,0,1) with zero mean     : 5345.552
 ARIMA(1,0,1) with non-zero mean : 5351.524
 ARIMA(1,0,2) with zero mean     : 5342.104
 ARIMA(1,0,2) with non-zero mean : 5348.064
 ARIMA(1,0,3) with zero mean     : 5232.239
 ARIMA(1,0,3) with non-zero mean : 5226.509
 ARIMA(1,0,4) with zero mean     : 5225.984
 ARIMA(1,0,4) with non-zero mean : 5217.275
 ARIMA(2,0,0) with zero mean     : 5269.305
 ARIMA(2,0,0) with non-zero mean : 5274.978
 ARIMA(2,0,1) with zero mean     : 5219.081
 ARIMA(2,0,1) with non-zero mean : 5215.115
 ARIMA(2,0,2) with zero mean     : 5191.175
 ARIMA(2,0,2) with non-zero mean : 5187.507
 ARIMA(2,0,3) with zero mean     : 5180.23
 ARIMA(2,0,3) with non-zero mean : 5179.316
 ARIMA(3,0,0) with zero mean     : 5272.762
 ARIMA(3,0,0) with non-zero mean : 5278.521
 ARIMA(3,0,1) with zero mean     : 5276.327
 ARIMA(3,0,1) with non-zero mean : 5282.059
 ARIMA(3,0,2) with zero mean     : 5164.91
 ARIMA(3,0,2) with non-zero mean : Inf
 ARIMA(4,0,0) with zero mean     : 5233.3
 ARIMA(4,0,0) with non-zero mean : 5238.784
 ARIMA(4,0,1) with zero mean     : 5135.163
 ARIMA(4,0,1) with non-zero mean : 5118.454
 ARIMA(5,0,0) with zero mean     : 5202.317
 ARIMA(5,0,0) with non-zero mean : 5207.485

 Now re-fitting the best model(s) without approximations...




 Best model: ARIMA(4,0,1) with non-zero mean
```

```
wave_auto_select_BIC
```

```
Series: wave_ts
ARIMA(4,0,1) with non-zero mean

Coefficients:
         ar1      ar2     ar3      ar4      ma1     mean
      1.3460  -1.3009  0.6457  -0.3901  -0.9465  -5.0761
s.e.  0.0477   0.0756  0.0759   0.0479   0.0170   0.5925

sigma^2 estimated as 21740:  log likelihood=-2538.54
AIC=5091.08   AICc=5091.37   BIC=5118.95
```

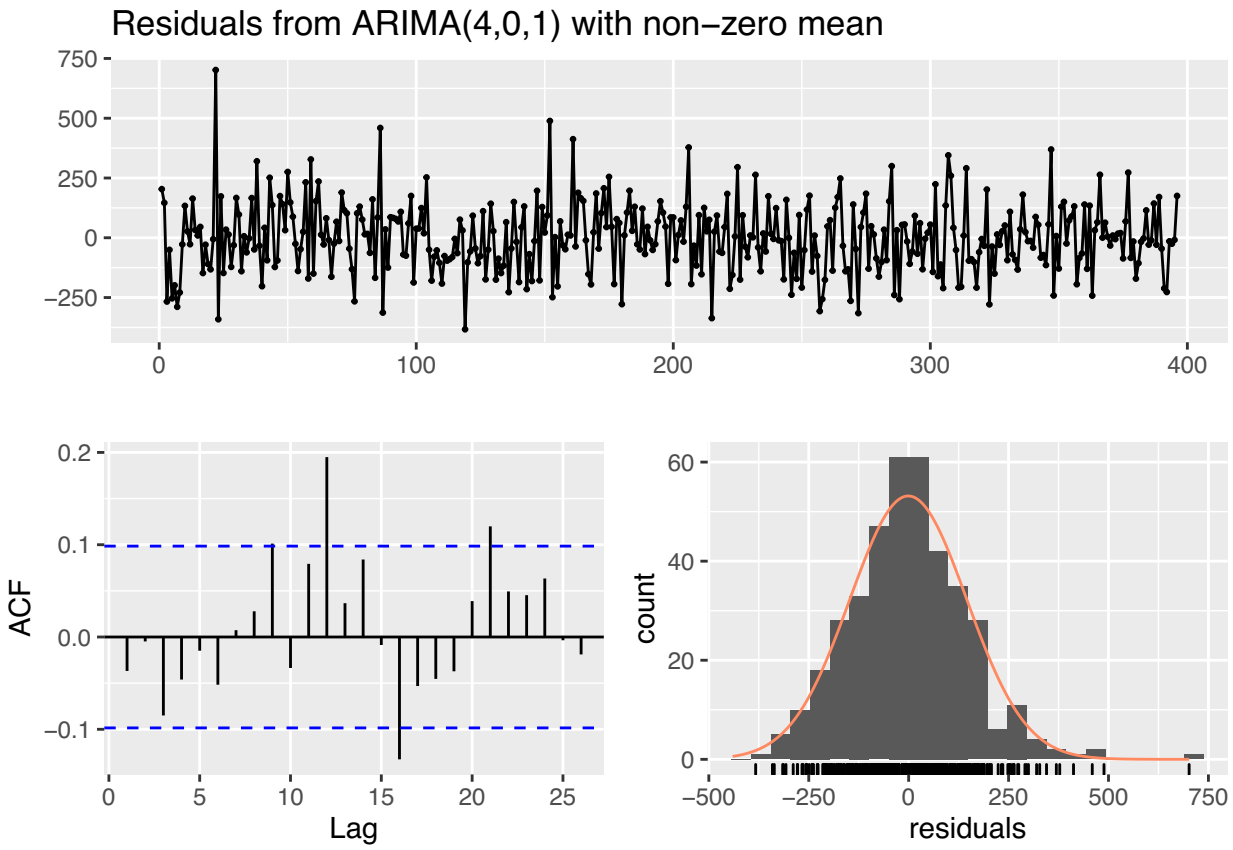
```
wave_auto_select_AICC = auto.arima(wave_ts,stepwise=FALSE,seasonal=FALSE,ic="aicc",trace=TRUE)
```

```
 Fitting models using approximations to speed things up...

 ARIMA(0,0,0) with zero mean     : 5548.684
 ARIMA(0,0,0) with non-zero mean : 5550.612
 ARIMA(0,0,1) with zero mean     : 5334.667
 ARIMA(0,0,1) with non-zero mean : 5336.657
 ARIMA(0,0,2) with zero mean     : 5316.801
 ARIMA(0,0,2) with non-zero mean : 5318.831
 ARIMA(0,0,3) with zero mean     : 5218.68
 ARIMA(0,0,3) with non-zero mean : Inf
 ARIMA(0,0,4) with zero mean     : 5186.054
 ARIMA(0,0,4) with non-zero mean : Inf
 ARIMA(0,0,5) with zero mean     : 5187.739
 ARIMA(0,0,5) with non-zero mean : Inf
 ARIMA(1,0,0) with zero mean     : 5449.971
 ARIMA(1,0,0) with non-zero mean : 5451.945
 ARIMA(1,0,1) with zero mean     : 5333.669
 ARIMA(1,0,1) with non-zero mean : 5335.701
 ARIMA(1,0,2) with zero mean     : 5326.281
 ARIMA(1,0,2) with non-zero mean : 5328.311
 ARIMA(1,0,3) with zero mean     : 5212.486
 ARIMA(1,0,3) with non-zero mean : 5202.837
 ARIMA(1,0,4) with zero mean     : 5202.311
 ARIMA(1,0,4) with non-zero mean : 5189.694
 ARIMA(2,0,0) with zero mean     : 5257.422
 ARIMA(2,0,0) with non-zero mean : 5259.155
 ARIMA(2,0,1) with zero mean     : 5203.257
 ARIMA(2,0,1) with non-zero mean : 5195.362
 ARIMA(2,0,2) with zero mean     : 5171.422
 ARIMA(2,0,2) with non-zero mean : 5163.835
 ARIMA(2,0,3) with zero mean     : 5156.557
 ARIMA(2,0,3) with non-zero mean : 5151.735
 ARIMA(3,0,0) with zero mean     : 5256.938
 ARIMA(3,0,0) with non-zero mean : 5258.768
 ARIMA(3,0,1) with zero mean     : 5256.574
 ARIMA(3,0,1) with non-zero mean : 5258.387
 ARIMA(3,0,2) with zero mean     : 5141.238
 ARIMA(3,0,2) with non-zero mean : Inf
 ARIMA(4,0,0) with zero mean     : 5213.547
 ARIMA(4,0,0) with non-zero mean : 5215.111
 ARIMA(4,0,1) with zero mean     : 5111.49
 ARIMA(4,0,1) with non-zero mean : 5090.873
 ARIMA(5,0,0) with zero mean     : 5178.644
 ARIMA(5,0,0) with non-zero mean : 5179.903

 Now re-fitting the best model(s) without approximations...



 Best model: ARIMA(4,0,1) with non-zero mean
```

```
wave_auto_select_AICC
```

```
Series: wave_ts
ARIMA(4,0,1) with non-zero mean

Coefficients:
         ar1      ar2     ar3      ar4      ma1     mean
      1.3460  -1.3009  0.6457  -0.3901  -0.9465  -5.0761
s.e.  0.0477   0.0756  0.0759   0.0479   0.0170   0.5925

sigma^2 estimated as 21740:  log likelihood=-2538.54
AIC=5091.08   AICc=5091.37   BIC=5118.95
```

```
checkresiduals(wave_auto_select_AIC)
```




```
    Ljung-Box test

data:  Residuals from ARIMA(4,0,1) with non-zero mean
Q* = 10.436, df = 4, p-value = 0.03369

Model df: 6.   Total lags used: 10
```