ARMA estimation

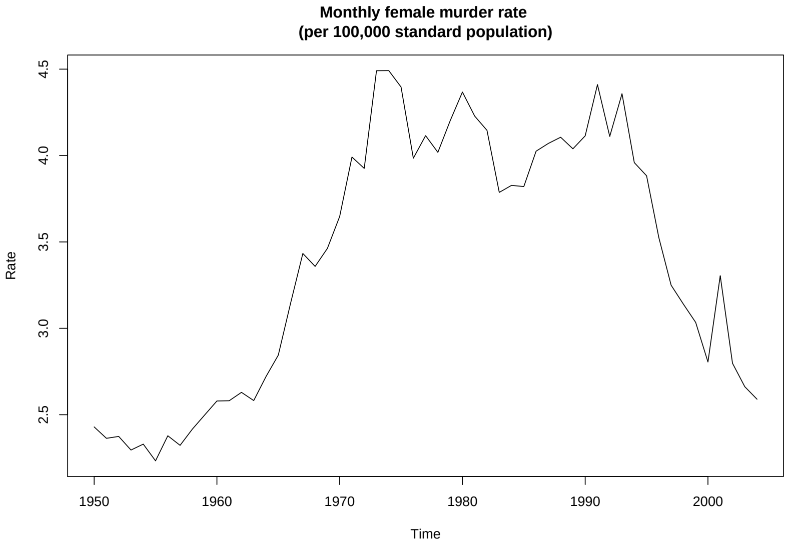
Giovani Valdrighi, Vitória Guardieiro

05/11/2020

Modelling process

- Plot the data and look for patterns.
- If necessary, remove seasonality.
- If necessary, use BoxCox transformation to stabilize the variance.
- If necessary, difference the data until is stationary. Use of Dickey Fuller test.
- Plot the ACF and PACF to identify the model order, p and q for $ARMA(p, q)$.
- Compare identified models, chose the one that minimize the AIC
- Analysis of the residuals of the model, with ACF and histogram.
- If the residuals look like white noise, make forecasts.

WMurders


Monthly female murder rate
(per 100,000 standard population)
Rate
4.5
4.0
3.5
3.0
2.5
1950
1960
1970
1980
1990
2000
Time

Variance stabilization and stationarity

- There is no need to stabilize data, we can look at the tendency with Dick-Fulley.

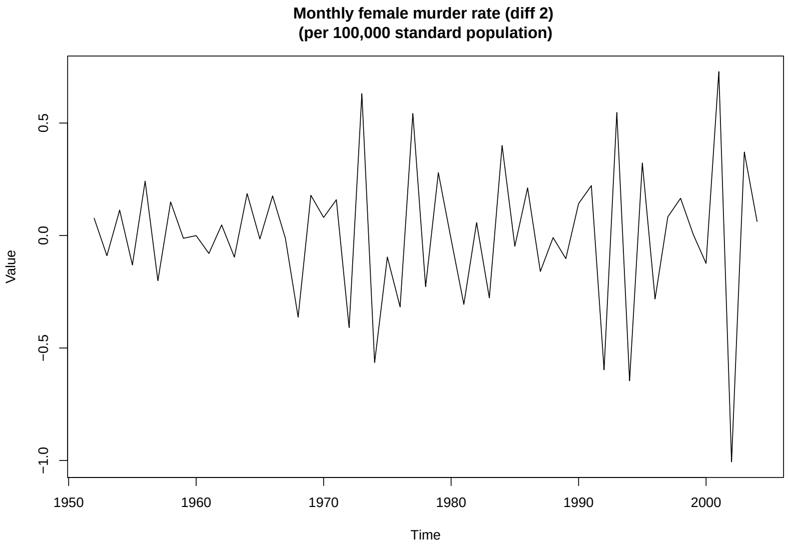
```
##
##  Augmented Dickey-Fuller Test
##
## data:  data
## Dickey-Fuller = -0.29243, Lag order = 3, p-value = 0.987
## alternative hypothesis: stationary
```

- Data is not stationary, we are going to test with one and two differences.

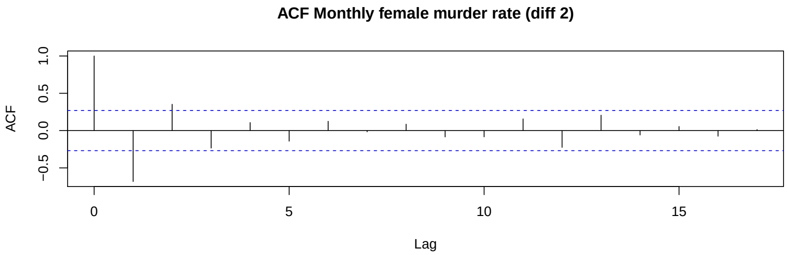
```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(data)
## Dickey-Fuller = -3.7688, Lag order = 3, p-value = 0.0272
## alternative hypothesis: stationary
```

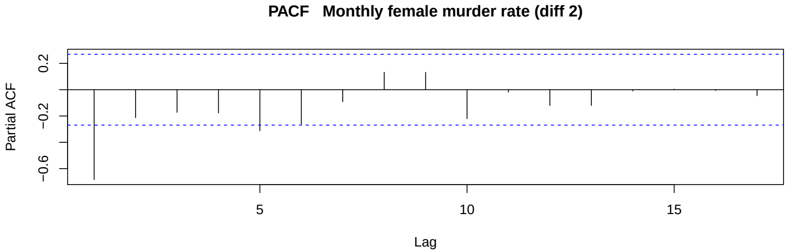
```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(data, differences = 2)
## Dickey-Fuller = -5.1646, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

- Diff 1 P-value: 0.02726
- Diff 2 P-value: less than 0.01 <- choosen.


Monthly female murder rate (diff 2)
(per 100,000 standard population)
Value
0.5
0.0
−0.5
−1.0
1950
1960
1970
1980
1990
2000
Time

ACF and PACF

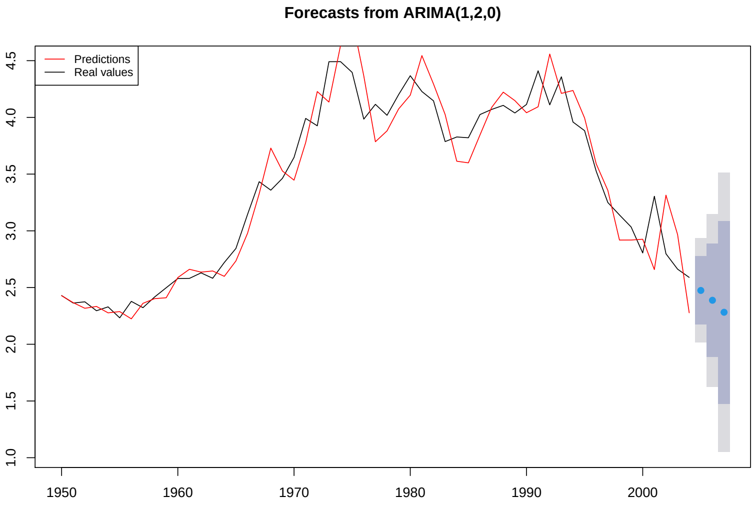




- Looks like AR(1) because the first spike of the PACF. The models that will be tested are AR(1), MA(1), ARMA(1, 1).

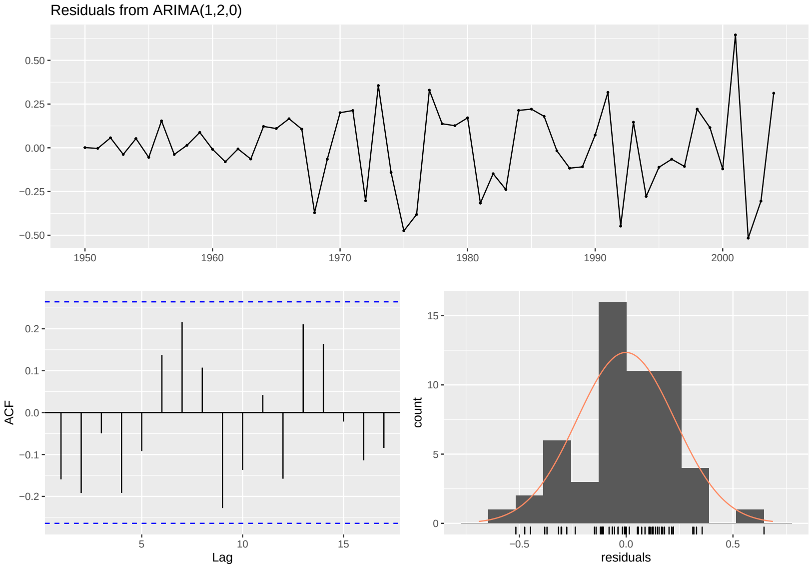
AR(1)

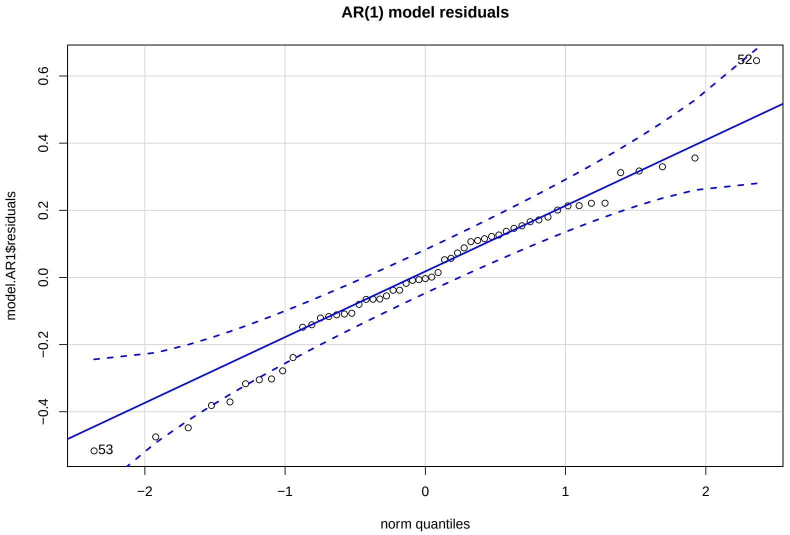
```
## Series: data
## ARIMA(1,2,0)
##
## Coefficients:
##           ar1
##       -0.6719
## s.e.   0.0981
##
## sigma^2 estimated as 0.05471:  log likelihood=2
## AIC=0   AICc=0.24   BIC=3.94
##
## Training set error measures:
##                        ME      RMSE       MAE         MP
## Training set -0.001376898 0.2274352 0.1777919 0.0014867
##                    ACF1
## Training set -0.1593845
```



Forecasts from ARIMA(1,2,0)
Predictions
Real values
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
1950
1960
1970
1980
1990
2000

Residuals analysis





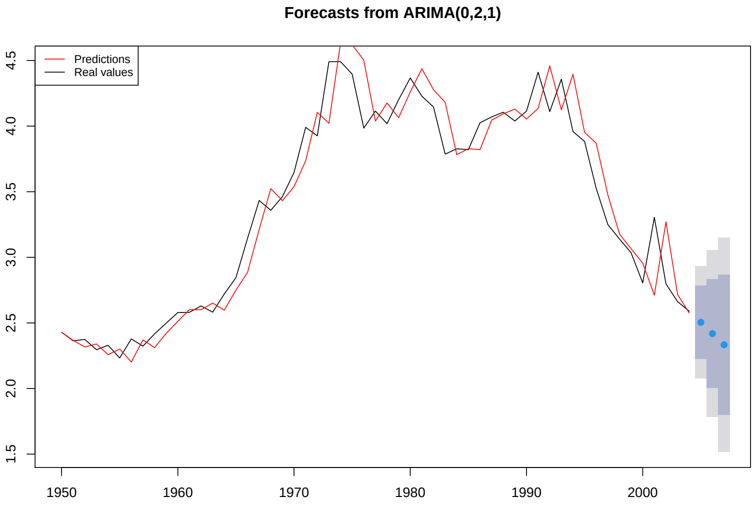
AR(1) model residuals
model.AR1$residuals
norm quantiles
0.6
0.4
0.2
0.0
−0.2
−0.4
−2
−1
0
1
2
52
53


```
## [1] 52 53
```

```
##
##  Jarque Bera Test
##
## data:  model.AR1$residuals
## X-squared = 0.24457, df = 2, p-value = 0.8849
```

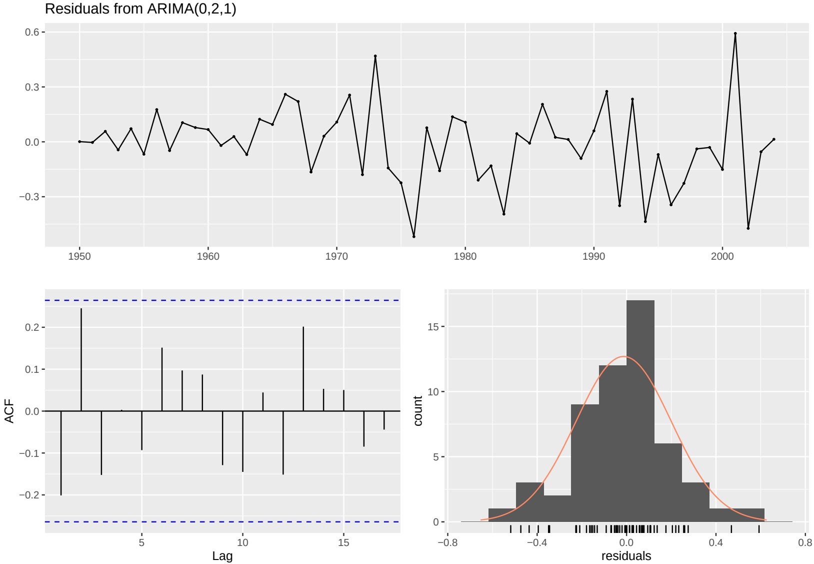
MA(1)

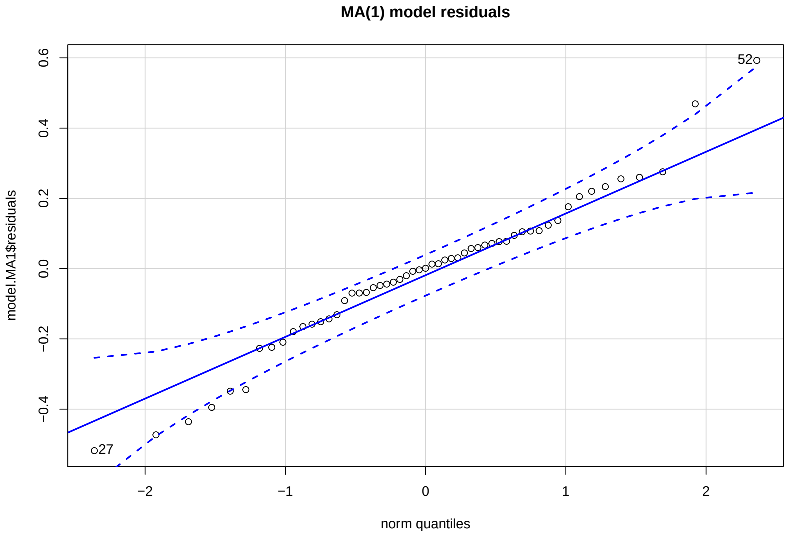
```
## Series: data
## ARIMA(0,2,1)
##
## Coefficients:
##           ma1
##       -0.8995
## s.e.   0.0669
##
## sigma^2 estimated as 0.04747:  log likelihood=5.24
## AIC=-6.48   AICc=-6.24   BIC=-2.54
##
## Training set error measures:
##                       ME      RMSE       MAE        MPE
## Training set -0.01306101 0.2118445 0.1559694 -0.3151353
##                    ACF1
## Training set -0.2011523
```



Forecasts from ARIMA(0,2,1)
Predictions
Real values
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1950
1960
1970
1980
1990
2000

Residuals analysis



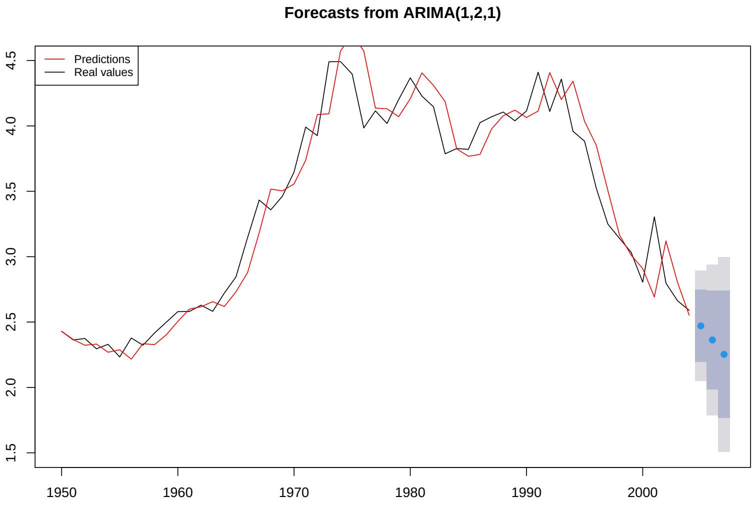

MA(1) model residuals
model.MA1$residuals
norm quantiles
0.6
0.4
0.2
0.0
−0.2
−0.4
−2
−1
0
1
2
52
27

```
## [1] 52 27
```

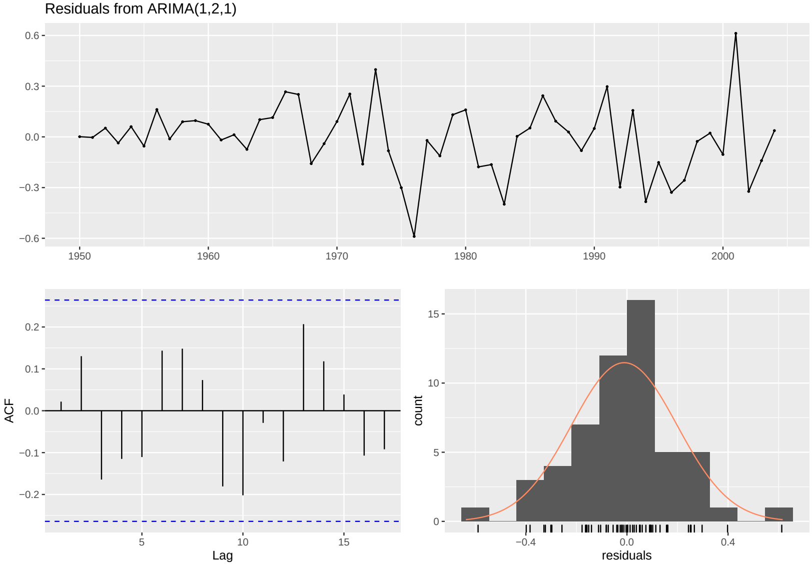
```
##
##  Jarque Bera Test
##
## data:  model.MA1$residuals
## X-squared = 1.3749, df = 2, p-value = 0.5028
```

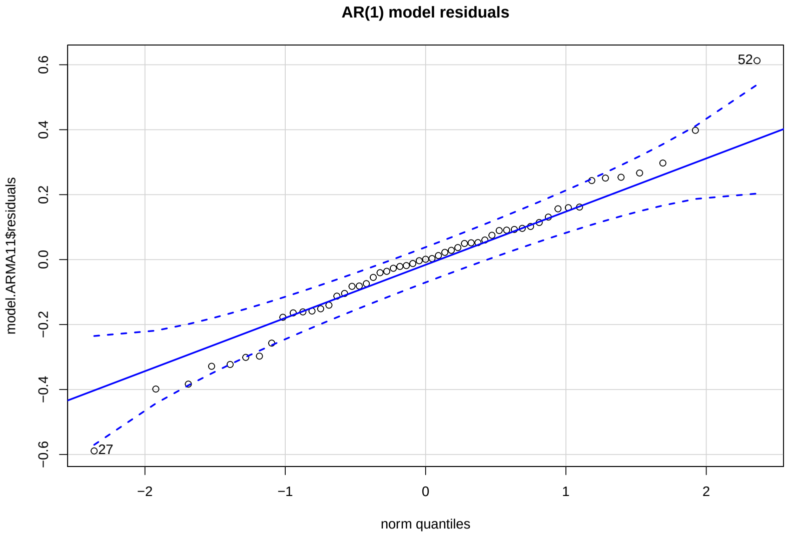
ARMA(1, 1)

```
## Series: data
## ARIMA(1,2,1)
##
## Coefficients:
##           ar1      ma1
##       -0.2434  -0.8261
## s.e.   0.1553   0.1143
##
## sigma^2 estimated as 0.04632:  log likelihood=6.44
## AIC=-6.88   AICc=-6.39   BIC=-0.97
##
## Training set error measures:
##                       ME      RMSE       MAE        MPE
## Training set -0.01065956 0.2072523 0.1528734 -0.2149476
##                    ACF1
## Training set 0.02176343
```


Forecasts from ARIMA(1,2,1)
Predictions
Real values
1.5
2.0
2.5
3.0
3.5
4.0
4.5
1950
1960
1970
1980
1990
2000

Residuals analysis



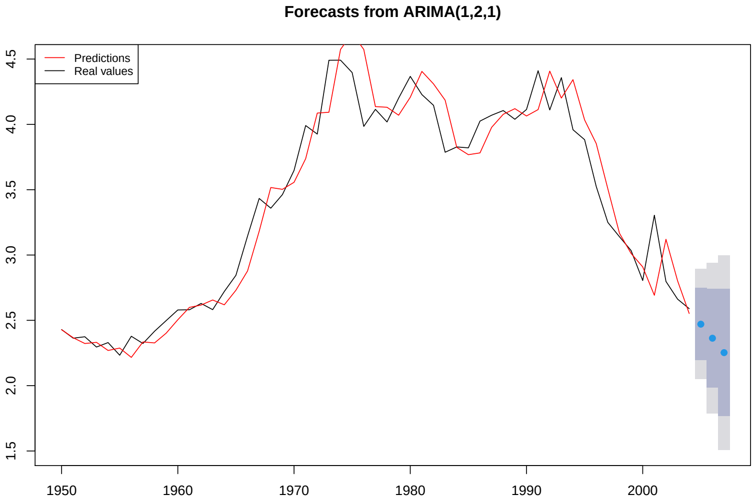

AR(1) model residuals
model.ARMA11$residuals
norm quantiles
52
27
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−2
−1
0
1
2

```
## [1] 52 27
```

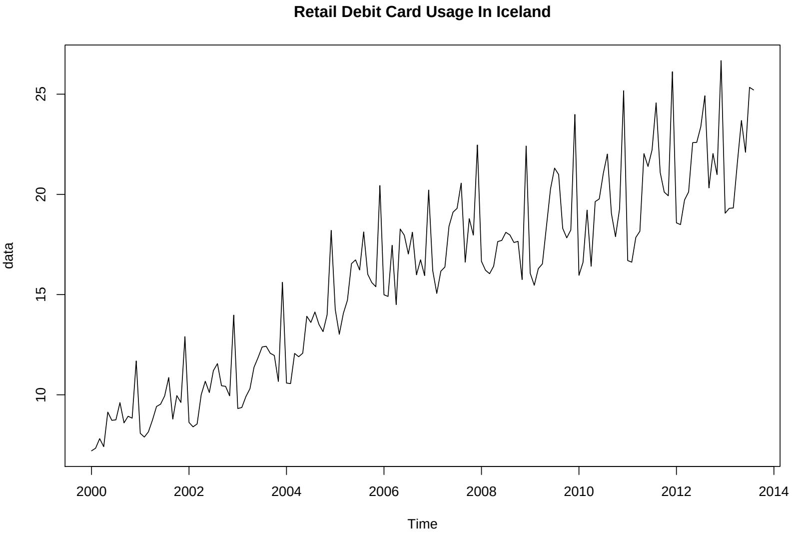
```
##
##  Jarque Bera Test
##
## data:  model.ARMA11$residuals
## X-squared = 2.3784, df = 2, p-value = 0.3045
```

Model selection

- The model with lowest AIC is the $ARMA(1,1)$. The final forecast is:

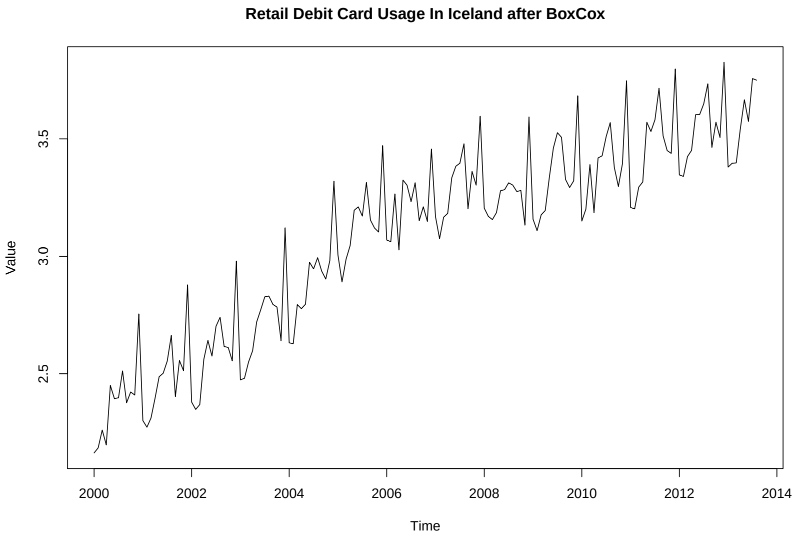


Debit Cards



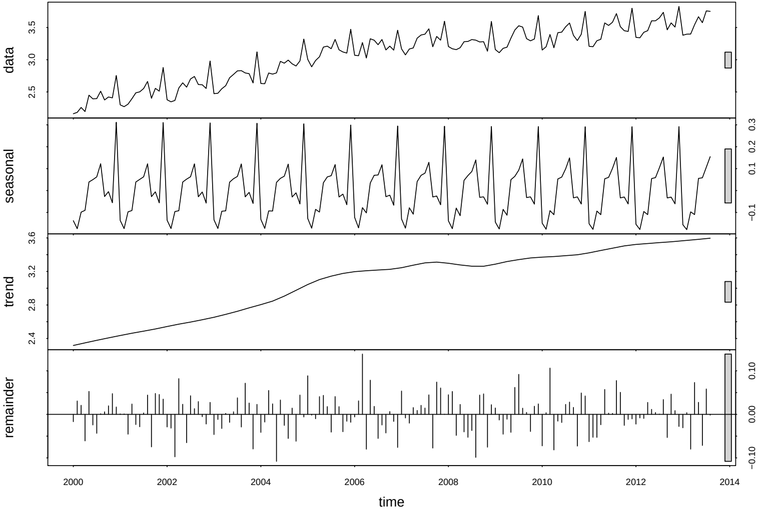
Retail Debit Card Usage In Iceland
data
10
15
20
25
Time
2000
2002
2004
2006
2008
2010
2012
2014

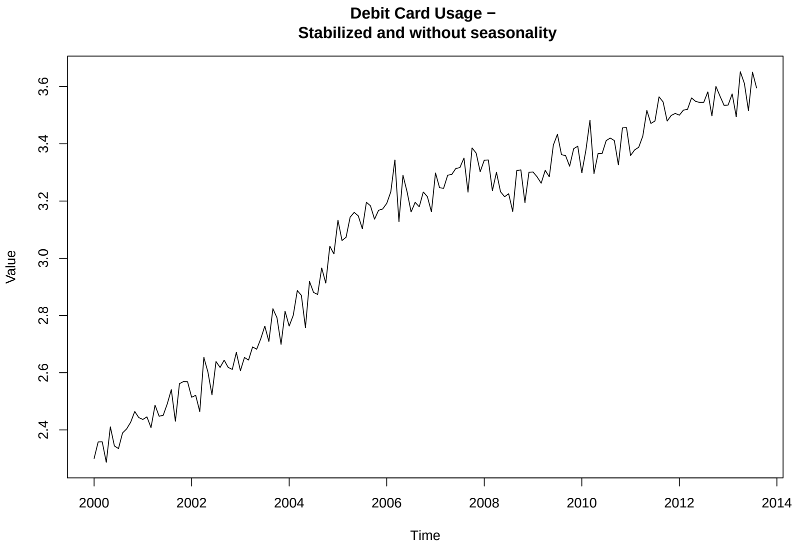
Stabilize variance



Remove seasonality

- Removing with STL (Lowess fit).



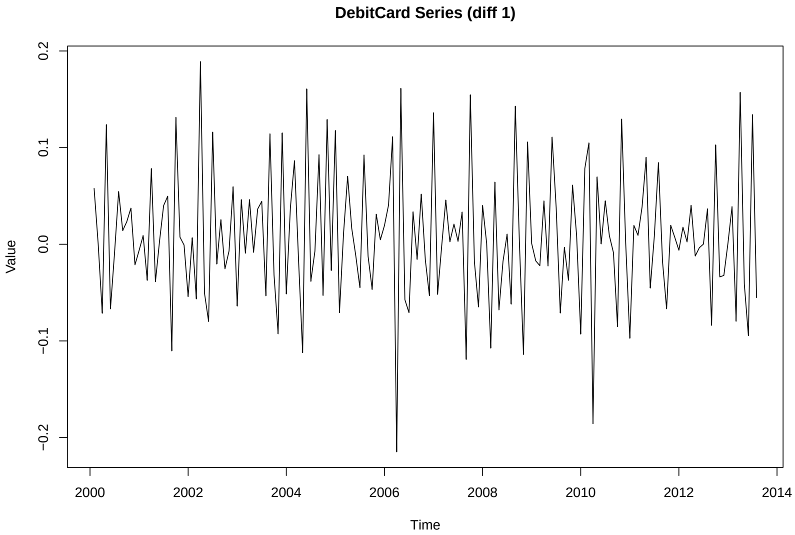

Debit Card Usage −
Stabilized and without seasonality
Value
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2000
2002
2004
2006
2008
2010
2012
2014
Time

Stationarity

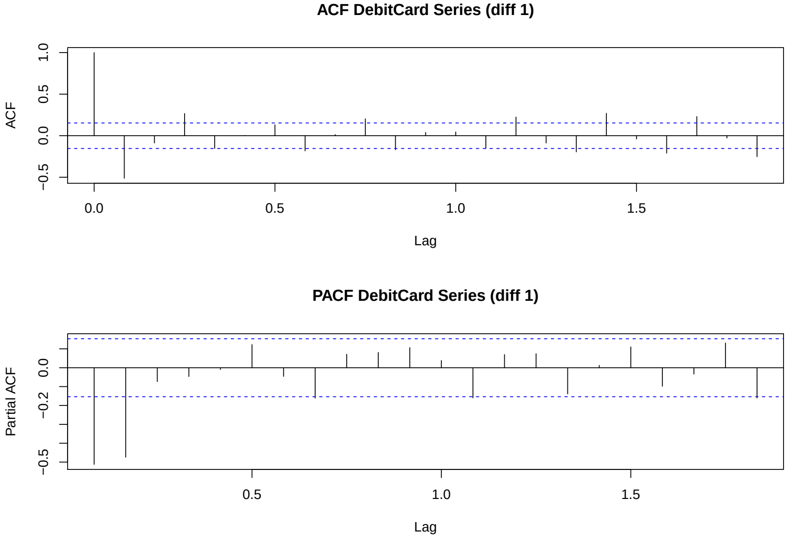
- Dickey Fuller test with series and differenced series.

```
##
##  Augmented Dickey-Fuller Test
##
## data:  data.trend
## Dickey-Fuller = -1.2005, Lag order = 5, p-value = 0.9038
## alternative hypothesis: stationary

##
##  Augmented Dickey-Fuller Test
##
## data:  diff(data.trend)
## Dickey-Fuller = -5.5839, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```


DebitCard Series (diff 1)
Value
Time
0.2
0.1
0.0
−0.1
−0.2
2000
2002
2004
2006
2008
2010
2012
2014

ACF and PACF

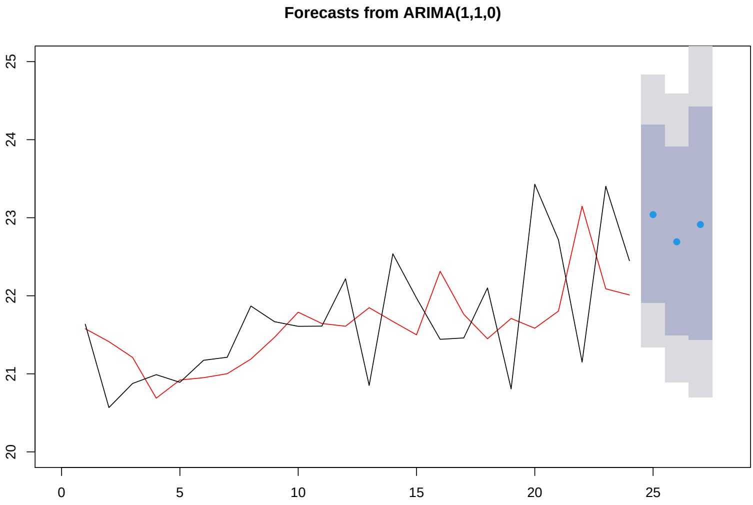


- Looks like AR(2) because of the first two lags. Models being tested are AR(1), AR(2) and ARMA(2, 1). The data is periodic with one year, because of that we are going to use a

AR(1)

- AR(1) forecast in the last window.

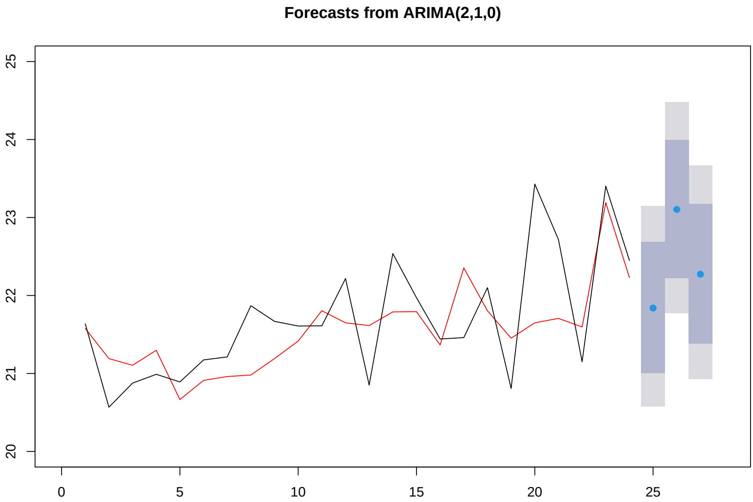
```
## Series: data.trend[141:164]
## ARIMA(1,1,0)
##
## Coefficients:
##           ar1
##       -0.6072
## s.e.   0.1647
##
## sigma^2 estimated as 0.002643:  log likelihood=35.91
## AIC=-67.82   AICc=-67.22   BIC=-65.55
##
## Training set error measures:
##                       ME       RMSE        MAE       MPE
## Training set 0.00539111 0.04922529 0.03849821 0.1375791
##                     ACF1
## Training set -0.3538996
```


Forecasts from ARIMA(1,1,0)
20
21
22
23
24
25
0
5
10
15
20
25

AR(2)

- AR(2) forecast in the last window.

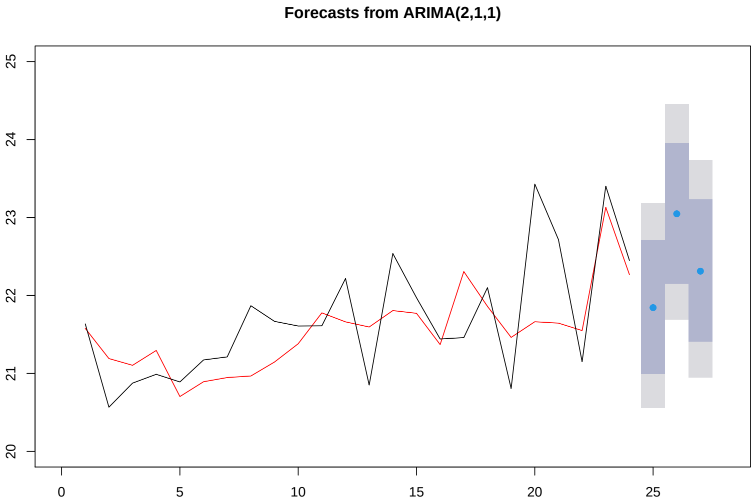
```
## Series: data.trend[141:164]
## ARIMA(2,1,0)
##
## Coefficients:
##           ar1      ar2
##       -0.9922  -0.6854
## s.e.   0.1554   0.1595
##
## sigma^2 estimated as 0.001578:  log likelihood=41.76
## AIC=-77.52   AICc=-76.26   BIC=-74.12
##
## Training set error measures:
##                       ME       RMSE        MAE       MPE
## Training set 0.008125179 0.03715651 0.02928903 0.2191975
##                    ACF1
## Training set -0.1017465
```


Forecasts from ARIMA(2,1,0)
20
21
22
23
24
25
0
5
10
15
20
25

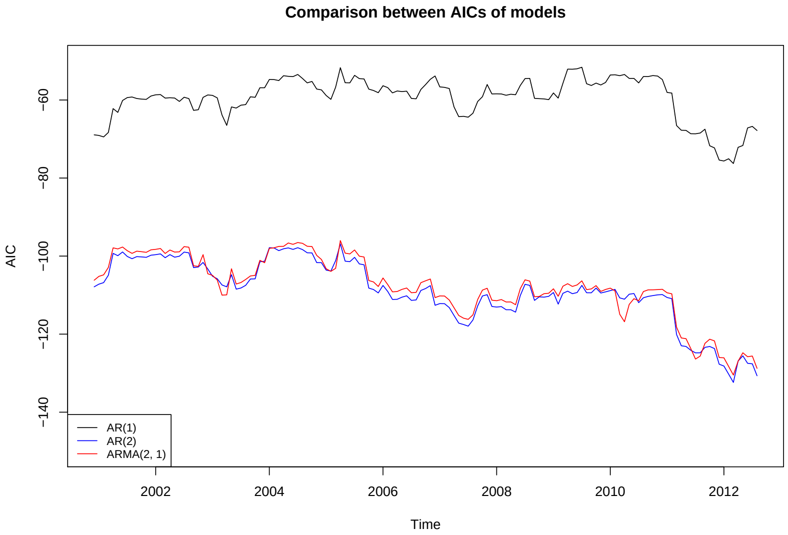
ARMA(2, 1)

- Forecast in the last window.

```
## Series: data.trend[141:164]
## ARIMA(2,1,1)
##
## Coefficients:
##          ar1      ar2      ma1
##      -0.9462  -0.6574  -0.0836
## s.e.   0.2282   0.1973   0.2809
##
## sigma^2 estimated as 0.00165:  log likelihood=41.81
## AIC=-75.62   AICc=-73.39   BIC=-71.07
##
## Training set error measures:
##                       ME       RMSE       MAE       MPE
## Training set 0.008656551 0.03708272 0.0291618 0.2341244
##                     ACF1
## Training set -0.06574234
```


Forecasts from ARIMA(2,1,1)
20
21
22
23
24
25
0
5
10
15
20
25

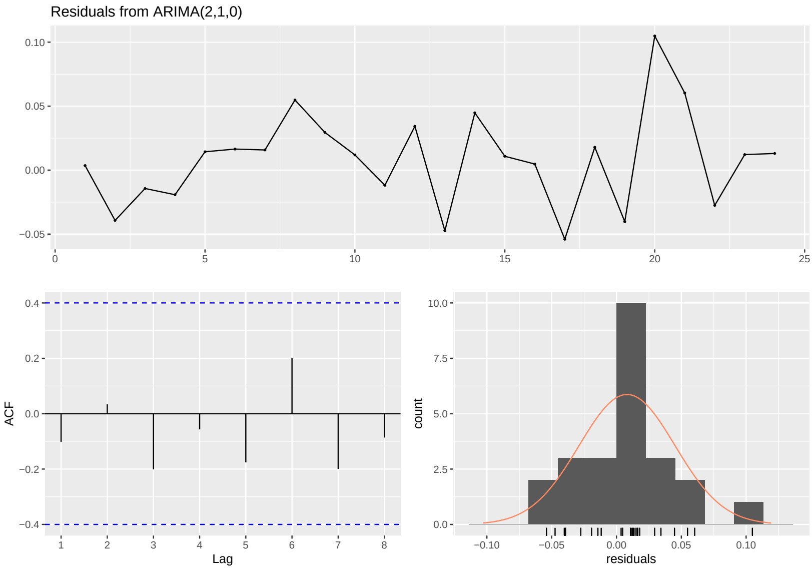
- Comparison between AIC of the three models in each of the windows.



- The best model is AR(2).

Residual analysis

- Analysis in the last window.

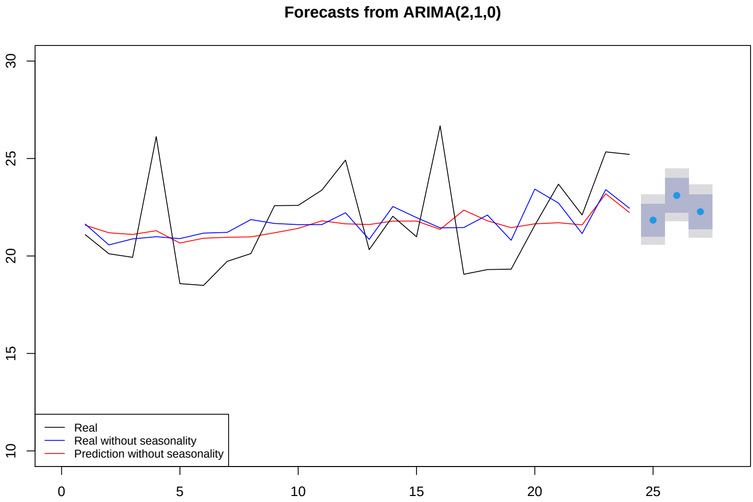


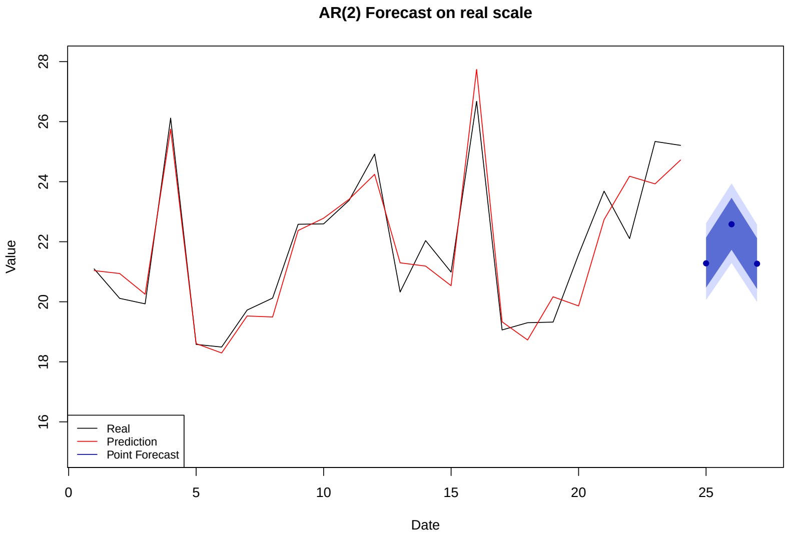
##

```
##
##  Jarque Bera Test
##
## data:  model.AR2$residuals
## X-squared = 0.93129, df = 2, p-value = 0.6277
```

Model selection

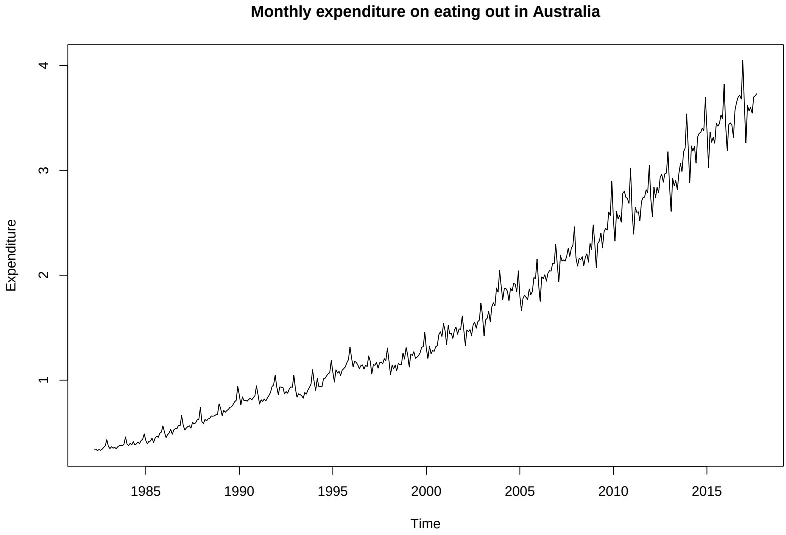
- Final model is AR(2). We are going to make predictions on the original scale, before BoxCox and STL.





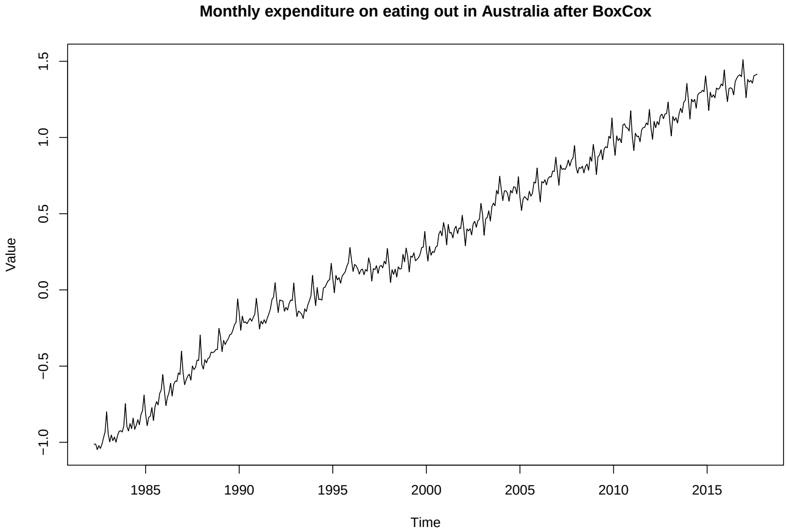
AR(2) Forecast on real scale
Value
Date
16
18
20
22
24
26
28
0
5
10
15
20
25
Real
Prediction
Point Forecast

Eating Out



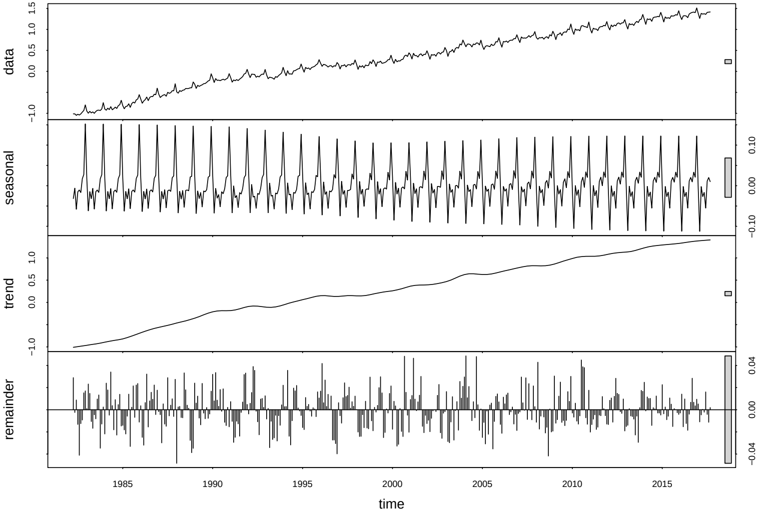
Monthly expenditure on eating out in Australia
Expenditure
Time
1
2
3
4
1985
1990
1995
2000
2005
2010
2015

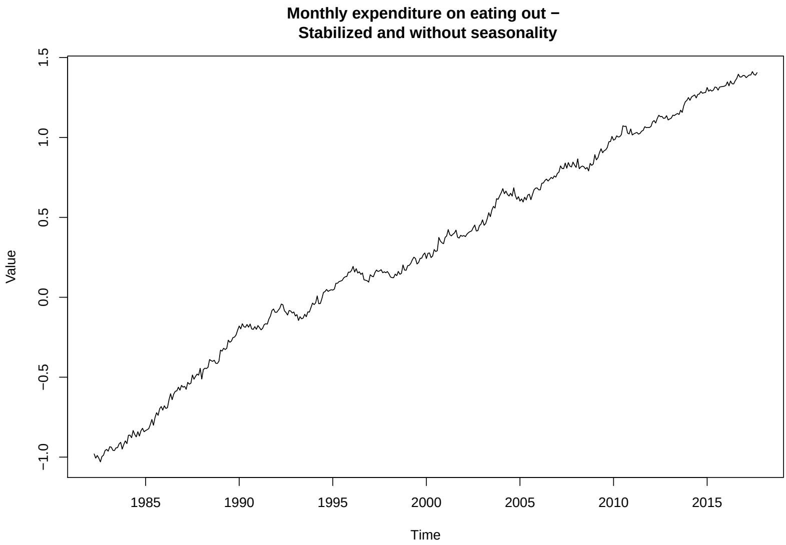
Stabilize variance



Remove seasonality

- Removing with STL (Lowess fit).



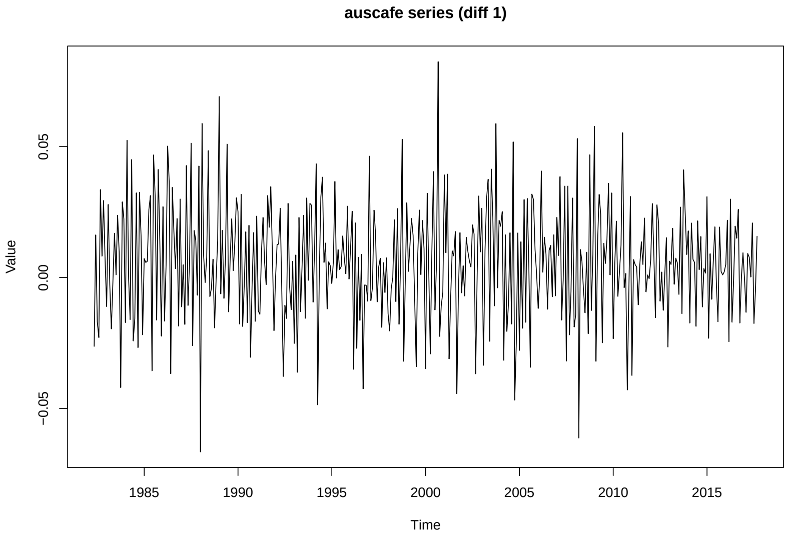

Monthly expenditure on eating out −
Stabilized and without seasonality
Value
Time
1.5
1.0
0.5
0.0
−0.5
−1.0
1985
1990
1995
2000
2005
2010
2015

Stationarity

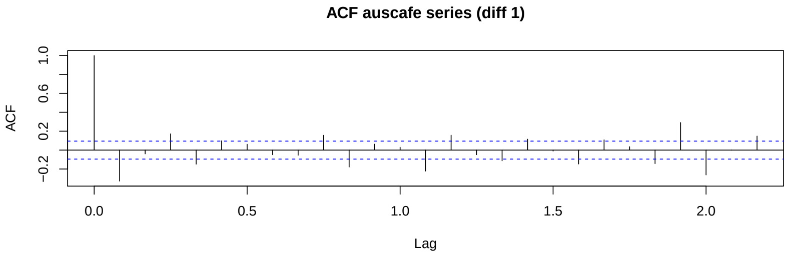
- Dickey Fuller test with series and differenced series.

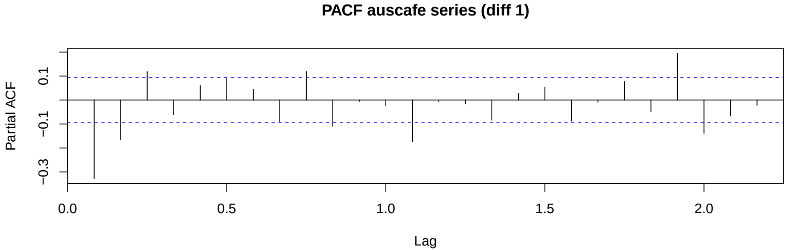
```
##
##  Augmented Dickey-Fuller Test
##
## data:  data.trend
## Dickey-Fuller = -2.9004, Lag order = 7, p-value = 0.197
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(data.trend)
## Dickey-Fuller = -7.1099, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```


auscafe series (diff 1)
Value
0.05
0.00
−0.05
1985
1990
1995
2000
2005
2010
2015
Time

ACF and PACF



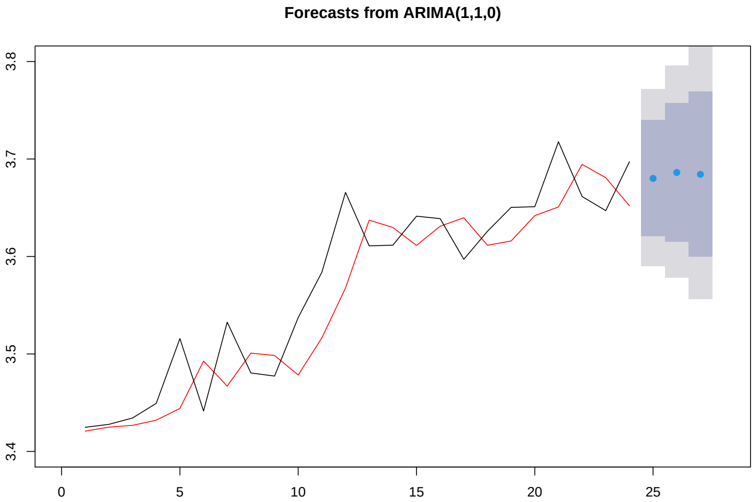


- Looks like AR(1) because of the first two lags. Models being tested are AR(1), AR(2), ARMA(1,1) and ARMA(2, 1). The data is periodic with one year, because of that we are going to

AR(1)

- AR(1) forecast in the last window.

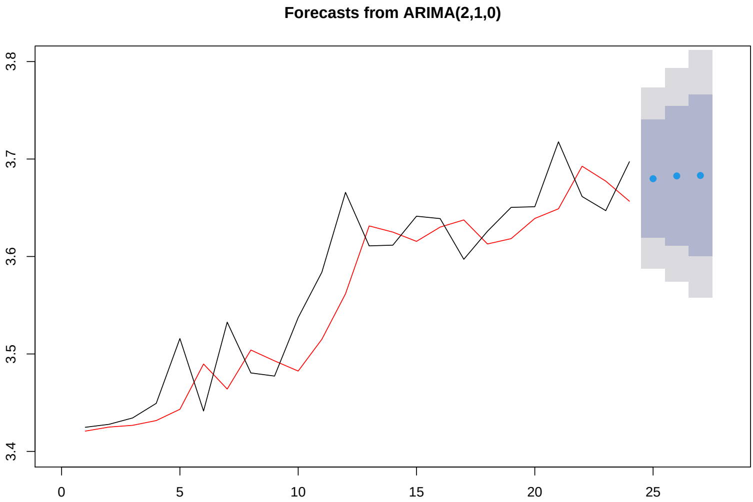
```
## Series: data.trend[403:426]
## ARIMA(1,1,0)
##
## Coefficients:
##           ar1
##       -0.3457
## s.e.   0.1945
##
## sigma^2 estimated as 0.0002114:  log likelihood=65.13
## AIC=-126.26   AICc=-125.66   BIC=-123.99
##
## Training set error measures:
##                       ME       RMSE        MAE       MPI
## Training set 0.004755535 0.01391924 0.01135628 0.343518:
##                    ACF1
## Training set -0.1673774
```


Forecasts from ARIMA(1,1,0)
3.4
3.5
3.6
3.7
3.8
0
5
10
15
20
25

AR(2)

- AR(2) forecast in the last window.

```
## Series: data.trend[403:426]
## ARIMA(2,1,0)
##
## Coefficients:
##           ar1      ar2
##       -0.3735  -0.0765
## s.e.   0.2080   0.2039
##
## sigma^2 estimated as 0.00022:  log likelihood=65.2
## AIC=-124.4   AICc=-123.13   BIC=-120.99
##
## Training set error measures:
##                       ME       RMSE        MAE       MPE
## Training set 0.005082026 0.01387307 0.01105395 0.3673927
##                    ACF1
## Training set -0.1521658
```

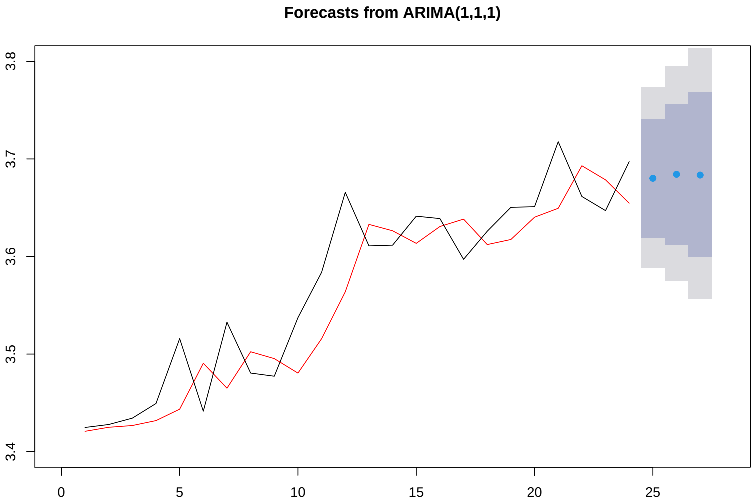


Forecasts from ARIMA(2,1,0)
3.4
3.5
3.6
3.7
3.8
0
5
10
15
20
25

ARMA(1, 1)

- Forecast in the last window.

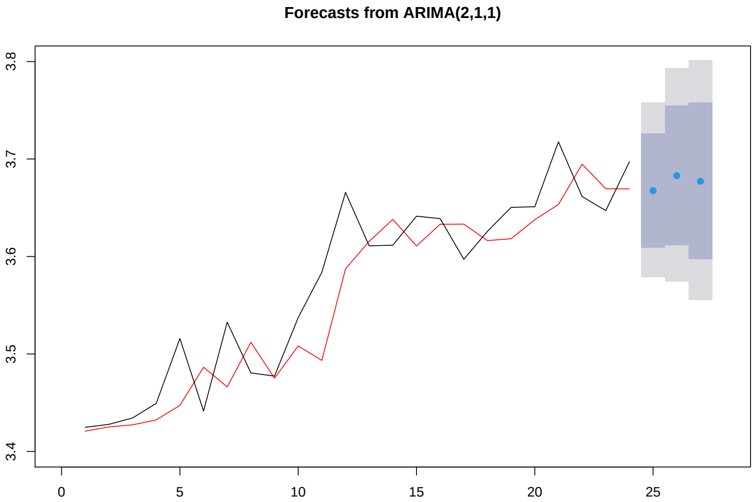
```
## Series: data.trend[403:426]
## ARIMA(1,1,1)
##
## Coefficients:
##           ar1      ma1
##       -0.2284  -0.1365
## s.e.   0.4659   0.4569
##
## sigma^2 estimated as 0.0002205:  log likelihood=65.17
## AIC=-124.35   AICc=-123.08   BIC=-120.94
##
## Training set error measures:
##                       ME       RMSE        MAE       MPE
## Training set 0.005001364 0.01389079 0.01116076 0.3614767
##                    ACF1
## Training set -0.1586
```


Forecasts from ARIMA(1,1,1)
3.4
3.5
3.6
3.7
3.8
0
5
10
15
20
25

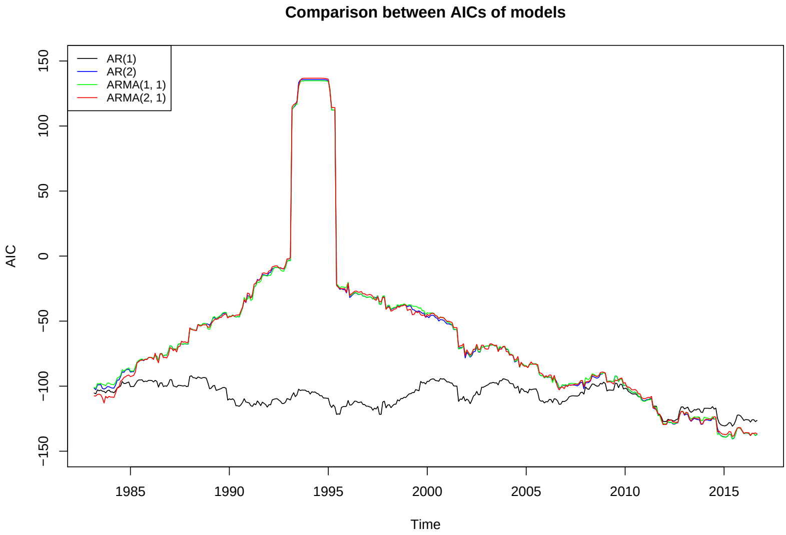
ARMA(2, 1)

- Forecast in the last window.

```
## Series: data.trend[403:426]
## ARIMA(2,1,1)
##
## Coefficients:
##           ar1      ar2     ma1
##       -1.2734  -0.4549  1.0000
## s.e.   0.1797   0.1932  0.2236
##
## sigma^2 estimated as 0.0001996:  log likelihood=66.03
## AIC=-124.06   AICc=-121.84   BIC=-119.52
##
## Training set error measures:
##                       ME       RMSE        MAE     MPE
## Training set 0.004699864 0.01289544 0.01002994 0.34023 (
##                    ACF1
## Training set -0.1205423
```


Forecasts from ARIMA(2,1,1)
3.4
3.5
3.6
3.7
3.8
0
5
10
15
20
25

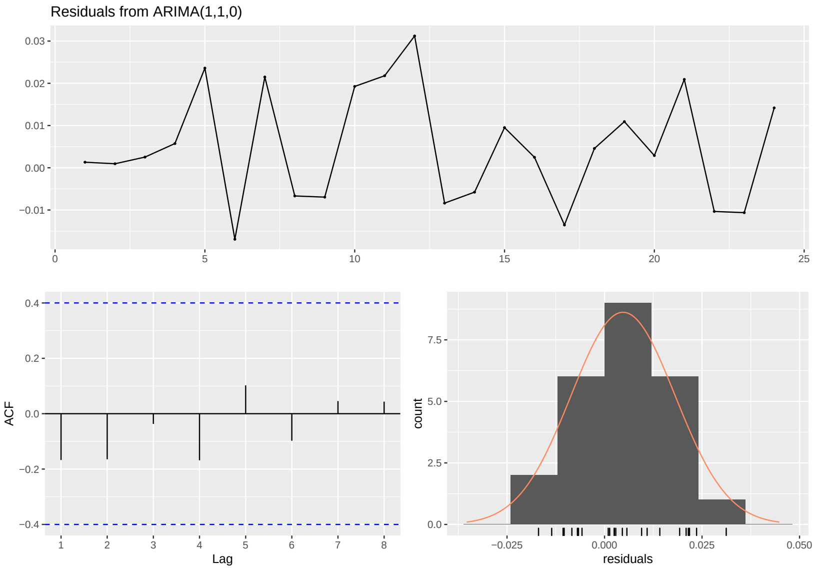
- Comparison between AIC of the three models in each of the windows.



- The best model is AR(1).

Residual analysis

- Analysis in the last window.



##

```
##
##  Jarque Bera Test
##
## data:  model.AR1$residuals
## X-squared = 1.2029, df = 2, p-value = 0.548
```

Model selection

- Final model is AR(1). We are going to make predictions on the original scale, before BoxCox and STL.