ARIMA applied to DEXCOM Equity

Guillaume Ostrom - ECP 2017

July 18th 2017

Introduction

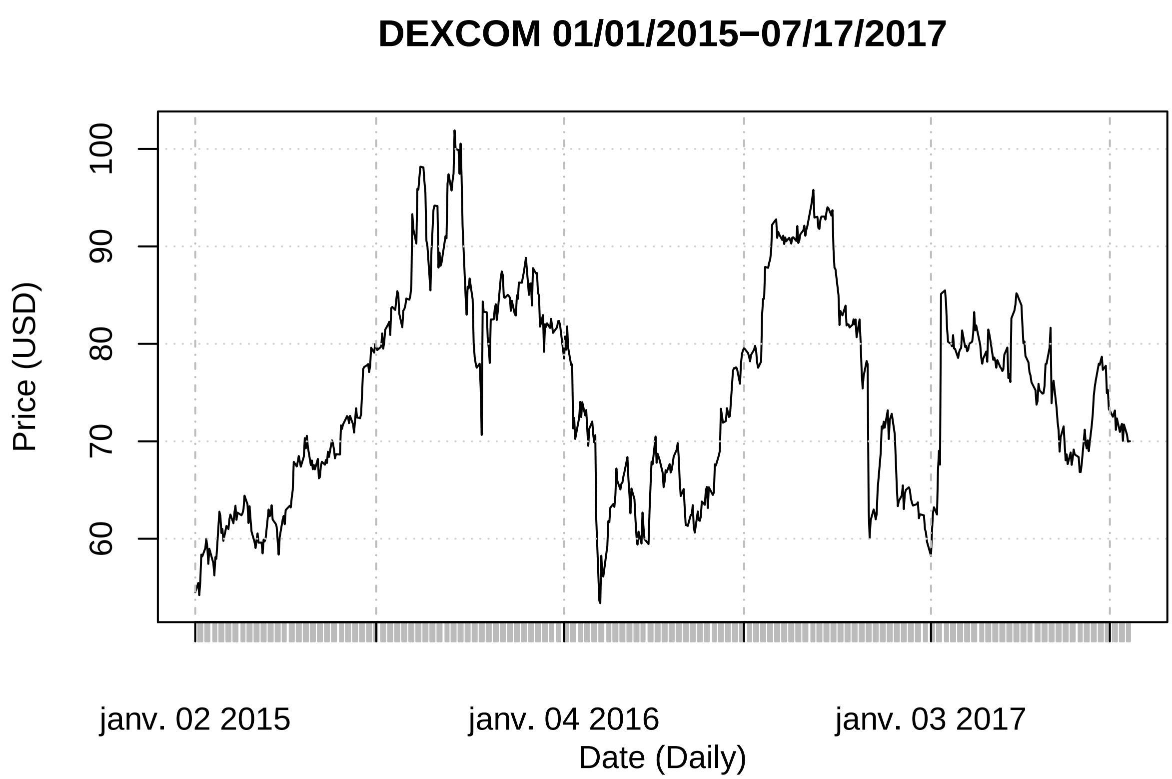
In this paper we will try to modeling and forecasting **DEXCOM Equity**. We will use a ARMA(p,q) (Auto Regression Moving Average) model. First we will differencing the daily price to convert the non-stationnarity time series in a stationnary time serie without trend. Then we will test the stationarity and find the p,q which reduce the AIC criteron. Finally we will estimate the accuracy of our model, forecast the next values and cross check the model.

Here is our ARMA model:

$$X_t = \mu + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

With X_t the time serie, μ the mean of X_t, ε_t the white noise of each X_t.
p and φ_i are the parameters of the AR model.
q and θ_i are the parameters of the MA model.

Here is the daily price of DexCom Inc. from Yahoo Finance.



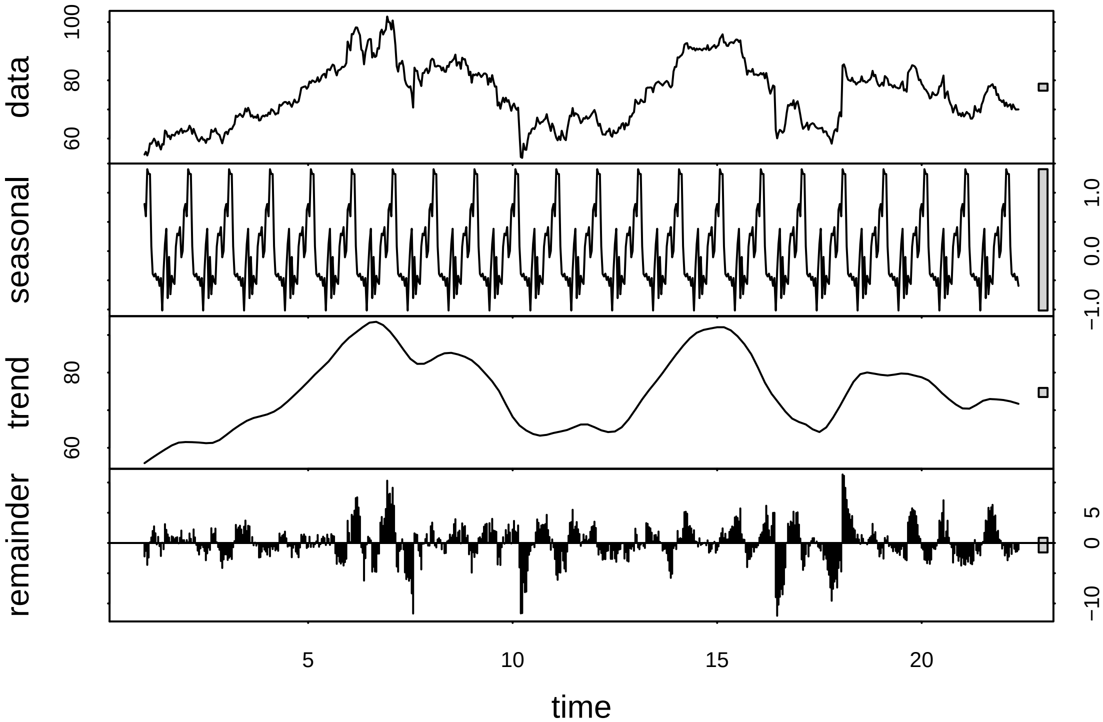
Here is a little summary:

```
##      Index              DXCM.Open        DXCM.High         DXCM.Low
##  Min.   :2015-01-02   Min.   : 52.92   Min.   : 55.61   Min.   :47.92
##  1st Qu.:2015-08-21   1st Qu.: 66.12   1st Qu.: 67.70   1st Qu.:64.87
##  Median :2016-04-12   Median : 75.00   Median : 76.03   Median :73.31
##  Mean   :2016-04-11   Mean   : 74.98   Mean   : 76.18   Mean   :73.73
##  3rd Qu.:2016-11-28   3rd Qu.: 82.49   3rd Qu.: 84.00   3rd Qu.:81.45
##  Max.   :2017-07-20   Max.   :101.09   Max.   :103.29   Max.   :99.91
##    DXCM.Close       DXCM.Volume        DXCM.Adjusted
##  Min.   : 53.38   Min.   :       0   Min.   : 53.38
##  1st Qu.: 66.12   1st Qu.:  578725   1st Qu.: 66.12
##  Median : 74.95   Median :  771850   Median : 74.95
##  Mean   : 75.02   Mean   :  968467   Mean   : 75.02
##  3rd Qu.: 82.56   3rd Qu.: 1061100   3rd Qu.: 82.56
##  Max.   :101.91   Max.   :10387200   Max.   :101.91
```

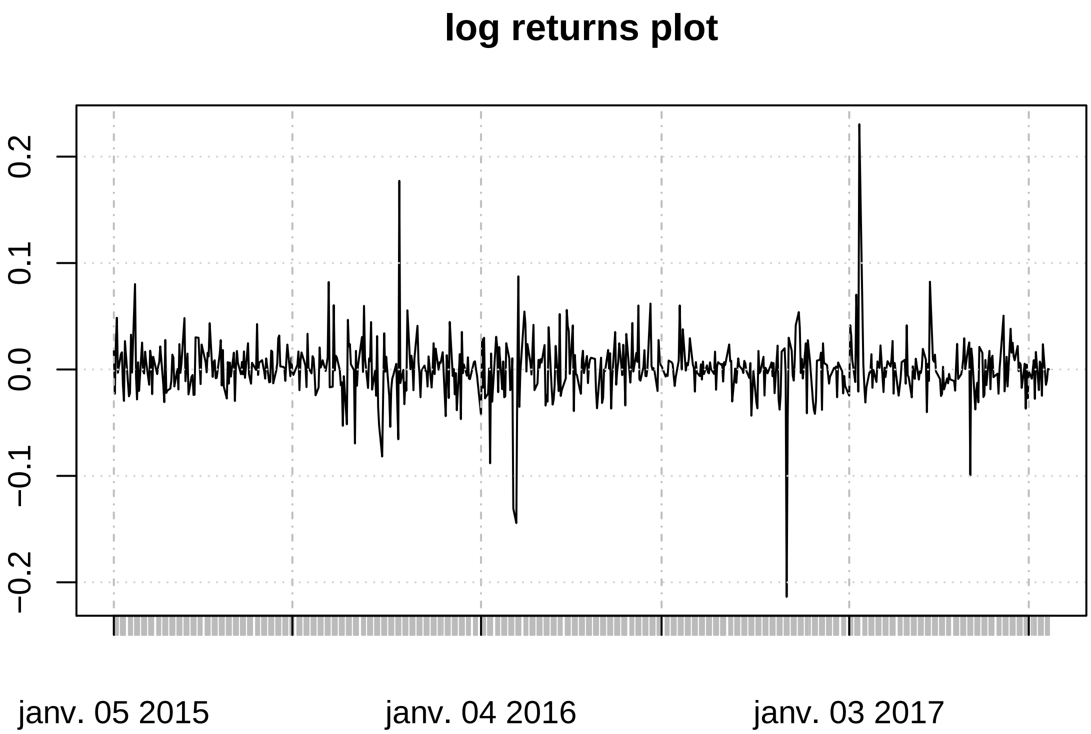
```
##            DXCM.Close
## 2017-07-13      70.04
## 2017-07-14      71.72
## 2017-07-17      70.69
## 2017-07-18      69.97
## 2017-07-19      70.00
## 2017-07-20      70.01
```

Part I : Differenciation

We need to decompose the time serie in different composants: his trend, his seasonality and the noise. We consider the low frequency variation as the trend, the middle frequencies as the seasonality and the high frequency variation as a noise.



We need to convert the non-stationnarity time serie in a stationnary time serie without trend. We tried 3 box-Cox transformations, finally the log difference gave the best results.



We can reject the hypothesis of non-stationnarity with the log differences.

```
##
##  Augmented Dickey-Fuller Test
##
## data:  stock
## Dickey-Fuller = -9.1225, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary
```

Now the time serie is stationnary. We can try an ARMA(p,q) model.

Part II : Modeling

To chose the p (AR) and the q (MA) we use the low AIC (Akaike information criterion) method.

Here we calculate the AIC for p=1,2,3,4 and q=1,2,3,4 :

	MA0	MA1	MA2	MA3	MA4
AR0	-2786.98	-2785.00	-2783.33	-2781.33	-2779.34
AR1	-2785.00	-2783.10	-2781.29	-2779.33	-2777.34
AR2	-2783.34	-2781.33	-2779.12	-2786.61	-2784.63
AR3	-2781.34	-2779.34	-2791.12	-2786.88	-2784.80
AR4	-2779.34	-2777.34	-2775.34	-2773.75	-2773.67

So we can select $ARMA(3, 2)$ according to the AIC method, now we calculate:

the φ_i for the AR(p=3)
The AR part involves regressing the variable on its own lagged/past values.

$$X_t = \mu_1 + \sum_{i=1}^{3} \varphi_i X_{t-i} + \varepsilon_t$$

the θ_i for the MA(p=2)
The MA part involves modeling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past.

$$X_t = \mu_2 + \sum_{i=1}^{2} \theta_i \varepsilon_{t-i} \ + \varepsilon_t$$

And the mean

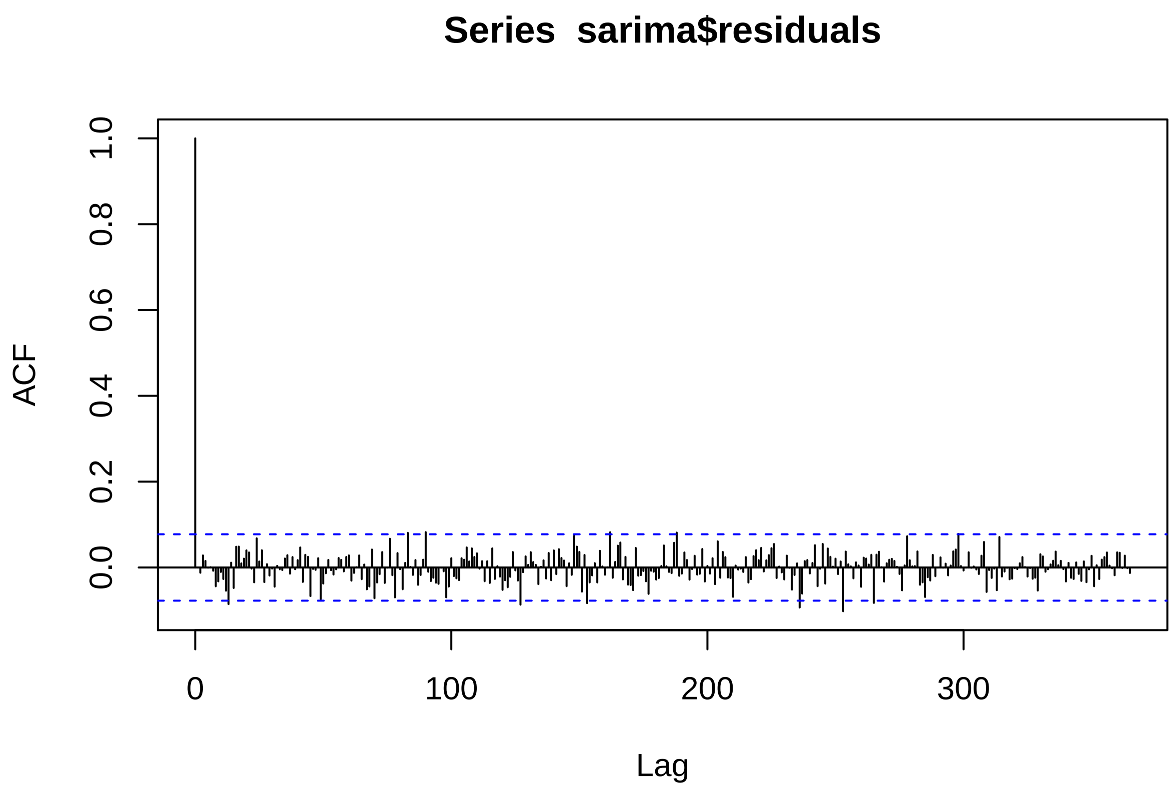
$$\mu = \mu_1 + \mu_2$$

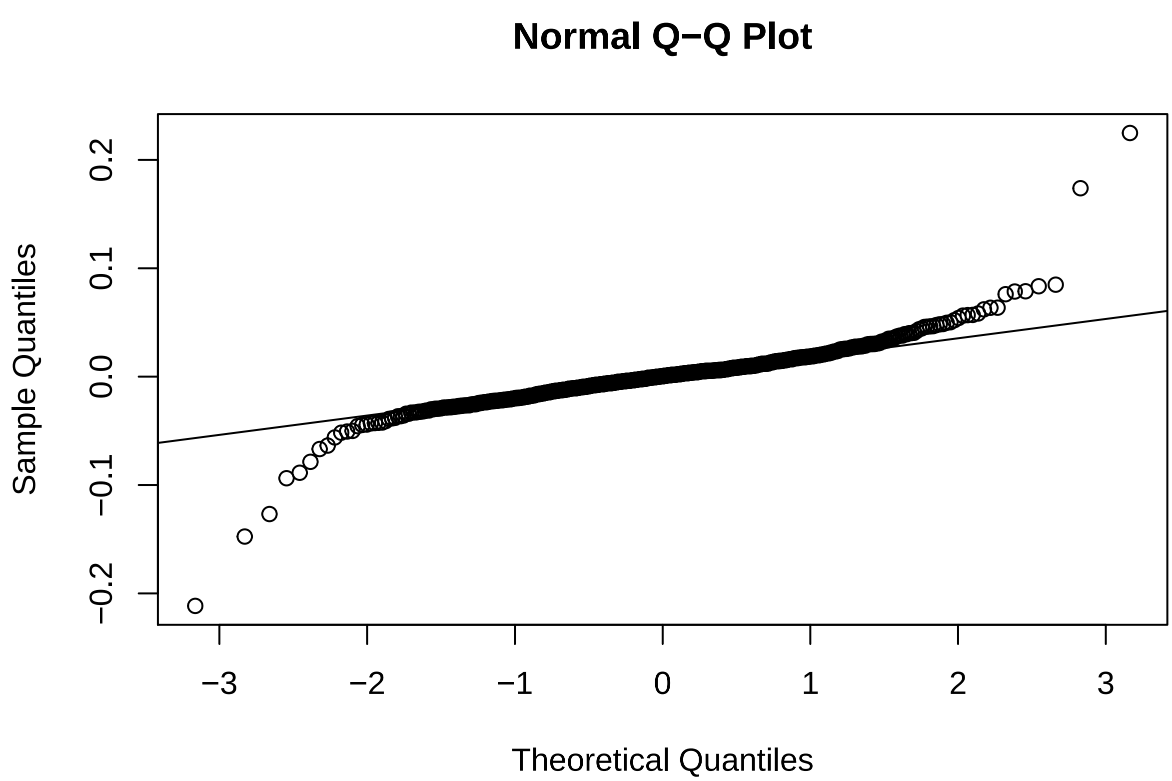
```
sarima=arima(stock,order=c(3,1,2))
sarima
```

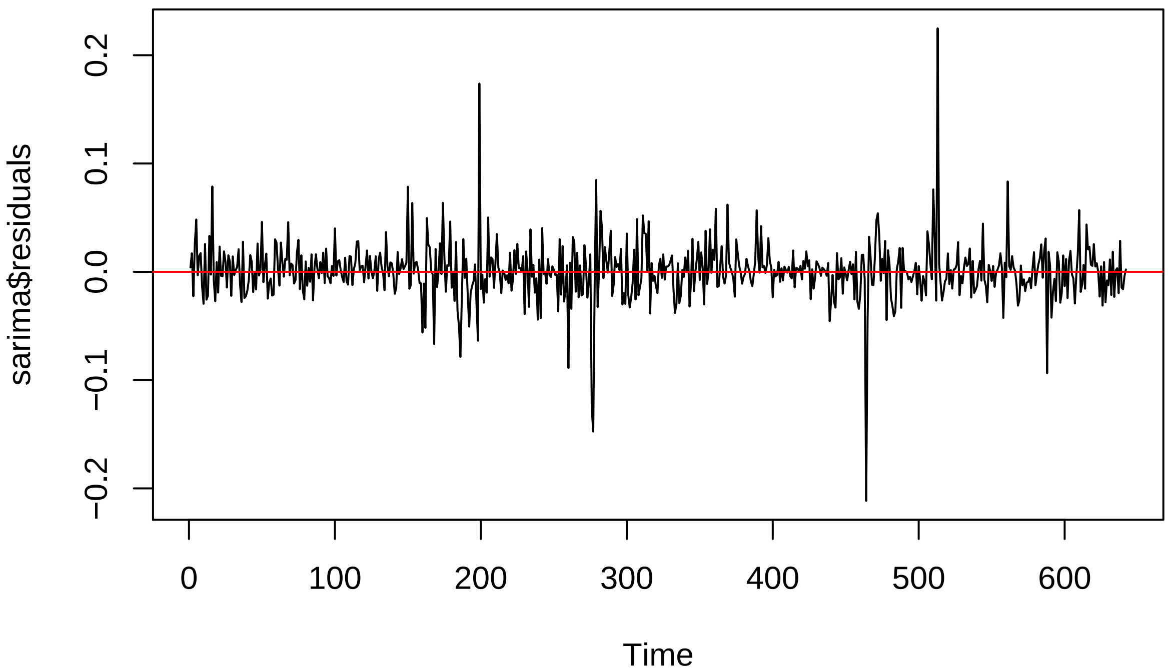
```
##
## Call:
## arima(x = stock, order = c(3, 1, 2))
##
## Coefficients:
##          ar1      ar2      ar3      ma1     ma2
##       0.8418  -0.9861  -0.0146  -0.8497  0.9995
## s.e.  0.0402   0.0342   0.0397   0.0091  0.0133
##
## sigma^2 estimated as 0.0007339:  log likelihood = 1401.56,  aic = -2791.12
```

Part III : Forecasting and back-testing

To check our model we need to check our residuals with the:
-ACF plot under the 5% level acceptation under the null hypothesis.
-QQplot to see if the residuals are normally distributed.
-check if there is not extremly high residual point.


Series sarima$residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
Lag

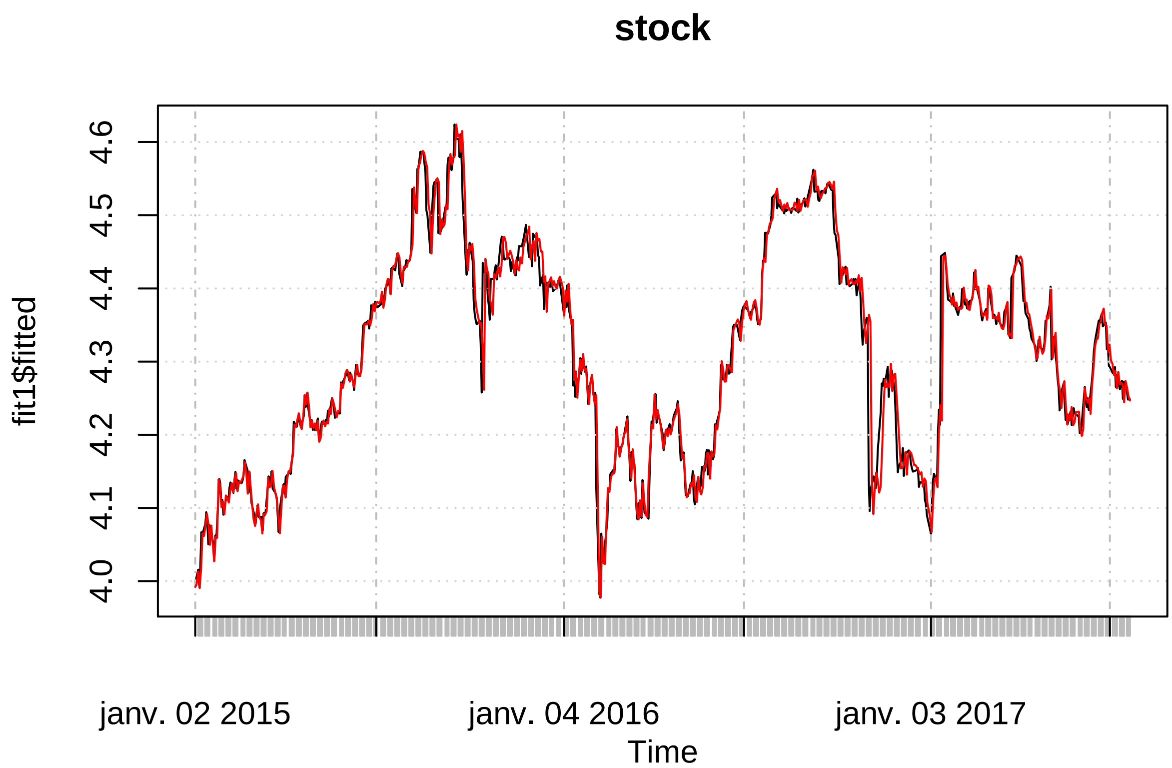

Normal Q−Q Plot
Sample Quantiles
0.2
0.1
0.0
−0.1
−0.2
−3
−2
−1
0
1
2
3
Theoretical Quantiles



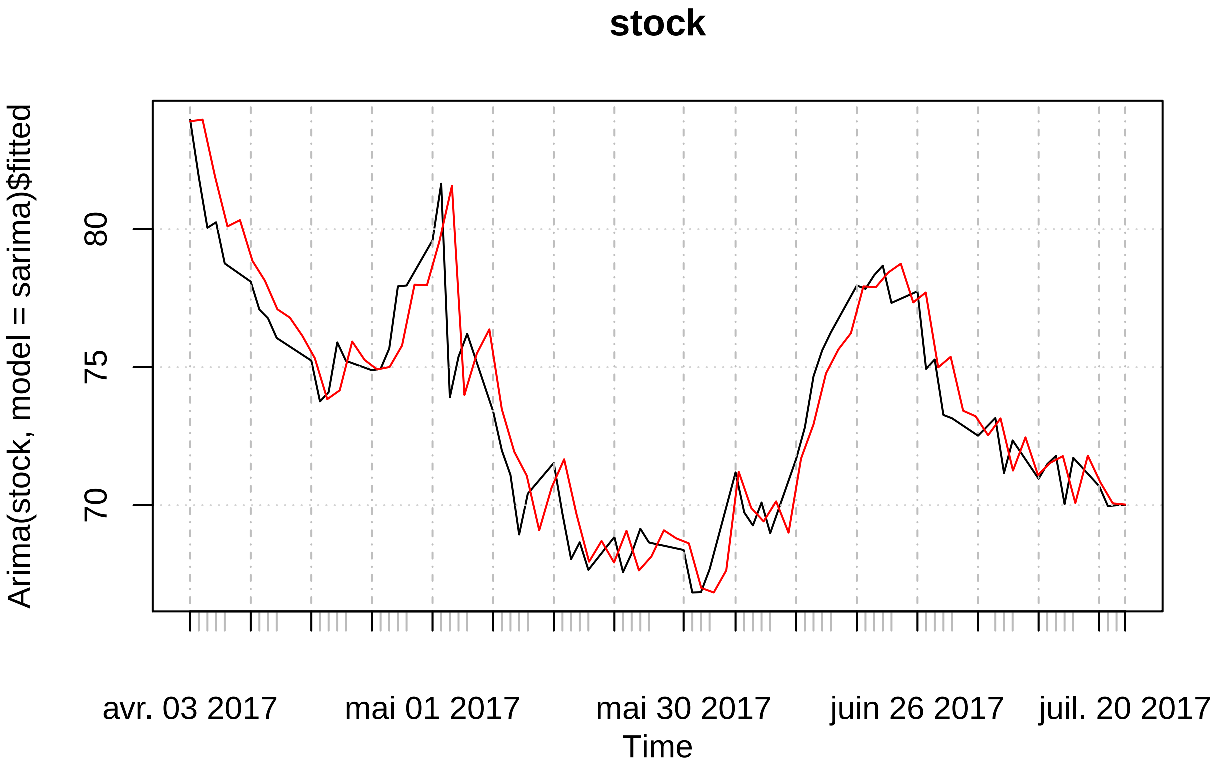
In conclusion we cannot reject the null hypothesis of Gaussiqn noise. So the the model is good to test.

```
## Series: stock
## ARIMA(3,1,2)
##
## Coefficients:
##          ar1      ar2      ar3      ma1     ma2
##       0.8418  -0.9861  -0.0146  -0.8497  0.9995
## s.e.  0.0000   0.0000   0.0000   0.0000  0.0000
##
## sigma^2 estimated as 0.0007339:  log likelihood=1401.56
## AIC=-2801.12   AICc=-2801.11   BIC=-2796.66
```

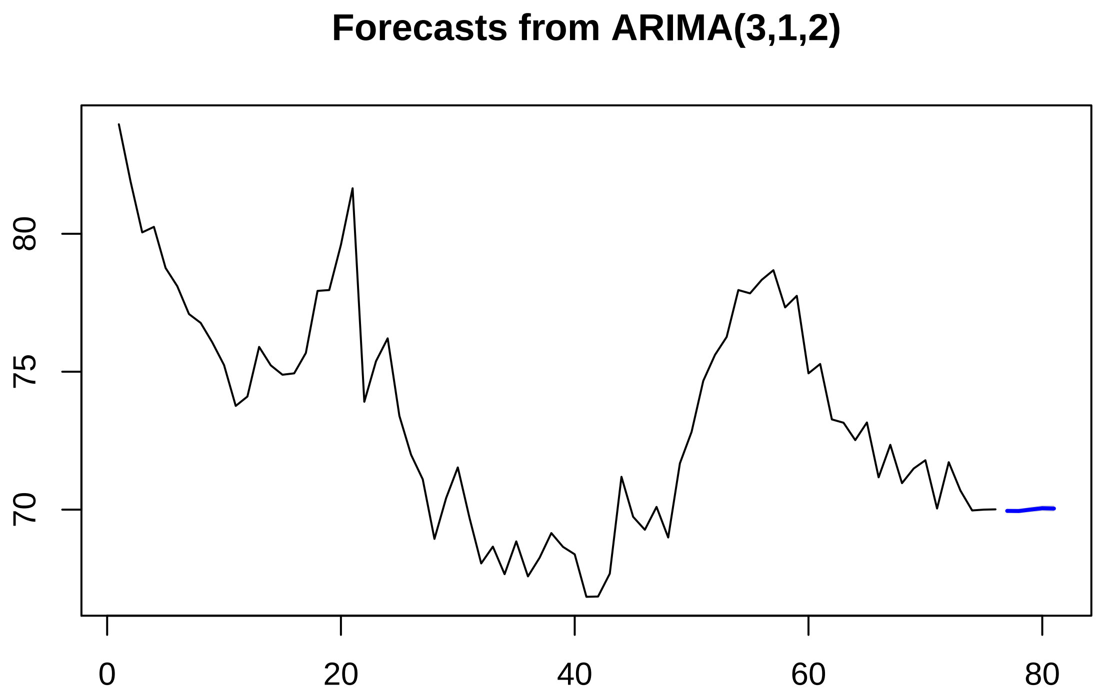
Now we can try to predict the price of this week.




```
## [1] "quadratic error"
## [1] 77.8153
## [1] "Cross validation (10^5)"
## [1] "Mean error (alpha=0.05) = 43%"
```



stock
Arima(stock, model = sarima)$fitted
70
75
80
avr. 03 2017
mai 01 2017
mai 30 2017
juin 26 2017
juil. 20 2017
Time




```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 77       69.95404 69.91911 69.98897 69.90063 70.00745
## 78       69.95057 69.90125 69.99989 69.87514 70.02600
## 79       70.00269 69.94234 70.06305 69.91039 70.09500

## [1] "Wednesday 19th July 2017 Close price: 69.954"

## [1] "Thursday 20th July 2017 Close price: 69.951"

## [1] "Friday 21th July 2017 Close price: 70.003"
```

Conclusion

We choose a ARMA$(3, 2)$ model with a low AIC criterion.
However the quadratic error is too high and the confidence interval is not optimal (57% for $\alpha = 0.05$).