Predict_413_Discussion_Week_6

Singh, Gurjeet

February 17, 2018

Contents

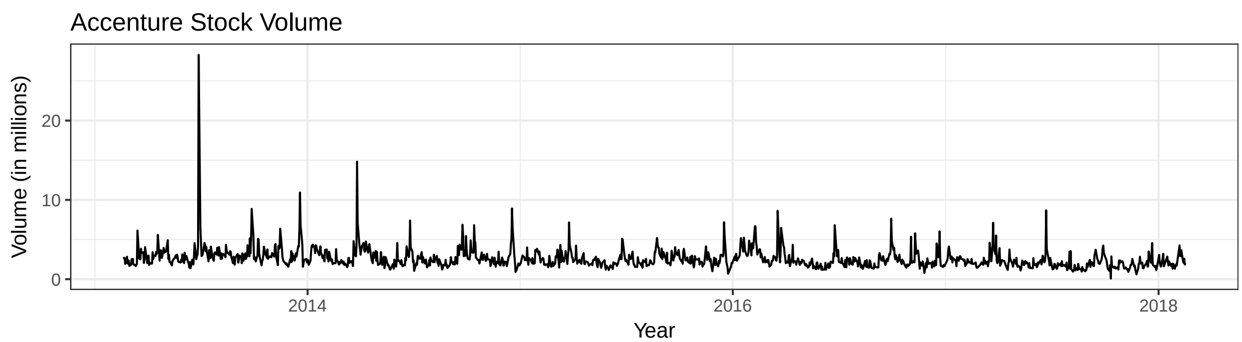
1. Introduction

For this week's discussion, we forecast the daily adjusted closing price of our choice using time series components and at least once external regressor (i.e. transaction volume at t-1).

2. Data

For this week's discussion, I chose ACN (Accenture) as my stock. I remember buying ACN shares for around $35 (after 15% discount) back in 2011/2012. I can't believe its stock price has gone so high.

The figure below shows the Accenture stock closing price and the stock volume. Closing Price chart shows that the stock price has been increasing over the years. As expected from the stocks, there's no seasonality in the data. Looking at the volume chart, I wonder what happened in Q4 of 2013.



3. Model

We created two models using auto.arima with external regressor. One of the models using the stepwise selection. The other model with the non-stepwise selection. The auto.arima with stepwise selection produced an ARIMA(1,1,1) with one AR term, one MA term, and one degree of differences. The auto.arima with non-stepwise selection produced an ARIMA(2,1,2) with two AR term, two MA term, and one degree of differences.

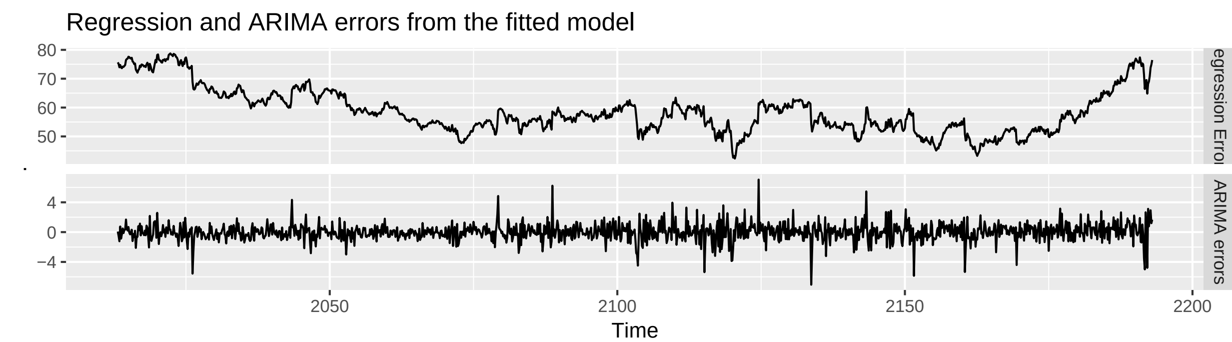
The information below shows the summary statistics of both the models. When comparing both the models using AIC and AICc, it appears that ARIMA(2,1,2) (i.e. the model with non-stepwise selection) performed better. However, when we look at other metrics such as BIC, MAPE, and MASE, it appears that ARIMA(1,1,1) performed better. Further investigation will be required to find the best model.

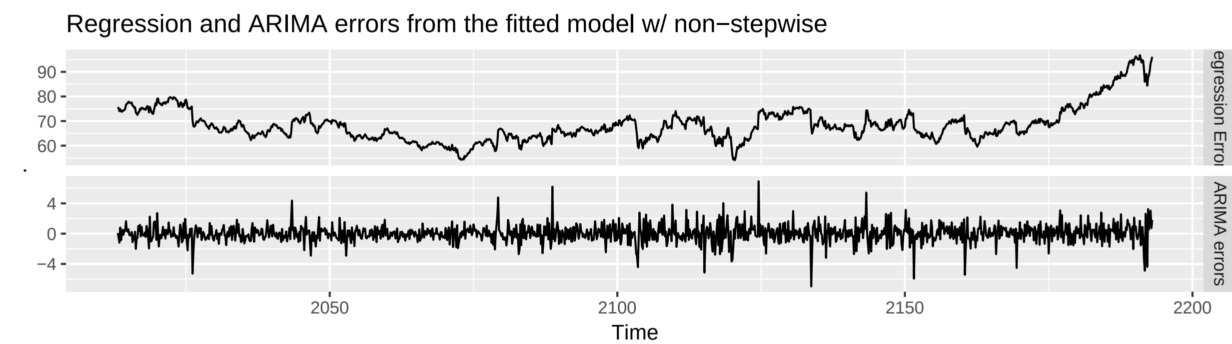
```
## Series: ts(ACN_Close_5yr, frequency = 7, start = c(2013, 2, 16))
## Regression with ARIMA(1,1,1) errors
##
## Coefficients:
##          ar1      ma1   drift  ACN.Volume
##       0.9136  -0.9454  0.0686           0
## s.e.  0.0401   0.0323  0.0000           0
##
## sigma^2 estimated as 1.37:  log likelihood=-1982.52
## AIC=3975.04   AICc=3975.09   BIC=4000.73
##
## Training set error measures:
##                        ME     RMSE       MAE         MPE      MAPE
## Training set -0.001139991 1.168036 0.8189111 -0.02582528 0.8189585
##                   MASE        ACF1
## Training set 0.3625254 0.006520938
```

```
## Series: ts(ACN_Close_5yr, frequency = 7, start = c(2013, 2, 16))
## Regression with ARIMA(2,1,2) errors
##
## Coefficients:
##           ar1      ar2     ma1     ma2  drift  ACN.Volume
##       -0.6165  -0.9578  0.6245  0.9354  0.053           0
## s.e.   0.0298   0.0249  0.0411  0.0303  0.000           0
##
## sigma^2 estimated as 1.366:  log likelihood=-1979.94
## AIC=3973.88   AICc=3973.97   BIC=4009.85
##
## Training set error measures:
##                      ME     RMSE       MAE           MPE      MAPE
## Training set 0.01636167 1.165605 0.8252545 -0.0004672985 0.8253481
##                   MASE        ACF1
## Training set 0.3653336 -0.02516963
```

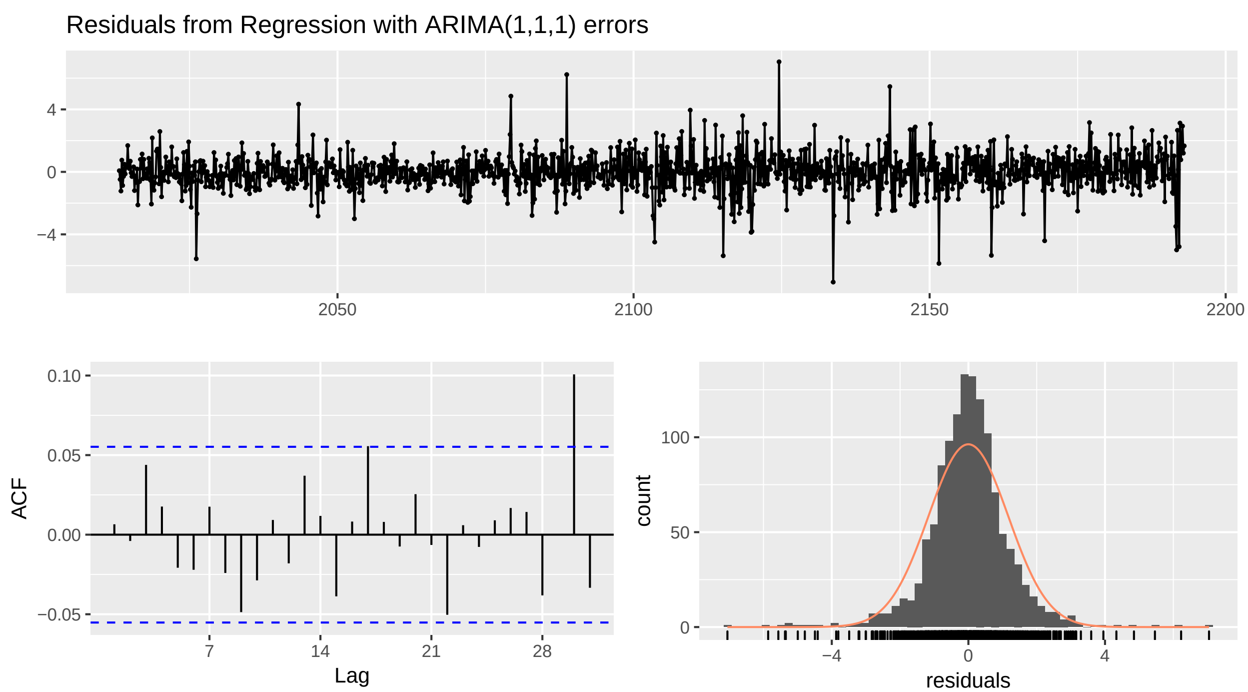
4. Forecast

The figures below show the regression errors and ARIMA errors from both the fitted models. Looking at the graphs, the errors seem to resemble for both models.



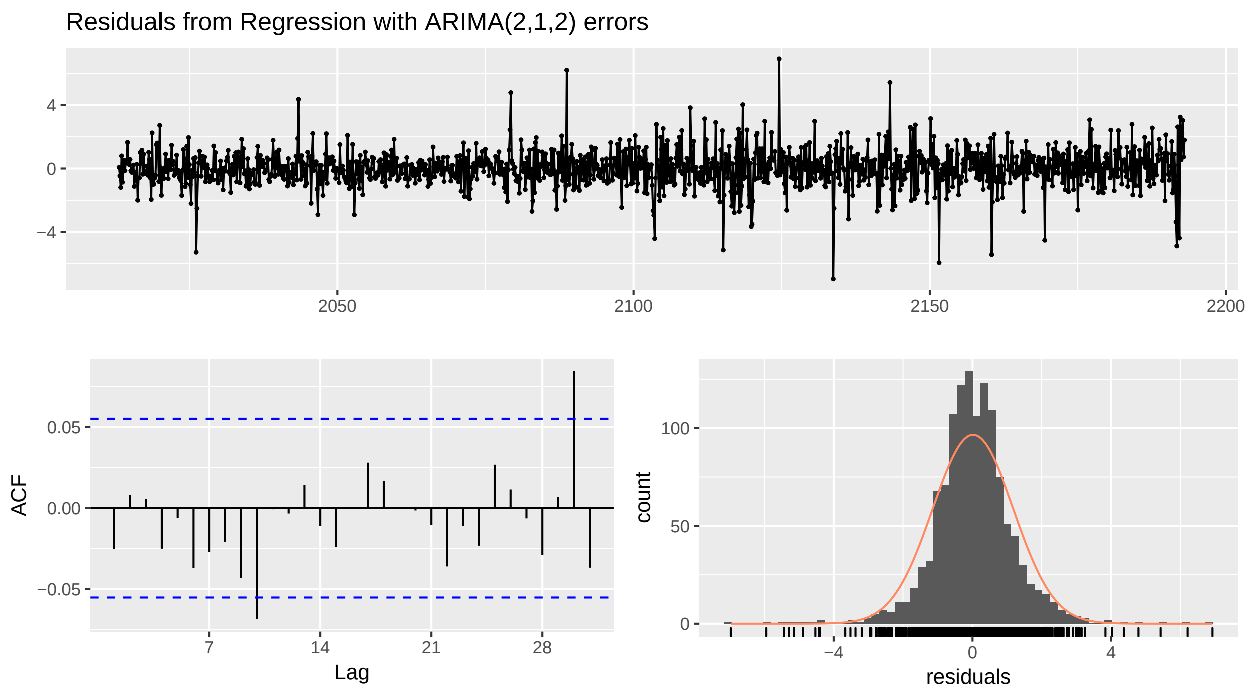


Next, we look at the residuals from the regression with ARIMA(1,1,1) errors and the portmanteau test results. The ACF plot in the figure below shows that all the correlations except one are within the threshold limits, indicating that the residuals are behaving like white noise. The portmanteau test below returns a large p-value, also suggesting that the residuals are white noise.



```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(1,1,1) errors
## Q* = 11.702, df = 10, p-value = 0.3055
##
## Model df: 4.   Total lags used: 14
```

Next, we look at the residuals from the regression with ARIMA(2,1,2) errors and the portmanteau test results. The ACF plot in the figure below shows that all the correlations except two are within the threshold limits, indicating that the residuals are also behaving like white noise but not better than residuals from the ARIMA(1,1,1). The portmanteau test below returns a smaller p-value, suggesting that the residuals are some white noise.



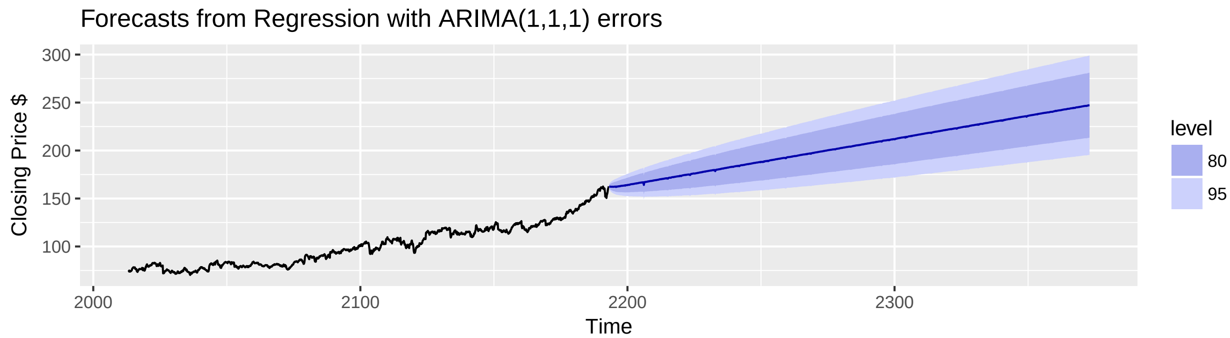

```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(2,1,2) errors
## Q* = 13.772, df = 8, p-value = 0.08789
##
## Model df: 6.   Total lags used: 14
```

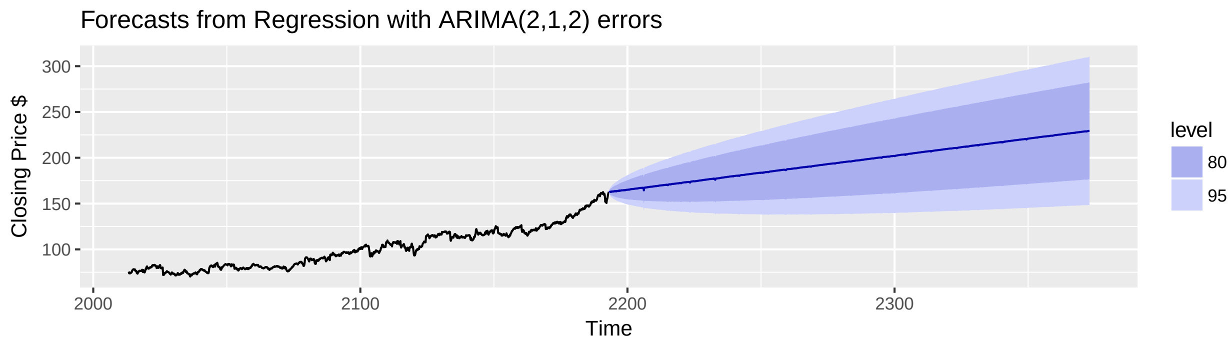
The table below shows the metrics from each model. It seems like default auto.arima with external regression performed better.

Table 1: Metrics for each model

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Auto Arima	-0.0011400	1.168036	0.8189111	-0.0258253	0.8189585	0.3625254	0.0065209
Auto Arima w/ Non-Stepwise	0.0163617	1.165605	0.8252545	-0.0004673	0.8253481	0.3653336	-0.0251696

The figure below shows the forecasts from both the models.





5. Conclusion

In conclusion, we would select auto.arima model with stepwise selection (ARIMA(1,1,1)) because we believe it performed better than the model with non-stepwise selection (ARIMA(2,1,2)).

RMarkdown code is available on my GitHub page (https://github.com/Singh-Gurjeet) in the Time Series and Forecasting repository.

6. Appendix I: R Code

```r
knitr::opts_chunk$set(echo = TRUE, warning = F, message = F)

library(tidyverse)
library(fpp)
library(quantmod) # use for gathering and charting Stock data
options(scipen = 999)

#Download the Accenture data from Google Finance.
K <- getSymbols("ACN", src="google", return.class = "xts", warnings = F)
#Extract only closing price of the stock.
ACN_Close <- ACN$ACN.Close # Get the closing rate
#Extract only closing price of the stock.
ACN_xreg <- ACN$ACN.Volume # Get the Volume
#filter data for 5 years only
ACN_Close_5yr <- ACN_Close["2013-02-16/"]
ACN_xreg_5yr <- ACN_xreg["2013-02-16/"]
```

```r
p1 <- autoplot(ACN_Close_5yr) + ggtitle("Accenture Stock Closing Price") +
        xlab("Year")+ ylab("Closing Price $") + theme_bw()
p2 <-autoplot(ACN_xreg_5yr/1000000) + ggtitle("Accenture Stock Volume") +
        xlab("Year")+ ylab("Volume (in millions)") + theme_bw()
gridExtra::grid.arrange(p1,p2, layout_matrix = rbind(1,2))

#default with xreg;
fitACN <- auto.arima(ts(ACN_Close_5yr, frequency = 7, start = c(2013,02,16)),
                    xreg = ACN_xreg_5yr)
summary(fitACN)

# # with Stepwise = False
fitACN2 <- auto.arima(ts(ACN_Close_5yr, frequency = 7, start = c(2013,02,16)),
                      xreg = ACN_xreg_5yr, stepwise = F, approximation = F)
summary(fitACN2)

p3 <- cbind("Regression Errors" = residuals(fitACN, type="regression"),
      "ARIMA errors" = residuals(fitACN, type="innovation")) %>%
        autoplot(facets=TRUE) +
          ggtitle("Regression and ARIMA errors from the fitted model")

p4 <- cbind("Regression Errors" = residuals(fitACN2, type="regression"),
      "ARIMA errors" = residuals(fitACN2, type="innovation")) %>%
        autoplot(facets=TRUE) +
        ggtitle("Regression and ARIMA errors from the fitted model w/ non-stepwise")
gridExtra::grid.arrange(p3, p4, nrow=2)

checkresiduals(fitACN)
checkresiduals(fitACN2)

fcast <- forecast(fitACN, xreg = ACN_xreg_5yr)
  acc <- accuracy(fcast)
fcast2 <- forecast(fitACN2, xreg = ACN_xreg_5yr)
  acc2 <- accuracy(fcast2)
both_acc <- rbind.data.frame(acc,acc2)
rownames(both_acc) <- c("Auto Arima","Auto Arima w/ Non-Stepwise")
knitr::kable(both_acc, caption = "Metrics for each model")

p5 <- fcast %>%
        autoplot() +
        ylab("Closing Price $")
p6 <- fcast2 %>%
        autoplot() +
        ylab("Closing Price $")
gridExtra::grid.arrange(p5, p6, nrow=2)
```