**Technical Specifications and Code**

# Determine the relationship between VN-Index and VCB on VN stock exchange

**A/ Introduction to analytical methods**

The analytical method is based on the correlation coefficient to determine the close correlation between the stock ticker VCB and VNI.

* Descriptive statistics method
* Tools: Rstudio
* Setup environment in Rstudio

*library(DT)*

*library(dplyr)*

*library(corrplot)*

*library(ggplot2)*

*library(quantmod)*

*library(ggplot2)*

*library(anytime)*

*library(lubridate)*

*library(rvest)*

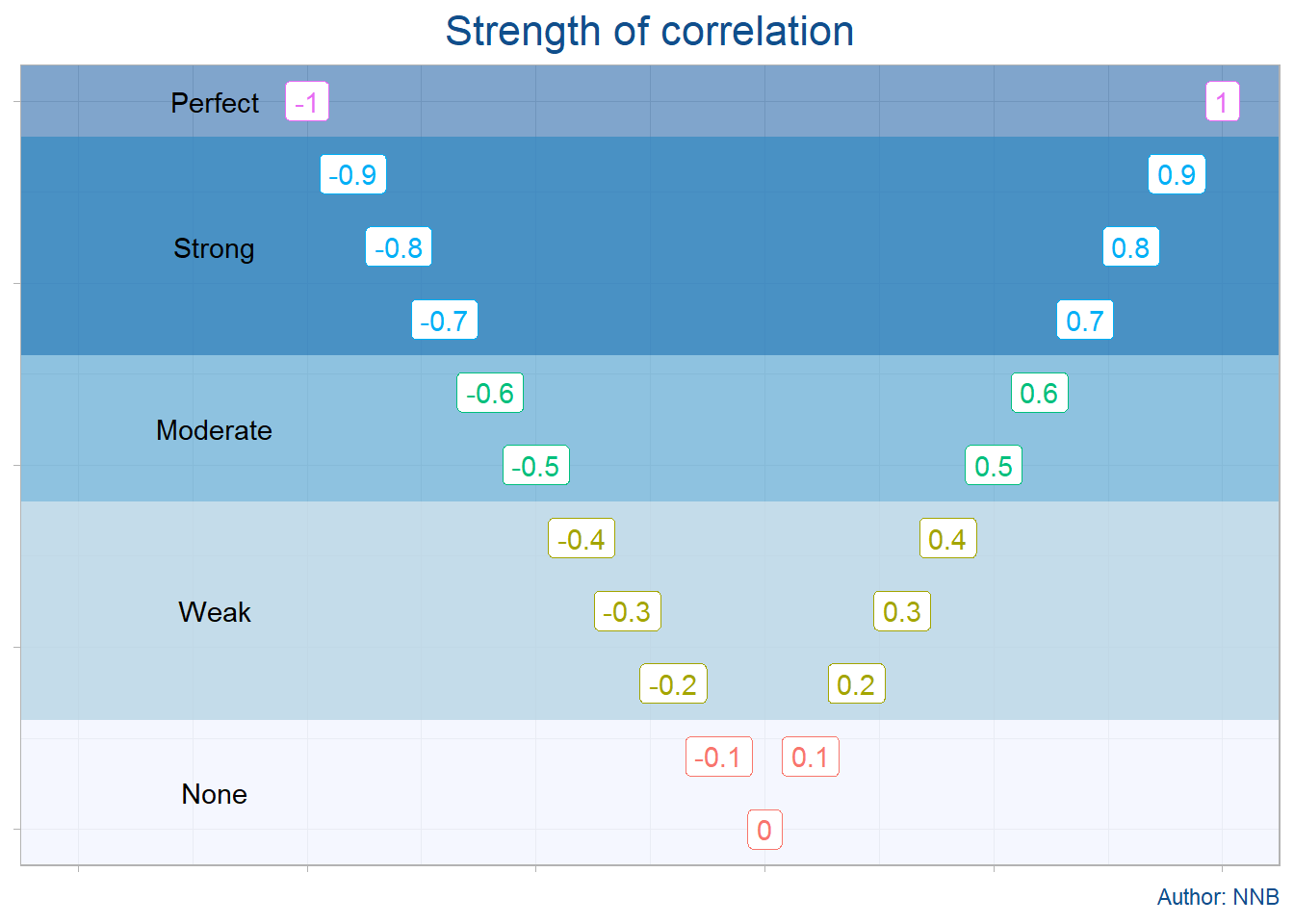
*library(httr)*

*library(stringr)*

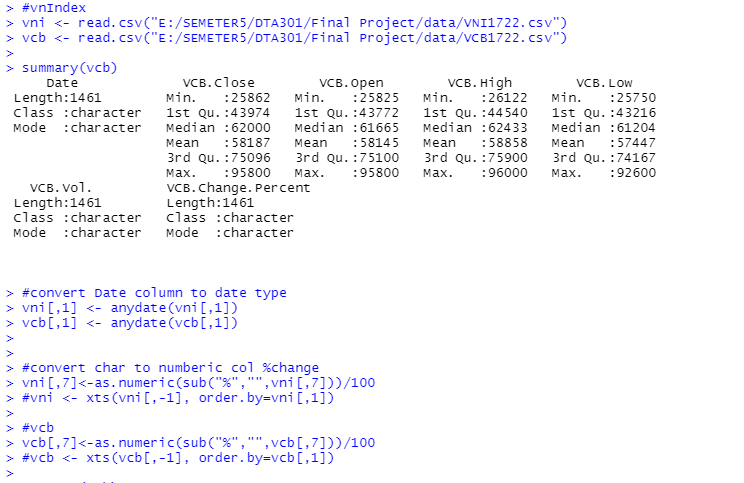
*library(magrittr)*

*library(dplyr)*

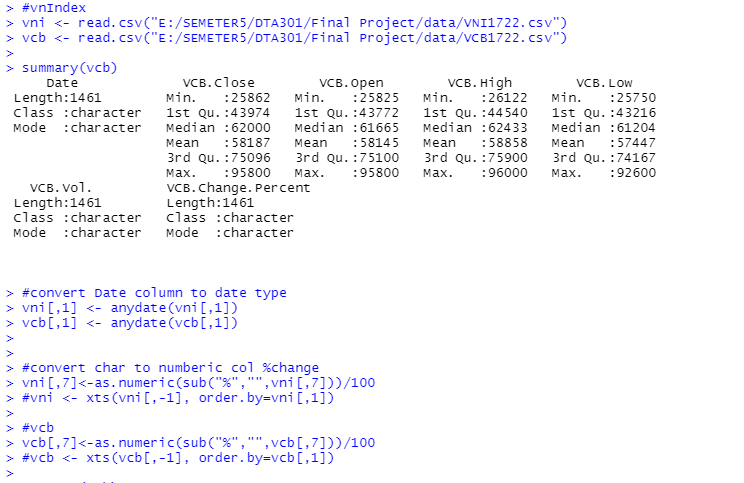
* Models used linear regression model with methods:
* Pearson
* Spearman
* Kendall

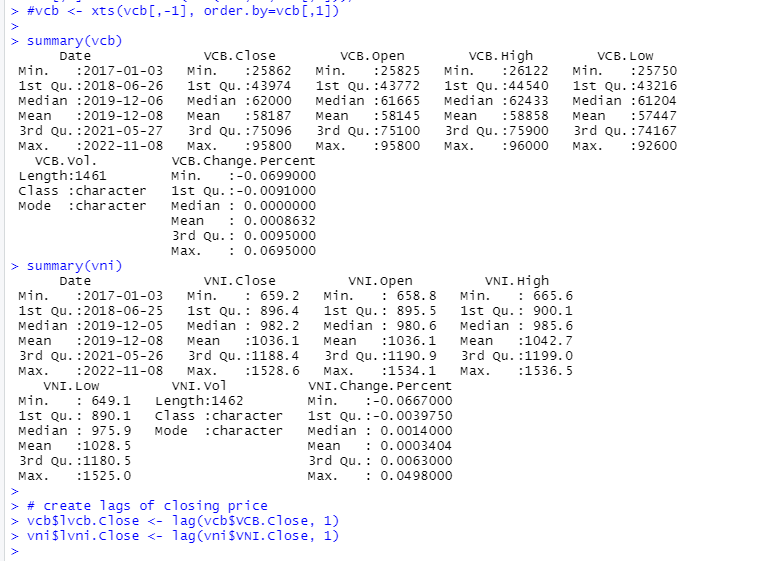


* Import data:

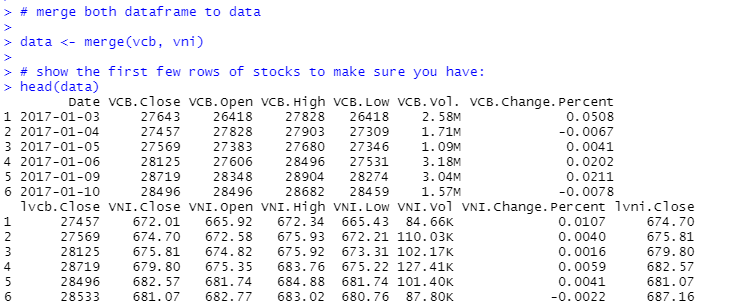


* Clean data

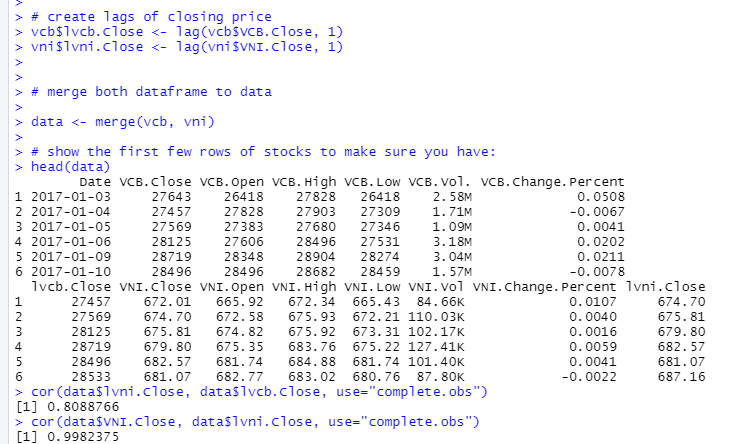




* Merge data:

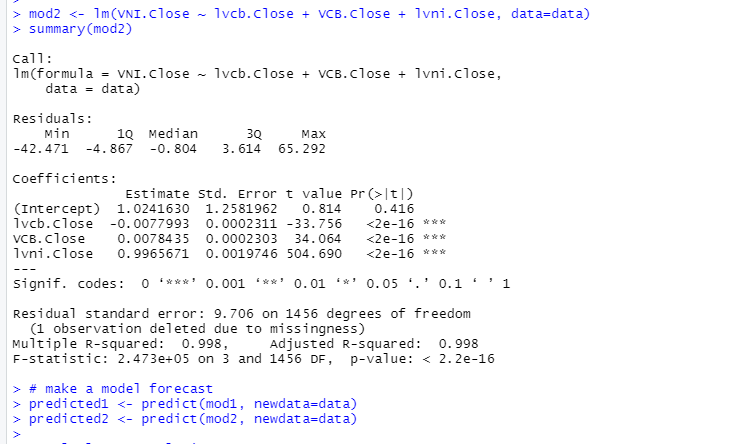
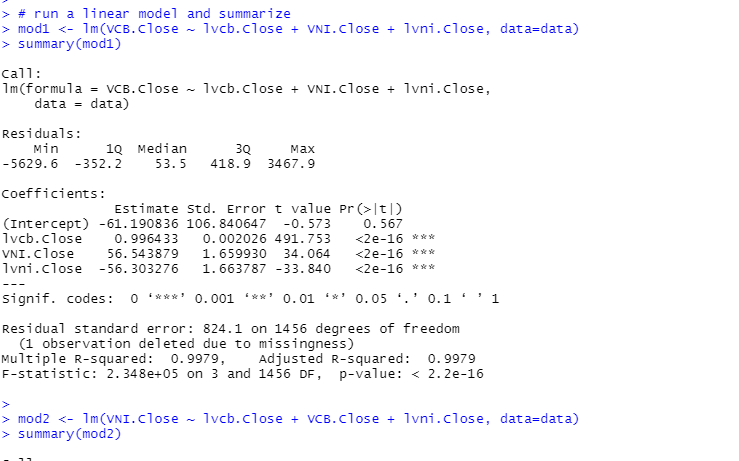


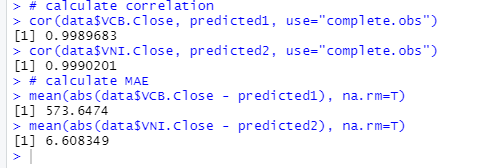
**B/ Technical Analyst**



***Based on the calculated correlation coefficients and plots, which stocks do you expect to be good predictors of each other stock price?***

Both VCB (VCB) and VN Index (VNI) can be expected to be good predictors of each others stock price because of the strong positive correlation coefficients of 0.8089 and 0.998, respectively. These strong positive correlations mean that most of the time when the respective company’s stock price increases, these stock price will increase as well.





**Which variables were significant in the model?**

The only variable that was significant in this model was the lag of each other’s close price. This is because this variable has a p-value of p < 0.001.

**Did the model results match your expectations from #8? Briefly discuss.**

The model did not match my expectations from #8. I expected VCB and VNI’s stick prices to be strong predictors of each other’s stock price. Even though there was a strong correlation between the two companies’ stock price and each other’s, the lack of significance for the variables lag of VCB stock price and lag of VNI stock price indicate that these two companies are not good indicators of each other’s stock price in our model.

**Evaluate the forecast accuracy using correlation and mean absolute error (MAE). Discuss your** **model’s performance in terms of each.**

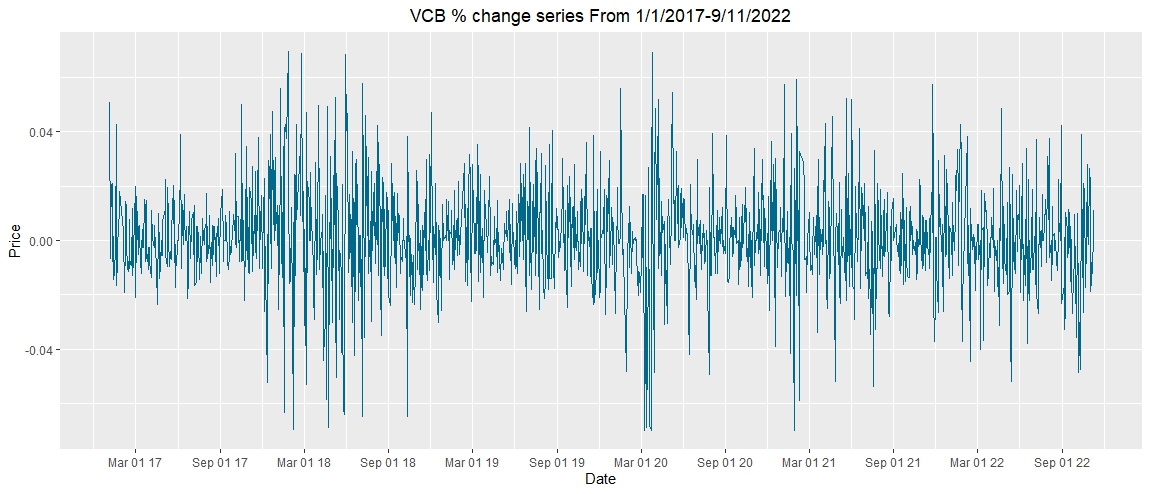
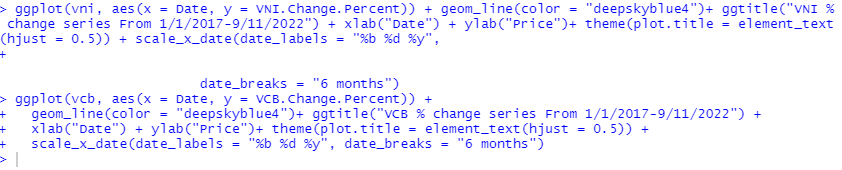
In terms of correlation, the forecast is accurate. There is a correlation coefficient of 0.999 indicating a very strong positive correlation.

The mean absolute error shows that even though there is a strong correlation between the predicted stock price and the true stock price, there is still variability. The MAE of VNI is 6.608 shows that on average the predicted stock price will be off by 6.608 VND. The MAE of VCB is 573.648 shows that on average the predicted stock price will be off by 573.648 VND7

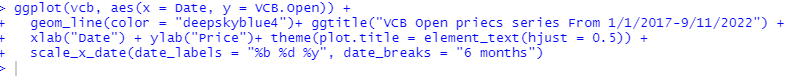
**C/ Code**

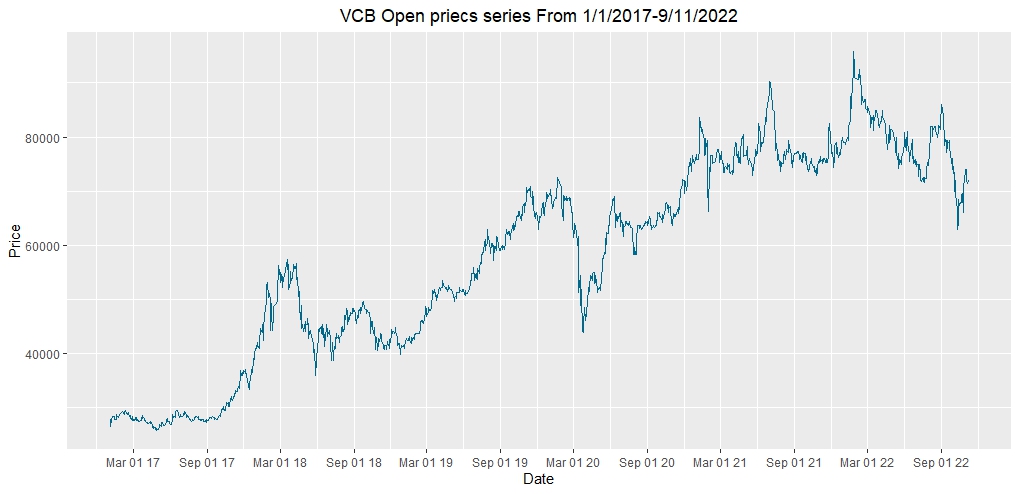
1. ***VCB***

* ***% change:***

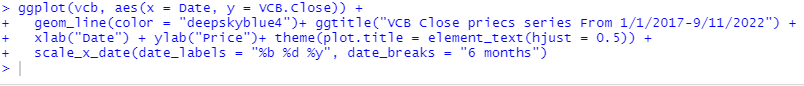
******

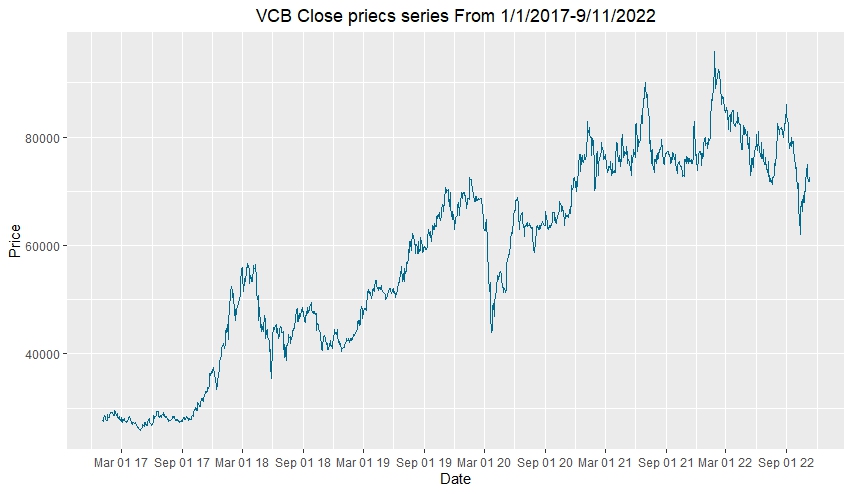
* ***Opening Prices:***

******

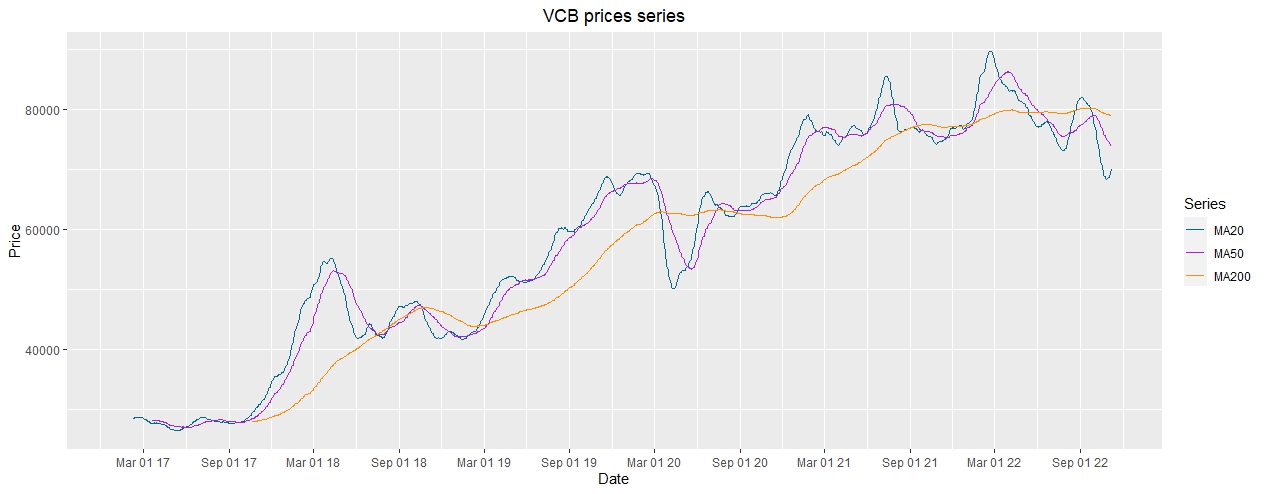
******

* ***Closing Price:***

******

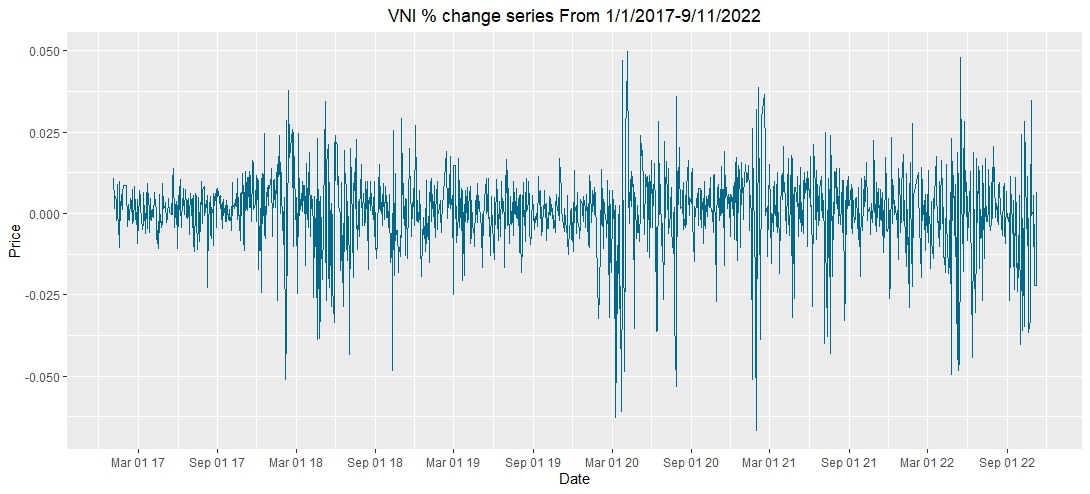
******

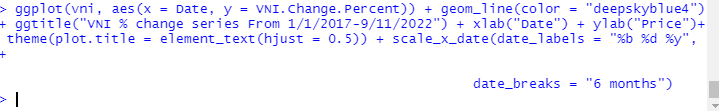
| *MA* | *#calculate three moving averages for the stock prices series,*  *#one with 20 days window and the other with 50 days and 200 days*  *vcb <- xts(vcb[,-1], order.by=vni[,1])*  *vcb\_mm20 <- rollmean(vcb[,7], 20, fill = list(NA, NULL, NA), align = "right")*  *vcb\_mm50 <- rollmean(vcb[,7], 50, fill = list(NA, NULL, NA), align = "right")*  *vcb\_mm200 <- rollmean(vcb[,7], 200, fill = list(NA, NULL, NA), align = "right")*  *vcb$mm20 <- coredata(vcb\_mm20)*  *vcb$mm50 <- coredata(vcb\_mm50)*  *vcb$mm200 <- coredata(vcb\_mm200)*  *#Ploting the prices series and the moving averages for all days since 2022:*  *ggplot(vcb, aes(x = index(vcb))) +*  *geom\_line(aes(y = vcb$mm20, color = "MA20")) + ggtitle("VCB prices series") +*  *geom\_line(aes(y = vcb$mm50, color = "MA50")) +*  *geom\_line(aes(y = vcb$mm200, color = "MA200")) + xlab("Date") + ylab("Price") +*  *theme(plot.title = element\_text(hjust = 0.5), panel.border = element\_blank()) +*  *scale\_x\_date(date\_labels = "%b %d %y", date\_breaks = "6 months") +*  *scale\_colour\_manual("Series", values=c("MA20"="deepskyblue4", "MA50"="purple", "MA200"="darkorange"))* | [*Picture*](https://drive.google.com/file/d/19Zzi0y-ZEUH1X0LuARNPtsFExThLP79w/view?usp=sharing)  *Moving Averages - the smoother the line, the less volatile*  *=> smoother the line* |
| --- | --- | --- |

******

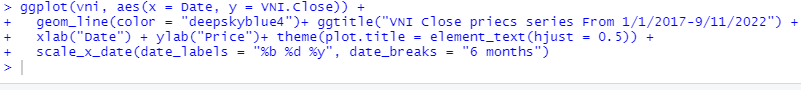
| *The volatility*  *(~20%low volatility*  *>40% is high volatility)* | *#volatility*  *>volatility=sqrt(252)\*sqrt(var(log(vcbOp[2:n]/vcbOp[1:n-1])))*  *>volatility*  *[1] 0.2948841* | *Average volatility (~29,5%)* |
| --- | --- | --- |

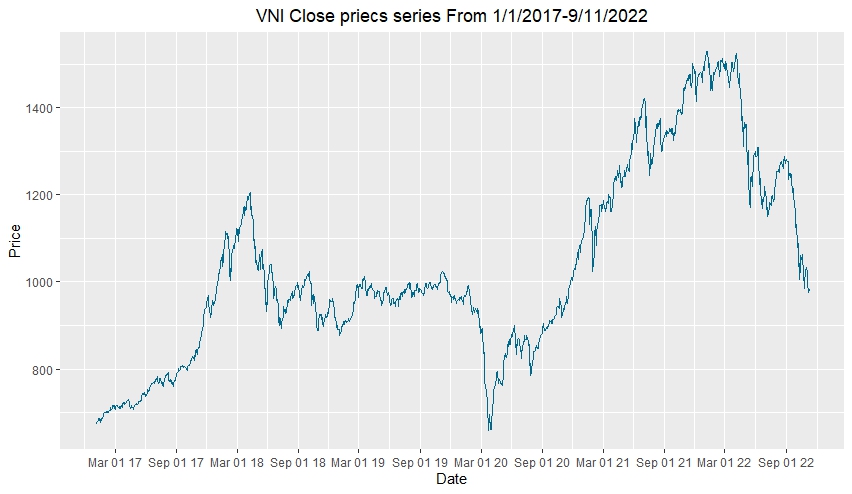
1. ***VNI***

* ***% change***

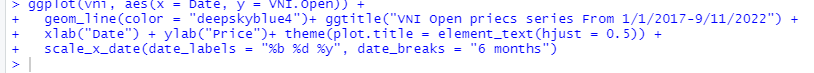
******

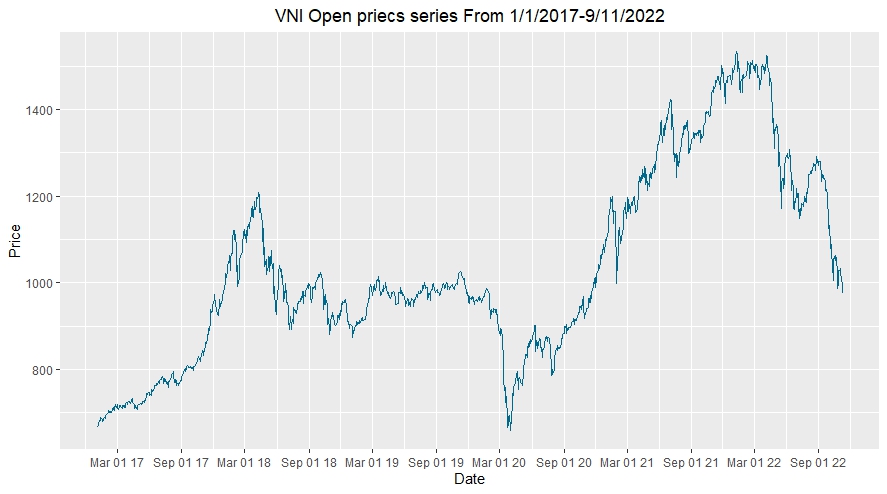
* ***Closing Price:***

******

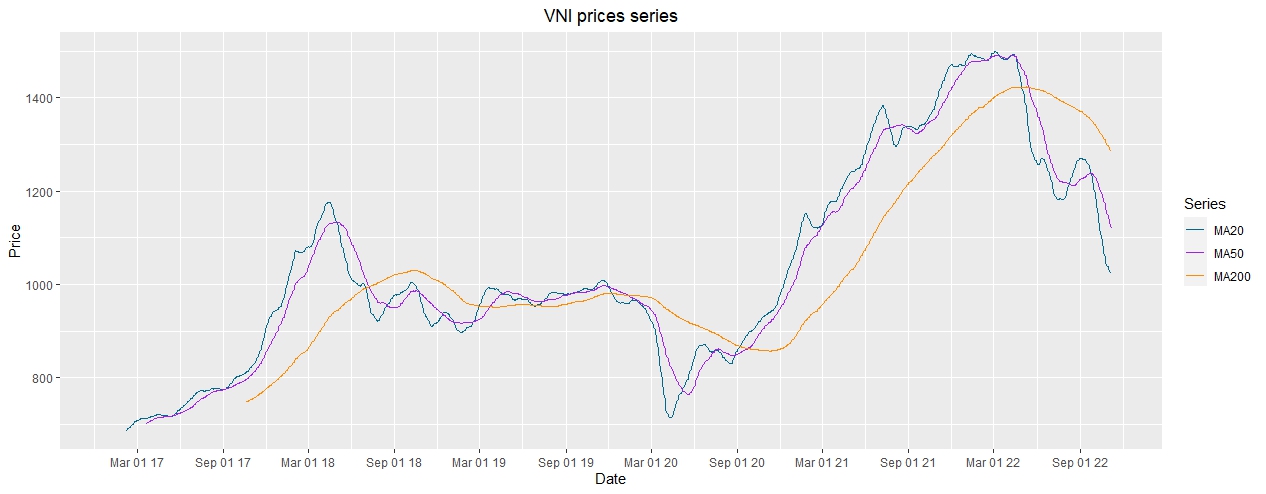
******

* ***Opening Price:***

******

******

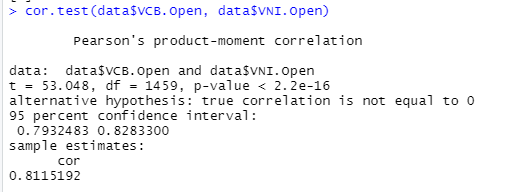
| *MA* | *#calculate three moving averages for the stock prices series,*  *#one with 20 days window and the other with 50 days and 200 days*  *vni <- xts(vni[,-1], order.by=vni[,1])*  *vni\_mm20 <- rollmean(vni[,7], 20, fill = list(NA, NULL, NA), align = "right")*  *vni\_mm50 <- rollmean(vni[,7], 50, fill = list(NA, NULL, NA), align = "right")*  *vni\_mm200 <- rollmean(vni[,7], 200, fill = list(NA, NULL, NA), align = "right")*  *vni$mm20 <- coredata(vni\_mm20)*  *vni$mm50 <- coredata(vni\_mm50)*  *vni$mm200 <- coredata(vni\_mm200)*  *#Ploting the prices series and the moving averages for all days since 2022:*  *ggplot(vni, aes(x = index(vni))) +*  *geom\_line(aes(y = vni$mm20, color = "MA20")) + ggtitle("VNI prices series") +*  *geom\_line(aes(y = vni$mm50, color = "MA50")) +*  *geom\_line(aes(y = vni$mm200, color = "MA200")) + xlab("Date") + ylab("Price") +*  *theme(plot.title = element\_text(hjust = 0.5), panel.border = element\_blank()) +*  *scale\_x\_date(date\_labels = "%b %d %y", date\_breaks = "6 months") +*  *scale\_colour\_manual("Series", values=c("MA20"="deepskyblue4", "MA50"="purple", "MA200"="darkorange"))* | [*Picture*](https://drive.google.com/file/d/1OucDVgqZel2BhVMHnT1wRASEwri8Ha3O/view?usp=sharing)  *Moving Averages - the smoother the line, the less volatile*  *=> smoother the line* |
| --- | --- | --- |

**

| *Volatility*  *(~20% low volatility*  *>40% is high volatility)* | *attach(vni)*  *vniOp = rev(vni$Open)*  *n=length(vcbOp)*  *#bien dong*  *volatility=sqrt(252)\*sqrt(var(log(vniOp[2:n]/vniOp[1:n-1])))*  *volatility*  *[1] 0.2323556* | *Average volatility (~23,23%)* |
| --- | --- | --- |

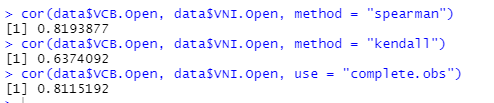
1. **Relationship between stock ticker of Joint Stock Commercial Bank for Foreign Trade of Vietnam (VCB) and VN index (VNI )**

* **Open prices:**

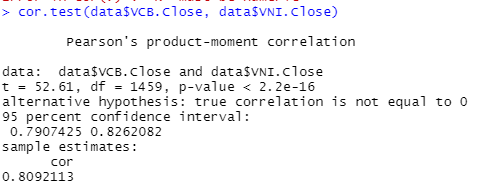


* Test hypothesis:
  + H0: No correlation (correlation coefficient = 0)
  + Ha: There is correlation
* Test method: Pearson
* => Result: There is correlation, the correlation coefficient is positive
* => Opening prices of VCB and VNI both increased/decreased

Similarly, when we use other testing methods such as kendall (rank), spreaman (rank) all give results of strong correlation or correlation.

**

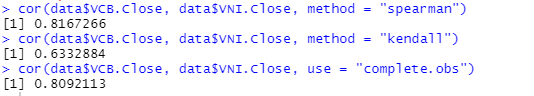
* **Close prices**

****

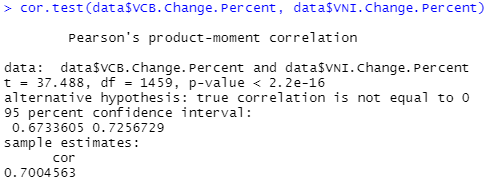
- Test hypothesis:

* + H0: No correlation (correlation coefficient = 0)
  + Ha: There is correlation
* Test method: Pearson
* => Result: There is correlation, the correlation coefficient is positive
* => Closing prices of VCB and VNI both increased/decreased

Similarly, when we use other testing methods such as kendall (rank), spreaman (rank) all give results of strong correlation or correlation.

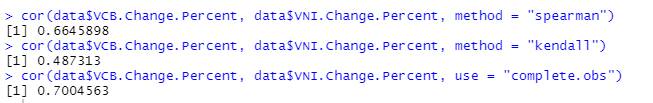
**

* **% Change:**

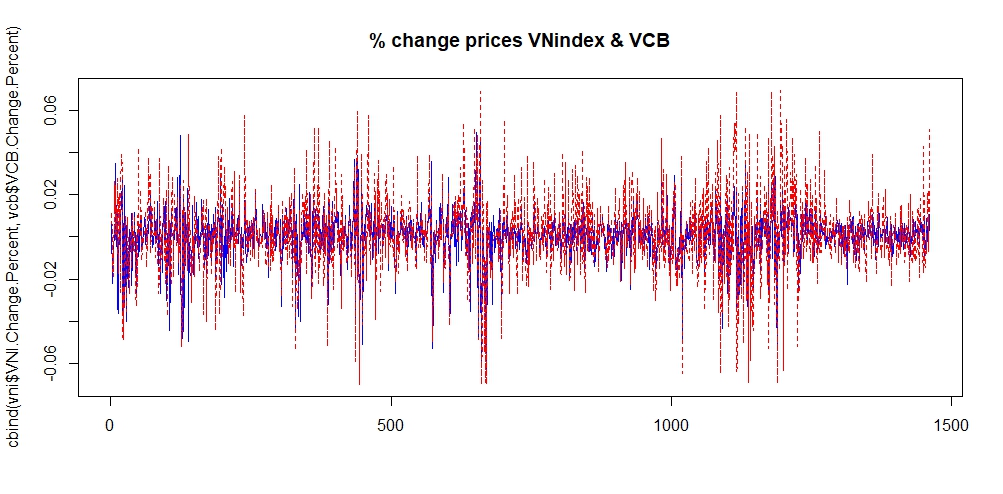
****

* Test hypothesis:
  + H0: No correlation (correlation coefficient = 0)
  + Ha: There is correlation
* Test method: Pearson
* => Result: There is correlation, the correlation coefficient is positive
* => % changing of VCB and VNI both increased/decreased

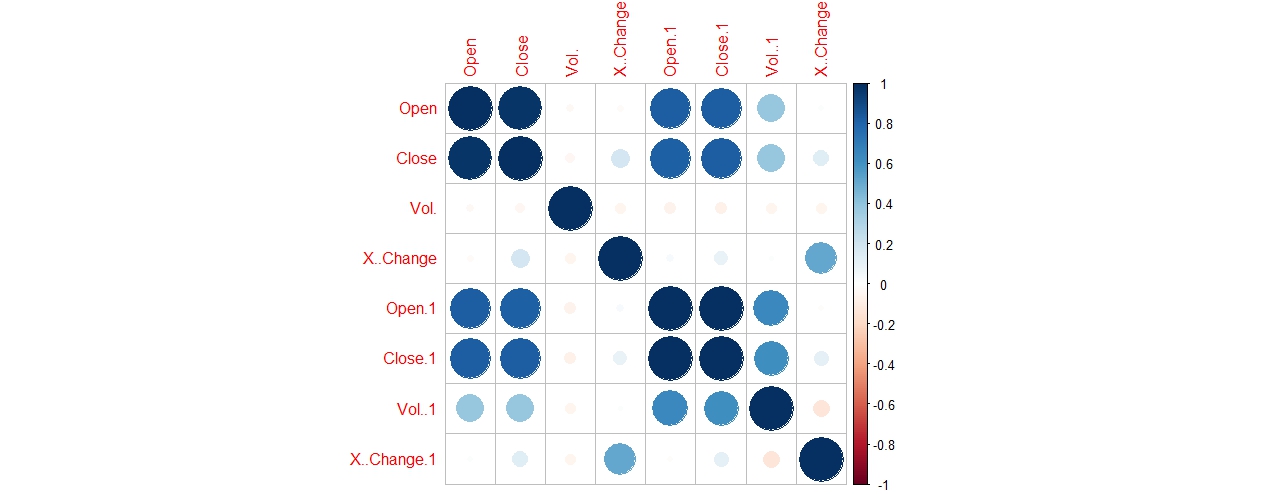
Similarly, when we use other testing methods such as kendall (rank), spreaman (rank) all give results of strong correlation or correlation.

**

| *Name* | *Code* |
| --- | --- |
| *% change* | *matplot(cbind(vni$VNI.Change.Percent,vcb$VCB.Change.Percent),type="l",col=c("blue","red") ,main="% change prices VNindex & VCB")* |



| *Correlation graph* | *stocks <- merge(vcb, vni)*  *corstocks <- as.data.frame(stocks[,c(1,4,5,7,8,11,12,14)])*  *corstocks %>% head()*  *M <- cor(corstocks)*  *corrplot(M, method = "circle")* |
| --- | --- |

****