COVID-19 effect on CBOE Volatility Index and real S&P500

Alisa Balakirska

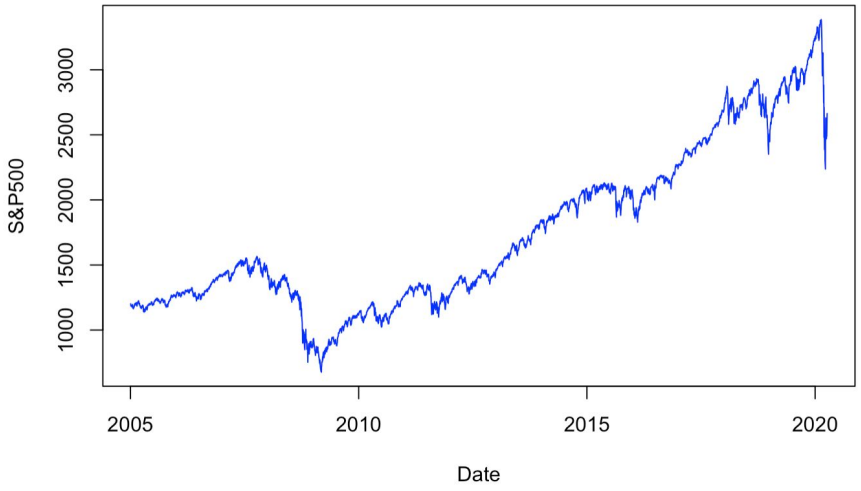
Idea

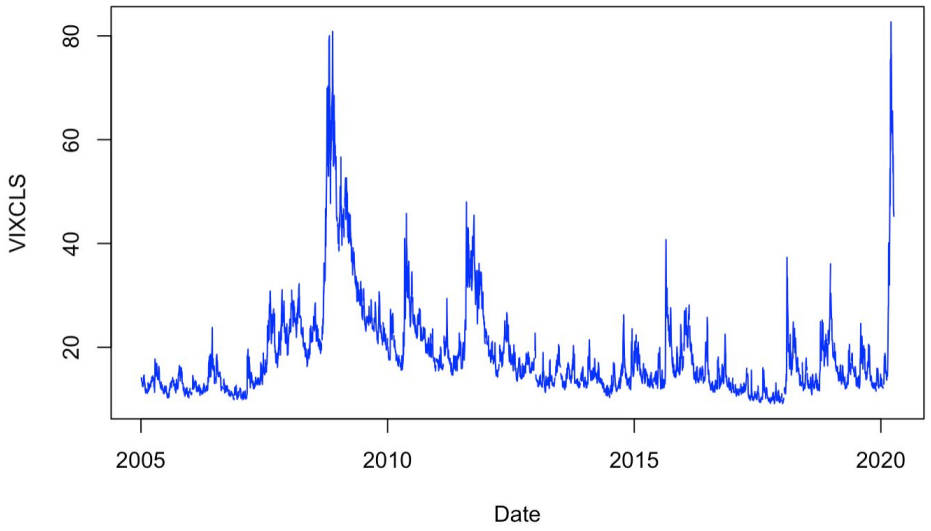
The main idea of my project is to research the COVID-19 effect on CBOE Volatility Index and real S&P500. I want to do this research to find out how can I predict future S&P500 considering tendencies of COVID-19.

Data description You can find in the appendix.

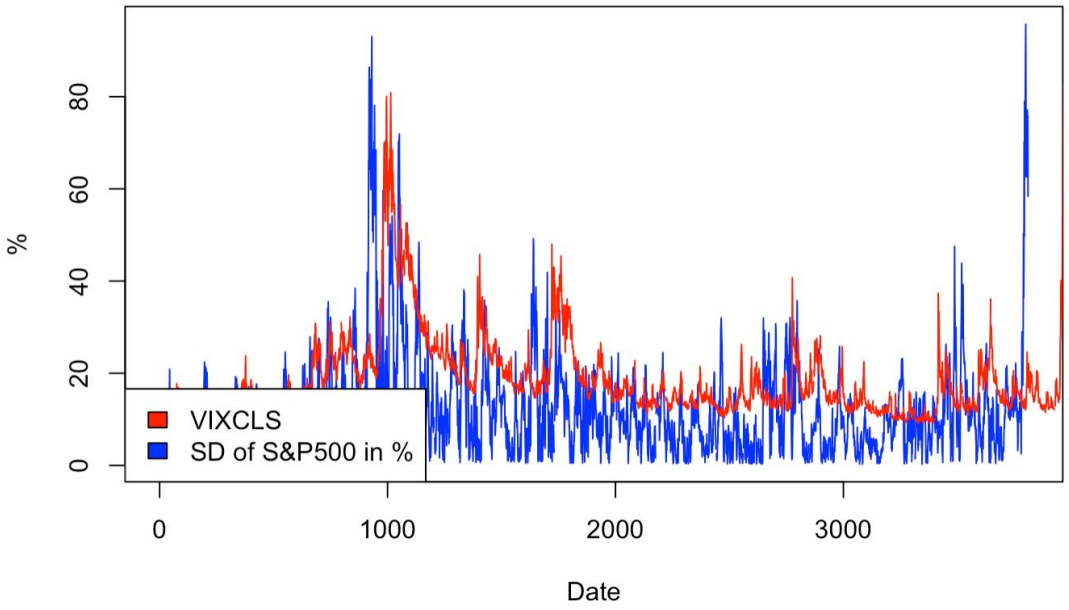
Research

I plotted S&P500 and VIXCLS to see periods of crises and usual good periods.





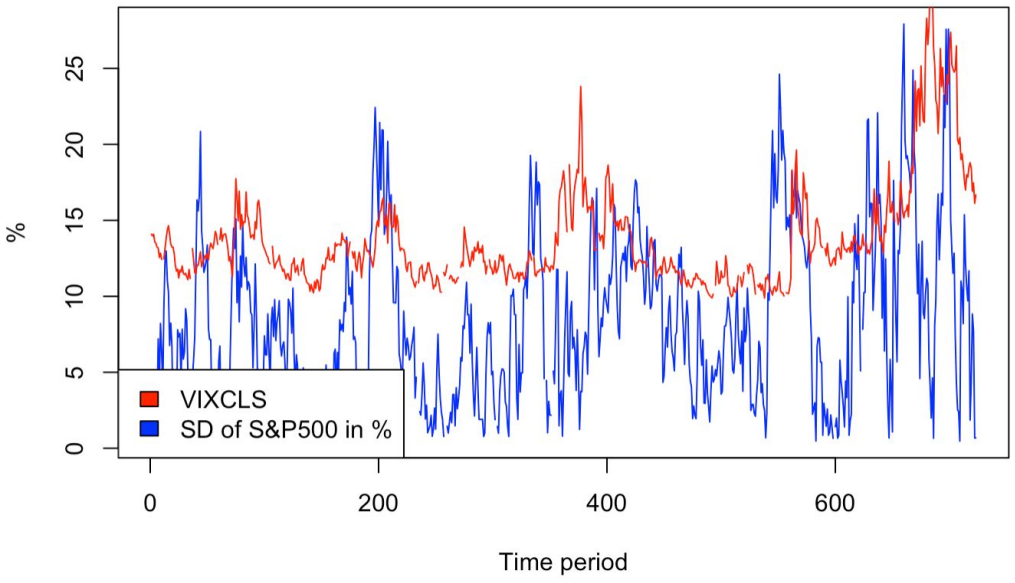
For each date in S&P500 data I found the percentage of standard deviation for the next 31 day - as like as in VIXCLS index - the prediction for percentage of volatility of S&P500 for the next month. I plotted both of them, You can see it below.



I saw that in different time periods this two indexes have different difference - they have peaks in different time, especially now - in the end of the plot.
That is why I wanted to research how good is VIXCLS to predict the volatility of S&P500 in crises and in usual good times.

I divided all data by 4 periods - from 2005 to 2008 and from 2013 to 2016 - time periods without crises; from 2008 to 2010 and now - time periods of crises - and did research for them.

Time period from 2005 to 2008

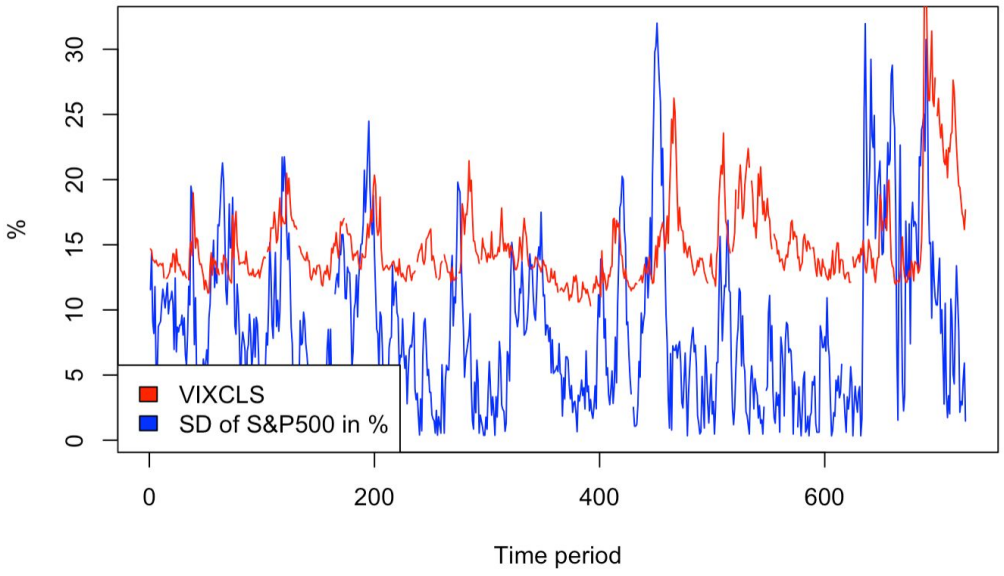


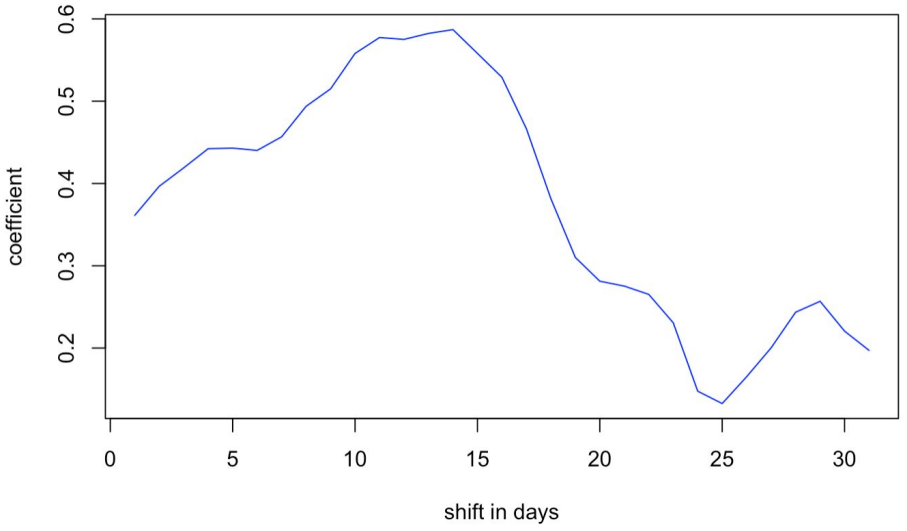
On this plot I noticed that VIXCLS prediction is good, but it has some lateness in comparison with percentage of SD of S&P500 for next month. That is why I decided

to shift the data by 1 day on each iteration, then build linear model, and then compare for how many days shift VIXCLS has the largest effect on percentage of SD of S&P500 for next month.
Here You can see the plot of coefficients for this shift.

As You can see, the largest effect is when we use shift equal in 12 days. Then I checked the significance and found out that it is significant(appendix 1).
I did the same actions with all time periods.

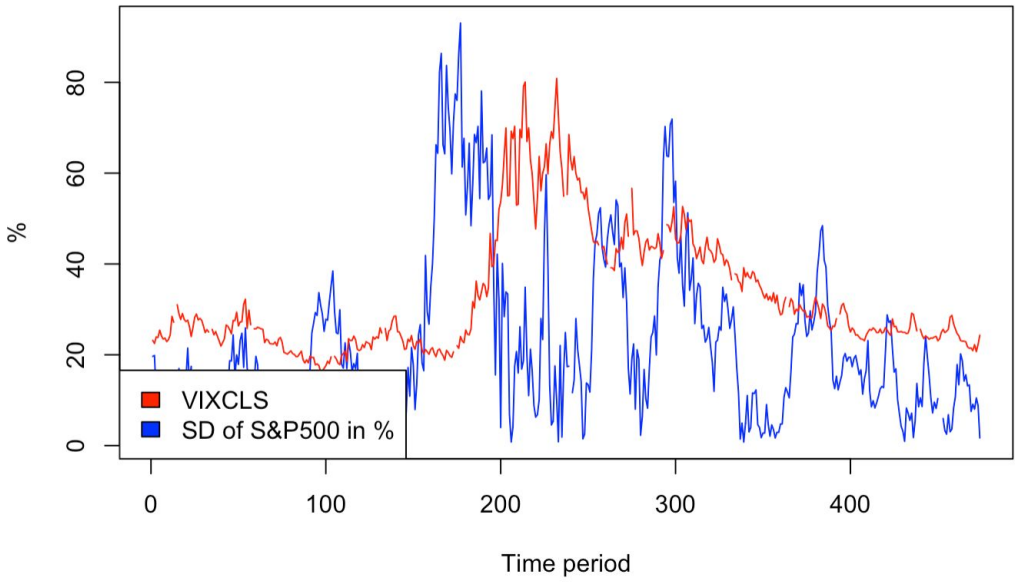
Time period from 2013 to 2016

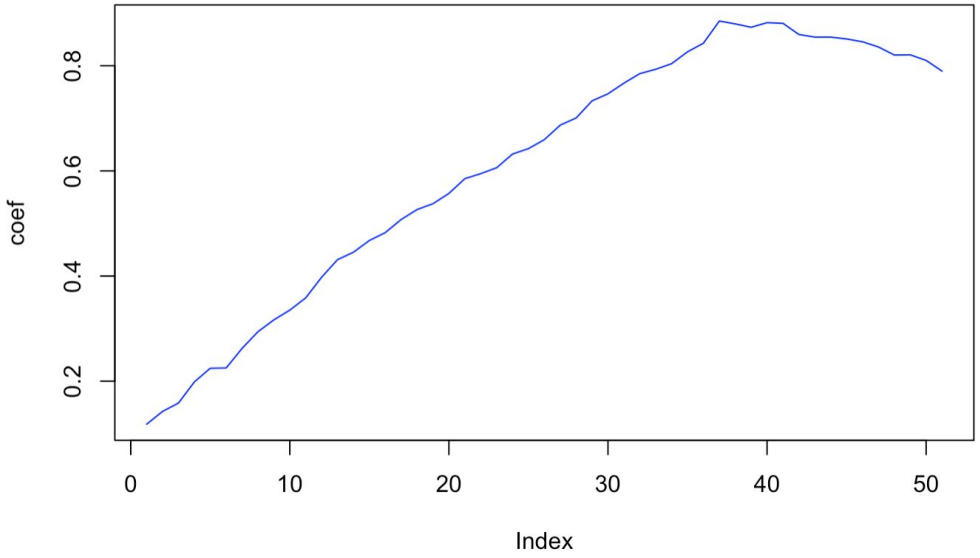




So, we can see that the largest coefficient is when the shift is equal to 14 days, it is also significant(appendix 2).

Time period from 2008 to 2010

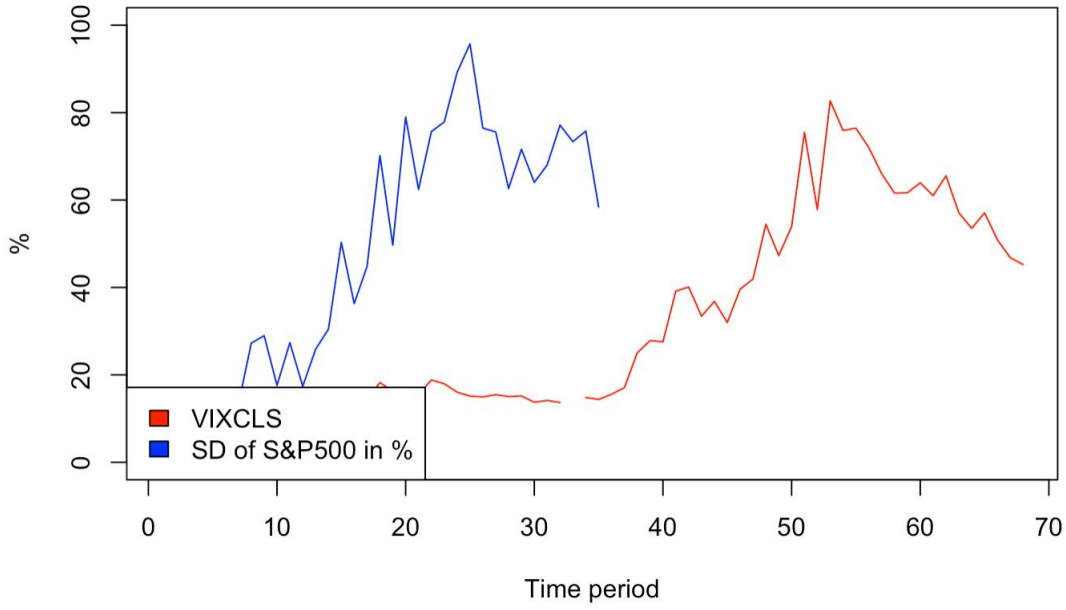




We can see that largest value is when the shift is equal to 37 and it is also significant(appendix 3).

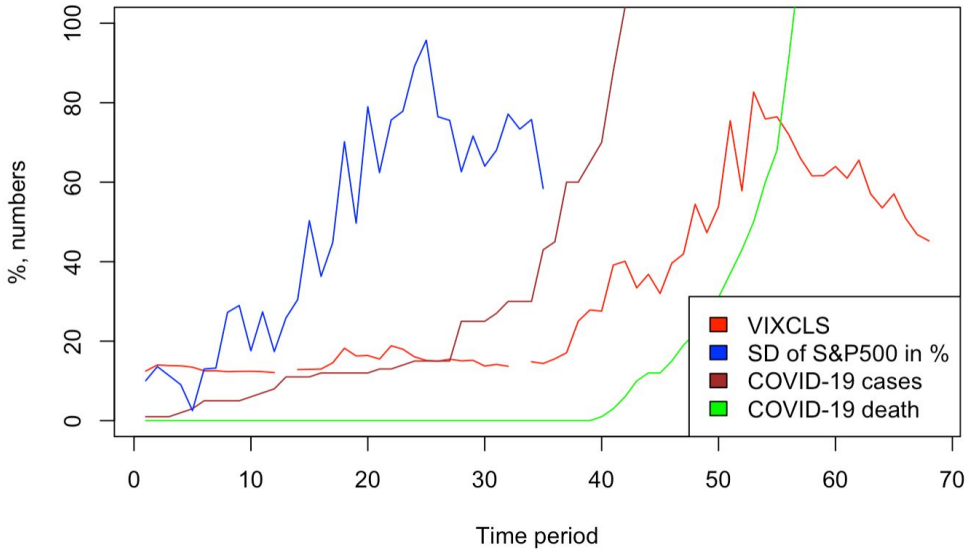
In the time of crisis the intercept is also significant and it has negative sign - it means that people have better expectations, than it is in reality.
Concluding the results: in usual good times the prediction is late by 12-15 days and during crisis the prediction is late for 37 days - so, the market crashes on the average by 37 days earlier, that people notice it. Let's take a look at the current situation.

Time period from January 2020 to 12th May 2020

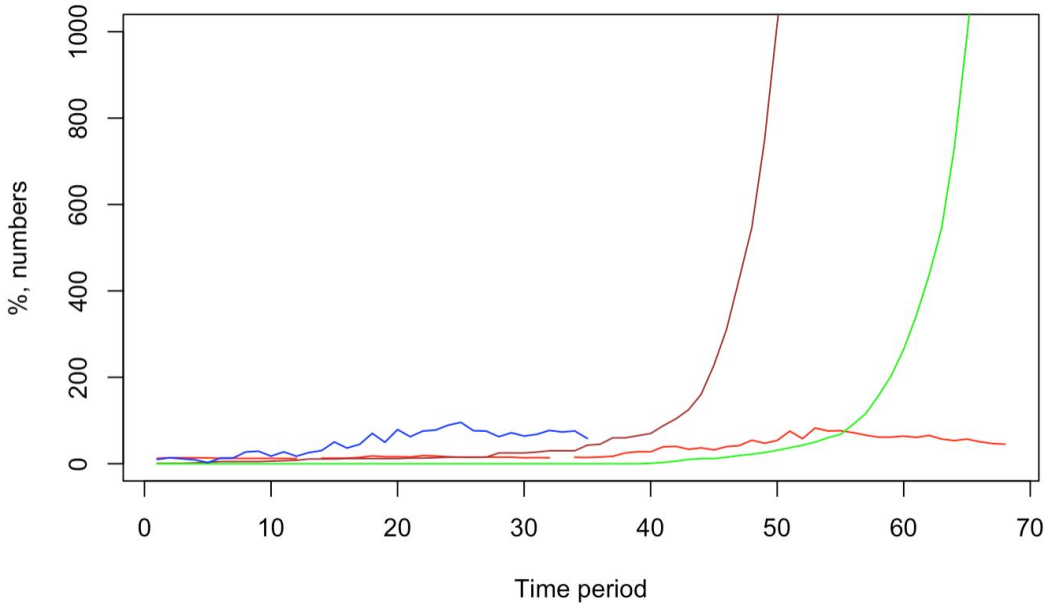


So, difference between peaks is equal to 30 days.

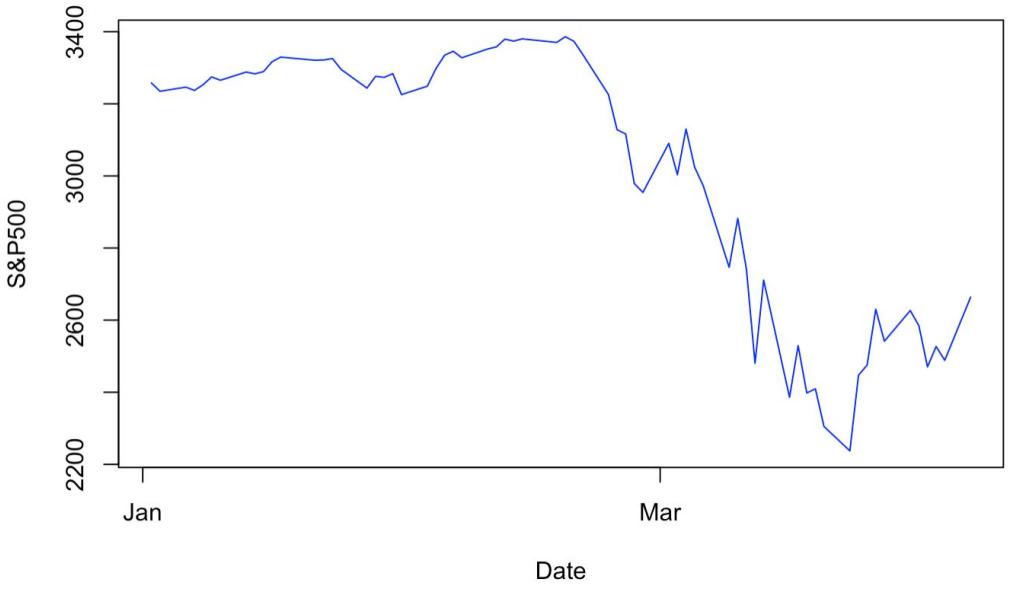
Let's take a look at number of the COVID-19 cases and death in USA.



Let's change the limit of y to see clearly:

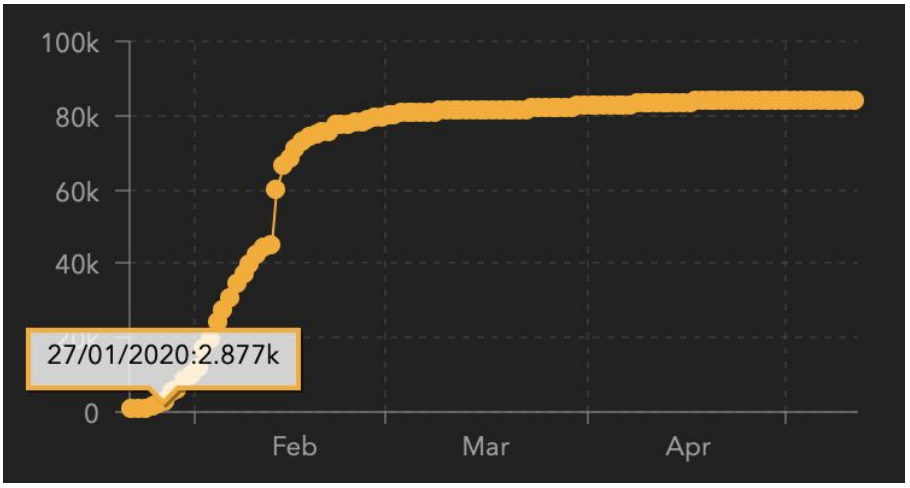


Below You can see the plot of S&P500 from January to May 12th:



So, the market crashed before COVID-19 came in USA - as Wikipedia says that the first case was confirmed on 20th January 2020 and the market started to crash on the January 27th.
The expectation of volatility increased only when it came in USA and the number of cases increased sharply.

I had a hypothesis that it crashed because of the number of cases in China, below You can see the plot of distribution of coronavirus in China(appendix 4)



On the January 27th in China there were not very much cases, China had not any transport restrictions or closed borders, but from January 27, China suspended sending groups of tourists abroad as part of measures to prevent the spread of coronavirus(appendix 5).

In all internet resources the decrease of market in USA is explained by decrease in price of the oil(appendix 6), because there were decrease in transportation influenced by COVID-19. But we can see that there were no restrictions due to COVID-19 when the prices started to fall.

Conclusion

So, we can notice the disbalance of information: prices started to fall before coronavirus started to spread, that is why we can assume that there are other reasons for this crisis.
As I saw that there are other factors which have impact on S&P500, I decided that the prediction of it only by using coronavirus will not be correct.

Appendix

Data description
In my project I used three dataset:

- S&P500 - a stock market index that tracks the stocks of 500 large-cap U.S. companies

  Source: https://fred.stlouisfed.org/series/SP500
- VIXCLS - VIX measures market expectation of near term volatility conveyed by stock index option prices

  Source: https://fred.stlouisfed.org/series/VIXCLS

- Data on COVID-19 in USA

  Source: https://github.com/nytimes/covid-19-data

1.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.48780    0.82812   0.589    0.556
befc2        0.57785    0.05846   9.884   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

2.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.22584    1.05017  -0.215     0.83
befc2        0.58699    0.06866   8.550   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

3.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -6.765      1.917  -3.529 0.000465 ***
befc2          0.885      0.052  17.017  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

4. Source: https://coronavirus.jhu.edu/map.html

5. Source: https://www.pravda.com.ua/rus/news/2020/01/27/7238502/

6. Source: https://www.vedomosti.ru/business/articles/2020/03/16/825314-tsena-nefti