New Study on Monte Carlo Based VaR: Managing Weekly Change Data with Linear Regression, k-NN and Developing an Alternative VaR

Hüseyin Albay, Furkan Demirbaş Betul Demirkol, Uğurcan Ocak

Summer 2019 – EC 48W Term Project

Bogazici University, Istanbul, Turkey

ABSTRACT

In this study, we as a team decided to develop an efficient calculation for Monte Carlo based VaR. Since the Monte Carlo Simulation with 95% confidence interval provides a wide area that can not be used for trading decisions, we have derived our own maximum expected loss with linear regression. We have used DXY and USD 3Y IRS values as our dataset, and we have shown some methods regarding linear regression an k-NN analysis on these datasets. As a result, Monte Carlo Simulation showed an efficient result for DXY but for IRS, its performance is weak due to different interest rates move environments due to different policies.

1.INTRODUCTION

In daily life, human beings are constantly faced with ambiguity and uncertainty. Since individuals cannot rule over all information, risk analysis become a basic process that we attend every day. In order to be included better decision-making process under uncertainty, Monte Carlo Simulation is widely being used, mostly in financial decision-making process. Monte Carlo Simulation which is a computerized mathematical technique helps us to clearly understand the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables and the technique is used by professionals in such widely disparate fields as finance, project management, energy, manufacturing, engineering, research and development, insurance, oil & gas, transportation, and the environment.

In this study, Monte Carlo Simulation is used for computing the expected weekly maximum loss with the change of DXY and IRS change within %95 confidence interval by using mean and standard deviation of weekly change of DXY and IRS. To understand how the logic was built in this project one should know the data that consist of 9 years DXY, IRS and US 3 years Treasury bond yield. An interest rate swap is the derivative contract between two parties (mainly banks, investors and corporations) that ensures the exchange of all future interest_rate payments forthcoming from a bond or loan. On the other hand, treasury yield is the interest rate that the government pays to borrow money for different lengths of time. Therefore, these two variables are positively correlated. Moreover, DXY represents the US dollar index which is a measurement of the dollar's value relative to foreign currencies. Therefore; these three parameters are interconnected with one and other. The study has conducted between DXY and IRS because in the first part, it had been discovered that IRS and USDT 3Y already have 0.99 correlation.

Thanks to machine learning algorithms, this project mainly aims to see whether machine learning algorithms help us to calculate a better Monte Carlo distribution. In the case of DXY, getting a narrower distribution helps us to have better decision. In IRS case, since interest rates does not have a graph like a currency or stock, wider Monte Carlo distribution does not necessarily say that results are bad. Moreover, it can give a hint on rates trading strategy in decreasing or increasing rate environments.

2.METHODS REVIEW AND BACKGROUND

Machine Learning refers to a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed and Python community has developed many modules to help programmers implement machine learning such as numpy, pandas, MatPlotLib, Scikit-Learn, Collections, SciPy etc. There are several types of machine learning algorithms but in this study, we used supervised learning that consist of a target, dependent variable which is to be predicted from a given set of independent variables. Using these set of variables, we generate a function that map inputs to desired outputs. KNN and logistic regressions are one of the examples of supervised learning. Hence; there are many Machine Learning algorithms and each of them have some advantages and disadvantages, in the study the most accurate algorithms were tried to use to predict an output. First of all; the linear regression tries to build a relationship between dependent and independent variable to find best fit and a linear equation: $Y= a *X + b$. These coefficients a and b are derived based on minimizing the sum of squared difference of distance between data points and regression line. It is mainly used for estimating the real values based on continuous variables. Since, change in DXY is a real value and based on continuous variable like time and change in IRS we conducted simple linear regression model in our study. On the other hand; we use also KNN (k-Nearest Neighbors) method to make a prediction with our data. KNN can be used for both classification and regression problems. KNN can easily be addressed to our real lives. For instance; if you want to learn about a person, of whom you have no information, you might like to find out about his close friends and the circles he moves in and gain access to her information. Therefore; this method was very suitable for our data set because we tried to learn which years have the same trends and which value is included to which year. By doing so, the study aimed to examine the weekly change predictions and their relevance with Monte Carlo results. In this study, it was used for grouping our data among years and understand the trend and differences between years.

3.DATA CLEANING AND RECONSTRUCTION

We take out data from Reuters and cleaning and restructuring data are commonly done through Microsoft Excel. DXY, IRS and USD 3Y Treasury Bond data were merged. The data is cleaned and reconstructed because time period of our parameters was different from each other. There were many #N/A due to the fact that our DXY and IRS data do not have values on weekend and holidays but USD 3Y Treasury Bond data has. Hence, we filled these #N/As with previous week's Friday data. In the end, we reached three parameters and 2349 days' data. Mainly we have the daily data between 2010 and 2019. To reach weekly data we find weekly change of DXY and IRS by subtracting Monday's data from Friday's data and dividing it to Monday's data. To do so we reach 469 weeks data and weekly change of our parameters. Since IRS and 3 years treasury yield have the same trend we do not use USD 3Y Treasury Bond data. To use machine learning algorithms, we import our data to Python. After that; we splitted the data to 40 percent to 60 percent as test and train data respectively. We tried to use first 261 weeks data to measure if we could predict the next, 262th week's change and so on like and compare with Monte Carlo's results. No outliers were detected and so not dropped.

4.DATA EXPLORATION

Below the features of the data we have used:

- The data frame named "data" shows the daily USD IRS, USD 3Y Treasury Bond yield and DXY values daily from 2019 back to 2010, 2345 days.

```
data = pd.read_excel("EC48W.xlsx")
```

```
data.head()
```

	Timestamp	IRS	USD T	DXY
0	2019-07-19	1.7799	1.780	97.151
1	2019-07-18	1.7311	1.750	96.794
2	2019-07-17	1.7848	1.807	97.223
3	2019-07-16	1.8350	1.844	97.395
4	2019-07-15	1.8090	1.812	96.933

```
data.describe()
```

	IRS	USD T	DXY
count	2345.000000	2345.000000	2345.000000
mean	1.320922	1.147684	88.204290
std	0.708818	0.720709	8.385531
min	0.420000	0.279000	72.933000
25%	0.809000	0.625000	80.252000
50%	1.099000	0.949000	89.623000
75%	1.689000	1.480000	96.077000
max	3.185000	3.045000	103.300000

- The data frame named “w_change” shows the weekly USD IRS and DXY change from 2019 back to 2010, 439 weeks.

```python
#WE HAVE COMPLETED DATA EXPLORATION
#THE WEEKLY CHANGE DATA WE HAVE CONSTITUTED ON EXCEL
w_change = pd.read_excel("change.xlsx")
```

```python
w_change.describe()
```

	Week	Weekly ChangeDXY	Weekly ChangeIRS
count	469.000000	469.000000	469.000000
mean	235.000000	0.000425	0.003932
std	135.532899	0.009357	0.065461
min	1.000000	-0.023852	-0.215608
25%	118.000000	-0.005527	-0.034243
50%	235.000000	0.000246	-0.001401
75%	352.000000	0.006502	0.033333
max	469.000000	0.029722	0.392801

```python
w_change.head()
```

	Date	Week	Weekly ChangeDXY	Weekly ChangeIRS
0	2010-07-26	1	-0.006676	-0.102456
1	2010-08-02	2	-0.006561	-0.039585
2	2010-08-09	3	0.027716	-0.035951
3	2010-08-16	4	0.006349	0.037037
4	2010-08-23	5	-0.002466	-0.009278

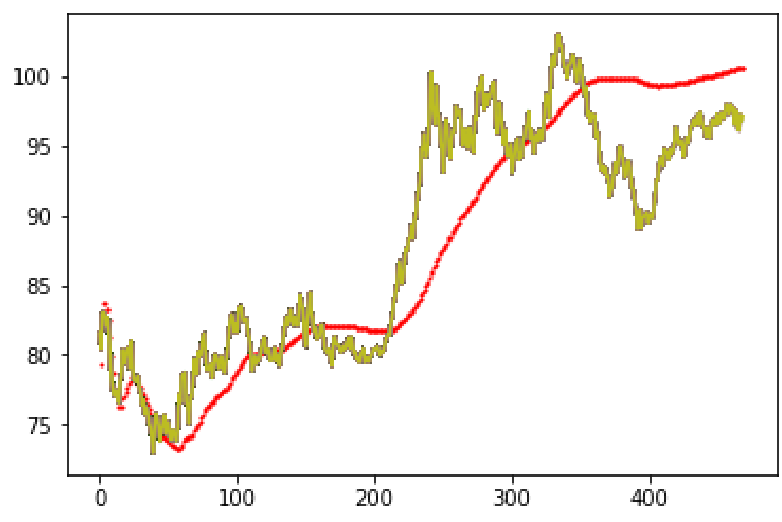
➢ The data frame named “w_change” shows the weekly USD IRS and DXY change from 2019 back to 2010, 439 weeks.

5.METHODS AND TOOLS

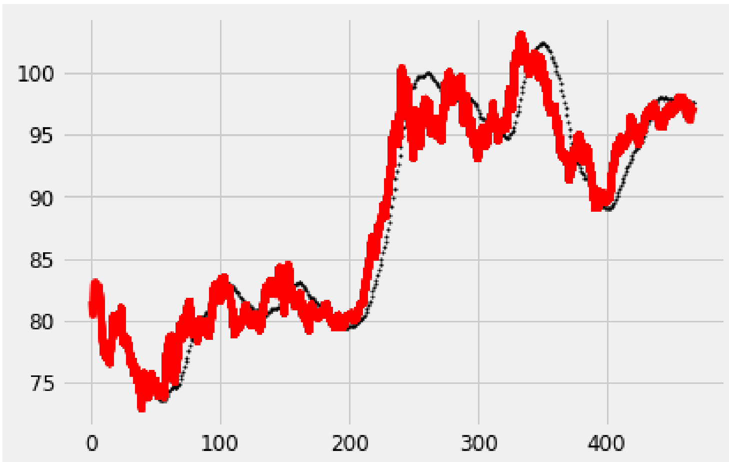
1.Linear Regression

Our goal in linear regression is to find the relation between two different data and see the general trend which can predict the future returns. First, we take IRS and USD Treasury Bond data. Since their correlation is too high it returns a line almost on the data. Then, we use IRS and DXY's day and week data to understand the relationship between them. Unlike the previous data, it has bigger deviation from the line. So, we take linear regression one step further.

In Week1 we find the linear regression from only that data and it predicts the next week same with the previous week. But with the new data, the regression improve itself and every week after adding a new data it finds next week's prediction again. First, we use all data from DXY column. Every week we evaluate the linear regression repeatedly and with this line it predicts the next week's price from all the data before that week. This gives us a system which learns from the new data and last result predicts from all weeks' data.



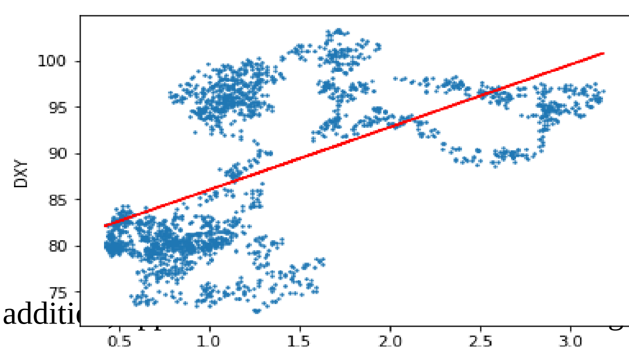
Yellow bins show the DXY's close in weeks which are in x-axis. Red points are the prediction of linear regression comes from previous weeks' data. However, we see some data will affect the result in a bad way. Because our data comes from almost 10 years and general trend can be changed. So, we use only 52 weeks from our data before the week we predict and with every new data it drops the first data and evaluate linear regression again.



From black points, we can see it follows trend in DXY better and quicker because it uses less but right data instead of more complex calculations.

2. k-NN

Comparing the relationship of IRS and DXY values according to the yearly periods will provide us with the overall knowledge about the economic and political decisions of these periods. This overview can be important while making prediction about those periods or and examining the data belong to those periods because it can be seen that there are obvious clusters.

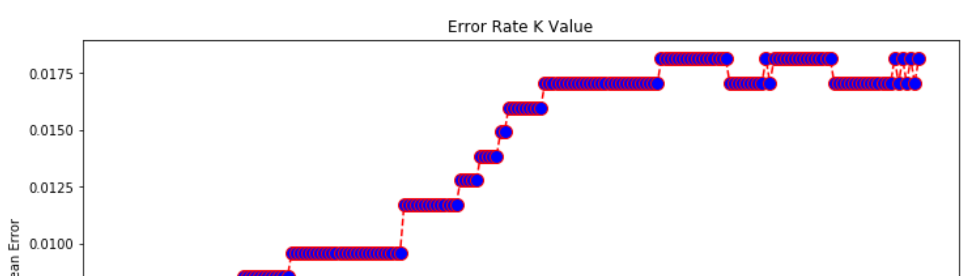


le to provide knowledge about the periods. In additi borhood method handy. When we set the periods as 2010-2015, 2015-2017, 2017-2019-year intervals, the obvious clusters are separated clearly. The specific feature of the period 2010-2015 is the permanent interest rate hike especially in United States and all over the world economies. At the period 2015-2017, the interest rates are more stable and more balanced economic policies are followed. Finally, at the period 2017-2019, the most deterministic characteristics are presidency of Donald Trump and the economic war that he provokes. As the result of those developments, economic growth slows down in this period. These specific features make these periods separable easily as we can see the figure below.

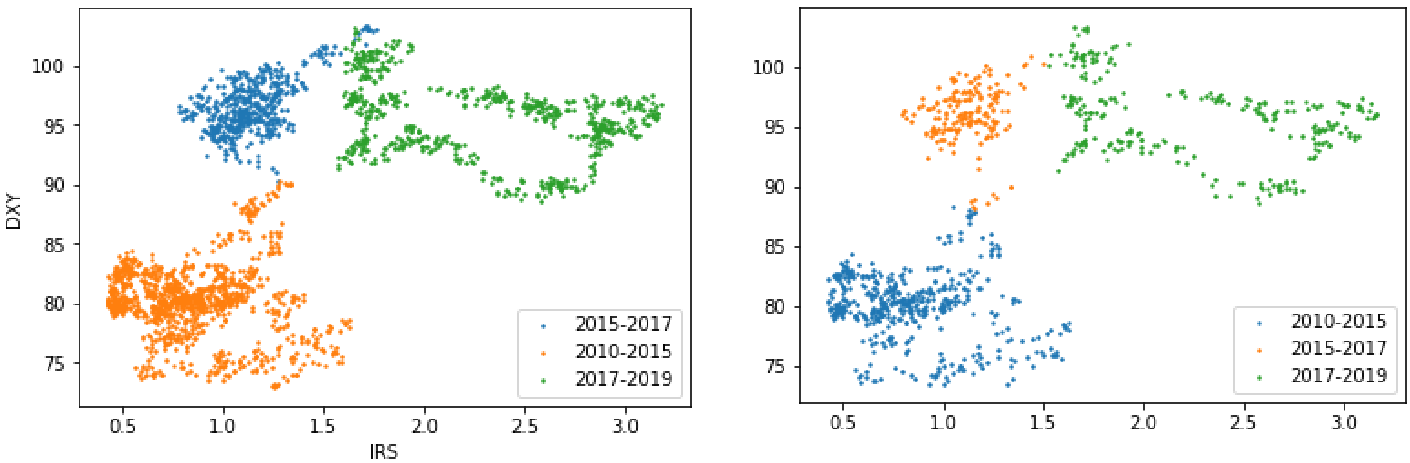
While using kNN, we set the test size as 40 percent and k-value as 200. The results we get are in the figure below.

```
[[487   9   0]
 [  0 178   8]
 [  0   0 256]]
              precision    recall  f1-score   support

   2010-2015       1.00      0.98      0.99       496
   2015-2017       0.95      0.96      0.95       186
   2017-2019       0.97      1.00      0.98       256

   micro avg       0.98      0.98      0.98       938
   macro avg       0.97      0.98      0.98       938
weighted avg       0.98      0.98      0.98       938
```

Also we can see that as the k-value increases, mean error count increases.

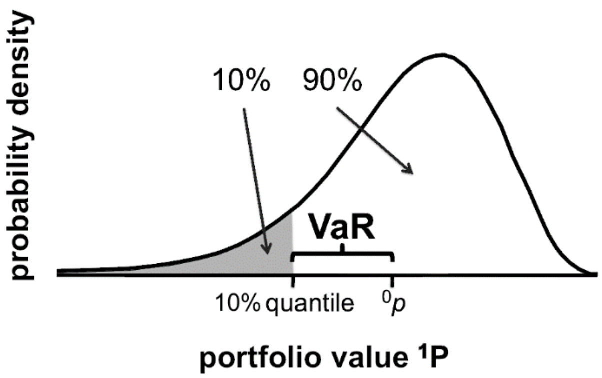


As we can expect by the success rate of the model, the graph of the predicted data and the whole DXY, IRS data are really similar.



3. Monte Carlo Simulation

Monte Carlo Simulation is a statistical tool that assumes normal distribution for the value change of an asset. After creating the regression and making k-NN test, we finally used linear regression to see whether it has provided us with a better distribution for to see Value at Risk. (VaR)



Our data has weekly changes of the IRS and DXY. Another dataset we have created by our own shows the average weekly change of the last 260 weeks (2010-2015) period, and maximum expected loss was calculated for the 261st, 262nd, 263rd… weeks and so on.

```
w_change.describe()
```

	Week	Weekly ChangeDXY	Weekly ChangeIRS
count	469.000000	469.000000	469.000000
mean	235.000000	0.000425	0.003932
std	135.532899	0.009357	0.065461
min	1.000000	-0.023852	-0.215608
25%	118.000000	-0.005527	-0.034243
50%	235.000000	0.000246	-0.001401
75%	352.000000	0.006502	0.033333
max	469.000000	0.029722	0.392801

That is to say, and basically, we have created a hypothetical question. Assume that we are in the first week of the 2015 and we have a data for DXY and IRS since 2010, last 260 weeks. Now we want to see that what is the maximum expected loss with 95% confidence interval in next week. In order to find the maximum expected loss, we first distributed the last 260 weeks' realized changes, and we have subtracted 1.6450*STDEV (1.6450 is the z-value for the one tail 95% confidence level) from mean. That is the maximum loss we can expect for the next week, and it has named as Value at Risk(VaR). VaR concept is very important mostly in financial institutions' positions on different assets.

	Mean 260 Week DXY	Standart Dev 260 Week DXY	Confidence Interval Max Loss DXY	Result DXY	Mean 260 Week IRS	Standart Dev 260 Week IRS	Confidence Interval Max Loss IRS	Result IRS
0	0.000671	0.010018	-0.015808	True	0.006037	0.077071	-0.120744	True
1	0.000737	0.010025	-0.015755	True	0.006393	0.076781	-0.119912	True
2	0.000731	0.010030	-0.015768	True	0.006414	0.076769	-0.119872	True
3	0.000657	0.009900	-0.015629	False	0.006530	0.076728	-0.119687	True
4	0.000636	0.009894	-0.015640	True	0.006645	0.076795	-0.119684	True

As it can be seen above, column 0 represents the first week of the 2010 and for every week, we have calculated the next 260 weeks' change, then we have compared the actual loss occurred in the week 260+i with maximum loss expectation derived from linear regression and actual maximum loss expectation computed by VaR modelling.

6. RESULTS

Assumption in this study is that we are hypothetically in the first week of 2015. Rather than classical VaR calculation with Monte Carlo Simulation, we have used the actual Monte Carlo

Simulation results to make next week's prediction with linear regression. In DXY, we have seen that linear regression-based Monte Carlo Simulation provide us narrower distribution in 200 weeks out of 208 and IRS provided us narrower distribution in 149 weeks out of 208. Even the linear regression is not the best ML algorithm in prediction, obviously, IRS performance can be explained by interest rates performance in the world from 2010 to today.

```
result_MCDXY = []
for i in range(0,208):
    a=w_change["Weekly ChangeDXY"][i+261] > linear_dxy_regressor.predict([[i+261]])
    result_MCDXY.append(a)
```

```
result_MCDXY.count(True)
```

```
200
```

```
result_MCDXY.count(False)
```

```
8
```

```
result_MCIRS = []
for i in range(0,208):
    a=w_change["Weekly ChangeIRS"][i+261] > linear_dxy_regressor.predict([[i+261]])
    result_MCIRS.append(a)
```

```
result_MCIRS.count(True)
```

```
149
```

```
result_MCIRS.count(False)
```

```
59
```

7. REFERENCES

- Monte Carlo Simulation Definition, Will Kenton, Jun 2019 https://www.investopedia.com/terms/m/montecarlosimulation.asp
- Monte Carlo Methods in Finance, Peter Jackel , 24th November 2001
- Efficient Monte Carlo Simulation of Security Prices, Darrel Duffie, Peter Glynn, 1995
- What is an Interest Rate Swap? Ferhad Malik,August 2018
- **https://medium.com/fintechexplained/what-is-an-interest-rate-swap-58450e1c539e**

- Interest Rate Swap Market Complexity and Its Risk Management İmplications, Steve Yang, Esen Onur, October 2018
- What Drives Interest Rate Swap Spreads?, Kodjo M. Apedjinou , November 2013