Parsa Mohammadpour

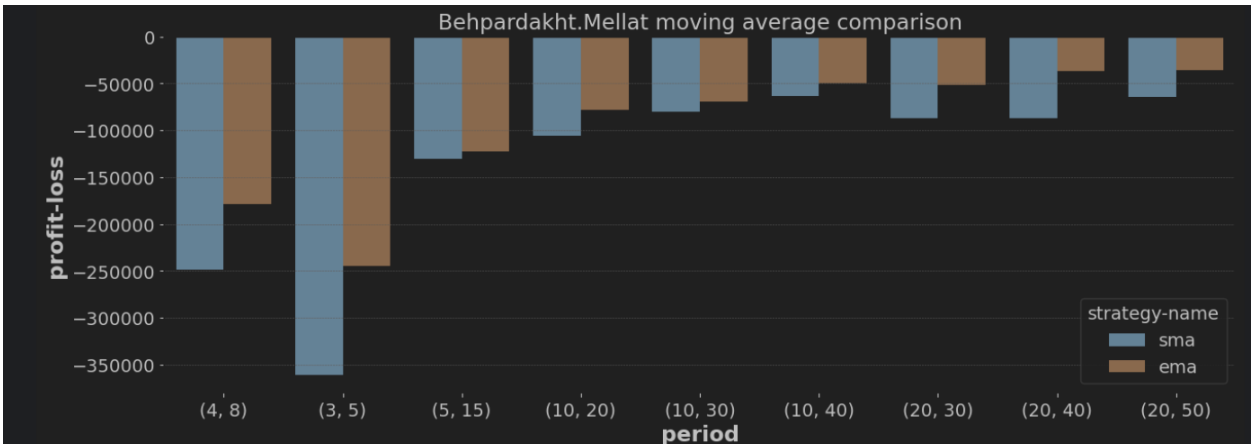
Questions:

1- Comparing different moving average periods for stocks and suggesting the best one.

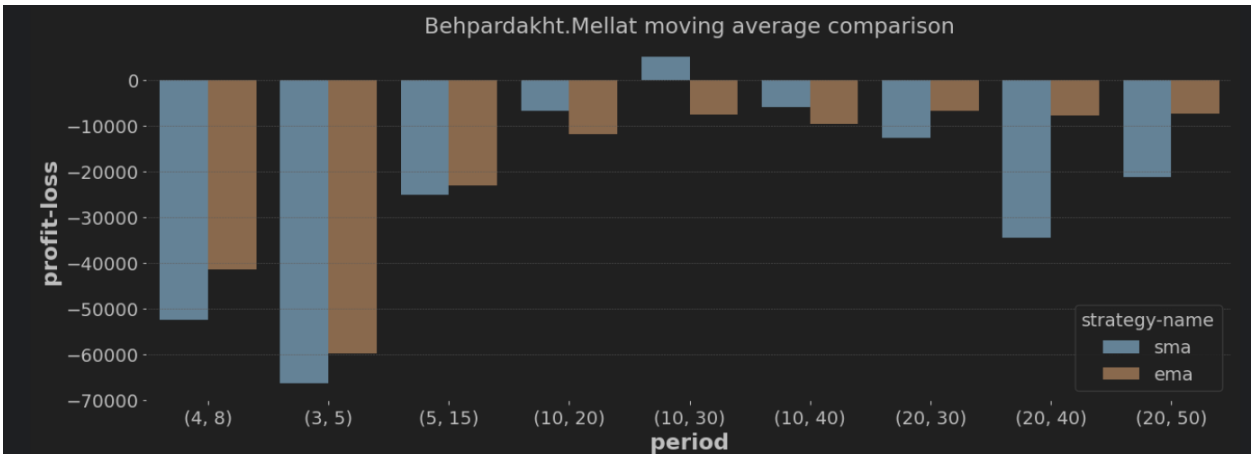
Here we went with two assumptions:

1- Considering market commission:

In this case usually the moving average algorithm didn't work very well and usually at the end, we finished we some loss. And as much as we decreased this amount, we got better results. As you can see in the following figure, when we set the market commission to 10%, almost every moving average test ended up we a huge amount of loss as you can for example see the following figure generated with this commission fee for the market for the Behpardakht.Mellat stock.



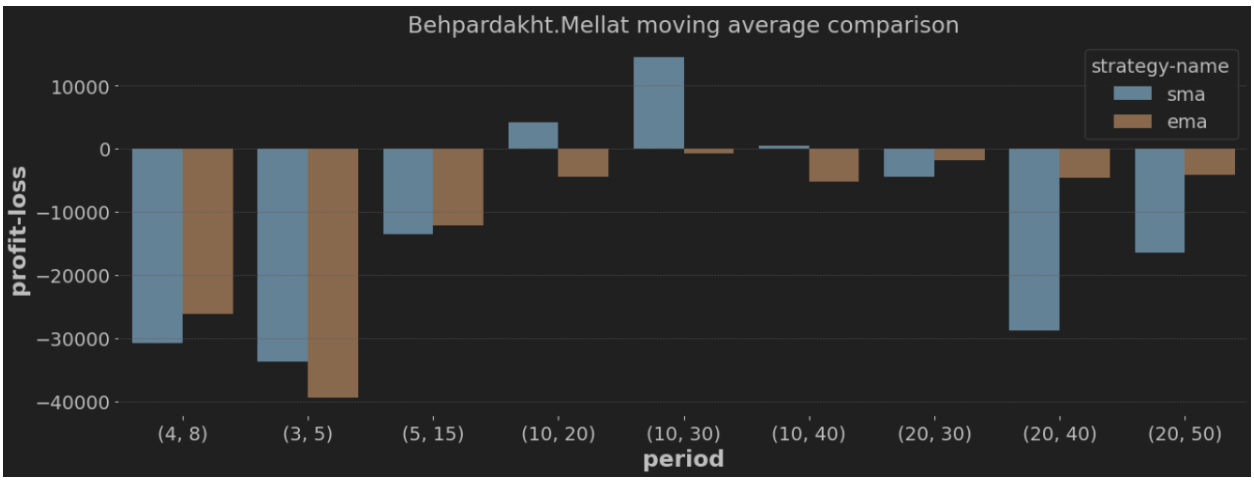
But as we started to reduce this commission fee for the market, wee saw better result. For example the following figure is the same stock with another simulation for moving average with market commission set to 1%. The following figure was the result of this simulation. And as you can see, it was improved.



As you can see in the above figure, results are improved.(we had less loss and we even had some positive results with 10, 30 period)

In the case that we set the market commission to zero percent. In this case the results got better and we even had more positive results.(we got some profit)

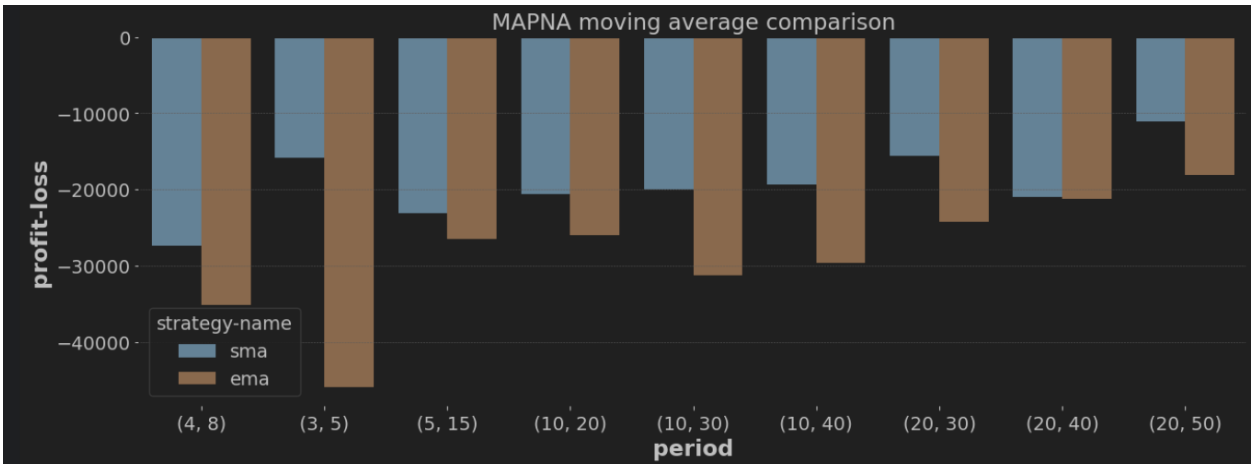
The figure below show the result of running this simulation with zero market commission percentage.



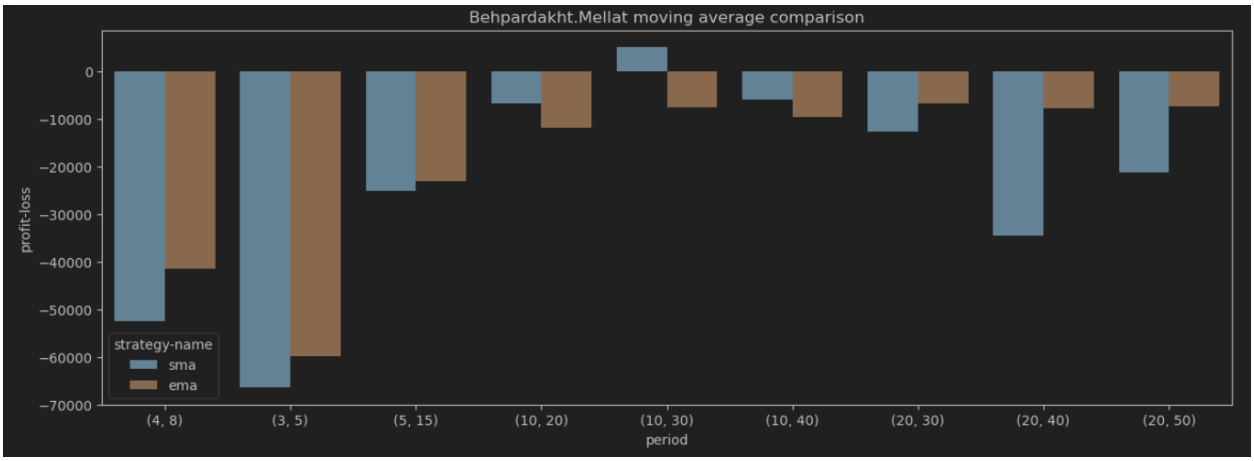
So in general, backtesting doesn't gives us much. Even for the best case with zero commission fee and also having future prices (which we don't and we argued about that in class, it is only used for few objectives such as testing a model in the worst and for worst stocks). In the best case, in some stocks like IranKhodro, Behpardakht and S_I..N..C..Ind. we had profit in best possible scenario (somehow imaginary scenario like as was mentioned above, with no market commission). And even In these cases, we had less than 1'500$ profit with starting capital of 1'000'000$ in a five year period (the simulation as was requested, during last five years). But in a more real simulation that we had some market commission fee, most of the time we didn't even had a profit from that.

Now lets jump into comparing SMA (Simple Moving Average) vs EMA (Exponential Moving Average). In general, we couldn't say that one of them works better than the other one, but as the results shows, SAM, was the one that provided some profit for us but not all of the time. Let's see some results for the previous statements.

The following figure, was the comparison of SAM and EMA with same period for MAPNA and as you can see, for all periods, SMA works better than EMA. The result is as follows:

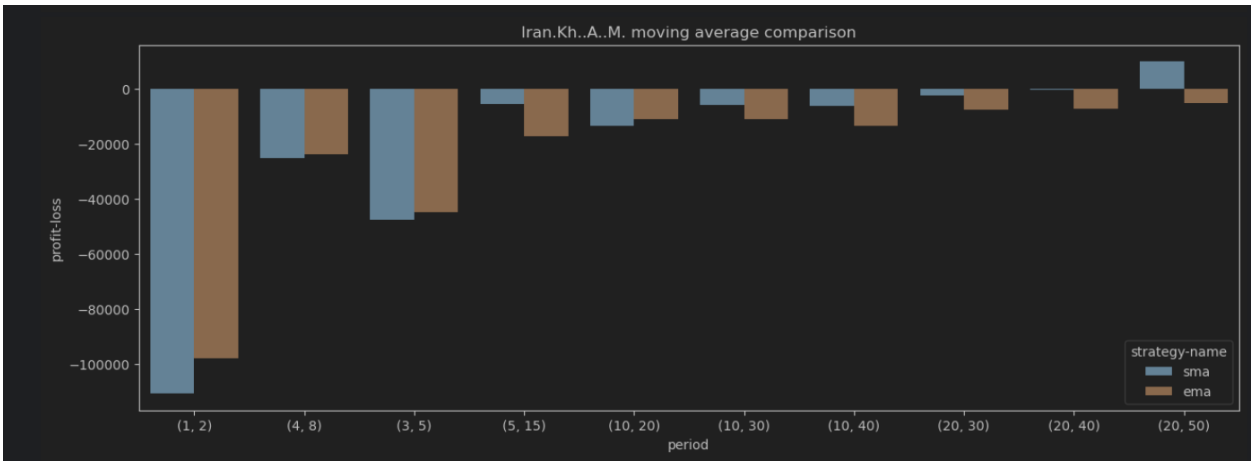


But that was for the case with zero market commission and one stock, but for some other stock with some market equal to 1 percent commission, we couldn't say that thing. In order to illustrate the point look at the following figure which the result of specified simulation.

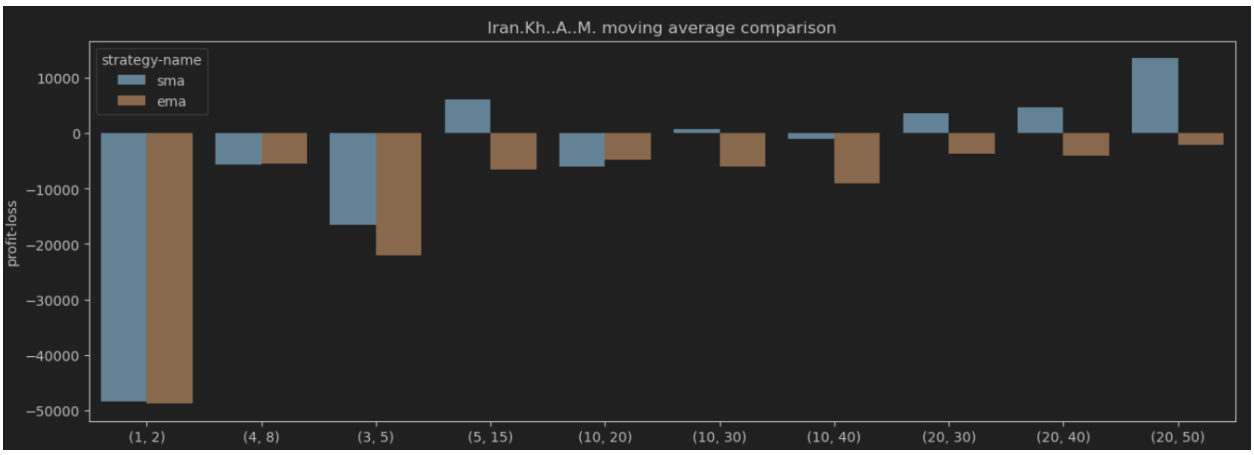


As you can see, EMA results were better in 6 periods in total of nine periods. (But the best result was for SMA with 10 and 30 period)

Here we specified both cases (one with having commission fee and one without it) because in real word, we have commission fee and also number of trades in SMA is usually more than EMA. As the following figures shows, number of trades with SMA cross strategy is more than number of trades with EMA strategy. We compared them to see if it makes a change or not. In following figures first we show the result of the simulation with the commission fee of 1 percent for IranKhodro stock:

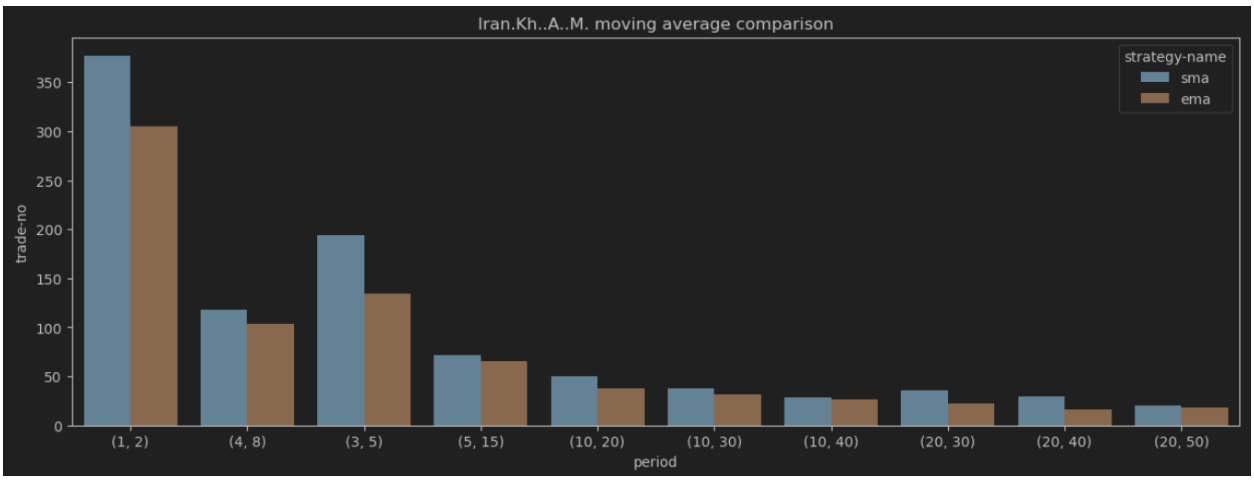


But as you can see in the following figure, which is a result of simulation for IranKhodro with the commission fee of 0.0 percent, The result of SMA for period (1, 2) was improved as the commission fee wes reduced. See the following figure to see the difference:

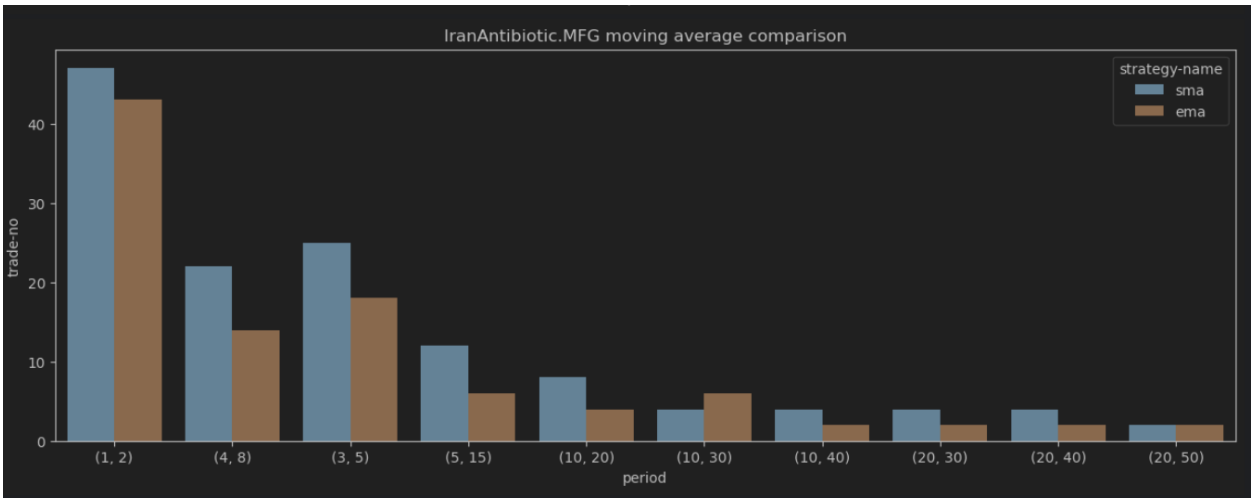


As you can see, SMA even works better than EMA for Irankhodro when we don't have any commission fee. (second case was without commission fee and the first case was with 1% commission fee)

This actually makes sense even more when we look at the comparison of number of trades of these two strategies(EMA and SMA). The following figure illustrates this point:



As the simulation results were, most of the time number of trades that were made by SMA was more than number of trades made by EMA. But that is not the case for all the time. For example the following figure shows the opposite of this:



As you can see, number of trades for period (10, 30) was more in EMA strategy.

2- At this part we were asked to apply different linear regression periods to the stock prices and compare the results with each other.

For this part we applied linear regression with different period o the input data and then for testing that we applied it for 20 percent of data as test. Then we calculated MAPE (mean absolute percentage error) and drew the plots. Then we could choose which of the specified period is best for the regression.

First lets consider the definition of MAPE error. The formula of MAPE error is as follows:

$$nAPE = \frac{\sum \frac{A - F}{A} * 100}{N}$$

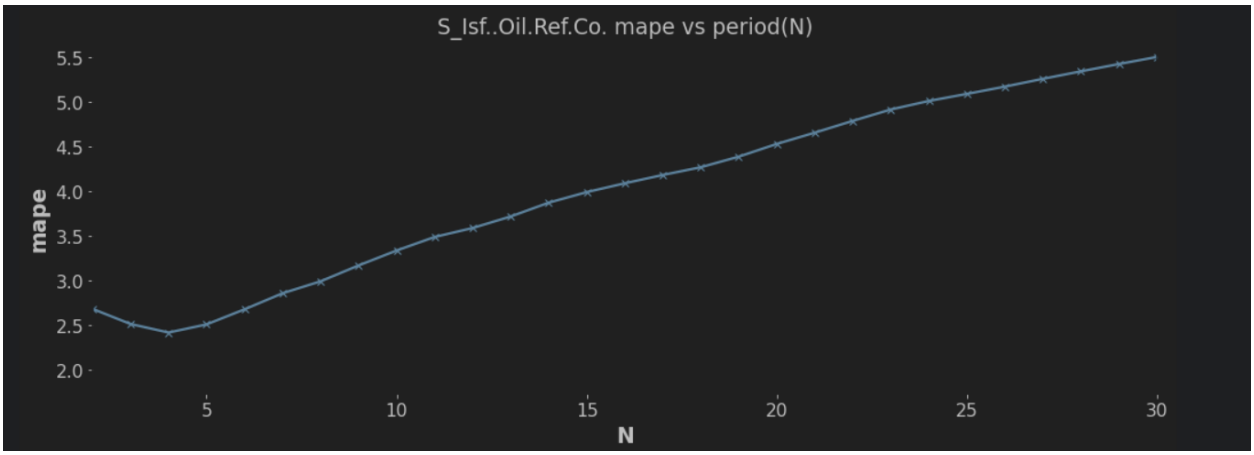
As the formula shows, it means somehow that calculate the mean percentage of difference between real values and estimated values. So our goal is to minimize this error.

So before going to the results, lest make our purpose by using period for regression. For predicting the day number t, t is our index for that day, we use N previous days for linear regression. Then as we go further, we use our own prediction for predicting t +1 day. It is easy, It's just like applying linear regression for N previous days. After that we add this new estimate of ours, to compute the prediction for next day **on the same linear regression model** but not with previous values, we train it again in our data's, but on the same LinearRegression object or in saying, on the same model. We don't change model unless we want to predict with another period.

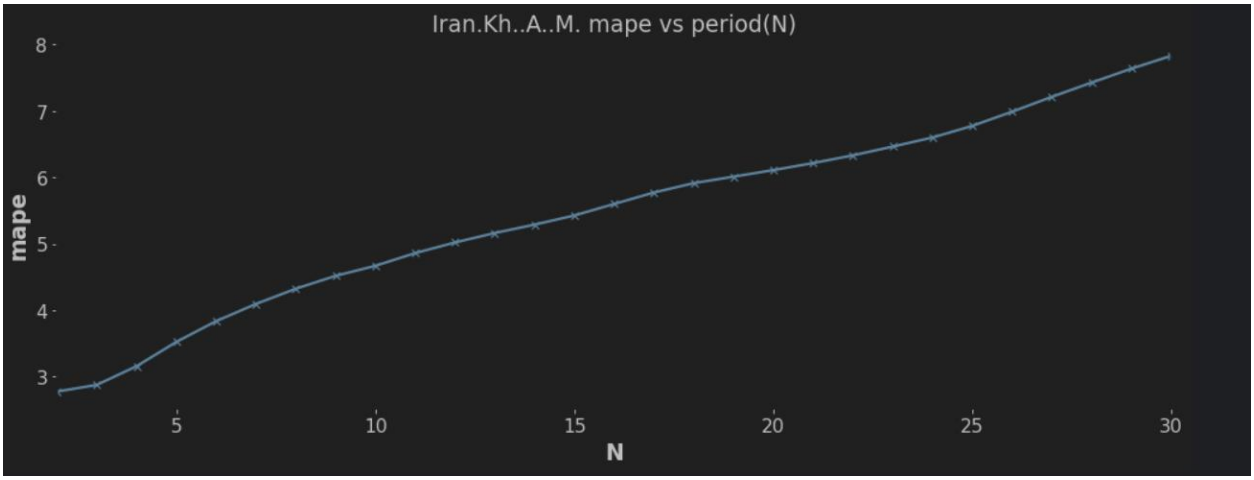
Now let's get to our simulation and results.

We did our simulation for periods between [1, 30] and we plotted the MAPE results (along side with the R2 and RMSE results, but as we were asked, we only concentrate on MAPE, but other error functions give us mostly the same result)

For example the following figure is the result of MAPE error for S_Isf..Oil.Ref.Co. for the specified period. The best period was for around 3 or 4.



But as you can see in the following figure, the best result for IranKhodro was for the 1 period, As each day value has a best relation only with its previous day.



Some notes on the notebook file:

- In the provided notebook, by installing the requirement.txt file, you can run this notebook.
- Datasets must be in a folder named “stock-samples” in root of the project.
- If you want to focus on only one stock, set SAMPLE_STOCK_NAME value to the name of that file in the “stock-samples” folder. Otherwise you will get too many plots and that might be confusing. This variable is in the “Defining some parameters” cell in notebook.
- If you want to only focus on the closing price, set “SAMPLE_COLUMN_NAME” value to “Close” (c is capital), otherwise you will get too many plots for some cells and that might be confusing. This variable is in the “Defining some parameters” cell in notebook.
- Set Initial value of the market commission in the same place. (This things can be simply done in function calls too)
- “simulator.best_stock_ma_comparison(SAMPLE_STOCK_NAME, figsize=(15, 7))” this line was commented and run at the end. Because it uses js code to run and changes some properties in pyplotlib and make notebook plots a mess, so it is better to run this code cell at the end.
- At first some visualization of data and data’s stats are provided in order to have a look and have a little sense about data’s. because we know it is very important to have a sense of our data’s and their stats.
- Although we said that this kind (plotting every price like closing, opening, high and low might be) might be a little messy and makes it hard to understand the result, but we did it because we wanted to save the results and figures. And it also help, for example if you want to find the best backtest result, you can simply just go to results folder, then go to cerebro, then in each folder there is a figure with the name of best moving average period in it.

References:

- Mainly class
- Class notes
- Some websites for coding
- Some website for reading about MAPE, R2 and RMSE

(I don’t have the exact websites right now, sorry)