Final project report: Portfolio Management

Fàtima El Baghdadi
Model Predictive Control subject
ETSEIB - UPC
Barcelona, Spain
fatima.el.baghdadi@estudiantat.upc.edu

David Pérez Ruiz
Model Predictive Control subject
ETSEIB - UPC
Barcelona, Spain
david.perez.ruiz@estudiantat.upc.edu

A. *Introduction*

With the current economic situation all companies as people are looking for alternatives to obtain an income. Therefore, there are many ways to earn income and one option is investing in assets. So, the aim of this project is how to invest in cryptocurrencies to maximize the profit.

First of all, the portfolio investment is a passive investment because the purpose is obtaining a financial profit without any implication in the internal decisions of the company. The factors that are considered to invest are: the invested amount, investor's risk tolerance and planning horizon. And related with this factors, there are different kind of investors that are more tolerant with the risk or not, short time inversion or large period,... Thus, in this report will be considered a linear model that will represent the evolution of the cryptos[1], a controller that will take the decision of where to invest taking into account as constraints different scenarios.

B. *Data*

The Dataset used has been extracted from Kaggle, a website, and has over 700 cryptocurrencies and a historical of 3 years, but we have selected only 5 cryptocurrencies: 'BTC', 'DASH', 'ETH', 'NEO' and 'ZEC'. The data of 2016 and 2017 has been used to obtain the linear model and the data of 2018 is reserved to compare it with the predictions of that same year.

The evolution of these cryptocurrencies of 2018 are represented in figure 1.

It can be seen that the cryptos have a variation during the year between $2.4 - 2.7 \cdot 10^5$ \$. At the beginning of the year the values change considerably, but then the variation is lower.

C. *Model of the evolution of the cryptocurrencies*

As it can be noted, the model does not have any input (it is an autonomous model), it depends on the previous values that capture the variation and a disturbance that is the uncertainty.

The first step for the linear model has been to take and process the data structure by fitting that data to a linear model, using Matlab. Several data historics have been used, and the ones that worked the best were the ones using the prices of the cryptocurrency at t-1, t-2 and t-4, The linear model is the following:

$$A(t) = \alpha A(t-1) + \beta A(t-2) + \tau A(t-4) + k \qquad (1)$$

[1]Crypto and cryptocurrency mean the same, the former is used for brevity.

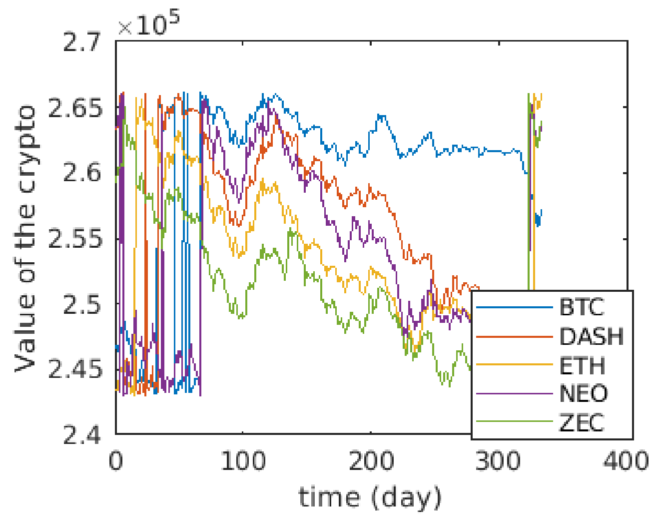

Fig. 1. Evolution of the crypto in 2018.

Where A(t) is the value of the crypto in the day 't' and is obtained by a model that fits the value of the day with the value of the previous 4 days by its constants and an independent value. This model has been used since it gives lower error than other combinations of the input that have been tested. The models obtained with respect to every crypto are: 'BTC', 'DASH', 'ETH', 'NEO', 'ZEC' are the following:

$$\begin{gathered} A_{BTC}(t) = 0.65284A(t-1)+ \\ 0.29961A(t-2) - 0.068003A(t-4) + 30102 \end{gathered} \qquad (2)$$

$$\begin{gathered} A_{DASH}(t) = 0.87731A(t-1)+ \\ 0.11302A(t-2) - 0.044311A(t-4) + 13753 \end{gathered} \qquad (3)$$

$$\begin{gathered} A_{ETH}(t) = 0.69391A(t-1) + 0.18807A(t-2) \\ +0.061978A(t-4) + 14263 \end{gathered} \qquad (4)$$

$$\begin{gathered} A_{NEO}(t) = 0.63792A(t-1) + 0.20566A(t-2) \\ +0.094792A(t-4) + 15672 \end{gathered} \qquad (5)$$

$$\begin{gathered} A_{ZEC}(t) = 0.93624A(t-1) + 0.034787A(t-2) \\ +0.0005393A(t-4) + 7154.4 \end{gathered} \qquad (6)$$

Note: the sum of the coefficients of every model sum approximately 1, it means that for the crypto 'ETH' the 69,4% of the prediction is the value of the previous day, the 18.8% is the value of before yesterday,... That model obtained will simulate the prediction of the values of 2018 that are shown in figure 2.

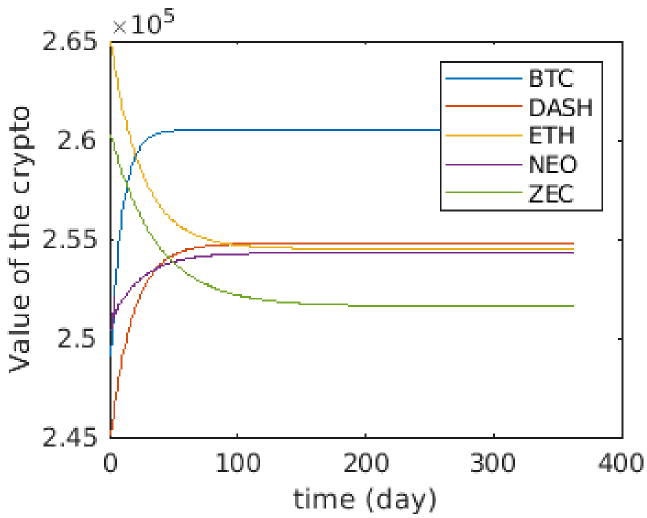

Fig. 2. Evolution of the prediction of cryptos in 2018

It can be observed in figure 2 that as it is a linear model it can't have the same shape as figure 1, but the values are between $2.45 - 2.65 \cdot 10^5$. These models are obtained using R to preprocess the data and separate the target and the values of the previous days to create a table that can be used in MATLAB to fit a linear model.

An example of the structure of the data set can be seen in figure 3. After, MATLAB is used to fit a linear model with

	type_crypto	target	date	import1	import2	import4
1:	DASH	245074	2018-01-05	244515	244946	243452
2:	DASH	244360	2018-01-06	245074	244515	243466
3:	DASH	245447	2018-01-07	244360	245074	244946
4:	DASH	245618	2018-01-08	245447	244360	244515
5:	DASH	243700	2018-01-09	245618	245447	245074

Fig. 3. Structure of the data set obtained with R.

the data for every crypto and, for example, the model obtained for the crypto 'DASH' can be seen in figure 4.

In order to improve the model, since the results of the investment depend a lot on how good is the model, a disturbance is inserted to simulate variations and noise to obtain a profile more similar to the real one. The linear models presented are now given a confidence interval of +/-5% (that can be observed in figure 5), that is, the value of the predictions may take any random value between 1,05 and 0,95 times the one the nominal prediction would yield. With this disturbance added, doing the same cryptocurrency prediction yields the evolution at the figure 6.

It can be seen that the shape of the evolution between the real one and the predicted one with the disturbance isn't similar but at least it has some variation.

D. Objective and constraints of the model of the controller

As it has been explained, the objective of the project is investing in the assets that are more profitable and/or less risky.

```
mdl_DASH =


Linear regression model:
    y ~ 1 + x1 + x2 + x3

Estimated Coefficients:
                   Estimate        SE         tStat       pValue  
                   _________    ________    _______    __________

    (Intercept)        13753        4893     2.8107     0.0052432
    x1               0.87731    0.055269     15.873    2.0787e-42
    x2               0.11302    0.066465     1.7004      0.090011
    x3             -0.044311    0.041822    -1.0595       0.29016


Number of observations: 329, Error degrees of freedom: 325
Root Mean Squared Error: 1.85e+03
R-squared: 0.89,  Adjusted R-Squared: 0.889
F-statistic vs. constant model: 875, p-value = 2.76e-155
```

Fig. 4. Model obtained with MATLAB for the crypto 'DASH'.

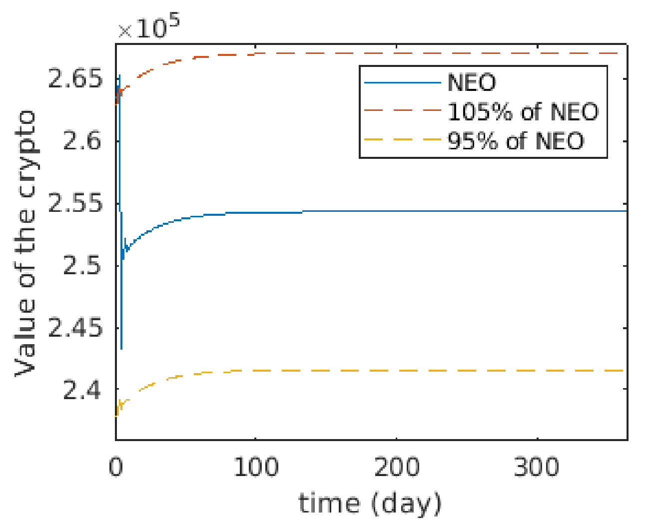

Fig. 5. The value of 'NEO' and the intervals of 105% and 95% that will represent the disturbance of the crypto.

The linear function that is used to control the benefits is the following:

$$benefit = (n \cdot A(hp) \cdot z(hp)) - I_0; \qquad (7)$$

where n is the number of cryptocurrencies that have been purchased with the initial investment for each crypto and I_0 is the value of the initial investment. So, with the cryptos that have been initially bought, it can be studied the evolution of the value of every one. With these numbers of cryptos by their predicted price in the future, the value of the investment along the horizon can be obtained. But, to choose in which crypto to invest, the controller $z(hp)$ has been used, which chooses which is the most profitable and/or less risky in the last day of the prediction horizon. The objective function depends on the last day of the prediction horizon because we are interested in the value of the investment the day that we sell the cryptos. It will be invested in every iteration only the initial inversion and the benefits will be for the investor.

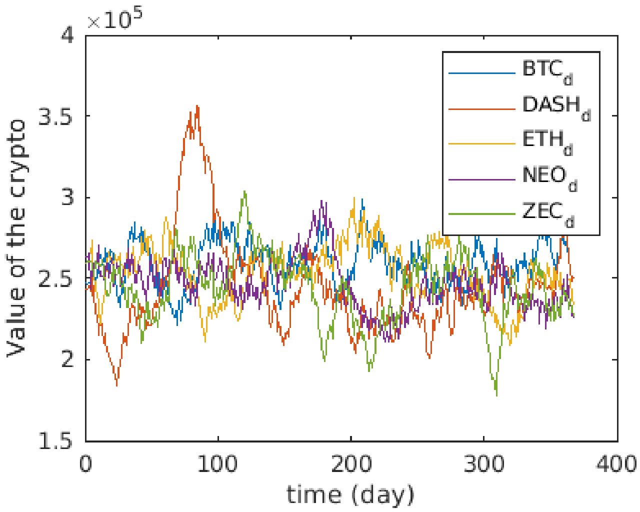

Fig. 6. Models with disturbances between 105% and 95% of the value of the crypto.

So, the objective is maximizing this benefit but it will take into account the risk in a simulation to see the effect. The formula that is used to quantify the risk uses the variance of the asset by its weights that are $z(hp)$, the equation [1] is the following:

$$\sigma_p^2 = \begin{bmatrix} z_1 & z_2 & z_3 & z_4 & z_5 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{15} \\ .... & \dots & .... & \dots \\ \sigma_{51} & \sigma_{52} & \dots & \sigma_5^2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \end{bmatrix} \tag{8}$$

Note: the covariance matrix has been scaled because the values are very big and affects negatively the objective function.

E. Simulation I

The first simulation explains how the process works. So, the goal is maximize the return:

$$-benefit = -(n \cdot A(hp) \cdot z(hp) - I_0); \tag{9}$$

s.t.

$$\begin{gathered} A(i+1) = \alpha A(i) + \beta A(i-1) + \tau A(i-3) + k + \bar{w}(i) \\ \textstyle\sum_{i=1}^{i=hp} z(i) \leq 1 \\ 0.3 \geq z(i) \geq 0 \end{gathered} \tag{10}$$

where $A(i) \in \mathcal{R}^{5Xhp}$, $n \in \mathcal{R}^{1X5}$ and $z(i) \in \mathcal{R}^{hpX5}$ is the controller that can take only one value in the prediction horizon, hp = 7 days, such that maximizing the benefit and the sum of the proportion of inversions must equal or lower to 1 and diversifying the investments. The hp means the horizon of the predicted evolution of every crypto and also the period in which the money can't be extracted from the initial investment[2].

Using a model with disturbances, the previous requirements and an initial investment of 5000 \$ the results are shown in figure 7.

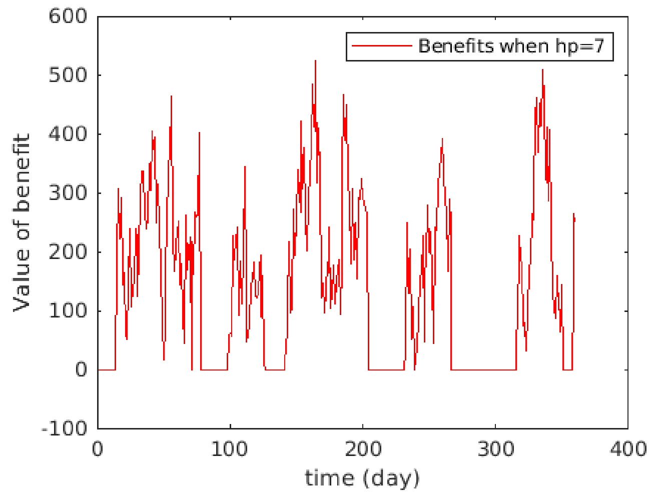

Fig. 7. The benefit obtained along the year using a hp=7.

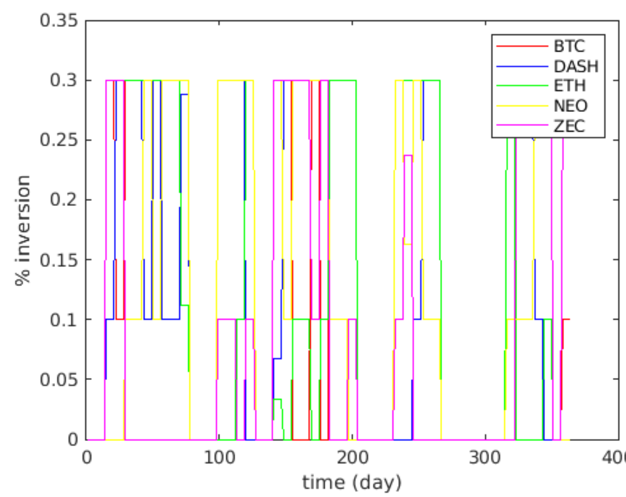

Fig. 8. Controller obtained that maximizes the benefit.

Now it is going to explain the procedure that is used choosing 7 days of the predicted values in figure 9 where the difference between the 1^{st} day and the 7^{th} day are $-0.0514 \cdot 10^5$\$ for BTC, $0.183 \cdot 10^5$\$ for DASH, $0.0665 \cdot 10^5$\$ for ETH, $0.0214 \cdot 10^5$\$ for NEO and $0.2335 \cdot 10^5$\$ for ZEC. The benefit is calculated by the following steps:

[2]It has been discussed with the professor that a simulation using a prediction horizon to predict how the cryptocurrencies change and a controller that changes every day with respect to this prediction would not be the optimal way to proceed. Instead, we proceed to generate simulations where the money can't be extracted after a certain period, which is the hp.

1.0e+05 *

'BTC'	'DASH'	'ETH'	'NEO'	'ZEC'	
2.4684	2.4345	2.6337	2.6342	2.5923	1st
2.4637	2.4347	2.6360	2.6369	2.6034	
2.4740	2.4495	2.6500	2.6528	2.6044	
2.4766	2.4451	2.6587	2.4334	2.6046	
2.5696	2.5511	2.5503	2.6078	2.6365	
2.4503	2.4878	2.5878	2.6883	2.7543	
2.4167	2.6175	2.7002	2.6556	2.8258	7th

Fig. 9. The value predicted with the linear model with disturbance.

- The initial value, for example, for the crypto 'DASH' is $2.6175 \cdot 10^5$. So, if it is invested 5000$ it can be bought $5000/2.4345 \cdot 10^5 \simeq 0.0205$ crypto of 'DASH'.
- With this quantity of cryptos is analysed the evolution of this cryptos along the horizon (hp=7).
- The last day of the horizon it can be seen the number of crypto by the value predicted of this day is the final value of the crypto that is $0.0205 \cdot 2.6175 \cdot 10^5 = 5365.785$\$ for 'DASH' and as the inversion is 5000, the benefit is 365.78$ that is the value obtained in figure 7. Note, that for this simulation, as the maximum proportion of the investment in a crypto is 0.3, it is selected different inversions in the same horizon.

In figures 7 and 8 it can be observed that when the controller can't find any investment to maximize the benefit and there are losses, then the investment is 0 and therefore, the benefit also is 0. The second step is comparing the effect of changing the prediction horizon to 15 and 30 days and see what happens with the benefit and the value the global benefit. Using the same problem for different prediction horizons, figure 10 has been obtained. It can be seen that there are differences because the time of reaction to change the proportion invested in every asset is different. The global benefit along the year using a hp=7 is 7113.9$, using a hp=15 is 4007.3$ and with a hp=30 is 2128.6$. As it was expected, a smaller hp horizon is better because the controller can react earlier to a drop or rise in the value of the cryptocurrency.

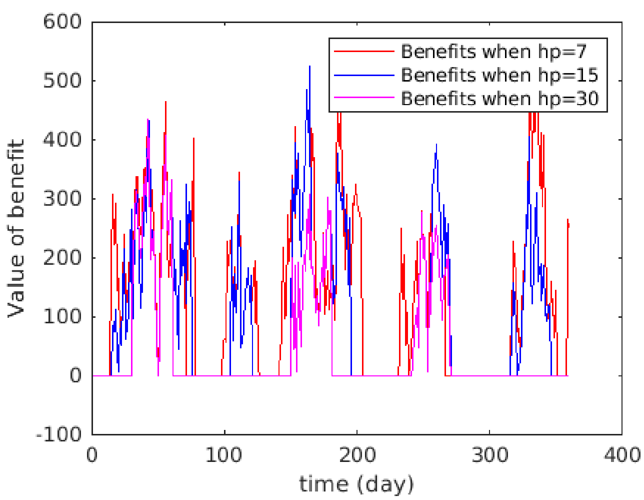

Fig. 10. Comparison between the different benefits respect the prediction horizon.

Now, it is presented another simulation that includes the risk by introducing the covariance matrix of the evolution of the cryptocurrencies at the selected horizon. This time, benefit has to be equal or larger than 0, and no diversification constraint is added (so, all money can go to one same type of coin). The objective function in this case is:

$$-(n \cdot A(hp) \cdot z(hp) - I_0) + z(hp)' S(i : i + hp - 1) z(hp) \quad (11)$$

s.t.

$$\begin{array}{c} A(i+1) = \alpha A(i) + \beta A(i-1) + \tau A(i-3) + k + \bar{w}(i) \\ \sum_{i=1}^{i=hp} z(i) \leq 1 \\ 1 \geq z(i) \geq 0 \\ benefit \geq 0 \end{array} \quad (12)$$

where $S(i)$ is the covariance matrix of the assets along the prediction horizon. The obtained results are:

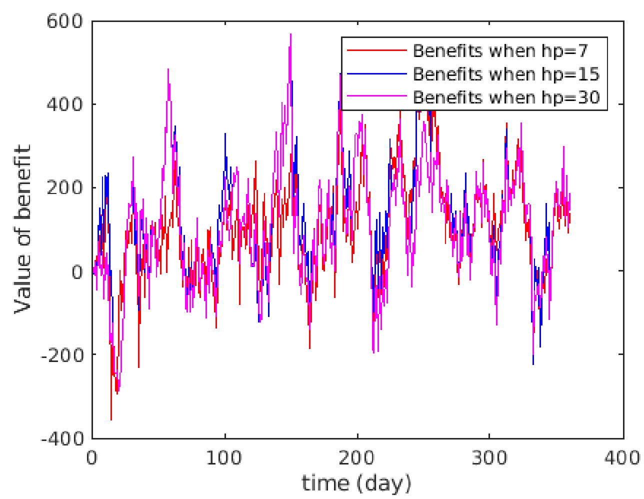

Fig. 11. Benefit obtained minimizing the risk and maximizing the benefit

The benefit obtained is 1557.3$ for hp=7, 1156.9$ for hp=15 and 894.4$ for hp=30. As it can be seen in figure 11 the global benefit is positive using any of the hp, but smaller than earlier. So, it can be concluded that taking into account the risk will yield the system very conservative and very risk avoidance. It has to be noted that the values on the covariance matrix had to be scaled down to avoid an over conservative behaviour.

Finally, we have given again to the system the ability to put all money into one same coin by not constraining with a maximum percentage for a coin. The results can be seen in figure 12 where the benefit is higher compared with the previous simulations: when hp= 7 is 14415.6$, when hp= 15 is 106611.1$ and when hp= 30 is 4342.2$.

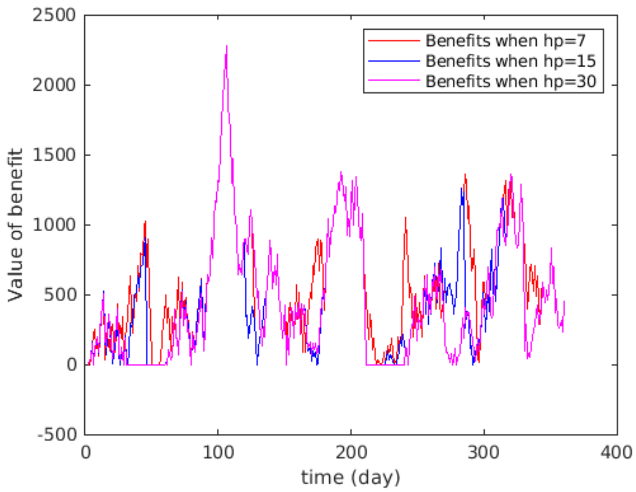

Fig. 12. Benefit obtained investing the 100% in only one crypto

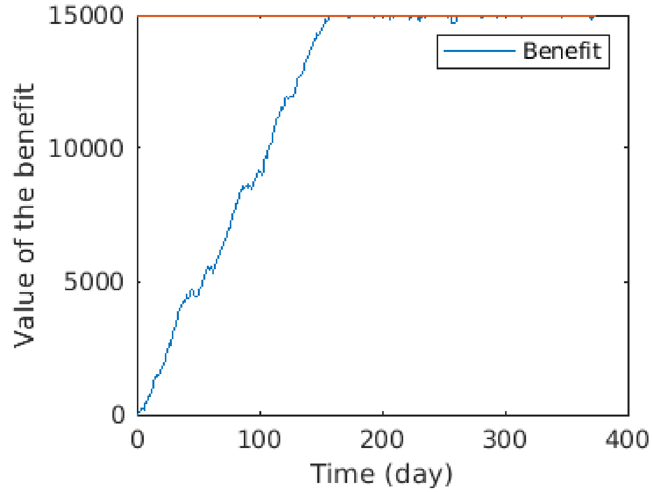

Fig. 13. Benefit obtained tracking the desired benefit 15000$ using a hp = 7

F. Simulation 5

Another way of formulating an MPC problem is thinking how to invest to obtain a certain benefit r. So, our state is the benefit and the controller returns $z(i)$. The formulation of the problem is the following:

$$J = \sum_{i=0}^{i=hp-1} (r - benefit(i+1))^T Q (r - benefit(i+1)) \quad (13)$$

s.t.

$$\begin{gathered} A(i+1) = \alpha A(i) + \beta A(i-1) + \tau A(i-3) + k + \bar{w}(i) \\ benefit(i+1) = benefit(i) + n \cdot A(i) \cdot z(i) - I_0 \\ \sum_{i=1}^{i=hp} z(i) \leq 1 \\ 1 \geq z(i) \geq 0 \end{gathered} \quad (14)$$

The value of the matrix Q is 1, the value of the reference $r = 15000$, and the initial investment I_0 is 5000. A new state has been added: the $benefit$ that depends of the previous benefit, and the number of cryptos bought by their price in the day i by the best proportion $z(i)$ to track the desired benefit.

The results obtained can be seen in figures 13 and 14. It can be seen that it reaches the desired value in the day 158 using a hp=7. Note that when the system achieves the desired value, there are small variations from the desired benefit because there is a disturbance in the model that makes variations in the value of the cryptocurrencies.

Now we are going to increase the number of hp to 15 to observe the difference using the previous formulation. As it can be seen in figure 15, the system reaches, after the day 83, the desired benefit because it can "see" more in the future the value of the coins and act accordingly with the controller. Finally, it has taken the previous simulation and tested it against the real data of the dataset for year 2018. In other words, we are no longer predicting the evolution of the price but rather we choose a cryptocurrency portfolio and then see how well the investment goes against the real data. As it can be observer in figure 16 the benefit is positive, but smaller than earlier, which makes sense.

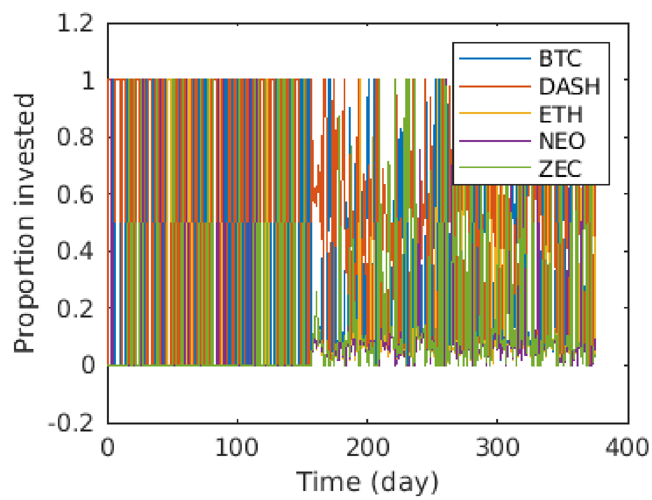

Fig. 14. Combination of controllers to track the desired benefit 15000$ using a hp = 7

G. Multi-scenario MPC

A useful tool to compute how will change a system during a certain horizon is using the multi-scenario model [2], but it is important to know how is the uncertainty of the model. As it can be observed in figures 18 and 17 the uncertainty of the same crypto in two different years changes a lot, so there is a different behavior of the same currency over the years. It can only be assured that it varies between $2.4 - 2.7 \cdot 10^5$ in all cryptocurrencies.

So, the formulation of this problem is:

$$-(n \cdot A(hp) \cdot z(hp)) \quad (15)$$

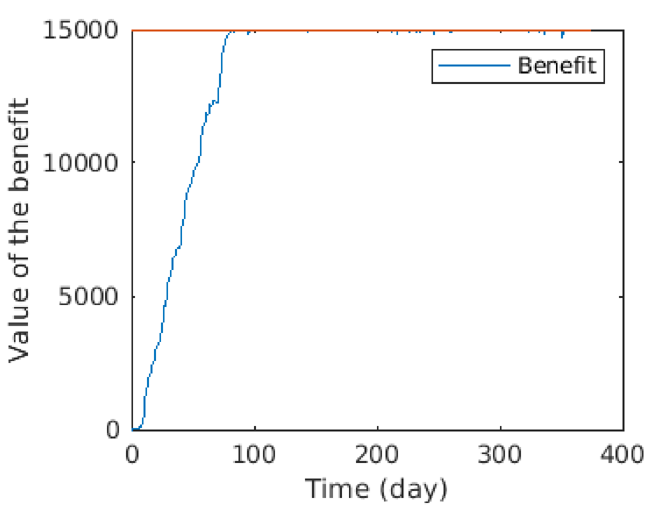

Fig. 15. Benefit obtained tracking the desired benefit 15000$ using a hp = 15

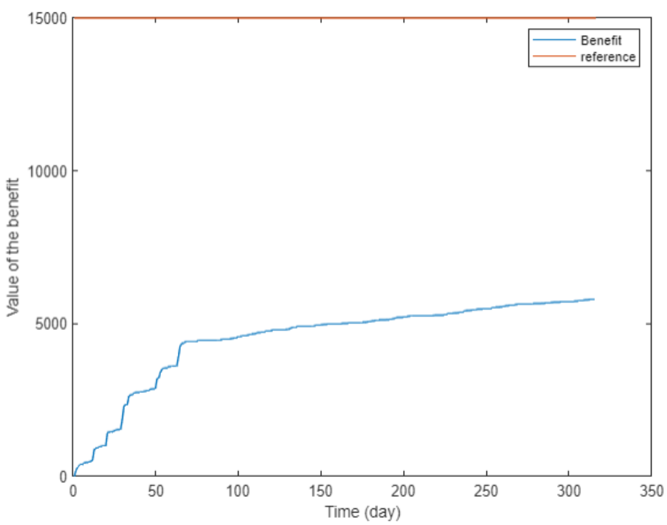

Fig. 16. Benefit obtained tracking the desired using the real data

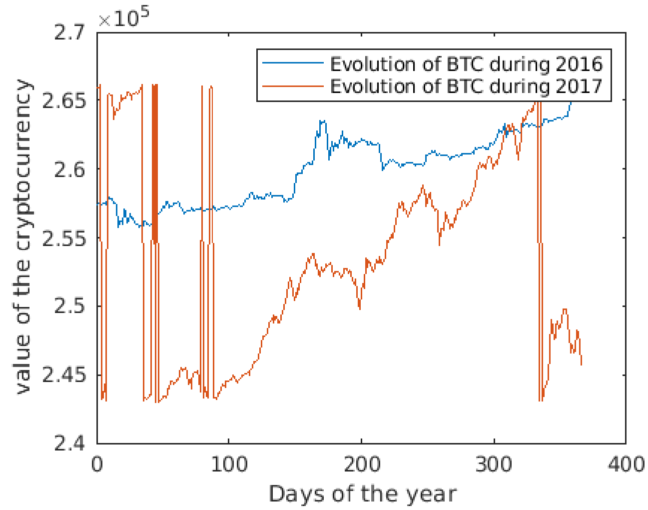

Fig. 17. Evolution of the cryptocurrency 'DASH' for 2 years

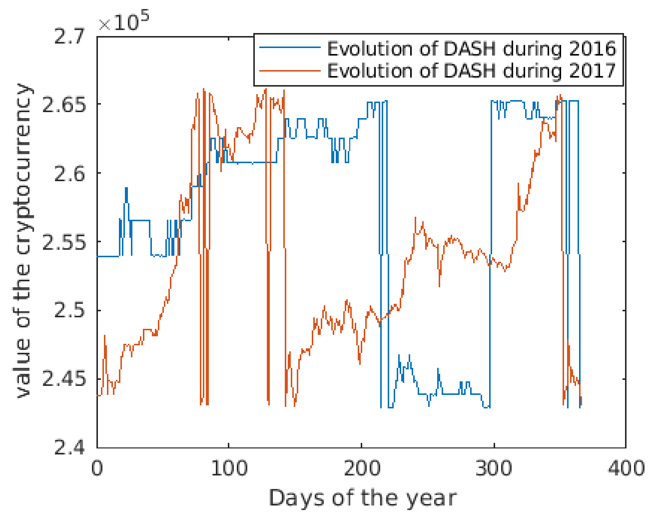

Fig. 18. Evolution of the cryptocurrency 'BTC' for 2 years

s.t.

$$\begin{gathered} A(i+1) = \alpha A(i) + \beta A(i-1) + \tau A(i-3) + k + \bar{w}(i) \\ \sum_{i=1}^{i=hp} z(i) \leq 1 \\ 1 \geq z(i) \geq 0 \\ benefit \geq I_0 \end{gathered} \tag{16}$$

The solver linprog has used and to make sure that there is a benefit, it has imposed a constraint where the benefit has to be equal or bigger than 5000 $. The different scenarios are independent of each other. It has selected the scenario of each crypto that are more similar to the real one using the relative error between the real value and the predicted one.

In figure 19 it can be seen an example of the prediction of 5 scenarios for the different cryptos, therefore the computation is more expensive.

As it can be seen in figure 20, the benefit behaves like "steps", because the solver linprog return a certain value of the benefit for each predicted horizon.

The more scenarios it generates the higher the probability to guess correctly the future and it can be observed in the following table:

Cryptocurrency	5	10	100
BTC	4.79%	4.79%	3.91%
DASH	6.14%	5.61%	4.8%
ETH	4.9%	4.73%	4.11%
NEO	5.82%	4.98%	3.6%
ZEC	7.37%	7.35%	6.34%

Finally, it has simulated a scenario with 5 cases in figure 21 to show the maximum and minimum benefit with the different scenarios that could happen and know the limits of our investment.

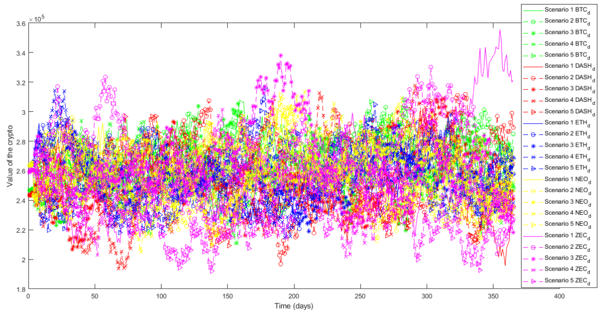

Fig. 19. Prediction using 5 different scenarios for every crypto.

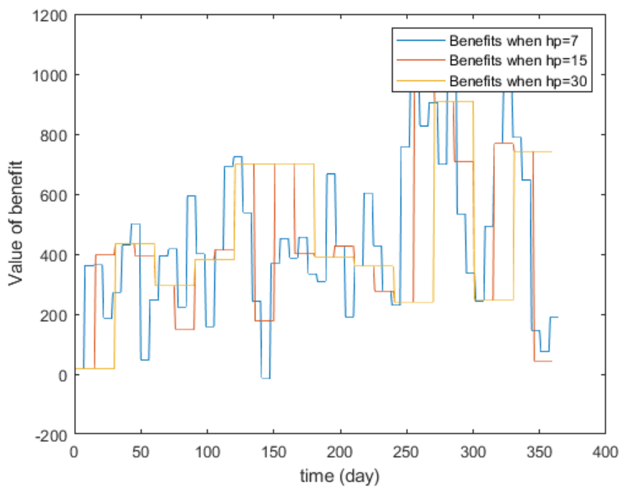

Fig. 20. Benefit using MS controller with 5 scenarios for each cryptocurrency.

H. Tree-based MPC

Another tool that is efficient is Tree-base MPC [2] because it generate trees where in every instant of the prediction horizon for every prediction it generates 2 different predicted values. In this model, it is important to have a better knowledge of the disturbance. In this case, it has been considered that in every step the predicted values are 5% upper or lower the value predicted.

It has been used the same objective function and constraints like the MS MPC case but instead of generating independents scenarios, it has build a tree considering in every iteration that means 2^i calculations. As the sizes grows exponentially and it was impossible the calculation the tree in the 365 days of the year. So, it has considered a hp of 7 days, then it has selected the iteration that are similar to the real as a good one and it started again the computation of the tree with the previous values which has been saved in the historical.

In figure 22 it can be seen the structure of the prediction and the maximum number of iterations that can computes MATLAB is 29 and then it stops. It is a trap because we are comparing the prediction with something that we haven't (the real scenario). But choosing the one that gives the most benefit is not a good idea because it would be chosen the

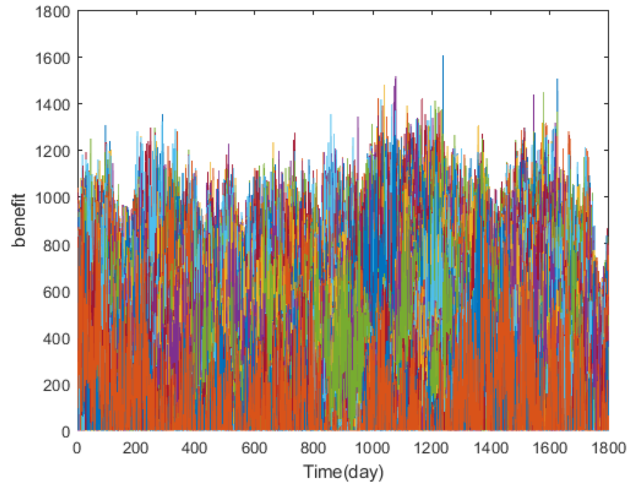

Fig. 21. Benefit using MS controller with 5 scenarios for each cryptocurrency.

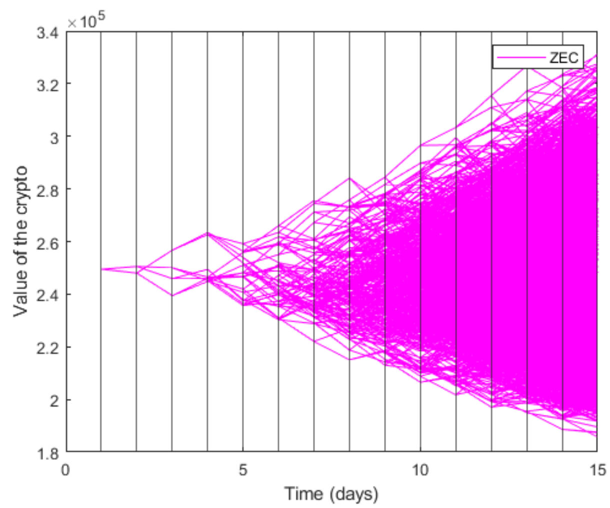

Fig. 22. Example of a Tree Based MPC using 15 iterations.

one that tends to grow and it hasn't sense in this case. The results that has obtained are shown in figures 23, 24 and 25 where it can be seen that it has invested 100% of the initial investment in only one crypto, the scenarios generated are between $2.35 - 2.8 \cdot 10^5$\$ (closer to the real scenario) and the median of the relative error calculated in the different prediction horizons are: 4.5% for 'BTC', 7.35% for 'DASH', 6.46% for 'ETH', 3.02% for 'NEO' and 2.14% for 'ZEC'. It can't be calculated the tree with 3 branches instead of 2 branches because the computer doesn't have the capacity.

I. Conclusions

In this project it has be seen that to generate a nonlinear behaviour using a linear model it is needed a bounded disturbance and the benefit depends completely how is the performance of these predictions.

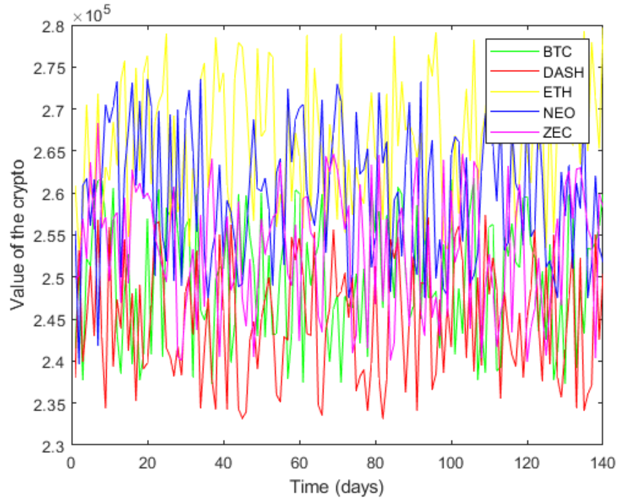

Fig. 23. Scenarios obtained generate lower error obtained using TB MPC during 140 days

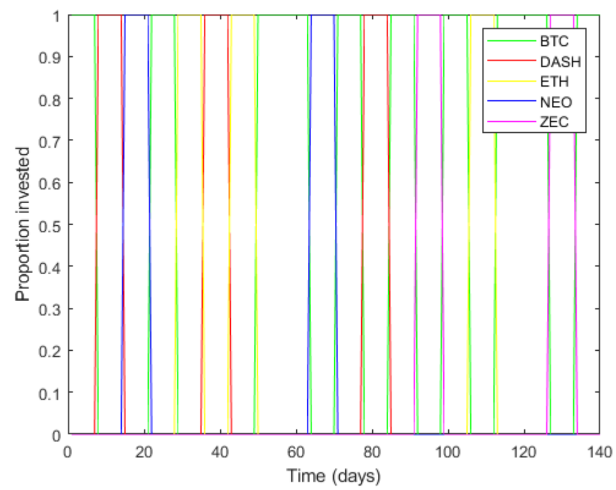

Fig. 25. Controllers obtained using TB MPC during 140 days

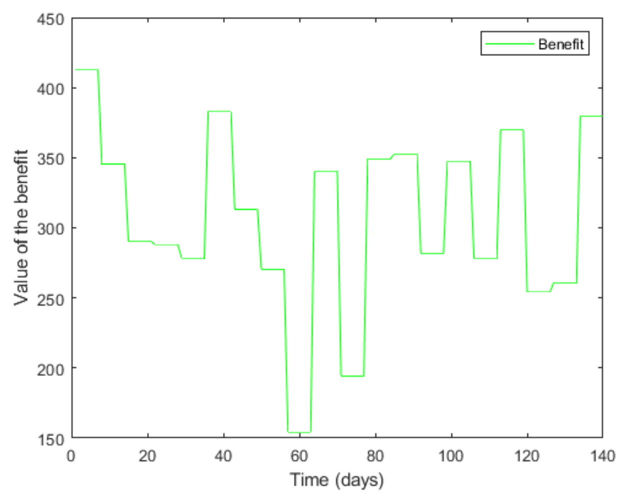

Fig. 24. Benefit obtained using TB MPC during 140 days.

The effect of changing the constraints also affects the benefit and the time that is imposed to extract the money from an investment. Also, it has seen that an investment less risky gives lower benefit.

Two types of scenarios have been generated: the scenarios where the hp is the horizon to ”see” how the value of the cryptocurrencies changes along the horizon and also is the period which the money can’t be extracted from the initial investment and it has be seen that a lower hp generates greater profit. The other scenario is to achieve the desired profit and the hp means only horizon of the predicted evolution of every crypto. In this case, a bigger hp generates greater profit.

With respect to multi-scenario MPC, the more scenarios it generates the higher the probability to guess correctly the future. We try to apply Tree-based MPC, but as it very expensive computationally and we couldn’t generate several TB MPC with different branches to see how the relative error changes.

Finally, as an improvement in the future it will be considered nonlinear dynamics model [3] that represents the behaviour of assets and apply Nonlinear MPC to treat improve the performance of the models.

Another interesting idea is to include different types of investments such as buying houses, shares in a company, jewelry, etc. The objective would be, if we are in a selected sector which would be the best option, if we are in the sector of shares of companies which would be the combination of investments that would provide a better profit, etc. Then it would be useful to apply MPC for hybrid systems because choosing a sector it will optimize the best investment.

References

[1] F. train. How to calculate portfolio risk and return. [Online]. Available: https://financetrain.com/how-to-calculate-portfolio-risk-and-return/

[2] P. Velarde, L. Valverde, J. Maestre, C. Ocampo-Martinez, and C. Bordons, “On the comparison of stochastic model predictive control strategies applied to a hydrogen-based microgrid,” *Journal of Power Sources*, vol. 343, pp. 161 – 173, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378775317300150

[3] T. Lux, “Chapter 3 - stochastic behavioral asset-pricing models and the stylized facts,” in *Handbook of Financial Markets: Dynamics and Evolution*, ser. Handbooks in Finance, T. Hens and K. R. Schenk-Hoppé, Eds. San Diego: North-Holland, 2009, pp. 161 – 215. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780123742582500075