Dipartimento di Economics and Finance

Cattedra Asset Pricing

Quantitative trading on cryptocurrencies:
A long/short strategy based on cointegration

RELATORE

Prof. Nicola Borri

CANDIDATO

Matr.

703531

CORRELATORE

Prof. Paolo Porchia

ANNO ACCADEMICO 2020/2021

INTRODUCTION

Nowadays the cryptocurrency are starting to gain a place among all other asset classes, they can be seen as an instrument which can be used in order to increase portfolio diversification since they are relatively new and offer fundamental characteristics which are different from any other asset class.

The typical strategy suggested when we try to approach this new market Is the classical "buy and hold" strategy, which merely consists into keep accumulating pieces of currencies exploiting their increasing value due to the initial expansion of the market.

Due to the relatively newness of the instruments there are few papers which try to treat this asset in a way which is more speculative than the classic buy-and-hold one, *Borri & Shakhnov* in their paper *"The Cross-Section of Cryptocurrency Returns"* try to present a more speculative approach presenting a decomposition of the asset movement into factor, which can be used to spot mispricing and to speculate on them.

This paper wants to propose a long/short strategy based on the assumption that a cointegrating relationship exists among a pair of cryptocurrency, in the first chapter I'll present the methodological background which is used in order to estimate the model, in the second one is presented the strategy and it's measured it's performance on the in-sample period and in the third chapter the performance on the out-of-sample period is presented, with a deep dive on the impact of COVID-19 and an attempt to get rid of this event.

1. Methodological Background

The central piece of the work is the cointegrating relation that is (or could be) present among a set of couples of cryptocurrencies. The idea would be to pairwise test a bunch of cryptocurrencies in order to spot a pair which is nonstationary and cointegrated.

1.1 Introduction to cointegration

The cointegrating relation refer to a particular characteristic of 2 (or more) stochastic processes: n non stationary random processes are said cointegrated with cointegrating vector β if there is at least one linear combination for which the quantity $\beta' y_t$ is stationary. It could be thought as if under the surface of the considered nonstationary time series the random walk process underlying all the series is the same.

Totally uncorrelated time series may appear to be related by using conventional testing procedure, suppose that we have the two following stochastic processes:

$$y_t = \rho y_{t-1} + \epsilon_t^1 \ ; \ \epsilon_t^1 \sim N(0, \sigma^2_{\epsilon_t^1})$$

And

$$x_t = \rho x_{t-1} + \epsilon_t^2 \ ; \ \epsilon_t^2 \sim N(0, \sigma^2_{\epsilon_t^2})$$

With the additional assumption that ϵ_t^1 and ϵ_t^2 are independent so:

1) $E(\epsilon_t^1 \epsilon_t^2) = 0$
2) $E(\epsilon_t^1 \epsilon_s^2) = 0 \ \forall \ (t, s)$

When $\rho = 1$, x_t and y_t are random walks:

$$y_t = y_{t-1} + \epsilon_t^1$$

$$x_t = x_{t-1} + \epsilon_t^2$$

Now, since they are totally independent one from the other it is reasonable to think that running the following regression will not give us any kind of statistically significant result:

$$y_t = \alpha + \beta x_t + \epsilon_t^3$$

But, despite the lack of causal relationship, running a t-test on the beta will give us a rejection of the null hypothesis; now recall that the hypothesis are formed as

$$\begin{cases} H_o : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$$

And rejecting the null means that the movements of the dependent variable y_t are somehow explained from the movements of the independent variable x_t, but we know from the assumptions that the two processes are uncorrelated.

To go to the root of this fake relation we have to give a look to the construction of t-test statistic and his asymptotical distribution, the t-stat is:

$$t = \frac{\hat{\beta}}{SE(\hat{\beta})} \sim \mathrm{T}_{n-2}$$

But this theory doesn't hold when we have $\beta = 1$, in this case we can't rely on the classic asymptotic distribution of the test and so we can't rely on the results of t-test.

In such cases we say that we have a spurious regression, another interesting element of spurious regression, which will be useful when we'll talk about Engle and Granger procedure, is a property of the residual, running the usual OLS setup we will have:

$$y_t \sim I(1) \quad x_t \sim I(1)$$

$$y_t = \alpha + \beta x_t + \epsilon_t^3 \quad \epsilon_t^3 \sim I(1)$$

Usually we would expect that the residuals of the OLS setup will be distributed as I (0) , but in case of spurious regression they are distributed as a random walk.

Generally speaking the linear combination of I (1) processes are themselves an I (1) process, but sometimes there is the possibility that a particular linear combination of two or more I(1) processes gives as result an I(0) process. This happens when the two I (1) processes taken into consideration shares a common stochastic trend. In this case we say that the processes are cointegrated CI (1,1). That is: $\mathrm{y_t} = (\mathrm{y_{1,t}}, \ldots, \mathrm{y_{k,t}})'$, $\mathrm{y_{j,t}} \sim \mathrm{I}(1)$, $\mathrm{j} = 1,2,\ldots,\mathrm{K}$ then $\mathrm{y_t}$ is cointegrated and it's indicated with C (1,1) if $\beta' \mathrm{y_t} = \beta_1 \mathrm{y_{1,t}} + \cdots + \beta_k \mathrm{y_{k,t}} \sim \mathrm{I}(0)$.

If we apply the Beveridge Nelson decomposition to the generic series x_t and y_t we can get to:

$$y_t = y_0 + \psi(1) \sum_{s=1}^{t} \epsilon_s^y + \psi^*(L) \sum_{s=1}^{t} \Delta\, \epsilon_s^y$$

$$x_t = x_0 + \psi(1) \sum_{s=1}^{t} \epsilon_s^x + \psi^*(L) \sum_{s=1}^{t} \Delta\, \epsilon_s^x$$

$Where$ $[\psi^*(L) \sum_{s=1}^{t} \Delta\, \epsilon_s^x\ ;\ \psi^*(L) \sum_{s=1}^{t} \Delta\, \epsilon_s^y]$ are the 2 stationary parts of the processes and $[\psi(1) \sum_{s=1}^{t} \epsilon_s^x\ ;\ \psi(1) \sum_{s=1}^{t} \epsilon_s^y]$ are the 2 nonstationary parts of the processes.

Now if we have that the 2 series are cointegrated at the end we are saying that the 2 nonstationary parts are the same one. Then if we spot some deviation from this common stochastic trend we could bet on the convergence of the series.

1.1.1 Non-stationarity and cointegration

Let's introduce 2 time series x_t and y_t which represent the prices of a pair of cryptocurrencies.

First of all I want to test the non-stationarity of the series, it's possible to test through various way, here I chose the Dickey-Fuller test, which is constructed as follows: taken into consideration the series x_t, we can rewrite it as the following AR (1) process:

$$x_t = \rho x_{t-1} + \epsilon_t$$

$$\epsilon_t \sim N(0, \sigma_\epsilon^2)$$

The hypothesis of the test are:

$$\begin{cases} H_0 : \ \rho = 1 \\ H_1 : \rho < 1 \end{cases}$$

The test statistic is the following:

$$t_T = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_t}}$$

Which converges to the following stochastic distribution:

$$t_T \xrightarrow{d} \frac{\left(\frac{1}{2}\right) [W(1)^2 - 1]}{\int_0^1 W(r) dr}$$

Where W(r) is the standard Brownian motion at time r.

This distribution has been tabulated by *Dickey and Fuller (1979)*[1]

Once we found that the processes contain a unit root we can proceed to the test for the cointegrating relation.

1.1.2. Vector Error Correction Model (VECM)

Let's start from the simplest case, where the system admits a VAR (1) representation

$$y_t = Ay_{t-1} + \epsilon_t$$

Subtracting y_{t-1} from both size we get

$$\Delta y_t = \Pi y_{t-1} + \epsilon_t$$

Where $\Pi = (A - I)$

In the univariate the matrix Π could be only 0 or different from 0, in the multivariate it's possible to have intermediate cases, which is precisely the one we want to study. Calling r the rank of the matrix Π we could have 3 scenarios:

1) r=0: in this case $\Pi = 0$ and so y_t is a multivariate random walk, there are no cointegrating relationship
2) r=n: y_t is not an I (1), but I (0) since Π is invertible
3) 0 < r < n: y_t is a cointegrating system

We want to focus on the last case: as we will see r is the cointegrating rank, furthermore Π could be rewritten as $\alpha\beta'$ where α and β are two (n x r) matrices; we call β the cointegrating matrix.

Put all these properties together we can rewrite the equation at the beginning of the paragraph as follows:

$$\Delta y_t = \alpha z_{t-1} + \epsilon_t$$

Where $z_T = \beta' y_t$ is an (r x 1) vector. Since β is the cointegrating matrix, we have that $z_t \sim I(0)$.

The variable z_t wants to capture the historical series of deviation from cointegrating relationship. Under these considerations it's easy to figure out that a cointegrating system is represented by 2 factors: one is a white noise (ϵ_t) while the other one (αz_{t-1}) is determined by the magnitude of the deviations from the cointegrating relation, captured at time t-1.

The matrix α is called loadings' matrix, his ij' s element tells us thich is the effect of i-th variable of the j-th element of z_{t-1}.

Example 1. *Let's take an example from quantity theory of money. Let's suppose that, in a time t, there is an excess of money supply with respect to the quantity given by the equilibrium relation. In this case we say that the circulation velocity is under his target value. Then, the variation between t and t+1 of the GDP and of the real money (in log) will be given by*

$$\begin{cases} \Delta y_{t+1} = \alpha_1(y_t - m_t) + \epsilon_{1t} \\ \Delta m_{t+1} = \alpha_2(y_t - m_t) + \epsilon_{2t} \end{cases}$$

Where $(m_t - y_t)$ can be seen as the log of the velocity of circulation; a lover value of the velocity ends up, in the successive period, in an adjustment of both the GDP and the real money supply.

If, let's say, α_2 is positive, this would mean that if we have an excess of money, Δm has to be negative, so m_t will be lower. This mechanism could happen, for example, for an increase of the prices.

This mechanism, where in the present time there is a relation to the difference from the same (lagged) variable and another (lagged) variable, is called Error Correction Mechanism; a VAR which is written in ECM form is called Vector Error Correction Model (VECM).

If a cointegrated system has cointegration rank r we can say that there exist r long run relations, so r stationary processes which are able to describe the path of the disequilibria from that relations over time. In case these disequilibria (which we called z_t) are different from zero there will be a movement of the vector y_{t+1} such that the disequilibrium will be reabsorbed.

We have to consider that the decomposition of the Π matrix is not unique, indeed for every non singular (r x r) matrix Ψ we can define $\alpha^* = \alpha\Psi'$ and $\beta^* = \beta\Psi^{-1}$ and get $\Pi = \alpha^*\beta^*$; from this peculiarity we can deduce that the cointegrating relations are not unique.

For this reason it is common to operate a normalization of the beta; typically a simple normalization is the following:

$$\beta^* = \begin{bmatrix} I_r \\ \beta_{K-r\,x\,r} \end{bmatrix}$$

In this way we are forcing the decomposition of the matrix Π to be unique and we will end up having $\beta' y_t = y_{1,t} - \beta_2 y_{2,t} - \cdots - \beta_k y_{k,t} = u_t \sim I(0)$. u_t is called cointegrating residual.

Generally speaking if y_t is a k-cointegrated element vector, there could be 0<r<k (linearly dependent) cointegrating relations, r, the number of cointegrating relations, is called cointegrating rank.

Taking as example K=3 and r=2, there will be a (k x r) matrix B such that:

$$\mathrm{B}' y_t = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} \sim \begin{bmatrix} I(0) \\ I(0) \end{bmatrix}$$

B is called cointegrating matrix.

1.1.3. Engle and Granger methodology

If we assume that the cointegrating matrix is feasible we can follow the Engle and Granger approach proposed in 1987 to estimate the cointegrating vector and an Error Correction Model to underline the long run relation and short-term dynamics.

The Engle and Granger approach contemplates an OLS estimation in the first step; then the second step consist into a unit-root testing in the residuals of the OLS estimation; if the residuals are I (1) then we have no cointegration in the series. In the case we reject the null of unit root the two series are C (1,1) and so the residuals are I(0). In this case the estimated coefficient previously obtained will form the cointegrating vector. If we apply this kind of procedure with ADF test we have to consider that we can't rely on usual critical values since the estimator is super-consistent, we have to rely to asymptotic distribution tabulated by Philips-Ouilaris (1990).

The OLS estimator for the cointegrating vector is super-consistent but we can obtain more accurate estimates with the Dynamic-OLS estimator (DOLS) proposed by Stock and Watson (1993) which is (asymptotically) efficient.

Once we obtained the cointegrating coefficient estimates we can proceed with the estimation of the following ECM:

$$\Delta \mathrm{y}_{\mathrm{i,t}} = \alpha_{\mathrm{i,0}} + \alpha_{\mathrm{i,1}}(\widehat{\mathrm{u}_{\mathrm{t}}}) + \sum\nolimits_{\mathrm{j}} \mathrm{a}_{\mathrm{i,j}} \Delta \mathrm{y}_{1,\mathrm{t-j}} + \cdots + \sum\nolimits_{\mathrm{j}} \mathrm{b}_{\mathrm{i,j}} \Delta \mathrm{y}_{\mathrm{k,t-j}} + \epsilon_{\mathrm{t}}$$

With i=1,2,…,k and $\widehat{u_t} \sim I(0)$ estimated cointegrating residual. the parameters $\alpha_{i,1}$ are called adjustment parameters and want to capture the adjustment speed with respect to the long-run equilibrium.

We have to consider that the Engle and Granger method, despite his simple and practice approach, presents a bunch of limitations:

- First of all we have to specify a variable as dependent and another one as independent, and so, as consequences, the results will depend on the normalization of the cointegrating vector, which is arbitrary
- This procedure could be used only in univariate case with 2 variables. This happen because we can estimate only 1 cointegrating vector, and so it's not possible to capture multiple cointegrating relations between more than 2 variables. If, for example, we would have 3 variables, X, Y and Z, it couldn't be possible to capture pairwise cointegrating relationship. Furthermore the cointegrating relationship could be present only in n<N variables of the system, this aspect is not captured properly from the Engle and Granger methodology since it wants to estimate a unique cointegrating vector, and if we insert in this unique cointegrating vector a parameter for a variable which is not cointegrated we are going to erase the consistency of the estimator. As we saw previously with N variable we could have r<N cointegrating relations, and so, r cointegrating vectors, which in the Engle and Granger methodology are not captured.

1.1.4 Granger Representation Theorem

When we switch from the univariate to the multivariate case then a new series of intermediate scenarios are opened: in the bivariate the rank of the cointegrating matrix could be only full or zero, while in the multivariate we could have rank=r<N. This dynamic open up a series of cases which are not present in the univariate. To start to talk about multivariate representation of non-stationary and cointegrated processes we have to introduce the Granger representation theorem.

To explore the properties of a cointegrated system y_t there are at least 2 possibilities: one linked to the fact that the system can be wrote as a n-order VAR (also infinite)

$$A(L)y_t = \epsilon_t$$

With the associate ECM representation, and linked to the fact that if y_t is an I (1), than Δy_t is an I(0), and so has to have a Wold representation of the type

$$\Delta y_t = C(L)\epsilon_t$$

To establish which kind of relations links these two representations we have to exploit the Granger representation theorem.

Granger Representation Theorem: *if a system of non-stationary processes can be written in ECM form than it has to be cointegrated, and if a system of non-stationary processes is cointegrated, then it has an ECM representation.*

Proof: *Assume, for simplicity, that* $\Gamma = I_N$, *given an ECM of the form*

$$y_t = \Psi \sum_{i=0}^{t} \epsilon_i + x_t$$

Where $\Psi = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp$. *It must hold that*

$$\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp + \alpha(\beta'\alpha)^{-1}\beta' = I_N$$

And so

$$y_t = [\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp + \alpha(\beta'\alpha)^{-1}\beta']y_t$$

$$y_t = \Psi y_0 + \Psi \sum_{i=0}^{t} \epsilon_i + \alpha(\beta'\alpha)^{-1} \sum_{i=0}^{\infty} (I_N + \beta'\alpha)^{-1}\beta'\epsilon_{t-i}$$

The condition $|\alpha'_\perp\beta_\perp| \neq 0$ *ensures that* $y_t \sim I(1)$ *and that* $\beta' y_t \sim I(0)$

The point of the theorem is that a cointegrated system can be expressed in a VAR form and in a MA form.

To derive the VAR form we start from the following VAR model for the variable y

$$\Phi(L)y_t = \epsilon_t$$

With L that is the usual Lag operator with the following characteristics:

$$L^k\,(X_t) = X_{t-k}$$

$$L^{(-k)}(X_t) = E[X_{t+k}]$$

The VAR can be rewritten as follows:

$$y_t = \Phi^*(L)y_{t-1} + \epsilon_t$$

Where

$$\Phi^*(a) = \big(I_N - \Phi(a)\big)a^{-1}$$

By applying the Beveridge-Nelson decomposition to $\Phi^*(L)$ we get

$$y_t = [\Phi^*(1) + \ \Phi^{**}(L)(1-L)]y_{t-1} + \epsilon_t$$

If we subtract y_{t-1} from both members we get

$$\Delta y_t = (\Phi^*(1) - I_N)y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} \ + \epsilon_t$$

Where

$$\Gamma(L) = \ I_N - [\big(\Phi_2 + \cdots + \Phi_p\big)L + \cdots + \Phi_p L^{p-1}]$$

Now denote that $\Phi^*(1) = I_N - \Phi(1)$ to rewrite

$$\Delta y_t = -\Phi(1)y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} \ + \epsilon_t$$

Which is the VECM form for the representation of a cointegrating system.

The Moving Average representation could be intended as a sort of Beveridge Nelson decomposition extension to the multivariate case, which appears as we previous saw:

$$y_t = [\beta_\perp \, \alpha'_\perp]\mu_t + \Psi^*(L)\epsilon_t$$

Where μ_t is defined by the property $\Delta\mu_t = \epsilon_t$ and so, since ϵ_t is a vectorial white noise, μ_t will be a vectorial random walk by definition.

$$\begin{cases} [\beta_\perp \, \alpha'_\perp]\mu_t \ \rightarrow non\ stationary\ component \\ \quad \Psi^*(L)\epsilon_t \rightarrow stationary\ component \end{cases}$$

It's important to denote that the matrix $[\beta_\perp \, \alpha'_\perp]$ has the property that it's killed if we multiply it for β', for the properties of orthogonal operator. In other words, the matrix $[\beta_\perp \, \alpha'_\perp]$ is nothing else than $\Psi(1)$, which is singular with rank (n – r).

Now we want to try to give a representation of a cointegrated system in order to spot his stationary and non-stationary part, let's start from the following series:

$$\eta_t = \alpha'\mu_t$$

η_t is a random walk with dimension (n – r) since the matrix $\alpha_\perp$ is an n x (n-r) matrix. So it's possible to rewrite the process y_t introduced before as

$$y_t = F\eta_t + \mu_t$$

Where $F = \beta_\perp H$ is a n x (n-r) matrix and $\mu_t = C^*(L)\epsilon_t$ is, by assumption, a stationary process. Writing y_t in this way allow us to clearly see that every cointegrated system can be seen as a system where there exists a certain number (n-r) of unobservable stochastic trends which are revealed through the matrix F, so the series observed through y_t contain an I(1) part, given by the linear combination of these common stochastic trends, and an I(0) part, given by μ_t. The cointegration relation exists because n - r < r, so the linear combination $\beta' y_t$ is a stationary because, for the construction and the properties of orthogonal operator, $\beta' F$ = 0. Practically, by multiplying β' by y_t we are killing the stochastic common trends.

Example 2: *let's assume that we have 2 cointegrated series y_t and x_t, with $\beta = (1, -1)'$; consequently, we will have that $z_t = x_t - y_t$ is an I (0). In this case H is a scalar, and $F = \beta_\perp H$ is proportional to the vector (1,1)'.*

The two series, so, can be described as a sum of I (1), which is common in both processes, plus a stationary part. It's clear that the cointegration derives from the fact that taking the difference from the two series we are going to kill the common stochastic trend.

What we have just seen needs an integration where we are treating with VAR or VECM which present a drift or a deterministic linear trend. Usually, in the univariate, a drift in the difference series generates a polynomial of order greater than 2 in the series in levels, in practice, having a deterministic linear trend in the first differences would mean to have a quadratic deterministic trend in levels. However this is not always true if we switch from univariate to multivariate for the representation in levels of VAR and in difference for VECM when we have to deal with a cointegrated system.

Let's imagine having a VAR with drift d_t

$$y_t = d_t + \beta y_{t-1} + \epsilon_t$$

$$\Delta y_t = \Pi y_{t-1} + d_t + \epsilon_t \,, \qquad with\, u_t \equiv d_t + \epsilon_t$$

$\Delta y_t = \Pi y_{t-1} + u_t$ is a VECM with the deterministic component linked with the error

Now recall the Beveridge Nelson decomposition of the Granger representation:

$$y_t = [\beta_\perp \, \alpha'_\perp]\mu_t + \Psi^*(L)\epsilon_t$$

For simplicity put $[\beta_\perp \, \alpha'_\perp] = \Psi(1)$ in order to get

$$y_t = \Psi(1)\mu_t + \Psi^*(L)\epsilon_t$$

Now we can rewrite the previous as

$$y_t = \Psi(1)\widehat{\mu_t} + \Psi^*(L)\epsilon_t$$

Where $u_t = \Delta\widehat{\mu_t} = d_t + \epsilon_t$. The process $\widehat{\mu_t}$ is then composed by a multivariate random walk plus a deterministic component which should be of order equal to the one in difference plus 1.

To understand why this is not always true let's recall the original Granger form of the process

$$y_t = [\beta_\perp \, \alpha'_\perp]\widehat{\mu_t} + \Psi^*(L)\epsilon_t$$

If $\Delta\widehat{u_t} = d_t + \epsilon_t$, the polynomial in $\widehat{u_t}$ will be of the typt $d_t t$. But now we have to take into account that in the cointegration this last polynomial, in levels, will be multiplied by $\alpha'_\perp$. In the case where $\widehat{u_t}$ is a linear combination of the columns of $\alpha_\perp$ the product $\alpha'_\perp u_t = 0$, so it will kill the polynomial $d_t t$.

In other words we will have a VECM with the intercept but that intercept doesn't impact on the levels like a deterministic trend. The process so will not present a linear trend in the time but will always move around a value different from zero, this implies that the deviations from cointegrating relations will have a non-zero mean and so the intercept will be present anyways in the cointegrating relation.

We could have the following cases:

1) $d_t = 0$. Here we have no problem, no linear trend.

2) $d_t \neq 0$ and $\alpha'_\perp u_t = 0$. Here we are facing an intercept in the differences, but that intercept will not generates a deterministic trend in levels.
3) $d_t \neq 0$ and $\alpha'_\perp u_t \neq 0$. Presence of intercept in the VECM and of deterministic trend in the VAR.
4) $d_t = d_0 + d_1 t$ and $\alpha'_\perp u_t \neq 0$. Deterministic trend in the VECM which will generate a quadratic deterministic trend in the VAR. So the common trend observed in the series, already stochastic by themselves, will also show a quadratic trend over time.
5) $d_t = d_0 + d_1 t$ and $\alpha'_\perp u_t = 0$. Here the deterministic trend in the VECM will not generate a quadratic trend in levels as we saw for point 4.

1.1.5 Johansen cointegration procedure

The Johansen estimation procedures is built on the assumption that the cointegrating system can be represented as a VAR(N) with $N < \infty$ with gaussian error.

The starting point is to rewrite the system as a VECM:

$$\Delta y_t = d_t + \Pi y_{t-1} + \sum_{i=1}^{p} \Gamma_i \Delta y_{t-i} + \epsilon_t$$

Here we are implicitly assuming that the P order is known; to find the optimal p one can freely use the usual hypothesis tests and information criteria to come up with an optimal number of lags.

Now, once is assumed that the optimal P is known, we can go through the estimation of VECM under a maximum likelihood context. Usually the maximum likelihood estimator under a linear regression setup with gaussian error is the classic OLS, but in this case we have to consider all the problems linked to the presence of the cointegrating relation: first of all the rank of the matrix Π is equal to the cointegrating rank r, so we want our estimator to come up with an estimation of Π with reduced rank, which the OLS is not able to give us.

The first problem to arise is to quantify the cointegrating rank r. The Johansen procedure establish 2 tests of the cointegrating rank of the matrix Π. The test are linked to the fact that in a positive semidefinite matrix the number of positive eigenvalues is equal to the rank of the matrix, the other eigenvalues are zeros.

The tests work as follows:

1) It's defined a positive definite matrix M which has the same rank as Π. We define a new matrix instead of working directly with Π in order to have all his eigenvalues real and non-negative.
2) It's built a consistent estimator of M, $\widehat{M}$, with the consequences that all eigenvalues of $\widehat{M}$, let's call them $\hat{\lambda}$, are consistent estimators of the n eigenvalues of M.
3) All eigenvalues of the matrix are sorted from the largest $\hat{\lambda}_1$, to the smallest $\hat{\lambda}_n$, and we test the positiveness of all of them.

If we reject the null, so we found that $\hat{\lambda}_n$ is positive, than we have that all eigenvalues are positive and the matrix Π has full rank and the system is stationary. Otherwise we consider $\hat{\lambda}_{n-1}$, and now we proceed by the following configuration:

1) We can build a test with the following hypothesis configuration:

$$\begin{cases} H_0 : \hat{\lambda}_{n-1} = 0 \\ H_1 : \hat{\lambda}_{n-1} \neq 0 \end{cases}$$

Here we are implicitly assuming that r < n and so we are testing the hypothesis that r < n-1; this is called λ-max test.

2) Alternatively we can go through the following hypothesis testing:

$$\begin{cases} H_0 : \hat{\lambda}_{n-1} = \hat{\lambda}_n = 0 \\ H_1 : \hat{\lambda}_n \neq 0 \ or \ \hat{\lambda}_{n-1} \neq 0 \end{cases}$$

This is called trace-test.

If the null is not rejected we keep moving on to test $\hat{\lambda}_{n-2} = 0$, and so on until we reject the null hypothesis $\hat{\lambda}_{n-k} = 0$; once we reject the null we found the rank of the matrix Π. If we don't reject the null in any case the rank of the matrix is null and so the system is not cointegrated.

To better capture the dynamics we are dealing with it's important to compare the equation of the VECM with the one of the dickey fuller test saw at the beginning of the chapter:

$$\text{The VECM equation is: } \Delta y_t = d_t + \Pi y_{t-1} + \sum_{i=1}^{p} \Gamma_i \Delta y_{t-i} + \epsilon_t$$

$$\text{The ADF} - \text{test equation is: } \Delta y_t = d_t + \rho y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \epsilon_t$$

As on can easily see the VECM equation represents nothing else than the extension of the ADF-test equation to the multivariate world. The only difference is that the ρ coefficient is a scalar in the ADF and so it is null or invertible, in the multivariate we have the intermediate case where the matrix Π is singular without being null, we are interested in this one.

The analogy with the dickey-fuller test is captured also by the asymptotic distribution of the trace-test, which, at his core, is nothing else than a likelihood-ratio test which is built as follows:

$$\begin{cases} H_0 : rank(\Pi) \leq r \\ H_1 : rank(\Pi) \leq N \end{cases}$$

Where the test statistic takes the form

$$LR(H_0|H_1) = 2\big(\mathcal{L}_t(H_N) - \mathcal{L}_t(H_r)\big)$$

Is distributed asymptotically as

$$LR(H_0|H_1) \xrightarrow{d} tr\left(\int_0^1 (dW)W'\left(\int_0^1 WW'du\right)^{-1}\int_0^1 w(dW)'\right)$$

Which is nothing else than a generalization of the ADF distribution to the multivariate (N-r) case.

Wanting to rewrite the ADF hypothesis under a multivariate look we will end up having the following situation:

$$\begin{cases} H_0 : rank(\Pi) = 0 \\ H_1 : rank(\Pi) = N \end{cases}$$

Recall that in the univariate N = 1, so the null is associated with I(1) process and the alternative with a stationary process.

In the Johansen test these distribution are not indifferent to the deterministic part of the VECM, which as we saw before is messy itself. Here it plays an important role the degree of the polynomial we are going to insert in the deterministic part. Usually the choice is between a plain constant or a constant plus a trend.

Once we found the optimal cointegrating rank we can proceed to the estimation of the β.

Johansen proposed a method called reduced rank regression (RRR) which is developed as follows:

Start from the multivariate generic regression:

$$y_t = Bx_t + Cz_t + e_t$$

Where B is a square matrix with reduced rank and C has full rank, in the VECM setup we have

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{p} \Gamma_i \Delta y_{t-i} + \epsilon_t$$

The analogy is clear:

$$\alpha\beta' = B$$

$$\Gamma_i = C$$

The RRR method, denoted as RRR$(y, x|z)$ relies on the following product matrices

$$S_{yx} = T^{-1} \sum_{t=1}^{T} y_t x_t'$$

$$S_{yx,z} = S_{yx} - S_{yz} S_{zz}^{-1} S_{zx}$$

The method is performed by applying the following steps:

1) Regress y and x on z with OLS, to obtain the following residuals

$$(y|z) = y_t - S_{yz}\, S_{zz}^{-1}\, z_t$$

$$(x|z) = x_t - \; S_{xz}\, S_{zz}^{-1}\, z_t$$

   And the product moments

$$S_{yx,z} = T^{-1} \sum_{t=1}^{T} (y|z)_t (x|z)_t' = S_{yx} - S_{yz} S_{zz}^{-1} S_{zx}$$

2) Solve the following eigenvalue problem

$$\left|\lambda S_{xx,z} - S_{xy,z}\, S_{yy,z}^{-1}\, S_{yx,z}\right| = 0$$

And sort the eigenvalues from largest to smallest in order to have $\Lambda = diag(\lambda_1 , \dots , \lambda_N)$ and the associated eigenvectors are $V = (v_1, \dots , v_N)$, where the eigenvectors are normalized so that $V'S_{xx,z}V = I_p$ and $V'S_{yx,z}S_{xx,z}^{-1}S_{xy,z}V = \Lambda$

3) Finally, the estimator of β is

$$\hat{\beta} = (v_1, \dots , v_r)$$

With $\hat{\alpha} = S_{yx,z}\hat{\beta}$ and $\widehat{\Omega} = S_{yy,z} - S_{yx,z}\hat{\beta}\left(\hat{\beta}'S_{xx,z}\hat{\beta}'\right)^{-1}\hat{\beta}'S_{xy,z}$ which are given by simple regression given $\hat{\beta}$.

An important thing to note is the difference between $\widehat{\Pi}_{OLS}$ and $\widehat{\Pi}_{RRR}$:

$$\widehat{\Pi}_{OLS} = S_{yx,z}S_{xx,z}^{-1}$$

$$\widehat{\Pi}_{RRR} = S_{yx,z}\hat{\beta}\left(\hat{\beta}'S_{xx,z}\hat{\beta}'\right)^{-1}\hat{\beta}'S_{xy,z}$$

The principal problem here is, as we saw in paragraph 1.1.2 the matrix β is not identified, in fact, if a generic matrix β is a cointegrating matrix, also the matrix $b = \beta K$ is a cointegrating matrix, with K which is a generic (r x r) non-singular matrix. So there will be an infinite number of $n\ x\ r$ matrices which are all equivalently cointegrating matrix.

This problem can be represented as follows:

Suppose we known Π; now we can represent the matrix as

$$\Pi = \alpha\beta'$$

Or, equivalently, as

$$\Pi = \alpha K^{-1}K\beta' = ab'$$

Both versions of Π are valid candidates to describe the cointegrating relations of the system, so we are under a classic problem of under-identification.

How does this problem can be solved?

In the original procedure proposed by Johansen the identification is obtained by imposing restriction on a quadratic of β. An alternative approach, already presented at the beginning of the chapter, is the so called triangular representation, which makes the constraint on the first *r* rows of the matrix β imposing them to be an identity matrix. So the matrix beta becomes:

$$\hat{\beta} = \begin{bmatrix} I \\ -\hat{\beta}_2 \end{bmatrix}$$

Where $\tilde{\beta}_2$ is the parameter found by estimation procedure.

If we are under the assumption of i.i.d. gaussian error, then the conditional log-likelihood is

$$\mathcal{L}_T(y_t|F_{t-1}) = -\frac{T}{2}\log|\Omega| \\ -\frac{1}{2}\sum_{t=1}^{T}\left(\mathrm{y_t} - \alpha\beta' y_{t-1} + \sum_{i=1}^{p}\Gamma_i \Delta y_{t-i}\right)' \Omega^{-1}\left(\mathrm{y_t} - \alpha\beta' y_{t-1} + \sum_{i=1}^{p}\Gamma_i \Delta y_{t-i}\right)$$

Johansen (1995) showed that in case of RRR the log-likelihood function reduces to

$$\mathcal{L}_T(y_t|F_{t-1}) = -\frac{T}{2}\left(\log|S_{yy,z}| + \sum_{i=1}^{r}\log(1-\lambda_i)\right)$$

The β estimator has some unusual asymptotic properties: it is super-consistent. Usually when an estimator $\hat{\theta}$ is consistent we have that, as T increases, $\hat{\theta} \xrightarrow{p} \theta_0$, and so $\hat{\theta} - \theta_0 \xrightarrow{p} 0$. Multiplying this quantity by $\sqrt{T}$ we obtain that $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \sigma_\theta^2)$. When we have this situation we say that the speed of convergence is $\sqrt{T}$. In our case, instead, in order to obtain a normal distribution the difference $(\hat{\beta} - \beta)$ has to be multiplied by T, instead that by $\sqrt{T}$. So we say that the speed at which the estimator $\hat{\beta}$ collapses to β is T, and so the dispersion of the estimator is proportional to T^{-1} and not to $T^{-\frac{1}{2}}$ anymore. This property doesn't touch the finite sample estimates, but has some important consequences on the long run properties of the estimator.

Johansen (1995) showed that $\hat{\alpha}_i, \hat{\Gamma}_i$ and $\hat{\Omega}$ are consistent and asymptotically Gaussian, once we found the $\hat{\beta}$ and inserted in the VECM, all the other parameters can be found wit OLS.

1.2. Long/short strategy based on cointegration

Now that the concept of cointegration and VECM have been clarified I can go through the creation of a trading strategy starting from the idea of cointegration.

The basic concept on which all the model is built up is that if 2 prices are cointegrated than they share a sort of equilibrium between them and so, it is possible to spot that equilibrium and to exploit the deviations from that trend.

1.2.1. Introduction to pair trading

The idea of this strategy is originally applied to stock market, since the cryptocurrency market is a relatively new environment,

Long/Short trading is a relatively straightforward investment strategy. The logic behind such a strategy is simple: some stocks have very similar fundamentals and their prices should move together in lock steps. When they move apart an opportunity for speculative profits exists as eventually such prices will move back to their equilibrium values. A sophisticated investor should then:

i) identify stocks which move closely together;

ii) take long/short positions in such assets when their prices diverge by a sufficiently wide margin;

iii) close their position when either the prices cross back or a stop-loss limit is reached.

Clearly the difficulties in implementing pair trading lie in identifying the pairs of stocks that do move closely together.

When pair trading was first introduced the main keys indicators which were used to identify these pairs was company fundamentals such as ratio metrics which was derived by accounting and financial statements. With advent of modern technologies and more advanced econometric techniques, such as the one saw in the last paragraph, it' s relatively easy and more accurate to spot pairs of stocks which moves together.

In his simplest formulation the pair trading wants to measure the tracking variance between 2 normalized prices, defined as follows:

$$Q_t^A = \frac{P_t^A}{P_1^A}$$

So the normalized price of the security A at time t is the price of the security divided the price at the beginning of the observation period.

The tracking variance of the prices is defied as:

$$TV = \frac{1}{T}\sum_{t=1}^{T}(\,Q_t^A - Q_t^B)^2$$

On his basis, the pair trading strategies wanted to bet on the deviation from the mean of this value in order to gain from his mean-reversal behaviour.

Gatev et al. (2006) proposed a model based on the tracking variance stat: whenever the spread of a pair exceed a given threshold a pair trading position is open, where the most expensive stock is sold and the least expensive one is bought; the position is closed when the normalized prices cross back.

The threshold value is usually defined in terms of standard deviation of tracking variance of average price distance. Thus, if TV is the tracking variance between the normalized prices of stocks A and B as defined before, the standard deviation of such estimated distance is:

$$SD = \left(\frac{1}{T-1}\sum_{t=1}^{T}[Q_t^A - Q_t^B - (\bar{Q}^A - \bar{Q}^B)\,]^2\right)^{\frac{1}{2}}$$

Usually the threshold value will be twice the standard deviation. Defined $\Delta_t = Q_t^A - Q_t^B$, then the pair trading position is triggered when $|\Delta_t|$ is larger than 2 times the standard deviation. The pair trading position will be closed when either the spread moves back to zero or it widens so much as to exceed some stop-loss maximum value.

1.2.2. Cointegrated prices and pair trading

Now that there are the basis for the understanding of what is the core of the pair trading strategy it's possible to expand the idea exploiting the long run relationship introduced by the cointegration.

The core concept is the following: one we spot a pair of stocks (or, in our cases, cryptocurrencies), defined the following relationships:

$$\begin{cases} P_t^X = Price\ of\ Stock\ X\ at\ time\ t \\ P_t^y = Price\ of\ Stock\ Y\ at\ time\ t \\ p_t^X = \log_e(P_t^X) \\ p_t^Y = \log_e(P_t^Y) \end{cases}$$

We know that there will be a matrix β with the form $\beta = [1\ ;\ -\beta_2]'$ such that the quantity $p_t^A - \beta_2 p_t^B$ will be stationary. On this assumption it's possible to built up a strategy which, whenever the quantity $\delta_t = p_t^A - \beta_2 p_t^B$ exceed a certain limit, it's triggered to bet on his tendency to come back around it's middle value. Now suppose the return on the strategy is calculated over the interval [t,t+1]. Using log returns it is straightforward to find that

$$r_{t+1}^{p} = [p_{t+1}^{A} - p_{t}^{A}] - \beta[p_{t+1}^{B} - p_{t}^{B}]$$

Which can be rearranged as

$$r_{t+1}^{p} = [p_{t+1}^{A} - \beta p_{t+1}^{B}] - [p_{t}^{A} - \beta p_{t}^{B}] = \delta_{t+1} - \delta_{t}$$

As long as the log-prices are cointegrated and $[p_t^A - \beta p_t^B]$ is stationary, the spread δ_t will have an expected value μ which can be estimated from data. A statistical arbitrage strategy is triggered when $|\delta_t - \mu|$ is larger than some threshold ξ^{open} and closed when the quantity $|\delta_t - \mu|$ is smaller than another threshold ξ^{close}.

A crucial point here is how to construct the threshold ξ: the simplest possibility is to set it as a multiple of standard deviation of δ. I started with this one as benchmark and tested it against a moving average of the variance and the result was less efficient. So the construction of the threshold has been made by picking multiples of standard deviation of δ.

In order to construct this strategy, whenever there is a deviation from the mean, once defined ξ^{open} *and* ξ^{close} it's necessary to operate as follows:

$$\begin{cases} \delta_t > \xi^{open} => open\ short\ \delta\colon \begin{cases} sell\ p_t^A \\ buy\ \beta_2 p_t^B \end{cases} \\ \delta_t < \xi^{close} => close\ short\ \delta\colon \begin{cases} buy\ p_t^A \\ sell\ \beta_2 p_t^B \end{cases} \end{cases}$$

And inversely:

$$\begin{cases} \delta_t < -\xi^{open} => open\ long\ \delta\colon \begin{cases} buy\ \ p_t^A \\ sell\ \beta_2 p_t^B \end{cases} \\ \delta_t > -\xi^{close} => close\ long\ \delta\colon \begin{cases} sell\ p_t^A \\ buy\ \beta_2 p_t^B \end{cases} \end{cases}$$

2. Empirical Analysis

In this chapter I am going to exploit the reasons why I choose the cryptocurrency as asset for the test of the strategy and the methodology applied for the empirical study.

First of all the crypto market is a very young market and so it could be possible that for this market there is a common trend moving at least a bunch of cryptocurrencies.

Since the principal cryptocurrency is the Bitcoin it is reasonable to think that it's this one a good candidate to be the common mover to the other currencies, the common trend, so I decided to take the prices of the currencies with respect to the Bitcoin for the analysis.

2.1. Data

In this work the dataset is composed by daily price which go from 30-09-2017 to 18-01-2020 for the building of the trading strategy and from 19-01-2020 to 23-10-2020 for the test of the model, the test sample size is approximately the 30% of the train sample size.

The training and trading sets are strictly non-overlapping to ensure that no look-ahead bias is introduced.

The source of the overall dataset is yahoo finance, but for the couple selected to construct the model I wanted to include also the liquidity of the cryptos so the test size (out of sample) is performed on the bid/ask price downloaded from binance.

For each day the only parameter took into consideration is the adjusted price since the other parameters (volume, daily high, daily low, open price, close price) are not useful for the development of the model.

2.2. Software

The code for this study is written in Python 3.5 (Python Software Foundation 2016). It involves the pre-processing and formatting of the data, the training of the models and the back testing engine, as well as the evaluation of the performance, i.e., the calculation of risk and return metrics.

Data preparation mostly relies on the packages and pandas, which are powerful tools for handling large amounts of data. Furthermore, the packages *SciPy* and *Empyrical* are deployed for the calculation of the statistical properties and performance analysis of the results *(Fisher & Krauss, 2019)*.

2.3. Methodology

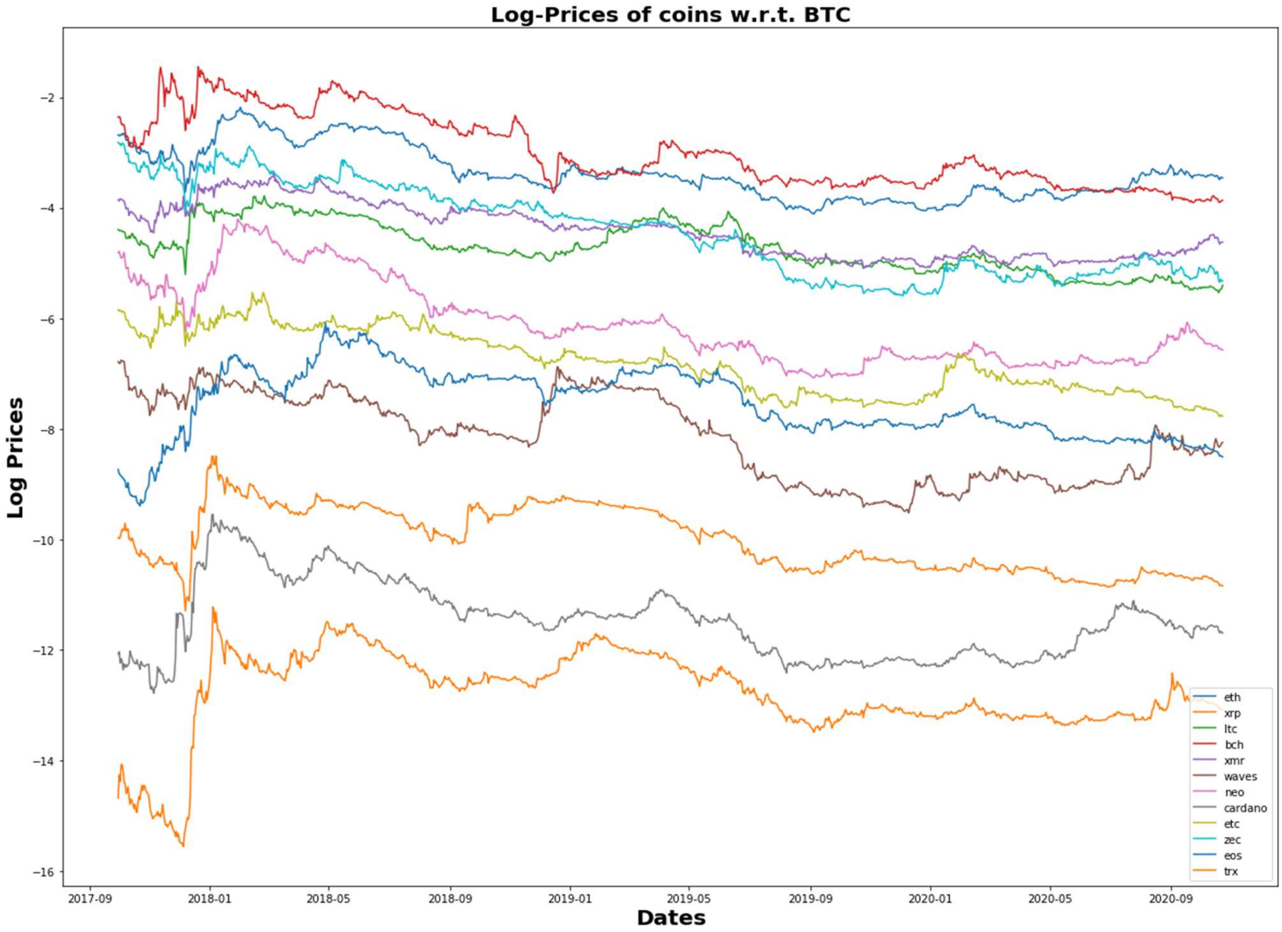
The methodology consists in the following steps:

1) Splitting of the data in train and test part.
2) Spotting of the cointegrated pairs.
3) Building the triggers for opening and close position.
4) Back test the strategy on the in-sample period.
5) Test the strategy on the out-of-sample period.
6) Evaluation of the results

Some consideration: this kind of trading strategy doesn’t takes into account all the fundamental characteristics that are properly of every cryptos, and so it’s not possible for the model to capture if there is a shift or a modification in that fundamentals. For this reason I choose to consider only the last part of 2017, in order to don’t let a one-time situation to influence all the study.

2.3.1. *Spotting Cointegrated Pairs*

The pairs from which the data frame is formed are taken from the most 50 capitalized cryptocurrencies, and the prices are taken, as we mentioned, with respect to bitcoin. The overall data frame appears as following:



Also from a qualitative analysis it's easy to spot that there is a common mover in these series, especially in the first part of the series there is a common increase in the prices, which correspond to the bubble of December 2017 and that it's common to the all series included in the data frame.

In order to spot the cointegrated pairs I run the Engle-Granger cointegration test saw in the previous chapter simplified for the bivariate case, so with the following hypothesis setting:

$$\begin{cases} H_0 : y_1 \text{ and } y_2 \text{ are not cointegrated} \\ H_1 : y_1 \text{ and } y_2 \text{ are cointegrated} \end{cases}$$

The p-values of the tests are resumed in the following table

	eth	xrp	ltc	bch	xmr	waves	neo	cardano	etc	zec	eos	trx
eth	0,00	0,42	0,37	0,07	0,13	0,40	0,02	0,16	0,47	0,13	0,08	0,32
xrp	0,37	0,00	0,12	0,30	0,30	0,10	0,19	0,04	0,11	0,17	0,03	0,01
ltc	0,34	0,12	0,00	0,26	0,16	0,15	0,15	0,01	0,16	0,23	0,03	0,12
bch	0,06	0,32	0,27	0,00	0,13	0,24	0,05	0,18	0,21	0,02	0,23	0,32
xmr	0,12	0,33	0,17	0,13	0,00	0,22	0,02	0,06	0,03	0,10	0,05	0,29
waves	0,39	0,12	0,16	0,25	0,23	0,00	0,48	0,07	0,37	0,32	0,02	0,14
neo	0,02	0,23	0,16	0,05	0,02	0,48	0,00	0,05	0,19	0,26	0,02	0,15
cardano	0,13	0,04	0,01	0,16	0,06	0,06	0,04	0,00	0,02	0,06	0,04	0,13
etc	0,44	0,12	0,16	0,21	0,02	0,36	0,18	0,02	0,00	0,38	0,01	0,12
zec	0,14	0,26	0,30	0,03	0,12	0,35	0,28	0,11	0,45	0,00	0,04	0,12
eos	0,05	0,02	0,02	0,18	0,04	0,01	0,01	0,04	0,01	0,01	0,00	0,01
trx	0,24	0,01	0,10	0,28	0,24	0,10	0,10	0,13	0,10	0,05	0,01	0,00

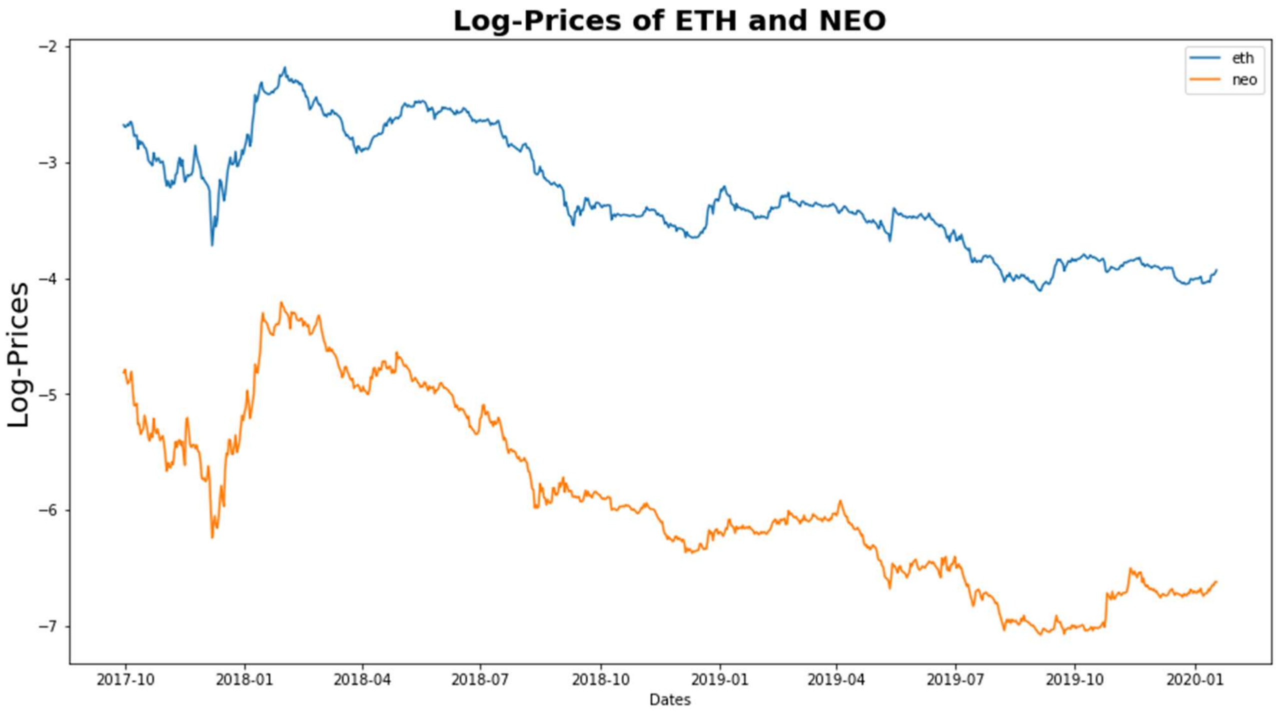
In order to don't have some border-line results I chose to pick the couples for which the statistical significance threshold is 4% instead of the canonical 5%.

Despite the more stringent threshold there are 17 pairs of cointegrated cryptocurrencies with the cointegrating relation which is statistically significant at 4%.

Among all the possible pairs I chose to take the couple ETH/NEO to build up the model and the strategy, but it could be followed as the same exact procedure using another couple, from now on I will imply that the couple used is that one.

2.3.2. Building the open and close triggers

Once selected the couple which will be used for the development of the trading strategy I want to operate a series of test in order to check whether this couple is appropriate to the application of the cointegration theory: the first step is to verify the non-stationarity.



Also from a qualitative point of view it's easy to spot that the two series can't have the same variance in all the observed periods: at the beginning of the series there is a huge spike in the prices which after becomes more smooth, so this is a great signal for the assessing of non-stationarity.

If we want to formalize the assessment of non-stationarity we have to operate a statistical test, the principal tests used to asses the non-stationarity of the series are the augmented Dickey-Fuller, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) and the Philips-Perron test: the DF test and the PP test assess the unit root under the null hypothesis while the KPSS test put the null as the stationarity.

The results of the tests are reported in the following table:

		Neo	Eth
PP	Test-Stat	-1.347	-1.185
	P-Value	0.608	0.68
KPSS	Test-Stat	3.766	3.594
	P-Value	0	0
ADF	Test-Stat	-1.141	-1.327
	P-Value	0.699	0.617

As we can denote all the p-values allows us to strongly assess the non-stationarity of the series, so we can proceed with the step 2 of the preliminary analysis for the construction of the model.

The second step is to check whether there is cointegration between the 2 series, since we picked them up from the results of Engle-Granger test, we know that there is a cointegrating relation among them so in theory it is possible to skip this step, anyway I wanted to test the presence of cointegration relation also with Johannsen trace-test and max eigenvalue test.

Since these test have a non-standard asymptotically distribution I will report only the test statistic and the critical values, recall the hypothesis distribution for this test:

$$\begin{cases} H_0 : rank(\Pi) = 0 \\ H_1 : rank(\Pi) = N \end{cases}$$

		R=0	R=1
Trace-Test	**Test-Stat**	13.11	2.08
	Critical Value (5%)	10.47	2.97
Max-Eig	**Test-Stat**	11.03	2.08
	Critical Value (5%)	9.47	2.97

Also in this case the results are not in contrast with what showed from the preliminary analysis, we can strongly asses that there is a cointegrating relationship between the two variables taken into consideration.

Once assessed that two variables are cointegrated the next step is to estimate the cointegrating coefficient and for that I exploited the Johansen procedure saw in the previous chapter, the model I want to estimate is the following VECM

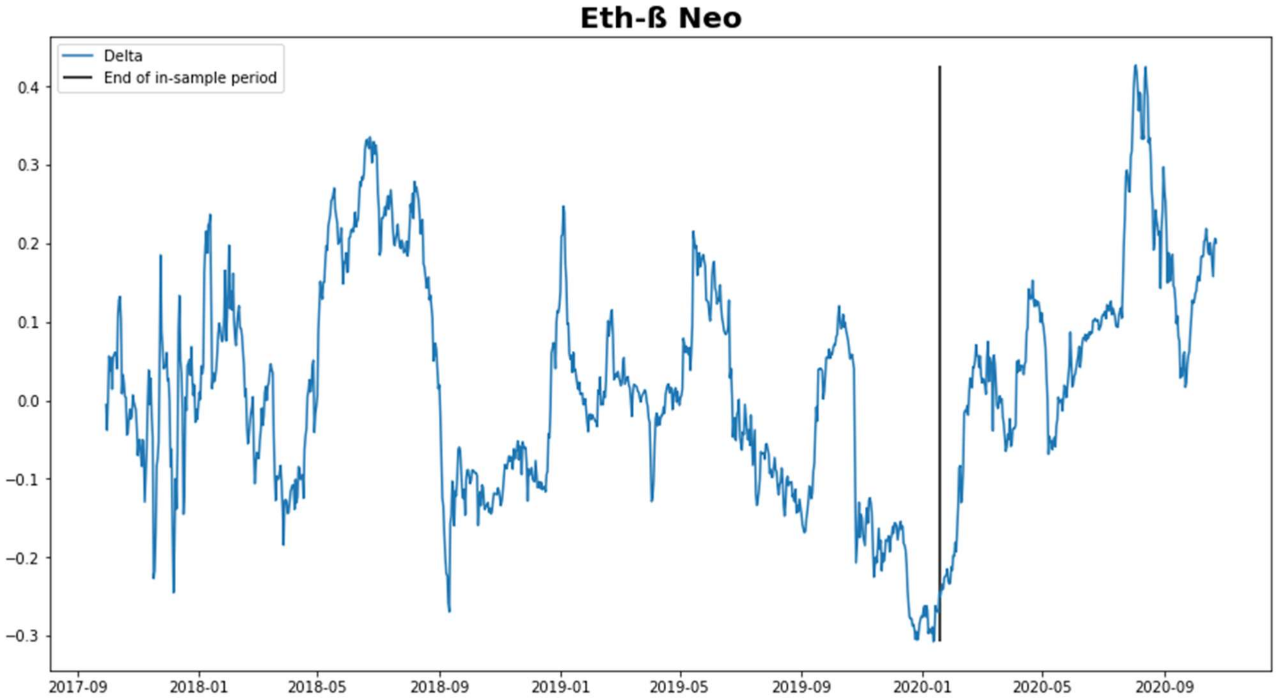
$$\begin{cases} y_t^{eth} - y_{t-1}^{eth} = \alpha^{\mathrm{eth}}\left(y_{t-1}^{eth} - \beta y_{t-1}^{neo}\right) + \epsilon_t^{eth} \\ y_t^{neo} - y_{t-1}^{neo} = \alpha^{neo}\left(y_{t-1}^{neo} - \beta y_{t-1}^{et}\right) + \epsilon_t^{neo} \end{cases}$$

Even if the model is bivariate for simplicity I will take in consideration only the first equation for the analysis and the performing of the strategy.

Now we have to make a consideration about the robustness of the analysis: since in our sample T=838 one can be worried about the efficiency of the parameter estimate; to analyse the variability of the estimate in this specific situation I performed an estimate on a simulated model which I will provide to explain in the following paragraph.

Once I estimated the cointegrating β it's possible to run an analysis on $\delta_t = y_t^{eth} - \beta y_t^{neo}$ in order to check whether itself satisfies some property which allows us to continue with our analysis.

First of all I want to check if the quantity is stationary, and for this I can use the same test used for the assessing of non-stationarity of the time series taken into consideration, for which the results, together with a visualization of the δ quantity, are reported below



		$\delta = y^{eth} - \widehat{\beta} y^{neo}$
PP	**Test-Stat**	-3.349
	P-Value	0.013
KPSS	**Test-Stat**	0.323
	P-Value	0.117
ADF	**Test-Stat**	-3.272
	P-Value	0.016

Both from a qualitative point ov view and from a quantitative analysis the variable taken into consideration seems to be stationary enough to allows us to build the trading model, the pvalues of the statistical test allows us to strongly accept the hypothesys of the stationarity (recall that for the KPSS test the null hypothesis is the stationarity, which is strongly accepted).

Once we ensured that the quantity we are working on presents the adequate characteristics it's possible to proceed with the next steps for the construction of the trigger buy and sell signals.

Once found that the series is stationary it's possible to build up the triggers for the buy and sell signal: the main assumption here, which is confirmed by the datas, Is that the δ quantity will be fluctuate always around it's middle value and so, in the case there is an excessive increase or decrease of the quantity, it will be possible to exploit this deviation from it's middle value buying or selling the δ, which will mean go long/short on the cryptos following the scheme below

$$\begin{cases} \delta_t > \xi^{open} => open\ short\ \delta\text{:} \begin{cases} sell\ p_t^{eth} \\ buy\ \beta_2 p_t^{neo} \end{cases} \\ \delta_t < \xi^{close} => close\ short\ \delta\text{:} \begin{cases} buy\ p_t^{eth} \\ sell\ \beta_2 p_t^{neo} \end{cases} \end{cases}$$

And inversely:

$$\begin{cases} \delta_t < -\xi^{open} => open\ long\ \delta\text{:} \begin{cases} buy\ \ p_t^{eth} \\ sell\ \beta_2 p_t^{neo} \end{cases} \\ \delta_t > -\xi^{close} => close\ long\ \delta\text{:} \begin{cases} sell\ p_t^{eth} \\ buy\ \beta_2 p_t^{neo} \end{cases} \end{cases}$$

Now it's easy to imagine that a crucial point in the strategy will be played by the quantification of the triggers ξ.

After some trials I chosed the ones which maximize the sharpe ratio of the train period (in-sample), I deliberately chose to do not watch the sharpe ratio of the out-of-sample period in order to put myself in a situation which is as similar as possible to the real one, where you are at the end of the train period and don't observe the out-of-sample period.

In order to pick up the opening and the closing threshold I defined a vector of possible open and a vector of possible closes as following:

$$\xi^{open} = \begin{bmatrix} 1 \\ \vdots \\ 2 \end{bmatrix} std(\delta)\ with\ step = 0.05$$

$$\xi^{close} = \begin{bmatrix} 0.05 \\ \vdots \\ 1 \end{bmatrix} std(\delta)\ with\ step = 0.05$$

After the run of this optimization algorithm the results are resumed in the following table:

	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95
1	1,0433	1,0091	0,9817	0,9287	0,8641	0,7795	0,7061	0,6469	0,6669	0,4106	0,4021	0,3489	0,3915	0,3647	0,0556	0,0341	-0,1138	-0,1638	-0,4723
1,05	1,0906	1,0581	1,0351	0,9819	0,9203	0,839	0,7692	0,7101	0,7293	0,4771	0,5226	0,4713	0,5242	0,4981	0,199	0,1535	0,0674	0,0255	-0,3452
1,1	1,0546	1,0222	0,9993	0,946	0,8848	0,8036	0,7342	0,6759	0,701	0,4502	0,4956	0,4449	0,4389	0,4118	0,1386	0,0915	0,0459	0,0033	-0,3291
1,15	1,0771	1,0455	1,0248	0,9713	0,9114	0,8318	0,7641	0,7057	0,7303	0,482	0,5289	0,479	0,4764	0,4496	0,1758	0,1295	0,0855	0,0427	-0,3178
1,2	1,4247	1,3932	1,3748	1,3256	1,2668	1,189	1,119	1,0544	1,0227	0,7555	0,8054	0,7685	0,8558	0,823	0,4882	0,466	0,4783	0,4268	-0,0035
1,25	1,3719	1,3427	1,3908	1,3421	1,2742	1,2204	1,1451	1,0772	1,0416	0,764	0,7476	0,7204	0,8016	0,8016	0,4601	0,4498	0,4618	0,4088	-0,0054
1,3	1,3924	1,3639	1,4146	1,3663	1,2998	1,2472	1,1739	1,1059	1,0712	0,7986	0,782	0,7554	0,8474	0,8474	0,4289	0,4394	0,4519	0,4257	0,02
1,35	1,4741	1,433	1,509	1,509	1,4179	1,3702	1,2888	1,2129	1,1698	0,9382	0,9211	0,885	0,7535	0,7535	0,3473	0,2986	0,3394	0,3131	-0,0352
1,4	1,4741	1,433	1,509	1,509	1,4179	1,3702	1,2888	1,2129	1,1698	0,9382	0,9211	0,885	0,7535	0,7535	0,3473	0,2986	0,3394	0,3131	0,04
1,45	1,5878	1,5477	1,5219	1,5219	1,4228	1,3699	1,3173	1,2392	1,1941	0,9894	0,9644	0,9644	0,8321	0,8321	0,5374	0,4851	0,4851	0,4581	0,1761
1,5	1,5878	1,5477	1,5219	1,5219	1,4228	1,3699	1,3173	1,2392	1,1941	0,9894	0,9644	0,9644	0,8321	0,8321	0,5374	0,4851	0,4851	0,4581	0,1761
1,55	1,4238	1,3893	1,4248	1,6247	1,5361	1,4879	1,4418	1,3688	1,3341	1,1391	1,1132	1,1132	0,9789	0,9789	0,7094	0,6603	0,6603	0,6345	0,237
1,6	1,6052	1,5685	1,5668	1,5668	1,558	1,512	1,4688	1,398	1,3694	1,1763	1,1498	1,1498	0,9747	0,9747	0,721	0,6953	0,6953	0,6698	0,3579
1,65	1,4154	1,3807	1,4114	1,4114	1,4028	1,3584	1,3164	1,3019	1,2738	1,0859	1,0859	1,0859	1,0859	1,0859	0,8609	0,836	0,836	0,8112	0,5438
1,7	1,6252	1,6252	1,6607	1,4248	1,4163	1,372	1,3301	1,3158	1,2879	1,1009	1,1009	1,1009	1,1009	1,1009	0,8779	0,853	0,853	0,8283	0,5636
1,75	1,3928	1,3428	1,3086	1,3086	1,2989	1,2476	1,2	1,1843	1,1549	0,952	0,952	0,952	0,952	0,952	0,6923	0,6632	0,6632	0,6343	0,5864
1,8	0,8324	0,7829	0,7492	0,7492	0,7492	0,7062	0,6434	0,6171	0,6196	0,5564	0,5564	0,5564	0,5564	0,5564	0,4625	0,432	0,432	0,4018	0,3511
1,85	0,8847	0,8345	0,8001	0,8001	0,8001	0,7561	0,6927	0,6672	0,672	0,6081	0,6081	0,6081	0,6081	0,6081	0,5139	0,4829	0,4829	0,452	0,4005
1,9	0,9312	0,8824	0,8487	0,8487	0,8487	0,8055	0,744	0,721	0,7282	0,666	0,666	0,666	0,666	0,666	0,5758	0,5454	0,5454	0,5152	0,4649
1,95	0,9484	0,8995	0,8655	0,8655	0,8655	0,822	0,7603	0,7376	0,7456	0,6833	0,6833	0,6833	0,6833	0,6833	0,5931	0,5625	0,5625	0,5321	0,4816

The results show that the combination of triggers which maximizes the sharpe ratio in the in-sample period Is the following:

$\boldsymbol{\xi^{open}}$	1.7 * std($\boldsymbol{\delta}$)
$\boldsymbol{\xi^{close}}$	0.15 * std(δ)
$\boldsymbol{\hat{\beta}}$	0.561

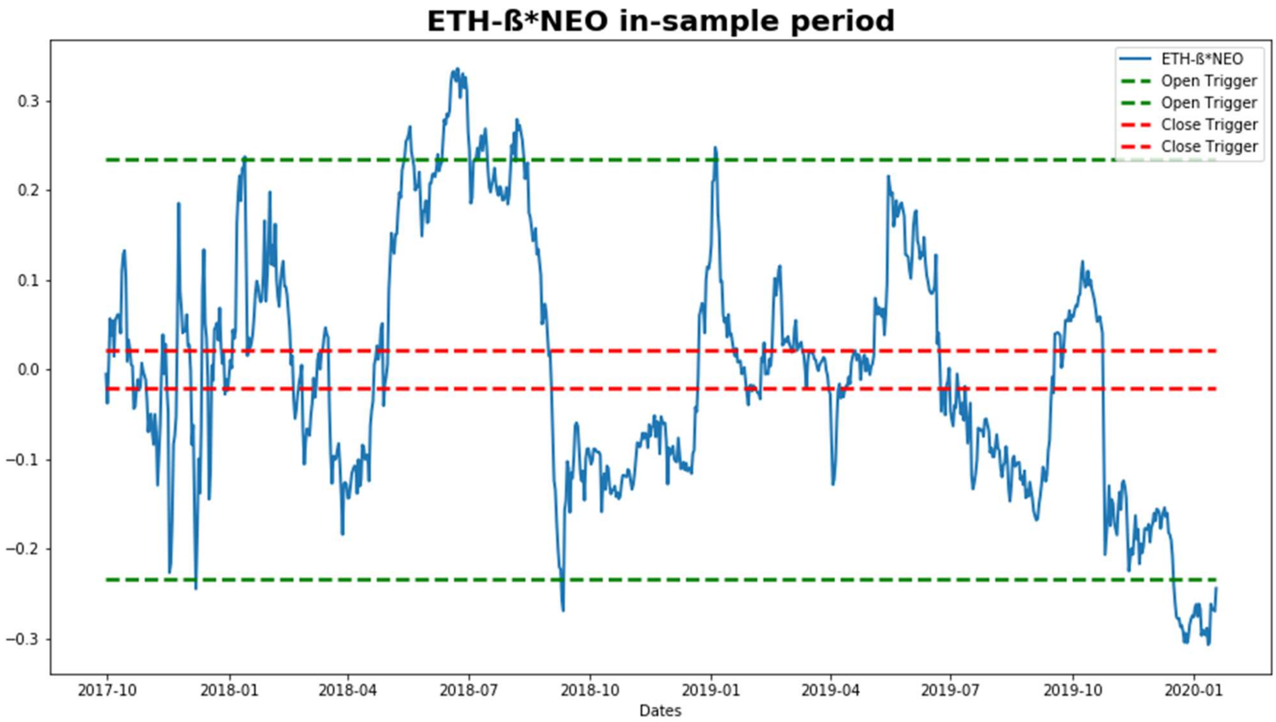
So the strategy will operate in the following way: if the value of δ_t will be greater than 1.75 times it's standard deviation on the overall in-sample period, than there will be a buy for $\hat{\beta}$ units of neo with respect to bitcoin and the selling of 1 unit of ethereum with respect to bitcoin, when the value of δ_t will be smaller than 0.15 times it's standard deviation on the overall in-sample period there will be the selling of the previously bought $\hat{\beta}$ units of neo with respect to bitcoin and the buy of the previously sold unit of ethereum with respect to bitcoin. The previous scheme becomes the following:

$$\begin{cases} \delta_t > 1.7 => open\ short\ \delta\colon \begin{cases} sell\ eth/btc \\ buy\ 0.561\ neo/btc \end{cases} \\ \delta_t < 0.15 => close\ short\ \delta\colon \begin{cases} buy\ eth/btc \\ sell\ 0.561\ neo/btc \end{cases} \end{cases}$$

And inversely:

$$\begin{cases} \delta_t < -1.7 => open\ long\ \delta\colon \begin{cases} buy\ \ eth/btc \\ sell\ 0.561\ neo/btc \end{cases} \\ \delta_t > -0.15 => close\ long\ \delta\colon \begin{cases} sell\ eth/btc \\ buy\ 0.561\ neo/btc \end{cases} \end{cases}$$

Graphically we will end up with the following configuration:



2.3.2.1. Deep Dive on the Robustness of the Beta

A possible point of weakness of the model could be in the robustness of an estimation with only 838 observation. In order to try to verify if the length of historical series represents a problem I run a simulated model with T=838 for 2500 times.

This allowed me to provide a quantification for the variance of the estimator in order to better quantify the uncertainty around the estimate.

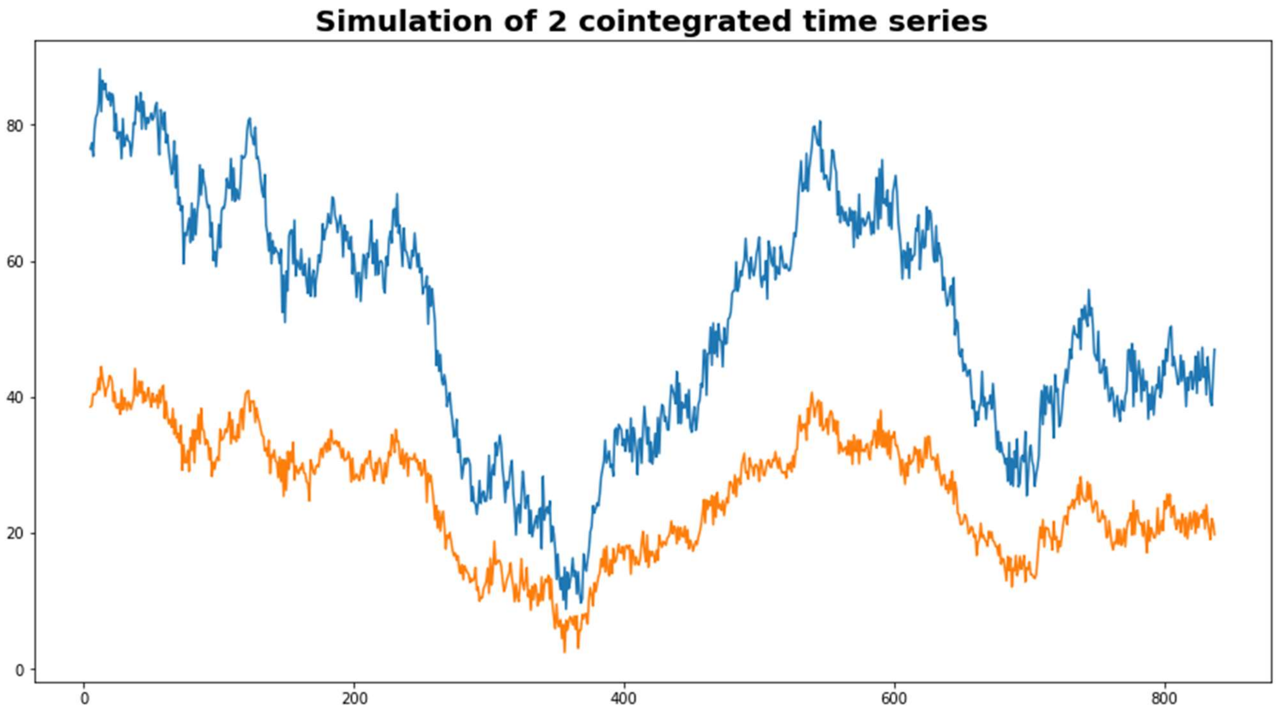
The setup of the simulated model is the following: the true model is

$$y_t^1 - y_{t-1}^1 = 0.7(y_{t-1}^1 - 2.0106\, y_{t-1}^2) + \epsilon_t^1$$

So the true value are:

$$\begin{cases} \alpha = 0.7 \\ \beta = \begin{bmatrix} 1 \\ -2.0106 \end{bmatrix} \\ \epsilon_t \sim N(0,1) \end{cases}$$

A couple of simulated series appears as the following

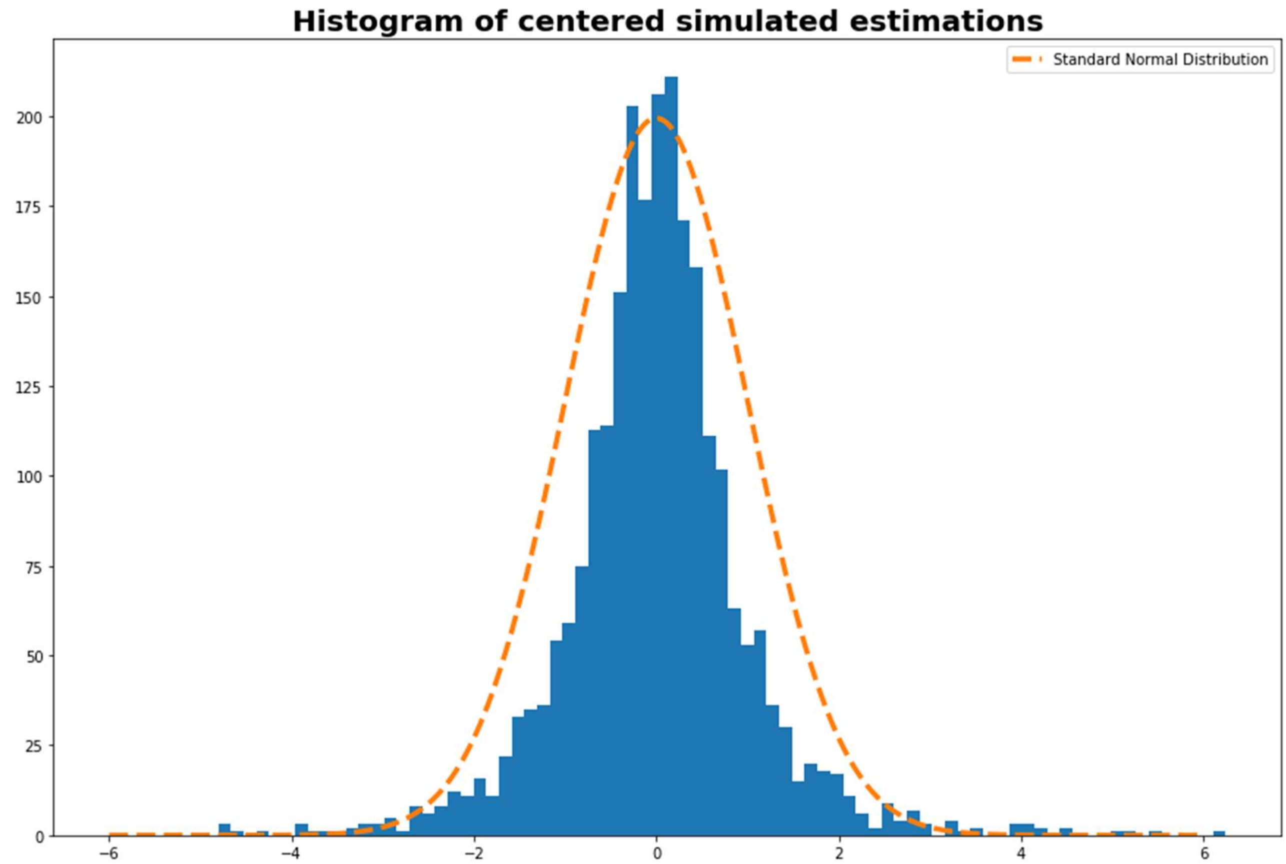


Once I got the true data I run the same estimation that I want to run on the real data in order to come with an estimate of the true parameter, which is called $\hat{\beta}$.

This process is repeated for a number of times large enough to exploit the consistency of the cointegrating estimator and to get a numerical value for the variance. In my case the simulation is run 2500 times and the results are presented in the following table.

True $\boldsymbol{\beta}$	2.01063
Mean ($\hat{\boldsymbol{\beta}}$)	2.01048
Std ($\hat{\boldsymbol{\beta}}$)	0.0073

From the result the variance doesn't seem to create too much problems.

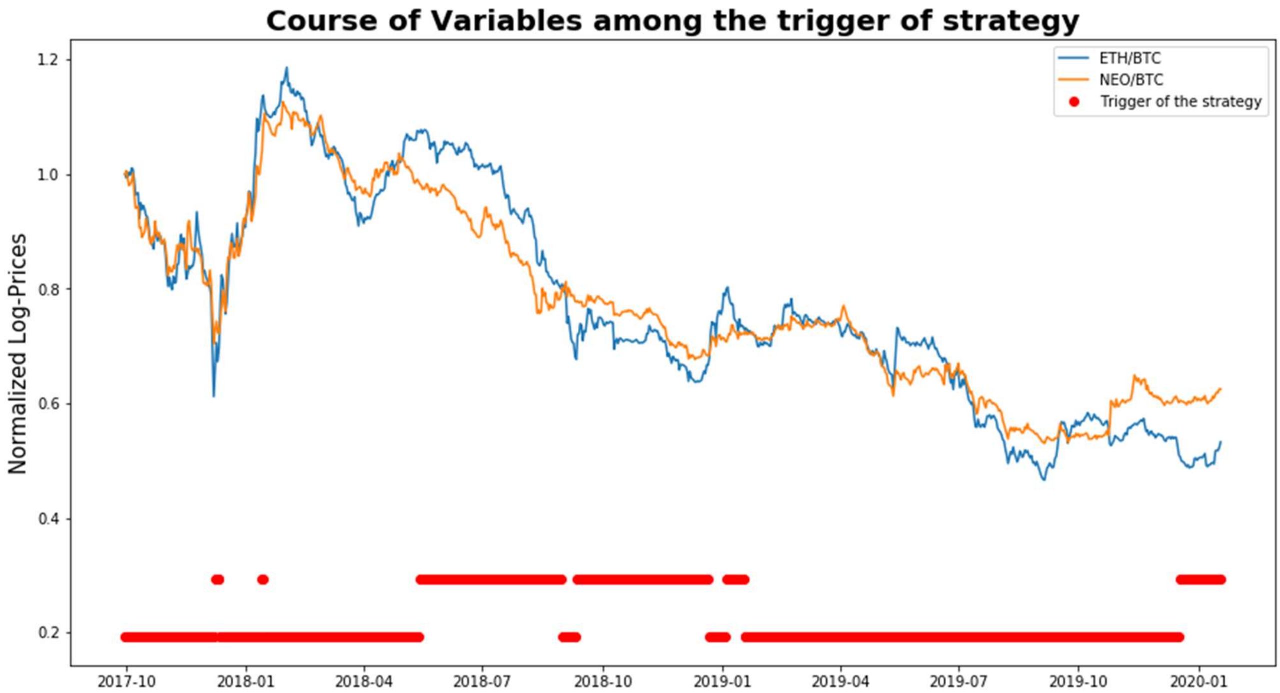


As we can see also from the compare with the standard normal curve, the distribution of the estimates it's more concentrated around his central value.

2.3.3. Back test of the strategy in the train period

Now that the thresholds are defined, and consequently the triggers for the opening and the close position the next step is to test the strategy on the in-sample period. In order to do that I generated a dummy variable which assumed 0 if the position is closed, than if there has to be a buy δ the variable assume 1, and if there has to be a sell of δ the variable will assume -1.

This procedure allows a clear visualization of the period in which the strategy is triggered, and so in which it's operates.



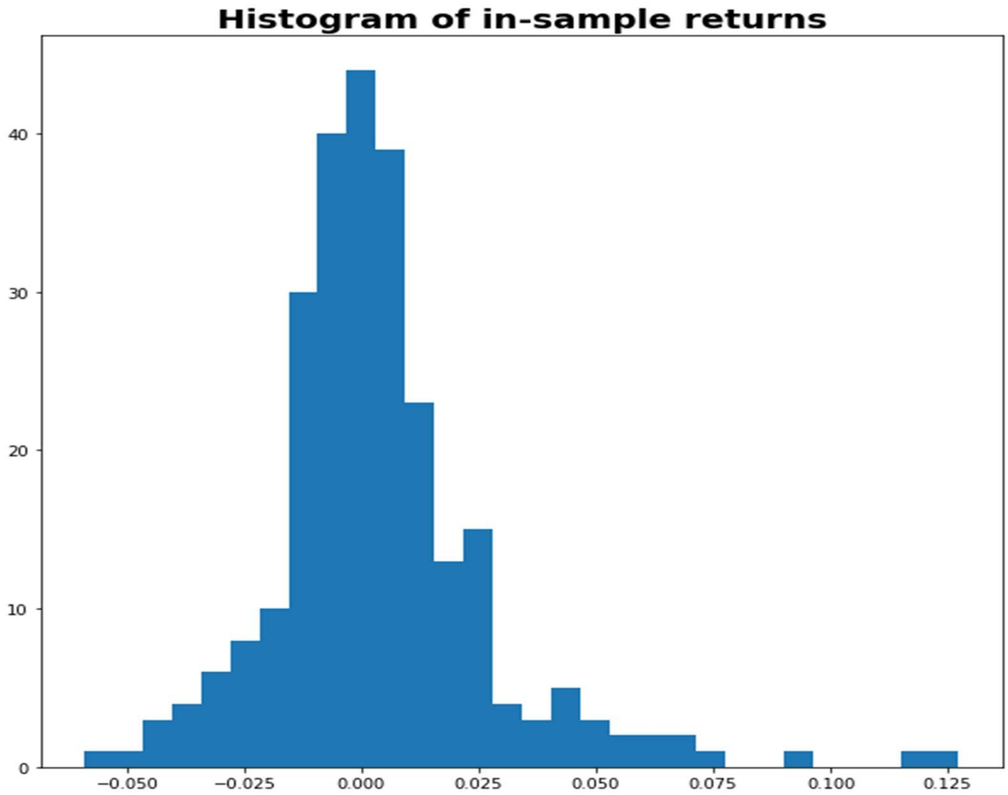
In this way it's possible to clearly visualize the moments in which the strategy is open, which are represented in the picture by the upper red lines, and the moment in which it's close, which are represented by the lower red lines.

Also graphically it's possible to see that when the discrepancy between the prices of the cryptocurrency becomes too large the trigger is open, for being closed if the difference between the 2 normalized log-prices comes back near to it's mean value.

Another issue to be solved for the tracking of a long/short strategy is how to track the evolving of a short-selling position, for this purpose I chose to simply invert the formulation for the classic return, so the tracking of the strategy will have the following setup:

$$\begin{cases} r^{long} = \dfrac{p_{t+1}}{p_t} \\ r^{short} = -\dfrac{p_{t+1}}{p_t} \end{cases}$$

There are some moments when the strategy will be close due to the fact that the spread between the two log-prices is too strict, another thing to take into account is that the triggers are build up on the log-prices but the buy and sell are operated on the standard prices without log.



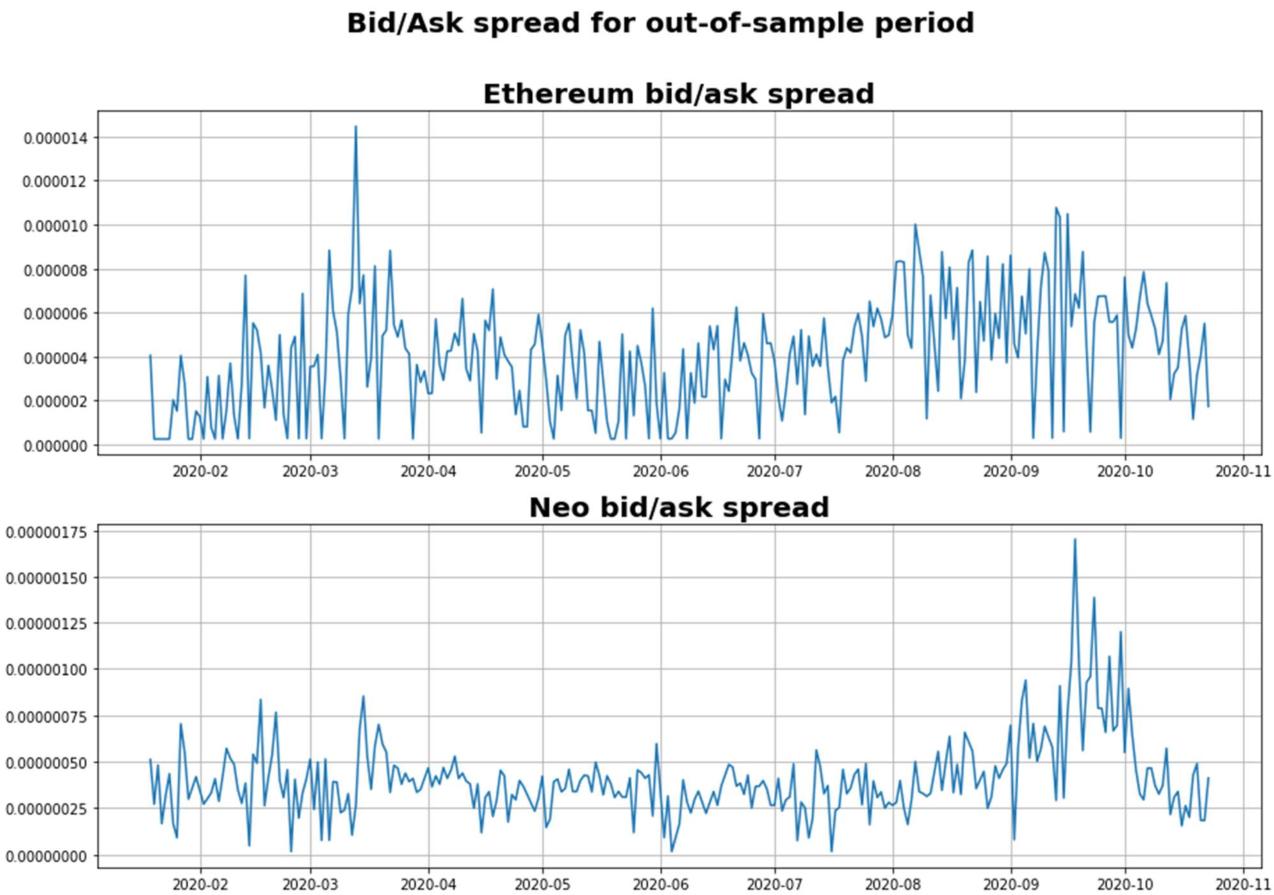
Mean	43,30%
Volatility	32,01%
Skewness	3,16
Kurtosis	25.39

As it can be denoted both from the histogram and from the table the strategy is very rewarding but also very risky both in terms of absolute volatility and in terms of tail risk, as can be denoted from the kurtosis.

3. Analysis and performances in the out-of-sample period

In this chapter there will be presented the performances for the out-of-sample period of the strategy, recall that there will be some differences for the out-of-sample strategy with respect to the in-sample data:

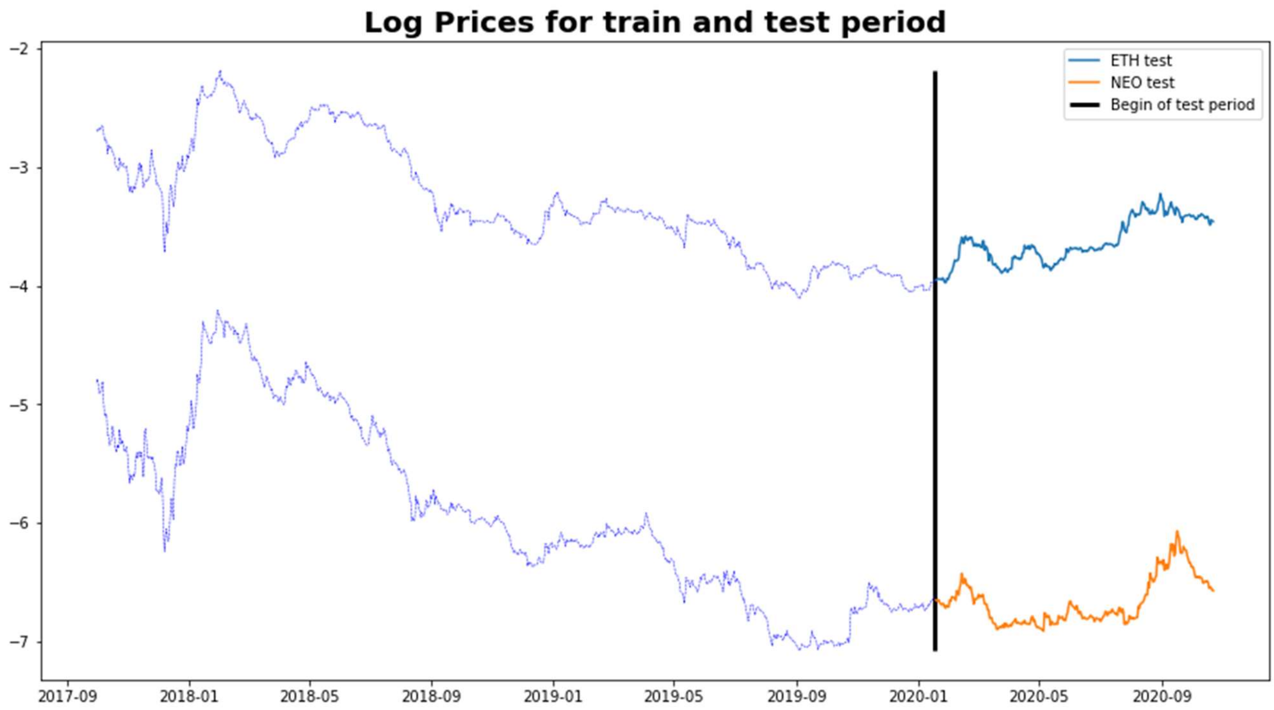
First of all here, in order to have more reliable results, the liquidity is taken into consideration, so every time there will be a buy the ask price will be used and vice versa every time there will be a sell the bid price will be used. The bid-ask prices are downloaded from quandl, which takes the data from binance.



The plot shows clearly that both the Ethereum and Neo spread are very tiny (they are expressed as percentage of prices), this is an indication that we are dealing with very liquid securities.

The data used into this sample are not used into the estimation of the parameter, they are "unknown" to the model.

The data go from 19-01-2020 to 23-10-2020 and cover approximately 1/4 of the overall length size



Even if it was possible to download data from other 2 months of 2020 I deliberately decided to stop at the end of October 2020 since in may the bitcoin had it's 3rd halving and this is a kind of event which could undermine the robustness of the model since it implies a disruption in the fundamentals of bitcoin, and consequently, of all cryptocurrencies market.

The halving of the bitcoin is, at it's core, a decrease of the supply of the currency: for it's nature the bitcoin is created whenever a block of the blockchain is created, approximately every 10 minutes, and an amount of currency is given as reward to the address that created the block.

In order to keep scarcity the creator of the currency decided to halve the amount of rewards for the creation of new blocks every 210.000 blocks, this event is called halving.

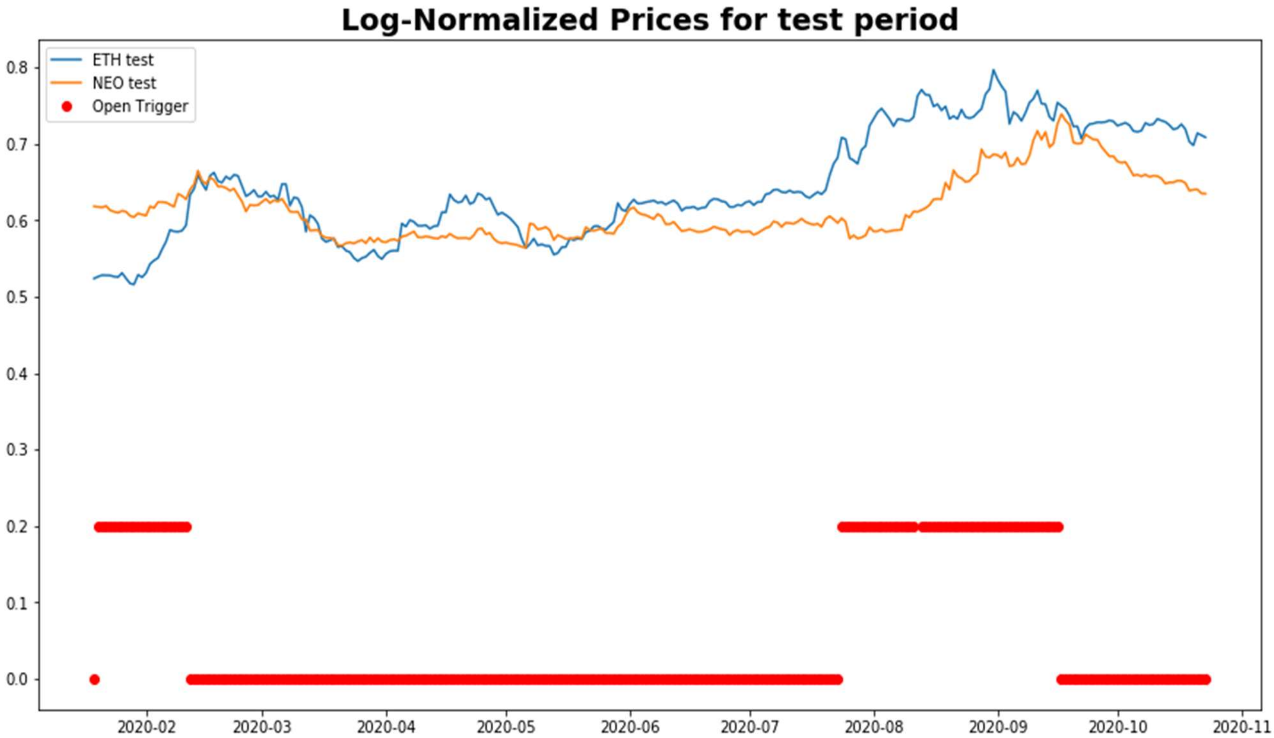
The halving is a functionality of the bitcoin protocol which represents the core of it's economic model since it influences the emission rate and the quantity of circulating currency, decreasing the supplied quantity, keeping into consideration that a new block is created approximately every 10 minutes we know that 210.000 blocks are created about every 4 years, which is the approximate distance between one halving and one other.

The last halving happened in July 2016 and, starting from the following year, there was a decrease of the supply and a consequent increase of the price of the currency, which ended up in the bubble at the end of 2017.

Given the previous facts I deliberately decided to stop the collection of the data to October 2020.

3.1. Performances on out-of-sample period

In the out-of-sample period the strategy has been triggered 2 times, one of them which derives from the opening which was present at the end of the in-sample period.



The graph above reports the log prices of the currencies normalized to the starting date of the train period.

Also from a qualitative point of view it's possible to see that there is a huge discrepancy between the values only in 2 occasions, the thick red line below the graph indicates the period when the strategy is trading: when it's high it means that the trading is open and when it's low it represent the closing periods of the strategy.

Another way to visualize the matter is to look at the $\delta_t = \frac{eth}{btc} - \hat{\beta}\frac{neo}{btc}$ together with the open and closing triggers, recall that the open and closing triggers are the ones defined on the train period.

As we can denote the δ quantity presents only one spike in the out-of-sample observation window.

Investing an hypothetical euro in this strategy, taking into account the bid/ask spread, we would end up having the following trajectory:

It's possible to denote a huge drawdown in the second open period, indicating that, as we saw from descriptive statistics in the previous chapter, this strategy is very rewarding but also very risky.

The investment in the out-of-sample period would have produced an overall return of 26,95%, which, in my humble opinion, is fairly rewarding for the amount of risk carried out in the period.

The representative statistics for the daily excess annualized returns of the out-of-sample period are the following:

Mean	42,97%
Std	0,3152
Skew	-0,76
Kur	8,44

As we saw in the in-sample period the overall performance is very rewarding but also very risky both for the volatility and for the tail risk.

3.2. Choice of the Benchmark and comparison

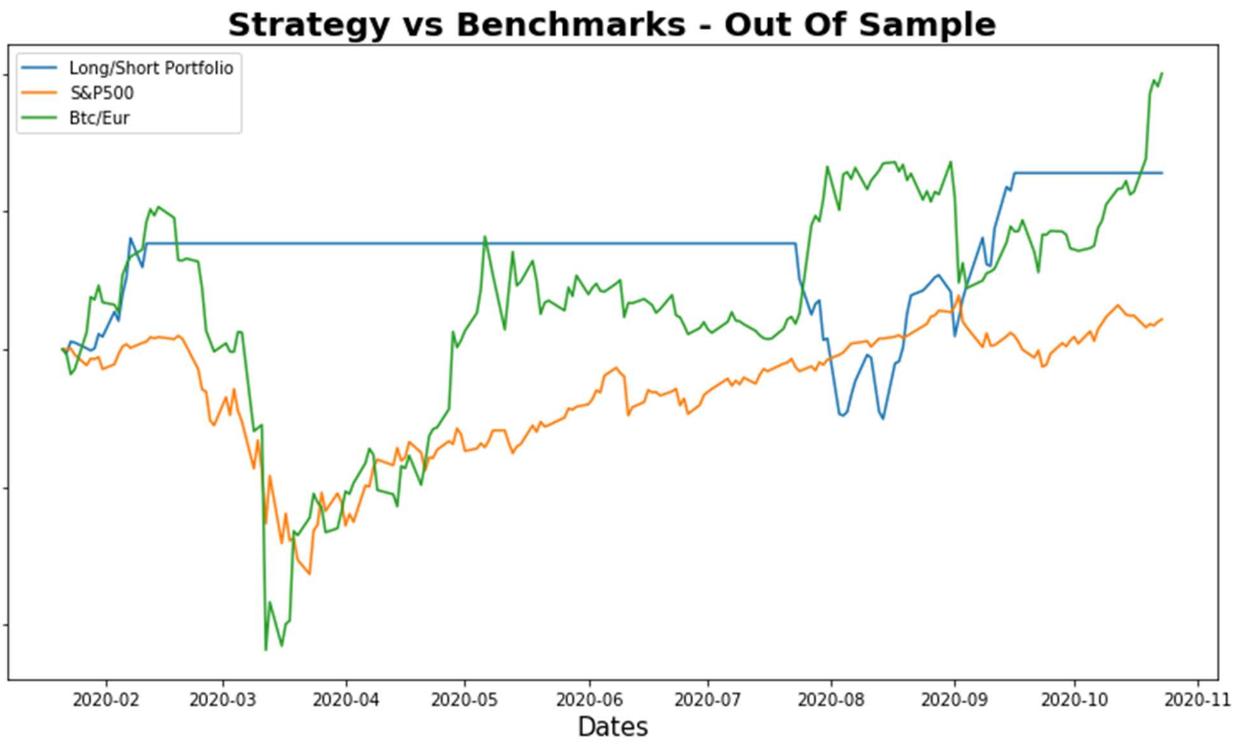
A legit question to be asked here is which benchmark is more appropriate when evaluating the performances of this strategy.

In my opinion the choice is between 2 candidates, one is the S&P 500, which is the usual benchmark on which evaluate investment strategies, and another is the price of the Bitcoin with respect to the euro.

This alternative could be useful since we are going to compare an active strategy in the market of cryptocurrencies with a more traditional passive one, which merely consists in a simple buy-and-hold strategy in the same market.

Instead the reasons for the comparisons with the S&P 500 it's to evaluate the strategy on an amplified point of view, and so to compare it with a passive strategy which merely replicates the evolution of the market, which is the canonical approach used in the evaluation of investment strategies.

I choose to compare the portfolio with both the price of the Bitcoin with respect to euro and the S&P 500 index in order to have a more complete point of view.



It's possible to see that the strategy presents an evolution which is totally detached from the ones of both it's comparisons, which could be a useful characteristic if we want to insert it in a portfolio of asset in order to increase the diversification.

The overall returns is in favour of the buy-and-hold Bitcoin strategy but it's possible to see, also from a visual check to the graph, which the buy-and-hold Bitcoin strategy is more volatile than the long/short portfolio.

	S&P500	Bitcoin	Portfolio
Mean	14%	182,57%	43,34%
Std	0,3857	0,8559	0,3152
Skewness	-0,4702	-2,4	-0,76
Kurtosis	6,159	26,99	8,44

As we can see the S&P 500 has been the lowest annualized yielding but even the lowest risky option, both in terms of volatility and in terms of tail risk, than we can see a particular thing, the long/short portfolio, even if is more risky in term of tail's thickness, as told by the kurtosis, it's less volatile even than the S&P 500 bringing nearly the double of it's annual return.

The bitcoin has been, in the period taken into consideration, the higher yielding asset, with a daily return doubled with respect to the long/short portfolio, but also the riskier, with the quadruple of volatility with respect to the long/short and the double of kurtosis.

Even if the absolute daily return seem nice it's good to have a look also to the risk-adjusted performances, first of all it's good to have a look to the alpha and beta of the strategy, since I choose to pick up both the Bitcoin price in euro and the S&P500 as benchmarks I took the values of alpha and betas regressed versus both the variables.

	S&P 500	Bitcoin / Euro
Alpha	0,495	0,4891
Beta	-0,0104	0.0038

The beta for both the Bitcoin and the S&P500 is very close to zero, meaning that the evolution of the market and of the principal cryptocurrency doesn't affect the evolution of the long/short portfolio.

Another way to capture the fact that the evolution of the strategy isn't affected by the trends which are present in the market and in the Bitcoin is to watch to the correlation matrix between the 3 variables:

	L/S Portfolio	Bitcoin	S&P 500
L/S Portfolio	1	-0,083912	-0,097327
Bitcoin		1	0,860979
S&P 500			1

The correlation between the long/short portfolio is negative, almost zero, with both it's benchmarks, a particular thing to denote is the high correlation between the S&P500 and the price of Bitcoin with respect to euro, this could be due to the large drawdown at the begin of the year, due to explosion of the covid-19 pandemic, which affected commonly the American stock market and the price of the Bitcoin.

Now that it's established that the comparison has to happen between S&P500 and Bitcoin price with respect to euro let's have a look to the risk-adjusted performances of the 3 elements, I will go trough the following statistics for each of the elements:

- $Sharpe\ Ratio$: $\frac{ret_{port}-rf}{\sigma(ret_{port})}$; it allows us to quantify the amount of excess return we are gaining for an extra unit of volatility.
- $Sortino\ Ratio$: $\frac{ret_{port}-r}{DSR}$; where DSR is defined as

$$DSR = \sqrt{\int_{-\infty}^{T} (T-x)^2 f(x) dx}$$

  And it wants to quantify the "bad" part of volatility, the downside one, with respect to the classical standard deviation the Downside Risk measures the negative deviations with respect to the minimum return acceptable, usually defined as the risk free rate. The Sortino ratio so wants to capture the amount of extra return we are gaining for the exposure to an extra amount of downside risk.
- *Maximum Drawdown:* it is simply the max amount of negative return that the historical series taken into consideration had for a single point in time. It wants to capture how badly the investment had performed in his worst period.

	L/S Portfolio	Bitcoin	S&P 500
Sharpe Ratio	1,1692	1,0137	0,3389
Sortino Ratio	1,6303	1,3238	0,4342
Max Drawdown	-22,61%	-53,23%	-33,925%

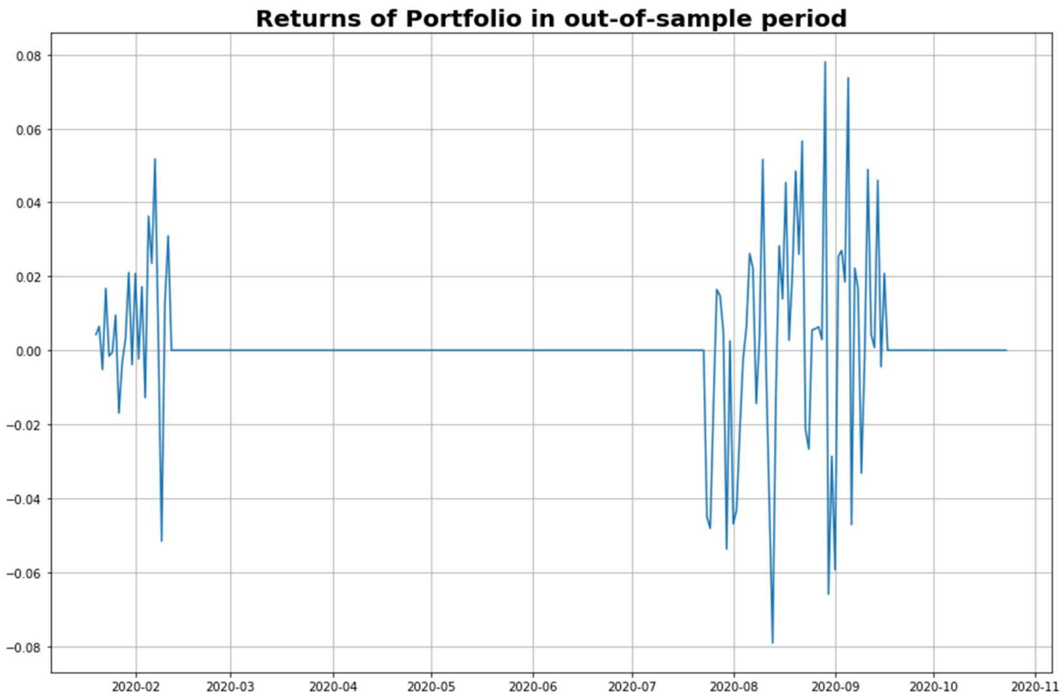
What these results tell us is that the portfolio presents an higher amount of reward for every source of risk we are exposing ourselves compared both with bitcoin and S&P 500, moreover, in the period taken into consideration both the assets we are comparing it to presented an higher drawdown.

Once assessed the risk-adjusted performances of the portfolio the next step could be to try to get a distribution for the returns of this strategy.

Since we have only approximately 280 observation for the returns it could be a little hard to try to assign a distribution with these few observation.

One way to bypass this issue could be, as like performed for the assessment of the consistency of the beta in the previous chapter, to perform a bootstrap analysis in order to come up with a distribution of the returns: the workflow I want to follow is splitted in the following steps:

i. I take the historical series of the returns of the portfolio, comprehensive of the period in which the strategy is closed, which appears as following

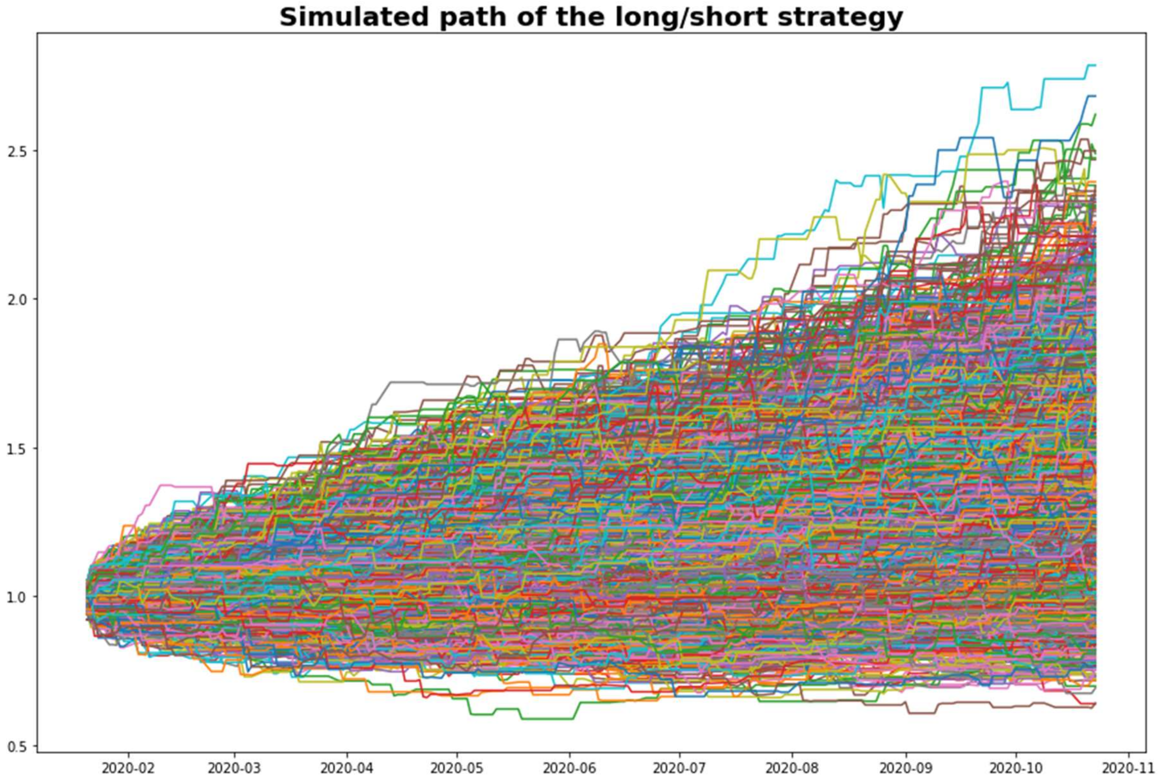


ii. In the second step I perform a bootstrap analysis randomly taking observation, with re-entry, in order to create a simulated historical series for the out-of-sample period.

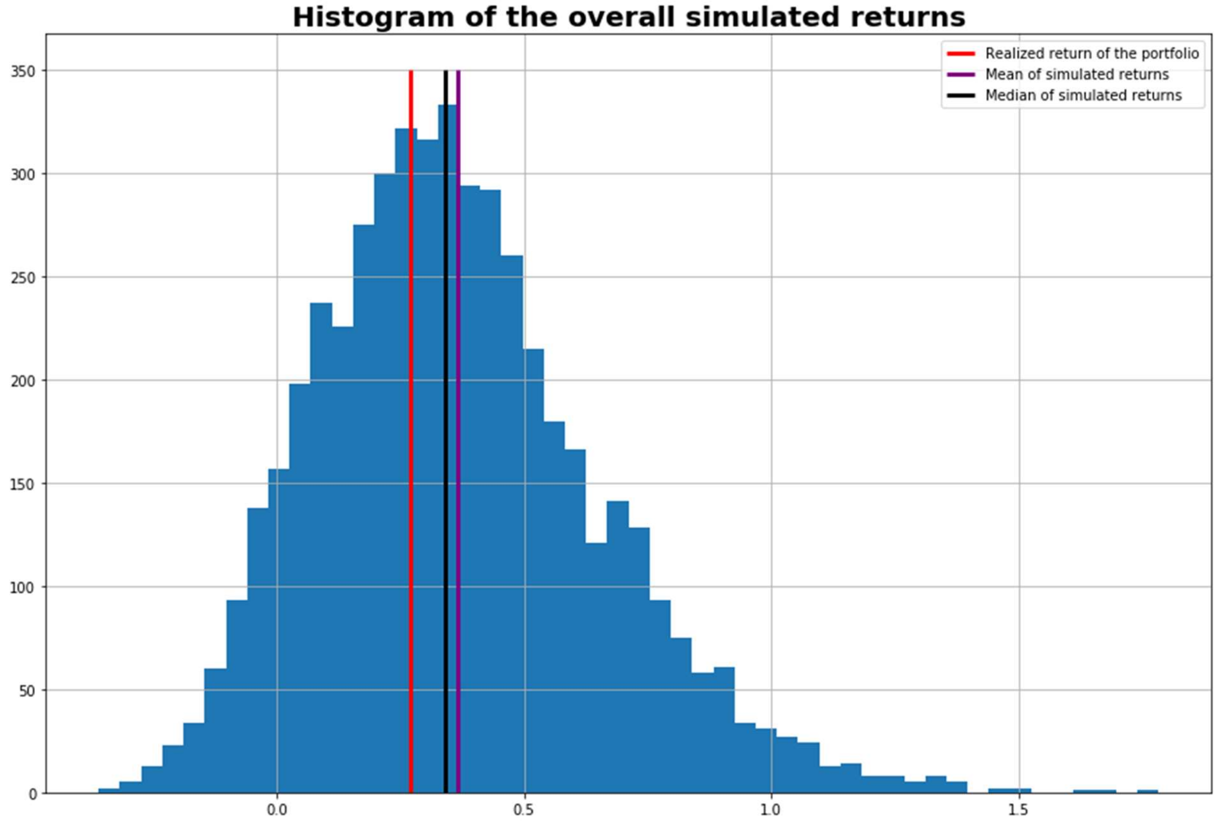
iii. I register the performance of the portfolio at the end of simulated out-of-sample period.

iv. I repeat the previous 2 step for a number of times large enough to get the shape of the distribution of overall out-of-sample returns.

Set n, the number of simulations to be performed, equal to 5000, the simulation appears as following:



While the histogram of the distributions for the overall returns at the end of the out-of-sample period appears as



As we can see the realized portfolio overall return is collocated below both the median and the mean of the simulated returns. The descriptive statistics of the distribution of simulated annualized returns are reported below, I decided to report both the mean and the median because the distribution Is not symmetric and so the mean could not correspond with the more frequent value.

Realized Return	36,79%
Median	46,63%
Mean	50,30%
Standard Deviation	5,4487
Skewness	0,5896
Kurtosis	0,5419

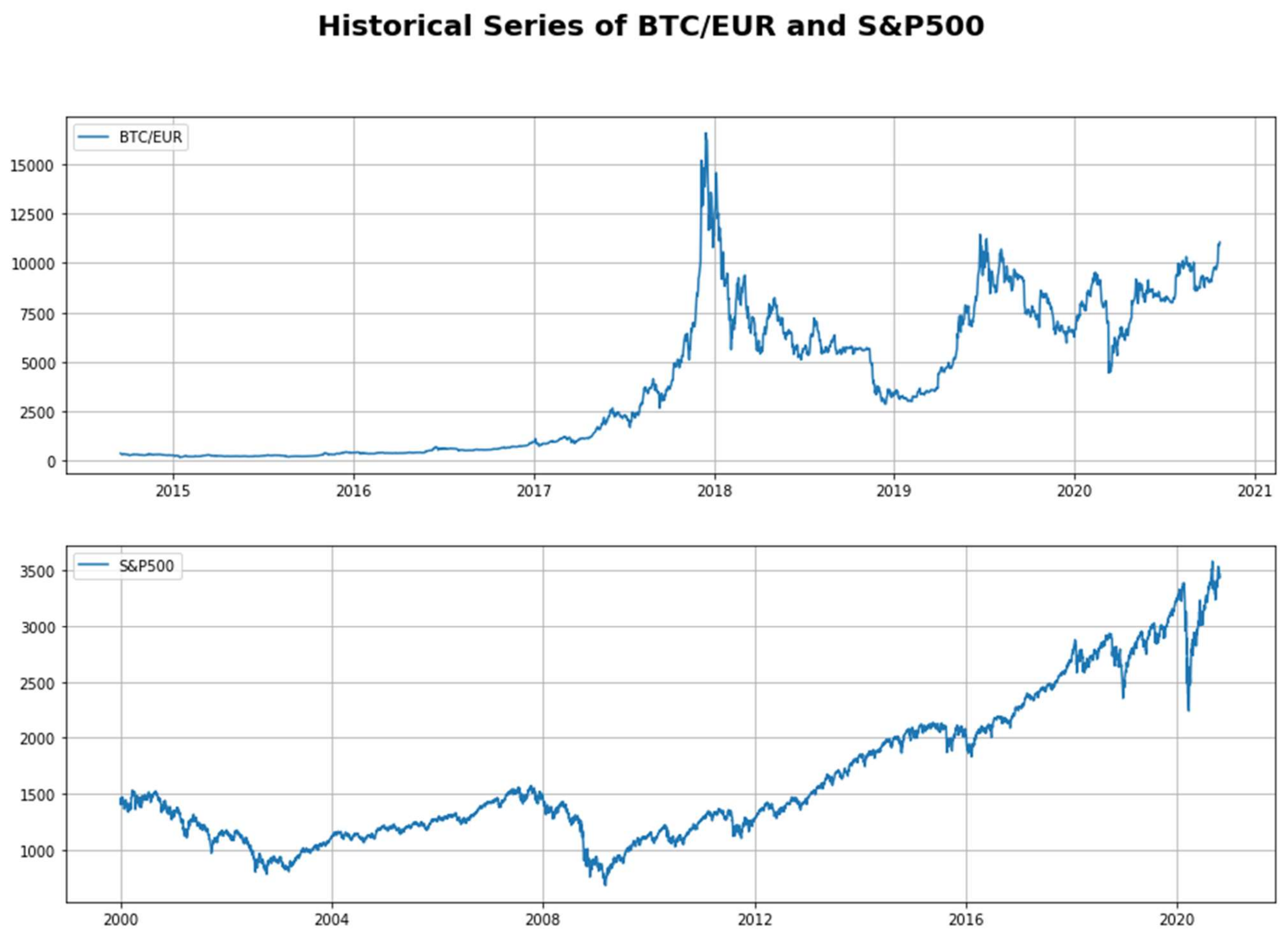
3.2.1. How did the covid-19 affected the analysis?

A consideration which is good to make is that the overall return of the S&P 500 and of the Bitcoin had been affected by the explosion of the covid-19 pandemic, which provoked the big downturn at the beginning of the 2020 year.

A possible stress test to better asses the performances of the long/short strategy could be to try to get rid of the covid-19 impact on the performances of the S&P500 and of the Bitcoin.

In order to try to compare the performances of the strategy to a more "normal" times I exploited the same bootstrap approach used for the estimation of the distribution of portfolio out-of-sample returns.

The historical series of S&P500 I starts from the 01/01/2000 and the historical series of the Bitcoin/Euro starts from the first available date of 09/16/2014.



From both of historical series I took the returns and bootstrapped n=5000 historical simulated series for the dates of out-of-sample period, which recall goes from 19/01/2020 to 23/10/2020, following this approach allowed me to took returns from a more true-to-real series.

Once simulated the series I took the simulated overall annualized returns for the out-of-sample period and compared them with the observed evolution of the long/short strategy.

The descriptive statistics for both the bootstrap simulations are reported in the following table:

	BTC/EUR Sims	S&P500 Sims
Mean	85,58%	9,82%
Median	73,89%	6,815%
Standard Dev	1,3499	0,2296
Skewness	4,533	0,6201
Kurtosis	71,81	0,6384

I choose to report both mean and median since we are dealing with non symmetric distributions and the mean could not always correspond to the more frequent value.

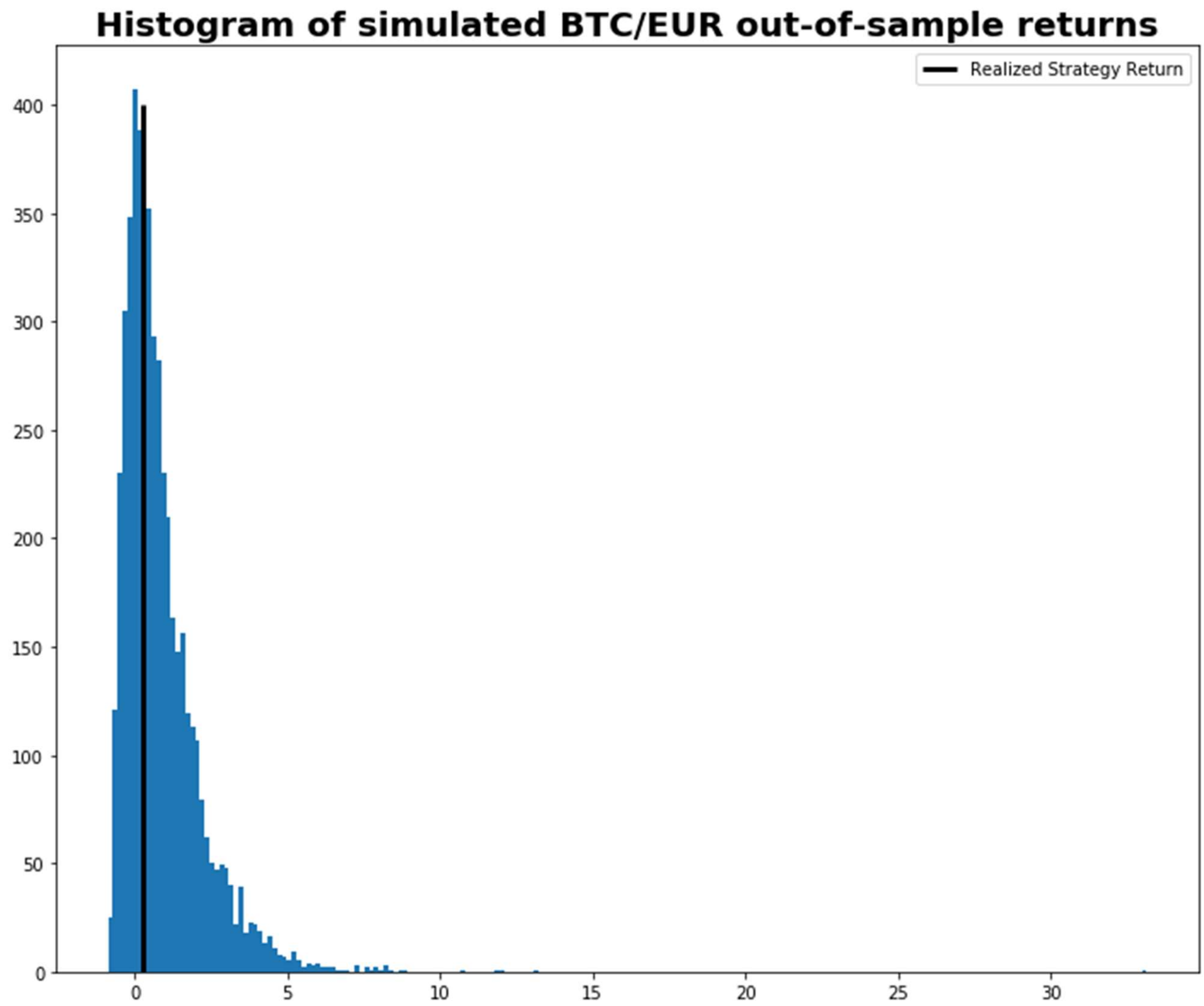
Also in the simulation the Bitcoin/Euro is the asset which gave more returns but It carries a huge amount of risk in terms of volatility and tail risk.

Comparing the realized annualized returns with the simulated annualized ones allows us to better quantify the impact of the covid-19 on the markets.

	Realized	Simulated	Delta
Bitcoin	55,63%	85,58%	29,95%
S&P 500	5,72%	9,82%	4,1%

Even if in the Bitcoin case it seems that the covid-19 brought a biggest loss we have to consider the uncertainty around the simulated value which for the BTC is very high compared to the one on the market.

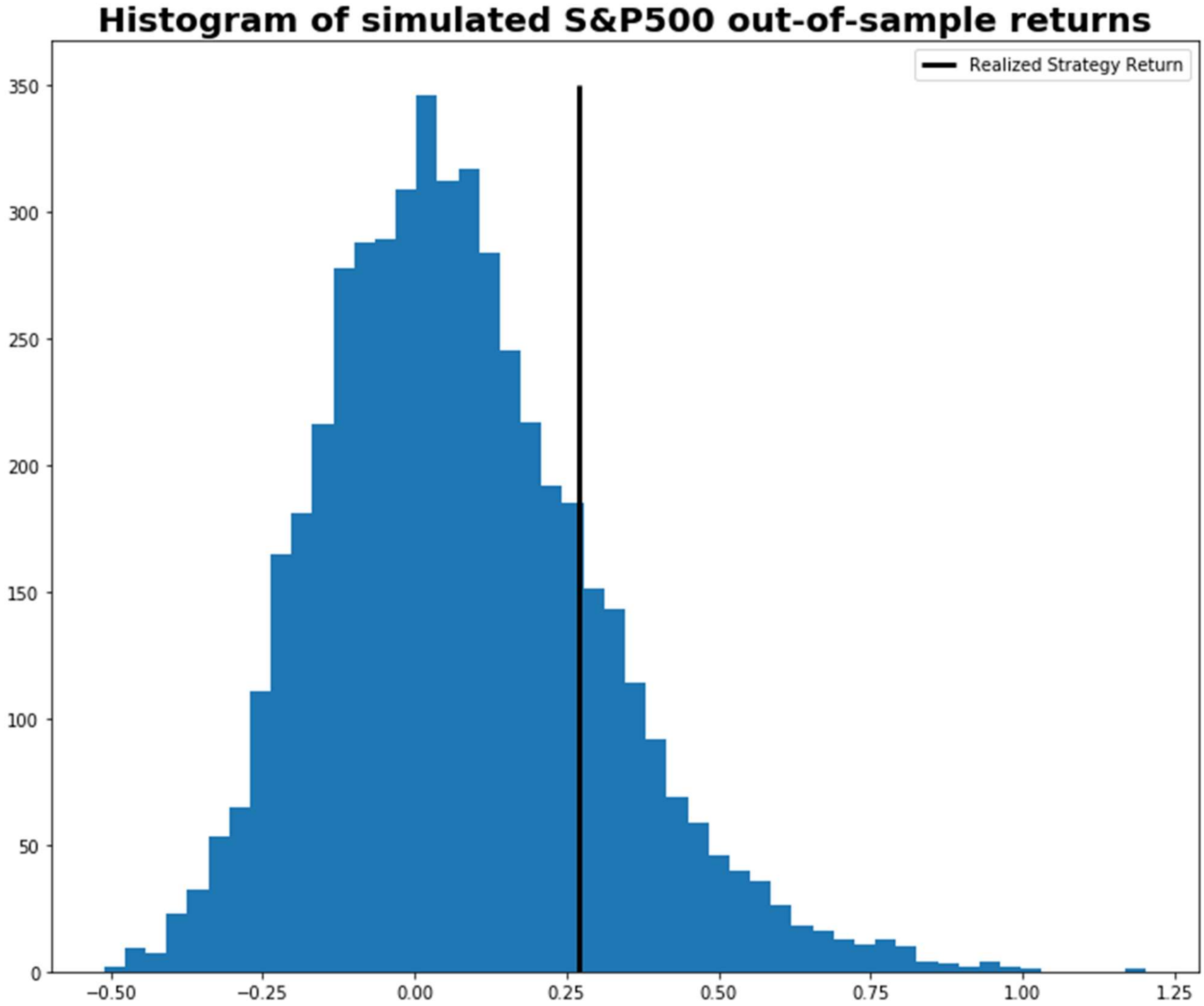
The next step is to compare the observed returns of the trading strategy with the simulated one in order to get a more reliable result, to do that I start plotting the histograms of the simulations and comparing them with the realized returns:



The overall return of the long/short trading strategy, compared with the simulated Bitcoins trajectories seems to fall in the middle of it's more frequent.

From the histogram of simulated trajectories it's possible to also observe the consideration about the riskiness of the asset, the histogram shows a really skewed distribution with the left tail, the loosing one, which is very thick, this factor is captured by the huge kurtosis showed in the descriptive statistics.

Plotting the same elements with the simulation of S&P500 allows us to observe a radically different story:



As we can see the distribution is totally different compared to the one of the bitcoin, less skewed and with less thick tails, here the returns obtained by the strategy seems to clearly overperform both the mean and the median of the simulated series.

The final step could be to compare the performances of the simulated series with the one realized by the portfolio also in terms of standard deviation of simulated series and risk adjusted performance (Sharpe ratio).

The data about comparison are summarized below, recall that here each statistic is referred to the simulated historical series, so for example the standard deviation will be the mean of the standard deviations of the simulated series:

	BTC/EUR Sims	S&P500 Sims	Long/Short Portfolio
Mean	85.58%	9,88%	36,79%
Mean(std)	0,7384	0,0841	0,3525
Mean(Sharpe Ratio)	0,94	1,043	1,16

Also for the simulated results the long/short portfolio seems to bring a nicer amount of returns for each unit of additional exposure to risk.

CONCLUSION

As we saw this kind of long/short strategy could bring a nice amount of return and also a decent amount of risk-adjusted return, anyway it has to be considered that it' s a very risky strategy as we saw in the work.

A possible point of improvement for the strategy could be to insert time varying parameters of to build a portfolio of long/short crypto.

Another issue to take into account is the fact that the statistical relationship could have some weaknesses during periods of high market turmoil like the bubble of end 2017, in these kind of period the best possible strategy is proved to be the classical buy-and-hold one, of course one has to be able to sell before the pop of the bubble.

Bibliography

1. David A. Dickey & Wayne A. Fuller (1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American statistical Association, 74:366a, 427-431, DOI: 0.1080/01621459.1979.10482531
2. Johansen S. (2009) Cointegration: Overview and Development. In: Mikosch T., Kreiß JP., Davis R., Andersen T. (eds) Handbook of Financial Time Series. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-71297-8_29
3. Engle, Robert F., and C. W. J. Granger. "Co-Integration and Error Correction: Representation, Estimation, and Testing." Econometrica, vol. 55, no. 2, 1987, pp. 251–276. JSTOR, www.jstor.org/stable/1913236. Accessed 20 Nov. 2020.
4. Stock, James H. "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors." Econometrica, vol. 55, no. 5, 1987, pp. 1035–1056. JSTOR, www.jstor.org/stable/1911260. Accessed 20 Nov. 2020.
5. Park, Joon Y. "Canonical Cointegrating Regressions." Econometrica, vol. 60, no. 1, 1992, pp. 119–143. JSTOR, www.jstor.org/stable/2951679. Accessed 20 Nov. 2020.
6. Bilgili, Faik: "Stationarity and cointegration tests: Comparison of Engle - Granger and Johansen methodologies." Erciyes University, Faculty of Economics and Administrative Sciences
7. Phillips, Peter C B & Ouliaris, S, 1990. "Asymptotic Properties of Residual Based Tests for Cointegration," Econometrica, Econometric Society, vol. 58(1), pages 165-193, January.
8. Stock, J.H. and Watson, M.W. (1993) A Simple Estimator of Cointegrating Vectors in Higher Order Integrated System. Economometrica, 61, 783-820. http://dx.doi.org/10.2307/2951763
9. Evan Gatev, William N. Goetzmann, K. Geert Rouwenhorst, Pairs Trading: Performance of a Relative-Value Arbitrage Rule, The Review of Financial Studies, Volume 19, Issue 3, Fall 2006, Pages 797–827, https://doi.org/10.1093/rfs/hhj020
10. Fischer, T.G.; Krauss, C.; Deinert, A. Statistical Arbitrage in Cryptocurrency Markets. *J. Risk Financial Manag.* **2019**, *12*, 31.