Non-negative LASSO

A Portfolio replication application

Giovanni Misseri, Andrea Della Vecchia

Business Economics and Financial Data

2019, February

Contents

1 Introduction

Portfolio management has always been one of the most used way investors start approaching to financial investments. An important task in finance is indeed optimizing the revenue with respect to the risk of investing in a portfolio.

Markowitz in the '50, stated that single minded pursuit of high returns results in a poor strategy and rational investors need to balance their desire for high returns with low risk. So investing in a portfolio seems reasonable, indeed it diversify our investment and if we are able to spot profitable companies on which invest, we are guaranteed good revenue taking relatively low risk.

There are mainly two strategies in portfolio management, active and passive managements. Active management tries to exploit the fluctuations of the market to gain as much as possible. On the other hand, passive management, tries to mimic the performance of an index. As one can easily notice, with active approach it's possible to loose great amount of money even investing on, on average, growing companies. On the other hand if we are able to well mimic a, on average, growing index, with passive approach we are guaranteed positive revenue.

Empirically is often the case passive approach outperforms active approach, also due to the transaction costs. In this work will try to propose two different solution to pursue respectively the passive and the active approach.

2 Shrinkage Methods

If we have a feature or some measurements of an interesting phenomenon, is often the case that, given some related variables, we try to explain the phenomenon as a function of the related variables. The use of linear model, in case of linear relation between phenomenon and variables, is a good choice for two reasons: it gives good prediction and on the other hand, given the linear relation, it's highly interpretable.

In general the more correlated variables we have, the "better" the model will be; but sometimes the presence of too much independent variables conditions the prediction error and also the presence of an high number of predictors limits the interpretability of the model. For these reasons one could decide to add a regularization term to its minimum least square problem.

$$\min_{\beta_0,\beta} \frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|^q$$

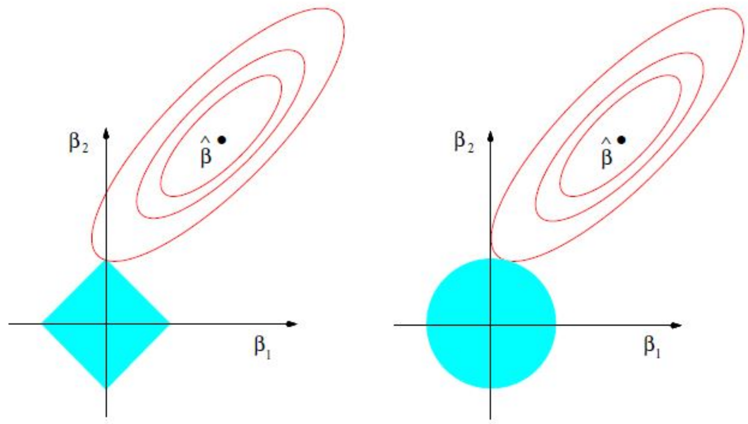
Where $|\beta|^p$ is the p-norm of β.

If in the regularized minimization problem we choose norm-two we obtain the ridge regression, if we choose the norm-one we obtain LASSO regression.

It's easy to prove that shrinkage methods also solve an other type of problem related to ordinary least square. With OLS indeed the solution of the minimization problem is $\beta_{ols} = (X^TX)^{-1}X^Ty$, decomposing X through SVD, $X = V\Sigma U^T$, we get $\beta_{ols} = U\Sigma^{-1}V^Ty$. From this we can easily see that if Σ is not invertible β_{ols} could not be found.

Shrinkage methods deal with this problem. Ridge regression ends up giving $\beta_{ridge} = U(\Sigma^2 + \lambda I)^{-1}\Sigma V^Ty$. Lasso regression's beta cannot be written in an explicit form, but the non-invertibility problem is solved due to the fact that, fixed λ, at the optimum, some beta will be equal to zero. Intuitively this is due to the feasible solutions space shape; indeed LASSO regression can be written as the following minimization problem.

$$\min_{\beta} \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \quad \textit{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq t \tag{1}$$



2.1 LASSO and Portfolio Replication

A replicating portfolio for a given asset or series of cash flows is a portfolio of assets with the same properties. So the principal problem of building a good replicating portfolio for a certain financial aggregate index is finding the right components of the index, the one able to represent the index itself.

With the problem posed like that, the link between lasso regression and portfolio replication it's evident. Indeed one interesting properties of lasso regression is the variable selection consistency, so using lasso regression we are guaranteed, under certain condition, that it will select an optimal subset of variables in order to explain our dependent variable. It's clear now that lasso could help on the selection of the subset of time series that will be part of the portfolio; so we will select a subset of an index' components able to explain the index itself.

Here comes one problem, how to use lasso, a cross-sectional regression method, on a dynamic dataset? We will use a moving window approach, so we will assume that the data inside a window of fixed length are time invariant, and we compute lasso on that.

Another problem is the interpretation of the resulting model. Ones we find the solution to the lasso minimization problem, we have the beta related to all the variables, in our case the components of the index. Some beta will be equal to zero, so they will be excluded by our portfolio, for all the other beta a nice interpretation is available.

If we carefully think to lasso as it has been defined in (1), we approximate an index, the used one will be $Sp500$, as a function of some of it's components, conditioning to the fact that the sum of the associated beta is less than a given t. Practically we will work on the daily log-returns, and so our response and regressors will be all daily log-return; said that it's easy to see that t can be seen as the initial budget and each beta is the volume of the "suggested" investment in a certain asset. Later on we will exploit this fact to run simulations using the different proposed methods.

For an active approach everything works fine, a positive beta will be an investment and a negative beta will be a short sell. On the other hand, for passive approach, it could be forbidden or in general doesn't make much sense to use short sell. For this reason to our estimates we need to add the constrain of non-negativity.

Considering that also elastic net regression has the consistency variable selection, we will try to select the assets both with lasso and elastic net regression. Some comparison will be done on that.

2.2 Dataset Analysis

The major problem dealing with this kind of financial data is the high variability of the index from one day to another. In our case, this means that the behaviour of $Sp500$ at time $t + 1$ is very difficult to predict at time t.

We start our analysis with the simpler case in which we want to replicate $Sp500$ at time t knowing the values of the 500 assets (our x variables) composing the index. Since there will be some correlation between the assets we can ask ourselves how many of them are really needed. Keeping the number of variables low is not only useful for having a less complex model but also, if we want really to invest in the assets and reproduce the index, to avoid the payment of a large number of fees (one for each purchase). On the other hand, the more we simplify the model ignoring some variables, the more our replication of the index will result to be inaccurate. With Lasso we can control the number of

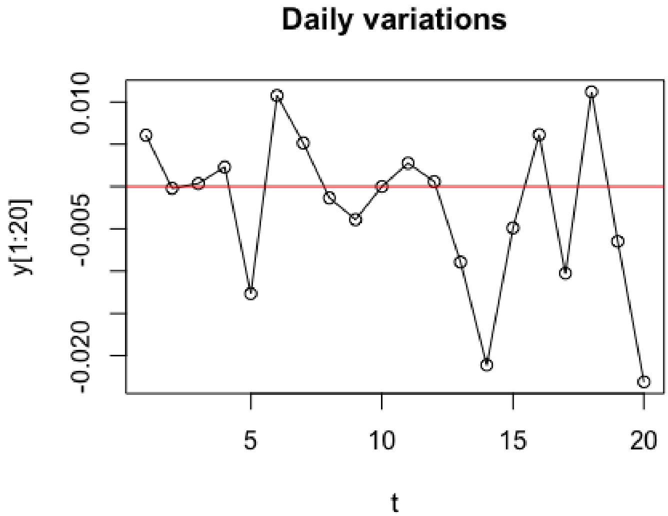


Figure 1: Daily values of $Sp500$ index

variables different from 0 adjusting the value of the regularization parameter λ, we show the results below:

lambda	#coeff	abs_err	rel_error
0.0001	194.0	0.000162	0.375579
0.0012	88.0	0.001827	4.145870
0.0023	57.0	0.002757	6.659445
0.0034	44.0	0.002785	6.503607
0.0045	31.0	0.003180	7.269857
0.0056	29.0	0.003693	8.413636
0.0067	25.0	0.004112	9.388155
0.0078	20.0	0.004444	10.099570
0.0089	18.0	0.004775	10.675141
0.0100	15.0	0.005058	11.193489

Figure 2: $Sp500$ replication varying λ

Unfortunately, this is not very useful. At time t we are interested in predicting $Sp500$ at time $t+1$ and according to that planning our financial strategy. The simplest way to do it is by using our linear model shifting y_t by a step h with respect to x_t:

$$\min_{\beta} \sum_{t=1}^{T} (y_{t+h} - \beta_0 - \sum_{j=1}^{p} x_{tj}\beta_j)^2 \quad \textit{subject to} \ \sum_{j=1}^{p} |\beta_j| \leq C. \tag{2}$$

As mentioned before, due to the high variance of the process this naive approach is quite inefficient and this will be a problem in the following sections where we will try to build up a financial strategy based on these predictions

	abs_err	rel_err
step		
0	0.001141	1.197119
1	0.006745	3.045187
2	0.006760	1.948486
3	0.006685	2.978954
4	0.006900	4.646308
5	0.006854	4.840236
6	0.006671	4.589862
7	0.006947	3.858749
8	0.006905	3.995296
9	0.006819	2.177310
10	0.006930	2.945953

Figure 3: Accuracy in predicting y_{t+h} given x_t

and focusing only on the one-step ahead prediction:

2.3 Parameter estimation

As we said our principal aim is to replicate the $Sp500$ with a relatively low number of assets. To select the right one we could use, as explained in the previous paragraph, lasso or elastic net regression; now we have to deal with the parameter estimation.

As we said we suppose, as actually happens due to the fact that our index is just a combination of 500 assets, that $Sp500$ has a linear relation with the

	y	y_pred	abs_err	rel_err
151	0.000016	-0.000356	0.000372	23.797417
152	-0.005572	0.002145	0.007717	1.385009
153	0.011465	-0.000673	0.012139	1.058711
154	0.002756	0.003383	0.000628	0.227798
155	-0.001638	0.001170	0.002808	1.714527
156	0.006685	-0.000845	0.007530	1.126447
157	0.004336	0.000721	0.003616	0.833805
158	-0.000061	0.002208	0.002270	36.980944
159	0.008496	0.000232	0.008264	0.972722
160	0.004988	0.002313	0.002676	0.536376
161	0.002475	0.002422	0.000053	0.021479
162	0.002946	0.001322	0.001623	0.551168
163	-0.001995	0.001159	0.003154	1.581262

Figure 4: One-step ahead prediction

selected assets. A first and natural choice for the β of the model are the one obtained by the non-negative lasso regression or non-negative elastic net regression on the whole set of $Sp500$ constituents. Actually this gives nice results, but we will see in experimental phase that selecting the assets and than estimate the model using non negative OLS works better.

As stated in the previous paragraph, we will use a moving window approach to build the models for both passive and active portfolio management. It's clear in fact that the size of the used window will affect all the further analysis. After different tests, presented before, we decided to use a window 250 observation long. In this way we end up having enough data to build meaningful models, but not too much, because we want to preserve the time invariance of the data inside the window.

3 Portfolio replication

In this section we will present the two developed strategies to manage the portfolio. One of the principal aim of portfolio management is the replication of a certain index performance, using only a relatively small number of assets. This is justified by the fact that a portfolio with 50 assets is much more flexible and easy to manage with respect to the one with 500 assets. Also an other good reason is that, depending of the type of taxation, could be necessary to pay a fixed amount of fee for each asset we decide to invest on. So if we are able to find a small subset of the 500 assets able to guarantee similar performances, we will gain in terms of management simplicity and lower taxation.

3.1 Beta

Using a moving window of length 250 and using an appropriate regularization parameter λ in order to have a number of variables (i.e number of β-parameters to estimate), we obtain the following result: We want to understand how rapidly

	intercept	0	1	2	3	...	489	490	491	492	number_coeff
0	0.000096	0.021194	0.019399	0.007723	0.015907	...	0.0	0.0	0.0	0.0	104.0
1	0.000096	0.021205	0.019832	0.007389	0.014799	...	0.0	0.0	0.0	0.0	111.0
2	0.000096	0.022710	0.018177	0.007424	0.018242	...	0.0	0.0	0.0	0.0	107.0
3	0.000096	0.022527	0.017844	0.007458	0.017294	...	0.0	0.0	0.0	0.0	107.0
4	0.000096	0.023029	0.017066	0.007755	0.016679	...	0.0	0.0	0.0	0.0	107.0
5	0.000096	0.023214	0.016186	0.007356	0.015540	...	0.0	0.0	0.0	0.0	107.0
6	0.000096	0.023670	0.016511	0.007017	0.014610	...	0.0	0.0	0.0	0.0	108.0
7	0.000096	0.024285	0.017344	0.007417	0.014157	...	0.0	0.0	0.0	0.0	112.0
8	0.000096	0.023799	0.017801	0.007603	0.015049	...	0.0	0.0	0.0	0.0	113.0
9	0.000096	0.023484	0.017825	0.007695	0.015908	...	0.0	0.0	0.0	0.0	113.0

Figure 5: Intercept and β-coefficients predicted by our model

these β-parameters change, in particular the turnover among the variables and, in case the model trained in the window$[t+1 : t+251]$ introduce a new variable with respect to the model trained in the window$[t : t+250]$, which is the weight of this new variable inside the model (and vice versa for a variable leaving the model). What we get is that on average only 5.49 new variables are added to the new model (circa 4%) and 5.47 are removed (circa 4%). Moreover, new assets enter the model with really low weight (so they still are almost irrelevant) and on the other hand leaving variables are unsurprisingly among the least important in the previous model. In fact, the mean value of the β-coefficients of the new variables is 0.000858 (compared with a 0.006668 mean value of all the β-coefficients of the chosen variables, only the 0.12% of the chosen variables has a lower weight). We have a similar result also for the variables leaving the model. These results show that we can assume that our "relevant" variables, the ones

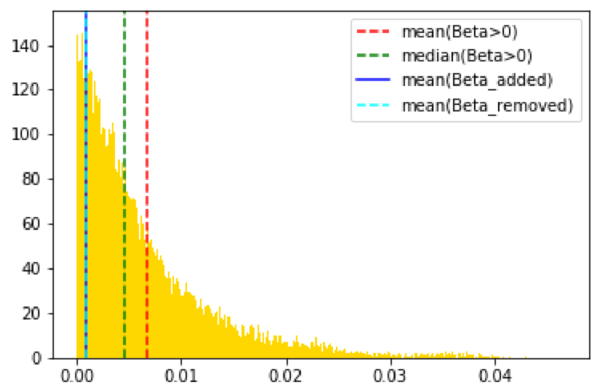
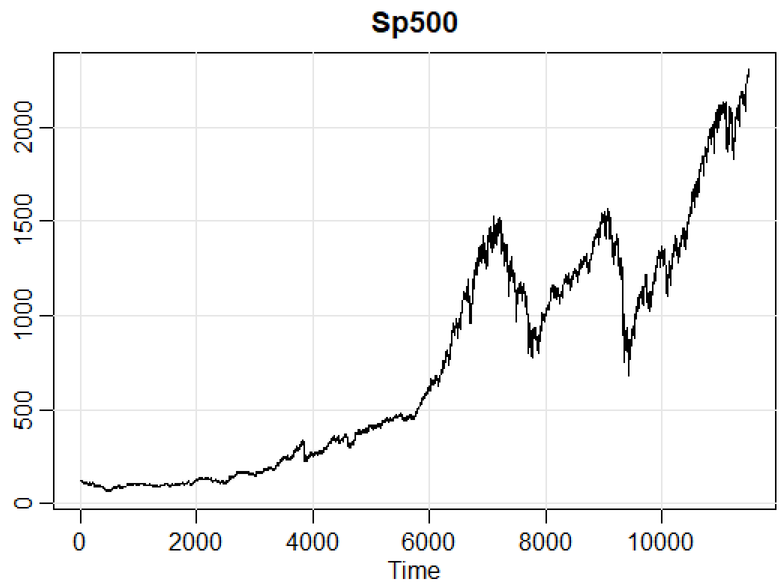


Figure 6: Relevance of the added/removed variables inside the model

chosen by our Lasso model at time t, will probably still be the important ones at time $t+1$ due to the fact that the only few turnovers we have are among the least important assets.

3.2 Passive approach

With passive approach, given that one tries to choose a growing index, we want to mimic as much as possible the index itself, in order to have its same performances. We will work with $Sp500$, below we report the $Sp500$ time series.



As it's clear by the plot, the considered index, on average, grows with time, so we would have been able to perform just like the index also our portfolio would have increased its value.

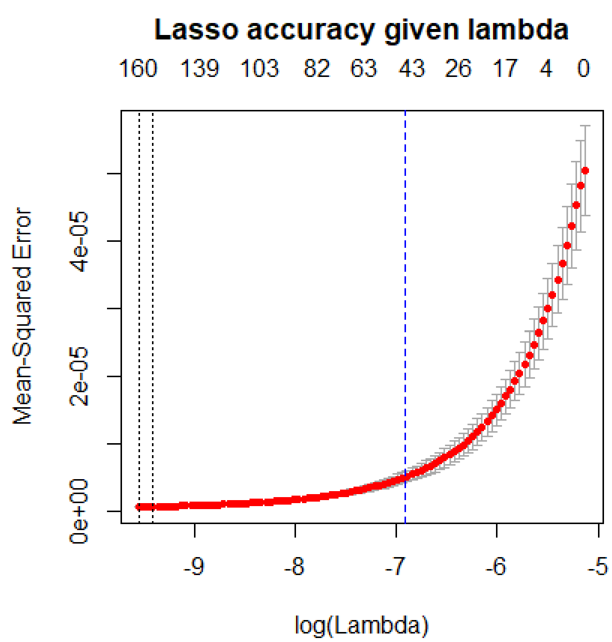
In previous paragraphs we already underlined how we could use Lasso in order to select the right subset of asset. So our passive approach to portfolio replication problem will be handled using lasso or elastic net.

A first way to build the portfolio is to use non negative lasso regression on the starting window, representing the time at which one decides to starts to invest, and distribute the budget proportionally to the β_{lasso}.. Given that we will work on daily log-returs, $\log(x_t) - \log(x_{t-1})$, both on dependent and independent variable. It's easy to see that $\exp(y_t)$, where y_t is the responce variable ($Sp500$), it's equal to the proportional value of $Sp500$ with respect to the value it take one lag before. For this reason, estimated the model on the first 250 observations, we can get the value of our portfolio after t day from our investment as $\exp(\sum_{i=1}^{t} \hat{y}_i)$, where $\hat{y}$ is the model's predicted value of the observation at time $250 + i$.

For this reason becomes computationaly easy to calculate the gain of a passive approach to portfolio management.

Said that we can actualy report the result of the passive approach to the studied problem.

For all the further presented simulations we will use for uor analysis the daily log-return of the $Sp500$ and its components form $07-01-2014$ till $09-02-2017$ (notice that some entries are missing, this is due to not collencted data). So we will start using the window made by the first 250 observations of our series and building a model using non negative lasso. As we said our aim here is to reduce the number of explanatory variable of the model; due to construction reason the best choise for λ would be 0, infact with all the 500 components we can perfectly explain $Sp500$. So we will choose our lambda in order to have a good trade-off between number of variable kept and prediction error.



The one reported above is the plot of the mean squared error value given a range of lambda going from $4.5\text{x}10^{-}5$ to 0.007. The blue line shows the selected value for λ, 0.001. This choise seems to be a good compromise between mean squared error and number of selected variables. Indeed the choosen λ guarantees small error being just before its explosion, and on the other hand empirical test on the moving window sais that, with that lambda, we select on average 51 assets. So seems reasonable to use it.

3.3 Active approach

Now we want to try to use our predictions to build up a financial strategy: we invest our money at time t in order to replicate $Sp500$ index only if we think that tomorrow the index will grow, otherwise we capitalize our portfolio and we don't invest money at time t. Every day we repeat this procedure trying to guess when the market will go up, investing only if this is the case. Of course, we must consider a variable fee (0.2%) we need to pay to make an investment. We can aggregate this tax to y, the daily log revenue, modifying its value $y \rightarrow y + log(1 - 0.002)$: in this way if the new y is still positive we want to reinvest our money even though this means to pay the additional fee.

3.3.1

	y	pred_y	real_gain
250	-0.000257	-0.003235	1.043583e-03
251	-0.002141	0.003040	5.218771e-04
252	-0.002002	-0.001770	1.857689e-04
253	0.001302	-0.001130	1.579566e-03
254	-0.001141	-0.001755	1.007089e-03
255	-0.006903	-0.001148	-3.253506e-03
256	-0.012366	-0.001203	-5.613125e-03
257	-0.002002	-0.002522	1.434524e-04
258	-0.002342	-0.001489	8.429011e-05
259	-0.020449	-0.000881	-1.790082e-02
260	-0.010935	-0.006411	-8.507162e-03

Figure 7: Portfolio's evolution with 69 variables

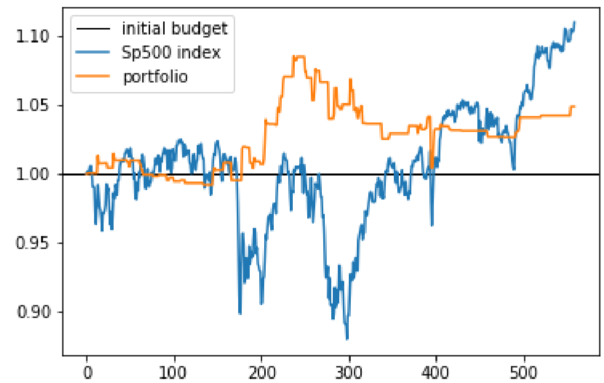


Figure 8: Portfolio's evolution compared with $Sp500$ index

3.3.2 Logistic prediction

Since our concern is about predicting every day which is the right choice between reinvest or capitalize our portfolio we can use a logistic model where if the prediction for day t is 0 we don't invest our budget at t, vice versa if the prediction for day t is 1 we invest our budget at t according to the β_{Lasso} coefficients estimated as before using $X[t-250:t-1]$.

3.3.3 Simulations

Non negative Lasso For the first simulation we will obtain our portfolio using non negative lasso. At this point we will invest in each asset proportionally its β.

Doing like that and calculating the relative value of the portfolio after 558 days, so keeping the same portfolio for the whole time period considered, we obtain that our portfolio now worth 1.191 times or initial portfolio, so we would have gained the $19,1\%$ percent of our initial investment. Below we report the result of the simulation. Note that there is inconsistency between the lambda choosed here due to sklearn and glmnet selection inconsistency, this is a known fact but we couldn figure out the reason, probably due to minimization problem definition.

	y	y_0/1	pred_y	real_gain
250	-0.000257	0.0	0.0	-0.000958
251	-0.002141	0.0	0.0	-0.001480
252	-0.002002	0.0	1.0	-0.001816
253	0.001302	1.0	0.0	-0.000422
254	-0.001141	0.0	0.0	-0.000995
255	-0.006903	0.0	0.0	-0.005256
256	-0.012366	0.0	1.0	-0.007615
257	-0.002002	0.0	0.0	-0.001859
258	-0.002342	0.0	0.0	-0.001918
259	-0.020449	0.0	0.0	-0.019903
260	-0.010935	0.0	0.0	-0.010509

Figure 9: Portfolio's evolution with 69 variables

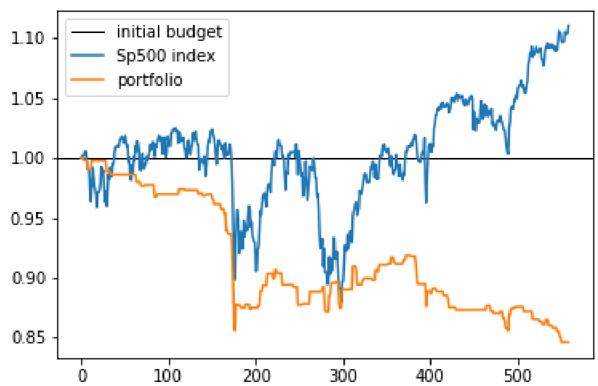
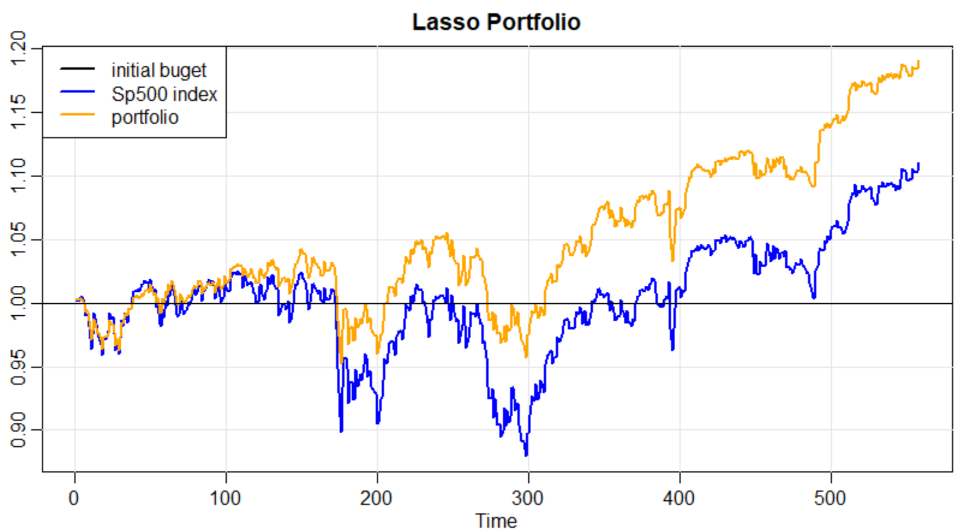


Figure 10: Portfolio's evolution compared with $Sp500$ index

	Lambda	# Coeff	R2
1	1e-04	147	0.9981005
2	3e-04	96	0.9900534
3	6e-04	70	0.9710203
4	9e-04	51	0.9448159
5	1e-03	47	0.9351245



So from the figure above we can see that our porfolio follows the $Sp500$ index really well initially and that as the time goes it differenciate from it. This is due to the fact that the model coefficient are not static in time and so even if we estimated a good model it stais good for a fixed period. This fact underlignes the needs of update the beta coefficient after a while. At the end of this simulation we gain 8.1% more than the real indx, so not a good approximation of the index but we selected the assets that on average grew more than the market.

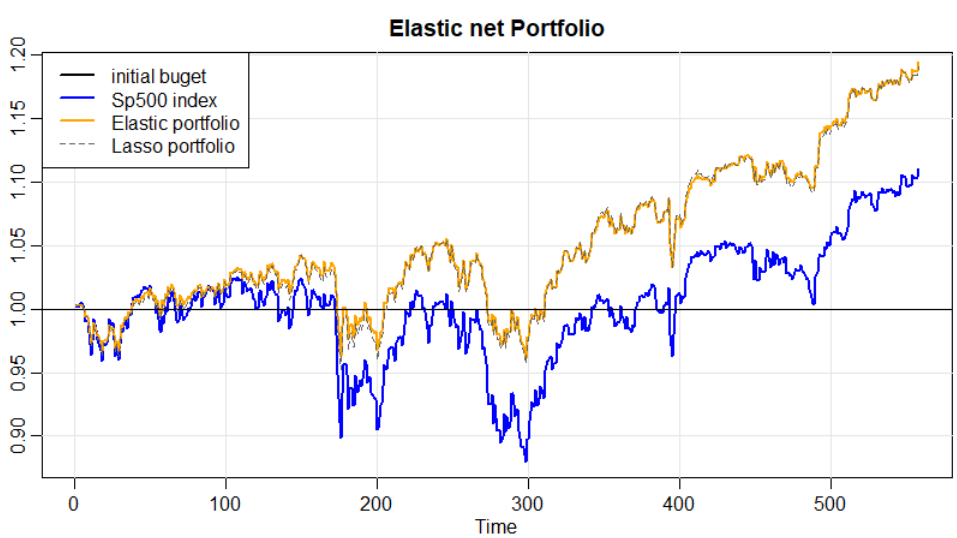
Non negative Elastic Net Non negative lasso seems to work pretty well, but as we can see from the picture above happend that our portfolio value is quite different from the $Sp500$ index, surelly this can bring advantages if one decides to exit from the market in a moment in which he is earning more than the general index, but also the opposite situation is equally likely. So we would like to stick as much as possible to the index.

If one thinks to the data we are working on, actually can happend to have an asset that is not really growing with time nor decreasing. This situation creates assests with very low variance and it brings to distorted model's coefficient for that specific asset. For this reson, also due to what we said in chapter (2), adding a ridge constrain to our minimization problem can help finding better beta.

Adding a ridge constrain ends up being elastic net regression. As for lasso, also here we will work only on non negative coefficient.

The general behaviour when varing the λ value is the same as before;the behaiour of the portfolio time series varing the α parameter is interesting. α represents on which percentage the l-1 regularization enters in the minimization problem, $1 - \alpha$ will be the percentage the ridge regularization enters. Infact mantaining almost constant the number of coefficients, adjusting properly α and β, the result is almost the same. So seems much more important the number of assets one uses whith respect the α and β parameter.

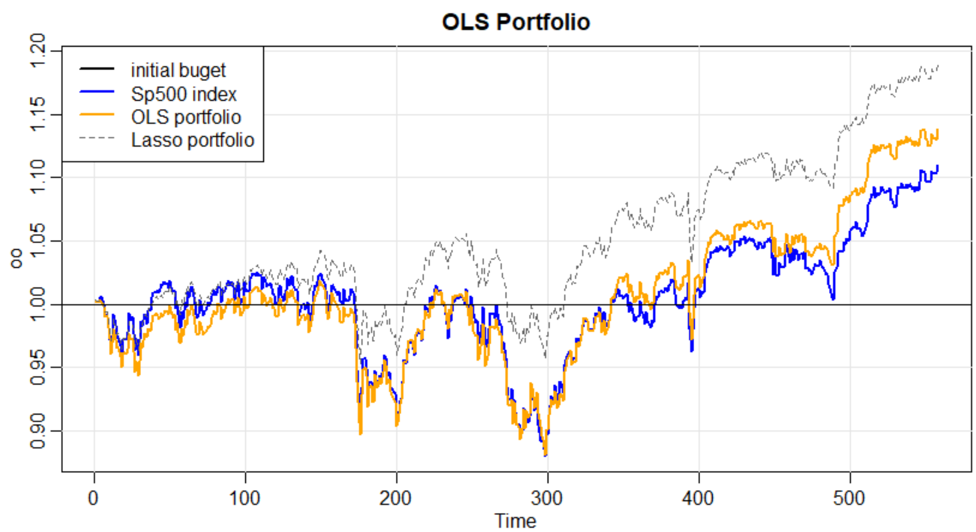
Below the plot containing the simulation results using non negative elastic net.



So it's clear that the two results are absolutely comparable. In the case reported above we used for elastic net $\lambda = 0.0025$ and $\alpha = 0.5$, but as previously said changing the two parameter the result doesn't change.

Non negative OLS In this last simulation we will estimate the model's parameters using ols estimate on the lasso-selected assets. Different paper reported that empirically seems to work better then simply using lasso estimate.

Below the plot with the simulation results.



Actually also in our case OLS approximate the original index much better that the other two approches. So at the end of our simulations we would say that at the moment the better strategy to build a portfolio is to use lasso an invest on the selected assets proportionally to the ols estimate of the corresponding β. With this approach at the end we gain less that using the others, but this can't be prooved to be a general characteristic. Here it happend that lasso and elastic net suggest to invest more on the assets that in future will grow more that the average, further work on investigating if this behaviour repeats also with other indexes would be interesting.