ISSN 0798 1015

REVISTA ESPACIOS

Vol. 39 (Number 33) Year 2018 • Page 32

Penalized regression approach to the portfolio selection problem considering parameter uncertainty

Enfoque de regresión penalizada para la selección de portafolio considerando incertidumbre en los parámetros

Diego H. SEGURA [1]; Carlos VALENCIA [2]

Received: 05/03/2018 • Approved: 12/04/2018

Content

1. Introduction
2. Methodology
3. Application
4. Results
5. Conclusions
Bibliographic references

ABSTRACT:
The portfolio selection problem can be viewed as a maximization of the risk-return relation based on the input parameters, expected returns and covariance between assets. As these parameters depend on historical market data with stochastic behavior, their real values are not achievable. The estimators haul error that result in under-performance of the selected portfolio. As a solution, we present and analyze a penalized linear regression methodology, which constrains the decision variables to limit the estimation risk of the parameters.
Keywords: Portfolio selection, penalized regression, elastic net, estimation risk.

RESUMEN:
El problema de selección de portafolio puede ser entendido como la maximización de la relación riesgo-retorno basada en los parámetros de entrada, retornos esperados y covarianzas entre activos. Estos parámetros dependen de datos históricos del mercado con componente aleatorio y sus valores reales no son conocidos. Los estimadores tienen errores que conllevan a un comportamiento sub-óptimo del portafolio. Como una solución, presentamos y analizamos una metodología de regresión penalizada, restringiendo las variables de decisión para limitar el riesgo de estimación.
Palabras clave: Selección de portafolio, regresión penalizada, red elástica, riesgo de estimación

1. Introduction

The art of making money in the stock market is indeed an art, as it seems impossible to follow a simple recipe. Investors are differentiated based on their risk aversion profile and their portfolios performances. As the risk aversion profile is intrinsic to the investor, it is possible to study how to choose a better portfolio selection strategy for a given risk. These studies can be agglomerated in what is called “Modern portfolio selection problem”, which has been studied since Harry Markowitz published his paper “Portfolio selection” in 1952

(Markowitz, 1952). In his seminal paper, it is proposed that the investor can select a group of assets of the stock market, wisely choosing a weight that is held in each asset, not only motivated by risk minimization or return maximization, but thinking in the diversification of his investment. Following these ideas, it is possible to think that Markowitz finally created a recipe, an optimization problem that can be *easily* solved, yet, there is art behind. Although Markowitz created a new framework, it remains conceptual because, as he recognizes, his model does not take into account the uncertainty of its parameters: expected returns of the assets and covariance between them. As we like to maximize the return-risk relation, these model parameters are significantly important. Pitifully, their real values are not easily estimable and then, we can only rely on historical data to predict future performances. However, new approaches have been formulated aiming to improve model's performance as they consider parameter's uncertainty and also, taking advantage of special statistical properties as sparsity, concept that will be discussed later. Interestingly, the key of success of newest models is the creation of additional constraints, which is counterintuitive in an optimization problem. As we have a vector of weights as our decision variable, some authors have focused on constraining different norms of the vector, specifically, L1 or L2 norms [3] (DeMiguel et al., 2009; Li, 2015).

The purpose of this article is to provide more detailed evidence on how the penalized regression model improve the performance of optimal portfolios, using back-testing with real data from the S&P 500, and under different comparison metrics and levels of risk aversion. Firstly, we present how to solve the risk-return Markowitz' relation as a linear regression model and then, how to penalize it, creating the elastic-net model. Finally, we implement the model with real data and explore the results showing that, due to the parameter's uncertainty, the constrained model has superior performance than the traditional one.

2. Methodology

2.1. Linear regression approximation for the portfolio selection problem

We know that each investor has a portfolio with an expected return of $E(w) = w^t\mu$ with some risk over his investment of $\sigma(w) = w^t\Sigma w$, knowing that w is a column vector of p weights, one for each asset, μ is a column vector with the expected return of each asset and Σ is the $p \times p$ variance-covariance matrix. Then, an investor would like to maximize the risk-return relation proposed in $U(w)$ where γ is a risk aversion coefficient.

$$U(w) = w^t\mu - \frac{\gamma}{2}w^t\Sigma w. \tag{1}$$

Thus, the optimal value of w is found with $\frac{\partial U(w)}{\partial w} = 0 \Rightarrow \mu - \gamma\Sigma w^* = 0 \Rightarrow w^* = \frac{1}{\gamma}\Sigma^{-1}\mu$. As it is proposed in Li (2015), knowing that the OLS estimator of a linear regression model $Y = Xw + \varepsilon$ is obtained minimizing $(Y - Xw)^t(Y - Xw)$, , that is, when $-X^tY + X^tX\widehat{w} = 0$;, and comparing this equation with the one obtained after deriving (1) with respect to w, we can easily see that both of them are of the form $a + bw = 0$;, therefore, we can match their coefficients, obtaining:

$$\mu = X^tY,$$

$$\gamma\Sigma = X^tX.$$

As Σ is positive semi-definite matrix, it can be expressed as $\Sigma = UDU^t$, and then $\Sigma^{\frac{1}{2}}$ is estimable and $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$; thus, from $\gamma\Sigma = X^tX$ we obtain

$$X = \sqrt{\gamma}\Sigma^{\frac{1}{2}}, \tag{2}$$

and from $\mu = X^tY$ we find that

$$Y = \frac{1}{\sqrt{\gamma}}\Sigma^{-\frac{1}{2}}\mu. \tag{3}$$

Concluding that applying these definitions of X and Y, the portfolio selection problem can be solved with the OLS estimator of the linear regression problem $Y = Xw + \varepsilon$.

2.2. Improvements to the linear regression estimators

Linear regression model is somehow restricted because of its assumptions. For instance, if variables X_i and X_j are highly correlated, the estimators $\hat{\beta}_i$ and $\hat{\beta}_j$ will not lead us to adequate conclusions. In order to manage this struggle, we would like to shrinkage the linear regression estimators correcting them. The elastic-net model showed in (4) solves the linear regression model but creating a constraint over the L1 norm and the L2 norm of the vector of coefficients, shrinking it.

$$\underset{(\beta_0,\beta)\in\mathbb{R}\times\mathbb{R}^p}{min}\left\{\frac{1}{2}\sum_{i=1}^{N}(Y_i-\beta_0-X_i^t\beta)^2+\lambda\left[\frac{1}{2}(1-\alpha)\|\beta\|_2^2+\alpha\|\beta\|_1\right]\right\}. \tag{4}$$

This model will help us correcting various things. First of all, there is an inevitable correlation between the assets that conform the stock market and so, a basic linear regression model could discard important information of highly correlated variables. Applying the elastic-net model, portfolio weights of similar assets are regulated. Then, this model aids us lessening the risk of our estimations due to the uncertainty of the expected returns, the covariance matrix and its inverse. Is it known that correlation between assets will lead to a covariance matrix that is "ill- conditioned" and a useful trick to cope that problem is to increase by a constant each element of the matrix diagonal, $\Sigma'=\Sigma+\delta I$. When we create the term $\|\beta\|_2^2$ in the elastic-net model, we are improving the estimation of Σ^{-1}(Li, 2015). According to Shen et al. (2014), this L2 constraint also makes that portfolio weights remain similar under rebalances on consecutive investment periods but it is beyond of this study to do so. On the other hand, we can measure the estimation risk as the difference between equation (1) evaluated with the real parameters, and the same equation evaluated with the estimated parameters, *i.e.*,$\left|U(w;\mu,\Sigma)-U\left(w;\hat{\mu},\hat{\Sigma}\right)\right|$. This calculation is regulated by the L1 norm of the vector of coefficients with the expression

$$\left|U(w;\mu,\Sigma)-U\left(w;\hat{\mu},\hat{\Sigma}\right)\right|\leq\|\hat{\mu}-\mu\|_\infty\|w\|_1+\frac{\gamma}{2}\left\|\hat{\Sigma}-\Sigma\right\|_\infty\|w\|_1^2. \tag{5}$$

Therefore, creating a constraint over the L1 norm value will reduce our estimation risk of μ and Σ (Fan et al., 2012) and a constraint over the L2 norm reduce the estimation risk of Σ^{-1}.

2.3. Sparsity

When we take statistical models to high dimensions (largenumber of predictors), we are moving to a world of sparsity. The general rule of this world is, roughly speaking, that when we have a large amount of predictors it is intuitive to think that only a group of them are relevant to describe our variable of interest.

In context of our problem, we have *p* as the number of assets that could be part of the portfolio. As we all are aware, not all stocks are good for the investor; therefore, he should not invest in all of them. The solution of (1), namely the Markowitz' portfolio (MP), will assign a weight for each of the proposed assets. Some of them will be small, even pretty near to zero, if the associated securities are not relevant to the portfolio; nevertheless, the model will never be able to actually shift those small weights to zero even if is desired.

Here again, elastic-net outstands over the MP. When the linear regression has L1 and L2 norm constraints, besides of having the advantages exposed in 2.2, it will also select relevant variables. In pursuance of satisfying the norm constraint, some of the portfolio weights will go to zero in optimality, helping the investor to choose only some *k* of *p* assets that are statistically relevant.

Furthermore, as portfolios are rebalanced regularly, the investor need to assume some transaction costs each time he buys or sells shares. While the MP will force him to pay those costs for each of the *p* assets, disregarding if the asset has a high or small weight in the portfolio; the elastic-net solution with *k*<*p* assets reduces the transaction cost.

2.4. Algorithm implementation

The algorithm that is used to solve the proposed model is the following: Using the train data, we found $\hat{\mu}$ and $\hat{\Sigma}$. Then, we created X and Y as proposed in 2.1, in equations (2) and (3). Afterward, the elastic-net is implemented. To do so, a grid of possible values for α and λ parameters is created and then the optimal values (those that generate the minimal MSE) (α^*, λ^*) are found using cross validation. Finally, we run an elastic-net model with α^* and λ^* obtaining 442 coefficients that are the portfolio weights. These weights are implemented along 2015 to evaluate the *out-of-sample* performance of the portfolio. In summary,

Table 1
Algorithm of implementation

1. Estimate $\hat{\mu}$ and $\hat{\Sigma}$ with the training data

2. Create Y and X with

$$Y = \frac{1}{\sqrt{\gamma}}\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}$$

$$X = \sqrt{\gamma}\hat{\Sigma}^{\frac{1}{2}}$$

3. Calibrate with cross validation the values of $\alpha \in A$ and $\lambda \in \Lambda$ for the elastic-net model using a grid of $A \times \Lambda$ where $A = \{0.1, 0.3, 0.5, 0.7\}$ and $\Lambda = \{\lambda \in \mathbb{R} | \lambda = 0.001 + n \cdot 0.01, n \in \mathbb{N}, n \in [1,500)\}$

4. Fit elastic-net with tuned parameters α^* and λ^*. Portfolio weights $\hat{w}$ are defined as the vector of the estimated coefficients of the elastic-net.

It is important to mention that $\hat{\Sigma}$ is estimated by Principal Orthogonal ComplEment Thresholding (POET) method, proposed in Fan et al. (2013) to do the estimation of large covariance matrices. On the other hand, $\hat{\mu}$ is simply calculated as the mean of the daily logarithmic returns during the training window for each asset, resulting in a p dimensional vector.

3. Application

3.1. Data selection

For this case of study, we evaluate the portfolio performance with stocks that belong to S&P 500. We obtained daily information -from 1998 to 2013- of the 500 stocks that were included in the index in 2013. Then, we also searched for data of those stocks up to 2016. As there is not full available data for all of the quotes, we removed some invalid entries to end up with data for 442 stocks. Having daily close prices P_t, we estimated the logarithmic returns, $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$ for each selected asset from 2010 to 2015. We then defined a time window of 5 years, form 2010 to 2014 as the training sample and 2015 as the test sample.

3.2. Model implementation

We use **R** to apply the algorithm proposed in 3.2. Firstly, POET package is used to create $\hat{\Sigma}$ using soft thresholding and $k = 15$. k is the number of factors and was chosen after analyzing a scree plot, viz. plotting the eigenvalues of $\tilde{\Sigma}$, where $\tilde{\Sigma}$ is the covariance matrix estimated traditionally. In this way, from the scree plot we saw that for $k > 15$, the plot is almost a straight line so a greater value of k is not necessary. Having $\hat{\Sigma}$ and $\hat{\mu}$ we can continue with step 2: creating X and Y. This transformation depends on γ, a risk aversion coefficient; in this study we are not interested on how γ should be estimated, as it deserves another paper. For this study, we are going to use a risk aversion coefficient of 3.7 to compare the elastic-net versus the MP.

It is assumable that a higher coefficient leads to a lower value of the utility function and a lower coefficient leads to a higher utility function in any of the portfolio selection strategies. To show that behavior, we are going to evaluate the elastic-net performance when γ is 1, 3, 3.7, 5 and 10.

Lastly, we use glmnet package to create the elastic-net model. This package was used in steps 3 and 4 because allow us to perform the necessary cross validation and estimate the final model.

3.3. Model validation

For this case of study, we are not going to perform portfolio rebalancing. Instead, after obtaining w from the training sample, we evaluate the behavior of that portfolio during 2015.

In the first place, we would like to see how the elastic-net model performs over the MP to see if we are obtaining a better result. As a tool of evaluation, we are going to plot the cumulative wealth index of each strategy (elastic-net and MP) and the passive index strategy (investing in SP500 index) during 2015.

Cumulative wealth (CW) is measured as

$$CW_t = CW_{t-1} \cdot (1 + R_t) \quad for\; t = 1, \dots, T$$
$$CW_0 = 1\,USD,$$

where for each t,

$$R_t = w^t r_t,$$

r_t being the column vector of logarithmic returns of the assets in the portfolio on day t and w^t the transposed vector of weights of the investment strategy.

That is to say, if we invest 1USD in the portfolio in the first stock day of 2015 and then, we reinvest our returns on the portfolio every day, how much money we will receive at a day t.

We are also interested in how many entries of w vector are different from zero in each strategy. So we will find the percentage $\text{NonZero} = \frac{|Z|}{p} * 100\%$, where $|\cdot|$ is the cardinality of a set $\cdot$, Z is a set that contains all the assets that conform the final portfolio and p is the number of total assets.

Besides, we are going to calculate the mean absolute deviation and annualized Sharpe Ratio (SR) for each strategy; a portfolio with higher SR would be desirable. In addition, the Information Ratio (IR) will be calculated to compare how the elastic-net and the MP performances over the benchmark, which will be the S&P 500 index for us. The IR is the rate between the active premium and the tracking error, *i.e.*,

$$IR = \frac{E[R_p - R_b]}{\sqrt{Var[R_p - R_b]}}, \qquad (6)$$

Where R_p is the vector of portfolio returns and R_b the vector of S&P 500 index returns.

Another measure of risk-return relation it is the Calmar Ratio. It helps us estimating the relation between the average return during some period and the maximum drawdown in the same period. We are going to evaluate this measure for both MP and elastic-net knowing that a Calmar ratio greater than one is good, greater than three is excellent and above five is more than desirable (Young, 1991). We will also provide plots of the daily return and drawdowns for each strategy.

4. Results

First of all, we are going to compare the performance of the elastic-net versus the traditional MP using a risk aversion coefficient of 3.7. The choice of this value will be discussed later.

As expected, the constraint over the L1 norm of the vector of coefficients in the elastic-net model helped us creating sparsity, as it is viewable in Table 2. The percentage of non-zero weights is 15.61% for the elastic-net compared to 100% of the MP. In this way, the elastic-net model shows that it can shift some weights to zero lessening the transaction costs.

Tabla 2
Percentage of non-zero weights for each strategy

	Elastic-net	**MP**
Number of non-zero weights	69	442
Percentage of non-zero weights	15.61%	100%

However, it is still important to see if having fewer assets in the portfolio yields better results. The plot of the cumulative wealth index is shown in Figure 1 where we compare the elastic-net portfolio, the MP and the CW investing directly in the S&P 500 index. During the chosen sample, the elastic-net model outstands over the others because, for the same level of risk aversion we obtain a higher plot of the CW. Figure 2 presents the same CW analysis with different levels of risk aversion.

It is viewable that starting with one dollar in each strategy and reinvesting profits daily, the elastic-net portfolio ends the year with 2.32 dollars while the MP ends with nearly 1.67 dollars.

Tabla 3
Portfolio performance indicators

	Elastic-net	**MP**
Mean absolute deviation	0.018	0.018
Annualized Sharpe Ratio		1.971
Information Ratio	3.564	4.106
Calmar Ratio	7.303	3.241

In addition to the CW analysis, we obtained some portfolio performance indicators in Table 3. Firstly, we observed that when we measure risk of each portfolio based on mean absolute deviation, similar values were found, as it is 0.018 for both strategies. The annualized Sharpe Ratio has values greater than one proving a good performance over the risk free rate in MP and elastic-net; a greater SR for elastic-net shows that the model outstands over the MP. On the other hand, when we compare the portfolios versus the benchmark, both of them are desirable over the passive index strategy. It is important to mention that the MP showed a better information ratio. Lastly, we found that Calmar ratio for the MP is greater than 3, giving us excellent results. Furthermore, the Calmar ratio for the elastic-net is greater than 7 showing even a better performance.

Figure 1
Cumulative wealth index comparison between portfolio selection strategies

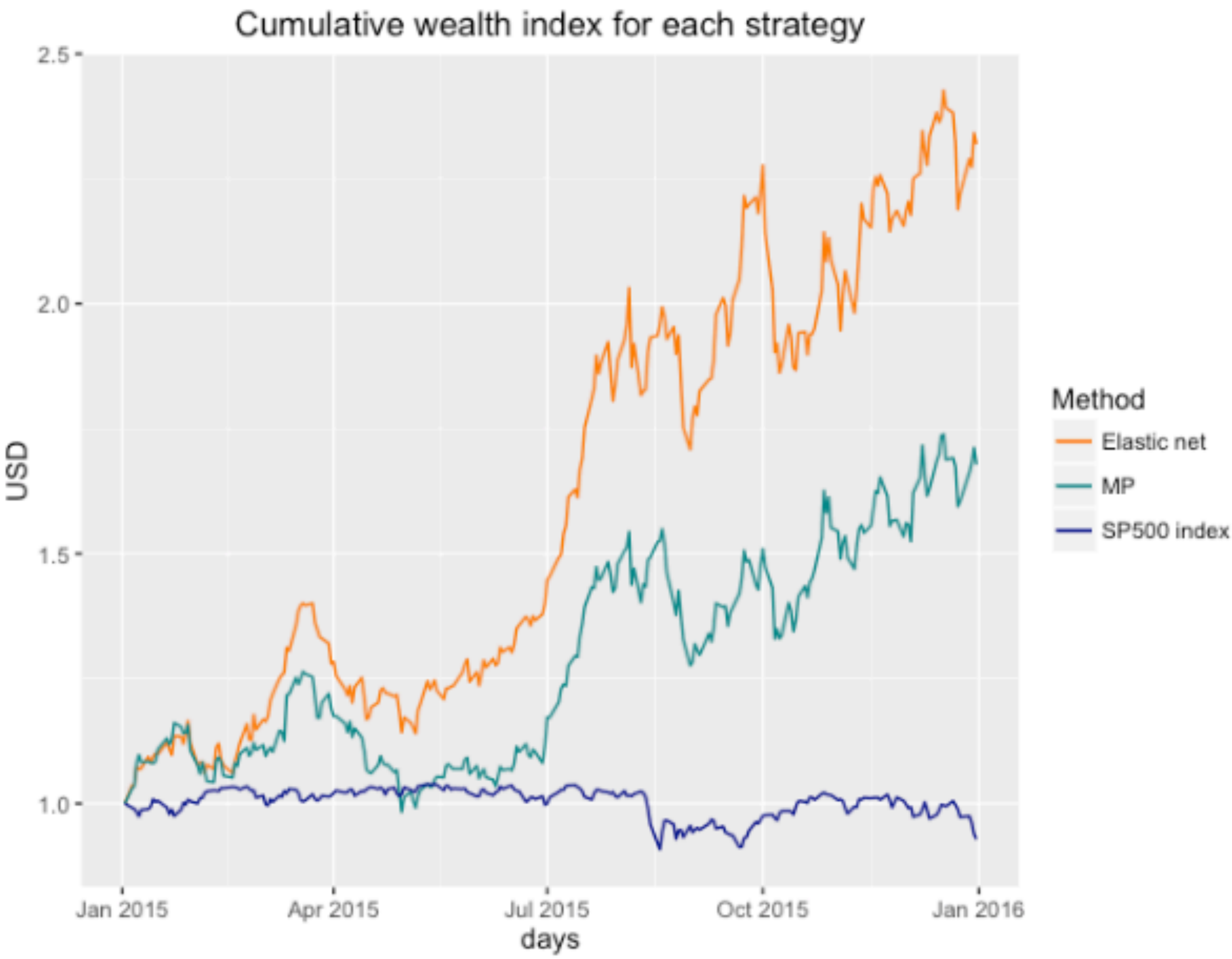
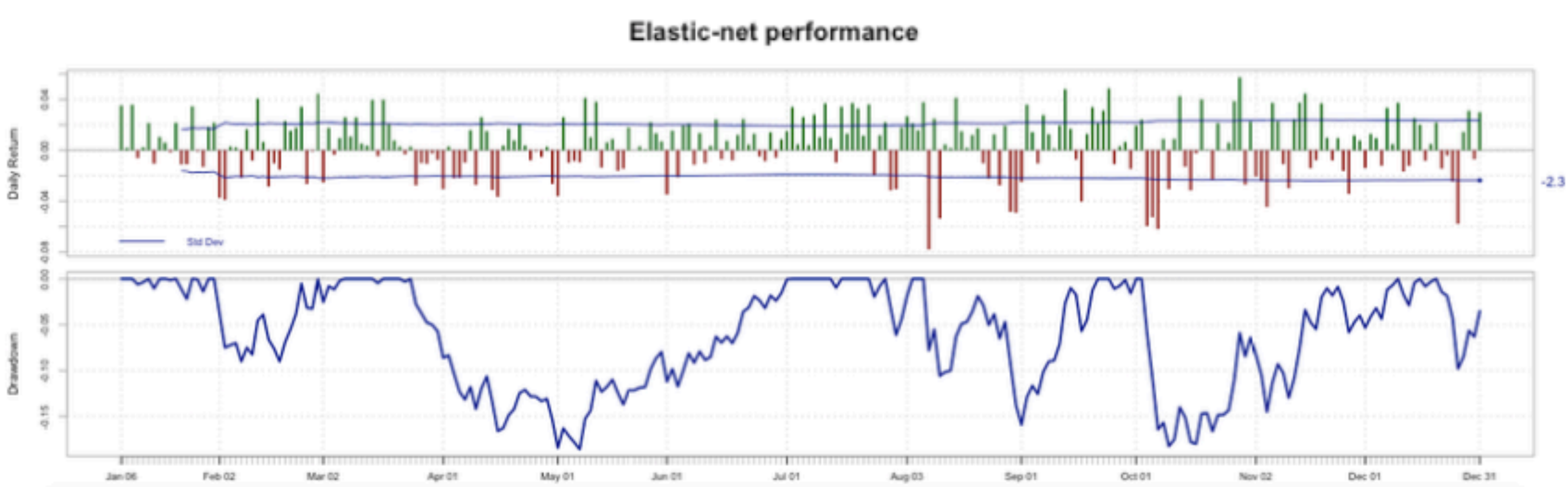


Figure 2
Cumulative wealth index using the elastic-net model for different type of investors

Considering these results, the elastic-net has a formidable behavior. Even though the MP scored a higher IR, the elastic-net has an excellent relation versus the S&P 500 index as well. It is important to mention that we are not taking into account transaction costs in any strategy. As the elastic-net has lesser assets, including transaction costs will reduce some performance indicators but not as much as it will for the MP; hence, creating a greater gap between the performances of both portfolios in favor of the elastic-net. Furthermore, we plotted daily returns and drawdowns for each strategy, as it is viewable in Figure 3 and Figure 4

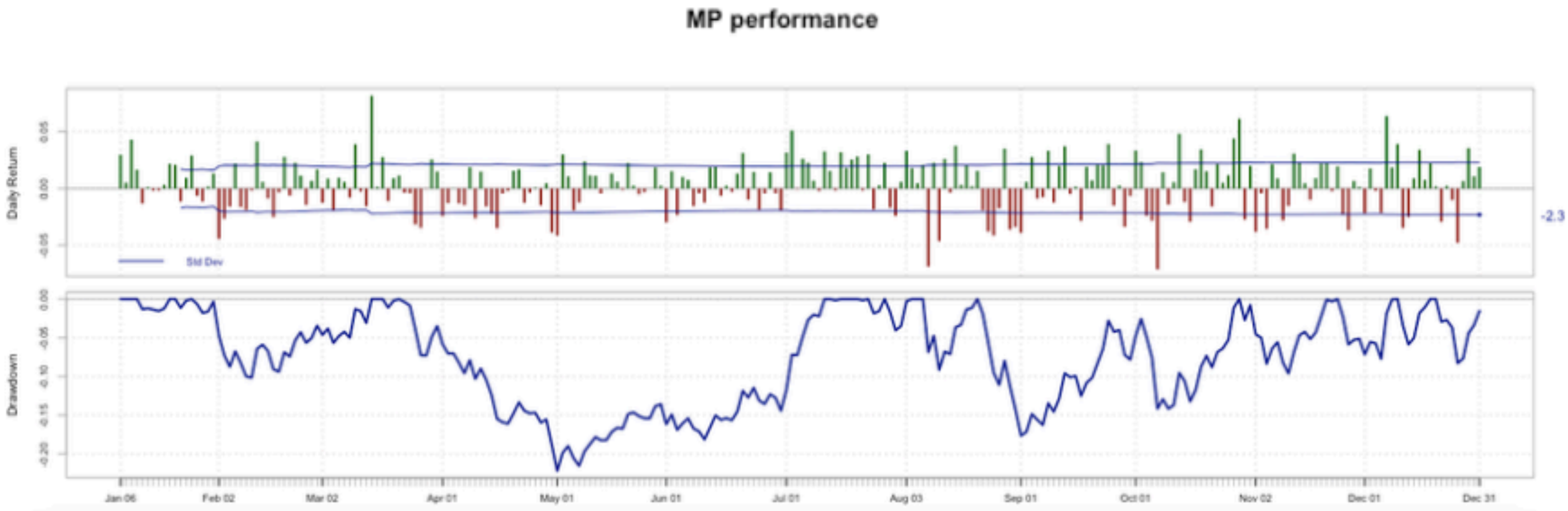
Figure 3 Elastic-net performance.



Considering these results, the elastic-net has a formidable behavior. Even though the MP scored a higher IR, the elastic-net has an excellent relation versus the S&P 500 index as

well. It is important to mention that we are not taking into account transaction costs in any strategy. As the elastic-net has lesser assets, including transaction costs will reduce some performance indicators but not as much as it will for the MP; hence, creating a greater gap between the performances of both portfolios in favor of the elastic-net.

Figure 4
MP performance



Furthermore, we plotted daily returns and drawdowns for each strategy, as it is viewable in Figure 3 and Figure 4. From these figures we can conclude that both of the time series have the same variance of their daily returns. We can see that the lower return for the elastic-net is -0.078 versus -0.071 for the MP during all 2015. Even though, in terms of drawdown, the maximum drawdown of the elastic-net is 18.5% while it is 22.2% for the MP; so the elastic-net has a better "worst case scenario" than the MP as the Calmar ratio showed in Table 3.

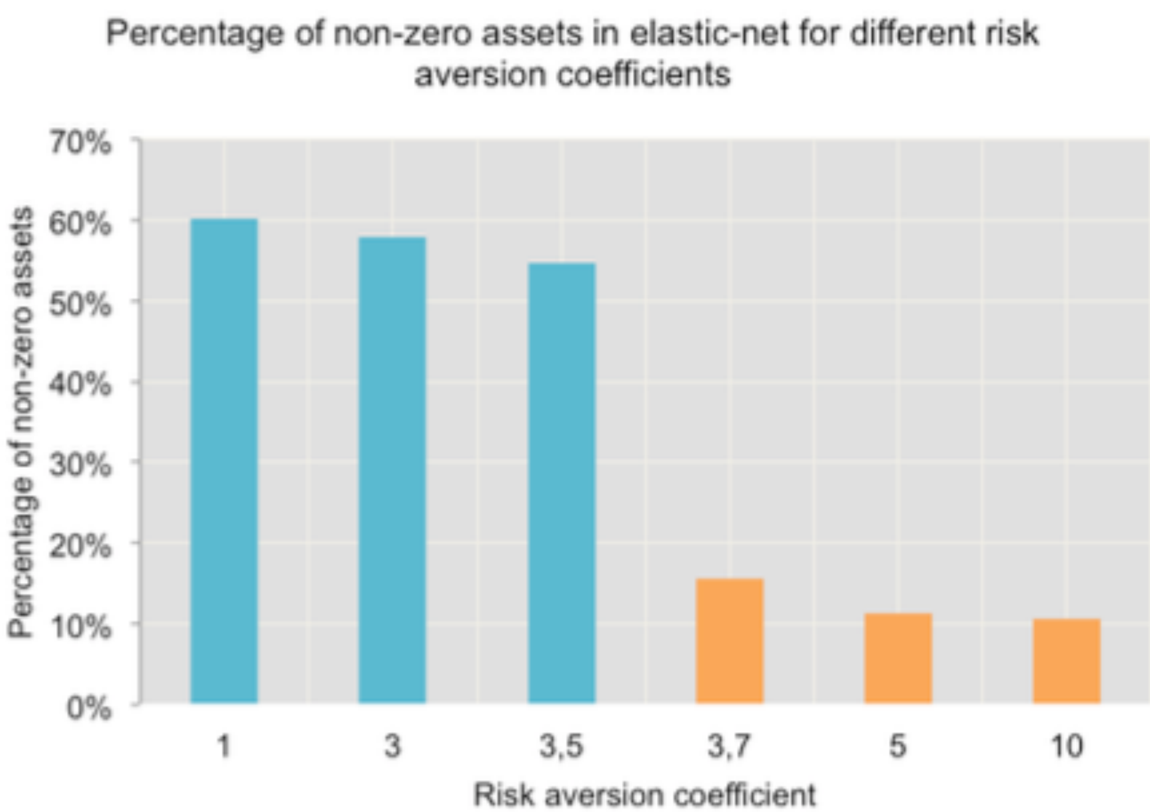
Finally, we analyze the behavior of elastic-net model when changes. Assuming that the risk aversion coefficient can be measured from 1 to 10 we evaluated the model with the values of 1, 3, 3.7, 5 and 10. From Figure 2 it is viewable that the CW is inversely proportional to the value of the risk aversion coefficient. As increases, we obtain lower curves of the cumulative wealth index

Table 4
Non-zero assets for elastic-net model for
different values of risk aversion coefficient

Risk aversion coefficient	Percentage of non-zero weights
$\gamma = 3$	$\frac{256}{442} = 57.92\%$
$\gamma = 3.5$	$\frac{242}{442} = 54.75\%$
$\gamma = 3.7$	$\frac{69}{442} = 15.61\%$
$\gamma = 5$	$\frac{50}{442} = 11.31\%$
$\gamma = 10$	$\frac{47}{442} = 10.63\%$

Figure 5

Non-zero percentages for elastic-net model and different values of γ



Two groups are clear in the plot, one for values of lesser than 3.7 and the other for values greater than 3.7. It is remarkable that for the group of the CW index line has more variance that the grouped lines of . Thus, we found that a risk aversion coefficient of 3.7 could be a good measure for an average investor and that justifies the value used for the comparison between MP and elastic-net.

Interestingly, when we evaluate the percentage of non-zero assets of each created portfolio when changes, the same two groups are easily recognizable. As Figure 5 shows, there are no significant changes in the percentage when the risk aversion coefficient changes between 1 and 3.7, and doesn’t change either when is between 3.7 and 10. Even though, the percentage is significantly different between the groups. When the percentage is near 50% and in other cases, it is approximately 10%. Real values are presented in Table 4. In any case, sparsity is present in the model; the elastic-net select assets and it is stricter as the risk aversion coefficient increases.

5. Conclusions

Some portfolio theories have been developed since Markowitz; nevertheless, some adjustments need to be done before implementing those theories in real data applications. It is impossible to invest without taking into account the randomness of the stocks prices in the market and then, portfolio selection models must include any control over parameter's uncertainty. To cope this problem, we proposed a penalized linear regression model known as elastic-net; which regulate the L1 norm and L2 norm of the vector of weights. Despite of the addition of new constraints to the optimization problem, it gives better results in empirical applications due to the uncertainty of the expected returns and the covariance between them. We showed that the elastic-net model has a better performance over the MP during a year of evaluation, without rebalancing, and using only 5 years of daily data to estimate the parameters.

Furthermore, using the constrained model helped us controlling the estimation risk of its parameters and it also helped us selecting which assets should or should not conform the portfolio.

The elastic-net model changes accordingly to the risk aversion coefficient and it remains for future works how to estimate an adequate risk aversion coefficient. It will be also interesting to study model's performance when some constraints are added; not the ones that control parameter uncertainty but portfolio constraints that are common in real financial applications. For instance, restricting the minimal percentage of health care assets in the portfolio. Some methods that have been proposed to work with sparsity, like non-parametrical graphical models, (e.g. Lafferty et al., 2012) can be applied to the portfolio selection problem and thus, as future work they could be compared with the elastic-net.

Bibliographic references

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance* 7 (1): 77–91.

Li, J. (2015). Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business & Economic Statistics* 33 (3): 381–92. doi:10.1080/07350015.2014.954708.

Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association* 107 (498): 592–606.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75 (4): 603–80.

Lafferty, J., Liu, H., and Wasserman, L. (2012). Sparse nonparametric graphical models. *Statistical Science* 27 (4): 519–37.

Young, T.W. (1991). Calmar ratio: A smoother tool. *Futures* 20(1), 40.

DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, *55*(5), 798-812.

Shen, W., Wang, J., & Ma, S. (2014, July). Doubly Regularized Portfolio with Risk Minimization. In *AAAI* (pp. 1286-1292).

1. Master student. Industrial Engineering Department. University of los Andes, Bogotá, Colombia. Email: dh.segura10@uniandes.edu.co

2. Professor. Industrial Engineering Department. University of los Andes, Bogotá, Colombia. Email: cf.valencia@uniandes.edu.co

3. The L1 norm of the vector w is the sum of the absolute value of each of its entries, this is, $\|w\|_1 = \sum_{i=1}^{p} |w_i|$. On the other hand, the L2 norm of the vector is $\|w\|_2 = \sqrt{\sum_{i=1}^{p} w_i^2}$. In general, the L$p$ norm is $\|w\|_p = \left(\sum_{i=1}^{p} |w_i|^p\right)^{\frac{1}{p}}$, when $p \to \infty$, $\|w\|_\infty = \max\{w_i\}$.