Improved portfolio selection using sparse statistics for parameter and portfolio weight estimation

Ernesto M. Cediel Cortés
Thesis Supervisor: Carlos Valencia Arboleda

Universidad de los Andes
Bogotá, Colombia

Abstract— **There are numerous ways to improve the empirical performance of the classic Markowitz mean-variance model. Designed to remediate the parameter uncertainty and estimation errors, these models use both stability and sparsity properties to achieve better out-of-sample performances. In this paper we will propose two different frameworks that take advantage of these two characteristics. The first uses L1 norm penalization to obtain sparse solutions with reduced estimation risk. The second approach focuses on the improvement of parameter covariance and precision matrix estimation: we first estimate the covariance matrix through a semi-parametric extension of the Gaussian graphical model using copulas, as the estimation is semi-parametric, it naturally becomes less dependent on data, achieving greater stability. Furthermore, this matrix is then penalized using Graphical Lasso optimization (GLASSO) to finally obtain sparse and stable covariance and precision matrices. With the results obtained in the empirical application, we show that these two frameworks outperform the benchmark models in the most common performance measures found in literature. Moreover, we find that the combination of both frameworks, yields the best performing model relative to other models evaluated.**

Keywords—portfolio selection; sparsity; stability; covariance matrix estimation; precision matrix estimation; penalization; graphical models; Graphical Lasso

1 Introduction

Since Nobel laurate in economics Harry Markowitz coined the mean-variance portfolio selection theory in 1952, the study of the portfolio selection problem has taken an important place in modern economics and finance. The portfolio optimization mean-variance model proposed by Markowitz guarantees optimal trade-off between mean (profit) and variance (risk) outputting weight solutions if true mean returns and covariance matrix are provided. However, since these two parameters are not obtainable in practice, they have to be estimated; investors naively use historical data as its main source, given its high variability, the estimated sample mean and sample covariance matrix has been proven to be far from the true values (Li, 2015). Furthermore, Carrasco and Noumon (2010) claim that this problem is aggravated mainly due to two issues, firstly, dimensionality significantly increases with the number of securities, and secondly, due to high correlation between these stocks. These limitations cause an ill-posed problem where optimal portfolio weights implicitly contain a huge estimation variance, where a slight change in the parameters can cause a greater than proportional change in these outputs. In consequence, parameter uncertainty and estimation errors lead to poor out-of-sample results (Kan and Zhou, 2007). Evidently, dimensionality has a key role in this modelling as of today, strictly in the New York Stock Exchange (NYSE) alone has more than 3000 stocks available to choose from.

Recent literature has shown to tackle the aforementioned issues, using multiple approaches. Ledoit and Wolf (2003) shrank the estimators of the covariance matrix by using penalization, and Carrasco and Noumon (2010) achieved similar results by performing L2-Penalization (Ridge) and spectral cut-off regularization. By shrinking these estimators, they achieve a sparse and stable estimation, obtaining better out-of-sample results. On the other side, some authors have attempted to find the optimal weighted averages between models, such as Tu and Zhou (2009) that combine both the naïve and the Markowitz portfolios. Jagannathan and Ma (2003) investigated the short-sale constraint, proving that this constraint improves empirical performance. Also, De Miguel et al. (2009) impose L1-norm and L2-norm constraints on portfolio weights in the minimum-variance portfolio. While it might seem unreasonable to consider sub-optimal solutions obtained by penalizing and altering the initial optimal model proposed by Markowitz, these yield better out-of-sample results as true parameters are not known and parameter uncertainty and estimation risk is drastically reduced, surpassing the effects of the bias imposed by these sub-optimal solutions.

The covariance matrix input in the portfolio selection problem has high dimensionality and high correlation across its components, typically resulting in an ill-posed matrix. Inverting an ill-posed matrix causes unstable results in estimating the stock weights (Carrasco and Noumon, 2010). Therefore, in this paper we focus on analyzing the impact of different methods of estimating the covariance matrix and computing its inverse in the portfolio selection problem.

Most of the papers mentioned before use the sample covariance matrix estimated with the traditional Pearson covariance approach, later on attempting to modify or penalize this matrix in a second step. In this paper, we propose a new train of thought, changing the estimation method of the covariance matrix in the first place. There are many methods to obtain a covariance matrix from a set of data, that can be categorized between parametric and semi/non-parametric approaches. To mention some, within the parametric approaches, we have Pearson, Kendall and Spearman covariance matrices, while on the non/semi-parametric approaches, Lafferty, Liu and Wasserman (2012) use the Kernel density or Gaussian copula estimation. Due to the high-dimensionality of the problem, we will estimate the covariance matrix using the semi-parametric extension of the Gaussian graphical model through the use of copulas. Later on, we will penalize the covariance matrix with the use of L1-regularization model graphical lasso (GLASSO), to achieve sparsity and stability in both the covariance and precision matrices. On the other hand, we will also use penalization techniques already explored such as L1-norm and L2-norm constraints in the penalization weights, finally obtaining multiple models whose performances will be compared with the base model of the classic Markowitz approach by obtaining their out-of-sample performances.

In this paper, in section 2, we show the thorough derivation and importance of the different models proposed, as well as some base models, such as Markowitz and Naïve approaches, that will work as reference to compare with the models that we are proposing. Afterwards, in section 3 we will specify model application in computer software, specifically in **R.** Then, we will use an out-of-sample real data application of these models to compare their performances using stocks included in the S&P500. Lastly, we will conclude and propose further investigation in section 4.

2 Model Formulation

It is important to mention that, from now onwards, all models will have a long-only constraint in order to make models feasible in real life applications.

2.1 *Naïve selection portfolio*

Our first and most simple selection model is the Naïve approach, where all weights have the same value $w = \frac{1}{P}$, of P risky assets considered. It is a good base model for comparison, as this model completely ignores the data, it also ignores any assumptions, estimations and optimizations that other models have to consider.

For models onwards, consider a portfolio problem with P risky-assets that have jointly normally distributed asset returns $R_1, R_2, \dots, R_p$ that follow a multivariate normal distribution with expected value P x 1 vector $\mu = [\mu_1, \mu_2, \dots, \mu_p]'$ and a P x P covariance matrix Σ.

2.2 *Markowitz Mean-Variance portfolio selection model*

The classical mean-variance model proposed by Harry Markowitz changed the way in which investors viewed their portfolios. Markowitz pioneered the idea of considering both return and risk, instead of only focusing on assets that returned the historically highest return in the market. This breakthrough discovery made investors and academics consider that it is important to acknowledge all interactions and relationships between assets. He proposed three different optimal models to solve this issue: the maximum expected return of a portfolio given a desired variance, the minimum variance portfolio given a desired expected return, and lastly, the optimal mean-variance relationship given an expected return and covariance matrix. Given that the third model is more robust, including a mean-variance relationship and a risk aversion coefficient in its objective function, we consider this as our base model from now onwards. Considering assumptions stated in the introduction of this section, the optimization problem is as follows:

$$\min_{w \epsilon \mathbb{R}^p} \frac{\gamma}{2} w'\Sigma \mathrm{w} - \mu'\mathrm{w} \qquad (1)$$
$$\text{subject to. } w'\mathbf{1} = 1$$

where $w'\Sigma \mathrm{w} = \sigma_p^2$ is the portfolio risk, $\mu'\mathrm{w}$ is the portfolio return, γ is the coefficient of relative risk aversion and $\mathbf{1} = [1,1,\dots,1]' \epsilon \mathbb{R}^p$. The constraint imposed is the budget constraint investors typically have (Markowitz 1952, Kremer et al., 2017, Das et al., 2010). An additional constraint is imposed to guarantee long-only (positive) weights in the model's output. This constraint is imposed with and additional line:

$$\mathrm{w} \geq 0$$

The first model, without long-only constraint, has a trivially derivable solution using Lagrange multipliers (Das et al., 2010):

$$w = \frac{1}{\gamma} \Sigma^{-1} \left[\mu - \left(\frac{\mathbf{1}'\Sigma^{-1}\mu - \gamma}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \right) \mathbf{1} \right] \epsilon \mathbb{R}^p \qquad (2)$$

However, when the long-only restriction is imposed, the solution becomes non-trivial and quadratic optimization must be used instead. Having this in mind, we propose two different approaches into imposing this constraint:

2.2.1 *Markowitz Mean-Variance model with quadratic optimization including long-only constraint*

We will run a quadratic optimization problem, imposing both budget and long-only restrictions as constraints into the model.

$$\min_{w \epsilon \mathbb{R}^p} \frac{\gamma}{2} w'\Sigma \mathrm{w} - \mu'\mathrm{w} \qquad (3)$$
$$\text{subject to. } w'\mathbf{1} = 1 \text{ and } w' \geq 0$$

2.2.2 *Markowitz Mean-Variance model with forced long-only constraint*

A two-step procedure, optimal weights are firstly obtained with the optimization problem without long-only constraint described in (2), then, we impose a forced long-only restriction by obliging all negative weights to 0. As in most cases this will change the budget value ($w'\mathbf{1}$): the weights are rebalanced by dividing by the absolute value of its sum (DeMiguel et al., 2009)

$$w = \frac{w}{|w'\mathbf{1}|} \tag{4}$$

2.3 *Penalized linear regression approximation of Markowitz portfolio selection model*

Considering the assumptions stated in the introduction to this section, and the Markowitz optimization model proposed in 2.2, investors are generally interested in maximizing the following mean-variance objective function, often called the utility function:

$$U(w,\mu,\Sigma) = w^T\mu - \frac{\gamma}{2}w^T\Sigma\mathrm{w} \tag{5}$$

Using optimization principles, we know that the optimal value for w is given by $\frac{\partial U(w)}{\partial w} = \mu - \gamma\Sigma\mathrm{w}^* = 0$. Yielding optimal solution, $w^* = \frac{1}{\gamma}\Sigma^{-1}\mu$. As Li (2015) proposes, the previous model allows us to formulate the initial portfolio selection problem as a linear regression problem. We have a linear regression model described by $Y = Xw + \varepsilon$, obtaining the OLS estimator of this model is given by the minimization of $(Y - Xw)^T(Y - Xw)$, so when $X^tX\widehat{w} - X^tY = 0$, remembering the optimal solution after deriving equation (5), we have:

$$\mu - \gamma\Sigma\mathrm{w}^* = 0.$$
$$X^tX\widehat{w} - X^tY = 0$$

Both of the equations have form $aw + b = 0$, so their coefficients are comparable, yielding:

$$\mu = X^tY$$
$$X^tX = \gamma\Sigma$$

It has been proven that in the Markowitz problem, Σ is a positive semi-definite matrix, meaning that all of its eigen values are non-negative and so its square root ($\Sigma^{\frac{1}{2}}$) is estimable. Considering this property, we can now solve for X and Y, our linear regression inputs:

$$X = \sqrt{\gamma}\Sigma^{\frac{1}{2}} \tag{6}$$

$$Y = X^{-1}\mu = \frac{1}{\sqrt{\gamma}}\Sigma^{-\frac{1}{2}}\mu \tag{7}$$

We have obtained the two necessary inputs to solve the portfolio selection problem as a linear regression problem, where the OLS estimators obtained, after solving the linear regression model, will represent the weights assigned to each corresponding stock. As mentioned before, portfolio selection problems have the characteristic of being high dimensional, with high correlation between their variables. When these two characteristics are present in the same data, there is a high probability that any data column might be nearly a linear combination of the other columns, meaning that there will be multicollinearity within the data. Considering that we now have a linear regression model, multicollinearity has proven to seriously increase the regression estimators' variances, leading to inadequate and unstable portfolio weights. For example, when returns of asset i and j are highly correlated, the precision matrix Σ^{-1} is highly volatile, containing extreme entries in positions $\Sigma^{-1}_{i,i}, \Sigma^{-1}_{j,j}, \Sigma^{-1}_{i,j}$ *and* $\Sigma^{-1}_{j,i}$, that variate dramatically over time (Ledoit and Wolf, 2003).

A common solution to get rid of multicollinearity in linear regression problems is to penalize the regression estimators. This method shrinks the regression estimators most when the estimator has a variable containing high correlation. While the penalized estimated portfolio weights are no longer unbiased, their variances and mean squared prediction errors are reduced (Li, 2015). Also, Fan, Zhang and Yu (2012) derived an inequality to measure the estimation risk for the portfolio problem shown in (5):

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

where $U(w,\hat{\mu},\hat{\Sigma}) = w^T\hat{\mu} - \frac{\gamma}{2}w^T\hat{\Sigma}\mathrm{w}$, $||.||_\infty$ are the component-wise estimation error for both μ and Σ, and $||w||_1 = \sum_{i=1}^N |w_i|$ is the L1-norm vector w. In this inequality, it is evident that the L1-norm constraint reduces the estimation risk of μ and Σ, while the L2-norm constraint will reduce the estimation risk of Σ and Σ^{-1}(Fan, Zhang and Yu, 2012). Hence, we will apply to the regression model, L2 and L1-norm penalization models to obtain portfolio weights with these improved characteristics.

2.3.1 *Improved linear regression estimators through L2-norm penalization (Ridge)*

The L2-norm penalization, also called Ridge regression penalization, penalizes the sum of squared coefficients with the following constraint:

$$\widehat{w}_{L2} = \underset{w}{\operatorname{argmax}}\left\{w^T\mu - \frac{\gamma}{2}w^T\Sigma\mathrm{w}\right\} \quad \text{subject to. } ||w||_2^2 < s_0 \tag{8}$$

where $s_0 > 0$ is a constant and $||.||_2^2 = \sqrt{\sum_{i=1}^N w_i{}^2}$ is the Euclidean or squared L2 norm of a vector. Using the linear regression approximation proposed before, where $X = \sqrt{\gamma}\Sigma^{\frac{1}{2}}$

and $Y = \frac{1}{\sqrt{\gamma}}\Sigma^{-\frac{1}{2}}\mu$ and using Lagrange multipliers, we derive the following equation:

$$\widehat{w}_{L2} = arg\min_{w}\{\|Y - Xw\|_2^2 + \lambda_1\|w\|_2^2\} \tag{9}$$

where $\lambda_1 > 0$. Finally, we obtain the OLS solution extension with L2-norm penalization. Note that when $\lambda_1 = 0$, we obtain the OLS estimator, but as λ_1 increases, portfolio weights are further shrunken towards zero, becoming increasingly biased results with less variance. A key characteristic of this model is that even if λ_1 is large, portfolio weights will not reach zero.

2.3.2 Improved linear regression estimators through L1-norm penalization (Lasso)

The L1-norm penalization, also known as Lasso regression penalization, penalizes the sum of absolute values of the coefficients by imposing the following constraint:

$$\widehat{w}_{L1} = \underset{w}{\operatorname{argmax}}\left\{w^T\mu - \frac{\gamma}{2}w^T\Sigma w\right\} \quad \text{subject to. } \|w\|_1 < s_1 \tag{10}$$

where $s_1 > 0$ is a constant and $\|w\|_1 = \sum_{i=1}^N |w_i|$ is the L1 norm of vector w. Using the linear regression approximation proposed before, where $X = \sqrt{\gamma}\Sigma^{\frac{1}{2}}$ and $Y = \frac{1}{\sqrt{\gamma}}\Sigma^{-\frac{1}{2}}\mu$ and using Lagrange multipliers, we derive the following:

$$\widehat{w}_{L1} = arg\min_{w}\{\|Y - Xw\|_2^2 + \lambda_2\|w\|_1\} \tag{11}$$

where $\lambda_2 > 0$. Similar to L2-norm, when $\lambda_2 = 0$, we obtain the OLS estimator, and as penalization parameter λ_2 increases, portfolio weights will be shrunken, but in this case, some weights will be exactly zero. Having a subset of assets with zero weights makes the portfolio rule sparse. Sparsity is a desired characteristic in these problems: firstly, they reduce transaction and portfolio management costs, secondly, as Li (2015) states, "since the number of historical asset returns is relatively small compared with the number of assets, estimation error is large. By setting a subset of small portfolio weights to zero, the estimated portfolio weights are no longer unbiased, but their variances and the mean squared prediction errors could be reduced". This is also known as bias-variance tradeoff.

Obviously, for both penalization models *2.3.1* and *2.3.2*, the variance-bias tradeoff has to be considered to determine how much the weights must be shrunken. Hence, for each model, its penalization parameter λ must be calibrated using techniques such as k-folds or Cross-Validation, where the objective is to minimize the Mean Squared Error (MSE) or some other similar criterion.

2.4 Covariance and precision matrix estimation using GLASSO penalized semi-parametric extension of the Gaussian graphical model, through the use of copulas.

In model 2.3.2, we achieved sparsity in the regression estimators. Nonetheless, stability is also a desired statistical property (Li, 2015). Stability is achieved when there is a reduction in estimation errors of the covariance matrix Σ and its inverse, which we will refer to from now onwards as the precision matrix Σ^{-1}. The Markowitz model has been proven to be unstable (Xing, H, Lai, T, & Chen, Z, 2010; Li, 2015), extreme positions are usually observed and the solution changes dramatically when new information is inputted to the model, meaning that the model is fully dependent on the historical data given. We will attempt to achieve a stable portfolio by making the model less dependent on the input data. This can be achieved by estimating the model's parameters with a semi-parametric approach. As in these models, stability depends on the covariance and precision matrices, these are the ones we will attempt to estimate semi parametrically.

Lafferty, Liu and Wasserman (2012) propose a semi-parametric estimation of the covariance and precision matrices through the use of arbitrary graphical models, making a distributional restriction imposed with the use of copulas. They mention that this approach is especially suitable for high-dimensional data as it also exploits sparsity.

The graph $G = (V, E)$ of a random vector $X = (X_1, X_2, \dots, X_d)$ with distribution P, consists of vertex set V containing d elements (one for each variable) and edge set E. Edge between (i,j) is excluded of set E if and only if X_i is independent of X_j, given the other variables $(X_s: 1 \le s \le d, s \ne i, j)$. Thus, the probability density inside an undirected graph G is given by:

$$p(x) = \frac{1}{Z(f)}\exp\left(\sum_{C \in Cliques(G)} f_C(x_C)\right) \tag{12}$$

where the sum is over all fully connected subsets of vertices (cliques) of graph G. Hence, they obtain a nonparametric graphical model, that is very general. In order to maintain its flexibility and nonparametric spirit, the random variable $X = (X_1, X_2, \dots, X_d)$ is replaced by a random variable $f(X) = (f_1(X_1), f_2(X_2), \dots, f_d(X_d))$, and assuming that $f(X)$ is multivariate Gaussian. This is a nonparametric extension of the Normal, the nonparanormal distribution. This distribution depends on the univariate functions $f_i\ \forall\, i \in \{1, \dots, d\}$, mean μ and covariance matrix Σ. The univariate marginals are estimated using the copula approach $f_j(x) = \mu_j + \sigma_j\Phi^{-1}\left(F_j(x)\right)$ where F_j is the distribution function for variable X_j. After each f_j is estimated, they transform to the assumed jointly Normal distribution by $Z = (f_1(X_1), f_2(X_2), \dots, f_d(X_d))$, finally applying methods for gaussian graphical models to estimate the graph. With this estimation, they achieve fully nonparametric univariate marginals, that are regulated with the use of graphical lasso (GLASSO) to finally obtain sparse covariance and precision matrices.

2.4.1 Covariance matrix estimation using the semi-parametric extension of the Gaussian, through the use of copulas.

We will show the two-step procedure proposed by Lafferty, Liu and Wasserman (2012) to estimate the aforementioned covariance and precision matrices:

Let $X^{(1)}, X^{(2)}, \dots, X^{(n)}$ be a sample of n rows, where $X^{(i)} = \left(X_1^{(i)}, X_2^{(i)}, \dots, X_d^{(i)}\right)' \in \mathbb{R}^d$.

They first desire to estimate the univariate marginals for each asset namely X_j, firstly, the distribution function for each asset must be found, a natural candidate for this is the marginal empirical distribution:

$$\hat{F}_j(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\left\{X_j^{(i)} \le t\right\}} \tag{13}$$

As the model is assumed Gaussian, the normal score of the marginal empirical distribution is obtained:

$$\hat{h}_j(x) = \Phi^{-1}\left(\hat{F}_j(x)\right) \tag{14}$$

But, considering that the normal scores approach negative and positive infinity with the smallest and largest values of $X_j^{(i)}$, the following Winzorized truncation must be performed before hand:

$$\tilde{F}_j(x) = \begin{cases} \delta_n, & \hat{F}_j(x) < \delta_n \\ \hat{F}_j(x), & \delta_n \le \hat{F}_j(x) \le (1-\delta_n) \\ (1-\delta_n), & \hat{F}_j(x) > (1-\delta_n) \end{cases} \tag{15}$$

Where δ_n is the truncation parameter, whose recommended value is:

$$\delta_n = \frac{1}{4n^{\frac{1}{4}}\sqrt{\pi \log n}}$$

The derivation of the truncation parameter value will not be shown, as it is outside of this study's scope. Now with the truncated distribution, the normal score can be obtained with:

$$\tilde{h}_j(x) = \Phi^{-1}\left(\tilde{F}_j(x)\right) \tag{16}$$

After calculating the estimate of the distribution of X_j, the transformation function f_j is derived:

$$\tilde{f}_j(x) = \hat{\mu}_j + \hat{\sigma}_j \tilde{h}_j(x) \tag{17}$$

were $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_j^{(i)}$ and $\hat{\sigma}_j = \sqrt{\frac{1}{n}\left(X_j^{(i)} - \hat{\mu}_j\right)^2}$ are the sample mean and standard deviation, respectively.

Finally, they obtain the sample covariance matrix of $\tilde{f}\left(X^{(1)}\right), \tilde{f}\left(X^{(2)}\right), \dots, \tilde{f}\left(X^{(n)}\right)$, namely $S_n(\tilde{f})$:

$$S_n(\tilde{f}) = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{f}(X^{(i)}) - \mu_n(\tilde{f})\right) \cdot \left(\tilde{f}(X^{(i)}) - \mu_n(\tilde{f})\right)' \tag{18}$$

$$where\ \mu_n(\tilde{f}) = \frac{1}{n}\sum_{i=1}^{n}\tilde{f}(X^{(i)})$$

The result in (18) was then used to estimate the precision matrix. While the maximum likelihood estimator is the direct inverse $\widehat{\Omega}_{\mathrm{n}}^{MLE} = S_n(\tilde{f})^{-1}$, to achieve sparsity and have more stable matrices, the GLASSO algorithm will come in handy.

2.4.2 Covariance and precision matrix estimation with Graphical Lasso Optimization (GLASSO)

As mentioned in the introduction to this section, in this model we are assuming a multivariate Gaussian distribution $\mathcal{N}_d(\mu, \Sigma)$. Graphical models shape multivariate distributions, representing its random variables as nodes. Specifically, in this case, we will use an undirected Gaussian graphical model (UGGM), $G = (V, E)$, where $V = \{1, 2, \dots, d\}$ represents set of variables and E is the set of undirected edges. The graph G represents the model where the precision or concentration matrix $\Omega = \Sigma^{-1}$ is a positive semi-definite matrix with $\Omega_{uv} = 0$ whenever there is no edge between vertices u and v in G (Højsgaard, Edwards, & Lauritzen, 2012, p.83). Hence, the partial correlation between two variables can be derived from the precision matrix:

$$\rho_{uv}|V \setminus \{u, v\} = \frac{-\Omega_{uv}}{\sqrt{\Omega_{uu}\Omega_{vv}}} \tag{19}$$

Thus, from the equation above, $\Omega_{uv} = 0$, if and only if these two variables are conditionally independent given all other variables (Højsgaard, Edwards, & Lauritzen, 2012, p.78). This model will help us to achieve a conditional independent problem, desired for high correlation problems such as the portfolio selection problem. They proceed to estimate the precision matrix through maximum likelihood estimator; considering the sample data described in the introduction to model 2.4 and an empirical covariance matrix **S**, the log-likelihood function is:

$$\widehat{\Omega} = Log\ \mathcal{L}(\Omega,\mu) = \frac{n}{2}\log\det(\Omega) - \frac{n}{2}tr(\Omega S) - \frac{n}{2}(\bar{X}-\mu)'K(\bar{X}-\mu) \quad (20)$$

For a fixed K, this function is maximized when $\hat{\mu} = \bar{X}$, after replacing this result in equation (20), we find that the last term is equal to 0:

$$\frac{n}{2}(\bar{X}-\bar{X},)'K(\bar{X}-\bar{X}) = 0 \quad (21)$$

Obtaining the following log-likelihood equation:

$$\widehat{\Omega} = Log\ \mathcal{L}(\Omega,\mu) = \frac{n}{2}\log\det(\Omega) - \frac{n}{2}tr(\Omega S) \quad (22)$$

The solution to this log-likelihood equation is obtained iteratively, for example, by using IPS (Iterative proportional scaling). To avoid this step-wise search and also to obtain sparse and stable matrices, Højsgaard, Edwards, & Lauritzen (2012) recommend to apply L1-norm penalization to the model, using a simple and remarkably fast algorithm known as the Graphical Lasso approach, also known as, GLASSO created by Friedman et al. (2007). First the log-likelihood function in (22) is penalized:

$$\widehat{\Omega}_{pen} = \underset{\Omega}{\operatorname{argmax}}\{\log\det(\Omega) - tr(\Omega S) - \rho||\Omega||_1\} \quad (23)$$

where $\widehat{\Omega}_{pen}$ is the estimated penalized precision matrix and ρ is a non-negative penalization parameter that can be calibrated (Friedman, Hastie, & Tibshirani, 2007). The sample covariance matrix **S** is replaced by the covariance matrix $S_n(\tilde{f})$ estimated in (18). Deriving the following GLASSO function:

$$\widehat{\Omega}_{pen} = \underset{\Omega}{\operatorname{argmin}}\{tr\left(\Omega S_n(\tilde{f})\right) - \log|\Omega| + \rho||\Omega||_1\} \quad (24)$$

where $||\Omega||_1 = \sum_{j=1}^{d}\sum_{k=1}^{d}|\Omega_{jk}|$. Friedman et al. (2007) mention that Banerjee et al. (2007) show that the problem in (24) is convex and can be estimated; this is equivalent to solving the following problem:

$$\min_{\beta}\{\frac{1}{2}\left\|W_{11}^{1/2}\beta - W_{11}^{-1/2}S_n(\tilde{f})_{12}\right\|^2 + \rho\|\beta\|_1\} \quad (25)$$

where W is the estimate of Σ and $\hat{\beta}$ is the vector of estimators for the model. Finally, the algorithm of implementation to obtain the solution is as follows. Notice that I is the identity matrix:

Graphical Lasso Algorithm

1. Start with $W = S_n(\tilde{f}) + \rho I$. The diagonal of W remains unchanged in what follows.
2. For each $j = 1,2,\dots p$ solve the LASSO problem in (25), which take as an input the inner products W_{11} and $S_n(\tilde{f})_{12}$. This gives a $p-1$ vector solution $\hat{\beta}$. Fill in the corresponding row and column of W using $w_{12} = W_{11}\hat{\beta}$.
3. Continue until convergence; when the average absolute change in W is less than $0.001 \cdot avg\left|S_n(\tilde{f})^{-diag}\right|$, where $\left|S_n(\tilde{f})^{-diag}\right|$ are the off-diagonal elements of $S_n(\tilde{f})$.

Note that this algorithm yields both penalized covariance and precision matrices. Due to penalization in both, sparsity and stability is induced into the model, decreasing estimation error of both uncertain parameters Σ and Σ^{-1}.

3 EMPIRICAL STUDIES

3.1 *Data selection*

To demonstrate an empirical application to the models, it is important to consider a market that at least complies with a semi-strong form within the definition of the efficient markets hypothesis (EMH) proposed by Eugene Fama in 1970. Thus, we decided to use the Standard and Poor's 500 index (S&P500), a popular index that operates in highly transacted markets such as the NYSE and NASDAQ. This renowned index has existed since 1957, but due to the fact that it consistently changes its 500 companies and that many of today's companies are from the 21st century, we chose to analyze prices of these stocks on an 11-year span, from 2008 to 2019. Not all 500 companies listed in today's S&P500 have consistent price data since 2008, hence, we removed these assets, finally having a pool of 448 assets to consider. On the other hand, these models require the coefficient of relative risk aversion as an input. Commonly accepted measures lie between 1 and 3. We choose to fix this coefficient to $\gamma = 2.5$ since estimating this value is outside the scope of this study. Also, as the standard, the annual risk-free rate to consider $R_f = 2.31\%$ was obtained by averaging the last year U.S. government Treasury Bill rates (Macrotrends.net, 2019).

3.2 *Model implementation*

Considering the data mentioned in section 3.1, the following model implementation was performed with the use of **R** software. The code is of public domain, it can be accessed in the following repo: https://github.com/emcediel/Improved-Portfolio-Selection-Paper-U-Andes under the name: Improved portfolio selection code.R.

3.2.1 *Data acquisition and parameter estimation*

To obtain 2008-2019 price data of the 500 stocks, we used the `quantmod` library in **R**. After removing all stocks with missing data, we obtained the complete 448 stock data. The logarithmic returns were calculated for each asset using the following equation:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (26)$$

To evaluate the model's out-of-sample performances, we chose to separate the data into two sets: a training set containing logarithmic returns from 2008 to 2018 and a testing set containing logarithmic returns from 2018 to 2019 data, specifically one trading year or 252 days of data.

For all models, except naïve selection, we used the train data to calculate the estimated mean return, the sample and semi-parametric covariance matrices, both their inverse (precision) matrices by using both the traditional inverse and the GLASSO approach inverse. The estimated mean return was calculated using the following:

$$\hat{\mu} = \frac{1}{n}\mathbf{1}'r \tag{27}$$

where r is the N x P matrix of logarithmic returns. The sample covariance matrix was estimated with the following equation:

$$\hat{S} = \frac{1}{n-1}X_C'X_C \tag{28}$$

where $X_C = (I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n')r$.

The semi-parametric covariance matrix $\widehat{S_n(f)}$ was estimated using the logarithmic returns matrix r and a created function `copula` that implements with **R** functions the algorithm shown in section 2.4.1. The traditional inverse for the sample covariance matrix $\hat{S}^{-1}$was calculated using `solve()` function in **R.** Finally, the GLASSO approach precision matrices for both the sample covariance matrix $\hat{\Omega}_{\hat{S}}$ and the semi-parametric copula covariance matrix $\hat{\Omega}_{\widehat{S_n(f)}}$ were estimated using the `CVglasso` library in **R**. In the documentation of function `CVglasso`, the author Matt Galloway specifies the same objective function as the one stated in equation (24) in section 2.4.2 (Galloway, 2019). Additionally, this function estimates 10 possible penalization parameters and outputs the optimal within these by performing Cross-Validation. Both estimations require an additional parameter `lam.min.ratio = 1e-3,` as the default `1e-2` results in an optimal tuning parameter on boundary. Note that this algorithm returns both penalized covariance and precision matrices, so additionally we obtain new sparse covariance matrices, $\hat{S}_{pen}$ and $\widehat{S_n(f)}_{pen}$ respectively.

In the next section, we will compare the performance of the four different models. Furthermore, within each model, we will compare the performance of the three estimated covariance and precision matrices pairs mentioned before, namely, $(\hat{S}, \hat{S}^{-1})$, $(\hat{S}_{pen}, \hat{\Omega}_{\hat{S}})$ and $(\widehat{S_n(f)}_{pen}, \hat{\Omega}_{\widehat{S_n(f)}})$.

3.2.2 *Naïve selection portfolio*

For this model, we calculated all weights using $w = \frac{1}{P}$, as $P = 448$, each weight is $w = 0.00223$, finally obtaining a 448 x 1 matrix with $w = 0.00223$ in each position.

3.2.3 *Markowitz Mean-Variance model with quadratic optimization including long-only constraint*

The Markowitz mean-variance model with quadratic optimization was implemented using the `solve.QP` of the `quadprog` library. This function minimizes quadratic optimization problems of the form $-d'b + \frac{1}{2}b'Db$ with constraints $A'b \geq b_0$ (Turlach & Weingessel, 2019). Considering the optimization problem in (3), the mapping is as follows:

$$d = r$$

$$D = \gamma\hat{\Sigma}$$

$$A = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

that is, a P+1 x P matrix, having the first row with 1's, and below a P x P identity matrix. Considering A, b_0 must be:

$$b_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

that is, a P+1 vector, with the first value equal to 1 followed by P zeros. Finally, function parameter `meq` indicates how many of the first meq constraints are equality statements instead of inequality statements. As we know, only the first constraint (budget constraint) is an equality, the long only constraints are inequalities, thus, `meq=1` (Turlach & Weingessel, 2019). A function, `traditionalMarkowitzOptimization` was created to receive as inputs a coefficient of relative risk aversion, a covariance matrix and a mean return vector, the function runs the optimization problem described in (3) with the function described above and outputs the optimal weights vector b.

3.2.4 *Markowitz Mean-Variance model with forced long-only constraint*

We created a function in **R**, `traditionalMarkowitzForced` that calculates the optimal weights using the analytical solution found in (2), afterwards, it forces the negative weights in the solution to 0, finally rebalancing the weights to comply with the budget constraint using (4).

3.2.5 *Improved linear regression estimators through L2 and L1 norm penalization (Ridge and Lasso)*

To implement the L2 and L1 norm penalized regression models, we created a generic **R** function `markowitzRegressionPenalization` with a binary parameter alpha, where $\alpha = 0$ runs a Ridge Regression model, while, on the other hand $\alpha = 1$ runs a Lasso Regression model. Additionally, the other inputs are: the coefficient of relative risk aversion, a covariance matrix, a precision matrix, a mean return vector and a parameter containing the number of folds desired for the Cross-Validation algorithm. Firstly, we use the linear regression approximation of the Markowitz models by finding the linear regression parameters using (6) and (7). Using **R** library `glmnet` and function `cv.glmnet`, we run a cross-validation algorithm with $K = \frac{P}{3}$ folds to find the best penalization parameter λ_1 or λ_2 depending on alpha being 0 or 1 respectively. Finally, using `glmnet` function, we use the λ_1 or λ_2 parameters found beforehand to find the optimal analytical solutions described in (9) and (11) for Ridge and Lasso regressions, respectively. To ensure that the solution adheres to the model restrictions, we add to both `glmnet` functions beforehand an additional parameter `lower.limits = 0` to ensure that solution weights are long-only. Also, after the optimal solution is obtained, we use (4) to obey budget restrictions.

3.3 Performance measures

Firstly, we would like to emphasize on the different models that we will perform and compare. As shown, we have the basic traditional models to use as benchmark for our comparison: The Naïve method and both the Markowitz mean-variance models proposed. The models that have been enhanced and improved with the use of sparse statistics are the penalization linear regression models, and the Ridge and Lasso approaches. Additionally, we propose the improvement in the estimation of the covariance and precision matrices using a semi-parametric approach and sparse statistics. The matrix pair that we propose consists on the penalized copula semi-parametric covariance matrix $\widehat{S_n(\hat{f})}_{pen}$ and its inverse precision matrix $\widehat{\Omega}_{\widehat{S_n(\hat{f})}}$, both estimated through penalized GLASSO optimization. To achieve a thorough performance measure of the improvement given by the proposed pair, we would like to compare this model's performance with two other benchmark pairs. The first pair is the commonly used standard sample covariance matrix $\hat{S}$ and its direct inverse precision matrix $\hat{S}^{-1}$. Due to the fact that the proposed matrix pair includes two improvements, the semi-parametric copula estimation of the covariance matrix and the GLASSO optimization algorithm, it would be desirable to create a second benchmark pair that only includes GLASSO optimization, enabling us to get a better understanding of the marginal improvements given by the GLASSO penalization and the semi-parametric estimation of the covariance matrix. Hence, we propose a second benchmark pair consisting of the penalized standard sample covariance matrix $\hat{S}_{pen}$ and its inverse precision matrix $\widehat{\Omega}_{\hat{S}}$, both estimated through penalized GLASSO optimization. Thus, we finally obtain 4 different models with 3 different inputs for each and the Naïve model, yielding a total of 13 different models to compare.

Furthermore, to prove that sparse statistic techniques used in this investigation yield better results than the traditional and quotidian models, we require model validation measures. Using the test set that consists on 252 days of data, we will use the Cumulative Wealth Index (CWI), the annualized Sharpe Ratio (SR), the Information Ratio (IR) and the Calmar Ratio (CR) to compare between the 12 models and conclude which has better characteristics and yield better out-of-sample results. Most of the following performance measures were implemented in **R** with package `PerformanceAnalytics.`

3.3.1 Cumulative Wealth Index (CWI)

The Cumulative Wealth Index is a standard graphical measure of comparison between models. The out-of-sample performance of each portfolio is evaluated by plotting the cumulative wealth obtained in a period of time if an investor invested 1 USD in that portfolio at time $t = 0$. The measure must be estimated in a specific time horizon from time 0 to time T, using the following equation:

$$CW_t = CW_{t-1}(1 + R_t) \quad \forall t \in \{1, \dots, T\} \qquad (29)$$
$$CW_0 = 1\ USD$$

where $R_t = w'r_t$, r_t is the row vector in time t within the matrix of daily logarithmic returns of the test set and w' is the transposed vector of portfolio weights to be evaluated. This measure assumes that all portfolios to compare have similar or equal mean return and mean volatility, hence, similar risk aversion coefficients. While theoretically we are comparing portfolios that have the same coefficient of relative risk aversion $\gamma = 2.5$, when we measure these portfolios in out-of-sample conditions, they will show differences in their mean return and mean volatility. Thus, we will explore other more robust measures for out-of-sample performance such as the Sharpe Ratio, the Information Ratio or the Calmar Ratio that consider these inherent characteristics.

3.3.2 Annualized Sharpe Ratio (SR)

The Sharpe Ratio developed by Nobel laureate William F Sharpe is used to understand the return of an investment compared to its risk. More specifically, the ratio measures the average return earned in excess of the risk-free rate per unit of volatility (risk). Hence, a portfolio that yields a higher SR is more attractive. The Sharpe Ratio can be calculated as follows:

$$Annualized\ SR = \frac{\frac{252}{T}\sum_{t=1}^{T}(R_t - Rf)}{\sigma_p * \sqrt{252}} \qquad (30)$$

where Rf is the daily risk-free rate and σ_p is the daily standard deviation of the excess return $(R_t - Rf)\ \forall t \in \{1, \dots, T\}$. This measure is more robust than CWI, as it considers the new mean returns and volatility for each portfolio, enabling all portfolios to become comparable.

3.3.3 Information Ratio (IR)

The Information Ratio (IR) measures the portfolio excess returns using an index as benchmark, compared to the volatility of those returns. This measure compares directly its performance against a benchmark and gives information regarding the consistency of the portfolio's performance throughout the year with the use of the Tracking Error.

$$Annualized\ IR = \frac{\frac{252}{T}\sum_{t=1}^{T}(R_t - RI_t)}{Tracking\ Error * \sqrt{252}} \quad (31)$$

where in equation (31) RI_t is the daily return rate of the benchmark index in time t and $Tracking\ Error$ is the daily standard deviation of the excess return $(R_t - RI_t)\ \forall t \in \{1, \dots, T\}$. Although similar to SR in its risk-return approach, its main difference is the benchmark. While SR uses a risk-free investment, IR uses an index or portfolio. In this case we will use as benchmark the S&P500 index. Always, the higher the IR index, the more attractive the portfolio. Also, positive IR show average excess returns over the benchmark, while negative IR show average loss over the benchmark.

3.3.4 Calmar Ratio (CR)

Developed by Terry W. Young in 1991, the Calmar Ratio (CR) is a measure of the annual average rate of return relative to the maximum drawdown risk in the same period. It allows investors to see potential opportunity but also potential loss risk. The ratio is calculated as follows:

$$Annualized\ CR = \frac{\frac{252}{T}\sum_{t=1}^{T}(R_t)}{Maximum\ Drawdown} \quad (32)$$

Where $Maximum\ Drawdown = \frac{Equity\ Peak\ High - Equity\ Trough}{Equity\ Peak\ High}$ represents in a specified period, the largest drop in equity, from the peak to the trough. Calmar Ratio above one is desirable; a portfolio with higher CR is more attractive.

3.4 Results

After implementing all 4 models, estimating the semi-parametric copula covariance matrix and obtaining the penalized covariance and precision matrices using the GLASSO algorithm, we parametrize each of the four models with the three pair combinations of covariance and precision matrices, finally obtaining 12 portfolios represented in weight vectors, to evaluate. From now onwards, we will distinguish between the three pairs of covariance and precision matrix parameters with the following abbreviations: Sample covariance matrix and standard precision matrix pair $(\hat{S}, \hat{S}^{-1})$ will be denoted **SS pair**, the penalized standard precision matrix and the penalized GLASSO precision matrix pair $(\hat{S}_{pen}, \hat{\Omega}_{\hat{S}})$ will be abbreviated to **GLASSO pair** and the penalized semi-parametric copula covariance matrix and the penalized GLASSO precision matrix pair $\left(\widehat{S_n(\tilde{f})}_{pen}, \hat{\Omega}_{\widehat{S_n(\tilde{f})}}\right)$ will be mentioned as **Copula GLASSO pair**. Note that all figures presented below are shown with more detail in Appendix A.

In **Figure 1** we can observe the CWI of the 12 portfolios obtained. The colors represent each model while the line type represent the input pair of matrices used. There are noticeable differences between inputs and also between models: each model has a clear range and pattern of CWI. The Markowitz Optimization model performed poorly, having a similar line to the Naïve approach with additional apparent high volatility, the Markowitz Forced model has a slight improvement in both CWI and apparent volatility. The Ridge and Lasso Regression models have an evident improvement relative to the benchmark Naïve and Markowitz models; at the end of the year (last 52 days), they appear to be in completely different range of values, while benchmark portfolios are within a 1.009 and 1.130 CWI range, the penalization models are within a 1.127 and 1.216 CWI range. Hence, while maintaining a relatively low volatility, they show higher CWI in the long-run. From **Figures 2-5** we explore in each, one of the four models, visualizing more clearly the differences within the three covariance and precision matrix pair inputs, these are referenced by color. In **Figure 2** the CWI of the Markowitz Forced models show a slight improvement relative to benchmark Naïve model. Conversely, between the input pairs, there is a barely noticeable improvement given by both the GLASSO and Copula GLASSO pairs relative to the SS pair. In **Figure 3**, we observe the CWI of the Markowitz Optimization models. As mentioned before, these portfolios perform similarly to the Naïve model, furthermore, they seem to have more volatility than the benchmark, the Copula GLASSO model has a better average performance with less apparent volatility than the other models, GLASSO and SS pair perform similarly. Now in **Figure 4**, we observe a clear improvement given by the Ridge Regression relative to the benchmark, this improvement aggrandizes as time passes. While similar performances are given by the three pairs, GLASSO and Copula GLASSO have a slight advantage. Finally, we evaluate the sparse Lasso Regression portfolios in **Figure 5.** This is clearly the best set of portfolios, having relatively low variance and high return in the long run. Specifically, there's an evident difference between the Copula GLASSO and the benchmark pairs, the Copula GLASSO implemented in the Lasso model outperforms all other portfolios in CWI.

While CWI results show that sparse models outperform the benchmark models, this measure, when used with out-of-sample data has a flaw, it doesn't consider volatility of the portfolios at hand, we might be comparing two very different portfolios in terms of risk, making the comparison biased and ultimately unfair. The average investor would not base his decisions only on CWI performances, they rather use measures that consider a ratio between return and risk. Considering this, we use three widely used performance measures that solve the issue at hand, these are the Sharpe Ratio, Calmar Ratio and Information Ratio.

The results obtained in the annualized Sharpe Ratio by applying equation (30) and considering an annual risk-free rate of 2.31% (average T-Bill rate of last year) were plotted in **Figure 6**. It is evident that penalized models outperformed the benchmark, while Markowitz Optimization had lacking results with the best at 0.437, Ridge Regression had its best at 0.99, meaning that the latter has more than double average return in excess of the risk-free rate per unit of risk compared to the former. Also, input pairs had evident improvements, the best SR for 3 out of 4 models

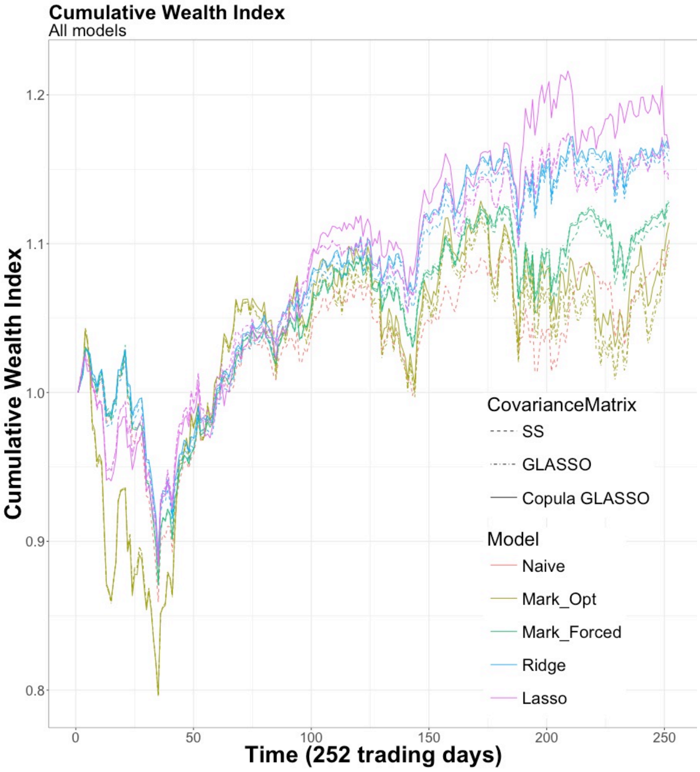

Figure 1: Cumulative Wealth Index (CWI) of all 12 models and Naïve model, with initial wealth of $1.

Figure 3: Cumulative Wealth Index of three Markowitz mean-variance Optimization Models, with different covariance precision pair inputs (index), relative to benchmark Naïve (red), all with initial wealth of $1.

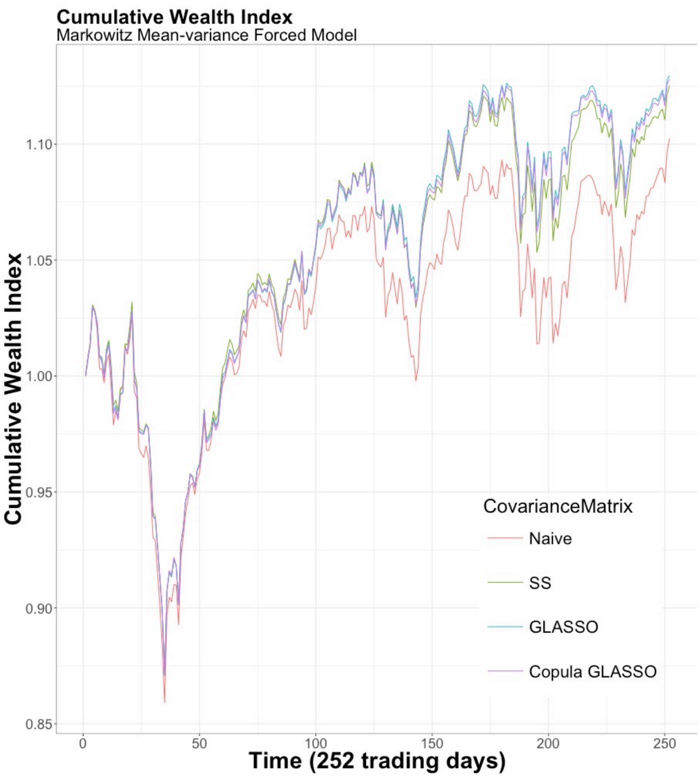

Figure 2: Cumulative Wealth Index (CWI) of three Markowitz mean-variance Forced Models, with different covariance precision pair inputs (index), relative to benchmark Naïve (red), all with initial wealth of $1.

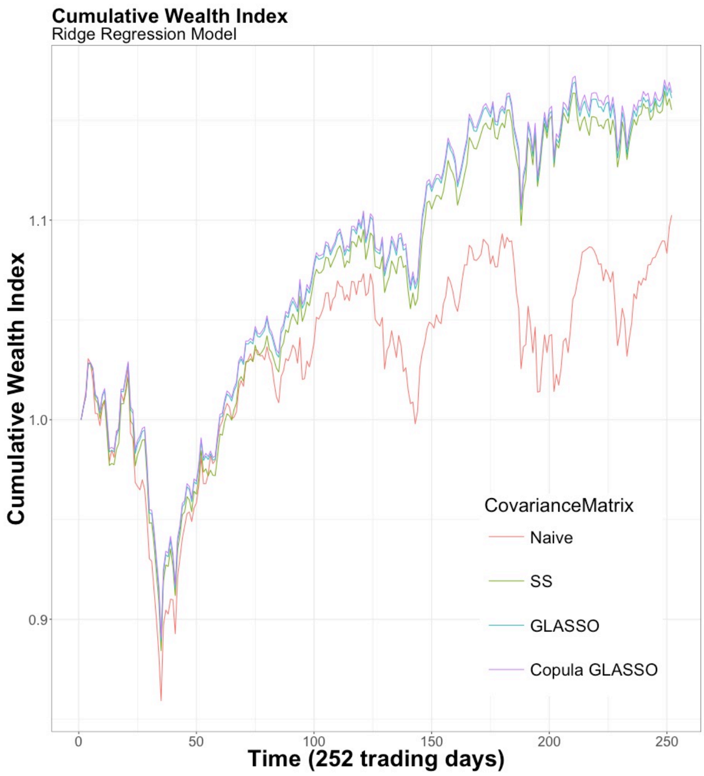

Figure 4: Cumulative Wealth Index (CWI) of three Ridge Regression Models, each having different covariance precision pair inputs (index), relative to benchmark Naïve (red), all with initial wealth of $1.

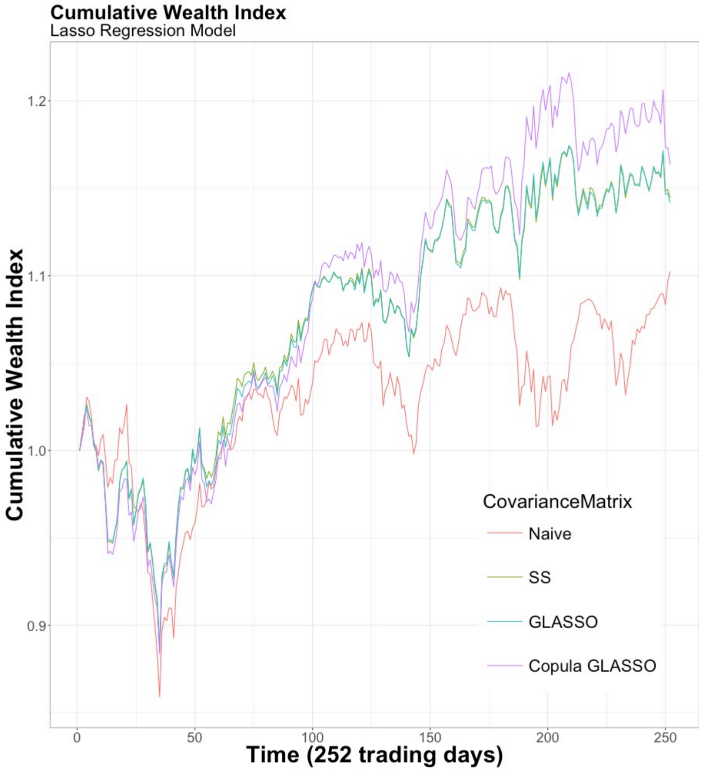

Figure 5: Cumulative Wealth Index (CWI) of three Lasso Regression Models, each having different covariance precision pair inputs (index), relative to benchmark Naïve (red), all with initial wealth of $1.

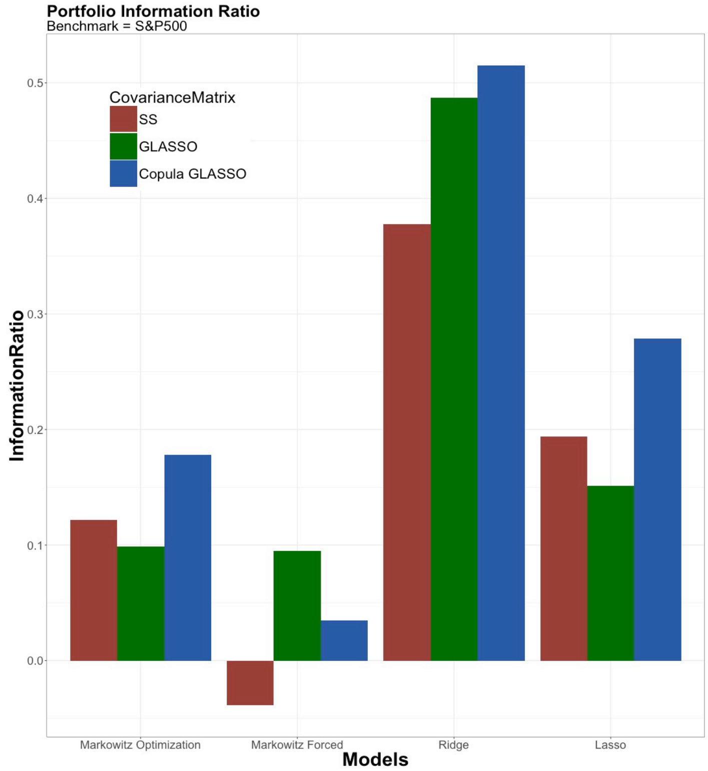

Figure 7: Annualized Information Ratio (IR) with S&P 500 as benchmark. Differentiation is given by model and by covariance precision pair inputs (index).

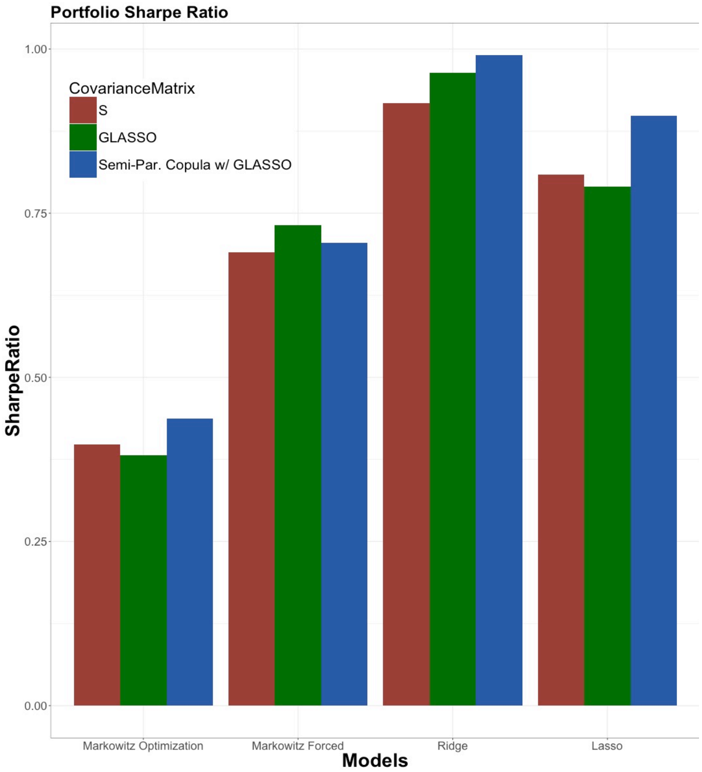

Figure 6: Annualized Sharpe Ratio (SR). Differentiation is given by model and by covariance precision pair inputs (index).

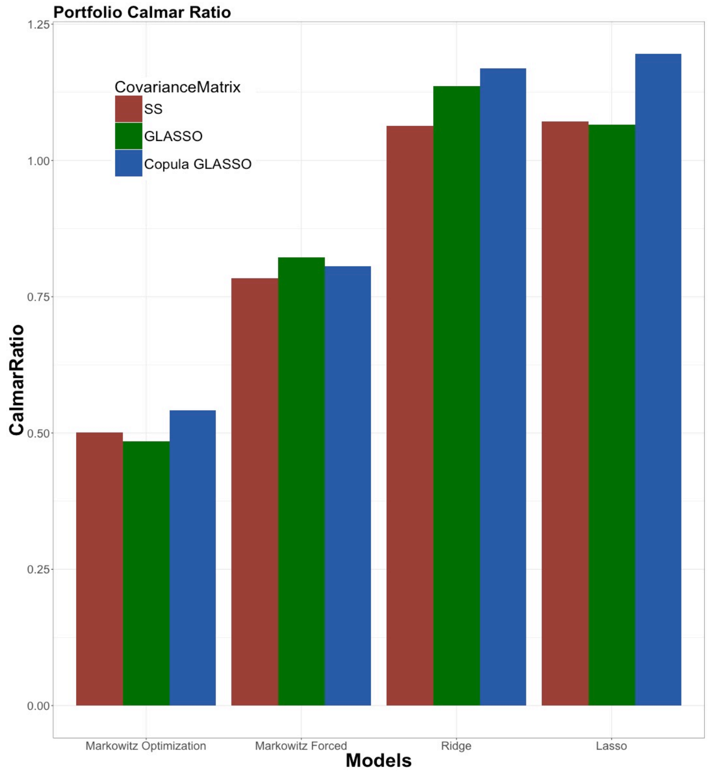

Figure 8: Annualized Calmar Ratio (CR). Differentiation is given by model and by covariance precision pair inputs (index).

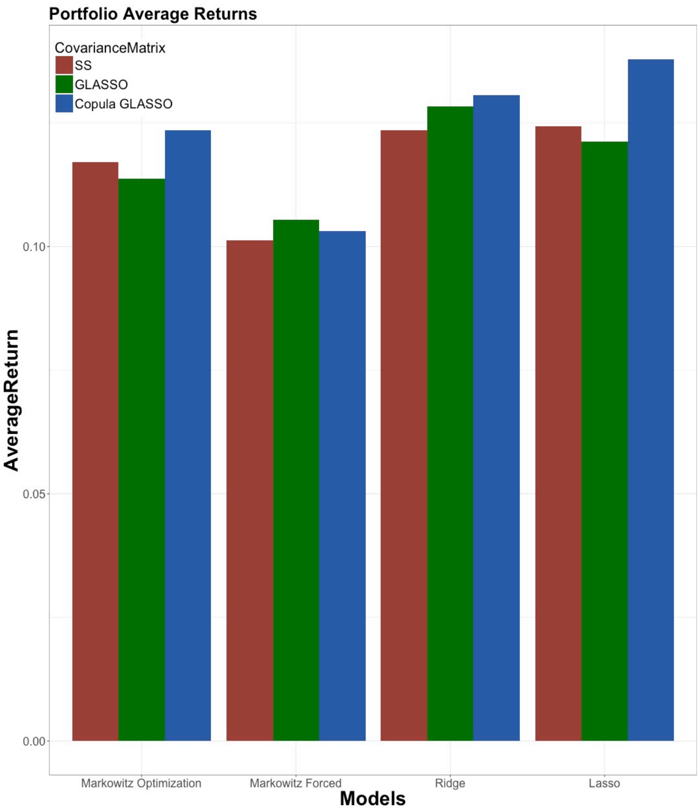

Figure 9: Out-of-Sample Annual Average Returns. Differentiation is given by model and by covariance precision pair inputs (index).

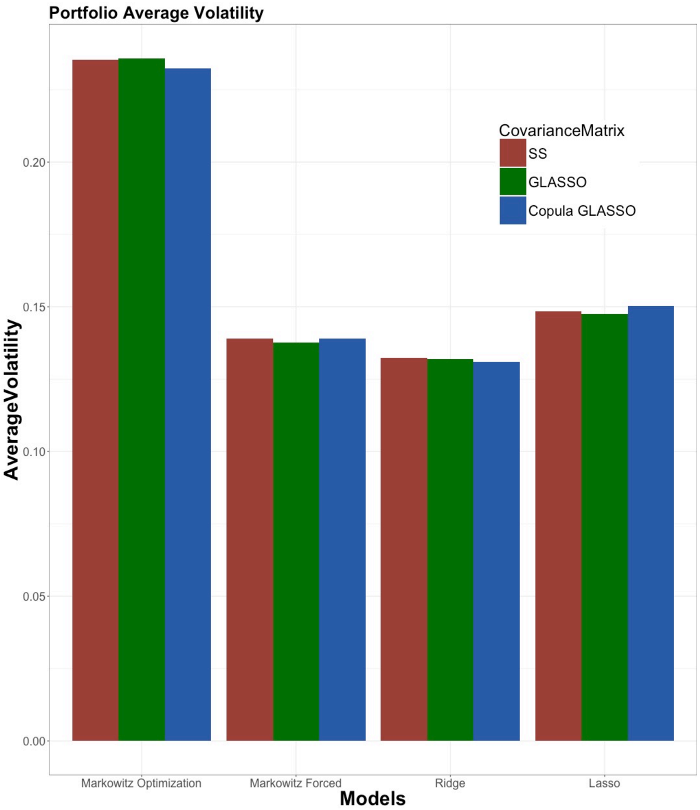

Figure 10: Out-of-Sample Annual Average Volatility. Differentiation is given by model and by covariance precision pair inputs (index)

were achieved by the Copula GLASSO pair. Furthermore, in Lasso, the outperformance of this pair is 11% greater than the SS benchmark pair.

It is desirable to compare portfolio performances with respect to any other investments that an investor could be interested in, the Information Ratio displayed in **Figure 7** compares our portfolios performances with respect to the S&P500 index. In this case, Ridge Regression has the best performance, showing high IR above 0.3, it is important to note that GLASSO and Copula GLASSO achieve IR higher than 0.4, meaning that they are portfolios that are able to consistently outperform the benchmark S&P500 for long periods of time. Additionally, similarly to SR, 3 out of 4 models were largely better off by having the Copula GLASSO pair as input.

Finally, in **Figure 8**, we observe the portfolio performances by using the Calmar Ratio, different to the others, this ratio considers maximum drawdown, showing both potential opportunity but more importantly potential risk in the investment. It is known that a Calmar Ratio's above 1 is desirable, both penalized models surpass this limit, while the benchmark models remain necessitous. Even more so, comparing the best of each model, we have CR of 0.541, 0.822, 1.169 and 1.195 for Markowitz Optimization, Markowitz Forced, Ridge and Lasso Regression respectively. In addition, between the best of the benchmark and the proposed models, there are improvements that range between 42 and 120%, a clear difference window in favor of the proposed models. It is important to mention, Copula GLASSO again shows dominance in 3 models, highlighting the improvement that it offers to each model.

In general, there is a clear improvement in all performance measures, both the penalized regression models and the enhancement in parameter estimation achieved better results. In all cases, the best portfolio was a combination between these two sparse statistic techniques, showing always a better out-of-sample performance. Consistently, Copula GLASSO, appeared as the best model in the whole range of portfolios. Furthermore, analyzing average annual returns and volatility in **Figures 9-10**, one can partly explain this outperformance; Copula GLASSO consistently increases the average return while maintaining a fairly equal volatility, relative to the other models. On the other hand, all models except Markowitz Optimization had similar average annual volatilities, around 13-15%, while Markowitz Optimization experienced a greater 23%, this partly explains its poor performance, especially in metrics such as Calmar and Sharpe Ratio where volatility is harshly punished. For more detailed graphs and additional results head to Appendix A.

4 CONCLUSIONS AND FUTURE WORK

We conclude by mentioning the highlights throughout our paper. Starting with the most simple and renowned models, we identify estimation errors, ill-posed matrices and multicollinearity that lead these models' solutions to poor out-of-sample performances. Therefore, we arrive to desired model characteristics such as sparsity and stability that can tackle this issue and improve empirical performance. We proposed two different approaches to achieve these desired characteristics in our models. First, we proposed penalization techniques such as L1 and L2 norm constraints in the penalization weights, that shrink the linear regression estimators, decreasing estimation errors to achieve stability and in the case of L1 norm penalization, generating sparsity by shrinking some estimators to zero. On the other hand, different to most solutions proposed in most papers, where issues are corrected *a posteriori*, we proposed a solution that corrects the issue *a priori* by improving the estimation of the input parameters, specifically the covariance and precision matrices. We achieve stability by making the parameters less reliable on data that could drastically change over time, this is done through semi-parametric estimations; then, we add sparsity with the use of GLASSO penalization within the covariance and precision matrices. Due to the fact that these two approaches are calculated separately, we noticed that they could be combined to create new models that proved to have additional desired qualities. Also, note that sparse portfolios pose an additional practical advantage regarding transaction costs, considering its nature of yielding a solution with a subset of zero weights, investors get to reduce their transactional fees, especially when they invest in periodically rebalanced portfolios.

In section 3, we applied these models in a real data scenario with 448 stocks to choose from. After obtaining the out-of-sample results, we noticed that they proved everything mentioned beforehand. As expected, Markowitz mean-variance models performed poorly, they presented high volatilities, with similar performance to the simple Naïve model. On the other hand, the L1 and L2-norm penalization models outperformed all benchmarks, doubling their results in all performance measures. While sparse input parameter covariance and precision matrices showed some slight improvement when used in Markowitz mean-variance models, they showed greater improvements when used with Ridge and Lasso models. Even though there's not a unanimous better performing model for all metrics, these metrics do clearly point out two best performing models, these are Ridge Regression with the Copula GLASSO pair and the Lasso Regression with the Copula GLASSO pair.

In this case, after reviewing the real data application results, we can conclude, that effectively these two proposed sparse and stability methods correct the undesirable characteristics that the optimal Markowitz mean-variance model contain. Hence, improving their empirical out-of-sample performance, generating considerably more attractive portfolios for investors. Additionally, the mixture of both these methods yielded an even more attractive portfolio with the best out-of-sample performance within the 13 models.

Having obtained desirable results for the proposed models, more questions arise for future work. Within the results, we observed an evident synergy between the sparse Copula GLASSO pair and the sparse penalization models (L1 and L2-norm), it would be interesting to investigate and understand why this relationship yield marginally greater results, we could ask: Is their relationship additive? Are they independent?

Also, given the improvement shown by the estimation of the covariance matrix using arbitrary graphs and restricting these with the semi-parametric extension of the Gaussian through Copulas, it would be interesting to make a similar investigation to estimate another covariance matrix using the other sparse nonparametric graphical model that Lafferty, Liu and Wasserman propose in their paper "Sparse Nonparametric Graphical Models" (2012). They derive a fully non-parametric model that uses kernel density estimation and restricts the graphs to trees and forests, it would be interesting to estimate the covariance matrix with this approach and evaluate its empirical performance.

Moreover, it would also be interesting to implement these models on different kinds of markets such as emerging markets, (where the semi-strong form might not apply) evaluating each models performance, finally concluding if these proposed models can be applied to a broader range of markets.

Finally, further enhancements can be applied to the covariance matrix estimation, for example, with the use of GARCH, ARCH, EWMA or implied volatility, the estimation of the matrix diagonal (variances) could be further improved. Note that these direct changes in the matrix might lead to non-invertible matrices due to violation of the semi-positive definite property, so its application must be handled finely.

5 References

Banerjee, O., El Ghaoui, L. & D'Aspremont, A. (2007). Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data. *Journal of Machine Learning Research (2008)*, (9), pp.485-516.

Carrasco, M. and Noumon, N. (2010). Optimal Portfolio Selection using Regularization. *Université de Montréal*.

Das, S., Markowitz, H., Scheid, J., & Statman, M. (2010). Portfolio Optimization with Mental Accounts. *Journal Of Financial And Quantitative Analysis*, *45*(2), 311-334. doi: 10.1017/s0022109010000141

De Miguel, V., Garlappi, L., & Uppal, R. (2009). Optimal Versus Naive Diversification: How Inefficient is the 1/NPortfolio Strategy?. *Review Of Financial Studies*, *22*(5), 1915-1953. doi: 10.1093/rfs/hhm075

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2).

Fan, J., Zhang, J. & Yu, K. (2012). Vast Portfolio Selection With Gross-Exposure Constraints. *Journal of the American Statistical Association*, 107(498), pp.592-606.

Friedman, J., Hastie, T. and Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), pp.432-441.

Galloway, M. (2019). Retrieved 13 November 2019, from https://cran.r-project.org/web/packages/CVglasso/CVglasso.pdf

Hojsgaard, S., Edwards, D., & Lauritzen, S. (2012). *Graphical models with R*. New York: Springer.

Jagannathan, R. and Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*, 58(4), pp.1651-1683.

Kan, R. & Zhou, G. (2007). Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3), pp.621-656.

Kremer, P., Lee, S., Bogdan, M. & Paterlini, S. (2017). Sparse Portfolio Selection Via the Sorted L1 - Norm. *SSRN Electronic Journal.*

Lafferty, J., Liu, H., & Wasserman, L. (2012). Sparse Nonparametric Graphical Models. *Statistical Science*, *27*(4), 519-537. doi: 10.1214/12-sts391

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal Of Empirical Finance*, *10*(5), 603-621. doi: 10.1016/s0927-5398(03)00007-0

Li, J. (2015). Sparse and Stable Portfolio Selection With Parameter Uncertainty. *Journal Of Business & Economic Statistics*, *33*(3). doi: 10.1080/07350015.2014.954708

Macrotrends.net. (2019). *1 Year Treasury Rate - 54 Year Historical Chart*. [online] Available at: https://www.macrotrends.net/2492/1-year-treasury-rate-yield-chart [Accessed 8 Nov. 2019].

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

Tu, J. and Zhou, G. (2009). Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies. *Journal of Financial Economics*, (99), pp.204-215.

Turlach, B., & Weingessel, A. (2019). Retrieved 13 November 2019, from https://cran.r-project.org/web/packages/quadprog/quadprog.pdf

Xing, H., Lai, T., & Chen, Z. (2010). Mean-Variance Portfolio Optimization When Means and Covariances are Unknown. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1694057

A APPENDIX: DETAILED GRAPHS AND FIGURES

Figure 1: Cumulative Wealth Index (CWI) of all 12 models and Naïve model, with initial wealth of $1.

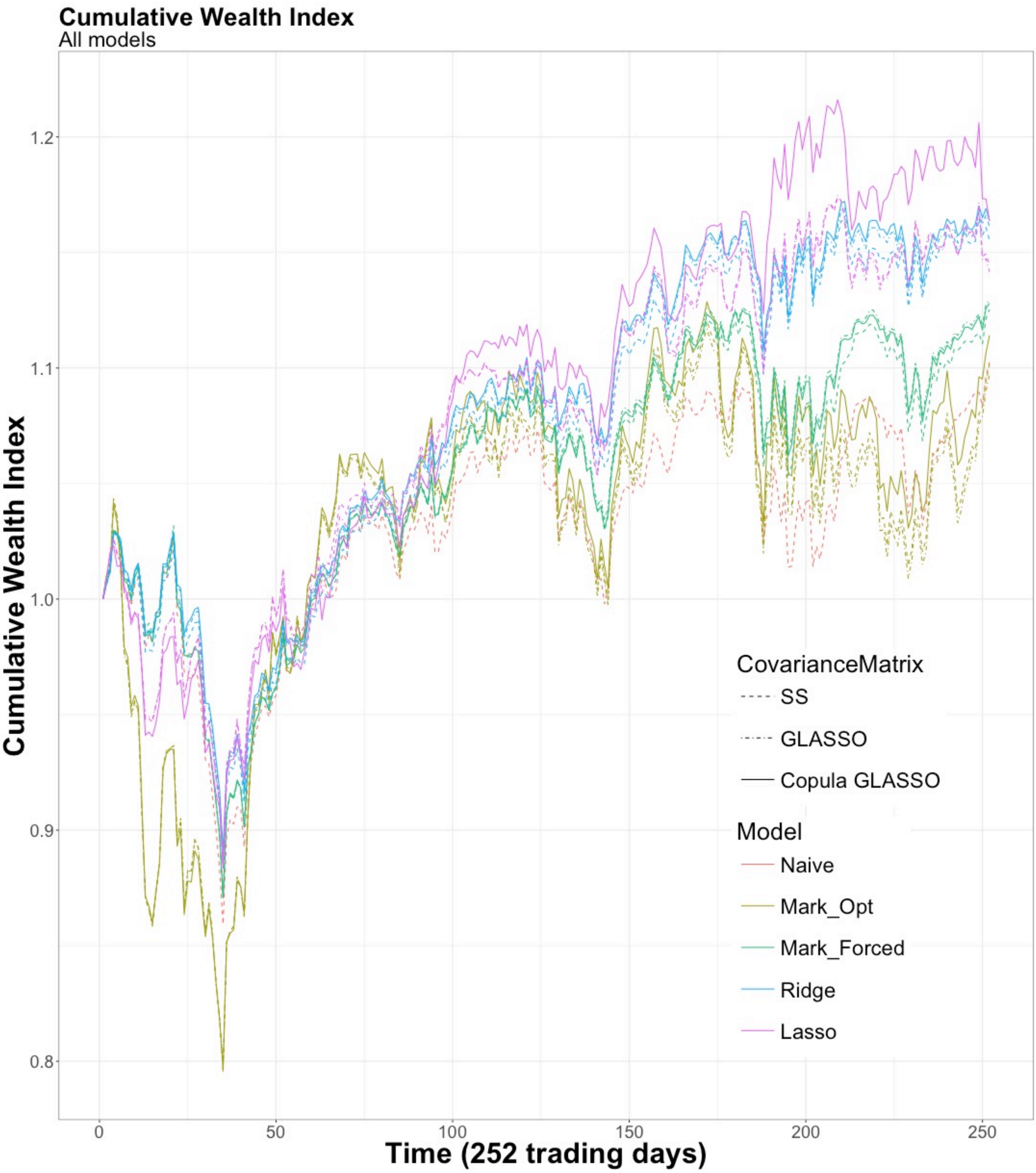

Figure 2: Cumulative Wealth Index (CWI) of three Markowitz mean-variance Forced Models, with different covariance precision pair inputs (index), relative to benchmark Naïve (red), all with initial wealth of $1.

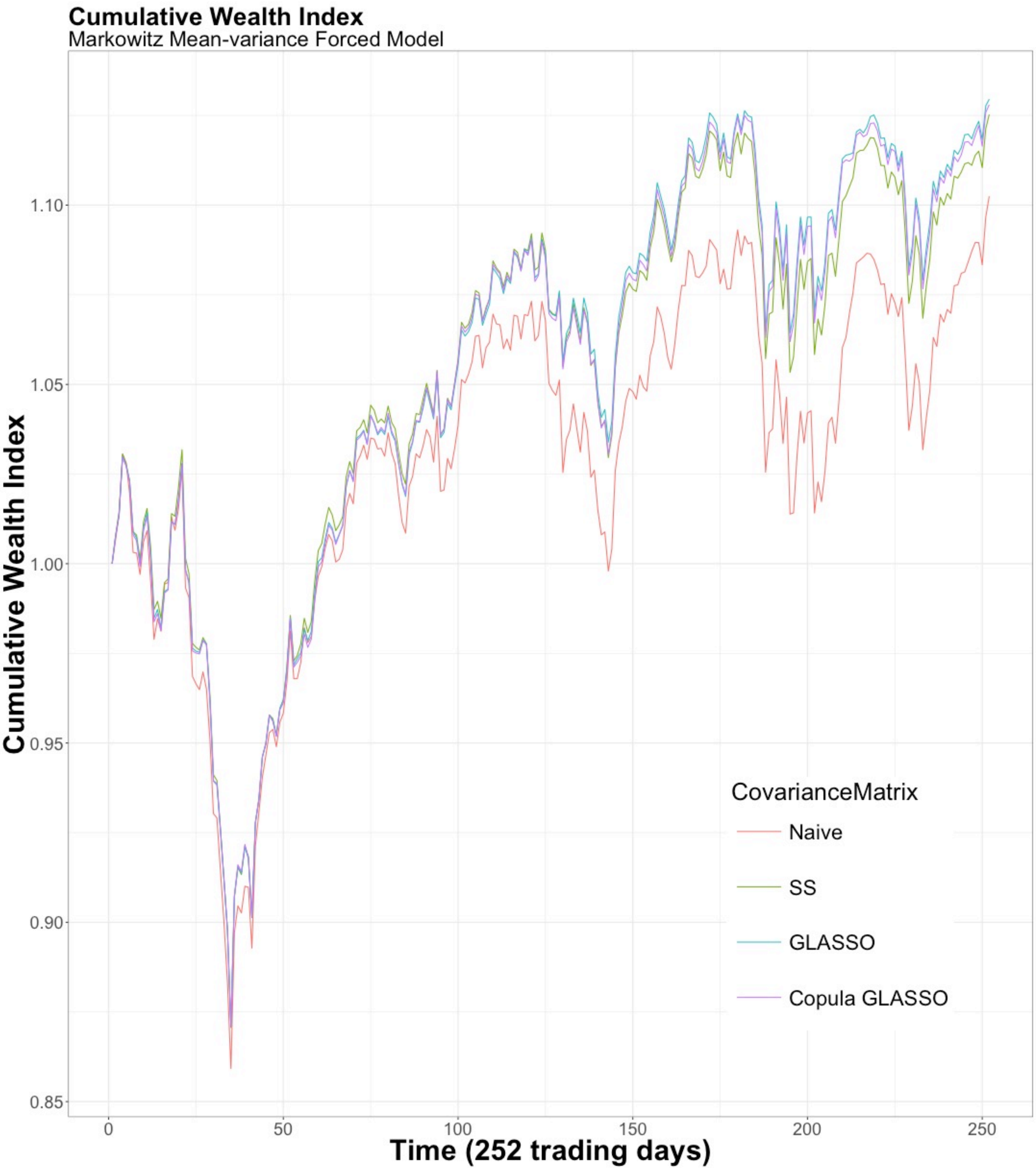

Figure 3: Cumulative Wealth Index (CWI) of three Markowitz mean-variance Optimization Models, with different covariance precision pair inputs (index), relative to benchmark Naïve (red), all with initial wealth of $1.

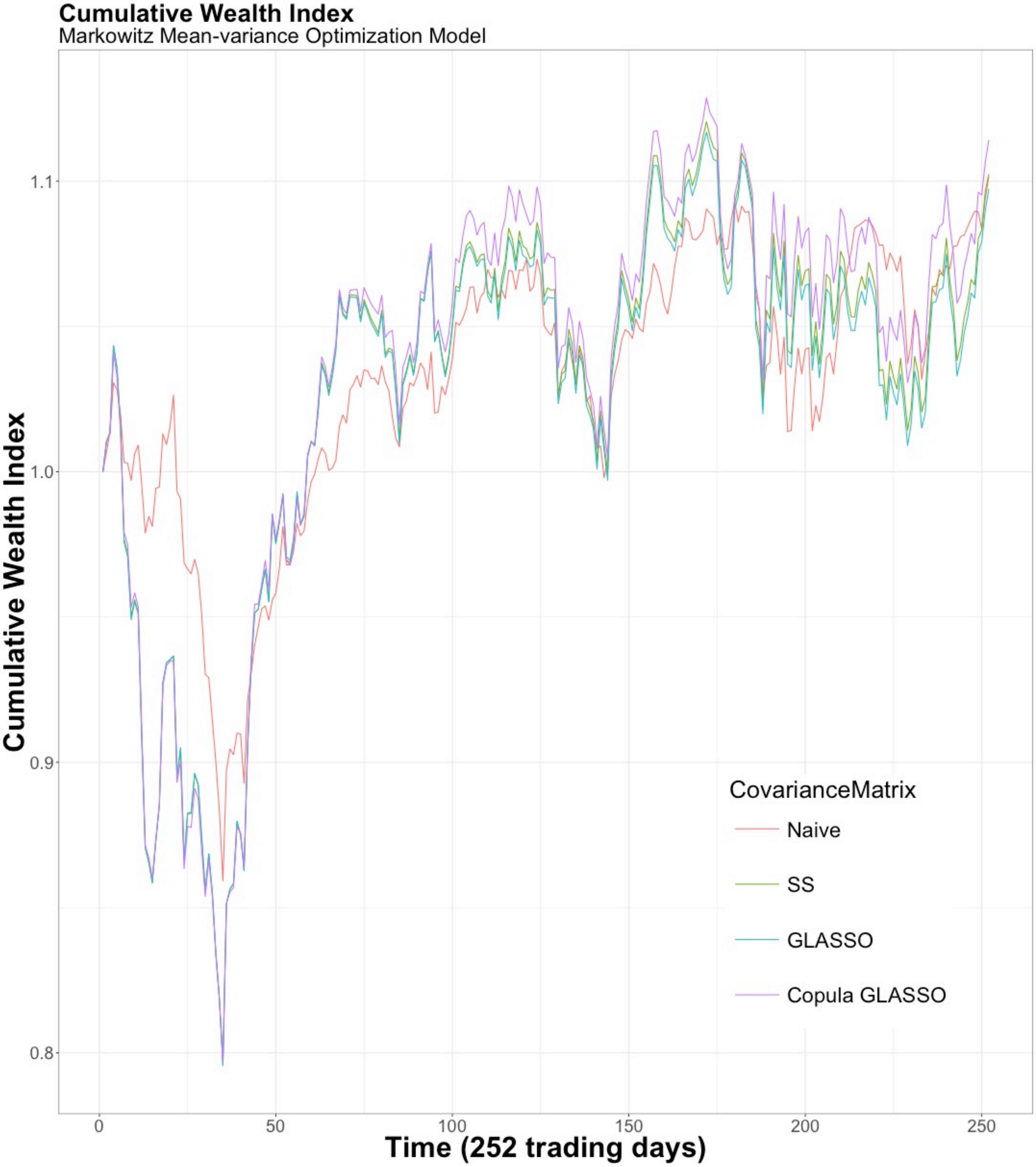

Figure 4: Cumulative Wealth Index (CWI) of three Ridge Regression Models, each having different covariance precision pair inputs (index), relative to benchmark Naïve (red), all with initial wealth of $1.

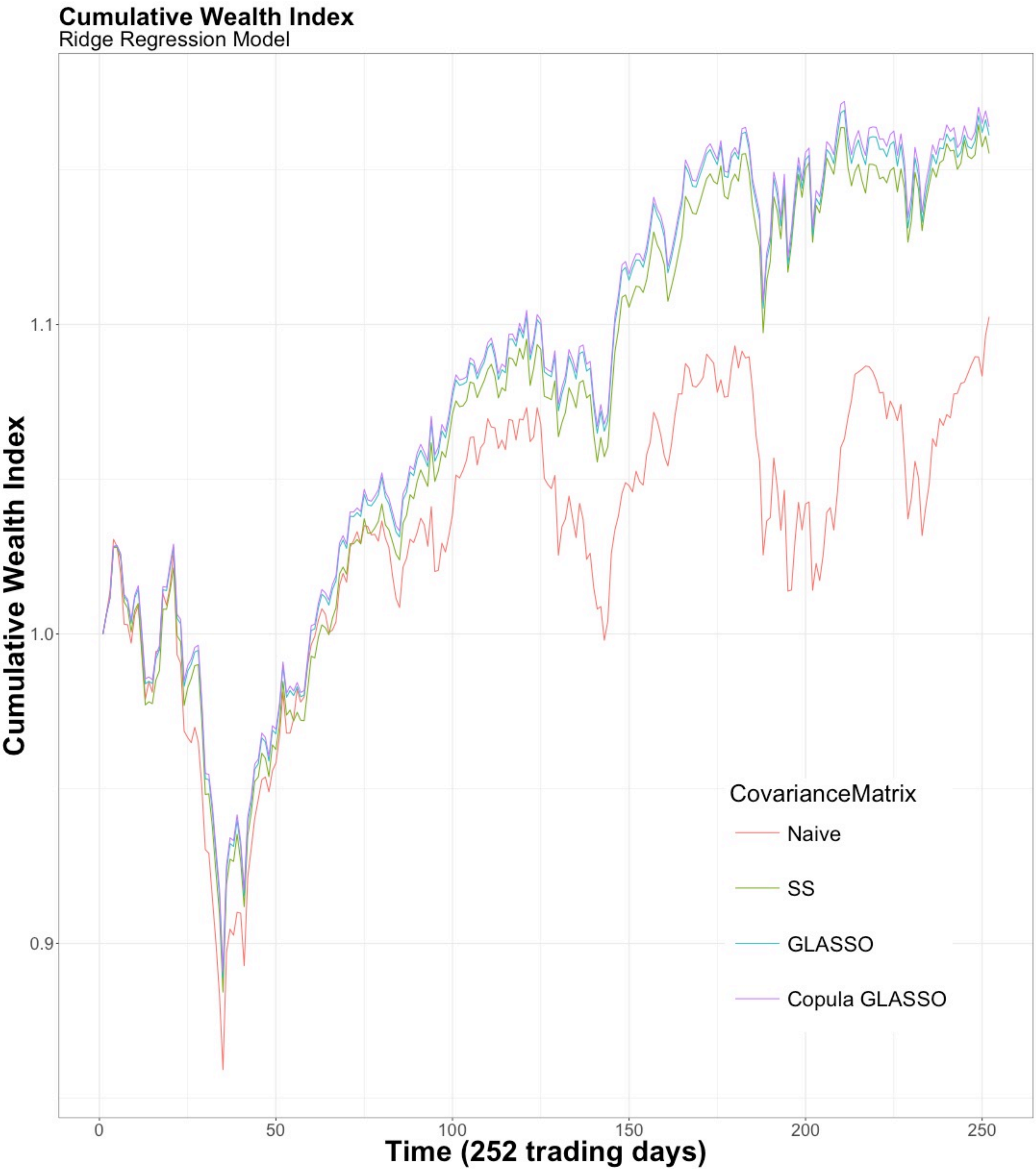

Figure 5: Cumulative Wealth Index (CWI) of three Lasso Regression Models, each having different covariance precision pair inputs (index), relative to benchmark Naïve (red), all with initial wealth of $1.

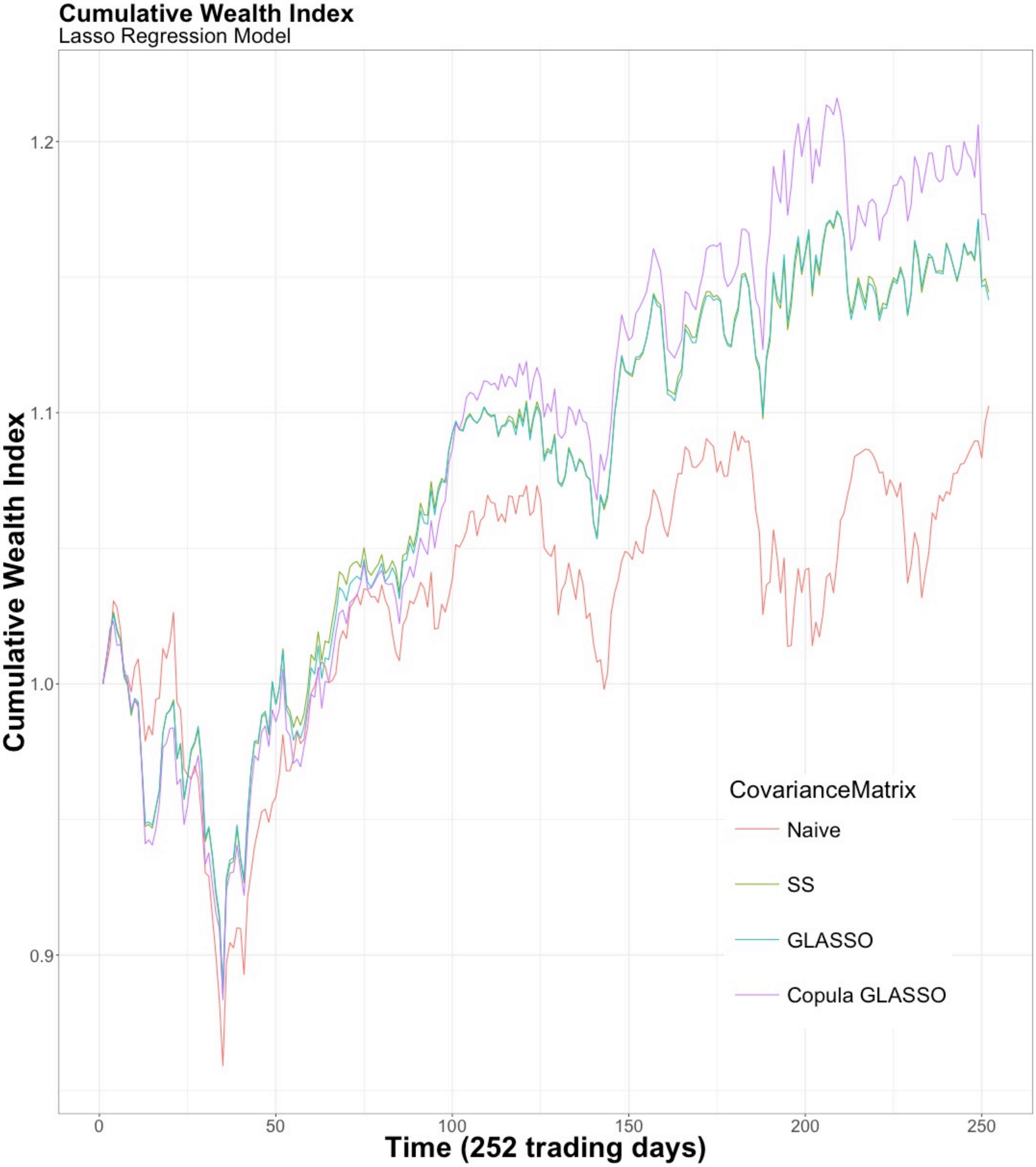

Figure 6: Annualized Sharpe Ratio (SR). Differentiation is given by model and by covariance precision pair inputs (index).

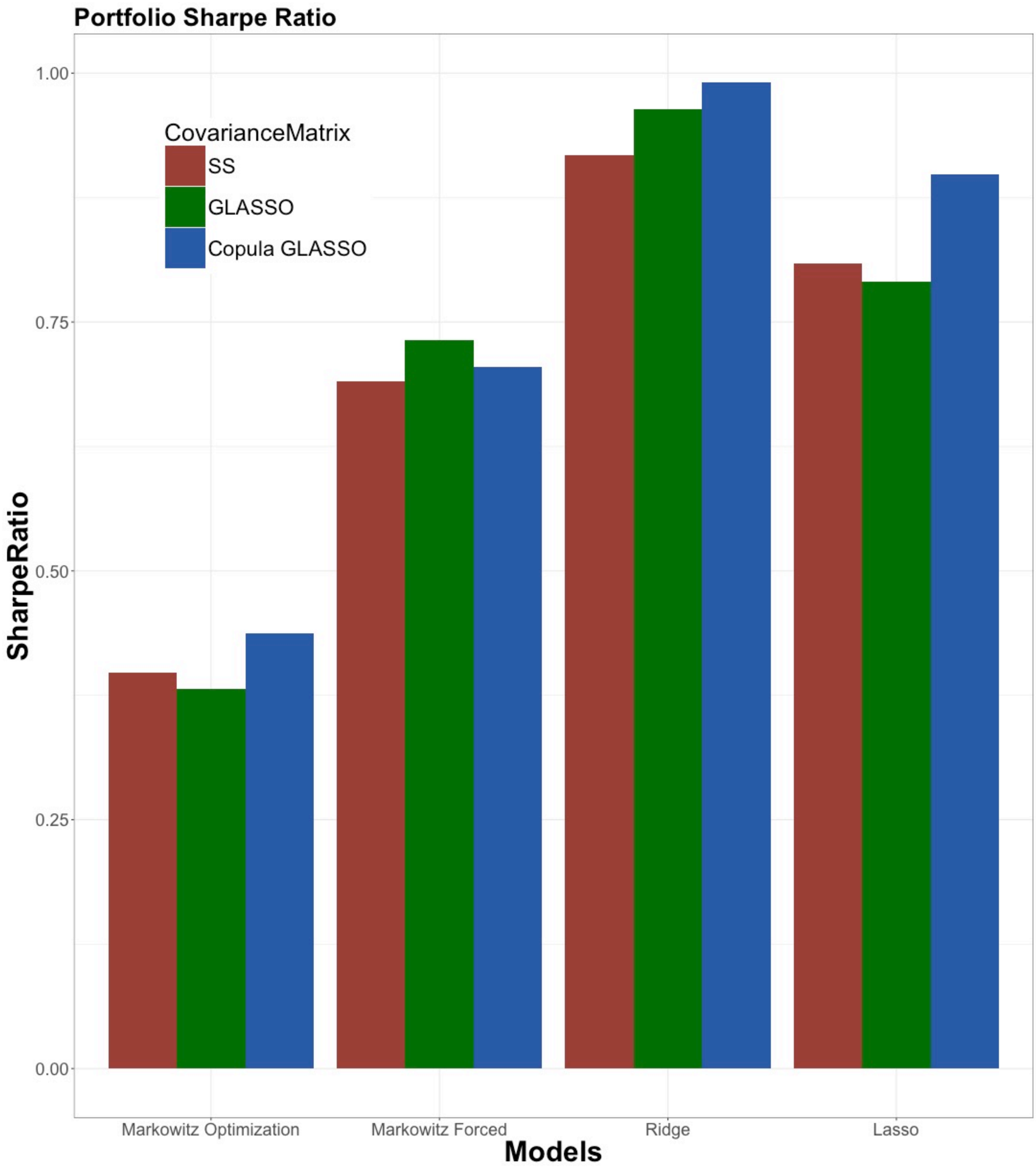

Figure 7: Annualized Information Ratio (IR) with S&P 500 as benchmark. Differentiation is given by model and by covariance precision pair inputs (index).

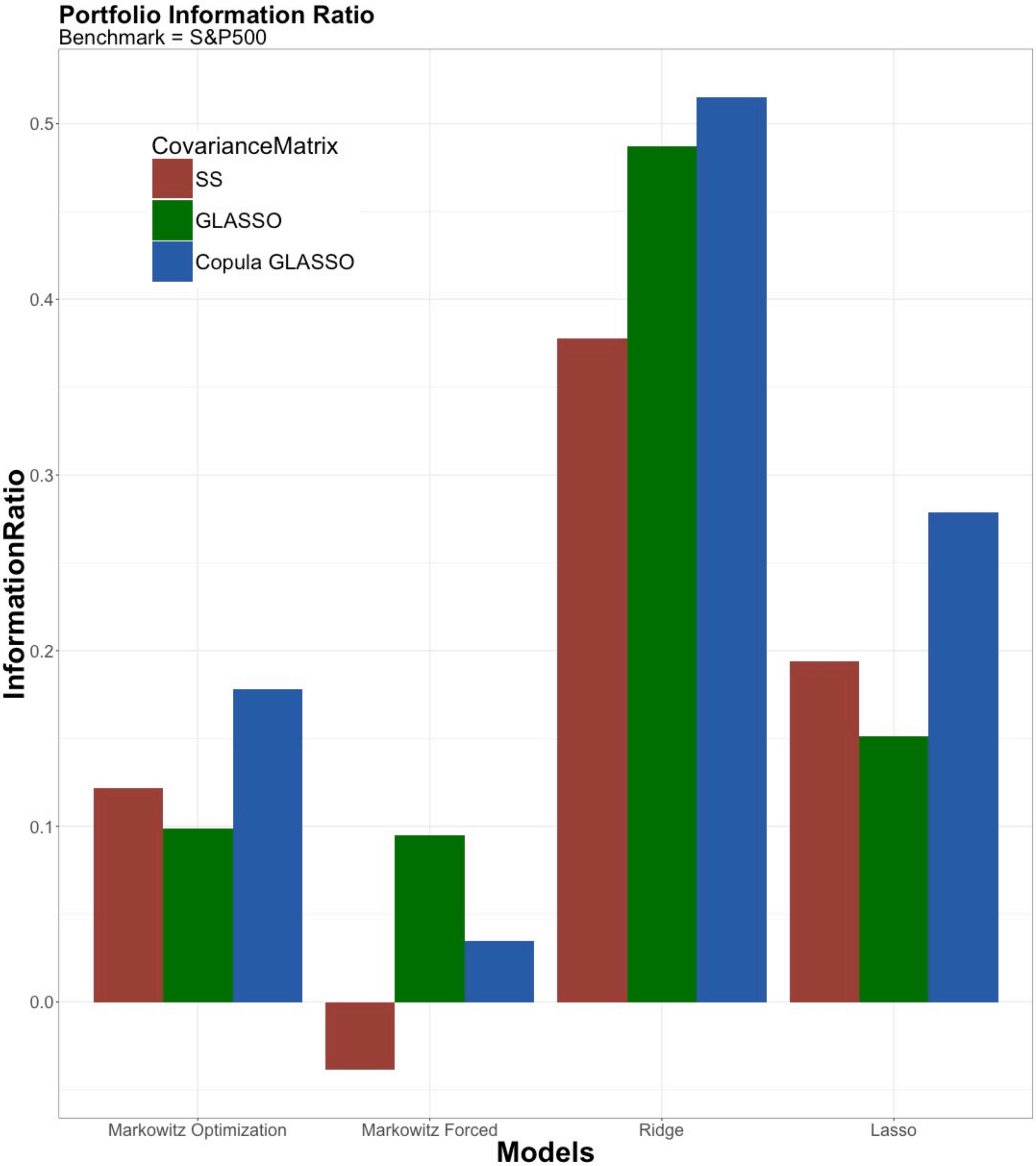

Figure 8: Annualized Calmar Ratio (CR). Differentiation is given by model and by covariance precision pair inputs (index).

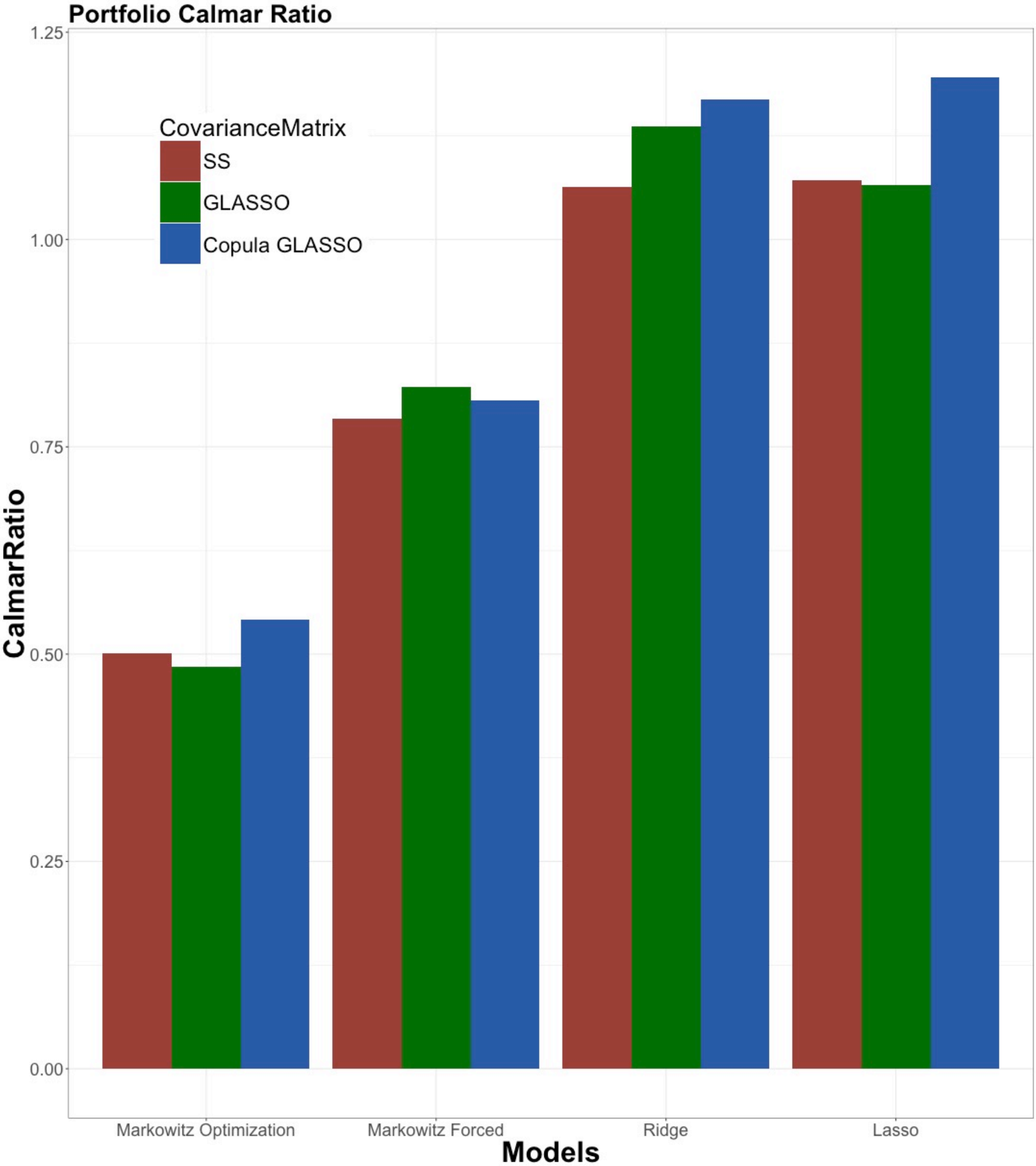

Figure 9: Out-of-Sample Annual Average Returns. Differentiation is given by model and by covariance precision pair inputs (index).

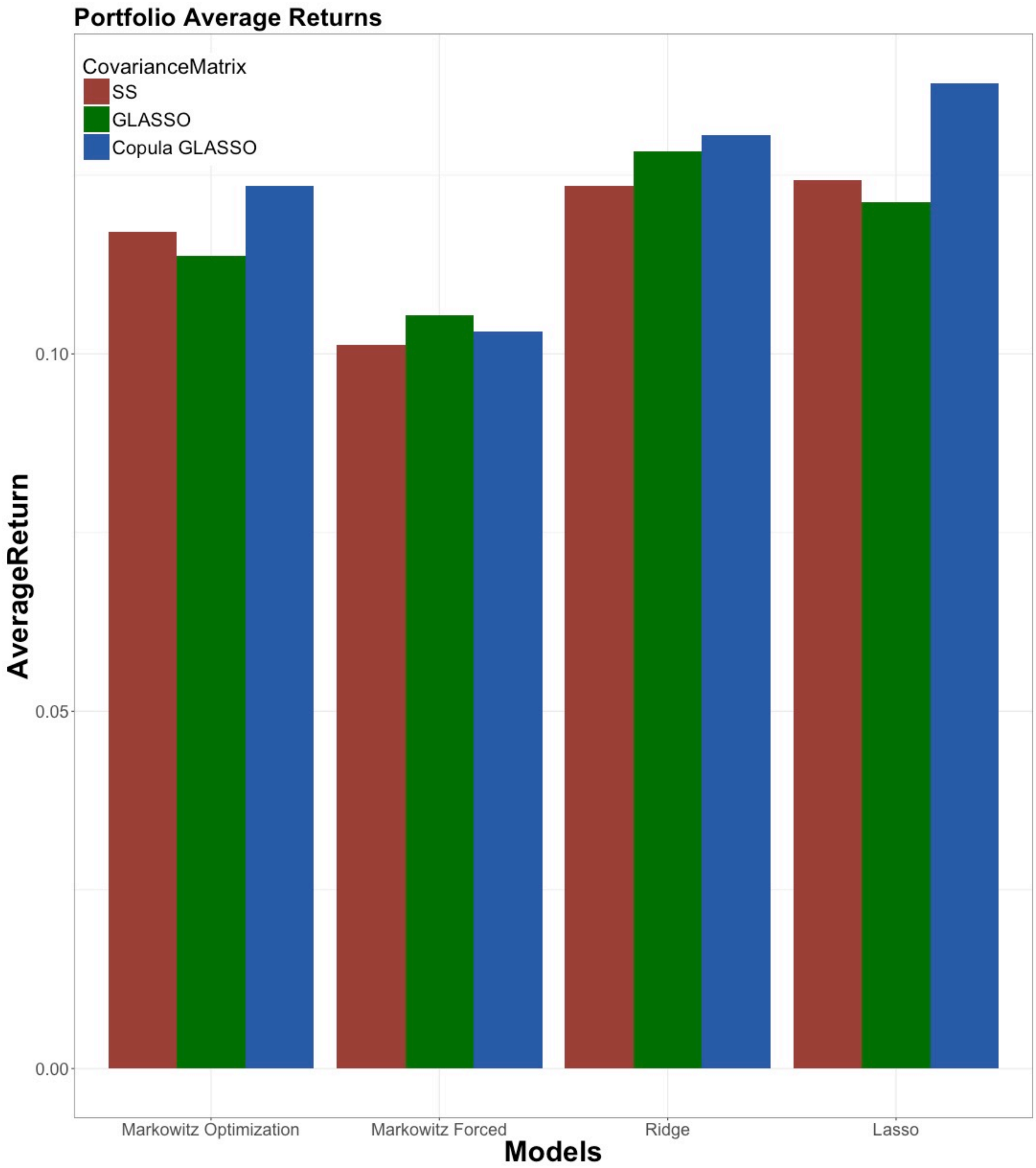

Figure 10: Out-of-Sample Annual Average Volatility. Differentiation is given by model and by covariance precision pair inputs (index).

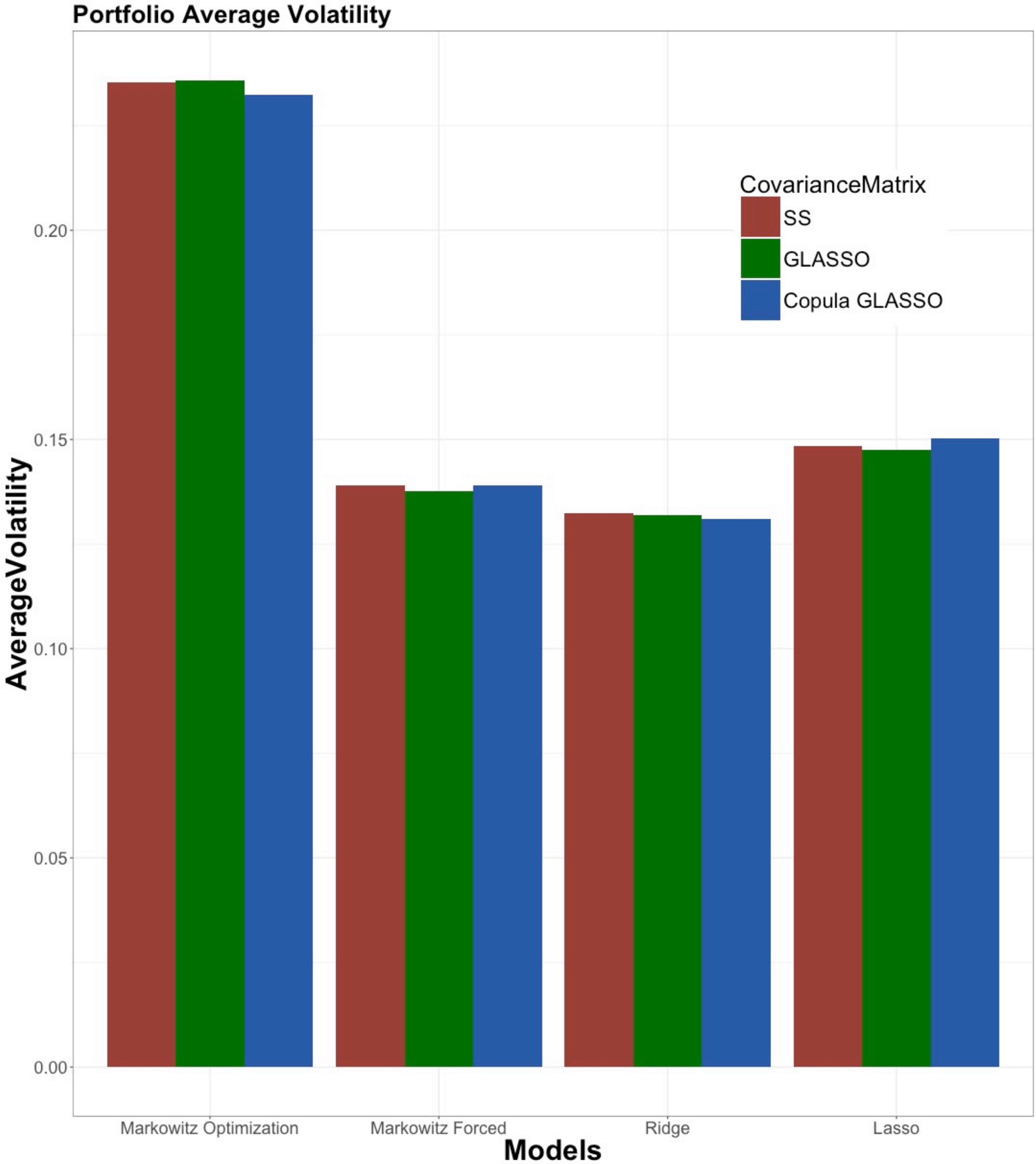

Figure 11: Annualized Information Ratio (IR) with Markowitz mean-variance Forced Model portfolio with sample correlation and traditional precision matrix as benchmark. Differentiation is given by model and by covariance precision pair inputs (index).

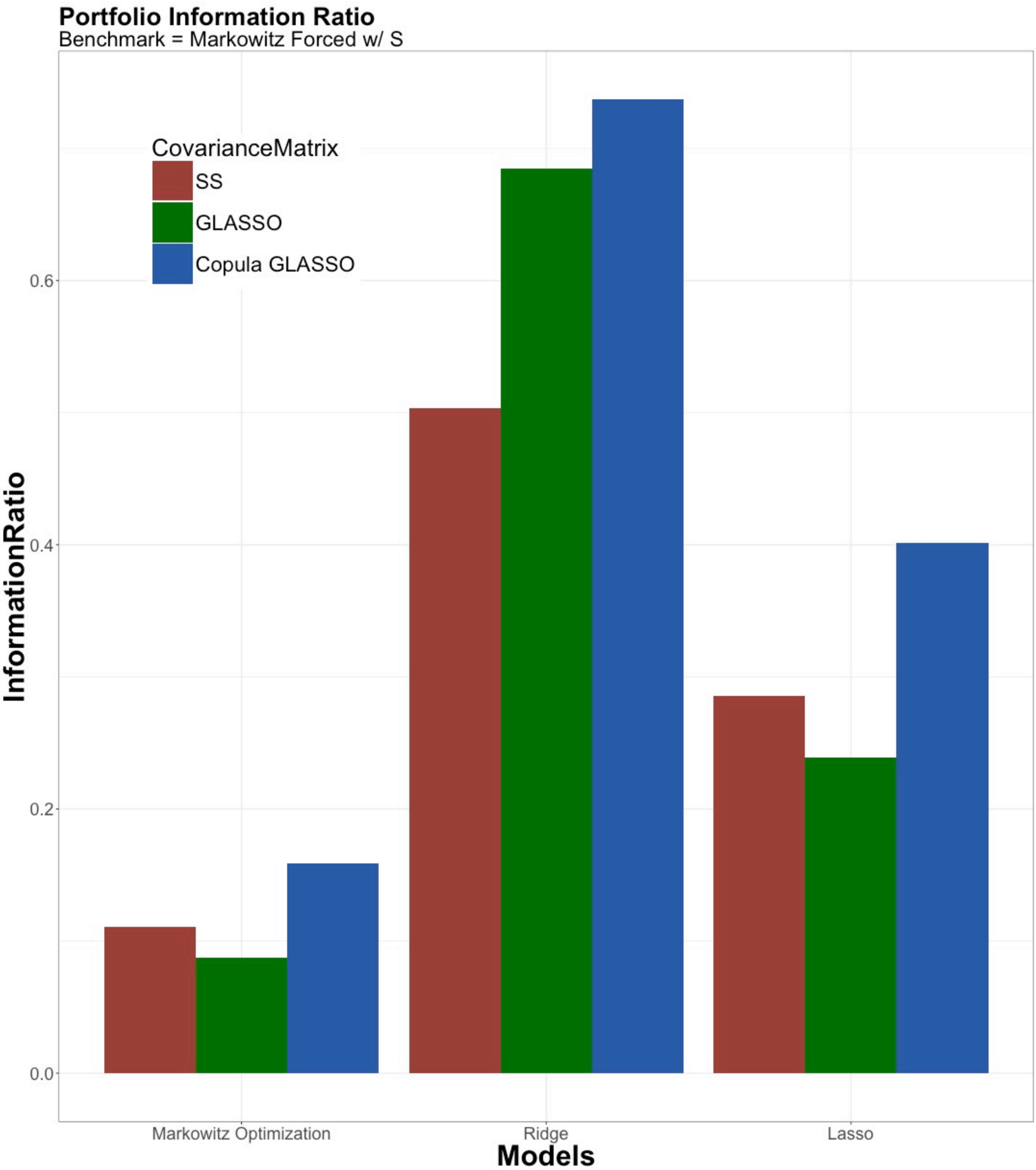

Figure 12: Annualized Information Ratio (IR) with Markowitz mean-variance Optimization Model portfolio with sample correlation and traditional precision matrix as benchmark. Differentiation is given by model and by covariance precision pair inputs (index).