Queen's Economics Department Working Paper No. 1453

Efficient mean-variance portfolio selection by double regularization

N'Golo Kone

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

2-2021

Efficient mean-variance portfolio selection by double regularization

N'Golo Koné
Department of Economics Queen's University
94 University Avenue Kingston, Kingston Ontario, Canada K7L 3N6

February 09, 2021

Abstract

This paper addresses the estimation issue that exists when estimating the traditional mean-variance portfolio. More precisely, the efficient mean-variance is estimated by a double regularization. These regularization techniques namely the ridge, the spectral cut-off, and Landweber-Fridman involve a regularization parameter or penalty term whose optimal value needs to be selected efficiently. A data-driven method has been proposed to select the tuning parameter. We show that the double regularized portfolio guarantees to investors the maximum expected return with the lowest risk. In empirical and Monte Carlo experiments, our double regularized rules are compared to several strategies, such as the traditional regularized portfolios, the new Lasso strategy of Ao, Yingying, and Zheng (2019), and the naive 1/N strategy in terms of in-sample and out-of-sample Sharpe ratio performance, and it is shown that our method yields significant Sharpe ratio improvements and a reduction in the expected utility loss.

Keywords: Portfolio selection, efficient mean-variance analysis, double regularization.
JEL subject Classification: C52, C58, G11

1 Introduction

In his seminal work, Markowitz (1952) stated that the optimal portfolio should be selected by an optimal trade-off between return and risk instead of an expected return maximization only. This principle continues to play a significant role in the financial market. The optimal solution obtained using this principle has a simple closed form that depends on the expected return and the inverse of the covariance matrix of asset returns. Nonetheless, these two quantities are unknown and need to be estimated in practice to obtain a feasible solution. The standard way to estimate this optimal solution is to replace the unknown parameters by their empirical counterpart. The plug-in portfolio obtained using the sample moments has been shown to give very poor in-sample and out-of-sample performance in the literature. This observed performance is essentially due to the large number of assets in the financial market compared with the sample size (see Kan and Zhou (2007), Bai, Liu, and Wong (2009) and El Karoui (2010) for instance). In fact, when the number of assets grows compared with the sample size, the sample covariance matrix used in the plug-in strategy is not appropriate because nearly singular. Hence, inverting such a matrix in the investment process may generate a strategy which is fare from the optimal one with very poor performance. This problem is amplified when using the sample mean to estimate the expected return in the optimal portfolio. Indeed, the estimation error in the expected return might be important, especially in a large financial market. Stein (1956) and Brown, Zhao, et al. (2012) even argue that the usual estimator of the expected return should be inadmissible if the dimension is sufficiently large.

Several methods have been proposed in the literature to deal with these issues in order to improve the performance of the selected strategy.

Some papers in the literature are focus on regularizing the covariance matrix of asset distribution or its inverse. Indeed, Ledoit and Wolf (2003, 2004a,b) propose to replace the covariance matrix by a weighted average of the sample covariance and some structured matrix. Ledoit and Wolf (2017, 2018) propose a nonlinear shrinkage estimator which is more flexible than the linear one by modifying each eigenvalue of the sample covariance matrix under the framework of Markowitz's portfolio selection. Several other papers are focus on estimating efficiently the covariance matrix or its inverse (see Rothman, Bickel, Levina, Zhu, et al. (2008), Fan, Liao, and Mincheva (2011, 2013), Fan, Fan, and Lv (2008), Touloumis (2015), Bodnar, Gupta, and Parolya (2016), Bauder, Bodnar, Parolya, and Schmid (2020), Koné (2021) among others). Recently, Carrasco, Koné, and Noumon (2019) investigate various regularization techniques from the literature of inverse problems to stabilize the inverse of the sample covariance matrix.

Jorion (1986) and Bodnar, Okhrin, and Parolya (2019) propose to use a shrinkage estimation for the expected return which seems to be more appropriate than the sample mean.

One other way to improve the performance of the selected strategy consists of imposing appropriate constraint in the optimization problem. Jagannathan and Ma (2003) impose a short-sale constraint in the investment process. They argue that this constraint could help to improve the performance by reducing the estimation risk in estimating the optimal portfolio. Brodie, Daubechies, De Mol, Giannone, and Loris (2009) and Fan, Zhang, and Yu (2008) generalize the short-sale constraint by using a method called Lasso which consists in imposing a constraint on the sum of the absolute values (l_1 norm) of the portfolio weights. This method generates sparse portfolios which degree of sparsity depends on a tuning parameter. This sparsity property should be very useful when trying to reduce transaction costs in the investment process. DeMiguel, Garlappi, Nogales, and Uppal (2009) propose a general method in terms of norm-constrained minimum-variance portfolio. They find that the norm-constrained portfolios often have a higher Sharpe ratio than the portfolio strategies in Jagannathan and Ma (2003), Ledoit and Wolf (2003, 2004a,b), the 1/N portfolio, and other strategies in the literature, such as factor portfolios. Fastrich, Paterlini, and Winker (2015) propose a new penalty that explicitly considers financial information to improve the l_1-regularization approach. Recently, Ao, Yingying, and Zheng (2019) introduce a new way to estimate the mean-variance portfolio based on an unconstrained regression representation of the optimization problem combined with high dimensional sparse regression method. Moreover, Kone (2020) proposes, in a dynamic setting, a temporal stability constraint in the investment process to guarantee that the optimal portfolio composition does not fluctuate wildly between periods. In addition to stabilize the inverse of the covariance matrix in the investment process, this constraint introduces a second level of regularization to control for the estimation errors in the expected return.

Brandt, Santa-Clara, and Valkanov (2009) and DeMiguel, Martin-Utrera, Nogales, and Uppal (2020) model directly the portfolio weights as a function of the assets characteristics to avoid the difficulties in the estimation of asset returns moments.

Our paper contributes to this vast literature by estimating the optimal portfolio in Proposition 1 of Ao, Yingying, and Zheng (2019) by a double regularization. This work is then related to the vast literature of linear invert problem (see Carrasco, Florens, and Renault (2007), Carrasco, Florens, and Renault (2014), Carrasco (2012), Carrasco and Tchuente (2015) among others).

Estimating this optimal solution involves a first step estimation of the square of the maximum Sharpe ratio of the optimal portfolio. The standard way to estimate this quantity will be to replace it by its empirical counterpart as it has to be done with the traditional regularization techniques of Carrasco, Koné, and Noumon (2019). Nonetheless, this estimation implies also estimating the covariance matrix of asset return and take its inverse. The sample covariance matrix used for this purpose in not an appropriate choice because nearly singular. Hence, the resulting estimation error by using the sample counterpart of the square of the maximum Sharpe ratio could considerably deteriorate the performance of the selected strategy. Therefore, Ao, Yingying, and Zheng (2019) propose an alternative estimator for this quantity which relies on a normal distribution assumption of assets return. To avoid imposing a normality assumption in our economy, we apply a regularization to estimate the square of the maximum Sharpe ratio by stabilizing the inverse of the covariance matrix that appears in this quantity. At each step of our double regularization, which consists of stabilizing the inverse of the covariance matrix of assets, several regularization techniques from inverse problem literature have been used. These regularization techniques namely the ridge, the spectral cut-off, and Landweber-Fridman involve a regularization parameter or penalty term whose optimal value is selected to minimize the expected distance between the inverse of the estimated covariance matrix and the inverse of the true covariance matrix at the first level. The tuning parameter of the second level of regularization is selected to minimize the expected loss in utility of a mean-variance investor. The optimality of these tuning parameters selection procedures has been largely augmented in the literature (see for instance Carrasco, Koné, and Noumon (2019)). Under appropriate regularity conditions, we show that the double regularized portfolio guarantees to investors the maximum expected return with the lowest risk. This implies that our selected portfolio achieves asymptotically the true Sharpe ratio.

To evaluate the performance of our procedures we implement a simulation exercise based on a three-factor model calibrated on real data from the US financial market from July 1980 to June 2016. We obtain by simulation that our procedure significantly improves the performance of the selected strategy with respect to the Sharpe ratio and the expected utility loss. The double regularized portfolios are compared to the new Lasso portfolio, the traditional regularized portfolio, and the naive 1/N strategy in terms of in-sample utility loss and the Sharpe ratio, and it is shown that our method yields significant Sharpe ratio improvements and considerably reduces the expected utility loss. To confirm our simulations, we do an empirical analysis using Kenneth R. French's 30-industry portfolios and 100 portfolios formed on size and book-to-market. According to this empirical result, by double regularizing the efficient mean-variance portfolio, we considerably improve the performance of the selected strategy in terms of maximizing the Sharpe ratio.

The rest of the paper is organized as follows. Section 2 presents the economy environment and the estimation method. Section 3 presents the first level of regularization and proposes a data-driven method to select the tuning parameter of this first regularization. Section 4 presents the second level of regularization and some asymptotic results. Some simulation results are given in section 5 and the empirical study in section 6. Section 7 concludes the paper. Proofs of all theoretical results are given section 8.

2 The economy environment and the estimation method

2.1 The economy environment

We consider a simple economy with N risky assets with random returns vector R_{t+1} and a risk-free asset where N is assumed to be large. We assume that the return on the risk-free asset is constant over time. Let denote by R_f the gross return on this risk-free asset. Empirically with monthly data, R_f will be calibrated to be the mean of the one-month Treasury-Bill (T-B) rate observed in the data.

Let denote by $r_{t+1} = R_{t+1} - R_f 1_N$ the vector of excess returns on the set of risky assets in the economy with 1_N the N-dimensional vector of ones. We assume as in Carrasco, Koné, and Noumon (2019) that the excess returns are independent and identically distributed with the mean and the covariance matrix given by μ and Σ respectively.

Let's denote by $\omega = (\omega_1, ..., \omega_N)'$ the share of the risky assets in the optimal portfolio. Hence, the investor allocates a fraction ω of wealth to the risky assets and the remainder $(1 - 1_N'\omega)$ to the risk-free asset.

Let r_t, $t = 1, \cdots, T$ be the observations of asset returns and R be the $T \times N$ matrix with tth row given by r_t'. $\Omega = E\left(r_t r_t'\right) = E\left(R'R\right)/T$. The mean-variance portfolio as proposed by Markowitz (1952) can be seen as a certain trade-off between the expected return and the variance of the asset returns. Therefore, the optimal solution of the mean-variance portfolio can be obtained by solving the following constrained optimization problem

$$\arg \max_{\omega:\omega^{'}\Sigma\omega^{'}<\sigma^2} E\left(\omega^{'} r_t\right) = \omega^{'}\mu \tag{1}$$

where σ is a given risk constraint. This way to formulate the classical mean-variance problem may be very useful in a situation of high uncertainty. It will help investors to select a strategy that correctly controls the global risk and maximizes at the same time the expected return of the selected portfolio. Ao, Yingying, and Zheng (2019) show that solving (1) is equivalent of solving the following unconstrained regression model

$$\arg \min_{\omega} E\left[\left(r_c - \omega^{'} r_t\right)^2\right] \tag{2}$$

where $r_c = \sigma \frac{1+\theta}{\sqrt{\theta}}$, $\theta = \mu^{'}\Sigma^{-1}\mu$ the square of the maximum Sharpe ratio of the optimal portfolio.

The optimal solution of this optimization problem is given by

$$\omega = r_c \Omega^{-1}\mu \tag{3}$$

which is unknown and needs to be estimated. The standard way to estimate this solution is to replace the unknown parameters by their counterpart empirical to obtained the so called plug -in rule.

$$\hat{\omega} = \hat{r}_c \hat{\Omega}^{-1}\hat{\mu} \tag{4}$$

Estimating this solution implies estimating the covariance matrix Ω and take its inverse. The choice of the sample covariance to form the plug-in rule may not be appropriate because it may be nearly singular, and sometimes not invertible. Moreover, we need to estimate r_c to form the optimal selected strategy. For a given level of risk, r_c is estimated based on an estimation of the parameter θ. Estimating θ implies estimating the covariance matrix Σ and take its inverse combined with an estimate of the expected return. The sample mean and the sample covariance matrix are not appropriate estimates for this purpose. In fact, when $N/T \to \rho \in (0,1)$ (the number of assets in the economy is large compared with the sample size) $\hat{\Sigma}$ and $\hat{\mu}$ may not be consistent estimates for Σ and μ. Hence, $\hat{\theta} = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$ may not appropriate in this situation.

To overcome all these problems related to the estimation of the solution in (3), Ao, Yingying, and Zheng (2019) propose to estimate it by solving the following optimization problem

$$\hat{\omega} = \arg\min_{\omega} \frac{1}{T}\sum_{t=1}^{T}\left(\hat{r}_c - \omega' r_t\right)^2 \tag{5}$$

subject to $\|\omega\| \leq \lambda$ where $\hat{r}_c = \sigma\frac{1+\hat{\theta}}{\sqrt{\hat{\theta}}}$, $\hat{\theta} = \frac{(T-N-2)\hat{\theta}_s - N}{T}$, $\hat{\theta}_s = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$. $\hat{\theta}$ is an unbiased estimator proposed by Kan and Zhou (2007) under the normality assumption of the excess return. The consistency of this estimator is obtained by assuming that the excess return are normally distributed which seems to be a restrictive assumption. Moreover, this estimator can take negative values so it is also an undesirable estimate for θ since $\theta \geq 0$.

In this paper, we address the estimation issue of the optimal solution through three regularization techniques (the ridge, spectral cut-off, Landweber-Fridman) borrowed from the linear invert literature (see Carrasco, Florens, and Renault (2007)). More precisely, we are going to use a double regularization technique to estimate consistently the efficient mean-variance portfolio. At each level of regularization, three different stabilization techniques will be used.

Before talking about these two level of regularization, let's briefly present the stabilization techniques we are interested in.

2.2 The regularization methods

The regularization methods used in this paper are drawn from the literature on inverse problems (see Kress (1999)). They are designed to stabilize the inverse of Hilbert-Schmidt operators (operators for which the eigenvalues are square summable). These regularization techniques will be applied to the sample covariance matrix of asset returns to stabilize the inverse of this covariance matrix in the selected strategy.

Let $\hat{\lambda}_1^2 \geq \hat{\lambda}_2^2 \geq ... \geq \hat{\lambda}_N^2 \geq 0$ be the eigenvalues of a sample covariance matrix $\hat{\Sigma}$. By spectral decomposition, we have that $\hat{\Sigma} = PDP'$ with $PP' = I_N$ where P is the matrix of eigenvectors and D the diagonal matrix with eigenvalues $\hat{\lambda}_j$ on the diagonal. Let also $\hat{\Sigma}^\alpha$ be the regularized inverse of $\hat{\Sigma}$.

$$\hat{\Sigma}^\alpha = PD^\alpha P'$$

where D^α is the diagonal matrix with elements $q(\alpha, \hat{\lambda}_j^2)/\hat{\lambda}_j^2$. The positive parameter α is the regularization parameter, a kind of smoothing parameter which is unknown and need to be selected efficiently. $q(\alpha, \hat{\lambda}_j^2)$ is the damping function which depends on the regularization scheme used.

2.2.1 Tikhonov regularization (TH)

This regularization scheme is close to the well known ridge regression used in presence of multicolinearity to improve properties of OLS estimators. In Tikhonov regularization scheme, the real function $q(\alpha, \hat{\lambda}_j^2)$ is given by

$$q(\alpha, \hat{\lambda}_j^2) = \frac{\hat{\lambda}_j^2}{\hat{\lambda}_j^2 + \alpha}$$

2.2.2 The spectral cut-off (SC)

It consists in selecting the eigenvectors associated with the eigenvalues greater than some threshold.

$$q(\alpha, \hat{\lambda}_j^2) = I\left\{\hat{\lambda}_j^2 \geq \alpha\right\}$$

The explosive influence of the factor $1/\hat{\lambda}_j^2$ is filtered out by imposing $q(\alpha, \hat{\lambda}_j^2) = 0$ for small $\hat{\lambda}_j^2$, that is $\hat{\lambda}_j^2 < \alpha$. α is a positive regularization parameter such that no bias is introduced when $\hat{\lambda}_j^2$ exceeds the threshold α. Another version of this regularization scheme is the Principal Components (PC) which consists in using a certain number of eigenvectors to compute the inverse of the operator. The PC and the SC are perfectly equivalent, only the definition of the regularization term α differs. In the PC, α is the number of principal components. In practice, both methods will give the same estimator.

2.2.3 Landweber Fridman regularization (LF)

In this regularization scheme, $\hat{\Sigma}^{\alpha}$ is computed by an iterative procedure with the formula

$$\begin{cases} \hat{\Sigma}_l^{\alpha} = \left(I_N - c\hat{\Sigma}^{\alpha}\right)\hat{\Sigma}_{l-1} + c\hat{\Sigma} \quad \text{for } l = 1, 2, ...1/\alpha - 1 \\ \hat{\Sigma}_0^{\alpha} = c\hat{\Sigma} \end{cases}$$

The constant c must satisfy $0 < c < 1/\hat{\lambda}_1^2$. Alternatively, we can compute this regularized inverse with

$$q(\alpha, \hat{\lambda}_j^2) = 1 - \left(1 - c\hat{\lambda}_j^2\right)^{\frac{1}{\alpha}}$$

The basic idea behind this procedure is similar to spectral cut-off but with a smooth bias function. See Carrasco, Florens, and Renault (2007) for more details on these regularization techniques.

3 First level of regularization

The fundamental objective of this first regularization is to consistently estimate r_c that appears in the selected portfolio without imposing the normality assumption on the return distribution. Kan and Zhou (2007) propose a consistent estimator for r_c under normality assumption on returns. In our framework, we do not impose such a normality instead we decide to stabilize the inverse of the covariance matrix by regularization when estimating r_c. By definition, we have that

$$\theta = \mu' \Sigma^{-1} \mu$$

Hence, this quantity can be estimated by regularization as follows

$$\hat{\theta}_{\alpha_1} = \hat{\mu}' \hat{\Sigma}^{\alpha_1} \hat{\mu} \tag{6}$$

where $\hat{\Sigma}^{\alpha_1}$ is the regularized inverse of the sample covariance matrix $\hat{\Sigma}$, α_1 the tuning parameter of the first regularization, and $\hat{\mu}$ the sample mean. Hence,

$$\hat{r}_{c,\alpha_1} = \sigma \frac{1 + \hat{\theta}_{\alpha_1}}{\sqrt{\hat{\theta}_{\alpha_1}}} \tag{7}$$

Let's now look at the consistency of $\hat{r}_{c,\alpha_1}$.

3.1 Consistency of $\hat{r}_{c,\alpha_1}$

To show the consistency of $\hat{r}_{c,\alpha_1}$ we will need some regularity conditions. In particular we need the following assumption.

Assumption A: $\frac{\Sigma}{N}$ is a trace class operator.

A a trace class operator K is a compact operator with a finite trace i.e $Tr\left(K\right) = O\left(1\right)$. This assumption is more realistic than assuming that Σ is a Hilbert-Schmidt operator. Moreover, Carrasco, Koné, and Noumon (2019) show that Assumption A holds for a standard factor model. This assumption implies in particular that

$$\left\| \frac{\Sigma}{N} \right\| = O\left(1\right).$$

Hence, under Assumption A, we have by Theorem 4 of Carrasco and Florens (2000) that

$$\left\| \frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N} \right\| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Proposition 1 *Under Assumption A we have that*

$$\hat{r}_{c,\alpha_1} = \sigma \frac{1 + \hat{\theta}_{\alpha_1}}{\sqrt{\hat{\theta}_{\alpha_1}}} \to_p r_c = \sigma \frac{1 + \theta}{\sqrt{\theta}} \tag{8}$$

if $\alpha_1 \sqrt{T} \to \infty$, *where* α_1 *is the tuning parameter.*

To show this proposition, we need only to show that $\hat{\theta}_{\alpha_1} \to_p \theta$ under assumption A so that we can use the continuous mapping theorem to conclude about the consistency of $\hat{r}_{c,\alpha_1}$. Proof of proposition 1 can be found in section 8. The regularity condition $\alpha_1\sqrt{T} \to \infty$ behind this proposition implies that the estimation window should go to infinity faster than the optimal tuning parameter α_1 goes to zero. We consistently estimate r_c without imposing the normality assumption on the return distribution. Moreover, we do not need a consistent estimate of μ to estimate r_c. In fact, to estimate r_c, we use the sample mean to estimate the expected return μ which is a non consistent estimator in or framework.

3.2 Data-driven selection of the tuning parameter α_1

We see in the previous subsection that $\hat{r}_{c,\alpha_1}$ depends on a certain smoothing parameter $\alpha_1 \in (0,1)$. We show the consistency of $\hat{r}_{c,\alpha_1}$ assuming that this tuning parameter is given. However, in practice, this regularization parameter is unknown and needs to be selected in an optimal way. Hence, we propose a data-driven selection procedure to obtain an approximation of this parameter.

Our objective here is to select the tuning parameter that minimizes the distance between the inverse of the estimated covariance matrix and the inverse of the true covariance matrix as in Koné (2021). More precisely, the following loss function has been used

$$\mu' \left[\left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1} \right)' \Sigma \left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1} \right) \right] \mu \tag{9}$$

where μ is the expected excess return. Hence, the objective is to select the tuning parameter that minimizes

$$E \left\{ \mu' \left[\left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1} \right)' \Sigma \left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1} \right) \right] \mu \right\} \tag{10}$$

which implies that

$$\hat{\alpha}_1 = arg \min_{\alpha_1 \in H_T} E \left\{ \mu' \left[\left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1} \right)' \Sigma \left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1} \right) \right] \mu \right\} \tag{11}$$

To obtain a better approximation of the tuning parameter based on a generalized cross-validation criterion, we need additional assumption. So, let start with some useful notations.

Let us denote $\beta = \Omega^{-1}\mu = E\left(R'R\right)^{-1} E\left(R'1_T\right)$. The following assumption help us to obtain a better approximation of the tuning parameter based on a generalized cross-validation criterion.

Assumption B: For some $\nu > 0$, we have that

$$N \sum_{j=1}^{N} \frac{<\beta, \phi_j>^2}{\eta_j^{2\nu}} < \infty$$

where ϕ_j and η_j^2 denote the eigenvectors and eigenvalues of $\frac{\Omega}{N}$.

The regularity condition in Assumption B can be found in Carrasco, Florens, and Renault (2007) and Carrasco (2012). Moreover, Carrasco, Koné, and Noumon (2019) show that assumption B hold if the returns are generated by a factor model. Assumption B is used combined with Assumption A to derive the rate of convergence of the mean squared error in the OLS estimator of β. These two assumptions imply in particular that $\|\beta\|^2 < +\infty$ such that we have the following relations

$$\|\beta - \beta_{\alpha_1}\|^2 = \begin{cases} O\left(\frac{\alpha_1^{\nu+1}}{N}\right) & \text{for } SC, LF \\ O\left(\frac{\alpha_1^{\min(\nu+1,2)}}{N}\right) & \text{for } T \end{cases}$$

β_{α_1} is the regularized version of β.

Using Assumption A combined with Assumption B, we obtain the following equivalent of the objective function

$$E\left\{\mu' \left[\left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1}\right)' \Sigma \left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1}\right)\right]\mu\right\} \sim \frac{1}{T} E\left\|X\left(\hat{\beta}_{\alpha_1} - \beta\right)\right\|^2 + \frac{\left(\mu'\left(\beta_{\alpha_1} - \beta\right)\right)^2}{(1-\mu'\beta)}$$

if $\frac{1}{\alpha_1^2 T} \to 0$ *as* T *goes to infinity.*

This equivalent is obtained using a combination of Proposition 2 from Koné (2021) and Proposition 1 in Carrasco, Koné, and Noumon (2019). We can easily apply a cross-validation approximation procedure on this expression of the objective function. It follows from this approximation that minimizing $E\left\{\mu' \left[\left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1}\right)' \Sigma \left(\hat{\Sigma}^{\alpha_1} - \Sigma^{-1}\right)\right]\mu\right\}$ with respect to α_1 is equivalent to minimizing

$$\frac{1}{T}E\left\|X\left(\hat{\beta}_{\alpha_1}-\beta\right)\right\|^2 \tag{12}$$

$$+\frac{\left(\mu'\left(\beta_{\alpha_1}-\beta\right)\right)^2}{\left(1-\mu'\beta\right)}. \tag{13}$$

Terms (12) and (13) depend on the unknown β and hence need to be approximated. These quantities will be approximated according to Carrasco, Koné, and Noumon (2019). In effect, the rescaled MSE

$$\frac{1}{T}E\left[\left\|X\left(\widehat{\beta}_{\alpha_1}-\beta\right)\right\|^2\right]$$

can be approximated by generalized cross validation criterion:

$$GCV\left(\alpha_1\right)=\frac{1}{T}\frac{\left\|\left(I_T-M_T\left(\alpha_1\right)\right)1_T\right\|^2}{\left(1-tr\left(M_T\left(\alpha_1\right)\right)/T\right)^2}.$$

Using the fact that

$$\hat{\mu}'\left(\beta_{\alpha_1}-\beta\right)=\frac{1_T'}{T}\left(M_T\left(\alpha_1\right)-I_T\right)X\beta,$$

(13) can be estimated by plug-in:

$$\frac{\left(1_T'\left(M_T\left(\alpha_1\right)-I_T\right)X\hat{\beta}_{\tilde{\alpha}}\right)^2}{T^2\left(1-\hat{\mu}'\hat{\beta}_{\tilde{\alpha}}\right)} \tag{14}$$

where $\hat{\beta}_{\tilde{\alpha}}$ is an estimator of β obtained for some consistent $\tilde{\alpha}$ ($\tilde{\alpha}$ can be obtained by minimizing $GCV\left(\alpha_1\right)$).

The optimal value of α_1 is defined as

$$\hat{\alpha_1}=\arg\min_{\alpha_1\in H_T}\left\{GCV\left(\alpha_1\right)+\frac{\left(1_T'\left(M_T\left(\alpha_1\right)-I_T\right)X\hat{\beta}_{\tilde{\alpha}}\right)^2}{T^2\left(1-\hat{\mu}'\hat{\beta}_{\tilde{\alpha}}\right)}\right\}$$

where $H_T=\{1,2,...,T\}$ for spectral cut-off and Landweber Fridman and $H_T=(0,1)$ for Ridge.

4 Second level of regularization

4.1 Estimation and asymptotic properties

In this first part we will focus on the estimation of the optimal portfolio and some asymptotic results such as the Sharpe ratio of the regularized portfolio.

Assume that r_c has been estimated by regularization according to section 3. We find in this situation that $\hat{r}_{c,\alpha_1}$ is consistent. We can then use this new estimator of r_c to form the selected portfolio in (3) combined with a second stabilization procedure to stabilize the inverse of the covariance matrix Ω. If we replace $\hat{r}_c$ by $\hat{r}_{c,\alpha_1}$ in (4), we still face an estimation issue due to the fact we use the sample counterpart of Ω^{-1} which is not appropriate in large dimensional problem as in our framework because may be nearly singular.

We use here three regularization methods to stabilize the inverse of Ω when estimating the optimal portfolio given in (3). Let us denote by $\hat{\Omega}^{\alpha_2}$ the regularized inverse of the sample covariance matrix $\hat{\Omega}$ with α_2 a positive tuning parameter specific to each regularization method. This parameter is also unknown and will need to be selected. Using this notation, our regularized portfolio is given by

$$\hat{\omega}_{\alpha_1,\alpha_2} = \hat{r}_{c,\alpha_1}\hat{\Omega}^{\alpha_2}\hat{\mu}. \tag{15}$$

$\hat{\Omega}^{\alpha_2}$ is computed according to the regularization technique considered in our estimation process. These regularization methods are summarized in section 2.2. We can notice that unlike what has been done in Carrasco, Koné, and Noumon (2019) and Koné (2021), the selected portfolio depends on two different tuning parameters α_1 and α_2.

The selected portfolio in (15) can then be use to compute the expected return on the optimal portfolio as follows

$$ER\left(\hat{\omega}_{\alpha_1,\alpha_2}\right) = \hat{\omega}'_{\alpha_1,\alpha_2}\mu = \hat{r}_{c,\alpha_1}\hat{\mu}'\hat{\Omega}^{\alpha_2}\mu. \tag{16}$$

Does this quantity converges to the true expected return $\omega'\mu$ under non- restrictive regularity conditions?

We obtain the following result under Assumption A.

Proposition 2 *Under Assumption A we have that*

$$ER\left(\hat{\omega}_{\alpha_1,\alpha_2}\right) = \hat{\omega}'_{\alpha_1,\alpha_2}\mu \to_p ER\left(\omega\right) = \omega'\mu \tag{17}$$

if $\alpha_1\alpha_2\sqrt{T} \to \infty$*, where* α_1 *and* α_1 *are the two tuning parameters of the double regularization. Moreover, we have that*

$$\hat{\omega}'_{\alpha_1,\alpha_2}\Sigma\hat{\omega}_{\alpha_1,\alpha_2} \to_p \omega'\Sigma\omega \tag{18}$$

Proof of this proposition can be found in the section 8. To obtain the result of Proposition 2, we do not need any assumption about the distribution of the excess return. We don't need the excess return to follow a normal distribution. The only useful assumption we need to derive this result is in Assumption A. Under this assumption, Proposition 2 gives us similar results as in Theorem 1 of Ao, Yingying, and Zheng (2019). According to this result, the double regularization portfolio that we implement guarantees to investors the maximum expected return with the lowest risk. In other words, the selected portfolio with our method achieves asymptotically the true Sharpe ratio.

Assumption A is less restrictive than imposing the normal distribution on the excess return. This assumption can easily be verified when the excess returns are generated by the three-factor model (see Carrasco, Koné, and Noumon (2019) for instance). The regularity condition behind the first part of Proposition 2 holds when the two tuning parameters are of order $1/T^{0.125}$. More precisely, if $\alpha_i \sim 1/T^{0.125}$ for $i = 1, 2$, we have that $\alpha_1\alpha_2\sqrt{T} \sim T^{0.125} \to \infty$ as $T \to \infty$.

In the next subsection we propose a way to select the tuning parameter of the second regularization method.

4.2 Selection of the tuning parameter of the second regularization technique

The results in Proposition 2 are derived assuming that the parameter α_2 is given. In practice, this parameter is unknown and needs to be estimated. Using the fact that $\beta = \Omega^{-1}\mu = E\left(R'R\right)^{-1} E\left(R'1_T\right)$, we have that

$$\hat{\omega}_{\alpha_1,\alpha_2} = \hat{r}_{c,\alpha_1}\hat{\Omega}^{\alpha_2}\hat{\mu} = \hat{r}_{c,\alpha_1}\hat{\beta}_{\alpha_2} \tag{19}$$

where $\hat{\beta}_{\alpha_2}$ is the regularized version of β.

According to Kan and Zhou (2007), the performance of a $\hat{\omega}_{\alpha_1,\alpha_2}$ can be measured by

$$U\left(\hat{\omega}_{\alpha_1,\alpha_2}\right) = \hat{\omega}^{'}_{\alpha_1,\alpha_2}\mu - \frac{\gamma}{2}\hat{\omega}^{'}_{\alpha_1,\alpha_2}\Sigma\hat{\omega}_{\alpha_1,\alpha_2}. \tag{20}$$

Hence, we propose as in Carrasco, Koné, and Noumon (2019) to select the tuning parameter α_2 that minimizes the following expected loss function

$$E\left(L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right]\right) = E\left[U\left(\omega\right) - U\left(\hat{\omega}_{\alpha_1,\alpha_2}\right)\right] \;\; with \tag{21}$$

$$L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right] = U\left(\omega\right) - U\left(\hat{\omega}_{\alpha_1,\alpha_2}\right) \tag{22}$$

More precisely, the tuning parameter α_2 is selected as follows

$$\hat{\alpha}_2 = \arg\min_{\alpha_2} E\left(L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right]\right). \tag{23}$$

The following result about the loss function $L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right]$ will be useful to correctly approximate the objective function of the optimization problem in (23)

Lemma 1 *Under Assumption A we have that*

$$L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right] \sim L\left[r_c\hat{\beta}_{\alpha_2}\right] \tag{24}$$

if $\alpha_1\alpha_2\sqrt{T} \to \infty$, *as* T *goes to infinity where* α_1 *and* α_1 *are the two tuning parameters of the double regularization.*

The result of this lemma comes directly from Proposition 2. According to this lemma, selecting α_2 with respect to $E\left(L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right]\right)$ is equivalent of selection this parameter with respect to $E\left(L\left[r_c\hat{\beta}_{\alpha_2}\right]\right)$. Hence, we will focus here on $E\left(L\left[r_c\hat{\beta}_{\alpha_2}\right]\right)$ as the objective function used to estimate the tuning parameter α_2.

$$
\begin{aligned}
L\left[r_c\hat{\beta}_{\alpha_2}\right] &= \left(r_c\beta'\mu - r_c\hat{\beta}'_{\alpha_2}\mu\right) - \frac{r_c^2\gamma}{2}\left(\hat{\beta}'_{\alpha_2}\Sigma\hat{\beta}_{\alpha_2} - \beta'\Sigma\beta\right) \\
&= r_c\left[\left(\beta - \hat{\beta}_{\alpha_2}\right)'\mu - \frac{r_c\gamma}{2}\left(\hat{\beta}'_{\alpha_2}\Sigma\hat{\beta}_{\alpha_2} - \beta'\Sigma\beta\right)\right] \\
&= \frac{r_c^2\gamma}{2}\left(\hat{\beta}_{\alpha_2} - \beta\right)'\Sigma\left(\hat{\beta}_{\alpha_2} - \beta\right)
\end{aligned}
$$

Hence, using this decomposition of the loss function $L\left[r_c\hat{\beta}_{\alpha_2}\right]$, we derive the following result that gives us a good approximation of the objective function $E\left(L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right]\right)$

Proposition 3 *Under Assumptions A and B we have that*

$$
\left(\frac{r_c^2\gamma}{2}\right)^{-1} E\left(L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right]\right) = \frac{1}{T}E\left\|R\left(\hat{\beta}_{\alpha_2} - \beta\right)\right\|^2 - \left(\mu'\left(\beta_{\alpha_2} - \beta\right)\right)^2 + rest\left(\alpha_2, N, T\right) \quad (25)
$$

where

$$
rest\left(\alpha_2, N, T\right) = O_p\left[\frac{1}{\sqrt{T}}\left(\frac{N}{\alpha_2^2 T} + \frac{\alpha_2^{\nu}}{N} + \frac{\alpha_2^{\nu/2-1}}{\sqrt{T}}\right)\right]. \quad (26)
$$

Moreover, if $\alpha_2^2\sqrt{T} \to \infty$ and $\alpha_2^{1-\nu/2}T \to \infty$ as T goes to infinity then we obtain the following approximation

$$
\left(\frac{r_c^2\gamma}{2}\right)^{-1} E\left(L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right]\right) \sim \frac{1}{T}E\left\|R\left(\hat{\beta}_{\alpha_2} - \beta\right)\right\|^2 - \left(\mu'\left(\beta_{\alpha_2} - \beta\right)\right)^2 \quad (27)
$$

Proof of this proposition can be found in section 8. To show this result, we use some results from Carrasco, Koné, and Noumon (2019) particularly the result of Lemma 3. According to Proposition 3, under appropriate regularity conditions, selecting the tuning parameter α_2 with respect to (23) is equivalent of selecting this parameter base on the right side of (27). This new objective function can easily be approximated by generalized cross-validation using the same technique as in Carrasco, Koné, and Noumon (2019). In fact, according to Proposition 3 $\hat{\alpha}_2$ is obtained by

$$
\hat{\alpha}_2 = \arg\min_{\alpha_2}\left\{\frac{1}{T}E\left\|R\left(\hat{\beta}_{\alpha_2} - \beta\right)\right\|^2 - \left(\mu'\left(\beta_{\alpha_2} - \beta\right)\right)^2\right\}. \quad (28)
$$

The objective function is unknown because depends on β hence, will be approximated by cross validation as follows

$$GCVF(\alpha_2) = \frac{1}{T}\frac{\|(I_T - M_T(\alpha_2))\,1_T\|^2}{(1 - tr(M_T(\alpha_2))/T)^2} - \frac{\left(1_T'(M_T(\alpha_2) - I_T)\,X\hat{\beta}_{\widetilde{\alpha}}\right)^2}{T^2}$$

where $\hat{\beta}_{\widetilde{\alpha}}$ is an estimator of β obtained for some consistent $\tilde{\alpha}$.

$$M_T(\alpha_2)\,v = \sum_{j=1}^{T} q(\alpha_2, \lambda_j^2)\left(\frac{v'\psi_j}{T}\right)\psi_j$$

for any T-dimensional vector v and $tr(M_T(\alpha_2)) = \sum_{j=1}^{T} q(\alpha_2, \lambda_j^2)$ and ψ_j the eigenvectors of RR'/T. Hence, the tuning parameter is selected as

$$\hat{\alpha}_2 = \arg\min_{\alpha_2} GCVF(\alpha_2)\,. \tag{29}$$

The optimality of this selection is proofed in Carrasco, Koné, and Noumon (2019).

4.3 Mean squared error

The aim here is to see if we can better control the estimation error when using a double regularization to estimate the mean-variance portfolio. For this purpose, we derive an approximation to the estimation error in the optimal portfolio to understand if this mean squared error vanishes asymptotically under less restrictive regularity conditions.

We define the mean squared error of the selected strategy as follows

$$MSE(\hat{\omega}_{\alpha_1,\alpha_2}) = \frac{1}{NT}E\left[\|R(\hat{\omega}_{\alpha_1,\alpha_2} - \omega)\|_2^2\right] \tag{30}$$

Using this definition of the MSE and under Assumptions A and B, we obtain the following result

Proposition 4 *Under Assumptions A and B we have the following result about the estimation error of the selected portfolio*

$$MSE(\hat{\omega}_{\alpha_1,\alpha_2}) \sim C(\alpha_1, N, T)\left[\frac{1}{T\alpha_2} + \alpha_2^{\nu+1}\right] + D(\alpha_1, N, T)\left[\frac{1}{T\alpha_2} + \alpha_2^{\nu+1}\right]^{1/2} + E(\alpha_1, N, T) \tag{31}$$

where

$$C(\alpha_1, N, T) = \frac{1}{N\alpha_1}\left(\frac{1}{\alpha_1\sqrt{T}} + 1\right)^2$$
$$D(\alpha_1, N, T) = \frac{1}{N\alpha_1}\left(1 + \frac{1}{\alpha_1\sqrt{T}}\right)$$
$$E(\alpha_1, N, T) = \frac{1}{NT\alpha_1^2}$$

Proof. In section 8.

This proposition implies that under appropriate regularity conditions, the MSE of the selected strategy by double regularization vanishes asymptotically. In other words, we have that

$$MSE(\hat{\omega}_{\alpha_1,\alpha_2}) \to 0$$

This implies that we asymptotically control the MSE of the selected portfolio by double regularization.

5 Simulations

We implement a simple simulation exercise to evaluate the performance of our procedure and compare it with the existing procedures. This comparison will be done using several statistics such as in-sample expected loss in utility and the Sharpe ratio.

Let us consider for this purpose a simple economy with $N \in \{10, 20, 40, 60, 80, 90, 100\}$ risky assets. We use several values of N to see how the size of the financial market (defined by the number of assets in the economy) could affect the performance of the selected strategy. Let T be the sample size used to estimate the unknown parameters in the investment process. Following Chen and Yuan (2016) and Carrasco, Koné, and Noumon (2019), we simulate the excess returns at each simulation step from the following three-factor model for $i = 1, ..., N$ and $t = 1, ..., T$

$$r_{it} = b_{i1}f_{1t} + b_{i2}f_{2t} + b_{i3}f_{3t} + \epsilon_{it} \tag{32}$$

$f_t = (f_{1t}, f_{2t}, f_{3t})'$ is the vector of common factors, $b_i = (b_{i1}, b_{i2}, b_{i3})'$ is the vector of factor loadings associated with the ith asset and ϵ_{it} is the idiosyncratic component of r_{it} satisfying $E(\epsilon_{it}|f_t) = 0$. We assume that $f_t \sim \mathcal{N}(\mu_f, \Sigma_f)$ where μ_f and Σ_f are calibrated on the monthly data of the market portfolio, the Fama-French size and the book-to-market portfolio from July 1980 to June 2016. Moreover, we assume that $b_i \sim \mathcal{N}(\mu_b, \Sigma_b)$ with μ_b and Σ_b calibrated using data of 30 industry portfolios from July 1980 to June 2016. Idiosyncratic terms ϵ_{it} are supposed to be normally distributed. The covariance matrix of the residual vector is assumed to be diagonal and given by Σ_ϵ=diag$(\sigma_1^2, ..., \sigma_N^2)$ with the diagonal elements drawn from a uniform distribution between 0.10 and 0.30 to yield an average cross-sectional volatility of 20%.

In the compact form (32) can be written as follows:

$$R = BF + \epsilon \tag{33}$$

where B is a $N \times 3$ matrix whose ith row is b_i'. The covariance matrix of the vector of excess return r_t is given by

$$\Sigma = B\Sigma_f B' + \Sigma_\epsilon.$$

The mean of the excess return is given by $\mu = B\mu_f$. The return on the risk-free asset R_f is calibrated to be the mean of the one-month T-B observed in the data from July 1980 to June 2016.

The calibrated parameters used in our simulation process are given in Table 1. The gross return on the risk-free asset calibrated on the data is given by $R_f = 1.0036$. Once generated, the factor loadings are kept fixed over replications, while the factors differ from simulations and are drawn from a trivariate normal distribution.

Let SR(ω) be the Sharpe ratio associated with the optimal portfolio ω, then SR(ω) is given as follows

$$SR(\omega) = \left[\mu' \Sigma \mu\right]^{1/2}$$

To evaluate the performance of our procedure in terms of the Sharpe ratio, we focus on

Table 1: **Calibrated parameters**

Parameters for factors loadings				Parameters for factors returns			
μ_b	Σ_b			μ_f	Σ_f		
1.0267	0.0422	0.0388	0.0115	0.0063	0.0020	0.0003	-0.0004
0.0778	0.0388	0.0641	0.0162	0.0011	0.0003	0.0009	-0.0003
0.2257	0.0115	0.0162	0.0862	0.0028	-0.0004	-0.0003	0.0009

the actual Sharpe ratio associated with the selected portfolio. If we have B replications in the simuation, the actual Sharpe ratio for a given replication is

$$SR(\hat{\omega}) = \frac{\hat{\omega}'\mu}{\left[\hat{\omega}'\Sigma\hat{\omega}'\right]^{1/2}}$$

Hence, the $SR(\hat{\omega})$ is estimated as follows

$$\hat{SR}(\hat{\omega}) = \frac{1}{B}\sum_{j=1}^{B} \frac{\hat{\omega^j}'\mu}{\left[\hat{\omega^j}'\Sigma\hat{\omega^j}\right]^{1/2}}$$

with $\hat{\omega^j}$ the selected portfolio at the replication j. To evaluate the performance of our procedure in terms of expected utility loss, we use the following loss fuction

$$L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right] = U(\omega) - U\left(\hat{\omega}_{\alpha_1,\alpha_2}\right)$$

where ω is the optimal portfolio and $\hat{\omega}_{\alpha_1,\alpha_2}$ the estimated portfolio by double regularization. Hence, in our simulation, the expected utility loss is estimated as follows

$$\hat{E}\left(L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right]\right) = \frac{1}{B}\sum_{j=1}^{B} L\left[\hat{\omega}^j_{\alpha_1,\alpha_2}\right]$$

We consider the following portfolio selection procedures: the naive portfolio (XoNP) which allocates a constant amount 1/N in each asset, our double regularized portfolios (DTH, DSC, DLF), the regularized portfolios proposed by Carrasco, Koné, and Noumon (2019) (TH, SC, LF), and the new Lasso estimator proposed by Ao, Yingying, and Zheng (2019) (AoP).

We did this analysis for 1000 replications. In Table 2 the mean squared error in estimating the aversion coefficient r_c for several number of risky assets in the economy.

Rows 2 to 4 contain the MSE when estimating r_c by regularization, row 5 the MSE obtained using the estimator of Ao, Yingying, and Zheng (2019) and row 6 the MSE for the sample-based estimate of r_c. The result in this table shows that by regularizing the sample covariance, one considerably reduces the estimation error when estimating the parameter r_c. When the number of assets in the economy increases, the regularized estimate of r_c considerably dominates the one proposed by Ao, Yingying, and Zheng (2019) in term of reducing the MSE. This in turn may help investors to improve the performance of the selected strategy that used the regularized estimation of r_c.

Table 2: **Mean squared error in estimating the parameter r_c for $T = 120$ and** $\sigma = 0.04$

N	10	20	40	60	80	90	100
Ridge	0.0025	0.0017	0.0022	0.0032	0.0067	0.0218	0.0230
SC	0.0023	0.0018	0.000437	0.0016	0.0045	0.0175	0.0210
LF	0.0013	0.0018	0.0013	0.0023	0.0047	0.0186	0.0214
Ao estimate	0.0017	0.0038	0.0104	0.0241	0.0436	0.0824	0.0945
Sample based estimate	0.0105	0.0559	0.2935	0.9053	4.8615	10.6255	12.702

Table 3 contains the Sharpe ratio obtained for several strategies. Rows 2 to 4 contain the result of our double regularized portfolio, row 5 contains the result of the new Lasso estimator of Ao, Yingying, and Zheng (2019), the result of the naive portfolio is in row 6 and the theoretical Sharpe ratio is in row 10.

The result of this simulation shows that the double regularization considerably improves the performance of the selected strategy in terms of the Sharpe ratio compared to what is obtained with the new Lasso estimator of Ao, Yingying, and Zheng (2019). In fact, the first level of the double regularization helps to reduce the estimation error when estimating r_c which in turn helps the investor to increase the Sharpe ratio compared to that of the Lasso estimator of Ao, Yingying, and Zheng (2019). Moreover, the fact that our tuning parameter is optimally selected can also explain the performance of our method over the new Lasso procedure.

Our method also outperforms the naive strategy for any set of risky assets considered in the simulation process.

The comparison of the double regularization with the traditional regularization will be done only for the expected utility loss.

Table 3: **The average monthly Sharpe ratio with $T = 120$ and $\sigma = 0.04$ over 1000 replications. Theo Sharpe R is the theoretical Sharpe ratio.**

N	10	20	40	60	80	90	100
DTH	0.4927	0.5201	0.5807	0.6272	0.6602	0.6509	0.7088
DSC	0.5189	0.5469	0.5705	0.6307	0.6482	0.6607	0.7185
DLF	0.4876	0.5581	0.5781	0.6289	0.6502	0.6508	0.7209
AoP	0.4562	0.4604	0.4687	0.5298	0.5589	0.5786	0.5980
XoNP	0.4021	0.4107	0.4209	0.4490	0.4705	0.4820	0.5072
Theo Sharpe R	0.5282	0.5652	0.6021	0.6521	0.6798	0.6887	0.7409

Another interesting result is the performance of the selected portfolio in term of minimizing the expected utility loss. The result of this analysis is given in Table 4.

Table 4: **The Actual Loss in utility $T = 120$ and $\sigma = 0.04$ over 1000 replications.**

N	10	20	40	60	80	90	100
DTH	0.0031	0.0028	0.0033	0.0048	0.0054	0.0078	0.0094
DSC	0.0028	0.0027	0.0034	0.0049	0.0056	0.0064	0.008
DLF	0.0027	0.0029	0.0032	0.0047	0.0057	0.0061	0.0078
TH	0.0076	0.0072	0.008	0.0086	0.0094	0.0102	0.0126
SC	0.0067	0.0071	0.0087	0.0092	0.0097	0.0120	0.0142
LF	0.0069	0.0070	0.0082	0.0085	0.0095	0.0102	0.0126
AoP	0.0078	0.008	0.0089	0.0101	0.0120	0.0150	0.0168
XoNP	0.009	0.0095	0.0107	0.0148	0.0165	0.0219	0.0263

The result of this table follows the same principle as in Table 3 where the Sharpe ratio is replaced by the expected utility loss. Rows 5 to 7 contain the result obtained by applying the traditional regularization as developed by Carrasco, Koné, and Noumon (2019) to estimate the optimal portfolio in the investment process.

In terms of comparison of the double regularization with the new Lasso procedure, we obtain similar results as with the actual Sharpe ratio. According to these results the investor gains more in terms of reducing the utility loss when using the double regularization to estimate the optimal portfolio in the investment process. The observed result seems to be plausible in the sense that in additional to considerably reducing the estimation error in estimating r_c, the tuning parameter of the second level of this double regularization is selected to minimize the expected utility loss. Hence, our method is highly recommended for investors because optimizes simultaneously the performance in term of utility cost and the Sharpe ratio.

Our method also outperforms the traditional regularized portfolio in terms of minimizing the expected utility loss. This observed performance of the double regularized portfolio can essentially be explained by the fact that a regularized estimate of r_c is used which significantly reduces the estimation error in the optimal solution. The result shows in particular that investors gain more in applying a double regularization when selecting the optimal portfolio in the financial market.

6 Empirical study

Our objective here is to use the real data (unlike in the simulation part) to estimate the unknown parameters of the optimal portfolio and then evaluate the performance of each estimation procedure.

We apply our method to several sets of portfolios from Kenneth R. French's website. We apply our procedure to the following portfolios: the 30-industry portfolios and the 100 portfolios formed on size and book-to-market. We allow investors to re-balance their portfolios every month. The investor holds this portfolio for one month, realizes gains and losses, updates information, and then recomputes the optimal portfolio weights for the next period using the same estimation window. This procedure is repeated each month, generating a time series of out-of-sample returns. This time series can then be used to analyze the out-of-sample performance of each strategy based on several statistics such as the out-of-sample Sharpe ratio. For this purpose, we use data from July 1980 to June 2016.

Table 5: **Out-of-sample performance in terms of the Sharpe ratio applied on the 30 industry portfolios (FF30) and the 100 portfolios formed on size and book-to-market (FF100) for two estimation windows. The risk constraint is $\sigma = 0.04$.**

P	EW	DTH	DSC	DLF	TH	SC	LF	AoP	XoNP
FF30	60	0.270	0.263	0.282	0.231	0.228	0.245	0.237	0.209
	120	0.289	0.284	0.320	0.248	0.235	0.268	0.258	0.219
FF100	120	0.328	0.316	0.339	0.243	0.242	0.254	0.249	0.247
	240	0.357	0.346	0.361	0.264	0.272	0.276	0.278	0.260

Table 5 contains the results of the out-of-sample analysis in terms of the Sharpe ratio for two different data sets: the FF30 and the FF100. These empirical results confirm what we have obtained in the simulation part. The result shows that by applying a double regularization to estimate the efficient mean-variance portfolio, we considerably improve the out-of-sample performance of the selected strategy in terms of maximizing the Sharpe ratio. Moreover, our regularized strategies outperform the new Lasso method of Ao, Yingying, and Zheng (2019), the traditional regularized portfolio proposed by Carrasco, Koné, and Noumon (2019) and the Equal-Weight portfolio for each data set. By outperforming the traditional regularized portfolio, these empirical results imply that the first level of regularization in estimating the efficient mean-variance portfolio plays a significant role in achieving a desire level of performance in the investment process. In other words, the estimation error from replacing r_c with its sample counterpart generate a significant loss in performance in the portfolio selection process. Therefore, it will be helpful to use a more appropriate estimator (which should replace the sample counterpart of r_c) to reduce the loss of performance due to the sample estimate particularly when the number of assets considered in the economy is large as in our framework. Hence, Ao, Yingying, and Zheng (2019) propose an alternative estimator for r_c which is unbiased and consistent under the normality assumption of asset return distribution. Nonetheless, we show by simulation that by stabilizing the inverse of the covariance matrix in estimating r_c one significantly reduces the estimation error compared to what has been proposed by

Ao, Yingying, and Zheng (2019) in a large financial market. Moreover, the properties of the proposed estimator Ao, Yingying, and Zheng (2019) depend on a distributional assumption: the normality of assets return. Therefore, the regularized estimate of r_c seems to be more appropriate since we do not need any distributional assumption for this estimation to perform. So, an interesting advantage of the double regularized portfolio is its ability to correctly perform well under a more general distributional assumption including the case of normality return.

Table 6: **Out-of-sample performance in terms of the risk applied on the 30 industry portfolios (FF30) and the 100 portfolios formed on size and book-to-market (FF100) for two estimation windows. The risk constraint is $\sigma = 0.04$.**

P	EW	DTH	DSC	DLF	TH	SC	LF	AoP	XoNP
FF30	60	0.0520	0.0473	0.0467	0.0530	0.0542	0.0523	0.0497	0.0517
	120	0.0471	0.0465	0.0455	0.0521	0.0510	0.0512	0.0482	0.0485
FF100	120	0.0502	0.0473	0.0461	0.0540	0.0530	0.0512	0.0512	0.0529
	240	0.0463	0.0462	0.0447	0.0527	0.0520	0.0471	0.0483	0.0498

In Table 6 we have the optimal risk taken by investors in the financial market. We find a similar result with the new Lasso technique in terms of controlling the risk when taking positions in the financial market. The optimal risk taken by the investor is very close to the risk constraint imposed in the economy.

Our regularized strategy may be very useful for investors especially during periods of high uncertainty in the financial as what we currently observe due to Covid-19.

7 Conclusion

The paper addresses the traditional estimation issue that exists in estimating the efficient mean-variance portfolio. More precisely, the efficient mean-variance is estimated by a double regularization where each level consists of stabilizing the inverse of the covariance matrix of assets using several regularization techniques from inverse problem literature. These regularization techniques namely the ridge, the spectral cut-off, and

Landweber-Fridman involve a regularization parameter or penalty term whose optimal value is selected to minimize the expected distance between the inverse of the estimated covariance matrix and the inverse of the true covariance matrix at the first level. The tuning parameter of the second level of regularization is selected to minimize the expected loss in utility of a mean-variance investor. The optimality of these tuning parameters selection procedures has been largely augmented in the literature.

Under appropriate regularity conditions, we show that the double regularized portfolio guarantees to investors the maximum expected return with the lowest risk. This implies that our selected portfolio achieves asymptotically the true Sharpe ratio. Unlike what has been obtained in the literature, we do not need the normality of asset returns to find this result. Moreover, we show based on a simple approximation of the mean squared error that the estimation error in estimating the efficient mean-variance portfolio vanishes asymptotically.

To evaluate the performance of our procedures we implement a simulation exercise based on a three-factor model calibrated on real data from the US financial market from July 1980 to June 2016.

We obtain by simulation that our procedure significantly improves the performance of the selected strategy with respect to the Sharpe ratio and the expected utility loss. The double regularized portfolios are compared to the new Lasso portfolio, the traditional regularized portfolio, and the naive 1/N strategy in terms of in-sample utility loss and the Sharpe ratio, and it is shown that our method yields significant Sharpe ratio improvements and considerably reduces the expected utility loss.

To confirm our simulations, we do an empirical analysis using Kenneth R. French's 30-industry portfolios and 100 portfolios formed on size and book-to-market. According to this empirical result, by double regularizing the efficient mean-variance portfolio, we considerably improve the performance of the selected strategy in terms of maximizing the Sharpe ratio.

Hence, we highly recommend this double regularized portfolio to the mean-variance investors in the sense that it optimizes simultaneously the performance in terms of utility loss and the Sharpe ratio with the lowest risk in the investment process.

References

Ao, M., L. Yingying, and X. Zheng (2019): "Approaching mean-variance efficiency for large portfolios," *The Review of Financial Studies*, 32(7), 2890–2919.

Bai, Z., H. Liu, and W.-K. Wong (2009): "Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory," *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 19(4), 639–667.

Bauder, D., T. Bodnar, N. Parolya, and W. Schmid (2020): "Bayesian inference of the multi-period optimal portfolio for an exponential utility," *Journal of Multivariate Analysis*, p. 104544.

Bodnar, T., A. K. Gupta, and N. Parolya (2016): "Direct shrinkage estimation of large dimensional precision matrix," *Journal of Multivariate Analysis*, 146, 223–236.

Bodnar, T., O. Okhrin, and N. Parolya (2019): "Optimal shrinkage estimator for high-dimensional mean vector," *Journal of Multivariate Analysis*, 170, 63–79.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009): "Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns," *The Review of Financial Studies*, 22(9), 3411–3447.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009): "Sparse and stable Markowitz portfolios," *Proceedings of the National Academy of Sciences*, 106(30), 12267–12272.

Brown, L. D., L. H. Zhao, et al. (2012): "A geometrical explanation of Stein shrinkage," *Statistical Science*, 27(1), 24–30.

Carrasco, M. (2012): "A regularization approach to the many instruments problem," *Journal of Econometrics*, 170(2), 383–398.

Carrasco, M., and J.-P. Florens (2000): "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, 16(6), 797–834.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization," *Handbook of Econometrics*, 6, 5633–5751.

——— (2014): "Asymptotic normal inference in linear inverse problems," *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, 73(74), 140.

Carrasco, M., N. Koné, and N. Noumon (2019): "Optimal portfolio selection using regularization," .

Carrasco, M., and G. Tchuente (2015): "Regularized LIML for many instruments," *Journal of Econometrics*, 186(2), 427–442.

Chen, J., and M. Yuan (2016): "Efficient portfolio selection in a large market," *Journal of Financial Econometrics*, 14(3), 496–524.

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

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2020): "A transaction-cost perspective on the multitude of firm characteristics," *The Review of Financial Studies*, 33(5), 2180–2222.

El Karoui, N. (2010): "High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation," *The Annals of Statistics*, 38(6), 3487–3566.

Fan, J., Y. Fan, and J. Lv (2008): "High dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, 147(1), 186–197.

Fan, J., Y. Liao, and M. Mincheva (2011): "High dimensional covariance matrix estimation in approximate factor models," *Annals of statistics*, 39(6), 3320.

——— (2013): "Large covariance estimation by thresholding principal orthogonal complements," *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75(4).

Fan, J., J. Zhang, and K. Yu (2008): "Asset allocation and risk assessment with gross exposure constraints for vast portfolios," .

Fastrich, B., S. Paterlini, and P. Winker (2015): "Constructing optimal sparse portfolios using regularization methods," *Computational Management Science*, 12(3), 417–434.

Jagannathan, R., and T. Ma (2003): "Risk reduction in large portfolios: Why imposing the wrong constraints helps," *The Journal of Finance*, 58(4), 1651–1684.

Jorion, P. (1986): "Bayes-Stein estimation for portfolio analysis," *Journal of Financial and Quantitative Analysis*, 21(3), 279–292.

Kan, R., and G. Zhou (2007): "Optimal portfolio choice with parameter uncertainty," *Journal of Financial and Quantitative Analysis*, pp. 621–656.

Kone, N. (2020): "A Multi-Period Portfolio Selection in a Large Financial Market," Discussion paper.

Koné, N. (2021): "Regularized Maximum Diversification Investment Strategy," *Econometrics*, 9(1), 1.

Kress, R. (1999): "Singular integral equations," in *Linear Integral Equations*, pp. 94–124. Springer.

LEDOIT, O., AND M. WOLF (2003): "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *Journal of Empirical Finance*, 10(5), 603–621.

——— (2004a): "Honey, I shrunk the sample covariance matrix," *The Journal of Portfolio Management*, 30(4), 110–119.

——— (2004b): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, 88(2), 365–411.

——— (2017): "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks," *The Review of Financial Studies*, 30(12), 4349–4388.

——— (2018): "Optimal estimation of a large-dimensional covariance matrix under Stein's loss," *Bernoulli*, 24(4B), 3791–3832.

MARKOWITZ, H. (1952): "Portfolio selection," *The Journal of Finance*, 7(1), 77–91.

ROTHMAN, A. J., P. J. BICKEL, E. LEVINA, J. ZHU, ET AL. (2008): "Sparse permutation invariant covariance estimation," *Electronic Journal of Statistics*, 2, 494–515.

STEIN, C. (1956): "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," Discussion paper, STANFORD UNIVERSITY STANFORD United States.

TOULOUMIS, A. (2015): "Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional settings," *Computational Statistics & Data Analysis*, 83, 251–261.

8 Proofs

8.1 Proof of proposition 1

Let's first recall the idea behind the continuous mapping theorem for the convergence in probability. According to this theorem, if X_n is sequence of random variable so that $X_n \rightarrow_p X$ then $g(X_n) \rightarrow_p g(X)$ where g is a continuous function. The continuous mapping theorem states that continuous functions preserve limits even if their arguments are sequences of random variables. Now let's look at the consistency of $\hat{\theta}_{\alpha_1}$ under the assumption A.

$$\begin{aligned}
\hat{\theta}_{\alpha_1} &= \hat{\mu}'\hat{\Sigma}^{\alpha_1}\hat{\mu} \\
&= (\hat{\mu} - \mu + \mu)' \hat{\Sigma}^{\alpha_1} (\hat{\mu} - \mu + \mu) \\
&= (\hat{\mu} - \mu)' \hat{\Sigma}^{\alpha_1} (\hat{\mu} - \mu) + 2(\hat{\mu} - \mu)' \hat{\Sigma}^{\alpha_1}\mu + \mu'\hat{\Sigma}^{\alpha_1}\mu
\end{aligned}$$

$$\begin{aligned}
\left\| (\hat{\mu} - \mu)' \hat{\Sigma}^{\alpha_1} (\hat{\mu} - \mu) \right\| &= \left\| \left(\frac{\hat{\mu} - \mu}{\sqrt{N}}\right)' \left(\frac{\hat{\Sigma}}{N}\right)^{\alpha_1} \left(\frac{\hat{\mu} - \mu}{\sqrt{N}}\right) \right\| \\
&= O_P\left(\frac{\left\| \left(\frac{\hat{\mu}-\mu}{\sqrt{N}}\right) \right\|^2}{\alpha_1}\right) \\
&= O_P\left(\frac{1}{\alpha_1 T}\right) \; \textit{because} \; \|\hat{\mu} - \mu\|^2 = O_P\left(\frac{N}{T}\right)
\end{aligned}$$

$$\begin{aligned}
(\hat{\mu} - \mu)' \hat{\Sigma}^{\alpha_1}\mu &= \left\| \left(\frac{\hat{\mu} - \mu}{\sqrt{N}}\right)' \left(\frac{\hat{\Sigma}}{N}\right)^{\alpha_1} \left(\frac{\mu}{\sqrt{N}}\right) \right\| \\
&= O_P\left(\frac{1}{\alpha_1\sqrt{T}}\right) \; \textit{because} \; \|\hat{\mu} - \mu\|^2 = O_P\left(\frac{N}{T}\right) \; \textit{and} \; \|\mu\|^2 = O(N)
\end{aligned}$$

$$
\begin{aligned}
\mu'\hat{\Sigma}^{\alpha_1}\mu &= \mu'\left(\hat{\Sigma}^{\alpha_1} - \Sigma^{\alpha_1} + \Sigma^{\alpha_1}\right) \\
&= \mu'\left(\hat{\Sigma}^{\alpha_1} - \Sigma^{\alpha_1}\right)\mu + \mu'\Sigma^{\alpha_1}\mu \\
&= \mu'\hat{\Sigma}^{\alpha_1}\left(\Sigma - \hat{\Sigma}_{\alpha_1}\right)\Sigma^{\alpha_1} + \mu'\Sigma^{\alpha_1}\mu \\
&= \left(\frac{\mu}{\sqrt{N}}\right)'\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha_1}\left[\frac{\Sigma}{N} - \left(\frac{\hat{\Sigma}}{N}\right)_{\alpha_1}\right]\left(\frac{\Sigma}{N}\right)^{\alpha_1}\left(\frac{\mu}{\sqrt{N}}\right) \\
&= \mu'\Sigma^{-1}\mu + O_P\left(\frac{1}{\alpha_1\sqrt{T}}\right)
\end{aligned}
$$

Hence, we obtain that,

$$
\hat{\theta}_{\alpha_1} = \theta + O_P\left(\frac{1}{\alpha_1\sqrt{T}}\right)
$$

Therefore, if the tuning parameter α_1 is selected in such a way that $\alpha_1\sqrt{T} \to \infty$ then we obtain that

$$
\hat{\theta}_{\alpha_1} \to_p \theta
$$

The continuous mapping theorem can then be used to show the consistency of $\hat{r}_{c,\alpha_1}$.

8.2 Proof of proposition 2

We have that

$$
\begin{aligned}
ER\left(\hat{\omega}_{\alpha_1,\alpha_2}\right) &= \hat{\omega}'_{\alpha_1,\alpha_2}\mu \\
&= \hat{r}_{c,\alpha_1}\hat{\mu}'\hat{\Omega}^{\alpha_2}\mu \\
&= \hat{r}_{c,\alpha_1}\left(\hat{\mu} - \mu\right)'\hat{\Omega}^{\alpha_2}\mu + \hat{r}_{c,\alpha_1}\mu'\hat{\Omega}^{\alpha_2}\mu \\
&= \left(\hat{r}_{c,\alpha_1} - r_c\right)\left(\hat{\mu} - \mu\right)'\hat{\Omega}^{\alpha_2}\mu + r_c\left(\hat{\mu} - \mu\right)'\hat{\Omega}^{\alpha_2}\mu + \left(\hat{r}_{c,\alpha_1} - r_c\right)\mu'\hat{\Omega}^{\alpha_2}\mu + r_c\mu'\hat{\Omega}^{\alpha_2}\mu
\end{aligned}
$$

$$
\left\|\left(\hat{r}_{c,\alpha_1} - r_c\right)\left(\hat{\mu} - \mu\right)'\hat{\Omega}^{\alpha_2}\mu\right\| \leq \left\|\left(\hat{r}_{c,\alpha_1} - r_c\right)\right\|\left\|\left(\frac{\hat{\mu} - \mu}{\sqrt{N}}\right)\right\|\left\|\frac{\mu}{\sqrt{N}}\right\|\left\|\left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\right\|
$$

By Proposition 1, we have that

$$\|(\hat{r}_{c,\alpha_1} - r_c)\| = o_p(1)$$

under the assumption that $\alpha_1\sqrt{T} \to \infty$.

$$\left\|\left(\frac{\hat{\mu}-\mu}{\sqrt{N}}\right)\right\| = O_p\left(\frac{1}{\sqrt{T}}\right) \; and$$
$$\left\|\frac{\mu}{\sqrt{N}}\right\| = O(1).$$

Moreover, by definition we have that

$$\left\|\left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\right\| = O_p\left(\frac{1}{\alpha_2}\right)$$

Hence,

$$\left\|(\hat{r}_{c,\alpha_1} - r_c)\,(\hat{\mu}-\mu)'\,\hat{\Omega}^{\alpha_2}\mu\right\| = o_p(1).O_p\left(\frac{1}{\alpha_1\sqrt{T}}\right)$$

Therefore, if $\max_{1\leq i\leq 2}\left\{\alpha_i\sqrt{T}\right\} \to \infty$ then,

$$\left\|(\hat{r}_{c,\alpha_1} - r_c)\,(\hat{\mu}-\mu)'\,\hat{\Omega}^{\alpha_2}\mu\right\| = o_p(1)$$

$$\left\|r_c\,(\hat{\mu}-\mu)'\,\hat{\Omega}^{\alpha_2}\mu\right\| \leq \|r_c\|\left\|\left(\frac{\hat{\mu}-\mu}{\sqrt{N}}\right)\right\|\left\|\frac{\mu}{\sqrt{N}}\right\|\left\|\left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\right\|$$

Since $0 \leq r_c < \infty$, then we have that

$$\left\|r_c\,(\hat{\mu}-\mu)'\,\hat{\Omega}^{\alpha_2}\mu\right\| = O_p\left(\frac{1}{\alpha_2\sqrt{T}}\right)$$

A similar argument can be used for $(\hat{r}_{c,\alpha_1} - r_c)\,\mu'\hat{\Omega}^{\alpha_2}\mu$ and we obtain that

$$(\hat{r}_{c,\alpha_1} - r_c)\,\mu'\hat{\Omega}^{\alpha_2}\mu = O_p\left(\frac{1}{\alpha_1\alpha_2\sqrt{T}}\right) \; because \; \|(\hat{r}_{c,\alpha_1} - r_c)\| = O_p\left(\frac{1}{\alpha_1\sqrt{T}}\right) \; and \; \left\|\left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\right\| = O_p\left(\frac{1}{\alpha_2}\right)$$

$$r_c\mu'\hat{\Omega}^{\alpha_2}\mu = r_c\mu'\left(\hat{\Omega}^{\alpha_2} - \Omega^{\alpha_2}\right)\mu + r_c\mu'\Omega^{\alpha_2}\mu$$

$$r_c\mu'\left(\hat{\Omega}^{\alpha_2}-\Omega^{\alpha_2}\right)\mu = r_c\left(\frac{\mu}{\sqrt{N}}\right)'\left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\left[\frac{\Omega}{N}-\left(\frac{\hat{\Omega}}{N}\right)_{\alpha_2}\right]\left(\frac{\Omega}{N}\right)^{\alpha_2}\left(\frac{\mu}{\sqrt{N}}\right)$$

$$\left\|\left(\frac{\mu}{\sqrt{N}}\right)'\left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\right\| = O_p\left(\frac{1}{\alpha_2}\right)$$

$$\left\|\left[\frac{\Omega}{N}-\left(\frac{\hat{\Omega}}{N}\right)_{\alpha_2}\right]\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

by Lemma 4 of Carrasco and Florens (2000).

$$\left(\frac{\Omega}{N}\right)^{\alpha_2}\left(\frac{\mu}{\sqrt{N}}\right) = O(1),$$

And because $\alpha_2 \to \infty$ as $T \to \infty$, we have that

$$ER\left(\hat{\omega}_{\alpha_1,\alpha_2}\right) = ER\left(\omega\right) + O_p\left(\frac{1}{\alpha_1\alpha_2\sqrt{T}}+\frac{1}{\alpha_1\alpha_2 T}+\frac{1}{\alpha_2\sqrt{T}}\right)$$

Therefore if $\alpha_1\alpha_2\sqrt{T} \to \infty$, then

$$ER\left(\hat{\omega}_{\alpha_1,\alpha_2}\right) \to_p ER\left(\omega\right)$$

$$\hat{\omega}'_{\alpha_1,\alpha_2}\Sigma\hat{\omega}_{\alpha_1,\alpha_2} = \hat{r}_{c,\alpha_1}\hat{\mu}'\hat{\Omega}^{\alpha_2}\Sigma\hat{\Omega}^{\alpha_2}\hat{\mu}\hat{r}_{c,\alpha_1}$$

$$\begin{aligned}\hat{\mu}'\hat{\Omega}^{\alpha_2} &= \left(\hat{\mu}-\mu\right)'\hat{\Omega}^{\alpha_2}+\mu'\hat{\Omega}^{\alpha_2}\\ &= \left(\hat{\mu}-\mu\right)'\Omega^{\alpha_2}+\left(\hat{\mu}-\mu\right)'\left(\hat{\Omega}^{\alpha_2}-\Omega^{\alpha_2}\right)+\mu'\Omega^{\alpha_2}+\mu'\left(\hat{\Omega}^{\alpha_2}-\Omega^{\alpha_2}\right)\end{aligned}$$

$$\left\|\left(\hat{\mu}-\mu\right)'\Omega^{\alpha_2}\right\| = \left\|\left(\frac{\hat{\mu}-\mu}{N}\right)'\left(\frac{\Omega}{N}\right)^{\alpha_2}\right\| = O_p\left(\frac{1}{\sqrt{NT}}\right)$$

$$\left\|\left(\hat{\mu}-\mu\right)'\left(\hat{\Omega}^{\alpha_2}-\Omega^{\alpha_2}\right)\right\| = \left\|\left(\frac{\hat{\mu}-\mu}{N}\right)'\left[\left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}-\left(\frac{\Omega}{N}\right)^{\alpha_2}\right]\right\| = O_p\left[\frac{\left\|\frac{\hat{\Omega}}{N}-\frac{\Omega}{N}\right\|}{\alpha_2\sqrt{NT}}\right] = O_p\left[\frac{1}{\alpha_2\sqrt{N}T}\right]$$

A similar argument may help to obtain that

$$\left\|\mu'\left(\hat{\Omega}^{\alpha_2}-\Omega^{\alpha_2}\right)\right\| = O_p\left[\frac{1}{\alpha_2\sqrt{NT}}\right]$$

Hence as $\alpha_2 \to 0$, as $T \to \infty$, we have that

$$\hat{\mu}'\hat{\Omega}^{\alpha_2} = \mu'\Omega^{-1} + O_p\left[\frac{1}{\alpha_2\sqrt{NT}}\right]$$

We can easily show that,

$$\hat{r}_{c,\alpha_1}\hat{\mu}'\hat{\Omega}^{\alpha_2} = r_c\mu'\Omega^{-1} + O_p\left[\frac{1}{\alpha_1\sqrt{NT}} + \frac{1}{\alpha_2\sqrt{NT}} + \frac{1}{\alpha_1\alpha_2 T\sqrt{N}}\right]$$

$$\hat{\omega}'_{\alpha_1,\alpha_2}\Sigma\hat{\omega}_{\alpha_1,\alpha_2} = \left\{r_c\mu'\Omega^{-1} + O_p\left[\frac{1}{\alpha_1\sqrt{NT}} + \frac{1}{\alpha_2\sqrt{NT}} + \frac{1}{\alpha_1\alpha_2 T\sqrt{N}}\right]\right\}\Sigma. \left\{r_c\Omega^{-1}\mu + O_p\left[\frac{1}{\alpha_1\sqrt{NT}} + \frac{1}{\alpha_2\sqrt{NT}} + \frac{1}{\alpha_1\alpha_2 T\sqrt{N}}\right]\right\}$$

$$\hat{\omega}'_{\alpha_1,\alpha_2}\Sigma\hat{\omega}_{\alpha_1,\alpha_2} = r_c^2\mu'\Omega^{-1}\Sigma\Omega^{-1}\mu + O_p\left[\left\|\mu'\Omega^{-1}\Sigma\right\|\psi_{N,T,\alpha_1,\alpha_2}\right] + O_p\left[\|\Sigma\|\,\psi^2_{N,T,\alpha_1,\alpha_2}\right]$$

where

$$\psi_{N,T,\alpha_1,\alpha_2} = \frac{1}{\alpha_1\sqrt{NT}} + \frac{1}{\alpha_2\sqrt{NT}} + \frac{1}{\alpha_1\alpha_2 T\sqrt{N}}$$

$$O_p\left[\|\Sigma\|\,\psi^2_{N,T,\alpha_1,\alpha_2}\right] = O_p\left[N\psi^2_{N,T,\alpha_1,\alpha_2}\right] = O_p\left[\left(\sqrt{N}\psi_{N,T,\alpha_1,\alpha_2}\right)^2\right] = O_p\left[\left(\frac{1}{\alpha_1\sqrt{T}} + \frac{1}{\alpha_2\sqrt{T}} + \frac{1}{\alpha_1\alpha_2 T}\right)^2\right]$$

$$O_p\left[\left\|\mu'\Omega^{-1}\Sigma\right\|\psi_{N,T,\alpha_1,\alpha_2}\right] = O_p\left[\sqrt{N}\psi_{N,T,\alpha_1,\alpha_2}\right] = O_p\left[\frac{1}{\alpha_1\sqrt{T}} + \frac{1}{\alpha_2\sqrt{T}} + \frac{1}{\alpha_1\alpha_2 T}\right]$$

Therefore, under the assumption that $\frac{1}{\alpha_1\sqrt{T}} + \frac{1}{\alpha_2\sqrt{T}} + \frac{1}{\alpha_1\alpha_2 T} \to 0$ which is obtained if $\max_{1\leq i\leq 2}\alpha_i\sqrt{T} \to \infty$ and $\alpha_1\alpha_2 T \to \infty$, then we obtain that

$$\hat{\omega}'_{\alpha_1,\alpha_2}\Sigma\hat{\omega}_{\alpha_1,\alpha_2} \to r_c^2\mu'\Omega^{-1}\Sigma\Omega^{-1}\mu = \omega'\Sigma\omega$$

8.3 Proof of Proposition 3

By Lemma 1, selecting α_2 with respect to $E\left(L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right]\right)$ is equivalent of selecting α_2 with respect to $E\left(L\left[r_c\hat{\beta}_{\alpha_2}\right]\right)$ where $L\left[\hat{\omega}_{\alpha_1,\alpha_2}\right] = U(\omega) - U\left(\hat{\omega}_{\alpha_1,\alpha_2}\right)$.

$$
\begin{aligned}
L\left[r_c\hat{\beta}_{\alpha_2}\right] &= U\left(r_c\beta\right) - U\left(r_c\hat{\beta}_{\alpha_2}\right) \\
&= r_c\left(\beta'\mu - \hat{\beta}'_{\alpha_2}\mu\right) - \frac{\gamma r_c^2}{2}\left(\hat{\beta}'_{\alpha_2}\Sigma\hat{\beta}_{\alpha_2} - \beta'\Sigma\beta\right) \\
&= r_c\left\{\left(\beta - \hat{\beta}_{\alpha_2}\right)'\mu - \frac{\tilde{\gamma}}{2}\left(\hat{\beta}'_{\alpha_2}\Sigma\hat{\beta}_{\alpha_2} - \beta'\Sigma\beta\right)\right\} \ where\ \tilde{\gamma} = \gamma r_c \\
&= \frac{r_c^2\gamma}{2}\left(\hat{\beta}_{\alpha_2} - \beta\right)'\Sigma\left(\hat{\beta}_{\alpha_2} - \beta\right).
\end{aligned}
$$

$$
\left(\hat{\beta}_{\alpha_2} - \beta\right)'\Sigma\left(\hat{\beta}_{\alpha_2} - \beta\right) = \left(\hat{\beta}_{\alpha_2} - \beta\right)'\hat{\Sigma}\left(\hat{\beta}_{\alpha_2} - \beta\right) + \left(\hat{\beta}_{\alpha_2} - \beta\right)'\left(\Sigma - \hat{\Sigma}\right)\left(\hat{\beta}_{\alpha_2} - \beta\right)
$$

$$
\left(\hat{\beta}_{\alpha_2} - \beta\right)'\left(\Sigma - \hat{\Sigma}\right)\left(\hat{\beta}_{\alpha_2} - \beta\right) = \left(\hat{\beta}_{N,\alpha_2} - \beta_N\right)'\left(\frac{\Sigma}{N} - \frac{\hat{\Sigma}}{N}\right)\left(\hat{\beta}_{N,\alpha_2} - \beta_N\right) \ with\ \beta_N = \left(\frac{\Omega}{N}\right)^{-1}\mu.
$$

$$
\left(\hat{\beta}_{\alpha_2} - \beta\right)'\left(\Sigma - \hat{\Sigma}\right)\left(\hat{\beta}_{\alpha_2} - \beta\right) \leq \left\|\hat{\beta}_{N,\alpha_2} - \beta_N\right\|^2\left\|\frac{\Sigma}{N} - \frac{\hat{\Sigma}}{N}\right\|
$$

$$
\left\|\hat{\beta}_{N,\alpha_2} - \beta_N\right\| = \left\|\hat{\beta}_{N,\alpha_2} - \beta_{N,\alpha_2}\right\| + \left\|\beta_{N,\alpha_2} - \beta_N\right\|
$$

$$
\begin{aligned}
\hat{\beta}_{N,\alpha_2} - \beta_{N,\alpha_2} &= \left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\hat{\mu} - \left(\frac{\Omega}{N}\right)^{\alpha_2}\mu \\
&= \left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\left(\hat{\mu} - \mu\right) + \left[\left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2} - \left(\frac{\Omega}{N}\right)^{\alpha_2}\right]\mu \\
&= \left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\left(\hat{\mu} - \mu\right) + \left(\frac{\hat{\Omega}}{N}\right)^{\alpha_2}\left[\left(\frac{\Omega}{N}\right)_{\alpha_2} - \left(\frac{\hat{\Omega}}{N}\right)_{\alpha_2}\right]\left(\frac{\Omega}{N}\right)^{\alpha_2}\mu
\end{aligned}
$$

Hence, we have that

$$
\left\|\hat{\beta}_{N,\alpha_2} - \beta_{N,\alpha_2}\right\| = O_p\left[\frac{\sqrt{N}}{\alpha_2\sqrt{T}} + \frac{1}{\alpha_2\sqrt{T}}\right] = O_p\left[\frac{\sqrt{N}}{\alpha_2\sqrt{T}}\right]
$$

Therefore, using Assumption B, we obtain that

$$\left\|\hat{\beta}_{N,\alpha_2}-\beta_N\right\|=O_p\left[\frac{\sqrt{N}}{\alpha_2\sqrt{T}}+\frac{\alpha_2^{\nu/2}}{\sqrt{N}}\right]$$

This imples that

$$\left\|\hat{\beta}_{N,\alpha_2}-\beta_N\right\|^2=O_p\left[\frac{N}{\alpha_2^2T}+\frac{\alpha_2^{\nu}}{N}+\frac{\alpha_2^{\nu/2-1}}{\sqrt{T}}\right]$$

So,

$$\left(\hat{\beta}_{\alpha_2}-\beta\right)'\left(\Sigma-\hat{\Sigma}\right)\left(\hat{\beta}_{\alpha_2}-\beta\right)=O_p\left[\frac{1}{\sqrt{T}}\left(\frac{N}{\alpha_2^2T}+\frac{\alpha_2^{\nu}}{N}+\frac{\alpha_2^{\nu/2-1}}{\sqrt{T}}\right)\right]$$

If $\alpha-2^2\sqrt{T}\to\infty$ and $\alpha_2^{1-\nu/2}T\to\infty$ then we have

$$\left(\hat{\beta}_{\alpha_2}-\beta\right)'\left(\Sigma-\hat{\Sigma}\right)\left(\hat{\beta}_{\alpha_2}-\beta\right)=o_p\left(1\right)$$

$$\left(\hat{\beta}_{\alpha_2}-\beta\right)'\Sigma\left(\hat{\beta}_{\alpha_2}-\beta\right)=\left(\hat{\beta}_{\alpha_2}-\beta\right)'\hat{\Sigma}\left(\hat{\beta}_{\alpha_2}-\beta\right)+rest\left(\alpha_2,N,T\right)$$

with

$$rest\left(\alpha_2,N,T\right)=O_p\left[\frac{1}{\sqrt{T}}\left(\frac{N}{\alpha_2^2T}+\frac{\alpha_2^{\nu}}{N}+\frac{\alpha_2^{\nu/2-1}}{\sqrt{T}}\right)\right]$$

Following the same procedure as in Carrasco, Koné, and Noumon (2019), we find that

$$\left(\hat{\beta}_{\alpha_2}-\beta\right)'\Sigma\left(\hat{\beta}_{\alpha_2}-\beta\right)=\frac{1}{T}E\left\|R\left(\hat{\beta}_{\tau}-\beta\right)\right\|^2-\left(\mu'\left(\beta_{\tau}-\beta\right)\right)^2+rest\left(\alpha_2,N,T\right)$$

8.4 Proof of Proposition 4

By definition we have that

$$MSE\left(\hat{\omega}_{\alpha_1,\alpha_2}\right)=\frac{1}{NT}E\left[\left\|R\left(\hat{\omega}_{\alpha_1,\alpha_2}-\omega\right)\right\|_2^2\right]$$

$$\begin{aligned}
\left\|R\left(\hat{\omega}_{\alpha_1,\alpha_2}-\omega\right)\right\|_2 &= \left\|R\left(\hat{r}_{c,\alpha_1}\hat{\beta}_{\alpha_2}-r_c\beta\right)\right\|_2 \\
&= \left\|R\hat{r}_{c,\alpha_1}\left[\hat{\beta}_{\alpha_2}-\beta\right]+R\left[\hat{r}_{c,\alpha_1}-r_c\right]\beta\right\|_2 \\
&\leq \left\|R\hat{r}_{c,\alpha_1}\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2+\left\|R\left[\hat{r}_{c,\alpha_1}-r_c\right]\beta\right\|_2 \\
&= \left\|R\hat{r}_{c,\alpha_1}\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2+O_p\left(\frac{1}{\alpha_1}\right)
\end{aligned}$$

$$\left\|R\left(\hat{\omega}_{\alpha_1,\alpha_2}-\omega\right)\right\|_2^2=\left\|R\hat{r}_{c,\alpha_1}\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2^2+O_p\left(\frac{1}{\alpha_1}\left\|R\hat{r}_{c,\alpha_1}\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2+\frac{1}{\alpha_1^2}\right)$$

Hence,

$$MSE\left(\hat{\omega}_{\alpha_1,\alpha_2}\right)=\frac{1}{NT}E\left[\left\|R\hat{r}_{c,\alpha_1}\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2^2\right]+O\left(\frac{1}{NT\alpha_1}E\left[\left\|R\hat{r}_{c,\alpha_1}\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2\right]+\frac{1}{NT\alpha_1^2}\right)$$

$$\begin{aligned}
\left\|R\hat{r}_{c,\alpha_1}\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2 &\leq \left\|R\left(\hat{r}_{c,\alpha_1}-r_c\right)\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2+\left\|Rr_c\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2 \\
&\leq \left\|\left(\hat{r}_{c,\alpha_1}-r_c\right)\right\|\left\|R\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2+r_c\left\|R\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2 \\
&= O_P\left[\left(1+\frac{1}{\alpha_1\sqrt{T}}\right)\left\|R\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2\right]
\end{aligned}$$

$$\begin{aligned}
MSE\left(\hat{\omega}_{\alpha_1,\alpha_2}\right) &= \frac{1}{NT}\left(1+\frac{1}{\alpha_1\sqrt{T}}\right)^2E\left[\left\|R\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2^2\right] \\
&+ O\left(\frac{1}{NT\alpha_1}\left(1+\frac{1}{\alpha_1\sqrt{T}}\right)E\left[\left\|R\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2\right]+\frac{1}{NT\alpha_1^2}\right)
\end{aligned}$$

By Lemma 3 in Carrasco, Koné, and Noumon (2019), we have that

$$\frac{1}{T}E\left[\left\|R\left[\hat{\beta}_{\alpha_2}-\beta\right]\right\|_2^2\right]\sim\frac{1}{T\alpha_2}+N\alpha_2^{\nu+1}$$

Therefore, we have that

$$\begin{aligned}
MSE\left(\hat{\omega}_{\alpha_1,\alpha_2}\right) &\sim \frac{1}{N}\left(1+\frac{1}{\alpha_1\sqrt{T}}\right)^2\left(\frac{1}{T\alpha_2}+N\alpha_2^{\nu+1}\right)+\frac{1}{NT\alpha_1}\left(1+\frac{1}{\alpha_1\sqrt{T}}\right)\left(\frac{1}{T\alpha_2}+N\alpha_2^{\nu+1}\right)^{1/2} \\
&+ \frac{1}{NT\alpha_1^2}
\end{aligned}$$