Queen's Economics Department Working Paper No. 1439

A Multi-Period Portfolio Selection in a Large Financial Market

N'Golo Kone

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

8-2020

A Multi-Period Portfolio Selection in a Large Financial Market*

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

September 22, 2020

Abstract

This paper addresses a multi-period portfolio selection problem when the number of assets in the financial market is large. Using an exponential utility function, the optimal solution is shown to be a function of the inverse of the covariance matrix of asset returns. Nonetheless, when the number of assets grows, this inverse becomes unreliable, yielding a selected portfolio that is far from the optimal one. We propose two solutions to this problem. First, we penalize the norm of the portfolio weights in the dynamic problem and show that the selected strategy is asymptotically efficient. Second, we penalize the norm of the difference of successive portfolio weights in the dynamic problem to guarantee that the optimal portfolio composition does not fluctuate widely between periods. This second method helps investors to avoid high trading costs in the financial market by selecting stable strategies over time. Extensive simulations and empirical results confirm that our procedures considerably improve the performance of the dynamic portfolio.

Keywords: Dynamic portfolio selection, Large Market, Asymptotic efficiency, Regularization.

JEL Subject Classification: G11, C16, C52

*I am greatly indebted to Marine Carrasco for her invaluable guidance. I am grateful to Benoit Perron, René Garcia, Georges Dionne, Tony Wirjanto, Louphou Coulibaly, Lucienne Papite Talba, Zerbo Souleymane, and the members of the atelier de discussion des doctorants en économie de l'Université de Montréal, CIREQ PhD Students 2018 in Montréal, the Marcel Dagenais Econometrics Seminar 2019, and the SCSE 2019 conference for their helpful comments.

1 Introduction

Understanding investors' behavior in a dynamic setting is very important for preventing losses from unexpected market downturns in the financial market. Therefore, several papers[1] have been interested in the multi-period portfolio selection problem since the seminal work of Markowitz (1959), who extends the mean-variance paradigm to the dynamic setting. In this paper, we also address a multi-period portfolio selection problem by developing a novel econometric method to consistently estimate the optimal solution of this dynamic problem. We use exponential utility functions as did Bodnar et al. (2015b) and Bauder et al. (2020), who derive a closed form solution to the dynamic portfolio problem. This optimal solution is shown to be a function of the inverse of the covariance matrix and the expected return, which are unknown and need to be estimated. When the number of assets grows, the inverse of the covariance matrix becomes unreliable, yielding a selected portfolio that is far from the optimal one. This problem is amplified by estimation errors in the financial market[2].

Hence, this paper proposes two solutions to stabilize the inverse of the covariance matrix in the optimal solution. These methods are particularly useful when the number of assets in the financial market increases considerably compared with the estimation window.

First, we penalize the norm of the portfolio weights in the dynamic problem and derive a closed-form solution to this new optimization problem. This optimal solution is closely related to a Ridge regularization, which consists of adding a diagonal matrix to the volatility's matrix to reduce estimation errors. Under appropriate regularity conditions, we show the consistency of the selected strategy by this procedure[3]. More importantly, we demonstrate that this regularized portfolio is asymptotically efficient in terms of the Sharpe ratio. However, this method partially controls the estimation error in the optimal solution because it ignores the estimation error in the expected return, which may also be important when the number of assets in the financial market increases considerably.

Second, we propose an alternative method that consists of penalizing the norm of the difference of successive portfolio weights in the dynamic problem to guarantee that the optimal portfolio composition does not fluctuate widely between periods. We show, under appropriate regularity conditions, that we better control the estimation error in the optimal portfolio with this new procedure. In fact, this procedure introduces a second level of regularization to control for the estimation error in the expected return. Moreover, this second method helps investors to avoid high trading costs in the financial market by selecting stable strategies over time.

Each strategy involves an unknown tuning parameter that needs to be selected in an optimal way at each time point. We propose, for each method, a data-driven method for selecting this parameter.

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. We obtain by

[1] See, for instance, the studies by Merton (1969), Samuelson (1975), Elton and Gruber (1974), Brandt and Santa-Clara (2006), Basak and Chabakauri (2010), Li and Ng (2000), Bodnar et al. (2015a), Penev et al. (2019), Ma et al. (2019) among others

[2] The estimation error in the expected return might be important, especially in a large financial market. Stein (1956) and Brown et al. (2012) even argue that the usual estimator of the expected return should be inadmissible if the dimension is sufficiently large.

[3] With respect to the norm induced by the inner product in $\mathbb{R}^N$.

simulation that by imposing an appropriate constraint on the dynamic problem we significantly improve the performance of the selected strategy with respect to the Sharpe ratio, the turnover that can be seen as a measure of transaction costs, the ability to predict the default probability and the dynamic of the optimal wealth. Moreover, our methods outperform the Bayesian procedure proposed by Bauder et al. (2020) in the large financial market. 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. We considerably reduce the turnover as a measure of transaction costs by imposing a temporal stability constraint on the dynamic portfolio selection problem.

This paper is related to the large literature on high dimensional estimation problems in the financial market. Ledoit and Wolf (2003, 2004) propose to replace the covariance matrix by a weighted average of the sample covariance matrix and some structured matrix. Moreover, Ledoit and Wolf (2017) 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. Brodie et al. (2009) use the lasso method which consists of imposing a constraint on the sum of the absolute value of the portfolio weights. Brandt et al. (2009) and DeMiguel et al. (2020) model the portfolio weights directly as a function of the assets' characteristics to avoid the difficulties in the estimation of asset return moments. Carrasco et al. (2019) investigate various regularization techniques found in the inverse problem literature to stabilize the inverse of the sample covariance matrix. Recently, Ao et al. (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. Other procedures have been proposed by Touloumis (2015), and Bodnar et al. (2016) to estimate the asset volatility matrix. Jorion (1986) and Bodnar et al. (2019) propose to use a shrinkage estimation for the expected return, which seems to be more appropriate than the sample mean. Moreover, in a recent paper, Bauder et al. (2020) propose a Bayesian method to estimate a multi-period portfolio but their method is not designed to handle a large number of assets. Our contribution to this literature is to provide a new method to consistently select the optimal portfolio in a dynamic setting with many assets.

Our work is also related to the vast literature on linear inverse problems. Carrasco et al. (2007), and Carrasco et al. (2014) use various regularization techniques for estimation issues in linear inverse problems. Carrasco (2012), and Carrasco and Tchuente (2015) handle the many instruments problem in linear models by regularization. Instead of using regularization, we propose a new way to stabilize the inverse of the covariance matrix by penalizing the norm of the difference of successive portfolio weights in the dynamic portfolio selection problem.

The rest of the paper is organized as follows. Section 2 presents the economy and shows that the dynamic portfolio selection problem can be seen as a linear inverse problem. Section 3 imposes a constraint on the portfolio weights in the dynamic problem and derives a closed-form solution to this new problem. In Section 4, we impose a temporal stability constraint on the dynamic portfolio optimization. Section 5 gives some asymptotic properties of the selected strategy and proposes data-driven methods to select the optimal tuning parameter. Section 6 presents some simulation results and an empirical study. Section 7 concludes the paper.

2 The model and an empirical fact

2.1 The economic environment

We consider a simple economy with N risky assets with random returns vector $\bar{R}_{t+1}$ 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 the investment horizon. Let R_f denote 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 $r_{t+1} = \bar{R}_{t+1} - R_f 1_N$ be 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 that the excess returns are independent over time with the mean and the covariance matrix given by μ_t and Σ_t respectively. This conditional distribution is assumed to be a normal distribution as in Croessmann (2017) and Bauder et al. (2020). This means that $r_t \sim \mathcal{N}(\mu_t, \Sigma_t)$. Let us assume also that the population covariance matrix Σ is positive definite such that the true and unknown optimal solutions are well defined.

We consider an investor with a finite life horizon (T) who can trade on a basket of assets available in the financial market. The investor has an initial wealth given by A_0. Without loss of generality we assume that $A_0 = 1$.

Let $\omega_t = (\omega_{t,1}, ..., \omega_{t,N})'$ be the vector of portfolio weights determined at the time point t.

Definition. *In our economy a portfolio is defined as a list of weights ω_t and $1 - \omega_t' 1_N$ that represent the amount of the capital to be invested in the risky assets and the risk-free asset respectively.*

Short-selling is allowed in the financial market, i.e. the optimal weights could also be negative or could contain negative weights for some assets.

The return on the optimal portfolio is given by

$$R_{p,t+1} = \omega_t' \bar{R}_{t+1} + R_f \left(1 - \omega_t' 1_N\right) = R_f + \omega_t' r_{t+1}. \tag{1}$$

We assume in our model that the investor does not receive other sources of income. Hence, the law of motion of the investor's total wealth is given by

$$A_{t+1} = A_t R_{p,t+1} = A_t (R_f + \omega_t' r_{t+1}) \tag{2}$$

for $t = 0, ..., T-1$ with $A_0 = 1$.

Moreover, let us assume that $R_f > 1$. This assumption implies that at each time point t we have that $A_t \geq 0$.

The investor has to select a sequence of portfolio weights $\{\omega_s\}_0^{T-1}$ in order to maximize the expected utility of final wealth, i.e. $E_0(U(A_T))$. Here we choose $U(x) = -\exp(-\gamma x)$ to be an exponential utility function with $\gamma > 0$, which represents the CARA and determines the investor's attitude towards risk. Note that the normality of the excess return could be abandoned in favor of a quadratic utility function, which seems to be less adapted than the exponential utility (which uses normality to find the closed-form solution) in a portfolio selection framework. In fact, according to Pratt (1964), the coefficient of absolute risk aversion (ARA) should decrease or at least should not increase with wealth. Therefore, because the quadratic utility function implies that the ARA is increasing in wealth, the exponential utility function (with a constant ARA) becomes a better choice than the quadratic utility function in the portfolio selection problem. Hence, there is a certain trade-off between using a less adapted utility function without the normality assumption and using an exponential utility function combined with the normality of the excess return.

The investor's optimization problem is then given by

$$V(0, A_0) = \max_{\{\omega_s\}_0^{T-1}} E_0(U(A_T)), \tag{3}$$

where $V(0, A_0)$ represents the value function. The solution of this problem is obtained recursively starting from the last period using the following Bellman equation associated with the optimization problem:

$$V(t, A_t) = \max_{\omega_t} E_t \{V(t, A_{t+1})\} = \max_{\omega_t} E_t \left\{ V(t+1, A_t(R_t^f + r_{t+1}\omega_t^{'})) \right\}, \tag{4}$$

$t = 0, ..., T-1$ with the following terminal condition $V(T, A_T) = -\exp(-\gamma A_T)$. Following Bauder et al. (2020) the solution of the optimization problem in Equation (4) is given by

$$\omega_t = \gamma_t \Sigma_t^{-1} \mu_t, \tag{5}$$

with $\gamma_t = \left(\gamma A_t R_f^{T-t-1}\right)^{-1}$ for $t = 0, ..., T-1$, which can be seen as an adjusted risk aversion used to capture the effect of previous actions on the selected portfolio.

This optimal portfolio is very close to Markowitz's strategy. The only difference comes from the constant γ_t. Hence, the relative share of the risky assets in the optimal portfolio is the same as in Markowitz's portfolio but the part allocated to the risk-free asset is different. In fact, if the investment horizon is reduced to a single period, the solution coincides with the mean-variance portfolio given by

$$\omega = \frac{1}{\gamma} \Sigma^{-1} \mu.$$

Hence, the investor's preferences enter in the solution only through the scalar term $1/\gamma$. Investors differ only in the overall scale of their risky asset position, not in the composition of that position. Therefore, conservative investors (with a high γ) hold more of the risk-free asset and less of all risky assets but they do not change the relative proportions of their risky assets determined by $\Sigma^{-1}\mu$. This is the mutual fund theorem of Tobin (1958). However, when the investment horizon covers more than one period, the solution also depends on a time-varying factor $\left(A_t R_f^{T-t-1}\right)^{-1}$, which is specific to each investor. The mutual fund theorem of Tobin (1958) is no longer verified in the dynamic setting.

Equation (5) shows that the optimal portfolio cannot be directly computed in practice since it depends on unknown parameters (Σ_t and μ_t) of the excess return distribution. As a result, these two quantities have to be estimated before we obtain an estimation of the optimal portfolio. The standard way to estimate the optimal portfolio consists of estimating Σ_t and μ_t by their sample counterpart at each period after updating information. More precisely, let $r_{t-n+1}, ..., r_t$ be the observations of the excess returns that are considered realizations of the corresponding random vector until the time point t. Then, the mean vector and the covariance matrix at time t are estimated traditionally by $\widehat{\mu}_t = \frac{1}{n}\sum_{i=t-n+1}^{t} r_i$ and $\widehat{\Sigma}_t = \frac{1}{n}\sum_{i=t-n+1}^{t}(r_i - \widehat{\mu}_t)(r_i - \widehat{\mu}_t)'$ respectively. The estimated portfolio at period t is obtained as follows $\widehat{\omega}_t = \widehat{\gamma}_t \widehat{\Sigma}_t^{-1} \widehat{\mu}_t$ where $\widehat{\gamma}_t = 1/\left(\gamma \widehat{A}_t R_f^{T-t-1}\right)$. In fact, γ_t is also an unknown parameter because it depends on A_t which is obtained as a function of $\{\omega_s\}_{s=0}^{t-1}$. Hence, γ_t should be estimated from an estimation of the sequence $\{\omega_s\}_{s=0}^{t-1}$. Using the traditional approach for estimating the optimal weights at time t, we obtain that

$$\widehat{\omega}_t = F\left(\gamma, R_f, \{\widehat{\mu}_s\}_{s=0}^{t}, \left\{\widehat{\Sigma}_s\right\}_{s=0}^{t}\right). \tag{6}$$

However, the choice of the sequence of sample covariance matrices $\left\{\widehat{\Sigma}_s\right\}_{s=0}^{t}$ to form the optimal strategy may not be appropriate. Indeed, the sample covariance matrices may be nearly singular. Inverting them may amplify the estimation errors and affect the performance of the selected strategy. Moreover, the estimation errors in the expected return might be also important, especially when the number of assets in the financial market is large.

2.2 The multi-period problem as a sequence of linear ill-posed problems

At each period t, the optimal portfolio weights are given by the relation (5) or equivalently by the following equation

$$\Sigma_t \omega_t = \gamma_t \mu_t. \tag{7}$$

Equation (7) can be seen as an inverse problem because it can be written as follows

$$\Sigma_t \omega_t = \eta_t, \tag{8}$$

where $\eta_t = \gamma_t \mu_t$. Equation (8) is said to be well-posed if it admits a unique and stable[4] solution ω_t. When one of these conditions, such as the existence of a solution, its uniqueness and its stability, is not satisfied, the problem is said to be ill-posed. If the population covariance matrix is not invertible, the relation defined in Equation (7) does not admit a unique solution. To ensure the uniqueness of the solution of Equation (7) at each period, it has been assumed in the previous section that the true and unknown covariance matrix is not singular. More importantly, when the number of assets in the financial market grows, even if Σ_t is not singular, it is likely to be ill-conditioned[5]. Therefore, the inverse of this matrix becomes unreliable yielding a selected portfolio far from optimal. Moreover, Σ_t and η_t are unknown and need to be estimated before solving the linear inverse problem. Any estimation error in $\hat{\Sigma}_t$ and $\hat{\eta}_t$ amplifies the error in the selected strategy. Hence, the sequence of portfolio weights $\{\omega_t\}$ can be seen as the solution of a sequence of ill-posed linear problems $\{\Sigma_t \omega_t = \eta_t\}$ over the investor's life cycle.

According to Carrasco et al. (2007) an interesting way to solve this problem is to regularize Equation (8) by dampening the explosive effect of the inversion of the singular values of $\hat{\Sigma}_t$. It consists in replacing the sequence $\{1/\lambda_{t,j}\}$ of explosive inverse singular values by a sequence $\{q(\alpha, \lambda_{t,j})/\lambda_{t,j}\}$ where the damping function $q(\alpha, \lambda)$ is chosen such that

1. $q(\alpha, \lambda)/\lambda$ remains bounded when $\lambda \to 0$,
2. for any λ, $\lim_{\alpha \to 0} q(\alpha, \lambda) = 1$

where α is the regularization parameter. The damping function is specific to each regularization.

[4]The solution of this problem is stable in the sense that it is continuous in η_t. In other words ω_t is stable with respect to a small change in η_t.

[5]The ratio of the largest eigenvalue over the smallest is large.

2.3 Empirical case to motivate our procedure

Assume that we have an economy with a professional investment management firm that administers a hedge fund. A hedge fund is an investment fund that pools capital from accredited investors or institutional investors and invests it in a variety of assets, often with complex portfolio construction and risk management techniques. Let us also assume that our investor is willing to invest the capital in one of the following industry portfolios from the US financial market: the 5-industry portfolios, 10-industry portfolios, 17-industry portfolios, and 30-industry portfolios. An industry portfolio provides information about the evolution of the shares of companies that compose a given sector based on a composite index. Hence, each sector included in the portfolio will be considered an asset in the financial market. For instance, the 5-industry portfolios contain information on 5 sectors (see Table 2 for more details) which may be considered as five risky assets and the Kenneth French data library provides information about the returns of those assets.

We estimate Equation (5) for each industry portfolio using the following procedures: the traditional method, which is based on both the sample covariance and the sample mean of asset returns, and the Bayesian method introduced by Bauder et al. (2020). We also consider a benchmark portfolio obtained by calibrating the covariance matrix of asset returns and the expected return using monthly data from July 1980 to January 2019. This benchmark will be considered as the true optimal solution. We will then evaluate the performance of the selected portfolio using the return per unit of risk and the turnover. Tables 3 and 4 present the results of this empirical analysis.

Note that the Bauder et al. (2020)'s procedure gives very nice results for the 5-and 10-industry portfolios with respect to several statistics, particularly the return per unit of risk. However, this method performs poorly for the 17-and 30-industry portfolios. This is due to the fact that their estimator involves the inverse of the sample covariance matrix and hence is not appropriate to handle a large number of assets.

Below we propose two methods to consistently estimate the optimal solution of the dynamic problem in order to improve the performance of the selected strategy.

3 Imposing a constraint on the portfolio weights

In this section, we impose a constraint on the portfolio weights when solving the dynamic problem. This new constraint may help improve the performance of the selected portfolio. In fact, portfolios constructed using sample moments generally involve taking extreme long and short positions, which may overestimate the optimal risk and negatively affect the performance of the selected strategy. However, imposing such a constraint is equivalent to shrinking the covariance matrix (toward the identity matrix) in order to avoid extreme positions in the selected portfolio and reduce, for instance, the risk in estimating the optimal strategy. More precisely, we impose the following constraint

$$\|\omega_t\|_2^2 = \sum_{j=1}^{N} \omega_{j,t}^2 \leq d_t$$

for $t = 0, ..., T-1$ where d_t is a non-random positive parameter. Hence, the new optimization problem we have to solve becomes

$$V(0, A_0) = \max_{\left\{\omega_s : \|\omega_s\|_2^2 \leq d_s\right\}_0^{T-1}} E_0(U(A_T)) \tag{9}$$

In the following subsection, we show that solving this problem is equivalent to solving a simple non-constrained dynamic problem.

3.1 From a constrained portfolio problem to a non-constrained problem

We transform the constrained problem into a non-constrained optimization problem and derive a closed-form solution.

In the recursive form the optimization problem in (9) is given by

$$V(t, A_t) = \max_{\left\{\omega_t : \|\omega_t\|_2^2 \leq d_t\right\}} E_t \left\{V(t+1, A_{t+1})\right\} = \max_{\left\{\omega_t : \|\omega_t\|_2^2 \leq d_t\right\}} E_t \left\{V(t+1, A_t(R_f + \omega_t^{'} r_{t+1}))\right\} \tag{10}$$

We then obtain the following result for this optimization problem.

Proposition 1. *Under the assumptions about the economy stated in Section 2.1, the solution of (10) can be obtained by solving the following unconstrained problem*

$$\max_{\{\omega_t\}} \left\{\exp\left(\lambda_t \|\omega_t\|^2\right) E_t \left[V\left(t+1, A_t\left(R_f + \omega_t^{'} r_{t+1}\right)\right)\right]\right\} \tag{11}$$

for $t = 0, ..., T-1$ with the terminal condition $V(T, A_T) = -\exp(-\gamma A_T)$ and λ_t the Lagrange multiplier associated with the constraint. Moreover, the portfolio weights that solve (11) can be written as follows

$$\omega_t^* = \left(\gamma A_t R_f^{T-t-1}\right)^{-1} \left(\Sigma_t + \alpha_t I_N\right)^{-1} \mu_t \tag{12}$$

where $\alpha_t \in (0, 1)$ is a smoothing parameter used to stabilize the optimal portfolio.

Proof. In Appendix.

The quantity $\left(\exp\left(\lambda_t \|\omega_t\|^2\right)\right)$ in Equation (11) is an additional term to the original portfolio selection problem that materializes the cost the investor has to pay in order to eliminate the constraint we impose in the optimization problem. This non-constrained problem is obtained by penalizing the objective function of the non-constrained portfolio problem in (4) with a penalty term that can be considered the additional cost the investor has to pay in order to reach a stable portfolio. α_t in Proposition 1 is related to the Lagrange multiplier associated with the constraint through the relation $\lambda_t = \frac{\alpha_t}{2}\left(\gamma R_f^{T-t-1}\right)^{T-t} A_t^2$. This implies that to obtain λ_t we need only select α_t. α_t can be seen as a smoothing parameter which helps us solve the problem of ill-posedness when estimating (5). $\frac{\alpha_t}{2}\left(\gamma R_f^{T-t-1}\right)^{T-t} A_t \|\omega_t\|^2$ can be interpreted as the trading cost associated with the optimal selected portfolio. It is, in fact, a quadratic trading cost, as Gârleanu and Pedersen (2013) assumed.

The resolution of the optimization problem is done assuming that α_t is given. However, since the portfolio depends on this parameter, we must select it in an optimal way. The main idea behind (12) is that with an appropriate constraint on the portfolio weights, we solve the problem of ill-posedness that arises when trying to estimate (5). Imposing such a constraint may thus improve the performance of the estimated portfolio. The closed-form solution of proposition 1 is in fact a particular regularized version (the Ridge regularization) of the optimal solution obtained in (5). It consists of adding to the covariance matrix a diagonal matrix in order to solve the problem of ill-posedness induced by the traditional method. Adding such a diagonal matrix may be helpful to stabilize the inverse of the covariance matrix that appears in the optimal solution.

However, the optimal solution obtained in (12) is unknown because it depends on the unknown parameters of the excess return distribution, and needs to be estimated in practice. We can easily estimate this solution by replacing the volatility matrix by the sample covariance and the expected return by the sample mean. More precisely, the estimated portfolio is given by

$$\hat{\omega}_{\alpha_t}^{RdgP} = \hat{\gamma}_t \left(\hat{\Sigma}_t + \alpha_t I_N\right)^{-1} \hat{\mu}_t \tag{13}$$

3.2 Comments on the result of the first procedure

The first thing to note about this method is that the selected strategy in (12) is closely related to a Ridge regularization. The general idea behind this procedure is to control the effect of asset volatility on the investment decision by stabilizing the inverse of the covariance matrix of asset returns. In fact, ridge regularization was first used in regressions in the context where there are too many regressors or when multicollinearity occurs (see Hoerl (1962), Hoerl and Kennard (1970), Mason and Brown (1975)). In this context, the ordinary least squares estimator is unbiased, but its variance is large, so it may be far from the true value. Hence, by adding a small bias to the regression estimates (replacing $X'X$ by $X'X + \alpha I$ where I is the identity matrix), ridge regression reduces the standard errors. More precisely, assume that we want to estimate a parameter θ from the following multiple linear regression model $y = X\theta + \epsilon$ then the standard OLS version of θ is given by $\hat{\theta}_{ols} = \left(X'X\right)^{-1} X'y$ and Ridge regularized version of θ is $\hat{\theta}_{ridge} = \left(X'X + \alpha I\right)^{-1} X'y$. The Ridge regression, as well as our procedure, involves an unknown regularization parameter α_t which needs to converge to zero with the sample size at a certain rate for the solution to converge. Moreover, a fixed α_t would result in a loss of efficiency. Hence, we need to optimally select this parameter based on a certain selection criterion.

More importantly, this procedure only controls for the estimation error in the covariance matrix of the asset returns through Ridge regularization and ignores estimation errors in the expected returns, which may also be important especially when the number of assets in the financial market increases. Nonetheless, a successful investment strategy is also based on investors' ability to well estimate the expected return.

In the next section, we propose an alternative method which imposes the temporal stability in the investment process and helps to control for the estimation error in the expected return.

4 Imposing a temporal stability constraint in the dynamic problem

In this section, instead of imposing a constraint on the optimal portfolio weights in the dynamic problem, we impose a temporal stability constraint. It consists of controlling the distance between two consecutive investment strategies. Hence, imposing such a constraint guarantees that the optimal portfolio composition remains stable over time. This new constraint will be very useful in the sense that it helps investors to avoid high transaction costs in their investment process. Moreover, with this second procedure, we introduce a second level of regularization to the sample expected return which helps to control for the estimation error in the expected return. We propose two different temporal stability constraints in this paper.

4.1 Imposing a L2 temporal stability constraint

We impose the following L2 stability constraint in our dynamic problem

$$\|\omega_t - \omega_{t-1}\|_2^2 = \sum_{i=1}^{N} (\omega_{i,t} - \omega_{i,t-1})^2 \leq d_t$$

for $t = 0, ..., T-1$ with $\omega_{-1} = 0_N$ and d_t a positive and non-random constant.

By imposing such a constraint at each period, the investor's new optimization problem becomes

$$V(0, A_0) = \max_{\left\{\omega_s : \|\omega_s - \omega_{s-1}\|_2^2 \leq d_s\right\}_0^{T-1}} E_0(U(A_T)). \tag{14}$$

In the recursive form we have that

$$V(t, A_t) = \max_{\left\{\omega_t : \|\omega_t - \omega_{t-1}\|_2^2 \leq d_t\right\}} E_t \left\{ V(t+1, A_t(R_f + \omega_t^{'} r_{t+1})) \right\} \tag{15}$$

with the terminal condition $V(T, A_T) = -\exp(-\gamma A_T)$. Solving this dynamic problem we obtain the following first order condition

$$\gamma_t^{-1} (\Sigma_t + \alpha_t I_N) \omega_t = \mu_t + \alpha_t \omega_{t-1}. \tag{16}$$

This equation gives the dynamics of the optimal portfolio over the investor's life cycle as a function of volatility and the expected return.

The following proposition provides an interesting way to estimate the optimal solution through (16).

Proposition 2. *The optimal solution of the optimization problem in (15) can be estimated as follows*

$$\hat{\omega}_{\alpha_t}^{L2TSP} = \hat{\gamma}_t \hat{\Sigma}_{\alpha_t}^{-1} \tilde{\mu}_t$$

for $t = 1, ..., T-1$ with $\hat{\omega}_{\alpha_0} = \hat{\Sigma}_{\alpha_0}^{-1} \hat{\mu}_0$ where $\hat{\Sigma}_{\alpha_t} = \hat{\Sigma}_t + \alpha_t I_N$, and

$$\tilde{\mu}_t = \hat{\mu}_t + \sum_{j=0}^{t-1} \left(\prod_{i=j}^{t-1} \hat{\gamma}_i \alpha_{i+1} \hat{\Sigma}_{\alpha_i}^{-1} \right) \hat{\mu}_j \tag{17}$$

is a shrinkage estimation for μ at the time point t. The sample mean $\hat{\mu}_j$ and the sample *covariance* $\hat{\Sigma}_j$ *are obtained by rolling windows.*

Proof. In Appendix.

This result implies that, instead of applying the usual estimator for the expected return to form the optimal portfolio, we propose to use a shrinkage estimator, which may be more appropriate than the standard one. In fact, according to Merton (1980), the expected stock returns are very hard to estimate and the estimated values differ strongly from the true value when using the sample mean. Therefore, the resulting estimation errors may induce a suboptimal portfolio composition with very poor performance. Hence, using a shrinkage estimation pioneered by Stein (1956) and James and Stein (1961) can be helpful to handle the error in estimating the expected return, and hence improve the performance of the estimated portfolio.

According to Proposition 2, our selected strategy depends on an unknown tuning parameter which need to be selected. We discuss the selection of this tuning parameter in Section 5.4.

4.2 Imposing a L1 temporal stability constraint in the dynamic problem

Although the L2 temporal stability constraint is helpful to better estimate the expected return through a shrinkage estimation, it does not guarantee that assets allocation remain stable over time (even if the tuning parameter is close to 1). In fact, this temporal stability constraint is equivalent to assuming a quadratic trading cost in our model such that investors trade in small quantities in each period. Moreover, as with the Ridge method, the L2 temporal stability procedure does not have a sparsity property, which may be particularly useful to eliminate irrelevant assets in the selected portfolio when $N > n$.

Instead of using a L2 stability constraint in the dynamic problem, we could use the following L1 temporal stability constraint

$$\|\omega_t - \omega_{t-1}\|_1 = \sum_{i=1}^{N} |\omega_{i,t} - \omega_{i,t-1}| \leq d_t$$

for $t = 0, ..., T-1$ with $\omega_{-1} = 0_N$ and d_t a positive and non-random constant.

This new constraint may be appropriate in particular if investors want to hold portfolios with a few active positions. With this L1 penalty, we will have a subset $N_c \subseteq N$ where $\omega_{jt} = \omega_{jt-1}$ $\forall j \in N_c$. The new optimization problem in such a situation becomes

$$V(t, A_t) = \max_{\left\{\omega_t : \|\omega_t - \omega_{t-1}\|_1 \leq d_t\right\}} E_t \left\{ V(t+1, A_t(R_f + \omega_t' r_{t+1})) \right\}. \tag{18}$$

Note that, unlike what we obtained in Subsection 4.1, there is no closed-form solution to this optimization problem. Hence, we need to solve it numerically. However, since we are in a large dimensional setting, it will be very difficult to solve this problem numerically in practice. Hence, in practice, we decide to use an approximation that helps us to relate this optimization problem to a constrained OLS estimation. More precisely, let n denote the rolling window which is the number of observations on assets returns used at each period to estimate the unknown parameters before solving the dynamic problem. At each period t let us denote by r_i for $i = t-n+1, ..., t$ the observations on the vector of excess returns of the n previous periods. R_t is a $n \times N$ matrix with the ith row given by r_i'. Let us also denote $\Omega_t = E\left(R_t' R_t\right)/n$, and $\theta_t = \Omega_t^{-1}\mu_t = E\left(R_t' R_t\right)^{-1} E\left(R_t' 1_n\right)$. With this notation we can easily compute the optimal portfolio in (5 as follows)

$$\omega_t = \gamma_t \frac{\theta_t}{1-\mu_t^{'}\theta_t} \tag{19}$$

This decomposition of the optimal portfolio has been obtained in Carrasco et al. (2019). θ_t can be obtained through the following OLS model $1_n = R_t\theta_t + u_t$ for $\forall$ t. We show in Lemma 2 in Appendix that $E\left(R_t^{'}u_t\right) = 0$.

A good way to estimate the optimal solution of the L1 temporal stability portfolio is by solving the following optimization problem

$$\hat{\theta}_{\alpha_t} = arg\min_{\theta_t} \frac{1}{n}\left\|1_n - R_t\theta_t\right\|_2^2 + \alpha_t\left\|\theta_t - \theta_{t-1}\right\|_1 \tag{20}$$

with $\alpha_t \in (0,1)$ and θ_t from $1_n = R_t\theta_t + u_t$.

Hence, this solution also depends on an unknown tuning parameter which needs to be selected reasonably in order to obtain a solution with good properties. When the α_t chosen is too large, the estimated solution may perform poorly. Moreover, if α_t is too close to zero, the estimated solution may be close to the standard sample-based portfolio, which is known to perform very poorly.

We can also use a generalization of this L1 penalty to include a Lasso penalty in our optimization problem. This penalty is a variant of the fused Lasso proposed by Tibshirani et al. (2005) and it consists of penalizing the L1-norm of both the portfolio weights and their successive changes over time. This procedure encourages sparse and stable portfolios and it may be particularly useful when $N >> n$ to eliminate irrelevant assets in the selected portfolio at each time point. With the fused Lasso method, the dynamic portfolio problem is given by

$$V(t, A_t) = \max_{\left\{\omega_t: \|\omega_t\|_1 \leq d_{1t}, \|\omega_t - \omega_{t-1}\|_1 \leq d_{2t}\right\}} E_t\left\{V(t+1, A_t(R_f + \omega_t^{'}r_{t+1}))\right\}. \tag{21}$$

5 Asymptotic properties of the selected portfolio

In this section, we derive some asymptotic properties of the selected strategy obtained with our procedures. Several asymptotic properties will be examined, such as consistency, efficiency and asymptotic distribution.

We denote by RdgP the selected strategy with an L2 norm on the portfolio weights, L2TSP the selected portfolio when imposing an L2 temporal stability constraint in the dynamic problem and by L1TSP the optimal selected strategy with an L1 temporal stability constraint. We will also need the notation in Subsection 4.2 to easily derive our asymptotic properties.

5.1 Consistency for L2 penalty

To obtain the consistency of the selected portfolio, we need to impose some regularity conditions.

Assumption A

A(i) For some $\tau_t > 0$, we have that

$$N\sum_j \frac{<\theta_t, \phi_{jt}>^2}{\lambda_{jt}^{2\tau_t}} < +\infty$$

where ϕ_{jt} and λ_{jt}^2 denote the eigenvectors and eigenvalues of $\frac{\Omega_t}{N}$.

A(ii) $\frac{\Sigma_t}{N}$ and $\frac{\Omega_t}{N}$ are Hilbert-Schmidt operators

The regularity condition in A(i) seems to be more restrictive than assuming that

$$\sum_j \frac{<\theta_t, \phi_{jt}>^2}{\lambda_{jt}^{2\tau_t}} < +\infty$$

as in Carrasco et al. (2007) and Carrasco (2012). Nonetheless, Carrasco et al. (2019) show that assumption A hold if the returns are generated by a factor model. Assumption A is used to derive the rate of convergence of the mean squared error in the OLS estimator of θ_t. These two assumptions imply in particular that $\|\theta_t\|^2 < +\infty$ such that we have the following relations $\|\theta_t - \theta_{\alpha_t}\|^2 = O_p(\alpha_t^{\min(\tau_t,2)})$. Note that $\theta_{\alpha_t} = \Omega_{\alpha_t}^{-1}\mu_t$ with $\Omega_{\alpha_t} = \Omega_t + \alpha_t I_N$ the Ridge regularized version of the covariance matrix Ω_t, and α_t the tuning parameter used to stabilize the inversion of the covariance matrix at the period t.

Let us denote by $\mathcal{F}_t$ the set of information at the time point t before the investor selects the optimal portfolio for period t. Using assumption A, we obtain the following result about the consistency of the estimated portfolio.

Proposition 3. *Given the set of information $\mathcal{F}_t$ and under assumption A, we have the following result*

$$\left\|\hat{\omega}_{\alpha_t}^{RdgP} - \omega_t\right\| = o_p(1) \tag{22}$$

$\hat{\omega}_{\alpha_t}^{RdgP}$ is the estimated version of the selected portfolio obtained by imposing the L2 norm on the portfolio weights.

Proof. In the appendix.

Proposition 3 implies the consistency of the estimated portfolio at each period with respect to $\|.\|$ under some regularity conditions. Here $\|.\|$ is the norm induced by the inner product in $\mathbb{R}^N$. According to this proposition, by imposing an appropriate constraint on the dynamic portfolio selection problem, we obtain a feasible strategy very close to the optimal portfolio if the estimation window is large enough and under reasonable regularity conditions.

5.2 Efficiency with respect to the Sharpe ratio for L2 penalty

Let us, now look at the asymptotic property of the Sharpe ratio associated with the selected portfolio. The Sharpe ratio measures the excess return (or the risk premium) per unit of deviation for a given trading strategy. It is a way to examine the performance of an investment by adjusting for its risk. The Sharpe ratio of a given portfolio allocation ω_t is expressed as follows:

$$s_t(\omega_t) = \frac{\mu_t^{'}\omega_t}{\left(\omega_t^{'}\Sigma_t\omega_t\right)^{1/2}}$$

The Sharpe ratio of the optimal portfolio at period t as defined in Equation (5) is thus given by

$$s_t(\omega_{opt,t}) = \left(\mu_t^{'}\Sigma_t\mu_t\right)^{1/2}.$$

However, as mentioned in Section 2, investors cannot reach the optimal portfolio in practice since neither μ_t nor Σ_t is known in advance. Because the optimal portfolio is estimated, the actual Sharpe ratio associated with this strategy may be different from the theoretical one. Hence, this paper aims to provide the investor with a feasible strategy whose Sharpe ratio is as close as possible to the theoretical and unknown Sharpe ratio.

The following proposition presents information about the asymptotic property of the Sharpe ratio associated with the selected portfolio.

Proposition 4. *Given the set of information $\mathcal{F}_t$ and under assumption A we have that*

$$s_t\left(\hat{\omega}_{\alpha_t}\right)^2 = s_t\left(\omega_t\right)^2 + O_p\left[\left(\frac{1}{\alpha_t\sqrt{n}} + \|\theta_t - \theta_{\alpha_t}\|\right)\right], \tag{23}$$

for the RdgP and the L2TSP if $\frac{1}{\alpha_t\sqrt{n}} \to 0$ *as* n *goes to infinity where* ω_t *is the optimal portfolio at the time point* t *given by the equation (5) with* $\|\theta_t - \theta_{\alpha_t}\|^2 = O_p(\alpha_t^{\min(\tau_t,2)})$.

Proof. In Appendix.

The regularity condition behind proposition 4 implies that $\alpha_t\sqrt{n} \to \infty$, which means that the estimation window should go to infinity faster than the optimal tuning parameter goes to zero. More importantly, under the regularity condition $\frac{1}{\alpha_t\sqrt{n}} \to 0$ the result of proposition 4 can be rewritten as follows

$$s_t\left(\hat{\omega}_{\alpha_t}\right)^2 = s_t\left(\omega_t\right)^2 + o_p\left(1\right) \tag{24}$$

since, $\|\theta_t - \theta_{\alpha_t}\|^2 = O\left(\alpha_t^{\min(\tau_t,2)}\right)$ by assumption A, and using the fact that α_t goes to zero as n goes to infinity. Hence, proposition 4 shows that the estimated portfolio is asymptotically efficient in terms of the Sharpe ratio for a wide choice of tuning parameters. Consequently, even if the optimal portfolio at the time point t is not practically available (due to the fact that μ_t and Σ_t are unknown) there exists a feasible portfolio (obtained by imposing an appropriate constraint on the dynamic problem) capable of reaching similar levels of performance in terms of the Sharpe ratio for a large estimation window and a wide choice of the regularization parameter. A similar result has been found by Chen and Yuan (2016) in a static mean-variance portfolio selection problem assuming that assets returns follow a K-factor model.

To show the consistency and the efficiency of the selected portfolio with a L1 temporal stability constraint we need an additional assumption (see assumption B in appendix).

5.3 Mean squared error

The aim of this subsection is to see if we can better control the estimation error by imposing a temporal stability constraint in the portfolio selection problem over investors' life cycle. For this purpose, we derive an approximation to the estimation error in the optimal portfolio at each period in order to understand if it could vanish asymptotically under less restrictive regularity conditions.

Here we define the mean squared error of the selected strategy as follows

$$MSE\left(\hat{\omega}_{\alpha_t}\right) = \frac{1}{Nn} E\left[\left\|\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t} - \omega_t\right)\right\|_2^2\right] \tag{25}$$

Under Assumption A we obtain the following result about the mean squared error of the L2 temporal stability portfolio.

Proposition 5.1 *Given the set of information $\mathcal{F}_t$ and under assumptions A we have the following result about the estimation error of the selected portfolio*

$$MSE\left(\hat{\omega}_{\alpha_t}^{L2TSP}\right) \sim \frac{N}{n^2\alpha_t^2} + \frac{N}{n}\alpha_t^{\min(\tau_t,2)} \tag{26}$$

which is minimized for α_t of order $\left(\frac{1}{n}\right)^{\frac{1}{\tau_t+2}}$. Moreover, we have that

$$MSE\left(\hat{\omega}_{\alpha_t}^{L2TSP}\right) \leq MSE\left(\hat{\omega}_{\alpha_t}^{RdgP}\right) \tag{27}$$

Proof. In Appendix.

The first point of this proposition implies that under appropriate regularity conditions, the estimation error of the selected strategy by imposing a L2 temporal stability constraint vanishes asymptotically. The second fact to notice about this proposition is that we better control the estimation error when imposing an L2 temporal stability constraint compared with the Ridge regularization procedure. Intuitively, this result can be explained by the fact that the RdgP method ignores estimation errors in the expected return while the L2TSP introduces a second level of regularization in the sample mean to control for estimation errors in the expected return.

To obtain a good approximation of the MSE of the selected portfolio obtained with a L1 temporal stability constraint, we need additional assumptions. More information about these assumptions (see assumption B in Appendix) are given in Section 8.2. We obtain the following result about the estimation error of the L1 temporal stability strategy under assumptions A and B.

Proposition 5.2 *Given the set of information $\mathcal{F}_t$ and under assumptions A and B we have the following result about the estimation error of the selected portfolio*

$$MSE\left(\hat{\omega}_{\alpha_t}^{L1TSP}\right) \sim N\alpha_t^2\left(s_t/\xi_{\Omega_t}^2\right) + N\alpha_t\left\|\theta_t - \theta_{t-1}\right\| \tag{28}$$

s_t and $\xi_{\Omega_t}^2$ are two time varying constants defined in Appendix.

Proof. In Appendix.

This proposition also implies that under appropriate regularity conditions, the estimation error of the selected strategy by imposing a L1 temporal stability constraint vanishes asymptotically. In other words, under appropriate regularity conditions we have that

$$MSE\left(\hat{\omega}_{\alpha_t}\right) \to 0$$

for L1TSP and L2TSP which implies that we asymptotically control the MSE for strategies obtained by imposing a temporal stability constraint.

We derive in appendix the asymptotic distribution of a certain linear combination of the estimated version of the Ridge regularized portfolio (see proposition 6) useful for constructing confidence intervals.

5.4 Data-driven Method for Selecting the Tuning Parameter

Sections 3 and 4 illustrate that the selected portfolio depends on a certain smoothing parameter $\alpha_t \in (0, 1)$. We have derived some asymptotic properties of the selected portfolio assuming that this tuning parameter is given. However, in practice, the regularization parameter is unknown and needs to be selected in an optimal way. Hence, for each method, we propose a data-driven selection procedure to obtain an approximation of this parameter.

5.4.1 Tuning parameter for the Ridge regularization

In a static mean-variance framework, Carrasco et al. (2019) propose a data-driven method to optimally select this parameter. This method is based on a cross-validation approximation of a loss function of the estimated portfolio.

In the dynamic setting, we base our procedure on a cross-validation approximation of the mean square error (MSE) of the estimated portfolio. The aim is to find an optimal α_t that minimizes the approximation MSE of $\mu^{'}\hat{\omega}_t$. This type of data-driven method for selecting the tuning parameter based on the MSE of a certain linear combination of the estimated parameter has been used by Carrasco (2012) and Carrasco and Tchuente (2015) for an arbitrary linear combination of the estimated parameter. Here, we select α_t, for which the following expected MSE $E\left[\left(\mu^{'}\left(\hat{\omega}_t-\omega_t\right)\right)^{'}\left(\mu^{'}\left(\hat{\omega}_t-\omega_t\right)\right)\right]$ is as small as possible. The idea behind this procedure is to select the value of α_t, which minimizes the distance between the expected return on the optimal portfolio and the return obtained with the regularized portfolio.

The following result gives us a very nice equivalent of the objective function. We can easily apply a cross-validation approximation procedure on this expression of the objective function.

Proposition 7 *Given the set of information $\mathcal{F}_t$ and under assumption A, we obtain the following result*

$$\left(1-\mu^{'}\theta_t\right)^4\gamma_t^{-2}E\left[\left(\mu^{'}\left(\hat{\omega}_t-\omega_t\right)\right)^{'}\left(\mu^{'}\left(\hat{\omega}_t-\omega_t\right)\right)\right]\sim\frac{1}{n}E\left[\left\|1_n^{'}R_t\left(\hat{\theta}_t-\theta_t\right)\right\|^2\right] \tag{29}$$

The proof of this proposition can be found in the Appendix.

It follows from Proposition 7 that minimizing $E\left[\left(\mu^{'}\left(\hat{\omega}_t-\omega_t\right)\right)^{'}\left(\mu^{'}\left(\hat{\omega}_t-\omega_t\right)\right)\right]$ with respect to α_t is also equivalent to minimizing $\frac{1}{n}E\left[\left\|1_n^{'}R_t\left(\hat{\theta}_t-\theta_t\right)\right\|^2\right]$ with respect to α_t. However, this new expression of the objective function is not feasible because it depends on β_t which is unknown. Hence, we investigate several cross-validation approximation techniques for $\frac{1}{n}E\left[\left\|1_n^{'}R_t\left(\hat{\theta}_t-\theta_t\right)\right\|^2\right]$ following Li (1986, 1987). See Section 8.4 for more details about these cross-validation approximation techniques.

The optimality of this data-driven procedure can be obtained following the same techniques as in the study by Carrasco et al. (2019).

5.4.2 Tuning parameter for the temporal stability constraint

A good way to approximate the optimal solution of the temporal stability portfolio consists of solving the following optimization problem

$$\hat{\theta}_{\alpha_t}=arg\min_{\theta_t}\frac{1}{n}\left\|1_n-R_t\theta_t\right\|_2^2+\alpha_tC\left(\theta_t,\theta_{t-1}\right) \tag{30}$$

where

$$C(\theta_t, \theta_{t-1}) = \begin{cases} \|\theta_t - \theta_{t-1}\|_2^2 & \text{for } L2TSP \\ \|\theta_t - \theta_{t-1}\|_1 & \text{for } L1TSP \end{cases}$$

with $\alpha_t \in (0,1)$ and θ_t from $1_n = R_t\theta_t + u_t$. To select the tuning parameter, the first thing is to transform this optimization as follows (see proof of Lemma 4 for more details)

$$\hat{\tilde{\theta}}_{\alpha_t} = arg\min_{\tilde{\theta}_t} \frac{1}{n}\left\|y_t - R_t\tilde{\theta}_t\right\|_2^2 + \alpha_t C\left(\tilde{\theta}_t\right) \tag{31}$$

where

$$C\left(\tilde{\theta}_t\right) = \begin{cases} \left\|\tilde{\theta}_t\right\|_2^2 & \text{for } L2TSP \\ \left\|\tilde{\theta}_t\right\|_1 & \text{for } L1TSP \end{cases}$$

$y_t = 1_n - R_t\theta_{t-1}$, $y_0 = 1_n$ for the first period. The tuning parameter $\alpha_t \in (0,1)$ at period t can then be selected by applying a cross validation procedure to a Ridge-type regression in (31) for the L2TSP and to a Lasso-type regression for the L1TSP. In practice, at each period t, we will use the following estimator $\hat{y}_{\alpha_{t-1}} = 1_n - R_t\hat{\theta}_{\alpha_{t-1}}$ for y_t in the OLS model in (31).

6 Simulations and empirical study

We start this section by a simulation exercise to set up the performance of our procedure and compare our result to the existing methods. In particular, we compare our method to the Bayesian procedures proposed by Bauder et al. (2020). More precisely, in this section, we focus our attention on how our procedure performs in terms of the Sharpe ratio and the default probability. Moreover, we are interested in how our procedure can perform in terms of minimizing the rebalancing cost at a given period. The rebalancing cost at the time point t can be naturally measured by

$$Cost_t = \sum_{j=1}^{N} |\omega_{t,j} - \omega_{t-1,j}|$$

This measure of the trading cost is, in fact, the turnover. The transaction cost can be measured using the turnover in the sense that these costs are positively related to the turnover. Therefore, in the rest of the paper the turnover will be called transaction costs. The average trading cost over the investment horizon is given by

$$TradingCost = \frac{1}{T}\sum_{t=0}^{T-1} Cost_t.$$

This quantity can be interpreted as the average percentage of wealth traded at each period. It can be assimilated to the transaction costs faced by the investor at a given period, who takes some positions in the financial market. By definition trading costs could be seen as all costs incurred by investors in the process of buying or selling an asset in the financial market. In other words trading costs include brokerage fees, cost of analysis, information cost and any expense incurred in the process of deciding upon and placing an order. Delay in execution, which causes prices at which one trades to be different from those at which one planned to trade, may be included as well.

We also analyze the out-of-sample performance of the selected portfolio from each procedure we have proposed.

6.1 Simulations

We implement a simple simulation exercise to set up the performance of our procedure and compare it with the existing procedures. This comparison will be done using several statistics such as the actual Sharpe ratio, the default probability, and the rebalancing cost. Let us consider for this purpose a simple economy with $N \in \{10, 20, 40, 60, 80, 90, 100\}$ risky assets and a risk-free asset. 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. We also consider a finite life ($T = 12$, which corresponds to one year or 12 months) investors who reallocate their portfolio monthly over their life cycle by maximizing an exponential utility function with the CARA parameter $\gamma = 3$. Let n be the rolling window used at each period to estimate, in particular, the covariance matrix of assets returns. So, at each simulation step, we have to generate $n+T$ excess returns and use them to form the dynamic portfolio over the last T periods of the data set. To form the optimal portfolio at the first period (which is $n+1$), we use the first n generated observations to estimate unknown parameters that appear in the optimal portfolio given in (5). For the second period ($n+2$), we also use the last n data from $t = 2, ..., n+1$ to estimate unknown parameters, and so on. Following Chen and Yuan (2016) and Carrasco et al. (2019), we simulate the excess returns at each simulation step from the following three-factor model for $i = 1, ..., N$ and $t = 1, ..., n+T$

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

$f_t = (f_{1t}, f_{2t}, f_{3t})'$ is the vector of common factors, $b_i = (b_{i1}, b_{i2}, b_{i3})'$ is the vector of factor loading associated with the i-th 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 $\Sigma_\epsilon = \text{diag}(\sigma_1^2, ..., \sigma_N^2)$ 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 Treasury-Bill (T-Bill) 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.

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

Let $\text{SR}(\omega_t)$ be the Sharpe ratio associated with the optimal portfolio ω_t, then $\text{SR}(\omega_t)$ is given as follows

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

To evaluate the performance of our procedure in terms of the Sharpe ratio, we focus our attention on the actual Sharpe ratio associated with the selected portfolio. The actual Sharpe ratio at time point t is given by

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

We also analyze the ability of our procedure to predict the default probability at each time point of the investment horizon. This default probability is defined as the probability of the event giving negative wealth. In fact, there is default at time point t if $A_t < 0$. Let $DP(t)$ denote the default probability at time point t. So, if B is the number of draws in our simulation, we have that

$$DP(t) = \frac{1}{B} \sum_{j}^{B} I\left(A_t(j) \leq 0\right)$$

where $A_t(j)$ is the wealth obtained at step j of our procedure. As we saw it in Section 2, we have N risky assets and a risk-free asset with a constant gross return calibrated by $R_f = 1.0036$. Since, $A_0 = 1$ and $R_f > 1$ then $DP(t) \approx 0 \; \forall \; t$. Hence, a procedure is said to perform well in terms of the default probability if the estimated default probability obtained using this procedure is close to zero, which is the theoretical default probability. The estimated default probability is given by

$$\hat{DP}(t) = \frac{1}{B} \sum_{j}^{B} I\left(\hat{A}_t(j) \leq 0\right)$$

with $\hat{A}_t(j)$ being the estimated wealth obtained at time point t and step j of our procedure.

Moreover, in this simulation we assess the performance of our procedure in terms of minimizing the rebalancing cost. The rebalancing cost at a given period t is estimated as follows:

$$\hat{Cost}_t = \frac{1}{B} \sum_{i=1}^{B} \left[\sum_{j=1}^{N} \left|\hat{\omega}_{t,j}(i) - \hat{\omega}_{t-1,j}(i)\right|\right]$$

Our procedures are compared with the Bayesian procedure introduced by Bauder et al.

(2020). We consider the following portfolio selection procedures: the sample-based portfolio (SbP), the naive portfolio (XoNP) which allocates a constant amount 1/N in each asset, the Ridge regularized portfolio (RdgP) obtained by penalizing the portfolio weights, the temporal stability L2 Norm portfolio (L2TSP), the temporal stability L1 Norm portfolio (L1TSP) obtained in Section 4 and the Bayesian portfolio (BP) proposed by Bauder et al. (2020).

In this analysis, we measure the degree of ill-posedness in our optimization problem by the condition number and the relative condition number defined as the ratio of the empirical condition number to the theoretical condition number. Tables 5 and 6 give the results of this analysis as a function of the number of risky assets in the financial market over several periods. Note that the higher the condition number, the more ill-posed our dynamic problem. As we can see from Tables 5 and 6, the condition number increases substantially when N exceeds 20, hence the sample-based strategy may not be appropriate to estimate the optimal solution which involves the inverse of the covariance matrix.

Therefore, we propose a way to improve the performance of the selected strategy in such a situation. We perform 1000 simulations and estimate our statistics over replications.

The result for the average monthly actual Sharpe ratio is given in Table 7. Several facts can be observed from these results. Indeed, the SbP performs poorly in terms of the actual Sharpe ratio when the number of assets in the financial market exceeds 10. For instance, we obtain an average bias in the actual Sharpe ratio of -0.0735, -0.1029, -0.1460, -0.1709, -0.1843, and -0.1965 for respectively 10, 20, 40, 60, 90, and 100 risky assets in the economy. In fact, when the number of assets in the financial market increases considerably compared with the estimation window, the estimation error resulting in the estimation of the optimal solution is amplified for several reasons. In particular, the sample covariance matrix used to form the SbP is close to a singular matrix. Hence, inverting such a matrix may increase the estimation error drastically such that the selected portfolio deviates strongly from the true one. Moreover, DeMiguel et al. (2007) show by simulation that the estimation window needed for the sample-based mean-variance strategy and its extensions to outperform the 1/N benchmark is around 3000 months for a portfolio with 25 assets, and about 6000 months for a portfolio with 50 assets. However, finding such historical data seems to be unrealistic in an empirical analysis. Hence, we propose a new way to improve the performance of the selected portfolio. The results in Table 7 show that by imposing an appropriate constraint on the dynamic problem we significantly improve the performance of the selected strategy compared with the SbP and Bauder et al. (2020)'s portfolio. The performance of those procedures seems to be independent of the size of the financial market (see Table 8, which contains the bias in the actual Sharpe ratio). In fact, with a reasonable choice of the tuning parameter, each of those methods can achieve satisfactory performance even if the number of assets in the economy is large. Moreover, our procedures outperform the 1/N portfolio, which is known to be a standard benchmark in the literature. More importantly, the L2TSP outperforms the RdgP. To explain this result, note that the RdgP is obtained by a simple Ridge regularization on the sample covariance matrix. However, in addition to this Ridge regularization of the sample covariance, the L2TSP introduces a shrinkage estimator for the expected returns. Hence, the fact that the shrinkage estimator is well known in the literature to reduce errors in estimating the expected returns can explain why the L2TSP outperforms the RdgP in terms of the Sharpe ratio. This result implies that a second level of regularization applied to the expected returns may be useful in some cases to improve the performance of the selected strategy. Similar results are obtained with the L1TSP.

We compute the Sharpe ratio as a function of the tuning parameter for the RdgP. The result of this simulation for $N = 60$ is given in Figure 1. The first interesting thing we can notice from this figure is that there is an optimal choice of the tuning parameter for which the actual Sharpe ratio is as close as possible to the theoretical and unknown Sharpe ratio. This implies that in a large financial market setting, this strategy can help investors to significantly improve the performance of the selected portfolio by selecting a reasonable tuning parameter. The second thing to point out from this graph is that the Sharpe ratio decreases faster as the tuning parameter approaches zero. In fact, the sample-based portfolio could be seen as a particular case of the ridge portfolio with $\alpha = 0$. Therefore, as α approaches zero, the ridge portfolio approaches the SbP and may perform poorly as mentioned. Moreover, the Sharpe ratio also decreases when the tuning parameter is large enough. Hence, investors should select a reasonable value of this smoothing parameter in order to obtain a performance that is as close as possible to the performance of the optimal strategy. Therefore, we propose a data-driven procedure based on cross-validation approximation of the mean square error to help investors to select the tuning parameter of the RdgP portfolio. The idea behind this procedure is to select the value of the regularization parameter that minimizes the distance between the return of the optimal portfolio and the return obtained with the RdgP portfolio. We show by simulations that the objective function used for this purpose is a convex function over the set of the regularization parameters. This property of convexity of the objective function ensures that there is a unique optimal choice of the tuning parameter that minimizes this function over the set of regularization parameters. The results of the cross-validation approximation analysis for $N = 60$ are given in Figure 2. More importantly, the cross-validation criterion increases drastically when the tuning parameter approaches zero. This result is plausible in the sense that the ridge procedure converges to the sample-based portfolio as α approaches zero. And since the number of assets in the economy is large, inverting the sample covariance amplifies the estimation error which creates a strong deviation of the selected portfolio from the true one. The ridge procedure is also known to perform poorly in such a situation (see Figure 1). This feature of the RdgP portfolio observed in Figure 2 gives us a new argument about the bad properties of the sample-based portfolio when the market size is large.

We analyze the ability of each strategy to predict the default probability over the investor's life cycle. For this purpose, we compute by simulation the average monthly default probability for each strategy. The result of this analysis can be found in Table 9. A strategy will be said to perform in terms of predicting the default probability if the default probability obtained with this strategy is as close as possible to the theoretical one. Note that the theoretical default probability is equal to zero. According to our simulations, the SbP and the Bayesian strategy give good results in terms of predicting the default probability only when the number of assets in the economy does not exceed 20. However, those procedures perform poorly when N exceeds 20. The Bayesian method does not perform well for large N because the number of hyper-parameters to be estimated with this procedure substantially increases when N is large (for $N \geq 20$). Nonetheless, by imposing an appropriate constraint in the dynamic problem, we obtain very nice results about investors' ability to predict the default probability. Indeed, the default probability obtained with those strategies is very close to the theoretical one. Moreover, this feature seems to be independent to the number of assets in the financial market.

Other interesting statistic is the monthly re-balancing cost. We show by simulations (see Table 10) that our procedures strongly reduce the re-balancing faced by investors over their life cycle compared with the sample-based portfolio and the Bayesian strategy. Using an appropriate constraint in the portfolio selection process, we obtain a more stable portfolio over time so that investors avoid several re-balancing costs. Our procedures may be appropriate for investors who want to take positions in the financial market in the sense that those strategies help them to avoid high trading costs on the selected portfolio with very good performance.

We also compute in Figure 3 the average transaction costs faced by investors as a function of the tuning parameter for the L1TSP and the L2TSP. This graph is obtained using 20 risky assets, an estimation window of 120 and a one-year investment horizon ($T = 12$). The first thing to notice about this result is that trading costs investors faced decrease as the tuning parameter approaches 1 for both the L1TSP and the L2TSP. α can be seen as the importance of the temporal stability constraint in the dynamic portfolio selection problem. It is, in fact, the additional cost the investor is willing to pay to change the composition of the portfolio between two consecutive time periods. Hence, as α increases, investors become less inclined to change their optimal portfolio to avoid large adjustment costs. The optimal investment policy becomes more stable over time as the tuning parameter increases. Moreover, the trading costs obtained using the L1TSP are always less than what we obtain with the L2TSP for each tuning parameter. This is essentially due to the fact that the L1TSP has a sparsity property that obliges investors to hold portfolios with few active positions. This result implies in particular that investors who fundamentally care about minimizing trading costs in the financial market should select strategies based on the L1TSP technique. Moreover, the rebalancing cost increases as the regularization parameter approaches zero for both the L1TSP and the L2TSP. This result is plausible in the sense that the temporal stability portfolio converges to the SbP as the tuning parameter goes to zero. However, the SbP generally involves taking extreme long and short positions, which may considerably increase the rebalancing cost of this strategy.

In Figure 4, we plot the evolution of the average stability rate as a function of the tuning parameter. Not surprisingly, the L2TSP is always non-stable over time for any $\alpha \in (0, 1)$. In fact, this method is equivalent to assuming a quadratic trading cost in the dynamic portfolio problem in such a way that investors trade at each period in small quantities (see Heaton and Lucas (1996), Gârleanu and Pedersen (2013)).

We also estimate the dynamics of the optimal wealth with our procedures and compute the bias in the optimal wealth. The results from this simulation exercise are given in Tables 11, 12 and 13. The bias in the optimal wealth at each period t is defined as follows

$$Bias(A_t) = \frac{1}{B} \sum_{j}^{B} \frac{\hat{A}_t(j) - A_t(j)}{A_t(j)}$$

The absolute value of this bias can be seen as the loss incurred in a dollar invested in the financial market by selecting a given strategy instead of the true one. Once again our procedures perform very well in terms of predicting the optimal wealth over investors' life cycle compared with the Bayesian method as well as the sample based portfolio. For instance, with 10 risky assets in the economy, we observe an average loss of 0.0422, 0.0132, 0.0104 and 0.0106 respectively for BP, RdgP, L2TSP, and L1TSP. In other words, for a billion dollars of investment in the financial market, the investor gains about 29, 31.8 and 31.6 million of dollars by using the RdgP, L2TSP and L1TSP strategies respectively instead of the Bayesian procedure. Similar results are obtained with 20 and 40 risky assets in the financial market.

Table 14 contains some results about the average bias in the actual Sharpe ratio obtained with several estimation windows for ridge regularization. The bias in the actual Sharpe ratio approaches 0 when the estimation window increases. For instance the bias is -0.0295 and -0.0098 for $n = 120$ and $n = 1000$ respectively. This result implies that the actual Sharpe ratio obtained using the ridge procedure approaches the true one as the estimation window increases. In other words, the ridge strategy is asymptotically efficient with respect to the actual Sharpe ratio, as mentioned in Section 5.

Our procedures involve some smoothing parameters selected using a data driven method. For each strategy, this tuning parameter is used to reduce the effect of the sample estimation errors on the selected portfolio performance. Table 15 and Figure 5 provide information about the optimal selected tuning parameter for each method. An interesting thing to point out is that the tuning parameter tends to increase over time for each strategy in order to mitigate the negative effect of previous estimation errors on the performance of the actual optimal selected portfolio. In fact, to obtain an estimation of the optimal portfolio in (5), we also have to estimate γ_t, whose accuracy depends on the previous estimation errors. Hence, an adjustment on the regularization parameter could help investors to reduce the effect of these estimation errors on the properties of the selected portfolio.

We do a comparative analysis between the RdgP and the L2TSP using the evolution of the mean squared error over the investment horizon. This analysis is done with 20 risky assets and an estimation window of 120 over 24 months. The results of this simulation exercise are given in Figure 6. The MSE of the selected portfolio is relatively stable for those two methods, with a slight increase over the investment horizon. Moreover, we observe an important gap between the MSE of the RdgP portfolio and the MSE of the L2TSP over the life cycle. Intuitively, this gap is plausible in the sense that the L2TSP introduces a second level of regularization in the expected return instead of using the sample mean used by the RdgP. Hence, this procedure also controls the estimation error in the expected return. This is why the global estimation error of the selected strategy is better controlled.

6.2 Empirical study

In this subsection, we investigate the performance of our procedures empirically. We apply our method to several sets of portfolios from Kenneth French's website: the monthly 30-industry portfolios and the monthly 100 portfolios formed on size and book-to-market. We allow investors to re-balance their portfolios every year, as did Barberis (2000). This implies that the optimal portfolio is constructed at the end of June every year for a given estimation window n by maximizing the expected utility. The investor holds this optimal portfolio for one year, realizes gains and losses, updates information and then recomputes optimal portfolio weights for the next period using the same estimation window. According to Brodie et al. (2009) this approach can be seen as an investment exercise to evaluate the effectiveness of investors who base their strategy on the last n periods. According to DeMiguel et al. (2009a,b), this procedure leads essentially to the same results as when holding optimal strategies for one month. This procedure is repeated each year, generating a time series of out-of-sample returns. Given a data set of size T^* and an estimation window of size n, we obtain a set of $T^* - n/12$ out-of-sample returns, each generated recursively using the n previous 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 and the rebalancing cost. For this purpose, we use data from July 1980 to June 2018. Therefore, if we choose the estimation window to be 108 and 120 then the first portfolio will be formed in June 1990 and June 1989 respectively and the last one in June 2017.

Table 16 contains some results of the out-of-sample analysis in terms of the Sharpe ratio for two different data sets: the FF30 and the FF100. For each data set, we compute the out-of-sample Sharpe ratio for two different rolling windows. We observe that the sample-based portfolio performs poorly in terms of the out-of-sample Sharpe ratio for both the FF30 and the FF100. The bad out-of-sample properties of this strategy are essentially due to errors in estimating the covariance matrix and the expected return. Moreover, this estimation error is amplified by the fact that one needs to invert the sample covariance matrix, which may be close to a singular matrix. Nonetheless, the estimation error could be limited using a large historical data set to estimate the unknown parameters. In fact, as seen in Tables 17 and 18, the condition number of the sample covariance matrix decreases when the rolling window increases from $n = 60$ to $n = 120$ and from $n = 120$ to $n = 240$ for the FF30 and FF100 respectively. Therefore, by improving the condition number we partly solve the problem of inversion of the sample covariance matrix of asset returns such that the estimation error is reduced significantly. However, to obtain a reasonable performance with this procedure, we need a very large historical data set in order to estimate the unknown parameters, which may be non-realistic in practice. For a portfolio with only 25 risky assets DeMiguel et al. (2007) show that one needs about 3000 months of historical data for the sample portfolio to achieve a similar performance to that of the 1/N benchmark. We cannot obtain such a rolling window in an empirical setting. Hence, to help investors to well allocate their resources, we focus on two ways to select the optimal portfolio over the life cycle. Each of those procedures significantly outperforms the SbP in terms of the Sharpe ratio. Nonetheless, the L2TSP and the L1TSP outperform the ridge portfolio for both the FF30 and the FF100, for each rolling window. As mentioned before, when the rolling window increases, we are able to estimate the unknown parameters more efficiently. Hence, for a given data set our procedures also tend to perform well for large estimation windows. We also compute in Table 16 the out-of-sample Sharpe ratio for Bauder et al. (2020)'s procedure. Our methods outperform this procedure for each data set.

We obtain similar results in terms of the out-of-sample analysis of the trading cost (see Table 19). More importantly, we obtain very nice results with the L1TSP for each data set across estimation windows. Those results imply that this procedure (the L1TSP) helps investors to select more stable portfolios over their life cycle (in order to avoid high trading costs) with very interesting performance compared with most existing procedures. When transaction costs are included in computing the optimal Sharpe ratio, we clearly see the advantage of using our method over the Bayesian portfolio and the sample based portfolio in the investment process (see Table 20). In particular, the L1 temporal stability portfolio gives very interesting results. The finding in Table 20 suggests that one over estimates the out-of-sample performance of the selected portfolio when trading costs are not included in the investment process. A similar result has been obtained by Carrasco and Koné (2019) who argue statistically the importance of accounting for transaction costs in the investment process.

In Figure 7, we plot the dynamic of the estimated wealth obtained with our procedures from 1990 to 2017. This graph is obtained using the 30 industry portfolios with an estimation window of 120. The evolution of this graph between 1990 and 2017 for each procedure reveals the existence of a period (from 2004 to 2009) with lower financial wealth, showing the negative effect of the financial crisis of 2007-2008 on the investment decision. This graph shows that imposing a temporal stability constraint improves the wealth. Similar results for the trading costs faced by the investor are obtained in Figure 8.

7 Conclusion

This paper addresses a dynamic portfolio selection problem in a large financial market by proposing two procedures for selecting the optimal strategy. First, we penalize the norm of the portfolio weights in the dynamic problem and derive a closed-form solution to this new optimization problem. This optimal solution is closely related to a Ridge regularization, which consists of adding to the volatility matrix a diagonal matrix to reduce estimation errors in the covariance matrix. Under appropriate regularity conditions, we show the consistency of the selected strategy and its efficiency in terms of the Sharpe ratio. This method partially controls the estimation errors in the optimal solution because it ignores estimation errors in the expected return which may also be important when the number of assets in the financial market increases considerably. Hence, we propose an alternative method that consists of penalizing the norm of the difference of successive portfolio weights in the dynamic problem to guarantee that the optimal portfolio composition does not fluctuate widely between periods. We show, under appropriate regularity conditions, that we better control estimation errors in the optimal portfolio with this new procedure. In fact, this procedure introduces a second level of regularization to control for the estimation error in the expected return. Moreover, this second method helps investors avoid high trading costs in the financial market by selecting stable strategies over time.

Each strategy involves an unknown tuning parameter that needs to be selected in an optimal way at each time point. Hence, for each strategy we propose a data-driven method for selecting this parameter.

To evaluate the performance of our procedures we implement a simulation exercise based on a three-factor model calibrated on the real data from US financial market. Simulations show that by imposing an appropriate constraint on the dynamic problem we significantly improve the performance of the selected strategy in terms of the Sharpe ratio, the trading cost, the ability to predict the default probability and the dynamic of the optimal wealth. 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. We considerably reduce the transaction cost by imposing a temporal stability constraint on the dynamic portfolio selection problem.

Therefore, our procedures are highly recommended for investors in the dynamic setting in the sense that those procedures help to avoid high trading costs in the financial market by selecting stable strategies that are very effective over time.

References

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

Barberis, Nicholas, "Investing for the long run when returns are predictable," *The Journal of Finance*, 2000, *55* (1), 225–264.

Basak, Suleyman and Georgy Chabakauri, "Dynamic mean-variance asset allocation," *The Review of Financial Studies*, 2010, *23* (8), 2970–3016.

Bauder, David, Taras Bodnar, Nestor Parolya, and Wolfgang Schmid, "Bayesian inference of the multi-period optimal portfolio for an exponential utility," *Journal of Multivariate Analysis*, 2020, p. 104544.

Bodnar, Taras, Arjun K Gupta, and Nestor Parolya, "Direct shrinkage estimation of large dimensional precision matrix," *Journal of Multivariate Analysis*, 2016, *146*, 223–236.

______, Nestor Parolya, and Wolfgang Schmid, "A closed-form solution of the multi-period portfolio choice problem for a quadratic utility function," *Annals of Operations Research*, 2015, *229* (1), 121–158.

______, ______, and ______, "On the exact solution of the multi-period portfolio choice problem for an exponential utility under return predictability," *European Journal of Operational Research*, 2015, *246* (2), 528–542.

______, Ostap Okhrin, and Nestor Parolya, "Optimal shrinkage estimator for high-dimensional mean vector," *Journal of Multivariate Analysis*, 2019, *170*, 63–79.

Brandt, Michael W and Pedro Santa-Clara, "Dynamic portfolio selection by augmenting the asset space," *The Journal of Finance*, 2006, *61* (5), 2187–2217.

______, ______, and Rossen Valkanov, "Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns," *The Review of Financial Studies*, 2009, *22* (9), 3411–3447.

Brodie, Joshua, Ingrid Daubechies, Christine De Mol, Domenico Giannone, and Ignace Loris, "Sparse and stable Markowitz portfolios," *Proceedings of the National Academy of Sciences*, 2009, *106* (30), 12267–12272.

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

Bühlmann, Peter and Sara Van De Geer, *Statistics for high-dimensional data: methods, theory and applications*, Springer Science & Business Media, 2011.

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

——— **and Barbara Rossi**, "In-sample inference and forecasting in misspecified factor models," *Journal of Business & Economic Statistics*, 2016, *34* (3), 313–338.

——— **and Guy Tchuente**, "Regularized LIML for many instruments," *Journal of Econometrics*, 2015, *186* (2), 427–442.

——— **and N'Golo Koné**, "Test for Trading Costs Effect in a Portfolio Selection Problem with Recursive Utility," 2019.

———**, Jean-Pierre Florens, and Eric Renault**, "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization," *Handbook of Econometrics*, 2007, *6*, 5633–5751.

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

———**, N'golo Koné, and Nérée Noumon**, "Optimal portfolio selection using regularization," 2019.

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

Croessmann, Roman, "A Sharpe Ratio Neutral Prior for Bayesian Portfolio Selection," *Available at SSRN 2939776*, 2017.

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

———**, Lorenzo Garlappi, and Raman Uppal**, "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?," *The Review of Financial Studies*, 2007, *22* (5), 1915–1953.

———**,** ———**, and** ———, "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?," *Review of Financial Studies*, 2009, *22* (5), 1915–1953.

———**,** ———**, Francisco J Nogales, and Raman Uppal**, "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," *Management Science*, 2009, *55* (5), 798–812.

Elton, Edwin J and Martin J Gruber, "On the optimality of some multiperiod portfolio selection criteria," *The Journal of Business*, 1974, *47* (2), 231–243.

Gârleanu, Nicolae and Lasse Heje Pedersen, "Dynamic trading with predictable returns and transaction costs," *The Journal of Finance*, 2013, *68* (6), 2309–2340.

Heaton, John and Deborah Lucas, "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, 1996, *104* (3), 443–87.

Hoerl, Arthur E, "Application of Ridge Analysis to Regression Problems,," *Chemical Engineering Progress*, 1962, *58* (3), 54–59.

——— **and Robert W Kennard**, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, 1970, *12* (1), 55–67.

James, William and Charles Stein, "Estimation with quadratic loss," in "Proceedings of the fourth Berkeley symposium on mathematical statistics and probability," Vol. 1 1961, pp. 361–379.

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

Ledoit, Olivier and Michael Wolf, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *Journal of Empirical Finance*, 2003, *10* (5), 603–621.

——— **and** ———, "Honey, I shrunk the sample covariance matrix," *The Journal of Portfolio Management*, 2004, *30*, 110–119.

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

Li, Duan and Wan-Lung Ng, "Optimal dynamic portfolio selection: Multiperiod mean-variance formulation," *Mathematical Finance*, 2000, *10* (3), 387–406.

Li, Ker-Chau, "Asymptotic optimality of CL and generalized cross-validation in ridge regression with application to spline smoothing," *The Annals of Statistics*, 1986, pp. 1101–1112.

———, "Asymptotic optimality for Cp, CL, cross-validation and generalized cross-validation: discrete index set," *The Annals of Statistics*, 1987, pp. 958–975.

Ma, Guiyuan, Chi Chung Siu, and Song-Ping Zhu, "Dynamic portfolio choice with return predictability and transaction costs," *European Journal of Operational Research*, 2019, *278* (3), 976–988.

Markowitz, Harry, "Portfolio Selection: Efficient Diversification of Investments," *Yale University Press*, 1959.

Mason, Robert and William G Brown, "Multicollinearity problems and ridge regression in sociological models," *Social Science Research*, 1975, *4* (2), 135–149.

Merton, Robert C, "Lifetime portfolio selection under uncertainty: The continuous-time case," *The Review of Economics and Statistics*, 1969, pp. 247–257.

———, "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 1980, *8* (4), 323–361.

Penev, Spiridon, Pavel V Shevchenko, and Wei Wu, "The impact of model risk on dynamic portfolio selection under multi-period mean-standard-deviation criterion," *European Journal of Operational Research*, 2019, *273* (2), 772–784.

Pratt, John W, "Risk aversion in the small and in the large," *Econometrica*, 1964, *32* (1/2), 122–136.

Samuelson, Paul A, "Lifetime portfolio selection by dynamic stochastic programming," in "Stochastic Optimization Models in Finance," Elsevier, 1975, pp. 517–524.

Stein, Charles, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," Technical Report, STANFORD UNIVERSITY STANFORD United States 1956.

Tibshirani, Robert, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, *67* (1), 91–108.

Tobin, James, "Liquidity preference as behavior towards risk," *The Review of Economic Studies*, 1958, *25* (2), 65–86.

Touloumis, Anestis, "Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional settings," *Computational Statistics & Data Analysis*, 2015, *83*, 251–261.

Online Appendix for

"A Multi-Period Portfolio Selection in a Large Financial Market"

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

September 22, 2020

8 Comment on the asymptotic properties of the selected portfolio

8.1 Consistency

The result of **Proposition 3** is obtained under the following regularity conditions $\max_{0\leq j\leq t-1}\left\{\frac{N^{1/2}}{\alpha_j\sqrt{n}}+\sqrt{N}\alpha_j^{min(\frac{\tau_j}{2},1)}\right\}\rightarrow 0,\ \sqrt{N}\alpha_t^{\min(\frac{\tau_t}{2},1)}\rightarrow 0\ and\ \frac{1}{\alpha_t\sqrt{n}}\rightarrow 0\ as\ n\rightarrow\infty$

$\sqrt{N}\alpha_t^{min(\frac{\tau_t}{2},1)}\rightarrow 0$ implies that α_t goes to zero faster than $\sqrt{N}$ goes to infinity.

8.2 Mean squared error

Proposition 5.1 is derived under the assumptions that $\max_{0\leq j\leq t-1}\left\{\frac{N^{1/2}}{\alpha_j\sqrt{n}}+\sqrt{N}\alpha_j^{min(\frac{\tau_j}{2},1)}\right\}\rightarrow 0,\ \sqrt{N}\alpha_t^{\min(\frac{\tau_t}{2},1)}\rightarrow 0\ as\ n\rightarrow\infty.$

The result of **proposition 5.2** needs additional assumption. Let us first start with the following useful notations. For each time point t

$$S_t=\{j\in\{1,...,N\}:\theta_{jt}\neq\theta_{jt-1}\}$$

with $s_t=|S_t|$. θ_t can be obtained through the following OLS model $1_n=R_t\theta_t+u_t$ for $\forall\ t$ with $E\left(R_t^{'}u_t\right)=0$ by Lemma 2 in Appendix. S_t will be called the active set at the time point t, which contains elements of θ_t different from their level of the previous period, and $N-s_t$ will be called the time stability index of θ_t. In fact, the main assumption that underlies our L1 procedure is that only a few of θ_t changes compared with their level of $t-1$. Hence, our L1 procedure may help investors to select a more stable portfolio over time in order to avoid high trading costs induced by continuous re-balancing in the optimal portfolio at each period. Moreover, we need the following assumption to obtain a nice result about the mean square error of the selected strategy using the L1 temporal stability constraint. In particular, with this assumption we can easily show the consistency of the L1 strategy. The following notations will also be used in this assumption: θ_{S_t} is a vector with zeros outside the set S_t and coincides with θ_t on S_t. In other words the jth element of this vector is given by:

$$\theta_{S_t,j}=\theta_{jt}1\{j\in S_t\}$$

Moreover, $\theta_{S_t^c}$ is a vector that coincides with θ_t outside S_t. It implies that

$$\theta_{S_t^c,j}=\theta_{jt}1\{j\notin S_t\}$$

and $\theta_t=\theta_{S_t}+\theta_{S_t^c}$

Assumption B.

B(i) At each time point t there is a positive constant α_t^0 with $2\alpha_t^0\leq\alpha_t$ such that we have, $\max_{1\leq j\leq N}\left\{2\left|u_t^{'}R_t^j\right|/n\right\}\leq\alpha_t^0$ where α_t is the smoothing parameter in (20).

B(ii) For some $\xi_{\Omega_t}>0$ and for all θ_t satisfying $\left\|\theta_{S_t^c}\right\|_1\leq\kappa_t\left\|\theta_{S_t}-\theta_{t-1}\right\|_1$ for $\kappa_t>1$, we have that

$\left\|\theta_{S_t}-\theta_{t-1}\right\|_1^2\leq\left(\theta_t^{'}\Omega_t\theta_t\right)s_t/\xi_{\Omega_t}^2$

The assumption B(i) can be found in the study by Bühlmann and Van De Geer (2011). B(ii) can be seen as a modified version of the compatibility condition in Bühlmann and Van De Geer (2011) with $\xi^2_{\Omega_t}$ being the compatibility constant of the matrix Ω_t. This assumption is useful to obtain the consistency of the L1 strategy. According to Bühlmann and Van De Geer (2011) is that if two matrices Σ_0 and Σ_1 are closed to each other, the Σ_0- compatibility condition implies the Σ_1- compatibility condition. This property will be useful when Σ_0 is the population covariance and Σ_1 its sample variance. For more detail about Assumption B see Bühlmann and Van De Geer (2011).

Using assumption B combined with assumption A, we obtain the result of **proposition 5.2**.

8.3 Asymptotic distributions

In this subsection, we derive the asymptotic distribution of a certain linear combination of the estimated version of the Ridge regularized portfolio. With this asymptotic distribution, we could easily construct a confidence interval for that linear combination. In particular, we could construct a confidence set for a given asset in the optimal selected portfolio based on this asymptotic distribution. We need the following assumption to find the asymptotic distribution.

Assumption C *for any given $N \times 1$ vector δ with $\|\delta\| = O(1)$, we have that*

C(i) $\left\|\hat{E}\left(R_t' 1_n\right) - \hat{\Omega}_t\theta_t\right\|^2 = O_p\left(\frac{1}{n}\right)$ with $\hat{\Omega}_t = R_t R_t'/n$.

C(ii) $r_t \sim \mathcal{N}(\mu, \Sigma)$

C(iii) $\delta' r_i u_i$ is independent and identically distributed with $E\left[\delta' r_i u_i\right] = 0$. Moreover $E\left[\delta' r_i r_i' u_i^2 \delta\right] < \infty$

Using this assumption combined with assumption A, we obtain the following result about the asymptotic distribution of $\delta' \hat{\omega}_{\alpha_t}$.

Proposition 6 *Given the set of information $\mathcal{F}_t$ and under assumptions A and C we have the following result*

$$\frac{\langle\sqrt{n}\left[\hat{\omega}_{\alpha_t} - \omega_t\right], \delta\rangle}{\left\|\left(E\left[\delta' r_i r_i' u_i^2 \delta\right]\right)^{1/2}\hat{\Omega}^{-1}_{\alpha_t}\right\|} \to_d \mathcal{N}\left(0, \frac{\gamma_t^2}{\left(1-\mu'\theta_t\right)^2}\right)$$

if $\max_{0\leq j\leq t-1}\left\{\frac{N^{1/2}}{\alpha_j\sqrt{n}} + \sqrt{N}\alpha_j^{min\left(\frac{\tau_j}{2},1\right)}\right\} \to 0,$ $\max\left(\sqrt{N}, \alpha_t\frac{\sqrt{n}}{\sqrt{N}}\right)\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)} \to 0,$ *and* $\frac{N^{1/4}}{\alpha_t n^{1/4}} + (Nn)^{1/4}\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)} \to 0$ *as n goes to infinity.*

Proof. In Section 9.

The result of Proposition 6 implies that under appropriate regularity conditions, the selected portfolio by Ridge regularization is asymptotically normal. This result can then be used in order to construct a confidence interval for $\delta'\omega_t$ or for any component of ω_t. More precisely, a confidence interval for $\delta'\omega_t$ can be obtained as follows:

$$I_{\delta'\omega_t} = \left[\delta'\hat{\omega}_{\alpha_t} - \frac{\hat{\sigma}_t}{\sqrt{n}}z_{\varphi/2}; \delta'\hat{\omega}_{\alpha_t} + \frac{\hat{\sigma}_t}{\sqrt{n}}z_{\varphi/2}\right]$$

where

$$\hat{\sigma}_t = \frac{\left\| \left(\hat{E}\left[\delta' r_i r_i' u_i^2 \delta \right] \right)^{1/2} \hat{\Omega}_{\alpha_t}^{-1} \right\|}{1 - \hat{\mu}_t' \hat{\theta}_{\alpha_t}} \hat{\gamma}_t$$

because $1 - \hat{\mu}_t' \hat{\theta}_{\alpha_t} > 0$. $z_{\varphi/2}$ is the quantile $1 - \varphi/2$ of the standard normal distribution with $\varphi \in (0, 1)$.

8.4 Tuning parameter for the Ridge regularization

Proposition 7 is obtained under the assumption that $\max_{0 \le j \le t-1} \left\{ \frac{N^{1/2}}{\alpha_j \sqrt{n}} + \sqrt{N} \alpha_j^{min(\frac{\tau_j}{2}, 1)} \right\} \to 0 \ as \ n \to \infty$

Following Li (1986, 1987), we investigate the following cross-validation approximation techniques for $\frac{1}{n} E\left[\left\| 1_n' R_t \left(\hat{\theta}_t - \theta_t \right) \right\|^2 \right]$:

(i) The generalized cross-validation (GCV) where:

$$\hat{\alpha}_t = arg \min_{\alpha_t \in H_n} \frac{n^{-1} \left\| \left(I_n - M_{t,n}(\alpha_t) \right) 1_n \right\|^2}{\left(1 - tr\left(M_{t,n}(\alpha_t) \right)/n \right)^2}$$

(ii) Mallow's C_L where:

$$\hat{\alpha}_t = arg \min_{\alpha_t \in H_n} n^{-1} \left\| \left(I_n - M_{t,n}(\alpha_t) \right) 1_n \right\|^2 + 2\sigma_u^2 n^{-1} tr\left(M_{t,n}(\alpha_t) \right)$$

with,

$$M_{t,n}(\alpha_t) v = \sum_{j=1}^{n} q(\alpha_t, \lambda_{jt}^2) \left(\frac{v' \psi_{jt}}{n} \right) \psi_{jt}$$

for any n-dimensional vector v and $tr\left(M_{t,n}(\alpha_t) \right) = \sum_{j=1}^{n} q(\alpha_t, \lambda_{jt}^2)$ and ψ_{jt} the eigenvectors of $R_t R_t'/n$.

9 Proofs

Lemma 2. We have that $E(R_t' u_t) = 0$ in the following ols estimation model

$$1_n = R_t \theta_t + u_t.$$

Proof of Lemma 2

$$\begin{aligned} E\left(R_t' u_t \right) &= E\left[R_t' \left(1_n - R_t \theta_t \right) \right] \\ &= E\left[R_t' 1_n \right] - E\left[R_t' R_t \theta_t \right] \\ &= E\left[R_t' 1_n \right] - E\left[R_t' R_t \right] \theta_t \\ &= E\left[R_t' 1_n \right] - E\left[R_t' R_t \right] E\left[R_t' R_t \right]^{-1} E\left[R_t' 1_n \right] \\ &= E\left[R_t' 1_n \right] - E\left[R_t' 1_n \right] = 0. \end{aligned}$$

Lemma 4. The optimization problem in (30) is equivalent to the optimization problem in

(31) for the L2 norm.

Proof of Lemma 4 The first order condition of (30) is given as follows

$$\begin{aligned} FOC30 &= \frac{2}{n}R_t^{'}\left(1_n - R_t\theta_t\right) + 2\alpha_t\left(\theta_t - \theta_{t-1}\right) = 0 \\ &= \frac{1}{n}R_t^{'}\left(1_n - R_t\theta_t\right) + \alpha_t\left(\theta_t - \theta_{t-1}\right) = 0. \end{aligned}$$

The first order condition of (31) is

$$\begin{aligned} FOC31 &= \frac{2}{n}R_t^{'}\left(y_t - R_t\tilde{\theta}_t\right) + 2\alpha_t\tilde{\theta}_t = 0 \\ &= \frac{1}{n}R_t^{'}\left\{1_n - R_t\theta_t - R_t\left(\theta_t - \theta_{t-1}\right)\right\} + \alpha_t\left(\theta_t - \theta_{t-1}\right) = 0 \\ &= \frac{1}{n}R_t^{'}\left(1_n - R_t\theta_t\right) + \alpha_t\left(\theta_t - \theta_{t-1}\right) = FOC30. \end{aligned}$$

Definition *We denote* $X_n = O_p(Y_n)$ *for positive sequence* $\{X_n\}$ *and* $\{Y_n\}$ *if the sequence* $\left\{\frac{X_n}{Y_n}\right\}$ *is bounded in probability. More precisely, it means that for all* $\epsilon > 0$ *there exists a constant* $B_\epsilon > 0$ *and an integer* N_ϵ *such that* $P\left[\frac{X_n}{Y_n} \leq B_\epsilon\right] \geq 1 - \epsilon \ \forall n \geq N_\epsilon$.

9.1 Proof of Proposition 1

Let's first look at a one period problem. Using the same assumptions as in Section 2, the optimal selection problem will be given as follows

$$\max_{\{\omega:\|\omega\|^2 \leq d\}} E\left(-\exp(-\gamma A_1)\right) = \max_{\{\omega:\|\omega\|^2 \leq d\}} E\left(V_1\right). \tag{34}$$

Since $A_1 = A_0\left(R_f + \omega' r_1\right)$, we have that

$$\begin{aligned} E\left(V_1\right) &= E\left(-\exp(-\gamma A_1)\right) \\ &= -\exp(-A_0\gamma R_f)E\left(\exp(-\gamma A_0\omega' r_1)\right) \\ &= -\exp(-A_0\gamma R_f)\exp\left[-\gamma A_0\left(\omega'\mu - \frac{A_0\gamma}{2}\omega'\Sigma\omega\right)\right] \\ &= -\exp\left[-\gamma A_0\left(R_f + \omega'\mu - \frac{A_0\gamma}{2}\omega'\Sigma\omega\right)\right]. \end{aligned}$$

where the third equality follows from the normality of r_1. Hence, (34) becomes as follows

$$\max_{\{\omega:\|\omega\|^2 \leq d\}} \left\{-\exp\left[-\gamma A_0\left(R_f + \omega'\mu - \frac{A_0\gamma}{2}\omega'\Sigma\omega\right)\right]\right\} \tag{35}$$

which is equivalent of solving the following problem

$$\max_{\{\omega:\|\omega\|^2 \leq d\}} \left\{\gamma A_0\left(R_f + \omega'\mu - \frac{A_0\gamma}{2}\omega'\Sigma\omega\right)\right\} \tag{36}$$

or equivalently,

$$\max_{\{\omega\}} \left\{\gamma A_0\left(R_f + \omega'\mu - \frac{A_0\gamma}{2}\omega'\Sigma\omega\right) - \lambda\|\omega\|^2\right\} \tag{37}$$

because $\gamma A_0 \geq 0$ by assumption, with $\lambda > 0$ the Lagrange multiplier associated with $\|\omega\|^2 \leq d$. Let α be the positive constant solution of $\lambda = \frac{\alpha}{2}\gamma^2 A_0^2$, then (37) becomes as follows

$$\max_{\{\omega\}} \left\{ \gamma A_0 \left(R_f + \omega' \mu - \frac{\alpha}{2}\gamma A_0 \|\omega\|^2 \right) - \frac{(A_0\gamma)^2}{2}\omega'\Sigma\omega \right\}. \tag{38}$$

The solution of this problem can be obtained by solving the following optimization problem

$$\max_{\{\omega\}} E_0 \left[-\exp\left(-\gamma A_0 \left(R_f + \omega' r_1 - \frac{\alpha}{2}\gamma A_0 \|\omega\|^2 \right) \right) \right] = \max_{\{\omega\}} \left\{ \exp\left(\frac{\alpha}{2}\gamma^2 A_0^2 \|\omega\|^2 \right) E_0 [V_1] \right\}$$

where $R_f + \omega' r_1 - \frac{\alpha}{2}\gamma A_0 \|\omega\|^2$ can be seen as the gross return on the optimal portfolio net of the trading cost with $\frac{\alpha}{2}\gamma A_0 \|\omega\|^2$ the transaction cost associated with the selected strategy. When solving this problem, we obtain that

$$\omega = (\gamma A_0)^{-1} (\Sigma + \alpha I_N)^{-1} \mu.$$

Let's now consider a two periods portfolio selection problem. At each period $t = 0, 1$ we solve the following constrained optimization problem starting from the last period with a terminal condition given in Section 2

$$V(t, A_t) = \max_{\left\{\omega_t : \|\omega_t\|^2 \leq d_t\right\}} E_t \left\{ V(t+1, A_{t+1}) \right\} = \max_{\left\{\omega_t : \|\omega_t\|^2 \leq d_t\right\}} E_t \left\{ V(t+1, A_t(R_f + \omega_t' r_{t+1})) \right\}. \tag{39}$$

Hence,

$$V(1, A_1) = \max_{\left\{\omega_1 : \|\omega_1\|^2 \leq d_1\right\}} E_1 \left\{ V(2, A_2) \right\} = \max_{\left\{\omega_1 : \|\omega_1\|^2 \leq d_1\right\}} E_1 \left\{ V\left(2, A_1\left(R_f + \omega_1' r_2\right)\right) \right\}.$$

And it follows from the one period problem that the solution of this optimization problem can be found by solving the following unconstrained problem

$$\begin{aligned}
\max_{\{\omega_1\}} \left\{ \exp\left(\frac{\alpha_1}{2}\gamma^2 A_1^2 \|\omega_1\|^2 \right) E_1 \left[V(2, A_2) \right] \right\} &= \max_{\{\omega_1\}} E_1 \left[-\exp\left\{ -\gamma A_1 \left(R_f + \omega_1' r_2 \right) + \frac{\alpha_1}{2}\gamma^2 A_1^2 \|\omega_1\|^2 \right\} \right] \\
&= \max_{\{\omega_1\}} E_1 \left[-\exp\left\{ -\gamma A_1 \left(R_f + \omega_1' r_2 - \frac{\alpha_1}{2}\gamma A_1 \|\omega_1\|^2 \right) \right\} \right] \\
&= \max_{\{\omega_1\}} \left\{ \exp\left(\frac{\alpha_1}{2}\gamma^2 A_1^2 \|\omega_1\|^2 \right) E_1 \left[-\exp\left\{ -\gamma A_1 \left(R_f + \omega_1' r_2 \right) \right\} \right] \right\} \\
&= \max_{\{\omega_1\}} \left\{ \exp\left(\frac{\alpha_1}{2}\gamma^2 A_1^2 \|\omega_1\|^2 \right) \left[-\exp\left\{ -\gamma A_1 \left(R_f + \omega_1' \mu - \frac{\gamma A_1}{2}\omega' \Sigma\omega \right) \right\} \right] \right\} \\
&= \max_{\{\omega_1\}} \left\{ -\exp\left(\frac{\alpha_1}{2}\gamma^2 A_1^2 \|\omega_1\|^2 - \gamma A_1 \left(R_f + \omega_1' \mu - \frac{\gamma A_1}{2}\omega' \Sigma\omega \right) \right) \right\} \\
&= \max_{\{\omega_1\}} \left\{ -\exp\left(-\gamma A_1 \left[R_f + \omega_1' \mu - \frac{\gamma A_1}{2}\omega_1' \Sigma\omega_1 - \frac{\alpha_1}{2}\gamma A_1 \|\omega_1\|^2 \right] \right) \right\}
\end{aligned}$$

where α_1 is a positive and non random parameter selected in such a way that the Lagrange multiplier λ_1 associated with the constraint $\|\omega_1\|^2 \leq d_1$ is given by $\lambda_1 = \frac{\alpha_1}{2}\gamma^2 A_1^2$.
$R_f + \omega_1' r_2 - \frac{\alpha_1}{2}\gamma A_1 \|\omega_1\|^2$ could be seen as the gross return net of the transaction cost on the optimal selected portfolio at $t = 1$ where $\frac{\alpha_1}{2}\gamma A_1 \|\omega_1\|^2$ is in fact the trading cost associated with the optimal selected strategy of this period. Since $\gamma A_1 \geq 0$, solving this problem is equivalent of solving the following optimization problem

$$\max_{\{\omega_1\}} \left\{ R_f + \omega_1' \mu - \frac{\gamma A_1}{2}\omega_1' \Sigma\omega_1 - \frac{\alpha_1}{2}\gamma A_1 \|\omega_1\|^2 \right\}.$$

The first order condition associated with this optimization is given by

$$\mu - \gamma A_1 \Sigma \omega_1 - \alpha_1 \gamma A_1 \omega_1 = 0.$$

Therefore, the solution of this problem is given by

$$\omega_1^* = (\gamma A_1)^{-1} (\Sigma + \alpha_1 I_N)^{-1} \mu.$$

Now look at the problem at $t = 0$

$$\begin{aligned} V(0, A_0) &= \max_{\{\omega_0 : \|\omega_0\|^2 \le d_0\}} E_0 \{V(1, A_1)\} \\ &= \max_{\{\omega_0 : \|\omega_0\|^2 \le d_0\}} E_0 \left\{ V(1, A_0(R_f + \omega_0' r_1)) \right\} \\ &= \max_{\{\omega_0\}} \left\{ \exp\left(\lambda_0 \|\omega_0\|^2\right) E_0 \left[V\left(1, A_0 \left(R_f + \omega_0' r_1\right)\right)\right] \right\}. \end{aligned}$$

meaning that solving the problem at $t = 0$ is equivalent of solving

$$\max_{\{\omega_0\}} \left\{ \exp\left(\lambda_0 \|\omega_0\|^2\right) E_0 \left[V\left(1, A_0 \left(R_f + \omega_0' r_1\right)\right)\right] \right\}.$$

with λ_0 the Lagrange multiplier associated with the constraint at this period.

Moreover, we have that

$$\begin{aligned} E_1 \left[-\exp\left\{ -\gamma A_1 \left(R_f + \omega_1' r_2 \right) \right\} \right] &= -\exp\{-\gamma A_1 R_f\} E_1 \left[-\exp\left\{ -\gamma A_1 \omega_1' r_2 \right\} \right] \\ &= -\exp\{-\gamma A_1 R_f\} \exp\left\{ -\gamma A_1 \left(\omega_1' \mu - \frac{\gamma A_1}{2} \omega_1' \Sigma \omega_1 \right) \right\} \\ &= -\exp\left\{ -\gamma A_1 \left(R_f + \omega_1' \mu - \frac{\gamma A_1}{2} \omega_1' \Sigma \omega_1 \right) \right\}. \end{aligned}$$

Therefore,

$$\begin{aligned} V^*(1, A_1) &= -\exp\left\{ -\gamma A_1 \left(R_f + (\omega_1^*)' \mu - \frac{\gamma A_1}{2} (\omega_1^*)' \Sigma \omega_1^* \right) \right\} \\ &= -\exp\left\{ -\gamma A_1 R_f - \mu' (\Sigma + \alpha_1 I_N)^{-1} \mu + \frac{1}{2} \mu' (\Sigma + \alpha_1 I_N)^{-1} \Sigma (\Sigma + \alpha_1 I_N)^{-1} \mu \right\} \\ &= -\exp\{-\gamma A_1 R_f + f_1(\mu, \Sigma, \gamma, R_f, \alpha_1)\} \end{aligned}$$

where

$$f_1(\mu, \Sigma, \gamma, R_f, \alpha_1) = -\mu' (\Sigma + \alpha_1 I_N)^{-1} \mu + \frac{1}{2} \mu' (\Sigma + \alpha_1 I_N)^{-1} \Sigma (\Sigma + \alpha_1 I_N)^{-1} \mu$$

We obtain the following problem at $t = 0$ given what is obtained at $t = 1$

$$\begin{aligned} V(0, A_0) &= \max_{\{\omega_0 : \|\omega_0\|^2 \le d_0\}} E_0 \{V^*(1, A_1)\} \\ &= \max_{\{\omega_0 : \|\omega_0\|^2 \le d_0\}} E_0 \{-\exp\{-\gamma A_1 R_f + f_1(\mu, \Sigma, \gamma, R_f, \alpha_1)\}\} \\ &= \max_{\{\omega_0 : \|\omega_0\|^2 \le d_0\}} \{-\exp\{f_1(\mu, \Sigma, \gamma, R_f, \alpha_1)\} E_0 \{\exp\{-\gamma A_1 R_f\}\}\} \\ &= \max_{\{\omega_0\}} \left\{ -\exp\{f_1(\mu, \Sigma, \gamma, R_f, \alpha_1)\} \exp\left(\lambda_0 \|\omega_0\|^2\right) E_0 \{\exp\{-\gamma A_1 R_f\}\} \right\} \\ &= \max_{\{\omega_0\}} \left\{ -\exp\{f_1(\mu, \Sigma, \gamma, R_f, \alpha_1)\} E_0 \left\{ \exp\left\{ -\gamma A_0 R_f \left(R_f + \omega_0' r_1 - \frac{\lambda_0}{\gamma R_f A_0} \|\omega_0\|^2 \right) \right\} \right\} \right\} \end{aligned}$$

where $R_f + \omega_0^{'} r_1 - \frac{\lambda_0}{\gamma R_f A_0} \|\omega_0\|^2$ can be interpreted as the gross return on the optimal portfolio at $t = 0$ net of the transaction cost.

$$\begin{aligned} V(0, A_0) &= \max_{\{\omega_0\}} \left\{ -\exp\left\{ f_1\left(\mu, \Sigma, \gamma, R_f, \alpha_1\right) \right\} E_0 \left\{ \exp\left\{ -\gamma A_0 R_f \left(R_f + \omega_0^{'} r_1 - \frac{\lambda_0}{\gamma R_f A_0} \|\omega_0\|^2 \right) \right\} \right\} \right\} \\ &= \max_{\{\omega_0\}} \left\{ -\exp\left\{ f_1\left(\mu, \Sigma, \gamma, R_f, \alpha_1\right) \right\} \exp\left\{ -\gamma A_0 R_f \left(R_f - \frac{\lambda_0}{\gamma R_f A_0} \|\omega_0\|^2 \right) \right\} * Rest \right\} \\ Rest &= E_0 \left\{ \exp\left\{ -\gamma A_0 R_f \omega_0^{'} r_1 \right\} \right\} \\ Rest &= \exp\left\{ -\gamma A_0 R_f \left(\omega_0^{'} \mu - \frac{\gamma R_f A_0}{2} \omega_0^{'} \Sigma \omega_0 \right) \right\} \end{aligned}$$

$$V(0, A_0) = \max_{\{\omega_0\}} \left\{ -\exp\left\{ f_1\left(\mu, \Sigma, \gamma, R_f, \alpha_1\right) - \gamma A_0 R_f \left(R_f - \frac{\lambda_0}{\gamma R_f A_0} \|\omega_0\|^2 + \omega_0^{'} \mu - \frac{\gamma R_f A_0}{2} \omega_0^{'} \Sigma \omega_0 \right) \right\} \right\}$$

Solving this optimization problem is equivalent to solving

$$\max_{\{\omega_0\}} \left\{ R_f - \frac{\lambda_0}{\gamma R_f A_0} \|\omega_0\|^2 + \omega_0^{'} \mu - \frac{\gamma R_f A_0}{2} \omega_0^{'} \Sigma \omega_0 \right\}$$

The first order conditions are given by

$$2 \frac{\lambda_0}{\gamma R_f A_0} \omega_0 + \gamma R_f A_0 \Sigma \omega_0 = \mu$$

Hence, by choosing $\lambda_0 = \frac{(\gamma R_f)^2 A_0^2 \alpha_0}{2}$, we obtain that

$$\omega_0^* = \left(\gamma A_0 R_f\right)^{-1} \left(\Sigma + \alpha_0 I_N\right)^{-1} \mu.$$

(10) will be solved recursively starting from $t = T - 1$ and using the following terminal condition $V(T, A_T) = -\exp(-\gamma A_T)$. More precisely,

$$\begin{aligned} V(T-1, A_{T-1}) &= \max_{\left\{\omega_{T-1} : \|\omega_{T-1}\|^2 \leq d_{T-1}\right\}} E_{T-1} \left\{ V(T, A_T) \right\} \\ &= \max_{\left\{\omega_{T-1} : \|\omega_{T-1}\|^2 \leq d_{T-1}\right\}} E_{T-1} \left\{ V(T, A_{T-1}(R_f + \omega_{T-1}^{'} r_T)) \right\}. \end{aligned}$$

Since $V(T, A_T) = -\exp(-\gamma A_T)$, we have that

$$\begin{aligned} E_{T-1} \left\{ V(T, A_{T-1}(R_f + \omega_{T-1}^{'} r_T)) \right\} &= E_{T-1} \left\{ -\exp\left[-\gamma A_{T-1} \left(R_f - \omega_{T-1}^{'} r_T \right) \right] \right\} \\ &= -\exp\left(-\gamma A_{T-1} R_f \right) E_{T-1} \left\{ \exp\left[-\gamma A_{T-1} \omega_{T-1}^{'} r_T \right] \right\} \\ &= -\exp\left(-\gamma A_{T-1} R_f \right) \exp\left[-\gamma \left(A_{T-1} \omega_{T-1}^{'} \mu - \frac{\gamma A_{T-1}^2}{2} \omega_{T-1}^{'} \Sigma \omega_{T-1} \right) \right] \\ &= -\exp\left\{ -\gamma \left(A_{T-1} R_f + A_{T-1} \omega_{T-1}^{'} \mu - \frac{\gamma A_{T-1}^2}{2} \omega_{T-1}^{'} \Sigma \omega_{T-1} \right) \right\}. \end{aligned}$$

Hence,

$$V(T-1, A_{T-1}) = \max_{\left\{\omega_{T-1} : \|\omega_{T-1}\|^2 \leq d_{T-1}\right\}} \left\{ -\exp\left\{ -\gamma \left(A_{T-1} R_f + A_{T-1} \omega_{T-1}^{'} \mu - \frac{\gamma A_{T-1}^2}{2} \omega_{T-1}^{'} \Sigma \omega_{T-1} \right) \right\} \right\}. \quad (40)$$

Since $\gamma > 0$, this optimization problem is also equivalent of solving the following problem

$$\max_{\left\{\omega_{T-1}:\|\omega_{T-1}\|^2 \leq d_{T-1}\right\}} \left\{ A_{T-1}R_f + A_{T-1}\omega_{T-1}^{'}\mu - \frac{\gamma A_{T-1}^2}{2}\omega_{T-1}^{'}\Sigma\omega_{T-1} \right\}. \tag{41}$$

If we denote by λ_{T-1} the Lagrange multiplier associated with $\|\omega_{T-1}\|^2 \leq d_{T-1}$, we have that, solving (41) with respect to ω_{T-1} is equivalent of solving the following unconstrained problem by assuming that λ_{T-1} is given

$$\max_{\omega_{T-1}} \left\{ A_{T-1}R_f + A_{T-1}\omega_{T-1}^{'}\mu - \frac{\gamma A_{T-1}^2}{2}\omega_{T-1}^{'}\Sigma\omega_{T-1} - \lambda_{T-1}\left\|\omega_{T-1}\right\|^2 \right\}. \tag{42}$$

Moreover, solving (42) with respect to ω_{T-1} is also equivalent of solving the following unconstrained problem with respect to ω_{T-1}

$$\max_{\omega_{T-1}} \left\{ -\exp\left(-\gamma\left(A_{T-1}R_f + A_{T-1}\omega_{T-1}^{'}\mu - \frac{\gamma A_{T-1}^2}{2}\omega_{T-1}^{'}\Sigma\omega_{T-1} - \frac{\lambda_{T-1}}{\gamma}\left\|\omega_{T-1}\right\|^2\right)\right)\right\}$$

$$= \max_{\omega_{T-1}} \left\{ \exp\left(\lambda_{T-1}\left\|\omega_{T-1}\right\|^2\right)\left\{-\exp\left(-\gamma\left(A_{T-1}R_f + A_{T-1}\omega_{T-1}^{'}\mu - \frac{\gamma A_{T-1}^2}{2}\omega_{T-1}^{'}\Sigma\omega_{T-1}\right)\right)\right\}\right\}$$

$$= \max_{\omega_{T-1}} \left\{\left\{\exp\left(\lambda_{T-1}\left\|\omega_{T-1}\right\|^2\right) E_{T-1}\left\{V(T, A_T)\right\}\right\}\right\}$$

with $V(T, A_T) = -\exp(-\gamma A_T)$.

Therefore, the solution of

$$V\left(T-1, A_{T-1}\right) = \max_{\left\{\omega_{T-1}:\|\omega_{T-1}\|^2 \leq d_{T-1}\right\}} E_{T-1}\left\{V(T, A_T)\right\} \tag{43}$$

can be obtained by solving the following non-constrained problem

$$\max_{\omega_{T-1}} \left\{\left\{\exp\left(\lambda_{T-1}\left\|\omega_{T-1}\right\|^2\right) E_{T-1}\left\{V(T, A_T)\right\}\right\}\right\}$$

First order conditions of the optimization problem in (42) with respect to ω_{T-1} are given by

$$A_{T-1}\mu - \gamma A_{T-1}^2\Sigma\omega_{T-1} - 2\lambda_{T-1}\omega_{T-1} = 0.$$

Hence, we obtain the following closed form to the solution at this time point

$$\omega_{T-1}^{*} = \left(\gamma A_{T-1}\right)^{-1}\left(\Sigma + \alpha_{T-1}I_N\right)^{-1}\mu$$

by choosing $\lambda_{T-1} = \frac{\gamma A_{T-1}^2 \alpha_{T-1}}{2}$ with α_{T-1} a smoothing parameter $\in (0, 1)$.

Let's look at now the problem at $T-2$. At this period, we have to solve the following optimization problem

$$V\left(T-2, A_{T-2}\right) = \max_{\left\{\omega_{T-2}:\|\omega_{T-2}\|^2 \leq d_{T-2}\right\}} E_{T-2}\left\{V^{*}\left(T-1, A_{T-1}\right)\right\}$$

with

$$V^*(T-1,A_{T-1}) = \underbrace{E_{T-1}\{-\exp(-\gamma A_T)\}}_{\omega_{T-1}=\omega^*_{T-1}}$$

$$= -\exp\left\{-\gamma A_{T-1}R_f - \gamma\left(A_{T-1}\left(\omega^*_{T-1}\right)'\mu - \frac{\gamma A^2_{T-1}}{2}\left(\omega^*_{T-1}\right)'\Sigma\omega^*_{T-1}\right)\right\}$$

$$= -\exp\left\{-\gamma A_{T-1}R_f - \gamma\left(\frac{1}{\gamma}\mu'\left(\Sigma+\alpha_{T-1}I_N\right)^{-1}\mu - \frac{1}{\gamma}\mu'\left(\Sigma+\alpha_{T-1}I_N\right)^{-1}\Sigma\left(\Sigma+\alpha_{T-1}I_N\right)^{-1}\mu\right)\right\}$$

$$= -\exp\left\{-\gamma A_{T-1}R_f + f_{T-1}\left(\mu,\Sigma,\gamma,R_f,\alpha_{T-1}\right)\right\}$$

with

$$f_{T-1}\left(\mu,\Sigma,\gamma,R_f,\alpha_{T-1}\right) = -\gamma\left(\frac{1}{\gamma}\mu'\left(\Sigma+\alpha_{T-1}I_N\right)^{-1}\mu - \frac{1}{\gamma}\mu'\left(\Sigma+\alpha_{T-1}I_N\right)^{-1}\Sigma\left(\Sigma+\alpha_{T-1}I_N\right)^{-1}\mu\right)$$

$$V\left(T-2,A_{T-2}\right) = \max_{\left\{\omega_{T-2}:\|\omega_{T-2}\|^2\le d_{T-2}\right\}} E_{T-2}\left\{-\exp\left\{-\gamma A_{T-1}R_f + f_{T-1}\left(\mu,\Sigma,\gamma,R_f,\alpha_{T-1}\right)\right\}\right\}.$$

Solving this problem with respect to ω_{T-2} is also equivalent to solve the following optimization problem (obtained using the same procedure as in the case with $t = T-1$) with respect to ω_{T-2} by also assuming that the Lagrange multiplier is given.

$$\max_{\omega_{T-2}}\left\{\exp\left(\lambda_{T-2}\|\omega_{T-2}\|^2\right)E_{T-2}\left\{-\exp\left\{-\gamma A_{T-1}R_f + f_{T-1}\left(\mu,\Sigma,\gamma,R_f,\alpha_{T-1}\right)\right\}\right\}\right\}$$

$$= \max_{\omega_{T-2}}\left\{\exp\left(\lambda_{T-2}\|\omega_{T-2}\|^2\right)E_{T-2}\left\{V\left(T-1,A_{T-1}\right)\right\}\right\}.$$

Hence, since

$$E_{T-2}\left\{V\left(T-1,A_{T-1}\right)\right\} = -\exp\left\{f_{T-1}\left(\mu,\Sigma,\gamma,R_f,\tilde{\alpha}_{T-1}\right)\right\}E_{T-2}\left\{\exp\left\{-\gamma A_{T-1}R_f\right\}\right\}$$

$$= -\exp\left\{f_{T-1}\left(\mu,\Sigma,\gamma,R_f,\tilde{\alpha}_{T-1}\right)\right\}E_{T-2}\left\{\exp\left\{-\gamma A_{T-2}R_f\left(R_f+\omega'_{T-2}r_{T-1}\right)\right\}\right\}$$

$$= -\exp\left\{f_{T-1}\left(\mu,\Sigma,\gamma,R_f,\tilde{\alpha}_{T-1}\right) - \gamma A_{T-2}R_f^2\right\}E_{T-2}\left\{\exp\left\{-\gamma A_{T-2}R_f\omega'_{T-2}r_{T-1}\right\}\right\}$$

$$= -\exp\left\{f_{T-1}\left(\mu,\Sigma,\gamma,R_f,\alpha_{T-1}\right) - \gamma A_{T-2}R_f^2\right\}\exp\left[-\gamma A_{T-2}R_f\left(\omega'_{T-2}\mu - \frac{\gamma A_{T-2}R_f}{2}\omega'_{T-2}\Sigma\omega_{T-2}\right)\right]$$

$$= -\exp\left\{f_{T-1}\left(\mu,\Sigma,\gamma,R_f,\tilde{\alpha}_{T-1}\right) - \gamma A_{T-2}R_f^2 - \gamma A_{T-2}R_f\left(\omega'_{T-2}\mu - \frac{\gamma A_{T-2}R_f}{2}\omega'_{T-2}\Sigma\omega_{T-2}\right)\right\}$$

$$\max_{\omega_{T-2}}\left\{\exp\left(\lambda_{T-2}\|\omega_{T-2}\|^2\right)E_{T-2}\left\{V\left(T-1,A_{T-1}\right)\right\}\right\}$$

$$= -\exp\left\{f_{T-1} - \gamma A_{T-2}R_f^2 - \gamma A_{T-2}R_f\left(\omega'_{T-2}\mu - \frac{\gamma A_{T-2}R_f}{2}\omega'_{T-2}\Sigma\omega_{T-2}\right) + \lambda_{T-2}\|\omega_{T-2}\|^2\right\}.$$

We then have that the first order conditions of the portfolio selection problem at this time point are given as follows

$$\gamma R_f A_{T-2}\mu - \left(\gamma R_f\right)^2 A^2_{T-1}\Sigma\omega_{T-2} - 2\lambda_{T-2}\omega_{T-2} = 0$$

which implies that

$$\omega^*_{T-2} = \left(\gamma A_{T-2}\right)^{-1}R_f^{-1}\left(\Sigma+\alpha_{T-2}I_N\right)^{-1}\mu$$

with $\lambda_{T-2} = \frac{(\gamma R_f)^2 A_{T-2}^2 \alpha_{T-2}}{2}$.

This procedure holds at each period for $t = 0, ..., T-1$.

9.2 Proof of Proposition 2

Using the same procedure as in the proof of Proposition 1, one can easily show that solving (15) is equivalent to solving the following non-constrained problem

$$\max_{\{\omega_t\}} \left\{ \underbrace{\exp\left(\lambda_t \|\omega_t - \omega_{t-1}\|^2\right)}_{B} E_t\left[V\left(t+1, A_t\left(R_f + \omega_t' r_{t+1}\right)\right)\right] \right\} \tag{44}$$

for $t = 0, ..., T-1$ with the following terminal condition $V(T, A_T) = -\exp(-\gamma A_T)$. And solving (44) at each period from $T-1$ one can easily obtain the following first order condition

$$A_t \gamma \left(\Sigma_t + \alpha_t I_N\right) \omega_t = \mu_t + \alpha_t \omega_{t-1} \tag{45}$$

for $t = 1, ..., T-1$ with

$$A_0 \gamma \left(\Sigma_0 + \alpha_0 I_N\right) \omega_0 = \mu_0. \tag{46}$$

Hence, to obtain a reasonable estimation for the optimal solution, we are going to apply a sequential estimation method. More precisely, at $t = 0$ ω_0 will be estimated as follows

$$\hat{\omega}_0 = \hat{\gamma}_0 \hat{\Sigma}_{\alpha_0}^{-1} \hat{\mu}_0. \tag{47}$$

At the $t = 1$ by combining (45) and (47) we obtain that

$$\hat{\gamma}_1^{-1} \hat{\Sigma}_{\alpha_1} \hat{\omega}_1 = \hat{\mu}_1 + \alpha_1 \hat{\omega}_0$$

which implies that

$$\hat{\omega}_1 = \hat{\gamma}_1 \hat{\Sigma}_{\alpha_1}^{-1} \left[\hat{\mu}_1 + \alpha_1 \hat{\gamma}_0 \hat{\Sigma}_{\alpha_0}^{-1} \hat{\mu}_0\right]. \tag{48}$$

Using the same procedure at $t = 2$ we obtain that

$$\hat{\omega}_2 = \hat{\gamma}_2 \hat{\Sigma}_{\alpha_2}^{-1} \left[\hat{\mu}_2 + \alpha_2 \hat{\gamma}_1 \hat{\Sigma}_{\alpha_1}^{-1} \hat{\mu}_1 + \alpha_1 \alpha_2 \hat{\gamma}_0 \hat{\gamma}_1 \hat{\Sigma}_{\alpha_0}^{-1} \hat{\Sigma}_{\alpha_1}^{-1} \hat{\mu}_0\right]. \tag{49}$$

Therefore, we have that

$$\hat{\omega}_t = \hat{\gamma}_t \hat{\Sigma}_{\alpha_t}^{-1} \tilde{\mu}_t \tag{50}$$

for $t = 1, ..., T-1$ where

$$\hat{\Sigma}_{\alpha_t} = \hat{\Sigma}_t + \alpha_t I_N \tag{51}$$

and

$$\tilde{\mu}_t = \hat{\mu}_t + \sum_{j=0}^{t-1} \left(\prod_{i=j}^{t-1} \hat{\gamma}_i \alpha_{i+1} \hat{\Sigma}_{\alpha_i}^{-1}\right) \hat{\mu}_j \tag{52}$$

9.3 Proof of Proposition 3

To prove this result we need first to show the following preliminary results. Let's recall that θ_t is from the following OLS estimation model $1_n = R_t \theta_t + u_t$. $\hat{\theta}_{\alpha_t}$ is the regularized version of θ_t.

Lemma 1 *Under assumption A the following results hold*

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\| = o_p(1) \tag{53}$$

$$\left\|\mu_t'\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\| = o_p(1) \tag{54}$$

if $\sqrt{N}\alpha_t^{min(\frac{\tau_t}{2},1)} \to 0$ *and* $\frac{1}{\alpha_t\sqrt{n}} \to 0$ *as* $n \to \infty$

Proof of Lemma 1

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\| = \left\|\hat{\theta}_{\alpha_t} - \theta_{\alpha_t} + \theta_{\alpha_t} - \theta_t\right\| \tag{55}$$

where $\theta_{\alpha_t} = \Omega_{\alpha_t}^{-1}\mu_t$. By (55), we have that

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\| \leq \underbrace{\left\|\hat{\theta}_{\alpha_t} - \theta_{\alpha_t}\right\|}_{(A)} + \underbrace{\left\|\theta_{\alpha_t} - \theta_t\right\|}_{(B)}. \tag{56}$$

The first term on the right side of this inequality is the bias corresponding to the estimation of the regularized solution and the second term corresponds to the regularization bias.

$$\begin{aligned}
\hat{\theta}_{\alpha_t} - \theta_{\alpha_t} &= \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\hat{\mu}_t - \left(\Omega_{\alpha_t}\right)^{-1}\mu_t \\
&= \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\hat{\mu}_t - \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\mu_t + \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\mu_t - \left(\Omega_{\alpha_t}\right)^{-1}\mu_t \\
&= \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\left[\hat{\mu}_t - \mu_t\right] + \left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1} - \left(\Omega_{\alpha_t}\right)^{-1}\right]\mu_t
\end{aligned}$$

$\theta_t = \Omega_t^{-1}\mu_t$, this implies that $\mu_t = \Omega_t\theta_t$

$$\hat{\theta}_{\alpha_t} - \theta_{\alpha_t} = \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\left[\hat{\mu}_t - \mu_t\right] + \left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1} - \left(\Omega_{\alpha_t}\right)^{-1}\right]\Omega_t\theta_t$$

$\left\|\hat{\mu}_t - \mu_t\right\|^2 = O_p\left(\frac{N}{n}\right)$ and

$$\left\|\left[\left(\frac{\hat{\Omega}_t}{N}\right)_{\alpha_t}\right]^{-1}\right\|^2 = \sqrt{\lambda_{\max}\left\{\left[\left(\frac{\hat{\Omega}_t}{N}\right)_{\alpha_t}\right]^{-2}\right\}} = \sup_j \frac{\hat{q}_{jt}^2}{\hat{\lambda}_{jt}^4} = O_p\left(\frac{1}{\alpha_t^2}\right).$$

Then,

$$\begin{aligned}
\left\|\left(\hat{\Omega}_{\alpha_t}\right)^{-1}\left[\hat{\mu}_t - \mu_t\right]\right\|^2 &= \left\|\left[\left(\frac{\hat{\Omega}_t}{N}\right)_{\alpha_t}\right]^{-1}\left[\frac{\hat{\mu}_t - \mu_t}{N}\right]\right\|^2 \\
&= O_p\left(\frac{1}{n\alpha_t^2}\right)
\end{aligned}$$

$$\left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1} - \left(\Omega_{\alpha_t}\right)^{-1}\right]\Omega_t\theta_t = \left[\left(\frac{\hat{\Omega}_t}{N}\right)_{\alpha_t}\right]^{-1}\left[\left(\frac{\Omega_t}{N}\right)_{\alpha_t} - \left(\frac{\hat{\Omega}_t}{N}\right)_{\alpha_t}\right]\left(\Omega_{\alpha_t}\right)^{-1}\Omega_t\theta_t.$$

Moreover, by assumption A, we have that $\left\|\Omega_{\alpha_t} - \hat{\Omega}_{\alpha_t}\right\| = O_p\left(\frac{N^2}{n}\right)$ and $\left\|\left(\Omega_{\alpha_t}\right)^{-1}\Omega_t\theta_t\right\| \leq$

$\|\theta_t\| = O\left(\frac{1}{N}\right)$. Hence, we obtain the following relation

$$\left\|\left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1} - \left(\Omega_{\alpha_t}\right)^{-1}\right]\Omega_t\theta_t\right\| = O_p\left(\frac{1}{\alpha_t N\sqrt{n}}\right)$$

$$\left\|\hat{\theta}_{\alpha_t} - \theta_{\alpha_t}\right\| = O_p\left(\frac{1}{\alpha_t\sqrt{n}}\right).$$

Hence, we have that

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\| = O_p\left(\frac{1}{\alpha_t\sqrt{n}} + \|\theta_{\alpha_t} - \theta_t\|\right)$$

where

$$\|\theta_{\alpha_t} - \theta_t\|^2 = O\left(\alpha_t^{min(\tau_t,2)}\right).$$

Therefore if $\frac{1}{\alpha_t\sqrt{n}} \to 0$ as $n \to \infty$ since $\alpha_t \to 0$, we obtain that

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\| \to 0$$

The proof of the second part of this lemma can be obtained using the same procedure as in Lemma 4 of Carrasco et al. (2019). We need also the following result

Lemma 3 *Under assumption A the following results hold*

$$\left\|\hat{\gamma}_t^{-1} - \gamma_t^{-1}\right\| = o_p(1) \tag{57}$$

if $\max_{0\le j\le t-1}\left\{\frac{N^{1/2}}{\alpha_j\sqrt{n}} + \sqrt{N}\alpha_j^{min(\frac{\tau_j}{2},1)}\right\} \to 0$ *as* $n \to \infty$

Proof of Lemma 3

By definition we have that

$$\gamma_t^{-1} = \gamma R_f^{T-t-1} A_t$$

Hence,

$$\hat{\gamma}_t^{-1} - \gamma_t^{-1} = \gamma R_f^{T-t-1}\left(\hat{A}_t - A_t\right)$$

We will show this result by induction. Let's consider the following statement

$$P(t) : \left\|\hat{\gamma}_t^{-1} - \gamma_t^{-1}\right\| = o_p(1)$$

This statement is trivially true for $t = 0$. If fact, at $t = 0$, $\gamma_0^{-1} = \gamma R_f^{T-1} A_0$ which is known. Therefore, $P(0)$ holds.

We will now look at the statement at $t = 1$. $\gamma_1^{-1} = \gamma R_f^{T-1} A_1 = \gamma R_f^{T-1} A_0\left(R_f + \omega_0' r_1\right)$. So,

$$\begin{aligned}
\hat{\gamma}_1^{-1} - \gamma_1^{-1} &= \gamma R_f^{T-2}\left(\hat{A}_1 - A_1\right) \\
&= \gamma R_f^{T-2} A_0 \left(\hat{\omega}_0 - \omega_0\right)' r_1
\end{aligned}$$

The quantity $\gamma R_f^{T-2} A_0$ is known. As in Carrasco et al. (2019) $\gamma_0^{-1}\left(\hat{\omega}_0 - \omega_0\right)$ can be written as follows

$$\begin{aligned}\gamma_0^{-1}(\hat{\omega}_0-\omega_0) &= \frac{\hat{\theta}_0}{1-\hat{\mu}_0'\hat{\theta}_0}-\frac{\theta_0}{1-\mu'\theta_0} \\ &= \frac{\hat{\theta}_0-\theta_0}{\left(1-\hat{\mu}_0'\hat{\theta}_0\right)(1-\mu'\theta_0)}-\frac{\left[\hat{\theta}_0\left(\mu_0'\theta_0\right)-\theta_0\left(\hat{\mu}_0'\hat{\theta}_0\right)\right]}{\left(1-\hat{\mu}_0'\hat{\theta}_0\right)(1-\mu_0'\theta_0)}.\end{aligned}$$

Using the proof of Proposition 1 in Carrasco et al. (2019) combined with the proof of the first part of Lemma 1, we can easily obtain that

$$\begin{aligned}\left\|\gamma_0^{-1}(\hat{\omega}_0-\omega_0)\right\| &= O_p\left(\left\|\hat{\theta}_0-\theta_0\right\|+\frac{1}{\sqrt{n}}\right) \\ &= O_p\left(\frac{1}{\alpha_0\sqrt{n}}+\alpha_0^{min(\frac{\tau_0}{2},1)}+\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Hence,

$$\begin{aligned}\left\|\hat{\gamma}_1^{-1}-\gamma_1^{-1}\right\| &= O_p\left(\frac{N^{1/2}}{\alpha_0\sqrt{n}}+\sqrt{N}\alpha_0^{min(\frac{\tau_0}{2},1)}+\frac{\sqrt{N}}{\sqrt{n}}\right) \\ &= O_p\left(\frac{N^{1/2}}{\alpha_0\sqrt{n}}+\sqrt{N}\alpha_0^{min(\frac{\tau_0}{2},1)}\right)\end{aligned}$$

Therefore, if $\frac{N^{1/2}}{\alpha_0\sqrt{n}}+\sqrt{N}\alpha_0^{min(\frac{\tau_0}{2},1)}\to 0$, $P(1)$ is true.

Let's assume that $P(t)$ is true for $t\geq 1$. This implies that $\left\|\hat{\gamma}_t^{-1}-\gamma_t^{-1}\right\|=o_p(1)$. We need now to show that if $P(t)$ is true, then $P(t+1)$ is also true.

$$\begin{aligned}\hat{\gamma}_{t+1}^{-1}-\gamma_{t+1}^{-1} &= \gamma R_f^{T-t-2}\left(\hat{A}_{t+1}-A_{t+1}\right) \\ &= \gamma R_f^{T-t-2}\left[\hat{A}_t\left(\hat{\omega}_t'r_{t+1}+R_f\right)-A_t\left(\omega_t'r_{t+1}+R_f\right)\right]\end{aligned}$$

By using the fact that the statement $P(t)$ is true, we will have that

$$\hat{\gamma}_{t+1}^{-1}-\gamma_{t+1}^{-1}\approx\gamma R_f^{T-t-2}A_t\left[\hat{\omega}_t-\omega_t\right]'r_{t+1}$$

$$\begin{aligned}\left\|\hat{\gamma}_{t+1}^{-1}-\gamma_{t+1}^{-1}\right\| &\approx \left\|\gamma R_f^{T-t-2}A_t\left[\hat{\omega}_t-\omega_t\right]'r_{t+1}\right\| \\ &\leq \left\|\gamma R_f^{T-t-2}A_t\right\|\left\|\hat{\omega}_t-\omega_t\right\|\left\|r_{t+1}\right\|.\end{aligned}$$

By using the proof of Proposition 1 in Carrasco et al. (2019) combined with the proof of the first part of Lemma 1, we can also obtain that

$$\left\|\hat{\omega}_t-\omega_t\right\|=O_p\left(\frac{1}{\alpha_t\sqrt{n}}+\alpha_t^{min(\frac{\tau_t}{2},1)}+\frac{1}{\sqrt{n}}\right).$$

Hence,

$$\left\|\hat{\gamma}_{t+1}^{-1}-\gamma_{t+1}^{-1}\right\|=O_p\left(\frac{N^{1/2}}{\alpha_t\sqrt{n}}+\sqrt{N}\alpha_t^{min(\frac{\tau_t}{2},1)}\right)$$

Therefore, if $\frac{N^{1/2}}{\alpha_t\sqrt{n}}+\sqrt{N}\alpha_t^{min(\frac{\tau_t}{2},1)}\to 0$, $P(t+1)$ is true.

The rest of the proof of Proposition 3

Using a decomposition similar to that of Carrasco et al. (2019), we obtain that

$$A = \hat{\gamma}_t^{-1}\hat{\omega}_t - \gamma_t^{-1}\omega_t = \frac{\hat{\theta}_t}{1-\hat{\mu}_t'\hat{\theta}_t} - \frac{\theta_t}{1-\mu'\theta_t}$$

$$A = \underbrace{\frac{\hat{\theta}_t - \theta_t}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)}}_{a} - \underbrace{\frac{\left[\hat{\theta}_t\left(\mu'\theta_t\right) - \theta_t\left(\hat{\mu}_t'\hat{\theta}_t\right)\right]}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)}}_{b}. \tag{58}$$

Note that $0 < \mu'\theta_t < 1$ by construction. In fact, since Σ and Ω_t are positive definite matrices, Σ^{-1} and Ω_t^{-1} are also two positive definite matrices. Therefore, $\mu'\Sigma^{-1}\mu > 0$ and $\mu'\Omega_t^{-1}\mu > 0$. Hence, $\mu'\Sigma^{-1}\mu > 0$ implies that $\frac{\mu'\Omega_t^{-1}\mu}{1-\mu'\Omega_t^{-1}\mu} > 0$. Since $\mu'\Omega_t^{-1}\mu > 0$ and $\mu'\Sigma^{-1}\mu > 0$, we have that $1-\mu'\Omega_t^{-1}\mu > 0$ which means that $0 < \mu'\Omega_t^{-1}\mu < 1$ with $\mu'\Omega_t^{-1}\mu = \mu'\theta_t$.

Therefore, we can apply the Taylor expansion on $\frac{1}{1-\hat{\mu}_t'\hat{\theta}_t}$. Hence, we obtain that

$$\frac{1}{1-\hat{\mu}_t'\hat{\theta}_t} = \frac{1}{1-\mu'\theta_t} + \frac{\mu'\left(\hat{\theta}_t-\theta_t\right)}{\left(1-\mu'\theta_t\right)^2} + o\left(\mu'\left(\hat{\theta}_t-\theta_t\right)\right)$$

$$\frac{\hat{\theta}_t-\theta_t}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\hat{\theta}_t-\theta_t}{\left(1-\mu'\theta_t\right)^2} + O_p\left(\left(\hat{\theta}_t-\theta_t\right)\mu'\left(\hat{\theta}_t-\theta_t\right)\right). \tag{59}$$

The second terms in (58) can be developed according to Carrasco et al. (2019) as follows

$$\frac{\hat{\theta}_t\left(\mu'\theta_t\right)-\theta_t\left(\hat{\mu}_t'\hat{\theta}_t\right)}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\left(\hat{\theta}_t-\theta_t\right)\mu'\theta_t - \theta_t\left(\hat{\mu}_t-\mu\right)'\left(\hat{\theta}_t-\theta_t\right) - \theta_t\left(\hat{\mu}_t-\mu\right)'\theta_t - \theta_t\mu'\left(\hat{\theta}_t-\theta_t\right)}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)}. \tag{60}$$

By (59) and because $\left|\mu'\theta_t\right| < 1$, we have that

$$\frac{\left(\hat{\theta}_t-\theta_t\right)\mu'\theta_t}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\left(\hat{\theta}_t-\theta_t\right)\mu'\theta_t}{\left(1-\mu'\theta_t\right)^2} + O_p\left(\left(\hat{\theta}_t-\theta_t\right)\mu'\left(\hat{\theta}_t-\theta_t\right)\right) \tag{61}$$

$$\frac{\theta_t\mu'\left(\hat{\theta}_t-\theta_t\right)}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\theta_t\mu'\left(\hat{\theta}_t-\theta_t\right)}{\left(1-\mu'\theta_t\right)^2} + O_p\left(\left(\hat{\theta}_t-\theta_t\right)\mu'\left(\hat{\theta}_t-\theta_t\right)\right) \tag{62}$$

$$\left|\left(\hat{\mu}_t-\mu\right)'\left(\hat{\theta}_t-\theta_t\right)\right|^2 \le \left\|\hat{\mu}_t-\mu\right\|^2\left\|\hat{\theta}_t-\theta_t\right\|^2$$

$$\left|\left(\hat{\mu}_t-\mu\right)'\theta_t\right|^2 \le \left\|\hat{\mu}_t-\mu\right\|^2\left\|\theta_t\right\|^2$$

$$\frac{\hat{\theta}_t\left(\mu'\theta_t\right)-\theta_t\left(\hat{\mu}_t'\hat{\theta}_t\right)}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\left(\hat{\theta}_t-\theta_t\right)\mu'\theta_t}{\left(1-\mu'\theta_t\right)^2} - \frac{\theta_t\mu'\left(\hat{\theta}_t-\theta_t\right)}{\left(1-\mu'\theta_t\right)^2} + O_p\left(\left(\hat{\theta}_t-\theta_t\right)\mu'\left(\hat{\theta}_t-\theta_t\right)\right) \tag{63}$$

$$+O_p\left[\frac{1}{\sqrt{n}} + \left\|\hat{\theta}_t-\theta_t\right\|\sqrt{\frac{N}{n}}\right] \tag{64}$$

Hence, by assumption A, we obtain that

$$A = \frac{\hat{\theta}_t - \theta_t}{(1 - \mu'\theta_t)} + \frac{\theta_t \mu' \left(\hat{\theta}_t - \theta_t\right)}{(1 - \mu'\theta_t)^2} + O_p\left(\left(\hat{\theta}_t - \theta_t\right)\mu'\left(\hat{\theta}_t - \theta_t\right)\right) + o_p(1) \tag{65}$$

Therefore, using the result of Lemma 3, we obtain that

$$\gamma_t A \approx (\hat{\omega}_t - \omega_t) = \frac{\gamma_t\left(\hat{\theta}_t - \theta_t\right)}{(1 - \mu'\theta_t)} + \frac{\gamma_t \theta_t \mu' \left(\hat{\theta}_t - \theta_t\right)}{(1 - \mu'\theta_t)^2} + o_p(1) \tag{66}$$

$$\|\hat{\omega}_t - \omega_t\| \leq \left\| \frac{\gamma_t\left(\hat{\theta}_t - \theta_t\right)}{(1 - \mu'\theta_t)} \right\| + \left\| \frac{\gamma_t \theta_t \mu' \left(\hat{\theta}_t - \theta_t\right)}{(1 - \mu'\theta_t)^2} \right\| + o_p(1) \tag{67}$$

$$\left\| \frac{\gamma_t\left(\hat{\theta}_t - \theta_t\right)}{(1 - \mu'\theta_t)} \right\| + \left\| \frac{\gamma_t \theta_t \mu' \left(\hat{\theta}_t - \theta_t\right)}{(1 - \mu'\theta_t)^2} \right\| = \frac{\gamma_t}{(1 - \mu'\theta_t)} \left\|\hat{\theta}_t - \theta_t\right\| + \frac{\gamma_t}{(1 - \mu'\theta_t)^2} \left\|\theta_t \mu' \left(\hat{\theta}_t - \theta_t\right)\right\| \tag{68}$$

$$\leq \frac{\gamma_t}{(1 - \mu'\theta_t)} \left\|\hat{\theta}_t - \theta_t\right\| + \frac{\gamma_t}{(1 - \mu'\theta_t)^2} \|\theta_t\| \left\|\mu' \left(\hat{\theta}_t - \theta_t\right)\right\|. \tag{69}$$

Because $\|\theta_t\| < \infty$ by assumption A, $\left\|\hat{\theta}_t - \theta_t\right\| = o_p(1)$ and $\left\|\mu'\left(\hat{\theta}_t - \theta_t\right)\right\| = o_p(1)$ by Lemma 1, we obtain that

$$\|\hat{\omega}_t - \omega_t\| = o_p(1). \tag{70}$$

9.4 Proof of Proposition 4

The actual Sharpe ratio associated with the estimated portfolio is given by

$$s(\hat{\omega}_{\alpha_t}) = \frac{\mu'\hat{\theta}_t}{\left(\hat{\theta}_t'\Sigma\hat{\theta}_t\right)^{1/2}}. \tag{71}$$

(1) What about $\mu'\hat{\theta}_t$?

Let us notice that, we have,

$$\left\|\mu'\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\| \leq \|\mu\| \left\|\hat{\theta}_{\alpha_t} - \theta_t\right\|$$

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\| \leq \left\|\hat{\theta}_{\alpha_t} - \theta_{\alpha_t}\right\| + \|\theta_{\alpha_t} - \theta_t\|.$$

Then,

$$\left\|\mu'\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\| \leq \|\mu\| \left[\left\|\hat{\theta}_{\alpha_t} - \theta_{\alpha_t}\right\| + \|\theta_{\alpha_t} - \theta_t\|\right]$$

$$\begin{aligned}\hat{\theta}_{\alpha_t}-\theta_{\alpha_t} &= \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\hat{\mu}_t-\left(\Omega_{\alpha_t}\right)^{-1}\mu \\ &= \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\hat{\mu}_t-\left(\hat{\Omega}_{\alpha_t}\right)^{-1}\mu+\left(\hat{\Omega}_{\alpha_t}\right)^{-1}\mu-\left(\Omega_{\alpha_t}\right)^{-1}\mu \\ &= \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\left[\hat{\mu}_t-\mu\right]+\left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1}-\left(\Omega_{\alpha_t}\right)^{-1}\right]\mu \\ &= \left(\hat{\Omega}_{\alpha_t}\right)^{-1}\left[\hat{\mu}_t-\mu\right]+\left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1}-\left(\Omega_{\alpha_t}\right)^{-1}\right]\Omega_t\theta_t.\end{aligned}$$

This implies that

$$\left\|\hat{\theta}_{\alpha_t}-\theta_{\alpha_t}\right\|\leq\left\|\left(\hat{\Omega}_{\alpha_t}\right)^{-1}\left[\hat{\mu}_t-\mu\right]\right\|+\left\|\left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1}-\left(\Omega_{\alpha_t}\right)^{-1}\right]\Omega_t\theta_t\right\|.$$

Therefore,

$$\left\|\hat{\theta}_{\alpha_t}-\theta_{\alpha_t}\right\| == O_p\left(\frac{1}{\alpha_t\sqrt{n}}\right).$$

Since $\|\mu\|^2=O(N)$, we have that

$$\left\|\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|=O_p\left(\frac{\sqrt{N}}{\alpha_t\sqrt{n}}+\sqrt{N}\left\|\theta_{\alpha_t}-\theta_t\right\|\right)$$

which implies that

$$\mu'\hat{\theta}_{\alpha_t}=\mu'\theta_t+O_p\left[\sqrt{N}\left(\frac{1}{\alpha_t\sqrt{n}}+\left\|\theta_{\alpha_t}-\theta_t\right\|\right)\right]. \tag{72}$$

(2) What about $\hat{\theta}_t'\Sigma\hat{\theta}_t$?

We know that

$$\hat{\theta}_t=\hat{\theta}_t-\theta_t+\theta_t$$

then,

$$\begin{aligned}\hat{\theta}_t'\Sigma\hat{\theta}_t &= \left(\hat{\theta}_t-\theta_t+\theta_t\right)'\Sigma\left(\hat{\theta}_t-\theta_t+\theta_t\right) \\ &= \left(\hat{\theta}_t-\theta_t\right)'\Sigma\left(\hat{\theta}_t-\theta_t\right)+\left(\hat{\theta}_t-\theta_t\right)'\Sigma\theta_t+\theta_t'\Sigma\left(\hat{\theta}_t-\theta_t\right)+\theta_t'\Sigma\theta_t \\ &= \left(\hat{\theta}_t-\theta_t\right)'\Sigma\left(\hat{\theta}_t-\theta_t\right)+2\left(\hat{\theta}_t-\theta_t\right)'\Sigma\theta_t+\theta_t'\Sigma\theta_t \\ \hat{\theta}_t'\Sigma\hat{\theta}_t-\theta_t'\Sigma\theta_t &= \left(\hat{\theta}_t-\theta_t\right)'\Sigma\left(\hat{\theta}_t-\theta_t\right)+2\left(\hat{\theta}_t-\theta_t\right)'\Sigma\theta_t\end{aligned}$$

$$\left(\hat{\theta}_t-\theta_t\right)'\Sigma\left(\hat{\theta}_t-\theta_t\right)\leq\|\Sigma\|\left\|\hat{\theta}_t-\theta_t\right\|^2.$$

By assumption A we have that $\|\Sigma\|=O(N)$. Moreover, we have that,

$$\left\|\hat{\theta}_{\alpha_t}-\theta_t\right\|=O_p\left(\frac{1}{\alpha_t\sqrt{n}}+\left\|\theta_{\alpha_t}-\theta_t\right\|\right).$$

Hence,

$$\left(\hat{\theta}_t-\theta_t\right)'\Sigma\left(\hat{\theta}_t-\theta_t\right)=O_p\left[N\left(\frac{1}{\alpha_t\sqrt{n}}+\left\|\theta_{\alpha_t}-\theta_t\right\|\right)^2\right]$$

$$\left\|\left(\hat{\theta}_t-\theta_t\right)'\Sigma\theta_t\right\| \leq \|\theta_t\|\,\|\Sigma\|\left\|\hat{\theta}_t-\theta_t\right\|.$$

Hence, by assumption A we obtain that

$$\left\|\left(\hat{\theta}_t-\theta_t\right)'\Sigma\theta_t\right\| = O_p\left[N\left(\frac{1}{\alpha_t\sqrt{n}}+\|\theta_{\alpha_t}-\theta_t\|\right)\right].$$

If $\frac{1}{\alpha_t\sqrt{n}}+\|\theta_{\alpha_t}-\theta_t\| \to 0$ then we have that

$$O_p\left[N\left(\frac{N}{\alpha_t\sqrt{n}}+\|\theta_{\alpha_t}-\theta_t\|\right)^2\right] = O_p\left[N\left(\frac{1}{\alpha_t\sqrt{n}}+\|\theta_{\alpha_t}-\theta_t\|\right)\right].$$

Therefore,

$$\hat{\theta}_t'\Sigma\hat{\theta}_t = \theta_t'\Sigma\theta_t + O_p\left[N\left(\frac{1}{\alpha_t\sqrt{n}}+\|\theta_{\alpha_t}-\theta_t\|\right)\right]. \tag{73}$$

Combining (72) and (73) we obtain that,

$$\begin{aligned} s\left(\hat{\omega}_{\alpha_t}\right)^2 &= \frac{\left(\mu'\hat{\theta}_t\right)^2}{\hat{\theta}_t'\Sigma\hat{\theta}_t} = \frac{\left(\mu'\theta_t\right)^2}{\theta_t'\Sigma\theta_t} + O_p\left[\left(\frac{1}{\alpha_t\sqrt{n}}+\|\theta_{\alpha_t}-\theta_t\|\right)\right] && (74)\\ &= s\left(\omega_t\right)^2 + O_p\left[\left(\frac{1}{\alpha_t\sqrt{n}}+\|\theta_{\alpha_t}-\theta_t\|\right)\right] && (75)\end{aligned}$$

9.5 Proof of Proposition 5.1

$$\hat{\omega}_{\alpha_t}-\omega_t = \left(\hat{\omega}_{\alpha_t}-\omega_{\alpha_t}\right)+\left(\omega_{\alpha_t}-\omega_t\right)$$

where $\omega_{\alpha_t} = \gamma_t\Sigma_{\alpha_t}^{-1}\mu$.

$$\hat{\omega}_{\alpha_t}-\omega_{\alpha_t} = \hat{\gamma}_t\hat{\Sigma}_{\alpha_t}^{-1}\tilde{\mu}_t - \gamma_t\Sigma_{\alpha_t}^{-1}\mu.$$

Using the fact that $\hat{\Sigma}_{\alpha_t}^{-1} = \hat{\Sigma}_{\alpha_t}^{-1}-\Sigma_{\alpha_t}^{-1}+\Sigma_{\alpha_t}^{-1}$, $\tilde{\mu}_t = \tilde{\mu}_t-\mu+\mu$ we obtain that

$$\begin{aligned}\hat{\omega}_{\alpha_t}-\omega_{\alpha_t} &= \hat{\gamma}_t\left(\hat{\Sigma}_{\alpha_t}^{-1}-\Sigma_{\alpha_t}^{-1}\right)\tilde{\mu}_t + \hat{\gamma}_t\Sigma_{\alpha_t}^{-1}\tilde{\mu}_t - \gamma_t\Sigma_{\alpha_t}^{-1}\mu\\ &= \hat{\gamma}_t\left(\hat{\Sigma}_{\alpha_t}^{-1}-\Sigma_{\alpha_t}^{-1}\right)\left(\tilde{\mu}_t-\mu\right) + \hat{\gamma}_t\left(\hat{\Sigma}_{\alpha_t}^{-1}-\Sigma_{\alpha_t}^{-1}\right)\mu + \hat{\gamma}_t\Sigma_{\alpha_t}^{-1}\tilde{\mu}_t - \gamma_t\Sigma_{\alpha_t}^{-1}\mu.\end{aligned}$$

Moreover, using Lemma 3 we obtain that

$$\begin{aligned}\hat{\omega}_{\alpha_t}-\omega_{\alpha_t} &\approx \gamma_t\left(\hat{\Sigma}_{\alpha_t}^{-1}-\Sigma_{\alpha_t}^{-1}\right)\left(\tilde{\mu}_t-\mu\right) + \gamma_t\left(\hat{\Sigma}_{\alpha_t}^{-1}-\Sigma_{\alpha_t}^{-1}\right)\mu + \gamma_t\Sigma_{\alpha_t}^{-1}\tilde{\mu}_t - \gamma_t\Sigma_{\alpha_t}^{-1}\mu\\ &\approx \gamma_t\left(\hat{\Sigma}_{\alpha_t}^{-1}-\Sigma_{\alpha_t}^{-1}\right)\left(\tilde{\mu}_t-\mu\right) + \gamma_t\left(\hat{\Sigma}_{\alpha_t}^{-1}-\Sigma_{\alpha_t}^{-1}\right)\mu + \gamma_t\Sigma_{\alpha_t}^{-1}\left(\tilde{\mu}_t-\mu\right).\end{aligned}$$

Using the following identity $B^{-1}-C^{-1} = B^{-1}\left(C-B\right)C^{-1}$, we have that

$$\hat{\omega}_{\alpha_t}-\omega_{\alpha_t} \approx \gamma_t\hat{\Sigma}_{\alpha_t}^{-1}\left(\Sigma-\hat{\Sigma}_t\right)\Sigma_{\alpha_t}^{-1}\left(\tilde{\mu}_t-\mu\right) + \gamma_t\hat{\Sigma}_{\alpha_t}^{-1}\left(\Sigma-\hat{\Sigma}_t\right)\Sigma_{\alpha_t}^{-1}\mu + \gamma_t\Sigma_{\alpha_t}^{-1}\left(\tilde{\mu}_t-\mu\right)$$

$$\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t}-\omega_{\alpha_t}\right) \approx \gamma_t\hat{\Sigma}_t\hat{\Sigma}_{\alpha_t}^{-1}\left(\Sigma-\hat{\Sigma}_t\right)\Sigma_{\alpha_t}^{-1}\left(\tilde{\mu}_t-\mu\right) + \gamma_t\hat{\Sigma}_t\hat{\Sigma}_{\alpha_t}^{-1}\left(\Sigma-\hat{\Sigma}_t\right)\Sigma_{\alpha_t}^{-1}\mu + \gamma_t\hat{\Sigma}_t\Sigma_{\alpha_t}^{-1}\left(\tilde{\mu}_t-\mu\right)$$

$$\left\|\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t}-\omega_{\alpha_t}\right)\right\|_2 \leq \left\|\gamma_t\hat{\Sigma}_t\hat{\Sigma}_{\alpha_t}^{-1}\left(\Sigma-\hat{\Sigma}_t\right)\Sigma_{\alpha_t}^{-1}\left(\tilde{\mu}_t-\mu\right)\right\|_2 + \left\|\gamma_t\hat{\Sigma}_t\hat{\Sigma}_{\alpha_t}^{-1}\left(\Sigma-\hat{\Sigma}_t\right)\Sigma_{\alpha_t}^{-1}\mu\right\|_2 + \left\|\gamma_t\hat{\Sigma}_t\Sigma_{\alpha_t}^{-1}\left(\tilde{\mu}_t-\mu\right)\right\|_2$$

Since $\left\|\hat{\Sigma}_t\hat{\Sigma}_{\alpha_t}^{-1}\right\|_2 \leq 1$, we have that

$$\left\|\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t}-\omega_{\alpha_t}\right)\right\|_2 \leq \gamma_t\left\|\Sigma-\hat{\Sigma}_t\right\|_2\left\|\Sigma_{\alpha_t}^{-1}\left(\tilde{\mu}_t-\mu\right)\right\|_2 + \gamma_t\left\|\Sigma-\hat{\Sigma}_t\right\|_2\left\|\Sigma_{\alpha_t}^{-1}\mu\right\|_2 + \gamma_t\left\|\tilde{\mu}_t-\mu\right\|_2$$

$\|\mu\| = O\left(\sqrt{N}\right)$, $\left\|\hat{\Sigma}_{\alpha_t}^{-1}\right\| = O_p\left(\frac{1}{\alpha_t}\right)$, $\|\tilde{\mu}_t-\mu\|_2 = O_p\left(\sqrt{\frac{N}{n}}\right)$, $\left\|\Sigma-\hat{\Sigma}_t\right\|_2 = O_p\left(\frac{N}{\sqrt{n}}\right)$ by Assumption A. Hence, we obtain that

$$\begin{aligned}\left\|\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t}-\omega_{\alpha_t}\right)\right\|_2 &= O_p\left(\gamma_t\frac{N}{\sqrt{n}}.\frac{1}{\alpha_t}.\frac{1}{\sqrt{n}}+\gamma_t.\frac{N}{\sqrt{n}}.\frac{1}{\alpha_t}+\gamma_t.\sqrt{\frac{N}{n}}\right)\\ &= O_p\left(\frac{N}{\alpha_t\sqrt{n}}\right).\end{aligned}$$

$$\left\|\hat{\Sigma}_t\left(\omega_t-\omega_{\alpha_t}\right)\right\|_2 \leq \gamma_t\left\|\hat{\Sigma}_t\right\|\left\|\Sigma^{-1}\mu-\Sigma_{\alpha_t}^{-1}\mu\right\|_2$$

since $\left\|\Sigma^{-1}\mu-\Sigma_{\alpha_t}^{-1}\mu\right\|_2 = O\left(\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)}\right)$ hence,

$$\left\|\hat{\Sigma}_t\left(\omega_t-\omega_{\alpha_t}\right)\right\|_2 = O_p\left(N\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)}\right).$$

Let's now recall the prediction error

$$MSE\left(\hat{\omega}_{\alpha_t}\right) = \frac{1}{Nn}E\left[\left\|\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t}-\omega_t\right)\right\|_2^2\right].$$

Using this definition of the prediction error, we obtain that

$$MSE\left(\hat{\omega}_{\alpha_t}\right) \sim \frac{N}{n^2\alpha_t^2}+\frac{N}{n}\alpha_t^{\min(\tau_t,2)}$$

9.6 Proof of Proposition 5.2

Let's first start with a simple example that verifies B(ii).

Example: Let us consider the following case where $N = 2$ with $\theta_{1t} = \theta_{1t-1} \neq 0$ and $\theta_{2t} \neq \theta_{2t-2}$

$$\Omega_t = \begin{pmatrix}\sigma^2 & \rho\\ \rho & \sigma^2\end{pmatrix}$$

$$\left\|\theta_{S_t^c}\right\|_1 = |\theta_{1t}| = |\theta_{1t-1}|$$
$$\|\theta_{S_t}-\theta_{t-1}\|_1 = |\theta_{2t}-\theta_{2t-1}| + |\theta_{1t-1}|$$

Hence, we have that $\left\|\theta_{S_t^c}\right\|_1 \leq \|\theta_{S_t}-\theta_{t-1}\|_1$ for any θ_t which implies that for any positive constant $\kappa_t > 1$ $\left\|\theta_{S_t^c}\right\|_1 \leq \kappa_t \left\|\theta_{S_t} - \theta_{t-1}\right\|_1$.

$$\begin{aligned}
\left\|\theta_{S_t} - \theta_{t-1}\right\|_1^2 &= \left[\left|\theta_{2t} - \theta_{2t-1}\right| + \left|\theta_{1t-1}\right|\right]^2 \\
&= \left(\theta_{2t} - \theta_{2t-1}\right)^2 + \theta_{1t-1}^2 + 2\left|\theta_{1t-1}\left(\theta_{2t} - \theta_{2t-1}\right)\right| \\
&= \theta_{2t}^2 + \theta_{2t-1}^2 + \theta_{1t-1}^2 - 2\theta_{2t}\theta_{2t-1} + 2\left|\theta_{1t-1}\left(\theta_{2t} - \theta_{2t-1}\right)\right| \\
&= \theta_{2t}^2 + \theta_{2t-1}^2 + \theta_{1t-1}^2 - 2\theta_{2t}\theta_{2t-1} + 2\theta_{1t-1}\theta_{2t} - 2\theta_{1t-1}\theta_{2t-1}
\end{aligned}$$

$$\theta_t^{'}\Omega_t\theta_t = \sigma^2\left(\theta_{1t}^2 + \theta_{2t}^2\right) + 2\rho\theta_{1t}\theta_{2t}$$

Let us now select $\xi_{\Omega_t}^2$ to be as follow

$$\xi_{\Omega_t}^2 = \frac{\frac{det(\Omega_t)}{\sigma^2}\theta_{1t-1}^2}{\Xi^2 + \theta_{2t-1}^2 + \theta_{1t-1}^2 + 2\left|\Xi\theta_{2t-1}\right| + 2\left|\Xi\theta_{1t-1}\right| - 2\theta_{1t-1}\theta_{2t-1}} > 0$$

where Ξ is a positive constant selected in such a way that $|\theta_{2t}| \leq \Xi$. In fact, it may be possible to find such a positive constant which verifies $|\theta_{2t}| \leq \Xi$ because the assumption A implies in particular that $\|\theta_t\| < +\infty$.

$$\begin{aligned}
\left(\theta_t^{'}\Omega_t\theta_t\right) s_t/\xi_{\Omega_t}^2 &= \frac{\Xi^2 + \theta_{2t-1}^2 + \theta_{1t-1}^2 + 2\left|\Xi\theta_{2t-1}\right| + 2\left|\Xi\theta_{1t-1}\right| - 2\theta_{1t-1}\theta_{2t-1}}{\frac{det(\Omega_t)}{\sigma^2}\theta_{1t-1}^2}\theta_t^{'}\Omega_t\theta_t \\
&= \frac{\theta_t^{'}\Omega_t\theta_t}{\frac{det(\Omega_t)}{\sigma^2}\theta_{1t-1}^2}\left(\Xi^2 + \theta_{2t-1}^2 + \theta_{1t-1}^2 + 2\left|\Xi\theta_{2t-1}\right| + 2\left|\Xi\theta_{1t-1}\right| - 2\theta_{1t-1}\theta_{2t-1}\right)
\end{aligned}$$

Moreover, since $\frac{det(\Omega_t)}{\sigma^2}\theta_{1t-1}^2 \leq \theta_t^{'}\Omega_t\theta_t$,

$$\frac{\theta_t^{'}\Omega_t\theta_t}{\frac{det(\Omega_t)}{\sigma^2}\theta_{1t-1}^2} \geq 1$$

Hence,

$$\left(\theta_t^{'}\Omega_t\theta_t\right) s_t/\xi_{\Omega_t}^2 \geq \Xi^2 + \theta_{2t-1}^2 + \theta_{1t-1}^2 + 2\left|\Xi\theta_{2t-1}\right| + 2\left|\Xi\theta_{1t-1}\right| - 2\theta_{1t-1}\theta_{2t-1}$$

Therefore,

$$\left(\theta_t^{'}\Omega_t\theta_t\right) s_t/\xi_{\Omega_t}^2 \geq \left\|\theta_{S_t} - \theta_{t-1}\right\|_1^2$$

because, $\left\|\theta_{S_t} - \theta_{t-1}\right\|_1^2 \leq \Xi^2 + \theta_{2t-1}^2 + \theta_{1t-1}^2 + 2\left|\Xi\theta_{2t-1}\right| + 2\left|\Xi\theta_{1t-1}\right| - 2\theta_{1t-1}\theta_{2t-1}$.

Now we are going to look at the MSE of the selected portfolio by imposing a L1 temporal stability constraint.

$$\begin{aligned}
\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t} - \omega_t\right) &= \left(\frac{R_t^{'}R_t}{n} - \frac{R_t^{'}1_n}{n}\left(\frac{R_t^{'}1_n}{n}\right)^{'}\right)\left(\hat{\omega}_{\alpha_t} - \omega_t\right) \\
&= R_t^{'}\left(\frac{I_n}{n} - \frac{1_n1_n^{'}}{n^2}\right)R_t\left(\hat{\omega}_{\alpha_t} - \omega_t\right)
\end{aligned}$$

$$
\begin{aligned}
\left\|\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t}-\omega_t\right)\right\| &= \left\|R_t^{'}\left(\frac{I_n}{n}-\frac{1_n1_n^{'}}{n^2}\right)R_t\left(\hat{\omega}_{\alpha_t}-\omega_t\right)\right\| \\
&\leq \left\|R_t^{'}\left(\frac{I_n}{n}-\frac{1_n1_n^{'}}{n^2}\right)\right\|\left\|R_t\left(\hat{\omega}_{\alpha_t}-\omega_t\right)\right\|
\end{aligned}
$$

$$
\begin{aligned}
\left\|R_t^{'}\left(\frac{I_n}{n}-\frac{1_n1_n^{'}}{n^2}\right)\right\| &\leq \left\|R_t^{'}\right\|\left\|\frac{I_n}{n}-\frac{1_n1_n^{'}}{n^2}\right\| \\
&\leq \left\|R_t^{'}\right\|\left(\left\|\frac{I_n}{n}\right\|+\left\|\frac{1_n1_n^{'}}{n^2}\right\|\right) \\
&\leq \frac{2}{n}\left\|R_t^{'}\right\| = \frac{2}{n}O_p\left(nN\right) = O_p\left(N\right)
\end{aligned}
$$

The last quantity is obtained using the same matrix norm definition as in Carrasco and Rossi (2016). Moreover, under appropriate regularity conditions we have that

$$
R_t\left(\hat{\omega}_{\alpha_t}-\omega_t\right) = R_t\Psi_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)+o_p(1)
$$

where

$$
\Psi_t = \gamma_t\left[\frac{I_N}{1-\mu_t^{'}\theta_t}+\frac{\theta_t\mu_t^{'}}{\left(1-\mu_t^{'}\theta_t\right)^2}\right]
$$

Hence,

$$
\begin{aligned}
\left\|R_t\left(\hat{\omega}_{\alpha_t}-\omega_t\right)\right\| &\sim \left\|R_t\Psi_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\| \\
&\leq \left\|\Psi_t\right\|\left\|R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\left\|\Psi_t\right\| &\leq \left\|\frac{\gamma_t I_N}{1-\mu_t^{'}\theta_t}\right\|+\left\|\frac{\gamma_t\theta_t\mu_t^{'}}{\left(1-\mu_t^{'}\theta_t\right)^2}\right\| \\
&\leq \frac{\gamma_t}{1-\mu_t^{'}\theta_t}\left\|I_N\right\|+\frac{\gamma_t}{\left(1-\mu_t^{'}\theta_t\right)^2}\left\|\theta_t\mu_t^{'}\right\| \\
&\leq \frac{\gamma_t}{1-\mu_t^{'}\theta_t}+\frac{\gamma_t}{\left(1-\mu_t^{'}\theta_t\right)^2} = \pi_t = O(1)
\end{aligned}
$$

$$
\begin{aligned}
\left\|R_t\left(\hat{\omega}_{\alpha_t}-\omega_t\right)\right\| &\leq \pi_t\left\|R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\| \\
\frac{1}{nN}E\left[\left\|R_t\left(\hat{\omega}_{\alpha_t}-\omega_t\right)\right\|^2\right] &\leq \frac{\pi_t^2}{nN}E\left[\left\|R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|^2\right]
\end{aligned}
$$

Let us now look at $\frac{1}{nN}E\left[\left\|R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|^2\right]$.

We want first to show the following inequality.

$$
\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|^2}{n}+\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1 \leq \frac{2u_t^{'}R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)}{n}+\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1 \tag{76}
$$

We have that

$$\frac{\left\|R_t\hat{\theta}_{\alpha_t}-R_t\theta_t\right\|^2}{n}+\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1\le\frac{2}{n}u_t^{'}\left(R_t\hat{\theta}_{\alpha_t}-R_t\theta_t\right)+\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1\Leftrightarrow$$

$$\frac{\left\|R_t\hat{\theta}_{\alpha_t}-R_t\theta_t-u_t+u_t\right\|^2}{n}+\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1\le\frac{2}{n}u_t^{'}\left(R_t\hat{\theta}_{\alpha_t}-R_t\theta_t-u_t+u_t\right)+\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1\Leftrightarrow$$

$$\frac{\left\|1_n-R_t\hat{\theta}_{\alpha_t}-u_t\right\|^2}{n}+\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1\le\frac{2}{n}u_t^{'}\left(R_t\hat{\theta}_{\alpha_t}-1_n+u_t\right)+\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1\Leftrightarrow$$

$$\begin{aligned}\frac{\left\|1_n-R_t\hat{\theta}_{\alpha_t}\right\|^2}{n}+\frac{\|u_t\|^2}{n}-\frac{2}{n}u_t^{'}\left(1_n-R_t\hat{\theta}_{\alpha_t}\right)+\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1 &\le -\frac{2}{n}u_t^{'}\left(1_n-R_t\hat{\theta}_{\alpha_t}\right)+\frac{2}{n}u_t^{'}u_t\\ &+ \alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1\Leftrightarrow\end{aligned}$$

$$\frac{\left\|1_n-R_t\hat{\theta}_{\alpha_t}\right\|^2}{n}+\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1\le\frac{u_t^{'}u_t}{n}+\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1$$

and using the fact that $u_t=1_n-R_t\theta_t$, we obtain that $\frac{u_t^{'}u_t}{n}+\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1=\frac{\|1_n-R_t\theta_t\|^2}{n}+\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1$, hence,

$$\frac{\left\|1_n-R_t\hat{\theta}_{\alpha_t}\right\|^2}{n}+\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1\le\frac{\|1_n-R_t\theta_t\|^2}{n}+\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1\qquad(77)$$

which is always true because we have that

$$\hat{\theta}_{\alpha_t}=\arg\min_{\theta_t}\frac{\|1_n-R_t\theta_t\|^2}{n}+\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1$$

Therefore, (76) and (77) are equivalent. Then, using (76), we have that

$$\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|^2}{n}\le\frac{2u_t^{'}R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)}{n}+\alpha_t\left[\left\|\theta_t-\theta_{t-1}\right\|_1-\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1\right]$$

Since, $\left\|\theta_t-\theta_{t-1}\right\|_1-\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1\le\left|\left\|\theta_t-\theta_{t-1}\right\|_1-\left\|\hat{\theta}_{\alpha_t}-\theta_{t-1}\right\|_1\right|\le\left\|\hat{\theta}_{\alpha_t}-\theta_t\right\|_1$ we have that,

$$\begin{aligned}\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|^2}{n} &\le \frac{2u_t^{'}R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)}{n}+\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_t\right\|_1\\ &\le \left\{\max_{1\le j\le N}2\left|u_j^{'}R_t^{(j)}\right|/n\right\}\left\|\hat{\theta}_{\alpha_t}-\theta_t\right\|_1+\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_t\right\|_1\\ &\le \left[\max_{1\le j\le N}2\left|u_j^{'}R_t^{(j)}\right|/n+\alpha_t\right]\left\|\hat{\theta}_{\alpha_t}-\theta_t\right\|_1.\end{aligned}$$

And using B(i), we obtain that

$$\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|^2}{n}\le\frac{3}{2}\alpha_t\left\|\hat{\theta}_{\alpha_t}-\theta_t\right\|_1.$$

Let's now look at $\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\|_1$.

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\|_1 = \left\|\hat{\theta}_{\alpha_t} - \theta_{t-1} + \theta_{t-1} - \theta_t\right\|_1$$
$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\|_1 \leq \left\|\hat{\theta}_{\alpha_t} - \theta_{t-1}\right\|_1 + \|\theta_t - \theta_{t-1}\|_1$$

by triangular inequality. Hence,

$$\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|^2}{n} \leq \frac{3}{2}\alpha_t \left\|\hat{\theta}_{\alpha_t} - \theta_{t-1}\right\|_1 + \frac{3}{2}\alpha_t \|\theta_t - \theta_{t-1}\|_1 .$$

Moreover, we have that $\hat{\theta}_{\alpha_t} = \hat{\theta}_{\alpha_t}^{S_t} + \hat{\theta}_{\alpha_t}^{S_t^c}$ which implies that

$$\left\|\hat{\theta}_{\alpha_t} - \theta_{t-1}\right\|_1 = \left\|\hat{\theta}_{\alpha_t}^{S_t} + \hat{\theta}_{\alpha_t}^{S_t^c} - \theta_{t-1}\right\|_1 \leq \left\|\hat{\theta}_{\alpha_t}^{S_t} - \theta_{t-1}\right\|_1 + \left\|\hat{\theta}_{\alpha_t}^{S_t^c}\right\|_1$$

by triangular inequality. And using Assumption B(ii) we obtain that

$$\begin{aligned}\left\|\hat{\theta}_{\alpha_t} - \theta_{t-1}\right\|_1 &\leq \tilde{\kappa}_t \left\|\hat{\theta}_{\alpha_t}^{S_t} - \theta_{t-1}\right\|_1 \\ &\leq \tilde{\kappa}_t \frac{\sqrt{s_t}}{\xi_{\hat{\Omega}_t}} \frac{\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|}{\sqrt{n}}\end{aligned}$$

see Bühlmann and Van De Geer (2011) p.105-106 for more details about the last inequality. Therefore,

$$\frac{3}{2}\alpha_t \left\|\hat{\theta}_{\alpha_t} - \theta_{t-1}\right\|_1 \leq \frac{3}{2}\alpha_t\tilde{\kappa}_t \frac{\sqrt{s_t}}{\xi_{\hat{\Omega}_t}} \frac{\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|}{\sqrt{n}}.$$

Using the fact that $4uv \leq u^2 + 4v^2$, we have that,

$$\frac{3}{2}\alpha_t\tilde{\kappa}_t \frac{\sqrt{s_t}}{\xi_{\hat{\Omega}_t}} \frac{\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|}{\sqrt{n}} \leq \frac{1}{4}\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|^2}{n} + \frac{9}{4}\alpha_t^2\tilde{\kappa}_t^2 \frac{s_t}{\xi_{\hat{\Omega}_t}^2}$$

which implies,

$$\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|^2}{n} \leq \frac{1}{4}\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|^2}{n} + \frac{9}{4}\alpha_t^2\tilde{\kappa}_t^2 \frac{s_t}{\xi_{\hat{\Omega}_t}^2} + \frac{3}{2}\alpha_t \|\theta_t - \theta_{t-1}\|_1 \Rightarrow$$

$$\frac{3}{4}\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|^2}{n} \leq \frac{9}{4}\alpha_t^2\tilde{\kappa}_t^2 \frac{s_t}{\xi_{\hat{\Omega}_t}^2} + \frac{3}{2}\alpha_t \|\theta_t - \theta_{t-1}\|_1 \Rightarrow$$

$$\frac{\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|^2}{n} \leq 3\alpha_t^2\tilde{\kappa}_t^2 \frac{s_t}{\xi_{\hat{\Omega}_t}^2} + 2\alpha_t \|\theta_t - \theta_{t-1}\|_1 .$$

Therefore,

$$\frac{1}{n}E\left[\left\|R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|^2\right] = O\left[\alpha_t^2\left(s_t/\xi_{\Omega_t}^2\right) + \alpha_t \|\theta_t - \theta_{t-1}\|_1\right] . \Rightarrow$$

$$\frac{1}{nN}E\left[\left\|\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t}-\omega_t\right)\right\|_2^2\right]=O\left[N\alpha_t^2\left(s_t/\xi_{\Omega_t}^2\right)+N\alpha_t\left\|\theta_t-\theta_{t-1}\right\|_1\right].$$

9.7 Proof of proposition 6

We need some intermediate results to show this proposition.

Proposition 6.1 *Given the set of information $\mathcal{F}_t$ and under assumptions A, and B, we have the following result*

$$\sqrt{n}\delta'\left(\hat{\omega}_{\alpha_t}-\omega_t\right)=\frac{\gamma_t\sqrt{n}\delta'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)}{\left(1-\mu'\theta_t\right)}+O_p\left[\sqrt{nN}\left(\frac{1}{\alpha_t\sqrt{n}}+\left\|\theta_{\alpha_t}-\theta_t\right\|\right)^2\right]$$

if $\max_{0\leq j\leq t-1}\left\{\frac{N^{1/2}}{\alpha_j\sqrt{n}}+\sqrt{N}\alpha_j^{min(\frac{\tau_j}{2},1)}\right\}\to 0$ *as* $n\to\infty$

Proposition 6.1 implies that under some regularity conditions $\delta'\hat{\omega}_{\alpha_t}$ and $\frac{\gamma_t\delta'\hat{\theta}_{\alpha_t}}{(1-\mu'\theta_t)}$ may have the same asymptotic distribution. Hence, in this situation, we need only to derive the asymptotic distribution of $\frac{\gamma_t\delta'\hat{\theta}_{\alpha_t}}{(1-\mu'\theta_t)}$ which depends only on the asymptotic distribution of $\delta'\hat{\theta}_{\alpha_t}$.

9.7.1 Proof of proposition 6.1

$$B=\hat{\gamma}_t^{-1}\delta'\hat{\omega}_t-\gamma_t^{-1}\delta'\omega_t=\frac{\delta'\hat{\theta}_t}{1-\hat{\mu}_t'\hat{\theta}_t}-\frac{\delta'\theta_t}{1-\mu'\theta_t}$$

$$\begin{aligned}\frac{1}{\left(1-\hat{\mu}_t'\hat{\theta}_{\alpha_t}\right)} &\equiv \frac{1}{1-\hat{\beta}}\simeq\frac{1}{1-\beta}+\frac{1}{(1-\beta)^2}\left(\beta-\hat{\beta}\right)\\ &= \frac{1}{1-\mu'\theta_t}-\frac{\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)}{\left(1-\mu'\theta_t\right)^2}+o\left(\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right)\end{aligned}$$

since $\mu'\theta_t\in(0,1)$. We then obtain that

$$\begin{aligned}B &= \frac{\delta'\hat{\theta}_{\alpha_t}-\delta'\theta_t}{1-\mu'\theta_t}-\frac{\delta'\hat{\theta}_{\alpha_t}\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)}{\left(1-\mu'\theta_t\right)^2}+O_p\left[\delta'\hat{\theta}_{\alpha_t}\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right]\\ &= \frac{\delta'\hat{\theta}_{\alpha_t}-\delta'\theta_t}{1-\mu'\theta_t}+O_p\left[\delta'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right].\end{aligned}$$

Since we assume that $\|\delta\|=O(1)$

$$\left\|\delta'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|=O\left(\left\|\hat{\theta}_{\alpha_t}-\theta_t\right\|\right).$$

Because,

$$\left\|\hat{\theta}_{\alpha_t}-\theta_t\right\|=O_p\left(\frac{1}{\alpha_t\sqrt{n}}+\left\|\theta_{\alpha_t}-\theta_t\right\|\right)$$

we have that,

$$\left\|\delta'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|=O_p\left(\frac{1}{\alpha_t\sqrt{n}}+\left\|\theta_{\alpha_t}-\theta_t\right\|\right).$$

Using Proof of Lemma 1, we have the following result

$$\left\|\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\| = O_p\left(\frac{\sqrt{N}}{\alpha_t\sqrt{n}}+\sqrt{N}\left\|\theta_{\alpha_t}-\theta_t\right\|\right)$$

Hence, using those two relations we obtain that

$$O_p\left[\delta'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right] = O_p\left[\sqrt{N}\left(\frac{1}{\alpha_t\sqrt{n}}+\left\|\theta_{\alpha_t}-\theta_t\right\|\right)^2\right]$$

Using Lemma 3 we obtain that

$$\gamma_t\sqrt{n}B \approx \sqrt{n}\delta'\left(\hat{\omega}_{\alpha_t}-\omega_t\right) = \frac{\gamma_t\sqrt{n}\delta'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)}{(1-\mu'\theta_t)} + O_p\left[\sqrt{nN}\left(\frac{1}{\alpha_t\sqrt{n}}+\left\|\theta_{\alpha_t}-\theta_t\right\|\right)^2\right].$$

Using assumption C, we obtain the following Lemma based on the standard central limit theorem.

Lemma 4 *Under assumption C, we have the following result*

$$\left\langle\sqrt{n}\left[\hat{E}\left(R_t'1_n\right)-\hat{\Omega}_t\theta_t\right],\delta\right\rangle \to_d \mathcal{N}\left(0,E\left[\delta' r_i r_i' u_i^2\delta\right]\right)$$

9.7.2 Proof of Lemma 4

In fact, $\hat{E}\left(R_t'1_n\right) = \frac{R_t'1_n}{n}$ with $1_n = R_t\theta_t + u_t$. This implies that,

$$\begin{aligned}\hat{E}\left(R_t'1_n\right) &= \frac{R_t'}{n}\left(R_t\theta_t+u_t\right)\\ &= \frac{R_t'R_t}{n}\theta_t+\frac{R_t'u_t}{n}\\ &= \hat{\Omega}_t\theta_t+\frac{R_t'u_t}{n}\end{aligned}$$

Hence,

$$\begin{aligned}\left\langle\sqrt{n}\left[\hat{E}\left(R_t'1_n\right)-\hat{\Omega}_t\theta_t\right],\delta\right\rangle &= \left\langle\frac{R_t'u_t}{\sqrt{n}},\delta\right\rangle\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta' r_i u_i.\end{aligned}$$

Therefore, using assumption C, the standard central limit theorem can be applied to obtain the result of lemma 4.

Proposition 6.2 *Given the set of information $\mathcal{F}_t$ and under assumptions A and C, we have the following result*

$$\frac{\left\langle\sqrt{n}\left[\hat{\theta}_{\alpha_t}-\theta_t\right],\delta\right\rangle}{\left\|\left(E\left[\delta' r_i r_i' u_i^2\delta\right]\right)^{1/2}\hat{\Omega}_{\alpha_t}^{-1}\right\|} \to_d \mathcal{N}(0,1)$$

if $\max\left(\sqrt{N},\alpha_t\frac{\sqrt{n}}{\sqrt{N}}\right)\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)} \to 0$ *as n goes to infinity.*

9.7.3 Proof of proposition 6.2

$$\begin{aligned}
\hat{\theta}_{\alpha_t} - \theta_t &= \hat{\theta}_{\alpha_t} - \theta_{\alpha_t} + \theta_{\alpha_t} - \theta_t \\
&= \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \frac{R_n^{'} 1_n}{n} - \Omega_{\alpha_t}^{-1} \Omega_t \theta_t + \theta_{\alpha_t} - \theta_t \\
&= \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \frac{R_n^{'} 1_n}{n} - \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \hat{\Omega}_t \theta_t + \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \hat{\Omega}_t \theta_t - \Omega_{\alpha_t}^{-1} \Omega_t \theta_t + \theta_{\alpha_t} - \theta_t \\
&= \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\frac{R_n^{'} 1_n}{n} - \hat{\Omega}_t \theta_t\right] + \left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \hat{\Omega}_t - \Omega_{\alpha_t}^{-1} \Omega_t\right] \theta_t + \theta_{\alpha_t} - \theta_t
\end{aligned}$$

$$\delta^{'}\left(\hat{\theta}_{\alpha_t} - \theta_t\right) = \delta^{'}\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\frac{R_n^{'} 1_n}{n} - \hat{\Omega}_t \theta_t\right] + \delta^{'}\left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \hat{\Omega}_t - \Omega_{\alpha_t}^{-1} \Omega_t\right] \theta_t + \delta^{'}\left(\theta_{\alpha_t} - \theta_t\right)$$

$$\begin{aligned}
\left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \hat{\Omega}_t - \Omega_{\alpha_t}^{-1} \Omega_t\right] \theta_t &= \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left\{\hat{\Omega}_t - \Omega_t\right\} \theta_t + \left[\left(\hat{\Omega}_{\alpha_t}\right)^{-1} - \Omega_{\alpha_t}^{-1}\right] \Omega_t \theta_t \\
&= \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left\{\hat{\Omega}_t - \Omega_t\right\} \theta_t + \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\Omega_{\alpha_t} - \hat{\Omega}_{\alpha_t}\right] \underbrace{\Omega_{\alpha_t}^{-1} \Omega_t \theta_t}_{\theta_{\alpha_t}} \\
&= \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left\{\hat{\Omega}_t - \Omega_t\right\} \theta_t + \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\Omega_{\alpha_t} - \hat{\Omega}_{\alpha_t}\right] \theta_{\alpha_t} \\
&= \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left\{\hat{\Omega}_t - \Omega_t\right\} \theta_t + \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\Omega_t - \hat{\Omega}_t\right] \theta_{\alpha_t} \\
&= \left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\Omega_t - \hat{\Omega}_t\right] \left(\theta_{\alpha_t} - \theta_t\right)
\end{aligned}$$

$$\delta^{'}\left(\hat{\theta}_{\alpha_t} - \theta_t\right) = \delta^{'}\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\frac{R_n^{'} 1_n}{n} - \hat{\Omega}_t \theta_t\right] + \delta^{'}\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\Omega_t - \hat{\Omega}_t\right] \left(\theta_{\alpha_t} - \theta_t\right) + \delta^{'}\left(\theta_{\alpha_t} - \theta_t\right)$$

$$\sqrt{n}\delta^{'}\left(\hat{\theta}_{\alpha_t} - \theta_t\right) = \sqrt{n}\delta^{'}\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\frac{R_n^{'} 1_n}{n} - \hat{\Omega}_t \theta_t\right] + \sqrt{n}\delta^{'}\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\Omega_t - \hat{\Omega}_t\right] \left(\theta_{\alpha_t} - \theta_t\right) + \sqrt{n}\delta^{'}\left(\theta_{\alpha_t} - \theta_t\right)$$

$$\begin{aligned}
\frac{\sqrt{n}\delta^{'}\left(\hat{\theta}_{\alpha_t} - \theta_t\right)}{\left\|\left(E\left[\delta' r_i r_i^{'} u_i^2 \delta\right]\right)^{1/2} \hat{\Omega}_{\alpha_t}^{-1}\right\|} &= \underbrace{\frac{\sqrt{n}\delta^{'}\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\frac{R_n^{'} 1_n}{n} - \hat{\Omega}_t \theta_t\right]}{\left\|\left(E\left[\delta' r_i r_i^{'} u_i^2 \delta\right]\right)^{1/2} \hat{\Omega}_{\alpha_t}^{-1}\right\|}}_{(a)} \\
&+ \underbrace{\frac{\sqrt{n}\delta^{'}\left(\hat{\Omega}_{\alpha_t}\right)^{-1} \left[\Omega_t - \hat{\Omega}_t\right] \left(\theta_{\alpha_t} - \theta_t\right)}{\left\|\left(E\left[\delta' r_i r_i^{'} u_i^2 \delta\right]\right)^{1/2} \hat{\Omega}_{\alpha_t}^{-1}\right\|}}_{(b)} \\
&+ \underbrace{\frac{\sqrt{n}\delta^{'}\left(\theta_{\alpha_t} - \theta_t\right)}{\left\|\left(E\left[\delta' r_i r_i^{'} u_i^2 \delta\right]\right)^{1/2} \hat{\Omega}_{\alpha_t}^{-1}\right\|}}_{(c)}.
\end{aligned}$$

By assumption C and using Lemma 4, we have that,

$$(a) \to_d \mathcal{N}(0,1)$$

$$\|(b)\| = \frac{1}{\left\|\left(E\left[\delta' r_i r_i' u_i^2 \delta\right]\right)^{1/2} \hat{\Omega}_{\alpha_t}^{-1}\right\|} \left\|\sqrt{n}\delta'\left(\hat{\Omega}_{\alpha_t}\right)^{-1}\left[\Omega_t - \hat{\Omega}_t\right](\theta_{\alpha_t} - \theta_t)\right\|$$

$$\left\|\sqrt{n}\delta'\left(\hat{\Omega}_{\alpha_t}\right)^{-1}\left[\Omega_t - \hat{\Omega}_t\right](\theta_{\alpha_t} - \theta_t)\right\| = O_p\left(\sqrt{n}\left\|\left(\hat{\Omega}_{\alpha_t}\right)^{-1}\right\|\left\|\Omega_t - \hat{\Omega}_t\right\|\|\theta_{\alpha_t} - \theta_t\|\right)$$

$$\|(b)\| = O_p\left(\frac{\sqrt{n}\left\|\Omega_t - \hat{\Omega}_t\right\|\|\theta_{\alpha_t} - \theta_t\|}{\left(E\left[\delta' r_i r_i' u_i^2 \delta\right]\right)^{1/2}}\right).$$

Since, $\left(E\left[\delta' r_i r_i' u_i^2 \delta\right]\right)^{1/2}$ is of order of $N^{1/2}$ then,

$$\|(b)\| = O_p\left(\sqrt{N}\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)}\right)$$

$$\|(c)\| = O_p\left(\alpha_t\frac{\sqrt{n}}{\sqrt{N}}\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)}\right).$$

Therefore, if $\max\left(\sqrt{N}, \alpha_t\frac{\sqrt{n}}{\sqrt{N}}\right)\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)} \to 0$, we obtain the result of proposition 6.2.

Combining the result of lemma 4 with proposition 6.1 and 6.2, we obtain the asymptotic distribution of $\delta'\hat{\omega}_{\alpha_t}$.

9.8 Proof of Proposition 7

We start by the fact that

$$\hat{\omega}_t - \omega_t = \frac{\gamma_t\left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)} + \frac{\gamma_t\theta_t\mu'\left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)^2} + O_p\left[\left(\hat{\theta}_t - \theta_t\right)'\mu'\left(\hat{\theta}_t - \theta_t\right)\right] + O_p\left[\frac{1}{\sqrt{n}} + \sqrt{\frac{N}{n}}\left\|\hat{\theta}_t - \theta_t\right\|\right].$$

This result is obtained using a similar decomposition as in Carrasco et al. (2019) combined with Lemma 3. Hence, we obtain that

$$\begin{aligned}
(\hat{\omega}_t - \omega_t)'\mu\mu'(\hat{\omega}_t - \omega_t) &= \frac{\gamma_t^2\left(\hat{\theta}_t - \theta_t\right)'\mu\mu'\left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)^2} + \frac{\gamma_t^2\left(\hat{\theta}_t - \theta_t\right)'\mu\theta_t'\mu\mu'\theta_t\mu'\left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)^4} \\
&+ 2\frac{\gamma_t^2\left(\hat{\theta}_t - \theta_t\right)'\mu\mu'\theta_t\mu'\left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)^3} + O_p\left[\left(\hat{\theta}_t - \theta_t\right)'\mu\mu'\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{N}{n}}\left\|\hat{\theta}_t - \theta_t\right\|\right)\right] \\
&+ O_p\left[\left(\hat{\theta}_t - \theta_t\right)'\mu\mu'\left(\hat{\theta}_t - \theta_t\right)'\mu'\left(\hat{\theta}_t - \theta_t\right)\right].
\end{aligned}$$

We know that $\mu^{'}\theta_t = \theta^{'}\mu$ and using the assumption A, we obtain that

$$
\begin{aligned}
(\hat{\omega}_t - \omega_t)^{'}\mu\mu^{'}(\hat{\omega}_t - \omega_t) &= \frac{\gamma_t^2\left(\hat{\theta}_t - \theta_t\right)^{'}\mu\mu^{'}\left(\hat{\theta}_t - \theta_t\right)}{\left(1-\mu^{'}\theta_t\right)^2} + \frac{\gamma_t^2\left(\mu^{'}\theta_t\right)^2\left(\hat{\theta}_t - \theta_t\right)^{'}\mu\mu^{'}\left(\hat{\theta}_t - \theta_t\right)}{\left(1-\mu^{'}\theta_t\right)^4} \\
&+ 2\frac{\gamma_t^2\left(\mu^{'}\theta_t\right)\left(\hat{\theta}_t - \theta_t\right)^{'}\mu\mu^{'}\left(\hat{\theta}_t - \theta_t\right)}{\left(1-\mu^{'}\theta_t\right)^3} + rest\left(\alpha_t\right)
\end{aligned}
$$

where

$$
rest\left(\alpha_t\right) = O_p\left[\left(\hat{\theta}_t - \theta_t\right)^{'}\mu\mu^{'}\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{N}{n}}\left\|\hat{\theta}_t - \theta_t\right\|\right)\right] + O_p\left[\left(\hat{\theta}_t - \theta_t\right)^{'}\mu\mu^{'}\left(\hat{\theta}_t - \theta_t\right)^{'}\mu^{'}\left(\hat{\theta}_t - \theta_t\right)\right]
$$

$$
(\hat{\omega}_t - \omega_t)^{'}\mu\mu^{'}(\hat{\omega}_t - \omega_t) = P_t\left(\hat{\theta}_t - \theta_t\right)^{'}\mu\mu^{'}\left(\hat{\theta}_t - \theta_t\right) + rest\left(\alpha_t\right)
$$

with

$$
\begin{aligned}
P_t &= \gamma_t^2\left[\frac{1}{\left(1-\mu^{'}\theta_t\right)^2} + \frac{2\left(\mu^{'}\theta_t\right)}{\left(1-\mu^{'}\theta_t\right)^3} + \frac{\left(\mu^{'}\theta_t\right)^2}{\left(1-\mu^{'}\theta_t\right)^4}\right] \\
&= \frac{\gamma_t^2}{\left(1-\mu^{'}\theta_t\right)^4}\left[\left(1-\mu^{'}\theta_t\right)^2 + 2\left(\mu^{'}\theta_t\right)\left(1-\mu^{'}\theta_t\right) + \left(\mu^{'}\theta_t\right)^2\right] \\
&= \frac{\gamma_t^2}{\left(1-\mu^{'}\theta_t\right)^4}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
(\hat{\omega}_t - \omega_t)^{'}\mu\mu^{'}(\hat{\omega}_t - \omega_t) &= \frac{\gamma_t^2}{\left(1-\mu^{'}\theta_t\right)^4}\left(\hat{\theta}_t - \theta_t\right)^{'}\mu\mu^{'}\left(\hat{\theta}_t - \theta_t\right) + rest\left(\alpha_t\right) \\
&= \frac{\gamma_t^2}{\left(1-\mu^{'}\theta_t\right)^4}\left\|\mu^{'}\left(\hat{\theta}_t - \theta_t\right)\right\|^2 + rest\left(\alpha_t\right) \\
\frac{\left(1-\mu^{'}\theta_t\right)^4}{\gamma_t^2}(\hat{\omega}_t - \omega_t)^{'}\mu\mu^{'}(\hat{\omega}_t - \omega_t) &= \left\|\mu^{'}\left(\hat{\theta}_t - \theta_t\right)\right\|^2 + rest\left(\alpha_t\right) \\
&= \frac{1}{n}\left\|1_n^{'}R_t\left(\hat{\theta}_t - \theta_t\right)\right\|^2 + rest\left(\alpha_t\right).
\end{aligned}
$$

Therefore,

$$
\left(1-\mu^{'}\theta_t\right)^4\gamma_t^{-2}E\left[(\hat{\omega}_t - \omega_t)^{'}\mu\mu^{'}(\hat{\omega}_t - \omega_t)\right] = \frac{1}{n}E\left[\left\|1_n^{'}R_t\left(\hat{\theta}_t - \theta_t\right)\right\|^2\right] + rest\left(\alpha_t\right)
$$

Let's now look at the properties of $rest\left(\alpha_t\right)$. Recall that

$$
rest\left(\alpha_t\right) = \underbrace{O_p\left[\left(\hat{\theta}_t - \theta_t\right)^{'}\mu\mu^{'}\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{N}{n}}\left\|\hat{\theta}_t - \theta_t\right\|\right)\right]}_{(k1)} + \underbrace{O_p\left[\left(\hat{\theta}_t - \theta_t\right)^{'}\mu\mu^{'}\left(\hat{\theta}_t - \theta_t\right)^{'}\mu^{'}\left(\hat{\theta}_t - \theta_t\right)\right]}_{(k2)}.
$$

$$
\begin{aligned}
(k2) &= O_p\left(\left\|\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|^3\right) \\
&= O_p\left[\sqrt{N}\left(\frac{1}{\alpha_t\sqrt{n}}+\alpha_t^{min(\frac{\tau_t}{2},1)}\right)\right].
\end{aligned}
$$

The last quantity is obtained using the proof of Lemma 4 in Carrasco et al. (2019).

$$
\begin{aligned}
(k1) &= O_p\left[\left(\hat{\theta}_t-\theta_t\right)'\mu\mu'\left(\frac{1}{\sqrt{n}}+\sqrt{\frac{N}{n}}\left\|\hat{\theta}_t-\theta_t\right\|\right)\right] \\
&= O_p\left[\left\|\mu'\left(\hat{\theta}_{\alpha_t}-\theta_t\right)\right\|\|\mu\|\left(\frac{1}{\sqrt{n}}+\sqrt{\frac{N}{n}}\left\|\hat{\theta}_t-\theta_t\right\|\right)\right].
\end{aligned}
$$

$\|\mu\| = O\left(\sqrt{N}\right)$. Moreover, in the proof of Lemma 1, we obtain that

$$
\left\|\hat{\theta}_t-\theta_t\right\| = O_p\left(\frac{1}{\alpha_t\sqrt{n}}+\alpha_t^{min(\frac{\tau_t}{2},1)}\right).
$$

Hence,

$$
(k1) = O_p\left[\left(\frac{\sqrt{N}}{\alpha_t\sqrt{n}}+\sqrt{N}\alpha_t^{min(\frac{\tau_t}{2},1)}\right)\left(\frac{\sqrt{N}}{\sqrt{n}}+\frac{N}{\alpha_t n}+\frac{N}{\sqrt{n}}\alpha_t^{min(\frac{\tau_t}{2},1)}\right)\right].
$$

Therefore,

$$
\begin{aligned}
rest\left(\alpha_t\right) &= O_p\left[\left(\frac{\sqrt{N}}{\alpha_t\sqrt{n}}+\sqrt{N}\alpha_t^{min(\frac{\tau_t}{2},1)}\right)\left(\frac{\sqrt{N}}{\sqrt{n}}+\frac{N}{\alpha_t n}+\frac{N}{\sqrt{n}}\alpha_t^{min(\frac{\tau_t}{2},1)}\right)\right] \\
&+ O_p\left[\sqrt{N}\left(\frac{1}{\alpha_t\sqrt{n}}+\alpha_t^{min(\frac{\tau_t}{2},1)}\right)\right].
\end{aligned}
$$

10 Tables and Figures

Table 2: Information about the 5 industry portfolios

Code	Composition of the sector
Cnsmr	Consumer Durables, NonDurables, Wholesale, Retail, and Some Services (Laundries, Repair Shops)
Manuf	Manufacturing, Energy, and Utilities
Hitec	Business Equipment, Telephone and Television Transmission
Hlth	Healthcare, Medical Equipment, and Drugs
Other	Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, Finance

Table 3: Out-of sample performance with an estimation window of 120 for 5 and 10 industry portfolios

	5 Industry Portfolios			10 Industry Portfolios		
	Risk	Return per unit of risk	Turnover	Risk	Return per unit of risk	Turnover
Sample based strategy	0.0509	0.0936	2.4937	0.0515	0.0747	2.7224
Bauder et al bayesian strategy	0.0451	0.1177	1.0866	0.0465	0.1107	1.2841
Approximation of the true solution	0.0441	0.1276	0.9812	0.0417	0.1765	1.0034

Table 4: Out-of sample performance with an estimation window of 120 for 17 and 30 industry portfolios

	17 Industry Portfolios			30 Industry Portfolios		
	Risk	Return per unit of risk	Turnover	Risk	Return per unit of risk	Turnover
Sample based strategy	0.5332	0.0462	15.2736	0.2703	0.0620	21.2963
Bauder et al bayesian strategy	0.0552	0.1089	2.3971	0.0726	0.0822	4.4491
Approximation of the true solution	0.0410	0.3152	0.9402	0.0501	0.3536	0.9168

Table 5: The condition number of the sample covariance matrix as a function of the number of assets in the economy. The sample size is given by $n = 120$ over 1000 replications. The investment horizon is $T = 12$. Standard errors of those statistics are given in bracket.

period/N	$\hat{\lambda}_{\max}/\hat{\lambda}_{\min}$						
	10	20	40	60	80	90	100
0	219.3835	622.9837	1629.7	3346.2	15842	27065	72148
	(29.1304)	(76.1754)	(202.0158)	(461.6792)	(3019.2)	(6250.2)	(21181)
2	220.8378	616.1889	1629.6	3376.7	15545	25885	72336
	(29.3274)	(73.7110)	(199.8609)	(459.1104)	(2896.0)	(5797.1)	(21344)
5	221.7853	627.2623	1621.9	3382.3	15233	26193	69306
	(29.3883)	(75.0828)	(200.7195)	(466.2505)	(2844.9)	(5791.3)	(19456)
7	222.2615	641.9389	1621.8	3260.8	15043	26714	68237
	(29.3190)	(75.9786)	(202.5438)	(457.5931)	(2836.5)	(5720.8)	(19483)
9	220.6481	632.5309	1628.0	3271.2	14895	26274	66849
	(29.0634)	(74.0857)	(200.7087)	(461.0739)	(2773.8)	(5662.3)	(19114)
11	218.7989	626.0016	1602.1	3266.8	15451	26382	66520
	(29.1129)	(73.3732)	(193.5361)	(461.2739)	(2840.9)	(5790.8)	(18701)

Table 6: The relative condition number of the sample covariance matrix as a function of the number of assets in the economy. The sample size is given by $n = 120$ over 1000 replications. The investment horizon is $T = 12$. Standard errors of those statistics are given in bracket.

Period/N	$\left(\hat{\lambda}_{\max}/\hat{\lambda}_{\min}\right) / (\lambda_{\max}/\lambda_{\min})$						
	10	20	40	60	80	90	100
0	1.1479	1.6661	2.4038	2.9609	9.8419	16.4828	37.7055
	(0.1524)	(0.2037)	(0.2980)	(0.4085)	(1.8757)	(3.8064)	(11.0693)
2	1.1555	1.16479	2.0437	2.9879	9.6575	15.7643	37.8039
	(0.1534)	(0.1971)	(0.2948)	(0.4062)	(1.7991)	(3.5305)	(11.1548)
5	1.1604	1.6775	2.3923	2.9928	9.4638	15.9516	36.2205
	(0.1538)	(0.2008)	(0.2961)	(0.4126)	(1.7674)	(3.5269)	(10.1681)
7	1.1629	1.7168	2.3922	2.8853	9.3455	16.2693	35.6716
	(0.1534)	(0.2032)	(0.2988)	(0.4049)	(1.7622)	(3.4840)	(10.1819)
9	1.1545	1.6916	2.4012	2.8946	9.2538	16.0013	34.9365
	(0.1521)	(0.1981)	(0.2960)	(0.4080)	(1.7232)	(3.4484)	(9.9892)
11	1.1448	1.6741	2.3630	2.8906	9.5993	16.0667	34.7646
	(0.1523)	(0.1962)	(0.2855)	(0.4082)	(1.7649)	(3.5266)	(9.7735)

Table 7: The average monthly Actual Sharpe ratio from optimal strategies using a three-factor model as a function of the number of assets in the economy with the sample size $n = 120$, the investment horizon given by $T = 12$ over 1000 replications. TSR is the true actual Sharpe ratio.

Strategy/N	10	20	40	60	80	90	100
SbP	0.1218	0.0878	0.0568	0.0341	0.0346	0.0213	0.0093
XoNP	0.1509	0.1554	0.1652	0.1559	0.1638	0.1639	0.1591
RdgP	0.1517	0.1777	0.1626	0.1736	0.1668	0.1800	0.1763
L2TSP	0.1625	0.1830	0.1742	0.1779	0.1706	0.1832	0.1769
L1TSP	0.1640	0.1791	0.1729	0.1838	0.1735	0.1817	0.1789
BP	0.1575	0.1195	0.0816	0.0769	0.0368	0.0266	0.0113
TSR	0.1953	0.1907	0.2028	0.2050	0.2052	0.2056	0.2058

Table 8: The average monthly bias in the Actual Sharpe ratio from optimal strategies using a three-factor model as a function of the number of assets in the economy with the sample size $n = 120$, the investment horizon given by $T = 12$ over 1000 replications.

Strategies	Number of risky assets						
	10	20	40	60	80	90	100
SbP	-0.0735	-0.1029	-0.1460	-0.1709	-0.1706	-0.1843	-0.1965
XoNP	-0.0444	-0.0353	-0.0376	-0.0491	-0.0417	-0.0417	-0.0467
RdgP	-0.0436	-0.013	-0.0402	-0.0314	-0.0384	-0.0256	-0.0295
L2TSP	-0.0313	-0.0077	-0.0286	-0.0271	-0.0346	-0.0224	-0.0289
L1TSP	-0.0313	-0.0116	-0.0299	-0.0212	-0.0317	-0.0239	-0.0269
BP	-0.0378	-0.0712	-0.1212	-0.1281	-0.1684	-0.1790	-0.1945

Table 9: The average monthly default probability from optimal strategies using a three-factor model as a function of the number of assets in the economy with the sample size $n = 120$, the investment horizon given by $T = 12$ over 1000 replications.

Strategy/N	10	20	40	60	80	90	100
SbP	0.0617	0.0763	0.1255	0.0816	0.1283	0.1291	0.1525
RdgP	0.0008	0.0001	0.0082	0.0065	0.0056	0.0000	0.0000
L2TSP	0.0002	0.0030	0.0036	0.0047	0.0065	0.0001	0.0000
L1TSP	0.0001	0.0025	0.0013	0.002	0.0011	0.0000	0.0000
BP	0.0000	0.0111	0.0631	0.0881	0.1133	0.1005	0.1232

Table 10: The average monthly Turnover from optimal strategies using a three-factor model as a function of the number of assets in the economy with the sample size $n = 120$, the investment horizon given by $T = 12$ over 1000 replications.

Strategy/N	10	20	40	60	80	90	100
SbP	8.4517	11.5957	11.7088	13.0296	13.5274	18.9334	21.3356
RdgP	0.7532	0.7689	0.7919	0.8067	0.9140	0.9317	0.989
L2TSP	0.6539	0.656	0.6350	0.7192	0.910	0.8845	0.9352
L1TSP	0.1648	0.1737	0.204	0.2774	0.2943	0.2668	0.2874
BP	0.4161	3.8001	6.1319	6.4807	7.0534	9.0807	9.3808

Table 11: The absolute bias in the optimal wealth using a three-factor model with the sample size $n = 120$, the investment horizon given by $T = 12$ over 1000 replications when $N = 10$.

Periods	Strategies				
	SbP	BP	RdgP	L2TSP	L1TSP
0	0.1538	0.0207	0.0043	0.0093	0.0007
2	0.12830	0.0129	0.0026	0.0031	0.0038
4	1.0379	0.0116	0.0154	0.0226	0.0284
6	0.3918	0.0180	0.0279	0.0246	0.0267
8	4.7723	0.0255	0.0199	0.0100	0.0052
10	9.9473	0.0753	0.0093	0.0010	0.0014
11	22.9089	0.1311	0.0130	0.0021	0.0082

Table 12: The absolute bias in the optimal wealth using a three-factor model with the sample size $n = 120$, the investment horizon given by $T = 12$ over 1000 replications when $N = 20$.

Periods	Strategies				
	SbP	BP	RdgP	L2TSP	L1TSP
0	0.0251	0.0237	0.0024	0.0017	0.0141
2	1.8049	0.3319	0.0071	0.0030	0.0153
4	2.9541	0.8699	0.0561	0.0145	0.0191
6	8.7675	3.0255	0.8404	0.6176	0.4716
8	10.4564	3.7632	0.8057	0.9080	0.7161
10	12.5781	6.9093	1.0164	0.8796	0.7369
11	31.0841	8.6879	1.0744	0.9373	0.8380

Table 13: The absolute bias in the optimal wealth using a three-factor model with the sample size $n = 120$, the investment horizon given by $T = 12$ over 1000 replications when $N = 40$.

Periods	Strategies				
	SbP	BP	RdgP	L2TSP	L1TSP
0	0.0588	0.0299	0.0185	0.0105	0.0077
2	32.5544	0.5531	0.0330	0.0263	0.0199
4	38.8939	1.3677	0.0463	0.0289	0.0679
6	38.5641	2.8210	0.0569	0.0569	0.006
8	57.8871	36.9814	0.0782	0.0625	0.0271
10	65.7681	47.7400	0.0983	0.0639	0.0127
11	153.7881	94.3946	0.1441	0.1181	0.0619

Table 14: The average bias in the actual Sharpe ratio and the average deviation between the true and the estimated portfolio for several sample sizes.

	Sample size			
	120	300	1000	2000
Average bias in the actual Sharpe ratio	-0.0295	-0.0259	-0.0098	-0.0084
Deviation between the estimated strategy and the true one	3.0303	2.9496	2.8107	2.8103

Table 15: The average tuning parameter using a three-factor model with the sample size $n = 120$, the investment horizon given by $T = 12$ over 1000 replications when $N = 40$.

Periods	Strategies		
	RdgP	L2TSP	L1TSP
0	0.0160	0.0188	0.0070
	(0.0123)	(0.0082)	(0.0026)
4	0.0168	0.0170	0.0198
	(0.0125)	(0.0083)	(0.0074)
8	0.0215	0.0178	0.0205
	(0.0120)	(0.0081)	(0.0071)
11	0.0249	0.0183	0.0205
	(0.0104)	(0.0079)	(0.0074)

Table 16: Out-of-sample performance in terms of Sharpe ratio applied on the 30 industry portfolios (FF30) and the 100 portfolios formed on size and book-to-market (FF100) for two different rolling windows.

Portfolios	Estimation Window	Strategies				
		SbP	BP	RdgP	L2TSP	L1TSP
FF30	60	0.0195	0.05195	0.0767	0.0963	0.1836
	120	0.0496	0.0822	0.1715	0.1878	0.1876
FF100	120	0.0569	0.1025	0.1697	0.1996	0.2424
	240	0.0973	0.1550	0.2050	0.2637	0.2837

Table 17: Some statistics on eigenvalues and condition number of the sample covariance matrix of the 30 industry portfolios for two different rolling windows.

Rolling window	Statistics	$\lambda_{\min}$	$\lambda_{\max}$	$\lambda_{\max}/\lambda_{\min}$
60	mean	5.6976E-05	0.0707	1.4073E+03
	std	2.5506E-05	0.0329	813.5847
	median	5.0329E-05	0.0636	1.2490E+03
120	mean	1.5189E-04	0.0696	510.4563
	std	4.553E-05	0.0174	217.2946
	median	1.2842E-04	0.0689	544.0167

Table 18: Some statistics on eigenvalues and condition number of the sample covariance matrix of the 100 industry portfolios for two different rolling windows.

Rolling window	Statistics	$\lambda_{\min}$	$\lambda_{\max}$	$\lambda_{\max}/\lambda_{\min}$
120	mean	3.3854E-06	0.2636	8.3722E+04
	std	9.4620E-07	0.0534	2.7365E+04
	median	3.2609E-06	0.2635	8.2893E+04
240	mean	4.4466E-05	0.2551	5.7491E+03
	std	3.6914E-06	0.0253	497.8610
	median	4.4371E-05	0.2522	5.7621E+03

Table 19: Out-of-sample performance in terms of re-balancing cost (turnover) applied on the 30 industry portfolios (FF30) and the 100 portfolios formed on size and book-to-market (FF100) for two different rolling windows.

Portfolios	Estimation Window	Strategies				
		SbP	BP	RdgP	L2TSP	L1TSP
FF30	60	4.6060	3.6181	1.9035	1.5590	0.2747
	120	2.1302	2.0560	1.770	1.2700	0.1916
FF100	120	7.9407	5.9596	3.9402	1.4065	0.6456
	240	5.6427	3.9562	2.7195	1.2516	0.5744

Table 20: Out-of-sample performance in terms of Sharpe ratio (when transaction costs are included) applied on the 30 industry portfolios (FF30) and the 100 portfolios formed on size and book-to-market (FF100) for two different rolling windows.

Portfolios	Estimation Window	Strategies				
		SbP	BP	RdgP	L2TSP	L1TSP
FF30	60	-0.0784	0.0328	0.0638	0.0783	0.1780
	120	-0.0325	0.0508	0.1610	0.1730	0.1748
FF100	120	-0.0880	0.0848	0.1303	0.1804	0.2301
	240	-0.0330	0.1156	0.1789	0.2484	0.2735

Figure 1: The Sharpe ratio as a function of the tuning parameter for the Ridge. We obtain this figure using a single sample when $N = 60$ and $N = 100$ and $n = 120$.

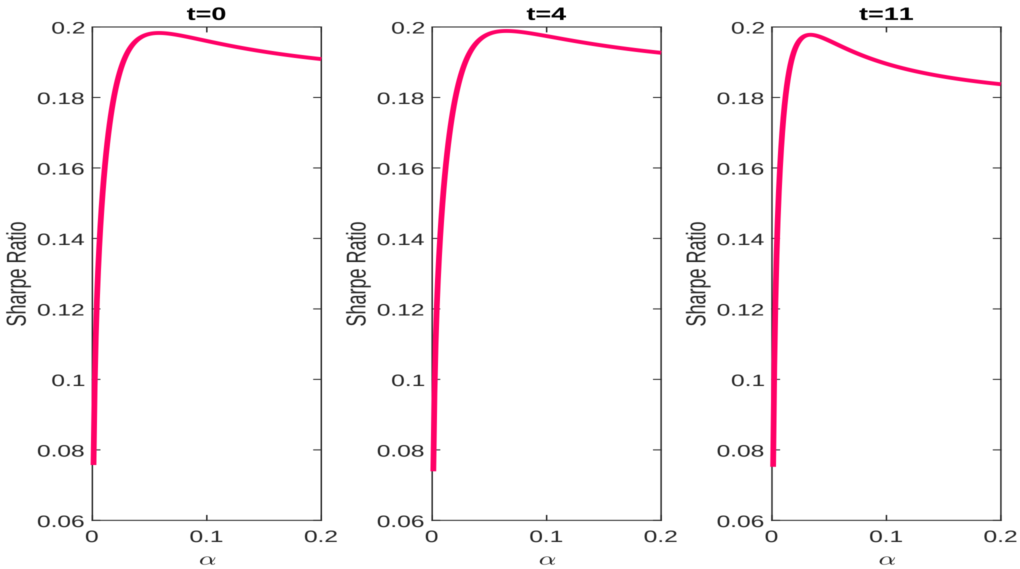

Figure 2: The GCV criterion as a function of the tuning parameter for the Ridge regularization using a single sample when $N = 60$ and $n = 120$.

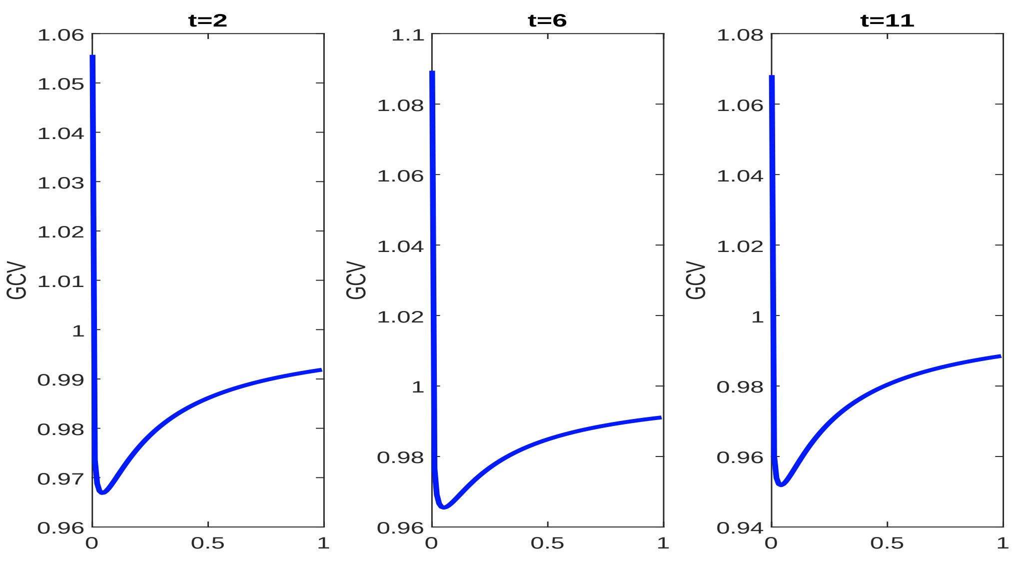

Figure 3: The transaction cost as a function of the tuning parameter for the L1TSP and L2TSP. We obtain this figure for $N = 20$, $T = 12$ with an estimation window of $n = 120$.

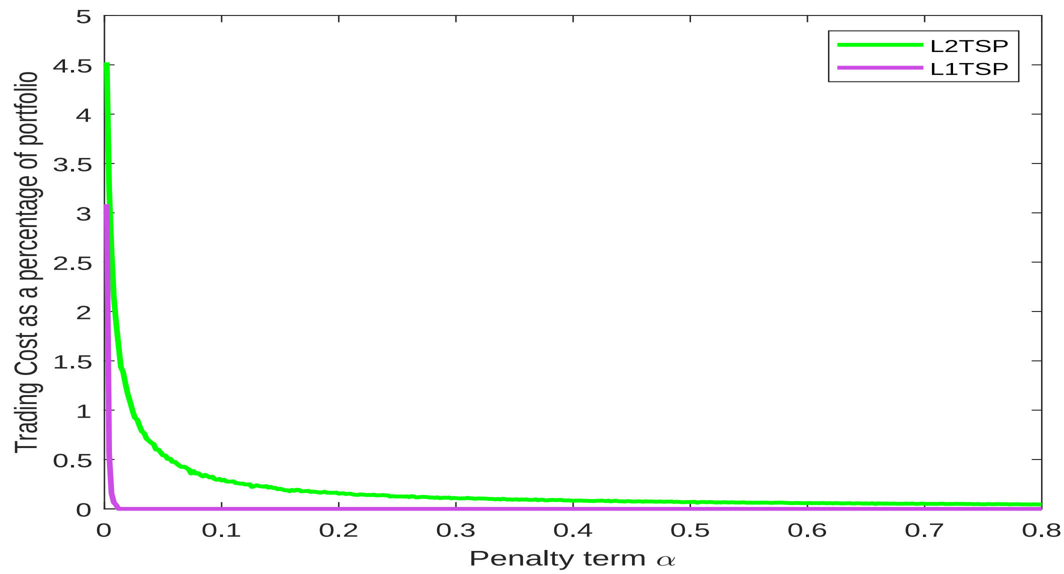

Figure 4: The average stability rate as a function of the tuning parameter for the L1TSP and L2TSP. We obtain this figure for $N = 20$, $T = 12$ with an estimation window of $n = 120$.

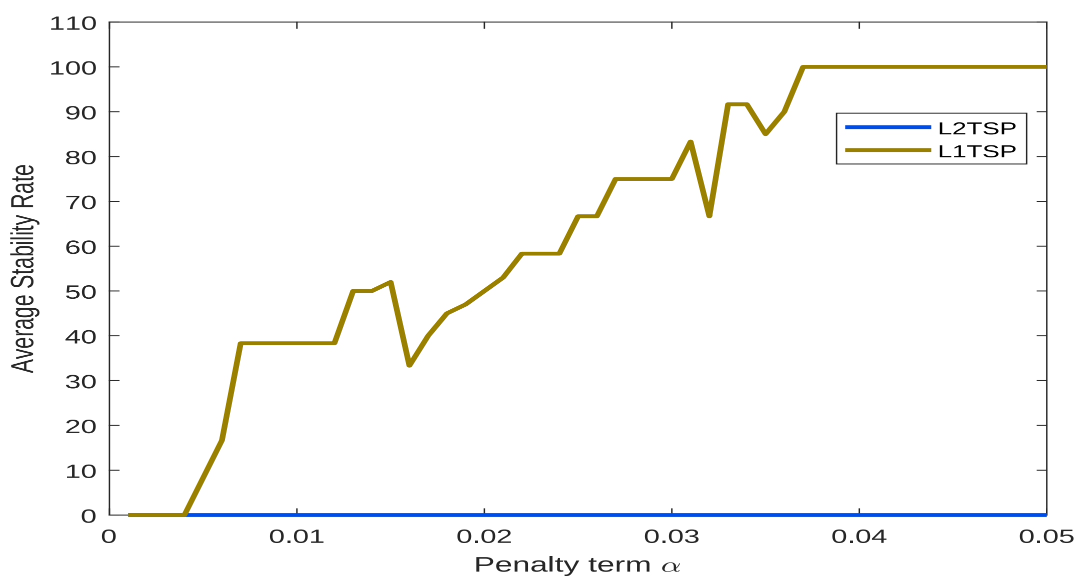

Figure 5: The Average Optimal selected tuning parameter for the RdgP, the L2TSP and the L1TSP over the life cycle when $N = 60$ and $n = 120$.

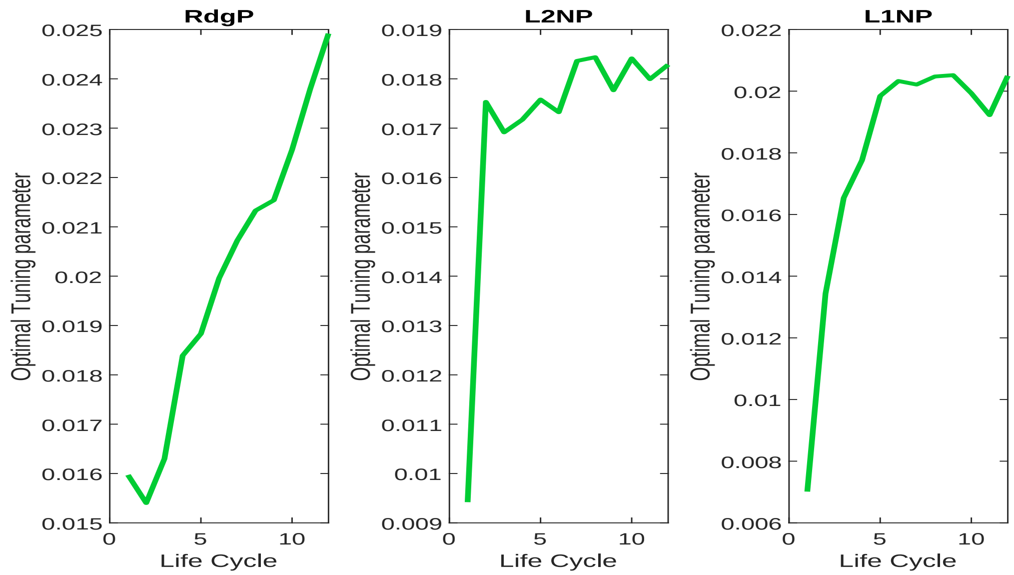

Figure 6: The Mean Squared Error of the selected strategy over the life cycle for the RdgP and the L2TSP with $N = 20$, $T = 24$ month and an estimation window of $n = 120$.

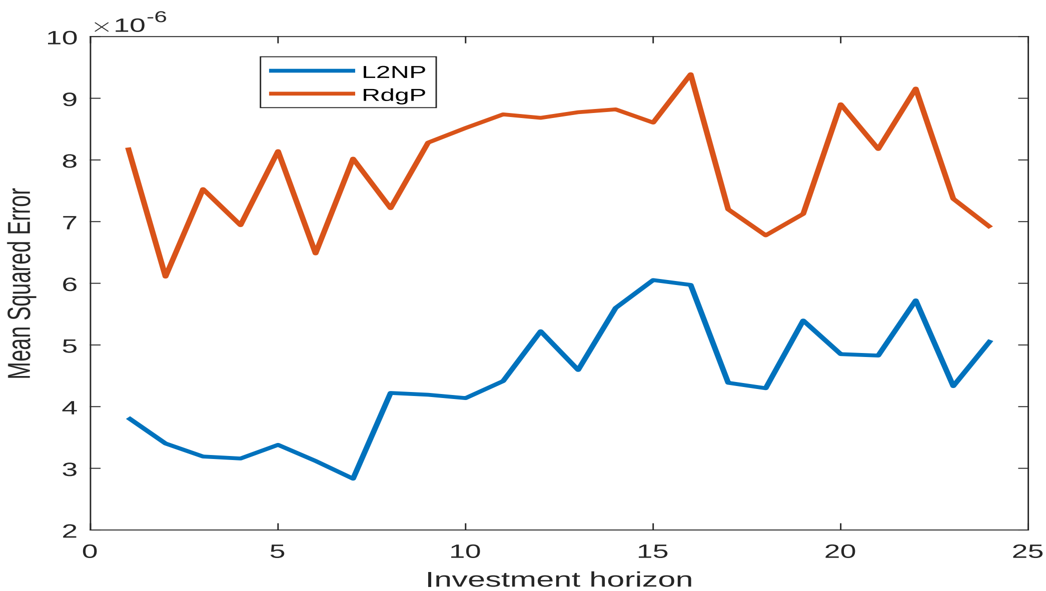

Figure 7: The optimal wealth over the life cycle for our procedures. We obtain this figure using the 30 industry portfolios with an estimation window of $n = 120$.

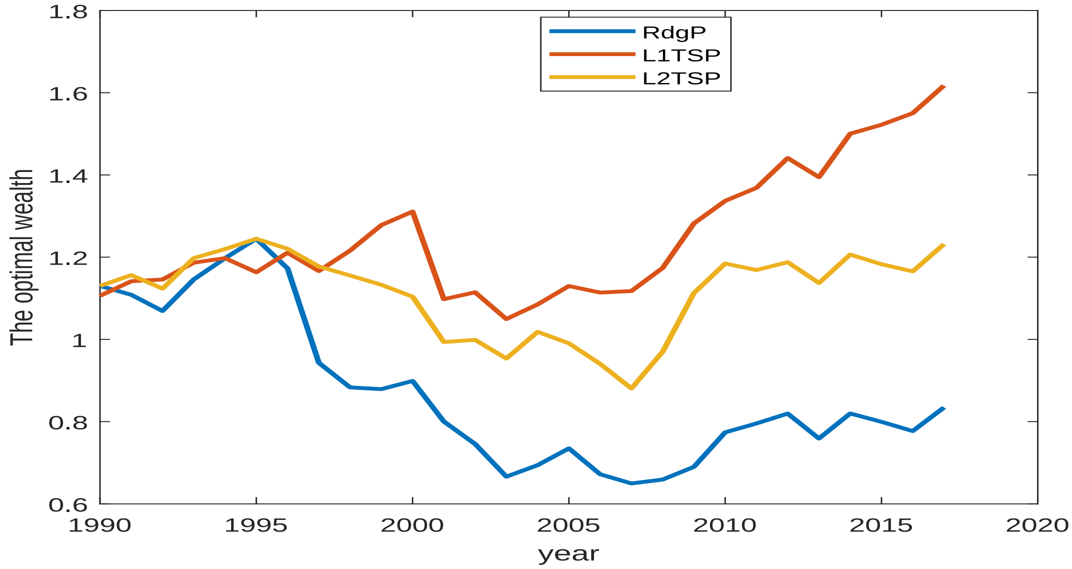

Figure 8: The re-balancing cost over the life cycle. We obtain this figure using the 30 industry portfolios with an estimation window of $n = 120$.